

Bioinformatics
TRIMESTER WISE DISTRIBUTION OF COURSES

I Trimester

		L	P
BI 501 / MBB 503	Molecular Cell Biology	3	0
BI 502 / CA 502	Introduction to Computer Application	1	1
BI 503 CA 551	Mathematical Foundations in Computer Application	4	0
BI 504 / MBB 501	Principles of Biotechnology	3	0
BI -505 CA 561	Principles of Computer Programming	2	1
BI 523	Advanced Techniques for Sequence and Structure Analysis	1	2
BI 524	Tools and Techniques for biological data mining	2	1
BI 525	Advanced Programming in Bioinformatics	2	1
BI 691	Seminar	1	0

II Trimester

BI-506/ CA 566	Database Management System	2	2
BI: 507/ GP 540/ MBB 509/ AS 571	Bioinformatics - I	3	1
BI 508/ BIO 602	Protein Biosynthesis	3	0
BI 526	Comparative Genomics	1	1
BI 527	Phylogenetic Analysis	1	1
BI 528	Chemo informatics and IPR issues	1	1
BI 691	Seminar	1	0

III Trimester

BI 509/ MBB 602	Genomics and Proteomics	3	0
BI 510	Biological Databases and Data Analysis	2	1
BI 511	RNA/Protein Structure Prediction & Molecular Modeling	1	2
BI 512/ AS 608	Bioinformatics – II	2	1
BI 691	Seminar	1	0

Core Courses

M.Sc.: BI 501, BI 502, BI 503, BI 504, BI 505, BI 506, BI 507, BI 508, BI 509, BI 510, BI 511.

BIOINFORMATICS

Major Field: Bioinformatics

Minor Field: M.Sc. student shall take one minor (9 credits of course work) from any of the other fields outside his / her own.

The Total minimum credit requirement of course work for M.Sc. in Bioinformatics is 45 including minor field.

BI 501 Molecular Cell Biology /MBB 503

(3L+0P) I

Objective: To provide insight into fundamentals of cell structure, organization and function

Unit I: General structure and constituent of cells; Similarities and distinction between plant and animal cells; Cell wall, cell membrane, cell surface related function.

Unit II: Structure and function of major organelles: Nucleus, Chloroplasts, Mitochondria, Endoplasmic reticulum, Microbodies, Golgi apparatus, Vacuoles

Unit III: Organellar genomes and their manipulation; Ribosome in relation to cell growth and cell division; Cyto-skeletal elements; Water, protein and ion transport;

Unit IV: Trafficking of biomolecules, Cell division and regulation of cell cycle; Signal transduction mechanisms.

Suggested Readings:

1. Molecular Cell Biology (2007) Bruce Alberts
2. Molecular Cell Biology & Solutions Manual (2007) by Harvey Lodish, Arnold Berk, Chris A. Kaiser, Monty Krieger, et al.
3. Plant Physiology (2008) by Taiz and Zieger

BI 502 / Introduction to Computer Application CA 502

(1L+1P) I

Objective

The course is aimed to provide fundamentals of networking and application protocols with emphasis on developing web based applications.

Theory

UNIT I

Computer organization; Software - System software and Application software.

UNIT II

Networking fundamentals, types of networking, network topology; File Transfer Protocol (FTP), Telnet, Simple Mail Transfer Protocol (SMTP).

UNIT III

Internet basics; Hyper Text Markup Language (HTML).

UNIT IV

Web designing; Web servers.

Practical

Network and mail configuration; Using Network Services; Browsing of Internet; Creation of web pages; Creation of websites using HTML and Creation of websites using DHTML.

Suggested Readings

- Buyens, Jim. 2002. *Microsoft FrontPage -Inside Out*. Microsoft Press.
Cox, V., Wermers L. and Reding E. E. 2006. *HTML Illustrated Complete*. Course Technology.
Niederst, J. 2001. *Web Design in a Nutshell*. O'Reilly Media, Inc.
Tanenbaum, A.S. 2003. *Computer Networks*. Prentice Hall of India, New Delhi.

BI 503 Mathematical Foundations in Computer Application (4L+0P) I CA 551

Objective

This course is designed to give basic foundations in mathematics that are needed to complement and improve the understanding of courses based on algorithm and problem solving.

Theory

UNIT I

Matrix algebra: Basic operations on matrices, Rank and inverse of matrices.

