

**SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY**

**(Established under section 3 of UGC Act, 1956)
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119.**



SYLLABUS

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOINFORMATICS

**BACHELOR OF SCIENCE IN BIOINFORMATICS AND DATA SCIENCE
[SIX SEMESTER DEGREE PROGRAMME]**

REGULATIONS 2019

Semester 5										
S.no	Course type	Course code	Course Title	L	T	P	C	Marks		Page No
								CAE	ESE	
1	Theory	SBIA1501	Modelling of molecules	3	0	0	3	50	50	45
2	Theory	SBIA1502	Drug design and development	3	0	0	3	50	50	46
3	Theory	SBIA1503	Systems Biology	3	0	0	3	50	50	47
4	Theory	SBIA1504	Cheminformatics	3	0	0	3	50	50	49
5	Theory		Elective	3	0	0	3	50	50	
6	Practical	SBIA2501	Systems biology lab	0	0	4	2	50	50	50
7	Practical	SBIA2502	Cheminformatics and Drug design Lab	0	0	4	2	50	50	50
8	Seminar	S738ASEM	Seminar	0	0	0	1		100	
Total credits for 5th Semester = 20										
Total marks for 5th Semester =800										

Semester 6										
S.no	Course type	Course code	Course Title	L	T	P	C	Marks		Page No
								CAE	ESE	
1	Theory	SBIA1601	Data mining and Machine Learning	3	0	0	3	50	50	52
2	Theory	SBIA1602	Microarray data analysis	3	0	0	3	50	50	54
3	Practical	S738APROJ	Project				15		100	
Total credits for 6th Semester = 21										
Total marks for 6th Semester =300										

Semester	Theory courses (including elective courses)			Practical Courses (including PT and project)		
	Total no.	Total Credits	Total Marks	Total no.	Total Credits	Total Marks
1	5	15	500	2	4	200
2	5	15	500	2	4	200
3	5	16	500	2	4	200
4	5	16	500	2	5	200
5	5	15	500	3	5	300
6	2	6	200	1	15	100
Overall Total	27	83	2700	10	37	1200

Overall total credits for B.Sc. programme	120
Overall total marks for B.Sc. programme	4000

L - LECTURE HOURS, T - TUTORIAL HOURS, P - PRACTICAL HOURS, C - CREDITS
 CAE - CONTINUOUS ASSESSMENT EXAMINATION, ESE - END SEMESTER EXAMINATION,
 VIVA - VIVA VOCE

SBIA1501	MODELLING OF MOLECULES	L	T	P	Credits	Total Marks
		3	1	0	4	100

COURSE OBJECTIVES

The course intends to impart knowledge on the various concepts of molecular modelling and its application in drug designing with a focus on identification of lead molecules and the concept of pharmacophore.

Unit I Introduction to quantum mechanics**9hrs**

Basic Concepts in Molecular Modelling: Introduction, Coordinate systems, Energy surfaces. Introduction to Quantum mechanics - Schrodinger wave equation, Time dependant and time independent forms.

Unit II Basics of molecular mechanics**9hrs**

Features of molecular mechanics, force fields; Bond structure and bending angles -electrostatic, van der Waals and non - bonded interactions, hydrogen bonding; Energy minimization; Application of energy minimization.

Unit III Molecular mechanics**9hrs**

Molecular Mechanics Introduction, Vander Waals and Dipole-Dipole Interaction, Types of Potential- Lennard- Jones, Truncated Lennard - Jones, Ionic and Polar potentials, Types of force field- AMBER, CHARMM, Merck Molecular Force Field, Consistent Force Field, MM2, MM3 and MM4 Force Field.

Unit IV Molecular dynamics**9hrs**

Molecular Dynamics using simple models; Molecular Dynamics with continuous potentials and at constant temperature and pressure; Time-dependent properties; Solvent effects in Molecular Dynamics; Conformational changes from Molecular Dynamics simulation. Introduction, Newton's equation of motion, equilibrium point, radial distribution function, pair correlation functions, MD methodology, periodic box, algorithm for time dependence; leapfrog algorithm, Verlet algorithm, Boltzman velocity, time steps, duration of the MD run.

Unit V Simulation and Modelling**9hrs**

Methods of computer aided drug design, ligand design methods, drug design approaches, Target identification and validation, lead optimization and validation, Structure and ligand based drug design, modelling of target-small molecule interactions.

