

AMS Common Exam Part B, Computational Biology Track, January Exam 2011

Name: _____

ID Num: _____

Part B: _____ / 75

DO THREE OUT OF FOUR QUESTIONS ONLY: Questions are based on AMS-535 (Questions 1-2) and CSE-549 (Questions 3-4). Each question is worth 25 points.

Question 1. Note this questions has two parts (a) and (b).

1a. Draw a thermodynamic cycle used to determine the *relative* free energy of hydration ($\Delta\Delta G_{\text{hyd}}$) between two molecules A and B. Clearly label all parts of the figure. Write the simple expression which relates how legs of the cycle (computed using techniques such as free energy perturbation or thermodynamic integration) are equivalent to the difference in experimental free energies of hydration. Indicate which two legs of the cycle correspond to experimental observables and which two are calculated.

1b: If the association of a ligand L with receptor R in the condensed phase (water) is defined as the *absolute* experimental binding free energy (ΔG_{expt}), clearly draw the thermodynamic cycle which can be used to computationally estimate the *absolute* binding free energy (ΔG_{calcd}). Neatly write the simple expression which relates how all legs of the cycle are equivalent to ΔG_{expt} and ΔG_{calcd} .

Which term best corresponds to the *absolute* hydration free energy of the ligand $\Delta G_{\text{hyd}}(\text{L})$?

Write a commonly used expression that could be used to computationally estimate the *absolute* hydration free energy.

Question 2: Note this question has multiple parts (a-i)

(a) What are the four major groups of biomolecules.

(b) List the three basic components of each amino acid (protein unit).

(c) Name two common secondary structure elements found in proteins.

(d) A diffraction pattern observed in a crystallographic experiment is due to X-rays scattered by the atomic nuclei or orbiting electrons?

(e) Model quality from crystallographic refinement can be assessed and validated in a number of ways. List three structural criteria or types of analysis commonly used to tell if a deposited structure is reasonable.

(f) List the 5 energy terms commonly used in most classical force fields for simulating macromolecular systems.

(g) List four experimental observables used to help guide force-field parameterization.

(h) Name two types of error in a comparative protein structure model.

(i) For the 20 naturally occurring amino acids fill in the following table. Indicate which of the following properties best-describes each amino acid. Properties = hydrophobic, hydrophilic, aromatic ring, 5-membered ring, negatively charged, positively charged, ring in protein backbone, disulphide bonds, smallest side chain.

	Name	3 letter code	1 letter code	Property
01				
02				
03				
04				
05				
06				
07				
08				
09				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

Question 3: Design/propose an experiment to determine which genes are expressed more heavily expressed in cancer cells than normal cells. What computational/algorithmic problems will you encounter?

Question 4: Certain DNA sequencing technologies tend to produce errors where repeats of characters are often added, ie. the sequence CAAT may well be read as CAT or CAAAT. For such data, we do not want to penalize an insertion/deletion within such a run as much with other characters.

Give an efficient algorithm to find the edit distance between two strings, where insertion/deletion within a run of 2 or more consecutive characters costs X , where a normal deletion costs Y . Assume a substitution costs Z .