SUMMARY

(In accordance with 40 CFR part 152, this summary is available
for public release after registration)

STUDY TITLE

Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Soybean Event DAS-68416-4 for Potential Protein Allergenicity and Toxicity

DATA REQUIREMENTS

N/A

AUTHOR(S)

P. Song

STUDY COMPLETED ON

28 – June – 2010

PERFORMING LABORATORY

Regulatory Sciences and Government Affairs—Indianapolis Lab
Dow AgroSciences LLC
9330 Zionsville Road
Indianapolis, Indiana 46268-1054

LABORATORY STUDY ID

101711

Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Soybean Event DAS-68416-4 for Potential Protein Allergenicity and Toxicity

SUMMARY

A plant-optimized *aad*-12 gene, originally from the soil bacterium *Delftia acidovorans*, was integrated into soybean (*Glycine max*) by *Agrobacterium*-mediated transformation of a variety "Maverick" with plasmid pDAB4468 to produce event DAS-68416-4. Aryloxyalkanoate dioxygenase-12 (AAD-12 protein), encoded by the *aad*-12 gene, provides tolerance to the herbicide 2,4-dicholorophenoxyacetic acid (2,4-D). Molecular characterization indicated that the event DAS-68416-4 contained a single insert including two intact expression cassettes, AAD-12 and PAT. DNA sequences flanking the insert in event DAS-68416-4 soybean were also cloned and characterized. The DNA sequence of the insert is identical to the corresponding portion in the T-DNA insert of plasmid pDAB4468 except for an extra 9 bp insertion at the 3' junction. All the junctions across the insert and its flanking borders were identified and screened for "novel" reading frames spanning the junction sites. A total of 12 "novel" reading frames were identified and 8 of them were evaluated for potential allergenicity and toxicity using bioinformatics tools since 4 of them are only 4 amino acids long. Searches of those putative "novel" reading frames against a peer reviewed allergen database (FARRP Allergen Database Version 10, Released in January, 2010) did not generate any significant amino acid sequence similarities with known allergens. Similarly, the search against the GenBank non-redundant protein sequences "nr" did not detect any protein sequence similarity with toxic proteins harmful to humans or animals.

STUDY TITLE

Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Soybean Event DAS-68416-4 for Potential Protein Allergenicity and Toxicity

DATA REQUIREMENTS

N/A

AUTHOR(S)

P. Song 317-337-3434
[psong@dow.com]

STUDY COMPLETED ON

28 – June – 2010

PERFORMING LABORATORY

Regulatory Sciences and Government Affairs—Indianapolis Lab
Dow AgroSciences LLC
9330 Zionsville Road
Indianapolis, Indiana 46268-1054

LABORATORY STUDY ID

101711

STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS

Compound: Soybean AAD-12 event DAS-68416-4

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Soybean Event DAS-68416-4 for Potential Protein Allergenicity and Toxicity

- STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS:

  No claim of confidentiality, on any basis whatsoever, is made for any information contained in this document. I acknowledge that information not designated as within the scope of FIFRA sec. 10(d)(1)(A), (B), or (C) and which pertains to a registered or previously registered pesticide is not entitled to confidential treatment and may be released to the public, subject to the provisions regarding disclosure to multinational entities under FIFRA sec. 10(g).

Company: Dow AgroSciences LLC

Company Agent: M. S. Krieger

Title: Regulatory Manager

Signature:

Date: 24 June 2010

THIS DATA MAY BE CONSIDERED CONFIDENTIAL IN COUNTRIES OUTSIDE THE UNITED STATES.

STATEMENT OF COMPLIANCE WITH GOOD LABORATORY PRACTICE STANDARDS

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Soybean Event DAS-68416-4 for Potential Protein Allergenicity and Toxicity

Study Initiation Date: 26/04/2010

This report represents data generated after the effective date of the EPA FIFRA Good Laboratory Practice Standards.

United States Environmental Protection Agency
Title 40 Code of Federal Regulations Part 160
FEDERAL REGISTER, August 17, 1989

Organisation for Economic Co-Operation and Development
ENV/MC/CHEM(98)17, Paris January 26, 1998

At the time this study was conducted, it was not subject to the Good Laboratory Practice Standards and was, therefore, not monitored by the quality assurance unit.

24 June 2010

M. S. Krieger
Sponsor
Dow AgroSciences LLC

Date

24 June 2010

M. S. Krieger
Submitter
Dow AgroSciences LLC

Date

Song Ping

28-June-2010

P. Song
Study Director/Author
Dow AgroSciences LLC

Study Completion Date

QUALITY ASSURANCE STATEMENT

Compound: Soybean AAD-12 event DAS-68416-4

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Soybean Event DAS-68416-4 for Potential Protein Allergenicity and Toxicity

Study Initiation Date: 26/04/2010 Study Completion Date: 28/06/2010

NON-GLP STUDY

SIGNATURE PAGE

Song Ping

P. Song
Author
Dow AgroSciences LLC

28-June-2010
Date

M. Zhuang
Peer Reviewer
Dow AgroSciences LLC

June-23-2010
Date

R. A. Herman
Peer Reviewer
Dow AgroSciences LLC

23-June-2010
Date

G. Shan
Science Leader
Dow AgroSciences LLC

23 June 2010
Date

Kathryn A Clayton

K.A. Clayton
Global Leader, Biotechnology Regulatory Sciences
Dow AgroSciences LLC

28 JUNE 2010
Date

STUDY PERSONNEL

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Soybean Event DAS-68416-4 for Potential Protein Allergenicity and Toxicity

Principal Analyst: N/A
(Principle Investigator)

Analysts: N/A

TABLE OF CONTENTS

Bioinformatics Evaluation of the Putative Reading Frames across the Junctions in Soybean Event DAS-68416-4 for Potential Protein Allergenicity and Toxicity

ABSTRACT

A plant-optimized *aad*-12 gene, originally from the soil bacterium *Delftia acidovorans*, was integrated into soybean (*Glycine max*) by *Agrobacterium*-mediated transformation of a variety "Maverick" with plasmid pDAB4468 to produce event DAS-68416-4. Aryloxyalkanoate dioxygenase-12 (AAD-12 protein), encoded by the *aad*-12 gene, provides tolerance to the herbicide 2,4-dicholorophenoxyacetic acid (2,4-D). Molecular characterization indicated that the event DAS-68416-4 contained a single insert including two intact expression cassettes, AAD-12 and PAT. DNA sequences flanking the insert in event DAS-68416-4 soybean were also cloned and characterized. The DNA sequence of the insert is identical to the corresponding portion in the T-DNA insert of plasmid pDAB4468 except for an extra 9 bp insertion at the 3' junction. All the junctions across the insert and its flanking borders were identified and screened for "novel" reading frames spanning the junction sites. A total of 12 "novel" reading frames were identified and 8 of them were evaluated for potential allergenicity and toxicity using bioinformatics tools since 4 of them are only 4 amino acids long. Searches of those putative "novel" reading frames against a peer reviewed allergen database (FARRP Allergen Database Version 10, Released in January, 2010) did not generate any significant amino acid sequence similarities with known allergens. Similarly, the search against the GenBank non-redundant protein sequences "nr" did not detect any protein sequence similarity with toxic proteins harmful to humans or animals.

INTRODUCTION

A plant-optimized *aad*-12 gene, originally from the soil bacterium *Delftia acidovorans*, was integrated into soybean (*Glycine max*) by *Agrobacterium*-mediated transformation of a variety "Maverick" with plasmid pDAB4468 to produce event DAS-68416-4. Aryloxyalkanoate dioxygenase-12 (AAD-12 protein), encoded by the *aad*-12 gene, provides tolerance to the herbicide 2,4-dicholorophenoxyacetic acid (2,4-D). Molecular characterization indicated that the event DAS-68416-4 contained a single insert containing two intact expression cassettes, AAD-12 and PAT (1). DNA sequences flanking the insert in event DAS-68416-4 soybean have been cloned and characterized (2). The DNA sequence of the insert is identical to the corresponding portion in the T-DNA insert of plasmid pDAB4468 except for an extra 9 bp insertion at the 3' junction.

Theoretically, the DNA sequences surrounding the junctions across the insert and its flanking borders could create potential "novel" reading frames spanning those junction sites. In the safety assessment of transgenic crops, one of the concerns is that "novel" proteins could be expressed that might have a potential to elicit allergic or toxic reactions in humans. Therefore, the potential novel reading frames spanning the junctions of the insert and border sequences can be analyzed for sequence similarity to known allergens or toxins as an indication of a safety concern should those reading frames actually be expressed. For this study, "novel" reading frames are defined very conservatively as any reading frame spanning the junctions regardless of the presence of a start codon and the number of amino acid residues.

