

**M.Sc. Biotechnology IV Semester**  
**BT - XVI: Computational Biology & Biostatistics**                      **Marks: 100**                      **Hours: 45**

**UNIT-I: Biological Data Bases**

The need for computation in Biology: An introduction to Bioinformatics, Historical overview, the principles involved, development of tools, internet based access. Introduction to Biological Databases, Database Browsing and Data Retrieval - Sequence databases, Structural databases, Literature and other databases

**UNIT-II: Applications of Bioinformatics**

Application of Bioinformatics Approaches for analysis and interpretation of Sequence Data and using: Homology Searches, Sequence Alignments, Pattern Searching. Application of Bioinformatics Approaches for analysis and interpretation of Genome data such as - Gene prediction, Full Genome comparison etc. Introduction to computational structural biology: Protein structure prediction using computational methods, Structure analysis, Classification of Proteins etc.

**UNIT-III : Proteomics**

Strategies in Proteomics: 2 D PAGE, Mass spectrometry. Databases and search engines in proteomics. Proteomics applications: Understanding the mechanism of pathogenesis, Drug discovery, Disease diagnosis, identification and characterization of novel proteins. Protein-Ligand Docking: Introduction; Docking problems, methods for protein- ligand docking, validation studies and applications

**UNIT-IV: Genomics**

Introduction sequencing strategies for whole genome analysis, sequence data analysis. Comparative Genomics: Protein evolution from exon shuffling, Protein structural genomics, Gene function by sequence comparison Global expression profiling : whole genome analysis of mRNA and protein expression, microarray analysis, types of microarrays and their applications Functional genomics, Toxicogenomics, Pharmacogenomics, Metagenomics. Metabolic engineering

**UNIT V: Biostatistics**

Brief description and tabulation of data and its graphical representation Measurement of central tendency and dispersion- mean, mode, median, range Mean deviation, standard deviation, variance . Idea of two types of errors and level of significance. Tests of significance- F-Test and chi-square test. Linear regression and correlation.

**Text & Reference:**

1. Teresa Attwood, David Parry-Smith - Introduction to Bioinformatics Prentice Hall
2. Jin Xiong- Essentials of Bioinformatics- Cambridge
3. Pierre Baldi, Søren Brunak -Bioinformatics : the Machine Learning Approach MIT Press
4. Rashidi H.H and Buahler L.K - Bioinformatics : Applications in Biological Science and Medicine-CRC
5. Andreas D. Baxevanis, B.F. Francis Ouellette - Bioinformatics : A Practical Guide, J. Wiley
6. Philip E. Bourne, Helge Weissig - Structural Bioinformatics Wiley,
7. Peter Clote, Rolf Backofen - Computational Molecular Biology : an Introduction, Wiley, 2000.
8. Warren J. Ewens, Gregory R. Grant - Statistical Methods in Bioinformatics : an Introduction Springer,
9. Timo Koski , Hidden Markov - Models for Bioinformatics Kluwer Academic Publishers
10. Arthur M Lesk - Introduction to Bioinformatics, Oxford University Press
11. David W. Mount - Bioinformatics : Sequence and Genome Analysis, Cold Spring Harbor Laboratory
12. Pavel A. Pevzner - Computational Molecular Biology : an Algorithmic Approach MIT Press
13. Fundamental of Statistics- S.C. Gupta
14. Statistical Method S. P. Gupta
15. Mathematical Statistics- S.C. Gupta & Kapoor

**Practical (Lab course work VII)**

1. NCBI/EBI : Data access – standard search engines : data retrievals tools – Entrez, DBGET and SRS, Pubmed etc
2. Software for data building. Submission of sequence to databases.
3. Sequence homology as product of molecular evolution, sequence similarity searches,
4. Sequence alignment-global, local, end free-space; measurement of sequence similarity
5. Homology Modeling
6. Phylogeny reconstruction by using biological data
7. Getting an amino acid sequence, nucleotide sequence by BLAST
8. Protein identification & characterization with peptide mass fingerprinting data.
9. Primary/ secondary structure analysis of proteins.
10. Tertiary structure analysis of proteins (3D structure prediction)
11. Experiments based of biostatistics and assumed data sets
12. Measures of central tendency and dispersion mean median, mode, range, standard deviation variance ,standard error