

M.Sc. BIOINFORMATICS (CBCS)
(Effective from the Academic Year 2012 – 2013 onwards)

Course Structure

Eligibility for Admission

A pass with 50% marks in Bachelors Degree in any one of the following as one of the major subjects : Agriculture, Applied Science, Zoology, Animal Science, Biochemistry, Biology, Biotechnology, Bioinformatics, Botany, Plant Biology, Biochemistry and Plant Biotechnology, Chemistry, Computer Science, Computer Applications (BCA), Information Technology, Electronics, Environmental Science, Food Science & Nutrition, Mathematics, Microbiology, Pharmacy, Physics, Statistics, Medical sciences (MBBS/BDS/B.V.Sc.) and B.Pharm. of any recognized Indian or Foreign university.

I Semester

Code	Subject	Hours	Credits
Core	1. Cell and Molecular Biology	6	5
	2. Introduction to Bioinformatics	6	5
	3. Programming in C	6	5
Elective (Any one)	4 (A). Structural Biology (OR)	6	5
	4 (B). Fundamentals of Algorithms		
Practical (Exam will be at II Semester)	1. Lab in Bioinformatics and Computational Biology	3	--
	2. Lab in Programming in C, PERL and R	3	--
Total (4 Courses)		30	20

II Semester

Code	Subject	Hours	Credits
Core	5. Computational Biology	6	4
	6. Mathematics and Statistics	6	4
	7. Programming in PERL	6	4
Elective (Any one)	8 (A). Applied Statistics for Bioinformatics using R (OR)	6	5
	8 (B). Soft Computing Methods		
Practical	1. Lab in Bioinformatics and Computational Biology	3	2
	2. Lab in Programming in C, PERL and R	3	2
Total (6 Courses)		30	25

III Semester

Code	Subject	Hours	Credits
Core	9. Molecular Modelling and Drug Design	6	5
	10. Genomics and Proteomics	6	5
	11. Programming in C++ and Java	6	5
Project	12. Project and Viva-Voce	6	5
Practical (Exam will be at IV Semester)	3. Lab in Programming in C++, Java, and Python	3	--
	4. Lab in Molecular Modelling, Genomics, Proteomics and Cheminformatics	3	--
Total (4 Courses)		30	20

IV Semester

Code	Subject	Hours	Credits
Core	13. Cheminformatics	6	4
	14. Systems Biology	6	4
	15. Programming in Python	6	4
Elective (Any One)	16 (A). Medical Informatics (OR)	6	5
	16 (B). Experimental Techniques of Biomolecules		
Practical	3. Lab in Programming in C++, Java, and Python	3	2
	4. Lab in Molecular Modelling, Genomics, Proteomics and Cheminformatics	3	2
Total (6 Courses)		30	25

Total Number of Subjects : 20 (15 Theory + 4 Practical + 1 Project)

Total Number of Hours : 120

Total Number of Credits : 90

1. Cell and Molecular Biology

Unit I:

Biology of Cells: Cells as a unit of life, structure of prokaryotic and eukaryotic cells. An overview of organelles (Mitochondria, chloroplasts, ER, Golgi, ribosomes, lysosomes and peroxysomes, nucleus and nucleolus). Differences and similarities in plant and animal cells. Cellular membrane: structure, transport, channels, carriers, receptors, endocytosis, membrane potentials.

Unit II:

DNA replication, Transcription and Translation. Cell-cell interactions and signal transductions: Intercellular junctions, signaling by hormones and neurotransmitters; receptors, G-proteins, protein kinases and second messengers. Protein traffic in cells.

Unit III:

Cell Cycle and Regulation: Mitosis, Meiosis. Mutation: Types of mutations, types of mutagenic agents and their molecular mechanism; DNA repair; Chromosomal types and structure; Mechanism by which genome undergoes changes, recombination, mutation, inversion, duplication, and transposition.

Unit IV:

Molecules of Life: Introduction to carbohydrates - Monosaccharides and their derivatives, Disaccharides, Polysaccharides. Proteins: Structure of aminoacids, Different levels of organization-Primary, secondary tertiary and Quaternary structures. Nucleic Acids: Purines, pyrimidines, Nucleosides and Nucleotides, Different structural form of DNA, denaturation and renaturation of DNA Lipids-Structure and function of Fatty acids, Triacylglycerols, sphingolipids, steroids and glycerophospholipids. Water, small molecules-Alkaloids, glycosides, phenols, oligopeptides, Flavonoids, and terpenoids

Unit V:

Enzymes:- Units of Activity, coenzymes and metal cofactors, temperature and pH effects, Michaelis – Menten kinetics, inhibitors and activators, active site and mechanism of enzyme action, Isoenzymes, allosteric enzymes. Metabolism of glucose: glycolysis, TCA cycle, glycogenesis, glycogenolysis and gluconeogenesis, pentophosphate shunt, ETC. Digestion of protein and protein metabolism, nitrogen balance: transamination, oxidative deamination and urea cycle. Lipid metabolism: beta oxidation. Interconnection of pathways, metabolic regulations.

REFERENCES:

1. E.D.P. De Robertis and E.M.F. De Robertis, Jr., Cell and Molecular Biology, Eighth Edition, B.I. Waverly Pvt Ltd, New Delhi, 1996.
2. Lehninger, A. L. 1984. Principles of Biochemistry. CBS publishers and distributors, New Delhi, India.
3. Horton, Moran, Ochs, Rawn, Scrimgeour. 2012 Principles of Biochemistry Prentice Hall Publishers.
4. David. E. Sadava. 1993. Cell Biology: Organelle Structure and Function. Jones & Bartlett publishers.

2. Introduction to Bioinformatics

Unit I:

Introduction to Bioinformatics: Aim and branches of Bioinformatics, Application of Bioinformatics, Role of internet and www in bioinformatics. Basic biomolecular concepts: Protein and amino acid, DNA & RNA, Sequence, structure and function. Forms of biological information, Types of Nucleotide Sequence: Genomic DNA, Complementary DNA (cDNA), Recombinant DNA (rDNA), Expressed sequence tags (ESTs), Genomic survey sequences (GSSs). DNA sequencing methods: Basic and Automated DNA sequencing, DNA sequencing by capillary array and electrophoresis, Gene expression data.

Unit II:

Bioinformatics Resources: NCBI, EBI, ExPASy, RCSB, DDBJ: The knowledge of databases and bioinformatics tools available at these resources, organization of databases: data contents, purpose and utility. Open access bibliographic resources and literature databases: PubMed, BioMed Central, Public Library of Sciences (PloS), CiteXplore.

