

Web-based Bioinformatics Applications in Proteomics

Chiquito Crasto
ccrasto@genetics.uab.edu

January 30, 2009

NCBI (National Center for Biotechnology Information)

<http://www.ncbi.nlm.nih.gov/>

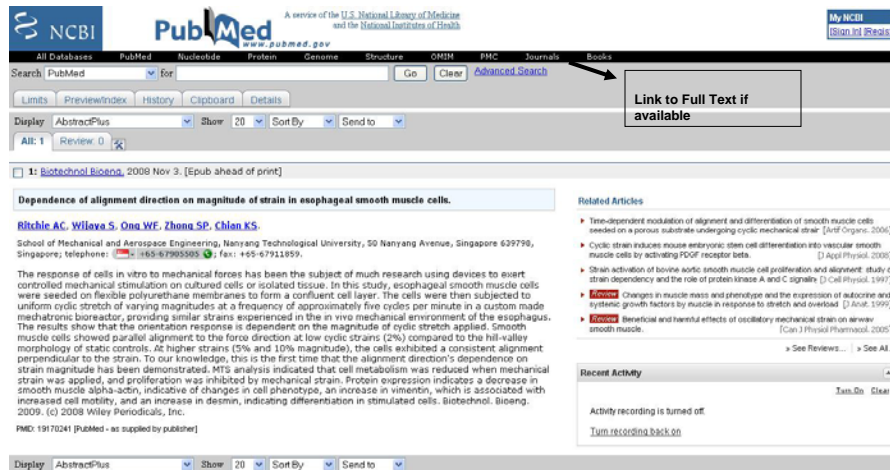
The image shows a screenshot of the NCBI website. On the left is a navigation menu with categories like 'SITE MAP', 'About NCBI', 'Genetics', 'Molecular Biology', 'Genomics', 'Tools', 'Search at NCBI', 'Software', 'Education', and 'FTP site'. The main content area is titled 'What does NCBI do?' and describes the center's mission. Below this, there are links to 'dbGaP: NCBI's Genome Wide Association Database', 'Pubmed Central', and 'NCBI News'. On the right, there is a search bar and a grid of database links. A red box highlights the 'Entrez' search engine, and another red box highlights the 'Protein Database' link. A black arrow points from the 'Protein Database' link in the grid to the 'Protein Database' link in the 'What does NCBI do?' section.

Pubmed—repository of biomedical abstracts



Information in Pubmed is available in several formats.
Abstracts can be downloaded 500 at a time
Abstracts can be specified in terms of date of publication, author lists, etc
If subscriptions are available, a user can access the full text of articles
NCBI has made several utility tools available to automatically download abstracts

A single Abstract in Pubmed



Nucleotide-Genbank's gene repository

NCBI Nucleotide

Search: Nucleotide for actin end Human

Found 15897 nucleotide sequences: Nucleotide [7975] EST [2921] GSS [1]

Display: Summary

Items 1 - 20 of 7975

This search in Gene shows 2876 results, including:

- ACTA1** (Homo sapiens): actin, alpha 1, skeletal muscle
- ELNA** (Homo sapiens): filamin A, alpha (actin binding protein 280)
- CORO1A** (Homo sapiens): coronin, actin binding protein, 1A

1: **NM_014688** Reports
Homo sapiens USP6 N-terminal like (USP6NL), transcript variant 1, mRNA
gi|212119073|ref|NM_014688.2|[212119073]

2: **NM_001080491** Reports
Homo sapiens USP6 N-terminal like (USP6NL), transcript variant 2, mRNA
gi|21219072|ref|NM_001080491.2|[21219072]

Accession Number

A Nucleotide Entry in Genbank

NM_001080491 Homo sapiens USP6 [gi212119072]

DEFINITION Homo sapiens USP6 N-terminal like (USP6NL), transcript variant 2, mRNA.

ACCESSION NM_001080491.2 GI21219072

FEATURES

Location/Qualifiers

- 1..4512
- /organism="Homo sapiens"
- /mol_type="mRNA"
- /db_xref="taxon:9606"
- /cblonumber="101"
- /map="10p11"
- 1..4512
- /gene="USP6NL"
- /gene_synonym="USP6L"
- /gene_synonym="USP6L2"
- /note="USP6 N-terminal like"
- /db_xref="GeneID:121211"
- /db_xref="WORM:121211"
- /db_xref="XREF:121211"
- 44..1250
- /gene="USP6NL"
- /gene_synonym="USP6L"
- /gene_synonym="USP6L2"
- /note="Exon 2 is encoded by transcript variant 2; related to the N terminus of the"
- /product="USP6 N-terminal like isoform 2"
- /protein_id="GI21219072.1"
- /db_xref="GI:21219072.1"
- /db_xref="GeneID:121211"
- /db_xref="GeneID:121211"

ORIGIN

1 actataaact aatataaaa tttggaata gaaatgacg tctatgaga ctgaaatgt
41 tctaatcc aagagaga ctgatact caaaagat taagcaag atgaaatc
111 caaatgag aagagaga ctgatact caaaagat taagcaag atgaaatc
181 aagatgaa atgtgaga atgtgata octtgctac aagatcag atagattg
241 atttatac agagagag tccatgaa taatgatt gtagaagg aagaagcc
301 gaaatgaa aagatgaa atgtgata caatgaga gtagagaa aatcaaga
361 aatgaaga tctatgag gaattaca aagataca ctccagca gaggtagc
421 ctgagcca atgtgaga tctataat gaagagaa aagagaga tgaatgaa
481 atataaac agagagag atgtgata tgaatgaa caaatgaa tgaatgaa
541 agcaatct agagagag atgtgata tgaatgaa caaatgaa tgaatgaa
601 atccatga atgtgata atgtgata tcaacaga gtagagaa atccatga

Links to Pubmed

Protein Sequence

Gene Sequence

Protein Sequence in Genbank (based on the search “Actin and Human:”

The screenshot displays the NCBI Protein database search results for the query "Actin and Human". The left pane shows a list of search results with columns for accession number, protein name, and links. The right pane shows the details for a specific protein entry, including its accession number (AA023103), protein name (Actin), and the full amino acid sequence in GenBank format.

Note that the protein sequence and the rest of the entries are formatted similar to that of the nucleotide sequences in Genbank. This is a database architecture issue.

Three-dimensional structure views in Genbank

The screenshot displays the NCBI Structure database search results for the query "actin". The search bar shows "actin" and the results page displays "Hints on finding a Structure" and "New structure viewer" sections.

