

**Study Title**

**Bioinformatics Analysis of CP4 EPSPS Protein Sequence  
Utilizing Toxin and Public Domain Genetic Databases**

**Authors**

**Ronald E. Hileman, Ph.D. and James D. Astwood, Ph.D.**

**Study Completed On**

**September 30, 1999**

**Performing Laboratory**

**Monsanto Company  
Biotechnology Regulatory Sciences  
700 Chesterfield Parkway North  
St. Louis, MO 63198**

**Laboratory Project ID**

**MSL Number: 16268**

**Study Number 99-01-62-08**

### 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).-

"We submit this material to the United States Environmental Protection Agency specifically under provisions contained in FIFRA as amended, and thereby consent to use and disclosure of this material by EPA according to FIFRA. In submitting this material to the EPA according to method and format requirements contained in PR Notice 86-5, we do not waive any protection of rights involving this material that would have been claimed by the company if this material had not been submitted to the EPA."

Company: Monsanto Company

Company Agent: \_\_\_\_\_

Title: \_\_\_\_\_

Signature: \_\_\_\_\_ Date: \_\_\_\_\_

**MSL Number:** 16268

**Title:** Bioinformatics Analysis of CP4 EPSPS Protein Sequence  
Utilizing Toxin and Public Domain Genetic Databases

**Facility:** Monsanto Company  
Biotechnology Regulatory Sciences  
700 Chesterfield Parkway North  
Chesterfield, Missouri 63198

**Sponsors:** Ravinder Sidhu

**Authors:** Ronald E. Hileman  
James D. Astwood

**Study Start Date:** September 13, 1999

**Study Completion Date:** September 30, 1999

**Records Retention:** All study specific raw data and final report will be retained at  
Monsanto-St. Louis.

**Signatures of Approval:**

Author: RW Hileman Date: 30 SEPT 99

Author: J D Astwood Date: Sept 30, 1999

Sponsor: Ravinder S Sidhu Date: 30 Sep 99

## Table of Contents

Section	Page
Title Page.....	1
Statement of No Data Confidentiality Claim .....	2
Signatures of Approval.....	3
Table of Contents .....	4
Abbreviations and Definitions .....	6
1.0 Summary .....	7
2.0 Introduction .....	7
3.0 Purpose .....	8
4.0 Methods.....	8
4.1 Toxin database preparation .....	8
4.2 Toxin database search for sequences similar to CP4 EPSPS protein.....	8
4.3 Public domain genetic database search for sequences similar to CP4 EPSPS protein .....	9
5.0 Results and Discussion.....	9
5.1 Toxin sequence database (TOXIN4) .....	9
5.2 CP4 EPSPS protein sequences are similar to insecticidal toxins .....	9
5.3 Public domain genetic database search for sequences similar to CP4 EPSPS protein .....	10
6.0 Conclusions .....	10
7.0 References .....	11

## Appendices

Appendix 1.	CP4 EPSPS protein sequence.....	12
Appendix 2.	FASTA sequence similarity alignments of CP4 EPSPS protein with toxins .....	13
Appendix 3.	FASTA sequence similarities of CP4 EPSPS protein with proteins in publicly available genetic databases.....	16

### Abbreviations and Definitions

aa	amino acid
<i>cp4 epsps</i>	gene encoding the <i>Agrobacterium</i> sp. strain CP4 5-enolpyruvylshikimate-3-phosphate synthase protein
CP4 EPSPS	Protein produced from the <i>cp4 epsps</i> gene
E.C.	Enzyme commission number
EMBL	European Molecular Biology Laboratory
FASTA	Algorithm used to find local high scoring alignments between a pair of protein or nucleotide sequences
GCG	Genetics Computer Group
GenBank	A public genetic database maintained by the National Center for Biotechnology Information
GenPept	GenBank peptide database
NRL3D	National Research Laboratories, 3D Protein database at Brookhaven
PIR	Protein Information Resource
RR	Roundup Ready®
SwissProt	Translated sequences from EMBL database
TOXIN4	Toxin database

## 1.0 Summary

In the safety assessment of proteins introduced into genetically modified crops, possible human health effects are addressed in a multifaceted approach which involves addressing the history of safe use, directly assessing the oral acute toxicity of the introduced protein, assessing the sensitivity to rapid digestion and comparing the amino acid sequence of the introduced protein with proteins associated with toxicity or other adverse health effects. A biologically relevant sequence similarity to a known toxin (*i.e.*, a sequence derived from a common ancestor gene) may suggest that additional toxicological assessments should be performed.

A database of protein sequences associated with toxicity was assembled from publicly available genetic databases (GenBank, EMBL, PIR and SwissProt). The keyword "toxin" was used to retrieve 4677 toxin sequences from the public domain databases. The amino acid sequence of the CP4 EPSPS protein was compared to protein sequences in the toxin database using the FASTA sequence alignment tool. In addition, the amino acid sequence of the CP4 EPSPS protein was compared to all protein sequences in the publicly available genetic databases to screen for structural similarity to other known proteins, including pharmacologically active proteins potentially relevant to animal and human health. The test sequence, CP4 EPSPS, shared sequence similarities to homologous EPSPS proteins, as expected. EPSPS proteins are ubiquitous, occurring in all plants and some algal and fungal species (Padgett et al., 1996). These data confirm that the CP4 EPSPS protein is homologous to other EPSPS proteins. No other significant structural homology was observed.

## 2.0 Introduction

Monsanto Company has developed Roundup Ready<sup>®</sup> corn line NK603 which is tolerant to glyphosate (the active ingredient in Roundup<sup>™</sup> herbicide) at the whole plant level. Corn line NK603 contains a 5-enolpyruvylshikimate-3-phosphate synthase protein from *Agrobacterium* sp. strain CP4 (CP4 EPSPS). Corn plants that demonstrate commercial level tolerance to Roundup herbicide are called Roundup Ready<sup>®</sup> (RR). The CP4 EPSPS gene from *Agrobacterium* sp. strain CP4 has been completely sequenced and encodes a 47.6-kDa protein consisting of a single polypeptide of 455 amino acids (Padgett et al., 1996). The CP4 EPSPS protein is functionally and structurally similar to plant EPSPS enzymes but has a much reduced affinity for glyphosate (Padgett et al., 1993). In non-transgenic plants, glyphosate binds to the plant EPSPS enzyme and blocks the biosynthesis of aromatic amino acids thereby starving plants of these essential nutrients (Haslam, 1993; Steinrucken and Amrhein, 1980). In RR plants, nutritional requirements for normal growth and development are met by the continued action of the tolerant CP4 EPSPS enzyme in the presence of glyphosate. A comprehensive safety assessment of the CP4 EPSPS protein has been described in the literature (Harrison et al., 1996).

