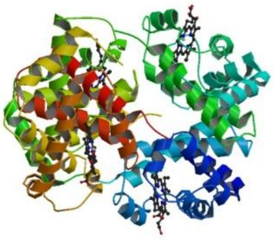
Εργασία Βιοπληροφορικής
Υπεύθυνη Καθηγήτρια: Χολή-Παπαδοπούλου Θεοδώρα

1HBS



ΠΑΠΑΖΟΓΛΟΥ ΒΑΣΙΛΙΚΗ

ΑΕΜ : 8355

Fasta sequence

>sp|P69905|HBA_HUMAN Hemoglobin subunit alpha OS=Homo sapiens GN=HBA1 PE=1 SV=2
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP
AVHASLDKFLASVSTVLTSKYR

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MVLSPADKTN VKAAWGKVGA HAGEYGAEAL ERMFLSFPTT KTYFPHFDLS HGSAQVKGHG

        70         80         90        100        110        120
KKVADALTNA VAHVDDMPNA LSALSDLHAH KLRVDPVNFK LLSHCLLVTL AAHLPAEFTP

       130        140
AVHASLDKFL ASVSTVLTSK YR
```

References and documentation are available.

Number of amino acids: 142

Molecular weight: 15257.5

Theoretical pI: 8.72

Amino acid composition: CSV format

Ala (A)	21	14.8%
Arg (R)	3	2.1%
Asn (N)	4	2.8%
Asp (D)	8	5.6%
Cys (C)	1	0.7%
Gln (Q)	1	0.7%
Glu (E)	4	2.8%
Gly (G)	7	4.9%
His (H)	10	7.0%
Ile (I)	0	0.0%
Leu (L)	18	12.7%
Lys (K)	11	7.7%
Met (M)	3	2.1%
Phe (F)	7	4.9%
Pro (P)	7	4.9%
Ser (S)	11	7.7%
Thr (T)	9	6.3%
Trp (W)	1	0.7%
Tyr (Y)	3	2.1%
Val (V)	13	9.2%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 12
Total number of positively charged residues (Arg + Lys): 14

