ENTRANCE EXAMINATION - 2020

Ph.D. Systems and Computational Biology

HALL TICKET NUMBER

Maximum Marks: 70

Time: 2 hours

INSTRUCTIONS: Please read the instructions carefully before answering the questions

1. Answers are to be marked on the OMR answer sheet.
2. Hand over the OMR answer sheet at the end of the examination to the invigilator.
3. The question paper contains 70 questions of multiple choices, printed in 14 pages (last three pages to be used for rough work), including this page.
4. OMR answer sheet provided separately.
5. All questions carry one mark each.
6. In case the candidates have equal marks, preference will be given towards the candidate who has obtained higher marks in Part-A.
7. **There is negative marking for wrong answer. Each wrong answer carries -0.33 mark.**
8. Non-programmable scientific calculators are permitted.
9. Cell,Mobile Phones are strictly prohibited in the examination hall.

PART A

1. Rotation about N-Ca bond in a peptide/protein is referred to as ----

   A. χ
   B. ϕ
   C. θ
   D. ψ

2. Collagen is rich in _____ amino acid residues

   A. Hydroxy-Pro and Gly
   B. Pro and Gly
   C. Cys and Gly
   D. Leu and Gly

3. Entrez is _____

   A. a global query cross-database search system available at NCBI Website
   B. a relational database of protein sequences available at NCBI website
   C. an information repository of protein structures available at NCBI website
   D. a query system connected to EMBL database

4. Spearman correlation coefficient is to determine how

   A. two random variables correlated to each other and the value can lie between -1 to +1
   B. two random variables are correlated to each other and the value can lie between 0 to +1
   C. the rank orders of two random variables are correlated to each other and the value lies between -∞ to +∞
   D. the rank orders of two random variables are correlated to each other and the value lies between -1 to +1

5. Speed of light in vacuum is equal to

   A. 300,000 km.sec-1
   B. 30,000 km.sec-1
   C. 300,000 mt.sec-1
   D. 3,000,000 km.sec-1

6. Consider two state variables X and Y, which are dynamically undergoing changes w.r.t. time t. In the X-Y Phase Plane, the X nullcline is that loci of (X,Y) points at

   A. $\frac{d(Y)}{dt} = 0$

   B. $\frac{d(X)}{dt} = 0$

   C. $\frac{d(Y)}{dX} = 0$

   D. $\frac{d(X)}{dY} = 0$

7. Isobaric tags in mass spectrometry are the peptides

   A. with identical composition
   B. of equal charge
   C. of equal mass
   D. at same pressure

8. The key characteristic feature of High Performance Computing (HPC) platform is

   A. multiple compute nodes
   B. multiple data threads
   C. online access and scalability
   D. multiple storage devices

9. A trial involving flipping of a (unbiased) coin for n times, the probability of finding x number of heads is given by:

   A. Hypergeometric distribution
   B. Binomial distribution
   C. Gumbel distribution
   D. Gamma distribution

10. Which one is a continuous periodic function?

   A. $f(x) = \sin x$
   B. $f(x) = \csc x$
   C. $f(x) = \tan x$
   D. $f(x) = \cot x$

11. A formula given by

   $$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$$ where, X_i is i^{th} random variable in a sample size of n and $\bar{X}$ is the mean. This (formula) is a measure of

   A. Covariance
   B. Correlation
   C. Sample standard deviation
   D. Sample variance

12. Consider two events A and B with non-zero probabilities, the probability P (B|A) is given by following formula
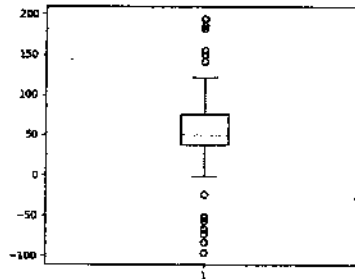
   $$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

