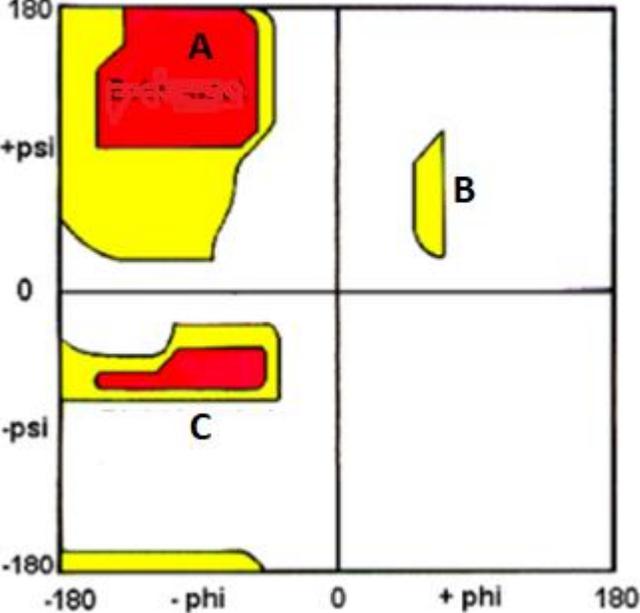# UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2023

Course: Computational Biology and Bioinformatics　　　　**Semester: IV**
Program: B.Sc. Microbiology　　　　　　　　　　　　　　**Time　 : 03 hrs.**
Course Code: HSMB2010　　　　　　　　　　　　　　　　**Max. Marks: 100**

Instructions: Attempt all the questions

Q.No	Section A Short answer questions/ MCQ/True&False	(20Q x 1.5M = 30 Marks)	COs
1.	Application of bioinformatics include a) Data storage and management b) Drug designing c) Understand relationships between organisms d) All of the above	1.5	CO1
2.	Literature database include a) MEDLINE & PubMED b) MEDLINE & PDB c) PubMED & PDB d) MEDLINE & PDS	1.5	CO1
3.	Which of the following database is a gene expression database a) GEO b) MMDB c) DDBJ d) EMBL	1.5	CO1
4.	Which of the following is the replacement of a single amino acid in the primary structure of a protein with another single amino acid, which is accepted by the processes of natural selection. a) PAM b) BLOSUM c) Pairwise d) Multiple	1.5	CO1
5.	What is FASTA file format?	1.5	CO1
6.	Which of the following is an example of Homology and similarity tool? a) BLAST b) RasMol c) EMBOSS d) PROSPECT	1.5	CO2

7.	What is the significance of PAM scoring matrix?	1.5	CO2
8.	What do you mean by unrooted phylogenetic tree?	1.5	CO2
9.	"Both rooted and unrooted trees can be either bifurcating or multifurcating". a) True b) False	1.5	CO2
10.	Which of the following statement is FALSE? a) In bioinformatics, the BLOSUM (BLOcks SUbstitution Matrix) matrix is a substitution matrix used for sequence alignment of proteins. b) BLOSUM matrices are used to score alignments between evolutionarily divergent protein sequences. c) All BLOSUM matrices are based on observed alignments; they are not extrapolated from comparisons of closely related proteins like the PAM Matrices. d) None of the above	1.5	CO2
11.	Proteomics is the study of a) Set of proteins b) Set of proteins in a specific region of the cell c) Entire set of expressed proteins in a cell d) None of these	1.5	CO3
12.	What is the approximate genome size of *Arabidopsis*?	1.5	CO3
13.	State the way in which increasing species diversity can affect the ecosystem a) it does not increase the efficiency and productivity of an ecosystem b) it increases only the efficiency and not the productivity of an ecosystem c) it increases the productivity and efficiency of an ecosystem d) it only increases the productivity of an ecosystem		CO3
14.	ClustalW a) Multiple Sequence alignment tool b) Protein prediction tool c) Data retrieval tool d) Protein sequence analysis tool	1.5	CO3
15.	Define genome.	1.5	CO3
16.	Which of the following does not affect the stability of an α-helix? a) Electrostatic repulsion b) Bulkiness c) Interaction between R groups spaced three residues apart d) Occurrence of alanine and glycine residues	1.5	CO4
17.	Which of the following is not true about secondary protein structure? a) The hydrophilic/hydrophobic character of amino acid residues is important to secondary structure b) The ability of peptide bonds to form intramolecular hydrogen bonds is important to secondary structure	1.5	CO4

