SYLLABUS FOR
POST GRADUATE DIPLOMA IN BIOINFORMATICS

A) Preamble:

Bioinformatics is a newly emerging branch in the field of life science. Bioinformatics is information technology applied to the management and analysis of biological data with the aid of computers. It is the science of using information to understand biology. It is a field in which biological information collected, compared, studied and analyses to find the interrelation between them for solving structural, functional and evolutionary problems using computational technologies. The biological information stored in various databases is available online through internet.

Bioinformatics refers to the creation and advancement of databses, softwares, computational and statistical techniques and theory to solve problems generated from the management and analysis of biological data. On the other hand, computational biology refers to the hypothesis based investigation of a specific biological problem using computers, carried out with experimental or simulated data, with the primary goal of discovery and the advancement of biological knowledge.

Bioinformatics solves the following problems and put more emphasis on understanding the disease related problems at molecular level.

- Protein sequencing, Nucleic acid sequencing and their analysis.
- Find proteins their interaction, activity, modification and function.
- Elucidation of function of a molecule based on its structure.
- Gene expression, analysis, prediction and establish genomic library.
- Find homology for studying evolutionary relationship among different species.
- Molecular modeling and molecular dynamics methods to study structure from sequence.
- Drug designing and discovery from data of functional genomics and proteomics.

In the recent years in this age of Internet and sequenced genome we have more information at our fingertips than ever before. Organizing this entire data and combating information overload is becoming more and more important. The advent of genetic engineering vastly increased size of the information. It is thus necessary for institutes like university evolve a system, which is most accurate and more student friendly. Keeping this view in mind we have decided to start a Post Graduate Diploma Course in Bioinformatics. After completing this one year diploma course students can be accommodated in any national/multinational drug designing pharmaceutical company. This course will provide common basic knowledge of Bioinformatics including Mathematics, Statistics, Computer Science, and Biochemistry. This will be the first of its own kind course in the southern part of Maharashtra. In a nutshell the following are the objectives of this course.

B) Objectives:

This course is expected to bring direct benefit to students of this university, strengthening on going University research in the area of life sciences.

i) To develop an expert manpower to help bioinformatics industry, academia and thereby society.

ii) Protein structure prediction and gene prediction methods.

iii) Drug designing and discovery.

iv) Create an advanced research facility to carry out research in frontier areas of bioinformatics and computational biology.

C) Course Fee: Rs. 20,000/- per year (examination and other fee will be applicable as per university norms).

D) Intake capacity : 20 students

E) Eligibility: B. Sc./ M. Sc. degree in Science (Physical, Mathematical, Chemical Computational, Statistics and Life Sciences), pharmacy, medical, engineering, agriculture and veterinary science.

F) Exam Pattern: Annual Pattern (80 External/20 Internal evaluations).

G) Teachers Qualification: M. Sc. Biochemistry/Biotechnology/Microbiology/ computer science with knowledge in Bioinformatics.

H) Teaching Assistant: M. Sc. Biochemistry/Biotechnology/Microbiology preferably doing research in Bioinformatics.

I) Course Structure: One year duration; Four Theory Papers and two practicals.

BIOINF 141: INFORMATICS SCIENCES AND DATABASES
BIOINF 142: GENERAL BIOLOGICAL CHEMISTRY
BIOINF 241: GENOMICS AND PROTEOMICS
BIOINF 242: STRUCTURAL BIOINFORMATICS AND DRUG DISCOVERY
BIOINF LAB I 141:
BIOINF LAB II 142:

………600 MARKS

BIOINF 141: INFORMATICS SCIENCES AND DATABASES: (60)

UNIT I: (15)

INTRODUCTION TO COMPUTERS:

History and development of computers; generations of computers; (I, II, III, IV and V), classifications of computers; analog computers, digital computers, mainframe computers, miniframe computers, microcomputers, fundamentals of logical concepts.

DIGITAL COMPUTERS:

Basic principle of operation of digital computers, structure of digital computers; arithmetic unit, central unit, memory unit, Input unit and output unit.

COMPUTER CODING:

Number system, decimal number system, binary number system, binary to decimal conversion, Binary arithmetic, octal number system, hexadecimal number system.

UNIT II: (15)

LANGUAGES AND FLOW CHARTS AND OPERATING SYSTEMS:

Machine level languages, assembly level languages, high level languages.

INPUT OUTPUT AND MEMORY:

Punched card reader, paper tape reader, magnetic tape, floppy disk, magnetic disk, optimal scanner, voice data, entry terminal, teleprocessing terminal, visual display unit, modern input devices, Output devices; CRT, printer, plotter.

