COURSE STRUCTURE OF THE “O LEVEL (BIOINFORMATICS)”

Objectives:

The course has been designed to be an entry level in Bioinformatics. It is introductory in nature and will provide an overview of the concepts and practices in Bioinformatics. The course structure has been designed such that students will acquire skills required to become Assistant Programmer/Technical Assistant in Bioinformatics. It would also help students to acquire a good foundation to take up further studies.

S. No.	Paper Code	Title
1.	BI- M1- R0	Foundation Course in Modern Biology
2.	BI- M2- R0	Basic Bioinformatics
3.	BI- M3- R0	IT Tools and Application
4.	BI- M4 – R0	(Electives) **Student has to opt one out of the following two:** BI –M4.1 –R0 Programming and Problem solving through ‘C’. BI –M4.2 –R0 Programming through Visual Basic.
BI –PJ	Project (On Bioinformatics)	

Practical

PR-I Basic Bioinformatics

COURSE STRUCTURE OF THE “A LEVEL (BIOINFORMATICS)”

S. No.	Paper Code	Title
Semester I		
1.	BI-A3-R0	IT Tools and Application
2.	BI-A1-R0	Basic Mathematics, Probability and Statistics
3.	BI-A4-R0	Programming and Problem solving through C
4.	BI-A2-R0	Basic Bioinformatics
5.	BI-A5-R0	Foundation Course in Modern Biology
Semester II		
6.	BI-A6-R0	Introduction to Database and Web enabling technologies
7.	BI-A7-R0	PERL/PYTHON Programming and applications to Bioinformatics
8.	BI-A8-R0	Introduction to Object Oriented programming through JAVA
9.	BI-A9-R0	Elements of protein Sequence, Structure and modelling
10.	BI-A10-R0	Basics of Genomics and Proteomics

PR-I	Basic Bioinformatics
PR- II	PERL & JAVA
BI- PJ	Project (On Bioinformatics)

COURSE STRUCTURE OF THE “B LEVEL (BIOINFORMATICS)”

S. No.	Paper Code	Title
Semester I		
1.	B1.1	IT Tools and Application
2.	B1.2	Basic Mathematics, Probability and Statistics
3.	B1.3	Programming and Problem solving through C
4.	B1.4	Basic Bioinformatics
5.	B1.5	Foundation Course in Modern Biology
Semester II		
6.	B2.1	Introduction to Database and Web enabling technologies
7.	B2.2	PERL/PYTHON Programming and applications to Bioinformatics
8.	B2.3	Introduction to Object Oriented programming through JAVA
9.	B2.4	Elements of protein Sequence, Structure and modelling
10.	B2.5	Basics of Genomics and Proteomics
Semester III		
11.	B3.1	Computer Organization and Distributed computing
12.	B3.2	Probability and Information theory
13.	B3.3	Computational methods in Biomolecular sequence analysis
14.	B3.4	Discrete Mathematics
Semester IV		
15.	B4.1	Statistical methods in Bioinformatics
16.	B4.2	Biomoleculer Structure and Dynamics
17.	B4.3	Data Structure and Algorithms
18.	B4.4	Computational Genomics
Semester V		
19.	B5.1	Optimisation, Machine Learning and Computational Intelligence
20.	B5.2	Object Technology for Bioinformatics
21.	B5.3	Computational Proteomics and Gene Expression studies
Optional Course		
22.	B5.4.1	Computer aided Molecular Modeling and Drug Discovery (OR)
	B5.4.2	Chemoinformatics
Semester VI		**Project**

Note: Theory = 60 hours and Practical = 60 hours

Practicals:

PR-I	Basic Bioinformatics
PR- II	PERL & JAVA
PR-III	Microbiology
PR-IV	Molecular Biology
PR-V	Biochemistry

SYLLABUS

BI- M3/A3/B1.1-R0: IT TOOLS AND APPLICATIONS

Objective of the Course

This course has been designed to provide an introduction to Information Technology and IT Tools. The student will become IT literate, and will understand the basic IT terminology. The students will be able to understand the role of Information Technology and more specifically computers, communication technology and software in the present social and economic scenario.

The course has focus on the following:

- Overview of the Information Technolgy
- Computer Organization
- Computer Operating System and Software
- MS Windows, LINUX, Word Processing and Spreadsheet tools
- Multimedia and Presentation Packages

Outline of Course

S. No.	Topic	Minimum No. of Hours
1.	Computer Appreciation	04
2.	Computer Organization	13
3.	Operating System	13
4.	Word Processing	10
5.	Spreadsheet Package	10
6.	Presentation Package	06
7.	Information Technology and Society	04

Lectures = 60
Practical / Tutorials = 60
Total = 120

Detailed Syllabus

1. Computer Appreciation 4 Hrs.

Characteristics of Computers, Input, Output, Storage units, CPU, Computer System, Binary Number System Binary to Decimal Conversion, Decimal to Binary Conversion, Binary Coded Decimal (BCD) Code, ASCII Code

2. Computer Organisation 13 Hrs.

2.1 Central Processing Unit 1 Hr.

Control Unit, Arithmetic Unit, Instruction Set, Register, Processor Speed

2.2 Memory 3 Hrs.

Main Memory: Storage Evaluation Criteria, Memory Organization, capacity, RAM, Read Only Memories. **Secondary Storage Devices**: Magnetic Disks, Floppy and Hard Disks, Optical Disks CD-ROM, Mass Storage Devices.

2.3 Input Devices 2 Hrs.

Keyboard, Mouse, trackball, joystick, Scanner, OMR, Bar-code reader, MICR Digitizer, Card Reader, Voice Recognition, web cam, video cameras.

2.4 Output Devices 2 Hrs.

Monitors, Printers – Dot matrix, inkjet, laser, plotters, Computer output Micro-Film (COM), Multimedia Projector, speech synthesizer, dumb, smart and intelligent terminal.

2.5 Multimedia: 2 Hrs.

What is Multimedia, Text, Graphics, Animation, Audio, images, Video, Multimedia Application in Education, Entertainment, Marketing.

2.6 Computer Software 3 Hrs.

Relationship between Hardware and Software ; system software, application software, compiler, names of some high level languages, free domain software.

3. Operating Systems 13 Hrs.

1.1 Disk Operating System

Simple DOS Commands, Simple File Operations, Directory related Commands

1.2 Microsoft Windows

An overview of different versions of Windows, Basic Windows elements, Files management through Windows.

Using essential accessories: Systems tools – Disk cleanup, Disk defragmenter, Entertainment, Games, Calculator, Imaging – Fax, Notepad, Paint, WordPad.

1.3 Linux

An overview of Linux , Basic Linux elements: System Features, Software Features. File Structure,
File handling in Linux, Installation of Linux: H/W, S/W requirements, Preliminary steps before installation, specifics on Hard drive repartitioning and booting a Linux System.

4. Word Processing 10 Hrs.

Word Processing concepts: Saving, Closing, Opening an existing document, Selecting text, Editing text, Finding and replacing text, printing documents, Creating and Printing Merged documents, Character and Paragraph Formatting, Page Design and Layout.

Editing and Proofing Tools: Checking and correcting spellings. Handling Graphics. Creating Tables and Charts. Document Templates and Wizards*.

5. Spreadsheet Package 10 Hrs.

Spreadsheet Concepts. Creating, Saving and Editing a workbook, Inserting, Deleting work Sheets, entering data in a cell/formula. Copying and Moving data from selected cells, Handling operators, in Formulae, Functions: Mathematical. Logical, Statistical, Text, Financial, Data ad Time Functions, Using Function wizard.

Formatting a Worksheet : Formatting Cells -- Changing data alignment, Changing date, number, character, or currency format, changing font, adding borders and colors, printing worksheets, Charts and Graphs – Creating, Previewing, Modifying Charts.

Integrating word processor, spread sheets, web pages*.

6. Presentation Package 6 Hrs.

Creating, Opening and Saving Presentations, Creating the Look of Your Presentation, Working in Different Views, Working with Slides, Adding and Formatting Text, Formatting Paragraphs, Checking Spelling and Correcting Typing Mistakes, Making Notes Pages and Handouts, Drawing and Working with Objects, Adding Clip Art and Other Pictures, Designing Slide Shows, Running and Controlling a Slide Show, Printing Presentations*.

7. Information Technology and Society 4 Hrs.

Application of Information Technology in Railways, Airlines, Banking, Insurance, Inventory Control, Financial Systems, Hotel Management, Education, Video Games, Telephone exchanges, Mobile Phones, Information Kiosks, Special effects in Movies.

* Note: The underlying concepts may be illustrated using MS Office Package.

RECOMMENDED BOOKS

MAIN READING

1. P. K. Sinha and P. Sinha (2002), “Foundations of Computing”, First Edition, BPB Publication.
2. S. Sagman, “Microsoft Office 2000 for Windows”, Second Indian Print, 2000, Pearson Education.

SUPPLEMENTARY READING

1. Turban, Mclean and Wetherbe, “Information Technology and Management”, Second Edition, 2001, John Wiley & Sons.

BI-M4.1/A4/B1.3-R0: PROGRAMMING AND PROBLEM SOLVING THROUGH ‘C’ LANGUAGE

Objective of the Course

The objectives of this course are to make the student understand programming language, programming, concepts of Loops, reading a set of Data, stepwise refinement, Functions, Control structure, Arrays. After completion of this course the student is expected to analyse the real life problem and write a program in ‘C’ language to solve the problem. The main emphasis of the course will be on problem solving aspect, i.e. developing proper algorithms.

- After completion of the course the student will be able to
- develop efficient algorithms for solving a problem
- Use the various constructs of a programming language viz. conditional, iteration and recursion.
- Implement the algorithms in ‘C’ language
- Use simple data structures like arrays, stacks and linked list in solving problems
- Handling File in ‘C’.

Outline of Course

S.No.	Topic	Minimum No. of Hours
1.	Introduction to Programming	04
2.	Algorithms for Problem solving	12
3.	Introduction to ‘C’ language	04
4.	Conditionals and loops	08
5.	Arrays	06
6.	Functions	06
7.	Structures and Unions	06
8.	Pointers	06
9.	Self Referential Structures and Linked Lists	04
10.	File Processing	04

Lectures = 60
Practicals/ Tutorials = 60
Total = 120

Detailed Syllabus

1. Introduction to Programming 4 hours

The Basic Model of Computation, Algorithms, Flow-Charts, Programming Languages, Compilation, Linking and Loading, Testing and Debugging, Documentation

2. Algorithms for Problem Solving 12 hours

Exchanging values of two variables, summation of a set of numbers, Decimal Base to Binary Base conversion, Reversing digits of an Integer, GCD (Greatest Common Division) of two numbers, Test whether a number is prime, Organise numbers in ascending order, Find square root of a number, factorial computation, Fibonacci sequence, Evaluate ‘sin x’ as sum of a series, Reverse order or elements of an array, find Largest number in an array, Print elements of upper triangular matrix, multiplication of two matrices, Evaluate a Polynomial.

