

Name : .....

Roll No. : .....

Invigilator's Signature : .....

**CS / M.TECH(BT) / SEM-2 / MBT-201 / 2012**

**2012**

**ADVANCED BIOINFORMATICS**

*Time Allotted : 3 Hours*

*Full Marks : 70*

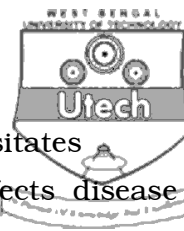
*The figures in the margin indicate full marks.*

*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 alternative for the following :  $10 \times 1 = 10$ 
  - i) The most acceptable phylogenetic marker is
    - a) cDNA
    - b) mRNA
    - c) 16srRNA
    - d) 23rRNA.
  - ii) A coiled coil is a secondary structural element with two or more
    - a) interacting alpha-helices
    - b) beta-sheets
    - c) knots
    - d) all of these.
  - iii) Fitch-Margoliash is a
    - a) statistical model
    - b) substitution matrix
    - c) motif evaluating software
    - d) phylogenetic method.



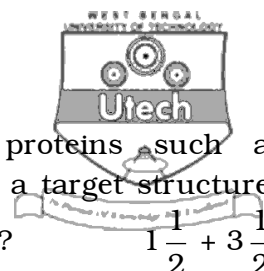
- iv) Target selection in drug discovery necessitates
- linking a target molecule that affects disease by affecting its function or expression
  - finding information about specific protein-protein interactions
  - knowledge about differential genomics and proteomics
  - all of these.
- v) Kimura model is also known as a ..... parameter model .
- one
  - two
  - three
  - $k^{\text{th}}$  .
- vi) For 6 taxa mathematically ..... rooted trees are possible .
- 105
  - 720
  - 136
  - 945 .
- vii) For a cladogram which of the following choices is true ?
- Branch lengths are proportional to number of evolutionary changes
  - Branch lengths are not proportional to the number of evolutionary changes
  - Both are correct
  - Branch lengths do not affect evolutionary choices at all.
- viii) For motif detection which is the better option ?
- Gibbs motif sampling
  - BLAST
  - Regular expression
  - FASTA.
- ix) If an overall candidate compound library size is  $10^{12}$  compounds, approximately how many compounds are typically forwarded for consideration for clinical trials ?
- 1—10
  - 10—100
  - $10^3$
  - $10^4$  .

- ### GROUP – B

Answer any *three* of the following.  $3 \times 5 = 15$

- $$\log \frac{1}{c} = -a\pi^2 + b\pi + \rho\sigma + cE_s + dS + e.$$

- [ Turn over



7. What biological characteristics make proteins such a valuable source for drug targets ? How is a target structure evaluated for structure based drug design ?  $1\frac{1}{2} + 3\frac{1}{2}$
8. What were the reasons that necessitated the technological developments in combinatorial chemistry ? What are the main utilities for molecular diversity in drug development ?  $2\frac{1}{2} + 2\frac{1}{2}$
9. Sorafenib is a multi-functional kinase inhibitor drug that has been in the very recent drug development news. Explain how drug polypharmacology in this inhibitor drug can be explained on the basis of the physico-chemical characteristics of the relevant binding sites on the pharmacophore.

### GROUP – C

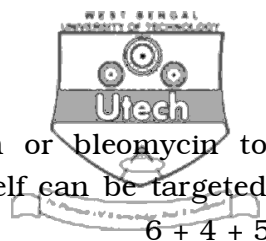
#### ( Long Answer Type Questions )

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

10. a) Use a flowchart to explain how a sequence profile/evolutionary information and multilayered filtering steps can be used in a neural networks based protein secondary structure prediction algorithm. Name two algorithms that work on the above principle briefly outlining differences in the sequence alignment step. What specific structural element filtering function does the final network/jury layer perform ?  $4 + 2 + 2$
- b) What structural/charge characteristics in transmembrane proteins necessitate the induction of the positive-inside rule ? Give a schematic representation of the positive-inside rule. A transmembrane protein secondary structure prediction program like HMMTOP, although accurate have impaired prediction accuracies for signal peptides. Explain the reasons for this impairment.  $2 + 2 + 3$



11. a) What type of phylogenetic tree would be suitable to construct when common ancestral knowledge is assumed ?
- b) How many rooted trees can be constructed for four taxa ?
- c) What is the role of a molecular clock in tree construction ?
- d) Itemize the steps involved in phylogenetic tree construction.
- e) Define bootstrapping. In perturbation of biological data sets, how do parametric and non-parametric bootstrapping methods differ ?  $1 + 2 + 2 + 2 + 5 + 3$
12. Briefly name and outline the computational approaches to 3D (tertiary) protein structural modelling and prediction. Draw a properly labelled flowchart that shows the steps involved in homology modelling. Why is loop modelling a necessary part of homology modelling ? What are the two general methods of doing homology modelling ? In tabular fashion, explain the differences in methodology between threading and homology modelling.  $3 + 5 + 2 + 2 + 3$
13. a) Name three biological force fields that are commonly used to describe the conformational behaviour of proteins and nucleic acids. Write the complete general energy expression in any one of these force fields explaining all the terms.
- b) More than 60% of marketed drugs are small molecule and protein based, target GPCRs. How many classes of GPCRs have been segmented to date ? Name 4 class I receptors and name the 3 segments of nuclear receptors.



- c) Use either the example of cisplatin or bleomycin to itemize how a nucleic acid /DNA itself can be targeted as part of drug design.

6 + 4 + 5

14. What are the advantages of multiple sequence alignment over pairwise alignment ? Describe briefly the methods for construction of a PSSM matrix. Use a comparative schematic diagram to outline the steps of BLAST and PSI-BLAST.

3 + 4 + 8

15. a) Use an "integrated flowchart" to depict the various areas here in silico computational procedures (viz. a computer and its associated hardware and software) can intervene in the drug discovery process. Explain in *concise technical terminology* how several items in this flowchart have evolved/changed the drug discovery process over the past decade .

6

- b) Explain the methodology of affinity ranking of candidate drugs. What is the basis of choice of *scoring functions* that are utilized in such comparative docking schemes ?

$2 \times 2 \frac{1}{2}$

- c) The  $\log \left( \frac{1}{k_i} \right)$  of two substituted phenyl-based inhibitors are determined and expected to be a linear function of hydrophobicity :  $\log \left( \frac{1}{k_i} \right) = a\pi + c$ ; use the data below to develop the QSAR equation :

4

Substituent	$\log \left( \frac{1}{k_i} \right)$	$\pi$
<i>n</i> -butyl	7.75	2.13
F	6.57	0.14



16. a) Outline the key assumptions in clustering based methods using the UPGMA method as a basis for your answer.
- b) Explain the operation of maximum parsimony tree building.
- c) Tabulate the advantages/disadvantages of ML (maximum likelihood) method.
- d) Find informative and non-informative sites between the following sequences. Cite reasons for your choice :
- AAGAGTGGA  
 AGCCGTGCG  
 AGATATCCA  
 AGAGATCCG 2 + 3 + 4 + 6
17. a) What are the major sub-branches of combichem and molecular diversity ? Briefly elaborate on any two of them.
- b) Explain why combinatorial chemistry has been compared to phage display. What are the major classes of compounds accessible to date through use of combinatorial chemistry techniques ?
- c) Use an example to highlight how a 96 well microtitre plate can be used to develop a combinatorial library of compounds. Your example should include approximate numbers of compounds that can be generated from the starting compound. 7 + 4 + 4

=====