Atomic composition:

Carbon	C	690
Hydrogen	H	1080

Nitrogen	N	188
Oxygen	O	195
Sulfur	S	4

Formula: $C_{690}H_{1080}N_{188}O_{195}S_4$
Total number of atoms: 2157

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 9970
Abs 0.1% (=1 g/l) 0.653, assuming all pairs of Cys residues form cystines

Ext. coefficient 9970
Abs 0.1% (=1 g/l) 0.653, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

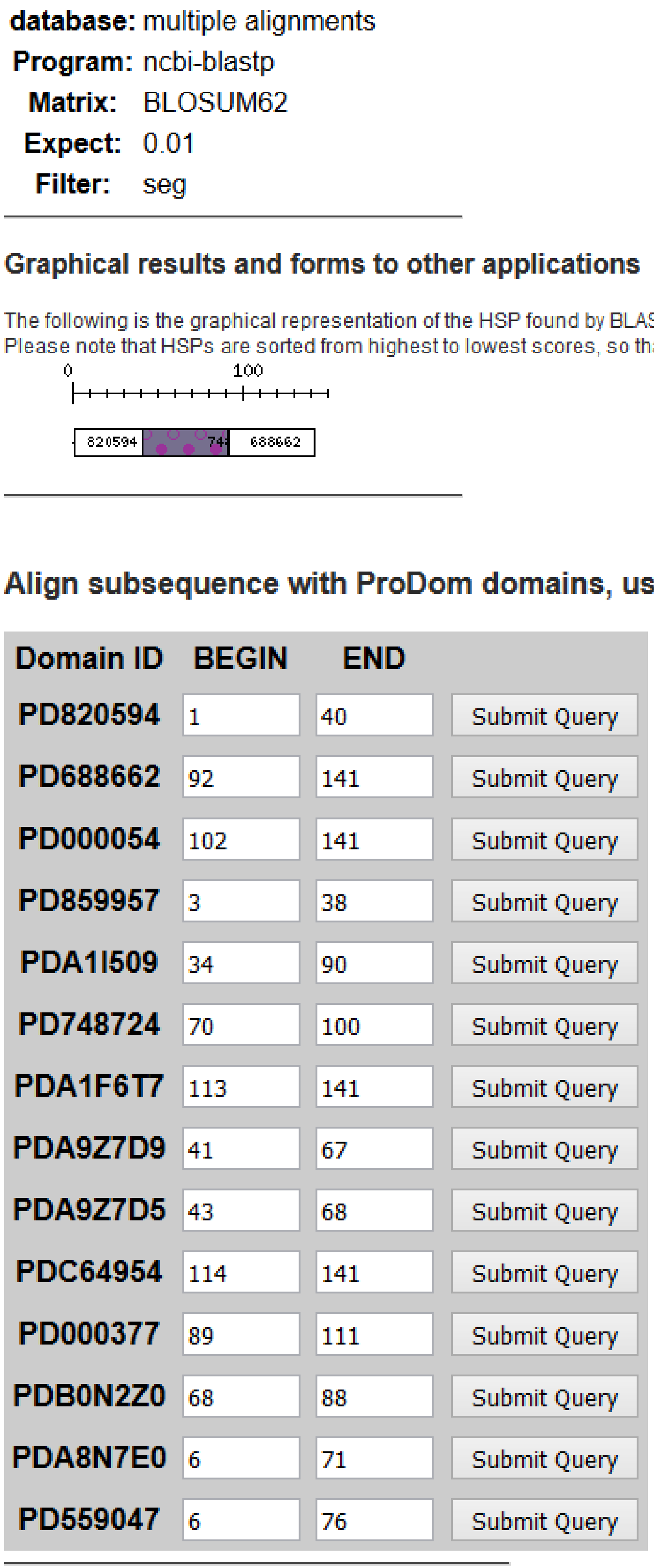
Instability index:

The instability index (II) is computed to be 6.97
This classifies the protein as stable.

Aliphatic index: 90.77

Grand average of hydropathicity (GRAVY): 0.048

ProDom



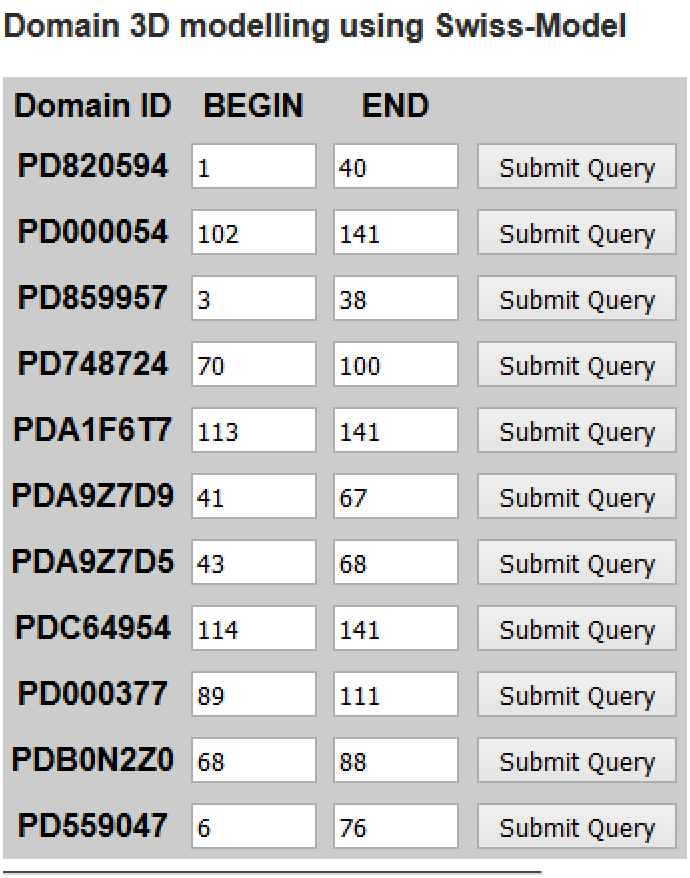
database: multiple alignments
Program: ncbi-blastp
Matrix: BLOSUM62
Expect: 0.01
Filter: seg

Graphical results and forms to other applications

The following is the graphical representation of the HSP found by BLAST.
Please note that HSPs are sorted from highest to lowest scores, so that lower scoring HSPs may be hidden.

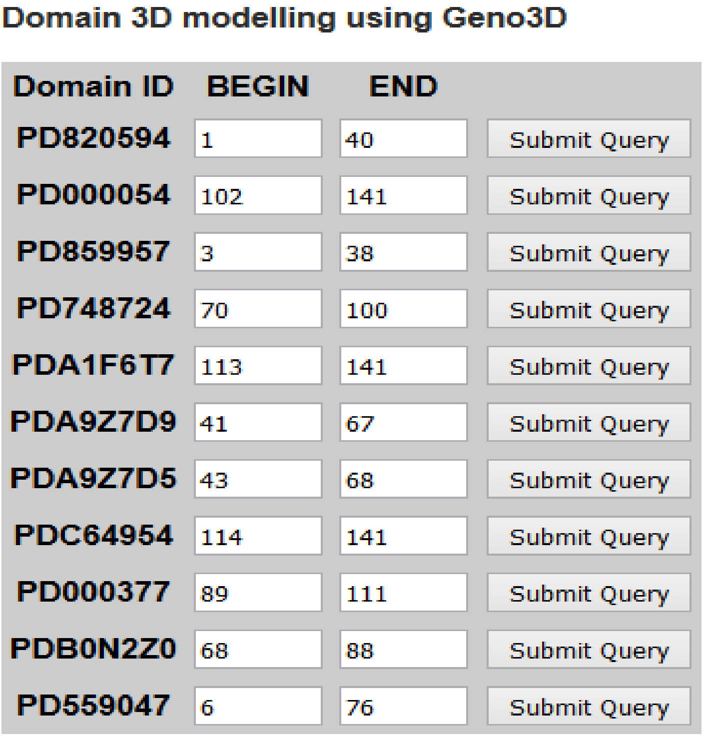
Align subsequence with ProDom domains, using Multalin

Domain ID	BEGIN	END	
PD820594	1	40	Submit Query
PD688662	92	141	Submit Query
PD000054	102	141	Submit Query
PD859957	3	38	Submit Query
