## SEMESTER III

**CSB2103 Fundamentals of Molecular Biology and Genetics [3-1-0-4]**

Cell and cell organelles, cell cycle and its regulation, cell division: mitosis and meiosis. Flow of genetic information, Organization of viral, prokaryotic and eukaryotic genome, DNA as the genetic material, DNA replication, transcription and translation in eukaryotes and prokaryotes, regulation of gene expression, RNA processing, post-translational modifications of proteins, DNA repair, Mutation: Types and detection. Basic principles of heredity, Mendelian genetics Genetic terminology, Deviations of Mendel's ratios - Genetic interactions, Epistasis, Pleiotropy, Penetrance and Expressivity, Multiple alleles, Linkage, Crossing over, Chromosome mapping, Gene mapping and Recombination frequencies.

**References**


1. B. Alberts, Molecular Biology of The Cell, 5th Edition, Bruce Alberts et al Garland Science, 2008.
2. H.F. Lodish, Molecular Cell Biology, 9th Edition, W.H Freeman, 2021.
3. G. Karp, Cell and Molecular Biology, 9th Edition, John Wiley and Sons, 2020.
4. T. A. Brown, Gene Cloning and DNA Analysis: An Introduction (7th ed.), Wiley-Blackwell, 2016.
5. B. Klug, M. Cummings, C. Spencer, and M. Palladino, Concepts of Genetics, 12th ed. Boston, MA, USA: Pearson, 2019.
6. D. Peter Snustad and M. J. Simmons, Principles of Genetics, 7th ed. Hoboken, NJ, USA: Wiley, 2021.

**CSB2132: Basic Analytical Lab** **[0-0-2-1]**

**List of Experiments**

1. Introduction to biological laboratory & its instruments
2. Study of Meiosis and Mitosis
3. Isolation of genomic DNA
4. Isolation of RNA
5. Agarose Gel Electrophoresis
6. Quantification of DNA/RNA/Protein
7. Demonstration of Polymerase Chain Reaction
8. Paper/Thin Layer Chromatography
9. Isolation and purification of bacteria from various environments
10. Morphological characterization of the bacteria by staining and non-staining methods
11. Plasmid DNA isolation from bacteria
12. Transformation
13. Verification of Beer - Lambert's law by UV-Vis spectrophotometer.
14. Gene mutation using UV treatment
15. Determination of linkage and cross-over analysis

**References**


1. Sambrook J, Russell D (2001) Molecular Cloning: A Laboratory Manual, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
2. Wilson, K., and Walker, J., Principles and Techniques of Biochemistry and Molecular Biology, (8th ed.), Cambridge University Press, UK, 2018.
3. Christian, G.D., Dasgupta, P.K. and Schug, K. A., Analytical Chemistry (Indian ed.), Wiley India Pvt. Ltd., India, 2020.
4. J. Sambrook and D. W. Russell, Molecular Cloning: A Laboratory Manual, 4th ed. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press, 2012.

## SEMESTER IV

**CSB2204: Bioinformatics** **[3-1-0-4]**

Introduction to Biological Databases, structure and sequence analysis of biomolecules. Sequence Analysis, types of Sequence alignment -Pairwise and Multiple sequence alignment, Global alignment, Local alignment, Dotplot, Different scoring methods, Substitution matrices (PAM and BLOSUM). Dynamic programming, NeedlemanWunsch algorithm, Smith-Waterman algorithm, Multiple Sequence Alignment methods (MSA), Scoring of a MSA, Progressive (CLUSTALW), Iterative and Hidden Markov Model (HMM), FASTA and BLAST algorithms. Phylogenetic tree and terminology, different methods of Phylogenetic tree prediction. Hidden Markov model, Overview of drug designing, molecular concept of disease, drug targets. Molecular and chemical properties of drugs, Small Molecule Databases: PubChem, DrugBank. Ligand selection using Lipinski's rule of five, Molecular Concept of Disease, DrugTargets; Types of Drug targets, Active sites and pharmacophores, allosteric sites, Intermolecular binding force of drugs with targets. Protein interaction database. Primer designing, Basics of artificial neural network, Applications of neural network for nucleotide and protein sequence prediction.

**References**


1. T K Attwood, D J Parry-Smith, Introduction to Bioinformatics, Pearson Education, 1st Edition, 11th Reprint 2005.
2. D.W. Mount, Bioinformatics: Genome and Sequence Analysis, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. 2001
3. A. M. Lesk, Introduction to Bioinformatics, 5th ed. Oxford, UK: Oxford University Press, 2019.
4. J. Pevsner, Bioinformatics and Functional Genomics, 3rd ed. Hoboken, NJ, USA: Wiley-Blackwell, 2015.