MEMORY: Primary memory or main memory; magnetic core memory, semi-conductor memory, RMA, ROM, PROM, EPROM, EEPROM. Secondary memory or auxillary memory or storage devices; Hard disk, discket, magnetic tape, ZIP, devices, digital tape, CD-ROM, DVD, virtual , memory, catche memory.

OPERATING SYSTEMS:

DOS, windows 98/XP/VISTA, UNIX/LINUX, Mac OS, VMS.

UNIT III: (15)

MODERN COMPUTERES: Workstations, parallel processing computers, super-computers.

INTERNET AND RELATED PROGRAMMES: WWW, HTML, HTTP, telnet, FTP, computer domain.

INTRODUCTION TO BIOSTATISTICS:

Applications and uses, sample variable, statistical sampling, population, primary and secondary data, screening and representation of data, frequency distribution, bar diagram, histogram. Pie diagram, cumulative frequency curves. Mean, median, mode, comparision between mean median and mode, measures of dispersion: range, variation, standard deviation, coefficient of variation, symmetry, probability distribution.

UNIT IV: (15)

BASIC MATHEMATICS:

SETS: Finite set, infinite set, null or void set, subset, Intervals; closed and open, universal set, operations of set. Relations and functions.

MATRICES: Types of matrices, properties of matrices, addition, substraction of matrices, matrix, multiplication, elementary transformation, inverse of matrices.

DETERMINANTS: Definition, properties associated with determinants. Trignometry and Derivatives.

INTRODUCTION OF DATABASES:

Concept of data, data models, data representation, mining, various types of databases, biological data and data analysis. Related programs; Oracle, SQL, VB, Database management System (DBMS). Introduction to protein and nucleic acid databases.

References:

1) Methods of Biostatistics by T. Bhaskara Rao.
2) Biostatistics- An indroductory text by Auram Gold Stein.
3) Introduction to database system by J. M. Martin, Prentice-Hall.
4) Developing Bioinformatics Computer Skills. (Cynthia Gibas and Per Jambeck).

BIOINF 142: GENERAL BIOLOGICAL CHEMISTRY: **(60)**

UNIT I: (15)

PROTEINS:

Amino acids structure and general properties, alpha helix and beta sheet, tertiary and quaternary structures. Classification of proteins; based on size, shape, biological functions (Enzyme, transport, storage, contractile, structural, defense and regulatory. Structure of peptide bond, restricted rotation, cis-trans, bending, Ramchandran plot, Peptides. Protein folding, dynamics of protein folding, membrane proteins, techniques for studying primary structure of proteins, prediction of conformation from amino acid sequence, ubiquitination, proteosome.

UNIT II . (15)

CELL ORGANELLES**:**

Cell as a basic unit of life, cell organization of prokaryotic and eukaryotic cells, structural and functional compartalization of cell; mitochondria, chloroplast, lysosomes, cell wall, nucleus and nucleolus, protein translocation in ER and vesicular transport into Golgi.

ENZYMES:

Classification of enzymes, IUB system, enzyme substrate complex, active site of enzymes, stereo specificity and ES complex formation, effect of temperature and pH and substrate concentration on reaction rate, activation energy, transition state theory, enzyme activity. Michaelis-Menten equation, significance of Vmax and Km, enzyme inhibition, types of inhibitors and mode of action. Chemical modification of enzymes, structure and functions of ribonuclease, trypsin, chymotrypsin, enzyme regulation, feedback control, product inhibition.

UNIT III (15)

NUCLEIC ACIDS**:**

Bases, nucleosides and nucleotides, purines and pyrimidines, synthesis of nucleic acids, experimental evidence of nucleic acids as genetic material, DNA and RNA, Watson-Crick model of DNA, A, B, & Z forms of DNA, Tm and its relation to GC content, chemical and enzymatic degradation of nucleic acids. DNA binding proteins, RNA-protein interactions.

CENTRAL DOGMA OF MOLEULAR BIOLOGY**:** Introduction to DNA Replication, transcription and translation.

UNIT IV: (15)

A) CARBOHYDRATES:

Classification, monosaccharides, disaccharides, polysaccharides, epimers, isomers, anomers, chiral carbon atoms, chair and boat forms, glycolysis, gluconeogenesis, glycogenolysis, glycopyranose, fructopyranose.

B) LIPIDS:

Definition and classification of lipids, fatty acids, general formula, nomenclature and chemical properties, structure function and properties of simple, complex, acylglycerols, phosphoglycerides, waxes, terpins, steroids and postaglands.

References:

1. Cells by David Prescott.
2. Cell structure and function by Loewy and Gallant.
3. Lehninger's Principles of Biochemistry by Nelson, D. L. and Cox, M. M.
4. Biochemistry by Lubert Stryer.
5. Principles of Protein Structure by Shulz and Schimer.
6. Fundamentals of Enzymology by Price and Steavenson.
7. Complex Carbohydrate by Nathan Sharon.