UNIT II

System of linear equations, Characteristic roots and equations, Eigen values and eigen vectors; Basic Differentiation, Integration and Differential Equations; Vector algebra: Double and Triple Product of vectors.

UNIT III

Coordinate geometry: circles and conic sections; Three dimensional geometry: point, straight line, plane and sphere.

UNIT IV

Sets: Set theory, subsets, operations on sets, set cardinality and counting; Functions: Bijective functions, pigeon-hole principle, Boolean functions, permutation functions, Boolean algebra, recursion relations.

UNIT V

Number Theory: Binary arithmetic, exponentiation, induction, sequences, Fibonacci sequence, big-oh notation, GCD, Euclidean algorithm, partially ordered sets, congruence and equivalence relation, encryption scheme, linear homogenous recurrence relations with constant coefficients.

UNIT VI

Graph Theory: Graphs, trees, LAN, Eulerian cycles, Hamiltonian cycles, graph coloring, graph algorithms; Mathematical Logic: Propositional calculus, proposition, logic connectives and compound statements, conjunction, disjunction, truth tables, duality, tautologies and fallacies; Turing Machine: DFA, NFA.

Suggested Readings

- Abertson, M.O. and Hutchinson, J.P. 1988. *Discrete Mathematics with Algorithms*. John Wiley.
Deo, N. 1984. *Graph Theory with Application to Engineering and Computer Science*. Prentice Hall of India, New Delhi.
Knuth, D.E. 1968. *Art of Computer Programming, Vol. I. Fundamental Algorithms*. Addison Wesley.
Tremblay, J.P. and Manohar, R.P. 1975. *Discrete Mathematical Structures with Applications to Computer Science*. McGraw Hill.

BI 504 / Principles of Biotechnology
MBB 501

(3L+0P) I

Objective

To provide insight into basics and application of general biotechnology.

Theory

UNIT I

The structure of DNA; Function of genes and genomes; Restriction enzymes and vectors; Methods of recombinant DNA technology; Nucleic acid hybridization; PCR and its applications.

UNIT II

Molecular markers and their applications; DNA sequencing.

UNIT III

Applications of gene cloning in basic and applied research; Genomics, transcriptomics and proteomics.

UNIT IV

Genetic engineering and transgenics; General application of biotechnology in agriculture, Medicine, Animal husbandry, Environmental remediation, Energy production and Forensics

UNIT V

Public perception of biotechnology; Bio-safety and bioethics issues; Intellectual Property Rights in biotechnology.

Suggested Readings

David P Clark. 2005. *Molecular Biology*

Bruce Alberts. 2008. *Molecular Biology of the Cell*

John, M Walker and Ralph Rapley. 2009. *Molecular Biology and Biotechnology*

B.D.Singh. 2010. *Biotechnology: Expanding Horizons*.

BI -505/ Principles of Computer Programming
CA 561

(2L+1P) I

Objective

The course is aimed to develop problem-solving strategies, techniques and skills, to help students develop the logic, ability to solve the problems efficiently using object oriented programming.

Theory

UNIT I

Techniques of problem solving, Algorithm development, Flowcharting, Stepwise refinement.

UNIT II

Structured programming; Object oriented programming, classes, objects, Abstract data types, Data types, Operators (Arithmetic, Logical and Comparison) and expressions.

UNIT III

Branching and iteration, Arrays, Object/Message paradigm.

UNIT IV

Data encapsulation- modules and interfaces; Polymorphism - Static and dynamic binding, Inheritance: class and object inheritance.

UNIT V

Object oriented software design; Generic and reusable classes, Debugging and testing of programs.

Practical

Programming constructs, control statements: branching and looping, file operations, Creation of classes with features - overloading, inheritance, data abstraction, polymorphism and a case study using and Object oriented language.

Suggested Readings

- Balaguruswamy, E. 1998. *Programming with ANSI C*. Tata McGraw Hill, New Delhi.
Balaguruswamy, E. 2001. *Programming with Object Oriented Programming using C++*. Tata McGraw Hill, New Delhi.
Sethi, R. 1996. *Programming Language Concepts*. Addison Wesley.
Arnold, Ken and Gosling, James 1996. *The Java Programming Language. The Java Series*. Addison Wesley.
Bergin, J. 1994. *Data Abstraction: The Object-Oriented Approach Using C++*. McGraw Hill.
Stroustrup, B. 1997. *The C++ Programming Language*. Addison Wesley.

