Ph.D in Bioinformatics (by course work)

With effect from 2014- 2015

Course -1: RESEARCH METHODOLOGY

Unit 1

Research Methodology: Introduction, Definition, Objectives of Research – Types - Methods, Research Process, Basic Principles of Experimental Design, Selection of Journal - preparation of scientific manuscript - Scientific Information's - Sites for journal and publications – Impact Factor – ICN Value - thesis writing – proof correction.

Unit 2

Techniques: Principles of Chromatography - Paper Chromatography – TLC, Gas Chromatography, HPLC, PCR, SDS-PAGE 2-D Electrophoresis, MALDI-TOF, Mass Spectrometry.

Unit 3

Bio-statistics: Collection and Classification of Data – Diagrammatic and Graphic Representation of Data – Measurement of Central Tendency – Standard Deviation – Normal Distribution – Test of Significance based on Small samples - Large samples – Student 'T' Test – Correlation and Regression – Chi Square Test for Independence of Attributes – ANOVA.

Unit 4

Genome Technology: High Throughput Genome and Protein sequencing, Genome and Proteome Assembly, DNA and Protein Micro Arrays, Gene and Protein Expression data sets.

Algorithms: Protein and Nucleic Acid Sequence Algorithms: Biological Databases, Use of the Algorithms BLAST, Multiple Sequence Alignments and Clustering Algorithms. Phylogeny: Evolutionary Trees, Pathway Analysis.

Unit 5

Macromolecules: DNA and RNA, Proteins – Primary, Secondary, Super Secondary, Tertiary and Quaternary Structure, Ramachandran Plot, Protein Structure Databases; Protein Structure Comparison; Fold Recognition; 3D-1D. Structure based drug design docking,

Systems Biology: Objectives of Systems Biology, Strategies relating to *insilico* Modeling of Biological Processes, Signal Transduction Pathways, Gene Expression Patterns.

Reference:

1. Mathews. "Successful scientific writing: A step-by-step guide for Biomedical Scientists", Second edition, Cambridge University Press, 2001.
2. Oster.G and A.W.Pollistol, Physical techniques in Biological research, Vol.III. Acadamic Press, Inc., New York. 1956
3. Lu, F.C. Basic Toxicology. Hemisphere, Pub, Crop. Washington, New York, Londan.1985
4. Andrews D.Baxevanis and BF. Francis Ouellette, Bioinformatics-A Practical guide to the Analysis of Genes and Proteins, A John Wiley & Sons Pub, 2001.
5. Mount, D."Bioinformatics: Sequence and Genome Analysis"; Cold Spring Harbor Laboratory Press, New York, 2004.

Course II : STRUCTURAL BIOLOGY AND PHYLOGENETICS

Unit 1

Structural Biology: Nucleic acid structures, RNA folding, RNA loops, conformational study, various ribose ring conformations, ribose-ring puckering, protein-protein interactions, protein-ligand interactions, NA binding proteins, RNA-binding proteins, Ramachandran plot, three dimensional structures of membrane proteins, importance of helix and loops, biophysical aspects of proteins and nucleic acids.

Unit 2

Biological Databases: Database- Organization of Databases: Data Contents, Purpose and Utility. Sources of primary data and basic principles of the method for deriving the Secondary data, Organization of data, contents and formats of database entries. Nucleic acid sequences. Sequence databases: GeneBank, European Molecular Biology Laboratory (EMBL) Nucleotide sequence databank, DNA Data Bank of Japan (DDBJ). Structural databases : Structural databases; Protein Data bank (PDB), Nucleic Acid Data Bank (NDB), Molecular modeling Data Bank (MMDB), Structure classification Databases; SCOP, CATH, PDBSum.

Unit 3

Genomics: Definition of Genomics, Gene map-Types of gene mapping, Molecular marker for mapping, conventional genome sequencing, Functional genomics-techniques. Sequence comparison,

Proteomics: Definition of proteomics, Tools of protein separation, analysis and characterization. Protein digestion techniques. Protein finger printing, 2D Gel Electrophoresis, Mass spectrometry and MALDI –TOF, SALDI –TOF (Surface enhanced Laser Desorption Time Of Flight). Protein Databases-PDB-PIR. Typical entries-Protein visualization tools-Swiss PDB Viewer, Expasy.

Unit 4

Molecular Modeling Drug Design-Introduction to molecular force fields: General features- bond stretching, angle bending, bond Interaction, types of force field. Molecular Energy minization. Drug discovery process- drug design target identification and validation, lead optimization and validation. Structure-based drug design and ligand based drug design.

