BHARATHIDASAN UNIVERSITY, TIRUCHIRAPPALLI – 620 024

P.G. DIPLOMA IN BIOINFORMATICS

ELIGIBILITY : Any Bachelor Degree in Science from a recognised University

(For students admitted from 2005-2006 onwards)

Semester I				
Paper No.	**Title of Paper**	**I.A.**	**U.E**	**Total**
01	Basics of Bioinformatics	25	75	100
02	Statistics for Bioinformatics	25	75	100
03	Basic Structural Biology	25	75	100
04	Object Oriented Programming and Web Publishing	25	75	100
05	Practical–I : C++ Programming and Web Publishing Lab	10	40	50
Semester - II				
06	Data Base Management Systems	25	75	100
07	Bio-Physics	25	75	100
08	Genomics and Proteomics	25	75	100
09	Computational Biology	25	75	100

10	Elective (anyone of the following)
	  a. Commercial Applications of Bio-informatics
	  b. Molecular Modelling and Drug Design
	  c. Data Communication Networks
		25	75	100

11	Practical – II Bioinformatics Lab	10	40	50

Paper I – Basics of Bioinformatics

Unit I

Bioinformatics – an Overview, Definition and History. Information Networks – Internet in Bioinformatics, EMBnet – Commercial Databases and Softwares, Intranet and Internet Packages, Bioinformatics Glossary.

Unit II

Languages – Basics of Programming
Perl – Basics, String handling, subroutines.
HTML – Basics, Tags, Text handling, Image handling, Links, Frames and Tables.
XML – Basics, data binding and record sets.

Unit III

Protein Information Resources – Biological Databases, Primary Sequence Databases, Composite Protein Sequence Databases, - Secondary Databases – Prosite, Prints, Blocks Profiles and Identity.

Unit IV

Genome Information Resources – DNA sequence Databases – EMBL DDBJ, Genbank GSDB (Genome, Sequence Database), UniGene.

Unit V

Evolution of Bioinformatics – Scope – Potentials of Bioinformatics Human Genome Project – Bioinformatics in India – Future of Bioinformatics.

Reference Books :

1. T.K. Attwood and D.J. Parry-Smith, *Introduction to Bioinformatics*, Pearson Education Ltd., New Delhi (2004).
2. Arthur M. Lesk, *Introduction to Bioinformatics,* Oxford University Press, New Delhi (2003).
3. D. Higgins and W. Taylor (Eds), *Bioinformatics- Sequence, structure and databank*s, Oxford University Press, New Delhi (2000).
4. A. Baxevanis and B.F. Ouellette. *Bioinformatics: A practical Guide to the Analysis of Genes and Proteins*, Wiley-Interscience, Hoboken, NJ (1998).
5. S. R. Swindell, R.R.Miller and G.S.A.Myers (Eds.), Internet for the Molecular Biologist, Horizon Scientific Press, Wymondham, UK, (1996).
6. Andrea Cabibbo, Richard Grant and Manuela Helmer-Citterich (Eds.), The Internet for Cell and Molecular Biologists (2nd Edn.), Horizon scientific Press, Norwich, UK (2004).

Paper II – Statistics for Bioinformatics

Unit I

Measures of central Tendency – Arithmetic Mean – Median – Mode – Quartiles – Range – Quartile deviation – Mean deviation – Standard Deviation.

Unit II

Probability – Addition Theorem – Multiplication Theorem - Baye's Theorem and related problems.

Unit III

Theoretical Distributions – Binomial, Poisson and Normal – Fitting of the Distributions and its properties - Z-score, P-value and E-value.

Unit IV

Theory of Attributes – Introduction – Dichotomy – Consistency of Data – Independence of Attributes – Association of Attributes – Rules coefficient of Association.

Unit V

Sampling Distributors – Large and small sample tests – Student's t^0 test, X^2 test, F-test – and Normal test and their applications.

Text Book :

1. S.C. Gupta and V.K. Kapoor, Fundamentals of Mathematical Statistics, 11th Edition, Sultan Chand & Sons, New Delhi, 2002.

Reference Books :

1. S.P.Gupta, Statistical Methods, Sultan Chand & Sons, 1996.
2. L. Forthofer, Introduction to Biostatistics, Academic Press, 1995.
3. Robert R. Sokal and F.J. Rohlf, Introduction to Biostatistics (Biology-Statistics Series), W.H. Freeman & Company, New York, 1987.

Paper III – Basic Structural Biology

Unit I

Cell Structure and Ultracell structure of Pro and Eukaryotic cells – Cell wall – cell membrane – Biomembranes – Organelles. – Diffusion – Active and Passive Transport – Osmoregulation

Unit II

Carbohydrates – Classification Types – Structure – Function.
Lipids – Classification Types – Structure – Function.

