

Web-based Bioinformatics Applications in Proteomics

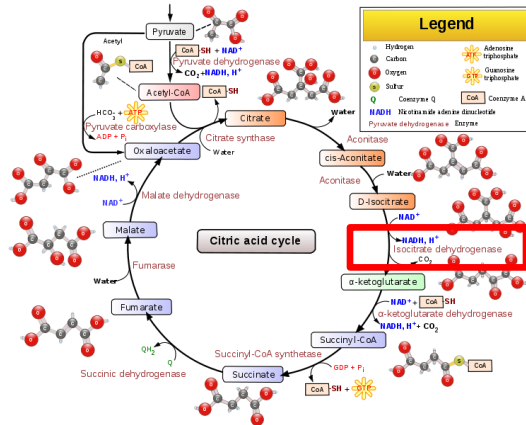
Chiquito Crasto
chiquito@uab.edu
January 28, 2011

Philosophical underpinnings ...

- Bioinformatics is here to stay—simply because computers are part of everyday life. This is not going to change in the near or distant future
- Students, researchers, etc., will be better served embracing bioinformatics ideas even if they do not necessarily want to pursue bioinformatics-driven careers, and opt to be “bench” scientists
- There is significant tool development that will allow scientists to access these to enhance their research (data-analysis, information dissemination, etc.) without having to recourse to collaborations with bioinformatics specialists—unless if specific tools have to be developed
- One should not ignore the intellectualism that goes into conceptualizing and developing tools
- It makes sense then to be able to access and understand how to use these tools

Theme of the today's class—web-based proteomics applications

- **Isocitrate dehydrogenase** ([EC 1.1.1.42](#)) and ([EC 1.1.1.41](#)), also known as **IDH**, is an enzyme that participates in the citric acid cycle. It catalyzes the third step of the cycle: the oxidative decarboxylation of isocitrate, producing alpha-ketoglutarate (α -ketoglutarate) and CO_2 while converting NAD^+ to NADH .



http://en.wikipedia.org/wiki/File:Citric_acid_cycle_with_aconitate_2.svg

NCBI (National Center for Biotechnology Information)

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

The screenshot shows the NCBI website homepage. The top navigation bar includes "NCBI", "Resources", and "How To". Below this is a search bar with "All Databases" selected. The main content area is divided into several sections:

- NCBI Home:** A vertical menu on the left with an arrow pointing to "All Resources". Other items include Site Map (A-Z), Chemicals & Bioassays, Data & Software, DNA & RNA, Domains & Structures, Genes & Expression, Genetics & Medicine, Genomes & Maps, Homology, Literature, Proteins, Sequence Analysis, Taxonomy, Training & Tutorials, and Variation.
- Welcome to NCBI:** A central section with a welcome message and links for "About the NCBI | Mission | Organization | Research | RSS Feeds".
- Get Started:** A section with links for "Tools: Analyze data using NCBI software", "Downloads: Get NCBI data or software", "How-To's: Learn how to accomplish specific tasks at NCBI", and "Submissions: Submit data to GenBank or other NCBI databases".
- Popular Resources:** A list of resources including BLAST, Bookshelf, Gene, Genome, Nucleotide, OMIM, Protein, PubChem, PubMed, PubMed Central, and SNP.
- Genotypes and Phenotypes:** A section with a sub-header "Data from Genome Wide Association studies that link genes and diseases. See study variables, protocols, and analysis." and a small image of a book.
- NCBI News:** A section with news items, including "NCBI Discovery Workshop: A Practical Hands-On Course" (18 Jan 2011) and "NAR's 2011 Database Issue is out with 9 NCBI-Authored Papers" (05 Jan 2011).

Selected Proteomics Applications through NCBI

- GenBank—resource for genes
- Pubmed
- Computational Resources from NCBI's Structure Group
- Conserved Domain Database (CDD)
- Peptidome
- Protein Clusters
- Protein Database
- Structure (Molecular Modeling Database)

Genbank (Search Nucleotide)

The screenshot shows the NCBI GenBank Overview page. At the top, there is a navigation bar with links for PubMed, Entrez, BLAST, OMIM, Books, Taxonomy, and Structure. Below this is a search bar with the text "Search Nucleotide" and a dropdown menu set to "Nucleotide". The search term "isocitrate dehydrogenase" is entered, and a "Go" button is visible. On the left side, there is a vertical menu with links for NCBI Home, NCBI Site Map, Submit to GenBank, Submit an update, Search GenBank, GenBank and RefSeq, a comparison, and BLAST. The main content area is titled "What is GenBank?" and contains the following text:

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2008 Jan 36(Database issue):D25-30). There are approximately 106,533,156,756 bases in 108,431,692 sequence records in the traditional GenBank divisions and 148,165,117,763 bases in 48,443,067 sequence records in the WGS division as of August 2009.

