Biophysics - BIO361
Mid-Semester Exam

Total Marks: 30
Duration: 1h

Part1- Multiple Choice Questions
Total marks - 10*1=10 marks
Each Question is for one mark. No negative marking. Attempt all questions.

1. Which of the following are basic amino acids?

   a) Glutamine
   b) Lysine
   c) Aspartic acid
   d) Arginine
   e) Histidine

   Options:

   1. a and c
   2. **b, d, and e**
   3. b and d
   4. a, c, and e

2. You are tasked with predicting the 3D structure of a newly discovered protein. The amino acid sequence of this protein shows high similarity to a protein with a known structure. Which of the following steps (in the order of steps) would you follow to successfully build a homology model?

   a) Identify a template protein with a known 3D structure that has high sequence similarity to your target protein.
   b) Run molecular dynamics simulations to explore various conformations of the target protein.
   c) Align the sequence of the target protein with the sequence of the template protein.
   d) Use the known structure of the template protein to predict the 3D structure of the target protein.
   e) Refine the model by energy minimization and structural validation.

   Options:

   1. **a, c, d**
   2. b, d, e
   3. a, b, d
   4. c, d, e

3. Which of the following secondary structure prediction methods utilizes PSI-BLAST profiles for prediction?

a) JPRED Consensus prediction
b) PSI-pred
c) PREDATOR
d) nnPredict

4. Which of the following correctly pairs the method with its application?

a) Chou–Fasman method – Sequence alignment
b) Smith–Waterman algorithm – Global sequence alignment
c) Needleman–Wunsch algorithm – Local sequence alignment
d) Chou–Fasman method – Secondary structure prediction

5. Does the central dogma of molecular biology still hold true? Choose the most accurate explanation.

A) Yes, because genetic information always flows in a linear sequence from DNA to RNA to protein, without any exceptions.

B) No, in reverse transcription, information flows from RNA back to DNA, but the overall framework of the central dogma remains relevant.

C) No, because proteins can directly transfer genetic information back to DNA, disproving the central dogma.

D) No, because the central dogma has been completely replaced by new models of molecular biology.

6. What is the use of GOR (Garnier–Osguthorpe–Robson) and Chow-Fasman

a) Secondary structure prediction
b) Multiple sequence alignment
c) Global sequence alignment
d) Quaternary structure prediction

7. A patient exhibited a single amino acid mutation in a critical protein, leading to a drastic effect. In which secondary structure is this mutation most likely to have occurred, and why?

A) Alpha helix, because mutations in this structure disrupt hydrogen bonds that stabilize the compact helical formation, potentially altering protein function.

B) Beta sheet, because mutations here lead to minimal impact on the protein, as beta sheets are more flexible and less critical for function.

C) Random coil, because this structure is highly ordered and even a minor mutation disrupts its extensive hydrogen bonding network.

D) Disulfide bridge, because this is the most common secondary structure, and a mutation directly affects the covalent bonds.

8. Which amino acid is known as a "helix breaker" due to its rigidity?

a) Glycine
b) Proline
c) Serine
d) Tyrosine

9. Which of the following statements is true about secondary structures in proteins?

a) They are stabilised by covalent bonds.
b) They refer to the 3D arrangement of the entire polypeptide chain.
c) They are formed by hydrogen bonds between backbone atoms.
d) They only include the alpha helix structure.

10. Which two amino acids are considered as exceptions in Ramachandran plot.

a. Alanine and valine
b. Glycine and proline
c. Glycine and Histidine
d. Methionine and Valine

Part 2 - Short Answer-type Questions (2 marks each)
Total marks - 5*2 = 10 marks

1. How are proteins classified according to their structure, write a few points about the properties of each, according to the CATH classification.

- **The CATH (Class, Architecture, Topology, Homology) protein structure classification system consists of four major levels of classification:**
- **Class:** The first level of classification is based on the secondary structure composition of proteins. Proteins are divided into four main classes: alpha, beta, alpha-beta, and few secondary structures.

