

Ref. No. : Ex/PG/BME/T/1210B/2022

Name of the Examinations: MASTER OF BIO-MEDICAL ENGINEERING FIRST YEAR SECOND SEMESTER - 2022

Subject : BIO-INFORMATICS

Time : 3Hrs

Full Marks : 100

**Instructions: Answer any twenty questions**

1. What are the types of databases? Illustrate with examples from biological databases. What is the difference between identifier and accession code? 2+1+2
2. What are the differences between global and local alignment methods? Are the beta and alpha chains of the hemoglobin molecule homologous, paralogous and /or orthologous to each other and why? Which algorithm method is used for local alignment by dynamic programming? 2+2+1
3. Why do you need multiple sequence alignments? Can they be done globally? What are conserved sequences? 2+2+1
4. Describe various methods to achieve MSA. What are the advantages of MSA over pair-wise comparisons? 3+2
5. Define BLAST. Which algorithm is used during BLAST search? What is a E-value? 2+1+2
6. Differentiate between a Markov Model and a Hidden Markov Model (HMM). What are the various applications of HMMs in bioinformatics? 2+3
7. Define a gene and compare the genomes of prokaryotes and eukaryotes. What is the purpose of gene prediction? 1+2+2
8. How can you find open reading frames (ORFs) in a given DNA sequence? Explain the gene prediction method by homology-based approaches. 3+2
9. Describe the RNA secondary structure in detail. What is the difference between minimization of energy approach and maximization of base pair approach? 3+2
10. What is DNA microarray technology? Define normalization step in microarray data analysis. Describe the clustering methods in microarray data analysis. 1+2+2
11. Which tool is recommended for obtaining information about the secondary structure of protein? What is Anfinsen's hypothesis? What properties you are likely to use for primary structure prediction of protein? 2+1+2
12. What are the prediction methods for secondary structure of protein? Explain any one method. 3+2
13. What is EST? Describe its role in genome analysis. Compare EMBL, DDBJ and GenBank. 2+3
14. What is database? What are the different types of protein databases? Differentiate between them. 1+2+2
15. From the sample entry from NCBI for a protein answer the following questions:

[ Turn over

GenPept ▾

**nsp4 [Severe acute respiratory syndrome coronavirus 2]**

NCBI Reference Sequence: YP\_009742611.1  
[Identical Proteins](#) [FASTA](#) [Graphics](#)

---

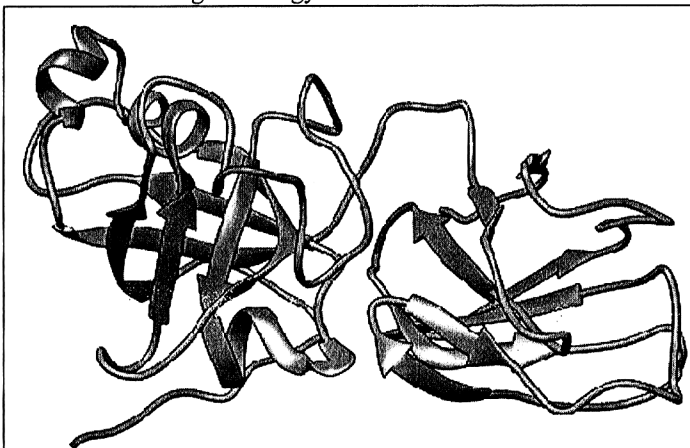
Go to:

LOCUS YP\_009742611 500 aa linear VRL 18-JUL-2020  
 DEFINITION nsp4 [Severe acute respiratory syndrome coronavirus 2].  
 ACCESSION YP\_009742611  
 VERSION YP\_009742611.1  
 DBLINK BioProject: [PRJNA485481](#)  
 DBSOURCE REFSEQ: accession [YP\\_009725295.1](#)  
 KEYWORDS RefSeq.  
 SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)  
 ORGANISM Severe acute respiratory syndrome coronavirus 2  
 Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;  
 Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;  
 Betacoronavirus; Sarbecovirus.  
 REFERENCE 1 (residues 1 to 500)  
 AUTHORS Wu,F., Zhao,S., Yu,B., Chen,Y.M., Wang,W., Song,Z.G., Hu,Y.,  
 Tao,Z.W., Tian,J.H., Pei,Y.Y., Yuan,M.L., Zhang,Y.L., Dai,F.H.,  
 Liu,Y., Wang,Q.M., Zheng,J.J., Xu,L., Holmes,E.C. and Zhang,Y.Z.  
 TITLE A new coronavirus associated with human respiratory disease in  
 China  
 JOURNAL Nature 579 (7798), 265-269 (2020)

- How many amino acids are in the protein?
- In which organism this protein present?
- What is the Accession ID for this protein?