The complete [release notes](#) for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

An example of a GenBank [record](#) may be viewed for a *Saccharomyces cerevisiae* gene.

In The News: 2009 H1N1 Flu Virus (Swine Flu)

The Centers for Disease Control and Prevention and other health officials are actively tracking the recent emergence of human cases of swine influenza A (H1N1) virus infection. Influenza A virus sequences from patients affected by this strain are being submitted to GenBank and can be accessed through the [NCBI Flu Resource](#).

► NLM/NCBI 2009 H1N1 Flu Resources:

- Newest [2009 H1N1 influenza A sequences](#)
- Citations [recently added](#) to PubMed
- [MedlinePlus](#) (consumer health information)
- [Enviro-Health Links](#)

At the bottom right, there is a "FLU.GOV" logo with the text "Know what to do about the flu." and a "VISIT FLU.GOV" button. Below the logo is a "SHARE THIS WISSET" button.

Nucleotide-Genbank's gene repository

NCBI Resources | How To | My NCBI Sign In

Nucleotide Alphabet of Life

Search: Nucleotide | Save search | Limits | Advanced search | Help

isocitrate dehydrogenase and Human | Search | Clear

Display Settings: Summary, 20 per page, Sorted by Default order | Send to

Found 1702 nucleotide sequences. Nucleotide (1227) EST (475)

This search in Gene shows 38 results, including:

- IGDH3** (Homo sapiens) isocitrate dehydrogenase 3 (NAD+) beta
- IGDH1** (Homo sapiens) isocitrate dehydrogenase 1 (NADP+), soluble
- IGDH2** (Homo sapiens) isocitrate dehydrogenase 2 (NADP+), mitochondrial

Results: 1 to 20 of 1227

1. **Streptococcus anginosus 1_2_62CV genomic scaffold supercont1.1 whole genome shotgun sequence**
1,137,758 bp linear DNA
GL63615.1 | Gi:30004214
[GenBank](#) [FASTA](#) [Graphical](#)
2. **Cocoribacillus sp. 29_1 genomic scaffold supercont1.1 whole genome shotgun sequence**
953,248 bp linear DNA
GL63677.1 | Gi:30004207
[GenBank](#) [FASTA](#) [Graphical](#)
3. **Mus musculus lactate dehydrogenase A (Ldha), transcript variant 2, mRNA**
1,854 bp linear mRNA
NM_001136069.2 | Gi:257743038
[GenBank](#) [FASTA](#) [Graphical](#) [Related Sequences](#)
4. **Homo sapiens isocitrate dehydrogenase 2 (NADP+), mitochondrial (IDH2), nuclear gene encoding mitochondrial protein, mRNA**
1,740 bp linear mRNA
NM_002168.2 | Gi:28178831
[GenBank](#) [FASTA](#) [Graphical](#) [Related Sequences](#)

Filter your results: All (1227)

- Bacteria (801)
- NCBI C. GenBank (696)
- mRNA (158)
- RefSeq (331)

Top Organisms [Drop]

- Homo sapiens (133)
- Coxiella burnetii (27)
- Mus musculus (23)
- Leishmania infantum (18)
- synthetic construct (18)
- All other taxa (1072)
- More...

Find related data

Database: Select

Find items

Search details

isocitrate dehydrogenase[All Fields] AND ("Homo sapiens"[Organism] OR Rnss[All Fields])

Accession Number

A Nucleotide Entry in Genbank

NCBI Resources | How To | My NCBI Sign In

Nucleotide Alphabet of Life

Search: Nucleotide | Limit | Advanced search | Help

isocitrate dehydrogenase and Human | Search | Clear

Display Settings: Summary, 20 per page, Sorted by Default order | Send to

Found 1702 nucleotide sequences. Nucleotide (1227) EST (475)

This search in Gene shows 38 results, including:

