

2mEPSPS PROTEIN

OVERALL AMINO ACID SEQUENCE

HOMOLOGY SEARCH WITH KNOWN

TOXINS AND ALLERGENS

DATA REQUIREMENT
No applicable guidelines

IN SILICO STUDY

AUTHOR: A. CAPT

TESTING FACILITY:

Bayer CropScience
355, rue Dostoïevski
BP 153
06903 Sophia Antipolis Cedex
France

SPONSOR:

Bayer AG
Bayer CropScience
Alfred Nobel Str. 50
40789 Monheim
Germany

STUDY COMPLETED ON: NOVEMBER 27, 2008
PAGE 1 OF 526



M-273792-02-1

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

STATEMENT OF NO DATA CONFIDENTIALITY CLAIM

No claim of confidentiality is made for any information contained in this study on the basis of its falling within the scope of FIFRA § 10 (d) (1) (A), (B) or (C).

Company Name:

Company Agent:

Title:

Signature:

Date: _____

These data are the property of Bayer CropScience, and as such, are considered to be confidential for all purposes other than compliance with FIFRA § 10. Submission of these data in compliance with FIFRA does not constitute a waiver of any right to confidentiality which may exist under any other statute or in any other country.

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

STATEMENT CONCERNING GOOD LABORATORY PRACTICE

The *in silico* search was conducted in the spirit of Good Laboratory Practices (GLP). The final report was not audited by the Quality Assurance Unit.

Company: Bayer CropScience

Submitter:

Author:

Date: November 27, 2008



A. CAPT

Sponsor Representative:

Date: NOVEMBER 27, 2008



C. HEROUET-GUICHENEY

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

FLAGGING STATEMENTS

This page is reserved for flagging statements as may be required by US EPA.

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

SIGNATURE

I, the undersigned, hereby declare that the work was performed under my supervision according to the procedures described and that this report provides a correct and faithful record of the results obtained.

There were no circumstances which affected the quality and integrity of the data.

Author:

Date: November 27, 2008



A. CAPT

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

TABLE OF CONTENTS

STATEMENT OF NO DATA CONFIDENTIALITY CLAIM	2
STATEMENT CONCERNING GOOD LABORATORY PRACTICE	3
FLAGGING STATEMENTS	4
SIGNATURE	5
TABLE OF CONTENTS	6
SUMMARY	8
INTRODUCTION	9
MATERIAL AND METHODS	10
1 - Amino acid query sequence	10
2 - Method	10
2.1 General search design	10
2.2 Protein databases	11
2.3 FASTA algorithm	12
2.4 Amino acid substitution matrices	13
2.5 Gap	13
2.6 Expect value	13
2.7 Filter	14
2.8 Bit score	14
2.9 Matching criterion	14
2.10 Expression of results	14
3 - Data storage	15
RESULTS	16
CONCLUSION	17
REFERENCES	18
ABBREVIATIONS AND ACRONYMS	19
GLOSSARY	21
TABLES	25
Table 1 - FASTA sequence similarities of 2mEPSPS protein with proteins from the Uniprot_Swissprot database	25
Table 2 - FASTA sequence similarities of 2mEPSPS protein with proteins from the Uniprot_TrEMBL database	32
Table 3 - FASTA sequence similarities of 2mEPSPS protein with proteins from the PDB database	38

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Table 4 - FASTA sequence similarities of 2mEPSPS protein with proteins from the DAD database	44
Table 5 - FASTA sequence similarities of 2mEPSPS protein with proteins from the GenPept database	50
APPENDICES	56
Appendix A - Amino acid codes	56
Appendix B - Detailed FASTA sequence similarity alignments of the 2mEPSPS protein with proteins from the Uniprot_Swissprot database	57
Appendix C - Detailed FASTA sequence similarity alignments of the 2mEPSPS protein with proteins from the Uniprot_TrEMBL database	164
Appendix D - Detailed FASTA sequence similarity alignments of the 2mEPSPS protein with proteins from the PDB database	262
Appendix E - Detailed FASTA sequence similarity alignments of the 2mEPSPS protein with proteins from the DAD database	329
Appendix F - Detailed FASTA sequence similarity alignments of the 2mEPSPS protein with proteins from the GenPept database	427
FINAL REPORT AMENDMENT	526

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

SUMMARY

This *in silico* study evaluated the potential amino acid sequence similarity between the double mutant maize 5-enol pyruvylshikimate-3-phosphate synthase (2mEPSPS) protein and known toxins or allergens.

An overall amino acid sequence identity search was carried out using FASTA algorithm by comparing the complete amino acid sequence of the 2mEPSPS protein with all protein sequences present in large general reference databases (Uniprot_Swissprot, Uniprot_TrEMBL, PDB, DAD and GenPept) and in an allergen reference database (AllergenOnline). The criterion indicating potential toxicity or allergenicity was a 35% identity over at least 80 consecutive amino acids with a toxin or allergenic protein.

In addition, an 80-mer sliding window search was performed to compare the query sequence, subdivided into 80 amino acid blocks, with all known allergens present in the allergen reference database AllergenOnline. The criterion indicating potential allergenicity was a 35 % identity on the window of 80 amino acids with an allergenic protein.

No identity was found with known toxins and known allergens. Only high similarities to other EPSPS proteins and to other enzymes from various organisms, which have good safety records were found. Therefore, it is unlikely that the 2mEPSPS protein could exhibit toxic or allergenic properties.

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

INTRODUCTION

One of the key endpoints evaluated for assessing the safety of a novel protein is the amino acid (aa) sequence similarity with known allergens. For this reason, this comparison is performed as part of current food safety evaluation strategies ([FAO/WHO, 2001](#); [CAC, 2003](#)).

The objective of this amino acid sequence homology search was to evaluate the potential amino acid sequence similarity of the 2mEPSPS protein to known toxins or allergens by using an *in silico* approach.

Date of search: November 24, 2008

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

MATERIAL AND METHODS

1 - AMINO ACID QUERY SEQUENCE

The query sequence was coded using the one-letter code adopted by the Commission on Biochemical Nomenclature of the [IUPAC-IUB \(1984\)](#) (see [Appendix A](#)).

As described in the document number M-234186-01-1 ([De Beuckeleer, 2003](#)), the query sequence corresponding to the double mutant maize 5-enol pyruvylshikimate-3-phosphate synthase (2mEPSPS) protein is as follows:

MAGAEIIVLQP	IKEISGTVKL	PGSKSLSNRI	LLLAALSEG	TVVDNLLNSE	DVHYMLGALR	61
TLGLSVEADK	AAKRAVVVGC	GGKFPVEDAK	EEVQLFLGNA	GIAMRSLTAA	VTAAGGNATY	121
VLDGVPRMRE	RPIGDLVVGL	KQLGADVDCF	LGTDCPPVRV	NGIGGLPGGK	VKLSGSISSQ	181
YLSALLMAAP	LALGDVEIEI	IDKLISIPYV	EMTLRLMERF	GVKAEHSDSW	DRFYIKGGQK	241
YKSPKNAYVE	GDASSASYFL	AGAAITGGTV	TVEGCGTTS	QGDVKFAEVL	EMMGAKVTWT	301
ETSVTVTGPP	REPFGRKHLK	AIDVNMNKMP	DVAMTLAVVA	LFADGPTAIR	DVASWRVKET	361
ERMVAIRTEL	TKLGASVEEG	PDYCIITPPE	KLNVTALD	DDHRMAMAFS	LAACAEVPVT	421
IRDPGCTRKT	FPDYFDVLST	FVKN				445

2 - METHOD

2.1 General search design

The rationale is based on the FAO/WHO expert panel ([FAO/WHO, 2001](#)) recommending a criteria of >35% identity over any segment of 80 or more amino acids as an indication of possible cross-reactivity for allergens. These criteria were adopted by the *Codex Alimentarius* Commission as the primary sequence search criteria for use in flagging proteins that might be of some concern of cross-reactivity for genetically modified plants ([C.A.C., 2003](#)). By extrapolation, the same criterion is used for conducting homology searches with known toxic proteins.

Two kinds of comparison were carried out:

- **an overall identity search:** comparison of the complete query amino acid sequence with all the sequences available in public reference databases and in the allergen database. Only the matches of at least 35% identity over 80 amino acids or more are considered relevant. The overall homology search used the FASTA program and BLOSUM62 scoring matrix.
- **a 80-mer sliding window search:** comparison of the query amino acid sequence subdivided in 80-mer blocks (segments of 1-80, then 2-81, 3-82, etc) with all known allergens present in the allergen database. The search is performed with every possible 80 amino acid segment of the query protein. Only the matches of at least 35% identity are considered as relevant. This 80-mer homology search used the FASTA program and BLOSUM50 scoring matrix.

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

Comparisons were made in a pairwise fashion. The similarity was shown by local alignments of the two sequences that included only the most similar local region(s). Each comparison was given a score reflecting the degree of similarity between the query and the comparing sequence. Discrimination between real and artefactual matches was done using an estimate of probability that the match might occur by chance. Finally, a matching criterion was selected for the identification of significant similarity to a toxin or an allergen.

In order to ensure that relevant and complete results were obtained, the following factors were considered:

- 1- the composition of the protein databases;
- 2- the sensitivity and match-scoring system of the alignment algorithm and;
- 3- the criterion used to define whether a particular match (or alignment) was biologically significant.

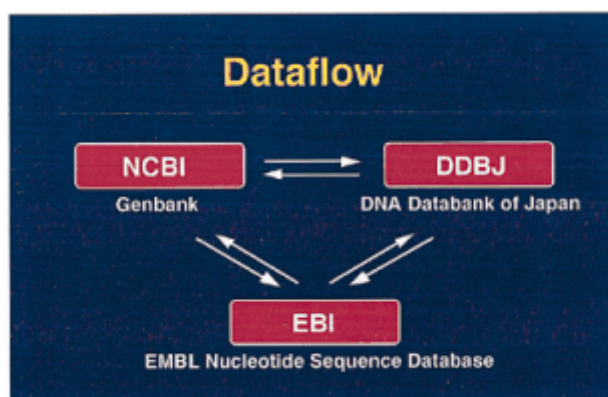
2.2 Protein databases

The query sequence will be compared to sequences present in the following large public reference databases:

- Uniprot_Swissprot (release 56.4, 2008; 400 771 sequences). Uniprot_Swissprot is a protein sequence database which strives to provide a high level of annotations (e.g. protein function, post-translational modification, etc.), a minimal level of redundancy and a high level of integration with other databases, notably the EMBL, PROSITE pattern and NRL_3D databases. Uniprot_Swissprot has been developed by SIB (Swiss Institute of Bioinformatics) and EBI (European Bioinformatics Institute).
- Uniprot_TrEMBL (release 38.6, 2008; 6 010 749 sequences). Uniprot_TrEMBL is a computer-annotated supplement of Uniprot_Swissprot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Uniprot_Swissprot. Uniprot_TrEMBL has been developed by SIB and EBI.
- PDB (2007; 56 020 sequences). PDB (Protein Data Bank) is the single international repository for public data on the 3-dimensional structures of biological macromolecules. The contents are primarily experimental data derived from X-ray crystallography and NMR experiments. This database is constantly updated as new structures are deposited by the international scientific community. It is also called NRL_3D.
- DAD (DDBJ Aminoacid sequence Database) (release 44.0, 2008; 12 561 319 sequences). The DAD is a protein database translated from the DDBJ (DNA Data Bank of Japan) which is officially certified to collect DNA sequences from researchers and to issue the internationally recognized accession number to data submitters. This database exchanges the collected data with EMBL/EBI and GenBank/NCBI on a daily basis.

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

- GenPept (release 166, 2008; 5 609 606 sequences). The GenPept is an amino acid sequence database translated from the GenBank DNA sequence of the NCBI (National Center for Biotechnology Information) as a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH). The GenBank is built by individual laboratories and by data exchange with the international nucleotide sequence databases, EMBL and DDBJ. Arrangements with the U.S. Patent and Trademark Office enable the incorporation of patent sequence data. A powerful and unique feature of this database is the ability to retrieve references that are available through PubMed, a Web search interface that provides access to over 11 million journal citations in MEDLINE and contains links to full-text articles at participating publishers' Web sites.



- AllergenOnline (release 8.0, 2008; 1 313 sequences). AllergenOnline allergen database (www.allergenonline.com) is a free, publicly available, archived resource list of known and putative allergens and sequences. The database is updated annually by searching NCBI and IUIS annotated sequences and by evaluating the candidate entries for evidence of protein allergenicity (i.e. IgE binding test) and food allergy (e.g. clinical test). A peer review panel of food allergy experts from academia is in charge of this curation. The exact list of experts is reported on the website. They identify whether proteins are allergens, putative allergens or unlikely to be allergenic based on predefined criteria, which are described on the website. Version 8.0 of the database includes 1 313 unique sequences that are clustered into 484 allergen groups based on species (n=230) and sequence identities (Thomas *et al.*, 2008).

2.3 FASTA algorithm

The FASTA program is a sequence comparison algorithm that is used to search sequence databases for optimal local alignments with a query sequence. The FASTA algorithm, developed by the University of Virginia (Pearson and Lipman, 1988), provides a quick search and local alignment of sequences contained within specified databases.

Results of matched sequences are compiled with the best optimum individual alignments between the query sequence and each of the database sequences based on the scoring matrix.

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

2.4 Amino acid substitution matrices

A key element in evaluating the quality of a pairwise sequence alignment is the substitution matrix that assigns a score for aligning any possible pair of residues. The BLOSUM series of matrices, created by [Henikoff and Henikoff \(1992\)](#), examine multiple alignments of distantly related protein regions directly, rather than extrapolate from closely related sequences. A number of tests have suggested that the BLOSUM matrices produced by this method are generally superior to the PAM matrices for detecting biological relationships.

The scoring matrix used in the overall identity search was BLOSUM62 because experimentation has shown that this matrix is one of the best for detecting low levels of protein similarities. In the 80-mer allergenic identity search, the matrix used was BLOSUM50.

2.5 Gap

Positions at which a letter is paired with a null are called gaps. Since a single mutational event may cause the insertion or deletion of more than one residue, the presence of a gap is frequently ascribed a higher significance than the length of the gap. Hence the gap is penalized heavily, whereas a lesser penalty is assigned to each subsequent residue in the gap.

To compensate for insertions and deletions, it is necessary to allow gaps to be introduced into an alignment. This was accomplished by deducting some value from the alignment score for each gap introduced - the gap cost. This value was the fixed deduction to the alignment score for the introduction of an inserted gap.

The gap costs used was (10,2) corresponding to a gap opening cost of 10 and a gap extension cost of 2. These decreased the score of the alignments which contains gaps and, thus, increased the stringency of the search. These gap costs were sufficiently large to expect that the optimal substitution scores did not change substantially.

2.6 Expect value

To assess whether a given alignment constitutes evidence for homology, it helps to know how strong an alignment can be expected by chance alone. The Expect (E) value describes the number of hits one can "expect" to see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E-value describes the random background noise that exists for matches between sequences. The E-value is used to create a significance threshold for reporting results. Lower expect thresholds are more stringent, leading to fewer chance matches being reported.

No E-value threshold was used for the overall identity search. For the 80-mer allergenic identity search, a threshold E-value of 10 was used, *i.e.* 10 matches were expected to be found merely by chance, in line with the stochastic model of [Karlin and Altschul \(1990\)](#). If the statistical significance ascribed to a match was greater than the E-value threshold, the match was not reported *i.e.* low-scoring hits were not reported.

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

2.7 Filter

Regions with "low-complexity" sequence (LCR) have an unusual composition that creates artefactual hits, with high scores, that owe virtually nothing to residue order in sequence similarity searching ([Wooton and Federhen, 1996](#)). Filters are used in the overall identity search to remove these low-complexity sequences and therefore eliminate statistically significant but biologically uninteresting data from the FASTA outputs (*e.g.* hits against common acidic-, basic- or proline-rich regions) and leave the more biologically interesting regions of the query sequence available for specific matching against database sequences.

A filter as determined by the SEG program ([Wooton and Federhen, 1993](#)) which masks off segments of the query sequence with low compositional complexity was used in the overall identity search. Filtering was only applied to the query sequence, not to database sequences. Low complexity sequence found by this filter program was substituted using the letter "X" in the protein sequence.

No filter was used in the 80-mer allergenic identity search.

2.8 Bit score

The bit score is derived from the raw alignment score in which the statistical properties of the scoring system used have been taken into account. Because bit scores have been normalized with respect to the scoring system, they can be used to compare alignment scores from different searches.

2.9 Matching criterion

The matching criterion was selected in line with the FAO/WHO report and the C.A.C. that recommend amino acid sequence searches be based on 35% identity over 80 or more amino acids of a novel protein for identification of significant similarity to a toxin or an allergen ([C.A.C., 2003](#); [FAO/WHO, 2001](#)).

2.10 Expression of results

Overall identity search: the first 100 best alignments with the query sequence are presented in the enclosed [Tables 1-5](#) and are identified by a name (accession number), a description (identification), a score in bits, a statistical significance of the match (E-value) and a percentage of identity (% identity).

Details of individual alignments between the query sequence and the database sequences are presented in the enclosed [Appendices B-F](#). The results of each individual alignment are presented as follows:

- 1- General information on the matching sequence: accession number, long form description and length of the comparing sequence in the database.

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

- 2- Overall results of the match: score, E-value, identity percentage, positive percentage, gaps.

The identity percentage was defined as the minimal percentage of identical residues between query sequence and sequence database hit.

The positive percentage was defined as the number of positive residues, dependent on the substitution matrix used.

- 3- Matching amino sequences: 'query' sequence and comparing ('Sbjct') sequence separated by a line showing the residues conserved between the two where
- ' ' (blank) indicates a mismatch
 - '+' indicates a positive match.
 - '-' in 'query' and/or 'sbjct' line indicates an insertion or a deletion.

80-mer allergenic identity search: each matching sequence from the allergen database was defined by general data on the matching sequence (accession number, long form description), the best identity percentage, and the number of 80-mers matching with more than 35% of identity. When no matches were identified between the query sequence blocks and the sequences from the entire allergen database, a computer message 'No hit found' appeared on the computer screen.

3 - DATA STORAGE

All raw data, supporting documents as well as final report are maintained in the document archive room. All of the above will be archived for at least 10 years in the designated areas at:

Bayer CropScience
355, rue Dostoïevski
BP 153
06903 Sophia Antipolis Cedex
France

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

RESULTS

The top scores for each of the protein databases were reported in [Tables 1-5](#) and [Appendices B-F](#).

Table Numbers	Appendix Numbers	Database Names	Number of matches
1	B	Uniprot_Swissprot	>100
2	C	Uniprot_TrEMBL	>100
3	D	DAD	>100
4	E	PDB	>100
5	F	GenPept	>100

In addition, no relevant similarity was found between the 2mEPSPS protein and known allergen from the AllergenOnline database.

As expected, only high similarities were observed to other EPSPS proteins (E.C. 2.5.1.19, also called 3-phosphoshikimate 1-carboxyvinyltransferase) from various origins, or with arom proteins (also called aroA and *ARO1* gene product) from various origins, known to have five enzymatic activities including the EPSPS activity. No record was found on potential toxicity or allergenicity associated with these matched proteins.

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

CONCLUSION

No identity was found with known toxins and known allergens. Only similarities to other EPSPS proteins and to other enzymes from various organisms, which have good safety records, were found.

Combined with the additional results from the epitope homology search, which are described by Capt ([2008](#)), it is unlikely that the 2mEPSPS protein could exhibit toxic or allergenic properties.

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

REFERENCES

DART Numbers	Author(s). Year. Title. Source. Edition. Pages
	Capt, A. 2008. 2mEPSPS protein - Epitope homology and N-Glycosylation searches. <i>In Silico</i> Study. Unpublished. Bayer CropScience. 22 pages.
M-276947-01-1	<i>Codex Alimentarius</i> commission (CAC). 2003. Codex principles and guidelines on food derived from biotechnology. <i>Codex Alimentarius</i> Commission 2003. CAC/GL44-2003 and CAC/GL45-2003.
M-234186-01-1	De Beuckeleer; M. 2003. Description of the amino acid sequence of the double mutated maize 5-enol pyruvylshikimate-3-phosphate synthase (2mEPSPS). Study number EPSPSaas/01. Bayer CropScience. June 23, 2003. 5 pages.
M-215805-01-1	FAO/WHO. 2001. Report of a joint Food and Agriculture Organization of the United Nations (FAO)/World Health Organization (WHO) expert consultation on foods derived from biotechnology. Rome, Italy, 22-25 January, 2001. 35 pages.
M-265207-01-1	Henikoff, S., Henikoff, J.G. 1992. Amino acid substitution matrices from protein blocks. <i>Proc. Natl. Acad. Sci. USA</i> . 89:10915-10919.
M-264607-01-1	IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). 1984. Nomenclature and symbolism for amino acids and peptides. Recommendations 1983. <i>Eur. J. Biochem.</i> 138:9-37. http://www.chem.qmw.ac.uk/iupac/AminoAcid/
M-265204-01-1	Karlin, S., Altschul, S.F. 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. <i>Proc. Natl. Acad. Sci. USA</i> . 87:2264-2268.
M-303903-01-1	Pearson, W.R. and Lipman, D.J. 1988. Improved tools for biological sequence comparison. <i>Proc. Natl. Acad. Sci. USA</i> . 85:2444-2448.
M-307889-01-1	Thomas, K., Herouet-Guicheney, C., Ladics, G., McClain, S., Macintosh, S., Privalle, L., and Woolhiser, M. 2008. Current and future methods for evaluating the allergenic potential of proteins: International workshop report 23-25 October 2007. <i>Food Chem. Toxicol.</i> 46: 3219-3225.
M-265197-01-1	Wootton, J. C., Federhen, S. 1993. Statistics of local complexity in amino acid sequences and sequence databases. <i>Computers in Chemistry</i> . 17:149-163.
M-265190-01-1	Wootton, J.C., Federhen, S. 1996. Analysis of compositionally biased regions in sequence databases. <i>Methods Enzymol.</i> 266:554-571.

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

ABBREVIATIONS AND ACRONYMS

%	Percent(age)
2mEPSPS	double mutant maize 5-enol pyruvylshikimate-3-phosphate synthase
λ	Lambda
aa	Amino acid(s)
BLOSUM	Blocks substitution matrix
bv.....	Biovar
cDNA	Complementary DNA
cds	Coding sequence
DAD	DDBJ aminoacid sequence database
DDBJ.....	DNA data bank of Japan
DNA.....	Desoxyribonucleic acid
E or E-value	Expect(ed) value
EBI	European Bioinformatics Institute
E.C.	Enzyme classification
EMBL	European Molecular Biology Laboratory sequence database
EPA	Environmental Protection Agency
EPSPS	5-enol pyruvylshikimate-3-phosphate synthase (2mEPSPS)
FAO.....	Food and Agriculture Organization
G.....	Gap score
GCG	Genetic Computer Group
HSP	High-scoring segments pairs
ID	Accession number
JCBN.....	Joint Commission on Biochemical Nomenclature
K.....	Statistical parameter for calculating BLAST scores
L	Gap opening penalty
LCR.....	Low complexity region
Ln	Natural logarithm
Log	Common logarithm
N°	Number

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

NCBI.....	National Center for Biotechnology Information
NIH.....	National Institute of Health
NLM.....	National Library of Medicine
PAM.....	Percent accepted mutation
PDB.....	Protein Data Bank
pp.	Pages
pv.....	Pathovar
S	Score
S'.....	Normalized score or bit score
Subjt	Subject (comparing) sequence
SEG	Filtering program for low complexity regions
SIB	Swiss Institute of Bioinformatics
sp.	Species
subsp.	Subspecies
str.	Strain
TrEMBL.....	Translated sequences from the EMBL nucleotide sequence database
US or USA	United States of America
WHO	World Health Organization
X.....	Masked residues (present in an alignment)

GLOSSARY

Alignment

The process of lining up two or more sequences to achieve maximal levels of identity (and conservation, in the case of amino acid sequences) for the purpose of assessing the degree of similarity and the possibility of homology.

Bit score

The score S' is derived from the raw alignment score S in which the statistical properties of the scoring system used have been taken into account. Since bit scores have been normalized with respect to the scoring system, they can be used to compare alignment scores from different searches.

If S is the raw score for a local alignment, the normalized score S' (in bits) is calculated by the formula $S' = (\lambda S - \ln K) / \ln 2$. A normalized bit score, S' with E value = E , is statistically significant if it exceeds $\log N/E$ where N is the size of the search space.

BLOSUM

Blocks Substitution Matrix. A substitution matrix in which scores for each position are derived from observations of the frequencies of substitutions in blocks of local alignments in related proteins. Each matrix is tailored to a particular evolutionary distance. In the BLOSUM62 matrix, for example, the alignment from which scores were derived was created using sequences sharing no more than 62% identity. Sequences with an identity higher than 62% are represented by a single sequence in the alignment in order to avoid over-emphasizing closely related family members.

Conservation

Changes at a specific position of an amino acid (or less commonly, DNA) sequence that preserve the physico-chemical properties of the original residue.

Description

General descriptive information about the protein sequence stored in the database. This information is generally sufficient to identify precisely the protein.

Epitope

Immunoreactive sequence of an allergen.

E-value

Expect(ed) value. The number of different alignments with scores equivalent to or better than S that are expected to occur in a database search by chance. The lower the E-value, the more significant the score.

The E-value corresponding to a given bit score is: $E = Kmne^{-\lambda S}$.

Filtering

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Also known as Masking. The process of hiding regions of (nucleic acid or amino acid) sequence having characteristics that frequently lead to spurious high scores.

Gap

A space introduced into an alignment to compensate for insertions and deletions in one sequence relative to another. To prevent the accumulation of too many gaps in an alignment, introduction of a gap causes the deduction of a fixed amount (the gap score) from the alignment score. Extension of the gap to encompass additional nucleotides or amino acid is also penalized in the scoring of an alignment.

Homology

Similarity attributed to descent from a common ancestor.

HSP

High-scoring segment pair. Local alignments with no gaps that achieve one of the top alignment scores in a given search.

In the limit of sufficiently large sequence lengths, the statistics of HSP scores are characterized by two parameters, K and λ . Most simply, the expected number of HSP with score S is given by the formula: $E = Kmne^{-\lambda S}$ called the E -value for the score S .

Identity

The extent to which two (nucleotide or amino acid) sequences are invariant. The identity percentage is the minimal percentage of identical residues between query sequence and sequence database hit.

K

A statistical parameter used in calculating BLAST scores that can be thought of as a natural scale for search space size. The value K is used in converting a raw score (S) to a bit score (S').

Lambda (λ)

A statistical parameter used in calculating BLAST scores that can be thought of as a natural scale for scoring system. The value λ is used in converting a raw score (S) to a bit score (S').

Local Alignment

The alignment of some portion of two protein (or nucleic acid) sequences.

LCR

Low Complexity Region. Regions of biased composition including homopolymeric runs, short-period repeats, and more subtle overrepresentation of one or a few residues. The SEG program is used to mask or filter LCR in amino acid queries.

Masking

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Also known as Filtering. The removal of repeated or low complexity regions from a sequence in order to improve the sensitivity of sequence similarity searches performed with that sequence.

Motif

A short conserved region in a protein sequence. Motifs are frequently highly conserved parts of proteins.

Multiple Sequence Alignment

An alignment of three or more sequences with gaps inserted in the sequences such that residues with common structural positions and/or ancestral residues are aligned in the same column.

Optimal Alignment

An alignment of two sequences with the highest possible score.

Positive

The Positive percentage is the number of positive residues, dependent on the substitution matrix used.

Query

The input sequence with which all of the entries in a database are to be compared.

Raw Score

The score of an alignment, S , calculated as the sum of substitution and gap scores. Substitution scores are given by a look-up table (see BLOSUM). Gap scores are typically calculated as the sum of G , the gap opening penalty and L , the gap extension penalty. For a gap of length n , the gap cost would be $G + Ln$. The choice of gap costs, G and L is empirical, but it is customary to choose a high value for G (10-15) and a low value for L (1-2).

Similarity

The extent to which nucleotide or protein sequences are related. The extent of similarity between two sequences can be based on percent sequence identity and/or conservation. In BLAST similarity refers to a positive matrix score.

SEG

A program for filtering low complexity regions in amino acid sequences. Residues that have been masked are represented as "X" in an alignment.

Substitution

The presence of a non-identical amino acid at a given position in an alignment. If the aligned residues have similar physico-chemical properties the substitution is said to be "conservative".

Substitution Matrix

A substitution matrix containing values proportional to the probability that amino acid i mutates into amino acid j for all pairs of amino acids. Such matrices are constructed by assembling a

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

large and diverse sample of verified pairwise alignments of amino acids. If the sample is large enough to be statistically significant, the resulting matrices should reflect the true probabilities of mutations occurring through a period of evolution.

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

TABLES

TABLE 1 - FASTA SEQUENCE SIMILARITIES OF 2MEPSPS PROTEIN WITH PROTEINS FROM THE UNIPROT SWISSPROT DATABASE

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Swissprot:AROA_PETHY	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE, CHLOROPLASTIC; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; FLAGS: PRECURSOR;	3475.1	3.8e-186	77.778
Uniprot_Swissprot:AROA_ARATH	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE, CHLOROPLASTIC; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; FLAGS: PRECURSOR;	3413.2	1e-182	75.169
Uniprot_Swissprot:AROA_BRANA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE, CHLOROPLASTIC; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; FLAGS: PRECURSOR;	3407.3	2.2e-182	75.395
Uniprot_Swissprot:AROA1_TOBAC	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE 1, CHLOROPLASTIC; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE 1; SHORT=EPSP SYNTHASE 1; FLAGS: PRECURSOR;	3403.3	3.8e-182	75.964
Uniprot_Swissprot:AROA_SOLLC	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE, CHLOROPLASTIC; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; FLAGS: PRECURSOR;	3365.3	4.9e-180	75.737
Uniprot_Swissprot:AROA2_TOBAC	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE 2; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE 2; SHORT=EPSP SYNTHASE 2; FLAGS: FRAGMENT;	2640.1	1.2e-139	79.077
Uniprot_Swissprot:AROA_RALSO	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1864.7	1.9e-96	48.284
Uniprot_Swissprot:AROA_BORA1	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1693.2	6.7e-87	45.805
Uniprot_Swissprot:AROA_BORPE	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1677.2	5.2e-86	45.495
Uniprot_Swissprot:AROA_BORPA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1673.2	8.7e-86	45.270
Uniprot_Swissprot:AROA_BORBR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1667.2	1.9e-85	45.270

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Swissprot:AROA_BLOFL	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1623.5	5.1e-83	43.187
Uniprot_Swissprot:AROA_BUCCC	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1611.5	2.4e-82	40.509
Uniprot_Swissprot:AROA_NEIMB	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1585.6	6.6e-81	44.344
Uniprot_Swissprot:AROA_RALEH	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1575.6	2.4e-80	47.846
Uniprot_Swissprot:AROA_RALEJ	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1575.6	2.4e-80	48.064
Uniprot_Swissprot:AROA_BURS3	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1575.6	2.4e-80	48.190
Uniprot_Swissprot:AROA_BURCA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1567.6	6.6e-80	48.182
Uniprot_Swissprot:AROA_BURCH	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1559.6	1.8e-79	48.190
Uniprot_Swissprot:AROA_BURVG	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1553.6	4e-79	47.738
Uniprot_Swissprot:AROA_BURCM	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1497.8	5.1e-76	47.059
Uniprot_Swissprot:AROA_BURXL	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1481.9	4e-75	47.368
Uniprot_Swissprot:AROA_THIDA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1366.3	1.1e-68	48.148
Uniprot_Swissprot:AROA_NITEC	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1362.2	1.8e-68	46.437
Uniprot_Swissprot:AROA_NITEU	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1336.3	5.1e-67	48.260
Uniprot_Swissprot:AROA_ACTSZ	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1330.3	1.1e-66	47.059
Uniprot_Swissprot:AROA_JANMA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1306.3	2.4e-65	45.330

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Swissprot:AROA_HERAR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1306.2	2.4e-65	45.747
Uniprot_Swissprot:AROA_ERWCT	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1290.5	1.8e-64	48.055
Uniprot_Swissprot:AROA_NEIMA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1284.4	3.9e-64	44.522
Uniprot_Swissprot:AROA_PASMU	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1268.4	3.1e-63	45.455
Uniprot_Swissprot:AROA_SODGM	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1224.6	8.5e-61	45.767
Uniprot_Swissprot:AROA_HAEIE	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1214.6	3e-60	47.126
Uniprot_Swissprot:AROA_HAEI8	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1214.6	3e-60	47.126
Uniprot_Swissprot:AROA_HAEIN	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1206.7	8.4e-60	46.789
Uniprot_Swissprot:AROA_BLOPB	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1200.7	1.8e-59	44.954
Uniprot_Swissprot:AROA_MANSM	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1166.8	1.4e-57	47.166
Uniprot_Swissprot:AROA_IDILO	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1156.9	5e-57	42.890
Uniprot_Swissprot:AROA_BAUCH	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1146.9	1.8e-56	42.431
Uniprot_Swissprot:AROA_SHIDS	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1135.0	8.3e-56	49.312
Uniprot_Swissprot:AROA_SHIFL	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1134.9	8.3e-56	49.541
Uniprot_Swissprot:AROA_SHISS	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1134.9	8.3e-56	49.312
Uniprot_Swissprot:AROA_ECOHS	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1134.9	8.3e-56	49.312

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Swissprot:AROAS_HIDY	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1134.9	8.3e-56	49.083
Uniprot_Swissprot:AROAS_HISO	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1134.9	8.3e-56	49.312
Uniprot_Swissprot:AROAS_ECOLI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1134.9	8.3e-56	49.312
Uniprot_Swissprot:AROAS_SHIBS	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1134.9	8.3e-56	49.312
Uniprot_Swissprot:AROAS_ECOLC	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1134.9	8.3e-56	49.312
Uniprot_Swissprot:AROAS_ECO57	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1134.9	8.3e-56	49.312
Uniprot_Swissprot:AROAS_ECO24	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1134.9	8.3e-56	49.312
Uniprot_Swissprot:AROAS_SHIF8	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1134.9	8.3e-56	49.541
Uniprot_Swissprot:AROAS_YERP3	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1130.9	1.4e-55	48.961
Uniprot_Swissprot:AROAS_YERPS	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1130.9	1.4e-55	48.961
Uniprot_Swissprot:AROAS_YERPA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1128.9	1.8e-55	48.961
Uniprot_Swissprot:AROAS_YERPE	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1128.9	1.8e-55	48.961
Uniprot_Swissprot:AROAS_YERPP	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1128.9	1.8e-55	48.961
Uniprot_Swissprot:AROAS_YERPN	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1128.9	1.8e-55	48.961
Uniprot_Swissprot:AROAS_ENT38	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1123.0	3.9e-55	49.312
Uniprot_Swissprot:AROAS_YERE8	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1119.0	6.5e-55	49.771

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Swissprot:AROA_KLEP7	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1117.0	8.3e-55	49.541
Uniprot_Swissprot:AROA_ECOL5	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1115.0	1.1e-54	49.083
Uniprot_Swissprot:AROA_ECOK1	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1115.0	1.1e-54	49.083
Uniprot_Swissprot:AROA_ECOUT	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1115.0	1.1e-54	49.083
Uniprot_Swissprot:AROA_PSEA6	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1113.0	1.4e-54	46.560
Uniprot_Swissprot:AROA_SALAR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1109.0	2.3e-54	47.706
Uniprot_Swissprot:AROA_SERP5	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1109.0	2.3e-54	49.083
Uniprot_Swissprot:AROA_SALPA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1107.0	3e-54	48.284
Uniprot_Swissprot:AROA_ECOL6	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1107.0	3e-54	48.853
Uniprot_Swissprot:AROA_PHOPR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1107.0	3e-54	48.624
Uniprot_Swissprot:AROA_VIBCH	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1105.0	3.9e-54	49.654
Uniprot_Swissprot:AROA_VIBC3	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1105.0	3.9e-54	49.423
Uniprot_Swissprot:AROA_SALT1	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1105.0	3.9e-54	48.284
Uniprot_Swissprot:AROA_SALGL	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1105.0	3.9e-54	48.284
Uniprot_Swissprot:AROA_SALPB	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1105.0	3.9e-54	48.284
Uniprot_Swissprot:AROA_SALCH	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1105.0	3.9e-54	48.284

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Swissprot:AROA_SALTY	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1105.0	3.9e-54	48.055
Uniprot_Swissprot:AROA_VIBVY	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1105.0	3.9e-54	49.199
Uniprot_Swissprot:AROA_ENTS8	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1103.0	5e-54	48.394
Uniprot_Swissprot:AROA_KLEPN	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1097.1	1.1e-53	49.541
Uniprot_Swissprot:AROA_CITK8	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1097.1	1.1e-53	48.853
Uniprot_Swissprot:AROA_SULNB	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1091.1	2.3e-53	49.195
Uniprot_Swissprot:AROA_PHOLL	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1091.1	2.3e-53	48.966
Uniprot_Swissprot:AROA_MARMS	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1089.1	3e-53	47.126
Uniprot_Swissprot:AROA_SHEHH	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1081.1	8.3e-53	47.685
Uniprot_Swissprot:ARO1_PNECA	PENTAFUNCTIONAL AROM POLYPEPTIDE; INCLUDES: 3-DEHYDROQUINATE SYNTHASE; EC=4.2.3.4; INCLUDES: 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS; INCLUDES: SHIKIMATE KINASE; EC=2.7.1.71; INCLUDES: 3-DEHYDROQUINATE DEHYDRATASE; SHORT=3-DEHYDROQUINASE; EC=4.2.1.10; INCLUDES: SHIKIMATE DEHYDROGENASE; EC=1.1.1.25;	1080.9	8.5e-53	36.036
Uniprot_Swissprot:AROA_SHEON	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1077.1	1.4e-52	46.528
Uniprot_Swissprot:AROA_SHESR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1075.1	1.8e-52	46.528
Uniprot_Swissprot:AROA_SHEPC	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1075.1	1.8e-52	46.065
Uniprot_Swissprot:AROA_SHESA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1075.1	1.8e-52	46.528
Uniprot_Swissprot:AROA_SHESW	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1073.1	2.3e-52	46.065

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Swissprot:AROA_SHESH	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1073.1	2.3e-52	46.882
Uniprot_Swissprot:AROA_SHESM	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1073.1	2.3e-52	46.296
Uniprot_Swissprot:AROA_EDWIC	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1073.1	2.3e-52	47.209
Uniprot_Swissprot:AROA_PSYIN	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1071.1	3e-52	48.284
Uniprot_Swissprot:AROA_SHEB9	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1069.1	3.8e-52	46.296
Uniprot_Swissprot:AROA_SHELP	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1069.1	3.8e-52	46.528
Uniprot_Swissprot:AROA_COLP3	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1067.2	5e-52	47.248
Uniprot_Swissprot:AROA_XENNE	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1067.1	5e-52	48.046
Uniprot_Swissprot:AROA_VIBPA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1065.2	6.4e-52	49.425
Uniprot_Swissprot:AROA_SHEB5	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; SHORT=EPSP SYNTHASE; SHORT=EPSPS;	1065.2	6.4e-52	46.528

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

TABLE 2 - FASTA SEQUENCE SIMILARITIES OF 2mEPSPS PROTEIN WITH PROTEINS FROM THE UNIPROT_TREMBL DATABASE

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Trembl:Q24566_MAIZE	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	4044.6	1.2e-216	91.216
Uniprot_Trembl:Q95AK0_ELEIN	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	3957.4	8.6e-212	89.213
Uniprot_Trembl:Q95AK1_ELEIN	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	3947.5	3e-211	89.213
Uniprot_Trembl:A0AR17_ELEIN	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	3939.5	8.4e-211	88.764
Uniprot_Trembl:A0AR18_ELEIN	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	3921.7	8.3e-210	88.764
Uniprot_Trembl:Q93VK6_ORYSA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3857.4	3.2e-206	86.517
Uniprot_Trembl:A2Y8X9_ORYSI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3857.4	3.2e-206	86.517
Uniprot_Trembl:Q5VNW0_ORYSJ	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3857.4	3.2e-206	86.517
Uniprot_Trembl:B5M6A1_WHEAT	5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE;	3792.0	1.4e-202	84.494
Uniprot_Trembl:Q3Y6S2_LOLMU	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	3765.3	4.3e-201	85.812
Uniprot_Trembl:Q1M157_CAMAC	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3548.2	5.3e-189	79.819
Uniprot_Trembl:Q6E6L5_CONCD	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3520.5	1.9e-187	78.912
Uniprot_Trembl:A8YTA6_9ASTR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	3517.5	2.7e-187	79.138
Uniprot_Trembl:Q946V0_9LAMI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3502.7	1.8e-186	79.224
Uniprot_Trembl:Q30CZ8_FAGSY	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3502.6	1.8e-186	78.005
Uniprot_Trembl:A7PYE6_VITVI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3500.6	2.4e-186	78.555
Uniprot_Trembl:A8YTA5_9ASTR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	3499.6	2.7e-186	78.685
Uniprot_Trembl:Q946U9_9LAMI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3490.8	8.4e-186	78.995

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Trembl:B2CM20_9ASTE	5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE; EC=2.5.1.19;	3490.7	8.4e-186	78.054
Uniprot_Trembl:Q6E6L4_CONCD	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3482.8	2.3e-185	77.602
Uniprot_Trembl:B3GGJ9_CONAR	5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE; EC=2.5.1.19;	3476.9	5e-185	78.054
Uniprot_Trembl:A8YTA7_9ASTR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	3469.9	1.2e-184	77.828
Uniprot_Trembl:A7Y7Y2_GOSHI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3463.0	3e-184	77.652
Uniprot_Trembl:B2C1I9_GOSHI	5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE;	3461.0	3.8e-184	77.427
Uniprot_Trembl:Q1KPV4_9ASPA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3413.4	1.7e-181	76.871
Uniprot_Trembl:O80428_ORYSJ	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3389.5	3.6e-180	85.934
Uniprot_Trembl:Q8W1M5_ORYVI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3381.8	9.9e-180	75.395
Uniprot_Trembl:Q6E6L3_CONCD	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	3368.7	5.2e-179	75.339
Uniprot_Trembl:A8YTA8_9ASTR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	3362.9	1.1e-178	75.113
Uniprot_Trembl:Q6JDV0_BRACM	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3362.0	1.2e-178	74.944
Uniprot_Trembl:Q9FVP6_ARATH	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3359.9	1.6e-178	74.718
Uniprot_Trembl:Q8LC97_ARATH	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3352.0	4.5e-178	74.492
Uniprot_Trembl:A8YTA9_9ASTR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	3297.6	4.8e-175	74.208
Uniprot_Trembl:Q0PKR0_PHAVU	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	3268.8	1.9e-173	74.150
Uniprot_Trembl:A3B841_ORYSJ	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	2963.8	1.9e-156	81.573
Uniprot_Trembl:Q9AT37_LOLRI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	2914.7	1e-153	84.104
Uniprot_Trembl:A9SYF9_PHYPA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	2829.5	5.7e-149	65.463
Uniprot_Trembl:Q6E6M2_ERIAN	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	2795.7	4.4e-147	77.654

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Trembl:Q3ECU2_ARATH	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	2760.0	4.3e-145	70.726
Uniprot_Trembl:Q6E6M1_ERIAN	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	2732.2	1.5e-143	75.000
Uniprot_Trembl:Q6E6M4_AMARU	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	2718.4	8.8e-143	75.630
Uniprot_Trembl:Q6YH15_VITVI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	2635.7	3.6e-138	79.077
Uniprot_Trembl:A4SB81_OSTLU	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	2495.2	2.4e-130	59.732
Uniprot_Trembl:Q6VVA4_ORYSJ	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	2447.4	1.1e-127	82.886
Uniprot_Trembl:A2T9R8_DUNSA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	2406.9	2e-125	58.239
Uniprot_Trembl:A8JH48_CHLRE	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	2363.4	5.3e-123	58.720
Uniprot_Trembl:A6MZG3_ORYSI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	2278.2	2.9e-118	83.150
Uniprot_Trembl:A9S302_PHYPY	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	2263.1	2e-117	66.219
Uniprot_Trembl:Q6E6M0_9ASTR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	1890.1	1.2e-96	71.970
Uniprot_Trembl:B5S6S6_RALSO	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EPSP SYNTHASE) PROTEIN; EC=2.5.1.19;	1880.9	3.9e-96	48.853
Uniprot_Trembl:Q6E6L6_PLALA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	1880.2	4.3e-96	71.321
Uniprot_Trembl:Q6E6L8_SARPR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	1880.2	4.3e-96	70.943
Uniprot_Trembl:B5SFX2_RALSO	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EPSP SYNTHASE) PROTEIN; EC=2.5.1.19;	1877.0	6.5e-96	48.853
Uniprot_Trembl:A3S0L8_RALSO	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1875.0	8.4e-96	48.853
Uniprot_Trembl:Q00VW9_OSTTA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1847.3	2.9e-94	60.991
Uniprot_Trembl:Q6E6L9_9ASTR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	1840.6	6.9e-94	69.697
Uniprot_Trembl:Q400R4_LACSA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	1831.5	2.2e-93	77.922

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Trembl:Q400R3_LACSA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	1823.6	6.1e-93	77.922
Uniprot_Trembl:Q6E6M3_ASITR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	1753.4	5e-89	66.288
Uniprot_Trembl:Q12CL6_POLSJ	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; EC=2.7.4.14;	1733.5	6.4e-88	46.120
Uniprot_Trembl:A1W8Q4_ACISJ	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; EC=2.7.4.14;	1713.6	8.1e-87	46.429
Uniprot_Trembl:A9BM07_DELAS	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; EC=2.7.4.14;	1711.6	1e-86	46.067
Uniprot_Trembl:B1XVQ3_POLNS	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1696.5	7.3e-86	46.517
Uniprot_Trembl:A4SW54_POLSQ	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1684.7	3.3e-85	46.517
Uniprot_Trembl:A1TSA6_ACIAAC	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; EC=2.7.4.14;	1681.9	4.8e-85	45.190
Uniprot_Trembl:Q2Y7A6_NITMU	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1678.8	7.1e-85	46.934
Uniprot_Trembl:Q47GK0_DECAR	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; EC=2.7.4.14;	1676.3	9.8e-85	48.268
Uniprot_Trembl:Q6E6L7_NYMAL	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; FLAGS: FRAGMENT;	1676.1	1e-84	65.098
Uniprot_Trembl:A9IJ17_BORPD	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1670.7	2e-84	45.880
Uniprot_Trembl:A2SI10_METPP	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; EC=2.7.4.14;	1636.4	1.6e-82	44.118
Uniprot_Trembl:Q21Y54_RHOFD	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; EC=2.7.4.14;	1578.9	2.6e-79	43.468
Uniprot_Trembl:B1DMY1_9RHO	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19; EC=2.7.4.14;	1577.1	3.3e-79	47.936
Uniprot_Trembl:B3R368_CUPTR	3-ENOLPYRUVYLSHIKIMATE-5-PHOSPHATE SYNTHETASE; EC=2.5.1.19;	1567.9	1.1e-78	47.846
Uniprot_Trembl:A1KUN6_NEIMF	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1565.9	1.4e-78	43.891
Uniprot_Trembl:B4RL93_NEIG2	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE;	1558.0	3.8e-78	44.626
Uniprot_Trembl:Q5F889_NEIG1	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1552.0	8.2e-78	43.764
Uniprot_Trembl:B1JXR9_BURCC	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1550.0	1.1e-77	48.190

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Trembl:A2VRY9_9BURK	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1549.4	1.1e-77	48.190
Uniprot_Trembl:B4EB42_BURCJ	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1532.2	1e-76	47.738
Uniprot_Trembl:Q1LQH1_RALME	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1516.1	8.2e-76	46.966
Uniprot_Trembl:A2WBS2_9BURK	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1512.4	1.3e-75	47.285
Uniprot_Trembl:B1FBZ3_9BURK	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1510.4	1.7e-75	47.511
Uniprot_Trembl:B1SYY8_9BURK	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE;	1508.4	2.2e-75	47.285
Uniprot_Trembl:A1K4D3_AZOSB	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1505.8	3.1e-75	47.368
Uniprot_Trembl:A7PYE5_VITVI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1501.6	5.2e-75	79.670
Uniprot_Trembl:B2U886_RALPJ	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE;	1500.5	6.1e-75	48.165
Uniprot_Trembl:A7CM60_BURPI	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1500.5	6.1e-75	48.402
Uniprot_Trembl:B2JF04_BURP8	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE;	1498.5	7.8e-75	47.368
Uniprot_Trembl:Q62M25_BURMA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1497.7	8.7e-75	48.055
Uniprot_Trembl:Q3JPX8_BURP1	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=1.3.1.12; EC=2.5.1.19;	1497.3	9.1e-75	48.055
Uniprot_Trembl:A8E7Y6_BURPS	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=1.3.1.12; EC=2.5.1.19;	1497.0	9.4e-75	48.055
Uniprot_Trembl:A5JCI2_BURMA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=1.3.1.12; EC=2.5.1.19;	1497.0	9.5e-75	48.055
Uniprot_Trembl:A3NXX1_BURP0	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=1.3.1.12; EC=2.5.1.19;	1497.0	9.5e-75	48.055
Uniprot_Trembl:A5XXG4_BURMA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=1.3.1.12; EC=2.5.1.19;	1497.0	9.5e-75	48.055
Uniprot_Trembl:A3NC32_BURP6	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=1.3.1.12; EC=2.5.1.19;	1497.0	9.5e-75	48.055
Uniprot_Trembl:A4LEC8_BURPS	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=1.3.1.12; EC=2.5.1.19;	1497.0	9.5e-75	48.055
Uniprot_Trembl:A9K5G6_BURMA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1497.0	9.5e-75	48.055

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Uniprot_Trembl:A2S4S0_BURM9	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1497.0	9.5e-75	48.055
Uniprot_Trembl:A5TQL4_BURMA	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=1.3.1.12; EC=2.5.1.19;	1497.0	9.5e-75	48.055
Uniprot_Trembl:A3MHN8_BURM7	3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE; EC=2.5.1.19;	1497.0	9.5e-75	48.055

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

TABLE 3 - FASTA SEQUENCE SIMILARITIES OF 2mEPSPS PROTEIN WITH PROTEINS FROM THE PDB DATABASE

Names	Descriptions	Bit scores	E-values	% Identity
PDB:2AA9_A	EPSP SYNTHASE LIGANDED WITH SHIKIMATE	1145.6	6.3e-57	49.312
PDB:1G6S_A	STRUCTURE OF EPSP SYNTHASE LIGANDED WITH SHIKIMATE-3- PHOSPHATE AND GLYPHOSATE	1145.6	6.3e-57	49.312
PDB:2QFT_A	<i>E. COLI</i> EPSP SYNTHASE PRO101SER LIGANDED WITH S3P AND GLYPHOSATE	1145.6	6.3e-57	49.541
PDB:2QFQ_A	<i>E. COLI</i> EPSP SYNTHASE PRO101LEU LIGANDED WITH S3P	1145.6	6.3e-57	49.312
PDB:2AAY_A	EPSP SYNTHASE LIGANDED WITH SHIKIMATE AND GLYPHOSATE	1145.6	6.3e-57	49.312
PDB:1X8R_A	EPSPS LIGANDED WITH THE (S)-PHOSPHONATE ANALOG OF THE TETRAHEDRAL REACTION INTERMEDIATE	1145.6	6.3e-57	49.312
PDB:1X8T_A	EPSPS LIGANDED WITH THE (R)-PHOSPHONATE ANALOG OF THE TETRAHEDRAL REACTION INTERMEDIATE	1145.6	6.3e-57	49.312
PDB:1G6T_A	STRUCTURE OF EPSP SYNTHASE LIGANDED WITH SHIKIMATE-3- PHOSPHATE	1145.6	6.3e-57	49.312
PDB:2QFS_A	<i>E. COLI</i> EPSP SYNTHASE PRO101SER LIGANDED WITH S3P	1145.6	6.3e-57	49.541
PDB:2PQ9_A	<i>E. COLI</i> EPSPS LIGANDED WITH (R)-DIFLUOROMETHYL TETRAHEDRAL REACTION INTERMEDIATE ANALOG	1145.6	6.3e-57	49.312
PDB:1MI4_A	GLYPHOSATE INSENSITIVE G96A MUTANT EPSP SYNTHASE LIGANDED WITH SHIKIMATE-3- PHOSPHATE	1145.6	6.3e-57	49.083
PDB:2QFU_A	<i>E. COLI</i> EPSP SYNTHASE PRO101LEU LIGANDED WITH S3P AND GLYPHOSATE	1145.6	6.3e-57	49.312
PDB:1EPS_A	STRUCTURE AND TOPOLOGICAL SYMMETRY OF THE GLYPHOSPHATE 5- ENOL-PYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE: A DISTINCTIVE PROTEIN FOLD	1143.6	8.2e-57	49.083
PDB:1Q36_A	EPSP SYNTHASE (ASP313ALA) LIGANDED WITH TETRAHEDRAL REACTION INTERMEDIATE	1129.5	5e-56	49.083
PDB:1P89_A	SUBSTRATE-INDUCED STRUCTURAL CHANGES TO THE ISOLATED N- TERMINAL DOMAIN OF 5- ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE	777.6	2e-36	42.857
PDB:1P88_A	SUBSTRATE-INDUCED STRUCTURAL CHANGES TO THE ISOLATED N- TERMINAL DOMAIN OF 5- ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE	777.6	2e-36	42.857

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
PDB:2O0Z_A	<i>MYCOBACTERIUM TUBERCULOSIS</i> EPSP SYNTHASE IN COMPLEX WITH PRODUCT (EPS)	437.8	1.7e-17	28.538
PDB:2O0D_A	<i>MYCOBACTERIUM TUBERCULOSIS</i> EPSP SYNTHASE IN COMPLEX WITH S3P	437.8	1.7e-17	28.538
PDB:2O0B_A	<i>MYCOBACTERIUM TUBERCULOSIS</i> EPSP SYNTHASE IN COMPLEX WITH S3P (PARTIALLY PHOTOLYZED)	437.8	1.7e-17	28.538
PDB:2O0E_A	<i>MYCOBACTERIUM TUBERCULOSIS</i> EPSP SYNTHASE IN COMPLEX WITH S3P AND PEP	437.8	1.7e-17	28.538
PDB:2O15_A	<i>MYCOBACTERIUM TUBERCULOSIS</i> EPSP SYNTHASE AFTER PARTIAL PRODUCTS WITHDRAWAL	437.8	1.7e-17	28.538
PDB:2O0X_A	<i>MYCOBACTERIUM TUBERCULOSIS</i> EPSP SYNTHASE IN COMPLEX WITH INTERMEDIATE	437.8	1.7e-17	28.538
PDB:2BJB_A	<i>MYCOBACTERIUM TUBERCULOSIS</i> EPSP SYNTHASE IN UNLIGANDED STATE	437.6	1.7e-17	28.538
PDB:1RF5_B	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE IN UNLIGANDED STATE	303.2	5.3e-10	24.541
PDB:1RF5_C	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE IN UNLIGANDED STATE	303.2	5.3e-10	24.541
PDB:1RF4_C	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE, TETRAHEDRAL INTERMEDIATE BOUND STATE	303.2	5.3e-10	24.541
PDB:1RF6_D	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE IN S3P-GLP BOUND STATE	303.2	5.3e-10	24.541
PDB:1RF5_D	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE IN UNLIGANDED STATE	303.2	5.3e-10	24.541
PDB:1RF4_D	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE, TETRAHEDRAL INTERMEDIATE BOUND STATE	303.2	5.3e-10	24.541
PDB:1RF5_A	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE IN UNLIGANDED STATE	303.2	5.3e-10	24.541
PDB:1RF6_C	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE IN S3P-GLP BOUND STATE	303.2	5.3e-10	24.541
PDB:1RF6_A	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE IN S3P-GLP BOUND STATE	303.2	5.3e-10	24.541
PDB:1RF4_A	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE, TETRAHEDRAL INTERMEDIATE BOUND STATE	303.2	5.3e-10	24.541

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
PDB:1RF4_B	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE, TETRAHEDRAL INTERMEDIATE BOUND STATE	303.2	5.3e-10	24.541
PDB:1RF6_B	STRUCTURAL STUDIES OF <i>STREPTOCOCCUS PNEUMONIAE</i> EPSP SYNTHASE IN S3P-GLP BOUND STATE	303.2	5.3e-10	24.541
PDB:2PQD_A	A100G CP4 EPSPS LIGANDED WITH (R)-DIFLUOROMETHYL TETRAHEDRAL REACTION INTERMEDIATE ANALOG	246.4	7.7e-07	24.374
PDB:2GGD_A	CP4 EPSP SYNTHASE ALA100GLY LIGANDED WITH S3P AND GLYPHOSATE	246.3	7.8e-07	24.374
PDB:2PQC_A	CP4 EPSPS LIGANDED WITH (R)-PHOSPHONATE TETRAHEDRAL REACTION INTERMEDIATE ANALOG	234.3	3.6e-06	24.146
PDB:2PQB_A	CP4 EPSPS LIGANDED WITH (R)-DIFLUOROMETHYL TETRAHEDRAL INTERMEDIATE ANALOG	234.3	3.6e-06	24.146
PDB:2GG4_A	CP4 EPSP SYNTHASE (UNLIGANDED)	234.2	3.7e-06	24.146
PDB:2GGA_A	CP4 EPSP SYNTHASE LIGANDED WITH S3P AND GLYPHOSATE	234.2	3.7e-06	24.146
PDB:2GG6_A	CP4 EPSP SYNTHASE LIGANDED WITH S3P	234.2	3.7e-06	24.146
PDB:1WY6_A	CRYSTAL STRUCTURE OF HYPOTHETICAL PROTEIN [ST1625P] FROM <i>HYPERTHERMOPHILIC ARCHAEON SULFOLOBUS TOKODAI</i>	126.2	3.8	37.778
PDB:1UAE_A	STRUCTURE OF UDP-N-ACETYLGLUCOSAMINE ENOLPYRUVYL TRANSFERASE	123.9	5.1	25.758
PDB:1A2N_A	STRUCTURE OF THE C115A MUTANT OF MURA COMPLEXED WITH THE FLUORINATED ANALOG OF THE REACTION TETRAHEDRAL INTERMEDIATE	123.9	5.1	25.758
PDB:2J5U_A	MREC <i>LYSTERIA MONOCYTOGENES</i>	121.4	7.1	32.394
PDB:2J5U_B	MREC <i>LYSTERIA MONOCYTOGENES</i>	121.4	7.1	32.394
PDB:1OXJ_A	CRYSTAL STRUCTURE OF THE SMAUG RNA BINDING DOMAIN	112.0	24	28.000
PDB:3C17_A	HEXAGONAL CRYSTAL STRUCTURE OF PRECURSOR <i>E. COLI</i> ISOASPARTYL PEPTIDASE/L-ASPARAGINASE (ECAIII) WITH ACTIVE- SITE T179A MUTATION	107.7	41	36.585
PDB:2ZAK_A	ORTHORHOMBIC CRYSTAL STRUCTURE OF PRECURSOR <i>E. COLI</i> ISOASPARTYL PEPTIDASE/L-ASPARAGINASE (ECAIII) WITH ACTIVE- SITE T179A MUTATION	107.7	41	36.585
PDB:2ZAK_B	ORTHORHOMBIC CRYSTAL STRUCTURE OF PRECURSOR <i>E. COLI</i> ISOASPARTYL PEPTIDASE/L-ASPARAGINASE (ECAIII) WITH ACTIVE- SITE T179A MUTATION	107.7	41	36.585

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
PDB:3C17_B	HEXAGONAL CRYSTAL STRUCTURE OF PRECURSOR <i>E. COLI</i> /ISOASPARTYL PEPTIDASE/L-ASPARAGINASE (ECAIII) WITH ACTIVE- SITE T179A MUTATION	107.7	41	36.585
PDB:2PZ5_A	CRYSTAL STRUCURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC N549T MUTATION RESPONSIBLE FOR PFEIFFER SYNDROME	107.6	42	32.051
PDB:2PZ5_B	CRYSTAL STRUCURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC N549T MUTATION RESPONSIBLE FOR PFEIFFER SYNDROME	107.6	42	32.051
PDB:1DD4_D	CRYSTAL STRUCTURE OF RIBOSOMAL PROTEIN L12 FROM <i>THERMOTOGA MARITIMA</i>	106.3	49	40.000
PDB:1DD4_C	CRYSTAL STRUCTURE OF RIBOSOMAL PROTEIN L12 FROM <i>THERMOTOGA MARITIMA</i>	106.3	49	40.000
PDB:2CCY_A	STRUCTURE OF FERRICYTOCHROME C(PRIME) FROM <i>RHODOSPIRILLUM MOLISCHIANUM</i> AT 1.67 ANGSTROMS RESOLUTION	104.1	65	35.897
PDB:2CCY_B	STRUCTURE OF FERRICYTOCHROME C(PRIME) FROM <i>RHODOSPIRILLUM MOLISCHIANUM</i> AT 1.67 ANGSTROMS RESOLUTION	104.1	65	35.897
PDB:2PZP_B	CRYSTAL STRUCURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC K526E MUTATION RESPONSIBLE FOR CROUZON SYNDROME	103.6	70	32.051
PDB:2PZP_A	CRYSTAL STRUCURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC K526E MUTATION RESPONSIBLE FOR CROUZON SYNDROME	103.6	70	32.051
PDB:1KZQ_A	CRYSTAL STRUCTURE OF A PARASITE PROTEIN	102.4	81	34.884
PDB:1KZQ_B	CRYSTAL STRUCTURE OF A PARASITE PROTEIN	102.4	81	34.884
PDB:1DD3_D	CRYSTAL STRUCTURE OF RIBOSOMAL PROTEIN L12 FROM <i>THERMOTOGA MARITIMA</i>	101.8	88	39.286
PDB:1DD3_C	CRYSTAL STRUCTURE OF RIBOSOMAL PROTEIN L12 FROM <i>THERMOTOGA MARITIMA</i>	101.8	88	39.286
PDB:2IN9_A	CRYSTAL STRUCTURE OF MTU RECA INTEIN, SPLICING DOMAIN	101.5	91	34.000
PDB:2CG4_A	STRUCTURE OF <i>E. COLI</i> ASNC	100.9	99	66.667
PDB:2CG4_B	STRUCTURE OF <i>E. COLI</i> ASNC	100.9	99	66.667
PDB:1MP5_A	Y177F VARIANT OF <i>S. ENTERICA</i> RMLA	100.3	1.1e+02	26.389
PDB:1MP3_B	L89T VARIANT OF <i>S. ENTERICA</i> RMLA	100.3	1.1e+02	26.389
PDB:1MP4_A	W224H VARIANT OF <i>S. ENTERICA</i> RMLA	100.3	1.1e+02	26.389

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
PDB:1MP5_C	Y177F VARIANT OF <i>S. ENTERICA</i> RMLA	100.3	1.1e+02	26.389
PDB:1MP5_D	Y177F VARIANT OF <i>S. ENTERICA</i> RMLA	100.3	1.1e+02	26.389
PDB:1MP5_B	Y177F VARIANT OF <i>S. ENTERICA</i> RMLA	100.3	1.1e+02	26.389
PDB:1MP4_B	W224H VARIANT OF <i>S. ENTERICA</i> RMLA	100.3	1.1e+02	26.389
PDB:1MP3_A	L89T VARIANT OF <i>S. ENTERICA</i> RMLA	100.3	1.1e+02	26.389
PDB:1ZUU_A	CRYSTAL STRUCTURE OF THE YEAST BZZ1 FIRST SH3 DOMAIN AT 0.97-Å RESOLUTION	99.6	1.2e+02	40.741
PDB:2Q0B_A	CRYSTAL STRUCTURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC E565A MUTATION RESPONSIBLE FOR PFEIFFER SYNDROME	99.5	1.2e+02	30.769
PDB:2Q0B_B	CRYSTAL STRUCTURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC E565A MUTATION RESPONSIBLE FOR PFEIFFER SYNDROME	99.5	1.2e+02	30.769
PDB:1A77_A	FLAP ENDONUCLEASE-1 FROM <i>METHANOCOCCUS JANNASCHII</i>	99.5	1.2e+02	46.667
PDB:1A76_A	FLAP ENDONUCLEASE-1 FROM <i>METHANOCOCCUS JANNASCHII</i>	99.5	1.2e+02	46.667
PDB:2GZA_A	CRYSTAL STRUCTURE OF THE VIRB11 ATPASE FROM THE <i>BRUCELLA SUI</i> S TYPE IV SECRETION SYSTEM IN COMPLEX WITH SULPHATE	98.8	1.3e+02	55.556
PDB:2GZA_B	CRYSTAL STRUCTURE OF THE VIRB11 ATPASE FROM THE <i>BRUCELLA SUI</i> S TYPE IV SECRETION SYSTEM IN COMPLEX WITH SULPHATE	98.8	1.3e+02	55.556
PDB:2GZA_C	CRYSTAL STRUCTURE OF THE VIRB11 ATPASE FROM THE <i>BRUCELLA SUI</i> S TYPE IV SECRETION SYSTEM IN COMPLEX WITH SULPHATE	98.8	1.3e+02	55.556
PDB:1J1N_A	STRUCTURE ANALYSIS OF ALGQ2, A MACROMOLECULE(ALGINATE)- BINDING PERIPLASMIC PROTEIN OF <i>SPHINGOMONAS</i> SP. A1., COMPLEXED WITH AN ALGINATE TETRASACCHARIDE	98.6	1.3e+02	28.571
PDB:1J1N_B	STRUCTURE ANALYSIS OF ALGQ2, A MACROMOLECULE(ALGINATE)- BINDING PERIPLASMIC PROTEIN OF <i>SPHINGOMONAS</i> SP. A1., COMPLEXED WITH AN ALGINATE TETRASACCHARIDE	98.6	1.3e+02	28.571
PDB:1KWH_A	STRUCTURE ANALYSIS ALGQ2, A MACROMOLECULE(ALGINATE)-BINDING PERIPLASMIC PROTEIN OF <i>SPHINGOMONAS</i> SP. A1.	98.6	1.3e+02	28.571
PDB:1IIN_D	THYMIDYLTRANSFERASE COMPLEXED WITH UDP-GLUCOSE	98.3	1.4e+02	26.389
PDB:1IIN_A	THYMIDYLTRANSFERASE COMPLEXED WITH UDP-GLUCOSE	98.3	1.4e+02	26.389

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
PDB:1IIM_B	THYMIDYLYLTRANSFERASE COMPLEXED WITH TTP	98.3	1.4e+02	26.389
PDB:1IIN_C	THYMIDYLYLTRANSFERASE COMPLEXED WITH UDP-GLUCOSE	98.3	1.4e+02	26.389
PDB:1IIM_A	THYMIDYLYLTRANSFERASE COMPLEXED WITH TTP	98.3	1.4e+02	26.389
PDB:1IIN_B	THYMIDYLYLTRANSFERASE COMPLEXED WITH UDP-GLUCOSE	98.3	1.4e+02	26.389
PDB:1H5T_D	THYMIDYLYLTRANSFERASE COMPLEXED WITH THYMIDYLYLDIPHOSPHATE- GLUCOSE	98.2	1.4e+02	26.389
PDB:1H5R_B	THYMIDYLYLTRANSFERASE COMPLEXED WITH THIMIDINE AND GLUCOSE- 1-PHOSPATE	98.2	1.4e+02	26.389
PDB:1H5T_C	THYMIDYLYLTRANSFERASE COMPLEXED WITH THYMIDYLYLDIPHOSPHATE- GLUCOSE	98.2	1.4e+02	26.389
PDB:1H5S_D	THYMIDYLYLTRANSFERASE COMPLEXED WITH TMP	98.2	1.4e+02	26.389
PDB:1H5S_B	THYMIDYLYLTRANSFERASE COMPLEXED WITH TMP	98.2	1.4e+02	26.389
PDB:1H5S_A	THYMIDYLYLTRANSFERASE COMPLEXED WITH TMP	98.2	1.4e+02	26.389
PDB:1H5R_C	THYMIDYLYLTRANSFERASE COMPLEXED WITH THIMIDINE AND GLUCOSE- 1-PHOSPATE	98.2	1.4e+02	26.389
PDB:1H5S_C	THYMIDYLYLTRANSFERASE COMPLEXED WITH TMP	98.2	1.4e+02	26.389

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

TABLE 4 - FASTA SEQUENCE SIMILARITIES OF 2mEPSPS PROTEIN WITH PROTEINS FROM THE DAD DATABASE

Names	Descriptions	Bit scores	E-values	% Identity
Dad:CAB42494.1	UNIDENTIFIED PROTEIN (SEQUENCE 5 FROM PATENT WO9802562.).	4162.2	0	91.685
Dad:CAA03526.1	ZEA MAYS PROTEIN (SEQUENCE 4 FROM PATENT WO9704114.).	4152.0	0	91.667
Dad:CAD42175.1	ZEA MAYS PROTEIN (SEQUENCE 4 FROM PATENT EP1217073.).	4152.0	0	91.667
Dad:CAA03539.1	ZEA MAYS PROTEIN (SEQUENCE 4 FROM PATENT WO9704103.).	4152.0	0	91.667
Dad:CAB42493.1	UNIDENTIFIED PROTEIN (SEQUENCE 3 FROM PATENT WO9802562.).	4141.9	0	91.236
Dad:CAA03538.1	ZEA MAYS PROTEIN (SEQUENCE 2 FROM PATENT WO9704103.).	4131.8	0	91.216
Dad:CAD42174.1	ZEA MAYS PROTEIN (SEQUENCE 2 FROM PATENT EP1217073.).	4131.8	0	91.216
Dad:CAA44974.1	ZEA MAYS EPSP-SYNTHASE PROTEIN.	4131.8	0	91.216
Dad:CAA03525.1	ZEA MAYS PROTEIN (SEQUENCE 2 FROM PATENT WO9704114.).	4131.8	0	91.216
Dad:CAD01096.1	ELEUSINE INDICA 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	4042.7	2.8e-216	89.213
Dad:AAN63155.1	ELEUSINE INDICA 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	4042.7	2.8e-216	89.213
Dad:CAD01095.1	ELEUSINE INDICA 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	4032.6	1e-215	89.213
Dad:AAN63156.1	ELEUSINE INDICA 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	4032.6	1e-215	89.213
Dad:AAR87844.1	ELEUSINE INDICA 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	4024.5	2.9e-215	88.764
Dad:AAR87845.1	ELEUSINE INDICA 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	4006.3	3e-214	88.764
Dad:BAB61062.1	ORYZA SATIVA 3-PHOSPHOSHAKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	3940.5	1.4e-210	86.517
Dad:AAL07437.1	ORYZA SATIVA EPSP SYNTHASE PROTEIN.	3940.5	1.4e-210	86.517
Dad:AAL06593.1	ORYZA SATIVA 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE PROTEIN.	3940.5	1.4e-210	86.517

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Dad:BAD67887.1	<i>ORYZA SATIVA</i> JAPONICA GROUP PUTATIVE 5-ENOLPYRUVYLSHIKIMATE- 3-PHOSPHATE SYNTHASE PROTEIN.	3940.5	1.4e-210	86.517
Dad:BAF18626.1	<i>ORYZA SATIVA</i> (JAPONICA CULTIVAR-GROUP) OS06G0133900 PROTEIN.	3940.5	1.4e-210	86.517
Dad:BAD68865.1	<i>ORYZA SATIVA</i> JAPONICA GROUP PUTATIVE 5-ENOLPYRUVYLSHIKIMATE- 3-PHOSPHATE SYNTHASE PROTEIN.	3940.5	1.4e-210	86.517
Dad:EAY99539.1	<i>ORYZA SATIVA</i> INDICA GROUP HYPOTHETICAL PROTEIN PROTEIN.	3940.5	1.4e-210	86.517
Dad:AAZ79230.2	<i>LOLIUM MULTIFLORUM</i> PLASTID 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE PROTEIN.	3846.4	2.4e-205	85.812
Dad:AAV64030.1	<i>CAMPTOTHECA ACUMINATA</i> 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE PROTEIN.	3624.6	5.5e-193	79.819
Dad:AAT45243.1	<i>CONYZA CANADENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	3596.2	2.1e-191	78.912
Dad:AAY40473.1	<i>CONYZA SUMATRENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	3593.2	3.1e-191	79.138
Dad:AAL27697.1	<i>DICLIPTERA CHINENSIS</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	3578.0	2.2e-190	79.224
Dad:ABA54869.1	<i>FAGUS SYLVATICA</i> PUTATIVE 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE PROTEIN.	3578.0	2.2e-190	78.005
Dad:CAO24534.1	<i>VITIS VINIFERA</i> PROTEIN (<i>VITIS VINIFERA</i> CHROMOSOME CHR15 SCAFFOLD_37, WHOLE GENOME SHOTGUN SEQUENCE.).	3576.0	2.8e-190	78.555
Dad:AAY40472.1	<i>CONYZA SUMATRENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	3575.0	3.2e-190	78.685
Dad:AAL27698.1	<i>DICLIPTERA CHINENSIS</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE B PROTEIN.	3565.9	1e-189	78.995
Dad:ACB37380.1	<i>CALYSTEGIA HEDERACEA</i> 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE PROTEIN.	3565.8	1e-189	78.054
Dad:AAT45244.1	<i>CONYZA CANADENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	3557.7	2.9e-189	77.602
Dad:AAY40474.1	<i>CONYZA SUMATRENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	3544.6	1.6e-188	77.828
Dad:ABV24481.1	<i>GOSSYPIUM HIRSUTUM</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	3537.5	3.9e-188	77.652
Dad:ABY61050.1	<i>GOSSYPIUM HIRSUTUM</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	3535.5	5.1e-188	77.427
Dad:AAA33699.1	PETUNIA X HYBRIDA PROTEIN (<i>P.HYBRIDA</i> 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE MRNA, COMPLETE CDS.).	3527.4	1.4e-187	77.778
Dad:CAB69241.1	PETUNIA X HYBRIDA PROTEIN (SEQUENCE 1 FROM PATENT WO9854330.).	3527.4	1.4e-187	77.778

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Dad:ABE77393.4	<i>ALLIUM MACROSTEMON</i> EPSP SYNTHASE PROTEIN.	3486.9	2.6e-185	76.871
Dad:BAE99170.1	<i>ARABIDOPSIS THALIANA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE (EPSP) SYNTHASE PROTEIN.	3464.6	4.5e-184	75.169
Dad:AAY25438.1	<i>ARABIDOPSIS THALIANA</i> AT2G45300 PROTEIN.	3464.6	4.5e-184	75.169
Dad:AAB82633.1	<i>ARABIDOPSIS THALIANA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE (EPSP) SYNTHASE PROTEIN.	3464.6	4.5e-184	75.169
Dad:BAA32276.1	<i>ORYZA SATIVA</i> JAPONICA GROUP 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFE RASE PROTEIN.	3462.5	5.9e-184	85.934
Dad:CAA35839.1	<i>BRASSICA NAPUS</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	3458.6	9.7e-184	75.395
Dad:AAA34071.1	<i>NICOTIANA TABACUM</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	3454.5	1.6e-183	75.964
Dad:AAL65913.1	<i>ORYCHOPHRAGMUS VIOLACEUS</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	3454.5	1.6e-183	75.395
Dad:AAT45245.1	<i>CONYZA CANADENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	3441.2	9e-183	75.339
Dad:CAA29828.1	<i>ARABIDOPSIS THALIANA</i> EPSP PROTEIN.	3438.3	1.3e-182	74.718
Dad:AAY40475.1	<i>CONYZA SUMATRENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	3435.3	1.9e-182	75.113
Dad:AAS80163.1	<i>BRASSICA RAPA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	3434.3	2.2e-182	74.944
Dad:AAK64123.1	<i>ARABIDOPSIS THALIANA</i> PUTATIVE 5-ENOLPYRUVYLSHIKIMATE-3- PHOSPHATE SYNTHASE PROTEIN.	3432.2	2.9e-182	74.718
Dad:AAG29739.1	<i>ARABIDOPSIS THALIANA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE (EPSP) SYNTHASE, PUTATIVE PROTEIN.	3432.2	2.9e-182	74.718
Dad:AAK25934.1	<i>ARABIDOPSIS THALIANA</i> PUTATIVE 5-ENOLPYRUVYLSHIKIMATE-3- PHOSPHATE (EPSP) SYNTHASE PROTEIN.	3432.2	2.9e-182	74.718
Dad:AAG50661.1	<i>ARABIDOPSIS THALIANA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE, PUTATIVE PROTEIN.	3432.2	2.9e-182	74.718
Dad:AAM63771.1	<i>ARABIDOPSIS THALIANA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE (EPSP) SYNTHASE, PUTATIVE PROTEIN.	3424.1	8.1e-182	74.492
Dad:AAA34136.1	<i>SOLANUM LYCOPERSICUM</i> PROTEIN (TOMATO 5-ENOLPYRUVYLSHIKIMATE- 3-PHOSPHATE SYNTHASE MRNA, COMPLETE CDS.).	3416.0	2.3e-181	75.737
Dad:AAY40476.1	<i>CONYZA SUMATRENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	3368.5	1e-178	74.208
Dad:ABG88197.1	<i>PHASEOLUS VULGARIS</i> EPSP SYNTHASE PROTEIN.	3339.1	4.4e-177	74.150

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Dad:EAZ35730.1	<i>ORYZA SATIVA</i> JAPONICA GROUP HYPOTHETICAL PROTEIN PROTEIN.	3027.5	1e-159	81.573
Dad:AAK20397.1	<i>LOLIUM RIGIDUM</i> 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE PROTEIN.	2977.3	6.2e-157	84.104
Dad:EDQ63766.1	<i>PHYSCOMITRELLA PATENS</i> SUBSP. <i>PATENS</i> PREDICTED PROTEIN PROTEIN.	2890.2	4.4e-152	65.463
Dad:AAT45236.1	<i>ERIGERON ANNUUS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	2855.6	3.7e-150	77.654
Dad:AAT45237.1	<i>ERIGERON ANNUUS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	2790.8	1.5e-146	75.000
Dad:AAT45234.1	<i>AMARANTHUS TUBERCULATUS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	2776.7	9.3e-146	75.630
Dad:AAN77867.1	<i>VITIS VINIFERA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	2692.2	4.8e-141	79.077
Dad:AAA34072.1	<i>NICOTIANA TABACUM</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	2679.9	2.3e-140	79.077
Dad:ABP00889.1	<i>OSTREOCOCCUS LUCIMARINUS</i> CCE9901 PREDICTED PROTEIN PROTEIN.	2548.6	4.7e-133	59.732
Dad:ECZ07785.1	MARINE METAGENOME HYPOTHETICAL PROTEIN PROTEIN.	2533.7	3.2e-132	59.910
Dad:AAQ75744.1	<i>ORYZA SATIVA</i> JAPONICA GROUP PUTATIVE EPSP1 PROTEIN.	2499.8	2.5e-130	82.886
Dad:ABM68632.1	<i>DUNALIELLA SALINA</i> PLASTID EPSP SYNTHASE PROTEIN.	2458.4	5e-128	58.239
Dad:EDO96795.1	<i>CHLAMYDOMONAS REINHARDTII</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	2413.9	1.5e-125	58.720
Dad:ABR25383.1	<i>ORYZA SATIVA</i> INDICA GROUP 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	2327.0	1e-120	83.150
Dad:EDQ74524.1	<i>PHYSCOMITRELLA PATENS</i> SUBSP. <i>PATENS</i> PREDICTED PROTEIN PROTEIN.	2311.4	7.7e-120	66.219
Dad:AAT45238.1	<i>HELIANTHUS SALICIFOLIUS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	1930.3	1.3e-98	71.970
Dad:AAT45242.1	<i>PLANTAGO LANCEOLATA</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	1920.2	4.8e-98	71.321
Dad:AAT45240.1	<i>SARRACENIA PURPUREA</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	1920.2	4.8e-98	70.943
Dad:CAD14609.1	<i>RALSTONIA SOLANACEARUM</i> PROBABLE 3-PHOSPHOSHAKIMATE 1-CARBOXYVINYLTRA NSFERASE (EPSP SYNTHASE) PROTEIN PROTEIN.	1892.6	1.6e-96	48.284
Dad:CAL56989.1	<i>OSTREOCOCCUS TAURI</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE (ISS) PROTEIN.	1886.6	3.5e-96	60.991

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Dad:AAT45239.1	<i>HELIANTHUS SALICIFOLIUS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	1879.7	8.5e-96	69.697
Dad:BAE20403.1	<i>LACTUCA SATIVA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	1870.5	2.8e-95	77.922
Dad:BAE20404.1	<i>LACTUCA SATIVA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	1862.4	7.9e-95	77.922
Dad:AAT45235.1	<i>ASIMINA TRILOBA</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	1790.6	7.8e-91	66.288
Dad:ABE43726.1	<i>POLAROMONAS</i> SP. JS666 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE / CYTIDYLATE KINASE PROTEIN.	1770.3	1.1e-89	46.120
Dad:ABM42629.1	<i>ACIDOVORAX</i> SP. JS42 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1750.0	1.4e-88	46.429
Dad:ABX37352.1	<i>DELFTIA ACIDOVORANS</i> SPH-1 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1747.9	1.9e-88	46.067
Dad:ACB44430.1	<i>POLYNUCLEOBACTER NECESSARIUS</i> STIR1 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1732.5	1.4e-87	46.517
Dad:ABP33718.1	<i>POLYNUCLEOBACTER</i> SP. QLW-P1DMWA-1 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1720.4	6.4e-87	46.517
Dad:CAJ48961.1	<i>BORDETELLA AVIUM</i> 197N 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1718.4	8.3e-87	45.805
Dad:ABM33844.1	<i>ACIDOVORAX AVENAE</i> SUBSP. CITRULLI AAC00-1 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1717.5	9.3e-87	45.190
Dad:ABB75365.1	<i>NITROSOSPIRA MULTIFORMIS</i> ATCC 25196 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1714.4	1.4e-86	46.934
Dad:AAZ46031.1	<i>DECHLOROMONAS AROMATICA</i> RCB CYTIDYLATE KINASE / 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1711.8	1.9e-86	48.268
Dad:AAT45241.1	<i>NYMPHAEA ALBA</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE PROTEIN.	1711.7	2e-86	65.098
Dad:CAP42227.1	<i>BORDETELLA PETRII</i> 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1706.1	4e-86	45.880
Dad:CAE41250.1	<i>BORDETELLA PERTUSSIS</i> 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1702.2	6.6e-86	45.495
Dad:CAE38415.1	<i>BORDETELLA PARAPERTUSSIS</i> 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1698.1	1.1e-85	45.270
Dad:AAF01290.1	<i>BORDETELLA BRONCHISEPTICA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE PROTEIN.	1694.1	1.9e-85	45.495
Dad:CAE33961.1	<i>BORDETELLA BRONCHISEPTICA</i> 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1692.0	2.4e-85	45.270
Dad:AAA22968.1	<i>BORDETELLA PERTUSSIS</i> PROTEIN (<i>B. PERTUSSIS</i> 5-ENOLPYRUVYLSHIKIMATE- 3-PHOSPHATE SYNTHASE (AROA) GENE, COMPLETE CDS.).	1690.0	3.2e-85	45.270

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
Dad:ABM95199.1	<i>METHYLIBIUM PETROLEIPHILUM</i> PM1 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE / CYTIDYLATE KINASE PROTEIN.	1671.0	3.6e-84	44.118
Dad:CAD83448.1	<i>CANDIDATUS BLOCHMANNIA FLORIDANUS</i> 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN.	1647.6	7.2e-83	43.187

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

TABLE 5 - FASTA SEQUENCE SIMILARITIES OF 2mEPSPS PROTEIN WITH PROTEINS FROM THE GENPEPT DATABASE

Names	Descriptions	Bit scores	E-values	% Identity
genpept:4774187	SEQUENCE 5 FROM PATENT WO9802562.	4134.5	0	91.685
genpept:21886544	SEQUENCE 4 FROM PATENT EP1217073.	4124.5	0	91.667
genpept:3714740	SEQUENCE 4 FROM PATENT WO9704103.	4124.5	0	91.667
genpept:3714674	SEQUENCE 4 FROM PATENT WO9704114.	4124.5	0	91.667
genpept:4774185	SEQUENCE 3 FROM PATENT WO9802562.	4114.4	0	91.236
genpept:3714672	SEQUENCE 2 FROM PATENT WO9704114.	4104.4	0	91.216
genpept:1524383	Z.MAYS MRNA FOR EPSP-SYNTHASE.	4104.4	0	91.216
genpept:21886542	SEQUENCE 2 FROM PATENT EP1217073.	4104.4	0	91.216
genpept:3714738	SEQUENCE 2 FROM PATENT WO9704103.	4104.4	0	91.216
genpept:16415786	<i>ELEUSINE INDICA</i> PLASTID PARTIAL MRNA FOR 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (EPSPS-S GENE).	4015.9	4e-215	89.213
genpept:24850309	<i>ELEUSINE INDICA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (EPSPS-S) GENE, PARTIAL CDS; NUCLEAR GENE FOR PLASTID PRODUCT.	4015.9	4e-215	89.213
genpept:24850311	<i>ELEUSINE INDICA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (EPSPS-R) GENE, PARTIAL CDS; NUCLEAR GENE FOR PLASTID PRODUCT.	4005.8	1.4e-214	89.213
genpept:16415784	<i>ELEUSINE INDICA</i> PLASTID PARTIAL MRNA FOR 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (EPSPS-R GENE).	4005.8	1.4e-214	89.213
genpept:40557113	<i>ELEUSINE INDICA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (EPSPS-R) MRNA, PARTIAL CDS.	3997.8	4e-214	88.764
genpept:40557115	<i>ELEUSINE INDICA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (EPSPS-S) MRNA, PARTIAL CDS.	3979.7	4.1e-213	88.764
genpept:15724392	<i>ORYZA SATIVA</i> 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE GENE, COMPLETE CDS.	3914.4	1.8e-209	86.517
genpept:15778436	<i>ORYZA SATIVA</i> EPSP SYNTHASE MRNA, COMPLETE CDS.	3914.4	1.8e-209	86.517
genpept:14549196	<i>ORYZA SATIVA</i> EPSPS, RPS20 GENES FOR 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE, 40S RIBOSOMAL PROTEIN S20, COMPLETE CDS.	3914.4	1.8e-209	86.517

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
genpept:55296169	<i>ORYZA SATIVA</i> JAPONICA GROUP GENOMIC DNA, CHROMOSOME 6, PAC CLONE:P0679C08.	3914.3	1.8e-209	86.517
genpept:113594752	<i>ORYZA SATIVA</i> (JAPONICA CULTIVAR-GROUP) GENOMIC DNA, CHROMOSOME 6.	3914.3	1.8e-209	86.517
genpept:55297191	<i>ORYZA SATIVA</i> JAPONICA GROUP GENOMIC DNA, CHROMOSOME 6, PAC CLONE:P0001H02.	3914.3	1.8e-209	86.517
genpept:113376729	<i>LOLIUM MULTIFLORUM</i> PLASTID 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE MRNA, PARTIAL CDS; NUCLEAR GENE FOR PLASTID PRODUCT.	3820.9	2.9e-204	85.812
genpept:55740769	<i>CAMPTOTHECA ACUMINATA</i> 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE (EPSPS) MRNA, COMPLETE CDS.	3600.5	5.4e-192	79.819
genpept:48526084	<i>CONYZA CANADENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS1) GENE, COMPLETE CDS.	3572.4	2e-190	78.912
genpept:63334354	<i>CONYZA SUMATRENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS-R1) MRNA, PARTIAL CDS.	3569.4	2.9e-190	79.138
genpept:16751567	<i>DICLIPTERA CHINENSIS</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE MRNA, COMPLETE CDS; NUCLEAR GENE FOR PLASTID PRODUCT.	3554.3	2e-189	79.224
genpept:76782198	<i>FAGUS SYLVATICA</i> PUTATIVE 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE MRNA, COMPLETE CDS.	3554.3	2e-189	78.005
genpept:63334331	<i>CONYZA SUMATRENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS-S1) MRNA, PARTIAL CDS.	3551.3	3e-189	78.685
genpept:16751569	<i>DICLIPTERA CHINENSIS</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE B MRNA, COMPLETE CDS; NUCLEAR GENE FOR PLASTID PRODUCT.	3542.3	9.5e-189	78.995
genpept:170783792	<i>CALYSTEGIA HEDERACEA</i> 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE MRNA, COMPLETE CDS.	3542.2	9.6e-189	78.054
genpept:48526086	<i>CONYZA CANADENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS2) GENE, COMPLETE CDS.	3534.1	2.7e-188	77.602
genpept:189170087	<i>CONVOLVULUS ARVENSIS</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE MRNA, COMPLETE CDS.	3528.1	5.8e-188	78.054
genpept:63334366	<i>CONYZA SUMATRENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS-S2) MRNA, PARTIAL CDS.	3521.1	1.4e-187	77.828
genpept:157142998	<i>GOSSYPIUM HIRSUTUM</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (EPSPS) MRNA, COMPLETE CDS.	3514.0	3.5e-187	77.652
genpept:164564335	<i>GOSSYPIUM HIRSUTUM</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (EPSPS) GENE, COMPLETE CDS.	3512.0	4.6e-187	77.427
genpept:6732247	SEQUENCE 1 FROM PATENT WO9854330.	3504.0	1.3e-186	77.778
genpept:169191	<i>P.HYBRIDA</i> 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE MRNA, COMPLETE CDS.	3504.0	1.3e-186	77.778

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
genpept:123965218	<i>ALLIUM MACROSTEMON</i> EPSP SYNTHASE MRNA, COMPLETE CDS.	3463.7	2.2e-184	76.871
genpept:110742505	<i>ARABIDOPSIS THALIANA</i> MRNA FOR 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE (EPSP) SYNTHASE, COMPLETE CDS, CLONE: RAFL09-78-J01.	3441.6	3.8e-183	75.169
genpept:2583124	<i>ARABIDOPSIS THALIANA</i> CHROMOSOME 2 CLONE F4L23 MAP CIC02E07, COMPLETE SEQUENCE.	3441.6	3.8e-183	75.169
genpept:63003818	<i>ARABIDOPSIS THALIANA</i> AT2G45300 GENE, COMPLETE CDS.	3441.6	3.8e-183	75.169
genpept:3410961	<i>ORYZA SATIVA</i> JAPONICA GROUP MRNA FOR 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE, COMPLETE CDS.	3439.5	5e-183	85.934
genpept:17815	<i>BRASSICA NAPUS</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE GENE.	3435.7	8.2e-183	75.395
genpept:18251236	<i>ORYCHOPHRAGMUS VIOLACEUS</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE MRNA, COMPLETE CDS.	3431.6	1.4e-182	75.395
genpept:170229	<i>N. TABACUM</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE MRNA, COMPLETE CDS.	3431.6	1.4e-182	75.964
genpept:48526088	<i>CONYZA CANADENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS3) MRNA, PARTIAL CDS.	3418.4	7.5e-182	75.339
genpept:295790	<i>ARABIDOPSIS THALIANA</i> GENE FOR 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (EPSP).	3415.5	1.1e-181	74.718
genpept:63334403	<i>CONYZA SUMATRENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS-S3) MRNA, PARTIAL CDS.	3412.5	1.6e-181	75.113
genpept:46095337	<i>BRASSICA RAPA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (EPSP) MRNA, COMPLETE CDS.	3411.6	1.8e-181	74.944
genpept:11094810	<i>ARABIDOPSIS THALIANA</i> CHROMOSOME 1 BAC F27K7 GENOMIC SEQUENCE, COMPLETE SEQUENCE.	3409.4	2.4e-181	74.718
genpept:12321113	<i>ARABIDOPSIS THALIANA</i> CHROMOSOME 1 BAC T24P22 GENOMIC SEQUENCE, COMPLETE SEQUENCE.	3409.4	2.4e-181	74.718
genpept:14532882	<i>ARABIDOPSIS THALIANA</i> PUTATIVE 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (AT1G48860) MRNA, COMPLETE CDS.	3409.4	2.4e-181	74.718
genpept:13430624	<i>ARABIDOPSIS THALIANA</i> PUTATIVE 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE (EPSP) SYNTHASE (AT1G48860) MRNA, COMPLETE CDS.	3409.4	2.4e-181	74.718
genpept:21555078	<i>ARABIDOPSIS THALIANA</i> CLONE 270032 MRNA, COMPLETE SEQUENCE.	3401.4	6.7e-181	74.492
genpept:170374	TOMATO 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE MRNA, COMPLETE CDS.	3393.4	1.9e-180	75.737
genpept:63334427	<i>CONYZA SUMATRENSIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS-R3) MRNA, PARTIAL CDS.	3346.2	8e-178	74.208

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
genpept:110729299	<i>PHASEOLUS VULGARIS</i> EPSP SYNTHASE MRNA, COMPLETE CDS.	3316.9	3.4e-176	74.150
genpept:13375567	<i>LOLIUM RIGIDUM</i> 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE (EPSP-S) MRNA, EPSP-S-S ALLELE, PARTIAL CDS.	2957.6	3.5e-156	84.104
genpept:48526070	<i>ERIGERON ANNUUS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS) MRNA, PARTIAL CDS.	2836.8	1.9e-149	77.654
genpept:48526072	<i>ERIGERON ANNUUS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS2) MRNA, PARTIAL CDS.	2772.4	7.3e-146	75.000
genpept:48526066	<i>AMARANTHUS RUDIS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS1) MRNA, PARTIAL CDS.	2758.3	4.4e-145	75.630
genpept:37359246	<i>VITIS VINIFERA</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE MRNA, PARTIAL CDS.	2674.4	2.1e-140	79.077
genpept:170231	<i>N. TABACUM</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE MRNA, 5' CDS.	2662.2	1e-139	79.077
genpept:144582815	<i>OSTREOCOCCUS LUCIMARINUS</i> CCE9901 CHROMOSOME 20, COMPLETE SEQUENCE.	2531.7	1.8e-132	59.732
genpept:34577025	<i>ORYZA SATIVA</i> (JAPONICA CULTIVAR-GROUP) PUTATIVE EPSP1 MRNA, PARTIAL CDS.	2483.3	9.2e-130	82.886
genpept:122937807	<i>DUNALIELLA SALINA</i> PLASTID EPSP SYNTHASE MRNA, COMPLETE CDS; NUCLEAR GENE FOR PLASTID PRODUCT.	2442.1	1.8e-127	58.239
genpept:149390731	<i>ORYZA SATIVA</i> (INDICA CULTIVAR-GROUP) CLONE OSE-97-192-D11 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE MRNA, PARTIAL CDS.	2311.6	3.3e-120	83.150
genpept:48526074	<i>HELIANTHUS SALICIFOLIUS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS1) MRNA, PARTIAL CDS.	1917.7	2.9e-98	71.970
genpept:48526082	<i>PLANTAGO LANCEOLATA</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS) MRNA, PARTIAL CDS.	1907.6	1.1e-97	71.321
genpept:48526078	<i>SARRACENIA PURPUREA</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS) MRNA, PARTIAL CDS.	1907.6	1.1e-97	70.943
genpept:17427919	<i>RALSTONIA SOLANACEARUM</i> GMI1000 CHROMOSOME COMPLETE SEQUENCE.	1880.1	3.6e-96	48.284
genpept:116054912	<i>OSTREOCOCCUS TAURI</i> STRAIN OTTH0595, *** SEQUENCING IN PROGRESS ***.	1874.2	7.7e-96	60.991
genpept:48526076	<i>HELIANTHUS SALICIFOLIUS</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS2) MRNA, PARTIAL CDS.	1867.4	1.9e-95	69.697
genpept:73912408	<i>LACTUCA SATIVA</i> MRNA FOR 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE, PARTIAL CDS, CLONE: EPSPS1.	1858.2	6e-95	77.922
genpept:73912410	<i>LACTUCA SATIVA</i> MRNA FOR 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE, PARTIAL CDS, CLONE: EPSPS2.	1850.2	1.7e-94	77.922

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
genpept:48526068	<i>ASIMINA TRILOBA</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS) MRNA, PARTIAL CDS.	1778.9	1.6e-90	66.288
genpept:91696897	<i>POLAROMONAS</i> SP. JS666, COMPLETE GENOME.	1758.6	2.1e-89	46.120
genpept:120606889	<i>ACIDOVORAX</i> SP. JS42, COMPLETE GENOME.	1738.4	2.8e-88	46.429
genpept:160365739	<i>DELFTIA ACIDOVORANS</i> SPH-1, COMPLETE GENOME.	1736.4	3.7e-88	46.067
genpept:171193469	<i>POLYNUCLEOBACTER NECESSARIUS</i> STIR1, COMPLETE GENOME.	1721.1	2.6e-87	46.517
genpept:145047091	<i>POLYNUCLEOBACTER</i> SP. QLW-P1DMWA-1, COMPLETE GENOME.	1709.1	1.2e-86	46.517
genpept:115422436	<i>BORDETELLA AVIUM</i> 197N COMPLETE GENOME.	1707.1	1.6e-86	45.805
genpept:120590404	<i>ACIDOVORAX AVENAE</i> SUBSP. CITRULLI AAC00-1, COMPLETE GENOME.	1706.2	1.8e-86	45.190
genpept:82411256	<i>NITROSOSPIRA MULTIFORMIS</i> ATCC 25196, COMPLETE GENOME.	1703.1	2.6e-86	46.934
genpept:71846535	<i>DECHLOROMONAS AROMATICA</i> RCB, COMPLETE GENOME.	1700.5	3.7e-86	48.268
genpept:48526080	<i>NYMPHAEA ALBA</i> 5-ENOL-PYRUVYLSHIKIMATE-PHOSPHATE SYNTHASE (EPSPS) MRNA, PARTIAL CDS.	1700.4	3.7e-86	65.098
genpept:163259926	<i>BORDETELLA PETRII</i> STRAIN DSM 12804, COMPLETE GENOME.	1694.9	7.5e-86	45.880
genpept:33571749	<i>BORDETELLA PERTUSSIS</i> STRAIN TOHAMA I, COMPLETE GENOME; SEGMENT 3/12.	1691.0	1.2e-85	45.495
genpept:33574090	<i>BORDETELLA PARAPERTUSSIS</i> STRAIN 12822, COMPLETE GENOME; SEGMENT 10/14.	1686.9	2.1e-85	45.270
genpept:6013215	<i>BORDETELLA BRONCHISEPTICA</i> PREPHENATE DEHYDRATASE (PHEA) GENE, PARTIAL CDS; PREPHENATE DEHYDROGENASE (TYRA), 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (ARO), AND CYTIDINE MONOPHOSPHATE KINASE (CMK) GENES, COMPLETE CDS; AND RIBOSOMAL PROTEIN S1 (RPSA) GENE, PARTIAL CDS.	1682.9	3.5e-85	45.495
genpept:33576881	<i>BORDETELLA BRONCHISEPTICA</i> STRAIN RB50, COMPLETE GENOME; SEGMENT 11/16.	1680.9	4.5e-85	45.270
genpept:144037	<i>B. PERTUSSIS</i> 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE (ARO) GENE, COMPLETE CDS.	1678.9	5.9e-85	45.270
genpept:124260205	<i>METHYLIBIUM PETROLEIPHILUM</i> PM1, COMPLETE GENOME.	1660.0	6.6e-84	44.118
genpept:33504186	<i>BLOCHMANNIA FLORIDANUS</i> COMPLETE GENOME.	1636.8	1.3e-82	43.187

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Names	Descriptions	Bit scores	E-values	% Identity
genpept:116256981	<i>BUCHNERA APHIDICOLA</i> STR. CC (<i>CINARA CEDRI</i>), COMPLETE GENOME.	1624.7	6.1e-82	40.509
genpept:89345096	<i>RHODOFERAX FERRIREDUCTENS</i> T118, COMPLETE GENOME.	1601.7	1.2e-80	43.468
genpept:7226672	<i>NEISSERIA MENINGITIDIS</i> MC58, COMPLETE GENOME.	1598.6	1.7e-80	44.344
genpept:120866826	<i>NEISSERIA MENINGITIDIS</i> SEROGROUP C FAM18 COMPLETE GENOME.	1588.5	6.3e-80	43.891
genpept:72119670	<i>RALSTONIA EUTROPHA</i> JMP134 CHROMOSOME 1, COMPLETE SEQUENCE.	1588.5	6.3e-80	48.064
genpept:113525597	<i>RALSTONIA EUTROPHA</i> H16 CHROMOSOME 1.	1588.5	6.3e-80	47.846

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

APPENDICES

APPENDIX A - AMINO ACID CODES

One-letter codes	Three-letter codes	Amino-acid names
A	Ala	Alanine
R	Arg	Arginine
N	Asn	Asparagine
D	Asp	Aspartic acid
C	Cys	Cysteine
Q	Gln	Glutamine
E	Glu	Glutamic acid
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
L	Leu	Leucine
K	Lys	Lysine
M	Met	Methionine
F	Phe	Phenylalanine
P	Pro	Proline
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
V	Val	Valine
B	Asx	Aspartic acid or Asparagine
Z	Glx	Glutamic acid or Glutamine
X	Xaa	Any amino acid

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

APPENDIX B – DETAILED FASTA SEQUENCE SIMILARITY ALIGNMENTS OF THE 2mEPSPS PROTEIN
WITH PROTEINS FROM THE UNIPROT_SWISSPROT DATABASE

AROA_PETHY | 3-phosphoshikimate 1-carboxyvinyltransferase, chloroplastic;
 EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase;
 Flags: Precursor; | *Petunia hybrida* (Petunia). | AA|516
 Length = 516
 initn: 1741 initl: 1406 opt: 1747 Z-score: 3475.1 bits: 652.4 E(): 3.8e-186
 Smith-Waterman score: 1747; 77.778% identity (90.023% similar) in 441 aa
 overlap (6-445:76-516)

			10	20	30
2MEPSP			MAGAEI	VLQPIKEIS	GTVKLP
			GSKSL	SNRILL	LLA
				
Unipro	MLVLK	KDSIF	MQKFC	SFRIS	ASVATA
	50	60	70	80	90
					100
	40	50	60	70	80
2MEPSP	ALSEGT	TVVDN	LLNSE	DVHYM	LGALRT

Unipro	ALSEGT	TVVDN	LLSSD	DIHYM	LGALKT
	110	120	130	140	150
					160
	100	110	120	130	140
2MEPSP	QLFLG	NAGIAM	RSLS	XXXXXXXXXXXX	YVL

Unipro	QLFLG	NAGTAM	RPLTA	AAVT	VAGGNS
	170	180	190	200	210
					220
	160	170	180	190	200
2MEPSP	DCPPV	RVNGI	GGGLP	GGKVK	LSGSI

Unipro	KCPPV	RIVSK	GGGLP	GGKVK	LSGSI
	230	240	250	260	270
					280
	220	230	240	250	260
2MEPSP	LRLME	RFVKA	EHSDS	WDRFY	IKGGQ

Unipro	LKLME	RFGIS	VEHSS	SWDRF	FVRGG
	290	300	310	320	330
					340
	280	290	300	310	320
2MEPSP	xxxxx	SLQGD	VKFAE	VLEMM	GAKVT

Unipro	GCGTN	SLQGD	VKFAE	VLEMM	GAKVT
	350	360	370	380	390
					400
	340	350	360	370	380
2MEPSP	MTLAV	VALF	ADGPT	AIRDV	ASWRV

Unipro	MTLAV	VALY	ADGPT	AIRDV	ASWRV
	410	420	430	440	450
					460

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	400	410	420	430	440
2MEPSP	VTAIDTYDDHRMAMAFSLAACAEVPTIRDPGCTRKTFFDYFDVLSTFVKN				
	:: ::				
Unipro	VTDIDTYDDHRMAMAFSLAACADVPVTINDPGCTRKTFFPNYFDVLQQYSKH				
	470	480	490	500	510

AROA_ARATH | 3-phosphoshikimate 1-carboxyvinyltransferase, chloroplastic;
 EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase;
 Flags: Precursor; | *Arabidopsis thaliana* (Mouse-ear cress). | AA|520
 Length = 520
 initn: 1688 initl: 1382 opt: 1716 Z-score: 3413.2 bits: 641.0 E(): 1e-182
 Smith-Waterman score: 1716; 75.169% identity (90.971% similar) in 443 aa
 overlap (4-445:78-520)

		10	20	30
2MEPSP		MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILL		
		: ::		
Unipro	SWGLKKSGMTLIGSELRPLKVMSSVSTA	EIVLQPIREISGLIKLPGSKSLSNRILL		
	50	60	70	80

	40	50	60	70	80	90
2MEPSP	LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVCGGKFPVE-DAKE					
	: ::					
Unipro	LAALSEGTTVVDNLLNSDDINYMILDALKRLGLNVETDSENNRAVVEGCGIFPASIDSKS					
	110	120	130	140	150	160

	100	110	120	130	140	150
2MEPSP	EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPRMRERPIGDLVVGLKQLGADVDCFL					
	...: :::::::::: :: :... .. : ::					
Unipro	DIELYLGNAGTAMRPLTAAVTAAGGNASYVLGDGVPRMRERPIGDLVVGLKQLGADVECTL					
	170	180	190	200	210	220

	160	170	180	190	200	210
2MEPSP	GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVE					
	: ::::::::::. : :::::::::::::::::::::::::: : ::::::::::::::::::::::::::					
Unipro	GTNCPVVRVNANGGLPGGKVKLSGSISSQYLTALLMSAPLALGDVEIEIVDKLISVPYVE					
	230	240	250	260	270	280

	220	230	240	250	260	270
2MEPSP	MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx					
	: ::::::::::. : ::::::::::::::::::. : ::::::::::::::::::::::::::::::::::					
Unipro	MTLKLMEFVGVSVEHSDSWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT					
	290	300	310	320	330	340

	280	290	300	310	320	330
2MEPSP	xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPPEPFGRKHLKAIDVNMNKMPD					
	..: :::::::::::::::::: : :::::::::::::::::::::::::: : : : ::::::::::::::::::					
Unipro	VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPPRDAFGMRHLRAIDVNMNKMPD					
	350	360	370	380	390	400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	340	350	360	370	380	390
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK					

Unipro	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGATVEEGSDYCVITPPKK					
	410	420	430	440	450	460

	400	410	420	430	440
2MEPSP	LNVTADITYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKV				

Unipro	VKTAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH				
	470	480	490	500	510

AROABRANA | 3-phosphoshikimate 1-carboxyvinyltransferase, chloroplastic;
 EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase;
 Flags: Precursor; | *Brassica napus* (Rape). | AA|516
 Length = 516
 initn: 1685 initl: 1379 opt: 1713 Z-score: 3407.3 bits: 639.8 E(): 2.2e-182
 Smith-Waterman score: 1713; 75.395% identity (90.519% similar) in 443 aa
 overlap (4-445:74-516)

	10	20	30
2MEPSP	MAGAEIIVLQPIKEISGTVKLPKSKLSNRILL		

Unipro	SWGLKKSGTMLNGSVIRPVKVTASVSTSEKASEIVLQPIREISGLIKLPKSKLSNRILL		
	50	60	70

	40	50	60	70	80	90
2MEPSP	LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVCGGKFPVE-DAKE					

Unipro	LAALSEGTTVVDNLLNSDDINMYLDALKKGLNVERDSVNNRAVVEGCGGIFPASLDSKS					
	110	120	130	140	150	160

	100	110	120	130	140	150
2MEPSP	EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL					

Unipro	DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVVGLKQLGADVECTL					
	170	180	190	200	210	220

	160	170	180	190	200	210
2MEPSP	GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVE					

Unipro	GTNCPVVRVNGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVPYVE					
	230	240	250	260	270	280

	220	230	240	250	260	270
2MEPSP	MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxx					

Unipro	MTLKLMEFVGSAEHSDSWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT					
	290	300	310	320	330	340

	280	290	300	310	320	330
2MEPSP	xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD					

Unipro	VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAVDVNMNKMPD					
	350	360	370	380	390	400

	340	350	360	370	380	390
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK					
	:	:	:	:	:	:
Unipro	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGATVEEGSDYCVITPPAK					
	410	420	430	440	450	460

	400	410	420	430	440
2MEPSP	LNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN				

Unipro	VKPAEIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFPDYFQVLESITKH				
	470	480	490	500	510

				10	20	30	
2MEPSP				MAGAE	EIVLQPIKEIS	GTVKLPGSKSLSNR	ILLLA
				:	:	:	:
Unipro	LWVSKEDSVLR	VAKSPFRISAS	VVTAQKPNEIVL	QPIKDIS	GTVKLPGSKSLSNR	ILLLA	
	50	60	70	80	90	100	
	40	50	60	70	80	90	
2MEPSP	ALSEGT	TVVDNLLN	SESDVHYMLG	ALRTLGLS	VEADKAAKRA	VVVGCGGKFPV	-EDAKEEV
	:	:	:	:	:	:	:
Unipro	ALSKGR	TVVDNLLS	DDIHYMLG	ALKTLGLH	VEDDNENQRA	IVEGCGGQFP	VGVGKKSEEEI
	110	120	130	140	150	160	

100 110 120 130 140 150
 2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVL DGVPRMRERP IGD LVVGLKQLGADVDCFLGT
 :
 Unipro QLFLGNAGTAMRPLTA AVTVAGGHSRYVL DGVPRMRERP IGD LV DGLKQLGA EVDCFLGT
 170 180 190 200 210 220

160 170 180 190 200 210
 2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
 .:.:. . :.:
 Unipro NCPPVRIVSKGGLPGGKVKLSGSISSQYL TALLMAAPLALGDVEIEIIDKLISVPYVEMT
 230 240 250 260 270 280

2MEPSP LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxx
 :.:
 Unipro LKLMERFGVSVEHTSSWDKFLVRGGQKYKSPGKAYVEGDASSASYFLAGAAVTGGTVTVE
 290 300 310 320 330 340

2MEPSP xxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGFRKHLKAIDVNMNKMMPDVA
 . :
Unipro GCGTSSLQGDVKFAEVLKMGAEVTTWTE NSVT VKGP PRNSSGMKHLRAVDVNMNKMMPDVA
 350 360 370 380 390 400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      280      290      300      310      320      330
2MEPSP xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA
      .. : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro GCGTSSLQGDVKFAEVLEKMGAEVTVTENSVTVKGPPRNSSGMKHLRAIDVNMNKMMPDVA
      350      360      370      380      390      400

```

```

      340      350      360      370      380      390
2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGATVVEGSDYCIITPPEKLN
      410      420      430      440      450      460

```

```

      400      410      420      430      440
2MEPSP VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VTEIDTYDDHRMAMAFSLAACADVPVTIKNPGCTRKTFPDYFEVLQKYSKH
      470      480      490      500      510      520

```

ARO2_TOBAC|3-phosphoshikimate 1-carboxyvinyltransferase 2; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase 2; Short=EPSP synthase 2; Flags: Fragment; |*Nicotiana tabacum* (Common tobacco). |AA|338
Length = 338
initn: 1327 initl: 1327 opt: 1327 Z-score: 2640.1 bits: 497.3 E(): 1.2e-139
Smith-Waterman score: 1327; 79.077% identity (90.462% similar) in 325 aa
overlap (121-445:14-338)

```

      100      110      120      130      140      150
2MEPSP KEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro          LTAAVAVAGGNSRYVLDGVPRMRERPIGDLVDGLKQLGAEVDC
      10      20      30      40

```

```

      160      170      180      190      200      210
2MEPSP FLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPY
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro FLGTCKPPVRIVSKGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVLY
      50      60      70      80      90      100

```

```

      220      230      240      250      260      270
2MEPSP VEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VEMTLKLMERFGISVEHSSSWDRFVVRGGQKYKSPGKAYVEGDASSASYFLAGAAVTGGT
      110      120      130      140      150      160

```

```

      280      290      300      310      320      330
2MEPSP xxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKM
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VTVEGCGTSSLQGDVKFAEVLEQMGAEVTVTENSVTVKGPPRNSSAMKHLRAIDVNMNKM
      170      180      190      200      210      220

```

```

      340      350      360      370      380      390
2MEPSP PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro PDVAMTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLGATVEEGPDYCIITPP
      230      240      250      260      270      280

```

	360	370	380	390	400	410
2MEPSP	VKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT	--	IDTYDDHRMAMAFSLAAC			
	:::::..:::	:::	:::::..:::	:::	..:	:::::::::::::::
Unipro	VKETDRLTAMATELRKLGAAVEEGTDYIRVTPPSHWTAPAGGIDTYDDHRMAMAFSLAAF					
	350	360	370	380	390	400
	420	430	440			
2MEPSP	AEVPVTIRDPGCTRKTFPDYFDVLSTFVK	N				
	.	:::	..:::	..	:::::..:::	
Unipro	GPVPVRINDPRCVAKTFPEYFTAFGGIAA					
	410	420	430			

	10	20	30	40	50	60
2MEPSP	MAGAEIEIVLQPIKEISGTVKLP	SGSKLSNRILL	LAL	SEGTTVVDN	LLNS	EDVHYMLGAL
	:	...	:
Unipro	MNALAYLDLPHIRQARG	LALPGSKSIS	NRVLLIAA	LAEGRTEIS	GLLDSD	DDTRVMLAAL
	10	20	30	40	50	60
	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRA	VVVGCGKFP	VEDAKEEVQ	LFLGNAGI	AMRSL	xxxxxxxxxxxx
	:	:
Unipro	RQLGVAV-TDLGQGR	VAVEGAR-RFPA	EKA---	ELFLGNAGT	AFRPLTA	ALALMGG--D
	70	80	90	100	110	

130 140 150 160 170 180
 2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
 : . : : : : : : : : : : . : : . : . : : : : : : : : : : : : : : : : : :
 Unipro YRLSGVPRMHERPIGDLVDALRAWGARIDYLGQAGYPPLHI-GRGDIRADRVRVQGSVSS
 120 130 140 150 160 170

190 200 210 220 230
 2MEPSP QYxxxxxxxxxxxxxxG---DVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
 : :
 Unipro QFLTALLLAAPIEAGASGRPVITIEVIGELISKPYIEITLNLMARYGVNVVR-DGWRAFTI
 180 190 200 210 220 230

2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDKVFAEVLMMGA
.: .:. : . : ::::: . . . : . : . : . : . : . : . : . : . : . :
Unipro EGDARYRSPGSIAVEGDASTASYLLALGLVGGPVRVTGVGEQSIQGDTAFAADTLAAMGA

240 250 260 270 280 290

Page 64 of 526

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      300      310      320      330      340      350
2MEPSP KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      : . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro DVRYGPGWIETRGVRVAEGGR--LKAFDADFNLIPDAAMTAATLALYADGPCRLRNIGSW
      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      : : : : . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro RVKETDRIHAMHTELEKLGAGVQSGADWLEVPPEPGGWRDAHIGTWDDHRMAMCFSLAA
      350      360      370      380      390      400

```

```

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
      410      420      430      440

```

AROABORPA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Bordetella parapertussis*. |AA|442
Length = 442
initn: 600 initl: 212 opt: 843 Z-score: 1673.2 bits: 318.8 E(): 8.7e-86
Smith-Waterman score: 843; 45.270% identity (69.369% similar) in 444 aa overlap (1-438:1-433)

```

      10      20      30      40      50      60
2MEPSP MAGAEFIVLQPIKEISGTVKLPKSGSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : . : . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGSTIEITGLLDSDTRVMLAAL
      10      20      30      40      50      60

```

```

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      : : : : . . : . . : : : : : : : : : : : : : : : : : : : : :
Unipro RQLGVSV--GEVADGCVTIEGVARFPIEQA---ELFLGNAGTAFRPLTAALALMGGD--
      70      80      90      100      110

```

```

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : . . . : : : : : : : : : : : : : :
Unipro YRLSGVPRMHERPIGDLVDALRQFGAGIEYLGQAGYPPLRIGGGSIRVDGPVRVEGSVSS
      120      130      140      150      160      170

```

```

      190      200      210      220      230
2MEPSP QYxxxxxxxxx---xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      : . . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMAAPVLARRSGQDITIEVVGELISKPYIEITLNLMARFGVSVRR-DGWRAFTI
      180      190      200      210      220      230

```

```

      240      250      260      270      280      290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      : . : . : : : : : : : : : : . : : : : : : : : : : : : : : : :
Unipro ARDAAYRGPGRMAIEGDASTASYFLALGAIGGPPVRVTGVGEDSIQGDVAFATLAAMGA
      240      250      260      270      280      290

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      300      310      320      330      340      350
2MEPSP KVTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASW
      : . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro DVRYGPGWIETRGVRVAEGGR--LKAFDADFNLIIPDAAMTAATLALYADGPCRLRNIGSW
      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro RVKETDRIHAMHTELEKLGAGVQSGADWLEVAPPAPGGWRDAHIGTWDDHRMAMCFSLAA
      350      360      370      380      390      400

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
      410      420      430      440

```

AROABORBR | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Bordetella bronchiseptica* (*Alcaligenes bronchisepticus*). |AA|442
Length = 442
initn: 600 initl: 212 opt: 840 Z-score: 1667.2 bits: 317.7 E(): 1.9e-85
Smith-Waterman score: 840; 45.270% identity (69.369% similar) in 444 aa overlap (1-438:1-433)

```

      10      20      30      40      50      60
2MEPSP MAGAEIIVLQPIKEISGTVKLPKSGSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : . : . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGSTEITGLLDSDTRVMLAAL
      10      20      30      40      50      60

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro RQLGVSV--GEVADGCVTIEGVARFPTEQA---ELFLGNAGTAFRPLTAALALMGGD--
      70      80      90      100      110

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : . . . : : : : : : : : : : : : :
Unipro YRLSGVPRMHERPIGDLVDALRQFGAGIEYLGQAGYPPLRIGGGSIRVDGPVRVEGSVSS
      120      130      140      150      160      170

      190      200      210      220      230
2MEPSP QYxxxxxxxxx---xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      : . . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMAAPVLARRSGQDITIEVVGELISKPYIEITLNLMARFGVSVRR-DGWRAFTI
      180      190      200      210      220      230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	240	250	260	270	280	290
2MEPSP	KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVL	EMMGA				
	: . . .	:	: . .	:	:	:
Unipro	ARDAAYRGPGRMAIEGDASTASYFLALGAIGGGPVRVTGVGEDSIQGDVAF	AATLAAMGA				
	240	250	260	270	280	290

	300	310	320	330	340	350
2MEPSP	KVTWTETSVTVTGPPREPFGFRKHLKAIDVNMNKM	PDVAMTLAVVALFADGPTAIRDVASW				
	:	:	:	:	:	:
Unipro	DVRYGPGWIE	TRGVRVAEGGR--LKA	FDADFN	LIPDAAMTAATL	ALYADGPCRLRNIGSW	
	300	310	320	330	340	

	360	370	380	390	400	410
2MEPSP	RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNV	TAIDTYDDHRMAMAFSLAA				
	:	:	:	:	:	:
Unipro	RVKETDRIHAMHTELEKLGAGVQSGADWLEVAPPAPGGWRDA	HIGTWDDHRMAMCFSLAA				
	350	360	370	380	390	400

	420	430	440
2MEPSP	CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN		
	:	:	
Unipro	FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD		
	410	420	430

AROABLOFL | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Blochmannia floridanus*. |AA|433
Length = 433
initn: 773 initl: 286 opt: 818 Z-score: 1623.5 bits: 309.5 E(): 5.1e-83
Smith-Waterman score: 818; 43.187% identity (70.670% similar) in 433 aa overlap (7-437:5-424)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILL	LAALSEGTTVVDNLLNSEDVHYMLGAL				
	:	:	:	:	:	:
Unipro	MENFITLNPIK	KINGTIYLP	GSKSISNRALL	LAQSVGATRLINLL	SDDTRYMLAAL	
	10	20	30	40	50	

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSL	xxxxxxxxxxxxxx				
	:	:	:	:	:	:
Unipro	LQLGVKYKLSTDHKICEIKGVGGS--LQSKNTQTTLFLGNAGTAIRPLIAALS	IKSHNI-				
	60	70	80	90	100	110

	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGS	SISS				
	:	:	:	:	:	:
Unipro	-VLTGDMRMKERPIFHLVDTLRQGGGRIEYIERDHHL	PVKL--YGGYCGGNIVIKGDISS				
	120	130	140	150	160	170

	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKA	EHSDSWDRFYIKGGQ				
	:	:	:	:	:	:
Unipro	QFLSAVLMMTPLASKNTYIEVIGSLVSKPYIDITLSVMRAFGICVQHD	KDYKFFYCEG	NR			
	180	190	200	210	220	230

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      .:. . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro IYKAPQEYVIEGDASSASYFLAAAIAKGGTIRVLGVGKNSKQGDIKFANILERMGSIIITW
                240          250          260          270          280          290

                310          320          330          340          350
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADG--PTAIRDVASWRV
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro GDNYI-----ECSKGIMLKSIDIDVNDIPDAAMTLAVIALFTTNNLPMILRNIYNWRV
                300          310          320          330          340

                360          370          380          390          400          410
2MEPSP KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEV
      .:. . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro KESDRLHAMATELRKVGAIVSEGYDLHIVPPVQIQSAFINTYNDHRIAMCFALVALSNV
                350          360          370          380          390          400

                420          430          440
2MEPSP PVTIRDPGCTRKTFPDYFDVLSTFVKN
      .:. . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro SVTINNPKCVCKTFPDFFDRFLSVGNL
                410          420          430

```

AROA_BUCCC | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Buchnera aphidicola* subsp. *Cinara cedri*. |AA|435
Length = 435
initn: 775 init1: 306 opt: 812 Z-score: 1611.5 bits: 307.3 E(): 2.4e-82
Smith-Waterman score: 812; 40.509% identity (70.139% similar) in 432 aa overlap (5-436:3-419)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro MQDSLTLKPVDYIQKINIPGSKSISNRVLLLSALSNGKTILKNLLYSDDIKYMLKAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro LKLGIIFYKLDKKKSKCTIYGISDAFSV--KNKIKLFLGNAGTAMRPLLAILSLKKNKI-
                60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro -ILTGEKRMKERPIHHLVDSLQGGANITYKNKKKFPLYIK--GGFKGGKIFIDGSISS
                120          130          140          150          160          170

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYEMTLRLMERFGVKAEHSDSWDRFYIKGGQ
      . . . . . : : : . . . : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLSSLLMAAPLAELDTEIIVKNQLVSKPYINLTINLMEKFGISVSILNDYKHFYIKGNQ
                180          190          200          210          220          230

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : : : . . : : : : : : : : : . . . : : : : : : : : : : : : : : : :
Unipro KYISPKKYYIESDLSSATYFLAAAIAKGGSIQINGIQKKSQIGDINFIKILKQMGVSIQW
                240          250          260          270          280          290

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . : : . . : : : : : : : : : . . . . . : : : : : : : : : : : : :
Unipro KKNSVICK-----KNKLLGITVDCNHIPDAAMTIAILGVFSKKKVYIKNIYNWRVKE
                300          310          320          330          340

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
      : : . . : : : : : : : : : . . . . . : : : : : : : : : : : : : :
Unipro TDRIYAMSTELKKIGARVITGKDYIKVYPVKNFIAKINTYNDHRIAMCFSLISLSGTSV
                350          360          370          380          390          400

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      : . : : : : : : : :
Unipro TLLNPKCVNKTFPSFFKNFY SICHYSNINKNI
                410          420          430

```

AROA_NEIMB | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Neisseria meningitidis* serogroup B. |AA|433
Length = 433
initn: 804 initl: 178 opt: 799 Z-score: 1585.6 bits: 302.5 E(): 6.6e-81
Smith-Waterman score: 799; 44.344% identity (70.136% similar) in 442 aa overlap (6-444:3-430)

```

                10          20          30          40          50
2MEPSP MAGAEEIVLQPIKEIS-GTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGA
      : : . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MTESVRLPVARLKPSTVALPGSKSISNRTLLLAALSDNACEIHSLLKSDDTDRMLEA
                10          20          30          40          50

                60          70          80          90          100          110
2MEPSP LRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro LDKLGVQIEY-LAEDRLKVHGTGGRFPNRTA---DLFLGNAGTAFRPLTAALAVLGGD-
                60          70          80          90          100          110

                120          130          140          150          160          170
2MEPSP xYVL DGVPRMRERPIGDLVVG LKQLGADVDCFLGTD-CPPVRVNGIGGLPGGKVKLSGSI
      : : : : : : : : : . . : : : . . : : . . : : . . : : . . : : .
Unipro -YHLHGVP RMHERPIGDLVDALRIAGADVE-YLGKEHY PPLHIGERQDNGERV IPIKGNV
                120          130          140          150          160

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      180      190      200      210      220      230
2MEPSP SSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG
      ::. . . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro SSQFLTALLMALPLTGQAFEIRMVGELISKPYIDITLKLMAQFGVQVIN-EGYRVFKIPA
      170      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
      : . . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro DAHYHAPEHLHVEGDASSASYFLAAGLIAATPVRVTGIGANSIQGDVAFARELEKIGADV
      230      240      250      260      270      280

      300      310      320      330      340      350
2MEPSP TWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      .: . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro VWGENFVEVSRPKE-----RAVQSFDDLDAHNPDAAMTLAIVALATGQTCTLRNIGSWRV
      290      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAACAE
      ::. . . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro KETDRIAAMANELRKLGAKEVEAEAIHITPPETLTPDAVIDTYDDHRMAMCFSLVSLLG
      350      360      370      380      390      400

      420      430      440
2MEPSP VPVTIRDPGCTRKTFPDYFDVLSTFVKV
      ::. . . :. . . . :. . . . .
Unipro VPVVINDPKCTHKTFTPTFDVFSSLTETAE
      410      420      430

```

AROA_RALEH | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Ralstonia eutropha* (strain ATCC 17699 / H16 / DSM 428 / Stanier 337) | AA|434
Length = 434
initn: 900 initl: 227 opt: 794 Z-score: 1575.6 bits: 300.7 E(): 2.4e-80
Smith-Waterman score: 948; 47.846% identity (74.150% similar) in 441 aa overlap (5-440:2-430)