This report describes the bioinformatics assessment of the CP4 EPSPS protein which demonstrated the absence of sequence similarity to proteins known to pose animal and human health risks. Potential for toxicity of the CP4 EPSPS protein was assessed by comparing the amino acid sequences to proteins associated with toxicity or other health effects. To facilitate these analyses, a database of protein sequences associated with toxicity was assembled from publicly available genetic databases (GenBank, EMBL, PIR and SwissProt). In addition, the amino acid sequence of the CP4 EPSPS protein was compared to all protein sequences resident in publicly available genetic databases to screen for structural similarity to other known proteins, including pharmacologically active proteins potentially relevant to animal and human health.

### 3.0 Purpose

The purpose of this study was to evaluate the amino acid sequence similarities of the CP4 EPSPS protein to protein sequences within a toxin database as well as to all protein sequences in the publicly available genetic databases using bioinformatics techniques. A biologically relevant sequence similarity (*i.e.* a sequence derived from a common ancestor gene) to a known toxin or pharmacologically active protein may cause concern and additional toxicological assessments may be initiated to address the impact of the apparent homology.

### 4.0 Methods

**4.1** *Toxin database preparation.* The toxin database was assembled from public domain databases (Genbank and EMBL GenPept version 108, PIR and NRL3D version 56 and SwissProt version 36). Protein sequences were retrieved using the STRINGSEARCH function (keyword = toxin) of the Wisconsin Package, version 9.1, Genetics Computer Group (GCG), Madison, Wisconsin. The DATASET function was used to compile the individual toxin sequences into a database, named TOXIN4, representing a total of 4677 sequences.

**4.2** *Toxin database search for sequences similar to CP4 EPSPS protein.* The SHUFFLE function of GCG (version 10.0) was used to generate a randomized sequence of the CP4 EPSPS protein. These randomized sequences were used to test for amino acid bias (negative control) in the similarity search of the parent CP4 EPSPS sequence against TOXIN4. TOXIN4 was searched using FASTA (Pearson and Lipman, 1988) for sequences similar to both the test proteins as well as their randomized sequences. Where sequence similarities were observed between the CP4 EPSPS protein and toxins, the complete data file of the TOXIN4 entry was retrieved. For the similar sequences identified, the scientific literature was consulted to verify that the similarity of "toxin" was either a true toxin or an



accidental (irrelevant) protein sequence retrieved by the STRINGSEARCH program.

- 4.3 *Public domain genetic database search for sequences similar to CP4 EPSPS protein.* The CP4 EPSPS protein sequence was compared to all known proteins in the public domain databases. The database used was ALLPEPTIDES and is comprised of all publicly available protein sequences (GenBank and EMBL GenPept version 108, PIR and NRL3D version 56 and SwissProt version 36). Sequence similarity was assessed using the FASTA program of GCG (version 10.0).

## 5.0 Results and Discussion

- 5.1 *Toxin sequence database (TOXIN4).* The keyword "toxin" was used to identify and retrieve 4677 amino acid sequences from the public domain databases. The 4677 amino acid sequences were assembled into a DATASET named TOXIN4. Since the STRINGSEARCH may retrieve multiple entries of the same toxin entered in separate databases, the actual number of unique toxin protein sequences in TOXIN4 is less than 4677. For example, two separate entries for a *Cochliobolous carbonum* protein were identified (described in Section 5.2).

- 5.2 *CP4 EPSPS protein sequence is not similar to protein toxins.* The complete FASTA output file for the CP4 EPSPS protein is shown in Appendix 2. The complete annotated sequence file for each protein identified from the TOXIN4 database was retrieved to make an initial toxicity assessment. A total of 6 sequence entries were identified. The best score was 154 aa overlap of a *Cochliobolous carbonum* protein, ToxD (Accession No. X92391, P54006). In the overlap, the CP4 EPSPS protein shared 24.7% amino acid identities and 53.2% similarity (percent residues having identical or conservative replacements) to the ToxD protein. Although the function of this protein has not been described, it is found in *C. carbonum* isolates that produce a cyclic peptide toxin. No toxicity has been described for the ToxD protein. This protein was assembled into the TOXIN4 database due to the reference of the cyclic peptide toxin in the annotation section of the sequence file. This degree of similarity does not likely represent bona fide homology (Doolittle, 1990).

A polygalacturonidase protein (Accession No. JC4049) was also identified in the FASTA output file (27.7% aa identity in a 94 aa overlap, Appendix 2). This pectinase protein, like the ToxD protein described above, is not a toxin. It is produced by *Aspergillus parasiticus*, an organism that produces aflatoxin and is responsible for fungal disease of cotton and other crops. The observed degree of

but see  
more of  
vol 1.

similarity to the polygalacturonidase protein does not likely represent homology (Doolittle, 1990).

The remaining proteins identified, *Escherichia coli* CDT-IIIa (Accession No. U89305, O32585) and *Serratia marcescens* shlA (Accession No. P15320) are bacterial toxins and share a low degree of similarity to the CP4 EPSPS protein. The observed degree of similarity to these proteins do not likely represent homology (Doolittle, 1990).

Thus no significant similarities were observed between the CP4 EPSPS protein and protein toxins in the TOXIN4 database. This observation was consistent with the previously reported equivalence of animal feed wholesomeness (Hammond et al., 1996) and lack of acute oral toxicity (Harrison et al., 1996).

A randomized amino acid sequence of the CP4 EPSPS protein was used to address potential amino acid bias. The negative control sequence (shuffled CP4 EPSPS protein sequence) also revealed no significant similarities to protein toxins. The CP4 EPSPS protein was no more likely to share sequence similarity with sequences in the TOXIN4 database than the randomized CP4 EPSPS protein sequence.

- 5.3 *Public domain genetic database search for sequences similar to CP4 EPSPS protein.* Sequence similarity of the CP4 EPSPS protein to the current public domain protein databases was assessed (Appendix 3). As anticipated, the vast majority of sequences similar to the CP4 EPSPS protein were EPSPS proteins (E.C. 2.5.1.19) from other organisms such as bacteria, yeast, plant and fungi. This observation has been reported previously (Padgett et al., 1996). In addition to the EPSPS proteins, a low degree of similarity was observed towards UDP-N-acetylglucosamine 1-carboxyvinyltransferase (E.C. 2.5.1.7, Accession No. P19670, S76347, Q55673) as well as the  $\alpha$ -chain of DNA-directed RNA polymerase (E.C. 2.7.7.6, Accession No. S71425, P74963). No toxicity has been expected for or associated with these proteins in the literature.

## 6.0 Conclusions

The results from this bioinformatics assessment reconfirm that the CP4 EPSPS protein is not homologous to protein toxins or proteins associated with animal and human health risks. As expected, the CP4 EPSPS protein was found to be homologous to EPSPS protein present in other organisms.

