FASTA Outputs of Putative Reading Frames Spanning the Junctions between the Insert and its Flanking Borders in Maize Event DAS-40278-9 against 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: 1 - 80 80 aa - 80 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      2     0:=
  28      1     0:=
  30     22     2:*=======
  32      8     8:==*
  34     45    21:======*========
  36     47    44:============*=
  38     64    72:==================== *
  40     96   101:============================= *
  42    109   123:==================================   *
  44    132   136:========================================== *
  46    168   138:================================================*=========
  48    109   132:==================================      *
  50     97   121:=============================        *
  52     82   106:=========================            *
  54    104    91:=======================*====
  56     77    76:=======================*
  58     82    62:==================*=======
  60     52    50:==============*=
  62     31    40:==========   *
  64     34    32:=========*=
  66     16    25:======   *
  68     13    20:==== *
  70     20    16:=====*=
  72     10    12:===*
  74      9    10:===*
  76      4     7:==*
  78      3     6:=*
  80      4     4:=*
  82      3     3:*
  84      6     3:*=
```

```
    86      1      2:*
    88      3      2:*                   inset = represents 1 library sequences
    90      3      1:*
    92      2      1:*          :*=
    94      3      1:*          :*==
    96      1      1:*          :*
    98      2      0:=          *==
   100      0      0:           *
   102      0      0:           *
   104      0      0:           *
   106      1      0:=          *=
   108      1      0:=          *=
   110      2      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))= 4.4474+/-0.00352; mu= 9.9402+/- 0.180

mean_var=40.4972+/-10.593, 0's: 2 Z-trim: 6 B-trim: 125 in 1/42

Lambda= 0.201540

Kolmogorov-Smirnov statistic: 0.0340 (N=29) 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: 0.080

The best scores are: opt bits E(1471)

```
gi|218059718|emb|CAT99612.1| thaumatin-like protei ( 158)   71 25.7    0.35
gi|218059715|emb|CAT99611.1| thaumatin-like protei ( 158)   71 25.7    0.35
gi|60418848|gb|AAX19851.1| thaumatin-like protein  ( 246)   71 25.7    0.52
gi|60418842|gb|AAX19848.1| thaumatin-like protein  ( 246)   71 25.7    0.52
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 24.0    1.5
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 25.0    1.7
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.7    2.2
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.8    3.1
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.8    3.1
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.8    3.1
gi|169971|gb|AAA33965.1| glycinin precursor [Glyci ( 240)   62 23.1    3.1
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.8    3.7
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.8    4.1
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 23.2    4.8
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.5    5.2
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.9    5.2
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.9    6.1
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.9    6.3
gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea m ( 252)   58 22.0    7.3
gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60 ( 392)   59 22.3    8.9
gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=P ( 139)   54 20.7    9.6
```

```
gi|741844|prf||2008179A major allergen Par j I      ( 143)   54 20.7     9.8

>>gi|218059718|emb|CAT99612.1| thaumatin-like protein [M  (158 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 110.6  bits: 25.7 E(): 0.35
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (6-51:99-141)

                                          10        20        30
RF_1_+                        ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                  .: : ::. . ...     .::       .::
gi|218 DVSLVDGFNLPMYVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
         70        80        90       100       110       120

            40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPART
         :   ::      :.::.  :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
             130       140       150


>>gi|218059715|emb|CAT99611.1| thaumatin-like protein [M  (158 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 110.6  bits: 25.7 E(): 0.35
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (6-51:99-141)

                                          10        20        30
RF_1_+                        ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                  .: : ::. . ...     .::       .::
gi|218 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
         70        80        90       100       110       120

            40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPART
         :   ::      :.::.  :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
             130       140       150


>>gi|60418848|gb|AAX19851.1| thaumatin-like protein prec  (246 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 107.5  bits: 25.7 E(): 0.52
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (6-51:159-201)

                                          10        20        30
RF_1_+                        ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                  .: : ::. . ...     .::       .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
        130       140       150       160       170       180

            40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPART
         :   ::      :.::.  :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
            190       200       210       220       230       240
```

```
>>gi|60418842|gb|AAX19848.1| thaumatin-like protein prec  (246 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 107.5  bits: 25.7 E(): 0.52
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (6-51:159-201)

                                 10        20        30
RF_1_+                 ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                            .: : ::. . ...     .::      .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
         130       140       150       160       170       180

              40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPART
         :   ::     :.::.  :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
           190       200       210       220       230       240


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score: 99.0  bits: 24.0 E():  1.5
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (17-73:114-176)

                             10        20        30        40
RF_1_+             ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                            .:. :. . ... .: :.    :::   : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
           90       100       110       120       130       140

                 50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPART
           . : : .::..: : :     : :::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
           150       160       170       180       190       200

gi|162 ENSEKTTISLW
           210


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly   (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score: 98.2  bits: 25.0 E():  1.7
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (22-80:301-356)

                         10        20        30        40        50
RF_1_+             ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                            :: .:   . .   . : .:..     :::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
             280       290       300       310       320

                  60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPART
         :.: :  ::: .  . ..:.  : .:
```

```
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          330       340       350       360       370       380


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score:  96.3  bits: 22.7 E():  2.2
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (17-73:29-91)

                          10        20        30        40
RF_1_+              ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                            .: . : . ... .: :.   ::: : :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
               10        20        30        40        50        60


              50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPART
          .   : :  .::..: :    :    : ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
               70        80        90        100       110       120


gi|159 SEKTTMPLW



>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score:  93.6  bits: 23.8 E():  3.1
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (22-80:216-271)

                        10        20        30        40        50
RF_1_+              ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                   :: .:  . .  . : .:.   :...:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


              60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPART
        :.:  :   ::: .  . ..:.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          250       260       270       280       290       300


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score:  93.6  bits: 23.8 E():  3.1
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (22-80:216-271)

                        10        20        30        40        50
RF_1_+              ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                   :: .:  . .  . : .:.   :...:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


              60        70        80
```

```
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPART
         :.:  :   :::  .   . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 93.6  bits: 23.8 E():  3.1
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (22-80:216-271)

                       10        20        30        40        50
RF_1_+          ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                   :: .:   .  .  . : .:.    :::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


                60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPART
         :.:  :   :::  .   . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|169971|gb|AAA33965.1| glycinin precursor [Glycine m  (240 aa)
 initn:  47 init1:  47 opt:  62  Z-score: 93.5  bits: 23.1 E():  3.1
Smith-Waterman score: 62;  26.087% identity (60.870% similar) in 46 aa overlap (3-46:127-172)

                       10        20        30
RF_1_+          ETYDPHVRMECPTVVYI-RSRGYP-IISIDHL
                            :.::   ..   .:.:.  :..:    ... .
gi|169 RISTLNSLTLPALRQFGLSAQYLVLYRNGIYSPHWNLNANSVIYVTRGKGRVRVVNCQGN
          100       110       120       130       140       150


                40        50        60        70        80
RF_1_+ PISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPART
         :. .. :  :.:   :
gi|169 PVFDGDLTRGQLLLVPQNFVVADQGGKQGLEYVVFKTQHNAVSSYIKDLFRAIPSEVLSN
          160       170       180       190       200       210


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 92.2  bits: 23.8 E():  3.7
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (22-80:292-347)

                       10        20        30        40        50
RF_1_+          ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                   :: .:   .  .   . : .:.     :::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            270       280       290       300       310


                60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPART
```

```
           :.:   :   :::  .  . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         320       330       340       350       360       370


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 91.3  bits: 22.8 E():  4.1
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (32-55:13-35)

               10        20        30        40        50        60
RF_1_+ TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASL
                                   .. .:: ::   :.:i.. :....
gi|898                  MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGK
                                10        20        30        40


               70        80
RF_1_+ SGPNLQKTAYPSLVRPART


gi|898 DWLDAKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIK
              50        60        70        80        90       100


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 90.1  bits: 23.2 E():  4.8
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (26-77:13-63)

               10        20        30        40        50        60
RF_1_+ ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYAS
                              :.  :: . :  : : .  . .: .::.  :    ::
gi|212                 MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTAS
                               10        20        30        40


               70        80
RF_1_+ LSGPNLQKTAYPSLVRPART
         ..    . :.  .: ::
gi|212 MGYKAGMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVW
          50        60        70        80        90       100


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 89.6  bits: 22.5 E():  5.2
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (32-55:19-41)

               10        20        30        40        50        60
RF_1_+ TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASL
                                   .. :::  :   :.:i.. :....
gi|115                  MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGS
                               10        20        30        40


               70        80
RF_1_+ SGPNLQKTAYPSLVRPART
```

```
gi|115 DWLDAKATWYGKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIK
           50        60        70        80        90        100


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 89.5  bits: 21.9 E():  5.2
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (17-73:72-134)

                               10        20        30        40
RF_1_+                 ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                               .:. :. ... .: :.   :::   :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
           50        60        70        80        90        100


               50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPART
       :.   : :  .::..: :   :      :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
           110       120       130       140       150       160


gi|159 ENSEKTTMPLW
             170


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 88.3  bits: 21.9 E():  6.1
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (17-73:105-167)

                               10        20        30        40
RF_1_+                 ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                               .:. :. ... .: :.   :::   :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
           80        90        100       110       120       130


                 50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPART
       :.   : :  .::..: :   :      :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
           140       150       160       170       180       190


gi|159 ENSEKTTMPLW
           200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 88.0  bits: 21.9 E():  6.3
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (17-73:114-176)

                               10        20        30        40
RF_1_+                 ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                               .:. :. ... .: :.   :::   :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
```

```
                90          100         110         120         130         140


                   50          60          70          80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPART
         :  .    : :  .::..: :   :       :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
              150         160         170         180         190         200


gi|162 ENSEKTTMPLW
           210


>>gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea mays]  (252 aa)
 initn:  52 init1:  52 opt:  58  Z-score: 86.9  bits: 22.0 E():  7.3
Smith-Waterman score: 58;  40.000% identity (75.000% similar) in 20 aa overlap (36-55:6-24)


            10          20          30          40          50          60
RF_1_+ HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPN
                              :: ::   :.::.. :.....
gi|898                        ARALVFLVSGAW-CGPPKVPPGKNITATYGKDWLD
                                     10          20          30


          70          80
RF_1_+ LQKTAYPSLVRPART

gi|898 AKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKP
              40          50          60          70          80          90


>>gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60S ri  (392 aa)
 initn:  45 init1:  45 opt:  59  Z-score: 85.4  bits: 22.3 E():  8.9
Smith-Waterman score: 59;  30.769% identity (53.846% similar) in 52 aa overlap (26-77:13-63)


            10          20          30          40          50          60
RF_1_+ ETYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYAS
                        :.  ::  .  :   : : .   . .:. ::.. :    ::
gi|833                  MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKDDPKKPVHLTAS
                              10          20          30          40


          70          80
RF_1_+ LSGPNLQKTAYPSLVRPART
          ..    . :.  .: ::
gi|833 MGYKAGMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVW
              50          60          70          80          90          100


>>gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=Proba  (139 aa)
 initn:  52 init1:  52 opt:  54  Z-score: 84.8  bits: 20.7 E():  9.6
Smith-Waterman score: 54;  39.286% identity (60.714% similar) in 28 aa overlap (7-33:87-113)


                             10          20          30
RF_1_+                 ETYDPHVRMECPTV-VYIRSRGYPIISIDHLPISSA
```

```
                                          :  :.:  ::  :     :.      :.  :.  :  :..
gi|391 MKTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPV-SLRHGPVTGP
            60        70        80        90       100       110

            40        50        60        70        80
RF_1_+ FLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPART


gi|391 SDPAHKARLERPQIRVPPPAPEKA
            120       130


>>gi|741844|prf||2008179A major allergen Par j I        (143 aa)
 initn:  52 init1:  52 opt:  54  Z-score: 84.6  bits: 20.7 E():  9.8
Smith-Waterman score: 54;  39.286% identity (60.714% similar) in 28 aa overlap (7-33:81-107)


                                    10        20        30
RF_1_+                   ETYDPHVRMECPTV-VYIRSRGYPIISIDHLPISSA
                                          :  :.:  ::  :     :.      :.  :.  :  :..
gi|741 MKTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPV-SLRHGPVTGP
              60        70        80        90       100

            40        50        60        70        80
RF_1_+ FLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPART


gi|741 SRSRPPTKHGWRDPRLEFRPPHRKKPNPAFSTLG
          110       120       130       140




80 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:36 2010 done: Fri Apr 30 23:42:37 2010
 Total Scan time:  0.080 Total Display time:  0.010

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_+1: 2 - 81 80 aa - 80 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    2    0:=
 28    4    0:==
 30   27    2:*=======
 32   13    8:==*==
 34   43   21:======*========
 36   46   44:=============*=
 38   58   72:=================    *
 40  103  101:================================*=
 42   95  123:============================         *
 44  138  136:============================================*
 46  169  138:============================================*==========
 48  112  132:===================================      *
 50   92  121:============================         *
 52   80  106:========================          *
 54  104   91:===========================*====
 56   74   76:======================*
 58   90   62:==================*========
 60   50   50:==============*
 62   31   40:==========   *
 64   36   32:========*=
 66   14   25:====    *
 68   12   20:====   *
 70   20   16:=====*=
 72    8   12:===*
 74    9   10:===*
 76    4    7:==*
 78    3    6:=*
 80    4    4:=*
 82    3    3:*
 84    6    3:*=
 86    1    2:*
 88    3    2:*               inset = represents 1 library sequences
 90    3    1:*
 92    2    1:*          :*=
 94    3    1:*          :*==
 96    1    1:*          :*
 98    2    0:=          *==
100    0    0:           *
102    0    0:           *
104    0    0:           *
106    1    0:=          *=
108    1    0:=          *=
110    2    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))= 4.5031+/-0.00353; mu= 9.7596+/- 0.180
 mean_var=40.0255+/-10.448, 0's: 2 Z-trim: 6  B-trim: 125 in 1/42
 Lambda= 0.202724
 Kolmogorov-Smirnov  statistic: 0.0408 (N=29) 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:  0.090
The best scores are:                                   opt bits E(1471)
gi|218059718|emb|CAT99612.1| thaumatin-like protei ( 158)   71 25.7    0.34
gi|218059715|emb|CAT99611.1| thaumatin-like protei ( 158)   71 25.7    0.34
gi|60418848|gb|AAX19851.1| thaumatin-like protein  ( 246)   71 25.8    0.51
gi|60418842|gb|AAX19848.1| thaumatin-like protein  ( 246)   71 25.8    0.51
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 24.0    1.5
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 25.0    1.7
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.8    2.1
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.8    3.1
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.8    3.1
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.8    3.1
gi|169971|gb|AAA33965.1| glycinin precursor [Glyci ( 240)   62 23.1    3.1
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.8    3.7
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.8    4.1
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 23.2    4.8
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.9    5.1
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.6    5.1
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.9     6
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.9    6.3
gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea m ( 252)   58 22.0    7.3
gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60 ( 392)   59 22.3    8.9
gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=P ( 139)   54 20.7    9.5
gi|741844|prf||2008179A major allergen Par j I     ( 143)   54 20.7    9.8


>>gi|218059718|emb|CAT99612.1| thaumatin-like protein [M  (158 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 110.8  bits: 25.7 E(): 0.34
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (5-50:99-141)

                                   10        20        30
RF_1_+                     TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                 .: : ::. . ...     .::      .::
gi|218 DVSLVDGFNLPMYVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
           70        80        90       100       110       120

              40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTI
         :  ::    :.::.  :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
           130       140       150
```

```
>>gi|218059715|emb|CAT99611.1| thaumatin-like protein [M  (158 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 110.8  bits: 25.7 E(): 0.34
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (5-50:99-141)

                                    10        20        30
RF_1_+                      TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                  .: : ::. . ...    .::      .::
gi|218 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
          70        80        90       100       110       120

              40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTI
          :   ::      :.::.  :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
            130       140       150


>>gi|60418848|gb|AAX19851.1| thaumatin-like protein prec  (246 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 107.6  bits: 25.8 E(): 0.51
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (5-50:159-201)

                                    10        20        30
RF_1_+                      TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                  .: : ::. . ...    .::      .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
          130       140       150       160       170       180

              40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTI
          :   ::      :.::.  :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
            190       200       210       220       230       240

>>gi|60418842|gb|AAX19848.1| thaumatin-like protein prec  (246 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 107.6  bits: 25.8 E(): 0.51
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (5-50:159-201)

                                    10        20        30
RF_1_+                      TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                  .: : ::. . ...    .::      .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
          130       140       150       160       170       180

              40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTI
          :   ::      :.::.  :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
            190       200       210       220       230       240
```

```
>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score:  99.1  bits: 24.0 E():  1.5
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (16-72:114-176)

                              10        20        30        40
RF_1_+               TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                               .:. :. ...  .: :.    :::   : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
            90       100       110       120       130       140

                      50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTI
         . : :  .::..: :  :    :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            150       160       170       180       190       200

gi|162 ENSEKTTISLW
            210


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly   (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score:  98.2  bits: 25.0 E():  1.7
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (21-79:301-356)

                              10        20        30        40        50
RF_1_+               TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                               :: .:   . .  . : .:..    ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
                 280       290       300       310       320

                      60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTI
        :.:  :   :::  .  . .::.  :  .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          330       340       350       360       370       380


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru   (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score:  96.4  bits: 22.8 E():  2.1
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (16-72:29-91)

                              10        20        30        40
RF_1_+               TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                               .:. :. ...  .: :.    :::   : :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
                  10        20        30        40        50        60

                      50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTI
        . : :  .::..: :  :    :     : :  ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
```

```
                    70        80        90       100       110       120


gi|159 SEKTTMPLW


>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 93.6  bits: 23.8 E():  3.1
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (21-79:216-271)

                         10        20        30        40        50
RF_1_+            TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                  :: .:    . .   . : .:..    :::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240


                 60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTI
         :.:   :   :::  .   . .:.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
           250       260       270       280       290       300


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 93.6  bits: 23.8 E():  3.1
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (21-79:216-271)

                         10        20        30        40        50
RF_1_+            TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                  :: .:    . .   . : .:..    :::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240


                 60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTI
         :.:   :   :::  .   . .:.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
           250       260       270       280       290       300


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 93.6  bits: 23.8 E():  3.1
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (21-79:216-271)

                         10        20        30        40        50
RF_1_+            TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                  :: .:    . .   . : .:..    :::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240


                 60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTI
```

```
            :.:   :     :::. .  . .::.   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|169971|gb|AAA33965.1| glycinin precursor [Glycine m  (240 aa)
 initn:  47 init1:  47 opt:  62  Z-score: 93.6  bits: 23.1 E():  3.1
Smith-Waterman score: 62;  26.087% identity (60.870% similar) in 46 aa overlap (2-45:127-172)

                                          10        20
RF_1_+                          TYDPHVRMECPTVVYI-RSRGYP-IISIDHL
                                :.::   ..   .:.:. :..:   ... .
gi|169 RISTLNSLTLPALRQFGLSAQYLVLYRNGIYSPHWNLNANSVIYVTRGKGRVRVVNCQGN
          100       110       120       130       140       150


        30        40        50        60        70        80
RF_1_+ PISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTI
          :.  .. :   :.:    :
gi|169 PVFDGDLTRGQLLLVPQNFVVADQGGKQGLEYVVFKTQHNAVSSYIKDLFRAIPSEVLSN
          160       170       180       190       200       210


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 92.2  bits: 23.8 E():  3.7
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (21-79:292-347)

                    10        20        30        40        50
RF_1_+              TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                            :: .:   . .  . : .:..     :!...:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
              270       280       290       300       310


              60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTI
        :.:   :     :::. .  . .::.   .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         320       330       340       350       360       370


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 91.3  bits: 22.8 E():  4.1
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (31-54:13-35)

              10        20        30        40        50        60
RF_1_+ TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASL
                            .. .:: ::   :.::.. :....
gi|898              MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGK
                              10        20        30        40


              70        80
RF_1_+ SGPNLQKTAYPSLVRPARTI
```

```
gi|898 DWLDAKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIK
                50        60        70        80        90       100


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 90.1  bits: 23.2 E():  4.8
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (25-76:13-63)

                        10        20        30        40        50        60
RF_1_+ TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASL
                            :.  :: . :  : : . . :. ::... :   ::.
gi|212            MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASM
                          10        20        30        40


                70        80
RF_1_+ SGPNLQKTAYPSLVRPARTI
        .    . :.  .: ::
gi|212 GYKAGMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWA
            50        60        70        80        90       100


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 89.6  bits: 21.9 E():  5.1
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (16-72:72-134)

                            10        20        30        40
RF_1_+             TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                       .: . : . ... .: :.   :::   :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
            50        60        70        80        90       100


                50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTI
        :.   : :  .::..: :  :       :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            110       120       130       140       150       160


gi|159 ENSEKTTMPLW
            170


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 89.6  bits: 22.6 E():  5.1
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (31-54:19-41)

                        10        20        30        40        50        60
RF_1_+ TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASL
                            .. ::: :   :.::.. :....
gi|115            MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGS
                          10        20        30        40


                70        80
```

```
RF_1_+ SGPNLQKTAYPSLVRPARTI


gi|115 DWLDAKATWYGKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIK
          50        60        70        80        90       100


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 88.4  bits: 21.9 E():     6
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (16-72:105-167)

                                10        20        30        40
RF_1_+              TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                .:. :. ... .: :.   ::: :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
          80        90       100       110       120       130


             50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTI
       :.   : :  .::..: :   :      :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
          140       150       160       170       180       190


gi|159 ENSEKTTMPLW
          200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 88.1  bits: 21.9 E():  6.3
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (16-72:114-176)

                                10        20        30        40
RF_1_+              TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                .:. :. ... .: :.   ::: :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
             90       100       110       120       130       140


             50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTI
       :.   : :  .::..: :   :      :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
          150       160       170       180       190       200


gi|162 ENSEKTTMPLW
          210


>>gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea mays]  (252 aa)
 initn:  52 init1:  52 opt:  58  Z-score: 86.9  bits: 22.0 E():  7.3
Smith-Waterman score: 58;  40.000% identity (75.000% similar) in 20 aa overlap (35-54:6-24)

              10        20        30        40        50        60
RF_1_+ HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPN
```

```
                                 ::  ::    :.::.. :....
gi|898                          ARALVFLVSGAW-CGPPKVPPGKNITATYGKDWLD
                                  10        20        30


         70         80
RF_1_+ LQKTAYPSLVRPARTI


gi|898 AKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKP
           40        50        60        70        80        90


>>gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60S ri  (392 aa)
 initn:  45 init1:  45 opt:  59  Z-score: 85.3  bits: 22.3 E():  8.9
Smith-Waterman score: 59;  30.769% identity (53.846% similar) in 52 aa overlap (25-76:13-63)

                  10        20        30        40        50        60
RF_1_+ TYDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASL
                        :.   :: . :   : : .   . :. ::.. :    ::.
gi|833            MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKDDPKKPVHLTASM
                     10        20        30        40


         70         80
RF_1_+ SGPNLQKTAYPSLVRPARTI
         .    . :.  .: ::
gi|833 GYKAGMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWA
           50        60        70        80        90        100


>>gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=Proba  (139 aa)
 initn:  52 init1:  52 opt:  54  Z-score: 84.8  bits: 20.7 E():  9.5
Smith-Waterman score: 54;  39.286% identity (60.714% similar) in 28 aa overlap (6-32:87-113)

                              10        20        30
RF_1_+              TYDPHVRMECPTV-VYIRSRGYPIISIDHLPISSA
                      : :.: :: :  :.    :. :. : :..
gi|391 MKTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPV-SLRHGPVTGP
           60        70        80        90        100       110


         40        50        60        70        80
RF_1_+ FLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTI


gi|391 SDPAHKARLERPQIRVPPPAPEKA
           120       130


>>gi|741844|prf||2008179A major allergen Par j I         (143 aa)
 initn:  52 init1:  52 opt:  54  Z-score: 84.6  bits: 20.7 E():  9.8
Smith-Waterman score: 54;  39.286% identity (60.714% similar) in 28 aa overlap (6-32:81-107)

                              10        20        30
RF_1_+              TYDPHVRMECPTV-VYIRSRGYPIISIDHLPISSA
                      : :.: :: :  :.    :. :. : :..
```

gi|741 MKTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPV-SLRHGPVTGP
                60   70   80   90   100

              40   50   60   70   80
RF_1_+ FLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTI

gi|741 SRSRPPTKHGWRDPRLEFRPPHRKKPNPAFSTLG
        110   120   130   140

80 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:37 2010 done: Fri Apr 30 23:42:37 2010
 Total Scan time: 0.090 Total Display time: 0.010

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_+1: 3 - 82 80 aa - 80 aa
 vs /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

        opt   E()
< 20   2   0:=
  22   0   0:   one = represents 3 library sequences
  24   2   0:=
  26   1   0:=
  28   19   0:======
  30   18   2:*=====
  32   30   8:==*=======
  34   42   21:======*=======
  36   55   44:==============*====
  38   61   72:==================== *
  40   98   101:================================*
  42   112   123:===================================== *
  44   133   136:==*
  46   138   138:==*
  48   101   132:================================   *
  50   103   121:===============================   *
  52   81   106:=========================   *
  54   80   91:========================   *

```
 56   114    76:========================*============
 58    83    62:====================*======
 60    39    50:=============    *
 62    28    40:=========    *
 64    34    32:=========*=
 66    19    25:======= *
 68     8    20:===    *
 70    17    16:=====*
 72     6    12:== *
 74    10    10:===*
 76     2     7:= *
 78     3     6:=*
 80     5     4:=*
 82     4     3:*=
 84     4     3:*=
 86     1     2:*
 88     4     2:*=            inset = represents 1 library sequences
 90     3     1:*
 92     2     1:*        :*=
 94     2     1:*        :*=
 96     1     1:*        :*
 98     2     0:=        *==
100     0     0:         *
102     0     0:         *
104     0     0:         *
106     1     0:=        *=
108     1     0:=        *=
110     2     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))= 4.5980+/-0.00351; mu= 9.7907+/- 0.180
 mean_var=38.7501+/-10.361, 0's: 2 Z-trim: 4 B-trim: 157 in 2/41
 Lambda= 0.206034
 Kolmogorov-Smirnov statistic: 0.0625 (N=29) 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: 0.090
The best scores are: opt bits E(1471)
gi|218059718|emb|CAT99612.1| thaumatin-like protei (158)   71 25.7   0.33
gi|218059715|emb|CAT99611.1| thaumatin-like protei (158)   71 25.7   0.33
gi|60418848|gb|AAX19851.1| thaumatin-like protein (246)   71 25.8   0.51
gi|60418842|gb|AAX19848.1| thaumatin-like protein (246)   71 25.8   0.51
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur (214)   65 24.0   1.5
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K (473)   68 24.9   1.7

```
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.7    2.2
gi|169971|gb|AAA33965.1| glycinin precursor [Glyci ( 240)   62 23.1    3.2
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.7    3.2
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.7    3.2
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.7    3.2
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.7    3.8
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.8    4.2
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 23.1      5
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.9    5.3
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.5    5.3
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.9    6.2
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.9    6.5
gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea m ( 252)   58 21.9    7.6
gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60 ( 392)   59 22.2    9.3
gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=P ( 139)   54 20.7    9.8

>>gi|218059718|emb|CAT99612.1| thaumatin-like protein [M  (158 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 110.9  bits: 25.7 E(): 0.33
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (4-49:99-141)

                                 10        20        30
RF_1_+                    YDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                               .: : ::. . ...    .::      .::
gi|218 DVSLVDGFNLPMYVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
          70        80        90       100       110       120

              40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIE
         :  ::     :.::.  :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
           130       140       150


>>gi|218059715|emb|CAT99611.1| thaumatin-like protein [M  (158 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 110.9  bits: 25.7 E(): 0.33
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (4-49:99-141)

                                 10        20        30
RF_1_+                    YDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                               .: : ::. . ...    .::      .::
gi|218 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
          70        80        90       100       110       120

              40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIE
         :  ::     :.::.  :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
           130       140       150


>>gi|60418848|gb|AAX19851.1| thaumatin-like protein prec  (246 aa)
```

```
 initn:  46 init1:  46 opt:  71  Z-score: 107.7  bits: 25.8 E(): 0.51
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (4-49:159-201)


                                     10        20        30
RF_1_+                      YDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                              .: : ::. . ...    .::      .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
           130       140       150       160       170       180


                 40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIE
          :  ::     :.::.  :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
            190       200       210       220       230       240


>>gi|60418842|gb|AAX19848.1| thaumatin-like protein prec  (246 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 107.7  bits: 25.8 E(): 0.51
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (4-49:159-201)


                                     10        20        30
RF_1_+                      YDPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                              .: : ::. . ...    .::      .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
           130       140       150       160       170       180


                 40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIE
          :  ::     :.::.  :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
            190       200       210       220       230       240


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score: 99.1  bits: 24.0 E():  1.5
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (15-71:114-176)


                                   10        20        30
RF_1_+                    YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                            .: . : ... .: :.   :::   : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
             90       100       110       120       130       140


      40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIE
         . : : .::..: :  :     : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             150       160       170       180       190       200


gi|162 ENSEKTTISLW
           210
```

```
>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly  (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score: 98.0  bits: 24.9 E():  1.7
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (20-78:301-356)

                          10        20        30        40
RF_1_+             YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                            :: .:   .  . : ..:.    ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            280       290       300       310       320

          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIE
        :.:   :   ::: .   . ..:.  : .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
        330       340       350       360       370       380


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score: 96.4  bits: 22.7 E():  2.2
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (15-71:29-91)

                          10        20        30        40
RF_1_+              YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                            .: . :. ... .: :.   ::: : :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
              10        20        30        40        50        60

          50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIE
        .   : :  .::..: :   :      : ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
              70        80        90       100       110       120

gi|159 SEKTTMPLW


>>gi|169971|gb|AAA33965.1| glycinin precursor [Glycine m  (240 aa)
 initn:  47 init1:  47 opt:  62  Z-score: 93.4  bits: 23.1 E():  3.2
Smith-Waterman score: 62;  26.087% identity (60.870% similar) in 46 aa overlap (1-44:127-172)

                          10        20
RF_1_+              YDPHVRMECPTVVYI-RSRGYP-IISIDHL
                     :.::  ..  .:.:. :..:   ... .
gi|169 RISTLNSLTLPALRQFGLSAQYLVLYRNGIYSPHWNLNANSVIYVTRGKGRVRVVNCQGN
            100       110       120       130       140       150

          30        40        50        60        70        80
RF_1_+ PISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIE
        :. .. : :.:   :
```

```
gi|169 PVFDGDLTRGQLLLVPQNFVVADQGGKQGLEYVVFKTQHNAVSSYIKDLFRAIPSEVLSN
           160       170       180       190       200       210


>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 93.3  bits: 23.7 E():  3.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (20-78:216-271)


                        10        20        30        40
RF_1_+             YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             :: .:   .  . . : .:..    :: ...:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


           50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIE
        :.:  :   :::  .  . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 93.3  bits: 23.7 E():  3.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (20-78:216-271)


                        10        20        30        40
RF_1_+             YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             :: .:   .  . . : .:..    :: ...:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


           50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIE
        :.:  :   :::  .  . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 93.3  bits: 23.7 E():  3.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (20-78:216-271)


                        10        20        30        40
RF_1_+             YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             :: .:   .  . . : .:..    :: ...:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


           50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIE
        :.:  :   :::  .  . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
```

```
                250         260         270         280         290         300


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 91.9  bits: 23.7 E():  3.8
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (20-78:292-347)


                        10        20        30        40
RF_1_+            YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                            ::  .:    .  .   . : .:..     ::..:
gi|187  GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            270       280       290       300       310


       50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIE
         :.:   :    :::  .   .  .::.   .  .:
gi|187  ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
     320       330       340       350       360       370


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 91.1  bits: 22.8 E():  4.2
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (30-53:13-35)


               10        20        30        40        50        60
RF_1_+ YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLS
                             .. .:: ::   :.:.:.. :....
gi|898                MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKD
                              10        20        30        40


               70        80
RF_1_+ GPNLQKTAYPSLVRPARTIE


gi|898  WLDAKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKC
                50        60        70        80        90       100


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 89.8  bits: 23.1 E():     5
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (24-75:13-63)


               10        20        30        40        50        60
RF_1_+ YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLS
                            :.  :: . :  : : .   . .:. ::.  :    ::..
gi|212                MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMG
                              10        20        30        40


               70        80
RF_1_+ GPNLQKTAYPSLVRPARTIE
           .  :.   .:  ::  ::
gi|212  YKAGMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAE
            50        60        70        80        90       100
```

```
>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 89.4  bits: 21.9 E():  5.3
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (15-71:72-134)

                              10        20        30        40
RF_1_+                YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                               .:. :. ... .: :.   :::    :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
              50        60        70        80        90       100

                   50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIE
       :. .  : :  .::..: :  :     : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
              110       120       130       140       150       160

gi|159 ENSEKTTMPLW
              170


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 89.3  bits: 22.5 E():  5.3
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (30-53:19-41)

                       10        20        30        40        50        60
RF_1_+ YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLS
                                 .. :::  :   :.:... :....
gi|115           MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSD
                         10        20        30        40

               70        80
RF_1_+ GPNLQKTAYPSLVRPARTIE

gi|115 WLDAKATWYGKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKC
           50        60        70        80        90       100


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 88.1  bits: 21.9 E():  6.2
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (15-71:105-167)

