# Zoology Legends

# Bif401

## Mid_Term Papers Spring 2023

Effort BY Taaha Khan

**Join our Legend Group**

https://chat.whatsapp.com/EVU415zOZ4S
IgCpc6fBAQX

**Team Zoology LEGENDS**

## Subjective part

## Question

### 1. PAM and Blosum difference.

**PAM** means "Point Accepted Mutation"  Point accepted mutations means the substitution of one amino acid in a sequence with another that protein function remain conserved.

**BLOSUM** matrices can be used to align the protein sequences. BLOSUM matrices was first purposed in 1992 by Henikoff et al.  BLOSUM matrices is also called the Block substitution matrix without any gap although it has mismatches in sequences.

### 2. Uniprot and swissprot.

Both UniProt and SwissProt are the online database for proteins.

**UniProt** is public database which is being used to search the sequence of proteins.

**SwissProt** is the manually annotated version of the UniProt Database.

## 3. Rooted and unrooted Phylogenetic tree

Phylogenetic specifies evolutionary relationship with the help of trees. Trees can be rooted or unrooted.

**Rooted trees** can show temporal evolutionary direction. Rooted trees are computationally expensive. Rooted trees shows the most basal ancestor of the tree.

**Unrooted** phylogenetic tree does not show an ancestral root. Unrooted trees represents the branching order but do not indicate the root or location of the last common ancestor. Unrooted trees shows the relatedness of organisms without indicating ancestry.

# Applications of bioinformatics in the field of GENOMICS and PROTEOMICS

## GENOMICS:

- Bioinformatics can help in assembling DNA sequencing data.
- It can help in gene finding (markers).
- Gene assembly can be performed using bioinformatics tools (nucleotide alignments)
- It can help transcribe the gene data to RNA data
- Also, databases can be generated from such data.

## PROTEOMICS :

- Bioinformatics can help us in decoding protein sequences.
- It can also help us in understanding protein structure.

- We can also understand post translational changes in proteins with the help of bioinformatics.
- We can better understand the protein-protein interaction in different biological reactions.
- It can also help us in generating databases of these sequences and structures.

## Translation and transcription

**Translation:** The process by which a cell makes proteins using the genetic information carried in messenger RNA (mRNA). The mRNA is made by copying DNA, and the information it carries tells the cell how to link amino acids together to form proteins.

**Transcription** is the process by which the information in a strand of DNA is copied into a new molecule of messenger RNA (mRNA). DNA safely and stably stores genetic material in the nuclei of cells as a reference, or template.

## How bulges are formed?

Bulges, are formed when a double-stranded region cannot form base pairs perfectly. Bulges can be asymmetric with varying number of base pairs on one side of the loop. Bulge loops are commonly found in helical segments of cellular RNAs and used to measure the helical twist of RNA in solution.

## Why RNA is less stable than DNA?

Unlike DNA, RNA in biological cells is predominantly a single-stranded molecule. While DNA contains deoxyribose, RNA contains ribose, characterized by the presence of the 2'-hydroxyl group on the pentose ring. This hydroxyl group make RNA less stable than DNA because it is more susceptible to hydrolysis.

## Uniprot and its features

Uniprot is a freely accessible database of protein sequence and functional information, many entries being derived from genome sequencing projects. It contains a large amount of information about the biological function of proteins derived from the research literature.

## Types and Function of RNA?

There are three major types of RNA which are as follows

- **mRNA**, or messenger RNA, that serve as temporary copies of the information found in DNA;
- **rRNA**, or ribosomal RNA, that serve as structural components of proteinmaking structures known as ribosome
- **tRNA**, or transfer RNA, that ferry amino acids to the ribosome to be assembled.

- **Micro RNAs (miRNA)**
- **Small Interfering RNA (siRNA**

## How Gibbs free energy effects the RNA structure:

- Gibbs free energy” is the free energy of an RNA molecule available for reaction
- The smaller, the better!
- RNA structure formation lowers the free energy

Energy of RNA structures

- RNA structures have the lowest (or close) quantity of free energy.
- In cases where RNA can take two structural forms, one can select the one with lower energy state.

## Substitution and indels

Indels lead to Gaps in alignment.

Substitutions lead to mismatches

Mutations are treated with substitution penalties Substitution penalties vary depending on the substitution. Indels were addition or removal of amino acids from protein sequences OR nucleotides from DNA or RNA sequences.