- IGDH3** (Homo sapiens) isocitrate dehydrogenase 3 (NAD+) beta
- IGDH1** (Homo sapiens) isocitrate dehydrogenase 1 (NADP+), soluble
- IGDH2** (Homo sapiens) isocitrate dehydrogenase 2 (NADP+), mitochondrial

Results: 1 to 20 of 1227

1. **Streptococcus anginosus 1_2_62CV genomic scaffold supercont1.1 whole genome shotgun sequence**
1,137,758 bp linear DNA
GL63615.1 | Gi:30004214
[GenBank](#) [FASTA](#) [Graphical](#)
2. **Cocoribacillus sp. 29_1 genomic scaffold supercont1.1 whole genome shotgun sequence**
953,248 bp linear DNA
GL63677.1 | Gi:30004207
[GenBank](#) [FASTA](#) [Graphical](#)
3. **Mus musculus lactate dehydrogenase A (Ldha), transcript variant 2, mRNA**
1,854 bp linear mRNA
NM_001136069.2 | Gi:257743038
[GenBank](#) [FASTA](#) [Graphical](#) [Related Sequences](#)
4. **Homo sapiens isocitrate dehydrogenase 2 (NADP+), mitochondrial (IDH2), nuclear gene encoding mitochondrial protein, mRNA**
1,740 bp linear mRNA
NM_002168.2 | Gi:28178831
[GenBank](#) [FASTA](#) [Graphical](#) [Related Sequences](#)

Filter your results: All (1227)

- Bacteria (801)
- NCBI C. GenBank (696)
- mRNA (158)
- RefSeq (331)

Top Organisms [Drop]

- Homo sapiens (133)
- Coxiella burnetii (27)
- Mus musculus (23)
- Leishmania infantum (18)
- synthetic construct (18)
- All other taxa (1072)
- More...

Find related data

Database: Select

Find items

Search details

isocitrate dehydrogenase[All Fields] AND ("Homo sapiens"[Organism] OR Rnss[All Fields])

U01112 Homo sapiens NM_002168.2
IDH2 CDS

LOCUS U01112 Homo sapiens 1740 bp mRNA linear PRE 21-08-2001
DEFINITION Homo sapiens isocitrate dehydrogenase 2 (IDH2), mitochondrial
FEATURES
ORIGIN
//

ORIGIN
1 cccagctgag ccagcagcag cagcagcagc aggagcagc agcagcagc cctctccagc
61 cctcctcctc cctcctcctc ctcagcagc agcagcagc agcagcagc cctcctcctc
121 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
181 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
241 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
301 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
361 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
421 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
481 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
541 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
601 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
661 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
721 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
781 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
841 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
901 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
961 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1021 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1081 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1141 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1201 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1261 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1321 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1381 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1441 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1501 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1561 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1621 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
1681 agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc agcagcagc
//

Protein Sequence

Links to Pubmed

Gene Sequence

Protein Sequence in Genbank (isocitrate dehydrogenase)

The screenshot shows the NCBI Protein search interface. The search query is 'isocitrate dehydrogenase and Human'. The results page lists 603 entries, showing the first four. Entry 1 is A00565.1 (419 aa protein) from Homo sapiens. Entry 2 is YP_251488.1 (742 aa protein) from Corynebacterium jeikeium. Entry 3 is YP_763190.1 (747 aa protein) from Corynebacterium jeikeium. Entry 4 is CA67870.1 (742 aa protein) from Corynebacterium jeikeium. On the right, a detailed view of the protein sequence is shown for A00565.1, including annotations for the protein name, accession number, and amino acid sequence. The sequence is: MDSK... IISK. Below the sequence is a reference to a PubMed article (PMID: 1581198) titled 'Expression of human mitochondrial NADP-dependent isocitrate dehydrogenase during lymphocyte activation'. The reference text includes authors, journal information, and a short abstract.

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.

Pubmed—repository of biomedical abstracts

The screenshot shows the PubMed website homepage. The search bar contains the query 'isocitrate dehydrogenase'. Below the search bar is a banner for PubMed, stating that PubMed comprises more than 20 million citations for biomedical literature from MEDLINE, life science journals, and online books. Below the banner are three columns of links: 'Using PubMed' (with links for Quick Start Guide, Full-Text Articles, PubMed FAQs, and PubMed Tutorials), 'PubMed Tools' (with links for Single Citation Matcher, Batch Citation Matcher, Clinical Queries, and Topic-Specific Queries), and 'More Resources' (with links for MeSH Database, Journals in NCBI Databases, Clinical Trials, E-Libraries, and LinkOut).