Unit III:

Sequence databases: Nucleic acid sequence databases: GenBank, EMBL, DDBJ; Protein sequence databases: Uniprot-KB: SWISS-PROT, TrEMBL, UniParc; Structure Databases: PDB, NDB, PubChem, ChemBank. Sequence file formats: Various file formats for biomolecular sequences: GenBank, FASTA, GCG, MSF etc. Protein and nucleic acid properties: Proteomics tools at the ExPASy server, GCG utilities and EMBOSS, Computation of various parameters

Unit IV:

Sequence Analysis: Basic concepts of sequence similarity, identity and homology, definitions of homologues, orthologues, paralogues and xenologues Scoring matrices: basic concept of a scoring matrix, Matrices for nucleic acid and proteins sequences, PAM and BLOSUM series, matrix derivation methods and principles.

Unit V:

Sequence alignment: Measurement of sequence similarity; Similarity and homology. Pairwise sequence alignment: Basic concepts of sequence alignment, Needleman and Wunsch, Smith and Waterman algorithms for pairwise alignments, gap penalties, use of pairwise alignments for analysis of Nucleic acid and protein sequences and interpretation of results.

REFERENCES:

1. Arthur M. Lesk, Introduction to Bioinformatics, Oxford University Press, New Delhi, 2003.
2. Bioinformatics: Sequence and Genome Analysis by Mount D., 2004 Cold Spring Harbor Laboratory Press, New York.
3. Bioinformatics - A practical guide to the analysis of Genes and Proteins by Baxevanis, A.D. and Francis Ouellette, B.F., 1998, John Wiley & Sons, UK.
4. Introduction to Bioinformatics by Teresa K. Attwood, David J. Parry-Smith, 1999, Pearson Education.

3. Programming in C

Unit I:

History of C - Character Set – C tokens – Keywords and Identifiers – Constants – Variables – Data Types – char, int, float, double Type qualifiers – short, long, signed, unsigned– Declaration of Variables – Declaration of storage Class – assigning Values to Variables – Defining Symbolic Constants – Declaring Variables as Constant – Declaring Variable as Volatile – Over flow and Underflow of Data

Unit II:

Operators – arithmetic, unary, relational, logical, assignment, increment, decrement, conditional, bitwise and special operators – arithmetic expressions - Type Conversions in Expressions - Evaluation of expressions - Hierarchy of operators - Input and output statements - character, string and Formatted input and output - Arrays – one- , two- and multi- dimensional arrays - character arrays and strings comparison of strings – mathematical and string library functions.

Unit III:

Control Structures - *if* and *switch* statements - *while*, *do-while* and *for* statements – *goto* statement - Arrays - Character strings - Simple programs.

Unit IV:

User defined Functions in C - Defining and accessing functions - Passing arguments - Function prototypes - Recursion - Passing Arrays to Functions – Passing Strings to Functions – Storage classes - Scope, Visibility and Lifetime of Variables – Multi-file Programs - Pointer Declarations - Passing pointers to functions - Pointers and arrays - Operations on pointers - Arrays of pointers - Dynamic memory allocation.

Unit V:

User defined data types in C – Structures – Declaring structures and Accessing members – Array of structures – Structure within structure – Unions – File operations – open, close, reading and writing – Random access files – Linked list – Preprocessor directives – Macros – Command line arguments.

REFERENCES:

1. S. Parthasarathy, Essentials of Computer Programming in C for Life Sciences, Ane Books, New Delhi, 2008
2. Byron S. Gottfried, Schaum's outline of Theory and Problems of Programming with C, Tata McGraw-Hill, New Delhi, 1991.
3. Brian W. Kernighan and Dennis. M. Ritchie, The C Programming Language, Second Edition, Printice-Hall of India, 1988.
4. Herbert Schildt, 2000. The Complete Reference C, Fourth Edition, Tata McGraw-Hill Publishing Company Limited.

4 (A). Structural Biology

Unit I:

DNA and RNA: types of base pairing – Watson-Crick and Hoogsteen; types of double helices A, B, Z and their geometrical as well as structural features; structural and geometrical parameters of each form and their comparison; various types of interactions of DNA with proteins, small molecules. RNA secondary and tertiary structures, t-RNA tertiary structure

Unit II:

Proteins: 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; Derivation, significance and applications of Ramachandran Map.

Unit III:

Carbohydrates: The various building blocks (monosaccharides), configurations and conformations of the building blocks; formations of polysaccharides and structural diversity due to the different types of linkages. Glyco-conjugates: various types of glycolipids and glycoproteins

Unit IV:

Structure Prediction Strategies: Secondary structure prediction: Algorithms viz. Chou Fasman, GOR methods; analysis of results and measuring the accuracy of predictions using Q3, Segment overlap, Mathew's correlation coefficient Identification/assignment of secondary structural elements from the knowledge of 3-D structure of macromolecule using DSSP and STRIDE methods

Unit V:

Tertiary Structure prediction: Fundamentals of the methods for 3D structure prediction (sequence similarity/identity of target proteins of known structure, fundamental principles of protein folding etc.) Homology Modeling, fold recognition, threading approaches, and ab-initio structure prediction methods – protocols/algorithms. Prediction of protein structure: PHD and PSI-PRED methods

REFERENCES:

1. P.E. Bourne and H. Weissig (Eds.) Structural Bioinformatics, John-Wiley and Sons, 2003.
2. Molecular Modeling Principles and Applications (2nd Ed.) by Andrew R. Leach, 2001, Prentice Hall, USA.
3. Principles of Protein Structure by G. E. Schulz, 2009, Springer.
4. Lehninger Principles of Biochemistry by David L. Nelson and Michael M. Cox, 2005, W. H. Freeman

4 (B). Fundamentals of Algorithms

Unit I:

Computing Algorithms:– Algorithms in Computing, Analyzing algorithms, Designing algorithms, Asymptotic notation, Standard notations, Big ‘O’ notations, Time and space complexity of algorithms and common functions. Sets: Union and Intersections, Differences, Disjoint Sets, Counting Elements, Relations Matrices: Adding and Multiplying, Extracting a sub-matrix, Combining, Inverting

Unit II:

