

BINC SYLLABUS for Paper-III

BINC BioInformatics Syllabus - Basic

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)

Sequence databases

Formats, querying and retrieval

Nucleic acid sequence databases: GenBank, EMBL, DDBJ;

Protein sequence databases: Uniprot-KB: SWISS-PROT, TrEMBL, PIR-PSD

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.)

3D Structure Database: PDB, NDB

Chemical Structure database: Pubchem

Gene Expression database: GEO, SAGE

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, Gene Ontology

Structure classification database: CATH, SCOP, FSSP

Protein-Protein interaction database: STRING

Compilation of resources: NAR Database and Web server Issues and other resources published in Bioinformatics related journals

Sequence Analysis

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

Basic concepts: Sequence similarity, identity and homology, definitions of homologues, orthologues, paralogues

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

Pairwise sequence alignments: Basic concepts of sequence alignment: local and global alignments, 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 dendrogram and its interpretation

Database Searches:

Keyword-based searches using tools like ENTREZ and SRS

Sequence-based searches: BLAST and FASTA

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 Phylogeny, Phylogenetic tree and its reconstruction. **Protein and nucleic acid properties:** Computation of various parameters using proteomics tools at the ExPASy server and EMBOSS

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

Structural Biology

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

Non-Covalent Interactions and their role in Biomolecular structure and function

Fundamentals of Receptor-ligand interactions.

Proteins: Principles of protein structure; Peptide bond, phi, psi and chi torsion angles, Ramachandran map, anatomy of proteins – Hierarchical organization of protein structure – Primary, Secondary, Super secondary, Tertiary and Quaternary structure; Hydrophobicity of amino acids, Packing of protein structure, Structures of oligomeric proteins and study of interaction interfaces

DNA and RNA: types of base pairing – Watson-Crick and Hoogsteen; types of double helices (A, B, Z), triple and quadruple stranded DNA structures, 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

Classification and comparison of protein 3D structures:

Purpose of 3-D structure comparison and concepts, Algorithms: CE, VAST and DALI, concept of coordinate transformation, RMSD, Z-score for structural comparison.

Databases of structure-based classification; CATH, SCOP and FSSP

Secondary structure prediction: Algorithms viz. Chou Fasman, GOR methods; nearest neighbor and machine learning based methods, analysis of results and measuring the accuracy of predictions.

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/comparative Modeling, fold recognition, threading approaches, and *ab initio* structure prediction methods

BINC Bioinformatics Syllabus - Advanced

Sequence analysis

Scoring matrices: Detailed method of derivation of the PAM and BLOSUM matrices

Pairwise sequence alignments: Needleman and Wunsch, Smith and Waterman algorithms and their implementation

Multiple sequence alignments (MSA):

Use of HMM-based Algorithm for MSA (e.g. SAM method)

Sequence patterns and profiles:

Repeats: Tandem and Interspersed repeats, repeat finding, Motifs, consensus, position weight matrices

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, PSIBLAST

Protein and nucleic acid properties: e.g. Proteomics tools at the ExPASy server and EMBOSS

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, Bayesian inference algorithm, Bootstrapping methods, use of tools such as PHYLIP, MEGA, PAUP

Analysis of regulatory RNA's: Databases and tools

Structural Biology

Experimental methods for Biomolecular structure determination: X-ray and NMR

Identification/assignment of secondary structural elements from the knowledge of 3-D structure of macromolecule using DSSP and STRIDE methods

Prediction of secondary structure: PHD and PSI-PRED methods

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 **Structure analysis and validation:** Pdbsum, Whatcheck, Procheck, Verify3D and ProsaII

Critical Assessment of Structure Prediction(CASP)

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: Molecular mechanics, conformational searches, Molecular Dynamics, Monte Carlo, genetic algorithm approaches, Rigid and Semi-Flexible Molecular Docking

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

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.

Next Generation sequencing & assembly: Elements of big data analysis, NGS Platforms based on pyrosequencing, sequencing by synthesis, emulsion PCR approach with small magnetic beads and single molecule real time (SMRT) sequencing; Genome assembly algorithms, *De-novo* assembly

algorithms, Sequence Alignment formats: Sequence Alignment/Map (SAM) format, Binary Alignment/Map (BAM) format.

