

CONFIDENTIAL ATTACHMENT

Study Title

**Additional Bioinformatic Evaluation of DNA Sequences Flanking the 5' Junction of
the Inserted DNA in MON 87705: Assessment of Putative Polypeptides**

Authors

**Andre Silvanovich, Ph.D.
HaidTu**

Study Completed On

October 20, 2009

Sponsor and Performing Laboratory

**Monsanto Company
Product Characterization Center
800 North Lindbergh Blvd.
St. Louis, MO 63167**

Laboratory Project ID

**MSL0022346
Study Number: REG-09-471**

Description of Confidential Attachment

The following sections of this report include *Confidential Business Information*:

Reason for deletion of each these sections listed above from the main body of report:
These sections disclose commercial information (description of product manufacturing or quality control processes) FIFRA reference 10(d)(1)(A).

<u>Deleted Pages</u>	<u>Reason for Deletion</u>	<u>FIFRA Reference</u>
5-24	Discloses manufacturing or quality control processes	10(d)

Table of Contents

The following sections of this report include *Confidential Business Information*:

Confidential Attachment	Page(s)
Figure 1. Reading frame assignment and DNA sequence at the additional 5' junction contained in MON 87705.....	4
Figure 2. Graphic mapping of the predicted sequence of polypeptides encoded by each reading frame at the additional 5' junction contained in MON 87705.....	5
Table 1. The predicted sequence of polypeptides encoded by each reading frame at the additional 5' junction contained in MON 87705.	6
Appendix 1. Bioinformatic analysis of polypeptide 5_1a	7
Appendix 2. Bioinformatic analysis of polypeptide 5_2a	10
Appendix 3. Bioinformatic analysis of polypeptide 5_3a	13
Appendix 4. Bioinformatic analysis of polypeptide 5_4a	16
Appendix 5. Bioinformatic analysis of polypeptide 5_5a	18
Appendix 6. Bioinformatic analysis of polypeptide 5_6a	21

Deleted pages are attached immediately following this page.

```
5'   CAAGATTCTCAAGACCGTGAAACAAAGGATtctcaggggtgttggttatcactgcggtttggcc   3'
5_1a  K I L K T V K Q R I s g c c y h c g l a   >
5_2a  R F S R P . N K G s q g v v i t a v w p   >
5_3a  Q D S Q D R E T K D l r v l l s l r f g   >

3'   GTTCTAAGAGTTCTGGCACTTTGTTTCCTAgagtcccacaacaatagtgacgccaacccgg   5'
5_4a  L N E L G H F L P D . p t t i v a t q g   <
5_5a  L I R L V T F C L I e p h q . . q p k a   <
5_6a  S E . S R S V F S r l t n n d s r n p r   <
```

Figure 1. Reading frame assignment and DNA sequence at the additional 5' junction contained in MON 87705.

Upper case characters refer to the genomic flank DNA and lower case characters refer to the so called likely duplicated DNA found in MON 87705. Likewise, uppercase amino acids are translated from genomic sequence and lowercase amino acids are translated from the inserted likely duplicated DNA. The carat (> or <) points towards the carboxyl terminal of each polypeptide. Stop codons are denoted as periods (.).

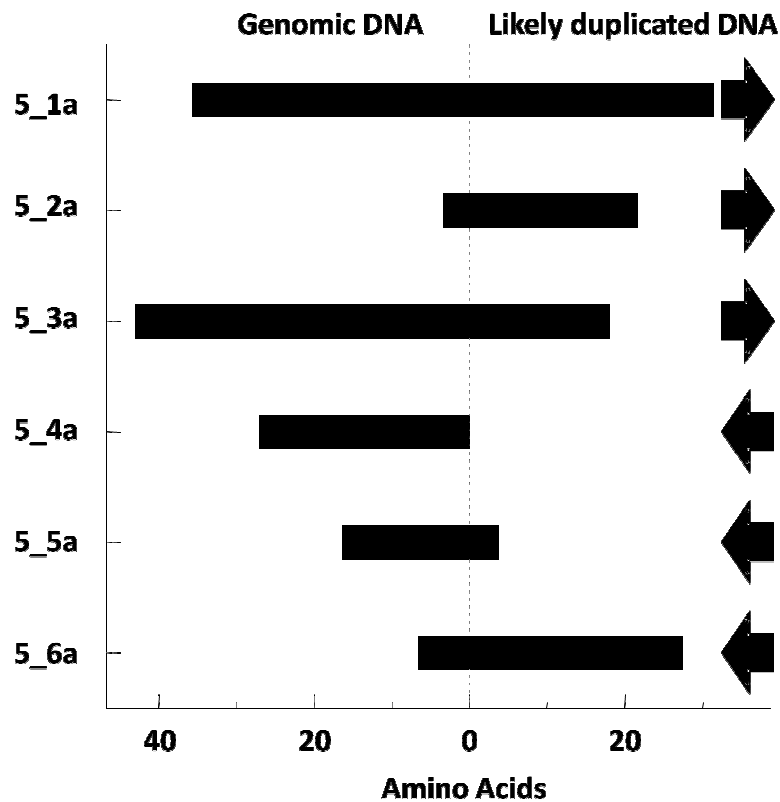


Figure 2. Graphic mapping of the predicted sequence of polypeptides encoded by each reading frame at the additional 5' junction contained in MON 87705.

The putative polypeptide coding sequences are mapped relative to the DNA sequence shown in Figure 2, and the amino acid sequences are tabulated in Table 1. The scale at the bottom of each map refers to amino acids from the junction. The arrow for each ORF points in the direction of N- to C-terminus.

Putative peptide ID	Putative peptide amino sequence
5_1a	CPNTHWWDSK STRGEQNVFY IINIQTILK TVKQRIsgcc yhcglafgpr hrcpeknmrv vilakde
5_2a	NKGsqgvvit avwplgqgtv vlkki
5_3a	GSLVVPIMPK HPLVGLKIYK GRAECLLHHQ YPNQDSQDRE TKDlrvllsl rfglwakapl s
5_4a	DPLFHGLENL GLDIDDVEDI LLSPCRF
5_5a	qhpeILCFTV LRILVWILMM
5_6a	rvlqlsyffq dngalaqrpn rsdnntlrSF VSRS

Table 1. The predicted sequence of polypeptides encoded by each reading frame at the additional 5' junction contained in MON 87705.

For display purposes, the predicted sequences are parsed into segments of ten amino acids in length. Uppercase characters refer to sequence encoded by genomic DNA. Lowercase characters refer to sequence encoded by the likely duplicated DNA.

