Mid-Semester Exam

Computer-Aided Drug Design (CADD) BIO563

Section A

Total Marks: 30 Total Time: 1 Hour

MCQ (Choose the most appropriate answer for each question.)

(1*15=15 marks)

1. Addition and removal of Beta-mercaptoethanol results in and respectively. A.
   Unfolding, Folding
   B. Folding, Unfolding
   C. Denaturation, Renaturation
   D. Renaturation, Denaturation

2. What makes proline unique among standard amino acids?
   A. It lacks a carboxyl group
   B. It lacks an amino group
   C. **Its side chain forms a cyclic structure with the backbone**
   D. It is the largest amino acid

3. Which of the following factors can cause deviations from expected regions in the Ramachandran plot?
   A. Mutations altering residue flexibility
   B. Incorrect protein structure refinement
   C. Extreme pH conditions affecting peptide bonds
   D. Sequence alignment errors
   E. A & B & C

F. B & C & D

4. Why do most peptide bonds in proteins adopt the trans conformation?

A. To minimize steric clashes between adjacent side chains

B. To stabilize hydrogen bonding in secondary structures

C. Because the cis conformation is more energetically favorable

D. Due to resonance stabilization

E. A & D

F. A & B

5. Why do mutations in the primary sequence of a protein sometimes affect its function?

A. They may alter key amino acids in the active site.

B. They can disrupt folding by introducing steric clashes.

C. They always prevent protein synthesis.

D. They may affect interactions with ligands or other proteins.

E. A & B

6. How does the van der Waals radius compare to the covalent radius of an atom?

A. It is generally smaller than the covalent radius

B. It is generally larger than the covalent radius

C. They are always equal

D. It depends only on the atomic number

7. In CATH classification, which category is associated with the type and number of secondary structure (SS) elements only?

A. Class (C)

B. Architecture (A)

C. Topology (T)

D. Homologous Superfamily (H)

8. Which of the following will be the best template for modeling a protein structure?

A. Template 1: 93% identity, 3.5 Å resolution

B. Template 2: 90% identity, 1.5 Å resolution

C. Template 3: 91% identity and 1.2 Å resolution

D. Template 4: 95% identity and 4.0 Å resolution

9. In a genetic disorder, a single mutation causes a normally soluble protein to aggregate. Which of the following best explains this phenomenon?

A. The mutation disrupts native interactions, exposing hydrophobic residues that drive aggregation.

B. The mutation enhances van der Waals interactions, making the protein more soluble. C. The mutation increases the protein's net charge, preventing aggregation. D. The mutation strengthens hydrogen bonding, leading to a more stable folded structure.

10. Which of the following are stereochemical constraints?

A. Bond lengths

B. Dihedral angles

C. Both A and B

D. None of the above

11. What type of bond contributes to the protein's ability to bind selectively to the ligand?

A. Covalent Bond

B. Peptide Bond

C. Hydrogen Bond

D. None of the above

12. Which amino acid is the most flexible due to the lack of a side chain in its structure?: A. Proline

B. Arginine

C. Glycine

D. Methionine

13. Which computational method is optimal for the study of protein that works on the principle of equation of motion?

A. Homology Modeling

B. Molecular Docking

C. MD Simulations

D. Multiple sequence Alignment

14. Which of the following servers is **NOT** used for homology modeling?

A. Modeller

B. Swiss-Model

C. Robetta

D. BLAST

15. Which of the following is the correct interpretation of Root Mean Square Deviation (RMSD) in homology modeling?

A. High RMSD indicates a more accurate model compared to the template **B. Low RMSD suggests a model that closely resembles the template structure** C. RMSD does not affect the quality of the homology model

D. A low RMSD always guarantees a biologically accurate model

Section B

Short Answer Questions

(10 marks)

1. With the help of a well labelled diagram highlight the local minima, global minima and transition state of the protein folding. **(2 mark)**

Refer to slides 57-61 of lecture 7

2. What are the Twilight Zone and Midnight Zone? What is their significance in modeling a protein structure? **(2 mark)**

- The Twilight Zone refers to sequence identity **between 20-30%**.In this range, sequence alignment becomes unreliable, and errors in model prediction increase.Structural similarity may still exist, but distinguishing true homologs from unrelated proteins becomes difficult.

- The Midnight Zone refers to sequence identity **below 20%**.In this range, it is nearly impossible to reliably model a protein structure based on homology due to a lack of detectable evolutionary relationships.Even if two proteins have similar structures, their sequences may have diverged too much for homology-based modeling to be effective.

3. If the sample φ (phi) and ψ (psi) angles are taken in 20-degree units, and the protein chain has 10 residues, how many possible states will the chain have? Explain the paradox associated with this problem. **(3 mark)**

(2 mark for calculation)

- If the φ (phi) and ψ (psi) torsion angles are taken a $20°$ increments, then each torsion angle can adopt 18 different states **(since $360°/20°$ = 18)**
- For a single residue, since it has both φ and ψ angles, the total number of conformational states per residue is: **$18*18 = 18^2$**

- For a 10-residue protein chain, the total number of possible conformations is: $(18^2)^{10} = (18)^{20}$

(Levinthal's Paradox) (1 mark)

This calculation highlights Levinthal's paradox, which states that if a protein were to randomly sample all possible conformations to reach its native folded state, it would take an astronomical amount of time—far longer than the age of the universe. However, in reality, proteins fold within milliseconds to seconds. This suggests that protein folding is not a random search but follows a directed pathway governed by thermodynamic and kinetic principles.

