PREPARED AND COMPOSED BY ZAINAB ARSHAD

BIF 101 ALL IMPORTANT QUESTIONS FOR FINAL TERM

FROM TOPICS 99 TO 150

If you don't design your own life plan, chances are you'll fall into someone else's plan. And guess what they have planned for you? Not much

You will got a surprise at the end of this file

Zainab Arshad Official

SQ: What is National Human Genome Research Initiative (NHGRI)?

Answer: National Human Genome Research Initiative (**NHGRI**) a federally funded organization (NIH) started in **1988** **(Francis Collin)**

- Celera, a commercial vendor **(Craig Venter)** joined the venture in 1998. Both

simultaneously announced the completion of the project in 2000.Total 3.4 billion bases sequenced at a cost of $1/base

SQ: Define: Gene Expression Omnibus (GEO)

A public repository for the archiving and distribution of gene expression data submitted by the scientific community

Sq: what are the uses and benefits of Gene Expression Omnibus (GEO)?

Convenient for deposition of gene expression data, as required by funding agencies and journals

Curated resource for gene expression data

Browsing, querying, analysis and retrieval

SQ: MCQs: How many records of GEO ?

GEO **has four kinds** of records

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

Zainab Arshad Official

Platform (GPL) technology used and the features detected

microarray or RNASeq

GEO has four kinds of records

2) **Series (GSE)**

defines a set of samples and how they are related

3) **Datasets (GDS)**

sample data collections assembled by GEO

SQ: Define health informatics?

- It deals with the resources, devices, and methods required to optimize the acquisition, storage, retrieval, and use of information in health and biomedicine.

Q:What is MEDLINE?

- MEDLINE is the primary resource for biomedical journal articles

Zainab Arshad Official

- Millions of citations to articles in biomedical journals
- MEDLINE uses the MeSH vocabulary

Q: How many options tools of NCBI ?

NCBI has two options

- **BANKIt**
  - For simple sequences and annotations
  - For submission through web
  - Do not require advanced tools
- **Sequin**
  - For Complex sequences and annotations
  - For off-line submissions
  - Do require advanced tools and graphical reports

SQ: What are different sequences formats?

Zainab Arshad Official

Answer:

FASTA

Genebank format

EMBL Format

SwissProt Format

XML Format

Q: What do mean by proteomics?

- **A complete description of proteins expressed in any given cell at any given time**

Is called as proteomics

Proteomics is the large-scale study of proteins, usually by biochemical methods. The word proteomics has been associated traditionally with displaying a large number of proteins from a given cell line or organism on two -dimensional polyacrylamide gels.

Zainab Ansrad Official

Q: Why should we study Proteomics? Or What is the importance and uses of proteomics?

- It directly contributes to drug development
- Verification of a gene product by proteomic methods
- Modifications of the proteins
- Protein expression level
- Many types of information cannot be obtained from the study of genes alone. For example, proteins, not genes, are responsible for the phenotypes of cells. It is impossible to elucidate
  mechanisms of disease, aging, and effects of the
  environment solely by studying the genome.
- **Q: What are the Advantages of study of proteomics?**

Zainab Arshad Official

- Shows that genetic alterations are not the reason
  for all types of diseases
  Helps in determining the proper treatment of diseases
  With the help of three dimensional analysis of proteins we have found that HIV protease is the enzyme which is responsible for AIDS.
  One of the most important use of proteomics in diagnosis is the identification of biomarkers. The study of drugs in proteomics is called pharmacoproteomics.

Q: Define pharmacoproteomics?

One of the most important use of proteomics in diagnosis is the identification of biomarkers. The study of drugs in proteomics is called pharmacoproteomics

Q: Name post-translational modifications (PTMs)?

Zainab Arshad Official

- C-terminal GPI anchor
- phosphorylation
- sulfation
- glycosylation
- N-myristoylation
- hydroxylation
- N-methylation
- carboxymethylation
- signal peptidase site.....

Q: What is Central Dogma?

The central dogma states that information in nucleic acid can be perpetuated or transferred, but the transfer of information form into protein is irreversible.

