

**Revised Course Structure and Syllabus of  
M.Sc. Bioinformatics Program  
(2016)**



**CENTER OF BIOINFORMATICS  
University of Allahabad  
Allahabad**

**UNIVERSITY OF ALLAHABAD**  
**CENTRE OF BIOINFORMATICS**

**Programme: M. Sc. Bioinformatics**

**Duration: 2 Years (4 Semester)**

**Total Credit: 89 (Semester Wise: 21 + 21 + 21 + 26 = 89)**

<b>Paper Code</b>	<b>Paper Name</b>	<b>Credit</b>
<b>SEMESTER – I</b>		
BIF501	Biochemistry and Biophysics	3
BIF502	Genomics and Proteomics	3
BIF503	Mathematics and Statistics	3
BIF504	Scripting Languages for Bioinformatics	3
BIF505	Sequence Analysis and Molecular Phylogeny	3
BIF531	<b>LAB – Biological Databases and Sequence Analysis</b>	3
BIF532	<b>LAB – Programming in PERL &amp; MATLAB</b>	3
<b>TOTAL</b>		<b>21</b>
<b>SEMESTER – II</b>		
BIF511	Algorithms for Bioinformatics	3
BIF512	Biological Networks and Systems Biology	3
BIF513	Biomolecular Structure, Modeling & Simulation	3
BIF514	Datamining and Warehousing of Biological Data	3
BIF515	Transcriptomics and Metabolomics	3
BIF533	<b>LAB – Biomolecular Modeling and Simulation</b>	3
BIF534	<b>LAB – JAVA and MySQL</b>	3
<b>TOTAL</b>		<b>21</b>
<b>SEMESTER – III</b>		
BIF521	Applied Bioinformatics	4
BIF522	Computer Aided Drug Designing (CADD)	4
BIF523	DNA Computing	4
BIF535	<b>LAB – Chemoinformatics &amp; CADD</b>	3
BIF536	<b>Mini Project (Synopsis-1+LiteratureReview-2+Final Assessment-3)</b>	6
<b>TOTAL</b>		<b>21</b>
<b>SEMESTER – IV</b>		
BIF537	<b>Major Project</b>	<b>26</b>

# Biochemistry and Biophysics

## Unit-I

**Water:** Hydrogen bonding and structure of water molecule, Ionization of water, pH and colligative properties of water.

**Carbohydrates:** general classification, Polysaccharides and proteoglycans: Starch, glycogen, cellulose, chitin & bacterial cell wall. Glycosaminoglycans and proteoglycans in extracellular matrix.

**Lipids** - Glycerophospholipids, sphingolipids, gangliosides, Eicosanoids & prostaglandins. Cholesterol and its biosynthesis.

**Proteins & amino acids** - Amino acids and their physio-chemical properties, peptide bond and peptides, disulphide cross links, various levels of structural organization of Proteins. Ramachandran plot, Alpha-helix, Beta sheet, Helix-coil transitions.

## Unit-II

**Enzymes:** Units of activity, coenzymes and metal cofactors, temperature and pH effects, Michaelis-Menten kinetics, inhibitors and activators, active site

**Organization of metabolic systems:** enzyme chains, multi-enzyme complexes, multifunctional enzymes and regulatory enzymes.

**Nucleic acids:** nucleosides, nucleotides, RNA and DNA. Nucleic acid conformation, A-DNA, B-DNA, Z-DNA and C-DNA, their geometrical and structural features, Types of base pairing – Watson-Crick and Hoogsteen. Denaturation and renaturation of DNA. Replication, Transcription and Translation process

## Unit III

**Mechanics:** Mechanics of particles and system of particles (conservation of energy, linear momentum, and angular momentum), Lagrangian and Hamiltonian, two body problems, Mechanics of small oscillations.

**Electromagnetism:** Electric charge and its properties, Coulomb's law, Electric field, potential, and superposition principles, Gauss's Law, Capacitance, Magnetic fields, Lorentz force, Ampère's laws, Electromotive force, Electromagnetic induction; Faraday and Lenz's laws, Forces on charge in Electric and Magnetic field.

## Unit IV

**Thermodynamics:** Concept of thermodynamic state, Heat and Work, Internal energy function and First law of thermodynamics, Entropy and Second law of thermodynamics, Entropy maximum and energy minimum principles, Thermodynamics free energies, First order phase transitions, and Triple point; Heat equation and separation of Variables.

## Unit V

**Properties of Matter:** Transport properties: effusion, diffusion, viscosity and conductivity, *van-der-Waals* equation, Structure of liquids, Surface energy, surface tension and capillary action, Flow properties and viscosity, Elasticity; Young's, shear and bulk modulus, Crystal structure and symmetry, X-ray diffraction and *Bragg's* law, Photoelectric effect.