                              10        20        30        40
RF_1_+                YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                               .:. :. ... .: :.   :::    :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
              80        90       100       110       120       130

                   50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIE
       :. .  : :  .::..: :  :     : ::::
```

```
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             140       150       160       170       180       190


gi|159 ENSEKTTMPLW
             200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58 Z-score: 87.8 bits: 21.9 E():  6.5
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (15-71:114-176)


                          10        20        30        40
RF_1_+               YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                             .:. .: ....  .: :.    :::   :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
              90       100       110       120       130       140


               50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIE
          :.  : : .::..: :   :    : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             150       160       170       180       190       200


gi|162 ENSEKTTMPLW
             210


>>gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea mays]  (252 aa)
 initn:  52 init1:  52 opt:  58 Z-score: 86.6 bits: 21.9 E():  7.6
Smith-Waterman score: 58;  40.000% identity (75.000% similar) in 20 aa overlap (34-53:6-24)


             10        20        30        40        50        60
RF_1_+ HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPN
                                     :: ::   :.::.. :....
gi|898                          ARALVFLVSGAW-CGPPKVPPGKNITATYGKDWLD
                                          10        20        30


             70        80
RF_1_+ LQKTAYPSLVRPARTIE

gi|898 AKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKP
              40        50        60        70        80        90


>>gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60S ri  (392 aa)
 initn:  45 init1:  45 opt:  59 Z-score: 84.9 bits: 22.2 E():  9.3
Smith-Waterman score: 59;  30.769% identity (53.846% similar) in 52 aa overlap (24-75:13-63)


             10        20        30        40        50        60
RF_1_+ YDPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLS
                     :.  :: . :  : :.   . .:. ::.. :   ::.
gi|833              MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKDDPKKPVHLTASMG
```

```
                          10        20        30        40

                70        80
RF_1_+ GPNLQKTAYPSLVRPARTIE
             .  :.   .:  ::
gi|833 YKAGMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAE
        50        60        70        80        90       100


>>gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=Proba  (139 aa)
 initn:  52 init1:  52 opt:  54  Z-score: 84.6  bits: 20.7 E():  9.8
Smith-Waterman score: 54;  39.286% identity (60.714% similar) in 28 aa overlap (5-31:87-113)


                                       10        20        30
RF_1_+                     YDPHVRMECPTV-VYIRSRGYPIISIDHLPISSA
                           :  :.:  ::  :    :.     :.  :.   :  :..
gi|391 MKTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPV-SLRHGPVTGP
           60        70        80        90       100       110


           40        50        60        70        80
RF_1_+ FLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIE


gi|391 SDPAHKARLERPQIRVPPPAPEKA
          120       130




80 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:37 2010 done: Fri Apr 30 23:42:38 2010
 Total Scan time:  0.090 Total Display time:  0.010


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 searchs 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: 4  - 83 80 aa - 80 aa
 vs   /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library


       opt      E()
< 20     2     0:=
  22     0     0:               one = represents 3 library sequences
  24     2     0:=
```

```
  26     1     0:=
  28    13     0:====
  30    22     2:*======
  32    31     8:==*=======
  34    43    21:=====*========
  36    57    44:=============*====
  38    58    72:==================    *
  40    97   101:===============================*
  42   113   123:=================================    *
  44   127   136:=========================================    *
  46   147   138:============================================*===
  48   108   132:==================================           *
  50    98   121:===============================          *
  52    82   106:=========================            *
  54    82    91:=========================   *
  56   108    76:========================*=========
  58    87    62:==================*=======
  60    37    50:============    *
  62    26    40:========       *
  64    37    32:=========*==
  66    18    25:=====   *
  68     9    20:===    *
  70    17    16:=====*
  72     6    12:== *
  74     6    10:== *
  76     5     7:==*
  78     3     6:=*
  80     3     4:=*
  82     4     3:*=
  84     4     3:*=
  86     1     2:*
  88     4     2:*=          inset = represents 1 library sequences
  90     2     1:*
  92     2     1:*         :*=
  94     2     1:*         :*=
  96     1     1:*         :*
  98     2     0:=         *==
 100     0     0:          *
 102     0     0:          *
 104     0     0:          *
 106     1     0:=         *=
 108     1     0:=         *=
 110     2     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))= 4.5111+/-0.00346; mu= 10.2165+/- 0.177
mean_var=38.8821+/-10.161, 0's: 2 Z-trim: 6 B-trim: 125 in 1/42
Lambda= 0.205683
Kolmogorov-Smirnov statistic: 0.0639 (N=29) 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: 0.090
The best scores are: opt bits E(1471)
gi|218059718|emb|CAT99612.1| thaumatin-like protei (158)   71 25.7   0.34
gi|218059715|emb|CAT99611.1| thaumatin-like protei (158)   71 25.7   0.34
gi|60418848|gb|AAX19851.1| thaumatin-like protein (246)   71 25.8   0.51
gi|60418842|gb|AAX19848.1| thaumatin-like protein (246)   71 25.8   0.51
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur (214)   65 24.0   1.5
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K (473)   68 24.9   1.7
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t (129)   61 22.7   2.2
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen (373)   64 23.7   3.2
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen (373)   64 23.7   3.2
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen (373)   64 23.7   3.2
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen (455)   64 23.8   3.8
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m (263)   61 22.8   4.2
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit (392)   62 23.1   5
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t (172)   58 21.9   5.3
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full (269)   60 22.5   5.3
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t (205)   58 21.9   6.2
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor (214)   58 21.9   6.5
gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea m (252)   58 21.9   7.5
gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60 (392)   59 22.3   9.2
gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=P (139)   54 20.7   9.9

>>gi|218059718|emb|CAT99612.1| thaumatin-like protein [M (158 aa)
 initn: 46 init1: 46 opt: 71 Z-score: 110.9 bits: 25.7 E(): 0.34
Smith-Waterman score: 71; 31.250% identity (52.083% similar) in 48 aa overlap (3-48:99-141)

                                  10   20   30
RF_1_+   DPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                 .: : ::. . ...   .::   .::
gi|218 DVSLVDGFNLPMYVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
           70   80   90   100   110   120

               40   50   60   70   80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIEL
         :   ::   :.::.   :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
            130   140   150

>>gi|218059715|emb|CAT99611.1| thaumatin-like protein [M (158 aa)
 initn: 46 init1: 46 opt: 71 Z-score: 110.9 bits: 25.7 E(): 0.34
Smith-Waterman score: 71; 31.250% identity (52.083% similar) in 48 aa overlap (3-48:99-141)

```
                                            10        20        30
RF_1_+                          DPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                  .: : ::. . ...    .::      .::
gi|218 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
        70        80        90       100       110       120


                             40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIEL
         :  ::      :.::.  :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
            130       140       150


>>gi|60418848|gb|AAX19851.1| thaumatin-like protein prec  (246 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 107.7  bits: 25.8 E(): 0.51
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (3-48:159-201)

                                            10        20        30
RF_1_+                          DPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                  .: : ::. . ...    .::      .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
         130       140       150       160       170       180


                             40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIEL
         :  ::      :.::.  :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
            190       200       210       220       230       240


>>gi|60418842|gb|AAX19848.1| thaumatin-like protein prec  (246 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 107.7  bits: 25.8 E(): 0.51
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (3-48:159-201)

                                            10        20        30
RF_1_+                          DPHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                  .: : ::. . ...    .::      .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
         130       140       150       160       170       180


                             40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIEL
         :  ::      :.::.  :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
            190       200       210       220       230       240


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score: 99.0  bits: 24.0 E():  1.5
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (14-70:114-176)
```

```
                          10        20        30
RF_1_+                DPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                               .:. : . ... .: :.   :::   : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
            90       100       110       120       130       140


       40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIEL
          .  : :  .::..: :  :       : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            150       160       170       180       190       200


gi|162 ENSEKTTISLW
           210


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly  (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score: 98.1  bits: 24.9 E():  1.7
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (19-77:301-356)


                          10        20        30        40
RF_1_+                DPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                               :: .:   . .  . : .:..    ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
                280       290       300       310       320


          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIEL
          :.:  :   :::  .  . .::.  : .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
           330       340       350       360       370       380


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score: 96.3  bits: 22.7 E():  2.2
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (14-70:29-91)


                          10        20        30        40
RF_1_+                DPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                               .:. : . ... .: :.   :::   : :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
                10        20        30        40        50        60


          50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIEL
          .  : :  .::..: :  :       : ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
                70        80        90        100       110       120


gi|159 SEKTTMPLW
```

```
>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 93.4  bits: 23.7 E():  3.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (19-77:216-271)

                          10        20        30        40
RF_1_+             DPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             :: .:   .  .  . : .:..      ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240

            50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIEL
        :.:   :    :::  .   . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 93.4  bits: 23.7 E():  3.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (19-77:216-271)

                          10        20        30        40
RF_1_+             DPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             :: .:   .  .  . : .:..      ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240

            50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIEL
        :.:   :    :::  .   . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 93.4  bits: 23.7 E():  3.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (19-77:216-271)

                          10        20        30        40
RF_1_+             DPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             :: .:   .  .  . : .:..      ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240

            50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIEL
        :.:   :    :::  .   . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300
```

```
>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 92.0  bits: 23.8 E():  3.8
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (19-77:292-347)

                      10        20        30        40
RF_1_+            DPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                            ::  .:    .   .   . :  .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            270       280       290       300       310

          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIEL
         :.:   :   :::  .   . .:..  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          320       330       340       350       360       370


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 91.1  bits: 22.8 E():  4.2
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (29-52:13-35)

                     10        20        30        40        50        60
RF_1_+ DPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSG
                            .. .:: ::   .:.::.. :....
gi|898            MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDW
                         10        20        30        40

             70        80
RF_1_+ PNLQKTAYPSLVRPARTIEL

gi|898 LDAKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCD
            50        60        70        80        90       100


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 89.8  bits: 23.1 E():    5
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (23-74:13-63)

                     10        20        30        40        50        60
RF_1_+ DPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSG
                          :.  :: . .:  : :. . . :. ::.. :   ::..
gi|212            MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGY
                         10        20        30        40

             70        80
RF_1_+ PNLQKTAYPSLVRPARTIEL
          . :.  .: ::
gi|212 KAGMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEH
            50        60        70        80        90       100


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
```

```
 initn:  37 init1:  37 opt:  58  Z-score: 89.4  bits: 21.9 E():  5.3
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (14-70:72-134)

                              10        20        30
RF_1_+                        DPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                    .:. :. ... .: :.   ::: 	 :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
            50        60        70        80        90       100

      40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIEL
        :.  :: .::..: :   :     : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             110       120       130       140       150       160

gi|159 ENSEKTTMPLW
              170


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp   (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 89.4  bits: 22.5 E():  5.3
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (29-52:19-41)

              10        20        30        40        50        60
RF_1_+ DPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSG
                         .. ::: :   :.:::.. :....
gi|115          MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDW
                         10        20        30        40

              70        80
RF_1_+ PNLQKTAYPSLVRPARTIEL

gi|115 LDAKATWYGKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCD
         50        60        70        80        90       100


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 88.1  bits: 21.9 E():  6.2
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (14-70:105-167)

                              10        20        30
RF_1_+                        DPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                    .:. :. ... .: :.   ::: 	 :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
             80        90       100       110       120       130

      40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIEL
        :.  :: .::..: :   :     : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            140       150       160       170       180       190
```

```
gi|159 ENSEKTTMPLW
          200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.8  bits: 21.9 E():  6.5
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (14-70:114-176)


                            10        20        30
RF_1_+                DPHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                             .:. :. ... .: :.   :::   :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
               90       100       110       120       130       140


         40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIEL
         :.   : :  .::..: :   :     : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
               150       160       170       180       190       200


gi|162 ENSEKTTMPLW
          210


>>gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea mays]  (252 aa)
 initn:  52 init1:  52 opt:  58  Z-score: 86.6  bits: 21.9 E():  7.5
Smith-Waterman score: 58;  40.000% identity (75.000% similar) in 20 aa overlap (33-52:6-24)


                 10        20        30        40        50        60
RF_1_+ HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPN
                              :: ::   :.:::.. :....
gi|898                   ARALVFLVSGAW-CGPPKVPPGKNITATYGKDWLD
                                     10        20        30


                 70        80
RF_1_+ LQKTAYPSLVRPARTIEL

gi|898 AKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKP
                40        50        60        70        80        90


>>gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60S ri  (392 aa)
 initn:  45 init1:  45 opt:  59  Z-score: 85.0  bits: 22.3 E():  9.2
Smith-Waterman score: 59;  30.769% identity (53.846% similar) in 52 aa overlap (23-74:13-63)


                 10        20        30        40        50        60
RF_1_+ DPHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSG
                        :.  :: . :  : :.   .:. ::.  :    :::.
gi|833            MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKDDPKKPVHLTASMGY
                          10        20        30        40
```

```
                     70          80
RF_1_+ PNLQKTAYPSLVRPARTIEL
           .  :.   .: ::
gi|833 KAGMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEH
         50        60        70        80        90       100


>>gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=Proba   (139 aa)
 initn:  52 init1:  52 opt:  54  Z-score: 84.5  bits: 20.7 E():  9.9
Smith-Waterman score: 54;  39.286% identity (60.714% similar) in 28 aa overlap (4-30:87-113)


                                    10        20        30
RF_1_+                      DPHVRMECPTV-VYIRSRGYPIISIDHLPISSA
                            :  :.: :: :   :.    :. :. : :..
gi|391 MKTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPV-SLRHGPVTGP
         60        70        80        90       100       110


          40        50        60        70        80
RF_1_+ FLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIEL


gi|391 SDPAHKARLERPQIRVPPPAPEKA
          120       130




80 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:38 2010 done: Fri Apr 30 23:42:38 2010
 Total Scan time:  0.090 Total Display time:  0.010

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_+1: 5  - 84 80 aa - 80 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library


      opt      E()
< 20     2    0:=
  22     0    0:            one = represents 3 library sequences
  24     2    0:=
  26     0    0:
  28    12    0:====
```

```
 30     22        2:*=======
 32     34        8:==*=========
 34     46       21:======*=========
 36     53       44:=============*===
 38     58       72:===================    *
 40     96      101:================================= *
 42    111      123:===================================    *
 44    129      136:=============================================   *
 46    149      138:===============================================*====
 48    105      132:===================================           *
 50    100      121:===============================              *
 52     85      106:==========================              *
 54     83       91:=========================   *
 56    105       76:=======================*=========
 58     88       62:==================*=========
 60     33       50:==========       *
 62     27       40:========      *
 64     33       32:=========*
 66     19       25:====== *
 68      9       20:===    *
 70     19       16:=====*=
 72      7       12:===*
 74      7       10:===*
 76      1        7:= *
 78      2        6:=*
 80      5        4:=*
 82      3        3:*
 84      5        3:*=
 86      2        2:*
 88      5        2:*=            inset = represents 1 library sequences
 90      3        1:*
 92      1        1:*          :*
 94      3        1:*          :*==
 96      1        1:*          :*
 98      2        0:=          *==
100      0        0:           *
102      0        0:           *
104      0        0:           *
106      1        0:=          *=
108      1        0:=          *=
110      2        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))= 4.5184+/-0.00358; mu= 10.1385+/- 0.183
mean_var=38.7303+/-10.428, 0's: 2 Z-trim: 4  B-trim: 157 in 2/41
```

```
 Lambda= 0.206086
 Kolmogorov-Smirnov  statistic: 0.0639 (N=29) 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:  0.090
The best scores are:                              opt bits E(1471)
gi|218059718|emb|CAT99612.1| thaumatin-like protei ( 158)   71 25.8    0.33
gi|218059715|emb|CAT99611.1| thaumatin-like protei ( 158)   71 25.8    0.33
gi|60418848|gb|AAX19851.1| thaumatin-like protein  ( 246)   71 25.8    0.5
gi|60418842|gb|AAX19848.1| thaumatin-like protein  ( 246)   71 25.8    0.5
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 24.0    1.5
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 25.0    1.7
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.8    2.1
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.8    3.1
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.8    3.1
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.8    3.1
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.8    3.7
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.8    4.2
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 23.2    4.9
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.9    5.2
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.5    5.2
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.9    6.1
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.9    6.4
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   58 21.9    6.6
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   58 21.9    6.6
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   58 21.9    6.6
gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea m ( 252)   58 21.9    7.4
gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60 ( 392)   59 22.3    9.1
gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=P ( 139)   54 20.7    9.7
gi|741844|prf||2008179A major allergen Par j I     ( 143)   54 20.7     10


>>gi|218059718|emb|CAT99612.1| thaumatin-like protein [M  (158 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 111.0  bits: 25.8 E(): 0.33
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (2-47:99-141)

                              10        20        30
RF_1_+                     PHVRMECPTVVYIRSRGYPIISIDHLPISSA
                            .: : ::. . ...    .::     .::
gi|218 DVSLVDGFNLPMYVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
          70        80        90       100       110       120

                   40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
         :  ::     :.::.  :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
            130       140       150


>>gi|218059715|emb|CAT99611.1| thaumatin-like protein [M  (158 aa)
```

```
 initn:  46 init1:  46 opt:  71  Z-score: 111.0  bits: 25.8 E(): 0.33
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (2-47:99-141)


                                      10        20        30
RF_1_+                         PHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                 .: : ::. . ...     .::      .::
gi|218 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
         70        80        90       100       110       120


                               40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
           :  ::     :.::.  :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
              130       140       150


>>gi|60418848|gb|AAX19851.1| thaumatin-like protein prec  (246 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 107.8  bits: 25.8 E():  0.5
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (2-47:159-201)


                                      10        20        30
RF_1_+                         PHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                 .: : ::. . ...     .::      .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
        130       140       150       160       170       180


                               40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
           :  ::     :.::.  :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
             190       200       210       220       230       240


>>gi|60418842|gb|AAX19848.1| thaumatin-like protein prec  (246 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 107.8  bits: 25.8 E():  0.5
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (2-47:159-201)


                                      10        20        30
RF_1_+                         PHVRMECPTVVYIRSRGYPIISIDHLPISSA
                                 .: : ::. . ...     .::      .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
        130       140       150       160       170       180


                               40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
           :  ::     :.::.  :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
             190       200       210       220       230       240


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]  (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score: 99.2  bits: 24.0 E():  1.5
```

```
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (13-69:114-176)

                                  10        20        30
RF_1_+                    PHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                                    .: . : . ... .: :.    :::   : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
              90       100       110       120       130       140

          40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQ
         . : : .:: .. : :  :      :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
              150       160       170       180       190       200

gi|162 ENSEKTTISLW
              210


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly  (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score:  98.3  bits: 25.0 E():  1.7
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (18-76:301-356)

                                  10        20        30        40
RF_1_+                    PHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                    :: .:   .  .   . : .:..     ::. .:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
              280       290       300       310       320

            50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
         :.:  :   ::: .   . . .:.  : .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          330       340       350       360       370       380


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score:  96.4  bits: 22.8 E():  2.1
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (13-69:29-91)

                                  10        20        30           40
RF_1_+                    PHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                                    .: . : . ... .: :.    :::   : :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
              10        20        30        40        50        60

            50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQ
         . : : .:: .. : :  :      :  ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
              70        80        90       100       110       120
```

gi|159 SEKTTMPLW

>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly (373 aa)
 initn: 36 init1: 36 opt: 64 Z-score: 93.6 bits: 23.8 E(): 3.1
Smith-Waterman score: 64; 28.333% identity (55.000% similar) in 60 aa overlap (18-76:216-271)

```
                       10        20        30        40
RF_1_+              PHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                            :: .:   .  .  . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240

          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
         :.:  :   :::  .   . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300
```

>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly (373 aa)
 initn: 36 init1: 36 opt: 64 Z-score: 93.6 bits: 23.8 E(): 3.1
Smith-Waterman score: 64; 28.333% identity (55.000% similar) in 60 aa overlap (18-76:216-271)

```
                       10        20        30        40
RF_1_+              PHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                            :: .:   .  .  . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240

          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
         :.:  :   :::  .   . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300
```

>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly (373 aa)
 initn: 36 init1: 36 opt: 64 Z-score: 93.6 bits: 23.8 E(): 3.1
Smith-Waterman score: 64; 28.333% identity (55.000% similar) in 60 aa overlap (18-76:216-271)

```
                       10        20        30        40
RF_1_+              PHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                            :: .:   .  .  . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240

          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
         :.:  :   :::  .   . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
```

```
              250        260        270        280        290        300

>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64 Z-score: 92.1 bits: 23.8 E():  3.7
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (18-76:292-347)

                              10        20        30        40
RF_1_+               PHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             ::  .:   .  .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           270        280        290        300        310

             50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
        :.:  :   ::: .  .  .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
      320        330        340        350        360        370

>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61 Z-score: 91.3 bits: 22.8 E():  4.2
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (28-51:13-35)

                  10        20        30        40        50        60
RF_1_+ PHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGP
                          .. .:: ::   :.:::.. :....
gi|898              MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWL
                          10        20        30        40

              70        80
RF_1_+ NLQKTAYPSLVRPARTIELQ

gi|898 DAKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDK
            50        60        70        80        90        100

>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62 Z-score: 90.0 bits: 23.2 E():  4.9
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (22-73:13-63)

                  10        20        30        40        50        60
RF_1_+ PHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGP
                        :.  :: . :  : : .  . .:. ::.. :   ::..
gi|212             MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYK
                          10        20        30        40        50

              70        80
RF_1_+ NLQKTAYPSLVRPARTIELQ
        .  :.  .: ::
gi|212 AGMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHL
              60        70        80        90        100       110
```

```
>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 89.5  bits: 21.9 E():  5.2
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (13-69:72-134)

                                  10        20        30
RF_1_+                   PHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                 .: . :. ... .: :.   ::: :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                50        60        70        80        90       100

        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQ
       :.  :: .::..: :  :      : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             110       120       130       140       150       160

gi|159 ENSEKTTMPLW
            170


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 89.5  bits: 22.5 E():  5.2
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (28-51:19-41)

                10        20        30        40        50        60
RF_1_+ PHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGP
                               .. ::: :   :.::.. :....
gi|115         MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWL
                        10        20        30        40        50

               70        80
RF_1_+ NLQKTAYPSLVRPARTIELQ

gi|115 DAKATWYGKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDK
                  60        70        80        90       100       110


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 88.3  bits: 21.9 E():  6.1
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (13-69:105-167)

                                  10        20        30
RF_1_+                   PHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                 .: . :. ... .: :.   ::: :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
               80        90       100       110       120       130

        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQ
       :.  :: .::..: :  :      : ::::
```

```
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            140       150       160       170       180       190


gi|159 ENSEKTTMPLW
            200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.9  bits: 21.9 E():  6.4
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (13-69:114-176)

                                 10        20        30
RF_1_+             PHVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                             .:. :. ... .: :.   :::   :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                90       100       110       120       130       140


        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQ
        :.   : :  .::..: :  :      :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            150       160       170       180       190       200


gi|162 ENSEKTTMPLW
            210


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]       (224 aa)
 initn:  42 init1:  42 opt:  58  Z-score: 87.6  bits: 21.9 E():  6.6
Smith-Waterman score: 58;  20.779% identity (51.948% similar) in 77 aa overlap (8-80:130-203)

                                 10        20        30
RF_1_+             PHVRMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                             :.   . .:..  .  ...   :::.    ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
        100       110       120       130       140       150


          40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
        :.       :  . : :.   :::  ..  .    ..  : :..:
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
        160       170       180       190       200       210


gi|162 RGPFPIIV
          220


>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta  (224 aa)
 initn:  42 init1:  42 opt:  58  Z-score: 87.6  bits: 21.9 E():  6.6
Smith-Waterman score: 58;  20.779% identity (51.948% similar) in 77 aa overlap (8-80:130-203)

                                 10        20            30
```

```
RF_1_+                              PHVRMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                                    :.  . .:..  . ...  :::.   ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
        100       110       120       130       140       150


             40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
       :.       :  . : : :.   :::  ..  .  ..  : : . .:
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
        160       170       180       190       200       210


gi|162 RGPFPIIV
          220


>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]    (224 aa)
 initn: 42 init1: 42 opt: 58  Z-score: 87.6  bits: 21.9 E(): 6.6
Smith-Waterman score: 58;  20.779% identity (51.948% similar) in 77 aa overlap (8-80:130-203)


                                  10        20        30
RF_1_+                              PHVRMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                                    :.  . .:..  . ...  :::.   ....
gi|459 PPFLQPEVMGVSKVKEAMAPKQKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
        100       110       120       130       140       150


             40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ
       :.       :  . : : :.   :::  ..  .  ..  : : . .:
gi|459 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
        160       170       180       190       200       210


gi|459 RGPFPIIV
          220


>>gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea mays]  (252 aa)
 initn: 52 init1: 52 opt: 58  Z-score: 86.8  bits: 21.9 E(): 7.4
Smith-Waterman score: 58;  40.000% identity (75.000% similar) in 20 aa overlap (32-51:6-24)


            10        20        30        40        50        60
RF_1_+ HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPN
                                 :: ::   :.::.. :.....
gi|898                           ARALVFLVSGAW-CGPPKVPPGKNITATYGKDWLD
                                        10        20        30


            70        80
RF_1_+ LQKTAYPSLVRPARTIELQ


gi|898 AKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKP
           40        50        60        70        80        90
```

```
>>gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60S ri  (392 aa)
 initn:  45 init1:  45 opt:  59  Z-score: 85.2  bits: 22.3 E():  9.1
Smith-Waterman score: 59;  30.769% identity (53.846% similar) in 52 aa overlap (22-73:13-63)

                    10        20        30        40        50        60
RF_1_+ PHVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGP
                         :.  ::  .  :   : : .   .  :.  ::.. :   ::..
gi|833          MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKDDPKKPVHLTASMGYK
                     10        20        30        40        50

               70        80
RF_1_+ NLQKTAYPSLVRPARTIELQ
          .  :.   .:  ::
gi|833 AGMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHL
               60        70        80        90       100       110


>>gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=Proba  (139 aa)
 initn:  52 init1:  52 opt:  54  Z-score: 84.7  bits: 20.7 E():  9.7
Smith-Waterman score: 54;  39.286% identity (60.714% similar) in 28 aa overlap (3-29:87-113)

                                    10        20        30
RF_1_+                       PHVRMECPTV-VYIRSRGYPIISIDHLPISSA
                             : :.: :: :  :.    :. :. : :..
gi|391 MKTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPV-SLRHGPVTGP
            60        70        80        90       100       110

               40        50        60        70        80
RF_1_+ FLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ

gi|391 SDPAHKARLERPQIRVPPPAPEKA
           120       130


>>gi|741844|prf||2008179A major allergen Par j I           (143 aa)
 initn:  52 init1:  52 opt:  54  Z-score: 84.4  bits: 20.7 E():   10
Smith-Waterman score: 54;  39.286% identity (60.714% similar) in 28 aa overlap (3-29:81-107)

                                    10        20        30
RF_1_+                       PHVRMECPTV-VYIRSRGYPIISIDHLPISSA
                             : :.: :: :  :.    :. :. : :..
gi|741 MKTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPV-SLRHGPVTGP
            60        70        80        90       100

               40        50        60        70        80
RF_1_+ FLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQ

gi|741 SRSRPPTKHGWRDPRLEFRPPHRKKPNPAFSTLG
          110       120       130       140
```

```
80 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:39 2010 done: Fri Apr 30 23:42:39 2010
 Total Scan time:  0.090 Total Display time:  0.010

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_+1: 6  - 85 80 aa - 80 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

        opt       E()
< 20      2     0:=
  22      0     0:                  one = represents 3 library sequences
  24      2     0:=
  26      0     0:
  28      6     0:==
  30     22     2:*=======
  32     34     8:==*=========
  34     47    21:======*=========
  36     58    44:=============*=====
  38     55    72:=================     *
  40     99   101:===============================*
  42    115   123:======================================= *
  44    135   136:==========================================*
  46    143   138:==========================================*==
  48    106   132:=================================     *
  50     99   121:=============================     *
  52     83   106:=========================         *
  54     84    91:=========================   *
  56     98    76:========================*=======
  58     94    62:==================*==========
  60     33    50:==========       *
  62     30    40:=========    *
  64     31    32:=========*
  66     19    25:======= *
  68     10    20:====   *
  70     17    16:=====*
  72      6    12:== *
  74      8    10:===*
```

```
  76    2     7:= *
  78    2     6:=*
  80    3     4:=*
  82    4     3:*=
  84    3     3:*
  86    1     2:*
  88    7     2:*==          inset = represents 1 library sequences
  90    2     1:*
  92    2     1:*        :*=
  94    2     1:*        :*=
  96    1     1:*        :*
  98    2     0:=        *==
 100    0     0:         *
 102    0     0:         *
 104    0     0:         *
 106    1     0:=        *=
 108    1     0:=        *=
 110    2     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))= 4.4328+/-0.00353; mu= 10.5318+/- 0.181

mean_var=39.3104+/-10.254, 0's: 2 Z-trim: 6 B-trim: 125 in 1/42

Lambda= 0.204560

Kolmogorov-Smirnov statistic: 0.0639 (N=29) 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: 0.090

The best scores are: opt bits E(1471)

```
gi|218059718|emb|CAT99612.1| thaumatin-like protei ( 158)   71 25.7    0.35
gi|218059715|emb|CAT99611.1| thaumatin-like protei ( 158)   71 25.7    0.35
gi|60418848|gb|AAX19851.1| thaumatin-like protein  ( 246)   71 25.7    0.52
gi|60418842|gb|AAX19848.1| thaumatin-like protein  ( 246)   71 25.7    0.52
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 24.0    1.6
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 24.9    1.7
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.7    2.2
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.7    3.2
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.7    3.2
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.7    3.2
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.8    3.8
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.8    4.2
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 23.1      5
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.5    5.3
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.9    5.3
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   59 22.2    5.5
```

```
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   59 22.2    5.5
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   59 22.2    5.5
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.9    6.3
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.9    6.5
gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea m ( 252)   58 21.9    7.5
gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60 ( 392)   59 22.3    9.2
gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=P ( 139)   54 20.6     10


>>gi|218059718|emb|CAT99612.1| thaumatin-like protein [M  (158 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 110.7  bits: 25.7 E(): 0.35
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (1-46:99-141)

                                       10        20        30
RF_1_+                         HVRMECPTVVYIRSRGYPIISIDHLPISSA
                               .: : ::. . ...    .::      .::
gi|218 DVSLVDGFNLPMYVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
          70        80        90       100       110       120

                     40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
         :  ::      :.::.  :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
           130       140       150


>>gi|218059715|emb|CAT99611.1| thaumatin-like protein [M  (158 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 110.7  bits: 25.7 E(): 0.35
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (1-46:99-141)

                                       10        20        30
RF_1_+                         HVRMECPTVVYIRSRGYPIISIDHLPISSA
                               .: : ::. . ...    .::      .::
gi|218 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
          70        80        90       100       110       120

                     40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
         :  ::      :.::.  :
gi|218 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
           130       140       150


>>gi|60418848|gb|AAX19851.1| thaumatin-like protein prec  (246 aa)
 initn:  46 init1:  46 opt:  71  Z-score: 107.5  bits: 25.7 E(): 0.52
Smith-Waterman score: 71;  31.250% identity (52.083% similar) in 48 aa overlap (1-46:159-201)

                                       10        20        30
RF_1_+                         HVRMECPTVVYIRSRGYPIISIDHLPISSA
                               .: : ::. . ...    .::      .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
         130       140       150       160       170       180
```

```
                    40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
          :   ::      :.::.   :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
           190       200       210       220       230       240
```

>>gi|60418842|gb|AAX19848.1| thaumatin-like protein prec (246 aa)
 initn: 46 init1: 46 opt: 71 Z-score: 107.5 bits: 25.7 E(): 0.52
Smith-Waterman score: 71; 31.250% identity (52.083% similar) in 48 aa overlap (1-46:159-201)

```
                                        10        20        30
RF_1_+                           HVRMECPTVVYIRSRGYPIISIDHLPISSA
                                 .: : ::.  . ...    .::      .::
gi|604 DVSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSA
           130       140       150       160       170       180

                    40        50        60        70        80
RF_1_+ FLHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
          :   ::      :.::.   :
gi|604 CLAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVIT
           190       200       210       220       230       240
```

>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn: 38 init1: 38 opt: 65 Z-score: 98.9 bits: 24.0 E(): 1.6
Smith-Waterman score: 65; 33.333% identity (52.381% similar) in 63 aa overlap (12-68:114-176)

```
                                        10        20        30
RF_1_+                           HVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                                 .:. : .  ...  .: :.    :::    : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
             90       100       110       120       130       140

              40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQS
          .  :  : .::..:  :  :     :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             150       160       170       180       190       200

gi|162 ENSEKTTISLW
           210
```

>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly (473 aa)
 initn: 38 init1: 38 opt: 68 Z-score: 98.1 bits: 24.9 E(): 1.7
Smith-Waterman score: 68; 30.000% identity (55.000% similar) in 60 aa overlap (17-75:301-356)

```
                                        10        20        30        40
RF_1_+                           HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .:   .  .   . : .:..    ::..:
```

```
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
               280       290       300       310       320


          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
        :.:   :    :::.   .  .:::.   :  .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         330       340       350       360       370       380


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score:  96.1  bits: 22.7 E():  2.2
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (12-68:29-91)


                          10        20        30
RF_1_+              HVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                      .: . :. .... .: :.    :::   :  :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
               10        20        30        40        50        60


