



# SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited A++ Grade by NAAC | 12B Status by UGC | Approved by AICTE

[www.sathyabama.ac.in](http://www.sathyabama.ac.in)

## School of Bio & Chemical Engineering

### Department of Biotechnology

#### 1.1.3 Courses having focus on Employability / Entrepreneurship / Skill Development

#### CONTENTS

#### M.Sc. Bioinformatics and Data science

| S. No.                  | Course Code | Course name                            | Page No |
|-------------------------|-------------|----------------------------------------|---------|
| <b>Employability</b>    |             |                                        |         |
| 1.                      | SBIA5101    | Structural and functional genomics     | 3       |
| 2.                      | SBIA5102    | Python for biologists                  | 4       |
| 3.                      | SBIA5103    | Biological databases and data analysis | 5       |
| 4.                      | SMTA5102    | Statistics and random process          | 7       |
| 5.                      | SBIA5203    | Biomolecular Modelling                 | 9       |
| 6.                      | SBIA5302    | Computer aided drug design             | 11      |
| 7.                      | SBIA5304    | Microarray data analysis               | 13      |
| 8.                      | SBIA5401    | R programming                          | 15      |
| 9.                      | SBIA7406    | Data mining in bioinformatics          | 16      |
| 10.                     | SBIA7408    | Next generation sequence data analysis | 17      |
| 11.                     | SBIA7409    | Health informatics                     | 18      |
| 12.                     | SBIA7412    | Cloud computing                        | 19      |
| <b>Entrepreneurship</b> |             |                                        |         |
| 13.                     | SBIA5102    | Python for biologists                  | 4       |



|                          |          |                                        |    |
|--------------------------|----------|----------------------------------------|----|
| 14.                      | SBIA5302 | Computer aided drug design             | 11 |
| 15.                      | SBIA5304 | Microarray data analysis               | 13 |
| 16.                      | SBIA5401 | R programming                          | 15 |
| 17.                      | SBIA7406 | Data mining in bioinformatics          | 16 |
| 18.                      | SBIA7408 | Next generation sequence data analysis | 17 |
| 19.                      | SBIA7409 | Health informatics                     | 18 |
| 20.                      | SBIA7412 | Cloud computing                        | 19 |
| <b>Skill Development</b> |          |                                        |    |
| 21.                      | SBIA5101 | Structural and functional genomics     | 3  |
| 22.                      | SBIA5102 | Python for biologists                  | 4  |
| 23.                      | SBIA5103 | Biological databases and data analysis | 5  |
| 24.                      | SMTA5102 | Statistics and random process          | 7  |
| 25.                      | SBIA6101 | Python lab                             | 8  |
| 26.                      | SBIA5203 | Biomolecular Modelling                 | 9  |
| 27.                      | SCSA6201 | Oracle and SQL lab                     | 10 |
| 28.                      | SBIA5302 | Computer aided drug design             | 11 |
| 29.                      | SBIA5304 | Microarray data analysis               | 13 |
| 30.                      | SBIA6301 | Computer Aided Drug Design Lab         | 14 |
| 31.                      | SBIA5401 | R programming                          | 15 |
| 32.                      | SBIA7406 | Data mining in bioinformatics          | 16 |
| 33.                      | SBIA7408 | Next generation sequence data analysis | 17 |
| 34.                      | SBIA7409 | Health informatics                     | 18 |
| 35.                      | SBIA7412 | Cloud computing                        | 19 |



| SBIA5101 | STRUCTURAL AND FUNCTIONAL GENOMICS | L | T | P | Credits | Total Marks |
|----------|------------------------------------|---|---|---|---------|-------------|
|          |                                    | 3 | * | 0 | 3       | 100         |

### COURSE OBJECTIVES

To have a knowledge on the conformational parameters and their influence on the structure of the macromolecule and to acquire a knowledge on the structure and activity relationship.

### UNIT I INTRODUCTION

Structure & Organization of Prokaryotic & Eukaryotic genome - Nucleotide and protein sequencing methods –Chemical, Enzymatic, high through put method – Automated sequencing methods – shotgun – chromosome walking, Contig assembly - Levels of structures in Biological macromolecules.

### UNIT II CONFORMATIONAL ANALYSIS

**Biomolecules and their interactions** - Forces that determine protein and nucleic acid structure, basic problems, polypeptide chains geometrics, potential energy calculations, observed values for rotation angles, hydrogen bonding, hydrophobic interactions and ionic interactions, disulphide bonds - **Prediction of proteins structure - Nucleic acids**, general characteristics of nucleic acid structure, geometrics, glycosidic bond rotational isomers - Ribose puckering - forces stabilizing ordered forms, base pairing, base stacking- tertiary structure of nucleic acids.

### UNIT III STRUCTURAL ANALYSIS OF MACROMOLECULES

Size and shape of macromolecules - methods of direct visualization - **X-ray crystallography – X-ray diffraction, determination of molecular structures, electron microscopy, NMR**. Protein structure databases - Protein Data Bank - SCOP - CATH - structure superposition - RMSD - TM-score- structure alignment - Different structure alignment algorithms - DALI, CE, VAST, TM-align - protein folds in PDB.

### UNIT IV STRUCTURE-FUNCTIONS RELATIONSHIP

DNA binding proteins, Prokaryotic and Eukaryotic transcription factors - DNA polymerases, Helix-turn-Helix motif in DNA binding, Trp repressor, Zn fingers, helix-turn helix motifs in homeodomain, Leucine zippers - DNA polymerases, Membrane proteins and receptors, Trans-membrane segments, bacterio-rhodopsin, photosynthetic centres, epidermal growth factor, insulin and PGDGF receptors and their interaction with effectors, protein phosphorylation, immunoglobulins, Serine proteases, ribonuclease, lysozyme.

