

Blast output

Database: dbs/dm6.44.fasta
17,874 sequences; 102,739,733 total letters

Query= sp|P08505|IL6_MOUSE Interleukin-6 OS=Mus musculus OX=10090 GN=IL6
PE=1 SV=1

Length=211

Sequences producing significant alignments:

FBgn0046706 type=gene; loc=2R:4014111..4049342; ID=FBgn0046706;...

Score (Bits)	E Value
30.0	4.0

Questions after the Blast search

Questions:

- How are the hits identified?
-
- What is the meaning of the score?

Blast output

```
Matrix: BLOSUM62
```

```
Gap Penalties: Existence: 11, Extension: 1
```

```
Neighboring words threshold: 13
```

```
Window for multiple hits: 40
```

Questions after the Blast search

Questions:

- How are the hits identified?

- What is the meaning of the score?

Observe & Practice: Scoring the similarity between two sequences.

How to measure the similarity between two sequences

Q: which one is a better match to the query ?

Query: M A T W L

Seq_A: M A T P P

Seq_B: M P P W I

Judging the match using “Scoring Matrix”

Q: which one is a better match to the query ?

Query: M A T W L

Seq_A: M A T P P

Score: 5 4 5 -4 -3

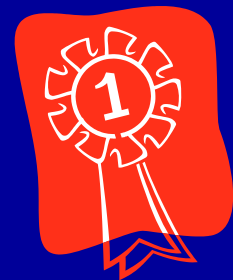
Total: 7

Query: M A T W L

Seq_B: M P P W I

Score: 5 -1 -1 1 2

Total: 16



**“Scoring Matrix” assigns a score to each pair
of amino acids**

	A	S	T	L	I	V	K	D	...
L	-1	-2	-2	4	3	1	-2	-4	

BLOSUM-62

BLOSUM - Blocks Substitution Matrices

Block: very well conserved region of a protein family. – perform the same (similar) function.

ASLDEFLL

SALEDFLL

ASLDDYLL

ASIDEFYL

ASIDEFYL

...

$$\text{Score}(a1/a2) = 2 * \log_2$$

**observed frequency of
 $a1/a2$**

**predicated frequency
of $a1/a2$**

AA: 6

AS: 4

AT: 0

BLOSUM - Blocks Substitution Matrices

Block: very well conserved region of a protein family. – perform the same (similar) function.

