Protein Information Resource

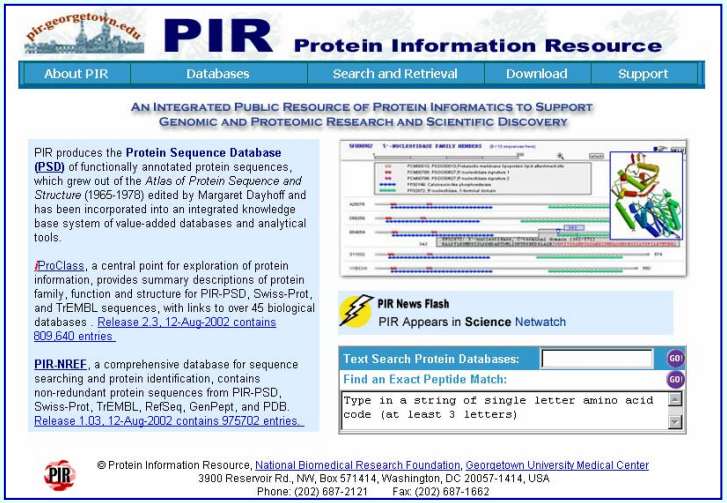
NCI caBIG Integrative Cancer Research Workspace
Pathways SIG & Genome Annotation SIG

July 6, 2004

Cathy H. Wu, Ph.D.
Director, PIR
Professor, Georgetown University Medical Center

Protein Information Resource

Integrated Public Resource of Protein Informatics for Genomic/Proteomic Research & Scientific Discovery



http://pir.georgetown.edu

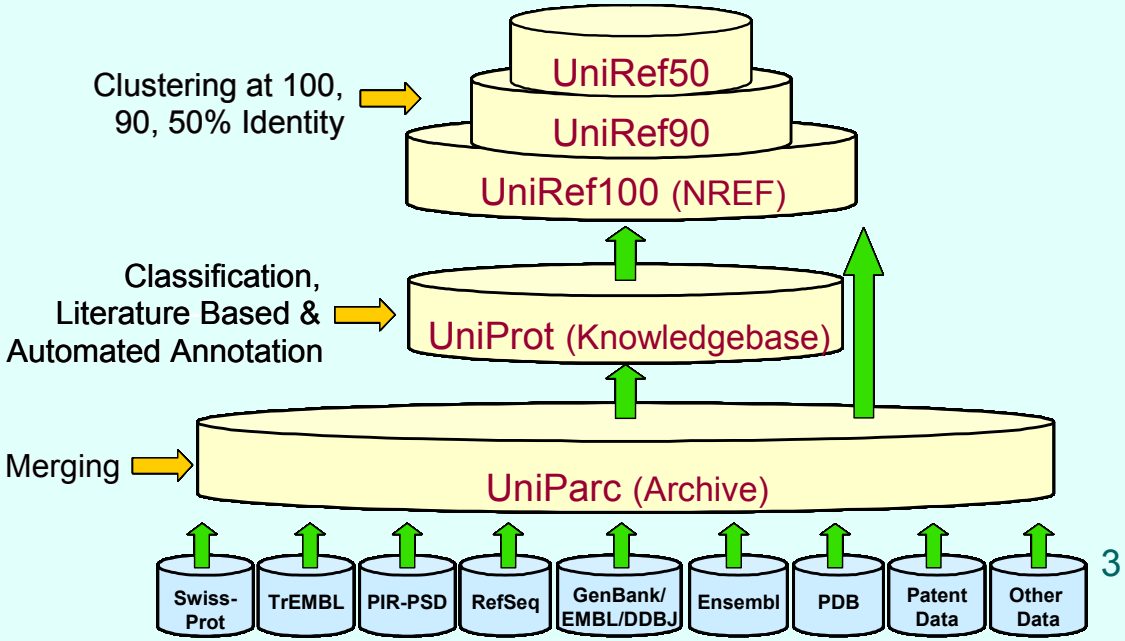
- UniProt Universal Protein Resource: Central Resource of Protein Sequence and Function
- PIRSF Family Classification System: Protein Classification and Functional Annotation
- iProClass Integrated Protein Knowledgebase: Data Integration and Functional Associative Analysis
- Interoperability and Dissemination: Ontology, XML, Relational Schema, Object Modeling

UniProt: Universal Protein Resource

Central Resource of Protein Sequence and Function

- International Consortium: PIR, EBI, SIB
- Unifies PIR-PSD, Swiss-Prot, TrEMBL
- UniParc: Comprehensive Sequence Archive with Sequence History
- UniRef: Non-redundant Reference Databases for Sequence Search
- UniProt: Knowledgebase with Full Classification and Functional Annotation

http://www.uniprot.org



UniProt Knowledgebase

- Objective: Stable, Comprehensive, Fully Classified, Richly and Accurately Annotated
- Describe in a single record all protein products derived from a certain gene in a given species
- Information Content
  - *Isoform Presentation*: Alternatively Spliced Forms, Proteolytic Cleavage, and Post-Translational Modification (each with FTid)
  - *Nomenclature*: Gene/Protein Names (Nomenclature Committees)
  - *Family Classification and Domain Identification*: InterPro and PIRSF
  - *Functional Annotation*: Function, Functional Site, Developmental Stage, Catalytic Activity, Modification, Regulation, Induction, Pathway, Tissue Specificity, Subcellular Location, Disease, Process

UniProt

PIR View

UniProt Entry: P00439

ENTRY INFORMATION	
ENTRY NAME	PH4H_HUMAN
ACCESSION NUMBERS	P00439; Q16717; Q8TC14
CREATED	Release 01, 21-JUL-1986
SEQUENCE UPDATE	Release 01, 21-JUL-1986
ANNOTATION UPDATE	Release 44, 05-JUL-2004

NAME AND ORIGIN OF THE PROTEIN	
PROTEIN NAME	Phenylalanine-4-hydroxylase
DESCRIPTION	(EC 1.14.16.1) . PAH; Phe-4-monooxygenase
GENE NAME	PAH
SOURCE ORGANISM	Homo sapiens
TAXONOMY ID	9606 [NCBI, NEWT]
LINEAGE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mam

REFERENCES

[1] Kwok SCM; Ledley FD; Dilella AG; Robson KJH; Woo SLC
Nucleotide sequence of a full-length complementary DNA clone an
1985, *Biochemistry*, 24, 556-561
Position: SEQUENCE FROM N.A.
Comments: tissue=Liver
PubMed: 2986678; Medline: 85199778;

COMMENTS	
CATALYTIC ACTIVITY	L-phenylalanine + tetrahydrobiopterin + O(2) = L-tyrosine + 4a-hydroxytetrahydrobio
COFACTOR	Ferrous ion.
ENZYME REGULATION	N-terminal region of PAH is thought to contain allosteric binding sites for phenylalanine and to constitute an "inhibitory" regulates the activity of a catalytic domain in the C-terminal portion of the molecule.
PATHWAY	Catabolism of phenylalanine; first (rate-limiting) step.
SUBUNIT	Homodimer.
POLYMORPHISM	The Glu-274 variant occurs on approximately 4% of African-American PAH alleles. The enzyme activity of the variant indistinguishable from that of the wild-type form.
DISEASE	Defects in PAH are the cause of phenylketonuria (PKU) [MIM:261600]. PKU is an autosomal recessive inborn error phenylalanine metabolism, due to severe phenylalanine hydroxylase deficiency. It is characterized by blood concentratio phenylalanine persistently above 1200 mumol (normal concentration 100 mumol) which usually causes mental retardatio low phenylalanine diet is introduced early in life). They tend to have light pigmentation, rashes similar to eczema, epileps hyperactivity, psychotic states and an unpleasant "mousy" odor.
DISEASE	Defects in PAH are the cause of non-phenylketonuria hyperphenylalaninemia (Non-PKU HPA) [MIM:261600]. Non- is a mild form of phenylalanine hydroxylase deficiency characterized by phenylalanine levels persistently below 600 mun allows normal intellectual and behavioral development without treatment. Non-PKU HPA is usually caused by the com of a mild hyperphenylalaninemia mutation and a severe one.
DISEASE	Defects in PAH are the cause of hyperphenylalaninemia (HPA) [MIM:261600]. HPA is the mildest form of phenylalani hydroxylase deficiency.
SIMILARITY	Belongs to the biopterin-dependent aromatic amino acid hydroxylase family.
ONLINE INFORMATION	PAHdb; Phenylalanine hydroxylase locus knowledgebase; Belongs to the biopterin-dependent aromatic amino acid hydroxylase family.

DATABASE CROSS-REFERENCES	
EC	1.14.16.1.
EMBL	K03020,AAA60082.1. [GenBank, DDBJ] U49897,AAC51772.1. [GenBank, DDBJ] S61296,AAD13926.1. [GenBank, DDBJ] BC026251,AAH26251.1. [GenBank, DDBJ]
GENEW	HGNC:8582,PAH.
GO	GO:0004505,F:phenylalanin GO:0008652,P:amino acid b
HSC_2DPAGE	P00439,HUMAN.
INTERPRO	IPR001273,Aaa_hydroxylas IPR002912,ACT. IPR005961,Phe4hydrxlaseT
MIM	261600.
PDB	1DMW,2001-03-24. 1J8T,2002-05-22. 1J8U,2002-05-22. 1KW0,2003-01-28. 1LRM,2002-06-12. 1MMK,2003-09-04. 1MMT,2003-09-04. 1PAH,1999-01-13. 2PAH,1999-10-06. 3PAH,1999-04-27. 4PAH,1999-04-27. 5PAH,1999-04-27.
PFAM	
PIR	
PRINTS	
PRODOM	
PROSITE	
REACTOME	
TIGRFAMS	

KEYWORDS

Oxidoreductase; Monooxygenase; Allosteric enzyme; Phenylketonuria; Phosphorylation; Phenylalanine catabolism; Iron; Disease mutation; Polymorphism; 3D-structure

FEATURES

Feature	Description	Begin Position	End Position	Length
MOD_RES	Phosphoserine (by PKA) (BY SIMILARITY)	16	16	1
METAL	Iron (BY SIMILARITY)	285	285	1
METAL	Iron (BY SIMILARITY)	290	290	1
METAL	Iron (BY SIMILARITY)	330	330	1
VARIANT	S -> P (in PKU) /FTId=VAR_000869	16	16	1
VARIANT	Q -> L (in HPA) /FTId=VAR_009239	20	20	1
VARIANT	F -> L (in PKU; haplotype 1) /FTId=VAR_000870	39	39	1
VARIANT	(in PKU; haplotypes 9,21) /FTId=VAR_000871	39	39	1
VARIANT	S -> L (in PKU) /FTId=VAR_000872	40	40	1
VARIANT	L -> F (in PKU) /FTId=VAR_000873	41	41	1
VARIANT	L -> P (in PKU; mild) /FTId=VAR_009240	41	41	1

ADDITIONAL INFORMATION FROM iProClass — Go to iProClass

CROSS-REFERENCE	RefSeq: NP_000268 phenylalanine hydroxylase LocusLink: 5053 phenylalanine hydroxylase(PAH) NCBI GI#: 2462722; 189937; 4557819
PIRSF FAMILY	PIRSF000336: phenylalanine 4-monooxygenase
GENE ONTOLOGY	*Molecular Function* GO:0004505: phenylalanine 4-monooxygenase activity [INTERPRO; evidence:IEA] [SPEC; evidence:IEA] [PMID:3856322; evidence:TAS] GO:0016597: amino acid binding [INTERPRO; evidence:IEA] GO:0005506: iron ion binding [INTERPRO; evidence:IEA] GO:0004497: monooxygenase activity [INTERPRO; evidence:IEA] [SPKW; evidence:IEA] GO:0003824: catalytic activity [SPKW; evidence:IEA] GO:0016491: oxidoreductase activity [SPKW; evidence:IEA] *Biological Process* GO:0009072: aromatic amino acid family metabolism [INTERPRO; evidence:IEA] GO:0008652: amino acid biosynthesis [PMID:3856322; evidence:TAS] GO:0008152: metabolism [INTERPRO; evidence:IEA] GO:0006559: phenylalanine catabolism [INTERPRO; evidence:IEA] [SPKW; evidence:IEA] [UniProt:P00439; evidence:none]
ENZYME/FUNCTION	EC 1.14.16.1 EC-IUBMB , KEGG, BRENDA, WIT, MetaCyc *Nomenclature:* Oxidoreductases; Acting on paired donors, with incorporation or reduction of molecular oxygen; With reduced pteridine as one donor, and incorporation of one atom of oxygen; phenylalanine 4-monooxygenase *Reaction:* L-phenylalanine + tetrahydrobiopterin + O_2 = L-tyrosine + 4a-hydroxytetrahydrobiopterin
PATHWAY	KEGG: Metabolism; Amino Acid Metabolism; Phenylalanine, tyrosine and tryptophan biosynthesis [PATH:hsa00400].
STRUCTURE	PDB: 1PAH(117-424,100.0%) ; 1DMW:A(118-424,100.0%) ; More 1DMW: SCOP CATH FSSP MMDB PDBsum 1IN9: SCOP CATH FSSP MMDB PDBsum More

http://www.pir.uniprot.org/cgi-bin/upEntry?id=PH4H_HUMAN

PIRSF Classification System

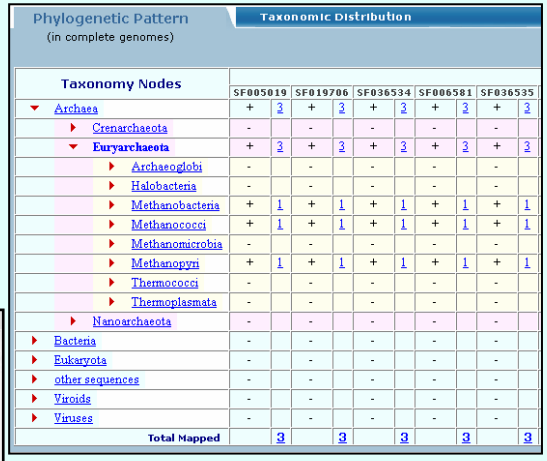
Protein Classification and Functional Annotation

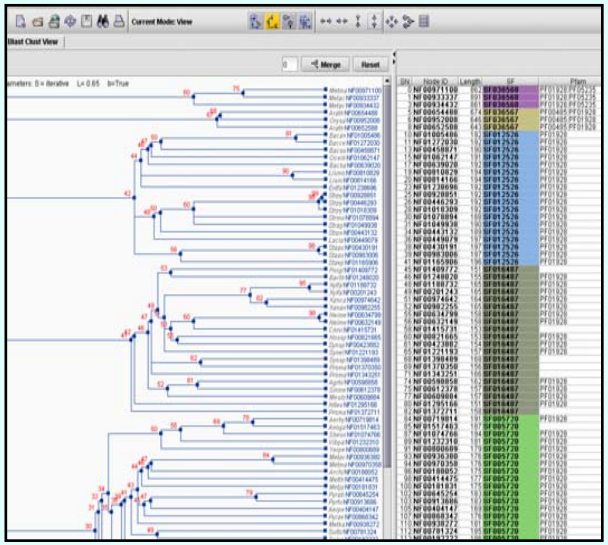
- A Network Structure from Superfamilies to Subfamilies to Reflect Evolutionary Relationships of Full-length Proteins and Domains
- Comprehensive Classification of All UniProt Proteins
- Curated Families
  - Integrated into InterPro
  - Functional Site Rules and Protein Name Rules
- Advantages
  - Annotation of generic biochemical and specific biological functions
  - Accurate propagation of annotation and development of standardized protein nomenclature and ontology
- PIRSF Web Site: http://pir.georgetown.edu/pirsf/

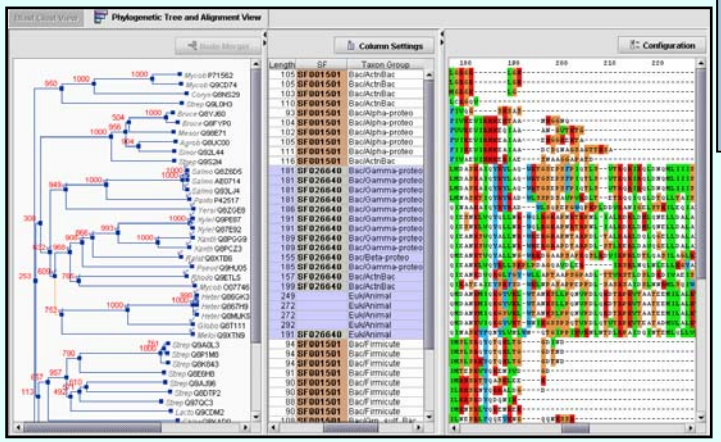
Domain Superfamily • One common Pfam domain	**PIRSF Superfamily** • 0 or more levels • One or more common domains	**PIRSF Homeomorphic Family** • Exactly one level • Full-length sequence similarity and common domain architecture	**PIRSF Homeomorphic Subfamily** • 0 or more levels • Functional specialization
PF02735: Ku70/Ku80 beta-barrel domain	PIRSF800001: Ku70/80 autoantigen	PIRSF003033: Ku70 autoantigen PIRSF016570: Ku80 autoantigen	
		PIRSF006493: Ku, prokaryotic type	
PF00219: Insulin-like growth factor binding protein (IGFBP)		PIRSF001969: IGFBP	PIRSF500001: IGFBP-1 … PIRSF500006: IGFBP-6
		PIRSF018239: IGFBP-related protein, MAC25 type	
PF01817: Chorismate mutase (CM)		PIRSF017318: CM of AroQ class, eukaryotic type PIRSF001501: CM of AroQ class, prokaryotic type PIRSF026640: Periplasmic CM PIRSF001500: Bifunctional CM/PDT (P-protein) PIRSF001499: Bifunctional CM/PDH (T-protein)	
PF02153: Prephenate dehydrogenase (PDH)		PIRSF001499: Bifunctional CM/PDH (T-protein) PIRSF006786: PDH, feedback inhibition-insensitive PIRSF005547: PDH, feedback inhibition-sensitive	

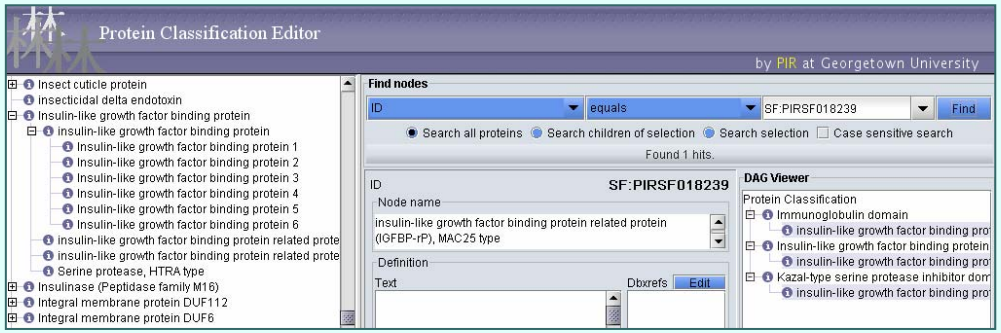
Classification & Visualization Tools

- Taxonomy Distribution and Phylogenetic Pattern
- Iterative BlastClust Tree with Annotation Table
- Multiple Alignment and Phylogenetic Tree
- PIRSF Classification in DAG









PIRSF Family Report

Summary Report for PIRSF000876

GENERAL INFORMATION

PIRSF Number	PIRSF000876 *Curation Status*: Full (with Description)
PIRSF Name	**chemotaxis response regulator methylesterase, CheB type [Validated]**
PIRSF Size	Total Families=18; Total Sequence Entries=80 (79 Proteins+1 Fragments)
Taxonomy Range	Eukaryotae=0; Prokaryotae=71; Archaea=9; Viruses=0; Other=0 (click on the image to see a detailed taxonomic distribution.)
Length Range	Minimum=184; Maximum=403; Average=341; Standard Deviation=52
Keyword	complete proteome(64); sensory transduction(38); phosphorylation(37); hydrolase(34); chemotaxis(30); carboxylic ester hydrolase(9); phosphoprotein(9); plasmid(3); hypothetical protein(2); 3d-structure(1)
Description	Members of this group are response regulators involved in chemoreceptor modification. In bacterial chemotaxis, cellular movement is directed in response to chemical gradients. Transmembrane chemoreceptors that sense the stimuli are coupled (via a coupling protein, CheW) with a signal transduction histidine kinase (CheA). CheA phosphorylates response regulators CheB and CheY. Phosphorylated CheY binds to FliM, a component of the flagellar motor switch complex, and modulates the direction of flagellar rotation (10049806) . Response regulator CheB (receptor modification enzyme, protein-glutamate methylesterase) modulates the signaling output of the chemotaxis receptors through control of the level of chemoreceptor methylation (10049806) . Specific glutamyl residues in the transmembrane chemoreceptor cytoplasmic domain are methylated by methyltransferase CheR to form gamma-carboxyl glutamyl methyl esters. These esters can be hydrolyzed by methylesterase CheB. Receptor modification resets the signaling states of receptors, allowing for responses to changes in concentration of the chemical stimuli irrespective of their absolute concentrations (10049806) . Response regulators of the microbial two-component signal transduction systems typically consist of an N-terminal CheY-like receiver domain and a C-terminal output (usually DNA-binding) domain (10066457) (11405410) . In members of this group, the output domain is an enzymatic domain, protein-glutamate methylesterase (demethylase, EC 3.1.1.61). In response to an environmental stimulus, a phosphoryl group is transferred from the His residue of a signal transduction histidine kinase to an Asp residue in the CheY-like receiver domain of the cognate response regulator. Phosphorylation of the receiver domain induces conformational changes that activate an associated output domain. Phosphorylation-induced conformational changes in the response regulator molecule have been demonstrated in direct structural studies (11851334) . In members of this group, phosphorylation of receiver domain activates the methylesterase (2677005) , resulting in the subsequent demethylation of the chemoreceptors.
Bibliography	PMID: 11406410; 9442881; 9760239; 10966457; 11851334; 2677005; 12196531; 10049806; 9687374; 8218244
Representative member	iProClass: P07330
Seed Members	iProClass: Q29221; O33558; O51376; O83639; O85128; O87002; O87717; P07330; P31758; Q05522; Q45047; Q8XCF9; Q8ZFM0; Q92YM4; Q97GZ3; Q9AAK0; Q9HQW6; Q9HXT8; Q9I6V9; Q9KA55; Q9KQD8; Q9KS59; Q9UYF3; Q9WYN9
Alignment and Tree	(click on the image to generate and display the multiple alignment and tree)
Domain Architecture	**PF00072:PF01339** (click on the image to display the seed member' domain architecture)

MEMBERSHIP

Prokaryotic Member	iProClass: O33558; O51376; O52262; O83639; O85128; O87002; O87125; O87717; P04042; P07330; P31758; P72253; Q05522; Q45047; Q52883; Q820K0; Q83R52; Q842G8; Q87MX5; Q88AU3; Q88658; Q88U8; Q8EEW5; Q8BGG5; Q8BM85; Q8BSQ1; Q8DAX6; Q8DG22; Q8DB67; Q8ECD1; Q8EF60; Q8EF61; Q8EQW0; Q8F3H4; Q8F3D8; Q8F6P9; Q8FGP1; Q8KLS5; Q8P3A6; Q8P9J7; Q8PBD9; Q8PIM5; Q8PLB4; Q8PMZ3; Q8RA12; Q8RAZ3; Q8VL08; Q8VR13; Q8XCF9; Q8XQ83; Q8YVX2; Q8Z5V2; Q8ZFM0; Q92YM4; Q93NZ9; Q97GZ3; Q98PD1; Q9AAK0; Q9FAD8; Q9HXT8; Q9I6V9; Q9KA55; Q9KQD8; Q9KS59; Q9L942; Q9PP09; Q9WYN9
Archaebacterial Member	iProClass: O29221; O58192; Q48300; Q8PX96; Q8Q009; Q8TLG9; Q8TUQ0; Q9HQW6; Q9UYF3
Model Organism	Escherichia coli:Q8XCF9

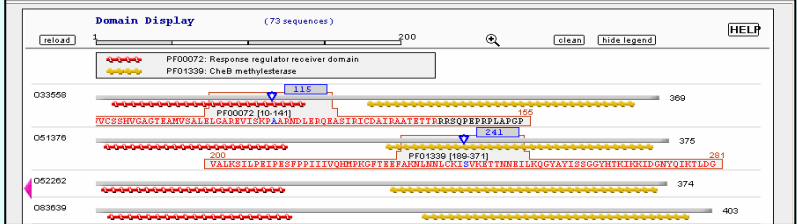
FUNCTION AND STRUCTURE

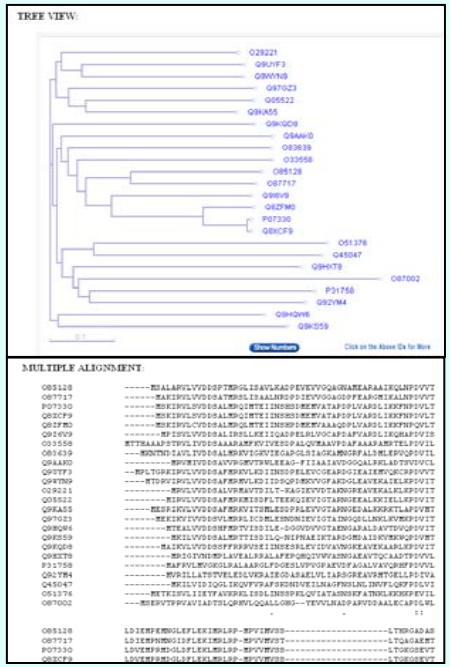
Ontology	*Molecular Function* GO:0004871: signal transducer activity (76) [HAMAP; evidence:IEA] [INTERPRO; evidence:IEA] GO:0003677: DNA binding (65) [INTERPRO; evidence:IEA] GO:0000156: two-component response regulator activity (65) [INTERPRO; evidence:IEA] GO:0016787: hydrolase activity (34) [SPKW; evidence:IEA] GO:0008984: protein-glutamate methylesterase activity (35) [HAMAP; evidence:IEA] [SPEC; evidence:IEA] *Biological Process* GO:0007606: sensory perception of chemical stimulus (29) [HAMAP; evidence:IEA] GO:0007165: signal transduction (76) [HAMAP; evidence:IEA] [INTERPRO; evidence:IEA] GO:0006355: regulation of transcription, DNA-dependent (65) [INTERPRO; evidence:IEA] GO:0006935: chemotaxis (76) [HAMAP; evidence:IEA] [INTERPRO; evidence:IEA] [SPKW; evidence:IEA] GO:0000160: two-component signal transduction system (phosphorelay) (66) [HAMAP; evidence:IEA] [INTERPRO; evidence:IEA] GO:0007600: sensory perception (38) [SPKW; evidence:IEA] *Cellular Component* GO:0046821: extrachromosomal DNA (3) [SPKW; evidence:IEA]
Enzyme/Function	EC 3.1.1.61 EC-IUBMB, KEGG, BRENDA, WIT, MetaCyc *Nomenclature*: Hydrolases; Acting on ester bonds; Carboxylic ester hydrolases; protein-glutamate methylesterase *Reaction*: protein L-glutamate O^2-methyl ester + H_2O = protein L-glutamate + methanol
Pathway	KEGG: Cellular Processes; Cell Motility; Bacterial chemotaxis [PATH: pho02030 afu02030 tma02030 hal02030 bsu02030 cac02030 bbu02030 bbu02030]. KEGG: Environmental Information Processing; Signal Transduction; Two-component system [PATH: afu02020 tma02020].
Structure	1A2O: PDB SCOP CATH FSSP MMDB PDBsum 1CHD: PDB SCOP CATH FSSP MMDB PDBsum

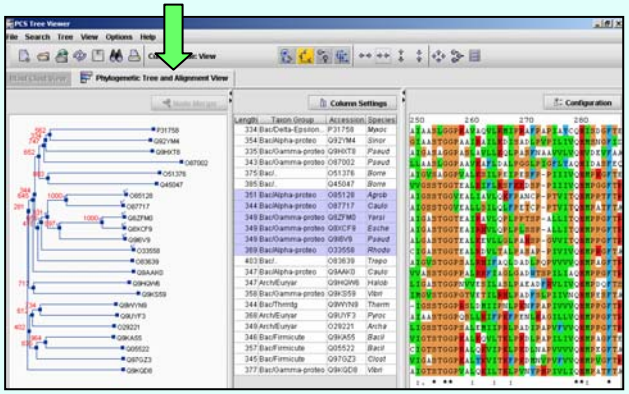
FAMILY RELATIONSHIP

Pfam Domain	Pfam: PF01339: CheB methylesterase(76) Pfam: PF00072: Response regulator receiver domain(63)
Prosite Motif	Prosite: PS50110: PDOC50110: Response regulatory domain profile.(66) Prosite: PS50122: PDOC50122: CheB-type methylesterase catalytic domain profile.(76)
COG	COG: COG2201
InterPro	InterPro: IPR008248: Chemotaxis response regulator, CheB type InterPro: IPR001789: Response regulator receiver InterPro: IPR001440: TPR repeat InterPro: IPR008262: Lipase, active site InterPro: IPR000673: CheB methylesterase
SCOP Fold	►*Class*: Alpha and beta proteins (a/b) ; *Fold*: Flavodoxin-like ; *Superfamily*: CheY-like ; *Family*: CheY-related [1A2O:A; 1A2O:B] ►*Class*: Alpha and beta proteins (a/b) ; *Fold*: Methylesterase CheB, C-terminal domain ; *Superfamily*: Methylesterase CheB, C-terminal domain ; *Family*: Methylesterase CheB, C-terminal domain [1A2O:A; 1A2O:B]
Other Classification	MetaFam: PIRSF000876 CDD: cd00156

DOMAIN/MOTIF DISPLAY







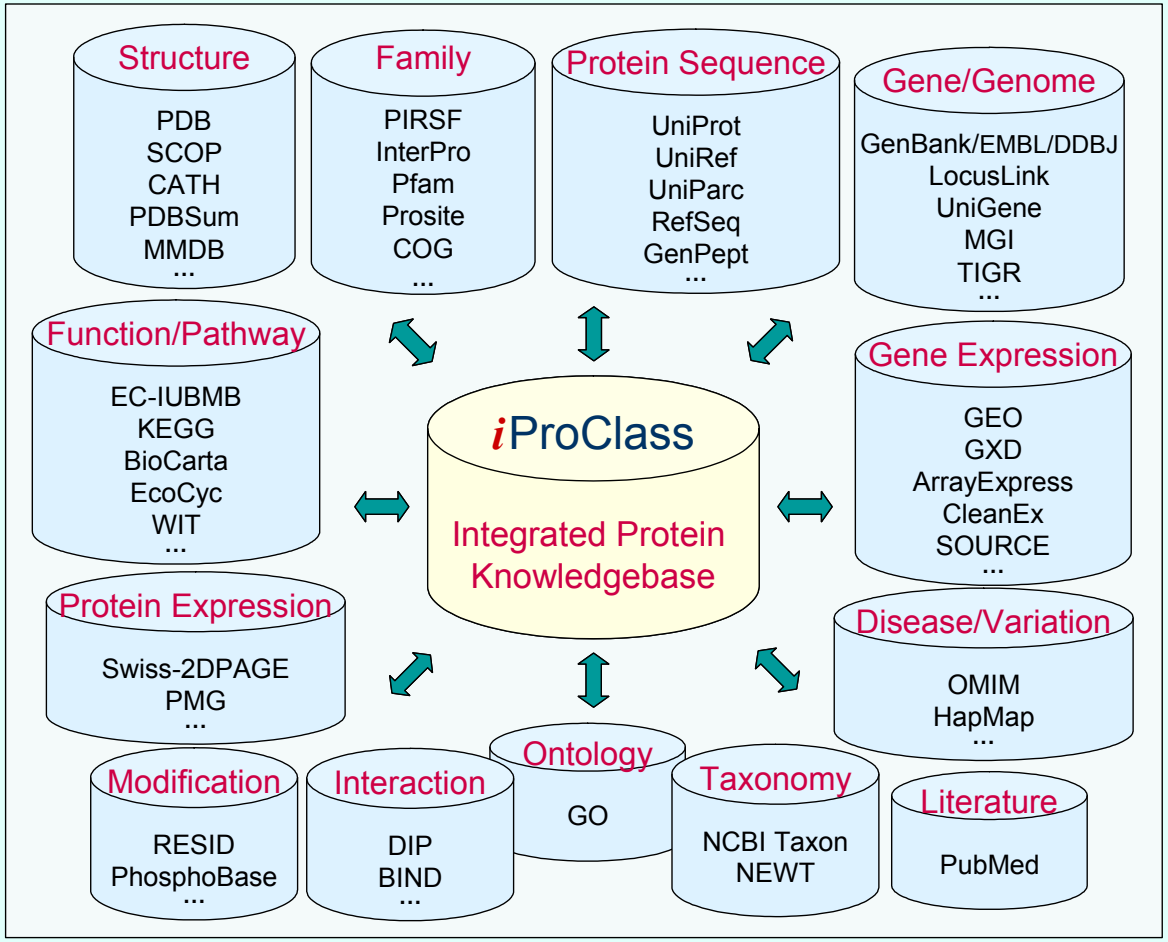
http://pir.georgetown.edu/cgi-bin/ipcSF?id=PIRSF000876

Biological Data Integration

- Challenge:
  - Voluminous, Complex, Dynamic Data from Heterogeneous Sources in Distributed Environment
- Data Warehouse
  - Local Copy of Databases in a Unified Database Schema
  - Allows Local Control of Data; Update Problem
- Hypertext Navigation
  - Browsing Model with Hypertext Links
  - Allows Direct Interaction; Easily Lost in Cyberspace
- iProClass Approach
  - Data Warehouse + Hypertext Navigation
  - Rich Links (Links + Executive Summaries)

iProClass Protein Knowledgebase

Data Integration and Functional Associative Analysis



- Integration of Protein Family, Function, Structure
- Rich Links to >80 Databases
- Value-Added Reports for 1.2 Million UniProt Proteins
- FTP: XML, Oracle, MySQL
- http://pir.georgetown.edu/iproclass

iProClass Protein Report

Summary Report for iProClass Entry: P35354 — *Related Sequences*

GENERAL INFORMATION

Protein Name and ID	Database / ID / Accession / Protein Name UniProt / PGH2_HUMAN / P35354; Q16876 / Prostaglandin G/H synthase 2 precursor (EC 1.14.99.1) (Cyclooxygen (Prostaglandin-endoperoxide synthase 2) (Prostaglandin H2 synthase 2 2) (PHS II) RefSeq: / g4506265 / NP_000954 / prostaglandin-endoperoxide synthase 2 precursor; prostaglandin G/H s PIR-NREF: NF00126043 GenPept: BAA05698.1 ; CAB41240.1 ; AAA57317.1 ; AAR23927.1 ; AAO38056.1 ; AAQ75702.1 ; AAH1373
Taxonomy	*Source Organism:* Homo sapiens(human) *Taxon Group:* Euk/mammal NCBI Taxon: 9606 *Lineage:* cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterost Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Euther Hominidae; Homo/Pan/Gorilla group; Homo
Gene Name	PTGS2; COX2
Keywords	dioxygenase; glycoprotein; heme; iron; membrane; oxidoreductase; peroxidase; polymorphism; prostaglandin biosynth
Function	May have a role as a major mediator of inflammationand/or a role for prostanoid signaling in activity-dependentplastic
Subunit	Homodimer (By similarity).

Enzyme/Function	EC 1.14.99.1 EC-IUBMB, KEGG, BRENDA, WIT, MetaCyc *Nomenclature:* Oxidoreductases; Acting on paired donors, with incorporation or reduction of molecular oxygen; Miscellaneous; prostaglandin-endoperoxide synthase *Reaction:* arachidonate + AH_2 + 2 O_2 = prostaglandin H_2 + A + H_2O
Pathway	KEGG: Metabolism; Metabolism of Complex Lipids; Prostaglandin and leukotriene metabolism [PATH:hsa00590].
Complex/Interaction	BIND: Complex 45349 2 subunits: P35354 P35354 [PMID:10811226] related BIND: 45348 COX-2 forms a homodimer. This interaction wa modelled on the demonstrated homodimerization of mouse COX-2. BIND: Interaction 45348 with P35354 by three-dimensional-structure [PMID:10811226]
Structure	PDB: 1DDX:C(18-568,88.4%) ; 1DDX:D(18-568,88.4%) ; 1DDX:A(18-568,88.4%) ; 1DDX:B(18-568,88.4%) ; 1DD0 (18-568,88.2%) ; 1DD0:B(18-568,88.2%) ; 1DCX:A(18-568,88.2%) ; 1DCX:B(18-568,88.2%) ; 1CVU:A(18-568,87.8%) ; 1CVU:B(18-568,87.8%) ; 3PGH:A(18-604,87.1%) ; 3PGH:B(18-604,87.1%) ; 3PGH:C(18-604,87.1%) ; 3PGH:D(18-604,87.1%) ; 1CX2:A(18-604,87.1%) ; 1CX2:B(18-604,87.1%) ; 1CX2:C(18-604,87.1%) ; 1CX2:D(18-604,87.1%) ; 4COX:A(18-604,87.1%) ; 4COX:B(18-604,87.1%) ; More 1CVU: SCOP CATH FSSP MMDB PDBsum 1CX2: SCOP CATH FSSP MMDB PDBsum 1DCX: SCOP CATH FSSP MMDB PDBsum 1DD0: SCOP CATH FSSP MMDB PDBsum

CROSS-REFERENCES

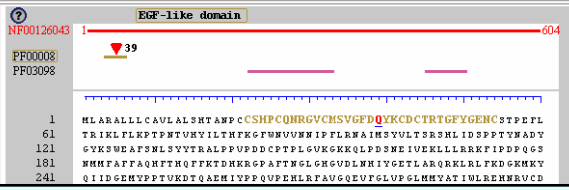
Bibliography	▸View Bibliography Information ▸Submit Bibliography PubMed: PMID: 1380156; 7945196; 8181472; 8473346; 11705852; 11712072; 11809691; 11809750; 11844819; 11847219; 11850206; 11867761; 11880271; 11891188; 11891209; 11897504; 11901151; 11906292; 11911965; 11913955; 11920472; 11920515; 11939728; 11948128; 11957147; 11959891; 11972392; 11981837; 11992399; 11994539; 12006564; 12016158; 12021045; 12032335; 12050157; 12050227; 12051953; 12067908; 12070598; 12072439; 12086404; 12124799; 12131767; 12143054; 12145315; 12162677; 12167656; 12174888; 12193665; 12195707; 12197220; 12205039; 12209724; 12209745; 12213900; 12218179; 12324457; 12349897; 12377741; 12378621; 12391274; 12391278; 12393863; 12393872; 12397176; 12401798; 12405290; 12444028; 12450219; 12452062; 12459168; 12462194; 12470703; 12477932; 12480928; 12485854; 12485915; 12488433; 12493747; 12498388; 12505789; 12519124; 12519748; 12532440; 12532441; 12533667; 12555215; 12578839; 12579269; 12591723; 12592372; 12598658; 12600292; 12626523; 12628757; 12636104; 12642362; 12654832; 12664576; More
DNA Sequence	GenBank: L15326; M90100; D28235; U04636; AY229989; AL033533; BC013734 EMBL: L15326; M90100; D28235; U04636; AY229989; AL033533; BC013734 DDBJ: L15326; M90100; D28235; U04636; AY229989; AL033533; BC013734
Genome/Gene	LocusLink: 5743 prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)(PTGS2) PTGS2: Genew; GeneCards; Ensembl [geneView; contigView]; GenAtlas Gene Index: human_THC1923125
Genetic Variation/Disease	*Gene Name:* PTGS2 HapMap OMIM: 600262

Gene Regulation	ERGDB: PTGS2: Up-regulated by estrogen (in vitro), in umbilical vein endothelial cells [PMID:12372818] ISG: PTGS2 Hormone. [PMID:9861020]
Ontology	*Molecular Function* GO:0016702: oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen [SPKW; evidence:IEA] GO:0004601: peroxidase activity [INTERPRO; evidence:IEA] [SPKW; evidence:IEA] [PMID:14511332; evidence:NAS] GO:0016491: oxidoreductase activity [SPKW; evidence:IEA] GO:0004666: prostaglandin-endoperoxide synthase activity [SPEC; evidence:IEA] [PMID:14511332; evidence:NAS] [PMID:9630216; evidence:TAS] *Biological Process* GO:0006954: inflammatory response [UniProt:P35354; evidence:none] GO:0006693: prostaglandin metabolism [PMID:9630216; evidence:TAS] GO:0006928: cell motility [PMID:9630216; evidence:TAS] GO:0050727: regulation of inflammatory response [PMID:14511332; evidence:NAS] GO:0019371: cyclooxygenase pathway [PMID:14511332; evidence:NAS] GO:0001516: prostaglandin biosynthesis [SPKW; evidence:IEA] GO:0007582: physiological process [PMID:9630216; evidence:TAS] *Cellular Component* GO:0016020: membrane [SPKW; evidence:IEA] GO:0005737: cytoplasm [PMID:10842192; evidence:TAS] [PMID:14511332; evidence:IDA]

FAMILY CLASSIFICATION

UniRef	UniRef100_P35354; UniRef90_P35354; UniRef50_P35354
PIRSF	PIRSF000342 human prostaglandin-endoperoxide synthase
Pfam Domain	Pfam: PF00008: EGF-like domain (22-54) Pfam: PF03098: Animal haem peroxidase (214-330, 450-507)
Prosite Motif	Prosite: PS00022: PDOC00021: EGF-like domain signature 1. Prosite: PS01186: PDOC00021: EGF-like domain signature 2. Prosite: PS50026: PDOC00021: EGF-like domain profile. Prosite: PS50292: PDOC00394: Animal heme peroxidase superfamily profile.
InterPro	*InterPro:* PGH2_HUMAN IPR002007: Animal haem peroxidase IPR006209: EGF-like domain IPR006210: Type I EGF IPR002016: Haem peroxidase, plant/fungal/bacterial
Other Classification	BLOCKS: IPB002007 Animal haem peroxidase signature BLOCKS: IPB006210 Type I EGF PRINTS: PR00457 ANPEROXIDASE SMART: SM00181 EGF

FEATURE & SEQUENCE DISPLAY



http://pir.georgetown.edu/cgi-bin/ipcEntry?id=P35354

PIR
Protein Information Resource

LIGCLRNCSAVTAAAKQLAE
VTGFSN----AKTTAQHVKK

Search | iProClass | for | Protein Name: protein kinase | AND | Organism Name: homo sapiens | AND | PIRSF ID: not null | AND | PDB ID: not null

Text Search

40 protein sequences in total.

For sequence analyses, pick a method (radiobutton) below, select a sequence(s) (checkbox) in *Protein ID* column, and GO.
BLAST | FASTA | HMM Search | Pattern Match | Multiple Alignment | Domain Display | GO

SAVE OPTIONS | SAVE ..
SAVE OPTIONS
Save All Retrievals
Save All Sequences
Save Selected Retrievals
Save Selected Sequences
MORE SAVE OPTION ...

Protein ID (check all)	Matched	NREF Entry Name	Length	Organism Name /Taxon Group	PIRSF ID /Family ID			PDB ID
NREF:NF00118722 iProClass:P36894 PIR-PSD:I37163 UniProt:BMRA_HUMAN	**NREF Entry Name** =>protein kinase **Organism Name** =>Homo sapiens	Bone morphogenetic protein receptor type IA precursor (EC 2.7.1.37) (Serine/threonine-protein kinase receptor R5) (SKR5) (Activin receptor-like kinase 3) (ALK-3)	532	Homo sapiens Euk/mammal	SF015229 FAM0012179	PF00069; PF01064	PCM00107; PCM00108; PCM00109	1ES7:B; 1ES7:D; 1ES7
NREF:NF00103479 iProClass:Q15078 PIR-PSD:S50861 UniProt:CD5R_HUMAN	NRE Prote							

Peptide Search

To perform a sequence analysis, make your selection below, check the box before the sequence ID, and then click on the Submit button. submit
BLAST; FASTA; Pattern Match; HMM Search; Multiple alignment; Domain Display;

ID	Protein Name	Organism	Taxon Group	PIRSF ID	length	Match Range
NREF:NF00298389 iProClass:O61308 PIR-PSD:T30934 SP/TR:PUMA_PARUN	227 kDa spindle- and centromere-associated protein (PUMA1)	Parascaris univalens	Euk/Animal		1955	1695- 1699: EKHSAELVISETASQ

BLAST/SSEARCH

For sequence analyses, pick a method (radiobutton) below, select a sequence(s) (checkbox) in *Protein ID* column
BLAST | FASTA | HMM Search | Pattern Match | Multiple Alignment | Domain Display | G

LEFLTELVISLEKLE

Protein ID (check all)	Protein Name	Organism	Taxon Group	PIRSF ID	e-value	Length	Ov.lap	% idn	Query Sequence
NREF:NF00828273 iProClass:Q96ME4 SP/TR:Q96ME4	Hypothetical protein FLJ32471	Homo sapiens(human)	Euk/mammal		e-124	246	246	100	
NREF:NF00114734 iProClass:Q13537 SP/TR:Q13537	Similar to pogo element	Homo sapiens(human)	Euk/mammal						
NREF:NF00829472 iProClass:Q96MW7 SP/TR:Q96MW7	Hypothetical protein FLJ31767	Homo sapiens(human)	Euk/mammal						
NREF:NF00085733 iProClass:S72481 PIR-PSD:S72481	probable transposase	Homo sapiens(human)	Euk/mammal	SF009339	2e-24	454	195	36	

```
Smith-Waterman score: 417;  36.923% identity in 195 aa overlap

                    10        20        30        40        50        60
NF0082 MSTKCPSVSPVSGEMKKRKAITLEMKLKIIAQHEGGKPVMAIARELGLWQSTISTILRDK
       :..::      :.: :.. ..::..::..:   : :     :...:::  .:.: ..  :
NF0008 MASKC------SSERKSHTSLTLNQKLEMIKLSEEGMSKAEIGQKLGLLCQTVSQVVNAK
                        10        20        30        40        50
```

Integrative Approach for Functional Genomics/Proteomics

- Homology Based
  - Sequence & Structural Families
- Functionally Linked
  - Genetic Association: Gene Clustering on Chromosomes, Multi-Domain Proteins
  - Function Association: Pathways, Biological Processes, Networks, Protein-Protein Interactions, Protein Complexes
  - Correlated Evolution: Related Phylogenetic Profile
  - Correlated Expression: mRNA/Protein Expression

Protein Family, Structure, Function

- Domain: Sharing Common Domain (Align with Pfam/SMART/CDD)
- Structure: Sharing Common Fold (Align with SCOP/CATH)
- Function: Functional Convergence/Divergence (Align with EC)

PIRSF ID	PIRSF Name	Avg Length	SCOP Superamily	Enzyme Classification	Pfam Domain
SF001490	cofactor-dependent PGM*	241	Phosphoglycerate mutase-like	EC 5.4.2.1; EC 5.4.2.4; EC 3.1.3.13; EC 3.1.3.-	PF00300
SF001492	cofactor-independent PGM	518	Alkaline phosphatase-like; 2,3-BPG*-independent PGM*, substrate-binding domain	EC 5.4.2.1	PF01676
SF006392	cofactor-independent PGM, archaeal type	411		EC 5.4.2.1	PF01676
SF001491	phosphopentomutase	403		EC 5.4.2.7	PF01676
SF000891	alkaline phosphatase	507	Alkaline phosphatase-like	EC 3.1.3.1	PF00245
SF000971	sulfatase	580	Alkaline phosphatase-like	EC 3.1.6.8; EC 3.1.6.1; EC 3.1.6.2; EC 3.1.6.12	PF00884

Genome Context and Phylogenetic Profile for Functional Inference of Conserved Hypothetical Proteins

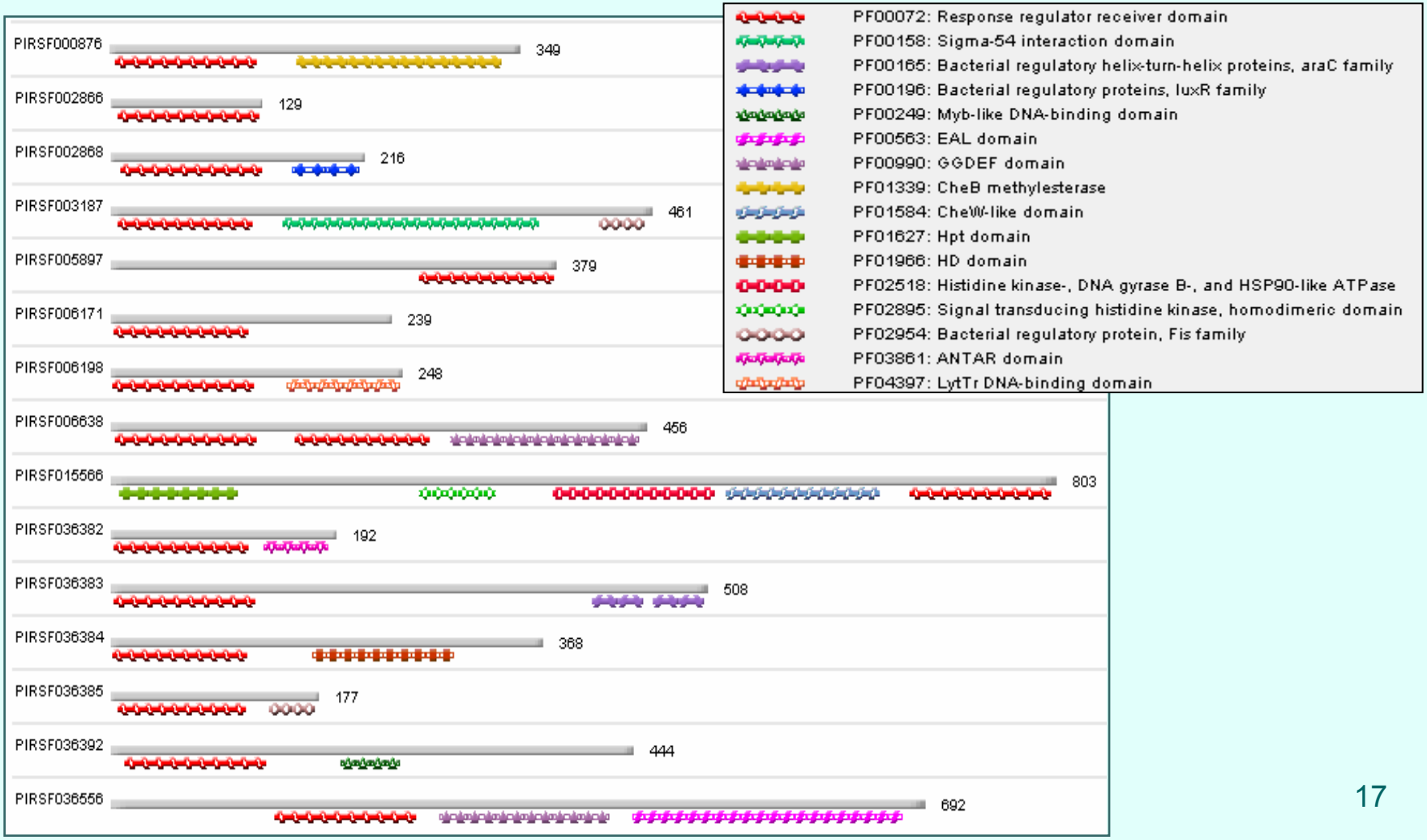
PIRSF ID and Name	Mt* Gene	Mk* Gene	Mj* Gene
PIRSF005019: [NiFe]-hydrogenase-3-type complex Eha, membrane protein EhaA	MTH384	MK0477	MJ0528
PIRSF019706: [NiFe]-hydrogenase-3-type complex Eha, membrane protein EhaB	MTH385	MK0476	MJ0527
PIRSF036534: [NiFe]-hydrogenase-3-type complex Eha, membrane protein EhaC	MTH386	MK0475	MJ0526.1
PIRSF006581: [NiFe]-hydrogenase-3-type complex Eha, membrane protein EhaD	MTH387	MK0474	MJ0526
PIRSF036535: [NiFe]-hydrogenase-3-type complex Eha, membrane protein EhaE	MTH388	MK0473	MJ0525
PIRSF019373: [NiFe]-hydrogenase-3-type complex Eha, membrane protein EhaF	MTH389	MK0472	MJ0524
PIRSF019136: [NiFe]-hydrogenase-3-type complex Eha, membrane protein EhaG	MTH390	MK0471	MJ0523
PIRSF036536: [NiFe]-hydrogenase-3-type complex Eha, membrane protein EhaH	MTH391	MK0470	MJ0522
PIRSF036537: [NiFe]-hydrogenase			
PIRSF000215: NADH:quinone membrane subunit C/D/J PIRSF500037: [NiFe]-hydroger			
PIRSF036538: [NiFe]-hydrogenase			
PIRSF004953: [NiFe]-hydrogenase			
PIRSF005292: [NiFe]-hydrogenase			
PIRSF002913: NADH:quinone complex, small subunit PIRSF500034: [NiFe]-hydroger			
PIRSF000230: NADH:quinone complex, large subunit PIRSF500033: [NiFe]-hydroger			

Phylogenetic Pattern (in complete genomes) | Taxonomic Distribution

Collapsed View | Expanded View Restart

Taxonomy Nodes	Pattern | Counts																							
	SF005019		SF019706		SF036534		SF006581		SF036535		SF019373		SF019136		SF036536		SF036537		SF036538		SF004953		SF005292	
Archaea	+	3	+	3	+	3	+	3	+	3	+	3	+	3	+	3	+	1	+	1	+	3	+	3
Crenarchaeota	-		-		-		-		-		-		-		-		-		-		-		-	
Euryarchaeota	+	3	+	3	+	3	+	3	+	3	+	3	+	3	+	3	+	1	+	1	+	3	+	3
Archaeoglobi	-		-		-		-		-		-		-		-		-		-		-		-	
Halobacteria	-		-		-		-		-		-		-		-		-		-		-		-	
Methanobacteria	+	1	+	1	+	1	+	1	+	1	+	1	+	1	+	1	+	1	+	1	+	1	+	1
Methanococci	+	1	+	1	+	1	+	1	+	1	+	1	+	1	+	1	-		-		+	1	+	1
Methanomicrobia	-		-		-		-		-		-		-		-		-		-		-		-	
Methanopyri	+	1	+	1	+	1	+	1	+	1	+	1	+	1	+	1	-		-		+	1	+	1
Thermococci	-		-		-		-		-		-		-		-		-		-		-		-	
Thermoplasmata	-		-		-		-		-		-		-		-		-		-		-		-	
Nanoarchaeota	-		-		-		-		-		-		-		-		-		-		-		-	
Bacteria	-		-		-		-		-		-		-		-		-		-		-		-	
Eukaryota	-		-		-		-		-		-		-		-		-		-		-		-	
other sequences	-		-		-		-		-		-		-		-		-		-		-		-	
Viroids	-		-		-		-		-		-		-		-		-		-		-		-	
Viruses	-		-		-		-		-		-		-		-		-		-		-		-	
Total Mapped		3		3		3		3		3		3		3		3		1		1		3		3

Domain Architecture for Functional Annotation of Response Regulators

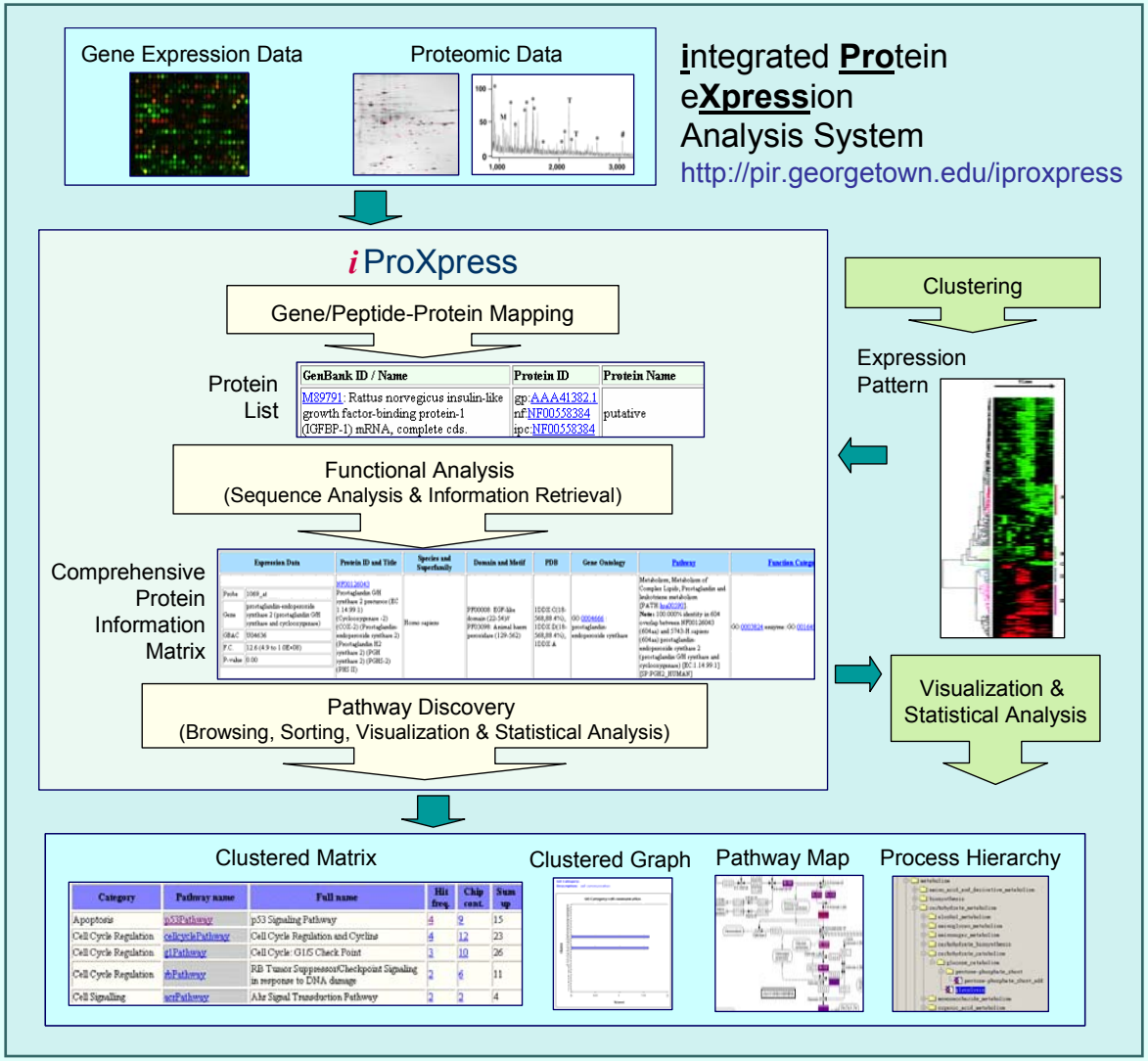


Protein Functional Analysis

Functional inference of uncharacterized hypothetical proteins
TIM-barrel signal transduction protein
metal-dependent hydrolase
nucleotidyltransferase
ATPase with chaperone activity and inactive LON protease domain
alpha/beta hydrolase
lipid carrier protein
[Ni,Fe]-hydrogenase-3-type complex, membrane protein EhaA
Correction, or improvement, of genome annotations
ligand-binding protein with an ACT domain
inactive homolog of metal-dependent protease
glycyl radical cofactor protein YfiD
chemotaxis response regulator methylesterase CheB
thioesterase, type II
bifunctional tetrapyrrole methylase and MazG NTPase
Enhanced understanding of structure, function, and evolutionary relationships
chorismate mutase, AroH class
chorismate mutase, AroQ class, prokaryotic type

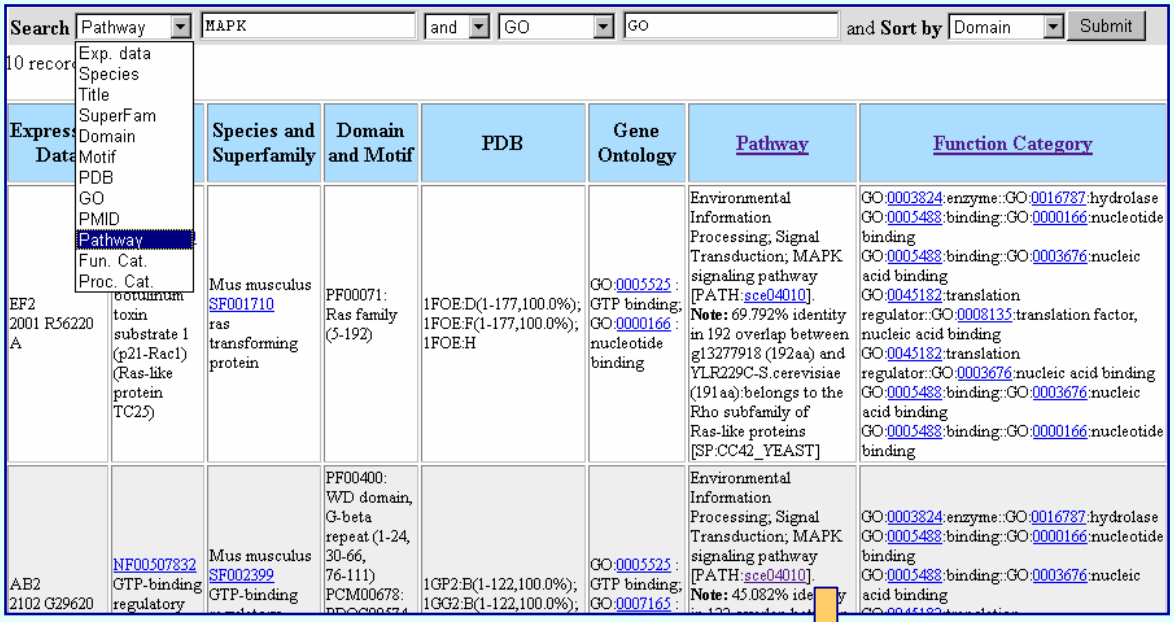
- Based on PIRSF Family Classification and iProClass Data Integration

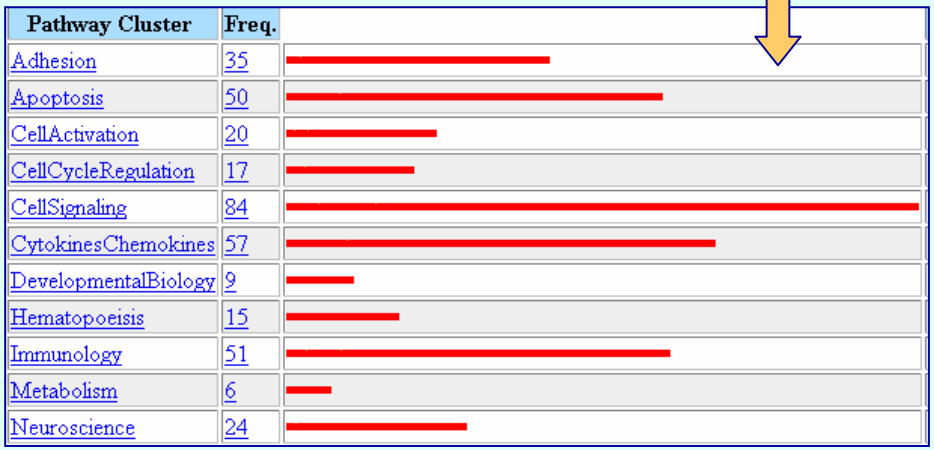
iProXpress Expression Data Analysis



- Global Bioinformatics Analysis of Genes and Proteins
- Gene/Peptide to Protein Mapping
- Sequence Analysis, Data Mining
- Display of Protein Information Matrix
- Categorization and Visualization for Pathway Discovery

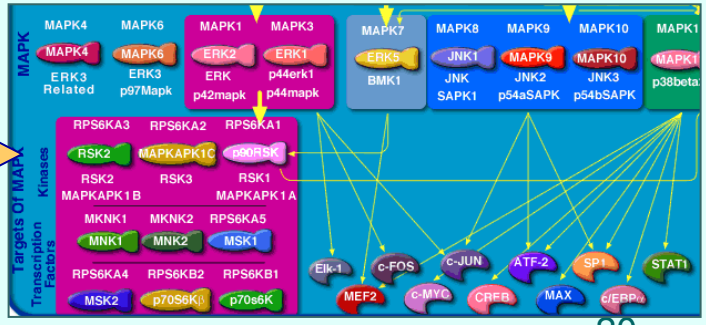
Expression Analysis





- Display of Protein Information Matrix
- Iterative Searching, Sorting, Categorization, Visualization
- Iterative Hypothesis Generation/Evaluation

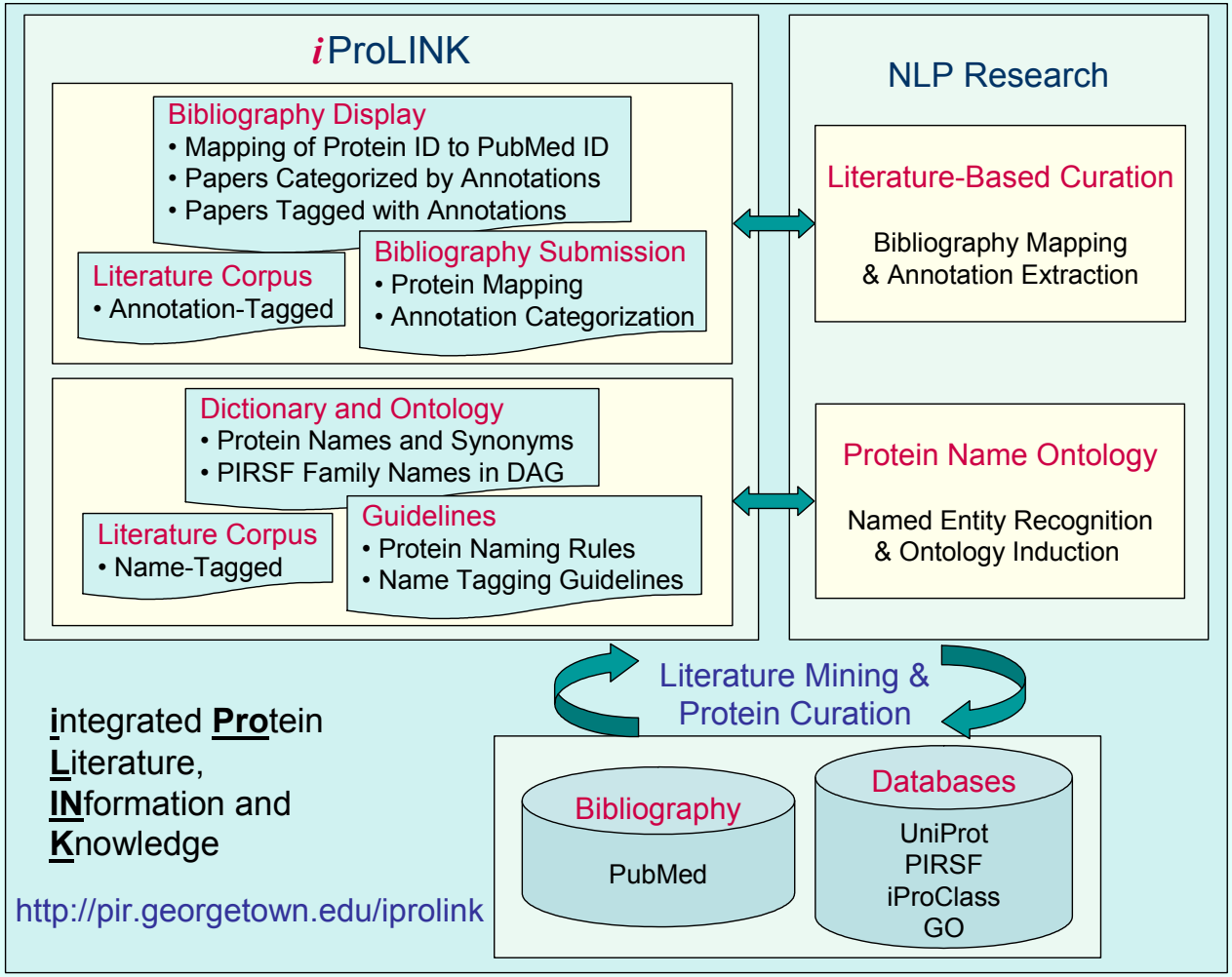
MAPKinase Signaling Pathway



iProXpress Dissemination Mechanisms

- Web Interface for Expression Data Analysis (http://pir.georgetown.edu/iproxpress)
  - Batch data submission (gene or protein lists, peptide sequences)
  - Gene/peptide to protein mapping
  - Generation of protein information matrix (both direct information retrieval from iProClass and homology-based functional inference)
  - Iterative searching, sorting, categorization and visualization
  - Data/results export in excel or tab-delimited table formats
  - Data import/export in formats interoperable with commonly used programs (e.g., Netaffx, Mascot)
- Web Services
  - Map gene list to proteins (based on gene IDs, probe IDs, gene names)
  - Map peptides to proteins (based on peptide sequences)
- Local Web Server (iProXpress Mirror Site)
  - Distribution files for local implementation: iProClass data files, data mining programs, graphical user interface

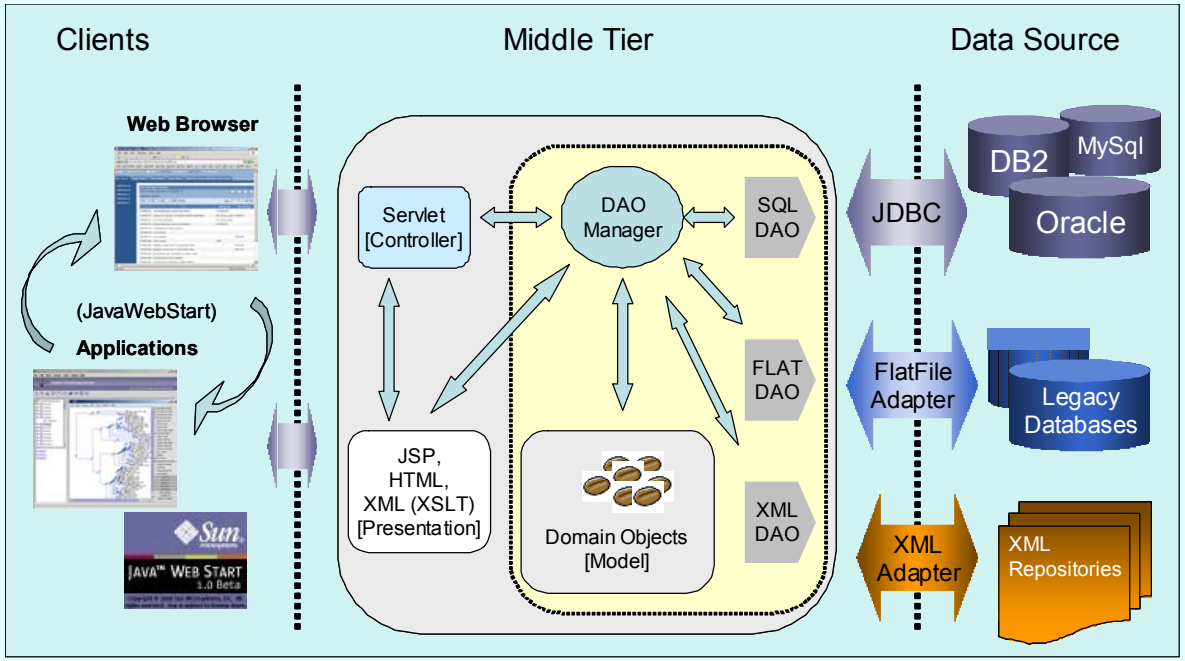
iProLINK Literature Mining / Ontology



- Literature-Based Curation and Evidence Attribution
- Protein Name Ontology for Standard Annotation and Query Expansion
- http://pir.georgetown.edu/iprolink

Bioinformatics Infrastructure

- Software Framework: J2EE *n*-Tier Architecture (Apache Struts, JSP, Servlets, JDBC, RMI, XML) with Object Models (Rational Rose)
- Database Distribution: XML/DTD, Flat File, FASTA, Relational Tables (Oracle 9i, MySQL, DB2) and Schema
- Planned Deliverables: Object Models, J2EE Application Programs, Web Services



PIR for caBIG ICR Workspace

- Downloadable Databases
  - UniProt and UniRef Databases
  - iProClass and PIRSF Databases
- Web Interfaces
  - PIR (iProClass/PIRSF): Iterative Text and Sequence Search
  - iProXpress: Expression Data Analysis
  - iProLINK: Name Dictionary, Annotated Bibliography and Corpus
- Web Services/Application Programs
  - Gene/Peptide to Protein Mapping
  - iProXpress Distribution Files for Local Web Mirror
  - PIRSF Viewers: Taxonomy, Iterative BlastClust, Tree/Alignment
- Involved in caBIO and CDE (Common Data Elements)
  - Protein Object Modeling and Controlled Vocabularies

Acknowledgments

- Funding
  - NIH: NHGRI/NIGMS/NLM/NIMH/NCRR/NIDCR (UniProt); NIAID (Proteomic Adm Resource)
  - NSF: BDI (iProClass); ITR (Ontology)
- Computing Resources
  - Sun Microsystems AEG grant (V880)
  - IBM SUR grant (P690)
- PIR Team

Dr. Cathy Wu
Dr. Hongzhan Huang
Dr. Darran Natale
Dr. CR Vinayaka
Dr. Leslie Arminski
Jing Zhang, M.S.
Vincent Hormoso
Dr. Winona Barker
Dr. Lai-Su Yeh
Dr. Zhangzhi Hu
Dr. Xianying Wei
Dr. Hsing-Kuo Hua
Sehee Chung, M.S.
Natalia Petrova
Dr. Robert Ledley
Dr. Anastasia Nikolskaya
Dr. Raja Mazumder
Baris Suzek, M.S.
Yongxing Chen, M.S.
Amar Kalelkar
Jorge Castro-Alvear