

## **BIOINFORMATICS AT NIAID BIODEFENSE PROTEOMICS ADMINISTRATIVE RESOURCE CENTER**

P. B. McGarvey, C. H. Wu

Protein Information Resource, Georgetown University Medical Center, Washington, DC,  
USA

The National Institute for Allergy and Infectious Diseases (NIAID) has created a biodefense proteomics program with the goal to “identify and validate therapeutic drug targets for the next generation of vaccines, therapeutics, and diagnostics” for infectious agents of concern in bioterrorism. The program consists of seven Proteomics Research Centers and an Administrative Resource Center. In addition to administrative support, the Admin Center is developing a bioinformatics infrastructure that includes a web portal for public dissemination, a central proteomic database hosting the data and technology protocols generated by each Proteomics Center, and a protein knowledge system for data integration, data mining and data analysis.

Administering such a program presents unique challenges as a large and highly diverse set of proteomic data will have to be integrated and managed in a way useful to knowledge discovery. The heterogeneous data types include: genome-scale protein identifications from mass spectroscopy, complete ORFeome cloning and expression, one and 2-D gel electrophoresis, protein interaction data from yeast two-hybrid experiments and protein array chips, antibody binding on western and protein arrays, microarray experiments, electron microscopy, and high resolution NMR and X-ray structures. An Interoperability Working Group has been established to discuss common data standards, exchange formats, standard operating procedures, controlled vocabularies and ontologies. Community standards, such as those developed at the HUPO PSI (Proteomics Standards Initiative) and NCI caBIG (cancer Biomedical Informatics Grid), are being evaluated for adoption. To facilitate function and pathway analysis of the proteomic data, a protein knowledge system is being developed at the Protein Information Resource (PIR) for expression data analysis, including gene/peptide to protein mapping, sequence analysis, protein information display, and function and pathway categorization and visualization.