BI 591

Seminar

(1L+0P)

**BI-506/ Database Management System
CA 566**

(2L+2P) II

Objective

Database systems are backbone of any information system, enterprise resource planning, research activities and other activity that require permanence of data storage. This course provides the basic introduction to database system technologies; design, concurrency, security and backup/recovery issues of database management systems. The major focus in this course is the Relational database model.

Theory

UNIT I

Database system - Operational Data, Characteristics of database approach, architecture.

UNIT II

Overview of DBMS; Data associations - Entities, Attributes and Associations, Relationship among Entities, Representation of Associations and Relationship, Data Model classification.

UNIT III

Entity Relationship model; Relational Data Structure- Relations, Domains and Attributes, Relational Algebra and Operations, Retrieval Operations.

UNIT IV

Relational Database Design - Anomalies in a Database, Normalization Theory, and Normal forms; Query processing and optimization; Security, backup and recovery.

UNIT V

Distributed Databases- concepts, architecture, design; Object Oriented databases; Structured Query Language (SQL) - Data Definition Language (DDL), Data Manipulation Language (DML), Query by example.

UNIT VI

PL/SQL - Stored procedure, Database triggers; Relational Data Base Management Package.

Practical

E-R diagram construction; SQL - Command Syntax, Data types, DDL Statements, DML Statements, integrity constraints; Triggers, creating stored procedures/ functions; Normalization of database and Case study on a database design and implementation.

Suggested Readings

- Date, C.J. 2000. *Introduction to Database System*. Addison Wesley.
Desai, B. C. 2000. *Introduction to Database Systems*. Galgotia Publications, New Delhi.
Elmasri and Navathe. 2006. *Fundamentals of Database Systems*. Addison Wesley.
Garcia-Molina, H., Ullman, J.D. and Widom J. 2002. *Database Systems: The Complete Book*. Prentice Hall.

Rob, P. and Coronel, C. 2006. Database Systems: Design, Implementation and Management. Thomson Learning.
Silberschartz, A., Korth, H. F. and Sudarshan, S. 1997. Database Systems Concepts. Tata McGraw Hill, India.

(Pre-requisite: CA-561)

BI: 507/ Bioinformatics - I (3L+1P) II
AS 571/
GP 540/
MBB 509

Objective: To provide information on basic principles of computational biology and statistical tools used for data analysis

Theory

Unit I:

Basic molecular biology; introduction to the basic principles of structure/function analysis of biological molecules; genome analysis; different types and classification of genome databases (e.g. HTGS, DNA, Protein, EST, STS, SNPs, Unigenes etc.)

Unit II

Statistical Techniques: MANOVA, Cluster analysis, Discriminant analysis, Principal component analysis, Principal coordinate analysis, Multidimensional scaling; Multiple regression analysis; Likelihood approach in estimation and testing; Resampling techniques – Bootstrapping and Jack-knifing; Markov Models. Hidden Markov Models, Bayesian estimation and Gibbs sampling

Unit III: DNA sequence retrieval system, various DNA and protein sequence file formats, Basic concepts of similarity searching and sequence alignments, pair wise and multiple sequence alignments, DNA sequence analysis, different gene prediction models and gene annotation tools,

Unit IV: Protein sequence analysis and structure prediction, comparative genome analysis, phylogenetic analysis, gene expression analysis tools, programming languages and their applications in bioinformatics.

Practicals

Different types of databases and database search and retrieval, DNA and protein sequence analysis, Similarity searching and multiple alignments, Gene annotation, Phylogenetic analysis, Sequence analysis, Protein structure prediction, Analysis of microarray data, Programming languages in bioinformatics

Suggested Readings:

DNA and protein sequence analysis. A Practical approach (1997) by Bishop M.J., Rawlings C.J. (Eds.)

Bioinformatics Basics: Applications in Biological Science and Medicine (2005) By Hooman Rashidi, Lukas K. Buehler

Bioinformatics in the Post-Genomic Era: Genome, Transcriptome, Proteome, and Information-Based Medicine (2004) By Jeffrey Augen

Frontiers in Computational Genomics (2003) Edited by: Michael Y. Galperin and Eugene V. Koonin

BI 508 / Protein Biosynthesis (3L+0P) II
BIO 602

Objective

To impart knowledge about the various components and processes involved in protein biosynthesis its regulation and the significance of post-translational modifications.