### Reference Books:

- Principles of Biochemistry : A.L. Lehninger, Nelson and Cox, McMillan Worth Publishers.
- Biochemistry : Voet and Voet, John Wiley and Sons, Inc. USA.
- Classical Mechanics by Gupta, Kumar and Sharma
- Thermodynamics by A.B. Gupta & H.P. Ray
- Introduction to Electrodynamics by David J. Griffiths
- Mathur, D.S. Elements of Properties of Matter, S. Chand & Company Limited, 1967

## Mathematics & Statistics

### Unit I

**Linear Algebra:** Matrices and Determinants, Minors and cofactors, Eigen values and Eigen vectors, Series (AP and GP), Limits, Logarithms.

**Vector Algebra:** Vector and Scalar, Dot and cross product, Vector differentiation, Gradient, divergent and curl, vector integrations (line integral, surface integral and volume integral).

### Unit II

**Calculus:** Differential Calculus, Maxima and minima, Integral Calculus, Ordinary Differential Equations (First order).

### Unit III

**Geometry:** Coordinate geometry, Straight line, Circle, Parabola, Ellipse, Hyperbola, Polar co-ordinates, Sphere.

### Unit IV

**Data Visualization and Descriptive Statistics:** Concept of Statistical population and sample, quantitative and qualitative data, ordinal and nominal data, time series and cross sectional data, bivariate data and multivariate data, Frequency distributions, cumulative frequency distributions and their graphical representation, histogram, frequency polygon, Box plot, scatter plot for bivariate data. Measures of central tendency- Mean, Median, Mode. Measures of dispersion- range, mean deviation, variance, standard deviation, skewness and kurtosis.

**Bivariate data:** Correlation and regression, method of least squares for fitting a regression line, Multiple linear regressions.

### Unit V

**Probability theory and probability distributions:** Prediction & Estimation; Concepts of random experiment, sample space and events, definition of probability and some elementary results of probability, conditional probability and Bayes theorem, random variable, probability mass function and probability distribution function, cumulative distribution function, Binomial distribution, Multinomial distribution, Poisson distribution, Normal or Gaussian distribution

**Hypothesis testing:** simple and composite hypothesis, null and alternative hypothesis, level of significance and power of a test, applications of t-test, Chi-square test and F-test.

### Reference books:

- Statistics (Schaum's Outline, Fourth Edition); by Murray Spiegel and Larry Stephens
- Probability and Statistics (Schaum's Outline) by Murray Spiegel, John Schiller, R. Alu Srinivasan, Debasree Goswami
- Theory and Problems of Probability, Random Variables, and Random Processes (Schaum's Outline); by Hwei P. Hsu
- Bayesian Statistics: An introduction; by Peter M. Lee
- Vector Analysis by M. R. Spiegel
- Statistical Methods in Bioinformatics: An Introduction; by Ewens and Grant

# Genomics and Proteomics

## Unit I

**Structural Genomics:** Genome Organization (intron, exon, promotor, intergenic region, ORF) Genome Sequencing: Maxam-Gilbert Method, Sanger Methods, Pyro-sequencing. Next Generation sequencing methods (NGS)- Illumina sequencing, 454 sequencing, Ion Torrent. NGS-file formats.

Genome Assembly and annotation

The Human Genome Project, Ethical issues in human Genome Research.

## Unit-II

Codon biasing, Codon-Usage and gene expression, C-Value Paradox, Exon shuffling, genome rearrangement

Prediction of genes, promoters, splices sites, regulatory regions: basic principles, application of methods to prokaryotic and eukaryotic genomes and interpretation of results.

DNA Polymorphism, Types of DNA Polymorphism, Single Nucleotide Polymorphism (SNPs), The nature of SNPs, mining of SNPs, distribution of SNPs, Applications of SNP technology. HapMap Project. Role of SNPs in Pharmacogenomics. Meta-genomics.

## Unit-III

**Comparative genomics:** Basic concepts and applications, BLAST2, MegaBlast algorithms, PipMaker, AVID, Vista, MUMmer, applications of suffix tree in comparative genomics, synteny and gene order comparisons

Comparative genomics databases: COG, VOG

**Functional genomics:** Application of sequence based and structure-based approaches to assignment of gene functions –e.g. sequence comparison, structure analysis (especially active sites, binding sites) and comparison, pattern identification, etc. Use of various derived databases in function assignment, use of SNPs for identification of genetic traits.

Gene/Protein function prediction using Machine learning tools viz. Neural network, SVM etc.

## Unit-IV

Principles of protein structure; anatomy of proteins – Hierarchical organization of protein structure – Primary. Secondary, Super secondary, Tertiary and Quaternary structure; Hydrophobicity of amino acids, Pacing of protein structure, van der Waal and Solvent accessible surface, Internal coordinates of proteins; Ramachandran Map, protein folding.