**CSB2232: Bioinformatics Lab** **[0-0-2-1]**

1. Explore NCBI, EMBL, DDBJ, and UniProt databases.
2. Perform sequence similarity searches.
3. Fetch and visualize protein 3D structures
4. Motif/Domain database search
5. Calculate GC content, ORFs, and codon usage.
6. Perform global alignment using Needleman-Wunsch algorithm.
7. Use Smith-Waterman algorithm for local alignment.
8. Visualize sequence similarity using dot plots.
9. Compare PAM and BLOSUM matrices in scoring alignments.
10. Perform progressive alignment of multiple sequences.
11. Generate and interpret phylogenetic trees.
12. Predict genes using Hidden Markov Models.
13. Identify protein domains using profile HMMs.
14. Predict secondary structure or splice sites.
15. Primer designing

**References**


1. S. R. Krishna, Bioinformatics: A Practical Approach, 1st ed. Chennai, India: MJP Publishers, 2019.
2. A. Mishra and A. Sharma, Practical Bioinformatics, 1st ed. New Delhi, India: Biotech Press, 2017.
3. J. Z. Kelemen, Practical Bioinformatics, 1st ed. Totowa, NJ, USA: Humana Press, 2003.
4. C. Gibas and P. Jambeck, Developing Bioinformatics Computer Skills, 1st ed. Sebastopol, CA, USA: O'Reilly Media, 2001

## SEMESTER V

**CSB3102: Computational and Structural Biology [3-1-0-4]**

Introduction to next generation sequencing, Transcriptomics: Introduction, types and functions of coding and non- coding RNAs. Commonly used Transcriptomic techniques-EST, SAGE, Microarray, RNAseq, miRbase, etc. Metagenomics: Introduction, Species level classification of metagenomics, strain level profiling and Bio-surveillance using metagenomics. Annotation and prediction of functional genes. Biomarker discovery from metagenomics. Protein structure prediction: Structure determination; Sample preparation for Cryo EM, X-ray crystallography, solution- and solid-state NMR spectroscopy, Single particle, Cryo Electron Microscopy, X-Ray Free-Electron Laser (XFEL). Protein Structure Prediction; DNA and RNA secondary structures (duplex, triplex, quadruplexes and aptamers), RNA secondary structure prediction. Aptamer databases, structure prediction and modelling, modifications in aptamers.

**References**


1. J. C. Barrett, C. A. Curtis, and C. A. Orengo, Computational Drug Discovery and Design, 2nd ed. New York, NY, USA: Springer, 2017.
2. A. D. Baxevanis and B. F. Ouellette, Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, 3rd ed. Hoboken, NJ, USA: Wiley-Interscience, 2005.
3. A. B. Taylor and K. C. Garcia, Structural Biology Techniques: A Practical Approach, 1st ed. New York, NY, USA: Oxford University Press, 2019.
4. M. Krawczuk and A. Pawlowski, Metagenomics: Methods and Protocols, 2nd ed. New York, NY, USA: Humana Press, 2018.


**CSB3105: Recombinant DNA Technology [3-1-0-4]**

Introduction, relevance and impact of rDNA technology, Applications of Recombinant DNA Technology in health, agriculture, food process, environment and energy sectors. Enzymes used in gene manipulation, host systems and Gene isolation, cloning and expression, DNA sequencing, oligonucleotide synthesis, Southern and Northern hybridization, FISH, PCR, RAPD, RFLP, DNA fingerprinting and their applications for diagnosis of disease, Site-directed mutagenesis, Gene silencing, Gene transfer technologies, Gene therapy, transgenic animals and gene knockout techniques. Genome sequencing projects, advancement of genome sequencing using Next generation Sequencing (NGS). Zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and RNA-guided engineered nucleases (RGENs) derived from the bacterial clustered regularly interspaced short palindromic repeat (CRISPR)-Cas (CRISPR-associated) system. Ethical issues in RDT, Regulatory frameworks for genetically modified organisms (GMOs), Ethical clearances for recombinant products.

