

Name :

Roll No. :

Invigilator's Signature :

CS/M.Tech (BT)/SEM-2/MBT-203/2010

2010

BIOINFORMATICS & DRUG DESIGN

Time Allotted : 3 Hours

Full Marks : 70

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

GROUP – A

(Multiple Choice Type Questions)

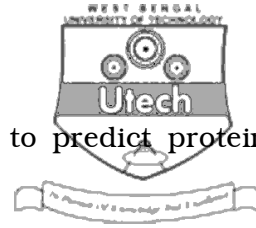
1. Choose the correct alternatives for any *ten* of the following : 10 × 1 = 10

i) A 3_{10} helix has atoms separating the amino hydrogen and carboxyl oxygen atoms that are hydrogen bonded together to form one complete turn of the helix.

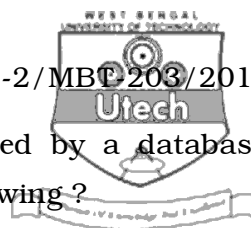
- | | |
|-------|-------|
| a) 10 | b) 12 |
| c) 5 | d) 6. |

ii) An example of a functional macromolecular fold is

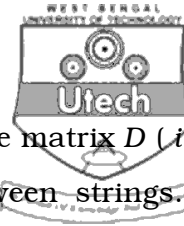
- | | |
|--------------------|------------------|
| a) TIM barrel | b) Rossmann fold |
| c) Flavodoxin-like | d) all of these. |



- iii) LINUS is an algorithm that is used to predict protein fold. The algorithm is based on
- dot-plot comparison
 - hidden neural network
 - hierarchic condensation
 - optimal local alignment.
- iv) A single ligand-multiple protein docking calculation provides which of the following specific pieces of information ?
- Binding energy estimation
 - Mode of binding
 - Specificity prediction
 - Ranking of affinities of ligand.
- v) The geometrical interpretation of a neuron that accepts two inputs x and y and fires if and only if $x + 2y \geq 2$ is that it selects
- points below and to the left to the line $x + 2y = 2$
 - points above and to the right of the line $x + 2y = 2$
 - points to the right of the line $x + 2y = 2$
 - points to the left of the line $x + 2y = 2$



- vi) A biologically significant 'hit' returned by a database search is easiest for which of the following ?
- Protein local alignment
 - Protein global alignment
 - DNA matches from a coding region
 - DNA matches from a non-coding region.
- vii) An example of a lead compound developed from the side effect of an existing drug is
- Penicillin
 - Interleukin-2
 - Minoxidil
 - all of these.
- viii) A good cross validation in QSAR is indicated by
- $r^2 < 0$
 - $r^2 < 0.5$
 - $r^2 = 0.5$
 - $r^2 > 0.5$.
- ix) The partition coefficient P between 1-octanol and water is given by
- $P = [\text{compound}]_{\text{oct}} / [\text{compound}]_{\text{aq}} (1 - \alpha)$
 - $P = [\text{compound}]_{\text{oct}} / [\text{compound}]_{\text{aq}}$
 - $P = [\text{compound}]_{\text{aq}} / [\text{compound}]_{\text{oct}} (1 - \alpha)$
 - $P = [\text{compound}]_{\text{oct}} \times [\text{compound}]_{\text{aq}}$.
- x) Given two character strings, a measure of the distance between them is given by
- Interpolation length
 - Forster distance
 - Hamming distance
 - Leventhal's distance.



- xi) In a dynamic programming algorithm the matrix $D(i, j)$ represents the minimum distance between strings. In which of the following ways does the algorithm compute $D(i, j)$?
- a) By a recursive operation
 - b) By an iterative one
 - c) By an exponential decay
 - d) By a parabolic dependence.
- xii) Structure prediction of α -helical transmembrane segment includes use of
- a) hydrophobicity
 - b) neural networks
 - c) evolutionary information
 - d) all of these.
- xiii) The key property(ies) of biological systems/modules is/are
- a) irreducibility
 - b) emergence
 - c) complexity
 - d) all of these.
- xiv) In beginning software platform/OS development which one of the following corporate entities was not involved ?
- a) Intel
 - b) Oracle
 - c) Sun Microsystems
 - d) Microsoft.



