AMS Common Exam Part B, Computational Biology Track, January Exam 2020

Name: _____ ID Num: _____

Part B: _____ / 75

DO THREE OUT OF FOUR QUESTIONS ONLY: Questions are based on AMS/CHE-535. Each question is worth 25 points.

Question 1. Note this question has multiple parts.

1a. Define database enrichment as it relates to virtual screening and explain how ROC curves can be used to assess the accuracy of a given computational method or scoring function. Draw and label three examples of ROC curves with (i) poor, (ii) reasonable, and (iii) good enrichment.

1b. Define what constitutes a pharmacophore, explain the difference between a ligand-based and a receptor-based pharmacophore, and give at least four examples of features (structural or functional) that could be used.

1c. Write the name, a key property, and draw the chemical structure for five common organic functional groups.

Name	Property	Structure

1d. Van der Waals (VDW) interactions in Molecular Mechanics force fields are usually computed using a Lennard-Jones potential energy function. On the axis below, sketch a curve which describes VDW interactions between two atoms A and B. Label each axis and indicate the attractive and the repulsive regions of the curve as well as the point which represents the equilibrium distance.

1e. Fill in the following table for the 20 naturally occurring amino acids and 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.

	Residue Name	**3 letter code**	**1 letter code**	**Residue Property**
01	glycine			
02		ALA		
03			V	
04	leucine			
05		ILE		
06			S	
07	threonine			
08		CYS		
09			M	
10	proline			
11		ASP		
12			N	
13	glutamic acid			
14		GLN		
15			K	
16	arginine			
17		HIS		
18			F	
19	tyrosine			
20		TRP		

Question 2. Note this question has multiple parts.

2a. Write the most common functional form (i.e. the actual equations) for each of the five terms that constitute 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.

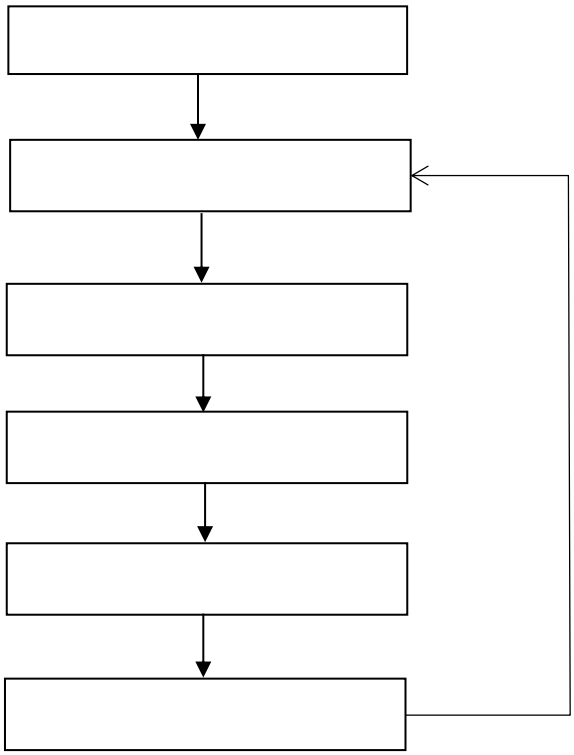
2b. Describe in detail how one would go about performing a virtual screening experiment for a newly discovered therapeutic target. Include in your answer an explanation of how the system would be setup, use of docking controls, and potential stumbling blocks.

2c. Describe the differences between virtual screening and *de novo design*, give pros and cons of each method, and list at least three challenges associated with de novo growth of small organic molecules.

2d. Describe Lipinski's rules and explain in detail why such "rules" are important.

Question 3. Note this question has multiple parts.

3a: 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.



3b: Write the Linear Response (LR) expression used to estimate binding free energy. Note LR is sometimes called the Linear Interaction Energy (LIE) method.

3c: Write the more “general” Extended Linear Response (ELR) expression used to estimate binding free energy.

3d: Descriptor sets in ELR methods are chosen so that a maximum r squared value is obtained (between experimental and theoretical results) using a maximize number of descriptors. (true or false)

3e: What is the physical meanings of the negative sign of the coefficient (-0.216) of the ΔHB_{total} term in the following ELR equation.

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

_____

3f: List four terms which could be considered as important for describing binding energy in ELR models, i.e. the “descriptors”

_____

3g: The sum of partial atomic charges for a ligand always yields a net formal charge of zero (true or false)?

_____

3h: Carbohydrates do not contain hydroxyl groups (true or false)?

_____

3i: Name 2 types of noncovalent interactions that help stabilize folded proteins

_____

3j: Electron distributions for a molecule are commonly modeled as a collection of "point charges" centered on the molecule electrons (true or false).

_____

3k: Give four examples of data plots (property versus time) often used in molecular dynamics simulation data analysis

_____

3l: Which four amino acid side chains are charged under most conditions ? Use THREE letter codes.

_____

Question 4: Note this question has multiple parts.

4a: Draw a thermodynamic cycle commonly used to compute the *relative* free energy of binding ($\Delta\Delta G_b$) between two ligands A and B with a protein target P using Free Energy Perturbation Methods. Clearly label all parts and terms of your figure.

Write the simple expression which shows how two legs of the cycle 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?

4b: Draw a thermodynamic cycle commonly used to compute the *absolute* free energy of binding (ΔG_{bind}) between a ligand L with a receptor target R using the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) Method. Clearly label all parts and terms of your figure.

Write the simple expression which relates which legs of the thermodynamic cycle are used to computationally estimate the *absolute* free energy of binding ΔG_{bind}, which, if the calculations were exact, would be equivalent to the *absolute* experimental free energy of binding ΔG_{expt}.

Indicate which leg best corresponds to the *absolute* hydration free energy of the ligand and provide a two term equation commonly used to estimate ΔG_{hyd}.