

Faculty of Science

Revised Syllabus for the

# M.Sc. Bioinformatics Credit Systems

(From Academic year: 2010 - 2011)



University of Pune  
Department of Bioinformatics  
Pune 411 007

## Present Course Structure of M. Sc. (Bioinformatics)

(T = Theory course, P = Practical course)

(Revised on April 26, 2010)

### Semester – I

<i>Course No.</i>	<i>Course Name</i>	<i>Total Credits</i>
BIM 101 (T)*	Basic Biology	2.5+0
BIM 102 (T+P)*	Mathematics For Bioinformatics	2+0.5
BIM 103 (T+P)	Statistical Techniques	2+0.5
BIM 104 (T)	Biological Chemistry	3+0
BIM 105 (T)	Genetic Information Flow & Processing	3+0
BIM 106 (T+P)	Basic Concepts in Computing	1+2
BIM 107 (T+P)	Introduction to Database Systems	2+2
BIM 108 (T+P)	Biological Databases and Data Analysis	2+2
BIM 109 (P)	Programming in Perl	0+3
		15+10
	Total credits	25

\* Students from Mathematics Stream will take BIM 101(T), while those from the Biology Stream will take BIM 102(T)

### Semester – II

<i>Course No.</i>	<i>Course Name</i>	<i>Total Credits T+P</i>
BIM 201 (T)	Cell Biology, Genetics & Genomics	3+0
BIM 202 (T)	Structural Biology & Bioinformatics	3+0
BIM 203 (T+P)	Chemoinformatics	2+1
BIM 204 (T+P)	Immunology and Bioinformatics	2+1
BIM 205 (T+P)	Programming in Object Oriented Languages	2+2
BIM 206 (P)	Programming in C	0+3
BIM 207 (P)	Computational Structural Biology	0+2
BIM 208 (P)	Laboratory Techniques in Modern Biology	0+2
BIM 209 (P)	UNIX & Linux Commands & Scripting	0+1
		12+12
	Total credits	24

**Semester – III**

<i>Course No.</i>	<i>Course Name</i>	<i>Total Credits</i>
BIM 301 (T+P)	Molecular Phylogenetics	1+1
BIM 302 (T+P)	Genome to Drug and Vaccine	3+2
BIM 303 (T+P)	Comparative Genomics and Proteomics	2+2
BIM 304 (T+P)	Molecular Modeling & Simulations	2+2
BIM 305 (P)	Database design & web-implementation	0+1
BIM 306 (T+P)	Scientific Data Visualization	1+2
BIM 307 (T+P)	Tools & Techniques for Biological Data Mining	2+2
BIM 308 (T)	Project (Phase I) & Review	2+0
		13+12
	Total credits	25

**Semester – IV**

<i>Course No.</i>	<i>Course Name</i>	<i>Total Credits</i>
BIM 401 (T+P)	Advanced Techniques for Sequence and Structure Analysis	2+2
BIM 402 (T+P)	Metabolic Engineering & Systems Biology	3+1
BIM 403 (T)	Emerging Areas in Bioinformatics	1+0
BIM 404 (P)	Project work	0+16
Any one of the following		
BIM 405 (T+P)	Seminars on Applications of Bioinformatics in Agriculture	1
BIM 406 (T+P)	Seminars on Applications of Bioinformatics in Human Health	1
BIM 407 (T+P)	Seminars on Applications of Bioinformatics in Environment	1
BIM 408 (T+P)	Seminars on Applications of Bioinformatics in Biotechnology	1
BIM 409 (T+P)	Seminars on Applications of Bioinformatics Molecular Biology	1
BIM 410 (T+P)	Seminars on Applications of Bioinformatics in Neurobiology	1
BIM 411 (T+P)	Seminars on Applications of Bioinformatics Drug Designing	1
BIM 412 (T+P)	Seminars on Applications of Bioinformatics in Veterinary Sciences	1+0
		7+19
	Total credits	26

(Numbers in the brackets indicate total credits for respective courses and also indicates number of hours of classroom teaching for respective theory courses per week. Credits for practical courses indicate number of laboratory sessions per week (each session is of 3 hours).

## Semester – I

### **BIM 101 (T): Basic Biology<sup>1</sup>**

**(2.5 credits)**

#### ***Objectives:***

This course will enable the students to:

- appreciate the basic nature and diversity of microbial, plant and animal life.
- understand the classification of organisms, their structure and function.

#### ***Theory***

##### ***Syllabus:***

- Origin of life: an introduction (2)
- Viruses: Biology of viruses; bacteriophages, plant and animal viruses (2)
- Level of organization: prokaryotic and eukaryotic cells, multicellularity (2)
- Five kingdom classification: salient features and outline classification with examples
  - Monera (3)
  - Protista (2)
  - Mycota (3)
  - Plantae (7)
  - Animalia (9)

#### ***References:***

- Wallace Robert A., Sanders Gerald P., Ferl Robert J. The science of life. Publisher: New York, NY : Harper Collins, 1991. ISBN: 0673380440.
- Solomon Eldra P., Berg Linda R., Martin Diana W. Biology 6<sup>th</sup> edition. Publisher: Pacific Grove, CA, Brooks/Cole Thomson Learning, 2002. ISBN: 0030335035.

### **BIM 102 (T+P) : Mathematics for Bioinformatics<sup>1</sup>**

**(2T+0.5P credits)**

#### ***Objectives:***

This course will enable the students to:

- achieve skills in mathematics that are essential for application in bioinformatics.

#### ***Theory***

##### ***Syllabus:***

- Trigonometry: (4)  
Trigonometric Functions, Series Expansion, Inverse, General Values, Graphs, Taylor series
- Calculus: (10)  
Limits, Continuity, Analysis, Differentiation (1D & Partial), Reimann Integration, Definite Integrals.

---

<sup>1</sup> Students from Mathematics Stream will take BIM 101(T), while those from the Biology Stream will take BIM 102(T)

- Ordinary & Partial Differential Equation (5)  
1'st Order & 2'nd Order Ordinary Differential Equations. Self Adjoint Equations, Method of Separation of Variables
- Vector & Matrices (3)  
Vector Algebra, Vector Calculus, Basic Computations, Matrices
- Introduction to Set Theory (2)
- Integral transform (6)  
Fourier Series, Fourier Transform, Laplace Transform

***Practicals***

***Objectives:***

This course will enable the students to:

- achieve skills in mathematics that are essential for application in bioinformatics.

***Syllabus:***

- Assignments/Tutorials based on theory topics as listed above: 8 (+2) sessions of 2 hours each (6<sup>th</sup> week onwards)

***References:***

- Isaev Alexander. Introduction to Mathematical Methods in Bioinformatics. Publisher: Berlin; New York: Springer, 2004. ISBN: 3540219730
- Raman K. V. & Pal Sourav. Mathematics in Chemistry. New Delhi, Vikas Publishing House Pvt Ltd., 2004. ISBN: 8125912886
- Jones D.S., Sleeman B.D.. Differential Equations and Mathematical Biology  
Publisher : Chapman & Hall. 2003. ISBN:1584882964
- Bracewell Ronald. The Fourier transform and it's applications 3<sup>rd</sup> edition. Publisher: New Delhi : McGraw Hill, 2000. ISBN: 0073039381.
- Stephenson G., Radmore P. M.. Advanced Mathematical Methods for Engineering and Science Students Cambridge: Cambridge University Press, 1990.
- Arfken George. Mathematical methods for physicists. Publisher: Orlando : Academic Press 1985. ISBN: 0120598205.

**BIM 103 (T+P): Statistical Techniques**

**(2T+0.5P credits)**

***Objectives:***

This course will enable the students to:

- understand and apply statistical techniques that are essential to process and interpret biological data.

***Theory***

***Syllabus:***

- Introduction to principles of statistical sampling from a population (2)
- Frequency Distributions and Statistical Measures: mean, mode, median, variance, standard deviation, coefficient of variation, measures of skewness and kurtosis (6)
- Introduction to theory of Probability, Conditional Probability, Bayesian Rules, Random variable, Distributions of random variables, Binomial, Poisson, Geometric, Normal and extreme value distribution, Vector valued Random Variables (8)

- Hypothesis testing: (10)
  - Test of significance viz. Z test, t test, pair t test,  $\chi^2$  test of goodness of fit
  - Non parametric tests: median test, sign test, Kruscal-Wallis test
- Basic of Cluster Analysis (4)
  - Search using stem numbers
  - Search using text signatures

***Practicals***

***Objectives:***

This course will enable the students to:

- understand and apply statistical techniques that are essential to process and interpret biological data

***Syllabus:***

- Assignments/Tutorials based on topics listed below: 8 (+2) sessions of 2 hours each (6<sup>th</sup> week onwards)
- Using any of the software like R programming/SAS etc. exercises will be done on -
  - Computation of Statistical Measures using given data (2)
  - Fitting distributions to given data (3)
  - PAM, BLOSUM, Substitution Rates etc. (1)
  - Standard parametric tests (1)
  - Non parametric tests (1)

***References:***

- Mathur Sunil K. Statistical bioinformatics with R. Publisher : Academic Press, 2010 ISBN: 9780123751041.
- Pal Nabendu, Sarkar Sahadeb. Statistics: Concepts and Applications. Publisher: PHI Learning Pvt. Ltd., 2005. ISBN: 8120326792.
- Gentle, James E.; Härdle, Wolfgang K. Mori, Yuichi (Eds.). Handbook of Computational Statistics Concepts and Methods. Publisher Springer 2004. ISBN: 354040464.
- Lange Kenneth. Mathematical and statistical methods for genetic analysis. Publisher: New York, Springer, 2002. ISBN: 8181281136.
- Joseph Glaz, Joseph Naus, Sylvan Wallenstein. Scan Statistics. Publisher: New York; Springer, 2001. ISBN: 038798819X.
- Murray R. Spiegel, Larry J. Stephens Schaum's Outline of Statistics 3rd edition Publisher: McGraw-Hill New Delhi 3<sup>rd</sup> edition 2000. ISBN:0070435103.
- Schaum's Outline of Introduction to Probability and Statistics. Publisher: McGraw-Hill, 1999. ISBN: 0071164936.
- Schaum's Outline of Statistics.
- Bilal M. Ayyub, McCuen Richard H.. Probability, statistics, & reliability for engineers. Publisher: Boca Raton : CRC Press, 1997. ISBN: 0849326907.
- Frank Harry, Althoen Steven C.. Statistics: Concepts and Applications. Publisher: Cambridge University Press. 1995. ISBN : 052144554X.
- Colin Richard Campbell Statistics for biologists. Publisher: Cambridge University Press, 1989 ISBN: 0521369320.

**Objectives:**

This course will enable the students to:

- be familiar with the chemistry of different classes of biomolecules and their interactions in an aqueous environment
- understand the structure-function relationships of macromolecules, the principles of enzyme catalysis and regulation,
- understand organization of metabolic systems with an in-depth account of the pathways for the complete oxidation of glucose.

