



Central University of Himachal Pradesh

(Established under Central Universities Act 2009)

PO BOX: 21, DHARAMSHALA, DISTRICT KANGRA – 176215, HIMACHAL
PRADESH

www.cuhimachal.ac.in

SEMESTER- II

Course Code: CBB 501

Course Name: Algorithms in Computational Biology

Course Instructor: Dr. Vikram Singh

Credits Equivalent: 4 Credits

(One credit is equivalent to 10 hours of lectures / organised classroom activity / contact hours; 5 hours of laboratory work / practical / field work / Tutorial / teacher-led activity and 15 hours of other workload such as independent individual/ group work; obligatory/ optional work placement; literature survey/ library work; data collection/ field work; writing of papers/ projects/dissertation/thesis; seminars, etc.)

Course Objectives: The course is designed to introduce students the algorithmic principles, central to the studies in Computational Biology and Bioinformatics. Concepts from computer science like dynamic programming and graph theory will enable students to understand a variety of concepts that are used in the theoretical studies of life sciences and expose them to the underlying mechanisms of widely used softwares. Students, who are familiar with at least one programming language, will be encouraged to write their own codes for various algorithms discussed in this course.

Pre-requisite: CBB 405 Basics of Bioinformatics

Attendance Requirements:

Students are expected to attend all lectures in order to be able to fully benefit from the course. A minimum of 75% attendance is a must failing which a student will not be permitted to appear in examination.

Evaluation Criteria:

1. Mid Term Examination: 25%
2. End Term Examination: 50%
3. Continuous Internal Assessment: 25%
 - Class Participation – 5%
 - Assignments – 10%
 - Course project – 10%

Course Contents:

Unit 1: Introduction to algorithms and complexity 8 Hours

1. Basic Concepts: Algorithmic complexity (spatial and temporal), Biological vs. Computer algorithms, Genetic algorithm.
2. Standard Notations: Big-Oh, Omega, Theta notations; Hardness of an algorithm.
3. Linear and non-linear data structures, Stack, Queues, Linked list.
4. Introduction to algorithm design techniques: Exhaustive search, Greedy Algorithms, Divide and conquer etc.
5. Searching algorithms: Linear and Binary search; Sorting algorithms: Selection, Bubble, Insertion, Merge, Quick, Heap.

Unit 2: Sequence Analysis 10 Hours

1. Models of DNA evolution: Jukes Cantor, Kimura and Tamura models
2. Derivation of protein evolution models: PAM and BLOSUM
3. Elements of dynamic programming: Edit distance, Longest Common Subsequences, Global and Local Sequence Alignment
4. Working of BLAST: Usage of Finite State Machine, E-value, etc.
5. Introduction to Multiple Sequence Alignment
6. Concepts of Joint, Conditional, Marginal probabilities
7. Markov Chains and Hidden Markov Models: CpG islands, Pairwise alignment using HMMs

Unit 3: Pattern matching 6 Hours

1. Tandem and Interspersed repeats.
2. Repeat finding: Motifs, consensus, position weight matrices
3. Algorithms for derivation of and searching sequence patterns: MEME, PHI-BLAST, SCanProsite and PRATT
4. Algorithms for generation of sequence profiles: PSIBLAST, HMMer,

Unit 4: Phylogenetics 8 Hours

1. Basics of Molecular Evolution,
2. Tree - terminologies, Binary trees, AVL trees.
3. Tree traversal: Pre-order, In-order, post-order; Breadth-first search, Depth-first search.
4. Evolutionary trees: (I) Distance based methods – UPGMA, NJ, Fitch Margoliash (FM), Minimum Evolution (ME); (II) Character based methods – MP, ML, Bayesian inference algorithm.
5. Tree Evaluation, Bootstrapping.

Unit 5: RNA and Protein Structure Prediction 8 Hours

1. RNA secondary structure: Nussinov algorithm, Energy minimization, Zuker's Algorithm, SCFG.
2. RNA tertiary structure: Basics of “minimum free energy” based methods.
3. Protein secondary structure: Chou-Fasman, GOR method.
4. Protein tertiary structure: Rosetta method, Contact potential method.

Text Books:

1. **Jones and Pevzner (2004)**, An Introduction to Bioinformatics Algorithms. MIT Press.
2. **Durbin *et al.* (1998)**, Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press
3. **Mount (2004)**, Bioinformatics: Sequence and Genome Analysis, CBS Publishers
4. **Pevsner (2015)**, Bioinformatics and Functional Genomics. Wiley.