BIOINF 241: GENOMICS AND PROTEOMICS: (60)

UNIT I: (15)

GENOMICS: NUCLEOTIDE SEQUENCE DATABASES AND ANALYSIS

Goals of the Human Genome Project, cloning vectors, concept of maps, physical maps, shotgun libraries, DNA polymorphism, nucleotides, DNA sequences. Sequence databases: GeneBank, EMBL Nucleotide sequence databank, DNA Data Bank of Japan (DDBJ), database formats. Recombinant DNA technology, restriction enzymes, resource for restriction enzyme (REBASE), similarity search. Polymerase chain reaction, primer selection for PCR, BLASTn, application of BioEdit.

GENOMICS: GENE IDENTIFICATION:

Genome information and special features, coding sequences (CDS), untranslated regions (UTR's), cDNA library, expressed sequence tags (EST). Approach to gene identification; masking repetitive DNA, database search, codon-bias detection, detecting functional sites in the DNA. Internet resources for gene identification, detection of functional sites, gene expression.

UNIT II: (15)

GENE EXPERIENCE AND DNA MICROARRAY:

Introduction, Basic steps for gene expression, concept of microarrays; spotted arrays, oligonucleotide arrays, designing the experiment, Two-color microarray experiments, Clustering gene expression profiles; Agglomerative, Hierarchical, Nearest neighbour (Single-linkage), complete-linkage, average-linkage, weighed pair-group average, k-means clustering, self-organizing maps (SOM) clustering, Tools for microarray analysis; soft-finder, xCluster, MADAM, SAGE, Microarray design, microarray experimentation, fabrication computational analysis of Microarray data, Applications of microarray technology.

PROTEOMICS: PROTEIN SEQUENCE DATABASES AND ANALYSIS:

Protein sequence information, composition and properties, physicochemical properties based on sequence, sequence comparison, Primary databases, Secondary databases. Pair-wise sequence alignment, gaps, gap-penalties, scoring matrices, PAM250, BLOSUM62, local and global sequence alignment, multiple sequence alignment, useful programs, ClustalW, BLASTp.

UNIT III: (15)

PROTEIN MICROARRAY:

Proteomics classification; Tools and techniques in proteomics; 2-D gel electrophoresis, gel filtration, PAGE, isoelectrick focusing, affinity chromatography, HPLC, ICAT, fixing and spot visualization, Mass spectroscopy for protein analysis, MALDI-TOF, Electrospray ionization (EST), Tandem mass spectroscopy (MS/MS) analysis; tryptic digestion and peptide fingerprinting (PMF), Protein Micro array in protein expression, profiling and diagnostics, drug target discovery. Database searching, 3-dimensional structure determination by X-ray and NMR.

PHYLOGENETIC ANALYSIS:

Evolution, elements of phylogeny, methods of phylogenetic analysis, Phylogenetic tree of life, comparison of genetic sequence of organisms, phylogenetic analysis tools-Phylip, ClustalW.

UNIT IV: (15)

BIODIVERSITY INFORMATICS:

Definition, genetic diversity, species diversity, ecosystem diversity. Biodiversity over geographic scales, α-diversity, β-diversity, γ-diversity. Distribution of biodiversity, hotspots of biodiversity, ecological and economic role of biodiversity, scientific role and environmental economics. Threats to biodiversity.

APPLICATIONS OF BIOINFORMATICS IN VARIOUS FIELDS:

Environment, biotechnology, molecular biology, neurobiology, agriculture, drug designing, biomedical genome medicines, medical microbiology.

References:

1. Introduction to Bioinformatics, (Atwood, T. K. and Parry-Smith, D. J).
2. An introduction to Computational Biochemistry. (C. Stain Tsai, A. John Wiley and Sons, Inc., publications).

3. Bioinformatics; Methods and applications; Genomics, Proteomics and Drug Discovery; (Rastogi, S. C. and Mendiratta and Rastogi, P.
4. Bioinformatics; A practical guide to the analysis of genes and proteins.; Edited by, Andreas D. Baxevanis and Francis Oulelette.
5. http://www.ncbi.nlm.nih.gov

BIOINF 242: STRUCTURAL BIOINFORMATICS AND DRUG DISCOVERY:

(60)

UNIT I **(15)**

STRUCTURAL BIOLOGY:

Nucleic acid structures, RNA folding, RNA loops, conformational study, various ribose ring conformations, ribose-ring puckering, protein-protein interactions, protein-ligand interactions, DNA-binding proteins, RNA-binding proteins, Ramachandran plot, , 3-dimensional structures of membrane proteins, importance of 3^{10} helix and loops, biophysical aspects of proteins and nucleic acids.