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

NCBI Resources How To

PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

Search: PubMed Limits Advanced search Help

isocitrate dehydrogenase Search Clear

Display Settings: Abstract Send to

Biochem Biophys Res Commun, 2011 Jan 22. [Epub ahead of print]

Ataxia telangiectasia mutated influences cytochrome c oxidase activity.

Patel AY, Macdonald TM, Spears LD, Ching JK, Fisher JS.
Department of Biology, Saint Louis University, St. Louis, MO 63103, USA.

Abstract
Cells lacking ataxia telangiectasia mutated (ATM) have impaired mitochondrial function. Furthermore, mammalian cells lacking ATM have increased levels of reactive oxygen species (ROS) as well as mitochondrial DNA (mtDNA) deletions in the region encoding for cytochrome c oxidase (COX). We hypothesized that ATM specifically influences COX activity in skeletal muscle. COX activity was ~40% lower in tibialis anterior from ATM-deficient mice than for wild-type mice (P<0.01, n=9/group). However, there were no ATM-related differences in activity of succinate dehydrogenase, isocitrate dehydrogenase, alpha-ketoglutarate dehydrogenase, mitochondrial glycerol 3-phosphate dehydrogenase, or complex III. Incubation of wild-type extensor digitorum longus muscles for 1 h with the ATM inhibitor KU55933 caused a ~50% reduction (P<0.05, n=5/group) in COX activity compared to muscles incubated with vehicle alone. Among the control muscles and muscles treated with the ATM inhibitor, COX activity was correlated (r=0.61, P<0.05) with activity of glucose 6-phosphate dehydrogenase, a key determinant of antioxidant defense through production of NADPH. Overall, the findings suggest that ATM has a protective role for COX activity.

Copyright © 2011. Published by Elsevier Inc.

PMID: 21266166 [PubMed - as supplied by publisher]

Computational Resources from
NCBI's Structure Group
<http://www.ncbi.nlm.nih.gov/Structure/index.shtml>

Three-dimensional structure views in Genbank-- STRUCTURE

The screenshot shows the NCBI Structure database search results for the query 'isopropylmalate dehydrogenase'. The search results are displayed in a list format, showing the first four items. Each item includes a 3D structure visualization, a checkbox, an accession number, and a brief description.

Item	Accession	Description
1:	2V42	Structure Of Isopropylmalate Dehydrogenase From Thermus Thermophilus - Complex With Nadh And Mn [Oxidoreductase, Ec: 1.1.1.85] Taxonomy: Thermus thermophilus Proteins: 4, Chemicals: 3 modified: 2011/01/20, MMDB ID: 87988
2:	2V41	Structure Of Isopropylmalate Dehydrogenase From Thermus Thermophilus - Complex With Ipm And Mn [Oxidoreductase, Ec: 1.1.1.85] Taxonomy: Thermus thermophilus Proteins: 2, Chemicals: 3 modified: 2011/01/20, MMDB ID: 87987
3:	2V40	Structure Of Isopropylmalate Dehydrogenase From Thermus Thermophilus - Complex With Mn [Oxidoreductase, Ec: 1.1.1.85] Taxonomy: Thermus thermophilus Proteins: 2, Chemicals: 2 modified: 2011/01/20, MMDB ID: 87986
4:	2V3Z	Structure Of Isopropylmalate Dehydrogenase From Thermus Thermophilus - Apo Enzyme [Oxidoreductase, Ec: 1.1.1.85] Taxonomy: Thermus thermophilus Proteins: 1, Chemicals: 4 modified: 2011/01/20, MMDB ID: 87985

Structure of Actin—Genbank Structure View

The screenshot shows the 'Structure Summary' page for Actin (PDB ID: 2CM1) in the NCBI Structure database. The page includes a 3D structure visualization, a description, deposition information, taxonomy, and related structure information. An arrow points from the 'PDB ID: 2CM1' link to the text 'Link to Protein Databank'. Another arrow points from the 'Structure View in RasMol' button to the text 'Visualization software'.

MMDB ID: 86995 | **PDB ID:** 2CM1 | **Search** | PDB or MMDB ID

Description: Crystal Structure Of Mouse Cytosolic Isocitrate Dehydrogenase.