3. Introduction to ‘C’ language 4 hours

3.1 Character set, Variable and Identifiers, Built-in-Data Types, Variable Definition
3.2 Arithmetic operators and Expressions, Constants and Literals
3.3 Simple assignment statement, Basic input/output statement
3.4 Simple ‘C’ programs

4. Conditional Statements and Loops 8 hours

4.1 Decision making within a program
4.2 Conditions, Relational Operators, Logical Connectives
4.3 *if* statement, *if-else* statement
4.4 Loops: *while* loop, do while, *for* loop, nested loops, infinite loops, *switch* statement, structured programming

5. Arrays 6 hours

One dimensional arrays: Array manipulation; Searching, Insertion, Deletion of an element from an array; Finding the largest /smallest element in an array; Two dimensional arrays, Addition/ Multiplication of two matrices, Transpose of a square matrix; Null terminated strings as array of characters, Representation sparse matrices

6. Functions 6 hours

Top down approach of problem solving, Modular programming and functions, Standard Library of C functions, Prototype of a function: Formal parameter list, Return Type, Function call, Block structure, Passing arguments to a function: call by reference, call by value, Recursive Functions, arrays as function arguments

7. Structures and Unions 6 hours

Structure variables, initialization, structure assignment, nested structure, structures and functions, structures and arrays: arrays of structures, structures containing arrays, unions

8. Pointers 6 hours

Address operators, pointer type declaration, pointer assignment, pointer initialization, pointer arithmetic, functions and pointers, Arrays and Pointers, pointer arrays

9. Self Referential Structures and Linked Lists 4 hours

Creation of a singly connected linked list, traversing a link list, Insertion into a linked list, Deletion from a linked list

10. File Processing 4 hours

Concepts of Files, File opening in various modes and closing of a file, Reading from a file, Writing onto a file

RECOMMENDED BOOKS

MAIN READING

1. Byron Gottfried, “Programming with C” Second Edition, Tata McGrawhill, 2000
2. R. G. Dromey, “How to solve it by Computer”, Seventh Edition, 2001, Prentice Hall of India

SUPPLEMENTARY READING

1. E. Balaguruswami, “Programming with ANSI-C”, First Edition, 1996, Tata McGrawhill
2. A. Kamthane, “Programming with ANSI & Turbo C”, First Edition, 2001, Pearson Education
3. Venugopal and Prasad, “Programming with ‘C’”, First Edition, 1997, Tata McGrawhill
4. B. W. Kernighan & D. M. Ritchie, “The C Programming Language”, Second Edition 2001, Prentice Hall of India

BI-M2/A2/B1.4-R0: BASIC BIOINFORMATICS

Objective of the Course

On completion of this course, students will have basic knowledge of sources of sequences and protein structure data, an understanding of the relevance and importance of this data, and some exposure to basic algorithms used for processing this data.

Outline of Course

S. No.	Topic	Minimum No. of Hours
1	Introduction to Genes and Proteins	10
2	Introduction to Internet Use and Search Engines	5
3	Genomic Sequence Information Sources	10
4	Protein Structure Information Sources	10
5	Introduction to Data Generating Methods	5
6	Sequence and Phylogeny Analysis	20

Lecture = 60
Practicals/Tutorials = 60
Total = 120

Detailed Syllabus

1. Introduction to Genes and Proteins 10 hours

Genome Sequences
ORFs, Genes, Introns, Exons, Splice Variants
DNA/RNA Secondary Structure Triplet Coding
Protein Sequences
Protein Structure: Secondary, Tertiary, Quaternary
The notion of Homology

2. Introduction to Internet Use and Search Engines 5 hours

WWW, HTML, URLs
Browsers: Netscape / Opera / Explorer
Search Engines Google, PUBMED

3. Sequence Information Sources 10 hours

EMBL
GENBANK
Entrez
Unigene
Understanding the structure of each source and using it on the web

4. Protein Information Sources 10 hours

PDB
SwissProt
TrEMBL
Understanding the structure of each source and using it on the web

5.	**Introduction to Data Generating Techniques**	**5 hours**
	Restriction Enzymes, Gel Electrophoresis, Chromatograms	
	Blots, PCR, Microarrays, Mass Spectrometry	
	What data each generates, and what bioinformatic problems they pose.	
6.	**Sequence and Phylogeny Analysis**	**20 hours**
	Detecting Open Reading Frames	**….2 hours**
	Outline of Sequence Assembly	**….3 hours**
	Mutation Matrices	**….2 hours**
	Pairwise Alignments	**….2 hours**
	Introduction to BLAST, using it on the web, interpreting results	**….4 hours**
	Multiple Sequence Alignment	**….3 hours**
	Phylogenetic Analysis	**….4 hours**

RECOMMENDED BOOKS

MAIN READING

Theory

1. D. Baxevanis and F. Oulette, (2002) “Bioinformatics : A practical guide to the analysis of genes and proteins”, Wiley Indian Edition

2. Cynthia Gibas and Per Jambeck (2001), “Developing Bioinformatics Computer Skills”. O’Reilly press, Shorff Publishers and Distributors Pvt. Ltd., Mumbai.

3. Bryan Bergeron MD (2003), “Bioinformatics Computing”. Prentice Hall India (Economy Edition)

4. Stuart Brown (2000) “Bioinformatics – A biologists guide to Biocomputing and Internet”. Eaton Publishing

Practical

1. Jean-Michel Claverie and Cedric Notredame (2003) Bioinformatics – A Beginners Guide. Wiley – Dreamtech India Pvt. Ltd.

SUPPLEMENTARY READING

Theory

1. T. K. Attwood & D. J. Parry-Smith (2001), “Introduction to Bioinformatics”, Pearson Education Ltd, Low Price Edition.

2. Bioinformatics: Sequence and Genome Analysis. D. W. Mount (2001) Cold Spring Harbor Laboratory Press.

3. Arthur M. Lesk (2002) “Introduction to Bioinformatis” Oxford University Press

BI-M1/A1/B1.5-R0: FOUNDATION COURSE IN MODERN BIOLOGY

Objective of the course

The objective of this foundation course is to familiarize students to basic terminology and concepts in the subject. Unifying concepts and theories pointed out and stressed wherever applicable.

After completion of this course students should be able to

- Understand the underlying basic concepts in biology
- Be familiar with basic terminology
- Relate to the functioning of cellular machinery
- Appreciate the role of heredity and evolution in our ecology

Outline of Course

S. No.	Topic	Minimum No. of Hours
1	General Biology	6
2	Basics of Cell Biology	8
3	Introductory Life Science – Botany and Zoology	6
4	Introduction to Human Biology	8
5	General Microbiology	6
6	Basics of Genetics and Evolution	10
7	Basic Molecular Biology I – Nucleic acids (DNA and RNA)	10
8	Basic Molecular Biology II – Proteins	6

Lecture = 60
Practicals/Tutorials = 60
Total = 120

Detailed Syllabus

1. General Biology **6 hours**

The nature of life, definition of life, characteristics of life, differences between animals and plants, principal divisions in biology, importance of biology

2. Basics of Cell Biology **8 hours**

Definition of Cell – fundamental cell types, difference between prokaryotic and eukaryotic cell types, cell structure, cell wall, plasma membrane, different organelles and their function. Cell division stages, chromosome structure, human chromosomes, sex chromosome.

3. Introductory Life Science -- Botany and Zoology **6 hours**

What is a tissue, classification of tissue, meristematic tissue, intercellular space.
Reproduction and development of seed and non-seed plants, levels of organization, form and function of systems, and a survey of major taxa.

4. **Introduction to Human Biology** 8 hours

Types of tissue in human : connective tissue, supporting tissues fluid tissue, muscular tissue, nerve tissue, secretory tissue.
Human respiration : respiratory organs, mechanism of breathing in man, composition of different respiratory gases, compartment of lung, gas exchange in lungs, transport of gases during respiration, diseases of respiratory system. Human nutrition: Salivary glands, function of saliva, swallowing process, digestion in stomach, secretory cells of stomach, digestion in small intestine, pancreas, liver, bile, role of bile in fat digestion.

Human circulation : heart, blood flow control, diseases of heart and circulatory system. Human excretion: Function of kidney, large intestine, summary of excretory substances, diseases related to excretory system. Movement and locomotion: Organs responsible for movement, joints, bones and muscles involved in walking, different sequences during locomotion.

Reproduction: Formation of sperm and ova, fertilization, Hormones: Regulation of hormone secretion

5. General Microbiology 6 hours

Microbes in Our Lives: Definition of microorganisms, role in ecological balance, microorganisms living in humans and animals, their role, microorganisms used to produce food and chemicals, disease causing microorganisms.

A Brief History of Microbiology: Robert Hookes observation in 1665, Antoni van Leeuwwenhoek sees microorganisms using a simple microscope (1673), Rudolf Virchow introduced the concept of biogenesis: (1858). Pasteur’s discoveries (1861). Alexander Fleming’s Penicillin discovery (1928). Penicillin has been used clinically an as antibiotic since the 1940s.

Fermentation and Pasteurization, Vaccination, Recombinant DNA Technology

Naming and Classifying of Microorganism

In a nomenclature system designed by Carolus Linnaeus (1735), each living organism is assigned two names, The two names consist of a genus and a specific epithet, both of which are underlined or italicized. All organisms are classified into Eubacteria, Archaea and Eucarya. Eucarya includes Protista, Fungi, Plantae and Animalia.

The diversity of Microorganisms

Bacteria, Fungi, Protozoa, Algae, Viruses, Multicelluer Animal Parasites

6. Basics of Genetics and Evolution 10 hours

Mendel’s work and experiments, gene bearer of heredity character, chemical basis of heredity.

Stages of evolution, origin of life, evidences of evolution from plant and animal kingdom. Genome sequencing techniques.

7. Basic Molecular Biology I – Nucleic Acids (DNA and RNA) 10 hours

Chemical structure, hybridization, double helical structures, replication, concepts of gene and genetic code, transcription and translation, mutations and its implications, Polymerase Chain Reaction.

8. Basic molecular Biology II – Proteins 6 hours

Amino acid structure, chemical nature of residues, polymeric chain and folding, concepts of pH, pKa, buffer, aqueous medium.

RECOMMENDED BOOKS

MAIN READING

1. Biological Science, Third Edition, N. P. O Green, G. W. Stout, D. J. Taylor, Editor, R. Soper, Cambridge Low priced Edition
2. Molecular Biology of the Cell, 3rd Edition, Bruce Alberts et al., Garland Publishing, New York.
3. Lehninger Principles of Biochemistry, 4th Edition, David Nelson and Michael M Cox, Freeman Publishers, 2004

SUPPLEMENTARY READING

1. DNA Structure and Function, R. R. Sinden, Academic Press, San Diego (1994)
2. Genes VIII Benjamin Lewin, Oxford University Press

BI-A5/B1.2-R0: BASIC MATHEMATICS, PROBABILITY AND STATISTICS

Objective of the Course

Mathematical and statistical frameworks are being increasingly employed to understand and investigate biological processes. These frameworks helps in analyzing vast amount of datasets generated from genome and related projects. It is thus essential to introduce basic concepts of mathematics, probability and statistics early within the Bioinformatics curriculum. This course will enable students to understand and appreciate computational problems in proper perspective. In addition, this course will provide a foundation for pursuing higher level courses in Computational Biology.