**Theory****Syllabus:**

- Water as the universal biological solvent; concepts of pH, buffer and osmolarity (2)
- Carbohydrates: monosaccharides, oligosaccharides, polysaccharides, glycosaminoglycans, proteoglycans and glycoproteins (3)
- Lipids: fatty acids, acylglycerols; phospholipids, sphingolipids, cholesterol and membranes; Isoprenoids, icosanoids and their biological importance. (4)
- Nucleic acids: bases, nucleotides, RNA and DNA; different structural forms of DNA; different types of RNA. (3)
- Proteins: amino acids and peptides; primary, secondary, tertiary and quaternary structures; structure, function and evolutionary relationships; protein – protein interactions, protein folding; allosteric proteins. (8)
- Enzymes: enzyme nomenclature and classification; units of enzyme activity; specificity; coenzymes and metal cofactors; temperature and pH effects; Michaelis-Menten kinetics; Inhibitors and activators; active site and catalytic mechanisms; covalent and non-covalent regulations; isoenzymes. (15)
- Organization of metabolic systems: enzyme chains, multi-enzyme complexes and multifunctional enzymes; anaplerotic sequences and amphibolic pathways; pacemaker enzymes and feedback control of metabolic pathways; shuttle pathways; energy charge. (3)
- Oxidation of glucose in cells: high energy bond, glycolysis, citric acid cycle and oxidative phosphorylation. (7)

**References:**

- Nelson David L., Cox Michale. Lehninger Principles of Biochemistry 5<sup>th</sup> Edition. Publisher: New York. W. H. Freeman. 2008. ISBN 978 0716771081.
- Berg, Jeremy M, Tymoczko, John L. Stryer, Lubert. Biochemistry 6<sup>th</sup> Edition. Publisher: New York : W.H. Freeman. 2007.ISBN: 071676766X.
- Hames David, Hooper Nigel. Instant Notes in Biochemistry 3<sup>rd</sup> Edition. Publisher . Nodia, Taylor & Francis. 2007. ISBN: 185996 2491.
- Voet, Donald, Voe Judith, Pratt, Charlotte W. Fundamentals of Biochemistry: Life at the molecular Level 2<sup>nd</sup> Edition. Publisher: Asia, John Wiley & Sons. 2006. ISBN: 0471753416.
- Horton, Robert, Moran, Laurence A, Scrimgeour, Perry Gray Marc, Rawn.David. Principles of biochemistry. Publisher: New Jersey, Pearson Prentice Hall, 2006. ISBN: 0131453068.

- Mathews, Christopher K. van Holde, K. E., Ahern Kevin G. Biochemistry. Publisher: San Francisco, Pearson Education 2000. ISBN: 81 29702150.
- Zubay, Geoffrey. Biochemistry 4<sup>th</sup> Edition, Publisher: Boston, Wm C. Brown, 1998, ISBN 0697219003.
- Garrett, Reginald H. , Grisham, Charles M. Principles of biochemistry: with a human focus. Publisher: Australia Brooks/Cole, Thomson Learning, 1997. ISBN:0030973694.

## **BIM105 (T): Genetic Information Flow & Processing**

**(3T credits)**

### ***Objectives:***

This course will enable the students to:

- understand the current concepts in gene organization, transcription, translation and regulation of gene function
- know biotechnological implications of recent developments in cloning and genome sequencing.

### ***Theory***

#### ***Syllabus:***

- Introduction: DNA as a genetic material – Experiments done to prove this (1)
- Nucleic acid structure: single stranded & double stranded; denaturation, renaturation and hybridization of DNA; secondary structures in single stranded molecules; alternative double helical structures in double stranded DNA; Closed DNA as supercoiled molecule (2)
- Genome organization: Prokaryotic and eukaryotic genomes – C value paradox, repetitive and non-repetitive DNA., transposons and retroposons; Exons and introns – organization of interrupted genes, one DNA sequence may code for multiple proteins; Gene numbers – essential genes and total gene number, gene clusters, pseudogenes; Gene families – globin and rDNA gene families; Organelle genome – mitochondrial and chloroplast. (6)
- Packaging of genome – Bacterial genome as nucleoid; Eukaryotic genome – nucleosomes, chromatin, solenoids, loops, domains, scaffolds, chromosomes. (3)
- Perpetuation of DNA: Prokaryotic DNA replication – DNA polymerases, origin of replication, initiation, elongation and termination of replication. Rolling circle model of replication; Eukaryotic DNA polymerases – multiple origins of replication, process of replication; Regulation of replication in both prokaryotes and eukaryotes. (5)
- DNA damage, repair and recombination: Different types of DNA damages; Variety of DNA repair systems in prokaryotes and eukaryotes – Base excision repair system, Nucleotide excision repair system, Mismatch repair system, Recombination repair system; Recombination – homologous and non-homologous recombination. (3)
- Gene Expression – Transcription and Translation: Transcription in prokaryotes – RNA polymerase, initiation, elongation and termination of transcription; regulation of transcription – operon concept, lactose and tryptophan operons (6)
- Transcription in eukaryotes: Different RNA polymerases – requirement of promoters by these RNA polymerases, initiation, elongation and termination by these polymerases; Processing of transcripts – 5' capping, 3' polyadenylation, splicing

and editing; Regulation of transcription – response elements, enhancers and silencers, protein-nucleic acid interaction, HLH, leucine zipper proteins, noncoding RNA (7)

- Translation in prokaryotes: protein synthesis machinery – mRNA, tRNA and rRNA molecules; initiation, elongation and termination of translation; Genetic code – interpreting genetic code; Accuracy of translation. (3)
- Eukaryotic translation: protein synthesis – initiation, elongation and termination; Post-translational modifications of proteins; protein degradation; Regulation of translation - mRNA stability, 5' and 3' UTRs, mRNA localization (5)
- Gene regulation by post-translational modifications of proteins (acetylation, methylation, ribosylation, phosphorylation etc.) and different intermediate RNAs (ribozymes, miRNAs, siRNAs etc.) (4)

### **References:**

- Krebs, Jocelyn E. Goldstein, Elliott S. Kilpatrick, Stephen T. Lewin's Essential GENES 2<sup>nd</sup> Ed. Publisher: Boston, Jones & Bartlett. 2010. ISBN: 978-0763759155.
- Krebs Jocelyn E., Goldstein Elliott S., Kilpatrick Stephen T. Lewin's Genes X Publisher: Jones & Bartlett 2009. ISBN: 9780763766320.
- Clark David P. Molecular Biology. Publisher: Boston Academic Press, Elsevier. 2009. ISBN: 9780123785893.
- Hood Leroy, Goldberg Michael L., Reynolds Ann E., Reynolds Lee M., Veres Ruth C. Hartwell Leland H. Genetics: From Genes to Genomes Publisher: McGraw/Hill, 2008. ASIN: B00210THP6.
- Watson James D., Baker Tania A., Bell Stephen P., Alexander Gann, Levine, Michael Losick Richard. Molecular Biology of the Gene 6<sup>th</sup> Edition. Publisher: New York, Cold Spring Harbor Laboratory Press. 2008. ISBN: 9780321507815
- Weaver Robert F. Molecular Biology, 3rd Edition. Publisher : Boston McGraw-Hill's 2005. ISBN:0071243445.
- Epstein Richard J. Human molecular biology: an introduction to the molecular basis of health and disease. Publisher: Cambridge, UK ; Cambridge University Press, 2003. ISBN: 052164481X.
- Russell Peter J. iGenetics: A Molecular Approach 3<sup>rd</sup> ed.: Pearson International Edition. Publisher : New York, Pearson ISBN: 9780321610225.
- Brown, T.A. Genomes 2 Publisher: New York, BIOS Scientific Publishers Ltd. 2002, ISBN: 1859960294

### **BIM 106 (T+P) : Basic Concepts in Computing**

**(1T+2P credits)**

#### **Objectives:**

This course will enable the students to:

- get acquainted with fundamentals of computers
- use various operating systems
- understand basic concepts in computing and networking

#### **Theory**

##### **Syllabus:**

- Overview and functions of a computer system, storage, devices, memory, etc (1)
  - Types of Processing: Batch, Real-Time, Online, Offline

- Types of modern computers: The workstation, The Minicomputer, Mainframe Computers, Parallel Processing Computer, The Super Computer, etc
- Introduction to operating systems: Windows/Unix/Linux (2)
- The Internet and its Resources, World Wide Web (WWW): associated tools, services, resources and various terminologies, advance search techniques (2)

### **Computer Networking; Network and Data security**

- Fundamentals of networking: OSI Reference Model, TCP/IP, topologies and protocols, designing networks (2)
- Networking gadgets (Router, Switch, etc); Data Communication (ISDN, VPN, DSL, cable modem, cellular modem, etc); Communication Links (Wire pairs, Coaxial cables, Fiber optics, Microwave, Satellite, etc) (2)
- Network security fundamentals: types of attacks, firewall, packet filtering, classification of data security threats, protection mechanism (authentication, access control, access rules) (2)
- Encryption/Decryptions techniques (1)
- An overview of Computer malware (1)
- Current & future technologies (Grid Computing, mobile computing, biometrics etc.) (2)

### ***Practicals***

#### ***Objectives:***

This course will enable the students to:

- use various operating systems, commands, text editors and Microsoft office tools
- use internet and associated tools

#### ***Syllabus:***

Practicals/Tutorials: 30 sessions of 2 hours each (2 sessions/week)

- Hands on Windows:
  - Windows Utilities (5)
- Basic word processing: Notepad, Wordpad, textpad, word, creation of computer presentations with graphics , spreadsheet applications (worksheet basics, working with workbooks, working with formulae and functions, graphs , an introduction to the use of advanced spreadsheet concepts) , various graphics formats, multimedia (8)
- Browsers, various search engines and metadata, E-Mail/Web mail, ftp, designing and creation of WebPages (7)
- Hands On Unix/Linux:
  - file Organization, file handling, directory manipulation, shell command, piping, etc (10)

#### ***References:***

- Tanenbaum Andrew S. Computer networks 4<sup>th</sup> edition. Publisher: Prentice Hall PTR, 2003. ISBN: 8120321758.
- Rajaraman V. Fundamentals of Computers. Publisher: Phi Learning 2001 ISBN: 8120315316.
- Sinha P. K. Computer Fundamentals: concepts system applications: Publisher : Delhi BPB publications 2001. ISBN: 8176563609.
- Forouzan Behrouz A., Coombs Catherine Ann, Fegan Sophia Chung. Data Communications and Networking 2nd edition. Publisher: Osborne Publishing, 2000. ISBN: 0072322047.

