Syllabus for Paper I – Basic: (All four sections)

1. BINC Biology Syllabus
2. BINC Mathematics, Statistics, Physics and Chemistry Syllabus
3. BINC Information Technology Syllabus
4. BINC Bioinformatics Syllabus

BINC Biology Syllabus – Basic

Cell Biology

- Basic aspects of Prokaryotic and eukaryotic cells (plant and animal cells); membranes and cellular compartments, cell organelles, structure and function
- Cell motility and shape: cytoskeletal elements, cilia and flagella; motor proteins
- Cell-cell interactions: Intercellular junctions
- Photosynthesis, transportation of proteins in cells, transpiration, Electron transport chain
- Cell cycle and its regulation; events during mitosis and meiosis

Genetics

- Mendelian principles of inheritance, sex linked inheritance
- Concept of linkage, linkage maps and recombination
- Mutations – molecular, point and chromosomal mutations, hotspots
- Phenotype and genotype relationships, role of environment, from gene to phenotype, gene interactions. Study of quantitative traits
- Genetics of populations, genetics and evolution

Immunology

- Immune systems: Innate and adaptive immunity in vertebrates
- Antigen processing and presentation
- Antibodies: Immunoglobulins, Immunoglobulin classes and subclasses, CDR and LDR regions and sequence numbering
- Concepts of generation of diversity and specificity in immune system; Immunological methods

Molecular Biology

- Prokaryotic genome organization and structure
- Prokaryotic gene expression, factors involved in gene regulation
- Eukaryotic genome organization and structure, Mechanism of gene expression in Eukaryotes, Basic mechanism of transcription and translation
- Mechanisms by which genome undergoes changes, recombination, mutation, inversion, duplication, transposition

Biochemistry

- Carbohydrates and lipids, their importance in cells
- Proteins: Amino acids and their physiocochemical properties, peptide bond and peptides

- Nucleic acids: nucleosides, nucleotides, RNA and DNA. Denaturation and renaturation of DNA
- 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, multienzyme complexes, multifunctional enzymes and regulatory enzymes
- Concept of biochemical regulation, feed back and feed forward systems, biochemical oscillations

Mathematics Statistics Physics and Chemistry Syllabus – Basic

- Functions and Graphs: Functions, Relations, notation and representation. Graphs. Review of basic functions. Functions of several variables
- 2D coordinate geometry: Equation of a line, circle, ellipse, parabola, hyperbola
- 3D geometry: Equation of sphere, cone, direction cosines, equation of line
- Basic trigonometric functions
- Matrix algebra: Addition, subtraction, multiplication, transpose
- Introduction to principles of statistical sampling from a population, random sampling
- Frequency distributions and associated statistical measures, Probability distributions – normal and binomial
- Particle dynamics, Newton's laws of motion, velocity, acceleration, momentum
- Pressure, temperature, volume relationship
- First law of thermodynamics, isothermal process, entropy and second law of thermodynamics, reversible and irreversible processes; Concepts of enthalpy, internal energy and potential energy; Inter-relation between potential energy and force
- Concept of pH, pK, chemical equilibrium, Henderson-Hasselbach equation, structure of water, chemical forces, hydrophilic and hydrophobic forces, hybridization states of atoms, electronic structure of molecules, and concept of bonding (chemical bonds, ionic bonds, covalent bonds, hydrogen bond, coordinate bonds)

Information Technology Syllabus – Basic

Concepts in Computing

- Overview and functions of a computer system
- Input and output devices
- Storage devices: Hard Disk, Diskette, Magnetic Tape, RAID, ZIP devices, Digital Tape, CD-ROM, DVD, etc (capacity and access time)
- Main Circuit Board of a PC: Chips, Ports, Expansion slots, etc
- Memory: Register, buffer, RAM, ROM, PROM, EPROM, EEPROM (comparison)
- Types of Processing: Batch, Real-Time, Online, Offline
- History of Computers: Evolution, Generation of computers (I, II, III, IV, V), Classification of computers (mainframes, mini computers, microcomputers, special purpose) – comparison with memory, power, cost, size – then and now
- Types of modern computing: Workstations, Servers
- An overview of computer viruses: What is a virus? Virus symptoms, How do they get transmitted? What are the dangers, General Precautions?