```

      10      20      30      40      50      60
2MEPSP MAGAEIIVLQPIKEISGTVKLPKSGKSLNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :. . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro MEHLTLGPLTRANGTVRLPGSKSISNRVLLLAALATGETVRDLLDSDDTRVMLQAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      ::. . . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro RTLGVAWRQE--GDDYIVTGAGGNFPNKS-----ELFMGNAGTAIRPLTAALALQGG--N
      60      70      80      90      100

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro YKLSGVPRMHERPIGDLVDGLRQVGAVIDYLGNEGFPPLHIQPAIRIDAPIRVRGDVSS
      110          120          130          140          150          160

                190          200          210          220          230
2MEPSP QYxxxxxxxx--xxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
      : . . . . . : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMSLPMAQSDSGRIEIEVVGELISKPYIEITLNLARFGIEIERQ-GWERFVLP
      170          180          190          200          210          220

                240          250          260          270          280          290
2MEPSP GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      . : : : . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro AGAAYRSPGEIFVEGDASSASYFLAAGAIGGGPVRVEGVGMASIQGDVRF AEALNRMGAN
      230          240          250          260          270          280

                300          310          320          330          340          350
2MEPSP VTWTETSVITVTGPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWR
      : . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VMAGDNWIEVRGTERDD-GRLLH--GIELDCNHIPDAAMTLAVAALFAEGTTTLTNIASWR
      290          300          310          320          330          340

                360          370          380          390          400          410
2MEPSP VKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT--IDTYDDHRMAMAFSLAAC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VKETDRIAAMATELRKLGAVVEEGADYLRVTPPQPWQTPADGIGTYDDHRMAMCFSLAAF
      350          360          370          380          390          400

                420          430          440
2MEPSP AEPVPTIRDPGCTRKTFPDYFDVLSTFVK
      . : : : : : : : : : : : : : : : :
Unipro GPLPVRINDPGCVAKTFPDYFSVFAGVTG
      410          420          430

```

AROA_RALEJ | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Ralstonia eutropha* (strain JMP134) | AA|434
Length = 434
initn: 892 initl: 224 opt: 794 Z-score: 1575.6 bits: 300.7 E(): 2.4e-80
Smith-Waterman score: 945; 48.064% identity (73.576% similar) in 439 aa overlap (5-438:2-428)

```

                10          20          30          40          50          60
2MEPSP MAGAEIEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MEHLTLGLPLTRAAGTVRLPGSKSISNRVLLLAALAGGETVRDLLDSDTRVMLQAL
      10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      : : : : . . . . : : : : : : : : : : : : : : : : : : :
Unipro KTLGVAWRQE--GSDYIVTGSGGNFPVKAA---ELFMGNAGTAIRPLTAALALQGGD--
      60          70          80          90          100

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  YKLSGVPRMHERPIGDLVDGLRQVGADIGYLANEGFPPLHIRPAQIRIDAPIRVRGDVSS
        110          120          130          140          150          160

                190          200          210          220          230
2MEPSP  QYxxxxxxxx--xxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
        : . . . . . : : : : : : : : : : : : : : : : : : : : : : :
Unipro  QFLTALLMTLPMAQAANGKIEIEVVGELISKPYIEITLNLARFGINVERQ-GWERFIVP
        170          180          190          200          210          220

                240          250          260          270          280          290
2MEPSP  GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
        . : : : . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  AGAVYRSPGEIYVEGDASSASYFLAAGAIGGGPVRVEGVGMSSIQGDVRF AEALNRMGAN
        230          240          250          260          270          280

                300          310          320          330          340          350
2MEPSP  VTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWR
        : . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  VMAGDNWIEVRGTERDD-GR LH--GIELDCNHIPDAAMTLAVAALFAGGTTTLTNIASWR
        290          300          310          320          330          340

                360          370          380          390          400          410
2MEPSP  VKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT A--IDTYDDHRMAMAFSLAAC
        : : : : . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  VKETDRISAMATELRKLGATVEEGADYLKVTTPPAQWQTPADGIGTYDDHRMAMCFSLAAF
        350          360          370          380          390          400

                420          430          440
2MEPSP  AEPVPTIRDPGCTRKTFPDYFDVLSTFVK N
        . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  GPLPVRINDPGCVAKTFPDYFAVFGGVTR
        410          420          430

```

ARO A_BURS3 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Burkholderia* sp. (strain 383) | AA | 434
Length = 434
initn: 950 initl: 251 opt: 794 Z-score: 1575.6 bits: 300.7 E(): 2.4e-80
Smith-Waterman score: 955; 48.190% identity (73.756% similar) in 442 aa overlap (9-444:6-433)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        : : . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  MDYLDLGPYSSASGTVRLPGSKSISNRVLLLAALAEGETTITNLLDSD DTRVMLDAL
        10          20          30          40          50

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  GTLGVKLARD--GDTCVVTGTRGAFTAKTA----DLFLGNAGTAVRPLTAALAVNGGD--
      60          70          80          90          100

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  YRVHGVPRMHERPIGDLVDGLRQIGAQIDYELNEGYPLRIKPANISVDAPIRVRGDVSS
      110          120          130          140          150          160

              190          200          210          220          230
2MEPSP  QYxxxxxxxx---xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  QFLTALLMTLPLVKAADGKIVVEVDGELISKPYIDITIRLMERFGVTVER-DGWQRFVVP
      170          180          190          200          210          220

              240          250          260          270          280          290
2MEPSP  GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  AGVRYRSPGRIMVEGDASSASYFLAAGALGGGPLRVEGVGRASIQGDVGAFANALMQMGAN
      230          240          250          260          270          280

              300          310          320          330          340          350
2MEPSP  VTWTETSVTVTGPPREPFGGRKH--LKAIDVNMNMKMPDVAMTLAVVALFADGPTAIRDVAS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  VTMGDDWIDVRG-----IGHDHGKLEPIDMDFNLIPDAAMTIAVAALFANGTSTLRNIAS
      290          300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  WRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  WRVKETDRIAAMATELRKVGAIIEEGPDYLVVTPPEKLTPNAAIDTYDDHRMAMCFSLVS
      350          360          370          380          390          400

              420          430          440
2MEPSP  CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  LGGVPVRINDPKCVGKTFPDYFDRFAALAKA
      410          420          430

```

AROA_BURCA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Burkholderia cenocepacia* (strain AU 1054). |AA|434
Length = 434
initn: 938 init1: 252 opt: 790 Z-score: 1567.6 bits: 299.2 E(): 6.6e-80
Smith-Waterman score: 945; 48.182% identity (73.409% similar) in 440 aa overlap (9-444:6-433)

```

              10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSGKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  MDYLDLGPYSSASGTVRLPGSKSISNRVLLLAALAEGETTITNLLDSDDTRVMLDAL
      10          20          30          40          50

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : . . . . : . . . . : : . . : . . . . . : : . . . .
Unipro  GKLGVKLARD--GDTCTVVTGTRGAFTAKTA---DLFLGNAGTAVRPLTAALAVNGGD--
      60          70          80          90          100

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : . . . . . : : : : : : : : : : : : . . . . . : . . . . .
Unipro  YRVHGVPRMHERPIGDLVDGLRQIGAQIDYELSEGYPPLRIKPATISVDAPIRVRGDVSS
      110          120          130          140          150          160

              190          200          210          220          230
2MEPSP  QYxxxxxxxx---xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
      : . . . . . : . . . . . : : : : : : : : : : : : : : : : : .
Unipro  QFLTALLMTLPLVKAQDGQAVVEVDGELISKPYVDITIRLMARFGVTVER-DGWQRFVVP
      170          180          190          200          210          220

              240          250          260          270          280          290
2MEPSP  GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      : . . . . : . . . . . : : : : : : : : : : : : : : : : : .
Unipro  AGVRYRSPGRIMVEGDASSASYFLAAGALGGGPLRVEGVGRASIQGDVGAFANALMQMGAN
      230          240          250          260          270          280

              300          310          320          330          340          350
2MEPSP  VTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWR
      : : . . : : . . : : : : : : : : : : : : : : : : : : : : :
Unipro  VTMGDDWIDVRGIGHD---RGKLEPIDMDFNLIPDAAMTIAVAALFANGTSTLRNIASWR
      290          300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  VKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAACA
      : : : . . . . : : : : : : : : : : : : : : : : : : : : : .
Unipro  VKETDRIAAMATELRKVGAIVEEGPDYLVVTPPEKLTNPAAIDTYDDHRMAMCFSLVSLG
      350          360          370          380          390          400

              420          430          440
2MEPSP  EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  GVPVRINDPKCVGKTFPDYFDRFAALAKA
      410          420          430

```

AROA_BURCH | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Burkholderia cenocepacia* (strain HI2424). |AA|434
Length = 434
initn: 938 initl: 252 opt: 786 Z-score: 1559.6 bits: 297.7 E(): 1.8e-79
Smith-Waterman score: 941; 48.190% identity (73.303% similar) in 442 aa overlap (9-444:6-433)

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEI	SGTVK	LPGSK	SLSNR
	:	:	:	:	:	:
Unipro	MDYLD	LGPYS	SASGT	TVRLP	GSKSIS	NRVLL
	:	:	:	:	:	:
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KAARAV	VVGCGG	KFPVED	AKEEVQ
	:	:	:	:	:	:
Unipro	GKLG	VKLAR	D--GD	TCVVT	GTRGA	FTAKTA
	:	:	:	:	:	:
	60	70	80	90	100	
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	PIGDL	VVGLK	QLGADV	DCFLGT
	:	:	:	:	:	:
Unipro	YRVHG	VPRMH	ERPIG	DLVDG	LRQIGA	QIDYEL
	:	:	:	:	:	:
	110	120	130	140	150	160
	190	200	210	220	230	
2MEPSP	QYxxxx	xxxxx	GDVEI	EIIDKL	ISIPY	VEMTL
	:	:	:	:	:	:
Unipro	QFLT	ALLMT	LPLVK	AKDQG	AVVEV	DGELIS
	:	:	:	:	:	:
	170	180	190	200	210	220
	240	250	260	270	280	290
2MEPSP	GGQKY	KSPKN	AYVEG	DASSA	SYFLA	GAIX
	:	:	:	:	:	:
Unipro	AGVRY	RSPGR	IMVEG	DASSA	SYFLA	AGALG
	:	:	:	:	:	:
	230	240	250	260	270	280
	300	310	320	330	340	350
2MEPSP	VTWT	ETSV	TVTG	PPREP	FGRKH	--LKA
	:	:	:	:	:	:
Unipro	VTMG	DDWID	VRG---	IGHD	HGKLE	PIDMD
	:	:	:	:	:	:
	290	300	310	320	330	340
	360	370	380	390	400	410
2MEPSP	WRVK	ETERM	VAIRTE	LTKL	GA	SVEEG
	:	:	:	:	:	:
Unipro	WRVK	ETDRI	AAMATE	LRKV	GAIVE	EGPDY
	:	:	:	:	:	:
	350	360	370	380	390	400
	420	430	440			
2MEPSP	CAEVP	VTIRD	PGCTR	KTFPD	YFDVL	STFVKN
	:	:	:	:	:	:
Unipro	LGGV	PVRIN	DPKCV	GKTFP	DYFDR	FAALAKA
	:	:	:	:	:	:
	410	420	430			

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AROA_BURVG | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Burkholderia vietnamiensis* (strain G4 / LMG 22486) | AA|434

Length = 434

initn: 910 initl: 248 opt: 783 Z-score: 1553.6 bits: 296.6 E(): 4e-79

Smith-Waterman score: 937; 47.738% identity (72.851% similar) in 442 aa overlap (9-444:6-433)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEI	SGTVK	LPGSK	SLSNR
	:	:	:	:	:	:
Unipro	MDYLD	LGPISS	ASGTV	RLPGS	KSISN	RVLLA
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KAAKRA	VVVGCG	GKFPVE	DAKEEV
	:	:	:	:	:	:
Unipro	GKLGV	RLTRD	--ADT	CVVAG	TRGAF	TARTA
	60	70	80	90	100	
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIGD	LVVGL	KQLGAD	VDCFL
	:	:	:	:	:	:
Unipro	YRVHG	VPRMR	HERPIG	DLVDG	LRQIGA	QIDYE
	110	120	130	140	150	160
	190	200	210	220	230	
2MEPSP	QYxxxx	xxxxx	GDVEI	EIIID	KLISIP	YVEMTL
	:	:	:	:	:	:
Unipro	QFLTAL	LMTLPL	VKAKDG	KIVVE	VDGEL	ISKPYI
	170	180	190	200	210	220
	240	250	260	270	280	290
2MEPSP	GGQKY	KSPKN	AYVEGD	ASSASY	FLAGAA	Ixxxxxxxx
	:	:	:	:	:	:
Unipro	AGVRY	RSPGR	IMVEGD	ASSASY	FLAAGAL	GGGPLR
	230	240	250	260	270	280
	300	310	320	330	340	350
2MEPSP	VTWTET	SVTVTG	PPREP	FGRKH	--LKA	IDVNMN
	:	:	:	:	:	:
Unipro	VSMGD	DWIEVR	G-----	IGHDH	GKLDPI	DMDFNL
	290	300	310	320	330	340
	360	370	380	390	400	410
2MEPSP	WRVKET	ERMVAI	RTLTKL	GASVEE	GPDYCI	IITPPE
	:	:	:	:	:	:
Unipro	WRVKET	DRIAAM	ATELRK	LGAIVE	EGPDYL	VVTPPQ
	350	360	370	380	390	400
	420	430	440			
2MEPSP	CAEVP	VTIRDP	GCTRKT	FPDYFD	VLSTF	VKN
	:	:	:	:	:	:
Unipro	LGGVP	VRINDP	KCVGKT	FPDYFD	RFAAL	AKA
	410	420	430			

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AROA_BURCM | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Burkholderia ambifaria* (strain ATCC BAA-244 / AMMD) |AA|434
Length = 434
initn: 901 initl: 237 opt: 755 Z-score: 1497.8 bits: 286.3 E(): 5.1e-76
Smith-Waterman score: 910; 47.059% identity (72.624% similar) in 442 aa overlap (9-444:6-433)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEI	SGTVK	LPGSK	SLSNR
	:	:	:	:	:	:
Unipro	MDYLD	LGPIY	SASGT	TVRLP	GSKSIS	NRVLL
	:	:	:	:	:	:
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KAARK	RAVVG	CGGKF	PVEDAK
	:	:	:	:	:	:
Unipro	GKLG	VKLAR	---GD	TCVVT	GTRGA	FTAKTA
	:	:	:	:	:	:
	60	70	80	90	100	
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIG	DLVVG	LKQLG	ADVDC
	:	:	:	:	:	:
Unipro	YRVHG	VPRMH	ERPIG	DLVDG	LRLQIG	AQIDYE
	:	:	:	:	:	:
	110	120	130	140	150	160
	190	200	210	220	230	
2MEPSP	QYxxxx	xxxxx	GDVEI	EIIDK	LISIP	YVEMTL
	:	:	:	:	:	:
Unipro	QFLT	ALLMT	LPLVK	AKDGR	TVVEI	DGELIS
	:	:	:	:	:	:
	170	180	190	200	210	220
	240	250	260	270	280	290
2MEPSP	GGQKY	KSPKN	AYVEG	DASSA	SYFLA	GAIXx
	:	:	:	:	:	:
Unipro	AGVRY	RSPGR	IMVEG	DASSA	SYFLA	AGALG
	:	:	:	:	:	:
	230	240	250	260	270	280
	300	310	320	330	340	350
2MEPSP	VTWT	ETSVT	VTGPP	REPFG	RKH--	LKAID
	:	:	:	:	:	:
Unipro	VSMG	DDWIE	VRG---	IGHD	HGKLD	PIDMDF
	:	:	:	:	:	:
	290	300	310	320	330	340
	360	370	380	390	400	410
2MEPSP	WRVK	ETERM	VAIRTE	LTKLG	ASVEE	GPDYCI
	:	:	:	:	:	:
Unipro	WRVK	ETDRI	AAMATE	LRKVG	AIVEE	GADYLV
	:	:	:	:	:	:
	350	360	370	380	390	400

420 430 440
 2MEPSP CAEVPVTIRDPGCTRKTFFDYFDVLSTFVKN
 . :: :: :: : ::: :: ::: :
 Unipro LGGVPVRINDPKCVGKTFPDYFNRFAALAKA
 410 420 430

(c) 107-126

	10	20	30	40	50	60
2MEPSP	MAGAEIEIVLQPIKEISGTVKLP	GSKSLSNRILL	LAALESGTTVVDNLLNS	EDVHYMLGAL		
	:	:	:	:	:	:
Unipro	MEFLDLGPFSRASGITRLPGSKS	SISNRVLLLAA	LAEGETTITNLLDSD	DTRVMLDAL		
	10	20	30	40	50	

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGC GGFVPEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxxxx
.  ::.  . . :  .  ::::  : :  . . :  .  :::::  :::  :...  .
Unipro  EKLGVRLKRD--GDTCVVTGTRGAFTARTA----DLFLGNAGTAVRPLTAALAVNGGD--
              60          70          80          90          100

```

130 140 150 160 170 180
 2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
 : . ::::::::::::::::::::
 Unipro YRIHGVPRMHERPIGDLVDGLRQLGAKIDYEENEGYPPLRIRPGQINADAPITVRGDVSS
 110 120 130 140 150 160

190 200 210 220 230
 2MEPSP QYxxxxxxxx--xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIK
 : : :
 Unipro QFLTSLMTLPLVRTGSGVSTVQVDGELISKPYIEITLKLMERFGIKVERH-GWHQFVVP
 170 180 190 200 210 220

```

240      250      260      270      280      290
2MEPSP  GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLLEMMGAK
       .:.:.:. .  :.:.:.:.:.:.:.:.:. .  :.:.:.:.:.:.:.:.:.
Unipro  AGQRYQSPGSIMVEGDASSASYFLAAGALGGGPLKVEGVGRASIQGDVGFDALIRMGAN
       230      240      250      260      270      280

```

300 310 320 330 340 350
 2MEPSP VTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWR
 . . : . : . : . : : : : : : : :
 Unipro LQMGDDWIEVRGVGHDS-GK--LEPIDMDCNLIPDAAMTIAVAALFADGATTLRNIASWR
 290 300 310 320 330 340

360 370 380 390 400 410
 2MEPSP VKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLVNTA-IDTYDDHRMAMAFSLAACA
 :::::::::: :: :::::::::: : : ::::: : ::::::::::: ::::::::::: ::::::::::
 Unipro VKETDRLAAMATELRKVGAKVKEGEDFIIIEPPEKLIPNAAIDTYDDHRMAMCFSLVSLG
 350 360 370 380 390 400

	360	370	380	390	400	410
2MEPSP	KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAACAE					
	::: :: : : : : :					
Unipro	KETDRIAAMATELRKVGATVEEGD DFLRVTPPETLRAGAVIDTYDDHRMAMCLSLVSLGG					
	350	360	370	380	390	400
	420	430	440			
2MEPSP	VPVTIRDPGCTRKTFPDYFDVLSTFVKN					
	: : : : : : :					
Unipro	VAVRINDPGCVAKTFPGYFNAFAE IA R					
	410	420				

		10	20	30	40	50	60		
2MEPSP	MAGAE	EIVLQPIKEIS	GTVKLP	SGSKLSNR	ILLALL	ALSEGTT	VVDNLLNS	EDVHY	MLGAL
	:	:	:	:	:	:	:	:	:
Unipro	MEWLD	LPLVRCAAG	KVRLPG	SKSISNR	VLLLSALA	EAGTTT	ISNLLES	DDTG	RMLDAL
		10	20	30	40	50			
		70	80	90	100	110	120		
2MEPSP	RTLGL	SVEADKAA	KRAVV	VGCGK	FPVEDA	KEEVQ	LFLGN	AGIAM	RSLxxxxxxxxxxxx
	:	:	:	:	:	:	:	:	:
Unipro	KMLGV	AVTRTDE	-DQYLIT	GCSGR	FSVKEA	---DLF	LGNAGT	AFRPLT	AVLSLMHGH--
	60	70	80	90	100	110			

130 140 150 160 170 180
 2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
 : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
 Unipro YHLSGVPRMHERPIGDLVDALRQVGAVITYLEREHFPPLFIHPAAIHPAD-ILINGNISS
 120 130 140 150 160

190 200 210 220 230 240
 2MEPSP QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
 : : :: : : : : : : : : : : : : : : :
 Unipro QFLSGLLMALPLAGEPATIIIVNGTLISQPYVTLTIAQMAYFGVQVER-ESWLRFIVPGNQ
 170 180 190 200 210 220

2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
 : : . :::::::::::::::::::: : :::::::::::::: :
 Unipro IYHSPGKIIVEGDASSASYFLAAGAIAGGPVRVDGVGRDSCQGDIRFVEALEAMGACIKM
 230 240 250 260 270 280

Page 81 of 526

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                370          380          390          400          410
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLVN--TAIDTYDDHRMAMAFSLAACAEV
      :. . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TDRITAMSTELRKLGAVESGDDFLRITPPDDPLVANTVIDTYDDHRMAMCFSLISLG-T
      350          360          370          380          390          400

                420          430          440
2MEPSP PVTIRDPGCTRKTFFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro PVRINDPHCVAKTFPDYFEKFTAITHQ
      410          420          430

```

AROA_NITEU | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Nitrosomonas europaea*. |AA|431
 Length = 431
 initn: 827 initl: 225 opt: 674 Z-score: 1336.3 bits: 256.4 E(): 5.1e-67
 Smith-Waterman score: 890; 48.260% identity (71.462% similar) in 431 aa overlap (9-437:6-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  MQWLDLPHVQRAQGNVRLPGSKSISNRILLLSALAEGTTMVSNNLLESDDTGRMLDAL
      10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  RLLGVAIVRTDDGKYRVA-GCKGKFPVREA---ELFLGNAGTAFRPLTAVLALMQGH--
      60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  YRLSGVPRMHMERPIGDLVDALRQIGAVITCLEHEGFPPLEIHPAVIRPGN-ISIKGNISS
      120          130          140          150          160

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . : : . : : : : : : : : : : : : : : : : : : : : : :
Unipro  QFLSGLLMALPLTGEPVTIVVSGTLISQPYVALTIAQMARFGVQVKQ-ESWQRFMLPENQ
      170          180          190          200          210          220

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : : : . : : : : : : : : : . . : : : : : : : : : : : : :
Unipro  TYRSPGKIAVEGDASSASYFLAAGAIAGGPVRIEGAGSDSCQGDIFRVEALEAMGARISM
      230          240          250          260          270          280

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . . :          : : : . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro GSDWIESGAPDGG-----LKAIDFDCNHIPDAAMTLATMALFARGTTTLRNIAASWRVKE
          290          300          310          320          330          340

```

```

          370          380          390          400          410
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPE-KLNVTA-IDTYDDHRMAMAFSLAACAEV
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TDRIAAMSAELRKLGRARVEAGDDFLRITPPDGPLTADAVIDTYDDHRMAMCFSLVSL-V
          350          360          370          380          390          400

```

```

          420          430          440
2MEPSP PVTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro PVRINDPGCVAKTFPDYFEKFAAITHTPF
          410          420          430

```

AROA_ACTSZ | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Actinobacillus succinogenes* (strain ATCC 55618 / 130Z). |AA|433
Length = 433
initn: 782 initl: 287 opt: 671 Z-score: 1330.3 bits: 255.3 E(): 1.1e-66
Smith-Waterman score: 899; 47.059% identity (72.398% similar) in 442 aa overlap (5-440:2-429)

```

          10          20          30          40          50          60
2MEPSP MAGAEFIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MEKITLNPRIARVEGTVNLPGSKSLSNRALLLAALAKGTTKVNTLLDSDVVRHMLNAL
          10          20          30          40          50

```

```

          70          80          90          100          110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxx--
      . : : : . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro KRLGVSYTLSDDKTVCEVQGLGRAFEWQNG--LSLFLGNAGTAMRPLTAALCLANSNGNE
          60          70          80          90          100          110

```

```

          120          130          140          150          160          170
2MEPSP --xxYVLGDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSG
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro SPAEIVLTGEPRMKERPIKHLVDALLQAGAEIEYLEQDGYPLAIRN-KGLHGKVKIDG
          120          130          140          150          160          170

```

```

          180          190          200          210          220          230
2MEPSP SISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      : : : : . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro SVSSQFLTALLMAAPMSAADTEIEIIGDLVSKPYIDITLNMKIFGVEVENRN-YQRFVT
          180          190          200          210          220          230

```

```

          240          250          260          270          280          290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro KGCQQYQSPKTFLVEGDASSASYFLAAGAIKGCVKVTGVGKN-SIQGDRFLFADVLTAAGA
          240          250          260          270          280          290

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      300      310      320      330      340      350
2MEPSP KVTWTETSVTVTGPPREPFGFRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASW
      :: : . . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro KITWGEDFIRAE-----QGELSGVDMMDMNHIPDAAMTIATAALFAKGETVIRNIYNW
      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPE--KLNVT AIDTYDDHRMAMAFSLAA
      ::::: . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro RVKETDRLAAMATELRKVGATVEEGEDFIRIQPLPLAEFKHAEIETYNDHRMAMCFALIA
      350      360      370      380      390      400

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro LSDTPVTILDPKCTAKTFPTFFDEFISRISQ
      410      420      430

```

AROJA_JANMA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Janthinobacterium* sp. (strain Marseille) |AA|441
Length = 441
initn: 815 initl: 318 opt: 659 Z-score: 1306.3 bits: 250.9 E(): 2.4e-65
Smith-Waterman score: 855; 45.330% identity (71.754% similar) in 439 aa overlap (9-440:10-433)

```

      10      20      30      40      50
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSKLSNRILL AALSEGTTVVDNLLNSEDVHYMLGA
      :: . . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MKHYPHYLDLKPAMHAKGVVRLPGSKSISNRTLLLAALA QGTTHIRDLLASDDTHVMLMA
      10      20      30      40      50      60

      60      70      80      90      100      110
2MEPSP LRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxx
      : . . . . : . . : : : : : : : : : : : : : : : : : : : : :
Unipro LQKLGVKWEQIGESQDYIVHGVDGSFPVHQA---DLFMGNAGTAIRPLTAALAVTGG--
      70      80      90      100      110

      120      130      140      150      160      170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro DYTLHGVS RMHERPIGDLVDALNAIGTHIEYTGE PGYPPLHIQR-GRIHAHEMSVRGNVS
      120      130      140      150      160      170

      180      190      200      210      220      230
2MEPSP SQYxxxxxxxx--xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKA EHSDSWDRFYIK
      : . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro SQFLTALLMAAPLMTREQDVIINVIGDLISKPYIEITLNL IQRFGVEVQRN-GWQSFTIK
      180      190      200      210      220      230

```

	240	250	260	270	280	290
2MEPSP	GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK					
	.::.: :: : .:::.....:::					
Unipro	AGQRYISPGNIHVEGDASSASYFLAAGAIAGGPVRVEGVGRDSIQGDVRFVEALEQMGAT					
	240	250	260	270	280	290
	300	310	320	330	340	350
2MEPSP	VTWTETSVTVTGPPREPFGGRKHLKAIDVNMNMKMPDVAMTLAVVALFADGPTAIRDVASWR					
	: . . . : . . ::::..: .:::.....:::					
Unipro	VRMGDNWI-----EAESNGVLRAIDADFNHIPDAAMTIAVAALYADGPSILRNIGSWR					
	300	310	320	330	340	
	360	370	380	390	400	410
2MEPSP	VKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAE					
	::::.: .: :: :::: :::: :::: .: .: .:::.....:::					
Unipro	VKETDRISAMATELRKLGAIVEEGEDYLVKVTTPAEILSAAIDTYDDHRMAMCFSLATLDG					
	350	360	370	380	390	400
	420	430	440			
2MEPSP	VP-----VTIRDPGCTRKTFPDYFDVLSTFVKN					
	. ::: :. ::::.....					
Unipro	AARRGNKERINDPQCVAKTFPEYFEAFKVTTEESLF					
	410	420	430	440		
ARO_A_HERAR 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.1 enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Sho Herminiimonas arsenicoxydans. AA 442 Length = 442 initn: 811 initl: 312 opt: 659 Z-score: 1306.2 bits: 250.9 E(): 2 Smith-Waterman score: 859; 45.747% identity (71.494% similar) in 4 (9-436:11-430)						
	10	20	30	40	50	
2MEPSP	MAGAE EIVLQPIKEISGTVKLPGSKSLSNRILL L AALSEGTTVVDNLLNSEDVHYMLG					
	::: . ::::.....:::					
Unipro	MNKHYPHYLDLQ PAMHAKGVVRLPGSKSISNR TLL L AALADGTTQIRDLLASDDTHVMLM					
	10	20	30	40	50	60
	60	70	80	90	100	110
2MEPSP	ALRTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxx					
	::. :... : .. :: : .: ::::..: .:::.....:::					
Unipro	ALQKIGVKWEQIGESQDYVVHGVNGAFPVHQA----DLFMGNAGTAIRPLTAALAVTGGD					
	70	80	90	100	110	
	120	130	140	150	160	170
2MEPSP	xxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTD CPPVRVNGIGGLPGGKVKLSGSI					
	::: :: ::::.....:::					
Unipro	--YTLHGVS RMHERPIGDLVDALNAIGTHIEYTGE PGYPPLHIQR-GRIHAQRMQVRGNV					
	120	130	140	150	160	170
	180	190	200	210	220	230
2MEPSP	SSQYxxxxxxxx--xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKA EHSDSWDRFYI					
	:::. . . : : :::: .: ::::.....:::					
Unipro	SSQFLTALLMAAPLMAREQDVTIDVVGELISKPYIEITLNLMRFRGVEVQR-DGWQSFTI					
	180	190	200	210	220	230

	240	250	260	270	280	290
2MEPSP	KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA					
	.::.: ::	.:::.....::.			.:::.....::.	
Unipro	AAGQRYISPGIIHVEGDASSASYFLAAGAIAGGPVRVEGVGKNSIQGDVRFVESLQQMGA					
	240	250	260	270	280	290
	300	310	320	330	340	350
2MEPSP	KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW					
	.: .. .	: ..	:::.....:	:::.....:::.....::	:::.....::	
Unipro	TIITMGDNWI-----EAKSNGALRAIDADFNHIPPDAAMTIAVAALYADGTSTLRNIGSW					
	300	310	320	330	340	
	360	370	380	390	400	410
2MEPSP	RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACA					
	:::..:: .	:: ::::.....::	::: :::	..	:::.....:::.....::	
Unipro	RVKETDRISAMTIELRLKLASVEEGEDYLRITPPAVIQPAAIDTYDDHRMAMCFSLATLD					
	350	360	370	380	390	400
	420	430	440			
2MEPSP	EV-----PVTIRDPGCTRKTFPDYFDVLSTFVKN					
	.	.::: .	:::.....			
Unipro	GAIRKGSKIRINDPQCVAKTFPDYFTAFKVTEDSLF					
	410	420	430	440		

		10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQPIKEIS	GTVKLPGSKSL	SNRILL	LAALSEG	TTVVDNLLN	SEDEVHYMLGAL
	:	:	:	:	:	:	:
Unipro	MQESL	TLHP	IKLINGT	LNLP	GSKSV	SNRALL	LAALSEGKTRLTNLLDSD
	:	:	:	:	:	:	:
		10	20	30	40	50	
		70	80	90	100	110	120
2MEPSP	RTLGLS	VEADKAAKRA	VVVGCGKFP	VEDAKEE	VQFLGN	NAGIAMRSL	xxxxxxxxxxxxxx
	:	:	:	:	:	:	:
Unipro	TALGVEY	HLSSDRTV	CEIIGL	GGAFA--	ASQPLEL	FLGNAGTAMR	PLAAALCLTDGDI-
	:	:	:	:	:	:	:
	60	70	80	90	100	110	
		130	140	150	160	170	180
2MEPSP	YVLDG	VPRMRER	PIGDLV	VGLKQL	GADVDC	FLGTDC	PPVRVNGIGGLPGGKVKLSGSISS
	:	:	:	:	:	:	:
Unipro	-VLTGE	PRMKER	PIGHLV	DALRQ	GGA	IDYLEQ	ENYPPLRLHG--GFQGG
	:	:	:	:	:	:	:
	120	130	140	150	160	170	

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQ
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMTAPLAAQDTQISIQGDLVSKPYIDITLHMMKAFGIDVRH-ENYQRFVAGRQ
                180          190          200          210          220          230

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      . : : . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QYRSPGDYLVEGDASSASYFLAAAIKGGVVRVTGVGRNSVQGDIRFADVLEKMGAIVRW
                240          250          260          270          280          290

                310          320          330          340          350
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADG-PTAIRDVASWRVK
      : . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro GEDYIECE-----RGELHAIDMDMNHIPDAAMTIATAALFAQGGTTTLRNIYNWRVK
                300          310          320          330          340

                360          370          380          390          400          410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro ETDRLAAMAIELRKVGAEVEEGNDYIRITPPAKLKAAEIGTYNDHRMAMCFSLVALSDTP
                350          360          370          380          390          400

                420          430          440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VTILDPKCTAKTFPDYFEQLARLSELA
                410          420

```

AROA_NEIMA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Neisseria meningitidis* serogroup A. | AA | 433
Length = 433
initn: 803 initl: 181 opt: 648 Z-score: 1284.4 bits: 246.8 E(): 3.9e-64
Smith-Waterman score: 783; 44.522% identity (69.930% similar) in 429 aa overlap (18-444:16-430)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MTESLRLPAASLKPSTVALPGSKSISNRTLLLAALSDNACEIHSLLKSDDTDRMLEAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro DKLGVIEFY-LAEDRLKVHGTGGRFPNRTA---DLFLGNAGTAFRPLTAALAVLGG--D
                60          70          80          90          100          110

                130          140          150          160          170
2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTD-CPPVRVNGIGGLPGGKVKLSGSIS
      : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro YRLHGVARMHERPIGDLADALRIAGADVE-YLGKEHYPLHIGERQDNGERVPIKGNVS
                120          130          140          150          160          170

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      180      190      200      210      220      230
2MEPSP SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      :: . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro SQFLTALLMALPLTGQAFEIRMGELISKPYIDITLKLMAQFGVQVIN-EGYRVFKIPAD
      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      : . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro AHYHAPHLHVEGDASSASYFLAAGLIAATPVRVTGIGANSIQGDVAFACELEKIGADV
      230      240      250      260      270      280

      300      310      320      330      340      350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      : : : : : : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro WGENFVEVSRPK-----ERAVRAFDLDANHIPDAAMTLAIVALATGQTCTLRNIGSWRVK
      290      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAACAEV
      : . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro ETDRIAAMANELRKLGAKVAEEAEAIHITPPETPTDAVIDTYDDHRMAMCFSLVSLLDV
      350      360      370      380      390      400

      420      430      440
2MEPSP PVTIRDPGCTRKTFPDYFDVLSTFVKN
      : . . . . : . . . . : . . . . .
Unipro PVVINDPKCTHKTFPTYFDVFSSLTETAE
      410      420      430

```

AROA_PASMU | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Pasteurella multocida*. | AA|440
Length = 440
initn: 865 initl: 252 opt: 640 Z-score: 1268.4 bits: 243.9 E(): 3.1e-63
Smith-Waterman score: 885; 45.455% identity (72.500% similar) in 440 aa overlap (4-437:5-430)

```

      10      20      30      40      50
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGKSLSNRILLALLAALSEGTTVVDNLLNSEDVHYMLGA
      : . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro MIKDATAITLNPISYIEGEVRLPGSKSLSNRALLLSALAKGKTTLTNLLDSDDVHRMLNA
      10      20      30      40      50      60

      60      70      80      90      100      110
2MEPSP LRTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxx---
      : . . . . : . . . . : . . . . : . . . . : . . . . : . . . . : . . . .
Unipro LKELGVTYQLSEDKSVCEIEGLGRAFEWQSG---LALFLGNAGTAMRPLTAALCLSTPNK
      70      80      90      100      110

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      120      130      140      150      160      170
2MEPSP -xxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLS
      :: : ::::: :: : : : :::: .      ::. . . : : :::::
Unipro EGKNEIVLTGEPRMKERPIQHLVDALCQAGAEIQYLEQEGYPPIAIRNTG-LKGGRIQID
      120      130      140      150      160      170

      180      190      200      210      220      230
2MEPSP GSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFY
      ::::: . . . . . ::::: ::: ::::: ::::: ::::: :::::
Unipro GSVSSQFLTALLMAAPMAEADTEIEIIGELVSKPYIDITLKMMQTFGVEVENQ-AYQRFL
      180      190      200      210      220      230

      240      250      260      270      280      290
2MEPSP IKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMG
      .:: :.::: . ::::: ::::: .      .::: ::::: ::
Unipro VKGHQQYQSPHRFLVEGDASSASYFLAAAAIKGKVKVTGVGKN-SIQGDRLFADVLEKMG
      240      250      260      270      280      290

      300      310      320      330      340      350
2MEPSP AKVTWTETSVITVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVAS
      : .:: . . : . ::::: ::::: ::::: ::::: ::::: .
Unipro AHITWGDDFIQVE-----KGNLKGIDMDMNHIPDAAMTIATTALFAEGETVIRNIYN
      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP WRVKETERMVAIRTELTKLGASVEEGPDYCIITPPE--KLNVTATIDTYDDHRMAMAFSLA
      ::::: .::: .:: :.::: ::: : : . . . . :.::: :::::
Unipro WRVKETDRLTAMATELRKVGAEEVEGEDFIRIQPLNLAQFQHAIEITYNDHRMAMCFALI
      350      360      370      380      390      400

      420      430      440
2MEPSP ACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      : . . .:: :.::: ::::: ::::: .::
Unipro ALSQTSVTILDPSCTAKTFPTFFDTFLRLTHAES
      410      420      430      440

```

AROA_SODGM|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Sodalis glossinidius* (strain morsitans). |AA|435
Length = 435
initn: 803 initl: 295 opt: 618 Z-score: 1224.6 bits: 235.7 E(): 8.5e-61
Smith-Waterman score: 893; 45.767% identity (72.082% similar) in 437 aa overlap (5-440:3-423)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      .....: .::: ::::: ::::: ::::: .::: . ::::: ::::: ::
Unipro MQDSLTLNPIARVDGTLNLPGSKSVSNRALLLAAQANGTTRLTNLLSDSDVRHMLTAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :. . . . . : : : : : : ::::: ::::: . . . . .
Unipro GQLGVNYRLSADRRSCEIDGLGG--PLR-ADEALTFLGNAGTAMRPLAALCLQAQDVT
      60      70      80      90      100      110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro --LTGEPRMKERPIGHLVDALRQGGAQIDYLEHEHYPLRLRG--GYQGGDITVDGSVSS
                120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro QFLTALLMMAPLAPQNSCIRIKGELVSRPYIDITLALMKSFGIEVRH-DNYQVFYLTGGS
                180          190          200          210          220          230

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      :. :. . : : : : : : : : : : . . : : : : : : : : : : : : :
Unipro VYRSPGEYLVVEGDASSASYFLAAAIRGGTVRVTGIGRHSVQGDIFADVLESMGATIRW
                240          250          260          270          280          290

                310          320          330          340          350
2MEPSP TETSVTVTGPPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADG-PTAIRDVASWRVK
      . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro GNDYIECS-----RASLHTIDMDMNHIPDAAMTIATTALFASGGTTTLRNIANWRVK
                300          310          320          330          340

                360          370          380          390          400          410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro ETDRLTAMATELRKVGATVIEGEDYLTVTPSARLSAARIGTYNDHRMAMCFALVALSDTP
                350          360          370          380          390          400

                420          430          440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : :
Unipro VTILDPQCTHKTFPDPFFARLSALSTPASPHSAP
                410          420          430

```

AROA_HAEIE | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Haemophilus influenzae* (strain PitTEE). |AA|432
Length = 432
initn: 748 initl: 215 opt: 613 Z-score: 1214.6 bits: 233.9 E(): 3e-60
Smith-Waterman score: 910; 47.126% identity (73.563% similar) in 435 aa overlap (5-436:2-422)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MEKITLAPISAVEGTINLP GSKSLSNRALLLAALAKGTTKVTNLLSDDIRHMLNAL
                10          20          30          40          50

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxx-xxxxxxx
      .:. . . . . : : : : : .: . . . . . : : : : : . .
Unipro  KALGVRYQLSDDKTICEVEGLGGTFNIQD---NLSLFLGNAGTAMRPLTAALCLKGKTES
              60          70          80          90          100          110

              120          130          140          150          160          170
2MEPSP  xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .: : :.:.:. .: .: .: .: . . . . . :. . . . . :. :.:.:.:.
Unipro  EIILTGEPRMKERPILHLVDALRQAGADIRYLENEGYPPLAIRN-KGIKGGKVKIDGSIS
              120          130          140          150          160          170

              180          190          200          210          220          230
2MEPSP  SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      :. . . . . :.:.:. .: .: :.:. .: .: :.:. .: .: :.:. .: .: :.:.
Unipro  SQFLTALLMSAPLAENDTEIEIIGELVSKPYIDITLAMMRDFGVKVENHH-YQKFQVKGN
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      :. :. . . :.:.:. .: .: .: .: . . . . . :.:. .: .: .: .: .: .:
Unipro  QSYISPKNYLVEGDASSASYFLAAGAIKGVKVTGIGKN-SIQGDRLFADVLEKMGAKIT
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      : : . . . . . :.:.:. .: .: :.:. .: .: :.:. .: .: :.:. .: .: :.:.
Unipro  WGEDFIQAE-----HAELNGIDMDMNHIPDAAMTIATTALFSNGETVIRNIYNWRVK
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  ETERMVAIRTELTKLGASVEEGPDYCIITPP--EKLNVTAIDTYDDHRMAMAFSLAACAE
      :.:.:. .: .: :.:. .: .: :. .: .: .: .: .: .: .: .: .: .: .: .:
Unipro  ETDRLTAMATELRKVGAEVEEGEDFIHIQPLPLNQFKHANIETYNDHRIAMCFSLIALSN
              350          360          370          380          390          400

              420          430          440
2MEPSP  VPVTIRDPGCTRKTFPDYFDVLSTFVK
      .: .: .: .: .: .:
Unipro  TPVTILDPKCTAKTFPTFFSEFEKICLRD
              410          420          430

```

AROHA_HAEI8 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Haemophilus influenzae* (strain 86-028NP). |AA|432
Length = 432
initn: 748 init1: 215 opt: 613 Z-score: 1214.6 bits: 233.9 E(): 3e-60
Smith-Waterman score: 910; 47.126% identity (73.563% similar) in 435 aa overlap (5-436:2-422)

```

              10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      .:. .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro  MEKITLAPISAVEGTINLPGSKSLSNRALLAALAKGTTKVTNLLDSDDIRHMLNAL
              10          20          30          40          50

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AROA_HAEIN | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Haemophilus influenzae*. |AA|432

Length = 432

initn: 753 init1: 215 opt: 609 Z-score: 1206.7 bits: 232.4 E(): 8.4e-60

Smith-Waterman score: 901; 46.789% identity (73.394% similar) in 436 aa overlap (5-437:2-423)

	10	20	30	40	50	60										
2MEPSP	MAGAE	EIVLQ	PIKEI	SGTVK	LPGSK	SLSNRILL	LAAL	SEGTT	VVDN	LLNS	EDVHY	MLGAL				
	:	:	:	:	:	:	:	:	:	:	:	:				
Unipro	MEKIT	LAPIS	AVEGT	INLPG	SKSLS	NRALL	LAAL	AKGTT	KVTN	LLDS	DDIRH	MNLAL				
	10	20	30	40	50											
	70	80	90	100	110											
2MEPSP	RTLGL	SVEAD	KAAK	RAVV	VVGCG	KFPVE	DAKEE	VQLFL	GNAG	IAMRS	Lxxxx	xxxxxxx				
	:	:	:	:	:	:	:	:	:	:	:	:				
Unipro	KALGV	RYQLS	DDKT	ICEVE	GLGGT	FNIQD	--N	LSLFL	GNAG	TAMR	PLTA	ALCLKG	KGTES			
	60	70	80	90	100	110										
	120	130	140	150	160	170										
2MEPSP	xYVLD	GVPR	MRER	PIGDL	VVGL	KQLGA	DVDC	FLGT	DCPP	VRVNG	IGGL	PGGK	VKL	SGSIS		
	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Unipro	EIILT	GEPR	MKER	PILHL	VDAL	RQAG	ADIR	YLENE	GYPPL	AIRN	-KGI	KGK	VK	IDGSIS		
	120	130	140	150	160	170										
	180	190	200	210	220	230										
2MEPSP	SQY	xxxxxxxxxxxx	GDVE	IEI	IDKL	ISIPY	VEMTL	RLMER	FGVKA	EHSD	SWDR	FYIK	GG			
	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Unipro	SQFLT	ALLMS	APLA	ENDT	EIEI	IGEL	VSKPY	IDITL	AMMR	DFGV	QVEN	HH	-YQK	FQVKGN		
	180	190	200	210	220	230										
	240	250	260	270	280	290										
2MEPSP	QKYK	SPKN	AYVE	GDASS	ASYFL	AGAAI	xxxxxxxxxxxx	SLQ	GDVK	FAEV	LEMM	GAKVT				
	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Unipro	QSYIS	PNKY	LVEG	DASS	ASYFL	AGAAI	KGKVK	VTGIG	KN	-SIQ	DRLF	ADVLE	KMGAKIT			
	240	250	260	270	280	290										
	300	310	320	330	340	350										
2MEPSP	WTETS	VTVT	GPPRE	PFG	RKHL	KAI	DVN	MN	KMPD	VAMT	LAV	VALF	ADGPT	AIRD	VAS	WRVK
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Unipro	WGED	FIQAE	-----	HAEL	NGID	MDM	NH	IPDA	AMT	IAT	TAL	FAN	SETV	IRNI	YN	NWRVK
	300	310	320	330	340	350										
	360	370	380	390	400	410										
2MEPSP	ETERM	VAIR	TELTK	L	GASV	EEGPD	YCIIT	PP--	EKL	NVT	AIDT	YDHR	MAMA	FSLA	ACAE	
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Unipro	ETDRL	TAMAT	ELRK	VGA	EV	EGGED	FIRI	QPL	PLNQ	FKH	ANI	ETY	NDHR	MAMC	FS	LIALSN
	350	360	370	380	390	400										
	420	430	440													
2MEPSP	VPVT	IRD	PGCTR	KTF	PDY	FDV	LST	FVKN								
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Unipro	TPVT	ILDP	KCTAK	T	FPT	F	NEF	EKIC	LKN							
	410	420	430													

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AROA_BLOPB | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Blochmannia pennsylvanicus* (strain BPEN). |AA|432
Length = 432
initn: 863 initl: 285 opt: 606 Z-score: 1200.7 bits: 231.3 E(): 1.8e-59
Smith-Waterman score: 872; 44.954% identity (70.872% similar) in 436 aa overlap (7-441:5-427)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEIS	GTVKLP	GSKSL	SNRILL
	:	:	:	:	:	:
Unipro	MNDFI	KLAPI	KIKIQ	GTHLP	SGSKS	SISNRA
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KAARAV	VVGCG	KFPVED	AKEEVQ
	:	:	:	:	:	:
Unipro	RNLGV	SYCLS	NNRKT	CEING	GIGGPI	QSKN--
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIG	DLVVGL	KQLGAD	VDCFLG
	:	:	:	:	:	:
Unipro	IVLTG	HPRMK	DRPIA	HLVDAL	RQGGA	RIEYME
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxx	xxxxxxx	GDVEIE	IIDKL	ISIPY	VEMTLR
	:	:	:	:	:	:
Unipro	QFLSS	VLMMT	PLAYRD	TLLIKV	DGALVS	RPYIDIT
	180	190	200	210	220	230
	250	260	270	280	290	300
2MEPSP	KYKSP	KNAYV	EGDASS	ASYFLA	GAAXxxxx	xxxxxxx
	:	:	:	:	:	:
Unipro	AYQSP	GDYLV	EGDASS	ASYFLA	ASAIRG	GTVRVIG
	240	250	260	270	280	290
	310	320	330	340	350	
2MEPSP	TETSV	TVTGT	GPPREP	FGRKHL	KAIDVN	NMNKMP
	:	:	:	:	:	:
Unipro	GDNYI	ECTRG	A-----	DLNAV	DLVDV	NNIPDA
	300	310	320	330	340	
	360	370	380	390	400	410
2MEPSP	ETERM	VAIRTE	LTKLG	ASVEEG	PDYCI	IITPPE
	:	:	:	:	:	:
Unipro	ESDRL	AAMATE	LKIGAE	IVEGYD	YLQITP	PFKIES
	350	360	370	380	390	400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFFPDYFDVLSTFVKN
      . . . . . : : : : : : : : .
Unipro  IIINNPKCTDKTFPDFFFTQLSSISVLQ
      410      420      430

```

AROAMANS | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Mannheimia succiniciproducens* (strain MBEL55E). |AA|433
Length = 433
initn: 791 initl: 253 opt: 589 Z-score: 1166.8 bits: 225.0 E(): 1.4e-57
Smith-Waterman score: 904; 47.166% identity (73.696% similar) in 441 aa overlap (5-440:2-428)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  MEKLTLPISHVEGTVNLPGSKSLSNRALLLAALAKGTTTRVTNLLDSDDVHRMLNAL
      10      20      30      40      50

```

```

      70      80      90      100      110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxx--xxxxxxx
      . : : . . . . . : : : : : . . . . : : : : : : : : : : . . .
Unipro  KQLGVNYSLSSEDKSVCEVQGLGKAFWQNG--LALFLGNAGTAMRPLTAALCLANADSV
      60      70      80      90      100      110

```

```

      120      130      140      150      160      170
2MEPSP xxxYVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGS
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  PAEIIILTGEPRMKERPIKHLVDALLQAGADVQYLEQEGYPPLAIRNTG-LKGGKVKIDGS
      120      130      140      150      160      170

```

```

      180      190      200      210      220      230
2MEPSP ISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
      . : : . . . . . : : : : : : : : : : : : : : : : : : : : :
Unipro  VSSQFLTALLMAAPMAERDTEIEIIGELVSKPYIDITLNMKIFAVDVDNQ-NYQRFVVK
      180      190      200      210      220      230

```

```

      240      250      260      270      280      290
2MEPSP GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      . : : . : : . : : : : : : : : : . : : : : : : : : : : :
Unipro  GNQQYQSPNIFLVEGDASSASYFLAAGAIKGVKRVTVGVGKN-SIQGDRLFDAEVLEKMGAK
      240      250      260      270      280      290

```

```

      300      310      320      330      340      350
2MEPSP VTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWR
      . : : . . . : : : : : : : : : : : : : : : : : : : : :
Unipro  ITWGEDYIEAE-----RGELNGIDMDMNHIPDAAMTIATTALFAQGETVIRNIYNWR
      300      310      320      330      340

```

```

      360      370      380      390      400      410
2MEPSP VKETERMVAIRTELTKLGASVEEGPDYCIITPP--EKLNVTAIDTYDDHRMAMAFSLAAC
      . : : . : : : : : : : : : : : : : : : : : : : : : : :
Unipro  VKETDRLSAMATELRKVGAEEVEGEDFIRIQPPASDQFKHAEIETYNDHRMAMCFALVAL
      350      360      370      380      390      400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          420          430          440
2MEPSP AEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      ... ::: :: :: :::: :. :.
Unipro SNTAVTICDPKCTAKTFPTFFDEFSAIATV
          410          420          430

```

AROA_IDILO | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Idiomarina loihiensis*. | AA | 429
Length = 429
initn: 777 init1: 354 opt: 584 Z-score: 1156.9 bits: 223.2 E(): 5e-57
Smith-Waterman score: 796; 42.890% identity (68.807% similar) in 436 aa overlap (7-440:5-424)

```

          10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  MVNSIHLEPRQHCRGTVTLPGSKSIANRALLMAALCQTPVILHNLLVSDDTSRMREAL
          10          20          30          40          50

```

```

          70          80          90          100          110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxx-
      ..::: : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  NALGVSFEDDKLITRVNGLGGGWNPASE-----LYLGNAGTAMRPLIAVLAATLKNEH
          60          70          80          90          100          110

```

```

          120          130          140          150          160          170
2MEPSP -xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSI
      . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  QAVVLKGDARMHERPVKHLIDAIQPRGAGVNYLGETGFPPLEMT--SGLKPGDFEIDGSV
          120          130          140          150          160          170

```

```

          180          190          200          210          220          230
2MEPSP SSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG
      :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  SSQFISALLMALPLLPGDSTLTLKGNVSRPYIELTLQMLSDFGISIKE-DSPQCYAIPG
          180          190          200          210          220

```

```

          240          250          260          270          280          290
2MEPSP GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
      : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  GQCYQSPGEYWVEGDASAASYWMAAALLGKGPVEIIGVGKNSIQGDKRFAEVIEAMGASV
          230          240          250          260          270          280

```

```

          300          310          320          330          340          350
2MEPSP TWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRV
      .. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  SYRKNSMTVSGT-----GSVQGIDQDFNDIPDAAMTVAPLALFANKPTTIRNVANWRV
          290          300          310          320          330          340

```



```

      360      370      380      390      400      410
2MEPSP KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEV
      ::::. . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro KETDRLHAMATELRKLGATVDEGEDFLRIEPLKHWHRHIAIDTYDDHRMAMCFSLVAFSSA
      350      360      370      380      390      400

      420      430      440
2MEPSP PVTIRDPGCTRKTFFPDYFDVLSTFVKN
      ::::: : : : : : : : : : : :
Unipro GVTINDPGCCAITYPDYFSEFSRLCHS
      410      420

```

		10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQPIKEIS	GTVKLPGSKSL	SNRILL	LAL	SEGTTVVDN	LLNS
	:
Unipro	MQELL	TTLQPIVRV	NGTIHLP	GSKSIS	NRALL	AAQALG	KTCLIN
		10	20	30	40	50	
		70	80	90	100	110	120
2MEPSP	RTLGL	SVEADKAAK	RAVVV	GCGGK	FPVEDA	KEEVQL	FLGNAG

Unipro	HKLGI	NYCLSID	RRSCEI	DGIGR	PLRVD	TALE---	LYLGN
	60	70	80	90	100	110	

130 140 150 160 170 180
 2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
 . . : : : : . . : : : . . : : . . : :
 Unipro -IITGDKRMKNRPIGHLVDALRQGSAQIHYLEKDNYPPLLLQ--GGFYNGDITIDCSLSS
 120 130 140 150 160 170

190 200 210 220 230 240
 2MEPSP QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
 : : : : : : : : :
 Unipro QFLTSLMMAPLASQDRCIFVKGRLVSKPYIDMTLAMMKSFGIVVQH-DQYKIFYIKGKS
 180 190 200 210 220 230

2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
 .:.:. . :.:.:.:.:.:.:.:.:. . . :.:.:.:.:.:.:.:.:. . :.:.:.
 Unipro QYRSPGHYLVEGDATNASYFLAAAAIRGGTVRVTGVGSNSIQGDIRFADILANMGAIIR
 240 250 260 270 280 290

Page 97 of 526

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      360      370      380      390      400      410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
      :::::  ::  ::::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Unipro ETNRLVAMATELRKVGAI VEGKEYLSIKPPNMFKIAKINTYDDHRIAMCFALVALSNVS
      350      360      370      380      390      400

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      .:: .: :: :::: .: .:
Unipro ITIVNPKCTYKTFPDPFFKLLKGISIT
      410      420

```

AROA_SHIDS | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shigella dysenteriae* serotype 1 (strain Sd197). |AA|424
Length = 424
initn: 955 initl: 398 opt: 573 Z-score: 1135.0 bits: 219.1 E(): 8.3e-56
Smith-Waterman score: 980; 49.312% identity (73.853% similar) in 436 aa overlap (5-440:2-421)

```

      10      20      30      40      50      60
2MEPSP MAGAEIIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :..... :..... :..... :..... :..... :.....
Unipro MESLTLQPIARVDGTINLP GSKSVSNRALLLAALAHGKTVLTNLLDSDVHRMLNAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      :::: . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro AALGVSYTLSADTRCEIIGNGGS LHAEGALE---LFLGNAGTAMRPLAAALCLGSNDI-
      60      70      80      90      100      110

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :: : :::: :::: :: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVGSVSS
      120      130      140      150      160      170

      190      200      210      220      230      240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEIENQH-YQQFVVKGGQ
      180      190      200      210      220

      250      260      270      280      290      300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      :::: . :::: :::: . . :::: :::: :::: :::: :::: :::: :::: ::::
Unipro SYQSPGTYLVEGDASSASYFLAAAIAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
      230      240      250      260      270      280

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE					
	. . . :	:	:	
Unipro	GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTLRNIYNWRVKE					
	290	300	310	320	330	340

	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVPV					
	:	:	:	:	:	:
Unipro	TDRLFAMATELRKVGAEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV					
	350	360	370	380	390	400

	430	440
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN	
	:: :: :: : . .	
Unipro	TILDPKCTAKTFPDYFEQLARIS	
	410	420

AROA_SHIFL|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shigella flexneri*. |AA|427
Length = 427
initn: 957 initl: 398 opt: 573 Z-score: 1134.9 bits: 219.1 E(): 8.3e-56
Smith-Waterman score: 982; 49.541% identity (73.624% similar) in 436 aa overlap (5-440:2-421)

	10	20	30	40	50	60
2MEPSP	MAGAEFIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL					
	:	:	:	:	:	:
Unipro	MESLTLQPIARVDGTINLPGSKSVSNRALLAALAHGKTVLTNLLDSDDVHRMLNAL					
	10	20	30	40	50	

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx					
	:	:	:	:	:	:
Unipro	TALGLSYTLSADRTRCEIIGNGGPLHAEGALE--LFLGNAGTAMRPLAALCLDSNDI-					
	60	70	80	90	100	110

	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS					
	: : :	: : . .	: : . .	:	:	:
Unipro	-VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS					
	120	130	140	150	160	170

	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ					
	:	:	:	:	:	:
Unipro	QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVKGGQ					
	180	190	200	210	220	

	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	:	:	:	:	:	:
Unipro	SYQSPGTYLVEGDASSASYFLAAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW					
	230	240	250	260	270	280

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTLRNINWVRVKE
      290          300          310          320          330          340

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPV
      : : . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
      350          360          370          380          390          400

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TILDPKCTAKTFPDYFEQLARISQAA
      410          420

```

AROA_SHISS | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shigella sonnei* (strain Ss046). |AA|427
Length = 427
initn: 954 initl: 398 opt: 573 Z-score: 1134.9 bits: 219.1 E(): 8.3e-56
Smith-Waterman score: 979; 49.312% identity (73.624% similar) in 436 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MESLTLQPIARVDGTINLPKSKSVSNRALLLAALAHGKTVLTNLLDSDDVHRMLNAL
      10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TALGVSYTLSADRTRCEIIGNGGPLHAEGALE--LFLGNAGTAMRPLAALCLGSNDI-
      60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVDDVGSVSS
      120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVKGGQ
      180          190          200          210          220

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        .:.:. .  :.:.:.:.:.:.:.:. .  .  .  :.:.:.:.:.:.:. . . . . :
Unipro  SYQSPGTYLVEGDASSASYFLAAAIAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
        230          240          250          260          270          280

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKE
        .  .  .  :  :.:.:.:. . . . . . . . . . . : . . . . . :.:.:.
Unipro  GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE
        290          300          310          320          330          340

```

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
        :.:. . . :.:.:.:. . . : . . . . . . . : . . . . . :.:.:.
Unipro  TDRLFAMATELRKVGAEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
        350          360          370          380          390          400

```

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  TILDPKCTAKTFPDYFEQLARISQAA
        410          420

```

AROAE | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Escherichia coli* O9:H4 (strain HS). |AA|427
Length = 427
initn: 952 initl: 398 opt: 573 Z-score: 1134.9 bits: 219.1 E(): 8.3e-56
Smith-Waterman score: 977; 49.312% identity (73.394% similar) in 436 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        :.:.:. . . :.:.:.:. . . :.:.:. . . :.:.:. . . :.:.:. . . :.:.:.
Unipro  MESLTLQPIARVDGTINLP GSKSVSNRALLAALAHGKTVLTNLLDSDDVHRMLNAL
        10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
        :.:. .  .  :. . . :. . . :. . . :. . . . :. . . . :. . . .
Unipro  TGLGVSYTSLADTRCEIIGNGGPLHAEGALE---LFLGNAGTAMRPLAAALCLGSNDI-
        60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        :. : . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
        120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :. . . . . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEIENQH-YQQFVVVKGGO
        180          190          200          210          220

```

	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	. . . : . : : : : : : : : : : . . : : : . :					
Unipro	SYQSPGYTLVEGDASSASYFLAAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGTICW					
	230	240	250	260	270	280
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPEPFGRKHLKAIDVNMNKMPPDVAMT LAVVALFADGPTAIRDVASWRVKE					
	. .. :					
Unipro	GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE					
	290	300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMVAI RTELT KLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMA FSLAACAEVPV					
	: . . : . : . . . :					
Unipro	TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNF AEIATYNDHRMAMCFSLVALSDTPV					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
	:: :: : : : : : : : . .					
Unipro	TILDPKCTAKTFPDYFEQLARISQAA					
	410	420				

		10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQPIKEIS	GTVKLPGSKSL	SNRILL	LAALSEG	TTVVDNLLN	SEDEVHYMLGAL
	:
Unipro	MES	LT	LQPIARVDGT	INLPGSKSV	SNRALL	LAALAHGKT	VL
		10	20	30	40	50	
		70	80	90	100	110	120
2MEPSP	RTLGLS	VEADKAAKRA	VVVGCGGK	FPVEDAKEE	VQLFLGNAGI	AMRSL	xxxxxxxxxxxx

Unipro	AALGV	SYTLSADR	TRCEIIGNGG	SLHAEGALE	---	LFLGNAGT	AMRPLAAALCLGSNDI-
	60	70	80	90	100	110	
		130	140	150	160	170	180
2MEPSP	YVLDG	VPRMRERPI	GD	LVVLGKQLG	ADVDCFLGT	DCPPVRVNGI	GG

Unipro	-VLTGE	PRMKERPI	GH	LV	DALRLGRA	KITYLEQEN	YPPLRLQG--GFTGGNVDVDG
	120	130	140	150	160	170	

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      . . . . . : . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEIENQH-YQQFVVKGGQ
                180          190          200          210          220

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      . : : : . : : : : : : : : : : : . . : : : : : : : : : : : : : : :
Unipro SYQSPGTYLVEGDASSASYFLAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
                230          240          250          260          270          280

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTLRNIYNWRVKE
                290          300          310          320          330          340

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
      : : . : : : : : : : : : : : : : : : . : : : : : : : : : : : : : :
Unipro TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
                350          360          370          380          390          400

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : . .
Unipro TILDPKCTAKTFPDYFEQLARISQAA
                410          420

```

AROA_SHISO | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shigella sonnei*. |AA|427
Length = 427
initn: 953 initl: 398 opt: 573 Z-score: 1134.9 bits: 219.1 E(): 8.3e-56
Smith-Waterman score: 978; 49.312% identity (73.394% similar) in 436 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEIIVLQPIKEISGTVKLP GSKSLNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MESLTLQPIARVDGTINLP GSKSVSNRALLLAALAHGKTVLTNLLDSDDVRLHMLNAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . : : . : . : : : . : : : : : : : : : : : : : : .
Unipro TALGVSYTLSADRTCEIIGNGGPLHAEGALE---LFLGNAGTAMRPLAAALCLGSNDI-
                60          70          80          90          100          110

```

	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS					
	:: : ::::: :: :: ::::: :: ::::: :::::					
Unipro	-VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVDVDGVS					
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQ					
	:: :: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::					
Unipro	QFLTALLMNAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFTGVEIENQH-YQQFVVKGGQ					
	180	190	200	210	220	
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	::::: . ::::: ::::: ::::: ::::: ::::: ::::: :::::					
Unipro	SYQSPGYTLVEGDASSASYFLAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW					
	230	240	250	260	270	280
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGGRKHLKAIDVNMNMKMPDVAMTLAVVALFADGPTAIRDVASWRVKE					
	. . . : : ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::					
Unipro	GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTLRNIYNWRVKE					
	290	300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVPV					
	::: . ::::: ::::: ::::: ::::: ::::: ::::: . : ::::: ::::: :::::					
Unipro	TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
	:: :: :: ::::: . .					
Unipro	TILDPKCTAKTFPDYFEQLARISQAA					
	410	420				

10 20 30 40 50 60
 2MEPSP MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL LLAALSEGTTVVDNLLNS EDVHYMLGAL
 : : : : : : :
 Unipro MESLTLQPIARVDGTINLP GSKSVSNRALL LLAALAHGKT VLTNNLLSDDDVRH MLNAL
 10 20 30 40 50

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFVPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx					
	. : . : . : . : . : . : . : . : . : . : . : . :					
Unipro	TALGVSYTLSADRTRCEIIGNGGPLHAEGALE--LFLGNAGTAMRPLAALCLGSNDI-					
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDGVPMRERPIGDLLVVLGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS					
	: :					
Unipro	-VLTGEPRMKERPIGHLVDALELRGGAKITYLEQENYPPLRLQG--GFTGGNVDDVDSVSS					
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ					
	: : . :					
Unipro	QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVKGGQ					
	180	190	200	210	220	
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLMMGAKVTW					
	. : . : : : : : : : : : : : . . : : : : : : : : : : : : :					
Unipro	SYQSPGYTLVEGDASSASYFLAAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW					
	230	240	250	260	270	280
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPEPFGRKHCLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE					
	. . :					
Unipro	GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNINWRVKE					
	290	300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMVAIKTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVPV					
	: . :					
Unipro	TDRLFAMATELRKVGAEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
	: : : : : : : : . .					
Unipro	TILDPKCTAKTFPDYFEQLARISQAA					
	410	420				

AROAO_SHIBS|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shigella boydii* serotype 4 (strain Sb227). |AA|427

Length = 427

```
initn: 954 init1: 398 opt: 573  Z-score: 1134.9  bits: 219.1  E(): 8.3e-56
```

Smith-Waterman score: 979; 49.312% identity (73.624% similar) in 436 aa overlap (5-440:2-421)

10 20 30 40 50 60
 2MEPSP MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL LLAALSEGTTVVDNLLNS EDVHYMLGAL
 : : : : : : :
 Unipro MESLTLQP IARVDGTINLP GSKSVSNRALL LAALAHGKT VLTNNLLD SDDVRHMLNAL
 10 20 30 40 50

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx					
	.:.: . :. .: :. .: : : :.:.: :.: :.. .					
Unipro	TALGVSYTLSADTRCEIIGNGGPLHAEGALE--LFLGNAGTAMRPLAAALCLGSNDI-					
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLGVPVRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS					
	:: : :.:.: :. :. :. . . :.:.: :. :.: :.:.: :					
Unipro	-VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS					
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ					
	:. .. . :. :.: :.: :.:.: :.: :. :.: :. :.: :					
Unipro	QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEIENQH-YQQFVVKGGQ					
	180	190	200	210	220	
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	.:.: . :.:.: :. . :.:.: :.: :.: :.: :.: :.: :.: :. :					
Unipro	SYQSPGYTLVEGDASSASYFLAAAIAKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW					
	230	240	250	260	270	280
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKE					
	. .. : : :.:.: :.:.: :.:.: :.: :.: :.: :.: :.: :.: :					
Unipro	GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE					
	290	300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADITYDDHRMAMAFSLAACAEVPV					
	:.. . :. :.:.: :. :.:.: :. :.:.: :. :.:.: :. :.:.: :.: :					
Unipro	TDRLFAMATELRKVGAEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
	:: :. :. :.:.: :. .					
Unipro	TILDPKCTAKTFPDYFEQLARISQAA					
	410	420				

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AROA_ECOLC | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Escherichia coli* (strain ATCC 8739 / DSM 1576 / Crooks). | AA|427

Length = 427

initn: 952 init1: 398 opt: 573 Z-score: 1134.9 bits: 219.1 E(): 8.3e-56

Smith-Waterman score: 977; 49.312% identity (73.394% similar) in 436 aa overlap (5-440:2-421)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEI	SGTVK	LPGSK	SLSNR

Unipro	MESLT	LQPIA	RVDGT	INLPG	SKSVS	NRALL
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KAAKR	AVVVG	CGGKF	PVEDA

Unipro	TGLGV	SYTSL	ADRTR	CEIIG	NGGGL	HAEGL
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIG	DLVVG	LKQLG	ADVDC

Unipro	-VLTG	EPRMK	ERPIG	HLVDL	ARLGG	AKITY
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxx	xxxxxxx	GDVEI	EIIDK	LISIP	YEMTL

Unipro	QFLT	ALLMT	APLAP	EDTVI	RIKGD	LVSKE
	180	190	200	210	220	
	250	260	270	280	290	300
2MEPSP	KYKSP	KNAVE	GDASS	ASYFL	AGAAI	xxxxxxxx

Unipro	SYQSP	GTYL	VEGD	ASSA	SYFL	AAAAI
	230	240	250	260	270	280
	310	320	330	340	350	360
2MEPSP	TETSV	TVTGP	PREPF	GRKHL	KAIDV	NMNKM

Unipro	GDDY	ISCT	-----	RGEL	NAID	MDMNH
	290	300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMV	AIRTE	LTKLG	ASVEE	GPDYC	IITP

Unipro	TDRLF	AMATE	LKRVG	AEVEE	GHDYI	IRITP
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDP	GCTRK	TFPDY	FDVL	STFV	KN

Unipro	TILDP	KCTAK	TFPDY	FEQL	ARIS	QAA
	410	420				

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AROAE057 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Escherichia coli* O157:H7. |AA|427
Length = 427
initn: 954 initl: 398 opt: 573 Z-score: 1134.9 bits: 219.1 E(): 8.3e-56
Smith-Waterman score: 979; 49.312% identity (73.624% similar) in 436 aa overlap (5-440:2-421)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEIS	GTVKLP	GSKSL	SNRILL

Unipro	MESLT	LQPIAR	VDGTIN	LPGSK	SVSNRA	LLAALAH
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEADK	AAKRAV	VVGCGG	KFPVED	AKEEVQ

Unipro	TALGV	SYTSLA	DRTRCE	IIGNGG	PLHAEGA	LE--LFL
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIGDL	VVGLK	QLGADV	CDCLGT

Unipro	-VLTGE	PRMKER	PIGHLV	DALRLG	GAKITY	LEQENYP
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxx	xxxxxx	GDVEIE	IIDKLIS	IPYVEM	TLRLMER

Unipro	QFLTAL	LMTAPL	APEDTV	IRIKGDL	VSKPYID	ITLNLMT
	180	190	200	210	220	
	250	260	270	280	290	300
2MEPSP	KYKSP	KNAYVE	GDASSA	SYFLAG	AAIxxxx	xxxxxxxx

Unipro	SYQSP	GTYLVE	GDASSA	SYFLAAA	AIKGGT	VKVTGIG
	230	240	250	260	270	280
	310	320	330	340	350	360
2MEPSP	TETSV	TVTGT	GPPREP	FGRKHL	KAIDVN	MNMKMP

Unipro	GDDYIS	CT-----	RGELNA	IDMDMN	HIPDAAM	TIATAAL
	290	300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMV	AIRTEL	TKLGAS	VEEGPD	YCIITP	PPEKLN

Unipro	TDRLF	FAMATE	LKVGAE	VEEGHD	YIRITP	PPEKLN
	350	360	370	380	390	400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        :: :: :: ::::::. :.
Unipro  TILDPKCTAKTFPDYFEQLARISQAA
                410          420

```

AROAE | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Escherichia coli* O139:H28 (strain E24377A / ETEC). | AA|427
Length = 427
initn: 954 initl: 398 opt: 573 Z-score: 1134.9 bits: 219.1 E(): 8.3e-56
Smith-Waterman score: 979; 49.312% identity (73.624% similar) in 436 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        :...::: :...::: :...::: :...::: :...::: :...:::
Unipro   MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLSDDDVRHMLNAL
                10          20          30          40          50

```

```

                70          80          90         100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
        :...: . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  TALGVSYTSLADTRCEIIGNGGPLHAEGALE---LFLGNAGTAMRPLAAALCLGSNDI-
                60          70          80          90         100         110

```

```

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        :: : ::::: : : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
                120         130         140         150         160         170

```

```

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :. . . . . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVVKGQ
                180         190         200         210         220

```

```

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        :...: . :...: :...: :...: :...: :...: :...: :...: :...: :...: :...:
Unipro  SYQSPGYTLVEGDASSASYFLAAAIAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
                230         240         250         260         270         280

```

```

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
        . . : : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE
                290         300         310         320         330         340

```

```

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVPV
        :.. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
                350         360         370         380         390         400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      :: :: :: ::::: . .
Unipro TILDPKCTAKTFPDYFEQLARISQAA
                410          420

```

AROAS_HIF8 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shigella flexneri* serotype 5b (strain 8401). |AA|427
Length = 427
initn: 957 initl: 398 opt: 573 Z-score: 1134.9 bits: 219.1 E(): 8.3e-56
Smith-Waterman score: 982; 49.541% identity (73.624% similar) in 436 aa overlap
(5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :...::: . . . . . :...::: :...::: . . . . . :...::: :...:::
Unipro  MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVHRMLNAL
                10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      .::: . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro  TALGLSYTSLADTRCEIIGNGGPLHAEGALE--LFLGNAGTAMRPLAAALCLDSNDI-
                60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :: : :...::: : : . . : : . . : :...::: : : : : : :...:::
Unipro  -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
                120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . . : . . . . . : . . . . . : . . . . . : . . . . . :
Unipro  QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVKGGQ
                180          190          200          210          220

```

```

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      .::: . :...::: : . . : :...::: : : . . : :...::: : : . . :
Unipro  SYQSPGTYLVEGDASSASYFLAAAIAKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
                230          240          250          260          270          280

```

```

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . : : : : : : : : : : : : : : : : : : : : : :
Unipro  GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTLRNIYNWRVKE
                290          300          310          320          330          340

```

	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT	TAIDTYDDHRMAMAFSLAACAEVPV				
	::: .: :::	:.::: :::	::: :::	. : :::	:.::: :::	:.::: :::
Unipro	TDRLFAMATELRKVGAEVEEGHDYIRITPPEKLNFAE	IATYNDHRMAMCFSLVALSDTPV				
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
	::: ::: :::	:.::: :::				
Unipro	TILDPKCTAKTFPDYFEQLARISQAA					
	410	420				

```

      10      20      30      40      50      60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :...:::  ..:::.....:::  :::::.....  ..:::.....:::
Unipro  MLESITLQPIALVNGTVNLPGSKSVSNRALLLAALAEGTTQLNNVLDSDDIRHMLNAL
      10      20      30      40      50

      70      80      90      100     110     120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      ..::..  .  :  :  :  :  :  .  .  :::::  :  :  :  .
Unipro  QALGVNFRLSADRTCCEVDGLGGKLVAE---QPLSLFLGNAGTAMRPLAAVLCLGNSDI-
      60      70      80      90      100     110

```

130 140 150 160 170 180
 2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
 ::
 Unipro -VLTGEPRMKERPIGHLVDALRQGGAQIDYLEQENYPPLRLR--GGFRGGELTVDGRVSS
 120 130 140 150 160 170

190 200 210 220 230 240
 2MEPSP QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
 : : : : : :
 Unipro QFLTALLMTAPLAEQDTTIRIMGDLVSKPYIDITLHLMKAFGIDVGH-ENYQIFHIKGGQ
 180 190 200 210 220 230

MEPSP 250 260 270 280 290 300
KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
.:. . :.:
Unipro TYRSPGYLVEGDASSASYFLAAAAIKGGTVRVTGIGKKSVQGDTKFADVLEKMGAKEWTW
 240 250 260 270 280 290

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro GDDYIECS-----RGELQGIDMDMNHIPDAAMTIATTALFATGPTTIRNIYNWRVKE
          300          310          320          330          340

          370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVPV
      : . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TDRLTAMATELRKVGAEVEEGEDYIRVVPPVQLTAADIGTYDDHRMAMCFSLVALSDTPV
          350          360          370          380          390          400

          430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : :
Unipro TILDPKCTAKTFPDYFEQFARLSQLA
          410          420

```

AROA_YERPS | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Yersinia pseudotuberculosis*. |AA|428
Length = 428
initn: 852 initl: 386 opt: 571 Z-score: 1130.9 bits: 218.4 E(): 1.4e-55
Smith-Waterman score: 981; 48.961% identity (75.058% similar) in 433 aa overlap (5-437:3-419)

```

          10          20          30          40          50          60
2MEPSP MAGAEFIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro MLESLTLQPIALVNGTVNLPGSKSVSNRALLLAALAEQTTQLNNVLDSDDIRHMLNAL
          10          20          30          40          50

          70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QALGVNFRLSADRTCCEVDGLGKLVAE---QPLSLFLGNAGTAMRPLAAVLC LGNSDI-
          60          70          80          90          100          110

          130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro -VLTGEPRMKERPIGHLVDALRQGAQIDYLEQENYPPLRLR--GGFRGGELTVDGRVSS
          120          130          140          150          160          170

          190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMTAPLAEQD TTIRIMGDLVSKPYIDITLHLMKAFGIDVGH-ENYQIFHIKGGQ
          180          190          200          210          220          230

          250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TYRSPGTYLVEGDASSASYFLAAAIAKGGTVRVTGIGKKSQVQGTKFADVLEKMGAKVTW
          240          250          260          270          280          290

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro GDDYIECS-----RGELQGIDMDMNHIPDAAMTIATTALFATGPTTIRNIYNWRVKE
          300          310          320          330          340

          370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPV
      : . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TDRLTAMATELRKVGAEVEEGEDYIRVVPVQLTAADIGTYDDHRMAMCFSLVALSDTPV
          350          360          370          380          390          400

          430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : :
Unipro TILDPKCTAKTFPDYFEQFARLSQLA
          410          420

```

AROA_YERPA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Yersinia pestis* bv. *Antiqua* (strain *Antiqua*). |AA|428
Length = 428
initn: 850 initl: 385 opt: 570 Z-score: 1128.9 bits: 218.0 E(): 1.8e-55
Smith-Waterman score: 979; 48.961% identity (75.058% similar) in 433 aa overlap (5-437:3-419)

```

          10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Unipro MLESLTLQPIALVNGTVNLPGSKSVSNRALLLAALAEQTTQLNNVLDSDDIRHMLNAL
          10          20          30          40          50

          70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : . . . . : . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QALGVDFRLSADRTCCEVDGLGGKLVAE---QPLSLFLGNAGTAMRPLAAVLCLGNSDI-
          60          70          80          90          100          110

          130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro -VLTGEPRMKERPIGHLVDALRQGGAQIDYLEQENYPPLRLR--GGFRGGELTVDGRVSS
          120          130          140          150          160          170

          190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMTAPLAEQD'TTIRIMGDLVSKPYIDITLHLMKAFGIDVGH-ENYQIFHIKGGQ
          180          190          200          210          220          230

```


	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	:.:. . : : : : : : : : : : : : : : . . :.:.:.:.:.:.:.:. : : : : :					
Unipro	TYRSPGTYLVEGDASSASYFLAAAAIKGGTVRVTGIGKKSVDGDTKTFADVLEKMGAKVTW					
	240	250	260	270	280	290
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGFRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE					
	. . . : .:.:.:.:.:.:.:. : : : : : : : : : : : : : : : : : : :					
Unipro	GDDYIECS-----RGELQGIDMDMNHIPDAAMTIATTALFATGPTTIRNIYNWRVKE					
	300	310	320	330	340	
	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVVPV					
	:.:.:.:. :					
Unipro	TDRLTAMATELRKVGAEEVEEGEDYIRVVPPLQLTAADIGTYDDHRMAMCFSLVALSDTPV					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
	:: : : : : : : : .					
Unipro	TILDPKCTAKTFPDYFEQFARLSQLA					
	410	420				

		10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQPIKEIS	GTVKLPGSKSL	SNRILL	LAALSEG	TTVVDNLLN	SEDVHYMLGAL
	:	:	:	:	:	:	:
Unipro	MLES	LTQLPIALV	NGTVNLPGSK	SVSNRALL	LAALAE	GTTQLNNV	LDSDDIRHMLNAL
		10	20	30	40	50	
		70	80	90	100	110	120
2MEPSP	RTLGLS	VEADKAAKRA	VVVGCGGK	FPVEDAKE	EEVQLFLGN	NAGIAMRSL	xxxxxxxxxxxx
	:	:	:	:	:	:	:
Unipro	QALGV	DFRLSADRT	CCEVDGLG	GKLVAE--	QPLSLFLGN	NAGTAMRPL	AAVLCCLGNSDI-
	60	70	80	90	100	110	
		130	140	150	160	170	180
2MEPSP	YVLDG	VPRMRERPI	GDVVG	LKQLGADV	DCFLGTDC	PPVRVNGI	GGGLPGGKVKLSGSISS
	:	:	:	:	:	:	:
Unipro	-VLTG	EPRMKERPI	GHVLDALR	QGGAQIDY	LEQENYP	PLRLR--	GGFRGGELTVDGRVSS
	120	130	140	150	160	170	

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      . . . . . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Unipro QFLTALLMTAPLAEQDTTIRIMGDLVSKPYIDITLHLMKAFGIDVGH-ENYQIFHIKGGQ
                180          190          200          210          220          230

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Unipro TYRSPGTYLVEGDASSASYFLAAAIAIKGGTVRVTGIGKKSQVQDGTKFADVLEKMGAKVTW
                240          250          260          270          280          290

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . . . . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Unipro GDDYIECS-----RGELQGIDMDMNHIPDAAMTIATTALFATGPTTIRNIYNWRVKE
                        300          310          320          330          340

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
      : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Unipro TDRLTAMATELRKVGAEVEEGEDYIRVVPPLQLTAADIGTYDDHRMAMCFSLVALSDTPV
                350          360          370          380          390          400

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Unipro TILDPKCTAKTFPDYFEQFARLSQLA
                410          420

```

AROA_YERPN|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Yersinia pestis* bv. *Antiqua* (strain Nepal516). |AA|428
Length = 428
initn: 850 initl: 385 opt: 570 Z-score: 1128.9 bits: 218.0 E(): 1.8e-55
Smith-Waterman score: 979; 48.961% identity (75.058% similar) in 433 aa overlap (5-437:3-419)

```

                10          20          30          40          50          60
2MEPSP MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Unipro MLESLTLQPIALVNGTVNLPGSKSVSNRALLLAALAEGTTQLNNVLDSDDIRHMLNAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Unipro QALGVDFRLSADRTCCEVDGLGGKLVAE---QPLSLFLGNAGTAMRPLAAVLCLGNSDI-
        60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP YVLGDGVP RMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Unipro -VLTGEPRMKERPIGHLVDALRQGGQAQIDYLEQENYPPLRLR--GGFRGGELTVDGRVSS
                120          130          140          150          160          170

```

AROAE000000030|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Enterobacter* sp. (strain 638). |AA|427
Length = 427
initn: 900 initl: 376 opt: 567 Z-score: 1123.0 bits: 216.9 E(): 3.9e-55
Smith-Waterman score: 997; 49.312% identity (75.688% similar) in 436 aa overlap (5-440:2-421)

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :: : :::::::::: :: :.: :.: :.: :. . :.: :.: :.: :.: :.: :.: :.: :.:
Unipro -VLTGEPRMKERPIGHLVDALRQGGAQIDYLEQKNYPPVRLRG--GFTGGNVEVDGSVSS
                120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . . :. :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro QFLTALLMTAPLAPQDTVISIKGDLVSKPYIDITLHLMKTFGVEVENQ-AYQRFVVRGAQ
                180          190          200          210          220

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      .: : : : : :.: :.: :.: :.: :. . :.: :.: :.: :.: :.: :.: :.: :.:
Unipro QYQSPGNLYVEGDASSASYFLAAGAIKGGTVKVTGIGRNSVQGDIFADVLEKMGAVVTW
                230          240          250          260          270          280

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . . : . :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro GDDFISCT-----QGELNAVDMMDMNHIPDAAMTIATAALFAKGTTLRNINRVRKE
                290          300          310          320          330          340

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIDTYDDHRMAMAFSLAACAEVPV
      :.: :. :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro TDRLFAMATELRKVGAQVEEGEDYIRVTPPAKLQVAEIGTYNDHRMAMCFSLVALSDTPV
                350          360          370          380          390          400

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      :: : : : :.: :.
Unipro TILDPKCTAKTFPDYFEQLARISTLA
                410          420

```

AROA_YERE8 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Yersinia enterocolitica* serotype O:8 / biotype 1B (strain 8081). |AA|428
Length = 428
initn: 852 initl: 388 opt: 565 Z-score: 1119.0 bits: 216.2 E(): 6.5e-55
Smith-Waterman score: 1001; 49.771% identity (75.229% similar) in 436 aa
overlap (5-440:3-422)

```

                10          20          30          40          50          60
2MEPSP MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro MLESLTLHPIALINGTVNLPGSKSVSNRALLLAALAE GTTQLNNLLDSDDIRHMLNAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      ..: :. . . :. :.: :.: :.: :. . :.: :.: :.: :.: :.: :.: :.:
Unipro QALGVKYRLSADRTRCEVDGLGGKLVAE---QPLELFLGNAGTAMRPLAALCLGKNDI-
                60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        :: : :::::::::: :: :.: :.: :.: :. . :.: :.: :. :. :.: :.: :.: :.:
Unipro  -VLTGEPRMKERPIGHLVDALRQGGAQIDYLEQENYPPLRIRG--GFRGGKLTVDGSVSS
                120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQ
        :. . . . . :.: :.: :.: :.: :.: :.: :.: :.: :. :. :. :. :.: :.: :.:
Unipro  QFLTALLMTAPLAEQDTEIQIQGELVSKPYIDITLHLMKAFGVDVVH-ENYQIFHIKGGQ
                180          190          200          210          220          230

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro  TYRSPGIYLVEGDASSASYFLAAAIAKGGTVRVTGIGKQSVQGDTKFADVLEKMGAKISW
                240          250          260          270          280          290

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
        . . . . . :. :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro  GDDYIECS-----RGELQGIDMDMNHIPDAAMTIATTALFADGPTVIRNIYNWRVKE
                300          310          320          330          340

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAIEVPV
        :.: :. :. :. :.: :.: :.: :.: :.: :. :. :. :. :.: :.: :.: :.: :.:
Unipro  TDRLSAMATELRKVGAEVEEGQDYIRVPPAQLIAAEIGTYNDHRMAMCFSLVALSDTPV
                350          360          370          380          390          400

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        :: :. :. :. :.: :.: :.
Unipro  TILDPKCTAKTFPDYFEQLARLSQIA
                410          420

```

AROA_KLEP7 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; *Klebsiella pneumoniae* subsp. *pneumoniae* (strain ATCC 700721 / MGH 78578). |AA|427

Length = 427

initn: 915 initl: 397 opt: 564 Z-score: 1117.0 bits: 215.8 E(): 8.3e-55

Smith-Waterman score: 990; 49.541% identity (75.688% similar) in 436 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro  MESLTLQPIARVEGTVNLPGSKSVSNRALLLAALARGTTVLTNLLSDDDVRHMLNAL
                10          20          30          40          50

```

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFVPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx					
	. : : . : . : . : . : . : . : . : . : . : .					
Unipro	SALGVQYTLSDADRTRCEVTGNNG--PLRSAAA-LELFLGNAGTAMRPLAALCLGSNDI-					
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDGVPMRERPIGDLLVVLGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS					
	: :					
Unipro	-VLTGEPRMKERPIGHLVDALRQGGAQIDCLEQENYPPLRLRG--GFQGGNVEVDGVS					
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ					
	: : . :					
Unipro	QFLTALLMTAPLAPQDTVIVIKGDLVSKPYIDITLHLMKFTFGVEVDNQ-SYQRFVVRGKQ					
	180	190	200	210	220	
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxXXXXXXXXXXXXXSLQGDVKFAEVLMMGAKVTW					
	. : : : . : : : : : : : : : : : . . : : : : : : : : : : : : : :					
Unipro	QQSPGDYLVEGDASSASYFLAAGAIAKG GTVKVTGIGRNSVQGDIRFADVLEKM GATV TW					
	230	240	250	260	270	280
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPPREPFGRKH LKAIDVN MNKMPDVAM TLAVVALFADGP TAIRDVASWRVKE					
	. . . : . :					
Unipro	GDDFIACT-----HGELKAVDMDMNHIPDAAMTIATAALFAQGTTTLRNINWRVKE					
	290	300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMA FSLAACAEVPV					
	: . . : . : : : : : : : : : : : : : . : : : : : : : : : : : :					
Unipro	TDRLFAMATELRKVGAEEVEEGEDYIRITPPAKLK YAEIGTYNDHRMAMCFSLVALSDTPV					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFFPDYFDVLSTFVK N					
	: : : : : : : : . .					
Unipro	TILDPKCTAKTFPDYFEQLARISTLA					
	410	420				

AROAO_EC05L|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Escherichia coli* O6:K15:H31 (strain 536 / UPEC). |AA|427

Length = 427

```
initn: 996 init1: 388 opt: 563  Z-score: 1115.0  bits: 215.4  E(): 1.1e-54
```

Smith-Waterman score: 981; 49.083% identity (74.312% similar) in 436 aa overlap (5-440:2-421)

10 20 30 40 50 60
 2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLALLALSEGTTVVDNLLNSEDVHYMLGAL
 :.....::::::::
 Unipro MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVHRHMLNAL
 10 20 30 40 50

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx					
	.:.: . :. .: :. . :.: :.: :.: :.: :.: :.: :.: :.: :.					
Unipro	TALGVSYTLSADTRCEIIGNGGPLHAESARE--LFLGNAGTAMRPLAAALCLGSNDI-					
	60	70	80	90	100	110

	130	140	150	160	170	180
2MEPSP	YVLGVPVRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS					
	:: : :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.					
Unipro	-VLTGEPRMKERPIGHLVDALRQGGAKITYLEQENYPPLRLQG--GFTGGNVDDVDSVSS					
	120	130	140	150	160	170

	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ					
	:. :. :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.					
Unipro	QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEIENQH-YQQFVVKGGQ					
	180	190	200	210	220	

	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	.:.: . :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.					
Unipro	SYQSPGYTLVEGDASSASYFLAAAIRGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW					
	230	240	250	260	270	280

	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKE					
	. .. : : :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.					
Unipro	GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE					
	290	300	310	320	330	340

	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVPV					
	:.. . :. :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.					
Unipro	TDRLFAMATELRKVGAEVEEGHDFIRITPPEKLFKFAEIATYNDHRMAMCFSLVALSDTPV					
	350	360	370	380	390	400

	430	440
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN	
	:: :. :. :.: :.: :.: :.: :.	
Unipro	TILDPKCTAKTFPDYFEQLARISQPG	
	410	420

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

ARO_A_ECOK1 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Escherichia coli* O1:K1 / APEC. |AA|427

Length = 427

initn: 996 init1: 388 opt: 563 Z-score: 1115.0 bits: 215.4 E(): 1.1e-54

Smith-Waterman score: 981; 49.083% identity (74.312% similar) in 436 aa overlap (5-440:2-421)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEI	SGTVK	LPGSK	SLSNR

Unipro	MESLT	LQPIA	RVDGT	INLPG	SKSVS	NRALL
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KAAKR	AVVVG	CGGKF	PVEDA

Unipro	TALGV	SYTTL	SADRTR	CEIIG	NGGGL	HAESAR
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIGD	LVLVGL	KQLGAD	VDCFL

Unipro	-VLTG	EPRMK	ERPIGHL	VDALR	QGGAK	ITYLEQ
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxx	xxxxxxx	GDVEI	EIIDKL	ISIPY	EMTLRL

Unipro	QFLT	ALLMT	APLAP	EDTVI	RIKGD	LVS
	180	190	200	210	220	
	250	260	270	280	290	300
2MEPSP	KYKSP	KNAAY	VEGDAS	ASYFL	AGAAI	xxxxxxxx

Unipro	SYQSP	GTYL	VEGDAS	ASYFL	AAAAI	RGGTV
	230	240	250	260	270	280
	310	320	330	340	350	360
2MEPSP	TETSV	TVTGP	PREPF	GRKHL	KAIDV	NMNK

Unipro	GDDY	ISCT	-----	RGEL	NAID	MDM
	290	300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMV	AIRTE	LTKLG	ASVEE	GPDYCI	IITPPE

Unipro	TDRLF	AMATE	LKRVG	AEVEE	GHDFI	RITPPE
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDP	GC	TRKTF	PDYF	DVLS	TFVKN

Unipro	TILDP	KCTAK	TFPDY	FEQL	ARIS	QPG
	410	420				

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AROAE_ECOUT|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Escherichia coli* (strain UTI89 / UPEC). |AA|427
Length = 427
initn: 996 initl: 388 opt: 563 Z-score: 1115.0 bits: 215.4 E(): 1.1e-54
Smith-Waterman score: 981; 49.083% identity (74.312% similar) in 436 aa overlap (5-440:2-421)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEI	SGTVK	LPGSK	SLSNR

Unipro	MESLT	LQPIA	RVDGT	INLPG	SKSVS	NRALL
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KAARK	RAVVV	GCGGK	FPVED

Unipro	TALGV	SYTSL	ADRTR	CEIIG	NGGPL	HAESAR
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIG	DLVVG	LKQLG	ADVDC

Unipro	-VLTG	EPRMK	ERPIG	HLVDL	ARQGG	AKITYL
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxx	xxxxxxxx	GDVEI	EIIDK	LISIP	YVEMTL

Unipro	QFLT	TALLM	TAPLA	PEDTV	IRIKG	DLVSK
	180	190	200	210	220	
	250	260	270	280	290	300
2MEPSP	KYKSP	KNAYV	EGDASS	ASYFL	AGAAI	xxxxxxxx

Unipro	SYQSP	GTYLVE	GDASS	ASYFL	AAAAI	RGGTVK
	230	240	250	260	270	280
	310	320	330	340	350	360
2MEPSP	TETSV	TVTGP	PREPF	GRKHL	KAIDV	NMNKM

Unipro	GDDY	ISCT-	-----	RGEL	NAIDM	DMNHIP
	290	300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMV	AIRTEL	TKLGAS	VEEGP	DYCIIT	PPEKLN

Unipro	TDRLF	FAMATE	LKVGAE	VEEGH	DFIRIT	PPEKLN
	350	360	370	380	390	400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP TIRDPGCTRKTFFPDYFDVLSTFVKN
      :: :: :: ::::::::::. :.
Unipro TILDPKCTAKTFFPDYFEQLARISQPG
                410          420

```

ARO_A_PSEA6 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Pseudoalteromonas atlantica* (strain T6c / BAA-1087). |AA|429
Length = 429
initn: 922 initl: 386 opt: 562 Z-score: 1113.0 bits: 215.1 E(): 1.4e-54
Smith-Waterman score: 941; 46.560% identity (72.936% similar) in 436 aa overlap (1-436:1-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL LLAALSEGTTVVDNLLN SEDVHYMLGAL
      :. . .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MSRSAPLLLNPISAISGTVNVP GSKSLSNRALL LLA AVANGETHLTNLLD SEDIRHMLKAL
                10          20          30          40          50          60

```

```

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      : : : . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TQLGVNYRLSDDKTECWVQGLGRGFSVDNLET---LFLGNAGTAMRPLCAVLATSIGE--
                70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro FELTGEPRMEERPIGALVDSL RQAGAQITYLKN EGYPLKMKGMA-LKGGNISVEGAVSS
                120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMAAPLFEHDSVINIVGELVSKPYIDITLNTMAQFGITVENNN-YQSFTVKGNGQ
                180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : : . : : : : : : : : : : . . : : : : : : : : : : : : : : : : : :
Unipro QYQAAGDFLVEGDASSASYFLAAGAIKGGTVRVTVGVGKKSIIQGDIFADVLEKMGAKITW
                240          250          260          270          280          290

```

```

                310          320          330          340          350          360
2MEPSP TETSVITVTGPPREPFGRKHLKAIDVNMNKM P DVAMTLAVVALFADGPTAIRDVASWRVKE
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro GDDYIEVTGAP-----LTAVDMDMNHIPDAAMTIATTALFAEGTTSIRNIYNWRVKE
                300          310          320          330          340

```

```

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
      : : : : . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TDRLAAMACELRKVGAEVIEGQDYITITPPNTLIQTDIDTYDDHRVAMCFSLVALSDTPV
                350          360          370          380          390          400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      :::: :: :::::
Unipro TINDPDCTAKTFPDYFTRLKQISA
      410          420

```

AROASALAR | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Salmonella arizonae* (strain ATCC BAA-731 / CDC346-86 / RSK2980). | AA|427
Length = 427
initn: 891 initl: 386 opt: 560 Z-score: 1109.0 bits: 214.3 E(): 2.3e-54
Smith-Waterman score: 956; 47.706% identity (73.394% similar) in 436 aa overlap
(5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Unipro  MESLTLQPIARVDGAINLPGSKSVSNRALLLAALACGKTVLTNLLDSDDVHRMLNAL
      10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      :::: . ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Unipro  SALGIDYTLADRTTRCDIIGNGGAL---RAPGDLELFLGNAGTAMRPLAAALCLGQNET-
      60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :: : ::::: ::::: ::::: ::::: . . ::::: ::::: :::::
Unipro  -VLTGEPRMKERPIGHLVDSLQGGANIDYLEQENYPPLRLRG--GFTGGDIEVDGSVSS
      120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :: . . . . ::::: ::::: ::::: ::::: ::::: . . . . :::::
Unipro  QFLTALLMTAPLAPKDTIIRVKGELVSKPYIDITLNLMTFGVEIMNHH-YQQFVVKGGQ
      180          190          200          210          220

```

```

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      .: :: . ::::: ::::: . . . . ::::: ::::: :::::
Unipro  QYHSPGRYLVEGDASSASYFLAAGAIKGGTVKVTGIGRKSMDIRFADVLEKMGATITW
      230          240          250          260          270          280

```

```

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . : : ::::: ::::: ::::: ::::: ::::: :::::
Unipro  GDDFIAC-----RGELHAIDMDMNHIPDAAMTIATTALFAKGTTLRNIYNWRVKE
      290          300          310          320          330          340

```

	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT	TAIDTYDDHRMAMAFSLAACAEVPV				
	::: :: ::::: :: ::::: ::: . : ::::: ::::: :::::					
Unipro	TDRLFAMATELRKVGAEVEEGDDYIRITPPAKLHHAD	IGTYNDHRMAMCFSLVALSDTPV				
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
	:: :: :: ::::: . .					
Unipro	TILDPKCTAKTFPDYFEQLARISTPA					
	410	420				

		10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQPIKEIS	GTVKLPGSKSL	SNRILL	LAALSEG	TTVVDNLLN	SEDEVHYMLGAL
	
Unipro	MVDSL	TLQPVALVNG	TVNLP	GSKSV	SNRALL	LAALAE	GTTTTLTNLLD
		10	20	30	40	50	
		70	80	90	100	110	120
2MEPSP	RTLGL	SVEADKAAKRA	VVVGCGKFP	VEDAKEE	VQLFLGN	AGIAMRSL	xxxxxxxxxxxx

Unipro	QALGV	SYQLSDDRT	TCKVDG	VGGPLV	--ASKP	LELFLGN	AGTAMRPLAAALCLGSNDV
	60	70	80	90	100	110	

130 140 150 160 170 180
 2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
 :: : ::::: :: :. : :::: : : :: ::::: : : : : :::::
 Unipro -VLTGEPRMKERPIGHLVDALRQGGAQIDYLEQTDYPPRLRLG--GFRGGDVTVDGSVSS
 120 130 140 150 160 170

190 200 210 220 230 240
 2MEPSP QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
 : : : : : : . : : : : : : : : : : : : : : : :
 Unipro QFLTALLMTAPLAEQDTHIHIGELVSKPYIDITLHLMRTFGVEVSH-DNYRVFHINGRQ
 180 190 200 210 220 230

2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
.: : . :
Unipro TYRSPGDYLVEGDASSASYFLAAAAIKGGTVRVGTIGIGKKSVQGDTKFADVLEKMGARITW
240 250 260 270 280 290

Page 126 of 526

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                370      380      390      400      410      420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTDAIDTYDDHRMAMAFSLAACAEVPV
      :. . . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro  TDRLAAMATELRKVGAEVDEGEDYIHVVPPAKLKFADIGTYNDHRMAMCFSLVALSDTPV
                350      360      370      380      390      400

                430      440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  TILDPKCTAKTFPDYFEQLARISQLA
                410      420

```

AROA_SALPA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Salmonella paratyphi* A. |AA|427
 Length = 427
 initn: 877 initl: 385 opt: 559 Z-score: 1107.0 bits: 214.0 E(): 3e-54
 Smith-Waterman score: 951; 48.284% identity (73.455% similar) in 437 aa overlap (5-440:2-421)

```

                10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :. . . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro  MESLTLQPIARVDGAINLPGSKSVSNRALLLAALACGKTVLTNLLDSDDVHRHMLNAL
                10      20      30      40      50

                70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :. . . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro  SALGINYTLSADTRCDITGNGGPLRASGALE---LFLGNAGTAMRPLAAALCLGQNEI-
                60      70      80      90      100      110

                130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  -VLTGEPRMKERPIGHLVDSLROGGANIDYLEQENYPPLRLRG--GFIGGDIEVDGSVSS
                120      130      140      150      160      170

                190      200      210      220      230
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVK-AEHSDSWDRFYIKGG
      :. . . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro  QFLTALLMTAPLAPEDTIIRVKGELVSKPYIDITLNLMTFGVEIANHH--YQQFVVKGG
                180      190      200      210      220

                240      250      260      270      280      290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  QQYHSPGRYLVEGDASSASYFLAAGAIKGGTVKVTGIGRKSMQGDIFADVLEKMGATIT
                230      240      250      260      270      280

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      300      310      320      330      340      350
2MEPSP WTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      : . . . :      : . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro WGDDFIACT-----RGELHAIDMDMNHIPDAAMTIATTALFAKGTTLRNINWRVK
      290      300      310      320      330      340

```

```

      360      370      380      390      400      410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVP
      : : : . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro ETDRLFAMATELRKVGAEEVEEGHDYIRITPPAKLHHADIGTYNDHRMAMCFSLVALSDTP
      350      360      370      380      390      400

```

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVK
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VTILDPKCTAKTFPDYFEQLARMSTPA
      410      420

```

AROA_ECOL6 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Escherichia coli* 06. |AA|427
Length = 427
initn: 992 initl: 384 opt: 559 Z-score: 1107.0 bits: 214.0 E(): 3e-54
Smith-Waterman score: 977; 48.853% identity (74.312% similar) in 436 aa overlap (5-440:2-421)

```

      10      20      30      40      50      60
2MEPSP MAGAEFIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MESLTLQPIARVDGTINLPGSKSVSNRALLAALAHGKTVLTNLLDSDVVRHMLNAL
      10      20      30      40      50

```

```

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . : : . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TALGVSYTLSADRTRCEIIGNGGPLHAESARE--LFLGNAGTAMRPLAALCLGSNDI-
      60      70      80      90      100      110

```

```

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro -VLTGEPRMKERPIGHLVDALRQGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
      120      130      140      150      160      170

```

```

      190      200      210      220      230      240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEIENQH-YQQFVVKGGQ
      180      190      200      210      220

```

```

      250      260      270      280      290      300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : : : . : : : : : : : : : : : : : : : : : : : : : :
Unipro SYQSPGTYLVEGDASSASYFLAAAIRGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
      230      240      250      260      270      280

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTLRNINWRVKE
      290          300          310          320          330          340

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTDAIDTYDDHRMAMAFSLAACAEVPV
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro SDRLFAMATELRKVGAEVEEGHDFIRITPPEKLFKFAEIATYNDHRMAMCFSLVALSDTPV
      350          360          370          380          390          400

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TILDPKCTAKTFPDYFEQLARISQPG
      410          420

```

AROA_PHOPR | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Photobacterium profundum* (*Photobacterium* sp. |AA|428
Length = 428
initn: 974 initl: 412 opt: 559 Z-score: 1107.0 bits: 214.0 E(): 3e-54
Smith-Waterman score: 985; 48.624% identity (75.917% similar) in 436 aa overlap (5-440:2-422)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MESLTLQPVSQINGEVNLPGSKSVSNRALLLAALAQGTTRLTNLLDSDDIRHMLNAL
      10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . : : . . . : : : : : : : : : : : : : : : : : : : : : :
Unipro KQLGVNYQLSTDKTVCEVEGLGSAF--NASQALELYLGNAGTAMRPLAALCLSQG--E
      60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro FVLTGEPRMKERPIGHLVDALRTAGADVTYLENENYPPLKITGTG-LYGGEVEIDGSISS
      120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTAFLMAAPLATADTVIRIKGDLVSKPYIDITLHIMAQFGVTVENRD-YQEFFIPAGQ
      180          190          200          210          220          230

```

	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	:
Unipro	TYQGAGDFLVEGDASSASYFLAAAAIKGGEVKVTGIGKKSIIQGDVQFAHALEMMGAEIEW					
	240	250	260	270	280	290
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE					
	.. :	:
Unipro	GDNYVIAR-----RGELKAVDMDFNHIPPDAAMTIAVAALFAEGTTSIRNVYNWRVKE					
		300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVPV					

Unipro	TDRLAAMATELRKVGAKVEEGNDYITIVPPTQLQHATIDTYDDHRMAMCFSLVALSDTPV					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
				
Unipro	TINDPKCTSKTFPDYFDKLAELSQPA					
	410	420				
ARO_A_VIBCH 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.1 enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Sho Vibrio cholerae. AA 426 Length = 426 initn: 1007 initl: 415 opt: 558 Z-score: 1105.0 bits: 213.6 E(): Smith-Waterman score: 983; 49.654% identity (73.903% similar) in 4 (5-437:2-419)						
	10	20	30	40	50	60
2MEPSP	MAGAEIEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL					

Unipro	MESLTLQPIELISGEVNLPGSKSVSNRALLLAALASGTRLTNLLDSDDIRHMLNAL					
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx					

Unipro	TKLGVNYRLSADKTTCEVEGLGQAF---HTTQPLELFLGNAGTAMRPLAALCLGQGD--					
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDGVPVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS					

Unipro	YVLTGEPVPRMKERPIGHLVDALRQAGAQIEYLEQENFPPLRIQGTG-LQAGTVTIDGSISS					
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ					

Unipro	QFLTAFLMSAPLAQKGKVTIKIVGELVSKPYIDITLHIMEQFGVQVINHD-YQEFVIPAGQ					
	180	190	200	210	220	230

ARO3_VIBC3 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Vibrio cholerae* serotype O1 (strain ATCC 39541 / Ogawa 395 / O395). | AA|426
Length = 426
initn: 1001 init1: 415 opt: 558 Z-score: 1105.0 bits: 213.6 E(): 3.9e-54
Smith-Waterman score: 977; 49.423% identity (73.903% similar) in 433 aa overlap (5-437:2-419)

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro QFLTAFLMSAPLAQGKVTIKIVGELVSKPYIDITLHIMEQFGVQVINHD-YQEFVIPAGQ
                180          190          200          210          220          230

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro SYVSPGQFLVEGDASSASYFLAAAIKGGEVKVTGIGKNSIQGDIQFADALEKMGAQIEW
                240          250          260          270          280          290

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro GDDYVIAR-----RGELNAVDLDFNHIPDAAMTIATTALFAKGTTAIRNVYNWRVKE
                300          310          320          330          340

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro TDRLAAMATELRKVGATVEEGEDFIVITPPTKLIHAAIDTYDDHRMAMCFSLVALSDTPV
                350          360          370          380          390          400

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro TINDPKCTSKTFPDYFDKFAQLSR
                410          420

```

AROA_SALTI | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Salmonella typhi*. | AA | 427
Length = 427
initn: 876 initl: 384 opt: 558 Z-score: 1105.0 bits: 213.6 E(): 3.9e-54
Smith-Waterman score: 949; 48.284% identity (73.913% similar) in 437 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro MESLTLQPIARVDGAINLP GSKSVSNRALLLAALACGKTVLTNLLDSDDVHMLNAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro SALGINYTLSADRTRCDITGNNG--PLR-ASGTLELFLGNAGTAMRPLAAALCLGQNEI-
                60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP YVLGDGVP RMRERPIGDLVVG LKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro -VLTGEPRMKERPIGHLVDSL RQG GANIDYLEQENYPPLRLRG--GFIGGDIEVDGSVSS
                120          130          140          150          160          170

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                190          200          210          220          230
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVK-AEHSDSWDRFYIKGG
      . . . . . : . . . . . : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMTAPLAPEDTIIRVKGELVSKPYIDITLNLTKTFGVEIANHH--YQQFVVKGG
                180          190          200          210          220

                240          250          260          270          280          290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      . : : . : : : : : : : : : . . : : : : : : : : : : : : : :
Unipro QQYHSPGRYLVEGDASSASYFLAAGAIKGGTVKVTGIGRKSMDIRFADVLEKMGATIT
                230          240          250          260          270          280

                300          310          320          330          340          350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      : . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro WGDDFIAC-----RDELHAIDMDMNHIPDAAMTIATTALFAKGTTLRNIYNWRVK
                290          300          310          320          330          340

                360          370          380          390          400          410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVP
      : : . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro ETDRLFAMATELRKVGAEVEEGHDYIRITPPAKLQHADIGTYNDHRMAMCFSLVALSDTP
                350          360          370          380          390          400

                420          430          440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VTILDPKCTAKTFPDYFEQLARMSTPA
                410          420

```

AROA_SALGL | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Salmonella gallinarum*. | AA|427
Length = 427
initn: 875 initl: 384 opt: 558 Z-score: 1105.0 bits: 213.6 E(): 3.9e-54
Smith-Waterman score: 949; 48.284% identity (73.227% similar) in 437 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEIIVLQPIKEISGTVKLPKSKSLNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MESLTLQPIARVDGAINLPKSKSVSNRALLLAALACGKTVLTNLLSDDDRHLNAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      . : . : : : : : : : : : : : : : : : : : : : : : : :
Unipro SALGINYTLSADRTCDITGNGGPLRAPGALE---LFLGNAGTAMRPLAAALCLGQNEI-
                60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :: : :::::::::: :: :.: : ::::: . . . ::::: :. :: ::::::::::
Unipro -VLTGEPRMKERPIGHLVDSLRLQGGANIDYLEQENYPPLRLRG--GFIGGDIEVDGSVSS
                120          130          140          150          160          170

                190          200          210          220          230
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVK-AEHSDSWDRFYIKGG
      :. . . . . :. :. . :.: : ::::::::::: :::: :.: : . . . : .: :
Unipro QFLTALLMTAPLAPKDTIIRVKGELVSKPYIDITLNLTKTFGVEIANHH--YQQFVVKGG
                180          190          200          210          220

                240          250          260          270          280          290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      :. : : . : :::::::::::::::::::: . . : ::::: :. : :
Unipro QQYHSPGRYLVEGDASSASYFLAAGAIKGGTVKVTGIGRKSMDIRFADVLEKMGATIT
                230          240          250          260          270          280

                300          310          320          330          340          350
2MEPSP WTETSVITVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVK
      : . . . : : : : ::::: : ::::: : : : : : : : : : : : : : : :
Unipro WGDDFIACT-----RGELHAIDMDMNHIPDAAMTIATTALFAKGTTLRNIYNWRVK
                290          300          310          320          330          340

                360          370          380          390          400          410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVP
      :.: : . : : : :.: : : : : : : : : : : : : : : : : : : : : :
Unipro ETDRLFAMATELRKVGAEVEEGHDYIRITPPAKLQHADIGTYNDHRMAMCFSLVALSDTP
                350          360          370          380          390          400

                420          430          440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      :.: : : : :.: : .
Unipro VTILDPKCTAKTFPDYFEQLARMSTPA
                410          420

```

AROA_SALPB | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Salmonella paratyphi* B (strain ATCC BAA-1250 / SPB7). |AA|427
Length = 427
initn: 875 initl: 384 opt: 558 Z-score: 1105.0 bits: 213.6 E(): 3.9e-54
Smith-Waterman score: 951; 48.284% identity (73.227% similar) in 437 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEIIVLQPIKEISGTVKLPGSKSLNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :.: :.: : :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :
Unipro MESLTLQPIARVDGAINLPGSKSVSNRALLLAALACGKTVLTLNLLSDSDVRHMLNAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :.: : . :. :. : :. : : : : :.: :.: :.: :.: :.: :.: :
Unipro SALGINYTLSADRTRCDITGNGGALRAPGALE--LFLGNAGTAMRPLAALCLGQNEI-
                60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro -VLTGEPRMKERPIGHLVDSLRLQGGANIDYLEQENYPPLRLRG--GFTGGDIEVDGSVSS
                120          130          140          150          160          170

                190          200          210          220          230
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVK-AEHSDSWDRFYIKGG
      :. . . . . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :
Unipro QFLTALLMTAPLAPKDTIIRVKGELVSKPYIDITLNLTKTFGVEIANHH--YQQFVVKGG
                180          190          200          210          220

                240          250          260          270          280          290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      :. : : . : : : : : : : : : : . . : : : : : : : : : : : : : : : :
Unipro QQYHSPGRYLVEGDASSASYFLAAGAIKGGTVKVTGIGRKSMDQDIRFADVLEKMGATIT
                230          240          250          260          270          280

                300          310          320          330          340          350
2MEPSP WTETSVTVTGPPREPFGRLKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      : . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro WGDDFIAC-----RGELHAIDMDMNHIPDAAMTIATTALFAKGTTLRNINWRVK
                290          300          310          320          330          340

                360          370          380          390          400          410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVP
      :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :
Unipro ETDRLFAMATELRKVGAEVEEGHDYIRITPPAKLQHADIGTYNDHRMAMCFSLVALSDTP
                350          360          370          380          390          400

                420          430          440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VTILDPKCTAKTFPDYFEQLARMSTPA
                410          420

```

AROA_SALCH | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Salmonella choleraesuis*. |AA|427
Length = 427
initn: 875 initl: 384 opt: 558 Z-score: 1105.0 bits: 213.6 E(): 3.9e-54
Smith-Waterman score: 951; 48.284% identity (73.227% similar) in 437 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSKSLNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :
Unipro MESLTLQPIARVDGAINLPKSKSVSNRALLLAALACGKTVLTNLLDSDDVHMLNAL
                10          20          30          40          50

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      .:. . . . :. .: :. . : : : : : : : : : : : : : : : : : : : :
Unipro  SALGINYTLSADTRCDITGNNGALRAPGALE---LFLGNAGTAMRPLAAALCLGQNEI-
              60          70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLGDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  -VLTGEPRMKERPIGHLVDSLRLQGGANIDYLEQENYPPLRLRG--GFTGGDIEVDGSVSS
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVK-AEHSDSWDRFYIKGG
      :. . . . . :. . . . . :. . : : : : : : : : : : : : : : : : : :
Unipro  QFLTALLMTAPLAPKDTIIRVKGELVSKPYIDITLNLTKTFGVEIANHH--YQQFVVKGG
              180          190          200          210          220

              240          250          260          270          280          290
2MEPSP  QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      :. : : . : : : : : : : : : : . . . : : : : : : : : : : : : : :
Unipro  QQYHSPGRYLVEGDASSASYFLAAGAIKGGTVKVTGIGRKSMQGDIFADVLEKMGATIT
              230          240          250          260          270          280

              300          310          320          330          340          350
2MEPSP  WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      : . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  WGDDFIACT-----RGELHAIDMDMNHIPDAAMTIATTALFAKGTTLRNIYNWRVK
              290          300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVP
      :.:. . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  ETDRLFAMATELRKVGAEVEEGHDYIRITPPAKLQHADIGTYNDHRMAMCFSLVALSDTP
              350          360          370          380          390          400

              420          430          440
2MEPSP  VTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : :
Unipro  VTILDPKCTAKTFPDYFEQLARMSTPA
              410          420

```

AROASALTY | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Salmonella typhimurium*. |AA|427
Length = 427
initn: 871 init1: 384 opt: 558 Z-score: 1105.0 bits: 213.6 E(): 3.9e-54
Smith-Waterman score: 947; 48.055% identity (73.227% similar) in 437 aa overlap (5-440:2-421)

```

              10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :.:.:. . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  MESLTLQPIARVDGAINLPKSKSVSNRALLAALACGKTALTNLLDSDDVRLHMLNAL
              10          20          30          40          50

```

[illegible]

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AROA_VIBVY | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Vibrio vulnificus* (strain YJ016). |AA|428

Length = 428

initn: 1016 initl: 412 opt: 558 Z-score: 1105.0 bits: 213.6 E(): 3.9e-54

Smith-Waterman score: 989; 49.199% identity (75.286% similar) in 437 aa overlap (5-441:2-423)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEIS	GTVKLP	GSKSL	SNRILL

Unipro	MESLT	LQPIK	KVNGE	VNLPG	SKSVS	NRALL
	10	20	30	40	50	

	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KAARAV	VVGCG	GKFPV	EDAKEE

Unipro	TKLGV	HYELS	ADKTV	CVVEGL	GRPFT	ATEAQE
	60	70	80	90	100	110

	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	PIGDL	VVGLK	QLGADV	CDCLGT

Unipro	FVLTG	EPRMK	ERPIGH	LVDAL	REAGAQ	IEYLEN
	120	130	140	150	160	170

	190	200	210	220	230	240
2MEPSP	QYxxxx	xxxxxxx	GDVEIE	IIDKL	ISIPY	EMTLRL

Unipro	QFLTA	FLMAA	PLAQGE	VKIHIV	GELVSK	PYIDIT
	180	190	200	210	220	230

	250	260	270	280	290	300
2MEPSP	KYKSP	KNAYVE	GDASSA	SYFLAG	AAIxxxx	xxxxxxxx

Unipro	SYVSP	GQFLVE	GDASSA	SYFLAAA	AIKGGE	IKVTGI
	240	250	260	270	280	290

	310	320	330	340	350	360
2MEPSP	TETSV	TVTGT	GPPREP	FGRKHL	KAIDVN	MNMKMP

Unipro	GEDYV	I-----	SRVGR	--LKG	IDMDYN	HIPDAAM
	300	310	320	330	340	

	370	380	390	400	410	420
2MEPSP	TERMV	AIRTEL	TKLGAS	VEEGPD	YCIITP	PPEKLN

Unipro	TDRLS	AMATEL	RKVGA	EVEEGE	DYLVNPP	QQLTHA
	350	360	370	380	390	400

	430	440
2MEPSP	TIRDP	GCTRKT

Unipro	TINDP	KCTSKT

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

410 420

AROA_ENTS8 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Enterobacter sakazakii* (strain ATCC BAA-894). | AA|428
Length = 428
initn: 910 initl: 378 opt: 557 Z-score: 1103.0 bits: 213.2 E(): 5e-54
Smith-Waterman score: 979; 48.394% identity (75.459% similar) in 436 aa overlap (5-440:3-422)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEI	SGTVK	LPGSK	SLSNR

Unipro	MQESL	TLPQI	ARVDG	TINLP	GSKSV	SNRALL
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KAARK	RAVVV	GCGKF	FPVED

Unipro	NALGV	HYSLS	DDRTR	CEIQG	GGPF--	NTLVE
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIG	DLVVG	LKQLG	ADVDC
	::	:	:	...
Unipro	-VLTG	EPRMK	ERPIG	HLVDL	RQGG	ADVTL
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxx	xxxxxxxx	GDVEI	EIIDK	LISIP	YVEMT
	:
Unipro	QFLT	TALLM	AAPLA	PGNTA	IDIKG	ELVSK
	180	190	200	210	220	230
	250	260	270	280	290	300
2MEPSP	KYKSP	KNAYV	EGDASS	ASYFL	AGAAI	xxxxxxxx

Unipro	QYQSP	GHYL	VEGD	ASSYF	LAAAA	IKGGT
	240	250	260	270	280	290
	310	320	330	340	350	360
2MEPSP	TETSV	TVTGT	PPREP	FGRKH	LKAID	VNMNK
	.	..	:	:
Unipro	GDDF	ISCT-	-----	RGEL	NAID	MDMNH
				300	310	320
	370	380	390	400	410	420
2MEPSP	TERMV	AIRTEL	TKLG	ASVEE	GPDYCI	IITPPE

Unipro	TDRLA	AMATE	LKVG	ATVEE	GHDFIT	VTPPA
	350	360	370	380	390	400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      :: :: :: ::::: :: :.
Unipro TILDPKCTAKTFPNYFAQLARISHSA
                410          420

```

AROA_KLEPN | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Klebsiella pneumoniae*. |AA|427
Length = 427
initn: 887 initl: 376 opt: 554 Z-score: 1097.1 bits: 212.1 E(): 1.1e-53
Smith-Waterman score: 976; 49.541% identity (75.000% similar) in 436 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :...::: :...::: :...::: :...::: :...::: :...:::
Unipro  MESLTLQPIARVDGTVNLPGSKSVSNRALLLAALARGTTVLTNLLDSDDVRLMLNAL
                10          20          30          40          50

```

```

                70          80          90         100         110         120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :... :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  SALGVHYVLSSDRTRCEVTGTGG--PLQ-AGSALELFLGNAGTAMRPLAAALCLGSNDI-
                60          70          80          90         100         110

```

```

                130         140         150         160         170         180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :: : :...::: :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  -VLTGEPRMKERPIGHLVDALRQGAQIDYLEQENYPPLRLRG--GFTGGDVEVDGSVSS
                120         130         140         150         160         170

```

```

                190         200         210         220         230         240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  QFLTALLMASPLAPQDTVIAIKGELVSRPYIDITLHLMKTFGVEVENQ-AYQRFIVRGNQ
                180         190         200         210         220

```

```

                250         260         270         280         290         300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      :... :. :...::: :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  QYQSPGDYLVEGDASSASYFLAAGAIKGGTVKVTGIGRNSVQGDIFADVLEKMGATVTW
                230         240         250         260         270         280

```

```

                310         320         330         340         350         360
2MEPSP TETSVTVTGPPEPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  GEDYIACT-----RGELNAIDMDMNHIPDAAMTIATAALFARGTTTLRNIYNWRVKE
                290         300         310         320         330         340

```

```

                370         380         390         400         410         420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVPV
      :... :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro  TDRLFAMATELRKVGAEEGEDYIRITPPLTLQFAEIGTYNDRMAMCFSLVALSDTPV
                350         360         370         380         390         400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        :: :: :: ::::: :.
Unipro  TILDPKCTAKTFPDYFGQLARISTLA
                410          420

```

ARO_A_CITK8 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Citrobacter koseri* (strain ATCC BAA-895 / CDC 4225-83 / SGSC4696). | AA|427
Length = 427
initn: 873 init1: 363 opt: 554 Z-score: 1097.1 bits: 212.1 E(): 1.1e-53
Smith-Waterman score: 978; 48.853% identity (74.312% similar) in 436 aa overlap
(5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        :...::: :...::: :...::: :...::: :...::: :...::: :...::: :...:::
Unipro   MESLTLQPIARVDGTINLPGSKSVSNRALLLAALASGTTVLTNLLDSDDVHRMLNAL
                10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
        .:::. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
Unipro   SALGVSYTSLADRTCEITGQGGVLHAEGALE---LFLGNAGTAMRPLAAALCLGANDI-
        60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        :: : ::::: : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro   -VLTGEPRMKERPIGHLVDALRQGGAKIDYLEQENYPPLRLRG--GFSGGHVEVDGSVSS
                120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :. . . . :. : : : : : : : : : : : : : : : : : : : : : : : :
Unipro   QFLTALLMTAPLAPQDTTIAIKGDLVSKPYIDITLNLMTFGVEVENQN-YQRFVVKGEQ
                180          190          200          210          220

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        .:::. : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro   QYRSPGQYLVEGDASSASYFLAAGAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATITW
        230          240          250          260          270          280

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKE
        . . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro   GDDFIAC-----RGELNAVDMMDMNHIPDAAMTIATAALFAKGTTLRNIYNWRVKE
        290          300          310          320          330          340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTDAIDTYDDHRMAMAFSLAACAEVPV					
	:.:. .: .: :.:.:.:.:. .: .: :.:. .: :.:.:.:.:.:.:. .:..:					
Unipro	TDRLFAMATELRKVGAEEVEEGHDFIRITPPAQLQFAEIGTYNDHRMAMCFSLVALSDTPV					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
	:.: .: .: :.:.:.:. .:					
Unipro	TILDPKCTAKTFPDYFEQLARISTPA					
	410	420				

AROASULNB | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Sulfurovum* sp. (strain NBC37-1). | AA|427
Length = 427
initn: 1017 initl: 351 opt: 551 Z-score: 1091.1 bits: 211.0 E(): 2.3e-53
Smith-Waterman score: 996; 49.195% identity (76.322% similar) in 435 aa overlap (7-441:4-423)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLPKSGKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL					
	:.:.:. .: .: :.:.:.:.:. .: :.:.:.:. .: :.:.:.:.:.:.:					
Unipro	MNSITLKPIRYIEGEVNLPGSKSLSNRALLIAALAEGTTRITNLLSDDTRHMLNAL					
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx					
	. :.:. .: .: :.:.:. .: : .: :.:.:. .: : .: :.:. .: .: .: .:					
Unipro	KLLGVEYTLSEDRTCTVVGNGGPF--HTKEPLELFLGNAGTAMRPLCAALTGSGT--					
	60	70	80	90	100	110

	130	140	150	160	170	180
2MEPSP	YVLGVPVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS					
	:.: :.:.:.:. .: .: .: .: .: .: .: .: :.:.:. .: :.:.:.:.:.:					
Unipro	YVLTGEPRMKERPIGHLVDALREAGAKITYLENEGYPPLKIEA-DGLKGGEVRIDGSISS					
	120	130	140	150	160	170

	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ					
	:. .: .: .: :.:.:. .: : .: :.:.:. .: :.:.:. .: :.:.:. .: :.:.:.:					
Unipro	QFLTALLMAAPMARGDMQIDIVGELVSKPYIDITLHIMKQFGVEVRN-ENYERFFIKGGQ					
	180	190	200	210	220	230

	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	:. .: .: :.:.:. .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:					
Unipro	IYQALETFMVEGDASSASYFLAAAIAIKGGTVKVTGIGKTSVQGDVAFADVLEKMGAKVEW					
	240	250	260	270	280	290

	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE					
	. :.:. .: .: :.:.:. .: :.:. .: :.:. .: :.:. .: :.:. .: :.:. .: :.:.:					
Unipro	GDDYVSVS-----RGELNAVNMDFNHIPDAAMTIATTALFAKGTTLRNIYNWRVKE					
	300	310	320	330	340	

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
      :. . . :. : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  TDRLHAMAMELRKVGAKVEEGEDYLTITPPVQLKHA AIDTYDDHRMAMCFSLALDPVSV
                350          360          370          380          390          400

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
      :. . . :. : : : : : : : : :
Unipro  TINEPECTAKTFPTYFEVLESISY
                410          420