Appendix 1. Bioinformatic analysis of polypeptide 5_1a

```
>5_1a
CPNTHWWDSKSTRGEQN VFYIINIQT KILKTVKQRISGCCYHCLAFGPRHRCPEKNMRVVILAKDE
```

```
Sliding 8 amino acid window search
Database searched = AD_2009
Query = 5_1a
```

```
Start time: Wed Sep 23 12:58:11 CDT 2009 Finish time: Wed Sep 23 12:58:11 CDT
2009
```

No 8 amino acid matches exist between 5_1a and the AD_2009 database

```
# fasta34 5_1a.pep /home/andre/db/AD_2009 -Q -E 1 -O 5_1a.pep_ad.fasta
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

```
5_1a, 67 aa
vs /home/andre/db/AD_2009 library
```

```

      opt      E()
< 20    11    0:====
      22     0    0:
      24     0    0:
      26     0    0:
      28     0    0:
      30     3    2:*
      32     8    7:==*
      34     3    20:=
      36    19    41:=====
      38    68    68:=====
      40    64    95:=====
      42   138   116:=====
      44   102   128:=====
      46   121   130:=====
      48    96   125:=====
      50    88   114:=====
      52   149   100:=====
      54   118    85:=====
      56    69    71:=====
      58    74    59:=====
      60    55    47:=====
```

```

62    40    38:=====*=
64    24    30:===== *
66    29    24:=====*=
68    19    19:=====*
70    16    15:=====*=
72    23    12:=====
74     9     9:==*
76    11     7:==*=
78     6     5:==*
80     3     4:==*
82     7     3:==*
84     4     3:==*
86     3     2:==*
88     2     2:==*
90     1     1:==*
92     1     1:==*
94     0     1:==*
96     0     1:==*
98     0     0:==*
100    0     0:==*
102    0     0:==*
104     1     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:==*
```

```

307888 residues in 1386 sequences
Expectation_n fit: rho(ln(x))= 5.57240.00354; mu= -2.3279 0.185
mean_var=41.965210.919, 0's: 11 Z-trim: 13 B-trim: 3 in 1/43
Lambda= 0.197984
Kolmogorov-Smirnov statistic: 0.0988 (N=29) at 50
```

```

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 2
join: 36, opt: 24, open/ext: -10/-2, width: 16
The best scores are:
gi|5059162|gb|AAD38942.1|AF144060_1 alpha-amylase ( 496) 69 26.3 0.54
gi|481397|pir||S38584 allergen Phl p Vb - common t ( 280) 64 25.0 0.77
```

```

>>gi|5059162|gb|AAD38942.1|AF144060_1 alpha-amylase [Der (496 aa)
initn: 39 initl: 39 opt: 69 Z-score: 106.7 bits: 26.3 E(): 0.54
Smith-Waterman score: 69; 26.087% identity (57.971% similar) in 69 aa
overlap (2-60:401-467)
```

```

10      20
5_1a    CPNTHWWDSKS-----TRGEQN VFYIINIQT
```


Confidential Attachment Page 8 of 24

Confidential Attachment

```
114      0      0:      *
116      0      0:      *
118      0      0:      *
>120     0      0:      *
1891534 residues in 7651 sequences
Expectation_n fit: rho(ln(x))= 0.79330.00072; mu= 24.6640 0.036
mean_var=67.046615.694, 0's: 63 Z-trim: 63 B-trim: 827 in 2/60
Lambda= 0.156634
Kolmogorov-Smirnov statistic: 0.0516 (N=29) at 42

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 2
join: 36, opt: 24, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000

67 residues in 1 query sequences
1891534 residues in 7651 library sequences
Scomplib [34t26]
start: Wed Sep 23 12:58:11 2009 done: Wed Sep 23 12:58:12 2009
Total Scan time: 0.240 Total Display time: 0.000

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

# fasta34 5_1a.pep /home/andre/db/PRT_2009 -Q -E 1 -O 5_1a.pep_prt.fasta
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

5_1a, 67 aa
vs /home/andre/db/PRT_2009 library

      opt      E()
< 20 258614    0:=====
 22  105      0:=          one = represents 22042 library sequences
 24  237     14:*
 26  527    309:*
 28 1691   3337:*
 30  9721 20271:*
 32 48950 78382:===*
 34 164656 212563:===== *
 36 395999 436554:===== *
 38 759290 721460:=====*=
 40 1078487 1006374:=====*=
 42 1271224
1230169:=====*=
 44 1322482
1356992:=====*
```

```
46 1279430
1382130:=====
48 1191912 1323230:=====
*
50 1096612 1207452:===== *
52 981272 1061552:===== *
54 857876 906752:===== *
56 762987 757416:=====*
58 640884 621824:=====*=
60 537618 503714:=====*=
62 425741 403829:=====*=
64 354240 321163:=====*=
66 279563 253837:=====*=
68 220592 199663:=====*=
70 169193 156468:=====*
72 133487 122265:=====*=
74 108957 95326:=====*
76 82933 74194:=====
78 65439 57670:=====
80 49372 44780:=====
82 38660 34256:=====
84 29248 27135:=====
86 21226 20995:=====
88 17386 16245:=====
90 12755 12570:=====
92 9759 9726:=====
94 8107 7525:=====
96 6069 5823:=====
98 4837 4505:=====
100 4151 3486:=====
102 2941 2697:=====
104 3076 2087:=====
106 2001 1615:=====
108 1063 1249:=====
110 858 967:=====
112 570 748:=====
114 505 579:=====
116 294 448:=====
118 218 347:=====
>120 603 268:=====
3787527556 residues in 14717352 library sequences
statistics sampled from 60000 to 14713875 sequences
Expectation_n fit: rho(ln(x))= 4.02290.000193; mu= 7.4170 0.010
mean_var=48.9261 9.742, 0's: 1025 Z-trim: 1025 B-trim: 0 in 0/64
Lambda= 0.183360
Kolmogorov-Smirnov statistic: 0.0337 (N=29) at 54

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 2
join: 36, opt: 24, open/ext: -10/-2, width: 16
```

The best scores are: opt bits
E(14717352)
gi|124360394|gb|ABN08407.1| Peptidase aspartic, ac (435) 135 42.9 0.051
gi|124360392|gb|ABN08405.1| Peptidase aspartic, ac (435) 135 42.9 0.051
gi|124359710|gb|ABN06064.1| RNA-directed DNA polym (1297) 137 43.9 0.078
gi|217073570|gb|ACJ85145.1| unknown [Medicago trun (185) 127 40.5 0.12

>>gi|124360394|gb|ABN08407.1| Peptidase aspartic, active (435 aa)
initn: 107 init1: 107 opt: 135 Z-score: 197.5 bits: 42.9 E(): 0.051
Smith-Waterman score: 135; 45.455% identity (75.758% similar) in 33 aa
overlap (34-66:58-90)