Unit 5

Molecular Phylogenetics : Concept of phylogenetics –Application of Phylogenetic trees- Basic terminology-taxa, taxonomy, clade, root, leaf, node graph & tree, Dendogram, cladogram, rooted tree, unrooted tree, scaled trees- Molecular clock hypothesis, Distance based methods-UPGMA, NJ algorithm, Character based methods-Maximum parsimony methods. Validating phylogenetic trees – jack knifing and bootstrapping – Study of Phylip, NJ plot, ClustalX and other popular softwares.

Reference:

1. Genes and genomes by Singer.M, and Berg. P, Blackwell Scientific Publication, Oxford, 1991
2. Discovering Genomics, Proteomics, and Bioinformatics by A.M. Campbell and L.J. Heyer.2003
3. Biological database modeling by Jake Chen, Amandeep S. Sidhu Artech House, 2008
4. Principles and applications of modelling by Leach, Second edition April 9, 2001
5. Molecular Evolution and Phylogenetics by Masatoshi Nei, Sudhir Kumar. Oxford University Press, 2000

Course III :ADVANCE IN BIOINFORMATICS

Unit 1

Overview of Bioinformatics application, Scope of Bioinformatics Importance of databases - nucleic acid sequence databases - protein sequence data bases - structure databases bibliographic databases and virtual library - specialized analysis packages

Unit 2

Sequence analysis of biological data- models for sequence analysis and their biological motivation- methods of alignment - methods for optimal alignments; using gap penalties and scoring matrices-multiple sequence alignment – introduction - tools for MSA - application of Multiple Sequence Alignment.

Unit 3

Gene predictions strategies - protein prediction strategies - molecular visualization. Phylogenetic Prediction, Relationship of phylogenetic analysis to sequence alignment, The concept of evolutionary tree, Maximum parsimony methods, Distance methods, Sequence alignment based on an evolutionary model, Complications from phylogenetic analysis.

Unit 4

Drug Discovery and Pharma Informatics, Discovering a drug - target identification and validation - identifying the lead compound - optimization of lead compound - chemical libraries.

Unit 5

Perl Programming: Introduction to Perl, Variables and Data types, Arrays and hashes, Control structures, Subroutines, Pattens and regular expressions, working with files, Perl applications for Bioinformatics, BioPerl.

References:

1. S.C. Rastogi & others, Bioinformatics- Concepts, Skills, and Applications, CBS Publishing, 2003.
2. S. Ignacimuthu, S.J., Basic Bioinformatics, Narosa Publishing House, 1995.
3. T K Attwood, D J parry-Smith, Introduction to Bioinformatics, Pearson Education, 1st Edition, 11 Reprint 2005.
4. C S V Murthy, Bioinformatics, Himalaya Publishing House, 1st Edition 2003.
5. Stephen A. Krawetz, David D. Womble, Introduction To Bioinformatics A Theoretical and Practical Approach, Humana Press, 2003.
6. Hooman H. Rashidi, Lukas K. Buehler, Bioinformatics Basics-Applications in Biological Science and Medicine, CRC press, 2005.
7. Bioinformatics, Biocomputing and Perl By: Michael Moorhouse and Paul Berry John Wiley & Sons publications

Course IV : Area of specialization
1: MOLECULAR MODELING AND DRUG DESIGN

Unit 1

Molecular modeling: Molecular Modeling and Drug Design, Introduction to molecular force fields: General features bond stretching, angle bending, bond Interaction, types of force field. Molecular energy minization. Drug discovery process: drug design target identification and validation. Structure based drug design and ligand based drug design.

Unit 2

Molecular Dynamics: Introduction, Molecular dynamics simulations, Molecular Dynamics Visualization. using simple models. Dynamics with continuous potentials. Constant temperature and constant dynamics. Conformation searching, Systematic search. Applications to protein folding.

Unit 3

Comparative protein modeling: Modelling by Homology the alignment, construction of frame work, selecting variable regions, side chain placement and refinement, validation of protein models Ramchandran plot, threading and *ab initio* modeling

Unit 4

Structure Based Drug Design: Determination of active site and hot spots, Receptor-Ligand interactions, Pharmacophore identification, De novo drug designing. Free energies and solvation, electrostatic and non-electrostatic contribution to free energies, Docking tools Autodock, Argus lab, Gold.

Unit 5

Analog Based Drug Design: Introduction to QSAR: Lead optimization and validation, linear and nonlinear modeled equations, biological activities, physicochemical parameter and molecular descriptors, molecular modeling in drug discovery.