```

AROA_PHOLL | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Photorhabdus luminescens* subsp. *laumondii*. |AA|428
Length = 428
initn: 984 initl: 383 opt: 551 Z-score: 1091.1 bits: 211.0 E(): 2.3e-53
Smith-Waterman score: 952; 48.966% identity (72.414% similar) in 435 aa overlap (5-439:2-420)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRI LLLAALSEGTTVVDNLLNSEDVHYMLGAL
      ..... : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  MQSLMLQPISYINGTINLP GSKSVSNRALL LAAFAKGATCLTNLLDSDDIRHMLNAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      :. . . :. : : : : : : : : : : : : : : : : : : : : : : :
Unipro  AALGISYRLSDDRTCCEVDGIGGLITHQGP---IELFLGNAGTAMRPLTAALCLGKNDV-
                60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  -VLTGEPRMKERPIGHLVDALRQGGAEIDYLEQENYPPLHVK--GGFVGGKVMVDGRVSS
                120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . :. : : : : : : : : : : : : : : : : : : : : : : :
Unipro  QFLTALLMAAPLAENDSEIHIQ GELVSKPYIDITLALMKSFGITINH-DQYQIFHIKGRQ
                180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : : :. : : : : : : : : : : : : : : : : : : : : : : :
Unipro  QYVSPGHYLVEGDASSASYFLAAA AIKGGTVRVTGIGKNSLQGDTKFANVLEKMGAKIRW
                230          240          250          260          270          280

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . :                      : : ..... : : ..... : .....
Unipro GDDFVECE-----RGTLTGIDMDMNEIPDAAMTIATTALFAAGETVIRNIYNWRVKE
      290                300          310          320          330          340

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVPV
      :. . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TDRLHAMATELRKVGAEVEEGVDYIRITPPRLLPAEIGTYNDHRMAMCFSLVALSDTPV
      350          360          370          380          390          400

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TILDPGCTAKTFPDYFNQLERLSQRKS
      410          420

```

AROA_MARMS | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Marinomonas* sp. (strain MWYL1). |AA|425
Length = 425
initn: 843 initl: 394 opt: 550 Z-score: 1089.1 bits: 210.6 E(): 3e-53
Smith-Waterman score: 938; 47.126% identity (74.713% similar) in 435 aa overlap (7-441:4-422)

```

                10          20          30          40          50          60
2MEPSP MAGAEFIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :. : . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MNSITLGPLSKANGEIQIPGSKSLSNRILLATLAKGTTKITNLLDSDDIRRMLES
      10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :. . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TKLGVSYSLEDNGTTCVLEGLG--PIQ--ADFGDLFLGNAGTAMRPLTAALCLGKGE--
      60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro FLLHGEPRMHERPIGDLVDALQALGVDITYEGEKNYPPLRIKA-NGLSGGEVSIKGNISS
      120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALMSAPLAKSDLTIKVDGELVSKPYIDITLHAMKQFGVEVENQN-YQAFVVKGQQ
      180          190          200          210          220

```

```

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      :. : . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TYQSPGEIMVEGDASSASYFLAAAAIAGGKIKVHGVGTDSVQGDVKFADVLAQMGAKITY
      230          240          250          260          270          280

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      :: ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro -----GPSWIEAERNELNGIDMDMNHIPDAAMTIATTALFAKGPTTIRNIYNWRVKE
                290          300          310          320          330          340

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPV
      :. : . : : : : : : : : : : . : : : . : : : : : : : : : : : : : : : : : : :
Unipro TDRLYAMATELKKLGADVIEGKDFITVTPVANLKHAAIDTYNDHRIAMCFSLVAFSDTPV
                350          360          370          380          390          400

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      :. : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TINDPGCTSKTFPTYFELFNTIAN
                410          420

```

AROA_SHEHH | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shewanella halifaxensis* (strain HAW-EB4). |AA|426
Length = 426
initn: 865 initl: 392 opt: 546 Z-score: 1081.1 bits: 209.2 E(): 8.3e-53
Smith-Waterman score: 969; 47.685% identity (74.537% similar) in 432 aa overlap (6-437:3-419)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      .. : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  MNQLRLEPIKKVSGTINIPGSKSISNRALLLATLASGTTTLTNLLDSDDIRYMLASL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . : : : . . . . : . : . : : : : : : : : : : : : : : : : : : .
Unipro  KQLGVSYSRLSNNTVCELDGLAGPL--NAGEPQTLFLGNAGTAMRPLCAALTLGQG--Q
                60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      .. : : : : : : : : : : : : : : : . . : : : : : : : : : : : : : :
Unipro  FTLTGEPRMEERPIGDLVDALRQLGAEVSYLKNEGFPLNITSTG-LNGGNVEIAGDLSS
                120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  QFLTALLMVAPLAKGDVNIKIKGELVSKPYIDITLALMAQFGVEVQNND-YASFVIKAGQ
                180          190          200          210          220          230

```

			10		20		30		40
2MEPSP		MAGAE	EIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVDNL						
				:	:	:	:	:	:
Unipro	SSVSDDDIRTILSQNILLYGIPLNAFQKH	TTTTITLP	GSKSISNRALILASLSNGICYLKNF						
	380	390	400	410	420	430			
	50	60	70	80	90	100			
2MEPSP	LNS	EDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFVPVEDAKEEVQLFLGNAGIAMR-							
	:	:	:	:	:	:	:	:	:
Unipro	LHSDDTYYMLS	ALEKLN-AAEFKW	EQDGDVLVVKGKSGYLENP--QMELYLGNSGTTF						
	440	450	460	470	480	490			
	110	120	130	140	150	160			
2MEPSP	--SLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVG	LKQLGADVDCFLGTD	CPPVRVNGI						
	:	:	:	:	:	:	:	:	:
Unipro	LTSICTLVQPNSRENHLILTGSNRMQRP	IGPLVDALKNNGCCIEYLELENCLPLLKP							
	500	510	520	530	540	550			

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      170      180      190      200      210      220
2MEPSP G-GLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIID-KLISIPYVEMTLRLMERFG
      :: :::::.....::: .. . . .: .::: : :: ::::: . . . ::
Unipro EIGLYGGININLSATVSSQYVSSILMCSPIYAKTQVTLSLIGGKPISQPYIDMTISMSSFG
      560      570      580      590      600      610

      230      240      250      260      270      280
2MEPSP VKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQ
      .:. .: : .. : : : .. .: :::::.....: :: :::. . . . .: :::
Unipro IKVTRSHSKENTYYIPKGCYTCPSEYIIIEGDATSATYPLAIAAITGGSCTISNVGSASLQ
      620      630      640      650      660      670

      290      300      310      320      330      340
2MEPSP GDVKFAE-VLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAI-DVNMNKMMPDVAMTLAVV
      :: :::. .: ::::: . . . . .: :::. .: .: .: .: .: .: .: .: .: .:
Unipro GDSKFSEYILKPMGCEVVQSPTTTYIKGPPK--GK--LKSLGSINMESMTDTFTLTA AVL
      680      690      700      710      720      730

      350      360      370      380      390
2MEPSP ALFA---DGP--TAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN-
      : : . : : : .: ::::: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro ASVAYEESKPYVTKITGISNQRKECNRINAMVCELKKFGIEAGELPDGIYVKALNTSNL
      740      750      760      770      780      790

      400      410      420      430      440
2MEPSP ---VTAIDTYDDHRMAMAFSLAAC-AEVPVTIRDPGCTRKTTFPDYFDVLSTFVKN
      : .: .: :::::.....: :: .. :::: : .: .: .: .: .: .: .:
Unipro PYSVEGINCYNDHRIAMSFSVLACISSKPTTILDKACVNKTWPYWWDILNSTFKVQMKGI
      800      810      820      830      840      850

Unipro EFDLNPTINSSILHHPSECTIFLIGMRGAGKTTLGQLAANFLGREFIDLDSIIIEDIKTT
      860      870      880      890      900      910

```

AROA_SHEON | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shewanella oneidensis*. |AA|426
Length = 426
initn: 947 initl: 384 opt: 544 Z-score: 1077.1 bits: 208.4 E(): 1.4e-52
Smith-Waterman score: 934; 46.528% identity (71.991% similar) in 432 aa overlap (5-436:2-418)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      ... :.:. . . : .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro MKQLRLPEVVQVRGEINIPGSKSISNRALLLATLAQGTTTLTNLLDSDDIRHMLASL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . :.:. . . . .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro KQLGVEYRLSHNNTVCELAGLGG--VMSSKQAQTLFLGNAGTAMRPLCAALTGQGE--
      60      70      80      90      100      110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  FTLTGEPRMEERPIGDLVDALRQLGANIVYLKNDGFPPLTINATG-LNGGDVEIVGDLSS
                120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  QFLTALLMVAPLAKGSVNIHVKGELVSKPYIDITLALMAQFGVKVINHH-YARFEIPAGQ
                180          190          200          210          220          230

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : : . . : : : : : : : : : : . : : : : : : : : : : : : : : : : :
Unipro  HYVSPGKVLVEGDASSASYFLAAGAIKGGEVKVTGVGRLSIQGDVKFADVLEKMGADIEW
                240          250          260          270          280          290

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  GDDYIIARGAP-----LTAVDLDMNHIPDAAMTIATAALFAKGTTVIRNIYNWRIKE
                300          310          320          330          340

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
      : : : : . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  TDRLAAMATELRKVGAEVEEGNDYIKITPPAVLNTAQIDTYNDHRMAMCF SMLAFADCGI
                350          360          370          380          390          400

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  TINDPDCTSKTFPDYFVQFASLKV
                410          420

```

AROA_SHESR | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shewanella* sp. (strain MR-7). |AA|426
Length = 426
initn: 848 initl: 382 opt: 543 Z-score: 1075.1 bits: 208.0 E(): 1.8e-52
Smith-Waterman score: 940; 46.528% identity (72.685% similar) in 432 aa overlap (5-436:2-418)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  MKQLRLPEPVVQVRGEINIPGSKSISNRALLLATLAKGTTTLTNLLDSDDIRHMLASL
                10          20          30          40          50

```

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFVPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx					
	. : : : . . . : . : : : : : : : : : . . .					
Unipro	KQLGVGYRLSQNNNTVCELTGLGGAISADTAQ---TLFLGNAGTAMRPLCAALTlGRGE--					
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLdGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLsgSISS					
	. . : : : : : : : : : . : : : . . . : : : : : : : : : : . : : : : . : :					
Unipro	FTLTGEPMERPEERPIGDLDALKQLGANIMYLKNEGFPLTINATG-LNGGDVEIAGDLSS					
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAehSDSWDRFYIKGGQ					
	: : : : : . : : : : : : : : : : . : . : : : : : :					
Unipro	QFLTALLMVAPLAKGSVNihVKELVSKPYIDITLALMAQFGVQVINHD-YARFEILAGQ					
	180	190	200	210	220	230
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxXXXXXXXXXXXXXSLQGdVKFAEVLemmGAkVTW					
	. : : : . . : : : : : : : : : . . : : : : : : : : : : : : : : . : :					
Unipro	QYVSPGKVLVEGDASSASYFLAAGAIKGGEVKVTGVGRLSIQGDVKFADVLEKMgADIEW					
	240	250	260	270	280	290
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGRKHLKAIDVNMNKMpdVAMTLAVVALFADGPTAIRDVASwRVKE					
 :					
Unipro	GDDYIIARGAP-----LTAVDldMNHIPDAAMTIATAALfAKGTTTTIRNIYNwRIKE					
	300	310	320	330	340	350
	370	380	390	400	410	420
2MEPSP	TERmVAIRTELTKLGASVEEGPDYCIITPPEKLNvTAIDTYDDHRMAMAfSLAACAEVpV					
	: : .					
Unipro	TDRLAAMATELRKVGAlVEEGHdyIQITPPAVLNtAEIDTYNDHRMAMCFsMMAFADcGI					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTfPDYFDVlSTfVKn					
	: :					
Unipro	TINDPDCTSkTFPDYfAQfASLKa					
	410	420				

Bayer CropScience

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx					
	. : : : : : :
Unipro	KQLGVNYRLSQNNTVCELDGIGGVISSASAE--LFLGNAGTAMRPLCAALTLGQGE--					
	60	70	80	90	100	110

	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS					
	. : . : . : . : . : .	. : . : . : . : .	. : . : . : . : .	. : . : . : . : .	. : . : . : . : .	. : . : . : . : .
Unipro	FTLTGEPRMEERPIGDLVDALRQLGANIVYLKNDGFPPLTINATG-LNGGDVEIAGDLSS					
	120	130	140	150	160	170

	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ					
	. : : . : . : .	. : . : . : .	. : . : . : .	. : . : . : .	. : . : . : .
Unipro	QFLTALLMVAPLAKGSVNIHVKGELVSKPYIDITIALMAQFGVNVINHD-YARFEIVAGQ					
	180	190	200	210	220	230

	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	. : . : . .	. : . : . : . : .	. : . : . : . : .	. : . : . : . : .	. : . : . : . : .	. : . : . : . : .
Unipro	RYISPGKVLVEGDASSASYFLAAGAIKGEVKVTGVGRLSIQGDVKFADVLEKMGADIEW					
	240	250	260	270	280	290

	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE					
	. : . . . : .	. : . : . : .	. : . : . : .	. : . : . : .	. : . : . : .	. : . : . : .
Unipro	GDDYIIARGSP-----LTAVDLDMNHIPDAAMTIATAALFAKGTTVIRNIYNWRIKE					
	300	310	320	330	340	

	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIDTYDDHRMAMAFSLAACAEVPV					
	. : . : . .	. : . : . : .	. : . : . : .	. : . : . : .	. : . : . : .	. : . : . : .
Unipro	TDRLAAMATELRKVGAEVEEGNDYIKITPPAVINTAEIDTYNDHRMAMCFMSMLAFADCGI					
	350	360	370	380	390	400

	430	440
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN	
	. : . : . .	. : . : . .
Unipro	TINDPDCTSKTFPDYFKQFASLQG	
	410	420

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AROAS_HESA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shewanella* sp. (strain ANA-3). |AA|426

Length = 426

initn: 862 init1: 381 opt: 543 Z-score: 1075.1 bits: 208.0 E(): 1.8e-52

Smith-Waterman score: 937; 46.528% identity (72.454% similar) in 432 aa overlap (5-436:2-418)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLP	GSKSLSNRILL	LAALSEGTTVVDNLLN	SEDEVHYMLGAL		

Unipro	MKQLRLPEVVQVRGEINIP	GSKSISNRALL	LATLAKGTTTLTNLL	SDDIRHMLASL		
	10	20	30	40	50	

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVV	VGCGKFPVEDAKEEV	QLFLGNAGIAMRSL	xxxxxxxxxxxx		

Unipro	KQLGVEYRLSQNNTV	CELTGLGGVISS	TAQ---TLFLGNAG	TAMRPLCAALT	TLGRGE--	
	60	70	80	90	100	110

	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIG	DLVVGLKQLGAD	VDCFLGTDCPP	VRVNGIGGLPG	GKVKLSGS	SISS

Unipro	FTLTGEPRMEERPI	GDLVDALKQLG	ANIVYLKNDG	FPPLTINATG	-LNGGDVEI	AGDLSS
	120	130	140	150	160	170

	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxx	GDVEIEIIDKL	SIPYVEMTLRL	MERFGVKA	EHSDSWDR	IFYIKGGQ

Unipro	QFLTALLMVAPL	AGSVNIHV	KGELVSKPY	IDITLALMA	QFGVQVIN	HD-YARFEIPAGQ
	180	190	200	210	220	230

	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGD	ASSASYFLAGAA	xxxxxxxxxxxx	SLQGDVKFA	EVLEMMGAK	VTW

Unipro	QYVSPGKVLVEGD	ASSASYFLAAG	AIKGGEVKVT	GVGRLSIQ	GDVKFADV	LEKMGADIEW
	240	250	260	270	280	290

	310	320	330	340	350	360
2MEPSP	TETSVTVTGPP	PREPFGRKHL	KAIDVNMN	KMPDVAMT	LAVVALFAD	GPTAIRDVASWRVKE

Unipro	GDDYIIARGAP	-----	LTAVDLDM	NHIPDAAM	TIATAAL	FAKGTTTIRNIYNWRIKE
	300		310	320	330	340

	370	380	390	400	410	420
2MEPSP	TERMVAIRTEL	TKLGASVEEG	PDYCIITP	PEKLNVT	AIDTYDDHR	MAMAFSLAACAEVPV

Unipro	TDRLAAMATEL	RKVGALVEE	GHDIQITP	PVVLNTAE	IDTYNDHR	MAMCFSMMAFADCGI
	350	360	370	380	390	400

	430	440
2MEPSP	TIRDPGCTRKT	FPDYFDVLSTFVKN

Unipro	TINDPDCTSKT	FPDYFAQFASLKA
	410	420

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AROA_SHESW|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shewanella* sp. (strain W3-18-1). |AA|426
Length = 426
initn: 944 initl: 382 opt: 542 Z-score: 1073.1 bits: 207.7 E(): 2.3e-52
Smith-Waterman score: 930; 46.065% identity (72.685% similar) in 432 aa overlap (5-436:2-418)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEIS	GTVKLP	GSKSL	SNRILL
	...	:..	..	:
Unipro	MKQLR	LEPVV	QVRGE	INIPGS	KSSISN	RALLAT
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KAARAV	VVGCG	KFPVED	AKEEVQ
	. :..	..	:
Unipro	KQLGV	NYRLS	QNNTV	CELDG	LGGVI	SSASAE
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIG	DLVVGL	KQLGAD	VDCFLG
	..	:
Unipro	FTLTG	EPRME	ERPIG	DLVDAL	RKLGAN	VVYLKN
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxx	xxxxxxx	GDVEIE	IIDKL	ISIPY	VEMTLR
	:.
Unipro	QFLTAL	LMVAP	LAKGS	VNIHV	KGELV	SKPYID
	180	190	200	210	220	230
	250	260	270	280	290	300
2MEPSP	KYKSP	KNAYV	EGDASS	ASYFLA	GAIIx	xxxxxxxx
	.: ::
Unipro	RYISP	CKVLV	EGDASS	ASYFLA	AGAIK	GGEVKV
	240	250	260	270	280	290
	310	320	330	340	350	360
2MEPSP	TETSV	TVTGT	PPREP	FGRKHL	KAIDVN	MNMKMP
	.	..	:	:
Unipro	GDDYII	ARGSP	-----	LTAVD	LDNMH	IPDAAM
	300	310	320	330	340	
	370	380	390	400	410	420
2MEPSP	TERMV	AIRTEL	TKLGAS	VEEGPD	YCIITP	PPEKLN
	:.....
Unipro	TDRLA	AMATEL	RKVGA	EVEEGN	DIKITP	PAVINT
	350	360	370	380	390	400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFFPDYFDVLSTFVKN
        :::: :: :::::
Unipro  TINDPDCTSKTFFPDYFKQFASLQG
                410          420

```

AROAS_HESH | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Shewanella sediminis* (strain HAW-EB3). | AA|426
Length = 426
initn: 863 initl: 379 opt: 542 Z-score: 1073.1 bits: 207.7 E(): 2.3e-52
Smith-Waterman score: 945; 46.882% identity (73.441% similar) in 433 aa overlap (5-437:2-419)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        ... ::::: ::::: ::::: ::::: ::::: :::::
Unipro   MKQLRLEPINKVQGTINIPGSKSISNRALLLATLAKGTTTLTNLLSDDIRYMLASL
                10          20          30          40          50

```

```

                70          80          90          100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
        . ::.. . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Unipro   KQLGVNYRLSNDNTVCELEGIGAPLNSEQAQ---TLFLGNAGTAMRPLCAALTGHGE--
                60          70          80          90          100         110

```

```

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        .. : ::::: ::::: . . . . . . . . . . . . . . . . . . . . . .
Unipro   FTLTGEPMEERPIGDLVDALRQLGANVTYLNKNDGFPPLTINATG-LDAGEVEIAGDLSS
                120         130         140         150         160         170

```

```

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :. . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Unipro   QFLTALLMVAPLATGDVNIKIKGELVSKPYIDITIALMAQFGVQVNVH-SYERFEIKAGQ
                180         190         200         210         220         230

```

```

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        : :: . . ::::: ::::: . . . . . . . . . . . . . . . . . . .
Unipro   GYVSPGKVLVEGDASSASYFLAAGAIKGEVKVTGVGRMSIQGDVKFADVLEKMGADIEW
                240         250         260         270         280         290

```

```

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
        . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Unipro   GDDYIISRGST-----LKAVDLDMNHIPDAAMTIATAALFATGTTTHIRNIYNWRIKE
                300         310         320         330         340

```

```

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADITYDDHRMAMAFSLAACAEVPV
        :. . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Unipro   TDRLAAMATELRKVGAIVDEGHDIYISITPPTKPHTADIDTYNDRMAMCFMSLAFADCGI
                350         360         370         380         390         400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        :::: :: :::::
Unipro  TINDPDCTSKTFPDYFNQFAALAQ
                410          420

```

AROA_SHESM | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shewanella* sp. (strain MR-4). |AA|426
Length = 426
initn: 860 init1: 380 opt: 542 Z-score: 1073.1 bits: 207.7 E(): 2.3e-52
Smith-Waterman score: 935; 46.296% identity (72.454% similar) in 432 aa overlap (5-436:2-418)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        ... ::. . . : ::::: ::::: ::::: ::::: ::::: ::::: :::::
Unipro   MKQLRLEPVVQVRGEINIPGSKSISNRALLATLAKGTTTLTNLLDSDDIRHMLASL
                10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
        . ::. . . . . . . : : . . . : : : : : : : : : : : . . .
Unipro   KQLGVEYRLSQNNTVCELIGLGGAISADTAQ--TLFLGNAGTAMRPLCAALTGRGE--
        60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        .. : : ::::: ::::: . . . . . : : . . : : : : : : : :
Unipro   FTLTGEPRMEERPIGDLVDALKQLGANIVYLKNEGFPPLAINATG-LNGGDVEIAGDLSS
        120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :. . . . . . . : : : : : : : : : : : : : : : : : : : : :
Unipro   QFLTALLMVAPLAKDSVNIHVKGELVSKPYIDITLALMAQFGVQVINHD-YARFEIPAGQ
        180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        .: : : . . : : : : : : : : . . . : : : : : : : : : : :
Unipro   QYVSPGKVLVEGDASSASYFLAAGAIKGGEVKVTGVGRLSIQGDVKFADVLEKMGADIEW
        240          250          260          270          280          290

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
        . . . : : : : : : : : : : : : : : : : : : : : : :
Unipro   GDDYIIARGAP-----LTAVDLDMNHIPDAAMTIATAALFAKGTTTIRNIYNWRIKE
        300          310          320          330          340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
        :.:.:. .: :.:. .: :.:. .: :.:. .: :.:. .: :.:. .: :.:. .: :.:.
Unipro  TDRLAAMATELRKVGALVEEGHDYIQITPPAVLNTAEIDTYNDHRMAMCFSMMAFANCGI
          350          360          370          380          390          400

          430          440
2MEPSP  TIRDPGCTRKTFFPDYFDVLSTFVKN
        :.:. .: :.:. .: :.:. .: :.:. .: :.:. .: :.:. .: :.:.
Unipro  TINDPDCTSKTFFPDYFAQFASLKV
          410          420

```

AROA_EDWIC | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Edwardsiella ictaluri*. |AA|428
Length = 428
initn: 954 initl: 377 opt: 542 Z-score: 1073.1 bits: 207.7 E(): 2.3e-52
Smith-Waterman score: 943; 47.209% identity (74.419% similar) in 430 aa overlap (7-436:5-418)

```

          10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL LLAALSEGTTVVDNLLN SEDVHYMLGAL
        ..:.:. .: :.:. .: :.:. .: :.:. .: :.:. .: :.:. .: :.:. .: :.:.
Unipro  MSSALTLPVRRFSGEINLP GSKSVSNRALL LAAQARGVTRLHNL LDSDDVRYMLDAL
          10          20          30          40          50

```

```

          70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
        ..:. .: :. .: :. .: :. .: :. .: :. .: :. .: :. .: :. .: :.
Unipro  KALGVRYQLSDCRTRCEVQGLGGT LSAHGA---LTLFLGNAGTAMRPLAAALSGLGRDV-
        60          70          80          90          100          110

```

```

          130          140          150          160          170          180
2MEPSP  YVLGVPVRMRERPIGDLVVG LKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        .: :. .: :. .: :. .: :. .: :. .: :. .: :. .: :. .: :. .: :.
Unipro  -ILTGEPRMKERPIAHLVTALRQGGAQVDYLESDGYPPVRLHG--GFNGGEISVDGSVSS
          120          130          140          150          160          170

```

```

          190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :. .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro  QFLTALLMAAPMAAEETRITILGELVSKPYIAITLMMRAFGVEVENH-AYRHFVVRGGQ
          180          190          200          210          220          230

```

```

          250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        :.:. .: :. .: :. .: :. .: :. .: :. .: :. .: :. .: :. .: :.
Unipro  VYQAPSDYLVEGDASSASYFLAGAAIAGGTVRVTGIGRHSMQGDIHFADVLEKMG AQVEW
          240          250          260          270          280          290

```

```

          310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKM PDVAMTLAVVALFADGPTAIRDVASWRVKE
        .: :. .: :. .: :. .: :. .: :. .: :. .: :. .: :. .: :. .: :.
Unipro  GNDYIACT-----RDSLHGIDMDMNAIPDAAMTIATTALFAKGPTTLRNIYNWRVKE
          300          310          320          330          340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTDAIDTYDDHRMAMAFSLAACAEVPV
      :. . . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro  TDRLAAMASELRKVGAVVEEGTDFLRIEPPAQLQAAQIATYNDHRMAMCFSLVALSGTPV
                350          360          370          380          390          400

                430          440
2MEPSP TIRDPGCTRKTFFPDYFDVLSTFVKN
      :. . . . . :. . . . .
Unipro  TICDPGCTAKTFFPDYFRQFSALCHPR
                410          420

```

AROA_PSYIN | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Psychromonas ingrahamii* (strain 37). | AA|428
Length = 428
initn: 929 initl: 388 opt: 541 Z-score: 1071.1 bits: 207.3 E(): 3e-52
Smith-Waterman score: 988; 48.284% identity (74.600% similar) in 437 aa overlap (5-441:2-425)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :. . . . . :. . . . . :. . . . . :. . . . . :. . . . . :. . . . .
Unipro  MEQLLLQPIAKVDGEINLPGSKSLSNRALLLAALAQTTLTNLLDSDDIRHMLNAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . :. . . . :. . . . :. . . . :. . . . :. . . . :. . . . :. . . .
Unipro  KKLGVNYQLSKDKKQCVVEGLGRAFNTAESGL-LLEFLGNAGTAMRPLCAALCLGQG--E
                60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :. . :. . :. . :. . :. . :. . :. . :. . :. . :. . :. . :. . :. .
Unipro  YILTGEARMFERPIGSLVDALQQAGAQVTYLENENYPPLKIKGTG-LKGGKIKIDGSVSS
                120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . :. . . . :. . . . :. . . . :. . . . :. . . . :. . . .
Unipro  QFLTAFLEMAAPMAMEDTEIEIVGELVSKPYIKITLQIMHDFGIDVDNHN-FERFIKKGKQ
                180          190          200          210          220          230

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      :. . :. . :. . :. . :. . :. . :. . :. . :. . :. . :. . :. . :. .
Unipro  TYSAPGHYLVEGDASSASYFLAAGAIAGGCIKVTGIGKKSQVQDQIFADALEAMGAKIEW
                240          250          260          270          280          290

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          310          320          330          340          350          360
2MEPSP TETSVTVTGPPEPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . . :          . . . . . : . . . . . : . . . . . : . . . . . :
Unipro GDDYIIARG-----SQLTAVDLDMNHIPDAAMTIATAALFAKGTTVIRNIYNWRIKE
          300          310          320          330          340

          370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTALDIDYDDHRMAMAFSLAACAEVPV
      : . . . . : : : : . . . . . : : : : : : . . . . . : : : : : :
Unipro TDRLAAMATELRKVGAEVEEGNDYIKITPPAVINTAEIDTYNDHRMAMCFMSLAFADCGI
          350          360          370          380          390          400

          430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      : . . . : : : : : :
Unipro TINDPDCTSKTFPDYFKQFASLQG
          410          420

```

AROA_SHELP | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shewanella loihica* (strain ATCC BAA-1088 / PV-4). |AA|426
Length = 426
initn: 852 initl: 373 opt: 540 Z-score: 1069.1 bits: 206.9 E(): 3.8e-52
Smith-Waterman score: 921; 46.528% identity (72.222% similar) in 432 aa overlap
(5-436:2-418)

```

          10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . . :
Unipro MKQLRLNPISKVHGTVNIPGSKSISNRALLLATLAEGKTRLTNLLDSDDIRHMLTAL
          10          20          30          40          50

          70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      . : . . . . . : . . : . : . : . : . : . : . : . : . : . : . : . : . :
Unipro KQLGVNYQLSDNNRVCEVEGLSGVINSDTAQ---TLFLGNAGTAMRPLCAALTGSGE--
          60          70          80          90          100          110

          130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . . : : . . . . . : . . : . . : . . : . . : . . : . . : . . : . . :
Unipro FTTLTGEPRMEERPIGDLVDALNALGADIRYLKQPGFPPLTINATG-LNGGDVEIAGDLSS
          120          130          140          150          160          170

          190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . . : . . : . . : . . : . . : . . : . . : . . : . . :
Unipro QFLTALLMVTPLAKAQVNIKIKGELVSKPYIDITIALMAQFGVTVINHD-YQRFEIPAGQ
          180          190          200          210          220          230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	:: :: .. ::::::::::::::::::::				:::::::::::::: ::::: :	
Unipro	KYVSPGTVLVEGDASSASYFLAAGAIQGGEVKVTGVGLKSIQGDVKFAEVLEAMGAQIEW					
	240	250	260	270	280	290
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE					
 :	:	::::::: ::::::::::::::::::::	:	::::::: :::::::	
Unipro	GDDFIIARSAP-----LHGVLDLMNHIPDAAMTIATAALFATGTTTLRNIYNWRIKE					
	300	310	320	330	340	
	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV					
	::::::: :::: :::::::::: ::	:	:	:::::: ::::::::::::::::::::	:	:: .
Unipro	TDRLAAMATELRKVGAEVEEGHDYIRVTAPAQLNTADIDTYNDHRMAMCFSLMAFADCGI					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
	:::: :: :: :::::::					
Unipro	TINDPDCTSKTFPDYFAQFAALAQ					
	410	420				

ARO_A_COLP3 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; | *Colwellia psychrerythraea* (strain 34H / ATCC BAA-681) | AA|426
Length = 426
initn: 911 initl: 370 opt: 539 Z-score: 1067.2 bits: 206.6 E(): 5e-52
Smith-Waterman score: 957; 47.248% identity (74.312% similar) in 436 aa overlap (5-440:2-422)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL					
	::::::: :::: . :::::::::: :::::::::::::: :: :::: ::::::::::::::					
Unipro	MEQLTLNPIGKINGEIFLP GSKSLSNRALLIAALANGVTKITNLLVSDDINHMLNAL					
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx					
 ::::: :::: :::::::::: :::: :					
Unipro	KSLGIEYTLSDCGTECTVIGNGGFF---NAKKPLELYLGNAGTAMRPLCAALAASEG--E					
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS					
	:: : :::::::::: :: : : :::: . . : ::::::: : : : :::::::					
Unipro	FILTGEPRMKERPIGHLVDALAQLDADIEYLENKDYPVKIKG-KALTGNTVTIDGSISS					
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ					
	:: ::::: ::::: :::::::::: ::::::::::::::: . : :::::					
Unipro	QFLTAIMIAPLLETNTTIEIDGELVSKPYIDITL DIMRRFNVSVQNND-YKSFIVNGKQ					
	180	190	200	210	220	230

	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
 : : : : : : : : : : . : : : : : : : : :					
Unipro	SYQALDKYMVEGDASSASYFLAAGAIGGGEVTVHGIGKLSVQGDKHFFADVLEKMGAIEIHW					
	240	250	260	270	280	290
	310	320	330	340	350	360
2MEPSP	TETSVTVTGTGPPREPFGFRKHLKAIDVN MNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE					
	. :					
Unipro	KDESITVIGKP-----LTAVDMDMNHIPDAAMTIAT'TALFATGTTTIRNIYNWRVKE					
	300	310	320	330	340	
	370	380	390	400	410	420
2MEPSP	TERMVAIARTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFLAACAEVPV					
	: : . :					
Unipro	TDR LNAMATELRKVGAEEVVEGKDYSISITPPKSLKHAEIDTYNDHRVAMCFSLVALSDTPV					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
	: :					
Unipro	TINDPKCTAKTFPDYFDKLAQVSC					
	410	420				

		10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQPIKEIS	GTVKLPGSKSL	SNRILL	LAALSEG	TTVVDNLLN	SEDEVHYMLGAL
	
Unipro	MQSL	TLPISR	INGTIN	LPGSKSV	SNRALL	AAFAKGT	TRLTNLLDSDDIRYMLNAL
		10	20	30	40	50	
		70	80	90	100	110	120
2MEPSP	RTLGLS	VEADKAAKRA	VVVGCGKFP	VEDAKEE	VQLFLGN	NAGIAMRSL	xxxxxxxxxxxx
	.:	.:	.:	.:	.:	.:	.:
Unipro	TALDIP	YRLSADRT	VCEVEGR	SGNI---	TGKSGLE	LFLGNAGT	AMRPLAAALCLGDNEI-
	60	70	80	90	100	110	
		130	140	150	160	170	180
2MEPSP	YVLDG	VPRMRER	PIGDLV	VLKQLG	ADVDCFL	GTDCPP	VRVNGIGGLPGGKVKLSGSISS
	::	::	::	::	::	::	::
Unipro	-VLTGE	PRMKER	PIGHLV	DALRQ	GGAKIDY	IEQENY	PPLHIK--GGFSGGKVTVDGSVSS
	120	130	140	150	160	170	

	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQ					
	: :					
Unipro	QFLTALLMAAPLAVNNTEIHIIQGDLVSKPYIDITLALMKSFGVTVENHQ-YQVFYIRGRQ					
	180	190	200	210	220	
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	: : . :					
Unipro	QYLSPGQYLVEGDASSASYFLAAAAIKGGIVRVGTIGIGKNSLQGDTKFANVLEQMGIATIRW					
	230	240	250	260	270	280
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGFRKHHLKAIDVNMMNKMMPDVAMT LAVVALFADGP TAI RDVASWRVKE					
	. :					
Unipro	GDDFVECE-----RGTLTGIDMDMNAIPDAAMTIAT'TALFAQGETVIRNIYNWRVKE					
	290	300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYD DHRMAMA FSLAACAEVPV					
	: . :					
Unipro	TDRLNAMATELRKVGAEEVEEGLDYIRVIPPEKIQHAEIETYNDHRVAMCFSLVALSNTFPV					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFFPDYFDVLSTFVKN					
	: :					
Unipro	TILDPGCTAKTFPDYFNQLKKLSEYTT					
	410	420				

		10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQPIKEIS	GTVKLPGSKSL	SNRILL	LAALSEG	TTVVDNLLN	SEDEVHYMLGAL
		:	:	:	:	:	:
Unipro	MES	LTLPINKIQGE	VNLPGSKSV	SNRALL	LSALAKG	TRLTNLLD	SDDIRHMLNAL
		10	20	30	40	50	
		70	80	90	100	110	120
2MEPSP	RTLGLS	VEADKAAKRA	VVVGCGKFP	VEDAKEE	VQFLGNAGI	AMRSL	xxxxxxxxxxxx
	:	:	:	:	:	:	:
Unipro	TKLGV	KYTLSADKTE	CVVEGLGR	PFSVS---	EPVELF	LGNA	GAMRPLAAALCVGQGE--
	60	70	80	90	100	110	
		130	140	150	160	170	180
2MEPSP	YVLDG	VPRMRERPI	GD	LVVLKQLG	ADVDCFL	GTDCPPVR	VNIGIGLPGGKVKLSGSISS
	:	:	:	:	:	:	:
Unipro	YVLTG	EPMRKERPI	GH	LV	TALQKAG	ADIEY	LENTNYPPLKIVGTG-LKAGTVSIDGSISS
	120	130	140	150	160	170	

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTAFLMSAPLAEGEVRIKIEGDLVSKPYIDITLHIMKQFGVEVINND-YQEFVIPAGQ
                180          190          200          210          220          230

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      . : . : : : : : : : : : : . : : : : : : : : : : : : :
Unipro HYVAPGDFLVEGDASSASYFLAAAAIKGGEVKVTGIGKNSIQGDIQFADALEKMGAEIEW
                240          250          260          270          280          290

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . : . : . : : : : : : : : : : : : : : : : : : : : :
Unipro GDDYVI-----SRVGK--LKIGIDMDYNHIPDAAMTIATTALFAEGTTAIRNVYNWRVKE
                300          310          320          330          340

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
      : . . : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TDRLSAMATELRKVGAEEVEEGEDYIIIVKVPVPHLKHA AIDTYDDHRMAMCFSL LALSDTPV
                350          360          370          380          390          400

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TINDPKCTSKTFPDYFDKLKALSC
                410          420

```

AROA_SHEB5 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase; Short=EPSP synthase; Short=EPSPS; |*Shewanella baltica* (strain OS155 / ATCC BAA-1091). |AA|426
Length = 426
initn: 836 initl: 382 opt: 538 Z-score: 1065.2 bits: 206.2 E(): 6.4e-52
Smith-Waterman score: 938; 46.528% identity (72.917% similar) in 432 aa overlap (5-436:2-418)

```

                10          20          30          40          50          60
2MEPSP MAGAEIIVLQPIKEISGTVKLP GSKSLNRILL LLAALSEGTTVVDNLLNSE DVHYMLGAL
      ... : . . : : : : : : : : : : : : : : : : : : : : :
Unipro MKQLRLPEVVQVRGEINIPGSKSISNRALLLATLAQGTTTLTNLLDSDDIRHMLASL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . : . . . . : : : . : : : : : : : : : : : : : . . .
Unipro KQLGVNYRLSQNNTVCELDGLGGVISSESAQE---LFLGNAGTAMRPLCAALT LGQGE--
                60          70          80          90          100          110

```

	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS					
	. . : : :::::.....: . :.....: . . :.: .: : : : :					
Unipro	FTLTGEPRMEERPIGDLVDALRQLGANVVYLKNDGFPLTINATG-LSGGDVEIAGDLSS					
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ					
	:. :.:... .:..:: :. :.: . . : . :.: : .:					
Unipro	QFLTALLMVAPLAKGSVNIIHVKGELVSKPYIDITLALMAQFGVTVINHD-YARFEIVAGQ					
	180	190	200	210	220	230
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxXXXXXXXXXXXXXSLQGDKFAEVLEMMGAKVTW					
	.: :. :. :.:.....:.....: . . :.:.....: :.:.....: :.: :.					
Unipro	RJVSPGKVLVEGDASSASYFLAAGAIAKGGEVKVTGVGRLSIQGDVKFADVLEKMADIEW					
	240	250	260	270	280	290
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGFRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE					
	. . . : .: :.:...: .:.....: :.:.....: :.:.....: :.:...					
Unipro	GDDYIIARG-----SQLTAVDLDMNHIPDAAMTIATAALFAKGTTVIRNIYNWRIKE					
	300	310	320	330	340	
	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVPV					
	:.....: :.: :.:.....: :. :.: .: .: :.:.....: :.: . . .					
Unipro	TDRLAAMATELRKVGAEEVEEGNDYIKITPPAVINTAEIDTYNDHRMAMCFSM LAFADCGI					
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFFPDYFDVLSTFVKN					
	::.: :. :.:.....:					
Unipro	TINDPDCTSKTFFPDYFKQFASLQG					
	410	420				

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

APPENDIX C – DETAILED FASTA SEQUENCE SIMILARITY ALIGNMENTS OF THE 2MEPSPS PROTEIN
WITH PROTEINS FROM THE UNIPROT_TREMBL DATABASE

O24566_MAIZE | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags: Fragment; |Zea mays (Maize). |AA|444
Length = 444
initn: 2045 initl: 2045 opt: 2045 Z-score: 4044.6 bits: 757.6 E(): 1.2e-216
Smith-Waterman score: 2045; 91.216% identity (95.270% similar) in 444 aa
overlap (2-445:1-444)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLP	GSKSLSNRILL	AALSEGTTVVDNLL	NS	EDVHYMLGAL	
					
Unipro	AGAEIIVLQPIKEISGTVKLP	GSKSLSNRILL	AALSEGTTVVDNLL	NS	EDVHYMLGAL	
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVG	CGGKFPVEDAKEEVQLFLGN	AGIAMRSL	xxxxxxxxxxxx		
 : : :					
Unipro	RTLGLSVEADKAAKRAVVVG	CGGKFPVEDAKEEVQLFLGN	AGTAMRPLTA	AVTAAGGNAT		
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIGDLV	VGLKQLGADVDCFLGTDC	PPVRVNGIGGLPGGKVK	LSGSIS		
					
Unipro	YVLDGVPRMRERPIGDLV	VGLKQLGADVDCFLGTDC	PPVRVNGIGGLPGGKVK	LSGSIS		
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxGDVEIEI	IDKLISIPYEMTLRLMER	FGVKA	EHSDSWDRFYIKGGQ		
	: : . . . :					
Unipro	QYLSALLMAAPLALGDVEIEI	IDKLISIPYEMTLRLMER	FGVKA	EHSDSWDRFYIKGGQ		
	180	190	200	210	220	230
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASY	FLAGAAIxxxxxxxxxxxx	SLQGDVKFAEVLEMMGAK	VTW		
 :					
Unipro	KYKSPKNAYVEGDASSASY	FLAGAAITGGTVTVEGCGT	TSLQGDVKFAEVLEMMGAK	VTW		
	240	250	260	270	280	290
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGRKHL	KAIDVNMNKM	PDVAMTLAVVALFADGPTA	IRDVASWRVKE		
					
Unipro	TETSVTVTGPPREPFGRKHL	KAIDVNMNKM	PDVAMTLAVVALFADGPTA	IRDVASWRVKE		
	300	310	320	330	340	350
	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEE	GPDYCIITPPEKLN	VT	AIDTYDDHRMAMAFSLAACAEVPV		
					
Unipro	TERMVAIRTELTKLGASVEE	GPDYCIITPPEKLN	VT	AIDTYDDHRMAMAFSLAACAEVPV		
	360	370	380	390	400	410

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          .....
Unipro  TIRDPGCTRKTFPDYFDVLSTFVKN
          420          430          440

```

Q95AK0_ELEIN | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags:
 Fragment; | *Eleusine indica* (Goosegrass). |AA|445
 Length = 445
 initn: 1606 initl: 1606 opt: 2001 Z-score: 3957.4 bits: 741.4 E(): 8.6e-212
 Smith-Waterman score: 2001; 89.213% identity (94.831% similar) in 445 aa
 overlap (2-445:1-445)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          .....
Unipro  AGAEEVVLQPIKEISGVVCLPGSKSLSNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
          10          20          30          40          50

                70          80          90          100         110
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
          .....
Unipro  KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNA
          60          70          80          90          100         110

                120         130         140         150         160         170
2MEPSP  xYVLDGVPVRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
          .....
Unipro  TYVLDGVPVRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSIS
          120         130         140         150         160         170

                180         190         200         210         220         230
2MEPSP  SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
          ... ..
Unipro  SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
          180         190         200         210         220         230

                240         250         260         270         280         290
2MEPSP  QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
          .....
Unipro  QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
          240         250         260         270         280         290

                300         310         320         330         340         350
2MEPSP  WTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
          .....
Unipro  WTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
          300         310         320         330         340         350

                360         370         380         390         400         410
2MEPSP  ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
          .....
Unipro  ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACADVP
          360         370         380         390         400         410

```

```

420          430          440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      ::::::::::::::::::::::
Unipro VTIRDPGCTRKTFPDYFDVLSTFVKN
      420          430          440

Q95AK1_ELEIN|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags:
Fragment; |Eleusine indica (Goosegrass). |AA|445
Length = 445
initn: 1601 init1: 1601 opt: 1996 Z-score: 3947.5 bits: 739.6 E(): 3e-211
Smith-Waterman score: 1996; 89.213% identity (94.831% similar) in 445 aa
overlap (2-445:1-445)

      10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Unipro AGAEEVVLQPIKEISGVVKLPGSKSLSNRILLLSALAEAGTTVVDNLLNSEDVHYMLGAL
      10          20          30          40          50

      70          80          90          100          110
2MEPSP RTLGLSVEADKAAKRAVVVGCCKGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .:::::::::::::::::::::::::::::::::::::::::::::::::::::.....
Unipro KTLGLSVEADKAAKRAVVVGCCKGKFPVEKDAKEEVQLFLGNAGTAMRSLTAAVTAAGGNA
      60          70          80          90          100          110

      120          130          140          150          160          170
2MEPSP xYVLGDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Unipro TYVLGDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVKIGIGGLPGGKVKLSGSIS
      120          130          140          150          160          170

      180          190          200          210          220          230
2MEPSP SQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYEMTLRLMERFGVKAHSDSWDRFYIKGG
      ::: .. .. . ::::::::::::::::::::::::::::::::::::::::::::::
Unipro SQYLSALLMAAPLALGDVEIEIIDKLISIPYEMTLRLMERFGVKAHSDSWDRFYIKGG
      180          190          200          210          220          230

      240          250          260          270          280          290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      ::::::::::::::::::::::::::::::. . . ::::::::::::::::::::::
Unipro QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
      240          250          260          270          280          290

      300          310          320          330          340          350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVK
      ::::::::::: ::::::::::::::::::::::::::::::::::::::::::::::
Unipro WTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVK
      300          310          320          330          340          350

      360          370          380          390          400          410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
      ::::::::::::::::::::::: :::::::::::::::::::::::::::::::
Unipro ETERMVAIRTELTKLGASVEEGLDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACADVP
      360          370          380          390          400          410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      ::::::::::::::::::::::
Unipro VTIRDPGCTRKTFPDYFDVLSTFVKN
      420      430      440

```

AOAR17_ELEIN | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags:
Fragment; | *Eleusine indica* (Goosegrass). | AA|445
Length = 445
initn: 1597 initl: 1597 opt: 1992 Z-score: 3939.5 bits: 738.1 E(): 8.4e-211
Smith-Waterman score: 1992; 88.764% identity (94.607% similar) in 445 aa
overlap (2-445:1-445)

```

              10              20              30              40              50              60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      ::::.::::::::::..:::::::::::::::::..:..::::::::::::::::::::::::::
Unipro  AGAEEVVLQPIKEISGVVGLPGSKSLSNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
              10              20              30              40              50

```

```

              70              80              90              100             110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      .:::::::::::::::::..:::::::::::::::::..:..:..:..:..:..:..:..:..:..
Unipro  KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRALTAAVTAAGGNA
      60              70              80              90              100             110

```

```

      120      130      140      150      160      170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .:::::::::::::::::..:::::::::::::::::..:..:..:..:..:..:..:..:..:..
Unipro  TYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSIS
      120      130      140      150      160      170

```

```

      180      190      200      210      220      230
2MEPSP SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      ::: .. .. . :::::::::::::::::::..:..:..:..:..:..:..:..:..:..:..
Unipro  SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      180      190      200      210      220      230

```

```

      240      250      260      270      280      290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      :::::::::::..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..
Unipro  QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLETMGAKVT
      240      250      260      270      280      290

```

```

      300      310      320      330      340      350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVK
      :::::::::::..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..
Unipro  WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVK
      300      310      320      330      340      350

```

```

      360      370      380      390      400      410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVP
      :::::::::::..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..:..
Unipro  ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACADVP
      360      370      380      390      400      410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      ::::::::::::::::::::::
Unipro VTIRDPGCTRKTFPDYFDVLSTFVKN
      420      430      440

```

AOAR18_ELEIN | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags:
Fragment; | *Eleusine indica* (Goosegrass). | AA|445
Length = 445
initn: 1588 initl: 1588 opt: 1983 Z-score: 3921.7 bits: 734.8 E(): 8.3e-210
Smith-Waterman score: 1983; 88.764% identity (94.382% similar) in 445 aa
overlap (2-445:1-445)

```

              10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      ::::.:::::::::.:::::::::::::::::.:.:::::::::::::::::::::::::::
Unipro  AGAEEVVLQPIKEISGVVGLPGSKSLSNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
              10      20      30      40      50

```

```

              70      80      90      100     110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .::::::::::::::::.:::::::::::::.: : : : : : : : : : : : : : : : .
Unipro  KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNA
      60      70      80      90      100     110

```

```

      120      130      140      150      160      170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .::::::::.: : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  TYVLDGVPRMLERPIGDLVVGLKQLGADVDCFLGTDYPPVRVKGIGGLPGGKVKLSGSIS
      120      130      140      150      160      170

```

```

      180      190      200      210      220      230
2MEPSP SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      : : : . . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      180      190      200      210      220      230

```

```

      240      250      260      270      280      290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      .::::::::.: : : . . . : : : : : : : : : : : : : : : : : : :
Unipro  QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
      240      250      260      270      280      290

```

```

      300      310      320      330      340      350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVK
      .::::::::.: : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVK
      300      310      320      330      340      350

```

```

      360      370      380      390      400      410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTaidtyddhrmamaflaacaevp
      .::::::::.: : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTaidtyddhrmamaflaacadvp
      360      370      380      390      400      410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      ::::::::::::::::::::::
Unipro VTIRDPGCTRKTFPDYFDVLSTFVKN
      420      430      440

```

Q93VK6_ORYSA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; | *Oryza sativa* (Rice). | AA|511
Length = 511
initn: 1950 initl: 1584 opt: 1951 Z-score: 3857.4 bits: 723.1 E(): 3.2e-206
Smith-Waterman score: 1951; 86.517% identity (94.382% similar) in 445 aa
overlap (2-445:67-511)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRI
                                : ::::::::::::::::::::::
Unipro GGMVRVRVRARGRREAVVVASASSSSVAAPAAKAEIVLQPIREISGAVQLPGSKSLSNRI
      40      50      60      70      80      90

```

```

                                40      50      60      70      80      90
2MEPSP LLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DA
      :::::::::::::::::::::: :::::::::::::::::::::: ::
Unipro LLLSALSEGTTVVDNLLNSEDVHYMLEALKALGLSVEADKVAKRAVVVGCGGKFPVEKDA
      100     110     120     130     140     150

```

```

                                100     110     120     130     140     150
2MEPSP KEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDC
      :::::::::::::: :: :... .. ::::::::::::::::::::::
Unipro KEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQLGADVDC
      160     170     180     190     200     210

```

```

                                160     170     180     190     200     210
2MEPSP FLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPY
      :::::::::::::: :: .. . ::::::::::::::
Unipro FLGTECPPVRVKIGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDKLISIPY
      220     230     240     250     260     270

```

```

                                220     230     240     250     260     270
2MEPSP VEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxx
      :::::::::::::: :::::::::::::: ::
Unipro VEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGAAITGGT
      280     290     300     310     320     330

```

```

                                280     290     300     310     320     330
2MEPSP xxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMK
      . . :::::::::::::: :::::::::::::: ::::::::::::::
Unipro VTVQCGGTTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPREPYGKKHLKAIDVNMNMK
      340     350     360     370     380     390

```

```

                                340     350     360     370     380     390
2MEPSP PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP
      :::::::::::::: :::::::::::::: ::::::::::::::
Unipro PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP
      400     410     420     430     440     450

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	400	410	420	430	440	
2MEPSP	EKLNVT	AI	DYDDHR	MAMAFSLAACAE	VPVTIRDPGCTRKT	FPDYFDVLSTFVK
	
Unipro	EKLNIT	AI	DYDDHR	MAMAFSLAACAD	VPVTIRDPGCTRKT	FPNYFDVLSTFVR
	460	470	480	490	500	510

A2Y8X9_ORYSI | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; | *Oryza sativa subsp. indica* (Rice). | AA|515
Length = 515
initn: 1950 initl: 1584 opt: 1951 Z-score: 3857.4 bits: 723.1 E(): 3.2e-206
Smith-Waterman score: 1951; 86.517% identity (94.382% similar) in 445 aa
overlap (2-445:71-515)

	10	20	30
2MEPSP	MAGAE	IVLQ	PIKEIS

Unipro	GGMRV	RV	RARGR
	50	60	70

	40	50	60	70	80	90
2MEPSP	LLLAAL	SE	GT	TVVDN	LLNS	EDVHY

Unipro	LLLSAL	SE	GT	TVVDN	LLNS	EDVHY
	110	120	130	140	150	160

	100	110	120	130	140	150
2MEPSP	KEEVQ	LFLG	NAGIAM	RS	Lxxxxxxxxxxxxx	YVLDG

Unipro	KEEVQ	LFLG	NAGTAM	RPLTAA	VTAA	AGGNAT
	170	180	190	200	210	220

	160	170	180	190	200	210
2MEPSP	FLGTD	CPPVR	VNGIG	GLPGG	KVKLSG	SISSQY

Unipro	FLGTE	CPPVR	VKGIG	GLPGG	KVKLSG	SISSQY
	230	240	250	260	270	280

	220	230	240	250	260	270
2MEPSP	VEMTL	RRLM	ERF	GVKAEH	SDSWDR	FYIKGG

Unipro	VEMTL	RRLM	ERF	GVKAEH	SDSWDR	FYIKGG
	290	300	310	320	330	340

	280	290	300	310	320	330
2MEPSP	xxxxxxx	SLQGD	VKFAE	VLEMM	GAKVT	TWTET

Unipro	VTVQG	CGTTS	LQGD	VKFAE	VLEMM	GAKVT
	350	360	370	380	390	400

	340	350	360	370	380	390
2MEPSP	PDVAM	TLAV	VALF	ADGPT	AIRDV	ASWRV

Unipro	PDVAM	TLAV	VALF	ADGPT	AIRDV	ASWRV
	410	420	430	440	450	460

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	400	410	420	430	440
2MEPSP	EKLNVT	AI	DYDDHRMAMAFSLAACAEVPVTIRDPGCTRKT	FPDYFDVLSTFVK	N

Unipro	EKLNIT	AI	DYDDHRMAMAFSLAACADVPVTIRDPGCTRKT	FPNYFDVLSTFVR	N
	470	480	490	500	510

Q5VNW0_ORYSJ | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; | *Oryza sativa subsp. japonica* (Rice). | AA|515
Length = 515
initn: 1950 initl: 1584 opt: 1951 Z-score: 3857.4 bits: 723.1 E(): 3.2e-206
Smith-Waterman score: 1951; 86.517% identity (94.382% similar) in 445 aa
overlap (2-445:71-515)

		10	20	30
2MEPSP		MAGAE	IVLQPIKEIS	GTVKLPGSKSLSNRI
		:
Unipro	GGMRVRVRARGRREAVVASASSSSVAAPAAKAE	IVLQPIREIS	GAVQLPGSKSLSNRI	
	50	60	70	80

	40	50	60	70	80	90
2MEPSP	LLLAAL	SE	GT	TVVDNLLN	SE	DVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DA

Unipro	LLLSAL	SE	GT	TVVDNLLN	SE	DVHYMLEALKALGLSVEADKVAKRAVVVGCGGKFPVEKDA
	110	120	130	140	150	160

	100	110	120	130	140	150
2MEPSP	KEEVQL	FLGNAGIAMRSL	xxxxxxxxxxxxxYVLDG	VPRMRERPI	IGDLVVGLKQLGADVDC	

Unipro	KEEVQL	FLGNAGTAMRPLTA	AVTAAGGNATYVLDG	VPRMRERPI	IGDLVVGLKQLGADVDC	
	170	180	190	200	210	220

	160	170	180	190	200	210
2MEPSP	FLGTDCPPVRVNGIGGLPGGKVKLSGSISSQY	xxxxxxxxxxxxxGDVEIEI	IDKLISIPY			

Unipro	FLGTECPPVRVKGIGGLPGGKVKLSGSISSQYLS	SALLMAAPLALGDVEIEI	IDKLISIPY			
	230	240	250	260	270	280

	220	230	240	250	260	270
2MEPSP	VEMTLRLMERFGVKA	EHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAI	xxxxx			

Unipro	VEMTLRLMERFGVKA	EHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGAAITGGT				
	290	300	310	320	330	340

	280	290	300	310	320	330
2MEPSP	xxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSV	TVTGTGPPREPFG	RKHLKAIDVNMNKM			

Unipro	VTVQCGTTS	LQGDVKFAEVLEMMGAKVTWTDTSVT	TVTGTGPPREPYGKKHLKAIDVNMNKM			
	350	360	370	380	390	400

	340	350	360	370	380	390
2MEPSP	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP					

Unipro	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP					
	410	420	430	440	450	460

	400	410	420	430	440
2MEPSP	EKLNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN				
				
Unipro	EKLNITAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPNYFDVLSTFVKN				
	460	470	480	490	500
					510
Q3Y6S2_LOLMU 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5. Fragment; <i>Lolium multiflorum</i> (Italian ryegrass) AA 437 Length = 437 initn: 1903 initl: 1570 opt: 1904 Z-score: 3765.3 bits: 705.9 E() Smith-Waterman score: 1904; 85.812% identity (93.364% similar overlap (10-445:1-437))					
	10	20	30	40	50
2MEPSP	MAGAEIEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL				
				
Unipro	QPIKEISGAVQLPGSKSLSNRILLLSALSEGTTVVDNLLNSEDVHYMLEAL				
	10	20	30	40	50
	70	80	90	100	110
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx				
				
Unipro	DALGLSVEADKVAKRAVVVGCGGRFPIEKDAKEEVKLFLGNAGTAMRPLTAAVVAAGGNA				
	60	70	80	90	100
					110
	120	130	140	150	160
2MEPSP	xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS				
				
Unipro	TYVLDGVPRMRERPTGDLVVGLKQLGANVDCFLGTDCPPVRINGIGGLPGGKVKLSGSIS				
	120	130	140	150	160
					170
	180	190	200	210	220
2MEPSP	SQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGG				
	::: .. .				
Unipro	SQYLSSLMAAPLALGDVEIEIIDKLISVPYVEMTLRLMERFGVTAEHSDSWDRFYIKGG				
	180	190	200	210	220
					230
	240	250	260	270	280
2MEPSP	QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT				
				
Unipro	QKYKSPGNAYVEGDASSASYFLAGAAITGGTVTVQGCGTTSLQGDVKFAEVLEMMGAKVT				
	240	250	260	270	280
					290
	300	310	320	330	340
2MEPSP	WTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVK				
	:::.....				
Unipro	WTDTSVTVTGPPRQPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVK				
	300	310	320	330	340
					350
	360	370	380	390	400
2MEPSP	ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP				
				
Unipro	ETERMVAICTELTKLGATVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP				
	360	370	380	390	400
					410

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFFPDYFDVLSTFVK
      .....
Unipro VTIRDPGCTRKTFFPNYFDVLSTFVK
      420      430

```

Q1M157_CAMAC | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Camptotheca acuminata* (Happy tree). | AA|519
Length = 519
initn: 1788 initl: 1446 opt: 1795 Z-score: 3548.2 bits: 665.9 E(): 5.3e-189
Smith-Waterman score: 1795; 79.819% identity (91.156% similar) in 441 aa
overlap (6-445:79-519)

```

                        10      20      30
2MEPSP                MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                        .....
Unipro VAVNSPVINVRVPLRVSASVVTTEKTSMTPEIVLQPIKEISGTVKLPGSKSLSNRILLLA
      50      60      70      80      90      100

```

```

      40      50      60      70      80      90
2MEPSP ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
      .....
Unipro ALSEGTTVVDNLLNSDDVHYMLGALRTLGLRVEEDSAIKRAIVEGCSGLFPVGKESTDEV
      110     120     130     140     150     160

```

```

      100     110     120     130     140     150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      .....
Unipro QLFLGNAGTAMRPLTAAVTAAGGNSSYILDGVPRMRERPIGDLVTGLKQLGADVDCFLGT
      170     180     190     200     210     220

```

```

      160     170     180     190     200     210
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      .....
Unipro NCPVVRVIGKGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISIPYVEMT
      230     240     250     260     270     280

```

```

      220     230     240     250     260     270
2MEPSP LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
      .....
Unipro LKLMKRFVGTVEHSDNWDRFLIQGGQKYKSPGNSYVEGDASSASYFLAGAAVTGGTITVE
      290     300     310     320     330     340

```

```

      280     290     300     310     320     330
2MEPSP xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVA
      .....
Unipro GCGSSSLQGDVKFAEVLEKMGAKVTWTENSVTVTGPPRNSSGRKHLRAVDVNMNKMPPDVA
      350     360     370     380     390     400

```

```

      340     350     360     370     380     390
2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      .....
Unipro MTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLGAATVEEGPDYCVITPPEKLN
      410     420     430     440     450     460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP  VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
          .....
Unipro  VTAVDTYDDHRMAMAFSLAACANVPVTIKDPGCTRKTFFPDYFEVLQRFKAH
          470          480          490          500          510

```

Q6E6L5_CONCD | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; | *Conyza canadensis* (Canadian horseweed) | AA|519
Length = 519
initn: 1776 initl: 1437 opt: 1781 Z-score: 3520.5 bits: 660.8 E(): 1.9e-187
Smith-Waterman score: 1781; 78.912% identity (91.383% similar) in 441 aa
overlap (6-445:79-519)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                                .....
Unipro  SVSSNNNRNLGKKCLIVSAVATTEKPSVPEIVLQPIKEISGTVNLPGSKSLSNRILLLA
          50          60          70          80          90          100

```

```

          40          50          60          70          80          90
2MEPSP  ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
          .....
Unipro  ALAEGTTIVDNLLNSDDVHYMLGALRTLGLNVEEDVAIKRAIVEGCGGVFPVGKEAKDDI
          110          120          130          140          150          160

```

```

          100          110          120          130          140          150
2MEPSP  QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPVRMRERPIGDLVVGLKQLGADVDCFLGT
          .....
Unipro  QLFLGNAGTAMRPLTAAVTAAGGNSSYILDGVPVRMRERPIGDLVTGLKQLGADVDCSLGT
          170          180          190          200          210          220

```

```

          160          170          180          190          200          210
2MEPSP  DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
          .....
Unipro  NCPVVRVVGGLPGGKVKLSGSISSQYLTALLMASPLALGDVEIEIIDKLISIPYVEMT
          230          240          250          260          270          280

```

```

          220          230          240          250          260          270
2MEPSP  LRLMERFVKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
          .....
Unipro  LKLMEFVGVSVEHSDSWDQFFIRGGQKYKSPGNAYVEGDASSASYFLAGAAITGGTITVE
          290          300          310          320          330          340

```

```

          280          290          300          310          320          330
2MEPSP  xxxxxSLQGDVKFAEVLGMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA
          .....
Unipro  GCGTSSLQGDVKFAEVLGQMGAEVTVTENSVTGKPPRDSSGRKHLRAVDVNMNKMMPDVA
          350          360          370          380          390          400

```

```

          340          350          360          370          380          390
2MEPSP  MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
          .....
Unipro  MTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGAIVVEEGPDYCVITPPEKLN
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP  VTAIDTYDDHRMAMAFSLAACAEVPTIRDPGCTRKTFPDYFDVLSTFVKN
          .....
Unipro  VTAIDTYDDHRMAMAFSLAACADVPVTIKDPSCTRKTFPDYFEVLQRFKAH
          470          480          490          500          510

```

A8YTA6_9ASTR | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags: Fragment; | *Conyza sumatrensis*. | AA|447
Length = 447
initn: 1774 initl: 1434 opt: 1779 Z-score: 3517.5 bits: 660.0 E(): 2.7e-187
Smith-Waterman score: 1779; 79.138% identity (91.156% similar) in 441 aa
overlap (6-445:7-447)

```

          10          20          30          40          50
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGA
          .....
Unipro  KPSTVPEIVLQPIKEISGTVNLP GSKSLSNRILLLAALAE GTTIVDNLLNSDDVHYMLGA
          10          20          30          40          50          60

```

```

          60          70          80          90          100          110
2MEPSP  LRTLGLSVEADKAAKRAVVVGC GCGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxx
          .....
Unipro  LRTLGLNVEEDGAIKRAIVEGCGMFPV GKEAKDDIQLFLGNAGTAMRPLTAAVTAAGGN
          70          80          90          100          110          120

```

```

          120          130          140          150          160          170
2MEPSP  xxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSI
          .....
Unipro  SSYVLDGVPRMRERPIGDLVTGLKQLGADVDCSLGTNCPVVRVVG GGLPGGKVKLSGSI
          130          140          150          160          170          180

```

```

          180          190          200          210          220          230
2MEPSP  SSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKA EHSWSWDRFYIKG
          ....
Unipro  SSQYL TALLMASPLALGDVEIEIIDKLISIPYVEMTLKLM EWFVGSVEHSWSWDRFFIRG
          190          200          210          220          230          240

```

```

          240          250          260          270          280          290
2MEPSP  GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
          .....
Unipro  GQKYKSPGNAYVEGDASSASYFLAGAAITGGTITVEGCGTSSLQGDVKFAEVLGQMGAEV
          250          260          270          280          290          300

```

```

          300          310          320          330          340          350
2MEPSP  TWTETSVTVTGPPREPFG RKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRV
          .....
Unipro  TWTENSVTVKGPPRDSSGRKHLRAVDVNMNKMPDVAMTLAVVALYADGPTAIRDVASWRV
          310          320          330          340          350          360

```

```

          360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEV
          .....
Unipro  KETERMIAICTELRKLGATVEEGPDYCVITPPEKLNVT AIDTYDDHRMAMAFSLAACADV
          370          380          390          400          410          420

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP PVTIRDPGCTRKTFPDYFDVLSTFVKN
      .....
Unipro PVTIKDPSCTRKTFPDYFEVLQRFK
      430      440

```

Q946V0_9LAMI | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Dicliptera chinensis*. | AA | 516
Length = 516
initn: 1527 initl: 824 opt: 1772 Z-score: 3502.7 bits: 657.5 E(): 1.8e-186
Smith-Waterman score: 1772; 79.224% identity (91.324% similar) in 438 aa
overlap (6-442:77-513)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                                .....
Unipro ISVSNVGKSRQLQLQVAAAKTAEKPPAVPEIVLQPIKDISGTVKLPGSKSLSNRVLLLA
      50      60      70      80      90     100

      40      50      60      70      80      90
2MEPSP ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
      .....
Unipro ALSEGTTVVENLLSSEDIHYMLGALRTLGLAVEEDKANQKAVVEGCVGQFPASKEGKDEV
      110     120     130     140     150     160

      100     110     120     130     140     150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      .....
Unipro QLFLGNAGTAMRPLTAAVVAAGGNARYVLDGVPRLRERPIGDLVTGLKQLGADVDCFLGT
      170     180     190     200     210     220

      160     170     180     190     200     210
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      .....
Unipro NCPVVRVVGKGLPGGKVKLSGVSQYLTALLMSAPLALGDVEIEIVDKLISVPYVEMT
      230     240     250     260     270     280

      220     230     240     250     260     270
2MEPSP LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
      .....
Unipro LKLMERFGVYVEHTDNWDRFLIRGGQKYKSPGNAFVEGDASSASYFLAGAAVTGGTVTVE
      290     300     310     320     330     340

      280     290     300     310     320     330
2MEPSP xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA
      .....
Unipro GCGTSSLQGDVKFAEVLEKMGAKVSWTENSVTVTGPPRVP-GKKHLRAIDVNMNKMMPDVA
      350     360     370     380     390     400

      340     350     360     370     380     390
2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      .....
Unipro MTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLATVEEGPDYCIITPPEKLN
      410     420     430     440     450     460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP  VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
          .....
Unipro  VTAIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFPNYFDVLSTYSNH
          470          480          490          500          510

```

Q30CZ8_FAGSY | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; | *Fagus sylvatica* (Beechnut). | AA | 520
Length = 520
initn: 1754 initl: 1409 opt: 1772 Z-score: 3502.6 bits: 657.5 E(): 1.8e-186
Smith-Waterman score: 1772; 78.005% identity (90.476% similar) in 441 aa
overlap (6-445:80-520)

```

                                10          20          30
2MEPSP                      MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLA
                                .....
Unipro  KHVCAVGS AKVGTFRVSASVATAEKPSTVPEIVLQPIKDISGTIKLPGSKSLSNRILLLA
          50          60          70          80          90          100

```

```

          40          50          60          70          80          90
2MEPSP  ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
          .....
Unipro  ALSEGTTVVDNLLNSDDIHYMLGALKTLGLRVEEDKAIKRAIVEGCGGLFPVGKESRDEI
          110          120          130          140          150          160

```

```

          100          110          120          130          140          150
2MEPSP  QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVP RMRERPIGDLVVGLKQLGADVDCFLGT
          .....
Unipro  QLFLGNAGTAMRPLTAAVTVAGGNSSYVLGDGVP RMRERPIGDLVDGLQQLGANVDCFLGT
          170          180          190          200          210          220

```

```

          160          170          180          190          200          210
2MEPSP  DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
          .....
Unipro  KCPVVRVFGKGLPGGKVKLSGSISSQYLTALVMAAPLALGDVEIEIIDKLISVPYVEMT
          230          240          250          260          270          280

```

```

          220          230          240          250          260          270
2MEPSP  LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
          .....
Unipro  LKLMERFGISVEHNDSWDRFLIRGGQKYKSPGNAYVEGDASSASYFLAGAAVTGGTVTVE
          290          300          310          320          330          340

```

```

          280          290          300          310          320          330
2MEPSP  xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKM PDVA
          .....
Unipro  GCGTSSLQGDVKFAEVLEKMGAKVTWSETSVTVTGPPQDSSKKKHLRAIDVNMNKM PDVA
          350          360          370          380          390          400

```

```

          340          350          360          370          380          390
2MEPSP  MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
          .....
Unipro  MTLAVVALFADGPTAIRDVASWRVKETERMIAVCTELRKL GATVEEGPDYCVITPPEKLN
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
      :: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
Unipro VTDIDTYDDHRMAMAFSLAACGDVPVTINDPGCTRKTFFPDYFEVLQRFTKH
          470          480          490          500          510          520

```

A7PYE6_VITVI | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; | *Vitis vinifera* (Grape). | AA|521
Length = 521
initn: 1772 initl: 1442 opt: 1771 Z-score: 3500.6 bits: 657.1 E(): 2.4e-186
Smith-Waterman score: 1771; 78.555% identity (90.971% similar) in 443 aa
overlap (4-445:79-521)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL
                        : ::::::::::::::: :::::::::::::::
Unipro HERAVSNSIVSVRAPFRVSASVATKEKPSTAPEIVLQPIKEISGTITLP GSKSLSNRILL
          50          60          70          80          90          100

```

```

          40          50          60          70          80          90
2MEPSP LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKE-
      ::::::::::::::::::::::::::::::::::::::: :: .. ::::: ::::: ..
Unipro LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLHVEEQSENKRIVVQCGGQFPAGNGSVG
          110          120          130          140          150          160

```

```

          100          110          120          130          140          150
2MEPSP EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVL DGVPRMRERPIGDLVVGLKQLGADVDCFL
      ::::::::::: :: :... .. :::::::::::::::::::::::::::::::::::::
Unipro EVQLFLGNAGTAMRPLTAAVTAAGGNASYVL DGVPRMRERPIGDLVTGLKQLGADVNCFL
          170          180          190          200          210          220

```

```

          160          170          180          190          200          210
2MEPSP GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      ::::::::::: :::::::::::::::::::: .. .. :::::::::::::::::::::
Unipro GTNCPVVRVNGGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISIPYVE
          230          240          250          260          270          280

```

```

          220          230          240          250          260          270
2MEPSP MTLRLMERFVGKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      ::::::::::: ::::::::::::::: :::::::::::::: :::::::::::::: ..
Unipro MTLKLMERFGVSVEHSNTWDRFLIRGGQKYKSPGNAFVEGDASSASYFLAGAAVTGGTVT
          290          300          310          320          330          340

```

```

          280          290          300          310          320          330
2MEPSP xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
      .. ::::::::::: ::::::::::: :::::::::::::: ::::::::::::::
Unipro VEGCGTSSLQGDVKFAEVLEQMGAKVSWMENSVTVTGPPRDSSGRKHLRAIDVNMNKMPD
          350          360          370          380          390          400

```

```

          340          350          360          370          380          390
2MEPSP VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      ::::::::::: ::::::::::::::: :::::::::::::: ::::::::::::::
Unipro VAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKL GATVEEGPDYCVITPPEK
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP PVTIRDPGCTRKTFPDYFDVLSTFVKN
      .....
Unipro PVTIKDPSCTRKTFPDYFEVLQRFK
      430      440

```

Q946U9_9LAMI | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Dicliptera chinensis*. | AA | 516
Length = 516
initn: 1514 initl: 812 opt: 1766 Z-score: 3490.8 bits: 655.3 E(): 8.4e-186
Smith-Waterman score: 1766; 78.995% identity (91.096% similar) in 438 aa
overlap (6-442:77-513)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                                .....
Unipro ISVSNVGKSRQLQLQVAAAAKTAEKPPAVPEIVLQPIKDISGTVKLPGSKSLSNRVLLLA
      50      60      70      80      90     100

      40      50      60      70      80      90
2MEPSP ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
      .....
Unipro ALSEGTTVVENLLSSEDIHYMLGALRTLGLHVEEDKANQKAVVEGCVGQFPASKEGKDEI
      110     120     130     140     150     160

      100     110     120     130     140     150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      .....
Unipro QLFLGKAGTAMRPLTAAVVAAGGNARYVLDGVPRMRERPIGDLVTGLKQLGADVDCFLGT
      170     180     190     200     210     220

      160     170     180     190     200     210
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      .....
Unipro NCPVVRVVGKGLPGGKVKLSGSVSSQYL TALLMSAPLALGDVEIEIVDKLISVPYVEMT
      230     240     250     260     270     280

      220     230     240     250     260     270
2MEPSP LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
      .....
Unipro LKLMERFGVYVEHTDNWDRFLIRGGQKYKSPGNAFVEGDASSASYFLAGAAVTGGTVTVE
      290     300     310     320     330     340

      280     290     300     310     320     330
2MEPSP xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA
      .....
Unipro GCGTSSLQGDVKFAEVLEKMGAKVSWTENSVTVTGPPRVP-GKKHLRAIDVNMNKMMPDVA
      350     360     370     380     390     400

      340     350     360     370     380     390
2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      .....
Unipro MTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKL GATVEEGPDYCIITPPEKLN
      410     420     430     440     450     460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP  VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFDYFDVLSTFVKN
          .....
Unipro  VTAIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFFPNYFDVLSTYSNH
          470          480          490          500          510

```

B2CM20_9ASTE | 5-enolpyruvylshikimate 3-phosphate synthase; EC=2.5.1.19;
| *Calystegia hederacea*. | AA | 520
Length = 520
initn: 1753 initl: 1428 opt: 1766 Z-score: 3490.7 bits: 655.3 E(): 8.4e-186
Smith-Waterman score: 1766; 78.054% identity (90.045% similar) in 442 aa
overlap (5-445:79-520)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLNRILL
                                .....
Unipro  TGKDSIFTAARSPLKVRASVATAKKPSMAPEEIVLQPIKEISGTVKLPGSKSLNRILL
          50          60          70          80          90          100

```

```

          40          50          60          70          80          90
2MEPSP  AALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKEE
          .....
Unipro  AALSQGTTVVDNLLSSDDIHYMLGALRTLGLRVEEDSAIQRATVEGSGGLFPASNESNDE
          110          120          130          140          150          160

```

```

          100          110          120          130          140          150
2MEPSP  VQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLG
          .....
Unipro  IQLFLGNAGTAMRPLTAAVVAAGGNARYVLDGVPRMRERPIGDLVEGLKQLGADIDCFLG
          170          180          190          200          210          220

```

```

          160          170          180          190          200          210
2MEPSP  TDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEM
          .....
Unipro  TNCPPVRVIGKGGPLGGKVKLSGSSVSSQYLTALLMAAPLALGDVEIEIVDKLISVPYVEM
          230          240          250          260          270          280

```

```

          220          230          240          250          260          270
2MEPSP  TLRIMERFGVKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
          .....
Unipro  TIKLIMERFGVSVSHSDSWDRFLIRGGQKYKSPGNAYVEGDASSASYFLAGAAVTGGTITV
          290          300          310          320          330          340

```

```

          280          290          300          310          320          330
2MEPSP  xxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMKMPDV
          .....
Unipro  ECGTSSSLQGDVKFAEVLEKMGAEVSWTENSVTVKGPPRGPSGRKHLRGIDVNMNMKMPDV
          350          360          370          380          390          400

```

```

          340          350          360          370          380          390
2MEPSP  AMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKL
          .....
Unipro  AMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGLATVEEGPDYCIITPPEKL
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP  NVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::: ::
Unipro  NMTEIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFPDYFDVLSKFSKH
          470          480          490          500          510          520

```

Q6E6L4_CONCD | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; | *Conyza canadensis* (Canadian horseweed) | AA|523

Length = 523

initn: 1757 initl: 1426 opt: 1762 Z-score: 3482.8 bits: 653.8 E(): 2.3e-185

Smith-Waterman score: 1762; 77.602% identity (90.950% similar) in 442 aa
 overlap (5-445:82-523)

```

                                10          20          30
2MEPSP                                MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                                ::::::::::::::::::::::::::::
Unipro  VSCNVKNNKNPFKVSATSTKEKPSKAPEEIVLKPIQEISGTVHLPKSKSLSNRILL
          60          70          80          90          100          110

```

```

          40          50          60          70          80          90
2MEPSP  AALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPV-EDAKEE
          ::::::::::::::::::::::::::::::::::::::::::::::: :: ::::: ::::: :::::
Unipro  AALSEGTTVVDNLLNSDDVHYMLGALRALGLNVEENSAIKRAIVEGCGGVFPVGKEAKDE
          120          130          140          150          160          170

```

```

          100          110          120          130          140          150
2MEPSP  VQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLG
          ::::::::::: :: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::
Unipro  IQLFLGNAGTAMRPLTAAVTAAGNSSYILDGVPRMRERPIGDLVTGLKQLGANVDCSLG
          180          190          200          210          220          230

```

```

          160          170          180          190          200          210
2MEPSP  TDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEM
          ::::::::::: : ::::::::::::::::::::::: .. .. . ::::::::::::::::::
Unipro  TNCPPVRVVGSGGLPGGKVKLSGSISSQYLTSLMAAPLALGDVEIEIVDKLISVPYVEM
          240          250          260          270          280          290

```

```

          220          230          240          250          260          270
2MEPSP  TLRRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
          ::::::::::: ::::::::::::::::::::::: ::::::::::::::::::::::::::: . .
Unipro  TLKLMEFVGVSVEHSDTWDRFHVGGQKYKSPGNAYVEGDASSASYFLAGAAITGGTVTV
          300          310          320          330          340          350

```

```

          280          290          300          310          320          330
2MEPSP  xxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMKMPDV
          ..::::::::: ::::::::::::::::::: ::::: ::: ::: :::::::::::
Unipro  EGCGTSSLQGDVKFAEVLGQMGAEVTTWENSVTVKGPPRNSSGRGHLRPVDVNMNMKMPDV
          360          370          380          390          400          410

```

```

          340          350          360          370          380          390
2MEPSP  AMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKL
          ::::::::::: ::::::::::::::::::: ::: ::::: :::::::::::
Unipro  AMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGLATVEEGPDYCVITPPEKL
          420          430          440          450          460          470

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      400      410      420      430      440
2MEPSP NVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .....
Unipro NVT AIDTYDDHRMAMAFSLAACAEVPVTIKDPGCTRKTFPDYFEVLERYTKH
      480      490      500      510      520

```

B3GGJ9_CONAR | 5-enolpyruvylshikimate-3-phosphate synthase; EC=2.5.1.19;
 | *Convolvulus arvensis* (Field bindweed). | AA|520
 Length = 520
 initn: 1746 initl: 1428 opt: 1759 Z-score: 3476.9 bits: 652.7 E(): 5e-185
 Smith-Waterman score: 1759; 78.054% identity (89.593% similar) in 442 aa
 overlap (5-445:79-520)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                                .....
Unipro SGKDSIFTAARSPLKVRASVATAKKPSMAPEEIVLQPIKEISGTVKLPGSKSLSNRILL
      50      60      70      80      90     100

```

```

      40      50      60      70      80      90
2MEPSP AALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKEE
      .....
Unipro AALSQGTTVVDNLLSSDDIHYMLGALGTLGLRVEEDSAIQRATVEGSGGLFPASNESNDE
      110     120     130     140     150     160

```

```

      100     110     120     130     140     150
2MEPSP VQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLG
      .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro IQLSLGNAGTAMRPLTAAVVAAGGNARYVLDGVPRMRERPIGDLVEGLKQLGADIDCFLG
      170     180     190     200     210     220

```

```

      160     170     180     190     200     210
2MEPSP TDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEM
      .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TNCPPVRVIGKGGPRGKVKLSGVSQYLTALLMAAPLALGDVEIEIVDKLISVPYVEM
      230     240     250     260     270     280

```

```

      220     230     240     250     260     270
2MEPSP TLRIMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
      .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro TIKLIMERFGVSVHSDSWDRFYIRGGQKYKSPGNAYVEGDASSASYFLAGAAVTGGTITV
      290     300     310     320     330     340

```

```

      280     290     300     310     320     330
2MEPSP xxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDV
      .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro ECGTSSLQGDVKFAEVLEKMGAEVSWTENSVTVKGPGRGPFGRKHLRGIDVNMNKMMPDV
      350     360     370     380     390     400

```

```

      340     350     360     370     380     390
2MEPSP AMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKL
      .....
Unipro AMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGAIVEEGPDYCIITPPEKL
      410     420     430     440     450     460

```

400 410 420 430 440
 2MEPSP NVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
 :: :: :: ::
 Unipro NVTEIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFPDYFDVLSKFSKH
 470 480 490 500 510 520

	10	20	30	40	50	60
2MEPSP	MAGAEIEIVLQPIKEISGTVKLP	GSKSLSNRILL	LAALSEGTTVVDN	LLNS	EDVHYMLGAL	

Unipro	SKAPEEIVLKPIQEI	SGTVHLP	GSKSLSNRILL	LAALSEGTTVVDN	LLNSDDVHYMLGAL	
	10	20	30	40	50	60

70 80 90 100 110

2MEPSP RTLGLSVEADKAAKRAVVVGC GGFVPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
.: : . :
Unipro RALGLNVEENAAIKRAIVEGCGGVFPVGKEAKDEIQLFLGNAGTAMRPLTA AVTAAGGNS

70 80 90 100 110 120

120 130 140 150 160 170
 2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
 . :
 Unipro SYVLDGVPRMRERPIGDLVTGLKQLGANVDCSLGTNCPVRVVGSGGLPGGKVKLSGSIS
 130 140 150 160 170 180

	180	190	200	210	220	230
2MEPSP	SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG					
	:::	..	.	:	:	:
Unipro	SQYLTSLLMAAPLALGDVEIEIIDKLISVPYVEMTLKLMERFGVSVEHSDTWDRFYVRGG					
	190	200	210	220	230	240

```

      240      250      260      270      280      290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLLEMMGAKVT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro QKYKSPGSAYVEGDASSASYFLAGAAITGGTIVTVEGCGTSSSLQGDVKFAEVLGQMGA EVT
      250      260      270      280      290      300

```

300 310 320 330 340 350
 2MEPSP WTETSVTVTGPPREPFGFRKHLKAIDVNMNMKMPDVAMTLAVVALFADGPTAIRDVASWRVK
 ::::: ::::: :: :: ::
 Unipro WTENSVTVKGPPRNSSRGELRPVDVNMNMKMPDVAMTLAVVALYADGPTAIRDVASWRVK
 310 320 330 340 350 360

360 370 380 390 400 410
 2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
 :::: .:
 Unipro ETERMIAICTELRKL GATVEEGPDYCVITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
 370 380 390 400 410 420

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFFPDYFDVLSTFVKN
      .....
Unipro VTIKDPGCTRKTFFPDYFEVLERYTKH
      430      440

```

A7Y7Y2_GOSHI | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Gossypium hirsutum* (Upland cotton) | AA|521
Length = 521
initn: 1740 initl: 1436 opt: 1752 Z-score: 3463.0 bits: 650.2 E(): 3e-184
Smith-Waterman score: 1752; 77.652% identity (90.519% similar) in 443 aa
overlap (4-445:79-521)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                                .....
Unipro VVKNNKGFGSIKARSLKVSASTATAEKPSRASEIVLQPIKEISGTVKLPGSKSLSNRILL
      50      60      70      80      90     100

```

```

      40      50      60      70      80      90
2MEPSP LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAK-E
      .....
Unipro LAALSEGTTVVENLLNSDDVHMLVALGKLGKLYVKHDSEKKQAIVEGCGGQFPVGKGEGQ
      110     120     130     140     150     160

```

```

      100     110     120     130     140     150
2MEPSP EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
      .....
Unipro EIELFLGNAGTAMRPLTAAITAAGGNSSYVLDGVPRMRERPIGDLVTGLKQLGADVDCIL
      170     180     190     200     210     220

```

```

      160     170     180     190     200     210
2MEPSP GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      .....
Unipro GTNCPVRIEGKGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISIPYVE
      230     240     250     260     270     280

```

```

      220     230     240     250     260     270
2MEPSP MTLRLMERFVGKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      .....
Unipro MTMKLMERFGVTVEHTDSWDRFFIRGGQKYMSPGNAYVEGDASSASYFLAGAAVTGGTVT
      290     300     310     320     330     340

```

```

      280     290     300     310     320     330
2MEPSP xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
      .....
Unipro VEGCGTSSLQGDVKFAEVLEMMGAKVTWTENSVTVTGPPRNSSGRKHLRAIDVNMNKMPD
      350     360     370     380     390     400

```

```

      340     350     360     370     380     390
2MEPSP VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      .....
Unipro VAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGAATVEEGPDYCVITPPEK
      410     420     430     440     450     460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
          .....
Unipro  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIKDPGCTRKTFFPDYFEVLDRVTKH
          470          480          490          500          510          520

```

B2C1I9_GOSHI |5-enolpyruvylshikimate-3-phosphate synthase; |*Gossypium hirsutum*
(Upland cotton) |AA|521
Length = 521
initn: 1739 initl: 1435 opt: 1751 Z-score: 3461.0 bits: 649.8 E(): 3.8e-184
Smith-Waterman score: 1751; 77.427% identity (90.519% similar) in 443 aa
overlap (4-445:79-521)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                              .....
Unipro  VVKNNKGFGSIKVRSLKVSASTATAEKPSRASEIVLQPIKEISGTVKLPGSKSLSNRILL
          50          60          70          80          90          100

```

```

                40          50          60          70          80          90
2MEPSP  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVCGGKFPVEDAK-E
          .....
Unipro  LAALSEGTTVVENLLNSDDVHMLVALGKLGKLYVKHDSEKKQAIVEGCGGQFPVGKGEGQ
          110          120          130          140          150          160

```

```

                100          110          120          130          140          150
2MEPSP  EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPVRMRERPIGDLVVGLKQLGADVDCFL
          .....
Unipro  EIELFLGNAGTAMRPLTAAITAAGGNSSYVLGDGVPVRMRERPIGDLVTGLKQLGADVDCFL
          170          180          190          200          210          220

```

```

                160          170          180          190          200          210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
          .....
Unipro  GTNCPVRIEGKGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISIPYVE
          230          240          250          260          270          280

```

```

                220          230          240          250          260          270
2MEPSP  MTLRLMERFVGKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
          .....
Unipro  MTIKLMERFGVTVEHTDSWDRFFIRGGQKYMSPGNAYVEGDASSASYFLAGAAVTGGTVT
          290          300          310          320          330          340

```

```

                280          290          300          310          320          330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
          .....
Unipro  VEGCGTSSLQGDVKFAEVLEMMGAKVTWTKNSVTVTGPPRNPSGRKHLRAIDVNMNKMPD
          350          360          370          380          390          400

```

```

                340          350          360          370          380          390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
          .....
Unipro  VAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGAATVEEGPDFCVITPPEK
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
          .....
Unipro  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIKDPGCTRKTFPDYFEVLARVTKH
          470          480          490          500          510          520

```

Q1KPV4_9ASPA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; | *Allium macrostemon*. | AA | 522

Length = 522

initn: 1725 initl: 1417 opt: 1727 Z-score: 3413.4 bits: 641.0 E(): 1.7e-181

Smith-Waterman score: 1727; 76.871% identity (90.930% similar) in 441 aa overlap (6-445:82-522)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                              .....
Unipro  RIGNCRRASVRVLASLAAVENAAEKVAVMPEITLQPIKEITGTVNLPGSKSLSNRILLLA
          60          70          80          90          100          110

```

```

                40          50          60          70          80          90
2MEPSP  ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
          .....
Unipro  ALAEGTTIVDNLLNSDDVSYMLAALKTLGLSVEDDRVNKRAVVGSGGLFPVGKESQKEV
          120          130          140          150          160          170

```

```

                100          110          120          130          140          150
2MEPSP  QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
          .....
Unipro  QLFLGNAGTAMRPLTAAVTAAGGNASYILDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
          180          190          200          210          220          230

```

```

                160          170          180          190          200          210
2MEPSP  DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
          .....
Unipro  DCPVVRVDANGGLPGGKVKLSDSISSQYLTALLMAAPLALGDVEIEIIDKLISIPYVEMT
          240          250          260          270          280          290

```

```

                220          230          240          250          260          270
2MEPSP  LRLMERFVGKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
          .....
Unipro  LKLMEFVGHVHDSSTWDRFFIKGGQKYKSPGNAYVEGDASSASYFLAGAAVTGGTVTVE
          300          310          320          330          340          350

```

```

                280          290          300          310          320          330
2MEPSP  xxxxxSLQGDVKFAEVLNMGAKVTWTETSVTVTGPPREPFGKHLKAIDVNMNKMMPDVA
          .....
Unipro  GCGTSSLQGDVKFAEVLNMGAKVTWTENSVTVTGPPQDPQKKRLKAVDVNMNKMMPDVA
          360          370          380          390          400          410

```

```

                340          350          360          370          380          390
2MEPSP  MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
          .....
Unipro  MTLAVVALYADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGPDYCIITPPEKLN
          420          430          440          450          460          470

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP  VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
          :: ::::::::::::::::::::::::::::::::::::::::::::::
Unipro  VTEIDTYDDHRMAMAFSLAACSDVPVTIKDPGCTRKTFFPDYFEVLERYAKH
          480          490          500          510          520

```

080428_ORYSJ | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; | *Oryza sativa subsp. japonica* (Rice). | AA | 391
Length = 391
initn: 1713 initl: 1584 opt: 1714 Z-score: 3389.5 bits: 636.2 E(): 3.6e-180
Smith-Waterman score: 1714; 85.934% identity (93.862% similar) in 391 aa
overlap (56-445:1-391)

```

          30          40          50          60          70          80
2MEPSP  SLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKF
          :: ::::::::::::::::::::::::::::::::::::::::::::::
Unipro                                     MLEALKALGLSVEADKVAKRAVVVGCGGKF
                                     10          20          30

```

```

          90          100          110          120          130          140
2MEPSP  PVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQL
          ::: :::::::::::::: :: :... .. ::::::::::::::::::::::::::::::
Unipro  PVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQL
          40          50          60          70          80          90

```

```

          150          160          170          180          190          200
2MEPSP  GADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDK
          :::::::::::::::::::: :: .. . ::::::::::::::
Unipro  GADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDK
          100          110          120          130          140          150

```

```

          210          220          230          240          250          260
2MEPSP  LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGA
          :::::::::::::::::::: ::::::::::::::::::::
Unipro  LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGA
          160          170          180          190          200          210

```

```

          270          280          290          300          310          320
2MEPSP  AIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGKHLKAID
          :: . . . ::::::::::::::::::::::::::::::::::::::
Unipro  AITGGTVTVQCGTTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPREPYGKHLKAID
          220          230          240          250          260          270

```

```

          330          340          350          360          370          380
2MEPSP  VNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDY
          :::::::::::::::::::: ::::::::::::::::::::
Unipro  VNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDY
          280          290          300          310          320          330

```

```

          390          400          410          420          430          440
2MEPSP  CIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVK
          :::::::::::::::::::: ::::::::::::::::::::
Unipro  CIITPPEKLNIT AIDTYDDHRMAMAFSLAACADVPVTIRDPGCTRKTFFPDYFDVLSTFVR
          340          350          360          370          380          390

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	400	410	420	430	440
2MEPSP	LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN				
	.. . : .				
Unipro	VKPAEIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFPDYFQVLESITKH				
	470	480	490	500	510

Q6E6L3_CONCD | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags: Fragment; | *Conyza canadensis* (Canadian horseweed) | AA|454
Length = 454
initn: 1699 initl: 1370 opt: 1704 Z-score: 3368.7 bits: 632.5 E(): 5.2e-179
Smith-Waterman score: 1704; 75.339% identity (90.045% similar) in 442 aa
overlap (5-445:13-454)

	10	20	30	40	50
2MEPSP	MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSED				
	: .				
Unipro	ANTTKEKPSKAPEEIVLKPIQEISGTVHLPGSKSLSNRILLLAALSEGTTVVDNLLNSDD				
	10	20	30	40	50

	60	70	80	90	100	110
2MEPSP	VHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEVQLFLGNAGIAMRSLxxx					
	: : : : : : : : : : . : .					
Unipro	VHYMLGALRALGLNIEENAAIKRAIVEGCGGLFPVGKEAKDEIQLFLGNAGTAMRTLTA					
	70	80	90	100	110	120

	120	130	140	150	160	170
2MEPSP	xxxxxxxxxxYVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGK					
	... : .					
Unipro	VTAAGGNLSYVLDGVPRMRERPIGDLVTGLKQLGVNVDCSLGTNCPVHVVGSGGLPGGK					
	130	140	150	160	170	180

	180	190	200	210	220	230
2MEPSP	VKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSW					
	: : : : : : : : : : . . . : .					
Unipro	VKLSGSISSVYLTSLLMAAPLALGDVEIEIIDKLISVPYVRMTLKLMMQRFVGSVEHSDTL					
	190	200	210	220	230	240

	240	250	260	270	280	290
2MEPSP	DRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVL					
	: . . . : : : : : : : : : : : : : .					
Unipro	DRFHVRGGQKYKSPGNAYVESDASSASYFLAGAAITGGTVTVEGCGTSSLQGDVKFAEVL					
	250	260	270	280	290	300

	300	310	320	330	340	350
2MEPSP	EMMGAKVTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIR					
	. : .					
Unipro	GQMGAEVTWTENSVTVKGPPRNSSGRGHLRPVDVNMNKMPDVAMTLAVVALYADGPTAIR					
	310	320	330	340	350	360

	360	370	380	390	400	410
2MEPSP	DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIAIDTYDDHRMAMAFS					
	: .					
Unipro	DVASWRVKETERMIAICTELRKLGAIVVEEGPDYCVTTTPEKLNVTIAIDTYDDHRMAMTFS					
	370	380	390	400	410	420

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          420          430          440
2MEPSP LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .....
Unipro LAACADVPTIKDPGCTRKSPFDYFEVLERYTKH
          430          440          450

```

A8YTA8_9ASTR | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags:
Fragment; | *Conyza sumatrensis*. | AA|446
Length = 446
initn: 1696 initl: 1369 opt: 1701 Z-score: 3362.9 bits: 631.4 E(): 1.1e-178
Smith-Waterman score: 1701; 75.113% identity (90.271% similar) in 442 aa
overlap (5-445:5-446)

```

          10          20          30          40          50          60
2MEPSP MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      .....
Unipro SKAPEEIVLKPIQEISGTVHLPKSGSKSLSNRILLLAALSEGTTVVDNLLNSDDVHYMLGAL
          10          20          30          40          50          60

```

```

          70          80          90          100          110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .....
Unipro RAVGLNVEENAAIKRAIVEGCGGVFPVGKEAKDEIQLFLGNSGTAMRTLTAAITAAGGNS
          70          80          90          100          110          120

```

```

        120        130        140        150        160        170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .....
Unipro SYVLDGVPRMRERPIGDLVTGLKQLGVNVDCSLGTNCPVVRVVGSGGLPGGKVRLSGSIS
        130        140        150        160        170        180

```

```

        180        190        200        210        220        230
2MEPSP SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      : : .. . : .....
Unipro SVYLTSLMLAAPVALGDVEIEIIDKLISVPYVQMTLKLKMQFGVSVEHSDTLDRFHVRRG
        190        200        210        220        230        240

```

```

        240        250        260        270        280        290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      .....
Unipro QKYKSPGNAYVEGDASSASYFLAGAAITGGTVTVEGCGTSSLQGDVKFAEVLGQMGA EVT
        250        260        270        280        290        300

```

```

        300        310        320        330        340        350
2MEPSP WTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVK
      .....
Unipro WTENSVTVKGPPRNSSGRGHLRPVDVNMNKMPDVAMTLAVVALYADGRTAIRDVASWRVK
        310        320        330        340        350        360

```

```

        360        370        380        390        400        410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
      .....
Unipro ETERMIAICTELRKL GATVEEGPDYCVITPPEKLNVT AIDTYDDHRMAMTFSLAACADV P
        370        380        390        400        410        420

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      .....
Unipro VTIKDPGCTRKSFPDYFEVLERYTKH
      430      440

```

Q6JDV0_BRACM | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 | *Brassica campestris* (Field mustard). | AA|514
 Length = 514
 initn: 1675 initl: 1375 opt: 1701 Z-score: 3362.0 bits: 631.5 E(): 1.2e-178
 Smith-Waterman score: 1701; 74.944% identity (90.068% similar) in 443 aa
 overlap (4-445:72-514)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                                .....
Unipro QISSWGLKKSNNGSVIRPVKVMASVSTA EKASEIVLQPIRQISGLIKLPGSKSLSNRILL
      50      60      70      80      90     100

```

```

      40      50      60      70      80      90
2MEPSP LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKE
      .....
Unipro LAALSEGTTVVDNLLNSDDINYM LDALNKLGLNVERDSENNRAVVEGCGGIFPASLDSKG
      110     120     130     140     150     160

```

```

      100     110     120     130     140     150
2MEPSP EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
      .....
Unipro DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGELVVGLKQLGADVECTL
      170     180     190     200     210     220

```

```

      160     170     180     190     200     210
2MEPSP GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      .....
Unipro GTNCPVVRV NANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVPYVE
      230     240     250     260     270     280

```

```

      220     230     240     250     260     270
2MEPSP MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      .....
Unipro MTLKLMERFGVSAEHSDSWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
      290     300     310     320     330     340

```

```

      280     290     300     310     320     330
2MEPSP xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
      .....
Unipro VEGCGTTS LQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAVDVNMNKMPD
      350     360     370     380     390     400

```

```

      340     350     360     370     380     390
2MEPSP VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      .....
Unipro VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKL GATVEEGSDYCVITPPAK
      410     420     430     440     450     460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .. :.....:
Unipro  VKPAEIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFPDYFQVLESITKH
                470          480          490          500          510

```

Q9FVP6_ARATH | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 | *Arabidopsis thaliana* (Mouse-ear cress). | AA|521
 Length = 521
 initn: 1672 initl: 1372 opt: 1700 Z-score: 3359.9 bits: 631.1 E(): 1.6e-178
 Smith-Waterman score: 1700; 74.718% identity (90.293% similar) in 443 aa
 overlap (4-445:79-521)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                        :.....:
Unipro  SWGLKKSDLMLNGSEIRPVKVRASVSTA EKASEIVLQPIREISGLIKLPGSKSLSNRILL
        50          60          70          80          90          100

```

```

                40          50          60          70          80          90
2MEPSP  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKE
      .....:.....:
Unipro  LAALSEGTTVVDNLLNSDDIN YMLDALKILGLNVETHSENNRAVVEGCGGVFPASIDSKS
        110         120         130         140         150         160

```

```

                100         110         120         130         140         150
2MEPSP  EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVL DGVPRMRERPIGDLVVGLKQLGADVDCFL
      .....:.....:
Unipro  DIELYLGNAGTAMRPLTAAVTAAGGNASYVL DGVPRMRERPIGDLVVGLKQLGADV ECTL
        170         180         190         200         210         220

```

```

                160         170         180         190         200         210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      .....:.....:
Unipro  GTNCPVVRV NANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIVDKLISVPYVE
        230         240         250         260         270         280

```

```

                220         230         240         250         260         270
2MEPSP  MTLRLMERFGVKA EHSWSDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxx
      .....:.....:
Unipro  MTLKLMERFGVSAEHS ESWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
        290         300         310         320         330         340

```

```

                280         290         300         310         320         330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
      .....:.....:
Unipro  VEGCGTTS LQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAIDVNMNKMPD
        350         360         370         380         390         400

```

```

                340         350         360         370         380         390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      .....:.....:
Unipro  VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKL GATVEEGSDYCVITPPKK
        410         420         430         440         450         460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . : : : : : : : : : : : : : : : : : : : : : : : : . . .
Unipro  VKPAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH
                470          480          490          500          510          520

```

Q8LC97_ARATH | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
|*Arabidopsis thaliana* (Mouse-ear cress). |AA|521
Length = 521
initn: 1668 initl: 1368 opt: 1696 Z-score: 3352.0 bits: 629.6 E(): 4.5e-178
Smith-Waterman score: 1696; 74.492% identity (90.293% similar) in 443 aa
overlap (4-445:79-521)

```

                                10          20          30
2MEPSP                                MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  SWGLKKSDLMLNGSEIRPVKVRASVSTA EKASEIVLQPIREISGLIKLPGSKSLSNRILL
                50          60          70          80          90          100

```

```

                40          50          60          70          80          90
2MEPSP  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKE
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  LAALSEGTTVVDNLLNSDDINMYLDALKILGLNVETHSENNRAVVEGCGGVFPASIDSKS
                110         120         130         140         150         160

```

```

                100         110         120         130         140         150
2MEPSP  EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPMRERPIGDLVVGLKQLGADV ECTL
                170         180         190         200         210         220

```

```

                160         170         180         190         200         210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      . : : : : : : : : : : : : : : : : : . . . : : : : : : : :
Unipro  GTNCPVVRVANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIVDKLISVPYVE
                230         240         250         260         270         280

```

```

                220         230         240         250         260         270
2MEPSP  MTLRLMERFVGKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      . : : : : : : : : : : : : : : : : : : : : : : : : : : .
Unipro  MTLKLMERFVSAEHSESWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
                290         300         310         320         330         340

```

```

                280         290         300         310         320         330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
      . : : : : : : : : : : : : : : : : : . : : : : : : : : :
Unipro  VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAIDVNMNKMPD
                350         360         370         380         390         400

```

```

                340         350         360         370         380         390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      . : : : : : : : : : : : : : : : : : . : : : : : : : :
Unipro  VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGA TVEEGSDYCVITPPKK
                410         420         430         440         450         460

```

	400	410	420	430	440
2MEPSP	LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN				
	.. .	:	:	:	:
Unipro	VKPAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH				
	470	480	490	500	510
					520
A8YTA9_9ASTR 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5. Fragment; <i>Conyza sumatrensis</i> . AA 444					
Length = 444					
initn: 1671 initl: 828 opt: 1668 Z-score: 3297.6 bits: 619.3 E():					
Smith-Waterman score: 1668; 74.208% identity (89.819% similar					
overlap (5-445:5-444)					
	10	20	30	40	50
2MEPSP	MAGAEIEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSD				
	:	:	:	:	:
Unipro	SKAPEEIVLKPIQKISGTVHLPGSKSLSNRILLLAALSEGTTVVDNLLNSD				
	10	20	30	40	50
					60
	70	80	90	100	110
2MEPSP	RTLGLSVEADKAAKRAVVVGC				
	:	:	:	:	:
Unipro	RAVGLNVEENAAIKRAIVEGCGGVFPVGKEAKDEIQFLGNSGTAMRTLTA				
	70	80	90	100	110
					120
	120	130	140	150	160
2MEPSP	xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS				
	:	:	:	:	:
Unipro	SYVLDGVPRMRERPIGDLVTGLKQLGVNVDCSLGTNCPPVRVVGSGGLPGGRVRLSGSIS				
	130	140	150	160	170
					180
	180	190	200	210	220
2MEPSP	SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYEMTLRLMERFGVKAHSDSWDRFYIKGG				
	:	:	:	:	:
Unipro	SVYLTSLMLAAPLALGDVEIEIIDKLISVPYVQMTLKLKMQFGVSVEHSDTLDRFHV				
	190	200	210	220	230
					240
	240	250	260	270	280
2MEPSP	QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVL				
	:	:	:	:	:
Unipro	Q--KSPGNAYVEGDASNASYFLAGAAITGGTVTVEGCGTSSLLGDVKFAEVL				
	250	260	270	280	290
	300	310	320	330	340
2MEPSP	WTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIR				
	:	:	:	:	:
Unipro	WAENSVTVKGPPRNSSGRGHLRPVDVNMNKMMPDVAMALAVVALYADGPTAIR				
	300	310	320	330	340
					350
	360	370	380	390	400
2MEPSP	ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT				
	:	:	:	:	:
Unipro	ETERMIAICTELRKLGATVEEGPDYCVITPPEKLNVT				
	360	370	380	390	400
					410

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYLSTFVKN
      :::::::::::::::
Unipro VTIKDPGCTRKTFPDYFEVLERYTKH
      420      430      440

```

QOPKRO_PHAVU | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Phaseolus vulgaris* (Kidney bean) | AA|522
Length = 522
initn: 1588 initl: 1296 opt: 1654 Z-score: 3268.8 bits: 614.2 E(): 1.9e-173
Smith-Waterman score: 1654; 74.150% identity (88.889% similar) in 441 aa
overlap (6-445:83-522)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                                :::::::::::::::
Unipro KDSMILSGAELSPFKVMSAVATAEKPSTSPSPEIEVEPIKDFSGWIKLPGTKSLSNRILLLA
      60      70      80      90     100     110

```

```

      40      50      60      70      80      90
2MEPSP ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEV
      :::::::::::::::
Unipro ALSEGTTVVDNLLNSDDIHYMLGALKWLGLRVETDKEINGAVVEGCGGIFPASIDSKSDI
      120     130     140     150     160     170

```

```

      100     110     120     130     140     150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      :::::::::::::::
Unipro ELYLGNAGTAMRPLTAAVTAAGGNAWYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      180     190     200     210     220     230

```

```

      160     170     180     190     200     210
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      :::::::::::::::
Unipro NCPVVRVNGGGLPGGKVKLFGSISSQYLTALLMSAPLALGDVEIEIIDKLISVPYVEVT
      240     250     260     270     280     290

```

```

      220     230     240     250     260     270
2MEPSP LRLMERFVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxx
      :::::::::::::::
Unipro LKLMEFVGSVEHWDSDRFLVHGGQKYKSPGNAYVEGDASSASYLLAGAAITGGTVTVE
      300     310     320     330     340     350

```

```

      280     290     300     310     320     330
2MEPSP xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVA
      . :::::::::::::::
Unipro GCGTKSLQGDVKFAEVLEKMGCKVWWTENSVTVTGPPRDLFGRRVLRAIDVNMNKMPPDVA
      360     370     380     390     400     410

```

```

      340     350     360     370     380     390
2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      :::::::::::::::
Unipro MTLAVVALFADGPTTIRDVASWRVKETEEMIAICTELRKLGAATVEEGPDYCVITPPKCLK
      420     430     440     450     460     470

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

A9SYF9_PHYPA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 | *Physcomitrella patens subsp. patens.* | AA|473
 Length = 473
 initn: 1433 initl: 984 opt: 1432 Z-score: 2829.5 bits: 532.8 E(): 5.7e-149
 Smith-Waterman score: 1432; 65.463% identity (83.296% similar) in 443 aa
 overlap (2-439:25-467)

	10	20	30		
2MEPSP	MAGAEIIVLQPIKEISGTVKLPKSKLSNRILLAL				
	.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.				
Unipro	MNKLNNATQKSVAGVARAHVAEKAASVEEITLQPIKTISGKIKLPKSKLSNRTLLAL				
	10	20	30	40	50
	60				
2MEPSP	40	50	60	70	80
	90				
2MEPSP	SEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVCGGKFPV-EDAKEE---				
	.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.				
Unipro	SEGETVVENLLDSEDEVRYMIGALKTLGYDIHEDRAENRCVIKSGGVPFPVARDAEERGQV				
	70	80	90	100	110
	120				
2MEPSP	100	110	120	130	140
	150				
2MEPSP	VQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLG				
	.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.				
Unipro	VKLFLGNAGTAMRPLTAAVTAAGGNASYELDGVPRMRERPIVDLVMGLQQLGADVTCTED				
	130	140	150	160	170
	180				
2MEPSP	160	170	180	190	200
	210				
2MEPSP	-TDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVE				
	.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.				
Unipro	YPNCPVVLINAKGGLPGGTVRLSGKVSSQYLSALLMAAPLALGDVEIIMVDKLVSPVPYD				
	190	200	210	220	230
	240				
2MEPSP	220	230	240	250	260
	270				
2MEPSP	MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxx				
	.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.				
Unipro	MTLRLMERFGVKVDRHDGWERFSIKGGQTYKSPGSAYVEGDASSASYFLAGAAVTGGTIT				
	250	260	270	280	290
	300				
2MEPSP	280	290	300	310	320
	330				
2MEPSP	xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGKHLKAIDVNMNKMPD				
	.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.				
Unipro	VEGCGTTSLQGDVKFAEVLEKMGATVQWGDHTVTVTGAPADFTTGKRLKAIDVDMNAMPD				
	310	320	330	340	350
	360				
2MEPSP	340	350	360	370	380
	390				
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK				
	.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.				
Unipro	VAMTLAVLGLFADGPVAIRDVENWRVKETERMRAIVDELTKLGAEEVEEGQDYCIVTPPKR				
	370	380	390	400	410
	420				
2MEPSP	400	410	420	430	440
	LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN				
	.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.				
Unipro	ITPAKVDTYDDHRMAMAFSLAACGDTAITIRDPGCTRKTFPTYFTELEKLCQH				
	430	440	450	460	470

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Q3ECU2_ARATH | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
|*Arabidopsis thaliana* (Mouse-ear cress). |AA|489
Length = 489
initn: 1659 initl: 1069 opt: 1397 Z-score: 2760.0 bits: 520.0 E(): 4.3e-145
Smith-Waterman score: 1640; 70.726% identity (85.470% similar) in 468 aa
overlap (4-445:22-489)

		10	20	30	40			
2MEPSP		MAGAE	EIVLQPIKEIS	GTVKLP	GSKSL	SNRILL	LAAL	SEGTT
		:	:	:	:	:	:	:
Unipro	MLNGSEIRPVKVRASV	TAEKASEIVLQPIREIS	GLIKLP	GSKSL	SNRILL	LAAL	SEGTT	
	10	20	30	40	50	60		

	50	60	70	80	90	100			
2MEPSP	VVDNLL	NSEDVHYMLGAL	RTLGLS	VEADKAAKRA	VVVGCGKFPVE	-DAKEEVQLFLGNA			
	:	:	:	:	:	:			
Unipro	VVDNLL	NSDDIN	YMLDALKILGL	NVETHSENNRA	VEGCGGVFPAS	IDS	SKSDI	ELYL	GN
	70	80	90	100	110	120			

	110	120	130	140	150	160			
2MEPSP	GIAMRSL	xxxxxxxxxxxxx	YVLDG	VPRMRERPI	GD	LVVGLKQLGAD	VDCFLGTDCPPVRV		
	:	:	:	:	:	:	:		
Unipro	GTAMRPL	TAAVTAAGGN	ASYVLDG	VPRMRERPI	GD	LVVGLKQLGAD	VECTLT	GNCP	PPVRV
	130	140	150	160	170	180			

	170	180	190	200	210	220
2MEPSP	NGIGGL	PGGKVKLSGS	ISSQYxxxxxxxxxxxxx	GDVEIE	IIDKLISIPY	VEMTLRLMERF
	:	:	:	:	:	:
Unipro	NANGGL	PGGKVKLSGS	ISSQYLTALL	MAAPLALGD	VEIEIVDKLIS	VPYVEMTLKLMERF
	190	200	210	220	230	240

	230	240	250	260	270	280
2MEPSP	GVKAEH	SDSWDRFYIKGG	QKYKSPKNAY	VEGDASSASY	FLAGAAI	xxxxxxxxxxxxxSL
	:	:	:	:	:	:
Unipro	GVSAEH	SESWDRFFVKG	GQKYKSPGNAY	VEGDASSASY	FLAGAAIT	GETVTVEGCGTTSL
	250	260	270	280	290	300

	290	300	310	320	330	340
2MEPSP	QGDVKFA	EVLEMMGAKVT	WTETSVTVTG	PPREPFRKHL	KAIDVNMNK	MPDVAMTLAVVA
	:	:	:	:	:	:
Unipro	QGDVKFA	EVLEKMGCKV	SWTENS	VTVTG	PSRDAFGMRHL	RAIDVNMNKMPDVAMTLAVVA
	310	320	330	340	350	360

	350	360	370			
2MEPSP	LFADGPTA	IRDVASWRVK	ETERMVAIRTELTK-----LGA			
	:	:	:			
Unipro	LFADGPTT	IRDVASWRVK	ETERMIAICTELRKVKFF	FLSLSV	YSAH	SKKFCGICVQLGA
	370	380	390	400	410	420

	380	390	400	410	420	430
2MEPSP	SVEEGPD	YCIITPPEK	LNVT	AIDTYDDHR	MAMAFSLA	ACAEVPVTIRDPGCTRKTFPDYF
	:	:	:	:	:	:
Unipro	TVEEGSD	YCVITPPK	KVKPAE	IDTYDDHR	MAMAFSLA	ACADVPITINDPGCTRKTFPDYF
	430	440	450	460	470	480

```

      440
2MEPSP DVLSTFVKN
      .:. .:.
Unipro OVLERITKH

```

2MEPSP ISGTVKLPGSKSLSNRILLLAALSEGTTTVVDNLLNSD¹⁰VHYMLGALRTLGLSVEADKA²⁰AKK³⁰
Unipro D¹⁰NLLNSDDVHYMLGALRALGLNVEENGA²⁰AIK³⁰

80 90 100 110 120 130
2MEPSP RAVVVGCGGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPMRERP
::: : ::
Unipro RAIVEGCGGVFPVGKEAKDEIQFLFLGNAGTAMRPLTAAVTAAGNSSSYVLGDGVPMRERP
 40 50 60 70 80 90

```

          140      150      160      170      180      190
2MEPSP IGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS SQYxxxxxxxxxxxxx
       ::::: ::::: ::::: ::::: ::::: ::::: : :::::::::::::::::::: .. .. .
Unipro IGDLVTGLKQLGANVDCSLGTNCPVRVVGSGGLPGGKVKLSGSIS QYLTSLLMAAPLA
           100         110         120         130         140         150

```

2MEPSP xGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGD

 Unipro LGDVEIEIIDKLISVPYVEMTLKLMERFGVSVEHNDTWDRFHVRRGGQKYKSPGNAYVEGD
 160 170 180 190 200 210

260 270 280 290 300 310
 2MEPSP ASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLLEMMGAKVTWTETSVTVTGPPRE
 :::::::::::::: : : : :
 Unipro ASSASYFLAGAAITGGTvtVEGCGTSSLQGDVKFAEVLGQMGAEVTTWENSVTVKGPPRN
 220 230 240 250 260 270

320 330 340 350 360 370
 2MEPSP PFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTK
 :: ::
 Unipro SSGRGHLRPVDVNMNKMMPDVAMTLAVVALYABGPTAIRDVASWRVKETERMIAICTELRK
 280 290 300 310 320 330

```

          380          390          400          410          420          430
2MEPSP LGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFF
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Unipro LGATVEEGTDYCVITPPEKLNVT AIDTYDD
          340          350          360

```

Q6E6M4_AMARU|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags:
Fragment; |*Amaranthus rudis* (Common waterhemp). |AA|357
Length = 357
initn: 1373 init1: 1249 opt: 1375 Z-score: 2718.4 bits: 511.8 E(): 8.8e-143
Smith-Waterman score: 1375; 75.630% identity (88.796% similar) in 357 aa
overlap (47-402:1-357)

Page 204 of 526

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGC GGFVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro  KVLGLSFTE DRENNILEITGCGGKLPVEGA---ELFLGNAGTAMRPLTA AVAAAGKGT-
              60          70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVG LKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro  FILDGVERMRERPIQDLVDGLVQLGVKAECTMG TGCPPVKVEA-NGLPGGRVELSGSVSS
              120          130          140          150          160          170

              190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKA EHSWSWDRFYIKGGQ
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro  QYLTALLMAAPLCEGSIEIVIVDELISKPYVEMTITLMERFGVKVEKADDLQSF KIQGGQ
              180          190          200          210          220          230

              250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro  KYISPGSAFVEGDASSASYFLAGATITGGTVTVIGCGSESIQGD TNFAYTMEQMGATLEW
              240          250          260          270          280          290

              310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFG RKHLKAIDVNMNKM P DVAMTLAVVALFADGPTAIRDVASWRVKE
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro  GPNSVTCTGP-----KGPLKAIDVNMNAMPDAAMTLAVAALFADGITTIRDVASWRVKE
              300          310          320          330          340

              370          380          390          400          410
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT A-----IDTYDDHRMAMAFSLA
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro  TERMIAICTELRKLGCDVFEGADYCVITPPHKLDPPAKMKANVDIDTYDDHRMAMAFALA
              350          360          370          380          390          400

              420          430          440
2MEPSP  ACAEVPVTIRDPGCTRKT FPDYFDVLSTFVKN
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro  ACGDVDVINDPKCTKKTFTPTFYFDVLKSVAK
              410          420          430

```

Q6VVA4_ORYSJ | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; | *Oryza sativa subsp. japonica* (Rice). | AA|331

Length = 331

initn: 1108 initl: 1108 opt: 1238 Z-score: 2447.4 bits: 461.6 E(): 1.1e-127

Smith-Waterman score: 1238; 82.886% identity (91.946% similar) in 298 aa overlap (56-352:1-298)

```

              30          40          50          60          70          80
2MEPSP  SLSNRILLLAALSEGTTVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGC GGFK
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro  MLEALKALGLSVEADKVAKRAVVVGC GGFK
              10          20          30

```

	90	100	110	120	130	140
2MEPSP	PVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGKQL					
	::: ::::::::::::::: ::: :... .. ::::::::::::::::::::					
Unipro	PVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGKQL					
	40	50	60	70	80	90
	150	160	170	180	190	200
2MEPSP	GADVDCFLGTDCCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDK					
	::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::					
Unipro	GADVDCFLGTCCPPVRVKIGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDK					
	100	110	120	130	140	150
	210	220	230	240	250	260
2MEPSP	LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGA					
	::::::::::: ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::					
Unipro	LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGA					
	160	170	180	190	200	210
	270	280	290	300	310	320
2MEPSP	AIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPPREPFGRKHLKAID					
	::. . . ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::					
Unipro	AITGGTVTVQGCGTTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPPREPYGKKHLKAID					
	220	230	240	250	260	270
	330	340	350	360	370	380
2MEPSP	VNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDY					
	::::::::::: ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::					
Unipro	VNMNKMPDVAMTLAVVALFADGPTAIRDGFLESKGNRKDGCNSDRANKGKFIRSRVLSFF					
	280	290	300	310	320	330
A2T9R8_DUNSA 3-phosphoshikimate 1-carboxyvinyltransferase; <i>Dunaliella salina</i> . AA 514 Length = 514 initn: 1124 initl: 460 opt: 1219 Z-score: 2406.9 bits: 454.7 E(): Smith-Waterman score: 1219; 58.239% identity (80.361% similar overlap (5-445:81-514)						
			10	20	30	
2MEPSP			MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILL			
			... ::::::::::: ::::::::::: ::::::::::: :::::::::::			
Unipro			RRSCSKSSIRSTRRLQTTVCSATLAAHSAPDQLVLQPIKQISGTVRLPGSKSISNRVLLL			
		60	70	80	90	100 110
	40	50	60	70	80	90
2MEPSP	AALSEGTTVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGC GGKFPVEDAKEEV					
	::: ::::::::::: ::::::::::: ::::::::::: ::::::::::: . :: ::::: : . .					
Unipro	AALAEGTTVVKNLLDSDDIRYMGALKGLGIELEERWDKGEMVVKGCGGQFSAEGG----					
	120	130	140	150	160	
	100	110	120	130	140	150
2MEPSP	QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGKQLGADVDCFLGT					
	::::::::::: :: ::::: ... ::::: ::::::::::: :: :: ::::: . :: ::					
Unipro	ELFLGNAGTAMRPLTAAVAAAAGRGK-FVLDGTARMRERPIQDLVDGLVQLGVDACPLGT					
	170	180	190	200	210	220

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      160      170      180      190      200      210
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGD--VEIEIIDKLISIPYVE
      :::::  :::::::::::::::  ..  .  :  .::  :::::  ::::
Unipro GCPVVEVNA-QGLPSGKVQLKGSVSSQYLTAALLMAAPLSKGTEGIEIVITDELVSQPYVD
      230      240      250      260      270      280

      220      230      240      250      260      270
2MEPSP MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxx
      :::::::::::  ::  ..  .  :  .:  ::  .:::::::::::::::::::  .
Unipro MTVQIMERFGVTVERLNLQHMRIPPNQTYKTSGEAFVEGDASSASYFLAGATITGGTVV
      290      300      310      320      330      340

      280      290      300      310      320      330
2MEPSP xxxxxxxxSLQGDVKFAEVLMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMKMPD
      ..:::::::::  :::::  :  .  .  ::::  .  ::  :  :::::  .  :  ::
Unipro VEGCGSASVQGDVRFVEMGLMGAKVEWSLYSIKITGP--SAFG-KPLQGIDHDCNDIPD
      350      360      370      380      390      400

      340      350      360      370      380      390
2MEPSP VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      :::::::::::  :::::  .:::::::::::::  .  :::::  :::::  :::::  .
Unipro AAMTLAVAALFADKPTTIRNVYNWRVKETERMVAIVNETRKLGAATVEEGRDYCVITPPKQ
      410      420      430      440      450      460

      400      410      420      430      440
2MEPSP LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
      ..  :::::::::::::::::::::  :::::::::::::::::::::  :  .  .
Unipro IQSAAIDTYDDHRMAMAFSLAACGPVPTINDPGCTRKTFFPDYFRVLESVTQH
      470      480      490      500      510

```

A8JH48_CHLRE | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Chlamydomonas reinhardtii*. | AA|512
Length = 512
initn: 907 initl: 484 opt: 1197 Z-score: 2363.4 bits: 446.7 E(): 5.3e-123
Smith-Waterman score: 1197; 58.720% identity (78.366% similar) in 453 aa
overlap (5-445:70-512)

```

      10      20      30
2MEPSP          MAGAEEIVLQPIKEISGTVKLPKSKSLSNRILL
      :::::::::::::::::::::
Unipro SVAPAPACSAPAGARRAVVVRASATKEKVEELTIQPVKKIAGTVKLPKSKSLSNRILL
      40      50      60      70      80      90

      40      50      60      70      80      90
2MEPSP AALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEV
      :::::::::::  :::::::::::::::::::::  .  :  .  ::  :::::  :
Unipro AALSEGTTLVKNLLDSDDIRYMGALKALNVKLEENWEAGEMVVHGCGRFDSAGA----
      100      110      120      130      140      150

      100      110      120      130      140      150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      :::::::::::  ::  .  .  .  .  :::::  :::::  ::  :  :::::  :  ::
Unipro ELFLGNAGTAMRPLTAAVVAAGRK-FVLDGVARMRERPIEDLVDGLVQLGVDAKCTMGT
      160      170      180      190      200      210

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      160      170      180      190      200
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxx-----xxxxGD-VEIEIIDKLISI
      :::::  :::  :::  :::::  ..  ..  .  ::  :::  :  ::::
Unipro GCPPVEVNS-KGLPTGKVYLSGKVSSQYLTAALLMAAPLAVPGGAGGDAIEIIKDELVSQ
      220      230      240      250      260      270

