BIOSC 1540 - Computational Biology

Computational Structural Biology Exam

Nov 14, 2024

100 points

Please read the following instructions carefully before beginning your assessment.

- **Time limit:** You have 75 minutes to complete and turn in this assessment.
- **Open note:** You may use notes, but with the following restrictions:
  - Notes must be hand-written on either (1) paper or (2) a tablet with a stylus, then printed.
  - You may use a maximum of one sheet of 8.5 $\times$ 11 in. paper for notes (front and back allowed).
  - Notes must be your own work. Sharing or copying notes from others is strictly prohibited.
  - Your name must be clearly written on your notes.
- **No digital devices:** The use of digital devices, including calculators, is not allowed.
- **Submission requirements:** You must submit both your completed assessment and all notes.

I agree to follow the above instructions. I affirm that all work on this assessment will be my own and that I will not give or receive any unauthorized assistance. To have your assessment graded, you must write your name, sign, and provide your student ID below.

_____ _____

Name Signature

_____

Student ID

Recommendation: There are 9 one-point, 11 two-point, 14 three-point, 3 four-point, and 3 five-point problems. More challenging questions are worth fewer points, so answer the easier problems first. Choose the best answer for each problem unless specified otherwise.

Key:

Full credit

Partial credit

Protein structure determination

Problem 1

Select the correct statement regarding the use of protein crystals in X-ray diffraction.
(3 points)

(A) Proteins cannot scatter X-rays unless crystallized.

(B) Crystallization ensures proteins adopt their native conformations.

(C) Crystals eliminate background noise in the diffraction pattern.

(D) Crystallization creates a lattice that produces measurable diffraction patterns.

Problem 2

In protein structure determination, electron density maps:
(3 points)

(A) Directly display the exact positions of all atoms in the protein.

(B) Are derived from processed diffraction data.

(C) Are used exclusively for small proteins.

(D) Represent theoretical electron distributions predicted from atomic models.

Problem 3

In X-ray crystallography, when will constructive interference occur? (Select all that apply)
(1 point)

(A) X-ray waves diffracted from parallel crystal planes meet Bragg's Law conditions.

(B) The path difference between scattered waves equals half a wavelength.

(C) The crystal lattice spacing is larger than the incident X-ray wavelength.

(D) The amplitude of diffracted waves cancels out.

Problem 4

The process of building a protein structure from electron density requires:
(2 points)

(A) Iterative refinement of atomic coordinates against experimental data and geometric restraints.

(B) Direct mapping of amino acid side chains based on characteristic electron density shapes.

(C) Sequential tracing of the backbone followed by automated side chain placement.

(D) Real-time modification of atomic positions guided by difference density maps.

Problem 5

Select all advantages that Cryo-EM has over X-ray crystallography:
(4 points)

- (A) Enables visualization of proteins in their native environment.
- (B) Enables capture of multiple conformational states.
- (C) Allows study of larger macromolecular complexes.
- (D) Does not require solidifying the sample.
- (E) Can always achieve superior resolution.
- (F) Captures membrane proteins in lipid environments.

Problem 6

At which structural level are hydrogen bonds between backbone atoms primarily responsible for stabilizing regular conformations?
(5 points)

- (A) Primary structure.
- (B) Secondary structure.
- (C) Tertiary structure.
- (D) Both primary and tertiary structure.

Problem 7

In Single Particle Analysis (SPA) for Cryo-EM, three-dimensional reconstruction requires:
(1 points) Question was thrown out.

- (A) Determination of particle orientations through projection matching and angular assignment.
- (B) Averaging of all particle images regardless of their conformational states.
- (C) Sequential merging of 2D class averages based on sample tilting angles.
- (D) Direct conversion of 2D micrographs into 3D volumes using Fourier transforms.

Problem 8

Explain the technical challenges of studying intrinsically disordered proteins (IDPs) versus well-ordered proteins using experimental techniques.

(5 points)

IDPs lack a stable 3D structure under physiological conditions, unlike well-ordered proteins that fold into specific shapes essential for their function. This inherent flexibility means that IDPs exist as dynamic ensembles of conformations rather than a fixed structure.

X-ray crystallography requires the formation of well-ordered crystals, which is nearly impossible with IDPs due to their structural heterogeneity. Similarly, nuclear magnetic resonance (NMR) spectroscopy faces difficulties because the multitude of overlapping signals from rapidly interconverting conformations complicates data interpretation. Other techniques like cryo-electron microscopy (cryo-EM) and circular dichroism (CD) spectroscopy also struggle with IDPs. Cryo-EM relies on averaging multiple images to resolve structures, but the conformational variability of IDPs leads to blurred results.