- DNA -> RNA -> Protein Synthesis
- **Q: Define transcription and translation?**

Zeinab Arshad Official

- Transcription:
  - Process of DNA serving as a template for RNA synthesis
- Translation:
  - Process of RNA serving as a template for protein synthesis

Q: What are the types of RNAs?

- mRNA: Messenger RNA
  - 1st RNA's made DIRECTLY from DNA template
  - Travel from nucleus to ribosome
- rRNA: Ribosomal RNA
  - Helps form ribosomes in cytoplasm
- tRNA: Transfer RNA
  - Brings amino acids from cytoplasm to ribosome so proteins can be made

Zaman Arshad Official

SQ: Write the Scope and types of Proteomics?

- Expression proteomics
- Structural proteomics
- Functional proteomics

SQ: Define EXPRESSION PROTEOMICS?

Expression proteomics is used to study the qualitative and quantitative expression of total proteins under two different conditions.
Normal and diseased state.
E.g. :tumor or normal cell.
It studied that protein is over expressed or under expressed. The 2-D electrophorasis technique is used.

SQ: Define STRUCTURAL PROTEOMICS?

¢ Structural proteomics helps to understand three
dimensional shape and structural complexities of

Zainab Arshad Official

functional proteins.

¢ It determine either by amino acid sequence in protein or

from a gene this process is known as **homology modeling**

SQ: Define FUNCTIONAL PROTEOMICS?

Functional proteomics explains understanding the protein functions as well as unrevealing molecular mechanisms within the cell that depend on the identification of the interacting protein partners.

SQ:LQ: What is structural proteomics and how we determine the structures of proteins? Explain the techniques?

- For High-throughput determination of the 3D structure of proteins we study the structural proteomics. It involves following techniques for determinations of protein structures.
-

Zainab Irshad Official

- 1) Direct determination - X-ray crystallography and
- 2) nuclear magnetic resonance (NMR).
- Prediction
  - 3) Comparative modeling -
  - 4) Threading/Fold recognition
  - 5) Ab initio

Q: How many different strategies for structure prediction of proteins? Explain them

Answer:

1. **Homology Modelling**
2. **Threading/Fold Recognition**
3. **Ab Initio Modelling**

Q: What is Swiss-Prot?

- **Swiss-Prot is Collaboration between the SIB and EBI**

Zainab Arshad Official

- Weekly releases from about 50 servers across the world, the main source being ExPASy in Geneva
- Q : What is Use of shotgun sequencing method?
- A: speeded up the process
- 1.8 million bp
- Took 9-months and cost was 1 million
- Paved the way for sequencing of many other organisms
-
- First attempt to sequence free living organism was launched in late
- 1990's

Q: the first published genome was

- A: *Haemophilus influenza*

Zeinab Arshad Official

- **Q: TIGR stand for**
- **A: The Institute of Genome Research**

Q: mcq: PIR stand for?

PIR ((Protein Information Resource)

MIPS stand for

Munich Center for Protein Sequences

JIPID stand for

Japan International Protein Information Database

SQ: What is Protein domain?

- "clearly recognizable portion of a protein that folds into a defined structure"is called as protein domain

Q: What are Main secondary structure elements?

Zainab Ahmad Official

- a-helix - right handed helical structure
- b-sheet - composed of two or more b-strands, conformation is more "zig-zag" than helical.

Q: what is crystallography?

- Make crystals of your protein
  - 0.3-1.0mm in size
  - Proteins must be in an ordered, repeating pattern.
- X-ray beam is aimed at crystal and data is collected.

  Structure is determined from the diffraction data

Q: Write the comparison or difference between Genomics and proteomics?

- Genomics and proteomics are closely-related fields.

Zainab Arshad Official

- 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.

  Genomics: The genome is constant. Every cell of an organism has the same set of genes.

  Proteomics: The proteome is dynamic and varies. The set of proteins produced in different tissues varies according to the gene expression

  Genomics: High throughput techniques are used in the genomics to map, sequence, and analyze genomes.

  Proteomics: In proteomics, characterization of the 3D structure and the function of proteins is carried out by the use of high throughput methods.

Zainab Arshad Official

Genomics: The techniques involved in genomics include gene sequencing strategies such as directed gene sequencing, whole-genome shotgun sequencing, construction of expressed sequence tags (ESTs), identification of single nucleotide polymorphisms (SNPs)

Extraction and electrophoretic separation of proteins.