**References**

1. T. A. Brown, Gene Cloning and DNA Analysis: An Introduction, 7th ed. Hoboken, NJ, USA: Wiley-Blackwell, 2016.
2. S. Primrose, R. Twyman, and R. Primrose, Principles of Gene Manipulation and Genomics, 7th ed. Hoboken, NJ, USA: Wiley-Blackwell, 2013.
3. B. R. Glick, J. J. Pasternak, and C. L. Patten, Molecular Biotechnology: Principles and Applications of Recombinant DNA, 4th ed. Washington, DC, USA: ASM Press, 2010.
4. D. S. Latchman, Gene Control, 2nd ed. London, U.K.: Garland Science, 2010.

**CSB3130: Computational and Structural Biology Lab [0-0-2-1]**

**List of experiments:**

1. Analysis of Disease Pathways Using KEGG
2. Retrieve protein structure of a disease target
3. Target Identification and Validation Using DrugBank or UniProt Databases
4. Active Site Identification and Pharmacophore Modeling
5. Water Molecule Removal and Structure Preparation for Docking
6. Molecular Docking and Binding Energy Prediction
7. Simulation study of docked complex
8. Demonstration of NGS
9. RNA-seq Differential Expression Analysis
10. Explore and Predict RNA Structures
11. miRNA Target Prediction and Database Mining
12. Taxonomic classification of metagenomic data
13. Biomarker discovery from microbial/other communities
14. Reconstruct microbial genomes from metagenomic data using assembly and binning.
15. Enrichment analysis of differentially expressed genes to understand biological functions and pathways.

**References**

1. C. W. Yap, Computational Chemical Biology: Methods and Protocols, 1st ed. New York, NY, USA: Humana Press, 2017.
2. H. Jin, Bioinformatics: A Practical Handbook of Next Generation Sequencing and Its Applications, 1st ed. Singapore: Springer, 2022.
3. J. Pevsner, Bioinformatics and Functional Genomics, 3rd ed. Hoboken, NJ, USA: Wiley-Blackwell, 2015.
4. M. Krawczuk and A. Pawlowski, Metagenomics: Methods and Protocols, 2nd ed. New York, NY, USA: Humana Press, 2018.

## SEMESTER VI

**CSB3203: Immunodiagnostics and Vaccine Manufacturing [3-1-0-4]**

Introduction to Immunology: Specificity, memory, discrimination of self from non- self, Innate and Acquired immunity, Humoral and cell-mediated immune response. Cells of the immune system, cytokines, complement system. Antibody, Antigens and Immune receptors: Immunoglobins: structure, classes and function. Immunogenicity, antigenicity, epitopes-B cell epitopes, T cell epitopes, haptens, Antigen Recognition by immune system: recognition of antigens by T and B Cells. Immunodiagnostics: Antigen – Antibody Interaction, ELISA, Radial Immunodiffusion, Double Diffusion, Immunoelectrophoresis, Western Blot, Immunofluorescence. Introduction to vaccines: Different types of vaccines, Sub-unit vaccines, Nucleic acid vaccines, recombinant vaccines, synthetic vaccines. Recent trends in vaccine production: Vaccine delivery system and approaches to enhance immunogenicity, Large-scale production of vaccines and automation.

**References**


1. R. A. Goldsby, T. J. Kindt, B. A. Osborne, and J. Kuby, Kuby Immunology, 7th ed. New York, NY, USA: W. H. Freeman and Company, 2013.
2. I. M. Roitt, P. J. Delves, S. J. Martin, and D. R. Burton, Roitt's Essential Immunology, 13th ed. Hoboken, NJ, USA: Wiley-Blackwell, 2017.
3. K. Murphy and C. Weaver, Janeway's Immunobiology, 10th ed. New York, NY, USA: W. W. Norton & Company, 2022.
4. A. K. Abbas, A. H. Lichtman, and S. Pillai, Basic Immunology: Functions and Disorders of the Immune System, 6th ed. Philadelphia, PA, USA: Elsevier, 2023.


**CSB3204: Genomics and Proteomics [3-1-0-4]**

Brief overview of prokaryotic and eukaryotic genome organization and methods of preparing genomic DNA. DNA sequence analysis methods, Identification and classification of organisms using molecular markers- 16S rRNA typing/sequencing; Gene variation and Single Nucleotide Polymorphisms (SNPs); Expressed sequenced tags (ESTs), Genetic and physical maps; methods and techniques used for gene mapping, physical mapping, linkage analysis, cytogenetic techniques, Recombinant DNA technology: DNA cloning basics, Human genome project and the genetic map. Introduction and scope of proteomics; Aims, strategies and challenges in proteomics. Transcriptome analysis for identification and functional annotation of gene, Contig assembly, chromosome walking and characterization of chromosomes, mining functional genes in genome, gene function- forward and reverse genetics, gene ethics; protein- protein and protein-DNA interactions; protein chips and functional proteomics; clinical and biomedical applications of proteomics; introduction to metabolomics, lipidomics, metagenomics and systems biology.

