CIBM PROGRAM RETREAT

Computation and Informatics in Biology and Medicine
Training Program Retreat

Michael Giddings

Assistant Professor
Departments of Microbiology & Immunology and Biomedical Engineering
The University of North Carolina at Chapel Hill

Genome Analysis by Proteomic Methods and Genome Fingerprint Scanning

Abstract: Proteomics is rapidly developing as a field. Major topics of research include understanding protein expression quantity and timing, as well as determining structural attributes of proteins on a cell-wide basis. These studies have intrinsic value for their ability to elucidate protein function. But such efforts also have another potential, which is to provide significant new information about the genome from whence the proteins are derived. In even the most thoroughly studied genomes, questions remain on the actual number of functional genes, on the number and identity of distinct functional products of those genes, on how genomes evolve to produce proteins that help them adapt, etc.

Though proteomics holds the promise of elucidating these questions, current technologies are limited when applied to such purposes. Protein separation and detection both require further improvement, but these are not the biggest limitation - the software is. Two significant challenges are present. The first is the non-linearity of the protein expression process. In almost all organisms, alternative protein-coding strategies have been demonstrated that result in products not conforming to the standard rules, and such products are quite difficult to predict from the gene sequence alone. These phenomena greatly complicate the process of tying an observed protein back to the gene that encoded it. Furthermore, the databases that proteomic methods rely on are incomplete, and proteomic methods that rely upon standard databases are often hobbled by these limitations.

We are working on new strategies and methods to interpret mass spectrometry-based protein measurements, linking those directly to the raw genomic sequence. Our goal is to gain insight on how a proteome relates to the genome that produced it, without reliance on prior interpretation or annotation of the genome sequence. A primary tool in this effort is "Genome-based peptide fingerprint scanning" (GFS) that utilizes both peptide-mass fingerprints and tandem mass spectrometry data to directly identify the genomic locus responsible for expressing a measured protein, and to elucidate gene-structure based on the observed protein. I will discuss the challenges of linking genome to proteome, the methods we are using to approach this, and results obtained to date applying them to both bacterial and mammalian studies.

Friday, October 3rd

3:30 p.m.
Pyle Center, Room 325/326
702 Langdon Street