## 7.0 References

- Doolittle, R. F. 1990. Searching through sequence databases. *Methods in Enzymology* 183, 99-110.
- Hammond, B. G., Vicini, J. L., Hartnell, G. F., Naylor, M. W., Knight, C. D., Robinson, E. H., Fuchs, R. L. and Padgett, S. R. 1996. The feeding value of soybeans fed to rats, chickens, catfish and dairy cattle is not altered by genetic incorporation of glyphosate tolerance. *J Nutr* 126, 717-27.
- Harrison, L. A., Bailey, M. R., Naylor, M. W., Ream, J. E., Hammond, B. G., Nida, D. L., Burnette, B. L., Nickson, T. E., Mitsky, T. A., Taylor, M. L., Fuchs, R. L. and Padgett, S. R. 1996. The expressed protein in glyphosate-tolerant soybean, 5-enolpyruvylshikimate-3-phosphate synthase from *Agrobacterium* sp. strain CP4, is rapidly digested *in vitro* and is not toxic to acutely gavaged mice. *J Nutr* 126, 728-40.
- Haslam, E. 1993. *Shikimic Acid: Metabolism and Metabolites* John Wiley and Sons, Chichester, England.
- Padgett, S., Re, D., Eichholtz, D., Delannay, X., Fuchs, R., Kishore, G. and Fraley, R. 1996. New weed control opportunities: Development of soybeans with a Roundup Ready™ gene, Duke, S. O., ed. CRC, Boca Raton, FL.
- Padgett, S. R., Barry, G. F., Re, D. B., Weldon, M., Eichholtz, D. A., Kolacz, K. and Kishore, G. M. 1993. Purification, cloning and characterization of a highly glyphosate-tolerant EPSP synthase from *Agrobacterium* sp. strain CP4. Study number MSL-12738, an unpublished study conducted by Monsanto Company.
- Pearson, W. and Lipman, D. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85, 2440-2448.
- Steinrucken, H. C. and Amrhein, N. 1980. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase. *Biochem Biophys Res Commun* 94, 1207-12.

**Appendix 1. CP4 EPSPS protein sequence.**

(filename = CP4\_EPSPS.pep)

```
1  MLHGASSRPA TARKSSGLSG TVRIPGDKSI SHRSFMFGGL ASGETRITGL
51  LEGEDVINTG KAMQAMGARI RKEGDTWIID GVGNGGLLAP EAPLDFGNAA
101 TGCRLTMGLV GVDYDFDSTFI GDASLTKRPM GRVLNPLREM GVQVKSEDGD
151 RLPVTLRGPK TPTPITYRVP MASAQVKS AV LLAGLNTPGI TTVIEPIMTR
201 DHTEKMLQGF GANLTVETDA DGVRTIRLEG RGKLTGQVID VPGDPSSTAF
251 PLVAALLVPG SDVTILNVLM NPTRTGLILT LQEMGADIEV INPRLAGGED
301 VADLRVRSST LKGVTVPEDR APSMIDEYPI LAVAAFAEG ATVMNGLEEL
351 RVKESDRLSA VANGLKLNGV DCDEGETSLV VRGRPDGKGL GNASGAAVAT
401 HLDHRIAMSF LVMGLVSENP VTVDDATMIA TSFPEFMDLM AGLGAKIELS
451 DTKAA
```

## Appendix 2. FASTA sequence similarity alignments of CP4 EPSPS protein with toxins

(Peptide) FASTA of: CP4\_EPSPS.pep from: 1 to: 455 September 13, 1999 12:07

TO: TOXIN4:\* \*  
Sequences: 4,677  
Symbols: 1,402,894  
Word Size: 2

Databases searched: Release 1.0, Released on 28Dec98, Formatted on 28Dec98

Scoring matrix: GenRunData:blosom50.cmp  
Variable pamfactor used  
Gap creation penalty: 12  
Gap extension penalty: 2

Histogram Key:  
Each histogram symbol represents 9 search set sequences  
Each inset symbol represents 1 search set sequences  
z-scores computed from opt scores

z-score	obs (=)	exp (*)
< 20	20	0:===
22	0	0:
24	0	0:
26	0	0:
28	0	1:*
30	16	6:*=
32	16	25:==*
34	35	67:===== *
36	139	138:=====*
38	257	228:=====*
40	367	318:=====*
42	401	389:=====*
44	429	429:=====*
46	522	437:=====*
48	354	419:=====*
50	315	382:=====*
52	337	336:=====*
54	240	287:=====*
56	236	240:=====*
58	184	197:=====*
60	178	159:=====*
62	171	128:=====*
64	99	102:=====*
66	60	80:=====*
68	58	63:=====*
70	53	49:=====*
72	35	39:=====*
74	41	30:=====*
76	40	23:=====*

78	6	18:==*
80	10	14:==*
82	16	11:==*
84	9	9:*
86	6	7:*
88	4	5:*
90	12	4:*=
92	4	3:*
94	2	2:*
96	0	2:*
98	0	1:*
100	2	1:*
102	0	1:*
104	0	1:*
106	0	1:*
108	1	0:*
110	0	0:*
112	1	0:*
114	1	0:*
116	0	0:*
118	0	0:*
>120	0	0:*

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are:

	init1	initn	opt	z-sc	E(4654)...
TO:CCTOXD_1					
! X92391 Cochliobolus carbonum C.carb...	36	36	91	113.8	0.73
TO:TOXD_COCCA					
! P54006 cochliobolus carbonum (bipol...	36	36	91	113.8	0.73
TO:JC4049					
! polygalacturonase (EC 3.2.1.15) pre...	67	67	88	108.7	1.4
TO:ECU89305_1					
! U89305 Escherichia coli Escherichia...	47	47	79	100.4	4.1
TO:O32585					
! O32585 escherichia coli. cytolethal...	47	47	79	100.4	4.1
TO:HLYA_SERMA					
! P15320 serratia marcescens. hemolys...	59	59	85	94.2	9

\\End of List

CP4\_EPSPS.pep  
TO:CCTOXD\_1

LOCUS CCTOXD\_1  
DEFINITION C.carbonum toxD gene;  
unique to isolates that make the cyclic peptide  
HC-toxin.  
DATE 26-OCT-1995  
ACCESSION X92391 . . .

SCORES Init1: 36 Initn: 36 Opt: 91 z-score: 113.8 E(): 0.73  
Smith-Waterman score: 91; 24.7% identity in 154 aa overlap

80 90 100 110 120  
 CP4\_EPSPS.pe EGDWTIIDGVNGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI-----GDASLT  
 CCTOXD\_1 PGCLVGC DYAGIVEEVGRSVKKPFKKGDRVCGFAHGGNAVFSDDGTF AEVITVKGDIQ-A  
 60 70 80 90 100 110

130 140 150 160 170 180  
 CP4\_EPSPS.pe KRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTIYRVPM-----ASAQVKS AVL  
 CCTOXD\_1 WIPENLSFQEAATLGVG IKT V-QGGLYQSLKLSWPTTPIEHAVPIL IYGGSTATGT LAIQ  
 120 130 140 150 160 170

190 200 210 220 230  
 CP4\_EPSPS.pe LAGLNTPGITTVIEPIMTRDHTKMLQGFGANLTVE---TDADGVRTIRLEGRGKLTGQ  
 CCTOXD\_1 LAKLSGYRVITTCSP----HHFELMKSLGADLVFDYHEITSADHIRRCT-QNK LKLVFD  
 180 190 200 210 220

240 250 260 270 280 290  
 CP4\_EPSPS.pe VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVINPRLAG  
 CCTOXD\_1 TISIDVSARFCDRAMSTEGGEYSALLDVS IARTNISSRWTLAYTVLGE GFTSEQIVFQPY  
 230 240 250 260 270 280

CP4\_EPSPS.pep  
 TO:TOXD\_COCCA

ID TOXD\_COCCA STANDARD; PRT; 297 AA.  
 AC P54006;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE TOXD PROTEIN. . . .

