BIF401 – BIOINFORMATICS-I

FINAL TERM MOST IMPORTANT QUESTIONS

PREPARED BY: SHAHZAIB SHABBIR (TEAM MEMBER, BS ZOOLOGY ANNOUNCEMENT GROUP)

(SHAHZEB MUGHAL, ADMIN BS ZOOLOGY ANNOUNCEMENT,)

Group link: https://chat.whatsapp.com/BjC6W0jZx8o9yaC2ISh1Gu

Q1: What is bioinformatics and how does it integrate various scientific disciplines to analyze biological information through computer algorithms and software?

Bioinformatics is an interdisciplinary science at the cross-roads of biology, mathematics, computer science, chemistry and physics. With the digitalization of the biological information, doors have been wide opened towards the analysis of this information using computer algorithms and software.

Q2.What challenges arise in the identification of genes and proteins.

Human genome has over 25,000 genes and these genes code for thousands of different proteins which perform day-to-day functions in the living cell.Furthermore, these proteins may take on various post-translational modifications leading to a very large number of functionally unique molecules. This presents us with a huge challenge in identification of genes and proteins.

Q4.Write scope of bioinformatics.

SCOPE OF BIOINFORMATICS:

Bioinformatics primarily deals with digitalized biological information as well as data reported from biology experiments. Computational methods, data processing techniques and algorithms are employed in addressing the following issues:

1. Storage of data.
2. Organization data.
3. Analysis of many experiments.
4. For representation of biological information

APPLICATIONS OF BIOINFORMATICS:

GENOMICS:

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

EVOLUTIONARY STUDIES:

1. Evolutionary relationships between different organisms can be derived from data.
2. Evolutionary distance among species can be computed by using bioinformatics tools.
3. Phylogenetic trees can be constructed to find relationships between species.
4. Ancestry can be better understood between several species and organisms.

PROTEOMICS:

1. Bioinformatics can help us in decoding protein sequences.
2. It can also help us in understanding protein structure.
3. We can also understand post translational changes in proteins with the help of bioinformatics.
4. We can better understand the protein-protein interaction in different biological reactions.
5. It can also help us in generating databases of these sequences and structures.

SYSTEMS BIOLOGY:

Bioinformatics can assist us in modelling regulatory mechanisms in gene and protein networks. Such models can be analyzed to identify the key regulators in these networks. Moreover, the models can help evaluate drugs to treat these key regulators.

Q4.Describe frontiers of bioinformatics.

Frontier in Bioinformatics includes:

FRONTIER IN PROTEIN STURUCTURE:

Bioinformatics helps us to understand the layer folding of proteins that how they are processed, and helps to know that how protein interact with each other and how a drug can affect or stimulate a protein.

FRONTIER IN SYSTEM BIOLOGY:

It helps us to understand the whole system of a single cell, in that cell how organelles, gene, proteins and metabolites are interconnected in a single unified system (cell). And bioinformatics also give us the idea how these models can be applied to real-time.

FRONTIER IN PERSONALIZED MEDICINE:

This is the important thing for this century and upcoming generation that personalize the medicine for exact cure of a disease. Because all the medicine cannot work exact some effect patient badly therefor with the help of Bioinformatics we are now able to personalize some medicines for some diseases. And bioinformatics helps us to evaluate the medicine.

Prepared by:
Shahzaib Shabbir
03308553704

Q5.Explain transcription and translation.

Transcription: This is the process where the DNA sequence of a gene is copied into messenger RNA (mRNA). It occurs in the nucleus, where RNA polymerase binds to the DNA and synthesizes a complementary RNA strand based on the DNA template.

Translation: This is the process where the mRNA is decoded to synthesize a specific protein. It occurs in the ribosome, where transfer RNA (tRNA) molecules bring amino acids that correspond to the codons on the mRNA, linking them together to form a polypeptide chain.

Q6.What is the composition of DNA and RNA, and how are the nucleotides classified into purines and pyrimidines?