Protein structure prediction

Problem 9

Which energy landscape feature presents the main difficulty for predicting protein structures?

(1 point)

- (A) The large number of possible conformations that increases exponentially with protein length.
- (B) The flat energy landscape lacking significant energy barriers between conformations.
- (C) The complex, rugged energy surface with numerous low-energy structures.
- (D) The influence of temperature on the stability of different conformational states.

Problem 10

When is homology modeling expected to give the most accurate structural predictions?

(4 points)

- (A) When the template and target share over 90% sequence identity across conserved regions.
- (B) When the template and target share less than 30% sequence identity but have similar functions.
- (C) When the template and target share >65% sequence identity across the full protein length.
- (D) When the template and target share 50% sequence identity with significant gaps and insertions.

Problem 11

When would protein threading be the most appropriate approach for structure prediction?
(1 point)

- (A) When the target has 15-25% sequence identity with known structures but predicted secondary structure elements match existing folds.
- (B) When the target sequence shows strong conservation of hydrophobic packing patterns despite low overall sequence identity.
- (C) When multiple sequence alignments reveal conserved structural motifs within a protein family.
- (D) When remote homologs exist but their evolutionary relationship cannot be detected by sequence comparison alone.

Problem 12

How does modern coevolutionary analysis identify meaningful residue-residue contacts in protein structures?
(3 points)

- (A) By detecting conserved residues that are identical across different species.
- (B) By separating direct evolutionary couplings from indirect correlations using statistical methods.
- (C) Through random sampling of residue pairs in protein sequences.
- (D) By predicting contacts based on amino acid likelihood for certain secondary structures.

Molecular simulations

Problem 13

How do MD simulations enhance our understanding of protein dynamics?
(3 points)

- (A) By providing static snapshots of proteins in their lowest energy states.
- (B) By sampling the global energy minimum conformation of proteins.
- (C) By using quantum mechanical methods to simulate bond-breaking and electron transfer events within proteins.
- (D) By generating time-resolved trajectories of each atom that capture both small-scale and large-scale protein motions.

Problem 14

How are protein force field parameters determined and optimized?
(2 points)

- (A) By iteratively adjusting them to match quantum mechanical calculations of small molecules and from experimental thermodynamic data.
- (B) By training machine learning models on experimental protein structures and spectroscopy.
- (C) By fitting them to high-level theoretical calculations and experimental vibrational spectra data.
- (D) By tuning parameters to align with protein folding and unfolding free energy measurements.

Problem 15

In force fields, chemical bonds are typically modeled as:
(3 points)

- (A) Springs that can stretch and break, accounting for the energy needed to break bonds.
- (B) Simple springs that can stretch and compress around their natural length.
- (C) Connected springs that affect both bond lengths and angles together.
- (D) Classical approximations based on quantum mechanical calculations.

Problem 16

Why is selecting an appropriate time step crucial in MD simulations?
(2 points)

- (A) The time step must be smaller than the shortest vibrational period to accurately capture atomic motions.
- (B) Larger time steps allow for faster simulations by skipping intermediate calculations without affecting accuracy.
- (C) The time step determines how efficiently the simulation explores the potential energy surface of the molecular system.
- (D) The time step does not have any impact on the physical accuracy of the simulation results.

Problem 17

How are dihedral angle potentials modeled in force fields?

(2 points)

(A) Using complex corrections based on quantum calculations between adjacent angles.

(B) With periodic energy functions that include patterns matching rotational symmetry.

(C) Through multiple cosine functions with optimized strengths and angles derived from detailed energy profiles.

(D) By employing flexible spline functions that connect quantum reference points while maintaining rotational consistency.

Molecular system representations

Problem 18

A researcher is simulating how a protein and a molecule bind together and notices that the molecule quickly leaves the binding site during the simulation. What change to the simulation settings could help keep the molecule in the binding area?

(1 point) Question was thrown out.

(A) Slowly lower the solvent's ability to reduce electric charges to make the attraction between the molecule and the binding site stronger.

(B) Use adaptive time steps that become smaller where there are strong forces around the binding site.

(C) Apply restrictions on the molecule's position or use advanced sampling methods to keep it in the binding site.

(D) Increase the range for non-bonded interactions and update the nearby particles list more often.

Problem 19

In simulations with periodic boundary conditions (PBC):

(3 points)

(A) The system uses elastic collisions at the edges to keep momentum but stops particles from escaping.

(B) Long-range forces are cut off at the box edges to save computing power and avoid errors from particles interacting with themselves.

(C) Atoms that leave one side of the simulation box re-enter from the opposite side, keeping interactions continuous.

