End sem 2025
Biophysics(BIOP)

Total Marks:40

Time: 1 Hour

Section A

Multiple Choice Questions(1 mark each)

1. Which of the following fundamental discoveries or concepts was NOT one of the three critical pieces of information used by Watson and Crick in 1953 to deduce the double helix structure of DNA?

a. Chargaff's rules (A=T, G=C).
b. X-ray diffraction data from Rosalind Franklin (helical nature and dimensions).
c. The knowledge of the chemical structure of a nucleotide (sugar, base, phosphate).
d. The Central Dogma of Molecular Biology (DNA to RNA to Protein).

2. Which factor increases the entropy component of free energy in biological macromolecule?

a. Formation of structured solvent cages around exposed hydrophobic side chains
b. Restriction of backbone dihedral angles due to cooperative secondary-structure formation
c. Desolvation of nonpolar surfaces leading to increased translational freedom of bulk water
d. Formation of long-range Coulombic interactions that constrain residue orientations

3. Why does placing a salt bridge deep inside a protein core dramatically increase its stabilizing contribution compared to the same salt bridge on the surface?

a.The dielectric constant is lower, strengthening charge–charge interactions
b. Entropy of ions in the core is higher
c.Water molecules penetrate the core to stabilize charges
d. Hydrogen bonds replace electrostatic interactions

4. Secondary structure prediction accuracy is often limited to 64–75% because:

a. Prediction algorithms cannot recognize hydrophobic residues
b. Local sequence alone does not fully determine secondary structure
c. All helices require long-range interactions
d. β-sheets never depend on sequence context

5. Which situation reflects a case where **enthalpic** forces dominate over entropic contributions?

a. Formation of hydrogen-bond networks in α-helices
b. Hydrophobic collapse in folding
c. Release of solvent after binding
d. Disorder–order transitions in intrinsically disordered proteins

6. Temperature in thermodynamics can be defined as:

a. Rate of Energy changes with entropy at constant volume (V) and number of molecules (N)
b. Rate of volume changes with energy at constant number of molecules (N) and entropy (S)
c. Rate of Entropy changes with volume at constant number of molecules (N) and Energy (E)
d. Rate of change energy with number of molecules at constant volume (V) and Entropy (S)

7. Which of the following expressions does **not** correctly describe the relationship between pressure, particle density, and height in a gravitational field?

a. $dp/dh = dn/dh * kT$
b. $dn/dh * kT = -mgV$
c. $dn/dh = -mg/kT * n$
d. $E(h) = mgh$

8. What is the fundamental assumption behind homology modelling?

a) Proteins with similar length have identical folds
b) Proteins with similar sequences adopt similar 3D structures
c) All proteins fold into α-helices
d) Hydrophobic residues are always surface-exposed

9. What is the primary difference between a .gro file and a .pdb file in GROMACS?
a. .gro contains coordinates only, .pdb contains topology
b. .gro stores coordinates with box vectors, .pdb follows a standardized structural format
c. Both are identical
d. .pdb stores velocities, .gro does not

10. Which GROMACS command is used to create custom index groups?
a. gmx trjcat
b. gmx genrestr
c. gmx make_ndx
d. gmx rms
11. During protein folding, molecular **chaperones** are used primarily to:
a. Provide energy to form covalent bonds in the protein

b. Increase the rate of translation of the protein synthesis
c. Prevent misfolding and aggregation of nascent polypeptides
d. Remove the disulfide bond for proper folding

12. Which command is used to process and convert trajectory files, such as removing PBC or extracting frames?

a. gmx trjconv
b. gmx energy
c. gmx solvate
d. gmx grompp

13. A single amino acid substitution that disrupts disulfide bonds affects which structural level the most?

a. Primary
b. Secondary
c. Tertiary
d. Quaternary

14. In the alignment files generated by MODELLER, which format is primarily meant for visual inspection?

a. PIR
b. PAP
c. FASTA
d. BLAST

15. For searching a query sequence with a database, which of the following statement is correct ?

a) Nucleotide query against a nucleotide sequence database is done by blastp
b) Protein query against a translated nucleotide sequence database is done by blastp
c) Translated nucleotide query against a protein database is done by blastx
d) Protein query against a protein database is done by tblast