PDA1I509	34	90	Submit Query
PD748724	70	100	Submit Query
PDA1F6T7	113	141	Submit Query
PDA9Z7D9	41	67	Submit Query
PDA9Z7D5	43	68	Submit Query
PDC64954	114	141	Submit Query
PD000377	89	111	Submit Query
PDB0N2Z0	68	88	Submit Query
PDA8N7E0	6	71	Submit Query
PD559047	6	76	Submit Query



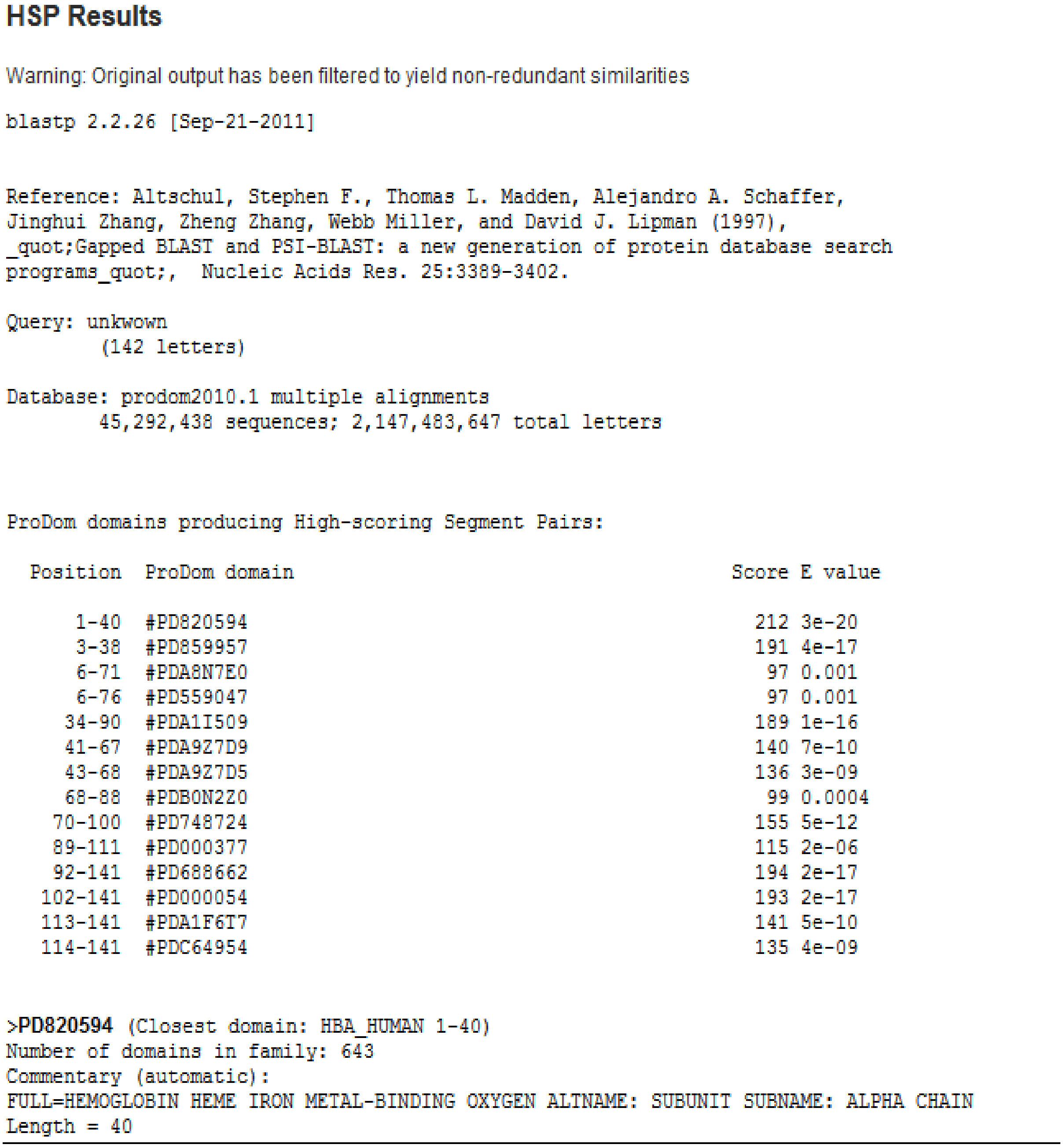
Domain 3D modelling using Swiss-Model

Domain ID	BEGIN	END	
PD820594	1	40	Submit Query
PD000054	102	141	Submit Query
PD859957	3	38	Submit Query
PD748724	70	100	Submit Query
PDA1F6T7	113	141	Submit Query
PDA9Z7D9	41	67	Submit Query
PDA9Z7D5	43	68	Submit Query
PDC64954	114	141	Submit Query
PD000377	89	111	Submit Query
PDB0N2Z0	68	88	Submit Query
PD559047	6	76	Submit Query



Domain 3D modelling using Geno3D

Domain ID	BEGIN	END	
PD820594	1	40	Submit Query
PD000054	102	141	Submit Query
PD859957	3	38	Submit Query
PD748724	70	100	Submit Query
PDA1F6T7	113	141	Submit Query
PDA9Z7D9	41	67	Submit Query
PDA9Z7D5	43	68	Submit Query
PDC64954	114	141	Submit Query
PD000377	89	111	Submit Query
PDB0N2Z0	68	88	Submit Query
PD559047	6	76	Submit Query



HSP Results

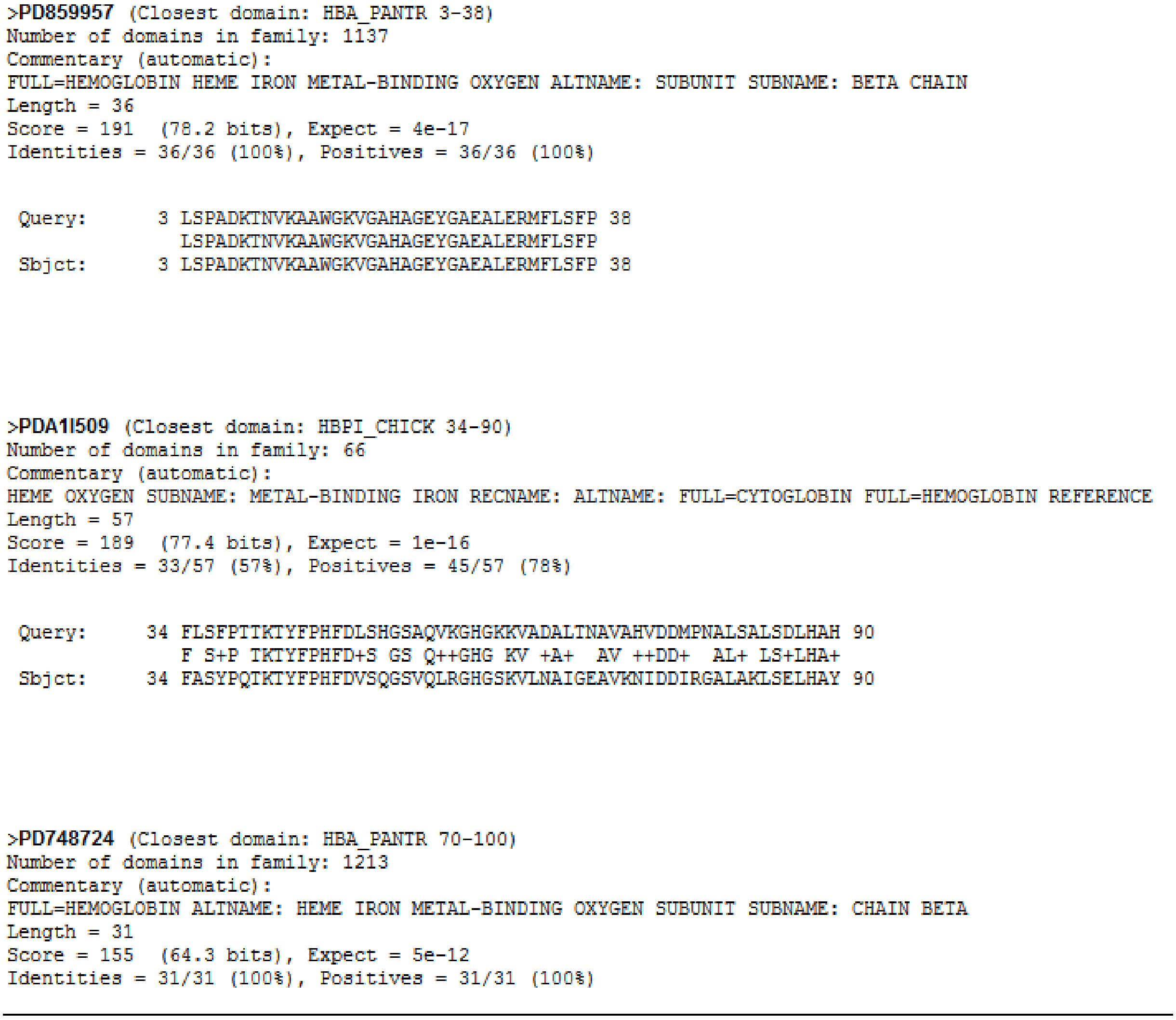
Warning: Original output has been filtered to yield non-redundant similarities

