

Bioinformatics II: BLOSUM Matrices

Dr Manaf A Guma
University of Anbar- college of applied sciences-Heet
Department of chemistry

1

What are BLOSUM matrices? Example?

- 1. Blocks Substitution Matrix.**
2. It is a matrix that calculates scores for each position which are obtained frequencies of substitutions in blocks of local alignments of protein sequences.
3. For example BLOSUM62 is derived from sequence alignments with no more than 62% identity.

2


BLOSUM Scoring Matrices

- BLOck SUbstitution Matrix
- It is based on comparisons of **blocks of sequences** derived from the Blocks database.
- The block length is 60 amino acids. (without any gaps or frequencies).
- Blocks database refers to the alignment not to the individual sequence.
- The Blocks database contains multiply aligned un-gapped segments belong to the most highly conserved regions of proteins (local alignment versus global alignment).

3

3

Example of Conserved blocks in alignments?



```

AABCD A . . . BBCDA
DABCD A . A . BBCBB
BBBCD ABA . BCCAA
AAACD AC . DCBCDB
CCBAD AB . DBBDCC
AAACA A . . . BBCCC

```

- Although there are mutations, but still in the same block,

4

4

What do matrices usually seek for?

- It seeks for measuring the score of the alignment.
- Why?
- Score measures the conservation.(what does that mean?)
- Which is the any substitution of amino acids in the nature.

5

What is identical sequence?

- It tells the % of matching.
- It can be 100% even if there is a substitution.
- It tells how much the sequence is conserved!
- As mentioned before, the score measures the % of alignment of the sequence to construct matrices.

6

What is the relation between?

- What is the relation between the identity and the revolution time?
- The highest the % of identity, the shorter the evolutionary time to separate the sequences.

7

How many?

- In the construction of the BLOSUM family of matrices, ~ 2000 block segments were analyzed for more than 500 groups of related proteins.
- These blocks were **clustered** according to their % of identity.

8

Give example for identity %?

Preservation of Characters Dominates in Homolog Sequences

K	K	V	K	T	P	G	A	G	V	M	K	V	P
K	V	V	T	T	T	P	A	P	V	K	K	V	M

Percentage of Identity: 7 Matches Over 14 Pairs = 50% Identity

K	L	V	K	T	P	G	A	G	V	M	K	V	P
K	V	L	T	P	T	P	K	P	V	K	K	V	M

Percentage of Identity: 4 Matches Over 14 Pairs = 29% Identity

↑ Percentage of Identity ↓ = ↓ Evolutionary Time ↓

Thus, the scores in BLOSUM45 and BLOSUM80 are generated from sequences of >45 % and >80 % similarity, respectively

BLOSUM in BLAST

Range 1: 127 to 501 [GenPept](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
431 bits(1108) 3e-147() Compositional matrix adjust. 203/375(54%) 278/375(74%) 8/375(2%)						
Query 32	EKKRRDREERQNI	VLRQPLITLQYFSLETLV	LKEWTSKLWHRQSI	VVSFLLLLAALVA	91	
Sbjct 127	KQERERLERGQLV	LWRRPLQTKYCGLELFTLL	RWTSTRLLQQRLL	LATLIVLSIVFSV	186	
Query 92	TYVVEGAHQYVQRI	EKQFLLYAYWIGLGLSSV	GLTGLHTFLLYLGP	HIASVTLAAYE	151	
Sbjct 187	IYKIDGPHQLAIEF	VRRNTWFFVYWLGLGVLSSV	GLTGLHTFLLYLGP	HIASVTLAAYE	246	
Query 152	CNSVNFPEPPYPDQI	ICPEEAGAEGAI	SLWSIISKVRIEAC	MWIGTIGELPPYF	MARA 211	
Sbjct 247	CNSLRFPPQPPYDDI	ICPEEPYDKHVPNI	WSIMSKVRLEAF	LWGAGTALGELPPY	MAKA 306	

The positives (+) in the alignment indicate good high scoring mismatches. Matrix scores >0.

Mis-matching !!!!!

What is BLOSUM K factor?

BLOSUMK

Substitution matrices provide a **likelihood for a character substitution**.

Constructed with **statistics** taken from block segments **percentages of identity** of at most K.



$$K = 1 / \text{limits of divergence}$$

<https://www.youtube.com/watch?v=xDUzRTx12ZE&t=456s>