Sorting, Searching & Strings Matching:– Sorting: Bubble Sort, Insertion sort, Selection sort, Quick Sort, Radix sort, Exchange sort, Shellsort, Mergesort. External sort (K-way mergesort, balanced mergesort, polyphase mergesort) Sorting in Linear time, Heaps (Binary Heaps, Janus Heap, Heap sort, Binomial Heaps, Fibonacci Heaps). Searching: Binary Search, Fibonacci Search, Hash Search, Lookup Searches, Generative Searches. String Matching: Naïve algorithm, Boyer-Moore algorithm, Knuth-Morris-Pratt algorithm

Unit III:

Graphs:– Representation of Graphs, Breadth First Search, Depth First Search, Topological Sort, Connected Components, Minimum Spanning Tree, Single-Source Shortest Path (Dijkstra’s and Bellman Fort Algorithm), All-Pairs Shortest Paths (Floyd-Warshall algorithm), Coloring of Graphs (Kruskal’s Algorithm, Prim’s Algorithm),

Unit IV:

Trees:– Forests, DAGs, Ancestors, and Descendants, Binary Search Trees, Querying a Binary search tree, Insertion and Deletion, Tree Traversals, Red-Black Trees, Properties of Red-Black Trees, AVL-Trees, Rotations, Insertion, Deletion, B+ Tree, B* Trees.

Unit V:

Algorithm Design and Analysis:– The substitution method, The iteration method, The master method, Divide and Conquer, Greedy Algorithms, Dynamic Programming (Traveling Sales Person Problem, Hamiltonian Path Problem), Backtracking Algorithms (8-queens Problem, Graph Coloring), Branch and Bound Algorithms

REFERENCES:

1. E. Horowitz and S. Sahani, “Fundamentals of Data structures”, Galgotia Booksource Pvt. Ltd., (1999)
2. Ellis Horwitz, Sartaz Sahani and Sanguthevar Rajasekaran, (1999), “Computer Algorithms”, Galgotia Publications
3. T .H. Cormen, C. E. Leiserson, R .L. Rivest (2001) “Introduction to Algorithms”, 3rd Ed PHI

Practical 1. Lab in Bioinformatics and Computational Biology

Bioinformatics:

1. Download a nucleotide/protein sequence from NCBI/EBI/DDBJ in *fasta* file format.
2. Download a protein structure from PDB. Also, visualize the downloaded structure using Jmol.
3. Perform query sequence search against biological databases using BLAST/FASTA.
4. Align pair of sequences in local/global format using online/offline EMBOSS tool.
5. Predict transmembrane region of protein sequence using Tmpred.
6. Predict domains, families and functional sites of a protein sequence using PROSITE.
7. Convert a file of a nucleotide/protein sequence into other file format using ReadSeq.
8. Design a primer for the DNA sequence using Primer3Plus.
9. Predict ORF from a DNA sequence using ORF Finder.
10. Compute physiochemical properties of the protein sequence using tools in ExPASy server.

Computational Biology:

1. Align three or more nucleotide/protein sequences using ClustalX/W. Also, construct phylogenetic tree from the result and display using Njplot.
2. Predict secondary structure of the protein using GOR/SOPMA/nnPredict.
3. Calculate bond length, bond angle, and dihedral angle from protein structure using RasMol/SwissPdbViewer.
4. Generate conserved patterns from a series of unaligned protein sequences using PRATT.
5. Reverse translate a protein sequence using EMBOSS Backtranseq.
6. Find homologous protein structure templates (PDB) for a protein sequence using BLASTp.
7. Predict secondary structure of RNA using RNAfold.
8. Perform molecular structure format conversion using E-BABEL.
9. Perform conserved domain search for a protein sequence using NCBI CDD.
10. Analyze the quaternary structure of a protein using PDBePISA.

Practical 2. Lab in Programming in C, PERL and R

Programming in C:

1. Convert the given Fahrenheit value to centigrade scale (or vice versa)
2. Compute the relative centrifugal force (RCF) using r_{max} (in cm) and rpm value
3. Find the biggest of three given numbers using if-else statement
4. Compute all possible roots of quadratic equation using if-else statement
5. Find the molecular weight of a DNA with n base pairs in length
6. Find the molecular weight of a given DNA sequence, after checking for phosphorylation
7. Find the sum of i) n natural numbers and ii) n odd numbers
8. Find the factorial of a given integer number.
9. Find the pH of a solution given the concentration of H^+ (or) OH^- ions and print the nature of the solution based on the pH value using if-else statement.
10. Compute the sum of individual digits up to a single digit of a given number
11. Calculate sedimentation time (in hrs. & mins.) using clearing factor and sedimentation coefficient
12. Compute the average of n given values
13. Arrange the n given numbers in ascending order
14. Reverse a given string (without using the built in string function) and Check for palindrome of a given string (without using the built in string function)
15. Arrange the given names in alphabetical order
16. Compute the mean, median, mode, variance, standard deviation for a set of given values
17. Compute the matrix addition, subtraction and multiplication (use different functions for each operation)
18. Generate n modelling numbers using 'static' storage class (define a function)
19. Swap two given numbers using pointers (use a separate function to swap)
20. Computing base composition of a given nucleotide sequence. Read the sequence from a data file.

Programming in PERL and R:

1. Write a PERL program to find transcription/translation/complement/reverse complement of a DNA/RNA/Protein sequence from user's choice.
2. Write a PERL program to translate a DNA sequence in all six reading frames.
3. Write a BioPERL program to download a nucleotide/protein sequence from a biological sequence database.
4. Write a program to find homologous sequences for a query sequence, from biological sequence database using RemoteBLAST using BioPERL.
5. Write a BioPERL program to predict secondary structure of a protein sequence.
6. Write a R program to align pair of sequences using Needleman-Wunsch algorithm.
7. Write a program to display dotplot from the pair of sequences using SeqinR in R.
8. Write a program to convert a file in GenBank file format to FASTA file format using R.
9. Write a R program to compute t -test value from two variables and conclude the hypothesis.
10. Write a R program to download a nucleotide/protein sequence from a biological sequence database.

4. Computational Biology

Unit I:

Analysis of DNA and Protein Sequences – distributions, frequency statistics– Protein and Nucleic Acid Sequence Databases – PIR, Swiss-prot, GenBank – pattern and motif searches – BLOCKS, PRINTS, PFAM – Structure Databases – PDB – Structure Classification, Alignment and Analysis – SCOP, CATH, FSSP

Unit II:

Sequence alignment: Scoring matrices - PAM and BLOSUM- Local and Global alignment concepts – dynamic programming methodology-Needleman and Wunsch algorithm, Smith-Waterman algorithm – Statistics of alignment score – Multiple sequence alignment – Progressive alignment – Databases searches for homologous sequences – FASTA and BLAST

Unit III:

Protein Secondary structure prediction – Chou-Fasman, Garnier-Osguthorpe-Robson (GOR) methods – Predicting 3D structure – comparative (homology) modelling, threading (fold recognition) and ab initio methods – Rosetta – CASP – Protein structure visualization tools – RasMol, Swiss PDB Viewer

Unit IV:

Fragment assembly – Genome sequence assembly – Gene finding methods: content and signal methods – Background of transform techniques – Fourier Transform and gene prediction – Analysis and prediction of regulatory regions – Neural network concepts and secondary structure prediction – Probabilistic models: Markov chain – random walk – Hidden Markov models – Gene identification and other applications.

Unit V:

Evolutionary analysis: distances – clustering methods – rooted and un-rooted tree representation – Bootstrapping strategies – phylogenetic trees – PHYLIP.

REFERENCES:

1. Arthur M. Lesk, Introduction to Bioinformatics, Oxford University Press, New Delhi (2003).
2. D. Higgins and W. Taylor (Eds), Bioinformatics- Sequence, structure and databanks, Oxford University Press, New Delhi (2000).
3. R. Durbin, S.R. Eddy, A. Krogh and G. Mitchison, Biological Sequence Analysis, Cambridge Univ. Press, Cambridge, UK (1998).
4. A. Baxevanis and B.F. Ouellette. Bioinformatics: A practical Guide to the Analysis of Genes and Proteins, Wiley- Interscience, Hoboken, NJ (1998).
5. Michael S. Waterman, Introduction to computational Biology, Chapman & Hall, (1995).
6. C. Gibas and P. Jambeck, Developing Bioinformatics Computer Skill, 1st Edition, O'Reilly, 2001.

6. Mathematics and Statistics

Unit I:

Basic Mathematics:- Matrix Algebra – Types – Determinants – Transpose – Conjugate – Inverse – Eigen values of matrices – Rank – Solving Simultaneous equations in three variables using matrices, Cayley – Hamilton theorem without proof – Verification and Computation of Inverse of a Matrix – Consistency of linear equations. Vector – Addition, subtraction – Dot product – Cross product (up to 3 vectors) – Scalar triple product – Gradient – Divergence and Curl.

Unit II:

Calculus:- Differentiation – Standard results – Derivatives of simple functions – Product Rule – Quotient Rule. Partial Differentiation – Partial derivative of simple functions (3 variables case only) – Euler's Theorem. Integration – Standard results – Integrals of simple functions – Definite Integrals – Indefinite Integrals – Integration by parts – Integration by substitution – Integration by partial fractions.

Unit III:

Basic Statistics:- Ungrouped data and Frequency distribution: Collection – Classification – Tabulation – graphical and diagrammatic representation of numerical data – Graphs – Histogram, Frequency curve. Statistical Averages: Mean, Median, Mode, SD, Variance and Coefficient of variation. Correlation and regression analysis: Types of correlation, Methods of studying correlation – Rank correlation – Simple linear regression – Regression Equations.

Unit IV:

Probability:- Random experiment – Definitions of probability – Theorems of Probability: Addition rule – Multiplication Rule – Properties of probability – Conditional probability – Bayes Theorem – Simple Problems. Random variables – Discrete and Continuous – Probability mass functions – Probability density functions – Cumulative density function and its properties – Distributions Function. Theoretical distributions – Binomial, Poisson and Normal distributions – Basic ideas and their applications.

Unit V:

Test of Hypothesis:- Sampling: Population – Sample – Parameter – Statistic – Standard error – Hypothesis-Null Hypothesis – Alternative Hypothesis – Critical Region – Level of Significance – Errors in Sampling – One tailed and two test statistic-test of significance and its test procedure. Test of significance for small samples: Tests based on normal distribution for Single mean, difference of two means – Tests based on t-distribution for single mean, difference of two means, paired test and observed correlation coefficient – F Test – Parametric and Non parametric tests – Chi-square (χ^2) test for goodness of fit. Analysis of variance: One way and two way classifications.

REFERENCES:

1. S.C. Gupta and V.K. Kapoor, Fundamentals of Mathematical Statistics, 11th Edition, Sultan Chand & Sons, New Delhi, 2002.
2. D.W. Jordan and P. Smith, Mathematical Techniques, 3rd Edn, Oxford University Press, New Delhi, 2002.
3. L. Forthofer, Introduction to Biostatistics, Academic Press, 1995.
4. Robert R. Sokal and F.J. Rohlf, Introduction to Biostatistics (Biology-Statistics Series), W.H. Freeman & Company, New York, 1987.
5. E. Batschelet, Introduction to Mathematics for Life Scientists, 2nd Edn., Springer International Student Edn., Narosa Publishing House, New Delhi, 1991.
6. Erwin Kreyszig, Advanced Engineering Mathematics (8th Ed.), Wiley Dreamtech India (P) Ltd., New Delhi, 1999.

7. Programming in Perl

Unit I:

Data Structure : Scalar Variables, Scalar Operations and Functions, Array Variables Literal Representation of Array, Array Operations and Functions, Scalar and List Context, Hash Variables, Literal Representation of a Hash, Hash Functions, Using Hashes for the Genetic Code, Gene Expression Data Using Hashes

Unit II:

Modular Programming: Subroutines, Advantage of Subroutines, Scoping and Subroutines, Arguments, Passing Data to Subroutines, Modules and Libraries of Subroutines, Concept about File handle, Opening and Closing a File handle, Opening and Closing a Directory Handle, Reading a Directory Handle, File and Directory Manipulation.

Unit III:

Regular Expression and Pattern Matching: Concepts about Regular Expressions, Simple uses of Regular Expressions, Patterns, Matching Operator, Substitutions, Split and Join functions.

Unit IV:

Common Gateway Interface (CGI) Programming: The CGI.pm Module, CGI program in Context, Simple CGI programs, Passing Parameters via CGI, Perl and the Web

Unit V:

Bioperl: Introduction to Bioperl, Installing procedures, Architectures, General Bioperl Classes, Sequences (Bio::Seq Class, Sequence Manipulation), Features and Location Classes (Extracting CDS), Alignments (AlignIO), Analysis (Blast, Genscan), Databases (Database Classes, Accessing a local database), Implementing REBASE

REFERENCES:

1. Beginning Perl for Bioinformatics (1st Edition) by Tisdall, J., 2004, O'Reilly Publishers.
2. Learning Perl (5th Edition) by Randal L. Schwartz, Tom Phoenix and Brain d Foy, 2008, O'Reilly Publishers.
3. Programming Perl (3rd Edition) by Wall, W., Christiansen, T. and Orwant, J., 2000, O'Reilly Publishers.
4. Official Site of Perl: <http://www.perl.org>
5. Official Site of BioPerl: <http://www.bioperl.org>

8 (A). Applied Statistics for Bioinformatics using R

Unit I:

Introduction: Installing R; R as a deluxe calculator, creating objects and assigning values.