Reference:

1. Principles and applications of modelling by Leach, Second edition April 9, 2001
2. Molecular Modeling by Hans Pieter,Heltje & Gerd Folkens, VCH.
3. Computational Chemistry by Guy H, Grant & W. Graham Richards, Oxford University Press, USA; 1 edition (May 11, 1995)
4. J.M.Haile, Molecular Dynamics Simulation Elementary Methods, John Wiley and Sons, 1997.
5. Satya Prakash Gupta, QSAR and Molecular Modeling, Springer - Anamaya Publishers, 2008.

2. PROTEIN ENGINEERING

Unit 1

Basics of protein science: Classification and properties of proteins, denaturation and coagulation of proteins, protein hydration, isolation, extraction and purification of proteins.

Unit 2

Macromolecular Structure: Protein - Primary, Secondary, Supersecondary, Tertiary and Quaternary structure, Potential energy maps, Ramachandran plot. Macromolecular interactions: Protein – Protein, Protein – Nucleic acids, Structure of Ribosome.

Unit 3

Mechanisms and Kinetics of Enzyme Action: Mechanisms of enzyme action, kinetics of single substrate reactions, turn over number, estimation of Michaelis-Menten parameters, Types of inhibition, Allosteric regulation of enzymes. Applications of Enzymes: Classification of Enzymes, Commercial applications of enzymes in food, pharmaceutical and other industries, Enzymes for analytical and diagnostic applications, Purification and characterization of enzymes from natural sources

Unit 4

Methods to study 3D structure: X-ray, NMR, Cryo-electronmicroscopy - Validation using Procheck, ProsaII - Principles of protein folding and methods - protein folding.

Unit 5

Prediction of protein structure: secondary structure prediction methods - First, second and third generation methods Tertiary structure prediction: Homology modeling, fold recognition and *ab initio* methods

References:

1. Forbes Burkowski. Structural bioinformatics: An algorithmic approach. Publisher: CRC Press, 2009.
2. Drenth Jan. Principles of Protein X-Ray Crystallography. Publisher: Netherlands, Springer Science. 2007.
3. Bourne Philip E., Weissig Helge. Structural Bioinformatics (Methods of Biochemical Analysis, V. 44), 2003. Publisher: Wiley-Liss.
4. Holtje Hans-Dieter, Sippl Wolfgang, Rognan Didier, Folkers Gerd. Molecular Modeling: Basic Principles and Applications. Publisher: New York, Wiley-VCH. 2003.
5. Leach, Andrew. Molecular Modelling: Principles and Applications. Publisher: Prentice Hall. 2001.
6. Friesner Richard A. Computational Methods for Protein Folding:advances in Chemical Physics Volume 120 Kindle Edition. Publisher: New York, John Wiley & Sons. 2002.
7. Heilmeyer L., Friedrich P. Protein Modules in Cellular Signalling. Publisher: Amsterdam, IOS Press. 2001.
8. Fasman G.D.Prediction of Protein Structure and the Principles of Protein Conformation. Publisher: New York, Plenum Press. 1989

Principles of Biochemistry BY A. Lehninger (1987)

3. GENOME TO DRUG AND VACCINE

Unit 1

Genome sequencing - Raw data & Assembly - Genome Databases and related data resources (EST, STS, GSS, HSS etc.) - Organization of data in databases; Genome Data Visualization (With emphasis on Human Genome) - Ensembl – MapViewer.

Unit 2

Tools for Genomic Data Mining - Basic Aspects of Genome Annotation - Prediction of ORFs - Prediction of Promoters splice sites, UTRs etc, Algorithms for gene prediction gene modelling.

Unit 3

Identification of Disease Genes & Drug Targets - (Genetics of Model organisms) - OMIM - Metabolic diseases and Pathogenic diseases - Gene Expression Analysis - Structural Genomics - Functional Genomics - Approaches to novel drug discovery

Unit 4

Pharmacokinetics – Classification - Case study - Comparative screening - ADMET: Drug metabolism; Elimination half-life, Toxicity screening.

Unit 5

Pharmacogenetics - The genetics of drug metabolism- The genetics of therapeutic targets Interactions of small molecules and gene-based drug targets

References:

1. Meyers Robert A. Genomics and Genetics: From Molecular Details to Analysis and Techniques. Volume 1 & 2.. Publisher: Wiley-VCH, 2007.
2. Falciani Francesco. Microarray Technology Through Applications. Publisher: New York, Taylor & Francis. 2007.
3. Primrose Sandy B., Twyman Richard. Principles of Genome Analysis and Genomics 3rd Edition. Publisher: Wiley-Blackwell, 2003.
4. Hiware Chandrashekhar J., Jadhav Babasaheb V., Mohekar Ashok D. Applied Parasitology: A Practical Manual Applied Parasitology: A Practical Manual. Publisher: Jaipur, Mangal Deep Publications. 2003.
5. Lengauer Thomas, Mannhold Raimund, Kubinyi Hugo, Timmerman Hendrik. Bioinformatics: From Genomes to Drugs Volume 1: Basic Technologies. Publisher: Weinheim, Wiley-VCH. 2002.
6. Baxevanis Andreas D., Ouellette B. F. Francis. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins 2nd Edition. Publisher: Singapore, John Wiley & Sons, 2002.
7. Licinio Julio Licinio, Wong Ma-Li. Pharmacogenomics: The Search for Individualized Therapies. Publisher: Germany, Wiley-VCH. 2002.

