

Project in Proteomic Bioinformatics at Bio21 (The university of Melbourne)

Supervisors: Dr David Perkins and Dr Nicholas Williamson (nawill@unimelb.edu.au)

We offer several projects in the area of Proteomic Bioinformatics (using computational methods to explore protein datasets). As proteomic bioinformatics is a rapidly expanding field with few practitioners (most bioinformaticians have worked on DNA or genomic problems), this project will appeal to those who would like to pursue a professional career in bioinformatics with the advantage of a rare but highly desirable skill set.

Definition of Protein Sequence Motifs

Important regions of protein sequences are conserved over different species and in some cases between proteins of different functions. A good example of this is the Major Histocompatibility Complex (MHC) binding regions of some proteins associated with auto-immune diseases such as arthritis. In this case there are two major classes, class I has a well defined conserved binding motif whereas class II is less well characterized. We have identified a large number of potentially novel MHC binding proteins and this project is designed to mine this data for prognostic amino acid patterns. Various alignment, motif definition and sequence searching algorithms will be tested and benchmarked to ultimately produce a system for the identification of predictive amino acid patterns within MHC binding proteins. Preferred knowledge: either C,C++ or Java programming and PERL scripting. Some biological background or interest would be beneficial

Improving Protein Identification from Mass Spectrometry Data

Modern methods for identifying and sequencing proteins associated with disease are based on the accurate mass measurements of protein fragments using mass spectrometers. From this very accurate mass data, peptide sequences are reconstituted and ultimately proteins are identified by combining the peptide sequences. A major concern is accurately identifying which peptide belongs to which protein, this project aims to address this issue using modern bioinformatics techniques and algorithms.

Preferred knowledge: either C,C++ or Java programming and PERL scripting. Some biological background or interest would be beneficial

Modeling Peptide Fragmentation

Proteins are sequenced and identified by reconstituting sequences from fragment mass data collected on mass spectrometers. An important part of this process is attempting to distinguish real protein data

from chemical and electronic noise. This project will model the fragmentation efficiency of proteins based on a number of parameters associated with their sequence. Using these models it will be possible to predict the actual fragmentation patterns of proteins and compare this to experimental data allowing protein identification to be both faster and have a higher degree of confidence.

Preferred knowledge: either C,C++ programming, R or SPSS or MATLAB. Some biological background or interest would be beneficial