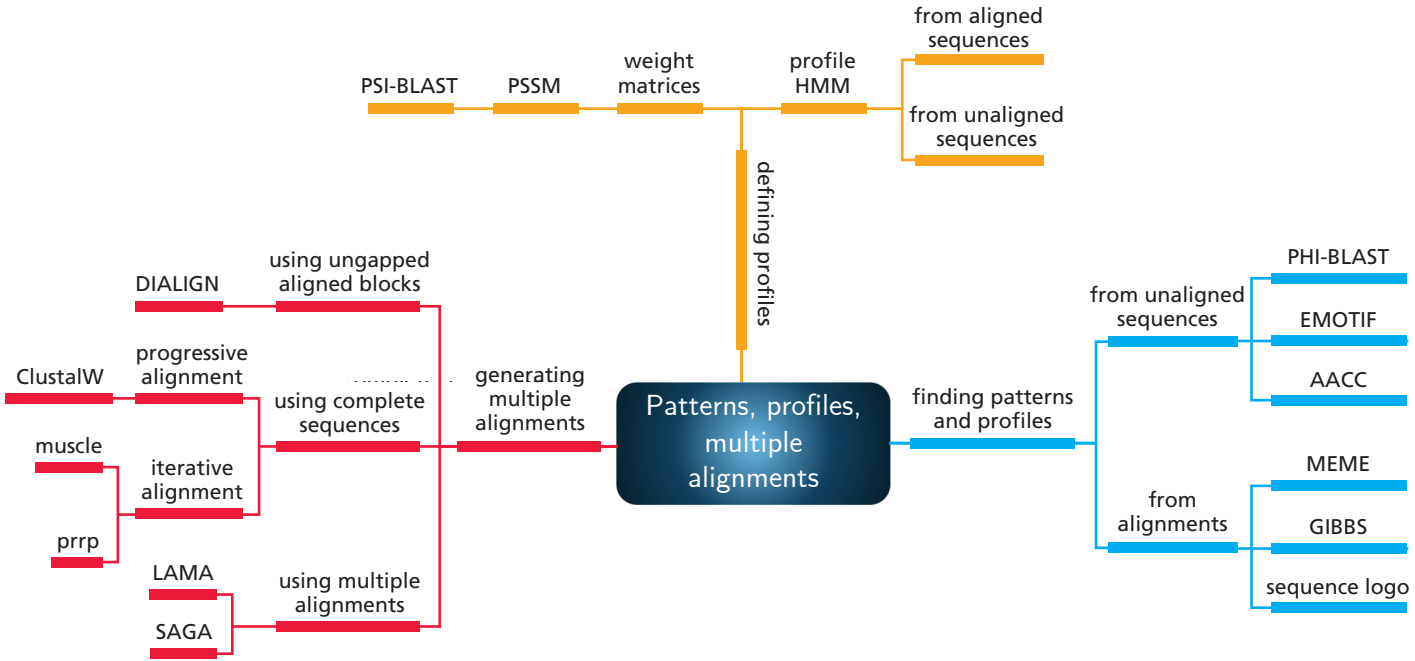
Position Specific Score Matrix (PSSM) & PSI-BLAST

Hamim Zafar

BSE633, Mar 04, 2020

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	0	0	0	0	0	0	0	0	0	0	0	2	1	0	2
C	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	5	0	1	0	0	0	1	0	0	0	0	1	0
F	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
G	5	0	0	2	0	5	1	0	1	0	2	3	1	1	0
H	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
K	0	0	0	1	1	0	1	1	0	1	0	0	0	0	0
L	0	0	0	1	0	0	0	0	0	0	1	0	2	0	0
M	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	1	0	0	1	0	1	1	0	0	0	0
S	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
T	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
V	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0

IIT Kanpur
Biological Sciences & Bioengineering


Patterns, profiles, multiple alignments
defining profiles
PSI-BLAST
PSSM
weight matrices
profile HMM
from aligned sequences
from unaligned sequences
generating multiple alignments
using ungapped aligned blocks
DIALIGN
using complete sequences
progressive alignment
ClustalW
iterative alignment
muscle
prrp
using multiple alignments
LAMA
SAGA
finding patterns and profiles
from unaligned sequences
PHI-BLAST
EMOTIF
AACC
from alignments
MEME
GIBBS
sequence logo

Sequence Families

- Functional biological sequences typically come in families – these sequences will be expected to have **similar properties at equivalent regions**
- Sequences in a family have **diverged during evolution**, but normally maintain the **same or a related function**
- Thus, identifying that a sequence belongs to a family tells about its function

