Table of Contents

1. Differences 1
1.1. Tertiary structure and Quaternary structure of protein 1
1.2. Greek key motif and Gelly roll motif 1
1.3. Domain and structure motif 2
1.4. Accession no of GenBank nucleotide sequence and protein sequence 2
1.5. Haemoglobin and myoglobin 2
2. Identify the type of domain in the given three proteins and give an example and write two features of the domain. 3
2.1. A. The TIM Barrel 3
2.2. B. The Complete g-crystallin Molecule 3
2.3. C. The Fold of IgG Domains 3
3. Question 3 3
3.1. Three different Databases in Bioinformatics and write their significance 3
3.1.1. NCBI 3
3.1.2. UniProt 4
3.1.3. EBI 4
3.2. Algorithm for sequence alignment 5
3.2.1. Determine the substitution matrix and the gap penalty scheme 5
3.2.2. Now make first row and column zero. 5
3.2.3. Now main work starts here. Fill the scoring matrix using following. 5

1. Differences

1.1. Tertiary structure and Quaternary structure of protein

Tertiary structure	Quaternary structure
A tertiary structure is knowing as 3-dimensional structure of polypeptide.	Some proteins are made up of multiple polypeptide chains, also known as subunits. When these subunits come together, they give the protein its quaternary structure.
Disulphides bond is special type of covalent bond that can contribute to tertiary structure	The same types of interactions that contribute to tertiary structure (mostly weak interactions, such as hydrogen bonding and London dispersion forces) also hold the subunits together to give quaternary structure.
Example: TIM Barrel	Example: Haemoglobin, DNA polymerase

1.2. Greek key motif and Gelly roll motif

Greek key motif	Gelly roll motif
• In Greek key motif antiparallel sheets are connected and one of them is not connected using “hair pin”. There are total 4 sheets in Greek key motifs. • The motif occurs when strand number n is connected to strand n + 3 or n - 3 instead of n + 1 or n - 1 in an eight-stranded antiparallel b sheet or barrel. The two different possible connections give two different hands of the Greek key motif.	• This is a fold topology that classically consists of four Greek key motifs that adopt an eight-stranded beta sandwich structure. • In this fold the hydrogen bonding pattern between adjacent strands is broken in two places, and as a consequence the structure comprises two four-stranded beta sheets. Both sheets are purely antiparallel, with strands adjacent in sequence appearing in different sheets with the exception of the fourth and fifth strands, which are in the same sheet. This leads to a structure with only one hairpin; all other beta-beta connections are arches.

1.3. Domain and structure motif

Domain	Motif
Domains are regions of a protein that has a specific function and can (usually) function independently of the rest of the protein	Motif is a certain grouping of the super secondary elements of proteins such as alpha helices and beta structures
Domain evolves, functions, and exists independently of the rest of the protein chain.	Motifs perform similar biological functions through a particular protein family
Domain is stable.	Motif is not stable by itself
For example, Pyruvate kinase	For example, Zinc fibre motif

1.4. Accession no of GenBank nucleotide sequence and protein sequence

Nucleotide sequence	Protein sequence
letter + 5 numerals 2 letters + 6 numerals 2 letters + 8 numerals	3 letters + 5 numerals 3 letters + 7 numerals
Example: Arabidopsis Thaliana Accession no. AF165912	Example: Arabidopsis Thaliana protein_id="AAD45922.1"

1.5. Haemoglobin and myoglobin

Haemoglobin	Myoglobin
Haemoglobin has 4 chains of two different types- alpha and beta, delta, gamma, or epsilon	It contains single polypeptide chains
A tetramer	A monomer.
Sigmoid binding curve	Hyperbolic curve.
Haemoglobin is transported along with blood to whole body and carry oxygen.	Myoglobin supplies oxygen to muscles only, which is helpful at the starving time of oxygen.