To assess potential allergenicity using bioinformatics tools, two criteria for evaluating structural similarities between query proteins and known allergens are currently used based on amino acid sequence alignments (3, 4, 5). The first criterion is a search over 80-amino-acid stretches (sliding window search) to detect >35% identity between a query protein and known allergens. The window size of 80 amino acids was selected to correspond with a typical domain size in a protein, and recognizes that single protein domains may contain epitopes that mediate antibody binding. The second criterion involves evaluating short amino-acid stretches for identity between the query protein and known allergens. As stated in the report of Codex Ad Hoc Working Group on Allergenicity (3), “the size of the contiguous amino acid search should be based on a scientifically justified rationale in order to minimize the potential for false negative or false positive results”. Window sizes of 6 to 8 amino acids have been suggested based on hypothetical epitope sizes, however, use of window sizes of less than 8 amino acids have been largely abandoned based on the high probability of random alignments that are of no predictive value (6, 7). The use of any short-alignment criteria for predicting the allergenic potential of proteins has also been recently criticized (8, 9, 10, 11). For evaluation of potential protein toxicity, structure similarity between a query protein and known protein toxins are identified using local sequence alignment search tools such as BLAST and FASTA algorithms against a database of all available protein sequences.

The purpose of this study is to identify the potential “novel” reading frames at the junction of the border and insert sequences of event DAS-68416-4 and evaluate them for potential allergenicity and toxicity using bioinformatics tools along with updated allergen and non-redundant protein databases.

METHODS

Search for “Novel” Reading Frames

DNA sequences including the whole insert and its border regions of soybean event DAS-68416-4 were analyzed with an in-house Perl script to search six-frame translations from stop codon to stop codon across all the identified junctions (Figure 1). For each reading frame (RF), the exact locations of 5’ and 3’ stop codons were identified. Those RFs spanning the junctions were considered as “novel” and evaluated for any potential protein allergenicity or toxicity using bioinformatics tools.

Query Sequence Preparation

Each putative reading frame sequence was prepared in FASTA format for the use of FASTA and BLASTp search programs.

Allergenicity Assessment

For the allergenicity assessment, the amino acid sequence of each RF was compared with a peer-reviewed database containing 1471 known and putative allergens as well as celiac-induction protein sequences residing in the FARRP dataset (Version 10, Released in January 2010, University of Nebraska, http://www.allergenonline.org/). Potential identities between the RF peptide sequences and proteins in the allergen database were evaluated with the FASTA program (v34) using the default algorithm parameters (Matrix = BLOSUM50; Expect = 10; Gap Penalties = -12/-2; *ktup*=2). The FASTA search was run by an in-house Perl script in a UNIX computer with a Linux operation system. If a query sequence is longer than 80 amino acids, the script parses the query sequence into a complete (overlapping) set of 80 amino acid long fragments and each fragment is subjected to a FASTA search. A greater than 35% identity threshold over any 80 or more amino acid sequences between a query sequence and an allergen was used to indicate the potential for cross-reactivity. To ensure that high identity over a short stretch (for example, 80% over 60 amino acids) will not be overlooked, a calculation, (Identity% × number of overlapped amino acids)/80, was implemented as a conversion to check the criteria of >35% over 80 amino acids when the FASTA alignment (overlapped amino acids) is less than 80 amino acids. Reading frames shorter than 29 amino acids were not evaluated using FASTA search since >35% identity requires at least a match of 29 amino acids over 80 amino acids. RF peptide sequences were also screened for any matches of 8 contiguous amino acids to the allergens contained in the database noted above as long as an RF is equal to or longer than 8 amino acids. This was done using an in-house Perl script that generates all sequentially possible (overlapping) 8-residue peptides from a query protein, followed by Fuzzpro program (Emboss Package v2.10.0) search that compares each query "word" with all allergen sequences in the database for perfect matches.

Toxicity Assessment

To assess potential toxicity of the *in silico* translated peptides of the RFs, a similarity search was conducted using the BLASTp algorithm. Reading frames were queried using the BLASTp 2.2.21 algorithm against non-redundant protein sequences "nr" (update to April 23, 2010), which incorporates non-redundant entries from all GenBank and RefSeq nucleotide translations, including non-redundant GenBank CDS translation along with protein sequences from SWISS-PROT (http://www.expasy.org/sprot/), PIR (http://pir.georgetown.edu/), PRF (http://www.prf.or.jp/aboutdb-e.html), and PDB (http://www.wwpdb.org/). BLASTp searches were done in the NCBI (National Center of Biotechnology Information) BLAST website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). BLASTp searches were performed in an internal UNIX computer using the default setting of algorithm parameters (Matrix = BLOSUM 62, Gap Costs: Existance: 11, Extension: 1, Word Size=3) except that the low complexity filter was off and Expectation =1. Although a statistically significant sequence similarity generally requires an alignment with an expectation value less than 0.01, a threshold of E-value < 1.0 ensures that proteins with even limited similarity will not be overlooked in the search (12).

RESULTS AND CONCLUSIONS

A total of 12 reading frames spanning the junctions across the insert and its border regions in the event DAS-68416-4 were identified (Table 1). Four of them are only 4 amino acids long, which are too short to be analyzed. When the amino acid sequences of the 8 reading frames were compared with the FARRP allergen dataset (Version 10, January 2010), no matches of eight or greater contiguous amino acids were observed in any of the translated sequences. Of those 8 reading frames, only RF 1_+2, 1_-1, and 2_+2 were subject to search against the allergen database using the FASTA program since the rest of the reading frames are less than 29 amino acids. No over threshold identities (greater than 35% identity over greater than or equal to 80 amino acid residues) were detected in the FASTA search outputs when using the peptide sequences from the 3 RFs as query (Table 2, Appendix 1).

When the 8 reading frames were subjected to BLASTp search against the GenBank non-redundant protein dataset, no alignments with E-values less than 1 were returned (Table 2, Appendix 2).

In conclusion, bioinformatics evaluation of the 8 putative “novel” reading frames did not generate any significant amino acid sequence similarities with known allergens or toxic proteins that are harmful to humans or animals.

REFERENCES

1. Song, P., Cruse, J., Thomas, A., 2009. Molecular Characterization of AAD-12 Soybean Event DAS-68416-4. Dow AgroSciences Study Report 081087.

2. Poorbaugh, J., Zhou, N., Mo, J., 2009. Cloning and Characterization of DNA Sequence in the Insert and the Flanking Border Regions of AAD-2 Soybean Event DAS-68416-4. Dow AgroSciences Study Report 091048.

3. Codex Alimentarius Special Publications, FOODS DERIVED FROM MODERN BIOTECHNOLOGY (Second Edition), 2009; GUIDELINE FOR THE CONDUCT OF FOOD SAFETY ASSESSMENT OF FOODS DERIVED FROM RECOMBINANT-DNA PLANTS; ANNEX 1: ASSESSMENT OF POSSIBLE ALLERGENICITY, 22-27.

4. Ladics G. S. 2008. Current Codex guidelines for assessment of potential protein allergenicity. Food Chem Toxicol 2008, 46:S20-S23.

5. FAO/WHO (World Health Organization): Evaluation of Allergenicity of Genetically Modified Foods. Report of Joint FAO/WHO Expert Consultation. Rome: Food and Agriculture Organization of the United Nations. 2001

6. Silvanovich A, Nemeth M.A., Song P, Herman R, Tagliani L, Bannon, G. A. 2006. The value of short amino acid sequence matches for prediction of protein allergenicity. Tox Sci., 90:252-258.

7. Stadler M. B., Stadler, B. M. 2003. Allergenicity prediction by protein sequence. FASEB J., 17:1141-1143.

8. Goodman R. E., Vieths S., Sampson H. A., Hill D., Ebisawa M., Taylor S. L., van Ree R. 2008. Allergenicity assessment of genetically modified crops – what makes sense? Nat Biotech, 26:73-81.

9. Thomas K., Herouet-Guicheney C., Ladics G., McClain S., MacIntosh S., Privalle L., Woolhiser M., 2008. Current and future methods for evaluating the allergenic potential of proteins: International workshop report 23-25 October 2007. Food Chem Tox, 46:3219-3225.

10. Cressman R. F., Ladics G., 2009. Further evaluation of the utility of “sliding window” FASTA in predicting cross-reactivity with allergenic proteins. Regul Toxicol Pharmacol, 54:S20-S25.

11. Herman R., Song P, ThirumalaiswamySekhar, A., 2009. Value of eight-amino-acid matches in predicting the allergenicity status of proteins: an empirical bioinformatic investigation. Clinical and Molecular Allergy, 7:9.