```
blastp 2.2.26 [Sep-21-2011]


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

Query: unkwown
        (142 letters)

Database: prodom2010.1 multiple alignments
        45,292,438 sequences; 2,147,483,647 total letters



ProDom domains producing High-scoring Segment Pairs:

  Position  ProDom domain                                      Score E value

      1-40  #PD820594                                            212 3e-20
      3-38  #PD859957                                            191 4e-17
      6-71  #PDA8N7E0                                             97 0.001
      6-76  #PD559047                                             97 0.001
     34-90  #PDA1I509                                            189 1e-16
     41-67  #PDA9Z7D9                                            140 7e-10
     43-68  #PDA9Z7D5                                            136 3e-09
     68-88  #PDB0N2Z0                                             99 0.0004
    70-100  #PD748724                                            155 5e-12
    89-111  #PD000377                                            115 2e-06
    92-141  #PD688662                                            194 2e-17
   102-141  #PD000054                                            193 2e-17
   113-141  #PDA1F6T7                                            141 5e-10
   114-141  #PDC64954                                            135 4e-09


>PD820594 (Closest domain: HBA_HUMAN 1-40)
Number of domains in family: 643
Commentary (automatic):
FULL=HEMOGLOBIN HEME IRON METAL-BINDING OXYGEN ALTNAME: SUBUNIT SUBNAME: ALPHA CHAIN
Length = 40
```

```
Score = 212  (86.3 bits), Expect = 3e-20
Identities = 40/40 (100%), Positives = 40/40 (100%)


 Query:      1 MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTT 40
               MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTT
 Sbjct:      1 MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTT 40
```

```
>PD688662 (Closest domain: HBPI_CHICK 91-141)
Number of domains in family: 19
Commentary (automatic):
HEME OXYGEN FULL=HEMOGLOBIN ALTNAME: METAL-BINDING IRON RECNAME: SUBNAME: SEQUENCING DIRECT
Length = 51
Score = 194  (79.3 bits), Expect = 2e-17
Identities = 35/50 (70%), Positives = 44/50 (88%)