Sequences from a Globin Family

```
Helix                AAAAAAAAAAAAAAAA   BBBBBBBBBBBBBBBBCCCCCCCCCCCC
HBA_HUMAN  ---------VLSPADKTNVKAAWGKVGA--HAGEYGAEALERMFLSFPTTKTYFPHF
HBB_HUMAN  --------VHLTPEEKSAVTALWGKV----NVDEVGGEALGRLLVVYPWTQRFFESF
MYG_PHYCA  ---------VLSEGEWQLVLHVWAKVEA--DVAGHGQDILIRLFKSHPETLEKFDRF
GLB3_CHITP ----------LSADQISTVQASFDKVKG------DPVGILYAVFKADPSIMAKFTQF
GLB5_PETMA PIVDTGSVAPLSAAEKTKIRSAWAPVYS--TYETSGVDILVKFFTSTPAAQEFFPKF
LGB2_LUPLU --------GALTESQAALVKSSWEEFNA--NIPKHTHRFFILVLEIAPAAKDLFS-F
GLB1_GLYDI ---------GLSAAQRQVIAATWKDIAGADNGAGVGKDCLIKFLSAHPQMAAVFG-F
Consensus            Ls....  v a W kv .  .    g . L.. f . P .   F  F

Helix          DDDDDDDEEEEEEEEEEEEEEEEEEE            FFFFFFFFFFFF
HBA_HUMAN  -DLS-----HGSAQVKGHGKKVADALTNAVAHV---D--DMPNALSALSDLHAHKL-
HBB_HUMAN  GDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHL---D--NLKGTFATLSELHCDKL-
MYG_PHYCA  KHLKTEAEMKASEDLKKHGVTVLTALGAILKK----K-GHHEAELKPLAQSHATKH-
GLB3_CHITP AG-KDLESIKGTAPFETHANRIVGFFSKIIGEL--P---NIEADVNTFVASHKPRG-
GLB5_PETMA KGLTTADQLKKSADVRWHAERIINAVNDAVASM--DDTEKMSMKLRDLSGKHAKSF-
LGB2_LUPLU LK-GTSEVPQNNPELQAHAGKVFKLVYEAAIQLQVTGVVVTDATLKNLGSVHVSKG-
GLB1_GLYDI SG----AS---DPGVAALGAKVLAQIGVAVSHL--GDEGKMVAQMKAVGVRHKGYGN
Consensus   .  t    .. . v..Hg kv. a   a...l   d   . a l. l   H  .

Helix       FFGGGGGGGGGGGGGGGGGG     HHHHHHHHHHHHHHHHHHHHHHHHHH
HBA_HUMAN  -RVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR------
HBB_HUMAN  -HVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH------
MYG_PHYCA  -KIPIKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG
GLB3_CHITP --VTHDQLNNFRAGFVSYMKAHT--DFA-GAEAAWGATLDTFFGMIFSKM-------
GLB5_PETMA -QVDPQYFKVLAAVIADTVAAG---------DAGFEKLMSMICILLRSAY-------
LGB2_LUPLU --VADAHFPVVKEAILKTIKEVVGAKWSEELNSAWTIAYDELAIVIKKEMNDAA---
GLB1_GLYDI KHIKAQYFEPLGASLLSAMEHRIGGKMNAAAKDAWAAAYADISGALISGLQS-----
Consensus    v.   f  l . ..  ....    f   . aa. k. .     l sky
```

Alignment of 7 globins

The 8 alpha helices are shown as A-H above the alignment

Helices are more conserved than the loop regions

Position Specific Score Matrix (PSSM)

- Substitution score matrix defines score $s_{a,b}$ for two residues without regard to their environment
- For finding all family members we need to account for known residue preferences at each alignment position.
- Inclusion of these position- specific preferences in the scoring scheme is achieved with the use of a scoring profile in which each alignment position has its own substitution scores.
- **Position-specific scoring matrices (PSSMs)**
  - frequencies of each residue in a specific position of a multiple alignment.