Outlines of Course

S. No.	Topics	Minimum No. of hours
1.	Real Numbers	2
2.	Set, Relation and Function	3
3.	Limits	4
4.	Binomial Theorem	2
5.	Calculus	12
6.	Matrices and Vectors	5
7.	Complex Numbers	2
8.	Elementary Statistics	6
9.	Regression and Correlation	4
10.	Probability	8
11.	Random Variables and Probability Distribution	12

Lectures = 60
Practicals/ Tutorials = 60
Total = 120

Detailed Syllabus

1. Real Numbers **2 Hours**

Different kinds of number Integer, Rational and Irrational
Surds and their properties, Fractional indices

2. Set, Relation and Function **3 Hours**

Set, Product sets
Relations
Functions (Polynomials, Trigonometric, Exponential)
Graphical Representation of Functions

3. Limits **4 Hours**

Sequences
Limits of sequences
Series
Limits of functions

4. Binomial Theorem **2 Hours**

Expanding $(x+y)^n$
Binomoal Coffiecients, Binomial Theorem

5. Calculus **12 Hours**

a) Differentiation **….5 Hours**
Calculating gradients of chords
First order and higher derivatives
Applications : Increasing and Decreasing
Functions, Maximum and Minimum points
Derivatives as rates of change

b) Integration **….5 Hours**
Finding a function from its derivative
Definite integral
Calculating Areas, Volumes

c) Differential Equations **….2 Hours**
Forming differential equations
First order differential equation, growth equation
Applications

6. Matrices and Vectors **5 hours**

Matrix algebra, Determinants
Applications
Vector in space
Applications

7. Complex Numbers **2 hours**

Extending the number system
Operations with complex numbers

8. Elementary Statistics **6 hours**

a) Representation of data :
Discrete Data, Continuous data
Histogram, Polygons, Frequency Curves
b) The Mean,
Variability of data –
The standard deviation
c) Median, quantiles, percentile
d) Skewness
e) Box and Whisker diagrams (box plots)

9. Regression and Correlation **4 hours**
Scatter diagrams
Regression function
Linear correlation and regression lines
Product moment correlation coefficient

10. Probability **8 hours**

Experimental probability ….. **2 hours**
Probability when outcomes are equally likely
Subjective probabilities
Probabilities law ….. **4 hours**
Probability rules for combined events
Conditional probability and independent events
Probability trees ….. **2 hours**
Bayes' theorem

11. Random Variables and Distributions **12 hours**

a) Discrete and Continuous Random Variables ….. **5 hours**
Cumulative distribution function
Probability mass function and Probability Density function
Expectation of random variables – Experimental Approach and theoretical approach
Expectation of X and variance of X
Expectation of function E[g(X)]

b) Bernoulli Distribution **3 hours**
Binomial Distribution
Poisson Distribution

c) Uniform Distribution **4 hours**
Normal Distribution
Normal approximation to Binomial Distribution
Central limit theorem

RECOMMENDED BOOKS

MAIN READING

1. H. Nell and D. Quadling, 'Pure Mathematics (Advanced Level Mathematics)'. Vol. 1,2,3, Cambridge University Press 2002.
2. Edward Batschelet. 'Introduction to Mathematics for Life Scientists', 3rd Edition, Springer – Verlag, 1992
3. J. Crawshaw and J. Chambers', 'Advanced level Statistics', 4th Edition, Nelson Thornes, 2002

SUPPLEMENTARY READING

1. Sheldon Ross, 'A first Course in Probability', Sixth Edition, Pearson Education Asia, 2002
2. S. Dobbs and J. Millser, 'Statistics (Advanced Level mathematics);, Cambridge University Press 2002

BI-A6/B2.1-R0: INTRODUCTION DATABASE AND WEB ENABLING TECHNOLOGIES

Objective of the course

On completion of this course students will have adequate knowledge of different data models, basic principles of data base management and basic web enabling technologies. Students will learn how to design a database system, various normalization techniques and some useful query language. Students will be able to develop web based applications to retrieve information from data bases. They should also be able to design their own web sites, install web server, and do the relevant administrative tasks.

Outline of Course

S. No.	Topic	Minimum No. of Hours
1	Introduction to databases	5
2	Entities Relationship model	4
3	Relational Algebra and Calculus	7
4	Issues in designing relational databases	6
5	Query language and query optimization	10
6	Database system architecture	4
7	Introduction to ASN.1 and NCBI data model	4
8	Basics of Internet and WWW	3
9	Web server	6
10	HTML	8
11	Introduction to XML and its difference with HTML	3

Lectures = 60
Practicals/Tutorials = 60
Total = 120

Detailed Syllabus

1. Introduction to Databases **5 hours**

Why Database systems, Data abstraction and data models, Instances and schemas, Database Administrator,

Data Definition and manipulation languages, brief introduction to network and hierarchical models.

2. Entitles Relationship model **4 hours**

Entity and entity sets, relationships and relationship sets,, E-R diagram, reducing E-R diagrams to Tables and trees.

3. Relational Algebra and Calculus **7 hours**

Relational algebraic operations such as select, project, union, set difference, Cartesian product,, intersections, natural join, division, generalized projection, outer join etc., tuple relational calculus, domain relational calculus.

4. Issues in designing relational databases **6 hours**

Pitfalls in relational database design, decomposition, importance of normalization, functional dependencies, Boyce-Code Normal form, third normal form and fourth normal form.

5. Query language and query optimization **10 hours**

Domain types in SQL, Schema definition in SQL, Types of SQL commands, SQL operators, tables, views, indexes, aggregate functions, insert, delete and update operations, join, union, intersection, minus etc. in SQL, queries, sub-queries, equivalence of queries.

6. Database system architecture **4 hours**

Introduction to centralized system, client server system, parallel system and distributed system.

7. Introduction to ASN.1 and NCBI data model **4 hours**

Why specialized data model is required for biological sequences, different data types supported by ASN.1and how they are used for storage of different types of information reading of NCBI data using freely available NCBI toolbox.

8. Basics of Internet and WWW **3 hours**

Introduction to Internet, TCP/IP, WWW, FTP, registration with ISP, Internet connection wizard, URL http.

9. Web server **6 hours**

Role of Web Server, a brief introduction to Apache, Introduction to PSW, capabilities of PSW, installation of PSW, role of CGI program, configuring PSW for Perl / CGI, configuring system Data Source name (DSN), publishing dynamic applications, creating web pages from information contained in a data base, creation of internet database connection file, simple CGI programming using Perl for simple form processing, use of ODBC drivers to connect a database – role of internet database connector (.idc)files and HTML (.htx)files.

10. HTML **8 hours**

Introduction, common tags, creation of hyperlink, incorporation of images, Tables, Frames, formatting text with font.

Dynamic HTML, cascading style sheets, creation of background images, HTML object model, dynamic positioning, direct animation path control.

11. Introduction to XML and its differences with HTML **3 hours**

RECOMMENDED BOOKS

MAIN READING

1. H. M. Dietel, P. J. Dietel and T. R. Nieto, Internet and World Wide Web – how to program, Pearson Education India
2. A. Silberschatz, H. F. Korth and S. Sudarshan, Database System Concepts, McGraw Hill International
3. Zoe'Lacroix and critchlow. Bioinformatics : Managing scientific data. Morgan Kaufmann Publishers 2004.

SUPPLEMENTARY READINGS

1. A. Leon and M. Leon: Database Management Systems, Leon Vicas
2. C. J. Date: Introduction to Database Systems

BI-A7/B2.2-R0: PERL PROGRAMMING AND APPLICATION TO BIOINFORMATICS

Objective of the Course

Perl is a popular programming language that is extensively used in Bioinformatics. The goal of this course is to provide a practice-oriented introduction to Perl programming language. The goal is two-fold: to teach programming skills and to apply them for the solution of interesting Bioinformatics problems. The programming concepts such as data type, arithmetic and logical operators, conditionals and loops, input/output, regular expression and pattern matching, functions and sub-routines, application in reading protein files, finding motifs, simulating DNA, parsing PDB record, BLAST and annotations in GenBank, etc. After completion of the course a student will have solid understanding of Perl basics and will be able to program for tackling tasks described above.

Outlines of Course

S. No.	Topics	Minimum No. of hours
1.	Introduction	02
2.	Data Types	06
3.	Arithmetic and Logical Operators	10
4.	Conditionals and Loops	04
5.	Input and Output	02
6.	Regular Expression and Pattern Matching	12
7.	Function and Subroutines	12
8.	Applications of Perl in Bioinformatics	12

Lectures = 60
Practicals/ Tutorials = 60
Total = 120

Detailed Syllabus

1. Introduction: 2 hours

Introduction to Perl, Downloading and installation from Website, Writing and Running a Perl Program, Editing, Advantages

2. Data Types 6 hours

Scalar data and scalar variables: Number, String, Conversion between Numbers and Strings, Variable Interpolation, Arithmetic and Decimal Precision, Arrays: Initialization, Manipulation of Array elements; Associative Array (Hashes): Initialization, Manipulation of Elements of Array.

3. Arithmetic and Logical Operators 10 hours

Arithmetic Operators, Assignment Operators, Increment and Decrement Operators, String Concatenation and Repetition, Operators precedence and Associativity, Conditional Operators, Logical Operators, Operators for manipulating arrays, Operators for Manipulating hashes.

4. Conditionals and Loops 4 hours

Conditional Statement; if, if…else, if and if-else, unless statement, Loops: while, for, until, do..while, do..until and foreach loop, last next, redo, continue and case switch statement.

5. Input and Output 2 hours

Creating a file, Reading Data from a file, Writing data to a file, Closing a file, Managing Files and Directories

6. **Regular Expressions and Pattern Matching** **12 hours**

Regular Expression, Pattern Matching, Meta Character, Simple Pattern, Matching Group of Characters, Matching multiple instances of Characters, Pattern Building, Pattern and Variable, Pattern and Loops, Using Pattern for Search and Replace, Matching Pattern over multiple Lines etc.

7. **Function and Subroutines** **12 hours**

Built-in Functions, Defining and calling subroutines, Returning Values from Subroutines, Using Local Variables in Subroutines, Passing Values into Subroutine, Perl References, Perl module and their uses.

8. **Applications of Perl in Bioinformatics** **12 hours**

Concatenating DNA Fragments, Transcription: DNA to RNA, Reading Protein Files, Finding Motifs, Simulating DNA, Generating Random DNA, Analysing DNA, Translating DNA to Proteins, Reading DNA from Files in FASTA format, Separating Sequence and Annotation, Parsing Annotation, Parsing PDB files, Parsing BLAST output, Bio-perl .

RECOMMENDED BOOKS

MAIN READING

1. James Tisdall, “Beginning Perl for Bioinformatics”, O’Reilly & Associates, 2001
2. James Tisdall, “Mastering Perl for Bioinformatics”, O’Reilly, 2003

SUPPLEMENTARY BOOKS

1. Cynthia Gibas & Per Jambeck, “Developing Bioinformatics Computer Skills”, O’Reilly & Associates, 2000.
2. Rex A. Dawyer, “Genomic Perl”, Cambridge University Press
3. Learning Perl, 3rd Edition , Author: Randal L. Schwartz and Tom Phoenix, O’Reilly

BI-A8/B2.3-R0: INTRODUCTION TO OBJECT ORIENTED PROGRAMMING THROUGH JAVA

Objective of the Course

The course is designed to impart knowledge and skills required to solve the real world problems using object – oriented approach utilizing JAVA language constructs. This course covers the subject in two parts, viz., Java Language and Java Library.

After completion of the course students are expected to understand the following:

- Java tokens for creating expressions and creating Datatypes
- The way various expressions and data types are assembled in packages
- Implementation of Inheritance, Exception handling and Multithreading in JAVA.
- JAVA I/O basics and Applets
- Setting up GUI using AWT/Swing
- Network Programming in JAVA
- Accessing relational databases from JAVA programs

Outlines of Course

S. No.	Topics	Minimum No. of hours
1.	The JAVA Language	30
2.	The JAVA Library	30

Lectures = 60
Practicals/ Tutorials = 60
Total = 120

Detailed Syllabus

1. The JAVA Language 30 hours

1.1 Introduction to JAVA 02 hours

An Overview of JAVA, JAVA Applets and Applications.
Difference between Java Script and JAVA
Object Oriented Programming features

1.2 Data Types, Variable & Arrays 03 hours

Java Token & Keywords. Integers types. Floating point types

The JAVA class libraries, Declaring a variable, Dynamic initialization, The scope and lifetime of variable. Type conversion and casting.

Arrays: One-dimensional arrays, Multidimensional arrays, Alternative array declaration syntax

1.3 Operators 02 hours

Arithmetic operators, The Bitwise operators, Relational operators, Boolean logical operators, the assignment operator, The ? Operator, Operator precedence

1.4 Control statements — 03 hours

Selection statements, Iteration statements, Jump Statements

1.5 Introduction classes and objects — 04 hours

Class fundamentals, Declaring objects, Assigning object reference variables, Introducing methods, Constructors, The this keyboard, Garbage collection, The Finalize () method, a stack class, Overloading constructors, Using objects as parameters, Arguments passing, Returning objects, Recursion.

