

# *PIR Integrated Resources and Data-Mining Tools for Functional Genomics and Proteomics*

---

**CODATA 2002**

**Sept. 29 - Oct. 3, 2002, Montreal, Canada**

**Zhang-Zhi Hu, M.D.**

**Bioinformatics Scientist,**

**Protein Information Resource**

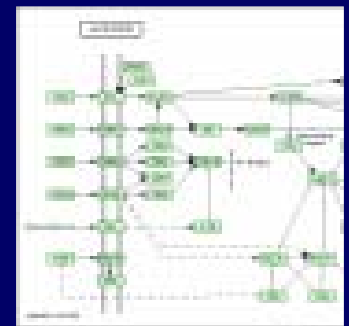
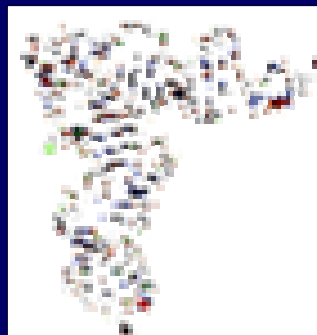
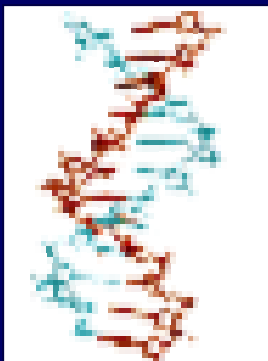
**Georgetown University Medical Center, Washington, DC**

# *Functional Genomics and Proteomics*

Study of Biological Systems Based on Global Knowledge of Genomes, Transcriptomes, Proteomes, Metabolomes

- **Genome:** All the Genetic Material in the Chromosomes
- **Transcriptome:** Entire Set of Gene Transcripts
- **Proteome:** Entire Set of Proteins
- **Metabolome:** Entire Set of Metabolites

**Genome** → **Transcriptome** → **Proteome** → **Metabolome**



# *Protein Information Resource (PIR)*

- **Goal: An Integrated Public Resource of Protein Informatics to Support Genomic/Proteomic Research & Scientific Discovery**
- **Components**
  - **Database: Data Organization & Information Retrieval**
  - **Software: Data Analysis & Sequence Annotation**
- **Challenges**
  - **Voluminous, Complex, Dynamic Data from Heterogeneous Sources**
- **Integrated, Classification Approach**
  - **Databases: PIR-PSD, PIR-NREF, iProClass**
  - **Integrated Analysis System: Knowledge Base System**
  - **Database Interoperability: Ontology, XML, Relational Schema, iProClass Framework**

# PIR Web Site (<http://pir.georgetown.edu>)



## PIR Protein Information Resource

[About PIR](#)[Databases](#)[Search and Retrieval](#)[Download](#)[Support](#)

AN INTEGRATED PUBLIC RESOURCE OF PROTEIN INFORMATICS TO SUPPORT  
GENOMIC AND PROTEOMIC RESEARCH AND SCIENTIFIC DISCOVERY

PIR produces the **Protein Sequence Database (PSD)** of functionally annotated protein sequences, which grew out of the *Atlas of Protein Sequence and Structure* (1965-1978) edited by Margaret Dayhoff and has been incorporated into an integrated knowledge base system of value-added databases and analytical tools.

**ProClass**, a central point for exploration of protein information, provides summary descriptions of protein family, function and structure for PIR-PSD, Swiss-Prot, and TrEMBL sequences, with links to over 45 biological databases. [Release 2.5, 9-Sep-2002 contains 826,766 entries](#)

**PIR-NREF**, a comprehensive database for sequence searching and protein identification, contains non-redundant protein sequences from PIR-PSD, Swiss-Prot, TrEMBL, RefSeq, GenPept, and PDB. [Release 1.05, 9-Sep-2002 contains 1011453 entries.](#)



### PIR News Flash

Press Release: Protein Information Resource Adds  
New Tools to Databases

Text Search Protein Databases:

GO!

Find an Exact Peptide Match:

GO!

Type in a string of single letter amino acid  
code (at least 3 letters)



© Protein Information Resource, [National Biomedical Research Foundation](#), [Georgetown University Medical Center](#)  
3900 Reservoir Rd., NW, Box 571414, Washington, DC 20057-1414, USA  
Phone: (202) 687-2121 Fax: (202) 687-1662  
[pirmail@nbrf.georgetown.edu](mailto:pirmail@nbrf.georgetown.edu)

[Site Map](#) [Site Search](#)





# *Protein Family Classification*

**Discovery of New Knowledge by  
Using Information Embedded within  
Families of Homologous Sequences and Their Structures**

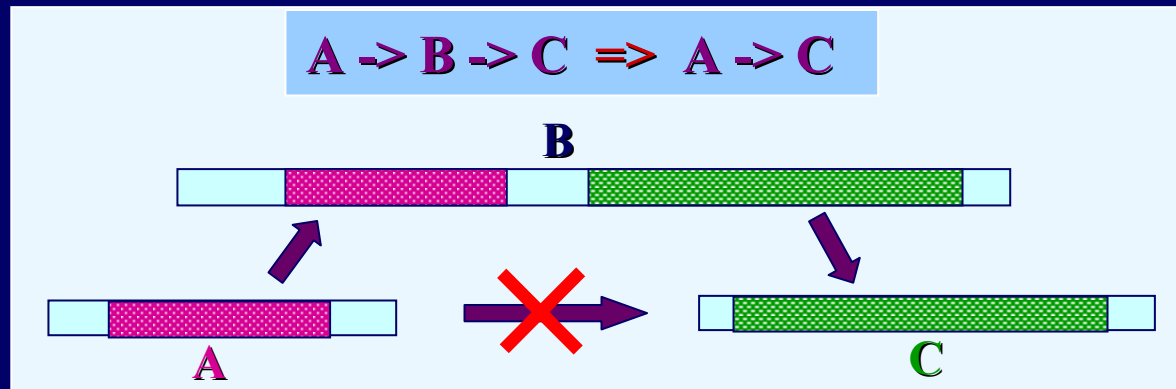
- **Superfamily, Domain, and Motif Classification**
- **Superfamily Concept**
  - **End-to-End Similarity & Same Overall Domain Architecture**
- **Significance**
  - **Improve Sensitivity of Protein Identification**
  - **Provide Complete Clustering for Database Organization**
  - **Detect and Correct Genome Annotation Errors Systematically**
  - **Drive Other Annotations**
  - **Stimulate Evolution, Genomics and Proteomics Research**

# Genome Sequence Annotation

	<input type="checkbox"/> H70468	SF001258	051440	<a href="#">phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) / phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) [similarity]</a>	<i>Aquifex aeolicus</i>	Prok/other	594.3	4.8e-26	205	39.086	197	
	<input type="checkbox"/> S76963	SF001258	039935	<a href="#">phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) / phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) [similarity]</a>	<i>Synechocystis</i> sp.	Prok/gram-	557.0	5.7e-24	230	39.175	194	
	<input type="checkbox"/> T35073	SF029243	005738	<a href="#">probable phosphoribosyl-AMP cyclohydrolase</a>	<i>Streptomyces coelicolor</i>	Prok/gram+	399.3	3.5e-15	128	42.157	102	
	<input type="checkbox"/> S53349	SF001257	001188	<a href="#">phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) / phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) / histidinol dehydrogenase (EC 1.1.1.23)</a>	<i>Saccharomyces cerevisiae</i>	Euk/fungi	384.1	2.5e-14	799	31.863	204	
	<input type="checkbox"/> E69493	SF029243	005738	<a href="#">phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) [similarity]</a>	<i>Archaeoglobus fulgidus</i>	Archae	396.8	4.8e-15	108	47.778	90	
	<input type="checkbox"/> G64337	SF006833	030827	<a href="#">phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) [similarity]</a>	<i>Methanococcus jannaschii</i>	Archae	246.9	1.1e-06	95	36.842	95	
	<input type="checkbox"/> D81178	SF006833	101491	<a href="#">phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) NMB0603 [similarity]</a>	<i>Neisseria meningitidis</i>	Prok/gram-	239.9	2.6e-06	107	35.227	88	
	<input type="checkbox"/> G81925	SF006833	101491	<a href="#">phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) NMA0807 [similarity]</a>	<i>Neisseria meningitidis</i>	Prok/gram-	245.2	1.4e-06	107	35.227	88	
	<input type="checkbox"/> S51513	SF001257	001188	<a href="#">phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) / phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) / histidinol dehydrogenase (EC 1.1.1.23)</a>	<i>Pichia pastoris</i>	Euk/fungi	225.6	1.7e-05	842	29.670	182	

	ID	Assigned		Correction	Superfamily
	E69493	3.5.4.19/3.6.1.31		3.5.4.19	SF029243
	H70468	3.6.1.31		3.5.4.19/3.6.1.31	SF001258
	G64337	3.5.4.19		3.6.1.31	SF006833

# Genome Era Challenges: Transitive Catastrophe



- **Error Propagation: At least 17 Sequences Incorrectly Named as IMP Dehydrogenase or Related (Propagated to KEGG & WIT)**

**A rational protein family classification, based on both global and local similarity, can prevent or correct many of such errors.**

# ***PIR-NREF Database***

- **Non-Redundant REFerence Protein Sequence Database**
  - **Comprehensiveness:** PIR-PSD, Swiss-Prot, TrEMBL, RefSeq, GenPept, PDB
  - **Timeliness:** Biweekly Updates (~ 1,000,000 Sequences)
  - **Non-Redundancy:** by Sequence Identity & Taxonomy (Species)
  - **Source Attribution:** Protein IDs and Names from Underlying Databases, Sequence, Taxonomy, Bibliography
  - **Related Sequences:** Identical Sequences from Different Species, Complete Substring,  $\geq 95\%$  Sequence Identity
- **Applications**
  - **Protein Identification:** Full-Scale or Species-Based Sequence Analysis and Text Search
  - **Detection of Annotation Errors**
  - **Development of Protein Name Ontology**
- **FTP Distribution:** XML and FASTA Formats



# PIR-NREF Report (I)

● NREF Entry: NF00104599

[iProClass View](#)



Last Updated: 25-Feb-2002

**Protein Name** DNA-directed RNA polymerase II 14.4 kDa polypeptide (EC 2.7.7.6) (RPB6) (RPB14.4)

**Taxonomy**  
**Homo sapiens** (human)  
 NCBI Taxon ID: [9606](#)  
 Lineage: cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Primates; Catarrhini; Hominidae; Homo/Pan/Gorilla group; Homo

**Source Organism** Homo sapiens (Taxon ID: [9606](#))

**Bibliography**  
 ▶ [View Bibliography information](#) ▶ [Submit Bibliography](#)  
 PubMed: PMID:[8786150](#); PMID:[7803819](#); PMID:[10784442](#); PMID:[8196653](#); PMID:[10591208](#)

Database	Protein ID	Accession	Taxon ID	Protein Name
PIR	<a href="#">S38627</a>	<a href="#">I38175</a> ; <a href="#">S38627</a>	<a href="#">9606</a>	DNA-directed RNA polymerase (EC 2.7.7.6) II 14.4K chain ALT_NAMES:DNA-directed RNA polymerase II chain F
SwissProt	<a href="#">RPB6_HUMAN</a>	<a href="#">P41584</a>	<a href="#">9606</a>	DNA-directed RNA polymerase II 14.4 kDa polypeptide (EC 2.7.7.6) (RPB6) (RPB14.4)
GenPept	<a href="#">g13097771</a>	<a href="#">AAH03582.1</a>	<a href="#">9606</a>	polymerase (RNA) II (DNA directed) polypeptide F
GenPept	<a href="#">g18254493</a>	<a href="#">CAB62981.1</a>	<a href="#">9606</a>	dJ1039K5.5 (polymerase (RNA) II (DNA directed) polypeptide F)
GenPept	<a href="#">g415388</a>	<a href="#">CAA81629.1</a>	<a href="#">9606</a>	RNA Polymerase II subunit 14.4 kD
RefSeq	<a href="#">g11527390</a>	<a href="#">NP_068809</a>	<a href="#">9606</a>	DNA directed RNA polymerase II polypeptide F; DNA directed RNA polymerase II 14.4 kda polypeptide
PDB	<a href="#">1OKL:A</a>	None	<a href="#">9606</a>	DNA-Directed RNA Polymerase II 14.4 Kd Polyp

**Protein Sequence**  
 MSDMEDNFDGDDFDDVEDEGLDDLENAEEEGQENVEILPSGERPQANQKRITTPYMTKY  
 ERARVLGTRALQIAMCAPVMVELEGETDPLLIAMKELKARKIPIIIRRYLPDGSYEDWGV  
 DELIITD

**Sequence Length** 127

**Related Sequence**  
 → **Identical NREF Sequence:**  
[NF00494607](#) [Cricetinae]; [NF00494911](#) [Cricetulus griseus]  
 → **Neighbor NREF sequence:**

Seq. ID	Organism	Seq. Length	% identity	Overlap	Matched region	
<a href="#">NF00104599</a>	Homo sapiens	127	100.000	127	1-127:1-127	
<a href="#">NF00494607</a>	Cricetinae	127	100.000	127	1-127:1-127	
<a href="#">NF00494911</a>	Cricetulus griseus	127	100.000	127	1-127:1-127	
<a href="#">NF00532685</a>	Mus musculus	127	99.213	127	1-127:1-127	

# PIR-NREF Report (II)

## Annotation Discrepancy of Multi-Domain Proteins

Schizosaccharomyces pombe (Taxon ID: [4896](#))

Database	Protein ID	Accession	Taxon ID	Protein Name
PIR	<a href="#">T40073</a>	<a href="#">T40073</a>	<a href="#">4896</a>	phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) / phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) [similarity]
TrEMBL	<a href="#">O59667</a>	<a href="#">O59667</a>	<a href="#">4896</a>	PROBABLE PHOSPHORIBOSYL-AMP CYCLOHYDROLASE (EC 3.5.4.19) / PHOSPHORIBOSYL-ATP PYROPHOSPHOHYDROLASE (EC 3.6.1.31) / HISTIDINOL DEHYDROGENASE (EC 1.1.1.23) (HDH)
GenPept	<a href="#">g3006138</a>	<a href="#">CAA18379.1</a>	<a href="#">4896</a>	probable phosphoribosyl-amp cyclohydrolase

MALLPFFDLTNFESDASEELGWLKYVGRVQTRVFPQHFKDNLEKVRKISETIDVIVDTTA  
 ELGPEACVLLNAGALAILVNEEMLNELADISPNRLVVKTDTTDIGKIEKLSQVAGSIQW  
 IGS AENYPPDFFERASKIIHKAVMPEGGGRTLYLEFPEQPSMEVLKFSVHVSVPVLSSS  
 FLTVKPAEEPKLSLADLILISANTDREDGLFSTLVVNELGIALGLVYSSKESVAESLKT  
 GTGVYQSRKRGLWYKGASSGAVQHLIHIDVDCDEDCLRFVYVYQTGKGFCHLDTLHCFGQA  
 SGLCQLEKTLIDRKNNAPESY TARLFSDPKLLRAKIMEEAEELCDATTKENVIWEMADL  
 MYFAITRCVGSVSLNDISRHLDLKHKVTRRKGDAKVAWQEKLKDKGGVANTS YTA



PF00815: Histidinol dehydrogenase (EC 1.1.1.23)



PF01502: Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19)



PF01503: Phosphoribosyl-ATP pyrophosphohydrolase (EC 3.6.1.13)

S51513



842

T40073



417

# PIR-NREF Database

(<http://pir.georgetown.edu/pirwww/search/pirnref.shtml>)

## Database Description

The PIR-NREF is a Non-redundant REFerence protein database designed to provide a timely and comprehensive collection of all protein sequence data, keeping pace with the genome sequencing projects and containing source attribution and minimal redundancy. The database contains all sequences in PIR-PSD, SwissProt, TrEMBL, RefSeq, GenPept, and PDB. Identical sequences from the same source organism (species) reported in different databases are presented as a single NREF entry with protein IDs and names from each underlying database, in addition to protein sequence, taxonomy, and composite bibliography. Related sequences identified by all-against-all FASTA search are listed for each NREF entry. The web site provides direct entry retrieval (based on protein IDs), text search (protein or species names), and sequence search (BLAST, peptide match, and pattern match) for full-scale and species-based protein identification. Species-based browsing and searching are supported for about 100 organisms, including over 70 complete genomes. PIR-NREF is available for free downloading and redistribution from our FTP site in XML format (data file) and FASTA format (sequence file). The database is updated biweekly and the release 1.03, 12-Aug-2002, contains 975702 entries from:

Database	Release#	Date	# of Entries
PIR	73.02	02-Aug-2002	283,224
SwissProt	40.25	02-Aug-2002	112,657
TrEMBL	21.6	02-Aug-2002	669,330
GenPept	130.0	15-Jun-2002	1,084,824
RefSeq		8-Aug-2002	375,001
PDB		5-Aug-2002	18,871

[Release History](#)

• More Description ↗

PIR-NREF current release 1.03, 12-Aug-2002 contains 975702 entries. [Download](#)  
Related NREF sequences identified by all-against-all FASTA search were pre-computed  
by [DuPont Bioinformatics Team](#).

## • Find Proteins by Name, Organism or UID

Retrieve a matching list (a summary report if only one entry found) by protein name and organism name using substring match, or by UID using exact match.

All UID Fields

## • Protein Sequence Similarity Search

▶ [Search NREF by Species/Organism](#)

▶ [BLAST Search](#)

Retrieve a matching list of entries by searching your query protein sequence against the NREF database.

Paste query sequence (single-letter amino acid code) or ">" followed by unique sequence identifier from any underlying database:

E-value:  Filter

▶ [Peptide Match](#)

Retrieve a list of entries with exact matches to your query peptide sequence.

Enter a string of single-letter amino acid codes below:

▶ [Pattern Match](#)

Retrieve a list of entries matching your query pattern or a ProSite pattern.

Insert a user-defined pattern below: [Click here for help on how to write a protein pattern](#)

Or, alternatively, type in a valid PROSITE code for a query pattern (e.g., PS00888):

# PIR Searches (I)

## Text Search

Search  for  AND  NOT  AND

37 entries found. Displaying page 1 ( up to 50 entries per page).

[New Search](#)

To perform a sequence analysis, make your selection below, check the box before the sequence ID, and then click on the Submit button.

BLAST;  FASTA;  Pattern Match;  HMM Search;  Multiple alignment;  Domain Display;

Protein ID	Matched	Protein Name	Length	Organism Name /Taxon Group	Superfamily# /Family#	Pfam ID	PC Motif ID	PDB ID
<input type="checkbox"/> NREF: <a href="#">NF00261325</a> iProClass: <a href="#">NF00261325</a> PIR-PSD: <a href="#">S45368</a>	Protein Name =>protein kinase Organism Name =>Zea mays	protein kinase C inhibitor	129	<a href="#">Zea mays</a> <a href="#">Euk/Plant</a>	<a href="#">SF000713</a> <a href="#">FAM0016285</a>		PCM00892;	3RHN; 4RHN; 5RHN; 6RHN; 1KPB:A;
<input type="checkbox"/> NREF: <a href="#">NF00261270</a> iProClass: <a href="#">NF00261270</a> PIR-PSD: <a href="#">T03024</a> SP/TR: <a href="#">Q41793</a>	Protein Name =>protein kinase Organism Name =>Zea mays	calcium-dependent protein kinase (EC 2.7.1.-), calmodulin-independent	465	<a href="#">Zea mays</a> <a href="#">Euk/Plant</a>	<a href="#">SF000592</a> <a href="#">FAM0007168</a>	PF00036; PF00069;	PCM00018; PCM00107; PCM00108;	1A06; 1PHK; 2PHK:A; 1QL6:A; 1KOA;
<input type="checkbox"/> NREF: <a href="#">NF00261218</a> iProClass: <a href="#">NF00261218</a> SP/TR: <a href="#">Q9AUC2</a>	Protein Name =>protein kinase Organism Name =>Zea mays	Receptor-like protein kinase 1	750	<a href="#">Zea mays</a> <a href="#">Euk/Plant</a>		PF00069; PF00560;		

Your input peptide ELVIS

matches 251 sequences in the NREF

Displaying page 1 (up to 50 entries per page) [Next page](#)

## Peptide Search

Sort by   Search

To perform a sequence analysis, make your selection below, check the box before the sequence ID, and then click on the Submit button.

BLAST;  FASTA;  Pattern Match;  HMM Search;  Multiple alignment;  Domain Display;

ID	Protein Name	Organism	Taxon Group	Superfamily	Family	length	Match Range
<input type="checkbox"/> NREF: <a href="#">NF00099273</a> iProClass: <a href="#">NF00099273</a> PIR-PSD: <a href="#">A55723</a> SP/TR: <a href="#">D3D2_HUMAN</a>	3,2-trans-enoyl-CoA isomerase, mitochondrial precursor (EC 5.3.3.8) (Dodecenoyl-CoA delta-isomerase)	<a href="#">Homo sapiens(human)</a>	<a href="#">Euk/Animal</a>	<a href="#">SF001331</a>	<a href="#">FAM0028565</a>	302	77-81: LEFLTELVISLEKLE
<input type="checkbox"/> NREF: <a href="#">NF00451499</a> iProClass: <a href="#">NF00451499</a> PIR-PSD: <a href="#">A47025</a> SP/TR: <a href="#">ABIC_LACLA</a>	Abortive phage resistance protein abiC	<a href="#">Lactococcus lactis</a>	<a href="#">Bact/Gram+</a>	<a href="#">SF019612</a>	<a href="#">FAM0019831</a>	344	191-195: KEQKVELVISRQFDD

# PIR Searches (II)

## BLAST 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;

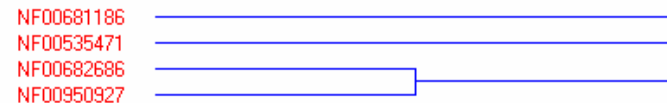
Query sequence: (T47769, length=170, Search NREF, e-value < 0.0001, filter=T)

>NF00682686 hypothetical protein F24I3.170 [Arabidopsis thaliana]  
MDAKIGQFFDVGTFSSGSDKIPWCDGDIAGCEREVREATD SGTEDLKKECLMRLSWAL  
VHSRQTEDVQRGIAMLEASLESSAPPLEDKREKLYLLAVGYRSGNYRSRQLVDRCIEMQ  
ADWRQALVLLKTTIEDKITKDGVIIGITATAFGAVGLIAGGIVAAMSRRK

9 match(es) shown in the following table: The color bar at the top of the table represents the BLAST alignment. %idn (%identity) and Ov lap (overlap length) are generated by SSEARCH.

Protein ID	Protein Name	
<input checked="" type="checkbox"/> NREF: <a href="#">NF00682686</a> iProClass: <a href="#">NF00682686</a> PIR-PSD: <a href="#">T47769</a> SP/TR: <a href="#">Q9M1J1</a>	hypothetical protein F24I3.170	<a href="#">Arabidopsis thaliana</a>
<input checked="" type="checkbox"/> NREF: <a href="#">NF00681186</a> iProClass: <a href="#">NF00681186</a> SP/TR: <a href="#">Q94CK3</a>	Hypothetical 17.9 kDa protein	<a href="#">Arabidopsis thaliana</a>
<input checked="" type="checkbox"/> NREF: <a href="#">NF00950927</a> iProClass: <a href="#">NF00950927</a> SP/TR: <a href="#">Q8RZ98</a>	B1147A04.28 protein	<a href="#">Oryza sativa</a>
<input checked="" type="checkbox"/> NREF: <a href="#">NF00535471</a> iProClass: <a href="#">NF00535471</a> SP/TR: <a href="#">Q9CQ92</a>	2010003O14Rik protein (RIKEN cDNA 2010003O14 gene)	<a href="#">Mus musculus</a>

### TREE VIEW:



### MULTIPLE ALIGNMENT:

```

NF00682686      -----MDAKIGQFFDVGTF
NF00950927      MRKAPERIEYRASQKSIATWPEILLHRLRAKPSRFLPHRSRRSAAMEAKIGRLVGAIGAF
NF00681186      -----MDAAIGKVFDSVDF
NF00535471      -----MEAVLMELVS-----
                                     *: * : : : : :
NF00682686      FSGS-----DKIPWCDGDIAGCEREVREATD SGTEDLKKECLMRLSWALVHSRQTEDVQ
NF00950927      FSGG-----DNVFWCGRDIAGVEREVAEA---ATEEHKNVSI MRLSWALVHSRNTDDVM
NF00681186      FSGAASASADEFPLCDSDIISGCEKELAE---QDEGRKKECIMRLSWALVHSKMPSDIQ
NF00535471      -----VEDLKNFERKQSE---QAAGSVSKSTQFEYAWCLVRSKYNEDIR
                                     : : . * : : . : : : * : * : * : * : * :
NF00682686      RGIAMLEASLESSAPPLEDKREKLYLLAVGYRSGNYRSRQLVDRCIEMQADWRQALVLK
NF00950927      RGIAMLQASLGGSKSPLEAREKLYLLAVGHYR--N-----VATCIQIQPGWGQALSILK
NF00681186      RGIAMLEALVWVNDTSAMKLRKLYLLALGYRSGDFSRSRDCIERCLEVEPESGQAQALK
NF00535471      RGIIVLLEELLPKG--SKEEQRDYVFFYLAVGNRYRLKEYEKALKYVVRGLLQTEPQMNQAKELE
                                     ***.:* : : . : * : : * : * * : : : : . * * * :
NF00682686      KTIEDKITKDGVIIGITATAFGAVGLIAGGIVAAMSRRK-
NF00950927      KTVEDKIAKDGVIIGIATT---AVGLLVG-IAAAVARKN-
NF00681186      KAIEDRIVKDGVIIGIATV---AVGVVAG-IAAAILRS--
NF00535471      RLIDKAMKDGVLGMAIVGGMALGVAGLAGLIGLAVSKSKS
                                     : : . : * : * : * : * . * : * * * : :
    
```

Multiple Alignment  
& Tree View

# *Data Integration*

## ● **Challenges**

- **Voluminous, Complex & Dynamic Data from Heterogeneous Sources in Distributed Networking Environment**

## ● **Data Warehouse**

- **Local Copy of Databases in a Unified Database Schema**
- **Allows Local Control of Data; Update Problem**

## ● **Hypertext Navigation**

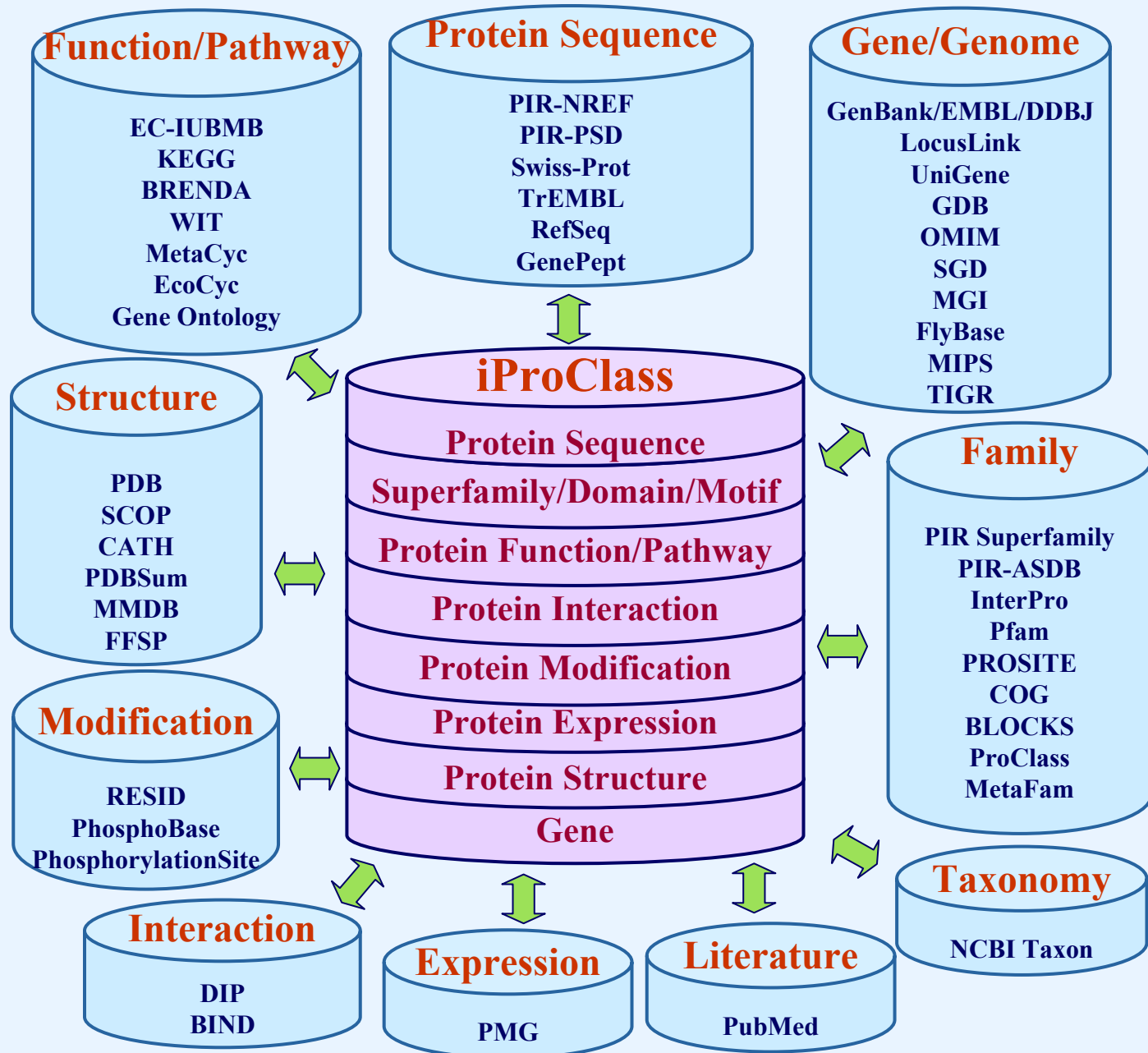
- **Browsing Model with Hypertext Links**
- **Allows Direct Interaction; Easily Lost in Cyberspace**

## ● **iProClass Approach**

- **Data Warehouse + Hypertext Navigation**
- **Rich Links (Links + Executive Summaries) between Database Objects**
- **An Integrated Platform for Describing Comprehensive Family Relationships and Structural and Functional Features of Proteins**

# *iProClass Database*

- **An Integrated Platform for Describing Comprehensive Family Relationships and Structural and Functional Features of Proteins**
- **Classification Scheme: Superfamily/Family & Domain/Motif**
  - **Superfamily/Family (Global): Full-Length Similarity with Same Domain Arrangement**
  - **Domain/Motif (Local): Structural/Functional Units & Sites**
- **Sequence and Family Data**
  - **Non-Redundant, Annotated PIR-PSD, Swiss-Prot, TrEMBL Sequences: ~827,000**
  - **Superfamilies (~36,000), Families (>145,000), Domains (>3700), Motifs (>1300), Post-Translational Modifications (>280)**
  - **Superfamily and Protein Summary Reports**
- **Modular Framework: Extensibility, Flexibility, Customization**



# *iProClass* Overview

*Cross-References:*  
Links to >50  
Databases



# iProClass - Sequence Report (I)





Summary Report for iProClass Entry: A26916+RIR2\_YEAST

[Related Sequences](#) 

## GENERAL INFORMATION

Protein Name and ID	PIR-NREF: <a href="#">NF00366352</a>	
	Database	ID
	PIR-PSD	<a href="#">A26916</a>
	SwissProt	<a href="#">RIR2_YEAST</a>
	RefSeq	<a href="#">NP_012508</a>
GenPept: <a href="#">CAA89317.1</a> ; <a href="#">AAA34988.1</a>		
Taxonomy	<i>Source Organism:</i> Saccharomyces cerevisiae(baker's yeast) NCBI Taxon: <a href="#">4932</a> <i>Lineage:</i> cellular organisms; Eukaryota; Fungi/Metazoa group; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces	
Gene Name	MIPS:YJL026w; SGD:RNR2	
Keywords	iron; oxidoreductase; pyrimidine deoxynucleotide metabolism	
Function	<i>Description:</i> oxidoreductase; pyrimidine deoxynucleotide metabolism	

## CROSS-REFERENCES

Bibliography		<a href="#">View Bibliography Information</a> <a href="#">Submit Bibliography</a> PubMed: PMID: <a href="#">3316984</a> ; <a href="#">3313004</a> ; <a href="#">8849441</a> ; <a href="#">8641269</a>
DNA Sequence		GenBank: <a href="#">M17221</a> ; <a href="#">M17789</a> ; <a href="#">Z49301</a> EMBL: <a href="#">M17221</a> ; <a href="#">M17789</a> ; <a href="#">Z49301</a> DDBJ: <a href="#">M17221</a> ; <a href="#">M17789</a> ; <a href="#">Z49301</a>
Genome/Gene		SGD: <a href="#">S0003563</a>
Ontology		GO: <a href="#">0016491</a> : oxidoreductase GO: <a href="#">0004748</a> : ribonucleoside-diphosphate reductase
Enzyme/Function		EC 1.17.4.1 <a href="#">EC-IUBMB</a> , <a href="#">KEGG</a> , <a href="#">BRENDA</a> , <a href="#">WIT</a> , <a href="#">MetaCyc</a> <i>Nomenclature:</i> Oxidoreductases; Acting on CH <sub>2</sub> groups; With a disulfide as acceptor; ribonucleoside-diphosphate reductase <i>Reaction:</i> 2'-deoxyribonucleoside diphosphate + oxidized thioredoxin + H <sub>2</sub> O = ribonucleoside diphosphate + reduced thioredoxin
Pathway		KEGG: Metabolism; Nucleotide Metabolism; Purine metabolism [PATH: <a href="#">sce00230</a> ]. KEGG: Metabolism; Nucleotide Metabolism; Pyrimidine metabolism [PATH: <a href="#">sce00240</a> ].
Complex/Interaction		BIND: <a href="#">11158</a> ; <a href="#">11211</a> ; <a href="#">11330</a> ; <a href="#">11348</a> ; <a href="#">11410</a> ; <a href="#">11447</a> ; <a href="#">3505</a> ; <a href="#">7604</a> ; <a href="#">7713</a> ; <a href="#">7874</a> ; <a href="#">8660</a> ; <a href="#">9200</a> ; <a href="#">9219</a> ; <a href="#">9484</a> ; <a href="#">9643</a> ; <a href="#">9998</a>

# Bibliography Information Display

- From Curated Databases (e.g., PIR-NREF, SGD)
- From User Submission
- From Computer-Mapping (e.g. Gene Symbol)

We have the following bibliography information for this protein: PIR ID:[A26916](#) NREF ID:[NF00366352](#)

**protein name** - Ribonucleoside-diphosphate reductase small chain 1 (EC 1.17.4.1) (Ribonucleotide reductase) - baker's yeast (*Saccharomyces cerevisiae*)

Search PubMed for

You may change the protein name and species for different searches (protein name - speceis).

## Curated Bibliography Information

authors Hurd HK, Roberts CW, Roberts JW,  
journal Mol Cell Biol. 1987 Oct; 7:3673-7  
title Identification of the gene for the yeast  
by methyl methanesulfonate. [PMID: 33169](#)  
information [DNA/RNA Sequence Features|Function/Pro](#)  
contained [SGD: S0003563](#);

authors Elledge SJ, Davis RW,  
journal Mol Cell Biol. 1987 Aug; 7:2783-93  
title Identification and isolation of the gene  
Saccharomyces cerevisiae: DNA damage-i  
information [DNA/RNA Sequence Features|Mapping|Muta](#)  
contained [Features|Protein Sequence Features|RNA](#)  
[SGD: S0003563](#);

- [Submission Form](#)  
Citation information  
Bibliography categories  
Contact information
- [View NREF Report](#)
- [View iProClass Report](#)
- [View PIR-PSD Report](#)
- [View Bibliography](#)  
for entry [NF00135498](#)

Welcome to PIR Bibliography Submission page for protein entry: PIR ID:[A29635](#), NREF ID:[NF00135498](#)  
**protein name** - transcription factor Sp1 - human(*Homo sapiens*)

Enter the bibliography information using the following form (information with \* is required).  
[Instructions for submission](#) [Sample Submission](#) [View Bibliography Information](#)

### Submission Form

**Citation information.**  from NCEI for PMID or MUID  if you know  
the citation's PubMed (PMID, e.g. 9407091) or Medline Unique Identifier (MUID, e.g. 98070508).

Authors\*:

Citation Source\*:

(e.g. Journal name)

Citation Type:  Year\*:  Volume:  Pages:

Title or Description\*:

Cross-Reference:  (e.g., PMID or MUID)

**Bibliography categories.** Please check below if the article you entered is applicable to any of the following categories:  
(choose one or more categories. Pred.= information is predicted, Exp.= information is experimentally determined)

Category*	Description/Method (please do NOT exceed 400 words)	N/A	Pred.	Exp.
<a href="#">1. Protein name</a>	<input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<a href="#">2. Organism</a>	<input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<a href="#">3. Genetic information</a>	<input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<a href="#">4. Tissue/cellular localization</a>	<input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<a href="#">5. Structure</a>	<input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<a href="#">6. Features</a>	<input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- ①anti-toxin +
- ①anticoagulant
- ①antifreeze +
- ①antioxidant +
- ①antisense RNA
- ①apoptosis regulator +
- ①cell adhesion molecule +
- ①cell cycle regulator +
- ①chaperone +
- ①cytoskeletal regulator
- ①defense/immunity protein +
- ①enzyme +
- ①enzyme activator +
- ①enzyme inhibitor +
- ①ligand binding or carrier +
- ①lysin +
- ①microtubule binding +
- ①molecular\_function\_unknown +
- ①motor +
- ①nucleic acid binding +
- ①obsolete +
- ①oxidative phosphorylation uncoupler +
- ①protein stabilization +
- ①protein tagging +
- ①ribosomal RNA
- ①signal transducer +
- ①small nuclear RNA
- ①small nucleolar RNA
- ①storage protein +
- ①structural protein +
- ①toxin +
- ①transfer RNA
- ①transporter +

# Functional Classification

## ● Gene Ontology (GO)

- Three Ontologies: Biological Process, Molecular Function, Cellular Component
- Consortium: FlyBase, SGD, MGI, TAIR, WormBase, Pombase

GO term:	<b>apoptosis activator</b>
GO id:	<b>GO:0016506</b>
Definition:	<b>A function held by products which directly activate any step in the process of apoptosis.</b>
Number of paths to term:	<b>1</b>

- ① denotes an 'is-a' relationship
- Ⓜ denotes a 'part-of' relationship

### Gene\_Ontology

#### Ⓜmolecular\_function

##### ①apoptosis regulator

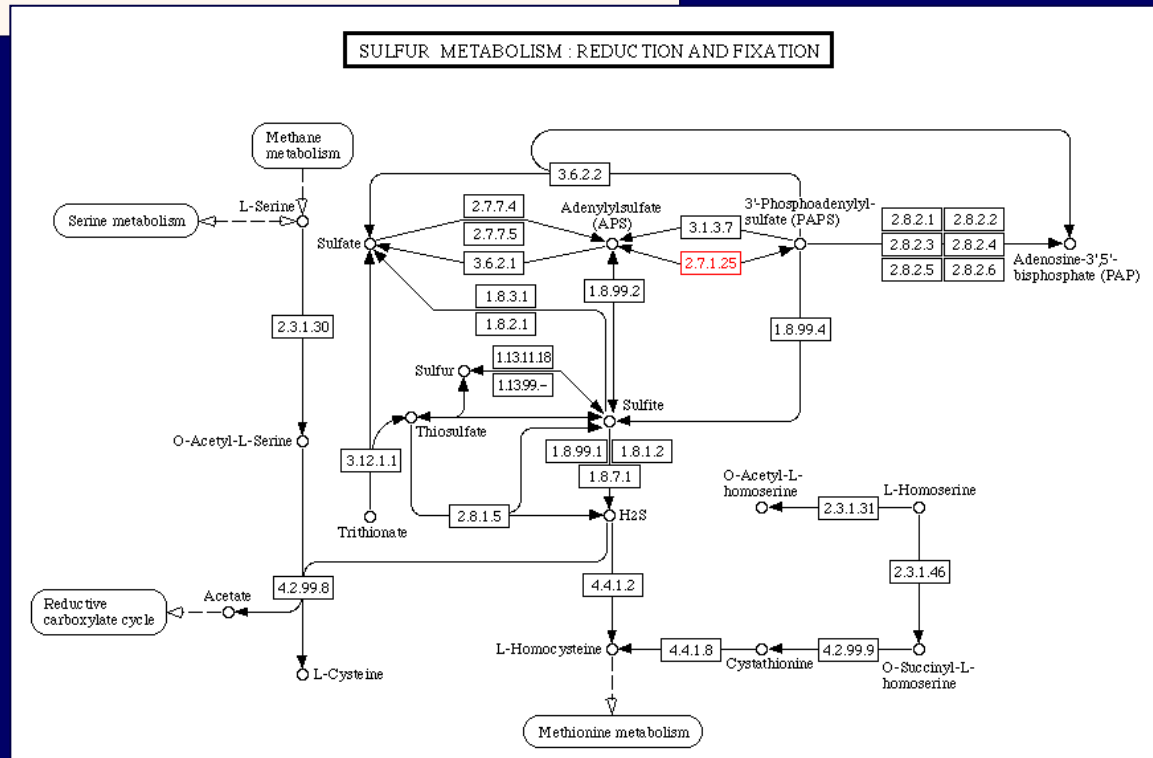
##### ①apoptosis activator [GO:0016506]

##### ①apoptotic protease activator

##### ①apoptosis inhibitor

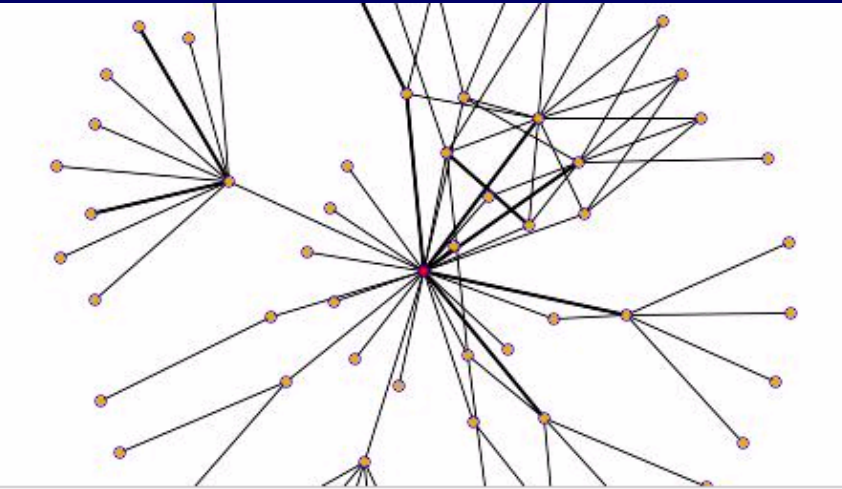
# KEGG Metabolic & Regulatory Pathways

ENTRY EC [2.7.1.25](#)  
 NAME Adenylylsulfate kinase  
 CLASS Transferases  
 Transferring phosphorus-containing groups  
 Phosphotransferases with an alcohol group as acceptor  
 SYSNAME ATP:adenylylsulfate 3'-phosphotransferase  
 REACTION ATP + Adenylylsulfate = ADP + 3'-Phosphoadenylylsulfate  
 SUBSTRATE [ATP](#)  
[Adenylylsulfate](#)  
 PRODUCT [ADP](#)  
[3'-Phosphoadenylylsulfate](#)  
 PATHWAY PATH: [MAPO0230](#) Purine metabolism  
 PATH: [MAPO0450](#) Selenoamino acid metabolism  
 PATH: [MAPO0920](#) Sulfur metabolism  
 GENES ECO: [b2750](#) (cysC)



# DIP Protein-Protein Interactions

Protein Information Data	
PIR	<a href="#">S55211</a>
SWISSPROT	<a href="#">YJZ2_YEAST</a>
Genpept (GI)	<a href="#">gi:1077873</a>
LENGTH	128
ORGANISM	Saccharomyces cerevisiae
DESCRIPTION	hypothetical protein YJR022w
FUNCTION	
EC	
LOCALIZATION	




## Interacting partners

This page contains every protein interacting with your protein of interest

- Retrieves information about a particular interaction
- Retrieves 1 level of interaction links.







### CENTERED ON

Links	Interaction information	PIR	SWISSPROT	GENBANK	NAME
	-	S55211	YJZ2_YEAST	1077873	hypothetical protein YJR022w

### INTERACTS WITH 52 PROTEINS

		S58329	HS49_YEAST	2131290	HSH49 protein
		S45760	SNP3_YEAST	626847	probable snRNP-related protein YBL026w

# iProClass - Sequence Report (II)

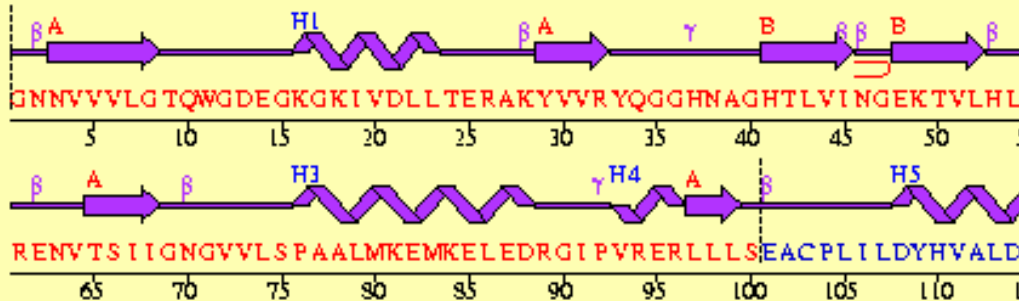
Structure	<p>PDB: <a href="#">1JK0:A</a>(1-399,100.0%)  ; <a href="#">1XSM</a>(74-399,68.7%)  ; <a href="#">1JK0:B</a>(53-399,54.5%) </p> <p>1JK0: <a href="#">SCOP</a> <a href="#">CATH</a> <a href="#">FSSP</a> <a href="#">MMDB</a> <a href="#">PDBsum</a></p> <p>1XSM: <a href="#">SCOP</a> <a href="#">CATH</a> <a href="#">FSSP</a> <a href="#">MMDB</a> <a href="#">PDBsum</a></p>
PIR Feature & Post Translational Modifications	<p>FEAT1; RESID: <a href="#">AA0019</a> (L-tyrosine)active site: Tyr (stable tyrosyl radical) (183) [predicted]</p> <p>FEAT2; binding site: 2Fe-O cluster (Asp, Glu, His, Glu, Glu, His) (145,176,179,239,273,276) [predicted]</p>
<b>FAMILY CLASSIFICATION</b>	
PIR FASTA Similarity	PIR-ASDB: <a href="#">A26916</a>
PIR Superfamily	iProClass: <a href="#">SF000355</a> herpesvirus ribonucleoside-diphosphate reductase small chain
PIR Family	PIR-MIPS: FAM0000390
PIR Motif	iProClass: <a href="#">PCM00368</a> : <a href="#">PDOC00317</a> : Ribonucleotide reductase small subunit signature (PST:175-191)
Other Classification	<p>BLOCKS: <a href="#">IPB000358</a> Ribonucleotide reductase</p> <p>Pfam: <a href="#">PF00268</a>: Ribonucleotide reductase, small chain (76-358)</p> <p>MetaFam: <a href="#">A26916</a></p> <p>InterPro: <a href="#">RIR2</a> <a href="#">YEAST</a></p>
<b>FEATURE &amp; SEQUENCE DISPLAY</b>	
<p><b>?</b> <b>PDOC00317, Ribonucleotide reductase small subunit signature</b></p> <p><b>NF00366352</b> 1  399</p> <p><b>PCM00368</b>  <b>183</b></p> <p><b>PF00268</b> </p> <pre> 1      MPKETPSK&amp;&amp;ADALSDLEIKDSKSNLNKELETLREENRUKSDMLKEKLSKDAENHKAYLK 61     SHQVHRHKLKEMEKEEPLLNEDKERTVLPFIKYHEIWQAYKRAEASFWTAAEIDLSDKIH 121    DWNRRMNENERFFISRVLAPFAASDGIUNENLVENPSTEVQIPEAKSFGYGFQIM<b>IENIHS</b> 181    <b>ETYSLIDITYI</b>KDPKSESEFLFNAIHTIPEIGEKAEWALRWIQDADALFGERLVAFASIEG 241    VFFSGSPASIFWLKKRGHMPGLTFSNELICRDEGLHTDFACLLFAHLKMKPDP&amp;IVEKIV 301    TEAVEIEQRYFLDALPVALLGMNADLMNQVEFVADRLLVAFGNKKYYKVENPFDPMENI 361    SLAGKTNPF EKRVSDYQKAGVMSKSTKQEAG&amp;FTFMEDF </pre>	

# Protein Structural Classification

## ● CATH Classification

Domain	Links	CATH no.	Class	Architecture
1	<a href="#">CATH</a> <a href="#">DHS</a>	<b>3.40.440.10</b> ->	Alpha Beta	3-Layer(aba) Sandwich
2	<a href="#">CATH</a> <a href="#">DHS</a>	<b>1.10.300.10</b> ->	Mainly Alpha	Orthogonal Bundle
3	<a href="#">CATH</a> <a href="#">DHS</a>	<b>3.90.170.10</b> ->	Alpha Beta	Complex

Domains:- 1 2 3



### 3.40.440.10

- ⓐ Alpha Beta
- ⓐ 3-Layer(aba) Sandwich
- ⓐ Adenylosuccinate Synthetase, subunit A, domain 1
- ⓐ LIGASE (SYNTHETASE)

Domain	CATH code	Length	Image
<a href="#">1adeA2</a>	<a href="#">1.10.300.10</a>	100	
<a href="#">1adeA1</a>	<a href="#">3.40.440.10</a>	165	
<a href="#">1adeA3</a>	<a href="#">3.90.170.10</a>	166	

# PIR-RESID

## Post-Translational Modification Database

[Annotation](#) or [Image](#) or [Model](#) ([instructions for viewing model](#))

PIRRES:AA0059

N-myristoyl-glycine

Alternate names: N-(1-oxotetradecyl)glycine

Systematic name: (tetradecanoylamino)ethanoic acid

Cross-references: CAS:14246-55-0

Formula: C 16 H 30 N 1 O 2 +

Formula weight: #chem 268.42 + #phys 268.2277 +

Correction formula: C 14 H 26 O 1 +

Correction weight: #chem 210.36 + #phys 210.1984

Date: 31-Mar-1995 #structure\_revision 31-Mar-1995 #t

Ozols, J.; Carr, S.A.; Strittmatter, P.

J. Biol. Chem. 259, 13349-13354, 1984

Title: Identification of the NH<sub>2</sub>-terminal blocked NADH-cytochrome b5 reductase as myristic acid acid sequence of the membrane-binding domain.

Reference number: [A22182](#); [MUID:85030460](#)

Note: GC and FAB mass spectrographic characteriz

Griffith, J.P.; Kim, J.L.; Kim, E.E.; Sintchak, M.D.; Thomson, J.A.; Fitzgibbon, M.J.; Fleming, M.A.; Caron, P.R.; Hsiao, K.; Navia, M.A. submitted to the Brookhaven Protein Data Bank, August 1996

Reference number: [A66708](#); [PDB:1TCO](#)

Note: X-ray crystallography, 2.5 angstroms

Griffith, J.P.; Kim, J.L.; Kim, E.E.; Sintchak, M.D.; Thomson, J.A.; Fitzgibbon, M.J.; Fleming, M.A.; Caron, P.R.; Hsiao, K.; Navia, M.A. Cell 82, 507-522, 1995

Title: X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex.

Reference number: [A56967](#); [MUID:95360994](#)

Note: X-ray crystallography, 2.5 angstroms

Comment: The myristyl group represents a mixture of saturated and unsaturated fatty acids.

Generating enzyme: glycylopeptide N-tetradecanoyltransferase (EC 2.3.1.97)

Sequence code: G

Conditions: amino-terminal

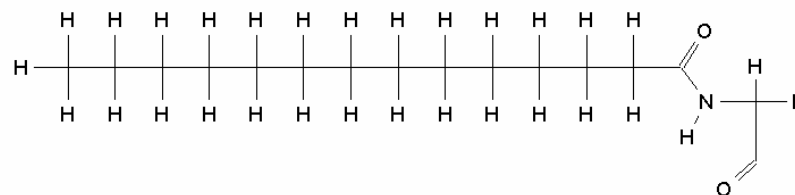
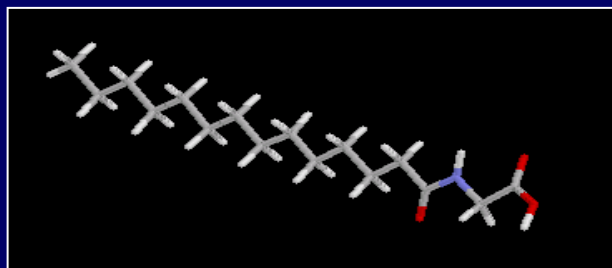
Keywords: [blocked amino end](#); [lipoprotein](#); [myristylation](#)

Residues

Feature

[Modified site: myristylated amino end \(Gly\)](#)

[Modified site: myristylated amino end \(Gly\) \(in mature form\)](#)





# iProClass - Superfamily Report

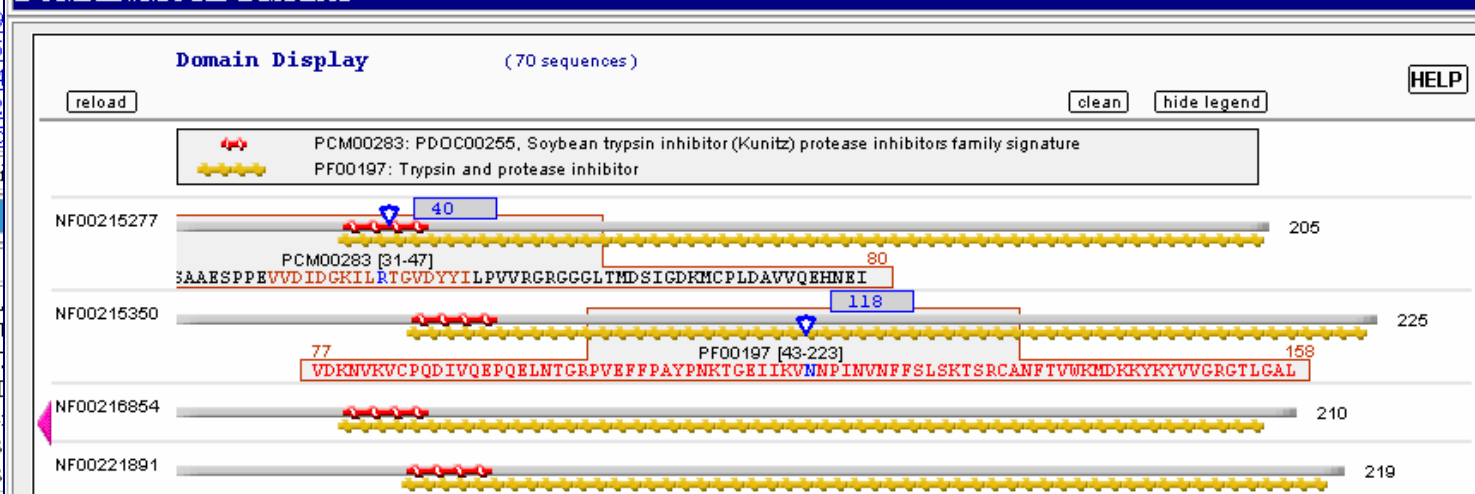
## GENERAL INFORMATION

Superfamily Number	SF001645
Superfamily Name	plant Kunitz-type proteinase inhibitor
Superfamily Size	Total Families=18; Total Sequence Entries=76 (63 Proteins+13 Fragments)
Taxonomy Range	Eukaryotae=76; Prokaryotae=0; Archaea=0; Viruses=0; Other=0
Length Range	Minimum=38; Maximum=225; Average=193; Standard Deviation=35
Keyword	serine proteinase inhibitor(29); seed(10); proteinase inhibitor(5); alpha-amylase inhibitor(3); glycoprotein(3); pyroglutamic acid(3); storage protein(2); chloroplast(1); homodimer(1); monomer(1); sweet taste(1); oxidoreductase(1); polymorphism(1)

## MEMBERSHIP

Eukaryotic Member	<a href="#">iProC</a> <a href="#">JQ09</a> <a href="#">JC55</a> <a href="#">A594</a> <a href="#">S042</a> <a href="#">A455</a>
Model Organism	Arabidopsis

## DOMAIN/MOTIF DISPLAY



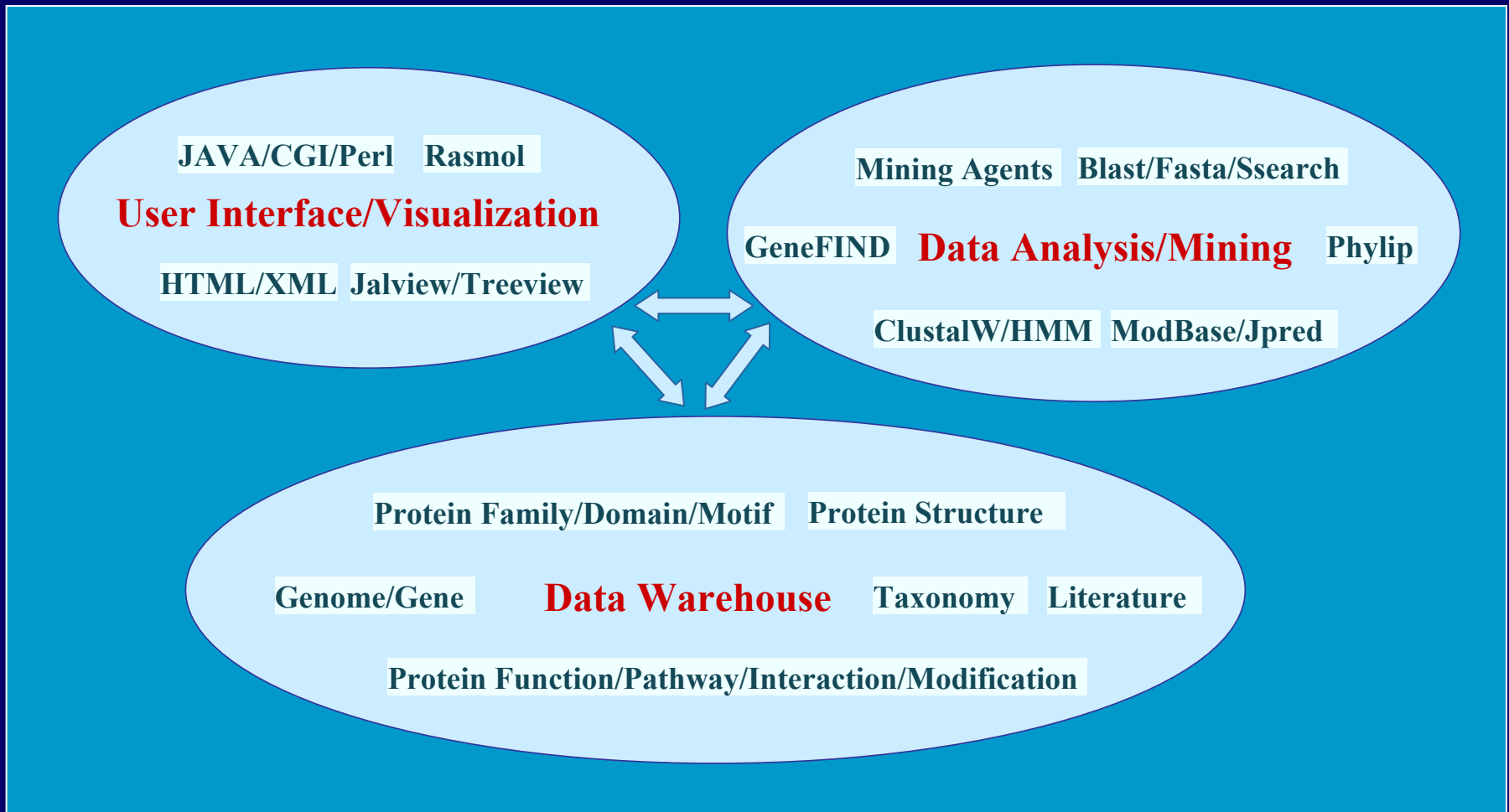
## FUNCTION AND STRUCTURE

Enzyme	EC 1
Structure	1AV 1AV 1BA 1EY 1TE 1WB 4WB

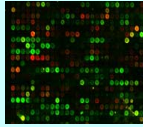
## FAMILY RELATIONSHIP

PIR Family	FAM0001666(16); FAM0001665(10); FAM0014361(7); FAM0025108(7); FAM0025748(5); FAM0025777(5); FAM0025122(4); FAM0025520(4); FAM0025178(3); FAM0025296(3); FAM0025570(2); FAM0073144(2); FAM0010462(1); FAM0810537(1); FAM0051174(1); FAM0026005(1); FAM0025367(1); FAM0019701(1)
PIR MOTIF	iProClass: <a href="#">PCM00283</a> : Soybean trypsin inhibitor (Kunitz) protease inhibitors family signature(59)
PIR Alignment	PIR-ALN: <a href="#">FA0909</a> : plant Kunitz-type proteinase inhibitor PIR-ALN: <a href="#">FA1318</a> : plant Kunitz-type proteinase inhibitor
Pfam Domain	PFAM: <a href="#">PF00197</a> : Trypsin and protease inhibitor(51)
Other Classification	MetaFam: <a href="#">SF001645</a>

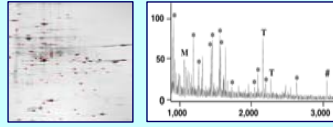
# *Integrated Protein Knowledge Base System*



## Gene Expression Data



## Proteomic Data



## Integrated Protein Knowledgebase

### Gene/Peptide-Protein Mapping

### Protein List

GeneLink ID / Name	Protein ID	Protein Name
M2291 Ratiaa neurogenin 2-like growth factor binding protein 1 (XPRP) mRNA, complete cds	IP:AAA63821	putative
M2291 Ratiaa neurogenin 2-like growth factor binding protein 1 (XPRP) mRNA, complete cds	IP:DBJ03338	
M2291 Ratiaa neurogenin 2-like growth factor binding protein 1 (XPRP) mRNA, complete cds	IP:DBJ03338	
M2291 Ratiaa neurogenin 2-like growth factor binding protein 1 (XPRP) mRNA, complete cds	IP:AAA21611	Tyrosine protein kinase
M2291 Ratiaa neurogenin 2-like growth factor binding protein 1 (XPRP) mRNA, complete cds	IP:DBJ03338	SHC (BC 2.1.11.2)
M2291 Ratiaa neurogenin 2-like growth factor binding protein 1 (XPRP) mRNA, complete cds	IP:DBJ03338	
M2291 Ratiaa neurogenin 2-like growth factor binding protein 1 (XPRP) mRNA, complete cds	IP:AAA68611	Artemin 1 (G-protein-coupled)
M2291 Ratiaa neurogenin 2-like growth factor binding protein 1 (XPRP) mRNA, complete cds	IP:DBJ03338	

### Functional Analysis (Sequence Analysis & Data Mining)

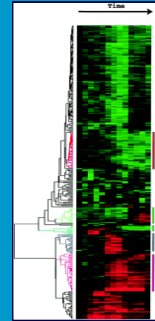
### Comprehensive Protein Information Matrix

GI (link to NCBI)	Length	%ID	Q.I. Length (link to alignment)	Matched sequence	Length	Definition (link to UniProt)	Class (link to pathway diagram)
1022724	314	90.852	117	4792-H spigone	317	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha (NF- $\kappa$ B, I $\kappa$ B, NFKB1)	Cellular Processes, Cell Death, Apoptosis (PATH:04941)
1049302	191	80.105	131	YLR229C-6 cerevisiae	191	belongs to the Ras subfamily of Ras-like proteins (GTPase, YEAH1)	Environmental Information Processing, Signal Transduction, MAPK signaling pathway (PATH:04941)
1049309	191	48.990	138	YPR165W-2 cerevisiae	209	GTP-binding protein of the rho subfamily of ras-like proteins (GTPase, YEAH1)	Environmental Information Processing, Signal Transduction, MAPK signaling pathway (PATH:04941)

### Pathway Discovery (Browsing, Sorting, Visualization & Statistical Analysis)

### Clustering

### Expression Pattern



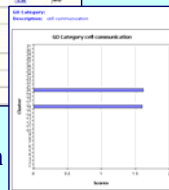
### Visualization & Statistical Analysis

# Integrated Knowledge Base

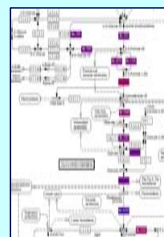
### Clustered Matrix

Category	Pathway name	Full name	Hit freq.	Chp. cont.	Sum up.
Apoptosis	apoptosis	p53 Signaling Pathway	4	2	15
Cell Cycle Regulation	cellcycle	Cell Cycle Regulation and Cyclins	4	12	23
Cell Cycle Regulation	cellcycle	Cell Cycle: G1/S Check Point	3		
Cell Cycle Regulation	cellcycle	Rb Tumor Suppressor/Checkpoint Signaling in response to DNA damage	2		
Cell Signaling	cellcycle	Akt Signal Transduction Pathway	2		
Immunology	cellcycle	Fc Epsilon Receptor 1 Signaling in Mast Cells	1		

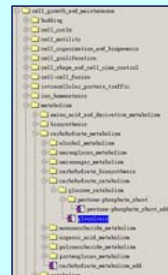
### Clustered Graph




### Pathway Map



### Process Hierarchy



# Protein Informatics for Expression Analysis



## Analysis of Gene Expression Data

About PIR
Databases
Search & Retrieval

Search   and   and Sort by

55 records found.

Expression Data	Protein ID and Title	Species and Superfamily	Domain and Motif	PDB	Category: Function	Freq.											
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>Probe</td><td>1069_at</td></tr> <tr><td>Gene</td><td>prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)</td></tr> <tr><td>GBAC</td><td>U04636</td></tr> <tr><td>F.C.</td><td>12.6 (4.9 to 1.0E+08)</td></tr> <tr><td>P-value</td><td>0.00</td></tr> </table>	Probe	1069_at	Gene	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	GBAC	U04636	F.C.	12.6 (4.9 to 1.0E+08)	P-value	0.00	<p><a href="#">NF00126043</a> Prostaglandin G/H synthase 2 precursor (EC 1.14.99.1) (Cyclooxygenase -2) (COX-2) (Prostaglandin-endoperoxide synthase 2) (Prostaglandin H2 synthase 2) (PGH synthase 2) (PGHS-2) (PHS II)</p>	Homo sapiens	PF00008: EGF-like domain (22-54)// PF03098: Animal haem peroxidase (129-562)	1DDX:C(18-568,88.4%); 1DDX:D(18-568,88.4%); 1DDX:A	GO:0000130 transcription factor GO:0000166 nucleotide binding GO:0003676 nucleic acid binding GO:0003677 DNA binding GO:0003700 transcription factor GO:0003712 transcription cofactor GO:0003750 cell cycle regulator GO:0003773 heat shock protein GO:0003801 blood coagulation factor GO:0003811 complement component	6 22 44 32 28 1 3 20 5 3	
Probe	1069_at																
Gene	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)																
GBAC	U04636																
F.C.	12.6 (4.9 to 1.0E+08)																
P-value	0.00																
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>Probe</td><td>1085_s_at</td></tr> <tr><td>Gene</td><td>phospholipase C, gamma 2 (phosphatidylinositol-specific)</td></tr> <tr><td>GBAC</td><td>M37238</td></tr> <tr><td>F.C.</td><td>5.4 (3.1 to 22.2)</td></tr> <tr><td>P-value</td><td>0.00</td></tr> </table>	Probe	1085_s_at	Gene	phospholipase C, gamma 2 (phosphatidylinositol-specific)	GBAC	M37238	F.C.	5.4 (3.1 to 22.2)	P-value	0.00	<p><a href="#">NF00081336</a> 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma 2 (EC 3.1.4.11) (PLC-gamma-2) (Phospholipase C-gamma-2) (PLC-IV)</p>	Homo sapiens <a href="#">SF000952</a>	PF00017: SH2 domain (532-617, 646-720)// PF00018: SH3 domain (772-827)// PF00168: C2 domain (1063-1152)// PF00169: PH domain (33-131, 849-908)// PF00387: Phosphatidylinositol-specific phospholipase C, Y domain (929-1044)// PF00388: Phosphatidylinositol-specific phospholipase C, X domain (313-457)	GO:0046666 : prostaglandin-endoperoxide synthase	[PATH: <a href="#">hsa00590</a> ]. <b>Note:</b> 100.000% identity in 604 overlap between NF00126043 (604aa) and S743-H.sapiens (604aa):prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase) [EC:1.14.99.1] [SP:PGH2_HUMAN]	GO:0003824:enzyme::GO:001649	
Probe	1085_s_at																
Gene	phospholipase C, gamma 2 (phosphatidylinositol-specific)																
GBAC	M37238																
F.C.	5.4 (3.1 to 22.2)																
P-value	0.00																
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>Probe</td><td>1104_s_at</td></tr> <tr><td>Gene</td><td>heat shock 70kD protein 1A</td></tr> <tr><td>GBAC</td><td>M11717</td></tr> <tr><td>F.C.</td><td>-5.2 (-4.6 to -5.9)</td></tr> <tr><td>P-value</td><td>0.00</td></tr> </table>	Probe	1104_s_at	Gene	heat shock 70kD protein 1A	GBAC	M11717	F.C.	-5.2 (-4.6 to -5.9)	P-value	0.00	<p><a href="#">NF00137040</a> dnaK-type molecular chaperone HSPA1L</p>	Homo sapiens <a href="#">SF002582</a>	PCM00329: PDC000269, Heat shock hsp70 proteins family signature 2 (PCT:197-210)	1HJO:A(3-382,99.5%); 1QQO:A(6-380,89.1%); 1HPM(1-3	GO:0016792 : phosphoric diester hydrolase; GO:0044435 : 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase	Metabolism; Metabolism of Complex Lipids; Inositol phosphate metabolism [PATH: <a href="#">hsa00562</a> ]. <b>Note:</b> 100.000% identity in 1252 overlap between NF00081336 (1252aa) and S336-H.sapiens (1252aa):phospholipase C, gamma 2 (phosphatidylinositol-specific) [EC:3.1.4.11] [SP:PIG2_HUMAN]	GO:0003824:enzyme::GO:001678
Probe	1104_s_at																
Gene	heat shock 70kD protein 1A																
GBAC	M11717																
F.C.	-5.2 (-4.6 to -5.9)																
P-value	0.00																
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>Probe</td><td>1104_s_at</td></tr> <tr><td>Gene</td><td>heat shock 70kD protein 1A</td></tr> <tr><td>GBAC</td><td>M11717</td></tr> <tr><td>F.C.</td><td>-5.2 (-4.6 to -5.9)</td></tr> <tr><td>P-value</td><td>0.00</td></tr> </table>	Probe	1104_s_at	Gene	heat shock 70kD protein 1A	GBAC	M11717	F.C.	-5.2 (-4.6 to -5.9)	P-value	0.00	<p><a href="#">NF00137040</a> dnaK-type molecular chaperone HSPA1L</p>	Homo sapiens <a href="#">SF002582</a>	PCM00329: PDC000269, Heat shock hsp70 proteins family signature 2 (PCT:197-210)	1HJO:A(3-382,99.5%); 1QQO:A(6-380,89.1%); 1HPM(1-3	GO:0030188 : chaperone regulator; GO:0003773 : heat shock protein;	Metabolism; Nucleotide Metabolism; Purine metabolism [PATH: <a href="#">dme00230</a> ]. <b>Note:</b> 81.106% identity in 651 overlap between NF00137040 (640aa) and CG4264-D melanogaster (651aa):adenosinetriphosphatase (heat shock protein cognate 4)	GO:0030234:enzyme regulator::GO:0004857:enzyme in GO:0030234:enzyme regulator::GO:0008047:enzyme ad GO:0003754:chaperone::GO:0009 protein GO:0003757:chaperone::GO:0009 protein GO:0003758:chaperone::GO:0009 protein GO:0003760:chaperone::GO:0009
Probe	1104_s_at																
Gene	heat shock 70kD protein 1A																
GBAC	M11717																
F.C.	-5.2 (-4.6 to -5.9)																
P-value	0.00																

# *Knowledge Base 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**

# *Acknowledgments*

## ● Sponsors

- NIH: NLM (PIR)
- NSF: BDI (iProClass); ITR (Ontology)

## ● PIR Team

- Cathy Wu, Winona Barker, Robert Ledley, Hongzhan Huang, Lai-Su Yeh, Bruce Orcutt, CR Vinayaka, Zhang-Zhi Hu, Baris Suzek, Yongxing Chen, Jim Zhang, Peter Kourtesis, Jorge L. Cardenas, Leslie Arminski