**BIM 107 (T+P) : Introduction to Database Systems****(2T+2P credits)****Objectives:**

This course will enable the students to:

- understand the concepts of data, data models and relationships
- be aware of various data representation techniques and various types of databases
- appreciate and implement relational database design
- create the database systems with user-friendly front-ends for fast and efficient data retrieval and storage
- acquire the skills of using MySQL, SQL and basic skills in creating front end applications

**Theory****Syllabus:**

- Database designing, data acquisition (1)
- Data Abstraction, Data Models, Instances & Schemes (1)
- E-R Model (1)
  - Entity and entity sets
  - Relations and relationship sets
  - E-R diagrams
  - Reducing E-R Diagrams to tables
- Relational Algebra and relational Calculus (2)
- Data Normalization (1)
- Basic concepts and applications of Network Data Model, Hierarchical Data Model, Multimedia Databases (2)
- Basic concepts of Indexing and Hashing (1)
- Textual Databases (1)
- Introduction to Distributed Database Processing (1)
- Data warehousing and Data mining (1)
- Organizing biological species information using various database techniques. (3)
  - Need of metadata standards

**SQL****(15)**

- MySQL Architecture
- objects - Tables, Views, Indexes, Sequences; Synonyms, Snapshots, Clusters
- Database - Table space, Data files, Blocks, Extents, Segments; SQL \*DBA - Export-Import/SQL Monitor Backup & Recovery (Archiving); Physical Storage & Logical Storage
- Reports - Reports Features; Full Integration with Forms and Graphics; Data Model and layout editors
- Layout Objects - Frames, Repeating Frames, Fields, Boiler Plate, Anchor; Interface Components; Report Formats; Example Reports; Single Query, Multi Query, Matrix, Master-Detail etc.; User Defined Columns; Menu - Default Menus; Custom Menus; Menu Objects; Menu Module, Main Menu, Individual Menus, Sub Menus, Menu Items; Menu Editor, SQL in Menu Modules, Menu Security
- Select Statements, Data Definition Statements, Data Manipulation Statements, Data Control Statements

## ***Practicals***

### ***Objectives:***

This course will enable the students to:

- appreciate and implement relational database design
- create the database systems with user-friendly front-ends for fast and efficient data retrieval and storage
- acquire the skills of using MySQL, SQL and basic skills in creating front end applications

### ***Syllabus:***

Practicals/Tutorials: 30 sessions of 2 hours each (2 sessions/week)

### **Structured Query Language**

- Assignment based on “Data Definition Language”. (5)
  - A set of SQL commands used to create table, modify table structure, drop table, rename table.
- Assignment based on “Data Manipulation Language” (5)
  - A set of SQL commands used to change the data within the database.
  - It consists of inserting of records in the tables, updation of all or specific set of records in tables, viewing the attributes of table’s column.
- Assignment based on “Data Query Language” (5)
  - It allows extracting the data out of the database. Selecting the data from table using
    - Arithmetic and logical operators.
    - Range searching and pattern matching.
    - Function, group function, scalar function.
- Assignment based on defining Constraints. (5)
  - Types: I/O constraints like Primary Key, Foreign key, Null and Unique constraints. Business constraints like check constraints.
  - Levels: Table level constraints, column level constraints, creating and deletion of constraints using the Alter Table clause.
- Assignment based on using joins. (4)
  - Joining multiple tables, joining a table to itself.
- Assignment based on using Indexes, Sequences (2)
- Security Management using SQL (2)
  - Granting rights on user objects such as Tables, Views, and Sequences.
  - Revoking rights on user objects such as Tables, Views, and Sequences.
- Studying datasets in biodiversity informatics: Species 2000, Tree of life, ATCC, NBII, Species analyst collaboration, ICTV, Animal Virus Information System etc (as examples of biological databases) (2)

### ***References:***

- Bayross Ivan. SQL, PL/SQL The Programming Language of Oracle. 2<sup>nd</sup> revised edition. Publisher: BPB New Delhi. 2003. ISBN: 8176560723.
- Kevin Loney, George Koch and experts at TUSC. Oracle9i: The Complete Reference. 2002. Publisher: Tata McGraw-Hill. ISBN: 0070499020.
- Silberschatz, A, Korth, H F & Sudarshan, S. Database system concepts; (4<sup>th</sup> Edition) Boston. McGraw-Hill higher education, 2002. ISBN: 007120413X.
- Date C. J. An Introduction to Database Systems. 1999. Publisher: Addison Wesley;

6<sup>th</sup> edition. ISBN:0201327546.

- Hanery Korth and Abraham Silberschatz. Database System Concepts. Publisher: New York, 1997 Tata Mac-Graw Hill. ISBN: 0071148108.

## **BIM 108 (T+P) : Biological Databases and Data Analysis (2T+2P credits)**

### ***Objectives:***

This course will enable the students to:

- understand the nature of biological data and need for biological databases
- to understand and explore major biomolecular sequence databases (organization and contents); search and retrieve data from the databases using their respective search engines
- understand and appreciate the need and significance of sequence analysis and the bioinformatics approaches for the same
- understand algorithms for sequence analysis
- understand the application of methods for analysis of the biomolecular sequence data

### ***Theory***

#### ***Syllabus:***

- Introduction to Bioinformatics (1)
- Nature of biological data (1)
- Overview of Bioinformatics resources on the web (1)
  - NCBI/EBI/EXPASY etc
- Biological Databases: Nucleic acid sequence databases (2)
  - GenBank/EMBL/DDBJ
- Biological Databases: Protein sequence databases (2)
  - UniProtKB
- Overview of concepts in sequence analysis (2)
- Pairwise sequence alignment algorithms (3)
  - Needleman & Wunsch
  - Smith & waterman
- Scoring matrices for Nucleic acids and proteins (3)
  - MDM
  - BLOSUM
  - CSW
- Database Similarity Searches (5)
  - BLAST
  - FASTA
- Multiple sequence alignment (3)
  - PRAS
  - CLUSTALW
- Derived databases (4)
  - Prosite
  - BLOCKS
  - Pfam/Prodom
  - Recent derived database (To be decided by faculty member)
- Biological databases: structural databases (3)
  - PDB
  - NDB

- CCSD

### ***Practicals***

#### ***Objectives:***

This course will enable the students to:

- to understand and explore major biomolecular sequence databases (organization and contents); search and retrieve data from the databases using their respective search engines
- apply bioinformatics methods for analysis and interpretation of biomolecular sequence data

**Practicals/Tutorials:** 30 sessions of 2 hours each (2 sessions/week)

#### ***Syllabus:***

- Exploring the integrated database system at NCBI server and querying the PUBMED and GenBank databases using the ENTREZ search engine (2)
  - Use of operators (AND, OR & NOT)
  - Use of limits
  - Exporting GI list
  - Batch retrieval
- Exploring the integrated database system at EBI server and searching the EMBL Nucleotide database using the SRS search engine (1)
- Exploring & querying SWISSPROT & UniProtKB (1)
- Sequence Formats & Format converters (1)
- Exploring tools on ExPASy (1)
- Exploring utilities in EMBOSS packages (1)
- Pair-wise global alignments of protein and DNA sequences using Needleman-Wunsch algorithm & interpretation of results to deduce homology between the sequences, use of scoring matrices (2)
- Pair-wise local alignments of protein and DNA sequences using Smith-Waterman algorithm and interpretation of results (1)
- Database (homology) searches using different versions of BLAST and interpretation of the results to derive the biologically significant relationships of the query sequences (proteins/DNA) with the database sequences (2)
- Database (homology) searches using different versions of FASTA & interpretation of the results to derive the biologically significant relationships of the query sequences (proteins/DNA) with the database sequences (1)
- Multiple sequence alignments of sets of sequences using web-based and stand-alone version of CLUSTAL. Interpretation of results to identify conserved and variable regions and correlate them with physico-chemical & structural properties (2)
- Exploring and using the derived databases: PROSITE, PRINTS, BLOCKS, Pfam and Prodom for pattern searching, domain searches etc. (2)
- Studying the format & content of structural databases & visualization of structures using Rasmol, Cn3D and other utilities (2)

#### ***References:***

- Baxevanis A.D., Davison D.B., Page R. D. M. & Petsko G.A. Current Protocols in Bioinformatics. New York, John Wiley & Sons Inc., 2004. ISBN: 0555015254

- Korf Ian, Yandell Mark, Bedell Joseph. BLAST: an essential guide to the basic local alignment search tool. Shroff Publishers and Distributors Pvt. Ltd., 2003. ISBN: 8173665125.
- Baxevanis Andreas D. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, 3rd Edition. Publisher: New York, John Wiley & Sons, Inc. 2002, ISBN: 9814126756
- Teresa Attwood, Parry-Smith David J. Introduction to Bioinformatics. Publisher: Pearson Education (Singapore) Pte.Ltd., 2001. ISBN:8178085070.
- Mount David W.. Bioinformatics: Sequence and Genome Analysis. Publisher: Cold Spring Harbor Laboratory Press; 1<sup>st</sup> edition 2001. ISBN: 0879695978
- Gibas Cynthia, Jambeck Per. Developing Bioinformatics Computer Skills. Publisher: Shroff Publishers and distributors O'Reilly Media, Inc., 2001. ISBN: 8173662428

### **BIM 109 (P) Programming in Perl**

**(3P credits)**

#### ***Practicals***

##### ***Objectives:***

This course will enable the students to:

- know basic concepts in programming
- develop Perl scripts to solve biological problems

**Practicals/Tutorials: 45 sessions of 2 hours each (2 sessions/week)**

#### ***Syllabus:***

- Concepts of flowcharting, algorithm development, pseudo codes etc.
- Introduction: What is Perl? Why use Perl in Bioinformatics? History of Perl, Availability, Support, Basic Concepts
- Scalar Data: What Is Scalar Data? Numbers, Strings, Scalar Operators, Scalar Variables, Scalar Operators and Functions (6)
- Arrays and List Data: What Is a List or Array? Literal Representation, Variables, Array Operators and Functions, Scalar and List Context
- Control Structures: Statement Blocks
- Hashes: What Is a Hash? Hash Variables, Literal Representation of a Hash, Hash Functions, Hash Slices
- Basic I/O (6)
- Regular Expressions: Concepts About Regular Expressions, Simple Uses of Regular Expressions, Patterns, More on the Matching Operator, Substitutions, The split and join Functions
- Subroutines: System and User Functions, The local Operator, Variable-length Parameter Lists, Notes on Lexical Variables (6)
- Miscellaneous Control Structures:
- File handles and File Tests: What Is a File handle? Opening and Closing a File handle, Using Pathnames and Filenames, A Slight Diversion: die, Using File handles, The -x File Tests, The stat Function
- Formats: What Is a Format? Defining a Format, Invoking a Format

(6)

- Directory Access: Moving Around the Directory Tree, Globbing, Directory Handles, Opening and Closing a Directory Handle, Reading a Directory Handle
- File and Directory Manipulation
- Process Management: Using system and exec, Using Backquotes,
- Other Data Transformation: Finding a Substring, Extracting and Replacing a Substring

(6)

- Formatting Data: Sorting, Transliteration
- System Information: Getting User and Machine Information, Packing and Unpacking Binary Data  
Getting Network Information
- Database Manipulation: DBM Databases and DBM Hashes, Opening and Closing DBM Hashes, Fixed-Length Random-Access Databases, Variable-Length (Text) Databases, Win32 Database Interfaces

(6)

- CGI Programming: The CGI.pm 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
- A tutorial to be developed to include the numerical techniques/algorithms as part of the programs.

(9)

### ***References:***

- Moorhouse Michael, Barry Paul. Bioinformatics, Biocomputing and Perl: an introduction to bioinformatics computing skills and practice. Publisher: John Wiley and Sons, 2004. ISBN:047085331X.
- Tisdall James D. Mastering Perl for bioinformatics. Publisher: Shroff Publishers and distributors O'Reilly Media, Inc. 2003. ISBN: 8173666768.
- Bal, H.P. Perl programming for Bioinformatics. Publisher: Tata McGraw-Hill New Delhi. 2003. 0070474478.
- Curtis Jamison D.. Perl programming for biologists. Publisher : John Wiley & sons, inc., 2003 ISBN 0471430595.
- Tisdall James D. Beginning Perl for bioinformatics. Publisher: Shroff Publishers and distributors O'Reilly Media, Inc. 2001 ISBN: 8173663955.
- Gibas Cynthia, Jambeck Per. Developing Bioinformatics Computer Skills. Publisher: Shroff Publishers and distributors O'Reilly Media, Inc., 2001. ISBN: 8173662428
- Schwartz Randal L., Phoenix Tom. Learning Perl. Publisher: Shroff Publishers and distributors O'Reilly Media, Inc.. ISBN: 8173663718
- Wall Larry, Christiansen Tom, Orwant Jon. Programming Perl. Publisher: Shroff Publishers and distributors O'Reilly Media, Inc. 1998. ISBN: 8173660328.