SCORES Initl: 36 Initn: 36 Opt: 91 z-score: 113.8 E(): 0.73  
 Smith-Waterman score: 91; 24.7% identity in 154 aa overlap

80 90 100 110 120  
 CP4\_EPSPS.pe EGDWTIIDGVNGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI-----GDASLT  
 TOXD\_COCCA PGCLVGC DYAGIVEEVGRSVKKPFKKGDRVCGFAHGGNAVFSDDGTF AEVITVKGDIQ-A  
 60 70 80 90 100 110

130 140 150 160 170 180  
 CP4\_EPSPS.pe KRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTIYRVPM-----ASAQVKS AVL  
 TOXD\_COCCA WIPENLSFQEAATLGVG IKT V-QGGLYQSLKLSWPTTPIEHAVPIL IYGGSTATGT LAIQ  
 120 130 140 150 160 170

190 200 210 220 230  
 CP4\_EPSPS.pe LAGLNTPGITTVIEPIMTRDHTKMLQGFGANLTVE---TDADGVRTIRLEGRGKLTGQ  
 TOXD\_COCCA LAKLSGYRVITTCSP----HHFELMKSLGADLVFDYHEITSADHIRRCT-QNK LKLVFD  
 180 190 200 210 220

240 250 260 270 280 290  
 CP4\_EPSPS.pe VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVINPRLAG  
 TOXD\_COCCA TISIDVSARFCDRAMSTEGGEYSALLDVS IARTNISSRWTLAYTVLGE GFTSEQIVFQPY  
 230 240 250 260 270 280

CP4\_EPSPS.pep  
 TO:JC4049

Pl:JC4049 - polygalacturonase (EC 3.2.1.15) precursor - Aspergillus  
 parasiticus  
 C:Species: Aspergillus parasiticus  
 C:Date: 27-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 06-Dec-1996  
 C:Accession: JC4049  
 R:Cary, J.W.; Brown, R.; Cleveland, T.E.; Whitehead, M.; Dean, R.A.  
 Gene 153, 129-133, 1995 . . .

SCORES Initl: 67 Initn: 67 Opt: 88 z-score: 108.7 E(): 1.4  
 Smith-Waterman score: 88; 27.7% identity in 94 aa overlap

10 20 30  
 CP4\_EPSPS.pe MLHGASSRPATARKSSGLSGTVRI PGDKSISHRS  
 JC4049 SVIAATVGAALVAAAPVELEARDSCFTTSAADAKSGKTSCTITL SNIEVPAGETI.DLTG  
 10 20 30 40 50 60

40 50 60 70 80  
 CP4\_EPSPS.pe FMFGGLA--SGETRIT-GLLEGE DVINTG---KAMQAMGARIRKEGDTWIIDGVG-NGGL  
 JC4049 LNDGTTVIFSGETTFPGYKEWEGPLISVSGTNIKVQASGAKIDGDSRWV-DGEGGNGGK  
 70 80 90 100 110 120

90 100 110 120 130 140  
 CP4\_EPSPS.pe LAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFIGDASLT KRPMGRVLNPLREMGVQVKSE  
 JC4049 TKPKFFYAHKLDSSSITGLQIYNTFPVQGF SIQSDNLNITDVITD NSAGTAEGHNTDAFDI  
 130 140 150 160 170 180

CP4\_EPSPS.pep  
 TO:ECU89305\_1

LOCUS ECU89305\_1  
 DEFINITION Escherichia coli plasmid pVIR cytolethal distending toxin-III  
 (cdt-III A, cdt-III B and cdt-III C) genes, complete cds;  
 form 1 of 3; CDT-III A.  
 DATE 26-JUN-1997  
 ACCESSION U89305 . . .

SCORES Initl: 47 Initn: 47 Opt: 79 z-score: 100.4 E(): 4.1  
 Smith-Waterman score: 79; 36.2% identity in 47 aa overlap

CP4\_EPSPS.pe PT--PITYRVPMASAVQKSAVLLAGLNTPGITTVEIPIMTRDHTKMLQGFGANLTVETD  
|| || || :: |:  
ECU89305\_1 PTNAPIPIPVPGTAPAVSLMNMDSGLTMTWSRGAGSSLWAYIISDSNSFGELRNWQIMPG  
70 80 90 100 110 120

```

ID 032585          PRELIMINARY;    PRT;    258 AA.
AC 032585;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CYPOTOETHAL DISTENDING TOXIN-IIIA.

```

CP4\_EPSPS.pe GCRLTMGLVGVDYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKT  
                  110       120       130       140       150       160  
                                ::| | | : | | : ||:  
O32585 KCTSILIVGILIPILLNGCSSRKNRAHLDPKVPFPQVEGGPTIPSPDEPGLPGLPGAGPAL  
                  10       20       30       40       50       60

CP4\_EPSPS.pe PT--PITYRVPMASAVQKSAVLLAGLNTPGITTVEIPIMTRDHTKMLQGFGANLTVETD  
O32585 PTNAPIPIPVPGTAPAVSLMNMMDGSLTMSRGAGSSLWAYIYISDSNSFGELRNWQIMPG

```

ID      HLYA_SERMA      STANDARD;      PRT;  1608 AA.
AC      P15320;
DT      01-APR-1990 (REL. 14, CREATED)
DT      01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT      01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
DE      HEMOLYSIN PRECURSOR.