Digestion of proteins with the use of trypsin into small fragments.

Determination of the amino acid sequence by mass spectrometry.

Identification of proteins using the information in the protein databases.

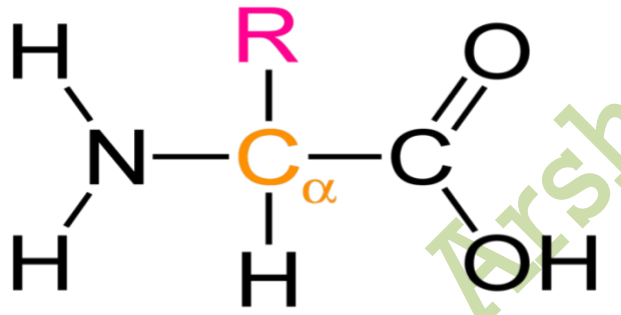
SQ: Which reactions occurs in Cytosol?

- most protein synthesis, glycolysis and metabolic pathways synthesizing amino acids, nucleotides.

Zainab Arshad Official

SQ: LQ: What is the Chemical composition of proteins?

- Proteins are polymers of amino acids.
- They range in size from small to very large.
- All the proteins are made up of Twenty different types of amino acids. So these amino acids are called standard amino acids.



The chemically proteins composed of an amine group, an carboxylic group , ahydrogen atom and an side alkyl group .

Sq: lq: what is Homology Modelling?

- **Protein sequence gives rise to its structure**

Zainab Arshad Official

- If another protein which has a similar sequence also has its structure known, the structure of an unknown protein can be predicted based on that similar protein
- So, it is then possible to identify unknown protein structures by just examining the homologous protein sequences
- In homology modelling, proteins with similar 1' sequences are considered
- Given that one of them has its 3' structure known, then the 3' structure of other protein can be predicted
- Homology modelling is used to predict structures of proteins having high

Zainab Arshad Official

sequence similarity with other proteins with known structures!

SQ: What are the steps of even Steps to Homology Modelling?