Course outcomes

At the end of the course the students will

- CO1** –Be able to apply and understand the modelling on the basis of quantum mechanical aspects.
- CO2** –To discuss the various types of interactions possible between the macromolecule and the ligand.
- CO3** –Identify the lead molecule by virtual screening.
- CO4** –Design the strategy towards drug design.
- CO5** –Design the strategy towards drug design based on the structural aspects of the macromolecule.
- CO6** –Adopt and implement the concepts of modelling to derive structures for molecules

Text / Reference books:

1. Andrew Leach - Molecular Modelling: Principles and Applications, 2nd Edition. Pearson Education EMA, 2001.
2. R.K. Prasad - Quantum Chemistry, 3rd Edition, 2006.

END SEMESTER EXAMINATION QUESTION PAPER PATTERN**Max. Marks : 100****Exam Duration : 3 Hrs.****PART A** : 10 questions of 2 marks each - No choice**20 Marks****PART B** : 2 questions from each unit of internal choice; each carrying 16 marks**80 Marks**

SBIA1502	DRUG DESIGN AND DEVELOPMENT	L	T	P	Credits	Total Marks
		3	1	0	4	100

COURSE OBJECTIVES

To understand the basics of drug design and to know the significance of pharmacophores in drug design.

UNIT I Introduction to Drugs**9hrs**

Drugs- Classification - Chemical basis of classification, Legal classification of drugs. Soft drugs and Hard drugs, Prodrugs - classification, Application

Unit II Drug Targets and Activity**9hrs**

Drug targets - Enzymes, Receptors, Proteins, Nucleic acids, Lipids. Drug solubility - Effect of pH, Effect of polarity and pKa. Drug Metabolism - Absorption, Distribution, Metabolism, Elimination. Pharmacokinetics and Pharmacokinetics of drugs

UNIT III Drug Discovery**9hrs**

Introduction to Drug Discovery and Development, stages of drug discovery, Natural sources of drugs, Computers in drug discovery, Rational approaches to lead discovery based on traditional medicine, ADME/T analysis, Oral bioavailability, Drug likeliness, Lipinski's rule.

Unit IV Drug Design**9hrs**

Types of drug design- Rational drug design, Structure based - Receptor based and ligand based. Virtual Screening, Hit to Lead identification, Lead optimisation. Pharmacophore analysis and mapping.

Unit V Docking and QSAR**9hrs**

Molecular docking - steps, types of docking, Pose identification, applications, software used. QSAR - origin & development, Physicochemical parameters, partition coefficient, Hammett's substitution constant, Taft steric constant, Hansch analysis, Free Wilson analysis, 3D QSAR methods CoMFA & CoMSIA.

Course Outcomes

At the end of the course the students will

- CO1** –realize the concept of drugs and their classification
- CO2** –explore the various targets of drugs and identify the factors influencing the activity
- CO3** –Comprehend the process of drug discovery.
- CO4** –Explore the drug design concept and they basics
- CO5** –Appreciate the significance of docking and QSAR in drug design.
- CO6** –Create an approach to design drugs based on the structural knowledge available

Text / Reference Books

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

END SEMESTER EXAMINATION QUESTION PAPER PATTERN**Max. Marks : 100****PART A** : 10 questions of 2 marks each - No choice**PART B** : 2 questions from each unit of internal choice; each carrying 16 marks**Exam Duration : 3 Hrs.****20 Marks****80 Marks**

SBIA1503	SYSTEMS BIOLOGY	L	T	P	Credits	Total Marks
		3	0	0	3	100

COURSE OBJECTIVES

To understand the basic concepts in Systems Biology and to apply them to real life problems in biology and medicine.

UNIT I Biological Basis**9hrs**

Introduction - systems understanding - what is systems biology? - Reductionism or holism? - Mechanism or emergentism? - Homeostasis or robustness? - What type of systems disease is cancer? - What are the challenges facing systems and cancer biology? - is systems biology a scientific revolution?