If we talk about the composition of DNA and RNA molecule than these are composed of four other molecules which are named as Nucleotides. These molecules are Adenine (A), Cytosine (C), Thymine (T), Uracil (U), and Guanine (G) Adenine and Guanine collectively called Purines while Cytosine, Uracil, and Thymine are called as Pyrimidine.

Q7.Decribe GenBank,Uniprot,Swissprot.

GenBank is online database where researcher can get access to the sequences of DNA, RNA and proteins.

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

Swiss Prot which contains human curated protein information, molecular mass, observed and predicted modifications etc.

Q8.What is the purpose of Pairwise Sequence Alignment in biological sequence. Also write it types.

Pairwise Sequence Alignment is used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences (protein or nucleic acid). There are two types of pair alignments:

1. Local

In Local alignment we compare one whole sequence with the one portion of other.

2. Global

In Global alignment we compare both sequence from end to end completely.

Q9.Decribe Dot-plots.

DOT PLOTS:

To visualize the sequence alignment we have a method called Dot Plots in this method the sequence is written top and left side of dot matrix grid. In dot plot the matching nucleotides are connected in diagonal way and represent the sequence alignments.

Q10.Why is RNA more stable than DNA.

When phosphate, nitrogen base and sugar come together if there is (OH) than molecule is RNA and if there is (H) in sugar than molecule is DNA.

Prepared by:
Shahzaib Shabbir
03308553704

Q11.What is domain shuffling.

DOMAIN SHUFFLING:

Aligned portions of sequence can be considered in varying orders and this process is called as domain shuffling

Q12.What is indels.

Removal and addition of amino acids in proteins and nucleotides in DNA, RNA by using Gaps named as Indels.

Q13.Describe dynamic programming with examples.

Dynamic programming helps us to reduce the computational cost in sequence comparisons and it works on the method of "scoring function".

For example:

Match = +a
Mismatch = -b
Gap = -c
Score = Match + Mismatch +Gaps

Q14.Briefly describe Needleman Wunch Algorithm.

Needleman Wunsch Algorithm is the way for alignment. The method is same like dot plot but it computes the scores in different way. We Start with a zero in the second row, second column. Move through the cells row by row calculating the score for each cell.

Q15. What is Fasta.

FASTA can search sequence databases and identify unknown sequences by comparing them to the known sequence databases. This can help obtain information on the parent organism, function and evolutionary history.

Q16. How many types of Fasta.

There are six types of FASTS:

fasts35: Compare unordered peptides to a protein sequence database.
fastm35: Compare ordered peptides (or short DNA sequences) to a protein (DNA) sequence database.
Fasta35: Scan a protein or DNA sequence library for similar sequences.
Fastx35: Compare a translated DNA sequence (6 ORFs) to a protein sequence database.
tfastx35: Compare a protein sequence to a DNA sequence database (6 ORFs).
fasty35: Compare a DNA sequence (6ORFs) to a protein sequence database.

Q17. How many types of Phylogenetic Trees.

Types of Phylogenetic Trees:

Scaled Trees: Branch lengths are equal to the magnitude of change in the nodes.
Unscaled Trees: Only representing the relationship between sequences.

Prepared by:
Shahzaib Shabbir
03308553704

Q18. How DNA gets modified.

DNA gets modified by:

- Mutation & Substitution
- Insertion
- Deletion

Q19. Why we use Zuker's Algorithm.

Energy based methods involve evaluating the free energy structures. To compute the RNA sequence for 1' or 2' optimal structure prediction we use Zuker's Algorithm.

Q20. How many types of alignments.

There are two types of alignments:

- Optimal Alignments
- Best Alignment

Q21. Why we use Blast? Also write it types.

BLAST can search sequence databases and identify unknown sequences by comparing them to the known sequences. This can help identify the parent organism, function and evolutionary history.

There are two main types of BLAST.

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

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

There are also many other types of BLAST:

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

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 sixframe translated proteins of a nucleotide sequence database.