- Introduction to operating systems: Operating System concept, Windows 2003/XP, Windows Vista, UNIX/LINUX
- The Internet and its Resources, World Wide Web (WWW): Associated tools, services, resources and various terminologies
- Programming in C

Introduction to Database Systems

- Concepts of various types of databases
- Data Abstraction
- Data Models
- Instances and Schemes
- E-R Model: Entity and entity sets; Relations and relationship sets; E-R diagrams; Reducing E-R Diagrams to tables
- Network Data Model: Basic concepts
- Hierarchical Data Model: Basic Concepts
- Text Databases
- Multimedia Databases – Basic Concepts and Applications; Indexing and Hashing

Bioinformatics Syllabus

Major Bioinformatics Resources: NCBI, EBI, ExPASy, RCSB

The knowledge of various databases and bioinformatics tools available at these resources, organization of databases: data contents and formats, purpose and utility in Life Sciences

Open access bibliographic resources and literature databases:

Open access bibliographic resources related to Life Sciences viz., PubMed, BioMed Central, Public Library of Sciences (PloS), CiteXplore

Sequence databases: Formats, querying and retrieval; Nucleic acid sequence databases: GenBank, EMBL, DDBJ; Protein sequence databases: Uniprot-KB: SWISS-PROT, TrEMBL, UniParc; Repositories for high throughput genomic sequences: EST, STS GSS, etc.; Genome Databases at NCBI, EBI, TIGR, SANGER – Viral Genomes; Archeal and Bacterial Genomes; Eukaryotic genomes with special reference to model organisms (Yeast, Drosophila, C. elegans, Rat, Mouse, Human, plants such as Arabidopsis thaliana, Rice, etc.)

Structure Database: PDB, NDB, PubChem, ChemBank

Derived Databases
Knowledge of the following databases with respect to: basic concept of derived databases, sources of primary data and basic principles of the method for deriving the secondary data, organization of data, contents and formats of database entries, identification of patterns in given sequences and interpretation of the same – Sequence: InterPro, Prosite, Pfam, ProDom; Structure: FSSP, DSSP

Extraction of knowledge from resources on Immunology, Plant, animal and infectious diseases: databases and servers published in the NAR Database and Web server Issues and other Bioinformatics journals viz. BMC Bioinformatics etc.

Sequence Analysis

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

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

Database Searches: Keyword-based Entrez and SRS; Sequence-based: BLAST & 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

Pairwise sequence alignments: 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

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 dandrogram 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, analysis and interpretation of profile-based searches

Taxonomy and phylogeny: Basic concepts in systematics, taxonomy and phylogeny; molecular evolution; nature of data used in Taxonomy and Phylogency, Definition and description of phylogenetic trees and various types of trees

Protein and nucleic acid properties: Computation of various parameters using proteomics tools at the ExPASy server, GCG utilities and EMBOSS

Comparative genomics: Basic concepts and applications, whole genome alignments: understanding significance. Artemis as an example

Structural Biology

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, protein folding DNA and RNA: types of base pairing – Watson-Crick and Hoogstein; 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

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

Structure analysis and validation: PDB Goodies, Procheck, ProsaII, PDBsum

3-D structure visualization and simulation: Visualization of structures using Rasmol or SPDBV or CHIME or VMD
Basic concepts in molecular modeling: different types of computer representations of molecules. External coordinates and Internal Coordinates
Concepts of force fields: representations of atoms and atomic interactions, potential energy representation

Classification and comparison of protein 3D structures:
Purpose of 3-D structure comparison and concepts, Algorithms such as FSSP, CE, VAST and DALI, Fold Classes
Databases of structure-based classification; CATH and SCOP

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

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
Fundamentals of docking small and macromolecules to proteins and nucleic acids

Xxxxxxxxxxxxxxx

Syllabus for Paper II – Advanced: (All four sections)