Deposition: Cho HJ, Kim KJ, Huh JY, Kang BS, 2006/5/10

Taxonomy: Mus musculus

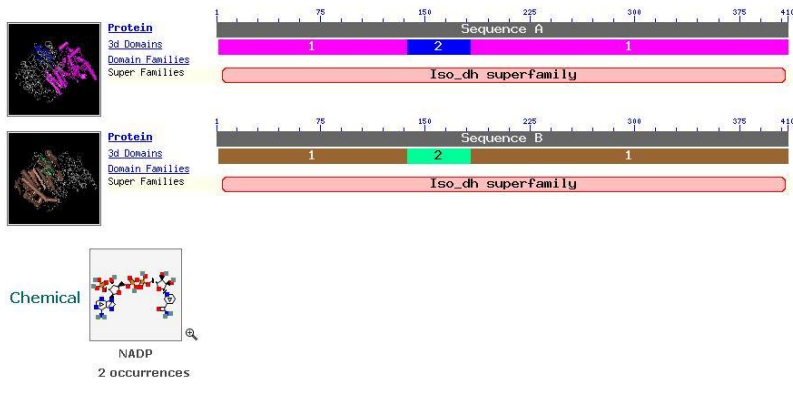
Related Structure: VAST

Structure View in RasMol ← Visualization software

Tasks: Display | Drawing: All Atoms | Download Cn3D | View Cn3D Tutorial

Molecular components in the MMDB 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.

Structure of Domains in Genbank



List of domains related to
or associated with
Isocitrate Dehydrogenase

Conserved domain database (CDD) in Genbank

The screenshot shows the NCBI Conserved Domains Database interface. At the top, there is a search bar with the text 'Conserved Domains' and a search button. Below the search bar, there are navigation tabs for 'All Databases', 'PubMed', 'Gene', 'HomoloGene', 'Protein', 'Protein Clusters', 'Structure', 'PubChem', and 'BioSys'. The 'Protein' tab is selected. Below the search bar, there are buttons for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The main content area is titled 'Conserved Domains Database' and contains a section 'Hints on Finding a Conserved Domain' with a list of search tips. Below this, there is a section 'About the Database' with a paragraph of text and a small 3D protein structure model. The text in 'About the Database' describes the CDD as a collection of multiple sequence alignments representing conserved domains, including NCBI-curated domains, which use 3D-structure information to explicitly define domain boundaries and provide insights into sequence/structure/function relationships. It also mentions that the data are used for putative functional annotation of protein query sequences based on matches to specific hits (illustrated example) or superfamilies, identification of proteins with similar domain architectures, and protein classification. The Conserved Domains and Protein Classification overview page provides more information about the resources available and how they can be used.

CDD ...

The screenshot shows the NCBI Conserved Domains search interface. The search term is 'isocitrate dehydrogenase'. The results are displayed in a list format, showing items 1 through 5. Item 1 is 'cd01891' (AceK isocitrate dehydrogenase kinase/phosphatase). Item 2 is 'pfam06316' (AceK isocitrate dehydrogenase kinase/phosphatase). Item 3 is 'pfam03971' (IDH Monomeric isocitrate dehydrogenase). Item 4 is 'TIGR02924' (ICDH_alpha isocitrate dehydrogenase). Item 5 is 'TIGR00178' (monomer_idh isocitrate dehydrogenase, NADP-dependent, monomeric type). A small 3D protein structure is visible next to item 5.

CDD...

The screenshot shows the detailed view of the 'TIGR02924: ICDH_alpha' domain. It includes a description: 'isocitrate dehydrogenase. This family of mainly alphaproteobacterial enzymes is a member of the isocitrate/isopropylmalate dehydrogenase superfamily described by pfam00180. Every member of the seed of this model appears to have a TCA cycle lacking only a determined isocitrate dehydrogenase. The precise identity of the cofactor (NADH -- 1.1.1.41 vs. NADPH -- 1.1.1.42) is unclear.' Below the description are sections for 'Links', 'Statistics', and 'Structure'. The 'Structure' section shows a sequence alignment with a format of 'Hypertext', row display of 'All 6 rows', and type selection of 'top listed sequences'. The alignment shows sequence identity between the domain and several protein sequences, with positions 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000.

CDD ... if structure is available

NCBI Conserved Domains

TIGR00178: monomer_idh

isocitrate dehydrogenase, NADP-dependent, monomeric type

The monomeric type of isocitrate dehydrogenase has been found so far in a small number of species, including *Azotobacter vinelandii*, *Commaebacterium glutamicum*, *Rhodospirillum rubrum*, and *Neisseria meningitidis*. It is NADP-specific.