## Semester – II

### **BIM 201 (T): Cell Biology, Genetics and Genomics**

**(3T credits)**

#### ***Objectives:***

This course will enable the students to:

- study the principles of structure and function of cells, membranes and organelles
- understand the laws of inheritance and principles of population genetics
- understand genome sequencing and related technologies

#### ***Theory***

#### ***Syllabus:***

##### **A. Cell Biology**

**(1 credit)**

- Prokaryotic and eukaryotic cells; membranes and cellular compartmentation; An overview of organelles, (chloroplasts, ER, Golgi, lysosomes and peroxisomes; nucleus) (4)
- Cellular membranes: Structure, transport, channels, carriers, receptors, membrane potentials. (3)
- Cell motility and shape: cytoskeletal elements, motor proteins. (2)
- Signal transduction: signaling by hormones and neurotransmitters; receptors, G-proteins, protein kinases and second messengers (3)
- Cell cycle and its regulation; events during mitosis and meiosis (3)

##### **B. Genetics**

**(1 credit)**

- Science of genetics – objectives, terminologies, methods (1)
- Mendelian principles of inheritance, sex linked inheritance (2)
- Chromosomes, linkage, linkage maps and recombination (2)
- Mutations – molecular, gene/point and chromosomal (2)
- Phenotype and genotype relationships, role of environment, from gene to phenotype, gene interactions (2)
- Study of quantitative traits (3)
- Genetics of populations, genetics and evolution (3)

##### **C. Genomics**

**(1 credit)**

- Recombinant DNA technology: cloning strategies, vectors, screening for recombinants (4)
- Genome projects – importance and strategies (1)
- Genome markers and mapping: STS, EST, RFLP, SNP, radiation hybrid mapping. (4)
- Genome sequencing – first, second and third generation methods (4)
- Functional genomics- Transcriptomics & Proteomics (2)

#### ***References:***

##### **Cell Biology**

- Alberts Bruce, Bray Dennis, Hopkin Karen, Johnson Alexander D, Julian, Lewis Raff Martin, Keith Roberts, Walter Peter. Essential Cell Biology. 3rd Edition Garland Science, 2010. ISBN-10: 0815341296
- Alberts Bruce, Johnson Alexander, Lewis Julian, Raff Martin, Keith Roberts, Walter Peter. Molecular Biology of the Cell. 5<sup>th</sup> edition. Publisher: Garland Science; 2008.

ISBN: 9780815341062.

- Cooper Geoffrey M., Hausman Robert E. The Cell: A Molecular Approach, 5<sup>th</sup> Edition. Publisher: Sinauer Associates Inc. ISBN: 9780878933006.
- Becker Wayne M., Kleinsmith Lewis J., Hardin Jeff, Bertoni Gregory Paul. The World of the Cell, 7<sup>th</sup> Edition. Publisher: Pearson Benjamin Cummings. 2009. ISBN: 9780321554185.
- Lodish Harvey, Berk Arnold, Matsudaira Paul, Kaiser Chris A., Krieger Monty, Scott Matthew P., Zipursky Lawrence, and Darnell James. Molecular Cell Biology, 5<sup>th</sup> edition. Publisher: W H Freeman & Co. 2003. ISBN : 0716743663.
- Wilson John. Hunt Tim. Molecular Biology of the Cell: A Problems Approach 4<sup>th</sup> Edition. Publisher: Garland Science, 2002. ISBN : 0815335776.

### **Genetics**

- Russell P. J. iGenetics: A molecular approach, Pearson International Edition 3<sup>rd</sup>. Publisher: Benjamin Cummings, 2010. ISBN: 9780321610225.
- Hickey G. I. , Fletcher H. L., Winter P. Instant Notes: Genetics 3<sup>rd</sup> Edn. Publisher: Taylor & Francis group, 2007. ISBN: 041537619X.
- Snustad, D.P., and Simmons, M. J. Principles of Genetics, 4<sup>th</sup> edition. John Wiley & Sons, 2006. ISBN :9780471725978.
- Klug William S., Cummings Michael R., Spencer Charlotte, Palladino Michael A. Concepts of Genetics : Pearson International Edition 9<sup>th</sup> Publisher : Benjamin Cummin. ISBN: 9780321540980.
- Hartwell Leland, Leroy Hood, Michael L. Goldberg, Ann E. Reynolds, Lee M. Silver, Ruth Veres. Genetics: From Genes to Genomes, 2<sup>nd</sup> Edition, 2004. Publisher: McGraw-Hill Higher Education. ISBN: 0071214682.
- Eldon John Gardner, Michael J. Simmons, D. Peter Snustad. Principles of Genetics, 8<sup>th</sup> Edition. Publisher: John Wiley & Sons Inc. 1983. ISBN: 8126510439.

### **Genomes**

- Dale Jeremy, Schantz Malcolm Von. From genes to genomes: concepts and applications of DNA technology. Publisher: John Wiley & Sons Ltd., 2007. ISBN: 9780470017340.
- Brown, T.A., Genomes, 3<sup>rd</sup> edition Garland Science publishers, 2006. ISBN: 1859960294.
- Primrose Sandy B., Twyman Richard. Principles of Gene Manipulation and Genomics. Publisher: Wiley-Blackwell. 7<sup>th</sup> edition 2006. ISBN: 1405135441.

## **BIM 202 (T): Structural biology & Bioinformatics**

**(3T credits)**

### ***Objectives:***

This course will enable the students to:

- understand the levels of structural organization of macromolecules and experimental methods of structure determination
- know the approaches for structure analysis
- acquire knowledge of various algorithms & methods of structure prediction
- understand the principles of macromolecular interactions

## ***Theory***

### ***Syllabus:***

- Macromolecular Structure (14)
  - Protein - Primary, Secondary, Supersecondary, Tertiary and Quaternary structure, Potential energy maps, Ramachandran map, Nucleic acid – DNA and RNA, Carbohydrates
    - Co-ordinate systems
- Overview of experimental techniques to study macromolecular structures (5)
  - Methods to study 3D structure: X-ray, NMR, Cryo-electronmicroscopy
  - Validation using Procheck, ProsaII
- Principles of protein folding and methods to study protein folding (5)
- Macromolecular interactions (4)
  - Protein – Protein
  - Protein – Nucleic acids
  - Protein - carbohydrates
- Structure of Ribosome (2)
- Prediction of protein structure (15)
  - secondary structure prediction methods
    - First, second and third generation methods
  - Tertiary structure prediction
    - Homology modeling, fold recognition and *ab initio* methods

### ***References:***

- Forbes Burkowski. Structural bioinformatics: An algorithmic approach. Publisher: CRC Press, 2009. ISBN: 9781584886839.
- Drenth Jan. Principles of Protein X-Ray Crystallography. Publisher: Netherlands, Springer Science. 2007. ISBN: 9780387333342.
- Bourne Philip E., Weissig Helge. Structural Bioinformatics (Methods of Biochemical Analysis, V. 44), 2003. Publisher: Wiley-Liss. ISBN: 0471202002.
- Hölftje Hans-Dieter, Sippl Wolfgang, Rognan Didier, Folkers Gerd. Molecular Modeling: Basic Principles and Applications. Publisher: New York, Wiley-VCH. 2003. ISBN: 3527305890.
- Leach, Andrew. Molecular Modelling: Principles and Applications. Publisher: Prentice Hall. 2001. ISBN: 0582239338.
- Friesner Richard A. Computational Methods for Protein Folding: advances in Chemical Physics Volume 120 Kindle Edition. Publisher: New York, John Wiley & Sons. 2002. ISBN: 0471209554.
- Heilmeyer L., Friedrich P. Protein Modules in Cellular Signalling. Publisher: Amsterdam, IOS Press. 2001. ISBN: 1586031805.
- Rhodes Gale. Crystallography Made Crystal Clear, Third Edition: A Guide for Users of Macromolecular Models. Publisher: USA, Academic Press 2000 ISBN: 0125870728.
- Branden ,Tooze John. Introduction to Protein Structure. Publisher: New York, Garland Publishing Inc. 1999. ISBN: 0815323050.
- Hill H.A.O. Sadler P.J., A.J. Ed. Metal Sites in Proteins and Models Redox Centres Publisher: New York, Springer 1999. ISBN: 3540655564.
- Sternberg Michael J. E. Protein Structure Prediction: A Practical Approach. Publisher: USA, Oxford University Press. 1997. ISBN: 0199634953.
- Fasman G.D. Prediction of Protein Structure and the Principles of Protein Conformation. Publisher: New York, Plenum Press. 1989 ISBN: 0306431319.

- Creighton T. E. Editor. Protein Structure: A Practical Approach. Publisher: IRL Press at Oxford University Press. 1989. ISBN: 0199630011.

## **BIM 203 (T+P) : Chemoinformatics**

**(2T+1P credits)**

### ***Objectives:***

This course will enable the students to:

- understand the fundamentals of chemoinformatics
- appreciate complementary aspects of chemoinformatics and bioinformatics for design of bioactive molecules.
- get hands-on experience in chemical structure representation, storage and analysis of small molecular data.

### ***Theory***

#### ***Syllabus:***

- Introduction to Chemoinformatics: aims, scope. Role of Chemoinformatics in pharmaceutical/chemical research (1)
- Chemical Structure representation: 1D, 2D and 3D structures (1)
- Molecular file formats (SMILES, WLN, SDF, MOL) (1)
- Molecular patterns- SMARTS, SMIRKS (1)
- Fragment based Molecular Fingerprints (ChemAxon, Daylight, MDL and BCI fingerprints- Daylight fingerprints). (1)
- Molecular Descriptors (1D, 2D and 3D) and MACCS Keys (2)  
Topological, electrotopological and shape indices.
- Molecular Similarity and Molecular Diversity Analysis. Similarity metrics:
- Tanimoto Coefficient, Euclidean distance and Tversky Index (1)
- Chemical Databases – Design, Storage and Retrieval methods (1)
- Molecular Database Screening: (Lipinski Rule: Drug/Lead like molecules) (1)
- Chemical Structure based Search techniques: Exact, Sub-structure and similar structure searches. (2)
- Clustering and Statistical methods for Molecular Informatics (PLS, PCA, PCR, kNN, ANN, Correlation and regression analysis) (2)
- Modeling of small molecules (Generation of lowest energy conformations from 2D structures) using Molecular Mechanics and Quantum mechanics methods (2)
- Quantitative Structure Activity/Property/Toxicity Relationship Studies (4)
  - Introduction to Molecular Properties, Activities and Toxicities
  - Training Data, Test Data and External validation Data
  - Applicability domain in QSAR
  - Cross validation techniques
  - Pubchem BioAssay data for QSAR Studies
  - Ligand based design of compounds
- In-silico ADMET Studies (1)
- Design & Analysis of combinatorial libraries (3)
  - Molecular Scaffolds, Linkers and Functional groups
  - Reagents and products based combinatorial library generation
- Pharmacophore Modeling (3)
  - Pharmacophore: Definition and classes (HBA, HBD, Aromatic etc.)
  - Identification of pharmacophore features

- Building 2D/3D pharmacophore hypothesis
  - Searching databases using pharmacophores
  - Application in HTS Analysis
- Docking Studies (2)
  - Target Selection
  - Active site analysis
  - Ligand preparation and conformational analysis
  - Rigid and flexible docking
  - Structure based design of lead compounds
  - Library docking
- Future perspectives in chemoinformatics (1)

### ***Practicals***

#### ***Objectives:***

This course will enable the students to:

- use chemical databases and storage systems
- retrieve and analyse data using physicochemical properties
- understand and model structure-activity activity

#### ***Syllabus:***

- Practical session for chemical structure representation and storage in special file formats (SMILES, WLN, sd and mol). (1)
- Importance of 3D structures and method of generation from 1D & 2D representations. (1)
- A brief introduction to building molecular databases with special emphasis on retrieval using structure input. (1)
- Substructure/Exact/similar structure based searching (1)
- Property based search of molecular databases (1)
- Quantitative Structure Activity/Property/Toxicity Relationship Studies (3)
- Pharmacophore hypothesis and searching (2)
- Docking studies (Rigid, Flexible & library based) (3)
- Design and analysis of focused combinatorial library (2)

#### ***References:***

- Bunin Barry A. Siesel Brian, Morales Guillermo, Bajorath Jürgen. Chemoinformatics: Theory, Practice, & Products Publisher: New York, Springer. 2006. ISBN: 1402050003.
- Gasteiger Johann, Engel Thomas. Chemoinformatics: A Textbook. Publisher: Wiley-VCH; 1<sup>st</sup> edition. 2003. ISBN: 3527306811.
- Leach Andrew R., Valerie J. Gillet. An introduction to chemoinformatics. Publisher: Kluwer academic, 2003. ISBN: 1402013477.
- Gasteiger Johann, Handbook of Chemoinformatics: From Data to Knowledge (4 Volumes), 2003. Publisher: Wiley-VCH. ISBN: 3527306803.