(D) Each particle interacts only with the nearest images of other particles within a set distance.

Problem 20

The minimum elements needed to simulate a protein inside a cell include:
(4 points)

- (A) Protein structure, explicit water molecules, and ions to balance the system's overall charge.
- (B) Protein structure using a model that doesn't show individual water molecules but includes normal salt levels.
- (C) Protein structure, explicit water molecules, ions at physiological concentration, and any required cofactors.
- (D) Protein structure with specific water molecules closely surrounding it and a variable that adjusts for the overall solvent effects.

Problem 21

Energy minimization is needed before running an MD simulation to:
(2 points)

- (A) Optimize the system's energy landscape to make sampling different shapes more efficient during the main simulation.
- (B) Relax local strains in bonds and angles while keeping the overall protein shape to start from a stable configuration.
- (C) Remove bad overlaps, unfavorable charge interactions, and high-energy shapes that could cause calculation problems.
- (D) Balance the water distribution around the protein while slowly removing restraints to keep the system stable.

Problem 22

When preparing a protein structure, choosing the right protonation states is important because:
(3 points)

- (A) They determine the strength of hydrogen bonds inside the protein and affect its local stability.
- (B) They control electric interactions, affect the acidity levels of nearby parts, and can change how the protein binds to other molecules.
- (C) They change the local electric environment around certain parts and influence how protons move in active sites.
- (D) They manage the formation of salt bridges where the protein meets the solvent and affect the overall charge in the simulation.

Problem 23

A complete protein structure determined by X-ray crystallography with a resolution of 3.0 Å is:
(3 points)

- (A) Suitable for MD simulations after refining side chain positions and backbone geometry.
- (B) Limited in reliability because atomic positions are uncertain and loops might be missing.
- (C) Good for MD simulations but needs additional modeling and validation.
- (D) Directly usable for MD if combined with other data like Cryo-EM or NMR.

Ensembles and atomistic insight

Problem 24

Observable properties in molecular systems are:
(3 points)

- (A) Average values calculated from the system's wave function over time for specific energy levels.
- (B) Averages taken over all possible states, weighted by their likelihood at a stable temperature.
- (C) Long-term averages of how variables change around their stable values.
- (D) States sampled based on the assumption that all accessible states are equally likely.

Problem 25

The Nosé-Hoover thermostat differs from the Berendsen thermostat because it:
(2 points)

- (A) Uses an extended system with a heat bath variable that ensures accurate temperature distribution through predictable movements.
- (B) Applies a simple relaxation method that slowly reaches the desired temperature by slightly interacting with an external heat source.
- (C) Maintains time-reversible motions by adjusting velocities, keeping both energy and momentum balanced in the extended system.
- (D) Creates regular temperature changes through a feedback system that ensures equal energy distribution among all movement types.

Problem 26

Multiple shorter simulations are often preferred over one long simulation because:
(3 points)

(A) They allow better exploration of different states through methods like parallel tempering and replica exchange while keeping the system's behavior consistent.

(B) They improve sampling of different shapes and structures by starting from various initial setups and running independent paths, which helps verify statistical results.

(C) They reduce the build-up of errors in calculations by periodically resetting velocities based on the Maxwell-Boltzmann distribution.

(D) They enable adaptive strategies that focus computing power on less-explored areas of the energy landscape.

Problem 27

The relationship between energy and probability in statistical mechanics means that:
(1 point)

(A) Higher energy states are less common because they require more energy to reach.

(B) States with higher energy are less likely to be occupied, following an exponential decrease.

(C) The system balances the tendency to spread out with the available energy.

(D) The system tries to occupy as many different states as possible while keeping its total energy fixed.

Problem 28

Barostats maintain constant pressure by:
(2 points)

(A) Using an extended system with a virtual piston that interacts with volume changes through a mass parameter.

(B) Dynamically adjusting the simulation box size to control internal forces while keeping external pressure steady.

(C) Changing the system's motion equations to include pressure-related terms that influence group particle movements.

(D) Making periodic changes to the volume based on current pressure differences from the target using a feedback system.

Structure-based drug design

Problem 29

Which interaction typically provides the strongest contribution to binding enthalpy?
(3 points)

- (A) Short-range quantum mechanical forces and dispersion interactions that maximize contact area between molecules.
- **(B)** Electrostatic interactions between charged groups, influenced by the surrounding environment and charge screening.
- (C) Directional hydrogen bonds that form organized networks and are stabilized by polarization effects.
- (D) Complex multipolar interactions, including higher-order electric moments and induced polarization between molecules.