Theory

Unit-I

RNA world; Diverse RNA functions in living cells and its significance.

Unit-II

Structure and function of tRNA, rRNA, mRNA; Pre mRNA splicing, tRNA processing, modification and transport.

Unit-III

Structure and function of amino acyl tRNA synthetases; tRNA identity; recognition and charging; proof reading mechanisms.

Unit-IV

Protein synthesis: structure and function of ribosomes; Genetic code: Elucidation, nature and properties; Initiation, elongation and termination cycles in prokaryotes and eukaryotes, Protein synthesis inhibitors and regulation.

Unit-V

Secretion and maturation of polypeptides: Signal sequences and secretion; Spontaneous and Chaperone mediated folding and transport to organelles like chloroplast, mitochondria and nucleus; Post translational modifications and their significance.

Suggested Readings:

- Alberts *et.al* (2006). *Molecular Biology of the Cell*. 6th Edition Garland Publ.
Freifelder, D. (Ed.) (1978). *Recombinant DNA* (Readings from Scientific American W.H.Freeman & Co. San Fransico).
Voet D, Voet JD, Prat CW 2007, *Fundamentals of Biochemistry*, John Wiley 3. G.
Zubay GL. 1998. *Biochemistry*. 4th Ed. WCB London
B. Lewin (2008). *Genes X*, Oxford Univ. Press.

BI 509/ Genomics MBB 602

(3L+0P) III

Objective: To provide insight into the functional aspects of cell function by studying the genome as a whole with special emphasis on structural and functional genomics

UNIT I:

Structural genomics: Classical ways of genome analysis, large fragment genomic libraries, physical mapping of genomes, genome sequencing, sequence assembly and annotation, comparative genomics

UNIT II:

Functional genomics: DNA chips and their use in transcriptome analysis, Mutants and RNAi in functional genomics, qPCR, SAGE, MPSS

UNIT III:

Application of genomics in crop improvement, protein structure and function, proteins as enzymes, protein purification, 2D, mass spectrometry

Suggested Readings:

- Genomes (2006) by TA Brown
Evolutionary Genomics and Systems Biology (2010) by Gustavo Caetano
Principles of Gene Manipulation and Genomics (2006) by Sandy B. Primrose and Richard Twyman
New and Emerging Proteomic Techniques (2010) by D. Nedelkov and R. Nelson

BI 510 Biological Databases and Data Analysis

(2L+1P) III

Objective: To know about different biological databases existing in the public domain and perform analysis on the data available in them

UNIT I:

Nature of biological data; Overview of available Bioinformatics resources on the web; NCBI/EBI/EXPASY etc; Biological Databases: Nucleic acid sequence

databases; GenBank/EMBL/DDBJ; Biological Databases: Protein sequence databases; PIR-PSD; SwissProt, UniProtKB; Database search engines: Entrez, SRS.

UNIT II:

Overview/concepts in sequence analysis; Pairwise sequence alignment algorithms: Needleman & Wunsch, Smith & waterman ; Scoring matrices for Nucleic acids and proteins: MDM, BLOSUM, CSW; Database Similarity Searches: BLAST, FASTA; Multiple sequence alignment: PRAS, CLUSTALW; Biological databases: Genome & genetic disorders;

UNIT III:

Genome databases: Human, model organisms, microbes & viral: OMIM; Biological databases: structural databases: PDB, NDB, CCSD; Derived databases: Prosite, BLOCKS, Pfam/Prodom.

Practicals:

Nucleic acid sequence databases, Protein sequence databases, Database search engines, Database Similarity Searches, Multiple sequence alignment, Genome databases, Structural databases, Derived databases

Suggested Readings:

Bioinformatics: A Practical Guide to the analysis of Genes and Proteins (2nd Ed.) by Baxevanis, A.D. & Ouellette, B., F. F., New York, John Wiley & Sons, Inc. Publications, 2002.

Introduction to Bioinformatics by Attwood, T.K. & Parry-Smith, D.J., Delhi, Pearson Education (Singapore) Pte.Ltd., 2001.

Bioinformatics: Sequence and Genome Analysis by Mount, David, New York, Cold Spring Harbor Laboratory Press, 2004.

Current Protocols in Bioinformatics by Baxevanis, A.D., Davison, D.B., Page, R. D. M. & Petsko, G.A., New York, John Wiley & Sons Inc., 2004.

