

<b>Semester-V/ Extra Credit V Value Added Course I (Multidisciplinary)</b>	<b>BIO INFORMATICS</b>	<b>Course Code: VAX:1A</b>
<b>Instruction Hours:</b>	<b>Credits:2</b>	<b>Exam Hours:</b>
<b>Internal Marks:-</b>	<b>External Marks:100</b>	<b>Total Marks: Grade</b>

### Course Objectives

- To understand the procedural and object oriented paradigm with concepts of streams, classes, functions, data and objects.
- Understand how to produce object-oriented software using C++
- Understand fundamentals of object oriented programming in Java, including defining classes, invoking methods, using class libraries, etc.
- Identify classes, objects, members of a class and the relationships among them needed for a finding the solution to specific problem
- It enables to understand object oriented programming concepts

### Theory

**TOTAL MARKS: 40**

#### Unit-1

History of Bioinformatics – Goal of bioinformatics as a separate discipline – Emerging branches of Bioinformatics-NCBI Pub Med-Genbank, Fasta and Swiss-PORT formats – Sequence Databases : Nucleotide Sequence – Protein Sequence Databases – SWISS-PORT, Protein Structure Visualization Tools: RasMol, Swiss PDB Viewer-Database searching tools– BLAST and FASTA algorithms – Various versions of basic BLAST and FASTA

#### Unit-2

Various definitions of bioinformatics, history of bioinformatics, applications of bioinformatics- Types of Databases Biological databases: Primary databases – GenBank, Protein sequence databases – Swissport, Blossum- Structural databases – PDB, Bibliographic databases:- Pubmed, Translate tool using DNA,RNA Prediction Tool.

#### Unit-3

Sequence databases – Nucleic acid sequence databases: GenBank, EMBL, DDBJ; Protein sequence databases: Uniport-KB: SWISS-Port. Structure Databases: PDB, NDB, PubChem, ChemBank. Scoring matrices: basic concept of a scoring matrix, Matrices for nucleic acid and proteins sequences, PAM and BLOSUM series, matrix derivation methods and principles. Pair wise sequence alignment: Basic concepts of sequence alignment

## Practical

**Total Marks: 60**

1. Explore the sitemap of NCBI and PUBMED and find the official gene symbol, alias name, chromosome number and ID for a particular sequence.
2. Retrieve the Genbank entry with an accession number AF375082 and save the sequence in FASTA format.
3. Retrieve Protein sequences from Protein Data Bank (PDB) and analyze the primary, secondary and tertiary protein structure using tools.
4. Retrieve nucleotide sequences and perform pair wise and multiple sequence alignment using BLAST tool and analyze the output.
5. Convert gene sequence into its corresponding amino acid sequence using Translate Tool.
6. Convert RNA gene sequence into its corresponding amino acid sequence using RNA prediction Tool.
7. Retrieve the structures of the compounds from PubChem: Xylitol, Saccharine, Aspartame
8. Perform the PHI-BLAST and PSI-BLAST for the protein sequence Q1A232. Write the top 4 E-scores values and the Sequence ID until convergence.
9. Perform Pair wise alignments for the proteins Insulin from the organism's homosapiens and Musmusculus. Calculate the Percent Similarity and Identity using BLOSUM 62 and PAM 250
10. Find the super secondary structure for any protein database.

## Course Outcomes

- After the completion of this course, the students will be able to develop applications.
- Introduces Object Oriented Programming concepts using the C++ language
- Discuss the various decisions making and construct statement of programming. Illustrate the essentials of the library.
- Apply object oriented programming concept in real time problems.
- Illustrate pattern matching and extraction using regular expression.