1.6 Inheritance — 04 hours

Inheritance basics, Member access and inheritance, Using super class, Creating a multilevel hierarchy, Method overriding, Dynamic method dispatch, Using abstract classes, Using final with inheritance, The object class

1.7 Packages and Interfaces — 03 hours

Packages ; Defining a package, Understanding classpath, Importing Packages
Interfaces : Defining an interface, Implementing interfaces, Applying Interfaces, Variable in interfaces

1.8 Exception handling — 03 hours

Exception handling fundamentals
Exception types: Uncaught exceptions. Using try and catch
JAVA's build-in exceptions. User defined exception subclasses

1.9 Multithreaded Programming — 03 hours

The JAVA thread model. The main thread, Creating a thread, Alive () and joint (), Suspend() and resume (), Thread priorities, Synchronization, Interthread communication

1.10 I/O, Applets and Other Topics — 03 hours

I/O Basics: Streams, The stream classes, The predefined streams, Reading console input, Writing console output, Reading and writing files, Applet fundamentals, The transient and volatile modifiers, Using instance of native methods

2. The JAVA Library — 30 hours

2.1 String handling — 02 hours

The string constructor, Special string operations, Character extraction, String Searching & Comparison, Data conversion using value of (), String buffer

2.2 Exploring JAVA Lang — 03 hours

Simple type wrappers, Runtime Memory management, Array copy, Object, Clone() and the cloneable interface, Class & Class loader

Math functions: Transcendental functions, Exponential functions, Rounding functions, Miscellaneous math methods

Compiler, Thread, Threadgroup and runnable, throwable, Security Manager

2.3 The utility classes 03 hours

The enumeration interface, Vector & Stack
Dictionary. Hash table, String tokenizer,
Bitset. Date: Date comparison, String and time zones
Random. Observer interface

2.4 Input/Output – Exploring JAVA I/O 03 hours

The JAVA I/O classes and interface
File Namefilter and Directories
I/O Stream classes: File Input Stream, File Output Stream, Byte array input stream, Byte array output stream, Filtered streams
Buffered Streams: Buffered Input Stream, Buffered Output Stream, Push back Input Stream, Sequence Input Stream, Print Stream, Random Access File

2.5 Networking 04 hours

Socket Overview, Reserved sockets, Proxy Servers
Internet Addressing: Domain naming services (DNS), JAVA and the net, The networking classes and interfaces, Inet Address: Factory methods, Introspection, TCP/IP server sockets
DataGrams: Datagram packet, Datagram server and client

2.6 The applet class 05 hours

The applet class, Applet architecture
An applet skeleton: Initialization and termination, Overriding update (), Status Window
Handling events: The event class, Processing mouse events
Handling keyboard events
HTML applet.tag, Passing parameters to applets
Applet context and show document (), The audioclip & appletstub interface, Outputting to the console

2.7 Swing 05 hours

Swing & its features
Text fields, Buttons, Toggle Buttons, Check boxes, and Radio Buttons, Viewports, Scrolling, Sliders and Lists, Combo Boxes, Progress Bars, Tooltips, Separators and Choosers, Layered Panes, Tabbed Panes, Split Panes and Layouts, Menus and Toolbars, Windows, Desktop Panes, Inner Frames, Dialogue Boxes, Tables and Trees, Text Components

2.8 Images 02 hours

File Formats
Image Fundamentals: Creating, Loading and Siaplaying
Image Observer: Double Buffering, Media Tracker

2.9 JAVA Database Connectivity (JDBC) 03 hours

Introduction to JDBC. Type of JDBC connectivity
Accessing relational database from Java programs
Establishing database connections

RECOMMENDED BOOKS

MAIN READING

1. H. Schildt, (2001), “The Complete Reference – Java 2”, Fourth Edition, Tata McGraw Hill
2. Dietel nad Dietel (2001), “Java: How to program Java 2”, Second Edition, Pearson Education

SUPPLEMENTARY READING

1. D. Hanagan (2001), “Java Examples in a Nutshell”, Third Edition, O’ Reilly
2. K. Mughal and R. W. Rasmuessen, “A Programmers Guide to Java Certification – First Edition, (1999), Pearson Education Comprehensive Primer” publication
3. M. T. Nelson, “ Java Foundation Classes”, Tata McGraw Hill

BI-A9/B2.4-R0 : ELEMENTS OF PROTEIN SEQUENCE, STRUCTURE & MODELLING

Objective of the Course

The course is formulated keeping in view first year Bachelor's students with no biology background. The essential focus is on a qualitative and elementary exposition of the idea that proteins are important biological polymer molecules made of amino acid units, that proteins possess unique structures which in turn dictate their function inside the cell and that it is possible to store, retrieve and represent this information on computers. Other biological macromolecules are introduced towards the end of the course. After attending the course, the student is expected to be acquainted with protein sequences, structures and their role inside the cells, their relation to diseases and drug targets, and how computers may help in understanding structure and function of proteins.

Outline of Course

S. No.	Topic	Minimum No. of Hrs.
1.	Protein Sequences & Structure	10
2.	Ramachandran Plots	03
3.	Protein databases & visualization	07
4.	Structure to function relationships	05
5.	Homology modeling of proteins	07
6.	Protein conformation	06
7.	Experimental techniques for proteins	04
8.	Introductory Proteomics	02
9.	Proteins as drug targets	06
10.	Structure and function of Nucleic acids, lipids & carbohydrates	10

Lectures	**=**	**60**
Practicals/Tutorials	**=**	**60**
Total	**=**	**120**

Detailed Syllabus

1. Protein Sequences & Structure 10 hrs.

a> Building Blocks : Amino acids and their classification, peptide bond, polypeptides, proteins.

b> Primary structure – protein sequence.

c> Secondary structures : alpha helix, beta sheet (parallel & antiparallel), loops and turns.

d> Tertiary structure

e> Quarternary structure (more than one polypeptide chain)

f> Classification of protein structures

2. Ramachandran Plots **3 hrs.**

a> Glycine

b> Alanine

c> An example of any polypeptide

3. Protein databases & visualization **7 hrs.**

a> Protein databases :

  i) Sequence databases : NCBI, TIGR, SWISS-PROT

  ii) Structure database : RCSB (and links therein)

  iii) Other databases of conserved domains, motifs, protein families, etc.

b> Sequence retrieval : web browsing, ftp

c> Structure retrieval and visualization of 3D structures using Rasmol and other public domain software.

4. Structure to function relationships **5 hrs**

Sequence to function and Structure to function relationships -

a> Some case studies

5. Homology modeling of proteins **7 hrs.**

Basic Concepts in modeling of proteins through homology

a> Sequence alignment

b> Coordinate assignment

c> Prediction of protein structure

d> Physico-chemical interactions in proteins in aqueous media

6. Protein conformation **6 hrs.**

a> Active forms & Zymogens

b> Denaturation and renaturation, effect of salts, solvents & temperature.

c> Protein misfolding and protein aggregation

7. Experimental techniques for proteins **4 hrs.**

a> Basic principles of amino acid sequencing

b> Protein synthesis

c> Isolation, purification, characterization & estimation

8. Introductory Proteomics **2 hrs.**

Proteomics : Comparative genome analysis using proteomics –
Differential expression on 2D- gel

9. Protein as drug targets **6 hrs.**

Some case studies

10. Structure and function of Nucleic acid, lipids & carbohydrates **10 hrs.**

Introduction to Nucleic acids, lipids & carbohydrates.

RECOMMENDED BOOKS

MAIN READING

1. C, Branden & C. Tooze, “Introduction to Protein Structure”, Garland Publishing Inc., New York, 1991.
2. T.E. Creighton, “Proteins – Structures and Molecular Properties”, W.H. Freeman & Co., New York, 1993.

SUPPLIMENTARY BOOKS

1. EE Conn & PK Stumpf, “Outlines of Biochemistry”, Wiley Eastern.
2. R.R. Sinden, “DNA Structure and Function”, Academic Press, San Diego, 1994.
3. Bioinformatics Sequence, Structure and Databanks Edited by D. Hidggins & W. Taylor, Oxford University Press, 2000.

BI-A10/B2.5-R0 : BASICS OF GENOMICS AND PROTEOMICS

Objective of the Course

The primary objective of this course is to impart advanced knowledge to students on a self learning mode such that students become familiar with the methods of searching for information and analyzing preliminary data.

After completion of this course the student should be able to

- Understand state-of-the-art molecular biology techniques
- Be conceptually familiar with metabolism
- Interpret preliminary gene expression and protein expression studies
- Be familiar with the various genome projects completed and ongoing

Outline of Course

S. No.	Topic	Minimum No. of hrs.
1.	Advanced Biochemistry	10
2.	Metabolism and Pathways	8
3.	Basic concepts in Genomics	10
4.	Micro array for gene expression	10
5.	Proteomics	10
6.	Genome Projects – The unfolding story	12

Lectures = 60
Practicals / Tutorials = 60
Total = 120

Detailed Syllabus

1. Advanced Biochemistry 10 hrs

An introduction to physical biochemistry, intermediary metabolism and molecular biology. Topics include a survey of structure, chemistry and function of proteins and nucleic acids; regulation of gene expression at the level of DNA, RNA, and protein synthesis.

2. Metabolism and Pathways 8 hrs.

Pathways of carbohydrate, lipid and nitrogen metabolism and their metabolic control.

3. Basic concepts in Genomics 10 hrs.

Whole genome analysis, Genome sequencing technology.

Comparative genomics – Paralogs and orthologs, Phylogeny, Human genetic disorders, Candidate gene identification, Concepts of Pharmacogenomics.

4. **Micro array for gene expression** **10 hrs.**

Target selection, customized microarray design, image processing and quantification, normalization and filtering, statistical analysis, public microarray data sources.

5. **Proteomics** **10 hrs.**

Basics of Protein structure, Introduction to basic Proteomics technology, Bio-informatics in Proteomics, Basics of Proteome Analysis, Concepts in Enzyme Catalysis.

6. **Genome Project – The unfolding story** **12 hrs.**

Introduction to the concepts cloning and mapping, Construction of Physical maps, Basics of radiation hybrid maps, Sequencing : Related discoveries and technology development, Implications of the Human Genome Project, Basic Human Inheritance Patterns, Basics of Single Nucleotide Polymorphism detection and its implication, Practical Application of medical Genetics Technology.

RECOMMENDED BOOKS

MAIN READING

1. Genomes. Author : T.A. Brown, John Wiley and Sons, Bios Scientific Publishers
2. Discovering Genomics, Proteomics and Bioinformatics Campbell AM and Heyer LJ Perason Education (Low priced Editions) 2003.

SUPPLIMENTARY BOOKS

1. Principles of Genome Analysis and Genomics (Third Edition) Primrose and Twyman Blackwell publishing (2003)
2. Reiner Westermeier and Tom Naven. Proteomics in practice. Wiley-vch. 3rd Edition 2002
3. Instant Notes-Bioinformatics Authors : D.R. Westheadn, J.H. Parish and R.M.Twyman. Publisher : Viva Books Pvt. Ltd. IndianEdition 2003.

B3.1-R0 : COMPUTER ORGANIZATION AND DISTRIBUTED SYSTEMS

Objective of the Course

Objective of the course is to familiarize students about computer hardware design and distributed systems, basic structure and behavior of the various functional modules of the computer and how they interact to provide the processing needs of the user.

This subject mainly focuses on the computer fundamentals, network and distributed systems, architectural models, fundamental models, networking and internetworking, operating systems & its supports for distributed systems. Students will be exposed to internet principles, different types of networking topologies. OSI (Open System Interconnection) seven layer model. The course also deals in memory management, methodologies for enhancement of processor throughout besides process control and interprocess communication.

