

Indian Institute of Technology (IIT) Delhi
Department of Biochemical Engineering and Biotechnology

II SEMESTER 2008 – 2009
BEL 418 – BIOINFORMATICS

Date	May 08, 2009
------	--------------

MAJOR TEST

Time = 2 hours

Answer all the questions

Max. Marks = 35

Question #	Question	Marks
1	<p>Study the GenBank feature table given in the Figure below and answer the following questions:</p> <p><u>Figure - Genbank Feature table</u></p> <pre> FEATURES Location/Qualifiers source 1..201523 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="6" /map="6p21.3" gene 501..18455 /gene="C2" /note="synonym: CO2" /db_xref="GeneID:717" /db_xref="LocusID:717" /db_xref="MIM:217000" mRNA join(501..582, 739..948, 1516..1701, 6394..6567, 6651..6749, 6950..7083, 8707..8845, 10103..10243, 2015..12104, 15743..15883, 16009..16103, 16200..16311, 16428..16593, 16701..16777, 16919..17010, 17511..17637, 17764..17813, 17962..18455) /gene="C2" /product="complement component 2" /note="unclassified transcription discrepancy" /transcript_id="NM_000063.3" /db_xref="GI:20631970" /db_xref="GeneID:717" /db_xref="LocusID:717" /db_xref="MIM:217000" 2 CDS join(537..582, 739..948, 1516..1701, 6394..6567, 6651..6749, 6950..7083, 8707..8845, 10103..10243, 12015..12104, 15743..15883, 16009..16103, 16200..16311, 16428..16593, 16701..16777, 16919..17010, 17511..17637, 17764..17813, 17962..18141) /gene="C2" /codon_start=1 /product="complement component 2" /protein_id="NP_000054.2" /db_xref="GI:14550407" /db_xref="GeneID:717" /db_xref="LocusID:717" /db_xref="MIM:217000" </pre>	<p>(5)</p>

	a) Name the gene described in the feature along with its chromosomal map location b) Identify the gene length in nucleotide units for this entry. c) Calculate the number of exons and introns in the gene d) Calculate the length of the third intron in nucleotide unit. e) What are the GI numbers used in the feature?	
2	a) What is the basic difference between Needleman-Wunsch and Smith-Waterman algorithm? b) Briefly explain the major differences between BLAST method and PSI-BLAST method.	(3 + 3 = 6)
3	In the laboratory, a large part of the genome is sequenced from eukaryote. Write steps to identify genes and their probable functions. Use available software tools and databases from open sources. Assume the availability of all resources from NCBI servers.	(4)
4	Structure-based drug design utilizes knowledge of a known 3-D structure for a biomolecular target to design lead molecules. Outline the steps involved in computer-aided drug design starting with input information of target structure and resulting in a drug.	(3)
5	Homology and Similarity are not two similar terms. Explain. What is a binary tree? Consider 4 species characterized by homologous sequences ATCC, ATGC, TTCG, TCGG. Derive the phylogenetic tree using the simple UPGMA method.	(1+1+4 = 6)
6	What is PSSM? How it is derived? What is its utility?	(2 + 2 + 2 = 6)
7	a) What does an E-value of 2×10^{-10} in BLAST search mean? Describe the two parameters on which the E-value depends. b) State the strength and weakness of similarity based gene-finding algorithm. c) Which one of the following pattern will match with a given PHI-BLAST query pattern A-[LIVM]-x(3)-G-[DE]. Pick one from (i) to (iii) below - (i) ADLVCGE (ii) AVSTTGE (iii) AMSTGDE (iv) ALAPSGLD	(1 + 1 = 2) (1 + 1 = 2) (1)