### UNIT V FUNCTIONAL GENOMICS

Introduction to Functional Genomics - cytological maps - Hap Map - SNPs and variation- Genotyping, microarray- analysis and applications. Microarray oligo sets and fabrication - integrative genomics and Meta genomics. **Quantitative Proteomics. Fluorescent imaging techniques. Quantitative RT-PCR analysis. High-throughput cloning and expression strategies. Next Generation Sequencing applications.**

### COURSE OUTCOMES

On completion of the course the student will be able to

- CO1:** The student will acquire a basic knowledge on the basics of genomics.
- CO2:** They will be able to appreciate the influence of conformational parameters on the structure of the macromolecule.
- CO3:** Will be able to select the particular method of structural analysis methodology, given the protein.
- CO4:** Assess the structure and activity relationship of the macromolecule.
- CO5:** Acquire knowledge on the functional aspects of genomics.
- CO6:** Explore the strategies used in high throughput screening of biological data

### TEXT / REFERENCE BOOKS

1. Cantor R., Schimmel P.R., Biophysical Chemistry, Vol. I, II, W.H. Freeman & Co., 1985.
2. Kensal E. van Holde, W. Curtis Johnson and P. Shing Ho, Principle of Physical Biochemistry, Prentice Hall, New York, 1998.
3. Pennington SR, Dunn MJ, Proteomics from Protein Sequence to Function, Viva Books Ltd, 2002.
4. G. Gibson and M. V. Muse. A primer of Genome Science. Sinauer Associates Inc; 2 edition, December 2004.

### END SEMESTER EXAM QUESTION PAPER PATTERN

| SBIA5102 | PYTHON FOR BIOLOGISTS | L | T | P | Credits | Total Marks |
|----------|-----------------------|---|---|---|---------|-------------|
|          |                       | 3 | 1 | 0 | 4       | 100         |

## COURSE OBJECTIVES

To empower the students in the life sciences, who may have a pertinent scientific question that can be readily answered by computational techniques

### UNIT I PROGRAMMING PRINCIPLES

Basic programming concepts - what is a program? - why python? - main features of python - comparing python with other languages - how is it used? - who uses python? - flavors of python - special python distributions - installing python - interactive mode - batch mode - other tools - additional resources - Basic Programming - Simple Values Booleans - Integers Floats Strings Expressions - Numeric Operators - Logical Operations - String Operations – Calls - Compound Expressions- Assigning Names - Defining Functions - Using Modules

### UNIT II COLLECTIONS AND CONTROL STATEMENTS

Sets - Sequences - Strings, Bytes, and Byte arrays - Ranges - Tuples - Lists - Mappings - Dictionaries - Streams - Files - Generators - Collection- Related Expression Features - Comprehensions - Functional Parameters - Conditionals – Loops – Iterations - Exception Handlers - Extended Examples

### UNIT III CLASSES, UTILITIES AND PATTERN MATCHING

Defining Classes - Class and Method Relationships - System Environment - Working with Text - Persistent Storage - Fundamental Syntax - The Actions of the re Module - Results of re Functions and Methods - Putting It All Together

### UNIT IV STRUCTURED TEXT, WEB PROGRAMMING, RELATIONAL DATABASES AND GRAPHICS

HTML – XML - An XML File for a Complete Genome - Manipulating URLs: urllib.parse - Opening Web Pages: web browser - Web Clients - Web Servers - Tips, Traps, and Tracebacks - Representation in Relational Databases - Using Relational Data - SQL Basics - SQL Queries - Introduction to Graphics Programming

### UNIT V BIOPYTHON AND PYTHON RECIPES

Project Organization - INSTALLING BIOPYTHON - Alphabet - Seq – Mutable Seq – Seq Record - Align - Align IO - ClustalW - SeqIO - AlignIO - BLAST - Biological Related Data - Entrez - PDB - PROSITE - Restriction - SeqUtils – Sequencing - SwissProt – Python recipes with commented source code – Collaborative Development: Version control with GitHub – PythonAnywhere – Scientific Python

## COURSE OUTCOMES

At the end of the course the students will be able

- CO 1- To understand why Python is a useful scripting language for developers.
- CO 2- To learn how to design and program Python applications.
- CO 3- To learn how to use lists, tuples, and dictionaries in Python programs.
- CO 4- To learn how to identify Python object types.
- CO 5- To apply python in real time informatics applications.
- CO 6- To learn how to use biopython and online resources for long – term learning.

## TEXT / REFERENCE BOOKS

1. Python Programming for Biology-Bioinformatics and Beyond, Tim J. Stevens, MRC Laboratory of Molecular Biology, Cambridge, Wayne Boucher, University of Cambridge
2. Managing Your Biological Data with Python by Allegra Via, Chapman and Hall/CRC; 1 edition (22 August 2017)
3. Practical Computing for Biologists by Steven H. D. Haddock, OUP USA; 2011 edition (5 November 2010)
4. Python for Biologists: A Complete Programming Course for Beginners, Dr Martin Jones, Createspace Independent Publishing Platform; 1 edition (7 September 2013)
5. Bioinformatics Programming Using Python by Mitchell L Model, Released December 2009, Publisher(s): O'Reilly Media, Inc. ISBN: 9780596154509
6. Python for Bioinformatics by Sebastian Bassi ISBN 9781138035263, Published July 30, 2017 by Chapman and Hall/CRC

| SBIA5103 | BIOLOGICAL DATABASES AND DATA ANALYSIS | L | T | P | Credits | Total Marks |
|----------|----------------------------------------|---|---|---|---------|-------------|
|          |                                        | 3 | * | 0 | 3       | 100         |

## COURSE OBJECTIVES

To have a knowledge on the conformational parameters and their influence on the structure of the macromolecule and to acquire a knowledge on the structure and activity relationship.