***Objectives:***

This course will enable the students to:

- understand the immune system, its components and their functions
- study informatics-based approaches for prediction of epitopes, design of vaccines and immuno-diagnostic tools

***Theory******Syllabus:***

- Immune systems: (4)
  - Innate and adaptive immunity in vertebrates
  - Antigen processing and presentation
  - Contemporary challenges to the immune system
  - Infectious diseases
  - Autoimmune diseases
  - Categories of vaccines
- Antibodies (3)
  - Immunoglobulin classes and subclasses
  - CDR and LDR regions and sequence numbering
  - Immunogenetics & immunogenomics
  - Hybridoma technology: applications
  - Humanization of antibodies by design
- Membrane receptors for antigen (4)
  - The B-cell surface receptor for antigen (BCR)
  - The T-cell surface receptor for antigen (TCR)
  - The major histocompatibility complex (MHC)
    - MHC polymorphism
    - Causes of MHC polymorphism
    - MHC supertypes
- Epitopes (2)
  - Affinity Maturation
  - Recognition of Antigen by B cells
  - Neutralizing Antibody
- Immunoinformatics overview (2)
  - Reverse vaccinology
  - Rational Vaccine design
  - Evolution and escape due to variations
- Immunoinformatics databases & tools: (3)
  - IMGT & IEDB
  - BciPep, Epiteome, CED, AgAbDb
- Molecular interactions (3)
  - Specificity, characterization
  - Ag-Ab: Sequential and conformational epitopes
  - MHC- peptide
  - HLA: nomenclature, HLA-peptide interactions & matrices
- Epitope prediction algorithms: (6)
  - B-cell epitope predictions using sequence-based approaches
  - B-cell epitope predictions using structure-based approaches

- T-cell epitope prediction methods
- Vaccine design: (3)
  - Pipeline & workflows
  - Prediction of immunogenicity
  - Case studies: JEV, Mumps etc

***Practicals***

**Objectives:**

This course will enable the students to:

- understand and explore sequence and structural databases relevant in the area of immunology
- explore sequence and structural features of antibodies using computational tools
- characterize and understand principles of antigen-antibody interactions
- understand algorithms and methods for prediction of epitopes
- explore and use approaches for vaccine design

***Syllabus:***

- Browsing and searching sequence & structure databases to retrieve data (1)
  - Immunoglobulin: sequence and structure
- Study of Antibody sequence and structure (2)
  - Antibody numbering: Kabat & Chothia methods
  - Equivalence between different numbering schemes & PDB numbering
  - Visualisation of structure & mapping CDRs on structure
  - Study of variable and constant domains
  - Drawing topology diagram
- Study of Ag-Ab interactions (2)
  - Searching for Ag-Ab complexes
  - Use of molsurfer, explorer etc
  - Characterisation of binding sites
    - Ab-protein complex
    - Ab-peptide complex
    - Geometric and molecular interaction
- Sequence analysis in immunology (1)
  - Alignments: antibody sequences
  - Alignments: antigen sequences, molecular evolution, variability analysis and phylogeny
- Prediction of sequential epitopes (1)
- Prediction of conformational epitopes (1)
- Estimation of accuracy of predicted epitopes: use of experimental data (2)
- Prediction of cytotoxic and helper T cell epitopes (1)
  - MHC Class I epitopes
  - MHC Class II epitopes
- Web-based tools for vaccine design (1)
- Predicting immunogenicity (1)
  - Combination of MHC and Proteasome predictions
  - Combination of MHC, TAP, and Proteasome predictions

**References:**

- Flower Darren R. Bioinformatics for Vaccinology Publisher: UK, John Wiley & Sons Inc. 2008. ISBN: 9780470027110.
- Flower Darren R. Immunoinformatics: Predicting Immunogenicity In Silico Publisher: New Jersey, Humana Press. 2007. ISBN: 9781603271189.
- Kindt, Thomas J., Osborne Barbara A., Goldsby Richard A. Kuby Immunology 6th Edition. Publisher: New York, W. H. Freeman. 2007. ISBN: 9780716785903.
- Lund Ole , Nielsen Morten , Lundegaard Claus, Kesmir Can, Brunak SÃ,ren. Immunological Bioinformatics. Publisher: London, MIT Press 2005 ISBN: 0262122804.
- Foundation Novartis. Immunoinformatics: Bioinformatic Strategies for Better Understanding of Immune Function. Publisher: Chichester, John Wiley & Sons Inc. 2003. ISBN: 0470853565.
- Roitt Ivan, Delves Peter. Roitt's Essential Immunology 10th Edition. Publisher: Canada, Blackwell. 2001, ISBN: 0632059028.
- Ellis Ronald W. Vaccines: New Approaches to Immunological Problems. Publisher: Stoneham, Butterworth-Heinemann 1992. ISBN: 0750692650.

**BIM 205 (T+P) : Programming in Object Oriented Languages (2T+ 2P credits)****Objectives:**

This course will enable the students to:

- Acquire skills in programming using object oriented language: JAVA
- Develop and implement programs to analyse biological data

**Theory + Practicals****Syllabus:**

- An introduction to JAVA programming (2)
- Object-oriented programming and Java
- Java Basics
- Working with objects (5)
- Arrays, Conditionals and Loops
- Creating Classes and Applications in Java (5)
- Java Applets Basics
- Graphics, Fonts and Color
- Simple Animation and Threads (5)
- Advanced Animation, Images and Sound
- Managing Simple Events and Interactivity
- Creating User Interfaces with AWT (2)
- Windows, Networking and other Tidbits
- Modifiers, Access Control and Class Design
- Packages and Interfaces (3)
- Exception
- Multithreading
- Streams and I/O (3)
- Using Native Methods and Libraries

- Java Programming Tools (5)
- Working with Data Structures and Java
- Image Filters

***Practicals***

***Objectives:***

This course will enable the students to:

- Acquire skills in programming using object oriented language: JAVA
- Develop and implement programs to analyse biological data

***Syllabus:***

- Programs JAVA programs to be developed for analysis of scientific data: (20 assignments)

***References:***

- Horstmann Cay S.Cornell Gary. Core Java Volume II Advanced Features. Publisher: New Delhi, Education. 2002. ISBN: 8178088460.
- Schildt Herbert. Java 2: The Complete Reference 5th Edition. Publisher: New Delhi, Tata McGraw-Hill Publishing Company. 2002. ISBN: 0070495432.
- Bruce Eckel. Thinking in Java 2nd Edition. Publisher:Pearson Education 2001 ISBN: 8178083302.
- Flanagan David. Java in a Nutshell: A Desktop Quick Reference. Publisher: Mumbai, Shroff Publishers & Dist. 1999.ISBN: 8173661065.
- Balagurusamy E.Programming With Java A Primer. Publisher: New Delhi, Tata Mgraw Hill 1999. ISBN: 0074635425.

**BIM 206 (P) Programming in C**

**(3P credits)**

***Objectives:***

This course will enable the students to:

- conceptualize and formulate logic and flow for the implementation of a computational task
- develop codes using the structured programming approach of ‘C’ programming language
- develop and implement programs to analyse biological data

***Syllabus:***

- Concepts of flowcharting, algorithm development, pseudo codes etc. (5 assignments)
- Laboratory assignments based on the following topics in ‘C’ programming: (10 assignments)  
Data types, operators and expressions, Hierarchy of operators, control statements including decision (if, if-else), loops (while, do-while, for), branching (switch, break, continue), functions, arrays (1D, 2D- all matrix operations including inverse of a matrix), strings, file handling, data structures etc.
- Writing C programs for Bioinformatics applications: (10 assignments)
  - Extract a protein or nucleic acid sequence from any of the databank files (GenBank entry, Swiss-Prot, EMBL entry etc.)

- Interconverting the sequence from one databank format to the other eg. GenBank format to FASTA format, FASTA to PIR format etc.
- Determining the base composition in a nucleic acid sequence and amino acid composition in a protein sequence.
- Generating the complimentary sequence of a DNA sequence
- Pattern search algorithms
- Search for a specific oligonucleotide pattern (eg. GAACATCC) in a given DNA sequence.
- Find the position where a specific sequence say “GGTCCCGAC” will hybridize a given DNA sequence.
- Find the restriction enzyme cleavage sites eg. where PVUZ, ECORI etc. will cut the DNA.
- Locate palindromic sequence stretches in a DNA sequence.
- Count the number of Open Reading frames (ORF's) in a DNA sequence.
- Calculate the codon usage in a nucleic acid sequence.
- Translate a DNA sequence into protein sequence in the forward and reverse frames.
- Implementation of the Needleman-Wunsch algorithm for pair wise alignment and testing alignment score with randomized pairs of sequences also.
- Numerical Techniques (4 assignments)
- Basic Formalism, Methods for Solving Equations, Finding Eigen values & Eigenvectors (5 assignments)

**References:**

- Kanetkar Yashavant. Let Us C 9<sup>th</sup> Edition. Publisher: New Delhi BPB Publications. 2009. ISBN: 9788183331630.
- Balagurusamy E. Programming In ANSI C. Publisher: New Delhi Tata McGraw Hill Publishing Company Ltd. 2007. ISBN: 9780070648227.
- Jonassen Inge, Kim Junhyong. Algorithms in Bioinformatics: 4th International Workshop, WABI 2004 Bergen, Norway, September 2004 Proceedings. Publisher: New York Springer. 2004. ISBN: 3540230181.
- Kernighan Brian W. Ritchie Dennis M. The C Programming Language 2<sup>nd</sup> Edition. Publisher: USA, Prentice-Hall, Inc. 1988. ISBN: 0876925964.

**BIM 207 (P): Computational Structural Biology**

**(2P credits)**

**Objectives:**

This course will enable the students to:

- explore primary and derived databases in the field of computational structural biology
- visualize macromolecular structures using various visualization tools
- acquire skills to use different approaches for prediction of protein structure

**Syllabus:**

- Structural data, databases and structure analysis (2)
- Exploring the Database & searches on PDB and CSD, WHATIF (3)
- Molecular visualization tools (3)

Visualization of tertiary structures, quaternary structures, architectures and topologies of proteins and DNA using molecular visualization softwares such as RasMol, Cn3D, SPDBV, Chime, Mol4D, PyMOL etc.