ASLDEFLL

ASLEDFLL

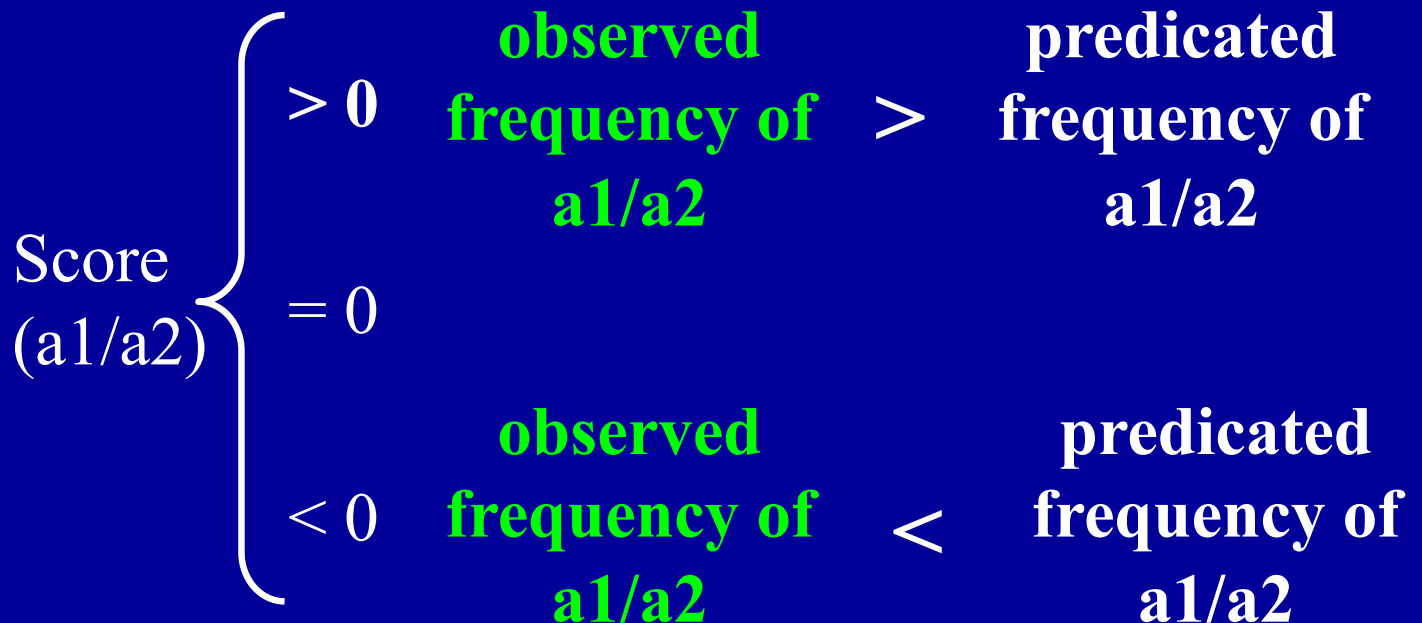
ASLDDYLL

SALEEFLL

ASLDDYLL

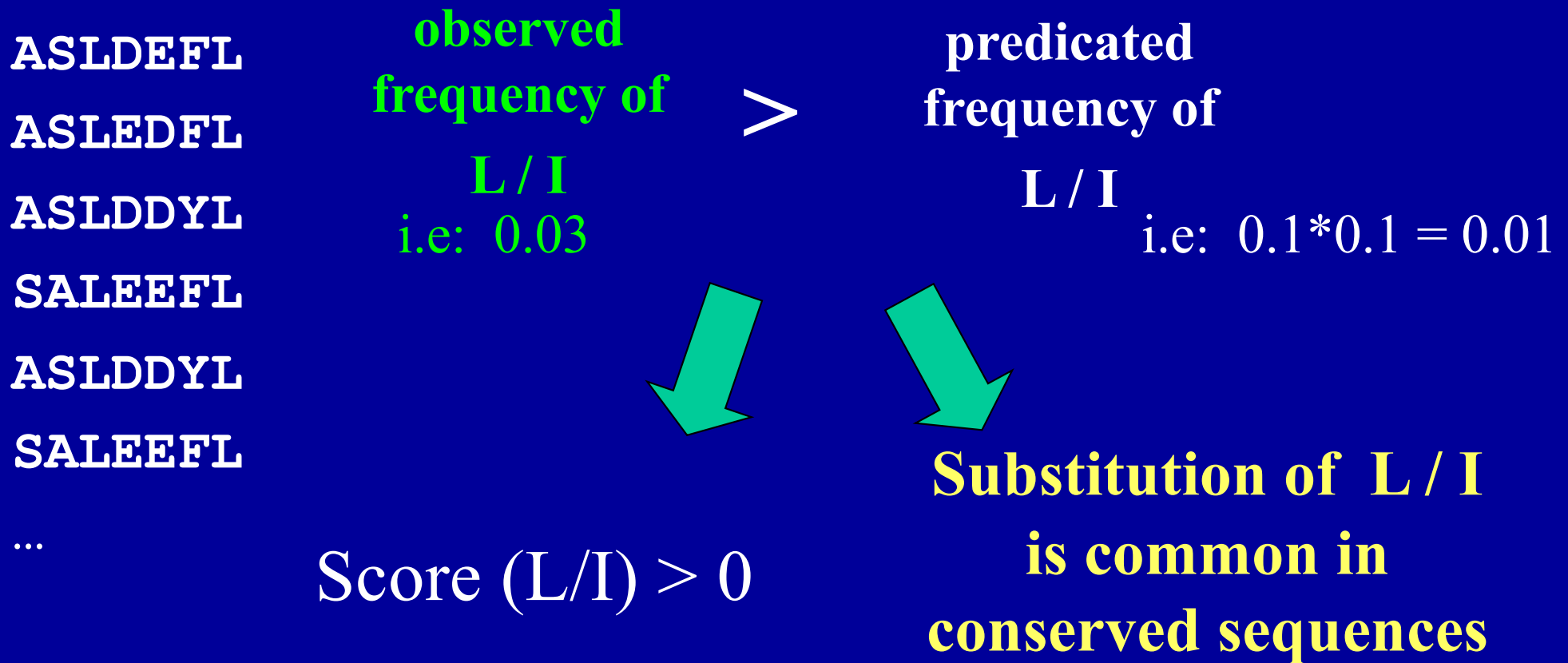
SALEEFLL

...