12. Pearson W. R., 2000. Flexible sequence similarity searching with the FASTA3 program package. Methods Mol Biol 132:185-219.

Table 1. Deduced Amino Acid Sequences of Reading Frames across the Junctions in Soybean Event DAS-68416-4

Reading Frame Name (Junction_Frame)	Nucleotide location	Number of amino acids	Deduced amino acid sequence
1_+1	2719−2766	16	IIQAPVSIITPKVRPE_
1_+2	2705−2818	38	KFIFKSFKHQSASSHQKLGPNSLKLESSQLRSTGQIRS_
1_+3	2700−2756	19	IKNLFLNHSSTSQHHHTKS_
1_-1	2817−2704	38	ERIWPVDLNCELSNFKLFGPNFWCDDADWCLNDLKINF_
1_-2	2798−2727	24	PSLYSASRYGAVDCNSMVRAGRTREG_
1_-3	2734−2723	4	LVLE_
2_+1	9082−9132	17	KRPQCVIKLSKRQYFNS_
2_+2	9068−9154	29	LQYIKNVRNVLLSCLSVNILILNNQYFNS_
2_+3	9114−9125	4	ASIF_
2_-1	9138−9073	22	LLRIKILTLRQLNNTLRTFLMY_
2_-2	9131−9120	4	ELKY_
2_-3	9124−9113	4	NIDA_

Table 2. Summary of Results from BLASTp Search for Sequence Similarities of Putative Reading Frames across Junctions in Soybean Event DAS-68416-4

Reading Frame (Junction_Frame)	**Length (aa)**	**Match of 8 or more residues with known allergens**	**FASTA Search (>35% identity over ≥ 80 residues)**	**Number of BLASTp hits (E()<1)**
1_+1	16	No	N/A	0
1_+2	38	No	No	0
1_+3	19	No	N/A	0
1_-1	38	No	No	0
1_-2	24	No	N/A	0
1_-3	4	N/A	N/A	N/A
2_+1	17	No	N/A	0
2_+2	29	No	No	0
2_+3	4	N/A	N/A	N/A
2_-1	22	No	N/A	0
2_-2	4	N/A	N/A	N/A
2_-3	4	N/A	N/A	N/A
1_+1	16	No	N/A	N/A

N/A = Not applicable;

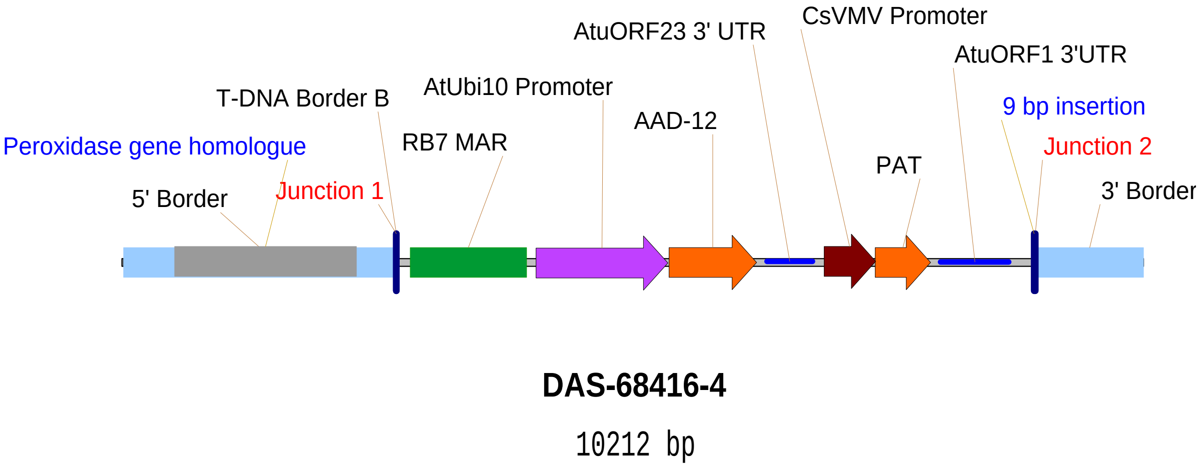

Figure 1. Diagram of the Insert, its Flanking Borders, and Junction Sites in Soybean Event DAS-68416-4

APPENDIX

1. FASTA Search Outputs of the Putative Reading Frames (>28 aa) in Soybean Event DAS-68416-4 against the Allergen Database V10

   There were 3 reading frames (>28 aa). FASTA search output files are electronically stored in a secured computer in Dow AgroSciences and are available for view in PDF format.

2. BLASTp Search Outputs Using Putative Reading Frames as Queries in Soybean Event DAS-68416-4

   BLASTp search output files of the 8 putative reading frames are electronically stored in a secured computer in Dow AgroSciences and are available for view in PDF format.