Peptide Sequencing, Protein expression analysis: 2D-PAGE, Protein microarray, Mass spectrometry, western blotting

## Unit-V

Molecular interactions: Protein-Protein interactions, Protein-DNA interactions. Methods to predict molecular interactions: Y2H method, Phage-Display method, Phylogenetic footprinting, Gene fusion method, Protein profiling, Molecular-Docking.

Database, Server and tools for analysis of protein-protein interaction, Docking.

Tools of analyzing Proteomics data (ExpASY server) and GCG utilities and EMBOSS

### Reference books:

- Principles of Biochemistry: A.L. Lehninger, Nelson and Cox, McMillan Worth Publishers.
- Introduction to Genomics by Arthur M. Leask, Oxford University Press
- Genes series, by Benjamin Lewin, Oxford University Press

## Scripting Languages for Bioinformatics

### Unit -I

Block diagram of computer, Its components and functions, Representation of data (integer, real, character, boolean), Number System, Introductory Boolean algebra, Concept of program, Programming languages, Introduction to operating systems; Linux OS, compilers, interpreters, Algorithms and flowcharts. Introduction to Parallel Computing

### Unit-II

Scalar data Numbers, strings, scalar operators, scalar variables, scalar operators and functions

Introduction to Data Structures: Lists, Arrays, Stack and Queue

Hashes: What is a Hash? Hash variables, Literal representation of a Hash, Hash Functions, Hash Slices

Basic Programming constructs – Sequence, Alteration, Indexed Alteration and Iteration

Document Object Model; Accessing different objects of HTML page; Programming using JavaScript, XML, ASP, CGI, PERL & PYTHON – an overview; Database connectivity using PHP and Data Retrieval; Introduction to DBMS.

### Unit-III

File handles and file tests,

File and directory manipulation

Perl language: Directories, subroutines, references, packages, libraries, modules, classes, objects

Bio-perl: Introduction, Installation procedures, Architecture, Uses of bioperl

Handling of bioinformatics data

### Unit-IV

Python language and syntax

Python data types: Simple values, Python expression: Numeric operators, Logical operators, String operators, compound expression. Name, function and modules in Python. Python's collections: set, sequences, mappings, strings, collection related expression features.

Control statements: conditional statement, loop. Classes-defining classes, class and method relationship, Python utilities.

### Unit-V

File management: reading/writing file in python

Pattern matching and regular expression: Fixed length matching, variable length matching, greedy versus non greedy matching,

Modules and Packages in python.

Introduction to Biopython: working with sequence, Parsing biological file format, connecting with biological databases; Web Programming using CGI interface.

### Reference Books:

- o Sinha P.K Computer Fundamentals, BPB publication, New Delhi
- o James D. Tisdall, Beginning Perl for Bioinformatics, 2001, O'REILLY
- o James D. Tisdall, Mastering Perl for Bioinformatics, 2003, O'REILLY
- o Bioinformatics programming: using python. Mitchel L. Model, 2009, O'REILLY

# Sequence analysis and Molecular phylogeny

## Unit-I

Various file formats for bio-molecular sequences: GenBank, FASTA, GCG, MSF etc

Basic concepts of sequence similarity, identity and homology, definitions of homologues, orthologues, paralogues and xenologues

Pairwise sequence alignment: Local and Global alignment concepts, Dynamic programming methodology- Needleman-Wunsch algorithm, Smith-Waterman algorithm, Substitution Scores and Gap penalties. Statistics of alignment score

Database Searches: Keyword-based Entrez and SRS; Sequence-based: BLAST and FASTA; Use of these methods for sequence analysis including the on-line use of the tools and interpretation of results from various sequence and structural as well as bibliographic databases.

Scoring matrices: basic concept of a scoring matrix, Matrices for nucleic acid and proteins sequences, PAM and BLOSUM series, principles based on which these matrices are derived

## Unit-II

Multiple sequence alignments (MSA): the need for MSA, basic concepts of various approaches for MSA (e.g. progressive, hierarchical etc.). Algorithm of CLUSTALW and PileUp and their application for sequence analysis (including interpretation of results), concept of dendrogram and its interpretation

Sequence patterns and profiles: Basic concept and definition of sequence patterns, motifs and profiles, various types of pattern representations viz. consensus, regular expression (Prosite-type) and sequence profiles; profile-based database searches using PSI-BLAST, HMMer, analysis and interpretation of profile-based searches

Computational method for promoter recognition: finding transcription factor binding sites, identifying motifs with unaligned sequences, Predicting transcriptional start sites (TSS)

Computational methods for sequence assembly

## Unit III

The concept of evolutionary tree, terminology of phylogenetics, definition of Dendrogram, Cladograms, & Phylograms, Types of phylogenetic trees (rooted vs. unrooted trees), gene tree & Species tree, Homologs, Orthologs, & Paralogs, Newick format of tree representation, True tree & Inferred tree.