   the term P(B|A) is known as

   A. Posterior probability
   B. Prior probability
   C. Conditional probability
   D. Joint probability

2

13. Chitin is linear polymer of?

    A. N-acetylfucosamine
    B. N-acetylgalactosamine
    C. N-acetylmannosamine .
    D. **N-acetylglucosamine**

14. In the box plot given below the outliers are those data points found
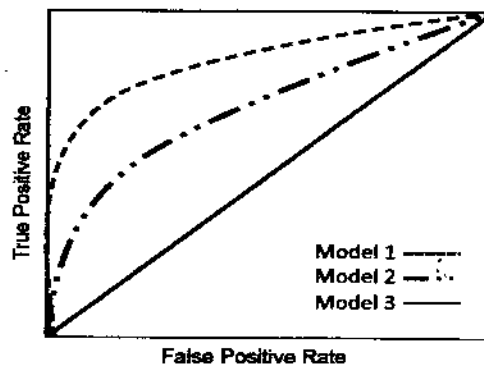

    i.   between the two whiskers
    ii.   below the bottom whisker
    iii. above the upper whisker

    Which of the above form the most appropriate answer?

    A. i and ii
    B. i and iii
    C. ii and iii
    D. Only i

15. The receiver operative characteristics (ROC) curves for three models are shown below.

    Which one of these statements is CORRECT?

    A. Model 2 is better than Model 1 and Model 3
    B. Model 3 is as good as a random guess
    C. Model 3 is better than Model 1 and Model 2
    D. Model 1 is better than Model 2 and Model 3

16. The angle swept when the length of the arc becomes equal to two times its radius is equal to

   A. Two radians
   B. Two degrees
   C. Two grads
   D. Two seconds

17. In a graph (X-Y plot) a parabola is drawn such that its vertex crosses the y-axis at (0,2) and is opened up. The quadratic function that represents this graph is

   A. $y=2x^2+2$
   B. $y=x^2+2$
   C. $y=2x^2-2$
   D. $y=x^2-2$

18. Your class strength is 150. A test was conducted for 100 marks and it was found that your score is 80 and you are at the 80th percentile level. How many students have scored less than you? (Please note: Every student has scored a different score).

   A. 80
   B. 120
   C. 100
   D. 90

19. Considering that one-week has 10 days, how many days are there in k weeks and k days?

   A. $10k^2$
   B. $10(k+1)$
   C. $10k+1$
   D. $11k$

20. There are 3 green, 4 orange and 5 white color balls in a box. If a ball is picked at random, what is the probability of having either a green or a white ball?

   A. 0.33
   B. 0.66
   C. 0.22
   D. 0.44

21. If all the six cysteine residues in a protein take part in making disulfide bond then the possible number of ways the three S-S bonds can be formed is

   A. $^6C_2$
   B. $^6P_2$
   C. 6!
   D. 3!

4

22. Cell 1 has radius of 1 μm and Cell 2 has radius of 2 μm. Assuming that both the cells are round shaped, the ratio of their surface areas to their volumes, and the cell that can support greater nutrient exchange, are

A. 3:1.5, cell 2
B. 3:4, cell 2
C. 3:1.5, cell 1
D. 3:4, cell 1

23. The molecular formula of imatinib is $C_{29}H_{31}N_7O \cdot CH_4SO_3$ and its molecular weight is 589.7. In order to prepare 10mM stock solution you will need

A. 58.97mg in 1ml
B. 5.897mg in 1ml
C. 5.897mg in 10ml
D. 0.5897mg in 10ml

24. Molecule X is hydrophobic and has a formula of $C_{28}H_{22} F_3N_7O$ and its molecular weight is 529.52. A company supplies the molecule X in powder form of 5mg. In order to prepare 5mM stock solution you will need

A. 3.78 ml of DMSO
B. 1.89 ml of water
C. 1.89 ml of DMSO
D. 3.78 ml of water

25. In a normal distribution, value of *Z-score* for a given random variable X was found to be 0. Which of the following statements is *CORRECT* about X?