Position Specific Score Matrix (PSSM)

frequencies of each residue
in a specific position of
a multiple alignment.

$$f_{u,b} = \frac{n_{u,b}}{N_{seq}}$$

G	H	E	G	V	G	K	V	V	K	L	G	A	G	A
G	H	E	K	K	G	Y	F	E	D	R	G	P	S	A
G	H	E	G	Y	G	G	R	S	R	G	G	G	Y	S
G	H	E	F	E	G	P	K	G	C	G	A	L	Y	I
G	H	E	L	R	G	T	T	F	M	P	A	L	E	C

→

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	0	0	0	0	0	0	0	0	0	0	0	2	1	0	2
C	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	5	0	1	0	0	0	1	0	0	0	0	1	0
F	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
G	5	0	0	2	0	5	1	0	1	0	2	3	1	1	0
H	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
K	0	0	0	1	1	0	1	1	0	1	0	0	0	0	0
L	0	0	0	1	0	0	0	0	0	0	1	0	2	0	0
M	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	1	0	0	1	0	1	1	0	0	0	0
S	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
T	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
V	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0

- Column 1: $f_{1,A} = \frac{0}{5} = 0, f_{1,G} = \frac{5}{5} = 1, \ldots$
- Column 2: $f_{2,A} = \frac{0}{5} = 0, f_{2,H} = \frac{5}{5} = 1, \ldots$
- ...
- Column 15: $f_{15,A} = \frac{2}{5} = 0.4, f_{15,C} = \frac{1}{5} = 0.2, \ldots$

Position Specific Score Matrix (PSSM)

- frequencies of each residue in a specific position of a multiple alignment.

$$f_{u,b} = \frac{n_{u,b}}{N_{seq}}$$

- Score Associated with row a and column u will be

$$m_{u,a} = \sum_{\substack{\text{residue} \\ \text{types } b}} f_{u,b} s_{a,b}$$

This is not a log-odds ratio

- Log-odds form for a PSSM element

$$m_{u,a} = \log \frac{q_{u,a}}{p_a}$$

$q_{u,a}$ → probability of a occurring at u of PSSM

p_a → background frequency

PSSM: Sequence Weighting

- PSSM should represent the full range of diversity within the sequence family in an unbiased fashion.
- A partial set of family sequences will most probably be biased toward a certain subgroup
- Different sequences must be weighted, the weighting should be reduced for very similar sequences.
  - More sensitive to distant relationships.

Sequence Weighting Scheme for MSA

- Tree-based weights assume that sequences are related by an evolutionary tree
  - Need not be the case for alignments of short and distantly related sequences, where root location can be uncertain.
- In a pairwise distance method, every sequence is assumed to lie some distance away from every other sequence, or from some generalized sequence
- The sequence weights are typically applied to PSSMs in which **each position** vector is considered **independently** of all others.
  - Useful sequence weights might be based on the **diversity observed at each position** in an alignment rather than on the **diversity measured for whole sequences**.

PSSM: Position-based Sequence Weight

- PSSM: Useful sequence weights based on the diversity observed at each position in an alignment rather than on the diversity measured for whole sequences.

J. Mol. Biol. (1994) **243**, 574–578

Position-based Sequence Weights

Steven Henikoff and Jorja G. Henikoff

Sequence weighting methods have been used to reduce redundancy and emphasize diversity in multiple sequence alignment and searching applications. Each of these methods is based on a notion of distance between a sequence and an ancestral or generalized sequence. We describe a different approach, which bases weights on the diversity observed at each position in the alignment, rather than on a sequence distance measure. These position-based weights make minimal assumptions, are simple to compute, and perform well in comprehensive evaluations.

PSSM: Position-based Sequence Weight

- In an alignment column, If there are m different residues, each is assigned a weight of $1/m$.
- for each residue type, the number of sequences that have this residue at this position is counted.
- If there are n sequences with this residue, the weight becomes $1/mn$.
- An overall weight for the whole sequence can be defined by adding the individual column weights and then normalizing by the number of columns

PSSM: Position-based Sequence Weight

1 2 3 4 5
H S A P L
H T A D V
H T A E V
H T G L I
H T G V I

Pos = 1	Pos = 2	Pos = 3	Pos = 4	Pos = 5
$w(H) = \frac{1}{5}$	$w(S) = \frac{1}{2}$	$w(A) = \frac{1}{6}$	$w(P) = \frac{1}{5}$	$w(L) = \frac{1}{3}$
	$w(T) = \frac{1}{8}$	$w(G) = \frac{1}{4}$	$w(D) = \frac{1}{5}$	$w(V) = \frac{1}{6}$
			$w(E) = \frac{1}{5}$	$w(I) = \frac{1}{6}$
			$w(L) = \frac{1}{5}$	
			$w(V) = \frac{1}{5}$	

PSSM: Position-based Sequence Weight

12345		Weight
HSAPL	$= (\frac{1}{5} + \frac{1}{2} + \frac{1}{6} + \frac{1}{5} + \frac{1}{3})/5 = 0.280$	0.280
HTADV		0.171
HTAEV		0.171
HTGLI		0.188
HTGVI		0.188

