JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST-3 EXAMINATIONS-2022

B.Tech-III Semester (BI)

COURSE CODE (CREDITS): 18B11BI313 (4) **MAX. MARKS: 35**

COURSE NAME: Biological computation

COURSE INSTRUCTORS: Dr. Shikha Mittal **MAX. TIME: 2 Hours**

Note: All questions are compulsory. Marks are indicated against each question in square brackets.

Q1. Compute the local alignment between the following sequences: GATTACA, TATACG with the following rules: match score = +5, mismatch = −3, gap penalty = −4 and discuss how the alignment depends on the choices of match scores, mismatch and gap penalty. **(5 marks) [CO-1]**

Q2. Hidden Markov models (HMM) are widely used in Bioinformatics. Explain with one example how you would detect CpG islands in a genome. **(4 marks) [CO-3]**

Q3. You have built a homology model based on the pairwise alignment of a *C. elegans* Glutaminyl-tRNA synthetase sequence to a template that you have identified with a BLAST search. Upon analysis of your model structure, you note that the presumed ATP-binding site of the model appears to be blocked by an arginine sidechain. **(4 marks) [CO-4]**

a) Briefly define the key steps involved in building a homology model.
b) Based on these steps, identify two probable sources of the discrepancy between the model and the function of the protein.

Q4. Define the UPGMA algorithm and state and justify its complexity. What is the output of the algorithm given the distance matrix of the species X1, X2, X3, X4 below? **(5 marks) [CO-6]**

$$\begin{pmatrix} species & X_1 & X_2 & X_3 \\ X_2 & 2 & & \\ X_3 & 4 & 4 & \\ X_4 & 6 & 6 & 6 \end{pmatrix}$$

Q5. Explain the concept of scoring matrices for aligning amino acid sequences. Briefly explain how PAM is derived? **(4 marks) [CO-2 & CO-3]**

Q6. Discuss with one example how an algorithm for RNA folding could use a Zuler folding algorithm and Nussinov folding algorithm. What are the cases when this will not work? **(4 marks) [CO-5 & CO-6]**

Q7. If you have a gene sequence of an organism whose genome is yet to be annotated, how will you proceed forward via *in silico* analysis to trace a specific region with a known function? Which tool will be useful? Explain the concept and different categories of that tool in detail. **(3 marks) [CO-5 & CO-6]**

Q8. Explain – (6 marks) [CO-1, CO-2, CO-5]

a. Homologs and Orthologs
b. FASTA
c. Bootstrapping