Outline of Course

S. No.	Topic	Minimum No. of hrs.
1.	Fundamentals	11
2.	Basic Computer Organization	12
3.	Introduction to Networks and Distributed Systems	5
4.	Distributed Systems Models	6
5.	Networking and Internetworking	6
6.	Interprocess Communication	6
7.	Operating Systems & its supports for DS	14

Lectures = 60
Practicals / Tutorials = 60
Total = 120

Detailed Syllabus

1. Fundamentals 11 hrs.

Binary number systems, Two's complement, floating point representation, addition, subtraction,overflow.

Logic gates, flip-flops, encoders, decoders, multiplexers, registers, counters, RAM, ROM.

2. Basic Computer Organization 12 hrs.

Bus and Memory transfers, Binary adder, counter, arithmetic circuit, Logical and shift operations Instruction codes, direct and indirect addressing, Instruction cycle, Instruction formats such as one address and two address instructions, memory reference instructions, input output instructions, arithmetic, logical and shift instructions, general register organization, Memory stack, brief introduction to interrupts.

3. Introduction to Networks and Distributed Systems 5 hrs.

The motivation and goals of network and distributed system, network topologies, layers in a network, OSI (Open Systems interconnection) seven layer model (introduction, e.g., Chapter 1 of Tanenbaum), characteristics of a distributed computing system, issues such as heterogeneity, openness, security, scalability, concurrency.

4. Distributed Systems Models 6 hrs.

Architectural Models : client server model, peer processes model, mobile agents, thin clients, spontaneous networking, issues in designing distributed systems,

Fundamental models : interaction model, failure model, security model and related issues.

5. Networking and Internetworking 6 hrs.

Networking issues, types of networks – local area networks, wide area networks, metropolitan area networks, wireless network and internetwork. Networks principles – protocols, routing, packet transmission, data streaming, switching schemes. TCP/IP protocols – IP addressing, IP routing, mobile IP.

6. Interprocess Communication 6 hrs.

Characteristics of interproces communication, sockets, UDP datagram communication, TCP stream communication, external data representation and marshalling. Client-server communication – the request-reply protocol and related issues, use of TCP streams to realize request-reply protocol. Group communication, use of IP-multicast to implement group communication.

7. Operating Systems & its supports for DS 14 hrs.

Tasks of OS, types of OS, OS structure.

Programs and process, process control, interacting processes, synchronization, interprocess communication,

Fork-Join primitives, parbegin-Parend control structure, Threads and their creation and termination, kernel levels threads, user level threads. – (chapters 9 and 10 of Dhamdhere).

Memory management – Memory allocation model, contiguous memory allocation and related issues, non-contiguous memory allocation – virtual memory management through demand paging – (15 of Dhamdhere).

Enhancement of throughput by multi-threading, architectures for multi-threaded servers, threads versus multiple processes, OS supports for communication and invocation, concurrent invocation, asynchronous invocation, desirable architecture of OS for distributed computing – (chapter 6 of Coulouris, Dollimore & Kindberg).

RECOMMENDED BOOKS

MAIN READING

1. George Coluouris Jean Dollimore Tim Kindberg, *Distributed Systems : Concepts and Design*, Pearson Education, 2001.

2. M.M. Mano *: Computer Systems Architecture*, 3rd Edition, Pearson Education, 2000.

3. D.M. Dhamdhere, *Systems Programming and Operating Systems*, Tata McGraw Hill, 1996.
4. A. Tannenbaum & Van Steen, *Distributed Systems Principles and Paradigms*, 1st Edition, 2002 Prentice Hall

SUPPLEMENTARY READING

1. H.F. Korth and A. Siberschatz : Database System Concepts, McGraw Hill.
2. J.P. Hayes : Computer Architecture and Organization, McGraw-Hill.
3. John D. Carpinelli, Computer Systems Organization & Architecture Pearson Low priced Edition 2004.

BI-B3.2-R0 : PROBABILITY AND INFORMATION THEORY

Objective of the Course

Biological processes are inherently random. A rational description would require study of probabilistic framework. Analysis of vast amount of biological data sets would require a through understanding of probability theory. For quantifying uncertainty embedded in a biological system, tools from information theory are useful. For simulating biological phenomena, Monte Carlo methods are generally employed. This course will provide foundational support to the field of Bioinformatics.

Outline of Course

S. No.	Topic	Minimum No. of hrs.
1.	Introduction to Probability Theory	4
2.	Random Variables	12
3.	Conditional Probability and Expectation	4
4.	Limit Theorems	4
5.	Elements of Stochastic process and Markov Chains	10
6.	Information Theory	8
7.	Information source :Memoryless, and with memory	8
8.	Monte Carlo methods	10

Lectures = 60
Practicals / Tutorials = 60
Total = 120

Detailed Syllabus

1. Introduction to Probability Theory 4 hrs.

Probability defined on events, Conditional Probability, Bayes' theorem

2. 12 hrs.

a> Random Variables

Discrete Random variables : - Bernoulli, Binomial, Geometric and Poisson distributions.

Continuous Random variables : - Uniform, Exponential, Normal, Gamma, and Chi-square distributions.

Expectation of random variables.

Chebyshev's inequality.

Jointly distributed random variables.

b> Functions of Random Variables

Minimum, Maximum sum of Random Variables

3. Conditional Probability and Expectation 4 hrs.

Discrete case, Continuous case, Computing Expectation by conditioning, Computing probabilities by conditioning.

4. Limit Theorems **4 hrs.**

Weak law of large number.
The Central Limit Theorem.
The Strong Law of Large Numbers.

5. Elements of Stochastic processes and Markov Chains **10 hrs.**

Introduction to stochastic processes.
How to characterize a stochastic process?
Markov processes and Markov chains.
Chapman – Kolmogorov equation.
Classification of states.
Limiting probabilities.

6. Information Theory **8 hrs.**

What is Information?
Surprise and Uncertainty.
Shannon's Information measure.
Conditional, Joint and Mutual Information measure.
Kullback-Leibler Directed Divergnce.
The communication model

7. a) Discrete memory-less Information source **8 hrs.**

The discrete information source.
Source coding.

b) Discrete Information source with memory

Markov Chain.
State transition diagram.
Amount of information in a first order Markov chain.

8. Monte Carlo Methods **10 hrs.**

Random number generator.
Linear Congruential Method
Techniques for generating continuous random variables.
- Inverse Transformation Method

Simulating from discrete distributions.
Variance reduction techniques.
Hit or Miss Monte Carlo method
- Approximate evaluation of

Note : Students have to familiarize with at least one statistical software like R for Linux, SYSTAT/MINITAB

RECOMMENDED BOOKS

MAIN READING

1. S.M. Ross, “A First course to Probability”, 6th Edition, Pearson Education.
2. J.L. Devore, “Probability and Statistics”,Thomson, 2002.
3. J.C.A. Vander Lubbe, ”Information Theory”, Cambridge University Press, 1997.

SUPPLEMENTARY READING

1. W.J. Ewens and G.R. Grant, “Statistical methods in Bioinformatics” : Anintroduction, Springer-Verlag 2001.
2. S.M. Ross, “Probability Models”, 7th Edition, Harcourt India Private Limited, 2001
3. D. Applebaum, “Probability and Information” – and integrated approach, Cambridge 1996.

BI-B3.3-R0 : COMPUTATIONAL METHODS IN BIOMOLECULAR SEQUENCE ANALYSIS

Objective of the Course

The aim of course is to introduce and impart skills to students about the computational tools and techniques which are needed to analyze Primary DNA and Protein sequences. This will enable them to understand the significance of Protein structure and function. Some of the major Computational issues and solutions will be covered with suitable examples. Focus will be on introducing few well known algorithms and methods which are widely used in analyzing sequence. Basic knowledge in Mathematics and Statistics is prerequisite. Students who undertake this course will get the necessary understanding, exposure and foundation to take up small or large scale software Development projects in the area of Bioinformatics.

After the end of this module the students can :

1. Use and apply state of the art tools pertaining to Sequence analysis
2. Will be able to analyze primary DNA and Protein sequences and interpret results
3. Will have the understanding of the methodology or algorithm employed in sequence analysis tools
4. Will be in a position to undertake software development project relate to sequence analysis

Outline of Course

S. No.	Topic	Minimum No. of hrs.
1.	Basic concepts of Molecular Biology	6
2.	Sequence composition and statistics	4
3.	Strings, Graphs and algorithms	6
4.	Pairwise and Multiple alignment	12
5.	Database homology Search	10
6.	Statistical Modeling of sequences	4
7.	Pattern recognition and motif analysis	6
8.	Secondary structure prediction	6
9.	Phylogenetic analysis	6

Lectures	**=**	**60**
Practicals / Tutorials	**=**	**60**
Total	**=**	**120**

Detailed Syllabus

Note : Appropriate References to Mainreading are suggested for teaching purpose.

1. Basic concepts of Molecular Biology [Refer MR-1] **6 hrs.**

- 1.1 Proteins (1 hour)
- 1.2 Nucleic Acids (1 hour)
  - 1.2.1 DNA
  - 1.2.2 RNA
- 1.3 Molecular Genetics (4 hours)
  - 1.3.1 Genes and Genetic code
  - 1.3.2 Transcription, Translation and Protein synthesis
  - 1.3.3 Chromosomes
  - 1.3.4 Prokaryotic and Eukaryotic Gene structure

2. Sequence composition and statistics [SR-1] **4 hrs.**

- 2.1 Nucleotide and amino acid compositions
- 2.2 Codon usage and statistics
- 2.3 AT and GC rich regions

3. Strings, Graphs and Algorithms [Refer MR-1] **6 hrs.**

- 3.1 Strings (1 hour)
- 3.2 Graphs (2 hours)
- 3.3 Algorithms (3 hours)

4. Pairwise and Multiple alignment [Refer MR-1, 2 & 3 **12 hrs.**

- 4.1 Pairwise alignment (6 hours)
  - 4.1.1 Global alignment
  - 4.1.2 Local alignment
  - 4.1.3 Scoring functions
  - 4.1.4 General gap and affine gap penalty
  - 4.1.5 Statistical significance
- 4.2 Multiple Sequence alignment (6 hours)
  - 4.2.1 SP (Sum of Pairs) measure
  - 4.2.2 Star alignments
  - 4.2.3 Tree alignments
  - 4.2.4 Motifs and Profile
  - 4.2.5 Alignment Representation and Applications
  - 4.2.6 ClustalW, clustalX, Hmmer and T-cofee

5. Database homology search [Refer MR2 & SR-1, 4] 10 hrs.

5.1 Scoring matrices
5.2 BLAST algorithm
5.3 Significance of alignments (6 hours)
5.4 Blast versions : Blastp, Blastx, tBlastn, tbalstx
5.5 PSI and PHI Blast

6. Statistical Modeling of Sequence [MR-2] 4 hrs.

6.1 Independent and indentically distributed sequences
6.2 Markov Chain models

7. Pattern Recognition and Motif analysis 6 hrs.

7.1 Restriction site
7.2 Finding repeats
7.3 General Pattern search
7.4 DNA and Protein motifs search
    Transfac
    Prosite

8. Secondary structure Analysis [MR-3] 6 hrs.

8.1 Stem and Loop structure
8.2 RNA fold algorithm

9. Phylogenetic analysis [MR-1 and MR-3] 6 hrs.

9.1 Distances and Parsimony methods
9.2 Clustering methods
9.3 Rooted and Unrooted tree
9.4 Bootstrap analysis
9.5 PHYLIP

RECOMMENDED BOOKS

MAIN READING

1. Setubal Joao and Meidanis Joao, “Introduction to Computational Molecular Biology”, PWS Publishing Company (An International Thomson Publishing Company), 1997, Indian low priced edition.