 Query:     92 LRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY 141
               LRVDPVNFKLLSHC+L ++AA  P++FTP VHA+ DKFL+S+S+VLT KY
 Sbjct:     92 LRVDPVNFKLLSHCILCSVAARYPSDFTPEVHAAWDKFLSSISSVLTEKY 141
```

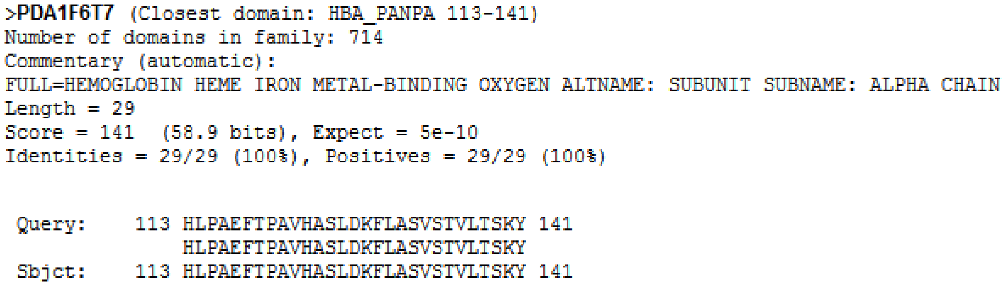
```
>PD000054 (Closest domain: G3V1N2_HUMAN 70-109)
Number of domains in family: 776
Commentary (automatic):
FULL=HEMOGLOBIN ALTNAME: HEME IRON METAL-BINDING OXYGEN BETA SUBUNIT CHAIN SUBNAME:
Length = 40
Score = 193  (79.0 bits), Expect = 2e-17
Identities = 40/40 (100%), Positives = 40/40 (100%)


 Query:    102 LSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY 141
               LSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY
 Sbjct:     70 LSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY 109
```



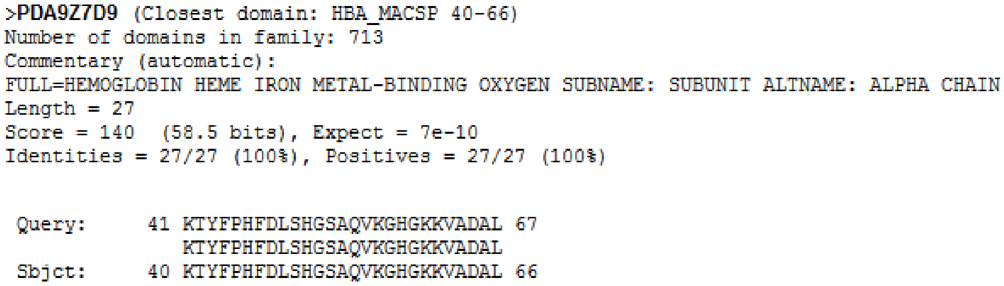
```
>PD859957 (Closest domain: HBA_PANTR 3-38)
Number of domains in family: 1137
Commentary (automatic):
FULL=HEMOGLOBIN HEME IRON METAL-BINDING OXYGEN ALTNAME: SUBUNIT SUBNAME: BETA CHAIN
Length = 36
Score = 191  (78.2 bits), Expect = 4e-17
Identities = 36/36 (100%), Positives = 36/36 (100%)


 Query:      3 LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFP 38
               LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFP
 Sbjct:      3 LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFP 38
```

```
>PDA1I509 (Closest domain: HBPI_CHICK 34-90)
Number of domains in family: 66
Commentary (automatic):
HEME OXYGEN SUBNAME: METAL-BINDING IRON RECNAME: ALTNAME: FULL=CYTOGLOBIN FULL=HEMOGLOBIN REFERENCE
Length = 57
Score = 189  (77.4 bits), Expect = 1e-16
Identities = 33/57 (57%), Positives = 45/57 (78%)


 Query:     34 FLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAH 90
               F S+P TKTYFPHFD+S GS Q++GHG KV +A+  AV ++DD+  AL+ LS+LHA+
 Sbjct:     34 FASYPQTKTYFPHFDVSQGSVQLRGHGSKVLNAIGEAVKNIDDIRGALAKLSELHAY 90
```

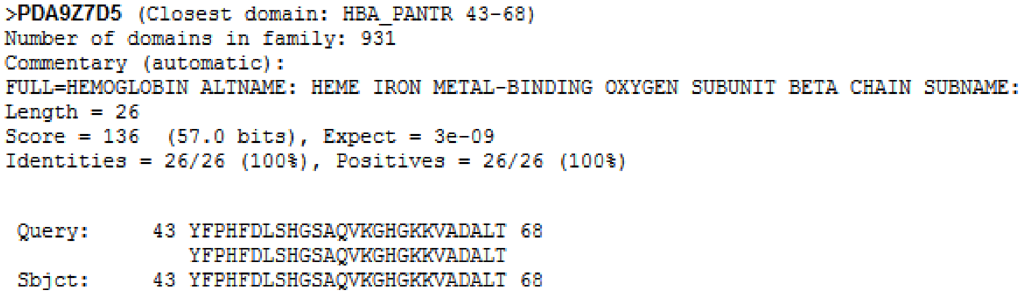
```
>PD748724 (Closest domain: HBA_PANTR 70-100)
Number of domains in family: 1213
Commentary (automatic):
FULL=HEMOGLOBIN ALTNAME: HEME IRON METAL-BINDING OXYGEN SUBUNIT SUBNAME: CHAIN BETA
Length = 31
Score = 155  (64.3 bits), Expect = 5e-12
Identities = 31/31 (100%), Positives = 31/31 (100%)
```



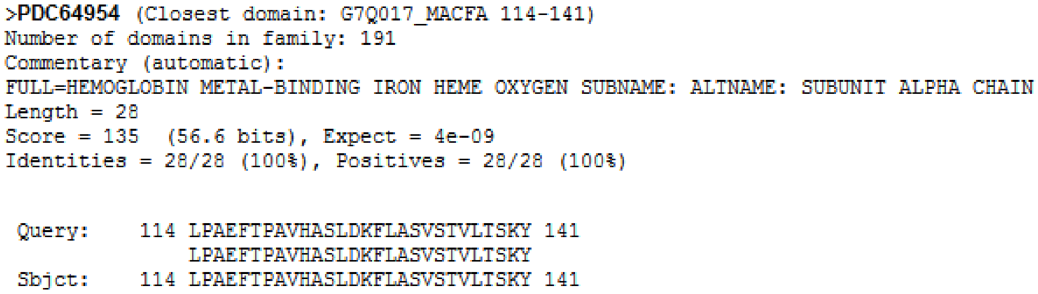
```
>PDA1F6T7 (Closest domain: HBA_PANPA 113-141)
Number of domains in family: 714
Commentary (automatic):
FULL=HEMOGLOBIN HEME IRON METAL-BINDING OXYGEN ALTNAME: SUBUNIT SUBNAME: ALPHA CHAIN
Length = 29
Score = 141  (58.9 bits), Expect = 5e-10
Identities = 29/29 (100%), Positives = 29/29 (100%)