## Unit IV

**Models of sequence evolution:** Introduction to evolutionary models. Models of Nucleotide substitutions: JC, K2P, F81, HKY85, TN93, GTR, JTT, Dayhoff.

**Phylogenetic analysis algorithms:** UPGMA, Neighbors-Relation, Neighbor-Joining, Least Square, Minimum Evolution, Maximum Parsimony, Maximum likelihood, and Bayesian Inference. Statistical assessments of phylogenetic methods (Consistency, Efficiency, Robustness, & Computational speed).

## Unit V

Tree evaluation methods: Bootstrapping, Randomized and jack-knifing methods,

Molecular Clock, Rate heterogeneity, LBA problem, Phylogenomics, Supertree & Super matrix method, Phylogenetic analysis software: PHYLIP, MEGA, PhyML, and MrBayes.

### Reference Books:

- Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids by Durbin et al., Cambridge University Press
- Molecular Evolution and Phylogenetics Masatoshi Nei and Sudhir Kumar
- Fundamentals of Molecular Evolution by Dan Gaur
- Inferring Phylogenies by Joseph Felsenstein
- Molecular Evolution: A Statistical Approach by Ziheng Yang
- Evolutionary Genomics by Naruya Saitou

## Data mining and warehousing of biological data

### Unit-I

Data Mining: Definition, Data mining task, Data mining process, architecture and component of a typical data mining system, Mining frequent pattern, associations and correlation: Pattern mining and Interestingness of pattern, Association Rules (mining multilevel association rules, multidimensional association rules, quantitative association rules, constraint based association mining).

Classification and Prediction: Decision tree induction, Bayesian classification, Rule based classification, HMM, ANN based classification (back-propagation), Support vector machines (SVM), associative classification

### Unit-II

Clustering: Portioning method, Hierarchical method, Density-Based Methods, Grid-Based Methods, Model-Based Clustering Methods, Clustering High-Dimensional Data, Constraint-Based Cluster Analysis, Outlier Analysis.

Mining Stream, Time-Series, and Sequence Data.

Mining Complex type of data: Sequence Mining, Web Mining, Text Mining, Spatial Data Mining.

Trends in data mining, Issues in data mining-Security and social issues, User interface issues, Performance issues, Data source issues

Integration of a Data Mining System with Data Warehouse System:

### Unit III

Data Warehousing: Basic Concepts, Data Warehouse Architecture, Benefits of a data warehouse, Three-tier Decision Support Systems (DSS), DataMart, Online Analytical Processing (OLAP) Engine, OLAP Servers (ROLAP, MOLAP, HOPAP), Multidimensional Data Model, Data Cube, Warehouse schema (Star schema, Snowflake schema); Enterprise Warehouse, Virtual Data Warehouse; Metadata; Data Preprocessing, Data Warehouse Design and Usage

### Unit IV

Application of Data Mining in Biodata analysis: DNA/protein sequence Analysis, Genome analysis, Protein Structure Analysis, Pathway analysis, microarray data analysis, annotation, gene ontology, gene mapping

Introduction to biological database: Designing of biological databases, Types of biological database: Primary database, Secondary database, Composite database.

Biological data mining tools: Entrez, Blast, sequence retrieval system (SRS)

### Unit V

Relational database management system (RDBMS), sequence query language (MySQL)- Overview, Tables, Queries, creating and using database

Design, implementation and updating of bioinformatics knowledge bases: Sources of data, design of knowledge bases, implementation of knowledge bases, updating of knowledge bases

Warehousing principal of NCBI taxonomy, Gene, Literature Databases and Gene Expression Omnibus (GEO). Warehousing architecture of UCSC Genome browser, miRNA data base, KEEG pathway database

### Reference Books:

- Data Mining: Concepts and Techniques by Jiawei Han and Micheline Kamber, 2000
- Data Mining Techniques, A. K. Pujari, University Press, Hyderabad, 2006
- Mount, D. W.: Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor. CSHL Press, 2001.
- Data mining in bioinformatics by Wang et al, Springer-Verlag, 2005



# Transcriptomics and Metabolomics

## Unit-I:

Transcription: Preprocessing of RNA, mRNA, Transcriptome and Transcriptomics, The Insights of Transcriptomics (mRNA regulation)

Types and function of RNA: rRNA, tRNA, mRNA, siRNA, miRNA, RNA interference (RNAi), RNA-induced silencing complex (RISC), Biogenesis of miRNA and siRNA.

Computational prediction of- miRNA genes and miRNA targets. Bioinformatics of siRNA designing

Transcriptome Project (Human, Mouse, Cancer, Fungal).

## Unit-II

EST-expressed sequence tags, EST Generation, EST data quality, EST Clustering and Assembly, EST clustering systems and algorithm -TIGR-Assembler, Gene Indices, STACK-Pack.