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: Clusters of Orthologous Groups (COGs), Ensembl

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: supervised/unsupervised learning, 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 STRINGS, DIP, PPI server and tools for analysis of protein-protein interactions

Modeling biological systems

Systems biology – Topology of biological networks: Random vs Scale-Free networks. Use of computers in simulation of cellular subsystems: Simulation and analysis of biochemical networks and their dynamics using ODEs and stochastic algorithm, Flux Balance Analysis (FBA), Boolean network simulations.

Metabolic networks, or network of metabolites and enzymes, Signal transduction networks, Gene regulatory networks, Metabolic pathways: databases such as KEGG, EMP, MetaCyc, AraCyc

Drug design

Drug discovery process

Role of Bioinformatics in drug design

Target identification and validation and lead optimization

Different systems for representing chemical structure of small molecules like SMILES etc

Generation of 3D coordinates of small molecules

Structure-based drug design: Identification and Analysis of Binding sites and virtual screening

Ligand based drug design: Structure Activity Relationship – QSARs and QSPRs, QSAR Methodology, Pharmacophore mapping

In silico prediction ADMET properties for Drug Molecules

Vaccine design:

Reverse vaccinology and immunoinformatics

Databases in Immunology

Principles of B-cell and T-cell epitope prediction

Suggested Books for Reading:

David W Mount, Bioinformatics: Sequence and Genome Analysis, 2nd Edition, Cold Spring Harbor Press

Durbin et al (2007) Biological Sequence Analysis: Probabilistic models of protein and Nucleic acids Cambridge University Press.

Stuart M. Brown (2013) Next-generation DNA sequencing Informatics. Cold Spring Harbor Press

M.E.J. Newman (2010) Networks: An Introduction, Oxford University Press.

Thomas E. Creighton, Proteins: structures and molecular properties

Chemoinformatics Edited by Johann Gasteiger and Thomas Engel

Structural Bioinformatics, Edited Philip E. Bourne and Helge Weissig

Lee A Segel (2008), Biological Kinetics, Cambridge University Press Cambridge

Cornish-Bowden (2012), Fundamentals of Enzyme Kinetics ,Wiley-Blackwell

Alberghina L (2005), System Biology : Definitions and Perspectives, Springer-Verlag Berlin Heidelberg.

Najarian K, Najarian S, Gharibzadeh S, Eichelberger CN (2009) Systems Biology and Bioinformatics: A Computational Approach, CRC Press.

Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H, Herwig R (2009) Systems Biology : A Text Book, Wiley-Blackwell

Integrative approaches for finding modular structure in biological networks, NATURE REVIEWS, GENETICS, VOLUME 14, OCTOBER 2013

BIOINFORMATICS, Vol. 19 no. 2, 2003

Nucleic Acids Research (2014), Vol. 42, Database issue D199–D205 doi:10.1093/nar/gkt1076

Nucleic Acids Research (2012), Vol. 40, Database issue D109–D114 doi:10.1093/nar/gkr988

An extended bioreaction database that significantly improves reconstruction and analysis of genome-scale metabolic networks (2011), Integrative Biology, 2011.3, 1071-1086.

Computational Systems Bioinformatics — Methods and Biomedical Applications By Xiaobo Zhou (Harvard Medical School and Brigham & Women's Hospital, USA), Stephen T C Wong (Harvard Medical School and Brigham & Women's Hospital, USA).

Bioinformatics for Systems Biology (2009) by Stephen Krawetz, Published by Humana Press.

BINC Information Technology Syllabus: Basic

Fundamentals in Computing

Types of Processing: Batch, Real-Time, Online, Offline.

Types of modern computing: Workstations, Servers, Parallel Processing Computing,

Cluster computing, Grid computing

Memory and Storage Devices, Network, Internet-Basics

Introduction to operating systems: Operating System concept, UNIX/LINUX.

Basic Programming Concepts – sequential, conditional and loop constructs, Arrays, Strings, Object Oriented Programming Concepts- Classes, Objects, Inheritance, Polymorphism; File Handling

Introduction to Database Systems- SQL Queries

BINC Information Technology Syllabus: Advanced

- A. Data Structures and Algorithm
Arrays, Link Lists, Stacks, Queues, Graphs, Trees, Sorting, Searching, string comparison
– Programs to be implemented using C or Python or Java or Perl
- B. Databases
SQL, indexing and Hashing.
- C. Elements of scripting languages.
- D. Elements of NoSQL

Suggested Books for Reading:

1. Database Management System – Ramakrishnan and Gehrke
2. Data Structure : Andrew S Tannenbaum
3. Complete Reference to C
4. Complete Reference to Java
5. Complete reference to Perl
6. Complete Reference to Python