Structure of Actin—Genbank Structure View

Structure Summary
MMDDB

MMDDB ID: 69126 | PDB ID: 2ZWH

Description: Model For The F-Actin Structure.
Deposition: Oda T, Iwasa M, Aihara T, Maeda Y, Nanta A, 2008/12/5
Taxonomy: *Cryptosporidium parvum*
Related Structure: VAST

Structure View in Cn3D | Structure View in RasMol

Tasks: Display | Drawing: All Atoms

Download Cn3D | View Cn3D Tutorial ← Visualization software

Molecular components in the MMDDB structure are listed below and may include macromolecular chains, 3D domains, protein classifications (domain families), and ligands, as available. Mouse over each icon for more information on the component.

Protein: ACTIN
M-Domains: ACTIN
Domain Families: ACTIN
Specific NITE: ACTIN
Superfamilies: ACTIN superFamily
Multi-Domains: ACTIN

Structure of Domains in Genbank

Structure Summary
MMDDB

MMDDB ID: 69143 | PDB ID: 3EDV

Reference: Davis L, Abd K, Machuga M, Braudgam C, Tomchick DF, Bennett V, Michalek P. Localization and structure of the ankyrin binding site on beta 2-spectrin. *J. Biol. Chem.* 267: 1111-1115, 1992.

Spectrins are tetrameric, actin-crosslinking proteins that form an elastic network, beneath the membrane skeleton, on the cytoplasmic surface of cellular membranes. At the plasma membrane, the membrane skeleton provides essential support, preventing loss of membrane material to environmental shear stresses. The skeleton also controls the location, abundance and activity of membrane proteins that are critical to cell and tissue function...

Description: Crystal Structure Of Repeats 14-16 Of Beta2-Spectrin.
Deposition: 2008/9/3
Taxonomy: *Homo sapiens*
Related Structure: VAST

Structure View in Cn3D | Structure View in RasMol

Tasks: Display | Drawing: All Atoms

Download Cn3D | View Cn3D Tutorial

Molecular components in the MMDDB structure are listed below and may include macromolecular chains, 3D domains, protein classifications (domain families), and ligands, as available. Mouse over each icon for more information on the component.

Protein: SPECT
M-Domains: SPECT
Domain Families: SPECT
Specific NITE: SPECT
Superfamilies: SPECT superFamily

Protein: SPECT
M-Domains: SPECT
Domain Families: SPECT
Specific NITE: SPECT
Superfamilies: SPECT superFamily

List of domains related to or associated with Actin →

Link to Protein Databank →

Conserved domain database (CDD) in Genbank

The screenshot shows the NCBI Conserved Domain Database (CDD) search results. The main list includes items 1 through 6, each with a small 3D protein structure icon and a text description. Item 1 is ACTIN, item 2 is F-actin_head, item 3 is ACTIN, item 4 is F-actin_rep_A, item 5 is F-actin_rep_B, and item 6 is Actin. To the right, a 'Detailed alignment' window is open, displaying a sequence alignment with residue numbers (e.g., 80, 100, 120, 140, 160, 180, 200, 220, 240) and domain boundaries indicated by brackets and labels like 'F-actin_head'.

Clustering Proteins in terms of Sequence Similarities--Genbank

The screenshot shows the NCBI Protein Clusters database homepage. The page has a search bar at the top with the text 'Search Protein Clusters' and a 'Go' button. Below the search bar are navigation tabs: 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The main content area is titled 'Protein Clusters' and contains an 'About the Database' section. This section includes a welcome message, information about the database's scope (related protein sequences), and instructions on how to search for clusters. Below the 'About' section is an 'Example Searches' section with three search queries: 'all clusters with ribosomal protein as the curated name', 'ribosomal protein[Protein Name]', and 'source:chloroplast[All Fields]'. The page also features a sidebar with various navigation and resource links.

Clustering Proteins in terms of Sequence Similarities--Genbank

PRK13410 molecular chaperone DnaK Gene name: None

Cluster Info: 858088
 Total proteins: 12
 Conserved in: **Cyanobacteria**
 Total genes: 9
 Total organisms: 12
 Putative Paralogs: 0
 Publications: 13

Cluster Tools: Show detailed alignment, Build tree, Genome PlotMap by PRK13410, Genome PlotMap by COG04430

Cross references: COG(s): COG04430, HAMAP: MF_00332, HEGG KO: MB043, InterPro: I, TrEMBL: T58802359, Domain(s): c08318(AC106), pfam02782(GGY_G)

Entrez Links

Organism	Protein name	Prev. Cluster	Accession	Next Cluster	Locus tag	Length	BLink	Alignment
Cyanobacteria								
<input type="checkbox"/> <i>Synechococcus maris neri</i> M8C1017	msl0241.ubaa000a	CL1121015	YP_001519959	CL1121030	AM1_0992	670aa	+	
<input type="checkbox"/> <i>Halobacterium salinarum</i> ATCC 29613	msl0241.ubaa000a	CL1121030	YP_321438	CL1121030	hsl_0919	680aa	+	
<input type="checkbox"/> <i>Cyanobacterium</i> sp. ATCC 61142	msl0241.ubaa000a	PR050296	YP_001802772	CL1121013	hsl_1398	730aa	+	
<input type="checkbox"/> <i>Micromastix aeruginosa</i> M81-843	msl0241.ubaa000a	CL1119292	YP_001555508	CL1121013	M81_1690	720aa	+	
<input type="checkbox"/> <i>Nautilia profundorum</i> FCC 72102	msl0241.ubaa000a	CL1121030	YP_001895211	CL1121030	Npna_09099	705aa	+	
<input type="checkbox"/> <i>Nautilia</i> sp. FCC 72102	msl0241.ubaa000a	CL1121030	MF_467030	CL1121030	na0900	680aa	+	
<input type="checkbox"/> <i>Synechococcus elongatus</i> FCC 8303	msl0241.ubaa000a	CL0902792	YP_172249	CL0902794	sy1630_4	740aa	+	
<input type="checkbox"/> <i>Synechococcus elongatus</i> FCC 7945	msl0241.ubaa000a	CL0902792	YP_481987	CL0902794	Synp07942_280	740aa	+	
<input type="checkbox"/> <i>Synechococcus</i> sp. FCC 7002	msl0241.ubaa000a	CL1121089	YP_001735395	CL1119292	SYNPC7002_42100	702aa	+	
<input type="checkbox"/> <i>Synechococcus</i> sp. FCC 8803	msl0241.ubaa000a	PR050011	MF_464943	CL1121012	sp1932	700aa	+	
<input type="checkbox"/> <i>Thermosynechococcus elongatus</i> M81	msl0241.ubaa000a	CL1121089	MF_852945	CL1121030	td1150	680aa	+	
<input type="checkbox"/> <i>Tricholimnium</i> <i>andersonii</i> M81501	msl0241.ubaa000a	CL1121030	YP_722605	CL1121030	Tav_4012	670aa	+	

ENSEMBL—European version of Genbank—now focused exclusively on genome wide applications

Ensembl Home

Search Ensembl: Search: Human for BRCA2 Go
 a.g. human gene BRCA2 or rat X:100000..200000 or insulin

Browse a Genome
 The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.
 Click on a link below to go to the species' home page.