ESTs and gene discovery, ESTs and sequence polymorphisms, SNP-prediction. EST database and web tools for ESTs project.

Serial analysis of gene expression (SAGE), SAGE data analysis, role of SAGE in gene discovery, Tissue Specific Transcriptomics and Expression Pattern Analysis. SAGE Bioinformatics. Virtual-SAGE.

## Unit-III

Genome Wide Gene Expression Analysis: cDNA-Microarrays - Technique of Micro array, Micro array design, Data representation, Data cleaning methods, Data Normalization methods, Analysis of Microarray data using, K-Means Clustering, Hierarchical Clustering, Self Organizing Maps (SOM), Principle Component Analysis (PCA), CLICK algorithm. Gene co-expression and co-regulation, Application of Micro array. Statistical test for differential expression in cDNA microarray experiments. Missing value estimation methods

## Unit-IV

Transcriptomics in functional genomics, Transcriptomics and Disorders, Transcriptomics in drug design, Transcriptomics in Human cancer hazard assessment, Transcriptomics and Phylogenetics applications, Impact of transcriptomics on Pharmaceutical Research.

Microarray databases (GEO, Array express etc.), data-file formats, Tools for Transcriptomics and Transcriptome Analysis, Bioconductor.

## Unit-V:

Metabolome and Metabolomics, Metabolic profiling and fingerprinting, Metabolic pathway analysis and metabolic networks, Single Cell Metabolomics, Metabotype Concept.

Computational Methods to Interpret and Integrate Metabolomic Data, Metabolomics data processing workflow, Chemical ontologies, Online metabolic databases (Human Metabolome Databases, KEGG, BioCyc) and pipelines.

Applications of Metabolomics: Metabolic Pathway as a target for Drug-screening, Metabolomics approach for hazard identification in human health assessment of environmental chemicals, Clinical implications of Metabolomics.

Plant metabolomics.

## Reference books:

- Transcriptomics: Expression pattern analysis. by Gomase, Virendra.
- Metabolomics, by Ute Roessner, ISBN 978-953-51-0046-1, Hard cover, 364 pages, Publisher: InTech, Published
- Microarray Analysis, by Mark Schena, Publisher: Wiley-Liss

# Algorithms in Bioinformatics

## Unit I

**Introduction:** Algorithms in Computing; Analyzing algorithms-Asymptotic notation, Standard notations, Big 'O' notations; Algorithm design techniques- Exhaustive Search, Branch-and-Bound Algorithms, Greedy Algorithms, Dynamic Programming, Divide-and-Conquer Algorithms, Machine Learning, Randomized Algorithms; Time and space complexity of algorithms, common Sort and Search algorithms

## Unit II

**Dynamic Programming Algorithms:** Needleman-Wunsch algorithm (Global Alignment Algorithm), Smith–Waterman algorithm (Local Alignment Algorithm), **Exhaustive Search-** Restriction Mapping, Finding Motifs; **Greedy Algorithms-** Genome Rearrangements, Sorting by Reversals, Finding Motifs; **Divide-and-Conquer Algorithms-** Divide-and-Conquer Approach to Sorting, Space-Efficient Sequence Alignment, Block Alignment. Clique detection algorithm. Ant colony optimization algorithm

## Unit III

**Combinatorial Pattern Matching-** Hash Tables, Repeat Finding, Exact Pattern Matching; **Expectation and Maximization (EM)** with forward and backward algorithms, discriminative learning; **Genetic Algorithm:** Basic Concepts, Reproduction, Cross over, Mutation, Fitness Value, Optimization using GAs; Applications in bioinformatics. Particle swarm algorithm

## Unit IV

**Hidden Markov Models:** Markov processes and Markov Models, Hidden Markov Models, Parameter estimation for HMMs, Optimal model construction, Applications of HMMs; **Artificial Neural Networks:** Historic evolution – Perceptron, NN Architecture, supervised and unsupervised learning, Back Propagation Algorithm, Training and Testing, Self-organizing Feature Map and Radial Basis Function Network; Overview of Support Vector Machines

## Unit V

**Structural alignment Algorithms:** DALI, Combinatorial extension, SSAP; **RNA Structure Prediction Algorithms:** Zuker & Nussinov folding Algorithm; **Spectral Analysis Algorithm:** Fourier analysis; **Graph Layout Algorithms:** Forced direct algorithm, Simulated annealing algorithm

### Reference Books:

- Fundamentals of Computer Algorithms by Horowitz, S. Sahni, and Rajasekharan. Galgotia Publications. 1984
- An introduction to bioinformatics algorithms by Neil C. Jones, Pavel Pevzner. MIT Press. 2004
- Biological Sequence Analysis by Richard Durbin, Sean Eddy, Anders Krogh. Cambridge. 1998
- Algorithms in Bioinformatics: A Practical Introduction by Wing-Kin Sung, Chapman and Hall. 2009.