Additional Readings:

1. **Gusfield (2005)**, Algorithms on Strings, Trees and Sequences. Cambridge University Press.
2. **Cormen *et al.* (2009)**, Introduction to Algorithms. MIT Press.
3. **Sung (2009)**, Algorithms in Bioinformatics: A Practical Introduction. Chapman & Hall/CRC.
4. **Neapolitan and Naimipour (2011)**, Foundations of Algorithms. Jones & bartlett.
5. **Korf *et al.* (2003)**, BLAST. O'Reilly
6. **Junker and Schreiber (2008)**, Analysis of Biological Networks. Wiley-Interscience, New Jersey.
7. **Mitchell (1998)**, An Introduction to Genetic Algorithms. MIT Press.



Central University of Himachal Pradesh

(Established under Central Universities Act 2009)

PO BOX: 21, DHARAMSHALA, DISTRICT KANGRA – 176215, HIMACHAL
PRADESH

www.cuhimachal.ac.in

SEMESTER- II

Course Code: CBB 504

Course Name: Genomics and Proteomics

Credits Equivalent: 4 Credits

(One credit is equivalent to 10 hours of lectures / organised classroom activity / contact hours; 5 hours of laboratory work / practical / field work / Tutorial / teacher-led activity and 15 hours of other workload such as independent individual/ group work; obligatory/ optional work placement; literature survey/ library work; data collection/ field work; writing of papers/ projects/dissertation/thesis; seminars, etc.)

Course Objectives: The course is designed to:

- Introduce students to the steps and key elements of the Sequence Analysis in biology.
- Utilization and application of genomics and association of genotype with phenotype.
- Hands-on- training in basic tools of genomics and their applications
- Study to understand proteome function, its quantification and understanding individual proteins and interacting partners including account of protein modifications.

Attendance Requirement:

Students are expected to attend all lectures in order to be able to fully benefit from the course. A minimum of 75% attendance is a must failing which a student may not be permitted to appear in examination.

Evaluation Criteria:

1. Mid Term Examination: 25%
2. End Term Examination: 50%
3. Continuous Internal Assessment : 25%
 - a. Presentation: 10%
 - b. Class Participation: 10%
 - c. Attendance: 5%

Course Contents:

UNIT-I: Introduction to Genomics

(8 hours)

1. Genes and Proteins
2. Genome sequencing and databases
3. Gene variation and Single Nucleotide Polymorphisms
4. Expressed sequenced tags

UNIT-II: Advances in Genomics**(8 hours)**

1. Genotype-phenotype associations
2. Genotyping tools
3. Comparative genomics
4. Functional genomics

UNIT-III: Genome analysis**(8 hours)**

1. Genetic and physical maps
2. Methods of molecular mapping
3. Bioinformatics analysis of genomes
4. Applications of genomics

UNIT-IV: Introduction to Proteomics**(8 hours)**

1. Introduction to proteins: structure and function
2. Methods of protein isolation, purification, quantification
3. Proteomics databases
4. Proteome & Interactome: Bacterial and Yeast two hybrid

UNIT-V: Proteome analysis**(8 hours)**

1. Mass-spec based analysis
2. High throughput proteome analysis: Qualitative & Quantitative
3. Applications of proteome analysis
4. Structural Proteomics

Text Books

1. Daniel Liebler, Introduction to Proteomics: Tools for the new Biology, 2012
2. Sandy B. Primrose, Richard Twyman. Principle of gene manipulation and genomics, 7th edition. 2006.
3. Introduction to Protein Structure by Carl-Ivar Branden, John Tooze

Additional Readings

4. Ehebauer MT, Wilmanns M. The progress made in determining the *Mycobacterium tuberculosis* structural proteome. Proteomics. 2011 Aug; 11(15): 3128-33.
5. Chim N, Habel JE, Johnston JM, Krieger I, Miallau L, Sankaranarayanan R, Morse RP, Bruning J, Swanson S, Kim H, Kim CY, Li H, Bulloch EM, Payne RJ, Manos-Turvey A, Hung LW, Baker EN, Lott JS, James MN, Terwilliger TC, Eisenberg DS, Sacchettini JC, Goulding CW. The TB Structural Genomics Consortium: a decade of progress. Tuberculosis (Edinb). 2011 Mar; 91(2):155-72.
6. Biochemistry and Molecular Biology, 7th edition, Keith Wilson and John Walker
7. Francki M, Appels R. Wheat functional genomics and engineering crop improvement.