2. Warren Ewens and Gregory R. Grant, “Statistical Methods in Bioinformatics : An Introduction” Springer-Gerlag 2001.

3. R. Durbin, S. Eddy, A. Krogh and G. Mitchison, “Biological sequence analysis : Probablistic models of proteins and nucleic acids”, Cambridge University Press, 1998 Indian low priced edition.

SUPPLEMENTARY READING

1. Mount “Bioinformatics” Cold Spring Harbor Laboratory.

2. Stephen A. Krawtz and David D. Womble “Introduction to Bioinformatics : a theoretical and practical approach. Humana Press.

3. Neil C. Jones and Pavel A Pevzner. An introduction to Bioinformatics algorithms MIT Press reprinted by Ane Books, New Delhi 2005.

4. Ian korf, Mark yandell & Joseph Bedell. “BLAST”. O’Reilly press 2003.

BI-B3.4-R0 : DISCRETE MATHEMATICS

Objective of the Course

The objective is to provide the necessary mathematical knowledge so that students can understand computational issues relating to practical problems. The paper intends to equip the students with the necessary mathematical ideas, notations and techniques that are required to formulate an algorithm for a problem, and to reason about its time and pace complexity. Many problems of bioinformatics are combinatorial in nature. This course will help to solve such problems also.

Outline of Course

S. No.	Topic	Minimum No. of hrs.
1.	Sets, graphs and trees	13
2.	Counting and probability	9
3.	Recurrences and generating functions	10
4.	Boolean algebra	5
5.	Graph algorithms	9
6.	Modeling computation : grammer, language and automata	9
7.	NP-completeness	5

Lectures	**=**	**60**
Practicals / Tutorials	**=**	**60**
Total	**=**	**120**

Detailed Syllabus

1. Sets, graphs and trees 13 hrs.

Sets, operations on sets, principle of inclusion and exclusion, countable and uncountable sets, Cartesian product, n-ary relation, binary relation, operations on relations, reflexivity, symmetry, transitivity, equivalence relation, equivalence class, partition, partial order, composite of relations, closures of relations, Warshall's algorithm for transitive closure.

Functions, domain, range, injection, surjection, bijection, composition of functions,growth of functions – asymptotic q-notation, O-notation, W-notation, o-notation, w-notation and their properties.

Simple graphs, multigraphs, directed graphs, undirected graphs, degree, reachability, paths, cycles, isomorphism, bipartite graphs, trees, rooted trees, forest, k-ary trees, basic properties of graphs and trees.

2. Counting and probability 9 hrs.

Summation and product rules, pigeonhole principle, permutation, combination, binomial coefficients and related identities, sample space, discrete probability distribution, conditional probability, independence, Baye's theorem, Bernoulli's trials, binomial distribution, simple probabilistic reasoning such as birthday paradox.

3. Recurrences and generating functions 10 hours

Linear homogeneous recurrence relations, linear non-homogeneous recurrence relations, solving recurrences by direct method, iteration method, master method, recurrence trees, generating functions for a sequence, solving simple counting problems using generating functions, solving recurrence relations using generating functions.

4. Boolean algebra 5 hours

Boolean expression and function, formal definition of Boolean algebra, identities, combinatorial circuits, Karnaugh maps.

5. Graph algorithms 9 hours

Graph representation scheme, breadth first search, depth first search and their properties, decomposing a directed graph into its strongly connected components, Euler tour, minimum spanning tree, Prim's algorithm, Kruskal's algorithm, shortest path problem, Dijkstra's algorithm and Bellman –Ford Algorithm for single source shortest paths problem.

6. Modeling computation: grammar, language and automata 9 hours

Terminals, non-terminals, grammar, language, types of grammar – context sensitive and context free grammar, parse tree, finite state automata, non-deterministic finite state automata, regular expression, regular sets, Kleene's theorem Turing machine.

7. NP-completeness 5 hours

Introduction to complexity class P, verification algorithm, complexity class NP, reduction function, reduction algorithm, NP-completeness

RECOMMENDED BOOKS

MAIN READING

1. Introduction of Algorithms, T. Coreman, C. E. Leiserson, and R. L. Rivest
2. Discrete Mathematics and its Applications, Kenneth H. Rosen, McGraw Hill Companies, 4th Edition

BI-B4.1-R0: STATISTICAL METHODS IN BIOINFORMATICS

Objective of the Course

Statistics as a discipline deals with collection and analysis of data. The collection of data involves techniques of sampling methods while the analysis involves inferences and estimation based on sampled data. In Bioinformatics the same principle regarding scientific data collection and analysis of data hold. To investigate, large scale data sets generated from various genomes, proteome, metabolomic projects and related sources, it is necessary to have a proper grounding in statistical techniques.

Outline of Course

S. No.	Topic	Minimum No. of Hours
1.	Review	04
2.	Sampling Distributions	05
3.	Estimation Theory	08
4.	Maximum Likelihood Esimation	04
5.	Inference – Tests of Hypotheses	16
6.	Simple Linear Regression and Correlation	06
7.	Analysis of Variance	06
8.	Statistical methods for Stochastic Processes	05
9.	Statistical Analysis of Stochastic Processes	06

Lectures = 60
Practical / Tutorials = 60
Total = 120

Detailed Syllabus

1. Review — 4 hours

Mean, Median, Mode
Standard Deviation, Variance and Correlation
(Emphasis to be placed on hands on approach with real data sets)
Probability limits and distribution theorems

2. Sampling Distributions — 5 hours

Statistic
Distribution of sample mean
Sample variance – (application of Central limit theorem)

3. Estimation Theory — 8 hours

Biased and unbiased estimator
Confidence interval: population mean, proportion
Variance

4. Maximum Likelihood Estimation **4 hours**

Discrete and Continuous distributions
Likelihood function
Log-likelihood functions
(use of package recommended)

5. Inference – Test of hypotheses **16 hours**

a) Formulation of Hypothesis ….**4 hours**
   Simple and Composite
   Type I and Type II errors
   Power of a test
   Significance of a test
   P-value
b) Testing **….6 hours**
   Normal, Chi-Square, t-test and F-test
c) Chi-Square goodness of fit **…..4 hours**
   Test of diversity based on entropic method
d) Non-parametric: Mann-Whitney test **….. 2 hours**

6. Simple Linear Regression and Correlation **6 hours**

Linear regression model
Least squares methods
Estimating model parameters
Residual sum of squares

7. Analysis of Variance **6 hours**

Single factor ANOVA
Multi-comparison ANOVA

8. Statistical Methods for Stochastic Processes **5 hours**

Testing of independence
Testing of Markov property
Association test

9. Analysis of Stochastic process **6 hours**

Long Repeat
Scan Statistics
Analysis of Patterns

RECOMMENDED BOOKS

MAIN READING

1. W. J. Ewens and G. R. Grant, “Statistical Methods in Bioinformatics”: An introduction Springer-Verlag 2001
2. J. L. Devore, “Probability and Statistics”, (Fifth Edition), Thomson Asia, 2002
3. S. M. Ross, “A First course to Probability”, 6th Edition, Pearson Education

BI-B4.2-R0: BIOMOLECULAR STRUCTURE & DYNAMICS

Objective of the Course

This course is formulated keeping in view first year Master's students with 'A' level (DOEACC) background in biology. The focus is on imparting a sound knowledge of structure and dynamics of proteins and nucleic acids and modelling them in silico. After attending the course, the students are expected to handle research projects in biomolecular structure and function.

Outline of Course

S. No.	Topic	Minimum No. of Hours
1.	Protein Structure, Classification of protein Structure, Membrane proteins, Viral Proteins, Proteins in immune system, Proteins as drug targets.	10
2.	Manipulating protein structures on computer, Cartesian and internal coordinate representations , ribbon diagrams, space filling models, topology diagram, searching for sequences & folds	10
3.	Nucleic acid structures – A review	10
4.	Database analyses of internal degrees of freedom of nucleic acid structures	5
5.	Forces stabilizing Protein & Nucleic Acid structures	5
6.	Energy minimization and Molecular Dynamics methods	15
7.	Structure and function of Carbohydrates and Lipids.	5

Lectures = 60
Practical / Tutorials = 60
Total = 120

Detailed Syllabus

1. a) Protein structure – A review of primary, secondary, tertiary and quaternary structures, motifs and folds.
   b) Classification of protein structures
   c) Membrane proteins
   d) Viral proteins
   e) Proteins in immune system
   f) Proteins as drug targets **10 hours**

2. Manipulating protein structures on computer, Cartesian and internal coordinate representations, ribbon diagrams, space filling models, topology diagrams, searching for sequences & folds.
10 hours

3. Nucleic acid structures – A review

   a) Right handed (A, B) and left handed (Z) double helical DNA.
   b) Structures of t-RNA and r-RNA
   c) DNA bending
   d) DNA supercoiling
   e) Cruciforms, triplexes and quadruplexes
   f) Nucleic acids as drug targets **10 hours**

4. Database analyses of internal degrees of freedom of nucleic acid structures **5 hours**

5. Forces stabilizing Protein & Nucleic Acid structures **5 hours**

6. Energy minimization and Molecular Dynamics methods **5 hours**
   a) Theory
   b) Application to Proteins **5 hours**
   c) Applications to nucleic acids **5 hours**

7. Structure and function of Carbohydrates and Lipids **5 hours**

RECOMMENDED BOOKS

MAIN READING

1. C. Branden & C. Tooze, “Introduction to Protein Structure”, Garland Publishing Inc. New York, 1991
2. R. R. Sinden, “DNA Structure & Function”, Academic Press, San Diego, 1994

SUPPLEMENTARY READING

1. Lehninger principles of Biochemistry, David Nelson and Michael M. Cox, Freeman Publishing, 4th Edition 2004
2. A. R. Leach, “Molecular Modelling – Principles & Applications” Addison Wesley Longman, Essex, 1996

BI-B4.3-R0: DATA STRUCTURES AND ALGORITHMS

Preamble

“Data Structures and Algorithms” can be considered as a classic, core topic of computer science. Data structures and algorithms are central to the development of good quality programs. Their study is very critical for a sound understanding and efficient implementation of Bioinformatics algorithms and programs.

Objective of the Course

(1) To gain a sound understanding of different types of data structures and their role in algorithm design in computer science (in general) and in bioinformatics (in particular).

(2) To be able to understand any typical algorithm design problem in computer science and bioinformatics and be able to visualize appropriate data structures for efficiently implementing the algorithm.

(3) To be able to prove why a particular data structure is best suited for a particular problem, by analyzing the algorithm in a sound and systematic way.

Outline of course

S. No.	Topic	Minimum No. of Hours
1.	Introduction	2
2.	Computational Complexity	6
3.	Essential Preliminaries	5
4.	Dictionaries	12
5.	Priority Queues	4
6.	Graphs	12
7.	Sorting Methods	6
8.	String Matching	4
9.	Algorithm Design Paradigms	9

Lectures = 60
Practical / Tutorials = 60
Total = 120

Detailed Syllabus

1. INTRODUCTION 2 hours

Abstract data type and data structures, Classes and objects

2. COMPUTATIONAL COMPLEXITY 6 hours

Complexity of algorithms: worst case, average case and amortized complexity, notation for algorithm complexity, Algorithm analysis, Recurrence relations, Introduction to NP-completeness

3. **ESSENTIAL PRELIMINARIES** **5 hours**

Overview of recursion, iteration, arrays, pointers, lists, stacks, queues

4. **DICTIONARIES** **12 hours**

Dictionaries: Hash tables, Binary search trees, splay trees, Balanced Trees, AVL trees, 2-3 trees, B-Trees

5. **PRIORITY QUEUES** **4 hours**

Priority Queues : Heaps, binomial queues, Applications

6. **GRAPHS** **12 hours**

Graphs: Shortest path algorithms (Dijkstra, Bellman-Ford, Floyd-Warshall), minimal spanning tree algorithms (Prim, Kruskall), depth-first, breadth-first search, and their applications.