     40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQS
        .   : :  .::..:  :   :       :   ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
               70        80        90       100       110       120


gi|159 SEKTTMPLW



>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score:  93.4  bits: 23.7 E():  3.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (17-75:216-271)


                          10        20        30        40
RF_1_+              HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                     ::  .:    .   .   .  :   .:..      ::...:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
        :.:   :    :::.   .  .::.   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
                250       260       270       280       290       300


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score:  93.4  bits: 23.7 E():  3.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (17-75:216-271)


                          10        20        30        40
```

```
RF_1_+                    HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                  :: .:   . .   . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


             50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
         :.:  :   :::.  . .:..  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          250       260       270       280       290       300


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 93.4  bits: 23.7 E():  3.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (17-75:216-271)

                              10        20        30        40
RF_1_+                    HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                  :: .:   . .   . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


             50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
         :.:  :   :::.  . .:..  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
             250       260       270       280       290       300


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 92.0  bits: 23.8 E():  3.8
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (17-75:292-347)

                              10        20        30        40
RF_1_+                    HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                  :: .:    . .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
             270       280       290       300       310


             50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
         :.:  :   :::.  . .:..  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          320       330       340       350       360       370


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 91.1  bits: 22.8 E():  4.2
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (27-50:13-35)

               10        20        30        40        50        60
RF_1_+ HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPN
```

```
                            ..  .::  ::   :.::..  :....
gi|898                   MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLD
                            10        20        30        40


              70        80
RF_1_+ LQKTAYPSLVRPARTIELQS


gi|898 AKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKP
          50        60        70        80        90        100


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62 Z-score: 89.9  bits: 23.1 E():    5
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (21-72:13-63)


              10        20        30        40        50        60
RF_1_+ HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPN
                    :.  :: . :  : : .  . :. ::.. :   ::..
gi|212                   MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKA
                            10        20        30        40        50


              70        80
RF_1_+ LQKTAYPSLVRPARTIELQS
        . :.  .: ::
gi|212 GMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLS
              60        70        80        90        100       110


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60 Z-score: 89.3  bits: 22.5 E(): 5.3
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (27-50:19-41)


              10        20        30        40        50        60
RF_1_+ HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPN
                            .. :::  :   :.::.. :....
gi|115                   MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLD
                            10        20        30        40        50


              70        80
RF_1_+ LQKTAYPSLVRPARTIELQS


gi|115 AKATWYGKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKP
                   60        70        80        90        100       110


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58 Z-score: 89.3  bits: 21.9 E(): 5.3
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (12-68:72-134)


                                10        20        30
RF_1_+                   HVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                .:. :. . ... .: :.   :::   :
```

```
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                50        60        70        80        90       100


         40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQS
        :  .   : :  .::..: :   :       :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
              110       120       130       140       150       160


gi|159 ENSEKTTMPLW
              170


>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta  (224 aa)
 initn:  42 init1:  42 opt:  59 Z-score: 89.0  bits: 22.2 E():  5.5
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (7-80:130-204)


                                     10        20        30
RF_1_+                      HVRMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                            :.  . .:..  . ...   :::.   ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
       100       110       120       130       140       150


              40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
        :.       :  . : : :.   :::  ..  .   ..   : :  .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
       160       170       180       190       200       210


gi|162 RGPFPIIV
          220


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]       (224 aa)
 initn:  42 init1:  42 opt:  59 Z-score: 89.0  bits: 22.2 E():  5.5
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (7-80:130-204)


                                     10        20        30
RF_1_+                      HVRMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                            :.  . .:..  . ...   :::.   ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
       100       110       120       130       140       150


              40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
        :.       :  . : : :.   :::  ..  .   ..   : :  .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
       160       170       180       190       200       210


gi|162 RGPFPIIV
          220
```

```
>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]   (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 89.0  bits: 22.2 E():  5.5
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (7-80:130-204)


                                   10        20        30
RF_1_+                     HVRMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                                 :.   . .:..  . ...  :::.    ....
gi|459 PPFLQPEVMGVSKVKEAMAPKQKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
           100       110       120       130       140       150


                  40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS
          :.      : . : : :.   :::  ..  .   ..  : : . .:.
gi|459 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
          160       170       180       190       200       210


gi|459 RGPFPIIV
          220


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 88.1  bits: 21.9 E():  6.3
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (12-68:105-167)


                                   10        20        30
RF_1_+                     HVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                 .: . : . ... .: :.   :::   :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
             80        90       100       110       120       130


           40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQS
         :.   : :  .:..: :  :     :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             140       150       160       170       180       190


gi|159 ENSEKTTMPLW
          200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.8  bits: 21.9 E():  6.5
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (12-68:114-176)


                                   10        20        30
RF_1_+                     HVRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                 .: . : . ... .: :.   :::   :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
             90       100       110       120       130       140
```

```
              40          50          60          70          80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQS
       :  .   :  :  .::..: :  :       :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
                 150         160         170         180         190         200

gi|162 ENSEKTTMPLW
              210


>>gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea mays]  (252 aa)
 initn:  52 init1:  52 opt:  58  Z-score: 86.6  bits: 21.9 E():  7.5
Smith-Waterman score: 58;  40.000% identity (75.000% similar) in 20 aa overlap (31-50:6-24)


                 10          20          30          40          50          60
RF_1_+ HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPN
                                      :: ::   :.::.. :....
gi|898                         ARALVFLVSGAW-CGPPKVPPGKNITATYGKDWLD
                                        10          20          30


                 70          80
RF_1_+ LQKTAYPSLVRPARTIELQS

gi|898 AKATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKP
              40          50          60          70          80          90


>>gi|83305621|sp|Q8NKF4.2|RL3_ASPFU RecName: Full=60S ri  (392 aa)
 initn:  45 init1:  45 opt:  59  Z-score: 85.1  bits: 22.3 E():  9.2
Smith-Waterman score: 59;  30.769% identity (53.846% similar) in 52 aa overlap (21-72:13-63)


                 10          20          30          40          50          60
RF_1_+ HVRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPN
                      :.   :: . :  : :  .   . :. ::..  :    ::..
gi|833         MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKDDPKKPVHLTASMGYKA
                      10          20          30          40          50


                 70          80
RF_1_+ LQKTAYPSLVRPARTIELQS
          .  :.   .:  ::
gi|833 GMTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLS
              60          70          80          90          100         110


>>gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=Proba  (139 aa)
 initn:  52 init1:  52 opt:  54  Z-score: 84.4  bits: 20.6 E():   10
Smith-Waterman score: 54;  39.286% identity (60.714% similar) in 28 aa overlap (2-28:87-113)


                                           10          20          30
RF_1_+                         HVRMECPTV-VYIRSRGYPIISIDHLPISSA
                                : :.: :: :  :.    :. :. : :..
gi|391 MKTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPV-SLRHGPVTGP
```

```
          60         70        80        90        100       110

                 40        50        60        70        80
RF_1_+ FLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQS


gi|391 SDPAHKARLERPQIRVPPPAPEKA
          120       130




80 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:39 2010 done: Fri Apr 30 23:42:40 2010
 Total Scan time:  0.090 Total Display time:  0.010

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_+1: 7  - 86 80 aa - 80 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

       opt      E()
< 20     2     0:=
  22     0     0:               one = represents 3 library sequences
  24     2     0:=
  26     0     0:
  28    22     0:=======
  30    13     2:*====
  32    34     8:==*=========
  34    42    21:======*=======
  36    62    44:==============*======
  38    64    72:===================== *
  40   112   101:==============================*====
  42   111   123:===================================    *
  44   135   136:============================================*
  46   133   138:============================================*
  48    95   132:=============================            *
  50    97   121:=============================          *
  52    82   106:========================          *
  54    84    91:=========================   *
  56   117    76:=========================*============
```

```
 58     73       62:==================*====
 60     39       50:============   *
 62     24       40:=======       *
 64     37       32:=========*==
 66     14       25:====   *
 68     12       20:====  *
 70     15       16:=====*
 72      9       12:===*
 74      7       10:===*
 76      3        7:= *
 78      1        6:=*
 80      2        4:=*
 82      5        3:*=
 84      2        3:*
 86      1        2:*
 88      8        2:*==          inset = represents 1 library sequences
 90      2        1:*
 92      3        1:*        :*==
 94      0        1:*        :*
 96      1        1:*        :*
 98      2        0:=        *==
100      0        0:         *
102      0        0:         *
104      1        0:=        *=
106      1        0:=        *=
108      1        0:=        *=
110      1        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))= 5.1045+/-0.00375; mu= 7.1370+/- 0.192
 mean_var=39.0310+/-10.482, 0's: 2 Z-trim: 4  B-trim: 157 in 2/41
 Lambda= 0.205291
 Kolmogorov-Smirnov  statistic: 0.0702 (N=29) at   40

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:  0.090
The best scores are:                            opt bits E(1471)
gi|218059718|emb|CAT99612.1| thaumatin-like protei ( 158)   70 25.4    0.41
gi|218059715|emb|CAT99611.1| thaumatin-like protei ( 158)   70 25.4    0.41
gi|60418848|gb|AAX19851.1| thaumatin-like protein  ( 246)   70 25.4    0.66
gi|60418842|gb|AAX19848.1| thaumatin-like protein  ( 246)   70 25.4    0.66
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 23.9    1.6
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 24.8     2
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.8    2.1
```

```
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)    64 23.6     3.5
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)    64 23.6     3.5
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)    64 23.6     3.5
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)    64 23.6     4.3
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)    61 22.7     4.5
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)    58 21.9     5.3
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)    62 23.0     5.5
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)    60 22.4     5.6
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)    59 22.1     5.7
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)    59 22.1     5.7
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)    59 22.1     5.7
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)    58 21.9     6.4
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)    58 21.8     6.7
gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea m ( 252)    58 21.8     7.9
gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=P ( 139)    54 20.7     9.6
gi|741844|prf||2008179A major allergen Par j I     ( 143)    54 20.7     9.9


>>gi|218059718|emb|CAT99612.1| thaumatin-like protein [M  (158 aa)
 initn:  45 init1:  45 opt:  70  Z-score: 109.3  bits: 25.4 E(): 0.41
Smith-Waterman score: 70;  31.915% identity (51.064% similar) in 47 aa overlap (1-45:100-141)

                                      10        20        30
RF_1_+                           VRMECPTVVYIRSRGYPIISIDHLPISSAF
                                  : : ::. . ...     .::       .::
gi|218 VSLVDGFNLPMYVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSAC
           70        80        90       100       110       120

                 40        50        60        70        80
RF_1_+ LHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
          :   ::       :.::.   :
gi|218 LAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
            130       140       150


>>gi|218059715|emb|CAT99611.1| thaumatin-like protein [M  (158 aa)
 initn:  45 init1:  45 opt:  70  Z-score: 109.3  bits: 25.4 E(): 0.41
Smith-Waterman score: 70;  31.915% identity (51.064% similar) in 47 aa overlap (1-45:100-141)

                                      10        20        30
RF_1_+                           VRMECPTVVYIRSRGYPIISIDHLPISSAF
                                  : : ::. . ...     .::       .::
gi|218 VSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSAC
           70        80        90       100       110       120

                 40        50        60        70        80
RF_1_+ LHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
          :   ::       :.::.   :
gi|218 LAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCP
            130       140       150
```

```
>>gi|60418848|gb|AAX19851.1| thaumatin-like protein prec  (246 aa)
 initn:  45 init1:  45 opt:  70  Z-score: 105.6  bits: 25.4 E(): 0.66
Smith-Waterman score: 70;  31.915% identity (51.064% similar) in 47 aa overlap (1-45:160-201)

                                        10        20        30
RF_1_+                          VRMECPTVVYIRSRGYPIISIDHLPISSAF
                                : : ::. . ...    .::      .::
gi|604 VSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSAC
          130       140       150       160       170       180

                     40        50        60        70        80
RF_1_+ LHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
          :  ::      :.::.  :
gi|604 LAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVITF
           190       200       210       220       230       240


>>gi|60418842|gb|AAX19848.1| thaumatin-like protein prec  (246 aa)
 initn:  45 init1:  45 opt:  70  Z-score: 105.6  bits: 25.4 E(): 0.66
Smith-Waterman score: 70;  31.915% identity (51.064% similar) in 47 aa overlap (1-45:160-201)

                                        10        20        30
RF_1_+                          VRMECPTVVYIRSRGYPIISIDHLPISSAF
                                : : ::. . ...    .::      .::
gi|604 VSLVDGFNLPMSVAPQGGTGECKPSSCPANVNMACPAQLQVKAADGSVISC-----KSAC
          130       140       150       160       170       180

                     40        50        60        70        80
RF_1_+ LHSGD--LACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
          :  ::      :.::.  :
gi|604 LAFGDSKYCCTPPNDTPETCPPTEYSEIFEKQCPQAYSYAYDDKNSTFTCSGGPDYVITF
           190       200       210       220       230       240


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score: 98.8  bits: 23.9 E():  1.6
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (11-67:114-176)

                                  10        20        30
RF_1_+                    VRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                          .:. :. ...  .: :.    :::    : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
               90       100       110       120       130       140

              40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQST
         .  : :  .::..: :  :      :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
                150       160       170       180       190       200

gi|162 ENSEKTTISLW
```

>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly (473 aa)
 initn: 38 init1: 38 opt: 68 Z-score: 97.1 bits: 24.8 E():   2
Smith-Waterman score: 68; 30.000% identity (55.000% similar) in 60 aa overlap (16-74:301-356)

```
                              10        20        30        40
RF_1_+             VRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                              ::  .:    . .   . : .:..     ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
                280       290       300       310       320

              50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
         :.:  :   ::: .  .  .::.  :  .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         330       340       350       360       370       380
```

>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru (129 aa)
 initn: 37 init1: 37 opt: 61 Z-score: 96.5 bits: 22.8 E(): 2.1
Smith-Waterman score: 61; 33.333% identity (52.381% similar) in 63 aa overlap (11-67:29-91)

```
                              10        20        30
RF_1_+             VRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                              .:  . :  ... .: :.    :::   :  :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
                  10        20        30        40        50        60

          40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQST
         .  : :  .::..: :  :      :  ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
                    70        80        90       100       110       120

gi|159 SEKTTMPLW
```

>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly (373 aa)
 initn: 36 init1: 36 opt: 64 Z-score: 92.6 bits: 23.6 E(): 3.5
Smith-Waterman score: 64; 28.333% identity (55.000% similar) in 60 aa overlap (16-74:216-271)

```
                              10        20        30        40
RF_1_+             VRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                              ::  .:    . .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
                190       200       210       220       230       240

              50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
```

```
            :.:   :    :::.  .  . .::.  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 92.6  bits: 23.6 E():  3.5
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (16-74:216-271)

                          10        20        30        40
RF_1_+              VRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                              :: .:   .  .  . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240


          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
           :.:   :    :::.  .  . .::.  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 92.6  bits: 23.6 E():  3.5
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (16-74:216-271)

                          10        20        30        40
RF_1_+              VRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                              :: .:   .  .  . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240


          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
           :.:   :    :::.  .  . .::.  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 91.0  bits: 23.6 E():  4.3
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (16-74:292-347)

                          10        20        30        40
RF_1_+              VRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                              :: .:   .  .  . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           270       280       290       300       310


          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
           :.:   :    :::.  .  . .::.  .   .:
```

```
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
           320        330       340       350       360       370


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 90.7  bits: 22.7  E():  4.5
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (26-49:13-35)

                   10        20        30        40        50        60
RF_1_+ VRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNL
                               .. .:: ::   :.::.. :....
       gi|898           MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDA
                                   10        20        30        40


               70        80
RF_1_+ QKTAYPSLVRPARTIELQST

gi|898 KATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPV
           50        60        70        80        90       100


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 89.4  bits: 21.9  E():  5.3
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (11-67:72-134)

                              10        20        30
RF_1_+                 VRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                            .:. : . ... .: :.    :::    :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                 50        60        70        80        90       100

           40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQST
        :.   : : .::..: :  :     :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
               110       120       130       140       150       160

gi|159 ENSEKTTMPLW
                 170


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 89.0  bits: 23.0  E():  5.5
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (20-71:13-63)

                   10        20        30        40        50        60
RF_1_+ VRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNL
                        :.  :: . :  : :.   . :. ::. :   ::..
gi|212          MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAG
                         10        20        30        40        50


               70        80
```

```
RF_1_+ QKTAYPSLVRPARTIELQST
          .  :.   .: ::
gi|212 MTTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSD
                60        70        80        90       100       110


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp   (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 88.9  bits: 22.4 E():  5.6
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (26-49:19-41)

                10        20        30        40        50        60
RF_1_+ VRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNL
                           ..  :::   :   :.::.. :....
gi|115         MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDA
                       10        20        30        40        50


                70        80
RF_1_+ QKTAYPSLVRPARTIELQST

gi|115 KATWYGKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPA
                60        70        80        90       100       110


>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta   (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 88.8  bits: 22.1 E():  5.7
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (6-79:130-204)

                                  10        20        30
RF_1_+                   VRMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                                  :.  . .:.. . ...  :::.  ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
           100       110       120       130       140       150


                40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
          :.        :  . : :.     :::  ..  .   ..  : : . .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
         160       170       180       190       200       210


gi|162 RGPFPIIV
          220


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]         (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 88.8  bits: 22.1 E():  5.7
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (6-79:130-204)

                                  10        20        30
RF_1_+                   VRMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                                  :.  . .:.. . ...  :::.  ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
           100       110       120       130       140       150
```

```
                   40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
         :.         :   . : : :.    :::  ..   .    ..   : : . .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
       160       170       180       190       200       210


gi|162 RGPFPIIV
          220


>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]    (224 aa)
 initn:  42 init1:  42 opt:  59 Z-score: 88.8  bits: 22.1 E():  5.7
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (6-79:130-204)


                                10        20        30
RF_1_+                 VRMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                           :.   . .:..   .  ...   :::.    ....
gi|459 PPFLQPEVMGVSKVKEAMAPKQKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
         100       110       120       130       140       150


                   40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST
         :.         :   . : : :.    :::  ..   .    ..   : : . .:.
gi|459 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
       160       170       180       190       200       210


gi|459 RGPFPIIV
          220


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58 Z-score: 87.9  bits: 21.9 E():  6.4
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (11-67:105-167)


                                10        20        30
RF_1_+                 VRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                            .:  . : . ... .: :.   :::    :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                80        90       100       110       120       130


                   40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQST
         :.   : : .::..: :   :      : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
                140       150       160       170       180       190


gi|159 ENSEKTTMPLW
          200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
```

```
 initn:  37 init1:  37 opt:  58  Z-score: 87.6  bits: 21.8 E():  6.7
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (11-67:114-176)

                                   10        20        30
RF_1_+                       VRMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                   .: . : . ... .: :.   ::: :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
               90       100       110       120       130       140

          40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQST
          : .   : :  .::..: :   :       : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
              150       160       170       180       190       200

gi|162 ENSEKTTMPLW
               210


>>gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea mays]  (252 aa)
 initn:  52 init1:  52 opt:  58  Z-score: 86.2  bits: 21.8 E():  7.9
Smith-Waterman score: 58;  40.000% identity (75.000% similar) in 20 aa overlap (30-49:6-24)

               10        20        30        40        50        60
RF_1_+ VRMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNL
                                   :: ::   :..::.. :....
gi|898                        ARALVFLVSGAW-CGPPKVPPGKNITATYGKDWLDA
                                      10        20        30

               70        80
RF_1_+ QKTAYPSLVRPARTIELQST

gi|898 KATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPV
                40        50        60        70        80        90


>>gi|3915783|sp|P43217.3|NLT11_PARJU RecName: Full=Proba  (139 aa)
 initn:  52 init1:  52 opt:  54  Z-score: 84.7  bits: 20.7 E():  9.6
Smith-Waterman score: 54;  39.286% identity (60.714% similar) in 28 aa overlap (1-27:87-113)

                                   10        20
RF_1_+                       VRMECPTV-VYIRSRGYPIISIDHLPISSA
                                   : :.: :: :  :.   :. :.  : :..
gi|391 MKTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPV-SLRHGPVTGP
               60        70        80        90       100       110

          30        40        50        60        70        80
RF_1_+ FLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST

gi|391 SDPAHKARLERPQIRVPPPAPEKA
              120       130
```

```
>>gi|741844|prf||2008179A major allergen Par j I          (143 aa)
 initn:  52 init1:  52 opt:  54  Z-score: 84.5  bits: 20.7 E():  9.9
Smith-Waterman score: 54;  39.286% identity (60.714% similar) in 28 aa overlap (1-27:81-107)


                                    10        20
RF_1_+                      VRMECPTV-VYIRSRGYPIISIDHLPISSA
                            : :.: :: :  :.    :. :. : :..
gi|741 MKTYSDIDGKLVSEVPKHCGIVDSKLPPIDVNMDCKTVGVVPRQPQLPV-SLRHGPVTGP
                 60        70        80        90       100


    30        40        50        60        70        80
RF_1_+ FLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQST


gi|741 SRSRPPTKHGWRDPRLEFRPPHRKKPNPAFSTLG
        110       120       130       140




80 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:40 2010 done: Fri Apr 30 23:42:40 2010
 Total Scan time:  0.090 Total Display time:  0.010

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_+1: 8 - 87 80 aa - 80 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library


        opt      E()
< 20     2    0:=
  22     0    0:               one = represents 3 library sequences
  24     2    0:=
  26     0    0:
  28    13    0:====
  30    22    2:*=======
  32    31    8:==*========
  34    38   21:======*======
  36    69   44:=============*========
  38    69   72:====================*
```

```
 40    97   101:================================*
 42   115   123:==================================== *
 44   136   136:==============================================*
 46   132   138:============================================ *
 48    91   132:============================                 *
 50    97   121:=============================            *
 52    89   106:==========================         *
 54    87    91:========================== *
 56   112    76:=======================*===========
 58    74    62:==================*====
 60    45    50:============== *
 62    25    40:========     *
 64    39    32:=========*==
 66    13    25:====    *
 68    13    20:==== *
 70    17    16:====*
 72     7    12:===*
 74     7    10:===*
 76     2     7:= *
 78     3     6:=*
 80     2     4:=*
 82     5     3:*=
 84     0     3:*
 86     2     2:*
 88     7     2:*==          inset = represents 1 library sequences
 90     2     1:*
 92     3     1:*          :*==
 94     0     1:*          :*
 96     2     1:*          :*=
 98     1     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))= 5.2496+/-0.00375; mu= 6.3042+/- 0.192
mean_var=40.1483+/-10.556, 0's: 2 Z-trim: 2  B-trim: 125 in 1/42
 Lambda= 0.202414
 Kolmogorov-Smirnov  statistic: 0.0680 (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:  0.090
The best scores are:                                    opt bits E(1471)
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 23.8    1.7
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 24.6    2.2
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.7    2.2
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.5    3.8
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.5    3.8
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.5    3.8
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.5    4.7
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.6    4.8
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.8    5.6
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 22.9     6
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.3     6
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   59 22.1    6.1
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   59 22.1    6.1
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   59 22.1    6.1
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.8    6.7
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.8    7.1
gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea m ( 252)   58 21.8    8.4


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score: 98.2  bits: 23.8 E():  1.7
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (10-66:114-176)

                          10        20        30
RF_1_+                    RMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                            .: . : ... .: :.    :::    : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
             90       100       110       120       130       140

               40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTA
          . : : .:..: :  :    :  :::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             150       160       170       180       190       200

gi|162 ENSEKTTISLW
             210


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly  (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score: 96.3  bits: 24.6 E():  2.2
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (15-73:301-356)

                            10        20        30        40
RF_1_+                      RMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                            :: .:  . .   . : .:.    ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            280       290       300       310       320
```

```
           50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTA
        :.:   :    :::.  .  . .::.   :  .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         330       340       350       360       370       380


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score:  96.1  bits: 22.7 E():  2.2
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (10-66:29-91)

                              10        20        30
RF_1_+              RMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                     .:  :. :. ... .: :.    :::    :  :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
          10        20        30        40        50        60

          40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTA
        .   : :  .::..: :   :      :  ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
                 70        80        90       100       110       120

gi|159 SEKTTMPLW


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score:  92.0  bits: 23.5 E():  3.8
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (15-73:216-271)

                              10        20        30        40
RF_1_+              RMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                          ::  .:   .   .  . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240

          50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTA
        :.:   :    :::.  .  . .::.   .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         250       260       270       280       290       300


>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score:  92.0  bits: 23.5 E():  3.8
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (15-73:216-271)

                              10        20        30        40
RF_1_+              RMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                          ::  .:   .   .  . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
```

```
                190       200       210       220       230       240


                50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTA
          :.:   :    :::.   .   . .:.   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
               250       260       270       280       290       300


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score:  92.0  bits: 23.5 E():  3.8
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (15-73:216-271)

                            10        20        30        40
RF_1_+            RMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                           ::  .:    .   .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240


                50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTA
          :.:   :    :::.   .   . .:.   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
               250       260       270       280       290       300


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score:  90.3  bits: 23.5 E():  4.7
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (15-73:292-347)

                            10        20        30        40
RF_1_+            RMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                           ::  .:    .   .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           270       280       290       300       310


                50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTA
          :.:   :    :::.   .   . .:.   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          320       330       340       350       360       370


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score:  90.2  bits: 22.6 E():  4.8
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (25-48:13-35)

                  10        20        30        40        50        60
RF_1_+ RMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQ
                            .. .:: ::   :.:.. :....
gi|898         MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDAK
                       10        20        30        40
```

```
                  70        80
RF_1_+ KTAYPSLVRPARTIELQSTA


gi|898 ATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVE
           50        60        70        80        90        100


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 88.9  bits: 21.8 E():  5.6
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (10-66:72-134)


                                  10        20        30
RF_1_+                 RMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                            .: . : . ... .: :.   :::   :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
           50        60        70        80        90        100


          40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTA
       :.   : :  .::..: :  :      :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
          110       120       130       140       150       160


gi|159 ENSEKTTMPLW
              170


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 88.4  bits: 22.9 E():   6
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (19-70:13-63)


              10        20        30        40        50        60
RF_1_+ RMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQ
                     :.   :: . :   : : .    . :. :: . :   : :::.     .
gi|212       MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAGM
               10        20        30        40        50


              70        80
RF_1_+ KTAYPSLVRPARTIELQSTA
        :.   .: ::
gi|212 TTVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSDE
            60        70        80        90        100       110


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 88.4  bits: 22.3 E():    6
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (25-48:19-41)


              10        20        30        40        50        60
RF_1_+ RMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQ
                     .. :::  :   :.:.:.. :.....
```

```
gi|115        MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDAK
                    10        20        30        40        50


                  70        80
RF_1_+ KTAYPSLVRPARTIELQSTA


gi|115 ATWYGKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAE
              60        70        80        90        100       110


>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta  (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 88.3  bits: 22.1 E():  6.1
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (5-78:130-204)


                                10        20        30
RF_1_+                   RMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                         :.   . .:..  . ...   :::.   ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
      100       110       120       130       140       150


                  40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTA
        :.       :  . : : :.   :::  ..  .   ..   : : . .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
      160       170       180       190       200       210


gi|162 RGPFPIIV
          220


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]        (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 88.3  bits: 22.1 E():  6.1
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (5-78:130-204)


                                10        20        30
RF_1_+                   RMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                         :.   . .:..  . ...   :::.   ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
      100       110       120       130       140       150


                  40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTA
        :.       :  . : : :.   :::  ..  .   ..   : : . .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
      160       170       180       190       200       210


gi|162 RGPFPIIV
          220


>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]     (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 88.3  bits: 22.1 E():  6.1
```

```
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (5-78:130-204)

                                       10        20        30
RF_1_+                         RMECPTVVYIRSRGYPIISID--HLPIS--SAFL
                                    :.   . .:..   . ...   :::.    ....
gi|459 PPFLQPEVMGVSKVKEAMAPKQKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
          100       110       120       130       140       150

                         40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTA
          :.        :   . : : :.    :::   ..   .    ..  : : . .:.
gi|459 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
          160       170        180       190       200       210

gi|459 RGPFPIIV
          220


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.5  bits: 21.8 E():  6.7
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (10-66:105-167)

                                       10        20        30
RF_1_+                         RMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                    .:. : . ... .: :.    :::    :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
              80        90       100       110       120       130

                         40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTA
          :.   : :  .::..: :  :       :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
              140       150       160       170       180       190

gi|159 ENSEKTTMPLW
            200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.1  bits: 21.8 E():  7.1
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (10-66:114-176)

                                       10        20        30
RF_1_+                         RMECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                    .:. : . ... .: :.    :::    :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                90       100       110       120       130       140

                         40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTA
          :.   : :  .::..: :  :       :  ::::
```

```
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
                150       160       170       180       190       200


gi|162 ENSEKTTMPLW
                210


>>gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea mays]  (252 aa)
 initn:  52 init1:  52 opt:  58  Z-score: 85.8  bits: 21.8 E():  8.4
Smith-Waterman score: 58;  40.000% identity (75.000% similar) in 20 aa overlap (29-48:6-24)

                      10        20        30        40        50        60
RF_1_+ RMECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQ
                                    ::  ::    :.::.. :....
gi|898                     ARALVFLVSGAW-CGPPKVPPGKNITATYGKDWLDAK
                                       10        20        30


               70        80
RF_1_+ KTAYPSLVRPARTIELQSTA


gi|898 ATWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVE
            40        50        60        70        80        90




80 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:41 2010 done: Fri Apr 30 23:42:41 2010
 Total Scan time:  0.090 Total Display time:  0.000


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_+1: 9  - 88 80 aa - 80 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library


      opt     E()
< 20     2    0:=
  22     0    0:             one = represents 3 library sequences
  24     2    0:=
  26     0    0:
  28    22    0:=======
```

```
 30    11     2:*===
 32    31     8:==*========
 34    30    21:======*===
 36    65    44:=============*=======
 38    72    72:====================*
 40    94   101:============================== *
 42   118   123:=====================================*
 44   138   136:=========================================*
 46   131   138:======================================= *
 48    88   132:===========================             *
 50    88   121:===========================         *
 52   103   106:==================================*
 54    84    91:=========================   *
 56   120    76:=====================*=============
 58    70    62:=================*===
 60    48    50:==============*
 62    32    40:==========  *
 64    31    32:=========*
 66    12    25:====     *
 68    17    20:=====*
 70    16    16:=====*
 72     8    12:===*
 74     8    10:===*
 76     1     7:= *
 78     3     6:=*
 80     2     4:=*
 82     5     3:*=
 84     2     3:*
 86     2     2:*
 88     7     2:*==          inset = represents 1 library sequences
 90     2     1:*
 92     3     1:*         :*==
 94     0     1:*         :*
 96     2     1:*         :*=
 98     1     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))= 5.4665+/-0.00371; mu= 5.2081+/- 0.190
mean_var=40.0743+/-10.737, 0's: 2 Z-trim: 2  B-trim: 125 in 1/42
```

```
 Lambda= 0.202601
 Kolmogorov-Smirnov  statistic: 0.0585 (N=29) 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:  0.080
The best scores are:                              opt bits E(1471)
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 23.8    1.7
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.7    2.2
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 24.6    2.3
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.4    3.9
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.4    3.9
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.4    3.9
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.4    4.9
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.6    4.9
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.8    5.6
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   59 22.0    6.1
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   59 22.0    6.1
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   59 22.0    6.1
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.3    6.1
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 22.8    6.2
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.8    6.8
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.8    7.1
gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea m ( 252)   58 21.7    8.5
gi|1398916|dbj|BAA07712.1| allergenic protein [Ory ( 157)   55 20.9    9.3
gi|23616954|dbj|BAC20657.1| allergen RA16 [Oryza s ( 157)   55 20.9    9.3


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score: 98.1  bits: 23.8 E():  1.7
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (9-65:114-176)


                              10        20        30
RF_1_+                MECPTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                               .:  :  . ... .: :.   :::   :  :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
           90       100       110       120       130       140

               40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAP
          .  :  : .::..: :  :     :   ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
           150       160       170       180       190       200


gi|162 ENSEKTTISLW
            210


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score: 96.2  bits: 22.7 E():  2.2
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (9-65:29-91)
```

```
                                 10        20        30
RF_1_+                     MECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                               .: . : . ... .: :. :  ::: : :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
                10        20        30        40        50        60

               40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAP
          .  : :  .:.. : : :     : ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
                70        80        90       100       110       120

gi|159 SEKTTMPLW


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly  (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score:  96.0  bits: 24.6 E():  2.3
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (14-72:301-356)

                                 10        20        30        40
RF_1_+                     MECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .: . . . : .:..    ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
                280       290       300       310       320

               50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAP
          :.: :  :::. . . .:. : .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          330       340       350       360       370       380


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score:  91.7  bits: 23.4 E():  3.9
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (14-72:216-271)

                                 10        20        30        40
RF_1_+                     MECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .: . . . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
               190       200       210       220       230       240

               50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAP
          :.: :  :::. . . .::. . .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
               250       260       270       280       290       300


>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
```

```
  initn:  36 init1:  36 opt:  64  Z-score: 91.7  bits: 23.4 E():  3.9
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (14-72:216-271)

                              10        20        30        40
RF_1_+               MECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .:   . .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
         190       200       210       220       230       240

              50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAP
          :.: :   ::: .  . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         250       260       270       280       290       300

>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
  initn:  36 init1:  36 opt:  64  Z-score: 91.7  bits: 23.4 E():  3.9
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (14-72:216-271)

                              10        20        30        40
RF_1_+               MECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .:   . .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
         190       200       210       220       230       240