 Query:    113 HLPAEFTPAVHASLDKFLASVSTVLTSKY 141
               HLPAEFTPAVHASLDKFLASVSTVLTSKY
 Sbjct:    113 HLPAEFTPAVHASLDKFLASVSTVLTSKY 141
```



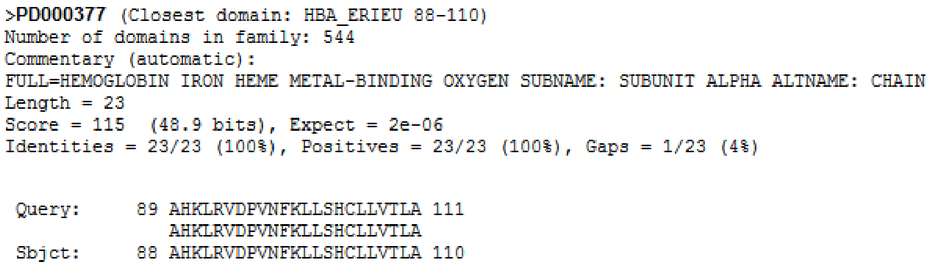
```
>PDA9Z7D9 (Closest domain: HBA_MACSP 40-66)
Number of domains in family: 713
Commentary (automatic):
FULL=HEMOGLOBIN HEME IRON METAL-BINDING OXYGEN SUBNAME: SUBUNIT ALTNAME: ALPHA CHAIN
Length = 27
Score = 140  (58.5 bits), Expect = 7e-10
Identities = 27/27 (100%), Positives = 27/27 (100%)


 Query:     41 KTYFPHFDLSHGSAQVKGHGKKVADAL 67
               KTYFPHFDLSHGSAQVKGHGKKVADAL
 Sbjct:     40 KTYFPHFDLSHGSAQVKGHGKKVADAL 66
```



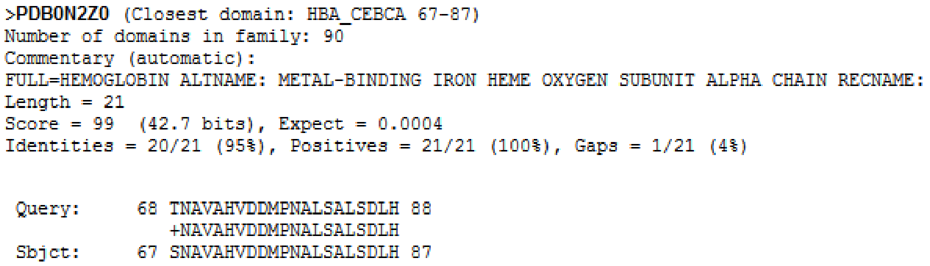
```
>PDA9Z7D5 (Closest domain: HBA_PANTR 43-68)
Number of domains in family: 931
Commentary (automatic):
FULL=HEMOGLOBIN ALTNAME: HEME IRON METAL-BINDING OXYGEN SUBUNIT BETA CHAIN SUBNAME:
Length = 26
Score = 136  (57.0 bits), Expect = 3e-09
Identities = 26/26 (100%), Positives = 26/26 (100%)


 Query:     43 YFPHFDLSHGSAQVKGHGKKVADALT 68
               YFPHFDLSHGSAQVKGHGKKVADALT
 Sbjct:     43 YFPHFDLSHGSAQVKGHGKKVADALT 68
```



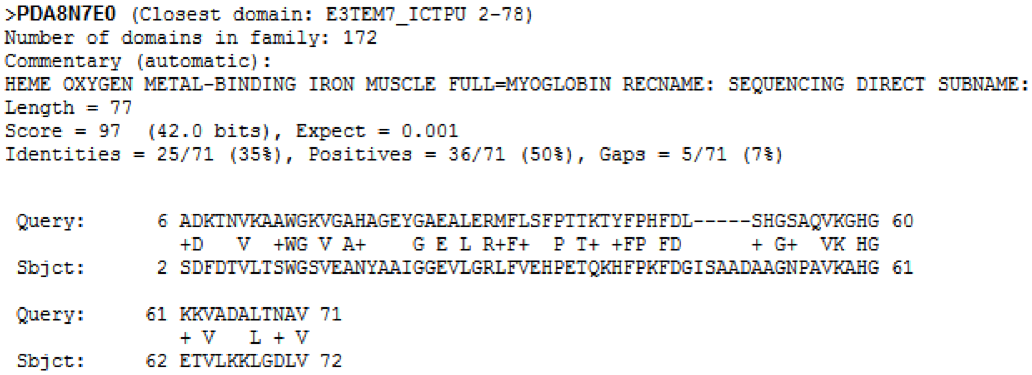
```
>PDC64954 (Closest domain: G7Q017_MACFA 114-141)
Number of domains in family: 191
Commentary (automatic):
FULL=HEMOGLOBIN METAL-BINDING IRON HEME OXYGEN SUBNAME: ALTNAME: SUBUNIT ALPHA CHAIN
Length = 28
Score = 135  (56.6 bits), Expect = 4e-09
Identities = 28/28 (100%), Positives = 28/28 (100%)