7. **SORTING METHODS** **6 hours**

Sorting: sorting methods and their analysis (shell sort, quicksort, merge sort, heap sort, radix sort), lower bound on complexity

8. **String Matching** **4 hours**

String matching algorithms (Rabin-Karp, Knuth-Morris-Pratt)

9. **ALGORITHM DESIGN PARADIGMS** **9 hours**

Algorithm Design Paradigms: Greedy methods, divide and conquer, dynamic programming, backtracking, local search methods, branch and bound technique. Travelling salesman problem.

RECOMMENDED BOOKS

MAIN READING

1. A. V. Aho, J. E. Hopcroft, and J. D. Ullman, Data Structures and Algorithms, Addison Wesley, Reading Massachusetts, USA, 1983 (Available in Indian Edition)

2. M. A. Weiss, Data Structures and Algorithms Analysis in C++, Benjamin / Cummins, Redwood City, California, USA, 1994 (Available in Indian Edition)

SUPPLEMENTARY BOOKS

1. T. H. Cormen, C. E. Leiserson, and R. L. Rivest, Introduction to Algorithms, the MIT Press, Cambridge, Massachusetts, USA, 1990 (Available in Indian Edition)

2. Y. Narahari, Web-based Lecture Notes on Data Structures and Algorithms. http://www.csa.iisc.ernet.in/~hari

BI-B4.4-R0: COMPUTATIONAL GENOMICS

Objective of the course

Computational approaches are adopted right from Genome sequencing to sequence characterization. This course module will provide the necessary basics and overview of the computational issues related to Genomics. Application of *in silico* methods to some of the complex biological problems such as Eukaryotic gene finding will be highlighted. Relevant and popular software tools and techniques employed in this domain will be covered with suitable examples. Students at the end of the course will be able to undertake pilot software project in related topics.

Outline of Course

S. No.	Topic	Minimum No. of Hours
1.	Prokaryotic and Eukaryotic Genomes	6
2.	Sequencing methods and technologies	4
3.	Computational methods in sequencing	8
4.	Computational approaches for Gene Identification	6
5.	Computational methods for gene prediction	12
6.	Analysis of regulatory regions	6
7.	Genome Characterization	6
8.	Comparative genomes	6
9.	Overview of Functional Genomics	6

Lectures = 60
Practical / Tutorials = 60
Total = 120

Detailed Syllabus

1. Prokaryotic and Eukaryotic Genomes 6 hours

Central Dogma of molecular biology
Prokaryotic Gene structure
Eukaryotic Gene structure
Genome Organization
Genome Databases

2. Sequencing methods and technologies 4 hours

Short-gun sequencing
Clone by Clone approach
Genome markers and maps

3. Computational methods in sequencing 8 hours

Sequencing errors
Base calling
Fragment assembly
Overlap and layout consensus

4.	**Computational approaches for Gene identification**	**6 hours**

Signal based methods
Content based methods
Homology based methods
Performance measures

5.	**Computational methods for gene prediction**	**12 hours**

Statistical methods
Compositional bias
Non-randomness
Machine learning and probabilistic methods
Artificial Neural Network
Markov Chain
Hidden Markov Model

6.	**Analysis of regulatory regions**	**6 hours**

Signals and Patterns
Promoter prediction
Transcription factor binding sites

7.	**Genome Characterization**	**6 hours**

EST comparison
Blast analysis

8.	**Comparative genomes**	**6 hours**

Gene order
Horizontal gene transfer
Transposable elements
Clusters of orthologus genes

9.	**Overview of Functional Genomics**	**6 hours**

Repeats and diseases
Gene expression
Applications

RECOMMENDED BOOKS

MAIN READING

1. Mount (2001), “Bioinformatics : Sequence and Genome analysis”, Cold Spring Harbor Laboratory press
2. Michael Waterman (1995), “Introduction to Computational Biology: Maps, Sequences and Genomes”
3. Warren Ewens and Gregory R. Grant, “Statistical Methods in Bioinformatics : An introduction”, Springer-Verlag 2001
4. Cecillia saccone and Grazianio pesole, Handbook of comparative Genomics. Principles and methodology

SUPPLEMENTARY READING

1. R. Durbin, S. Eddy, A Krogh and G. Mitchison, “Biological sequence analysis: Probabilistic models of proteins and nucleic acids”, Cambridge University Press, 1998
2. Thomas Lengauer (editor)”Bioinformatics – from Genomes to drugs”, Volume 1 “Basic Technologies” Wiley – Vch Germany
3. Tao Jiang, Ying Xu and Michael Q. Zhnag (Editors) Current topics in computational Molecular biology. Ane Books, New Delhi 2004 .
4. Pierre Baldi and Soren Brunak (2003) “Bioinformatics: The machine learning approach”, East West Press, Low priced Edition, Delhi

BI-B5.1-R0 ; OPTIMIZATION, MACHINE LEARNING AND COMPUTATIONAL INTELLIGENCE

Objective of the course
BI-B5.1-R0 : OPTIMIZATION, MACHINE LEARNING AND COMPUTATIONAL INTELLIGENCE

BI-B5.1-R0 ; OPTIMIZATION, MACHINE LEARNING AND COMPUTATIONAL INTELLIGENCE

Objective of the Course

To provide the basic concepts and terminology of machine learning and to give an overview on the main approaches to machine learning along with an understanding of the differences, advantages and problems of the various approaches. On completion of the course students should be able to understand and apply various clustering and classification algorithms in classical and computational intelligence framework for bioinformatic problems. The intention is to provide the necessary theoretical background also so that they can avoid blind use of such tools for practical problems. It is a self-contained course which will also equip the students with the necessary knowledge of optimization theory.

Outline of Course

S. No.	Topic	Minimum No. of Hours
1.	Optimization techniques	12
2.	Genetic algorithms	7
3.	Basic concepts of machine learning	5
4.	Clustering and classification	14
5.	Introduction to neural networks	3
6.	Feed forward neural networks	12
7.	Self-organizing maps and recurrent neural networks	7

Lectures = 60
Practicals/Tutorials = 60
Total = 120

Detailed Syllabus

1. Optimization techniques 12 hours

Concepts of local and global optimum, constrained and unconstrained optimization, gradient, Hessian of a function, first order and second order necessary conditions for local minimizers, method of steepest descent, steepest descent for quadratic function, Newton’s method, Levenberg-Marquardt algorithm, and Lagrange multipliers methods.

2. Genetic algorithms 7 hours

Introduction to Genetic algorithms, exploitation and exploration, general structure of the basic genetic algorithm, different coding schemes, fitness function, chromosome, population, basic genetic operators, different selection mechanism, mutation, single point and multi-point crossover, termination criteria, elitist strategy.

3. **Basic concepts of machine learning** **5 hours**

Basic notions of learning,, introduction to learning algorithms, incremental learning, supervised learning, unsupervised learning, reinforcement learning, instance based learning and analytical learning.

4. **Clustering and classification** **14 hours**

Introduction to clustering, maximum likelihood decomposition and its application to normal mixtures, k-means clustering algorithms, hierarchical clustering algorithms.

Introduction to classifier design, linear discriminant analysis – two category and multi category cases, perceptron criterion and its minimization, decision trees (ID3, C4.5), impurity functions, pruning methods, rule extraction from decision trees, nearest neighbour classifier, k-nearest neighbour classifier, Bayes decision rule, lose function, minimum error rate classification, Bayes classifier with multivariate normal density.

5. **Introduction to neural networks** **3 hours**

Introduction to neural networks, introduction to biological neural network, motivation for artificial neural network (ANN), significance of massive parallelism and characteristics of (ANN), various types of architectures.

6. **Feed forward neural network** **12 hours**

Layered networks, perceptron and motivation for multilayered perceptron (MLP), back-propagation learning , use of Levenberg-Marquardt method in training of MLP's, online vas batch learning, issues relating to initialization, termination, choice of architecture. Radial basis function networks and related training issues, properties of MLP and RBF.

7. **Self-organizing maps and Recurrent neural networks** **7 hours**

Self-organizing feature map, recurrent neural networks, Hopfield network and its applications, simulated annealing.

RECOMMENDED BOOKS

MAIN READING

1. R. O. Duda, P. E. Hart and D. G. Stork, Pattern classification, John Wiley Sons, Second Edition, 2000

2. D. E. Goldberg, Genetic algorithms in search, optimization and machine learning, Pearson Education (Paper back)

3. Christopher M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995

4. E. K. P. Chong and S. H. Zak, An introduction to optimization, John Wiley & Sons, 2001

5. Tom Mitchell, Machine Learning, McGraw Hill, 1997

ADDITIONAL BOOKS

1. J. T. Touand R. C. Ganzales : Pattern Recognition Principles, Addison-Wesley, 1977, W. I. Zangwill: Non-linear Programming, Prentice Hall, 1969
2. P. E. Gill, W. Murray and M. H. Wright, Practical optimization, Academic Press, 1982
3. S. Haykin, Neural networks – a comprehensive foundation, Prentice Hall, 1998 Vojislav Kecman, Learning and soft computing, Pearson Low priced Edition 2004

BI-B5.2-R0 : OBJECT TECHNOLOGY FOR BIOINFORMATICS

Preamble

Object technology has emerged as a sound paradigm for modelling, analysis, design, and implementation of complex software intensive systems. Object technology is indeed an attractive methodology to use in Bioinformatics in the areas of software modelling, software design and database design. Emerging software technologies such as component technology and distributed object technology, which are quire important for Bioinformatics are all founded on the object oriented paradigm. What "Data Structures and Algorithms" is to Bioinformatics programming, "Object Oriented Modelling and Design" is to Bioinformatics software development.

Objective of the course

The objective of the course is to provide a sound technical exposure to the concepts, principles, methods, and best practices in object oriented modelling and design as applied to Bioinformatics.

The version of the course is to produce "Object Technologists" with sound knowledge and superior competence in building robust, scalable, and reliable software intensive systems in Bioinformatics. They would have a clear appreciation of the role of abstraction, modelling and design patterns in the development of a Bioinformatics product. They would be able to make optimal design choices and employ the most relevant methods, best practices, and technologies for building a software intensive bioinformatics product, regardless of its complexity and scale.

Outline of Course

S. No.	Topic	Minimum No. of Hours
1.	Introduction to Object Technology	10
2.	Unified Modelling Language (UML)	10
3.	Object Oriented Modelling	10
4.	Object Oriented Design	10
5.	Distributed Objects and Component Technology	10
6.	Object Oriented Databases	10

Lectures = 60
Practical / Tutorials = 60
Total = 120

Detailed Syllabus

1. INTRODUCTION TO OBJECT TECHNOLOGY 10 hours

Meaning of the following terms: model, notation, method, methodology, analysis, design, process with examples, Evolution of object technology, difference between structured methodology and OO methodology

Object characteristics: abstraction, encapsulation, modularity, hierarchy (aggregation, composition, inheritance), polymorphism, concurrency, typing, persistence. Software development process, Rational unified process.

2. UNIFIED MODELING LANGUAGE (UML) 10 hours

Need for UML. UML Diagrams: Class diagrams, object diagrams, use case diagrams, component diagrams, deployment diagrams, state chart diagrams, activity diagrams, sequence diagrams, collaboration diagrams, other essential UML features. Examples and case studies

3. OBJECT ORIENTED MODELIGN 10 hours

Use case analysis, Analysis level UML diagrams, Analysis best practices, Role of aggregation, composition, association, inheritance, delegation, interaces, Examples and case studies.

4. OBJECT ORIENTED DESIGN 10 hours

Design best practices, Design patterns, Creational patterns, Structural patterns, Behavioural patterns, Role of patterns in object oriented design, Examples and case studies

5. DISTRIBUTED OBJECTS AND COMPONENT TECHNOLOGY 10 hours

Distributed object computing, Concept of object request broker, Interface definition language (IDL). Common object request broker architecture (CORBA). Component technology, Reusing components, Bioinformatics components, Object oriented frameworks.