Section B (5*2)marks

1. A loop in a protein shows α-helical conformation in a homolog's X-ray crystal structure. MD simulation of the same protein segment in solution shows 50% α-helix, 40% β-strand, and 10%

coil. Explain the discrepancy between X-ray and MD data. How could solvent and thermal fluctuations explain this observation?

Ans: **Discrepancy reason:** X-ray structure represents a static, crystal environment (low temperature, packing forces). MD simulates protein in **solution at 300 K**, capturing thermal fluctuations and solvent effects.

Implication: The loop is flexible; both α-helix and β-sheet are accessible. MD reveals dynamic behavior not captured in static X-ray.

2. Briefly explain safe zone and twilight zone in sequence modeling?

Ans: **Safe Zone (≥ 40% sequence identity)** When the target–template sequence identity is **40% or higher**, homology modeling is considered **reliable and safe**. Structural similarity is almost guaranteed because proteins above this threshold generally share the same fold. Alignment accuracy is high → fewer modeling errors.

Twilight Zone (20–35% sequence identity) When sequence identity drops into the **20–35% range**, structural similarity is **uncertain**.This is called the **twilight zone** because: Alignment becomes error-prone, Structural homology is difficult to confirm and Template selection requires additional evidence (motifs, secondary structure prediction, profile–profile alignments)

3. For a protein that is 230 amino acids long and contains 35 cysteines and 14 methionines, what types of covalent bonds can form between side chains of cysteines, and how many such covalent bonds are possible? How many peptide bonds can form in this protein??

Ans: disulfide bonds (S–S) between **side chains of** cysteines, The **maximum** number of disulfide bonds in one molecule = (number_of_cys / 2. (1.5)

#peptide bonds = length − 1 (0.5)

4. **a.) What is the expected behavior of the temperature profile during NVT equilibration in MD simulations, and why is this important?(1)**
b.) what if temperature fluctuates continuously, suggest two correct actions one should consider.(1)

a.The temperature must rise to 300k then form a stable plateau, which indicates that system has reached equilibrium(0.5). it is important because it is essential for reliable and reproducible results of simulations(0.5)

b. Two correct actions are:
Increase the equilibrium time in nvt.mdp file (0.5)
Check for bad atomics contacts or unstable ions.(0.5)

5. Briefly mention the steps involved in Homology modelling. Explicitly mention in what scenario we go for advanced modelling.

1. Template identification

2. Sequence Alignment

3. Model building

4. Model Refinement

5. Model Validation

Advanced techniques (like multiple-template modelling, loop modelling, ab initio modelling for missing regions, molecular dynamics refinement) are used when:

a. Sequence identity with available templates is low (<30%)

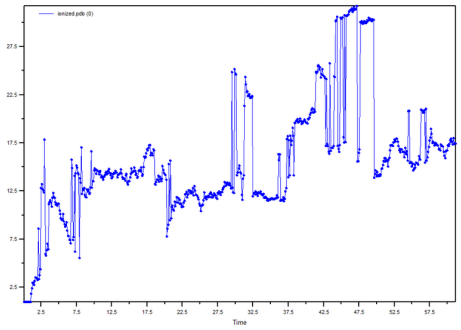
b.Twilight zone → simple homology modelling becomes unreliable

c Template has missing regions or poor resolution

d. Target contains long loops, insertions, or unique domains not present in templates

Section C (3*5)marks

1. After performing a Molecular Dynamics (MD) simulation, you generated the RMSD plot for your protein backbone relative to the minimized starting structure. The plot displays a clear initial rise ("jump") followed by a stable region. When the RMSD is recalculated using a structure from the stable phase as the reference, another large, unphysical "jump" appears later in the trajectory (as shown in the figure).