Popular genomes

Human (NCBI96)	Dog (ENSDOG2)
Chimpanzee (ENSMUS1)	Chicken (ENSGAL2)
Macaque (MIMUL1)	Zebrafish (ZFIN07)
Mouse (NCBIM87)	Fruitfly (DROME4)
Rat (RNOEC4)	C.elegans (WORM2)

All genomes: - Select a species -

Other pre-built species are available in [Ensembl Perl](#)

Other Ensembl websites

- [archivio](#) - past releases of Ensembl
- [VEGA](#) - Vertebrate Genome Annotation
- [Ensembl mirror in Bio-ST, Release 50](#)
- [Ensembl mirror in Bio-ST, Release 51](#)

New to Ensembl?

- Did you know you can:
 - [Add custom tracks](#) using our new Control Panel
 - [Upload your own data](#) and save it to your Ensembl account
 - [Search for a RNA or protein sequence](#) using BLAST or BLAT
 - [Fetch only the data you want](#) from our public database, using the Ensembl Perl API
 - [Download our databases via FTP](#) in FASTA, MySQL, and other formats
 - [Mine Ensembl with BioMart](#) and export sequences or tables in text, html, or Excel format
- Still got questions? [Try our FAQs](#)

What's New in Release 52 (5 December 2008)

- Homo sapiens core database (Human)
- Gorilla 2x assembly and genebuild (Gorilla)
- ncRNA update (multiple species)
- Mus musculus core (Mouse)
- Cow otherfeatures (Cow)

[More news](#)

Sanger Ensembl is a joint project between EMBL, EBI and the Wellcome Trust Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes.
 Ensembl receives major funding from the Wellcome Trust. Our [acknowledgements page](#) includes a list of additional current and previous funding bodies.

EMBL-EBI

Sample Ensembl Result—Chromosomal location and other features for downloading information

Gene: BRCA2 (ENSG00000139618)
Breast cancer type 2 susceptibility protein (Fanconi anemia group D1 protein) [Ensembl: UniProt/RefSeq/Protein/Prot PS1987](#)

Location: [Chromosome 13: 31,797,617-31,871,808](#) forward strand

Transcripts: There is one transcript in this gene: [BRCA2-001 \(ENST00000380152\)](#), with protein product [ENSP00000369497](#)

[Gene summary](#) [/help](#) [Splice variants »](#)

Name: [BRCA2](#) (HGNC (curated))

Synonyms: BRCC2, FACD, FAD, FAD1, FANCD, FANCD1 [[To view all Ensembl genes linked to the name \[click here\]\(#\)](#)]

CCDS: This gene is a member of the Human CCDS set: [CCDS3344](#)

Gene type: Known protein coding

Prediction Method: Gene containing both Ensembl genebuild transcripts and [Vega](#) manual curation, see [article](#)

Transcripts

Configuring the display
Tip: use the "Configure this page" link on the left to show additional data in this region.

Ensembl release 52 - Dec 2008 © [WTSI](#) / [EBI](#). Ensembl is available to [download for public use](#) - please see the [code licence](#) for details. This is a mirror site of Ensembl from [BGI-SZ](#).
[Permanent link](#) - [View in archive site](#) [About Ensembl](#) | [Contact Us](#) | [Help](#)

SWISSPROT--<http://www.expasy.ch/>

– UniProt combines SwissProt and TrEMBL

“UniProtKB/TrEMBL (unreviewed) contains protein sequences associated with computationally generated annotation and large-scale functional characterization. UniProtKB/Swiss-Prot (reviewed) is a high quality manually annotated and non-redundant protein sequence database, which brings together experimental results, computed features and scientific conclusions” --<http://www.uniprot.org/help/uniprotkb>

In the very near future, UniProt will replace SwissProt

Mirror Sites

Switzerland: <http://www.expasy.org/> at [Swiss Institute of Bioinformatics, Geneva](#)

Australia: <http://au.expasy.org/> at [Australian Proteome Analysis Facility, Sydney](#)


Brazil: <http://br.expasy.org/> at [Laboratório Nacional de Computação Científica, Petrópolis](#)

Canada: <http://ca.expasy.org/> at [Canadian Bioinformatics Resource, Halifax](#)

China: <http://cn.expasy.org/> at [Peking University](#)

Korea: <http://kr.expasy.org/> at [Yonsei Proteome Research Center, Seoul](#)

SwissProt--http://www.expasy.org/



ExpASY Proteomics Server
The ExpASY (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (SIB) is dedicated to the analysis of protein sequences and structures as well as 2D PAGE. (Disclaimer / References / Linking to ExpASY)

[Database] [Tools & Software] [Education & Services] [Links] [Publications] [Direct Sites] [Job openings]