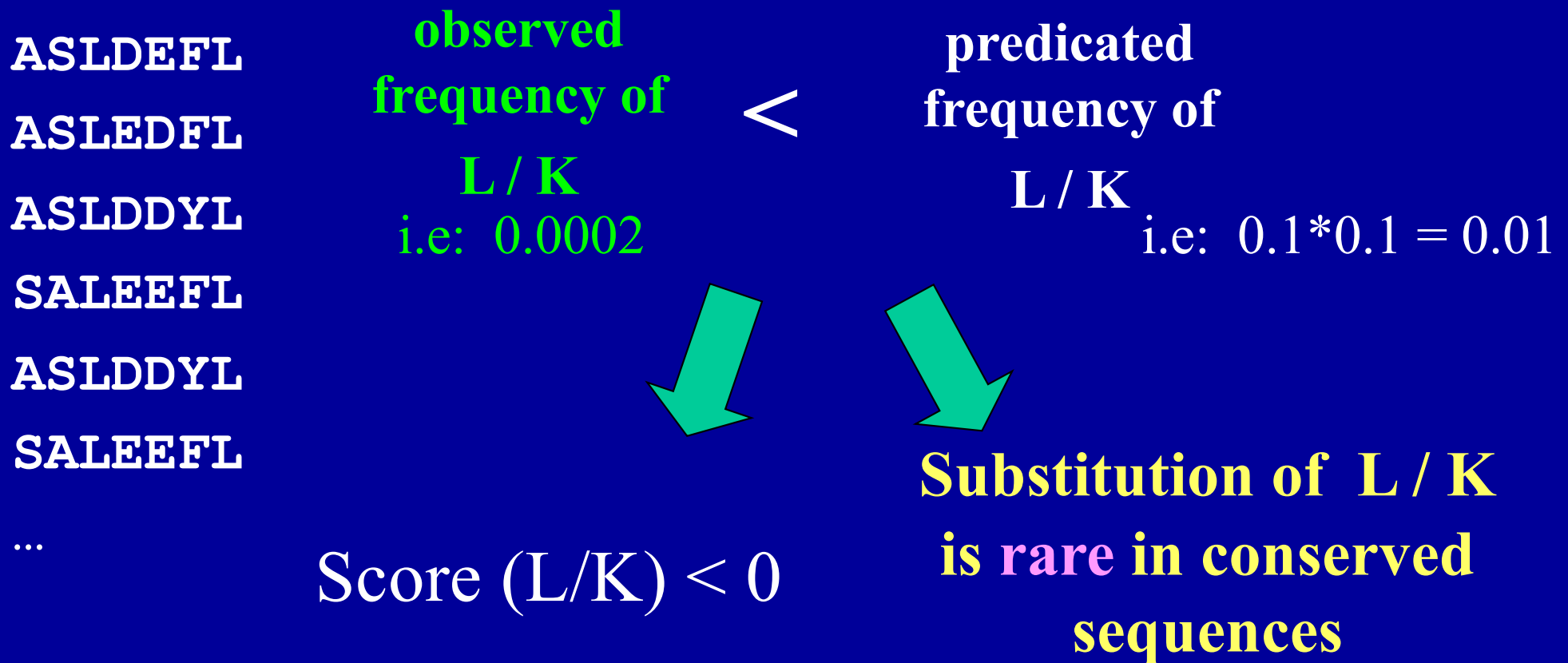
BLOSUM - Blocks Substitution Matrices

Block: very well conserved region of a protein family. – perform the same (similar) function.



BLOSUM - Blocks Substitution Matrices

Block: very well conserved region of a protein family. – perform the same (similar) function.