**References**


1. Brown, T. A., Gene Cloning and DNA Analysis, An Introduction. Wiley –Blackwell publication. 2010.
2. Sandy B. Primrose, Richard M. Twyman. Principles of Gene Manipulation and Genomics, Blackwell Scientific Publication. 2009.
3. Old, R. W., Primrose, S. B., & Twyman, R. M., Principles of Gene Manipulation: An Introduction to Genetic Engineering. Oxford: Blackwell Scientific Publications. 2001.
4. Liebler, D. C. (2002). Introduction to Proteomics: Tools for the New Biology. Totowa, NJ: Humana Press.
5. Campbell, A. M., & Heyer, L. J. (2003). Discovering Genomics, Proteomics, and Bioinformatics. San Francisco: Benjamin Cummings.

## Program Electives

**CSBXXXX: Python Programming for Biotechnologists [3-0-0-3]**

Features of python, installing python packages via PIP, Variables, Assignment, Keywords, Input-output, Indentation, Basic data types, Operators and Expressions, Control structures; Python data structures. Strings - creation, accessing, operators, methods; Sets - creation, accessing, operators, methods; List comprehensions; Functions - defining functions, calling functions, passing arguments - keyword arguments, default arguments, variable-length arguments, anonymous functions (lambda); Classes, 'self-variable' methods, Constructor method, Inheritance, Overriding Methods, Data hiding. BIOPYTHON: Tetra nucleotide Frequency: Counting Things, Transcribing DNA into mRNA: Mutating Strings, Reading and Writing Files, Reverse Complement of DNA: String Manipulation, Computing GC Content: Parsing. FASTA and Analyzing Sequences, Finding the Hamming Distance: Counting Point Mutations, translating mRNA into Protein: More Functional Programming, find a Motif in DNA: Exploring Sequence Similarity.

**References**


1. Mitchell L. Model, "Bioinformatics Programming Using Python: Practical Programming for Biological Data (Animal Guide)", 1st edition, O'Reilly Media, 2009.
2. Tim J. Stevens, Wayne Boucher, "Python Programming for Biology: Bioinformatics and Beyond", 1st edition, Cambridge University Press, 2015.
3. David Beazley, "Python essential reference", 2nd edition, New Riders,2001.


**CSBXXXX: R programming for Bioscience [3-0-0-3]**

Basic fundamentals, installation and use of software, data editing, use of R as a calculator, functions and assignments, functions and matrix operations, missing data and logical operators, conditional executions and loops, data management with sequences, Data management (DM), Data frames, import of external data in various file formats, statistical functions, data compilation, Graphics and plots, statistical functions for central tendency, variation, skewness and kurtosis, handling of bivarite data through graphics, correlations, programming and illustration with examples. Multivariate EDA, and Principal Components Analysis; Cluster Analysis; Discriminant Analysis; Decision Tree models (Tree-based models), A Simple PCA using Vegan, Time series analysis, false discovery. Tools for genomic analysis in R, comparative phylogenetic tests, Data Manipulation in R with dplyr, Exploratory Biological Data Analysis with R, RNAseq data analysis in R, Chip-Seq data analysis.

**References**


1. Norman Matloff, "The Art of R Programming", 1st edition, No Starch Press, 2017.
2. Torsten Hothorn and Brian S. Everitt. A Handbook of Statistical Analyses Using R. (3rd edition). Chapman & Hall/CRC Press, Boca Raton, Florida, USA, 2014. ISBN 978-1-4822-0458-2.
3. Johannes Ledolter, "Data Mining and Business Analytics with R", 1st edition, Wiley publishers, 2014.

**CSBXXXX: Data mining and Machine Learning for Bioinformatics [3-0-0-3]**

Data warehousing, OLAP, and preprocessing, followed by data stream mining techniques like sampling, filtering, and moment estimation. It introduces machine learning fundamentals, including supervised, unsupervised, and reinforcement learning, with algorithms such as regression models, decision trees, random forests, SVMs, KNN, Naïve Bayes, and ANNs. It also includes data collection from sources like UCSC, OMIM, PDB, IEDB, and Kaggle API, along with model optimization and evaluation using ChEMBL, ZincBind, and BindingDB. Numpy, Pandas, Matplotlib, Seaborn, Sklearn. Biopython libraries: MSA, BLAST, NCBI's ENTREZ, ENCODE. Tensor Flow and its applications. Introduction to chem-informatics, Database screening for identification of potential drug candidates, Pharrmacophore-modeling, Chemical structure and property relationships, Quantitative structure property relationship, Linear Regression Theory, Lead Optimization and identification. Clustering and Classification. Prediction of gene expression from chromatin information: Deep learning for expression and chromatin prediction: DNA methylation and gene expression. Strong enhancers and Weak enhancers for gene expression. Splicing prediction: Tissue-specific splicing.