1. BINC Biology Syllabus
2. BINC Mathematics, Statistics, Physics and Chemistry Syllabus
3. BINC Information Technology Syllabus
4. BINC Bioinformatics Syllabus

Biology Syllabus – Advanced

Cell Biology and Genetics

- Vesicular transport and protein traffic in cells
- Different mechanisms of signal transduction, concepts in signal network, second messenger, molecules involved in various signaling pathways such as G-protein coupled receptors, protein kinases, calcium binding proteins

- X-linked and autosomal diseases, mitochondrial related disease, QTL methods for diagnostics
- Extra-chromosomal inheritance
- Immune response, autoimmune disorders, ELISA method
- Molecular genetics and genetic disorders

Molecular Biology

- Genome organization, Initiation, elongation and termination of transcription-template and enzyme properties, promoter and regulator sequences. Regulation of translation, Post transcriptional modifications
- Methods for studying gene expression and regulatory sequences, large-scale expression analysis, use of microarrays
- Genetic information transfer, details of regulation in eukaryotes and prokaryotes, horizontal gene transfer
- Operons – positive and negative regulation, Processing of RNA and Proteins – Transport and Stability
- Organization of eukaryotic genome, Methods for studying variation and polymorphism at genome level, PCR, northern, southern, western blotting, RFLP, fingerprinting, RAPDs, DNA and protein sequencing methods
- Epigenetic mechanisms of inheritance regulatory RNA molecules (RNA; miRNA, siRNA), antisense RNA and their applications

Biochemistry

- Enzyme kinetics, Lineweaver-Burk plot, Competitive and non competitive inhibition
- Molecular mechanisms of interactions of small and large molecules including ions, regulation of protein pathways, mechanism of enzyme action, ribozyme and abzymes
- Isoenzymes, allosteric enzymes, regulation by covalent modification
- Carbohydrate metabolism: Glycolysis, gluconeogenesis, glycogenolysis, glycogenesis, TCA cycle and oxidative phosphorylation
- Pentose phosphate pathway; hormonal control, ß-oxidation and biosynthesis of fatty acids
- Transamination and deamination of amino acids, ketogenic and glycogenic amino acids, urea cycle
- Purine and pyrimidine biosynthesis

Mathematics, Statistics, Physics and Chemistry Syllabus – Advanced

- Number integration. Interpolation and approximate methods
- Vector – addition, subtraction, dot, cross, scalar triple product, divergence and curl.
- System of linear equations. Matrix inverse, eigenvalue, eigenvector, principalm component analysis
- Mathematical modeling and simulation
- Methods of least squares, chi-square test, systematic and random sampling, accidental and systematic errors, correlation and regression analysis. Poisson and extreme value distributions
- Multivariate analysis, Hypothesis testing, Markov process
- Bayesian Statistics
- Basics of classical mechanics and quantum mechanics
- Laws of motion
- Refraction of light, focal length of lens, magnification. Definition of resolution, optical and electron microscope
- Principles of lasers
- Luminescence, fluorescence and phosphorescence (basic concepts and applications)
- Non-covalent bonding in protein structure

- Biophysical techniques for determining size and shape of macromolecules – ultra centrifugation, electrophoresis and chromatography. Application of spectroscopy (fluorescence and absorption spectroscopy) and X-ray diffraction for determination of biomolecular secondary and tertiary structure – CD, NMR, X-ray crystallography, mass spectroscopy of biological molecules
- Basic principle of chemical kinetics – Zero order and first order kinetics, energy of activation. Reversible and irreversible thermodynamics

BINC Information Technology Syllabus – Advanced

- Local area networking, network devices, IP address, computational cluster
- Parallel processing/computing, cluster computing, grid computing, etc
- Java and Perl programming
- Introduction to distributed database processing; understand, appreciate and implement relational database design

SQL and front end development

- Select statement
- Data definition statements
- Data manipulation statements
- Data control statements
- Other database objects: Views, Sequences, Synonyms
- Application development using visual basic
- Working with code and forms
- Variables, procedures and controlling program executor
- Standard controls
- Data access using Data control
- Connecting to Oracle database using visual basic
- Using Oracle DBMS as backend, SQL skills and basic skill in using VB as a front end

