BIF(101) CURRENT SOLVED PAPER SPRING 2021 BY SULMAN ALI

1. Modification in protein lifecycle:

Protein synthesis occurs during a process called 'translation'. Posttranslational modification (PTM) of proteins, being one of the later stages in protein biosynthesis, refers to the reversible or irreversible chemical changes proteins may undergo after translation. In other words, PTMs are chemical modifications of a polypeptide chain that occur after DNA has been transcribed into RNA and translated into protein.

2. Types of proteins:

There are seven types of proteins: antibodies, contractile proteins, enzymes, hormonal proteins, structural proteins, storage proteins, and transport proteins.

3. Sequence analysis and its types:

in bioinformatics, sequence analysis is the process of subjecting a DNA, RNA or peptide sequence to any of a wide range of analytical methods to understand its features, function, structure, or evolution.

4. Types:

1= Pair-wise Sequence Alignment 2= Multiple Sequence Alignment.

5. Data retrieval:

Data retrieval means obtaining data from a Database Management System (DBMS) such as ODBMS. ... Then the DBMS selects the demanded data from the database. The retrieved data may be stored in a file, printed, or viewed on the screen. A query language, such as Structured Query Language (SQL), is used to prepare the queries.

6. Modeling for Homology AND Ab initio methods:

Homology modeling of the target structure can be done as follows:

1. Template recognition and initial alignment 2. Alignment correction 3. Backbone generation 4. Loop modeling 5. Sidechain modeling 6. Model optimization 7. Model validation.
2. Ab initio methods have Anfinsen's thermodynamic hypothesis at the center • These methods attempt to identify the structure with minimum free energy. Need for Ab Initio Modelling • Applicable to any sequence • Not very accurate biologically • Accuracy and applicability are limited by our understanding of the protein folding problem.

7. Function of lysosomes:

Degradation of defunct intracellular organelles and material taken in from the outside of the cell by endocytosis.

8. Threading:

Threading is a knowledge-based tertiary structure prediction method. With this method, the software attempts to align the full backbone plus side chains of a sequence with an unknown structure to just the backbone (no side chains) of a known structure.

9. Limitations of structural Modelling:

1= Large Bias towards structure of template 2= Cannot study conformational changes 3= Cannot elicit new catalytic/binding sites.

10. Applications of computations:

Investigating Interactions Between Proteins and Nucleic Acids by Computational Approaches. Bioinformatics Analysis of MicroRNA and Protein-Protein Interaction in Plant Host-Pathogen Interaction System. Computational Modelling of the Alu-Carrying RNA Network in Th17-Mediated Autoimmune Diseases.

11. Principle of the homology modeling:

Homology modeling is also known as comparative modeling predicts protein structures based on sequence homology with known structures. It is based on the principle that "if two proteins share a high enough sequence similarity, they are likely to have very similar three-dimensional structures."

12. Process of gene expression:

It consists of two major steps: transcription and translation. Together, transcription and translation are known as gene expression.

- During the process of transcription, the information stored in a gene's DNA is passed to a similar molecule called RNA (ribonucleic acid) in the cell nucleus.
- Translation, the second step in getting from a gene to a protein, takes place in the cytoplasm. The mRNA interacts with a specialized complex called a ribosome, which "reads" the sequence of mRNA nucleotides.

13. Folding process:

Protein folding is a process by which a polypeptide chain folds to become a biologically active protein in its native 3D structure.

The amino acids in the chain eventually interact with each other to form a well-defined, folded protein. The amino acid sequence of a protein determines its 3D structure.

14. Types of protein sequence techniques:

The two major direct methods of protein sequencing are mass spectrometry and Edman degradation using a protein sequenator (sequencer). Mass spectrometry methods are now the most widely used for protein sequencing and identification but Edman degradation remains a valuable tool for characterizing a protein's N-terminus.

15. Types of GEO:

1)Sample (GSM) preparation and description of the samples.

2) Platform (GPL) technology used and the features detected – microarray or RNASeq

3) Series (GSE) defines a set of samples and how they are related

4) Datasets (GDS) sample data collections assembled by GEO.

16. Applications of computational genomics:

graph algorithms for genome mapping (the process of locating fragments of DNA on chromosomes) and for certain types of DNA and peptide sequencing methods, clustering algorithms for gene expression analysis and phylogenetic reconstruction, and pattern matching for various sequence search problems.

17. Basic modes of protein Localization:

1. Gated transport 2. Transmembrane transport 3. Vesicular transport

18. Alignment correction in structural modeling:

1= Fine tune and adjust the BLAST alignments 2= Use MSA tools (e.g. ClustalW), to find the residues and properties that need to be conserved. 3= Examine the template structure to check which residues are in the core hence less likely to change than the residues at the outside. 4= Insertions and deletions can be made in those parts of the sequence which are highly variable. 5= Gaps have to be shifted around until they are as small as possible.

19. What is database:

A biological database is a large, organized body of persistent data, usually associated with computerized software designed to update, query, and retrieve components of the data stored within the system

20. What is algorithm:

Bioinformatics is the study of living organisms viewed as information processors. We study algorithms for sequence alignment, motif finding and gene finding, construction of phylogenetic trees, and structure prediction.