**References**


1. Pierre Baldi, Soren Brunak, "Bioinformatics: The machine learning approach", 2nd edition, MIT Press, 2001.
2. Pramod Singh, AvinashManur, "Learn Tensor Flow 2.0: Implement Machine Learning and Deep Learning Models with Python",1st edition, Academic Press. 2019.
3. Müller, Andreas C., and Sarah Guido, "Introduction to machine learning with Python: a guide for data scientists", 1st edition, O'Reilly, 2016.


**CSBXXXX: Neural Networks [3-0-0-3]**

Biological neurons and Neural Networks; Artificial Intelligence (AI) - artificial neurons, networks of artificial neurons, concept of perceptrons, Artificial Neural Networks (ANN). Hebbian learning; Gradient descent learning; Generalized delta rule; Practical considerations - learning in single-layer, multilayer perceptrons; Back-propagation; Learning with momentum; Conjugate gradient learning. Supervised and unsupervised learning - under- fitting and over-fitting, improving generalization. Radial basis function networks - introduction, radial basis function networks, algorithms and applications; Committee machines; Applications of AI in medical. Fundamentals, Self-Organizing Maps (SOM), N-N clustering, Algorithms and applications, Learning vector quantisation.

**References**


1. Daniel Graupe, "Principles of Artificial Neural Networks", 3rd edition, World Scientific Publishing Co. Pte. Ltd., 2013.
2. Raul Rojas, "Neural Networks: A Systematic Introduction", 1st edition, Springer, 2013.
3. David W. Pearson, Nigel C. Steele and Rudolf F. Albrecht, "Artificial Neural Nets and Genetic Algorithms", 2nd edition, Springer, 2012.

**CSBXXXX: Bioperl [3-0-0-3]**

Perl overview, variables and data types, control Structure, loops- while loop, for loop, until loop, File handles - opening and closing files, reading and writing file handles, Library Functions: String specific functions, User defined functions. Arithmetic Operators, Assignment Operators, Logical operators, Equality Operators, Increment and Decrement Operators, String Concatenation and Repetition, Operators precedence and Associativity, Conditional Operators, Logical Operators, Operators for manipulating arrays, Operators for Manipulating hashes. Simple characters, * special character, . character, | character, grouping with () s, anchor characters, pattern matching, regular expression shortcuts, defining subroutines, returning values, using arguments, inheritance in Perl, polymorphism in Perl. Bio-pearl Concatenating DNA Fragments, Transcription: DNA to RNA, Reading Protein Files, Finding Motifs, Simulating DNA, Generating Random DNA, Analyzing DNA, Translating DNA to Proteins, Reading DNA from Files in FASTA format, Separating Sequence and Annotation, Parsing Annotation, Parsing PDB files, Parsing BLAST output, Bio-pearl.

**References**


1. Tisdall, James, “Beginning Perl for Bioinformatics: an introduction to Perl for Biologists “, O’Reilly, 1st edition, 2001.
2. Dwyer, Rex A, “Genomic Perl: From Bioinformatics Basics to Working Code”, Cambridge University Press, 2nd edition, 2003.
3. Tisdall, James, “Mastering Perl for Bioinformatics: Perl Programming for Bioinformatics” O’Reilly, 1st edition, 2003.


**CSBXXXX: Data Management in Healthcare Analytics [3-0-0-3]**

Use of Databases in Health Care, Health Care Settings and the Relevancy of Database Technology, Private solo practice, Group practice, Specialty practice, Hospitals, Clinical research, non-patient databases. Current Health Care Applications of Databases, Reimbursement databases, Disease-specific shared databases, Databases used in HMO's, Surveillance databases Specialty clinical databases General clinical databases, Databases in research. The Future Use of Databases in Health Care, Cost-effectiveness issues, Initiatives and innovation due to technology push, the human element, Sharing of information, Privacy in Databases, The Effect of Databases on Health Care Costs, Quality, and Access

**References**


1. J. A. Magnuson and P. Fu Jr., Eds., Public Health Informatics and Information Systems, 2nd ed. London, UK: Springer, 2014.
2. S. Biedermann and D. Olson, Introduction to Healthcare Informatics, 2nd ed. Chicago, IL: AHIMA Press, 2017.
3. American Health Information Management Association (AHIMA), Health Information Management Technology: An Applied Approach, 6th ed. Chicago, IL: AHIMA Press, 2020.