- xv) Monophyletic group is
- a) a group of taxa descended from a single common ancestor
 - b) a pair of taxa descended from a single common ancestor
 - c) five taxa descended from a single common ancestor
 - d) a group of species descended from a single common ancestor.
- xvi) PSSM stands for
- a) Position Specific Scoring Models
 - b) Positional Specific Scoring Models
 - c) Position Specific Scoring Matrix
 - d) Position Specific Scoring Match.
- xvii) Histamine is
- a) a bronchodilator
 - b) a small compound that becomes immunogenic under specific conditions
 - c) a vasoactive organic compound released from granules within mast cells
 - d) an enzyme that unwinds the DNA double helix.



GROUP – B

(Short Answer Type Questions)

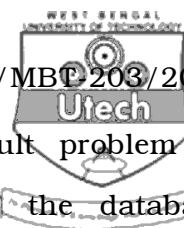
Answer any *three* of the following.

3 × 5 = 15

2. a) Briefly describe the two cladistic methods that deal with patterns of ancestry.

b) Why are cladistic methods more accurate than simple clustering methods ? 2 + 3
3. What two sub-disciplines form the basis of systems biology ?
How does systems biology lead to a better understanding of genotype-phenotype relationships ? 2 + 3
4. Many proteins from pathogens have human homologous. Suppose you had devised a method for comparing the factors that determine specificity in the binding sites of two homologous proteins. How could you use this method to select specific targets for drug design ?
5. a) Define different types of gap penalties in sequence alignment programmes, with suitable examples.

b) Mention the difference between global and local alignments. 3 + 2



6. Why is loop modelling a relatively difficult problem in homology modelling ? Pointwise describe the database method for modelling loops. Name two public domain web servers that model loops. 2 + 2 + 1
7. What is the main objective of molecular phylogenetics ? Use a simple diagram to represent why finding a correct tree topology is computationally difficult. Write out the mathematical formula for the number of trees (N_R) for n taxa. Explain the formula. 2 + 1 + 2
8. Many empirical methods have been developed/being developed for accurate correlation of physico-chemical parameters with biological activity. Describe the Free and Wilson equation to address this issue and the modifications to this. Validate this approach with one hypothetical example of a lead compound.
9. a) The overall base composition of the *E.coli* genome is $A = T = 49.2\%$; $G = C = 50.8\%$. In a random sequence of 4 639 221 nucleotides with these proportions, what is the expected number of occurrences of the sequence CTAG ?
- b) Depict the construction of a PSSM from a multiple alignment of nucleotides. $2\frac{1}{2} + 2\frac{1}{2}$

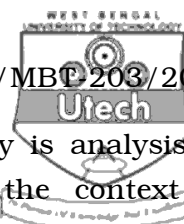


GROUP – C

(Long Answer Type Questions)

Answer any *three* of the following. $3 \times 15 = 45$

10. a) Briefly describe how the UPGMA method of tree building works. Why was the UPGMA method modified to the Neighbour joining method ? 4
- b) Consider 4 species characterized by homologous sequences ATCC, ATGC, TTCG and TCGG. Taking the number of differences as the measure of dissimilarity between each pair of species, use a simple clustering procedure to derive a phylogenetic tree. (Hint : You will be using the UPGMA method here). 7
- c) From the final tree graph obtained in (b) above how was the branch length of the nodes joining the clusters (ATCC, ATGC) and (TTCG, TCGG) arrive at ? 4
11. a) What are the major biochemical classes of drug targets ? Use a histogram to illustrate the approximate percentage of these. 4
- b) Give one example of existing/in development drugs in each of these categories which involved the active use of principles of structure-based design. 3
- c) Itemize 6 key scientific-technological developments in the field of drug discovery and development that led to the developing and increasingly widespread acceptance of this field. 4
- d) Use one of the examples you have cited above to draw up a flowchart of the steps involved in structure based drug design. How does structure based drug design fit in to simplify the clinical trials part of the drug development process. 4