      210      220      230      240      250      260
2MEPSP PYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxx
      :::::  :::::  :::  ..  ...  :  :::  :::  .  :::::  :::::  :::
Unipro PYVDMTVKLMERFGVVVERLNLGLQHLRIPAGQTYKTPGEAYVEGDASSASYFLAGATITG
      280      290      300      310      320      330

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMN
      .  .  .  :::::  :::::  .  :::::  .  :::  :  .  :::  .  :
Unipro GTVTVEGCGSDSLQGDVRF AEVMGLLGAKVEWSPYSITITGP--SAFG-KPITGIDHDCN
      340      350      360      370      380      390

      330      340      350      360      370      380
2MEPSP KMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT
      .  :::::  :::::  :::::  .  :::::  :::::  :::  :::::  :::::  :::::
Unipro DIPDAAMTLAVAALFADRPTAIRNVYNWRVKETERMVAIVTELRLKLGAEVEEGRDYCIIVT
      400      410      420      430      440      450

      390      400      410      420      430      440
2MEPSP PPE-----KLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTF
      ::  :  :  .  :::::  :::::  :::  :::::  :::::  :::  :  :  .  .  .
Unipro PPPGGVKGVKANV-GIDTYDDHRMAMAFSLVAAAGVPVIVIRDPGCTRKTFPTYFKVFESV
      460      470      480      490      500

2MEPSP VKN
...
Unipro AQH
510

```

A6MZG3_ORYSI | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags: Fragment; | *Oryza sativa subsp. indica* (Rice). | AA|273
Length = 273
initn: 1152 initl: 1152 opt: 1152 Z-score: 2278.2 bits: 430.0 E(): 2.9e-118
Smith-Waterman score: 1152; 83.150% identity (91.941% similar) in 273 aa
overlap (96-368:1-273)

```

      70      80      90      100      110      120
2MEPSP SVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDG
      :::::  :::  :::  ...  .  :::::
Unipro LFLGNAGTAMRPLTAAVTAAGGNATYVLDG
      10      20      30

      130      140      150      160      170      180
2MEPSP VPMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxx
      :::::  :::::  :::::  :::::  :::::  :::::  ..
Unipro VPMRERPIGDLVVGLKQLGADVDCFLGT ECPVVRVKIGIGGLPGGKVKLSGSISSQYLSA
      40      50      60      70      80      90

```

	190	200	210	220	230	240
2MEPSP	xxxxxxxxxx	GDVEIEI	IDKLISIPY	EMTLRLMER	FGVKA	EHSDSWDRFYIKGGQKYKSP
	..	.	:	:	:	:
Unipro	LLMAAPLAL	GDVEIEI	IDKLISIPY	EMTLRLMER	FGVKA	EHSDSWDRFYIKGGQKYKSP
	100	110	120	130	140	150
	250	260	270	280	290	300
2MEPSP	KNAYVEGDASSASY	FLAGAAI	xxxxxxxxxxxx	SLQGDVK	FAEVLEMMGAKVT	TWTETSV
	:	:	:	:	:	:
Unipro	GNAYVEGDASSASY	FLAGAAIT	GGTVTV	QCGGTT	SLQGDVK	FAEVLEMMGAKVTWTDTSV
	160	170	180	190	200	210
	310	320	330	340	350	360
2MEPSP	TVTGP	PREPFGRKHL	KAIDVNMN	KMPDVAM	TLAVVALFADG	PTAIRDVASWRVKETERMV
	:	:	:	:	:	:
Unipro	TVTGP	PREPYGKKHL	KAIDVNMN	KMPDVAM	TLAVVALFADG	PTAIRDVASWRVKETERMV
	220	230	240	250	260	270
	370	380	390	400	410	420
2MEPSP	AIRTELTKLGAS	VEEGPDYCI	ITPPEKLN	VT	AIDTYDDHR	MAMAFSLAACAEVPVTIRDP
	:	:	:	:	:	:
Unipro	AIR					

A9S302_PHYPA|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
|*Phycomitrella patens subsp. patens.* |AA|454
Length = 454
initn: 1434 initl: 970 opt: 1146 Z-score: 2263.1 bits: 427.9 E(): 2e-117
Smith-Waterman score: 1431; 66.219% identity (81.655% similar) in 447 aa
overlap (5-439:2-448)

		10	20	30	40	50	60
2MEPSP	MAGAE	EEIVLQPIKEIS	GTVKLPGSKSL	SNRILL	LAALSEG	TTVVDNLL	NSEDVHYMLGAL
		::: :::	::: :::	::: :::	::: :::	::: :::	::: :::
Unipro	MEEIT	LQPIKTIS	GTIKLPGSKSL	SNRTLL	LAALSEG	TTVVENLL	DSEDVRYMVAAL
		10	20	30	40	50	
		70	80	90	100	110	
2MEPSP	RTLGLS	VEADKAAKRA	VVGCGGKFPV	----	EDAKEEVQL	FLGNAGIAMR	SLxxxxxxxxxx
	::: :::	::: :::	::: :::	::: :::	::: :::	::: :::	::: :::
Unipro	QTLGFK	IEEDRAANRL	VIEGEGGVFP	IAARDAEK	SGETVNL	FLGNAGTAMR	PLTAAVVAAG
	60	70	80	90	100	110	
	120	130	140	150	160		
2MEPSP	xxxxxY	VL	DG	VPRMRER	PIGDLV	VGLKQLGAD	VDCFLG-TDCPPVRVN-----GIGGLP
	::: :::	::: :::	::: :::	::: :::	::: :::	::: :::	::: :::
Unipro	GNVSY	VL	DG	VPRMRER	PIIDL	VAGLQQLGAD	VKCTEEYPNCP
	120	130	140	150	160	170	
	170	180	190	200	210	220	
2MEPSP	GGKVK	LSGS	ISSQYxxxxxxxxxxxxxx	GDVEIE	IIDKL	SIPYVEM	TLRLMERFGVKAHS
	::: :::	::: :::	::: :::	::: :::	::: :::	::: :::	::: :::
Unipro	GGTVH	LSGK	VSSQYLSALLMA	APLALGD	VEIV	MIDKLVS	VPYVDMTLRLMERFGV
	180	190	200	210	220	230	

```

          410          420          430          440
2MEPSP  AFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
        : : : : : : : : : : : : : : : : : : :
Unipro  AFSLAACGDPVITILDPGCTRKTFPTYFTELEKLCQH
          420          430          440          450

```

210 220 230 240 250 260
 2MEPSP YVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxx
 ::
 Unipro YVDMTIKLMERFGVSV EHSWDRFFIKGGQKYKSPGNAYVEGDASSASYFLAGAAITGG
 160 170 180 190 200 210

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

270	280	290	300	310	320
2MEPSP	xxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFG	GRKHLKAIDVNMNK			

Unipro	TITVEGCGTSSSLQGDVKFAEVLGQMGAEVTWTENSVTVRGPPRNASGRKHLHAV				
	220	230	240	250	260

330	340	350	360	370	380
2MEPSP	MPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITP				

B5S6S6_RALSO | 3-phosphoshikimate 1-carboxyvinyltransferase (Epsp synthase) protein; EC=2.5.1.19; | *Ralstonia solanacearum* (*Pseudomonas solanacearum*). | AA|435
Length = 435
initn: 909 initl: 231 opt: 953 Z-score: 1880.9 bits: 357.2 E(): 3.9e-96
Smith-Waterman score: 953; 48.853% identity (72.936% similar) in 436 aa overlap (5-436:2-427)

10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLP	GSKSLSNRILL	LAALSEGTTVVDNLLN	SEDEVHYMLGAL	

Unipro	MEHLDVGPLKAARGTVKLP	GSKSISNRVLL	LAALAEGETTVVRD	LLSDDDTRV	MLAAL
	10	20	30	40	50

70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSL	xxxxxxxxxxxx			

Unipro	DKLGVRCEPLGTANAYRVVGTGGRFPVKS	---	DLFMGNAGTAIRPLTAALALQ	GGA--	
	60	70	80	90	100

130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS				

Unipro	YTLHGVPRMHERPIGDLVDGLRQVGARIDYAGNAGFPPLAIHAAPVKIDAPIRVRGDVSS				
	120	130	140	150	160

190	200	210	220	230
2MEPSP	QYxxxxxxxx--xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG			

Unipro	QFLTALLMALPLVESVGQVTIEVVGELISKPYIEITLNLMARFGVQVAR-DGWASFTVPT			
	180	190	200	210

240	250	260	270	280	290
2MEPSP	GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV				

Unipro	GVRYPAPGEMFVEGDASSASYFLAAGALGGGPVRVEGVGMSSIQGDVRFADALNRMGANV				
	240	250	260	270	280

300	310	320	330	340	350
2MEPSP	TWTETSVTVTGPPREPFG	GRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRV			

Unipro	MAGDNWIEVRGVERDD-GKLH--ALELDCNHIPDAAMTLAVAALFADGTTTLTNIGSWRV				
	300	310	320	330	340

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      360      370      380      390      400      410
2MEPSP KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT--IDTYDDHRMAMAFSLAACA
      ::::::::::: :: ::::::::::: :: ::::: .. : ::::::::::::::::::: .
Unipro KETDRLTAMATELRKLGAAVEEGTDYIRVTPPSHWTAPAGGIDTYDDHRMAMAFSLAAFG
      350      360      370      380      390      400

      420      430      440
2MEPSP EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      ::: :.:: :. ::::::::::
Unipro PVPVRINDPRCVAKTFPDYFTAFGEIAD
      410      420      430

```

Q6E6L6_PLALA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags:
 Fragment; | *Plantago lanceolata* (English plantain). |AA|265
 Length = 265
 initn: 861 initl: 861 opt: 951 Z-score: 1880.2 bits: 356.3 E(): 4.3e-96
 Smith-Waterman score: 951; 71.321% identity (86.038% similar) in 265 aa overlap
 (60-323:1-265)

```

      30      40      50      60      70      80
2MEPSP RILLLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-E
      ::::::::::: :: ::::: ::::: ::::: .
Unipro LRTLGLDVEEDKANQRATVGGCGGLFPVSK
      10      20      30

```

```

      90      100      110      120      130      140
2MEPSP DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADV
      .....: :: :::: .. ..: ::::::::::::::::::::::::::::::
Unipro ESKDEIQFLFLGNAGTAMRPLTAAVVAAGGDATYVLDGVPRMRERPIGDLVTGLKQLGADI
      40      50      60      70      80      90

```

```

      150      160      170      180      190      200
2MEPSP DCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISI
      ::::: ::::: : ::::::::::::::::::: .. .. . ::::::::::::::
Unipro DCFLGTCKPPVRVIGKGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISV
      100      110      120      130      140      150

```

```

      210      220      230      240      250      260
2MEPSP PYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxx
      ::::::::::: ::::: ::::: ..: ::::::::::: ::::::::::::::
Unipro PYVEMTLKLMERFGVYVEHTDSWDRFLVRGGQKYKSPGKAYVEGDASSASYFLAGAAVTG
      160      170      180      190      200      210

```

```

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMN
      . . .: ::::::::::: ::: ::::: ::::: ::::: :::::
Unipro GTVTVEGCGTSSLQGDVKFAEVLEKMGAEVTWTENSVTVKGPPRNSSGRKHLRAI
      220      230      240      250      260

```

```

      330      340      350      360      370      380
2MEPSP KMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      .:. :   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .
Unipro  DKLGVRCPEPLGTANAYRVVGTGGRFPVKSA---DLFMGNAGTAIRPLTAALALQGGA--
              60          70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .
Unipro  YTLHGVPRMHERPIGDLVDGLRQVGARIDYAGNAGFPPLAIHAAPVKIDAPIRVRGDVSS
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxx--xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG
      .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .
Unipro  QFLTALLMALPLVESVGQVTIEVVGELISKPYIEITLNLMARFGVQVAR-DGWASFAVPT
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
      :   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .
Unipro  GVRYPAPGEMFVEGDASSASYFLAAGALGGGPVRVEGVGMSSIQGDVRFADALNRMGANV
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  TWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRV
      .. :   :   :   :   :   :   :   :   :   :   :   :   :   :
Unipro  MAGDNWIEVRGVERDD-GKLH--ALELDCNHIPDAAMTLAVAALFADGTTTTLTNIGSWRV
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT--IDTYDDHRMAMAFSLAACA
      .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .
Unipro  KETDRLTAMATELRKLGAAVEEGIDYIRVTPPSHWTAPAGGIDTYDDHRMAMAFSLAAFG
              350          360          370          380          390          400

              420          430          440
2MEPSP  EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .
Unipro  PVPVRINDPRCVAKTFPDYFTAFGEIAD
              410          420          430

```

Q00VW9_OSTTA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
|*Ostreococcus tauri*. |AA|316
Length = 316
initn: 958 init1: 386 opt: 935 Z-score: 1847.3 bits: 350.5 E(): 2.9e-94
Smith-Waterman score: 935; 60.991% identity (78.638% similar) in 323 aa overlap
(129-444:1-316)

```

              100          110          120          130          140          150
2MEPSP  GNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPP
      .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .:.   .
Unipro  MRERPIADLVGLVQLGVKAECTMGTCPP
              10          20          30

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      160      170      180      190      200      210
2MEPSP VRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLM
      :.... :.....:.....:.. .. :...: :.....: :.....: :.
Unipro VKIEA-DGLPGGRVELSGSVSSQYL TALLMAAPLCQGSIEIVIVDELISKPYVEMTITLM
      40      50      60      70      80

      220      230      240      250      260      270
2MEPSP ERFGVKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxx
      :.....: : . . : :.....: : : :.....: :.....: :. . .
Unipro ERFGVKVEKSADLQSFKIQGGQKYVSPGSAFVEGDASSASYFLAGATITGGTVTVIGCGS
      90      100      110      120      130      140

      280      290      300      310      320      330
2MEPSP xSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLA
      :.....: :.....: . : :...: :...: . :.....: :.....: :
Unipro ESIQGD TNFAYTMEQMGATLEWGPNSVKCTGP-----QGPKAIDVNMNAMPDAAMTLA
      150      160      170      180      190      200

      340      350      360      370      380      390
2MEPSP VVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT-
      :.....: :.....: :.....: : : :...: : : :.....: : :
Unipro VAALFADGVTTIRDVASWRVKETERMIAICTELRKLGCDFEGSDYCVITPPHKLNPAPK
      210      220      230      240      250      260

      400      410      420      430      440
2MEPSP -----IDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      :.....: :.....: : : :...: :.....: :.....: :.
Unipro MRANVDIDTYDDHRMAMAFALAACGDVDVIINDPTCTKKTFTPTYFDVLKSVVQ
      270      280      290      300      310

Q6E6L9_9ASTR|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags:
Fragment; |Helianthus salicifolius. |AA|264
Length = 264
initn: 926 initl: 857 opt: 931 Z-score: 1840.6 bits: 349.0 E(): 6.9e-94
Smith-Waterman score: 931; 69.697% identity (85.985% similar) in 264 aa overlap
(61-323:1-264)

      40      50      60      70      80
2MEPSP ILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-ED
      :.....: . :...: :...: :...: :.
Unipro RALGLNVEENGEIKRATVEGCGGVFPVPGKE
      10      20      30

      90      100      110      120      130      140
2MEPSP AKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVD
      :.....: : : :...: : : :.....: :.....: :.....: :
Unipro AKDEIQFLFLGNAGTAMRPLTAAVTAAGNSSYILDGVPRMRERPIGDLVTGLKQLGADVD
      40      50      60      70      80      90

      150      160      170      180      190      200
2MEPSP CFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIP
      :.....: : : :.....: :.. .. :.....: :.....: :
Unipro CFLGTNCPVRAANGGLPGGKVKLSGSISSQYL TALLMAAPLALGDVEIEIIDKLISVP
      100      110      120      130      140      150

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      210      220      230      240      250      260
2MEPSP YVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxx
      .....
Unipro YVEMTLKLMERFGVSVSHSDSWDKFYVRGGQKYKSPGNAYVEGDASSASYFLAGAAITGG
      160      170      180      190      200      210

```

```

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNK
      . . . . .
Unipro TVTVEGCGTSSLQGDVKFAEVLGQMGAEVTWTENSVTVRGPPRNASGRGHLRPV
      220      230      240      250      260

```

```

      330      340      350      360      370      380
2MEPSP MPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITP

```

Q400R4_LACSA|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags:
Fragment; |*Lactuca sativa* (Garden lettuce). |AA|231
Length = 231
initn: 926 initl: 926 opt: 926 Z-score: 1831.5 bits: 347.1 E(): 2.2e-93
Smith-Waterman score: 926; 77.922% identity (90.476% similar) in 231 aa overlap
(174-404:1-231)

```

      150      160      170      180      190      200
2MEPSP LGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIID
      .....
Unipro LSGSISSQYLTALLMASPLALGDVEIEIID
      10      20      30

```

```

      210      220      230      240      250      260
2MEPSP KLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAG
      .....
Unipro KLISIPYVEMTLKLMERFGVSVSHSDSWDRFFVRGGQKYKSPGNAYVEGDASSASYFLAG
      40      50      60      70      80      90

```

```

      270      280      290      300      310      320
2MEPSP AAxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAI
      . . . . .
Unipro AAITGGTITVEGCGTSSLQGDVKFAEVLGQMGAEVTWTENSVTVKGPPRNPSGRKHLRAV
      100      110      120      130      140      150

```

```

      330      340      350      360      370      380
2MEPSP DVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPD
      .....
Unipro DVNMNKMMPDVAMTLAVVALYADGPTTIRDVASWRVKETERMIAICTELRKLGLATVEEGAD
      160      170      180      190      200      210

```

```

      390      400      410      420      430      440
2MEPSP YCIITPPEKLNVTAITDYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFV
      . . . . .
Unipro YCVITPPEKLNVAIDTYDDH
      220      230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Q400R3_LACSA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags: Fragment; | *Lactuca sativa* (Garden lettuce). | AA|231
Length = 231
initn: 922 init1: 922 opt: 922 Z-score: 1823.6 bits: 345.7 E(): 6.1e-93
Smith-Waterman score: 922; 77.922% identity (89.610% similar) in 231 aa overlap (174-404:1-231)

	150	160	170	180	190	200
2MEPSP	LGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIID					
				::::::::::	::::::::::
Unipro				LSGSISSQXLTALLMAAPLALGDVEIEIID		
				10	20	30

	210	220	230	240	250	260
2MEPSP	KLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAG					
	::::::::::
Unipro	KLISIPYVEMTLKLMERFGVSVQHSDTWDRFHVQGGQKYKSPGNAYVEGDASSASYFLAG					
	40	50	60	70	80	90

	270	280	290	300	310	320
2MEPSP	AAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAI					
	:::
Unipro	AAITGGTITVEGCGTSSLQGDVKFAEVLGQMGQAQVTWTENSVTVKGPPRDPGRKHLRPV					
	100	110	120	130	140	150

	330	340	350	360	370	380
2MEPSP	DVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPD					
	::::::::::
Unipro	DVNMNKMPDVAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGLATVEEGPD					
	160	170	180	190	200	210

	390	400	410	420	430	440
2MEPSP	YCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFV					
	::::::::::	::::::::::				
Unipro	YCIITPPXKLNVT AIDTYDDH					
	220	230				

Q6E6M3_ASITR | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags: Fragment; | *Asimina triloba* (Pawpaw) | AA|264
Length = 264
initn: 897 init1: 872 opt: 887 Z-score: 1753.4 bits: 332.9 E(): 5e-89
Smith-Waterman score: 887; 66.288% identity (83.712% similar) in 264 aa overlap (60-323:1-264)

	30	40	50	60	70	80
2MEPSP	RILLLAALSEGTTTVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVED					
				::::::::::	::::::::::	::::::::::
Unipro				LQTLGLNVEEDSAANKATVEGCGGQFPVGK		
				10	20	30

	90	100	110	120	130	140
2MEPSP	AKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADV					

Unipro	DAKDIQLFLGNAGTAMRPLTAAVVAAGGKSRYVLDGVPRMRERPIGDLVSGLKQLDADV					
	40	50	60	70	80	90

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      150      160      170      180      190      200
2MEPSP CFLGTDCCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIP
      :  :::::  ::  ::  ::::::::::::::  ..  ..  .  :::::::::::::::
Unipro  CVLGTNCPVVSINANGGLRGKVKLSGTLSSQFLTSILMAAPLALGDVEIEIIDKLISVP
              100      110      120      130      140      150

```

```

      210      220      230      240      250      260
2MEPSP YVENTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxx
      :::::  ::::::::::  ::::  ::::  ..::::::::  ..:::::::::::::
Unipro  YVENTLKLMEFGVSAEHSGSWDRFLVRGGQKYKSPGTAFVEGDASSASYFLAGAAVTGG
              160      170      180      190      200      210

```

```

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGKHLKAIDVNMNK
      .  .  ..::::::::::::  :::::  ::::::::::::::  .:  .:::  .:
Unipro  TVTVEGCGTSSLQGDVKFAEVLEKMGAKVSWTENSVTVTGPPPLDPSRKKRLHGI
              220      230      240      250      260

```

```

      330      340      350      360      370      380
2MEPSP MPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITP

```

Q12CL6_POLSJ | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 EC=2.7.4.14; | *Polaromonas* sp. (strain JS666 / ATCC BAA-500). | AA|668
 Length = 668
 initn: 875 initl: 313 opt: 880 Z-score: 1733.5 bits: 330.5 E(): 6.4e-88
 Smith-Waterman score: 880; 46.120% identity (71.840% similar) in 451 aa overlap
 (1-439:1-440)

```

              10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :  :  .  .  .  .  ::::::::::::::  :::  :::  :::  ::::
Unipro  MFDIEYLDIPPLARAGGTVRLPGSKSISNRVLLLAALSRGQTTVHDLLASDDTAVMLAAL
              10      20      30      40      50      60

```

```

              70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .  ::  ::  .  .  :::  :  :::  :  :::  .  ::::::  :::  ....
Unipro  KQLGCSVA--QHGTTAVIDGLGGQ--VGQAK--ATLFMGNAGTAMRPLTAALALLGGE--
              70      80      90      100      110

```

```

              130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCCPPVRVNGIGGLPGGKVKLSGSISS
      .  :::  ::::::::::  ::::::::::  .  .  :::::  ...  ::::
Unipro  FELSGVARMHERPIGDLVDALRQLGCSIEYLGNEGYPLRLRPAQLKIEEPIVRGVDVSS
              120      130      140      150      160      170

```

```

              190      200      210      220      230
2MEPSP QYxxxxxxxxxxxxG-DVEIEIIDKLISIPYVENTLRLMERFGVKAH---SDSWDRFYI
      :  ..  .  .:  :::::  :::  ::::::::::::::  ...  :::::  :
Unipro  QFLTALLMALPLVAGQDIHIEVVGELISRPIEITLNLKRFGIHVQRPLGADNWQRFTH
              180      190      200      210      220      230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	240	250	260	270	280	290
2MEPSP	KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxx-----xxSLQGDVKFAEVL					

Unipro	PAGSQYQSPGDIHVEGDASSASYFIALGAISEPAAGQNHIEILGVGADSIQGDIFIEAA					
	240	250	260	270	280	290

	300	310	320	330	340	350
2MEPSP	EMMGAKVTWTETSVTVT-GPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAI					

Unipro	RMMGAQIESTPNSLRISRGKPGQGW---PLKAIDLDCNHIPDAAMTLAVMALYADGTTTL					
	300	310	320	330	340	

	360	370	380	390	400	410
2MEPSP	RDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIDTYDDHRMAMAF					

Unipro	RNIASWRVKETDRIAAMACELQKLGAATVEEGADYKITPPLAWKSAAIHTYDDHRIAMCF					
	350	360	370	380	390	400

	420	430	440			
2MEPSP	SLAAC--AEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN					
			
Unipro	SLAAFNPAPQPIRILDPKCVAKTFPDYFEALFAVSEAETAQIPVICIDGPTASGKGTLLA					
	410	420	430	440	450	460

	470	480	490	500	510	520
Unipro	LAAHRLGYHYLDGALYRLSAFAATRAGVALENGADVARIARTLPVFRFRGDRIFLGTEDEV					

A1W8Q4_ACISJ | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 EC=2.7.4.14; |*Acidovorax* sp. (strain JS42). |AA|673
 Length = 673
 initn: 834 initl: 238 opt: 870 Z-score: 1713.6 bits: 326.9 E(): 8.1e-87
 Smith-Waterman score: 870; 46.429% identity (72.545% similar) in 448 aa overlap
 (9-439:9-441)

	10	20	30	40	50	60
2MEPSP	MAGAEIEIVLQPIKEISGTVKLPKSGKSLNRIALLAALSEGTTVVDNLLNSEDVHYMLGAL					
	:
Unipro	MFSTAFLDLPLASVQGSVQLPGSKSISNRVLLAALSQGTTEVRDLLASDDTRVMDAL					
	10	20	30	40	50	60

	70	80	90	100	110	
2MEPSP	RTLGLSVEADKAAKRAV-VVGCCKGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx					
	: :
Unipro	RQLGCTV--DEAGGTVRITGLGSSTP---PSPTQLFMGNAGTAMRPLTAALALLGGE-					
	70	80	90	100	110	

	120	130	140	150	160	170
2MEPSP	xYVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPG---GKVKLS					
	: :
Unipro	-YELSGVPRMHERPIGDLVDALRQLGCRIDYLGNGQFPPLRIAHAGGLPPLQLQAPIRVR					
	120	130	140	150	160	170

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      180      190      200      210      220      230
2MEPSP GSISSQYxxxxxxxx--xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDR
      :...: . . . . . : : : . : : : : : : : : : : : : : : : : : :
Unipro GDVSSQFLTALLMALPLVARQQDVVIEVVGELISKPYIHITLELLARFGIAVQH-EQWQR
      180      190      200      210      220      230

      240      250      260      270      280
2MEPSP FYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxx-----SLQGDVKFA
      : : . : . : : : . : : : : : : : : . . . : : : : : : : : : :
Unipro FTIPAGSRYQSPGLIHVEADASSASYFIAAGAIAGASGQKGIKILGVGLDSIQGDIRFV
      240      250      260      270      280      290

      290      300      310      320      330      340
2MEPSP EVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPT
      : . . : : : : : . . : . . : : : : : : : : : : : : : : : : :
Unipro EAAQAMGARVTGGPNWLEVA---RDTWP---LKAIDLDCNHIPDAAMTLAVMALYARGTT
      300      310      320      330      340

      350      360      370      380      390      400
2MEPSP AIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN--VTAIDTYDDHR
      . : . : : : : : : : : : : : : : : : : : : : . : . : : : : : : :
Unipro VLRNIASWRVKETDRIAAMTCELRLKLGATVEEGADYLCVTPPAQSGDWRAASIHTYDDHR
      350      360      370      380      390      400

      410      420      430      440
2MEPSP MAMAFSLAA--CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . : : : : . : . : : : : . : : : : : : : : : : : : : :
Unipro VAMCFSLAAFNAAGLPVRIEDPKCVAKTFPDYFEALFQVAQTDADRIPVICIDGPTASGK
      410      420      430      440      450      460

Unipro GTVAAAVARALGYHFLDSGALYRVTALAATRAGLEISPDNEPRIAELARQLPVRFAGDRI
      470      480      490      500      510      520

```

A9BM07_DELAS | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 EC=2.7.4.14; | *Delftia acidovorans* (strain DSM 14801 / SPH-1). | AA|675
 Length = 675
 initn: 632 initl: 244 opt: 869 Z-score: 1711.6 bits: 326.5 E(): 1e-86
 Smith-Waterman score: 869; 46.067% identity (73.258% similar) in 445 aa overlap
 (9-439:9-441)

```

      10      20      30      40      50      60
2MEPSP MAGAEIIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      . : . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MFSTAFLDIPPLSRAQGTVSLPGSKSISNRVLLLAALSQGTTVVDLLDSDDTRVMLRAL
      10      20      30      40      50      60

      70      80      90      100      110
2MEPSP RTLGLSVE-ADKAAKRAV-VVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxx
      . : : . : : . . : . : : : : . . : : : : : : : : : : : : .
Unipro QQLGCTVEPATITLGQPV RITGLGGVLP---SGASAELFMGNAGTAMRPLTAALSVLGGD
      70      80      90      100      110

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
:  :::::  .:  :  .:  :::::.....:  .:.....:  :::  ....
Unipro  RQLGLSV-INKENKACIVEGCGGKLPVQEA----DLFMGNAGTAIRPLTAALAMQGGN--
              70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
:  :::  :::::  :::  :::::..  :  :.....  :  :  :::  :::::
Unipro  YRLSGVARMRERPIRDLDVGLRQVGAKIGYELQEGYPPIKILAADIQIKDVVKVRGDVSS
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxxxxxxGD-VEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG-
:  .  .  .  .  .  :::::  :::  :::::.....:  :::  .  :  :  :
Unipro  QFLTALLMALPLVANEPVRIEVIGELISRPYIDITLKLMAFGVTVACPDM-QSFVIPAK
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  --GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
:  ::::  .  :.....:  :::  :.....:  :::  :::  :::  :::
Unipro  TSGAVYKSPGQLLVEGDASSASYFLALGAIGGGPVRVLGVGKESIQGDVAFADALALMGA
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
..:  :  .  :::  ..  .:  :::  ..  .....:.....:.....:.....:
Unipro  NITAGEDWIEVAGV-KNANGK--LNGITIDCTEIPDAAMTLAVAALFAEGPTRLNSIASW
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  RVKETERMVAIRTELTKLGASVEEGPDYCIITPPE----KLNVTATIDTYDDHRMAMAFS
:.....:.....:  ::  :::::.....:  ..  :  :  :.....:
Unipro  RVKETDRIAAAKEKLVGANVEEGADYIVVQAPVLQSDWKSPEGVDTYDDHRMAMCFS
              350          360          370          380          390          400

              420          430          440
2MEPSP  LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
:::  .  .  :::::  :::  ::
Unipro  LAAFGLNALKINDPNCVAKTFPTYFAEFAKVVN
              410          420          430          440

```

A4SW54_POLSQ | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Polynucleobacter* sp. (strain QLW-P1DMWA-1). | AA | 442
Length = 442
initn: 780 initl: 205 opt: 854 Z-score: 1684.7 bits: 320.9 E(): 3.3e-85
Smith-Waterman score: 854; 46.517% identity (69.663% similar) in 445 aa overlap
(1-436:1-434)

```

              10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
:::  :::  .  .  .  .....:.....:.....:.....:  :  ::
Unipro  MSGIPEITIGPFTRAQGSIVLPGSKSISNRALLLAALASGTTTLKNLLDADDTQVMRNAL
              10          20          30          40          50          60

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :  :::::  ::  ::  :::::  :::::  :::::  ::  :::::  ::  :::::  ::
Unipro  RQLGLSV-IDQANHVCVVKGCGGQFPVREA---DLFMGNAGTAIRPLTAALAMQGGN--
              70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :  ::::  :::::  ::  :::::  ::  :::::  ::  :::::  ::  :::::  ::
Unipro  YRLSGVARMHERPIRDLVDGLRQVGAKIDYELQEGYPPIKILAADIQIKDVVKVRGDVSS
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxxxxxxGD-VEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      :  .  .  .  .  .  :::::  ::  :::::  :::::  :::::  .  .  :  :  .
Unipro  QFLTALLMALPLVAQEAVRIEVIGELISRPYIDITLKLMAFVGNVACPDA-QSFVIPAK
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  QK---YKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      .  ::::  :  :::::  :::::  .  .  .  .  .  :::::  :::::  :::::
Unipro  TSDAVYKSPGNLSVEGDASSASYFLALGALGAGPVRVLGVGKDSIQGDVAFADALALMGA
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  KVTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASW
      ::  :  .  ::  .  :  :  ::  .  .  :::::  :::::  :::::  :  :::::
Unipro  KITAGEDWIEVSGVKNA--GGK-LNGITLDCTEIPDAAMTLAVAALFAEGQTRLNNIASW
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN-----VTAIDTYDDHRMAMAFS
      :::::  :::::  ::  :::::  :::::  ::  .  :  .  .  :::::  :::::
Unipro  RVKETDRIAAMAKELKKIGAIVEEGADYIVVQAPALGDWKSPSEGIDTYDDHRMAMCFS
              350          360          370          380          390          400

              420          430          440
2MEPSP  LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      ::  .  .  :::::  :::::  ::
Unipro  LATLGPNTLKINDPNCVAKTFPTYFAEFAKIVS
              410          420          430          440

AlTSA6_ACIAC | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
EC=2.7.4.14; |Acidovorax avenae subsp. citrulli (strain AAC00-1). |AA|679
Length = 679
initn: 786 init1: 215 opt: 854 Z-score: 1681.9 bits: 321.0 E(): 4.8e-85
Smith-Waterman score: 854; 45.190% identity (70.917% similar) in 447 aa overlap
(9-439:9-443)

              10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSGSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :  .  .  :::::  :::::  :::::  ::  ::  :::::  ::  ::
Unipro  MYSTAFDLPLPLDTAGGAVRLPGSKSISNRVLLLAALSEGTTVEVDLLASDDTRVMDLAL
              10          20          30          40          50          60

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGC GGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:
Unipro  REIGCGVDEGAAAQGTVRITGLGTSP---ARSPSKLFLGNAGTAMRPLTAALALLGGE--
              70          80          90          100          110

              130          140          150          160          170
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLP---GGKVKLSG
      .  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:
Unipro  FELSGVPRMHERPIGDLVEALLQLGCHISYLGNGFPPLRIAHAGGVPPLALDAPVRVRG
              120          130          140          150          160          170

              180          190          200          210          220          230
2MEPSP  SISSQYxxxxxxxx--xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRF
      .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:
Unipro  DVSSQFLTALLMALPLVAREKDVVIEVVGELISRPYIHITLQLLERFGIRVRH-DEWQRF
              180          190          200          210          220          230

              240          250          260          270          280
2MEPSP  YIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxx-----xxxxxxxxxSLQGDVKFAE
      :  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:
Unipro  TIPAGSRYRSPGTIHVEADASSASYFIALGALAAPAPGQEPLRILGVGLDSIQGDIRFAE
              240          250          260          270          280          290

              290          300          310          320          330          340
2MEPSP  VLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTA
      .  .  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:
Unipro  AARAMGA EVTGGPNWLVVRR-----GAWPLRAVDLDCNHIPDAAMTLAVMALYAQGT TT
              300          310          320          330          340

              350          360          370          380          390          400
2MEPSP  IRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK---LNVTAIDTYDDHRM
      .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:
Unipro  LRNIASWRVKETDRIAAMAQGCRRLGATVEEGSDFLRVTPPASPADWRAASIHTYDDHRI
              350          360          370          380          390          400

              410          420          430          440
2MEPSP  AMAFSLAAC--AEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:  .:
Unipro  AMCFSLAAFNPALPVRIEDPKCVAKTFPDYFEALFSVAGTPVEHVPVICIDGPTASGKG
              410          420          430          440          450          460

Unipro  TVAAAVAQRLGYRFLDSGAMYRITALAALRAGLSIDA AHQDRIAALARTLPVRFESGRIW
              470          480          490          500          510          520

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Q2Y7A6_NITMU | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
|*Nitrosospora multiformis* (strain ATCC 25196 / NCIMB 11849). |AA|437
Length = 437
initn: 787 init1: 220 opt: 851 Z-score: 1678.8 bits: 319.8 E(): 7.1e-85
Smith-Waterman score: 851; 46.934% identity (71.934% similar) in 424 aa overlap
(21-437:1-412)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLP	GSKSLSNRILL	LLAALSEGTTVVDNLL	NSEDVHYMLGAL		
Unipro		MPGSKSISNRILL	LLAALSEGVT	DVCDLLASDD	TARMLDAL	
		10	20	30	40	

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVV	VGCGKFPVEDAKEEV	QLFLGNAGIAMRSL	xxxxxxxxxxxx		
Unipro	STLGVSI-LQIGRDHYRL	QGVGDQFPLRLPT	TEADLFLGNAGTV	FRPLTAMLALA	QGH--	
	50	60	70	80	90	

	130	140	150	160	170
2MEPSP	YVLDGVPRMRERPIGDL	VVGLKQLGADVDC	FLGTD-CPPVRVNG	IIGGLPGGKVK	LSGSIS
Unipro	YRLSGVPRMHERPIGDL	VDALRQVGADIT-	YLGKEGFPPLQIKP-	GRIHPGEITVR	GEVS
	100	110	120	130	150

	180	190	200	210	220	230
2MEPSP	SQYxxxxxxxxxxxx	GDVE----IEIIDKL	SIPYVEMTLRLMER	FGVKAHSDSWDR	FY	
Unipro	SQFLTALLMVLPLFL	RAEMDEL	PVITVAGELISRPY	IDLTIALMARFGV	QVER--E	EWRRFT
	160	170	180	190	200	210

	240	250	260	270	280	290
2MEPSP	IKGGQKYKSPKNAYVE	GDASSASYFLAGAAI	xxxxxxxxxxxx	SLQGDVKFAEV	LEMMG	
Unipro	VPADQRYRSPGQVF	VEGDASSASYFLA	AGAIRGPVRVEGL	GRDSVQGDIRFA	EALERM	G
	220	230	240	250	260	270

	300	310	320	330	340	350
2MEPSP	AKVTWTETSVTVTGPP	REPFRGKHLKAIDV	NMKNMPDVAMTL	AVVALFADGPTA	IRDVAS	
Unipro	ADIRFGDNWIEASGP	-----GPGGLRAI	DLDCNHIPDAAM	TAVTALFARGNT	VLNRNAS	
	280	290	300	310	320	

	360	370	380	390	400	410
2MEPSP	WRVKETERMVAIRTEL	TKLGASVEEGPDYCI	ITPP--EKLNVTAI	DYDDHRMAMAF	SLA	
Unipro	WRVKETDRIAAMAQEL	RKLGAEEVAGSDFL	QISPPRGELVANA	AIDTYDDHRMAM	CFSLV	
	330	340	350	360	370	380

	420	430	440		
2MEPSP	ACAEVPVTIRDPGCTR	KTFPDYFDVLSTF	VKN		
Unipro	SFG-APVRINDPRCV	SKTFPDYFEKFAAI	AYADPGQGKFAAR		
	390	400	410	420	430

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Q47GK0_DECAR | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 EC=2.7.4.14; | *Dechloromonas aromatica* (strain RCB). | AA|643
 Length = 643
 initn: 803 init1: 220 opt: 851 Z-score: 1676.3 bits: 319.9 E(): 9.8e-85
 Smith-Waterman score: 911; 48.268% identity (72.055% similar) in 433 aa overlap
 (5-437:2-418)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEIS	GTVKLP	SGSKSL	SNRILL	LAALSEG	TTVVDNLLNSEDVHYMLGAL
	:	:	:	:	:	:
Unipro	MEFLDLPQLLSAS	GTVRLP	SGSKSI	SNRVLL	LAALSGGETEVRD	LLGSDDTERMLEAL
	10	20	30	40	50	

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRA	VVVGCGGKFPVEDA	KEEVQLFLGNAGIAM	RSLxxxxxxxxxxxx		

Unipro	KALGVGVQ-HLGGEN	WSIKGCGGHFPV---	KQIELFLGNAGTAF	RPLTAALALAGGD--		
	60	70	80	90	100	110

	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIG	DLVVGLKQLGADV	DCFLGTDCPPVRV	NGIGGLPGGKVKLS	SGSISS	
	:::	:::	:::	:::	:::	:::
Unipro	YVLKGVARMHHERPI	GDLDVGLRQLGAD	ITYLGNEGFPLHL	KPASILPNGVVTVR	GDVSS	
	120	130	140	150	160	170

	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxx	GDVEIEIIDKLIS	IPYVEMTLRLMER	FGVKAHSDSWDR	IFYIKGGQ	

Unipro	QFLTGLLMALPLT	GETVVVEVLGEL	ISKPYIEITLAT	MARFGVIVER-EG	WQRFTVLAGS	
	180	190	200	210	220	

	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDAS	ASYFLAGAAIxxxx	xxxxxxxxxxxxx	SLQGDVKFAEVLE	MMGAKVTW	

Unipro	RYVSPGTIFVEGDAS	ASYFLALGAIGGP	VRVEGVGRDSIQ	GDVKFAEALVKMG	AQIEM	
	230	240	250	260	270	280

	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREP	FGRKHLKAIDVN	MNMKMPDVAMT	LAVVALFADGPTA	IRDVASWRVKE	

Unipro	GPNWMEARAP-----	QAGLVAVDLDCN	HIPDAAMTLAT	TALFAKGTTTLR	NIASWRVKE	
	290	300	310	320	330	340

	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKL	GASVEEGPDYCI	ITPPEKLNVT	AIDTYDDHRMAMA	FLAACA	AEVPV

Unipro	TDRIAAMATELRKL	GAVVDEGDDYI	RVTPAD-LKPAA	IDTYDDHRMAMC	FLA	AFG-TPL
	350	360	370	380	390	400

	430	440
2MEPSP	TIRDPGCTRKT	FPDYFDVLSTFVKN

Unipro	RINDPKCAKT	FPDYFERFAGVT
	410	420

	430	440	450	460
2MEPSP	TIRDPGCTRKT	FPDYFDVLSTFVKN		

Unipro	RINDPKCAKT	FPDYFERFAGVT	KAAPVIAIDGPS	ASGKGTVAARIAL
	410	420	430	440

Q6E6L7_NYMAL|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19; Flags: Fragment; |*Nymphaea alba* (White water-lily). |AA|264
Length = 264
initn: 838 initl: 757 opt: 848 Z-score: 1676.1 bits: 318.6 E(): 1e-84
Smith-Waterman score: 848; 65.098% identity (82.745% similar) in 255 aa overlap
(60-313:1-255)

330 340 350 360 370 380
2MEPSP KMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT

A9IJI7_BORPD|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 |*Bordetella petrii* (strain ATCC BAA-461 / DSM 12804 / CCUG 43448). |AA|447
 Length = 447
 initn: 755 initl: 230 opt: 847 Z-score: 1670.7 bits: 318.3 E(): 2e-84
 Smith-Waterman score: 847; 45.880% identity (69.710% similar) in 449 aa overlap
 (1-443:1-438)

10 20 30 40 50 60
 2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLALLALSEGTTVVDNLLNSEDVHYMLGAL
 : : : : :
 Unipro MTGQPYLDLPRARQARGQVALPGSKSISNRVLLLLAALAGRTDISGLLDSDDTRVMLAAL
 10 20 30 40 50 60

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGC GGFVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Unipro  RQLG--VELAEAGEGRVTVGGAGRFPVKQA---DLFLGNAGTAFRPLTAALALMGG--Q
              70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Unipro  YRLSGVPRMHERPIGDLVDALRQWGARIDYLGQAGYPPLAVGEGRIRADAPARVQGAVSS
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxxx----xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Unipro  QFLTALLLAAPVLAQGSDRPVVEVAGELISKPYIEITLNLMARYGVQVRR-DGWRTFTI
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Unipro  EPGAAYRSPGAIAVEGDASSASYFLALGAIGGGPVRVTGVGADSIQGDVAFARTLADMGV
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Unipro  QIDYGPDWIEARGVRVDQGG--LKAFTDFTNLIIPDAAMTAAALALYADGPCRLRNIGSW
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Unipro  RVKETDRIHAMHTELAKLGAEVESGPDWLRITPPADGGWRDAHIGTWDDHRMAMCFSLAA
              350          360          370          380          390          400

              420          430          440
2MEPSP  CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Unipro  FGPAAVRILDPGCVSKTFPDYFDVYAGLVSGAPDSYDD
              410          420          430          440

```

A2SI10_METPP | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 EC=2.7.4.14; |*Methylibium petroleiphilum* (strain PM1). |AA|674
 Length = 674
 initn: 847 init1: 268 opt: 831 Z-score: 1636.4 bits: 312.6 E(): 1.6e-82
 Smith-Waterman score: 831; 44.118% identity (71.267% similar) in 442 aa overlap
 (9-439:10-438)

```

              10          20          30          40          50
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGA
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Unipro  MATPRAFLDIPPLQAAGGTVRLPGSKSISNRVLLLAGLCAGRTRVLDLLDSDDTQVMLDA
              10          20          30          40          50          60

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Q21Y54_RHOFD | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 EC=2.7.4.14; | *Rhodoferrax ferrireducens* (strain DSM 15236 / ATCC BAA-621 / T118).
 |AA|669
 Length = 669
 initn: 533 initl: 239 opt: 802 Z-score: 1578.9 bits: 301.9 E(): 2.6e-79
 Smith-Waterman score: 802; 43.468% identity (72.072% similar) in 444 aa overlap
 (9-439:9-436)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLP	GSKSLSNRILL	LLAALSEGTTVVDNLL	NSDEVHYMLGAL		
	: : : : :	: : : : :	: : : : :	: : : : :	: : : : :
Unipro	MFSTAFLDIPPLTGVSGTVVLP	GSKSISNRVLL	SALSAGRTTLIDLL	SDDTRVMFEAL		
	10	20	30	40	50	60

	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVG	CGGKFPVEDAKEEVQLFLGN	AGIAMRSLxxxxxxxxxxxx			
	: : :	: : :	: : :	: : :	: : :	: : :
Unipro	RALGCGVR--QSGSTVEIEGL	GGQL---GHRQAALFMGN	AGTAIRPLTAALAVLGGD--			
	70	80	90	100	110	

	130	140	150	160	170	
2MEPSP	YVLDGVPRMRERPIGDLVVGL	KQLGADVDCFLGTD-CPPVR	VNGIGGLPGGKVKLSGSIS			
	. : : : : : : : . : : : .	. : : : . : : : .	. : : : . : : : .	. : : : . : : : .	
Unipro	FELRGVPRMHERPIGDLVDAL	RQIGCHID-YLGQDGFPLH	IGRPSLKLDTPIRVRGDVS			
	120	130	140	150	160	170

	180	190	200	210	220	230
2MEPSP	SQYxxxxxxxx--xxxxGDVEI	EIIDKLISIPYVEMTLRLMER	FGVKAHSDSWDRFYIK			
	: :	: :	: :	: :	: :	: :
Unipro	SQFLTALLMALPLVAATRDIV	EVVGEISRPYIEITLNL	LARFNVVRVVR-DGWQRFTIP			
	180	190	200	210	220	230

	240	250	260	270	280	290
2MEPSP	GGQKYKSPKNAYVEGDASSA	SYFLAGAAIxxxxxxxxxxxx	-----xxSLQGDVKFAE	VLE		
 : : : : : : : : : : : : : : : : : : : :
Unipro	AASQFRSPGTLFVEADASSA	SYFIAAGAIATGAGGQKGL	RIEGVGADSIQGDIFIEAAQ			
	240	250	260	270	280	290

	300	310	320	330	340	350
2MEPSP	MMGAKVTWTETSVTVTGPP	REPFGRKHLKAIDVNMNK	MPDVAMTLAVVALFADGPT	AIRD		
	: : : : .	. : : .	: : : : .	: : : : .	: : : : .	: : : : .
Unipro	MMGAQVQSGPNWLEVSR----	GCWPLKAIDLDCNHIPDA	AMTLAMMALYAQGSTTLRN			
	300	310	320	330	340	

	360	370	380	390	400	
2MEPSP	VASWRVKETERMVAIRTEL	TKLGASVEEGPDYCIITP	PEKL--NVT	AIDTYDDHRMAMA		
	. : : : :	: : : : : : : : . : : : .	. : : : . : : : .	. : : : . : : : .	
Unipro	IASWRVKETDRIAAMACEL	RKLGATVEEGADFIRVTP	PAVAADWRAASIHTYDD	HRMAMC		
	350	360	370	380	390	400

	410	420	430	440		
2MEPSP	FSLAAC--AEVPVTIRDPG	CTRKTFPDYFDVLSTFVKN				
	: : : : .	: : : . : : : .	: : : : .	: : : : .		
Unipro	FSLAAFNPAGLPVRIVDPK	CAKTFPDYFEALFSLAQ	TSQAIPVICVDGPTASG	KGTLA		
	410	420	430	440	450	460

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Unipro GAVAQALGYHLLDSGSLYRITALAALQAGLALEFAHEQAIADLIPGLSIRFSQDQVFLNG
470 480 490 500 510 520

B1DMY1_9RHOO | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
EC=2.7.4.14; | *Thauera* sp. MZ1T. | AA | 653
Length = 653
initn: 952 initl: 388 opt: 801 Z-score: 1577.1 bits: 301.5 E(): 3.3e-79
Smith-Waterman score: 967; 47.936% identity (74.312% similar) in 436 aa overlap
(5-440:2-425)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEIS	GTVKLP	SGSKSL	SNRILL	LAAL	SEGTTVVDNLLN
	: . : .	: . : .	: . : .	: . : .	: . : .	: . : .
Unipro	MEFLDLP	PMLGAAGSV	RPLPGSKS	SISNRV	LLLAAL	AEGETDIRD
	10	20	30	40	50	

	70	80	90	100	110	120
2MEPSP	RTLGLS	VEADKAAK	RAVVVG	CGGKFP	VEDAKE	EEVQLFLGN
	: . : .	: . : .	: . : .	: . : .	: . : .	: . : .
Unipro	RALGVD	WRREGDSL	NYRVCG	VGGPFP	VKTG---	DLFLGNAG
	60	70	80	90	100	110

	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMRER	PIGDLV	VGLKQL	GADVDC	FLGTDCPP
	: . : .	: . : .	: . : .	: . : .	: . : .	: . : .
Unipro	YRLSG	VPRMHER	PIGDLV	DALRQL	GADITCT	ANEGYPPL
	120	130	140	150	160	170

	190	200	210	220	230	240
2MEPSP	QYxxxxxxx	GDVEIEI	IDKLIS	IPYVEM	TLRLMER	FGVKAHSD
	: . : .	: . : .	: . : .	: . : .	: . : .	: . : .
Unipro	QFLTAL	LMLPLTG	VETTIE	VGELISK	PIRITL	ELMARFGV
	180	190	200	210	220	230

	250	260	270	280	290	300
2MEPSP	KYKSP	KNAYVE	GDASSA	SYFLAG	AAIxxxxxxx	SLQGDVK
	: . : .	: . : .	: . : .	: . : .	: . : .	: . : .
Unipro	RYRSP	GTVFVE	GDASSA	SYFLAAG	AIGGGP	VRVEGVGR
	240	250	260	270	280	290

	310	320	330	340	350	360
2MEPSP	TETSV	TVTGP	PREPF	GRKHLK	AIDVNM	NKMPDV
	: . : .	: . : .	: . : .	: . : .	: . : .	: . : .
Unipro	GDNW	IEAAAP	-----	AGGVLK	AFDLDL	NHIPDA
	300	310	320	330	340	

	370	380	390	400	410	420
2MEPSP	TERMV	AIRTEL	TKLGAS	VEEGPD	YCIITP	PPEKLN
	: . : .	: . : .	: . : .	: . : .	: . : .	: . : .
Unipro	TDRIA	AMATEL	RKLGA	VEEGAD	YLVVQR	PPRLQP
	350	360	370	380	390	400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFFPDYFDVLSTFVKN
        :.: :.: :.: :.: :.:
Unipro  RINDPKCVNKTFPGYFEAFAQVARPVPVLAIDGPSASGKGTVAARVAETLGWHYLDSGSL
        410          420          430          440          450          460

```

B3R368_CUPTR | 3-enolpyruvylshikimate-5-phosphate synthetase; EC=2.5.1.19;
| *Cupriavidus taiwanensis* (strain R1 / LMG 19424) | AA | 434
Length = 434
initn: 893 initl: 227 opt: 795 Z-score: 1567.9 bits: 299.2 E(): 1.1e-78
Smith-Waterman score: 954; 47.846% identity (75.283% similar) in 441 aa overlap
(5-440:2-430)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro   MEHLTLGPLTRATGTVRLPGSKSISNRVLLLAALANGETVRDRLDSDDTRVMLQAL
        10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
        :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro   RTLGVAWRQD--GPDYIVTGAGGNFPVKSA---ELFMGNAGTAIRPLTAALALQGG--S
        60          70          80          90          100

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro   YKLSGVPRMHERPIGDLVDGLRQVGAVIDYLGNEGFPPLHIQPASLRIDAPIRVRGDVSS
        110          120          130          140          150          160

```

```

                190          200          210          220          230
2MEPSP  QYxxxxxxxx--xxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
        :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro   QFLTALLMSLPLAQSASGRIEIEVVGELISKPYIEITLNLARFGIEVERQ-GWERFILP
        170          180          190          200          210          220

```

```

                240          250          260          270          280          290
2MEPSP  GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
        :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro   AGTAYRSPGEIFVEGDASSASYFLAAGAIGGGPVRVEGVGMASIQGDVRFADALNRMGAN
        230          240          250          260          270          280

```

```

                300          310          320          330          340          350
2MEPSP  VTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWR
        :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro   VMAGDNWIEVRGTERDD-GRLH--GIELDCNHIPDAAMTLAVAALFAEGTTTLTNIASWR
        290          300          310          320          330          340

```

```

                360          370          380          390          400          410
2MEPSP  VKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT--IDTYDDHRMAMAFSLAAC
        :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Unipro   VKETDRIAAMATELRKLGAVVEEGADYLRVTPPQPWQTPADGIGTYDDHRMAMCFSLAAF
        350          360          370          380          390          400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP AEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . .:: :.:::. :.:::.::..
Unipro GPLPVRINDPGCVAKTFPDYFSVFAGVTR
      410      420      430

```

A1KUN6_NEIMF | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 | *Neisseria meningitidis* serogroup C / serotype 2a (strain ATCC 700532 / FAM18).
 | AA|433
 Length = 433
 initn: 797 initl: 175 opt: 794 Z-score: 1565.9 bits: 298.9 E(): 1.4e-78
 Smith-Waterman score: 794; 43.891% identity (70.136% similar) in 442 aa overlap
 (6-444:3-430)

```

      10      20      30      40      50
2MEPSP MAGAEEIVLQPIKEIS-GTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGA
      : : . . . .: :.:::.::: :.:::.::: . .:::.::: :.:::.:::
Unipro  MTESVRLPVARLKPSTVALPGSKSISNRTLLLAALSDNACEIHSLLKSDDTDRMLEA
      10      20      30      40      50

```

```

      60      70      80      90     100     110
2MEPSP LRTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      : :.:::.::: : : : :.:::.::: : :.:::.::: :.:::.::: :.:::.:::
Unipro  LDKLGVQIEY-LAEDRLKVHGTGGRFPNRTA---DLFLGNAGTAFRPLTAALAVLGGD-
      60      70      80      90     100     110

```

```

      120     130     140     150     160     170
2MEPSP xYVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTD-CPPVRVNGIGGLPGGKVKLSGSI
      : : :.:::.::: :.:::.::: :.:::.::: :.:::.::: :.:::.::: :.:::.:::
Unipro  -YHLHGVPRMHERPIGDLADALRIAGADVE-YLGKEHYPLHIGKRQDNGERVIPIKGNV
      120     130     140     150     160

```

```

      180     190     200     210     220     230
2MEPSP SSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG
      :.:::.::: :.:::.::: :.:::.::: :.:::.::: :.:::.::: :.:::.:::
Unipro  SSQFLTALLMALPLTGQAFEIRMVGEISKPYIDITLKLMAQFGVQVAN-ENYRVFKIPA
      170     180     190     200     210     220

```

```

      240     250     260     270     280     290
2MEPSP GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
      : :.:::.::: :.:::.::: :.:::.::: :.:::.::: :.:::.::: :.:::.:::
Unipro  DAHYHAPEHLHVEGDASSASYFLAAGLIAATPVRVTGIGANSIQGDVAFARELEKIGADV
      230     240     250     260     270     280

```

```

      300     310     320     330     340     350
2MEPSP TWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      .: :.:::.::: :.:::.::: :.:::.::: :.:::.::: :.:::.::: :.:::.:::
Unipro  VWGENFVEVSRPKE-----RAVQSFDLDANHIPDAAMTLAIVALATGQTCTLRNIGSWRV
      290     300     310     320     330     340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      360      370      380      390      400      410
2MEPSP KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAACAE
      ::::: .:: :::: : . :::: . : ::::
Unipro KETDRIAAMANELRKLGAKEVEEAEAIHITPPKTLTPDAVIDTYDDHRMAMCFSLVSLLG
      350      360      370      380      390      400

      420      430      440
2MEPSP VPVTIRDPGCTRKTFPDYFDVLSTFVK
      :::.::: ::::: :::.:::
Unipro VPVINDPKCTHKTFPTYFDVFSSLTETTE
      410      420      430

```

B4RL93_NEIG2 | 3-phosphoshikimate 1-carboxyvinyltransferase; | *Neisseria gonorrhoeae* (strain NCCP11945). | AA | 433
Length = 433
initn: 788 initl: 182 opt: 790 Z-score: 1558.0 bits: 297.4 E(): 3.8e-78
Smith-Waterman score: 790; 44.626% identity (69.860% similar) in 428 aa overlap (18-444:16-430)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGKSLNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :: :::: :::: :::: :::: ::::
Unipro MTESVRLPAASLKPSTVALPGSKSISNRTLLLAALSDNVCEIHSLLKSDDTDRMLEAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      ::::: : : : : :::: . :::: :::: ::::
Unipro DKLGVQIE-HLAEGRLKVHGTGGRFPNRS- ---DLFLGNAGTAFRPLTAALAVLGG--G
      60      70      80      90      100      110

```

```

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : :::: :::: :::: :::: ::::
Unipro YHLHGVP RMHERPIGDLVDALRIAGADVEYLGNEHYPLHIGKRQDCGERVIPIKGNVSS
      120      130      140      150      160      170

```

```

      190      200      210      220      230      240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . :::: :::: :::: :::: :::: ::::
Unipro QFLTALLMALPLTGQAFEIRMV GELISKPYIDITLKLMAQFGVQVAN-EGYRVFKIPADA
      180      190      200      210      220      230

```

```

      250      260      270      280      290      300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . . . :::: :::: :::: :::: :::: ::::
Unipro HYHAPEHLHVEGDASGASYFLAAGLIAATPVRVTGIGANSIQGDVAFARELEKIGADVW
      240      250      260      270      280      290

```

```

      310      320      330      340      350      360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      : : : : : : : : : : : : : : : :
Unipro GENFVEVSRPK----GRA-VQAFDL DANHIPDAAMTLAIVALATRQTCTLRNIGSWRVKE
      300      310      320      330      340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                370          380          390          400          410
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAACAEVP
      :. . . . . :. : : : : : : . : : : : : : : : : : : : : : : : : : :
Unipro TDRIAAMANELRKLGAKVVEEAIAIHITPPETPTPDAVIDTYDDHRMAMCFSLLSLGVP
      350          360          370          380          390          400

          420          430          440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      :. . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VVINDPKCTHKTFPTYFEVFSSLTETAE
      410          420          430

```

Q5F889_NEIG1 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 | *Neisseria gonorrhoeae* (strain ATCC 700825 / FA 1090). | AA|433
 Length = 433
 initn: 784 initl: 182 opt: 787 Z-score: 1552.0 bits: 296.3 E(): 8.2e-78
 Smith-Waterman score: 787; 43.764% identity (69.161% similar) in 441 aa overlap
 (5-444:3-430)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :. . : . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MTESVRLPAASPKPSTVALPGSKSISNRTLLLAALSDNVCEIHSLLKSDDTDRMLEAL
      10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      :. . . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro DKLGVQIE-HLAEGRLKVHGTGGRFPNRS- ---DLFLGNAGTAFRPLTAALAVLGGD--
      60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP YVLGVPVPMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro YHLHGVPVPMRERPIGDLVDALRIAGADVEYLGNEHYPLHIGKRQDCGERVIPIKGNVSS
      120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . : : : : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMALPLTGQAFEIRMVGELISKPYIDITLKLMAQFGVQVAN-EGYRVFKIPADA
      180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : :. . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro HYHAPEHLHVEGDASGASYFLAAGLIAATPVRVTGIGANSIQGDVAFARELEKIGADVWV
      240          250          260          270          280          290

```

```

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      :. : : . : : : : : : : : : : : : : : : : : : : : : :
Unipro GENFVEVS---RPKGRA-VQAFDLDANHIPDAAMTLAIVALATRQTCTLRNIGSWRVKE
      300          310          320          330          340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                370          380          390          400          410
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAACAEVP
      : . . . . . : : : : : : : . : : : : : : : : : : : : : : : : : : :
Unipro TGRIAAMANELRKLGAKVVEEAIAIHITPPETPTPDAVIDTYDDHRMAMCFSLLSLLGVP
                350          360          370          380          390          400

```

```

                420          430          440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      : . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VVINDPKCTHKTFPTYFEVFSSSLTETAE
                410          420          430

```

B1JXR9_BURCC | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Burkholderia cenocepacia* (strain MC0-3). |AA|434
Length = 434
initn: 940 initl: 252 opt: 786 Z-score: 1550.0 bits: 295.9 E(): 1.1e-77
Smith-Waterman score: 941; 48.190% identity (73.077% similar) in 442 aa overlap
(9-444:6-433)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro MDYLDLGPYSSASGTVRLPGSKSISNRVLLLAALAEGETTITNLLDSDTRVMDAL
                10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro GKLGVKLARD--GDTCVVTGTRGAFTAKTA---DLFLGNAGTAVRPLTAALAVNGGD--
                60          70          80          90          100

```

```

                130          140          150          160          170          180
2MEPSP YVLGCVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro YRVHGVPRMHERPIGDLVDGLRQIGAQIDYELNEGYPLRIKPATISVDAPIRVRGDVSS
                110          120          130          140          150          160

```

```

                190          200          210          220          230
2MEPSP QYxxxxxxxxx---xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
      : . . . . : : . . : : : : : : : : : : : : : : : : : : : :
Unipro QFLTALLMTLPLVKAKDGRTVVEVDGELISKPYIDITIRLMARFGVTVER-DGWQRFVVP
                170          180          190          200          210          220

```

```

                240          250          260          270          280          290
2MEPSP GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      . : : : : . : : : : : : : : : : : : : : : : : : : : : :
Unipro AGVRYKSPGRIMVEGDASSASYFLAAGALGGGPLRVEGVGRASIQQGDVGFANALMQMGAN
                230          240          250          260          270          280

```

```

                300          310          320          330          340          350
2MEPSP VTWTETSVTVTGPPREPFGGRKH--LK AIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVA
      : : . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro VTMGDDWIDVRG-----IGHDHGKLEPIDMDFNLIPDAAMTIAVAALFANGTSTLRNIAS
                290          300          310          320          330          340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      360      370      380      390      400      410
2MEPSP WRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAA
      :::::..... :::: :::: :::::..... :::::.....
Unipro WRVKETDRIAAMATELRKVGAIVEEGPDYLVVTPPEKLTNPAAIDTYDDHRMAMCFSLVS
      350      360      370      380      390      400

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . :::: :::: :::::..... :::::
Unipro LGGVPVRINDPKCVGKTFPDYFDRFAALAKA
      410      420      430

```

A2VRY9_9BURK | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Burkholderia cenocepacia* PC184. | AA | 479
Length = 479
initn: 940 initl: 252 opt: 786 Z-score: 1549.4 bits: 296.0 E(): 1.1e-77
Smith-Waterman score: 941; 48.190% identity (73.077% similar) in 442 aa overlap
(9-444:51-478)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALS
                                : : .. :::::.....
Unipro PRARHGRNAAARPLPNRSRNNRTIPMDYLDLGPYSSASGTVRLPGSKSISNRVLLLAALA
                                30      40      50      60      70      80

      40      50      60      70      80      90
2MEPSP EGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFL
      :: .. :::::..... :::: ::::: .. ::::: .. ::::: .. :::::
Unipro EGETTITNLLSDSDTRVMDALGKLGKVLARD--GDTCVVTGTRGAFTAKTA---DLFL
      90      100      110      120      130

      100      110      120      130      140      150
2MEPSP GNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPP
      :::: ::::: ..:: . : . :::::..... :::::..... ::::: .. :::::
Unipro GNAGTAVRPLTAALAVNGGD--YRVHGVPRMHERPIGDLVDGLRQIGAQIDYELNEGYP
      140      150      160      170      180      190

      160      170      180      190      200      210
2MEPSP VRVNGIGGLPGGKVKLSGSISSQYxxxxxxxx--xxxxGDVEIEIIDKLISIPYVEMTL
      .::.. . ... :::::..... . . : . .::: ..:::
Unipro LRIKPATISVDAPIRVRGDVSSQFLTALLMTLPLVKAKDGRTVVEVDGELISKPYIDITI
      200      210      220      230      240      250

      220      230      240      250      260      270
2MEPSP RLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
      :::: ::::: ..::: ..::: ..::: ..::: ..::: ..::: ..:::
Unipro RLMARFGVTVER-DGWQRFVVPAGVRYKSPGRIMVEGDASSASYFLAAGALGGGPLRVEG
      260      270      280      290      300      310

      280      290      300      310      320      330
2MEPSP xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGGRKH--LKAIDVNMNMKMPDV
      .:::..... ::::: ..:::..... . . : : .::: ..::: ..:::
Unipro VGRASIQGDVGAFANALMQMGANVTMGDDWIDVRG-----IGHDHGKLEPIDMDFNLIPDA
      320      330      340      350      360

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          340          350          360          370          380          390
2MEPSP  AMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKL
          ::::::::::: ..... :::: :::: ::::: :::::
Unipro  AMTIAVAALFANGTSTLRNIASWRVKETDRIAAMATELRKVGAIVEEGPDYLVVTPPEKL
          370          380          390          400          410          420

```

```

          400          410          420          430          440
2MEPSP  NVTA-IDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
          . .: :::::::::::..... . :::: :::: .: ::::: .....
Unipro  TPNAADITYDDHRMAMCFSLSLGGVPVRINDPKCVGKTFFPDYFDRFAALAKA
          430          440          450          460          470

```

B4EB42_BURCJ | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Burkholderia cepacia* (strain J2315 / LMG 16656) | AA | 434
Length = 434
initn: 931 initl: 246 opt: 777 Z-score: 1532.2 bits: 292.6 E(): 1e-76
Smith-Waterman score: 932; 47.738% identity (72.851% similar) in 442 aa overlap
(9-444:6-433)

```

          10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          : : .. :::::::::::.....: .. :::::..... ::::
Unipro  MDYLDLGPYSSASGTVRLPGSKSISNRVLLLAALAEGETTITNLLDSDTRVMDAL
          10          20          30          40          50

```

```

          70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
          ::... : . ::::: : : .. : ..... ::::: ....
Unipro  GKLGVKLARD--GDTCVVTGTRGAFTAKTA---DLFLGNAGTAVRPLTAALAVNGGD--
          60          70          80          90          100

```

```

          130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          : . :::::.....: :::::.....: .. :::::..... . ... :::::
Unipro  YRVHGVPRMHERPIGDLVDGLRQIGAQIDYELNEGYPLRIKPATISVDAPIRVRGDVSS
          110          120          130          140          150          160

```

```

          190          200          210          220          230
2MEPSP  QYxxxxxxxxx---xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
          :. . . . : . . . : :::: :::::..... ::::: .: ::::: .
Unipro  QFLTALLMTLPLVKAKDGRTVVEVDGELISKPYIDITIRLMARFGVTVER-DGWQRFVVP
          170          180          190          200          210          220

```

```

          240          250          260          270          280          290
2MEPSP  GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
          .: ::::: . : :::::::::::.....: ::::: ::::: :::::
Unipro  AGVRYRSPGRIMVEGDASSASYFLAAGALGGGPLRVEGVGRASIQQGDVGFANALMQMGAN
          230          240          250          260          270          280

```

```

          300          310          320          330          340          350
2MEPSP  VTWTETSVTVTGPPREPFGGRKH--LKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVAS
          :: . . : : .: : : .: ::::: :::::.....: :::::
Unipro  VTMGDDWIDVRG-----IGHDHGKLEPIDMDFNLIPDAAMTIAVAALFANGTSTLRNIAS
          290          300          310          320          330          340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      360      370      380      390      400      410
2MEPSP WRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAA
      ::::: . . . . . ::::: ::::: ::::: . . . . . ::::: :::::
Unipro WRVKETDRIAAMATELRKVGAIVEEGPDYLVVTPPAKLTPNAAIDTYDDHRMAMCFSLV
      350      360      370      380      390      400

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . ::::: . . . . . ::::: . . . . .
Unipro LGGVPVRINDPKCVGKTFPDYFDRFAALAKA
      410      420      430

```

Q1LQH1_RALME | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 | *Ralstonia metallidurans* (strain CH34 / ATCC 43123 / DSM 2839). | AA | 452
 Length = 452
 initn: 841 initl: 226 opt: 769 Z-score: 1516.1 bits: 289.7 E(): 8.2e-76
 Smith-Waterman score: 919; 46.966% identity (74.382% similar) in 445 aa overlap
 (5-440:17-448)

```

      10      20      30      40
2MEPSP      MAGAEEIVLQPIKEISGTVKLPKSKSLSNRILLLAALSEGTTVVDNLL
      ::::: . . . . . ::::: ::::: ::::: . . . . .
Unipro MPPARAARNGHAAENMEQLTLGLPLTRASGTVRLPGSKSISNRVLLLSALATGETVRDLL
      10      20      30      40      50      60

```

```

      50      60      70      80      90      100
2MEPSP NSEDVHYMLGALRTLGLSVEADKAAKRAVVVGC GGKFPVEDAKEEVQLFLGNAGIAMRSL
      . . . . . ::::: . . . . . . . . . . ::::: . . . . . :::::
Unipro DSDDTRVMLEALKVLGVAWR--REGNDCIVSGVGGNFPSKAA---DLFMGNAGTAIRPL
      70      80      90      100      110

```

```

      110      120      130      140      150      160
2MEPSP xxxxxxxxxxxxYVL DGVPRMRERPIGDLVVLKQLGADVDCFLGTD-CPPVRVNGIGGL
      . . . . . . . . . . ::::: ::::: ::::: ::::: :::::
Unipro TAALALQGG--TYKLSGVQRMHERPIGDLVDGLRQVGASVD-YLGTGPGYPPLQINPAQIR
      120      130      140      150      160      170

```

```

      170      180      190      200      210      220
2MEPSP PGGKVKLSGSISSQYxxxxxxx---xxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVK
      . . . . . ::::: . . . . . . . . . . ::::: ::::: :::::
Unipro IDAPIRVRGDVSSQFLTALLMTLPMAQAQSGRIEIEVVGELISKPYIEITLNLRSRFGIA
      180      190      200      210      220      230

```

```

      230      240      250      260      270      280
2MEPSP AEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGD
      . . . . . ::::: . . . . . . . . . . ::::: ::::: :::::
Unipro VERQ-GWERFILPAGARYRSPGEIYVEGDASTASYFLAAGAIGGGPVRVEGVGMASIQGD
      240      250      260      270      280      290

```

```

      290      300      310      320      330      340
2MEPSP VKFAEVLEMMGAKVTWTETSVTVTGPPREPFG RKHLKAIDVNMNKMPPDVAMTLAVVALFA
      . . . . . ::::: . . . . . . . . . . ::::: ::::: :::::
Unipro VRFADALNRMGANVMAGDNWIEVRGTERDD-GR--LNGIELDCNHIPDAAMTLAVAALFA
      300      310      320      330      340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      350      360      370      380      390
2MEPSP DGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN-----TAID
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro EGTTLTNIASWRVKETDRISAMATELRKLGAEVEEGADYLVKVPPTSPATWQTPDAGIG
      350      360      370      380      390      400

```

```

      400      410      420      430      440
2MEPSP TYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro TYDDHRMAMCFSLAAGPLPVKINDPGCVAKTFPEYFEVFAGVAR
      410      420      430      440      450

```

A2WBS2_9BURK | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Burkholderia dolosa* AU0158. | AA | 434
Length = 434
initn: 921 initl: 234 opt: 767 Z-score: 1512.4 bits: 289.0 E(): 1.3e-75
Smith-Waterman score: 926; 47.285% identity (73.077% similar) in 442 aa overlap
(9-444:6-433)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro MDYLDLGPYSSASGTVRLPGSKSISNRVLLLAALAEGLTTITNLLDSDTRVMLDAL
      10      20      30      40      50

```

```

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro AKLGVKLARDRT--CVVTGTRGAFTAKTA---DLFLGNAGTAVRPLTAALAVNGGD--
      60      70      80      90      100

```

```

      130      140      150      160      170      180
2MEPSP YVLGVPVPMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro YRVHGVPMRHERPIGDLVDGLRQIGARIDYEQNAGYPPLRIRPATIAVDASIRVRGDVSS
      110      120      130      140      150      160

```

```

      190      200      210      220      230
2MEPSP QYxxxxxxxxx---xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro QFLTALLMTLPLVKAKDGKIVVEVDGELISKPYIDITIRLMERFGVVVER-DGWQRFTVP
      170      180      190      200      210      220

```

```

      240      250      260      270      280      290
2MEPSP GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro AGVYRSPGRIMVEGDASSASYFLAAGALGGGPLRVEGVGRASIQQGDVGFANALMQMGAN
      230      240      250      260      270      280

```

```

      300      310      320      330      340      350
2MEPSP VTWTETSVTVTGPPREPFGGRKH--LKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVAS
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Unipro VTMGDDWIEVRG-----IGHDHGKLEPIDMDFNLIPDAAMTIAVAALFASGTSTLRNIAS
      290      300      310      320      330      340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      360      370      380      390      400      410
2MEPSP WRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAA
      :::::..... :: :.....: : ..::: :. .: :.....:
Unipro WRVKETDRIAAMATELRKVGATVEEGADYLVVTPPAALTPNAAIDTYDDHRMAMCFSLV
      350      360      370      380      390      400

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . ::: :.:: :. :.....: .....:
Unipro LGGVPVRINDPKCVGKTFPDYFDRFAALAKA
      410      420      430

```

B1FBZ3_9BURK | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Burkholderia ambifaria* IOP40-10. | AA | 434
Length = 434
initn: 911 initl: 241 opt: 766 Z-score: 1510.4 bits: 288.6 E(): 1.7e-75
Smith-Waterman score: 921; 47.511% identity (72.851% similar) in 442 aa overlap
(9-444:6-433)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : .. :.....: :.....: :.. :.....: : : :
Unipro MDYLDLGPYSSASGTVRLPGSKSISNRVLLLAALAEEDTTITNLLDSDTRVMDAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      :... : . :...: : : .. : :.....: : : :...
Unipro GKLGVKLARD--GDTCVVTGTRGAFTAKTA---DLFLGNAGTAVRPLTAALAINGGD--
      60      70      80      90      100

```

```

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : . :.....: :.....: . :.....: . ... :...
Unipro YRVHGVPRMHERPIGDLVDGLRQIGAQIDYEQNEGFPLRIKPGTIAVDAPIRVRGDVSS
      110      120      130      140      150      160

```

```

      190      200      210      220      230
2MEPSP QYxxxxxxxx---xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
      :. .. . . : . : : :.....: :. :.....:
Unipro QFLTALLMTLPLVKAKDGRTVVEIDGELISKPYIDITIRLMERFGVTVER-DGWQRFVVP
      170      180      190      200      210      220

```

```

      240      250      260      270      280      290
2MEPSP GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      .: :...: . :.....: :.....: :.....: :...:
Unipro AGVYRSPGRIMVEGDASSASYFLAAGALGGGPLRVEGVGRASIQQGDVGFANALMQMGAN
      230      240      250      260      270      280

```

```

      300      310      320      330      340      350
2MEPSP VTWTETSVTVTGPPREPFGGRKH--LKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVA
      :. . . : : : : : : : : : : : : : : : :
Unipro VSMGDNWIEVRG-----IGHDHGKLDPIDMDFNLIPDAAMTIAVVALFASGTSTLRNIAS
      290      300      310      320      330      340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      360      370      380      390      400      410
2MEPSP WRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAA
      :::::..... :::: :::: :::: :::: :::::.....
Unipro WRVKETDRIAAMATELRKVGAIVEEGADYLVVTPPQRLTPNAAIDTYDDHRMAMCFSLVS
      350      360      370      380      390      400

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . :::: :::: :::: :::::.....
Unipro LGGVPVRINDPKCVGKTFPDYFDRFAALAKA
      410      420      430

```

B1SYY8_9BURK | 3-phosphoshikimate 1-carboxyvinyltransferase; | *Burkholderia ambifaria* MEX-5. | AA | 434

Length = 434

initn: 911 initl: 236 opt: 765 Z-score: 1508.4 bits: 288.2 E(): 2.2e-75

Smith-Waterman score: 920; 47.285% identity (72.851% similar) in 442 aa overlap (9-444:6-433)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : .. :::::::::::::::::::::::::::: :::::..... ::::
Unipro MDYLDLGPYSSASGTVRLPGSKSISNRVLLLAALAEEDTTITNLLDSDDTRVMLDAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      ::... : . ::::: : : .. : . ::::: ::::: :::::
Unipro GKLGVKLARD--GDTCVVTGTRGAFTAKTA---DLFLGNAGTAVRPLTAALAINGGD--
      60      70      80      90      100

```

```

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : . :::::..... :::::..... . :::::.....
Unipro YRVHGVPRMHERPIGDLVDGLRQIGAQIDYEQNEGFPLRIKPAAIADVAPIRVRGDVSS
      110      120      130      140      150      160

```

```

      190      200      210      220      230
2MEPSP QYxxxxxxxxx---xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
      :. .. . . : . :::: :::::::::::::::::::: :::::.....
Unipro QFLTALLMTLPLVKAKDGRTVVEIDGELISKPYIDITIRLMERFGVTVER-DGWQRFVVP
      170      180      190      200      210      220

```

```

      240      250      260      270      280      290
2MEPSP GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      .: ::::: . ::::::::::::::::::::. ::::: ::::: :::::
Unipro AGVYRSPGRIMVEGDASSASYFLAAGALGGGPLRVEGVGRASIQQGDVGFANALMQMGAN
      230      240      250      260      270      280

```

```

      300      310      320      330      340      350
2MEPSP VTWTETSVTVTGPPREPFGGRKH--LKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVA
      :. . . : : .: : : ::::: :::::::::::::::::::: :::::.....
Unipro VSMGDDWIEVRG-----IGHDHGKLDPIDMDFNLIPDAAMTIAVAALFASGTSTLRNIAS
      290      300      310      320      330      340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      360      370      380      390      400      410
2MEPSP WRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAA
      :::::..... ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Unipro  WRVKETDRIAAMATELRKVGAIVEEGADYLVVTPPQQLTPNAAIDTYDDHRMAMCFSLVS
      350      360      370      380      390      400

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . ::::: ::::: ::::: ::::: ::::: :::::
Unipro  LGGVPVRINDPKCVGKTFPDYFDRFAALAKA
      410      420      430

```

A1K4D3_AZOSB | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Azoarcus sp.* (strain BH72). | AA | 653
Length = 653
initn: 880 initl: 359 opt: 765 Z-score: 1505.8 bits: 288.3 E(): 3.1e-75
Smith-Waterman score: 922; 47.368% identity (72.082% similar) in 437 aa overlap
(5-441:4-428)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : . : . : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro  MTMEFLDLPPMLAARGQVRLPGSKSISNRSLLLAALAEGETDIRDLLASDDVERMLEAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      ..::: . .... : : : : : : : : : : : : : : : : : : : : :
Unipro  QALGVRWSREEGTDNYRVHGVGGPFPVKKG---DLFLGNAGTAFRPLTAALALSGG--D
      60      70      80      90      100      110

```

```

      130      140      150      160      170      180
2MEPSP YVLGDVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Unipro  YRLSGVARMHERPIGDLVDALRQAGADIEYVGNEGFPPLHIRPATIRPGGVLVKVRGDVSS
      120      130      140      150      160      170

```

```

      190      200      210      220      230      240
2MEPSP QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . . : : : : : : : : : : : : : : : : : : :
Unipro  QFLTALLMALPLTGVETTIEVVGELISKPYIAITLDLMARFGVDVVRED-WQRFVTPGGA
      180      190      200      210      220      230

```

```

      250      260      270      280      290      300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      .::: : : : : : : : : : : : : : : : : : : : : :
Unipro  RYRSPGVLYVEGDASSASYFLAAGAIGGGPVRVEGVGRDSIQGDVRFADALAQLGAVITV
      240      250      260      270      280      290

```

```

      310      320      330      340      350      360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . . . : : : : : : : : : : : : : : : : : :
Unipro  GDNWIEAAAPAG---GR--LRAFDLDLNHIPDAAMTLAVAALFADGPTLRNIASWRVKE
      300      310      320      330      340

```

	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT	AIDTYDDHRMAMAFSLAACAEV	PV			
	: . : . : . : :	:	:	:	:	:
Unipro	TDRIAAMATELRKVGAEEVGADYLRI	MPPAVL	RPA	AIDTYDDHRMAMCFSLVSLGGCRV		
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVK	N				
	: . : :	:	:	:	:	:
Unipro	RINDPRCVNKTFPSYFEAFATVAAPVPVVAIDGPSASGKGTVGARVAAALGWHLDSGSL					
	410	420	430	440	450	460

	230	240	250	260	270	280
2MEPSP	GVKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSL					
				:	:	:
Unipro	MGPPILDPRRLKILLGDASSASYFLAGAAVTGGTVTVEGCGTSSL					
		10	20	30	40	
	290	300	310	320	330	340
2MEPSP	QGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVA					
	:	:	:	:	:	:
Unipro	QGDVKFAEVLEQTGAKVSWTENSVTVTGPPRDSSGRKHLRAIDVNMNEMPDAITLAVVA					
	50	60	70	80	90	100

```

                350          360          370          380          390          400
2MEPSP LFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIDTY
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro LYADGPTAIRDVASWRVKETERMIAICTELRKLGATVEEGPDYCVITPPEKLNVTSIDTY
      110          120          130          140          150          160

                410          420          430          440
2MEPSP DDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Unipro DDHRMAMAFSLAACADVPVTIKDPGCTKEKLP
      170          180          190

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

B2U886_RALPJ | 3-phosphoshikimate 1-carboxyvinyltransferase; | *Ralstonia pickettii*
(strain 12J). |AA|434
Length = 434
initn: 873 init1: 234 opt: 761 Z-score: 1500.5 bits: 286.8 E(): 6.1e-75
Smith-Waterman score: 927; 48.165% identity (72.018% similar) in 436 aa overlap
(5-436:2-426)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEIS	GTVKLP	SGSKSL	SNRILL	LAAL	SEGTTVVDNLLN
	:.	:.	:.	:.	:.	:.
Unipro	MEHLDVG	PLKTAR	GTIKLP	SGSKSI	SNRVLL	LAALAQGETVVR
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGLS	VEADKA	KRAVVV	GCGKFP	VEDAKE	EEVQLFLGN
	:.	:.	:.	:.	:.	:.
Unipro	GKLGVS	VEG-QGE	NAYRV	TGTGGR	FPNTSA	---DLFMGN
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDGV	PRMRER	PIGDLV	VGLKQL	GADVDC	FLGTDCPP
	:.	:.	:.	:.	:.	:.
Unipro	YTLHGV	PRMHER	PIGDLV	DGLRQV	GARIDY	TGNEGYPPL
	120	130	140	150	160	170
	190	200	210	220	230	
2MEPSP	QYxxxxxxx	--xxxx	GDVEIE	IIDKLIS	IPYVEM	TLRLMERF
	:.	:.	:.	:.	:.	:.
Unipro	QFLTALL	MALPLV	ESAGNV	TIEVVG	ELISKPY	IEITLNL
	180	190	200	210	220	
	240	250	260	270	280	290
2MEPSP	GQKYKSP	KNAYVE	GDASSA	SYFLAG	AAIxxxx	xxxxxxxxxxxx
	:.	:.	:.	:.	:.	:.
Unipro	GVAYTAP	GEIFVE	GDASSA	SYFLAA	GGGPPV	RVVEGVGM
	230	240	250	260	270	280
	300	310	320	330	340	350
2MEPSP	TWTETSV	TVTGP	PREPF	GRKHLK	AIDVNM	NKMPDVAM
	:.	:.	:.	:.	:.	:.
Unipro	MAGGNW	IEVRGA	ERDD-G	KLH--A	VELDCN	HIPDAAM
	290	300	310	320	330	340
	360	370	380	390	400	410
2MEPSP	KETERM	VAIRTE	LTKLGAS	VEEGPD	YCIITP	PPEKLN
	:.	:.	:.	:.	:.	:.
Unipro	KETDRL	SAMATE	LRKLGA	EVEEG	ADYIRV	TPPSQW
	350	360	370	380	390	400
	420	430	440			
2MEPSP	EVPVTIR	DPGCTR	KTFPDY	FDVLST	FTVKN	
	:.	:.	:.	:.	:.	
Unipro	PVPVRIN	DPRCVA	KTFPEY	FTAFGG	GITA	
	410	420	430			

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

A7CM60_BURPI | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
 | *Ralstonia pickettii* 12D. | AA|434
 Length = 434
 initn: 892 init1: 234 opt: 761 Z-score: 1500.5 bits: 286.8 E(): 6.1e-75
 Smith-Waterman score: 924; 48.402% identity (72.146% similar) in 438 aa overlap
 (5-436:2-426)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLP	SGSKSLSNRILL	LAALSEGTTVVDNLL	NSSE	VHYMLGAL	
	:.. . :.:	:.:.:.:.:.:.:.:.:.:	:.:.:.:.:.:.:.:.:.:	:.:.:.:.:.:.:.:.:	:.:.:.:.:.:.:.:	:.:.:
Unipro	MEHLDVGPLKTARGTIKLP	SGSKSISNRVLL	LAALAQGETVVRD	LLSD	DTRVMLEAL	
	10	20	30	40	50	

	70	80	90	100	110
2MEPSP	RTLGLSVEA--DKAAKRAV	VVGCGGKFPVEDAKEEV	QLFLGNAGIAMRSL	xxxxxxxxxx	
	:.:.:.:. . :.:	:.:.:.:. . :.:	:.:.:.:. . :.:	:.:.:.:. . :.:	:.:.:. .
Unipro	GKLGVSVEGLGDNAYR---	VTGTAGRFPNKSA---	DLFMGNAGTAIRPL	TAAALQGG	
	60	70	80	90	100

	120	130	140	150	160	170
2MEPSP	xxYVL	DGVPRMRERPIGDL	VVGLKQLGADV	DCFLGTDCPPVR	VNGIGGLPGGK	VKLSGSI
	:.:	:.:.:.:.:.:.:.:.:	:.:.:.:. .	:.:	:.:. .	:.:. .
Unipro	--YTLHG	VPRMHERPIGDL	VLDGLRQVGAR	IDYTGNEGYPPL	AIHAAPVKIDAP	IRVRGDV
	120	130	140	150	160	

	180	190	200	210	220	230
2MEPSP	SSQYxxxxxxxx--xxxx	GDVEIEIIDKL	ISIPYVEMTLRL	MERFGVKA	EHSDSWDRFYI	
	:.:. .	:.:. .	:.:. .	:.:.:.:.:.:.:.:.:	:.:. .	:.:
Unipro	SSQFLTALLMALPLVES	AGNVTIEVVGEL	ISKPYIEITLNL	MARFGVQVAR	-DGWASFTV	
	170	180	190	200	210	220

	240	250	260	270	280	290
2MEPSP	KGGQKYKSPKNAYVEGD	ASSASYFLAGAAI	xxxxxxxxxxxxxx	SLQGDVKFAE	VLEMMGA	
	:	:.:. .	:.:.:.:.:.:.:.:.:	:.:.:.:.:.:.:.:	:.:.:	
Unipro	PTGVAYKAPGEIFVEGD	ASSASYFLAAGAL	GGGPVRVEGV	GMSSIQGDVRF	ADALNRMGA	
	230	240	250	260	270	280

	300	310	320	330	340	350
2MEPSP	KVTWTETSVTVTGPP	REPFGRKHLKAID	VNMNKM	PDVAMTLAV	VALFADGPTAIR	DVASW
	:.	:. .	:.:	:.:	:.:.:.:.:.:.:.:.:	:.:. .
Unipro	NVMAGGNWIEVRGAER	DD-GKLH--A	VELDCNHIPDA	AMTLAVAAL	FADGTTTTLT	NIASW
	290	300	310	320	330	340

	360	370	380	390	400	410
2MEPSP	RVKET	RMVAIRTELTKL	GASVEEGPDYCI	ITPPEKLN	VT--IDTYDD	HRMAMAFSLAA
	:.:.:.:. .	:.:	:.:.:.:.:.:.:.:	:.:	:.:. .	:.:.:.:.:.:.:.:
Unipro	RVKET	DRLSAMATELR	KLGAEEVEEGADY	IRVTPPSQW	TPPAGGIDTY	DDHRMAMAFSLAA
	350	360	370	380	390	400

	420	430	440
2MEPSP	CAEVPVTIRDPG	CTRKTFPDY	FDVLSTFVKN
	:.:	:.:. .	:.:.:.:
Unipro	FGPVPVRINDPR	CAKTFPEY	FTAFGGITA
	410	420	430

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

B2JF04_BURP8 | 3-phosphoshikimate 1-carboxyvinyltransferase; | *Burkholderia phymatum* (strain DSM 17167 / STM815). | AA|434
Length = 434
initn: 934 init1: 236 opt: 760 Z-score: 1498.5 bits: 286.4 E(): 7.8e-75
Smith-Waterman score: 922; 47.368% identity (73.227% similar) in 437 aa overlap (5-437:2-426)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEI	SGTVK	LPGSK	SLSNR
	:	:	:	:	:	:
Unipro	MEHLD	LGPF	SRASG	TVRLP	GSKS	SISNR
	:	:	:	:	:	:
	10	20	30	40	50	

	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KAAKRA	VVVGCG	KFPVED	AKEEVQ
	:	:	:	:	:	:
Unipro	EKLGV	KFKRD	--GDT	CVVTG	TRGAL	PAARA
	:	:	:	:	:	:
	60	70	80	90	100	

	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIG	DLVVGL	KQLGAD	VDCFLG
	:	:	:	:	:	:
Unipro	YRIHG	VPRMH	ERPIG	DLVDGL	RQIGAK	IDYEEN
	:	:	:	:	:	:
	110	120	130	140	150	160

	190	200	210	220	230
2MEPSP	QYxxxx	xxxxx	GDVEI	EIIDKL	ISIPY
	:	:	:	:	:
Unipro	QFLTAL	LMTLPL	VKTES	GETIVE	VSSEL
	:	:	:	:	:
	170	180	190	200	210

	240	250	260	270	280	290
2MEPSP	GGQKY	KSPKN	AYVEGD	ASSASY	FLAGAA	Ixxxx
	:	:	:	:	:	:
Unipro	SGVRY	QSPGK	IMVEGD	ASSASY	FLAAG	ALGGG
	:	:	:	:	:	:
	230	240	250	260	270	280

	300	310	320	330	340	350
2MEPSP	VTWTET	SVTVTG	PPREP	FGRKHL	KAIDVN	MNMKMP
	:	:	:	:	:	:
Unipro	VTMGDD	WIEVR	GVGND	-HGK--	LDPID	MDFNLI
	:	:	:	:	:	:
	290	300	310	320	330	340

	360	370	380	390	400	410
2MEPSP	VKETER	MVAIR	TELTKL	GASVEE	GPDYCI	IITPPE
	:	:	:	:	:	:
Unipro	VKETDR	IAAMAT	ELSKVG	AKVQAG	EDFLV	VTTPPE
	:	:	:	:	:	:
	350	360	370	380	390	400

	420	430	440
2MEPSP	EVPVT	IRDPG	CTRKTF
	:	:	:
Unipro	GVPVR	INDPK	CVGKTF
	:	:	:
	410	420	430

Q62M25_BURMA|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
|Burkholderia mallei (Pseudomonas mallei). |AA|673
Length = 673
initn: 927 initl: 246 opt: 761 Z-score: 1497.7 bits: 286.9 E(): 8.7e-75
Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
(5-437:241-665)

400 410 420 430 440
 2MEPSP KLVNTA-IDTYDDHRMAMAFSLAACAEVPTIRDPGCTRKTFPDYFDVLSTFVKN
 . : . : :::::::::::::::::::: . : : : : : : : : : :
 Unipro QLTPNASIDTYDDHRMAMCFSLVSLGGVPVRINDPKCVGKTFPDYFDRFLALANA
 620 630 640 650 660 670

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

Q3JXP8_BURP1 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=1.3.1.12;
 EC=2.5.1.19; | *Burkholderia pseudomallei* (strain 1710b). | AA|714
 Length = 714
 initn: 927 initl: 246 opt: 761 Z-score: 1497.3 bits: 286.9 E(): 9.1e-75
 Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
 (5-437:282-706)

		10	20	30	
2MEPSP		MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILL			
		:.. : :... :... :... :... :... :			
Unipro	SRAARA	AWRERGA	KPAPAVR	SDTSGT	GSHMEHLDLGPFSHAQGT
	260	270	280	290	300 310

	40	50	60	70	80	90
2MEPSP	AALSEGTTVVDNLLNSEDEVHYMLGALRTLGLSVEADKAAKRAVVVGC					
	:.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :					
Unipro	AALAEGETTVTNLLD	SDDTRV	MLDALTKL	GVKLSRD	GGT--CVVGGT	TRGAFTAKTA---
	320	330	340	350	360	

	100	110	120	130	140	150
2MEPSP	QLFLGNAGIAMRSLXXXXXXXXXXXXYVL					
	:.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :					
Unipro	DLFLGNAGTAVRPL	TAALAVNGGE--	YRIHGV	PRMHERPI	GDLDV	GLRQIGARIDYEGNE
	370	380	390	400	410	420

	160	170	180	190	200	210
2MEPSP	DCPPVRVNGIGGLPGGKVKLSGSISSQYXXXXXXXX--XXXXGDVEIEI					
	:.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :					
Unipro	GFPPLRIRPATIS	VDAPIR	VRGDVSSQ	FLTALLMTL	PLVKAKD	GASVVEIDGELISKPYI
	430	440	450	460	470	480

	220	230	240	250	260	270
2MEPSP	EMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAI					
	:.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :					
Unipro	EITIKLMARFGVT	VER-DGWQR	FTVPAGV	RYRSPGT	IMVEGD	ASSASYFLAAGALGGGPL
	490	500	510	520	530	540

	280	290	300	310	320	330
2MEPSP	XXXXXXXXSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFG					
	:.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :					
Unipro	RVEGVGRASI	QGDVGF	ANALMQM	GANVTMG	DDWIEVR	GIGHD-HGR--LAPIDMDFNLIP
	550	560	570	580	590	

	340	350	360	370	380	390
2MEPSP	DVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPE					
	:.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :					
Unipro	DAAMTI	AVAALF	ADGPST	LRNIGS	WRVKET	DRIAAMATELRKVGATVEEGADYLVVTPPA
	600	610	620	630	640	650

	400	410	420	430	440
2MEPSP	KLNVT--IDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN				
	:.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :				
Unipro	QLTPNASIDTYDD	HRMAMCF	SLVSLGG	VPVRI	NDPKCVGKTFPDYFDRFLALANA
	660	670	680	690	700 710

A8E7Y6_BURPS|3-phosphoshikimate 1-carboxyvinyltransferase; EC=1.3.1.12;
EC=2.5.1.19; |*Burkholderia pseudomallei* 406e. |AA|740
Length = 740
initn: 927 initl: 246 opt: 761 Z-score: 1497.0 bits: 286.9 E(): 9.4e-75
Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
(5-437:308-732)

Page 252 of 526

A5JC12_BURMA|3-phosphoshikimate 1-carboxyvinyltransferase; EC=1.3.1.12;
EC=2.5.1.19; |Burkholderia mallei FMH. |AA|749
Length = 749
initn: 927 initl: 246 opt: 761 Z-score: 1497.0 bits: 286.9 E(): 9.5e-75
Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
(5-437:317-741)

Bayer CropScience

A3NXX1_BURPO | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=1.3.1.12;
EC=2.5.1.19; | *Burkholderia pseudomallei* (strain 1106a). | AA|749
Length = 749
initn: 927 initl: 246 opt: 761 Z-score: 1497.0 bits: 286.9 E(): 9.5e-75
Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
(5-437:317-741)

Bayer CropScience

A5XXG4_BURMA|3-phosphoshikimate 1-carboxyvinyltransferase; EC=1.3.1.12;
EC=2.5.1.19; |Burkholderia mallei JHU. |AA|749
Length = 749
initn: 927 initl: 246 opt: 761 Z-score: 1497.0 bits: 286.9 E(): 9.5e-75
Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
(5-437:317-741)

400 410 420 430 440
 2MEPSP KLVNVT-IDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
 . : . : : : : : : : : : : : . : : : . : : : : : : :
 Unipro QLTPNASIDTYDDHRMAMCFSLVSLGGVPVRINDPKCVGKTFPDYFDRLALANA
 700 710 720 730 740

A3NC32_BURP6 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=1.3.1.12;
EC=2.5.1.19; |*Burkholderia pseudomallei* (strain 668). |AA|749
Length = 749
initn: 927 initl: 246 opt: 761 Z-score: 1497.0 bits: 286.9 E(): 9.5e-75
Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
(5-437:317-741)

Page 256 of 526

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

A4LEC8_BURPS | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=1.3.1.12;
 EC=2.5.1.19; | *Burkholderia pseudomallei* 305. | AA|749
 Length = 749
 initn: 927 initl: 246 opt: 761 Z-score: 1497.0 bits: 286.9 E(): 9.5e-75
 Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
 (5-437:317-741)

		10	20	30	
2MEPSP		MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILL			
		:.. : :... :... :... :... :... :...			
Unipro	SRAARA	AWRER	GAKPA	PAVRSD	TSGTGSH
	290	300	310	320	330
					340
	40	50	60	70	80
2MEPSP	AALSEGTTVVDNLLNSEDEVHYMLGALRTLGLSVEADKAAKRAVVVGC				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	AALAEGETTVTNLLD	SDDTRV	MLDAL	TKLGV	KLSRD
	350	360	370	380	390
					400
	100	110	120	130	140
2MEPSP	QLFLGNAGIAMRSLXXXXXXXXXXXXXVLDGVPRMRERPIGDLVVGLKQLG				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	DLFLGNAGTAVRPL	TAALAV	NGGE--	YRIHG	VPRMH
	410	420	430	440	450
	160	170	180	190	200
2MEPSP	DCPPVRVNGIGGLPGGKVKLSGSISSQYXXXXXXXX--XXXXGDVEIEI				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	GFPPLRIRPATIS	VDAPIR	VRGDV	SSQFL	TALLMT
	460	470	480	490	500
					510
	220	230	240	250	260
2MEPSP	EMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASY				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	EITIKLMARFGV	TVTVER	-DGWQR	FTVPAG	VRYRSP
	520	530	540	550	560
					570
	280	290	300	310	320
2MEPSP	XXXXXXXXSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFG				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	RVEGVGRASI	QGDVGF	ANALMQ	MGANVT	MGDDW
	580	590	600	610	620
					630
	340	350	360	370	380
2MEPSP	DVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGAS				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	DAAMTI	AAALF	ADGPT	SLRNIG	SWRVK
	640	650	660	670	680
					690
	400	410	420	430	440
2MEPSP	KLNVT- IDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDY				
	:.. : :... :... :... :... :... :... :... :...				
Unipro	QLTPNASIDTY	DDHRM	AMCFS	LVSLG	GVPRIN
	700	710	720	730	740

A9K5G6_BURMA|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
|*Burkholderia mallei* ATCC 10399. |AA|749
Length = 749
initn: 927 initl: 246 opt: 761 Z-score: 1497.0 bits: 286.9 E(): 9.5e-75
Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
(5-437:317-741)

Page 258 of 526

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

A2S4S0_BURM9 | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
| *Burkholderia mallei* (strain NCTC 10229). | AA|749
Length = 749
initn: 927 initl: 246 opt: 761 Z-score: 1497.0 bits: 286.9 E(): 9.5e-75
Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
(5-437:317-741)

			10	20	30
2MEPSP			MAGAEIIVLQPIKEISGTVKLP	GSKSLSNRILL	
			:.. : :... .	:...:.....:.....	
Unipro	SRAARA	AWRERGA	KPAPAVRSD	TSGTGSHMEH	LDLGPFSHAQ
	290	300	310	320	330 340

	40	50	60	70	80	90
2MEPSP	AALSEGTTVDNLLN	SEDEVHYMLGALRT	LGLSVEADKAAKRA	VVVGCGKFPVEDAKEE	V	
	:::..: .:	:...:..: ::	:...: : ..	:...: : : : ..	:	
Unipro	AALAEGETTVTNLL	SDDTRVMLDALTK	LGVKLSRDGGT--	CVVGGTRGAFTAKTA---		
	350	360	370	380	390 400	

	100	110	120	130	140	150
2MEPSP	QLFLGNAGIAMRSL	xxxxxxxxxxxxxY	VL	DGVPRMRERPIG	DLVVGLKQLGADV	DCFLGT
	:...:..: .:	:...: .	:	:...:.....: :...: .	:	
Unipro	DLFLGNAGTAVRPL	TAALAVNGGE--	YRIHGVP	RMHERPIGDLVD	GLRQIGARIDYEGNE	
	410	420	430	440	450	

	160	170	180	190	200	210
2MEPSP	DCPPVRVNGIGGLP	GGKVKLSGSISSQY	xxxxxxxxxx--	xxxxGDVEIEI	IDKLISIPYV	
	:...: .	:...: :...: .	:	: : :...: :...	:	
Unipro	GFPPLRIRPATISV	DAPIRVRGDVSSQ	FLTALLMTLPLV	KAKDGASVVEIDG	ELISKPYI	
	460	470	480	490	500 510	

	220	230	240	250	260	270
2MEPSP	EMTLRLMERFGVKA	EHSDSWDRFYIKG	GQKYKSPKNAYVE	GDASSASYFLAGAA	Ixxxxx	
	:...:~: :...: .	:...: :...: .	:...: :...: .	:...:.....: :...: .	:	
Unipro	EITIKLMARFGVT	VER-DGWQRFTVP	AGVRYRSPGTIM	VEGDASSASYFLA	AGALGGGPL	
	520	530	540	550	560 570	

	280	290	300	310	320	330
2MEPSP	xxxxxxxxxSLQGD	VKFAEVLEMMGAK	VTWTETSVTVTG	PPREPFGRKHLKA	IDVNMNMKMP	
	:...:~: :...: .	:...: :...: .	: : : : ..	: : :...: .	:	
Unipro	RVEGVGRASIQGD	VGFANALMQMGAN	VTMGDDWIEVRG	IGHD-HGR--	LAPIDMDFNLIP	
	580	590	600	610	620 630	

	340	350	360	370	380	390
2MEPSP	DVAMTLAVVALFAD	GPTAIRDVASWRV	KETERMVAIRTEL	TKLGASVEEGPDY	CIITPPE	
	:...:~: :...: .	:...: :...: .	:...: :...: .	:...: :...: .	:...: .	
Unipro	DAAMTIAVAALFAD	GPSTLRNIGSWRV	KETDRIAAMATEL	RKVGATVEEGADY	LVVTPPA	
	640	650	660	670	680 690	

	400	410	420	430	440
2MEPSP	KLNVT	A-IDTYDDHRMAMA	FSLAACAEVPVT	IRDPGCTRKTFPDY	FDVLSTFVKN
	:. .:	:...:~: :...: .	: : : : ..	:...: .	:...: .
Unipro	QLTPNASIDTYDD	HRMAMCFSLVSL	GGVPRINDPKCVG	KTFPDYFDRFLA	LANA
	700	710	720	730	740

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

A5TQL4_BURMA | 3-phosphoshikimate 1-carboxyvinyltransferase; EC=1.3.1.12;
 EC=2.5.1.19; | *Burkholderia mallei* 2002721280. | AA|749
 Length = 749
 initn: 927 initl: 246 opt: 761 Z-score: 1497.0 bits: 286.9 E(): 9.5e-75
 Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
 (5-437:317-741)

		10	20	30	
2MEPSP		MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILL			
		:.. : :... :... :... :... :... :...			
Unipro	SRAARA	AWRER	GAKPA	PAVRSD	TSGTGSH
	290	300	310	320	330
					340
	40	50	60	70	80
2MEPSP	AALSEGTTVVDNLLNSEDEVHYMLGALRTLGLSVEADKAAKRAVVVGC				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	AALAEGETTVTNLLD	SDDTRV	MLDAL	TKLGV	KLSRDGGT
	350	360	370	380	390
					400
	100	110	120	130	140
2MEPSP	QLFLGNAGIAMRSLXXXXXXXXXXXXXVYLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	DLFLGNAGTAVRPL	TAALAV	NGGE--	YRIHG	VPRMHER
	410	420	430	440	450
	160	170	180	190	200
2MEPSP	DCPPVRVNGIGGLPGGKVKLSGSISSQYXXXXXXXX--XXXXGDVEIEIIDKLISIPYV				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	GFPPLRIRPATIS	VDAPIR	VRGDV	SSQFLT	ALLMTL
	460	470	480	490	500
					510
	220	230	240	250	260
2MEPSP	EMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxx				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	EITIKLMARFGVT	VER-DG	WQRFT	VPAGV	RYRSPG
	520	530	540	550	560
					570
	280	290	300	310	320
2MEPSP	xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMKMP				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	RVEGVGRASI	QGDVGF	ANALM	QMGAN	VTMGDD
	580	590	600	610	620
					630
	340	350	360	370	380
2MEPSP	DVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPE				
	:... :... :... :... :... :... :... :... :... :...				
Unipro	DAAMTI	AAALF	ADGPT	SLRNIG	SWRVK
	640	650	660	670	680
					690
	400	410	420	430	440
2MEPSP	KLNVT--IDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN				
	:.. : :... :... :... :... :... :... :... :...				
Unipro	QLTPNASIDTY	DDHRM	AMCF	SLVSL	GGVP
	700	710	720	730	740

A3MHN8_BURM7|3-phosphoshikimate 1-carboxyvinyltransferase; EC=2.5.1.19;
|*Burkholderia mallei* (strain NCTC 10247). |AA|749
Length = 749
initn: 927 initl: 246 opt: 761 Z-score: 1497.0 bits: 286.9 E(): 9.5e-75
Smith-Waterman score: 919; 48.055% identity (72.769% similar) in 437 aa overlap
(5-437:317-741)

Page 261 of 526

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

APPENDIX D – DETAILED FASTA SEQUENCE SIMILARITY ALIGNMENTS OF THE 2MEPSPS PROTEIN
WITH PROTEINS FROM THE PDB DATABASE

2AA9_A | EPSP SYNTHASE LIGANDED WITH SHIKIMATE | *ESCHERICHIA COLI* | AA | 427
Length = 427
initn: 954 initl: 398 opt: 573 Z-score: 1145.6 bits: 221.1 E(): 6.3e-57
Smith-Waterman score: 979; 49.312% identity (73.624% similar) in 436 aa overlap
(5-440:2-421)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEIS	GTVKLP	GSKSL	SNRILL

PDB: 2A	MESLT	LQPIAR	VDGTIN	LPGSKS	SVSNR	ALLAAL
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEADK	AAKRAV	VVGCGG	KFPVED	AKEEVQ

PDB: 2A	TALGV	SYTSL	ADRT	RCEI	IGNGG	PLHAEG
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIG	DLVVGL	KQLGAD	VDCFLG

PDB: 2A	-VLTG	EPRMK	ERPIG	HLVDAL	RLGGAK	ITYLEQ
	120	130	140	150	160	170
	190	200	210	220	230	240
2MEPSP	QYxxxx	xxxxxxx	GDVEI	EIIDKL	ISIPY	EMTLRL

PDB: 2A	QFLT	ALLMT	APLAP	EDTVIR	IKGDLV	SKPYID
	180	190	200	210	220	
	250	260	270	280	290	300
2MEPSP	KYKSP	KNAYV	EGDASS	ASYFLA	GAIIxxxx	xxxxxxxx

PDB: 2A	SYQSP	GTYLVE	GDASS	ASYFLA	AAAIK	GGTVK
	230	240	250	260	270	280
	310	320	330	340	350	360
2MEPSP	TETSV	TVTGP	PREPF	GRKHL	KAIDVN	MNMKMP

PDB: 2A	GDDY	ISCT	-----	RGEL	NAIDM	DMNHIP
	290	300	310	320	330	340
	370	380	390	400	410	420
2MEPSP	TERMV	AIRTEL	TKLGAS	VEEGPD	YCIITP	PPEKLN

PDB: 2A	TDRLF	FAMATE	LRKVGA	EEVEEG	HDYIRI	TPPEKL
	350	360	370	380	390	400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

430 440
2MEPSP TIRDPGCTRKTFFPDYFDVLSTFVKN
:: :: :: ::::: . .
PDB:2A TILDPKCTAKTFFPDYFEQLARISQAA
410 420

1G6S_A | STRUCTURE OF EPSP SYNTHASE LIGANDED WITH SHIKIMATE-3- PHOSPHATE AND GLYPHOSATE | *ESCHERICHIA COLI* | AA | 427
Length = 427
initn: 954 initl: 398 opt: 573 Z-score: 1145.6 bits: 221.1 E(): 6.3e-57
Smith-Waterman score: 979; 49.312% identity (73.624% similar) in 436 aa overlap (5-440:2-421)

10 20 30 40 50 60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
: : : : : :
PDB:1G MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVRLHMLNAL
10 20 30 40 50

70 80 90 100 110 120
2MEPSP RTLGLSVEADKAAKRAVVVGC GGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
. : : : : :
PDB:1G TALGVSYTSLADTRCEIIGNGGPLHAEGALE---LFLGNAGTAMRPLAAALCLGSNDI-
60 70 80 90 100 110

130 140 150 160 170 180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
: : : : : . . : . . : : :
PDB:1G -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
120 130 140 150 160 170

190 200 210 220 230 240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
: : : : : :
PDB:1G QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVVKGQ
180 190 200 210 220

250 260 270 280 290 300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
. : : : : :
PDB:1G SYQSPGYTLVEGDASSASYFLAAAIAKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
230 240 250 260 270 280

310 320 330 340 350 360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
. . . : : : : : :
PDB:1G GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE
290 300 310 320 330 340

370 380 390 400 410 420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVPV
: . . : . : : : : :
PDB:1G TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIIATYNDHRMAMCFSLVALSDTPV
350 360 370 380 390 400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        :: :: :: :::::::::: ::
PDB:1G  TILDPKCTAKTFPDYFEQLARISQAA
                410          420

2QFT_A|E. COLI  EPSP  SYNTHASE  PRO101SER  LIGANDED  WITH  S3P  AND
GLYPHOSATE|ESCHERICHIA COLI|AA|427
Length = 427
initn: 954 initl: 398 opt: 573  Z-score: 1145.6  bits: 221.1 E(): 6.3e-57
Smith-Waterman score: 984;  49.541% identity (73.853% similar) in 436 aa overlap
(5-440:2-421)

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        :.....: .....: .....: .....: .....: .....:
PDB:2Q   MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVHRMLNAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
        :...:  .  :.  ..:  ::  .  .:  :  :  :  :  :  :  :  :  :  :  :  :  :
PDB:2Q   TALGVSYTLSADTRCEIIGNGGPLHAEGALE---LFLGNAGTAMRSLAAALCLGSNDI-
                60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        ::  :  :::::  ::  :.  ::  .  .  .  :  :  :  :  :  :  :  :  :  :
PDB:2Q   -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
                120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :.  ..  .  .  .  :.  :.  :.  :  :  :  :  :  :  :  :  :  :  :  :
PDB:2Q   QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVKGGQ
                180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        :...:  .  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
PDB:2Q   SYQSPGYTLVEGDASSASYFLAAAIAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
                230          240          250          260          270          280

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
        .  ..  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
PDB:2Q   GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE
                290          300          310          320          330          340

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADITYDDHRMAMAFSLAACAEVPV
        :..  .  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
PDB:2Q   TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
                350          360          370          380          390          400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      :: :: :: ::::: . :.
PDB:2Q TILDPKCTAKTFPDYFEQLARISQAA
                410          420

```

2QFQ_A|*E. COLI* EPSP SYNTHASE PRO101LEU LIGANDED WITH S3P|*ESCHERICHIA COLI*|AA|427
Length = 427
initn: 954 initl: 398 opt: 573 Z-score: 1145.6 bits: 221.1 E(): 6.3e-57
Smith-Waterman score: 978; 49.312% identity (73.624% similar) in 436 aa overlap
(5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :...::: . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:2Q MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVRLHMLNAL
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      .::: . . . . . : . . . : : : : : : : : : : : : : : : : : : .
PDB:2Q TALGVSYTLSADTRCEIIGNGGPLHAEGALE---LFLGNAGTAMRLLAALCLGSNDI-
                60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:2Q -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
                120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . . : . : : : : : : : : : : : : : : : : : : : : : : :
PDB:2Q QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEIENQH-YQQFVVKGGQ
                180          190          200          210          220

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      . : : . : : : : : : : : : : . . : : : : : : : : : : : : : : : :
PDB:2Q SYQSPGYTLVEGDASSASYFLAAAIAKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
                230          240          250          260          270          280

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPEPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:2Q GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE
                290          300          310          320          330          340

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVPV
      : . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:2Q TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
                350          360          370          380          390          400

```

```

      430      440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      :: :: :: :..... . .
PDB:2Q TILDPKCTAKTFPDYFEQLARISQAA
      410      420

2AA_Y_A|EPSP SYNTHASE LIGANDED WITH SHIKIMATE AND GLYPHOSATE|ESCHERICHIA
COLI|AA|427
Length = 427
initn: 954 initl: 398 opt: 573 Z-score: 1145.6 bits: 221.1 E(): 6.3e-57
Smith-Waterman score: 979; 49.312% identity (73.624% similar) in 436 aa overlap
(5-440:2-421)

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :...::: . . . . . :...::: :...::: :...::: :...::: :...:::
PDB:2A MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTLNLLSDDDRHMLNAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCCKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      .:::. . . . . :. . . :. . . :. . . :. . . :. . . :. . .
PDB:2A TALGVSYTLSADTRCEIIGNGGPLHAEGALE--LFLGNAGTAMRPLAALCLGSNDI-
      60      70      80      90      100      110

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :: : :...::: :. . . :. . . :. . . :. . . :. . . :. . .
PDB:2A -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
      120      130      140      150      160      170

      190      200      210      220      230      240
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . . :. . . :. . . :. . . :. . . :. . . :. . . :. . .
PDB:2A QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEIENQH-YQQFVVKGGQ
      180      190      200      210      220

      250      260      270      280      290      300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      .:::. . :...::: :. . :. . . :. . . :. . . :. . . :. . .
PDB:2A SYQSPGYTLVEGDASSASYFLAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
      230      240      250      260      270      280

      310      320      330      340      350      360
2MEPSP TETSVTVTGPPREPFGGRKHLKAIDVNMNMKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . . : : :...::: :...::: :...::: :...::: :...::: :...:::
PDB:2A GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTLRLNIYNWRVKE
      290      300      310      320      330      340

      370      380      390      400      410      420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVPV
      :. . . :. . . :. . . :. . . :. . . :. . . :. . . :. . . :. . .
PDB:2A TDRLFAMATELRKVGAEVEEGHDYIRITPPEKLNFAEIIATYNDHRMAMCFSLVALSDTPV
      350      360      370      380      390      400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          :: :: :: ::::: . .
PDB:2A  TILDPKCTAKTFPDYFEQLARISQAA
                410          420

```

1X8R_A|EPSPS LIGANDED WITH THE (S)-PHOSPHONATE ANALOG OF THE TETRAHEDRAL REACTION INTERMEDIATE|*ESCHERICHIA COLI*|AA|427

Length = 427

initn: 954 initl: 398 opt: 573 Z-score: 1145.6 bits: 221.1 E(): 6.3e-57

Smith-Waterman score: 979; 49.312% identity (73.624% similar) in 436 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          ::::: . . . . . ::::: . . . . . ::::: . . . . . ::::: . . . . .
PDB:1X   MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVHRMLNAL
                10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
          :::: . . . . . ::::: . . . . . ::::: . . . . . ::::: . . . . .
PDB:1X  TALGVSYTLSADTRCEIIGNGGPLHAEGALE---LFLGNAGTAMRPLAAALCLGSNDI-
                60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          :: : ::::: . . . . . :: . . . . . . . . . . . . . . . . . . . . .
PDB:1X  -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
                120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
PDB:1X  QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVKGGQ
                180          190          200          210          220

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
PDB:1X  SYQSPGYTLVEGDASSASYFLAAAIAKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
                230          240          250          260          270          280

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          . . . : . . . . . : . . . . . : . . . . . : . . . . .
PDB:1X  GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE
                290          300          310          320          330          340

```

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADITYDDHRMAMAFSLAACAEVPV
          : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
PDB:1X  TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
                350          360          370          380          390          400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        :: :: :: :::::.. :.
PDB:1X  TILDPKCTAKTFPDYFEQLARISQAA
                410          420

```

1X8T_A|EPSPS LIGANDED WITH THE (R)-PHOSPHONATE ANALOG OF THE TETRAHEDRAL REACTION INTERMEDIATE|*ESCHERICHIA COLI*|AA|427

Length = 427

initn: 954 initl: 398 opt: 573 Z-score: 1145.6 bits: 221.1 E(): 6.3e-57

Smith-Waterman score: 979; 49.312% identity (73.624% similar) in 436 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        :...::: :...::: :...::: :...::: :...::: :...:::
PDB:1X   MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVHRMLNAL
                10          20          30          40          50

```

```

                70          80          90         100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
        :...: . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1X  TALGVSYTLSADTRCEIIGNGGPLHAEGALE---LFLGNAGTAMRPLAAALCLGSNDI-
                60          70          80          90         100         110

```

```

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        :: : ::::: : : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1X  -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
                120         130         140         150         160         170

```

```

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :. . . . . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1X  QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVKGGQ
                180         190         200         210         220

```

```

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        :...: . :...::: :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1X  SYQSPGYTLVEGDASSASYFLAAAIAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
                230         240         250         260         270         280

```

```

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
        . . : : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1X  GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE
                290         300         310         320         330         340

```

```

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADITYDDHRMAMAFSLAACAEVPV
        :.. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1X  TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
                350         360         370         380         390         400

```

```

      430      440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      :: :: :: :..... . .
PDB:1X TILDPKCTAKTFPDYFEQLARISQAA
      410      420

1G6T_A|STRUCTURE OF EPSP SYNTHASE LIGANDED WITH SHIKIMATE-3-
PHOSPHATE|ESCHERICHIA COLI|AA|427
Length = 427
initn: 954 initl: 398 opt: 573 Z-score: 1145.6 bits: 221.1 E(): 6.3e-57
Smith-Waterman score: 979; 49.312% identity (73.624% similar) in 436 aa overlap
(5-440:2-421)

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :...::: ..:::.....::: .....::: .....::: .....:::
PDB:1G MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTLNLLSDDDVRHMLNAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCCKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      .:::. . :. .: :: . : : : : : : : : : : : : : : : : : : : .
PDB:1G TALGVSYTLSADRTRCEIIGNGGPLHAEGALE--LFLGNAGTAMRPLAAALCLGSNDI-
      60      70      80      90      100      110

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :: : :::::.....::: :. .: :: . . . : : : : : : : : : : : : : :
PDB:1G -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVDVDGSVSS
      120      130      140      150      160      170

      190      200      210      220      230      240
2MEPSP QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . :. .: :. : : : : : : : : : : : : : : : : : : : : : :
PDB:1G QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEIENQH-YQQFVVKGGQ
      180      190      200      210      220

      250      260      270      280      290      300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      .:::. . : : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:1G SYQSPGTYLVEGDASSASYFLAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
      230      240      250      260      270      280

      310      320      330      340      350      360
2MEPSP TETSVTVTGPPREPFGKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:1G GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTLRLNIYNWRVKE
      290      300      310      320      330      340

      370      380      390      400      410      420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVPV
      :. . :. : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:1G TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIIATYNDHRMAMCFSLVALSDTPV
      350      360      370      380      390      400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        :: :: :: ::::: . .
PDB:1G  TILDPKCTAKTFPDYFEQLARISQAA
                410          420

```

2QFS_A|*E. COLI* EPSP SYNTHASE PRO101SER LIGANDED WITH S3P|*ESCHERICHIA COLI*|AA|427
Length = 427
initn: 954 initl: 398 opt: 573 Z-score: 1145.6 bits: 221.1 E(): 6.3e-57
Smith-Waterman score: 984; 49.541% identity (73.853% similar) in 436 aa overlap
(5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        :...::: . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:2Q   MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVRLHMLNAL
                10          20          30          40          50

                70          80          90          100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
        .::: . . . . . : . . : : . . : : : : : : : : : : : : : . . . . .
PDB:2Q   TALGVSYTLSADTRCEIIGNGGPLHAEGALE---LFLGNAGTAMRSLAAALCLGSNDI-
                60          70          80          90          100         110

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        :: : : : : : : : : : : : : : : . . . . . : : : : : : : : : : : : :
PDB:2Q   -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
                120         130         140         150         160         170

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        : . . . . . : . . : : : : : : : : : : : : : : : : : : : : : : :
PDB:2Q   QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVKGGQ
                180         190         200         210         220

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        .::: . : : : : : : : : : : : . . . : : : : : : : : : : : : : : :
PDB:2Q   SYQSPGYTLVEGDASSASYFLAAAIAKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
                230         240         250         260         270         280

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKE
        . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:2Q   GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE
                290         300         310         320         330         340

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADITYDDHRMAMAFSLAACAEVPV
        : . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:2Q   TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
                350         360         370         380         390         400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      :: :: :: :::::.. ::
PDB:2Q TILDPKCTAKTFPDYFEQLARISQAA
                410          420

```

2PQ9_A|*E. COLI* EPSPS LIGANDED WITH (R)-DIFLUOROMETHYL TETRAHEDRAL REACTION
 INTERMEDIATE ANALOG|*ESCHERICHIA COLI*|AA|427

Length = 427

initn: 954 initl: 398 opt: 573 Z-score: 1145.6 bits: 221.1 E(): 6.3e-57

Smith-Waterman score: 979; 49.312% identity (73.624% similar) in 436 aa overlap
 (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :...::: :...::: :...::: :...::: :...::: :...:::
PDB:2P MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVHRMLNAL
                10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      :...: . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:2P TALGVSYTLSADTRCEIIGNGGPLHAEGALE---LFLGNAGTAMRPLAAALCLGSNDI-
                60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :: : ::::: : : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:2P -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
                120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:2P QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVKGGQ
                180          190          200          210          220

```

```

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      :...: . :...: :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:2P SYQSPGYTLVEGDASSASYFLAAAIAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
                230          240          250          260          270          280

```

```

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPEPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . : : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:2P GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE
                290          300          310          320          330          340

```

```

                370          380          390          400          410          420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVPV
      :...: . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:2P TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
                350          360          370          380          390          400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        :: :: :: ::::::::::. :.
PDB:2P  TILDPKCTAKTFPDYFEQLARISQAA
                410          420

```

1MI4_A|GLYPHOSATE INSENSITIVE G96A MUTANT EPSP SYNTHASE LIGANDED WITH SHIKIMATE-3-PHOSPHATE|*ESCHERICHIA COLI*|AA|427
Length = 427
initn: 954 initl: 398 opt: 573 Z-score: 1145.6 bits: 221.1 E(): 6.3e-57
Smith-Waterman score: 973; 49.083% identity (73.624% similar) in 436 aa overlap (5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        :...::: :...::: :...::: :...::: :...::: :...:::
PDB:1M   MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVHRMLNAL
                10          20          30          40          50

```

```

                70          80          90         100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
        :...: . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1M  TALGVSYTLSADTRCEIIGNGGPLHAEGALE---LFLGNAATAMRPLAAALCLGSNDI-
                60          70          80          90         100         110

```

```

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        :: : ::::: : : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1M  -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
                120         130         140         150         160         170

```

```

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :. . . . . :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1M  QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVKGGQ
                180         190         200         210         220

```

```

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        :...: . :...::: :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1M  SYQSPGYTLVEGDASSASYFLAAAIAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
                230         240         250         260         270         280

```

```

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
        . . : : :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1M  GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLRNIYNWRVKE
                290         300         310         320         330         340

```

```

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADITYDDHRMAMAFSLAACAEVPV
        :.. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :. :.
PDB:1M  TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
                350         360         370         380         390         400

```

```

      430      440
2MEPSP TIRDPGCTRKTFPDYFDVLSTFVKN
      :: :: :: :..... . .
PDB:1M TILDPKCTAKTFPDYFEQLARISQAA
      410      420

2QFU_A|E. COLI EPSP SYNTHASE PRO101LEU LIGANDED WITH S3P AND
GLYPHOSATE|ESCHERICHIA COLI|AA|427
Length = 427
initn: 954 initl: 398 opt: 573 Z-score: 1145.6 bits: 221.1 E(): 6.3e-57
Smith-Waterman score: 978; 49.312% identity (73.624% similar) in 436 aa overlap
(5-440:2-421)

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :...::: ..:::.....::: .....::: .....::: .....:::
PDB:2Q MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTLNLLSDDDRHMLNAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCCKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      .:::. . .: .: :: . : : : : : : : : : : : : : : : : : : : :
PDB:2Q TALGVSYTLSADTRCEIIGNGGPLHAEGALE--LFLGNAGTAMRLLAAALCLGSNDI-
      60      70      80      90      100      110

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :: : :::::.....::: .: .: ::: . . . :.....: .: ::::: .....:
PDB:2Q -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
      120      130      140      150      160      170

      190      200      210      220      230      240
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :. . . . .: .: .: : : : : : : : : : : : : : : : : : : : :
PDB:2Q QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEIENQH-YQQFVVKGGQ
      180      190      200      210      220

      250      260      270      280      290      300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      .:::. . :.....: . . :.....: : : : : : : : : : : : : :
PDB:2Q SYQSPGYTLVEGDASSASYFLAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
      230      240      250      260      270      280

      310      320      330      340      350      360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNMKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      . . : : :.....: :.....: : :.....: .....:
PDB:2Q GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTLRNIYNWRVKE
      290      300      310      320      330      340

      370      380      390      400      410      420
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVPV
      :..: . : : :.....: : :.....: . : :.....: : : : : : : : :
PDB:2Q TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPV
      350      360      370      380      390      400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        :: :: :: ::::: . .
PDB:2Q  TILDPKCTAKTFPDYFEQLARISQAA
                410          420

