

**Department of Computer Science and Software Engineering  
Concordia University  
COMP 6811 — Bioinformatics Algorithms  
Fall 2022**

**Instructor:** Dr Gregory Butler, Office ER-1049, email [gregb@cse.concordia.ca](mailto:gregb@cse.concordia.ca)

## Curriculum Description

**COMP 6811 Bioinformatics Algorithms** (4 credits)

The principal objectives of the course are to cover the major algorithms used in bioinformatics; sequence alignment, multiple sequence alignment, phylogeny; classifying patterns in sequences; secondary structure prediction; 3D structure prediction; analysis of gene expression data. This includes dynamic programming, machine learning, simulated annealing, and clustering algorithms. Algorithmic principles will be emphasized. A project is required.

## Outline of Topics

The course will focus on algorithms for *protein sequence analysis*. It will not cover genome assembly, genome mapping, or gene recognition.

Lectures will present a general introduction into biology, genomics, biotechnology, and bioinformatics. The introduction will include a discussion of the major resources for protein sequences, protein families, and their curation using ontologies.

The lectures on algorithms will assume a general background in algorithms and data structures, especially specification of algorithm requirements by pre-conditions and postconditions, simple complexity analysis for resource usage (time, memory, disk); and familiarity with hashing, string algorithms, dynamic programming, and greedy algorithm.

Bioinformatics algorithms will focus on algorithms for protein sequence analysis with the aim of understanding the algorithms required for the eggNOG system construction and search with eggNOG-mapper. This includes sequence alignment; profile Hidden Markov models; k-mer techniques for sequence analysis; clustering and phylogenetic tree construction; multiple sequence analysis; and phylogenomics.

There is no set textbook. Pointers to supporting material on the web will be provided. You will be required to read seminal papers on each of the algorithms.

Topics:

- ▶ Background in Biology and Genomics
- ▶ Sequence Alignment: Pairwise and Multiple
- ▶ Representation of Protein Amino Acid Composition
- ▶ Profile Hidden Markov Models
- ▶ Profile-Based Alignments
- ▶ Algorithms Based on  $k$ -mers
- ▶ Protein Families, Phylogenomics, and Orthologous Groups

## Recommended Texts — in Library

D. Higgins and W. Taylor (editors). **Bioinformatics: Sequence, Structure and Databanks**, Oxford University Press, 2000.

A. D. Baxevanis and B. F. F. Ouellette. **Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins**, Wiley, 1998.

Richard Durbin, Sean R. Eddy, Anders Krogh, and Graeme Mitchison. **Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids**. Cambridge University Press, 1998.

## Other Reading Material

The course website <http://users.encs.concordia.ca/~gregb/home/comp6811-f2022.html> will contain links to seminal publications for each topic, and to relevant software and databases.

## Learning Outcomes

Specific learning outcomes covering

- ▶ Knowledge,
- ▶ Critical Thought, and
- ▶ Skills

are given on the course website.

# Evaluation

## Assignments

There will be four written assignments. Assignments are individual work.

## Projects

There will be two projects: a small project due in the first half of the course; and a larger project with presentation and report due at the end of the course. Both projects are individual work.

The large project will require a 15-minute presentation showing your results and a technical report of up to 10 pages in IEEE two-column format in Latex.

## Examinations

There will be a midterm examination.

There will **not** be a final examination.

## Marks

Assignment 1	10%
Assignment 2	10%
Assignment 3	10%
Assignment 4	10%
<hr/>	
Small Project	10%
Large Project	20%
<hr/>	
Midterm Exam	30%
<hr/>	
Total	100%

## Note

There is no standard correspondence between the numerical marks and the final letter grades.

Students must pass each component of the course — assignments, projects, and examination — in order to pass the course.

## **What if ... Pandemic**

We reserve the right to modify the evaluation schema in the light of circumstances outside the control of the University.

## **Academic Honesty**

Violation of the Academic Code of Conduct in any form will be severely dealt with. This includes copying (even with modifications) of program segments. You must demonstrate independent thought through your submitted work. The Academic Code of Conduct is available at: <http://www.concordia.ca/students/academic-integrity/code.html>.