## UNIT I OVERVIEW AND CHALLENGES

Problem and Scope - Biological Data Integration - Developing a Biological Data Integration System - Specifications - Translating Specifications into a Technical Approach - Development Process - Evaluation of the System - The Life Science Discovery Process - An Information Integration Environment for Life Science Discovery - The Nature of Biological Data - Diversity - Variability - Data Sources in Life Science - Biological Databases Are Autonomous - Heterogeneous in Data Formats - Biological Data Sources Are Dynamic - Computational Analysis Tools Require Specific Input/Output Formats and Broad Domain Knowledge - Challenges in Information Integration - **Data Integration - Meta-Data Specification - Data Provenance and Data Accuracy - Ontology**

## UNIT II DATA MANAGEMENT AND INTEGRATION

Data Management - Basics - **Two Popular Data Management Strategies and Their Limitations** - Traditional Database Management - Dimensions Describing the Space of Integration Solutions - - Browsing vs. Querying - Syntactic vs. Semantic Integration - Warehouse vs. Federation - Declarative vs. Procedural Access - Generic vs. Hard-Coded - Relational vs. Non- Relational Data Model - Use Cases of Integration Solutions - Browsing-Driven Solutions - Data Warehousing Solutions - Federated Database Systems Approach - Semantic Data Integration - Strengths and Weaknesses of the Various Approaches to Integration - Browsing and Querying: Strengths and Weaknesses - Warehousing and Federation: Strengths and Weaknesses - Generic and Hard-Coded Approaches: Strengths and Weaknesses - **Relational and Non-Relational Data Models**: Strengths and Weaknesses - A Hybrid Approach to Integration Is Ideal

## UNIT III DESIGNING A BIOLOGICAL LEGACY

Biological Data - **Biological Tools and Workflows** - A Domain in Constant Evolution - Traditional Database Management and Changes - Data Fusion - Fully Structured vs. Semi-Structured - Scientific Object Identity - Concepts and Ontologies - Biological Queries - Searching and Mining - Browsing - Semantics of Queries - **Tool-Driven vs. Data-Driven Integration** - Query Processing - Biological Resources - Query Planning - Query Optimization - Visualization - Multimedia Data - Browsing Scientific Objects

## UNIT IV MODEL DATABASES

**SRS - An Integration and Analysis Tools** Platform for Databanks - Integrating Flat File Databanks - Integration of XML Databases - Integrating Relational Databases - The SRS Query Language - Linking Databanks - The Object Loader - Scientific Analysis Tools - Interfaces to SRS - **Building biological databases with SQL** - Common database types - Relational databases -

## UNIT V DATA MINING

Biomarker discovery - An introduction to biomarker discovery - Data preprocessing - Modeling - Validation - Case study - Discussion and future perspective

## COURSE OUTCOMES

On completion of the course the student will be able to

- CO1-** Acquire the basic understanding of a biological database
- CO2-** Explore the various platforms for constructing the database.
- CO3-** Design the biological database and defining its attributes.
- CO4-** Assess the structure and modelling of complex databases.
- CO5-** Acquire knowledge on data mining and futuristic perspectives on biological data mining.
- CO6-** Explore the data mining role in biomarker discovery.

### TEXTBOOKS / REFERENCE BOOKS

1. Zengyou, He. Data Mining for Bioinformatics Applications, Elsevier Science & Technology, 2015.
2. Bioinformatics : Managing Scientific Data, edited by Zoé Lacroix, and Terence Critchlow, Elsevier Science & Technology, 2003.
3. Bessant, Conrad, et al. Building Bioinformatics Solutions, Oxford University Press USA - OSO, 2014.

### END SEMESTER EXAM QUESTION PAPER PATTERN

**Max. Marks : 100**

**Exam Duration : 3 Hrs.**

**PART A :** 6 Questions for 5 Marks each –without choice

**30 Marks**

**PART B :** 2 Questions from each unit with internal choice, each carrying 14 marks

**70 Marks**



| SMTA5102 | STATISTICS AND RANDOM PROCESS | L | T | P | CREDITS | TOTAL MARKS |
|----------|-------------------------------|---|---|---|---------|-------------|
|          |                               | 3 | 0 | 0 | 3       | 100         |

## COURSE OBJECTIVES

To acquire skills in handling situations involving more than one random variable and functions of random variables. This gives exposure to statistical methods designed to contribute to the process of making scientific judgments in the face of uncertainty and variation.

### UNIT 1 POINT ESTIMATION

Properties of estimators, Unbiasedness, Efficiency, Consistency, Sufficiency - Rao - Blackwell theorem, C.R. Rao Inequality - Methods of estimation: Method of Likelihood Estimation, Method of Minimum  $\chi^2$  Modified  $\chi^2$  - Method of moments.

### UNIT II THEORY OF SAMPLING AND TEST OF HYPOTHESIS

Test of Hypothesis - large sample - test of significance - Single mean, difference of means, single proportion, difference of proportions - small sample - students 't' test - single mean - difference of means - Fisher's test or difference of variance - Chi square test - goodness of fit - independence of attributes.

### UNIT III NON - PARAMETRIC TESTS

Sign test, Wilcoxon Signed Rank Test, Median test, Mann-Whitney U test, Kruskal - Wallis H test.

### UNIT IV INTRODUCTION TO STOCHASTIC PROCESSES

Definition, Classification of Stochastic processes, Random walk, Markov chains, Transition Probability Matrix, Graphical representation of Markov chains, Classification of states and chains- Simple Problems- Chapman Kolmogorov Theorem.

### UNIT V MARKOV PROCESS

Introduction to Markov Process- Poisson process-Postulates and properties-Birth and Death process-Hidden Markov Model - Introduction - Vitebi Algorithm, Forward and Backward Algorithm, Estimation algorithms.

## TEXT / REFERENCE BOOKS

1. Veerarajan T, Probability, Statistics and Random Process, 4th edition, Tata McGraw Hill, 2008.
2. Medhi J, Stochastic Processes, Second Edition, New Age International Publishers. 2002.
3. Stephen Bernstein Schaum's Outline of Elements of Statistics I: Descriptive Statistics and Probability, McGraw-Hill, 1998.
4. Jerrold H zar, Bio statistical Analysis, 5th edition, Prentice Hall, 2010.
5. Irvin Miller John E Freund, Probability and Statistics for Engineers, Prentice Hall, 1977.