FASTA Search Outputs of the Putative Reading Frames (>28 aa) in Soybean Event DAS-68416-4 against the Allergen Database V10

```
RF_1_-1
# fasta -Q -d 500 -E 10 fasta_input.txt /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta 1
FASTA searches a protein or DNA sequence data bank
 version 3.4t26 July 7, 2006
Please cite:
 W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

Query library fasta_input.txt vs /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library
searching /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

  1>>>RF_1_-1 38 aa - 38 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

       opt      E()
< 20     2     0:=
  22     0     0:          one = represents 3 library sequences
  24     0     0:
  26     0     0:
  28     2     0:=
  30     4     2:*=
  32     7     8:==*
  34    22    21:======*=
  36    21    44:=======       *
  38    36    72:============           *
  40    99   101:=================================*
  42    96   123:================================        *
  44    96   136:================================             *
  46   171   138:=============================================*===========
  48   145   132:===========================================*=====
  50   150   121:========================================*=========
  52    99   106:=================================  *
  54    76    91:==========================    *
  56    84    76:=========================*==
  58    89    62:====================*=========
  60    29    50:==========      *
  62    40    40:=============*
  64    40    32:==========*===
  66    28    25:========*=
  68    30    20:======*===
  70    19    16:=====*=
  72    14    12:===*=
  74    15    10:===*=
  76    13     7:==*==
  78    15     6:=*===
  80    13     4:=*===
  82     1     3:*
  84     5     3:*=
  86     2     2:*
  88     3     2:*          inset = represents 1 library sequences
  90     0     1:*
  92     3     1:*         :*==
  94     1     1:*         :*
  96     0     1:*         :*
  98     0     0:          *
 100     0     0:          *
 102     0     0:          *
 104     0     0:          *
 106     0     0:          *
 108     0     0:          *
 110     0     0:          *
 112     0     0:          *
 114     0     0:          *
 116     0     0:          *
 118     0     0:          *
>120     1     0:=         *=
 331323 residues in  1471 sequences
  Expectation_n fit: rho(ln(x))= 4.0707+/-0.00321; mu= 5.7296+/- 0.165
 mean_var=31.6493+/- 7.810, 0's: 2 Z-trim: 3  B-trim: 52 in 1/42
 Lambda= 0.227978
 Kolmogorov-Smirnov  statistic: 0.0838 (N=29) at  44

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
 join: 42, opt: 30, open/ext: -10/-2, width:  32
 Scan time:  0.030
The best scores are:                                      opt bits E(1471)
gi|188572343|gb|ACD65081.1| eukaryotic translation ( 325)   70 27.8   0.077
gi|73535415|pdb|1WKX|A Chain A, Crystal Structure  (  43)   46 19.7     2.8
gi|3367714|emb|CAA08836.1| BDAI-1; Barley dimeric  ( 152)   50 21.2     3.6
gi|146737976|gb|ABQ42566.1| thaumatin-like protein ( 201)   51 21.5     3.7
gi|71057064|emb|CAI38795.2| thaumatin-like protein ( 225)   51 21.5     4.1
gi|204324083|gb|ACI01048.1| arginine kinase [Bomby ( 355)   51 21.6     6.3
gi|29292272|emb|CAD82911.1| precursor Can f II [Ca ( 177)   48 20.5     6.6
gi|29292274|emb|CAD82912.1| precursor Can f II [Ca ( 179)   48 20.5     6.6
gi|3121746|sp|O18874.1|ALL2_CANFA RecName: Full=Mi ( 180)   48 20.5     6.7
gi|76782247|gb|ABA54897.1| hydrophobic seed protei ( 134)   46 19.8       8
gi|9087163|sp|Q96385.1|MPAC1_CHAOB RecName: Full=M ( 375)   50 21.3     8.4
gi|15886861|emb|CAC85911.1| arginine kinase [Plodi ( 355)   49 20.9      10

>>gi|188572343|gb|ACD65081.1| eukaryotic translation ini  (325 aa)
 initn:  61 init1:  61 opt:  70  Z-score: 122.4  bits: 27.8 E(): 0.077
Smith-Waterman score: 70;  36.364% identity (60.606% similar) in 33 aa overlap (3-34:36-67)

                                           10         20        30
RF_1_-                             ERIWPVDLNCE-LSNFKLFGPNFWCDDADWCL
                                     .:  .:: : :..:.  .   :: :.::
gi|188 MQGHERAITQIKYNREGDLLFSCAKDHKPNVW-FSLNGERLGTFNGHAGAVWCVDVDWTT
```

```
          10        20        30        40        50        60

RF_1_- NDLKINF
       . :
gi|188 TKLITGSGDMSVRLWDVETGTSVACIPCKSSARTVGFSFSGNQAAYSTDRAMGHICELFV
           70        80        90       100       110       120

>>gi|73535415|pdb|1WKX|A Chain A, Crystal Structure Of A  (43 aa)
 initn:  46 init1:  46 opt:  46  Z-score: 94.4  bits: 19.7 E():  2.8
Smith-Waterman score: 46;  40.000% identity (60.000% similar) in 15 aa overlap (16-30:10-24)

               10        20        30
RF_1_- ERIWPVDLNCELSNFKLFGPNFWCDDADWCLNDLKINF
                      ::   :. :..  ::
gi|735       EQCGRQAGGKLCPDNLCCSQWGWCGSTDEYCSPDHNCQSNCKD
                     10        20        30        40

>>gi|3367714|emb|CAA08836.1| BDAI-1; Barley dimeric alph  (152 aa)
 initn:  50 init1:  50 opt:  50  Z-score: 92.3  bits: 21.2 E():  3.6
Smith-Waterman score: 50;  71.429% identity (85.714% similar) in 7 aa overlap (19-25:32-38)

                           10        20        30
RF_1_-             ERIWPVDLNCELSNFKLFGPNFWCDDADWCLNDLKINF
                                     :: .:::
gi|336 GAMWMKSMLLVLLLCMLMVTPMTGARSDNSGPWMWCDPEMGHKVSPLTRCRALVKLECVG
              10        20        30        40        50        60

gi|336 NRVPEDVLRDCCQEVANISNEWCRCGDLGSMLRSVYAALGVGGGPEEVFPGCQKDVMKLL
              70        80        90       100       110       120

>>gi|146737976|gb|ABQ42566.1| thaumatin-like protein [Ac  (201 aa)
 initn:  44 init1:  44 opt:  51  Z-score: 92.1  bits: 21.5 E():  3.7
Smith-Waterman score: 51;  47.368% identity (63.158% similar) in 19 aa overlap (9-27:158-175)

                                     10        20        30
RF_1_-                       ERIWPVDLNCELSNFKLFGPNFWCDDADWCLNDLKINF
                                     :: :.::. :  .  : ::
gi|146 GQCPNELRAPGGCNNPCTVFKTDQFCCNSGNCGLTNFSKFFKDR-CPDAYSYPKDDQTST
       130       140       150       160       170        180

gi|146 FTCPAGTNYKVVFCP
        190       200

>>gi|71057064|emb|CAI38795.2| thaumatin-like protein [Ac  (225 aa)
 initn:  44 init1:  44 opt:  51  Z-score: 91.3  bits: 21.5 E():  4.1
Smith-Waterman score: 51;  47.368% identity (63.158% similar) in 19 aa overlap (9-27:182-199)

                                     10        20        30
```

```
RF_1_-                       ERIWPVDLNCELSNFKLFGPNFWCDDADWCLNDLKINF
                                     :: :.::. :  .  : ::
gi|710 GQCPNELRAPGGCNNPCTVFKTDQYCCNSGNCGLTNFSKFFKDR-CPDAYSYPKDDQTST
             160       170       180       190        200       210

gi|710 FTCPAGTNYKVVFCP
              220

>>gi|204324083|gb|ACI01048.1| arginine kinase [Bombyx mo  (355 aa)
 initn:  39 init1:  39 opt:  51  Z-score: 88.0  bits: 21.6 E():  6.3
Smith-Waterman score: 51;  29.630% identity (55.556% similar) in 27 aa overlap (2-28:201-225)

                                            10        20        30
RF_1_-                              ERIWPVDLNCELSNFKLFGPNFWCDDADWCL
                                     :.::.  .   .. : :    ::.. :
gi|204 GMSKETQQQLIDDHFLFKEGDRFLQAANACRFWPTGRGIYHNENKTFL--VWCNEEDHLR
              180       190       200       210         220


RF_1_- NDLKINF

gi|204 IISMQMGGDLQQVYKRLVSAVNEIEKKIPFSHHDRLGFLTFCPTNLGTTVRASVHIKLPK
      230       240       250       260       270       280

>>gi|29292272|emb|CAD82911.1| precursor Can f II [Canis   (177 aa)
 initn:  32 init1:  32 opt:  48  Z-score: 87.7  bits: 20.5 E():  6.6
Smith-Waterman score: 48;  39.394% identity (48.485% similar) in 33 aa overlap (5-34:74-105)

                                         10           20        30
RF_1_-                           ERIWPVDLNCE---LSNFKLFGPNFWCDDADWCL
                                     : : .::   :. ::    : . :   :
gi|292 SDLIKPWGHFRVFIHSMSAKDGNLHGDILIPQDGQCEKVSLTAFKTATSNKF-DLEYWGH
            50        60        70        80        90        100


RF_1_- NDLKINF
       :::
gi|292 NDLYLAEVDPKSYLILYMINQYNDDTSLVAHLMVRDLSRQQDFLPAFESVCEDIGLHKDQ
            110       120       130       140       150       160

>>gi|29292274|emb|CAD82912.1| precursor Can f II [Canis   (179 aa)
 initn:  32 init1:  32 opt:  48  Z-score: 87.6  bits: 20.5 E():  6.6
Smith-Waterman score: 48;  39.394% identity (48.485% similar) in 33 aa overlap (5-34:76-107)

                                         10           20        30
RF_1_-                           ERIWPVDLNCE---LSNFKLFGPNFWCDDADWCL
                                     : : .::    :. ::    : . :   :
gi|292 SDLTKPWGHFRVFIHSMSAKDVNLHGDILIPQDGQCEKVSLTAFKTATSNKF-DLEYWGH
          50        60        70        80        90        100
```

```
RF_1_- NDLKINF
       :::
gi|292 NDLYLAEVDPKSYLILYMINQYNDDTSLVAHLMVRDLSRQQDFLPAFESVCEDIGLHKDQ
          110       120       130       140       150       160

>>gi|3121746|sp|O18874.1|ALL2_CANFA RecName: Full=Minor   (180 aa)
 initn:  32 init1:  32 opt:  48  Z-score: 87.6  bits: 20.5 E():  6.7
Smith-Waterman score: 48;  39.394% identity (48.485% similar) in 33 aa overlap (5-34:77-108)

                                         10           20        30
RF_1_-                           ERIWPVDLNCE---LSNFKLFGPNFWCDDADWCL
                                     : : .::   :. ::    : . :   :
gi|312 SDLIKPWGHFRVFIHSMSAKDGNLHGDILIPQDGQCEKVSLTAFKTATSNKF-DLEYWGH
         50        60        70        80        90        100


RF_1_- NDLKINF
       :::
gi|312 NDLYLAEVDPKSYLILYMINQYNDDTSLVAHLMVRDLSRQQDFLPAFESVCEDIGLHKDQ
         110       120       130       140       150       160

>>gi|76782247|gb|ABA54897.1| hydrophobic seed protein pr  (134 aa)
 initn:  41 init1:  41 opt:  46  Z-score: 86.1  bits: 19.8 E():    8
Smith-Waterman score: 46;  50.000% identity (62.500% similar) in 16 aa overlap (19-34:61-74)

                           10        20        30
RF_1_-             ERIWPVDLNCELSNFKLFGPNFWCDDADWCLNDLKINF
                                     ::.  : : . ::: :
gi|767 PPPPPCYVPAPLTPPPSLSPPPSLSPPPPSGPS--CPDLSVCLNILDGSPADDCCALIAD
               40        50        60          70        80

gi|767 LVDLEASVCLCIQLRVLGIVNLDLNLQLILNACGPSYPSNATCPRT
       90       100       110       120       130

>>gi|9087163|sp|Q96385.1|MPAC1_CHAOB RecName: Full=Major  (375 aa)
 initn:  50 init1:  50 opt:  50  Z-score: 85.8  bits: 21.3 E():  8.4
Smith-Waterman score: 50;  42.857% identity (64.286% similar) in 14 aa overlap (23-36:29-42)

                     10        20        30
RF_1_-       ERIWPVDLNCELSNFKLFGPNFWCDDADWCLNDLKINF
                                   :  ::.:  : .:.
gi|908 MASCTLLAVLVFLCAIVSCFSDNPIDSCWRGDANWDQNRMKLADCAVGFGSSAMGGKGGA
               10        20        30        40        50        60

gi|908 FYTVTSSDDDPVNPAPGTLRYGATRERSLWIIFSKNLNIKLNMPLYIAGNKTIDGRGAEV
               70        80        90        100       110       120
```

```
>>gi|15886861|emb|CAC85911.1| arginine kinase [Plodia in  (355 aa)
 initn:  39 init1:  39 opt:  49  Z-score: 84.4  bits: 20.9 E():   10
Smith-Waterman score: 49;  29.630% identity (51.852% similar) in 27 aa overlap (2-28:201-225)

                                            10        20        30
RF_1_-                              ERIWPVDLNCELSNFKLFGPNFWCDDADWCL
                                     :.::   .   .. : :    ::.. :
gi|158 GMSKETQQQLIDDHFLFKEGDRFLQAANACRFWPSGRGIYHNENKTFL--VWCNEEDHLR
              180       190       200       210         220


RF_1_- NDLKINF

gi|158 LISMQMGGDLKQVYKRLVRGVNDIAKRIPFSHNERLGFLTFCPTNLGTTVRASVHIKLPK
      230       240       250       260       270       280



38 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Mon Mar  1 23:40:37 2010 done: Mon Mar  1 23:40:38 2010
 Total Scan time:  0.030 Total Display time:  0.010
```

