BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024,
Tamil Nadu, India.

Programme : M.Sc., Biomedical Science

Course Title : Bioinformatics
Course Code : BM35S1BI

Unit-III

TOPIC: PAM AND BLOSUM

Dr. P. JEGANATHAN
Guest Lecturer
Department of Biomedical Science

SCORING MATRICES OR SUBSTITUTION MATRICES

(PAM AND BLOSUM)

SCORING MATRICES :

- Scoring matrices are used to determine the relative score made by matching two characters in a sequence alignment.
- For both nucleic acids and proteins ,the alignment score is calculated using scoring matrix .
- A scoring matrix is a set of values representing the likelihood of one residue being substituted by another during sequence divergence through evolution . This is why the scoring matrix is also known as the substitution matrices .

CONT……

- A Scoring matrix for comparing DNA sequence can be simple because there are only four nucleotides and the mutation frequencies are assumed to be equal .
- For generating a DNA sequence alingment score ,the simple scoring matrix is still used, such as the NUC4.2 and NUC4.4 These matrices can be obtained from the NCBI .

CONT……..

- Scoring matrices for amino acid substitution are more complex ,reflecting the similarity of physico chemical properties ,as well as the likelihood of one amino acid being substituted by another at a particular position in homologous proteins .
- Two well known types of scoring matrices for proteins are PAM and BLOSUM .

PAM

- Pam (Point Accepted Mutation –that is ,accepted point mutation) matrices were first developed by Margaret dayhoff and collaegues in 1978 and hence are also known as Dayhoff PAM matrices .
- A PAM represents a substitution of one aminoacid by another that has been fixed by natural selection because either it does not alter the protein function or it is beneficial to the organisms .

CONT…

- Calculated from very similar sequences by measuring their differences.
- Sequence are defined as "one PAM unit diverged "if the series of accepted mutations converted S1 and S2 with an average of one point mutation per 100 aminoacids .

CONT…

- The number with the matrix (PAM120,PAM90),refers to the evolutionary distance . Greater numbers are greater distances .
- To derive PAM250 you multiply PAM1 250 times .
- PAM250 is the matrix derived of sequences with 250 PAMs.

BLOSUM :

- BLOSUM (BLOcks SUbstitution Matrices)were proposed by Steven Henikoff and Jorja Henikoff in 1992 .BLOSUM represents an alternative set of scoring matrices, which are widely used in sequence alignment algorithms .
- BLOSUM matrices were developed based on multiple alignment of 500groups of related protein sequences, which yielded > 2000 blocks of conserved amino acid patterns .
- In each alignment the sequences similar at some threshold value of percent identity were clustered into groups and averaged .

CONT…

- Therefore ,the blocks of ungapped multiple sequence alignments ,which are the cornerstone of BLOSUM matrices, reveal the evolutionary relationship between proteins .
- Thus ,BLOSUM62 means that the sequences clustered in this block are at least 62% identical .
- This allows detection of more distantly related sequences ,as it downplays the role of the more related sequences in the block when building the matrix .

DIFFERENCE BETWEEN :

PAM	BLOSUM
Based on mutational model of evolution	Based on the multiple alignment of blocks .
PAM1is based on sequence of 85% similarity .	Good to be used to compare distant sequences
Designed to track the evolutionary origins	Designed to find protein’s conserved domains .

REFERENCES:

Papadopoulos JS ,Agarwala R. Bioinformatics.

Larkin MA ,et al. Bioinformatics.

THANK YOU !!!