Q22.Write types of protein structure and how do they differ in terms of composition and function?.

Types of Protein Structures

1.***Primary Structure***:

•The primary structure refers to the unique sequence of amino acids in a polypeptide chain linked by peptide bonds. This sequence determines the protein's identity and function, with a diversity represented by 20^n combinations where n is the number of amino acids.

2.***Secondary Structure:***

•The secondary structure involves local folding patterns within the polypeptide chain primarily stabilized by hydrogen bonds. Common forms include α-helices and β-pleated sheets which contribute to the overall stability and shape of the protein.

3.***Tertiary Structure:***

•The tertiary structure is the overall three-dimensional shape of a single polypeptide chain formed by the bending and folding of the chain due to interactions among side chains (R groups). This structure is

crucial for the protein's functionality and is stabilized by various interactions, including hydrogen bonds, ionic bonds, and hydrophobic interactions.

4.***Quaternary Structure:***

•The quaternary structure refers to the arrangement of multiple folded protein molecules (subunits) in a multi-subunit complex. This structure is important for the function of many proteins as the interactions between subunits can affect the protein's activity and regulation.

Q23.What are the two categories of RNA?

There are two categories of RNA:

1. Coding RNA
2. Non-Coding RNA

Coding RNA perform their coded function in protein synthesis. And Non-coding RNA helps in translation process.

Q24.What are the different types of RNA, and what are their roles in protein synthesis and gene regulation?

Types of RNA:

- ***Messenger RNA (mRNA):*** Carries genetic information from DNA to the ribosome, where it serves as a template for protein synthesis.
- ***Transfer RNA (tRNA):*** Transports specific amino acids to the ribosome during translation, matching them to the corresponding codons on the mRNA.
- ***Ribosomal RNA (rRNA):*** Combines with proteins to form ribosomes, the cellular machinery that facilitates the translation of mRNA into proteins.
- ***Micro RNAs (miRNA):*** Small, non-coding RNA molecules that regulate gene expression by binding to complementary mRNA sequences, leading to their degradation or inhibition of translation.
- ***Small Interfering RNA (siRNA):*** Involved in the RNA interference (RNAi) pathway, siRNAs can silence gene expression by promoting the degradation of target mRNA.

Q25. Why we use PAM

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.

Q26. What is PAM Unit.

PAM unit is actually that time during which 1% amino acid undergo for acceptable mutation. If two sequences diverge by 100 PAM units, it does not mean that they will be at totally different positions.

Q28. How many steps to compute the BLOSUM matrics.

There are three steps to compute the BLOUSM 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

Prepared by:
Shahzaib Shabbir
03308553704

Q29.What are the different types of RNA structures.

1. ***Single-Stranded RNA (ssRNA)***: RNA that consists of a single strand, which can fold into various secondary structures.
2. Double-stranded RNA (dsRNA) is a helical structure formed by stacked base pairs from two complementary RNA strands. .
3. ***Stem and Loop (Hairpin Loop)***: A common secondary structure where a segment of RNA folds back on itself, forming a double-stranded "stem" and a single-stranded "loop" at the end.
4. ***Bulge Loop***: A structure where one or more unpaired nucleotides are present in a double-stranded region, creating a bulge.
5. ***Interior Loop***: Similar to a bulge loop, but involves unpaired nucleotides on both sides of a double-stranded region, forming a loop within the helix.
6. ***Junction or Intersection***: Refers to points where multiple RNA strands or structures converge, often seen in complex RNA molecules.

Q30.What is EXPASY.

EXPASY It is developed by Swiss Bioinformatics Institute (SIB). Website provides access to databases and tools Proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc can be searched.

Q31.What are BLOSUM matrices.

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.

Q32.What is the significance of Gibbs Free Energy in RNA folding? how does it relate to the stability of RNA structures.