	c) The alpha helix, beta pleated sheet and beta turns are examples of protein secondary structure d) The steric influence of amino acid residues is important to secondary structure		
18.	Which of the following is untrue about Ab initio prediction? a) The limited knowledge of protein folding forms the basis of ab initio prediction b) The ab initio prediction method attempts to produce all-atom protein models based on sequence information alone without the aid of known protein structures c) The ab initio prediction method attempts to produce all-atom protein models based on sequence information alone with some aid of known protein structures d) The perceived advantage of this method is that predictions are not restricted by known folds and that novel protein folds can be identified	1.5	CO4
19.	Ramachandran plot can be used to predict which of the following structure? a) Quaternary structure b) Tertiary structure c) Primary structure d) Secondary structure	1.5	CO4
20.	In protein sequence analysis, Twilight zone refers to the evolutionary distance corresponding to about a) 60% identity between two proteins b) 90% identity between two proteins c) 30% identity between two proteins d) 85% identity between two proteins	1.5	CO4
	Section B	**(4Qx5M =** **20 Marks)**	CO
1.	Explain the BLAST tool at NCBI. What is the significance of the E value or expected value resulting from a blast?	5	CO1
2.	Compare pairwise and multiple sequence alignment	5	CO2
3.	Discuss the Anfinsen experiment on RibonucleaseA for determination of 3-D folding of proteins.	5	CO3
4.	The below given figure is Ramachandran plot. In below given diagram, label A, B, and C. Please also mention what red, yellow and white colour represents.	5	CO4

The Ramachandran Plot.

	Section C	(2Qx15M= 30 Marks)	

1.	1. a) Differentiate between genomics, transcriptomics and proteomics.
	  c) When genomics is there then what was the need for proteomics? Why is proteomics considered more complicated than genomics?
	2. a) Why protein structure prediction is important?
	  d) Give full account of structure prediction using ab-initio, fold recognition and homology modelling methods.
		3+3+2+7	CO4

2.	As many other developed countries, Finland is strongly betting to the opportunities generated by the availability of huge amounts of data and gene-based collections. Big data is expected to actively participate in the enhancement of medical research and the consequent generation of economic wealth, in the path toward the total access of the population to health and wellbeing. The imperative relies on the usage of massive genomic data justified by the moral principle of improving health (Snell, 2019). Such imperative of health also establishes the legal regulation of data collection, infrastructure development and genomic innovation. Finnish data-driven medicine is based on the moral principle of health, with policies related to privacy and autonomy subjected to this national goal. On that note, the highly restrictive and conscious government regulatory policies of data privacy and security become the major challenge for leveraging the entire potential of big data analytics in the country. However, Finland has found a potential solution through the institutionalisation of automatic systemic data anonymisation policies in order to make healthcare data safe from cyber-attacks and accessible to life scientists (Taiwo, 2019). At the opposite corner, violating the ethical consensus of scientists all over the world, a team of Chinese scientists led by He Jiankui. He presented during the second World Summit of Human Gene Editing in early 2019, the gene-editing	6 + 6 + 3	CO3

project that led to the birth of two baby girls with man-made C-C chemokine receptor type 5 (CCR5) mutations. This research was strongly repudiated by the worldwide scientific community as an extremely irresponsible behaviour that exhibits a lack of both medical ethics and understanding of gene editing science. The gene editing on a human germline constitutes a severe violation of both the Chinese regulations and the consensus reached by the international science community (Wang and Yang, 2019). How- ever, this incident triggered a wide-ranging discussion about the criteria and standards for genome editing in the human germline for reproductive purposes, attempting to reach a solid consensus about clear and strict policies at an international level.

Based on this case study, answer the following:
a) How has genomics changed medicine? Discuss your views on genomic research, particularly genome sequencing which has dramatically been translated into clinical practice.
b) Here, two extreme case studies are unveiled: Finland and China. Discuss how countries are evolving their medical systems through genomic medicine approaches. What are the challenges and concerns associated with genomic research
c) What are the various ethical issues in human genetics and genomics?

	Section D	(2Qx10M= 20 Marks)	
1.	a) What is a phylogenetic tree? b) Explain different approaches of phylogenetic tree construction (Hint: UPGMA, Neighbor-joining, Maximum Parsimony, Maximum likelihood)	2 + 8	CO2
2.	Write the principle of mass spectrometry? Define molecular ions and fragment ions. Explain MALDI in detail.	2 + 3+5	CO3