## COURSE OUTCOMES

- CO1 Define characteristics of good estimators, List the methods of estimators, and Evaluate the MLE of estimators.
- CO2 Understand the basis and applications of sampling; Choose the appropriate tests of significance, Formulate the hypothesis for testing the significance of difference, Identify the size of the sample, level of significance, degrees of freedom of the samples.
- CO3 Advantages of non-parametric tests Choose the appropriate tests and Evaluate the various methods of Non-Parametric tests.
- CO4 List the classification of stochastic process, evaluate problems in tpm, Identify the classification of states and chains.
- CO5 Enable students to understand the tools used in bioinformatics and their corresponding algorithms used for evaluating the bio models.
- CO6 Explore the statistical methods designed to contribute to making scientific judgments in the face of uncertainty and to design suitable models.

| SBIA6101 | PYTHON LAB | L | T | P | Credits | Total Marks |
|----------|------------|---|---|---|---------|-------------|
|          |            | 0 | 0 | 4 | 2       | 100         |

#### SUGGESTED LIST OF EXPERIMENT

1. Counting DNA Nucleotides
2. Transcribing DNA into RNA
3. Computing GC Content
4. Counting Point Mutations
5. Translating RNA into Protein
6. Finding a Motif in DNA and Finding a Protein Motif
7. Consensus and Profile
8. Inferring mRNA from Protein
9. Open Reading Frames
10. Calculating Protein Mass
11. Locating Restriction Sites
12. RNA Splicing
13. Longest Increasing Subsequence and Genome Assembly as Shortest Superstring
14. Finding a Spliced Motif
15. Transitions and Transversions
16. Biopython - Parsing GenBank records from the net and SwissProt sequences
17. Biopython - Converting between sequence file formats
18. Biopython - Multiple Alignments
19. Biopython - Determining the secondary structure
20. Biopython - Annotated Chromosomes
21. Biopython - Parsing KEGG records
22. Biopython - Simple quality filtering for FASTQ files
23. Biopython - Calculating a substitution matrix from a multiple sequence alignment
24. Biopython - Sequence parsing plus simple plots



| SBIA5203 | BIOMOLECULAR MODELLING | L | T | P | Credits | Total Marks |
|----------|------------------------|---|---|---|---------|-------------|
|          |                        | 3 | 1 | 0 | 4       | 100         |

## COURSE OBJECTIVES

To enable the student to analyze and discuss the different constraints for a macromolecular modelling and debate the predicted structures of the models.

### UNIT I BASIC CONCEPTS IN MOLECULAR MODELLING

Internal parameters - Z-matrix – Introduction to quantum chemistry –basic postulates – Schrodinger wave equation – Derivation – Hydrogen atom – Born-Oppenheimer approximation - Laws of thermodynamics - entropy – enthalpy - free energy calculations-chemical potential - Calculating thermodynamic properties using force field; Transferability of force field parameters, treatment of de-localized pi system; Force field for metals and inorganic systems – Application of energy minimization.

### UNIT II MOLECULAR GEOMETRY

Conformational parameters - Potential energy Surface - Molecular mechanics: empirical forces fields - bond stretching - angle bending - torsional terms – non-bonded and electrostatic interaction.

### UNIT III TYPES OF FORCE FIELDS AND ENERGY MINIMIZATION

Simplex – sequential univariate method - steepest descent - conjugate gradient method - Newton–Raphson method – Molecular Dynamics with continuous potentials and at constant temperature and pressure; Time dependent properties; Solvent effects in Molecular Dynamics - Conformational analysis - Molecular visualization - Molecular graphics – Rendering - Rasmol.

### UNIT IV MOLECULAR DYNAMICS (MD) SIMULATION OF BIOPOLYMERS

Time steps - Setting up MD - energy conservation in MD Simulation -continuous potentials and constraint dynamics - MD at constant temperature and pressure - incorporating solvent effects - examples-random number generator - Monte Carlo simulation of biological macromolecules - MD softwares

### UNIT V PREDICTION OF SECONDARY STRUCTURE

Membrane prediction – Comparative modeling - Sequence Alignment Homolog's - analogs - Homology modeling - steps in homology modeling – side chain modeling – loop modeling - fold recognition – ab initio prediction – Predicting protein structures by threading protein folding – active site/binding site prediction – tools – databases - CASP.

## COURSE OUTCOMES

At the end of the course

- CO1-** Students will learn the basics of molecular modeling aspects.
- CO2-** They will be able to relate the conformational parameters to the conformation of the molecule.
- CO3-** They will be able to predict the effect of the force field applied on a molecule.
- CO4-** Students will be able to examine and analyze the significance of simulation of macromolecules.
- CO5-** Students will be able to appreciate the predicted models.
- CO6-** Assessment of novel algorithm in structure prediction

## TEXTBOOKS / REFERENCE BOOKS

1. Leach A.R., Molecular Modelling - Principles and Applications, 2nd Edition, Prentice Hall, 2001.
2. Vinter J.G. and Mark Gardener, Molecular modeling and drug design, Mac Millan, 1994.
3. Prasad, R.K., Quantum Chemistry, Halsted Press, 1992.
4. Ramachandran, K.I, Deepa, G, Namboori, K, Computational Chemistry and Molecular Modeling: Principles & Application, Springer 2008.
5. 3. Young, D.C., Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems, Wiley-Indescience 2001

| SCSA6201 | ORACLE AND SQL LAB | L | T | P | CREDITS | TOTAL MARKS |
|----------|--------------------|---|---|---|---------|-------------|
|          |                    | 0 | 0 | 4 | 2       | 100         |

### LIST OF EXPERIMENTS

#### COURSE OBJECTIVES

1. Learn and practice data modelling using the entity-relationship and developing database designs.
2. Understand the use of Structured Query Language (SQL) and learn SQL syntax.
3. Apply normalization techniques to normalize the database
4. Understand the needs of database processing and learn techniques for controlling the consequences of concurrent data access.

