Dr. C. Rout

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
END SEMESTER EXAMINATION-2015
B.Tech. (BI) VI Semester

COURSE CODE: 10B11BI611 MAX. MARKS: 45
COURSE NAME: ADVANCED CHEMOINFORMATICS
COURSE CREDITS: 04 MAX. TIME: 3 HRS

Note: *All questions are compulsory. Carrying of mobile phone during examinations will be treated as a case of unfair means. Provide example and schematic diagram wherever applicable.*

Section A (9 Marks)

1. Each question carries 1 mark.
   a. Tanimoto coefficient is calculated using fragments/fingerprints. If you were asked to calculate this coefficient using 3D descriptors then what types of problems do you face and how do you overcome them?
   b. Suppose you would like to use quantitative structure activity relationship (QSAR) method to find new lead. When the QSAR model developed from inhibitors will fail and why?
   c. What are the advantages of Caco-2 cell lines used for predicting clogP as compared to fragment-based approach?
   d. How do pharmacokinetics and pharmacodynamics principles help to decide initial lead molecules?
   e. How do you determine active molecules against a target using recursive partitioning method?
   f. What is the major application of distance matrix and why do the upper and lower halves of the matrix differ?
   g. What is the major application of SMILES and how do you use it for virtual combinatorial library preparation?
   h. What is the difference between q^2 and r^2 in QSAR? For a better model, how are these interrelated?
   i. Suppose you have a set of active and inactive molecules and you have the descriptors. Out of QSAR and support vector machine (SVM) approaches, which one generally gives higher accuracy and why?

Section B (13.5 Marks)

2. Each question carries 4.5 marks.
   a. Suppose you have a 3D molecular database. How do you convert it into a pharmacophoric database (explain where do you get the data and how do you annotate them)? What types of options do you provide to the users so that the user can explore it maximally? (3.5+1)
   b. Four molecules and their number of conformations will be provided. How do you determine pharmacophore using clique detection technique? What are the advantages of this method as compared to ensemble distance-geometry technique? (3+1.5)

P.T.O.

c. Suppose an active site is known to be flexible and inhibitors against this active site are also known. More than 3000 descriptors from different categories can be determined for the inhibitors. What types of descriptors do you need to consider initially and why? How do you come up with reasonable set of descriptors that are related to activity? How do you modify or design new ligands based on your data? (1.5+2+1)

Section C (22.5 Marks)

3. Each question carries 4.5 marks.
   a. How do you identify lead molecules using mix and split synthesis method? How do you create a diverse virtual library using combinatorial chemistry principle and how do you computationally identify potential compounds from this library? (2+2.5)
   b. How do you generate pharmacophore using protein-protein interactions and how do you ascertain the accuracy of the model? What is the structure-based pharmacophore (SBP) method and discuss the unique applications of both the methods? (2.5+2)
   c. The absorption, distribution, metabolism, excretion and toxicity (ADMET) prediction is more biology than chemistry of a drug molecule. How do you develop models for cytochrome P450 isoenzymes, so that metabolism of a drug can be predicted? How do you use pharmacokinetics modeling and empirical scoring method to predict metabolism? (2.5+2)
   d. Why are the grid-based QSAR methods so popular today and what values are stored at each grid points? Make a comparative analysis among G-WHIM, CoMFA and 3D-Morse descriptors? What is chemometrics and how is it correlated with predictive power of a QSAR model? (1.5+2+1)
   e. Classification models are important in chemoinformatics, and suppose you have active and inactive molecules. How do you develop a model which classifies the data using principal component analysis (PCA) method? How do you validate the model and also how do you use it for new lead discovery and refinement? (2.5+2)