BI 511 RNA/Protein Structure Prediction and Molecular Modeling (1L+2P) III

Objective: To get insight into various techniques and tools available for protein structure prediction, visualization and validation.

UNIT I:

Structural data, databases and structure analysis: Exploring the Database searches on PDB and CSD, WHATIF Molecular visualization tools; Visualization of tertiary structures, quaternary structures, architectures and topologies of proteins and DNA using molecular visualization softwares such as RasMol, Cn3D, SPDBV, Chime, Mol4D, etc;

UNIT II:

Structure prediction tools and homology modeling: Prediction of secondary structures of proteins using different methods with analysis and interpretation of the results; Comparison of the performance of the different methods for various classes of proteins. (Fasman method, Garnier Osguthorpe Robson (GOR), Neural Network based; methods); NLP approach for secondary structure prediction of RNA; Introduction to mfold and Vienna packages; Prediction of tertiary structures of proteins using Homology Modeling approach: SWISSMODEL, SWISS-PDB Viewer; Prediction of tertiary structures of proteins different methods for fold recognition along with analysis and interpretation of results (Threading techniques; Homology Modeling and *abinitio* methods);

UNIT III:

Molecular dynamics simulation and docking: Basic principles of theoretical modeling, Empirical force fields for biomolecular simulations, Energy minimization, Molecular dynamics, Monte Carlo simulation Peptide building (PYMOL / DStools).

Practicals:

Structural data, databases and structure analysis, Molecular visualization tools, Structure prediction tools and homology modeling, Molecular dynamics simulation and docking

Suggested Readings:

Wilkins, M.R., Williams, K.L., Appel, R.D., Hochstrasser, D.F. (Editors) 1997 Proteome Research: New Frontiers in Functional Genomics. Springer Verlag Berlin Heidelberg.

Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins by Baxevanis, A.D. and Francis Ouellette, B.F., Second Edition, Wiley, 2004.

Fundamentals of Molecular Evolution by Graur, D. and Li, W-H., Sinauer Ass., USA, 2000.

Protein structure prediction: methods and protocols by Webster, D. M., Ed. Totowa Humana Press, 2000.

Modular protein domains by Gimona, G. Cesareni. & Yaffe, M. Sudol (EDS.), USA., Wiley-vch verlag gmbh & co.,3-527-30813-X , Aug. 2004.

Molecular modeling: basic principles and applications by Holtje, H.D. & Folkers, G., Weinheim, VCH, 1997

Molecular Modeling: Basic Principles and application by Hans Dieter & Didier Rognan, Wiley VeH Gmbh and Co. KGA, 2003

BI 512/ Bioinformatics – II

(2L+1P) III

AS 608

Objective: To aim at exposing the students to advanced statistical and computational techniques related to bioinformatics. The course would prepare the students in understanding bioinformatics principles and their applications.

Theory

UNIT- I:

Genomic databases and analysis of high-throughput data sets, Analysis of DNA sequence, Sequence annotation, ESTs, SNPs. BLAST and related sequence comparison methods. EM algorithm and other statistical methods to discover common motifs in biosequences. Multiple alignment and database search using motif models, ClustalW and others. Concepts in phylogeny. Gene prediction based on codons, Decision trees, Classificatory analysis, Neural Networks, Genetic algorithms, Pattern recognition, Hidden Markov models.

UNIT- II:

Computational analysis of protein sequence, structure and function. Modeling protein families. Expression profiling by microarray/gene chip, proteomics etc., Multiple alignment of protein sequences, Modeling and prediction of structure of proteins, Designer proteins, Drug designing.

UNIT- III:

Markov chains (MC with no absorbing states; Higher order Markov dependence; patterns in sequences; Markov chain Monte Carlo – Hastings-Metropolis algorithm, Simulated Annealing, MC with absorbing States), Bayesian techniques and use of Gibbs Sampling, Advanced topics in design and Analysis of DNA microarray experiments.

UNIT- IV:

Computationally intensive methods (Classical estimation methods, Bootstrap estimation and Confidence Intervals, Hypothesis testing, Multiple Hypothesis testing), Evolutionary models (Models of Nucleotide substitution), Phylogenetic tree estimation (Distances: Tree reconstruction – Ultrametric and Neighbor-Joining cases, Surrogate distances, Tree reconstruction, Parsimony and Maximum Likelihood, Modeling, Estimation and Hypothesis Testing), Neural Networks (Universal Approximation Properties, Priors and Likelihoods, Learning Algorithms – Back propagation, Sequence encoding and output interpretation, Prediction of Protein Secondary Structure, Prediction of Signal Peptides and their cleavage sites, Application for DNA and RNA Nucleotide Sequences), Analysis of SNPs and Haplotypes.