```

1EPS_A | STRUCTURE AND TOPOLOGICAL SYMMETRY OF THE GLYPHOSPHATE 5- ENOL-
PYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE: A DISTINCTIVE PROTEIN FOLD | *ESCHERICHIA*
COLI | AA | 427
Length = 427
initn: 950 init1: 397 opt: 572 Z-score: 1143.6 bits: 220.7 E(): 8.2e-57
Smith-Waterman score: 975; 49.083% identity (73.394% similar) in 436 aa overlap
(5-440:2-421)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        :...::: . . . . . :...::: . . . . . :...::: . . . . . :...::: . . . . .
PDB:1E   MESLTLQPIARVDGTINLPGSKTVSNRALLLAALAHGKTVLTNLLDSDDVHRHMLNAL
                10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
        .::: . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
PDB:1E   TALGVSYTLSADRTCEIIGNGGPLHAEGALE---LFLGNAGTAMRPLAAALCLGSNDI-
                60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        :: : ::::: : : . . : : . . . . . : . . . . . : . . . . .
PDB:1E   -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVVDVDSVSS
                120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :. . . . . : . . . . . : . . . . . : . . . . . : . . . . .
PDB:1E   QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQH-YQQFVVKGGQ
                180          190          200          210          220

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        .::: . : . . . . . : . . . . . : . . . . . : . . . . .
PDB:1E   SYQSPGTYLVEGDASSASYFLAAAIAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICW
                230          240          250          260          270          280

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKE
        . . . : : . . . . . : . . . . . : : . . . : . . . . .
PDB:1E   GDDYISCT-----RGELNAIDMDMNHIPDAAMTIATAALFAKGTTRLRNIYNWRVKE
                290          300          310          320          330          340

```

```

          370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPV
       . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:1E  TDRLFAMATELRKVGAEVEEGHDYIRITPPEKLNFAEIIATYNDHRMAMCFSLVALSDTPV
          350          360          370          380          390          400

          430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
       : : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:1E  TILDPKCTAKTFPDYFEQLARISQAA
          410          420

```

Smith-Waterman score: 971; 49.083% identity (73.394% similar) in 436 aa overlap (5-440:2-421)

```

                10      20      30      40      50      60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLLAALSEGTTVVDNLLNSEDVHYMLGAL
        :...::: . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:1Q   MESLTQLPIARVDGTINLPKSKSVSNRALLLAALAHGKTVLTLNLLDSDDVHRHMLNAL
                10      20      30      40      50

                70      80      90      100     110     120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFVPEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
        .:. . . . . : . . . : : : : : : : : : : : : : : : : .
PDB:1Q  TALGVSYTLSADRTRCEIIGNGGPLHAEGALE--LFLGNAGTAMRPLAALCLGSNDI-
                60      70      80      90      100     110

```

130 140 150 160 170 180
 2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
 :: : ::::: :: :. :: . . ::::: :. :::: ::::
 PDB:1Q -VLTGEPMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVDDGVS
 120 130 140 150 160 170

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      . . . . . : . : : : : : : : : : : : : : : : : : : : : : : :
PDB:1Q QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMKTFGVEIENQH-YQQFVVKGGQ
                180          190          200          210          220

```

2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
 . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
PDB:1Q SYQSPGTYLVEGDASSASYFLAAAAIKGGTVKVKTIGRNSMQGDIRFADVLEKMGATICW
 230 240 250 260 270 280

```

          310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGGRKHLK AIDVNMNKM PDMATLAVVALFADGPTAIRDVASWRVKE
      . . . : : . : : : : : : : : : : : : : : : : : : : : : : : : :
PDB:1Q GDDYISCT-----RGELNAIDMDMNHIPAAAMTIATAALFAKGTTLRNINWRVKE
      290          300          310          320          330          340

```

	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT	AIDTYDDHRMAMAFSLAACAEVPV				
	:::	:	:::	::::	::::	::::
PDB:1Q	TDRLFAMATELRKVGAEEVEEGHDYIRITPPEKLNFAE	IATYNDHRMAMCFSLVALSDTPV				
	350	360	370	380	390	400
	430	440				
2MEPSP	TIRDPGCTRKTFPDYFDVLSTFVKN					
	::	::	::	:::::	:	:
PDB:1Q	TILDPKCTAKTFPDYFEQLARISQAA					
	410	420				

10 20 30 40 50 60
 2MEPSP MAGAEFIVLQPIKEISGTVKLPGSKSLSNRILLALLAALSEGTTVVDNLLNSEDVHYMLGAL
 PDB:1P SNRALLLAALAHGKTVLTLNLLSDDDVRHMLNAL
 10 20 30

70 80 90 100 110 120
 2MEPSP RTLGSLVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
 : . . . :
 PDB:1P TALGVSYTSLADSTRCEIIGNGGPLHAEGA--LELFLGNAGTAMRPLAALCLGSNDI-
 40 50 60 70 80

130 140 150 160 170 180
 2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
 :: : : ::
 PDB:1P -VLTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPLRLQG--GFTGGNVDDVGSVSS
 90 100 110 120 130 140

190 200 210 220 230 240
 2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
 :
 PDB:1P QFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTKTFGVEIENQH-YQQFVVKGGQ
 150 160 170 180 190 200

```

          250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        .:.:. . :.
PDB:1P  SYQSPGTYLVE
          210

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1P88_A | SUBSTRATE-INDUCED STRUCTURAL CHANGES TO THE ISOLATED N- TERMINAL DOMAIN OF 5-ENOLPYRUVYL SHIKIMATE-3-PHOSPHATE SYNTHASE | *ESCHERICHIA COLI* | AA | 216
Length = 216
initn: 365 initl: 141 opt: 388 Z-score: 777.6 bits: 152.0 E(): 2e-36
Smith-Waterman score: 388; 42.857% identity (69.643% similar) in 224 aa overlap (28-251:1-216)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLP	GSKSLSNRILL	LLAALSEGTTVVDNLLN	SEDVHYMLGAL		
			::: ::::: .: ::: ::::: ::::: :::::			
PDB:1P			SNRALLLAALAHGKTVL	TNLLDSDDV	RHMLNAL	
			10	20	30	
	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVV	GCGKFPVEDAKEEV	QLFLGNAGIAMRSL	xxxxxxxxxxxx		
	..: .	:. .: :: . .: :	. .: : . .: ::::: ::: ::::			
PDB:1P	TALGVSYTLSADRTRCEI	IIGNGGPLHAEGA---	LELFLGNAGTAMRPLA	AALCLGSNDI-		
	40	50	60	70	80	
	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIGDL	VVGLKQLGADVDCFL	GTDCPPVRVNGIGGL	PGGKVKLSGSISS		
	:: : ::::: ::: .: .: .	. .: . .: . .: .	. .: .: .: .: .: .: .	.: .: .: .: .: .: .		
PDB:1P	-VLTGEPRMKERPIGHL	VDALRLGGAKITYLE	QENYPPLRLQG--	GFTGGNV	DVDSVSS	
	90	100	110	120	130	140
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxGDVE	IEIIDKLISIPVEM	TLRLMERFGVKA	EHSDSWDRFYIKGGQ		
	:.: .: .: .: .: .: .	.: .: .: .: .: .: .: .	.: .: .: .: .: .: .	.: .: .: .: .: .: .	.: .: .: .: .: .: .
PDB:1P	QFLTALLMTAPLAPED	TVIRIKGDLVSKPY	IDITLNLTKTFGVE	IENQH-YQQFVVKGGQ		
	150	160	170	180	190	200
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASS	ASYFLAGAAI	xxxxxxxxxxxxxSLQ	GDVKFAEVLEMMGAK	VTW	
	.: .: . .: .: .					
PDB:1P	SYQSPGTYLVE					
	210					

200Z_A | *MYCOBACTERIUM TUBERCULOSIS* EPSP SYNTHASE IN COMPLEX WITH PRODUCT (EPS) | *MYCOBACTERIUM TUBERCULOSIS* | AA | 450
Length = 450
initn: 282 initl: 72 opt: 222 Z-score: 437.8 bits: 90.2 E(): 1.7e-17
Smith-Waterman score: 330; 28.538% identity (59.629% similar) in 431 aa overlap (15-435:13-415)

	10	20	30	40	50	
2MEPSP	MAGAEIIVLQPIKEISGTVKLP	GSKSLSNRILL	LAALS---EGTTVVDNLLN	SEDVHYM		
		. .: .: .: .: .: .: .	.: .: .: .: .: .: .	.: .: .: .: .: .: .	.: .: .: .: .: .: .	
PDB:20	MKTWPAPTAPT	PVRATVTVPGSKSQT	NRALVLAALAA	AQGRGASTIS	GALRSRDTELM	
	10	20	30	40	50	
	60	70	80	90	100	110
2MEPSP	LGALRTLGLSVEADKAAKRAVV	GCGKFPVEDAKEE--	VQLFLGNAGIAMRSL	xxxxx		
	: .: .: .: .: .	.: .: .: .: .: .	.: .: .: .: .: .	.: .: .: .: .: .	.: .: .: .: .: .	.: .: .: .: .: .
PDB:20	LDALQTLGLRVD-----	GVGSELT	VS	GRIE	PGPGARVDCGLAGT	VLRFVPLAA
	60	70	80	90	100	

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      120      130      140      150      160      170
2MEPSP xxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVK
      .  .  ::  .  :  ::::  .  .  :::::  ::  ::  :  ::::  :::  ::  :
PDB:2O LGSVPVTF--DGDQQARGRPIAPLLDALRELGVAVD--GTGLP-FRVRGNSLAGGTVA
      110      120      130      140      150      160

      180      190      200      210      220      230
2MEPSP LSGSISSQYxxxxxxxxxxxxGDVEIEII-DKLISIPYVEMTLRLMERFGVKAHSDSWD
      ...:  :::.  .  ....  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O IDASASSQFVSGLLLSAASFDTGLTVQHTGSSSLPSAPHIAMTAAMLRQAGVDIDDSTP-N
      170      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP RFYIKGGQYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVL
      :.  .  :  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O RWQVRPGPV--AARRWDIEPDLTNAVAFLSAAVVSGGTVRITGWPRVSVQPADHILAILR
      230      240      250      260      270

      300      310      320      330      340      350
2MEPSP MMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIR-
      ...:  :  ....  :  ::  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O QLNAVVIHADSSLEVRGPTG-----YDGFVDLRAVGELTPSVAALAALASPGSVSRL
      280      290      300      310      320      330

      360      370      380      390      400      410
2MEPSP -DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIDTYDDHRMAMAF
      .:  :  :::::  .  :::::  .  :  ::  .  .  .  .  .  .  .  .  .
PDB:2O SGIAHLRGHETDRLAALSTEINRLGGTCRETPDGLVITATP-LRPGIWRAYADHRMAMAG
      340      350      360      370      380      390

      420      430      440
2MEPSP SLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      ..  .  .  :  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O AIIGLRVAGVEVDDIAATTKTLPEFPRLWAEMVGPQGQGWGYPQPRSGQRARRATGQSGG
      400      410      420      430      440      450

```

200D_A|*MYCOBACTERIUM TUBERCULOSIS* EPSP SYNTHASE IN COMPLEX WITH
S3P|*MYCOBACTERIUM TUBERCULOSIS*|AA|450

Length = 450

initn: 282 initl: 72 opt: 222 Z-score: 437.8 bits: 90.2 E(): 1.7e-17

Smith-Waterman score: 330; 28.538% identity (59.629% similar) in 431 aa overlap
(15-435:13-415)

```

      10      20      30      40      50
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGSKLSNRILLLAALS---EGTTVVDNLLNSEDEVHYM
      .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O MKTWPAPTAPTVPVRATVTVPGSKSQTNRALVLAALAAQGRGASTISGALRSRDELM
      10      20      30      40      50

      60      70      80      90      100      110
2MEPSP LGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEE--VQLFLGNAGIAMRSLxxxxx
      :  :::::  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O LDALQTLGLRVD-----GVGSELTVSGRIEPPGARVDCGLAGTVLRFVPLPAA
      60      70      80      90      100

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      120      130      140      150      160      170
2MEPSP xxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVK
      .  .  ::  .  :  ::::  .  .  :::::  ::  ::  :  ::::  :::  ::  :
PDB:2O LGSVPVTF--DGDQQARGRPIAPLLDALRELGVAVD--GTGLP-FRVRGNSLAGGTVA
      110      120      130      140      150      160

      180      190      200      210      220      230
2MEPSP LSGSISSQYxxxxxxxxxxxxxxGDVEIEII-DKLISIPYVEMTLRLMERFGVKAHSDSWD
      ...:  :::.  .  ....  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O IDASASSQFVSGLLLSAASFDTGLTVQHTGSSSLPSAPHIAMTAAMLRQAGVDIDDSTP-N
      170      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP RFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVL
      :.  .  :  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O RWQVRPGPV--AARRWDIEPDLTNAVAFLSAAVVSGGTVRITGWPRVSVQPADHILAILR
      230      240      250      260      270

      300      310      320      330      340      350
2MEPSP MMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIR-
      ...:  :  ....:  :  ::  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O QLNAVVIHADSSLEVRGPTG-----YDGFVDLRAVGELTPSVAALAALASPGSVSRL
      280      290      300      310      320      330

      360      370      380      390      400      410
2MEPSP -DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIDTYDDHRMAMAF
      .:  :  :::::  .  :::::  .  :  ::  .:  .  .  .  .  .  .  .  .  .
PDB:2O SGIAHLRGHETDRLAALSTEINRLGGTCRETPDGLVITATP-LRPGIWRAYADHRMAMAG
      340      350      360      370      380      390

      420      430      440
2MEPSP SLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O AIIGLRVAGVEVDDIAATTKTLPEFPRLWAEMVGPQGQGWGYPQPRSGQRARRATGQSGG
      400      410      420      430      440      450

```

200B_A|*MYCOBACTERIUM TUBERCULOSIS* EPSP SYNTHASE IN COMPLEX WITH S3P (PARTIALLY PHOTOLYZED)|*MYCOBACTERIUM TUBERCULOSIS*|AA|450

Length = 450

initn: 282 initl: 72 opt: 222 Z-score: 437.8 bits: 90.2 E(): 1.7e-17

Smith-Waterman score: 330; 28.538% identity (59.629% similar) in 431 aa overlap (15-435:13-415)

```

      10      20      30      40      50
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALS---EGTTVVDNLLNSEDEVHYM
      .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O MKTWPAPTAPTVPVRATVTVPGSKSQTNRALVLAALAAQGRGASTISGALRSRDTELM
      10      20      30      40      50

      60      70      80      90      100      110
2MEPSP LGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEE--VQLFLGNAGIAMRSLxxxxxx
      :  :::::  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O LDALQTLGLRVD-----GVGSELTVSGRIEPPGARVDCGLAGTVLRFVPPPLAA
      60      70      80      90      100

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      120      130      140      150      160      170
2MEPSP xxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVK
      .  .  ::  .  :  ::::  .  .  :::::  ::  ::  :  ::::  :::  ::  :
PDB:2O LGSVPVTF--DGDQQARGRPIAPLLDALRELGVAVD--GTGLP-FRVRGNSLAGGTVA
      110      120      130      140      150      160

      180      190      200      210      220      230
2MEPSP LSGSISSQYxxxxxxxxxxxxxxGDVEIEII-DKLISIPYVEMTLRLMERFGVKAHSDSWD
      ...:  :::.  .  ....  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O IDASASSQFVSGLLLSAASFDTGLTVQHTGSSLPSAPHIAMTAAMLRQAGVDIDDSTP-N
      170      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP RFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVL
      :.  .  :  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O RWQVRPGPV--AARRWDIEPDLTNAVAFLSAAVVSGGTVRITGWPRVSVQPADHILAILR
      230      240      250      260      270

      300      310      320      330      340      350
2MEPSP MMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIR-
      ...:  :  ....:  :  ::  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O QLNAVVIHADSSLEVRGPTG-----YDGFVDLRAVGELTPSVAALAALASPGSVSRL
      280      290      300      310      320      330

      360      370      380      390      400      410
2MEPSP -DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAF
      .:  :  :::::  .  :::::  .  :  ::  .  .  .  .  .  .  .  .  .
PDB:2O SGIAHLRGHETDRLAALSTEINRLGGTCRETPDGLVITATP-LRPGIWRAYADHRMAMAG
      340      350      360      370      380      390

      420      430      440
2MEPSP SLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      ..  .  .  :  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O AIIGLRVAGVEVDDIAATTKTLPFPRWLAEVMVGPQGQGWGYPQPRSGQRARRATGQSGG
      400      410      420      430      440      450

```

200E_A|*MYCOBACTERIUM TUBERCULOSIS* EPSP SYNTHASE IN COMPLEX WITH S3P AND PEP|*MYCOBACTERIUM TUBERCULOSIS*|AA|450

Length = 450

initn: 282 initl: 72 opt: 222 Z-score: 437.8 bits: 90.2 E(): 1.7e-17

Smith-Waterman score: 330; 28.538% identity (59.629% similar) in 431 aa overlap (15-435:13-415)

```

      10      20      30      40      50
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALS---EGTTVVDNLLNSEDEVHYM
      .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O MKTWPAPTAPTVPVRATVTVPGSKSQTNRALVLAALAAQGRGASTISGALRSRDT
      10      20      30      40      50

      60      70      80      90      100      110
2MEPSP LGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEE--VQLFLGNAGIAMRSLxxxxx
      :  :::::  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O LDALQTLGLRVD-----GVGSELTVSGRIEPPGARVDCGLAGTVLRFVPLPAA
      60      70      80      90      100

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      120      130      140      150      160      170
2MEPSP xxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVK
      .  .  ::  .  :  ::::  .  .  :::::  ::  ::  :  ::::  :::  ::  :
PDB:2O LGSVPVTF--DGDQQARGRPIAPLLDALRELGVAVD--GTGLP-FRVRGNSLAGGTVA
      110      120      130      140      150      160

      180      190      200      210      220      230
2MEPSP LSGSISSQYxxxxxxxxxxxxxxGDVEIEII-DKLISIPYVEMTLRLMERFGVKAHSDSWD
      ...:  :::.  .  ....  .  .  .  .  .  .  ::  ::  ....  ::  .  :  .
PDB:2O IDASASSQFVSGLLLSAASFTDGLTVQHTGSSSLPSAPHIAMTAAMLRQAGVDIDDSTP-N
      170      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP RFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVL
      :.  .  :  .  .  .  :  :  .  :  :::::  .  .  .  .  ::  .  .  .
PDB:2O RWQVRPGPV--AARRWDIEPDLTNAVAFLSAAVVSGGTVRITGWPRVSVQPADHILAILR
      230      240      250      260      270

      300      310      320      330      340      350
2MEPSP MMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIR-
      ...:  :  ....  :  ::  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O QLNAVVIHADSSLEVRGPTG-----YDGFVDLRAVGELTPSVAALAALASPGSVSRL
      280      290      300      310      320      330

      360      370      380      390      400      410
2MEPSP -DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAF
      .:  :  :::::  .  :::::  .  :  ::  .  .  .  .  .  :  :::::
PDB:2O SGIAHLRGHETDRLAALSTEINRLGGTCRETPDGLVITATP-LRPGIWRAYADHRMAMAG
      340      350      360      370      380      390

      420      430      440
2MEPSP SLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      ..  .  .  :  .  .  :  :::::
PDB:2O AIIGLRVAGVEVDDIAATTKTLPFPRWLAEVVGPGQGWPQPRSGQRARRATGQSGG
      400      410      420      430      440      450

```

2015_A|*MYCOBACTERIUM TUBERCULOSIS* EPSP SYNTHASE AFTER PARTIAL PRODUCTS
WITHDRAWAL|*MYCOBACTERIUM TUBERCULOSIS*|AA|450

Length = 450

initn: 282 initl: 72 opt: 222 Z-score: 437.8 bits: 90.2 E(): 1.7e-17

Smith-Waterman score: 330; 28.538% identity (59.629% similar) in 431 aa overlap
(15-435:13-415)

```

      10      20      30      40      50
2MEPSP MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL AALS---EGTTVVDNLLNSEDVHYM
      .  .  :  :::::  .  :  :::::  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O MKTWPAPTAPTVPVRATVTVPGSKSQTNRALVLAALAAQGRGASTISGALRSRDT ELM
      10      20      30      40      50

      60      70      80      90      100      110
2MEPSP LGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEE--VQLFLGNAGIAMRSLxxxxx
      :  :::::  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
PDB:2O LDALQTLGLRVD-----GVGSELTVSGRIE PGPGARVDCGLAGTVLRFVPLPAA
      60      70      80      90      100

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      120      130      140      150      160      170
2MEPSP xxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVK
      .  .  ::  .  :  ::::  .  .  :::::  ::  ::  :  ::::  :::  ::  :
PDB:2O LGSVPVTF--DGDQQARGRPIAPLLDALRELGVAVD--GTGLP-FRVRGNSLAGGTVA
      110      120      130      140      150      160

      180      190      200      210      220      230
2MEPSP LSGSISSQYxxxxxxxxxxxxxxGDVEIEII-DKLISIPYVEMTLRLMERFGVKAHSDSWD
      ...:  :::.  .  ....  .  .  .  .  ::  :  :..  ::  ....  ::  .  :  .
PDB:2O IDASASSQFVSGLLLSAASFDTGLTVQHTGSSSLPSAPHIAMTAAMLRQAGVDIDDSTP-N
      170      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP RFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVL
      :.  .  :  .  .  .  :  :  .  :  :::::  .  .  .  :  :  .  .  .
PDB:2O RWQVRPGPV--AARRWDIEPDLTNAVAFLSAAVVSGGTVRITGWPRVSVQPADHILAILR
      230      240      250      260      270

      300      310      320      330      340      350
2MEPSP MMGAKVTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIR-
      ...:  :  ....  :  ::  .  :  :  .  .  .  .  .  .  :  .  .  :
PDB:2O QLNAVVIHADSSLEVRGPTG-----YDGFVDLRAVGELTPSVAALAALASPGSVSRL
      280      290      300      310      320      330

      360      370      380      390      400      410
2MEPSP -DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIDTYDDHRMAMAF
      .:  :  :::::  .  :::::  .  :  :  :  :  :  .  .  :  :  :  :  :
PDB:2O SGIAHLRGHETDRLAALSTEINRLGGTCRETPDGLVITATP-LRPGIWRAYADHRMAMAG
      340      350      360      370      380      390

      420      430      440
2MEPSP SLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .  .  .  :  :  .  :  :  :  :
PDB:2O AIIGLRVAGVEVDDIAATTKTLPEFPRLWAEMVGPQGQGWGYPQPRSGQRARRATGQSGG
      400      410      420      430      440      450

```

200X_A | *MYCOBACTERIUM TUBERCULOSIS* EPSP SYNTHASE IN COMPLEX WITH
INTERMEDIATE | *MYCOBACTERIUM TUBERCULOSIS* | AA | 450

Length = 450

initn: 282 initl: 72 opt: 222 Z-score: 437.8 bits: 90.2 E(): 1.7e-17

Smith-Waterman score: 330; 28.538% identity (59.629% similar) in 431 aa overlap
(15-435:13-415)

```

      10      20      30      40      50
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALS---EGTTVVDNLLNSEDEVHYM
      .  .  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
PDB:2O MKTWPAPTAPTVPVRATVTVPGSKSQTNRALVLAALAAQGRGASTISGALRSRDTELM
      10      20      30      40      50

      60      70      80      90      100      110
2MEPSP LGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEE--VQLFLGNAGIAMRSLxxxxxx
      :  :::::  .  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
PDB:2O LDALQTLGLRVD-----GVGSELTVSGRIEPPGARVDCGLAGTVLRFVPLPAA
      60      70      80      90      100

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      120      130      140      150      160      170
2MEPSP xxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVK
      .  .  ::  .  :  ::::  .  .  :::::  ::  ::  :  ::::  :::  ::  :
PDB:2O LGSVPVTF--DGDQQARGRPIAPLLDALRELGVAVD--GTGLP-FRVRGNSLAGGTVA
      110      120      130      140      150      160

      180      190      200      210      220      230
2MEPSP LSGSISSQYxxxxxxxxxxxxGDVEIEII-DKLISIPYVEMTLRLMERFGVKAHSDSWD
      ...:  :::.  .  ....  .  .  .  .  ::  :  :..  ::  ....  ::  .  :  .
PDB:2O IDASASSQFVSGLLLSAASFDTGLTVQHTGSSSLPSAPHIAMTAAMLRQAGVDIDDSTP-N
      170      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP RFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVL
      :.  .  :  .  .  .  :  :  .  :  :::::  .  .  .  :  :  .  .  .
PDB:2O RWQVRPGPV--AARRWDIEPDLTNAVAFLSAAVVSGGTVRITGWPRVSVQPADHILAILR
      230      240      250      260      270

      300      310      320      330      340      350
2MEPSP MMGAKVTWTETSVTVTGPPREPFGKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIR-
      ...:  :  ....  :  ::  .  :  :  .  .  .  .  .  .  :  .  .  :
PDB:2O QLNAVVIHADSSLEVRGPTG-----YDGFVDLRAVGELTPSVAALAALASPGSVSRL
      280      290      300      310      320      330

      360      370      380      390      400      410
2MEPSP -DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAF
      .:  :  :::::  .  :::::  .  :  :  .  :  :  .  :  :  :  :  :
PDB:2O SGIAHLRGHETDRLAALSTEINRLGGTCRETPDGLVITATP-LRPGIWRAYADHRMAMAG
      340      350      360      370      380      390

      420      430      440
2MEPSP SLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      ..  .  .  :  .  :  .  :  :  :
PDB:2O AIIGLRVAGVEVDDIAATTKTLPFEPRLWAEMVGPQGQGWGYPQPRSGQRARRATGQSGG
      400      410      420      430      440      450

```

2BJB_A | *MYCOBACTERIUM TUBERCULOSIS* EPSP SYNTHASE IN UNLIGANDED STATE | *MYCOBACTERIUM TUBERCULOSIS* | AA | 462

Length = 462

initn: 282 initl: 72 opt: 222 Z-score: 437.6 bits: 90.2 E(): 1.7e-17

Smith-Waterman score: 330; 28.538% identity (59.629% similar) in 431 aa overlap (15-435:13-415)

```

      10      20      30      40      50
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGSKLSNRILLLAALS---EGTTVVDNLLNSEDVHYM
      .  .  :  .  :  :  .  :  :  .  :  :  .  :  :  .  :  :  .  :  :
PDB:2B MKTWPAPTAPTVPVRATVTVPGSKSQTNRALVLAALAAQGRGASTISGALRSRDT
      10      20      30      40      50

      60      70      80      90      100      110
2MEPSP LGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEE--VQLFLGNAGIAMRSLxxxxx
      :  :::::  .  .  :  :  :  .  :  :  .  :  :  .  :  :  .  :  :
PDB:2B LDALQTLGLRVD-----GVGSELTVSGRIEPPGARVDCGLAGTVLRFVPLPAA
      60      70      80      90      100

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      120      130      140      150      160      170
2MEPSP xxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVK
      .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
PDB:2B LGSVPVTF--DGDQQARGRPIAPLLDALRELGVAVD---GTGLP-FRVRGNSLAGGTVA
      110      120      130      140      150      160

      180      190      200      210      220      230
2MEPSP LSGSISSQYxxxxxxxxxxxxxxGDVEIEII-DKLISIPYVEMTLRLMERFGVKAHSDSWD
      .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
PDB:2B IDASASSQFVSGLLLSAASFDTGLTVQHTGSSSLPSAPHIAMTAAMLRQAGVDIDDSTP-N
      170      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP RFYIKGGQYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVL
      .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
PDB:2B RWQVRPGPV--AARRWDIEPDLTNAVAFLSAAVVSGGTVRITGWPRVSVQPADHILAILR
      230      240      250      260      270

      300      310      320      330      340      350
2MEPSP MMGAKVTWTETSVTVTGPPREPFGKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIR-
      .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
PDB:2B QLNAVVIHADSSLEVRGPTG-----YDGFVDLRAVGELTPSVAALAALASPGSVSRL
      280      290      300      310      320      330

      360      370      380      390      400      410
2MEPSP -DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAF
      .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
PDB:2B SGIAHLRGHETDRLAALSTEINRLGGTCRETPDGLVITATP-LRPGIWRAYADHRMAMAG
      340      350      360      370      380      390

      420      430      440
2MEPSP SLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
PDB:2B AIIGLRVAGVEVDDIAATTKTLPEFPRLWAEMVGPQGQGWGYPQPRSGQRARRATGQSGG
      400      410      420      430      440      450

```

1RF5_B | STRUCTURAL STUDIES OF *STREPTOCOCCUS PNEUMONIAE* EPSP SYNTHASE IN UNLIGANDED STATE | *STREPTOCOCCUS PNEUMONIAE* | AA | 427

Length = 427

initn: 252 initl: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10

Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap (12-443:7-425)

```

      10      20      30      40      50      60
2MEPSP MAGAEIIVLQPIKEISGTVKLPKSGSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
PDB:1R MKLKTNIRHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
PDB:1R RDLGVEIEDKDGVITVQGVGMAGLKAPQNA-----LNMGNSGTSIRLISGVLAGADFEVE
      60      70      80      90      100      110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : . .:. . . : : : : . : : : : . : . : .
PDB:1R  --MFGDDSLSKRPMDRVTLP LKKMGVSISGQTERDLPPLRLKGTKNLRPIHYELP-IASA
                120          130          140          150          160

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . : : : : . . : . . : : . . : . . : :
PDB:1R  QVKSALMFAALQAKG--ESVIEKEYTRNHT--DMLQQFG--GHLSVDGKKITVQGPQ
                170          180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . : . : : : : : . . . . . . . . . . : : : :
PDB:1R  KLTGQK-VVVPDISAFAFWLVAGLIAPNSRLVLQNVGIN-ETRTGIIDVIRAMGGKLEI
                230          240          250          260          270

                310          320          330          340          350
2MEPSP  TETSVTVTGPPR--EPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      : : . . . : : . . . : . : : : : : : : : :
PDB:1R  TEIDPVAKSATLIVESSDLKGTEICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKV
                280          290          300          310          320          330

                360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAAC--A
      : : : . . : : : : : : : : . : . . . : : : : :
PDB:1R  KETDRIQVVADALNSMGADITPTADGMIKGSALHGARVNTFGDHRIGMMTAIAALLVA
                340          350          360          370          380          390

                420          430          440
2MEPSP  EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : : : : :
PDB:1R  DGEVELDRAEAINTSYPSFFDDLES LIHG
                400          410          420

```

1RF5_C | STRUCTURAL STUDIES OF *STREPTOCOCCUS PNEUMONIAE* EPSP SYNTHASE IN UNLIGANDED STATE | *STREPTOCOCCUS PNEUMONIAE* | AA | 427

Length = 427

initn: 252 initl: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10

Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap (12-443:7-425)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL AALSEGTTVVDNLLNSEDVHYMLGAL
      : . . : : : : : : : : : : : : : : :
PDB:1R  MKLKTNIRHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : . . : : . : . : : : : : : . .
PDB:1R  RDLGVEIEDKDGVITVQGVGMAGLKAPQNA-----LNMGNSGTSIRLISGVLAGADFEVE
                60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
PDB:1R  --MFGDDSLSKRPMDRVTLP LKKMGVSISGQTERDLPPLRLKGTKNLRPIHYELP-IASA
                120          130          140          150          160

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQ
      : . . . . : : : : . . : . . : . : : . . : . . : . :
PDB:1R  QVKSALMFAALQAKG--ESVIEKEYTRNHT--DMLQQFG--GHLSVDGKKITVQGPQ
                170          180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . : . : : : : : . . . . . : . . . . . : . : . : . :
PDB:1R  KLTGQK-VVVPGLISSAAFVWVAGLIAPNSRLVLQNVGIN-ETRTGIIDVIRAMGGKLEI
                230          240          250          260          270

                310          320          330          340          350
2MEPSP  TETSVTVTGPPR--EPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      : . . . : : . . . : . : : : : : : : : : . :
PDB:1R  TEIDPVAKSATLIVESSDLKGTEICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKV
                280          290          300          310          320          330

                360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAAC--A
      : : : . . : : : : : : : : . : . . . : : : : : : :
PDB:1R  KETDRIQVVADALNSMGADITPTADGMIKGSALHGARVNTFGDHRIGMMTAIAALLVA
                340          350          360          370          380          390

                420          430          440
2MEPSP  EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : : : : :
PDB:1R  DGEVELDRAEAINTSYPSFFDDLES LIHG
                400          410          420

```

1RF4_C | STRUCTURAL STUDIES OF *STREPTOCOCCUS PNEUMONIAE* EPSP SYNTHASE, TETRAHEDRAL INTERMEDIATE BOUND STATE | *STREPTOCOCCUS PNEUMONIAE* | AA | 427

Length = 427

initn: 252 initl: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10

Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap (12-443:7-425)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL AALSEGTTVVDNLLNSEDVHYMLGAL
      : . . : . : : : : : : : : : : : : : : : . .
PDB:1R  MKLKTNIRHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : . . : : . : . : : : : : : : . .
PDB:1R  RDLGVEIEDKDGVITVQGVGMAGLKAPQNA-----LNMGNSGTSIRLISGVLGADFEVE
                60          70          80          90          100          110

```

```

1RF6_D|STRUCTURAL STUDIES OF STREPTOCOCCUS PNEUMONIAE EPSP SYNTHASE IN S3P-GLP
BOUND STATE|STREPTOCOCCUS PNEUMONIAE|AA|427
Length = 427
initn: 252 init1: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10
Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap
(12-443:7-425)

          10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          :... : .....: : : .....: ..
PDB:1R   MKLKTNIIRHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
          10          20          30          40          50

          70          80          90          100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCCKGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
          : :...: .. . : : .. : :...: .. . .
PDB:1R  RDLGVEIEDKDGVITVQGVMAGLKAPQNA-----LNMGNSGTSIRLISGVLGADFEVE
          60          70          80          90          100         110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : . . : . . . : : : : : : : : : : : : : : : : : : : :
PDB:1R  --MFGDDSLSKRPMDRVTLP LKKMGVSISGQTERDLPPLRLKGTKNLRPIHYELP-IASA
                120          130          140          150          160

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . : : : : . . : . . : : : : : . . : . . . : :
PDB:1R  QVKSALMFAALQAKG--ESVIEKEYTRNHT--DMLQQFG--GHLSDVGKKITVQGPQ
                170          180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . : . : : : : : : : . . . . . . . . . . . . : : :
PDB:1R  KLTGQK-VVVPDISAFAFWLVAGLIAPNSRLVLQNVGIN-ETRTGIIDVIRAMGGKLEI
                230          240          250          260          270

                310          320          330          340          350
2MEPSP  TETSVTVTGPPR--EPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      : : . . . : : . . . . : . : : : : : : : : : : :
PDB:1R  TEIDPVAKSATLIVESSDLKGTEICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKV
                280          290          300          310          320          330

                360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAAC--A
      : : : : . . : : : : : : : : . . . . : : : : : : : :
PDB:1R  KETDRIQVVADALNSMGADITPTADGMIKKGKSALHGARVNTFGDHRIGMMTAIAALLVA
                340          350          360          370          380          390

                420          430          440
2MEPSP  EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : : : : : : :
PDB:1R  DGEVELDRAEAINTSYPSFFDDLES LIHG
                400          410          420

```

1RF5_D | STRUCTURAL STUDIES OF *STREPTOCOCCUS PNEUMONIAE* EPSP SYNTHASE IN UNLIGANDED STATE | *STREPTOCOCCUS PNEUMONIAE* | AA | 427

Length = 427

initn: 252 initl: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10

Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap (12-443:7-425)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL AALSEGTTVVDNLLNSEDVHYMLGAL
      : . . : : : : : : : : : : : : : : : : : : : :
PDB:1R  MKLKTNI RHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      : : : : : . . : : . . : . : : : : : : . . . .
PDB:1R  RDLGVEIEDKDG VITVQGVGMAGLKAPQNA-----LNMGNSGTSIRLISGVLGADFEVE
                60          70          80          90          100          110

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : . .:. . . : : : : . : : : : : : : : : : : : : : : :
PDB:1R  --MFGDDSLSKRPMDRVTLPKKMGVSISGQTERDLPPLRLKGTKNLRPIHYELP-IASA
                120          130          140          150          160

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . : : : : . . : . . . : : . . : . . . : :
PDB:1R  QVKSALMFAALQAKG--ESVIEKEYTRNHT--DMLQQFG--GHLSVDGKKITVQGPQ
                170          180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . : . : : : : : : : . . . . . . . . . . . . : : :
PDB:1R  KLTGQK-VVVPDISAFAFWLVAGLIAPNSRLVLQNVGIN-ETRTGIIDVIRAMGGKLEI
                230          240          250          260          270

                310          320          330          340          350
2MEPSP  TETSVTVTGPPR--EPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      : : . . . : : . . . . : . : : : : : : : : : : :
PDB:1R  TEIDPVAKSATLIVESSDLKGTEICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKV
                280          290          300          310          320          330

                360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAAC--A
      : : : : . . : : : : : : : : . : . . . . : : : : : :
PDB:1R  KETDRIQVVADALNSMGADITPTADGMIKKGKSALHGARVNTFGDHRIGMMTAIAALLVA
                340          350          360          370          380          390

                420          430          440
2MEPSP  EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : : : : : : :
PDB:1R  DGEVELDRAEAINTSYPSFFDDLES LIHG
                400          410          420

```

1RF4_D | STRUCTURAL STUDIES OF *STREPTOCOCCUS PNEUMONIAE* EPSP SYNTHASE, TETRAHEDRAL INTERMEDIATE BOUND STATE | *STREPTOCOCCUS PNEUMONIAE* | AA | 427

Length = 427

initn: 252 initl: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10

Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap (12-443:7-425)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL AALSEGTTVVDNLLNSEDVHYMLGAL
      : . . : . : : : : : : : : : : : : : : : : : : :
PDB:1R  MKLKTNIRHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : : . . : : . : . : : : : : : : : : . .
PDB:1R  RDLGVEIEDKDG VITVQGVGMAGLKAPQNA-----LNMGNSGTSIRLISGVLGADFEVE
                60          70          80          90          100          110

```

```

1RF5_A|STRUCTURAL STUDIES OF STREPTOCOCCUS PNEUMONIAE EPSP SYNTHASE IN
UNLIGANDED STATE|STREPTOCOCCUS PNEUMONIAE|AA|427
Length = 427
initn: 252 initl: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10
Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap
(12-443:7-425)

          10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSGSKLSNRILLALLAALSEGTTVVDNLLNSEDVHYMLGAL
          :... : .....: : : .....: : ..
PDB:1R      MKLKTNIIRHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
          10          20          30          40          50

          70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCCKGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
          : :...: .. . : : .. : :...: .. . .
PDB:1R  RDLGVEIEDKDGVITVQGVMAGLKAPQNA-----LNMGNSGTSIRLISGVLGADFEVE
          60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : . .:. . . : : : : : . : : : : : . : . : .
PDB:1R  --MFGDDSLSKRPMDRVTLPKKMGVSISGQTERDLPPLRLKGTKNLRPIHYELP-IASA
                120          130          140          150          160

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . : : : : . . : . . : : . . : . . : :
PDB:1R  QVKSALMFAALQAKG--ESVIEKEYTRNHT--DMLQQFG--GHLSVDGKKITVQGPQ
                170          180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . : . : : : : : . . . . . . . . . . : : . . . . : :
PDB:1R  KLTGQK-VVVPDISAFAFWLVAGLIAPNSRLVLQNVGIN-ETRTGIIDVIRAMGGKLEI
                230          240          250          260          270

                310          320          330          340          350
2MEPSP  TETSVTVTGPPR--EPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      : : . . . : : . . . . : . : : : : : : : : : . .
PDB:1R  TEIDPVAKSATLIVESSDLKGTEICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKV
                280          290          300          310          320          330

                360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAAC--A
      : : : . . : : : : : : : : . : . . . : : : : : : :
PDB:1R  KETDRIQVVADALNSMGADITPTADGMIKGSALHGARVNTFGDHRIGMMTAIAALLVA
                340          350          360          370          380          390

                420          430          440
2MEPSP  EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : : : : : : :
PDB:1R  DGEVELDRAEAINTSYPSFFDDLES LIHG
                400          410          420

```

1RF6_C | STRUCTURAL STUDIES OF *STREPTOCOCCUS PNEUMONIAE* EPSP SYNTHASE IN S3P-GLP BOUND STATE | *STREPTOCOCCUS PNEUMONIAE* | AA | 427

Length = 427

initn: 252 initl: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10

Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap (12-443:7-425)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL AALSEGTTVVDNLLNSEDVHYMLGAL
      : . . : : : : : : : : : : : : : : : : : : . .
PDB:1R  MKLKTNIRHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      : : : : . . . : : . . : : : : : : : . . . .
PDB:1R  RDLGVEIEDKDGVITVQGVGMAGLKAPQNA-----LNMGNSGTSIRLISGVLAGADFEVE
                60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : . .:. . . : : : : : . : : : : : . : . : .
PDB:1R  --MFGDDSLSKRPMDRVTLP LKKMGVSISGQTERDLPPLRLKGTKNLRPIHYELP-IASA
                120          130          140          150          160

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . : : : : . . : . . : : . . : . . : :
PDB:1R  QVKSALMFAALQAKG--ESVIEKEYTRNHT--DMLQQFG--GHLSVDGKKITVQGPQ
                170          180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . : . : : : : : : . . . . . . . . . . : : :
PDB:1R  KLTGQK-VVVPGLISSAAFVWVAGLIAPNSRLVLQNVGIN-ETRTGIIDVIRAMGGKLEI
                230          240          250          260          270

                310          320          330          340          350
2MEPSP  TETSVTVTGPPR--EPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      : : . . . : : . . . . : . : : : : : : : : :
PDB:1R  TEIDPVAKSATLIVESSDLKGTEICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKV
                280          290          300          310          320          330

                360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAAC--A
      : : : : . . : : : : : : : : . : . . . : : : : :
PDB:1R  KETDRIQVVADALNSMGADITPTADGMIKGSALHGARVNTFGDHRIGMMTAIAALLVA
                340          350          360          370          380          390

                420          430          440
2MEPSP  EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : : : : : : :
PDB:1R  DGEVELDRAEAINTSYPSFFDDLES LIHG
                400          410          420

```

1RF6_A | STRUCTURAL STUDIES OF *STREPTOCOCCUS PNEUMONIAE* EPSP SYNTHASE IN S3P-GLP BOUND STATE | *STREPTOCOCCUS PNEUMONIAE* | AA | 427

Length = 427

initn: 252 initl: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10

Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap (12-443:7-425)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL AALSEGTTVVDNLLNSEDVHYMLGAL
      : . . : : : : : : : : : : : : : : : . .
PDB:1R  MKLKTNIRHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : . . : : . : . : : : : : . . . .
PDB:1R  RDLGVEIEDKDGVITVQGVGMAGLKAPQNA-----LNMGNSGTSIRLISGVLAGADFEVE
                60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : . .:. . . : : : : . : : : : : : : . : . : .
PDB:1R  --MFGDDSLSKRPMDRVTLP LKKMGVSISGQTERDLPPLRLKGTKNLRPIHYELP-IASA
                120          130          140          150          160

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQ
      : . . . . : : : : . . : . . : : . . : . . : :
PDB:1R  QVKSALMFAALQAKG--ESVIEKEYTRNHT--DMLQQFG--GHLSDVGKKITVQGPQ
                170          180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . : . : : : : : : . . . . . . . . . . : : .
PDB:1R  KLTGQK-VVVPGLISSAAFVWVAGLIAPNSRLVLQNVGIN-ETRTGIIDVIRAMGGKLEI
                230          240          250          260          270

                310          320          330          340          350
2MEPSP  TETSVTVTGPPR--EPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      : : . . . : : . . . . : . : : : : : : : .
PDB:1R  TEIDPVAKSATLIVESSDLKGTEICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKV
                280          290          300          310          320          330

                360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAAC--A
      : : : . . : : : : : : : : . : . . . : : : : :
PDB:1R  KETDRIQVVADALNSMGADITPTADGMIKGSALHGARVNTFGDHRIGMMTAIAALLVA
                340          350          360          370          380          390

                420          430          440
2MEPSP  EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : : : : : .
PDB:1R  DGEVELDRAEAINTSYPSFFDDLES LIHG
                400          410          420

```

1RF4_A | STRUCTURAL STUDIES OF *STREPTOCOCCUS PNEUMONIAE* EPSP SYNTHASE, TETRAHEDRAL INTERMEDIATE BOUND STATE | *STREPTOCOCCUS PNEUMONIAE* | AA | 427

Length = 427

initn: 252 initl: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10

Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap (12-443:7-425)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL AALSEGTTVVDNLLNSEDVHYMLGAL
      : . . : : : : : : : : : : : : : : : . .
PDB:1R  MKLKTNI RHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : : . . : : . . : : : : : . . . .
PDB:1R  RDLGVEIEDKDG VITVQGVGMAGLKAPQNA-----LNMGNSGTSIRLISGVLGADFEVE
                60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : . .:. . . : : : : . . : : : : : . : . : .
PDB:1R  --MFGDDSLSKRPMDRVTLP LKKMGVSISGQTERDLPPLRLKGTKNLRPIHYELP-IASA
                120          130          140          150          160

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQ
      : . . . . : : : : . . : . . : : . . : . . : :
PDB:1R  QVKSALMFAALQAKG--ESVIEKEYTRNHT--DMLQQFG--GHLSDVGKKITVQGPQ
                170          180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . : . : : : : : : . . . . . : . . . . : : :
PDB:1R  KLTGQK-VVVPDISAFAFWLVAGLIAPNSRLVLQNVGIN-ETRTGIIDVIRAMGGKLEI
                230          240          250          260          270

                310          320          330          340          350
2MEPSP  TETSVTVTGPPR--EPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      : : . . . : : . . . : . : : : : : : : : . .
PDB:1R  TEIDPVAKSATLIVESSDLKGTEICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKV
                280          290          300          310          320          330

                360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAAC--A
      : : : . . : : : : : : : . : . . . : : : : : :
PDB:1R  KETDRIQVVADALNSMGADITPTADGMIKGSALHGARVNTFGDHRIGMMTAIAALLVA
                340          350          360          370          380          390

                420          430          440
2MEPSP  EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : : : : : :
PDB:1R  DGEVELDRAEAINTSYPSFFDDLES LIHG
                400          410          420

```

1RF4_B | STRUCTURAL STUDIES OF *STREPTOCOCCUS PNEUMONIAE* EPSP SYNTHASE, TETRAHEDRAL INTERMEDIATE BOUND STATE | *STREPTOCOCCUS PNEUMONIAE* | AA | 427

Length = 427

initn: 252 initl: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10

Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap (12-443:7-425)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL AALSEGTTVVDNLLNSEDVHYMLGAL
      : . . : : : : : : : : : : : : : : :
PDB:1R  MKLKTNIRHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : . . : : . . : : : : : : : . .
PDB:1R  RDLGVEIEDKDGVITVQGVGMAGLKAPQNA-----LNMGNSGTSIRLISGVLGADFEVE
                60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : . .:. . . : : : : : . : : : : : . : . : .
PDB:1R  --MFGDDSLSKRPMDRVTLP LKKMGVSISGQTERDLPPLRLKGTKNLRPIHYELP-IASA
                120          130          140          150          160

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . : : : : . . : . . : : . . : . . : :
PDB:1R  QVKSALMFAALQAKG--ESVIEKEYTRNHT--DMLQQFG--GHLSVDGKKITVQGPQ
                170          180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . : . : : : : : : . . . . . : . . . . . : : :
PDB:1R  KLTGQK-VVVPGLISSAAFVLVAGLIAPNSRLVLQNVGIN-ETRTGIIDVIRAMGGKLEI
                230          240          250          260          270

                310          320          330          340          350
2MEPSP  TETSVTVTGPPR--EPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      : : . . . : : . . . : . : : : : : : : : : .
PDB:1R  TEIDPVAKSATLIVESSDLKGTEICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKV
                280          290          300          310          320          330

                360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAAC--A
      : : : . . : : : : : : : : . : . . . : : : : :
PDB:1R  KETDRIQVVADALNSMGADITPTADGMIKGSALHGARVNTFGDHRIGMMTAIAALLVA
                340          350          360          370          380          390

                420          430          440
2MEPSP  EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : : : : :
PDB:1R  DGEVELDRAEAINTSYPSFFDDLES LIHG
                400          410          420

```

1RF6_B | STRUCTURAL STUDIES OF *STREPTOCOCCUS PNEUMONIAE* EPSP SYNTHASE IN S3P-GLP BOUND STATE | *STREPTOCOCCUS PNEUMONIAE* | AA | 427

Length = 427

initn: 252 initl: 113 opt: 155 Z-score: 303.2 bits: 65.2 E(): 5.3e-10

Smith-Waterman score: 281; 24.541% identity (58.028% similar) in 436 aa overlap (12-443:7-425)

```

                10          20          30          40          50          60
2MEPSP  MAGAE EIVLQPIKEISGTVKLP GSKSLSNRILL AALSEGTTVVDNLLNSEDVHYMLGAL
      : . . : : : : : : : : : : : : : : : . .
PDB:1R  MKLKTNIRHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVF
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : . . . : : . . : : : : : : . . .
PDB:1R  RDLGVEIEDKDG VITVQGVGMAGLKAPQNA-----LNMGNSGTSIRLISGVLGADFEVE
                60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : . . : . . . : : : : : : : : : : : : : : : : : : : :
PDB:1R --MFGDDSLSKRPMDRVTLPLKKMGVSISGQTERDLPPLRLKGTKNLRPIHYELP-IASA
                120          130          140          150          160

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . : : : : . . : . . . : . . : . . . : :
PDB:1R QVKSALMFAALQAKG--ESVIEKEYTRNHT--DMLQQFG--GHLSVDGKKITVQGPQ
                170          180          190          200          210          220

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . : . : : : : : : : . . . . . . . . . . : : : :
PDB:1R KLTGQK-VVVPDISAFAFWLVAGLIAPNSRLVLQNVGIN-ETRTGIIDVIRAMGGKLEI
                230          240          250          260          270

                310          320          330          340          350
2MEPSP TETSVTVTGPPR--EPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASRV
      : : . . . : : . . . . : . : : : : : : : : :
PDB:1R TEIDPVAKSATLIVESSDLKGTEICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKV
                280          290          300          310          320          330

                360          370          380          390          400          410
2MEPSP KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADITYDDHRMAMAFSLAAC--A
      : : : . . : : : : : : : : . . . . : : : : : : :
PDB:1R KETDRIQVVADALNSMGADITPTADGMIKGSALHGARVNTFGDHRIGMMTAIAALLVA
                340          350          360          370          380          390

                420          430          440
2MEPSP EVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : : : : : : :
PDB:1R DGEVELDRAEAINTSYPSFFDDLES LIHG
                400          410          420

```

2PQD_A|A100G CP4 EPSPS LIGANDED WITH (R)-DIFLUOROMETHYL TETRAHEDRAL REACTION
 INTERMEDIATE ANALOG|AGROBACTERIUM SP.|AA|445

Length = 445

initn: 202 initl: 103 opt: 127 Z-score: 246.4 bits: 54.8 E(): 7.7e-07

Smith-Waterman score: 200; 24.374% identity (53.303% similar) in 439 aa overlap
 (15-440:13-436)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      . : : : : : : : : : : : : : : : : : : : : : : :
PDB:2P SSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTGKAM
                10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      . . : . . : : . : : : : : : : : : : : : :
PDB:2P QAMGARIRKEGDTWIIDGVGNGLLA-----PEAPLDFGNAGTGCRLTMGLVGVDYDFDST
                60          70          80          90          100          110

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . . : . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  FI--GDASLTKRPMGRVLNPLREMGVQVKSEHG--DRLPVTLRG--PKTPTPITYRVPMASA
                120          130          140          150          160

                190          200          210          220          230
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWD--RFYIKG
      : . . . . : . . . . : . . . . : . . . . : . . : . . : . . : . .
PDB: 2P  QVKSAVLLAGLNTPGITT--VIEPIMTRDHT--KMLQGFGANLTVETDADGVRTIRLE
                170          180          190          200          210          220

                240          250          260          270          280          290
2MEPSP  GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
      : . . . . : . . . . : . . . . : . . . . : . . . . : . . . . : . .
PDB: 2P  GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMN--PTRTGLILTLQEMGADI
                230          240          250          260          270          280

                300          310          320          330          340          350
2MEPSP  TWTETSVTVTGPPREPFGRKH--LKAIDVNMNKM---DVAMTLAVVALFADGPTAIRDVA
      . . . . . : . . . . : . . . . : . . . . : . . . . : . . . . : . .
PDB: 2P  EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLE
                290          300          310          320          330          340

                360          370          380          390          400
2MEPSP  SWRVKETERMVAIRTELTKLGASVEEGPDYCIIT--PPEK----LNVTAIDTYDDHRMAM
      . : . . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  ELRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAM
                350          360          370          380          390          400

                410          420          430          440
2MEPSP  AFS--LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  SFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELS
                410          420          430          440

```

2GGD_A | CP4 EPSP SYNTHASE ALA100GLY LIGANDED WITH S3P AND GLYPHOSATE | *AGROBACTERIUM SP.* | AA | 455

Length = 455

initn: 202 initl: 103 opt: 127 Z-score: 246.3 bits: 54.8 E(): 7.8e-07

Smith-Waterman score: 200; 24.374% identity (53.303% similar) in 439 aa overlap (15-440:18-441)

```

                10          20          30          40          50
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYML
      . : . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2G  MSHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
                10          20          30          40          50          60

                60          70          80          90          100          110
2MEPSP  GALRTLGLSVEADKAAKRAVVVGCCKGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxx
      . : . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2G  KAMQAMGARIRKEGDTWIIDGVNGGLLA-----PEAPLDFGNAGTGCRLTMGLVGVDYDF
                70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

120      130      140      150      160      170
2MEPSP xxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGS
      . . . . : . . . . . : . . . . . : : : : : : : .
PDB: 2G DSTFI--GDASLTKRPMGRVLNPLREMGVQVKSEDG--DRLPVTLRG-PKTPTPITYRVPM
      120      130      140      150      160      170

180      190      200      210      220      230
2MEPSP ISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWD--RFY
      . : . . . : . . . . . : . . . . . : . . . : .
PDB: 2G ASAQVKSAVLLAGLNTPGITT--VIEPIMTRDHT--KMLQGFGANLTVETDADGVRTI
      180      190      200      210      220

240      250      260      270      280      290
2MEPSP IKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMG
      . : . . : : : . . . . . : . . . . . : . . . : .
PDB: 2G RLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMN-PTRTGLILTLQEMG
      230      240      250      260      270      280

300      310      320      330      340      350
2MEPSP AKVTWTETSVTVTGPPREPFGGRKH-LKAIDVNMNKMP--DVAMTLAVVALFADGPTAIR
      : . . . . : . : : : : : : : : : : : : : : : .
PDB: 2G ADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMN
      290      300      310      320      330      340

360      370      380      390      400
2MEPSP DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT--PPEK---LNVTAIDTYDDHR
      . . : : : . : . : : : . : : : : . : : : : . : :
PDB: 2G GLEELRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHR
      350      360      370      380      390      400

410      420      430      440
2MEPSP MAMAFS-LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : : . . : : : : : : : : : . : : : .
PDB: 2G IAMSFLVMGLVSENPNVTDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
      410      420      430      440      450

```

2PQC A | CP4 EPSPS LIGANDED WITH (R)-PHOSPHONATE TETRAHEDRAL REACTION INTERMEDIATE
 ANALOG | *AGROBACTERIUM SP.* | AA | 445

Length = 445

initn: 202 initl: 103 opt: 121 Z-score: 234.3 bits: 52.5 E(): 3.6e-06

Smith-Waterman score: 194; 24.146% identity (53.303% similar) in 439 aa overlap
 (15-440:13-436)

```

10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGKSLSNRILLAAALSEGTTVVDNLLNSEDVHYMLGAL
      . : : : : : : : : : : : : : : : : .
PDB: 2P SSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTGKAM
      10      20      30      40      50

70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . . : . . : : . : : : . : : : . : .
PDB: 2P QAMGARIRKEGDTWIIDGVGNGLLA-----PEAPLDFGNAATGCRLTMGLVGVDYDFDST
      60      70      80      90      100      110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . . : . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  FI--GDASLTKRPMGRVLNPLREMGVQVKSEDG--DRLPVTLRG--PKTPTPITYRVPMASA
                120          130          140          150          160

                190          200          210          220          230
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWD--RFYIKG
      : . . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  QVKSAVLLAGLNTPGITT--VIEPIMTRDHT--KMLQGFGANLTVETDADGVRTIRLE
        170          180          190          200          210          220

                240          250          260          270          280          290
2MEPSP  GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
      : . . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMN--PTRTGLILTQEMGADI
        230          240          250          260          270          280

                300          310          320          330          340          350
2MEPSP  TWTETSVTVTGPPREPFGRKH--LKAIDVNMNKM---DVAMTLAVVALFADGPTAIRDVA
      . . . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNGLE
        290          300          310          320          330          340

                360          370          380          390          400
2MEPSP  SWRVKETERMVAIRTELTKLGASVEEGPDYCIIT--PPEK----LNVTAIDTYDDHRMAM
      . : . . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  ELRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAM
        350          360          370          380          390          400

                410          420          430          440
2MEPSP  AFS--LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  SFLVMGLVSENVPVTVDDATMIATSFPEFMDLMAGLGAKIELS
        410          420          430          440

```

2PQB_A | CP4 EPSPS LIGANDED WITH (R)-DIFLUOROMETHYL TETRAHEDRAL INTERMEDIATE ANALOG | *AGROBACTERIUM SP.* | AA | 445

Length = 445

initn: 202 initl: 103 opt: 121 Z-score: 234.3 bits: 52.5 E(): 3.6e-06

Smith-Waterman score: 194; 24.146% identity (53.303% similar) in 439 aa overlap (15-440:13-436)

```

                10          20          30          40          50          60
2MEPSP  MAGAEIEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      . : . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  SSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTGKAM
        10          20          30          40          50

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . . : . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  QAMGARIRKEGDTWIIDGVGNGLLA-----PEAPLDFGNAATGCRLTMGLVGVDYDFDST
        60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . . : . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  FI--GDASLTKRPMGRVLNPLREMGVQVKSEGD--DRLPVTLRG--PKTPTPITYRVPMASA
                120          130          140          150          160

                190          200          210          220          230
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWD--RFYIKG
      : . . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  QVKSAVLLAGLNTPGITT--VIEPIMTRDHT--KMLQGFGANLTVETDADGVRTIRLE
        170          180          190          200          210          220

                240          250          260          270          280          290
2MEPSP  GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
      : . . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMN--PTRTGLILTQEMGADI
        230          240          250          260          270          280

                300          310          320          330          340          350
2MEPSP  TWTETSVTVTGPPREPFGRKH--LKAIDVNMNKM---DVAMTLAVVALFADGPTAIRDVA
      . . . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNGLE
        290          300          310          320          330          340

                360          370          380          390          400
2MEPSP  SWRVKETERMVAIRTELTKLGASVEEGPDYCIIT--PPEK----LNVTAIDTYDDHRMAM
      . : . . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  ELRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAM
        350          360          370          380          390          400

                410          420          430          440
2MEPSP  AFS-LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2P  SFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELS
        410          420          430          440

```

2GG4_A | CP4 EPSP SYNTHASE (UNLIGANDED) | *AGROBACTERIUM SP.* | AA | 455
Length = 455
initn: 202 initl: 103 opt: 121 Z-score: 234.2 bits: 52.5 E(): 3.7e-06
Smith-Waterman score: 194; 24.146% identity (53.303% similar) in 439 aa overlap
(15-440:18-441)

```

                10          20          30          40          50
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSKLSNRILLLAALSEGTTVVDNLLNSEDVHYML
      . : . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2G  MSHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
        10          20          30          40          50          60

                60          70          80          90          100          110
2MEPSP  GALRTLGLSVEADKAAKRAVVVGCCKGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxx
      . : . . . : . . : . . : . . : . . : . . : . . : . . : . . : . .
PDB: 2G  KAMQAMGARIRKEGDTWIIDGVNGGLLA-----PEAPLDFGNAATGCRLTMGLVGVDYDF
        70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

120      130      140      150      160      170
2MEPSP xxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGS
      . . . . : . . . . . : . . . . : : : : : : .
PDB: 2G DSTFI--GDASLTKRPMGRVLNPLREMGVQVKSEDG--DRLPVTLRG--PKTPTPITYRVPM
      120      130      140      150      160      170

```

```

180      190      200      210      220      230
2MEPSP ISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWD--RFY
      .: . . . . : . . . . . : . . . . : . . . . : :
PDB: 2G ASAQVKSAVLLAGLNTPGITT--VIEPIMTRDHT--KMLQGFGANLTVETDADGVRTI
      180      190      200      210      220

```

```

240      250      260      270      280      290
2MEPSP IKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMG
      . . . . : : : . . . . . . . . . . . . . . . : :
PDB: 2G RLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMN--PTRTGLILTQEMG
      230      240      250      260      270      280

```

```

300      310      320      330      340      350
2MEPSP AKVTWTETSVTVTGPPREPFGGRKH-LKAIDVNMNKMP--DVAMTLAVVALFADGPTAIR
      : . . . . . : . . . . : : : : : : : : : : : :
PDB: 2G ADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMN
      290      300      310      320      330      340

```

```

360      370      380      390      400
2MEPSP DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT--PPEK----LNVTAIDTYDDHR
      . . : . . . . : . . : : . . : : . . . . : : :
PDB: 2G GLEELRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHR
      350      360      370      380      390      400

```

```

410      420      430      440
2MEPSP MAMAFS-LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .: . . . : : : : : : : : : : :
PDB: 2G IAMSFLVMGLVSENPNVTDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
      410      420      430      440      450

```

2GGA_A | CP4 EPSP SYNTHASE LIGANDED WITH S3P AND GLYPHOSATE | *AGROBACTERIUM SP.* | AA | 455
Length = 455
initn: 202 initl: 103 opt: 121 Z-score: 234.2 bits: 52.5 E(): 3.7e-06
Smith-Waterman score: 194; 24.146% identity (53.303% similar) in 439 aa overlap (15-440:18-441)

```

10      20      30      40      50
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYML
      .: . . . . : : : . . . . . : : . . . . : :
PDB: 2G MSHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
      10      20      30      40      50      60

```

```

60      70      80      90      100      110
2MEPSP GALRTLGLSVEADKAAKRAVVVGCCKGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxx
      .: . . . : : : . : : : : : : : : : :
PDB: 2G KAMQAMGARIRKEGDTWIIDGVNGGLLA-----PEAPLDFGNAATGCRLTMGLVGVDYDF
      70      80      90      100      110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

120      130      140      150      160      170
2MEPSP xxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGS
      . . . . : . . . . . : . . . . : : : : : : .
PDB: 2G DSTFI--GDASLTKRPMGRVLNPLREMGVQVKSEGD--DRLPVTLRG--PKTPTPITYRVPM
      120      130      140      150      160      170

180      190      200      210      220      230
2MEPSP ISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWD--RFY
      . . : . . : . . . . . : . . . . : . . : :
PDB: 2G ASAQVKSAVLLAGLNTPGITT--VIEPIMTRDHT--KMLQGFGANLTVETDADGVRTI
      180      190      200      210      220

240      250      260      270      280      290
2MEPSP IKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMG
      . . . . : : : . . . . . : . . . . : . . . . :
PDB: 2G RLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNM--PTRTGLILTLQEMG
      230      240      250      260      270      280

300      310      320      330      340      350
2MEPSP AKVTWTETSVTVTGPPREPFGGRKH-LKAIDVNMNKMP--DVAMTLAVVALFADGPTAIR
      : . . . . : . : : : : : : : : : : : : :
PDB: 2G ADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMN
      290      300      310      320      330      340

360      370      380      390      400
2MEPSP DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT--PPEK---LNVTAIDTYDDHR
      . . : . . . : . : : . . : : . . . : : :
PDB: 2G GLEELRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHR
      350      360      370      380      390      400

410      420      430      440
2MEPSP MAMAFS-LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . . : . : : : : : . . . . .
PDB: 2G IAMSFLVMGLVSENPNVTDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
      410      420      430      440      450

```

2GG6_A | CP4 EPSP SYNTHASE LIGANDED WITH S3P | *AGROBACTERIUM SP.* | AA | 455
Length = 455
initn: 202 initl: 103 opt: 121 Z-score: 234.2 bits: 52.5 E(): 3.7e-06
Smith-Waterman score: 194; 24.146% identity (53.303% similar) in 439 aa overlap
(15-440:18-441)

```

10      20      30      40      50
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYML
      . . . . . : . . . . : : . . . . :
PDB: 2G MSHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
      10      20      30      40      50      60

60      70      80      90      100      110
2MEPSP  GALRTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxx
      . . . . : . : : . : : : . : : :
PDB: 2G KAMQAMGARIRKEGDTWIIDGVNGGLLA-----PEAPLDFGNAATGCRLTMGLVGVDYF
      70      80      90      100      110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

120      130      140      150      160      170
2MEPSP xxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGS
      . . . . : . . . . . : . . . . : : : : : : .
PDB: 2G DSTFI--GDASLTKRPMGRVLNPLREMGVQVKSEDG--DRLPVTLRG--PKTPTPITYRVMG
      120      130      140      150      160      170

180      190      200      210      220      230
2MEPSP ISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWD--RFY
      . : . . . : . . . . . : . . . . . : . . : :
PDB: 2G ASAQVKSAVLLAGLNTPGITT--VIEPIMTRDHTe--KMLQGFGANLTVETDADGVRTI
      180      190      200      210      220

240      250      260      270      280      290
2MEPSP IKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMG
      . : . . : : : . . . . . : . . . . . : . . . :
PDB: 2G RLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMN--PTRTGLILTQEMG
      230      240      250      260      270      280

300      310      320      330      340      350
2MEPSP AKVTWTETSVTVTGPPREPFGGRKH--LKAIDVNMNKMP--DVAMTLAVVALFADGPTAIR
      : . . . . : . : : : : : : : : : : : : :
PDB: 2G ADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMN
      290      300      310      320      330      340

360      370      380      390      400
2MEPSP DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT--PPEK----LNVTAIDTYDDHR
      . . : : : . : . : : . . : : . . . : : :
PDB: 2G GLEELRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHR
      350      360      370      380      390      400

410      420      430      440
2MEPSP MAMAFS--LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : : . . : : : : : : : : : . . . . .
PDB: 2G IAMSFLVMGLVSENPNVTDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
      410      420      430      440      450

```

1wy6_A | CRYSTAL STRUCTURE OF HYPOTHETICAL PROTEIN [ST1625P] FROM
HYPERTHERMOPHILIC ARCHAEON SULFOLOBUS TOKODAI | *SULFOLOBUS TOKODAI* | AA | 172

Length = 172

initn: 64 initl: 64 opt: 64 Z-score: 126.2 bits: 31.2 E(): 3.8

Smith-Waterman score: 64; 37.778% identity (60.000% similar) in 45 aa overlap
(38-82:40-84)

```

10      20      30      40      50      60
2MEPSP VLQPIKEISGTVKLPKSLSNRILLLAALSEGTTVDNLLNSEDVHYMLGALRTLGLSV
      . : . . : : : : : : : : : :
PDB: 1W KLMDAKKFLLDGYIDEGVKIVLEITKSSTKSEYNWFICNLLESIDCRYMFQVLDKIGSYF
      10      20      30      40      50      60

70      80      90      100      110      120
2MEPSP EADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVP
      . : . . : : :
PDB: 1W DLDKCQNLKSVVECGVINNTLNEHVNKALDILVIQGRDKLEEIGREILKNNEVSASILV
      70      80      90      100      110      120

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1UAE_A | STRUCTURE OF UDP-N-ACETYLGLUCOSAMINE ENOLPYRUVYL TRANSFERASE | *ESCHERICHIA COLI* | AA | 419
Length = 419
initn: 45 init1: 45 opt: 66 Z-score: 123.9 bits: 32.0 E(): 5.1
Smith-Waterman score: 68; 25.758% identity (63.636% similar) in 66 aa overlap
(7-71:352-417)

```

                                10      20      30
2MEPSP      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLAA
              .. . ....::. . . : ..: ..
PDB:1U FITETVFENRFMHVPELSRMGAHAEIESNTVICHGVEKLSGAQVMATDLRASASLVLAGC
              330      340      350      360      370      380

              40      50      60      70      80      90
2MEPSP LSEGTTVVDNLLNSEDVHYML-GALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQ
              ..:::..... . . . :::: ..: :
PDB:1U IAEGTTVVDRIYHIDRGYERIEDKLRALGANIERVKGE
              390      400      410

              100      110      120      130      140      150
2MEPSP LFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTD

```

1A2N_A | STRUCTURE OF THE C115A MUTANT OF MURA COMPLEXED WITH THE FLUORINATED ANALOG OF THE REACTION TETRAHEDRAL INTERMEDIATE | *ESCHERICHIA COLI* | AA | 419
Length = 419
initn: 45 init1: 45 opt: 66 Z-score: 123.9 bits: 32.0 E(): 5.1
Smith-Waterman score: 68; 25.758% identity (63.636% similar) in 66 aa overlap
(7-71:352-417)

```

                                10      20      30
2MEPSP      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLAA
              .. . ....::. . . : ..: ..
PDB:1A FITETVFENRFMHVPELSRMGAHAEIESNTVICHGVEKLSGAQVMATDLRASASLVLAGC
              330      340      350      360      370      380

              40      50      60      70      80      90
2MEPSP LSEGTTVVDNLLNSEDVHYML-GALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQ
              ..:::..... . . . :::: ..: :
PDB:1A IAEGTTVVDRIYHIDRGYERIEDKLRALGANIERVKGE
              390      400      410

              100      110      120      130      140      150
2MEPSP LFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTD

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

2J5U_A | MREC *LYSTERIA MONOCYTOGENES* | *LISTERIA MONOCYTOGENES* | AA | 255

Length = 255

initn: 44 initl: 44 opt: 63 Z-score: 121.4 bits: 30.8 E(): 7.1

Smith-Waterman score: 63; 32.394% identity (61.972% similar) in 71 aa overlap
 (16-86:120-186)

	10	20	30	40
2MEPSP	MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVD			
		:.:.:	:.:.:	:.:
PDB: 2J	DKGSSDGVKPDMAVTTT	PSGLIGKVTTT	GAKSATVELLTSSDVKNRV---	SAKVQGENAF
	90	100	110	120

	50	60	70	80	90	100
2MEPSP	NLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAM					
	...:	:	:	:	:	:
PDB: 2J	GIINGYDS	TKLLELKQLPYDMKFKKGQK-VVT	SGLGGKFPAGIFIGTIEKVETDKMGLS			
	150	160	170	180	190	200

	110	120	130	140	150	160
2MEPSP	RSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIG					
PDB: 2J	QTAFIKPGADMYDLNHVTVLKRSAGTDDDDTTSSD					
	210	220	230	240	250	

2J5U_B | MREC *LYSTERIA MONOCYTOGENES* | *LISTERIA MONOCYTOGENES* | AA | 255

Length = 255

initn: 44 initl: 44 opt: 63 Z-score: 121.4 bits: 30.8 E(): 7.1

Smith-Waterman score: 63; 32.394% identity (61.972% similar) in 71 aa overlap
 (16-86:120-186)

	10	20	30	40
2MEPSP	MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVD			
		:.:.:	:.:.:	:.:
PDB: 2J	DKGSSDGVKPDMAVTTT	PSGLIGKVTTT	GAKSATVELLTSSDVKNRV---	SAKVQGENAF
	90	100	110	120

	50	60	70	80	90	100
2MEPSP	NLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAM					
	...:	:	:	:	:	:
PDB: 2J	GIINGYDS	TKLLELKQLPYDMKFKKGQK-VVT	SGLGGKFPAGIFIGTIEKVETDKMGLS			
	150	160	170	180	190	200

	110	120	130	140	150	160
2MEPSP	RSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIG					
PDB: 2J	QTAFIKPGADMYDLNHVTVLKRSAGTDDDDTTSSD					
	210	220	230	240	250	

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

10XJ_A|CRYSTAL STRUCTURE OF THE SMAUG RNA BINDING DOMAIN|*DROSOPHILA MELANOGASTER*|AA|173

Length = 173

initn: 50 initl: 50 opt: 57 Z-score: 112.0 bits: 28.5 E(): 24

Smith-Waterman score: 57; 28.000% identity (60.000% similar) in 50 aa overlap (281-330:77-126)

	260	270	280	290	300	310
2MEPSP	EGDASSASYFLAGAAI	xxxxxxxxxxxxxx	SLQGDVKFAEVLEMMGAKVTWTETSVTVTGP			
			:.:.:.:.:.:	.:	:	::
PDB:1O	VTKGASHKLALCIDK	LKERANILNRVEQEL	LSGQME	LSTAVEELTNIVL	TPMKPLESPGP	
	50	60	70	80	90	100

	320	330	340	350	360	370
2MEPSP	PREPFGRKHLKAIDV	NMNKMPDVAMTLA	VVALFADGPTAIR	DVASWRVKETERM	VAIRTE	
	:.:	:.:	.:	::.:	:	.
PDB:1O	PEENIGLRFLKVIDI	VNTNLQQDPYAVQ	DDETLGVL	MWILDRSIHNEAF	MNHASQLKDLK	
	110	120	130	140	150	160

3C17_A|HEXAGONAL CRYSTAL STRUCTURE OF PRECURSOR *E. COLI* ISOASPARTYL PEPTIDASE/L-ASPARAGINASE (ECAIII) WITH ACTIVE- SITE T179A MUTATION|*ESCHERICHIA COLI*|AA|320

Length = 320

initn: 57 initl: 57 opt: 57 Z-score: 107.7 bits: 28.6 E(): 41

Smith-Waterman score: 57; 36.585% identity (70.732% similar) in 41 aa overlap (33-73:152-192)

	10	20	30	40	50	60
2MEPSP	GAEIIVLQPIKEISG	TVKLPGSKSL	SNRILLLAAL	SEGTTVVDNLL	NSEDVHYMLG	ALRT
				::::	:.:.:.:.:	:.:.:
PDB:3C	IGEGAENFAFARGM	ERVSP	EIFSTSLRYEQ	LLAARKEGATV	LDHSGAPLDEK	QKMGAVGA
	130	140	150	160	170	180

	70	80	90	100	110	120
2MEPSP	LGLSVEADKAAKRA	VVGCGGKFP	VEDAKEEVQL	FLGNAGIAMR	SLxxxxxxxxxxx	YV
	..:.:.:.:	::				
PDB:3C	VALDLGNLAAATST	GGMTNKL	PGRVGD	SPLVGAGCYAN	NASVAVSCTGT	GEVFIRALAA
	190	200	210	220	230	240

2ZAK_A|ORTHORHOMBIC CRYSTAL STRUCTURE OF PRECURSOR *E. COLI* ISOASPARTYL PEPTIDASE/L-ASPARAGINASE (ECAIII) WITH ACTIVE- SITE T179A MUTATION|*ESCHERICHIA COLI*|AA|320

Length = 320

initn: 57 initl: 57 opt: 57 Z-score: 107.7 bits: 28.6 E(): 41

Smith-Waterman score: 57; 36.585% identity (70.732% similar) in 41 aa overlap (33-73:152-192)

	10	20	30	40	50	60
2MEPSP	GAEIIVLQPIKEISG	TVKLPGSKSL	SNRILLLAAL	SEGTTVVDNLL	NSEDVHYMLG	ALRT
				::::	:.:.:.:.:	:.:.:
PDB:2Z	IGEGAENFAFARGM	ERVSP	EIFSTSLRYEQ	LLAARKEGATV	LDHSGAPLDEK	QKMGAVGA
	130	140	150	160	170	180

```

              70              80              90              100             110             120
2MEPSP  LGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxxxxxYV
      ..:..... ::
PDB: 2Z  VALDLDGNLAAATSTGGMTNKLPGRVGDSPLVGAGCYANNASVAVSCTGTGEVFI RALAA
              190              200              210              220             230             240

```

Smith-Waterman score: 57; 36.585% identity (70.732% similar) in 41 aa overlap (33-73:152-192)

[illegible]

```

              70              80              90              100             110             120
2MEPSP  LGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxxxxxYV
      ..:..... ::
PDB: 2Z  VALDLDGNLAAATSTGGMTNKLPGRVGDSPLVGAGCYANNASVAVSCTGTGEVFIKALAA
              190             200             210             220             230             240

```

Smith-Waterman score: 57; 36.585% identity (70.732% similar) in 41 aa overlap (33-73:152-192)

10 20 30 40 50 60
 2MEPSP GAEEIVLQPIKEISGTVKLPGSKSLSNRILLALLAALSEGTTVVDNLLNSEDVHYMLGALRT
 :::: : . : . : . : . :
 PDB:3C IGEGAENFAFARGMERVSPEIFSTSLRYEQLLAAKRGATVLHDHSGAPLDEKQKMGA VAG
 130 140 150 160 170 180

```

              70              80              90              100             110             120
2MEPSP  LGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxxYV
      ..:..... ::
PDB:3C  VALDLDGNLAAATSTGGMTNKLPGRVGDSPLVGAGCYANNASVAVSCTGTGEVFIKALAA
              190              200              210              220             230             240

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

2PZ5_A|CRYSTAL STRUCURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC N549T MUTATION RESPONSIBLE FOR PFEIFFER SYNDROME|*HOMO SAPIENS*|AA|324
Length = 324

initn: 35 initl: 35 opt: 57 Z-score: 107.6 bits: 28.6 E(): 42
Smith-Waterman score: 57; 32.051% identity (60.256% similar) in 78 aa overlap
(321-395:55-128)

	300	310	320	330	340	350
2MEPSP	LEMMGAKVTWTETSVTVTGPPREPFG	GRKHLKAIDVNMNKM	PDVAMTLAVVALFADGPTAI			
				.: .: .: .: .: .:	.: .: .: .: .: .:	.: .: .: .: .: .:
PDB:2P	PEDPKWEFPRDKLTLGKPLGEGAFGQV	MAEAVGIDKDK-PKEAVTVAVKMLKDDATE--				
	30	40	50	60	70	80

	360	370	380	390	400
2MEPSP	RDVASWRVKETERMVAI---RTELTKLGASVEEGPDYCIITPPEKLNVT	AIDTYDDHRMA			
	.: .: .: .: .: .:	.: .: .: .: .: .:	.: .: .: .: .: .:	.: .: .: .: .: .:	.: .: .: .: .: .:
PDB:2P	KDLSDL-VSEMEMMKMIGKHKNIITLLGACTQDGPLYVIVEYASKGNLREYLRARRPPGM				
	90	100	110	120	130

	410	420	430	440
2MEPSP	MAFSLAACAEVPVTIRDPGCTRKT	FPDYFDVLSTFVK		
PDB:2P	EYSYDINRVPEEQMTFKDLVSC	TYQLARGMEYLASQKCIHRDLAARNVLVTENNV	KIAD	
	150	160	170	180

2PZ5_B|CRYSTAL STRUCURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC N549T MUTATION RESPONSIBLE FOR PFEIFFER SYNDROME|*HOMO SAPIENS*|AA|324
Length = 324

initn: 35 initl: 35 opt: 57 Z-score: 107.6 bits: 28.6 E(): 42
Smith-Waterman score: 57; 32.051% identity (60.256% similar) in 78 aa overlap
(321-395:55-128)

	300	310	320	330	340	350
2MEPSP	LEMMGAKVTWTETSVTVTGPPREPFG	GRKHLKAIDVNMNKM	PDVAMTLAVVALFADGPTAI			
				.: .: .: .: .: .:	.: .: .: .: .: .:	.: .: .: .: .: .:
PDB:2P	PEDPKWEFPRDKLTLGKPLGEGAFGQV	MAEAVGIDKDK-PKEAVTVAVKMLKDDATE--				
	30	40	50	60	70	80

	360	370	380	390	400
2MEPSP	RDVASWRVKETERMVAI---RTELTKLGASVEEGPDYCIITPPEKLNVT	AIDTYDDHRMA			
	.: .: .: .: .: .:	.: .: .: .: .: .:	.: .: .: .: .: .:	.: .: .: .: .: .:	.: .: .: .: .: .:
PDB:2P	KDLSDL-VSEMEMMKMIGKHKNIITLLGACTQDGPLYVIVEYASKGNLREYLRARRPPGM				
	90	100	110	120	130

	410	420	430	440
2MEPSP	MAFSLAACAEVPVTIRDPGCTRKT	FPDYFDVLSTFVK		
PDB:2P	EYSYDINRVPEEQMTFKDLVSC	TYQLARGMEYLASQKCIHRDLAARNVLVTENNV	KIAD	
	150	160	170	180

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1DD4_D|CRYSTAL STRUCTURE OF RIBOSOMAL PROTEIN L12 FROM *THERMOTOGA*
 MARITIM|*THERMOTOGA MARITIMA*|AA|40

Length = 40

initn: 48 initl: 48 opt: 49 Z-score: 106.3 bits: 25.4 E(): 49
 Smith-Waterman score: 49; 40.000% identity (63.333% similar) in 30 aa overlap
 (196-225:4-33)

170	180	190	200	210	220
2MEPSP	GLPGGKVKLSGSISSQYxxxxxxxxxxxx	GDVEIEIIDKLISIPYVEMTLRLMERFGVKA			
		:	::	:::: .:::
PDB:1D		MTIDEIIEAIEKLTVSELAELVKKLEDKFGVTA			
		10	20	30	

230	240	250	260	270	280
2MEPSP	EHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxx	SLQGDV			

PDB:1D AAPVAVA
 40

1DD4_C|CRYSTAL STRUCTURE OF RIBOSOMAL PROTEIN L12 FROM *THERMOTOGA*
 MARITIM|*THERMOTOGA MARITIMA*|AA|40

Length = 40

initn: 48 initl: 48 opt: 49 Z-score: 106.3 bits: 25.4 E(): 49
 Smith-Waterman score: 49; 40.000% identity (63.333% similar) in 30 aa overlap
 (196-225:4-33)

170	180	190	200	210	220
2MEPSP	GLPGGKVKLSGSISSQYxxxxxxxxxxxx	GDVEIEIIDKLISIPYVEMTLRLMERFGVKA			
		:	::	:::: .:::
PDB:1D		MTIDEIIEAIEKLTVSELAELVKKLEDKFGVTA			
		10	20	30	

230	240	250	260	270	280
2MEPSP	EHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxx	SLQGDV			

PDB:1D AAPVAVA
 40

2CCY_A|STRUCTURE OF FERRICYTOCHROME C(PRIME) FROM *RHODOSPIRILLUM MOLISCHIANUM* AT
 1.67 ANGSTROMS RESOLUTION|*RHODOSPIRILLUM MOLISCHIANUM*|AA|128

Length = 128

initn: 51 initl: 51 opt: 52 Z-score: 104.1 bits: 26.6 E(): 65
 Smith-Waterman score: 52; 35.897% identity (61.538% similar) in 39 aa overlap
 (345-383:67-105)

320	330	340	350	360	370
2MEPSP	FGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKL				
		:	:	:	..: ..: ..: :::
PDB:2C	PADAAQRAENMAMVAKLAPIGWAKGTEALPNGETKPEAFGSKSAEFLEGWKALATESTKL				
	40	50	60	70	80

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          380          390          400          410          420          430
2MEPSP GASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKT FPD
      .: . . . . : : :
PDB: 2C AAAAKAGPDALKQA AATGKVCKACHEEFKQD
          100          110          120

```

2CCY_B | STRUCTURE OF FERRICYTOCHROME C (PRIME) FROM *RHODOSPIRILLUM MOLISCHIANUM* AT 1.67 ANGSTROMS RESOLUTION | *RHODOSPIRILLUM MOLISCHIANUM* | AA | 128
Length = 128
initn: 51 initl: 51 opt: 52 Z-score: 104.1 bits: 26.6 E(): 65
Smith-Waterman score: 52; 35.897% identity (61.538% similar) in 39 aa overlap (345-383:67-105)

```

          320          330          340          350          360          370
2MEPSP FGRKHLKAIDVNMNKM P DVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKL
      .: : . .: . . : .: : : :
PDB: 2C PADAAQRAENMAMVAKLAPIGWAKGTEALPNGETKPEAFGSKSAEFLEGWKALATESTKL
          40          50          60          70          80          90

```

```

          380          390          400          410          420          430
2MEPSP GASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKT FPD
      .: . . . . : : :
PDB: 2C AAAAKAGPDALKQA AATGKVCKACHEEFKQD
          100          110          120

```

2PZP_B | CRYSTAL STRUCURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC K526E MUTATION RESPONSIBLE FOR CROUZON SYNDROME | *HOMO SAPIENS* | AA | 324
Length = 324
initn: 30 initl: 30 opt: 55 Z-score: 103.6 bits: 27.9 E(): 70
Smith-Waterman score: 55; 32.051% identity (57.692% similar) in 78 aa overlap (321-395:55-128)

```

          300          310          320          330          340          350
2MEPSP LEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKM P DVAMTLAVVALFADGPTAI
      .: . . .: : : : : : : : :
PDB: 2P PEDPKWEFPRDKLT LGKPLGEGAFGQVVM AEAVGIDKDK-PKEAVTVAVKMLKDD---AT
          30          40          50          60          70          80

```

```

          360          370          380          390          400
2MEPSP RDVASWRVKETERMVAI---RTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMA
      .. : : : : : : .. . . : : : : : : : : : : :
PDB: 2P EEDLSDLVSEMEMMKMIGKHKNIINLLGACTQDGPLYVIVEYASKGNLREYLRARRPPGM
          90          100          110          120          130          140

```

```

          410          420          430          440
2MEPSP MAFSLAACAEVPVTIRDPGCTRKT FPDYFDVLSTFVKN
PDB: 2P EYSYDINRVPEEQMTFKDLV SCTYQLARGMEYLASQKCIHRDLAARNVLVTENNVMKIAD
          150          160          170          180          190          200

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

2PZP_A|CRYSTAL STRUCTURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC K526E MUTATION RESPONSIBLE FOR CROUZON SYNDROME|*HOMO SAPIENS*|AA|324
Length = 324

initn: 30 initl: 30 opt: 55 Z-score: 103.6 bits: 27.9 E(): 70
Smith-Waterman score: 55; 32.051% identity (57.692% similar) in 78 aa overlap
(321-395:55-128)

	300	310	320	330	340	350
2MEPSP	LEMMGAKVTWTETSVTVTGPPREPFG	GRKHLKAIDVNMNKM	PDVAMTLAVVALFADGPTAI			
					
PDB:2P	PEDPKWEFPRDKLTLGKPLGEGAFGQVMAEAVGIDKDK-PKEAVTVAVKMLKDD--AT					
	30	40	50	60	70	80

	360	370	380	390	400
2MEPSP	RDVASWRVKETERMVAI---RTELTKLGASVEEGPDYCIITPPEKLNVT	TAIDTYDDHRMA			
				
PDB:2P	EEDLSDLVSEMEMMKMIGKHKNIINLLGACTQDGPLYVIVEYASKGNLREYLRARRPPGM				
	90	100	110	120	130

	410	420	430	440
2MEPSP	MAFSLAACAEVPTIRDPGCTRKTFPDYFDVLSTFVKN			
PDB:2P	EYSYDINRVPEEQMTFKDLVSCITYQLARGMEYLASQKCIHRDLAARNVLVTENNVMKIAD			
	150	160	170	180

1KZQ_A|CRYSTAL STRUCTURE OF A PARASITE PROTEIN|*TOXOPLASMA GONDII*|AA|289
Length = 289

initn: 54 initl: 54 opt: 54 Z-score: 102.4 bits: 27.5 E(): 81
Smith-Waterman score: 54; 34.884% identity (53.488% similar) in 43 aa overlap
(62-104:219-261)

	40	50	60	70	80	90
2MEPSP	LLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVV	CGGKFPVEDAK				
					
PDB:1K	TGCNEKSFKDILPKLTENPWQGNASSDKGATLTIKKEAFPAESKSVIIGCTGG	SPEKHHK				
	190	200	210	220	230	240

	100	110	120	130	140	150
2MEPSP	EEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADV	DCF				
					
PDB:1K	TVKLEFAGAAGSAKSAAGTASHVSIFAMVIGLIGSIAACVA					
	250	260	270	280		

1KZQ_B|CRYSTAL STRUCTURE OF A PARASITE PROTEIN|*TOXOPLASMA GONDII*|AA|289
Length = 289

initn: 54 initl: 54 opt: 54 Z-score: 102.4 bits: 27.5 E(): 81
Smith-Waterman score: 54; 34.884% identity (53.488% similar) in 43 aa overlap
(62-104:219-261)

	40	50	60	70	80	90
2MEPSP	LLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVV	CGGKFPVEDAK				
					
PDB:1K	TGCNEKSFKDILPKLTENPWQGNASSDKGATLTIKKEAFPAESKSVIIGCTGG	SPEKHHK				
	190	200	210	220	230	240

100 110 120 130 140 150
 2MEPSP EEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPMRRERPIGDLVVGLKQLGADVDCF
 : : : : :
 PDB:1K TVKLEFAGAAGSAKSAAGTASHVSIFAMVIGLIGSIAACVA
 250 260 270 280

initn: 45 initl: 45 opt: 46 Z-score: 101.8 bits: 24.2 E(): 88
Smith-Waterman score: 46; 39.286% identity (64.286% similar) in 28 aa overlap
(196-223:4-31)

170 180 190 200 210 220
 2MEPSP GLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKA
 : :: :: : . . .
 PDB:1D MTIDEIIIEAIEKLTVSELAELVKKLEDKFQVT
 10 20 30

230 240 250 260 270 280
2MEPSP EHSDSWDRFYIKGGQYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDV

initn: 45 initl: 45 opt: 46 Z-score: 101.8 bits: 24.2 E(): 88
Smith-Waterman score: 46; 39.286% identity (64.286% similar) in 28 aa overlap
(196-223:4-31)

```

      170      180      190      200      210      220
2MEPSP GLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKA
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
PDB:1D                               MTIDEIEIEAIEKLTVSELAELVKKLEDKFGVT
                                10      20      30

```

2MEPSP EHSDSWDRFYIKGGOKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLOGDV

initn: 38 initl: 38 opt: 51 Z-score: 101.5 bits: 26.3 E(): 91
Smith-Waterman score: 51; 34.000% identity (54.000% similar) in 50 aa overlap
(339-388:31-77)

310 320 330 340 350 360
2MEPSP GPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIR
PDB:2I CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAKDGTLHARPVVSWFDQGTQTRDVIQLR
10 20 30 40 50 60

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          370          380          390          400          410          420
2MEPSP TELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCT
          :: :      ::. .:
PDB:2I I---AGGAIVWATPDHKVLTEYGWRAAGELRKGDRVAVRDVETGELRYSVIREVLPTRRA
          70          80          90          100          110

```

2CG4_A | STRUCTURE OF *E. COLI* ASNC | *ESCHERICHIA COLI* | AA | 152

Length = 152

initn: 50 initl: 50 opt: 51 Z-score: 100.9 bits: 26.3 E(): 99

Smith-Waterman score: 51; 66.667% identity (80.000% similar) in 15 aa overlap
(139-153:59-73)

```

          110          120          130          140          150          160
2MEPSP xxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLP
          :. .... :: :.:
PDB:2C AKQFGVSPETIHVRVEKMKQAGIITGARIDVSPKQLGYDVGCFIGIILKSAKDYP SALAK
          30          40          50          60          70          80

```

```

          170          180          190          200          210          220
2MEPSP GGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKA EHS
PDB:2C LESLDEVTEAYYTTGHYSIFIKVMCRSIDALQHVLINKIQTIDEIQSTETLIVLQNPIMR
          90          100          110          120          130          140

```

2CG4_B | STRUCTURE OF *E. COLI* ASNC | *ESCHERICHIA COLI* | AA | 152

Length = 152

initn: 50 initl: 50 opt: 51 Z-score: 100.9 bits: 26.3 E(): 99

Smith-Waterman score: 51; 66.667% identity (80.000% similar) in 15 aa overlap
(139-153:59-73)

```

          110          120          130          140          150          160
2MEPSP xxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLP
          :. .... :: :.:
PDB:2C AKQFGVSPETIHVRVEKMKQAGIITGARIDVSPKQLGYDVGCFIGIILKSAKDYP SALAK
          30          40          50          60          70          80

```

```

          170          180          190          200          210          220
2MEPSP GGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKA EHS
PDB:2C LESLDEVTEAYYTTGHYSIFIKVMCRSIDALQHVLINKIQTIDEIQSTETLIVLQNPIMR
          90          100          110          120          130          140

```

1MP5_A | Y177F VARIANT OF *S. ENTERICA* RMLA | *SALMONELLA ENTERICA* | AA | 292

Length = 292

initn: 28 initl: 28 opt: 53 Z-score: 100.3 bits: 27.1 E(): 1.1e+02

Smith-Waterman score: 53; 26.389% identity (58.333% similar) in 72 aa overlap
(353-424:135-202)

```

          330          340          350          360          370          380
2MEPSP IDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP
          : ..... :: ... . . :.:
PDB:1M CALVLGDNIFYGHDLPKLMEEAVNKESGATVFAYHVNDPERYGVVEFDQAGTAVSLEEKP
          110          120          130          140          150          160

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          390          400          410          420          430          440
2MEPSP  DYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKT FPDYFDVLSTF
          . : .. :... :.. . : : : :. . : :
PDB:1M  ----LQPKSNYAVTGLFFYDNSV VEMAKNLKPSARGELEITDINRIYMEQGRLSVAMMGR
          170          180          190          200          210          220

```

2MEPSP VKN

```

PDB:1M  GYAWLDTGTHQSLIEASNFIATIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPLSKND
          230          240          250          260          270          280

```

1MP3_B | L89T VARIANT OF *S. ENTERICA* RMLA | *SALMONELLA ENTERICA* | AA | 292
Length = 292
initn: 28 initl: 28 opt: 53 Z-score: 100.3 bits: 27.1 E(): 1.1e+02
Smith-Waterman score: 53; 26.389% identity (58.333% similar) in 72 aa overlap
(353-424:135-202)

```

          330          340          350          360          370          380
2MEPSP  IDVNMNKM P DVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP
                               : ..... : : ..... : : ..... :
PDB:1M  CALVLGDNIFYGHDLPKLMEAAVNKESGATVFAYHVNDPERYGVVEFDQAGTAVSLEEKP
          110          120          130          140          150          160

```

```

          390          400          410          420          430          440
2MEPSP  DYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKT FPDYFDVLSTF
          . : .. :... :.. . : : : :. . : :
PDB:1M  ----LQPKSNYAVTGLYFYDNSV VEMAKNLKPSARGELEITDINRIYMEQGRLSVAMMGR
          170          180          190          200          210          220

```

2MEPSP VKN

```

PDB:1M  GYAWLDTGTHQSLIEASNFIATIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPLSKND
          230          240          250          260          270          280

```

1MP4_A | W224H VARIANT OF *S. ENTERICA* RMLA | *SALMONELLA ENTERICA* | AA | 292
Length = 292
initn: 28 initl: 28 opt: 53 Z-score: 100.3 bits: 27.1 E(): 1.1e+02
Smith-Waterman score: 53; 26.389% identity (58.333% similar) in 72 aa overlap
(353-424:135-202)

```

          330          340          350          360          370          380
2MEPSP  IDVNMNKM P DVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP
                               : ..... : : ..... : : ..... :
PDB:1M  CALVLGDNIFYGHDLPKLMEAAVNKESGATVFAYHVNDPERYGVVEFDQAGTAVSLEEKP
          110          120          130          140          150          160

```

```

          390          400          410          420          430          440
2MEPSP  DYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKT FPDYFDVLSTF
          . : .. :... :.. . : : : :. . : :
PDB:1M  ----LQPKSNYAVTGLYFYDNSV VEMAKNLKPSARGELEITDINRIYMEQGRLSVAMMGR
          170          180          190          200          210          220

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

2MEPSP VKN

PDB:1M GYAHLDTGTHQSLIEASNFIAETIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPLSKND
230 240 250 260 270 280

1MP5_C|Y177F VARIANT OF *S. ENTERICA* RMLA|*SALMONELLA ENTERICA*|AA|292

Length = 292

initn: 28 initl: 28 opt: 53 Z-score: 100.3 bits: 27.1 E(): 1.1e+02

Smith-Waterman score: 53; 26.389% identity (58.333% similar) in 72 aa overlap
(353-424:135-202)

330	340	350	360	370	380
2MEPSP IDVNMNKM	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP				
			: ::
PDB:1M CALVLGDNIFYGHDLPKLMEAAVNKESGATVFAYHVNDPERYGVVEFDQAGTAVSLEEKP					
110	120	130	140	150	160

390	400	410	420	430	440
2MEPSP DYCIIITPPEKLNVT	AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTF				
. : . .	: : : . .	. : . .	. : . .	. : . .
PDB:1M ----LQPKSNYAVTGLFFYDNSVVEAMAKNLKPSARGELEITDINRIYMEQGRLSVAMMGR					
170	180	190	200	210	220

2MEPSP VKN

PDB:1M GYAWLDTGTHQSLIEASNFIAETIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPLSKND
230 240 250 260 270 280

1MP5_D|Y177F VARIANT OF *S. ENTERICA* RMLA|*SALMONELLA ENTERICA*|AA|292

Length = 292

initn: 28 initl: 28 opt: 53 Z-score: 100.3 bits: 27.1 E(): 1.1e+02

Smith-Waterman score: 53; 26.389% identity (58.333% similar) in 72 aa overlap
(353-424:135-202)

330	340	350	360	370	380
2MEPSP IDVNMNKM	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP				
			: ::
PDB:1M CALVLGDNIFYGHDLPKLMEAAVNKESGATVFAYHVNDPERYGVVEFDQAGTAVSLEEKP					
110	120	130	140	150	160

390	400	410	420	430	440
2MEPSP DYCIIITPPEKLNVT	AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTF				
. : . .	: : : . .	. : . .	. : . .	. : . .
PDB:1M ----LQPKSNYAVTGLFFYDNSVVEAMAKNLKPSARGELEITDINRIYMEQGRLSVAMMGR					
170	180	190	200	210	220

2MEPSP VKN

PDB:1M GYAWLDTGTHQSLIEASNFIAETIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPLSKND
230 240 250 260 270 280

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1MP5_B|Y177F VARIANT OF *S. ENTERICA* RMLA|*SALMONELLA ENTERICA*|AA|292

Length = 292

initn: 28 initl: 28 opt: 53 Z-score: 100.3 bits: 27.1 E(): 1.1e+02

Smith-Waterman score: 53; 26.389% identity (58.333% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMN	KMPDVAM	TLAVVAL	FADGPTA	IRDVASW	RVKETERM
						VAIRTELTK
						LGASVEEGP
						: :: : :
PDB:1M	CALVLGD	NIFYGHD	LPKLME	AAVNKE	SGATVF	FAYHVND
						PERYGVVE
						FDQAGTAV
						SLEEKP
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITP	PPEKLN	VTAIDTY	DDHRMA	MAFSLA	ACAEVP
						VTIRDPG
						CTRKTFP
						DPDYFDV
						LSTF
						. : . . : . . : . . : . : : . . : :
PDB:1M	----	LQPKS	NYAVT	GLFFY	DNSV	EMAKNL
						KPSARGE
						LEITDIN
						RIYMEQ
						GRLSVAM
						MGR
	170	180	190	200	210	220

2MEPSP VKN

	230	240	250	260	270	280
PDB:1M	GYAWLD	TGTHQS	LIEASN	FIATIE	ERQGLK	VSCPEE
						IAFRKN
						FINAQQ
						VIELAG
						PLSKND

1MP4_B|W224H VARIANT OF *S. ENTERICA* RMLA|*SALMONELLA ENTERICA*|AA|292

Length = 292

initn: 28 initl: 28 opt: 53 Z-score: 100.3 bits: 27.1 E(): 1.1e+02

Smith-Waterman score: 53; 26.389% identity (58.333% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMN	KMPDVAM	TLAVVAL	FADGPTA	IRDVASW	RVKETERM
						VAIRTELTK
						LGASVEEGP
						: :: : :
PDB:1M	CALVLGD	NIFYGHD	LPKLME	AAVNKE	SGATVF	FAYHVND
						PERYGVVE
						FDQAGTAV
						SLEEKP
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITP	PPEKLN	VTAIDTY	DDHRMA	MAFSLA	ACAEVP
						VTIRDPG
						CTRKTFP
						DPDYFDV
						LSTF
						. : . . : . . : . . : . : : . . : :
PDB:1M	----	LQPKS	NYAVT	GLYFY	DNSV	EMAKNL
						KPSARGE
						LEITDIN
						RIYMEQ
						GRLSVAM
						MGR
	170	180	190	200	210	220

2MEPSP VKN

	230	240	250	260	270	280
PDB:1M	GYAHLDT	GTGTHQ	SLIEASN	FIATIE	ERQGLK	VSCPEE
						IAFRKN
						FINAQQ
						VIELAG
						PLSKND

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1MP3_A|L89T VARIANT OF *S. ENTERICA* RMLA|*SALMONELLA ENTERICA*|AA|292

Length = 292

initn: 28 initl: 28 opt: 53 Z-score: 100.3 bits: 27.1 E(): 1.1e+02

Smith-Waterman score: 53; 26.389% identity (58.333% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNKM	PDVAMTLAV	VALFADG	PTAIRDVAS	WRVKETERM	VAIRTELTKLGASVEEGP
				:	::
PDB:1M	CALVLGDN	IFYGHDLP	KLMEAAVN	KESGATVF	FAYHVNDP	ERYGVVEFDQAGTAVSLEEK
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITPPE	KLNVTAID	TYDDHRMA	MAFSLAAC	AEVPTIRDP	GCTRKTFPDYFDVLSTF
	. : . .	::. .	::. .	. :: .	: . .	: . .
PDB:1M	----LQPK	SNYAVTGL	YFYDNSV	EMAKNLK	PSARGELE	ITDINRIYMEQGRLSVAMMGR
	170	180	190	200	210	220

2MEPSP VKN

PDB:1M	GYAWLDTG	THQSLIEA	SNFIAITIE	ERQGLKV	SCPEEIAFR	KNFINAQQVIELAGPLSKND
	230	240	250	260	270	280

1ZUU_A|CRYSTAL STRUCTURE OF THE YEAST BZZ1 FIRST SH3 DOMAIN AT 0.97-Å RESOLUTION|*SACCHAROMYCES CEREVISIAE*|AA|58

Length = 58

initn: 37 initl: 37 opt: 47 Z-score: 99.6 bits: 24.7 E(): 1.2e+02

Smith-Waterman score: 47; 40.741% identity (66.667% similar) in 27 aa overlap (374-400:7-33)

	350	360	370	380	390	400
2MEPSP	ADGPTAIR	DVASWRVK	ETERMVAIR	TELTKLGAS	VEEGPDYCI	ITPPEKLNVT
				: : . .	: : . .	: : . .
PDB:1Z				GMENKVL	YAYVQKDD	EITITPGDKISLVARDTGSG
				10	20	30

	410	420	430	440
2MEPSP	HRMAFSLAAC	AEVPTIRDP	GCTRKTFPDY	FDVLSTFVKN

PDB:1Z	WTKINND	TTGETGLV	PTTYIRI
	40	50	

2Q0B_A|CRYSTAL STRUCURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC E565A MUTATION RESPONSIBLE FOR PFEIFFER SYNDROME|*HOMO SAPIENS*|AA|324

Length = 324

initn: 30 initl: 30 opt: 53 Z-score: 99.5 bits: 27.1 E(): 1.2e+02

Smith-Waterman score: 53; 30.769% identity (61.538% similar) in 78 aa overlap (321-395:55-128)

	300	310	320	330	340	350
2MEPSP	LEMMGAKV	TWTETSV	TGPPREPF	GRKHLKAID	VNMNKM	PDVAMTLAV
			
PDB:2Q	PEDPKWE	FPRDKLT	LGLKPLGE	GAFGQV	MAEAVGID	KDK-PKEAVTVAVKMLKDDATE--
	30	40	50	60	70	80

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                360          370          380          390          400
2MEPSP RDVASWRVKETERMVAI---RTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMA
      . . . . . : : : : : : . . . . : : : : : : : : : : : :
PDB: 2Q KDLSDL-VSEMEMMKMIGKHKNIINLLGACTQDGPLYVIVAYASKGNLREYLRARRPPGM
                90          100          110          120          130          140

```

```

                410          420          430          440
2MEPSP MAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN

PDB: 2Q EYSYDINRVPEEQMTFKDLVSCITYQLARGMEYLASQKCIHRDLAARNVLVTENNVMKIAD
                150          160          170          180          190          200

```

2Q0B_B | CRYSTAL STRUCURE OF FGF RECEPTOR 2 (FGFR2) KINASE DOMAIN HARBORING THE PATHOGENIC E565A MUTATION RESPONSIBLE FOR PFEIFFER SYNDROME | *HOMO SAPIENS* | AA | 324
Length = 324
initn: 30 initl: 30 opt: 53 Z-score: 99.5 bits: 27.1 E(): 1.2e+02
Smith-Waterman score: 53; 30.769% identity (61.538% similar) in 78 aa overlap (321-395:55-128)

```

                300          310          320          330          340          350
2MEPSP LEMMGAKVTWTETSVTVTGPPREPFGFRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAI
      . . . . . : : : : : : : : : : : : : : : : :
PDB: 2Q PEDPKWEFPRDKLTGLGKPLGEGAFGQVMAEAVGIDKDK-PKEAVTVAVKMLKDDATE--
                30          40          50          60          70          80

```

```

                360          370          380          390          400
2MEPSP RDVASWRVKETERMVAI---RTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMA
      . . . . . : : : : : : . . . . : : : : : : : : : : : :
PDB: 2Q KDLSDL-VSEMEMMKMIGKHKNIINLLGACTQDGPLYVIVAYASKGNLREYLRARRPPGM
                90          100          110          120          130          140

```

```

                410          420          430          440
2MEPSP MAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN

PDB: 2Q EYSYDINRVPEEQMTFKDLVSCITYQLARGMEYLASQKCIHRDLAARNVLVTENNVMKIAD
                150          160          170          180          190          200

```

1A77_A | FLAP ENDONUCLEASE-1 FROM *METHANOCOCCUS JANNASCHII* | *METHANOCOCCUS JANNASCHII* | AA | 326
Length = 326
initn: 49 initl: 49 opt: 53 Z-score: 99.5 bits: 27.1 E(): 1.2e+02
Smith-Waterman score: 53; 46.667% identity (66.667% similar) in 30 aa overlap (136-165:206-234)

```

                110          120          130          140          150          160
2MEPSP RSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIG
      . . . . . : : : : : : : : : : : : : : : : :
PDB: 1A ALLYGAPRVVRNLTTTKEMPELIELNEVLEDLRISLDDL-IDIAIFMGTDYNPGGVKGIG
                180          190          200          210          220          230

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

PDB:2G SKSLTPFEQELLALKEAGDYMSFLRRVQLERVIVVAGETGSGKTTLMKALMQEIPFDQR
150 160 170 180 190 200

2GZA_C | CRYSTAL STRUCTURE OF THE VIRB11 ATPASE FROM THE *BRUCELLA SUIIS* TYPE IV SECRETION SYSTEM IN COMPLEX WITH SULPHATE | *BRUCELLA SUIIS* | AA | 361

Length = 361

initn: 53 initl: 53 opt: 53 Z-score: 98.8 bits: 27.2 E(): 1.3e+02

Smith-Waterman score: 53; 55.556% identity (77.778% similar) in 18 aa overlap (418-435:115-132)

390 400 410 420 430 440
2MEPSP TPPEKLNVTADTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
. :. :.

PDB:2G ISDSRPVLSAILPMGERIQIVRPPACEHGTISVTIRKPSFTRRTLEDYAQQGFFKHVRPM
90 100 110 120 130 140

PDB:2G SKSLTPFEQELLALKEAGDYMSFLRRVQLERVIVVAGETGSGKTTLMKALMQEIPFDQR
150 160 170 180 190 200

1J1N_A | STRUCTURE ANALYSIS OF ALGQ2, A MACROMOLECULE (ALGINATE)- BINDING PERIPLASMIC PROTEIN OF *SPHINGOMONAS* SP. A1., COMPLEXED WITH AN ALGINATE TETRASACCHARIDE | *SPHINGOMONAS* SP. | AA | 492

Length = 492

initn: 46 initl: 46 opt: 54 Z-score: 98.6 bits: 27.6 E(): 1.3e+02

Smith-Waterman score: 54; 28.571% identity (67.347% similar) in 49 aa overlap (216-263:193-241)

190 200 210 220 230 240
2MEPSP xxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ-KYKS
:.. :..... :.. :.. :.. :.

PDB:1J VLKAFKEKDPNGNGKADEVFPIDRHPDEVFRLVNFVGARSSGSDNYMDFYIDNGRVKHPW
170 180 190 200 210 220

250 260 270 280 290 300
2MEPSP PKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETS
... :. :.. :

PDB:1J AETAFRDGMKHVAQWYKEGLIDKEIFTRKAKAREQMFGGNLGGFTHDWFASTMTFNEGLA
230 240 250 260 270 280

1J1N_B | STRUCTURE ANALYSIS OF ALGQ2, A MACROMOLECULE (ALGINATE)- BINDING PERIPLASMIC PROTEIN OF *SPHINGOMONAS* SP. A1., COMPLEXED WITH AN ALGINATE TETRASACCHARIDE | *SPHINGOMONAS* SP. | AA | 492

Length = 492

initn: 46 initl: 46 opt: 54 Z-score: 98.6 bits: 27.6 E(): 1.3e+02

Smith-Waterman score: 54; 28.571% identity (67.347% similar) in 49 aa overlap (216-263:193-241)

190 200 210 220 230 240
2MEPSP xxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ-KYKS
:.. :..... :.. :.. :.. :.