Databases	Tools and software packages
<ul style="list-style-type: none"> UniProt Knowledgebase (Swiss-Prot and TrEMBL) - Protein knowledgebase ViewZone - Portal to view UniProtKB/Swiss-Prot entries PROSITE - Protein families and domains SWISS-2DPAGE - Two-dimensional polyacrylamide gel electrophoresis World-2DPAGE Repository - A public standards-compliant repository for gel-based proteomics data published in the literature MAPE@SIB - A public repository for MAPE Gel electrophoresis documents ENZYME - Enzyme nomenclature UniPathway - Metabolic pathways SWISS-MODEL Repository - Automatically generated protein models Links to many other molecular biology databases 	<ul style="list-style-type: none"> Proteomics and sequence analysis tools <ul style="list-style-type: none"> Identification and characterization (Abolite, FindMot, Popdam, Pheny, pMv, ProScan...) DNA -> Protein Similarity searches (BLAST...) Pattern and profile searches (ScanProse...) Post-translational modification and topology prediction Primary structure analysis Secondary and tertiary structure tools (Swiss-PdbViewer...) Alignment and Phylogenetic analysis Melanin ImageMaster - Software for 2-D PAGE analysis MSight - Mass Spectrometry Imager Roche Applied Science's Biochemical Pathways
Education and services	
<ul style="list-style-type: none"> The ExpASY FTP server Swiss-Shop - automatically obtain (by email) new sequence entries relevant to your field(s) of interest Popular Science <ul style="list-style-type: none"> Protein Spotlight Proteins à la Une e-Proxemis - Bioinformatics Learning Portal for Proteomics 	<ul style="list-style-type: none"> Bioinformatics Core Facility for Proteomics Proteomics Core Facility (formerly known as SWISS-2DSERVICE) SFS Digest - the Swiss Proteomics Society selection of proteomics articles Master's degree in Proteomics and Bioinformatics Swiss-Quiz
Lists of molecular biology resources	
<ul style="list-style-type: none"> ExpASY Life Science Directory - The ExpASY list of biomolecular servers BioLinks - Search the internet for molecular biology information WORLD-2DPAGE List - Links to 2-D PAGE database servers and 2-D PAGE related servers and services World-2DPAGE Portal - A dynamic portal to query simultaneously World-Wide gel-based proteomics databases 	
Swiss Institute of Bioinformatics (SIB) services	
<ul style="list-style-type: none"> VitalIT - The HPC Center for Life Sciences Swiss node of EMBnet Astha Genome Database Cancer Immunome Database 	<ul style="list-style-type: none"> ClearIT - Gene expression database EPD - Eukaryotic Promoter Database GermOnLine - Knowledgebase on germ cell differentiation smRNAdb - a database of small RNAs SwissRegulon - a database of regulatory sites

The links here are the set of tools for various proteomic analysis. Some of these are replicated in Genbank.

SwissProt—search for Proteins

Notice: This page will be replaced with www.uniprot.org. Please send us your feedback!

Search for

UniProt
Protein knowledgebase
TrEMBL
Computer-annotated supplement to Swiss-Prot

Search for

Search in UniProt Knowledgebase (Swiss-Prot and TrEMBL) for: actin

UniProtKB/Swiss-Prot Release 56.7 of 20-Jan-2009
UniProtKB/TrEMBL Release 39.7 of 20-Jan-2009

- Number of sequences found in UniProt Knowledgebase (Swiss-Prot₄₃₉₆ and TrEMBL₂₈₈₁₅): **33010**
- Note that the selected sequences can be saved to a file to be later retrieved, to do so, go to the bottom of this page.
- For more directed searches, you can use the Sequence Retrieval System SRS.

Search in UniProtKB/Swiss-Prot: There are matches to 4395 out of 408099 entries

6PGD_ACTAC (**P70718**)
6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44) [GENE: Name=gnp] - Actinobacillus actinomycetemcomitans (Haemophilus actinomycetemcomitans)

6PGL_ACTAC (**P70715**)
6-phosphogluconolactonase (EC 3.1.1.31) (6PGL) [GENE: Name=pgl; Synonyms=devE] - Actinobacillus actinomycetemcomitans (Haemophilus actinomycetemcomitans)

A26CB_HUMAN (**A5A3E0**)
ANKRD26-like family C member 1B (Chimeric POTE-actin protein) [GENE: Name=A26CB1B] - Homo sapiens (Human)

AAC_ACTUT (**P29958**)
Aculeacin-A acylase precursor (EC 3.5.1.-) [Contains: Aculeacin-A acylase small subunit, Aculeacin-A acylase large subunit] [GENE: Name=aac] - Actinoplanes utahensis

AB140_YEAST (**O08641**)
Uncharacterized methyltransferase ABP140 (EC 2.1.1.-) (140 kDa actin-binding protein) [GENE: Name=ABP140; Ordered_ocusNames=YOR239W/YOR240W] - Saccharomyces cerevisiae (Baker's yeast)

AB1P_CHICK (**O52L23**)
Amyloid beta A4 precursor protein-binding family B member 1-interacting protein (AFBB1-interacting protein 1) [GENE: Name=APBB1P; ORFNames=RCJM04_8c12] - Gallus gallus (Chicken)

AB1P_DANRE (**G5PFT9**)
Amyloid beta A4 precursor protein-binding family B member 1-interacting protein (AFBB1-interacting protein 1) [GENE: Name=agbb1ip] - Danio rerio (Zebrafish) (Brachydanio rerio)

AB1P_HUMAN (**Q725R6**)

Swissprot-Actin (Results in a Single Page)

The screenshot displays the UniProt entry for Actin (P0ABF1). Key sections include:

- General:** Name: Actin, Gene: ACTA1, Organism: Homo sapiens.
- Amino acid sequence:** A table showing the full sequence of the protein, with a highlighted segment.
- Protein names:** A list of names and synonyms, including 'Actin', 'F-actin', and 'G-actin'.
- References:** A list of scientific publications related to the protein.

Swiss 2DPAGE

SWISS-2DPAGE
Two-dimensional polyacrylamide gel electrophoresis database

SWISS-2DPAGE contains data on proteins identified on various 2-D PAGE and SDS-PAGE reference maps. You can locate these proteins on the 2-D PAGE maps of 80% of the region of a 2-D PAGE map where one might expect to find a protein from UniProt/Swiss-Prot (More details | References | Linking to SWISS-2DPAGE | Commercial start | Disclaimer)

Release 18.6, September 2006 and updates up to 31 January 2008 (contains 1262 entries in 36 reference maps from human, mouse, Arabidopsis thaliana, Dictyostelium discoideum, Escherichia coli, Saccharomyces cerevisiae and Staphylococcus aureus (N718)).