#### COURSE OUTCOMES

**CO 1:** The learner will be able:

**CO 2:** To describe data models and schemas in DBMS.

**CO 3:** To understand the features of database management systems and Relational database

**CO 4:** To use SQL-the standard language of relational databases

**CO 5:** To understand the functional dependencies and design of the database

**CO 6:** To understand the concept of Transaction and Query processing.

1. Data Definition Language (DDL) commands in RDBMS
2. Data Manipulation Language (DML) and Data Control Language (DCL)
3. High level language extensions with cursors
4. High level language extension with Triggers
5. Procedures and Functions
6. Embedded SQL
7. Database design using E-R model and Normalization
8. Design and implementation of payroll processing system
9. Design and implementation of Banking system
10. Design and implementation of Library Information System
11. Design and implementation of Student Information System
12. Automatic Backup of Files and Recovery of Files



| SBIA5302 | COMPUTER AIDED DRUG DESIGN | L | T | P | CREDITS | TOTAL MARKS |
|----------|----------------------------|---|---|---|---------|-------------|
|          |                            | 3 | 1 | 0 | 4       | 100         |

### COURSE OBJECTIVES

To understand the basic concepts in drug design and to know the significance of computational methods and algorithm in drug design. Henceforth, to appreciate the logical reasoning and intellectual thinking in bridging the gap in drug discovery.

### UNIT I INTRODUCTION TO DRUG DESIGN AND DEVELOPMENT

9hrs

Drug Discovery as a Process, Stages of drug discovery and development, Pathophysiology of a disease, Target Identification and Validation, 3-dimensional structure of the receptor-protein, nucleic acid, protein-nucleic acid complex, Structure Determination- X-ray crystallography, NMR spectroscopy, and mass spectrometry, Drug Target Validation.

### UNIT II DRUG TARGETS

9hrs

Membrane proteins- receptors, ion channels and transporters, advantages and disadvantages of membrane proteins as drug targets, Nucleic acids-DNA, messenger RNA, and ribosomal RNA are important molecular targets for cancer, viral, and microbial chemotherapy, Rational design of DNA groove and recognition of RNA binding sites. Enzymes in cellular processes, enzymes as drug targets, development of agents that selectively block mRNA- antisense oligonucleotides, ribozymes, siRNA, gene silencing

### UNIT III LEAD IDENTIFICATION AND MODIFICATION

9hrs

Biological assays- High-throughput screening, virtual screening-random , non-random screening, Combinatorial Library Design, pharmacophore modeling, functional group optimization, QSAR and QSPR methods- Partition coefficient, Hammett's substituent constant and Taft's steric constant. 3DQSAR- COMFA, COMSIA, bioisosteric replacement, fragmentation, rigid analogs, drug binding kinetics- competitive, non-competitive inhibition, binding parameters-binding affinity, dissociation constants, inhibitory constants

### UNIT IV COMPUTER-AIDED DRUG DESIGN 9hrs

Molecular modeling- molecular mechanics and quantum mechanics Force field methods ,Minimum energy conformation-energy minimization, Lead based drug design-geometry optimisation, conformational searching. Pharmacophore mapping, molecular docking molecular dynamics

### UNIT V DRUG DELIVERY

9hrs

Bioavailability, pharmacodynamics and pharmacokinetics. ADME/Toxicity predictions, Pre-Clinical and Clinical testing, Clinical Trials- Clinical Trial Design, Ethics of Human and Animal Experimentation, Intellectual Property, Commercial Considerations in Drug Development

Max 45 hrs

### COURSE OUTCOMES

The student will be able to:

- CO 1-** Apply and understand the modelling on the basis of quantum mechanical aspects.
- CO 2-** Discuss the various types of interactions possible between the macromolecule and the ligand. **CO 3-** Analyse the improvement of current anticancer and antibacterial drugs through rational design. **CO 4-** Explore the interaction of drug target and drug
- CO 5-** Develop knowledge and techniques to identify drug targets
- CO 6-** Critically evaluate the strategy for lead identification and optimizing pharmacokinetic properties

#### **TEXT / REFERENCE BOOKS**

1. Shayne C. Gad, (2005), Drug Discovery Handbook, 2nd edition, John Wiley & Sons: New York.
2. Tudor I. Opera, (2005), Chemoinformatics in Drug Discovery, 5th edition, John Wiley & Sons: New York.



| SBIA5304 | MICROARRAY DATA ANALYSIS | L | T | P | Credits | Total Marks |
|----------|--------------------------|---|---|---|---------|-------------|
|          |                          | 3 | 1 | 0 | 4       | 100         |

## COURSE OBJECTIVES

Students are allowed to understand the key concepts of class comparison and class discovery of genes.

### UNIT I GENE EXPRESSION

Basics of Gene expression- definition , gene expression studies, gene expression patterns- Applications of gene expression studies- Microarrays – definition , discovery, technique, making microarrays, spotted microarrays, In-situ synthesized oligonucleotide arrays, inkjet array synthesis, Affymetrix techniques, DNA CHIP technology, photolithography, spot quality, sample preparation and labelling, washing, image acquisition Sequencing by Hybridization Arrays- DNA Mass Array™ Technology- Printing DNA Microarrays-Types of microarrays - Designing a microarray experiment.

### UNIT II IMAGE PROCESSING

Image processing, feature extraction, identifying positions of features- Normalization – data cleaning and transformation, within array normalization, between array normalization, measuring and quantifying microarray variability –variability between replicate features on an array-, variability between hybridizations to different arrays. Analysis of differentially expressed genes- significance analysis of microarrays.

### UNIT III PREDICTION

Prediction of cross hybridization to related genes, thermodynamics of nucleic acid duplexes, prediction of T<sub>m</sub>- probe secondary structure. Analysis of relationships between genes, tissues or treatments- similarity of gene or sample profiles – dimensionality reduction, principal component analysis, hierarchical clustering, machine learning methods for cluster analysis. Classification of tissues and samples – validation.

