

AMS Common Exam Part B, Computational Biology Track, January Exam 2008

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: Draw a thermodynamic cycle commonly used to determine the *relative* free energy of binding ($\Delta\Delta G_b$) between two ligands A and B with a protein receptor P. Clearly label all parts and terms of your figure.

Write the simple expression which shows how two legs of the cycle (computed using techniques such as free energy perturbation) are equivalent to the difference in the experimental binding energies ΔG_b (A) and ΔG_b (B) between the two ligands.

Which term most closely corresponds to the *relative* free energy of hydration between ligands A and B?

Question 2: This question has multiple parts. For structural drawings label all atoms and indicate stereo chemistry with hashed or wedged lines.

(a) The sum of partial atomic charges for a ligand always yields a net formal charge of zero (true or false)?

(b) Carbohydrates do not contain hydroxyl groups (true or false)?

(c) Which energy terms in classical Molecular Mechanics force fields are considered the "bonded" terms?

(d) Name 2 types of noncovalent interactions that help stabilize folded proteins

(e) Electron distributions for a molecule are commonly modeled as a collection of "point charges" centered on the molecule electrons (true or false).

(f) Write the name and the ONE letter code for each of the 20 amino acids.

01:	11:
02:	12:
03:	13:
04:	14:
05:	15:
06:	16:
07:	17:
08:	18:
09:	19:
10:	20:

(g) Which four amino acid side chains are charged under most conditions ? Use THREE letter codes.

(h) Draw a simple line representation showing the side chain for lysine in the standard protonation state at physiological pH.

(i) Write the names of the two most commonly used implicit water models

(j) Give four examples of data plots (property versus time) often used in molecular dynamics simulation data analysis

Questions Based on CSE 549:

Question 3: A new organism has been discovered, and the nucleotide (DNA) sequence for the entire genome has been determined. Explain how you might find all the genes in the genome. Be sure to explain both the rationale for the method and the algorithmic approach.

Question 4: DNA microarrays have been used to measure the expression of 10,000 genes in 1000 different types of cells. Explain how you might find clusters of genes whose expression levels vary in similar ways across all cell types. Be sure to describe all steps of the process, from input to output. Briefly comment on how your approach will scale in terms of number of genes and number of cell types.