```
      10      20      30      40      50      60
5_1a  THWWDSKSTRGEQNVFYIINIQTILKTVKQRIISGCCYHCLAFGPRHRCPEKNMRVIL
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
gi|124 NVGQNKTHTINTANWRDKNVRSLSQEIADRRQKGLCFKCGGPYHPRHQCPDKNLSVMVL
      30      40      50      60      70      80
```

```
5_1a  AKDE
      :
gi|124 EDDSEDENEVRVLNDEDVDTGAEELQLNVLTFFENALTFDRQTEYYQDRFQCIRFQGVKRE
      90      100      110      120      130      140
```

>>gi|124360392|gb|ABN08405.1| Peptidase aspartic, active (435 aa)
initn: 107 init1: 107 opt: 135 Z-score: 197.5 bits: 42.9 E(): 0.051
Smith-Waterman score: 135; 45.455% identity (75.758% similar) in 33 aa
overlap (34-66:58-90)

```
      10      20      30      40      50      60
5_1a  THWWDSKSTRGEQNVFYIINIQTILKTVKQRIISGCCYHCLAFGPRHRCPEKNMRVIL
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
gi|124 NVGQNKTHTINTANWRDKNVRSLSQEIADRRQKGLCFKCGGPYHPRHQCPDKNLSVMVL
      30      40      50      60      70      80
```

```
5_1a  AKDE
      :
gi|124 EDDSEDENEVRVLNDEDVDTGAEELQLNVLTFFENALTFDRQTEYYQDRFQCIRFQGVKRE
      90      100      110      120      130      140
```

>>gi|124359710|gb|ABN06064.1| RNA-directed DNA polymeras (1297 aa)
initn: 137 init1: 137 opt: 137 Z-score: 194.0 bits: 43.9 E(): 0.078
Smith-Waterman score: 137; 41.667% identity (80.556% similar) in 36 aa
overlap (32-67:105-140)

```
      10      20      30      40      50      60
5_1a  PNTHWWDSKSTRGEQNVFYIINIQTILKTVKQRIISGCCYHCLAFGPRHRCPEKNMRV
      .... .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:

```

```
gi|124 GPKGEKQAQYDKKKSGPRDRSFTHLSYNELMERKQKGLCFKCGGPFHMPHQCPDKQLRVL
      80      90      100      110      120      130
```

```
5_1a  ILAKDE
      .: .: .:
gi|124 VLEEDDEEGEPEGKLLAVEVDDEEEDGEMCMMEFFHLGHSRPSIKLMGVIKEVPVVVLV
      140      150      160      170      180      190
```

>>gi|217073570|gb|ACJ85145.1| unknown [Medicago truncatu (185 aa)
initn: 106 init1: 68 opt: 127 Z-score: 190.9 bits: 40.5 E(): 0.12
Smith-Waterman score: 127; 44.444% identity (77.778% similar) in 36 aa
overlap (33-67:3-38)

```
      10      20      30      40      50      60
5_1a  NTHWWDSKSTRGEQNVFYIINIQTILKTVKQRIISGCCYHCLAFGPR-HRCPEKNMRVV
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
gi|217 MAERRAKGLCFKCGGKYHPTLHKCPEKSLRVL
      10      20      30
```

```
5_1a  ILAKDE
      .: .: .:
gi|217 ILGEGGVNEEGEIVSLETQEVLEEEEEIESECKVIGVLGSMGEYNTMKIGGKLENIDV
      40      50      60      70      80      90
```

67 residues in 1 query sequences
3787527556 residues in 14717352 library sequences
Scomplib [34t26]
start: Wed Sep 23 12:58:12 2009 done: Wed Sep 23 13:05:50 2009
Total Scan time: 437.780 Total Display time: 0.000

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

Appendix 2. Bioinformatic analysis of polypeptide 5_2a

>5_2a
NKGSGQVVITAVWPLQGQTVVLKKI

Sliding 8 amino acid window search
Database searched = AD_2009
Query = 5_2a

Start time: Wed Sep 23 13:05:53 CDT 2009 Finish time: Wed Sep 23 13:05:54 CDT 2009

No 8 amino acid matches exist between 5_2a and the AD_2009 database

```
# fasta34 5_2a.pep /home/andre/db/AD_2009 -Q -E 1 -O 5_2a.pep_ad.fasta
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

5_2a, 25 aa

vs /home/andre/db/AD_2009 library

```

      opt      E()
< 20      3      0:==
 22       2      0:=
 24       6      0:==
 26      10      0:====
 28       8      0:====
 30      11      2:*===
 32      14      7:==*==
 34      17     20:=====*
 36      32     41:===== *
 38      46     68:===== *
 40      61     95:===== *
 42      77    116:===== *
 44      89    128:===== *
 46     117    130:===== *
 48     147    125:===== *
 50     153    114:===== *
 52     108    100:===== *
 54      77     85:===== *
 56      79     71:===== *
 58      46     59:===== *
 60      78     47:===== *
 62      40     38:===== *
 64      16     30:===== *
 66      15     24:===== *
 68      19     19:===== *
 70      11     15:===== *
 72      22     12:===== *
 74      15      9:===== *
 76      11      7:===== *
 78      11      5:===== *
 80      12      4:===== *
 82      14      3:===== *
 84       1      3:*
 86       1      2:*
 88       4      2:*
inset = represents 1 library sequences

```

```

90      11      1:*===
92       0      1:*
94       2      1:*
96       0      1:*
98       0      0:
100      0      0:
102      0      0:
104      0      0:
106      0      0:
108      0      0:
110      0      0:
112      0      0:
114      0      0:
116      0      0:
118      0      0:
>120     0      0:

```

307888 residues in 1386 sequences

Expectation_n fit: rho(ln(x))= 4.12290.00294; mu= 3.5383 0.154
mean_var=23.4566 6.223, 0's: 3 Z-trim: 3 B-trim: 216 in 1/43
Lambda= 0.264815
Kolmogorov-Smirnov statistic: 0.0849 (N=29) at 46

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
join: 42, opt: 30, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000

25 residues in 1 query sequences

307888 residues in 1386 library sequences

Scomplib [34t26]

start: Wed Sep 23 13:05:53 2009 done: Wed Sep 23 13:05:53 2009

Total Scan time: 0.040 Total Display time: 0.000

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

```
# fasta34 5_2a.pep /home/andre/db/TOX_2009 -Q -E 1 -O 5_2a.pep_tx.fasta
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

5_2a, 25 aa

vs /home/andre/db/TOX_2009 library