(Pre-requisite: AS 571)

Practical

Genomic databases and analysis of high-throughput data sets, BLAST and related sequence comparison methods, Statistical methods to discover common motifs in biosequences, Multiple alignment and database search using motif models, ClustalW, Classificatory analysis, Neural Networks, Genetic algorithms, Pattern recognition, Hidden Markov models, Computational analysis of protein sequence, Expression profiling by microarray/gene chip, proteomics, Modelling and prediction of structure of proteins, Bayesian techniques and use of Gibbs Sampling, Analysis of DNA microarray experiments, Analysis of one DNA sequence, Analysis of multiple DNA or protein sequences, Computationally intensive methods, Multiple Hypothesis testing, Phylogenetic tree estimation, Analysis of SNPs and Haplotypes.

Suggested Readings

- Baldi, P. and Brunak, S. 2001. *Bioinformatics: The Machine Learning Approach*. MIT Press.
- Baxevanis, A.D. and Francis, B.F. 2004. *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*. John Wiley.
- Duda, R.O., Hart, P.E. and Stork, D.G. 1999. *Pattern Classification*. John Wiley.
- Ewens, W.J. and Grant, G.R. 2001. *Statistical Methods in Bioinformatics*. Springer.
- Jones, N.C. and Pevzner, P.A. 2004. *Introduction to Bioinformatics Algorithms*. The MIT Press.
- Koskinen, T. 2001. *Hidden Markov Models for Bioinformatics*. Kluwer Academic Publishers.
- Krane, D.E. and Raymer, M.L. 2002. *Fundamental Concepts of Bio-informatics*. Benjamin / Cummings.
- Krawetz, S.A. and Womble, D.D. 2003. *Introduction to Bioinformatics: A Theoretical and Practical Approach*. Humana Press.
- Lesk, A.M. 2002. *Introduction to Bio-informatics*. Oxford University Press.
- Linder, E. and Seefeld, K. 2005. *R for Bioinformatics*. O'Reilly and Associates.
- Percus, J.K. 2001. *Mathematics of Genome Analysis*. Cambridge University Press.
- Sorensen, D. and Gianola, D. 2002. *Likelihood, Bayesian and MCMC Methods in Genetics*. Springer.
- Tisdall, J.D. 2001. *Mastering Perl for Bioinformatics*. O'Reilly and Associates.
- Wang, J.T.L., Zaki, M.J., Toivonen, H.T.T. and Shasha, D. 2004. *Data Mining in Bioinformatics*. Springer.
- Wu, C.H. and McLarty, J.W. 2000. *Neural Networks and Genome Informatics*. Elsevier.
- Wunschiers, R. 2004. *Computational Biology Unix/Linux, Data Processing and Programming*. Springer.
- Yang, M.C.C. 2000. *Introduction to Statistical Methods in Modern Genetics*. Taylor and Francis.

BI 523 Advanced Techniques for Sequence and Structure Analysis (1L+2P) I

Objective: To teach various approaches meant for sequence alignment and docking simulation in proteomics.

UNIT- I:

Advanced Techniques for Sequence Analysis: Sequence Profiles: Derivation, Databases, Application, Gapped BLAST, PSI-BLAST, PHI-BLAST;

UNIT- II:

Advanced Techniques for Structure Analysis: Molecular replacement method, Direct method & Fiber diffraction, Methods for Comparison of 3D structures; Application of the optimization techniques: Sequence Alignments, Prediction of Protein Structure, Docking Simulations; Advance techniques in Prediction of 3D Structure: Rossetta Stone Hidden Markov Model, Neural networks.

UNIT- III:

Molecular Dynamics Simulations & Monte Carlo Methods: Electrostatics of biomolecules, Simulations of Bio-macromolecular Structures in Water & membrane, Free energy perturbation method; Simulated Annealing: Multiple Sequence Alignments, Simulations of Bio-macromolecular Structures; Designing of molecules like drug, inhibitors using: Structure based and ligand based docking methods, Different Scoring schemes.