UNIT II Integrative Analysis of Omics Data**9hrs**

Integrative Analysis of Omics Data - Summary - Introduction - Omics Data and Their Measurement Platforms - Data Processing: Quality Assessment, Quantification, Normalization, and Statistical Analysis - Data Integration: From a List of Genes to Biological Meaning - Outlook and Perspectives

UNIT III Flux Analysis**9hrs**

Flux Analysis in Biotechnology and Medicine - Introduction - Theoretical Foundations of Emerging Challenges - MFA - Metabolic Flux Analysis in Biotechnology - Metabolic Flux Analysis in Medicine - Metabolic Modelling for Design of Cell Factories - Building and Refining Genome-Scale Metabolic Models - Strain Design Algorithms - Case Studies

UNIT III Genome-Scale Metabolic Modelling**9hrs**

Genome-Scale Metabolic Modelling and In silico Strain Design of Escherichia coli - The COBRA Approach - In silico Model-Based Strain Design of E. coli Cell Factories - Future Directions of Model-Guided Strain Design in E. coli References - Accelerating the Drug Development Pipeline with Genome-Scale Metabolic Network Reconstructions - Metabolic Reconstructions in the Drug Development Pipeline - The Human Reconstruction - Community Models - Personalized Medicine

UNIT IV Computational Modelling of Microbial Communities**9hrs**

Computational Modelling of Microbial Communities - Ecological Models - Genome-Scale Metabolic Models - Drug Targeting of the Human Microbiome - The Human Microbiome - Association of the Human Microbiome with Human Diseases - Drug Targeting of the Human Microbiome - Future Perspectives.

UNIT V Modelling Biological Processes**9hrs**

Signal Transduction - Toward Genome-Scale Models of Signal Transduction Networks - The Potential of Network Reconstruction - Information Transfer Networks - Approaches to Reconstruction of ITNs - Aging - Systems Biology of Aging - The Biology of Aging - The Mathematics of Aging - Future Challenges - Immune response - Modelling the Dynamics of the Immune Response - Background - Dynamics of NF- κ B Signalling - JAK/STAT Signalling

Course outcomes

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

- CO1** –To understand the basic concepts in Systems Biology
- CO2** –To learn to apply systems biology to omics data.
- CO3** –To appreciate the importance of flux balance analysis and its importance.
- CO4** –To understand and perform genome scale modelling.
- CO5** –To learn model microbial communities and their interactions
- CO6** –To model specific biological processes.

Text / Reference books

- Nielsen, Jens, and Stefan Hohmann. Systems Biology, John Wiley & Sons, Incorporated, 2017.
- Systems Biology, edited by Robert A. Meyers, John Wiley & Sons, Incorporated, 2012.
- Systems Biology: Philosophical Foundations, edited by Fred C. Boogerd, et al., Elsevier Science & Technology, 2007.
- Lodhi, Huma M., and Stephen H. Muggleton. Elements of Computational Systems Biology, John Wiley & Sons, Incorporated, 2010.
- Ingalls, Brian P. Mathematical Modelling in Systems Biology: An Introduction, MIT Press, 2013.

6. Juan, Hsueh-Fen. Systems Biology: Applications in Cancer-Related Research, edited by Hsuan-Cheng Huang, World Scientific Publishing Co Pte Ltd, 2012.
Young, Daniel L., and Seth Michelson. Systems Biology in Drug Discovery and Development, John Wiley & Sons, Incorporated, 2011.

END SEMESTER EXAMINATION QUESTION PAPER PATTERN**Max. Marks : 100****PART A :** 10 questions of 2 marks each - No choice**PART B :** 2 questions from each unit of internal choice; each carrying 16 marks**Exam Duration : 3 Hrs.****20 Marks****80 Marks**

SBIA1504	CHEMINFORMATICS	L	T	P	Credits	Total Marks
		3	0	0	3	100

COURSE OBJECTIVES

The course is designed to make the students realize the role of chemical molecules in biology with an emphasis on its physical parameters and conformations and their role in drug discovery.

UNIT I Introduction To Cheminformatics**9 Hrs.**

Introduction to Cheminformatics - History and evolution of Cheminformatics - Chemical representation - Sequence, 2D, 3D structure, Types of chemical representation - linear notation, tabular storage, graphical representation

UNIT II Chemical Databases**9 Hrs.**

Chemical Databases - Cambridge Structural Database CCDC CSD, Crystallographic Open Database COD Protein Data Bank PDB Ligand Explorer, Chempider.

UNIT III Combinatorial Chemistry**9 Hrs.**

Combinatorial chemistry: Introduction, Liquid Phase synthesis, Solid phase synthesis, Designing Combinatorial Synthesis

UNIT IV Library Design**9 Hrs.**

Combinatorial library design - Rational principles of compound selection for combinatorial library design and optimization approach, Modelling toxicity, Computer Assisted Synthesis design and structure based library design.

