Dr. C. Rout

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
TEST 2 EXAMINATIONS - DECEMBER 2017
B.Tech III Semester (BI)

COURSE CODE: 10B11BI311 MAX. MARKS: 35
COURSE NAME: BIOLOGICAL COMPUTATION
COURSE CREDITS: 04 MAX. TIME: 2 HOUR

Q1. Each question carries 1 mark. Answer any two. (1 x 2 = 2)

i. What are the major applications of probabilistic modeling and why is it being frequently used in bioinformatics?
ii. How family of a sequence is determined using PFam server?
iii. What is hidden in hidden Markov model (HMM) and why?

Q2. Each question carries 2 marks. Answer any two. (2 x 2 = 4)

i. Progressive methods are used for multiple sequence alignment (MSA), why? Suppose you use different software for MSA using progressive methods, how do you compare their outputs? How do you determine accuracy of MSA? (0.5+1+0.5)
ii. Why double and triple affine penalties are used in MSA than pairwise alignments? How do you implement affine penalty in dynamic programming matrix?
iii. Describe about the branch and bound method with their applications?

Q3. Each question carries 3 marks. (3 x 3 = 9)

i. Suppose you have two RNA sequences. How do you carry out their alignment? In contrast to proteins, RNA secondary structure prediction is important, why? How do you use secondary structure information of RNA to predict its tertiary structures?
ii. Suppose you have sequences of a genome. How different modules of BLAST can be used to annotate genes? If you are unable to identify functions of few genes then how do you use local patterns to get more information about them? (2+1)
iii. Why BLOCK is considered as characteristic of a protein family than Prosite pattern? How do you develop an evolutionary profile using BLOCK of a family and compare the performance of this profile with the one developed using profile? (1+2)

Q4. Each question carries 5 marks. (5 x 4 = 20)

i. Describe the methodologies used to develop PAM250 scoring matrix (Data collection, mutations, odd-ratio, probabilities, etc.)? Why does performance of this matrix lesser than BLOSUM? Which matrix is used as default matrix in BLAST and why? (3+1+1)
ii. States and sequence data for a protein family will be provided. Determine the transition and emission probabilities for this family? Take any amino acid sequence of length five and calculate its probability of emission? (4+1)
iii. What is Markov principle and why is it being used? Transition probability of CpG and non-CpG will be provided. How was this table developed? Determine the sequence "ACGTGCAGG" is from CpG or non-CpG? In CpG island table, the C and G column values are significantly higher than A& T, why? (1+1+2+1)
iv. RNA sequence is given as "GGGAAAUCC". Use DMP to determine stem and loop present in this sequence? Use energy minimization method to determine secondary structure of the same sequence? How dot matrix plot is used to determine secondary structure in large RNA sequence? (2+2+1)