

AMS Common Exam Part B, Computational Biology Track, June 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: Write the most common form of the classical potential energy expression used in computer simulations that employ a Molecular Mechanics force field. Explicitly label all variables and constants that appear in this standard expression.

Question 2:

(a) Write the Linear Response (LR) expression used to estimate binding free energy.

(b) Write the more “general” Extended Linear Response (ELR) expression used to estimate binding free energy.

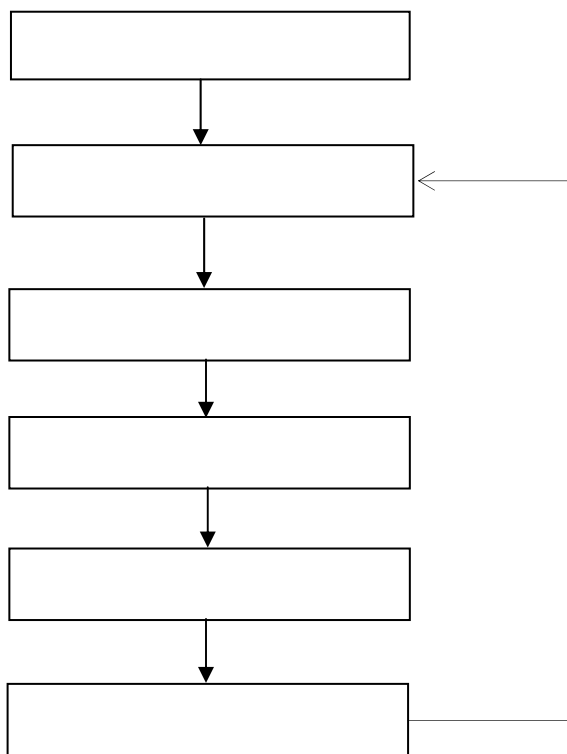
(c) Descriptor sets in ELR methods are chosen so that a maximum r^2 value will be obtained (between experimental and theoretical results) using a maximum number of descriptors. (true or false)

(d) What is the physical meaning of the negative sign of the coefficient (-0.216) of the ΔHB_{total} term in the following ELR equation.

$$\Delta G_{calcd} = 0.100\langle EXX - C \rangle + 0.110\langle EXX - LJ \rangle - 0.216\langle \Delta HB_{total} \rangle - 1.350$$

(e) List four terms which could be considered as important for describing binding energy in ELR models, i.e. the “descriptors”

- (f) 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 usually determined in the context of de novo design.



- (g) Breeding in the genetic algorithm cycle for de novo design involves which two primary steps?

- (h) For a related chemical series, smaller values for IC_{50} indicate more potent ligands (true or false)?

- (i) The smallest IC_{50} value for a molecule in a related chemical series translates into a *better* drug (true or false)?

(j) In DOCK 4.0, one of the sampling methods used was called incremental construction (also known as anchor and grow). List the three preprocessing steps the algorithm automatically does prior to docking each molecule?

(k) List the two *intermolecular* energy terms used in DOCK 4.0 to rank-order candidate orientations of a ligand with a receptor.

Questions Based on CSE 549:

Question 3: Describe how you might write a de novo protein folding program.

Question 4: Give an efficient algorithm for finding the shortest non-repeating string in a sequence, that is the shortest string which appears in the sequence only once. For example, in the sequence \$ACACCACCA\$ the shortest unique string is of length 3, namely \$ACA\$.