- Methods for prediction of secondary structure of proteins (10)  
Prediction of secondary structures of proteins using at least 5 different methods with analysis and interpretation of the results. Comparison of the performance of the different methods for various classes of proteins.
- Methods for prediction tertiary structure of proteins along with analysis and interpretation of results
  - Homology modeling: InsightII, Discovery Studio, SWISSMODEL, SWISS-PDB Viewer (7)
  - Fold recognition methods: PHYRE, TOPITS, GenThreader (or other equivalent methods) (8)

**References:**

- Cesareni Giovanni, Gimona Mario, Sudol Marius, Yaffe Michael (Editors). Modular Protein Domains. Publisher: Weinheim Wiley-VCH. 2005. ISBN: 352730813X.
- Höltje Hans-Dieter, Sippl Wolfgang, Rognan Didier, Folkers Gerd. Molecular Modeling: Basic Principles and Applications. Publisher: New York, Wiley-VCH. 2003. ISBN: 3527305890.
- Webster David (Editor). Protein Structure Prediction: Methods and Protocols (Methods in Molecular Biology) Volume 143. Publisher: New Jersey Humana Press. 2000. ISBN: 0896036375.
- Sternberg Michael J. E. Protein Structure Prediction: A Practical Approach. Publisher: USA, Oxford University Press. 1997. ISBN: 0199634953.

**BIM 208(P): Laboratory Techniques in Modern Biology**

**(2P credits)**

**Objectives:**

This course will enable the students to:

- gain hands-on experience with some of the laboratory techniques necessary to understand genomics and proteomics.

**Syllabus:**

- A. General (2)
  - Washing, Cleaning and sterilization of apparatus, chromic acid preparation.
  - Distillation of water
  - Media and buffer preparation
- B. Basics of Molecular Biology and Biophysics (8)
  - Overnight culture inoculation, pouring of agar plates
  - Streaking and spreading of plates
  - Bacterial growth curve
  - Transformation by TSS method
  - Plasmid isolation
  - Gel analysis
  - DNA spectrum analysis
  - Restriction enzyme digestion
  - PH titration of glycine, histidine and glutamic acid
  - CD spectra

- C. Basics of Proteomics (6)
- Extraction and column (affinity, ion-exchange, gel filtration) purification of protein
  - Lowry/Bradford estimation of fractions
  - SDS-PAGE
  - PCR amplification of gene
  - PCR product analysis on agarose gel

**References:**

- Harris E. L. V., Angal S. Protein Purification Applications: A Practical Approach 1990. Publisher: Oxford University Press, USA. ISBN: 0199630232.
- Ladisch Michael R., Builder Stuart E., Painton Chih-duen C., Willson Richard C. Protein Purification: From Molecular Mechanisms to Large-Scale Processes (Acs Symposium Series). Publisher: An American Chemical Society Publication. 1990. ISBN: 0841217904.

**BIM 209(P): Unix and Linux: Commands & Scripting (1P credits)**

**Objectives:**

This course will enable the students to:

- make effective use of advanced Linux/Unix commands
- perform advanced tasks in parallel computing environment

**Syllabus:**

- Using Vim, g-Vim editor (2 assignments)
- Advanced Unix commands (3 assignments)
- Grep, SED, AWK (6 assignments)
- Shell scripting (2 assignments)
- Configuring services in Unix (2 assignments)
- Package management in Red hat like systems (RHEL, fedora etc.) (2 assignments)

**References:**

- Neil Matthew, Stones Richard. Beginning Linux Programming 3rd Edition. Publisher: New Delhi Wiley Dreamtech India Pvt. Ltd. 2004. ISBN: 8126504846.

## Semester III

### **BIM 301 (T+P) : Molecular Phylogenetics**

**(1T+1P credits)**

#### ***Objectives:***

This course will enable the students to:

- understand concepts of molecular evolution and the nature of data for deriving molecular phylogeny
- learn and apply the statistical approaches and models for phylogenetic analysis and tree reconstruction

#### ***Theory***

##### ***Syllabus:***

- Molecular Evolution: (2)
  - overview & protocols
  - Nature of data used
- Probabilistic models and associated algorithms (3)
  - Probabilistic models of evolution
  - Maximum likelihood algorithm
- Phylogenetic analysis algorithms (5)
  - Distance-based: UPGMA, Transformed Distance, Neighbors-Relation, Neighbor-Joining
  - Character optimization; delayed and accelerated transformation
  - Maximum Parsimony
  - Reliability of trees: Bootstrap, jackknife, decay, randomization tests.
- Phylogenetic trees and their comparison: (2)
  - Definition and description, various types of trees;
  - Consensus (strict, semi-strict, Adams, majority rule, Nelson).
  - Data partitioning and combination.
  - Tree to tree distances, similarity
- Case studies (2)
- Numerical taxonomy & bacterial identification (1)

#### ***Practicals***

##### ***Objectives:***

This course will enable the students to:

- use computational approaches for phylogenetic analysis
- explore and use packages available for molecular phylogeny

##### ***Syllabus:***

- Compilation & curation of dataset, format conversion (1)
- Survey of software programs available for phylogenetic analysis (1)
  - Installation of at least 2 public domain packages for both Windows & Unix environment: Phylip, PAUP, MEGA
- MSA using ClustalW: writing files for phylogenetic analysis (1)
  - Informative and variable sites; Singleton sites etc.
  - Difference between dendrogram & phylogenetic tree
  - Converting sequence data into distance data
- Reconstruction of phylogenetic trees using molecular data (1 datasets)

- Distance-based methods: UPGMA, Neighbor-joining, Neighbor-relations & Transformed distance (2)
  - Maximum Parsimony (1)
  - Maximum likelihood (1)
- Using bootstrapping tool to generate multiple datasets from the original input data & generation of consensus tree (1)
- Plotting, visualizing & printing phylogenetic trees: TreeView and other tools (1)
  - Various rendering
  - Formatting & labeling
  - Interpretation of trees
- Comparison of trees drawn using (1)
  - RNA, Nucleotide & protein data
  - Gene trees & species tree
- Un-rooted & rooted tree
  - Rooting un-rooted tree using an out group (1)
- Reconstruction of phylogenetic trees using whole genome data of viruses (1)
- Numerical taxonomy & Bacterial identification using matrices (1)

**References:**

- Bromham Lindell. Reading the Story in DNA: A Beginner's Guide to Molecular Evolution. Publisher: USA, Oxford University Press. 2008. ISBN: 9780199290918.
- Bernardi Giorgio. Structural and Evolutionary Genomics, Volume 37: Natural Selection in Genome Evolution (New Comprehensive Biochemistry). Publisher: Netherlands, Elsevier Science.2005. ISBN: 9780444521361.
- Marco Salemi, Anne-Mieke Vandamme. The phylogenetic handbook: a practical approach to DNA and protein phylogeny. Publisher: Cambridge University Press, 2003. ISBN: 052180390X.
- Pathy Laszlo. Protein Evolution. Publisher: London, Blackwell Science Ltd. 1999. ISBN: 0632047747.
- Takahata Naoyuki, Clark Andrew G.(Editor). Mechanisms of Molecular Evolution: Introduction to Molecular Paleopopulation Biology. Publisher: Japan, Japan Scientific Societies Press and Sinauer Associates, Inc. 1993. ISBN: 476226718X.
- Graur Dan, Wen-Hsiung Li. Fundamentals of molecular evolution Publisher: Sinauer Associates, 1991. ISBN: 0878932666.

**BIM 302 (T+P) : Genome to Drug and Vaccine**

**(3T+2P credits)**

**Objectives:**

This course will enable the students to:

- appreciate and understand the changes in the approaches for computational analysis between the pre- and post-genomic era
- understand the role of Bioinformatics in sequence assembly and annotation
- appreciate the role of Bioinformatics in post-genomics technologies and areas such as DNA micro-array experiments, proteomics, protein-protein interactions, pharmacogenetics, identification of disease genes, drug and vaccine design etc.

## ***Theory***

### ***Syllabus:***

- Genome sequencing, Raw data & Assembly (1)
- Genome Databases and related data resources (EST, STS, GSS, HSS etc.) (4)
- Organization of data in databases; (2)
  - Genome Data Visualization (With emphasis on Human Genome)
    - Ensembl
    - MapViewer
  - Biology of the following: (5)
    - M. tuberculosis
    - Plasmodium, Leishmania, Entamoeba histolytica
    - Mosquito
- Genome databases of the following (To be explored by the students for presentation in a seminar)
  - E. coli
  - M. tuberculosis
  - Yeast
  - Entamoeba histolytica
  - Leishmania
  - Plasmodium
  - C. elegans
  - Drosophila
  - Anopheles gambiae
  - Zebra fish
  - Fugu
  - Human
  - Mouse
  - Arabidopsis thaliana
  - Rice
  - Any other
- Tools for Genomic Data Mining (10)
  - Basic Aspects of Genome Annotation
  - Prediction of ORFs
  - Prediction of Promoters, splice sites, UTRs etc
  - Algorithms for gene prediction gene modeling
- Identification of Disease Genes & Drug Targets (5)  
(in the context of Human Genetics and Genetics of Model organisms)
  - OMIM
  - Metabolic diseases and Pathogenic diseases
- High throughput data & its analysis (6)
  - Gene Expression Analysis
  - Structural Genomics
  - Functional Genomics
- Host-Parasite and Host-Vector-Parasite Interactions (3)
- Multi-Drug Resistance (3)

- Mechanism of MDR: genomic, molecular, cellular
- Identification of genes responsible for MDR
- Approaches to novel drug discovery
- Pharmacokinetics (4)
  - Classification
  - Case study
  - Comparative screening
  - ADMET: Drug metabolism; Elimination half-life, Toxicity screening
- Pharmacogenetics (2)
  - The genetics of drug metabolism
  - The genetics of therapeutic targets
  - Interactions of small molecules and gene-based drug targets

### ***Practicals***

#### ***Objectives:***

This course will enable the students to:

- browse, navigate through and search genomic databases
- understand the process of assembly and annotations
- learn & use methods for prediction of various genomic regions
- learn & apply various methods in functional genomics

#### ***Syllabus:***

- Browsing & viewing genome data (1)
  - Ensembl@EBI
  - MapViewer@NCBI
- Viewing regions exhibiting Synteny
  - Genome assembly (1)
  - Genome Annotation (2)
- Using integrated genome annotation servers such as the server developed at IMTech, Chandigarh (<http://imtech.res.in/raghava/gp.html>) (2)
- Gene Prediction and Gene Modeling
  - Prediction of genes and gene structures (gene modeling) using online (web) servers of different methods tailored for prokaryotic and eukaryotic organisms such as GLIMMER, GeneMark, Grail, GENSCAN etc. Interpretation of results and comparison with known gene models (where available). Evaluation of accuracy of the methods. (2)
  - Prediction of promoters using methods such as Neural Network Promoter Prediction (NNPP) at Berkeley Drosophila Genome Project server, Genome inspector for combined analysis of multiple signals in genomes etc. Using Promoter databases. (1)
  - Prediction of alternate splice sites using methods such as Splice Site Prediction by Neural Network (at Berkeley Drosophila Genome Project server), GenScan, NetGene2, GeneSplicer etc. Prediction of PCR primers using Primer 3, ePCR etc. (1)
- Functional Genomics
  - Using primary databases (such as UniProt) and derived databases such as InterPro, PRINTS, BLOCKS, PRODOM, Pfam etc. along with advanced sequence analysis tools such as profiles searches, pattern searches for function