RF_1_+2

```
Function used was FASTA [version 3.4t26 July 7, 2006]
# fasta -Q -d 500 -E 10 fasta_input.txt /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta 1
FASTA searches a protein or DNA sequence data bank
 version 3.4t26 July 7, 2006
Please cite:
 W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

Query library fasta_input.txt vs /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library
searching /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

  1>>>RF_1_+2 38 aa - 38 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

       opt      E()
< 20     2     0:=
  22     0     0:          one = represents 3 library sequences
  24     0     0:
  26     0     0:
  28     1     0:=
  30     9     2:*==
  32     8     8:==*
  34    53    21:======*===========
  36    49    44:==============*==
  38    71    72:=======================*
  40    89   101:==============================   *
  42   105   123:===================================     *
  44    93   136:===============================              *
  46   157   138:=============================================*=======
  48   152   132:===========================================*=======
  50   101   121:==================================      *
  52   117   106:===================================*===
  54    86    91:============================= *
  56    74    76:=========================*
  58    41    62:==============      *
  60    76    50:================*=========
  62    29    40:==========   *
  64    24    32:========  *
  66    17    25:======  *
  68    35    20:======*=====
  70    35    16:=====*======
  72    13    12:===*=
  74     4    10:== *
  76     5     7:==*
  78     2     6:=*
  80    11     4:=*==
  82     1     3:*
  84     8     3:*==
  86     1     2:*
  88     0     2:*          inset = represents 1 library sequences
  90     0     1:*
  92     0     1:*         :*
  94     2     1:*         :*=
  96     0     1:*         :*
  98     0     0:          *
 100     0     0:          *
 102     0     0:          *
 104     0     0:          *
 106     0     0:          *
 108     0     0:          *
 110     0     0:          *
 112     0     0:          *
 114     0     0:          *
 116     0     0:          *
 118     0     0:          *
>120     0     0:          *
 331323 residues in  1471 sequences
  Expectation_n fit: rho(ln(x))= 4.8574+/-0.00316; mu= 2.6829+/- 0.164
 mean_var=29.2344+/- 8.102, 0's: 2 Z-trim: 2  B-trim: 32 in 1/42
 Lambda= 0.237207
 Kolmogorov-Smirnov  statistic: 0.0305 (N=28) at  36

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
 join: 42, opt: 30, open/ext: -10/-2, width:  32
 Scan time:  0.040
The best scores are:                                      opt bits E(1471)
gi|162927|gb|AAA30478.1| alpha-s1-casein [Bos taur (  76)   48 20.6     2.6
gi|75107131|sp|P82952.1|ALL2_PRUDU RecName: Full=S (  25)   42 18.8       3
gi|22090|emb|CAA43361.1| HMW glutenin subunit 1By9 ( 705)   54 22.2     8.2
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   47 20.1     8.5
gi|33149333|gb|AAP96759.1| group 1 allergen Dac g  ( 240)   48 20.4     9.8

>>gi|162927|gb|AAA30478.1| alpha-s1-casein [Bos taurus]   (76 aa)
 initn:  48 init1:  48 opt:  48  Z-score: 94.9  bits: 20.6 E():  2.6
Smith-Waterman score: 48;  32.000% identity (76.000% similar) in 25 aa overlap (5-29:19-43)

                             10        20        30
RF_1_+               KFIFKSFKHQSASSHQKLGPNSLKLESSQLRSTGQIRS
                         :... .: :: ... :::.. .. :
gi|162 LSKDIGSESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKQIQKEDVPSERYLGYLEQLL
               10        20        30        40        50        60

gi|162 RLKKYKVPQLEIVPNS
               70

>>gi|75107131|sp|P82952.1|ALL2_PRUDU RecName: Full=Seed   (25 aa)
 initn:  42 init1:  42 opt:  42  Z-score: 93.8  bits: 18.8 E():    3
Smith-Waterman score: 42;  31.818% identity (72.727% similar) in 22 aa overlap (12-33:3-24)

               10        20        30
RF_1_+ KFIFKSFKHQSASSHQKLGPNSLKLESSQLRSTGQIRS
                  ...:  . ::.: :. .. :.:
gi|751          TKSQTHVPIRPNKLVLKVQKDRATN
                        10        20

>>gi|22090|emb|CAA43361.1| HMW glutenin subunit 1By9 [Tr  (705 aa)
 initn:  47 init1:  47 opt:  54  Z-score: 86.0  bits: 22.2 E():  8.2
Smith-Waterman score: 54;  37.037% identity (74.074% similar) in 27 aa overlap (9-35:174-199)

                                     10        20        30
RF_1_+                       KFIFKSFKHQSASSHQKLGPNSLKLESSQLRSTGQIRS
                                     :::....:   :.::. . .: .. ::
gi|220 ASPQQPGQGQQPGKWQELGQGQQGYYPTSLHQSGQGQQGYYPSSLQ-QPGQGQQIGQGQQ
           150       160       170       180        190       200

gi|220 GYYPTSLQQPGQGQQIGQGQQGYYPTSPQHPGQRQQPGQGQQIGQGQQLGQGRQIGQGQQ
            210       220       230       240       250       260

>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  47 init1:  47 opt:  47  Z-score: 85.7  bits: 20.1 E():  8.5
Smith-Waterman score: 47;  32.000% identity (72.000% similar) in 25 aa overlap (5-29:31-55)

                                         10        20        30
```

```
RF_1_+                           KFIFKSFKHQSASSHQKLGPNSLKLESSQLRSTG
                                     :... .: :: ... :::.. .  :
gi|159 FSEVFGKEKVNELSKDIGSESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQKEDVP
               10        20        30        40        50        60


RF_1_+ QIRS

gi|159 SERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQQKEPMIGVNQELAYFYP
               70        80        90       100       110       120

>>gi|33149333|gb|AAP96759.1| group 1 allergen Dac g 1.01  (240 aa)
 initn:  45 init1:  45 opt:  48  Z-score: 84.6  bits: 20.4 E():  9.8
Smith-Waterman score: 48;  35.714% identity (67.857% similar) in 28 aa overlap (9-36:104-130)

                                     10        20        30
RF_1_+                       KFIFKSFKHQSASSHQKLGPNSLKLESSQLRSTGQIRS
                                     : . :.:  .:  . : : ..:::.:..
gi|331 EIKCTKPESCSGEAVTVHITDDNEEPIAPYHFDLSGHA-FGSMAKKGEEQKLRSAGELEL
            80        90       100       110        120       130

gi|331 QFRRVKCKYPEGTKLTFHVEKGSNPNYLALLVKYVDGDGDVVAVDIKEKGKDKWIALKES
            140       150       160       170       180       190



38 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:48:22 2010 done: Fri Apr 30 23:48:22 2010
 Total Scan time:  0.040 Total Display time:  0.000
```