Computer graphics and visualization

- Introduction
- Scientific and Engineering opportunities
- Visualization techniques: Software, Hardware, Color representation – RGB, CMY, gray-scale
- Interactive graphics
- Interaction devices and techniques
- Geometric transformations
- Viewing in three dimensions. Stereo-pairs, perspective, depth-cue
- Rendering
- Standards – CGI, GKS, PHIGS

Programming in C

Concepts of flowcharts, algorithm development, pseudo codes etc

Computer assignments based on the following topics in 'C' programming: Data types, operatos and expressions, Hierarchy of operators, control statements including decision (if, if-else), loops (while, do-

while, for), branching (switch, break, continue), functions, arrays (1D, 2D- all matrix operations including inverse of a matrix), strings, Pointers, file handling, data structures etc,

OR

Programming in Object Oriented Languages
JAVA

- An introduction to JAVA programming
- Object-oriented programming and Java
- Java basics
- Working with objects
- Arrays, Conditionals and Loops
- Creating classes and applications in Java
- More about methods
- Java applets basics
- Graphics, Fonts and Color
- Simple Animation and threads
- Advanced animation, images and sound
- Managing simple events and interactivity
- Creating user interfaces with AWT
- Modifiers, Access control and class design
- Packages and Interfaces
- Exception
- Multithreading
- Streams and I/O
- Using Native methods and libraries
- Java programming tools
- Working with data structures and Java
- Image filters,

OR

Perl

- What is Perl? Why use Perl in Bioinformatics? History of Perl, Availability, Support, Basic concepts.
- Scalar data: What is Scalar Data? Numbers, strings, scalar operators, scalar variables, scalar operators and functions
- Arrays and list data: What is a list or array? Literal representation, variables, arrays operators and functions, scalar and list context
- Control structures: Statement blocks
- Hashes: What is a Hash? Hash variables, Literal representation of a Hash, Hash Functions, Hash Slices
- Basic I/O
- Regular expressions: Concepts about regular expressions, simple uses of regular expressions, patterns, matching operator, substitutions, the split and join functions

- Subroutines: System and user functions, the local operator, variable-length parameter lists, lexical variables
- Miscellaneous control structures
- Filehandles and file tests: What is a filehandle? Opening and closing a filehandle, using pathnames and filenames, die, using filehandles, The –x file tests, the stat function
- Formats: What is a format? Defining a format, invoking a format
- Directory access: Directory tree, globbing, directory handles, opening and closing a directory handle, reading a directory handle
- File and directory manipulation
- Process management: Using system and exec, using backquotes
- Other data transformation: Finding a substring, extracting and replacing a substring
- Formatting data: Sorting, Transliteration
- System information: Getting User and Machine information, Packing and Unpacking Binary data, getting network information
- Database manipulation: DBM databases and DBM Hashes, Opening and closing DBM Hashes, Fixed-length random-access databases, Variable-Length (Text) Databases, Win32 Database Interfaces
- CGI programming: The CGI.pm Module, CGI program in context, simple CGI programs, passing parameters via CGI, Perl and the Web
- Object oriented perl: Introduction to modules, Creating Objects
- Bioperl: Introduction, Installation procedures, Architecture, Uses of bioperl

BINC Bioinformatics Syllabus- Advanced

- **Sequence analysis**
- **Scoring matrices:** Detailed method of derivation of the PAM and BLOSUM matrices
- **Pairwise sequence alignments:** Needleman and Wuncsh, Smith and Waterman algorithms and their implementation
- **Multiple sequence alignments (MSA): Use of HMM-based Algorithm** for MSA (e.g. SAM method)

- **Taxonomy and phylogeny:** Phylogenetic analysis algorithms such as maximum Parsimony, UPGMA, Transformed Distance, Neighbors-Relation, Neighbor-Joining, Probabilistic models and associated algorithms such as Probabilistic models of evolution and maximum likelihood algorithm, Bootstrapping methods, use of tools such as Phylip, Mega, PAUP