PubMed References

Cloning, sequence analysis, expression, and inactivation of the *Commaebacterium glutamicum icd* gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme. *J. Bacteriol.* 1995 Feb; 177(3):774-782.

TIGR00178 is classified as a model that may span more than one domain.
TIGR00178 is not assigned to any domain superfamily.

Sequence Alignment

Reformat: Format: Hypertext Raw Display: All 5 rows Color Bits: 2.0 bit Type Selection: top listed sequences

```

          10      20      30      40      50      60      70      80
117W_A  1  RDTP--RIITLTDSEAPALATYSLLP I IKAFTGSSGIAVTRD ISLAGRLATFPEVLTQKISDGLAELGELATFPA 78
gi|1708408 2  RDTP--RIITLTDSEAPALATYSLLP I IKAFTGSSGIAVTRD ISLAGRLATFPEVLTQKISDGLAELGELATFPA 78
gi|1708389 1  ----KRIITVTDSEAPALATYSLLP I IKAFTGSSGIAVTRD ISLAGRLATFPEVLTQKISDGLAELGELATFPA 76
gi|15607208 1  RDIA@TIITLTDSEAPALATYSLLP I IKAFTGSSGIAVTRD ISLAGRLATFPEVLTQKISDGLAELGELATFPA 80
gi|217188 1  RDTP--RIITLTDSEAPALATYSLLP I IKAFTGSSGIAVTRD ISLAGRLATFPEVLTQKISDGLAELGELATFPA 80
          90     100     110     120     130     140     150     160
117W_A  79  NIILKPHISAVPQLAAIKELQQVYKLPDPEEFKTDTEKQVARTKIGQAVNPFVLEQNSRRAPLSDVIVAKES 158
gi|1708408 79  NIILKPHISAVPQLAAIKELQQVYKLPDPEEFKTDTEKQVARTKIGQAVNPFVLEQNSRRAPLSDVIVAKES 158
gi|1708389 77  NIILKPHISAVPQLAAIKELQQVYKLPDPEEFKTDTEKQVARTKIGQAVNPFVLEQNSRRAPLSDVIVAKES 156
gi|15607208 81  NIILKPHISAVPQLAAIKELQQVYKLPDPEEFKTDTEKQVARTKIGQAVNPFVLEQNSRRAPLSDVIVAKES 160
gi|217188 81  NIILKPHISAVPQLAAIKELQQVYKLPDPEEFKTDTEKQVARTKIGQAVNPFVLEQNSRRAPLSDVIVAKES 160
          170     180     190     200     210     220     230     240
117W_A  159  PRKQKASADSESVYARMDGDFPGRKALIQAPQVVEILAEQSSVTKARTVQAGRIISVSDKALRPIILAA 238
gi|1708408 159  PRKQKASADSESVYARMDGDFPGRKALIQAPQVVEILAEQSSVTKARTVQAGRIISVSDKALRPIILAA 238
gi|1708389 157  PRKQKASADSESVYARMDGDFPGRKALIQAPQVVEILAEQSSVTKARTVQAGRIISVSDKALRPIILAA 236
gi|15607208 161  PRKQKASADSESVYARMDGDFPGRKALIQAPQVVEILAEQSSVTKARTVQAGRIISVSDKALRPIILAA 240
gi|217188 161  PRKQKASADSESVYARMDGDFPGRKALIQAPQVVEILAEQSSVTKARTVQAGRIISVSDKALRPIILAA 240
          250     260     270     280     290     300     310     320
    
```

Clustering Proteins in terms of Sequence Similarities--Genbank

NCBI Protein Clusters

Search: Protein Clusters For: Go Clear

Limits: Preview/Index History Clipboard Details

Protein Clusters

About the Database

Welcome to Entrez Protein Clusters. This collection of related protein sequences (clusters) consists of Reference Sequence proteins encoded by complete genomes. The database contains both curated and non-curated clusters. For release-specific information check the [stats page](#).

The Protein Clusters database provides easy access to annotation information, publications, domains, structures, and external links and analysis tools including multiple alignments, phylogenetic trees, and genomic neighborhoods ([ProMap](#)).

Protein Clusters can be searched like any other Entrez database. For more information on how to use Entrez please examine the [Entrez Help Document](#).

A specialized BLAST service is accessible ([Concise Protein BLAST](#)).