 Query:    114 LPAEFTPAVHASLDKFLASVSTVLTSKY 141
               LPAEFTPAVHASLDKFLASVSTVLTSKY
 Sbjct:    114 LPAEFTPAVHASLDKFLASVSTVLTSKY 141
```



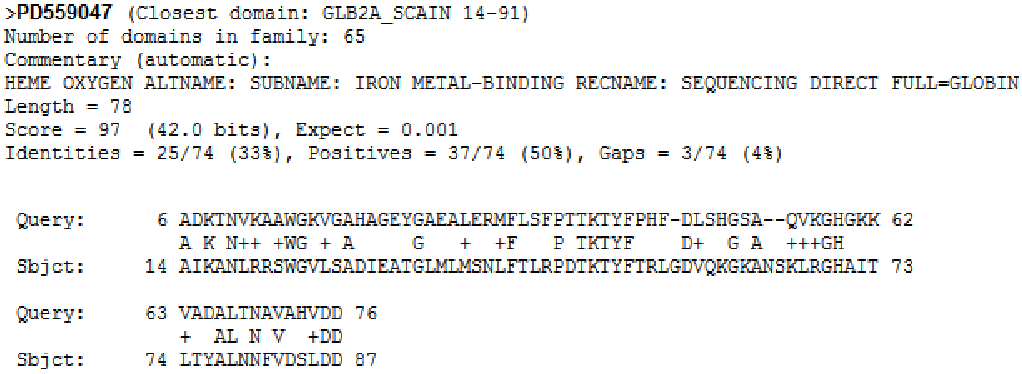
```
>PD000377 (Closest domain: HBA_ERIEU 88-110)
Number of domains in family: 544
Commentary (automatic):
FULL=HEMOGLOBIN IRON HEME METAL-BINDING OXYGEN SUBNAME: SUBUNIT ALPHA ALTNAME: CHAIN
Length = 23
Score = 115  (48.9 bits), Expect = 2e-06
Identities = 23/23 (100%), Positives = 23/23 (100%), Gaps = 1/23 (4%)


 Query:     89 AHKLRVDPVNFKLLSHCLLVTLA 111
               AHKLRVDPVNFKLLSHCLLVTLA
 Sbjct:     88 AHKLRVDPVNFKLLSHCLLVTLA 110
```



```
>PDB0N2Z0 (Closest domain: HBA_CEBCA 67-87)
Number of domains in family: 90
Commentary (automatic):
FULL=HEMOGLOBIN ALTNAME: METAL-BINDING IRON HEME OXYGEN SUBUNIT ALPHA CHAIN RECNAME:
Length = 21
Score = 99  (42.7 bits), Expect = 0.0004
Identities = 20/21 (95%), Positives = 21/21 (100%), Gaps = 1/21 (4%)


 Query:     68 TNAVAHVDDMPNALSALSDLH 88
               +NAVAHVDDMPNALSALSDLH
 Sbjct:     67 SNAVAHVDDMPNALSALSDLH 87
```

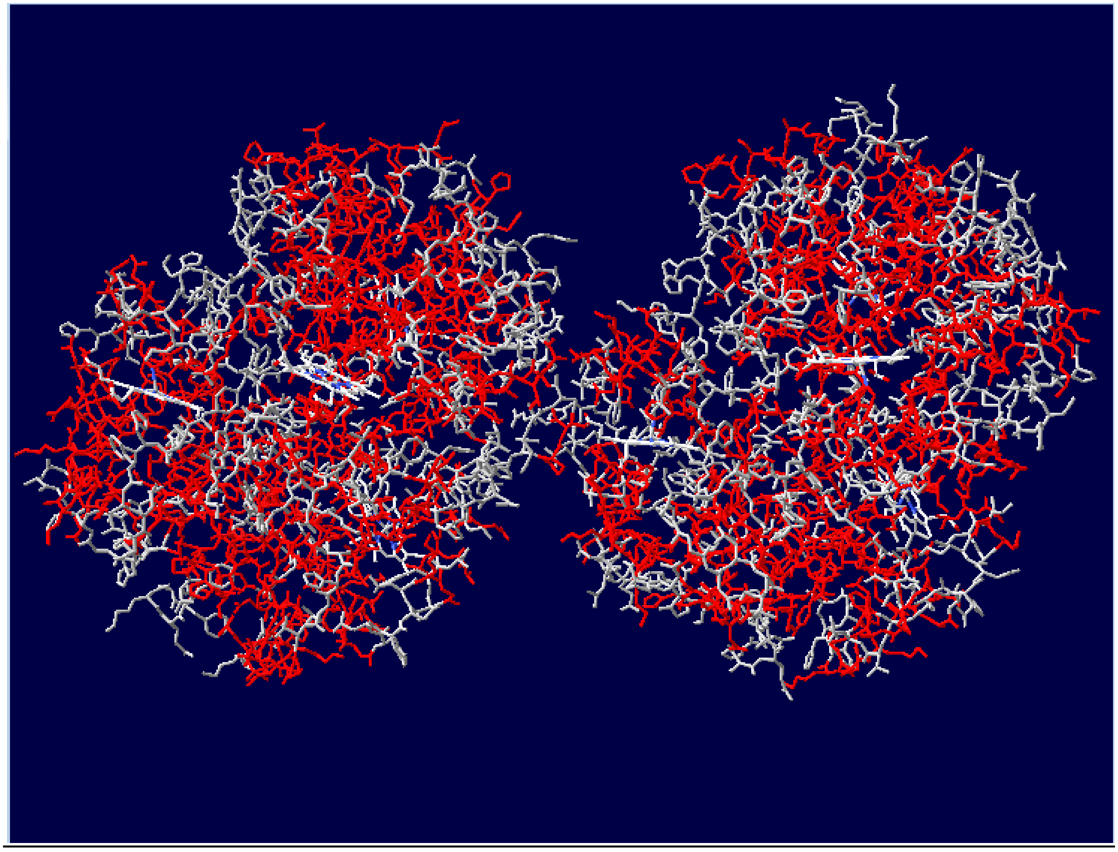