PDB:1J VLKAFKEKDPNGNGKADEVFPIDRHPDEVFRLVNFVGARSSGSDNYMDFYIDNGRVKHPW
170 180 190 200 210 220

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      250      260      270      280      290      300
2MEPSP PKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETS
      ... .:      :... :
PDB:1J AETAFRDGMKHVAQWYKEGLIDKEIFTRKAKAREQMFGGNLGGFTHDWFASMTTFNEGLA
      230      240      250      260      270      280

```

1KWH_A | STRUCTURE ANALYSIS ALGQ2, A MACROMOLECULE (ALGINATE)-BINDING PERIPLASMIC PROTEIN OF *SPHINGOMONAS* SP. A1. | *SPHINGOMONAS* SP. | AA | 492
Length = 492
initn: 46 initl: 46 opt: 54 Z-score: 98.6 bits: 27.6 E(): 1.3e+02
Smith-Waterman score: 54; 28.571% identity (67.347% similar) in 49 aa overlap (216-263:193-241)

```

      190      200      210      220      230      240
2MEPSP xxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ-KYKS
      :... :..... :... :... :.
PDB:1K VLKAFKEKDPNGNGKADEVFIDRHPDEVFRLVNFVGARSSGSDNYMDFYIDNGRVKHPW
      170      180      190      200      210      220

```

```

      250      260      270      280      290      300
2MEPSP PKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETS
      ... .:      :... :
PDB:1K AETAFRDGMKHVAQWYKEGLIDKEIFTRKAKAREQMFGGNLGGFTHDWFASMTTFNEGLA
      230      240      250      260      270      280

```

1IIN_D | THYMIDYLTRANSFERASE COMPLEXED WITH UDP-GLUCOSE | *SALMONELLA ENTERICA* | AA | 292
Length = 292
initn: 28 initl: 28 opt: 52 Z-score: 98.3 bits: 26.7 E(): 1.4e+02
Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

```

      330      340      350      360      370      380
2MEPSP IDVNMNKMPPVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP
      : ..... :... : ..... :
PDB:1I CALVLGDNIFYGHDLPKLMEEAVNKEGATVFAYHVNDPERYGVVEFDQKGTAVSLEEKP
      110      120      130      140      150      160

```

```

      390      400      410      420      430      440
2MEPSP DYCIIPTPEKLNVTADTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTF
      . : .. :... :.. . : : : :. . : :
PDB:1I ----LQPKSNYAVTGLYFYDNSVEMAKNLKPSARGELEITDINRIYMEQGRLSVAMMGR
      170      180      190      200      210      220

```

2MEPSP VKN

```

PDB:1I GYAWLDTGTHQSLIEASNFATIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPLSKND
      230      240      250      260      270      280

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1IIN_A | THYMIDYLYLTRANSFERASE COMPLEXED WITH UDP-GLUCOSE | *SALMONELLA ENTERICA* | AA | 292

Length = 292

initn: 28 initl: 28 opt: 52 Z-score: 98.3 bits: 26.7 E(): 1.4e+02

Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP					
				: :: :
PDB:1I	CALVLGDNIFYGHDLPKLMEEAAVNKESGATVFAYHVNDPERYGVVEFDQKGTAVSLEEKP					
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTF					
	. : . . : . . . : . : . : . : . : . : . :					
PDB:1I	----LQPKSNYAVTGLYFYDNSVVEAMAKNLKPSARGELEITDINRIYMEQGRLSVAMMGR					
	170	180	190	200	210	220

2MEPSP VKN

PDB:1I	GYAWLDTGTHQSLIEASNFATIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPLSKND
	230 240 250 260 270 280

1IIM_B | THYMIDYLYLTRANSFERASE COMPLEXED WITH TTP | *SALMONELLA ENTERICA* | AA | 292

Length = 292

initn: 28 initl: 28 opt: 52 Z-score: 98.3 bits: 26.7 E(): 1.4e+02

Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP					
				: :: :
PDB:1I	CALVLGDNIFYGHDLPKLMEEAAVNKESGATVFAYHVNDPERYGVVEFDQKGTAVSLEEKP					
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTF					
	. : . . : . . . : . : . : . : . : . : . :					
PDB:1I	----LQPKSNYAVTGLYFYDNSVVEAMAKNLKPSARGELEITDINRIYMEQGRLSVAMMGR					
	170	180	190	200	210	220

2MEPSP VKN

PDB:1I	GYAWLDTGTHQSLIEASNFATIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPLSKND
	230 240 250 260 270 280

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1IIN_C | THYMIDYLYLTRANSFERASE COMPLEXED WITH UDP-GLUCOSE | *SALMONELLA ENTERICA* | AA | 292

Length = 292

initn: 28 initl: 28 opt: 52 Z-score: 98.3 bits: 26.7 E(): 1.4e+02

Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP					
				: :: :
PDB:1I	CALVLGDNIFYGHDLPKLMEEAAVNKESGATVFAYHVNDPERYGVVEFDQKGTAVSLEEKP					
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTF					
	. : . . : . . . : . : . : . : . : . : .					
PDB:1I	----LQPKSNYAVTGLYFYDNSVVEMAKNLKPSARGELEITDINRIYMEQGRLSVAMMGR					
	170	180	190	200	210	220

2MEPSP VKN

PDB:1I	GYAWLDTGTHQSLIEASNFATIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPLSKND
	230 240 250 260 270 280

1IIM_A | THYMIDYLYLTRANSFERASE COMPLEXED WITH TTP | *SALMONELLA ENTERICA* | AA | 292

Length = 292

initn: 28 initl: 28 opt: 52 Z-score: 98.3 bits: 26.7 E(): 1.4e+02

Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP					
				: :: :
PDB:1I	CALVLGDNIFYGHDLPKLMEEAAVNKESGATVFAYHVNDPERYGVVEFDQKGTAVSLEEKP					
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTF					
	. : . . : . . . : . : . : . : . : . : .					
PDB:1I	----LQPKSNYAVTGLYFYDNSVVEMAKNLKPSARGELEITDINRIYMEQGRLSVAMMGR					
	170	180	190	200	210	220

2MEPSP VKN

PDB:1I	GYAWLDTGTHQSLIEASNFATIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPLSKND
	230 240 250 260 270 280

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1IIN_B | THYMIDYLYLTRANSFERASE COMPLEXED WITH UDP-GLUCOSE | *SALMONELLA ENTERICA* | AA | 292

Length = 292

initn: 28 initl: 28 opt: 52 Z-score: 98.3 bits: 26.7 E(): 1.4e+02

Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNKM	PDVAMTLAV	VALFADGPTA	IRDVASWRV	KETERMVA	IRTELTKLGASVEEGP
				:
PDB:1I	CALVLGDN	IFYGHDLP	KLMEAAVN	KESGATVF	FAYHVNDP	ERYGVVEFDQKGTAVSLEEKP
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITPPE	KLNVTAID	TYDDHRMA	MAFSLAAC	AEVPVTIR	DPGCTRKT
	.	:	:	:
PDB:1I	----	LQPKSNYA	VTGLYFYD	NSVVEMA	KNLKPSAR	GELEITDINRIYMEQGRLSVAMMGR
	170	180	190	200	210	220

2MEPSP VKN

PDB:1I	GYAWLDTG	THQSLIEA	SNFIATIE	ERQGLKV	SCPEEIAF	RKNFINAQ
	230	240	250	260	270	280

1H5T_D | THYMIDYLYLTRANSFERASE COMPLEXED WITH THYMIDYLYLDIPHOSPHATE-GLUCOSE | AA | 293

Length = 293

initn: 28 initl: 28 opt: 52 Z-score: 98.2 bits: 26.8 E(): 1.4e+02

Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNKM	PDVAMTLAV	VALFADGPTA	IRDVASWRV	KETERMVA	IRTELTKLGASVEEGP
				:
PDB:1H	CALVLGDN	IFYGHDLP	KLMEAAVN	KESGATVF	FAYHVNDP	ERYGVVEFDKNGTAISLEEKP
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITPPE	KLNVTAID	TYDDHRMA	MAFSLAAC	AEVPVTIR	DPGCTRKT
	.	:	:	:
PDB:1H	----	LEPKSNYA	VTGLYFYD	NDVVQMA	KNLKPSAR	GELEITDINRIYLEQGRLSVAMMGR
	170	180	190	200	210	220

2MEPSP VKN

PDB:1H	GYAWLDTG	THQSLIEA	SNFIATIE	ERQGLKV	SCPEEIAF	RKGFIDVEQVRKLAVPLIKNN
	230	240	250	260	270	280

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1H5R_B | THYMIDYLYLTRANSFERASE COMPLEXED WITH THIMIDINE AND GLUCOSE- 1-PHOSPHATE | |AA| 293
Length = 293
initn: 28 init1: 28 opt: 52 Z-score: 98.2 bits: 26.8 E(): 1.4e+02
Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP					
				: ::	:
PDB:1H	CALVLGDNIFYGHDLPKLMEEAVNKESGATVFAYHVNDPERYGVVEFDKNGTAISLEEKP					
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTF					
	. : . . : . . : . . . : . . : . . : . . : . .					
PDB:1H	----LEPKSNYAVTGLYFYDNDVVQMAKNLKPSARGELEITDINRIYLEQGRLSVALMGR					
	170	180	190	200	210	220

2MEPSP VKN

PDB:1H	GYAWLDTGTHQSLIEASNFATIEERQGLKVSCPEEIAFRKGFIDVEQVRKLAVPLIKNN					
	230	240	250	260	270	280

1H5T_C | THYMIDYLYLTRANSFERASE COMPLEXED WITH THYMIDYLYLDIPHOSPHATE-GLUCOSE | |AA| 293
Length = 293
initn: 28 init1: 28 opt: 52 Z-score: 98.2 bits: 26.8 E(): 1.4e+02
Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP					
				: ::	:
PDB:1H	CALVLGDNIFYGHDLPKLMEEAVNKESGATVFAYHVNDPERYGVVEFDKNGTAISLEEKP					
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTF					
	. : . . : . . : . . . : . . : . . : . . : . .					
PDB:1H	----LEPKSNYAVTGLYFYDNDVVQMAKNLKPSARGELEITDINRIYLEQGRLSVAMMGR					
	170	180	190	200	210	220

2MEPSP VKN

PDB:1H	GYAWLDTGTHQSLIEASNFATIEERQGLKVSCPEEIAFRKGFIDVEQVRKLAVPLIKNN					
	230	240	250	260	270	280

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1H5S_D|THYMIDYLYLTRANSFERASE COMPLEXED WITH TMP||AA|293

Length = 293

initn: 28 initl: 28 opt: 52 Z-score: 98.2 bits: 26.8 E(): 1.4e+02

Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNKM	PDVAMTLAV	VALFADG	PTAIRDVAS	WRVKETERM	VAIRTELTKLGASVEEGP
				:	:: :
PDB:1H	CALVLGDN	IFYGHDLP	KLMEAAVN	KESGATVF	FAYHVNDP	ERYGVVEFDKNGTAISLEEK
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITP	PPEKLNVT	AIDTYDD	HRMAMAF	SLAACAEV	PVTIRDPGCTRKT
	. : . .	: . . .	: . .	. : . .	. : . .	. : . .
PDB:1H	----	LEPKSN	YAVTGLY	FYDNDV	VQMAKNL	KPSARGELEITDINRIYLEQGRLSVALMGR
		170	180	190	200	210 220

2MEPSP VKN

PDB:1H	GYAWLDTG	THQSLIE	ASNFIAT	IEERQGLK	VSCPEE	IAFRKGFIDVEQVRKLAVPLIKNN
	230	240	250	260	270	280

1H5S_B|THYMIDYLYLTRANSFERASE COMPLEXED WITH TMP||AA|293

Length = 293

initn: 28 initl: 28 opt: 52 Z-score: 98.2 bits: 26.8 E(): 1.4e+02

Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNKM	PDVAMTLAV	VALFADG	PTAIRDVAS	WRVKETERM	VAIRTELTKLGASVEEGP
				:	:: :
PDB:1H	CALVLGDN	IFYGHDLP	KLMEAAVN	KESGATVF	FAYHVNDP	ERYGVVEFDKNGTAISLEEK
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITP	PPEKLNVT	AIDTYDD	HRMAMAF	SLAACAEV	PVTIRDPGCTRKT
	. : . .	: . . .	: . .	. : . .	. : . .	. : . .
PDB:1H	----	LEPKSN	YAVTGLY	FYDNDV	VQMAKNL	KPSARGELEITDINRIYLEQGRLSVALMGR
		170	180	190	200	210 220

2MEPSP VKN

PDB:1H	GYAWLDTG	THQSLIE	ASNFIAT	IEERQGLK	VSCPEE	IAFRKGFIDVEQVRKLAVPLIKNN
	230	240	250	260	270	280

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1H5S_A | THYMIDYLYLTRANSFERASE COMPLEXED WITH TMP | |AA| 293

Length = 293

initn: 28 initl: 28 opt: 52 Z-score: 98.2 bits: 26.8 E(): 1.4e+02

Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP					
				: :: : :
PDB:1H	CALVLGDNIFYGHDLPKLMEEAVNKESGATVFAYHVNDPERYGVVEFDKNGTAISLEEKP					
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTF					
	. : . . : . . . : . . : . : . : . : . : .					
PDB:1H	----LEPKSNYAVTGlyFYDNDVVLMAKNLKPSARGELEITDINRIYLEQGRLSVALMGR					
	170	180	190	200	210	220

2MEPSP VKN

PDB:1H GYAWLDTGTHQSLIEASNFATIEERQGLKVSCPEEIAFRKGFIDVEQVRKLAVPLIKNN

	230	240	250	260	270	280
--	-----	-----	-----	-----	-----	-----

1H5R_C | THYMIDYLYLTRANSFERASE COMPLEXED WITH THIMIDINE AND GLUCOSE- 1-PHOSPATE | |AA| 293

Length = 293

initn: 28 initl: 28 opt: 52 Z-score: 98.2 bits: 26.8 E(): 1.4e+02

Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap (353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGP					
				: :: : :
PDB:1H	CALVLGDNIFYGHDLPKLMEEAVNKESGATVFAYHVNDPERYGVVEFDKNGTAISLEEKP					
	110	120	130	140	150	160

	390	400	410	420	430	440
2MEPSP	DYCIITPPEKLNVTALDITYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTF					
	. : . . : . . . : . . : . : . : . : . : .					
PDB:1H	----LEPKSNYAVTGlyFYDNDVVMQMAKNLKPSARGELEITDINRIYLEQGRLSVAMMGR					
	170	180	190	200	210	220

2MEPSP VKN

PDB:1H GYAWLDTGTHQSLIEASNFATIEERQGLKVSCPEEIAFRKGFIDVEQVRKLAVPLIKNN

	230	240	250	260	270	280
--	-----	-----	-----	-----	-----	-----

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

1H5S_C | THYMIDYLYLTRANSFERASE COMPLEXED WITH TMP | |AA|293

Length = 293

initn: 28 initl: 28 opt: 52 Z-score: 98.2 bits: 26.8 E(): 1.4e+02

Smith-Waterman score: 52; 26.389% identity (56.944% similar) in 72 aa overlap
(353-424:135-202)

	330	340	350	360	370	380
2MEPSP	IDVNMNKM	PDVAMTLA	VVALFADG	PTAIRDVA	SWRVKETER	MVAIRTELTKLGASVEEGP
				:
PDB:1H	CALVLGDN	IFYGHDLP	KLMEAAVN	KESGATVF	FAYHVNDP	ERYGVVEFDKNGTAISLEEKP
	110	120	130	140	150	160
	390	400	410	420	430	440
2MEPSP	DYCIITPPE	KLNVTAID	TYDDHRMA	MAFSLAAC	AEVPVTIR	DPGCTRKTFPDYFDVLSTF
	.	:	.	:	.	:
PDB:1H	----	LEPKS	NYAVTGL	YFYDNDV	VQMAKNL	KPSARGELEITDINRIYLEQGRLSVAMMGR
	170	180	190	200	210	220

2MEPSP VKN

	230	240	250	260	270	280
PDB:1H	GYAWLDTG	THQSLIEA	SNFIATIE	ERQGLKV	SCPEEIAF	RKGFIDVEQVRKLAVPLIKNN

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

APPENDIX E – DETAILED FASTA SEQUENCE SIMILARITY ALIGNMENTS OF THE 2MEPSPS PROTEIN
WITH PROTEINS FROM THE DAD DATABASE

CAB42494.1 | unidentified protein (Sequence 5 from Patent WO9802562.
). | *unidentified* | AA | 445
 Length = 445
 initn: 2060 initl: 2060 opt: 2060 Z-score: 4162.2 bits: 779.3 E(): 0
 Smith-Waterman score: 2060; 91.685% identity (95.730% similar) in 445 aa
 overlap (1-445:1-445)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLP	GSKSLSNRILL	LAALSEGTTVVDNLL	NS	EDVHYMLGAL	
	:::.....					
Dad:CA	MAGAEIIVLQPIKEISGTVKLP	GSKSLSNRILL	LAALSEGTTVVDNLL	NS	EDVHYMLGAL	
	10	20	30	40	50	60
	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVG	CGGKFPVEDAKEEVQLFLGN	AGIAMRSL	xxxxxxxxxxxx		
	:::.....					
Dad:CA	RTLGLSVEADKAAKRAVVVG	CGGKFPVEDAKEEVQLFLGN	AGIAMRSL	TAAVTAAGGNAT		
	70	80	90	100	110	120
	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIGDLVVGLK	QLGADVDCFLGTDCPPVRVNG	IIGGLPGGKVKLSGS	ISS		
	:::.....					
Dad:CA	YVLDGVPRMRERPIGDLVVGLK	QLGADVDCFLGTDCPPVRVNG	IIGGLPGGKVKLSGS	ISS		
	130	140	150	160	170	180
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxGDVEIEI	IDKLISIPYEMTLRLMERF	GVKAEHSDSWDRFYIKGGQ			
	:: .. . :::.....					
Dad:CA	QYLSALLMAAPLALGDVEIEI	IDKLISIPYEMTLRLMERF	GVKAEHSDSWDRFYIKGGQ			
	190	200	210	220	230	240
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFL	AGAAIxxxxxxxxxxxxxSLQ	GDVKFAEVLEMMGAKVTW			
	:::.....					
Dad:CA	KYKSPKNAYVEGDASSASYFL	AGAAITGGTVTVEGCGTTS	LQGDVKFAEVLEMMGAKVTW			
	250	260	270	280	290	300
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREPFGRKHL	KAIDVNMNKM	PDVAMTLAVVALFADGPTA	IRDVASWRVKE		
	:::.....					
Dad:CA	TETSVTVTGPPREPFGRKHL	KAIDVNMNKM	PDVAMTLAVVALFADGPTA	IRDVASWRVKE		
	310	320	330	340	350	360
	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKLGASVEEG	PDYCIITPPEKLN	VT	AIDTYDDHRMAMAFSLAACAEVPV		
	:::.....					
Dad:CA	TERMVAIRTELTKLGASVEEG	PDYCIITPPEKLN	VT	AIDTYDDHRMAMAFSLAACAEVPV		
	370	380	390	400	410	420

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          ::::::::::::::::::::
Dad:CA  TIRDPGCTRKTFPDYFDVLSTFVKN
                430          440

```

CAA03526.1 | *Zea mays* protein (Sequence 4 from Patent WO9704114.) . | *Zea mays* | AA | 444

Length = 444

initn: 2055 initl: 2055 opt: 2055 Z-score: 4152.0 bits: 777.4 E(): 0
 Smith-Waterman score: 2055; 91.667% identity (95.721% similar) in 444 aa
 overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          ::::::::::::::::::::
Dad:CA  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
                10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
          ::::::::::::::::::::
Dad:CA  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLTAAVTAAGGNAT
                60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          ::::::::::::::::::::
Dad:CA  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
                120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          :: .. .. . ::::::::::::::::::::
Dad:CA  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
                180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          ::::::::::::::::::::. . . ::::::::::::::::::::
Dad:CA  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
                240          250          260          270          280          290

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          ::::::::::::::::::::
Dad:CA  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
                300          310          320          330          340          350

```

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          ::::::::::::::::::::
Dad:CA  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
                360          370          380          390          400          410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          ::::::::::::::::::::
Dad:CA  TIRDPGCTRKTFPDYFDVLSTFVKN
          420          430          440

```

CAD42175.1|*Zea mays* protein (Sequence 4 from Patent EP1217073.).|*Zea mays*|AA|444
Length = 444
initn: 2055 initl: 2055 opt: 2055 Z-score: 4152.0 bits: 777.4 E(): 0
Smith-Waterman score: 2055; 91.667% identity (95.721% similar) in 444 aa
overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          ::::::::::::::::::::
Dad:CA  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
                10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
          ::::::::::::::::::::
Dad:CA  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLTAAVTAAGGNAT
                60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          ::::::::::::::::::::
Dad:CA  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
                120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          :: .. .. . ::::::::::::::::::::
Dad:CA  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
                180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          ::::::::::::::::::::. . . ::::::::::::::::::::
Dad:CA  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
                240          250          260          270          280          290

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          ::::::::::::::::::::
Dad:CA  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
                300          310          320          330          340          350

```

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          ::::::::::::::::::::
Dad:CA  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
                360          370          380          390          400          410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          ::::::::::::::::::::
Dad:CA  TIRDPGCTRKTFPDYFDVLSTFVKN
          420          430          440

```

CAA03539.1 | *Zea mays* protein (Sequence 4 from Patent WO9704103.) . | *Zea mays* | AA | 444
Length = 444
initn: 2055 initl: 2055 opt: 2055 Z-score: 4152.0 bits: 777.4 E(): 0
Smith-Waterman score: 2055; 91.667% identity (95.721% similar) in 444 aa
overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          ::::::::::::::::::::
Dad:CA  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
          ::::::::::::::::::::
Dad:CA  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLTAAVTAAGGNAT
          60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          ::::::::::::::::::::
Dad:CA  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          :: .. .. . ::::::::::::::::::::
Dad:CA  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          ::::::::::::::::::::. . . ::::::::::::::::::::
Dad:CA  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
          240          250          260          270          280          290

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          ::::::::::::::::::::
Dad:CA  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          300          310          320          330          340          350

```

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          ::::::::::::::::::::
Dad:CA  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          360          370          380          390          400          410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          ::::::::::::::::::::
Dad:CA  TIRDPGCTRKTFPDYFDVLSTFVKN
          420          430          440

```

CAB42493.1 | unidentified protein (Sequence 3 from Patent WO9802562.
). | *unidentified* | AA | 445
 Length = 445
 initn: 2050 initl: 2050 opt: 2050 Z-score: 4141.9 bits: 775.6 E(): 0
 Smith-Waterman score: 2050; 91.236% identity (95.281% similar) in 445 aa
 overlap (1-445:1-445)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          ::::::::::::::::::::
Dad:CA  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          10          20          30          40          50          60

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
          :::::::::::::::::::: : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNAT
          70          80          90          100          110          120

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          ::::::::::::::::::::
Dad:CA  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          130          140          150          160          170          180

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          : : . . . : ::::::::::::::::::::
Dad:CA  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          190          200          210          220          230          240

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          :::::::::::::::::::: . . . : ::::::::::::::::::::
Dad:CA  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
          250          260          270          280          290          300

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          ::::::::::::::::::::
Dad:CA  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          310          320          330          340          350          360

```

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          ::::::::::::::::::::
Dad:CA  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          370          380          390          400          410          420

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          ::::::::::::::::::::
Dad:CA  TIRDPGCTRKTFPDYFDVLSTFVKN
                430          440

```

CAA03538.1 | *Zea mays* protein (Sequence 2 from Patent WO9704103.) . | *Zea mays* | AA | 444

Length = 444

initn: 2045 initl: 2045 opt: 2045 Z-score: 4131.8 bits: 773.7 E(): 0
 Smith-Waterman score: 2045; 91.216% identity (95.270% similar) in 444 aa
 overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          ::::::::::::::::::::
Dad:CA  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
                10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
          :::::::::::::::::::: : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNAT
                60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          ::::::::::::::::::::
Dad:CA  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
                120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
                180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          :::::::::::::::::::: : : : : : : : : : : : : : : : : : : : : :
Dad:CA  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
                240          250          260          270          280          290

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          ::::::::::::::::::::
Dad:CA  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
                300          310          320          330          340          350

```

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIDTYDDHRMAMAFSLAACAEVPV
          ::::::::::::::::::::
Dad:CA  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIDTYDDHRMAMAFSLAACAEVPV
                360          370          380          390          400          410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          ::::::::::::::::::::
Dad:CA  TIRDPGCTRKTFPDYFDVLSTFVKN
          420          430          440

```

CAD42174.1 | *Zea mays* protein (Sequence 2 from Patent EP1217073.) | *Zea mays* | AA | 444
Length = 444
initn: 2045 initl: 2045 opt: 2045 Z-score: 4131.8 bits: 773.7 E(): 0
Smith-Waterman score: 2045; 91.216% identity (95.270% similar) in 444 aa
overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          ::::::::::::::::::::
Dad:CA  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
                10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
          :::::::::::::::::::: : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNAT
                60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          ::::::::::::::::::::
Dad:CA  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
                120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
                180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          :::::::::::::::::::: . . . : : : : : : : : : : : : : : : : : : :
Dad:CA  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
                240          250          260          270          280          290

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          ::::::::::::::::::::
Dad:CA  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
                300          310          320          330          340          350

```

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          ::::::::::::::::::::
Dad:CA  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
                360          370          380          390          400          410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          ::::::::::::::::::::
Dad:CA  TIRDPGCTRKTFPDYFDVLSTFVKN
          420          430          440

```

CAA44974.1 | *Zea mays* EPSP-synthase protein. | *Zea mays* | AA | 444
Length = 444
initn: 2045 initl: 2045 opt: 2045 Z-score: 4131.8 bits: 773.7 E(): 0
Smith-Waterman score: 2045; 91.216% identity (95.270% similar) in 444 aa
overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          ::::::::::::::::::::
Dad:CA  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
          :::::::::::::::::::: : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNAT
          60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          ::::::::::::::::::::
Dad:CA  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          :::::::::::::::::::: . . . : : : : : : : : : : : : : : : : : :
Dad:CA  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
          240          250          260          270          280          290

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          ::::::::::::::::::::
Dad:CA  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          300          310          320          330          340          350

```

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          ::::::::::::::::::::
Dad:CA  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          360          370          380          390          400          410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          ::::::::::::::::::::
Dad:CA  TIRDPGCTRKTFPDYFDVLSTFVKN
          420          430          440

```

CAA03525.1|*Zea mays* protein (Sequence 2 from Patent WO9704114.).|*Zea mays*|AA|444
Length = 444
initn: 2045 initl: 2045 opt: 2045 Z-score: 4131.8 bits: 773.7 E(): 0
Smith-Waterman score: 2045; 91.216% identity (95.270% similar) in 444 aa
overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          ::::::::::::::::::::
Dad:CA  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
          :::::::::::::::::::: : : : : : : : : : : : : : : : : : : : :
Dad:CA  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNAT
          60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          ::::::::::::::::::::
Dad:CA  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          :::::::::::::::::::: : : : : : : : : : : : : : : : : : :
Dad:CA  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
          240          250          260          270          280          290

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          ::::::::::::::::::::
Dad:CA  TETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          300          310          320          330          340          350

```

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          ::::::::::::::::::::
Dad:CA  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          360          370          380          390          400          410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          ::::::::::::::::::::
Dad:CA  TIRDPGCTRKTFPDYFDVLSTFVKN
          420          430          440

CAD01096.1|Eleusine indica 5-enolpyruvylshikimate-3-phosphate synthase
protein.|Eleusine indica|AA|445
Length = 445
initn: 1606 initl: 1606 opt: 2001 Z-score: 4042.7 bits: 757.2 E(): 2.8e-216
Smith-Waterman score: 2001; 89.213% identity (94.831% similar) in 445 aa
overlap (2-445:1-445)

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          ::::.:::::::::.:::::::::::::.:.:::::::::::::::::::
Dad:CA  AGAEEVVLQPIKEISGVVGLPGSKSLSNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
          10          20          30          40          50

                70          80          90          100         110
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxx
          .::::::::::::.:::::::::::::.: : : : : : : : : : : : : : : : .
Dad:CA  KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNA
          60          70          80          90          100         110

                120         130         140         150         160         170
2MEPSP  xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
          .::::::::::::.:::::::::::::.:::::::::::::.: : : : : : : : : : .
Dad:CA  TYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSIS
          120         130         140         150         160         170

                180         190         200         210         220         230
2MEPSP  SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
          : : : . . . . . : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
          180         190         200         210         220         230

                240         250         260         270         280         290
2MEPSP  QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
          .::::::::::::.: : : . . . : : : : : : : : : : : : : : : :
Dad:CA  QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
          240         250         260         270         280         290

                300         310         320         330         340         350
2MEPSP  WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
          .::::::::::::.: : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
          300         310         320         330         340         350

                360         370         380         390         400         410
2MEPSP  ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVP
          .::::::::::::.: : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACADVP
          360         370         380         390         400         410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      .....
Dad:CA VTIRDPGCTRKTFPDYFDVLSTFVKN
      420      430      440

```

AAN63155.1 | *Eleusine indica* 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Eleusine indica* | AA | 445
Length = 445
initn: 1606 initl: 1606 opt: 2001 Z-score: 4042.7 bits: 757.2 E(): 2.8e-216
Smith-Waterman score: 2001; 89.213% identity (94.831% similar) in 445 aa
overlap (2-445:1-445)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      .....
Dad:AA AGAEEVVLQPIKEISGVVGLPGSKSLNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
      10      20      30      40      50

```

```

      70      80      90      100     110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .....
Dad:AA KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNA
      60      70      80      90      100     110

```

```

      120     130     140     150     160     170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .....
Dad:AA TYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSIS
      120     130     140     150     160     170

```

```

      180     190     200     210     220     230
2MEPSP SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      ... ..
Dad:AA SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      180     190     200     210     220     230

```

```

      240     250     260     270     280     290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      .....
Dad:AA QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
      240     250     260     270     280     290

```

```

      300     310     320     330     340     350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      .....
Dad:AA WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      300     310     320     330     340     350

```

```

      360     370     380     390     400     410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVP
      .....
Dad:AA ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACADVP
      360     370     380     390     400     410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      ::::::::::::::::::::::::::::
Dad:AA VTIRDPGCTRKTFPDYFDVLSTFVKN
      420      430      440

```

CAD01095.1 | *Eleusine indica* 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Eleusine indica* | AA | 445
Length = 445
initn: 1601 initl: 1601 opt: 1996 Z-score: 4032.6 bits: 755.3 E(): 1e-215
Smith-Waterman score: 1996; 89.213% identity (94.831% similar) in 445 aa
overlap (2-445:1-445)

```

              10              20              30              40              50              60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      ::::.::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:CA AGAEEVVLQPIKEISGVVCLPGSKSLSNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
              10              20              30              40              50

```

```

              70              80              90              100             110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      .::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::. . . . .
Dad:CA KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRSLTAAVTAAGGNA
      60              70              80              90              100             110

```

```

              120             130             140             150             160             170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:CA TYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSIS
      120             130             140             150             160             170

```

```

              180             190             200             210             220             230
2MEPSP SQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      ::: .. .. . ::::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:CA SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      180             190             200             210             220             230

```

```

              240             250             260             270             280             290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      ::::::::::::::::::::::::::::::. . . . . ::::::::::::::::::::::::::::
Dad:CA QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
      240             250             260             270             280             290

```

```

              300             310             320             330             340             350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      :::::::::::::: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:CA WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      300             310             320             330             340             350

```

```

              360             370             380             390             400             410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVP
      :::::::::::::::::::::: ::::::::::::::::::::::::::::::::::::::::::::::
Dad:CA ETERMVAIRTELTKLGASVEEGLDYCIITPPEKLNVTADTYDDHRMAMAFSLAACADVP
      360             370             380             390             400             410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      ::::::::::::::::::::::
Dad:CA VTIRDPGCTRKTFPDYFDVLSTFVKN
      420      430      440

```

AAN63156.1 | *Eleusine indica* 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Eleusine indica* | AA | 445
Length = 445
initn: 1601 initl: 1601 opt: 1996 Z-score: 4032.6 bits: 755.3 E(): 1e-215
Smith-Waterman score: 1996; 89.213% identity (94.831% similar) in 445 aa
overlap (2-445:1-445)

```

              10              20              30              40              50              60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      ::::.::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:AA AGAEEVVLQPIKEISGVVLPGSKSLSNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
              10              20              30              40              50

```

```

              70              80              90              100             110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .::::::::::::::::::::::::::::::::::::::::::::::::::::::::::.
Dad:AA KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRSLTAAVTAAGGNA
      60              70              80              90              100             110

```

```

              120             130             140             150             160             170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:AA TYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSIS
      120             130             140             150             160             170

```

```

              180             190             200             210             220             230
2MEPSP SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      ::: .. .. . ::::::::::::::::::::::::::::::::::::::::::::::
Dad:AA SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      180             190             200             210             220             230

```

```

              240             250             260             270             280             290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      ::::::::::::::::::::::::::::::. . . . ::::::::::::::::::::::
Dad:AA QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
      240             250             260             270             280             290

```

```

              300             310             320             330             340             350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:AA WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      300             310             320             330             340             350

```

```

              360             370             380             390             400             410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVP
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:AA ETERMVAIRTELTKLGASVEEGLDYCIITPPEKLNVTADTYDDHRMAMAFSLAACADVP
      360             370             380             390             400             410

```

	420	430	440	
2MEPSP	VTIRDPGCTRKTFPDYFDVLSTFVKN			
			
Dad:AA	VTIRDPGCTRKTFPDYFDVLSTFVKN			
	420	430	440	
AAR87844.1 <i>Eleusine indica</i> 5-enolpyruvylshikimate-3-phosphate synthase protein. <i>Eleusine indica</i> AA 445				
Length = 445				
initn: 1597 initl: 1597 opt: 1992 Z-score: 4024.5 bits: 753.8 E(): 2.9e-215				
Smith-Waterman score: 1992; 88.764% identity (94.607% similar) in 445 aa				
overlap (2-445:1-445)				
	10	20	30	40
2MEPSP	MAGAEIEVLQPIKEISGTVKLP	SGSKSLSNRILL	LAALSEGTTVVDNLL	NSEDVHYMLGAL

Dad:AA	AGAEIEVLQPIKEISGTVKLP	SGSKSLSNRILL	LSALAE	GTTVVDNLLNSEDVHYMLGAL
	10	20	30	40
	70	80	90	100
2MEPSP	RTLGLSVEADKAAKRAVVV	GC	GKFPVE-DAKEEVQLFLGNAGIAMRSL	xxxxxxxxxxxx

Dad:AA	KTLGLSVEADKAAKRAVVV	GC	GKFPVEKDAKEEVQLFLGNAGTAMRALTA	AVTAAGGNA
	60	70	80	90
	120	130	140	150
2MEPSP	xYVLDGVP	MRERPIGDLVVGLKQLGADVDCFLGTDCPPVRV	NGIGGLP	GGKVKLSGSIS

Dad:AA	TYVLDGVP	MRERPIGDLVVGLKQLGADVDCFLGTDCPPVRV	KGIGGLP	GGKVKLSGSIS
	120	130	140	150
	180	190	200	210
2MEPSP	SQYxxxxxxxxxxxxGDVEIEI	IDKLISIPYEMTLRLMERFGVKA	EHSDSWDRFYIKGG	
	:::
Dad:AA	SQYLSALLMAAPLALGDVEIEI	IDKLISIPYEMTLRLMERFGVKA	EHSDSWDRFYIKGG	
	180	190	200	210
	240	250	260	270
2MEPSP	QKYKSPKNAYVEGDASSASYFLAGAAI	xxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT		

Dad:AA	QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTV	EGCGTTS	LQGDVKFAEVLETMGAKVT	
	240	250	260	270
	300	310	320	330
2MEPSP	WTETSVTVTGPPREPFGRKHLKAIDVNMN	KMPDVAMTLAVVALFADGPTAIRDVASWRVK		

Dad:AA	WTETSVTVTGPPREPFGRKHLKAIDVNMN	KMPDVAMTLAVVALFADGPTAIRDVASWRVK		
	300	310	320	330
	360	370	380	390
2MEPSP	ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN	VT	AIDTYDDHRMAMAFSLAACAEVP	

Dad:AA	ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN	VT	AIDTYDDHRMAMAFSLAACADVP	
	360	370	380	390

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      ::::::::::::::::::::::
Dad:AA VTIRDPGCTRKTFPDYFDVLSTFVKN
      420      430      440

```

AAR87845.1 | *Eleusine indica* 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Eleusine indica* | AA | 445
Length = 445
initn: 1588 initl: 1588 opt: 1983 Z-score: 4006.3 bits: 750.5 E(): 3e-214
Smith-Waterman score: 1983; 88.764% identity (94.382% similar) in 445 aa
overlap (2-445:1-445)

```

              10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      ::::.::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:AA AGAEEVVLQPIKEISGVVGLPGSKSLNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
              10      20      30      40      50

```

```

              70      80      90      100     110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .:::::::::::::::::::::::::::::::::::::::::::: :::::::::::::: :::: :... ..
Dad:AA KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNA
              60      70      80      90      100     110

```

```

              120     130     140     150     160     170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .:::::::::::: :::::::::::::::::::::::::::::: ::::::.::::::::::::::::::
Dad:AA TYVLDGVPRMLERPIGDLVVGLKQLGADVDCFLGTDYPPVRVKGIGGLPGGKVKLSGSIS
              120     130     140     150     160     170

```

```

              180     190     200     210     220     230
2MEPSP SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      ::: .. .. . ::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:AA SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
              180     190     200     210     220     230

```

```

              240     250     260     270     280     290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      ::::::::::::::::::::::::::::::. . . . ::::::::::::::::::::::::::::
Dad:AA QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
              240     250     260     270     280     290

```

```

              300     310     320     330     340     350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      :::::::::::::: :::::::::::::::::::::::::::::: ::::::::::::::::::::
Dad:AA WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
              300     310     320     330     340     350

```

```

              360     370     380     390     400     410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADITYDDHRMAMAFSLAACAEVP
      ::::::::::::::::::::::::::::::. . . . ::::::::::::::::::::::::::::
Dad:AA ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADITYDDHRMAMAFSLAACADVP
              360     370     380     390     400     410

```

BAB61062.1|*Oryza sativa* 3-phosphoshikimate 1-carboxyvinyltransferase protein.|*Oryza sativa*|AA|511
Length = 511
initn: 1950 init1: 1584 opt: 1951 Z-score: 3940.5 bits: 738.5 E(): 1.4e-210
Smith-Waterman score: 1951; 86.517% identity (94.382% similar) in 445 aa
overlap (2-445:67-511)

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      :
Dad:AA VTIRDPGCTRKTFPDYFDVLSTFVKN
      420      430      440

      10      20      30
2MEPSP      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRI
      :
Dad:BA GGMVRVRVRARGRREAVVVASASSSSVAAPAAKAEIVLQPIREISGAVQLPGSKSLSNRI
      40      50      60      70      80      90

      40      50      60      70      80      90
2MEPSP LLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVVGCGGKFPVE-DA
      :
Dad:BA LLLSALSEGTTVVDNLLNSEDVHYMLEALKALGLSVEADKVAKRAVVVVGCGGKFPVEKDA
      100      110      120      130      140      150

      100      110      120      130      140      150
2MEPSP KEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPMRERPIGDLVVGLKQLGADVDC
      :
Dad:BA KEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVLGDGVPMRERPIGDLVVGLKQLGADVDC
      160      170      180      190      200      210

      160      170      180      190      200      210
2MEPSP FLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPY
      :
Dad:BA FLGTECPPVRVKIGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDKLISIPY
      220      230      240      250      260      270

      160      170      180      190      200      210
2MEPSP FLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPY
      :
Dad:BA FLGTECPPVRVKIGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDKLISIPY
      220      230      240      250      260      270

      220      230      240      250      260      270
2MEPSP VEMTLRLMERFGVKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxx
      :
Dad:BA VEMTLRLMERFGVKAEHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGAAITGGT
      280      290      300      310      320      330

      280      290      300      310      320      330
2MEPSP xxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGKHLKAIDVNMNMK
      :
Dad:BA VTVQGCGTTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPREPYPGKHLKAIDVNMNMK
      340      350      360      370      380      390

      340      350      360      370      380      390
2MEPSP PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP
      :
Dad:BA PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP
      400      410      420      430      440      450
```


340 350 360 370 380 390
 2MEPSP PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP
 :
 Dad:AA PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP
 400 410 420 430 440 450

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	400	410	420	430	440	
2MEPSP	EKLNVT	AI	DYDDHR	MAMAFSLAACAE	VPVTIRDPGCTRKT	FPDYFDVLSTFVKN

Dad:BA	EKLNIT	AI	DYDDHR	MAMAFSLAACAD	VPVTIRDPGCTRKT	FPNYFDVLSTFVRN
	470	480	490	500	510	

BAF18626.1|*Oryza sativa* (japonica cultivar-group) Os06g0133900 protein.|*Oryza sativa* (japonica cultivar-group)|AA|515
Length = 515
initn: 1950 initl: 1584 opt: 1951 Z-score: 3940.5 bits: 738.5 E(): 1.4e-210
Smith-Waterman score: 1951; 86.517% identity (94.382% similar) in 445 aa
overlap (2-445:71-515)

	10	20	30
2MEPSP	MAGAE	EIVLQPIKE	ISGTVKLP

Dad:BA	GGMRV	RV	RARGR
	50	60	70

	40	50	60	70	80	90	100
2MEPSP	LLLAAL	SE	GT	TTVVDN	LLNS	EDVHYMLG	ALRTLG

Dad:BA	LLLSAL	SE	GT	TTVVDN	LLNS	EDVHYMLE	ALKALG
	110	120	130	140	150	160	

	100	110	120	130	140	150
2MEPSP	KEEVQ	LFLGN	AGIAMR	SLxxxxxxxxxxxx	YVLDG	VPRMRER

Dad:BA	KEEVQ	LFLGN	AGTAMR	PLTAAV	TAAAGG	NATYVLDG
	170	180	190	200	210	220

	160	170	180	190	200	210
2MEPSP	FLGTDC	PPVRV	NGIGGL	PGGKVK	LSGSIS	SQYxxxxxxxxxxxx

Dad:BA	FLGTE	CPPVR	VKGIGGL	PGGKVK	LSGSIS	QYLSALLMA
	230	240	250	260	270	280

	220	230	240	250	260	270
2MEPSP	VEMTL	RRLMER	FGVKA	EHSDS	WDRFY	IKGGQKYK

Dad:BA	VEMTL	RRLMER	FGVKA	EHSDS	WDRFY	IKGGQKYK
	290	300	310	320	330	340

	280	290	300	310	320	330
2MEPSP	xxxxxxx	SLQGD	VKFAE	VLEMM	GAKVT	TWTET

Dad:BA	VTVQG	CGTTS	LQGD	VKFAE	VLEMM	GAKVT
	350	360	370	380	390	400

	340	350	360	370	380	390
2MEPSP	PDVAM	TLAV	VALF	ADGPT	AIRDV	ASWRV

Dad:BA	PDVAM	TLAV	VALF	ADGPT	AIRDV	ASWRV
	410	420	430	440	450	460

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	400	410	420	430	440
2MEPSP	EKLNVT	AI	DYDDHRMAMAFSLAACAEVPVTIRDPGCTRKT	FPDYFDVLSTFVK	N

Dad:BA	EKLNIT	AI	DYDDHRMAMAFSLAACADVPVTIRDPGCTRKT	FPNYFDVLSTFVR	N
	470	480	490	500	510

BAD68865.1 | *Oryza sativa Japonica Group* putative 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Oryza sativa Japonica Group* | AA | 515
Length = 515
initn: 1950 initl: 1584 opt: 1951 Z-score: 3940.5 bits: 738.5 E(): 1.4e-210
Smith-Waterman score: 1951; 86.517% identity (94.382% similar) in 445 aa
overlap (2-445:71-515)

		10	20	30
2MEPSP		MAGAE	IVLQPIKEIS	GT
	
Dad:BA	GGMRV	RV	RARGR	EA
	50	60	70	80

	40	50	60	70	80	90
2MEPSP	LLLAAL	SE	GT	TV	VD	N

Dad:BA	LLLSAL	SE	GT	TV	VD	N
	110	120	130	140	150	160

	100	110	120	130	140	150
2MEPSP	KEEVQ	L	FLGNAGIAMRSL	XXXXXXXXXXXX	YVLDG	V

Dad:BA	KEEVQ	L	FLGNAGTAMRPL	TAAVTAAGGNAT	YVLDG	V
	170	180	190	200	210	220

	160	170	180	190	200	210
2MEPSP	FLGTD	CPP	VRVNGIGGLPGGK	VKL	SGSISSQY	XXXXXXXXXXXX

Dad:BA	FLGTE	CPP	VRVKGIGGLPGGK	VKL	SGSISSQYLSALL	MA
	230	240	250	260	270	280

	220	230	240	250	260	270
2MEPSP	VEMTL	R	L	M	E	R

Dad:BA	VEMTL	R	L	M	E	R
	290	300	310	320	330	340

	280	290	300	310	320	330
2MEPSP	xxxxxxx	SLQGD	V	K	FAE	V

Dad:BA	VTVQ	G	CGT	T	SLQGD	V
	350	360	370	380	390	400

	340	350	360	370	380	390
2MEPSP	PDVAM	T	L	A	V	A

Dad:BA	PDVAM	T	L	A	V	A
	410	420	430	440	450	460

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	400	410	420	430	440
2MEPSP	EKLNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN				

Dad:BA	EKLNITAIDTYDDHRMAMAFSLAACADVPVTIRDPGCTRKTFPNYFDVLSTFVRN				
	470	480	490	500	510

EAY99539.1 | *Oryza sativa* Indica Group hypothetical protein protein. | *Oryza sativa* Indica Group | AA | 515
Length = 515
initn: 1950 initl: 1584 opt: 1951 Z-score: 3940.5 bits: 738.5 E(): 1.4e-210
Smith-Waterman score: 1951; 86.517% identity (94.382% similar) in 445 aa
overlap (2-445:71-515)

	10	20	30
2MEPSP	MAGAEIIVLQPIKEISGTVKLPGSKSLSNRI		
		
Dad:EA	GGMRVRVRARGRREAVVASASSSSVAAPAAKAEIIVLQPIREISGAVQLPGSKSLSNRI		
	50	60	70

	40	50	60	70	80	90
2MEPSP	LLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DA					

Dad:EA	LLLSALSEGTTVVDNLLNSEDVHYMLEALKALGLSVEADKVAKRAVVVGCGGKFPVEKDA					
	110	120	130	140	150	160

	100	110	120	130	140	150
2MEPSP	KEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDC					

Dad:EA	KEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQLGADVDC					
	170	180	190	200	210	220

	160	170	180	190	200	210
2MEPSP	FLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPY					

Dad:EA	FLGTECPPVRVKIGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDKLISIPY					
	230	240	250	260	270	280

	220	230	240	250	260	270
2MEPSP	VEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxx					

Dad:EA	VEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGAAITGGT					
	290	300	310	320	330	340

	280	290	300	310	320	330
2MEPSP	xxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKM					

Dad:EA	VTVQCGTTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPREPYGKKHLKAIDVNMNKM					
	350	360	370	380	390	400

	340	350	360	370	380	390
2MEPSP	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP					

Dad:EA	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP					
	410	420	430	440	450	460

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	400	410	420	430	440	
2MEPSP	EKLNVT	AI	DYDDHR	MAMAFSLAACAE	VPVTIRDPGCTRKT	FPDYFDVLSTFVKN

Dad:EA	EKLNIT	AI	DYDDHR	MAMAFSLAACAD	VPVTIRDPGCTRKT	FPNYFDVLSTFVRN
	470	480	490	500	510	

AZ79230.2 | *Lolium multiflorum* plastid 5-enolpyruvylshikimate 3-phosphate synthase protein. | *Lolium multiflorum* | AA | 437

Length = 437

initn: 1903 initl: 1570 opt: 1904 Z-score: 3846.4 bits: 720.9 E(): 2.4e-205

Smith-Waterman score: 1904; 85.812% identity (93.364% similar) in 437 aa
 overlap (10-445:1-437)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVL	QPIKEIS	GT	VKLPGSK	SLSNRILL

Dad:AA	QPIKEIS	GA	VQLPGSK	SLSNRILL	LS	ALSEGTT
	10	20	30	40	50	

	70	80	90	100	110
2MEPSP	RTLGLS	VEADKAA	KRAVV	VGCGGK	FPVE-DAKEEV

Dad:AA	DALGLS	VEADK	VAKRAVV	VGCGGR	FP
	60	70	80	90	100

	120	130	140	150	160	170
2MEPSP	xYVLD	GVPRMR	ERPIG	DLVVGL	KQLGAD	VDCFLGT

Dad:AA	TYVLD	GVPRMR	ERTG	DLVVGL	KQLGAN	VDCFLGT
	120	130	140	150	160	170

	180	190	200	210	220	230
2MEPSP	SQY	xxxxxxxx	xxGDVE	IEI	IDKLIS	IPYVEM

Dad:AA	SQYL	SSLL	MAAP	LALGD	VEIEI	IDKLIS
	180	190	200	210	220	230

	240	250	260	270	280	290
2MEPSP	QKYK	SPKNAY	VEGD	ASSASY	FLAGAAI	xxxxxxxx

Dad:AA	QKYK	SPGNAY	VEGD	ASSASY	FLAGAAI	TGGT
	240	250	260	270	280	290

	300	310	320	330	340	350
2MEPSP	WTETS	VT	VTGPP	RP	FG	RKHLKA

Dad:AA	WTDTS	VT	VTGPP	RQ	PF	GRKHLKA
	300	310	320	330	340	350

	360	370	380	390	400	410
2MEPSP	ETERM	VAIR	TELTK	L	GASVEE	GP

Dad:AA	ETERM	VAIC	TELTK	L	GATVEE	GP
	360	370	380	390	400	410

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFFPDYFDVLSTFVK
      .....
Dad:AA VTIRDPGCTRKTFFPNYFDVLSTFVK
      420      430

```

AAV64030.1 | *Camptotheca acuminata* 5-enolpyruvylshikimate 3-phosphate synthase protein. | *Camptotheca acuminata* | AA | 519
Length = 519
initn: 1788 initl: 1446 opt: 1795 Z-score: 3624.6 bits: 680.1 E(): 5.5e-193
Smith-Waterman score: 1795; 79.819% identity (91.156% similar) in 441 aa
overlap (6-445:79-519)

```

                        10      20      30
2MEPSP                MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                        .....
Dad:AA VAVNSPVINVRVPLRVSASVVTTEKTSMTPEIVLQPIKEISGTVKLPGSKSLSNRILL
      50      60      70      80      90      100

```

```

      40      50      60      70      80      90
2MEPSP ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPV-EDAKEEV
      .....
Dad:AA ALSEGTTVVDNLLNSDDVHYMLGALRTLGLRVEEDSAIKRAIVEGCSGLFPVGKESTDEV
      110     120     130     140     150     160

```

```

      100     110     120     130     140     150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      .....
Dad:AA QLFLGNAGTAMRPLTAAVTAAGGNSSYILDGVPRMRERPIGDLVTGLKQLGADVDCFLGT
      170     180     190     200     210     220

```

```

      160     170     180     190     200     210
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      .....
Dad:AA NCPVVRVIGKGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISIPYVEMT
      230     240     250     260     270     280

```

```

      220     230     240     250     260     270
2MEPSP LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
      .....
Dad:AA LKLMKRFVGTVEHSDNWDRFLIQGGQKYKSPGNSYVEGDASSASYFLAGAAVTGGTITVE
      290     300     310     320     330     340

```

```

      280     290     300     310     320     330
2MEPSP xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA
      .....
Dad:AA GCGSSSLQGDVKFAEVLEKMGAKVTWTENSVTVTGPPRNSSGRKHLRAVDVNMNKMMPDVA
      350     360     370     380     390     400

```

```

      340     350     360     370     380     390
2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      .....
Dad:AA MTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLGAATVEEGPDYCVITPPEKLN
      410     420     430     440     450     460

```

```

      400          410          420          430          440
2MEPSP VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      ... ..
Dad:AA VTAVDTYDDHRMAMAFSLAACANVPVTIKDPGCTRKTFPDYFEVLQRFAPH
      470          480          490          500          510

AAT45243.1|Conyza canadensis 5-enol-pyruvylshikimate-phosphate synthase
protein.|Conyza canadensis|AA|519
Length = 519
initn: 1776 initl: 1437 opt: 1781 Z-score: 3596.2 bits: 674.8 E(): 2.1e-191
Smith-Waterman score: 1781; 78.912% identity (91.383% similar) in 441 aa
overlap (6-445:79-519)

                        10           20           30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLA
                        .....
Dad:AA SVSSNNNRNLGKKCLIVSAVATTEKPSTVPEIVLQPIKEISGTVNLP GSKSLSNRILLLA
      50           60           70           80           90          100

      40           50           60           70           80           90
2MEPSP ALSEGT TVVDNLLNS EDVHYMLGALRTLGLSVEAD KAAKR AVVVGC GGKFPV -E DAK EEV
      ::..... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA ALAEGT TIVDNLLNS DDVHYMLGALRTLGLNVEED VAIKRA IVEGCGGV FVPVG KEAKDDI
      110          120          130          140          150          160

      100          110          120          130          140          150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVP RMRERPI GD LVVGLKQLGADVDC FLGT
      ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA QLFLGNAGTAMRPLTAAVTAAGNSSYILDGVP RMRERPI GD LV TG LKQLGADVDC SLGT
      170          180          190          200          210          220

      160          170          180          190          200          210
2MEPSP DCP PVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA NCP PVRVVG GGLPGGKVKLSGSISSQYL TALLMAS PLALGDVEIEIIDKLISIPYVEMT
      230          240          250          260          270          280

      220          230          240          250          260          270
2MEPSP LRLMERFGVKA EHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxx
      . : : : : : . : : : : : . : : : : : : : : : : : : : : : . .
Dad:AA LKL MERFGVSVEHSDSWD QFFIRGGQKYKSPGNAYVEGDASSASYFLAGAAITGGTITVE
      290          300          310          320          330          340

      280          290          300          310          320          330
2MEPSP xxxxxSLQG DVKF AEVLEMMGA KVTTW TETSVTVTGP PREPFGRKHLKAIDVN MNKM PDVA
      .. : : : : : : : : : : . : : : : : : : : : : : : : : : : : : : : :
Dad:AA GCGTSSLQGD VKFAEV LGQMGA EVTWTENS VT VK GP PRDSS GRKHLRAVDVN MNKM PDVA
      350          360          370          380          390          400

      340          350          360          370          380          390
2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA MTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKL GATVEEGPDYCVITPPEKLN
      410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP  VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
          .....
Dad:AA  VTAIDTYDDHRMAMAFSLAACADVPVTIKDPSCSTRKTFFPDYFEVLQRFKAH
          470          480          490          500          510

```

AA40473.1 | *Conyza sumatrensis* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Conyza sumatrensis* | AA | 447

Length = 447

initn: 1774 initl: 1434 opt: 1779 Z-score: 3593.2 bits: 674.0 E(): 3.1e-191
 Smith-Waterman score: 1779; 79.138% identity (91.156% similar) in 441 aa
 overlap (6-445:7-447)

```

          10          20          30          40          50
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGA
          .....
Dad:AA  KPSTVPEIVLQPIKEISGTVNLPGSKSKLSNRILLLAALAEGETTIVDNLLNSDDVHYMLGA
          10          20          30          40          50          60

```

```

          60          70          80          90          100          110
2MEPSP  LRTLGLSVEADKAAKRAVVVGCCKGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxx
          .....
Dad:AA  LRTLGLNVEEDGAIKRAIVEGCGGMFPVGKEAKDDIQLFLGNAGTAMRPLTAAVTAAGGN
          70          80          90          100          110          120

```

```

          120          130          140          150          160          170
2MEPSP  xxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSI
          .....
Dad:AA  SSVVLDGVPRMRERPIGDLVTGLKQLGADVDCSLGTNCPVVRVVGGLPGGKVKLSGSI
          130          140          150          160          170          180

```

```

          180          190          200          210          220          230
2MEPSP  SSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG
          .....
Dad:AA  SSQYLTALLMASPLALGDVEIEIIDKLISIPYVEMTLKLMWFGVSVEHSDSWDRFFIRG
          190          200          210          220          230          240

```

```

          240          250          260          270          280          290
2MEPSP  GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
          .....
Dad:AA  GQKYKSPGNAYVEGDASSASYFLAGAAITGGTITVEGCGTSSLQGDVKFAEVLGQMGAEV
          250          260          270          280          290          300

```

```

          300          310          320          330          340          350
2MEPSP  TWTETSVTVTGPPREPFGKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRV
          .....
Dad:AA  TWTENSVTVKGPPRDSSGRKHLRAVDVNMNKMPPDVAMTLAVVALYADGPTAIRDVASWRV
          310          320          330          340          350          360

```

```

          360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEV
          .....
Dad:AA  KETERMIAICTELRKLGATVEEGPDYCVITPPEKLNVT AIDTYDDHRMAMAFSLAACADV
          370          380          390          400          410          420

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP PVTIRDPGCTRKTFPDYFDVLSTFVKN
      .....
Dad:AA PVTIKDPSCTRKTFPDYFEVLQRFK
      430      440

```

AAL27697.1 | *Dicliptera chinensis* 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Dicliptera chinensis* | AA | 516
Length = 516
initn: 1527 initl: 824 opt: 1772 Z-score: 3578.0 bits: 671.4 E(): 2.2e-190
Smith-Waterman score: 1772; 79.224% identity (91.324% similar) in 438 aa
overlap (6-442:77-513)

```

                        10      20      30
2MEPSP                MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                        .....
Dad:AA ISVSNVGKSRQLQLQVVAAAKTAEKPPAVPEIVLQPIKDISGTVKLPGSKSLSNRVLLLA
      50      60      70      80      90      100

```

```

      40      50      60      70      80      90
2MEPSP ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
      .....
Dad:AA ALSEGTTVVENLLSSEDIHYMLGALRTLGLAVEEDKANQKAVVEGCVGQFPASKEGKDEV
      110     120     130     140     150     160

```

```

      100     110     120     130     140     150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      .....
Dad:AA QLFLGNAGTAMRPLTAAVVAAGGNARYVLDGVPRLRERPIGDLVTGLKQLGADVDCFLGT
      170     180     190     200     210     220

```

```

      160     170     180     190     200     210
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      .....
Dad:AA NCPVVRVVGKGLPGGKVKLSGSVSSQYLTALLMSAPLALGDVEIEIVDKLISVPYVEMT
      230     240     250     260     270     280

```

```

      220     230     240     250     260     270
2MEPSP LRLMERFVGKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
      .....
Dad:AA LKLMERFGVYVEHTDNWDRFLIRGGQKYKSPGNAFVEGDASSASYFLAGAAVTGGTVTVE
      290     300     310     320     330     340

```

```

      280     290     300     310     320     330
2MEPSP xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVA
      .....
Dad:AA GCGTSSLQGDVKFAEVLEKMGAKVSWTENSVTVTGPPRVP-GKKHLRAIDVNMNKMPPDVA
      350     360     370     380     390     400

```

```

      340     350     360     370     380     390
2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      .....
Dad:AA MTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLATVEEGPDYCIITPPEKLN
      410     420     430     440     450     460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
          .....
Dad:AA VTAIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFPNYFDVLSTYSNH
          470          480          490          500          510

```

ABA54869.1 | *Fagus sylvatica* putative 5-enolpyruvylshikimate 3-phosphate synthase protein. | *Fagus sylvatica* | AA | 520
Length = 520
initn: 1754 initl: 1409 opt: 1772 Z-score: 3578.0 bits: 671.4 E(): 2.2e-190
Smith-Waterman score: 1772; 78.005% identity (90.476% similar) in 441 aa
overlap (6-445:80-520)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                                .....
Dad:AB KHVCAVGS AKVGTFRVSASVATAEKPSTVPEIVLQPIKDISGTIKLPGSKSLSNRILLLA
          50          60          70          80          90          100

```

```

          40          50          60          70          80          90
2MEPSP ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
          .....
Dad:AB ALSEGTTVVDNLLNSDDIHYMLGALKTLGLRVEEDKAIKRAIVEGCGGLFPVGKESRDEI
          110          120          130          140          150          160

```

```

          100          110          120          130          140          150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVP MRERP IGD LVVGLKQLGADVDCFLGT
          .....
Dad:AB QLFLGNAGTAMRPLTAAVTVAGGNSSYVLGDGVP MRERP IGD LV DGLQQLGANVDCFLGT
          170          180          190          200          210          220

```

```

          160          170          180          190          200          210
2MEPSP DCP PVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
          .....
Dad:AB KCP PVRVFGKGLPGGKVKLSGSISSQYLTALVMAAPLALGDVEIEIIDKLISVPYVEMT
          230          240          250          260          270          280

```

```

          220          230          240          250          260          270
2MEPSP LRLMERFGVKA EHS DSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
          .....
Dad:AB LKLMERFGISVEHNDSWDRFLIRGGQKYKSPGNAYVEGDASSASYFLAGAAVTGGTVTVE
          290          300          310          320          330          340

```

```

          280          290          300          310          320          330
2MEPSP xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFG RKHLKAIDVNMNKM PDVA
          .....
Dad:AB GCGTSSLQGDVKFAEVLEKMGAKVTWSETSVTVTGPPQDSSKKKHLRAIDVNMNKM PDVA
          350          360          370          380          390          400

```

```

          340          350          360          370          380          390
2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
          .....
Dad:AB MTLAVVALFADGPTAIRDVASWRVKETERMIAVCTELRKL GATVEEGPDYCVITPPEKLN
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP  VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
          :: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:AB  VTDIDTYDDHRMAMAFSLAACGDVPVTINDPGCTRKTFFPDYFEVLQRFTKH
          470          480          490          500          510          520

```

CAO24534.1|*Vitis vinifera* protein (*Vitis vinifera* chromosome chr15 scaffold_37, whole genome shotgun sequence.).|*Vitis vinifera*|AA|521
Length = 521
initn: 1772 initl: 1442 opt: 1771 Z-score: 3576.0 bits: 671.1 E(): 2.8e-190
Smith-Waterman score: 1771; 78.555% identity (90.971% similar) in 443 aa
overlap (4-445:79-521)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                                : ::::::::::::::: :::::::::::::::
Dad:CA  HERAVSNSIVSVRAPFRVSASVATKEKPSTAPEIVLQPIKEISGTITLPGSKSLSNRILL
          50          60          70          80          90          100

```

```

          40          50          60          70          80          90
2MEPSP  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKE-
          ::::::::::::::::::::::::::::::::::::::: :: .. ::::: ::::: ..
Dad:CA  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLHVEEQSENKRIVVQCGGQFPAGNGSVG
          110          120          130          140          150          160

```

```

          100          110          120          130          140          150
2MEPSP  EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
          ::::::::::: :: :... .. :::::::::::::::::::::::::::::::::::::
Dad:CA  EVQLFLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVTGLKQLGADVNCFL
          170          180          190          200          210          220

```

```

          160          170          180          190          200          210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
          ::::::::::: ::::::::::::::: .. .. ::::::::::::::::::::::
Dad:CA  GTNCPVVRVNGNGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISIPYVE
          230          240          250          260          270          280

```

```

          220          230          240          250          260          270
2MEPSP  MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
          ::::::::::: ::::::::::::::: :::::::::::::: :::::::::::::: ..
Dad:CA  MTLKLMERFGVSVEHSNTWDRFLIRGGQKYKSPGNAFVEGDASSASYFLAGAAVTGGTVT
          290          300          310          320          330          340

```

```

          280          290          300          310          320          330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
          .. ::::::::::: ::::::::::: :::::::::::::: ::::::::::::::
Dad:CA  VEGCGTSSLQGDVKFAEVLEQMGAKVSWMENSVTVTGPPRDSSGRKHLRAIDVNMNKMPD
          350          360          370          380          390          400

```

```

          340          350          360          370          380          390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
          ::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Dad:CA  VAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGAATVEEGPDYCVITPPEK
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
        .....
Dad:CA  LNVTSIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFFPDYFEVLQRFTKH
        470          480          490          500          510          520

```

AA40472.1 | *Conyza sumatrensis* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Conyza sumatrensis* | AA | 447
Length = 447

initn: 1765 initl: 1430 opt: 1770 Z-score: 3575.0 bits: 670.7 E(): 3.2e-190
Smith-Waterman score: 1770; 78.685% identity (90.930% similar) in 441 aa
overlap (6-445:7-447)

```

                10          20          30          40          50
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGA
        .....
Dad:AA  KPSTVPEIVLQPIKEISGAVNLPGSKLSNRILLLAALAEGTTIVDNLLNSDDVHYMLGA
        10          20          30          40          50          60

```

```

        60          70          80          90          100          110
2MEPSP  LRTLGLSVEADKAAKRAVVVGCCKGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxx
        .....
Dad:AA  LRTLGLNVEEDGAIKRAIVEGCGMFPVGKEAKDDIQLFLGNAGTAMRPLTAGVTAAGGN
        70          80          90          100          110          120

```

```

        120          130          140          150          160          170
2MEPSP  xxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSI
        .....
Dad:AA  SSYVLDGVPRMRERPIGDLVTGLKQLGADVDCSLGTNCPVVRVVGGLPGGRVKLSGSI
        130          140          150          160          170          180

```

```

        180          190          200          210          220          230
2MEPSP  SSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG
        .....
Dad:AA  SSQYLTALLMASPLALGDVEIEIIDKLISIPYVEMTLKLEWFGVSVEHSDSWDRFFIRG
        190          200          210          220          230          240

```

```

        240          250          260          270          280          290
2MEPSP  GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
        .....
Dad:AA  GQKYKSPGNAYVEGDASSASYFLAGAAITGGTITVEGCGTSSLQGDVKFAEVLGQMGAEV
        250          260          270          280          290          300

```

```

        300          310          320          330          340          350
2MEPSP  TWTETSVTVTGPPREPFGKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRV
        .....
Dad:AA  TWTENSVTVKGPPRDSSGRKHLRAVDVNMNKMPPDVAMTLAVVALYADGPTAIRDVASWRV
        310          320          330          340          350          360

```

```

        360          370          380          390          400          410
2MEPSP  KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEV
        .....
Dad:AA  KETERMIAICTELRKL GATVEEGPDYCVITPPEKLNVT AIDTYDDHRMAMAFSLAACADV
        370          380          390          400          410          420

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFDYFDVLSTFVKN
          .....
Dad:AA VTAIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFFPNYFDVLSTYSNH
          470          480          490          500          510

```

ACB37380.1 | *Calystegia hederacea* 5-enolpyruvylshikimate 3-phosphate synthase protein. | *Calystegia hederacea* | AA | 520
Length = 520
initn: 1753 initl: 1428 opt: 1766 Z-score: 3565.8 bits: 669.2 E(): 1e-189
Smith-Waterman score: 1766; 78.054% identity (90.045% similar) in 442 aa
overlap (5-445:79-520)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                                .....
Dad:AC TGKDSIFTAARSPLKVRASVATAKKPSMAPEEIVLQPIKEISGTVKLPGSKSLSNRILL
          50          60          70          80          90          100

```

```

          40          50          60          70          80          90
2MEPSP AALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKEE
          .....
Dad:AC AALSQGTTVVDNLLSSDDIHYMLGALRTLGLRVEEDSAIQRATVEGSGGLFPASNESNDE
          110          120          130          140          150          160

```

```

          100          110          120          130          140          150
2MEPSP VQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLG
          .....
Dad:AC IQLFLGNAGTAMRPLTAAVVAAGGNARYVLDGVPRMRERPIGDLVEGLKQLGADIDCFLG
          170          180          190          200          210          220

```

```

          160          170          180          190          200          210
2MEPSP TDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEM
          .....
Dad:AC TNCPPVRVIGKGGPLGGKVKLSGSSVSSQYLTALLMAAPLALGDVEIEIVDKLISVPYVEM
          230          240          250          260          270          280

```

```

          220          230          240          250          260          270
2MEPSP TLRIMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
          .....
Dad:AC TIKLIMERFGVSVHSDSWDRFLIRGGQKYKSPGNAYVEGDASSASYFLAGAAVTGGTITV
          290          300          310          320          330          340

```

```

          280          290          300          310          320          330
2MEPSP xxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPDV
          .....
Dad:AC ECGTSSLQGDVKFAEVLEKMGAEVSWTENSVTVKGPPRGPSGRKHLRGIDVNMNKMPDV
          350          360          370          380          390          400

```

```

          340          350          360          370          380          390
2MEPSP AMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKL
          .....
Dad:AC AMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGAATVEEGPDYCIITPPEKL
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      400      410      420      430      440
2MEPSP NVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AC NMTEIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFPDYFDVLSKFSKH
      470      480      490      500      510      520

```

AAT45244.1 | *Conyza canadensis* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Conyza canadensis* | AA | 523
Length = 523
initn: 1757 initl: 1426 opt: 1762 Z-score: 3557.7 bits: 667.7 E(): 2.9e-189
Smith-Waterman score: 1762; 77.602% identity (90.950% similar) in 442 aa
overlap (5-445:82-523)

```

                        10      20      30
2MEPSP                MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                        . : : . : : . : : . : : . : : . : :
Dad:AA VSCNVKNNKNPFKVSATSTKEKPSKAPEEIVLKPIQEISGTVHLPGSKSLSNRILL
      60      70      80      90     100     110

```

```

      40      50      60      70      80      90
2MEPSP AALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPV-EDAKEE
      . : : . : : . : : . : : . : : . : : . : : . : : . : :
Dad:AA AALSEGTTVVDNLLNSEDVHYMLGALRALGLNVEENSAIKRAIVEGCGGVFPVGKEAKDE
      120     130     140     150     160     170

```

```

      100     110     120     130     140     150
2MEPSP VQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLG
      . : : . : : . : : . : : . : : . : : . : : . : : . : :
Dad:AA IQLFLGNAGTAMRPLTAAVTAAGNSSYILDGVPRMRERPIGDLVTGLKQLGANVDCSLG
      180     190     200     210     220     230

```

```

      160     170     180     190     200     210
2MEPSP TDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEM
      . : : . : : . : : . : : . : : . : : . : : . : : . : :
Dad:AA TNCPPVRVVGSGGLPGGKVKLSGSISSQYLTSLMAAPLALGDVEIEIVDKLISVPYVEM
      240     250     260     270     280     290

```

```

      220     230     240     250     260     270
2MEPSP TLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
      . : : . : : . : : . : : . : : . : : . : : . : : . : :
Dad:AA TLKLMEFVGVSVEHSDTWDRFHVRRGGQKYKSPGNAYVEGDASSASYFLAGAAITGGTVTV
      300     310     320     330     340     350

```

```

      280     290     300     310     320     330
2MEPSP xxxxxxSLQGDVKFAEVLLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDV
      . : : . : : . : : . : : . : : . : : . : : . : : . : :
Dad:AA ECGTSSSLQGDVKFAEVLGQMGAEVTTWTE NSVTVKGPPRNSSGRGHLRPVDVNMNKMMPDV
      360     370     380     390     400     410

```

```

      340     350     360     370     380     390
2MEPSP AMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKL
      . : : . : : . : : . : : . : : . : : . : : . : : . : :
Dad:AA AMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGLATVEEGPDYCVITPPEKL
      420     430     440     450     460     470

```

	400	410	420	430	440	
2MEPSP	NVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTTFPDYFDVLSTFVKN					

Dad:AA	NVTAIDTYDDHRMAMAFSLAACAEVPVTIKDPGCTRKTTFPDYFEVLERYTKH					
	480	490	500	510	520	
AA0474.1 <i>Conyza sumatrensis</i> 5-enol-pyruvylshikimate-phosphate synthase protein. <i>Conyza sumatrensis</i> AA 446						
Length = 446						
initn: 1750 initl: 1420 opt: 1755 Z-score: 3544.6 bits: 665.0 E(): 1.6e-188						
Smith-Waterman score: 1755; 77.828% identity (90.724% similar) in 442 aa						
overlap (5-445:5-446)						
	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL					

Dad:AA	SKAPEEIVLKPIQEISGTVHLPGSKSLSNRILLLAALSEGTTVVDNLLNSDDVHYMLGAL					
	10	20	30	40	50	60
	70	80	90	100	110	
2MEPSP	RTLGLSVEADKAAKRAVVVGC GGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx					

Dad:AA	RALGLNVEENAAIKRAIVEGCGGVFPVGKEAKDEIQFLGNAGTAMRPLTAAVTAAGGNS					
	70	80	90	100	110	120
	120	130	140	150	160	170
2MEPSP	xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS					

Dad:AA	SYVLDGVPRMRERPIGDLVTGLKQLGANVDCSLGTNCPVVRVVGSGGLPGGKVKLSGSIS					
	130	140	150	160	170	180
	180	190	200	210	220	230
2MEPSP	SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG					

Dad:AA	SQYLTSLLMAAPLALGDVEIEIIDKLISVPYVEMTLKLMERFGVSEHSDTWDRFYVRGG					
	190	200	210	220	230	240
	240	250	260	270	280	290
2MEPSP	QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLMMGAKVT					

Dad:AA	QKYKSPGSAYVEGDASSASYFLAGAAITGGTVTVEGCGTSSLQGDVKFAEVLGQMGA EVT					
	250	260	270	280	290	300
	300	310	320	330	340	350
2MEPSP	WTETSVTVTGPPREPFGRKHLKAIDVNMNKM PDMATLAVVALFADGPTAIRDVASWRVK					

Dad:AA	WTENSVTVKGPPRNSSGRGELRPVDVNMNKM PDMATLAVVALYADGPTAIRDVASWRVK					
	310	320	330	340	350	360
	360	370	380	390	400	410
2MEPSP	ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP					

Dad:AA	ETERMIAICTELRKL GATVEEGPDYCVITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP					
	370	380	390	400	410	420

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYLSTFVKN
      .....
Dad:AA VTIKDPGCTRKTFPDYFEVLERYTKH
      430      440

```

ABV24481.1 | *Gossypium hirsutum* 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Gossypium hirsutum* | AA | 521
Length = 521
initn: 1740 initl: 1436 opt: 1752 Z-score: 3537.5 bits: 664.0 E(): 3.9e-188
Smith-Waterman score: 1752; 77.652% identity (90.519% similar) in 443 aa
overlap (4-445:79-521)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPKSKLSNRILL
                                .....
Dad:AB VVKNNKGFGSIKARSLKVSASTATAEKPSRASEIVLQPIKEISGTVKLPKSKLSNRILL
      50      60      70      80      90     100

```

```

      40      50      60      70      80      90
2MEPSP LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVCGGKFPVEDAK-E
      .....
Dad:AB LAALSEGTTVVENLLNSDDVHMLVALGKLGKLYVKHDSEKKQAIVEGCGGQFPVGKGEQ
      110     120     130     140     150     160

```

```

      100     110     120     130     140     150
2MEPSP EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
      .....
Dad:AB EIELFLGNAGTAMRPLTAAITAAGGNSSYVLDGVPRMRERPIGDLVTGLKQLGADVDCIL
      170     180     190     200     210     220

```

```

      160     170     180     190     200     210
2MEPSP GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYE
      .....
Dad:AB GTNCPVRIEGKGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISIPYE
      230     240     250     260     270     280

```

```

      220     230     240     250     260     270
2MEPSP MTLRLMERFVGKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      .....
Dad:AB MTMKLMERFGVTVEHTDSWDRFFIRGGQKYMSPGNAYVEGDASSASYFLAGAAVTGGTVT
      290     300     310     320     330     340

```

```

      280     290     300     310     320     330
2MEPSP xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPD
      .....
Dad:AB VEGCGTSSLQGDVKFAEVLEMMGAKVTWTENSVTVTGPPRNSSGRKHLRAIDVNMNKMMPD
      350     360     370     380     390     400

```

```

      340     350     360     370     380     390
2MEPSP VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      .....
Dad:AB VAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGAIVVEEGPDYCVITPPEK
      410     420     430     440     450     460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
          .....
Dad:AB  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIKDPGCTRKTFPDYFEVLDRVTKH
          470          480          490          500          510          520

```

ABY61050.1 | *Gossypium hirsutum* 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Gossypium hirsutum* | AA | 521
Length = 521
initn: 1739 initl: 1435 opt: 1751 Z-score: 3535.5 bits: 663.6 E(): 5.1e-188
Smith-Waterman score: 1751; 77.427% identity (90.519% similar) in 443 aa
overlap (4-445:79-521)

```

                                10          20          30
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSKLSNRILL
          .....
Dad:AB  VVKNNKGFGSIKVRSLKVSASTATAEKPSRASEIVLQPIKEISGTVKLPKSKLSNRILL
          50          60          70          80          90          100

```

```

                40          50          60          70          80          90
2MEPSP  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVEDAK-E
          .....
Dad:AB  LAALSEGTTVVENLLNSDDVHMLVALGKLGKLYVKHDSEKKQAIVEGCGGFVPVGKGEQG
          110          120          130          140          150          160

```

```

                100          110          120          130          140          150
2MEPSP  EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
          .....
Dad:AB  EIELFLGNAGTAMRPLTAAITAAGGNSSYVLDGVPRMRERPIGDLVTGLKQLGADVDCFL
          170          180          190          200          210          220

```

```

                160          170          180          190          200          210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
          .....
Dad:AB  GTNCPVRIEGKGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISIPYVE
          230          240          250          260          270          280

```

```

                220          230          240          250          260          270
2MEPSP  MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
          .....
Dad:AB  MTIKLMEFVGTVTEHTDSWDRFFIRGGQKYMSPGNAYVEGDASSASYFLAGAAVTGGTVT
          290          300          310          320          330          340

```

```

                280          290          300          310          320          330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
          .....
Dad:AB  VEGCGTSSLQGDVKFAEVLEMMGAKVTWTKNSVTVTGPPRNPSGRKHLRAIDVNMNKMPD
          350          360          370          380          390          400

```

```

                340          350          360          370          380          390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
          .....
Dad:AB  VAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGAATVEEGPDFCVITPPEK
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
          .....
Dad:AB  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIKDPGCTRKTFFPDYFEVLARVTKH
          470          480          490          500          510          520

```

AAA33699.1|*Petunia x hybrida* protein (*P.hybrida* 5-enolpyruvylshikimate 3-phosphate synthase mRNA, complete cds.).|*Petunia x hybrida*|AA|516
Length = 516
initn: 1741 initl: 1406 opt: 1747 Z-score: 3527.4 bits: 662.1 E(): 1.4e-187
Smith-Waterman score: 1747; 77.778% identity (90.023% similar) in 441 aa
overlap (6-445:76-516)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                              .....
Dad:AA  MLVLKKDSIFMQKFCSFRISASVATAQKPSEIVLQPIKEISGTVKLPGSKSLSNRILLLA
          50          60          70          80          90          100

          40          50          60          70          80          90
2MEPSP  ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
          .....
Dad:AA  ALSEGTTVVDNLLSSDDIHYMLGALKTLGLHVEEDSANQRAVVEGCGGLFPVGKESKEEI
          110          120          130          140          150          160

          100          110          120          130          140          150
2MEPSP  QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPVRMRERPIGDLVVGLKQLGADVDCFLGT
          .....
Dad:AA  QLFLGNAGTAMRPLTAAVTVAGGNSRYVLGDGVPVRMRERPISDLVDGLKQLGAEVDCFLGT
          170          180          190          200          210          220

          160          170          180          190          200          210
2MEPSP  DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
          .....
Dad:AA  KCPVVRIVSKGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVPYVEMT
          230          240          250          260          270          280

          220          230          240          250          260          270
2MEPSP  LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
          .....
Dad:AA  LKLMERFGISVEHSSSWDRFFVRGGQKYKSPGKAFVEGDASSASYFLAGAAVTGGTITVE
          290          300          310          320          330          340

          280          290          300          310          320          330
2MEPSP  xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA
          .....
Dad:AA  GCGTNSLQGDVKFAEVLEKMGAEVTWTENSVTVKGPPRSSSGRKHLRAIDVNMNKMMPDVA
          350          360          370          380          390          400

          340          350          360          370          380          390
2MEPSP  MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
          .....
Dad:AA  MTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLKATVEEGPDYCIITPPEKLN
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP  VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
          :: :::::::::::::::::::::::::::::::::::::::::::::::::::: . :.
Dad:AA  VTDIDTYDDHRMAMAFSLAACADVPVTINDPGCTRKTFFPNYFDVLQQYSKH
          470          480          490          500          510

```

CAB69241.1|*Petunia x hybrida* protein (Sequence 1 from Patent WO9854330.
).|*Petunia x hybrida*|AA|516
 Length = 516
 initn: 1741 initl: 1406 opt: 1747 Z-score: 3527.4 bits: 662.1 E(): 1.4e-187
 Smith-Waterman score: 1747; 77.778% identity (90.023% similar) in 441 aa
 overlap (6-445:76-516)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                                ::::::::::::::::::::::::::::::::::::
Dad:CA  MLVLKKDSIFMQKFCSFRISASVATAQKPSEIVLQPIKEISGTVKLPGSKSLSNRILLLA
          50          60          70          80          90          100

          40          50          60          70          80          90
2MEPSP  ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
          :::::::::::::::::::::::::::::::::::: :: :: .:::: :::: :: .::::.
Dad:CA  ALSEGTTVVDNLLSSDDIHYMLGALKTLGLHVEEDSANQRAVVEGCGGLFPVGKESKEEI
          110          120          130          140          150          160

          100          110          120          130          140          150
2MEPSP  QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPVRMRERPIGDLVVGLKQLGADVDCFLGT
          :::::::::: :: :... . . . :::::::::::::::::::::: ::::::::::::::::::::
Dad:CA  QLFLGNAGTAMRPLTAAVTVAGGNSRYVLGDGVPVRMRERPISDLVDGLKQLGAEVDCFLGT
          170          180          190          200          210          220

          160          170          180          190          200          210
2MEPSP  DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
          ::::: . :::::::::::::::::::::: . . . . ::::::::::::::::::::::::::::
Dad:CA  KCPVVRIVSKGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVPYVEMT
          230          240          250          260          270          280

          220          230          240          250          260          270
2MEPSP  LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
          ::::::::::::::::::::::::::::::::::: .: :::::::::::::::::::::::::::: . .
Dad:CA  LKLMERFGISVEHSSSWDRFFVRGGQKYKSPGKAFVEGDASSASYFLAGAAVTGGTITVE
          290          300          310          320          330          340

          280          290          300          310          320          330
2MEPSP  xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA
          . :::::::::::::: :::::::::::::: ::::: ::::::::::::::::::::::::::::
Dad:CA  GCGTNSLQGDVKFAEVLEKMGAEVTWTENSVTVKGPPRSSSGRKHLRAIDVNMNKMMPDVA
          350          360          370          380          390          400

          340          350          360          370          380          390
2MEPSP  MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
          :::::::::: :::::::::::::::::::::: ::::: ::::::::::::::::::::::::::::
Dad:CA  MTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLKLGATVEEGPDYCIITPPEKLN
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP  VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
          :: :::::::::::::::::::::::::::::::::::::::::::::::::::: . :.
Dad:CA  VTDIDTYDDHRMAMAFSLAACADVPVTINDPGCTRKTFFPNYFDVLQQYSKH
          470          480          490          500          510

```

ABE77393.4 | *Allium macrostemon* EPSP synthase protein. | *Allium macrostemon* | AA | 522
Length = 522
initn: 1725 initl: 1417 opt: 1727 Z-score: 3486.9 bits: 654.6 E(): 2.6e-185
Smith-Waterman score: 1727; 76.871% identity (90.930% similar) in 441 aa
overlap (6-445:82-522)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                                :: ::::::::::::::::::::::::::::::::::::
Dad:AB  RIGNCRRASVRVLASLAAVENAAEKVAVMPEITLQPIKEITGTVNLPGSKSLSNRILLLA
          60          70          80          90          100          110

```

```

          40          50          60          70          80          90
2MEPSP  ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPV-EDAKEEV
          :: :::::::::::::::::::: :::::::::::::::::::: :. :::::::::: :: :: :: :: :: :: ::
Dad:AB  ALAEGTTIVDNLLNSDDVSYMLAALKTLGLSVEDDRVNKRAVVGSGGLFPVGKESQKEV
          120          130          140          150          160          170

```

```

          100          110          120          130          140          150
2MEPSP  QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPVRMRERPIGDLVVGLKQLGADVDCFLGT
          :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Dad:AB  QLFLGNAGTAMRPLTAAVTAAGGNASYILDGVPVRMRERPIGDLVVGLKQLGADVDCFLGT
          180          190          200          210          220          230

```

```

          160          170          180          190          200          210
2MEPSP  DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
          :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Dad:AB  DCPVVRVDANGGLPGGKVKLSDSISSQYLTALLMAAPLALGDVEIEIIDKLISIPYVEMT
          240          250          260          270          280          290

```

```

          220          230          240          250          260          270
2MEPSP  LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxx
          :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Dad:AB  LKLMEFVGVDHSSWTDRFFIKGGQKYKSPGNAYVEGDASSASYFLAGAAVTGGTVTVE
          300          310          320          330          340          350

```

```

          280          290          300          310          320          330
2MEPSP  xxxxxxSLQGDVKFAEVLNMGAKVTTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA
          .. :::::::::::::::::::: :::::::::::::::::::: :. ::::::::::::::::::::
Dad:AB  GCGTSSLQGDVKFAEVLNMGAKVTTWTENSVTVTGPPQDPQKKRLKAVDVNMNKMMPDVA
          360          370          380          390          400          410

```

```

          340          350          360          370          380          390
2MEPSP  MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
          :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Dad:AB  MTLAVVALYADGPTTIRDVASWRVKETERMIAICTELRKLKATVVEGPDYCIITPPEKLN
          420          430          440          450          460          470

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
      :: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:AB VTEIDTYDDHRMAMAFSLAACSDVPVTIKDPGCTRKTFFPDYFEVLERYAKH
          480          490          500          510          520

```

BAE99170.1 | *Arabidopsis thaliana* 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase protein. | *Arabidopsis thaliana* | AA | 520
Length = 520
initn: 1688 initl: 1382 opt: 1716 Z-score: 3464.6 bits: 650.5 E(): 4.5e-184
Smith-Waterman score: 1716; 75.169% identity (90.971% similar) in 443 aa
overlap (4-445:78-520)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                        :::::::::::::: ::::::::::::::
Dad:BA SWGLKKSGMTLIGSELRLPKVMSSVSTA EKASEIVLQPIREISGLIKLPGSKSLSNRILL
          50          60          70          80          90          100

```

```

          40          50          60          70          80          90
2MEPSP LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKE
      :::::::::::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Dad:BA LAALSEGTTVVDNLLNSDDINMYLDALKRLGLNVETDSENNRAVVEGCGGIFPASIDSKS
          110          120          130          140          150          160

```

```

          100          110          120          130          140          150
2MEPSP EVQLFNLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
      ..... :: :::: .. :::::::::::::::::::::::::::::::::::::: ::
Dad:BA DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVVGLKQLGADVECTL
          170          180          190          200          210          220

```

```

          160          170          180          190          200          210
2MEPSP GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      :::::::::: :::::::::::::: :: .. .. ::::::::::::::
Dad:BA GTNCPVVRVANGGLPGGKVKLSGSISSQYLTALLMSAPLALGDVEIEIVDKLISVPYVE
          230          240          250          260          270          280

```

```

          220          230          240          250          260          270
2MEPSP MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: ::
Dad:BA MTLKLMERFGVSVEHSDSWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
          290          300          310          320          330          340

```

```

          280          290          300          310          320          330
2MEPSP xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPD
      .. :::::::::::::: :::::::::::::: :::::::::::::: :: ::::::::::::::
Dad:BA VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPPRDAFGMRHLRAIDVNMNKMMPD
          350          360          370          380          390          400

```

```

          340          350          360          370          380          390
2MEPSP VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      :::::::::::::: :::::::::::::: :::::::::::::: :: ::::::::::::::
Dad:BA VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGA TVEEGSDYCVITPPKK
          410          420          430          440          450          460

```

AA25438.1 | *Arabidopsis thaliana* At2g45300 protein. | *Arabidopsis thaliana* | AA | 520
Length = 520
initn: 1688 init1: 1382 opt: 1716 Z-score: 3464.6 bits: 650.5 E(): 4.5e-184
Smith-Waterman score: 1716; 75.169% identity (90.971% similar) in 443 aa
overlap (4-445:78-520)

VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
Dad::AA VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGLATVEEGSDYCVITPPKK

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
        .....
Dad:AA  VKTAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH
        470          480          490          500          510          520

```

AAB82633.1 | *Arabidopsis thaliana* 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase protein. | *Arabidopsis thaliana* | AA | 520
Length = 520
initn: 1688 initl: 1382 opt: 1716 Z-score: 3464.6 bits: 650.5 E(): 4.5e-184
Smith-Waterman score: 1716; 75.169% identity (90.971% similar) in 443 aa
overlap (4-445:78-520)

```

                                10          20          30
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
        :.....
Dad:AA  SWGLKKSGMTLIGSELRLPKVMSSVSTA EKASEIVLQPIREISGLIKLPGSKSLSNRILL
        50          60          70          80          90          100

```

```

                40          50          60          70          80          90
2MEPSP  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKE
        .....
Dad:AA  LAALSEGTTVVDNLLNSDDINMYLDALKRLGLNVETDSENNRAVVEGCGGIFPASIDSKS
        110         120         130         140         150         160

```

```

                100         110         120         130         140         150
2MEPSP  EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
        .....
Dad:AA  DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVVGLKQLGADVECTL
        170         180         190         200         210         220

```

```

                160         170         180         190         200         210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
        .....
Dad:AA  GTNCPVVRVANGGLPGGKVKLSGSISSQYLTALLMSAPLALGDVEIEIVDKLISVPYVE
        230         240         250         260         270         280

```

```

                220         230         240         250         260         270
2MEPSP  MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
        .....
Dad:AA  MTLKLMERFGVSVEHSDSWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
        290         300         310         320         330         340

```

```

                280         290         300         310         320         330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPD
        .....
Dad:AA  VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPPRDAFGMRHLRAIDVNMNKMMPD
        350         360         370         380         390         400

```

```

                340         350         360         370         380         390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
        .....
Dad:AA  VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGA TVEEGSDYCVITPPKK
        410         420         430         440         450         460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVK
      .... :.....
Dad:AA  VKTAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFFPDYFQVLERITKH
                470          480          490          500          510          520

```

BAA32276.1|*Oryza sativa* Japonica Group 3-phosphoshikimate 1-carboxyvinyltransfe
rase protein.|*Oryza sativa Japonica Group*|AA|391

Length = 391

initn: 1713 initl: 1584 opt: 1714 Z-score: 3462.5 bits: 649.7 E(): 5.9e-184

Smith-Waterman score: 1714; 85.934% identity (93.862% similar) in 391 aa
overlap (56-445:1-391)

```

                30          40          50          60          70          80
2MEPSP  SLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVVGCGKF
      :: :.....
Dad:BA                                     MLEALKALGLSVEADKVAKRAVVVVGCGKF
                                   10          20          30

```

```

                90          100          110          120          130          140
2MEPSP  PVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQL
      ::: :.....
Dad:BA  PVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQL
                40          50          60          70          80          90

```

```

                150          160          170          180          190          200
2MEPSP  GADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDK
      :::::.....
Dad:BA  GADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDK
                100          110          120          130          140          150

```

```

                210          220          230          240          250          260
2MEPSP  LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGA
      :::::.....
Dad:BA  LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGA
                160          170          180          190          200          210

```

```

                270          280          290          300          310          320
2MEPSP  AIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGKRHLKAID
      ::. . . :.....
Dad:BA  AITGGTVTVQCGTTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPREPYGKKHLKAID
                220          230          240          250          260          270

```

```

                330          340          350          360          370          380
2MEPSP  VNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDY
      :::::.....
Dad:BA  VNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDY
                280          290          300          310          320          330

```

```

                390          400          410          420          430          440
2MEPSP  CIITPPEKLNVTAEIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVK
      :::::.....
Dad:BA  CIITPPEKLNITAEIDTYDDHRMAMAFSLAACADVPVTIRDPGCTRKTFFPNYFDVLSTFVR
                340          350          360          370          380          390

```

2MEPSP N
:
Dad:BA N

	10	20	30
2MEPSP	MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILL		
Dad:CA	SWGLKKSGTMLNGSVIRPVKVTASVSTSEKASEIVLQPIREISGLIKLPGSKSLSNRILL		
	50	60	70
	40	50	60
2MEPSP	LAALSEGTTVVDNLLNSSEVDHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKE		
Dad:CA	LAALSEGTTVVDNLLNSDDINYMLDALKKLGLNVERDSVNNRAVVEGCGGIFPASLDSKS		
	110	120	130
	100	110	120
2MEPSP	EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGKQLGADVDCFL		
Dad:CA	DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVVGKQLGADVCTL		
	170	180	190
	160	170	180
2MEPSP	GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVE		
Dad:CA	GTNCPVVRVNANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVPYVE		
	230	240	250
	220	230	240
2MEPSP	MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxx		
Dad:CA	MTLKLMEFRFGVSAEHSDSWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT		
	290	300	310
	280	290	300
2MEPSP	xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMKMPD		
Dad:CA	VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAVDVNMNMKMPD		
	350	360	370
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK		
Dad:CA	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCVITPPAK		
	410	420	430
	340	350	360
2MEPSP			

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .. :.....
Dad:CA  VKPAEIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFPDYFQVLESITKH
                470          480          490          500          510

```

AAA34071.1 | *Nicotiana tabacum* 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Nicotiana tabacum* | AA | 518
Length = 518
initn: 1713 initl: 1389 opt: 1711 Z-score: 3454.5 bits: 648.6 E(): 1.6e-183
Smith-Waterman score: 1711; 75.964% identity (89.116% similar) in 441 aa
overlap (6-445:78-518)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                        .....
Dad:AA  LWVSKEDSVLRVAKSPFRISASVVTAQKPNEIVLQPIKDISGTVKLPGSKSLSNRILLLA
                50          60          70          80          90          100

```

```

                40          50          60          70          80          90
2MEPSP  ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
      :: :..... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  ALSKGRTVVDNLLSSDDIHYMLGALKTLGLHVEDDNNQRAIVEGCGGQFPVGKKSEEEI
                110          120          130          140          150          160

```

```

                100          110          120          130          140          150
2MEPSP  QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPVRMRERPIGDLVVGLKQLGADVDCFLGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  QLFLGNAGTAMRPLTAAVTVAGGHSRYVLGDGVPVRMRERPIGDLVDGLKQLGAEVDCFLGT
                170          180          190          200          210          220

```

```

                160          170          180          190          200          210
2MEPSP  DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  NCPVVRIVSKGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVPYVEMT
                230          240          250          260          270          280

```

```

                220          230          240          250          260          270
2MEPSP  LRLMERFVGKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  LKLMERFVGSVEHTSSWDKFLVRGGQKYKSPGKAYVEGDASSASYFLAGAAVTGGTVTVE
                290          300          310          320          330          340

```

```

                280          290          300          310          320          330
2MEPSP  xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMKMPDVA
      .. :..... : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  GCGTSSLQGDVKFAEVLEKMGAEVTVTENSVTVKGPPRNSSGMKHLRAVDVNMNMKMPDVA
                350          360          370          380          390          400

```

```

                340          350          360          370          380          390
2MEPSP  MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  MTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLGATVVEGSDYCIITPPEKLN
                410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP  VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
          :: :::::::::::::::::::::::::::::::::::::::::::::::::::: . :.
Dad:AA  VTEIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFFPNYFDVLQQYSKH
          470          480          490          500          510

```

AAL65913.1 | *Orychophragmus violaceus* 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Orychophragmus violaceus* | AA | 518
Length = 518
initn: 1683 initl: 1375 opt: 1711 Z-score: 3454.5 bits: 648.6 E(): 1.6e-183
Smith-Waterman score: 1711; 75.395% identity (90.293% similar) in 443 aa
overlap (4-445:76-518)

```

                                10          20          30
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                                :. :::::::::::::: . ::::::::::::::
Dad:AA  SWGLKKSGMMLNDSVIRPVTVTASVSTA EKASEIVLQPIKEISGLIKLPGSKSLSNRILL
          50          60          70          80          90          100

```

```

          40          50          60          70          80          90
2MEPSP  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKE
          :::::::::::::::::::::::::::: :: ::::: :. : :::: :::: :... :...
Dad:AA  LAALSEGTTVVDNLLNSDDIN YMLDALKKLG LNV ERDSENNRAVVEGCGGIFPASVDSKS
          110          120          130          140          150          160

```

```

          100          110          120          130          140          150
2MEPSP  EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
          ... ::::: ::: :... :... :. :::::::::::::::::::::::::::::::::::: :
Dad:AA  DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVVGLKQLGADVECTL
          170          180          190          200          210          220

```

```

          160          170          180          190          200          210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
          ::::::::::: :::::::::::::::::::: .. .. : ::::::::::::::::::::
Dad:AA  GTNCPVVRVNANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVPYVE
          230          240          250          260          270          280

```

```

          220          230          240          250          260          270
2MEPSP  MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
          ::::::::::: :::::::::::::::::::: : :::::::::::::::::::: . .
Dad:AA  MTLKLMERFGVSAEHSESWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
          290          300          310          320          330          340

```

```

          280          290          300          310          320          330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPD
          .. :::::::::::::: :::::::::::::::::::: :. : : ::::::::::::::::::::
Dad:AA  VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAVDVNMNKMMPD
          350          360          370          380          390          400

```

```

          340          350          360          370          380          390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
          :::::::::::::: :::::::::::::::::::: : : :::::::::::::::::::: :
Dad:AA  VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKL GATVEEGSDYCVITPPAK
          410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  VKPAEIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFPDYFQVLESITKH
                470          480          490          500          510

```

AAT45245.1 | *Conyza canadensis* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Conyza canadensis* | AA | 454
Length = 454
initn: 1699 initl: 1370 opt: 1704 Z-score: 3441.2 bits: 645.9 E(): 9e-183
Smith-Waterman score: 1704; 75.339% identity (90.045% similar) in 442 aa
overlap (5-445:13-454)

```

                10          20          30          40          50
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSED
      . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  ANTTKEKPSKAPEEIVLKPIQEISGTVHLPGSKSLSNRILLLAALSEGTTVVDNLLNSDD
                10          20          30          40          50          60

```

```

                60          70          80          90          100          110
2MEPSP  VHYMLGALRTLGLSVEADKAAKRAVVVGCCKFPV-EDAKEEVQLFLGNAGIAMRSLxxx
      . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  VHYMLGALRALGLNIEENAAIKRAIVEGCGGLFPVGKEAKDEIQLFLGNAGTAMRTLTA
                70          80          90          100          110          120

```

```

                120          130          140          150          160          170
2MEPSP  xxxxxxxxxxxYVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGK
      . . . : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  VTAAGGNLSYVLDGVPRMRERPIGDLVTGLKQLGVNVDCSLGTNCPVHVVGSGGLPGGK
                130          140          150          160          170          180

```

```

                180          190          200          210          220          230
2MEPSP  VKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSW
      . : : : : : : : : : . . . : : : : : : : : : : : : : :
Dad:AA  VKLSGSISSVYLTSLMAAPLALGDVEIEIIDKLISVPYVRMTLKLMMQRFVGSVEHSDTL
                190          200          210          220          230          240

```

```

                240          250          260          270          280          290
2MEPSP  DRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVL
      . : : . . : : : : : : : : : : . . . : : : : : : : :
Dad:AA  DRFHVRGGQKYKSPGNAYVESDASSASYFLAGAAITGGTVTVEGCGTSSLQGDVKFAEVL
                250          260          270          280          290          300

```

```

                300          310          320          330          340          350
2MEPSP  EMMGAKVTWTETSVTVTGPPREPFGKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIR
      . : : . : : : : : : : : : . : : . : : : : : : : : :
Dad:AA  GQMGAEVTTWTVNSVTGKPPRNSSGRGHLRPVDVNMNKMPPDVAMTLAVVALYADGPTAIR
                310          320          330          340          350          360

```

```

                360          370          380          390          400          410
2MEPSP  DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFS
      . : : : : : : : : : . : : : : : : : : : . : : : : :
Dad:AA  DVASWRVKETERMIAICTELRKLGAIVVEEGPDYCVTTTPEKLNVTADTYDDHRMAMTFS
                370          380          390          400          410          420

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          420          430          440
2MEPSP LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .....
Dad:AA LAACADVPTIKDPGCTRKSPFDYFEVLERYTKH
          430          440          450

```

CAA29828.1 | *Arabidopsis thaliana* EPSP protein. | *Arabidopsis thaliana* | AA | 520
Length = 520
initn: 1675 initl: 1369 opt: 1703 Z-score: 3438.3 bits: 645.6 E(): 1.3e-182
Smith-Waterman score: 1703; 74.718% identity (90.519% similar) in 443 aa
overlap (4-445:78-520)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                        .....
Dad:CA SWGLKKSGMTLIGSELRLPKVMSSVSTA EKASEIVLQPIREISGLIKLPGSKSLSNRILL
          50          60          70          80          90         100

```

```

          40          50          60          70          80          90
2MEPSP LAALSEGTTTVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKE
      .....
Dad:CA LAALSEGTTTVDNLLNSDDIN YMLDALKRLGLNVETDSENNRAVVEGCGGIFPASIDSKS
          110         120         130         140         150         160

```

```

          100         110         120         130         140         150
2MEPSP EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
      .....
Dad:CA DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVVGLKQLGADV ECTL
          170         180         190         200         210         220

```

```

          160         170         180         190         200         210
2MEPSP GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      .....
Dad:CA GTNCPVVRV NANGGLPGGKVKLSGSISSQYLTALLMSAPLALGDVEIEIVDKLISVPYVE
          230         240         250         260         270         280

```

```

          220         230         240         250         260         270
2MEPSP MTLRLMERFGVKA EHSWSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      .....
Dad:CA MTLKLMERFGVSVEHSWSWDRFFVKGGQKYKSPGNAYVEGDASSACYFLAGAAITGETVT
          290         300         310         320         330         340

```

```

          280         290         300         310         320         330
2MEPSP xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFG RKHLKAIDVNMNKMPD
      .....
Dad:CA VEGCGTTS LQGDVKFAEVLEKMGCKVSWTENSVTVTGPPRDAFGMRHLRAIDVNMNKMPD
          350         360         370         380         390         400

```

```

          340         350         360         370         380         390
2MEPSP VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      .....
Dad:CA VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKL GATVEEGSDYCVITPPKK
          410         420         430         440         450         460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
        .... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA  VKTAEIDTYDDHRMAMAFSLAACADVPITINDSGCTRKTFPDYFQVLERITKH
        470          480          490          500          510          520

```

AA40475.1 | *Conyza sumatrensis* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Conyza sumatrensis* | AA | 446
Length = 446
initn: 1696 initl: 1369 opt: 1701 Z-score: 3435.3 bits: 644.8 E(): 1.9e-182
Smith-Waterman score: 1701; 75.113% identity (90.271% similar) in 442 aa
overlap (5-445:5-446)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  SKAPEEIVLKPIQEISGTVHLPKSLSNRILLLAALSEGTTVVDNLLNSDDVHYMLGAL
        10          20          30          40          50          60

```

```

                70          80          90          100         110
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
        ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  RAVGLNVEENAAIKRAIVEGCGGVFPVGKEAKDEIQLFLGNSGTAMRTLTAAITAAGGNS
        70          80          90          100         110         120

```

```

               120          130          140          150          160          170
2MEPSP  xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
        ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  SYVLDGVPRMRERPIGDLVTGLKQLGVNVDCSLGTNCPVVRVVGSGGLPGGKVRLSGSIS
        130          140          150          160          170          180

```

```

               180          190          200          210          220          230
2MEPSP  SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGG
        : : .. .. : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  SVYLTSLMLAAPVALGDVEIEIIDKLISVPYVQMTLKLKMKQFGVSVEHSDTLDRFHVRRG
        190          200          210          220          230          240

```

```

               240          250          260          270          280          290
2MEPSP  QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
        ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  QKYKSPGNAYVEGDASSASYFLAGAAITGGTVTVEGCGTSSLQGDVKFAEVLGQMGAEVT
        250          260          270          280          290          300

```

```

               300          310          320          330          340          350
2MEPSP  WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
        ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  WTENSVTVKGPPRNSSGRGHLRPVDVNMNKMPPDVAMTLAVVALYADGRTAIRDVASWRVK
        310          320          330          340          350          360

```

```

               360          370          380          390          400          410
2MEPSP  ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVP
        ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  ETERMIAICTELRKLGAIVVEEGPDYCVITPPEKLNVTADTYDDHRMAMTFSLAACADVP
        370          380          390          400          410          420

```

420	430	440	
2MEPSP	VTIRDPGCTRKTFPDYFDVLSTFVKN		
		
Dad:AA	VTIKDPGCTRKSFPDYFEVLERYTKH		
	430	440	
AAS80163.1 <i>Brassica rapa</i> 5-enolpyruvylshikimate-3-phosphate synthase protein. <i>Brassica rapa</i> AA 514			
Length = 514			
initn: 1675 init1: 1375 opt: 1701 Z-score: 3434.3 bits: 644.8 E(): 2.2e-182			
Smith-Waterman score: 1701; 74.944% identity (90.068% similar) in 443 aa			
overlap (4-445:72-514)			
		10	20
2MEPSP		MAGAE	EIVLQPIKEISGTVKLP
	
Dad:AA	QISSWGLKKSNNGSVIRPVKVMASVSTA	EIVLQPIRQISGLIKLP	PGSKSLSNRILL
	50	60	70
		80	90
		100	
	40	50	60
2MEPSP	LAALSEGTTVDNLLNS	EDVHYMLGALRTLGLS	VEADKAAKRAVVVGC

Dad:AA	LAALSEGTTVDNLLNS	DDINYMILDALNKLGLN	VERDSENNRAVVEGCGGIFPASLDSKG
	110	120	130
		140	150
		160	
	100	110	120
2MEPSP	EVQLF	LG	NAGIAMRSLxxxxxxxxxxxxxYVLDG

Dad:AA	DIELY	LG	NAGTAMRPLTAAVTAAGGNASYVLDG
	170	180	190
		200	210
		220	
	160	170	180
2MEPSP	GTDCPPVRVNGIGGL	PGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVE	

Dad:AA	GTNCP	PPVRVNANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVPYVE	
	230	240	250
		260	270
		280	
	220	230	240
2MEPSP	MTLRLMERFGVKA	EHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx	

Dad:AA	MTLKL	MERFGVSAEHSDSWDRFFVKG	GQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
	290	300	310
		320	330
		340	
	280	290	300
2MEPSP	xxxxxxxSLQGDVK	FAEVLEMMGAKVTWTETSVTVTGPPREP	FGRKHLKAIDVNMNKM

Dad:AA	VEGCG	TTSLQGDVKFAEVLEKMGCKVSWTENS	SVTVTGPSRDAFGMRHLRAVDVNMNKM
	350	360	370
		380	390
		400	
	340	350	360
2MEPSP	VAMTLAVVALFADG	P	TAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK

Dad:AA	VAMTLAVVALFADG	P	TTIRDVASWRVKETERMIAICTELRKL
	410	420	430
		440	450
		460	

400 410 420 430 440
 2MEPSP LNVTAIDITYDDHRMAMAFSLAACAEVPTIRDPGCTRKTFPDYFDVLSTFVKN
 . . . :
 Dad:AA VKPAEIDTYDDHRMAMAFSLACADVPTIKDPGCTRKTFPDYQVLESITKH
 470 480 490 500 510

2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKLSNRRILL
 Dad: AA SWGLKKSDLMLNGSEIRPVKVRASVSTAEKASEIVLQPIREISGLIKLPGSKLSNRRILL

40 50 60 70 80 90
 2MEPSP LAALSEGTTVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKE
 :
 Dad: AA LAALSEGTTVDNLLNSDDINYM L DALKILGLNVETHSENNRAVVEGCGGVFPASIDSKS
 110 120 130 140 150 160

100 110 120 130 140 150
 2MEPSP EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
 :
 Dad:AA DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVVGLKQLGADVECTL
 170 180 190 200 210 220

2MEPSP MTLRLMERFGVKA²²⁰EHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxx
 :.
 Dad:AA MTLKL²⁹⁰MERFGVSAEHS³⁰⁰ESWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT³⁴⁰

2MEPSP xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKM
Dad:AA VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAIDVNMNKM

340 350 360 370 380 390
 2MEPSP VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
 :
 Dad:AA VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGLATVEEGSDYCVITPPKK
 410 420 430 440 450 460

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : .
Dad:AA  VKPAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH
                470          480          490          500          510          520

```

AAG29739.1 | *Arabidopsis thaliana* 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase, putative protein. | *Arabidopsis thaliana* | AA | 521
Length = 521
initn: 1672 initl: 1372 opt: 1700 Z-score: 3432.2 bits: 644.5 E(): 2.9e-182
Smith-Waterman score: 1700; 74.718% identity (90.293% similar) in 443 aa
overlap (4-445:79-521)

```

                                10          20          30
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
      . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  SWGLKKSDLMLNGSEIRPVKVRASVSTA EKASEIVLQPIREISGLIKLPGSKSLSNRILL
                50          60          70          80          90          100

```

```

                40          50          60          70          80          90
2MEPSP  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKE
      . : : : : : : : : : : : : : : : : : : : : : : : : : : .
Dad:AA  LAALSEGTTVVDNLLNSDDIN YMLDALKILGLNVETHSENNRAVVEGCGVFPASIDSKS
                110         120         130         140         150         160

```

```

                100         110         120         130         140         150
2MEPSP  EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
      . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVVGLKQLGADVECTL
                170         180         190         200         210         220

```

```

                160         170         180         190         200         210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      . : : : : : : : : : : : : : : : : : : : : : : : : : : .
Dad:AA  GTNCPVVRV NANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIVDKLISVPYVE
                230         240         250         260         270         280

```

```

                220         230         240         250         260         270
2MEPSP  MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      . : : : : : : : : : : : : : : : : : : : : : : : : : : .
Dad:AA  MTLKLME RFVSAEHSESWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
                290         300         310         320         330         340

```

```

                280         290         300         310         320         330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
      . : : : : : : : : : : : : : : : : : : : : : : : : : : .
Dad:AA  VEGCGTTS LQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAIDVNMNKMPD
                350         360         370         380         390         400

```

```

                340         350         360         370         380         390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      . : : : : : : : : : : : : : : : : : : : : : : : : : : .
Dad:AA  VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKL GATVEEGSDYCVITPPKK
                410         420         430         440         450         460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . : : : : : : : : : : : : : : : : : : : : : : : : . . .
Dad:AA  VKPAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH
                470          480          490          500          510          520

```

AAK25934.1 | *Arabidopsis thaliana* putative 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase protein. | *Arabidopsis thaliana* | AA | 521
Length = 521
initn: 1672 initl: 1372 opt: 1700 Z-score: 3432.2 bits: 644.5 E(): 2.9e-182
Smith-Waterman score: 1700; 74.718% identity (90.293% similar) in 443 aa
overlap (4-445:79-521)

```

                                10          20          30
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  SWGLKKSDLMLNGSEIRPVKVRASVSTA EKASEIVLQPIREISGLIKLPGSKSLSNRILL
                50          60          70          80          90          100

```

```

                40          50          60          70          80          90
2MEPSP  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKE
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  LAALSEGTTVVDNLLNSDDIN YMLDALKILGLNVETHSENNRAVVEGCGVFPASIDSKS
                110         120         130         140         150         160

```

```

                100         110         120         130         140         150
2MEPSP  EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVL DGVPRMRERPIGDLVVGLKQLGADVDCFL
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  DIELYLGNAGTAMRPLTAAVTAAGGNASYVL DGVPRMRERPIGDLVVGLKQLGADV ECTL
                170         180         190         200         210         220

```

```

                160         170         180         190         200         210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      . : : : : : : : : : : : : : : : : : . . . : : : : : : : : :
Dad:AA  GTNCPVVRV NANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIVDKLISVPYVE
                230         240         250         260         270         280

```

```

                220         230         240         250         260         270
2MEPSP  MTLRLMERFGVKA EHSWSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      . : : : : : : : : : : : : : : : : : : : : : : : : : : . .
Dad:AA  MTLKLMERFGVSAEHS ESWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
                290         300         310         320         330         340

```

```

                280         290         300         310         320         330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
      . . : : : : : : : : : : : : : : : : : : . . : : : : : : : :
Dad:AA  VEGCGTTS LQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAIDVNMNKMPD
                350         360         370         380         390         400

```

```

                340         350         360         370         380         390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      . : : : : : : : : : : : : : : : : : . : : : : : : : : :
Dad:AA  VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKL GATVEEGSDYCVITPPKK
                410         420         430         440         450         460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : .
Dad:AA  VKPAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH
                470          480          490          500          510          520

```

AAG50661.1 | *Arabidopsis thaliana* 5-enolpyruvylshikimate-3-phosphate, putative protein. | *Arabidopsis thaliana* | AA | 521
Length = 521
initn: 1672 initl: 1372 opt: 1700 Z-score: 3432.2 bits: 644.5 E(): 2.9e-182
Smith-Waterman score: 1700; 74.718% identity (90.293% similar) in 443 aa
overlap (4-445:79-521)

```

                                10          20          30
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
      . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  SWGLKKSDLMLNGSEIRPVKVRASVSTA EKASEIVLQPIREISGLIKLPGSKSLSNRILL
                50          60          70          80          90          100

```

```

                40          50          60          70          80          90
2MEPSP  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKE
      . : : : : : : : : : : : : : : : : : : : : : : : : : : .
Dad:AA  LAALSEGTTVVDNLLNSDDIN YMLDALKILGLNVETHSENNRAVVEGCGVFPASIDSKS
                110         120         130         140         150         160

```

```

                100         110         120         130         140         150
2MEPSP  EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  DIELYLG NAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVVGLKQLGADV ECTL
                170         180         190         200         210         220

```

```

                160         170         180         190         200         210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      . : : : : : : : : : : : : : : : : : : : : : : : : : : .
Dad:AA  GTNCPVVRV NANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIVDKLISVPYVE
                230         240         250         260         270         280

```

```

                220         230         240         250         260         270
2MEPSP  MTLRLMERFGVKA EHSWSDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      . : : : : : : : : : : : : : : : : : : : : : : : : : : .
Dad:AA  MTLKLMERFGVSAEHS ESWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
                290         300         310         320         330         340

```

```

                280         290         300         310         320         330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
      . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  VEGCGTTS LQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAIDVNMNKMPD
                350         360         370         380         390         400

```

```

                340         350         360         370         380         390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      . : : : : : : : : : : : : : : : : : : : : : : : : : : .
Dad:AA  VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKL GATVEEGSDYCVITPPKK
                410         420         430         440         450         460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : . . .
Dad:AA  VKPAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH
                470          480          490          500          510          520

```

AAM63771.1 | *Arabidopsis thaliana* 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase, putative protein. | *Arabidopsis thaliana* | AA | 521
Length = 521
initn: 1668 initl: 1368 opt: 1696 Z-score: 3424.1 bits: 643.0 E(): 8.1e-182
Smith-Waterman score: 1696; 74.492% identity (90.293% similar) in 443 aa
overlap (4-445:79-521)

```

                                10          20          30
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  SWGLKKSDLMLNGSEIRPVKVRASVSTA EKASEIVLQPIREISGLIKLPGSKSLSNRILL
                50          60          70          80          90          100

```

```

                40          50          60          70          80          90
2MEPSP  LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKE
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  LAALSEGTTVVDNLLNSDDIN YMLDALKILGLNVETHSENNRAVVEGCGVFPASIDSKS
                110         120         130         140         150         160

```

```

                100         110         120         130         140         150
2MEPSP  EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  DIELYLG NAGTAMRPLTAAVTAAGGNASYVLDGV PQMRERPIGDLVVGLKQLGADV ECTL
                170         180         190         200         210         220

```

```

                160         170         180         190         200         210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  GTNCPVVRV NANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIVDKLISVPYVE
                230         240         250         260         270         280

```

```

                220         230         240         250         260         270
2MEPSP  MTLRLMERFGVKA EHSWSDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  MTLKLMERFGVSAEHS ESWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
                290         300         310         320         330         340

```

```

                280         290         300         310         320         330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPD
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  VEGCGTTS LQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAIDVNMNKMPD
                350         360         370         380         390         400

```

```

                340         350         360         370         380         390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKL GATVEEGSDYCVITPPKK
                410         420         430         440         450         460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . : : : : : : : : : : : : : : : : : : : : : : : : . . .
Dad:AA  VKPAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH
                470          480          490          500          510          520

```

AAA34136.1|*Solanum lycopersicum* protein (Tomato 5-enolpyruvylshikimate-3-phosphate synthase mRNA, complete cds.).|*Solanum lycopersicum*|AA|520
Length = 520
initn: 1694 initl: 1372 opt: 1692 Z-score: 3416.0 bits: 641.5 E(): 2.3e-181
Smith-Waterman score: 1692; 75.737% identity (88.435% similar) in 441 aa
overlap (6-445:80-520)

```

                                10          20          30
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
      . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  GVFKKDSVLRVVRKSSFRISASVATAEKPHEIVLXPIKDISGTVKLPGSKSLSNRILLLA
                50          60          70          80          90          100

```

```

                40          50          60          70          80          90
2MEPSP  ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
      . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  ALSEGRTVVDNLLSSDDIHYMLGALKTLGLHVEDDNNQRAIVEGCGGQFPVGKKSEEEI
                110          120          130          140          150          160

```

```

                100          110          120          130          140          150
2MEPSP  QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  QLFLGNAGTAMRPLTAAVTVAGGHSRYVLDGVPRMRERPIGDLVDGLKQLGAEVDCSLGT
                170          180          190          200          210          220

```

```

                160          170          180          190          200          210
2MEPSP  DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  NCPPVRIVSKGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVPYVEMT
                230          240          250          260          270          280

```

```

                220          230          240          250          260          270
2MEPSP  LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
      . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  LKLMERFGVFVEHSSGWDRLFVKGGQKYKSPGKAFVEGDASSASYFLAGAAVTGGTVTVE
                290          300          310          320          330          340

```

```

                280          290          300          310          320          330
2MEPSP  xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVA
      . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  GCGTSSLQGDVKFAEVLEKMGAEVTVTENSVTVKGPPRNSSGMKHLRAIDVNMNKMPPDVA
                350          360          370          380          390          400

```

```

                340          350          360          370          380          390
2MEPSP  MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  MTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEGSDYCIITPPEKLN
                410          420          430          440          450          460

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          400          410          420          430          440
2MEPSP VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
      :: ::::::::::::::::::::::::::::::::::::::::::::::::::::::. . ::
Dad:AA VTEIDTYDDHRMAMAFSLAACADVPVTIKNPGCTRKTFFPDYFEVLQKYSKH
          470          480          490          500          510          520

```

AA40476.1 | *Conyza sumatrensis* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Conyza sumatrensis* | AA | 444

Length = 444

initn: 1671 initl: 828 opt: 1668 Z-score: 3368.5 bits: 632.5 E(): 1e-178

Smith-Waterman score: 1668; 74.208% identity (89.819% similar) in 442 aa overlap (5-445:5-444)

```

          10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      ::::.:::.::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
Dad:AA SKAPEEIVLKPIQKISGTVHLPKSGSKSLSNRILLLAALSEGTTVVDNLLNSDDVHYMLGAL
          10          20          30          40          50          60

```

```

          70          80          90          100          110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      ::::.::. . : ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: .
Dad:AA RAVGLNVEENAAIKRAIVEGCGGVFPVGKEAKDEIQLFLGNSGTAMRTLTAAITAAGGNS
          70          80          90          100          110          120

```

```

          120          130          140          150          160          170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      ::::::::::::::::::::::::::::::::::::::: ::::::::::: : :::::::::::
Dad:AA SYVLDGVPRMRERPIGDLVTGLKQLGVNVDCSLGTNCPVVRVVGSGGLPGGRVRLSGSIS
          130          140          150          160          170          180

```

```

          180          190          200          210          220          230
2MEPSP SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      : : .. .. . ::::::::::::::::::::::::::::::::::: :::::::::::
Dad:AA SVYLTSLMLAAPLALGDVEIEIIDKLISVPYVQMTLKLKMKQFGVSVEHSDTLDRFHVRRG
          190          200          210          220          230          240

```

```

          240          250          260          270          280          290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      : :::: ::::::::::: ::::::::::: . . . . . ::::::::::: :::::
Dad:AA Q--KSPGNAYVEGDASNASYFLAGAAITGGTVTVEGCGTSSLLGDVKFAEVLGQMGA EVT
          250          260          270          280          290

```

```

          300          310          320          330          340          350
2MEPSP WTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      :.::: ::::: : : : . :::::::::::::::::::::::::::::::::::
Dad:AA WAENSVTVKGPPRNSSGRGHLRPVDVNMNKMPPDVAMALAVVALYADGPTAIRDVASWRVK
          300          310          320          330          340          350

```

```

          360          370          380          390          400          410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
      ::::.::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Dad:AA ETERMIAICTELRKLGATVEEGPDYCVITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
          360          370          380          390          400          410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYLSTFVKN
      ::::::::::::::::::::
Dad:AA VTIKDPGCTRKTFPDYFEVLERYTKH
      420      430      440

```

ABG88197.1|*Phaseolus vulgaris* EPSP synthase protein.|*Phaseolus vulgaris*|AA|522
Length = 522
initn: 1588 initl: 1296 opt: 1654 Z-score: 3339.1 bits: 627.2 E(): 4.4e-177
Smith-Waterman score: 1654; 74.150% identity (88.889% similar) in 441 aa
overlap (6-445:83-522)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLA
                                ::::::::::::::::::::
Dad:AB KDSMILSGAELSPFKVMSAVATAEKPSTSPSPEIEVEPIKDFSGWIKLPGTKSLSNRILLLA
      60      70      80      90     100     110

      40      50      60      70      80      90
2MEPSP ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEV
      ::::::::::::::::::::: ::::: . ::::: ::::: ::::: :::::
Dad:AB ALSEGTTVVDNLLNSDDIHYMLGALKWLGLRVETDKEINGAVVEGCGGIFPASIDSKSDI
      120     130     140     150     160     170

      100     110     120     130     140     150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPVRMRERPIGDLVVGLKQLGADVDCFLGT
      ::::: ::::: ::::: . ::::::::::::::::::::
Dad:AB ELYLGNAGTAMRPLTAAVTAAGGNAWYVLGDGVPVRMRERPIGDLVVGLKQLGADVDCFLGT
      180     190     200     210     220     230

      160     170     180     190     200     210
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      ::::: ::::: ::::: ::::: .. .. . ::::::::::::::::::::
Dad:AB NCPVVRVNAGGGLPGGKVKLFGSISSQYLTALLMSAPLALGDVEIEIIDKLISVPYVEVT
      240     250     260     270     280     290

      220     230     240     250     260     270
2MEPSP LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxx
      ::::: ::::: ::::: . ::::::::::::::::::::: . .
Dad:AB LKLMEFVGVSVEHWDSWDRFLVHGGQKYKSPGNAYVEGDASSASYLLAGAAITGGTVTVE
      300     310     320     330     340     350

      280     290     300     310     320     330
2MEPSP xxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA
      . ::::::::::::::: ::::: ::::::::::::::: ::::: :::::::::::::::
Dad:AB GCGTKSLQGDVKFAEVLEKMGCKVWWTENSVTVTGPPRDLFGRRVLRAIDVNMNKMMPDVA
      360     370     380     390     400     410

      340     350     360     370     380     390
2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      ::::::::::::::: ::::: ::::::::::::::: ::::: :::::::::::::::
Dad:AB MTLAVVALFADGPTTIRDVASWRVKETEEMIAICTELRKLKLGATVEEGPDYCVITPPKKLK
      420     430     440     450     460     470

```

400 410 420 430 440
 2MEPSP VTAIDTYDDHRMAMAFSLAACAEVPTIRDPGCTRKTFPDYFDVLSTFVKN
 :
 Dad:AB VAEIDTYDDHRIPMAFSLAACADVPVTINDP-CTRKTFFPDYFEVLERLTKH
 480 490 500 510 520

Smith-Waterman score: 1801; 81.573% identity (90.562% similar) in 445 aa overlap (2-445:71-506)

```

                                10      20      30
2MEPSP      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRI
              : ::::::::::: ::::::::::: :::::::::::
Dad:EA  GGMRVVRVRARGRREAVVVASASSSSVAAPAAKAAEEIVLQPIREISGAVQLPGSKSLSNRI
              50      60      70      80      90     100

```

40 50 60 70 80 90
 2MEPSP LLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DA
 :::::::::::::::::::::::::::::: :::::::::::::::::::::::::::::::::: ::
 Dad:EA LLLSALSEGTTVVDNLLNSEDVHYMLEALKALGLSVEADKVAKRAVVVGCGGKFPVEKDA
 110 120 130 140 150 160

	100	110	120	130	140	150
2MEPSP	KEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVL	DGVPRMRERPIGDLVVGLKQLGADVDC				
	: : : : : : : : : : : :	:
Dad:EA	KEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVL	DGVPRMREKPIGDLVFGLKQLGADVDC				
	170	180	190	200	210	220

```

                160      170      180      190      200      210
2MEPSP FLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxxGDVEIEIIDKLISIPY
      :::::  :  .  .  .  .  ::  .  .  .  .  ::::::::::::::
Dad:EA FLGTECP-----LFVSRELEDFLVAQYLSALLMAAPLALGDVEIEIIDKLISIPY
                230      240      250      260      270

```

2MEPSP VEMTLRLMERFGVKA²²⁰EHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxx
Dad:EA VEMTLRLMERFGVKA²⁸⁰EHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGAAITGTT³³⁰

2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGKHLKAIDVNMNMK
 Dad:EA VTVQCGCTTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPREPYGKHLKAIDVNMNMK

340 350 360 370 380 390
 2MEPSP PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP
 :
 Dad:EA PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP
 400 410 420 430 440 450

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                400          410          420          430          440
2MEPSP  EKLNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
        .....
Dad:EA  EKLNITAIDTYDDHRMAMAFSLAACADVPVTIRDPGCTRKTFFPNYFDVLSTFVRN
                460          470          480          490          500

```

AAK20397.1 | *Lolium rigidum* 5-enolpyruvylshikimate 3-phosphate synthase protein. | *Lolium rigidum* | AA | 347
Length = 347
initn: 1473 initl: 1379 opt: 1474 Z-score: 2977.3 bits: 559.7 E(): 6.2e-157
Smith-Waterman score: 1474; 84.104% identity (91.908% similar) in 346 aa
overlap (65-409:2-347)

```

                40          50          60          70          80          90
2MEPSP  AALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEE
        .....
Dad:AA                                     PLSVEADKVAKRAVVVGCGGRFPXEKDAKEE
                                   10          20          30

```

```

                100          110          120          130          140          150
2MEPSP  VQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLG
        .....
Dad:AA  VKLFLGNAGTAMRPLTAAVVAAGGNATYVLDGVPRMRERPIGDLVVGLKQLGANVDCFLG
                40          50          60          70          80          90

```

```

                160          170          180          190          200          210
2MEPSP  TDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEM
        .....
Dad:AA  TDCPPVRINGIGGLPGGKVKLSGSISSQYLSSLLMAAPLALGDVEIEIIDKLISVPYVEM
                100          110          120          130          140          150

```

```

                220          230          240          250          260          270
2MEPSP  TLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
        .....
Dad:AA  TLRLMERFGVTAHSDSWDRFYIKGGQKXKSPGNAYVEGDASSASYFLAGAAITGGTVTV
                160          170          180          190          200          210

```

```

                280          290          300          310          320          330
2MEPSP  xxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDV
        .....
Dad:AA  QGCGTTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPRQPFGRKHLKAIDVNMNKMMPDV
                220          230          240          250          260          270

```

```

                340          350          360          370          380          390
2MEPSP  AMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKL
        .....
Dad:AA  AMTLAVVALFADGPTAIRDVASWRVKETERMVAICTELTKLGATVEEGPDYCIITPPEKL
                280          290          300          310          320          330

```

```

                400          410          420          430          440
2MEPSP  NVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
        .....
Dad:AA  NVT AIDTYDDHRMAMA
                340

```

EDQ63766.1|*Physcomitrella patens* subsp. *patens* predicted protein
protein.|*Physcomitrella patens* subsp|AA|473
Length = 473
initn: 1433 init1: 984 opt: 1432 Z-score: 2890.2 bits: 544.1 E(): 4.4e-152
Smith-Waterman score: 1432; 65.463% identity (83.296% similar) in 443 aa
overlap (2-439:25-467)

Page 389 of 526

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AAT45236.1 | *Erigeron annuus* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Erigeron annuus* | AA | 358
Length = 358
initn: 1412 initl: 1257 opt: 1414 Z-score: 2855.6 bits: 537.2 E(): 3.7e-150
Smith-Waterman score: 1414; 77.654% identity (89.665% similar) in 358 aa
overlap (47-403:1-358)

	20	30	40	50	60	70	
2MEPSP	GTKVLP	SGSKSL	SNRILL	LAALSE	GTTVVD	NLLNSE	DVHYMLGALRTLGLSVEADKAAKRA
						: : : :
Dad:AA				LLNSDD	VHYMLGALRTLGLNVEED	GAIKRA	
				10	20	30	
	80	90	100	110	120	130	
2MEPSP	VVVGCG	GKFPV	-EDAKE	EVQLFL	GNAGIAM	RLSxxxxxx	YVLDGVP
	.: : : :	.: : : :	.: : : :	.: : : :	.: : : :	.: : : :	.: : : :
Dad:AA	IVEGCG	GVFPVG	KEAKDD	IQFLGN	AGTAMR	PLTAAVTA	AGGNSSVLDGVP
	40	50	60	70	80	90	
	140	150	160	170	180	190	
2MEPSP	DLVVGL	KQLGAD	VDCFLG	TDCTPP	VRVNGIG	GLPGGK	VKLSGSISSQYxxxxxx
	...: : : :	...: : : :	...: : : :	...: : : :	...: : : :	...: : : :	...: : : :
Dad:AA	DLVMGL	KQLGAD	VDCSLG	TNCPVR	VVGGLPG	GKVKLSG	SISSQYLTALLMASPLALG
	100	110	120	130	140	150	
	200	210	220	230	240	250	
2MEPSP	DVEIEI	IDKLIS	IPYVEM	TLRLMER	FGVKA	AEHSDSW	DRFYIKGGQKYKSPKNAYVEGDAS
: : : :: : : :: : : :: : : :: : : :: : : :: : : :
Dad:AA	DVEIEI	IDKLIS	IPYVEM	TLKLME	RVGSVE	HSDSWDR	FFIRGGQKYKSPGNAYVEGDAS
	160	170	180	190	200	210	
	260	270	280	290	300	310	
2MEPSP	SASYFL	AGAAI	xxxxxxxxxxxx	SLQGDV	KFAEVLE	MMGAKVT	TWTETSVTVTGPPREPF
: : : :: : : :: : : :: : : :: : : :: : : :: : : :
Dad:AA	SASYFL	AGAAIT	GGTITV	EGCGT	SSLQGD	VKFAEVL	GQMGAETWTENSVTVKGPPRDSS
	220	230	240	250	260	270	
	320	330	340	350	360	370	
2MEPSP	GRKHLK	AIDVNM	NKMPDV	AMTLAV	VALFAD	GPTAIR	DVASWRVKETERMVAIRTELTKLG
: : : :: : : :: : : :: : : :: : : :: : : :: : : :
Dad:AA	GRKHLR	AVDVM	NKMPDV	AMTLAV	VALYAD	GPTAIR	DVASWRVKETERMIAICTELRKL
	280	290	300	310	320	330	
	380	390	400	410	420	430	
2MEPSP	ASVEEG	PDYCI	IITPPE	KLNVT	AITDYD	DHRMAM	AFSLAACAEVPVTIRDPGCTRKTFPDY
	.: : : :	.: : : :	.: : : :	.: : : :	.: : : :	.: : : :	.: : : :
Dad:AA	ATVEEG	PDYCV	ITPPE	KLNVT	AITDYD		
	340	350					

AAT45237.1|*Erigeron annuus* 5-enol-pyruvylshikimate-phosphate synthase
 protein.|*Erigeron annuus*|AA|360
 Length = 360
 initn: 1380 init1: 1223 opt: 1382 Z-score: 2790.8 bits: 525.3 E(): 1.5e-146
 Smith-Waterman score: 1382; 75.000% identity (88.889% similar) in 360 aa
 overlap (45-403:1-360)

```

          380          390          400          410          420          430
2MEPSP LGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFP
      ::::: ::::::::::::::::::::
Dad:AA LGATVEEGTDYCVITPPEKLNVT AIDTYDD
          340          350          360

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

AAT45234.1 | *Amaranthus tuberculatus* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Amaranthus tuberculatus* | AA | 357
Length = 357
initn: 1373 initl: 1249 opt: 1375 Z-score: 2776.7 bits: 522.6 E(): 9.3e-146
Smith-Waterman score: 1375; 75.630% identity (88.796% similar) in 357 aa
overlap (47-402:1-357)

	20	30	40	50	60	70	
2MEPSP	GTKVLPGSKSLSNRILL	LAALSEGTTVVDNLL	NSEDVHYMLGALRT	LGLSVEADKAAKRA			
				:: ::.	::: ::	::: ::	::: ::
Dad:AA				LLYSDDILYMLDALRT	LGLKVEDDNTDKRA		
				10	20	30	
	80	90	100	110	120	130	
2MEPSP	VVVGCGGKFPV-EDAK	EEVQLFLGNAGIAMR	SLxxxxxxxxxxxxx	YVLDGVPRMRERPIG			
	:: :::	::: ::.	::: ::	::: ::	::: ::	::: ::	::: ::
Dad:AA	VVEGCGGLFPVGKDG	KEEIQLFLGNAGTAMR	PLTAAVAVAGGNSS	YVLDGVPRMRERPIG			
	40	50	60	70	80	90	
	140	150	160	170	180	190	
2MEPSP	DLVVGLKQLGADVDC	FLGTDCPPVRVNGIG	GLPGGKVKLSGSISS	QYxxxxxxxxxxxxx	G		
	::: :::	::: ::.	::: ::	::: ::	::: ::	::: ::	::: ::
Dad:AA	DLVAGLKQLGSDVDC	FLGTNCPVRVNAKG	GLPGGKVKLSGVS	SSQYLTALLMATPLGLG			
	100	110	120	130	140	150	
	200	210	220	230	240	250	
2MEPSP	DVEIEIIDKLISIPY	VEMTLRLMERFGVKA	EHSDSWDRFYIKGG	QKYKSPKNAYVEGDAS			
	::: :::	::: ::.	::: ::	::: ::	::: ::	::: ::	::: ::
Dad:AA	DVEIEIVDKLISVPY	VEMTIRLMERFGVS	VEHSDSWDRFFIR	GGQKYKSPGKAYVEGDAS			
	160	170	180	190	200	210	
	260	270	280	290	300	310	
2MEPSP	SASYFLAGAAIxxxx	xxxxxxxxxxxxxSL	QGDVKFAEVLEMMG	AKVTWTETSVTVTG	PPREPF		
	::: :::	::: ::.	::: ::	::: ::	::: ::	::: ::	::: ::
Dad:AA	SASYFLAGAAVTGGT	VTVKGCGTSSLQGD	VKFAEVLEKMGCK	VTWTDNSVTVTGPP	RESS		
	220	230	240	250	260	270	
	320	330	340	350	360	370	
2MEPSP	GRKHLKAIDVNMNKM	PDVAMTLAVVALFAD	GPTAIRDVASWRVK	ETERMVAIRTELTKLG			
	::: :::	::: ::.	::: ::	::: ::	::: ::	::: ::	::: ::
Dad:AA	GRKHLRAIDVNMNKM	PDVAMTLAVVALYAD	GPTAIRDVASWRVK	ETERMIAICTELRKL	G		
	280	290	300	310	320	330	
	380	390	400	410	420	430	
2MEPSP	ASVEEGPDYCIITP	PEKLNVT	AITDYDDHRMAMAF	SLAACAEVPVTIRDP	GCTRKTFPDY		
	::: :::	::: ::.	::: ::	::: ::	::: ::	::: ::	::: ::
Dad:AA	ATVEEGSDYCVITP	PEKLNPTAIETYD					
	340	350					

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                160      170      180      190      200      210
2MEPSP  FLGTDCCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPY
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  FLGTKCPPVRIVSKGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVLY
        50      60      70      80      90      100

                220      230      240      250      260      270
2MEPSP  VEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxx
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  VEMTLKLMERFGISVEHSSSWDRFVVRGGQKYKSPGKAYVEGDASSASYFLAGAAVTGGT
        110     120     130     140     150     160

                280      290      300      310      320      330
2MEPSP  xxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMK
        . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  VTVEGCGTSSSLQGDVKFAEVLEQMGAEVTWTENSVTVKGPPRNSSAMKHLRAIDVNMNMK
        170     180     190     200     210     220

                340      350      360      370      380      390
2MEPSP  PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  PDVAMTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLGAATVEEGPDYCIITPP
        230     240     250     260     270     280

                400      410      420      430      440
2MEPSP  EKLNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  EKLNVTEIDTYDDHRMAMAFSLAACADVPVTINDPGCTRKTFFPNYFDVLQQYSKH
        290     300     310     320     330

```

ABP00889.1 | *Ostreococcus lucimarinus* CCE9901 predicted protein.
protein. | *Ostreococcus lucimarinus* CCE9901 | AA | 436
Length = 436
initn: 1122 initl: 396 opt: 1263 Z-score: 2548.6 bits: 480.7 E(): 4.7e-133
Smith-Waterman score: 1263; 59.732% identity (80.089% similar) in 447 aa
overlap (5-444:2-436)

```

                10      20      30      40      50      60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSGKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB  MEQLTLKPMKKIEGTVRLPGSKSLSNRILLLAALAEGLTKVENLLSDDIRYMVDAL
        10      20      30      40      50

                70      80      90      100     110     120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB  KVLGLSFTEDRENNILEITGCGGKLPVEGA---ELFLGNAGTAMRPLTAAVAAAGKGT-
        60      70      80      90      100     110

                130     140     150     160     170     180
2MEPSP  YVLDGVPRMRERPIQDLVGLKQLGADVDCFLGTDCCPPVRVNGIGGLPGGKVKLSGSISS
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB  FILDGVERMRERPIQDLVDGLVQLGVKAECTMGTCPPVKVEA-NGLPGGRVELSGSVSS
        120     130     140     150     160     170

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      :: .. .. :...: :...: :...: :...: :...: :...: :...: :...: :...:
Dad:AB QYLTALLMAAPLCEGSIEIVIVDELISKPYVEMTITLRLMERFGVKVEKADDLQSFQKIQGGQ
                180          190          200          210          220          230

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      :: :: .:.....:.....:.....:.....:.....:.....:.....:.....:.....:
Dad:AB KYISPGSAFVEGDASSASYFLAGATITGGTVTVIGCGSESIQGDTNFAYTMEQMGATLEW
                240          250          260          270          280          290

                310          320          330          340          350          360
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      .::: :: . :...: :...: :...: :...: :...: :...: :...: :...: :...:
Dad:AB GPNSVTCTGP-----KGPLKAIDVNMNAMPDAAMTLAVAALFADGITTIRDVASWRVKE
                300          310          320          330          340

                370          380          390          400          410
2MEPSP TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT-----IDTYDDHRMAMAFSLA
      ::::: :: :...: :...: :...: :...: :...: :...: :...: :...: :...:
Dad:AB TERMIAICTELRKLGCDFEGADYCVITPPHKLDPKMKANVDIDTYDDHRMAMAFALA
                350          360          370          380          390          400

                420          430          440
2MEPSP ACAEVPVTIRDPGCTRKTFPDYFDVLSTFVK
      ::::: :...: :...: :...: :...: :...: :...: :...: :...: :...:
Dad:AB ACGDVDVVINDPKCTKKTFPTYFDVLKSVAK
                410          420          430

```

ECZ07785.1 |marine metagenome hypothetical protein protein. |marine metagenome |AA|482
Length = 482
initn: 1134 initl: 397 opt: 1256 Z-score: 2533.7 bits: 478.1 E(): 3.2e-132
Smith-Waterman score: 1256; 59.910% identity (79.955% similar) in 444 aa
overlap (5-441:48-479)

```

                                10          20          30
2MEPSP MAGAEIIVLQPIKEISGTVKLPKSKSLSNRILL
      :...: :...: :...: :...: :...: :...: :...: :...: :...: :...:
Dad:EC SFAPAKCGSTRRRPSRSSRTLTRANATGAMEQLTKPIKKVEGTVRLPGSKSLSNRILL
                20          30          40          50          60          70

                40          50          60          70          80          90
2MEPSP AALSEGTTVVDNLLNSQDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEV
      ::::: :...: :...: :...: :...: :...: :...: :...: :...: :...:
Dad:EC AALAEGTTKVENLLSDDIRYMVDALKVLGLDFTEDRENNILEITGCGGEFPVEGA----
                80          90          100          110          120          130

                100          110          120          130          140          150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      :...: :...: :...: :...: :...: :...: :...: :...: :...: :...:
Dad:EC DLFLGNAGTAMRPLTAAVAAAGKGT-FVLDGVERMRERPIQDLVDGLVQLGVKAECTMGT
                140          150          160          170          180          190