```

CP4\_EPSPS.pe LSGTVRIPGDKSISHRSFMFGLGASGETRITGLLEGEDVINTGKAMQAMGARIRKEGDTW  
HLYA\_SERMA VDPNTGKDTHTLGLGKLVGVGEQDKTHTANGITAGDVTLSNGKDLRLAGRV--DADS-

CP4\_EPSPS.pe IIDGVGNGGLLAPEAPLDFGNAATGCRLTMGLVGVDYDFDSTFIGDASL--TKRPMGRVLN  
::| | | | : | : : : | : : | : : : | : | : | :  
HLYA\_SERMA -VQG KVGGDHLVESRKDVENGVK-VDDVDAGLSHSDPGSSITSKLSKVGTPRYAGKVKE  
1360 1370 1380 1390 1400 1410

CP4\_EPSPS.pe 140 150 160 170 180 190  
PLERMGV-QVKSEGDRLPVTLR--GPKTTPITYRVPMASQVKSAVLLAGLNTPGITT  
| | | : : : | : | : : : | : : |  
HLYA\_SERMA KL-EAGVNKVADATTDKYNVSARRLDPOQDTTGAVSFSKAEGKVTLPTATPAGEKPGPLW  
1420 1430 1440 1450 1460 1470

CP4\_EPSPS.pe VIEPIMTRDHTKMLQGFGANLTVETDADGVRTIRLEGRGKLTGVQIVDVGDPSPSTAFPL  
 HLYA\_SERMA DRGARTVGGAVKDSITGPAGRQGHLLKVNADVNNNAVGEQSAIAGKNGVALQVGGQTQLT

```

! Distributed over 1 thread.
!      Start time: Mon Sep 13 12:07:22 1999
! Completion time: Mon Sep 13 12:07:32 1999

! CPU time used:
!      Database scan: 0:00:02.8
! Post-scan processing: 0:00:00.4
!      Total CPU time: 0:00:03.2
! Output File: CP4_EPSPS.TOXIN4

```

### Appendix 3. FASTA sequence similarities of CP4 EPSPS protein with proteins in publicly available genetic databases

(Peptide) FASTA of: CP4\_EPSPS.pep from: 1 to: 455 September 13, 1999 12:13

TO: ALLPEPTIDES: \*  
Sequences: 840,315  
Symbols: 263,233,319  
Word Size: 2

Databases searched:

SWISS-PROT, Release 38.0, Released on 23Jul1999, Formatted on 1Sep1999  
SWISS-PROT, Release 11.0, Released on 3Aug1999, Formatted on 23Aug1999  
PIR, Release 60.0, Released on 31Mar1999, Formatted on 9May1999  
GenBank, Release 113.0, Released on 25Aug1999, Formatted on 3Sep1999

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12

Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1311 search set sequences

Each inset symbol represents 9 search set sequences

z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	4397	0:====
22	10	0:=
24	32	1:*
26	71	18:*
28	326	190:*
30	1539	1157:*=
32	5114	4475:====*
34	13116	12134:=====*
36	27318	24921:=====*
38	46685	41185:=====*
40	63777	57450:=====*
42	75427	70225:=====*
44	78649	77465:=====*
46	77834	78900:=====*
48	75579	75538:=====*
50	66150	68928:=====*
52	58330	60600:=====*
54	50478	51763:=====*
56	40222	43238:=====*
58	33711	35497:=====*
60	27099	28755:=====*
62	20470	23053:=====*
64	16409	18334:=====*
66	13319	14491:=====*
68	10211	11398:=====*
70	7719	8932:=====*
72	5770	6980:=====*
74	4697	5442:=====*
76	3692	4235:=====*
78	2630	3292:=====*

80	2045	2556:==*
82	1648	1956:==*
84	1213	1549:==*
86	957	1199:==*
88	720	927:==*
90	618	718:==*
92	424	555:==*
94	353	430:==*
96	271	332:==*
98	259	257:==*
100	214	199:==*
102	217	154:==*
104	104	119:==*
106	69	92:==*
108	62	71:==*
110	29	55:==*
112	34	43:==*
114	26	33:==*
116	29	26:==*
118	20	20:==*
>120	222	15:==*

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are:  
E(839954)...

	initl	initn	opt	z-sc
PIR2:S76218				
! 3-phosphoshikimate 1-carboxyvinyltr...	480	977	1149	1239.3 2.7e-61
GP_BCT1:SSAROAGEN_1				
! X75325 Synechocystis sp. Synechocys...	480	977	1149	1239.3 2.7e-61
GP_BCT1:D90914_87				
! D90914 Synechocystis sp. Synechocys...	480	977	1149	1239.3 2.7e-61
SW:AROA_SYNY3				
! Q59975 synechocystis sp. (strain pc...	480	977	1149	1239.3 2.7e-61
SP_PRO:Q46550				
! Q46550 bacteroides nodosus (dichelo...	442	961	1065	1149.1 2.8e-56
PIR2:S44096				
! enolpyruvylshikimate 3-phosphate sy...	442	961	1065	1149.1 2.8e-56
GP_BCT1:DNEPS3PS_1				
! Z29339 Dichelobacter nodosus D.nodo...	442	961	1065	1149.1 2.8e-56
PIR2:D70433				
! 5-enolpyruvylshikimate-3-phosphate ...	453	565	1015	1095.5 2.7e-53
GP_BCT2:AE000744_15				
! AE000744 Aquifex aeolicus Aquifex a...	453	565	1015	1095.5 2.7e-53
SP_PRO:O67494				
! O67494 aquifex aeolicus. 5-enolpyru...	453	565	1015	1095.5 2.7e-53
GP_BCT2:AF169483_1				
! AF169483 Streptococcus pneumoniae S...	385	900	971	1048.3 1.2e-50
PIR2:S52580				
! 3-phosphoshikimate 1-carboxyvinyltr...	400	969	957	1033.2 8.1e-50
SW:AROA_LACLA				
! P43905 lactococcus lactis (subsp. 1...	400	969	957	1033.2 8.1e-50
GP_BCT1:LLTYRAPH_2				
! X78413 Lactococcus lactis L.lactis ...	400	969	957	1033.2 8.1e-50
PIR2:C26532				
! 5-enolpyruvylshikimate-3-phosphate...	525	943	930	1004.3 3.3e-48
GP_BCT1:BSUB0012_200				