A. X is same as the mean.
B. X lies beyond ±3times of standard deviation
C. X is same as standard deviation
D. X is same as dispersion

. 26. It has been reported that transcript and protein abundances of a large number of genes of a genome are correlated. The average correlation coefficient value would likely to be

A. 0.25 to 0.75
B. 0.0
C. -0.25 to -0.75
D. 1.0

27. A whole genome re-sequencing study of fruit-fly (size: 175 Mb) involved generation of 3.5 million reads of length 500 bp. The coverage (number of times the fruit-fly genome is covered) would be

A. 175
B. 100
C. 500
D. 10

28. Given a task to translate large number of coding sequences (CDS) into peptide sequences using codon table (having information about all codons and the corresponding amino acids

5

they encode), which data type/structure would be most suited for storing a codon table in python or perl?

A. List
B. Array
C. Hash/Dictionary
D. Scalar variable

29. The function $y = f(x)$ with an initial value of 0 (at $x=0$) having its roots at $x = n\pi$ $(n \varepsilon N_1)$ having the amplitude = 2 units is

A. $2y = \sin(x)$
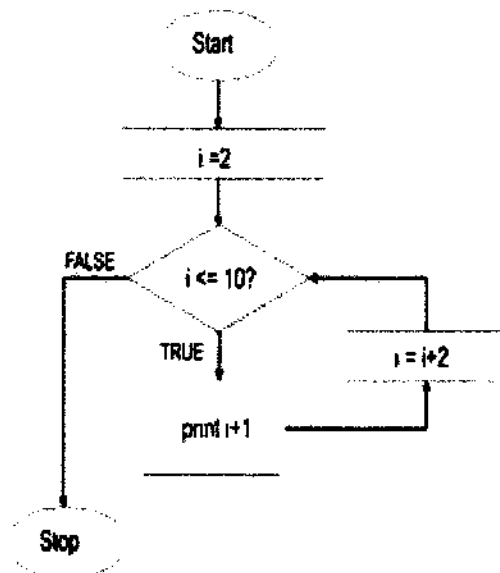B. $y = 2\sin(x) + 2$
C. $y = \sin(2x)$
D. $y = 2\sin(x)$

30. For a given array, bubble sort is used to sort the array elements.
        Arr = {6, 4,1, 2, 5}
How many iterations are needed to sort the array?

A. 4
B. 2
C. 1
D. 0

31. What will be the output for the below given algorithm represented as flowchart

A. 3, 6, 8, 10,12
B. 3, 5, 7, 9,11
C. 3, 4, 6, 8,10
D. 3, 5, 8, 9,11

6

32. For the antibiotic ampicillin, approximately 40% of the drug is eliminated every hour. Assuming that a typical dose of ampicillin for kids is 125 mg, the amount of drug left in body after 6 hours is given by following exponential function

A. $125 * (0.4)^6$
B. $125 * 0.4 * 6$
C. $125 * (0.6)^6$
D. $125 * 0.6 * 6$

33. In a C/perl/python-like pseudocode, a computer program for translation of a coding sequence into amino acid sequence has following condition to stop the translation ·

A. if ((codon == 'UAA') || (codon == 'UAG') || (codon == 'UGA')) {#stop translation}
B. if ((codon == 'UAA') && (codon == 'UAG') && (codon == 'UGA')) {#stop translation}
C. if ((codon = 'UAA') || (codon = 'UAG') || (codon = 'UGA')) {#stop translation }
D. if ((codon = 'UAA') && (codon = 'UAG') && (codon = 'UGA')) {#stop translation }

34. In a sequencing based gene expression experiment, the read counts for gene X (size 2 Kb) and gene Y (size 1 kb) are exactly same, i.e., 800. One can infer that

A. both X and Y have same expression level
B. expression level of X is higher than Y
C. expression level of Y is higher than X
D. data is insufficient to infer anything

35. In order to discover genes whose activity changed after a particular treatment, transcriptome profiling of control and treated samples was done. One of the genes had -2 as $\log_2$ fold change (relative to control sample), which indicates that

A. transcript abundance of control sample was -2
B. transcript abundance of treated sample was one-fourth of control sample
C. transcript abundance of control sample was 2 less than treated sample
D. transcript abundance of control sample was one-fourth of treated sample