2+2+1

16. In the following homology modeled structure



- What types of secondary structures are present?
- How many numbers of different types of secondary structures are present in this structure? 2+3

17. The BLAST search result for nsp4 protein of SARS CoV2 and MERS CoV shows the following results

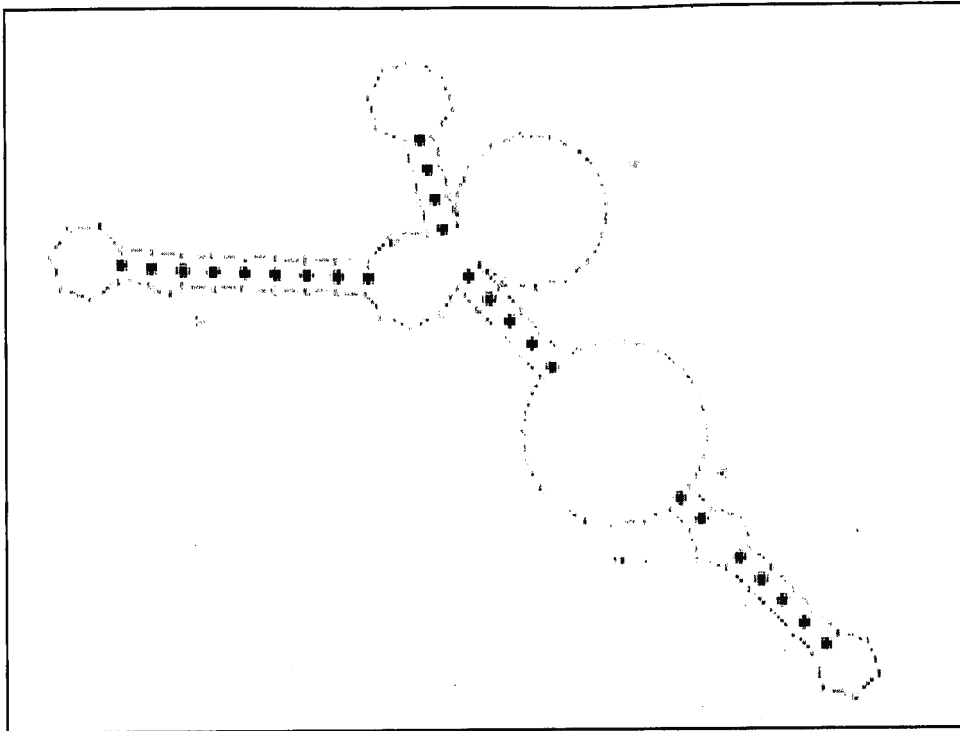
**nsp4 protein [Middle East respiratory syndrome-related coronavirus]**  
 Sequence ID: [YP\\_009047216.1](#) Length: 507 Number of Matches: 1  
[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 13 to 507 [GenPept](#) [Graphics](#)

Score	Expect	Method	Identities	Positives	Gaps		
397 bits(1020)	2e-134	Compositional matrix adjust.	199/498(40%)	298/498(59%)	7/498(1%)		
Query 7	LKQLIKVTLVFLFVAAI	FYLITPVHMSKHTD	FSSEIIGYKAIDGG	VTRDIAS	TDTCFAN 66		
Sbjct 13	LKGYVLATIIIVFLCA	VLYLCLPTFSMAP	VEFYEDRIIDFKV	LNDNGIIRDV	NPDDKCFAN 72		
Query 67	KHADFDTWFSQR-GGS	YTNDAKACPLIAA	VITREVGFWVGLP	GTILRTTNGD	FLHFLPRV 125		
Sbjct 73	KHRSFTQNYHEHVGG	VDNSITCPLTVA	VIAGVAGARIPD	VPTTLAHWNNQ	IIFVSRV 131		
Query 126	FSAVGNICYTPSKL	IEYTDFAVSACV	LAAECTIFKDA	SGKPVPCYD	TNVLEGSVAYESL 185		
Sbjct 132	FANTGSVCYTPIDE	IPYKSFSDSGC	ILPSECTMFRD	AEGRMPYCHD	PTVLPGAFAYSQM 191		
Query 186	RPDTRYVLMDSGI-	IQFPNTYLEGS	VRVVTTFDSE	YCRHGTCE	SEAGVCVSTSGR	WVLN 244	
Sbjct 192	RPHVRYDLYDGM	FIKFPVVFEST	LRITRTLSTQ	YCRFSGSCE	YAEQEGVCIT	THGSHWIF 251	
Query 245	NDYYRSLPGVFC	GVDVAVNLLT	NMFTPLIOP	IGALDisasi	vaggiavivTCL	AYYFMRF 304	
Sbjct 252	NDHHLNRPGV	YCGSDFIDIV	RRLAVSLFQ	PITYFQLT	LSLVLGIGL	CAFLTLFFYYINKV 311	
Query 305	RRAFGEYSHVAF	NLTLFLNSF	TVLCLTPV	YSLPGVYS	VIYLYLTFYLT	NDVSLAHIQ 364	
Sbjct 312	KRAFADYQCA	AVIAVVAV	LNSLCICF	VTSIPLC	IVPYTALYY	ATFYFTNEPA	FINHVS 371
Query 365	KNMFMFPLVP	FWITIIAYI	ICISTKHF	YHFFSNYL	KRRV-VF-	NGVSFSTF	EAAALCTFLL 422
Sbjct 372	WYIMFGPIV	PIWMTCVY	TVAACFRH	FFWVLA	YFSSKHHVE	VFTDGLNCS	FQDAASNIFVI 431
Query 423	NKEMYLKLRSD	VLLPLTQYN	RYLALYN	KYKYGAM	DTTSYREA	ACCHLAKAL	NDFSNSG 482
Sbjct 432	NKDTYAALRNS	--LTNDAYS	RFLGLFN	KYKYGAM	EAAAYREA	AAACHLAKAL	QTYSETG 489
Query 483	SDVLYQPPQTS	SITS	AVLQ 500				
Sbjct 490	SDLLYQPPNCS	SITSGVLQ 507					