Problem 30

The entropic contribution to binding:
(3 points)

- (A) Results from changes in how molecules rotate and move, along with the rearrangement of water molecules at the binding interface.
- **(B)** Represents the change in the number of possible configurations the system can adopt when the complex forms, including the movement of solvent molecules.
- (C) Arises from the balance between decreased flexibility of the bound molecules and the favorable release of water molecules due to the hydrophobic effect.
- (D) Occurs through changes in the vibrational movements and quantum states of the bound complex compared to the free molecules.

Problem 31

In alchemical free energy simulations:
(2 points)

- **(A)** The system is changed in a series of small steps while measuring energy differences between each step.
- **(B)** A control parameter gradually turns interactions on or off in a smooth, reversible way.
- (C) Multiple copies of the system are simulated at different states to improve sampling.
- (D) Energy differences are calculated by comparing the initial and final states directly.

Problem 32

The Gibbs free energy of binding (ΔG_{bind}) represents:

(1 point)

(A) The potential of mean force integrated along the binding coordinate, incorporating solvent-mediated effective interactions between partners.

(B) The reversible work performed during the association process under constant temperature and pressure conditions, including reorganization effects.

(C) The difference in chemical potential between bound and unbound states that combines enthalpic interactions and entropic penalties at equilibrium.

(D) The ensemble-averaged energy change weighted by the ratio of partition functions for complexed versus dissociated states.

Docking

Problem 33

Docking simplifies binding free energy calculations by:

(5 points)

(A) Using empirical scoring methods that estimate energy contributions based on atomic interactions and simplified solvent models.

(B) Finding favorable binding positions by systematically rotating rigid molecules and ignoring complex entropy factors.

(C) Applying knowledge-based scoring from analyzing protein-ligand structures in databases.

(D) Breaking down interaction energies on a grid and assuming the protein is rigid.

Problem 34

Stochastic algorithms in pose optimization:

(2 points)

(A) Use adaptive energy barriers to guide the sampling of molecular shapes while ensuring proper transition probabilities.

(B) Group similar molecular shapes based on their structural differences to identify key representative structures.

(C) Explore the energy landscape by making random changes to molecular poses and accepting them based on probability criteria.

(D) Apply evolutionary techniques that select and improve binding poses using fitness scores from scoring functions.

Problem 35

Scoring functions in docking:

(2 points)

(A) Break down binding interactions into distance-based atomic potentials calibrated with experimental and quantum data.

(B) Estimate binding energies using weighted terms that combine physical forces, statistical data, and solvation effects.

(C) Use machine learning models trained on structural data to predict binding strengths based on geometric and chemical features.

(D) Assess how well proteins and ligands fit together using energy calculations enhanced by combining multiple scoring methods.

Problem 36

Grid-based pocket detection involves:

(1 point)

(A) Examining local surface shapes and angles to find inward-facing areas.

(B) Dividing the protein space into a grid of cubes and categorizing points based on their relation to protein atoms and solvent exposure.

(C) Using energy probes to map areas where binding is energetically favorable.

(D) Calculating shapes and spaces within the protein using computational topology.

Ligand-based drug design

Problem 37

Which molecular property is the best indicator of a molecule's ability to cross cell membranes?

(3 points)

(A) The ratio of polar surface area to volume combined with the number of rotatable bonds.

(B) The octanol-water partition coefficient (LogP) along with ionization states.

(C) Dynamic surface area measurements that consider different shapes and hydrogen bonding sites.

(D) Molecular shape based on mass distribution and electrostatic potential maps.

Problem 38

The Tanimoto similarity coefficient:

(2 points)

(A) Measures the normalized inner product of molecular fingerprints, accounting for bit density.

(B) Calculates the overlap of binary feature sets by dividing the number of shared features by the total unique features.

(C) Uses weighted feature matching based on the frequency of specific substructural patterns.

(D) Computes distance metrics in chemical space by comparing pharmacophoric elements.

Problem 39

The purpose of hashing in generating molecular fingerprints is to:

(1 point)

(A) Convert various molecular substructures into fixed-size feature vectors.

(B) Assign unique, fixed-length codes to different molecular substructures.

(C) Reduce the dimensionality of complex molecular data while preserving structural similarities.

(D) Create consistent numerical identifiers for molecular fragments using encoding methods.

Problem 40

In ECFP (Extended-Connectivity Fingerprints) generation, each subsequent iteration:

(3 points)

(A) Combines information from neighboring atoms while retaining stereochemical details.

(B) Updates atom identifiers by aggregating data from connected atoms using hashing functions.

(C) Applies information theory to retain important structural features in the molecular graph.

(D) Develops more complex connectivity patterns with atomic features within defined radius shells.