AMS Foundation Exam Part B, Computational Biology Track, January Exam 2025

Name: _____ ID Num: _____

Part B: _____ / 75

Please complete ALL three questions which are based on AMS/CHE-535. Each question is worth 25 points.

Question 1. Note this question has multiple parts.

1a: Describe in DETAIL how ROC curves can be used to assess database enrichment, and therefore the potential accuracy of a given computational method or scoring function for virtual screening. What commonly used database construction protocols help to avoid artificial enrichment? Draw and label three examples of ROC curves with (i) poor, (ii) reasonable, and (iii) good enrichment.

1b: Describe in DETAIL how atomic level Molecular Modeling force fields for Molecular Dynamics (MD) and Monte Carlo (MC) simulations were developed and tested. Include in your answer the direct connection between experimental observables and computational outcomes and what key properties were compared (give multiple examples) when developing force fields such as OPLS, AMBER, and CHARMM.

1c: A colleague asked you to collaborate on identification of drug-lead candidates for a newly discovered therapeutic protein target. Explain in DETAIL how you would go about performing a large-scale virtual screening experiment to prioritize compounds for purchase and experimental testing. Be specific and detailed in how you would go about the setups and the calculations and use of controls. Include in your answer potential problems (and remedies) that you might encounter.

1d: 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.

	3 letter code	**1 letter code**	**Residue Name**	**Residue Property**
01			proline	
02	TRP			
03			glycine	
04	ALA			
05		S		
06			glutamic acid	
07	LYS			
08			valine	
09	ASP			
10			arginine	
11			glutamine	
12		L		
13			isoleucine	
14		T		
15			cysteine	
16	ASN			
17	MET			
18		H		
19		F		
20			tyrosine	

Question 2. Note this question has multiple parts.

2a: Discuss the primary differences with regards to employing a *receptor-based* method versus a *ligand-based* method for drug discovery projects. For each method give one example of the types of "features" that that could be employed if a pharmacophore-based model was being used.

2b: Discuss the primary differences between use of *virtual screening* and *de novo design* for lead discovery and give one challenge (or con) associated with each method.

2c: What algorithmic strategies or protocols could be used to help mitigate known combinatorial explosion issues when using *de novo design* to construct ensembles of small organic molecules in a protein drug target site.

2d. Describe radial distribution functions, how are the computed, and why are they useful?

2e. What are protein point mutations and how do they arise? How can mutations alter drug potency from a structure-based perspective?

2f: List four properties commonly plotted as a function of time when interrogating the outcomes of molecular dynamics simulations.

_____

2g: Write the simple two-term Linear Response (LR) expression (equation) used to estimate binding free energy. Note LR is sometimes called the Linear Interaction Energy (LIE) method.

_____

2h: Write the more "general" Extended Linear Response (ELR) expression used to estimate binding free energy.

_____

2i: Descriptor sets in ELR methods are chosen so that a maximum r squared value is obtained (between experimental and theoretical results) using a minimum number of descriptors. (true or false)

_____

2j: List four types of descriptors (terms) which could be employed to describe binding energy in terms of multiple linear regression equation models.

_____

2k: Carbohydrates contain hydroxyl groups (true or false)?

_____

2l: Electron distributions for a molecule are commonly modeled as a collection of "point charges" centered on the molecule nuclei (true or false).

_____

2m: The sum of the partial atomic charges for a given ligand always yields a net formal charge of zero (true or false)?

_____

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

_____

Question 3. Note this question has multiple parts.

3a: 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 the "generic" two-term equation commonly used to estimate ΔG_{hyd}.

3b: Draw a thermodynamic cycle used to determine the *relative* free energy of hydration ($\Delta\Delta G_{hyd}$) between two molecules A and B. Clearly label all parts of the figure. Write the simple expression which relates how two legs of the cycle (computed using techniques such as free energy perturbation or thermodynamic integration) are equivalent to the *difference* in two experimental free energy of hydration values ΔG_{hyd} (A) and ΔG_{hyd} (B).