Recent advances in biological sciences make it possible to study the genome (all the genes of an organism) and the proteome (all of the proteins in an organism), making biology an information science. Computational algorithms and software tools then enable one to study this data to understand the processes of life are designed at the molecular level. This increased understanding has attendant practical benefits to mankind: a scientific basis for the design of drugs, the causes of diseases, better ways to decompose cellulose to make ethanol, an understanding of how all organisms are related, etc. Bioinformatics or computational biology is the moniker for this area of research, and it involves the development of string algorithms, combinatorial algorithms, databases, statistical methods, and software.

This year, we will consider several topics in this course: pairwise global alignments of genes and proteins; pairwise local alignment of proteins and database search; multiple alignment of proteins and phylogenetic trees; secondary and tertiary structures of proteins; profile-based database search; hidden Markov models; data mining in proteomic networks; ontologies in proteomics; protein families; and other topics of interest to participating students.

Grades will be based on homework problems, programming assignments, and projects. Students will be required to present the results of their projects in class.

One project that you will have the opportunity to work on is with Professor Lesley Greene of the Chemistry department. She is trying to understand how proteins fold after they are synthesized to assume their native shape, which is critical to their function. She conjectures that interactions between amino acids in the proteins that are far apart along the protein sequence, but which are close to each other in the 3-dimensional geometry are critical to the way in which a protein folds. In particular, those interactions that are conserved across protein families should determine some of the significant steps along this folding pathway. She will try to verify this conjecture using both computations and experiments.

No previous knowledge of bioinformatics will be assumed in the course. We will begin with a quick review of some of the biological concepts needed for
bioinformatics, and will learn more as we go along. Pre-requisite for computer science students will be a course in algorithms, preferably at the graduate level. Students from other departments could be accommodated, but should contact the instructor before signing up for the course.

You could use one of two textbooks for the course. The first is *Protein Bioinformatics: An Algorithmic Approach to Sequence and Structure Analysis*, by Ingvard Eidhammer, Inge Jonassen, and William R. Taylor, John Wiley and Sons, 2004. The second is *Understanding Bioinformatics* by Marketa Zvelebil and Jeremy O. Baum, Garland Science, 2008. These are available on Amazon and you should be able to obtain a used or new copy by next week.