Answers for Sample Final Problems

Problem I
Definitions

Find definitions/ descriptions of these terms in your textbook or notes.

Problem II
Secondary sequence analysis

a. The Chou-Fasman method is based on the propensity of different amino acids to form different types of secondary structure. It then looks for patterns of these different types of residues to determine the presence of helices and strands etc. These propensity values reflect the physics of the individual amino acids ie prolines are likely to break helices because of its sidechain.

b. The GOR method is based on the idea that neighboring amino acids influence the likelihood that a particular residue will adopt a particular secondary structure. It computes something like an information content by comparing the logodds ratio of probability that an amino is in a particular structure to the probability that it is not.

c. JPRED is a consensus method because it uses 6 different methods for prediction and compares/combines the results of all of them. Ultimately, it is the most accurate because if a particular secondary structure pattern is predicted using a variety of different methods then it is likely real rather than an artifact of a particular method.

Problem III
Protein Structure Prediction

a. 1) Comparative modeling
   Uses protein(s) of similar sequence(s) and known structure as a template to construct a model for the target. Used when a template structure(s) of > ~30% sequence identity is available. This cutoff is a rough measure that depends on what is known about both the target and the template.
   2) Fold recognition and threading
   Tests the “fit” of a sequence to each fold in a library of
possibilities using an energy function that is usually based on statistical properties of proteins in the pdb. Used when there is no obviously similar template but when it is likely that the target adopts a known fold (as is usually the case).

3) Ab initio methods
Constructs a model without the use of an existing structural template for the fold. ROSETTA, the best method for ab inition prediction, assembles structures from short peptide fragments.
Used when there is no homolog of known structure and when threading fails to identify a good hit.

b.
One is tempted to do comparative modeling of the N-terminal segment using CspA as the template; however, based on the abstract, it sounds like this synthetic protein could potentially have a different fold from that of CspA. Thus, it would probably be better to do fold recognition.

1) Choose a library of possible folds, e.g. one each of ~800 found in the pdb.
2) Align the target sequence with the template(s), using multiple sequence alignment if possible, and drawing on information from secondary structure prediction.
3) Score each alignment using a statistical potential that takes as input information about amino acid identity and spatial relationships (e.g. inter-residue distances) and returns an energy.
4) Determine whether any of the folds have statistically significant scores (e.g. use a Z-score)
5) For the best template, build a model by transferring the target sequence onto the template structure, using the alignment from (2)
6) Refine the structure by adjusting side chain positions, building missing loops
7) Validate and evaluate the structure by checking, e.g., if it has good stereochemistry, good hydrophobic/polar patterning

c.
1) Methods for sequence/structure alignment need to be improved. There is no good way to recover from a bad alignment in homology modeling or threading. Frequently, even when a good template is available, the best alignment of sequence to structure cannot be identified.
2) Better refinement methods are needed for homology modeling. One can rarely do better than simply taking the coordinates of the best template structure. Ideally, molecular simulations would better capture the physics of proteins and thereby generate better final models.
Problem IV
Protein Structure Modeling Approaches

a.
- Analyzing the dynamics of a protein in solution. The goal is to simulate the motion of a protein in solution (for example, the folding trajectory or some thermodynamic properties). An example scoring function to use in this case would be a an all atom molecular mechanics potential, which takes into account atomic van der Waals interactions, electrostatics in explicit water (using Coulomb’s law), harmonic bond stretching, bending and torsion energies. One search algorithm, which can be used in this case is Mote Carlo sampling (random moves are evaluated according to the change in energy they bring about) with simulated annealing (temperature decreases with time to the energies are allowed to jump around less and less and the search propagates – simulated a cooling down process).
- Protein design. Energy function similar to that for molecular mechanics e.g. including van der Waals interactions, some sort of solvation model terms for bond stretching, bending and torsion energies, and electrostatics. No explicit water. Usually, the energy function can be expressed a sum of single-residue and residue-pair energies. Some search algorithms appropriate for this case are: Monte Carlo/Simulated Annealing, Self Consistent Mean Field, Dead End Elimination.
- De novo structure prediction. Ideally, one would like the scoring function here also to be heavily physics based. However, most successes in this field have been with very much knowledge based potentials. For example, a potential which includes some sort of a close contact term, an empirical hydrogen bond term and a database derived residue to residue distance term (as a function of burial) may be appropriate in this case. The search algorithm can be some Rosetta style Monte Carlo searching where random segments of the structure are replaced with segments from the PDB and the move evaluated with the scoring function.

b.
Threading using an energy function that depends on the distance of residue-residue interactions cannot be combined with alignment using dynamic programming (a search technique) because the energy depends on the alignment, but the alignment depends on the energy.
Problem V
Modeling of simple reactions

(sorry about image quality)

a.

\[ \frac{[A^*]}{([A]+[A^*])} \]

This system is monostable.

b.
This is a picture of only the linear feedback ... not the basal forward reaction
Rate

$\frac{[A^*]}{([A]+[A^*])}$
c. The arrows indicate the two steady states. The one on the left is unstable and the one on the right is stable.

Simple linear feedback cannot generate a bistable system because it does not have two STABLE steady states.