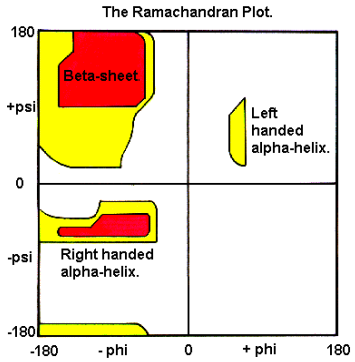
- **Architecture:** The second level of classification is based on the overall shape and arrangement of secondary structure elements in the protein. E.g. alpha-helical or beta-sheet structure, or a combination of both.
- **Topology:** The third level of classification is based on the connectivity of secondary structure elements within the protein.
- **Homology: T**he fourth level of classification is based on sequence similarity and evolutionary relationships between proteins. Proteins with similar sequences and structures are grouped together into families and superfamilies based on their evolutionary relationships, even if they have different functions.

2.Discuss the use of a Ramachandran plot in protein structure validation. Answer the following:

A. What does a Ramachandran plot represent?

The **Ramachandran plot** represents the distribution of the **ϕ** and **ψ** dihedral angles of the polypeptide backbone. It shows which combinations of these angles are sterically allowed or disallowed based on their spatial relationships. The plot is divided into regions:

a. **Allowed regions**: These correspond to favourable ϕ, ψ angle combinations that lead to stable and energetically favourable conformations.
b. **Disallowed regions**: These are combinations that result in steric clashes or unfavourable interactions, leading to unstable or impossible conformations.



B. What indicates a high-quality protein model on this plot?

A Ramachandran plot for a well-refined protein structure typically shows:

a. A large number of points in the favoured regions.

b. A smaller number of points in the allowed regions.
c. Very few or no points in the disallowed regions.
d. Distinct clustering of points corresponding to common secondary structures.

C. What might suggest errors in the protein model based on the plot?

Errors can be identified if a significant number of residues fall into disallowed regions or if the overall distribution deviates from typical secondary structure expectations.

3. Water boiling at the mountain versus deep under the sea. Which boils faster? Explain with Boltzmann distribution

- Mountain (Lower Pressure): At high altitudes, because the pressure is lower, fewer molecules need to have high energy to overcome atmospheric pressure and boil. The Boltzmann distribution shows that even at a lower temperature, there are enough molecules with sufficient energy to escape into the gas phase, leading to a faster boiling process (though at a lower temperature).

- Deep Underwater (Higher Pressure): At high pressure, the boiling point increases, meaning that the temperature needs to be much higher for a sufficient number of molecules to have the energy required to transition to the gas phase. The Boltzmann distribution would shift to higher energy values at higher temperatures, but it takes longer to reach the point where enough molecules can boil, so boiling is delayed.

4. What is the difference between sequence similarity, sequence identity and what is the significance of e-value in BLAST results?

1. Sequence Similarity: Sequence similarity refers to the percentage of aligned positions between two sequences where the residues are either the same or have similar chemical properties. **Example**: A substitution of leucine (L) for isoleucine (I) might be counted as a similar residue due to both being hydrophobic.

2. Sequence Identity: Sequence identity refers to the percentage of exact matches between two sequences in a pairwise alignment. This is a more stringent measure than similarity, as it counts only positions where the amino acids or nucleotides are exactly the same. **Example**: If two sequences have 50 out of 100 positions aligned with the same amino acid, their sequence identity would be 50%.

3. E-value (Expected Value) **Definition**: The **e-value** represents the number of times one would expect to see a particular alignment by chance in a database of a given size. A lower e-value indicates a more significant alignment (less likely to be due to random chance).**Low e-value (close to 0)**: Highly significant alignment, meaning the match is likely not random, **High e-value**: Indicates a less significant match, and it could be a result of random sequence similarity rather than true biological relevance.

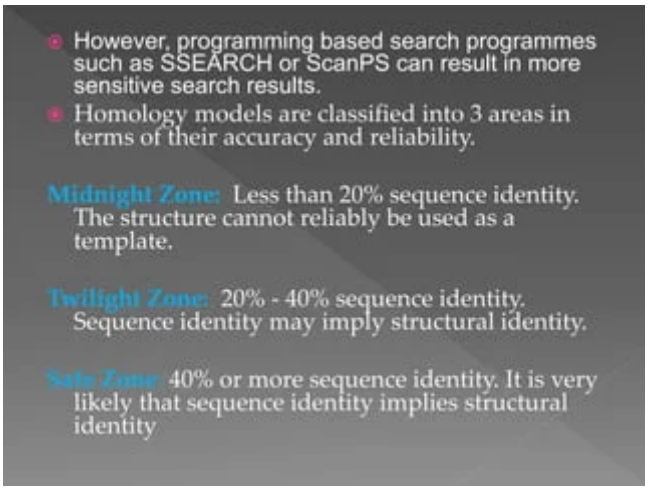
5. A DNA sample contains 20% adenine (A). According to Chargaff's rules, what percentage of thymine (T), guanine (G), and cytosine (C) would be present?

