Final Exam Practice Questions

CS/BIOE/CME/BIOPHYS/BIOMEDIN 279

Fall 2020

Final Assessment Logistics

The final assessment will be released on Thursday November 19, 2020. It is an 80-minute timed assessment. You will be able to start and finish the assessment at any time on Nov. 19, provided you finish within 80 minutes of when you start. The final will be open-book and open-internet, but we stress the importance of answers being written in your own words.

Instructions

These are practice questions in the style of questions you might expect on the final assessment. Each question should be answerable in a few sentences (that is, you're not required to provide a great deal of detail).

Question 1: Compare and contrast the energy functions used for molecular dynamics simulations and those used for ab initio protein structure prediction.

Solution: Molecular dynamics simulations use a molecular mechanics force field. This is an approximate potential energy function — it specifies the energy of each precise arrangement of atoms, and therefore allows one to compute the force acting on each atom. The terms in most molecular mechanics force fields are physics-based. Ab initio protein structure prediction generally uses an (approximate) free energy function, which specifies the energy associated with a set of atomic arrangements, and thus allows one to pose protein structure prediction as a minimization problem. It is usually knowledge-based to some degree.

Question 2: How would you go about estimating how long it would take to run an MD simulation? What information would you need to consider?

Solution: You would need to consider

- Number of time steps (depends on total time being simulated.)
- Total number of atoms in the system (both protein and environment) being simulated.
- The average number of non-bonded interactions to be computed for each atom at each time step (non-bonded interactions will dominate the overall computation). This will depend on the algorithm being used to compute the non-bonded interactions.
- Time to compute each non-bonded interaction.

Question 3: We would like to estimate how tightly a particular drug candidate binds to a particular target protein.

Suppose we have a single molecular dynamics simulation in which the drug candidate binds to the target and stays bound for the remainder of the simulation. Can we accurately estimate the binding affinity from that simulation? Why or why not?

Solution: No, you would need to see the ligand binding and unbinding many times in order to accurately estimate the fraction of time it will remain bound. Note that simulating multiple binding/unbinding reactions may be infeasible to model through MD simulations (note the short time scales that we can simulate using MD).

Question 4: Describe one common approximation made by ligand docking methods, and explain why it helps simplify the problem to be solved.

Solution: A common approximation is that the protein is rigid. This simplifies the problem because one doesn't need to consider all possible arrangements of the protein atoms (and weight the binding energy/score across possible arrangements).

Question 5: X-ray crystallography and single-particle electron microscopy are both techniques for determining the structure of a molecule or molecular complex. Why is single-particle electron microscopy typically used for larger molecules or complexes than x-ray crystallography?

Solution: 1) It is generally harder to form crystals of larger molecules/complexes than of smaller ones
2) In single-particle electron microscopy, its usually harder to solve the computational reconstruction problem for smaller molecules (smaller molecules tend to have more similar looking cross-section densities from different orientations).

Question 6: Discuss the trade offs of a stochastic particle-based reaction-diffusion simulation versus a continuum approach (in which concentrations of each type of molecule are represented in each voxel).

Solution: A stochastic reaction-diffusion equation will be more appropriate for simulations involving a small numbers of particles. This is because we can more precisely simulate single-particle movement, which would not be possible with a continuum approach. If we have a large number of particles however, the continuum approach would be more appropriate, as we would likely be concerned with the overall diffusion behavior of all the particles, as opposed to a small handful. This behavior would be computed significantly faster and with less memory using a continuum based approach. This is because a stochastic model would require us to iterate through each particle in the stochastic simulation, a computationally expensive step that is not necessary in the continuum approach.

Question 7: There is a very efficient algorithm for computing the Fourier Transform known as the Fast Fourier Transform (FFT). Describe how this algorithm is useful for one of the methods covered in this course.

Solution:

- The FFT allows one to compute convolutions more efficiently. This is useful for filtering images, signals, etc.
- In x-ray crystallography, the FFT allows one to calculate efficiently the diffraction pattern associated with a hypothesized electron density.