```
>PDA8N7E0 (Closest domain: E3TEM7_ICTPU 2-78)
Number of domains in family: 172
Commentary (automatic):
HEME OXYGEN METAL-BINDING IRON MUSCLE FULL=MYOGLOBIN RECNAME: SEQUENCING DIRECT SUBNAME:
Length = 77
Score = 97  (42.0 bits), Expect = 0.001
Identities = 25/71 (35%), Positives = 36/71 (50%), Gaps = 5/71 (7%)


 Query:      6 ADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDL-----SHGSAQVKGHG 60
               +D   V  +WG V A+    G E L R+F+  P T+ +FP FD      + G+  VK HG
 Sbjct:      2 SDFDTVLTSWGSVEANYAAIGGEVLGRLFVEHPETQKHFPKFDGISAADAAGNPAVKAHG 61

 Query:     61 KKVADALTNAV 71
               + V   L + V
 Sbjct:     62 ETVLKKLGDLV 72
```

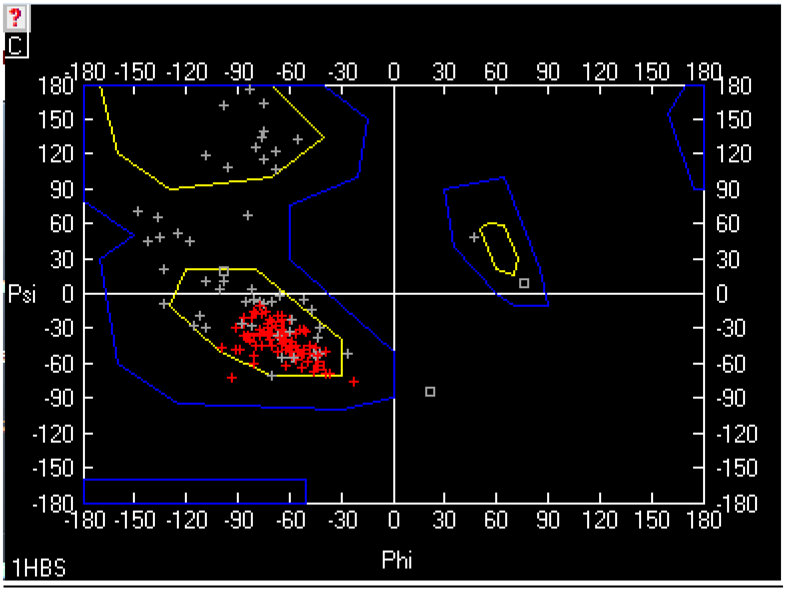


```
>PD559047 (Closest domain: GLB2A_SCAIN 14-91)
Number of domains in family: 65
Commentary (automatic):
HEME OXYGEN ALTNAME: SUBNAME: IRON METAL-BINDING RECNAME: SEQUENCING DIRECT FULL=GLOBIN
Length = 78
Score = 97  (42.0 bits), Expect = 0.001
Identities = 25/74 (33%), Positives = 37/74 (50%), Gaps = 3/74 (4%)


 Query:      6 ADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLSHGSA--QVKGHGKK 62
               A K N++ +WG + A     G   +   +F  P TKTYF    D+  G A  +++GH
 Sbjct:     14 AIKANLRRSWGVLSADIEATGLMLMSNLFTLRPDTKTYFTRLGDVQKGKANSKLRGHAIT 73

 Query:     63 VADALTNAVAHVDD 76
               +  AL N V  +DD
 Sbjct:     74 LTYALNNFVDSLDD 87
```

1HBS
(render by 3D/colored by secondary structure)



Ramachandran Plot



Οι διαμορφώσεις των πεπτιδίων προσδιορίζονται από τις τιμές των γωνιών φ και ψ. Εφικτές θεωρούνται μόνο οι διαμορφώσεις που περιλαμβάνουν μικρή ή καθόλου στερεοτακτική παρεμβολή, με βάση υπολογισμούς που χρησιμοποιούν γνωστές ακτίνες van der Waals και γωνίες δεσμών. Οι τιμές που επιτρέπονται μπορούν να τοποθετηθούν σε ένα διδιάστατο σχήμα που ονομάζεται διάγραμμα Ramachandran. Από το Ramachandran plot παρατηρούμε ότι η GLY18 είναι εκτός των επιτρεπτών μεταβολών. Επειδή η πλευρική αλυσίδα της γλυκίνης είναι αρκετά μικρή αφού έχει ένα μόνο άτομο υδρογόνου, η γλυκίνη μπορεί να συμμετέχει σε πολλές διαμορφώσεις μη αποδεκτές για άλλα αμινοξέα και να κάνει πολλές κινήσεις. Στο Ramachandran οι επιτρεπτές μεταβολές των αμινοξέων είναι αυτές που βρίσκονται εντός των κίτρινων και μπλέ καμπυλών.