- understand the basis and nature of protein-protein interactions and related databases

### ***Theory***

#### ***Syllabus:***

- Objective and Overview of Genome Comparisons (1)
- Genome Alignments (5)
  - BLAST2
  - MUMmer
  - PipMaker
  - VISTA
- Comparison of Gene Order (1)
  - GeneOrder
- Comparative Genomics (6)
  - Viruses
  - Microbes
  - Pathogens
  - Eukaryotes
- Comparative Genomics Databases (6)
  - COG
  - VirGen
  - CORG
  - HOBACGEN
  - Homophila
  - XREFdb
  - Gramene
- Single Nucleotide Polymorphism, dbSNP and other SNP-related databases (2)
- Overview of Proteomics (3)
  - Experimental Techniques
  - Bioinformatics Approaches
- Protein-Protein Interaction Networks, databases and software (6)
  - DIP (Database of Interacting Proteins)
  - PPI Server
  - BIND - Biomolecular Interaction Network Database
  - PIM -Hybrigenics
  - PathCalling Yeast Interaction Database
  - MINT - a Molecular Interactions Database\
  - GRID - The General Repository for Interaction Datasets
  - InterPreTS - protein interaction prediction through tertiary structure

### ***Practicals***

#### ***Objectives:***

This course will enable the students to:

- browse and search the comparative genomics & proteomics databases
- use the methods in comparative genomics and interpret the results.
- learn and apply methods in characterization of protein-protein interactions

**Syllabus:**

- Explore comparative genomics resources and NCBI and EBI (3)
- Comparison of full / partial genomic sequences using following methods to identify conserved genes and map/compare the annotations of the two sequences
  - BLAST2 (1)
  - MegaBLAST and Discontiguous MegaBLAST (2)
  - MUMmer (2)
  - PipMaker (2)
  - VISTA (1)
  - Artemis (3)
- Compare gene order of given genomic sequences using the GeneOrder tool (1)
- Explore and query the comparative genomics databases: COG, VirGen, CORG, HOBACGEN, Homophila, XREFdb, Gramene etc. (3)
- Explore and query SNP and SNP-related databases (1)
- Explore and query the protein-protein interaction databases: DIP, PPI Server, BIND, PIM, PathCalling, MINT, GRID, InterPreTS (6)
- Comparative genomics case studies of - (5)
  - model organisms
  - plants
  - microbes
  - viruses
  - host-vestor-parasites

**References:**

- Choudhuri Supratim & Carlson David B. Genomics fundamentals and applications. Publisher: New York. Informa healthcare, 2009. 9781420067057.
- Primrose Sandy B., Twyman Richard. Principles of Gene Manipulation and Genomics, 7<sup>th</sup> edition. Publisher: Australia, Wiley-Blackwell. 2006. ISBN: 1405135441.
- Mount David W. Bioinformatics: Sequence and Genome Analysis, Second Edition Publisher: New York, Cold Spring Harbor Laboratory Press, 2004. ISBN10: 0879697121.
- Hecker Michael, Müllner Stefan. Proteomics of Microorganisms: Fundamental Aspects and Application. Publisher: Springer, 2003. ISBN:3540005463.
- Liebler Daniel C. Introduction to Proteomics: Tools for the New Biology. Publisher: New Jersey, Humana Press. 2002. ISBN: 0896039919.
- Pennington S., Dunn M. J. Proteomics: From Protein Sequence to Function. Publisher: New Delhi, Viva books Pvt. Ltd. 2002. ISBN: 8176492906.
- Sankoff D., Nadeau J.H., Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment and the Evolution of Gene Families. Publisher: Netherlands, Kluwer Academic Publishers, 2000. ISBN:0792365836.

**BIM 304 (T+P) : Molecular Modelling & Simulations****(2T+2P credits)****Theory****Objectives:**

This course will enable the students to:

- understand basic concepts in molecular mechanics, modeling & simulations
- understand application of classical mechanics for biomolecules
- understand application of simulation techniques to study molecular dynamics & derive properties of thereof
- understand principles of docking simulations

**Syllabus:**

- Overview of molecular modelling (1)
- Molecular modelling methods (2)
  - Semi-empirical methods, Empirical methods
- Molecular Mechanics (5)
  - Conformations: global vs. local
  - Force fields: expressions for stretch, bond, torsion, etc.
  - Description of various force fields: MM3, Dreiding, AMBER, CHARMM
  - Mechanics of Bio-macromolecules
- Molecular Dynamics (10)
  - Newton's equations for many particles
  - Verlet and related algorithms
  - Types of dynamics simulations: adiabatic, constant T, simulated annealing, etc.
  - Conformational searching using MD and other methods
  - Free energy calculations
  - Dynamics of Bio-macromolecules
  - Electrostatics of biomolecules
- Energy Minimisation (6)
  - Golden section
  - derivative based method (SD, CG, Newton-Raphson)
- Docking simulations (6)
  - Rigid docking
  - Flexible docking
  - Different Scoring schemes

**Practicals:****Objectives:**

This course will enable the students to:

- use energy minimization methods for optimization of geometry
- use molecular dynamics methods to sample conformational space/predict structures of oligo-peptides
- use methods of molecular docking & its applications

**Syllabus:**

- Use of SD and CG for energy minimization of biomolecule (3)
- Oligopeptide/oligonucleotide model building (PYMOL / DStools) (4)
- MD using any one of the SW (GROMACS, AMBER & CHARMM): (10)
  - MD in vacuo
  - MD using implicit solvent

- MD using explicit solvent with PBC/layer of solvent
- Visualisation and analysis of MD trajectory data (Dstoools & VMD) (3)
- Docking simulations (10)

**References:**

- Höltje Hans-Dieter, Sippl Wolfgang, Rognan Didier, Folkers Gerd. Molecular Modeling: Basic Principles and Applications. Publisher: New York, Wiley-VCH. 2003. ISBN: 3527305890.
- Friesner Richard A. Computational Methods for Protein Folding: advances in Chemical Physics Volume 120 Kindle Edition. Publisher: New York, John Wiley & Sons. 2002 ISBN: 0471209554
- Leach, Andrew. Molecular Modelling: Principles and Applications. Publisher: Prentice Hall. 2001. ISBN: 0582239338.
- Höltje Hans-Dieter, Folkers Gerd. Molecular Modeling: Basic Principles and Applications (Methods and Principles in Medicinal Chemistry) Vol. 5. Publisher: New York, Wiley-VCH, 1997. ISBN: 3527293841.
- McCammon Andrew J., Harvey Stephen C. Dynamics of Proteins and Nucleic Acids Publisher: New York, Cambridge University Press, 1987. ISBN: 0521356520.

**BIM 305 (P) : Database design & web-implementation (1P credit)**

**Objectives:**

This course will enable the students to:

- parse and transform large scale biological data.
- set the data into formats suitable for feeding in traditional RDBMS
- develop appropriate web based front end to browse, query and manipulate the data.

**Syllabus:**

- Identify relevant problem, Design and create MySQL database. (2)
- Parsing large scale data into “csv”(OR other parseable) format and using MySQL import or similar tools to populate database tables in bulk. (3)
- Design html and css codes for web based front end. (3)
- Connecting the front end and database with server side language OR cgi script. (3)
- Using javascript for client side user interaction. (5)
- Brief introduction to web2 technologies and Ajax. (1)

**References:**

- Tremblett Paul, Instant JavaServer Pages. Publisher: New Delhi, McGraw-Hill Companies, 2000. ISBN: 0070411883.
- Graham Ian S. The HTML sourcebook, 2nd ed.: a complete guide to HTML 3.0. Publisher : New York, John Wiley & Sons, Inc., 1996. ISBN: 0471142425.

**BIM 306 (T+P) : Scientific Data Visualization****(1T+2P credits)****Objectives:**

This course will enable the students to:

- visualize the data using various techniques and derive knowledge out of it

**Theory & Practicals****Syllabus:**

- Introduction to different statistical plots: histogram, line plot, pie, box-Whisker plot (1 hr, 2 assignments)
- 3-D interactive plots (3 hrs, 6 assignments)
- Heat maps, contour plots (5 hrs, 10 assignments)
- Time series data visualization (3 hrs, 6 assignments)
- Introduction to processing for custom visualization: (3hrs, 10 assignments)
  - Transformation: Translation, Rotation, Scaling
  - Forms, Motions, Animation.

**References:**

- Pickover Clifford A. Visualizing biological information. Publisher: Singapore, World Scientific. 1995. ISBN 9810214278.

**BIM 307 (T+P) : Tools & Techniques for biological data mining****(2T+2P credits.)****Theory****Objectives:**

This course will enable the students to:

- Understand, implement and apply various advanced mathematical, physical and statistical techniques for analysis of biological data.

**Syllabus:**

- Quality of Biological Data & Data Accuracy (2)
  - General issues regarding Biological Databases; Representation of errors due to (machines, 3D structural and sequence data of proteins and nucleic acid, Proteomics and Microarray data)
- Optimization Techniques (7)
  - Random walk, Monte carlo
  - Simulated annealing
  - Genetic Algorithms
- Information theory, entropy and relative entropy (2)
- Clustering & Classification Algorithms (4)
  - Hierarchical and non-hierarchical Clustering
  - K-Means clustering
  - Grid based clustering
  - SOM
- Machine learning Techniques: (8)
  - Bayesian modeling
  - Hidden Markov Models
  - Neural Networks

- Support Vector machine & Ant colony optimization applied to (4)
  - Identification of gene and protein function
  - Biomolecular Structure Prediction
- Fuzzy logic & applications in Clustering and classifications (3)

### ***Practicals***

#### ***Objectives:***

This course will enable the students to:

- implement and apply techniques for biological data mining using various packages/self generated codes
- analyse and interpret outcome of data mining techniques

#### ***Syllabus:***

- Machine learning techniques (5\*3)
  - Use of WEKA for NN to secondary structure/gene prediction.
  - Estimator of transition probabilities for markov models based on various sample sizes.
  - Hidden Markov model implementation in R/MATLAB
  - Ant colony algorithm for mining classification rules for sequence data
  - Implementation of grid techniques for tuning SVM parameters
  - Determine the entropy in a DNA sequence, relative entropy in 2 sequences etc
- Optimization Algorithms: (5\*3)
  - Programming in 'C' for implementation of Golden section algorithm, Steepest descent for energy minimization applications.
  - Implementation of random walk and Monte Carlo algorithm.
  - Simulated Annealing algorithm for energy minimization.