              50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAP
          :.: :   ::: .  . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         250       260       270       280       290       300

>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
  initn:  36 init1:  36 opt:  64  Z-score: 90.0  bits: 23.4 E():  4.9
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (14-72:292-347)

                              10        20        30        40
RF_1_+               MECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .:   . .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
         270       280       290       300       310

              50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAP
          :.: :   ::: .  . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
        320       330       340       350       360       370

>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
  initn:  52 init1:  52 opt:  61  Z-score: 90.0  bits: 22.6 E():  4.9
```

```
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (24-47:13-35)

                    10        20        30        40        50        60
RF_1_+ MECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQK
                            .. .:: ::   :.::.. :....
gi|898           MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKA
                        10        20        30        40

              70        80
RF_1_+ TAYPSLVRPARTIELQSTAP

gi|898 TWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVEC
          50        60        70        80        90       100


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 88.9  bits: 21.8 E():  5.6
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (9-65:72-134)

                                  10        20        30
RF_1_+                 MECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                              .:. : . ... .: :.   ::: :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                  50        60        70        80        90       100

              40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAP
         :. . : : .::..: :  :     : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
              110       120       130       140       150       160

gi|159 ENSEKTTMPLW
              170


>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta  (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 88.2  bits: 22.0 E():  6.1
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (4-77:130-204)

                                    10        20
RF_1_+                     MECPTVVYIRSRGYPIISID--HLPIS--SAFL
                                  :. . .:.. . ... :::.  ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
          100       110       120       130       140       150

       30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAP
         :.       : . : :.  :::  ..  .   ..  : : .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
          160       170       180       190       200       210
```

```
gi|162 RGPFPIIV
        220


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]      (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 88.2  bits: 22.0 E():  6.1
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (4-77:130-204)


                                        10        20
RF_1_+                        MECPTVVYIRSRGYPIISID--HLPIS--SAFL
                              :.  . .:..  . ...  :::.    ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
         100       110       120       130       140       150


        30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAP
        :.       :  . : : :.    :::  ..  .   ..  : : . .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
         160       170       180       190       200       210


gi|162 RGPFPIIV
        220


>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]     (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 88.2  bits: 22.0 E():  6.1
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (4-77:130-204)


                                        10        20
RF_1_+                        MECPTVVYIRSRGYPIISID--HLPIS--SAFL
                              :.  . .:..  . ...  :::.    ....
gi|459 PPFLQPEVMGVSKVKEAMAPKQKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
         100       110       120       130       140       150


        30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAP
        :.       :  . : : :.    :::  ..  .   ..  : : . .:.
gi|459 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
         160       170       180       190       200       210


gi|459 RGPFPIIV
        220


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 88.2  bits: 22.3 E():  6.1
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (24-47:19-41)


              10        20        30        40        50        60
RF_1_+ MECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQK
                            .. :::  :   :.:.. :....
gi|115       MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDAKA
```

```
                    10        20        30        40        50

                  70        80
RF_1_+ TAYPSLVRPARTIELQSTAP


gi|115 TWYGKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAEC
           60        70        80        90       100       110


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 88.1  bits: 22.8 E():  6.2
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (18-69:13-63)


               10        20        30        40        50        60
RF_1_+ MECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQK
                    :.   ::  . :  : ::.   . :.  ::..  :    ::..     .
gi|212        MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAGMT
                      10        20        30        40        50


               70        80
RF_1_+ TAYPSLVRPARTIELQSTAP
        :.    .:  ::
gi|212 TVVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSDEV
                60        70        80        90       100       110


>>gi|159793197|gb|ABW98943.1|  alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.4  bits: 21.8 E():  6.8
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (9-65:105-167)


                              10        20        30
RF_1_+                 MECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                          .: . : . ... .: :.    :::    :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
           80        90       100       110       120       130


           40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAP
        :.   : :  .:..: :  :       :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
           140       150       160       170       180       190


gi|159 ENSEKTTMPLW
           200


>>gi|162792|gb|AAA30428.1|  alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.1  bits: 21.8 E():  7.1
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (9-65:114-176)


                              10        20        30
RF_1_+                 MECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
```

```
                                 .: :. .  ... .: :.    :::    :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
               90        100       110       120       130       140


              40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAP
       :.  :: .::..: :  :      : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
               150       160       170       180       190       200


gi|162 ENSEKTTMPLW
               210


>>gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea mays]  (252 aa)
 initn:  52 init1:  52 opt:  58  Z-score: 85.6  bits: 21.7 E():  8.5
Smith-Waterman score: 58;  40.000% identity (75.000% similar) in 20 aa overlap (28-47:6-24)


              10        20        30        40        50        60
RF_1_+ MECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQK
                                   :: ::    :.::.. :....
gi|898                      ARALVFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKA
                                      10        20        30


              70        80
RF_1_+ TAYPSLVRPARTIELQSTAP


gi|898 TWYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVEC
           40        50        60        70        80        90


>>gi|1398916|dbj|BAA07712.1| allergenic protein [Oryza s  (157 aa)
 initn:  32 init1:  32 opt:  55  Z-score: 85.0  bits: 20.9 E():  9.3
Smith-Waterman score: 55;  32.653% identity (51.020% similar) in 49 aa overlap (34-80:19-66)


              10        20        30        40        50        60
RF_1_+ PTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA--SLSGPNLQKT
                                :: .    :  :: : :.   .: :...
gi|139                 MASNKVVIPALLVVVVSVLAATTTMADHHQEQVVYTPGQLCQPGIGYP
                              10        20        30        40


              70        80
RF_1_+ AYPSLVRPARTIELQSTAP
       .:: :  :   .. : .::
gi|139 TYP-LPRCRAFVKRQCVAPGTLDEQVRRGCCRQLAGIDSSWCRCDALNHMLRIIYREERA
           50        60        70        80        90       100

>>gi|23616954|dbj|BAC20657.1| allergen RA16 [Oryza sativ  (157 aa)
 initn:  32 init1:  32 opt:  55  Z-score: 85.0  bits: 20.9 E():  9.3
Smith-Waterman score: 55;  32.653% identity (51.020% similar) in 49 aa overlap (34-80:19-66)
```

```
                10        20        30        40        50        60
RF_1_+ PTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA--SLSGPNLQKT
                                     ::  .      :   :: : :.   .:  :..
gi|236           MASNKVVISALLVVVVSVLAATTTMADHHQEVVYTPGQLCQPGIGYP
                      10        20        30        40


                  70         80
RF_1_+ AYPSLVRPARTIELQSTAP
        .:: : :    .. : .::
gi|236 TYP-LPRCRAFVKRQCVAPGTVDEQVRRGCCRQLAAIDSSWCRCDALNHMLRIIYRESGA
         50        60        70        80        90        100




80 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:41 2010 done: Fri Apr 30 23:42:42 2010
 Total Scan time:  0.080 Total Display time:  0.010


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_+1: 10  - 89 80 aa - 80 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library


       opt      E()
< 20      2    0:=
  22      0    0:              one = represents 3 library sequences
  24      2    0:=
  26      0    0:
  28     22    0:=======
  30     11    2:*===
  32     32    8:==*========
  34     31   21:======*====
  36     69   44:=============*========
  38     75   72:====================*=
  40     98  101:==============================*
  42    128  123:========================================*==
  44    134  136:==========================================*
  46    121  138:======================================    *
  48     83  132:=========================            *
```

```
  50    104    121:================================         *
  52     85    106:============================          *
  54     84     91:===========================   *
  56    123     76:==========================*=============
  58     71     62:==================*===
  60     35     50:===========     *
  62     28     40:=========    *
  64     32     32:=========*
  66     19     25:====== *
  68     21     20:=====*
  70     19     16:=====*=
  72     10     12:===*
  74      4     10:== *
  76      1      7:= *
  78      1      6:=*
  80      4      4:=*
  82      3      3:*
  84      3      3:*
  86      8      2:*==
  88      2      2:*            inset = represents 1 library sequences
  90      3      1:*
  92      0      1:*         :*
  94      2      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))= 5.5162+/-0.00378; mu= 5.1661+/- 0.193
 mean_var=41.7162+/-11.144, 0's: 2 Z-trim: 2  B-trim: 125 in 1/42
  Lambda= 0.198574
  Kolmogorov-Smirnov  statistic: 0.0661 (N=29) at  42

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:  0.090
The best scores are:                          opt bits E(1471)
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 23.6      2
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.5    2.6
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 24.3    2.7
```

```
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.2    4.6
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.2    4.6
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.2    4.6
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.4    5.7
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.2    5.7
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.6    6.4
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   59 21.8    7.1
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   59 21.8    7.1
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   59 21.8    7.1
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.1    7.1
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 22.6    7.2
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.6    7.8
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.6    8.2
gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea m ( 252)   58 21.5    9.8


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]    (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score: 96.8  bits: 23.6 E():    2
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (8-64:114-176)

                          10        20        30
RF_1_+                   ECPTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                           .: . : . ... .: :.   :::   : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
            90       100       110       120       130       140

               40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPY
          .  : :  .::..: :   :      :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            150       160       170       180       190       200

gi|162 ENSEKTTISLW
          210


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score: 94.9  bits: 22.5 E():  2.6
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (8-64:29-91)

                                10        20        30
RF_1_+                   ECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                           .: . : . ... .: :.   :::   : :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
                 10        20        30        40        50        60

              40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPY
          .  : :  .::..: :   :      :  ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
                 70        80        90       100       110       120
```

gi|159 SEKTTMPLW

>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly (473 aa)
 initn: 38 init1: 38 opt: 68 Z-score: 94.7 bits: 24.3 E(): 2.7
Smith-Waterman score: 68; 30.000% identity (55.000% similar) in 60 aa overlap (13-71:301-356)

```
                            10        20        30        40
RF_1_+                  ECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                  :: .:   .  .   . : .:..     ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            280       290       300       310       320
```

```
              50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPY
         :.:  :   :::.   .  . .::.   :   .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
        330       340       350       360       370       380
```

>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly (373 aa)
 initn: 36 init1: 36 opt: 64 Z-score: 90.5 bits: 23.2 E(): 4.6
Smith-Waterman score: 64; 28.333% identity (55.000% similar) in 60 aa overlap (13-71:216-271)

```
                            10        20        30        40
RF_1_+                  ECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                  :: .:   .  .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            190       200       210       220       230       240
```

```
              50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPY
         :.:  :   :::.   .  . .::.   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300
```

>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly (373 aa)
 initn: 36 init1: 36 opt: 64 Z-score: 90.5 bits: 23.2 E(): 4.6
Smith-Waterman score: 64; 28.333% identity (55.000% similar) in 60 aa overlap (13-71:216-271)

```
                            10        20        30        40
RF_1_+                  ECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                  :: .:   .  .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            190       200       210       220       230       240
```

```
              50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPY
         :.:  :   :::.   .  . .::.   .   .:
```

```
gi|187  ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 90.5  bits: 23.2 E():  4.6
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (13-71:216-271)


                           10        20        30        40
RF_1_+                 ECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                           ::  .:     .  .   .  :  .:..      ::..:
gi|187  GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240


              50        60        70        80
RF_1_+  -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPY
           :.:   :    :::  .   . .::.   .   .:
gi|187  ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 88.9  bits: 22.4 E():  5.7
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (23-46:13-35)


               10        20        30        40        50        60
RF_1_+ ECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKT
                          .. .:: ::   :.::.. :....
gi|898         MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKAT
                       10        20        30        40


              70        80
RF_1_+ AYPSLVRPARTIELQSTAPY

gi|898 WYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVECS
          50        60        70        80        90       100


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 88.8  bits: 23.2 E():  5.7
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (13-71:292-347)


                           10        20        30        40
RF_1_+                 ECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                           ::  .:     .  .   .  :  .:..      ::..:
gi|187  GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            270       280       290       300       310


              50        60        70        80
RF_1_+  -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPY
           :.:   :    :::  .   . .::.   .   .:
gi|187  ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
```

>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru (172 aa)
 initn: 37 init1: 37 opt: 58 Z-score: 87.8 bits: 21.6 E(): 6.4
Smith-Waterman score: 58; 33.333% identity (52.381% similar) in 63 aa overlap (8-64:72-134)

```
                                   10        20        30
RF_1_+                    ECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                             .: . :. ... .: :.   :::   :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
            50        60        70        80        90       100

            40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPY
       : .   : :  .::..: :  :       : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
          110       120       130       140       150       160

gi|159 ENSEKTTMPLW
            170
```

>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta (224 aa)
 initn: 42 init1: 42 opt: 59 Z-score: 87.1 bits: 21.8 E(): 7.1
Smith-Waterman score: 59; 20.513% identity (52.564% similar) in 78 aa overlap (3-76:130-204)

```
                                   10        20
RF_1_+                    ECPTVVYIRSRGYPIISID--HLPIS--SAFL
                             :.  . .:..  .  ...  :::.   ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
          100       110       120       130       140       150

          30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPY
       :.       : . : : :.   :::  ..  .   .. : : . .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
          160       170       180       190       200       210

gi|162 RGPFPIIV
          220
```

>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]   (224 aa)
 initn: 42 init1: 42 opt: 59 Z-score: 87.1 bits: 21.8 E(): 7.1
Smith-Waterman score: 59; 20.513% identity (52.564% similar) in 78 aa overlap (3-76:130-204)

```
                                   10        20
RF_1_+                    ECPTVVYIRSRGYPIISID--HLPIS--SAFL
                             :.  . .:..  .  ...  :::.   ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
          100       110       120       130       140       150
```

```
             30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPY
          :.        :   . : : :.    :::  ..  .   ..  : : . .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
          160       170       180       190       200       210


gi|162 RGPFPIIV
          220


>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]    (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 87.1  bits: 21.8 E():  7.1
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (3-76:130-204)

                                        10        20
RF_1_+                     ECPTVVYIRSRGYPIISID--HLPIS--SAFL
                             :.  . .:..  . . ...   :::.    ....
gi|459 PPFLQPEVMGVSKVKEAMAPKQKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
          100       110       120       130       140       150


            30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPY
          :.        :  . : : :.    :::  ..  .   ..  : : . .:.
gi|459 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
          160       170       180       190       200       210


gi|459 RGPFPIIV
          220


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 87.1  bits: 22.1 E():  7.1
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (23-46:19-41)

                10        20        30        40        50        60
RF_1_+ ECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKT
                         .. :::  :   :.:i.. :.....
gi|115     MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDAKAT
                    10        20        30        40        50


            70        80
RF_1_+ AYPSLVRPARTIELQSTAPY

gi|115 WYGKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAECS
             60        70        80        90        100       110


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 87.0  bits: 22.6 E():  7.2
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (17-68:13-63)
```

```
                10        20        30        40        50        60
RF_1_+  ECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKT
                    :.    :: . :  : : .    . :. ::.. :   ::..   . :
gi|212      MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAGMTT
                    10        20        30        40        50


                70        80
RF_1_+  AYPSLVRPARTIELQSTAPY
             .  .: : ::
gi|212  VVRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSDEVK
             60        70        80        90       100       110


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn: 37 init1: 37 opt: 58  Z-score: 86.3  bits: 21.6 E(): 7.8
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (8-64:105-167)

                                    10        20        30
RF_1_+                     ECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                              .: . : . ... .: :.   :::   :
gi|159  EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
             80        90       100       110       120       130


                40        50        60        70        80
RF_1_+  ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPY
          :. . : : .::..: :   :     :  ::::
gi|159  AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            140       150       160       170       180       190


gi|159  ENSEKTTMPLW
            200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn: 37 init1: 37 opt: 58  Z-score: 86.0  bits: 21.6 E(): 8.2
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (8-64:114-176)

                                    10        20        30
RF_1_+                     ECPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                              .: . : . ... .: :.   :::   :
gi|162  EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
             90       100       110       120       130       140


                40        50        60        70        80
RF_1_+  ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPY
          :. . : : .::..: :   :     :  ::::
gi|162  AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            150       160       170       180       190       200


gi|162  ENSEKTTMPLW
            210
```

```
>>gi|89892723|gb|ABD79095.1| Zea m 1 allergen [Zea mays]  (252 aa)
 initn:  52 init1:  52 opt:  58  Z-score: 84.6  bits: 21.5 E():  9.8
Smith-Waterman score: 58;  40.000% identity (75.000% similar) in 20 aa overlap (27-46:6-24)


                  10        20        30        40        50        60
RF_1_+ ECPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKT
                                   :: ::   :.::.. :....
gi|898                     ARALVFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKAT
                                     10        20        30


                  70        80
RF_1_+ AYPSLVRPARTIELQSTAPY


gi|898 WYGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVECS
            40        50        60        70        80        90




80 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:42 2010 done: Fri Apr 30 23:42:42 2010
 Total Scan time:  0.090 Total Display time:  0.010

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_+1: 11  - 90 80 aa - 80 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library


      opt      E()
< 20     2    0:=
  22     1    0:=            one = represents 3 library sequences
  24     3    0:=
  26     2    0:=
  28    16    0:======
  30    12    2:*===
  32    31    8:==*========
  34    33   21:======*====
  36    64   44:=============*=======
  38    81   72:====================*===
```

```
 40     98    101:================================*
 42    125    123:======================================*=
 44    126    136:=======================================    *
 46    128    138:=========================================   *
 48     82    132:=========================                    *
 50     95    121:=============================              *
 52     85    106:==========================            *
 54     97     91:=============================*==
 56    120     76:========================*=============
 58     70     62:==================*===
 60     35     50:===========      *
 62     34     40:=========== *
 64     30     32:=========*
 66     24     25:=======*
 68     16     20:=====*
 70     18     16:====*
 72      8     12:===*
 74      3     10:=   *
 76      3      7:= *
 78      4      6:=*
 80      4      4:=*
 82      2      3:*
 84      3      3:*
 86      8      2:*==
 88      2      2:*             inset = represents 1 library sequences
 90      3      1:*
 92      0      1:*         :*
 94      2      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))= 5.6689+/-0.00372; mu= 4.3527+/- 0.190
mean_var=42.6297+/-11.386, 0's: 2 Z-trim: 2  B-trim: 125 in 1/42
 Lambda= 0.196435
 Kolmogorov-Smirnov  statistic: 0.0652 (N=29) at  38

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:  0.090
The best scores are:                                    opt bits E(1471)
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 23.5    2.2
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.4    2.7
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 24.2    2.9
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.1    4.9
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.1    4.9
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.1    4.9
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.3      6
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.1    6.1
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.5    6.8
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   59 21.8    7.4
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   59 21.8    7.4
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   59 21.8    7.4
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.0    7.5
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 22.5    7.7
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.5    8.2
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.5    8.6


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]    (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score: 96.3  bits: 23.5 E():  2.2
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (7-63:114-176)

                                    10        20        30
RF_1_+                       CPTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                               .:  .  :  . ... .: :.    :::   : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
           90       100       110       120       130       140

                      40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
          .  :  :   .::..:  :    :       :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
          150       160       170       180       190       200

gi|162 ENSEKTTISLW
          210


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score: 94.6  bits: 22.4 E():  2.7
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (7-63:29-91)

                                    10        20        30
RF_1_+                       CPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                               .:  .  :  . ... .: :.    :::   :  :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
              10        20        30        40        50        60

             40        50        60        70        80
```

```
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
         .   : :  .:!..: :  :      :  ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
                70        80        90       100       110       120


gi|159 SEKTTMPLW


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly  (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score:  94.0  bits: 24.2 E():  2.9
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (12-70:301-356)

                                10        20        30        40
RF_1_+                   CPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .:    . .   . : .:..    ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
               280       290       300       310       320

             50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
          :.:  :    ::: .  . .:!.  :  .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          330       340       350       360       370       380


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score:  89.9  bits: 23.1 E():  4.9
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (12-70:216-271)

                                10        20        30        40
RF_1_+                   CPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .:    . .   . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
               190       200       210       220       230       240

             50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
          :.:  :    ::: .  . .:!.   .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
               250       260       270       280       290       300


>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score:  89.9  bits: 23.1 E():  4.9
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (12-70:216-271)

                                10        20        30        40
RF_1_+                   CPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .:    . .   . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
               190       200       210       220       230       240
```

```
                  50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
         :.:   :    :::.   .  .:.   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.9  bits: 23.1 E():  4.9
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (12-70:216-271)


                                  10        20        30        40
RF_1_+                 CPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             :: .:   .  .  . : .:..    :::.:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


                  50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
         :.:   :    :::.   .  .:.   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 88.4  bits: 22.3 E():    6
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (22-45:13-35)


                  10        20        30        40        50        60
RF_1_+ CPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTA
                       .. .:: ::    :.::.. :....
gi|898          MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKATW
                         10        20        30        40        50


                  70        80
RF_1_+ YPSLVRPARTIELQSTAPYL

gi|898 YGKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVECSG
                 60        70        80        90       100       110


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 88.2  bits: 23.1 E():  6.1
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (12-70:292-347)


                                  10        20        30        40
RF_1_+                 CPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             :: .:   .  .  . : .:..    :::.:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          270       280       290       300       310
```

```
                 50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
          :.:  :    :::.  .  . .::.  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          320       330       340       350       360       370


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.5  bits: 21.5 E():  6.8
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (7-63:72-134)

                                    10        20        30
RF_1_+                     CPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                            .:. :. ...  .: :.    :::    :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                  50        60        70        80        90       100

            40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
         :.   : :  .::..: :   :      :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
                 110       120       130       140       150       160

gi|159 ENSEKTTMPLW
                 170


>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta  (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 86.7  bits: 21.8 E():  7.4
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (2-75:130-204)

                                    10        20
RF_1_+                     CPTVVYIRSRGYPIISID--HLPIS--SAFL
                            :.  . .:..  .  ...   :::.   ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
          100       110       120       130       140       150

           30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
         :.    :   : : :.   :::  ..  .   ..  : : .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
          160       170       180       190       200       210

gi|162 RGPFPIIV
          220


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]       (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 86.7  bits: 21.8 E():  7.4
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (2-75:130-204)

                                    10        20
```

```
RF_1_+                                    CPTVVYIRSRGYPIISID--HLPIS--SAFL
                                          :.   . .:..  . ...   :::.   ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
          100       110       120       130       140       150


           30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
         :.        :  . : : :.    :::  ..   .   ..  : : . .:.
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
          160       170       180       190       200       210


gi|162 RGPFPIIV
           220


>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]    (224 aa)
 initn:  42 init1:  42 opt:  59  Z-score: 86.7  bits: 21.8 E():  7.4
Smith-Waterman score: 59;  20.513% identity (52.564% similar) in 78 aa overlap (2-75:130-204)


                                          10        20
RF_1_+                                    CPTVVYIRSRGYPIISID--HLPIS--SAFL
                                          :.   . .:..  . ...   :::.   ....
gi|459 PPFLQPEVMGVSKVKEAMAPKQKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
          100       110       120       130       140       150


           30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
         :.        :  . : : :.    :::  ..   .   ..  : : . .:.
gi|459 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
          160       170       180       190       200       210


gi|459 RGPFPIIV
           220


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 86.7  bits: 22.0 E():  7.5
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (22-45:19-41)


                 10        20        30        40        50        60
RF_1_+ CPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTA
                          .. :::  :   :.:::..  :....
gi|115    MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDAKATW
                  10        20        30        40        50


                 70        80
RF_1_+ YPSLVRPARTIELQSTAPYL


gi|115 YGKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAECSG
           60        70        80        90       100       110
```

```
>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 86.4  bits: 22.5 E():  7.7
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (16-67:13-63)

                  10        20        30        40        50        60
RF_1_+ CPTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTA
                :.   ::. . :   : :.  . . :. ::..  :   ::..    . :.
gi|212     MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAGMTTV
                  10        20        30        40        50

               70        80
RF_1_+ YPSLVRPARTIELQSTAPYL
          .: ::
gi|212 VRDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSDEVKR
             60        70        80        90       100       110


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.9  bits: 21.5 E():  8.2
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (7-63:105-167)

                              10        20        30
RF_1_+                 CPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                          .: . : . ... .: :.   :::    :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                80        90       100       110       120       130

                  40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
         :.   : :  .:.... :   :       : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
                140       150       160       170       180       190

gi|159 ENSEKTTMPLW
             200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.6  bits: 21.5 E():  8.6
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (7-63:114-176)

                              10        20        30
RF_1_+                 CPTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                          .: . : . ... .: :.   :::    :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                90       100       110       120       130       140

                  40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYL
         :.   : :  .:.... :   :       : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
```

```
              150         160         170         180         190         200

gi|162 ENSEKTTMPLW
              210




80 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:43 2010 done: Fri Apr 30 23:42:43 2010
 Total Scan time:  0.090 Total Display time:  0.010


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_+1: 12  - 91 80 aa - 80 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library


        opt      E()
< 20     2     0:=
  22     0     0:                one = represents 3 library sequences
  24     1     0:=
  26     1     0:=
  28    14     0:=====
  30    19     2:*======
  32    28     8:==*=======
  34    28    21:======*===
  36    54    44:=============*====
  38   101    72:====================*==========
  40   105   101:=============================*=
  42   118   123:====================================*
  44   113   136:===================================         *
  46   131   138:========================================= *
  48    90   132:=============================                *
  50    95   121:==========================            *
  52    87   106:=========================          *
  54    90    91:=============================*
  56   120    76:=======================*==============
  58    68    62:==================*==
  60    40    50:=============  *
  62    40    40:============*
```

```
  64    27    32:========= *
  66    24    25:=======*
  68    15    20:==== *
  70    16    16:=====*
  72    10    12:===*
  74     2    10:=   *
  76     1     7:= *
  78     2     6:=*
  80     8     4:=*=
  82     2     3:*
  84     4     3:*=
  86     5     2:*=
  88     2     2:*           inset = represents 1 library sequences
  90     2     1:*
  92     3     1:*         :*==
  94     2     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))= 5.7922+/-0.00372; mu= 3.6888+/- 0.190
 mean_var=43.6479+/-11.625, 0's: 2 Z-trim: 2  B-trim: 125 in 1/42
 Lambda= 0.194130
 Kolmogorov-Smirnov  statistic: 0.0702 (N=29) at  40

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:  0.080
The best scores are:                                opt bits E(1471)
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 23.4    2.3
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.3    2.9
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 24.1    3.2
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   63 22.8    3.6
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   63 22.8    3.6
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   63 22.8    3.6
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.3
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.3
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.3
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.2    6.4
```

```
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)    64 22.9     6.6
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)    58 21.4     7.1
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)    60 21.9      8
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)    62 22.4     8.3
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)    58 21.4     8.7
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)    58 21.4     9.1


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]    (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score: 95.8  bits: 23.4 E():  2.3
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (6-62:114-176)

                                   10        20        30
RF_1_+                    PTVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                                  .: .:. ... .: :.   :::   : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
               90       100       110       120       130       140

                40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
          . : :  .::..: :  :      :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
               150       160       170       180       190       200

gi|162 ENSEKTTISLW
               210


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score: 94.1  bits: 22.3 E():  2.9
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (6-62:29-91)

                                   10        20        30
RF_1_+                    PTVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                                  .: .: ... .: :.   :::   :  :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
               10        20        30        40        50        60

                40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
        .  : :  .::..: :  :      :  ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
               70        80        90       100       110       120

gi|159 SEKTTMPLW


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly  (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score: 93.3  bits: 24.1 E():  3.2
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (11-69:301-356)
```

```
                                    10        20        30        40
RF_1_+                        PTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                    :: .:   .  .   . : .:.    ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
               280       290       300       310       320


                             50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
          :.:   :    ::: .   . .:.   :   .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         330       340       350       360       370       380


>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta  (224 aa)
 initn:  42 init1:  42 opt:  63  Z-score: 92.3  bits: 22.8 E():  3.6
Smith-Waterman score: 63;  21.429% identity (51.190% similar) in 84 aa overlap (1-80:130-210)

                                             10        20
RF_1_+                               PTVVYIRSRGYPIISID--HLPIS--SAFL
                                     :.  . .:..  . ...   :::.    ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
           100       110       120       130       140       150


            30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
       :.       :  . : :.    ::: ..  .   ..  : : . .:.   : :
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
         160       170       180       190       200       210


gi|162 RGPFPIIV
          220


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]       (224 aa)
 initn:  42 init1:  42 opt:  63  Z-score: 92.3  bits: 22.8 E():  3.6
Smith-Waterman score: 63;  21.429% identity (51.190% similar) in 84 aa overlap (1-80:130-210)

                                             10        20
RF_1_+                               PTVVYIRSRGYPIISID--HLPIS--SAFL
                                     :.  . .:..  . ...   :::.    ....
gi|162 PPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
           100       110       120       130       140       150


            30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
       :.        :  . : : :.    ::: ..  .   ..   : : . .:.   : :
gi|162 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
         160       170       180       190       200       210


gi|162 RGPFPIIV
          220
```

```
>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]   (224 aa)
 initn:  42 init1:  42 opt:  63  Z-score: 92.3  bits: 22.8 E():  3.6
Smith-Waterman score: 63;  21.429% identity (51.190% similar) in 84 aa overlap (1-80:130-210)


                               10        20
RF_1_+                 PTVVYIRSRGYPIISID--HLPIS--SAFL
                       :.  . .:.. . ... :::. ....
gi|459 PPFLQPEVMGVSKVKEAMAPKQKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWM
          100       110       120       130       140       150


             30        40        50        60        70        80
RF_1_+ HSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
       :.      : . : : :.   :::  .. .   .. : : . .:.   : :
gi|459 HQPHQPLPPTVMFPPQSVL---SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPV
          160       170       180       190       200       210


gi|459 RGPFPIIV
          220


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.4  bits: 23.0 E():  5.3
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (11-69:216-271)


                               10        20        30        40
RF_1_+                 PTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .:   . .  . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


                 50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
         :.:  :   :::.  . . .::.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.4  bits: 23.0 E():  5.3
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (11-69:216-271)


                               10        20        30        40
RF_1_+                 PTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .:   . .  . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


                 50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
         :.:  :   :::.  . . .::.  .  .:
```

```
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.4  bits: 23.0 E():  5.3
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (11-69:216-271)


                                 10        20        30        40
RF_1_+                     PTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                              :: .:   .  .  . : .:..     ::...:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


               50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
         :.:  :   ::: .  . .::.  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 87.9  bits: 22.2 E():  6.4
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (21-44:13-35)


               10        20        30        40        50        60
RF_1_+ PTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAY
                          .. .:: ::   :.::.. :....
gi|898        MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKATWY
                     10        20        30        40        50


               70        80
RF_1_+ PSLVRPARTIELQSTAPYLE

gi|898 GKPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVECSGK
                 60        70        80        90       100       110


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 87.6  bits: 22.9 E():  6.6
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (11-69:292-347)


                                 10        20        30        40
RF_1_+                     PTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                              :: .:   .  .  . : .:..     ::...:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          270       280       290       300       310


               50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
         :.:  :   ::: .  . .::.  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
```

>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru (172 aa)
 initn: 37 init1: 37 opt: 58 Z-score: 87.1 bits: 21.4 E(): 7.1
Smith-Waterman score: 58; 33.333% identity (52.381% similar) in 63 aa overlap (6-62:72-134)

```
                                          10        20        30
RF_1_+                    PTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                             .: . :. ... .: :.   ::: :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
               50        60        70        80        90       100

                  40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
         :. : : .::..: : :     : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
              110       120       130       140       150       160

gi|159 ENSEKTTMPLW
              170
```

>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp (269 aa)
 initn: 54 init1: 54 opt: 60 Z-score: 86.2 bits: 21.9 E():   8
Smith-Waterman score: 60; 33.333% identity (70.833% similar) in 24 aa overlap (21-44:19-41)

```
                 10        20        30        40        50        60
RF_1_+ PTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAY
                         .. ::: :   :.::.. :....
gi|115    MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDAKATWY
                  10        20        30        40        50

               70        80
RF_1_+ PSLVRPARTIELQSTAPYLE

gi|115 GKPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAECSGK
              60        70        80        90       100       110
```

>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib (392 aa)
 initn: 48 init1: 48 opt: 62 Z-score: 85.9 bits: 22.4 E(): 8.3
Smith-Waterman score: 62; 30.769% identity (53.846% similar) in 52 aa overlap (15-66:13-63)

```
                 10        20        30        40        50        60
RF_1_+ PTVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAY
                      :.  :: . :  : : .  . :. ::.. :   ::.   . :.
gi|212    MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAGMTTVV
                  10        20        30        40        50

               70        80
RF_1_+ PSLVRPARTIELQSTAPYLE
```

```
               .: ::
gi|212 RDLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSDEVKRR
           60        70        80        90       100       110


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.5  bits: 21.4 E():  8.7
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (6-62:105-167)

                                    10        20        30
RF_1_+                       PTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                              .:. :. ... .: :.   :::   :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
            80        90       100       110       120       130


               40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
         :.   : :  .::..: :   :      :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
           140       150       160       170       180       190


gi|159 ENSEKTTMPLW
           200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.2  bits: 21.4 E():  9.1
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (6-62:114-176)

                                    10        20        30
RF_1_+                       PTVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                              .:. :. ... .: :.   :::   :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
            90       100       110       120       130       140


               40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLE
         :.   : :  .::..: :   :      :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
           150       160       170       180       190       200


gi|162 ENSEKTTMPLW
           210




80 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:46 2010 done: Fri Apr 30 23:42:46 2010
 Total Scan time:  0.080 Total Display time:  0.010
```

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_+1: 13 - 92 80 aa - 80 aa
 vs /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