RF_2+2

```
Function used was FASTA [version 3.4t26 July 7, 2006]
# fasta -Q -d 500 -E 10 fasta_input.txt /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta 1
FASTA searches a protein or DNA sequence data bank
 version 3.4t26 July 7, 2006
Please cite:
 W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

Query library fasta_input.txt vs /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library
searching /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

  1>>>RF_2_+2 29 aa - 29 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

       opt      E()
< 20     2     0:=
  22     0     0:          one = represents 3 library sequences
  24     0     0:
  26     0     0:
  28     1     0:=
  30     1     2:*
  32     9     8:==*
  34    14    21:===== *
  36    81    44:==============*============
  38    83    72:=======================*====
  40    70   101:========================         *
  42   106   123:====================================    *
  44   127   136:===========================================  *
  46   148   138:=============================================*====
  48   100   132:==================================         *
  50    97   121:=================================       *
  52   104   106:===================================*
  54    93    91:==============================*
  56   111    76:=========================*===========
  58    67    62:====================*==
  60    41    50:==============  *
  62    27    40:=========    *
  64    52    32:==========*=======
  66    32    25:========*==
  68    14    20:===== *
  70    34    16:=====*======
  72     4    12:== *
  74    10    10:===*
  76    14     7:==*==
  78     4     6:=*
  80     8     4:=*=
  82     3     3:*
  84     1     3:*
  86     4     2:*=
  88     2     2:*          inset = represents 1 library sequences
  90     3     1:*
  92     2     1:*         :*=
  94     1     1:*         :*
  96     1     1:*         :*
  98     0     0:          *
 100     0     0:          *
 102     0     0:          *
 104     0     0:          *
 106     0     0:          *
 108     0     0:          *
 110     0     0:          *
 112     0     0:          *
 114     0     0:          *
 116     0     0:          *
 118     0     0:          *
>120     0     0:          *
```

```
 331323 residues in  1471 sequences
  Expectation_n fit: rho(ln(x))= 3.1982+/- 0.003; mu= 7.1394+/- 0.157
 mean_var=25.4479+/- 7.427, 0’s: 2 Z-trim: 2  B-trim: 52 in 1/42
 Lambda= 0.254243
 Kolmogorov-Smirnov  statistic: 0.0428 (N=29) at  52

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
 join: 42, opt: 30, open/ext: -10/-2, width:  32
 Scan time:  0.040
The best scores are:                                      opt bits E(1471)
gi|8843917|gb|AAF80164.1| pollen major allergen 1- ( 367)   49 22.6     2.4
gi|22796153|emb|CAD42710.1| hydrophobin [Davidiell ( 105)   44 20.5     3.1
gi|8101715|gb|AAF72627.1|AF257493_1 Cup s 1 pollen ( 367)   47 21.9       4
gi|8101717|gb|AAF72628.1|AF257494_1 Cup s 1 pollen ( 367)   47 21.9       4
gi|8101713|gb|AAF72626.1|AF257492_1 Cup s 1 pollen ( 367)   47 21.9       4
gi|8101711|gb|AAF72625.1|AF257491_1 Cup s 1 pollen ( 367)   47 21.9       4
gi|8101719|gb|AAF72629.1|AF257495_1 Cup s 1 pollen ( 367)   47 21.9       4
gi|9087167|sp|Q9SCG9.1|MPAC1_CUPAR RecName: Full=M ( 346)   45 21.2     6.3
gi|118197955|gb|ABK78766.1| major allergen Cup a 1 ( 347)   45 21.2     6.3
gi|9087152|sp|P81294.1|MPAJ1_JUNAS RecName: Full=M ( 367)   45 21.2     6.6
gi|15139849|emb|CAC48400.1| putative allergen jun  ( 367)   45 21.2     6.6
gi|19069497|emb|CAC37790.2| putative allergen Cup  ( 367)   45 21.2     6.6
gi|8843921|gb|AAF80166.1| pollen major allergen 1- ( 367)   45 21.2     6.6
gi|9087163|sp|Q96385.1|MPAC1_CHAOB RecName: Full=M ( 375)   44 20.8     8.7

>>gi|8843917|gb|AAF80164.1| pollen major allergen 1-2 [J  (367 aa)
 initn:  49 init1:  49 opt:  49  Z-score: 95.5  bits: 22.6 E():  2.4
Smith-Waterman score: 49;  53.846% identity (92.308% similar) in 13 aa overlap (16-28:203-215)

                              10        20
RF_2_+                LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     :..: :.::..::
gi|884 AITMRNVTNAWIDHNSLSDCSDGLIDVTLGSTGITIFNNHFFNHHKVMLLGHDDTYDDDK
            180       190       200       210       220       230

gi|884 SMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDPWNIYAIGGSSNPTILSEGNSFTAP
            240       250       260       270       280       290

>>gi|22796153|emb|CAD42710.1| hydrophobin [Davidiella ta  (105 aa)
 initn:  44 init1:  44 opt:  44  Z-score: 93.6  bits: 20.5 E():  3.1
Smith-Waterman score: 44;  50.000% identity (71.429% similar) in 14 aa overlap (12-25:60-73)

                                  10        20
RF_2_+                    LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     :.:::. :: .  :
gi|227 KVEIDGQDSAPVCGNGQKVACCNSGEDLIGLNCLSIPILAIPIQKACGSNIAACCQTGDS
      30        40        50        60        70        80

gi|227 EGNLLNLEANCLAIPL
```

```
      90       100

>>gi|8101715|gb|AAF72627.1|AF257493_1 Cup s 1 pollen all  (367 aa)
 initn:  45 init1:  45 opt:  47  Z-score: 91.6  bits: 21.9 E():    4
Smith-Waterman score: 47;  50.000% identity (75.000% similar) in 20 aa overlap (9-28:198-215)

                                     10        20
RF_2_+                       LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     .: ::  :..: : ::..::
gi|810 AQDGDAITMRNVTNAWIDHNSLPDCSDGLIDVTLS--STGITISNNHFFNHHKVMLLGHD
       170       180       190       200         210       220

gi|810 DTYDDDKSMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDQWNIYAIGGSSNPTILSE
         230       240       250       260       270       280

>>gi|8101717|gb|AAF72628.1|AF257494_1 Cup s 1 pollen all  (367 aa)
 initn:  45 init1:  45 opt:  47  Z-score: 91.6  bits: 21.9 E():    4
Smith-Waterman score: 47;  50.000% identity (75.000% similar) in 20 aa overlap (9-28:198-215)

                                     10        20
RF_2_+                       LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     .: ::  :..: : ::..::
gi|810 AQDGDAITMRNVTNAWIDHNSLSDCSDGLIDVTLS--STGITISNNHFFNHHKVMLLGHD
       170       180       190       200         210       220

gi|810 DTYDDDKSMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDQWNIYAIGGSSNPTILSE
         230       240       250       260       270       280

>>gi|8101713|gb|AAF72626.1|AF257492_1 Cup s 1 pollen all  (367 aa)
 initn:  45 init1:  45 opt:  47  Z-score: 91.6  bits: 21.9 E():    4
Smith-Waterman score: 47;  50.000% identity (75.000% similar) in 20 aa overlap (9-28:198-215)

                                     10        20
RF_2_+                       LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     .: ::  :..: : ::..::
gi|810 AQDGDAITMRNVTNAWIDHNSLSDCSDGLIDVTLS--STGITISNNHFFNHHKVMLLGHD
       170       180       190       200         210       220

gi|810 DTYDDDKSMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDQWNIYAIGGSSNPTILSE
         230       240       250       260       270       280

>>gi|8101711|gb|AAF72625.1|AF257491_1 Cup s 1 pollen all  (367 aa)
 initn:  45 init1:  45 opt:  47  Z-score: 91.6  bits: 21.9 E():    4
Smith-Waterman score: 47;  50.000% identity (75.000% similar) in 20 aa overlap (9-28:198-215)

                                     10        20
RF_2_+                       LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     .: ::  :..: : ::..::
gi|810 AQDGDAITMRNVTNAWIDHNSLSDCSDGLIDVTLS--STGITISNNHFFNHHKVMLLGHD
```

```
       170       180       190       200         210       220

gi|810 DTYDDDKSMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDQWNIYAIGGSSNPTILSE
         230       240       250       260       270       280

>>gi|8101719|gb|AAF72629.1|AF257495_1 Cup s 1 pollen all  (367 aa)
 initn:  45 init1:  45 opt:  47  Z-score: 91.6  bits: 21.9 E():    4
Smith-Waterman score: 47;  50.000% identity (75.000% similar) in 20 aa overlap (9-28:198-215)

                                     10        20
RF_2_+                       LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     .: ::  :..: : ::..::
gi|810 AQDGDAITMRNVTNAWIDHNSLSDCSDGLIDVTLS--STGITISNNHFFNHHKVMLLGHD
       170       180       190       200         210       220

gi|810 DTYDDDKSMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDQWNIYAIGGSSNPTILSE
         230       240       250       260       270       280

>>gi|9087167|sp|Q9SCG9.1|MPAC1_CUPAR RecName: Full=Major  (346 aa)
 initn:  45 init1:  45 opt:  45  Z-score: 88.0  bits: 21.2 E():  6.3
Smith-Waterman score: 45;  53.846% identity (84.615% similar) in 13 aa overlap (16-28:182-194)

                              10        20
RF_2_+               LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     :..: : ::..::
gi|908 AITMRNVTNAWIDHNSLSDCSDGLIDVTLGSTGITISNNHFFNHHKVMLLGHDDTYDDDK
             160       170       180       190       200       210

gi|908 SMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDQWNIYAIGGSSNPTILSEGNSFTAP
             220       230       240       250       260       270

>>gi|118197955|gb|ABK78766.1| major allergen Cup a 1 [Cu  (347 aa)
 initn:  45 init1:  45 opt:  45  Z-score: 88.0  bits: 21.2 E():  6.3
Smith-Waterman score: 45;  53.846% identity (84.615% similar) in 13 aa overlap (16-28:183-195)

                              10        20
RF_2_+               LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     :..: : ::..::
gi|118 AITMRNVTNAWIDHNSLSDCSDGLIDVTLGSTGITISNNHFFNHHKVMLLGHDDTYDDDK
            160       170       180       190       200       210

gi|118 SMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDQWNIYAIGGSSNPTILSEGNSFTAP
            220       230       240       250       260       270

>>gi|9087152|sp|P81294.1|MPAJ1_JUNAS RecName: Full=Major  (367 aa)
 initn:  45 init1:  45 opt:  45  Z-score: 87.6  bits: 21.2 E():  6.6
Smith-Waterman score: 45;  53.846% identity (84.615% similar) in 13 aa overlap (16-28:203-215)

                              10        20
```

```
RF_2_+                LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     :..: : ::..::
gi|908 AITMRNVTNAWIDHNSLSDCSDGLIDVTLGSTGITISNNHFFNHHKVMLLGHDDTYDDDK
            180       190       200       210       220       230

gi|908 SMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDPWNIYAIGGSSNPTILSEGNSFTAP
            240       250       260       270       280       290

>>gi|15139849|emb|CAC48400.1| putative allergen jun o 1   (367 aa)
 initn:  45 init1:  45 opt:  45  Z-score: 87.6  bits: 21.2 E():  6.6
Smith-Waterman score: 45;  53.846% identity (84.615% similar) in 13 aa overlap (16-28:203-215)

                              10        20
RF_2_+                LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     :..: : ::..::
gi|151 AITMRNVTNAWIDHNSLSDCSDGLIDVTLGSTGITISNNHFFNHHKVMLLGHDDTYDNDK
            180       190       200       210       220       230

gi|151 SMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDPWNIYAIGGSSNPTILSEGNSFTAP
            240       250       260       270       280       290

>>gi|19069497|emb|CAC37790.2| putative allergen Cup a 1   (367 aa)
 initn:  45 init1:  45 opt:  45  Z-score: 87.6  bits: 21.2 E():  6.6
Smith-Waterman score: 45;  53.846% identity (84.615% similar) in 13 aa overlap (16-28:203-215)

                              10        20
RF_2_+                LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     :..: : ::..::
gi|190 AITMRNVTNAWIDHNSLSDCSDGLIDVTLGSTGITISNNHFFNHHKVMLLGHDDTYDDDI
            180       190       200       210       220       230

gi|190 SMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDQWNIYAIGGSSNPTILSEGNSFTAP
            240       250       260       270       280       290

>>gi|8843921|gb|AAF80166.1| pollen major allergen 1-1 [J  (367 aa)
 initn:  45 init1:  45 opt:  45  Z-score: 87.6  bits: 21.2 E():  6.6
Smith-Waterman score: 45;  53.846% identity (84.615% similar) in 13 aa overlap (16-28:203-215)

                              10        20
RF_2_+                LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     :..: : ::..::
gi|884 AITMRNVTNAWIDHNSLSDCSDGLIDVTLGSTGITISNNHFFNHHKVMLLGHDDTYDDDK
            180       190       200       210       220       230

gi|884 SMKVTVAFNQFGPNAGQRMPRARYGLVHVANNNYDPWNIYAIGGSSNPTILSEGNSFTAP
            240       250       260       270       280       290

>>gi|9087163|sp|Q96385.1|MPAC1_CHAOB RecName: Full=Major  (375 aa)
 initn:  44 init1:  44 opt:  44  Z-score: 85.5  bits: 20.8 E():  8.7
```

```
Smith-Waterman score: 44;  46.154% identity (84.615% similar) in 13 aa overlap (16-28:203-215)

                              10        20
RF_2_+                LQYIKNVRNVLLSCLSVNILILNNQYFNS
                                     :... : ::..::
gi|908 AITMRNVTDVWIDHNSLSDSSDGLVDVTLASTGVTISNNHFFNHHKVMLLGHSDIYSDDK
            180       190       200       210       220       230

gi|908 SMKVTVAFNQFGPNAGQRMPRARYGLIHVANNNYDPWSIYAIGGSSNPTILSEGNSFTAP
            240       250       260       270       280       290



29 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:48:48 2010 done: Fri Apr 30 23:48:49 2010
 Total Scan time:  0.040 Total Display time:  0.000

Function used was FASTA [version 3.4t26 July 7, 2006]
```