Overcoming lack of Data: Pseudocounts

- If any residue type is not observed in the column, the score for aligning that residue type in that column will have the value $-\infty$
  - lack of sequence alignment data
  - exclusion of the corresponding residue at this position
- Add a small number to all observed frequencies.

$$q_{u,a} = \frac{n_{u,a} + 1}{N_{seq} + 20} \qquad q_{u,a} = \frac{\alpha f_{u,a} + \beta p_a}{\alpha + \beta}$$

$$q_{u,a} = \frac{\alpha f_{u,a} + \beta g_{u,a}}{\alpha + \beta}$$ → Pseudocount based on Substitution matrix

PSSM Example

G	H	E	G	V	G	K	V	V	K	L	G	A	G	A
G	H	E	K	K	G	Y	F	E	D	R	G	P	S	A
G	H	E	G	Y	G	G	R	S	R	G	G	G	Y	S
G	H	E	F	E	G	P	K	G	C	G	A	L	Y	I
G	H	E	L	R	G	T	T	F	M	P	A	L	E	C

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	0	0	0	0	0	0	0	0	0	0	2	1	0	2
C	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	5	0	1	0	0	0	1	0	0	0	0	1	0
F	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
G	5	0	0	2	0	5	1	0	1	0	2	3	1	1	0
H	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
K	0	0	0	1	1	0	1	1	0	1	0	0	0	0	0
L	0	0	0	1	0	0	0	0	0	0	1	0	2	0	0
M	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	1	0	0	1	0	1	1	0	0	0	0
S	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
T	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
V	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	1.3	0.7	-0.2	1.3
C	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	0.7
D	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
E	-0.2	-0.2	2.3	-0.2	0.7	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	0.7	-0.2
F	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
G	2.3	-0.2	-0.2	1.3	-0.2	2.3	0.7	-0.2	0.7	-0.2	1.3	1.7	0.7	0.7	-0.2
H	-0.2	2.3	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
I	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7
K	-0.2	-0.2	-0.2	0.7	0.7	-0.2	0.7	0.7	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	-0.2
L	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	1.3	-0.2	-0.2
M	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	-0.2
N	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
P	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	0.7	-0.2	-0.2
Q	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
R	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	0.7	-0.2	0.7	0.7	-0.2	-0.2	-0.2	-0.2
S	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	0.7	-0.2
T	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	0.7	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
V	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	0.7	0.7	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
W	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
Y	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2

PSI-BLAST (Position-Specific Iterated BLAST)

- Carefully constructed PSSM can find many distant members of a protein sequence family not easily found by a standard sequence search
- PSI-BLAST enhances the BLAST database searching method to incorporate PSSMs.
- Involves series of repeated steps or iterations

PSI-BLAST Algorithm

1. Perform standard BLAST search using a substitution matrix with a single query sequence.
2. Obtain initial set of related sequences whose BLAST score gives an *E*-value smaller than a predetermined cut-off
3. Create a PSSM from alignments of these significant matches with the query sequence
4. Scan PSSM against the database using a variant of the BLAST program to identify new sequences with suitably small *E*-values.
5. If this second search finds some newly identified related sequences, use them to update the PSSM