[Search] [Documents] [Services] [Substrains] [Related servers] [Other databases] [Job openings]

Access to SWISS-2DPAGE	SWISS-2DPAGE documents
<ul style="list-style-type: none"> By description (any word in the DE, OS, GN and DI lines) By accession number (AC line) By clicking on a spot: select one of our 2-D PAGE or SDS-PAGE reference maps, click on a spot and then get the corresponding information from the SWISS-2DPAGE database By author (RA lines) By spot serial number (2D and 1D lines) By experimental gel image By experimental identification methods By full text search Retrieve all the protein entries identified on a given reference map Complex queries (SPS line) compute estimated location of reference maps for a user-entered sequence 	<ul style="list-style-type: none"> User manual Release notes (September 26, 2006) FAQ (Frequently Asked Questions about SWISS-2DPAGE) Protocols: <ul style="list-style-type: none"> Technical information about 2-D PAGE (2D), silver staining, protocols, etc. High performance 2-D gel comparison Figure captions of SWISS-2DPAGE maps available from publications: <ul style="list-style-type: none"> Human CBF, ELK, HEPG2, HEPG2SP, LIVER, LYMPHOMA, PLASMA, PLATELET, RBC, LYST, CDC, K562/1 Dictyostelium discoideum, Escherichia coli, Saccharomyces cerevisiae
Software	<ul style="list-style-type: none"> ImageMaster 2D - Software package for 2-D PAGE analysis Max2D-DB package ver 2.50 - A package preparing the data and the programs necessary to build a hierarchical 2-DE database on one's own web site.
Outways to other 2-D PAGE related servers and services	<ul style="list-style-type: none"> The SWISS-2DPAGE Consortium - Promote and publish gel-based proteomics data through the EFP-Italy server WORLD-2DPAGE List - Index to other European 2-D PAGE databases SWISS-2DPAGE Portal - A dynamic portal to query simultaneously words-wide gel-based proteomics databases SWISS-2DPAGE Repository - A public standards-compliant repository for gel-based proteomics data published in the literature 2D-HUNT - 2-D electrophoresis web site finder Proteomics tools

SWISS-2DPAGE


Search by description (DE), entry name (ID), gene name (GN) or UniProtKB/Swiss-Prot keywords (KW)

Enter search keywords:

Limit to: All fields DE ID GN KW

KEGG (Kyoto Encyclopedia of Genes and Genomes)

http://www.genome.jp/kegg/



Search KEGG Get Entry

KEGG Home
Introduction
Overview
Release notes
Current statistics

KEGG Identifiers
KGML
KEGG API
KEGG FTP
KegTools

Feedback
GenomeNet

KEGG: Kyoto Encyclopedia of Genes and Genomes

A grand challenge in the post-genomic era is a complete computer representation of the cell, the organism, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviors from genomic and molecular information. Towards this end we have been developing a bioinformatics resource named KEGG as part of the research projects of the Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo.

- Main entry point to the KEGG web service**
 - KEGG2 KEGG Table of Contents Update notes Help
- Data-oriented entry points**
 - KEGG Atlas Global maps of cell/organism functions
 - KEGG PATHWAY Pathway maps and pathway modules
 - KEGG BRITE Functional hierarchies and ontologies
 - KEGG ORTHOLOGY KO system and ortholog annotation
 - KEGG GENES Genomes, genes, and proteins
 - KEGG LIGAND Chemical compounds, drugs, glycans, and reactions
- Organism-specific entry points**
 - KEGG Organisms Select (example) hsa
- Subject-specific entry points**
 - KEGG DISEASE Gene/molecule based disease information resource
 - KEGG DRUG Chemical structure based drug information resource
 - KEGG GLYCAN Glycome informatics resource
 - KEGG COMPOUND Knowledge base for biochemical compounds
 - KEGG REACTION Knowledge base for biochemical reactions
 - KEGG PLANT Knowledge base for plant natural products
 - KAAS KEGG automatic annotation server

Copyright 1995-2009 Kanehisa Laboratories

Kegg Atlas

KEGG2 ATLAS PATHWAY BRITE KO GENES BIRD LIGAND DRUG

KEGG Atlas

KEGG Atlas is a new graphical interface to the KEGG suite of databases, especially to the systems information in the PATHWAY and BRITE databases. It currently consists of a global metabolism map with **highly developed viewers and 8 cancer maps with the traditional KEGG map viewer.**

Metabolism map (version 5.1, to be phased out)

Metabolism map viewer

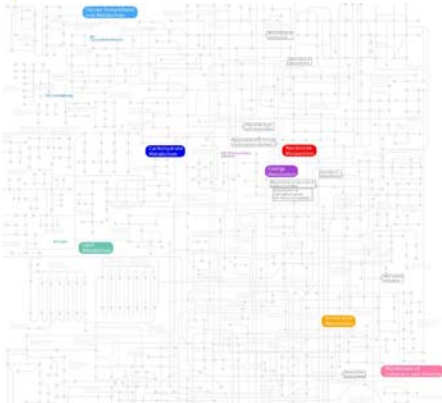
Plant secondary metabolism map

Cancer map

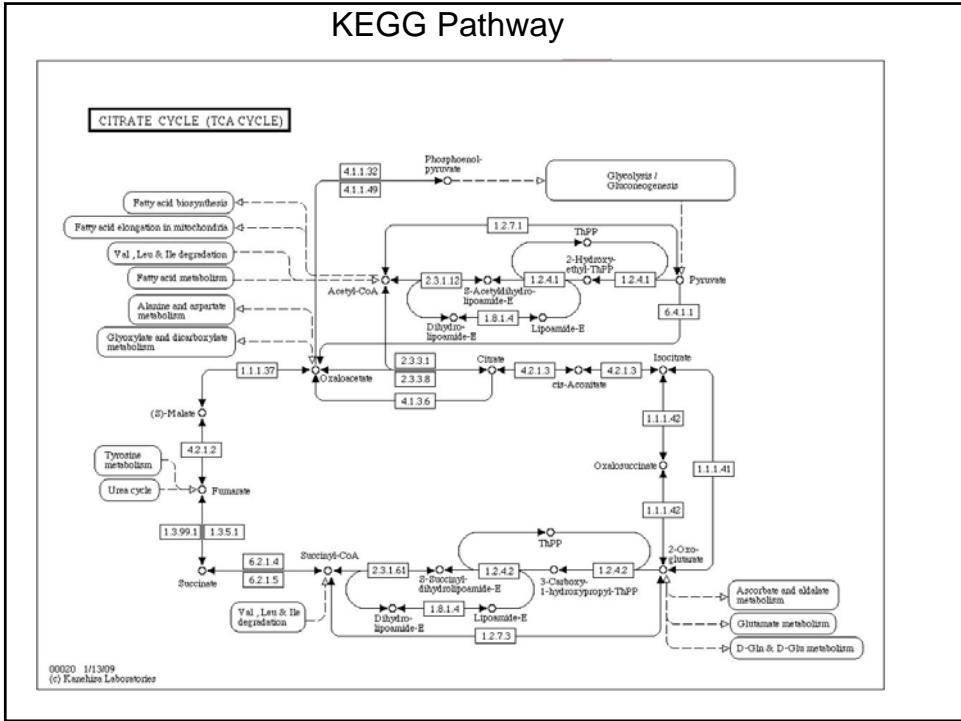
Other global maps are being developed or planned including:

- Cell map
- Body map
- Brain map

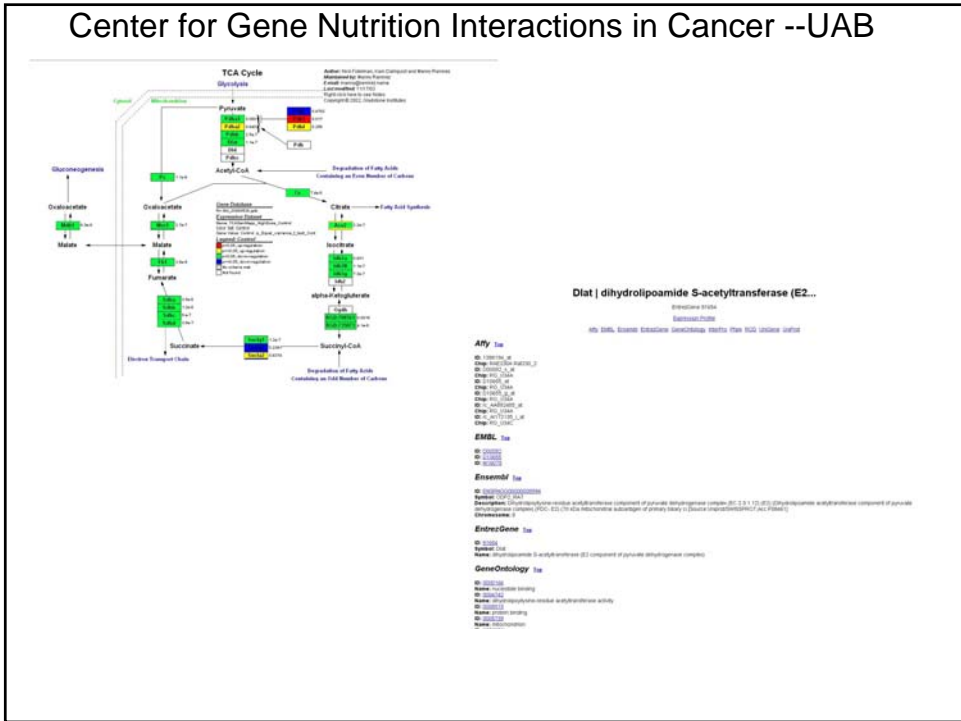
The new **KEGG metabolism map** is created as an SVG file by manually combining about 120 existing metabolic pathway maps. Each node (circle) is a chemical compound identified by the C number. Each line (curved or straight) connecting two nodes is manually defined as a segment lacking branches in the existing maps, named **MetElement**, and identified by the N number. Each **MetElement** corresponds to one or several KOs (such as this) in the reference pathway view, or one to several genes (such as this) in an organism-specific view.



KEGG Pathway



Center for Gene Nutrition Interactions in Cancer --UAB



MASCOT—Protein Identification from Mass Spectrometry Data

- Peptide Mass Fingerprinting
- Sequence Query
- MS/MS Ion Search

Welcome

This site features MASCOT, a powerful search engine that uses mass spectrometry data to identify proteins from primary sequence databases. To assist you, the help text for MASCOT forms a substantial knowledge base concerning protein identification by MS.

If this is your first visit, please check for browser compatibility and read the [read me](#) page. If you include results from MASCOT in a publication, please cite either [www.matrixscience.com](#) or [Electrophoresis](#), **20**(18):3553-67 (1999) (abstract).

We value your feedback and suggestions for new features. If you find any problems, errors, omissions, or just get unexpected results then please let us know.

For information on licensing MASCOT for in-house use, please refer to our [Products](#) and [Support](#) pages. For recent news, check [What's New](#).

Matrix Science develops and markets software products which integrate mass spectrometry into bioinformatics. Our interests extend to all aspects of mass spectrometry in the life sciences. Please contact us to discuss:

- Developing new applications
- Consultancy in mass spectrometry and bioinformatics
- Systems analysis and integration

Collaborations

MASCOT incorporates code from Mosaic, developed by Carol Pappin and David Perkins when working at the former Imperial Cancer Research Fund, and licensed from its technology transfer subsidiary, Cancer Research Technology.

LabVantage Solutions and Matrix Science are working together to develop data management and data mining solutions for proteomics.

We are grateful to the Swiss Institute of Bioinformatics for permission to make SwissProt available on this web site for searching with MASCOT.

MASCOT Search Results

MASCOT Search Results

Peptide View

MS/MS Fragmentation of QAGLSYIR

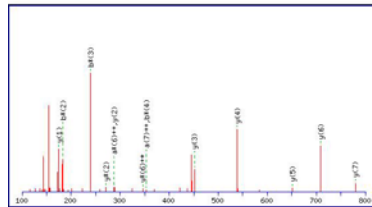
Found in [gi258789](#), H(+)-ATP synthase epsilon-subunit [rats, liver, Peptide Mitochondrial, 50 aa]

Match to Query 39: 906.598884 from(454:306718,2+)

File: NanoSET1.wcf, Sample: Complex V Lane-1-1 (sample number 1), Elution: 23.07 min, Period: 1, Cycle(s): 2586 (Experiment 4)
From data file C:\DOCUME~1\MASSSP~1\LOCALS~1\Temp\nms2C5.tmp