RNA molecules form many structures for stability and different functions. “Gibs Free Energy” (LANGRIDGE and KOLLMAN 1987) is the free energy available for RNA molecule for reactions and RNA structure formation takes place at this lower energy. Incase if RNA has two structure we can select the one with lowest energy state.We can calculate the overall energy of RNA structures by summing up energies given out duringthe process of folding. For knowing the positive and negative values of calculations of stabilizing and destabilizing energies we may factor in ways in which RNA can be destabilized.

Q33.Write types of protein structure and how do they differ in terms of composition and function?.

Types of Protein Structures:

1.Primary Structure: The primary structure refers to the unique sequence of amino acids in a polypeptide chain linked by peptide bonds. This sequence determines the protein's identity and function, with a diversity represented by 20^n combinations where n is the number of amino acids.

2.Secondary Structure: The secondary structure involves local folding patterns within the polypeptide chain primarily stabilized by hydrogen bonds. Common forms include α-helices and β-pleated sheets which contribute to the overall stability and shape of the protein.

Prepared by:
Shahzaib Shabbir
03308553704

3.Tertiary Structure: The tertiary structure is the overall three-dimensional shape of a single polypeptide chain formed by the bending and folding of the chain due to interactions among side chains (R groups). This structure is crucial for the protein's functionality and is stabilized by various interactions, including hydrogen bonds, ionic bonds, and hydrophobic interactions.

4.Quaternary Structure: The quaternary structure refers to the arrangement of multiple folded protein molecules (subunits) in a multi-subunit complex. This structure is important for the function of many proteins as the interactions between subunits can affect the protein's activity and regulation.

Q34.What is CustalW also write its features.

ClustalW is used to align multiple nucleotide or protein sequences. It helps identify regions of similarity that may indicate functional, structural, or evolutionary relationships among the sequences.

Features of CustalW:

1. Create multiple alignments
2. Optimize existing alignments
3. Profile analysis
4. Create phylogenetic trees

Q35.Write steps of homology modeling?

1. Template recognition and initial alignment.
2. Alignment correction.
3. Backbone generation.
4. Loop modeling.
5. Side-chain modeling.
6. Model optimization.
7. Model validation

Q36.Define MGF?

MGF(Mascot Generic Format) is a simple human-readable format for MS/MS data developed by matrix science.

Q37.Write the forces involved in protein folding?

The forces involved in protein folding include:

- Electrostatic interactions
- van der Waals interactions
- Hydrogen bonds
- Hydrophobic interactions

Q38.Why The Tertiary Structure Of RNA Is Less Stable Over The Secondary Structure Of RNA?

The tertiary structure of RNA is less stable than its secondary structure due to its complexity which involves long-range interactions that can lead to misfolding or disruption. Additionally tertiary structures are more sensitive to environmental factors like pH and temperature making them less stable. Secondary structures such as hairpins and loops are more energetically favorable and require less energy to maintain resulting in greater stability.

Prepared by:
Shahzaib Shabbir
03308553704

Q39.What Happend When Ph Is Less Than The Pk Value?

When the pH is less than the pKa value the environment is more acidic leading to the protonation of acidic groups in molecules such as amino acids or functional groups in organic compounds.

Q40.Define Ionization In Terms Of Proteins.

The ionization is a process by which amino acid side chains gain or lose protons (H^+ ions) depending on the pH of the surrounding environment. This process affects the charge of the protein influencing its structure, stability and interactions with other molecules. The ionization state of specific residues can play a crucial role in enzyme activity, protein folding and overall biological function.

Q41.How homology modeling is used to sequence unknown protein sequence?

Homology modeling is used to predict the structure of an unknown protein sequence by comparing it to known protein structures with similar sequences. Researchers identify homologous proteins in databases align the unknown sequence with these templates and use the structural information from the templates to build a model of the unknown protein's 3D structure. This predicted structure can provide insights into the protein's function and interactions guiding further experimental validation.

Q42.Discuss chou fasman algorithm its improvements and limitations?