**“Scoring Matrix” assigns a score to each pair
of amino acids**

	A	S	T	L	I	V	K	D	...
L	-1	-2	-2	4	3	1	-2	-4	

BLOSUM-62

Scoring matrix –BLOSUM 62

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

BLOSUM - Blocks Substitution Matrices

-- Clustering threshold

BLOSUM 90 – Blocks with \geq **90%** identity are counted as one to compute the substitution score

⋮

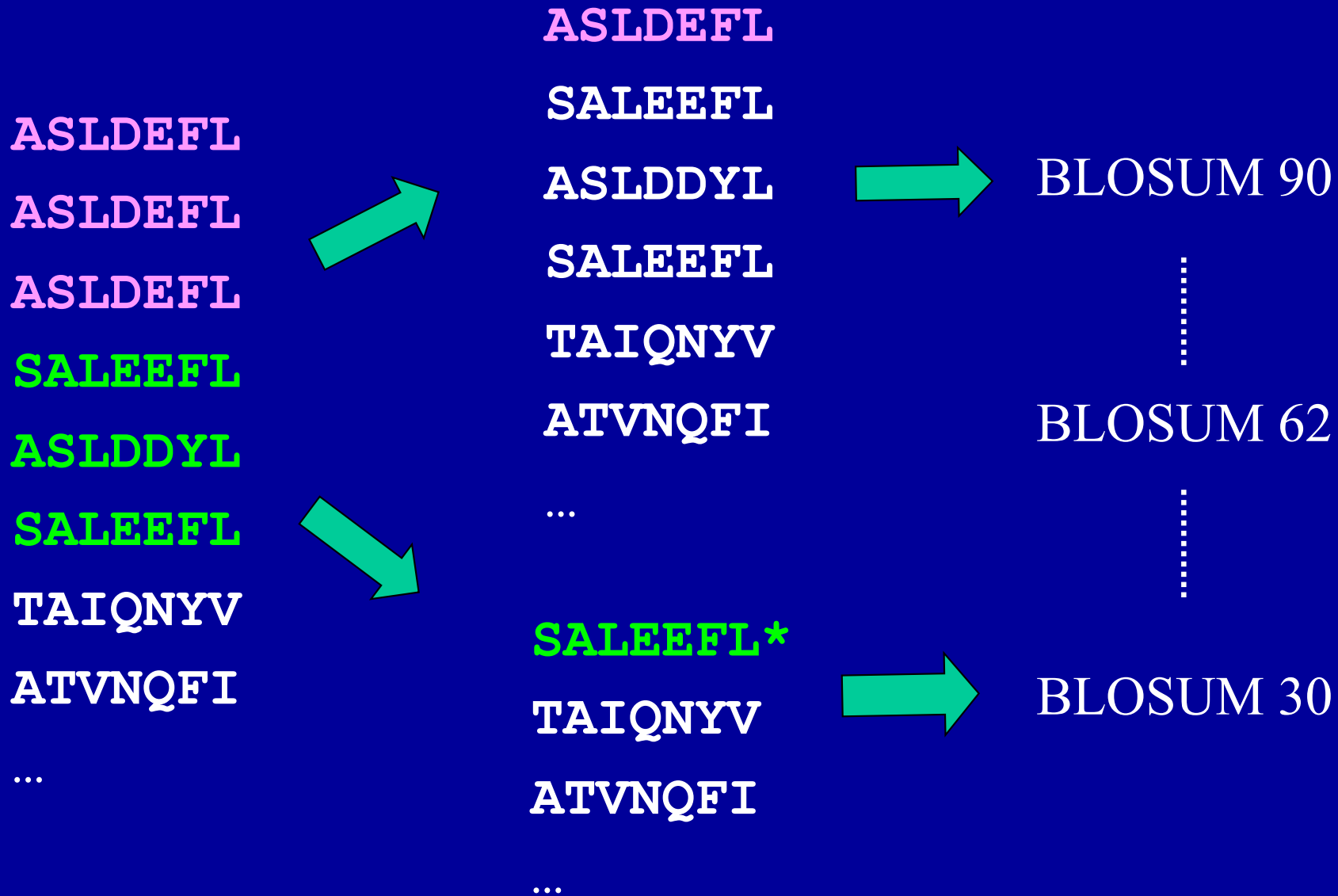
BLOSUM 62

⋮

BLOSUM 30 – Blocks with \geq **30 %** identity are counted as one to compute the substitution score