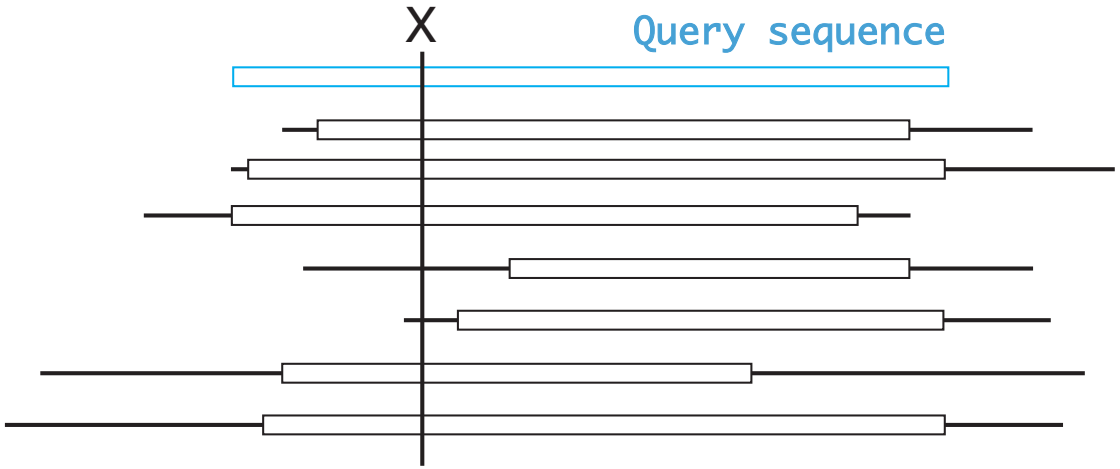
PSI-BLAST: Profile Length

- The profile constructed by PSI-BLAST has **exactly the same length** as the **query sequence**. **Insertions** with respect to the query are simply **ignored**

```
>sp|Q99728.2|BARD1_HUMAN
Length=777

 Score = 53.1 bits (126),  Expect = 3e-07, Method: Composition-based stats.
 Identities = 32/111 (29%), Positives = 55/111 (50%), Gaps = 15/111 (14%)

Query  24   THVVMKTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDFEVRGDVVNGR  83
            THVV+  DA    + TLK  LGI  G W++ + WV   ++ +    E  +E+
Sbjct  605  THVVVPGDA---VQSTLKCMLGILNGCWILKFEWVKACLRRKVCEQEEKYEIP-------  654

Query  84   NHQGPKRARESQDR---KIFRGLEICCYGPFTNMPTDQLEWMVQLCGASVV  131
              +GP+R+R ++++   K+F G     +G F + P D L  +V   G  ++
Sbjct  655  --EGPRRSRLNREQLLPKLFDGCYFYLWGTFKHHPKDNLIKLVTAGGGQIL  703
```

These aligned letters are ignored

PSI-BLAST: PSSM Construction



at position X only six residues are considered

- PSSM is restricted to those residues that have been aligned to a residue in the query sequence
- At each residue position of the query, a PSSM is constructed using only those sequences whose BLAST alignments involve that position.
- Number of sequences in the alignment changes from column to column

The Corruption of Profiles

- PSI-BLAST E-values are calculated for the profiles PSI-BLAST produces, and can not be interpreted as referring to the original query sequence

- Once a sequence unrelated to the query is included in a PSI-BLAST multiple alignment, and thus in the construction of PSI-BLAST's profile, it will bring in many of its "neighbors" on the next iteration, and this process can snowball. **Sequence weighting** will exacerbate this process.

Profile corruption is a major problem for Iterative approaches such as PSI-BLAST

PSI-BLAST vs BLAST

- Because of its **cycling nature**, PSI-BLAST allows to find **more distant homologs** than a simple BLAST search.

- PSI-BLAST uses two E-values:

  - the threshold E-value for the initial BLAST (-e option). The default is 10 as in the standard BLAST;
  - the inclusion E-value to accept sequences (-h option) in the PSSM construction (default is 0.001).

PSI-BLAST Advantages

- Fast because of the BLAST heuristic.
- Allows PSSMs searches on large databases.
- A particularly efficient algorithm for sequence weighting.
  - position-based sequence weight scheme, slightly modified to include gaps as another residue type, and to ignore fully conserved residues.
- A very sophisticated statistical treatment of the match scores.

PSI-BLAST Caution

- Avoid too close sequences ⇒ overfit!
- Can include false homologs! Therefore check the matches carefully: include or exclude sequences based on biological knowledge.
- The E-value reflects the significance of the match to the previous training set not to the original sequence!

NIH U.S. National Library of Medicine | NCBI National Center for Biotechnology Information | Sign in to NCBI

BLAST® » blastp suite

Home | Recent Results | Saved Strategies | Help

Standard Protein BLAST

blastn | **blastp** | blastx | tblastn | tblastx

BLASTP programs search protein databases using a protein query. more...

Reset page | Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

Query subrange

From

To

Or, upload file Choose File No file chosen

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

BLAST results will be displayed in a new format by default

You can always switch back to the Traditional Results page.

Choose Search Set

Database Non-redundant protein sequences (nr)

Organism Optional

Enter organism name or id--completions will be suggested exclude +

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Optional

Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Program Selection

Algorithm

- Quick BLASTP (Accelerated protein-protein BLAST)
- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)
- DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

Algorithm parameters

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

Restore default search parameters

General Parameters

Max target sequences: 500
Select the maximum number of aligned sequences to display

Short queries: Automatically adjust parameters for short input sequences

Expect threshold: 10

Word size: ♦ 2

Max matches in a query range: 0

Scoring Parameters

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Compositional adjustments: Conditional compositional score matrix adjustment

Filters and Masking

Filter: Low complexity regions

Mask: Mask for lookup table only
Mask lower case letters

PSI/PHI/DELTA BLAST

Upload PSSM Optional: Choose File No file chosen

PSI-BLAST Threshold: 0.005

Pseudocount: 0

BLAST

Search database nr using PSI-BLAST (Position-Specific Iterated BLAST)

Show results in a new window