Chou fasman algorithm Predicts protein secondary structure from amino acid sequences using empirical rules.

Improvements:

1.Incorporation of solvent accessibility data to enhance predictions.
2.Use of sequence context to improve accuracy in structural predictions.
3.Integration of machine learning techniques for better predictive models.
4.Development of hybrid methods combining Chou-Fasman with other algorithms.
5.Utilization of larger and more diverse protein databases for training.

Limitations:

1.Relies heavily on statistical data which may not represent all structural variations.
2.Limited ability to predict tertiary structure or complex folding patterns.
3.May produce false positives or negatives in secondary structure predictions.
4.Does not account for dynamic conformational changes in proteins.
5.Performance can vary significantly across different protein families.

Q43.How Edman Degradation Work?

Edman degradation is a method used to identify the amino acid sequence of a protein or peptide by sequentially removing one amino acid at a time from the N-terminus. The process begins with the reaction of the N-terminal amino acid with phenylisothiocyanate (PITC) under alkaline conditions forming a phenylthiohydantoin (PTH) derivative. The PTH-amino acid is then cleaved from the peptide allowing for its identification through chromatographic techniques such as high-performance liquid chromatography (HPLC) and this process is repeated for each subsequent amino acid until the entire sequence is determined.

Prepared by:
Shahzaib Shabbir
03308553704

Q44.Advantage and disadvantage of Ab initio?

Ab initio methods are computational techniques used in bioinformatics and structural biology to predict protein structures from amino acid sequences without relying on existing structural templates.

Advantages of Ab Initio Methods:

1. ***No Template Required:*** Ab initio methods do not rely on existing structural templates, making them useful for predicting structures of novel proteins.
2. ***Comprehensive:*** They can provide insights into the full range of possible conformations and interactions within a protein.
3. ***Flexibility:*** Capable of modeling complex folds and topologies that may not be represented in known structures.
4. ***Theoretical Basis:*** Based on physical and chemical principles, allowing for a deeper understanding of protein folding mechanisms.
5. ***Advancements in Computational Power:*** Improved algorithms and computational resources have enhanced the accuracy of ab initio predictions.

Disadvantages of Ab Initio Methods:

1. ***Computationally Intensive:*** Requires significant computational resources and time especially for large proteins.
2. ***Lower Accuracy:*** Predictions may be less accurate compared to template-based methods particularly for complex structures.
3. ***Limited by Force Fields:*** The accuracy is dependent on the quality of the force fields used to model interactions.
4. ***Difficulty with Large Proteins:*** Struggles with accurately predicting the structures of larger proteins due to increased complexity.
5. ***Ambiguity in Results:*** May yield multiple plausible structures making it challenging to determine the most accurate conformation.

Q45.MS1 And MS2 Discuss?

MS1 (Mass Spectrometry 1):

MS1 refers to the first stage of mass spectrometry where ions generated from a sample are analyzed based on their mass-to-charge ratio (m/z). In this stage, the mass spectrometer separates ions allowing for the identification of the molecular weight of the analytes present in the sample. MS1 data provides a broad overview of the components in a mixture which is essential for initial characterization.

MS2 (Mass Spectrometry 2):

MS2, or tandem mass spectrometry follows MS1 and involves further fragmentation of selected ions from the MS1 stage. This fragmentation allows for the analysis of the structure and composition of specific ions providing detailed information about the molecular structure and sequence of the analytes. MS2 is particularly useful for identifying specific compounds elucidating structures and quantifying components in complex mixtures enhancing the overall analytical capability of mass spectrometry.

Q46.In silico and In vitro?

In silico refers to computer-based simulations and analyses. In vitro refers to experiments conducted in a controlled laboratory environment outside of living organisms.

Prepared by:
Shahzaib Shabbir
03308553704

Q47.What Are Uses Of Mass Spectrometry?