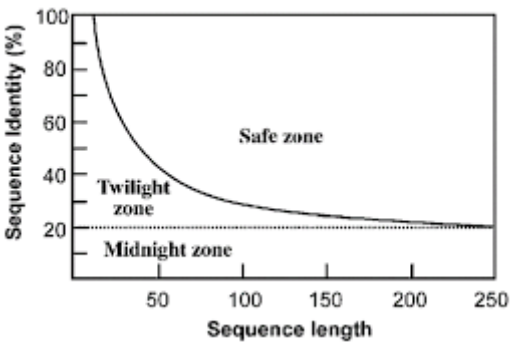
A=T=20%

G=C=30%

5 mark each - Answer any 2

1. **With the help of a well labelled plot explain the relevance of different zones of sequence identity in homology modelling? The R.M.S.D values between your template and the 3 models that you predicted are as follows : 0.5 A, 1.2A and 0.8 A. Which of the above models will you choose and why ? Mention one tool which will help you visualise the modelled structure.(2+2+1)**





We will choose structure with RMSD 0.5A, as it shows least deviation from our template structure.

Chimera, pymol

2. **Given a nucleotide sequence 5'-AAGTCGGA-3' , write its complementary sequence, mention sense and antisense strand. Mention the main enzyme involved in replication. What is the drawback of the structure suggested by Pauling?(1+1+1+2)**

   5’----->3’ sense strand (5'-AAGTCGGA-3’)

   3’----->5’ antisense strand(TTCAGCCT)

   Enzyme - DNA polymerase. Pauling suggested outward facing bases.

   Pauling placed the **nucleotide bases** (adenine, thymine, cytosine, guanine) on the **outside** of the structure.

   Problem: DNA’s nucleotide bases are hydrophobic, they prefer to avoid water, which is abundant in the cellular environment. Placing the bases on the outside, exposed to water, is energetically unfavorable. In contrast, Watson and Crick’s model, which placed the bases inside the helix, allows the hydrophobic bases to interact with each other, forming hydrogen bonds.

3. **5’-AAGCATGACTCTGTTTCGGTAGGCGTAATAG-3’. What will be the RNA and protein sequence generated from this DNA strand? Mention the enzyme involved in transcription and the site of protein synthesis.(1.5+1.5+2)**

RNA : 5’ AAGC AUG ACU CUG UUU CGG UAGGCGUAAUAG-3’

Protein : Met-–Thr—Leu—Phe—Arg— (STOP)

RNA polymerase

Ribosomes

FIRST LETTER	SECOND LETTER U	SECOND LETTER C	SECOND LETTER A	SECOND LETTER G	THIRD LETTER
U	UUU Phenylalanine (F) UUC	UCU UCC Serine (S) UCA UCG	UAU Tyrosine (Y) UAC	UGU Cysteine (C) UGC	U C A G
	UUA Leucine (L) UUG		**UAA** stop codon **UAG** stop codon	**UGA** stop codon UGG Tryptophan (W)	
C	CUU CUC Leucine (L) CUA CUG	CCU CCC Proline (P) CCA CCG	CAU Histidine (H) CAC	CGU CGC Arginine (R) CGA CGG	U C A G
			CAA Glutamine CAG (Q)		
A	AUU AUC Isoleucine (I)	ACU ACC Threonine ACA (T) ACG	AAU Asparagine AAC (N)	AGU Serine (S) AGC	U C A G
	AUA		AAA Lysine (K) AAG	AGA Arginine (R) AGG	
	AUG start codon* (M)				
G	GUU GUC Valine (V) GUA GUG	GCU GCC Alanine (A) GCA GCG	GAU Aspartic acid GAC (D)	GGU GGC Glycine (G) GGA GGG	U C A G
			GAA Glutamic acid GAG (E)		

* The start codon encodes the amino acid methionine