

**Indraprastha Institute of Information Technology Delhi (IIITD)**

**Department of Computational Biotechnology**

**BIO213 – Introduction to Quantitative Biology**

**Quiz-2 (April 18, 2022)**

---

1. In Prosite, which of the following patterns is wrongly matched with its significance? **(1 mark)**
- a. > – C-terminus
  - b. { } – Any but not
  - c. ( ) – Repeats
  - d. Z – Any

Correct answer: d

2. How many bonded atoms are required to constitute a dihedral (torsion) angle, such as phi or psi? **(1 mark)**
- a. None
  - b. 1
  - c. 2
  - d. 3
  - e. 4
  - f. 5

Correct answer: e

3. Which of the following wrongly describes protein domains? **(1 mark)**
- a. A single domain comprises of only a single type of secondary structure
  - b. Defined as independently foldable units
  - c. They are stable structures as compared to motifs
  - d. They are separated by linker regions

Correct answer: a

4. Even if many pseudocounts are added in comparison to real sequence counts, the amino acid frequencies will not have any effect or influence. **(1 mark)**
- a. True
  - b. False

Correct answer: b

5. What is iterated in PSI-BLAST? **(1 mark)**
- a. PSSM
  - b. Z-score
  - c. E-value
  - d. BLOSUM62

Correct answer: a

6. The overall height of a logo position reflects how conserved the position is, and the \_\_\_\_\_ of each letter in a position reflects the \_\_\_\_\_ of the residue in the alignment. **(1 mark)**

- a. height, relative frequency
- b. width, relative frequency
- c. height, amplitude
- d. width, amplitude

Correct answer: a

7. What does this representation mean -  $\langle A-x-[ST](2)-x(0,1)-V-\{ED\} \rangle$ ? **(2 mark)**
- a. C terminal- Ala-any-[Ser or Thr]-[Ser or Thr]-(any or none)-Val-{any except Glu or Asp}
  - b. N terminal-Ala-any-[Ser or Thr]-[Ser or Thr]-(any or none)-Val-{any except Glu or Asp}
  - c. C terminal- Ala-any-[Ser or Thr]-[Ser or Thr]-(any or none)-Val-{Glu or Asp}
  - d. N terminal-Ala-any-[Ser or Thr]-[Ser or Thr]-(any or none)-Val-{Glu or Asp}

Correct answer: b

8. PSSM stands for **(1 mark)**
- a. Position specific scanning matrix
  - b. Point specific scoring matrix
  - c. Position specific scoring matrix
  - d. Protein specific scoring matrix

Correct answer: c

9. Which of the following regular expressions would be matched by sequence MAVDEKC? **(1 mark)**
- a.  $\{MG\}-A-x(2)-[DE]-[ILK]-x$
  - b.  $M-A-x-[DE]-[ILK]-x$
  - c.  $\{HS\}-A-x(2)-[DE]-[ILK]-x$
  - d.  $M-A-x-[DE]-\{LK\}-C$

Correct answer: c

10. Which of the following program uses hidden markov model for predicting the family of a protein? **(1 mark)**
- a. LAMA
  - b. PSI-BLAST
  - c. COACH
  - d. PROSITE

Correct answer: c

11. An invariant column in a multiple sequence alignment will be have \_\_\_\_\_ entropy. **(1 mark)**
- a. Less
  - b. High

Correct answer: a

12. The maximum height of the entire stack of residues in case of protein sequence logo is \_\_\_\_\_ bits. **(2 mark)**

Correct answer: 4.3

13. Which of the following is NOT true for GOR method of secondary structure prediction? **(1 mark)**
- 1) It was developed by Garnier, Osguthorpe and Robson.
  - 2) It is an information theory-based method.

- 3) It considers a window size of 15- one central amino acid and 7 neighboring amino acids on each flanking side.
- 4) It predicts four states of secondary structure- helix, strand, turn and coil.
- a. All the statements are correct
- b. All the statements are wrong
- c. 3
- d. 4
- e. 2 and 3
- f. 1 and 4

Correct answer: c

14. Calculate the contribution of each residue for making the sequence logo using the column information as given below: **(4 mark)**

A  
T  
T  
A  
A  
T  
T  
G  
T  
T

Give the value for: Entropy, Information index and contribution.

Correct answer:

Entropy:  $-(0.3 \cdot \log_2 0.3 + 0.6 \cdot \log_2 0.6 + 0.1 \cdot \log_2 0.1) = 1.298$  (1.5 mark)

Information:  $\log_2 4 - \text{Entropy} = 2 - 1.298 = 0.702$  (1 mark)

Contribution: i)  $A = 0.3 \cdot \text{Information} = 0.3 \cdot 0.702 = 0.2106$  (0.5 mark)

ii)  $T = 0.6 \cdot 0.702 = 0.4212$  (0.5 mark)

iii)  $G = 0.1 \cdot 0.702 = 0.0702$  (0.5 mark)

15. \_\_\_\_\_ algorithm is used to find the most likely sequence of states in an HMM model. **(1 mark)**

- a. Greedy
- b. Viterbi
- c. Genetic
- d. PSSM

Correct answer: b