12. a) What is a hydrophobicity profile ? Why is analysis of hydrophobicity profiles important in the context of bioinformatics and drug discovery ? Draw a prototypical hydrophobicity profile for HEWL (hen egg white lysozyme). What is the significance of the minima in a hydrophobicity profile ? 1 + 2 + 1 + 2

b) The $\log (1/K_i)$ of two substituted phenyl-based inhibitors was determined and expected to be a simple linear function of hydrophobicity : $\log (1/K_i) = a\pi + c$. Use the data below to develop the corresponding QSAR equation. 5

Substituent	$\log (1/K_i)$	π
<i>n</i> -butyl	8.24	2.52
<i>F</i>	7.06	0.63

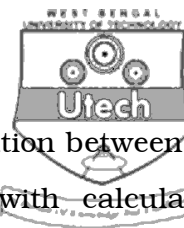
c) Interpret the answer in (b) above in terms of what you have written in (a) above. 4

[Be specific in answering parts (a) and (c) of this question]

13. a) For finding treatments for diseases, the importance of structure prediction in proteins has acquired significance. Pointwise explain 3 key reasons how structure prediction helps in this process. 5

b) What are the methods normally adopted for prediction of protein structure ? 5

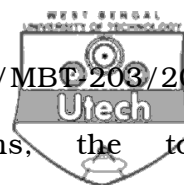
c) Explain the principles and algorithm of any one method of predicting protein-protein interaction that uses Monte Carlo simulation. 5



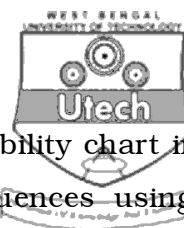
14. A classical QSAR attempts to set up a correlation between an experimental property A (e.g. activity) with calculated structural parameters a, b, c in an equation of the form

$$\log A = x_1 a + x_2 b + x_3 c + \dots + \text{constant.} \quad \text{Equation 1}$$

- a) What are the properties of molecules such as A based on ? 2
- b) What sort of mathematical representations does a, b, c have ? 2
- c) If Equation 1 is found to be valid, then what can it be used for ? 2
- d) What is the primary use of Equation 1 in the context of drug discovery ? What physico-chemical properties of a drug do the parameters in Equation 1 above attempt to measure normally ? 3
- e) What standard statistical methods have been classically used to validate QSAR equations like 1 above ? 2
- f) What is the calculated $\log P$ value for the long known anti-cancer drug diethylstilbestrol (hint : it has 2 methyl groups, two CH_2 linkages, one ethylenic linkage, and a phenolic moiety; π for methyl and methylene = 0.50, for ethylenic linkage = 0.69 and $\log P$ is 1.46 for phenolic group; use standard correction factor). 4



15. a) In molecular dynamics simulations, the total conformational entropy of a biomolecule is typically computed by the method of isomer counting. An alternative method of computation is normal mode analysis to find out how rigid a particular structure is. How is kinetic energy specifically defined in this model? Which biological macromolecular system has the normal mode analysis been applied to? 3 + 2
- b) What are 3 force fields that have been commonly employed for biological applications? What macromolecular systems were these force fields utilized for? How are electrostatic interactions treated in these force fields? 2 + 1 + 2
- c) Gatifloxacin is a methoxyfluoroquinolone that is used as an ophthalmic solution. What are the normal pharmacokinetic parameters that are used to decide dosage/frequency of administration for this drug? How can these terms be incorporated in a generalized QSAR equation? $2\frac{1}{2} + 2\frac{1}{2}$



16. a) How is a transition and emission probability chart in a 2-state set-up for analyzing DNA sequences using a partial HMM ? 4
- b) Draw a typical architecture of a hidden Markov model that represents a multiple sequence alignment. Explain the meaning of the symbols used. 5
- c) A score matrix is constructed from a simple HMM to define optimal score paths. Different dynamic programming algorithms are used to construct a score matrix. Explain the procedure by which the Viterbi algorithm develops a score matrix for multiple sequence alignment. 6
17. What is the basis for DNA interactive drugs ? What is the toxicity of DNA interactive drugs ? Outline one numerical method for measuring the toxicity profile of a DNA-interactive drug. Name the three classes of drugs that interact with DNA. Give one example of a drug within each category. Use one example from within any of the three classes of DNA-interactive drugs to highlight how principles of structure based drug design have been used/could be used for lead modification and optimization. 2 + 2 + 3 + 4 + 4
-