UNIT V Applications**9 Hrs.**

Application of cheminformatics - QSPR Drug design - Target identification and Validation, lead finding and optimization - Pharmacophore-Based Drug Design - Structure-Based Drug design

Course Outcomes

At the end of the course the students will

- CO1** –Describe the basics of cheminformatics and understand the representation of molecules
- CO2** –Access the various chemical databases that form the basis for designing and studying new molecules.
- CO3** –They will be able to appreciate the methods of synthesising new molecules.
- CO4** –screen various compounds in the combinatorial library for the identification of lead molecules
- CO5** –Analyse role of chemical molecules in drug design
- CO6** –Comprehend the role of cheminformatics in drug design

Text / Reference Books

1. Andrew R Leach, V J Gillet An Introduction to Chemoinformatics Springer 2007
2. Jurgen Bajorath Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery Humana Press 2010
3. Jurgen Bajorath Chemoinformatics for Drug Discovery Wiley 2013
4. Daria Bove Computational Chemistry: Theories, Methods and Applications Nova Science Publishers Inc, New York 2014

END SEMESTER EXAMINATION QUESTION PAPER PATTERN**Max. Marks : 100****Exam Duration : 3 Hrs.****PART A** : 10 questions of 2 marks each - No choice**20 Marks****PART B** : 2 questions from each unit of internal choice; each carrying 16 marks**80 Marks**

SBIA2502	CHEMINFORMATICS AND DRUG DESIGN LAB	L	T	P	Credits	Total Marks
		0	0	4	2	100

1. Visualization of Bio molecules - RASMOL, Pymol
2. Chemical databases
3. Drawing and Visualization of molecules - 2D and 3D - ChemSketch
4. Interconversion of file formats -OpenBabel
5. Molecular Modelling: Homology & Ab initio
6. SAVS - Validation of the Modelled Structure
7. ADMET properties
8. Energy Minimization
9. Interaction of protein with small molecules - Docking

SBIA2501	SYSTEMS BIOLOGY LAB	L	T	P	Credits	Total Marks
		0	0	4	2	100

SUGGESTED LIST OF EXPERIMENTS

1. Modelling and Visualization
2. E cell
3. Copasi
4. Construction and Analysis of gene expression networks
5. Mat lab I
6. Mat Lab II
7. Cell Designer
8. Cytoscape

SIXTH SEMESTER

SBIA1601	DATAMINING AND MACHINE LEARNING	L	T	P	Credits	Total Marks
		3	0	0	3	100

COURSE OBJECTIVES

The main objective is to understand various concepts of datamining and machine learning and implement various advanced mathematical, physical and statistical techniques that are being applied for analyzing information biological data.

Unit I Data warehousing

9hrs

Introduction - Operational data sources- data warehousing - Data warehousing design - Guidelines for data warehousing implementation - Data warehousing metadata An introduction to Genome databases - resources, file formats, functional annotations. Advance Database Systems and Applications, Data Mining Functionalities, Classification of Data Mining Systems, Major issues in Data Mining.

Unit II Data mining

9hrs

Data mining application - data mining techniques - data mining case studies- the future of data mining - data mining software - Association rules mining: Introduction- basics- task and a naïve algorithm- apriori algorithm - improve the efficient of the apriori algorithm - mining frequent pattern without candidate generation

Unit III Classification

9hrs

Classification : Introduction - decision tree - over fitting and pruning - DT rules- estimation predictive accuracy of classification methods - other evaluation criteria for classification method - classification software

Unit IV Cluster analysis

9hrs

cluster analysis - types of data - computing distances types of cluster analysis methods - partitioned methods - hierarchical methods - density based methods - dealing with large databases - quality and validity of cluster analysis methods - cluster analysis software.

Unit V Machine Learning

9hrs

Machine Learning Techniques - Introduction to Support Vector Machines and their applications, - Introduction to Neural Networks Classifying Samples from two populations using Multilayer Perceptron and back propagation - Using genetic algorithm and Perceptron for feature selection and supervised classification - Applications of ANN, SVM etc. Case studies on Identifying gene coding regions, Structure prediction, AI in healthcare

Course outcomes

At the end of the course the students will

- CO1** - Understand various data collection methods.
- CO2** - Comprehend the role of data mining in the field of biology
- CO3** - Approach data ware house and data mining as a process of data collection and processing.
- CO4** - Discover patterns from large amounts of data to analyze and extract patterns
- CO5** - Apply suitable methods of data analysis for a given problem.
- CO6** - Make predictions from the given data sets using classification and clustering algorithms

References

1. J. Han and M. Kamber, Jian Pei (2012) *Data mining: concepts and techniques*. 3rd edition, Elsevier
2. T. Hastie, R. Tibshirani, and J. Friedman (2001) *The Elements of Statistical Learning: data mining, inference and prediction*. Springer Verlag.
3. Data Mining Concepts and Techniques - Jiawei Han, Micheline Kamber, Academic Press Morgan Kaufman Publishers. 2006 Reference Books:
4. Data Mining: Practical machine learning tools Techniques with java implementation by Ian H.Witten, Eibe Frank, 2005.