4. What is the difference between sequence identity and sequence similarity? Among the following given sequences, which pair will have the best similarity, which will have the best identity, and what will be the percentage of identity? **(3 mark)**

- Sequence 1: AGAACTGAT
- Sequence 2: ACTGATG
- Sequence 3: GAACTGAAT

Answer:

Sequence Identity: The percentage of exact matches between two sequences at the same positions when aligned. It considers only identical residues (nucleotides or amino acids).**(0.5 mark)**

Sequence Similarity: the degree of resemblance between two sequences when they are compared. A broader measure that includes both identical and chemically similar residues (for proteins). It is often used in protein sequences, where substitutions of similar amino acids (e.g., leucine to isoleucine) are considered similar. **(0.5 mark)**

Sequence Identity calculation (1 mark) (even if only 1 vs 3 is done then give 1 mark)

Sequence Identity = (no of identical nucleotides / min(length(A),length(B))).

Sequence pair	1 vs 2	1 vs 3	2 vs 3
Sequence align	AGAACTGAT - - - - ACTGATG	AGAACTGA - T - GAACTGAAT	- - ACTGA -TG GAACTGAAT -
Matching positions	6	8	6
Min length	7	9	7
Sequence Identity	6/7 *100=**85.71**	8/9 *100 = **88.89**	6/7 *100 =**85.71**

Best similarity and identity (1 mark)

Best Identity: Sequence 1 & Sequence 3

Best Similarity: Sequence 1 & Sequence 3

Section C

Long Answer Questions (Attempt any one)

(1*5=5 marks)

1. A research team is working to develop a new drug for a rare genetic disorder caused by a malfunctioning enzyme .

(a) However, unavailability of enzyme's experimental structure poses a big challenge. As a computational biologist, what approach would you follow? **(1 mark)**

As a computational biologist, I would use homology modeling or AI-based protein structure prediction (e.g., AlphaFold or RoseTTAFold) to predict the enzyme's 3D structure. If homologous structures are available, comparative modeling can be used to build a reliable model.

(b) During preclinical testing, researchers find that a promising inhibitor has low bioavailability. What properties should be optimized? **(2 marks)**

Solubility – Improve aqueous solubility through prodrug design, salt formation, or nanocarriers.

Lipophilicity – Adjust logP values to ensure good membrane permeability without excessive hydrophobicity.

Metabolic Stability – Modify chemical structure to reduce rapid metabolism by liver enzymes

Permeability – Improve intestinal absorption by optimizing molecular weight and hydrogen bond donors/acceptors (Lipinski's Rule of Five).

Efflux Transporter Avoidance – Reduce recognition by efflux pumps

(c)What ethical considerations should be taken into account when testing a drug for a population? **(2 mark)**

c) Ethical Considerations in Drug Testing for a Population (2 marks)

1. Informed Consent – Ensure participants fully understand the study's risks, benefits, and purpose before enrolling.
2. Equitable Selection of Participants – Avoid exploiting vulnerable populations and ensure fair representation of affected groups.
3. Minimization of Harm – Conduct rigorous preclinical testing to minimize potential risks to participants.
4. Transparency and Data Integrity – Ensure accurate reporting of results and avoid bias or data manipulation.
5. Post-Trial Access – Ensure that participants, especially those in low-income settings, have access to the drug if it proves effective.

OR

2. A pharmaceutical company is developing a new drug for a rare genetic disorder. Before the drug reaches clinical trials, it must go through several stages in the drug discovery pipeline.
   (a) name the key stages involved in the drug discovery pipeline before clinical trials. **(1 marks)**
   (b) Discuss two common myths associated with drug discovery and explain why they are incorrect. **(2 mark)**
   (c) Briefly describe the four phases of clinical trials and their purpose. **(2 mark)**

Target Identification – Identifying a biological target (e.g., protein, gene) linked to the disease. Target Validation – Confirming that modulating the target has therapeutic potential. Hit Identification – Screening compounds to find potential drug candidates (hits). Hit-to-Lead Optimization – Refining hits to improve efficacy, selectivity, and safety. Lead Optimization – Further modifying leads to enhance pharmacokinetics and reduce toxicity. Preclinical Testing – Conducting in vitro (cell-based) and in vivo (animal) studies to assess safety and efficacy.
(b) Two Common Myths in Drug Discovery and Why They Are Incorrect (2 marks)
Myth: Drug discovery is a fast process.
Reality: Drug discovery is time-consuming, often taking 10–15 years before a drug reaches the market. Each stage, from target identification to preclinical and clinical trials, involves extensive testing, regulatory approvals, and optimization.

Myth: A single promising molecule will always succeed as a drug.
Reality: Most drug candidates fail due to toxicity, poor bioavailability, or lack of efficacy. Only about 1 in 10,000 compounds that enter early research ultimately become an approved drug.

(c) Four Phases of Clinical Trials and Their Purpose (2 marks)
Phase I – Tests the drug's safety, dosage, and side effects in a small group of healthy volunteers (or patients for severe diseases).
Phase II – Evaluates the efficacy and further safety in a larger group of patients with the target disease.
Phase III – Conducts large-scale randomized trials to confirm efficacy, monitor adverse reactions, and compare the drug to existing treatments.
Phase IV – Post-market surveillance after approval, monitoring long-term effects and rare side effects in the general population.