! Z99115 *Bacillus subtilis* *Bacillus* s... 525 943 930 1004.3 3.3e-48  
 SW:AROA\_BACSU  
 ! P20691 *bacillus subtilis*. putative ... 525 943 930 1004.3 3.3e-48  
 GP\_BCT1:BACVARGNS\_18  
 ! M80245 *Bacillus subtilis* *B.subtilis*... 525 943 930 1004.3 3.3e-48  
 GP\_BCT2:AE001715\_8  
 ! AE001715 *Thermotoga maritima* *Thermo*... 326 792 835 902.5 1.5e-42  
 GP\_BCT1:STAAROA\_2  
 ! L05004 *Staphylococcus aureus* *Staphy*... 271 526 770 832.3 1.3e-38  
 SW:AROA\_STAAU  
 ! Q05615 *staphylococcus aureus*. 3-pho... 271 526 770 832.3 1.3e-38  
 PIR2:G71863  
 ! 3-phosphoshikimate 1-carboxyvinyl t... 222 573 713 771.1 3.2e-35  
 SP\_PRO:Q9ZKF7  
 ! Q9Zkf7 *helicobacter pylori* j99. 3-p... 222 573 713 771.1 3.2e-35  
 GP\_BCT2:AE001527\_6  
 ! AE001527 *Helicobacter pylori* J99 He... 222 573 713 771.1 3.2e-35  
 PIR2:A64570  
 ! 3-phosphoshikimate 1 carboxyvinyltr... 220 570 706 763.6 8.4e-35  
 GP\_BCT2:AE000556\_2  
 ! AE000556 *Helicobacter pylori* 26695 ... 220 570 706 763.6 8.4e-35  
 SW:AROA\_HELPY  
 ! P56197 *helicobacter pylori* (campylo... 220 570 706 763.6 8.4e-35  
 SW:AROA\_CAMJE  
 ! P52312 *campylobacter jejuni*. 3 phos... 203 676 648 701.3 2.5e-31  
 GP\_BCT1:CJDNAPALG\_2  
 ! X89371 *Campylobacter jejuni* *C.jejun*... 203 676 648 701.3 2.5e-31  
 PIR2:JC5338  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 203 676 648 701.3 2.5e-31  
 SW:AROA\_HAESO  
 ! P52310 *haemophilus somnus*. 3-phosph... 133 304 466 505.7 2e-20  
 GP\_BCT1:HEA3P1C\_1  
 ! L47538 *Haemophilus somnus* *Haemophil*... 133 304 466 505.7 2e-20  
 GP\_BCT2:PHU89948\_1  
 ! U89948 *Pasteurella haemolytica* *Past*... 141 299 448 486.4 2.3e-19  
 SP\_PRO:P96968  
 ! P96968 *pasteurella haemolytica*. 3-p... 141 299 448 486.4 2.3e-19  
 GP\_BCT1:PHU03068\_1  
 ! U03068 *Pasteurella haemolytica* *Past*... 141 299 443 480.9 4.7e-19  
 SW:AROA\_PASHA  
 ! P54220 *pasteurella haemolytica*. 3-p... 141 299 443 480.9 4.7e-19  
 PIR2:JN0758  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 170 312 442 479.9 5.3e-19  
 GP\_BCT1:HEAAROAUR\_2  
 ! L04686 *Haemophilus influenzae* *Haemo*... 170 312 442 479.9 5.3e-19  
 SW:AROA\_YERPE  
 ! Q60112 *yersinia pestis*. 3-phosphosh... 176 359 430 467.1 2.7e-18  
 GP\_BCT1:YEPAROA\_1  
 ! L46372 *Yersinia pestis* *Yersinia pes*... 176 359 430 467.1 2.7e-18  
 PIR1:XUEBY  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 185 373 425 461.7 5.5e-18  
 GP\_BCT1:YEPSERCARO\_2  
 ! M32213 *Yersinia enterocolitica* *Y.en*... 185 373 425 461.7 5.5e-18  
 SW:AROA\_YEREN  
 ! P19688 *yersinia enterocolitica*. 3-p... 185 373 425 461.7 5.5e-18  
 GP\_BCT2:U32833\_2

! U32833 *Haemophilus influenzae* *Rd Ha*... 170 315 425 461.6 5.6e-18  
 SW:AROA\_HAEIN  
 ! Q03421 *haemophilus influenzae*. 3-ph... 170 315 425 461.6 5.6e-18  
 PIR2:F64131  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 170 315 425 461.6 5.6e-18  
 SW:AROA\_KLEPN  
 ! P24497 *klebsiella pneumoniae*. 3-pho... 151 403 420 456.3 1.1e-17  
 GP\_BCT1:KPAROA\_1  
 ! X82415 *Klebsiella pneumoniae* *K.pneu*... 151 403 420 456.3 1.1e-17  
 PIR2:S13266  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 151 403 420 456.3 1.1e-17  
 PIR2:S41329  
 ! 5-enolpyruvylshikimate-3-phosphate ... 178 367 412 448.0 3.2e-17  
 GP\_BCT1:PPAROA\_1  
 ! X77019 *Burkholderia pseudomallei* *P...* 178 367 412 448.0 3.2e-17  
 SW:AROA\_BURPS  
 ! P39915 *burkholderia pseudomallai* (p... 178 367 412 448.0 3.2e-17  
 GP\_BCT2:AE000855\_4  
 ! AE000855 *Methanobacterium thermoaut*... 136 465 399 433.9 1.9e-16  
 PIR2:D69202  
 ! 5-enolpyruvylshikimate 3-phosphate ... 136 465 399 433.9 1.9e-16  
 SP\_ARC:O26860  
 ! O26860 *methanobacterium thermoautot*... 136 465 399 433.9 1.9e-16  
 SW:AROA\_BORPE  
 ! P12421 *bordetella pertussis*. 3-phos... 123 224 396 430.3 3.1e-16  
 GP\_BCT1:BPEAROA\_1  
 ! M20023 *Bordetella pertussis* *B.pertu*... 123 224 396 430.3 3.1e-16  
 PIR1:XUBRVS  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 123 224 396 430.3 3.1e-16  
 PIR1:XUECVS  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 163 413 392 426.3 5.2e-16  
 GP\_BCT1:D90729\_6  
 ! D90729 *Escherichia coli* *Escherichia*... 163 413 392 426.3 5.2e-16  
 GP\_BCT2:AE000193\_3  
 ! AE000193 *Escherichia coli* *Escherich*... 163 413 392 426.3 5.2e-16  
 GP\_BCT1:D90728\_13  
 ! D90728 *Escherichia coli* *Escherichia*... 163 413 392 426.3 5.2e-16  
 SW:AROA\_ECOLI  
 ! P07638 *escherichia coli*. 3-phosphos... 163 413 392 426.3 5.2e-16  
 GP\_BCT2:AF101225\_1  
 ! AF101225 *Shigella sonnei* *Shigella s*... 163 413 391 425.2 6e-16  
 SP\_PRO:Q9ZFF7  
 ! Q9zff7 *shigella sonnei*. 3-phosphosh... 163 413 391 425.2 6e-16  
 GP\_BCT1:ECAROA\_1  
 ! X00557 *Escherichia coli* *E.coli* gene... 160 407 386 419.8 1.2e-15  
 NRL3D:1EPS  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 160 407 386 419.8 1.2e-15  
 GP\_BCT2:AF110153\_2  
 ! AF110153 *Edwardsiella ictaluri* *Edwa*... 136 357 386 419.8 1.2e-15  
 SP\_PRO:O87006  
 ! O87006 *shigella dysenteriae*. 3-phos... 163 413 382 415.5 2.1e-15  
 GP\_BCT2:SDU82268\_1  
 ! U82268 *Shigella dysenteriae* *Shigell*... 163 413 382 415.5 2.1e-15  
 SW:AROA\_SALTY  
 ! P07637 *salmonella typhimurium*. 3-ph... 142 376 372 404.8 8.2e-15  
 GP\_BCT1:STYAROAPM\_1

