PIR and *i*ProClass for Functional Genomics/Proteomics

Barker, W.C., Castro-Alvear, J., Chen, Y., Hu, Z., Huang, H., Ledley, R.S., Lewis, K.C., Orcutt, B.C., Suzek, B., Vinayaka, C.R., Wu, C.H., Yeh, L.L., and Zhang, J.
Protein Information Resource, National Biomedical Research Foundation, Georgetown University Medical Center, 3900 Reservoir Road, NW, Washington, DC 20007-2195

The genome projects have revolutionized the practice of biology and made the computational approaches increasingly important for deriving and evaluating hypotheses. The Protein Information Resource (PIR) is an integrated public resource for protein sequences and knowledge management based on family classification, an effective means for large-scale functional characterization of genes. The PIR Web site provides many useful databases such as the Protein Sequence Database and iProClass as well as data analysis and mining tools to facilitate the functional identification of proteins. The *i*ProClass database provides comprehensive family relationships at both global and local levels, as well as structural/functional classifications and features of proteins. Integrating information from different source databases, *i*ProClass currently consists of about 266,000 non-redundant PIR and SwissProt proteins organized with more than 30,000 superfamilies, 100,000 families, 3000 domains, 1300 motifs, 280 post-translational modification sites, and links to over 40 databases of protein families, structures, functions, genes, genomes, literature, and taxonomy. Future releases will be based on a new PIR non-redundant reference sequence database (NREF) containing more than 800,000 protein sequences. The iProClass employs a modular structure for scalability and extendibility, thereby providing a framework for integration of new data and/or software components in a distributed networking environment.

Presented at the Fifth Annual Conference On Computational Genomics, Baltimore, Maryland, 2001

[Back to Publications Page](#)