Sequence patterns and profiles:

- Algorithms for derivation of and searching sequence patterns: MeMe, PHI-BLAST, SCanProsite and PRATT
- Algorithms for generation of sequence profiles: Profile Analysis method of Gribskov, HMMer, PSI-BLAST
- **Protein and nucleic acid properties:** e.g. Proteomics tools at the ExPASy server and GCG utilities and EMBOSS

Structural Biology

- Identification/assignment of secondary structural elements from the knowledge of 3-D structure of macromolecule using DSSP and STRIDE methods
- Prediction of protein structure: PHD and PSI-PRED methods
- Tertiary structure: Detailed protocols/algorithms for Homology modeling, fold recognition and ab-initio approaches
- Structures of oligomeric proteins and study of interaction interfaces

Molecular modeling and simulations

- Macro-molecular force fields, salvation, long-range forces
- Geometry optimization algorithms: Steepest descent, conjugate gradient
- Various simulation techniques: MD, Monte Carlo, docking strategies etc
- Molecular mechanics, conformational searches

Genomics

- 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
- 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)

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

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

Modeling biological systems

- Systems biology – Use of computers in simulation of cellular subsystems
- Metabolic networks, or network of metabolites and enzymes
- Metabolic pathways: databases such as KEGG, EMP
- Study of plant pathways –MetaCyc, AraCyc
- Signal transduction networks
- Gene regulatory networks

Bioinformatics Resources at the species level

- ICTV Database, AVIS, VirGen, Viral genomes at NCBI, VBRC, VBCA, PBRC and Subviral RNA database, Species 2000, TreeBASE etc

Drug 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

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
- 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

Paper III will be computer based practical and will include questions on Bioinformatics.

BioInformatics National Certification (BINC) Examination 2009
Instituted by: Department of Biotechnology, Government of India, New Delhi
Coordinated by: Bioinformatics Centre, University of Pune

Sample Questions for Paper I (Objective type)
Answers are shown in bold type & are underlined

Subject: BIOINFORMATICS

1. What is PROSITE?
A. A database of protein structures
B. A database of interacting proteins
C. A database of protein motifs
D. A search tool

2. You have two distantly related proteins. Which BLOSUM or PAM matrix would you choose to compare them?
A. BLOSUM45 or PAM250
B. BLOSUM60 or PAM1
C. BLOSUM80 or PAM120
D. BLOSUM80 or PAM1

3. What is the difference between RefSeq and GenBank?
A. RefSeq includes publicly available DNA sequences
B. GenBank includes nonredundant curated data
C. GenBank sequences are derived from RefSeq
D. RefSeq sequences are derived from GenBank

4. The two main features of any phylogenetic tree are the
A. clades and the nodes
B. topology and the branch lengths
C. clades and the root
D. alignment and the bootstrap

5. The approach that can be used to predict the 3D structure of a protein which has no detectable sequence similarity with the available templates is
A. homology modeling
B. comparative modeling
C. fold recognition
D.**ab initio modeling**

Subject: Biology

1. Phase of the cell cycle during which replication occurs is
A. G0 phase
B. M phase

C. G1/G2 phase
D. S phase

2. Polycistronic mRNA refers to
A. mRNA which is transcribed by multiple RNA polymerases
B. mRNAs that are simultaneously translated
C. mRNA that is translated by many ribosomes simultaneously
D. mRNA with multiple open reading frames

3. Total number of carbon atoms in lauric acid is
A. 10
B. 12
C. 14
D. 16

4. An enzyme was purified from liver extract. The total activity and total protein were estimated at each step of purification. During this process the
A. specific activity increases and total activity increases
B. specific activity decreases and total activity increases
C. specific activity increases and total activity decreases
D. specific activity and total activity remain unchanged

5. Which pair of amino acids absorbs the most UV light at 280nm?
A. Thr and His
B. Trp and Tyr
C. Cys and Asp
D. Phe and Pro