BLOSUM - Blocks Substitution Matrices

-- Clustering threshold



Comparison of Blosum matrixes

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F
L	-2	-3	-4	-5	-2	-3	-4	-5	-4	1	5	-3	2	0

Blosum 90

L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0
---	----	----	----	----	----	----	----	----	----	---	---	----	---	---

Blosum 62

L	-1	-2	-2	-1	0	-2	-1	-2	-1	2	4	-2	2	2
---	----	----	----	----	---	----	----	----	----	---	---	----	---	---

Blosum 30

Q:

Which substitution matrix will you use to identify a distant ortholog ?

a.) Blosum 40

b.) Blosum 60

c.) Blosum 90

Why BLAST uses BLOSUM62 as the Default.

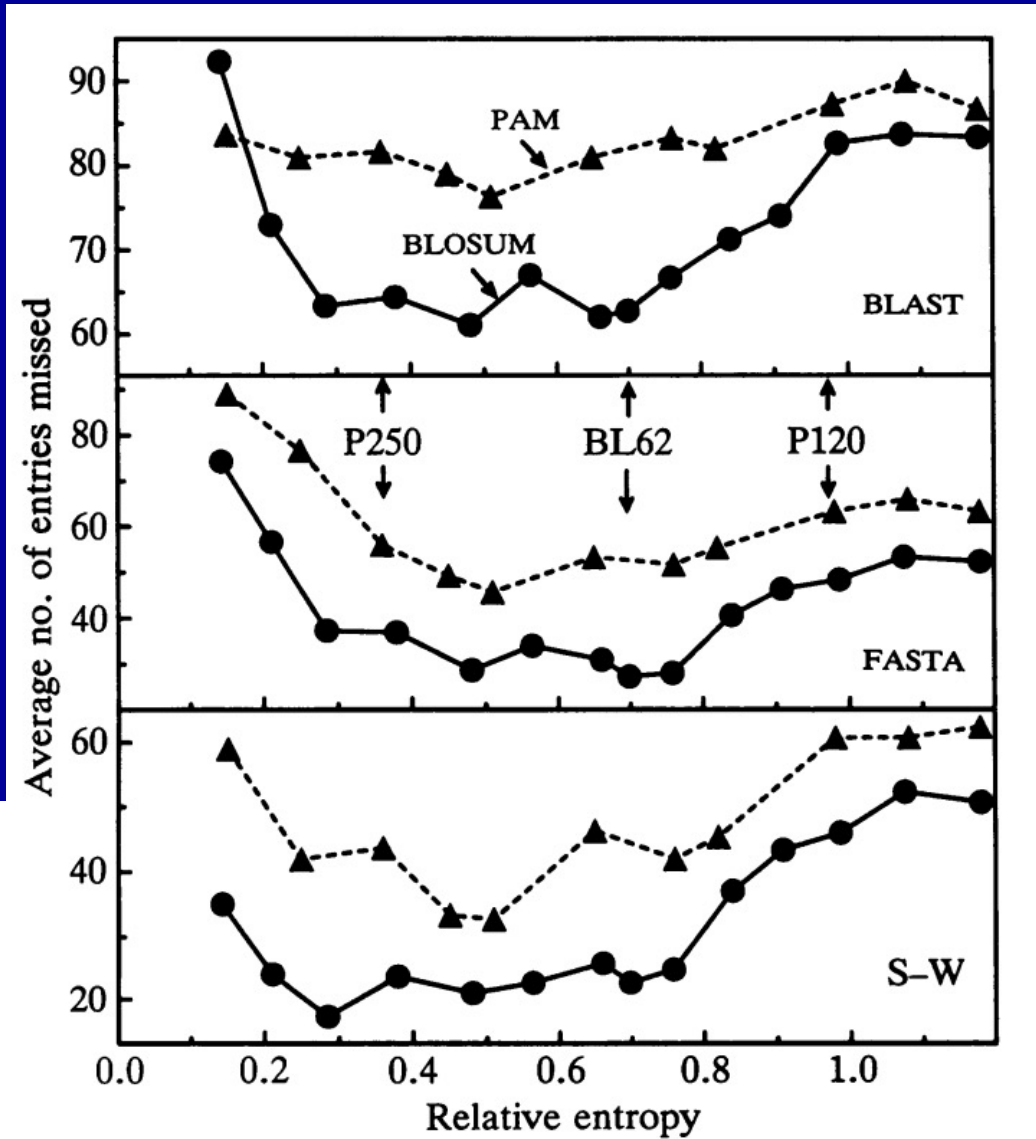
[Proc Natl Acad Sci U S A.](#) 1992 Nov 15; 89(22): 10915–10919.

