Quiz 3 (CADD) BIO563

Section A (1 mark each)

1. Which of the following best describes the primary goal of molecular docking in drug discovery?

   A. To measure the solubility of a drug molecule in water
   B. To predict the binding affinity and orientation of a ligand within a target's active site
   C. To determine the sequence of a mino acids in a protein
   D. To simulate the folding pathway of a protein

2. While calculating Tanimoto similarity from a SMILES string, two common types of molecular fingerprints can be generated: **Morgan** and **RDKit**. What type of fingerprint does each represent?

   A. Morgan – linear; RDKit – radial
   B. Morgan – path-based; RDKit – circular
   C. Morgan – circular; RDKit – path-based
   D. Morgan – hashed; RDKit – uncompressed

3. Which of the following is the correct Open Babel command to convert all SDF files in the current directory into PDBQT format?

   A. obabel -isdf *.sdf -opdb -O*.pdbqt
   B. obabel -isdf *.sdf -opdbqt -O*.pdbqt
   C. obabel -ipdb *.pdb -osdf -O*.sdf
   D. obabel -isdf *.sdf -opdbqt -O*.pdb

4. In the following GROMACS command, What does the highlighted part represent?

   "gmx grompp **-f minim.mdp** -c 1AKI_solv_ions.gro -p topol.top -o em.tpr"

   A. Parameter file
   B. Structure after solvation and ion addition
   C. Output binary run file
   D. Topology file

5. When adding ions to your system in GROMACS, you are prompted to select a group where ions will be placed. Which group should you choose, and why?

   A. Protein, to neutralize the active site directly
   B. SOL, so that water molecules are replaced instead of protein atoms

C. System, to distribute ions randomly
D. DNA, to maintain charge balance near nucleic acids

6. Identify the correct vina command.
   A. vina -receptor protein.pdbqt -ligand ligand.pdbqt -config conf.txt -out output.pdbqt
   B. vina --receptor protein.pdbqt --ligand ligand.pdbqt --center 0 0 0 --size 20 20 20 --output output.pdbqt
   C. vina --config conf.txt --log docking.log --out output.pdbqt
   D. vina config=conf.txt log=docking.log out=output.pdbqt

7. Which file extension represents the GROMACS binary input run used by **genion**?
   A. .gro
   B. .top
   C. .tpr
   D. .mdp

8. What insight does the radius of the gyration (Rg) plot provide during the MD simulation?

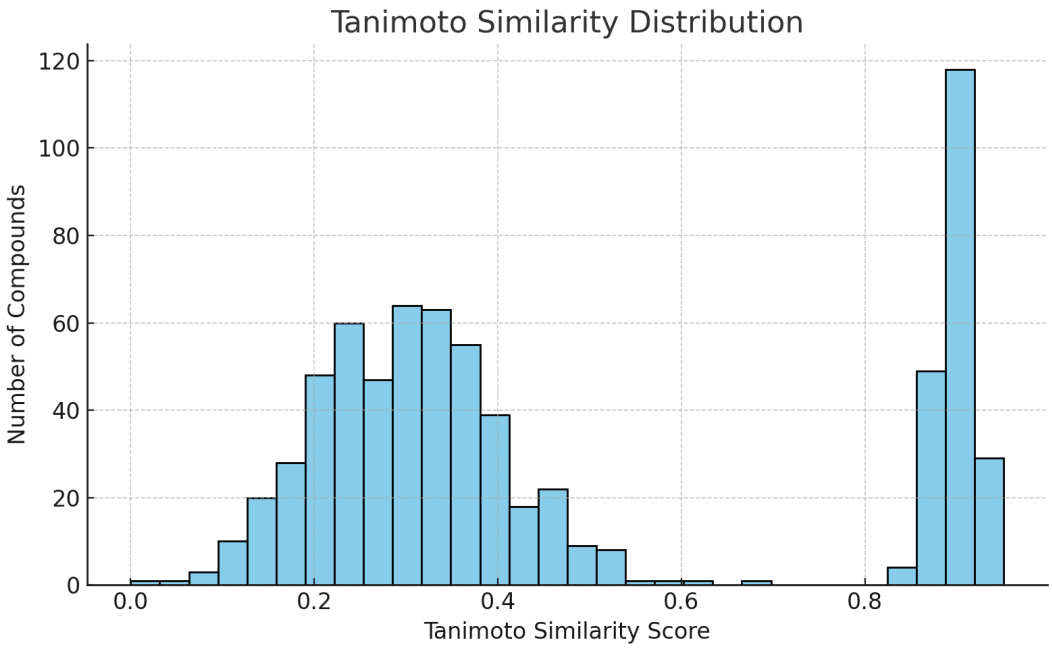
   A. Compactness and folding of the protein structure
   B. Stability of the electrostatic energy
   C. Local residue flexibility
   D. Total number of H-bonds

9. What is the impact of using the **-conc flag** when adding an ion in **genion**?

   A. It replaces protein residues with ions
   B. It adds ions randomly
   C. It ensures fixed neutralizing ions only
   D. To add user-defined concentrations of ions beyond neutrality.

10. Which interpretation is most accurate for Tanimoto similarity plot of a given query molecule against a given database?

   A. Most compounds in the database are chemically unrelated to the query molecule.
   B. The query molecule has many near-identical analogs in the database.
   C. The plot gives no insight into the similar compounds found in the database.
   D. The database compounds are highly diverse, with no clear similarity trends.



Section B (2 marks each)

1. How does the receptor file in **.pdb** format differ from the receptor file in **.pdbqt** format during the docking preparation process?
   Ans: receptor.pdbqt includes polar hydrogens and Gasteiger/kollman charges and water (HOH) is removed, while receptor.pdb does not.

2. Mention the key output files generated during energy minimization step in a molecular dynamics simulation using GROMACS.
   Ans: key output files
   - em.gro : Final energy-minimized structure.
   - em.log : EM log file.
   - em.edr : Energy data (used for plotting).
   - em.trr : Trajectory file (optional but useful for visualizing EM process).

3. What is the purpose of pdb2gmx in GROMACS and what does topol.top file contains?
   Ans: The purpose of pdb2gmx is to generate three files:
   - The topology for the molecule.

- A position restraint file.
- A post-processed structure file.

The system topology is a file that describes all the components of your system, including atoms, bonds, charges, angles, and interactions, according to a chosen force field. In simple terms it defines how your molecules are built, how they interact, and how they move during the simulation.

4. Given the binary fingerprints of two molecules A and B, Calculate the **Tanimoto similarity coefficient** between molecule A and B. what inference can be drawn from the score?
    - **Fingerprint A:** 1 0 1 1 0 1 0 0
    - **Fingerprint B:** 1 1 1 0 0 1 0 1

    Ans: Tanimoto similarity is calculated as = c/(a+b-c)

    - **a** = number of 1s in A
    - **b** = number of 1s in B
    - **c** = number of 1s in **common** (intersection of A and B)
    - A = 1 0 1 1 0 1 0 0 → a = 4
    - B = 1 1 1 0 0 1 0 1 → b = 5
    - Common = 1 0 1 0 0 1 0 0 → c = 3
      T= 3/(4+5-3) = 0.5

5. For the given config file below, identify the error and correct them to generate 15 mode of orientation while docking.

```
protein = hsg1.pdb
lig = ligand.sdf

center_x =  2
center_y =  6
center_z = -7

size_x = 25
size_y = 25
size_z = 25
```

```
receptor = hsg1.pdbqt
ligand = lisuride.pdbqt

center_x = 2
center_y = 6
center_z = -7

size_x = 25
size_y = 25
size_z = 25

num_modes = 15
```