**CSBXXXX: Forensic Science and Technology [3-0-0-3]**

Introduction to crime laboratories, Responsibilities of the forensic scientist, Securing and Searching the Crime Scene, Recording and collection of crime scene evidence, Document examination, Ethics and Integrity. History of Forensic Genetics, Biological sample collections, The Autopsy and handling of a Dead Body, The Stages and factors of decomposition, Determining the Age and Provenance of Remains, Asphyxia, Gunshot Wounds, Bite Marks. Pattern Analysis: Biological Evidence Overview, Body Fluids - Peripheral blood, Saliva, Semen, Urine, and Sweat, Blood, Markers for Evidence, Study of Hair, Study of Fibre. Detecting the Presence of Blood, Bloodstain Pattern Analysis. Methods of Identification: Forensic anthropology, Paleontology, Drug Identification and Toxicology, Methods used in forensic for human identification: Autosomal STR Profiling, Analysis of Y chromosome, Analysis of Mitochondrial DNA. Sequencing Methods in Forensics: Rules and Principles of Identification under Criminal Justice System, Autosomal single- nucleotide polymorphisms (SNP) typing, Biomarkers in forensic identification, Polymorphic Enzymes, DNA Finger Printing- RFLP. PCR directed Y chromosome sequences, PCR Amelogenein Gene, Next generation Sequencing. Forensic Case Studies: Case studies of Royal Romanov Family, Ice Man, King Tut (Tuttenkhamun), The Hitler Diaries, Criminal investigations revolutionized by DNA. Study of Kinship by DNA Profiling, Paternity disputes, Illegal hunting case identification using Molecular markers; detection of narcotics in body fluids.

**References**


1. P. J. Lincoln and J. Thomson, Forensic DNA profiling protocols. New Jersey: Humana, 1998.
2. S. H. James and J. J. Nordby, Forensic science: An introduction to scientific and investigative techniques. Boca Raton, Flor.: Crc Press, 2005.
3. S. B. Primrose and R. Twyman, Principles of Gene Manipulation and Genomics. John Wiley & Sons, 2013.
4. N. Rudin and K. Inman, An introduction to forensic DNA analysis. Boca Raton, Fla.: Crc Press, 2002.
5. T. A. Brown, Gene Cloning and DNA Analysis : an Introduction. Somerset: Wiley, 2013


**CSBXXXX: Cheminformatics and QSAR [3-0-0-3]**

The domain and scope of Cheminformatics, Learning in Cheminformatics, Structure Elucidation -Quantitative Structure - Activity Relationships - Chemical Reaction and synthesis design. Data, Information and Knowledge Data Preprocessing - Variable selection - Preparation of datasets for validation of the model quality - Databases in the Information System-Catalogs of Chemical Compound - ChemInform RX-Reaction Retrieval. Empirical approaches to the Calculation of Properties - Drug Receptor Binding energies - Quantitative Descriptors of Chirality - BCUT Descriptors - HYBOT Descriptors - 4D QSAR. Prediction of Properties of Compounds - Linear Free Energy Relationship (LFER) - Quantitative Structure-Property Relationship (QSAPR) model - Target Identification and Validation - Lead Finding and Optimization.

**References**


1. Johann Gasteiger and Thomas Engel, ‘’Cheminformatics: A textbook’’, 1st edition, WILEY VCH Publisher, Germany, 2003.

2. Andrew R. Leach and Valerie J. Gillet, ''Introduction to Cheminformatics'', 1st edition, Kluwer Academic Publisher, Netherlands, 2003.
3. Andrew R. Leach, "An Introduction to Chemoinformatics", 1st edition, Springer, 2020.

**CSBXXXX: Artificial Intelligence and Cloud Computing in Healthcare [3-0-0-3]**

Overview of Health Informatics, Introduction to Artificial Intelligence and Cloud Computing, Healthcare data and challenges, Introduction to language models. Cloud architecture and deployment models, Cloud services: IaaS, PaaS, and SaaS, Cloud security and compliance, Data warehousing and data mining, Predictive modelling and decision support, Population health management. Supervised and unsupervised learning, Data pre-processing and feature extraction, Model evaluation and selection, Natural language processing (NLP) for machine learning in healthcare, Text classification and sentiment analysis, Named Entity Recognition (NER), Applications of language models in healthcare. Image processing and feature extraction, Object detection and recognition, Medical imaging analysis. Clinical decision support systems, Electronic Health Records (EHR) and telemedicine, Precision medicine and genomics, Privacy, security, and confidentiality, Bias and fairness, Regulatory and compliance issues.