Unit III

Classification of amino Acids. Classification and three-Dimensional structure of proteins. Overview of protein structure – Primary, Secondary, Tertiary and Quartenary structures – α helix, β Pleated sheet, Ramachandran plot and bonds stabilizing protein structure.

Unit IV

Principles of Structural Organization and Conformational Analysis.

Prediction of Pprotein structure – Fold Recognition (threading), Comparative Modeling (homology), The Chou and Fasman Sheme. Basic Principles of X-ray Diffraction Studies, NMR, Mass Spectroscopy in Identifying Protein Conformation.

Unit V

Introduction and Physiochemical Properties of Nucleic Acids – DNA and RNA. Watson and Crick Model of DNA and the different forms of DNA. RNA structure – Principles and Prediction. Gene Structure.

Reference Books :

1. Principles of Physical Biochemistry – Van Holde, Prentice Hall.
2. Basic one & two dimensional NMR sepectroscopy, Horst Friebolin Witey – VCH – 1990.
3. Lechninger Principles of Biochemistry – David L.Nelson, Michael M.cox. 3rd Edition Macmillan worth Publishers.
4. Principles of Protein structure, G.E.Shuttz and R.H.Schirmer, Springer – verlag, New York.

Paper IV – Object Oriented Programming and Web Publishing

Object Oriented Programming with C++

Unit I

Introduction to the basic concepts of C++ language – Tokens, Keywords, Identifiers, Data types, Variables, Manipulators – Expression and Control structures.

Functions : Main function – function prototyping – call by reference – function overloading – friend and inline functions.

Unit II

Classes and objects – Constructors and Destructors – Operator overloading – Type conversions.

Unit III

Inheritance – Single inheritance – Multiple inheritance – Hierarchical, Hybrid inheritance – Polymorphism – Pointers – Console I/O operations.

Unit IV

Files – classes for file stream operations – Opening, closing and processing files – End of file detection – File pointers – Error handling during file operations – Command line arguments – Exception handling.

WEB PUBLISHING :

Internet – WWW – Browser – Designing web site – Basics of creating a web page with HTML – Linking – Text formatting – Adding Images and background to HTML pages – Tables – Image maps.

Text Books :

1. For Units 1 to 4 "Object Oriented Programming with C++" – E Balagurusamy – TMH.
2. For Unit 5 : "Web Publishing" – Monica D'Souza & Jude D'Souza – TMH – 2001. (Chapters 1 to 7 and 9 only).

Reference Books :

1. Object Oriented Programming in C++ - Robert Lafore, Galgotia
2. Let us C++ - Yeshwant Kanetkar – BPB.
3. Instant HTML Programmer's Reference – Steve Wright – WROX Press Ltd.

Practical – I   C++ Programming and Web Publishing Lab

C++

1. Factorial of a given number
2. Sorting and Merging two arrays of numbers.
3. Sorting names in alphabetical order
4. Reversing a given string
5. Checking for palindrome
6. String Manipulation – String copy, String length, String concatenation, String comparison and String reverse.
7. Mean, median, mode and standard deviation calculations.
8. Storing and retrieving amino acid sequences using structure data types
9. Swapping two values using pointers
10. File Processing – Creation and simple processing
11. Computing amino acid composition of a given protein sequence.

Web Publishing

1.   Create a web page for your University / College using HTML. The opening page should provide hyperlinks to other pages (add animation and sound effects appropriately).

2.   Creating a web page to get protein sequence data and compute and display amino acid composition

3.   Creating a web page to get nucleic acid sequence data and compute and display base composition

SEMSTER II

Paper VI - Database Management Systems

Unit I

Data base : Introduction – Basic Technology and Data Basics – Objective of a Data base organisation – Entities & Attributes.

Unit II

Data Models : Introduction – Schemes & Sub Schemes – Data base Management Systems – Tree Structures – Plex Structures – Relational Databases – Normal Forms.

Unit III

Data Base Languages : Introduction – Data Description Languages – the CODASYL Data Description Language – Query Languages – Data Dictionaries.

Unit IV

Physical Organization : Introduction – Criteria affecting physical organization – Differences between physical & logical organization – Addressing Techniques – Index sequential organisation – Hashing.

Unit V

Pointers – Chains & Ring Structures : Introduction – Physical representation of tree structures – Physical representation of plex structures – Virtual & Storage hierarchies.

Text Book :

1. James Martin, Computer Database Organisation, Prentice Hall of India.

Reference Books :

1. James Martin, Principles of Database Management, Prentice Hall of India
2. C.J.Date, An Introduction to Database systems, 3rd edition, Narosa Publishing House.