PART B

36. α-helix is characterized by the presence of hydrogen bonds of the type

A. $NH_{i+2} \rightarrow O_i$
B. $NH_{i+1} \rightarrow O_i$
C. $NH_{i+3} \rightarrow O_i$
D. $NH_{i+4} \rightarrow O_i$

37. PSI-BLAST stands for

A. Pattern Specific Interactive BLAST
B. Position Specific Iterative BLAST
C. Partial Score Interactive BLAST
D. Position Scoring Iterative BLAST

38. Which one of the following is *CORRECT* for dynamic programming based local alignment

    A. Matrix value is set to zero for negative values
    B. Matrix values can be negative, positive or zero
    C. Alignment always begins from lower right and end at top left cell
    D. First row and first column are set as gap penalty

39. Match the following

Column A	Column B
a) Uniprot/Swissprot	i) Genetic information of human diseases
b) OMIM	ii) Primary/secondary database of proteins
c) DIP	iii) 3D bio-molecular structure database
d) PDB	iv) Protein-protein interaction database

    A. a - iv, b – iii, c – ii, d – I
    B. a – ii, b – i, c – iv, d – iii
    C. a – ii, b – iii, c- I, d – iv
    D. a – iv, b – iii, c – ii, d - i

40. Most of the miRNA are known to regulate expression of target genes by binding to __ region

    A. 5' UTR
    B. ORF
    C. 3' UTR
    D. polyA

41. The steps involved in Basic Local Alignment Search Tool (BLAST) are given below. Arrange the steps in order:

    i) Reporting high-scoring segment pairs
    ii) Pairwise alignment by extending from the words in both directions
    iii) Creating a list of words from the query sequence
   iv) Searching a sequence database for the occurrence of words

    A. iii → i → iv → ii
    B. iii → iv → i → ii
    C. iii → ii → iv → i
    D. iii → iv → ii → i

42. Microarray, used for profiling genome-wide gene expression, is based on:

    A. intensity of fluorescence of labeled cDNA hybridized with oligomer probes
    B. intensity of binding of mRNA with different fluorescent dyes
    C. concentration of florescent dye labeled mRNA in microwells
   D. concentration of florescent dye labeled oligomer probes in microwells

43. Genetic map of a genome, which helps in assembly of genome, can be obtained by

A. information of chromosome bands by Giemsa staining
B. relative position of restriction sites detected by digestion by combination of restriction endonucleases
C. frequency of recombinants from a dihybrid or trihybrid cross
D. position of unique sequences identified by fluorescent in-situ hybridization (FISH)

44. In order to profile composition of a microbial community, (amplicon) sequencing of marker gene is carried out. The choice of marker genes for bacteria is

A. DNA polymerase
B. 16s rRNA
C. Inter-transcribed spacer (ITS)
D. 18s rRNA

45. Which of the cyclin and cyclin dependent kinase (CDK) is required for G2 to M transition?

A. Cyclin D/ CDK4
B. Cyclin E/ CDK4
C. Cyclin B/ CDK1
D. Cyclin A/ CDK2

46. The leader sequence present in the Histidine operon plays a major role in the attenuation of transcription in the absence of Histidine. In this leader sequence how many codons are present for Histidine?

A. 0
B. 2
C. 4
D. 7

47. Parkinson's disease is due to deficiency in production of
A. GABA
B. Acetylcholine
C. Serotonin
D. Dopamine

48. In a family of proteins some positions were found to be absolutely conserved and some others with minimal changes/conservative changes. Using this observation following statements were made by different people.

a) These regions mostly correspond to conserved secondary structures
b) These regions are only of structural importance and of no functional significance
c) Any changes in such regions are deleterious
d) Disease-causing mutations are expected to be found in such regions

Which of these statements are correct? Pick the most appropriate answer.
A. a, c and d
B. a, b, c and d
C. b, c and d
D. a and d