2. Identify the type of domain in the given three proteins and give an example and write two features of the domain.

2.1. A. The TIM Barrel

- A core of twisted parallel b strands arranged close together, like the staves of a barrel.
- Here shows a closed barrel exemplified by schematic and topological diagrams for the enzyme triosephosphate isomerase.

2.2. B. The Complete g-crystallin Molecule

- The two domains of the complete molecule have the same topology; each is composed of two Greek key motifs that are joined by a short loop region.
- There is a greater amino acid sequence homology between the domains than the motifs within each domain, suggesting that the four Greek Key motifs in g-crystallin are evolutionarily related by gene duplication and fusion.

2.3. C. The Fold of IgG Domains

- Beta strands labelled A-G of the constant and variable domains of immunoglobulins have the same topology and similar structures. There are two extra b strands, C' and C" (red) in the variable domain. The loop between these strands contains the hyper-variable region CDR2. The remaining CDR regions are at the same end of the barrel in the loops connecting b strands.

3. Question 3

3.1. Three different Databases in Bioinformatics and write their significance

There are many databases in bioinformatics, some of them are:

1. **NCBI** (National Centre for Biotechnology Information)
2. **Uniprot**
3. **EBI**

3.1.1. NCBI

The NCBI houses a series of databases relevant to biotechnology and biomedicine and is an important resource for bioinformatics tools and services. Major databases include GenBank for DNA sequences and PubMed, a bibliographic database for the biomedical literature.
In NCBI we can search the sequence of different organisms from different databases

NCBI consist of databases like

- Conserved Domain Database (CDD)
- Database of Genomic Structural Variation (dbVar)
- Database of Genotypes and Phenotypes (dbGaP)
- Database of Short Genetic Variations (dbSNP)
- GenBank
- Gene

NCBI Popular Resources

- PubMed
- BLAST
- Nucleotide
- Genome
- SNP

3.1.2. UniProt

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

Uniprot Functions

- BLAST
- Align
- Retrieve/ID mapping
- Peptide Search
- SPARQL

UniProt Searches

- UniProtKB
- UniREf
- UniParc
- Proteomes
- Annotation
- Sequence Alignment

3.1.3. EBI

The European Bioinformatics Institute (EMBL-EBI) is part of EMBL, Europe's flagship laboratory for the life sciences. More about EMBL-EBI and our impact.

EBI Genomics Resources

- Ensembl
- Ensembl Genomes (ivertebrate metazoans, fungi, plants, bacteria)
- European Nucleotide Archive (ENA)
- European Genome Phenome Archive (EGA)
    - EMBL-Bank
    - Trace Archive
    - Sequence Read Archive (SRA)

3.2. Algorithm for sequence alignment

Let A=a1, a2, a3… and B=b1, b2, b3…. be the sequence to be aligned then algorithm is given by.

3.2.1. Determine the substitution matrix and the gap penalty scheme

Match=3; Mismatch=-3; (you can choose your own)

3.2.2. Now make first row and column zero.

M(x,0) =0; and MN (0, y) =0 where {x, y=0 to 'n', 0 to m} n, m length of sequence respectively.

3.2.3. Now main work starts here. Fill the scoring matrix using following.

Match M (I, j) = {M (i-1, j-1) +3] and draw arrow from that cell to its diagonal or from M (I, j) to M (i-1, j-1). If (value < 0 then algorithms 0 and no arrow).

For gap or mismatch =M (I, j) =Max {(M (i-1, j)-3) and (M (i, j-1)-3}
If M (i-1, j)-3=M (I, j-1) then M (I, j) =M (i-1, j)
(draw arrows from filled box to directing box from which we have taken value)
If both < 0, then M(Imp)=0 By algorithm incase no arrow drawn.

1. Fill Scoring matrix using above formulas till no box is left and draw arrows from where we are giving box a value.
2. Traceback. Starting at the highest score in the scoring matrix M and ending at a matrix cell that has a score of 0, traceback based on the source of each score recursively to generate the best local alignment.