BLASTp Search Outputs Using Putative Reading Frames as Queries in Soybean Event DAS-68416-4 against GenBank Non Redundant Protein Sequences “nr”

```
RF_1_+1
BLASTP 2.2.21 [Jun-14-2009]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, 
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), 
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs",  Nucleic Acids Res. 25:3389-3402.

Reference for compositional score matrix adjustment: Altschul, Stephen F., 
John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis,
Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches
using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

Query= RF_1_+1
         (16 letters)

Database: /usr/local/blast/db/blastlibs/nr 
           10,862,569 sequences; 3,701,345,023 total letters

Searching..................................................done

 ***** No hits found ******


  Database: /usr/local/blast/db/blastlibs/nr
    Posted date:  Apr 19, 2010 11:58 AM
  Number of letters in database: 3,701,345,023
  Number of sequences in database:  10,862,569
  
Lambda     K      H
   0.319    0.139    0.388 

Gapped
Lambda     K      H
   0.267   0.0410    0.140 


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 10862569
Number of Hits to DB: 105,906,678
Number of extensions: 838480
Number of successful extensions: 476
Number of sequences better than  1.0: 0
Number of HSP's gapped: 476
Number of HSP's successfully gapped: 0
```

```
Length of query: 16
Length of database: 3,701,345,023
Length adjustment: 0
Effective length of query: 16
Effective length of database: 3,701,345,023
Effective search space: 59221520368
Effective search space used: 59221520368
Neighboring words threshold: 11
Window for multiple hits: 40
X1: 16 ( 7.4 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 41 (21.7 bits)
S2: 81 (35.8 bits)
BLASTP 2.2.21 [Jun-14-2009]
```

RF_1_+2

```
BLASTP 2.2.21 [Jun-14-2009]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs",  Nucleic Acids Res. 25:3389-3402.

Reference for compositional score matrix adjustment: Altschul, Stephen F.,
John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis,
Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches
using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

Query= RF_1_+2
         (38 letters)

Database: /usr/local/blast/db/blastlibs/nr
           10,862,569 sequences; 3,701,345,023 total letters

Searching..................................................done

 ***** No hits found ******


  Database: /usr/local/blast/db/blastlibs/nr
    Posted date:  Apr 19, 2010 11:58 AM
  Number of letters in database: 3,701,345,023
  Number of sequences in database:  10,862,569

Lambda     K      H
   0.312    0.123    0.322
```

```
Gapped
Lambda     K      H
   0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 10862569
Number of Hits to DB: 198,329,733
Number of extensions: 2870352
Number of successful extensions: 4657
Number of sequences better than  1.0: 0
Number of HSP's gapped: 4657
Number of HSP's successfully gapped: 0
Length of query: 38
Length of database: 3,701,345,023
Length adjustment: 12
Effective length of query: 26
Effective length of database: 3,570,994,195
Effective search space: 92845849070
Effective search space used: 92845849070
Neighboring words threshold: 11
Window for multiple hits: 40
X1: 16 ( 7.2 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 42 (21.9 bits)
S2: 83 (36.6 bits)
```