Paper VII - Biophysics

Unit I   Introduction

Levels of structure in biological macromolecules, Central questions in biophysics, basic strategies in biophysics.

Unit II   Conformational Analysis

Forces that determine protein and nucleic acid structure, basic problems, polypeptide chain geometrics, potential energy calculations, observed values for rotation angles, hydrogen bonding, hydrophobic interactions and water structure ionic interactions, disulphide bonds.

Unit III   Structural Analysis of Macromolecules

Prediction of protein structure nucleic acids, general characteristics of nucleic acid structure, geometrics, glycosidic bond rotational isomers and ribose puckering backbone rotational isomers ane ribose puckering forces stabilizing ordered forms, base pairing, base stacking tertiary structure of nucleic acids.

Unit IV   Kinetics of ligand Interactions

Biochemical kinetics studies, unimolecular reactions, simple biomolecular multiple intermediates, steady stake kinetics, catalytic efficiency, relaxation spectrometry, ribonuclease as an example.

Unit V   Techniques for Study of Biological Structure and Function

Size and shape of macromolecules – method of direct visualization – macromolecules as hydrodynamic particles – macromolecular diffusion – ultracentrifugation – viscometry x-ray crystallography determination of molecular structure x-ray fibre diffraction electron minoscopy neutron scattering – light scattering.

Reference:

1. C.R.Cantor & P.R.Schimmel, Biophysical Chemistry Part – I, II & III, W.H. Freeman & Co., in San Fransisco, 1980.
2. C. Branden and J. Tooze, Introduction to Protein Structure, Garland Publishing Inc., New York., 1991.
3. R. Glaser, Biophysics, Springer, 2000.

Paper VIII – Genomics and Proteomics

Unit I

The genetic material – Identification of genetic material, genetic code, concept of gene – operon concept – *lac* and *trp* operons, promoters and repressors.

Unit II

Genomics – definitions – pharmacogenomics – taxicogenomics – prokaryotic and eukaryotic genome – genome relationships – human genomics.

Unit III

Genomics methodologies : Whole genome analysis – Physical methods of sequencing – automated sequencing – genome expression and analysis – serial analysis – oligo NT array technology – code micro assay and micro chips.

Unit IV

Proteomics : Definition – Transcriptomics; Proteomics, metabolomics. Techniques of proteomics – 2D PAGE, Multidimensional protein identification (Mud PIT) Isotopically coded affinity Tag (ICAT), Mass spectrophotometer – (MALDI – TOF) MS. Application of functional genomics in basic biology, target / marker identification, target valediction / toxicology, microbial drug resistance, tumour immunology, vaccine discovery, drug design.

Unit V

A brief account of genetic engineering and protein engineering.

References :

1. Baxevanis AD and B.F.F. Ouellette, Wiley Bioinformatics – A practical guide to the analysis of genes and proteins. (ed) – Interscience, New York, 2001.
2. Doolittle RF (1990). Molecular evolution. Computer Analysis of Protein and Nucleic acid Sequences, Methods in Enzymology, Academic Press, New York.
3. Faber K (1992) Biotransformation in Organic Chemistry, Springer Verlag.
4. Gerbardt P.Murray RG, Wood WA, Kreig NR (1994) Methods for General and Molecular Bacteriology – American Society for Microbiology Washington D.C.

Paper IX – Computational Biology

Unit I

Structure of DNA & Protein – Sequence analysis – pairewise sequence comparison – sequence queries against biological databases – BLAST and FASTA – multifunctional tools for analysis.

Unit II

Multiple sequence alignments, Phylogenetic alignment – profiles and motifs – distance and similarity – evolutionary basis of sequence alignment – scores and gaps.

Unit III

Protein structure visualisation – tools structure – classification, alignment and analysis. Solvent accessibility and Interactions – Physico chemical properties, structure optimisation. Protein resource databases.

Unit IV

Predicting Protein structure and function from sequence – Determination of structure – feature detection – secondary structure prediction – Predicting 3 D structure. Protein modeling.

Unit V

Genomics and Proteomics – Sequencing genomes – sequence assembly – genome on the web – annotating and analyzing genome sequences. Proteomics – biochemical pathway databases – submitting sequence to the databases.

Reference:

1. Bioinformatics : A practical guide to the analysis of genes and proteins – 2001 – AD Baxevanis & BFF Ouellette – Wiley Interscience – New York.
2. Bioinformatics : Methods and Protocols – 2000 – Stephen Misener & Stephen A. Krawetz, Humana Press, New Jersey.
3. Bioinformatics : Sequence, structure and databanks – 2000 – Des Higgins & Willie Taylor – Oxford University Press.