4. **GENOMICS**

Unit 1

Genes, Genomes, evolution of genomes, Structure & organization of Prokaryotic and Eukaryotic genome, mitochondrial & chloroplast genome, gene structure, SNP's, protein coding genes, repeated sequences.

Unit 2

Physical map of the genome – Chromosome walking and jumping, Restriction enzyme fingerprinting. Sequence tagged sites, radiation hybrid mapping, ESTs. Hybridization assays-Hybridization mapping, Hybridization reference library, *in-situ* hybridization.

Unit 3

Genome information resources, Brief outlook of various genome projects and their outcome, Genome rearrangements, the breakpoint graph, expected reversal distance, signed permutations, interleaving graphs and hurdles, algorithm for sorting by reversals, duality theorem and genomic distance.

Unit 4

Comparative genomics- Genomes of Escherichia coli, Mycoplasma genitalium, Saccharomyces cervisiae, Caenorhabditis elegans, Drosophila melanogaster, Arabidopsis thaliana, Homosapiens, Computational Approaches in comparative genomics.

Unit 5

Genome analysis- sequencing and mapping-physical & high resolution maps, restriction mapping, Introduction to phylogenetic trees, relationship of Phylogenetic analysis to sequence alignment, unweighted pair group method with arithmetic mean (UPGMA), evaluation of phylogenetic methods.

References:

1. Genes and genomes by Singer.M, and Berg.P, Blackwell Scientific Publication, Oxford ,1991
2. Gene Structure and Transcription by Beebe.T, and Burke.T, Oxford Univ Press.
3. Bioinformatics Methods And Applications: Genomics Proteomics And Drug Discovery 3Rd Ed by Rastogi PHI Learning Pvt. Ltd., 2008
4. Metagenomics: Theory, methods and applications, edited by Diana Marco, Caister Academic Press.2010
5. Computational Molecular Biology – An algorithmic approach – Pavel A. Pevzner, 2000.
6. Handbook of Computational Molecular Biology by Srinivas Aluru CRC Press, 21-Dec-2005 - Science

5.PROTEOMICS

Unit 1

Proteins and Enzymes; Proteomics classification; Need, scope and challenges of proteomics, strategies for Protein Separation: 2D gel electrophoresis- principle and applications, Liquid chromatography- principle and applications, multidimensional liquid chromatography

Unit 2

Protein Identification and Quantitation- Strategies for Protein Identification: Protein Identification with antibodies, protein sequence determination by chemical degradation, Mass Spectrometry - principles, instrumentation and applications in proteomics, Strategies for Protein Quantitation: Quantitative proteomics with standard 2D gels, multiplexed proteomics, quantitative with mass spectrometry

Unit 3

Protein structure prediction- primary, secondary and tertiary structure of proteins-SCOP, DALIDD, CATH classification. Interatomic forces and protein structure, covalent interaction, hydrogen bonds, hydrophobic and hydrophilic interaction, charge/dipole interaction, Vander waals forces, steric interaction, Determining protein structure with X ray Crystallography & NMR spectroscopy.

Unit 4

Structural Proteomics**-** Protein structures, techniques for solving protein structures, Protein- protein interactions- principles and methods to study them, Protein Modification in Proteomics.

Unit 5

Protein Databases-Proteins; Protein sequence databases; primary databases and secondary databases, different formats of databases, composite databases.

Reference:

1. R. M. Twyman, Principles of Proteomics (Advanced Text Series), Bios Scientific, 2004.
2. David W Mount, Bioinformatics- Sequence and genome analysis, Cold Spring Harbor Laboratory Press, second edition, 2004.
3. S. R. Pennington, M. J. Dunn, Proteomics: from Protein Sequence to Function, Springer publications, first edition, 2001.
4. Timothy Palzkill, Proteomics, Springer, 2002.
5. Bioinformatics; Methods and applications; Genomics, Proteomics and Drug Discovery; (Rastogi, S. C. and Mendiratta and Rastogi, P.2013