**References**

1. E. H. Shortliffe and J. J. Cimino, Eds., Biomedical Informatics: Computer Applications in Health Care and Biomedicine, 5th ed., New York, NY: Springer, 2021.
2. R. Hoyt and A. Yoshihashi, Health Informatics: Practical Guide, 8th ed., Pensacola, FL: Informatics Education, 2022.
3. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA: MIT Press, 2016.
4. T. M. Mitchell, Machine Learning, 1st ed., New York, NY: McGraw-Hill, 1997.

**CSBXXXX: Legal and Ethical Aspects of Health Informatics [3-0-0-3]**

Ethical Issues in Biomedical and Health Informatics, Health-Informatics Applications: Appropriate Use, Users, and Contexts, Privacy, Confidentiality, and Data Sharing, Social Challenges and Ethical Obligations, Legal and Regulatory Matters. Ethical Standards, Ethical Decisions and Challenges, Bioethics Issues. Patient Record Requirements, Confidentiality and Informed Consent, Access to Health Information, Specialized Patient Records. Risk Management, Quality Management, and Utilization Management, Information Systems, Health Care Fraud and Abuse, Law and Ethics in the Workplace. Professional ethics for public health practitioners: Developing empathy, resolving conflicts and building consensus, conflict of interest, issues of integrity, transparency and accountability, communication skills, etc.

**References**

1. R. A. Spinello, Cyberethics: Morality and Law in Cyberspace, 6th ed., Burlington, MA: Jones & Bartlett Learning, 2020.
2. F. G. Miller, J. C. Moreno, and A. Caplan, Eds., The Ethics of Biomedical Big Data, Cham, Switzerland: Springer, 2016.

3. L. E. Terman, Biomedical Ethics in a Secular Society, 1st ed., West Conshohocken, PA: Templeton Press, 2020.
4. A. M. Rothstein, Health Privacy: The Medical and Genetic Information Privacy Handbook, Ann Arbor, MI: University of Michigan Press, 2020.

**Faculty of Science, Technology and Architecture | School of Core Engineering**

**Department of Biotechnology and Chemical Engineering**

**Honors Course List** – 2023 onwards

| Semester | Course code | Name | L-T-P-C |
|---|---|---|---|
| V | CSB3180 | Research Methodology | 1-0-0-1 |
| VI | CSB3280 | Structural Biology | 3-0-0-3 |
| VII | CSB4180 | Genome Editing and Engineering | 3-0-0-3 |
| VII | CSB4181 | Introduction to Proteogenomics | 3-0-0-3 |
| VIII | CSB4280 | Honors Project | 0-0-16-8 |

| | | | |
|---|---|---|---|
| VI | CSB3280 | Structural Biology | 3-0-0-3 |

**Course Name: Structural Biology**

**Link for the course**: https://onlinecourses.nptel.ac.in/noc21_bt14/preview

**Course Instructors:**

Dr. Saugata Hazra
IIT Roorkee

**Duration: 12 weeks (Jan - April)**

**BRIEF COURSE OUTLINE:**

This course introduces the post-NGS era and the central role of structural biology in translating sequence to structure to function, using proteins as the focal macromolecules. Students explore protein architecture: amino acids, primary to quaternary structures, motifs, domains, conformational analysis, and folding alongside the chemical forces that stabilize macromolecules (covalent, ionic/coordinate, hydrophobic, and van der Waals interactions). Enzyme structure–function principles are developed through active-site ("magic pocket") concepts and ligand interactions. Core experimental techniques are surveyed comparatively—macromolecular crystallography, NMR, and cryo-EM—followed by deeper practice in crystallization, data collection, diffraction theory (Bragg's law, scattering/structure factors), phase determination, and full structure solution. Learners then gain fluency with 3D structural data through the PDB/RCSB, perform calculations, and visualize models in PyMOL and Coot. The course culminates in an introduction to molecular dynamics simulations, atomistic modeling, force fields, bonded/non-bonded terms, topology/parameters, periodic boundaries, solvation, and workflow with standard packages—and applies all skills to case studies in structure-based drug design and protein engineering.