49. What do the cross lines in the dot plot given below represent when a sequence is compared with itself?
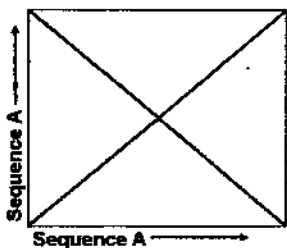
A. Inversion
B. Palindrome
C. Indel
D. Translocation

50. Based on the given phylogenetic tree, fish is most closely related to
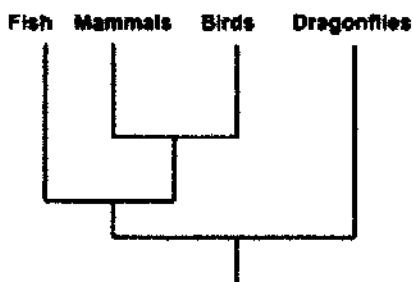
A. Birds
B. Mammals
C. Mammals and Birds
D. Dragonflies

51. Classic gene regulatory models have regulatory proteins such as, Lactose repressor (lacI) and bacteriophage lambda repressor (cI), as key regulators in their respective operons. In the last two decades, such regulatory sequences have been used to design novel genetic switches such as, toggle switch in E. coli. Study of such aspects are covered under an emerging discipline named

A. Synthetic biology
B. Mathematical biology
C. Computational biology
D. Regulatory biology

52. The point mutations, generally random in nature, are subjected to varying selection pressure(s). The eukaryotic genomic regions arranged in order of increasing mutation rate will be

A. non-degenerate sites of coding region < untranslated region (UTR) < intron
B. intron < untranslated region (UTR) < non-degenerate sites of coding region

C. intron < non-degenerate sites of coding region < untranslated region (UTR)

D. untranslated region (UTR) < non-degenerate sites of coding region < intron

53. A novel protein has been isolated from amphibian oocytes and BLAST sequence analysis results suggested that particular protein has leucine zipper motif in N-terminal of the protein. In order to identify the specific sequence of DNA bound by this particular protein which of the following methods is used?

A. Mobility Shift Assay Microarray
B. Luciferase assay
C. Chromatin Immunoprecipitation
D. Microarray

54. Which mutation *CANNOT* occur in a tRNA encoding gene?

A. Transversion
B. Nonsense mutation
C. Transition
D. Deletion

55. Western blot data suggests a mutant protein is ten times more than the wild type protein. In order to show that mutant version of the protein is stabilized by post-translational modification and not due to transcription, which of the following will confirm?

A. Southern blotting
B. Western blotting
C. Quantitative PCR
D. MALDI-TOF

56. Structural biologists validate quality of protein structures by computing its Ramachandran Map which is a graphical representation of fraction of amino acid residues

A. Having no contacts
B. Having bad geometries (bond lengths and bond angles)
C. Found in "allowed" and "disallowed" regions
D. In unconducive environments such as polar residues surrounded by apolar residues

57. A protein with 80% of its amino acid residues found in α-helices and β-sheets and the two secondary structures are interspersed along the sequence, then this protein's structural class is said to be

A. β -class
B. α -class
C. α+β class
D. α /β class

58. You did a BLAST search of a nucleotide sequence and it gave hits only for exons. The query is a
A. Genomic sequence of the gene
B. Transcribed sequence of the gene
C. Regulatory sequence of the gene
D. Non-coding sequence of the gene

59. A novel protein X is extracted from the homogenate of muscle cells and *in vivo* biochemical experiment suggests that protein X has kinase activity and is able to phosphorylate protein Y either directly or indirectly. In order to check whether protein X can directly phosphorylate protein Y, an *in vitro* kinase assay is performed with the help of purified protein X, Y and radio labeled ATP. In order to label the phosphorylation of substrate protein Y which position of ATP is labeled with P^{32}?

   A. Beta P^{32} ATP
   B. Alpha P^{32} ATP
   C. Gamma P^{32} ATP
   D. B and C

60. Protein X is a cell membrane protein acts as a cytokine receptor and activates signaling molecules after binding to a cytokine, however due to oncogenic mutation the mutant version of the protein is located in endoplasmic reticulum instead of cell surface and triggers oncogenic signaling pathway. In order to show that ER localization can trigger the oncogenic pathway which of the following experiments can be done to prove it?