   a. What is the purpose of RMSD, and how do you interpret the large "jump" observed in the provided plot? **(2 marks)**

Ans: It evaluates the structural stability during the simulation of the protein structure.(1)
Interpretation: A large, sudden jump typically indicates an *artifact* rather than a real conformational change. The most common cause is a **Periodic Boundary Condition (PBC) jump**, where the protein crosses the box boundary and is wrapped to the opposite side, artificially increasing RMSD. This produces a discontinuous spike in the trajectory. (1)

   b. How to remove these jumps in the plot? (2)

Ans: To remove jumps from the trajectory, we must remove PBC, and the protein must be recentered and made whole before calculating the RMSD. These steps reconstruct continuous

molecular motion and eliminate artificial coordinate discontinuities, thereby eliminating the false RMSD jumps.

c. Mention two other analyses that can be performed to analyse the trajectories. (other than rmsd)1

Ans: **Radius of Gyration (Rg):** Measures protein compactness and folding state.
Hydrogen Bond Analysis: Tracks stability of intramolecular or intermolecular H-bonds.
Root Mean Square Fluctuation (RMSF) of ligand or domain (if RMSF for backbone excluded).
Secondary Structure Analysis (DSSP): Monitors helix/β-sheet evolution. (Any two receive full credit.)

2. A computational biologist generated a homology model for a target protein and used the DOPE (Discrete Optimized Protein Energy) Score Plot to assess its quality.
   a. What is the principle behind homology modeling? (2)

Ans: Homology modeling is based on the principle that **proteins with similar amino-acid sequences generally adopt similar 3D structures**, allowing a target protein's structure to be predicted from a homologous template with known structure.

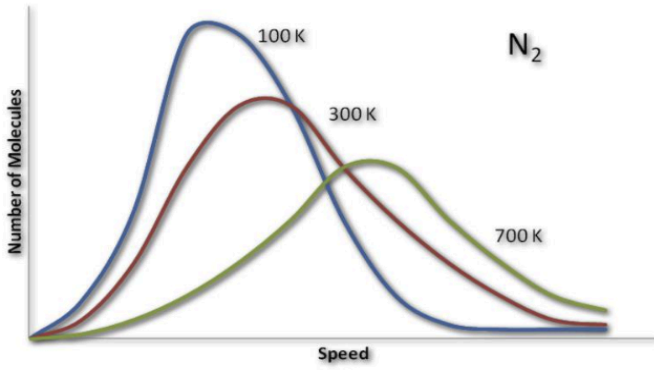
   b. What are the criteria to choose a template? (2)

Ans: High sequence identity, high coverage, low e-value

   c. Mention any two web servers available for homology modeling (1)

Ans: itasser, swissmodel, alphafold

3. **This image shows the Maxwell–Boltzmann distribution of molecular speeds for nitrogen gas (N_2) at three different temperatures:**



Answer the following based on the plot given:

a. Which temperature curve shows the highest average molecular speed? Why?(1.5)

Ans: The 700 K curve shows the highest average molecular speed because at higher temperatures, the molecules have more kinetic energy, causing the distribution to shift toward higher speeds.

b. Why does the curve become flatter as temperature increases? (1)

Ans: At higher temperatures, the speeds of molecules vary widely, so the distribution spreads out. This "flattens" the curve because the number of molecules at any specific speed decreases.

c. Which curve has the sharpest peak, and what does that tell us about the molecular speeds? (1.5)

Ans: The 100 K curve has the sharpest peak. This indicates that at low temperatures, most molecules have similar, low speeds, resulting in a narrow distribution of speeds.

d. Which temperature curve has the greatest number of very fast-moving molecules? (1)

Ans: The 700 K curve shows the longest tail at high speeds, indicating that a larger fraction of molecules move very fast.