# Biomolecular Structure, Modeling and Simulation

## Unit-1

**Nuclear Magnetic Resonance:** Basic principles of NMR, chemical shift, The Nuclear Overhauser Effect (NOE), Correlation Spectroscopy (COSY), Nuclear Overhauser Effect Spectroscopy (NOESY), Protein 3D structure determination using NMR.

**X-ray Crystallography:** Principles, Bragg's Law, Protein Crystallization Techniques

Protein data bank (PDB), PDB-file format (ATOM, HETATM, coordinates etc.)

## Unit-II

RNA secondary structure prediction: combinatorial approach, energy minimization method, phylogenetic comparative methods  
Protein Secondary structure prediction methods - Chou Fasman, Garnier-Osguthorpe-Robson (GOR), Neural network based methods-PHD, PSIPRED

Comparative modeling: Homology modeling, Model evaluation and optimization methods, Model validation methods.

Protein threading and fold recognition, score functions in threading. Rotamer libraries.

*Ab-initio* method for protein structure prediction: Introduction, Energy functions, Conformational search methods, model selection.

Critical Assessment of Structure Prediction (CASP). Protein Contact Maps, Protein conformation and folding problems: Theory of Protein folding and misfolding, folding problems, folding kinetics, folding mechanism.

## Unit-III

Molecular Mechanics: Underlying principle, Model and energy formulation.

Types of Potentials- Lennard-Jones, Truncated Lennard-jones, Exponential-6, Ionic and Polar potentials.

Types of Force Fields: AMBER, CHARMM, Merck Molecular Force Field, Consistent Force Field, MM2, MM3 and MM4 force fields.

Potential Energy Surface, Energy minimization and related methods for exploring the Energy surface

Optimization: multivariable Optimization Algorithms, level Sets, Level Curves, Gradients, Optimization Criteria, Unidirectional Search, Finding Minimum Point, Gradient based Methods-Steepest Descent and Conjugate Gradient Methods.

## Unit-IV

**Computer Simulation Methods:** Introduction, Calculations of Simple Thermodynamics properties, Phase space, Practical aspects of computer simulation, analyzing the results of simulation and estimating errors.

**Molecular Dynamics Simulation:** Introduction, Radial distribution functions, Pair Correlation function, Newtonian dynamics, Integrators -Verlet algorithm, Potential truncation and shifted-force potentials, Implicit and explicit Solvation models, Periodic boundary conditions, Temperature and pressure control in molecular dynamics simulations. Molecular dynamics-Ensembles

Introduction to semi-empirical, molecular mechanics and *ab initio* techniques, potential energy surfaces

## Unit-V

Monte Carlo Simulation Methods: Introduction, Theory of Metropolis Method, Monte-carlo algorithm, Implementation of the Metropolis Monte-Carlo method, Monte-Carlo Simulation of molecules. Monte-Carlo simulation for protein folding.

Application of Molecular simulations - Protein folding modeling, substrate receptor interactions modeling

Bioinformatics tools for simulation: Gromacs, VMD, AMBER, CHARMM

Classification of 3D structure of proteins (SCOP, CATH, HSSP, FSSP), Structure comparison of proteins (DALI, VAST)

### Reference books:

- Molecular Modelling and Simulation: An Interdisciplinary Guide. An Interdisciplinary Guide. Authors: Schlick, Tamar (Publisher: Springer-Verlag New York)
- Molecular modeling: Principles and application. Author: Andrew R. Leach (Publisher: Pearson)

# Biological Networks & Systems Biology

## Unit I

Introduction to Graph & Network, Types of networks (small world, random, scale-free networks, and Hierarchical networks), Introduction to biological networks, Importance of biological networks, Types of biological networks

**Network parameters:** Node degree, Node degree distribution, Scale-free networks and the degree exponent, Shortest path, Mean path length, Clustering coefficient, Node centrality and network centrality

**Network decomposition:** Subgraphs, Motifs, Motif clusters, and Modules

**Network robustness:** Topological, Functional and dynamical robustness

## Unit- II

Protein interaction network, Types of Protein-Protein interactions (PPI): Stable, Transient, Physical, and Genetic interactions, **Growth mechanism of biological networks**, Prediction of Protein-Protein interactions: Experimental methods, Computational methods- Phylogenetics profiling method, Interolog & regulog method, Co-expression method,

Database of biological networks (STRING, BioGRID, STITCH and KEEG), Designing of network circuitry (CYTOSCAPE), Network layouts

## Unit III

Gene Regulatory network (GRN): Methods for regulatory network reconstruction, GRN model: Ordinary Differential Equation Model (ODE), Boolean network model, Probabilistic Boolean Network model (PBN), Bayesian network model.