```

      opt      E()
< 20      60      0:=====
 22       1      0:=
 24       0      0:
one = represents 14 library sequences

```

```
26 5 0:=
28 16 2:*=
30 16 11:*=
32 64 41:==*==
34 190 110:=====*=
36 268 227:=====*=
38 313 375:===== *
40 391 523:===== *
42 407 639:===== *
44 743 705:=====*=
46 792 718:=====*=
48 602 688:===== *
50 770 627:=====*=
52 755 552:=====*=
54 465 471:===== *
56 370 394:===== *
58 232 323:===== *
60 347 262:=====*=
62 135 210:===== *
64 99 167:===== *
66 103 132:===== *
68 112 104:===== *
70 88 81:===== *
72 42 64:===== *
74 35 50:===== *
76 36 39:===== *
78 41 30:===== *
80 33 23:===== *
82 31 18:===== *
84 13 14:*
86 5 11:*
88 6 8:*
90 9 7:*
92 3 5:*
94 9 4:*
96 8 3:*
98 13 2:*
100 1 2:*
102 0 1:*
104 12 1:*
106 5 1:*
108 0 1:*
110 0 1:*
112 0 0:*
114 0 0:*
116 0 0:*
118 0 0:*
>120 0 0:*
1891534 residues in 7651 sequences
Expectation_n fit: rho(ln(x))= 3.00000.000575; mu= 9.1741 0.029
```

```
mean_var=18.8650 3.913, 0's: 59 Z-trim: 60 B-trim: 508 in 1/61
Lambda= 0.295288
Kolmogorov-Smirnov statistic: 0.0338 (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: 16
!! No sequences with E() < 1.000000
```

```
25 residues in 1 query sequences
1891534 residues in 7651 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:05:54 2009 done: Wed Sep 23 13:05:54 2009
Total Scan time: 0.180 Total Display time: 0.000
```

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

```
# fasta34 5_2a.pep /home/andre/db/PRT_2009 -Q -E 1 -O 5_2a.pep_prt.fasta
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

5_2a, 25 aa
vs /home/andre/db/PRT_2009 library

```
opt E()
< 20 220741 0:=====
22 761 0:= one = represents 22586 library sequences
24 1803 14:*
26 4516 309:*
28 12834 3337:*
30 37029 20272:*
32 97391 78384:==*=
34 223954 212569:=====*
36 410662 436566:=====*
38 661187 721481:===== *
40 913960 1006402:===== *
42 1159619 1230204:===== *
44 1287563
1357030:===== *
46 1355153
1382169:===== *
48 1311896
1323267:===== *
50 1238984 1207487:=====*=
52 1116498 1061582:=====*=
54 963477 906777:=====*=
```

```
56 785774 757437:=====*=
58 619884 621841:=====*
60 482694 503728:=====*
62 391616 403840:=====*
64 316035 321172:=====*
66 253318 253845:=====*
68 188880 199669:=====*
70 153690 156472:=====*
72 118175 122268:=====*
74 88407 95328:=====*
76 74785 74196:=====*
78 55931 57672:=====*
80 45974 44781:=====*
82 32502 34257:=====*
84 22946 27135:=====*
86 17538 20996:=====*
88 13300 16246:=====*
90 9632 12570:=====*
92 7538 9726:=====*
94 4993 7525:=====*
96 3644 5823:=====*
98 2596 4505:=====*
100 2107 3486:=====*
102 1244 2697:=====*
104 998 2087:=====*
106 749 1615:=====*
108 449 1249:=====*
110 279 967:=====*
112 221 748:=====*
114 149 579:=====*
116 113 448:=====*
118 75 347:=====*
>120 154 268:=====*

inset = represents 151 library sequences

3787527556 residues in 14717352 library sequences
statistics sampled from 60000 to 14714292 sequences
Expectation_n fit: rho(ln(x))= 3.90070.000177; mu= 5.2272 0.010
mean_var=25.8007 5.101, 0's: 906 Z-trim: 911 B-trim: 0 in 0/63
Lambda= 0.252499
Kolmogorov-Smirnov statistic: 0.0203 (N=29) at 48

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
join: 42, opt: 30, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000

25 residues in 1 query sequences
3787527556 residues in 14717352 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:05:54 2009 done: Wed Sep 23 13:12:26 2009
Total Scan time: 372.670 Total Display time: 0.000
```

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

Appendix 3. Bioinformatic analysis of polypeptide 5_3a

>5_3a
GSLVVPIMPKHPLVGLKIYKGRAECLLHHQYPNQDSQDRETKDLRVLLSLRFGWLWAKAPLS

Sliding 8 amino acid window search
Database searched = AD_2009
Query = 5_3a

Start time: Wed Sep 23 13:12:29 CDT 2009 Finish time: Wed Sep 23 13:12:30 CDT 2009

No 8 amino acid matches exist between 5_3a and the AD_2009 database

fasta34 5_3a.pep /home/andre/db/AD_2009 -Q -E 1 -O 5_3a.pep_ad.fasta
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

5_3a, 61 aa
vs /home/andre/db/AD_2009 library

	opt	E()
< 20	9	0:===
22	0	0:
24	0	0:
26	1	0:=
28	0	0:
30	0	2:*
32	1	7:= *
34	12	20:==== *
36	14	41:===== *
38	68	68:=====*
40	93	95:=====*
42	105	116:===== *
44	98	128:===== *
46	135	130:=====*
48	162	125:=====*
50	90	114:===== *
52	115	100:=====*
54	69	85:===== *
56	60	71:===== *

one = represents 3 library sequences

```
58 62 59:=====*=
60 32 47:===== *
62 61 38:=====*=
64 33 30:=====*=
66 28 24:=====*=
68 19 19:=====*
70 23 15:=====*
72 23 12:=====*
74 17 9:=====*
76 6 7:=====*
78 8 5:=====*
80 4 4:=====*
82 2 3:=====*
84 23 3:=====*
86 4 2:=====*
88 4 2:=====* inset = represents 1 library sequences
90 2 1:=====*
92 2 1:=====*
94 0 1:=====*
96 0 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:=====*
307888 residues in 1386 sequences
Expectation_n fit: rho(ln(x))= 4.48860.00349; mu= 2.3001 0.182
mean_var=34.6177 9.319, 0's: 9 Z-trim: 9 B-trim: 106 in 1/42
Lambda= 0.217984
Kolmogorov-Smirnov statistic: 0.0665 (N=29) at 60

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 2
join: 36, opt: 24, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000

61 residues in 1 query sequences
307888 residues in 1386 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:12:28 2009 done: Wed Sep 23 13:12:29 2009
Total Scan time: 0.030 Total Display time: 0.000

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