! M10947 Salmonella typhimurium S.typ... 142 376 372 404.8 8.2e-15  
 PIR1:XUEBVS  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 142 376 372 404.8 8.2e-15  
 GP\_BCT1:STSERARO\_2  
 ! Y10355 Salmonella typhimurium S.typ... 142 376 372 404.8 8.2e-15  
 PIR2:A49746  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 147 348 370 402.6 1.1e-14  
 GP\_BCT1:STYSERARO\_2  
 ! M62801 Salmonella gallinarum Salmon... 147 348 370 402.6 1.1e-14  
 SW:AROA\_SALGL  
 ! P22299 salmonella gallinarum. 3-pho... 147 348 370 402.6 1.1e-14  
 GP\_BCT1:ST5E3PS\_1  
 ! X54545 Salmonella typhi S.typhi gen... 147 348 361 393.0 3.7e-14  
 SW:AROA\_SALTI  
 ! P19786 salmonella typhi. 3-phosphos... 147 348 361 393.0 3.7e-14  
 PIR2:S12096  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 147 348 361 393.0 3.7e-14  
 SW:AROA\_PASMU  
 ! Q04570 Pasteurella multocida. 3-pho... 124 283 344 374.5 4e-13  
 PIR2:S28063  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 124 283 344 374.5 4e-13  
 GP\_BCT1:PMAROA\_1  
 ! Z14100 Pasteurella multocida P.mult... 124 283 344 374.5 4e-13  
 GP\_BCT2:AE000999\_8  
 ! AE000999 Archaeoglobus fulgidus Arc... 103 283 323 352.3 6.8e-12  
 PIR2:H69436  
 ! 5-enolpyruvylshikimate 3-phosphate ... 103 283 323 352.3 6.8e-12  
 SP\_ARC:O28775  
 ! O28775 archaeoglobus fulgidus. 5-en... 103 283 323 352.3 6.8e-12  
 GP\_BCT2:AE001310\_1  
 ! AE001310 Chlamydia trachomatis Chla... 67 202 311 339.0 3.7e-11  
 PIR2:G71522  
 ! probable phosphoshikimate vinyltran... 67 202 311 339.0 3.7e-11  
 SP\_PRO:O84371  
 ! O84371 chlamydia trachomatis. phosp... 67 202 311 339.0 3.7e-11  
 GP\_BCT1:APL012748\_1  
 ! AJ012748 Actinobacillus pleuropneum... 137 164 290 318.3 5.3e-10  
 SP\_PRO:Q9ZEQ0  
 ! Q9zeq0 actinobacillus pleuropneumon... 137 164 290 318.3 5.3e-10  
 PIR2:S18354  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 105 161 289 317.1 6.2e-10  
 SW:ARO2\_TOBAC  
 ! P23281 nicotiana tabacum (common to... 105 161 289 317.1 6.2e-10  
 GP\_PLN1:TOBEPSPS2\_1  
 ! M61905 Nicotiana tabacum N.tabacum ... 105 161 289 317.1 6.2e-10  
 SP\_PLN:O80428  
 ! O80428 oryza sativa (rice). 3-phosp... 102 178 288 315.1 8.1e-10  
 GP\_PLN1:AB016765\_1  
 ! AB016765 Oryza sativa Oryza sativa ... 102 178 288 315.1 8.1e-10  
 GP\_BCT2:U67500\_3  
 ! U67500 Methanococcus jannaschii Met... 156 294 280 305.9 2.6e-09  
 SW:AROA\_METJA  
 ! Q57925 methanococcus jannaschii. pu... 156 294 280 305.9 2.6e-09  
 PIR2:F64362  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 156 294 280 305.8 2.7e-09  
 GP\_BCT1:AP000060\_34

! AP000060 Aeropyrum pernix Aeropyrum... 130 249 275 300.6 5.2e-09  
 GP\_BCT1:MSGARO\_1  
 ! M62708 Mycobacterium tuberculosis M... 76 222 264 288.4 2.5e-08  
 GP\_BCT1:MTY20B11\_2  
 ! Z95121 Mycobacterium tuberculosis M... 76 222 264 288.4 2.5e-08  
 SW:AROA\_MYCTU  
 ! P22487 mycobacterium tuberculosis. ... 76 222 264 288.4 2.5e-08  
 GP\_BCT1:MTAROA\_1  
 ! X52269 Mycobacterium tuberculosis M... 76 222 264 288.4 2.5e-08  
 PIR2:E70590  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 76 222 264 288.4 2.5e-08  
 GP\_PLN1:TOBEPSPS1\_1  
 ! M61904 Nicotiana tabacum N.tabacum ... 134 272 250 272.4 1.9e-07  
 SW:ARO1\_TOBAC  
 ! P23981 nicotiana tabacum (common to... 134 272 250 272.4 1.9e-07  
 PIR2:S18353  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 134 272 250 272.4 1.9e-07  
 GP\_BCT1:SSAROA\_1  
 ! X72784 Synechocystis PCC6803 Synech... 173 214 238 270.9 2.4e-07  
 SW:ARO1\_PNECA  
 ! Q12659 p pentafunctional arom polyp... 56 122 255 270.5 2.5e-07  
 GP\_PLN1:PMCAROMX\_1  
 ! L18918 Pneumocystis carinii Pneumoc... 56 122 255 270.5 2.5e-07  
 PIR1:XUTOVS  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 133 243 243 264.9 5.1e-07  
 GP\_PLN1:TOMAROA\_1  
 ! M21071 Lycopersicon esculentum Toma... 133 243 241 262.7 6.7e-07  
 SW:AROA\_LYCES  
 ! P10748 lycopersicon esculentum (tom... 133 243 241 262.7 6.7e-07  
 PIR2:I39539  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 134 162 237 259.7 9.8e-07  
 GP\_BCT1:AEOROA\_1  
 ! L05002 Aeromonas salmonicida Aeromo... 134 162 237 259.7 9.8e-07  
 SW:AROA\_AERSA  
 ! Q03321 aeromonas salmonicida. 3-pho... 134 162 237 259.7 9.8e-07  
 SW:AROA\_PETHY  
 ! P11043 petunia hybrida (petunia). 3... 136 279 238 259.6 1e-06  
 PIR1:XUPJVS  
 ! 3-phosphoshikimate 1-carboxyvinyltr... 136 279 238 259.6 1e-06  
 GP\_PLN1:PETAROA\_1  
 ! M21084 Petunia x hybrida P.hybrida ... 136 279 238 259.6 1e-06  
 SP\_PRO:Q9Z470  
 ! Q9z470 corynebacterium glutamicum. ... 97 226 237 258.7 1.1e-06  
 GP\_BCT2:AF114233\_1  
 ! AF114233 Corynebacterium glutamicum... 97 226 237 258.7 1.1e-06  
 GP\_PAT:A59345\_1  
 ! A59345 Zea mays Sequence 2 from Pat... 136 300 230 252.0 2.7e-06  
 SP\_PLN:O24566  
 ! O24566 ze mays (maize). 3-phosphos... 136 300 230 252.0 2.7e-06  
 GP\_PLN1:ZMEPSPS\_1  
 ! X63374 Zea mays Z.mays mRNA for EPS... 136 300 230 252.0 2.7e-06  
 GP\_PAT:A59404\_1  
 ! A59404 Zea mays Sequence 2 from Pat... 136 300 230 252.0 2.7e-06  
 GP\_PAT:A69535\_1  
 ! A69535 unidentified Sequence 3 from... 136 300 230 251.9 2.7e-06  
 PIR1:XUMUVS