Genome Biol. 2002; 3(5).

8. Kaur S, Francki MG, Forster JW. Identification, characterization and interpretation of single-nucleotide sequence variation in allopolyploid crop species. *Plant Biotechnol J.* 2012 Feb;10(2): 125-38.
9. Nagaraj SH, Gasser RB, Ranganathan S. A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform.* 2007 Jan;8(1): 6-21.



Central University of Himachal Pradesh

(Established under Central Universities Act 2009)

PO BOX: 21, DHARAMSHALA, DISTRICT KANGRA – 176215, HIMACHAL
PRADESH

www.cuhimachal.ac.in

SEMESTER- II

Course Code: CBB-429

Course Name: Practical Course on Structure Prediction and Modelling

Credits Equivalent: 2 credits

(One credit is equivalent to 10 hours of lectures / organized classroom activity / contact hours; 5 hours of laboratory work / practical / field work / Tutorial / teacher-led activity and 15 hours of other workload such as independent individual/ group work; obligatory/ optional work placement; literature survey/ library work; data collection/ field work; writing of papers/ projects/dissertation/thesis; seminars, etc.)

Course Objectives:

This practical course will teach aspects of computational prediction of protein structures. Molecular modelling and simulation methods will be introduced

Attendance Requirement:

Students are expected to attend all lectures in order to be able to fully benefit from the course. A minimum of 75% attendance is a must failing which a student may not be permitted to appear in examination.

Evaluation Criteria:

4. Mid Term Examination: 25%
5. End Term Examination: 50%
6. Continuous Internal Assessment : 25%
 - d. Practical record: 10%
 - e. Class room participation: 15%

Course Contents:

Lab I – Structure prediction by homology modelling

Lab II – Structure refinement by energy minimisation methods

Lab III – Structure refinement by molecular dynamics simulations

Lab IV – Analysis of MD trajectories

Lab V – Chemical structure softwares

Lab VI – Small molecule structure refinement, and conformer generation

Lab VII – Information from chemical structures



Central University of Himachal Pradesh

(Established under Central Universities Act 2009)

PO BOX: 21, DHARAMSHALA, DISTRICT KANGRA – 176215, HIMACHAL
PRADESH

www.cuhimachal.ac.in

SEMESTER- II

Course Code: CBB-410

Course Name: Computational methods in Structure Analysis

Credits Equivalent: 2 credits

(One credit is equivalent to 10 hours of lectures / organized classroom activity / contact hours; 5 hours of laboratory work / practical / field work / Tutorial / teacher-led activity and 15 hours of other workload such as independent individual/ group work; obligatory/ optional work placement; literature survey/ library work; data collection/ field work; writing of papers/ projects/dissertation/thesis; seminars, etc.)

Course Objectives:

This introductory course will provide a broad overview of Computational methods for macromolecule structure analysis. This course will introduce various bioinformatics methods and tools for detailed understanding of macromolecular structures. It will focus on ways to obtain useful information from available structural data and apply for various research scenarios.

Attendance Requirement:

Students are expected to attend all lectures in order to be able to fully benefit from the course. A minimum of 75% attendance is a must failing which a student may not be permitted to appear in examination.

Evaluation Criteria:

7. Mid Term Examination: 25%
8. End Term Examination: 50%
9. Continuous Internal Assessment : 25%
 - f. Assignment: 5%
 - g. Class room participation: 10%
 - h. Class test: 10%

Course Contents:

Unit I –Overview of Structural Bioinformatics (4 Hours)

- Introduction
- Importance of analysis of bio-macromolecule data
- Primary and derived 3D structure databases

Unit II – Bio-Macromolecule Structure

(4 Hours)

- Protein structure
- Secondary structure assignment and prediction
- Secondary and supersecondary structures
- Fundamentals of DNA and RNA structures

Unit III – Tools for analysis of Protein Structure

(4 Hours)

- Protein folds and structural family resources
- Structure based signatures
- Prediction of function from structure

Unit IV–Structure Prediction methods

(4Hours)

- Homology Modelling
- Threading

Unit V –Applications

(4Hours)

- Methods to predict binding sites in proteins
- Analysis of protein-protein, protein-DNA and protein-RNA complexes