Unit II:

Graphics; simple plotting, advanced plotting, using color in plots, using subscripts and superscripts in graph labels, interactive graphics, saving graphical output, loops.

Unit III:

Working with data sets: Data structures, moving to and from files, character data, generating random values.

Unit IV:

Statistical distributions: The use of statistical distributions in epidemiology, statistical analyses, writing functions, closing a session.

Unit V:

Sequence Analysis: Frequency Plot, Pairwise Alignment, Microarray Data Analysis, Markov Model

REFERENCES:

1. Robert Gentleman, 2008 R Programming for Bioinformatics, Chapman & Hall/CRC Computer Science & Data Analysis
2. Crawley, M.J. 2007. The R Book. Wiley.
3. Gentleman, R., Carey, V.J., Huber, W., Irizarry, R.A., and Dudoit, S., eds. 2005. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York.
4. Murrell, P. 2006. R Graphics. Chapman & Hall/CRC, Boca Raton.
5. Venables, W.N., Smith, D.M., and the R Development Core Team. 2007. An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics. Version 2.6.1. <http://cran.rproject.org/doc/manuals/R-intro.pdf>.
6. R Gentleman, V. Carey, W Huber, R Irizarry, S Dudoit. 2005 Bioinformatics and computational biology solutions using R and BioConductor, Springer
7. G. Parmigiani, E. Garrett, R. Irizarry and S. Zeger. 2003 The analysis of gene expression data, methods and software. Springer
8. Wim P. Krijnen. 2009 Applied Statistics for Bioinformatics using R. cran.r-project.org/doc/contrib/Krijnen-IntroBioInfStatistics.pdf
9. Kim Seefeld, and Ernst Linder. 2007 Statistics Using R with Biological Examples. cran.r-project.org/doc/contrib/Seefeld_StatsRBio.pdf

8 (B). Soft Computing Methods

Unit I:

Introduction to Soft Computing: History - The need for tolerance for imprecision, uncertainty, approximate reasoning etc for low cost solutions; Biological motivation in soft computing – comparison of human brain and CPU.

Unit II:

Artificial Neural Networks: Historic evolution – Perceptron, Features of NNs – Fault Tolerance, Parallelism, Flexibility and Adaptivity, Learning etc, Disadvantages; supervised and unsupervised networks, Multi – layer Perceptrons, Back Propagation Algorithm (Derivation not required), Learning & Momentum Parameters, selecting Hidden nodes, Training & Testing, Overview of other ANNs - Kohonen's networks, Boltzman Machine, ART, ANN applications bioinformatics, overview of Support Vector Machines.

Unit III:

GA, and ACO: Genetic Algorithms, Basic Concepts, Reproduction – Cross over – Mutation - Fitness Value- Optimization using GAs; Applications in bioinformatics; Ant Colony Optimization: Swarm Intelligence – Basic motivation – artificial ants – the bridge crossing problem – theory and applications.

Unit IV:

Fuzzy Logic: Fuzzy sets, Membership functions, Logical Operations, Fuzzification & Defuzzification – Linguistic Variables, Fuzzy rule – based reasoning; Applications in Bioinformatics; overview of Neuro – Fuzzy Systems. Dimensionality Reduction: Overview of Dimensionality Reductions techniques, K-means clustering, Principal Component Analysis - Eigen values & Eigen vectors of co-variance matrix, choosing Components

Unit V:

Hidden Markov Models: Markov processes and Markov Models, Hidden Markov Models, 3 basic HMM problems & corresponding algorithms – Application in bioinformatics

REFERENCES:

1. Sushmita Mitra, Yoichi Hayashi, "Bioinformatics with Soft Computing", IEEE Transactions on systems, Man and Cybernetics, Vol.56, No.5, September 2006.
2. T.D. Seely, "The Wisdom of the Hive", Harvard University Press, 1995, pages 295-296.
3. Marco Dorigo and Thomas Stutzle "Ant Colony Optimization", Prentice-Hall of India, India, 2001
4. David E Goldberg, 'Genetic algorithm', Pearson Education.
5. Timo Koshi, 'Hidden Markov model for Bioinformatics', Kluwer Academic Publishers.
6. M Ganesh, 2006 'Introduction to Fuzzy sets and Fuzzy logic', Phi Publications, Inc.
7. Martin Gollery, 'Hand book for Hidden Markov model for Bioinformatics', CPC Press.

9. Molecular Modelling and Drug Design

Unit I:

Introduction to molecular force fields: General features - bond stretching, angle bending, improper torsions, out of plane bending, cross terms, non-bonded interactions, point charges, calculation of atomic charges, polarization, van der waals interactions, hydrogen bond interactions, Water models. Types of force field - all atoms force field, united atom force field, etc.

Unit II:

Molecular Energy minimization: Steepest descent, conjugate gradient – Derivatives, First order steepest decent and conjugate gradients. Second order derivatives Newton-Raphson, Minima, maxima saddle points and convergence criteria.-non derivatives minimization methods: the simplex and sequential univariate methods.

Unit III:

Molecular Dynamics Simulation methods: Classical Molecular Dynamics: Newtonian dynamics, Integration algorithm, Periodic boundary conditions and minimum image convention, Potential truncation and shifted-force potentials, Neighbour list, Force calculations, Long range interactions, MD code for liquid Argon. Classical Monte Carlo: Random numbers, Evaluating integrals using random numbers, Importance sampling, Metropolis algorithm, Smart MC techniques. Analysis of simulated trajectories: Radial distribution functions, Self diffusion coefficient, Time correlation functions

Unit IV:

Drug and Vaccine design: Drug discovery process. Role of Bioinformatics in drug design. Target identification and validation, lead optimization and validation. Structure-based drug design and ligand based drug design. Modeling of target-small molecule interactions. Fundamentals of docking small and macromolecules to proteins and nucleic acids. Vaccine design: Reverse vaccinology and immunoinformatics. Databases in Immunology. B-cell epitope prediction methods. T-cell epitope prediction methods. Resources to study antibodies, antigen-antibody interactions

Unit V:

Structure Activity Relationship: QSARs and QSPRs, QSAR Methodology, Various Descriptors used in QSARs: Electronics; Topology; Quantum Chemical based Descriptors. Use of Genetic Algorithms, Neural Networks and Principle Components Analysis in the QSAR equations

REFERENCES:

1. Molecular Modeling Principles and Applications (2nd Ed.) by Andrew R. Leach, 2001, Prentice Hall, USA.
2. Molecular Modeling and Simulation – An interdisciplinary Guide by Tamar Schlick, 2000, Springer-verlag.
3. Computational medicinal chemistry for drug discovery edited by Patrick Bultinck, 2004, Marcel Dekker Inc.