```

	160	170	180	190	200	210
2MEPSP	DCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT					
	: : : :	: : : : : . : : : : : . : : : : .	: : : : : .	: : : : : .	: : : : : .	: : : : : .
Dad:EC	GCPPVKVEA-NGLPGGRVELSGSVSSQYLTALLMAAPLCQGSIEIVIVDELISKPYVEMT					
	200	210	220	230	240	250
	220	230	240	250	260	270
2MEPSP	LRLMERFGVKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxx					
	. : : : : : : .	. : . : : : : : : : : : : : : : : : : .	. : : : : : : : : : : : : : : : : .	. : : : : : : : : : : : : : : : : .	. : : : : : : : : : : : : : : : : .	. : .
Dad:EC	ITLMERFGVKVEKADDLQSFKIQQGQKYVSPGSAFVEGDASSASYFLAGATITGGTVTVI					
	260	270	280	290	300	310
	280	290	300	310	320	330
2MEPSP	xxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA					
	. : : : : : : .	. : : : : : : .	. : : : : : : .	. : : : : : : .	. : : : : : : .	. : : : : : : .
Dad:EC	GCGSESIQGD TNFAYTMEQMGATLEWGPNSVTCTGP-----QGPLKAIDVNMNAMPDAA					
	320	330	340		350	360
	340	350	360	370	380	390
2MEPSP	MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN					
	: : : : . : : : : .	: : : : : : : : : : : : : : : : .	: : : : : : : : : : : : : : : : .	: : : : : : : : : : : : : : : : .	: : : : : : : : : : : : : : : : .	: : : : : : : : : : : : : : : : .
Dad:EC	MTLAVAALFADGVTTIRDVASWRVKETERMIAICTELRKLGCDFVFGSDYCVITPPHKLE					
	370	380	390	400	410	420
	400	410	420	430	440	
2MEPSP	VTA-----IDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN					
	:	: : : : : : : : : : : : : : : : .	: : : : : : : : : : : : : : : : .	: : : : : : : : : : : : : : : : .	: : : : : : : : : : : : : : : : .	
Dad:EC	PPATMKANVDIDTYDDHRMAMAFALAACGDVDVIINDPKCTKKTFTPTYWDLSSVSA					
	430	440	450	460	470	480

[illegible]

150 160 170 180 190 200
 2MEPSP GADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxxGDVEIEIIDK
 : : : : : : :
 Dad:AA GADVDCFLGTETCPPVRVKGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDK
 100 110 120 130 140 150

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      210      220      230      240      250      260
2MEPSP LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGA
      .....
Dad:AA LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGA
      160      170      180      190      200      210

```

```

      270      280      290      300      310      320
2MEPSP AIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAID
      :. . . . .
Dad:AA AITGGTVTVQGCSTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPREPYGKKHLKAID
      220      230      240      250      260      270

```

```

      330      340      350      360      370      380
2MEPSP VNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDY
      .....
Dad:AA VNMNKMPPDVAMTLAVVALFADGPTAIRDGFLESKGNRKDGCNSDRANKGKFIRSRVLSFF
      280      290      300      310      320      330

```

ABM68632.1 | *Dunaliella salina* plastid EPSP synthase protein. | *Dunaliella salina* | AA | 514
Length = 514
initn: 1124 initl: 460 opt: 1219 Z-score: 2458.4 bits: 464.3 E(): 5e-128
Smith-Waterman score: 1219; 58.239% identity (80.361% similar) in 443 aa
overlap (5-445:81-514)

```

                        10      20      30
2MEPSP                MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                        .....
Dad:AB RRSCSKSSIRSTRLQTTVCSATLAHSAAPDQLVLQPIKQISGTVRLPGSKSISNRVLLL
      60      70      80      90      100      110

```

```

      40      50      60      70      80      90
2MEPSP AALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEV
      :. . . . .
Dad:AB AALAEGTTVVKNLDSDDIRYMGALKGLGIELEERWDKGEMVVKCGGQFSAEGG----
      120      130      140      150      160

```

```

      100      110      120      130      140      150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPMRERPIGDLVVGLKQLGADVDCFLGT
      .....
Dad:AB ELFLGNAGTAMRPLTAAVAAAGRGK-FVLDGTARMRERPIQDLVDGLVQLGVDAKCPLGT
      170      180      190      200      210      220

```

```

      160      170      180      190      200      210
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGD--VEIEIIDKLISIPYVE
      :. . . . .
Dad:AB GCPPVEVNA-QGLPSGKVQLKGSVSSQYLTAALLMAAPLSKGTEGIEIVITDELVSQPYVD
      230      240      250      260      270      280

```

```

      220      230      240      250      260      270
2MEPSP MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxx
      :. . . . .
Dad:AB MTVQIMERFGVTVERLNLQHMRIPPNQTYKTSGEAFVEGDASSASYFLAGATITGGTVV
      290      300      310      320      330      340

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                280          290          300          310          320          330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMKMPD
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Dad:AB  VEGCGSASVQGDVRFDAEVMGLMGAKVEWSLYSIKITGP--SAFG-KPLQGIDHDCNDIPD
                350          360          370          380          390          400

```

```

                340          350          360          370          380          390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Dad:AB  AAMTLAVAALFADKPTTIRNVYNWRVKETERMVAIVNETRKLGLATVEEGRDYCVITPPKQ
                410          420          430          440          450          460

```

```

                400          410          420          430          440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .. . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Dad:AB  IQSAAIDTYDDHRMAMAFSLAACGPVPVTINDPGCTRKTFPDYFRVLESVTQH
                470          480          490          500          510

```

EDO96795.1 | *Chlamydomonas reinhardtii* 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Chlamydomonas reinhardtii* | AA | 512
Length = 512
initn: 907 initl: 484 opt: 1197 Z-score: 2413.9 bits: 456.0 E(): 1.5e-125
Smith-Waterman score: 1197; 58.720% identity (78.366% similar) in 453 aa overlap (5-445:70-512)

```

                                10          20          30
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Dad:ED  SVAPAPACSAPAGAGRRVVRASATKEKVEELTIQPVKKIAGTVKLPGSKSLSNRILL
                40          50          60          70          80          90

```

```

                40          50          60          70          80          90
2MEPSP  AALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEV
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Dad:ED  AALSEGTTLVKNLLSDDIRYMGALKALNVKLEENWEAGEMVVHGCGRFDSAGA----
                100          110          120          130          140          150

```

```

                100          110          120          130          140          150
2MEPSP  QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPVRMRERPIGDLVVGLKQLGADVDCFLGT
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Dad:ED  ELFLGNAGTAMRPLTAAVVAAGR GK-FVLDGVARMRERPIEDLVDGLVQLGVDAKCTMG
                160          170          180          190          200          210

```

```

                160          170          180          190          200
2MEPSP  DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxx-xxxxGD-VEIEIIDKLISI
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Dad:ED  GCPPVEVNS-KGLPTGKVYLSGKVSSQYL TALLMAAPLAVPGGAGGDAIEIIKDELVSQ
                220          230          240          250          260          270

```

```

                210          220          230          240          250          260
2MEPSP  PYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxx
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
Dad:ED  PYVDMTVKLMERFGVVVERLNLGLHLRIPAGQTYKTPGEAYVEGDASSASYFLAGATITG
                280          290          300          310          320          330

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMN
      . . . : : : : : . . : : : . : : : . : : :
Dad:ED GTVTVEGCGSDSLQGDVRF AEVMGLLGAKVEWSPYSITITGP--SAFG-KPITGIDHDCN
      340      350      360      370      380      390

```

```

      330      340      350      360      370      380
2MEPSP KMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT
      . : : : : : : : : : : : : : : : : : : : : : : :
Dad:ED DIPDAAMTLAVAALFADRPTAIRNVYNWRVKETERMVAIVTELRLKGAEVEEGRDYCIIVT
      400      410      420      430      440      450

```

```

      390      400      410      420      430      440
2MEPSP PPE-----KLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTF
      : : : : : : : : : : : : : : : : : : : : : : :
Dad:ED PPPGGVKGVKANV-GIDTYDDHRMAMAFSLVAAAGVPVIRDPGCTRKTFPTYFKVFESV
      460      470      480      490      500

```

```

2MEPSP VKN
      ...
Dad:ED AQH
      510

```

ABR25383.1|*Oryza sativa* Indica Group 5-enolpyruvylshikimate-3-phosphate synthase protein.|*Oryza sativa Indica Group*|AA|273
Length = 273
initn: 1152 initl: 1152 opt: 1152 Z-score: 2327.0 bits: 439.0 E(): 1e-120
Smith-Waterman score: 1152; 83.150% identity (91.941% similar) in 273 aa
overlap (96-368:1-273)

```

      70      80      90      100      110      120
2MEPSP SVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDG
      : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB LFLGNAGTAMRPLTAAVTAAGGNATYVLDG
      10      20      30

```

```

      130      140      150      160      170      180
2MEPSP VPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxx
      : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB VPRMRERPIGDLVVGLKQLGADVDCFLGTECPPVRVKIGIGGLPGGKVKLSGSISSQYLSA
      40      50      60      70      80      90

```

```

      190      200      210      220      230      240
2MEPSP xxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSP
      . . : : : : : : : : : : : : : : : : : : : : :
Dad:AB LLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSP
      100      110      120      130      140      150

```

```

      250      260      270      280      290      300
2MEPSP KNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSV
      : : : : : : : : : . . . : : : : : : : : : :
Dad:AB GNAYVEGDASSASYFLAGAAITGGTVTVQGC GTTSLQGDVKFAEVLEMMGAKVTWTDTSV
      160      170      180      190      200      210

```

```

      310      320      330      340      350      360
2MEPSP TVTGPPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMV
      ::::::::::.:.:::.:::.:::.:::.:::.:::.:::.:::.:::.:::.:::.:::.:::.:::.::
Dad:AB TVTGPPPREPYGKKHLKAVDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMV
      220      230      240      250      260      270

      370      380      390      400      410      420
2MEPSP AIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDP
      :::
Dad:AB AIR

```

2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLALLAALSEGTTVVDNLLNSEDVHYMLGAL
 Dad:ED MEEITLQPIKTIISGTIKLPGSKSLSNRTLLALLAALSEGTTVVENLLDSEDVRYMVAAL

```

              70          80          90          100         110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPV---EDAKEEVQLFLGNAGIAMRSLxxxxxxxxx
      .::.: :.: :. : : :. : . : :.: :.: :.: :.: :.: :.: :.: :.:
Dad:ED QTLGFKIEEDRAANRLVIEGGGVFPIARDAEKSGETVNLFLGNAGTAMRPLTA AVVAAG
        60          70          80          90          100         110

```

120 130 140 150 160
 2MEPSP xxxxyVLDGVPRMRERPIGDLVVLGLKQLGADVDCFLG-TDCPPVRVN-----GIGGLP
 . :
 Dad:ED GNVSYVLDGVPRMRERPIIDLVLGLLQQLGADVKTTEEYPNCPVVIINDKSSGSKSKGGLP
 120 130 140 150 160 170

```

      170      180      190      200      210      220
2MEPSP  GGKVKLSGSISSQYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:ED  GGTVHLSGKVSSQYLSALLMAAPLALGDVEIVMIDKLVSVPYVDMTLRLMERFGVKVERH
      180      190      200      210      220      230

```

290 300 310 320 330 340
2MEPSP EVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKM P D V A M T L A V V A L F A D G P T
 :
Dad:ED AVLEKMGAKVEWTEHSVTVTGAPVDPR TGRLKS IDVD MNAMPDVAMTLAVLGLFSDGPV
 300 310 320 330 340 350

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      350      360      370      380      390      400
2MEPSP AIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAM
      ::::: . ::::: :: :: ::::: ::::: ::::: ::::: :::::
Dad:ED AIRDVENWRVKETERMRAIVDELTKLGAEEVEEGQDYCIVTPPKRITPAKIETYDDHRMAM
      360      370      380      390      400      410

      410      420      430      440
2MEPSP AFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
      ::::: . ::::: ::::: ::::: ::::: :::::
Dad:ED AFSLAACGDVPITILDPGCTRKTFFPTYFTELEKLCQH
      420      430      440      450

```

AAT45238.1 | *Helianthus salicifolius* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Helianthus salicifolius* | AA | 264
Length = 264
initn: 951 initl: 866 opt: 956 Z-score: 1930.3 bits: 365.6 E(): 1.3e-98
Smith-Waterman score: 956; 71.970% identity (86.364% similar) in 264 aa overlap (61-323:1-264)

```

      40      50      60      70      80
2MEPSP ILLLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-ED
      ::::: :: : : ::::: ::::: ::::: ..
Dad:AA RTLGLRVEEDGAIKRAIVEGCGGVFPVGKE
      10      20      30

      90      100      110      120      130      140
2MEPSP AKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPRMRERPIGDLVVGLKQLGADVD
      ::::: ::::: . ::::: . . . . . ::::: ::::: ::::: :::::
Dad:AA AKDEIQLFLGNAGTSMRPLTAAVTAAGNSSYILDGVPRMRERPIGDLVTGLKQLGADVD
      40      50      60      70      80      90

      150      160      170      180      190      200
2MEPSP CFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIP
      ::::: ::::: : ::::: ::::: . . . . . ::::: ::::: :::::
Dad:AA CFLGTNCPVVRVVGGLPGGKVKLSGSISSQYLTALLMASPLALGDVEIEIIDKLISIP
      100      110      120      130      140      150

      210      220      230      240      250      260
2MEPSP YVENTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxx
      ::::: . ::::: . . . . . ::::: ::::: ::::: :::::
Dad:AA YVDMTIKLMERFGVSEHSDSWDRFFIKGGQKYKSPGNAYVEGDASSASYFLAGAAITGG
      160      170      180      190      200      210

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNK
      . . . . . ::::: . . . . . ::::: . . . . .
Dad:AA TITVEGCGTSSLQGDVKFAEVLGQMGAEVTWTENSVTVRGPPRNASGRKHLHAV
      220      230      240      250      260

      330      340      350      360      370      380
2MEPSP MPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITP

```

AAT45242.1 | *Plantago lanceolata* 5-enol-pyruvylshikimate-phosphate
 synthase protein. | *Plantago lanceolata* | AA | 265
 Length = 265
 initn: 861 initl: 861 opt: 951 Z-score: 1920.2 bits: 363.7 E(): 4.8e-98
 Smith-Waterman score: 951; 71.321% identity (86.038% similar) in 265 aa overlap
 (60-323:1-265)

330 340 350 360 370 380
2MEPSP KMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT

AAT45240.1 | *Sarracenia purpurea* 5-enol-pyruvylshikimate-phosphate
synthase protein. | *Sarracenia purpurea* | AA | 265
Length = 265
initn: 949 initl: 857 opt: 951 Z-score: 1920.2 bits: 363.7 E(): 4.8e-98
Smith-Waterman score: 951; 70.943% identity (86.792% similar) in 265 aa overlap
(60-323:1-265)

30 40 50 60 70 80
2MEPSP RILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-E
Dad:AA LRTLGLSVEEDSVIKRAIVEGCGGVFPVGK
10 20 30

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      90      100      110      120      130      140
2MEPSP DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADV
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA ESRDEIPLFLGNAGTAMRPLTAAVTAAGGNSSYILDGVPRMRERPIGDLVTGLKQLGADV
      40      50      60      70      80      90

```

```

      150      160      170      180      190      200
2MEPSP DCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISI
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA DCFLGTNCPVVRVIGKGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISI
      100      110      120      130      140      150

```

```

      210      220      230      240      250      260
2MEPSP PYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxx
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA PYVEMTLKLMKRFVSVSEHTDSWDKFLIRGGQKYKSPGSAYVEGDASSASYFLAGAAVTG
      160      170      180      190      200      210

```

```

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMN
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA GTITVEGCGTSSLQGDVKFAEVLEKMGAKVTWTENSVTVRGPPRSSSGRKHLRAI
      220      230      240      250      260

```

```

      330      340      350      360      370      380
2MEPSP KMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT

```

CAD14609.1 | *Ralstonia solanacearum* probable 3-phosphoshikimate 1-carboxyvinyltransferase (epsp synthase) protein protein. | *Ralstonia solanacearum* | AA | 436
Length = 436
initn: 901 initl: 231 opt: 939 Z-score: 1892.6 bits: 359.3 E(): 1.6e-96
Smith-Waterman score: 939; 48.284% identity (72.540% similar) in 437 aa overlap
(5-436:2-428)

```

      10      20      30      40      50      60
2MEPSP MAGAEIIVLQPIKEISGTVKLPKSGKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA MEHLDVGPLKAARGTVKLPKSGKSISNRVLLLAALAEGETTVVRDLLDSDTRVMLAAL
      10      20      30      40      50

```

```

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA DTLGVRCEPLGTANAYRVGTGTGGRFPAKSA---DLFMGNAGTAIRPLTAALALQGGE--
      60      70      80      90      100      110

```

```

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA YTLHGVPRMHERPIGDLVDGLRQVGARIDYTGNEGFPLAIRAASIRIDAPIRVRGDVSS
      120      130      140      150      160      170

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                190          200          210          220          230
2MEPSP QYxxxxxxxxxxxxGD---VEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
      : . . . . . : . : : . . : : : : : : : : : : : : : : : : : :
Dad:CA QFLTALLMALPLVEGSGRPVTIEVVGELISKPYIEITLNLMARFGVQVERN-GWASFSVP
                180          190          200          210          220          230

                240          250          260          270          280          290
2MEPSP GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      : . . : . . : : : : : : : : : . . : : : : : : : : : : : :
Dad:CA TGVAYRAPGEIFVEGDASSASYFLAAGALGGGPVRVEGVGMSSIQGDVRFADALNRMGAN
                240          250          260          270          280          290

                300          310          320          330          340          350
2MEPSP VTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWR
      : . . . : : : . : : : . . . : : : : : : : : : : : . . . : :
Dad:CA VMAGDNWIEVRGVERDD-GKLH--ALELDCNHIPDAAMTLAVAALFADGTTTLTNIGSWR
                300          310          320          330          340

                360          370          380          390          400          410
2MEPSP VKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT--IDTYDDHRMAMAFSLAAC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA VKETDRLTAMATELRKLGAAVEEGTDYIRVTPPSHWTAPAGGIDTYDDHRMAMAFSLAAF
                350          360          370          380          390          400

                420          430          440
2MEPSP AEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : : : : : : : : : : : : : : : : :
Dad:CA GPVPVRINDPRCVAKTFPEYFTAFGGIAA
                410          420          430

```

CAL56989.1 | *Ostreococcus tauri* 5-enolpyruvylshikimate-3-phosphate
 (ISS) protein. | *Ostreococcus tauri* | AA | 316
 Length = 316
 initn: 958 initl: 386 opt: 935 Z-score: 1886.6 bits: 357.8 E(): 3.5e-96
 Smith-Waterman score: 935; 60.991% identity (78.638% similar) in 323 aa overlap
 (129-444:1-316)

```

                100          110          120          130          140          150
2MEPSP GNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPP
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA MRERPIADLVDGLVQLGVKAECTMGTCPP
                        10          20          30

                160          170          180          190          200          210
2MEPSP VRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLM
      : . . . : : : : : : : : : . . . : : : : : : : : : : : : : :
Dad:CA VKIEA-DGLPGGRVELSGSVSSQYL TALLMAAPLCQGSIEIVIVDELISKPYVEMTITLM
                40          50          60          70          80

                220          230          240          250          260          270
2MEPSP ERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA ERFGVKVEKSADLQSFKIQGGQKYVSPGSAFVEGDASSASYFLAGATITGGTVTVIGCGS
                90          100          110          120          130          140

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      280      290      300      310      320      330
2MEPSP xSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA ESIQGD TNFAYTMEQMGATLEWGPNSVKCTGP-----QGPKAIDVNMNAMPDAAMTLA
      150      160      170      180      190      200

      340      350      360      370      380      390
2MEPSP VVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT-
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA VAALFADGVTTIRDVASWRVKETERMIAICTELRKLGCDVFEGSDYCVITPPHKLNPAPK
      210      220      230      240      250      260

      400      410      420      430      440
2MEPSP -----IDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA MRANVDIDTYDDHRMAMAFALAACGDVDVIINDPTCTKKTFTFYFDVLKSVVQ
      270      280      290      300      310

```

AAT45239.1 | *Helianthus salicifolius* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Helianthus salicifolius* | AA | 264
Length = 264
initn: 926 initl: 857 opt: 931 Z-score: 1879.7 bits: 356.2 E(): 8.5e-96
Smith-Waterman score: 931; 69.697% identity (85.985% similar) in 264 aa overlap (61-323:1-264)

```

      40      50      60      70      80
2MEPSP ILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCCKFPV-ED
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA RALGLNVEENGEIKRATVEGCGGVFPVGKE
      10      20      30

      90      100      110      120      130      140
2MEPSP AKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVD
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA AKDEIQFLGNAGTAMRPLTAAVTAAGNSSYILDGVPRMRERPIGDLVTGLKQLGADVD
      40      50      60      70      80      90

      150      160      170      180      190      200
2MEPSP CFLGTD CPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIP
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA CFLGTN CPPVRVAANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVP
      100      110      120      130      140      150

      210      220      230      240      250      260
2MEPSP YVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA YVEMTLKLMERFGVSVESDSDWKFYVRGGQKYKSPGNAYVEGDASSASYFLAGAAITGG
      160      170      180      190      200      210

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA TVTVEGCGTSSLQGDVKFAEVLGQMGAEVTWTENSVTVRGPPRNASGRGHLRPV
      220      230      240      250      260

```

330 340 350 360 370 380
2MEPSP MPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITE

[illegible]

2MEPSP KLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAG
Dad : BA KLISIPYVEMTLKLMERFVSV EHTDSWDRFFVRGGQKYKSPGNAYVEGDASSASYFLAG

270 280 290 300 310 320

2MEPSP AAIxxxxxxxxxxxxxxSLQGDVKFAEVLMMGAKVTWTETSVTVTGPPEPFGRKHLKAI
:::. :

Dad:BA AAITGGTITVEGCGTSSLQGDVKFAEVLGMGA EVTW TENS VTVKGPPRNP SGRKH LRAV

100 110 120 130 140 150

330 340 350 360 370 380
 2MEPSP DVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPD
 :
 Dad:BA DVNMNKMPPDVAMTLAVVALYADGPTTIRDVASWRVKETERMIAICTELRLKGATVEEGAD
 160 170 180 190 200 210

390 400 410 420 430 440
 2MEPSP YCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV TIRDPGCTRKTFPDYFDVLSTFV
 :.: .: .: .: .: .: .: .: .: .: .: .: .:
 Dad:BA YCVITPPEKLNVA AIDTYDDH
 220 230

BAE20404.1|*Lactuca sativa* 5-enolpyruvylshikimate-3-phosphate synthase
protein.|*Lactuca sativa*|AA|231
Length = 231
initn: 922 initl: 922 opt: 922 Z-score: 1862.4 bits: 352.8 E(): 7.9e-95
Smith-Waterman score: 922; 77.922% identity (89.610% similar) in 231 aa overlap
(174-404:1-231)

[illegible]

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      210      220      230      240      250      260
2MEPSP KLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAG
      .....
Dad:BA KLISIPYVEMTLKLMERFGVSVQHSDTWDRFHVQGGQKYKSPGNAYVEGDASSASYFLAG
      40      50      60      70      80      90

```

```

      270      280      290      300      310      320
2MEPSP AAIXXXXXXXXXXXXXSLQGDVKFAEVLLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAI
      ...
Dad:BA AAITGGTITVEGCGTSSLQGDVKFAEVLGQMGAAQVTWTENSVTVKGPPRDPGRKHLRPV
      100     110     120     130     140     150

```

```

      330      340      350      360      370      380
2MEPSP DVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPD
      .....
Dad:BA DVNMNKMPPDVAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGATVEEGPD
      160     170     180     190     200     210

```

```

      390      400      410      420      430      440
2MEPSP YCIITPPEKLNVTDAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFV
      .....
Dad:BA YCIITPPXKLNVTDAIDTYDDH
      220     230

```

AAT45235.1 | *Asimina triloba* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Asimina triloba* | AA | 264
Length = 264
initn: 897 initl: 872 opt: 887 Z-score: 1790.6 bits: 339.8 E(): 7.8e-91
Smith-Waterman score: 887; 66.288% identity (83.712% similar) in 264 aa overlap (60-323:1-264)

```

      30      40      50      60      70      80
2MEPSP RILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVCGGKFPVED
      .....
Dad:AA LQTLGLNVEEDSAANKATVEGCGGQFPVVGK
      10      20      30

```

```

      90      100     110     120     130     140
2MEPSP AKEEVQLFLGNAGIAMRSLXXXXXXXXXXXXXVLDGVPRMRERPIGDLVVGLKQLGADVD
      ...
Dad:AA DAKDIQLFLGNAGTAMRPLTAAVVAAGGKSRYVLDGVPRMRERPIGDLVSLKQLDADVD
      40      50      60      70      80      90

```

```

      150     160     170     180     190     200
2MEPSP CFLGTD CPPVRVNGIGGLPGGKVKLSGSISSQYXXXXXXXXXXXXGDVEIEIIDKLISIP
      :
Dad:AA CVLGTN CPPVSINANGGLRGKVKLSGTLSSQFLTSILMAAPLALGDVEIEIIDKLISVP
      100     110     120     130     140     150

```

```

      210      220      230      240      250      260
2MEPSP YVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxx
      .....
Dad:AA YVEMTLKLMERFGVSAEHSGSWDRFLVRGGQKYKSPGTAFVEGDASSASYFLAGAAVTGG
      160     170     180     190     200     210

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNK
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA TVTVEGCGTSSLQGDVKFAEVLEKMGAKVSWTENSVTVTGPPLDPSRKKRLHGI
      220      230      240      250      260

```

```

      330      340      350      360      370      380
2MEPSP MPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITP

```

ABE43726.1 | *Polaromonas* sp. JS666 3-phosphoshikimate 1-carboxyvinyltransferase / cytidylate kinase protein. | *Polaromonas* sp | AA | 668
Length = 668
initn: 875 initl: 313 opt: 880 Z-score: 1770.3 bits: 337.3 E(): 1.1e-89
Smith-Waterman score: 880; 46.120% identity (71.840% similar) in 451 aa overlap (1-439:1-440)

```

      10      20      30      40      50      60
2MEPSP MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB MFDIEYLDIPPLARAGGTVRLPGSKSISNRVLLLAALSRGQTTVHDLLASDDTAVMLAAL
      10      20      30      40      50      60

```

```

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . : : : . . : : : : : : : : : : : : : : : : : : : : : :
Dad:AB KQLGCSVA--QHGTTAVIDGLGGQ--VGQAK--ATLFMGNAGTAMRPLTAALALLGGE--
      70      80      90      100      110

```

```

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB FELSGVARMHERPIGDLVDALRQLGCSIEYLGNEGYPPLRLRPAQLKIEEPIRVRGDVSS
      120      130      140      150      160      170

```

```

      190      200      210      220      230
2MEPSP QYxxxxxxxxxxxxxG-DVEIEIIDKLISIPYVEMTLRLMERFGVKAEH--SDSWDRFYI
      : . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB QFLTALLMALPLVAGQDIHIEVVGELISRPIEITLNLKRFGIHVQRPLGADNWQRFTH
      180      190      200      210      220      230

```

```

      240      250      260      270      280      290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxx----xxSLQGDVKFAEVL
      . : : : : : . : : : : : : : : : : : : : : : : : : : : : :
Dad:AB PAGSQYQSPGDIHVEGDASSASYFIALGAISEPAAGQNHIEILGVGADSIQGDIFIEAA
      240      250      260      270      280      290

```

```

      300      310      320      330      340      350
2MEPSP EMMGAKVTWTETSVTVT-GPPREPFGRKHLKAIDVNMNMKMPDVAMTLAVVALFADGPTAI
      . : : : . : . . : : . . : : : : : : : : : : : : : : : :
Dad:AB RMMGAQIESTPNSLRISRGKPGQGW--PLKAIDLDCNHIPDAAMTLAVMALYADGTTTL
      300      310      320      330      340

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                360          370          380          390          400          410
2MEPSP RDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAF
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB RNIASWRVKETDRIAAMACELQKLGATVEEGADYLKITPPLAWKSAAIHTYDDHRIAMCF
      350          360          370          380          390          400

```

```

                420          430          440
2MEPSP SLAAC--AEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB SLAAFNPARQPIRILDPKCVAKTFPDYFEALFAVSEAETAQIPVICIDGPTASGKGTLLA
      410          420          430          440          450          460

```

```

Dad:AB LAAHRLGYHYLD SGALYRLSAFAATRAGVALENGADVARIARTLPVFRGRDRIFLGTEDV
      470          480          490          500          510          520

```

ABM42629.1 | *Acidovorax* sp. JS42 3-phosphoshikimate 1-carboxyvinyltransferase protein. | *Acidovorax* sp | AA | 673
Length = 673
initn: 834 initl: 238 opt: 870 Z-score: 1750.0 bits: 333.6 E(): 1.4e-88
Smith-Waterman score: 870; 46.429% identity (72.545% similar) in 448 aa overlap (9-439:9-441)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB MFSTAFLDLPLASVQGSVQLPGSKSISNRVLLLAALSQGTTEVRDLLASDDTRVMLDAL
      10          20          30          40          50          60

```

```

                70          80          90          100          110
2MEPSP RTLGLSVEADKAAKRAV-VVGC GGFVVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB RQLGCTV--DEAGGGTVRITGLGSSTP----PSPTQLFMGNAGTAMRPLTAALALLGGE-
      70          80          90          100          110

```

```

                120          130          140          150          160          170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPG---GKVKLS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB -YELSGVPRMHERPIGDLVDALRQLGCRIDYLGNGQGFPLRIAHAGGLPPLQLQAPIRVR
      120          130          140          150          160          170

```

```

                180          190          200          210          220          230
2MEPSP GSISSQYxxxxxxxx--xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDR
      : : : : : . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB GDVSSQFLTALLMALPLVARQQDVVIEVVGELISKPYIHITLELLARFGIAVQH-EQWQR
      180          190          200          210          220          230

```

```

                240          250          260          270          280
2MEPSP FYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxx-----SLQGDVKFA
      : : . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB FTIPAGSRYQSPGLIHVEADASSASYFIAAGAIAGASGQKGIKILGVGLDSIQGDIRFV
      240          250          260          270          280          290

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      240      250      260      270      280      290
2MEPSP YIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxx---SLQGDVKFAEVL
      : . . . . . : . . . . . : . . . . . : . . . . . : . . . . . :
Dad:AB TIPAGSRYSPPGEIHVEADASSASYFIALGAIASDSAEGIRILGVGLDSIQGDIRFVEAA
      240      250      260      270      280      290

      300      310      320      330      340      350
2MEPSP EMMGAKVTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIR
      . : . . . : . . . : . : . . . : . : . . . : . : . . . : . :
Dad:AB RAMGAEEVEGGPNWLRIRR-----GAWPLKAIDLDCNHIPDAAMTLAVMALYADGTTTLR
      300      310      320      330      340

      360      370      380      390      400
2MEPSP DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKL---NVT AIDTYDDHRMAM
      . . . . . : . : . : . : . : . : . : . : . : . : . : . : . :
Dad:AB NIASWRVKETDRIAAMTKELRKLGADEEGADYIRITPPASTAAWRAASIHTYDDHRVAM
      350      360      370      380      390      400

      410      420      430      440
2MEPSP AFSLAAC--AEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . . . : . : . : . : . : . : . : . : . : . :
Dad:AB CFSLAAFNPALPVRIEDPKCVAKTFPDYFEALFVCETARARIPVICIDGPTASGKGTI
      410      420      430      440      450      460

Dad:AB AAEVAQALGYHLLDSGALYRLTGLAASRAGLVLDDEPHAQQIAVLARELPVRFDAEQRIWL
      470      480      490      500      510      520

```

ACB44430.1 | *Polynucleobacter necessarius* STIR1 3-phosphoshikimate
carboxyvinyltransferase protein. | *Polynucleobacter necessarius* STIR1 | AA | 442
Length = 442
initn: 781 initl: 214 opt: 860 Z-score: 1732.5 bits: 329.7 E(): 1.4e-87
Smith-Waterman score: 860; 46.517% identity (69.213% similar) in 445 aa overlap
(1-436:1-434)

```

      10      20      30      40      50      60
2MEPSP MAGAE EIVLQPIKEISGTVKLP GSKSLSNRILL AALSEGTTVVDNLLN SEDVHYMLGAL
      : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Dad:AC MSGLPDIQIGPFKQAQGSIVLP GSKSISNRALL AALSTGTTTLKNLLDADDTQVMRNAL
      10      20      30      40      50      60

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : . . : . : : . : . : . : . : . : . : . : . : . : . : . :
Dad:AC RQLGLSV-INKENKACIVEGCGGKLPVQEA---DLFMGNAGTAIRPLTAALAMQGGN--
      70      80      90      100      110

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : . : . : . : . : . : . : . : . : . : . : . : . : . :
Dad:AC YRLSGVARMRERPIRDLVDGLRQVGAKIGYELQEGYPPIKILAADIQIKDVVKVRGDVSS
      120      130      140      150      160      170

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                190          200          210          220          230
2MEPSP QYxxxxxxxxxxxxGD-VEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG-
      . . . . . : : : : : : : : : : : : : : : : : : : : : : :
Dad:AC QFLTALLMALPLVANEPVRIEVIGELISRPYIDITLKL MARFGVTVACPD M-QSFVIPAK
                180          190          200          210          220          230

                240          250          260          270          280          290
2MEPSP --GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AC TSGAVYKSPGQLLVEGDASSASYFLALGAIGGPPVRVLGVGKESIQGDVAFADALALMGA
                240          250          260          270          280          290

                300          310          320          330          340          350
2MEPSP KVTWTETSVTVTGPPREPFG R KHLKAIDVNMNKM P DVAMTLAVVALFADGPTAIRDVASW
      . . : . : : . . : . : : . . : . : : : : : : : : : : : :
Dad:AC NITAGEDWIEVAGV-KNANGK--LNGITIDCTEIPDAAMTLAVAALFAEGPTRLNSIASW
                300          310          320          330          340

                360          370          380          390          400          410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPE-----KLNVT AIDTYDDHRMAMAFS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AC RVKETDRIAAMAKELKKVGANVEEGADYIVVQAPVLQSDWKSPSEGVDTYDDHRMAMCFS
                350          360          370          380          390          400

                420          430          440
2MEPSP LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : . . : : : : : : : : : : :
Dad:AC LAAFGLNALKINDPNCVAKTFPTYFAEFAKVVN
                410          420          430          440

```

ABP33718.1 | *Polynucleobacter* sp. QLW-P1DMWA-1 3-phosphoshikimate 1-carboxyvinyltransferase protein. | *Polynucleobacter* sp | AA | 442
Length = 442
initn: 780 initl: 205 opt: 854 Z-score: 1720.4 bits: 327.5 E(): 6.4e-87
Smith-Waterman score: 854; 46.517% identity (69.663% similar) in 445 aa overlap (1-436:1-434)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL LLAALSEGTTVVDNLLN SEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB MSGIPEITIGPFTRAQGSIVLP GSKSISNRALL LLAALASGTTTLKNLLDADDTQVMRNL
                10          20          30          40          50          60

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB RQLGLSV-IDQANHVCVVGKCGGQFPVREA----DLFMGNAGTAIRPLTAALAMQGGN--
                70          80          90          100          110

                130          140          150          160          170          180
2MEPSP YVL DGVPRMRERPIGDLVVG LKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB YRLSGVARMHERPIRDLVDGLRQVGAKIDYELQEGYPPIKILAADIQIKDVVKVRGDVSS
                120          130          140          150          160          170

```

CAJ48961.1|*Bordetella avium* 197N 3-phosphoshikimate 1-carboxyvinyltransferase protein.|*Bordetella avium* 197N|AA|439
Length = 439
initn: 635 initl: 230 opt: 853 Z-score: 1718.4 bits: 327.1 E(): 8.3e-87
Smith-Waterman score: 853; 45.805% identity (71.655% similar) in 441 aa overlap (9-443:9-437)

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                190                200                210                220                230
2MEPSP QYxxxxxxxxxxxxG----DVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA QFLTALLLAAPIEAGASGRPVTVIEVIGELISKPYIEITLNLMARYGVNVVR-DGWRAFTI
                180                190                200                210                220                230

                240                250                260                270                280                290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      . : . : : . : : : : : : : : : . : : : : : : : : : : : : : : :
Dad:CA EGDARYRSPGSIAVEGDASTASYLLALGVLGGGPVRVTGVGEQSIQGDTAFAADTLAAMGA
                240                250                260                270                280                290

                300                310                320                330                340                350
2MEPSP KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASW
      . : . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA NITKGSDWIEASGQAVAEGR--IKAFDADFNLIPDAAMTAATMALFADGPCRLRNIGSW
                300                310                320                330                340

                360                370                380                390                400                410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLVN--TAIDTYDDHRMAMAFSLAA
      : : : : . : : : : : : : : : . : : . . : : : : : : : : : : : :
Dad:CA RVKETDRIHAMHTELAKLGAKVESGPDWLSLTPPADSDWRDAHIGTWDDHRMAMCFSLAA
                350                360                370                380                390                400

                420                430                440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : : : : : : : . :
Dad:CA FGPAAVRILDPGCVSKTFPDYFDVYAGLVSA
                410                420                430

```

ABM33844.1 | *Acidovorax avenae* subsp. *citrulli* AAC00-1 3-phosphoshikimate 1-carboxyvinyltransferase protein. | *Acidovorax avenae* subsp | AA | 679
Length = 679
initn: 786 initl: 215 opt: 854 Z-score: 1717.5 bits: 327.6 E(): 9.3e-87
Smith-Waterman score: 854; 45.190% identity (70.917% similar) in 447 aa overlap (9-439:9-443)

```

                10                20                30                40                50                60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGSKLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB MYSTAFDLPLDPTAGGAVRLPGSKSISNRVLLLAALSEGTTVEVDLLASDDTRVMDAL
                10                20                30                40                50                60

                70                80                90                100                110                120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      : : . : . : : . : : : : : : : : : : : : : : : : : :
Dad:AB REIGCGVDEGAAAQGTVRITGLGTSP---ARSPSKLFLGNAGTAMRPLTAALALLGGE--
                70                80                90                100                110

                130                140                150                160                170
2MEPSP YVLGCVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLP---GGKVKLSG
      . : : : : : : : : : . : : : . . . . : : : : : : : : : : : :
Dad:AB FELSGVPRMHERPIGDLVEALLQLGCHISYLGNGPFPPLRIAHAGGVPLALDAPVRVRG
                120                130                140                150                160                170

```

	180	190	200	210	220	230
2MEPSP	SISSQYxxxxxxxx--xxxxG	DVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRF				
	:	: : . .	:	:
Dad:AB	DVSSQFLTALLMALPLVAREKD	VVIEVVGELISRPIIHITLQLLERFGIRVRH-DEWQRF				
	180	190	200	210	220	230
	240	250	260	270	280	
2MEPSP	YIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxx-	-----xxxxxxxxxSLQGDKFAE				
	:	:	:	:
Dad:AB	TIPAGSRYSRPGTIHVHEADASSASYFIALGALAAPAPGQEPLRILGVGLDSIQGDIRFAE					
	240	250	260	270	280	290
	290	300	310	320	330	340
2MEPSP	VLEMMGAKVTWTETSVTVTGPPREPFRGRKHLKAIDVN	MNKMMPDVAMTLAVVALFADGPTA				
	:	:	:	:	:
Dad:AB	AARAMGA EVTGGPNLVVRR-----GAWPLRAVDLCDNHIPDAAMTLAVMALYAQGT	TTT				
	300	310	320	330	340	
	350	360	370	380	390	400
2MEPSP	IRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK---	LNVTAIDTYDDHRM				
	:	:	:	:	:
Dad:AB	LRNIASWRVKETDRIAAMAQGCRRLGATVEEGSDFLRVTPPASPADWRAASIHTYDDHRI					
	350	360	370	380	390	400
	410	420	430	440		
2MEPSP	AMAFSLAAC--AEVPVTIRDPGCTRKTFPDYFDVLSTFVK	N				
	:	:	:	:		
Dad:AB	AMCFSLA AFNP AKLPVRIEDPKCVAKTFPDYFEALFSVAGTPVEHVPVICIDGPTASGKG					
	410	420	430	440	450	460
Dad:AB	TVAAAVAQRLGYRFLDSGAMYRITALAALRAGLSIDA AHQDRIAALARTLPVRFESGRIW					
	470	480	490	500	510	520

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          . : : : . : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB          MPGSKSISNRILLLAALSEGVTDVCDLLASDDTARMLDAL
                10          20          30          40

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxxxx
          : : : . . . . . : : : : : : : : : : : : : : : : : : :
Dad:AB  STLGVSI-LQIGRDHYRLQGVGDQFPLRLPTTEADLFLGNAGTVFRPLTAMLALAQGH--
                50          60          70          80          90

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTD-CPPVRVNGIGGLPGGKVKLSGSIS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB YRLSGVPRMHERPIGDLVDALRQVGADIT-YLGKEGFPPLQIKP-GRIHPGEITVRGEVS
      100          110          120          130          140          150

                180          190          200          210          220          230
2MEPSP SQYxxxxxxxxxxxxxxGDVE----IEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFY
      : : . . . . . : . : : : : : : : : : : : : : : : : : : : : :
Dad:AB SQFLTALLMVLPLFLRAEMDELPVITVAGELISRPIYDLTIALMARFGVQVER-EEWRRFT
      160          170          180          190          200          210

                240          250          260          270          280          290
2MEPSP IKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMG
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB VPADQRYRSPGQVFVEGDASSASYFLAAGAIGRGPVRVEGLGRDSVQGDIRFAEALERMG
      220          230          240          250          260          270

                300          310          320          330          340          350
2MEPSP AKVTWTETSVITVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVAS
      : . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB ADIRFGDNWIEASGP-----GPGGLRAIDLDCNHIPDAAMTLAVTALFARGNTVLRNIAS
      280          290          300          310          320

                360          370          380          390          400          410
2MEPSP WRVKETERMVAIRTELTKLGASVEEGPDYCIITPP--EKLNVTAIDTYDDHRMAMAFSLA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB WRVKETDRIAAMAQELRKLGAEEVAGSDFLQISPPRGELVANAAIDTYDDHRMAMCFSLV
      330          340          350          360          370          380

                420          430          440
2MEPSP ACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : : : : : : : : : : : : : : : : : : :
Dad:AB SFG-APVRINDPRCVSKTFPDYFEKFAAIAYADPGQGKFAARIDSSDIS
      390          400          410          420          430

```

AAZ46031.1 | *Dechloromonas aromatica* RCB cytidylate kinase / 3-phosphoshikimate 1-carboxyvinyltransferase protein. | *Dechloromonas aromatica* RCB | AA | 643
Length = 643
initn: 803 initl: 220 opt: 851 Z-score: 1711.8 bits: 326.4 E(): 1.9e-86
Smith-Waterman score: 911; 48.268% identity (72.055% similar) in 433 aa overlap (5-437:2-418)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : . : . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA MEFLDLPQLLSASGTVRLPGSKSISNRVLLLAALSGGETEVRDLLGSDDTERMLEAL
      10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      . : : : : . . . . : : : : : . . . . : : : : : : : : : . . . .
Dad:AA KALGVGVQ-HLGGENWSIKGCGGHFPV---KQIELFLGNAGTAFRPLTAALALAGGD--
      60          70          80          90          100          110

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        ::: : : ::::::::::: ::::::::::. . . : : : : : : : : : : : : : :
Dad:AA  YVLKGVARMHERPIGDLVDGLRQLGADITYLGNEGFPLHLKPASILPNGVVTVRGDVSS
                120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :. . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  QFLTGLLMALPLTGETVVVEVLGELISKPYIEITLATMARFGVIVER-EGWQRFTVLAGS
                180          190          200          210          220

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        .: : : . : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  RYVSPGTIFVEGDASSASYFLALGAIGGPPVRVEGVGRDSIQGDVKFAEALVKMGAQIEM
                230          240          250          260          270          280

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
        . . . . : . : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  GPNWMEARAP-----QAGLVAVDLDCNHIPDAAMTLATTALFAKGTTTLRNIASWRVKE
                290          300          310          320          330          340

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  TDRIAAMATELRKLGAVVDEGDDYIRVTPAD-LKPAAIDTYDDHRMAMCFSLAAGF-TPL
                350          360          370          380          390          400

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        : : : : : : : :
Dad:AA  RINDPKCVAKTFPDYFERFAGVTKAAPVIAIDGPSASGKGTVAARIALALGYDYLDSGAL
                410          420          430          440          450          460

```

AAT45241.1 | *Nymphaea alba* 5-enol-pyruvylshikimate-phosphate synthase protein. | *Nymphaea alba* | AA | 264

Length = 264

initn: 838 initl: 757 opt: 848 Z-score: 1711.7 bits: 325.1 E(): 2e-86

Smith-Waterman score: 848; 65.098% identity (82.745% similar) in 255 aa overlap (60-313:1-255)

```

                30          40          50          60          70          80
2MEPSP  RILLLAALSEGTTVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-E
        : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  LRTLGLRVDEDRDMKRAIVEGCSGQFPVAK
                10          20          30

                90          100          110          120          130          140
2MEPSP  DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADV
        :. : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA  DSAKEVELFLGNAGTAMRPLTAAVVAAGGNTRYILDGVPRMRERPIGDLVSLQLQLGADI
                40          50          60          70          80          90

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      150      160      170      180      190      200
2MEPSP DCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISI
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA GCTLGTNCPVYINGKGGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEVEMADKLVS
      100      110      120      130      140      150

```

```

      210      220      230      240      250      260
2MEPSP PYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA PYVEMTLKLMERFGVAVEHGGGWDRLIRGRQMYMSPGSAYVEGDASSASYFLAGAAITG
      160      170      180      190      200      210

```

```

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMN
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA GTVTVEGCGTSSSLQGDVKFAEVLEKMGAKVTWSQNSVTVTGPPKDGSRTRLRDI
      220      230      240      250      260

```

```

      330      340      350      360      370      380
2MEPSP KMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT

```

CAP42227.1 | *Bordetella petrii* 3-phosphoshikimate 1-carboxyvinyltransferase protein. | *Bordetella petrii* | AA | 447
Length = 447
initn: 755 initl: 230 opt: 847 Z-score: 1706.1 bits: 324.9 E(): 4e-86
Smith-Waterman score: 847; 45.880% identity (69.710% similar) in 449 aa overlap (1-443:1-438)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : . : . : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA MTGQPYLDLPRARQARGQVALPGSKSISNRVLLLAALAAGRTDISGLLDSDDTRVMLAAL
      10      20      30      40      50      60

```

```

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      : : : : . : . : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA RQLG--VELAEAGEGRVTVGGAGRFPVKQA---DLFLGNAGTAFRPLTAALALMG--Q
      70      80      90      100      110

```

```

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA YRLSGVPRMHERPIGDLVDALRQWGARIDYLGQAGYPPLAVGEGRIRADAPARVQGA VSS
      120      130      140      150      160      170

```

```

      190      200      210      220      230
2MEPSP QYxxxxxxxxx-----xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      : . . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA QFLTALLLAAPVLAQGSDRPVVIEVAGELISKPYIEITLNLMARYGVQVRR--DGWRTFTI
      180      190      200      210      220      230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      240      250      260      270      280      290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Dad:CA EPGAAYRSPGAIAVEGDASSASYFLALGAIGGGPVRVTGVGADSIQGDVAFARTLADMGV
      240      250      260      270      280      290

```

```

      300      310      320      330      340      350
2MEPSP KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      . . . . . : . : . : . : . : . : . : . : . : . : . : . : . :
Dad:CA QIDYGPDWIEARGVRVDQGG--LKAFDTDFNLIPDAAMTAAALALYADGPCRLRNIGSW
      300      310      320      330      340

```

```

      360      370      380      390      400      410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Dad:CA RVKETDRIHAMHTELAKLGAEVESGPDWLRITPPADGGWRDAHIGTWDDHRMAMCFSLAA
      350      360      370      380      390      400

```

```

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
      . . : : . : . : . : . : . : . : . : . : . : . : . : . : . :
Dad:CA FGPAAVRILDPGCVSKTFFPDYFDVYAGLVSGAPDSYDD
      410      420      430      440

```

CAE41250.1|*Bordetella pertussis* 3-phosphoshikimate 1-carboxyvinyltransferase protein.|*Bordetella pertussis*|AA|442
Length = 442
initn: 605 initl: 217 opt: 845 Z-score: 1702.2 bits: 324.1 E(): 6.6e-86
Smith-Waterman score: 845; 45.495% identity (69.595% similar) in 444 aa overlap (1-438:1-433)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Dad:CA MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGSTIEITGLLDSDDTRVMLAAL
      10      20      30      40      50      60

```

```

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Dad:CA RQLGVSV--GEVADGCVTIEGVARFPTEQA---ELFLGNAGTAFRPLTAALALMGGD--
      70      80      90      100      110

```

```

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
Dad:CA YRLSGVPRMHERPIGDLVDALRQFGAGIEYLGQAGYPPLRIGGGSIRVDGPVRVEGSVSS
      120      130      140      150      160      170

```

```

      190      200      210      220      230
2MEPSP QYxxxxxxxxx----xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      . . . . . : . : . : . : . : . : . : . : . : . : . : . : . :
Dad:CA QFLTALLMAAPVLARRSGQDITIEVVGELISKPYIEITLNLMARFGVSVRR-DGWRAFTI
      180      190      200      210      220      230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      240      250      260      270      280      290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      . . . . . : : : : : : : : : : . . . . . : : : : : : : : : :
Dad:CA ARDAVYRGPGRMAIEGDASTASYFLALGAIGGGPVRVTGVGEDSIQGDVAFAATLAAMGA
      240      250      260      270      280      290

      300      310      320      330      340      350
2MEPSP KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      : . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA DVRYGPGWIETRGRVVAEGGR--LKAFDADFNLIIPDAAMTAATLALYADGPCRLRNIGSW
      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      : : : . . . : : : : : : : . . . . . : . . : : : : : : : :
Dad:CA RVKETDRIHAMHTELEKLGAGVQSGADWLEVAPPEPGGWRDAHIGTWDDHRMAMCFSLAA
      350      360      370      380      390      400

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : . : : : : : : : : :
Dad:CA FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
      410      420      430      440

CAE38415.1 | Bordetella parapertussis 3-phosphoshikimate 1-carboxyvinyltransferas
e protein. | Bordetella parapertussis | AA | 442
Length = 442
initn: 600 initl: 212 opt: 843 Z-score: 1698.1 bits: 323.4 E(): 1.1e-85
Smith-Waterman score: 843; 45.270% identity (69.369% similar) in 444 aa overlap
(1-438:1-433)

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGSTEITGLLDSDDTRVMLAAL
      10      20      30      40      50      60

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      : : : : . . . . . : : : : : : : : : : : : : : : : : : :
Dad:CA RQLGVSV--GEVADGCVTIEGVARFPFIEQA---ELFLGNAGTAFRPLTAALALMGGD--
      70      80      90      100      110

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : . . . . . : : : : : : : : : :
Dad:CA YRLSGVPRMHERPIGDLVDALRQFGAGIEYLGQAGYPPLRIGGGSIRVDGPVRVEGSVSS
      120      130      140      150      160      170

      190      200      210      220      230
2MEPSP QYxxxxxxxxx----xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      : . . . . . : . . . . . : : : : : : : : : : : : : : : : :
Dad:CA QFLTALLMAAPVLARRSGQDITIEVVGELISKPYIEITLNLMARFGVSVRR-DGWRAFTI
      180      190      200      210      220      230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      240      250      260      270      280      290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      . . . . . : : : : : : : : : : . . . . . : : : : :
Dad:CA ARDAAYRGPGRMAIEGDASTASYFLALGAIGGGPVRVTGVGEDSIQGDVAFATLAAMGA
      240      250      260      270      280      290

```

```

      300      310      320      330      340      350
2MEPSP KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      : . . . : : : : : : : : : : : : : : : : :
Dad:CA DVRYGPGWIETRGVRVAEGGR--LKAFDADFNLIIPDAAMTAATLALYADGPCRLRNIGSW
      300      310      320      330      340

```

```

      360      370      380      390      400      410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      : : : . . : : : : : : : : . . : : : : : : : :
Dad:CA RVKETDRIHAMHTELEKLGAGVQSGADWLEVAPPAPGGWRDAHIGTWDDHRMAMCFSLAA
      350      360      370      380      390      400

```

```

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : . : : : : :
Dad:CA FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
      410      420      430      440

```

AAF01290.1 | *Bordetella bronchiseptica* 5-enolpyruvylshikimate-3-phosphate synthase protein. | *Bordetella bronchiseptica* | AA | 442
Length = 442
initn: 600 initl: 212 opt: 841 Z-score: 1694.1 bits: 322.6 E(): 1.9e-85
Smith-Waterman score: 841; 45.495% identity (69.595% similar) in 444 aa overlap (1-438:1-433)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : . . : : : : : : : : : : : : : : : : :
Dad:AA MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGSTIEITGLLDSDDTRVMLAAL
      10      20      30      40      50      60

```

```

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      : : : : . . : : . . : : : : : : : : : : :
Dad:AA RQLGVSV-GEVADGRVTIEGVA-RFPTEQA---ELFLGNAGTAFRPLTAALALMGGD--
      70      80      90      100      110

```

```

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : . . . : : : : :
Dad:AA YRLSGVPRMHERPIGDLVDALRQFGAGIEYLGQAGYPPLRIGGGSIRVDGPVRVEGSVSS
      120      130      140      150      160      170

```

```

      190      200      210      220      230
2MEPSP QYxxxxxxxxx-----xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      : . . . . . : : : : : : : : : : : : : :
Dad:AA QFLTALLMAAPVLARRSGQDITIEVVGELISKPYIEITLNLMARFGVSVRR-DGWRAFTI
      180      190      200      210      220      230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      240      250      260      270      280      290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      . . . . . : : : : : : : : : : . . . . . : : : : : : : : : :
Dad:AA ARDAAYRGPGRMAIEGDASTASYFLALGAIGGGPVRVTGVGEDSIQGDVAFATLAAMGA
      240      250      260      270      280      290

      300      310      320      330      340      350
2MEPSP KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      : . . . . : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA DVRYGPGWIETRGVRVAEGGR--LKAFDADFNLIPDAAMTAATLALYADGPCRLRNIGSW
      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      : : : : . . : : : : : : : : : : . . . . : : : : : : : : : :
Dad:AA RVKETDRIHAMHTELEKLGAGVQSGADWLEVAPPAPGGWRDAHIGTWDDHRMAMCFSLAA
      350      360      370      380      390      400

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : . : : : : : : : : :
Dad:AA FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
      410      420      430      440

CAE33961.1|Bordetella bronchiseptica 3-phosphoshikimate 1-carboxyvinyltransfera
se protein.|Bordetella bronchiseptica|AA|442
Length = 442
initn: 600 initl: 212 opt: 840 Z-score: 1692.0 bits: 322.2 E(): 2.4e-85
Smith-Waterman score: 840; 45.270% identity (69.369% similar) in 444 aa overlap
(1-438:1-433)

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGSTEITGLLDSDDTRVMLAAL
      10      20      30      40      50      60

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      : : : : . . : : . . : : : : : : : : : : : : : : : : : : .
Dad:CA RQLGVSV--GEVADGCVTIEGVARFPTEQA---ELFLGNAGTAFRPLTAALALMGGD--
      70      80      90      100      110

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : . . . . : : : : : : : : : : : :
Dad:CA YRLSGVPRMHERPIGDLVDALRQFGAGIEYLGQAGYPPLRIGGGSIRVDGPVRVEGSVSS
      120      130      140      150      160      170

      190      200      210      220      230
2MEPSP QYxxxxxxxxx----xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      : . . . . . : : : : . . : : : : : : : : : : : : : : : : : :
Dad:CA QFLTALLMAAPVLARRSGQDITIEVVGELISKPYIEITLNLMARFGVSVRR-DGWRAFTI
      180      190      200      210      220      230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      240      250      260      270      280      290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      . . . . . : : : : : : : : : : . . . . . : : : : : : : : : :
Dad:CA ARDAAYRGPGRMAIEGDASTASYFLALGAIGGGPVRVTGVGEDSIQGDVAFAATLAAMGA
      240      250      260      270      280      290

      300      310      320      330      340      350
2MEPSP KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      : . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:CA DVRYGPGWIETRGRVVAEGGR--LKAFDADFNLIPDAAMTAATLALYADGPCRLRNIGSW
      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      : : : . . . : : : : : : : : . . . . . : : : : : : : : : :
Dad:CA RVKETDRIHAMHTELEKLGAGVQSGADWLEVAPPAPGGWRDAHIGTWDDHRMAMCFSLAA
      350      360      370      380      390      400

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : . : : : : : : : : :
Dad:CA FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
      410      420      430      440

AAA22968.1|Bordetella pertussis protein ( B. pertussis 5-enolpyruvylshikimate-
3-phosphate synthase (aroA) gene, complete cds. ).|Bordetella pertussis|AA|442
Length = 442
initn: 605 initl: 217 opt: 839 Z-score: 1690.0 bits: 321.9 E(): 3.2e-85
Smith-Waterman score: 839; 45.270% identity (69.369% similar) in 444 aa overlap
(1-438:1-433)

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGSTEITGLLDSDDTRVMLAAL
      10      20      30      40      50      60

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      : : : : . . . . . : : : : : : : : : : : : : : : : : : : :
Dad:AA RQLGVSV--GEVADGCVTIEGVARFPTEQA---ELFLGNAGTAFRPLTAALALMGGD--
      70      80      90      100      110

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : . . . . . : : : : : : : : : :
Dad:AA YRLSGVPRMHERPIGDLVDALRQFGAGIEYLQAGYPPLRIGGGSIRVDGPVRVEGSVSS
      120      130      140      150      160      170

      190      200      210      220      230
2MEPSP QYxxxxxxxxx----xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      : . . . . . : . . . . . : : : : : : : : : : : : : : : : : :
Dad:AA QFLTALLMAAPVLARRSGQDITIEVVGELISKPYIEITLNLMARFGVSVRR-DGWRAFTI
      180      190      200      210      220      230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      240      250      260      270      280      290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA ARDAVYRGPGRMAIEGDASTASYFLALGAIGGGPVRVTGVGEDSIQGDVAFATLAAMGA
      240      250      260      270      280      290

      300      310      320      330      340      350
2MEPSP KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      : . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA DVRYGPGWIETRGRVRAEGGR--LKAFDADFNLIIPDAAMTAATLALYADGPCRLRNIGSW
      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      : : : . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA RVKETDRIHAMHTELEKLGAGVQSGADWLEVAPPEPGGWRDAHIGTWDDHRMAMCFLAA
      350      360      370      380      390      400

      420      430      440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AA FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
      410      420      430      440

```

ABM95199.1 | *Methylibium petroleiphilum* PM1 3-phosphoshikimate 1-carboxyvinyltransferase / cytidylate kinase protein. | *Methylibium petroleiphilum* PM1 | AA | 674
Length = 674
initn: 847 initl: 268 opt: 831 Z-score: 1671.0 bits: 319.0 E(): 3.6e-84
Smith-Waterman score: 831; 44.118% identity (71.267% similar) in 442 aa overlap (9-439:10-438)

```

      10      20      30      40      50
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGA
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB MATPRAFLDIPPLQAAGGTVRLPGSKSISNRVLLLAGLCAGRTRVLDLLSDDTQVMLDA
      10      20      30      40      50      60

      60      70      80      90      100      110
2MEPSP LRTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : . . : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB LRALGCDIETDGAAR--VVTGLGGRLAVREAR---LFLGNAGTAMRPLAALALLAADQ
      70      80      90      100      110

      120      130      140      150      160      170
2MEPSP x--YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGS
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dad:AB GGRFELSGVPRMHERPIGDLVDALRPLGCTITCLANEGYPPLRLERGTCLKLDAPIRVRGD
      120      130      140      150      160      170

      180      190      200      210      220      230
2MEPSP ISSQYxxxxxxxxxxxxxG--DVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFY
      : : : . . . . . : : : : : : : : : : : : : : : : : : : : :
Dad:AB VSSQFLTALLMALPLVAARQSITIEVDGELISKPYIEITLALLARFGISVQR-EGWQRFV
      180      190      200      210      220      230

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	240	250	260	270	280	290
2MEPSP	IKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxx--xxxxxxxSLQGDVKFAEVLEM					
	:	:	:	:	:	:
Dad:AB	IPQGSAYRSPGEIAVEGDASSASYFIAAGAAIAAADTPLRIEGVGSASIQGDIRFVEAARA					
	240	250	260	270	280	290
	300	310	320	330	340	350
2MEPSP	MGAKVTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDV					
	::: .:	: : : : .	: : : : .	: : : : .	: : : : .	: : : : .
Dad:AB	MGADITEEANALVVRR-----GAWPLTAITLDCNHIPDAAMTLAAMALYATGTTRLTNI					
	300	310	320	330	340	
	360	370	380	390	400	410
2MEPSP	ASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLA					
	:	: :	: :	: :	: :	: :
Dad:AB	ASWRVKETDRIAAMAIELRKFGATVLEGTDFIEVTPPARWQAAAIHTYDDHRMAMCASLA					
	350	360	370	380	390	400
	420	430	440			
2MEPSP	AC-----AEVPVTIRDPGCTRKTFPDYFDVLSTFVKN					
	:	: :	: :	: :	: :	: :
Dad:AB	AFNPLAGGDVPVRILDPKCNKTFPAYFDALFGVTRARTDRVPVLTVDGPTASGKGTLS					
	410	420	430	440	450	460
Dad:AB	ALAERLGYHHLDSGALYRATALAALRQGVPADEAAVAIAIALPLRFENQQTLLAGEDV					
	470	480	490	500	510	520

CAD83448.1 | *Candidatus Blochmannia floridanus* 3-phosphoshikimate 1-carboxyvinyltransferase protein. | *Candidatus Blochmannia floridanus* | AA | 433
Length = 433
initn: 773 initl: 286 opt: 818 Z-score: 1647.6 bits: 314.0 E(): 7.2e-83
Smith-Waterman score: 818; 43.187% identity (70.670% similar) in 433 aa overlap (7-437:5-424)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLPKSKLSNRILLAAALSEGTTVVDNLLNSEDVHYMLGAL					
	:	:	:	:	:	:
Dad:CA	MENFITLNPPIKINGTIYLPKSKSISNRALLAAQSVGATRLINLLSDDDTRYMLAAL					
	10	20	30	40	50	
	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx					
	:	:	:	:	:	:
Dad:CA	LQLGVKYKLSTDHKICEIKGVGGS--LQSKNTQTTLFLGNAGTAIRPLIAALSISKSHNI-					
	60	70	80	90	100	110
	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGKVKLSGSISS					
	: : :	: : :	: : :	: : :	: : :	: : :
Dad:CA	-VLTGDMRMKERPIFHLVDTLRQGGRIEYIERDHHLVPKL--YGGYCGGNIVIKGDISS					
	120	130	140	150	160	170

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQ					
	:. :. . . :. :. :. :.					
Dad:CA	QFLSAVLMMTPLASKNTYIEVIGSLVSKPYIDITLSVMRAFGICVQHDKDYKFFYCEGNR					
	180	190	200	210	220	230
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW					
	:. :. :. :. :. :.					
Dad:CA	IYKAPQEYVIEGDASSASYFLAAAIKGGTIRVLGVGKNSKQGDIKFANILERMGSIIITW					
	240	250	260	270	280	290
	310	320	330	340	350	
2MEPSP	TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADG--PTAIRDVASWRV					
	. . . :. :. :. :.					
Dad:CA	GDNYI-----ECSKGIMLKSIDIDVNDIPDAAMTLAVIALFTTNNLPMILRNIYNWRV					
	300	310	320	330	340	
	360	370	380	390	400	410
2MEPSP	KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEV					
	:. :. . . :. . . :. . . :. . . :. . . :. . . :. . . :. . . :. . .					
Dad:CA	KESDRLHAMATELRKVGAIVSEGYDYLHIVPPVQIQSAFINTYNDHRIAMCFALVALSNV					
	350	360	370	380	390	400
	420	430	440			
2MEPSP	PVTIRDPGCTRKTFPDYFDVLSTFVKN					
	:. . . :. . :. . . :. .					
Dad:CA	SVTINNPKCVCKTFPDDFDRFLSVGNNL					
	410	420	430			

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

APPENDIX F – DETAILED FASTA SEQUENCE SIMILARITY ALIGNMENTS OF THE 2mEPSPS PROTEIN
WITH PROTEINS FROM THE GENPEPT DATABASE

4774187 | Sequence 5 from Patent WO9802562. | *unidentified* | AA | 445
Length = 445
initn: 2060 initl: 2060 opt: 2060 Z-score: 4134.5 bits: 774.2 E(): 0
Smith-Waterman score: 2060; 91.685% identity (95.730% similar) in 445 aa
overlap (1-445:1-445)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEIS	GTVKLPGSKSL	SNRILLLAAL	SEGTTVVDNLL	NSEDVHYML	GAL
					
genpep	MAGAEIIVLQPIKEIS	GTVKLPGSKSL	SNRILLLAAL	SEGTTVVDNLL	NSEDVHYML	GAL
	10	20	30	40	50	60
	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRA	VVVGCGGKFP	VEDAKEEVQL	FLGNAGIAMRSL	xxxxxxxxxxxx	
					
genpep	RTLGLSVEADKAAKRA	VVVGCGGKFP	VEDAKEEVQL	FLGNAGIAMRSL	TAAVTAAGGNAT	
	70	80	90	100	110	120
	130	140	150	160	170	180
2MEPSP	YVLDGVPRMRERPIG	DLVVGLKQLG	ADVDCFLGT	DCPPVRVNGIG	GLPGGKVKLS	SGSISS
					
genpep	YVLDGVPRMRERPIG	DLVVGLKQLG	ADVDCFLGT	DCPPVRVNGIG	GLPGGKVKLS	SGSISS
	130	140	150	160	170	180
	190	200	210	220	230	240
2MEPSP	QYxxxxxxxxxxxx	GDVEIEIIDKL	ISIPYVEMTLRL	MERFGVKA	EHSDSWDRFY	IKGGQ
	:: .. . :					
genpep	QYLSALLMAAPLAL	GDVEIEIIDKL	ISIPYVEMTLRL	MERFGVKA	EHSDSWDRFY	IKGGQ
	190	200	210	220	230	240
	250	260	270	280	290	300
2MEPSP	KYKSPKNAYVEGD	ASSASYFLAGAAI	xxxxxxxxxxxx	SLQGDVKFAE	VLEMMGAKVT	W
					
genpep	KYKSPKNAYVEGD	ASSASYFLAGAAI	TGGTVTVEGCG	TTSLQGDVKFAE	VLEMMGAKVT	W
	250	260	270	280	290	300
	310	320	330	340	350	360
2MEPSP	TETSVTVTGPPREP	FGRKHLKAIDV	NMNKMPDVAM	TLAVVALFAD	GPTAIRDVAS	WRVKE
					
genpep	TETSVTVTGPPREP	FGRKHLKAIDV	NMNKMPDVAM	TLAVVALFAD	GPTAIRDVAS	WRVKE
	310	320	330	340	350	360
	370	380	390	400	410	420
2MEPSP	TERMVAIRTELTKL	GASVEEGPDYCI	IITPPEKLN	VT	AITD	YDDHRMAMAFSLAACAEVPV
					
genpep	TERMVAIRTELTKL	GASVEEGPDYCI	IITPPEKLN	VT	AITD	YDDHRMAMAFSLAACAEVPV
	370	380	390	400	410	420

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          .....
genpep  TIRDPGCTRKTFPDYFDVLSTFVKN
                430          440

```

21886544 | Sequence 4 from Patent EP1217073. | *Zea mays* | AA | 444
Length = 444
initn: 2055 initl: 2055 opt: 2055 Z-score: 4124.5 bits: 772.3 E(): 0
Smith-Waterman score: 2055; 91.667% identity (95.721% similar) in 444 aa
overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          .....
genpep  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
                10          20          30          40          50

                70          80          90         100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
          .....
genpep  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLTAAVTAAGGNAT
        60          70          80          90         100         110

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          .....
genpep  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        120         130         140         150         160         170

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          :: .. .. . : .....
genpep  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        180         190         200         210         220         230

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          .....
genpep  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
        240         250         260         270         280         290

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          .....
genpep  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
        300         310         320         330         340         350

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          .....
genpep  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
        360         370         380         390         400         410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          .....
genpep  TIRDPGCTRKTFPDYFDVLSTFVKN
          420          430          440

```

3714740 | Sequence 4 from Patent WO9704103. | *Zea mays* | AA | 444

Length = 444

initn: 2055 initl: 2055 opt: 2055 Z-score: 4124.5 bits: 772.3 E(): 0
 Smith-Waterman score: 2055; 91.667% identity (95.721% similar) in 444 aa
 overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          .....
genpep  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          10          20          30          40          50

```

```

                70          80          90         100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
          .....
genpep  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLTAAVTAAGGNAT
          60          70          80          90         100         110

```

```

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          .....
genpep  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          120         130         140         150         160         170

```

```

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          :: .. .. . : .....
genpep  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          180         190         200         210         220         230

```

```

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          .....
genpep  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
          240         250         260         270         280         290

```

```

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          .....
genpep  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          300         310         320         330         340         350

```

```

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          .....
genpep  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          360         370         380         390         400         410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          .....
genpep  TIRDPGCTRKTFPDYFDVLSTFVKN
                420          430          440

```

3714674 | Sequence 4 from Patent WO9704114. | *Zea mays* | AA | 444
Length = 444
initn: 2055 initl: 2055 opt: 2055 Z-score: 4124.5 bits: 772.3 E(): 0
Smith-Waterman score: 2055; 91.667% identity (95.721% similar) in 444 aa
overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          .....
genpep  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
                10          20          30          40          50

                70          80          90          100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
          .....
genpep  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLTAAVTAAGGNAT
                60          70          80          90          100         110

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          .....
genpep  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
                120         130         140         150         160         170

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          :: .. .. . : .....
genpep  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
                180         190         200         210         220         230

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          .....
genpep  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
                240         250         260         270         280         290

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          .....
genpep  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
                300         310         320         330         340         350

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          .....
genpep  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
                360         370         380         390         400         410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          .....
genpep  TIRDPGCTRKTFPDYFDVLSTFVKN
                420          430          440

```

4774185 | Sequence 3 from Patent WO9802562. | *unidentified* | AA | 445
Length = 445
initn: 2050 initl: 2050 opt: 2050 Z-score: 4114.4 bits: 770.5 E(): 0
Smith-Waterman score: 2050; 91.236% identity (95.281% similar) in 445 aa
overlap (1-445:1-445)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          .....
genpep  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
                10          20          30          40          50          60

                70          80          90         100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
          .....
genpep  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNAT
                70          80          90         100         110         120

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          .....
genpep  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
                130         140         150         160         170         180

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          :: .. .. . : .....
genpep  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
                190         200         210         220         230         240

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          .....
genpep  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
                250         260         270         280         290         300

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          .....
genpep  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
                310         320         330         340         350         360

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          .....
genpep  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
                370         380         390         400         410         420

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
        .....
genpep  TIRDPGCTRKTFPDYFDVLSTFVKN
                430          440

```

3714672 | Sequence 2 from Patent WO9704114. | *Zea mays* | AA | 444
Length = 444
initn: 2045 initl: 2045 opt: 2045 Z-score: 4104.4 bits: 768.6 E(): 0
Smith-Waterman score: 2045; 91.216% identity (95.270% similar) in 444 aa
overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
        .....
genpep  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
                10          20          30          40          50

                70          80          90          100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
        .....
genpep  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNAT
        60          70          80          90          100         110

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        .....
genpep  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
        120         130         140         150         160         170

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        :: .. .. . : .....
genpep  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
        180         190         200         210         220         230

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
        .....
genpep  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
        240         250         260         270         280         290

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
        .....
genpep  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
        300         310         320         330         340         350

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
        .....
genpep  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
        360         370         380         390         400         410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          .....
genpep  TIRDPGCTRKTFPDYFDVLSTFVKN
                420          430          440

```

1524383 | Z.mays mRNA for EPSP-synthase. | *Zea mays* | AA | 444

Length = 444

initn: 2045 initl: 2045 opt: 2045 Z-score: 4104.4 bits: 768.6 E(): 0
 Smith-Waterman score: 2045; 91.216% identity (95.270% similar) in 444 aa
 overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          .....
genpep  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
                10          20          30          40          50

```

```

                70          80          90         100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
          .....
genpep  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNAT
                60          70          80          90         100         110

```

```

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          .....
genpep  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
                120         130         140         150         160         170

```

```

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          .. .. . .....
genpep  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
                180         190         200         210         220         230

```

```

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          .....
genpep  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
                240         250         260         270         280         290

```

```

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          .....
genpep  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
                300         310         320         330         340         350

```

```

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          .....
genpep  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
                360         370         380         390         400         410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          .....
genpep  TIRDPGCTRKTFPDYFDVLSTFVKN
          420          430          440

```

21886542 | Sequence 2 from Patent EP1217073. | *Zea mays* | AA | 444
Length = 444
initn: 2045 initl: 2045 opt: 2045 Z-score: 4104.4 bits: 768.6 E(): 0
Smith-Waterman score: 2045; 91.216% identity (95.270% similar) in 444 aa
overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          .....
genpep  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          10          20          30          40          50

```

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
          .....
genpep  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNAT
          60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          .....
genpep  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          :: .. .. . : .....
genpep  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          .....
genpep  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
          240          250          260          270          280          290

```

```

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          .....
genpep  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          300          310          320          330          340          350

```

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          .....
genpep  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          360          370          380          390          400          410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          .....
genpep  TIRDPGCTRKTFPDYFDVLSTFVKN
                420          430          440

```

3714738 | Sequence 2 from Patent WO9704103. | *Zea mays* | AA | 444
Length = 444
initn: 2045 initl: 2045 opt: 2045 Z-score: 4104.4 bits: 768.6 E(): 0
Smith-Waterman score: 2045; 91.216% identity (95.270% similar) in 444 aa
overlap (2-445:1-444)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          .....
genpep  AGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
                10          20          30          40          50

                70          80          90          100         110         120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
          .....
genpep  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNAT
                60          70          80          90          100         110

                130         140         150         160         170         180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
          .....
genpep  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
                120         130         140         150         160         170

                190         200         210         220         230         240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
          :: .. .. . : .....
genpep  QYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
                180         190         200         210         220         230

                250         260         270         280         290         300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
          .....
genpep  KYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTW
                240         250         260         270         280         290

                310         320         330         340         350         360
2MEPSP  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
          .....
genpep  TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
                300         310         320         330         340         350

                370         380         390         400         410         420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
          .....
genpep  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPV
                360         370         380         390         400         410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
          .....
genpep  TIRDPGCTRKTFPDYFDVLSTFVKN
          420          430          440

```

16415786 | *Eleusine indica* plastid partial mRNA for 5-enolpyruvylshikimate-3-phosphate synthase (epsps-S gene). | *Eleusine indica* | AA | 445
Length = 445
initn: 1606 initl: 1606 opt: 2001 Z-score: 4015.9 bits: 752.2 E(): 4e-215
Smith-Waterman score: 2001; 89.213% identity (94.831% similar) in 445 aa
overlap (2-445:1-445)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
          .....
genpep  AGAEEVVLQPIKEISGVVGLPGSKSLSNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
          10          20          30          40          50

                70          80          90          100         110
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
          .....
genpep  KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNA
          60          70          80          90          100         110

                120         130         140         150         160         170
2MEPSP  xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
          .....
genpep  TYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSIS
          120         130         140         150         160         170

                180         190         200         210         220         230
2MEPSP  SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
          ... ..
genpep  SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
          180         190         200         210         220         230

                240         250         260         270         280         290
2MEPSP  QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
          .....
genpep  QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
          240         250         260         270         280         290

                300         310         320         330         340         350
2MEPSP  WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
          .....
genpep  WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
          300         310         320         330         340         350

                360         370         380         390         400         410
2MEPSP  ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
          .....
genpep  ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACADVP
          360         370         380         390         400         410

```

	420	430	440			
2MEPSP	VTIRDPGCTRKTFPDYFDVLSTFVKN					
					
genpep	VTIRDPGCTRKTFPDYFDVLSTFVKN					
	420	430	440			
24850309 <i>Eleusine indica</i> 5-enolpyruvylshikimate-3-phosphate synthase (epsps-S)						
gene, partial cds; nuclear gene for plastid product. <i>Eleusine indica</i> AA 445						
Length = 445						
initn: 1606 init1: 1606 opt: 2001 Z-score: 4015.9 bits: 752.2 E(): 4e-215						
Smith-Waterman score: 2001; 89.213% identity (94.831% similar) in 445 aa						
overlap (2-445:1-445)						
	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLP	SGSKSLSNRILL	LAALSEGTTVVDNLL	NS	EDVHYMLGAL	

genpep	AGAEIIVLQPIKEISGVVKLP	SGSKSLSNRILL	LSALAEGTTVVDNLL	NS	EDVHYMLGAL	
	10	20	30	40	50	
	70	80	90	100	110	
2MEPSP	RTLGLSVEADKAAKRAVVVG	CGGKFPVE-DAKEEVQL	FLGNAGIAMRSL	xxxxxxxxxxxx		

genpep	KTLGLSVEADKAAKRAVVVG	CGGKFPVEKDAKEEVQL	FLGNAGTAMRPL	TAAVTAAGGNA		
	60	70	80	90	100	110
	120	130	140	150	160	170
2MEPSP	xYVLDGVPRMRERPIGDLV	VGLKQLGADVDCFL	GTDCPPVRVNGIGGL	PGGKVKL	SGSIS	

genpep	TYVLDGVPRMRERPIGDLV	VGLKQLGADVDCFL	GTDCPPVRVKGIGGL	PGGKVKL	SGSIS	
	120	130	140	150	160	170
	180	190	200	210	220	230
2MEPSP	SQYxxxxxxxxxxxxGDVEI	EIIDKLISIPYEMTLRL	MERFGVKA	EHSDSWDRFYIKGG		
	:::
genpep	SQYLSALLMAAPLALGDVEI	EIIDKLISIPYEMTLRL	MERFGVKA	EHSDSWDRFYIKGG		
	180	190	200	210	220	230
	240	250	260	270	280	290
2MEPSP	QKYKSPKNAYVEGDASS	ASYFLAGAAI	xxxxxxxxxxxxxSLQGD	VKFAEVL	EMMGAKVT	

genpep	QKYKSPKNAYVEGDASS	ASYFLAGAAITGGTVT	VEGCGTTSLQGD	VKFAEVL	EMMGAKVT	
	240	250	260	270	280	290
	300	310	320	330	340	350
2MEPSP	WTETSVTVTGPPREPFGR	KHLKAIDVNMN	KMPDVAMTLAV	VALFADGPTAIR	DVASWRVK	

genpep	WTETSVTVTGPPREPFGR	KHLKAIDVNMN	KMPDVAMTLAV	VALFADGPTAIR	DVASWRVK	
	300	310	320	330	340	350
	360	370	380	390	400	410
2MEPSP	ETERMVAIRTELTKLGAS	VEEGPDYCIITPPEK	LNVTAIDTYDDHR	MAMAFSLAACA	EV	

genpep	ETERMVAIRTELTKLGAS	VEEGPDYCIITPPEK	LNVTAIDTYDDHR	MAMAFSLAAC	ADVP	
	360	370	380	390	400	410

```

420          430          440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      .....
genpep VTIRDPGCTRKTFPDYFDVLSTFVKN
      420          430          440

24850311|Eleusine indica 5-enolpyruvylshikimate-3-phosphate synthase (epsps-R)
gene, partial cds; nuclear gene for plastid product. |Eleusine indica|AA|445
Length = 445
initn: 1601 initl: 1601 opt: 1996 Z-score: 4005.8 bits: 750.4 E(): 1.4e-214
Smith-Waterman score: 1996; 89.213% identity (94.831% similar) in 445 aa
overlap (2-445:1-445)

      10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      .....
genpep AGAEEVVLQPIKEISGVVKLPGSKSLSNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
      10          20          30          40          50

      70          80          90          100          110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .....
genpep KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRSLTAAVTAAGGNA
      60          70          80          90          100          110

      120          130          140          150          160          170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .....
genpep TYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVKIGIGGLPGGKVKLSGSIS
      120          130          140          150          160          170

      180          190          200          210          220          230
2MEPSP SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      ::: .. .. . .....
genpep SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      180          190          200          210          220          230

      240          250          260          270          280          290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      .....
genpep QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
      240          250          260          270          280          290

      300          310          320          330          340          350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      .....
genpep WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      300          310          320          330          340          350

      360          370          380          390          400          410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
      .....
genpep ETERMVAIRTELTKLGASVEEGLDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACADVP
      360          370          380          390          400          410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYPFDVLSTFVKN
      ::::::::::::::::::::::
genpep VTIRDPGCTRKTFPDYPFDVLSTFVKN
      420      430      440

```

16415784 | *Eleusine indica* plastid partial mRNA for 5-enolpyruvylshikimate-3-phosphate synthase (epsps-R gene). | *Eleusine indica* | AA | 445
Length = 445
initn: 1601 initl: 1601 opt: 1996 Z-score: 4005.8 bits: 750.4 E(): 1.4e-214
Smith-Waterman score: 1996; 89.213% identity (94.831% similar) in 445 aa
overlap (2-445:1-445)

```

              10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      ::::.::::::::::::::::::::::::::::::::::::::::::::::::::::
genpep AGAEEVVLQPIKEISGVVLPGSKSLSNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
              10      20      30      40      50

```

```

              70      80      90      100     110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .::::::::::::::::::::::::::::::::::::::::::::::::::::.
genpep KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRSLTAAVTAAGGNA
      60      70      80      90      100     110

```

```

      120     130     140     150     160     170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .::::::::::::::::::::::::::::::::::::::::::::::::::::
genpep TYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSIS
      120     130     140     150     160     170

```

```

      180     190     200     210     220     230
2MEPSP SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      ::: .. . ::::::::::::::::::::::::::::::::::::::::::::::
genpep SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      180     190     200     210     220     230

```

```

      240     250     260     270     280     290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      ::::::::::::::::::::::::::::::. . . ::::::::::::::::::::::
genpep QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
      240     250     260     270     280     290

```

```

      300     310     320     330     340     350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
genpep WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      300     310     320     330     340     350

```

```

      360     370     380     390     400     410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVP
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
genpep ETERMVAIRTELTKLGASVEEGLDYCIITPPEKLNVTADTYDDHRMAMAFSLAACADVP
      360     370     380     390     400     410

```

```

420          430          440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      :
genpep VTIRDPGCTRKTFPDYFDVLSTFVKN
      420          430          440

40557113|Eleusine indica 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS-R)
mRNA, partial cds. |Eleusine indica|AA|445
Length = 445
initn: 1597 initl: 1597 opt: 1992 Z-score: 3997.8 bits: 748.9 E(): 4e-214
Smith-Waterman score: 1992; 88.764% identity (94.607% similar) in 445 aa
overlap (2-445:1-445)

      10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :
genpep AGAEEVVLQPIKEISGVVKLPGSKSLSNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
      10          20          30          40          50

      70          80          90          100          110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .
genpep KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRALTAAVTAAGGNA
      60          70          80          90          100          110

      120          130          140          150          160          170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .
genpep TYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSIS
      120          130          140          150          160          170

      180          190          200          210          220          230
2MEPSP SQYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      :
genpep SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      180          190          200          210          220          230

      240          250          260          270          280          290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      :
genpep QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLETMGAKVT
      240          250          260          270          280          290

      300          310          320          330          340          350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      :
genpep WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      300          310          320          330          340          350

      360          370          380          390          400          410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIAIDTYDDHRMAMAFSLAACAEVP
      :
genpep ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIAIDTYDDHRMAMAFSLAACADVP
      360          370          380          390          400          410

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      ::::::::::::::::::::::
genpep VTIRDPGCTRKTFPDYFDVLSTFVKN
      420      430      440

```

40557115|*Eleusine indica* 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS-S)
mRNA, partial cds. |*Eleusine indica*|AA|445
Length = 445
initn: 1588 initl: 1588 opt: 1983 Z-score: 3979.7 bits: 745.5 E(): 4.1e-213
Smith-Waterman score: 1983; 88.764% identity (94.382% similar) in 445 aa
overlap (2-445:1-445)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      .....
genpep AGAEEVVLQPIKEISGVVLPGSKSLSNRILLLSALAEGTTVVDNLLNSEDVHYMLGAL
      10      20      30      40      50

```

```

      70      80      90      100     110
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      .....
genpep KTLGLSVEADKAAKRAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNA
      60      70      80      90      100     110

```

```

      120     130     140     150     160     170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
      .....
genpep TYVLDGVPRMLERPIGDLVVGLKQLGADVDCFLGTDYPPVRVKGIGGLPGGKVKLSGSIS
      120     130     140     150     160     170

```

```

      180     190     200     210     220     230
2MEPSP SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      ... ..
genpep SQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      180     190     200     210     220     230

```

```

      240     250     260     270     280     290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      .....
genpep QKYKSPKNAYVEGDASSASYFLAGAAITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVT
      240     250     260     270     280     290

```

```

      300     310     320     330     340     350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVK
      .....
genpep WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVK
      300     310     320     330     340     350

```

```

      360     370     380     390     400     410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACAEVP
      .....
genpep ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTADTYDDHRMAMAFSLAACADVP
      360     370     380     390     400     410

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	400	410	420	430	440
2MEPSP	EKLNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN				
	::: . :::				
genpep	EKLNITAIDTYDDHRMAMAFSLAACADVPVTIRDPGCTRKTFFPNYFDVLSTFVRN				
	460	470	480	490	500
				510	

14549196 | *Oryza sativa* EPSPs, rps20 genes for 3-phosphoshikimate 1-carboxyvinyltransferase, 40S ribosomal protein S20, complete cds. | *Oryza sativa Japonica Group* | AA | 511
Length = 511
initn: 1950 initl: 1584 opt: 1951 Z-score: 3914.4 bits: 733.7 E(): 1.8e-209
Smith-Waterman score: 1951; 86.517% identity (94.382% similar) in 445 aa
overlap (2-445:67-511)

	10	20	30
2MEPSP	MAGAEIIVLQPIKEISGTVKLPGSKSLSNRI		
	: ::::::::::: . ::::::::::: . ::::::::::: .		
genpep	GGMRVRVRARGRREAVVVASASSSSVAAPAAKAEIIVLQPIREISGAVQLPGSKSLSNRI		
	40	50	60
			70
			80
			90

	40	50	60	70	80	90
2MEPSP	LLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCCKFPVE-DA					
	::: . ::::::::::: . ::::::::::: . ::::::::::: . ::::::::::: . ::					
genpep	LLLSALSEGTTVVDNLLNSEDVHYMLEALKALGLSVEADKVAKRAVVVGCCKFPVEKDA					
	100	110	120	130	140	150

	100	110	120	130	140	150
2MEPSP	KEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDC					
	::: . ::::::::::: . :: ::::::::::: . ::::::::::: .					
genpep	KEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQLGADVDC					
	160	170	180	190	200	210

	160	170	180	190	200	210
2MEPSP	FLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPY					
	::: . ::::::::::: . ::::::::::: ::::::::::: . ::::::::::: .					
genpep	FLGTECPPVRVKIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDKLISIPY					
	220	230	240	250	260	270

	220	230	240	250	260	270
2MEPSP	VEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxx					
	::: . ::::::::::: . ::::::::::: . ::::::::::: . ::::::::::: . .					
genpep	VEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGAAITGGT					
	280	290	300	310	320	330

	280	290	300	310	320	330
2MEPSP	xxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMK					
	. . . ::::::::::: . ::::::::::: . ::::::::::: . ::::::::::: .					
genpep	VTVQGCCTTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPREPYGKKHLKAIDVNMNMK					
	340	350	360	370	380	390

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	340	350	360	370	380	390
2MEPSP	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP					

genpep	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP					
	400	410	420	430	440	450

	400	410	420	430	440
2MEPSP	EKLNVTIAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFDYFDVLSTFVKN				

genpep	EKLNITIAIDTYDDHRMAMAFSLAACADVPVTIRDPGCTRKTFFPNYFDVLSTFVRN				
	460	470	480	490	510

55296169 | *Oryza sativa Japonica Group* genomic DNA, chromosome 6, PAC clone:P0679C08. | *Oryza sativa Japonica Group* | AA | 515
Length = 515
initn: 1950 initl: 1584 opt: 1951 Z-score: 3914.3 bits: 733.7 E(): 1.8e-209
Smith-Waterman score: 1951; 86.517% identity (94.382% similar) in 445 aa
overlap (2-445:71-515)

		10	20	30
2MEPSP		MAGAEIIVLQPIKEISGTVKLPGSKSLSNRI		
		:		
genpep	GGMRVRVRARGRREAVVVASASSSSVAAPAAKAEIIVLQPIREISGAVQLPGSKSLSNRI			
	50	60	70	100

	40	50	60	70	80	90
2MEPSP	LLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DA					

genpep	LLLSALSEGTTVVDNLLNSEDVHYMLEALKALGLSVEADKVAKRAVVVGCGGKFPVEKDA					
	110	120	130	140	150	160

	100	110	120	130	140	150
2MEPSP	KEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDC					

genpep	KEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQLGADVDC					
	170	180	190	200	210	220

	160	170	180	190	200	210
2MEPSP	FLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPY					

genpep	FLGTECPPVRVKGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDKLISIPY					
	230	240	250	260	270	280

	220	230	240	250	260	270
2MEPSP	VEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxx					

genpep	VEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGAAITGGT					
	290	300	310	320	330	340

	280	290	300	310	320	330
2MEPSP	xxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGKKHLKAIDVNMNMK					

genpep	VTVQGCGETTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPREPYGKKHLKAIDVNMNMK					
	350	360	370	380	390	400

```
113594752|Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6.
|Oryza sativa Japonica Group|AA|515
Length = 515
initn: 1950 init1: 1584 opt: 1951 Z-score: 3914.3 bits: 733.7 E(): 1.8e-209
Smith-Waterman score: 1951; 86.517% identity (94.382% similar) in 445 aa
overlap (2-445:71-515)
```

	340	350	360	370	380	390
2MEPSP	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP					
	:	:	:	:	:	:
genpep	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP					
	410	420	430	440	450	460
	400	410	420	430	440	
2MEPSP	EKLNVTIAIDTYDDHRMAMAFSLAACAEVPTIRDPGCTRKTFPDYFDVLSTFVKN					
	:	:	:	:	:	:
genpep	EKLNITAIIDTYDDHRMAMAFSLAACADVPVTIRDPGCTRKTFPNYFDVLSTFVRN					
	470	480	490	500	510	

```

Length = 515
initn: 1950 init1: 1584 opt: 1951  Z-score: 3914.3  bits: 733.7 E(): 1.8e-209
Smith-Waterman score: 1951;   86.517% identity (94.382% similar) in 445 aa
overlap (2-445:71-515)

```

```

                                10          20          30
2MEPSP      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRI
              : : : : : : : : : : : : : : : :
genpep      GGMRVRVRARGRREAVVVASASSSSVAAPAAKAEIIVLQPIREISGAVQLPGSKSLSNRI
              50          60          70          80          90         100

```

40 50 60 70 80 90
 2MEPSP LLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DA
 :::::::::::::::::::::::::::::::::::::: :::::::::::::::::::::::::::::::::::::: ::
 genpep LLLSALSEGTTVVDNLLNSEDVHYMLEALKALGLSVEADKVAKRAVVVGCGGKFPVEKDA
 110 120 130 140 150 160

100 110 120 130 140 150
 2MEPSP KEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDC
 ::::::::::: :: :... .. :::
 genpep KEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQLGADVDC
 170 180 190 200 210 220

```

                160          170          180          190          200          210
2MEPSP  FLGTD CPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxxGDVEIEIIDKLISIPY
        :::::::::::::::::::::::::::::::::::  .  .  .  ::::::::::::::::::::
genpep  FLGTECPPVRVKIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDKLISIPY
                230          240          250          260          270          280

```

2MEPSP VEMTLRLMERFGVKA²²⁰EHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAI²³⁰xxxx
 ::
 genpep VEMTLRLMERFGVKA²⁹⁰EHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGAAITGGT³⁴⁰

	280	290	300	310	320	330
2MEPSP	xxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGKRHLKAIDVNMNMK					
	.	:	:	:	:	:
genpep	VTVQGCGETTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPREPYGKKHLKAVDVNMNMK					
	350	360	370	380	390	400

	340	350	360	370	380	390
2MEPSP	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP					
	:	:	:	:	:	:
genpep	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP					
	410	420	430	440	450	460
	400	410	420	430	440	
2MEPSP	EKLNVTIAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN					
	:	:	:	:	:	:
genpep	EKLNITIAIDTYDDHRMAMAFSLAACADVPVTIRDPGCTRKTFPNYFDVLSTFVRN					
	470	480	490	500	510	

2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLLAALSEGTTVVDNLLNSEDVHYMLGAL
 ::::::::::: . . ::::::::::: . . ::::::::::: . . ::::::::::: . . ::::::::::: . . ::
 genpep QPIKEISGAVQLPGSKSLSNRILLLLSALSEGTTVVDNLLNSEDVHYMLEAL
 10 20 30 40 50

Sequence alignment

```
                70          80          90         100        110  
2MEPSP RTLGSLVEADKAAKRAVVVGC GGFPE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx  
. :.: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:  
genpep DALGSLVEADKVAKRAVVVGC GRFPIEKDAKEEVLKFLGNAGTAMRPLTAAVVAAGGNA  
              60            70           80          90         100        110
```

120 130 140 150 160 170
 2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSIS
 . :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
 genpep TYVLDGVPRMRERPTGDLVVGLKQLGANVDCFLGTDCPPVRINGIGGLPGGKVKLSGSIS
 120 130 140 150 160 170

```

      180      190      200      210      220      230
2MEPSP  SQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      ::  ..  .  .  ::::::::::::::::::::::::::::::::::::::::::::::
genpep  SQYLLSLLMAAPLALGDVEIEIIDKLISVPYVEMTLRLMERFGVTAHSDSWDRFYIKGG
      180      190      200      210      220      230

```

2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
::: : : : : : : : : : : : : : : . . . : : : : : : : : : : : : : :
genpep QKYKSPGNAYVEGDASSASYFLAGAAITGGTVTVQGCGTTSLQGDVKFAEVLEMMGAKVT
 240 250 260 270 280 290

300 310 320 330 340 350
 2MEPSP WTETSVTVTGPPREPFGGRKHLKAIDVNMNMKMPDVAMTLAVVALFADGPTAIRDVASWRVK
 :::.....
 genpep WTDTSVTVTGPPRQPFGRKHLKAIDVNMNMKMPDVAMTLAVVALFADGPTAIRDVASWRVK
 300 310 320 330 340 350

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      360      370      380      390      400      410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
      :::::::::: :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
genpep ETERMVAICTELTKLGATVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
      360      370      380      390      400      410

```

```

      420      430      440
2MEPSP VTIRDPGCTRKTFPDYFDVLSTFVKN
      :::::::::: ::::::::::
genpep VTIRDPGCTRKTFPNYFDVLSTFVKN
      420      430

```

55740769 | *Camptotheca acuminata* 5-enolpyruvylshikimate 3-phosphate synthase (EPSPS) mRNA, complete cds. | *Camptotheca acuminata* | AA | 519
Length = 519
initn: 1788 initl: 1446 opt: 1795 Z-score: 3600.5 bits: 675.6 E(): 5.4e-192
Smith-Waterman score: 1795; 79.819% identity (91.156% similar) in 441 aa
overlap (6-445:79-519)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLA
                                :::::::::: ::::::::::
genpep VAVNSPVINVRVPLRVSASVVTTEKTSMTPEIVLQPIKEISGTVKLP GSKSLSNRILLLA
      50      60      70      80      90     100

```

```

      40      50      60      70      80      90
2MEPSP ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
      :::::::::: :::::::::: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
genpep ALSEGTTVVDNLLNSDDVHYMLGALRTLGLRVEEDSAIKRAIVEGCSGLFPVGKESTDEV
      110     120     130     140     150     160

```

```

      100     110     120     130     140     150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      :::::::::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
genpep QLFLGNAGTAMRPLTAAVTAAGGNSSYILDGVPRMRERPIGDLVTGLKQLGADVDCFLGT
      170     180     190     200     210     220

```

```

      160     170     180     190     200     210
2MEPSP DCPFVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      .:::::::: : :::::::::: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
genpep NCPPVRVIGKGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISIPYVEMT
      230     240     250     260     270     280

```

```

      220     230     240     250     260     270
2MEPSP LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx
      :::::::::: :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
genpep LKLMKRFGVTVEHSDNWDRFLIQGGQKYKSPGNSYVEGDASSASYFLAGAAVTGGTITVE
      290     300     310     320     330     340

```

```

      280     290     300     310     320     330
2MEPSP xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA
      ..:::::::: :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
genpep GCGSSSLQGDVKFAEVLEKMGAKVTWTENSVTVTGPPRNSSGRKHLRAVDVNMNKMMPDVA
      350     360     370     380     390     400

```

340 350 360 370 380 390
 2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
 :::::::::::::::::::::::::::: : : : : : : : : : : : : : : : :
 genpep MTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLGLATVEEGPDYCVITPPEKLN
 410 420 430 440 450 460

400 410 420 430 440
 2MEPSP VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTTFPDYFDVLSTFVKN
 :::::::::::::::::::::::::::: : : : : : : : : : : : : : : : :
 genpep VTAVDITYDDHRMAMAFSLAACANVPVTIKDPGCTRKTTFPDYFEVLQRFAPH
 470 480 490 500 510

				10	20	30
2MEPSP				MAGAE	EIVLQPIKEIS	GTVKLPGSKSL
				:	:	:
genpep	SVSSNNNRNLGKKCLIVSA	VATTEK	PSTVPEIVLQPIKEIS	GTVNLPGSKSL	SNRILL	LLA
	50	60	70	80	90	100
	40	50	60	70	80	90
2MEPSP	ALSEG	TTVVDNLLNS	EDVHYMLGALRT	LGLSVEADKA	AKRAVVVGC	GKFPV-EDAKEEV
	:	:	:	:	:	:
genpep	ALAEG	TTIVDNLNSDDVHYMLGALRT	LGLNVEEDVAIKRAIVE	GCGGVFPV	VGKEAKDDI	
	110	120	130	140	150	160

100 110 120 130 140 150
 2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
 :::::::::: ::: :... .. .::: :::::::::::::::::::::::::::::::::::: :::
 genpep QLFLGNAGTAMRPLTAAVTAAGGNSSYILDGVPRMRERPIGDLVTGLKQLGADVDCSLGT
 170 180 190 200 210 220

```

      160      170      180      190      200      210
2MEPSP DCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
genpep NCPPVRVVGGLPGGKVKLSGSISSQYLTALLMASPLALGDVEIEIIDKLISIPYVEMT
      230      240      250      260      270      280

```

2MEPSP LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxx
 :.:
 genpep LKLMERFGVSVEHSDSWDQFFIRGGQKYKSPGNAYVEGDASSASYFLAGAAITGGTITVE

Page 450 of 526

	340	350	360	370	380	390
2MEPSP	MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN					

genpep	MTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGATVEEGPDYCVITPPEKLN					
	410	420	430	440	450	460

	400	410	420	430	440
2MEPSP	VT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTTFPDYFDVLSTFVKN				

genpep	VT AIDTYDDHRMAMAFSLAACADVPVTIKDPSCTRKTTFPDYFEVLQRFKHH				
	470	480	490	500	510

2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL¹⁰L²⁰AALSEGTTVVDNLLNS³⁰EDVHYMLGA⁴⁰
 :::::::::::::::::::::
 genpep KPSTVPEIVLQPIKEISGTVNLPGSKSLSNRILL¹⁰AALAE²⁰GTTIVD³⁰NLLNSDDVHYMLGA⁴⁰
 10 20 30 40 50 60

```

      60      70      80      90      100      110
2MEPSP  LRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      ::::::::::: : ::::: : ::::: : ::::: : ::::: : ::::: : ::::: : ::::: :
genpep  LRTLGLNVEEDGAIKRAIVEGCGGMFPVGKEAKDDIQFLFLGNAGTAMRPLTAAVTAAGGN
      70      80      90      100      110      120

```

120 130 140 150 160 170
 2MEPSP xxYVLGDVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSI
 . :
 genpep SSYVLGDVPRMRERPIGDLVTGLKQLGADVDCSLGTNCPVVRVVGGGGLPGGKVKLSGSI
 130 140 150 160 170 180

180 190 200 210 220 230
2MEPSP SSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG
::: . . . :::::::::::::::::::::::::::
genpep SSQYL TALLMASPLALGDVEIEIIDKLISIPYVEMTLKLMWFGVSV EHS DSWDRFFIR
190 200 210 220 230 240

```

      240      250      260      270      280      290
2MEPSP  GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLLEMMGAKV
      ::::::::::::::::::::::::::::::. . .:::::::::::::
genpep  GQKYKSPGNAYVEGDASSASYFLAGAAITGGTITVEGCGTSSLQGDVKFAEVLGQMGAEV
      250      260      270      280      290      300

```

300 310 320 330 340 350
 2MEPSP TWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
 ::::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
 genpep TWTENSVTVKGPPRDSSGRKHLRAVDVNMNKMMPDVAMTLAVVALYADGPTAIRDVASWRV
 310 320 330 340 350 360

	360	370	380	390	400	410	
2MEPSP	KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT	IAIDTYDDHRMAMAFSLAACAEV					
	:	:	:	:	:	:	:
genpep	KETERMIAICTELRKL	GATVEEGPDYCVITPPEKLNVT	IAIDTYDDHRMAMAFSLAACADV				
		370	380	390	400	410	420
	420	430	440				
2MEPSP	PVTIRDPGCTRKT	FPDYFDVLSTFVKN					
	:	:	:	:	:	:	:
genpep	PVTIKDPSC	TRKT	FPDYFEVLQRF	AKH			
		430	440				

		10	20	30
2MEPSP		MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLA		
	
genpep	ISVSNVGKSRQLQLQVVAAKTAEKPPAVPEIVLQPIKDISGTVKLPGSKSLSNRVLLLA			
	50	60	70	80
	90	100		
	40	50	60	70
2MEPSP	ALSEGTTVVDNLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV			

genpep	ALSEGTTVVENLLSSEDIHYMLGALRTLGLAVEEDKANQKAVVEGCVGQFPASKEGKDEV			
	110	120	130	140
	150	160		
	100	110	120	130
2MEPSP	QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPMRERPIGDLVVGLKQLGADVDCFLGT			

genpep	QLFLGNAGTAMRPLTAAVVAAGGNARYVLDGVPRLRERPIGDLVTGLKQLGADVDCFLGT			
	170	180	190	200
	210	220		
	160	170	180	190
2MEPSP	DCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT			

genpep	NCPPVRVVGKGLPGGKVKLSGSVSSQYLTALLMSAPLALGDVEIEIVDKLISVPYVEMT			
	230	240	250	260
	270	280		
	220	230	240	250
2MEPSP	LRLMERFGVKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxx			

genpep	LKLMERFGVYVEHTDNWDRFLIRGGQKYKSPGNAFVEGDASSASYFLAGAAVTGGTVTVE			
	290	300	310	320
	330	340		
	280	290	300	310
2MEPSP	xxxxxxSLQGDVKFAEVLEMMGAKVWTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA			

genpep	GCGTSSLQGDVKFAEVLEKMGAKVSWTENSVTVTGPPRVP-GKKHLRAIDVNMNKMMPDVA			
	350	360	370	380
	390	400		

	340	350	360	370	380	390
2MEPSP	MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN					
	:	:	:	:	:	:
genpep	MTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLGATVEEGPDYCIITPPEKLN					
	410	420	430	440	450	460
	400	410	420	430	440	
2MEPSP	VT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN					
	:	:	:	:	:	:
genpep	VT AIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFPNYFDVLSTYSNH					
	470	480	490	500	510	

Smith-Waterman score: 1772; 78.005% identity (90.476% similar) in 441 aa overlap (6-445:80-520)

2MEPSP MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLA
:::
genpep KHVCAVGSAKVGTFRV SASVATAEKPSTVPEI VLQPIKD ISGTIKLP GSKSLSNRILLLA

50 60 70 80 90 100

40 50 60 70 80 90
 2MEPSP ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV
 ::::::::::::::::::::
 genpep ALSEGTTVVDNLLNSDDIHYMLGALKTLGLRVEEDKAIKRAIVEGCGGLFPVGKESRDEI
 110 120 130 140 150 160

100 110 120 130 140 150
 2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
 :::::::::: : : :::::::::::::::::::: : ::::::::::::::::::::
 genpep QLFLGNAGTAMRPLTAAVTVAGNSSYVLDGVPRMRERPIGDLVDGLQQLGANVDCFLGT
 170 180 190 200 210 220

160 170 180 190 200 210
 2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
 ::::: : :::::::::::::::::::: . . . : ::::::::::::::::::::
 genpep KCPPVRVFGKGGGLPGGKVKLSGSISSQYLTALVMAAPLALGDVEIEIIDKLISVPYVEMT
 230 240 250 260 270 280

2MEPSP LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxx
 :.:
 genpep LKLMERFGISVEHNDSWDRFLIRGGQKYKSPGNAYVEGDASSASYFLAGAAVTGGTVTVE
 290 300 310 320 330 340

2MEPSP xxxxxxSLQGDVKFAEVLMMGAKVTWTETSVTVTGPPPREPFGRKHLKAIDVNMNKMMPDVA
 .:
genpep GCGTSSSLQGDVKFAEVLKMGAKVTWSETSVTVTGPPQDSSKKKHLRAIDVNMNKMMPDVA

 280 290 300 310 320 330
 350 360 370 380 390 400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          340          350          360          370          380          390
2MEPSP  MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
genpep  MTLAVVALFADGPTAIRDVASWRVKETERMIAVCTELRKLGATVEEGPDYCVITPPEKLN
          410          420          430          440          450          460

```

```

          400          410          420          430          440
2MEPSP  VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
genpep  VTDIDTYDDHRMAMAFSLAACGDVPVTINDPGCTRKTFFPDYFEVLQRFTKH
          470          480          490          500          510          520

```

63334331 | *Conyza sumatrensis* 5-enol-pyruvylshikimate-phosphate synthase (EPSPS-S1) mRNA, partial cds. | *Conyza sumatrensis* | AA | 447
Length = 447
initn: 1765 initl: 1430 opt: 1770 Z-score: 3551.3 bits: 666.3 E(): 3e-189
Smith-Waterman score: 1770; 78.685% identity (90.930% similar) in 441 aa
overlap (6-445:7-447)

```

          10          20          30          40          50
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGA
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
genpep  KPSTVPEIVLQPIKEISGAVNLPGSKSLSNRILLLAALAEAGTTIVDNLLNSDDVHYMLGA
          10          20          30          40          50          60

```

```

          60          70          80          90          100          110
2MEPSP  LRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxx
          ::::::::::: : : ::::: : : : : : : : : : : : : : : : : : : : :
genpep  LRTLGLNVEEDGAIKRAIVEGCGGMFPVGKEAKDDIQLFLGNAGTAMRPLTAGVTAAGGN
          70          80          90          100          110          120

```

```

          120          130          140          150          160          170
2MEPSP  xxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSI
          ..::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
genpep  SSVLDGVPRMRERPIGDLVTGLKQLGADVDCSLGTNCPVRVVGGLPGGRVKLSGSI
          130          140          150          160          170          180

```

```

          180          190          200          210          220          230
2MEPSP  SSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG
          :::: .. . : : : : : : : : : : : : : : : : : : : : : : : :
genpep  SSQYL TALLMASPLALGDVEIEIIDKLISIPYVEMTLKLMWFGVSVEHSDSWDRFFIRG
          190          200          210          220          230          240

```

```

          240          250          260          270          280          290
2MEPSP  GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
          :::::: : : : : : : : : : : . . . : : : : : : : : : : : :
genpep  GQKYKSPGNAYVEGDASSASYFLAGAAITGGTITVEGCGTSSLQGDVKFAEVLGQMGAEV
          250          260          270          280          290          300

```

```

          300          310          320          330          340          350
2MEPSP  TWTETSVTVTGPPREFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRV
          ::::: : : : : : : : : : : : : : : : : : : : : : : :
genpep  TWTENSVTVKGPPRDSSGRKHLRAVDVNMNKMPPDVAMTLAVVALYADGPTAIRDVASWRV
          310          320          330          340          350          360

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      360      370      380      390      400      410
2MEPSP KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEV
      .....: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep KETERMIAICTELRKL GATVEEGPDYCVITPPEKLNVT AIDTYDDHRMAMAFSLAACADV
      370      380      390      400      410      420

      420      430      440
2MEPSP PVTIRDPGCTRKTFPDYFDVLSTFVKN
      ....: : : : : : : : : : : : : : : : : : : : : : : :
genpep PVTIKDPSCTRKTFPDYFEVLQRF A KH
      430      440

```

16751569 | *Dicliptera chinensis* 5-enolpyruvylshikimate-3-phosphate synthase B mRNA, complete cds; nuclear gene for plastid product. | *Dicliptera chinensis* | AA | 516
Length = 516
initn: 1514 initl: 812 opt: 1766 Z-score: 3542.3 bits: 664.8 E(): 9.5e-189
Smith-Waterman score: 1766; 78.995% identity (91.096% similar) in 438 aa
overlap (6-442:77-513)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLA
                                .....: : : : : : : : : : : : : :
genpep ISVSNVGKSRQLQLQVAAA A KTA EKP PAVPEIVLQPIKDISGTVKLP GSKSLSNRVLLLA
      50      60      70      80      90      100

      40      50      60      70      80      90
2MEPSP ALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPV-EDAKEEV
      .....: : : : : : : : : : : : : : : : : : : : : : : :
genpep ALSEGTTVVENLLSSEDIHYMLGALRTLGLHVEEDKANQKAVVEGCVGQFPASKEGKDEI
      110      120      130      140      150      160

      100      110      120      130      140      150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      .....: : : : : : : : : : : : : : : : : : : : : : : :
genpep QLFLGKAGTAMRPLTA AVVAAGGNARYVLDGVPRMRERPIGDLVTGLKQLGADVDCFLGT
      170      180      190      200      210      220

      160      170      180      190      200      210
2MEPSP DCP PVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      .....: : : : : : : : : : : : : : : : : : : : : : : :
genpep NCP PVRVVGKGLPGGKVKLSGSVSSQYLTALLMSAPLALGDVEIEIVDKLISVPYVEMT
      230      240      250      260      270      280

      220      230      240      250      260      270
2MEPSP LRLMERFGVKA EHS DSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxx
      .....: : : : : : : : : : : : : : : : : : : : : : : :
genpep LKLMERFGVYVEHTDNWDRFLIRGGQKYKSPGNAFVEGDASSASYFLAGAAVTGGT VTVE
      290      300      310      320      330      340

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      280      290      300      310      320      330
2MEPSP xxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDV
      .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep EGCCTSSLQGDVKFAEVLEKMGAEVSWTENSVTVKGPPRGPSGRKHLRGIDVNMNKMMPDV
      350      360      370      380      390      400

```

```

      340      350      360      370      380      390
2MEPSP AMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKL
      .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep AMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLKATVEEGPDYCIITPPEKL
      410      420      430      440      450      460

```

```

      400      410      420      430      440
2MEPSP NVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTTFPDYFDVLSTFVKV
      .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep NMTEIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTTFPDYFDVLSKFSKH
      470      480      490      500      510      520

```

48526086 | *Conyza canadensis* 5-enol-pyruvylshikimate-phosphate synthase (EPSPS2)
gene, complete cds. | *Conyza canadensis* | AA | 523
Length = 523
initn: 1757 initl: 1426 opt: 1762 Z-score: 3534.1 bits: 663.3 E(): 2.7e-188
Smith-Waterman score: 1762; 77.602% identity (90.950% similar) in 442 aa
overlap (5-445:82-523)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                                .: : : : : : : : : : : : : : : : : :
genpep VSCNVKNNKNPFKVSFAFSATSTKEKPSKAPEEIVLKPIQEISGTVHLPGSKSLSNRILL
                                60      70      80      90     100     110

```

```

      40      50      60      70      80      90
2MEPSP AALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPV-EDAKEE
      .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep AALSEGTTVVDNLLNSDDVHYMLGALRALGLNVEENSAIKRAIVEGCGGVFPVGKEAKDE
      120     130     140     150     160     170

```

```

      100     110     120     130     140     150
2MEPSP VQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLG
      .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep IQFLFLGNAGTAMRPLTAAVTAAGNSSYILDGVPRMRERPIGDLVTGLKQLGANVDCSLG
      180     190     200     210     220     230

```

```

      160     170     180     190     200     210
2MEPSP TDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEM
      .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep TNCPPVRVVGSGGLPGGKVKLSGSISSQYLTSLMAAPLALGDVEIEIVDKLISVPYVEM
      240     250     260     270     280     290

```

```

      220     230     240     250     260     270
2MEPSP TLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      .: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep TLKLMERFGVSVEHSDTWDRFHVRRGGQKYKSPGNAYVEGDASSASYFLAGAAITGGTVTV
      300     310     320     330     340     350

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

300	310	320	330	340	350
2MEPSP	WTETSVTVTGPPREPFG	GRKHLKAIDVNMNKM	PDVAMTLAVVALFADG	PTAIRDVASWRVK	

genpep	WTENSVTVKGPPRNSS	GRGELRPVDVNMNKM	PDVAMTLAVVALYADG	PTAIRDVASWRVK	
	310	320	330	340	350
360	370	380	390	400	410
2MEPSP	ETERMVAIRTELTKLG	ASVEEGPDYCIITPPE	KLNVTAIDTYDDHRM	AMAFSLAACAEVP	

genpep	ETERMIAICTELRKL	GATVEEGPDYCVITP	PEKLNVTADTYDDHR	MAMAFSLAACAEVP	
	370	380	390	400	410
420	430	440			
2MEPSP	VTIRDPGCTRKTFPD	YFDVLSTFVKN			
			
genpep	VTIKDPGCTRKTFPD	YFEVLERYTKH			
	430	440			

157142998 | *Gossypium hirsutum* 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) mRNA, complete cds. | *Gossypium hirsutum* | AA | 521
Length = 521
initn: 1740 initl: 1436 opt: 1752 Z-score: 3514.0 bits: 659.6 E(): 3.5e-187
Smith-Waterman score: 1752; 77.652% identity (90.519% similar) in 443 aa
overlap (4-445:79-521)

			10	20	30
2MEPSP			MAGAE	EIVLQPIKE	ISGTVKLPGSKSL
		
genpep	VVKNN	GKFGSIKARSLKVS	ASTATAEKPSRASE	IVLQPINEISGTVKL	PGSKSLSNRILL
	50	60	70	80	90
40	50	60	70	80	90
2MEPSP	LAALSE	GTTVVDNLLN	SEDVHYMLGALRT	LGLSVEADKAAKRA	VVVGCGKFPVEDAK-E

genpep	LAALSE	GTTVVENLLN	SDDVHMLVALGKL	GLYVKHDSEKKQA	IVEGCGGQFPVGKGEQ
	110	120	130	140	150
100	110	120	130	140	150
2MEPSP	EVQLFL	GNAGIAMRSL	XXXXXXXXXXXXX	YVL	DGVPRMRERPIGDL

genpep	EIELFL	GNAGTAMRPL	TAAITAAGGNSS	YVL	DGVPRMRERPIGDL
	170	180	190	200	210
160	170	180	190	200	210
2MEPSP	GTDCPP	VRVNGIGGLPGG	KVKLSGSISSQY	XXXXXXXXXXXXX	GDVEIEIIDKLISIPYVE

genpep	GTNCP	PVRIEGKGLPGG	KVKLSGSISSQY	L	TALLMAAPLALGDVEIEIIDKLISIPYVE
	230	240	250	260	270
220	230	240	250	260	270
2MEPSP	MTLRL	MERFGVKA	EHSDSWDRFYIK	GGQYKYS	PKNAYVEGDASSASYFLAGAAI

genpep	MTMKL	MERFGVTVEHT	DSWDRFFIRGG	QKYMSPGNAY	VEGDASSASYFLAGAAVTGGT
	290	300	310	320	330
					340


```
6732247|Sequence 1 from Patent WO9854330. |Petunia x hybrida|AA|516
Length = 516
initn: 1741 initl: 1406 opt: 1747 Z-score: 3504.0 bits: 657.8 E(): 1.3e-186
Smith-Waterman score: 1747; 77.778% identity (90.023% similar) in 441 aa
overlap (6-445:76-516)
```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	280	290	300	310	320	330
2MEPSP	xxxxxSLQGDVKFAEVL	EMMGAKVTWTETSV	TVTGPPREPFG	RKHLKAIDV	NMNKMPDVA	

genpep	GCGTNSLQGDVKFAEVL	EKMGAEVTWTENS	VTVKGPPRSS	SGRKHLRAIDV	NMNKMPDVA	
	350	360	370	380	390	400
	340	350	360	370	380	390
2MEPSP	MTLAVVALFADGPTAIR	DVASWRVKETERM	VAIRTELTKLGAS	VEEGPDYCIITP	PPEKLN	

genpep	MTLAVVALYADGPTAIR	DVASWRVKETERM	IAICTELRKLGA	TVEEGPDYCIITP	PPEKLN	
	410	420	430	440	450	460
	400	410	420	430	440	
2MEPSP	VTADIDYDDHRMAMAF	SLAACAEVPVTIR	DPGCTRKTFPDY	FDVLSTFVKN		
	::
genpep	VTDIDYDDHRMAMAF	SLAACADVPVTIN	DPGCTRKTFPNY	FDVLQQYSKH		
	470	480	490	500	510	

169191 | *P.hybrida* 5-enolpyruvylshikimate 3-phosphate synthase mRNA, complete cds.
| *Petunia x hybrida* | AA | 516
Length = 516
initn: 1741 initl: 1406 opt: 1747 Z-score: 3504.0 bits: 657.8 E(): 1.3e-186
Smith-Waterman score: 1747; 77.778% identity (90.023% similar) in 441 aa
overlap (6-445:76-516)

		10	20	30
2MEPSP		MAGAE	EIVLQPIKEIS	GTVKLPGSKSLSNRILL
	
genpep	MLVLKKDSIFMQKFC	SFRISASVATAQ	KPSEIVLQPIKEIS	GTVKLPGSKSLSNRILL
	50	60	70	80
	40	50	60	70
2MEPSP	ALSEGTTVVDNLLNS	EDVHYMLGALRT	LGLSVEADKAAKRA	VVVGCGKFPV-EDAKEEV

genpep	ALSEGTTVVDNLLSS	DDIHMYMLGALK	TLGLHVEEDSANQRA	VVEGCGGLFPVGKESKEEI
	110	120	130	140
	100	110	120	130
2MEPSP	QLFLGNAGIAMRSL	xxxxxxxxxxxxxY	VL	DGVPRMRERPIGDLVVGLKQLGADVDCFLGT

genpep	QLFLGNAGTAMRPL	TAAVTVAGGNSR	YVL	DGVPRMRERPI
	170	180	190	200
	160	170	180	190
2MEPSP	DCPPVRVNGIGGLPG	GKVKLSGSISSQY	xxxxxxxxxxxxxG	DVEIEIIDKLISIPYVEMT

genpep	KCPPVRIVSKGGLPG	GKVKLSGSISSQY	L	TALLMAAPLALGDVEIEIIDKLISVPYVEMT
	230	240	250	260
	220	230	240	250
2MEPSP	LRLMERFGVKA	EHSDSWDRFYIK	GGQYKSPKNAY	VEGDASSASYFLAGAAIxxxxxxxx

genpep	LKL	MERFGISVEHSS	SWDRFFV	RGQYKSPGKAFVEGDASSASYFLAGAAVTGGTITVE
	290	300	310	320
	280	290	300	310

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	280	290	300	310	320	330
2MEPSP	xxxxxSLQGDVKFAEVL	EMMGAKVTWTETSV	TVTGPPREPFG	RKHLKAIDV	NMNKMPDVA	

genpep	GCGTNSLQGDVKFAEVL	EKMGAEVTWTENSV	TVKGPPRSSSG	RKHLRAIDV	NMNKMPDVA	
	350	360	370	380	390	400
	340	350	360	370	380	390
2MEPSP	MTLAVVALFADGPTAIR	DVASWRVKETERMVA	IRTELTKLGASVEEG	PDYCIITPPEKLN		

genpep	MTLAVVALYADGPTAIR	DVASWRVKETERMIA	ICTELRKL	GATVEEGPDYCIIT	PPEKLN	
	410	420	430	440	450	460
	400	410	420	430	440	
2MEPSP	VT	AIDTYDDHRMAMAFSL	AACAEVPVTIRDPG	CTRKTFPDYFDVL	STFVKN	
	::
genpep	VT	DIDTYDDHRMAMAFSL	AACADVPVTINDPG	CTRKTFPNYFDVL	QQYSKH	
	470	480	490	500	510	

123965218 | *Allium macrostemon* EPSP synthase mRNA, complete cds. | *Allium macrostemon* | AA | 522
Length = 522
initn: 1725 initl: 1417 opt: 1727 Z-score: 3463.7 bits: 650.3 E(): 2.2e-184
Smith-Waterman score: 1727; 76.871% identity (90.930% similar) in 441 aa
overlap (6-445:82-522)

		10	20	30
2MEPSP		MAGAE	EIVLQPIKEIS	GTVKLPGSKSLSNRILL
	
genpep	RIGNCRRASVRVLASLA	AVENAAEKVAVMPEIT	LQPIKEITGT	VNLPGSKSLSNRILL
	60	70	80	90 100 110
	40	50	60	70 80 90
2MEPSP	ALSE	GTTVVDNLLNSEDV	HYMLGALRTLGLS	VEADKAAKRAVVVGCGKFPV-EDAKEEV
	::
genpep	ALAEG	TTIVDNLNSDDVSY	MLAALKTLGLS	VEDDRVNKRATVVGSGGLFPVGKESQKEV
	120	130	140	150 160 170
	100	110	120	130 140 150
2MEPSP	QLFLGNAGIAMRSL	xxxxxxxxxxxxxY	VL	DGVPRMRERPIGDLVVGLKQLGADVDCFLGT

genpep	QLFLGNAGTAMRPL	TAAVTAAGGNASY	ILDGVPRMRERPIG	DLVVGLKQLGADVDCTLTGT
	180	190	200	210 220 230
	160	170	180	190 200 210
2MEPSP	DCPPVRVNGIGGLPGG	KVKLSGSISSQY	xxxxxxxxxxxxxG	DVEIEIIDKLISIPYVEMT

genpep	DCPPVRVDANGGLPGG	KVKLSDSISSQYL	TALLMAAPLALG	DVEIEIIDKLISIPYVEMT
	240	250	260	270 280 290
	220	230	240	250 260 270
2MEPSP	LRLMERFGVKA	EHSDSWDRFYIKGG	QKYKSPKNAYVEGD	ASSASYFLAGAAIxxxxxxxx

genpep	LKLMERFGVHVDH	SSTWDRFFIKGG	QKYKSPGNAYVEGD	ASSASYFLAGAAVTGGTVTVE
	300	310	320	330 340 350

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      280      290      300      310      320      330
2MEPSP xxxxxSLQGDVKFAEVLMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVA
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep GCGTSSLQGDVKFAEVLNMGAKVTWTENSVTVTGPPQDPQKKKRLKAVDVNMNKMMPDVA
      360      370      380      390      400      410

      340      350      360      370      380      390
2MEPSP MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep MTLAVVALYADGPTTIRDVASWRVKETERMIAICTELRKLGA TVVEGPDYCIITPPEKLN
      420      430      440      450      460      470

      400      410      420      430      440
2MEPSP VTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep VTEIDTYDDHRMAMAFSLAACSDVPVTIKDPGCTRKTFPDYFEVLERYAKH
      480      490      500      510      520

```

110742505 | *Arabidopsis thaliana* mRNA for 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase, complete cds, clone: RAFL09-78-J01. | *Arabidopsis thaliana* | AA | 520
Length = 520
initn: 1688 initl: 1382 opt: 1716 Z-score: 3441.6 bits: 646.2 E(): 3.8e-183
Smith-Waterman score: 1716; 75.169% identity (90.971% similar) in 443 aa
overlap (4-445:78-520)

```

                                10      20      30
2MEPSP                                MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
                                . : : : : : : : : : : : : : : : : : : : : : :
genpep SWGLKKSGMTLIGSELRLPKVMSSVSTA EKASEIVLQPIREISGLIKLPGSKSLSNRILL
      50      60      70      80      90     100

      40      50      60      70      80      90
2MEPSP LAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGC GGFVE-DAKE
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep LAALSEGTTVVDNLLNSDDINYMLDALKRLGLNVETDSENNRAVVEGCGGIFPASIDSKS
      110     120     130     140     150     160

      100     110     120     130     140     150
2MEPSP EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL
      . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVVGLKQLGADV ECTL
      170     180     190     200     210     220

      160     170     180     190     200     210
2MEPSP GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVE
      : : : : : : : : : : : : : : : : . . . : : : : : : : : : : : : : : : :
genpep GTNCPVVRVNANGGLPGGKVKLSGSISSQYLTALLMSAPLALGDVEIEIVDKLISVPYVE
      230     240     250     260     270     280