   A. Addition of ER signal sequence to the wild type protein and measure oncogenic signaling
   B. Addition of ER retention signal KDEL to the wild type protein and measure oncogenic signaling
   C. Addition of Mannose 6-phosphate to the wild type protein and measure oncogenic signaling
   D. Addition of Lys and Arg repeats to the to the wild type protein and measure oncogenic signaling

61. Which of the following methods is used to identify expression of cell surface proteins?

   A. FACS
   B. Chromatin immunoprecipitation
   C. Western blotting
   D. Both A and B

62. Which of the following protein purification methods are used to separate proteins of same molecular weight and charge?

   A. Gel filtration and Ion-exchange chromatography
   B. Isoelectric focusing and Hydrophobic Interaction Chromatography (HIC)
   C. Hydrophobic Interaction Chromatography (HIC) and Ion-exchange chromatography
   D. Isoelectric focusing and Ion-exchange chromatography

63. Which of the following does *NOT* happen during RNA splicing?

   A. Mutually exclusive exons
   B. Alternative 3' splice site
   C. RNA editing
   D. Exon skipping

64. DNA, protein, DNase and Proteinase K were incubated in a test tube at appropriate conditions where reaction takes place. Which of the following will stay at the end?

   A. DNA
   B. Protein
   C. Proteinase K
   D. DNase

65. VLA-4 is an integrin family member, which is important for the cell-cell adhesion. In certain kinds of leukemia mainly Acute Myeloid leukemia (AML) and Acute lymphoblastic leukemia (ALL), over expression of this particular integrin is due to SDF1 alpha and CXCR4 signaling axis. Which of the following will be a specific therapeutic agent in order to treat these leukemia?

   A. Anti-VLA-4 treatment
   B. Inhibition of SDF1 alpha and CXCR4 signaling axis
   C. Downregulation of VLA-4 expression by bromodomain inhibitors
   D. Treatment with DNA methylating agents

66. Genomic DNA of which organism is likely to give bands of satellite DNA on being subjected to density gradient centrifugation:

   A. *Mycobacterium lapre*
   B. *E. coli*
   C. SARS-coV-2
   D. *Plasmodium falciparum*

67. A network comprising of 8 nodes is with the following degree distribution. Which of the statements is the most appropriate one in describing the nature of this network?

Node	Degree
1	1
2	1
3	2
4	10
5	2
6	3
7	8
8	1

   A. This is a scale-free network.
   B. This is a random network.
   C. This is a scale free network with two hubs (nodes 4 and 7).
   D. This is a random network with two hubs (nodes 4 and 7).

68. Consider an open decay reaction A→ with a decay rate of k. The ODE representing this reaction is:
(Here, [A](t) = concentration of A at time t, Ao= initial concentration]

A. $\frac{d}{dt}[A](t) = +k[A](t)$

B. $\frac{d}{dt}[A](t) = -k[A](t)$

C. $[A](t) = Aoe^{-kt}$

D. $[A](t) = Aoe^{kt}$

69. In Basic Local Alignment Search Tool (BLAST), how is the raw alignment score (S) related to the Expectation-value

    A. Linearly
    B. Not related
    C. Inversely
    D. Exponentially

70. In a hypothetical population, an estimate of frequency of two alleles 'A' and 'a' was 0.2 and 0.8, respectively. A genotyping survey of a sample of size 1000 reported 648 individuals of 'aa' genotype, 200 'Aa' genotype, and 152 of AA genotype. Which statement is *CORRECT* about this population

    A. The population is in Hardy-Weinberg equilibrium
    B. The population is not in Hardy-Weinberg equilibrium
    C. The population is unlikely to attain Hardy-Weinberg equilibrium even in subsequent generations
    D. Data is not sufficient to draw inference about Hardy-Weinberg equilibrium