Click mouse within plot area to zoom in by factor of two about that point

Or, Plot from to Da



Monoisotopic mass of neutral peptide H(calc): 906.49

Ion Score: 47 Expect: 0.043

Matches (Bold Red): 14/84 fragment ions using 23 most intense peaks

#	a	a ⁺⁺	a ⁺	a ⁺⁺⁺	b	b ⁺⁺	b ⁺	b ⁺⁺⁺	Seq.	y	y ⁺⁺	y ⁺	y ⁺⁺⁺	#
1	101.07	51.04	84.04	42.53	129.07	65.04	112.04	56.52	Q					8
2	172.11	86.56	155.08	78.04	200.10	100.56	183.08	92.04	A	779.44	390.22	762.41	381.71	7
3	229.13	115.07	212.10	106.56	257.12	129.07	240.10	120.55	G	708.40	354.71	691.38	346.19	6
4	342.21	171.61	325.19	163.10	370.21	185.61	353.18	177.09	L	651.38	326.19	634.36	317.68	5
5	429.25	215.13	412.22	206.61	457.24	229.12	440.21	220.61	S	538.30	269.65	521.27	261.14	4
6	592.31	296.66	575.28	288.14	620.30	310.66	603.28	302.14	Y	451.27	226.14	434.24	217.62	3
7	705.39	353.20	688.37	344.69	733.39	367.20	716.36	358.68	I	388.20	144.61	371.18	136.09	2
8									R	175.12	88.06	158.09	79.55	1

Other Mass Spectrometry Analysis Software

- Phenyx--<http://phenyx.vital-it.ch/pwi/login/login.jsp>
- Aldente--<http://www.genebio.com/aldente/index.html>

Protein Data Bank-PDB

- <http://www.rcsb.org/pdb/home/home.do>
- **“A Resource for Studying Biological Macromolecules**

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the [wwPDB](http://www.rcsb.org/pdb/home/home.do), the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.”

Bioinformatics and Computational Biology as a Drive or Discovery

- **Recognition of the mechanism of olfaction depends upon understanding the sequence-structure-function relationships of olfactory receptors. We used bioinformatics methods to identify a new structural sub-class of olfactory receptors and GPCRs. We used available statistical methods to predict trans-membrane helical domains in olfactory receptor hOR17-210, a receptor that has been shown to be variably functional and pseudogenic in humans.**
- **TM domain identification was undertaken as a prelude to modeling this olfactory receptor in order to understand its interaction with ligands that have been experimentally shown to bind to this receptor. Our analyses revealed that there are only five typically observed TM regions in this protein with an additional orphan TM. The C-terminus is extra-cellular. This reversed polarity in the termini does not disrupt the positions of typical OR-motifs that initiate the signal transduction process at the membrane.**
- **Our observations are contrary to conventional structural knowledge about ORs and GPCRs. Preliminary sequence analysis studies have shown that such a structure is observed in a limited number of olfactory receptors distributed across different mammalian species.**
- **We also present a methodology for modeling such olfactory receptors. When combined with experimental data, we believe that this information will further our understanding of olfaction.**

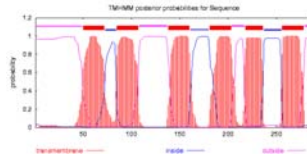
- **Sequence Features for OR17-210**
- **This protein sequence for olfactory receptor OR17-210 appears as a pseudogene in the HORDE1 database:**

```

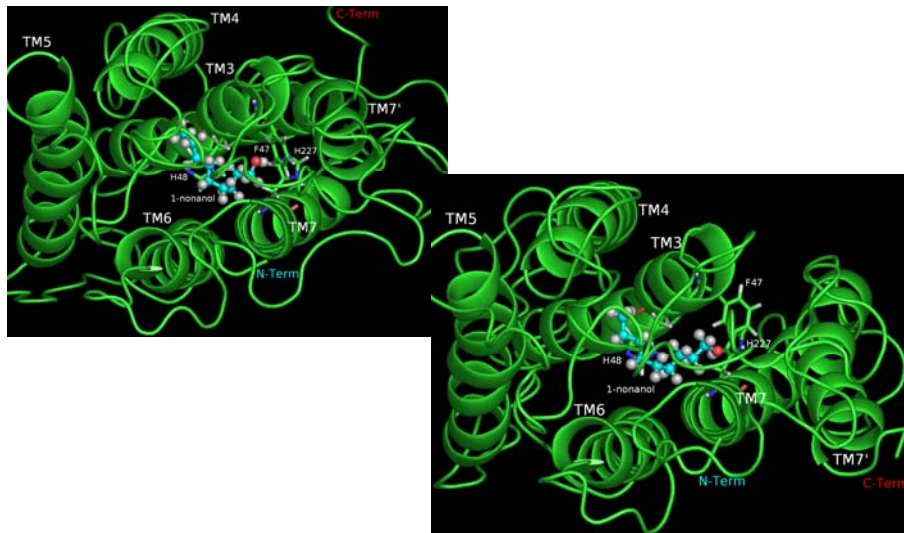
• ATGATGAAGAAGAACCACCAACCATGATCTCAGAGTTCCTGCTCCTGGGCCTTCCATCCAACCTGAGCAGCAGAATCTGTTCTATGCGT
TGTTCTTGGCCGTGATCTTACCACCCTCCTGGGGAACCTCCTCGTCATTGCTCATTGCACTGGACTCCACCTCCACATGCGCTA
TGTAATTTGTCTCAGCAACTTGTCTTCTCTGACCTCTGCTTTCCCTCGGTACAAATGCCCAAATTTGCTGCAGAAATGCAGAGCCCA
AAACCCATCCATCCCTTTCCGAGTGGCTGGCTGAGATGACTTTGATCTGTTTATGGAGTTCTGGAGAGCTTCCTCCTTTGGT
CATGGCTTACACTGCTATGTGGCTATTGCTTTCCTCTGCACTACACCACTATCATGAGCCCAAGTGTTCCTTGGTCTGCTGACA
CTCTCCTGGCTGTGACCACTGCCATGCCACGTTGCACACTTGCCTATGGCCAGGCTGTCTTTTGTGTGAGAAATGTGATTCTCT
CACTTTTTCTGTGATACATCTACTTTGTGAAGTGGCTGCTCCAAACCCAACTCAATGGTGGTGTGATGTTTTCTATGGGGGG
CTCATCCTTGTGATCCCATCTCTACTCTCATCATGCTCTGTGCAAGAATCGTCTCCACCATCTCAGGGTCCCTTCCACTGGGGCC
ATCCAGAAGGCTTCTCCACTGTGGCCCCACCTCTCTGTGGTGTCTCTCTCTATGGGACAATATTGGTCTACTTGTGGCCA
TTGACGAATCATAACTGTGAAGGACACTGCTATGGCTGTGATGTACACTGGGGTGAACCCATGCTGAACCCCTTCTATGACAG
CTGAGGAACAGAGACATGAGGGGAACCTGGCCAGAGTCTTACGCACAAAGAAATTTTTTTGTCTTAAAAATGATAATGTTGGC
ATTTTACCCTTATGAAAT
    
```

- **Intuitively Translated as:**
- **MMKKNOTMISEFLLGL/PIOPEQONLFYALFLAVYLTLLGNLLVIVLIRLDSHLHMPMYLCLSNLSFSDLCFSSVTMPKLLONMOSQNPSPFADCLAQMYFHLFYGVLESFLLVVMAYHCYVAICFPLHYTTIMSPKCLGLLTLWLLTAHATLHLLMARLSFCAENVIPHFCDTSTLLK LACSNTQVNGWVMMFIMGGLIVIPFLLIMSCARIVSTILRVPSTGGIQKAFSTCGPHLSVSLFYGTIGLYLCLTNHNTVKDVTMAMVMTGVTHMLNPFYISLRNRDMRGNPQSLQHKENFFVKIVIVGILPLL**
- **A two nucleotide frame shift however results in a functional protein with the following sequence:**
- **MPMYLCLSNLSFSDLCFSSVTMPKLLONMOSQNPSPFADCLAQMYFHLFYGVLESFLLVVMAYHCYVAICFPLHYTTIMSPKCLGLLTLWLLTAHATLHLLMARLSFCAENVIPHFCDTSTLLK LACSNTQVNGWVMMFIMGGLIVIPFLLIMSCARIVSTILRVPSTGGIQKAFSTCGPHLSVSLFYGTIGLYLCLTNHNTVKDVTMAMVMTGVTHMLNPFYISLRNRDMRGNPQSLQHKENFFVKIVIVGILPLL**

- **OR17-210 is an Atypical Olfactory Receptor**
- **OR17-210 begins with MPMY---**. This sequence PMY is strongly conserved in most ORs. This sequence typically marks the beginning of the second transmembrane region. Hidden Markov Models² have predicted that in OR17-210, this region is not a TM3. Furthermore, an HA-epitope tag experiment revealed this region of the protein to be extra-cellular.



- What is typically helix 3 in ORs is helix 1 in OR17-210. This TM has the MAYD(E)RY motif, which marks the intracellular side and (part of intracellular loop 2) of TM3. The directionality of this TM1 is extracellular to intracellular. This correctly positions the DRY region of the TM intracellularly—where structural changes following activation may be necessary for signal transduction in GPCRs⁴
- This allows only five typically observed in TMs in OR17-210. HMM strongly predicts that the cDNA sequence has an additional TM helix in the long C-terminus following what would be the seventh TM in most OR sequences. We call this the 7' TM. OR17-210 has a homolog in chimpanzee with greater than 95% sequence similarity. A BLAST search of the 7' sequence, "FVFKI VIVGILPLLN LVGVVKLI" does not return any matches in other ORs, GPCRs or any other protein sequence in GENBANK.
- TM 7' can then occupy either the position of the missing TM1 or TM2 in order to maintain the TM scaffold and protect the ligand and the binding pocket from the surrounding lipid layer
- If one follows the progression of N-terminus-TM1-IC1-TM2-EC1-TM3 .. etc, the C-terminus of this receptor is extra-cellular



Frame shifts, stop and start codons

- <http://ca.expasy.org/tools/dna.html>
- http://bioportal.weizmann.ac.il/HORDE/search_horde.html
- <http://bip.weizmann.ac.il/cgi-bin/HORDE/showGene.pl?key=symbol&value=OR1E3P>
- TMHMM--
<http://www.cbs.dtu.dk/services/TMHMM/>

Problems during Protein Identification

- No sequence in database --- nothing to correlate with
- Problems with entries in database: human errors in entering information (typographical errors and curation); sequencing errors; errors during transcription
- Modifications in large proteins: degradation, oxidation of methionine, deamidation of N and Q, remember glycosylations, phosphorylations, and acetylations
<http://www.unimod.org/> lists the possible modifications that can occur

Sources for Protein Identification

- [EBI, European Bioinformatics Institute, EBI Download site](#) "The EBI is a centre for research and services in bioinformatics. The Institute manages databases of biological data including nucleic acid, protein sequences and macromolecular structures."
 - [Expasy](#) "The ExpASy (**Expert Protein Analysis System**) [proteomics](#) server of the [Swiss Institute of Bioinformatics](#) (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE Databases, Tools and Software Packages.
 - [INFOBIOGEN, DBCAT, The Public Catalog of Databases](#)
 - [NCBI, The National Center for Biotechnology Information](#) "provides an integrated approach to the use of gene and protein sequence information" [Databases and Tools](#)
 - [PIR Protein Information Resource](#) "An integrated public resource of protein informatics to support genomic and proteomic research and scientific discovery." Located at Georgetown University.
 - [Plant Genome Database \(PlantGDB\)](#) Resource for Plant Comparative Genomics

Protein Information Resource-- <http://pir.georgetown.edu/>

The screenshot shows the homepage of the Protein Information Resource (PIR) website. The header includes the PIR logo and navigation links: About PIR, Databases, Search/Analysis, Download, and Support. Below the header, the main heading reads "INTEGRATED PROTEIN INFORMATICS RESOURCE FOR GENOMIC AND PROTEOMIC RESEARCH".

The main content area features several key resources:

- UniProt**: The Universal Protein Resource (UniProt) provides the scientific community with a single, centralized, authoritative resource for protein sequences and functional information. It includes UniProtKB, UniProt, and UniParc. The current release is 14.7.
- PIRSF** (Protein Family Classification System): Offers classification reflecting evolutionary relationships of full-length proteins, functional site and protein name rules, and a sample family report.
- iProClass** (Integrated Protein Classification): Provides value-added reports for UniProtKB and unique UniParc proteins, functional analysis and protein ID mapping, and a sample protein report.
- iProLINK** (Integrated Protein & Knowledge): Serves as a source for text mining and ontology development, featuring a BLAST-E text mining tool, BioThesaurus, and Eoscan Ontology, and bibliography mapping.

At the bottom, there are sections for "OTHER RESOURCE" (listing Proteomics: NIAD Biodefense Proteomics Admin. Center and PIR Grid-Enablement Data node on NCTX LABS), "PEPTIDE SEARCH" (with a database dropdown set to UniProtKB and a search input field for "one single letter amino acid code"), and "TEXT SEARCH" (with a database dropdown set to iProClass and a search input field).

Footer information includes the Bioinformatics Graduate Program (M.S. program at Georgetown University), which is now recruiting new students for Spring & Fall of 2009, with scholarships available.