5. Machine Learning and data mining in pattern recognition in third International conference MLDM, by Petra Perner and Azriel Rosenfield, Springer.2003
6. Witten and Frank, Data Mining: Practical Machine Learning Tools and Techniques by, Elsevier.

END SEMESTER EXAMINATION QUESTION PAPER PATTERN**Max. Marks : 100****Exam Duration : 3 Hrs.****PART A : 10 questions of 2 marks each - No choice****20 Marks****PART B : 2 questions from each unit of internal choice; each carrying 16 marks****80 Marks**

SBIA1602	MICROARRAY DATA ANALYSIS	L	T	P	Credits	Total Marks
		3	0	0	3	100

COURSE OBJECTIVES

This course gives the technical knowledge on Microarray techniques and data analysis DNA Microarray and its statistical analysis of RNA data Statistical computing and Statistical Genetics

UNIT I DNA Microarrays**9 hrs**

The Technical Foundations, Why are Microarray Important? What is a DNA Microarray?, Designing a Microarray Experiment-The Basic steps, Types..

UNIT II Microarray Databases**9 hrs**

NCBI and Microarray Data Management, GEO (Gene Expression Omnibus), MAML, The benefits of GEO and MAML, The Promise of Microarray Technology in Treating Disease

UNIT III Microarray Data Normalization**9 hrs**

Microarray Data Pre-processing, Data-Data normalization, Measuring Dissimilarity of Expression Pattern-Distance Motifs and Dissimilarity measures, Visualizing Microarray Data-Principal Component Analysis, Microarray Data.

UNIT IV Microarray Data Analysis**9 hrs**

K Means Clustering, Hierarchical Clustering, Self-Organization Maps (SOM), Identifying Genes: Expressed usually in a sample- Expressed significantly in population-Expressed differently in two populations, Classifying Samples from two populations using Multilayer Perceptron, Support Vector Machines and their applications, Using genetic algorithm and Perceptron for feature selection and supervised classification.

UNIT V Microarray Applications**9 hrs**

Gene Ontology and pathway analysis, Promoter analysis and gene regulatory network, Co expression analysis, CGH & Genotyping chips, Chromosome aberration and polymorphism via genome-wide scanning, Future direction of microarray approach, Pharmacogenomics, Toxicogenomics, Data mining.

Course outcomes

At the end of the course the students will

- CO1** –To familiarize students with the DNA microarray technology and with the analysis of gene expression data
- CO2** –To ensure familiarity with the different microarray techniques and their everyday use in a biological laboratory setting.
- CO3** –To present some of the important low level analysis (Normalization, quality control, etc.).
- CO4** –To provide a sound introduction to a good experimental design.
- CO5** –To enable the students to access and analyse microarray data
- CO6** – To understand the applications of microarray analysis in various fields of research

TEXT / REFERENCES BOOKS

1. Arun Jogota, "Microarray Data Analysis and Visualization", The Bay Press, 2001.
2. Ernst Wit and John McClure, "Statistics for Microarrays Design", Analysis and Inference, John Wiley & Sons, 2004.
3. Steen Knudsen, "Guide to analysis of DNA Microarray data", John Wiley & Sons, 2004.
4. Dov Stekel, "Microarray Bioinformatics", Cambridge University Press, 2003.
5. Uwe R. Müller, Dan V. Nicolau, "Microarray Technology and Its Applications", Springer, 2005.
6. Emanuele de Rinaldis, Armin Lahm, "DNA Microarrays: Current applications", Horizon Scientific Press, 2007

END SEMESTER EXAMINATION QUESTION PAPER PATTERN**Max. Marks : 100****PART A :** 10 questions of 2 marks each - No choice**PART B :** 2 questions from each unit of internal choice; each carrying 16 marks**Exam Duration : 3 Hrs.****20 Marks****80 Marks**