UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Biologia Computazionale

2526-1-F0602Q124

Aims

This course aims to provide students with a solid theoretical and practical foundation in computational biology. It will focus on predicting protein structure and function from amino acid sequences using classical approaches and artificial intelligence techniques. Students will learn to analyze major databases of biological sequences and structures, and they will be taught to integrate and critically interpret results generated by various computational methods.

The general objectives include:
Knowledge and understanding
Students will have acquired knowledge of the primary computational methods used to analyze biological sequences. This learning will encompass a theoretical understanding of the underlying algorithms and the ability to apply them in practice. Students will also learn to interpret computational data within a biological context and develop the ability to formulate coherent, testable hypotheses.

Application of Knowledge and Understanding
Students will be able to design and implement an independent bioinformatics analysis pipeline for the structural and functional prediction of a protein with an unknown function.

Autonomy of Judgment
Students will be encouraged to critically evaluate the reliability of computational predictions and recognize the main limitations of the methods used. Students will also be encouraged to integrate various sources of information, such as structural and functional data, to achieve a coherent, comprehensive, scientifically sound biological interpretation.

Communicative Skills
Students will master the appropriate terminology to accurately describe a computational analysis, including its objectives, methods, results, and limitations.

Learning Skills

Students will be able to independently apply acquired skills and flexibly use methods learned during the course to tackle new biological problems not covered during course.

Contents

1. Biological Databases: Sequences, Structures, and Query Strategies
2. Sequence Alignment
3. Molecular Phylogenetics
4. Functional Prediction: From Amino Acid Sequence to Function
5. Structural Prediction: From Amino Acid Sequence to Structure
6. Advanced Computational Methods for Protein-Ligand Interactions
7. Application of Computational Methods to Predict the Structure and Function of an Unknown Protein.

Detailed program

1. Biological Databases: Sequences, Structures, and Query Strategies
Query strategies and data retrieval from major bioinformatics resources.
Exploration of sequence databases (GenBank, UniProt) and metagenomics databases (MGnify).
Exploration of structural databases (PDB, AlphaFold DB) and understanding of standard formats (PDB, mmCIF).
Exploration of databases based on families, domains, and structural classification (Pfam, PROSITE, CATH).

2. Sequence Alignment
Principles and application of scoring matrices for alignment (PAM, BLOSUM, Position Specific Scoring Matrix).
Pairwise (local) and multiple alignment techniques for identifying conserved regions.
Large-scale similarity searches using BLAST: critical interpretation of results (E-value, bit-score).

3. Molecular Phylogenetics
Methods for constructing phylogenetic trees (Neighbor-Joining, Maximum Likelihood).
Tree interpretation: reading topology, assessing statistical support (bootstrap), and identifying evolutionary events.
Principles of ancestral sequence reconstruction to infer properties of ancient proteins.

4. Functional Prediction: From Sequence to Function
Functional annotation methods based on homology and phylogenetic analysis.
Strategies and limitations of automatic functional annotation.
Application of Hidden Markov Models (HMMs) for recognizing functional motifs and assigning proteins to families.
Prediction of post-translational modifications.

5. Structural Prediction: From Sequence to 3D Structure
Prediction of secondary structures and intrinsically disordered regions (IDRs).
Homology modeling: from template search to model construction.
The deep learning revolution: ab initio structural prediction with AlphaFold.
Critical evaluation of 3D model quality: analysis of pLDDT, QMEAN scores, and use of the Ramachandran plot.
Principles of structural alignment for protein comparison.

6. Computational Methods for Studying Protein-Ligand Dynamics and Interactions
Introduction to molecular docking.
Introduction to molecular dynamics.

7. Laboratory: Prediction of the Structure and Function of an Unknown Protein

Perform in silico translation and identify the correct Open Reading Frame (ORF).
Conduct a complete functional analysis through homolog search (BLAST), multiple sequence alignment, and identification of conserved domains and motifs (INTERPRO).
Perform 3D structural prediction using Homology Modeling and AlphaFold approaches.
Assess the reliability of the generated structural model and visualize it using dedicated software.

Prerequisites

Biochemistry, Protein Biochemistry and Molecular Biology

Teaching form

The course consists of two integrated modules totaling 45 hours (6 CFU), combining theoretical foundations with practical activities.

Theoretical Module (18 classroom lectures, 35 hours - 5 CFU): Integrates frontal sessions with slides and videos, and interactive sessions where students analyze computational data, applying concepts learned.

Practical Module (5 computer laboratory sessions, mandatory attendance, 10 hours - 1 CFU): Includes introductory phases with slides and software tutorials, followed by practical application of skills to extract and interpret data.

Textbook and teaching resource

Slides and video lectures are available on the course Moodle page (http://elearning.unimib.it/).

Recommended texts:
M.H. Citterich et al. “Fondamenti di Bioinformatica” Zanichelli
Pascarella e Paiardini “Bioinformatica, dalla sequenza alla struttura delle proteine” Zanichelli

Semester

Second semester

Assessment method

Oral examination. The examination will focus:
- Presentation using PowerPoint slides of laboratory module results (50% of evaluation). Evaluation considers understanding of computational tools used, critical interpretation of results, and appropriate language use.

- Interview on theoretical module contents (50% of evaluation). The evaluation is based on knowledge of topics covered, ability to connect theoretical content with computational tools used in the laboratory, language use, and clarity of presentation.

Office hours

Upon request by e-mail to lecturer

Sustainable Development Goals

GOOD HEALTH AND WELL-BEING | QUALITY EDUCATION | RESPONSIBLE CONSUMPTION AND PRODUCTION