Multi-layer hierarchical structure of regulatory network, Network motifs, integrated network.

Metabolic Network, Signaling networks and their identification methods

## Unit IV

Biological network comparison: Pairwise network alignment, Multiple Network alignment, Network querying. Software and tools for network structure analysis (Network Analyzer).

Modeling and Simulation of Biological Systems (Ordinary Differential Equation Model, Stochastic Model

Flux Balance Analysis, Pathway analysis, Biochemical and metabolic pathways, Cell Designer

## Unit V

Methods in system Biology: Interaction based method, Construction based methods, and Mechanism based methods.

Visual representations and notations for systems biology, Metabolic Pathway visualization and editing software (MyBioNet, MetaViz, Cytoscape).

System Biology Informatics Infrastructure (SBII), Future for system Biology. Standards for Computational System Biology. Synthetic biology and artificial gene circuits.

### Reference books:

- Principles of Computational Cell Biology By Volkhard Helms
- Protein-protein Interactions and Networks: Identification, Computer Analysis, and Prediction (Computational Biology) Anna Panchenko (Editor), Teresa M. Przytycka (Editor)
- An Introduction to Systems Biology: Design Principles of Biological Circuits (Chapman & Hall/CRC Mathematical & Computational Biology) By Uri Alon (Author)

## Computer Aided-Drug designing (CADD)

### Unit-I

Introduction to Drug discovery: Drug Discovery methods, Basic Steps of drug designing process, Drug target identification, clinical Trials,

Pharmacogenomics: SNP and Pharmacogenomics (CYP450, TPMT), The Multi Drug Resistance proteins (MDR, P-gp), VIP genes, PharmGKB

Pharmacokinetics: Fundamentals of Pharmacokinetics, Absorption- bio-availability, Distribution-volume of distribution, drug barriers, Metabolism, Elimination-drug elimination kinetics

Pharmacodynamics: Drugs affinity and efficacy, Dose response relationship, Response of Drug on receptor: Agonist, antagonist. Quantification of drug safety: Effectiveness, Toxicity, Lethality, Therapeutic Index

Impact of pharmacogenomics, pharmacokinetics and pharmacodynamics on drug discovery and development process

### Unit-II

Representation and manipulation of 2D and 3D molecular structure, Molecular descriptors: Based on 2D structure-Physiochemical properties, Molar refractivity, topological indices, 2D fingerprints, BCUT descriptors. Based on 3D structure-3D fragment screens, Pharmacophore keys, other 3D descriptors. Data verification and manipulation.

QSAR: Deriving a QSAR equation, Designing of QSAR experiment, Principle component regression, Partial least square (PLS). Molecular field analysis and PLS. QSAR models: Free Wilson analysis, Hansch constant, Hammett Substituent Constant

3D QSAR: common 3D molecular field, Designing of 3D QSAR experiment, 3D-QSAR model- Comparative Molecular Field Analysis (CoMFA), Comparative Molecular Similarity Index Analysis (CoMSIA)

### Unit-III

Pharmacophore: Pharmacophore modeling, Pharmacophore mapping- constrained systemic search approach, Clique detection, Maximum likelihood approach, Genetic algorithm based approach,

Pharmacophore generation: Ligand based pharmacophore generation (Hiphop and hypogen theories), receptor based pharmacophore generation. Pharmacophore fingerprints.

Pharmacophore applications: Pharmacophore searching, Pharmacophore based De-novo design of ligands, Pharmacophore based Lead optimization, Pharmacophore in 3D QSAR

### Unit-IV

Drug-likeness and compound filters- Lipinski's 'rule of five'. Virtual screening: Ligand based virtual screening (LBSV)- Pharmacophore based screening, Descriptor based screening

Structure based virtual screening (SBVS) -Docking: Introduction, basic steps of docking, Types of docking-Rigid and flexible docking

Search algorithm: Random searching (GA, Tabu search, Monte-Carlo), systematic searching (incremental construction), simulation based searching (Monte Carlo, Simulated annealing), shape matching algorithm

Scoring/evaluation functions- Force-Field Based scoring, Empirical Scoring Function, Knowledge-Based scoring function, Consensus Scoring

### Unit-V

Prediction of ADMET properties: Hydrogen bonding descriptors, Polar surface area, Descriptors based on 3D fields, Toxicity prediction,

Introduction to combinatorial chemistry and Chemoinformatics, Combinatorial libraries: Diverse and focused libraries, Library enumeration, combinatorial library design strategies, Approaches to product based library design, multi-objective library design

Bioactive chemical database: DrugBank, PubChem, KEGG DRUG

### Reference books:

- Bioinformatics-from Genomes to drugs- Thomas Lengauer.
- Molecular Modeling-Principles and applications-Andrew R. Leach.
- Fundamentals of Medicinal Chemistry by Gareth Thomas.
- Pharmacogenomics: An Approach to New Drug Development-Chiranjib Chakraborty, Atanu Bhattacharyya.
- An introduction to chemoinformatics- Andrew R. Leach and Valerie J. Gillet (springer)

# Applied Bioinformatics

## Unit I

Role of bioinformatics in human genome project: Handling and maintenance of genomic data, genome assembly, tagging, analysis-visualization and interpretation of genomic data.