## Frontiers in Personalized Medicine

• Not all medicines work all the time! • Some medicines have side effects on certain patients • How to evaluate drugs using Bioinformatics tools?

## BLOSUM long 5 marks

BLOSUM matrices were first proposed in 1992 by Henikoffet al.

- Protein sequences can be aligned using BLOSUM scoring matricesBLOSUM or Block Substitution Matrix

- A block is an a multiple sequence alignment without any gaps

- Let’s take a look at a block of protein sub-sequences.

**Computing BLOSUM-x matrices**

Step 1: Eliminate sequences that are identical in x% positions

Step 2: Compute observed frequency f i,j of aligned pair Ai to Aj. Hence, f i,j becomes the probability of aligning Ai and Aj in the selected blocks.

Step 3: Compute fi which is the frequency of observing Ai in the entire block.

## Identity and similarity

**Identity** is the count of exact matches between two sequences Gaps are excluded

Identity = Num. of matches/Smaller Legth * 100% = 100%

**Similarity** is the comparison between sequences calculated by using alignment approach.

Can be calculated by pairwise sequence alignment.

## Uses of Phylogenetic trees

- Phylogenetics is the science of studying evolutionary relationships.
- Phylogenetics has led to the creation of relationship trees between various species of Bacteria, Archaea, and eukariya.
- Phylogenetic trees can be used to show phylogenetic relationships between sequences.

## Difference between Smith and needleman Algorithm

The Smith Waterman Algorithm allows us to elicit local alignments. Local alignments can be performed using Smith Waterman algorithm

Traceback can start from any position in the scoring matrix. Local alignments can be extracted by starting from a high score till reaching ‘0’.

Smith Waterman helped find best local alignment.

Needleman-Wunsch calculates global alignment.

Needleman Wunsh algorithm traceback begins from the bottom right element. Proceeds progressively until we reach the top left. This provides us with a global alignment.

## Blast and its types

BLAST stands for basic local alignment search tool. First developed in 1990.

BLAST can search sequence databases and identify unknown sequences by comparing them to the known sequences BLAST performs searches by quick alignments on sequences.

**Types:**

Nucleotides

**Blastn**: Compares a nucleotide query sequence against a nucleotide database. Proteins

**Blastp**: Compares an amino acid query sequence against a protein database.

**Blastx**: Compares a nucleotide query sequence against a protein sequence database. Helps find potential translation products of unknown nucleotide sequences.

Reverse of **Blastx tblastn:** Compares a protein query sequence against a nucleotide sequence database. Nucleotide sequence dynamically translated into all reading frames.

**tblastx**: Compares the six-frame translated proteins of a nucleotide query sequence against the six-frame translated proteins of a nucleotide sequence database.

## How RNA becomes more stable?

RNA comprises of 4 nucleotide bases (A, U, G & C).

These bases are attached to Ribose sugars in the RNA backbone. Hydrogen bonds can form between complementary nucleotide bases Complementary nucleotides i.e. G&C or A&U form hydrogen bond. Upon formation of the bond, energy is released.  Thereby, RNA molecule becomes more stable.

5 nucleotides formed H-Bonds.  This bond formation released energy (-12.0 kcal/mol).  RNA molecule took up a 2’ structure. Hence became more stable.

## Note on bulges

Bulges, are formed when a double-stranded region cannot form base pairs perfectly.

- Bulges can be asymmetric with varying number of base pairs on one side of the loop.

## How are Junctions Formed?

Junctions include two or more double-stranded regions converging to form a closed structure. The unpaired bases appear as a bulge.

## Why we use 4th sequence in NJ algorithm?

NJ Algorithm is a DP approach to predict RNA 2’ structures A scoring matrix is initialized to record scores in NJ Algo.

For filling scoring matrix, the maximum score from 4 matrix positions is chosen.

## NJ algorithm long 5 marks

Nussinov-Jacobson (NJ) Algorithm is a Dynamic Programming (DP) strategy to predict optimal RNA 2’ structures. Proposed in 1980.

- Computes 2’ structures with most nucleotide coupling.
- Create a matrix with RNA sequences on top and right.
- Set diagonal & lower tri-diagonal to zero.
- Start filling each empty position in matrix by choosing the maximum of 4 scores

## Difference between FASTA35 and FASTy35

**Fasta35** Scan a protein or DNA sequence library for similar sequences

**fasty35** Compare a DNA sequence (6ORFs) to a protein sequence database.

## Expasy and the fields that are explores by expasy?

Expasy provides access to a variety of online databases and tools

Developed by Swiss Bioinformatics Institute (SIB) Website provides access to databases and tools The fieldsthat explores by expasy are Proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc.