| | | | |
|---|---|---|---|
| VII | CSB4180 | Genome Editing and Engineering | 3-0-0-3 |

**Course Name:** Genome Editing and Engineering

**Link for the course**: https://onlinecourses.nptel.ac.in/noc22_bt35/preview

**Course Instructors:**

Prof. Utpal Bora
IIT Guwahati

**Duration: 12 weeks (July - October)**

**BRIEF COURSE OUTLINE:**

This course introduces foundational genetics and genetic engineering, building from these concepts to cutting-edge genome editing and its real-world impact. We begin with core principles and the cellular pathways of DNA breakage and repair, then explore recombination and strategies for targeted genetic modification. Students gain a comparative understanding of the major editing platforms—ZFN, TALEN, and CRISPR/Cas9—including design logic, mechanisms, and delivery considerations. Applications span therapeutic genome editing for human diseases, genome-engineered disease models, and engineered immune cells for cancer therapy, culminating in the promise of personalized interventions. Throughout, we critically evaluate technical challenges—safety, specificity, off-target effects—and conclude with a rigorous discussion of ethical, legal, and social implications, with special attention to germline gene editing.

| | | | |
|---|---|---|---|
| VII | CSB4181 | Introduction to Proteogenomics | 3-0-0-3 |

**Course Name:** Introduction to Proteogenomics

**Link for the course**: Introduction to Proteogenomics - Course

**Course Instructors:**

Prof. Sanjeeva Srivastava
IIT, Bombay

**Duration: 12 weeks (July - October)**

**BRIEF COURSE OUTLINE:**

Cancer research has been significantly aided by advancements in proteogenomics technologies, where proteomics information derived from mass spectrometry is used to complement genomics using next generation sequencing. With the recent advent of Cancer Moonshot Project, the critical role that proteogenomics can play in improving cancer patient treatment is increasingly being recognized. This course will utilize advanced genomic and proteomic technologies and their data from high-quality human biospecimens to identify potentially actionable therapeutic molecular targets. This course is a part of a workshop by experts in the fields of proteomics and proteogenomics in cancer research from the Broad Institute of MIT and Harvard and Indian Institute of Technology Bombay. The course will comprise interactive lectures with case studies, hands-on sessions and demonstrations on proteogenomics aimed at accelerated understanding of cancer and will cover the principles of proteogenomics followed by experimental sessions, where proteomics data using LC-MS/MS will be processed and analyzed. The next step will be to integrate the proteomics data with genomics data, from The Cancer Genome Atlas for the proteogenomics analysis. Lectures and demonstrations on different computational methods will be performed for statistical data analysis of proteogenomics data.

**“Big Data, Genes and Medicine”** as a value- added course is offered by Coursera. The details of the course are:

**Link for the course:** https://www.coursera.org/learn/data-genes-medicine

**Instructor:** Isabelle Bichindaritz, Associate Professor, The State University of New York

Dr. Isabelle Bichindaritz is an associate professor in Computer Science and Director of Biomedical Informatics at SUNY Oswego. Following receiving a PhD in Computer Science from the University René Descartes – Paris V, Dr. Bichindaritz served as a research scientist at the Fred Hutchinson Cancer Research Center in Seattle and as assistant professor and professor at the University of Washington, Tacoma Institute of Technology and the National School of Public Health. Her research focuses on intelligent learning systems and biomedical data science for decision analytics and data analytics in healthcare and biomedical research, as well as more broadly artificial intelligence in medicine. She has authored or co-authored more than 100 papers in peer reviewed journals and conferences.

**Description:** This course distills students for expert knowledge and skills mastered by professionals in Health Big Data Science and Bioinformatics. Students will learn exciting facts about the human body biology and chemistry, genetics, and medicine that will be intertwined with the science of Big Data and skills to harness the avalanche of data openly available at fingertips and which are just starting to make sense of. They will investigate the different steps required to master Big Data analytics on real datasets, including Next Generation Sequencing data, in a healthcare and biological context, from preparing data for analysis to completing the analysis, interpreting the results, visualizing them, and sharing the results.

Needless to say, when they master these high-demand skills, they will be well positioned to apply for or move to positions in biomedical data analytics and bioinformatics. No matter what their skill levels are in biomedical or technical areas, they will gain highly valuable new or sharpened skills that will make them stand-out as a professional and want to dive even deeper in biomedical Big Data. It is hoped that this course will spark student interest in the vast possibilities offered by publicly available Big Data to better understand, prevent, and treat diseases.

There are 6 modules in this course which can be completed in 4 weeks with 10 hours per week.