RF_1_-1

```
BLASTP 2.2.21 [Jun-14-2009]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs",  Nucleic Acids Res. 25:3389-3402.

Reference for compositional score matrix adjustment: Altschul, Stephen F.,
John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis,
Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches
using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

Query= RF_1_-1
         (38 letters)

Database: /usr/local/blast/db/blastlibs/nr
           10,862,569 sequences; 3,701,345,023 total letters
```

```
Searching..................................................done

 ***** No hits found ******


  Database: /usr/local/blast/db/blastlibs/nr
    Posted date:  Apr 19, 2010 11:58 AM
  Number of letters in database: 3,701,345,023
  Number of sequences in database:  10,862,569

Lambda     K      H
   0.328    0.148    0.576

Gapped
Lambda     K      H
   0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 10862569
Number of Hits to DB: 317,428,271
Number of extensions: 7119911
Number of successful extensions: 12293
Number of sequences better than  1.0: 0
Number of HSP’s gapped: 12296
Number of HSP’s successfully gapped: 0
Length of query: 38
Length of database: 3,701,345,023
Length adjustment: 12
Effective length of query: 26
Effective length of database: 3,570,994,195
Effective search space: 92845849070
Effective search space used: 92845849070
Neighboring words threshold: 11
Window for multiple hits: 40
X1: 15 ( 7.1 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 40 (21.7 bits)
S2: 83 (36.6 bits)
```

RF_1_-2

```
BLASTP 2.2.21 [Jun-14-2009]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
″Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs″,  Nucleic Acids Res. 25:3389-3402.
```

```
Reference for compositional score matrix adjustment: Altschul, Stephen F.,
John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis,
Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches
using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

Query= RF_1_-2
         (24 letters)

Database: /usr/local/blast/db/blastlibs/nr
           10,862,569 sequences; 3,701,345,023 total letters

Searching..................................................done

 ***** No hits found ******


  Database: /usr/local/blast/db/blastlibs/nr
    Posted date:  Apr 19, 2010 11:58 AM
  Number of letters in database: 3,701,345,023
  Number of sequences in database:  10,862,569

Lambda     K      H
   0.324    0.134    0.363

Gapped
Lambda     K      H
   0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 10862569
Number of Hits to DB: 130,268,777
Number of extensions: 1380559
Number of successful extensions: 3369
Number of sequences better than  1.0: 0
Number of HSP's gapped: 3369
Number of HSP's successfully gapped: 0
Length of query: 24
Length of database: 3,701,345,023
Length adjustment: 0
Effective length of query: 24
Effective length of database: 3,701,345,023
Effective search space: 88832280552
Effective search space used: 88832280552
Neighboring words threshold: 11
Window for multiple hits: 40
X1: 15 ( 7.0 bits)
```

```
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 40 (21.6 bits)
S2: 83 (36.6 bits)
```

RF_1_+3

```
BLASTP 2.2.21 [Jun-14-2009]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs",  Nucleic Acids Res. 25:3389-3402.

Reference for compositional score matrix adjustment: Altschul, Stephen F.,
John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis,
Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches
using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

Query= RF_1_+3
         (19 letters)

Database: /usr/local/blast/db/blastlibs/nr
           10,862,569 sequences; 3,701,345,023 total letters

Searching..................................................done

 ***** No hits found ******


  Database: /usr/local/blast/db/blastlibs/nr
    Posted date:  Apr 19, 2010 11:58 AM
  Number of letters in database: 3,701,345,023
  Number of sequences in database:  10,862,569

Lambda     K      H
   0.313    0.117    0.343

Gapped
Lambda     K      H
   0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 10862569
Number of Hits to DB: 105,320,210
Number of extensions: 767916
Number of successful extensions: 1958
Number of sequences better than  1.0: 0
Number of HSP's gapped: 1958
Number of HSP's successfully gapped: 0
Length of query: 19
Length of database: 3,701,345,023
Length adjustment: 0
Effective length of query: 19
Effective length of database: 3,701,345,023
Effective search space: 70325555437
Effective search space used: 70325555437
Neighboring words threshold: 11
Window for multiple hits: 40
X1: 16 ( 7.2 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 41 (21.6 bits)
S2: 82 (36.2 bits)
```

RF_2_+1

```
BLASTP 2.2.21 [Jun-14-2009]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs",  Nucleic Acids Res. 25:3389-3402.

Reference for compositional score matrix adjustment: Altschul, Stephen F.,
John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis,
Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches
using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

Query= RF_2_+1
         (17 letters)

Database: /usr/local/blast/db/blastlibs/nr
           10,862,569 sequences; 3,701,345,023 total letters

Searching..................................................done

 ***** No hits found ******


  Database: /usr/local/blast/db/blastlibs/nr
    Posted date:  Apr 19, 2010 11:58 AM
  Number of letters in database: 3,701,345,023
  Number of sequences in database:  10,862,569

Lambda     K      H
   0.325    0.134    0.406
```

```
Gapped
Lambda     K      H
   0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 10862569
Number of Hits to DB: 97,056,364
Number of extensions: 575487
Number of successful extensions: 344
Number of sequences better than  1.0: 0
Number of HSP's gapped: 344
Number of HSP's successfully gapped: 0
Length of query: 17
Length of database: 3,701,345,023
Length adjustment: 0
Effective length of query: 17
Effective length of database: 3,701,345,023
Effective search space: 62922865391
Effective search space used: 62922865391
Neighboring words threshold: 11
Window for multiple hits: 40
X1: 15 ( 7.0 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 40 (21.7 bits)
S2: 82 (36.2 bits)
```

RF_2_-1

```
BLASTP 2.2.21 [Jun-14-2009]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs",  Nucleic Acids Res. 25:3389-3402.

Reference for compositional score matrix adjustment: Altschul, Stephen F.,
John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis,
Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches
using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

Query= RF_2_-1
         (22 letters)

Database: /usr/local/blast/db/blastlibs/nr
           10,862,569 sequences; 3,701,345,023 total letters

Searching..................................................done
```

```
 ***** No hits found ******


  Database: /usr/local/blast/db/blastlibs/nr
    Posted date:  Apr 19, 2010 11:58 AM
  Number of letters in database: 3,701,345,023
  Number of sequences in database:  10,862,569

Lambda     K      H
   0.341    0.151    0.406

Gapped
Lambda     K      H
   0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 10862569
Number of Hits to DB: 105,235,511
Number of extensions: 887243
Number of successful extensions: 1706
Number of sequences better than  1.0: 0
Number of HSP's gapped: 1706
Number of HSP's successfully gapped: 0
Length of query: 22
Length of database: 3,701,345,023
Length adjustment: 0
Effective length of query: 22
Effective length of database: 3,701,345,023
Effective search space: 81429590506
Effective search space used: 81429590506
Neighboring words threshold: 11
Window for multiple hits: 40
X1: 15 ( 7.4 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 39 (21.9 bits)
S2: 83 (36.6 bits)
```

RF_2_+2

```
BLASTP 2.2.21 [Jun-14-2009]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs",  Nucleic Acids Res. 25:3389-3402.
```

```
Reference for compositional score matrix adjustment: Altschul, Stephen F.,
John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis,
Alejandro A. Schaffer, and Yi-Kuo Yu (2005) ″Protein database searches
using compositionally adjusted substitution matrices″, FEBS J. 272:5101-5109.

Query= RF_2_+2
         (29 letters)

Database: /usr/local/blast/db/blastlibs/nr
           10,862,569 sequences; 3,701,345,023 total letters

Searching..................................................done

 ***** No hits found ******


  Database: /usr/local/blast/db/blastlibs/nr
    Posted date:  Apr 19, 2010 11:58 AM
  Number of letters in database: 3,701,345,023
  Number of sequences in database:  10,862,569

Lambda     K      H
   0.327    0.140    0.391

Gapped
Lambda     K      H
   0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 10862569
Number of Hits to DB: 173,293,224
Number of extensions: 2417020
Number of successful extensions: 7164
Number of sequences better than  1.0: 0
Number of HSP’s gapped: 7165
Number of HSP’s successfully gapped: 0
Length of query: 29
Length of database: 3,701,345,023
Length adjustment: 4
Effective length of query: 25
Effective length of database: 3,657,894,747
Effective search space: 91447368675
Effective search space used: 91447368675
Neighboring words threshold: 11
Window for multiple hits: 40
X1: 15 ( 7.1 bits)
X2: 38 (14.6 bits)
```

X3: 64 (24.7 bits)
S1: 40 (21.7 bits)
S2: 83 (36.6 bits)