**AMS Common Exam Part B, Computational Biology Track, May Exam 2021**

**Name: _____ ID Num: _____**

**Part B: _____ / 75**

**Please complete ALL 3 questions which are based on AMS/CHE-535. Each question is worth 25 points. This exam has 10 pages, please make sure you have all pages.**

**Question 1. Note this question has multiple parts.**

**1a.** 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.

**1b.** Describe in DETAIL how one would setup and execute a large-scale virtual screen to identify potential inhibitors targeting a recently crystallized SARS-CoV-2 protein. Include in your answer an explanation of how the system would be setup, use of docking controls, how compounds would be prioritized for purchase and experimental testing, and potential stumbling blocks. Also include in your discussion what follow-up computational experiments would be useful (and why) once initial hits with experimental activity were identified.

**1c.** 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.

**Question 2. Note this question has multiple parts.**

**2a.** Explain why pharmacophores are useful in the context of drug design. Explain the difference between ligand-based and receptor-based pharmacophore methods. Which of the two methods is more powerful and why ? Give at least **5** examples of features (structural or functional) that can be used to define a pharmacophore.

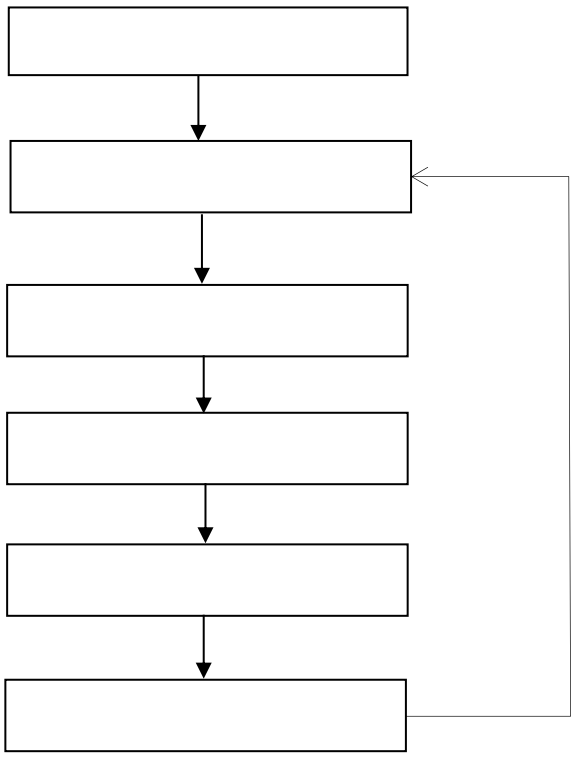
**2b.** Fill in the following table for the 20 naturally occurring amino acids and indicate which of the following properties best-describes each amino acid. Properties = hydrophobic, hydrophilic, aromatic ring, 5-membered ring, negatively charged, positively charged, ring in protein backbone, disulphide 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 | | |

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

**2d.** Explain how a Genetic Algorithm (GA) could be used to help identify potential inhibitors in the context of receptor-based structure-based drug design.

If the GA consists of 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.



**Question 3. Note this question has multiple parts.**

**3a.** Draw a thermodynamic cycle commonly used to compute the *absolute* free energy of binding ($\Delta G_{bind}$) 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 $\Delta G_{bind}$, which, if the calculations were exact, would be equivalent to the *absolute* experimental free energy of binding $\Delta G_{expt}$.

Indicate which leg best corresponds to the *absolute* hydration free energy of the ligand AND provide a two-term equation commonly used to estimate $\Delta G_{hyd}$.

**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.** List four terms which could be considered as important for describing binding energy in ELR models, i.e. the "descriptors"

---

**3e.** What is the physical meanings of the negative sign of the coefficient (-0.216) of the $\Delta HB_{total}$ 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.** Define database enrichment as it relates to validating virtual screening methods and explain how ROC curves can be used to assess the accuracy of a given computational method or scoring function.