

# BIOINFORMATICS PROGRAM

Academic Adviser: Kalle M. Levon

Bioinformatics is a relatively new branch of science that organizes and interprets the overwhelming amount of data recently available on living organisms as a result of the human genome project and follow-up projects. The need for this branch of science is very real; while it is becoming increasingly simple to obtain experimental information on DNA, RNA and protein sequences and expression levels, the interpretation of this information is far from straightforward. There is a critical scarcity of individuals with the training in biology, chemistry, computer science and the specific methods of bioinformatics to interpret this data. The need is very real in academia and especially in industry. The need for individuals with a focused expertise in bioinformatics emerged a few years ago and is growing rapidly.

Polytechnic's MS in Bioinformatics is designed to provide a solid, carefully tailored program of study. It is responsive to the needs of (1) persons with BS degrees seeking entry into the field and (2) persons with advanced degrees who may be employed in sectors that now value and require such expertise. The latter typically have primary (PhD level) expertise in chemistry, molecular biology or chemical modeling, and may be employed in the biotechnology or pharmaceutical sectors. A master's program, with flexible on-site requirements and extensive use of Web-based or other remote-access training concepts, such as multimedia instruction, addresses the needs of both groups.

The curriculum is flexible, allowing sufficient exposure to and mastery of computer science skills to enable students to be effective in their home environments or competitive in the job market. In addition to traditional computer science, students gain expertise with such commercial databases as Oracle and Psibase, in addition to theory and hands-on experience with widely used methods in bioinformatics, including dynamic programming, profiles, hidden Markov chains and neural nets to predict protein secondary structure.

## GOALS AND OBJECTIVES

The Master of Science program in Bioinformatics is a crucial component in offering a mechanism to train individuals seeking reassignment to fast growing bioinformatics positions within the pharmaceutical or biotechnological industries. The goal of the MS program is to fill the present educational needs with flexible on-site requirements and extensive use of web-based or other remote-access training concepts such as multimedia instruction. In addition to the traditional computer science components, the program entails training in commercial databases and exposure to theory and to hands-on experience with widely used methods in bioinformatics such as dynamic programming, profiles, hidden Markov chains and neural nets to predict protein secondary structure. Thus, the basic understanding of sequence analysis, protein structure prediction and the structure/function relationship are the core educational goals.

## REQUIREMENTS FOR THE MASTER OF SCIENCE

The Master of Science degree is intended for students from various backgrounds seeking depth knowledge in informatics in chemical and biological sciences.

Admission to the master's program requires a bachelors' degree from an accredited institution, with superior undergraduate academic record with completion of all prerequisite courses. Students not meeting all these requirements will be considered for admission on an individual basis and may be admitted subject to completing appropriate undergraduate courses to remove deficiencies.

Applicants who are otherwise sufficiently prepared for admission without undergraduate deficiencies may nevertheless be required to take specified undergraduate and introductory level graduate courses. Such courses count towards a master's degree.

To satisfy the requirement for the degree Master of Science, students must complete a total of 36 units and hold an overall B grade in all graduate courses. Students must take all four of the required

core courses and at least two of the basic core courses. The knowledge in all basic core courses is required, thus all these courses are required only when such knowledge can not be proven. Computational proficiency is expected.

Students may elect research and a thesis (12 units). An oral defense of the thesis is held after the typed written thesis has been submitted. A grade of A or B is required. Students not electing to write a thesis may elect to take from 3 to 6 units of guided studies (BI 871, BI 872) with submission of a written report. Electives can be selected from the existing courses.

## Main Courses for the MS degree:

Basic Core Courses		Units
BI 751	Chemical Foundation for Bioinformatics	3
BI 752	Biological Foundation for Bioinformatics	3
MA 5414	Stringology: Mathematics of String Comparisons in Computational Biology	4
CM 5714	Molecular Modeling and Simulation	4

Required Core Courses		Units
BI 753	Bioinformatics I: Sequence Analysis	3
BI 754	Bioinformatics II: Protein Structure	3
BI 755	Bioinformatics III: Functional Prediction	3
BI 756	Cheminformatics	3

## GRADUATE COURSES

### BI 751 Chemical Foundation for Bioinformatics 2/0:0:3

An intensive review of those aspects of organic chemistry and biochemistry necessary to begin research in bioinformatics and to enter graduate courses in biology. Covalent bonding, quantum mechanical basis of bond formation, three-dimensional structure of molecules, reaction mechanisms, catalysis, polymers, enzymes, thermodynamic and kinetic foundations, metabolic pathways, sequence and structure of macromolecules. This course makes extensive use of computer approaches to convey the essential computational and visual nature of material to be covered. *Prerequisites:* CM102, general chemistry, general physics, organic chemistry and calculus.

