Name: Enrolment No:	

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2022

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/T&F	(20Qx1.5M= 30 Marks)	COs
1.	Define relational database with example.	1.5	CO1
2.	Literature database include a) MEDLINE & PubMED b) MEDLINE & PDB c) PubMED & PDB d) MEDLINE & PDS	1.5	CO1
3.	For a protein search, which of the following is the wrong choice for fields? a) Accession number b) EC number c) Issue d) Journal number	1.5	CO1
4.	_____ 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 FTP mode of data transfer?	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.	_____ indicates the hypothetical common ancestor, or ancestral lineage, of the tree. A. unrooted tree diagram B. rooted tree diagram C. Both A and B a) D. None of the above	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 size of *E.coli* genome?	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.	What is the principle of 2D gel electrophoresis?	1.5	CO3
15.	What is MALDI TOF used for?	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 c) The alpha helix, beta pleated sheet and beta turns are examples of protein secondary structure	1.5	CO4

	d) The steric influence of amino acid residues is important to secondary structure		
18.	Which of the following is also known as dihedral angles? a) Right angles b) Obtuse angles c) Acute angles d) Torsion angles	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.	Left-handed alpha-helix allowed region is present in which of the following quadrants of Ramachandran plot? a) Fourth quadrant b) Third quadrant c) Second quadrant d) First quadrant	1.5	CO4
	Section B	**(4Qx5M=20 Marks)**	CO
1.	What type of database is PDB? How is PDB used in bioinformatics?	5	CO1
2.	Discuss local and global sequence alignment with the help of example.	5	CO2
3.	Write a short note on diversity of genomes	5	CO3
4.	Explain energy minimizations and evaluation by Ramachandran plot. How this plot is involved in validation of protein structure?	5	CO4
	Section C	**(2Qx15M=30 Marks)**	
1.	The World Health Organization estimates that some 170 million people worldwide suffer from diabetes mellitus. Although only a minority of these suffer from insulin-dependant (type 1) diabetes, world demand for purified insulin approached 5 metric tons in 2000 and is continually increasing. Healthy humans secrete insulin continuously at a low basal level, with rapid but transient increases triggered by elevated blood glucose concentrations. A combination of fast acting and slow acting insulin thus must usually be administered to diabetics to mimic the natural state. Insulin consists of a 21-amino acid A chain linked to a 30-amino acid B chain via two interchain disulfide linkages. At physiological concentrations (10^{-10} M), insulin molecules exist in monomeric form. However, when present at therapeutic dose concentrations typical of commercial products ($\sim 10^{-3}$ M), individual	3 marks each	CO4

	insulin molecules dimerize, with subsequent oligomerization of three dimers to form a hexamer, often coordinated with zinc ions. When administered via subcutaneous or intramuscular injection, individual insulin molecules must first disassociate from one another before leaking from the site of injection into the bloodstream. As a practical consequence, even such fast acting therapeutic insulins must be administered 30–45 min before a planned meal, and the subsequent mealtime must not be altered or the diabetic risks hypoglycemia. Faster acting insulins that could be administered concurrently with a meal would afford far greater flexibility to diabetics, and the development of such products was made possible by protein engineering. The principle amino acids contributing to dimer formation reside in the B chain at positions B 8, 9, 12, 13, 16, and particularly 23–28. The main engineering strategies centered around introducing amino acid substitutions that will discourage monomer interaction in a subset of these positions believed to lie outside the insulin receptor binding site (specifically B 9, 12, and 26–28). Novorapid and Novolog (**www.novolog.com**) are trade names of one such engineered product in which the proline at position B 28 has been replaced with an aspartate residue, thereby introducing charge repulsion at the monomer-monomer surface. The product shows greatly decreased propensity for oligomerization and, as a practical consequence, can be administered directly prior to or during a meal. **Based on the above case study, please answer the following:** a) Discuss briefly the structure of human insulin b) Justify the statement "the administered time of fast acting therapeutic insulin is about 30-45 min before a planned meal". c) What approaches could be used to discourage dimer formation? d) Can you see any potential therapeutic disadvantages/complications potentially caused by such engineering? e) If provided in the laboratory with (unlabeled) samples of both native insulin and the engineered fast acting insulin, can you think of any potential experiments you could undertake to figure out which is naïve and which is fast acting engineered insulin?		
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	6 + 6 + 3	CO3

	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 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) Discuss various approaches of building phylogenetic trees (Hint: UPGMA, Neighbor joining, Maximum Parsimony, Maximum likelihood)	10	CO2
2.	a) What is MALDI TOF spectroscopy? Explain its working principle. b) Discuss the applications of MALDI-TOF mass spectrometry in clinical diagnosis	5+5	CO3