```
        opt      E()
< 20     2     0:=
  22     0     0:                  one = represents 3 library sequences
  24     1     0:=
  26     1     0:=
  28     6     0:==
  30    19     2:*======
  32    23     8:==*=====
  34    36    21:======*=====
  36    62    44:=============*======
  38    95    72:====================*========
  40   103   101:============================*=
  42   125   123:========================================*=
  44   121   136:======================================     *
  46   135   138:==============================================*
  48    96   132:==============================                *
  50    77   121:=======================                   *
  52    87   106:==========================           *
  54    91    91:============================*
  56   103    76:========================*========
  58    74    62:==================*====
  60    47    50:==============*
  62    33    40:==========  *
  64    35    32:=========*=
  66    20    25:====== *
  68    18    20:=====*
  70    17    16:====*
  72    11    12:===*
  74     2    10:=  *
  76     1     7:= *
  78     2     6:=*
  80     8     4:=*=
  82     1     3:*
  84     4     3:*=
  86     5     2:*=
```

```
 88      6      2:*=             inset = represents 1 library sequences
 90      1      1:*
 92      1      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))= 5.9252+/-0.00368; mu= 2.7260+/- 0.188
mean_var=44.7868+/-12.139, 0's: 2 Z-trim: 2  B-trim: 125 in 1/42
 Lambda= 0.191646
 Kolmogorov-Smirnov  statistic: 0.0682 (N=29) at  42


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:  0.080
The best scores are:                                    opt bits E(1471)
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 23.3    2.4
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.3    2.9
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K   ( 473)   68 24.0    3.3
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   61 22.2    5.4
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   61 22.2    5.4
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   61 22.2    5.4
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 22.9    5.5
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 22.9    5.5
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 22.9    5.5
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.2    6.5
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 22.9    6.8
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.4    7.2
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 21.9    8.1
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 22.4    8.5
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.4    8.7
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.4    9.2


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn:  38 init1:  38 opt:  65 Z-score:  95.5  bits: 23.3 E():  2.4
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (5-61:114-176)
```

```
RF_1_+                           TVVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                                 .: . : . ... .: :.   :::   : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
           90        100       110       120       130       140


       30        40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
         . : : .::..: : :    : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
           150       160       170       180       190       200


gi|162 ENSEKTTISLW
          210


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn: 37 init1: 37 opt: 61 Z-score: 94.0 bits: 22.3 E(): 2.9
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (5-61:29-91)


                           10        20        30
RF_1_+                     TVVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                           .: . : . ... .: :.   :::   : :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
               10        20        30        40        50        60


              40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
         . : : .::..: : :    : ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
                 70        80        90       100       110       120


gi|159 SEKTTMPLW


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly  (473 aa)
 initn: 38 init1: 38 opt: 68 Z-score: 93.0 bits: 24.0 E(): 3.3
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (10-68:301-356)


                           10        20        30
RF_1_+                     TVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                           :: .:   . .  . : .:..    ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
               280       290       300       310       320


       40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
         :.: :   ::: .  . .::.  : .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          330       340       350       360       370       380
```

```
>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta  (224 aa)
 initn:  42 init1:  42 opt:  61  Z-score: 89.2  bits: 22.2 E():  5.4
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (18-79:152-210)

                             10        20        30        40
RF_1_+             TVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                                 ::. ....:.        :  . : : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
               130       140       150       160       170


           50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
          :::  ..  .   ..   : : . .:.    : :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
          180       190       200       210       220


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]       (224 aa)
 initn:  42 init1:  42 opt:  61  Z-score: 89.2  bits: 22.2 E():  5.4
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (18-79:152-210)


                             10        20        30        40
RF_1_+             TVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                                 ::. ....:.        :  . : : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
               130       140       150       160       170


           50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
          :::  ..  .   ..   : : . .:.    : :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
          180       190       200       210       220


>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]    (224 aa)
 initn:  42 init1:  42 opt:  61  Z-score: 89.2  bits: 22.2 E():  5.4
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (18-79:152-210)


                             10        20        30        40
RF_1_+             TVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                                 ::. ....:.        :  . : : :.
gi|459 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
               130       140       150       160       170


           50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
          :::  ..  .   ..   : : . .:.    : :
gi|459 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
          180       190       200       210       220


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
```

```
 initn:  36 init1:  36 opt:  64  Z-score: 89.1  bits: 22.9 E():  5.5
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (10-68:216-271)

                              10        20        30
RF_1_+                        TVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                :: .:   . .  . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            190       200       210       220       230       240

           40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
         :.:  :   ::: .  . .:.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.1  bits: 22.9 E():  5.5
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (10-68:216-271)

                              10        20        30
RF_1_+                        TVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                :: .:   . .  . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            190       200       210       220       230       240

           40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
         :.:  :   ::: .  . .:.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
              250       260       270       280       290       300


>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.1  bits: 22.9 E():  5.5
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (10-68:216-271)

                              10        20        30
RF_1_+                        TVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                :: .:   . .  . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            190       200       210       220       230       240

           40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
         :.:  :   ::: .  . .:.  .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
              250       260       270       280       290       300


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 87.7  bits: 22.2 E():  6.5
```

```
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (20-43:13-35)

                   10        20        30        40        50        60
RF_1_+ TVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYP
                          .. .:: ::    :.::.. :....
   gi|898        MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKATWYG
                          10        20        30        40        50


                   70        80
RF_1_+ SLVRPARTIELQSTAPYLEA

   gi|898 KPTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVECSGKP
                  60        70        80        90       100       110


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 87.4  bits: 22.9 E():  6.8
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (10-68:292-347)

                                10        20        30
RF_1_+                     TVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                :: .:  .  .  . :  .:..   :::..:
   gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
               270       280       290       300       310


           40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
         :.:   :    :::  .   . .::.   .   .:
   gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          320       330       340       350       360       370


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.0  bits: 21.4 E():  7.2
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (5-61:72-134)

                                10        20        30
RF_1_+                     TVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                 .: . : . ... .: :.   :::   :
   gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                   50        60        70        80        90       100


               40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
         :. . : :  .::..: :  :       :  ::::
   gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
                  110       120       130       140       150       160


   gi|159 ENSEKTTMPLW
                170
```

```
>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 86.0  bits: 21.9 E():  8.1
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (20-43:19-41)

               10        20        30        40        50        60
RF_1_+ TVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYP
                        .. :::  :    :.::.. :....
gi|115  MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDAKATWYG
               10        20        30        40        50

               70        80
RF_1_+ SLVRPARTIELQSTAPYLEA

gi|115 KPTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAECSGKP
           60        70        80        90       100       110


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 85.7  bits: 22.4 E():  8.5
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (14-65:13-63)

               10        20        30        40        50        60
RF_1_+ TVVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYP
                 :.  :: . :   : : .  . :. ::.. :   :::.    . :.
gi|212  MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAGMTTVVR
                10        20        30        40        50

               70        80
RF_1_+ SLVRPARTIELQSTAPYLEA
          .: ::
gi|212 DLDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSDEVKRRF
           60        70        80        90       100       110


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.5  bits: 21.4 E():  8.7
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (5-61:105-167)

                                  10        20        30
RF_1_+                     TVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                            .: . :. ... .: :.   :::    :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
            80        90       100       110       120       130

               40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
         :.   : :  .:..: :   :      :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            140       150       160       170       180       190


gi|159 ENSEKTTMPLW
```

>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo (214 aa)
 initn: 37 init1: 37 opt: 58 Z-score: 85.1 bits: 21.4 E(): 9.2
Smith-Waterman score: 58; 33.333% identity (52.381% similar) in 63 aa overlap (5-61:114-176)

```
                                          10        20            30
RF_1_+                         TVVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                     .:. : . ... .: :.   :::   :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
             90        100       110       120       130       140


                  40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEA
        :  .  : :  .:.:.: :   :       :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            150       160       170       180       190       200


gi|162 ENSEKTTMPLW
          210
```

80 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:47 2010 done: Fri Apr 30 23:42:48 2010
 Total Scan time: 0.080 Total Display time: 0.010

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_+1: 14 - 93 80 aa - 80 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    4    2:*=
```

```
  32    42      8:==*===========
  34    35     21:======*=====
  36    63     44:=============*======
  38    91     72:=====================*======
  40   109    101:============================*===
  42   126    123:==================================*=
  44   118    136:===================================      *
  46   137    138:=====================================*
  48    97    132:==================================            *
  50    77    121:=======================              *
  52    91    106:============================     *
  54    91     91:============================*
  56   100     76:=======================*=======
  58    74     62:==================*====
  60    46     50:==============*
  62    33     40:==========   *
  64    36     32:=========*=
  66    20     25:====== *
  68    18     20:=====*
  70    17     16:====*
  72    10     12:===*
  74     2     10:=  *
  76     1      7:= *
  78     2      6:=*
  80     8      4:=*=
  82     1      3:*
  84     4      3:*=
  86     5      2:*=
  88     4      2:*=           inset = represents 1 library sequences
  90     3      1:*
  92     0      1:*         :*
  94     2      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))= 5.9097+/-0.00367; mu= 2.8333+/- 0.187
 mean_var=44.4068+/-12.027, 0's: 2 Z-trim: 2  B-trim: 125 in 1/42
 Lambda= 0.192464
```

```
  Kolmogorov-Smirnov  statistic: 0.0682 (N=29) at  42

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:  0.080
The best scores are:                                   opt bits E(1471)
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 23.4    2.4
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.3    2.9
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 24.0    3.3
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   61 22.2    5.3
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   61 22.2    5.3
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   61 22.2    5.3
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.4
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.4
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.4
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.2    6.4
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 22.9    6.7
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.5    7.1
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 21.9      8
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 22.4    8.3
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.4    8.6
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.4      9


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn: 38 init1: 38 opt:  65 Z-score: 95.7  bits: 23.4 E():  2.4
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (4-60:114-176)

                                  10        20
RF_1_+                   VVYIRSRGYPIISIDHLPISSAF-LHS---G-D
                          .:. .:. ... .: :.   :::    : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
            90        100       110       120       130       140


      30        40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
        . : :  .::..: :   :      :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            150       160       170       180       190       200


gi|162 ENSEKTTISLW
            210


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn: 37 init1: 37 opt:  61 Z-score: 94.2  bits: 22.3 E():  2.9
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (4-60:29-91)

                                  10        20        30
RF_1_+                   VVYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                          .:. .:. ... .: :.   :::    : :
```

```
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
                  10        20        30        40        50        60

                  40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
        .   : :  .::..: :  :      :  ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
                  70        80        90       100       110       120


gi|159 SEKTTMPLW


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly   (473 aa)
 initn: 38 init1: 38 opt: 68  Z-score: 93.2  bits: 24.0 E():  3.3
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (9-67:301-356)

                                    10        20        30
RF_1_+                     VVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                           :: .:   .  . . : .:..    ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
               280       290       300       310       320

           40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
        :.:  :   ::: .  . .:. :   : .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
           330       340       350       360       370       380


>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta   (224 aa)
 initn: 42 init1: 42 opt:  61  Z-score: 89.3  bits: 22.2 E():  5.3
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (17-78:152-210)

                                    10        20        30        40
RF_1_+                     VVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                           ::. ....:.       :  . : : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
               130       140       150       160       170

           50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
        :::  ..  .  .. : : . .:.   : :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
           180       190       200       210       220


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]        (224 aa)
 initn: 42 init1: 42 opt:  61  Z-score: 89.3  bits: 22.2 E():  5.3
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (17-78:152-210)

                                    10        20        30        40
```

```
RF_1_+                         VVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                                ::. ....:.        :  . : : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
              130       140       150       160       170


          50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
        :::  ..  .   ..  : : .:.    : :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
         180       190       200       210       220


>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]    (224 aa)
 initn:  42 init1:  42 opt:  61  Z-score: 89.3  bits: 22.2 E():  5.3
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (17-78:152-210)


                                  10        20        30        40
RF_1_+                         VVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                                ::. ....:.        :  . : : :.
gi|459 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
              130       140       150       160       170


          50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
        :::  ..  .   ..  : : .:.    : :
gi|459 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
         180       190       200       210       220


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.3  bits: 23.0 E():  5.4
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (9-67:216-271)


                                  10        20        30
RF_1_+                         VVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                ::  .:   .  .  . : .:..    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            190       200       210       220       230       240


          40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
        :.:  :   :::  .   . .:..   .  .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.3  bits: 23.0 E():  5.4
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (9-67:216-271)


                                  10        20        30
RF_1_+                         VVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
```

```
                                  :: .:   .   .  . :  .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


           40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
          :.:   :    :::  .    .  .:.   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
              250       260       270       280       290       300


>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn: 36 init1: 36 opt: 64  Z-score: 89.3  bits: 23.0 E():  5.4
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (9-67:216-271)


                            10        20        30
RF_1_+                   VVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                  :: .:   .   .  . :  .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190       200       210       220       230       240


           40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
          :.:   :    :::  .    .  .:.   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
              250       260       270       280       290       300


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn: 52 init1: 52 opt: 61  Z-score: 87.9  bits: 22.2 E():  6.4
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (19-42:13-35)


             10        20        30        40        50        60
RF_1_+ VVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPS
                       .. .:: ::   :.:.. .....
gi|898       MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKATWYGK
                      10        20        30        40        50


             70        80
RF_1_+ LVRPARTIELQSTAPYLEAE

gi|898 PTGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVECSGKPV
                60        70        80        90       100       110


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn: 36 init1: 36 opt: 64  Z-score: 87.5  bits: 22.9 E():  6.7
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (9-67:292-347)


                            10        20        30
RF_1_+                   VVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                  :: .:   .   .  . :  .:..     ::..:
```

```
gi|187  GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
             270       280       290       300       310


          40        50        60        70        80
RF_1_+  -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
          :.:   :    :::  .   . .:::.   .   .:
gi|187  ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          320       330       340       350       360       370


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.1  bits: 21.5 E():  7.1
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (4-60:72-134)


                                   10        20
RF_1_+                     VVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                           .: . :. . ... .: :.   ::: :
gi|159  EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                50        60        70        80        90       100


     30        40        50        60        70        80
RF_1_+  ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
          :  .   : :  .::..: :   :      :  ::::
gi|159  AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
                110       120       130       140       150       160


gi|159  ENSEKTTMPLW
             170


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 86.2  bits: 21.9 E():    8
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (19-42:19-41)


             10        20        30        40        50        60
RF_1_+  VVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPS
                         .. :::  :   :.:i.. :....
gi|115  MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDAKATWYGK
             10        20        30        40        50


             70        80
RF_1_+  LVRPARTIELQSTAPYLEAE

gi|115  PTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAECSGKPV
             60        70        80        90       100       110


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 85.8  bits: 22.4 E():  8.3
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (13-64:13-63)


             10        20        30        40        50        60
```

```
RF_1_+ VVYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPS
               :.  :: . :  : : .   . :. ::.. :   ::..    . :.  .
gi|212 MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAGMTTVVRD
               10        20        30        40        50

               70        80
RF_1_+ LVRPARTIELQSTAPYLEAE
         : ::
gi|212 LDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSDEVKRRFY
        60        70        80        90       100       110


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.6  bits: 21.4 E():  8.6
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (4-60:105-167)

                                        10        20
RF_1_+                      VVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                            .: . :. . ... .: :.   :::   :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
              80        90       100       110       120       130

     30        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
        :.   : : .::..: :  :      : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            140       150       160       170       180       190

gi|159 ENSEKTTMPLW
            200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.2  bits: 21.4 E():    9
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (4-60:114-176)

                                        10        20
RF_1_+                      VVYIRSRGYPIISIDHLPISSAF-LHS---GDL
                            .: . :. . ... .: :.   :::   :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
              90       100       110       120       130       140

     30        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAE
        :.   : : .::..: :  :      : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            150       160       170       180       190       200

gi|162 ENSEKTTMPLW
            210
```

80 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:48 2010 done: Fri Apr 30 23:42:48 2010
 Total Scan time: 0.080 Total Display time: 0.010

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_+1: 15 - 94 80 aa - 80 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   4   2:*=
  32   36   8:==*=========
  34   37   21:======*======
  36   69   44:=============*========
  38   87   72:=====================*=====
  40   109   101:===============================*===
  42   126   123:====================================*=
  44   114   136:===================================   *
  46   142   138:==*==
  48   94   132:==============================   *
  50   77   121:========================   *
  52   91   106:============================   *
  54   89   91:============================*
  56   101   76:=======================*========
  58   75   62:==================*====
  60   50   50:===============*
  62   36   40:=========== *
  64   34   32:=========*=
  66   20   25:======= *
  68   17   20:======*
  70   16   16:=====*
  72   11   12:===*

```
   74     2    10:=  *
   76     1     7:= *
   78     2     6:=*
   80     7     4:=*=
   82     2     3:*
   84     4     3:*=
   86     5     2:*=
   88     4     2:*=           inset = represents 1 library sequences
   90     3     1:*
   92     0     1:*        :*
   94     2     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))= 5.9037+/-0.00363; mu= 2.8737+/- 0.185
 mean_var=44.3620+/-12.012, 0's: 2 Z-trim: 2  B-trim: 125 in 1/42
  Lambda= 0.192561
  Kolmogorov-Smirnov  statistic: 0.0668 (N=29) at   42
```

```
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:  0.080
The best scores are:                              opt bits E(1471)
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 23.4    2.3
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.3    2.9
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 24.0    3.2
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   61 22.2    5.3
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   61 22.2    5.3
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   61 22.2    5.3
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.4
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.4
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.4
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.2    6.4
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 22.9    6.7
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.5    7.1
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 21.9      8
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 22.4    8.3
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.4    8.6
```

```
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)    58 21.4       9

>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score: 95.7  bits: 23.4 E():  2.3
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (3-59:114-176)

                                       10        20
RF_1_+                      VYIRSRGYPIISIDHLPISSAF-LHS---G-D
                              .: . :. ... .: :.   ::: : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
              90       100       110       120       130       140

        30        40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
         . : : .::..: : :    : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             150       160       170       180       190       200

gi|162 ENSEKTTISLW
          210

>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score: 94.2  bits: 22.3 E():  2.9
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (3-59:29-91)

                                       10        20        30
RF_1_+                      VYIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                              .: . :. ... .: :.   :::   : :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
              10        20        30        40        50        60

          40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
         .  : : .::..: : :    : ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
             70        80        90       100       110       120

gi|159 SEKTTMPLW

>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly   (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score: 93.2  bits: 24.0 E():  3.2
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (8-66:301-356)

                                       10        20        30
RF_1_+                      VYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                              :: .:   . .  . : .:..    ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
              280       290       300       310       320
```

```
        40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
        :.:   :    :::.   .   . .::.   :   .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
      330       340       350       360       370       380


>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta  (224 aa)
 initn: 42 init1: 42 opt: 61  Z-score: 89.3  bits: 22.2 E():  5.3
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (16-77:152-210)

                          10        20        30        40
RF_1_+             VYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                             ::. ....:.       :  . : : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
            130       140       150       160       170


        50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
        :::  ..  .   ..  : : ..:.    : :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
       180       190       200       210       220


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]      (224 aa)
 initn: 42 init1: 42 opt: 61  Z-score: 89.3  bits: 22.2 E():  5.3
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (16-77:152-210)

                          10        20        30        40
RF_1_+             VYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                             ::. ....:.       :  . : : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
            130       140       150       160       170


        50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
        :::  ..  .   ..  : : ..:.    : :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
       180       190       200       210       220


>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]    (224 aa)
 initn: 42 init1: 42 opt: 61  Z-score: 89.3  bits: 22.2 E():  5.3
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (16-77:152-210)

                          10        20        30        40
RF_1_+             VYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                             ::. ....:.       :  . : : :.
gi|459 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
            130       140       150       160       170
```

```
               50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
       :::  ..  .   ..  : : . .:.    : :
gi|459 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
           180       190       200       210       220
```

\>\>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly (373 aa)
 initn: 36 init1: 36 opt: 64 Z-score: 89.3 bits: 23.0 E(): 5.4
Smith-Waterman score: 64; 28.333% identity (55.000% similar) in 60 aa overlap (8-66:216-271)

```
                               10        20        30
RF_1_+                 VYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                       ::  .:    .  .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240

          40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
       :.:  :   :::  .   . .:..  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
             250       260       270       280       290       300
```

\>\>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly (373 aa)
 initn: 36 init1: 36 opt: 64 Z-score: 89.3 bits: 23.0 E(): 5.4
Smith-Waterman score: 64; 28.333% identity (55.000% similar) in 60 aa overlap (8-66:216-271)

```
                               10        20        30
RF_1_+                 VYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                       ::  .:    .  .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240

          40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
       :.:  :   :::  .   . .:..  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
             250       260       270       280       290       300
```

\>\>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly (373 aa)
 initn: 36 init1: 36 opt: 64 Z-score: 89.3 bits: 23.0 E(): 5.4
Smith-Waterman score: 64; 28.333% identity (55.000% similar) in 60 aa overlap (8-66:216-271)

```
                               10        20        30
RF_1_+                 VYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                       ::  .:    .  .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240

          40        50        60        70        80
```

```
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
        :.:  :   :::  .  .  .:.  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 87.9  bits: 22.2 E():  6.4
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (18-41:13-35)


             10        20        30        40        50        60
RF_1_+ VYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSL
              .. .:: ::   :.::.. :....
gi|898      MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKATWYGKP
              10        20        30        40        50


             70        80
RF_1_+ VRPARTIELQSTAPYLEAEY

gi|898 TGAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVECSGKPVV
            60        70        80        90       100       110


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 87.5  bits: 22.9 E():  6.7
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (8-66:292-347)


                                    10        20        30
RF_1_+                     VYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                               ::  .:   .  .   . :  .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            270       280       290       300       310


         40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
        :.:  :   :::  .  .  .:.  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
        320       330       340       350       360       370


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.1  bits: 21.5 E():  7.1
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (3-59:72-134)


                                    10        20
RF_1_+                     VYIRSRGYPIISIDHLPISSAF-LHS---GDL
                              .: . : . ... .: :.    :::    :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
            50        60        70        80        90       100


        30        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
```

```
            :  .    :  :   .::..:  :   :       :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
                110       120       130       140       150       160


gi|159 ENSEKTTMPLW
              170


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 86.2  bits: 21.9 E():    8
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (18-41:19-41)

                  10        20        30        40        50
RF_1_+   VYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPS
                         ..  :::  :    :.::.. :....
gi|115 MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDAKATWYGK
               10        20        30        40        50


      60        70        80
RF_1_+ LVRPARTIELQSTAPYLEAEY

gi|115 PTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAECSGKPV
          60        70        80        90       100       110


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 85.8  bits: 22.4 E():  8.3
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (12-63:13-63)

                  10        20        30        40        50
RF_1_+   VYIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPS
                 :.  :: . :  : : .   . :. ::.. :    ::.     . :.  .
gi|212 MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAGMTTVVRD
               10        20        30        40        50


      60        70        80
RF_1_+ LVRPARTIELQSTAPYLEAEY
        : ::
gi|212 LDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSDEVKRRFY
          60        70        80        90       100       110


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.6  bits: 21.4 E():  8.6
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (3-59:105-167)

                                    10        20
RF_1_+                   VYIRSRGYPIISIDHLPISSAF-LHS---GDL
                            .:. :. :  ... .: :.    :::   :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
           80        90       100       110       120       130
```

```
        30        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
          :  .   :  :  .::..:  :   :       :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             140       150       160       170       180       190


gi|159 ENSEKTTMPLW
          200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo   (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.2  bits: 21.4 E():     9
Smith-Waterman score: 58;   33.333% identity (52.381% similar) in 63 aa overlap (3-59:114-176)

                                        10        20
RF_1_+                          VYIRSRGYPIISIDHLPISSAF-LHS---GDL
                                 .: . : . ...  .: :.    :::    :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
             90       100       110       120       130       140


        30        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEY
          :  .   :  :  .::..:  :   :       :  ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             150       160       170       180       190       200


gi|162 ENSEKTTMPLW
          210




80 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:48 2010 done: Fri Apr 30 23:42:49 2010
 Total Scan time:  0.080 Total Display time:  0.010


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_+1: 16  - 95 80 aa - 80 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    28      8:==*=======
  34    37     21:======*======
  36    65     44:==============*=======
  38    92     72:=====================*=======
  40   109    101:===============================*===
  42   119    123:=======================================*
  44   118    136:=====================================         *
  46   130    138:===========================================  *
  48    97    132:================================           *
  50    81    121:=========================                *
  52    94    106:==============================        *
  54    97     91:==========================*==
  56    97     76:========================*=======
  58    75     62:==================*====
  60    57     50:==============*==
  62    40     40:============*
  64    32     32:=========*
  66    16     25:======   *
  68    17     20:======*
  70    14     16:=====*
  72    15     12:===*=
  74     3     10:=   *
  76     1      7:= *
  78     0      6: *
  80     8      4:=*=
  82     2      3:*
  84     3      3:*
  86     6      2:*=
  88     1      2:*                inset = represents 1 library sequences
  90     6      1:*=
  92     0      1:*          :*
  94     2      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))= 5.9996+/-0.00358; mu= 2.2424+/- 0.183
 mean_var=43.5343+/-11.755, 0's: 2 Z-trim: 2  B-trim: 125 in 1/42
 Lambda= 0.194383
 Kolmogorov-Smirnov  statistic: 0.0607 (N=28) at  40

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:  0.080
The best scores are:                                     opt bits E(1471)
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 23.5    2.2
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.5    2.6
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 24.1    3.1
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   61 22.3      5
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   61 22.3      5
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   61 22.3      5
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.1      5
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.1      5
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.1      5
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.3      6
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.0    6.4
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.6    6.6
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.0    7.5
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 22.5    7.9
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.5    8.1
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.5    8.5


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]    (214 aa)
 initn: 38 init1: 38 opt:  65  Z-score: 96.3  bits: 23.5 E():  2.2
Smith-Waterman score: 65;  33.333% identity (52.381% similar) in 63 aa overlap (2-58:114-176)

                                    10        20
RF_1_+                     YIRSRGYPIISIDHLPISSAF-LHS---G-D
                           .:. : ... .: :.    :::   : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
              90       100       110       120       130       140

           30        40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
         . : : .::..: : :     : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
            150       160       170       180       190       200

gi|162 ENSEKTTISLW
          210
```

```
>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score:  94.9  bits: 22.5 E():  2.6
Smith-Waterman score: 61;  33.333% identity (52.381% similar) in 63 aa overlap (2-58:29-91)

                                      10        20
RF_1_+                      YIRSRGYPIISIDHLPISSAF-LHS---GDLAC
                            .:  :. ... .: :.   :::    :  :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
                10        20        30        40        50        60

           30        40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
         .   : :  .::..: :  :       :  ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
              70        80        90       100       110       120

gi|159 SEKTTMPLW


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly  (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score:  93.7  bits: 24.1 E():  3.1
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (7-65:301-356)

                                      10        20        30
RF_1_+                      YIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                            ::  .:    . .   . : .:..     ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
                  280       290       300       310       320

              40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
        :.:  :    ::: .  . . .::.   :  .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          330       340       350       360       370       380


>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta  (224 aa)
 initn:  42 init1:  42 opt:  61  Z-score:  89.8  bits: 22.3 E():    5
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (15-76:152-210)

                                      10        20        30        40
RF_1_+                      YIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                            ::. ....:.      :   :  . : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
                 130       140       150       160       170

              50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
        :::  ..  .  .. : : .:.    : :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
```

```
          180         190         200         210         220


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]      (224 aa)
 initn:  42 init1:  42 opt:  61  Z-score: 89.8  bits: 22.3 E():    5
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (15-76:152-210)


                                 10        20        30        40
RF_1_+                   YIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                                 ::. ....:.        :  . : : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
              130       140       150       160       170


             50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
         :::  ..  .   ..  : :  .:.    : :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
          180       190       200       210       220


>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]      (224 aa)
 initn:  42 init1:  42 opt:  61  Z-score: 89.8  bits: 22.3 E():    5
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (15-76:152-210)


                                 10        20        30        40
RF_1_+                   YIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                                 ::. ....:.        :  . : : :.
gi|459 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
              130       140       150       160       170


             50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
         :::  ..  .   ..  : :  .:.    : :
gi|459 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
          180       190       200       210       220


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.8  bits: 23.1 E():    5
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (7-65:216-271)


                                 10        20        30
RF_1_+                   YIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                 :: .:  .  .  . : .:.     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
            190       200       210       220       230       240


             40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
         :.: :   :::. .  . .::.  . .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300
```

```
>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.8  bits: 23.1 E():    5
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (7-65:216-271)

                                 10        20        30
RF_1_+                    YIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             ::  .:   .   .   :  .:.    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240

          40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
         :.:   :    :::  .   . .::.  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
             250       260       270       280       290       300


>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.8  bits: 23.1 E():    5
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (7-65:216-271)

                                 10        20        30
RF_1_+                    YIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             ::  .:   .   .   :  .:.    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240

          40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
         :.:   :    :::  .   . .::.  .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
             250       260       270       280       290       300


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 88.4  bits: 22.3 E():    6
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (17-40:13-35)

               10        20        30        40        50        60
RF_1_+ YIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLV
                      .. .:: ::   :.::. :....
gi|898      MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKATWYGKPT
                    10        20        30        40        50

               70        80
RF_1_+ RPARTIELQSTAPYLEAEYN

gi|898 GAGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVECSGKPVVV
             60        70        80        90       100       110
```

>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly (455 aa)
 initn: 36 init1: 36 opt: 64 Z-score: 87.9 bits: 23.0 E(): 6.4
Smith-Waterman score: 64; 28.333% identity (55.000% similar) in 60 aa overlap (7-65:292-347)

                                       10   20   30
RF_1_+   YIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                               :: .:   .   .   . : .:..   ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
             270   280   290   300   310

              40   50   60   70   80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
         :.:   :   ::: .   . .:. .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         320   330   340   350   360   370

>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru (172 aa)
 initn: 37 init1: 37 opt: 58 Z-score: 87.7 bits: 21.6 E(): 6.6
Smith-Waterman score: 58; 33.333% identity (52.381% similar) in 63 aa overlap (2-58:72-134)

                                       10   20
RF_1_+   YIRSRGYPIISIDHLPISSAF-LHS---GDL
                                 .: : . ... .: :.   :::   :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
                 50   60   70   80   90   100

          30   40   50   60   70   80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
         : .   : : .:.. : :   :   : : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
                 110   120   130   140   150   160

gi|159 ENSEKTTMPLW
                 170

>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp (269 aa)
 initn: 54 init1: 54 opt: 60 Z-score: 86.7 bits: 22.0 E(): 7.5
Smith-Waterman score: 60; 33.333% identity (70.833% similar) in 24 aa overlap (17-40:19-41)

                 10   20   30   40   50
RF_1_+   YIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPS
                       .. :::   :   :.::.. :....
gi|115 MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDAKATWYGK
                 10   20   30   40   50

          60   70   80
RF_1_+ LVRPARTIELQSTAPYLEAEYN

gi|115 PTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAECSGKPV

```
                60        70        80        90       100       110

>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 86.3  bits: 22.5  E():  7.9
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (11-62:13-63)

                       10        20        30        40        50
RF_1_+    YIRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPS
               :.   ::   . :   : : .   . :.  ::..  :    ::..    . :.  .
gi|212 MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAGMTTVVRD
              10        20        30        40        50

          60        70        80
RF_1_+ LVRPARTIELQSTAPYLEAEYN
        :  ::
gi|212 LDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSDEVKRRFY
          60        70        80        90       100       110


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 86.1  bits: 21.5  E():  8.1
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (2-58:105-167)

                                      10        20
RF_1_+                        YIRSRGYPIISIDHLPISSAF-LHS---GDL
                              .:. .:. ... .: :.    :::     :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
             80        90       100       110       120       130

          30        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
        :  .  : :  .::..: :   :       :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
             140       150       160       170       180       190

gi|159 ENSEKTTMPLW
           200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.7  bits: 21.5  E():  8.5
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (2-58:114-176)

                                      10        20
RF_1_+                        YIRSRGYPIISIDHLPISSAF-LHS---GDL
                              .:. .:. ... .: :.    :::     :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
             90       100       110       120       130       140

          30        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYN
```

```
              : .    :  :   .::..:  :   :       :   ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
              150       160       170       180       190       200


gi|162 ENSEKTTMPLW
          210




80 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:49 2010 done: Fri Apr 30 23:42:49 2010
 Total Scan time:  0.080 Total Display time:  0.010


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_+1: 17  - 96 80 aa - 80 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library


       opt      E()
< 20     2    0:=
  22     0    0:                  one = represents 2 library sequences
  24     0    0:
  26     0    0:
  28     2    0:=
  30     3    2:*=
  32    21    8:===*=======
  34    34   21:=========*======
  36    66   44:====================*===========
  38   101   72:===================================*==============
  40   111  101:====================================================*=====
  42   115  123:======================================================== *
  44   120  136:========================================================*
  46   119  138:========================================================*
  48    97  132:================================================        *
  50    90  121:=========================================               *
  52    90  106:=========================================            *
  54   102   91:========================================*=====
  56    98   76:===================================*==========
  58    73   62:==========================*=====
```

```
  60    61    50:=====================*======
  62    40    40:=================*
  64    31    32:==============*
  66    18    25:========    *
  68    15    20:======= *
  70    15    16:======*
  72    15    12:====*==
  74     2    10:=    *
  76     1     7:=   *
  78     0     6:  *
  80     8     4:=*==
  82     2     3:=*
  84     4     3:=*
  86     5     2:*==
  88     2     2:*
  90     5     1:*==
  92     0     1:*
  94     2     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))= 6.0663+/-0.00362; mu= 1.9148+/- 0.184
 mean_var=43.8374+/-11.714, 0's: 2 Z-trim: 2  B-trim: 125 in 1/42
 Lambda= 0.193710
 Kolmogorov-Smirnov  statistic: 0.0614 (N=28) at  40

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:  0.090
The best scores are:                              opt bits E(1471)
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   65 23.4    2.2
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   61 22.4    2.7
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K  ( 473)   68 24.1    3.2
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   61 22.3    5.1
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   61 22.3    5.1
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   61 22.3    5.1
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.2
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.2
```

```
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.2
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.3    6.2
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.0    6.6
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   58 21.5    6.7
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.0    7.7
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 22.4    8.1
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t ( 205)   58 21.5    8.2
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor ( 214)   58 21.5    8.7


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn:  38 init1:  38 opt:  65  Z-score:  96.1  bits: 23.4 E():  2.2
Smith-Waterman score: 65;   33.333% identity (52.381% similar) in 63 aa overlap (1-57:114-176)


                                  10        20
RF_1_+                    IRSRGYPIISIDHLPISSAF-LHS---G-D
                          .: . : . ... .: :.   :::   : :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGID
            90        100       110       120       130       140


           30        40        50        60        70        80
RF_1_+ LACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
         . : : .::..: :  :      : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
           150       160       170       180       190       200


gi|162 ENSEKTTISLW
          210


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  61  Z-score:  94.7  bits: 22.4 E():  2.7
Smith-Waterman score: 61;   33.333% identity (52.381% similar) in 63 aa overlap (1-57:29-91)


                                  10        20
RF_1_+                    IRSRGYPIISIDHLPISSAF-LHS---GDLAC
                          .: . : . ... .: :.   :::   : :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
            10        20        30        40        50        60


     30        40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
       . : : .::..: :  :      : ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
           70        80        90       100       110       120


gi|159 SEKTTMPLW


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly  (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score:  93.4  bits: 24.1 E():  3.2
```