Computational resources for biotechnology: Primer designing, Probe designing, Vector screening, siRNA designing, e-PCR, site directed mutagenesis, virtual-SAGE, translated protein sequence, handling and distribution of biological data-NCBI

Agriculture Bioinformatics: Identification of biotic-abiotic stress related gene, screening of genomic response against pesticides and herbicides test

## Unit II

Genome-wide gene expression analysis: Identification of co-expressed and coregulated genes expression and its role in function, gene promoter, orthologus prediction and pathway analysis

Statistical assessment of functional categories of genes deregulated in pathological condition using microarray data

Use of Bioinformatics in modern clinical trial analysis; Clinical Bioinformatics, management of clinical data, Translational Bioinformatics; Synthetic Biology: promises and challenges; Biological Big Data Analysis.

## Unit III

Bioinformatics for genomic studies: Genome assessment, genome organization, micro-mini satellite identification, development of synteny map, Functional Diversity identification, Functional Analysis of Inter- genic Regions, Horizontal-Vertical gene transfer, genetic polymorphism, evolutionary relationship

Structural bioinformatics: Secondary structure prediction of protein and RNAs, structural classification of protein and RNAs, Protein folding problems, 3D structure modeling of biomolecules, Structure Comparison Methods, Protein-ligand interaction (Docking, pharmacophore, virtual screening), Sequence data reduction (Subtractive genomics)

## Unit IV

Characterizing protein and microbes, and epidemiological investigation; Data-repository of model organism in biological sciences (molecular Features, annotation, importance);

Systemic studies: Networks Alignment and evolutionary study of biological networks, development of gene regulatory network, application of Gene Ontologies (GOs) and its usefulness;

Application of Bioinformatics in - Meta-Genomics, Microbes & Environmental Protection,

## Unit V

Role of bioinformatics in drug discovery process: Drug target identification (subtractive genomic approach, microarray based approach, system biology based approach), Lead identification and optimization (HTS, QSAR etc.), Designing of ligand using combinatorial approach, success stories of docking

Bioinformatics methods of identifying bimolecular relationships and networks associated with complex disease/ phenotypes and their applications to drug treatments

Cancer bioinformatics: Identification of cancer causing and related genes, effect of chemotherapy on genome wide gene expression, cancer related databases, success stories of cancer bioinformatics

### Reference Books:

- Applied Bioinformatics: An Introduction By Paul Maria Selzer, Richard Marhöfer, Andreas Rohwer
- Applied Computational Genomics By Yin Yao Shugart, Andrew Collins (auth.), Yin Yao Shugart (eds.)
- Modern Clinical Trial Analysis By Tang, Wan, Tu, Xin (Eds.)
- Computational Biology: Issues and Applications in Oncology By Ying Chen, Danh V. Nguyen (auth.), Tuan Pham (eds.)
- Bioinformatics for Diagnosis Prognosis and Treatment of Complex Diseases By Shen, Bairong (Ed.)

# DNA Computing

## Unit I

Basic Structure of DNA; Role of DNA; DNA & its characteristics; Uniqueness of DNA Computing; Motivation for DNA Computing; Concepts of DNA Computing and DNA Computer; Applications of DNA Computing

## Unit II

Operations on DNA strand (Hybridization, Ligation, Polymerase Chain Reaction (PCR), Gel electrophoresis, Affinity separation), Beginnings of Molecular Computing, Adleman's Experiment, Boolean SATISFIABILITY problem, Breaking DES code, P, NP, NP hard & NP complete, Turing machine. Computational complexity, Brute force algorithm,

## Unit III

DNA Computing vs Bio-Computing, DNA Computer vs Conventional computers, Information Storage & Processing capabilities; Difficulties of DNA computers,

## Unit IV

Models of Molecular Computation (Filtering, Sticker, Constructive - principle of self-assembly, Tile Assembly models, Splicing, Membrane), Watson-Crick automata

## Unit V

Cellular Computing, Ciliate Computing, Applications of DNA Computer Computational Genes, Basic Model, Diagnostic Rules, Diagnosis and Therapy, & Future prospects.

### Reference Books:

- DNA Computing: New Computing Paradigms by Gheorghe Paun, Grzegorz Rozenberg, and Arto Salomaa
- Theoretical and Experimental DNA Computation by Martyn Amos
- DNA Computing Models by Zoya Ignatova, Israel Martínez-Pérez, Karl-Heinz Zimmermann [Springer 2008]