```
# fasta34 5_3a.pep /home/andre/db/TOX_2009 -Q -E 1 -O 5_3a.pep_tx.fasta
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
```

```
5_3a, 61 aa
vs /home/andre/db/TOX_2009 library
```

```
opt      E()
< 20      70      0:=====
22         0      0:===== one = represents 14 library sequences
24         1      0:=====
26         14     0:=====
28         2      2:=====
30         5      11:=====
32         14     41:===== *
34         56     110:===== *
36         187    227:===== *
38         553    375:===== *
40         697    523:===== *
42         550    639:===== *
44         560    705:===== *
46         784    718:===== *
48         635    688:===== *
50         463    627:===== *
52         520    552:===== *
54         421    471:===== *
56         309    394:===== *
58         465    323:===== *
60         290    262:===== *
62         227    210:===== *
64         210    167:===== *
66         186    132:===== *
68         81     104:===== *
70         59     81:===== *
72         46     64:===== *
74         37     50:===== *
76         31     39:===== *
78         36     30:===== *
80         28     23:===== *
82         14     18:===== *
84         13     14:===== *
86         6      11:===== *
88         50     8:===== *
90         5      7:===== *
92         5      5:===== *
```

```
94 3 4:* :==*
96 6 3:* :==*==
98 3 2:* :==
100 0 2:* : *
102 1 1:* :*
104 0 1:* :*
106 1 1:* :*
108 0 1:* :*
110 2 1:* :*
112 0 0: *
114 0 0: *
116 0 0: *
118 0 0: *
>120 0 0: *
1891534 residues in 7651 sequences
Expectation_n fit: rho(ln(x))= 3.97160.000569; mu= 5.9351 0.029
mean_var=36.5481 7.805, 0's: 70 Z-trim: 70 B-trim: 151 in 1/60
Lambda= 0.212149
Kolmogorov-Smirnov statistic: 0.0411 (N=29) at 56

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 2
join: 36, opt: 24, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000

61 residues in 1 query sequences
1891534 residues in 7651 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:12:30 2009 done: Wed Sep 23 13:12:30 2009
Total Scan time: 0.180 Total Display time: 0.000

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

```
# fasta34 5_3a.pep /home/andre/db/PRT_2009 -Q -E 1 -O 5_3a.pep_prt.fasta
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

5_3a, 61 aa
vs /home/andre/db/PRT_2009 library

opt E()
< 20 265159 0:=====
22 133 0:== one = represents 22371 library sequences
24 301 14:*
26 690 309:*
28 2107 3337:*
```

```
30 8244 20271:*
32 38462 78383:== *
34 137073 212565:===== *
36 365311 436558:===== *
38 692156 721468:===== *
40 1024628 1006385:=====*=
42 1272738
1230183:=====*=
44 1342251
1357007:=====*
46 1304265
1382145:=====*
48 1213666 1323244:=====
*
50 1106597 1207466:===== *
52 995154 1061564:===== *
54 883674 906762:=====*
56 753860 757424:=====*
58 653639 621831:=====*=
60 543762 503719:=====*=
62 435641 403833:=====*=
64 368135 321167:=====*=
66 291944 253840:=====*=
68 235871 199665:=====*=
70 176145 156470:=====*=
72 138517 122266:=====*=
74 105806 95327:=====*
76 86280 74195:=====*
78 66864 57671:=====*
80 50765 44780:=====*
82 39014 34256:=====*
84 27938 27135:=====*
86 21698 20996:=====*
88 15542 16245:=====*
90 13516 12570:=====*
92 9476 9726:=====*
94 7437 7525:=====*
96 5494 5823:===== *
98 4125 4505:===== *
100 2801 3486:===== *
102 1987 2697:===== *
104 1468 2087:===== *
106 1106 1615:===== *
108 815 1249:===== *
110 623 967:===== *
112 405 748:===== *
114 303 579:===== *
116 241 448:===== *
118 144 347:===== *
>120 447 268:===== *
```

inset = represents 190 library sequences

3787527556 residues in 14717352 library sequences
statistics sampled from 60000 to 14714039 sequences
Expectation_n fit: rho(ln(x))= 4.16730.000187; mu= 6.1776 0.010
mean_var=42.8967 8.630, 0's: 1127 Z-trim: 1127 B-trim: 0 in 0/63
Lambda= 0.195822
Kolmogorov-Smirnov statistic: 0.0390 (N=29) at 56

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 2
join: 36, opt: 24, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000

61 residues in 1 query sequences
3787527556 residues in 14717352 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:12:30 2009 done: Wed Sep 23 13:19:48 2009
Total Scan time: 419.190 Total Display time: 0.000

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

Appendix 4. Bioinformatic analysis of polypeptide 5_4a

>5_4a
DPLFHGLENLGLDIDDVEDILLSPCRF

Sliding 8 amino acid window search
Database searched = AD_2009
Query = 5_4a

Start time: Wed Sep 23 13:19:50 CDT 2009 Finish time: Wed Sep 23 13:19:51 CDT 2009

No 8 amino acid matches exist between 5_4a and the AD_2009 database

fasta34 5_4a.pep /home/andre/db/AD_2009 -Q -E 1 -O 5_4a.pep_ad.fasta
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

5_4a, 27 aa
vs /home/andre/db/AD_2009 library

	opt	E()
< 20	3	0:==
22	1	0:==

one = represents 3 library sequences

24	9	0:==	
26	17	0:=====	
28	0	0:	
30	3	2:*	
32	11	7:==*=	
34	12	20:===== *	
36	25	41:===== *	
38	35	68:===== *	
40	104	95:=====*	
42	68	116:===== *	
44	150	128:=====*	
46	105	130:===== *	
48	132	125:=====*	
50	126	114:=====*	
52	140	100:=====*	
54	76	85:===== *	
56	72	71:=====*	
58	61	59:=====*	
60	40	47:===== *	
62	51	38:=====*	
64	22	30:===== *	
66	23	24:=====*	
68	27	19:=====*	
70	12	15:=====*	
72	10	12:=====*	
74	8	9:=====*	
76	8	7:=====*	
78	8	5:=====*	
80	18	4:=====*	
82	0	3:*	
84	2	3:*	
86	2	2:*	
88	0	2:*	inset = represents 1 library sequences
90	1	1:*	
92	0	1:*	:*
94	1	1:*	:*
96	3	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:	*