6. OBJECT ORIENTED DATABASES 10 hours

Limitations of flat files, Advantages and disadvantages of relational databases, Need for object oriented databases, Object relational databases, Mapping from objects to relational or object relational databases, Examples and case studies in bioinformatics.

RECOMMENDED BOOKS

MAIN READING

1. J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy, and W. Lorenson, Object-oriented Modelling and Design, PHI, EEE, 1997 (Available in Indian Edition)
2. G. Booch, J. Rumbaugh, and I. Jacobson, I. The Unified Modeling Language User guide. Addison-Wesley 1999 (Available in Indian Edition)

SUPPLEMENTARY READING

1. E.Gamma, R Helm, R. Johnson, and J. Vlissides, Design Patterns : Elements of Reusable Object Oriented Software. Addison-Wesley, 1995 (Available in Indian Edition)

2. G. Booch, Object Oriented Analysis and Design with Applications, Second Edition, Benjamin Cummings, 1994 (Available in Indian Edition)

3. Jim Conallen, Building Web Applications with UML. Addison-Wesley, 2000 (available in Indian Edition)

BI-B5.3-R0 : COMPUTATIONAL PROTEOMICS AND GENE EXPRESSION STUDIES

Objective of the Course

On completion of this course, students will have been exposed to computational problems which arise in the analysis for data from high throughput proteomic and genomic platforms.

.

Outline of Course

S. No.	Topic	Minimum No. of Hours
1.	Introduction to Gel Electrophoresis	6
2.	Introduction to Mass Spectrometry	6
3.	Introduction to Microarrays	6
4.	Sequencing Hybridization and Mass Spectrometry	7
5.	Image Analysis Basis	15
6.	Statistics for Hypothesis Testing	6
7.	Clustering Algorithms	8
8.	Class Prediction Algorithms	6

Lectures = 60
Practicals / Tutorials = 60
Total = 120

Detailed Syllabus

1. Introduction to Gel Electrophoresis 6 hours

Physical Principles behind Electrophoresis
Application to Genomics and Proteomics
Overview of Computational Problems in Electrophoresis:
Lane Detection, Spot recognition, Image Matching

2. Introduction to Mass Spectrometry 6 hours

Physical Principles of Spectrometry
Applications to Genomics and Spectrometry
Overview of Computational Problems in spectrometry:
Peptide Sequencing, Peak Picking

3. Introduction to Microarrays 6 hours

Physical Principles of Microarrays
Applications to Genomics and Sequencing
Overview of Computational Problems in Microarrays:
Gridding, Spot Recognition and Segmentation, Hypothesis Testing
Clustering, Class Prediction

4. Sequencing by Hybridization and Mass Spectrometry 7 hours

Sequencing Algorithms

5.	**Image Analysis Basics**	**15 hours**

Working with Image Formats: TIFF, PGM, JPEG
Edge Detection
Segmentation
Line Detection
Image Alignment
Microarray Gridding

6.	**Statistical Hypothesis Testing**	**6 hours**

Normal Distribution
T and F Distributions
T-Test and ANOVA
Permutation Testing

7.	**Clustering Algorithms**	**8 hours**

Distance Metrics
K-Means, Hierarchical Clustering
Visualizing Clustering Results
Application of Clustering

8.	**Class prediction Algorithms**	**6 hours**

Decision Trees and Bayesian Methods
Introduction to Neural Networks

RECOMMENDED BOOKS

MAIN READING

1. Microarray Gene Expression Data Analysis: A Beginner's Guide, Helen. C Caustn Blackwell Publishers, 2003

2. Microarray Analysis, Mark Schena, Wiley-Liss, 2002

3. DNA Microarraya and Gene Expression: From Experiments to Data Analysis and Modeling. Pierre Baldi (Author), G. Wesley Hatfield (Author) and Wesley G. Hatfield, Cambridge University Press, 2002

4. Microarray Bioinformatics, Dov Stekel, Cambridge University Press, 2003

SUPPLEMENTARY READING

1. Microarrays for an integrative Genomics, Issac S. Kohare, Alvin T. Kho and Atul J. Butte, MIT Press reprinted by Ane Books, New Delhi 2004

2. DNA Array Image Analysis by Gerda Kamberova, Shishir Shah.

BI-B5.4.1-R0: COMPUTER AIDED MOLECULAR MODELLING AND DRUG DISCOVERY

Objective of the Course

The course is designed keeping in view an undergraduate student with plus two level knowledge in Physics, Chemistry and Mathematics, and elementary knowledge in Biology. The student also needs to have some exposure to use of computers. The objective of the course is to familiarize the student with basic concepts in analog and target structure based (rational) drug designing. The focus is on comparisons of structural and electronic parameters amongst different molecules and scoring methods to rank the ligand making use of “Principle of similarity” with other known ligand or “complimentary” with the active site of the target molecule. To achieve this, the student would be familiarized with building small molecules, computation of structure based and electronic parameters, docking a small molecule in the active cavity and setting up of QSAR (quantitative structure activity relationship) method.

Learning outcome

At the end of the course, the student will be equipped with different skills involved in computer aided drug designing. He would be able to do molecular modelling, quantum chemical calculations, compute electronic indices, identify active sites of target molecules and dock a ligand to the target molecule. He would also be able to optimise a ligand using different scoring methods.

Outline of Course

S. No.	Topic	Minimum No. of Hours
1.	Introduction to Drug Designing	5
2.	Small molecular structures	10
3.	Introduction to quantum chemical calculations	15
4.	Geometry optimization	10
5.	Basic concepts in quantitative structure activity relationship (OSAR)	5
6.	Basic principles of target structure based (rational) drug designing	15

Lectures = 60
Practicals / Tutorials = 60
Total = 120

Detailed Syllabus

1. Introduction to drug designing **5 hours**

a. Different approaches to drug designing
b. Basic principle of similarity and complimentarity
c. High throughput vs rational drug designing
d. Use of computer modelling technique to drug designing

2. Small molecular structures **10 hours**

a. Different coordinate systems and transformations amongst them
b. Basic Principle 2D and 3D Graphics and use of molecular graphics packages (e.g. RasMol, RasTop, Qmol, MolMol)
c. Cambridge Structural Data base
d. Building small molecules using chemical information

e. Use of Builders and Sketchers (eq. ISIS Draw, HyperChem)

3. Introduction to quantum chemical calculations 15 hours

a. Scrodinger equation, Variation method
b. Roothan's equation, Uniform & non uniform methods
c. Introduction to molecular orbital methods as : CNDO, INDO, MINDO, PCILO and *ab initio.*
d. Computation of HOMO (Highest occupied molecular orbital), LEMO (Lowest empty molecular orbital), Bond orders, Energies & Electronic indices using packages
e. Computation of Molecular Electrostatic Potential (MEP) maps using packages

4. Geometry optimization 10 hours

a. Use of classical (Empirical potential energy / function and grid search) methods
b. Use of Quantum mechanical approach
c. Use of molecular mechanics methods

5. Basic concepts in quantitative structure activity relationship (QSAR) 5 hours

a. Objective of QSAR
b. Development of Hansch QSAR equation
c. QSAR Descriptors
d. Regression analysis

6. Basic principles of target structure based (rational) drug designing 15 hours

a. Target identification and validation
b. Active site analysis
c. Basic principle of Docking a ligand in active site of a target
d. Builders & Growers
e. Different methods of scoring and Lead optimization

Methodology

The course would be taught through lectures, demonstration and tutorials

RECOMMENDED BOOKS

MAIN READING

1. Molecular Modeling: Basic principles and applications. Holtje HD, sippl W, Rognan D and Folkers G. Wiley-VFH 2nd Edition (2003)
2. Quantum Biology: S. P. Gupta, New Age publishers
3. Molecular modelling and drug design. Andrew Vinter and Mark Gardner and Boca Raton, CRC Press, 1994
4. Molecular Similarity in drug design, Dean PM, Chapman and Hall, 1995

SUPPLEMENTARY READING

1. Computer assisted Lead finding and optimization: current tools for medicinal chemistry. Han Van De Waterbeemd (Ed.) Wiley-VCH, 1997
2. Chemical Applications of Molecular Modelling, Goodman, JM Royal Society of Chemistry , 1998
3. Molecular Modelling: Principles and Applications by Andrew R. Leach, Longman, 1996
4. Computer Aided drug design: Methods and applications: Ed by Thomas J. Perun and CL Propst. Marcel Dekker, 1989
5. 3D QSAR in Drug Design Theory Methods and Applications – Hugo Kubinyi. Kluwer, 1993
6. Advances in Drug Discovery Techniques Alan L. Harvey (Ed.) John Wiley & Sons, 1998

BI-B5.4.2-R0: CHEMOINFORMATICS

Objective of the Course

The course aims to introduce the fundamentals of Chemoinformatics which are required in computer aided design. The students will also be exposed to issues related to chemical data processing

Outline of Course

S. No.	Topic	Minimum No. of Hours
1.	Basics of Chemoinformatics	4
2.	Chemoinformatics and computational Tools and Techniques	25
3.	Molecular similarity measures	6
4.	High throughput screening	10
5.	Quantitative structure-activity relationship	15

Lectures = 60
Practicals / Tutorials = 60
Total = 120

Detailed Syllabus

1. Basics of Chemoinformatics 4 hours

Introduction and Scope of Chemoinformatics – Chemical Information Sources- Structural representations and file formats – Chemical data formats and standard databases – Area of application of Chemoinformatics – Property prediction, QSAR, QSPR etc.

2. Chemoinformatics and computational Tools and Techniques 25 hours

Introduction to Chemoinformatics related algorithms – Graph theoretical applications in chemistry – Chemical information retrieval – 2D searches (structural and substructure searches, similarity searches), 3D searches (virtual screening, conformational searches, flexible searches), building queries (pharmacophore generation and validation), Diversity searches and diversity analysis, Algorithms for diversity analysis – Calculation of structure descriptors – Estimation of Physical and Chemical properties – Methods of data analysis - Machine learning techniques – Chemometrics, Statistical techniques – Design of chemical (combinatorial) libraries – Examples of chemoinformatics software.

Web based chemoinformatics : Molecular data generated in modern drug discovery efforts – Data – structured numerical annotation /text – graphical overview of the architecture, database access, similarity and subtractive searches, Library design and reactant solution and filtering

3. Molecular similarity measures 6 hours

Relavance chemical reasoning and analysis: Molecular representations and their similarity measures, chemical graphs, discrete and continuous- valued feature vectors, Field based functions – representations of Molecular fields, chemistry spaces.

4. High throughput screening 10 hours

Practical aspects, clustering and similarity, recursive partitioning, rule based evaluation, pharmacophore modelling, data sharing issues associated with NCI dataset

Information compound sets – High throughput screening, Methods, techniques and strategies implemented at Pharmacia, computational methods, custom prioritization software, score components – q score, molecular complexity estimated stability, scoring strategy Biological score, important features of scores and statistics

5. Quantitative structure – activity relationship 15 hours

Design of compounds for medicinal use, predicted bioactivity, QSAR virtual throughput sequencing (VHTS), bioactive data Chiral centres Force fields and partial changes, molecular confirmations systematic conformation feature – Molecular alignment.

Alignment – free 3D description software and hardware – Description calculation, Molecular surface, Molecular Lipophilicity, potential calculations

RECOMMENDED BOOKS

MAIN READING

1. Chemoinformatics (Methods in Molecular Biology Vol. 275 Ed. By Jurgen bajorath. Humana Press 2004
2. Structural Bioinformatics. Ed. By P. E. Bourne and H. Weissig. Wiley-Liss 2003
3. J. Gasteiger “Chemoinformatics: A text book” John Wiley and Sons 2003