! 3-phosphoshikimate 1-carboxyvinyltr...	125	241	229	249.8	3.5e-06	! S76668 Pneumocystis carinii arom=en...	69	69	144	168.8	0.11
GP_PAT:A18838_1						SP_FUN:Q12733					
! A18838 Aeromonas salmonicida aroA g...	124	152	227	249.0	3.9e-06	! Q12733 pneumocystis carinii. 3-phos...	69	69	144	168.8	0.11
GP_PLN1:BNPEPSG_1						GP_PLN2:PCU31056_1					
! X51475 Brassica napus Brassica napu...	124	268	228	248.8	4e-06	! U31056 Pneumocystis carinii f. sp. ...	69	69	144	168.8	0.11
SW:AROA_BRANA						GP_BCT1:BSUB0019_207					
! P17688 brassica napus (rape). 3-pho...	124	268	228	248.8	4e-06	! Z99122 Bacillus subtilis Bacillus s...	45	73	149	165.2	0.18
PIR2:S12744						PIR2:G32354					
! 3-phosphoshikimate 1-carboxyvinyltr...	124	268	228	248.8	4e-06	! UDP-N-acetylglucosamine 1-carboxyvi...	45	73	149	165.2	0.18
GP_PLN2:ATAC002387_18						GP_BCT1:BSUB0020_7					
! AC002387 Arabidopsis thaliana Arabi...	125	241	227	247.7	4.6e-06	! Z99123 Bacillus subtilis Bacillus s...	45	73	149	165.2	0.18
SP_PLN:O22142						GP_BCT1:BSDNA320D_15					
! O22142 arabidopsis thaliana (mouse-...	125	241	227	247.7	4.6e-06	! Z49782 Bacillus subtilis B.subtilis...	45	73	149	165.2	0.18
GP_PLN1:ATEPSPS_1						SW:MURA_BACSU					
! X06613 Arabidopsis thaliana Arabido...	125	241	227	247.7	4.6e-06	! P19670 bacillus subtilis. probable ...	45	73	149	165.2	0.18
SW:AROA_ARATH						SP_FUN:Q12654					
! P05466 arabidopsis thaliana (mouse-...	125	241	227	247.7	4.6e-06	! Q12654 pneumocystis carinii. 3-phos...	55	55	140	164.5	0.2
GP_PAT:A59347_1						GP_PLN2:PCU31053_1					
! A59347 Zea mays Sequence 4 from Pat...	136	276	225	246.6	5.3e-06	! U31053 Pneumocystis carinii f. sp. ...	55	55	140	164.5	0.2
GP_PAT:A59406_1						GP_PLN2:PCU31054_1					
! A59406 Zea mays Sequence 4 from Pat...	136	276	225	246.6	5.3e-06	! U31054 Pneumocystis carinii f. sp. ...	55	86	138	162.4	0.26
GP_PAT:A69537_1						SP_FUN:Q12655					
! A69537 unidentified Sequence 5 from...	136	276	225	246.6	5.3e-06	! Q12655 pneumocystis carinii. 3-phos...	55	86	138	162.4	0.26
GP_BCT2:BUHTPIA_5						PIR2:S76347					
! L43549 Buchnera aphidicola Buchnera...	103	298	223	244.7	6.8e-06	! UDP-N-acetylglucosamine 1-carboxyvi...	79	105	138	153.2	0.84
SW:AROA_BUCAP						GP_BCT1:SYCSLRB_90					
! Q59178 buchnera aphidicola. 3-phosp...	103	298	223	244.7	6.8e-06	! D64000 Synechocystis sp. Synechocys...	79	105	138	153.2	0.84
GP_BCT1:AP000061_266						SW:MURA_SYNY3					
! AP000061 Aeropyrum pernix Aeropyrum...	82	105	210	230.7	4e-05	! Q55673 synechocystis sp. (strain pc...	79	105	138	153.2	0.84
GP_PLN1:SCARO1_1						GP_PLN2:PCU31055_1					
! X06077 Saccharomyces cerevisiae Yea...	82	122	212	224.3	9.3e-05	! U31055 Pneumocystis carinii f. sp. ...	55	55	125	148.4	1.6
GP_PLN1:SC9302X_2						SP_FUN:Q12656					
! Z48179 Saccharomyces cerevisiae S.c...	82	122	212	224.3	9.3e-05	! Q12656 pneumocystis carinii. 3-phos...	55	55	125	148.4	1.6
PIR1:BVBAY1						PIR2:S71425					
! ARO1 protein - yeast (Saccharomyces...	82	122	212	224.3	9.3e-05	! DNA-directed RNA polymerase (EC 2.7...	56	56	131	147.6	1.7
SW:ARO1_YEAST						SP_PRO:P74963					
! P08566 s pentafunctional arom polyp...	82	122	212	224.3	9.3e-05	! P74963 shewanella sp. rna polymeras...	56	56	131	147.6	1.7
GP_PLN1:ANAROMG_1						GP_BCT1:D83194_1					
! X05204 Emericella nidulans Aspergil...	110	158	210	222.1	0.00012	! D83194 Shewanella sp. Barophilic ba...	56	56	131	147.6	1.7
PIR1:BVASA1						GP_BCT1:ECU31523_1					
! aroM protein - Emericella nidulans	110	158	210	222.1	0.00012	! U31523 Escherichia coli Escherichia...	47	80	114	141.9	3.6
SW:ARO1_EMENI						\\End of List					
! P07547 e pentafunctional arom polyp...	110	158	210	222.1	0.00012						
SP_PRO:Q926M0											
! Q926M0 chlamydia pneumoniae. phosph...	115	297	180	198.2	0.0026						
GP_BCT2:AE001684_6											
! AE001684 Chlamydomonada pneumoniae C...	115	297	180	198.2	0.0026						
SP_FUN:Q12430											
! Q12430 pneumocystis carinii. epsp s...	91	91	154	179.6	0.029						
GP_PLN2:PCU31057_1											
! U31057 Pneumocystis carinii f. sp. ...	91	91	154	179.6	0.029						
GP_PLN2:S76670_1											
! S76670 Pneumocystis carinii arom=en...	91	91	154	179.6	0.029						
GP_BCT1:CNSPAX02_158											
! AJ248284 Pyrococcus abyssi Pyrococc...	79	268	153	169.8	0.1						
GP_PLN2:S76668_1											