AMS Common Exam Part B, Computational Biology Track, January Exam 2010

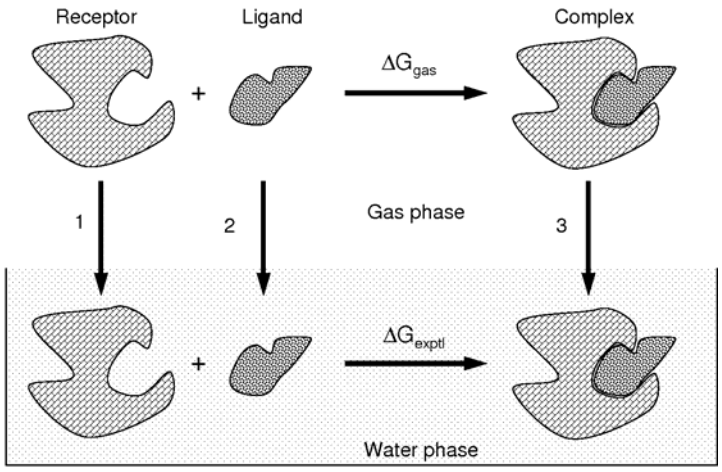
Name: _____ **ID Num:** _____

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 that this question has multiple parts (a-g)

(a) The thermodynamic cycle below shows how binding energy in solution can be estimated from ΔG_{gas} and three other terms labeled arrows 1, 2 and 3. What physical quantities are represented by the three vertical arrows?



arrow 1 = _____ arrow 2 = _____ arrow 3 =_____

(b) Write the most commonly used two-term expression used to computationally estimate the physical observables defined by the vertical arrows in the thermodynamic cycle.

_____

(c) Describe how each of the two terms listed as the answer for part (b) could be computed.

_____

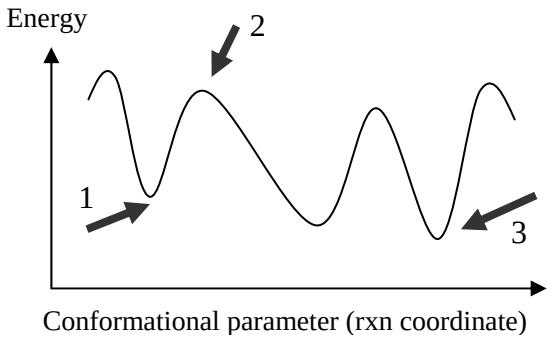
(d) Using the thermodynamic cycle above, write the expression for estimating ΔG_{exptl} as a function of the other terms.

ΔG_{exptl} = _____

(e) Using the thermodynamic cycle above, given three *separate* simulations representing the unbound receptor, the unbound ligand and the complex respectively, ΔG_{exptl} can be estimated as shown above. However, if only a single MD simulation of the complex is performed, how would one go about calculating the individual energy terms required to estimate ΔG_{exptl}. What are the pros and cons of using such an approach.

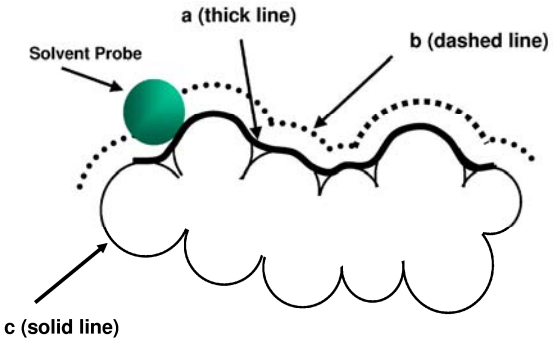
_____

(f) Write the common name given to each of the positions on the potential energy diagram marked by the numbered arrows.



arrow 1 _____ arrow 2 _____ arrow 3 _____

(g) List the names given to the surfaces defined in the following figure. (3 point)



a = _____ b = _____ c =_____

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 assed 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.

_____

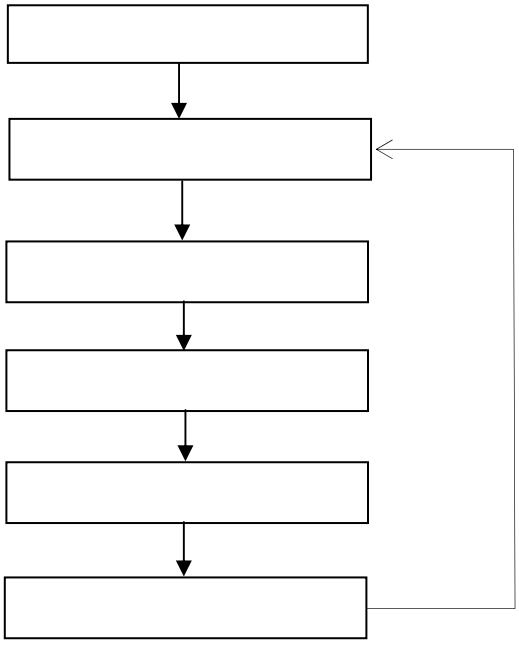
(g) List four experimental observables used to help guide force-field parameterization.

_____

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

_____

(i) A typical genetic algorithm cycle includes six steps: Initial Generation, New generation, Breeding, Survivors, Fitness pressure, and Termination Check (not necessarily in order). Fill in the following flow chart with these 6 terms in the correct order and write down how ‘Survivors’ are determined in the context of using such an approach for de novo ligand design.



Question 3: Describe how the genomes are sequenced. First describe the laboratory work, and then the computational procedures/issues that arise.

Question 4: Describe how you might write a program to identify genes in DNA sequence data from a database.