307888 residues in 1386 sequences

Expectation_n fit: rho(ln(x))= 3.84480.00321; mu= 6.0246 0.165
mean_var=24.3126 6.302, 0's: 3 Z-trim: 3 B-trim: 81 in 2/42
Lambda= 0.260111
Kolmogorov-Smirnov statistic: 0.0488 (N=27) at 46

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
join: 42, opt: 30, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000

27 residues in 1 query sequences
307888 residues in 1386 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:19:50 2009 done: Wed Sep 23 13:19:50 2009
Total Scan time: 0.030 Total Display time: 0.000

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

fasta34 5_4a.pep /home/andre/db/TOX_2009 -Q -E 1 -O 5_4a.pep_tx.fasta
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

5_4a, 27 aa
vs /home/andre/db/TOX_2009 library

	opt	E()
< 20	60	0:=====
22	0	0: one = represents 14 library sequences
24	1	0: =
26	1	0: =
28	14	2: *
30	42	11: * ==
32	72	41: == * ==
34	129	110: ===== * ==
36	221	227: ===== *
38	266	375: ===== *
40	439	523: ===== *
42	822	639: ===== * =====
44	581	705: ===== *
46	601	718: ===== *
48	595	688: ===== *
50	521	627: ===== *
52	412	552: ===== *
54	637	471: ===== * =====
56	571	394: ===== * =====
58	309	323: ===== *

60	233	262:===== *
62	314	210:===== * =====
64	182	167:===== * =
66	172	132:===== * ==
68	68	104:===== *
70	75	81:===== *
72	69	64:===== *
74	52	50:===== *
76	41	39:===== *
78	27	30:===== *
80	20	23:===== *
82	17	18:===== *
84	41	14:===== *
86	11	11:===== *
88	6	8:===== * inset = represents 1 library sequences
90	14	7:===== *
92	3	5:===== * =
94	1	4:===== * = *
96	0	3:===== * = *
98	4	2:===== * = * ==
100	2	2:===== * = *
102	0	1:===== * = *
104	0	1:===== * = *
106	0	1:===== * = *
108	0	1:===== * = *
110	0	1:===== * = *
112	0	0:===== *
114	0	0:===== *
116	0	0:===== *
118	0	0:===== *
>120	0	0:===== *

1891534 residues in 7651 sequences
Expectation_n fit: rho(ln(x))= 3.69500.000567; mu= 6.8625 0.029
mean_var=20.1285 4.222, 0's: 59 Z-trim: 59 B-trim: 235 in 1/61
Lambda= 0.285870
Kolmogorov-Smirnov statistic: 0.0660 (N=29) at 52

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
join: 42, opt: 30, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000

27 residues in 1 query sequences
1891534 residues in 7651 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:19:51 2009 done: Wed Sep 23 13:19:51 2009
Total Scan time: 0.190 Total Display time: 0.000

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

```
# fasta34 5_4a.pep /home/andre/db/PRT_2009 -Q -E 1 -O 5_4a.pep_prt.fasta
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

5_4a, 27 aa
vs /home/andre/db/PRT_2009 library

      opt      E()
< 20 219441    0:=====
22   593      0:=          one = represents 23183 library sequences
24  1272     14:*
26  3155     309:*
28 10947    3337:*
30 39019   20272:*
32 104755   78383:===*
34 248956  212566:=====*=
36 479475  436561:=====*=
38 756120  721473:=====*=
40 1033409 1006392:=====*=
42 1267743 1230192:=====*=
44 1359477
1357016:=====*
46 1390932
1382155:=====*
48 1342293
1323254:=====*
50 1169660 1207474:===== *
52 1007463 1061571:===== *
54 860673  906768:===== *
56 690594  757429:===== *
58 564292  621835:===== *
60 443479  503723:===== *
62 365605  403836:===== *
64 297765  321169:===== *
66 233907  253842:===== *
68 190288  199667:===== *
70 156057  156471:===== *
72 118243  122267:===== *
74 88435   95327:===== *
76 67857   74196:===== *
78 52908   57671:===== *
80 37910   44781:===== *
82 29780   34256:===== *
84 20673   27135:===== *
86 15647   20996:===== *
```

```
88 11672 16245:*      inset = represents 136 library sequences
90 9604 12570:*
92 6786 9726:*      :=====
94 4740 7525:*      :=====
96 3484 5823:*      :=====
98 2401 4505:*      :=====
100 1741 3486:*      :=====
102 1341 2697:*      :=====
104 1061 2087:*      :=====
106 764 1615:*      :=====
108 512 1249:*      :=====
110 366 967:*      :=====
112 285 748:*      :=====
114 229 579:*      :=====
116 157 448:*      :=====
118 134 347:*      :=====
>120 318 268:*      :=====
```

3787527556 residues in 14717352 library sequences
statistics sampled from 60000 to 14714140 sequences
Expectation_n fit: rho(ln(x))= 4.26050.000177; mu= 4.9497 0.010
mean_var=28.3993 5.692, 0's: 904 Z-trim: 910 B-trim: 0 in 0/65
Lambda= 0.240669
Kolmogorov-Smirnov statistic: 0.0184 (N=29) at 48

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
join: 42, opt: 30, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000

27 residues in 1 query sequences
3787527556 residues in 14717352 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:19:51 2009 done: Wed Sep 23 13:26:55 2009
Total Scan time: 405.110 Total Display time: 0.000

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

Appendix 5. Bioinformatic analysis of polypeptide 5_5a

```
>5_5a
QHPEILCFTVLRILVWILMM
```

Sliding 8 amino acid window search
Database searched = AD_2009
Query = 5_5a

Start time: Wed Sep 23 13:26:58 CDT 2009 Finish time: Wed Sep 23 13:26:59 CDT 2009

No 8 amino acid matches exist between 5_5a and the AD_2009 database

```
# fasta34 5_5a.pep /home/andre/db/AD_2009 -Q -E 1 -O 5_5a.pep_ad.fasta
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

5_5a, 20 aa

vs /home/andre/db/AD_2009 library

```

      opt      E()
< 20      3      0:==
22         0      0:
24         0      0:
26         1      0:==
28         0      0:
30         4      2:*
32         5      7:==*
34        10     20:==== *
36        41     41:=====*
38        70     68:=====*=
40        53     95:=====
42        60    116:=====
44        97    128:=====
46       128    130:=====
48       141    125:=====
50       151    114:=====
52       101    100:=====
54        84     85:=====
56        92     71:=====
58        73     59:=====
60        57     47:=====
62        26     38:=====
64        48     30:=====
66        51     24:=====
68        24     19:=====
70         10     15:=====
72         12     12:=====
74          6      9:==*
76          6      7:==*
78         17     5:==*
80          1      4:=*
82          0      3:*
84          3      3:*
86          1      2:*
88          2      2:*
      inset = represents 1 library sequences
```