### UNIT IV CARBOHYDRATE MICROARRAYS

Carbohydrate microarrays- Carbohydrate sources- Synthesis of oligosaccharides - Isolation of oligosaccharides from natural sources- Arrays of monosaccharides and disaccharides- Arrays of polysaccharides- Arrays of oligosaccharides-Perspectives- DNA micro array analysis-clustering-immunological applications.

### UNIT V DATABASES AND TOOLS FOR MICROARRAYS

Bioinformatics in Arrays- Databases and tools for microarrays- Bioconductor, expression profiler, EST databases- Assessing levels of gene expression using EST's, TIGR gene indices, STACK, SAGE, CGAP, Xprofiler, ARRAY DB, cluster, tree view, Scanalyze, gene cluster, informatics aspects of microarray production- MGED and gene-ontology, description of MIAME (Minimum Information About a Microarray Experiment), Business Aspects of Biochip Technologies- Microarray Technology in Treating Disease.

## COURSE OUTCOMES

At the end of the course the students will be able to

- CO 1- To learn the basics of DNA Microarray and its applications
- CO 2- To learn how to Analyze of the raw RNA sequence data
- CO 3- To understand why Statistical computing and Statistical Genetics is required
- CO 4- To apply data analysis to microarray related concepts
- CO 5- To apply the finding of the analytics in healthcare
- CO 6- To develop novel algorithms in the area

## TEXT / REFERENCE BOOKS

1. Dov Stekel - Microarray bioinformatics, Cambridge,
2. Steen Knudsen - A Biologist's Guide to analysis of DNA microarray data, second edition, John Wiley & Sons, Inc.,
3. Andreas D Baxevanis and BF Francis Ouellette - Bioinformatics: A Practical Guide to the Analysis of Gene and Proteins , Wiley-Interscience,
4. David W. Mount - Bioinformatics: Sequences and genome analyses, Cold Spring Harbor Laboratory press,

| SBIA6301 | COMPUTER AIDED DRUG DESIGN LAB | L | T | P | Credits | Total Marks |
|----------|--------------------------------|---|---|---|---------|-------------|
|          |                                | 0 | 0 | 4 | 2       | 100         |

1. Molecular Visualization Softwares: Pymol and RasMol
2. Geometry Optimization
3. Binding Site Identification - Cast-P, POCASA, 3D ligand site, Metapocket,
4. Molecular Docking using discovery studio
5. Protein-ligand complex MD simulation using DS
6. Ligand Based drug design – QSAR



| SBIA5401 | R PROGRAMMING | L | T | P | Credits | Total Marks |
|----------|---------------|---|---|---|---------|-------------|
|          |               | 3 | 1 | 0 | 4       | 100         |

### COURSE OBJECTIVES

This course provides knowledge to handle biological problems in Bioinformatics and computational biology

#### UNIT I – R LANGUAGE

Motivation of R, Attributes and some special values, Types of objects, Sequence generating and vector subsetting, Types of functions, Data Structures, Managing R session, Evaluations.

#### UNIT II - OBJECT ORIENTED PROGRAMMING IN R

OOP in R, Basics of OOP, S3 OOP, S4 OOP, Documentation, Debugging, Managing S3 and S4 together.

#### UNIT III - INPUT AND OUTPUT IN R

Basic file handling, connections, File input and output, Tools for accessing files on the Internet.

#### UNIT IV - WORKING WITH CHARACTER DATA

Built-in Capabilities, Regular Expression, Prefix, Suffix and Substrings, Biological sequences, Matching patterns.

#### UNIT V - R PACKAGES

Package basics, Package Management and authoring, Initialization, Debugging and Profiling

### COURSE OUTCOMES

At the end of the course the students will be able to

- CO 1- Providing general properties and unique aspects of the R language
- CO 2- To teach Object Oriented programming in R
- CO 3- To teach methods for getting of data in and out
- CO 4- To provide knowledge how to how to write own software packages
- CO 5- To port this understanding to healthcare problems
- CO 6- To develop novel pipelines for genomic research

### TEXT / REFERENCE BOOKS

- Robert G, "R programming in Bioinformatics", CRC press, Taylor and Francis Group, USA, 2008.
- Own J, Robert. M, and Andrew R., "Introduction to Scientific programming and simulation using R", CRC Press, Taylor and Francis Group, USA, 2014.
- Norman M. And Norman S. Matloff "The Art of R programming: A tour of statistical software Design", No Starch Press Inc, USA, 2011.

### END SEMESTER EXAM QUESTION PAPER PATTERN

Max. Marks: 100

Exam Duration: 3 Hrs.

PART A: 6 Questions for 5 Marks each – without choice

30 Marks

PART B: 2 Questions from each unit with internal choice, each carrying 14 marks

70 Marks

| SBIA7406 | DATA MINING IN BIOINFORMATICS | L | T | P | Credits | Total Marks |
|----------|-------------------------------|---|---|---|---------|-------------|
|          |                               | 3 | 0 | 0 | 3       | 100         |

### COURSE OBJECTIVES

This course will introduce the concepts of data ware house and data mining, which gives a complete description about the principles, used, architectures, applications, design and implementation of data mining and data ware housing concepts

### UNIT I INTRODUCTION

Introduction, Importance of Data Mining, Relational Databases, Data Warehouses, Transactional Databases, Advance Database Systems and Applications, Data Mining Functionalities, Classification of Data Mining Systems, Major issues in Data Mining.

### UNIT II PRIMITIVES AND SYSTEM ARCHITECTURES

Data Mining Primitives, Data Mining Query Language, Designing Graphical User, Interfaces Based on a Data Mining Query Language, Architectures of Data Mining Systems.

### UNIT III CONCEPT DESCRIPTION AND ASSOCIATION RULES

Concept Description, Characterization and comparison, Data Generalization and Summarization-Based Characterization, Analytical Characterization, Mining Class Comparisons, Mining Association Rules in Large Databases, Association Rule Mining, Mining Single Dimensional Boolean Association Rules from Transactional Databases.