10. Genomics and Proteomics

Unit I:

Genomics and Metagenomics: Large scale genome sequencing strategies. Genome assembly and annotation. Genome databases of Plants, animals and pathogens. Metagenomics: Gene networks: basic concepts, computational model such as Lambda receptor and lac operon. Prediction of genes, promoters, splice sites, regulatory regions: basic principles, application of methods to prokaryotic and eukaryotic genomes and interpretation of results. Basic concepts on identification of disease genes, role of bioinformatics-OMIM database, reference genome sequence, integrated genomic maps, gene expression profiling; identification of SNPs, SNP database (DbSNP). Role of SNP in Pharmacogenomics, SNP arrays. Basic concepts in identification of Drought stress response genes, insect resistant genes, nutrition enhancing genes

Unit II:

Epigenetics: DNA microarray: database and basic tools, Gene Expression Omnibus (GEO), ArrayExpress, SAGE databases DNA microarray: understanding of microarray data, normalizing microarray data, detecting differential gene expression, correlation of gene expression data to biological process and computational analysis tools (especially clustering approaches)

Unit III:

Comparative genomics: Basic concepts and applications, whole genome alignments: understanding the significance; Artemis, BLAST2, MegaBlast algorithms, PipMaker, AVID, Vista, MUMmer, applications of suffix tree in comparative genomics, synteny and gene order comparisons Comparative genomics databases: COG, VOG

Unit IV:

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 V:

Proteomics: Protein arrays: basic principles. Computational methods for identification of polypeptides from mass spectrometry. Protein arrays: bioinformatics-based tools for analysis of proteomics data (Tools available at ExPASy Proteomics server); databases (such as InterPro) and analysis tools. Protein-protein interactions: databases such as DIP, PPI server and tools for analysis of protein-protein interactions

REFERENCES:

1. Principles of Genome Analysis and Genomics (3rd Ed.) by Primrose, S.B. and Twyman, R.M., 2003, Blackwell Publishing Company, Oxford, UK.
2. Introduction to proteomics – Tools for the new biology (1st Ed.) by Liebler, D.C., 2002, Human Press Inc., New Jersey, USA.
3. Bioinformatics and Functional Genomics by Pevsner, J., 2003, John Wiley and Sons, New Jersey, USA.
4. Bioinformatics: Sequence and Genome Analysis by Mount, D., 2004, Cold Spring Harbor Laboratory Press, New York.

11. Programming in C++ and Java

Unit I:

Object Oriented Programming (OOP) - Basic concepts and applications - Differences between C and C++ - Functions in C++ - *inline* Functions - Default arguments.

Unit II:

Function overloading/polymorphism - Classes and objects - Constructors and destructors - Operator overloading and type conversions.

Unit III:

Extending classes - Inheritance and its types - Single level, multilevel, multiple and hybrid inheritance - Pointers to objects and derived classes - Virtual functions - C++ stream classes - Console I/O operations - Simple Data File operations

Unit IV:

Java Basics - Importance and features of java, Modifiers, Access Controls, Data types, Expressions, Declarations, Statements & Control Structures, Program Structures, Packages, Interfaces, Working with java util Package, Garbage Collection- Multithreaded Programming- Java Thread Model - Creating a Thread and Multiple Threads - Thread Priorities- String handling - I/O – Applets - the Applet class.

Unit V:

BioJava - Installing BioJava, Symbols, Basic Sequence Manipulation (DNA to RNA, Reverse Complement, motif as regular expression), Translation (DNA to Protein, Codon to amino acid, Six frame translation), Proteomics (Calculate the mass and pI of a peptide), Sequence I/O (File Formats conversions), Locations and Features (PointLocation, RangeLocation, Feature modifications), BLAST and FASTA (Blast and FastA Parser, extract information from parsed results), Counts and Distributions, Weight Matrices and Dynamic Programming, User Interfaces.

REFERENCES:

1. E. Balagurusamy, Programming in C++, Tata McGraw-Hill Publishing Company Ltd, New Delhi, 2004.
2. Robert Lafore, Object-Oriented Programming in Turbo C++, Galgotia Publications, New Delhi, 1991.
3. Herbert Schildt, Java - A Beginners Guide (4th Ed.), 2007, Tata Mc-Graw-Hill publication
4. Computing Concepts with Java 2 Essentials (2nd Ed.) by Horstmann, C.S., 2000, John Wiley Publishers.
5. Object Oriented Design and Applications (2nd Ed.) by Benjamin, Cummings and Booch, G., 1994, Addison Wesley Publishers.

12. Project and Viva-Voce

Project Report - 60 Marks

Viva-voce - 40 Marks

Total - 100 Marks

Instructions:

1. The Project for PG students shall be “**Individual Project**”.
2. Project report evaluation will be done **centrally** and Viva-Voce will be conducted by both the External examiner and the guide at the end of **third semester**.

13. Cheminformatics

Unit I:

Introduction: Introduction to chemoinformatics, History and Evolution of chemoinformatics, Use of chemoinformatics, Prospects of chemoinformatics, Molecular Modeling and Structure Elucidation

Unit II:

Representation of Molecules and Chemical Reactions: Nomenclature; Different types of Notations; SMILES Coding; Matrix Representations; Structure of Molfiles and Sdfiles; Libraries and toolkits; Different electronic effects; Reaction classification

Unit III:

Searching Chemical Structure: Full structure search; sub structure search; basic ideas; similarity search; Three dimensional search methods; Basics of Computation of Physical and Chemical Data and structure descriptors; Data visualization and Non-linear Mapping

Unit IV:

Computer Assisted Virtual screening design: Structure Based Virtual Screening- Protein Ligand Docking, Scoring Functions for Protein Ligand docking, Practical aspects of structure based Virtual Screening; Prediction of ADMET Properties, 2 D and 3D data searching, Chemical databases, Role of computers in Chemical Research.