1. ***Protein Identification:*** Analyzes proteins and peptides to determine their molecular weight and sequence, aiding in proteomics research.
2. ***Metabolomics:*** Identifies and quantifies metabolites in biological samples providing insights into metabolic pathways and disease states.
3. ***Drug Development:*** Assists in the characterization of drug compounds including their structure, purity, and stability during the development process.
4. ***Environmental Analysis:*** Detects and quantifies pollutants and contaminants in environmental samples such as water, soil, and air.
5. ***Clinical Diagnostics:*** Used in medical laboratories to analyze biological fluids for biomarkers aiding in disease diagnosis and monitoring.
6. ***Forensic Science:*** Analyzes substances in forensic investigations, such as drugs, toxins, and biological samples to provide evidence in legal cases.
7. ***Structural Elucidation:*** Determines the structure of unknown compounds by analyzing fragmentation patterns and isotopic distributions.
8. ***Quality Control:*** Monitors the quality and consistency of pharmaceutical products and other chemical substances in manufacturing processes

Q48.Differentiate Between Bottom Up Proteomics And Top Down Proteomics?

Bottom-up proteomics:

Bottom-up proteomics and top-down proteomics differ primarily in their approach to protein analysis. Bottom-up proteomics involves digesting proteins into smaller peptides using enzymes which allows for easier handling of complex mixtures and provides detailed information about peptide sequences and post-translational modifications.

Top-down proteomics:

Top-down proteomics analyzes intact proteins directly offering insights into the complete protein structure, including isoforms and modifications but faces challenges with complex samples and may have lower sensitivity due to the difficulty in ionizing larger proteins.

Q49.What Is Hydrophobicity?

Hydrophobicity refers to the tendency of a substance to repel water or not interact favorably with it. Hydrophobic molecules or regions of proteins are typically nonpolar and avoid contact with water influencing protein folding and interactions. This property is crucial for understanding membrane formation, protein structure and the behavior of biomolecules in aqueous environments.

Q50.Role of Amino Acid?

Amino acids are the building blocks of proteins and play a critical role in various biological processes. They link together through peptide bonds to form polypeptides which fold into specific three-dimensional structures essential for protein function.Amino acids participate in metabolic pathways serve as precursors for neurotransmitters and hormones and contribute to cellular signaling and regulation.

Prepared by:
Shahzaib Shabbir
03308553704

Q51.Role of Global Protein?

Global proteins often referred to as global or housekeeping proteins play essential roles in maintaining cellular functions and homeostasis across various biological systems. They are involved in fundamental processes such as metabolism, cell structure, and signaling ensuring that cellular activities are carried out efficiently. Global proteins often serve as reference points in studies of protein expression and function providing insights into cellular health and disease states.

Q52.Advantage of Folded Protein over Denatured Protein.

Folded proteins have a specific three-dimensional structure that is crucial for their biological function allowing them to interact correctly with other molecules and perform their roles effectively. Denatured proteins lose their functional shape leading to a loss of activity and inability to participate in biological processes. The stability and functionality of folded proteins are essential for maintaining cellular processes, enzymatic reactions and overall organism health.

Q53.Shot Gun and Peptide Mass Finger Printing?

Shotgun Proteomics: a method used to analyze complex protein mixtures by digesting proteins into smaller peptides which are then separated and identified using mass spectrometry. This approach allows for the simultaneous identification and quantification of thousands of proteins in a sample.

Peptide Mass Fingerprinting (PMF): a specific technique within shotgun proteomics that involves generating a mass spectrum of peptide fragments from a digested protein. The resulting mass data is compared against known protein databases to identify the protein based on its unique mass profile providing insights into its identity and structure.

Q54.What is Smith-Waterman algorithm?

The Smith-Waterman algorithm is a dynamic programming algorithm used for local sequence alignment it compares segments of all possible lengths and optimizes the similarity measure. It is commonly used in bioinformatics to align DNA, RNA or protein sequences.

Prepared by: Shahzaib Shabbir
Whatsapp no: 03308553704

Group Admin: Shahzeb Mughal
Whatsapp no:03132310940