```

90         1      1:*
92         0      1:*
94         0      1:*
96         3      1:*
98         4      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:
```

307888 residues in 1386 sequences

Expectation_n fit: rho(ln(x))= 3.19180.00258; mu= 4.5308 0.135

mean_var=19.0760 4.958, 0's: 3 Z-trim: 3 B-trim: 18 in 1/43

Lambda= 0.293651

Kolmogorov-Smirnov statistic: 0.1001 (N=28) at 46

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1

join: 42, opt: 30, open/ext: -10/-2, width: 16

!! No sequences with E() < 1.000000

20 residues in 1 query sequences

307888 residues in 1386 library sequences

Scomplib [34t26]

start: Wed Sep 23 13:26:57 2009 done: Wed Sep 23 13:26:58 2009

Total Scan time: 0.020 Total Display time: 0.000

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

```
# fasta34 5_5a.pep /home/andre/db/TOX_2009 -Q -E 1 -O 5_5a.pep_tx.fasta
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

5_5a, 20 aa

vs /home/andre/db/TOX_2009 library

```

      opt      E()
< 20      59      0:=====
22         2      0:==
24         1      0:==
      one = represents 13 library sequences
```

```
26      3      0:=
28      68      2:*=====
30      21      11:*=
32      91      41:====*==
34      81      110:===== *
36      363     227:=====*=====
38      416     375:=====*====
40      256     523:=====
42      695     639:=====*====
44      612     705:=====
46      742     718:=====*====
48      577     688:=====
50      471     627:=====
52      422     552:=====
54      606     471:=====*=====
56      391     394:=====
58      474     323:=====
60      281     262:=====
62      228     210:=====
64      228     167:=====
66      144     132:=====
68      86      104:=====
70      73      81:=====
72      55      64:=====
74      61      50:=====
76      52      39:=====
78      23      30:=====
80      11      23:=====
82      18      18:=====
84      12      14:=====
86      5       11:*
88      5       8:*
90      1       7:*
92      0       5:*
94      4       4:*
96      1       3:*
98      0       2:*
100     1       2:*
102     0       1:*
104     5       1:*
106     0       1:*
108     0       1:*
110     1       1:*
112     0       0:*
114     0       0:*
116     0       0:*
118     0       0:*
>120    0       0:*
1891534 residues in 7651 sequences
Expectation_n fit: rho(ln(x))= 1.38100.000538; mu= 14.8547 0.027
```

```
mean_var=21.9742 4.864, 0's: 59 Z-trim: 59 B-trim: 48 in 1/61
Lambda= 0.273601
Kolmogorov-Smirnov statistic: 0.0523 (N=29) at 52
```

```
FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
join: 42, opt: 30, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000
```

```
20 residues in 1 query sequences
1891534 residues in 7651 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:26:59 2009 done: Wed Sep 23 13:26:59 2009
Total Scan time: 0.140 Total Display time: 0.000
```

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

```
# fasta34 5_5a.pep /home/andre/db/PRT_2009 -Q -E 1 -O 5_5a.pep_prt.fasta
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

5_5a, 20 aa
vs /home/andre/db/PRT_2009 library

```
      opt      E()
< 20 220265    0:=====
22 869      0:=
24 1579     14:*
26 4347     309:*
28 12104    3337:*
30 35436    20272:*=
32 90076    78383:====
34 199942   212566:=====
36 366470   436560:=====
38 590767   721471:=====
40 872018   1006389:=====
42 1102396  1230188:=====
44 1262980
1357012:=====
46 1321703
1382151:=====
48 1312311
1323250:=====
50 1218019  1207471:=====
52 1084813  1061568:=====
54 946344   906766:=====
```

```
56 808341 757427:=====*=
58 674642 621833:=====*=
60 547894 503722:=====*=
62 445336 403835:=====*=
64 351899 321168:=====*=
66 269371 253841:=====*=
68 215247 199666:=====*=
70 167264 156470:=====*=
72 133602 122267:=====*=
74 103678 95327:=====*=
76 78538 74195:=====*=
78 65594 57671:=====*=
80 51573 44781:=====*=
82 37364 34256:=====*=
84 27602 27135:=====*=
86 24553 20996:=====*=
88 18483 16245:=====*=
90 11441 12570:=====*=
92 8782 9726:=====*=
94 7356 7525:=====*=
96 5284 5823:=====*=
98 3609 4505:=====*=
100 2945 3486:=====*=
102 2089 2697:=====*=
104 1456 2087:=====*=
106 2227 1615:=====*=
108 2720 1249:=====*=
110 1248 967:=====*=
112 762 748:=====*=
114 407 579:=====*=
116 187 448:=====*=
118 104 347:=====*=
>120 381 268:=====*=

inset = represents 176 library sequences

3787527556 residues in 14717352 library sequences
statistics sampled from 60000 to 14714101 sequences
Expectation_n fit: rho(ln(x))= 3.18550.000177; mu= 6.9701 0.010
mean_var=23.2784 4.682, 0's: 879 Z-trim: 888 B-trim: 2689 in 1/62
Lambda= 0.265826
Kolmogorov-Smirnov statistic: 0.0413 (N=29) at 48

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
join: 42, opt: 30, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000

20 residues in 1 query sequences
3787527556 residues in 14717352 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:26:59 2009 done: Wed Sep 23 13:32:23 2009
Total Scan time: 298.990 Total Display time: 0.000
```

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

Appendix 6. Bioinformatic analysis of polypeptide 5_6a

>5_6a
RVLQLSYFFQDNGALAQRPNRSDNNTLRSFVSRS

Sliding 8 amino acid window search
Database searched = AD_2009
Query = 5_6a

Start time: Wed Sep 23 13:32:25 CDT 2009 Finish time: Wed Sep 23 13:32:26 CDT 2009

No 8 amino acid matches exist between 5_6a and the AD_2009 database

fasta34 5_6a.pep /home/andre/db/AD_2009 -Q -E 1 -O 5_6a.pep_ad.fasta
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

5_6a, 34 aa
vs /home/andre/db/AD_2009 library

	opt	E()
< 20	3	0:=
22	0	0:=
24	2	0:=
26	0	0:=
28	2	0:=
30	5	2:*=
32	17	7:==*==
34	39	20:=====*
36	41	41:=====*
38	72	68:=====*=
40	54	95:=====*
42	112	116:=====*
44	99	128:=====*
46	124	130:=====*
48	142	125:=====*
50	145	114:=====*
52	93	100:=====*
54	95	85:=====*
56	75	71:=====*

one = represents 3 library sequences