Unit V:

Application of Chemoinformatics in Drug Design: Quantitative Structure-Property Relations; Descriptor Analysis; Computer Assisted Structure elucidations; Target Identification and Validation; Lead Finding and Optimization; Analysis of HTS data; Design of Combinatorial Libraries; Ligand-Based and Structure Based Drug design

REFERENCES:

1. Andrew R. Leach, Valerie J. Gillet, Introduction to Chemoinformatics, Kluwer Academic Publishers, Netherlands, 2003.
2. Lisa B. English, Combinatorial Library Methods and Protocols, Humana Press Inc, Volume 2, 2002.
3. Frank Jensen, Introduction to Computational Chemistry, Wiley Publisher, Second Edition, 2006.

14. Systems Biology

Unit I:

Introduction: Systems Biology Networks- basics of computer networks, Biological uses and Integration. Micro array – definition, Applications of Micro Arrays in systems biology. Selforganizing maps and Connectivity maps - definition and its uses. Networks and Pathways – Types and methods. Metabolic networks.

Unit II:

Simulation of pathways: Whole cell: Principle and levels of simulation – Virtual Erythrocytes. Pathological analysis. Flux Balance Analysis. Biochemical metabolic pathways, Metabolomics and enzymes. Interconnection of pathways, metabolic regulation. Translating biochemical networks into linear algebra. Cellular models. Networks and Motifs: Gene Networks: basic concepts, computational models. Lambda receptor and lac operon as an example. All types of networks and its uses.

Unit III:

Signalling & Experimental methods in systems biology: slow and auto-regulation The coherent FFL- temporal order, FIFO, DOR, Global, Development, memory and irreversibility signalling networks and neuron circuits-robust adaptation-any model. Robustness and optimality in Biology: model and integral feedback-signaling/bifunctional enzymes. Perfect robustness- Role and its measurement. Linking models and measurement, concepts, calibration and identification, data Vs metadata.

Unit IV:

Design of Circuits and Databases: Introduction- databases KEGG, EMP, MetaCyc, AraCyc etc., Expression databases and various databases related to systems biology. Optional design of gene circuits I- cost and benefit: gene circuits II- selection of regulation. Stochasticity in gene expression.

Unit V:

Synthetic Biology: Introduction, definition and Basics, Synthetic Oligonucleotide/DNA-based, RNA-based, Peptide-based and polyketide Technologies and Applications, Technologies and Applications of Directed Evolution and Microbial Engineering, Potential Hazards of Synthetic Biology

REFERENCES:

1. Systems Biology: Definitions and perspectives by L. Alberghina H. V. Westerhoff, 2005, Springer
2. Synthetic Biology, A New Paradigm for Biological Discovery, a report by Beachhead Consulting, Feb 2006
3. Computational systems biology by A.Kriete, R.Eils, 2005, Academic press.
4. Systems Biology in practice: Concepts, Implementation and applications by E. Klipp, R. Herwig, A. Kowlad, C. Wierling and H. Lehrach, 2005, Wiley InterScience
5. Systems Biology and Synthetic Biology by Pengcheng Fu, Sven Panke, 2009, Wiley InterScience

15. Programming in Python

Unit I:

Introduction to Python: Data types, variables, expressions, operators. Sequence, set, dictionary, print statement, control-flow statements, functions.

Unit II:

Objects and classes, metaclasses. Decorators, special methods. Exception handling. Modules sys, os, etc. Strings and regular expressions. File operations. Working with processes and threads. Pipes and signals

Unit III:

Graphical user interface design in Python (including the Tkinter module), Widgets and basic components, Layout options, Event handling

Unit IV:

Network scripting (sockets, FTP, and e-mail clients), Server-side scripting. Databases and persistence in Python (including pickled objects and shelf files)

Unit V:

Custom and built-in data structures in Python. C integration with Python (including the SWIG module), Embedding Python calls within C. Introduction to BioPython.

REFERENCES:

1. Programming Python by Mark Lutz, O'Reilly
2. Learning Python, 3rd Edition by Mark Lutz, O'Reilly
3. Python in a Nutshell by Alex Martelli, O'Reilly
4. An Introduction to Python by Guido van Rossum and Jr. Fred L. Drake, Network Theory Ltd
5. Official site of BioPython: <http://www.biopython.org>

16 (A). Medical Informatics

Unit I:

Introduction to medical informatics: Basic concepts in health informatics and its history, definitions, sub disciplines and professional organizations and activities.

Unit II:

Major applications and commercial vendors: Major health informatics applications including electronic medical records (EHR) and computerized physician order entry (CPOE). Data interoperability of records; Medical controlled vocabularies and electronic data interchange standards. Commercial and open source applications.

Unit III:

Good clinical practices (GCP): GCP history and guidelines; India and the world in GCP; Standard operating procedure and guidelines; Current issues in India and abroad. Institutional review boards and protection of human health.

Unit IV:

Information systems design and engineering: Planning and implementing issues of healthcare information systems. Software engineering principles, human factors and human-computer interaction issues, and evaluation methods of end user acceptance and outcomes.

Unit V:

New opportunities and emerging trends: Information technologies in healthcare; wireless and handheld devices, social computing paradigms, and eHealth applications, web-based tools.

REFERENCES:

1. Taylor Paul, From Patient Data to Medical Knowledge: The Principles and Practice of Health Informatics, Blackwell Scientific Publishing, 2006.

16 (B). Experimental Techniques of Biomolecules

Unit I:

Isolation and purification of proteins - Crystallization of protein – Crystal Structure – Bravais Lattice – Symmetry elements and operations – Point groups – Space groups – Bragg's law – X-ray diffraction - Proteins structure determination by X-ray diffraction - Phase determination - Calculation of electron density map - Interpretation of electron density map - Refinement of the structures - Electron crystallography of proteins – High throughput techniques in Crystallography

Unit II:

Electronic energy levels – electronic transitions – selection rules – types of spectra – IR, UV – visible spectroscopy - Measurement of Infrared (IR) spectrum – Theory of IR spectroscopy – IR spectra of polyatomic molecules – biological examples – Theory of UV - visible spectroscopy – application of UV spectra to proteins – measurement of molecular dynamics by fluorescence spectroscopy

Unit III:

The principle of Nuclear Spin – Spin flipping – theory of Nuclear Magnetic Resonance – spectral parameters in NMR – intensity, chemical shift, spin-spin coupling, relaxation times, line widths, nuclear Overhauser effect (NOE), chemical exchange, paramagnetic centers – application of NMR in biomolecular structure determination.