Smith-Waterman score: 68; 30.000% identity (55.000% similar) in 60 aa overlap (6-64:301-356)

```
                                10        20        30
RF_1_+                    IRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             ::  .:     .  .   .  : .:..    ::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
              280       290       300       310       320

              40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
         :.:   :    :::  .   .  .:.  :   .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
          330       340       350       360       370       380
```

>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta (224 aa)
 initn: 42 init1: 42 opt: 61 Z-score: 89.7 bits: 22.3 E(): 5.1
Smith-Waterman score: 61; 24.194% identity (50.000% similar) in 62 aa overlap (14-75:152-210)

```
                                10        20        30        40
RF_1_+                    IRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                             ::.  ....:.         :   .  : : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
             130       140       150       160       170

              50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
         :::   ..   .   ..   :  :  .  .:.    :  :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
          180       190       200       210       220
```

>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]   (224 aa)
 initn: 42 init1: 42 opt: 61 Z-score: 89.7 bits: 22.3 E(): 5.1
Smith-Waterman score: 61; 24.194% identity (50.000% similar) in 62 aa overlap (14-75:152-210)

```
                                10        20        30        40
RF_1_+                    IRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                             ::.  ....:.         :   .  : : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
             130       140       150       160       170

              50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
         :::   ..   .   ..   :  :  .  .:.    :  :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
          180       190       200       210       220
```

>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]   (224 aa)
 initn: 42 init1: 42 opt: 61 Z-score: 89.7 bits: 22.3 E(): 5.1
Smith-Waterman score: 61; 24.194% identity (50.000% similar) in 62 aa overlap (14-75:152-210)

```
                                 10        20        30        40
RF_1_+                   IRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                                ::. ....:.      :  . : : :.
gi|459 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
            130       140       150       160       170


              50        60        70        80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
         :::  ..  .  ..  : : . .:.   : :
gi|459 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
        180       190       200       210       220


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.5  bits: 23.0 E():  5.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (6-64:216-271)

                                      10        20        30
RF_1_+                       IRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                :: .:   .  .  . : .:.     ::.:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240


              40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
         :.:  :   ::: .  . . .:. .  . .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.5  bits: 23.0 E():  5.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (6-64:216-271)

                                      10        20        30
RF_1_+                       IRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                                :: .:   .  .  . : .:.     ::.:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240


              40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
         :.:  :   ::: .  . . .:. .  . .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
            250       260       270       280       290       300


>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.5  bits: 23.0 E():  5.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (6-64:216-271)
```

```
                                      10        20        30
RF_1_+                       IRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             ::  .:    .  .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190       200       210       220       230       240


               40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
        :.:   :   :::  .   .  .:..   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
           250       260       270       280       290       300


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 88.2  bits: 22.3 E():  6.2
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (16-39:13-35)


                 10        20        30        40        50        60
RF_1_+ IRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVR
                    .. .:: ::    :.::.. :....
gi|898     MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKATWYGKPTG
               10        20        30        40        50


               70        80
RF_1_+ PARTIELQSTAPYLEAEYNE

gi|898 AGPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVECSGKPVVVH
           60        70        80        90       100       110


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 87.7  bits: 23.0 E():  6.6
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (6-64:292-347)


                                      10        20        30
RF_1_+                       IRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             ::  .:    .  .   . : .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           270       280       290       300       310


               40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
        :.:   :   :::  .   .  .:..   .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         320       330       340       350       360       370


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 87.5  bits: 21.5 E():  6.7
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (1-57:72-134)

                                      10        20
```

```
RF_1_+                                 IRSRGYPIISIDHLPISSAF-LHS---GDL
                                       .: . : . ... .: :.    :::    :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
               50        60        70        80        90       100


          30        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
       : .   : :  .::..: :   :      :  ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
              110       120       130       140       150       160


gi|159 ENSEKTTMPLW
              170


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 86.5  bits: 22.0 E():  7.7
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (16-39:19-41)


                10        20        30        40        50
RF_1_+     IRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPS
                          .. :::   :   :.::.. :....
gi|115 MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDAKATWYGK
               10        20        30        40        50


          60        70        80
RF_1_+ LVRPARTIELQSTAPYLEAEYNE

gi|115 PTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAECSGKPV
           60        70        80        90       100       110


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 86.0  bits: 22.4 E():  8.1
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (10-61:13-63)


                10        20        30        40        50
RF_1_+     IRSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPS
                   :.   :: . :  : :.  . .: :::. :   ::..   . :.  .
gi|212 MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAGMTTVVRD
               10        20        30        40        50


          60        70        80
RF_1_+ LVRPARTIELQSTAPYLEAEYNE
         : ::
gi|212 LDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSDEVKRRFY
          60        70        80        90       100       110


>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru  (205 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.9  bits: 21.5 E():  8.2
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (1-57:105-167)
```

```
                                        10        20
RF_1_+                           IRSRGYPIISIDHLPISSAF-LHS---GDL
                                 .: .: . ... .: :.   :::   :
gi|159 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
           80        90       100       110       120       130


            30        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
        :.  : : .::..: :  :      : ::::
gi|159 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
          140       150       160       170       180       190


gi|159 ENSEKTTMPLW
          200


>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo  (214 aa)
 initn:  37 init1:  37 opt:  58  Z-score: 85.5  bits: 21.5 E():  8.7
Smith-Waterman score: 58;  33.333% identity (52.381% similar) in 63 aa overlap (1-57:114-176)


                                        10        20
RF_1_+                           IRSRGYPIISIDHLPISSAF-LHS---GDL
                                 .: .: . ... .: :.   :::   :
gi|162 EEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIH
           90       100       110       120       130       140


            30        40        50        60        70        80
RF_1_+ ACNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNE
        :.  : : .::..: :  :      : ::::
gi|162 AQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGS
          150       160       170       180       190       200


gi|162 ENSEKTTMPLW
          210




80 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:50 2010 done: Fri Apr 30 23:42:50 2010
 Total Scan time:  0.090 Total Display time:  0.000

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_+1: 18 - 97 80 aa - 80 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      3     2:*
  32     21     8:==*====
  34     34    21:======*=====
  36     69    44:=============*========
  38    102    72:====================*==========
  40     99   101:==============================*
  42    123   123:=======================================*
  44    115   136:======================================          *
  46    123   138:=======================================      *
  48     98   132:===================================          *
  50     91   121:============================           *
  52     87   106:==========================          *
  54     99    91:=========================*==
  56     97    76:========================*======
  58     76    62:==================*=====
  60     61    50:===============*====
  62     40    40:============*
  64     29    32:=========*
  66     18    25:=====   *
  68     19    20:======*
  70     16    16:=====*
  72     15    12:===*=
  74      2    10:=  *
  76      1     7:= *
  78      0     6: *
  80      8     4:=*=
  82      4     3:*=
  84      4     3:*=
  86      3     2:*
  88      2     2:*           inset = represents 1 library sequences
  90      5     1:*=
  92      2     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     0     0:              *
   331323 residues in  1471 sequences
  Expectation_n fit: rho(ln(x))= 6.0914+/-0.00361; mu= 1.8176+/- 0.184
  mean_var=43.5858+/-11.611,  0's: 2 Z-trim: 2  B-trim: 125 in 1/42
  Lambda= 0.194268
  Kolmogorov-Smirnov  statistic: 0.0570 (N=28) at  38

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:   0.090
The best scores are:                              opt bits E(1471)
gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K ( 473)   68 24.1    3.2
gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur ( 214)   63 22.9    3.3
gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos t ( 129)   59 21.9    3.9
gi|162931|gb|AAA30480.1| beta-casein precursor [Bo ( 224)   61 22.3    5.1
gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]  ( 224)   61 22.3    5.1
gi|459292|gb|AAB29137.1| beta-casein A3 [Bos tauru ( 224)   61 22.3    5.1
gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.2
gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.2
gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen  ( 373)   64 23.0    5.2
gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea m ( 263)   61 22.3    6.1
gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen  ( 455)   64 23.0    6.5
gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full ( 269)   60 22.0    7.6
gi|21215170|gb|AAM43909.1|AF464911_1 large subunit ( 392)   62 22.4    8.1
gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos t ( 172)   56 21.0    9.8


>>gi|12697782|dbj|BAB21619.1| allergen Gly m Bd 28K [Gly  (473 aa)
 initn:  38 init1:  38 opt:  68  Z-score: 93.4  bits: 24.1 E():  3.2
Smith-Waterman score: 68;  30.000% identity (55.000% similar) in 60 aa overlap (5-63:301-356)

                                 10        20        30
RF_1_+                  RSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                              ::  .:   .  .   . : .::.    :::..:
gi|126 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
             280       290       300       310       320

          40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNEG
          :.:   :   :::  .   . .  .::.   :  .:
gi|126 ISDE--YTIVLSGYGELHIGYPNGSRAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
```

```
            330         340         350         360         370         380


>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn:  38 init1:  38 opt:  63  Z-score: 93.2  bits: 22.9 E():  3.3
Smith-Waterman score: 63;  33.871% identity (51.613% similar) in 62 aa overlap (1-56:115-176)


                                            10        20
RF_1_+                             RSRGYPIISIDHLPISSAF-LHS---G-DL
                                   : . : . ... .: :.    :::    : :
gi|162 EIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIDA
            90        100       110       120       130       140


            30        40        50        60        70        80
RF_1_+ ACNPPHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNEG
          . : : .::..: : :      : ::::
gi|162 QQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSE
            150       160       170       180       190       200


gi|162 NSEKTTISLW
            210


>>gi|159793217|gb|ABW98953.1| alpha S1 casein [Bos tauru  (129 aa)
 initn:  37 init1:  37 opt:  59  Z-score: 91.8  bits: 21.9 E():  3.9
Smith-Waterman score: 59;  33.871% identity (51.613% similar) in 62 aa overlap (1-56:30-91)


                                            10        20
RF_1_+                             RSRGYPIISIDHLPISSAF-LHS---GDLAC
                                   : . : . ... .: :.    :::    : :
gi|159 IVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQ
                 10        20        30        40        50        60


          30        40        50        60        70        80
RF_1_+ NP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNEG
         . : : .::..: : :      : ::::
gi|159 QKEPMIGVNQELAYFYPELFXQFYQPDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSEN
              70        80        90        100       110       120


gi|159 SEKTTMPLW


>>gi|162931|gb|AAA30480.1| beta-casein precursor [Bos ta  (224 aa)
 initn:  42 init1:  42 opt:  61  Z-score: 89.7  bits: 22.3 E():  5.1
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (13-74:152-210)


                                      10        20        30        40
RF_1_+                             RSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                                   ::. ....:.      : . : : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
              130       140       150       160       170
```

```
                50          60          70          80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNEG
         :::  ..   .   ..  : : ..:.    : :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
          180         190         200         210         220


>>gi|162805|gb|AAA30431.1| beta-casein [Bos taurus]       (224 aa)
 initn:  42 init1:  42 opt:  61  Z-score: 89.7  bits: 22.3 E():  5.1
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (13-74:152-210)

                             10          20          30          40
RF_1_+                  RSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                           ::. ....:.       :  . : : :.
gi|162 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
               130         140         150         160         170


                50          60          70          80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNEG
         :::  ..   .   ..  : : ..:.    : :
gi|162 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
          180         190         200         210         220


>>gi|459292|gb|AAB29137.1| beta-casein A3 [Bos taurus]      (224 aa)
 initn:  42 init1:  42 opt:  61  Z-score: 89.7  bits: 22.3 E():  5.1
Smith-Waterman score: 61;  24.194% identity (50.000% similar) in 62 aa overlap (13-74:152-210)

                             10          20          30          40
RF_1_+                  RSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYA
                           ::. ....:.       :  . : : :.
gi|459 KEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVL---
               130         140         150         160         170


                50          60          70          80
RF_1_+ SLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNEG
         :::  ..   .   ..  : : ..:.    : :
gi|459 SLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
          180         190         200         210         220


>>gi|187766749|gb|ACD36975.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.6  bits: 23.0 E():  5.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (5-63:216-271)

                             10          20          30
RF_1_+                  RSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                           :: .:   .  .   . : .:.     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
          190         200         210         220         230         240
```

```
           40          50          60          70          80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNEG
          :.:  :    :::.  .   . .::.    .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
           250         260         270         280         290         300


>>gi|187766747|gb|ACD36974.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.6  bits: 23.0  E():  5.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (5-63:216-271)

                                  10          20          30
RF_1_+                     RSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                           :: .:    .  .   . :  .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190         200         210         220         230         240


           40          50          60          70          80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNEG
          :.:  :    :::.  .   . .::.    .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
           250         260         270         280         290         300


>>gi|187766751|gb|ACD36976.1| Gly m Bd 28K allergen [Gly  (373 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 89.6  bits: 23.0  E():  5.2
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (5-63:216-271)

                                  10          20          30
RF_1_+                     RSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                           :: .:    .  .   . :  .:..     ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           190         200         210         220         230         240


           40          50          60          70          80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNEG
          :.:  :    :::.  .   . .::.    .   .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
           250         260         270         280         290         300


>>gi|89892721|gb|ABD79094.1| Zea m 1 allergen [Zea mays]  (263 aa)
 initn:  52 init1:  52 opt:  61  Z-score: 88.2  bits: 22.3  E():  6.1
Smith-Waterman score: 61;  33.333% identity (75.000% similar) in 24 aa overlap (15-38:13-35)

               10          20          30          40          50          60
RF_1_+ RSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPSLVRP
                      .. .:: ::    :.::.. :.....
gi|898    MWSLMQAQVAMVVALSFLVSGAW-CGPPKVPPGKNITATYGKDWLDAKATWYGKPTGA
                  10          20          30          40          50


               70          80
```

```
RF_1_+ ARTIELQSTAPYLEAEYNEG


gi|898 GPDDNGGGCGYKDVNKPPFNSMGACGNIPIFKDGLGCGSCFEIKCDKPVECSGKPVVVHI
           60        70        80        90        100       110


>>gi|187766755|gb|ACD36978.1| Gly m Bd 28K allergen [Gly  (455 aa)
 initn:  36 init1:  36 opt:  64  Z-score: 87.7  bits: 23.0 E():  6.5
Smith-Waterman score: 64;  28.333% identity (55.000% similar) in 60 aa overlap (5-63:292-347)

                                      10        20        30
RF_1_+                        RSRGYPIISIDHLPISSAFLHSGDLACNPPHIDP
                             :: .:   .  .  . : .::.    ::..:
gi|187 GSPASYNLYDDKKADFKNAYGWSKALHGGEYPPLSEPDIGVLLVKLSAGSMLA--PHVNP
           270       280       290       300       310

          40        50        60        70        80
RF_1_+ -SQEVVYYASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNEG
        :.:   :   ::: .  . .::.   . .:
gi|187 ISDE--YTIVLSGYGELHIGYPNGSKAMKTKIKQGDVFVVPRYFPFCQVASRDGPLEFFG
         320       330       340       350       360       370


>>gi|115502168|sp|P0C1Y5.1|EXB11_MAIZE RecName: Full=Exp  (269 aa)
 initn:  54 init1:  54 opt:  60  Z-score: 86.5  bits: 22.0 E():  7.6
Smith-Waterman score: 60;  33.333% identity (70.833% similar) in 24 aa overlap (15-38:19-41)


                  10        20        30        40        50
RF_1_+     RSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPS
                    .. :::  :   :.:::.. :....
gi|115 MTVVSIMWSLVQVQVLVAVALAFL-VGGAWCGPPKVPPGKNITAKYGSDWLDAKATWYGK
                  10        20        30        40        50

          60        70        80
RF_1_+ LVRPARTIELQSTAPYLEAEYNEG


gi|115 PTGAGPDDNGGGCGYKDVNKAPFNSMGACGNVPIFKDGLGCGSCFEIKCDKPAECSGKPV
           60        70        80        90        100       110


>>gi|21215170|gb|AAM43909.1|AF464911_1 large subunit rib  (392 aa)
 initn:  48 init1:  48 opt:  62  Z-score: 86.1  bits: 22.4 E():  8.1
Smith-Waterman score: 62;  30.769% identity (53.846% similar) in 52 aa overlap (9-60:13-63)


                  10        20        30        40        50
RF_1_+     RSRGYPIISIDHLPISSAFLHSGDLACNPPHIDPSQEVVYYASLSGPNLQKTAYPS
                 :. :: . :  : :. . .: .::.. :   ::.    . :.  .
gi|212 MSHRKYEAPRHGSLAFLPRKRAARHRGKVK-SFPKYDPKKPVHLTASMGYKAGMTTVVRD
                  10        20        30        40        50

          60        70        80
RF_1_+ LVRPARTIELQSTAPYLEAEYNEG
```

```
             :  ::
gi|212 LDRPGAKMHKKEIVEAVTIIETPPLVAVGVVGYIETPRGLRSLTTVWAEHLSDEVKRRFY
          60        70        80        90       100       110


>>gi|159793201|gb|ABW98945.1| alpha S1 casein [Bos tauru  (172 aa)
 initn:  37 init1:  37 opt:  56  Z-score: 84.6  bits: 21.0 E():  9.8
Smith-Waterman score: 56;  33.871% identity (51.613% similar) in 62 aa overlap (1-56:73-134)


                                          10        20
RF_1_+                         RSRGYPIISIDHLPISSAF-LHS---GDLA
                               : . : . ... .: :.    :::    :  :
gi|159 EIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHA
                 50        60        70        80        90       100


           30        40        50        60        70        80
RF_1_+ CNP-PHIDPSQEVVY-YASLSGPNLQKTAYPSLVRPARTIELQSTAPYLEAEYNEG
         .   : :  .::..: :   :     :  ::::
gi|159 QQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSE
               110       120       130       140       150       160


gi|159 NSEKTTMPLW
           170




80 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:42:51 2010 done: Fri Apr 30 23:42:51 2010
 Total Scan time:  0.090 Total Display time:  0.000
```

RF_1_-1

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_-1 69 aa - 69 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     3     0:=
  30     7     2:*==
  32    21     8:==*====
  34    37    21:=====*======
  36    50    44:=============*==
  38    69    72:====================*
  40    80   101:=======================            *
  42   103   123:===============================         *
  44   106   136:==================================           *
  46   176   138:====================================================*============
  48   162   132:==============================================*=========
  50    99   121:=============================            *
  52    96   106:=============================         *
  54    96    91:==========================*=
  56    77    76:======================*
  58    43    62:==============        *
  60    60    50:==============*===
  62    46    40:===========*==
  64    22    32:=======   *
  66    38    25:=======*====
  68    12    20:====   *
  70    15    16:=====*
  72    18    12:===*==
  74     3    10:=   *
  76     7     7:==*
  78     9     6:=*=
  80     3     4:=*
  82     1     3:*
  84     4     3:*=
  86     0     2:*
  88     2     2:*           inset = represents 1 library sequences
  90     1     1:*
  92     1     1:*         :*
  94     0     1:*         :*
  96     1     1:*         :*
  98     0     0:          *
 100     0     0:          *
 102     1     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))= 3.1174+/-0.00309; mu= 13.0294+/- 0.160
 mean_var=25.7958+/- 6.263, 0's: 2 Z-trim: 2  B-trim: 0 in 0/43
 Lambda= 0.252523
 Kolmogorov-Smirnov  statistic: 0.0292 (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:  0.070
The best scores are:                              opt bits E(1471)
gi|62240390|gb|AAX77383.1| 11S globulin precursor ( 510)   59 25.6      1
gi|62240392|gb|AAX77384.1| 11S globulin precursor ( 523)   56 24.5     2.2
gi|169929|gb|AAB01374.1| beta-conglycinin storage ( 639)   55 24.2     3.3
gi|25991543|gb|AAN76862.1|AF453947_1 allergen Ana ( 457)   53 23.4     4.2
gi|18635|emb|CAA33215.1| glycinin subunit G1 [Glyc ( 495)   52 23.0     5.8
gi|18615|emb|CAA26723.1| unnamed protein product [ ( 495)   52 23.0     5.8
gi|22595342|gb|AAN02510.1|AF409110_1 serine protea ( 244)   48 21.4     9.1
gi|9087167|sp|Q9SCG9.1|MPAC1_CUPAR RecName: Full=M ( 346)   49 21.8     9.4
gi|18639|emb|CAA33217.1| glycinin subunit G3 [Glyc ( 481)   50 22.3     9.4
gi|19069497|emb|CAC37790.2| putative allergen Cup  ( 367)   49 21.9     9.8


>>gi|62240390|gb|AAX77383.1| 11S globulin precursor [Sin  (510 aa)
 initn:  41 init1:  41 opt:  59  Z-score: 102.2  bits: 25.6 E():    1
Smith-Waterman score: 59;  26.087% identity (56.522% similar) in 46 aa overlap (24-68:335-380)

                      10        20        30        40        50
RF_1_-           IVANLHCIPLQGTVLLIVTQWFVLDARERDRRFSASLGRLERRNTLLLGMDLY
                            .:   :   .. .:::.   :.  : .  :
gi|622 EQWRHPRGPPQSPQDNGLEETICSMRTHENIDDPARADIYKPNLGRVTSVNSYTLPILQY
           310       320       330       340       350       360


              60
RF_1_- V-VGYKRGLLNGEKLMR
       . ..  ::.:.:  ..
gi|622 IRLSATRGILQGSAMVLPKYNMNANEILYCTQGQARIQVVNDNGQNVLDQQVQKGQLVVI
           370       380       390       400       410       420


>>gi|62240392|gb|AAX77384.1| 11S globulin precursor [Sin  (523 aa)
 initn:  36 init1:  36 opt:  56  Z-score: 96.2  bits: 24.5 E(): 2.2
Smith-Waterman score: 56;  23.913% identity (56.522% similar) in 46 aa overlap (24-68:348-393)

                      10        20        30        40        50
RF_1_-           IVANLHCIPLQGTVLLIVTQWFVLDARERDRRFSASLGRLERRNTLLLGMDLY
                            .:   :   .. .:::.   :.  . . :
gi|622 EQWRHPRGPPQSPQDNGLEETICSMRTHENIDDPARADIYKPNLGRVTSVNSYTIPILQY
           320       330       340       350       360       370


              60
RF_1_- V-VGYKRGLLNGEKLMR
       . ..  ::.:.:  ..
```

```
gi|622  IRLSATRGILQGSAMVLPKYNMNANEILYCTQGQARIQVVNDNGQNVLDQQVQKGQLVVI
             380       390       400       410       420       430


>>gi|169929|gb|AAB01374.1| beta-conglycinin storage prot  (639 aa)
 initn:  37 init1:  37 opt:  55  Z-score: 93.0  bits: 24.2 E():  3.3
Smith-Waterman score: 55;  35.897% identity (61.538% similar) in 39 aa overlap (29-62:446-484)

                         10        20        30        40        50
RF_1_-     IVANLHCIPLQGTVLLIVTQWFVLDARERDRRFSASLGRL---ERRNTLLLGMDLY--
                             ::  .: .::.:     .:: :  .:..
gi|169 KKQIRELSKHAKSSSRKTISSEDKPFNLGSRDPIYSNKLGKLFEITQRNPQLRDLDVFLS
             420       430       440       450       460       470


               60
RF_1_- VVGYKRGLLNGEKLMR
        :: ...: :
gi|169 VVDMNEGALFLPHFNSKAIVVLVINEGEANIELVGIKEQQQRQQQEEQPLEVRKYRAELS
             480       490       500       510       520       530


>>gi|25991543|gb|AAN76862.1|AF453947_1 allergen Ana o 2   (457 aa)
 initn:  53 init1:  53 opt:  53  Z-score: 91.1  bits: 23.4 E():  4.2
Smith-Waterman score: 53;  42.105% identity (73.684% similar) in 19 aa overlap (15-33:372-390)

                                 10        20        30        40
RF_1_-             IVANLHCIPLQGTVLLIVTQWFVLDARERDRRFSASLGRLERRN
                                 .:.: : :..   : :..::
gi|259 IIYGCKGKGQVQVVDNFGNRVFDGEVREGQMLVVPQNFAVVKRAREERFEWISFKTNDRA
                 350       360       370       380       390       400


                   50        60
RF_1_- TLLLGMDLYVVGYKRGLLNGEKLMR

gi|259 MTSPLAGRTSVLGGMPEEVLANAFQISREDARKIKFNNQQTTLTSGESSHHMRDDA
                 410       420       430       440       450


>>gi|18635|emb|CAA33215.1| glycinin subunit G1 [Glycine   (495 aa)
 initn:  51 init1:  51 opt:  52  Z-score: 88.6  bits: 23.0 E():  5.8
Smith-Waterman score: 52;  29.508% identity (49.180% similar) in 61 aa overlap (10-61:406-466)

                             10        20        30
RF_1_-             IVANLHCIPLQGTVLLIVTQWFVLDARERDRRF------
                             ::   .::: : ::. :: .. :
gi|186 LNANSIIYALNGRALIQVVNCNGERVFDGELQEGRVLIVPQNFVVAARSQSDNFEYVSFK
             380       390       400       410       420       430


                   40        50        60
RF_1_- ---SASLGRLERRNTLLLGMDLYVVGYKRGLLNGEKLMR
          .  .: :   :.:: ..   :. .  .:
gi|186 TNDTPMIGTLAGANSLLNALPEEVIQHTFNLKSQQARQIKNNNPFKFLVPPQESQKRAVA
```

```
              440         450         460         470         480         490


>>gi|18615|emb|CAA26723.1| unnamed protein product [Glyc   (495 aa)
 initn:  51 init1:  51 opt:  52  Z-score: 88.6  bits: 23.0 E():  5.8
Smith-Waterman score: 52;  29.508% identity (49.180% similar) in 61 aa overlap (10-61:406-466)


                                    10        20        30
RF_1_-                    IVANLHCIPLQGTVLLIVTQWFVLDARERDRRF------
                                  ::    .::: : ::. :: ..   :
gi|186 LNANSIIYALNGRALIQVVNCNGERVFDGELQEGRVLIVPQNFVVAARSQSDNFEYVSFK
            380       390       400       410       420       430


                   40        50        60
RF_1_- ---SASLGRLERRNTLLLGMDLYVVGYKRGLLNGEKLMR
          .   .: :    :.:: ..    :.  .   .:
gi|186 TNDTPMIGTLAGANSLLNALPEEVIQHTFNLKSQQARQIKNNNPFKFLVPPQESQKRAVA
            440       450       460       470       480       490


>>gi|22595342|gb|AAN02510.1|AF409110_1 serine protease [   (244 aa)
 initn:  39 init1:  39 opt:  48  Z-score: 85.1  bits: 21.4 E():  9.1
Smith-Waterman score: 48;  24.615% identity (50.769% similar) in 65 aa overlap (1-65:59-117)


                                    10        20        30
RF_1_-                    IVANLHCIPLQGTVLLIVTQWFVLDARERD
                                 :..   ::.  : .  :: ..   ::  .
gi|225 SNASPGDAVYQIALQSASHFCGGSILDEYWILTAAHCVDGQTVSKLIRSK--VLGEKISV
            30        40        50        60        70        80


                   40        50        60
RF_1_- RRFSASLGRLERRNTLLLGMDLYVVGYKRGLLNGEKLMR
          .. :      :. .. :: :. ..   :    ::..
gi|225 SKIFAH----EKYDSRLLDNDIALIKLKSPRLNSKNARVLPGSDVVKDGQVQSVWGYLEE
            90       100       110       120       130       140


gi|225 GSYSLPPELRRVDIGGASRKECNELYSKVNAEVTDNMICGGDVANGGKDSCQGDSGGPLV
               150       160       170       180       190       200


>>gi|9087167|sp|Q9SCG9.1|MPAC1_CUPAR RecName: Full=Major   (346 aa)
 initn:  49 init1:  49 opt:  49  Z-score: 84.9  bits: 21.8 E():  9.4
Smith-Waterman score: 49;  72.727% identity (81.818% similar) in 11 aa overlap (48-58:80-90)


           20        30        40        50        60
RF_1_- VTQWFVLDARERDRRFSASLGRLERRNTLLLGMDLYVVGYKRGLLNGEKLMR
                            : :  :::.:::
gi|908 PVNPTPGTLRYGATREKALWIIFSQNMNIKLQMPLYVAGYKTIDGRGAVVHLGNGGPCLF
         50        60        70        80        90       100


gi|908 MRKASHVILHGLHIHGCNTSVLGDVLVSESIGVEPVHAQDGDAITMRNVTNAWIDHNSLS
          110       120       130       140       150       160
```

>>gi|18639|emb|CAA33217.1| glycinin subunit G3 [Glycine   (481 aa)
 initn: 46 init1: 46 opt: 50 Z-score: 84.9 bits: 22.3 E(): 9.4
Smith-Waterman score: 50; 30.909% identity (49.091% similar) in 55 aa overlap (10-55:392-446)

```
                              10        20        30
RF_1_-                 IVANLHCIPLQGTVLLIVTQWFVLDARERDRRFS-----
                              ::   .::: : :.. :: ..  :
gi|186 LNANSIIYALNGRALVQVVNCNGERVFDGELQEGQVLIVPQNFAVAARSQSDNFEYVSFK
             370       380       390       400       410       420


                40        50        60
RF_1_- ----ASLGRLERRNTLLLGMDLYVVGYKRGLLNGEKLMR
           :.: :   :.:: ..   :.
gi|186 TNDRPSIGNLAGANSLLNALPEEVIQQTFNLRRQQARQVKNNNPFSFLVPPKESQRRVVA
            430       440       450       460       470       480
```

>>gi|19069497|emb|CAC37790.2| putative allergen Cup a 1   (367 aa)
 initn: 49 init1: 49 opt: 49 Z-score: 84.6 bits: 21.9 E(): 9.8
Smith-Waterman score: 49; 72.727% identity (81.818% similar) in 11 aa overlap (48-58:101-111)