Template recognition and initial alignment

Alignment correction

Backbone generation

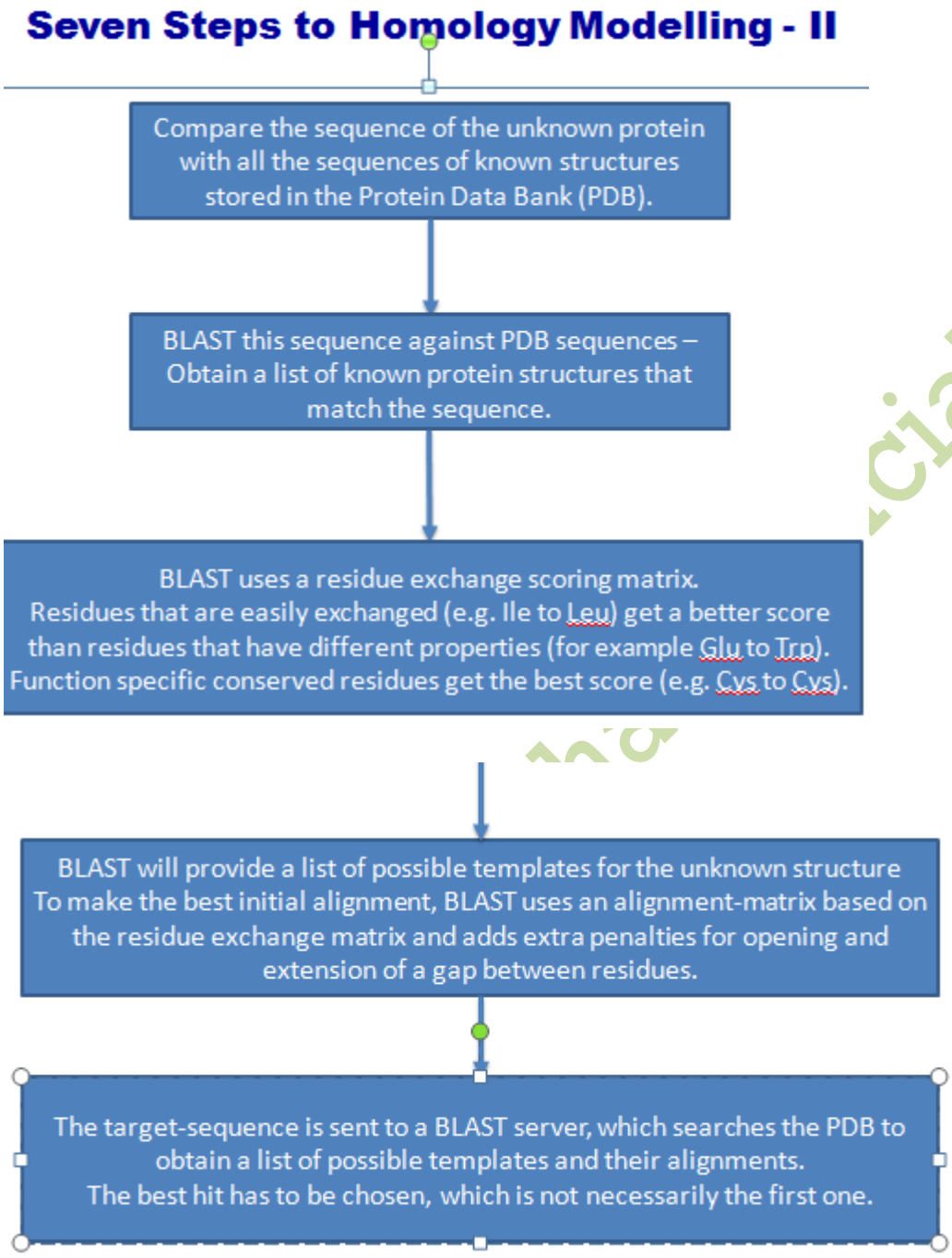
Loop modeling

Side-chain modeling

Model optimization

Model validation

Zainab Arshad Official



SQ: HAT ARE THE MODELLER for Homology Modelling?

.log : log output from the run.

.B* : model generated in the PDB format.

.D* : progress of optimisation.

.V* : violation profile.

.ini : initial model that is generated.

.rsr : restraints in user format.

.sch : schedule file for the optimisation process

SQ:LQ: What is Fold Recognition – Threading?

- **Fold recognition: Finding the best fit of a sequence to a set of candidate folds**
- **A protein fold is defined by the way the secondary structure elements of the structure are arranged relative to each other in space.**
- **Common folds include 4-helix bundle and the TIM barrel.**

Zainab Arshad Official

- **Fold recognition is also called Threading**
- **Technique for predicting protein structures**
- **Employed when homology modelling cannot predict quality structures**

- In the process of "Threading", we mount an amino acid sequence on to the backbone of template structures in a folds library
- Each step is "drag" along the sequence (MQVKLFTY…) through each location of each template fold
- Then, for each fold, we must compute the fitness of sequence matching that fold!
  - **Combinations of secondary structures come together to form the best prediction**

Zainab Arshad Official

- **Scoring typically involves using a Z-Score function based on energy of a molecule**

SQ:LQ: What is the Iterative threading assembly refinement (I-TASSER) server ?

- Software for automated protein structure &function prediction based on the sequence-to-structure-to-function.
- Steps:

1) Starts from amino acid sequence
2) i-TASSER first generates 3D atomic models from multiple threading alignments and iterative structural assembly simulations.
3) The function of the protein is then inferred by structurally matching the 3D models with other known proteins.

Zainab Arshad Official

4) Outputs full-length secondary & tertiary structures and functional annotations on ligand-binding sites

5) An estimate of accuracy of the predictions is provided based on the confidence score of the modeling

6) **iTASSER helps thread amino acid sequences on fold and secondary structure databases**

7) **It also helps predict function of structures output**

SQ: What are Advantages and Disadvantages of Threading ?