### UNIT IV CLASSIFICATION AND PREDICTION

Classification and Prediction, Issues: Data preparation for classification and Prediction, Comparing classification Methods Classification by Decision Tree Induction: Decision Trees and Decision Tress induction

### UNIT V CLUSTERING METHODS

Clustering Analysis, Types data in clustering analysis: Scaled variable, Binary variables, Variables of Mixed Types, Partitioning Methods: K-means and K-Medoids, Model-Based Methods, Data Mining Applications: Data mining for Biomedical and DNA Data Analysis

### COURSE OUTCOMES

- CO1- Be familiar with mathematical foundations of data mining tools.
- CO2- Understand and implement classical models and algorithms in data warehouses and data mining
- CO3- Characterize the kinds of patterns that can be discovered by association rule mining, classification and clustering.
- CO4- Master data mining techniques in various applications like social, scientific and environmental context.
- CO5- Develop skill in selecting the appropriate data mining algorithm for solving practical problems.
- CO6-

### TEXT/ REFERENCES BOOKS

1. Data Mining Concepts and Techniques – Jiawei Hen, Micheline Kambler, Academic Press Morgan Kaufman Publishers. 2006
2. Data Mining: Practical machine learning tools Techniques with java implementation by Ian H.Witten, Eibe Frank, 2005.
3. Machine Learning and data mining in pattern recognition in third International conference MLDM, by Petra Perner and Azriel Rosenfield, Springer.2003

### END SEMESTER EXAM QUESTION PAPER PATTERN

Max. Marks : 100

Exam Duration : 3

Hrs. PART A : 6 Questions for 5 Marks each –without choice

30 Marks PART B : 2 Questions from each unit with internal choice, each carrying 14 marks

70 Marks

| SBIA7408 | NEXT GENERATION SEQUENCE DATA ANALYSIS | L | T | P | Credits | Total Marks |
|----------|----------------------------------------|---|---|---|---------|-------------|
|          |                                        | 3 | 0 | 0 | 3       | 100         |

### COURSE OBJECTIVES

- NGS bioinformatics workflow steps following sequence generation
- NGS for transcriptomics; QC, mapping, visualisation tools
- NGS for genomics; assembly, alignment, QC and variant calling tools
- Exploring EMBL-EBI resources

### UNIT I DNA SEQUENCING

DNA Sequencing, first generation DNA sequences, Drawbacks of the first generation sequencing method- Emergence of Next generation sequencing -Platform overview- Biological applications- Basic concepts- Recent scientific breakthroughs using NGS technology

### UNIT II OVERVIEW OF NGS TECHNOLOGY

Pyro-sequencing, Illumina Genome Analyzer, Applied Biosystems Sequencing, Ion Torrent Sequencing, Polonator Technology, Nanopore Sequencing, Single Molecule Real Time DNA sequencing- Comparison of Next generation sequencing techniques, & applications-Drawbacks of NGS, NGS File formats- FASTA, FASTQ, FNA, CSFASTA, GFF, SAM and BAM

### UNIT III NGS DATABASES AND ANALYSIS OF NGS DATA

NGS databases, Sequence Analysis: Pairwise and multiple sequence alignment methods and tools- Sample data processing with annotation

### UNIT IV ALIGNMENT AND ANALYSIS

Genome alignment and analysis tools- BWA (Burrows-Wheeler Aligner), SAMtools, GATK (The Genome Analysis Toolkit), IGV (Integrative Genomics Viewer), HISAT, StringTie, Cuffcompare, Velvet, Oases, Trinity

### UNIT V TRANSCRIPTOMICS ANALYSIS

Transcriptome (RNA) sequencing, Exome sequencing, Genome Annotation, Using NGS to detect sequence variants, ChIP-sequence, biological theories on ChIP-sequence analysis, Understanding the non – coding genome, Disease gene identification, DNA fragment evaluation, Peak identification, two condition comparison, Saturation analysis, Motif finding and related theories

### COURSE OUTCOMES

After this course the students will be able to

- CO 1: Discuss a variety of applications and workflow approaches for NGS technologies
- CO 2: Use a range of bioinformatics software and tools to undertake basic analysis of NGS data
- CO 3: Understand the advantages and limitations of NGS analyses
- CO 4: Submit, browse and access a range of NGS data available in public repositories using EBI resources
- CO 5: Comprehend the different areas of NGS analysis utility
- CO 6: Apply the knowledge of NGS technology in transcriptome data analysis

### TEXT BOOKS /REFERENCES

1. Ali Masoudi-Nejad, Zahra Narimani, Nazanin Hosseinkhan; "Next Generation Sequencing and Sequence Assembly", Methodologies and Algorithms, Springer; 2013.
2. Stuart M. Brown, "Next-Generation DNA Sequencing Informatics", Cold Spring Harbor Laboratory Press, 2013.
3. Mark I. Rees, "Challenges and Opportunities of Next-generation Sequencing for Biomedical Research", Academic Press, 2012.

### END SEMESTER EXAM QUESTION PAPER PATTERN

Max. Marks : 100

Exam Duration : 3 Hrs.

PART A : 6 Questions for 5 Marks each –without choice

30 Mar

PART B : 2 Questions from each unit with internal choice, each carrying 14 marks

70 Marks

| SBIA7409 | HEALTH INFORMATICS | L | T | P | Credits | Total Marks |
|----------|--------------------|---|---|---|---------|-------------|
|          |                    | 3 | 0 | 0 | 3       | 100         |

### COURSE OBJECTIVES

This course provides an insight into the fundamental aspects and applications of health informatics.