STRUTURAL DATABASES:

Structural databases; Protein Data bank (PDB), Nucleic Acid Data Bank (NDB), Molecular modeling Data Bank (MMDB). Secondary structure, three-dimensional structure prediction, protein folding and functional sites, protein folding classes.

UNIT II **(15)**

A) PROTEIN STRUCTURE PREDICTION:

Protein Structure Prediction; Homology modeling, prediction of protein structure from sequences, functional sites, Protein folding problem, protein folding classes, protein identification and characterization; AACompIdent, TagIdent, PepIdent and MultiIdent; PROSEARCH, PepSea, PepMAPPER, FindPept, Predicting transmembrane helices, Primary structure analysis and prediction, Secondary structure analysis and prediction, motifs, profiles, patterns and fingerprints search. Methods of sequence based protein prediction.

UNIT III **(15)**

B) MOLECULAR MODELING:

Introduction, force field, quantum chemistry, Schrödinger equation, potential energy functions, energy minimization, local and global minima, saddle point, grid search, , various approximations; LCAO, HF, semi-empirical calculations; single point calculations, full-geometry optimization methods, ZDO, MNDO, CNDO, NDDO, AM1, PM3, RM1, conformational search, Z-matrix, docking, molecular modeling packages.

MOLECULAR MECHANICS:

Definition, balls and springs, force fields, bond-stretching, bond-bending, dihedral motions, out of plane angle potential, non-bonded interaction, coulomb interactions, conformational search, united atoms and cut-offs, Derivative methods; First-order methods; Steepest descent, conjugate gradient, Second order methods; Newton-Raphson method.

UNIT IV **(15)**

MOLECULAR DYNAMICS:

Introduction, Newton's equation of motion, equilibrium point, radial distribution function, pair correlation functions, MD methodology, periodic box, algorithm for time dependence; leapfrog algorithm, Verlet algorithm, Boltzman velocity, time steps, duration of the MD run. Starting structure, analysis of MD job, uses in drug designing, ligand protein interactions.

DRUG DISCOVERY AND COMPUTER-AIDED DRUG DESIGNING:

Introduction, drug discovery area, pharmacogenetics and pharmacogenomics applications, SNPs, parameters in drug discovery identification of drug target molecules, drug design and its approaches, computer-aided drug designing methods; computer aided molecular design (CAMD), Quantum CAChe and project leader, ligand design methods, docking programs; De novo design.

References:

1. An introduction to Computational Biochemistry. (C. Stain Tsai, A. John
   Wiley and Sons, Inc., publications).

2. Bioinformatics; Methods and applications; Genomics, Proteomics and Drug Discovery; (Rastogi, S. C. and Mendiratta and Rastogi, P.
3. Developing Bioinformatics Computer Skills. (Cynthia Gibas and Per Jambeck).

BIOINF Lab 141: **(60)**

1. Computer basic knowledge; hardware, connection cables, typing, Windows 98/XP, Internet browsers, search engines.
2. LAN connections, setting up the IP address, network security.
3. Internet surfing and searching information, downloading and installing software (Acrobat Reader, etc).
4. Hands on session with Microsoft Word, Microsoft Excel (Spreadsheet Application).
5. Hands on session with Microsoft Access (Database related applications).
6. Creation of computer presentation with MS Power Point.
7. Introduction to Oracle for creation of database.
8. Learning C language.
9. Introduction to literature database at NCBI and querying the PUBMED database using the ENTREZ search engine.
10. Getting the amino acid sequences by exploring and querying the protein sequence database.
11. Getting the gene sequences by exploring and querying the nucleic acid databases.
12. 3-D Protein structure visualization and measurement of bond length, bond angle and torsion angles using RasMol.
13. Secondary and tertiary structural analysis.

PROJECT WORK:

BIOINF Lab 242. (60)

1. Using RasMol through command line.
2. Quaternary structural analysis.
3. Investigation of molecular interactions using the program KineMage.
4. Similarity search using the Blast and interpretation of the results.
5. Pair-wise and multiple sequence sequence alignment by using ClustalW.
6. Introduction of BioEdit.
7. Phylogenetic analysis using web tools.
8. Protein Structure Prediction (Homology Modeling) using SPDBV.
9. Molecular modeling using SPARTAN.
10. Model Building and Energy minimization.
11. Quantum chemical and molecular mechanics practicals.
12. Basic UNIX commands, pine, telnet, ftp.
13. Molecular dynamics simulation using GROMACS etc.
14. Molecular Docking and Drug designing by using Chimera.

PROJECT WORK: