

Protein Information Resource

Integrated Protein Informatics Resource for Genomic & Proteomic Research

For four decades the Protein Information Resource (PIR) has provided databases and protein sequence analysis tools to the scientific community, including the Protein Sequence Database, which grew out from the Atlas of Protein Sequence and Structure, edited by Margaret Dayhoff [1965-1978]. Currently, PIR major activities include: i) **UniProt** (Universal Protein Resource) development, ii) **iProClass** protein data integration and ID mapping, iii) **PIRSF** protein classification, and iv) **iProLINK** protein literature mining and ontology development.



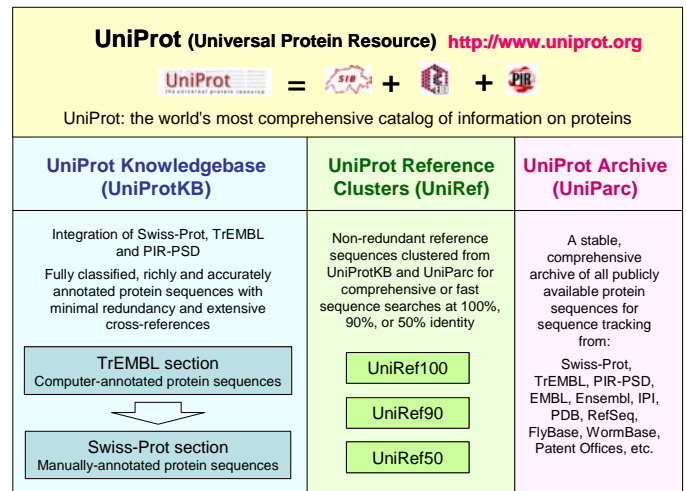
UniProt – Universal Protein Resource

What is UniProt?

UniProt is the central resource for storing and interconnecting information from large and disparate sources and the most comprehensive catalog of protein sequence and functional annotation.

When to use UniProt databases

Use UniProtKB to retrieve curated, reliable, comprehensive information on proteins. Use UniRef to decrease redundancy and speed up sequence similarity searches. Use UniParc to access to archived sequences and their source databases.



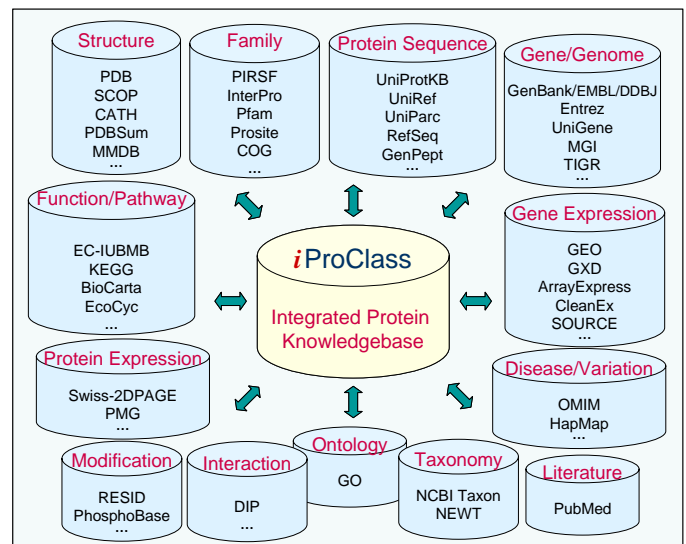
iProClass – Integrated Protein Knowledgebase

What is iProClass?

iProClass provides extensive data integration of over 90 biological databases, with protein ID mapping service, and executive summary descriptions of proteins for UniProtKB and selected UniParc protein sequences.

When to use iProClass

Use iProClass to retrieve comprehensive, up-to-date information about a protein, including function, pathway, interactions, family classification, structure and structural classification, gene and genomes, ontology, literature, and taxonomy. Use iProClass to access to ID mapping, protein BioThesaurus and related sequences.






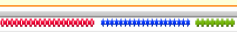

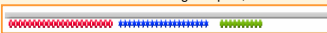






PIRSF – Protein Family Classification System

What is PIRSF?

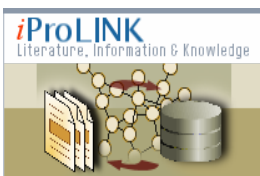
The PIRSF protein classification system is a network with multiple levels of sequence diversity from superfamilies to subfamilies that reflects the evolutionary relationship of full-length proteins and domains. The primary PIRSF classification unit is the homeomorphic family, whose members are both homologous (evolved from a common ancestor) and homeomorphic (sharing full-length sequence similarity and a common domain architecture). PIRSFs are manually curated for membership, annotation of specific biological functions, biochemical activities, and sequence features. In addition, rules for functional sites and protein names are created to assist in the propagation and standardization of protein annotation and the systematic detection of annotation errors.

When to use PIRSFs

The PIRSF report offers a single platform for studying evolutionarily-related proteins. It summarizes distinctive features of the family such as family name, taxonomic distribution, hierarchy and domain architecture, as well as information about family members, including function, structure, pathway, ontology and family classification, with extensive links to the corresponding databases. Use this information to obtain accurate functional data on your protein of interest and/or to predict function and other properties of uncharacterized members of the family.

Domain Superfamily	PIRSF Superfamily	PIRSF Homeomorphic Family	PIRSF Homeomorphic Subfamily
PF02735: Ku70/Ku80 beta-barrel domain 	PIRSF800001: Ku DNA-binding complex, Ku70/80 subunits 	PIRSF003033: Ku DNA-binding complex, Ku70 subunit  PIRSF016570: Ku DNA-binding complex, Ku80 subunit 	
		PIRSF006493: Ku DNA-binding protein, prokaryotic type 	
PF00219: Insulin-like growth factor binding protein (IGFBP) 		PIRSF001969: insulin-like growth factor binding protein 	PIRSF500001: IGFBP-1  ... PIRSF500006: IGFBP-6 
		PIRSF018239: IGFBP-related protein, MAC25 type 	

pir.georgetown.edu/pirsf



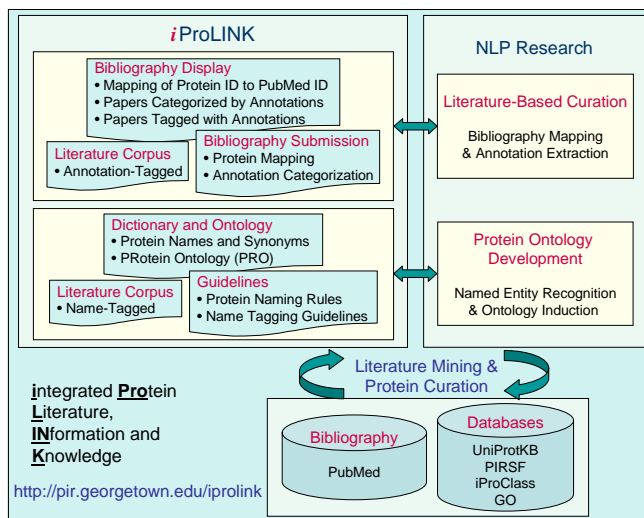
iProLINK – Integrated Protein Literature, Information and Knowledge

What is iProLINK?

iProLINK provides annotated literature, protein name dictionary, and other information to facilitate Natural Language Processing technology development in literature mining, database curation, and protein name tagging and ontology.

When to use iProLINK

Use iProLINK to obtain literature sources that describe protein entries (Bibliography Mapping), to map protein/gene names to UniProtKB entries (BioThesaurus), to obtain annotated data sets for developing text mining algorithms, to mine literature for protein phosphorylation (RLIMS-P), and to obtain information on protein ontology (PRO).



pir.georgetown.edu/iprolink

