

## **Protein Information Resource (PIR): An Integrated Bioinformatics Resource for Functional Proteomics**

McGarvey P, Hu ZZ, Wu CH. Protein Information Resource, Department of Biochemistry & Molecular Biology, Georgetown University Medical Center, 3900 Reservoir Road, NW, Washington, DC 20007-2195

The Protein Information Resource (PIR) is an integrated bioinformatics resource that supports functional proteomics. PIR is a member of UniProt—Universal Protein Resource—the central repository of protein sequence and function, which provides the UniProt Knowledgebase (UniProtKB) with extensively curated annotations, UniProt Reference clusters (UniRef) to speed sequence searches, and UniProt Archive (UniParc) to reflect sequence history. The PIRSF protein family classification system based on evolutionary relationships of full-length proteins is central to the PIR/UniProt functional annotation. iProClass is an integrated database of protein family, function, and structure that provides rich links to over 90 molecular biology databases, and serves as a gateway for data integration and functional associative analysis. PIR is developing a knowledge system for expression data analysis, including gene/peptide to protein mapping, sequence analysis, protein information display, and function and pathway categorization and visualization.

As part of the newly-established Administrative Center for the NIAID National Proteomics Research program, PIR will support data standards and interoperability, as well as functional analysis of data generated from the Proteomic Centers. PIR also participates in the NCI caBIG (cancer Biomedical Informatics Grid) initiative that aims at integrating biomedical informatics infrastructure for cancer research by developing protein object models, web services, controlled vocabularies and common data elements for a grid-enabled UniProt. PIR is collaborating with several text mining groups and has developed the iProLINK resource for literature mining and RLIMS-P literature mining system for extracting protein phosphorylation data from MEDLINE abstracts. PIR and UniProt are accessible from <http://pir.georgetown.edu/> and <http://www.uniprot.org/>, respectively.