```
           20        30        40        50        60
RF_1_- VTQWFVLDARERDRRFSASLGRLERRNTLLLGMDLYVVGYKRGLLNGEKLMR
                              : : :::.:::
gi|190 PVNPTPGTLRYGATREKALWIIFSQNMNIKLQMPLYVAGYKTIDGRGADVHLGNGGPCLF
                  80        90       100       110       120       130


gi|190 MRTASHVILHGLHIHGCNTSVLGDVLVSESIGVEPVHAQDGDAITMRNVTNAWIDHNSLS
                 140       150       160       170       180       190
```


69 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:46:11 2010 done: Fri Apr 30 23:46:11 2010
 Total Scan time: 0.070 Total Display time: 0.000

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 63 aa - 63 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      0    0:
  30      0    2:*
  32      2    8:= *
  34      5   21:==     *
  36     22   44:=======        *
  38     58   72:==================    *
  40     87  101:==========================     *
  42    109  123:===================================    *
  44    120  136:======================================       *
  46    148  138:================================================*====
  48    173  132:============================================*==============
  50    112  121:=====================================  *
  52    112  106:=================================*==
  54    116   91:=============================*=======
  56     84   76:========================*==
  58     64   62:===================*=
  60     33   50:==========      *
  62     47   40:============*==
  64     45   32:==========*====
  66     23   25:=======*
  68     17   20:======*
  70     23   16:=====*==
  72     15   12:===*=
  74      5   10:== *
  76     10    7:==*=
  78      9    6:=*=
  80      6    4:=*
  82      7    3:*==
  84      1    3:*
  86      5    2:*=
  88      2    2:*              inset = represents 1 library sequences
  90      2    1:*
  92      2    1:*          :*=
  94      3    1:*          :*==
  96      1    1:*          :*
  98      1    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))= 5.1702+/-0.00378; mu= 3.2801+/- 0.194
 mean_var=50.9101+/-13.298, 0's: 2 Z-trim: 2  B-trim: 3 in 1/42
  Lambda= 0.179751
  Kolmogorov-Smirnov  statistic: 0.0705 (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:  0.060
The best scores are:                              opt bits E(1471)
gi|42559514|sp|Q967Z0.1|MYSP_DERFA RecName: Full=P ( 692)   71 25.0    1.9
gi|21954740|gb|AAM83103.1| paramyosin allergen [Bl ( 875)   72 25.3    1.9
gi|2735118|gb|AAB93839.1| ABA-1 allergen [Ascaris  ( 134)   61 22.3    2.4
gi|2735114|gb|AAB93837.1| ABA-1 allergen [Ascaris  ( 134)   61 22.3    2.4
gi|37778944|gb|AAO73464.1| HDM allergen [Dermatoph ( 875)   70 24.8    2.8
gi|2735108|gb|AAD13650.1| ABA-1 allergen [Ascaris  ( 267)   63 22.9    3.2
gi|21673|emb|CAA35238.1| unnamed protein product [ ( 307)   63 22.9    3.7
gi|2735112|gb|AAD13652.1| ABA-1 allergen [Ascaris  ( 267)   61 22.3    4.6
gi|2735110|gb|AAD13651.1| ABA-1 allergen [Ascaris  ( 267)   61 22.3    4.6
gi|75317968|sp|O22116|O22116_WHEAT LMM glutenin 3  ( 373)   61 22.4    6.3
gi|75107016|sp|P80208.1|2SS3_BRANA RecName: Full=N ( 125)   55 20.7    6.7
gi|3703107|gb|AAC63045.1| glycinin [Arachis hypoga ( 507)   62 22.7     7
gi|224036293|pdb|3C3V|A Chain A, Crystal Structure ( 510)   62 22.7    7.1
gi|199732457|gb|ACH91862.1| arachin Arah3 isoform  ( 530)   62 22.7    7.3
gi|1304264|dbj|BAA12318.1| alpha-gliadin [Triticum ( 259)   58 21.6    7.7
gi|119390336|pdb|2J23|B Chain B, Cross-Reactivity  ( 121)   54 20.5    7.8
gi|110349081|gb|ABG73108.1| Pis v 1 allergen 2S al ( 149)   54 20.5    9.5


>>gi|42559514|sp|Q967Z0.1|MYSP_DERFA RecName: Full=Param  (692 aa)
 initn:  65 init1:  65 opt:  71  Z-score: 97.5  bits: 25.0 E():  1.9
Smith-Waterman score: 71;  27.119% identity (59.322% similar) in 59 aa overlap (1-57:308-366)

                                    10        20        30
RF_1_+                    AAQTCRKPPIPLSCVQHEPLSYNQQHRTLK
                          :  .  :     :. .::: .   .:.   :
gi|425 QQLEKRVAQLEKINLDLKNKLEEVTMLMEQAQKELRVKIAELQKLQHEYEKLRDQRDQLA
          280       290       300       310       320       330

                     40        50        60
RF_1_+  RNTMKVSYDLQQSQNTMN--HKVIEARNIRRNKYF
        :.. :.. :: .... .: :. :. ..:
gi|425 RENKKLTDDLAEAKSQLNDAHRRIHEQEIEIKRLENERDELSAAYKEAETLRKQEEAKNQ
          340       350       360       370       380       390
```

```
>>gi|21954740|gb|AAM83103.1| paramyosin allergen [Blomia  (875 aa)
 initn:  66 init1:  66 opt:  72  Z-score: 97.2  bits: 25.3 E():  1.9
Smith-Waterman score: 72;  29.167% identity (70.833% similar) in 48 aa overlap (12-57:405-452)

                              10        20        30        40
RF_1_+                AAQTCRKPPIPLSCVQHEPLSYNQQHRTLKRNTMKVSYDLQ
                              :. .::: .  .:..::.:.. :.. ::
gi|219 KINLDLKSKLEEVSMLLEQTQKDLRVKIADLQKLQHEYEKLRDQKEALARENKKLADDLA
          380       390       400       410       420       430


                       50        60
RF_1_+ QSQNTMN--HKVIEARNIRRNKYF
        .... .: :. :. ..:
gi|219 EAKSQLNDAHRRIHEQEIEIKRLENEREELAAAYKEAETLRKQEEAKNQRLTAELAQTRH
          440       450       460       470       480       490


>>gi|2735118|gb|AAB93839.1| ABA-1 allergen [Ascaris lumb  (134 aa)
 initn:  38 init1:  38 opt:  61  Z-score: 95.4  bits: 22.3 E():  2.4
Smith-Waterman score: 61;  29.167% identity (60.417% similar) in 48 aa overlap (15-58:35-82)

                              10        20        30        40
RF_1_+                  AAQTCRKPPIPLSCVQHEPLS-YNQQHRTLKRNT---MKVSYDL
                              .::: .  :.: :   :...   .::.  .
gi|273 HYLKTYLSWLTEEQKEKLKEMKEAGKTKAEIQHEVMHFYDQLHGEEKQQATEKLKVGCKM
            10        20        30        40        50        60


                     50        60
RF_1_+ QQSQNTMNHKVIEARNIRRNKYF
        .     ..::.: ::..
gi|273 LLKGVIGEEKVVELRNMKEAGADIQELQQKVEKMLSEVTDEKQKEKVHEYGPACKKIFGA
          70        80        90       100       110       120


>>gi|2735114|gb|AAB93837.1| ABA-1 allergen [Ascaris lumb  (134 aa)
 initn:  39 init1:  39 opt:  61  Z-score: 95.4  bits: 22.3 E():  2.4
Smith-Waterman score: 61;  29.167% identity (60.417% similar) in 48 aa overlap (15-58:35-82)

                              10        20        30        40
RF_1_+                  AAQTCRKPPIPLSCVQHEPLSY-NQQHRTLKRNT---MKVSYDL
                              .::: .  : .: :   :...   .::.   .
gi|273 HYLKTYLSWLTEEQKEKLKEMKEAGKTKAEIQHEVMRYYDQLHGEEKQQATEKLKVGCKM
            10        20        30        40        50        60


                     50        60
RF_1_+ QQSQNTMNHKVIEARNIRRNKYF
        .     ..::.: ::..
gi|273 LLKGIIGEEKVVELRNMKEAGADIQELRQKVEKMLSEVTDEKQKEKVHEYGPACKKIFGA
          70        80        90       100       110       120
```

```
>>gi|37778944|gb|AAO73464.1| HDM allergen [Dermatophagoi  (875 aa)
 initn:  64 init1:  64 opt:  70  Z-score: 94.4  bits: 24.8 E():  2.8
Smith-Waterman score: 70;  29.167% identity (66.667% similar) in 48 aa overlap (12-57:405-452)

                                  10        20        30        40
RF_1_+                    AAQTCRKPPIPLSCVQHEPLSYNQQHRTLKRNTMKVSYDLQ
                                  :.  .:::   .   .:.   :  :.. :.. ::
gi|377 KINLDLKDKLEEVTMLMEQAQKELRIKIGELQKLQHEYEKVRDQRDQLARENKKLTDDLA
             380       390       400       410       420       430

                           50        60
RF_1_+ QSQNTMN--HKVIEARNIRRNKYF
          ....  .:  :.  :.  ..:
gi|377 EAKSQLNDAHRRIHEQEIEIKRLENERDELSAAYKEAETLRKQEEAKNQRLIAELAQVRH
          440       450       460       470       480       490


>>gi|2735108|gb|AAD13650.1| ABA-1 allergen [Ascaris lumb  (267 aa)
 initn:  40 init1:  40 opt:  63  Z-score: 93.2  bits: 22.9 E():  3.2
Smith-Waterman score: 63;  29.167% identity (60.417% similar) in 48 aa overlap (15-58:35-82)

                                  10        20        30        40
RF_1_+                    AAQTCRKPPIPLSCVQHEPLSY-NQQHRTLKRNT---MKVSYDL
                                  .:::  .  :  .:  :    :...   .:.   .
gi|273 HYLKTYLSWLTEEQKEKLKEMKEAGQTKAEIQHEVMHYYDQLHGEEKQQATEKLKVGCKM
             10        20        30        40        50        60

                           50        60
RF_1_+ QQSQNTMNHKVIEARNIRRNKYF
          .      ..:.: .  :..
gi|273 LLKGIIGEEKVVELRNVKEAGADIQELQQKVEKMLSEVTDEKQKEKVHEYGPACKKIFGA
             70        80        90       100       110       120


>>gi|21673|emb|CAA35238.1| unnamed protein product [Trit  (307 aa)
 initn:  52 init1:  52 opt:  63  Z-score: 92.2  bits: 22.9 E():  3.7
Smith-Waterman score: 63;  23.214% identity (60.714% similar) in 56 aa overlap (7-57:111-166)

                                    10        20        30
RF_1_+                     AAQTCRKPPIPLSCVQH----EPLSYNQQHRTLKRN
                               .  :  :  :.    .:.:  .:..  ...
gi|216 FPQPQLPYPQPQLPYPQPQLPYPQPQPFRPQQPYPQSQPQYSQPQQPISQQQQQQQQQQQ
             90       100       110       120       130       140

                     40        50        60
RF_1_+ TMKVSYDLQQS-QNTMNHKVIEARNIRRNKYF
          .  .  ::  :.  .....:  :..
gi|216 QKQQQQQQQQQILQQILQQQLIPCRDVVLQQHSIAYGSSQVLQQSTYQLVQQLCCQQLWQI
             150       160       170       180       190       200


>>gi|2735112|gb|AAD13652.1| ABA-1 allergen [Ascaris lumb  (267 aa)
```

```
 initn:  39 init1:  39 opt:  61 Z-score: 90.4  bits: 22.3 E():  4.6
Smith-Waterman score: 61;  29.167% identity (60.417% similar) in 48 aa overlap (15-58:35-82)

                            10        20        30        40
RF_1_+                AAQTCRKPPIPLSCVQHEPLSY-NQQHRTLKRNT---MKVSYDL
                                .::: . : .: :  :...   .::.  .
gi|273 HYLKTYLSWLTEEQKEKLKEMKEAGKTKAEIQHEVMHYYDQLHGEEKQQATEKLKVGCKM
          10        20        30        40        50        60


               50        60
RF_1_+ QQSQNTMNHKVIEARNIRRNKYF
            .     ..::.: ::...
gi|273 LLKGIIGEEKVVELRNMKEAGADIQELQQKVEKMLSEVTDEKQKEKVHEYGPACKKIFGA
          70        80        90       100       110       120


>>gi|2735110|gb|AAD13651.1| ABA-1 allergen [Ascaris lumb  (267 aa)
 initn:  39 init1:  39 opt:  61 Z-score: 90.4  bits: 22.3 E():  4.6
Smith-Waterman score: 61;  29.167% identity (60.417% similar) in 48 aa overlap (15-58:35-82)

                            10        20        30        40
RF_1_+                AAQTCRKPPIPLSCVQHEPLSY-NQQHRTLKRNT---MKVSYDL
                                .::: . : .: :  :...   .::.  .
gi|273 HYLKTYLSWLTEEQKEKLKEMKEAGKTKAEIQHEVMRYYDQLHGEEKQQATEKLKVGCKM
          10        20        30        40        50        60


               50        60
RF_1_+ QQSQNTMNHKVIEARNIRRNKYF
            .     ..::.: ::...
gi|273 LLKGIIGEEKVVELRNMKEAGADIQELQQKVEKMLSEVTDEKQKEKVHEYGPACKKIFGA
          70        80        90       100       110       120


>>gi|75317968|sp|O22116|O22116_WHEAT LMM glutenin 3        (373 aa)
 initn:  59 init1:  59 opt:  61 Z-score: 88.0  bits: 22.4 E():  6.3
Smith-Waterman score: 61;  32.432% identity (59.459% similar) in 37 aa overlap (6-42:11-46)

                      10        20        30        40        50
RF_1_+        AAQTCRKPPIPLSCVQHEPLSYNQQHRTLKRNTMKVSYDLQQSQNTMNHKVIEAR
                   ..:: :.:  :. :.: .::    ...     : . ::
gi|753 ASAVAQISQQQQPP-PFSQQQQPPFSQQQQPPFSQQQQSPFSQQQQQPPFSQQQQPPFSQ
              10        20        30        40        50


          60
RF_1_+ NIRRNKYF

gi|753 QPLISQQQQLPFSQQQQPQFSQQQQPPYSQQQQPPYSQQQQPPFSQQQQPPFSQQQQPSF
          60        70        80        90       100       110


>>gi|75107016|sp|P80208.1|2SS3_BRANA RecName: Full=Napin   (125 aa)
 initn:  36 init1:  36 opt:  55 Z-score: 87.5  bits: 20.7 E():  6.7
```

Smith-Waterman score: 55; 26.000% identity (60.000% similar) in 50 aa overlap (6-51:43-89)

```
                                       10        20        30
RF_1_+                         AAQTCRKPPIPLSCV----QHEPLSYNQQHRTLKR
                                 ..::.  .:     :.:::        :::
gi|751 EFQQAQHLRACQQWLHKQAMQSGSGPQGPQQRPPLLQQCCNELHQEEPLCVCP---TLKG
               20        30        40        50        60

                   40        50        60
RF_1_+ NTMKVSYDLQQSQNTMNHKVIEARNIRRNKYF
         .  :.  ...:.:.  .....
gi|751 ASRAVKQQVRQQQGQQGQQLQQVISRIYQTATHLPKVCNIPQVSVCPFQKTMPGPS
          70        80        90       100       110       120
```

>>gi|3703107|gb|AAC63045.1| glycinin [Arachis hypogaea]   (507 aa)
 initn: 61 init1: 61 opt: 62 Z-score: 87.2 bits: 22.7 E():   7
Smith-Waterman score: 62; 29.167% identity (58.333% similar) in 48 aa overlap (16-61:180-227)

```
                                       10        20        30        40
RF_1_+                         AAQTCRKPPIPLSCVQHEPLSYNQQHRTLKRNTMKVSYDLQQSQN
                                               ..: : :.:: :  .: ..  :      :::
gi|370 DVVAVSLTDTNNNDNQLDQFPRRFNLAGNTEQEFLRYQQQSRQSRRRSLPYSPYSPQSQP
           150       160       170       180       190       200

                   50        60
RF_1_+ TMNHKVIEAR--NIRRNKYF
         ....  .   :   .  ::..
gi|370 RQEEREFSPRGQHSRRERAGQEEENEGGNIFSGFTPEFLEQAFQVDDRQIVQNLRGETES
          210       220       230       240       250       260
```

>>gi|224036293|pdb|3C3V|A Chain A, Crystal Structure Of   (510 aa)
 initn: 61 init1: 61 opt: 62 Z-score: 87.1 bits: 22.7 E(): 7.1
Smith-Waterman score: 62; 29.167% identity (58.333% similar) in 48 aa overlap (16-61:183-230)

```
                                       10        20        30        40
RF_1_+                         AAQTCRKPPIPLSCVQHEPLSYNQQHRTLKRNTMKVSYDLQQSQN
                                               ..: : :.:: :  .: ..  :      :::
gi|224 DVVAVSLTDTNNNDNQLDQFPRRFNLAGNHEQEFLRYQQQSRQSRRRSLPYSPYSPQSQP
               160       170       180       190       200       210

                   50        60
RF_1_+ TMNHKVIEAR--NIRRNKYF
         ....  .   :   .  ::..
gi|224 RQEEREFSPRGQHSRRERAGQEEEHEGGNIFSGFTPEFLAQAFQVDDRQIVQNLRGENES
                  220       230       240       250       260       270
```

>>gi|199732457|gb|ACH91862.1| arachin Arah3 isoform [Ara   (530 aa)
 initn: 61 init1: 61 opt: 62 Z-score: 86.8 bits: 22.7 E(): 7.3
Smith-Waterman score: 62; 29.167% identity (58.333% similar) in 48 aa overlap (16-61:203-250)

```
                                  10        20        30        40
RF_1_+                    AAQTCRKPPIPLSCVQHEPLSYNQQHRTLKRNTMKVSYDLQQSQN
                                  ..: : :.:: :  .: .. :    :::
gi|199 DVVAVSLTDTNNNDNQLDQFPRRFNLAGNHEQEFLRYQQQSRQSRRRSLPYSPYSPQSQP
           180       190       200       210       220       230


             50        60
RF_1_+ TMNHKVIEAR--NIRRNKYF
        .... . :  . ::..
gi|199 RQEEREFSPRGQHSRRERAGQEEENEGGNIFSGFTPEFLAQAFQVDDRQIVQNLRGENES
           240       250       260       270       280       290


>>gi|1304264|dbj|BAA12318.1| alpha-gliadin [Triticum aes  (259 aa)
 initn:  51 init1:  51 opt:  58  Z-score: 86.4  bits: 21.6 E():  7.7
Smith-Waterman score: 58;  24.528% identity (56.604% similar) in 53 aa overlap (9-61:83-133)


                                  10        20        30
RF_1_+                    AAQTCRKPPIPLSCVQHEPLSYNQQHRTLKRNTMKVSY
                                  : :    .::.: .::..  ... ..
gi|130 QPYLQLQPFPQPQLPYSQPQPFRPQQPYPQPQPQYSQPQEPISQQQQQQQQQQQQILQ--Q
                 60        70        80        90       100       110


           40        50        60
RF_1_+ DLQQSQNTMNHKVIEARNIRRNKYF
        :::.          :.. .:: ...
gi|130 ILQQQLIPCMDVVLQQHNIAHGRSQVLQQSTYQLLQELCCQHLWQIPEQSQCQAIQNVVH
                 120       130       140       150       160       170


>>gi|119390336|pdb|2J23|B Chain B, Cross-Reactivity And    (121 aa)
 initn:  37 init1:  37 opt:  54  Z-score: 86.3  bits: 20.5 E():  7.8
Smith-Waterman score: 54;  35.294% identity (73.529% similar) in 34 aa overlap (21-52:4-36)


              10        20        30        40        50
RF_1_+ AAQTCRKPPIPLSCVQHEPLSYNQQHRTLKRNTMKV--SYDLQQSQNTMNHKVIEARNIR
                    :.....:. . :.....: ::: :.: : . ::.
gi|119             MRGSHHHHHHLVPRGSVQVISSYD-QFKQVTGGDKVVVIDFWA
                              10        20        30        40


          60
RF_1_+ RNKYF


gi|119 TWCGPCKMIGPVFEKISDTPAGDKVGFYKVDVDEQSQIAQEVGIRAMPTFVFFKNGQKID
                 50        60        70        80        90       100


>>gi|110349081|gb|ABG73108.1| Pis v 1 allergen 2S albumi  (149 aa)
 initn:  38 init1:  38 opt:  54  Z-score: 84.8  bits: 20.5 E():  9.5
Smith-Waterman score: 54;  28.846% identity (65.385% similar) in 52 aa overlap (12-63:34-81)
```

```
                              10        20        30        40
RF_1_+                        AAQTCRKPPIPLSCVQHEPLSYNQQHRTLKRNTMKVSYDLQ
                              ::  :    ....:.. .:.  : :. ..:
gi|110 LVLLLSAFAFLILAANASIYRATVEVEGENLSSGQSCQKQFEEQQK-FKHCQMYVQQEVQ
            10        20        30        40        50        60


                50        60
RF_1_+ QSQNTMNHKVIEARNIRRNKYF
       .::.   .:. . ::  .:.. :
gi|110 KSQD--GHS-LTARINQRQQCFKQCCQELQEVDKKCRCQNLEQMVKRQQQQGQFRGEKLQ
                70        80        90        100       110
```

63 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:44:24 2010 done: Fri Apr 30 23:44:24 2010
 Total Scan time: 0.060 Total Display time: 0.000

RF_2_+1

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_+1 38 aa - 38 aa
 vs /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

```
       opt      E()
< 20    2     0:=
  22    0     0:                 one = represents 4 library sequences
  24    0     0:
  26    0     0:
  28    5     0:==
  30   13     2:*===
  32   15     8:=*==
  34   31    21:=====*==
  36   56    44:=========*===
  38   78    72:================*==
  40   68   101:================         *
  42   78   123:=================          *
  44  106   136:========================        *
  46  123   138:===========================   *
```

```
 48   183   132:==================================*=============
 50   118   121:==============================*
 52    78   106:====================        *
 54    80    91:===================   *
 56    61    76:===============   *
 58    82    62:==============*=====
 60    57    50:===========*==
 62    54    40:========*====
 64    63    32:======*=======
 66    33    25:=====*==
 68    14    20:====*
 70    18    16:===*=
 72    23    12:==*===
 74    12    10:==*
 76     5     7:=*
 78     3     6:=*
 80     4     4:*
 82     4     3:*
 84     1     3:*
 86     0     2:*
 88     0     2:*            inset = represents 1 library sequences
 90     0     1:*
 92     0     1:*         :*
 94     0     1:*         :*
 96     0     1:*         :*
 98     2     0:=         *==
100     0     0:          *
102     0     0:          *
104     0     0:          *
106     1     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))= 4.5927+/-0.00303; mu= 4.5473+/- 0.158
 mean_var=23.4163+/- 6.198, 0's: 2 Z-trim: 3  B-trim: 11 in 1/42
 Lambda= 0.265042
 Kolmogorov-Smirnov  statistic: 0.0525 (N=27) at   56

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:  0.050
The best scores are:                          opt bits E(1471)
gi|1885350|emb|CAA72273.1| serpin [Triticum aestiv ( 399)   59 25.0    0.65
gi|66845554|gb|EAL85888.1| molecular chaperone and ( 706)   58 24.5     1.7
```

```
gi|83303658|sp|P40292.3|HSP90_ASPFU RecName: Full= ( 706)    58 24.5     1.7


>>gi|1885350|emb|CAA72273.1| serpin [Triticum aestivum]    (399 aa)
 initn:  50 init1:  50 opt:  59  Z-score: 105.7  bits: 25.0 E(): 0.65
Smith-Waterman score: 59;  30.556% identity (66.667% similar) in 36 aa overlap (1-31:241-276)


                                      10         20
RF_2_+                         KLYIYIFYPKAPQGVA-----LGVRTDSKH
                               .. .::. :.:: :..      :... :  .
gi|188 MSSMDDDQYISSSDGLKVLKLPYKQGGDNRQFSMYILLPEAPGGLSSLAEKLSAEPDFLE
               220        230        240        250        260        270


           30
RF_2_+ RQLARQVGGVSLI
       :.. ::
gi|188 RHIPRQRVAIRQFKLPKFKISFGMEASDLLKCLGLQLPFSDEADFSEMVDSPMPQGLRVS
               280        290        300        310        320        330


>>gi|66845554|gb|EAL85888.1| molecular chaperone and all  (706 aa)
 initn:  57 init1:  57 opt:  58  Z-score: 98.2  bits: 24.5 E():  1.7
Smith-Waterman score: 58;  36.000% identity (72.000% similar) in 25 aa overlap (7-30:414-438)


                                      10         20         30
RF_2_+                         KLYIYIFYPKAPQGVALGVRTDSKHRQ-LARQVGGV
                                 ::     ... ::.. :....:: ::.
gi|668 KIMKVIKKNIVKKTLELFNEIAEDREQFDKFYSAFSKNIKLGIHEDAQNRQTLAKLLRYQ
           390        400        410        420        430        440



RF_2_+ SLI

gi|668 STKSGDEATSLADYVTRMPEHQKQIYYITGESIKAVAKSPFLDSLKQKNFEVLFLVDPID
           450        460        470        480        490        500


>>gi|83303658|sp|P40292.3|HSP90_ASPFU RecName: Full=Heat  (706 aa)
 initn:  57 init1:  57 opt:  58  Z-score: 98.2  bits: 24.5 E():  1.7
Smith-Waterman score: 58;  36.000% identity (72.000% similar) in 25 aa overlap (7-30:414-438)


                                      10         20         30
RF_2_+                         KLYIYIFYPKAPQGVALGVRTDSKHRQ-LARQVGGV
                                 ::     ... ::.. :....:: ::.
gi|833 KIMKVIKKNIVKKTLELFNEIAEDREQFDKFYSAFSKNIKLGIHEDAQNRQTLAKLLRYQ
           390        400        410        420        430        440



RF_2_+ SLI

gi|833 STKSGDEATSLADYVTRMPEHQKQIYYITGESIKAVAKSPFLDSLKQKNFEVLFLVDPID
           450        460        470        480        490        500
```

38 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:43:54 2010 done: Fri Apr 30 23:43:54 2010
 Total Scan time: 0.050 Total Display time: 0.000

RF_2_+3
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_+3 39 aa - 39 aa
 vs /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

         opt   E()
< 20   2   0:=
  22   0   0:   one = represents 4 library sequences
  24   0   0:
  26   2   0:=
  28   1   0:=
  30   8   2:*=
  32   29   8:=*======
  34   30   21:=====*==
  36   48   44:=========*=
  38   71   72:===============*
  40   75   101:=================   *
  42   75   123:=================   *
  44   120   136:============================   *
  46   207   138:=============================*=================
  48   98   132:======================   *
  50   109   121:========================== *
  52   131   106:======================*======
  54   47   91:===========   *
  56   77   76:================*=
  58   65   62:==============*=
  60   54   50:===========*=
  62   45   40:========*==
  64   15   32:====   *
  66   35   25:=====*==
  68   46   20:====*=======

```
   70    19    16:===*=
   72    12    12:==*
   74    25    10:==*====
   76     4     7:=*
   78     4     6:=*
   80     7     4:*=
   82     3     3:*
   84     1     3:*
   86     1     2:*
   88     2     2:*              inset = represents 1 library sequences
   90     0     1:*
   92     1     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))= 3.0904+/-0.00261; mu= 11.7041+/- 0.139
 mean_var=20.5690+/- 5.268, 0's: 2 Z-trim: 2  B-trim: 113 in 1/42
 Lambda= 0.282792
 Kolmogorov-Smirnov  statistic: 0.0328 (N=29) at  64


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:  0.050
The best scores are:                              opt bits E(1471)
gi|51093373|gb|AAT95008.1| allergen Sol i 1 precur ( 346)   50 22.8    2.7
gi|160962611|gb|ABX54876.1| Ole e 5 olive pollen a ( 152)   47 21.4    3.1
gi|4538529|emb|CAB39376.1| Cop c1 allergen [Coprin (  81)   45 20.5    3.1
gi|33113263|gb|AAP94213.1| Humj1 [Humulus japonicu ( 155)   45 20.6    5.5
gi|23616947|dbj|BAC20650.1| putative allergenic pr ( 160)   45 20.6    5.7
gi|14423730|sp|P78983.2|HSP70_ALTAL RecName: Full= ( 152)   44 20.2    7.2


>>gi|51093373|gb|AAT95008.1| allergen Sol i 1 precursor   (346 aa)
 initn:  50 init1:  50 opt:  50  Z-score: 94.6  bits: 22.8 E():  2.7
Smith-Waterman score: 50;  28.000% identity (68.000% similar) in 25 aa overlap (15-39:12-36)

                10        20        30
RF_2_+ IFFKENSIKSLKGIYIKTIYLYFLPKSTARGSPGCADGL
```

```
                  ...  .:: : .. :. .:: .. :
gi|510    MRKFAAIFVVFFVQCTHLYSLAQARAEPDPGVVEYLKQSCVYGNSSYINVYLYNSRF
                 10        20        30        40        50


gi|510 QGKNLGNQQSCQDINASLPVVFITHGFTSSAQVSTFKDLANAFVQKGHTAFIVDWSEAAC
          60        70        80        90       100       110


>>gi|160962611|gb|ABX54876.1| Ole e 5 olive pollen aller  (152 aa)
 initn:  47 init1:  47 opt:  47  Z-score: 93.6  bits: 21.4 E():  3.1
Smith-Waterman score: 47;  46.154% identity (76.923% similar) in 13 aa overlap (22-34:62-74)


                          10        20        30
RF_2_+          IFFKENSIKSLKGIYIKTIYLYFLPKSTARGSPGCADGL
                          .::: .  .:.::
gi|160 GNLSGLKPGLHGFHVHALGDTTNGCMSTGPHFLPVGKEHGAPGDENRHAGDLGNITVGED
               40        50        60        70        80        90


gi|160 GTAAINIVDKQIPLTGPHSIIGRAVVVHSDPDDLGRGGHELSKSTGNAGGRVACGIIGLQ
              100       110       120       130       140       150


>>gi|4538529|emb|CAB39376.1| Cop c1 allergen [Coprinus c  (81 aa)
 initn:  45 init1:  45 opt:  45  Z-score: 93.5  bits: 20.5 E():  3.1
Smith-Waterman score: 45;  38.462% identity (76.923% similar) in 13 aa overlap (24-36:60-72)


                          10        20        30
RF_2_+          IFFKENSIKSLKGIYIKTIYLYFLPKSTARGSPGCADGL
                               ::   ..:.:.:.
gi|453 PQLKPVAHPLLLLPLDTTTLHMPPLLLQLQLPPLLSQGNPACSPKWLQLLVP
          30        40        50        60        70        80


>>gi|33113263|gb|AAP94213.1| Humj1 [Humulus japonicus]    (155 aa)
 initn:  45 init1:  45 opt:  45  Z-score: 89.0  bits: 20.6 E():  5.5
Smith-Waterman score: 45;  26.316% identity (63.158% similar) in 19 aa overlap (17-35:122-140)


                          10        20        30
RF_2_+          IFFKENSIKSLKGIYIKTIYLYFLPKSTARGSPGCADGL
                               :. ..::.  . ..   : :
gi|331 SESSPPPATTSRESQYRGAEKDHVLVVGGCKSCFMYFMVPKQVQECPKCCGQPLHFDRSQ
               100       110       120       130       140       150


gi|331 SGSQ



>>gi|23616947|dbj|BAC20650.1| putative allergenic protei  (160 aa)
 initn:  45 init1:  45 opt:  45  Z-score: 88.8  bits: 20.6 E():  5.7
Smith-Waterman score: 45;  38.889% identity (61.111% similar) in 18 aa overlap (21-38:48-65)


                          10        20        30
RF_2_+          IFFKENSIKSLKGIYIKTIYLYFLPKSTARGSPGCADGL
```

```
                                    .: ::.   :  ..   :: :
gi|236 MLVATTTMADHRGQVVYTPGQLCAAGRGYPMYPLPRCRALAKRQCAGGAVDEQVRQDCCR
            20        30        40        50        60        70


gi|236 QLAAIDDSFCRCPALSHMLVGMYKELGAPAKGQPMDEVFPGCRRGDMKRVAASLPAFCNV
            80        90       100       110       120       130


>>gi|14423730|sp|P78983.2|HSP70_ALTAL RecName: Full=Heat  (152 aa)
 initn:  35 init1:  35 opt:  44  Z-score: 87.0  bits: 20.2 E():  7.2
Smith-Waterman score: 44;  29.412% identity (52.941% similar) in 34 aa overlap (5-38:97-129)

                                        10        20        30
RF_2_+                           IFFKENSIKSLKGIYIKTIYLYFLPKSTARGSPG
                                 :..  :  :.:.   .  :..  :    .    : ::
gi|144 DAGDKQKLTAEIDKTVQWLDDNQTATKDEYESQQKELEGV-ANPIMMKFYGAGGEGGMPG
            70        80        90       100       110       120



RF_2_+ CADGL
            :
gi|144 GMPGGGMPGGAPGGAAGDDGPTVEEVD
           130       140       150




39 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:45:19 2010 done: Fri Apr 30 23:45:20 2010
 Total Scan time:  0.050 Total Display time:  0.000
```

RF_2_-1

```
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_-1 39 aa - 39 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library


        opt       E()
 < 20     2      0:=
   22     1      0:=             one = represents 4 library sequences
   24     2      0:=
```

```
 26      2      0:=
 28      7      0:==
 30     16      2:*===
 32     21      8:=*====
 34     23     21:=====*
 36     27     44:======    *
 38     58     72:=============   *
 40    102    101:======================*
 42    117    123:==========================*
 44    112    136:=========================        *
 46    137    138:=================================*
 48    183    132:================================*============
 50    109    121:=========================  *
 52     94    106:====================   *
 54    114     91:====================*=====
 56     40     76:=========          *
 58     42     62:==========       *
 60     30     50:=======       *
 62     42     40:========*=
 64     23     32:=====  *
 66     36     25:======*==
 68     21     20:====*=
 70     16     16:===*
 72     14     12:==*=
 74     12     10:==*
 76      5      7:=*
 78     42      6:=*========
 80      2      4:*
 82      3      3:*
 84      7      3:*=
 86      3      2:*
 88      2      2:*          inset = represents 1 library sequences
 90      3      1:*
 92      0      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      0      0:          *
 331323 residues in  1471 sequences
```

  Expectation_n fit: rho(ln(x))= 2.0726+/-0.00286; mu= 16.5871+/- 0.154
 mean_var=22.7128+/- 5.587, 0's: 2 Z-trim: 2 B-trim: 219 in 1/42
 Lambda= 0.269116
 Kolmogorov-Smirnov statistic: 0.0362 (N=29) at 64