Paper X – Elective

a. Commercial Applications of Bioinformatics

Unit I

Commercial Bioinformatics, Definition of Bioinformatics company, Genome Technology; high throughput sequencing and assembly.

Unit II

Micro arrays and genome wide expression analysis; transcriptome, proteome, Genomics in medicine, diseases monitoring, profiles for therapeutic molecular targeting.

Unit III

Diagnostics drug discovery and genomic. Pharmacogenomics and its applications. SNPs and their applications.

Unit – IV

Proteomic in medicine and therapeutic target identification. Comparative proteomics and its applications.

Unit V

Patenting and data generation from patent literature for commercial benefits. IPR and Bioinformatics. Bioinformatics patents.

Reference:

1. Drug Discovery Today, Trends journals.
2. Functional Genomics, http://www.functionalgenomics.org.uk/

b. Molecular Modeling and Drug Design

Unit I

Recent advances in drug design methodologies, Biomolecular structure, Structure activity relationship, Pharmacokinetics, structure-based drug design.

Unit II

Pharmacophoric pattern, ADME Properties, quantitative structure activity relationship, Use of genetic algorithms and principle component analysis in the OSAR equations.

Unit III

Molecular modeling, quantum mechanical and molecular orbital methods, introduction to semi-empritical, molecular mechanics and ab intio techniques. Simulation techniques, potential energy surfaces, docking and modeling substrate – receptor interactions.

Unit IV

Software tools for modeling bio-molecules. Molecular electrostatic potentials, charge analyses. Protein conformations, folding and mutation through modeling.

Unit V

Use of Genomics and Proteomics for understanding diseases at molecular level strategies for target identification and lead design.

Reference Books:

1. Andrew Leach, Molecular Modelling: Principles and Applications (2nd Edition), Addison Wesley Longman, Essex, England, 1996.
2. Alan Hinchliffe, Molecular Modelling for Beginners, John-Wiley, 2003.
3. N. Cohen (Ed.), Guide Book on Molecular Modeling in Drug Design, Academic Press, San Diego, 1996.

c. Data Communication Networks

Unit I

Data Communication : Introduction – Data Communication System – Data Communication Software – Data Communication concepts : Digital and analog Communication – Parallel and Serial Communication – Synchronous and Asynchronous Communication – Simplex, Half-duplex, Full-Duplex Communication.

Unit II

Communication Hardware : Communication Adapter – Acoustic Couplers and Modems – Communication Processing – Communication Channels : Wired Transmission – Optical, Fiber, Microwave, Laser, Radio and Satellite Transmission – Applications of Data Communication.

Unit III

Topologies and Transmission media; Bus Topology – Ring topology – Star Broad band coaxial cable topology - Star – Ring topology – Twisted Pair Wire – Base band Coaxial Cable – Fiber optic cables.

Unit IV

Networks : Introduction – Use of Computer Networks – Network Hardware : Local Area Network – Metropolitan Area Network – wide Area Network – Wireless Network – Internet works.

Unit V

Network Software : Introduction to Protocol – LAN protocols, Protocol Hierarchies – Open System Interconnection Reference Model – Design issues for the layer – Service Primitives.

Text Books :

1. For Unit – I, II & III : "Computer Networks", Andrews S.Tannenbaum, Prentice Hall of India, New Delhi, July 1998 – 3rd Edition.
2. For Unit – IV & V : "Local Area Networks", by S.K.Basandra Jaiswal, Galgotia Publishing.

Reference Books :

1. "Data and Computer Communication" William Stallings 5th Ed-PHI, 2000.
2. "Lan Tutorial, A Complete Introduction to Local Area Networks" 3rd Edition Comdex Computer Publishing by the Editors of LAN Magazine.

Practical – II   Bioinformatics Lab

1. Study of Internet resources in Bioinformatics.   E.g. NCBI, CGEB, EMBL.

2. Searches on MEDLINE, PubMed and CDROM bibliographic databases. Concept of boolean operators in searching.

3. Introduction to sequence data bases

   Protein sequence databank, NBRF-PIR, SWISSPROT, EMBL.
   Nucleic acid sequence databank – Gene bank, EMBL

4. Pair wise alignment- Needleman-Wunsch and Smith-Waterman algorithms

5. Multiple alignment- CLUSTALW & PRINTS

6. BLAST, FASTA programs for sequence database search

7. Genome data bank – study the features of human genome.

8. Evaluation of protein structure by Swiss PDB viewer and by other molecular visualization tools.

9. Calculation of phi – psi angles - Ramachandran plot.

10. Homology modeling of a given protein sequence.

****