## Domain and domain shuffling

**Domains** are semi-independent functional structures in a protein  Have a stable structure Over ~40 residues

**Domain Shuffling:**  Aligned portions of sequence can be considered in varying orders and the process is called domain shuffling.

## Scoring matrices and derive it's formula

Scoring matrices accurately reward and penalize matches and mismatches.

**Formula:**

Name the three ways that modified DNA?

DNA gets modified by: • Mutation & Substitution • Insertion • Deletion

## Application of Bioinformatics

Bioinformatics has wide range of applications in a variety of feilds which are as follows

- Genomics
- Transcriptomics
- Proteomics
- Metabolomics Structural
- Proteomics
- Drug Design
- Systems Biology
- Personalized Medicine

## What is DNA

DNA is a hereditary material. It transfers the genetic information from one generation to next generation. It is double stranded. It is made up of three components

- Sugar,
- nucleotides and
- phosphate group .

## NCBI Genebank

Collaboratively maintained by:

• DNA DataBank of Japan (DDBJ)

• European Molecular Biology Laboratory (EMBL)

• GenBank at NCBI

Entries can be a contiguous stretch of DNA or RNA sequences and their annotations

• Entries are updated every two months

## Significant features of uniprot?

**Comprehensive protein sequence database:**

UniProt provides a vast collection of protein sequences from various organisms, including reviewed (UniProtKB/Swiss-Prot) and unreviewed (UniProtKB/TrEMBL) entries.

**Functional annotations:** UniProt offers functional information about proteins through detailed annotations obtained from literature, expert curation, and automated systems.

**Proteome completeness and quality**:

UniProt employs new methods to assess proteome completeness and quality, ensuring the inclusion of accurate and reliable protein sequences.

**Integration of data and resources**: UniProt integrates, interprets, and standardizes data from multiple resources, acting as a central hub for protein-related information and linking out to other relevant databases and tools.

## Role of Genetic Code in Nucleotides:

The genetic code is a set of rules that determines the correspondence between nucleotide triplets (codons) in DNA or RNA and the amino acids they encode during protein synthesis.

The genetic code plays a crucial role in translating the sequence of nucleotides into the sequence of amino acids, thus determining the primary structure of proteins.

It ensures the accurate and consistent translation of genetic information from DNA to protein.

## Bulges Matrices and Types:

In the context of your query, the term "bulges matrices" is not clear. It may refer to bulge loops, which are structural features in RNA molecules. Bulge loops are non-canonical structural elements formed when an unpaired nucleotide disrupts a regular helical region in an RNA sequence. The types of bulge loops include single-nucleotide bulges, internal loops, and multiple-nucleotide bulges.

## Use of Databases and Online Bioinformatics Tools:

Databases and online bioinformatics tools play crucial roles in biological research. They facilitate the storage, organization, and retrieval of biological data, allowing researchers to access and analyze large volumes of information efficiently.

Databases provide repositories of genomic, proteomic, and other biological data, enabling researchers to explore and compare various datasets.

Bioinformatics tools aid in data analysis, sequence alignment, protein structure prediction, and other computational analyses, assisting researchers in gaining insights and making discoveries.

## Structures Formed in RNA Sequences:

RNA sequences can form various secondary and tertiary structures. Some common structures include:

**Hairpins:** Formed when a region of RNA folds back on itself, creating a double-stranded stem and an unpaired loop.

**Bulge loops**: As mentioned earlier, bulge loops are formed by unpaired nucleotides disrupting a regular helical region.

**Internal loops:** These occur when unpaired nucleotides are present in the interior of a helical region, creating additional loops.

**Pseudoknots:** Complex RNA structures in which two or more helical regions intertwine, often with base pairing interactions between non-adjacent segments.

## Difference between Essential and Non-essential Amino Acids:

Essential amino acids are those that cannot be synthesized by the body and must be obtained from the diet.

There are nine essential amino acids: histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine.

Non-essential amino acids, on the other hand, can be synthesized by the body and are not required in the diet for protein synthesis. Non-essential amino acids include alanine, asparagine, aspart.

**Muhammad Zaman**

**03044756496**

**Remember me in you Precious Prayer**

**Jazak Allah**