### UNIT I INTRODUCTION

An Introduction to Structural Equation Modeling - Privacy and Data Protection in Healthcare - Biomedical Technology – Case - Allergic Severity Under Conjunctival Provocation - Texture Analysis of Heterogeneous MRI Data for Diagnostic Classification of Childhood Brain Tumors - Establishing a National Medical Device Registry

### UNIT II CLINICAL INFORMATICS

Endoscopy Information System (EIS) - Remote Patient Monitoring System Implementation at a Cardiac Care Centre - Vagal Nerve Stimulation (VNS) Therapy System Implementation - Improving Emergency Room Performance - Automation of Fugl- Meyer Assessment Scale for Neurorehabilitation - Extensible Neuroimaging e - Repository for Clinical Trials of Tumours - Query Tool Enabling Clinicians and Researchers to Explore Patient Cohorts

### UNIT III INTELLIGENT SYSTEMS

Clinical Decision Support Knowledge Management - Prediction of 5-Year Survival with Data Mining Algorithms – Abstractive Summarization of Drug Dosage Regimens - Exploitation of Medical Crisp Database for Fuzzy Systems - Spatial Epidemiology - E-Learning and Education - Role for Education - Proactive Support of Internet Browsing - Biomedical Informatics Education Programs and Ecosystems - e-Learning Approach to Hospital Information Systems

### UNIT IV FORMALISATION AND INFORMATICS

Formalisation of Knowledge, Ontologies, Clinical Guidelines and Standards of Healthcare Clinical Pathways: Identifying Development, Implementation and Evaluation Challenges - “HoPRO” - Data Warehouse Design - Innovative Information Systems in Intensive Care Unit - A Framework to Enrich EHR System - e ENERCA) - Based Stroke Rehabilitation - Greek Public Hospitals' Websites

### UNIT V ADVANCES

Health Information Management - Emergency Physician Productivity - Taming Big Data: An Information Extraction Strategy for Large Clinical Text Corpora - Institutional Repository of Digital Health Information Content - Healthcare IT - Neural Network in Healthcare - Healthcare Management - Public Health Informatics - Social Media and Health - Telemedicine

### COURSE OUTCOMES

The students will be able to

- CO 1- Understand the basics concepts of health informatics.
- CO 2- To understand the intricacies of clinical informatics.
- CO 3- To elaborate on the understanding of decision support systems in a hospital.
- CO 4- Comprehend clinical documentation in the real world.
- CO 5- To ponder and on the various aspects of telemedicine, healthcare, and social networks in the practice of medicine

### TEXT BOOKS / REFERENCES

1. Enabling Health Informatics Applications, edited by J. Mantas, et al., IOS Press, 2011.
2. Sullivan, Frank, and Jeremy Wyatt. ABC of Health Informatics, John Wiley & Sons, Incorporated, 2006.
3. Health Informatics : Improving Patient Care, edited by The Chartered Institute for IT BCS, BCS Learning & Development Limited, 2012.
4. International Perspectives in Health Informatics, edited by E.M. Borycki, et al., IOS Press, 2011.
5. Brender, Jytte. Handbook of Evaluation Methods for Health Informatics, Elsevier Science & Technology, 200

| SBIA7412 | CLOUD COMPUTING | L | T | P | Credits | Total Marks |
|----------|-----------------|---|---|---|---------|-------------|
|          |                 | 3 | 0 | 0 | 3       | 100         |

### COURSE OBJECTIVES

- To gain knowledge in basics of Cloud computing.
- To understand Cloud Computing architecture and industry frameworks such as Map Reduce.
- To discuss practical applications of cloud computing.

### UNIT I UNDERSTANDING CLOUD COMPUTING

History of Cloud computing -Cloud Computing Reference Model – NIST Architectural Framework -Types of Clouds- pros and cons of cloud computing –difference between web 2.0 and cloud - key challenges in cloud computing - Major Cloud players - Cloud Deployment Models - Virtualization in Cloud Computing- types of virtualization – web services – key concepts.

### UNIT II CLOUD SERVICE MODELS

Software as a Service (SaaS) - Infrastructure as a Service (IaaS)- Platform as a Service (PaaS)- Service Oriented Architectur (SoA) - Elastic Computing - On Demand Computing - Parallelization in Cloud Computing – cloud resource management – dynamic resource allocation- Optimal allocation of cloud models.

### UNIT III CLOUD DEPLOYMENT MODELS

Deployment models: Public cloud – Private Cloud –Hybrid cloud – Community cloud - Deployment of applications on the cloud - Hypervisor -Case studies- Xen, Hyper V, Virtual Box, Eucalyptus, Amazon Cloud Computing - Amazon S3 - Amazon EC2-, Windows Azure – creating deploying cloud services, Working with Google Appengine, Disaster Recovery.

### UNIT IV VIRTUALIZATION

Cloud data centers – Energy efficiency in data centre – Mobile cloud computing service models– Need for virtualization – Types of Virtualization – Virtualization OS – VMNware, KVM – System VM – Process VM - Virtual Machine Monitor – Properties Interpretation and Binary Translation – HLL VM.

### UNIT V CLOUD SECURITY

Cloud security – Security threats and solutions in clouds – Auditing protocols – dynamic auditing – storage security –Privacy preserving — Fully Homomorphic Encryption – Big data security- Cloud availability- DoS attacks – Fault tolerance management in cloud computing- Cloud computing in India.

### COURSE OUTCOMES

At the end of the course the students will be able to

- CO1- Understand the basic concepts of cloud computing.
- CO2- Use cloud service models as infrastructure.
- CO3- Deploy cloud models in a real – time environment
- CO4- To efficiently use data centres.
- CO5- Grasp virtualization.
- CO6- To comprehend security issues involved.

### TEXT / REFERENCE BOOKS

1. Cloud computing a practical approach - Anthony T.Velte, Toby J. Velte Robert Elsenpeter TATA McGraw- Hill , New Delhi, 2010.
2. Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online – Michael Miller – Que, 2008.
3. Cloud Computing Bible, Barrie Sosinsky, Wiley-India, 2010.
4. Cloud Computing: Principles and Paradigms, Editors: Rajkumar Buyya, James Broberg, Andrzej M. Goscinski, Wiley, 2011.
5. Cloud Computing: Principles, Systems and Applications, Editors: Nikos Antonopoulos, Lee Gillam, Springer, 2012.
6. Cloud Security: A Comprehensive Guide to Secure Cloud Computing, Ronald L. Krutz, Russell Dean Vines, Wiley-India, 2010.