Unit IV:

Principles of electrophoresis – SDS PAGE – Molecular weight determination of proteins - 2D-gel electrophoresis – capillary electrophoresis - principles of chromatography – Gel & ion exchange chromatography – applications

Unit V:

Micro array techniques and their applications in biology - Mass spectroscopy - ESI and MALDI-TOF - protein finger printing.

REFERENCES:

1. W. Kemp, Organic Spectroscopy, 3rd edition, ELBS, McMillan, London, 1991.
2. C. N. Banwell and E.M.McCash, Fundamentals of molecular spectroscopy, 4th edition, Tata McGraw Hill, New Delhi, 1995.
3. I. Howe, D. H. Williams and R. D. Bowen, Mass Spectrometry, principles and applications, 2nd edition, McGraw Hill, London, 1981.
4. Gary Siuzdak, Mass Spectroscopy for Biotechnology, Academic Press, 1995.
5. Cunico, Gooding and Wehr Bay, Basic HPLC and CE of Biomolecules, Bioanalytical Lab, 1998.
6. Van Holde, Principles of Physical Biochemistry, Prentice Hall, 2000.
7. Helen C. Causton, John Quackenbush and Alvis Brazma, A Beginner's Guide: Microarray Gene Expression Data Analysis, Blackwell Publishing, USA, First Indian Reprint, 2004.
8. Vasantha Pattabhi and N. Gautham, Biophysics, Narosa Publishing House, New Delhi, 2002.

Practical 3. Lab in Programming in C++, Java, and Python

Programming in C++:

1. Write a C++ program to convert temperature given in Fahrenheit into Celsius and ii) Celsius to Fahrenheit.
2. Write a C++ program to swap two values using reference variable.
3. Write a C++ program to find all possible roots of a quadratic equation.
4. Write a C++ program to compute the mean and standard deviation of a given n values.
5. Write a C++ program to compute matrix addition, subtraction and multiplication of two matrices using functions
6. Write a C++ program by defining two functions with the same name `power()` to raise a number m to a power n . One function takes a double value for m and int value for n and the other function having the same name takes a *int* value for m and *int* value for n . Write a main function that calls both the functions to demonstrate the function overloading.
7. Write a C++ program using class to find the smallest of two numbers.
8. Write a C++ program with the following specifications:
 - a. Define a class to represent a gene sequence data. Include the following members:
 - Data members:
 - Name of the gene
 - Gene id
 - Length
 - A, T, G, C content
 - Member functions:
 - To read data for a gene
 - To compute A, T, G, C content
 - To display all the details of a gene
 - b. Write a main program to test the program by reading n gene sequences data.
9. Write a C++ program to exchange the private values of two classes using a common friend function.
10. Write a C++ program to find the sum of two complex numbers using overloaded constructors for data input and operator overloading

Programming in Java:

1. Write a java program to find addition of two matrices using arrays.
2. Write a java program to display a multiplication table in the format $n \times i = m$
3. Write a java program to check whether a string is palindrome or not
4. Write a java program to find greatest and smallest element of an array
5. Write a java program to create a simple circle using java applet
6. Write a BioJava program to compute physicochemical properties of a protein sequence
7. Create a simple circle using java applets

Programming in Python:

1. Write a Python program to calculate addition, subtraction, multiplication, and division
2. Write a simple CGI program which contains all html form fields using Python
3. Create a simple application using Python Tkinter
4. Write a program using BioPython to find complement, reverse, reverse complement, transcription, and translation of a DNA sequence.
5. Write a Python program to draw a protein hydrophobicity plot using SciPy/MatPlotLib package

Practical 4. Lab in Molecular Modelling, Genomics, Proteomics and Cheminformatics

Molecular Modelling and Cheminformatics:

1. To perform the consensus secondary structure prediction for a given protein sequence at NSP@ by selecting six different methods
2. To identify the fold for a given protein sequence using 3-D PSSM fold recognition server
3. To find the structural neighbours of a given protein (2TRX) according to SCOP, CATH, FSSP and CE. To find out if any particular structure is identified by all these classifications.
4. Protein structure prediction and validation
 - a. Primary feature computing by PROTPARAM
 - b. Secondary structure by SOPMA
 - c. 3D structure by PSI-BLAST tool, SWISS-MODEL and SAVS server (MODELER software)
5. Protein structural alignment and classification
 - a. Pairwise structural alignment by DALITE server
 - b. Multiple structural alignments by DALI/ConSurf server
 - c. Structural classification by SCOP and CATH servers
6. Retrieval chemical information from PUBCHEM and Ligand databases
7. Retrieving pharmacological information from Pharma base and MSDchem database
8. Prediction of binding affinity of ligand by protein-ligand interaction database tools
9. Ligand design and analysis by ISIS ChemDraw, VMD software
10. Protein-ligand interaction prediction by ArgusLab and Discovery Studio software
11. Binding site identification of target by Q-site finder server
12. Molecular properties prediction by VEGAZ software
13. Molecular dynamics simulation by GROMAS software
14. Molecular force field analysis by TINKER software
15. Drug activity test by ADMETox tools

Genomics and Proteomics:

1. Comparative genome annotation by VISTA tools
2. Bacterial operon prediction by OperonDB tools
3. Gene prediction by WebGene, ORF finder and COG database and GenoCluster software
4. Promoter and regulan prediction by Virtual FootPrint
5. Prediction of secondary structure of rRNA by rRNA project and mFOLD server
6. Cloning vector construction by FastPCR, pDRAW software, WebCutter server and Plasmid Genome Database
7. Identification of coding region by CRITICA and CodanDB tools
8. Identification of mutations in genes by GeneSNP-VISTA software
9. Recombination frequency analysis by MEGA, RAS and RAT software
10. Metabolic pathway prediction by UB-BBD and Pathway Hunter Tool
11. Protein network prediction by Sting and KEGG Net servers
12. Protein bulk properties prediction by WinGene/WinPep software
13. 2D gel data analysis by SWISS-2D GEL DB and NCI Flicker web server/software
14. Microarray data analysis by NCBI-OMNIBUS, TIGR Archive Viewer and TH4 / BioConductor / BASE software
15. Mass spectroscopy data analysis by MSDB and Mascot/PeptIdent/GFS server