Practicals:

Advanced Techniques for Sequence Analysis, Molecular replacement method, Methods for Comparison of 3D structures, Optimization techniques: Sequence Alignments, Prediction of Protein Structure, Docking Simulations; Advance techniques in Prediction of 3D Structure.

Suggested Readings:

Hidden markov models for Bioinformatics by Koski, T, Dordrecht Kluwer, Academic Publishers, 2001.

Neural Networks: A Comprehensive Foundation (2nd Edition) by Simon Haykin.

Protein Bioinformatics: An Algorithmic Approach to Sequence and Structure Analysis by Ingvar Eidhammer, Inge Jonassen, William R. Taylor John Wiley, 2003.

Genetic algorithms in molecular modeling by Devillers, J., Ed, 1996.

Simulation Modeling and Analysis by Averill M. Law d and W. David Kelton, Tata MacGraw – Hill, 2000.

BI 524 Tools and Techniques for biological data mining (2L+1P) I

Objective: To understand various algorithms of machine learning approaches

UNIT- I:

Quality of Biological Data & Data Accuracy; General issues regarding Biological Databases: Representation of errors due to (machines, 3D structural and sequence data of proteins and nucleic acid, Proteomics and Micro array data);

UNIT- II:

Optimization Techniques: Steepest Descent, Conjugate Gradient, Newton-Raphson, Simulated annealing in Biomolecular Structure Optimization; Genetic Algorithms: *Ab initio* methods for structure prediction; Lattice, SOM, etc., Information theory, entropy and relative entropy, Stochastic Grammars & natural languages processing techniques;

UNIT- III:

Clustering & Classification Algorithms: Hierarchical and non-hierarchical Clustering, K-Means clustering, Grid based clustering, Analysis of MD trajectories, Protein Array data Analysis;

UNIT- IV:

Dynamic Programming and application in bioinformatics: Sequence Alignments, Structure Alignments; Foundations for Machine learning Techniques: Hidden Markov Model, Neural Network, Bayesian modeling, The Cox-Jaynes Axiomes; Support Vector machine & Ant colony optimization: Multiple Sequence Alignments, Biomolecular Structure Prediction; Fuzzy logic system & application in bioinformatics; Introduction to WEKA package; Clustering and classifications, Protein Array data Analysis.

Suggested Readings:

Data Mining: Concepts and Techniques by Han and Kamber, Morgan Kaufmann.
Data Mining: Practical Machine Learning Tools and Techniques by Witten and Frank, Elsevier.
Exploration and analysis of DNA microarray and protein array data by Amaratunga, D. & Cabrera, J. New Jersey. John Wiley & Sons Inc., 2004.
Data mining: introductory and advanced topics by Dunham, M.H.: New Delhi, Pearson Education, 2003
Fuzzy sets and fuzzy logic: theory and applications by Klir, G.J. & Yuan Bo, New Delhi. Printice Hall of India, 2002. 81-203-1136-1.

BI 525 Advanced Programming in Bioinformatics (2L+1P) I

Objective: To learn programming skills for parsing biological data, database connectivity and web-interface

Unit – I: Perl: Introduction, Scalar Data, Arrays and List Data, Control Structures Hashes, Regular Expressions; Subroutines, File handles and File Tests; Function: Formats, Directory Access, Process Management, Other Data Transformation: Finding a Sub string, Extracting and Replacing a Sub string, Formatting Data: Sorting, Transliteration Database Manipulation: DBM Databases and DBM Hashes, Opening and Closing DBM Hashes, Fixed-Length Random-Access Databases, Variable-Length (Text) Databases, Win32 Database Interfaces

Unit – II: CGI Programming: The CGI Module, Your CGI Program in Context, Simplest CGI Program, Passing Parameters via CGI, Perl and the Web, Object oriented perl: Introduction to modules, Creating Objects; Bioperl: Introduction, Installation procedures, Architecture, Uses of bioperl; introduction to Python/ CORBA.

Suggested Readings:

Beginning Perl for Bioinformatics by James Tisdall, O-Reilly.
Learning Perl by Randal L. Schwartz, Tom Phoenix, O-Reilly.
Programming the Perl DBI by Alligator Descartes, Tim Bunce, O-Reilly.
Advanced Perl Programming by Sriram Srinivasan, O-Reilly.
Client/Server Programming with JAVA and CORBA by Orfali and Harkey, Wiley .

BI 591 Seminar (1L+0P)

BI 526

Comparative Genomics

(1L+1P) II

Objective: To understand comparative genomics analysis among various species and their role in nucleotide variations.