#### ***References:***

- Fogel Gary B., Corne David W., Pan Yi Edt. Computational Intelligence in Bioinformatics Publisher: New Jersey, John Wiley & Sons, Inc. 2008 ISBN: 978-0470105269.
- Parida Laxmi. Pattern Discovery in Bioinformatics: Theory & Algorithms Publisher: New York, Chapman and Hall/CRC 2008. ISBN: 9781584885498.
- Jones Neil C., Pevzner Pavel A. An Introduction to Bioinformatics Algorithms Publisher: New Delhi, Ane Books 2005. ISBN: 8180520781.
- Keedwell Edward, Narayanan Ajit Intelligent Bioinformatics: The Application of Artificial Intelligence Techniques to Bioinformatics Problems. Publisher: England, John Wiley & Sons Ltd. 2005. ISBN: 0470021756.
- Amaratunga Dhammika , Cabrera Javier. Exploration and Analysis of DNA Microarray and Protein Array Data. Publisher: New Jersey, John Wiley & Sons, Inc. 2004. ISBN: 0471273988.
- Dorigo Marco, Stutzle Thomas. Ant Colony Optimization. Publisher: New Delhi, Prentice Hall of India Pvt. Ltd. 2004. ISBN: 8120326849.
- Dunham Margaret H. Data Mining: Introductory and Advanced Topics Publisher: New Delhi, Pearson Education. 2003 ISBN: 8178089963.
- Durbin Richard, Eddy Sean R., Anders Krogh. Graeme Mitchison Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Publisher: New York, Cambridge University Press, 2003. ISBN: 0521540798.

- Lin, Simon M. Johnson Kimberly F. Methods of Microarray Data Analysis II. Publisher: USA Kluwer Academic Publishers 2002. ISBN: 1402071116.
- Teodorescu Horia-Nicolai L ,Kandel Abraham , Jain Lakhmi C. Fuzzy and Neuro-Fuzzy Systems in Medicine. Publisher: New York, CRC Press.1999 ISBN: 0849398061.
- Mitchell Tom. Machine Learning. Publisher: The McGraw-Hill Company, Inc., 1997 ISBN: 0071154671.
- Klir George J. Yuan Bo Fuzzy Sets and Fuzzy Logic: Theory and Applications. Publisher: New Delhi, Prentice Hall 1995. ISBN: 8120311361.
- Witten, Ian H. Frank Eibe. Data Mining: Practical Machine Learning Tools and Techniques, 2<sup>nd</sup> Edition Publisher: Morgan Kaufmann; ISBN: 8131200507.
- Rao S. S Optimization: Theory and applications 2nd ed. Publisher: Pune, Wiley eastern Ltd., 1991. ISBN-10: 0852267568.
- Brian D. Ripley Stochastic Simulation (Wiley Series in Probability and Statistics) Publisher: New York, John- Wiley. 1987 ISBN: 0471818844.

**BIM 308 (T) : Project (Phase I) & Review**

**(2T credits)**

***Objectives:***

This course will enable the students to:

- Identify a topic for project work
- Carry out literature survey and background reading
- write a review on the topic of research project

## Semester IV

### **BIM 401 (T+P) : Advanced Techniques for Sequence and Structure Analysis (2T+2P credits)**

#### *Theory*

##### **Objectives:**

This course will enable the students to:

- understand and implement various advanced level techniques for mining biomolecular sequence and structure data.

##### **Syllabus:**

- Advanced Techniques for Sequence Analysis (5)
  - Sequence Profiles: Derivation, Databases, Application
  - PSI-BLAST, PHI-BLAST
- Advanced data mining Techniques (4)
  - Random forest : ligand binding site prediction
  - Decision trees: prediction of post-translational modifications
- Advanced sequence analysis (6)
  - QTL, CNV, LoH
- Image analysis (7)
- Graph theory (8)

#### *Practicals*

##### **Objectives:**

This course will enable the students to:

- Implement/apply various advanced level techniques for mining biomolecular sequence and structure data.

##### **Syllabus:**

- Profile based searches & analysis (5\*3)
- Random forest: Prediction of Galactose binding sites
- Decision trees: prediction of n-linked glycosylations
- High throughput data analysis using R: QTL, CNV, LoH
- DNA & Protein Array data Analysis

##### **References:**

- Keith Jonathan M. Bioinformatics: Structure, Function and Applications Volume II. Publisher: Australia, Humana Press. 2008. ISBN: 9781603274289.
- Parida Laxmi. Pattern Discovery in Bioinformatics: Theory & Algorithms Publisher: New York, Chapman and Hall/CRC. 2008. ISBN 9781584885498
- Keedwell Edward, Narayanan Ajit Intelligent Bioinformatics: The Application of Artificial Intelligence Techniques to Bioinformatics Problems. Publisher: England, John Wiley & Sons Ltd. 2005. ISBN: 0470021756.
- Eidhammer Ingvar, Jonassen Inge, Taylor William R. Protein Bioinformatics: An Algorithmic Approach to Sequence and Structure Analysis. Publisher: New Delhi, Wiley India Pvt. Ltd. 2004. ISBN: 9788126522729.
- K. V. Raman & Pal Sourav. Mathematics in Chemistry. New Delhi, Vikas Publishing House Pvt. Ltd., 2004. ISBN: 8125912886.
- Durbin Richard, Eddy Sean R., Anders Krogh. Graeme Mitchison Biological

Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Publisher: New York, Cambridge University Press, 2003. ISBN: 0521540798.

- Goldberg David E. Genetic Algorithms in Search, Optimization, and Machine Learning. Publisher: Delhi, Pearson Education Pvt. Ltd. 2002. ISBN: 817808130X.
- Goodfellow Julia M. Molecular Dynamics: Applications in Molecular Biology. Publisher: London, The Macmillan Press Ltd. 1991. ISBN: 0333498860
- Davies E. R. Machine Vision: Theory, Algorithms, Practicalities. Publisher: New York, Academic Press Publishers. 1990. ISBN: 0122060903.
- Glassner Andrew S. Graphics Gems. Publisher: New York, Academic Press Inc. 1990. ISBN: 0122861655.
- Heijne Gunnar Von. Sequence Analysis in Molecular Biology: Treasure Trove or Trivial Pursuit Publisher: London, Academic Press. 1987. ISBN: 0127251308.

### **BIM 402 (T+P) : Metabolic Engineering & Systems Biology (3T+1P credits)**

#### ***Theory***

##### ***Objectives:***

This course will enable the students to:

- learn the basic biochemical concepts of metabolic pathways
- understand the role of Bioinformatics in the study of metabolic pathways & pathway databases
- learn the Bioinformatics-based approaches for predicting and engineering metabolic pathways

##### ***Syllabus:***

- Major Metabolic Pathways (10)
  - Gluconeogenesis, Pentose phosphate pathway, Glycogen synthesis and degradation, Fatty acid oxidation and synthesis, Amino acid catabolism, Purine and pyrimidine nucleotide synthesis etc.
- Enzyme Nomenclature and Classification (5)
- Metabolic Pathways databases (5)
  - KEGG, EcoCyc and MetaCyc
  - EMP, Malaria Parasite Metabolic Pathways, Boehringer Mannheim - Biochemical Pathways
- Enzymes, Compounds and Reactions databases (2)
  - LIGAND - Biochemical Compounds and Reactions
  - ENZYME - Enzymes
  - BRENDA - Comprehensive Enzyme Information System
- Mathematical modeling of metabolic pathways (4)
- Prokaryotes Vs Eukaryotes – single cell Vs multi cell species biology from a modeling perspective (1)
- Systems approach to Metabolism and integration of different data types (2)
- Properties of biochemical networks: structural, topological and dynamic (8)
- Determining Optimal states and applications (2)
- Metabolic Engineering- Basics (2)
- *E. coli* as a Host for metabolic engineering (2)
- Rational Metabolic engineering design (2)

### ***Practicals***

#### ***Objectives:***

This course will enable the students to:

- use metabolic pathway databases and associated tools
- use advanced methodologies to model pathways

#### ***Syllabus:***

- Browse & search metabolic pathway databases (2)
- MATLAB introduction (3)
- Structural and topological properties of biochemical networks (3)
- Setting up Genome scale models (3)
- Computations of phenotypes and interpretation using genome scale models (3)
- Mini project (2)

#### ***References:***

- Klipp Edda, Liebermeister Wolfram, Wierling Christoph, Kowald Axel, Lehrach Hans, Herwig Ralf. Systems Biology: A Textbook. Publisher: Germany, Wiley-VCH. 2009. ISBN: 9783527318742.
- Nelson David L., Cox Michale. Lehninger Principles of Biochemistry 5th Edition. Publisher: New York. W. H. Freeman. 2008. ISBN: 9780716771081.
- Alterovitz Gil, Ramoni Marco F. Systems Bioinformatics: An Engineering Case-Based Approach. Publisher: Boston, Artech House Publishers. 2007. ISBN: 9781596931244.
- Palsson Bernhard O. Systems Biology: Properties of Reconstructed Networks. Publisher: New York, Cambridge University Press. 2007. ISBN: 9780521859035.
- Voet, Donald, Voe Judith, Pratt, Charlotte W. Fundamentals of Biochemistry: Life at the molecular Level 2nd Edition. Publisher: Asia, John Wiley & Sons. 2006. ISBN: 0471753416.
- Stephanopoulos Gregory N., Aristidou Aristos A., Nielsen Jens. Metabolic Engineering: Principles and Methodologies. Publisher: New Delhi, Reed Elsevier India Pvt. Ltd. 2006. ISBN: 9788131203330.
- Purich Daniel L., Allison R. Donald. The Enzyme Reference: A Comprehensive Guidebook to Enzyme Nomenclature, Reactions, and Methods. Publisher: California, Academic Press. 2002. ISBN: 0125680414.
- Collado-Vides Julio, Hofstadt Ralf. Gene Regulation and Metabolism: Post-Genomic Computational Approaches. Publisher: England, The MIT Press. 2002. ISBN: 026203297X.
- Anton Howard. Elementary Linear Algebra 5th Edition. Publisher: New York, John Wiley & Sons. 1987. ISBN: 0471852236.

### **BIM 403 (T) : Emerging Areas in Bioinformatics.**

**(1 credits)**

#### ***Theory***

#### ***Objectives:***

This course will enable the students to:

- appreciate recent developments in Life sciences in the context of Bioinformatics through a series of lectures on appropriate topics.

- understand the fundamental principles and practises of intellectual property rights and their application to bioinformatics and related areas

**Syllabus :**

Series of talks on any 2-3 topics such as Biodiversity informatics, Neuro bioinformatics, Epigenetics, Advanced Glycobiology, Nano–biotechnology, System Biology, Agro bioinformatics etc.

- Fundamentals of IPR: (5)
  - Patents and Reading the Patent
  - International Framework
  - Patent Searching – Sources and Techniques
  - Patinformatics
  - Patent Prosecution
  - Types of Claims in Biotech Inventions
  - IPR Management
- IPR in bioinformatics, computer science, medical research, pharma, biodiversity, genetic resources & biotechnology (3)

**BIM 404 (P) : Project work**

**(16P credits)**

**Objectives:**

This course will enable the students to:

- understand and implement research methodology
- compile datasets, develop codes to process data , analyse and interpret results obtained
- acquire skills in scientific writing

**Any one of the followings:**

<i>Course No.</i>	<i>Course Name</i>	<i>No. of credits</i>
BIM 405 (T+P)	Seminars on Applications of Bioinformatics in Agriculture	1
BIM 406 (T+P)	Seminars on Applications of Bioinformatics in Human Health	1
BIM 407 (T+P)	Seminars on Applications of Bioinformatics in Environment	1
BIM 408 (T+P)	Seminars on Applications of Bioinformatics in Biotechnology	1
BIM 409 (T+P)	Seminars on Applications of Bioinformatics in Molecular Biology	1
BIM 410 (T+P)	Seminars on Applications of Bioinformatics in Neurobiology	1
BIM 411 (T+P)	Seminars on Applications of Bioinformatics in Drug Designing	1
BIM 412 (T+P)	Seminars on Applications of Bioinformatics in Veterinary Sciences	1