doi: [10.1073/pnas.89.22.10915](https://doi.org/10.1073/pnas.89.22.10915)

Amino acid substitution matrices from protein blocks.

[S Henikoff](#) and [J G Henikoff](#)

FIG. 3. Searching performance of programs using members of the guanine nucleotide-binding protein-coupled receptor family as queries and matrices from the BLOSUM and PAM series scaled in half-bits (11). Removal of this family from the BLOCKS data base led to a nearly identical matrix with similar performance. Matrices represented (left to right) are BLOSUM (BL) 30, 35, 40, 45, 50, 55, 60, 62, 65, 70, 75, 80, 85, and 90 and PAM (P) 400, 310, 250, 220, 200, 160, 150, 140, 120, 110, and 100. The average numbers of true positive Swiss-Prot entries missed are shown for LSHR\$RAT, RTA\$RAT, and UL33\$HCMVA versus Swiss-Prot 20. Results using BLAST and FASTA or SSEARCH (S-W) are not comparable to each other, since different detection criteria were used for the three programs.



Finding the best alignment = Get the highest score

The consideration on whether to open/extend a gap is weighed by its effect on the **total score** of the alignment.

Optimization - Dynamic programming

Effect of matrices on Local Alignment

Observe: effect of matrices on the outcome of local alignment

First name initial > L -- Align seq1 and seq 2 with “blosum62”

Others -- Align seq1 and seq 2 with “blosum35”

Effect of matrices on Local Alignment

Score: **156** at (seq1) [10..36] :
(seq2) [64..90]

```
10 EPTEVFMDLWPEDHSNWQEELSPLEPSD
   |||||||||||||||||||||||||
64 EPTEVFMDLWPEDHSNWQEELSPLEPSD
```

Blosum 62:

P / H: -2

L/M: 2

Score: **206** at (seq1) [10..38] :
(seq2) [64..92]

```
10 EPTEVFMDLWPEDHSNWQEELSPLEPSDPL
   ||||||||||||||||||||||||| .
64 EPTEVFMDLWPEDHSNWQEELSPLEPSDHM
```

Blosum 35:

P / H: -1

L/M: 3

Introducing a gap

Q: M A T W L I .

A: M A W T V A .

Scr: 5 4 -2 -2 1 -1

Total: 5

Q: M A T W L I .

A: M A - W T V .

Scr: 5 4 -? 11 -1 3

Total = 22 - ?

Blosum 62:

Gap opening: -6 ~ -15

Gap Extension: -2 ~ -6

Effect of gap penalty on Local Alignment

Practice : effect of gap penalty on local alignment

Set matrix to “blosum62”

Column 1,3,5, align seq1 and seq2 with “gap=15, ext=3,”

Column 2 and 4, align seq1 and seq2 with “gap=5, ext=1”

Effect of gap penalty on Local Alignment

Blosum 62

Score: 156 at (seq1) [10..36] :
(seq2) [64..90]

```
10 EPTEVFMDLWPEDHSNWQELSPLPSD
   |||
64 EPTEVFMDLWPEDHSNWQELSPLPSD
```

Gap: -15

Ex: -3

Gap: -5 Ex: -1

Score: 161 at (seq1) [2..36] : (seq2) [53..90]

```
2  ASTV----TSCLEPTEVFMDLWPEDHSNWQELSPLPSD
   || | | |||
53 ASSVSVGATEA-EPTEVFMDLWPEDHSNWQELSPLPSD
```