Unit I:

Objective and Overview of Genome Comparisons; Genome Alignments: BLAST2, MUMmer, PipMaker, VISTA, Gene Order; Comparative Genomics: Synteny among Prokaryotes and Eukaryotes;

Unit II:

Comparative Genomics Databases: COG, VirGen, CORG, HOBACGEN, Homophila, XREFdb, Gramene; Single Nucleotide Polymorphisms: dbSNP and other SNP-related databases;

Suggested Readings:

Bioinformatics: sequence and genome analysis by David Mount, cold springer harbour press, 2004.

Comparative genomics: empirical and analytical approaches to gene order dynamics, map alignment and the evolution of gene families by Sankoff, D. & Nadeau, J.H., Netherlands, Kluwer Academic Publishers, 2000.

Campbell, A.M. & Heyer, L.J. 2002 Discovering Genomics, Proteomics and Bioinformatics. Benjamin/Cummings.

Bioinformatics approach Guide to the analysis of genes and proteins by Andreas Baxevanis and B.F. Francis Ouellette. John Wiley, 2004

BI 527

Phylogenetic Analysis

(1L+1P) II

Objective: To find out the evolutionary relationship among various species by using different phylogenetic algorithms.

Unit I:

Phylogenetic trees and their comparison: Definition and description, various types of trees; Consensus (strict, semi-strict, Adams, majority rule, Nelson); Data partitioning and combination. Tree to tree distances, similarity; Phylogenetic analysis algorithms: Maximum Parsimony, Distance-based: UPGMA, Transformed Distance, Neighbors-Relation, Neighbor-Joining,

Unit II:

Probabilistic models of evolution, Maximum likelihood algorithm ; Approaches for tree reconstruction: Character optimization; delayed and accelerated transformation, Reliability of trees, Bootstrap, jackknife, decay, randomization tests; Applications of phylogeny analyses: Comparison of Phylogenetic Trees obtained using DNA seq. Vs. protein seq. Vs. Full genomes. Need for addition of other properties towards total phylogenetic analysis, Comparative methods for detection of species / organism relationships, Gene duplication, Horizontal transfer, Domain evolution, Study of co-evolution: Plant-insect interactions. Host-parasite interactions, Viral evolution.

Suggested Readings:

Comparative genomics: empirical and analytical approaches to gene order dynamics, map alignment and the evolution of gene families by Sankoff, D. & Nadeau, J.H., Netherlands, Kluwer Academic Publishers, 2000.

Molecular Evolution and Phylogenetics by Nei, M. and Kumar, S., Oxford University Press, 2000.

Phylogenetic Trees Made Easy: A How to Manual for Molecular Biologists by Hall, B.G. Sinauer Ass., USA, 2001.

Molecular Evolution and Phylogenetics by Nei, M. and Kumar, S., Oxford University Press, 2000.

BI 528

Chemo informatics and IPR issues

(1L+1P) II

Objective: To get insight into chemoinformatics and its role in drug discovery.

Unit – I:

Chemo informatics: History, Current activities and Challenges in the Chemo informatics, Chemical information and sources, Major chemical databases and information retrieval, Chemical Structure drawing tools.

Unit – II:

Chemo informatics and drug discovery, Pharmacodynamics and pharmacokinetics, Drug potency and Efficacy, Docking, Active site, Absorption, Distribution, Development of a drug: Classical steps, Chemical Parameters in drug design, Structure based drug discovery, Quantitative Structure Activity Relationships.

Unit – III:

IPR issues: Definitions, Production of plant varieties and farmer's rights authority and registry. Registration of plant varieties and essentially derived varieties. Duration and effect of registration of benefit sharing. Surrender and revocation of certificate and recertification and correction of register, farmer's rights, Compulsory license, Plant varieties protection appellate tribunal, Finance, accounts and audit, Infringement, offences, penalties and procedure.

Suggested Readings:

Chemoinformatics by Johann Gasteiger and Thomas Engel, 2004.

An introduction to Chemoinformatics by Andrew R. Leach and Valerie J. Gillet, Kluwer Academic Publisher, 2003.

Handbook of Chemoinformatics. From Data to Knowledge by Johann Gasteiger.

Chemometrics and Chemoinformatics by Barry K. Lavine, ACS Symposium series 894.

BI 591

Seminar

(1L+0P)