AdvantageS:

1) **Threading is a technique for predicting protein structures**

Zainab Arshad Official

2) It is useful in cases where homology modelling fails to predict quality structures

3) Threading helps predict secondary structures of proteins towards tertiary structure prediction

4) For the "Twilight Zone" with low alignment quality and identity, threading is useful

Disadvantage:

1) Novel proteins cannot be predicted using threading
2) Fewer than 30% of the predicted first hits are true remote homologues

3) Validation of each result is necessary

- **Question: what are the Strategies for Ab Initio Modelling?**

Zainab Arshad Official

1) Energy Optimization in Ab Initio Modelling

2) Start with a rough initial model.
3) Define an energy function mapping structures to energy values. We have to minimize this later!!
4) Solve the computational problem of finding the global minimum.
5) Simulation of the Folding Process
6) Build an accurate initial model (including energy and forces).
7) Accurately simulate the dynamics of the protein folding process.
8) The native structure will steadily emerge

- Question: why proteins folde?
- Protein Fold structures in order to obtain the most stable thermodyanamic structure

Zainab Arshad Official

- The protein structure reporting lowest energy is selected to be the optimal structure
- The protein structure reporting lowest energy is selected to be the optimal structure
- Q: What is Ab initio methods?
- Ab initio methods predict protein structures by folding proteins based on each constituent atom's volume, charge, mass etc.

  Sq: DESCRIBE THE Pros and Cons of Ab Initio Modelling?
- The Native structure not always at the global minimum
- No clear way of choosing among alternative structures that are generated
- Advantages

Zainab Arshad Official

-

1) Ab Initio methods can fold any target sequence using only physical atomic properties
2) Predictions are mostly accurate and correctly describe the natural folding process

- Disadvantages
-

1) Ab initio methods are the very difficult to design (energy function)
2) These methods are slow due to the huge possibilities
3) An order of 10^{12} steps are needed to simulate protein folding for medium sized protein structures

- Sq: what are the Challenges in Ab Initio Modelling ?

1) Very hard to accurately describe energy functions that can reliably

Zainab Arshad Official

discriminate native and non-native structures.

2) Enormous amount of computations.

- **Sq: how many Types of Alignments:**
- 1) Global Alignment
- (Needle Wunsch) is a tool gives global alignment of sequences
- 2) Local Alignment
- (Smith Waterman) is a tool gives the local alignment of sequence
- **SQ: How many types of Phylogenetic Trees?**
- Phylogenetic Trees

1) Scaled Trees
2) Unscaled Trees
3) Rooted Trees
4) Unrooted Trees

- **SQ: What are the Techniques of protein sequencing**

Zainab Arshad Official

1) Edman Degradation
2) Mass Spectrometry
3) Protein Ionization
4) Mass Analysis
5) Protein Fragmentation

Sq: What are the online tools for Composite Scoring Schemes

Online tools:

1) Mascot
2) Sequest
3) Prosight PC

SQ: How many trypes of protein structures?

Protein Structures are generally of four types:

- Primary
- Secondary

Zainab Arshad Official

- Tertiary
- Quaternary

- **SQ: Write the name of Types of Protein Secondary Structures**
  -
  - Types of Protein Secondary Structures
  - Helices
  - Beta Sheets
  - Coils
  - Loops

SQ: Write the name of Algorithm for predicting protein structures

1) Chou Fasman Algorithm
2) Protein Structure Database - PDB
3) Online tools for predicting structures by using proteins sequences

Zaraab Arshad Official

SQ: What are the carrer in bioinformatics?

Careers in Bioinformatics

Unique opportunity for us in Pakistan

Bioinformatics requires two things

1) **Smart mind**
2) **Internet connected computer**
3) **You can take public databases and design drugs**
4) **You can make a startup company which manages and process health BIGDATA!**
5) **All it needs is basic software development skills coupled with Bioinformatics**
6) **The next Google, Facebook and Uber is going to emerge from Health and Bioinformatics**
7) **Pharmaceutical companies are investing into bioinformatics human resource development**

Zainab Arshad Official

JOBS :

1)) Pharmaceutical Giants
2) Research Centers & Universities
3) Hospital & Diagnostic IT departments
4) Your own startup company

Hmm if u has just scrolled this file to see surprise its means you gain nothing

But when you come down this files by studying and understanding it , it means you are prepare for your exams and I think your success is enough for your surprise. Got it?

As **John D. Rockefeller** said

If you want to succeed you should strike out on new paths, rather than travel the worn paths of accepted success.

Effort by Zainab Arshad