MCB*4050 Protein and Nucleic Acid Structure

Midterm Exam (Practice)

Fall Semester
October 16th, 2008
2:30 to 3:50 p.m.
MacKinnon 116
Examiner: Dr. Matthew Kimber

Instructions: Please record all your answers in the examination booklet(s) provided. **Answers written in the test book will not be marked.**
Answer questions in **pen** using any colour except red.
Hand in all materials including this exam booklet

You have **80 minutes** to complete the exam.

There are **50 total marks.**
This midterm exam is worth **20% of your final course grade**
Part A: Answer all of the following questions.
(1 mark per question, 25 marks total)

1) What is meant if two proteins are described as being “orthologs”?

2) What does the “45” refer to in BLOSUM45?

3) Why might you use BLAST to search for a protein structure in the PDB instead of simply using the protein’s name?

4) Assuming that you are omniscient (all knowing), what criteria would you use to decide if a proposed phylogenetic tree was a “true” one?

5) Give the full name (e.g. histidine) of the residue in the following sequence that would be considered to be the strongest helix breaker: GAFRSYDTPN

6) How do you calculate the most favourable distance for two atoms to interact if the atoms are uncharged, and not connected by a covalent or a hydrogen bond?

7) Bond vibrations are the fastest atomic motions that occur in a protein. What time scale should we then use in a realistic molecular dynamics simulation of a protein?

8) What fraction of Xaa1-Xaa2 peptide bonds are found in cis configuration when Xaa2 is not proline.

9) What Greek letter (α, β, γ…) used to designate the depicted protein dihedral angle?

10) What is an alternate name for an α-helix (notation - Xγ where X and Y are numbers)?
11) Name the secondary structural element depicted below:

![Image]

12) Which residue(s) has a Phi/Psi distribution like the one shown in the Ramachandran plot below?

![Ramachandran Plot]

13) What portion of β-strands interact with one strand a parallel fashion, and another strand on the opposite side in an anti-parallel?

14) Why are negatively charged residues more common towards the N-terminus of α-helices?

15) Give one circumstance where it would be advantageous for a protein to be monomer.

16) Define the term “domain” as it applies to protein structure.

17) Green Fluorescent protein is unusual in that it forms an eleven-stranded β-barrel. What is the structural and functional role of the missing twelfth strand?
18) Where is the catalytic site of $\alpha\beta_8$ barrels invariably situated?

19) What name is given for the fold of the protein shown below?

![Protein Diagram](image)

20) What does it mean for a protein to be “folded” (as opposed to unfolded)?

21) Name a factor that enthalpically disfavors protein folding.

22) What is the ultimate fate of proteins that persistently fail to fold in vivo?

23) What Phi-Psi relationship gives rise to the typical twist of $\beta$-sheets?

24) Name one reagent that can be used to (reversibly) unfold proteins chemically.

25) Name one type of experimentally derived data that Rosetta can use to improve its prediction success over what would be achieved with just the primary sequence input?
Part B) Answer all of the following questions. Marks are as indicated. (25 marks total)

1) Describe the steps BLAST follows as it sets about finding “high scoring segment pairs” (HSPs).
(4 marks)

2) Describe an experiment and the associated calculation that would allow hydrophobicity values to be assigned to individual amino acids?
(4 marks)

3) Protein structures are often represented by contact surfaces in order to give a sense of the shape of the molecule. Describe how these contact surfaces are calculated.
(2 marks)

4) Why do coiled coils form a left handed superhelix?
(2 marks)

5) Describe the steps you would go through (using the Chou-Fasman secondary structure prediction scheme) before predicting that a stretch of amino acids will form an $\alpha$-helix.
(5 marks)

6) Explain how we know that proteins do not exhaustively search all possible conformations before settling into their native conformation. What is the name of this argument?
(4 marks)

7) Rosetta is a state of the art program which has had success in predicting tertiary structures of small proteins from the primary sequence. It breaks the problem down into a candidate fold generating stage, where structures are evaluated using a simplified “pseudo-energy function”, and then refines the best candidate folds using an all atom minimization procedure. Describe the component terms of the “pseudo-energy function” that is used to score the models Rosetta generates in its initial fold generation run.
(5 marks)

Bonus question (1)

Which protein we discussed in class was the subject of the 2008 Nobel Prize for Chemistry?
Practice midterm – answers

Part A

1) Two proteins that are homologs in different organisms and have the same function; they are related by a gene duplication event

2) It refers to the fact that sequences with 45% identity were used to derive this scoring matrix

3) Because protein names are not always standardized (more than one name for the same thing)
   Because analogs can have the same name while being completely different proteins
   Because this can sometimes be the only way of finding poorly annotated structural genomics structures.