```

	220	230	240	250	260	270
2MEPSP	MTLRLMERFGVKA	EHSDSWDRFYIKGGQKYK	SPKNAYVEGDASSASYFLAGAAI	xxxxxxx		
genpep	MTLKLME	RFVGSVEHSDSWDRFFVKG	GQKYKSPGNAYVEGDASSASYFLAGAAITGETVT			
	290	300	310	320	330	340
	280	290	300	310	320	330
2MEPSP	xxxxxxxSLQGDVKFAEVL	EMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKM	PD			
genpep	VEGCGTTS	LQGDVKFAEVL	EKMCKVSWTENS	VTVTGPPRDAFGMRHLRAIDVNMNKM	PD	
	350	360	370	380	390	400
	340	350	360	370	380	390
2MEPSP	VAMTLAVVALFADGPTAIR	DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK				
genpep	VAMTLAVVALFADGPTTIR	DVASWRVKETERMIAICTELRKL	GATVEEGSDYCVITPPKK			
	410	420	430	440	450	460
	400	410	420	430	440	
2MEPSP	LNVTAIDTYDDHRMAMAFSLAACAEV	PVTIRDPGCTRKTFPDYFDVLSTFVK	N			
genpep	VKTAEIDTYDDHRMAMAFSLAACADVPITIN	DPGCTRKTFPDYFQVLERITKH				
	470	480	490	500	510	520
2583124 <i>Arabidopsis thaliana</i> chromosome 2 clone F4L23 map CIC021 sequence. <i>Arabidopsis thaliana</i> AA 520 Length = 520 initn: 1688 initl: 1382 opt: 1716 Z-score: 3441.6 bits: 646.2 E() Smith-Waterman score: 1716; 75.169% identity (90.971% similar overlap (4-445:78-520)						
			10	20	30	
2MEPSP			MAGAE	EIVLQPIKEIS	GTVKLPGSKSLSNRILL	
genpep	SWGLKKSGMTLIGSEL	RPLKVMSSVSTAEKASEIVLQPIREISGLIKLPGSKSLSNRILL				
	50	60	70	80	90	100
	40	50	60	70	80	90
2MEPSP	LAALSEGTTVVDNLLN	SEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKE				
genpep	LAALSEGTTVVDNLLN	SDDINYMLDALKRGLGNVETDSENNRAVVEGCGGIFPASIDS	KS			
	110	120	130	140	150	160
	100	110	120	130	140	150
2MEPSP	EVQLF	LGNAGIAMRSLxxxxxxxxxxxxxYVLDGVP	MRERP	IGDLVVGLKQLGADV	DCFL	
genpep	DIELYLGNAGTAMRPL	TAAVTAAGGNASYVLDGVP	MRERP	IGDLVVGLKQLGADV	ECTL	
	170	180	190	200	210	220
	160	170	180	190	200	210
2MEPSP	GTDCPPVRVNGIGGLPGGKVK	LSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYE				
genpep	GTNCP	PPVRVNANGGLPGGKVKLSGSISSQYLTALLMSAPLALGDVEIEIVDKLISVPYE				
	230	240	250	260	270	280

```

      220      230      240      250      260      270
2MEPSP MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep MTLKLMEFVGVSVEHSDSWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
      290      300      310      320      330      340

      280      290      300      310      320      330
2MEPSP xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPPREPFGRKHLKAIDVNMNMKMPD
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPPRDAFGMRHLRAIDVNMNMKMPD
      350      360      370      380      390      400

      340      350      360      370      380      390
2MEPSP VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEEGSDYCVITPPKK
      410      420      430      440      450      460

      400      410      420      430      440
2MEPSP LNVTAIDTYDDHRMAMAFSLAACAEVPTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep VKTAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH
      470      480      490      500      510      520

63003818 | Arabidopsis thaliana At2g45300 gene, complete cds.
thaliana | AA|520
Length = 520
initn: 1688 initl: 1382 opt: 1716 Z-score: 3441.6 bits: 646.2 E()
Smith-Waterman score: 1716; 75.169% identity (90.971% similar
overlap (4-445:78-520)

      10      20      30
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep SWGLKKSGMTLIGSELRPLKVMSSVSTAEKASEIVLQPIREISGLIKLPGSKSLSNRILL
      50      60      70      80      90      100

      40      50      60      70      80      90
2MEPSP LAALSEGTTVVDNLLNSDEVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKE
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep LAALSEGTTVVDNLLNSDDINYMLDALKRLGLNVETDSENNRAVVEGCGGIFPASIDSKS
      110      120      130      140      150      160

      100      110      120      130      140      150
2MEPSP EVQLFGLGNAGIAMRSLxxxxxxxxxxxxxxxxxYVLGDGVPRMRERPIGDLVVGLKQLGADVDCFL
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep DIELYLGNAGTAMRPLTAAVTAAGGNASYVLGDGVPRMRERPIGDLVVGLKQLGADVECTL
      170      180      190      200      210      220

      160      170      180      190      200      210
2MEPSP GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxxxxxGDVEIEIIDKLISIPYE
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep GTNCPVVRVNANGGLPGGKVKLSGSISSQYLTALLMSAPLALGDVEIEIVDKLISVPYE
      230      240      250      260      270      280

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	220	230	240	250	260	270
2MEPSP	MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxx					
	::: :: . . .					
genpep	MTLKLMERFGVSVEHSDSWDRFFVKGGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT					
	290	300	310	320	330	340

	280	290	300	310	320	330
2MEPSP	xxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGKRHLKAIDVNMNKMPD					
	.. :::::::::::::::::::::: :::::::::::::::::::::::::::::: :: ::::::::::::::::::::::					
genpep	VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPPRDAFGMRHLRAIDVNMNKMPD					
	350	360	370	380	390	400

	340	350	360	370	380	390
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK					
	:: :: :::::::::::::: ::::::::::::::					
genpep	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAATVEEGSDYCVITPPKK					
	410	420	430	440	450	460

	400	410	420	430	440
2MEPSP	LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN				
 ::				
genpep	VKTAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH				
	470	480	490	500	510

3410961|*Oryza sativa Japonica* Group mRNA for 3-phosphoshikimate 1-carboxyvinyltransferase, complete cds. |*Oryza sativa Japonica* Group|AA|391
Length = 391
initn: 1713 initl: 1584 opt: 1714 Z-score: 3439.5 bits: 645.4 E(): 5e-183
Smith-Waterman score: 1714; 85.934% identity (93.862% similar) in 391 aa
overlap (56-445:1-391)

	30	40	50	60	70	80
2MEPSP	SLSNRILLLAALSEGTTVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVVGCGGKF					
	:: ::					
genpep	MLEALKALGLSVEADKVAKRAVVVVGCGGKF					
		10	20	30		

	90	100	110	120	130	140
2MEPSP	PVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQL					
	::: :::::::::::::::::::::: :: ::... .. ::::::::::::::::::::::::::::::::::::::					
genpep	PVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQL					
	40	50	60	70	80	90

	150	160	170	180	190	200
2MEPSP	GADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDK					
	:: ::::::::::::::					
genpep	GADVDCFLGTGECPPVRVKIGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDK					
	100	110	120	130	140	150

	210	220	230	240	250	260
2MEPSP	LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGA					
	::					
genpep	LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGA					
	160	170	180	190	200	210

18251236|*Orychophragmus violaceus* 5-enolpyruvylshikimate-3-phosphate synthase mRNA, complete cds. |*Orychophragmus violaceus*|AA|518
Length = 518
initn: 1683 initl: 1375 opt: 1711 Z-score: 3431.6 bits: 644.4 E(): 1.4e-182
Smith-Waterman score: 1711; 75.395% identity (90.293% similar) in 443 aa
overlap (4-445:76-518)

400 410 420 430 440
 2MEPSP LNVTAIDTYDDHRMAMAFSLAACAEVPTIRDPGCTRKTFPDYFDVLSTFVKN
 . . :
 genpep VKPAEIDTYDDHRMAMAFSLAACADVPTIKDPGCTRKTFPDYQVLESITKH
 470 480 490 500 510

160 170 180 190 200 210
 2MEPSP DCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
 .:.:. . :.:
 genpep NCPPVRIVSKGGGLPGGKVKLSGSISSQYLTAALLMAAPLALGDVEIEIIDKLISVPYVEMT
 230 240 250 260 270 280

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	220	230	240	250	260	270
2MEPSP	LRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxx					
	.: . .					
genpep	LKLMERFGVSVEHTSSWDKFLVRGGQKYKSPGKAYVEGDASSASYFLAGAAVTGGTVTVE					
	290	300	310	320	330	340

	280	290	300	310	320	330
2MEPSP	xxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVA					
	.: . .					
genpep	GCGTSSLQGDVKFAEVLEKMGAEVTWTENSVTVKGPPRNSSGMKHLRAVDVNMNKMPPDVA					
	350	360	370	380	390	400

	340	350	360	370	380	390
2MEPSP	MTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN					
	.: . .					
genpep	MTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLGA TVVEGSDYCIITPPEKLN					
	410	420	430	440	450	460

	400	410	420	430	440
2MEPSP	VTAIDTYDDHRMAMAFSLAACAEVPTIRDPGCTRKTFFPDYFDVLSTFVKN				
	.: . .				
genpep	VTEIDTYDDHRMAMAFSLAACADVPVTIKDPGCTRKTFFPNYFDVLQQYSKH				
	470	480	490	500	510

48526088 | *Conyza canadensis* 5-enol-pyruvylshikimate-phosphate synthase (EPSPS3) mRNA, partial cds. | *Conyza canadensis* | AA | 454
Length = 454
initn: 1699 initl: 1370 opt: 1704 Z-score: 3418.4 bits: 641.7 E(): 7.5e-182
Smith-Waterman score: 1704; 75.339% identity (90.045% similar) in 442 aa
overlap (5-445:13-454)

	10	20	30	40	50
2MEPSP	MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSD				
	.: . .				
genpep	ANTTKEKPSKAPEEIVLKPIQEISGTVHLPKSGSKSLSNRILLLAALSEGTTVVDNLLNSDD				
	10	20	30	40	50

	60	70	80	90	100	110
2MEPSP	VHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEVQLFLGNAGIAMRSLxxx					
	.: : : : : : : : : : . : : : : : : : : : : : : : : : : : : : . .					
genpep	VHYMLGALRALGLNIEENAAIKRAIVEGCGGLFPVGKEAKDEIQLFLGNAGTAMRTLTA					
	70	80	90	100	110	120

	120	130	140	150	160	170
2MEPSP	xxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGK					
 : . .					
genpep	VTAAGGNLSYVLDGVPRMRERPIGDLVTGLKQLGVNVDCSLGTNCPVHVVGSGGLPGGK					
	130	140	150	160	170	180

	180	190	200	210	220	230
2MEPSP	VKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSW					
	.: : : : : : : : . . . : : : : : : : : : : : : : : : : : : : . .					
genpep	VKLSGSISSVYLTSLLMAAPLALGDVEIEIIDKLISVPYVRMTLKLMMQRFGVSVESDTL					
	190	200	210	220	230	240

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	240	250	260	270	280	290
2MEPSP	DRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVL					

genpep	DRFHVRGGQKYKSPGNAYVESDASSASYFLAGAAITGGTVTVEGCGTSSLQGDVKFAEVL					
	250	260	270	280	290	300
	300	310	320	330	340	350
2MEPSP	EMMGAKVTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIR					

genpep	GQMGAEVTWTENSVTVKGPPRNSSGRGHLRPVDVNMNKMPPDVAMTLAVVALYADGPTAIR					
	310	320	330	340	350	360
	360	370	380	390	400	410
2MEPSP	DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFS					

genpep	DVASWRVKETERMIAICTELRKL GATVEEGPDYCVTTPPEKLNVT AIDTYDDHRMAMTFS					
	370	380	390	400	410	420
	420	430	440			
2MEPSP	LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN					

genpep	LAACADVPVTIKDPGCTRKSF PDYFEVLERYTKH					
	430	440	450			

295790 | *Arabidopsis thaliana* gene for 5-enolpyruvylshikimate-3-phosphate synthase (EPSP). | *Arabidopsis thaliana* | AA | 520
Length = 520
initn: 1675 initl: 1369 opt: 1703 Z-score: 3415.5 bits: 641.4 E(): 1.1e-181
Smith-Waterman score: 1703; 74.718% identity (90.519% similar) in 443 aa
overlap (4-445:78-520)

			10	20	30
2MEPSP			MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILL		
		
genpep	SWGLKKSGMTLIGSELRLKVMSSVSTAEKASEIIVLQPIREISGLIKLPGSKSLSNRILL				
	50	60	70	80	90
	40	50	60	70	80
2MEPSP	LAALSEGTTVVDNLLNSD VHYMLGALRTLGLSVEADKAAKRAVVVGC				

genpep	LAALSEGTTVVDNLLNSDDIN YMLDALKRLGLNVETDSENNRAVVEGCGIFPASIDSKS				
	110	120	130	140	150
	100	110	120	130	140
2MEPSP	EVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFL				

genpep	DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPRMRERPIGDLVVGLKQLGADVECTL				
	170	180	190	200	210
	160	170	180	190	200
2MEPSP	GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYE				

genpep	GTNCPVVRVANGGLPGGKVKLSGSISSQYLTALLMSAPLALGDVEIEIVDKLISVPYE				
	230	240	250	260	270

	240	250	260	270	280	290
2MEPSP	QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT					

genpep	QKYKSPGNAYVEGDASSASYFLAGAAITGGTVTVEGCGTSSLQGDVKFAEVLGQMGA EVT					
		250	260	270	280	290
						300
	300	310	320	330	340	350
2MEPSP	WTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVK					

genpep	WTENSVTVKGPPRNSSGRGHLRPVDVNMNKMMPDVAMTLAVVALYADGRTAIRDVASWRVK					
		310	320	330	340	350
						360
	360	370	380	390	400	410
2MEPSP	ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP					

genpep	ETERMIAICTELRKLGATVEEGPDYCVITPPEKLNVT AIDTYDDHRMAMTFSLAACADVP					
		370	380	390	400	410
						420
	420	430	440			
2MEPSP	VTIRDPGCTRKTFPDYFDVLSTFVKN					

genpep	VTIKDPGCTRKSFDPDYFEVLERYTKH					
		430	440			

46095337|*Brassica rapa* 5-enolpyruvylshikimate-3-phosphate synthase (EPSP) mRNA.
complete cds. |*Brassica rapa*|AA|514
Length = 514
initn: 1675 initl: 1375 opt: 1701 Z-score: 3411.6 bits: 640.6 E(): 1.8e-181
Smith-Waterman score: 1701; 74.944% identity (90.068% similar) in 443 aa
overlap (4-445:72-514)

Bayer CropScience


```

220      230      240      250      260      270
2MEPSP  MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      :::::::::::::::::::::
genpep  MTLKLMEFVGSAEHSESWDRFFVKGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
      290      300      310      320      330      340

280      290      300      310      320      330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMKMPD
      ..:::::::::::::::::::
genpep  VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAIDVNMNMKMPD
      350      360      370      380      390      400

340      350      360      370      380      390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      :::::::::::::::::::::
genpep  VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAIVVEEGSDYCVITPPKK
      410      420      430      440      450      460

400      410      420      430      440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      ..:
genpep  VKPAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH
      470      480      490      500      510      520

12321113|Arabidopsis thaliana chromosome 1 BAC T24P22 genomic sequence.
|Arabidopsis thaliana|AA|521
Length = 521
initn: 1672 initl: 1372 opt: 1700 Z-score: 3409.4 bits: 640.3 E()
Smith-Waterman score: 1700; 74.718% identity (90.293% similar
overlap (4-445:79-521)

10      20      30
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
      :::::::::::::::::::::
genpep  SWGLKKSDLMLNGSEIRPVKVRASVSTAEKASEIVLQPIREISGLIKLPGSKSLSNRILL
      50      60      70      80      90      100

40      50      60      70      80      90
2MEPSP  LAALSEGTTVVDNLLNSDEVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKE
      :::::::::::::::::::::
genpep  LAALSEGTTVVDNLLNSDDINYMLDALKILGLNVETHSENNRAVVEGCGGVFPASIDSKS
      110      120      130      140      150      160

100      110      120      130      140      150
2MEPSP  EVQLFGLGNAGIAMRSLxxxxxxxxxxxxxYVLGDGVPRMRERPIGDLVVGLKQLGADVDCFL
      ...:
genpep  DIELYLGNAGTAMRPLTAAVTAAGGNASYVLGDGVPRMRERPIGDLVVGLKQLGADVECTL
      170      180      190      200      210      220

160      170      180      190      200      210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYE
      :::::::::::::::::::::
genpep  GTNCPVVRVNANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIVDKLISVPYE
      230      240      250      260      270      280

```

	220	230	240	250	260	270
2MEPSP	MTLRLMERFGVKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx					

genpep	MTLKLMEFVGSAEHSESWDRFFVKGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT					
	290	300	310	320	330	340
	280	290	300	310	320	330
2MEPSP	xxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMKMPD					

genpep	VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAIDVNMNMKMPD					
	350	360	370	380	390	400
	340	350	360	370	380	390
2MEPSP	VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK					

genpep	VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGAATVEEGSDYCVITPPKK					
	410	420	430	440	450	460
	400	410	420	430	440	
2MEPSP	LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN					

genpep	VKPAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH					
	470	480	490	500	510	520
14532882 <i>Arabidopsis thaliana</i> putative 5-enolpyruvylshikimat synthase (Atlg48860) mRNA, complete cds. <i>Arabidopsis thaliana</i> AA 5						
Length = 521						
initn: 1672 initl: 1372 opt: 1700 Z-score: 3409.4 bits: 640.3 E()						
Smith-Waterman score: 1700; 74.718% identity (90.293% similar overlap (4-445:79-521))						
			10	20	30	
2MEPSP			MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILL			
			
genpep	SWGLKKSDLMLNGSEIRPVKVRASVSTAEKASEIIVLQPIREISGLIKLPGSKSLSNRILL					
	50	60	70	80	90	100
	40	50	60	70	80	90
2MEPSP	LAALSEGTTVVDNLLNSDEVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKE					

genpep	LAALSEGTTVVDNLLNSDDINYMLDALKILGLNVETHSENNRAVVEGCGGVFPASIDSKS					
	110	120	130	140	150	160
	100	110	120	130	140	150
2MEPSP	EVQLFGLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPVPRMRERPIGDLVVGLKQLGADVDCFL					

genpep	DIELYLGNAGTAMRPLTAAVTAAGGNASYVLVGVPVPRMRERPIGDLVVGLKQLGADVECTL					
	170	180	190	200	210	220
	160	170	180	190	200	210
2MEPSP	GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYE					

genpep	GTNCPVVRVNANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIVDKLISVPYE					
	230	240	250	260	270	280

	220	230	240	250	260	270
2MEPSP	MTLRLMERFGVKA	EHSDSWDRFYIKGGQKYK	SPKNAYVEGDASSASYFLAGAAI	xxxxxx		

genpep	MTLKLME	RGVSAEHSESWDRFFVKG	GQKYKSPGNAYVEGDASSASYFLAGAAITGETVT			
	290	300	310	320	330	340
	280	290	300	310	320	330
2MEPSP	xxxxxxxSLQGDVKFAEVL	EMMGAKVTWTETSVTVTGPP	PREPFGRKHLKAIDVNMN	KMPD		

genpep	VEGCGTTS	LQGDVKFAEVL	EKMCKVSWTENS	TVTGPSRDAFGMRHLRAIDVNMN	KMPD	
	350	360	370	380	390	400
	340	350	360	370	380	390
2MEPSP	VAMTLAVVALFADGPTA	IRDVASWRVKETERMVAIRTE	LT	KL	GSVEEGPDYCIITPPEK	

genpep	VAMTLAVVALFADGPTT	IRDVASWRVKETERMIAICTEL	RKL	GATVEEGSDYCVITPPK		
	410	420	430	440	450	460
	400	410	420	430	440	
2MEPSP	LNVTAIDTYDDHRMAMA	FLSAAEVPVTIRDPGCTR	KTFPDYFDVLSTFVK	N		

genpep	VKPAEIDTYDDHRMAMA	FLSAA	CDVPITINDPGCTR	KTFPDYFQVLERITKH		
	470	480	490	500	510	520
13430624 <i>Arabidopsis thaliana</i> putative 5-enolpyruvylshikimate-3-phosphatase (Atlg48860) mRNA, complete cds. <i>Arabidopsis thaliana</i> AA 5521 Length = 521 initn: 1672 initl: 1372 opt: 1700 Z-score: 3409.4 bits: 640.3 E() Smith-Waterman score: 1700; 74.718% identity (90.293% similar overlap (4-445:79-521))						
			10	20	30	
2MEPSP			MAGAE	EIVLQPIKEIS	GT	VKLPGSKSL
		
genpep	SWGLKKSD	LM	LN	GS	EIRPVKVRASV	STA
	50	60	70	80	90	100
	40	50	60	70	80	90
2MEPSP	LAALSEGTTVVDN	LLNS	EDVHYMLGAL	RTLGLSVEADKAAKRA	VVVGCGGKFPVE-DAKE	

genpep	LAALSEGTTVVDN	LLNSDD	IN	YMLDALKILGLNV	ETHSENNRAV	VEGCGGVFPASIDSKS
	110	120	130	140	150	160
	100	110	120	130	140	150
2MEPSP	EVQLF	LGNAGIAMRSL	xxxxxxxxxxxxxxxxxY	VL	DGVPRMRERPIGDLV	VGLKQLGADVDCFL

genpep	DIELYLGNAGTAMRPL	TA	AVTAAGGNASY	VL	DGVPRMRERPIGDLV	VGLKQLGADV
	170	180	190	200	210	220
	160	170	180	190	200	210
2MEPSP	GTDCPPVRVNGIGGL	PGGKVKLSGS	SISSQY	xxxxxxxxxxxxxxxxxG	DVEIEIIDKLISIPYVE	

genpep	GTNCP	PPVRVNANGGL	PGGKVKLSGS	SISSQYLTALLMAAP	LALGDVEIEIVDKLISVPYVE	
	230	240	250	260	270	280

```

220      230      240      250      260      270
2MEPSP  MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxx
      :::::::::::::::::::::
genpep  MTLKLMEFVGSAEHSESWDRFFVKGQKYKSPGNAYVEGDASSASYFLAGAAITGETVT
      290      300      310      320      330      340

      280      290      300      310      320      330
2MEPSP  xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPPREPFGRKHLKAIDVNMNKMPD
      ..:::::::::::::::::::
genpep  VEGCGTTSLQGDVKFAEVLEKMGCKVSWTENSVTVTGPSRDAFGMRHLRAIDVNMNKMPD
      350      360      370      380      390      400

      340      350      360      370      380      390
2MEPSP  VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      :::::::::::::::::::::
genpep  VAMTLAVVALFADGPTTIRDVASWRVKETERMIAICTELRKLGATVEEGSDYCVITPPKK
      410      420      430      440      450      460

      400      410      420      430      440
2MEPSP  LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      ..:
genpep  VKPAEIDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH
      470      480      490      500      510      520

21555078|Arabidopsis thaliana clone 270032 mRNA, complete sequence.
thaliana|AA|521
Length = 521
initn: 1668 initl: 1368 opt: 1696 Z-score: 3401.4 bits: 638.8 E()
Smith-Waterman score: 1696; 74.492% identity (90.293% similar
overlap (4-445:79-521)

      10      20      30
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILL
      :::::::::::::::::::::
genpep  SWGLKKSDLMLNGSEIRPVKVRASVSTAEKASEIVLQPIREISGLIKLPGSKSLSNRILL
      50      60      70      80      90      100

      40      50      60      70      80      90
2MEPSP  LAALSEGTTVVDNLLNSDEVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVE-DAKE
      :::::::::::::::::::::
genpep  LAALSEGTTVVDNLLNSDDINYMLDALKIILGLNVETHSENNRAVVEGCGGVFPASIDSKS
      110      120      130      140      150      160

      100      110      120      130      140      150
2MEPSP  EVQLFGLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPVPRMRERPIGDLVVGKQLGADVDCFL
      ...:
genpep  DIELYLGNAGTAMRPLTAAVTAAGGNASYVLDGVPVQMRERPIGDLVVGKQLGADVECTL
      170      180      190      200      210      220

      160      170      180      190      200      210
2MEPSP  GTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYE
      :::::::::::::::::::::
genpep  GTNCPVVRVNANGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIVDKLISVPYE
      230      240      250      260      270      280

```


	220	230	240	250	260	270
2MEPSP	MTLRLMERFGVKA	EHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAI	xxxxxxx			

genpep	MTLKLME	RGVSAEHS	ESWDRFFVKG	GQKYKSPGNAYVEGDASSASYFLAGAAITGETVT		
	290	300	310	320	330	340
	280	290	300	310	320	330
2MEPSP	xxxxxxxSLQGDVKFAEVL	EMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKM	PD			

genpep	VEGCGTTS	LQGDVKFAEVL	EKMCKVSWTENS	VTVTGPSRDAFGMRHLRAIDVNMNKM	PD	
	350	360	370	380	390	400
	340	350	360	370	380	390
2MEPSP	VAMTLAVVALFADGPTA	IRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK				

genpep	VAMTLAVVALFADGPTT	IRDVASWRVKETERMIAICTELRKL	GATVEEGSDYCVITPPKK			
	410	420	430	440	450	460
	400	410	420	430	440	
2MEPSP	LNVT	AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVK	N			

genpep	VKPAE	IDTYDDHRMAMAFSLAACADVPITINDPGCTRKTFPDYFQVLERITKH				
	470	480	490	500	510	520
170374 Tomato 5-enolpyruvylshikimate-3-phosphate synthase mRNA, c						
<i>Solanum lycopersicum</i> AA 520						
Length = 520						
initn: 1694 initl: 1372 opt: 1692 Z-score: 3393.4 bits: 637.3 E()						
Smith-Waterman score: 1692; 75.737% identity (88.435% similar						
overlap (6-445:80-520)						
			10	20	30	
2MEPSP			MAGAE	EIVLQPIKEISGTVKLP	GSKSLSNRILL	LA
		
genpep	GVFKK	DSVLRVVRKSSFRISASVATAEKPHEIVL	XPIKDISGTVKLP	GSKSLSNRILL	LA	
	50	60	70	80	90	100
	40	50	60	70	80	90
2MEPSP	ALSEG	TTVDNLLNS	EDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-EDAKEEV			

genpep	ALSEG	RTVDNLLSSDDIHYMLGALKTLGLHVEDDNENQRAIVEGCGGQFPVGKKSEEEI				
	110	120	130	140	150	160
	100	110	120	130	140	150
2MEPSP	QLFLGNAGIAMRSL	xxxxxxxxxxxxxxxxxYVLDGVP	MRERP	IGDLV	VGLKQLGADVDC	FLGT

genpep	QLFLGNAGTAMRPL	TAAVTVAGGHSRYVLDGVP	MRERP	IGDLVDGLKQLGA	EVDCSLGT	
	170	180	190	200	210	220
	160	170	180	190	200	210
2MEPSP	DCPPVRVNGIGGLPGGKVK	LSGSISSQYxxxxxxxxxxxxxxxxxGDVEIEI	IDKLISIPYVEMT			

genpep	NCPPVRIVSKGGLPGGKVK	LSGSISSQYLTALLMAAPLALGDVEIEI	IDKLISVPYVEMT			
	230	240	250	260	270	280

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	220	230	240	250	260	270
2MEPSP	LRLMERFGVKA	EHSDSWDRFYIK	GGQKYKSPKNAY	VEGDASSASYFLA	GAAIxxxxxxxx	
	:	:	:	:	:	:
genpep	LKLMERFGVFE	HSSGWDRFLVK	GGQKYKSPGKAF	VEGDASSASYFLA	AAVTGGTVTVE	
	290	300	310	320	330	340

	280	290	300	310	320	330
2MEPSP	xxxxxSLQGDVK	FAEVLEMMGAKV	TWTETSVTVTG	PPREPFGRKHLK	AIDVNMNKMPDVA	
	:	:	:	:	:
genpep	GCGTSSLQGDVK	FAEVLEKMGAEV	TWTENSVTVKG	PPRNSSGMKHLRA	IDVNMNKMPDVA	
	350	360	370	380	390	400

	340	350	360	370	380	390
2MEPSP	MTLAVVALFADG	PTAIRDVASWRV	KETERMVAIRTEL	TKLGASVEEGPD	YCIITPPEKLN	
	:	:	:	:	:	:
genpep	MTLAVVALFADG	PTTIRDVASWRV	KETERMIAICTEL	RKLGATVVEGSD	YCIITPPEKLN	
	410	420	430	440	450	460

	400	410	420	430	440
2MEPSP	VTAIDTYDDHRMA	MAFSLAACAEVP	TIRDPGCTRKT	FPDYFDVLSTFV	KN
	:	:	:	:	:
genpep	VTEIDTYDDHRMA	MAFSLAACADVP	TIKNPGCTRKT	FPDYFEVLQKYS	KH
	470	480	490	500	510

63334427 | *Conyza sumatrensis* 5-enol-pyruvylshikimate-phosphate synthase (EPSPS-R3) mRNA, partial cds. | *Conyza sumatrensis* | AA | 444
Length = 444
initn: 1671 initl: 828 opt: 1668 Z-score: 3346.2 bits: 628.3 E(): 8e-178
Smith-Waterman score: 1668; 74.208% identity (89.819% similar) in 442 aa
overlap (5-445:5-444)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPI	KEISGTVKLPGS	KSLSNRILLLA	AALSEGTTVVDN	LLNSSEVDHYML	LGAL
	:	:	:	:	:	:
genpep	SKAPEEIVLKPI	QKISGTVHLP	GSKSLSNRILL	LAALSEGTTVVD	NLLNSDDVHYML	LGAL
	10	20	30	40	50	60

	70	80	90	100	110
2MEPSP	RTLGLSVEADKAA	KRAVVVGCGGK	FPV-EDAKEEVQ	LFLGNAGIAMRSL	xxxxxxxxxxxx
	:	:	:	:	:
genpep	RAVGLNVEENAAI	KRAIVEGCGGV	FPVGKEAKDEIQ	LFLGNSGTAMRTL	TAAITAAGGNS
	70	80	90	100	110

	120	130	140	150	160	170
2MEPSP	xYVLDGVPRMRER	PIGDLVVGLKQL	GADVDCFLGTDC	PPVRVNGIGGLP	GGKVKLSGSIS	
	:	:	:	:	:	:
genpep	SYVLDGVPRMRER	PIGDLVTGLKQL	GVNVDCSLGTNC	PPVRVVGSGGLP	GGRVRLSGSIS	
	130	140	150	160	170	180

	180	190	200	210	220	230
2MEPSP	SQYxxxxxxxxxxx	GDVEIEIIDKLIS	IPYVEMTLRLMER	FGVKA	EHSDSWDRFYIK	GG
	:	:	:	:	:	:
genpep	SVYLTSLMLA	PLALGDVEIEI	IDKLISVPYVQ	MTLKLKMQFGV	SVEHSDTLDRFH	VRGG
	190	200	210	220	230	240

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      240      250      260      270      280      290
2MEPSP QKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVT
      :   ::  ::::::::::::::::::::::. . . . . ::::::::::: :::::
genpep Q--KSPGNAYVEGDASNASYFLAGAAITGGTVTVEGCGTSSLLGDVKFAEVLGQMGA EVT
      250      260      270      280      290

      300      310      320      330      340      350
2MEPSP WTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVK
      :.:::: ::::. :: ::. ::::::::::::::::::::: :::::::::::
genpep WAENSVTVKGPPRNSSGRGHLRPVDVNMNKMPPDVAMALAVVALYADGPTAIRDVASWRVK
      300      310      320      330      340      350

      360      370      380      390      400      410
2MEPSP ETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
      :::::.:: :: ::::: ::::::::::::::::::::: :::::::::::
genpep ETERMIAICTELRKL GATVEEGPDYCVITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVP
      360      370      380      390      400      410

      420      430      440
2MEPSP VTIRDPGCTRKTFPD YFDVLSTFVKN
      :::: ::::::::::::::: . . .
genpep VTIKDPGCTRKTFPD YFEVLERYTKH
      420      430      440

```

110729299 | *Phaseolus vulgaris* EPSP synthase mRNA, complete cds. | *Phaseolus vulgaris* | AA | 522
Length = 522
initn: 1588 initl: 1296 opt: 1654 Z-score: 3316.9 bits: 623.1 E(): 3.4e-176
Smith-Waterman score: 1654; 74.150% identity (88.889% similar) in 441 aa
overlap (6-445:83-522)

```

                                10      20      30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLA
                                ::  ::::::::::: :::::::::::
genpep KDSMILSGAELSPFKVMSAVATAEKPSTSP EIEVEPIKDFSGWIKLP GTKSLSNRILLLA
      60      70      80      90      100     110

      40      50      60      70      80      90
2MEPSP ALSEGTTVVDNLLN SEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPVE-DAKEEV
      :::::::::::::: .: ::::: . ::: :::: :... :.....
genpep ALSEGTTVVDNLLN SDDIHYMLGALKWLGLRVETDKEINGAVVEGCGGIFPASIDSKSDI
      120     130     140     150     160     170

      100     110     120     130     140     150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      .: ::::: :: :... . . : :::::::::::::::::::::
genpep ELYLGNAGTAMRPLTAAVTAAGGNAWYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      180     190     200     210     220     230

      160     170     180     190     200     210
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMT
      .: ::::: ::::::::::: ::::: . . . : :::::::::::
genpep NCPVVRVNAGGGLPGGKVKLFGSISSQYLTALLMSAPLALGDVEIEIIDKLISVPYVEVT
      240     250     260     270     280     290

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	220	230	240	250	260	270
2MEPSP	LRLMERFGVKA	EHSDSWDRFYIK	GGQKYKSPKNAY	VEGDASSASYFL	AGAAIxxxxxxx	
	:. :. :. :. :. :. :.	:. :. :. :. :. :.	:. :. :. :. :. :.	:. :. :. :. :. :.	:. :. :. :. :. :.	:. :. :. :. :. :.
genpep	LKLMERFGVS	VEHWDSWDRFL	VHGGQKYKSPGN	AYVEGDASSASY	LLAGAAITGGT	VTVE
	300	310	320	330	340	350

	280	290	300	310	320	330
2MEPSP	xxxxxSLQGD	VKFAEVLEMMG	AKVTWTETSV	TVTGTGPPREP	FGRKHLKAID	VNMNKM
	. :. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.
genpep	GCGTKSLQGD	VKFAEVLEKMG	CKVWWTENSV	TVTGTGPPRDL	FGRRLRAID	VNMNKM
	360	370	380	390	400	410

	340	350	360	370	380	390
2MEPSP	MTLAVVALFAD	GPTAIRDVAS	WRVKETERMVA	IRTELTCLGAS	VEEGPDYCIIT	PPEKLN
	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.
genpep	MTLAVVALFAD	GPTTIRDVAS	WRVKETEEMIA	ICTELRKLGA	TVEEGPDYCVI	TPPKLK
	420	430	440	450	460	470

	400	410	420	430	440
2MEPSP	VT	AITDYDDHRMA	MAFSLAACAE	VPVTIRDPGCT	RKTFPDYFDVL
	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.
genpep	VAEIDTYDDH	RIPMAFSLAAC	ADVPVTINDP	-CTRKTFPDYF	EVLERLTKH
	480	490	500	510	520

13375567 | *Lolium rigidum* 5-enolpyruvylshikimate 3-phosphate synthase (epsp-s) mRNA, epsp-s-S allele, partial cds. | *Lolium rigidum* | AA | 347
Length = 347
initn: 1473 initl: 1379 opt: 1474 Z-score: 2957.6 bits: 556.1 E(): 3.5e-156
Smith-Waterman score: 1474; 84.104% identity (91.908% similar) in 346 aa
overlap (65-409:2-347)

	40	50	60	70	80	90
2MEPSP	AALSEGTTV	VDNLLNSEDV	HYMLGALRTL	GLSVEADKAA	KRAVVVGCGG	KFPVE-DAKEE
				:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.
genpep				PLSVEADKVA	KRAVVVGCGG	RFPXEKDAKEE
				10	20	30

	100	110	120	130	140	150
2MEPSP	VQLFLGNAGI	AMRSLxxxxxxx	xxxxxxxYVLD	GVPRMRERPI	GDLVVGLKQL	GADVDCFLG
	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.
genpep	VKLFLGNAGT	AMRPLTAAVVA	AGGNATYVLD	GVPRMRERPI	GDLVVGLKQL	GANVDCFLG
	40	50	60	70	80	90

	160	170	180	190	200	210
2MEPSP	TDCPPVRVNG	IGGLPGGKVK	LSGSISSQYxxxx	xxxxxxxGDVE	IEIIDLKLI	SIPYVEM
	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.	:. :. :. :. :.
genpep	TDCPPVRING	IGGLPGGKVK	LSGSISSQYL	SSLLMAAPLAL	GDVEIEIIDL	KLISVPYVEM
	100	110	120	130	140	150

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      260      270      280      290      300      310
2MEPSP SASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPF
      ::::::::::. . . . . ::::::::::: . ::::::::::: ::::.
genpep SASYFLAGAAITGGTITVEGCGTSSLQGDVKFAEVLGQMGAEVTWTENSVTVKGPPRDSS
      220      230      240      250      260      270

      320      330      340      350      360      370
2MEPSP GRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLG
      ::::. . ::::::::::: . ::::::::::: . ::::: ::::
genpep GRKHLRAVDVNMNKMPPDVAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLG
      280      290      300      310      320      330

      380      390      400      410      420      430
2MEPSP ASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDY
      . ::::::::::. :::::::::::
genpep ATVEEGPDYCVITPPEKLNVTATIDTYDD
      340      350

```

48526072|*Erigeron annuus* 5-enol-pyruvylshikimate-phosphate synthase (EPSPS2)
mRNA, partial cds. |*Erigeron annuus*|AA|360
Length = 360
initn: 1380 initl: 1223 opt: 1382 Z-score: 2772.4 bits: 521.8 E(): 7.3e-146
Smith-Waterman score: 1382; 75.000% identity (88.889% similar) in 360 aa
overlap (45-403:1-360)

```

      20      30      40      50      60      70
2MEPSP ISGTVKLPKSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAK
      ::::. ::::::::::: . : :
genpep DNLNLSDDVHYMLGALRALGLNVEENGAIK
      10      20      30

```

```

      80      90      100      110      120      130
2MEPSP RAVVVGCGGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERP
      ::: ::::: . . . . . ::::: ::::: . . . . . :::::::::::
genpep RAIVEGCGGVFPVGKEAKDEIQFLGNAGTAMRPLTAAVTAAGGNSSYVLDGVPRMRERP
      40      50      60      70      80      90

```

```

      140      150      160      170      180      190
2MEPSP IGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxx
      ::::. ::::::::::: . ::::: . ::::::::::: . . . . .
genpep IGDLVVGLKQLGANVDCSLGTNCPVRVVGSGGLPGGKVKLSGSISSQYLTSLMAAPLA
      100      110      120      130      140      150

```

```

      200      210      220      230      240      250
2MEPSP xGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGD
      ::::::::::: . ::::::::::: . ::::::::::: . ::::::::::: . :::::::::::
genpep LGDVEIEIIDKLISVPYVEMTLKLMEFVGVSVEHNDTWDRFHVGGQKYKSPGNAYVEGD
      160      170      180      190      200      210

```

```

      260      270      280      290      300      310
2MEPSP ASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPRE
      ::::::::::: . . . . . ::::::::::: . ::::::::::: ::::.
genpep ASSASYFLAGAAITGGTVTVEGCGTSSLQGDVKFAEVLGQMGAEVTWTENSVTVKGPPRN
      220      230      240      250      260      270

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          320      330      340      350      360      370
2MEPSP PFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTK
      :: :: . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep SSGRGHLRPVDVNMNKMPPDVAMTLAVVALYABGPTAIRDVASWRVKETERMIAICTELRK
          280      290      300      310      320      330

```

```

          380      390      400      410      420      430
2MEPSP LGASVEEGPDYCIITPPEKLNVTDAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTTFP
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep LGATVEEGTDYCVITPPEKLNVTDAIDTYDD
          340      350      360

```

48526066 | *Amaranthus rudis* 5-enol-pyruvylshikimate-phosphate synthase (EPSPS1) mRNA, partial cds. | *Amaranthus tuberculatus* | AA | 357
Length = 357
initn: 1373 initl: 1249 opt: 1375 Z-score: 2758.3 bits: 519.2 E(): 4.4e-145
Smith-Waterman score: 1375; 75.630% identity (88.796% similar) in 357 aa
overlap (47-402:1-357)

```

          20      30      40      50      60      70
2MEPSP GTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRA
      :: : : . : : : : : : : : : : : : : : : : : : : : : : : :
genpep LLYSDDILYMLDALRTLGLKVEDDNTDKRA
          10      20      30

```

```

          80      90      100      110      120      130
2MEPSP VVVGCGGKFPV-EDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIG
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep VVEGCGGLFPVGKDGKEEIQFLGNAGTAMRPLTA AVAVAGGNSSYVLDGVPRMRERPIG
          40      50      60      70      80      90

```

```

          140      150      160      170      180      190
2MEPSP DLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxG
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep DLVAGLQLGSDVDCFLGTNCPVVRVNAKGGLPGGKVKLSGSSQYLTALLMATPLGLG
          100      110      120      130      140      150

```

```

          200      210      220      230      240      250
2MEPSP DVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDAS
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep DVEIEIIVDKLISVPYVEMTIRLMERFGVSVSHSDSWDRFFIRGGQKYKSPGKAYVEGDAS
          160      170      180      190      200      210

```

```

          260      270      280      290      300      310
2MEPSP SASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPF
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep SASYFLAGAAVTGGTVTVKGCGTSSLQGDVKFAEVLEKMGCKVTWTDNSVTVTGPPRESS
          220      230      240      250      260      270

```

```

          320      330      340      350      360      370
2MEPSP GRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLG
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep GRKHLRAIDVNMNKMPPDVAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLK
          280      290      300      310      320      330

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          380          390          400          410          420          430
2MEPSP ASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDY
      :.:::: :.:.::::: :.:.::::
genpep ATVEEGSDYCVITPPEKLNPTA IETYD
          340          350

```

37359246 | *Vitis vinifera* 5-enolpyruvylshikimate-3-phosphate synthase mRNA, partial cds. | *Vitis vinifera* | AA | 330
Length = 330
initn: 1336 initl: 687 opt: 1333 Z-score: 2674.4 bits: 503.6 E(): 2.1e-140
Smith-Waterman score: 1333; 79.077% identity (91.692% similar) in 325 aa
overlap (121-445:7-330)

```

          100          110          120          130          140          150
2MEPSP KEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDC
      :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.:::::
genpep                AGGNASYVLDGVPRMRERPIGDLVTGLKQLGADVNC
                        10          20          30

```

```

          160          170          180          190          200          210
2MEPSP FLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPY
      :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.:::::
genpep FLGTNCPVVRVSGNGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISIPY
          40          50          60          70          80          90

```

```

          220          230          240          250          260          270
2MEPSP VEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxx
      :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.:::::
genpep VEMTLKLMERFGVSVEHSNTWDRFLIRGGQKYKSPGNAFVEGDASSASYFLAGAAVTGGT
          100          110          120          130          140          150

```

```

          280          290          300          310          320          330
2MEPSP xxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMK
      . :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.:::::
genpep VTVEGCGTSSLQGDVKFVEFLSM-GAKVSWTENSVTVTGPPQDSSGRKHLRAIDVNMNMK
          160          170          180          190          200          210

```

```

          340          350          360          370          380          390
2MEPSP PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP
      :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.:::::
genpep PDVAMTLAVVALYAEGPTAIRDVASWRVKETERMIAICTELRKLKLGATVEEGPDYCVITPP
          220          230          240          250          260          270

```

```

          400          410          420          430          440
2MEPSP EKLNVT AIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.::::: :.:.:::::
genpep EKLNVT SIDTYDDHRMAMAFSLAACADVPVTIKDPGCIKRTSPDYFEVLQRFTHK
          280          290          300          310          320          330

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

170231|*N.tabacum* 5-enolpyruvylshikimate-3-phosphate synthase mRNA, 5' cds.
|*Nicotiana tabacum*|AA|338
Length = 338
initn: 1327 initl: 1327 opt: 1327 Z-score: 2662.2 bits: 501.4 E(): 1e-139
Smith-Waterman score: 1327; 79.077% identity (90.462% similar) in 325 aa
overlap (121-445:14-338)

	100	110	120	130	140	150
2MEPSP	KEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDC					
					
genpep		LTAAVAVAGGNSRYVLDGVPRMRERPIGDLVDGLKQLGAEVDC				
		10	20	30	40	

	160	170	180	190	200	210
2MEPSP	FLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPY					
	:
genpep	FLGTKCPPVRIVSKGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISVLY					
	50	60	70	80	90	100

	220	230	240	250	260	270
2MEPSP	VENTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxx					

genpep	VENTLKLMERFGISVEHSSSWDRFVVRGGQKYKSPGKAYVEGDASSASYFLAGAAVTGGT					
	110	120	130	140	150	160

	280	290	300	310	320	330
2MEPSP	xxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNMK					

genpep	VTVEGCGTSSLQGDVKFAEVLEQMGAEVTWTENSVTVKGPPRNSSAMKHLRAIDVNMNMK					
	170	180	190	200	210	220

	340	350	360	370	380	390
2MEPSP	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPP					

genpep	PDVAMTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLGAATVEEGPDYCIITPP					
	230	240	250	260	270	280

	400	410	420	430	440
2MEPSP	EKLNVTIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN				

genpep	EKLNVTIDTYDDHRMAMAFSLAACADVPVTINDPGCTRKTFPNYFDVLQQYSKH				
	290	300	310	320	330

144582815|*Ostreococcus lucimarinus* CCE9901 chromosome 20, complete sequence.
|*Ostreococcus lucimarinus* CCE9901|AA|436
Length = 436
initn: 1122 initl: 396 opt: 1263 Z-score: 2531.7 bits: 477.6 E(): 1.8e-132
Smith-Waterman score: 1263; 59.732% identity (80.089% similar) in 447 aa
overlap (5-444:2-436)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL					

genpep	MEQLTLKPMKKIEGTVRLPGSKSLSNRILLLAALAEGTTKVENLLDSDDIRYMVDAL					
	10	20	30	40	50	

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
genpep  KVLGLSFTEARENILEITGCGGKLPVEGA---ELFLGNAGTAMRPLTAAVAAAGKGT-
                60          70          80          90          100          110

                130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
genpep  FILDGVERMRERPIQDLVDGLVQLGVKAECTMGTCPPVKVEA-NGLPGGRVELSGSVSS
                120          130          140          150          160          170

                190          200          210          220          230          240
2MEPSP  QYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
genpep  QYLTALLMAAPLCEGSIEIVIVDELISKPYVEMTITLMERFGVKVEKADDLQSFQGGQ
                180          190          200          210          220          230

                250          260          270          280          290          300
2MEPSP  KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
genpep  KYISPGSAFVEGDASSASYFLAGATITGGTVTVIGCGSESIQGDTNFAYTMEQMGATLEW
                240          250          260          270          280          290

                310          320          330          340          350          360
2MEPSP  TETSVTVTGPPREPFGKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
genpep  GPNSVTCTGP-----KGPLKAIDVNMNAMPDAAMTLAVAALFADGITTIRDVASWRVKE
                300          310          320          330          340

                370          380          390          400          410
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT-----IDTYDDHRMAMAFSLA
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
genpep  TERMIAICTELRKLGCDVFEGADYCVITPPHKLDPPAKMKANVDIDTYDDHRMAMAFALA
                350          360          370          380          390          400

                420          430          440
2MEPSP  ACAEVPVTIRDPGCTRKTFFDYFDVLSTFVKN
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
genpep  ACGDVDVVINDPKCTKKTFFDYFDVLKSVAK
                410          420          430

```

34577025|*Oryza sativa* (japonica cultivar-group) putative EPSP1 mRNA, partial cds. |*Oryza sativa Japonica Group*|AA|331
Length = 331
initn: 1108 initl: 1108 opt: 1238 Z-score: 2483.3 bits: 468.2 E(): 9.2e-130
Smith-Waterman score: 1238; 82.886% identity (91.946% similar) in 298 aa
overlap (56-352:1-298)

```

                30          40          50          60          70          80
2MEPSP  SLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKF
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
genpep  MLEALKALGLSVEADKVAKRAVVVGCGGKF
                10          20          30

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          90          100          110          120          130          140
2MEPSP PVE-DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQL
      :: :::::::::::::: :: :... .. ::::::::::::::::::::
genpep PVEKDAKEEVQLFLGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQL
          40          50          60          70          80          90

          150          160          170          180          190          200
2MEPSP GADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDK
      ::::::::::: ::::::::::: ::::::::::: :: .. .. : :::::::::::
genpep GADVDCFLGTDCPPVRVKGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEIIDK
          100          110          120          130          140          150

          210          220          230          240          250          260
2MEPSP LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGA
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
genpep LISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPGNAYVEGDASSASYFLAGA
          160          170          180          190          200          210

          270          280          290          300          310          320
2MEPSP AIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGKHLKAID
      :: . . . ::::::::::: ::::::::::: ::::::::::: :::::::::::
genpep AITGGTVTVQCGTTSLQGDVKFAEVLEMMGAKVTWTDTSVTVTGPPREPYGKHLKAID
          220          230          240          250          260          270

          330          340          350          360          370          380
2MEPSP VNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDY
      ::::::::::: ::::::::::: :::::::::::
genpep VNMNKMPPDVAMTLAVVALFADGPTAIRDGFLESKGNRKDGCNSDRANKGKFIRSRVLSFF
          280          290          300          310          320          330

```

122937807 | *Dunaliella salina* plastid EPSP synthase mRNA, complete cds; nuclear gene for plastid product. | *Dunaliella salina* | AA | 514
Length = 514
initn: 1124 initl: 460 opt: 1219 Z-score: 2442.1 bits: 461.3 E(): 1.8e-127
Smith-Waterman score: 1219; 58.239% identity (80.361% similar) in 443 aa
overlap (5-445:81-514)

```

                                10          20          30
2MEPSP                      MAGAEEIVLQPIKEISGTVKLPKSKSLSNRILL
                                .....
genpep RRSCSKSSIRSTRLQTTVCSATLAHSAPDQLVLQPIKQISGTVRLPGSKSISNRVLLL
                                60          70          80          90          100          110

          40          50          60          70          80          90
2MEPSP AALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEV
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: . :: ::::::::::: . :
genpep AALAEGTTVVKNLDSDDIRYMGALKGLGIELEERWDKGEMVVKGCGGQFSAEGG----
          120          130          140          150          160

          100          110          120          130          140          150
2MEPSP QLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGT
      ::::::::::: :: :... .. :::::: ::::::::::: :: : ::::::::::: : ::
genpep ELFLGNAGTAMRPLTAAVAAAGRK-FVLDGTARMRERPIQDLVDGLVQLGVDKACPLGT
          170          180          190          200          210          220

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      160      170      180      190      200      210
2MEPSP DCPVVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGD--VEIEIIDKLISIPYVE
      :::::  :::::::::::::::  ..  .  :  .:: : :::::  ::::
genpep GCPPVEVNA-QGLPSGKVQLKGSVSSQYL TALLMAAPLSKGTEGIEIVITDELVSQPYVD
      230      240      250      260      270      280

      220      230      240      250      260      270
2MEPSP MTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxx
      :::::::::::  ::  ..  .  :  .:  ::  .:::::::::::::::::::  .
genpep MTVQIMERFGVTVERLNLQHMRIIPNQTYKTSGEAFVEGDASSASYFLAGATITGGTVV
      290      300      310      320      330      340

      280      290      300      310      320      330
2MEPSP xxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPD
      ..:::::::::  :::::  :  .  .  ::::  .  :: : :::::  .  :  ::
genpep VEGCGSASVQGDVRF AEVMGLMGAKVEWSLYSIKITGP--SAFG-KPLQGIDHDCNDIPD
      350      360      370      380      390      400

      340      350      360      370      380      390
2MEPSP VAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK
      :::::::::::  :::::  .:::::::::::::  :  :::::  :::::  :::::  .
genpep AAMTLAVAALFADKPTTIRNVYNWRVKETERMVAIVNETRKL GATVEEGRDYCVITPPKQ
      410      420      430      440      450      460

      400      410      420      430      440
2MEPSP LNVTAIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN
      ..  :::::::::::::::::::::  :::::::::::::::::::::  :...  ...
genpep IQSAAIDTYDDHRMAMAFSLAACGPVPVTINDPGCTRKTFFPDYFRVLESVTQH
      470      480      490      500      510

```

149390731 | *Oryza sativa* (indica cultivar-group) clone OSE-97-192-D11 5-enolpyruvylshikimate-3-phosphate synthase mRNA, partial cds. | *Oryza sativa Indica Group* | AA | 273
Length = 273
initn: 1152 initl: 1152 opt: 1152 Z-score: 2311.6 bits: 436.2 E(): 3.3e-120
Smith-Waterman score: 1152; 83.150% identity (91.941% similar) in 273 aa
overlap (96-368:1-273)

```

      70      80      90      100      110      120
2MEPSP SVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLVDG
      :::::  ::  :...  ...  :::::
genpep LFLGNAGTAMRPLTAAVTAAGGNATYVLVDG
      10      20      30

      130      140      150      160      170      180
2MEPSP VPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxx
      :::::::::::::::::::::  :::::::::::::::::::::  ..
genpep VPRMRERPIGDLVVGLKQLGADVDCFLGT ECPPVRVKIGIGGLPGGKVKLSGSISSQYLSA
      40      50      60      70      80      90

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      190      200      210      220      230      240
2MEPSP xxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSP
      .. . :.....:.....:.....:.....:.....:.....:.....:.....:.....:
genpep LLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSP
      100      110      120      130      140      150

      250      260      270      280      290      300
2MEPSP KNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSV
      :.....:.....:.....:.....:.....:.....:.....:.....:.....:
genpep GNAYVEGDASSASYFLAGAAITGGTVTVQGCGTTSLQGDVKFAEVLEMMGAKVTWTDTSV
      160      170      180      190      200      210

      310      320      330      340      350      360
2MEPSP TVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMV
      :.....:.....:.....:.....:.....:.....:.....:.....:.....:
genpep TVTGPPREPYPGKKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMV
      220      230      240      250      260      270

      370      380      390      400      410      420
2MEPSP AIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVPVTIRDP
      :::
genpep AIR

```

48526074 | *Helianthus salicifolius* 5-enol-pyruvylshikimate-phosphate synthase (EPSPS1) mRNA, partial cds. | *Helianthus salicifolius* | AA | 264
Length = 264
initn: 951 initl: 866 opt: 956 Z-score: 1917.7 bits: 363.3 E(): 2.9e-98
Smith-Waterman score: 956; 71.970% identity (86.364% similar) in 264 aa overlap (61-323:1-264)

```

      40      50      60      70      80
2MEPSP ILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-ED
      :.....:.....:.....:.....:.....:.....:.....:.....:.....:
genpep                                RTLGLRVEEDGAIKRAIVEGCGGVFPVGKE
      10      20      30

      90      100      110      120      130      140
2MEPSP AKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVD
      :.....:.....:.....:.....:.....:.....:.....:.....:.....:
genpep AKDEIQFLFLGNAGTSMRPLTAAVTAAGNSSYILDGVPRMRERPIGDLVTGLKQLGADVD
      40      50      60      70      80      90

      150      160      170      180      190      200
2MEPSP CFLGTD CPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxxGDVEIEIIDKLISIP
      :.....:.....:.....:.....:.....:.....:.....:.....:.....:
genpep CFLGTN CPPVRVVGGLPGGKVKLSGSISSQYLTALLMASPLALGDVEIEIIDKLISIP
      100      110      120      130      140      150

      210      220      230      240      250      260
2MEPSP YVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxx
      :.....:.....:.....:.....:.....:.....:.....:.....:.....:
genpep YVDMTIKLMERFGVSVSHSDSWDRFFIKGGQKYKSPGNAYVEGDASSASYFLAGAAITGG
      160      170      180      190      200      210

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNK
      . .      . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep TITVEGCGTSSLQGDVKFAEVLGQMGAEVTWTENSVTVRGPPRNASGRKHLHAV
      220      230      240      250      260

```

```

      330      340      350      360      370      380
2MEPSP MPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITP

```

48526082 | *Plantago lanceolata* 5-enol-pyruvylshikimate-phosphate synthase (EPSPS) mRNA, partial cds. | *Plantago lanceolata* | AA | 265
Length = 265
initn: 861 initl: 861 opt: 951 Z-score: 1907.6 bits: 361.4 E(): 1.1e-97
Smith-Waterman score: 951; 71.321% identity (86.038% similar) in 265 aa overlap (60-323:1-265)

```

      30      40      50      60      70      80
2MEPSP RILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCCKFPV-E
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep LRTLGLDVEEDKANQRATVGGCGGLFPVSK
      10      20      30

```

```

      90      100      110      120      130      140
2MEPSP DAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADV
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep ESKDEIQFLFLGNAGTAMRPLTAAVVAAGGDATYVLDGVPRMRERPIGDLVTGLKQLGADI
      40      50      60      70      80      90

```

```

      150      160      170      180      190      200
2MEPSP DCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISI
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep DCFLGTCKPPVRVIGKGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEIIDKLISV
      100      110      120      130      140      150

```

```

      210      220      230      240      250      260
2MEPSP PYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxx
      . : : : : : : : : : . : : : : : : : : : . : : : : : : : : : . : : : : : : : :
genpep PYVEMTLKLMERFGVYVEHTDSWDRFLVRGGQKYKSPGKAYVEGDASSASYFLAGAAVTG
      160      170      180      190      200      210

```

```

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMN
      . .      . : : : : : : : : : : : : : : : : : : : : : : :
genpep GTVTVGCGTSSLQGDVKFAEVLKMGAEVTWTENSVTVKGPPRNSSGRKHLRAI
      220      230      240      250      260

```

```

      330      340      350      360      370      380
2MEPSP KMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGC G G K F P V E D A K E E V Q L F L G N A G I A M R S L x x x x x x x x x x x x
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep  DTLGVRCEPLGTANAYRV T G T G G R F P A K S A - - - - D L F M G N A G T A I R P L T A A L A L Q G G E - -
              60          70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep  YTLHGVP RM H E R P I G D L V D G L R Q V G A R I D Y T G N E G F P P L A I R A A S I R I D A P I R V R G D V S S
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxxxxxxGD---VEIEIIDKLISIPYVEMTLRLMERFGVKA E H S D S W D R F Y I K
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep  QFLTALLMALPLVEGSGRPVTIEVVGELISKPYIEITLNLMARFGVQVERN-GWASFSVP
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      : . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep  TGVAYRAPGEIFVEGDASSASYFLAAGALGGGPVRVEGVGMSSIQGDVRFADALNRMGAN
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  VTWTETSVTVTGPPREPFGRKHLKAIDVNMNKM P D V A M T L A V V A L F A D G P T A I R D V A S W R
      : . . . . : : . . . . : . . . . : . . . . . : . . . . . : . . . . .
genpep  VMAGDNWIEVRGVERDD-GKLH--ALELDCNHIPDAAMTLAVAALFADGTTTLTNIGSWR
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  VKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT A - - I D T Y D D H R M A M A F S L A A C
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep  VKETDRLTAMATELRKLGA A V E E G T D Y I R V T P P S H W T A P A G G I D T Y D D H R M A M A F S L A A F
              350          360          370          380          390          400

              420          430          440
2MEPSP  AEVPVTIRDPGCTRKTFPDYFDVLSTFVK N
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep  GPVPVRINDPRCVAKTFPEYFTAFGGIAA
              410          420          430

```

116054912 | *Ostreococcus tauri* strain OTTH0595, *** SEQUENCING IN PROGRESS ***.
 |*Ostreococcus tauri*|AA|316
 Length = 316
 initn: 958 initl: 386 opt: 935 Z-score: 1874.2 bits: 355.5 E(): 7.7e-96
 Smith-Waterman score: 935; 60.991% identity (78.638% similar) in 323 aa overlap
 (129-444:1-316)

```

              100          110          120          130          140          150
2MEPSP  GNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPP
      . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep  MRERPIADLV D G L V Q L G V K A E C T M G T G C P P
              10          20          30

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

160      170      180      190      200      210
2MEPSP VRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLM
      :.... :.....:.....:.. .. :...: :.....: :.....: ..
genpep VKIEA-DGLPGGRVELSGSVSSQYLTAALLMAAPLCQGSIEIVIVDELISKPYVEMTITLM
      40      50      60      70      80

220      230      240      250      260      270
2MEPSP ERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxx
      :.....: : . . : :.....: : : :.....: :.....: : . . .
genpep ERFGVKVEKSADLQSFQIQQGQKYVSPGSAFVEGDASSASYFLAGATITGGTVTVIGCGS
      90      100      110      120      130      140

280      290      300      310      320      330
2MEPSP xSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLA
      :.....: :.....: . : :..: :..: . :.....: :.....:
genpep ESIQGDNTFAYTMEQMGATLEWGPNSVKCTGP-----QGPKAIDVNMNAMPDAAMTLA
      150      160      170      180      190      200

340      350      360      370      380      390
2MEPSP VVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT-
      :.....: :.....: :.....: : : :.....: : : :.....: : :
genpep VAALFADGVTTIRDVASWRVKETERMIAICTELRKLGCDFEGSDYCVITPPHKLNPAPK
      210      220      230      240      250      260

400      410      420      430      440
2MEPSP -----IDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      :.....: :.....: : : : : : : : : : : : : : : : :
genpep MRANVDIDTYDDHRMAMAFALAACGDVDVIINDPTCTKKTFTPTFDVLKSVVQ
      270      280      290      300      310

48526076 | Helianthus salicifolius 5-enol-pyruvylshikimate-phosphate synthase
(EPSPS2) mRNA, partial cds. | Helianthus salicifolius | AA | 264
Length = 264
initn: 926 initl: 857 opt: 931 Z-score: 1867.4 bits: 353.9 E(): 1.9e-95
Smith-Waterman score: 931; 69.697% identity (85.985% similar) in 264 aa overlap
(61-323:1-264)

40      50      60      70      80
2MEPSP ILLLAALSEGTTVVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPV-ED
      :.....: . :.....: :.....: :.....: ..
genpep RALGLNVEENGEIKRATVEGCGGVFPVGKE
      10      20      30

90      100      110      120      130      140
2MEPSP AKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADVD
      :.....: : : :... .. :.....: :.....: :.....: :.....:
genpep AKDEIQFLFLGNAGTAMRPLTAAVTAAGNSSYILDGVPRMRERPIGDLVTGLKQLGADVD
      40      50      60      70      80      90

150      160      170      180      190      200
2MEPSP CFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxGDVEIEIIDKLISIP
      :.....: : : :.....: :.....: : .. . :.....: :.....:
genpep CFLGTNCPVRAANGGLPGGKVKLSGSISSQYLTAALLMAAPLALGDVEIEIIDKLISVP
      100      110      120      130      140      150

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      210      220      230      240      250      260
2MEPSP YVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxx
      .....
genpep YVEMTLKLMERFGVSVSHSDSWDKFYVRGGQKYKSPGNAYVEGDASSASYFLAGAAITGG
      160      170      180      190      200      210

```

```

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNK
      . . . . .
genpep TVTVEGCGTSSLQGDVKFAEVLGQMGAEVTWTENSVTVRGPPRNASGRGHLRPV
      220      230      240      250      260

```

```

      330      340      350      360      370      380
2MEPSP MPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITP

```

73912408 | *Lactuca sativa* mRNA for 5-enolpyruvylshikimate-3-phosphate synthase, partial cds, clone: EPSPS1. | *Lactuca sativa* | AA | 231
Length = 231
initn: 926 initl: 926 opt: 926 Z-score: 1858.2 bits: 352.1 E(): 6e-95
Smith-Waterman score: 926; 77.922% identity (90.476% similar) in 231 aa overlap (174-404:1-231)

```

      150      160      170      180      190      200
2MEPSP LGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIID
      .....
genpep LSGSISSQYLTALLMASPLALGDVEIEIID
      10      20      30

```

```

      210      220      230      240      250      260
2MEPSP KLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAG
      .....
genpep KLISIPYVEMTLKLMERFGVSVSHSDSWDRFFVRGGQKYKSPGNAYVEGDASSASYFLAG
      40      50      60      70      80      90

```

```

      270      280      290      300      310      320
2MEPSP AAxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAI
      . . . . .
genpep AAITGGTITVEGCGTSSLQGDVKFAEVLGQMGAEVTWTENSVTVKGPPRNPSGRKHLRAV
      100      110      120      130      140      150

```

```

      330      340      350      360      370      380
2MEPSP DVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPD
      .....
genpep DVNMNKMMPDVAMTLAVVALYADGPTTIRDVASWRVKETERMIAICTELRKLGLATVEEGAD
      160      170      180      190      200      210

```

```

      390      400      410      420      430      440
2MEPSP YCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFV
      . . . . .
genpep YCVITPPEKLNVAIDTYDDH
      220      230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

73912410 | *Lactuca sativa* mRNA for 5-enolpyruvylshikimate-3-phosphate synthase, partial cds, clone: EPSPS2. | *Lactuca sativa* | AA | 231
Length = 231
initn: 922 initl: 922 opt: 922 Z-score: 1850.2 bits: 350.6 E(): 1.7e-94
Smith-Waterman score: 922; 77.922% identity (89.610% similar) in 231 aa overlap (174-404:1-231)

	150	160	170	180	190	200
2MEPSP	LGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIID					
				::::::::::	::::::::::
genpep				LSGSISSQXLTALLMAAPLALGDVEIEIID		
				10	20	30

	210	220	230	240	250	260
2MEPSP	KLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAG					
	::::::::::
genpep	KLISIPYVEMTLKLMERFGVSVQHSDTWDRFHVQGGQKYKSPGNAYVEGDASSASYFLAG					
	40	50	60	70	80	90

	270	280	290	300	310	320
2MEPSP	AAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAI					
	::::
genpep	AAITGGTITVEGCGTSSLQGDVKFAEVLGQMGQAQVTWTENSVTVKGPPRDPGRKHLRPV					
	100	110	120	130	140	150

	330	340	350	360	370	380
2MEPSP	DVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPD					
	::::::::::
genpep	DVNMNKMPPDVAMTLAVVALYADGPTAIRDVASWRVKETERMIAICTELRKLGLATVEEGPD					
	160	170	180	190	200	210

	390	400	410	420	430	440
2MEPSP	YCIITPPEKLNVTATIDTYDDHRMAMAFSLAACAEVPVTIRDPGCTRKTFPDYFDVLSTFV					
	::::::::::				
genpep	YCIITPPXKLNVTATIDTYDDH					
	220	230				

48526068 | *Asimina triloba* 5-enol-pyruvylshikimate-phosphate synthase (EPSPS) mRNA, partial cds. | *Asimina triloba* | AA | 264
Length = 264
initn: 897 initl: 872 opt: 887 Z-score: 1778.9 bits: 337.6 E(): 1.6e-90
Smith-Waterman score: 887; 66.288% identity (83.712% similar) in 264 aa overlap (60-323:1-264)

	30	40	50	60	70	80
2MEPSP	RILLLAALSEGTTVDNLLNSEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGGKFPVED					
				::::::::::
genpep				LQTLGLNVEEDSAANKATVEGCGGQFPVGK		
				10	20	30

	90	100	110	120	130	140
2MEPSP	AKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxYVLDGVPRMRERPIGDLVVGLKQLGADV					

genpep	DAKDIQLFLGNAGTAMRPLTAAVVAAGGKSRYVLDGVPRMRERPIGDLVSGLKQLDADV					
	40	50	60	70	80	90

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      150      160      170      180      190      200
2MEPSP CFLGTDPCPPVRVNGIGGLPGGKVKLSGSISSQYxxxxxxxxxxxxxGDVEIEIIDKLISIP
      :  :::::  ::  ::  ::::::::::::::  ..  ..  .  :::::::::::::::
genpep  CVLGTNCPVPSINANGGLRGKVKLSGTLSSQFLTSILMAAPLALGDVEIEIIDKLISVP
      100      110      120      130      140      150

```

```

      210      220      230      240      250      260
2MEPSP YVENTLRLMERFGVKAHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxx
      :::::  ::::::::::  ::::  ::::  ..::::::::  ..::::::::::::::::::
genpep  YVENTLKLMERFGVSAEHSGSWDRFLVRGGQKYKSPGTAFVEGDASSASYFLAGAAVTGG
      160      170      180      190      200      210

```

```

      270      280      290      300      310      320
2MEPSP xxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFGKHLKAIDVNMNK
      .  .  ..::::::::::::  :::::  ::::::::::::::  .:  .:::  .:
genpep  TVTVEGCGTSSLQGDVKFAEVLEKMGAKVSWTENSVTVTGPPPLDPSRKKRLHGI
      220      230      240      250      260

```

```

      330      340      350      360      370      380
2MEPSP MPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITP

```

91696897 | *Polaromonas* sp. JS666, complete genome. | *Polaromonas* sp JS666 | AA | 668
Length = 668
initn: 875 initl: 313 opt: 880 Z-score: 1758.6 bits: 335.2 E(): 2.1e-89
Smith-Waterman score: 880; 46.120% identity (71.840% similar) in 451 aa overlap
(1-439:1-440)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :  :  .  .  .  .  ::::::::::::::  ::  ::  ::  ::  ::::
genpep  MFDIEYLDIPPLARAGGTVRLPGSKSISNRVLLLAALSRGQTTVHDLLASDDTAVMLAAL
      10      20      30      40      50      60

```

```

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      .  ::  ::  .  .  ::  :  ::  :  ::  .  ::::::  ::  ::::
genpep  KQLGCSVA--QHGTTAVIDGLGGQ--VGQAK--ATLFMGNAGTAMRPLTAALALLGGE--
      70      80      90      100      110

```

```

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDPCPPVRVNGIGGLPGGKVKLSGSISS
      .  ::::  ::::::::::  ..::::::::  .  .  :::::  ...  ::::
genpep  FELSGVARMHERPIGDLVDALRQLGCSIEYLGNEGYPLRLRPAQLKIEEPIRVRGDVSS
      120      130      140      150      160      170

```

```

      190      200      210      220      230
2MEPSP QYxxxxxxxxxxxxxG-DVEIEIIDKLISIPYVENTLRLMERFGVKAH---SDSWDRFYI
      :  ..  .  .:  :::::  ::::  ::::::::::::::  ...  :::::  :
genpep  QFLTALLMALPLVAGQDIHIEVVGELISRPIEITLNLKRFGIHVQRPLGADNWQRFITI
      180      190      200      210      220      230

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

	240	250	260	270	280	290
2MEPSP	KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxx-----xxSLQGDVKFAEVL					

genpep	PAGSQYQSPGDIHVEGDASSASYFIALGAISEPAAGQNHIEILGVGADSIQGDIFIEAA					
	240	250	260	270	280	290
	300	310	320	330	340	350
2MEPSP	EMMGAKVTWTETSVTVT-GPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAI					

genpep	RMMGAQIESTPNSLRISRGKPGQG---PLKAIDLDCNHIPDAAMTLAVMALYADGTTTL					
	300	310	320	330	340	
	360	370	380	390	400	410
2MEPSP	RDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTIDTYDDHRMAMAF					

genpep	RNIASWRVKETDRIAAMACELQKLGAATVEEGADYDKITPPLAWKSAAIHTYDDHRIAMCF					
	350	360	370	380	390	400
	420	430	440			
2MEPSP	SLAAC--AEVPVTIRDPGCTRKTFFPDYFDVLSTFVKN					

genpep	SLAAFNPAPQPIRILDPKCVAKTFPDYFEALFAVSEAETAQIPVICIDGPTASGKGTLLA					
	410	420	430	440	450	460
genpep	LAAHRLGYHYLDGALYRLSAFAATRAGVALENGADVARIARTLPVFRFRGDRIFLGTEDEV					
	470	480	490	500	510	520

120606889 | *Acidovorax* sp. JS42, complete genome. | *Acidovorax* sp JS42 | AA | 673
Length = 673
initn: 834 initl: 238 opt: 870 Z-score: 1738.4 bits: 331.4 E(): 2.8e-88
Smith-Waterman score: 870; 46.429% identity (72.545% similar) in 448 aa overlap
(9-439:9-441)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLPKSGKSLSNRILLLAALSEGTTVDNLLNSEDVHYMLGAL					

genpep	MFSTAFLDLPLASVQGSVQLPGSKSISNRVLLLAALSQGTTEVRDLLASDDTRVMDLAL					
	10	20	30	40	50	60
	70	80	90	100	110	
2MEPSP	RTLGLSVEADKAAKRAV-VVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx					

genpep	RQLGCTV--DEAGGTVRITGLGSSTP---PSPTQLFMGNAGTAMRPLTAALALLGGE-					
	70	80	90	100	110	
	120	130	140	150	160	170
2MEPSP	xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPG----GKVKLS					

genpep	-YELSGVPRMHERPIGDLVDALRQLGCRIDYLGNGQGFPLRIAHAGGLPPLQLQAPIRVR					
	120	130	140	150	160	170

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      120      130      140      150      160      170
2MEPSP xxYVLGDGVPVRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVN--GIGGLPGGKVKLSG
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep --FKLSGVPRMHERPIGDLVDALRQLGCRIDYLGNEGFPPLHIRQPDFSQLPHSPIQVRG
      120      130      140      150      160      170

      180      190      200      210      220      230
2MEPSP SISSQYxxxxxxxxxxxxxG--DVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRF
      . : : : . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep DVSSQFLTSLLMALPLLANTRDITIEVVDELISKPYIHITLELLARFGI-AVHNENWQRF
      180      190      200      210      220      230

      240      250      260      270      280      290
2MEPSP YIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxx--SLQGDVKFAEVL
      : . : : : : . : : : : : : : : : . . : : : : : : : : : : : : : : : : :
genpep TIPAGSRYSSPGEIHVEADASSASYFIALGAIASDSAEGIRILGVGLDSIQGDIRFVEAA
      240      250      260      270      280      290

      300      310      320      330      340      350
2MEPSP EMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIR
      . : : : : . . . : : : : : : : : : : : : : : : : : : : : : : : : :
genpep RAMGAEVEGGPNWLRIRR-----GAWPLKAIDLDCNHIPDAAMTLAVMALYADGTTTLR
      300      310      320      330      340

      360      370      380      390      400
2MEPSP DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKL--NVT AIDTYDDHRMAM
      . : : : : : : : : : : : : : : : : : : : : . : : : : : : : : : : : :
genpep NIASWRVKETDRIAAMTKELRKLGADVEEGADYIRITPPASTAAWRAASIHTYDDHRVAM
      350      360      370      380      390      400

      410      420      430      440
2MEPSP AFSLAAC--AEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep CFSLAAFNPALPVRIEDPKCVAKTFPDYFEALFVSVETARARIPVICIDGPTASGKGTI
      410      420      430      440      450      460

genpep AAEVAQALGYHLLDSGALYRLTGLAASRAGLVLDDEPHAQQIAVLARELPVRFDAEQRIWL
      470      480      490      500      510      520

```

171193469 | *Polynucleobacter necessarius* STIR1, complete genome. | *Polynucleobacter necessarius* STIR1 | AA | 442
Length = 442
initn: 781 initl: 214 opt: 860 Z-score: 1721.1 bits: 327.6 E(): 2.6e-87
Smith-Waterman score: 860; 46.517% identity (69.213% similar) in 445 aa overlap
(1-436:1-434)

```

      10      20      30      40      50      60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSGKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : . : . : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep MSGLPDIQIGPFKQAQGSIVLPKSGKSISNRALLLAALSTGTTTLKNLLDADDTQVMRNAL
      10      20      30      40      50      60

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
:  :::::  .:  :  .:  :::::  :::::  .:  :::::  :::  :...  .
genpep  RQLGLSV-INKENKACIVEGCGGKLPVQEA----DLFMGNAGTAIRPLTAALAMQGGN--
              70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
:  :::  :::::  :::  :::::  .  :  :::::  .  :...  .  :::  :::::
genpep  YRLSGVARMRERPIRDLDGLRQVGAKIGYELQEGYPPIKILAADIQIKDVVKVRGDVSS
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxxxxxxGD-VEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKG-
:  .  .  .  .  .  :::::  :::  :::::  :::  .  :  .  :  :  .
genpep  QFLTALLMALPLVANEPVRIEVIGELISRPYIDITLKLMAFGVTVACPDM-QSFVIPAK
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  --GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
:  :::::  .  :::::  :::::  .  :  :::::  :::::  :::  :...  .
genpep  TSGAVYKSPGQLLVEGDASSASYFLALGAIGGGPVRVLGVGKESIQGDVAFADALALMGA
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  KVTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
..:  :  .  :::  ..  .  :::  ..  :::::  :::::  :::::  :::::  :::::
genpep  NITAGEDWIEVAGV-KNANGK--LNGITIDCTEIPDAAMTLAVAALFAEGPTRLNSIASW
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  RVKETERMVAIRTELTKLGASVEEGPDYCIITPPE----KLNVTATIDTYDDHRMAMAFS
:::::  :::::  :::  :::::  :::  .  :  :  :::::  :::::
genpep  RVKETDRIAAMAKELKKVGANVEEGADYIVVQAPVLQSDWKSPEGVDTYDDHRMAMCFS
              350          360          370          380          390          400

              420          430          440
2MEPSP  LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKV
:::  .  .  :::::  :::  :::
genpep  LAAFGLNALKINDPNCVAKTFPTYFAEFAKVVN
              410          420          430          440

```

145047091 | *Polynucleobacter* sp. QLW-P1DMWA-1, complete genome. | *Polynucleobacter* sp QLW-P1DMWA-1 | AA | 442
Length = 442
initn: 780 initl: 205 opt: 854 Z-score: 1709.1 bits: 325.4 E(): 1.2e-86
Smith-Waterman score: 854; 46.517% identity (69.663% similar) in 445 aa overlap (1-436:1-434)

```

              10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
:::  :::::  .  .  .  :::::  :::::  :::::  :::::  :::::  :  :::
genpep  MSGIPEITIGPFTRAQGSIVLPGSKSISNRALLLAALASGTTTLKNLLDADDTQVMRNAL
              10          20          30          40          50          60

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGC GGFVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :  ::::  ::  ::  :::::  ::  :::::  ::  ::  ::  ::  ::  ::  ::
genpep  RQLGLSV-IDQANHVCVVKGCGGQFPVREA---DLFMGNAGTAIRPLTAALAMQGGN--
              70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :  ::::  :::::  ::  :::::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
genpep  YRLSGVARMHERPIRDLVDGLRQVGAKIDYELQEGYPPIKILAADIQIKDVVKVRGDVSS
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxxxxxxGD-VEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGG
      :  .  .  .  .  .  ::::  ::  ::  :::::  ::  ::  ::  ::  ::  ::  ::
genpep  QFLTALLMALPLVAQEAVRIEVIGELISRPYIDITLKLMAFGVNVACPDQ-SFVIPA
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  QK---YKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      .  ::::  :  :::::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
genpep  TSDAVYKSPGNLSVEGDASSASYFLALGALGAGPVRVLGVGKDSIQGDVAFADALALMGA
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASW
      ::  :  .  ::  .  :  :  ::  .  .  :::::  ::  ::  ::  ::  ::  ::  ::
genpep  KITAGEDWIEVSGVKNA--GGK-LNGITLDCTEIPDAAMTLAVAALFAEGQTRLNNIASW
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLN-----VTAIDTYDDHRMAMAFS
      :::::  ::::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
genpep  RVKETDRIAAMAKELKKIGAIVEEGADYIVVQAPALGDWKSPSEGIDTYDDHRMAMCFS
              350          360          370          380          390          400

              420          430          440
2MEPSP  LAACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      ::  .  .  :::::  ::  ::  ::
genpep  LATLGPNTLTKINDPNCVAKTFPTYFAEFAKIVS
              410          420          430          440

```

115422436 | *Bordetella avium* 197N complete genome. | *Bordetella avium* 197N | AA | 439
Length = 439
initn: 635 initl: 230 opt: 853 Z-score: 1707.1 bits: 325.0 E(): 1.6e-86
Smith-Waterman score: 853; 45.805% identity (71.655% similar) in 441 aa overlap
(9-443:9-437)

```

              10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :  .  .  .  :::::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
genpep  MNALAYLDLPHIRQARGLAALPGSKSISNRVLLIAALAEGRTEISGLLDSDDTRVMLAAL
              10          20          30          40          50          60

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :  :::::  .:  .  :::::  .:  .:  :::::  :  .:  :::::  :::  :::::  .
genpep  RQLGVAV-TDLGQGRVAVEGAR-RFPAEKA---ELFLGNAGTAFRPLTAALALMGG--D
              70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :  :::::  :::::  :::  :::  .  .  :::::  :  :  .  .  :::::  :::
genpep  YRLSGVPRMHERPIGDLVDALRAWGARIDYLGQAGYPPLHI-GRGDIRADRVRVQGSVSS
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxxxxxxxG----DVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      :  .  .  .  .  .  :  :  :::  :::  :::::  :::  :::::  .  :::  :  :
genpep  QFLTALLLAAPIEAGASGRPVITIEVIGELISKPYIEITLNLMARYGVNVVR-DGWRAFTI
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      .:  .:  :::  .  :::::  :::::  .  .  .  .  .  .  .  .  .  .  .  .  .
genpep  EGDARYRSPGSIAVEGDASTASYLLALGVLGGGPVRVTGVGEQSIQGDATFADTLAAMGA
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      .:  .  .  .  .  .  :  :  :::  :::  :::::  :::  :::::  .  :::  :  :
genpep  NITKGSWDWIEASGQAVAEGGR--IKAFDADFNLIPDAAMTAATMALFADGPCRLRNIGSW
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLVN--TAIDTYDDHRMAMAFSLAA
      :::::  .:  .:  :::  :::  :::  .  .  .  .  .  .  .  .  .  .  .  .
genpep  RVKETDRIHAMHTELAKLGAKVESGPDWLSLTPPADSDWRDAHIGTWDDHRMAMCFSLAA
              350          360          370          380          390          400

              420          430          440
2MEPSP  CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .  .  :  :  :::  .  :::::  .  .  .
genpep  FGPAAVRILDPGCVSKTFPDYFDVYAGLVSA
              410          420          430

```

120590404 | *Acidovorax avenae* subsp. *citrulli* AAC00-1, complete genome.
 | *Acidovorax avenae* subsp. *citrulli* AAC00-1 | AA | 679
 Length = 679
 initn: 786 init1: 215 opt: 854 Z-score: 1706.2 bits: 325.5 E(): 1.8e-86
 Smith-Waterman score: 854; 45.190% identity (70.917% similar) in 447 aa overlap
 (9-439:9-443)

```

              10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPKSGKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
genpep  MYSTAFDLPLDPTAGGAVRLPGSKSISNRVLLLAALSEGTTVEVDLLASDDTRVMDLAL
              10          20          30          40          50          60

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGC GGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep  REIGCGVDEGAAAQGTVRITGLGTSP---ARSPSKLFLGNAGTAMRPLTAALALLGGE--
              70          80          90          100          110

              130          140          150          160          170
2MEPSP  YVLDGVPRMRERPIGDLVVLKQLGADVDCFLGTDCPPVRVNGIGGLP---GGKVKLSG
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep  FELSGVPRMHERPIGDLVEALLQLGCHISYLGNGFPPLRIAHAGGVPPLALDAPVRVRG
              120          130          140          150          160          170

              180          190          200          210          220          230
2MEPSP  SISSQYxxxxxxxx--xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRF
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep  DVSSQFLTALLMALPLVAREKDVVIEVVGELISRPIYIHITLQLLERFGIRVRH-DEWQRF
              180          190          200          210          220          230

              240          250          260          270          280
2MEPSP  YIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxx-----xxxxxxxxxSLQGDVKFAE
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep  TIPAGSRYRSPGTIHVEADASSASYFIALGALAAPAPGQEPLRILGVGLDSIQGDIRFAE
              240          250          260          270          280          290

              290          300          310          320          330          340
2MEPSP  VLEMMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPDVAMTLAVVALFADGPTA
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep  AARAMGA EVTGGPNWLVVRR-----GAWPLRAVDLDCNHIPDAAMTLAVMALYAQGT TT
              300          310          320          330          340

              350          360          370          380          390          400
2MEPSP  IRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK---LNVTAIDTYDDHRM
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep  LRNIASWRVKETDRIAAMAQGCRRLGATVEEGSDFLRVTTPASPADWRAASIHTYDDHRI
              350          360          370          380          390          400

              410          420          430          440
2MEPSP  AMAFSLAAC--AEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
genpep  AMCFSLAAFNPALPVRIEDPKCVAKTFPDYFEALFSVAGTPVEHVPVICIDGPTASGKG
              410          420          430          440          450          460

genpep  TVAAAVAQRLGYRFLDSGAMYRITALAALRAGLSIDA AHQDRIAALARTLPVRFESGRIW
              470          480          490          500          510          520

```

82411256 | *Nitrosospira multiformis* ATCC 25196, complete genome. | *Nitrosospira multiformis* ATCC 25196 | AA | 437
Length = 437
initn: 787 initl: 220 opt: 851 Z-score: 1703.1 bits: 324.3 E(): 2.6e-86
Smith-Waterman score: 851; 46.934% identity (71.934% similar) in 424 aa overlap (21-437:1-412)

	10	20	30	40	50	60
2MEPSP	MAGAEIIVLQPIKEISGTVKLPGSKSLSNRILL	LAALSEGTTVVDNLLNS	EDVHYMLGAL			
genpep		MPGSKSISNRILL	LAALSEGVT	DVCDLLASDDT	ARMLDAL	
		10	20	30	40	
	70	80	90	100	110	120
2MEPSP	RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSL	xxxxxxxxxxxxxx				
genpep	STLGVSI-LQIGRDHYRLQGVGDQFPLRLPTTEADLFLGNAGTVFRPLT	AMLALAQGH--				
	50	60	70	80	90	
	130	140	150	160	170	
2MEPSP	YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTD-CPPVRVNGIGGLPGGKVKLSGSIS					
genpep	YRLSGVPRMHERPIGDLVDALRQVGADIT-YLGKEGFPPLQIKP-GRIHPGEITVRGEVS					
	100	110	120	130	140	150
	180	190	200	210	220	230
2MEPSP	SQYxxxxxxxxxxxxxGDVE----	IEIIDKLISIPYVEMTLRLMERFGVKA	EHSDSWDRFY			
genpep	SQFLTALLMVLPLFLRAEMDELPVITVAGELISR	PYIDLTIALMARFGVQVER-EEWRRFT				
	160	170	180	190	200	210
	240	250	260	270	280	290
2MEPSP	IKGGQKYKSPKNAYVEGDASSASYFLAGAAI	xxxxxxxxxxxxxxSLQGDVKFAEVLEMMG				
genpep	VPADQRYRSPGQVFVEGDASSASYFLAAGAIGRGPVRVEGLGRDSVQGD	IRFAEALERMG				
	220	230	240	250	260	270
	300	310	320	330	340	350
2MEPSP	AKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKM	PDVAMTLAVVALFADGPTAIRDVAS				
genpep	ADIRFGDNWIEASGP-----GPGGLRAIDLDCNHIPDAAMTLAVTALFARGNTVLRNIAS					
	280	290	300	310	320	
	360	370	380	390	400	410
2MEPSP	WRVKETERMVAIRTELTKLGASVEEGPDYCIITTP--EKLNVTAIDTYDDHRMAMAFSLA					
genpep	WRVKETDRIAAMAQELRKLGAEVEAGSDFLQISPPRGELVANAAIDTYDDHRMAMCFSLV					
	330	340	350	360	370	380
	420	430	440			
2MEPSP	ACAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN					
genpep	SFG-APVRINDPRCVSKTFPDYFEKF	AAIAYADPGQGKFAARIDSSDIS				
	390	400	410	420	430	

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

48526080 | *Nymphaea alba* 5-enol-pyruvylshikimate-phosphate synthase (EPSPS) mRNA, partial cds. | *Nymphaea alba* | AA | 264
Length = 264
initn: 838 initl: 757 opt: 848 Z-score: 1700.4 bits: 323.1 E(): 3.7e-86
Smith-Waterman score: 848; 65.098% identity (82.745% similar) in 255 aa overlap (60-313:1-255)

	30	40	50	60	70	80
2MEPSP	RILL	LAAL	SEGT	TVVD	NLLN	SEDVHYMLGALRTLGLSVEADKAAKRAVVVGCGKFPV-E
				 :... :.....
genpep				LRTL	GLRV	DEDRDMKRAIVEGCSGQFPVAK
				10	20	30

	90	100	110	120	130	140
2MEPSP	DAKEEVQLFLGNAGIAMRSL	xxxxxxxxxxxxx	YVLD	GVPR	MRERPI	IGDLVVGLKQLGADV

genpep	DSAKEVELFLGNAGTAMRPLTA	AAVVAAGGNTRYILD	GVPR	MRERPI	IGDLV	SGLQLGADI
	40	50	60	70	80	90

	150	160	170	180	190	200
2MEPSP	DCFLGTDCPPVRVNGIGGLPGGKVKLSGS	ISSQYxxxxxxxxxxxxx	GDVEIEIIDKLISI			
	:
genpep	GCTLTGNCPPVYINGKGG	LPGGKVKLSGS	ISSQYLTALLMAAPLALGDVEVEMADKLVS			
	100	110	120	130	140	150

	210	220	230	240	250	260
2MEPSP	PYVENTLRLMERFGVKA	EHSDSWDRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAIxx				

genpep	PYVENTLKLMERFGVAVEHGGG	WDRFLIRGRQMYMSPGSAYVEGDASSASYFLAGAAITG				
	160	170	180	190	200	210

	270	280	290	300	310	320
2MEPSP	xxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTWTETSVTVTGPPREPFG	GRKHLKAIDVNMN				

genpep	GTVTVEGCGTSSSLQGDVKFAEVLEKMGAKVTWSQNSVTVTGPPKDGSRTRLRDI					
	220	230	240	250	260	

	330	340	350	360	370	380
2MEPSP	KMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT					

163259926 | *Bordetella petrii* strain DSM 12804, complete genome. | *Bordetella petrii* | AA | 447
Length = 447
initn: 755 initl: 230 opt: 847 Z-score: 1694.9 bits: 322.8 E(): 7.5e-86
Smith-Waterman score: 847; 45.880% identity (69.710% similar) in 449 aa overlap (1-443:1-438)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVL	QPIKEIS	GTVKLP	SGSKSL	SNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL

genpep	MTGQPYLDLPRARQARGQ	VALPGSKSISNRVLLLAALAAGRTDISGLLDSD	DDTRV	MLAAL		
	10	20	30	40	50	60

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
genpep  RQLG--VELAEAGEGRVTVGGAGRFPVKQA---DLFLGNAGTAFRPLTAALALMGG--Q
              70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
genpep  YRLSGVPRMHERPIGDLVDALRQWGARIDYLGQAGYPPLAVGEGRIRADAPARVQGAVSS
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxxx----xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
genpep  QFLTALLLAAPVLAQGS DRPVVIEVAGELISKPYIEITLNLMARYGVQVRR-DGWRTFTI
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
genpep  EPGAAYRSPGAIAVEGDASSASYFLALGAIGGGPVRVTGVGADSIQGDVAFARTLADMGV
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
genpep  QIDYGPDWIEARGVRVDQGR--LKAFTDTFNLIPDAAMTAAALALYADGPCRLRNIGSW
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
genpep  RVKETDRIHAMHTELAKLGAEVESGPDWLRITPPADGGWRDAHIGTWDDHRMAMCFSLAA
              350          360          370          380          390          400

              420          430          440
2MEPSP  CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
genpep  FGPAAVRILDPGCVSKTFPDYFDVYAGLVSGAPDSYDD
              410          420          430          440

```

33571749 | *Bordetella pertussis* strain Tohama I, complete genome; segment 3/12.
 | *Bordetella pertussis* Tohama I | AA | 442
 Length = 442
 initn: 605 initl: 217 opt: 845 Z-score: 1691.0 bits: 322.0 E(): 1.2e-85
 Smith-Waterman score: 845; 45.495% identity (69.595% similar) in 444 aa overlap
 (1-438:1-433)

```

              10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
genpep  MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGST EITGLLDSDDTRVMLAAL
              10          20          30          40          50          60

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGC GGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      :  ::::  ..:  ....  .....:  .....:  ::  ....
genpep  RQLGVSV--GEVADGCVTIEGVARFPTEQA---ELFLGNAGTAFRPLTAALALMGGD--
              70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :  :::::  :::::  :::::  ..  .  .  :::::  .  :  :::::
genpep  YRLSGVPRMHERPIGDLVDALRQFGAGIEYLQQAGYPPLRIGGGSIRVDGPVRVEGSVSS
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxxx----xxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      :  ..  ..  .  :  ....  ....  :::::  :::::  :::::  :::::
genpep  QFLTALLMAAPVLARRSGQDITIEVVGELISKPYIEITLNLMARFGVSVRR-DGWRAFTI
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      :  :  .  :::::  :::::  :  .  :  :::::  ::  :  :  :
genpep  ARDAVYRGPGRMAIEGDASTASYFLALGAIGGPPVRVTGVGEDSIQGDVAFAATLAAMGA
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      :  .  .  :  :  :::::  :::::  :::::  :::::  :::::
genpep  DVRYGPGWIETRGRVVAEGGR--LKAFDADFNLIIPDAAMTAATLALYADGPCRLRNIGSW
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      :::::  :  :::::  :::::  :  .  :  .  :  :::::  :::::
genpep  RVKETDRIHAMHTELEKLGAGVQSGADWLEVAPPEPGGWRDAHIGTWDDHRMAMCFSLAA
              350          360          370          380          390          400

              420          430          440
2MEPSP  CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .  .  :  :  :::::  :::::
genpep  FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
              410          420          430          440

33574090 | Bordetella parapertussis strain 12822, complete genome; segment 10/14.
| Bordetella parapertussis | AA | 442
Length = 442
initn: 600 init1: 212 opt: 843 Z-score: 1686.9 bits: 321.3 E(): 2.1e-85
Smith-Waterman score: 843; 45.270% identity (69.369% similar) in 444 aa overlap
(1-438:1-433)
              10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      :  :  .  :  .  :  :  :::::  :::::  :::::  :::::  :::::
genpep  MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGSTAITGLLDSDDTRVMLAAL
              10          20          30          40          50          60

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

              70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGC GGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      :  ::::  ..:  ....  :::::  :::::  :::  ....
genpep  RQLGVSV--GEVADGCVTIEGVARFPIEQA---ELFLGNAGTAFRPLTAALALMGGD--
              70          80          90          100          110

              130          140          150          160          170          180
2MEPSP  YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      :  :::::  :::::  :::::  ..  .  .  :::::  .  :  :::::
genpep  YRLSGVPRMHERPIGDLVDALRQFGAGIEYLGQAGYPPLRIGGGSIRVDGPVRVEGSVSS
              120          130          140          150          160          170

              190          200          210          220          230
2MEPSP  QYxxxxxxxxx----xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      :  ..  ..  .  :  ....  ::::  :::::  :::::  :::  :
genpep  QFLTALLMAAPVLARRSGQDITIEVVELISKPYIEITLNLMARFGVSVRR-DGWRAFTI
              180          190          200          210          220          230

              240          250          260          270          280          290
2MEPSP  KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      :  :  .  :::::  :::::  :  .  :  ::::  :  :  :
genpep  ARDAAYRGPGRMAIEGDASTASYFLALGAIGGPPVRVTGVGEDSIQGDVAFAATLAAMGA
              240          250          260          270          280          290

              300          310          320          330          340          350
2MEPSP  KVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASW
      :  .  .  :  :  :::::  :::::  :::::  :::::  :
genpep  DVRYGPGWIETRGRVVAEGGR--LKAFDADFNLIPDAAMTAATLALYADGPCRLRNIGSW
              300          310          320          330          340

              360          370          380          390          400          410
2MEPSP  RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      :::::  :  :::::  :::::  :  ....  .  :  :::::
genpep  RVKETDRIHAMHTELEKLGAGVQSGADWLEVAPPAPGGWRDAHIGTWDDHRMAMCFSLAA
              350          360          370          380          390          400

              420          430          440
2MEPSP  CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      .  .  :  :  ::::  :::::
genpep  FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
              410          420          430          440

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

6013215 | *Bordetella bronchiseptica* prephenate dehydratase (pheA) gene, partial cds; prephenate dehydrogenase (tyrA), 5-enolpyruvylshikimate-3-phosphate synthase (aroA), and cytidine monophosphate kinase (cmk) genes, complete cds; and ribosomal protein S1 (rpsA) gene, partial cds. | *Bordetella bronchiseptica* | AA | 442

Length = 442

initn: 600 initl: 212 opt: 841 Z-score: 1682.9 bits: 320.6 E(): 3.5e-85

Smith-Waterman score: 841; 45.495% identity (69.595% similar) in 444 aa overlap (1-438:1-433)

	10	20	30	40	50	60
2MEPSP	MAGAE	EIVLQ	PIKEIS	GTVKLP	GSKSL	SNRILL
	..	.	:	:	:	:
genpep	MSGLA	YLDLP	PAARL	ARGEVA	LP	PGSKS
	10	20	30	40	50	60
	70	80	90	100	110	120
2MEPSP	RTLGL	SVEAD	KA	KRAV	VVGCG	GKFP
	:	:	:	:	:	:
genpep	RQLGV	SV-GE	VADGR	VTIE	GV	-RFPTE
	70	80	90	100	110	
	130	140	150	160	170	180
2MEPSP	YVLDG	VPRMR	ERPIG	DLVVG	LKQLG	ADVDC
	:	:	:	:	:	:
genpep	YRLSG	VPRMH	ERPIG	DLVD	ALRQF	GAGIE
	120	130	140	150	160	170
	190	200	210	220	230	
2MEPSP	QYxxxx	xxxx	GDVE	IEI	IDKL	ISIPY
	:	:	:	:	:	:
genpep	QFLT	TALL	MAAP	VLARR	SGQD	ITIEV
	180	190	200	210	220	230
	240	250	260	270	280	290
2MEPSP	KGGQ	KYKSP	KNAY	VEGD	ASSA	SYFLA
	:	:	:	:	:	:
genpep	ARDA	AYRGP	GRMA	IEGD	ASTA	SYFLA
	240	250	260	270	280	290
	300	310	320	330	340	350
2MEPSP	KVTWT	ETSV	TGTG	PPREP	FGRK	HLKA
	:	:	:	:	:	:
genpep	DVRY	GPGW	IETRG	VRVA	EGGR	--LKA
	300	310	320	330	340	
	360	370	380	390	400	410
2MEPSP	RVKET	ERMVA	IRTE	LTKL	GASV	EEGPD
	:	:	:	:	:	:
genpep	RVKET	DRIH	AMHTE	LEKLG	AGVQ	SGADW
	350	360	370	380	390	400

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

          420          430          440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : . : : : : :
genpep FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
          410          420          430          440

```

33576881 | *Bordetella bronchiseptica* strain RB50, complete genome; segment 11/16.
 | *Bordetella bronchiseptica* RB50 | AA | 442
 Length = 442
 initn: 600 initl: 212 opt: 840 Z-score: 1680.9 bits: 320.2 E(): 4.5e-85
 Smith-Waterman score: 840; 45.270% identity (69.369% similar) in 444 aa overlap
 (1-438:1-433)

```

          10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : . : . : . : : : : : : : : : : : : : : : : : : : : : : :
genpep MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGSTEITGLLDSDDTRVMLAAL
          10          20          30          40          50          60

          70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      : : : : : . : . : . : : : : : : : : : : : : : : : .
genpep RQLGVSV--GEVADGCVTIEGVARFPTEQA---ELFLGNAGTAFRPLTAALALMGGD--
          70          80          90          100          110

          130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : . . . . : : : : : : : : : : : :
genpep YRLSGVPRMHERPIGDLVDALRQFGAGIEYLGQAGYPPLRIGGGSIRVDGPVRVEGSVSS
          120          130          140          150          160          170

          190          200          210          220          230
2MEPSP QYxxxxxxxxx----xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      : . . . . . : . : . . : : : : : : : : : : : : : : : : : :
genpep QFLTALLMAAPVLARRSGQDITIEVVGELISKPYIEITLNLMARFGVSVRR-DGWRAFTI
          180          190          200          210          220          230

          240          250          260          270          280          290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      : . : . : : : : : : : : : : . : : : : : : : : : : : : : :
genpep ARDAAYRGPGRMAIEGDASTASYFLALGAIGGPVRVTGVGEDSIQGDVAFAATLAAMGA
          240          250          260          270          280          290

          300          310          320          330          340          350
2MEPSP KVTWTETSVTVTGPPEPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      : . . . : : : : : : : : : : : : : : : : : : : : : : :
genpep DVRYGPGWIEITRGVRVAEGGR--LKAFDADFNLIIPDAAMTAATLALYADGPCRLRNIGSW
          300          310          320          330          340

          360          370          380          390          400          410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      : : : : . : : : : : : : : : . . : : : : : : : : : : : :
genpep RVKETDRIHAMHTELEKLGAGVQSGADWLEVAPPAPGGWRDAHIGTWDDHRMAMCFSLAA
          350          360          370          380          390          400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                420          430          440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : . : : : : :
genpep FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
      410          420          430          440

144037 | B. pertussis 5-enolpyruvylshikimate-3-phosphate synthase (aroA) gene,
complete cds. | Bordetella pertussis | AA | 442
Length = 442
initn: 605 initl: 217 opt: 839 Z-score: 1678.9 bits: 319.8 E(): 5.9e-85
Smith-Waterman score: 839; 45.270% identity (69.369% similar) in 444 aa overlap
(1-438:1-433)

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : . : . : : : : : : : : : : : : : : : : : : : : : : : : :
genpep MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGSTEITGLLDSDDTRVMLAAL
      10          20          30          40          50          60

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : : . : . : . : : : : : : : : : : : : : : .
genpep RQLGVSV--GEVADGCVTIEGVARFPTEQA---ELFLGNAGTAFRPLTAALALMGGD--
      70          80          90          100          110

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : . . . : : : : : : : : : : : :
genpep YRLSGVPRMHERPIGDLVDALRQFGAGIEYLGQAGYPPLRIGGGSIRVDGPVRVEGSVSS
      120          130          140          150          160          170

                190          200          210          220          230
2MEPSP QYxxxxxxxxx----xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYI
      : . . . . . : . : . . : : : : : : : : : : : : : : : : : :
genpep QFLTALLMAAPVLARRSGQDITIEVVELISKPYIEITLNLMARFGVSVRR-DGWRAFTI
      180          190          200          210          220          230

                240          250          260          270          280          290
2MEPSP KGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGA
      : . : . : : : : : : : : : : . : : : : : : : : : : : : : :
genpep ARDAVYRGPGRMAIEGDASTASYFLALGAIGGPPVRVTGVGEDSIQGDVAFAATLAAMGA
      240          250          260          270          280          290

                300          310          320          330          340          350
2MEPSP KVTWTETSVTVTGPPEPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASW
      : . . . : : : : : : : : : : : : : : : : : : : : : : : :
genpep DVRYGPGWIETRGVRVAEGGR--LKAFDADFNLIIPDAAMTAATLALYADGPCRLRNIGSW
      300          310          320          330          340

                360          370          380          390          400          410
2MEPSP RVKETERMVAIRTELTKLGASVEEGPDYCIITPPEK--LNVTAIDTYDDHRMAMAFSLAA
      : : : : . : : : : : : : : : . . : : : : : : : : : : : :
genpep RVKETDRIHAMHTELEKLGAGVQSGADWLEVPPEPGGWRDAHIGTWDDHRMAMCFLLA
      350          360          370          380          390          400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                420          430          440
2MEPSP CAEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . . : : : : : : : : : : : : : :
genpep FGPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
        410          420          430          440

124260205 | Methylibium petroleiphilum PM1, complete genome. | Methylibium
petroleiphilum PM1 | AA | 674
Length = 674
initn: 847 initl: 268 opt: 831 Z-score: 1660.0 bits: 316.9 E(): 6.6e-84
Smith-Waterman score: 831; 44.118% identity (71.267% similar) in 442 aa overlap
(9-439:10-438)

                10          20          30          40          50
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGA
      . . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep  MATPRAFLDIPPLQAAGGTVRLPGSKSISNRVLLLAGLCAGRTRVLDLLDSDDTQVMLDA
        10          20          30          40          50          60

                60          70          80          90          100          110
2MEPSP  LRTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      : : : : . . . : : : : : : : : : : : : : : : : : : : : :
genpep  LRALGCDIETDGAAR--VVTGLGGRLAVREAR----LFLGNAGTAMRPLAALALLAADQ
        70          80          90          100          110

                120          130          140          150          160          170
2MEPSP  x--YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGS
      . : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep  GGRFELSGVPRMHERPIGDLVDALRPLGCTITCLANEGYPPLRLERGTKLDAPIRVRGD
        120          130          140          150          160          170

                180          190          200          210          220          230
2MEPSP  ISSQYxxxxxxxxxxxxG--DVEIEIIDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFY
      : : : : . . . . : : : : : : : : : : : : : : : : : : : :
genpep  VSSQFLTALLMALPLVAARQSITIEVDGELISKPYIEITLALLARFGISVQR-EGWQRFV
        180          190          200          210          220          230

                240          250          260          270          280          290
2MEPSP  IKGGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxx--xxxxxxxSLQGDVKFAEVLEM
      : . : : : . : : : : : : : : : : . : : : : : : : :
genpep  IPQGSAYRSPGEIAVEGDASSASYFIAAGAAIAAADTPLRIEGVGSASIQGDIRFVEAARA
        240          250          260          270          280          290

                300          310          320          330          340          350
2MEPSP  MGAKVTWTETSVTVTGPPEPFGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDV
      : : . : : : : : : : : : : : : : : : : : : : : : :
genpep  MGADITEEANALVVRR-----GAWPLTAITLDCNHIPDAAMTLAAMALYATGTTRLTNI
        300          310          320          330          340

                360          370          380          390          400          410
2MEPSP  ASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVTATIDTYDDHRMAMAFSLA
      : : : : : : : : : : : : : : : : : : : : : : : :
genpep  ASWRVKETDRIAAMAIELRKFGATVLEGTDFIEVTPPARWQAAAIHTYDDHRMAMCASLA
        350          360          370          380          390          400

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                420          430          440
2MEPSP AC-----AEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      :      . . . . . : : : . . . . . : . . . . .
genpep AFNPLAGGDVPVRILDPKCNKTFPAYFDALFGVTRARTDRVPVLTVDGPTASGKGTLS
      410          420          430          440          450          460

genpep ALAERLGYHHLDGSGALYRATALAALRQGVPADEAAVAAIARALPLRFENQQTLLAGEDV
      470          480          490          500          510          520

```

33504186 | *Blochmannia floridanus* complete genome. | *Candidatus Blochmannia floridanus* | AA | 433
Length = 433
initn: 773 initl: 286 opt: 818 Z-score: 1636.8 bits: 312.0 E(): 1.3e-82
Smith-Waterman score: 818; 43.187% identity (70.670% similar) in 433 aa overlap
(7-437:5-424)

```

                10          20          30          40          50          60
2MEPSP MAGAEEIVLQPIKEISGTVKLPKSKLSNRILLAAALSEGTTVVDNLLNSEDVHYMLGAL
      : . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep MENFITLNPICKINGTIYLPKSKLSNRALLAAQSVGATRLINLLDSDDTRYMLAAL
      10          20          30          40          50

                70          80          90          100          110          120
2MEPSP RTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep LQLGVKYKLSTDHKICEIKGVGGS--LQSKNTQTTLFLGNAGTAIRPLIAALSISKSHNI-
      60          70          80          90          100          110

```

```

                130          140          150          160          170          180
2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : . . . . . : : : . . . . . : . . . . . : . . . . . : . . . . .
genpep -VLTGDMRMKERPIFHLVDTLRQGGRIEYIERDHHLPVKL--YGGYCGGNIVIKGDISS
      120          130          140          150          160          170

```

```

                190          200          210          220          230          240
2MEPSP QYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
      : . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep QFLSAVLMMTPLASKNTYIEVIGSLVSKPYIDITLSVMRAFGICVQHDKDYKFFYCEGNR
      180          190          200          210          220          230

```

```

                250          260          270          280          290          300
2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
      : . . . . . : . . . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep IYKAPQEYVIEGDASSASYFLAAAIAIKGGTIRVLGVGKNSKQGDIFANILERMGSIIITW
      240          250          260          270          280          290

```

```

                310          320          330          340          350
2MEPSP TETSVTVTGPPREPFGRKHLKAIDVNMNKMMPDVAMTLAVVALFADG--PTAIRDVASWRV
      . . . : . . . . . : . . . . . : . . . . . : . . . . .
genpep GDNYI-----ECSKGIMLSIDIDVNDIPDAAMTLAVIALFTTNNLPMILRNIYNWRV
      300          310          320          330          340

```

	360	370	380	390	400	410
2MEPSP	KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT	AIDTYDDHRMAMAFSLAACAEV				
	:::::	: :	:::	:	:	:
genpep	KESDRLHAMATELRKVGAIVSEGYDYLHIVPPVQIQSAFINTYNDHRIAMCFALVALSNV					
	350	360	370	380	390	400
	420	430	440			
2MEPSP	PVTIRDPGCTRKTFFPDYFDVLSTFVKN					
	:::::	: :	:::::			
genpep	SVTINNPKCVCKTFPDDFDRFLSVGNNL					
	410	420	430			

Smith-Waterman score: 812; 40.509% identity (70.139% similar) in 432 aa overlap (5-436:3-419)

```

      10      20      30      40      50      60
2MEPSP  MAGAEEIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      . . . . . : : : : : : : : : : : : : : : : : : : : : : : :
genpep  MQDSLTLKPVVDYIQGKINIPGSKSISNRVLLLSALSNGKTILKNLLYSDDIKYMLKAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP  RTLGLSVEADKAAKRAVVVGC GGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep  LKLGIFYKLDKKKSKCTIYGISDAFSV---KNKIKFLGNAGTAMRPLLAILSLKKNKI-
      60      70      80      90      100      110

```

130 140 150 160 170 180
 2MEPSP YVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
 . : : :::::: :: .::: ::.. :: .. ::: :::. :::::
 genpep -ILTGEKRMKERPIHHLVDSLRQGGANITYKNKKKFPPLYIK--GGFKGGKIFIDGSISS
 120 130 140 150 160 170

190 200 210 220 230 240
 2MEPSP QYxxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKGGQ
 : : : : . . . : : : : : : : : : : . . . : : : : :
 genpep QFLSSLLMAAPLAELDT E I I V K N Q L V S K P Y I N L T I N L M E K F G I S V S I L N D Y K H F Y I K G N Q
 180 190 200 210 220 230

2MEPSP KYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKVTW
:: :::: :.:
genpep KYISPKKYIIESDLSSATYFLAAAAIKGGSIQINGIQKKSIOGDINFIKILKQMGSVSIQW

240 250 260 270 280 290

310 320 330 340 350 360
 2MEPSP TETSVTVTGPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRVKE
 ..:: . : . : . : . : . : . : . : . : . : . : . :
 genpep KKNSVICK-----KNKLLGITVDCNHIPDAAMTIAILGVFSKKKVYIKNIYNWRVKE
 300 310 320 330 340

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                370          380          390          400          410          420
2MEPSP  TERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT AIDTYDDHRMAMAFSLAACA E V P V
      . . . . . : : : : : : : : : : . : . . . . . : : : : : : : : : : . . . :
genpep  TDRIYAMSTELKKIGARVITGKDYIKVYPVKNF I HAKINTYNDHRIAMCFSLISLSGTSV
                350          360          370          380          390          400

                430          440
2MEPSP  TIRDPGCTRKTFPDYFDVLSTFVKN
      . . . : : : : : : : :
genpep  TLLNPKCVNKT FPSFFKNFY SICHYSNINKNI
                410          420          430

```

89345096 | *Rhodoferax ferrireducens* T118, complete genome. | *Rhodoferax ferrireducens* T118 | AA | 669

Length = 669

initn: 533 initl: 239 opt: 802 Z-score: 1601.7 bits: 306.1 E(): 1.2e-80

Smith-Waterman score: 802; 43.468% identity (72.072% similar) in 444 aa overlap (9-439:9-436)

```

                10          20          30          40          50          60
2MEPSP  MAGAEEIVLQPIKEISGTVKLP GSKSLSNRILL L AALSEGTTVVDNLLN SEDVHYMLGAL
      . . . : : : : : : : : : : : : : : : : : : : : : . . . : : : : :
genpep  MFSTAFLDIPPLTGVSGTVVLP GSKSISNRVLL SALSAGRTTLIDLLSD DTRVMFEAL
                10          20          30          40          50          60

                70          80          90          100          110          120
2MEPSP  RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : . . . . . : : : . . . : : : : : : : : : : : : : : . . .
genpep  RALGCGVR--QSGSTVEIEGLGGQL---GHRQAALFMGNAGTAIRPLTAALAVLGGD--
                70          80          90          100          110

```

```

                130          140          150          160          170
2MEPSP  YVL DGVPRMRERPIGDLVVG LKQLGADVDCFLGTD-CPPVRVNGIGGLPGGKVKLSGSIS
      . : : : : : : : : : : : : : : : : : : : : : : : : . . . : : :
genpep  FELRGVPRMHERPIGDLVDALRQIGCHID-YLGQDGFPLHIGRPSLKLDTPIRVRGDVS
                120          130          140          150          160          170

```

```

                180          190          200          210          220          230
2MEPSP  SQYxxxxxxxx--xxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKA EHS DSWDRFYIK
      : : . . . . . : : : : : : : : : : : : : : : . : : : : : : :
genpep  SQFLTALLMALPLVAATRDIVIEVVGELISRPYIEITLNL LARFNVRVVR-DGWQRFTIP
                180          190          200          210          220          230

```

```

                240          250          260          270          280          290
2MEPSP  GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxx-----xxSLQGDVKFAEVLE
      . . . . . : : : : : : : : : : . . . . . : : : : : : : : . .
genpep  AASQFRSPGTLFVEADASSASYFIAAGAIATGAGGQKGLRIEGVGADSIQGDIRFIEAAQ
                240          250          260          270          280          290

```

```

                300          310          320          330          340          350
2MEPSP  MMGAKVTWTETSVTVTGPPREPFGRKHLKAIDVNMNKM P DVAMTLAVVALFADGPTAIRD
      : : : : : . . . . : : : : : : : : : : : : : : : : : : : :
genpep  MMGAQVQSGPNWLEVSR-----GCWPLKAIDLDCNHIPDAAMTLAMMALYAQGSTTLRN
                300          310          320          330          340

```


2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

                360          370          380          390          400
2MEPSP VASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKL--NVT AIDTYDDHRMAMA
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep IASWRVKETDRIAAMACELRKLGATVEEGADFIRVTPPAVAADWRAASIHTYDDHRMAMC
                350          360          370          380          390          400

                410          420          430          440
2MEPSP FSLAAC--AEVPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep FSLAAFNPAGLPVRIVDPKCVAKTFPDYFEALFSLAQ TSAQAIPVICVDGPTASGKGTLA
                410          420          430          440          450          460

genpep GAVAQALGYHLLDSGSLYRITALAALQAGLALEFAHEQAIADLIPGLSIRFSQDQVFLNG
                470          480          490          500          510          520

```

7226672 | *Neisseria meningitidis* MC58, complete genome. | *Neisseria meningitidis* MC58 | AA | 433
Length = 433
initn: 804 initl: 178 opt: 799 Z-score: 1598.6 bits: 304.9 E(): 1.7e-80
Smith-Waterman score: 799; 44.344% identity (70.136% similar) in 442 aa overlap (6-444:3-430)

```

                10          20          30          40          50
2MEPSP MAGAEEIVLQPIKEIS-GTVKLP GSKSLSNRILL LLAALSEGTTVVDNLLNS EDVHYMLGA
      : : : . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep MTESVRLPVARLKPSTVALPGSKSISNRTL LLAALSDNACEIHSLLKSDDTDRMLEA
                10          20          30          40          50

                60          70          80          90          100          110
2MEPSP LRTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep LDKLGVQIEY-LAEDRLKVHGTGGRFPNRTA---DLFLGNAGTAFRPLTAALAVLGGD-
                60          70          80          90          100          110

                120          130          140          150          160          170
2MEPSP xYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTD-CPPVRVNGIGGLPGGKVKLSGSI
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep -YHLHGVPRMHERPIGDLVDALRIAGADVE-YLGKEHYPPPLHIGERQDNGERVIPIKGNV
                120          130          140          150          160

                180          190          200          210          220          230
2MEPSP SSQYxxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG
      : : . . . . : : . . : : : : : : : : : : : : : : : : : : : : : :
genpep SSQFLTALLMALPLTGQAFEIRMV GELISKPYIDITLKLMAQFGVQVIN-EGYRVFKIPA
                170          180          190          200          210          220

                240          250          260          270          280          290
2MEPSP GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
      : . . . : : : : : : : : . . . . : : : : : : : : : : : : : : :
genpep DAHYHAPEHLHVEGDASSASYFLAAGLIAATPVRVTGIGANSIQGDVAFARELEKIGADV
                230          240          250          260          270          280

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      300      310      320      330      340      350
2MEPSP TWTETSVTVTGPPREPFGFRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWRV
      . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
genpep VWGENFVEVSRPKE-----RAVQSFDDLDAHNPDAAMTLAIVALATGQTCTLRNIGSWRV
      290      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAACAE
      . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
genpep KETDRIAAMANELRKLGAQVVEEAEAIHITPPTLTPDAVIDTYDDHRMAMCFSLVSLLG
      350      360      370      380      390      400

      420      430      440
2MEPSP VPVTIRDPGCTRKTFPDYFDVLSTFVKN
      . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
genpep VPVVINDPKCTHKTFTPTFYFDVFSSTLTETAE
      410      420      430

```

120866826 | *Neisseria meningitidis* serogroup C FAM18 complete genome. | *Neisseria meningitidis* FAM18 | AA | 433

Length = 433

initn: 797 initl: 175 opt: 794 Z-score: 1588.5 bits: 303.1 E(): 6.3e-80

Smith-Waterman score: 794; 43.891% identity (70.136% similar) in 442 aa overlap (6-444:3-430)

```

      10      20      30      40      50
2MEPSP MAGAEFIVLQPIKEIS-GTVKLPGSKSLSNRILLALLAALSEGTTVVDNLLNSDEVHYMLGA
      : : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
genpep MTESVRLPVARLKPSTVALPGSKSISNRTLLAALSDNACEIHSLLKSDDTDRMLEA
      10      20      30      40      50

      60      70      80      90      100      110
2MEPSP LRTLGLSVEADKAAKRAVVVGCGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
genpep LDKLGVQIEY-LAEDRLKVHGTGGRFPNRTA---DLFLGNAGTAFRPLTAALAVLGGD-
      60      70      80      90      100      110

      120      130      140      150      160      170
2MEPSP xYVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTD-CPPVRVNGIGGLPGGKVKLSGSI
      : : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
genpep -YHLHGVPRMHERPIGDLADALRIAGADVE-YLGKEHYPLHIGKRQDNGERVPIKGNV
      120      130      140      150      160

      180      190      200      210      220      230
2MEPSP SSQYxxxxxxxxxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIKG
      : : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
genpep SSQFLTALLMALPLTGQAFEIRMVGEISKPYIDITLKLMAQFGVQVAN-ENYRVFKIPA
      170      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP GQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAKV
      : : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
genpep DAHYHAPPEHLHVEGDASSASYFLAAGLIAATPVRVTGIGANSIQGDVAFARELEKIGADV
      230      240      250      260      270      280

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      300      310      320      330      340      350
2MEPSP TWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMMPDVAMTLAVVALFADGPTAIRDVASWRV
      . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :
genpep VWGENFVEVSRPKE-----RAVQSFDDLDAHNPDAAMTLAIVALATGQTCTLRNIGSWRV
      290      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP KETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT- IDTYDDHRMAMAFSLAACAE
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep KETDRIAAMANELRKLGAKEVEEAEAIHITPPKTLTPDAVIDTYDDHRMAMCFSLVSLLG
      350      360      370      380      390      400

      420      430      440
2MEPSP VPVTIRDPGCTRKTFPDYFDVLSTFVKN
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep VPVVINDPKCTHKTFTPTFYFDVFSSLTETTE
      410      420      430

```

72119670 | *Ralstonia eutropha* JMP134 chromosome 1, complete sequence. | *Ralstonia eutropha* JMP134 | AA | 434
Length = 434
initn: 892 initl: 224 opt: 794 Z-score: 1588.5 bits: 303.1 E(): 6.3e-80
Smith-Waterman score: 945; 48.064% identity (73.576% similar) in 439 aa overlap (5-438:2-428)

```

      10      20      30      40      50      60
2MEPSP MAGAEFIVLQPIKEISGTVKLPKSGKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep MEHLTLGLPLTRAAGTVRLPGSKSISNRVLLLAALAGGETVRDLDSDTRVMLQAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxxx
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep KTLGVAWRQE--GSDYIVTGSGGNFPVKAA---ELFMGNAGTAIRPLTAALALQGGD--
      60      70      80      90      100

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGKLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep YKLSGVPRMHERPIGDLVDGLRQVGADIGYLANEGFPPLHIRPAQIRIDAPIRVRGDVSS
      110      120      130      140      150      160

      190      200      210      220      230
2MEPSP QYxxxxxxxx--xxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
      : . . . . : : : : : : : : : : : : : : : : : : : : : : :
genpep QFLTALLMTLPMAQAANGKIEIEVVGELISKPYIEITLNLARFGINVERQ-GWERFIVP
      170      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      . : : : . : : : : : : : : : : : : : : : : : : : : : :
genpep AGAVYRSPGEIYVEGDASSASYFLAAGAIGGGPVRVEGVGMSSIQGDVRF AEALNRMGAN
      230      240      250      260      270      280

```

2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS

```

      300      310      320      330      340      350
2MEPSP VTWTETSVTVTGPPREPFGGRKHLKAIDVNMNKMPPDVAMTLAVVALFADGPTAIRDVASWR
      :   . . . : :   : :   :   . . . . :   . . . . . . . . . :   . . . . .
genpep VMAGDNWIEVRGTERDD-GRLH--GIELDCNHIPDAAMTLAVAALFAGGTTTTLTNIASWR
      290      300      310      320      330      340

      360      370      380      390      400      410
2MEPSP VKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNVT--IDTYDDHRMAMAFSLAAC
      : : : . :   : :   : : : . : : : : : : : : : : : : : : : : : : :
genpep VKETDRISAMATELRKLGATVEEGADYLKVTTPPAQWQTPADGIGTYDDHRMAMCFSLAAF
      350      360      370      380      390      400

      420      430      440
2MEPSP AEPVPTIRDPGCTRKTFPDYFDVLSTFVK
      . . : : . : : : . : : : : : : :
genpep GPLPVRINDPGCVAKTFPDYFAVFGGVTR
      410      420      430

```

113525597 | *Ralstonia eutropha* H16 chromosome 1. | *Ralstonia eutropha* H16 | AA | 434
Length = 434
initn: 900 initl: 227 opt: 794 Z-score: 1588.5 bits: 303.1 E(): 6.3e-80
Smith-Waterman score: 948; 47.846% identity (74.150% similar) in 441 aa overlap
(5-440:2-430)

```

      10      20      30      40      50      60
2MEPSP MAGAEFIVLQPIKEISGTVKLPGSKSLSNRILLLAALSEGTTVVDNLLNSEDVHYMLGAL
      : . . : :   : : : : : : : : : : : : : : : : : : : : : :
genpep MEHLTLGLPLTRANGTVRLPGSKSISNRVLLLAALATGETRVRDLLDSDTRVMLQAL
      10      20      30      40      50

      70      80      90      100      110      120
2MEPSP RTLGLSVEADKAAKRAVVVGCGGKFPVEDAKEEVQLFLGNAGIAMRSLxxxxxxxxxxxx
      : : : . . . . : : : : : : : : : : : : : : : : : : : :
genpep RTLGVAWRQE--GDDYIVTGAGGNFPNKSA---ELFMGNAGTAIRPLTAALALQGG--N
      60      70      80      90      100

      130      140      150      160      170      180
2MEPSP YVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISS
      : : : : : : : : : : : : : : : : : : : : : : : : : :
genpep YKLSGVPRMHERPIGDLVDGLRQVGAVIDYLGNEGFPPLHIQPAIRIDAPIRVRGDVSS
      110      120      130      140      150      160

      190      200      210      220      230
2MEPSP QYxxxxxxxx--xxxxxGDVEIEIIDKLISIPYVEMTLRLMERFGVKAHSDSWDRFYIK
      : . . . . . : : : : : : : : : : : : : : : : : : :
genpep QFLTALLMSLPMAQSDSGRIEIEVVGELISKPYIEITLNLARFGIEIERQ-GWERFVLP
      170      180      190      200      210      220

      240      250      260      270      280      290
2MEPSP GGQKYKSPKNAYVEGDASSASYFLAGAAIxxxxxxxxxxxxxxxxSLQGDVKFAEVLEMMGAK
      . : : : . : : : : : : : : : : : : : : : : : : :
genpep AGAAYRSPGEIFVEGDASSASYFLAAGAIGGGPVRVEGVGMASIQGDVRF AEALNRMGAN
      230      240      250      260      270      280

```

**2mEPSPS PROTEIN
OVERALL AMINO ACID SEQUENCE HOMOLOGY SEARCH
WITH KNOWN TOXINS AND ALLERGENS**

FINAL REPORT AMENDMENT

There is no final report amendment at this time.