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|162794|gb|AAA30429.1| alpha-S1-casein [Bos taur (214)   49 22.1   2.7
gi|218059730|emb|CAT99618.1| profilin [Malus x dom (77)   45 19.9   4.5
gi|47606043|sp|Q8SAE6.1|PROF_DAUCA RecName: Full=P (134)   46 20.7   4.6
gi|28569698|emb|CAD68071.1| manganese superoxide d (237)   47 21.4   4.9
gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos t (205)   46 20.9   5.9
gi|162792|gb|AAA30428.1| alpha-s1-casein precursor (214)   46 21.0   6
gi|2580504|gb|AAB82404.1| Cr-PII [Periplaneta amer (395)   47 21.7   6.5
gi|1184668|gb|AAA87456.1| beta-1,3-glucanase [Heve (374)   46 21.3   8.2
gi|124365251|gb|ABN09654.1| beta-1,3-glucanase [He (374)   46 21.3   8.2
gi|164510840|emb|CAK93704.1| PR-10 protein [Malus (159)   44 20.0   8.7
gi|164510824|emb|CAK93667.1| PR-10 protein [Malus (159)   44 20.0   8.7
gi|164510828|emb|CAK93677.1| PR-10 protein [Malus (159)   44 20.0   8.7
gi|2833325|sp|Q25641.1|CRPI_PERAM RecName: Full=Al (685)   47 22.1   8.8
gi|1173557|gb|AAA86533.1| Ory s 1 [Oryza sativa]   (263)   45 20.7   8.8
gi|85701160|sp|Q00002.2|PDI_ALTAL RecName: Full=Pr (436)   46 21.4   8.9
gi|21748153|emb|CAD38167.1| putative nuclear trans (124)   43 19.4   9.9

>>gi|162794|gb|AAA30429.1| alpha-S1-casein [Bos taurus]   (214 aa)
 initn: 49 init1: 49 opt: 49 Z-score: 94.7 bits: 22.1 E(): 2.7
Smith-Waterman score: 49; 33.333% identity (58.333% similar) in 24 aa overlap (2-25:20-43)

                            10   20   30
RF_2_-   SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFSLKNI
                    : : ::: :: . ...: ..
gi|162 MKLLILTCLVAVALARPKHPIKHQGLPQEVLNENLLRFFVALFPEVFGKEKVNELSKDIG
               10   20   30   40   50   60

gi|162 SESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYK
               70   80   90   100   110   120

>>gi|218059730|emb|CAT99618.1| profilin [Malus x domesti (77 aa)
 initn: 45 init1: 45 opt: 45 Z-score: 90.7 bits: 19.9 E(): 4.5
Smith-Waterman score: 45; 58.333% identity (75.000% similar) in 12 aa overlap (6-17:43-54)

                            10   20   30
RF_2_-   SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFS
                    :: : .:. :::
gi|218 EVTGVMNEFNEPGSLAPTGLYFGGTKYMVIPGEPGVVIRGKKGPGGVTVKKSTMALLIGI
              20   30   40   50   60   70

RF_2_- LKNI

gi|218 YDEPM

>>gi|47606043|sp|Q8SAE6.1|PROF_DAUCA RecName: Full=Profi (134 aa)
 initn: 46 init1: 46 opt: 46 Z-score: 90.4 bits: 20.7 E(): 4.6
Smith-Waterman score: 46; 63.636% identity (81.818% similar) in 11 aa overlap (7-17:80-90)

```
                                 10        20        30
RF_2_-                   SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFSL
                                   : :.::. :::
gi|476 ITGIMKNFDEPGHLAPTGLYLGGTKYMVIQGEPIAVIRGKKGSGGVTIKKTGQALVFGVY
          50        60        70        80        90       100
```


RF_2_- KNI

gi|476 DEPVTPGQCNLIVERLGDYLIEQGL
         110   120   130

>>gi|28569698|emb|CAD68071.1| manganese superoxide dismu (237 aa)
 initn: 34 init1: 34 opt: 47 Z-score: 90.0 bits: 21.4 E(): 4.9
Smith-Waterman score: 47; 30.556% identity (52.778% similar) in 36 aa overlap (2-36:6-41)

```
               10        20        30
RF_2_-     SPSAHP-GLPLAVLLGKKYKYIVFIYMPLRLFIEFSLKNI
             ::: :  ::. :.... . .  :     :   : .:
gi|285 PFYPIPSALPFPLPIHSLFSRRTRLFRFSRTAARAGTEHTLPPLPYEYNALEPFISADIM
            10        20        30        40        50        60
```

```
gi|285 MVHHGKHHQTYVNNLNASTKAYNDAVQAQDVLKQMELLTAVKFNGGGHVNHALFWKTMAP
            70        80        90       100       110       120
```

>>gi|159793197|gb|ABW98943.1| alpha S1 casein [Bos tauru (205 aa)
 initn: 46 init1: 46 opt: 46 Z-score: 88.6 bits: 20.9 E(): 5.9
Smith-Waterman score: 46; 38.095% identity (57.143% similar) in 21 aa overlap (2-22:11-31)

```
               10        20        30
RF_2_-         SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFSLKNI
                 :  : :::  ::  .  ...:
gi|159 VSAALARPKHPIKHQGLPQEVLNENLLRFFVAPFPEVFGKEKVNELSKDIGSESTEDQAM
            10        20        30        40        50        60
```

```
gi|159 EDIKQMEAESISSSEEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPN
            70        80        90       100       110       120
```

>>gi|162792|gb|AAA30428.1| alpha-s1-casein precursor [Bo (214 aa)

```
   initn:  46 init1:  46 opt:  46  Z-score: 88.4  bits: 21.0 E():    6
Smith-Waterman score: 46;  38.095% identity (57.143% similar) in 21 aa overlap (2-22:20-40)

                                    10        20        30
RF_2_-                    SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFSLKNI
                           :   : ::: :: .   ...:
gi|162 MKLLILTCLVAVALARPKHPIKHQGLPQEVLNENLLRFFVAPFPEVFGKEKVNELSKDIG
                 10        20        30        40        50        60

gi|162 SESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYK
                 70        80        90       100       110       120

>>gi|2580504|gb|AAB82404.1| Cr-PII [Periplaneta american  (395 aa)
 initn:  47 init1:  47 opt:  47  Z-score: 87.8  bits: 21.7 E():  6.5
Smith-Waterman score: 47;  36.000% identity (64.000% similar) in 25 aa overlap (10-34:58-82)

                                    10        20        30
RF_2_-                    SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFSLKNI
                            .: : :.... :::     :. : .:
gi|258 LNDFLALLPVDEITAIVMDYLANDAEVQEAVAYLQGEEFHKIVFTVEGLQEFGNFVQFLE
            30        40        50        60        70        80

gi|258 DHGLDAVGYINRLHSVFGWDPYVPSSKRKHTRRGVGVDGLIDDIIAILPIDDLKALFQEK
            90       100       110       120       130       140

>>gi|1184668|gb|AAA87456.1| beta-1,3-glucanase [Hevea br  (374 aa)
 initn:  46 init1:  46 opt:  46  Z-score: 86.0  bits: 21.3 E():  8.2
Smith-Waterman score: 46;  29.167% identity (66.667% similar) in 24 aa overlap (13-36:209-232)

                                    10        20        30
RF_2_-                    SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFSLKNI
                           ::.. : :... : :  . . ..:
gi|118 SYPPSAGAFRDDVRSYLDPIIGFLSSIRSPLLANIYPYFTYAYNPRDISLPYALFTSPSV
          180       190       200       210       220       230

gi|118 VVWDGQRGYKNLFDATLDALYSALERASGGSLEVVVSESGWPSAGAFAATFDNGRTYLSN
          240       250       260       270       280       290

>>gi|124365251|gb|ABN09654.1| beta-1,3-glucanase [Hevea    (374 aa)
 initn:  46 init1:  46 opt:  46  Z-score: 86.0  bits: 21.3 E():  8.2
Smith-Waterman score: 46;  29.167% identity (66.667% similar) in 24 aa overlap (13-36:209-232)

                                    10        20        30
RF_2_-                    SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFSLKNI
                           ::.. : :... : :  . . ..:
gi|124 SYPPSAGAFRDDVRSYLDPIIGFLSSSXSPLLANIYPYFTYAYNPRDISLPYALFTSPSV
          180       190       200       210       220       230

gi|124 VVWDGQRGYKNLFDATLDALYSALERASGGSLEVVVSESGWPSAGAFAATFDNGRTYLSN
```

```
>>gi|164510840|emb|CAK93704.1| PR-10 protein [Malus x do  (159 aa)
 initn:  44 init1:  44 opt:  44  Z-score: 85.5  bits: 20.0 E():  8.7
Smith-Waterman score: 44;  37.500% identity (81.250% similar) in 16 aa overlap (6-21:53-68)

                                        10        20        30
RF_2_-                       SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFS
                                 :.   ::..  :..:.:.
gi|164 LFLMLTISFRRLLHKQSKLLKFSREMAVLEPSRKLALVKGSEYSYVKHKVEGIDKDNFDY
            30        40        50        60        70        80


RF_2_- LKNI

gi|164 SYSLIEGDAISDKIEKISYEIKLVASGSGSIIKNTSHYHTKGDVEIKEEHVKAGKDKAHG
            90        100       110       120       130       140

>>gi|164510824|emb|CAK93667.1| PR-10 protein [Malus x do  (159 aa)
 initn:  44 init1:  44 opt:  44  Z-score: 85.5  bits: 20.0 E():  8.7
Smith-Waterman score: 44;  37.500% identity (81.250% similar) in 16 aa overlap (6-21:53-68)

                                        10        20        30
RF_2_-                       SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFS
                                 :.   ::..  :..:.:.
gi|164 LFLMLTISFRRLLHKQSKLLKFSREMAVLEPSRKLALVKGSEYSYVKHKVEGIDKDNFDY
            30        40        50        60        70        80


RF_2_- LKNI

gi|164 SYSLIEGDAISDKIEKISYEIKLVASGSGSIIKNTSHYHTKGDVEIKEEHVKVGKDKAHG
            90        100       110       120       130       140

>>gi|164510828|emb|CAK93677.1| PR-10 protein [Malus x do  (159 aa)
 initn:  44 init1:  44 opt:  44  Z-score: 85.5  bits: 20.0 E():  8.7
Smith-Waterman score: 44;  37.500% identity (81.250% similar) in 16 aa overlap (6-21:53-68)

                                        10        20        30
RF_2_-                       SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFS
                                 :.   ::..  :..:.:.
gi|164 LFLMLTISFRRLLHKQSKLLKFSREMAVLEPSRKLALVKGSEYSYVKHKVEGIDKDNFDY
            30        40        50        60        70        80


RF_2_- LKNI

gi|164 SYSLIEGDAISDKIEKISYEIKLVASGSGSIIKNTSHYHTKGDVEIKEEHVKAGKDKAHG
            90        100       110       120       130       140
```

```
>>gi|2833325|sp|Q25641.1|CRPI_PERAM RecName: Full=Allerg  (685 aa)
 initn:  47 init1:  47 opt:  47  Z-score: 85.4  bits: 22.1 E():  8.8
Smith-Waterman score: 47;  58.333% identity (83.333% similar) in 12 aa overlap (10-21:503-514)

                              10        20        30
RF_2_-                 SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFSLKNI
                           .::.:: :: :: :.
gi|283 NVDVQLAVRLNHKPFTYNIEVSSDKAQDVYVAVFLGPKYDYLGREYDLNDRRHYFVEMDR
           480       490       500       510       520       530

gi|283 FPYHVGAGKTVIERNSHDSNIIAPERDSYRTFYKKVQEAYEGKSQYYVDKGHNYCGYPEN
           540       550       560       570       580       590

>>gi|1173557|gb|AAA86533.1| Ory s 1 [Oryza sativa]        (263 aa)
 initn:  45 init1:  45 opt:  45  Z-score: 85.4  bits: 20.7 E():  8.8
Smith-Waterman score: 45;  55.556% identity (88.889% similar) in 9 aa overlap (1-9:20-28)

                              10        20        30
RF_2_-                 SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLFIEFSLKNI
                          :::..:: .:
gi|117 MASSSLLLACVVVAAMVSPSPAGHPKVPPGPNITTSYGDKWLEARPPGMVRPRVLAPKDN
               10        20        30        40        50        60

>>gi|85701160|sp|Q00002.2|PDI_ALTAL RecName: Full=Protei  (436 aa)
 initn:  38 init1:  38 opt:  46  Z-score: 85.3  bits: 21.4 E():  8.9
Smith-Waterman score: 46;  30.303% identity (69.697% similar) in 33 aa overlap (2-34:231-261)

                              10        20        30
RF_2_-                   SPSAHPGLPLAVLLGKKYKYIVFIYMPLRLF
                          : .. : :..:.....:: .: :     ..
gi|857 EAKITEKEIGKFVDDFLAGKIDPSIKSEPIPESNDG-PVTVVAHNYKDVV-IDNDKDVL
             210       220       230       240       250

RF_2_- IEFSLKNI
        .::
gi|857 VEFYAPWCGHCKALAPKYEELGQLYASDELSKLVTIAKVDATLNDVPDEIQGFLPSSLFP
          260       270       280       290       300       310

>>gi|21748153|emb|CAD38167.1| putative nuclear transport  (124 aa)
 initn:  33 init1:  33 opt:  43  Z-score: 84.5  bits: 19.4 E():  9.9
Smith-Waterman score: 43;  34.146% identity (51.220% similar) in 41 aa overlap (2-38:72-105)

                              10                20
RF_2_-                 SPSAHPGLPL----AVLLGKKYKYIVFIYMP
                          ::: :. .    :.::: . :        :
gi|217 GTQGSAAIVEKLQNLPFQEIQHRTDTVDAQPSADDGILVLVTGALLLGGESK-------P
             50        60        70        80        90
```

```
          30
RF_2_- LRLFIEFSLKNI
          . .    :.:::
gi|217 MSFTQAFQLKNAEGNWFVLNDVFRLVYPAA
           100      110      120




39 residues in 1 query   sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:45:46 2010 done: Fri Apr 30 23:45:46 2010
 Total Scan time:  0.040 Total Display time:  0.010


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 39 aa - 39 aa
 vs  /bioinformatics/Allergenicity/FARRPV10/FARRPV10.fasta library

       opt      E()
< 20     2     0:=
  22     0     0:                 one = represents 4 library sequences
  24     0     0:
  26     0     0:
  28     0     0:
  30     1     2:*
  32     6     8:=*
  34    14    21:==== *
  36    27    44:======   *
  38    51    72:============    *
  40    70   101:================        *
  42   114   123:=========================== *
  44   104   136:========================       *
  46   147   138:================================*==
  48   194   132:===============================*================
  50   143   121:=============================*=====
  52   132   106:========================*=====
  54   106    91:====================*====
  56    55    76:============     *
```

```
 58    40    62:==========       *
 60    37    50:==========   *
 62    51    40:========*===
 64    21    32:====== *
 66    35    25:======*==
 68    24    20:====*=
 70    17    16:===*=
 72    13    12:==*=
 74    13    10:==*=
 76    12     7:=*=
 78     4     6:=*
 80    10     4:*==
 82     5     3:*=
 84     6     3:*=
 86     0     2:*
 88     5     2:*=            inset = represents 1 library sequences
 90     2     1:*
 92     1     1:*        :*
 94     1     1:*        :*
 96     1     1:*        :*
 98     1     0:=       *=
100     0     0:        *
102     1     0:=       *=
104     2     0:=       *==
106     2     0:=       *==
108     0     0:        *
110     0     0:        *
112     1     0:=       *=
114     0     0:        *
116     0     0:        *
118     0     0:        *
>120     0     0:        *
 331323 residues in  1471 sequences
  Expectation_n fit: rho(ln(x))= 4.7787+/-0.00256; mu= 1.2920+/- 0.134
 mean_var=22.2534+/- 5.123, 0's: 2 Z-trim: 5  B-trim: 47 in 1/42
 Lambda= 0.271879
 Kolmogorov-Smirnov  statistic: 0.0814 (N=28) 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|125987805|sp|P08819.2|CBP2_WHEAT RecName: Full= ( 444)   60 26.5    0.27
gi|76097507|gb|ABA39436.1| Der f 1 allergen precur ( 276)   55 24.8    0.56
gi|38326693|gb|AAR17475.1| unknown [Penicillium ci ( 228)   54 24.4    0.57
gi|27530349|dbj|BAC53948.1| Der f 1 allergen prepr ( 321)   55 24.7    0.68
gi|156106765|gb|ABU49605.1| Der f 1 allergen [Derm ( 321)   55 24.7    0.68
gi|730035|sp|P16311.2|PEPT1_DERFA RecName: Full=Pe ( 321)   54 24.3    0.89
gi|18772|emb|CAA45778.1| trypsin inhibitor subtype ( 217)   50 22.9     1.6
```

```
gi|2506771|sp|P16968.2|IAA1_HORVU RecName: Full=Al ( 146)    47 21.9    2.1
gi|129614|sp|P00784.1|PAPA1_CARPA RecName: Full=Pa ( 345)    50 22.7    2.9
gi|4826572|emb|CAB42886.1| polygalacturonase [Phle ( 394)    50 22.6    3.5
gi|83305645|sp|Q92450.3|SODM_ASPFU RecName: Full=S ( 210)    46 21.3    4.5
gi|1648970|gb|AAB60779.1| manganese superoxide dis ( 221)    46 21.3    4.8
gi|14423687|sp|Q9LEI9.1|ENO2_HEVBR RecName: Full=E ( 445)    49 22.2    5.3
gi|14423688|sp|Q9LEJ0.1|ENO1_HEVBR RecName: Full=E ( 445)    49 22.2    5.3
gi|1168171|gb|AAB35353.1| allergenic peptide=low-m (  30)    36 18.3    5.5
gi|11124572|emb|CAC14917.1| triosephosphat-isomera ( 253)    46 21.3    5.8
gi|170726|gb|AAA34283.1| pre-alpha-/beta-gliadin A ( 282)    46 21.2    6.6
gi|1228078|emb|CAA33034.1| kappa-casein [Bos tauru ( 190)    43 20.2    9
gi|162811|gb|AAA30433.1| kappa-casein precursor [B ( 190)    43 20.2    9


>>gi|125987805|sp|P08819.2|CBP2_WHEAT RecName: Full=Seri  (444 aa)
 initn:  47 init1:  47 opt:  60  Z-score: 112.7  bits: 26.5 E(): 0.27
Smith-Waterman score: 60;  34.483% identity (65.517% similar) in 29 aa overlap (12-39:73-101)


                            10        20        30
RF_2_-                  ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCF-WVKNINI
                              ...:..:::.:.  :  :  . : :  :.
gi|125 LQEAPEDAQPAPLVLWLNGGPGCSSVAYGASEELGAFRVKPRGAGLVLNEYRWNKVANVL
              50        60        70        80        90       100


gi|125 FLDSPAGVGFSYTNTSSDIYTSGDNRTAHDSYAFLAKWFERFPHYKYRDFYIAGESYAGH
             110       120       130       140       150       160


>>gi|76097507|gb|ABA39436.1| Der f 1 allergen precursor   (276 aa)
 initn:  49 init1:  49 opt:  55  Z-score: 106.9  bits: 24.8 E(): 0.56
Smith-Waterman score: 55;  31.034% identity (62.069% similar) in 29 aa overlap (4-32:171-197)


                            10        20        30
RF_2_-                  ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFW
                              ..:  ::.::.  .    ::..: : .  .
gi|760 QELVDCASQHGCHGDTIPRGIEYIQQNGVVEERSYPYVAREQQCR--RPNSQHYGISNYC
            150       160       170       180       190


RF_2_- VKNINI

gi|760 QIYPPDVKQIREALTQTHTAIAVIIGIKDLRAFQHYDGRTIIQRDNGYQPNYHAVNIVGY
          200       210       220       230       240       250


>>gi|38326693|gb|AAR17475.1| unknown [Penicillium citrin  (228 aa)
 initn:  49 init1:  49 opt:  54  Z-score: 106.7  bits: 24.4 E(): 0.57
Smith-Waterman score: 54;  30.303% identity (54.545% similar) in 33 aa overlap (7-37:28-60)


                            10        20        30
RF_2_-                  ASSDQRHTPYLARQLS--VFRVRPHTQGYPLRCFWVKNI
                           :.:  :  ..   .:.:.:..  ::    : :.:
```

```
gi|383 MGFTDFVSDAGLSLANNYLATRSYIVGHAPSQADVVTYKAFTASPDAEKYPHVARWYKHI
                10        20        30        40        50        60


RF_2_- NI

gi|383 ASYESEFPTLPGDASKAFTAYGPEGSEASANPKDKPAEEEEEEDLFASDSEDEDPAVVAE
                70        80        90        100       110       120

>>gi|27530349|dbj|BAC53948.1| Der f 1 allergen preproenz  (321 aa)
 initn:  49 init1:  49 opt:  55  Z-score: 105.4  bits: 24.7 E(): 0.68
Smith-Waterman score: 55;  31.034% identity (62.069% similar) in 29 aa overlap (4-32:189-215)

                                          10        20        30
RF_2_-                             ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFW
                                   ..:  ::.::. .     :::..: : .  .
gi|275 QELVDCASQHGCHGDTIPRGIEYIQQNGVVEERSYPYVAREQQCR--RPNSQHYGISNYC
         160       170       180       190       200       210


RF_2_- VKNINI

gi|275 QIYPPDVKQIREALTQTHTAIAVIIGIKDLRAFQHYDGRTIIQHDNGYQPNYHAVNIVGY
          220       230       240       250       260       270

>>gi|156106765|gb|ABU49605.1| Der f 1 allergen [Dermatop  (321 aa)
 initn:  49 init1:  49 opt:  55  Z-score: 105.4  bits: 24.7 E(): 0.68
Smith-Waterman score: 55;  31.034% identity (62.069% similar) in 29 aa overlap (4-32:189-215)

                                          10        20        30
RF_2_-                             ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFW
                                   ..:  ::.::. .     :::..: : .  .
gi|156 QELVDCASQHGCHGDTIPRGIEYIQQNGVVEERSYPYVAREQQCR--RPNSQHYGISNYC
         160       170       180       190       200       210


RF_2_- VKNINI

gi|156 QIYPPDVKQIREALTQTHTAIAVIIGIKDLRAFQHYDGRTIIQHDNGYQPNYHAVNIVGY
          220       230       240       250       260       270

>>gi|730035|sp|P16311.2|PEPT1_DERFA RecName: Full=Peptid  (321 aa)
 initn:  48 init1:  48 opt:  54  Z-score: 103.3  bits: 24.3 E(): 0.89
Smith-Waterman score: 54;  31.034% identity (58.621% similar) in 29 aa overlap (4-32:189-215)

                                          10        20        30
RF_2_-                             ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFW
                                   ..:  ::.::.       :::..: : .  .
gi|730 QELVDCASQHGCHGDTIPRGIEYIQQNGVVEERSYPYVAREQRCR--RPNSQHYGISNYC
```

```
         160        170        180        190        200        210


RF_2_- VKNINI

gi|730 QIYPPDVKQIREALTQTHTAIAVIIGIKDLRAFQHYDGRTIIQHDNGYQPNYHAVNIVGY
          220        230        240        250        260        270


>>gi|18772|emb|CAA45778.1| trypsin inhibitor subtype B [  (217 aa)
 initn:  50 init1:  50 opt:  50  Z-score: 98.8  bits: 22.9 E():  1.6
Smith-Waterman score: 50;  53.333% identity (73.333% similar) in 15 aa overlap (16-30:85-99)


                              10        20        30
RF_2_-            ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWVKNINI
                              :  ::.:   ..:  :::
gi|187 RAAPTGNERCPLTVVQSRNELDKGIGTIISSPFRIRFIAEGNPLRLKFDSFAVIMLCVGI
          60        70        80        90        100       110


gi|187 PTEWSVVEDLPEGPAVKIGENKDAVDGWFRIERVSDDEFNNYKLVFCTQQAEDDKCGDIG
          120       130       140       150       160       170


>>gi|2506771|sp|P16968.2|IAA1_HORVU RecName: Full=Alpha-  (146 aa)
 initn:  47 init1:  47 opt:  47  Z-score: 96.4  bits: 21.9 E():  2.1
Smith-Waterman score: 47;  38.462% identity (69.231% similar) in 13 aa overlap (22-34:122-134)


                         10        20        30
RF_2_-          ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWVKNINI
                              :.:.:     :.:.
gi|250 ATVAEVFPGCRTEVMDRAVASLPAVCNQYIPNTNGTDGVCYWLSYYQPPRQMSSR
                100       110       120       130       140


>>gi|129614|sp|P00784.1|PAPA1_CARPA RecName: Full=Papain  (345 aa)
 initn:  50 init1:  50 opt:  50  Z-score: 94.1  bits: 22.7 E():  2.9
Smith-Waterman score: 50;  38.889% identity (77.778% similar) in 18 aa overlap (9-26:235-252)


                              10        20        30
RF_2_-                ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWVKNIN
                              :: :.   .: .:.:...:
gi|129 LQLVAQYGIHYRNTYPYEGVQRYCRSREKGPYAAKTDGVRQVQPYNEGALLYSIANQPVS
          210       220       230       240       250       260


RF_2_- I

gi|129 VVLEAAGKDFQLYRGGIFVGPCGNKVDHAVAAVGYGPNYILIKNSWGTGWGENGYIRIKR
          270       280       290       300       310       320


>>gi|4826572|emb|CAB42886.1| polygalacturonase [Phleum p  (394 aa)
 initn:  50 init1:  50 opt:  50  Z-score: 92.7  bits: 22.6 E():  3.5
```

Smith-Waterman score: 50; 28.571% identity (60.714% similar) in 28 aa overlap (2-29:275-302)

```
                                      10        20        30
RF_2_-                      ASSDQRHTPYLARQLSVFRVRPHTQGYPLRC
                            :  .. ..:  : .:.    :. .   :::.
gi|482 LGRYKDEKDVTDITVKNCVLKKSTNGLRIKSYEDAKSPLTASKLTYENVKMEDVGYPIII
          250       260       270       280       290       300


RF_2_- FWVKNINI

gi|482 DQKYCPNKICTSKGDSARVTVKDVTFRNITGTSSTPEAVSLLCSDKQPCNGVTMNDVKIE
          310       320       330       340       350       360
```

>>gi|83305645|sp|Q92450.3|SODM_ASPFU RecName: Full=Super   (210 aa)
 initn: 46 init1: 46 opt: 46 Z-score: 90.6 bits: 21.3 E(): 4.5
Smith-Waterman score: 46; 31.579% identity (68.421% similar) in 19 aa overlap (9-27:19-37)

```
                       10        20        30
RF_2_-          ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWVKNINI
                    ::...:.  .. . : : :
gi|833 MSQQYTLPPLPYPYDALQPYISQQIMELHHKKHHQTYVNGLNAALEAQKKAAEANDVPKL
                10        20        30        40        50        60


gi|833 VSVQQAIKFNGGGHINHSLFWKNLAPEKSGGGKIDQAPVLKAAIEQRWGSFDKFKDAFNT
                70        80        90       100       110       120
```

>>gi|1648970|gb|AAB60779.1| manganese superoxide dismuta   (221 aa)
 initn: 46 init1: 46 opt: 46 Z-score: 90.1 bits: 21.3 E(): 4.8
Smith-Waterman score: 46; 31.579% identity (68.421% similar) in 19 aa overlap (9-27:30-48)

```
                             10        20        30
RF_2_-                ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWVKNINI
                          ::...:.  .. . : : :
gi|164 GTSPIQTPINTMSQQYTLPPLPYPYDALQPYISQQIMELHHKKHHQTYVNGLNAALEAQK
                10        20        30        40        50        60


gi|164 KAAEATDVPKLVSVQQAIKFNGGGHINHSLFWKNLAPEKSGGGKIDQAPVLKAAIEQRWG
                70        80        90       100       110       120
```

>>gi|14423687|sp|Q9LEI9.1|ENO2_HEVBR RecName: Full=Enola   (445 aa)
 initn: 49 init1: 49 opt: 49 Z-score: 89.4 bits: 22.2 E(): 5.3
Smith-Waterman score: 49; 42.105% identity (73.684% similar) in 19 aa overlap (11-29:118-136)

```
                             10        20        30
RF_2_-                ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWVKNINI
                          :: .:.:  ..  :..:  ::
gi|144 VGIDNFMVQQLDGTVNEWGWCKQKLGANAILAVSLAVCKAGAHVKGIPLYKHVANLAGNK
             90       100       110       120       130       140
```

```
gi|144 NLVLPVPAFNVINGGSHAGNKLAMQEFMILPVGASSFKEAMKMGAEVYHHLKSVIKKKYG
             150       160       170       180       190       200


>>gi|14423688|sp|Q9LEJ0.1|ENO1_HEVBR RecName: Full=Enola  (445 aa)
 initn:  49 init1:  49 opt:  49  Z-score: 89.4  bits: 22.2 E():  5.3
Smith-Waterman score: 49;  42.105% identity (73.684% similar) in 19 aa overlap (11-29:118-136)


                                  10        20        30
RF_2_-                     ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWVKNINI
                                  ::  .:.:  ..   :..: ::
gi|144 VGIDNFMVQQLDGTVNEWGWCKQKLGANAILAVSLAVCKAGAHVKGIPLYEHIANLAGNK
           90       100       110       120       130       140


gi|144 NLVLPVPAFNVINGGSHAGNKLAMQEFMILPVGASSFKEAMKMGAEVYHHLKSVIKKKYG
             150       160       170       180       190       200


>>gi|1168171|gb|AAB35353.1| allergenic peptide=low-molec  (30 aa)
 initn:  31 init1:  31 opt:  36  Z-score: 89.1  bits: 18.3 E():  5.5
Smith-Waterman score: 36;  28.571% identity (57.143% similar) in 28 aa overlap (3-28:1-28)


                 10        20        30
RF_2_- ASSDQRHTPYLARQ--LSVFRVRPHTQGYPLRCFWVKNINI
          :.:.. :.   .:  .:  .:  .   : .:   :
gi|116    SQQQQPPFSQQQPPFSQQQQPPFSQQQPPF
                  10        20        30


>>gi|11124572|emb|CAC14917.1| triosephosphat-isomerase [  (253 aa)
 initn:  33 init1:  33 opt:  46  Z-score: 88.7  bits: 21.3 E():  5.8
Smith-Waterman score: 46;  33.333% identity (48.148% similar) in 27 aa overlap (9-35:45-70)


                                  10        20        30
RF_2_-                     ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWVKNIN
                                  ::.      ..::. :    : :::
gi|111 GTVEQVESIVNTLNAGQIASTDVVEVVVSPPYVFLPTVKGKLRPEIQVAAQNC-WVKKGG
           20        30        40        50        60        70


RF_2_- I


gi|111 AFTGEVSAEMLVNLGVPWVILGHSERRSLMGESSEFVGEKVAYALAQGLKVIACVGETLE
                  80        90       100       110       120       130


>>gi|170726|gb|AAA34283.1| pre-alpha-/beta-gliadin A-III  (282 aa)
 initn:  46 init1:  46 opt:  46  Z-score: 87.6  bits: 21.2 E():  6.6
Smith-Waterman score: 46;  29.167% identity (54.167% similar) in 24 aa overlap (5-28:48-71)


                                  10        20        30
RF_2_-                     ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWV
```

```
                                     :.. :   .:.   .  :: : .:
gi|170 TSAVRVPVPQLQPQNPSQQQPQEQVPLMQQQQQFPGQQEQFPPQQPYPHQQPFPSQQPYP
          20        30        40        50        60        70


RF_2_- KNINI


gi|170 QPQPFPPQLPYPQTQPFPPQQPYPQPQPQYPQPQQPISQQQAQQQQQQQQQTLQQILQQQL
          80        90       100       110       120       130

>>gi|1228078|emb|CAA33034.1| kappa-casein [Bos taurus]    (190 aa)
 initn: 43 init1: 43 opt: 43 Z-score: 85.3 bits: 20.2 E():    9
Smith-Waterman score: 43;  33.333% identity (61.905% similar) in 21 aa overlap (11-31:12-32)


                 10        20        30
RF_2_-  ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWVKNINI
                   ::  :  .  .. ..:  :.::
gi|122 MMKSFFLVVTILALTLPFLGAQEQNQEQPIRCEKDERFFSDKIAKYIPIQYVLSRYPSYG
          10        20        30        40        50        60


gi|122 LNYYQQKPVALINNQFLPYPYYAKPAAVRSPAQILQWQVLSNTVPAKSCQAQPTTMARHP
          70        80        90       100       110       120


>>gi|162811|gb|AAA30433.1| kappa-casein precursor [Bos t  (190 aa)
 initn: 43 init1: 43 opt: 43 Z-score: 85.3 bits: 20.2 E():    9
Smith-Waterman score: 43;  33.333% identity (61.905% similar) in 21 aa overlap (11-31:12-32)


                 10        20        30
RF_2_-  ASSDQRHTPYLARQLSVFRVRPHTQGYPLRCFWVKNINI
                   ::  :  .  .. ..:  :.::
gi|162 MMKSFFLVVTILALTLPFLGAQEQNQEQPIRCEKDERFFSDKIAKYIPIQYVLSRYPSYG
          10        20        30        40        50        60


gi|162 LNYYQQKPVALINNQFLPYPYYAKPAAVRSPAQILQWQVLSNTVPAKSCQAQPTTMARHP
          70        80        90       100       110       120




39 residues in 1 query    sequences
331323 residues in 1471 library sequences
 Scomplib [34t26]
 start: Fri Apr 30 23:46:54 2010 done: Fri Apr 30 23:46:54 2010
 Total Scan time:  0.030 Total Display time:  0.010

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