a) What is the significance of values shown here as Identities, Positives and Gaps?  
 18. In the secondary structure of a mRNA is shown below

2+2+1



How many (a) hairpin-loop (b) interior loop and (c) multiloop (junction) are present in the above structure?

2+2+1

19. How many base pairs are present in human genome? Why human genome sequencing is important for mankind?

1+4

20. What is DNA sequencing? Write down the principle of dideoxy sequencing by Sanger method.

2+3

21. The gene search result for p53 gene in human genome shows the following:

**TP53 tumor protein p53 [ *Homo sapiens* (human) ]**

Gene ID: 7157, updated on 26-Jun-2022

**Summary**

**Official Symbol** TP53

**Official Full Name** tumor protein p53

**Primary source** [HGNC:HGNC:11998](#)

**Gene type** Protein coding

REVIEWED

[Homo sapiens](#)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae;

Homo

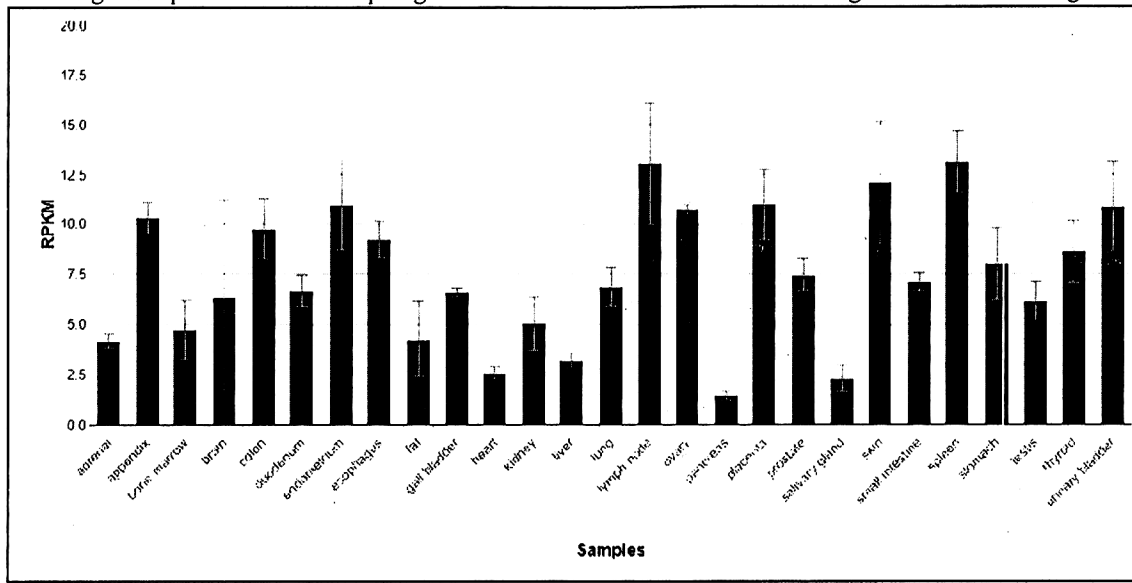
This gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. Mutations in this

gene are associated with a variety of human cancers, including hereditary cancers such as Li-Fraumeni syndrome. Alternative splicing of this gene and the use of alternate promoters result in multiple transcript variants and isoforms. Additional isoforms have also been shown to result from the use of alternate translation initiation codons from identical transcript variants (PMIDs: 12032546, 20937277). [provided by RefSeq, Dec 2016]

(a) What is the function of the gene?

(b) In which disease this gene is related? 4+1

22. The gene expression details of p53 gene from different tissues of human being shows the following results:



(a) The gene expression levels are the highest and the lowest in which tissues?

(b) The gene expression value is 5 RPKM in which tissue? 4+1