```
58 51 59:===== *
60 30 47:===== *
62 38 38:=====*
64 28 30:=====*
66 26 24:=====*=
68 34 19:=====*=
70 13 15:=====*
72 7 12:=====*
74 7 9:=====*
76 6 7:=====*
78 3 5:=====*
80 8 4:=====*
82 2 3:=====*
84 0 3:=====*
86 0 2:=====*
88 0 2:=====*
90 0 1:=====*
92 5 1:=====*:=====
94 7 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 1 0:=====*:=====
114 0 0:=====*:=====
116 0 0:=====*:=====
118 0 0:=====*:=====
>120 0 0:=====*:=====
307888 residues in 1386 sequences
Expectation_n fit: rho(ln(x))= 3.67260.00286; mu= 6.7188 0.148
mean_var=21.9777 6.501, 0's: 3 Z-trim: 4 B-trim: 50 in 1/43
Lambda= 0.273579
Kolmogorov-Smirnov statistic: 0.0292 (N=26) at 46

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
join: 42, opt: 30, open/ext: -10/-2, width: 16
The best scores are:
gi|729979|sp|P39673.1|MAG_DERFA Allergen Mag (341) 57 25.7 0.29

>>gi|729979|sp|P39673.1|MAG_DERFA Allergen Mag (341 aa)
initn: 55 initl: 55 opt: 57 Z-score: 111.6 bits: 25.7 E(): 0.29
Smith-Waterman score: 57; 32.258% identity (70.968% similar) in 31 aa
overlap (3-33:169-198)
```

10 20 30

```
5_6a RVLQLSYFFQDNGALAQRPNSRSDNNTLRSFVS
gi|729 DIDGTLIEGHAQGTIREGKIHIKGRQTDFFEIESNYRYED-GKLIIEPVKSENGKLEGVLS
140 150 160 170 180 190

5_6a RS
:
gi|729 RKVPSHLTLETPRVKMNMKYDRYAPVKVFKLDYDGIHFEKHTDIEYEPGVRYKIIGNGKL
200 210 220 230 240 250
```

34 residues in 1 query sequences
307888 residues in 1386 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:32:25 2009 done: Wed Sep 23 13:32:25 2009
Total Scan time: 0.040 Total Display time: 0.000

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

fasta34 5_6a.pep /home/andre/db/TOX_2009 -Q -E 1 -O 5_6a.pep_tx.fasta
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

5_6a, 34 aa
vs /home/andre/db/TOX_2009 library

```
opt E()
< 20 59 0:=====
22 3 0:===== one = represents 13 library sequences
24 2 0:=====
26 3 0:=====
28 21 2:=====
30 30 11:=====
32 58 41:=====
34 150 110:=====*=
36 194 227:===== *
38 357 375:=====*=
40 373 523:===== *
42 571 639:===== *
44 693 705:=====*=
46 632 718:===== *
48 668 688:=====*=
50 701 627:=====*=
```

```
52 749 552:=====*=====
54 556 471:=====*=====
56 354 394:===== *
58 235 323:===== *
60 210 262:===== *
62 183 210:===== *
64 190 167:=====*==
66 130 132:=====*
68 63 104:===== *
70 74 81:=====*
72 57 64:=====*
74 74 50:=====*==
76 38 39:=====*
78 38 30:=====*
80 33 23:=====*
82 66 18:=====*
84 15 14:=====*
86 24 11:=====*
88 10 8:*          inset = represents 1 library sequences
90 11 7:*
92 2 5:*          :== *
94 2 4:*          :== *
96 2 3:*          :==*
98 6 2:*          :==*====
100 5 2:*          :==*====
102 0 1:*          :*
104 0 1:*          :*
106 3 1:*          :*==
108 0 1:*          :*
110 0 1:*          :*
112 0 0:*          :*
114 1 0:*          :*
116 0 0:*          :*
118 0 0:*          :*
>120 0 0:*          :*
1891534 residues in 7651 sequences
Expectation_n fit: rho(ln(x))= 4.69370.000628; mu= 2.1077 0.032
mean_var=19.5607 3.978, 0's: 59 Z-trim: 59 B-trim: 169 in 1/61
Lambda= 0.289989
Kolmogorov-Smirnov statistic: 0.0374 (N=29) at 48

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 1
join: 42, opt: 30, open/ext: -10/-2, width: 16
!! No sequences with E() < 1.000000

34 residues in 1 query sequences
1891534 residues in 7651 library sequences
Scomplib [34t26]
start: Wed Sep 23 13:32:26 2009 done: Wed Sep 23 13:32:26 2009
```

Total Scan time: 0.210 Total Display time: 0.000

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

```
# fasta34 5_6a.pep /home/andre/db/PRT_2009 -Q -E 1 -O 5_6a.pep_prt.fasta
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

5_6a, 34 aa

vs /home/andre/db/PRT_2009 library

```

      opt      E()
< 20 221349    0:=====
22 472    0:===== one = represents 22748 library sequences
24 888    14:*
26 2298   309:*
28 7576  3337:*
30 28092 20271:*
32 90606 78383:====*
34 220787 212565:=====*
36 412481 436559:=====*
38 662109 721470:===== *
40 932950 1006387:===== *
42 1148496 1230185:===== *
44 1297975
1357009:===== *
46 1364861
1382148:===== *
48 1323878
1323247:===== *
50 1215325 1207468:===== *
52 1063042 1061566:===== *
54 898602 906763:===== *
56 736329 757425:===== *
58 599824 621832:===== *
60 494125 503720:===== *
62 406130 403834:===== *
64 337995 321167:===== *
66 263248 253841:===== *
68 221434 199666:===== *
70 172094 156470:===== *
72 139492 122266:===== *
74 104681 95327:===== *
76 82199 74195:===== *
78 63589 57671:===== *
80 50707 44780:===== *
```



```
Wed Sep 23 13:42:28 CDT 2009      a155ebc7632842e917ba5dd4cce1dc09
/home/andre/db/PRT 2009
```

```
Wed Sep 23 13:40:18 CDT 2009      5c91759664b0377022cf35e7d5a0d23c
/home/andre/db/AD_2009

Wed Sep 23 13:40:18 CDT 2009      fa2987b358e662a0d03802f0cfba9676
/home/andre/db/TOX_2009
```