**BI 752 Biological Foundation for Bioinformatics** 2½:0:0:3

An intensive review of those aspects of biochemistry, molecular biology and cell biology necessary to begin research in bioinformatics and to enter graduate courses in biology. Areas covered include cell structure, intracellular sorting, cellular signaling (receptors), Cytoskeleton, cell cycle, DNA replication, transcription and translation. This course makes extensive use of computer approaches to convey the essential computational and visual nature of the material to be covered. *Prerequisites: CM102, general chemistry, general physics, organic chemistry, calculus or instructor's permission.*

**BI 753 Bioinformatics I: Sequence Analysis** 1¼:2:0:3

Computer representations of nucleic acid and protein sequences, pairwise and multiple alignment methods, available databases of nucleic acid and protein sequences, database search methods, scoring functions for assessment of alignments, nucleic acid to protein sequence translation and codon usage, genomic organization and gene structure in prokaryotes and eukaryotes, introns and exons, prediction of open reading frames, alternative splicing, existing databases of mRNA, DNA protein and genomic information. An overview of available programs and of resources on the web. Lab is an integral part of this course.

**BI 754 Bioinformatics II: Protein Structure** 1¼:2:0:3

Protein folding representations, databases of protein folding classes, secondary structure prediction, tertiary structure prediction via computer folding experiments threading, and homology model building, prediction of post translation modification sites, active and binding sites in proteins, representations of contiguous and non-contiguous epitopes on protein surfaces at the sequence level, representations of functional motifs at the three dimensional and at the sequence level. Lab is an integral part of this course. *Prerequisite: BI 753.*

**BI755 Bioinformatics III: Functional Prediction** 1¼:2:0:3

Functional classifications of proteins, prediction of function from sequence and structure, Orthologs and paralogs, representations of biological pathways, available systems for the analysis of whole genomes and for human-assisted and automatic functional prediction. Lab is an integral part of this course. *Prerequisite: BI 754.*

**BI 756 Chemoinformatics** 1¼:2:0:3

Review of Database Theory, Chemical Structure Representation; connection tables, line notations, structure diagrams, Representations of Chemical Reactions, Structure manipulation: Graph Theory, Structure Analysis: ring perception, structural fingerprints, symmetry perception, Molecular Modeling Algorithms, Genetic Algorithms, Simulated Annealing, QSAR historical approaches, Structural Search of Chemical Databases, Commercial Chemical Information Databases, Combinatorial Chemistry and diversity assessment.

**BI 757 Special Topics in Bioinformatics Presentation at intervals of various advanced or specialized topics in chemo- or bioinformatics.****BI758/9 Guided Studies in Bioinformatics** as arranged**BI 760 Research in Bioinformatics** as arranged

Original research, which serves as basis for master's degree. Minimum research registration requirements for the master's thesis: 12 units. Registration for research required each semester consecutively until students have completed adequate research projects and acceptable theses and have passed required oral examinations. Research credits registered for each semester realistically reflect time devoted to research. *Prerequisites for MS candidates: degree status and consent of graduate adviser and thesis director.*

**CM 5714 Molecular Modeling and Simulation** 3:3:0:4

Introduces students to principles and applications of modern molecular modeling and simulations methods, using commercially available software packages on powerful computer workstations. Algorithms for visualization and prediction of structural and physical properties of molecules and molecular aggregates are taught, based on the principles of quantum, classical and statistical mechanics, which will be reviewed in a mathematically simplified form. The accompanying laboratory part focuses on acquiring practical skills in application of commercial software packages to illustrative problems in physical chemistry, chemical engineering, biology and medicine. *Prerequisites: completion of core undergraduate courses in mathematics and science (grade C or better) in CM, CH, ME, EE, CS, PH, CE or equivalent.*

**MA 5414 Stringology: Mathematics of String Comparisons in Computational Biology** 3:3:0:4

Addresses basic combinatorial problems of string manipulation - string matching, string editing, string distance computations, arising from areas of text processing, computational biology and genomics. Classical, modern and entirely new approaches to these problems will be presented with all necessary mathematical and computer science backgrounds. Emphasis is on practical and effective algorithm implementations.