4) A true phylogenetic tree is one that accurately recapitulates the evolutionary history of the group of sequences (or organisms) it depicts.

5) Proline

6) Add up their van der Waals radii

7) Femtoseconds (10e-15 s)

8) 0.05%

9) $\Phi$

10) 3.613

11) Pi helix

12) Proline

13) 20%

14) Forming an a-helix aligns the backbone dipole moments so that the positive vector points towards the N-terminus, the resulting partial positive charge at the N-terminus can be neutralized by placing a negatively charged residues at the N-terminus.

15) Where fast diffusion is required (e.g. plastocyanin)
   or Where the protein is secreted and required to operate at low concentration where oligomers would dissociate.
   or Where maximal surface area needs to be freed up for interacting with other ligands (e.g. signal transducing proteins)
16) A domain is a compact part of a structure that folds independently; single functions often map onto single domains.

17) The other strand passes through the center of the barrel, where three residues spontaneously react to form the fluorophor.

18) At the C-terminal end of the β-strands

19) αβ horseshoe fold

20) A folded protein is in a single (or small set of closely related) conformation, the structure normally solved by structure determination methods.

21) loss of hydrogen bonds as waters are stripped of residues being buried or van der Waals clashes or Strain introduced in protein covalent bonds

22) They are targeted for degradation

23) Phi is slightly more negative than Psi is positive

24) Urea or guanadinium hydrochloride or strong acid or strong base

25) Long range NMR NOE constraints or knowledge of a few disulfide bonds or the structure of a known homolog

Part B)

1) In the first stage of a sequence-sequence comparison, BLAST breaks the query sequence into a series of short words (normally 3 letters, defined by the word length W). It then generates all possible (3 letter) combination of related sequences that would align with a score (generated from the scoring matrix, e.g. BLOSUM45) better than or equal to some pre-specified minimum score (the neighbourhood word threshold, T). These words are then used to scan the sequence of interest residue by residue for an exact match. Exact matches are termed “High scoring segment pairs”.

2) A series of terminally blocked peptides with a short (5 a.a.) defined sequence are made with each of the 20 a.a. in turn occupying the central position. These are allowed to partition between water and an organic liquid, such as cyclooctanol. The free energy of partition is then calculated by $\Delta G = -RT \ln(\text{fraction of peptide in water} / \text{fraction of peptide in organic solvent})$.

3) Contact surfaces are calculated by rolling a 1.4 Å virtual sphere (the radius of a water
molecule) over the van der Waal spheres of the protein atoms. The contact surface is the surface where the surfaces of the spheres meet.

4) α-helices form right handed helices with 3.6 residues per turn. Coiled coils repeat every seven residues. Wrapping the two helices around one another with a slight left handed twist compensates for the extra 0.2 residues per two full turns, allowing the two α-helices to meet up every seven residues in the exact same relative configuration every seven residues.

5) Find a stretch of six amino acids where four of the six have P(α) larger than 100; this forms a helix nucleus. Extend this nucleus in each direction until you reach four amino acids in a row which the average P(α) is less than a hundred.
For this region, sum up the P(α) and then the P(β) values. If the total P(α) is larger than the total of P(β) and the run is more than 5 amino acids long, then these amino acids are predicted to form an α-helix.

6) Take, for example, a small protein of 100 amino acids. Lets assume that the side chains are inflexible, and the backbone can take on only 3 configurations – α-helical, β-stranded or “L”. The number of potential configurations of this protein is then 3^{100}. Even if the protein spends only femtoseconds in each configuration, it would still take significantly longer than the total age of the universe to fold. Since proteins typically fold in milliseconds, we can conclude that they do not explore every possible conformation. This is known as Levinthal’s paradox.

7) Local interactions are assumed to be favourable - interactions within a pdb derived fragment are therefore ignored.
Side chains are depicted as a centroid at the average side chain conformation (the “lollipop” model) - burying hydrophobic lollipops is rewarded.
Hydrogen bonding and electrostatics are not explicitly evaluated.
β-sheet geometry is probabilistically evaluated (strands collinear and close are rewarded).
Self intersection (van der Waals clash) is disallowed.
Global compactness is evaluated (approximates favourable van der Waals contribution).

**Bonus question**

Green fluorescent protein