



INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

END SEMESTER EXAMINATION- BT 305: Computational Biology

May 6, 2023. Maximum Marks: 100. Time: 3 hours

All questions carry 10 marks each, last question has 20marks; this question paper has 3 pages.

1. Explain the Hansch equation which describes the quantitative structure activity relationships that rationalize the biological activity of a molecule. Elaborate the possible reasons for the parabolic dependence of the activity of a molecule upon $\log P$ (P is partition coefficient).
2. Scoring functions for molecular docking attempt to approximate the binding free energy of the ligand binding to the receptor. Write at least two additive equations with components, reflecting their individual contributions to binding, with details of terms involved.
- ✓ 3. The SMILES representation of four molecules are as follows. Attempt to draw an approximate 2D structure of these molecules.

c1c(N(=O)=O)cccc1 ✓

CC(=O)C ✓

N[C@@H](C)C(=O)O

CC(O)C ✓

4. A) Bit strings for query and test molecules in a typical fingerprinting exercise is given below.

Mol 1	1 ✓	0	1 ✓	1	1 ✓	0	0	1	0
Mol 2	0	0	1 ✓	0	0	1	1	1	1
Mol 3	0	1 ✓	1 ✓	0	1 ✓	0	1 ✓	1 ✓	1 ✓
Mol 4	1 ✓	0	0	1 ✓	0	0	1	1	0

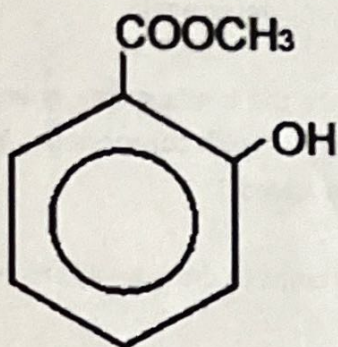
Calculate Tanimoto coefficients for structure comparison and attempt the following question:

If receptors of the four candidate structures Mol 1 Mol 2 Mol 3 and Mol 4 are respectively Rec 1, Rec 2, Rec 3 and Rec 4. Which molecule (other than Mol1) is more likely to bind with Rec 1? Rationalize your answer.

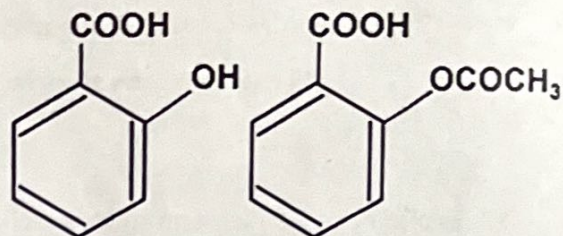
- ✓ B) Show how Euclidean and city block distances vary while measuring dissimilarity between two individual patterns.

5.

- A) Suppose we are performing an experiment of structure based molecular profiling for methyl salicylate (below structure) in a structure pool of four members on the right hand side, how do we determine the closest structure by calculating Tanimoto coefficient?

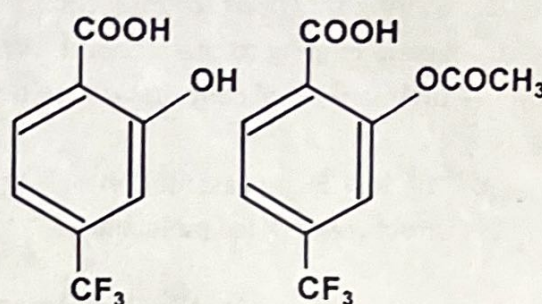


Salicylic acid Aspirin



HTB

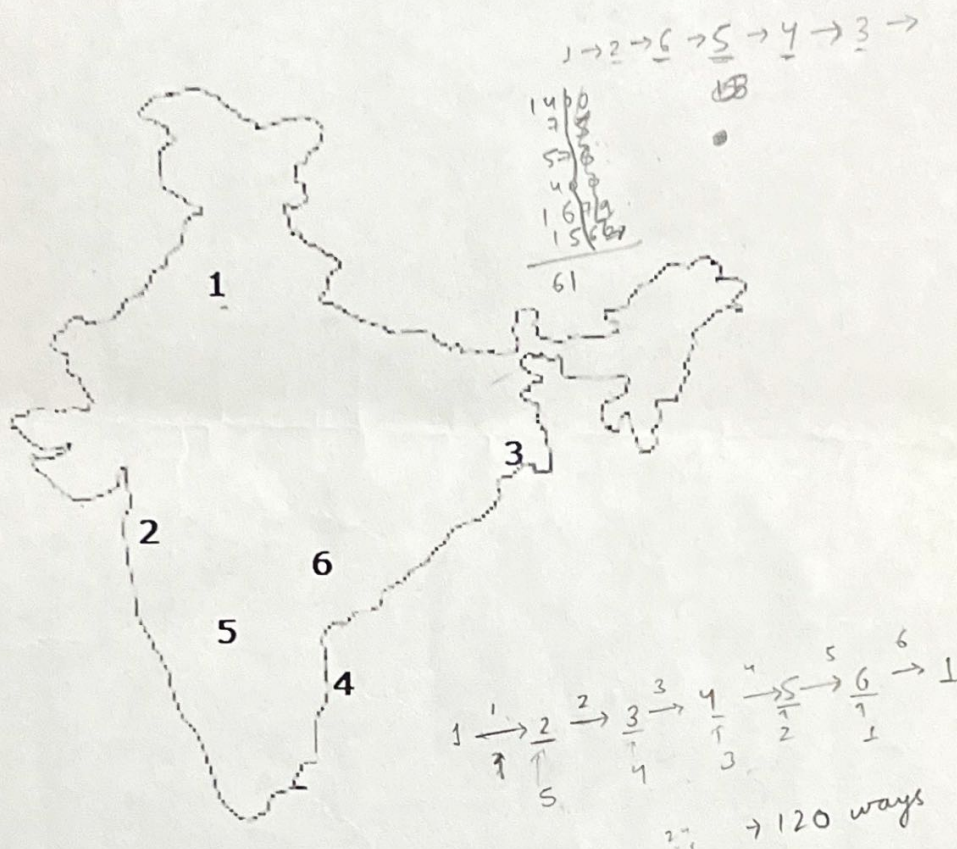
Triflusal



6. In a Monte carlo simulation, energy is calculated at each step and compared with the previous step. What is the probability of accepting the step? How are the difference in energy between subsequent steps and temperature affects the likelihood of acceptance?
7. Using all three genetic operators (Selection, Crossover and Mutation), maximize the function $f(x) = x^2$, with 'x' lies in the interval 0 to 25. Generate initial population as 6, 11, 18 and 23.
8. Explain briefly the following concepts.
A) Simulated Annealing B) Energy minimization C) RMSD based clustering D) Solvent accessible surface area.
9. A travelling salesman is required to cover all the shown 6 cities (only once) in India with minimum cost and get back to the starting point Delhi. Using Kruskal's minimum spanning tree algorithm (and if needed the methods of optimizing the route employing operators in genetic algorithm), attempt to find the most economical route (20 Marks).

Distance Matrix

	1	2	3	4	5	6
1	0					
2	1400	0				
3	1562	1892	0			
4	2208	1316	1679	0		
5	2174	985	1868	400	0	
6	1560	715	1487	626	576	0



Notes:

1. All distances in km.
2. You may approximate 100 km as one unit. For example, Delhi – Chennai distance of 2208 as 22 OR Chennai – Hyderabad distance of 626 as 6.

END OF QUESTIONS