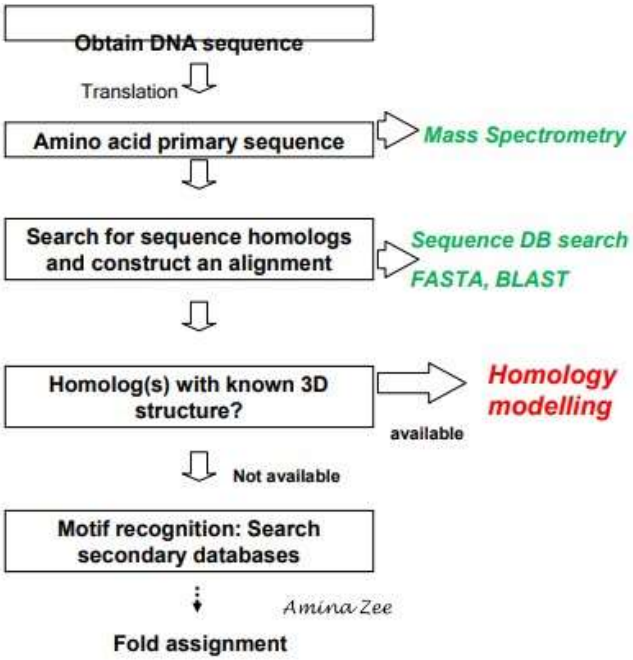
21. Difference between genomics and proteomics:

The main difference between genomics and proteomics is that genomics is the study of the entire set of genes in the genome of a cell whereas proteomics is the study of the entire set of proteins produced by the cell.

22. Applications of proteomics:

Proteomics-based technologies are utilized in various capacities for different research settings such as detection of various diagnostic markers, candidates for vaccine production, understanding pathogenicity mechanisms, alteration of expression patterns in response to different signals and interpretation of functional protein pathways in different diseases. Proteomics is practically intricate because it includes the analysis and categorization of overall protein signatures of a genome. Its also applicable in personalized medicines, biomarker discovery, drug discovery and development, systems biology, agriculture, food sciences, poleo proteomics and in astrobiology.

23. Workflow of Structural Modelling:



24. Difference between the I-tasser and modeler:

I-TASSER (Iterative Threading ASSEmbly Refinement) is a bioinformatics method for predicting three-dimensional structure model of protein molecules from amino acid sequences.

MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2).

25. Force field:

The force field is a collection of equations and associated constants designed to reproduce molecular geometry and selected properties of tested structures.

26. TYPES OF PROTEOMICS:

Expression proteomics • Structural proteomics • Functional proteomics

1. Expression proteomics is used to study the qualitative and quantitative expression of total proteins under two different conditions.
2. Structural proteomics helps to understand three dimensional shape and structural complexities of functional proteins.
3. Functional proteomics explains understanding the protein functions well as unrevealing molecular mechanisms within the cell that depend on identification of the interacting protein partners.

27. Optical energy function rationale:

Ab initio methods rely on computing the energies of folded proteins • The protein structures with the lowest energy are declared as plausible predictions. Sometimes it so happens that even slightly homologous proteins may not be available.

28. FASTA and gene bank:

FASTA is a sequence Format. DNA sequences are stored in specified formats. FASTA is the most frequently used format to present DNA and Protein sequences.

Gene banks are a type of biorepository that preserves genetic material. ... In an effort to conserve agricultural biodiversity, gene banks are used to store and conserve the plant genetic resources of major crop plants and their crop wild relatives.

29. Entrez:

is an integrated search engine which allows users to search and retrieve different data from the NCBI. Entrez is a molecular biology database system that provides integrated access to nucleotide and protein sequence data, gene-centered and genomic mapping information, 3D structure data, PubMed MEDLINE.

30. Function of mitochondria:

Mitochondria role in ATP synthesis. Mitochondria are membrane-bound cell organelles (mitochondrion, singular) that generate most of the chemical energy needed to power the cell's biochemical reactions. Chemical energy produced by the mitochondria is stored in a small molecule called adenosine triphosphate (ATP).

31. NCBI: (National Center for Biotechnology Information)

NCBI NCBI has two options:

1) BANKIt – For simple sequences and annotations – For submission through web – Do not require advanced tools 2) Sequin – For Complex sequences and annotations – For off-line submissions – Do require advanced tools and graphical reports.

32. BLAST:

Basic Local Alignment Search Tool BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

33. Modeller of software and its use:

it is a Online tool for predicting protein structures Uses: Template recognition and initial alignment Alignment correction Backbone generation Loop modeling Side-chain modeling Model optimization Model validation Template recognition and initial alignment Alignment correction Backbone generation Loop modeling Side-chain modeling Model optimization Model validation.

34. Z-Score:

A Z-score is simply the comparison of an actual alignment score with the scores obtained on a set of random sequences by a Monte-Carlo process.

35. Protein database DNA:

The Protein-DNA Interface database (PDIdb) is a repository containing relevant structural information of Protein-DNA complexes solved by X-ray crystallography and available at the Protein Data Bank.