Data is available for download via [Protein Clusters FTP](#).

Example Searches

- all clusters with ribosomal protein as the curated name
"ribosomal protein"[protein_name]
- all clusters that are encoded by chloroplasts
"source:chloroplast"[all_fields]

Check the [limits page](#) and the [help document](#) for more information.

Clustering Proteins in terms of Sequence Similarities--Genbank

PTZ00435
isocitrate dehydrogenase
Gene name: **None**

(Curated - Fructosebait)

▼ Cluster Info

ID: **2743117**

Total proteins: **31**

Conserved in: **Eukaryota**

Total genera: **12**

Total organisms: **29**

Putative Paralogs: **0**

Publications: **20**

▼ Cluster Tools

Show detailed alignment

Build tree

▼ Cross references

EC Number: **1.1.1.42**

KEGG KO: **K00031**

InterPro: **3**

TIGRFAM: **TIGR00127**

Domain(s): **c00445(isc_dh)**

► Entrez Links

Catalyzes the oxidative decarboxylation of isocitrate into alpha-ketoglutarate, dependent on NADP+. Cytoplasmic localization.

Domain description: **Isocitrate isopropylmalate dehydrogenase**

BRITe hierarchy:

Metabolism; Carbohydrate Metabolism; Citrate cycle (TCA cycle)

Energy Metabolism; Reductive carboxylate cycle (CO2 fixation)

Metabolism of Other Amino Acids; Glutathione metabolism

► Publications by categories (only one publication per category is shown) (Show all 20)

- Curated [3]** | Influence of compartmental localization on the function of yeast NADP+-specific isocitrate dehydrogenases. *Arch Biochem Biophys* 2004 Mar 15 [more...](#)
- SwissProt [2]** | Proteomics fingerprinting of phagosome maturation and evidence for the role of a Galpha during uptake. *Mol Cell Proteomics* 2006 Dec [more...](#)
- CDD [15]** | The high-resolution Structure of LeuR (Rv2995c) from *Mycobacterium tuberculosis*. *J Mol Biol* 2005 Feb 11 [more...](#)

Organism (click)	Protein name	Prev. Cluster	Accession	Next Cluster	Locus_tag	Length	UniProt	Blink	Alignment
a.Eukaryota									
<input type="checkbox"/> <i>Babesia bovis</i> T2B6	isocitrate dehydrogenase, NADP-dependent family protein		CL522434767 XP_001618855	CL522434773	BBOV_0007260	440aa	A7AP30	◆	All sequences have the same domain structure (Eukaryot)
<input type="checkbox"/> <i>Cryptosporidium parvum</i>	isocitrate dehydrogenase		CL522436108 XP_002141313	CL522435904	CMU_033490	412aa	B6AF38	◆	
<input type="checkbox"/> <i>Dicostelium</i>	isocitrate dehydrogenase (NADP+)		CL522430407 XP_545263	P1700498	DBB_00272208	412aa	P75J83	◆	
<input type="checkbox"/> <i>Dicostelium</i>	isocitrate dehydrogenase (NADP+)		P1700498 XP_545264	P1700227	DBB_00272210	428aa	P75J82	◆	
<input type="checkbox"/> <i>Leishmania</i>	isocitrate dehydrogenase		CL522456728 XP_001562802	CL522456727	LiM10_V2_0110	436aa	AH6112	◆	
<input type="checkbox"/> <i>Leishmania infantum</i>	isocitrate dehydrogenase		CL522456728 XP_001496380	CL522456727	LiM10_0610	436aa	AH6106	◆	
<input type="checkbox"/> <i>Leishmania major</i>	isocitrate dehydrogenase		CL522456728 XP_001496381	CL522456727	LiM10_0200	436aa	P49407	◆	
<input type="checkbox"/> <i>Neisseria meningitidis</i>	isocitrate dehydrogenase (NADP-dependent)		CL522737184 XP_000872653	CL522736938	NA608894T_82731	390aa		◆	
<input type="checkbox"/> <i>Paraburkholderia</i>	haemolysal protein		CL522448920 XP_001428548	CL522448918	oSPAT700029791001	411aa	A06K13	◆	

Peptidome

NCBI Resources (2) How To (2)
My NCBI Sign In

Peptidome
NCBI Peptide Data Resource

Search: Peptidome Limits Advanced search Help

Peptidome

Peptidome is a public repository that archives and freely distributes tandem mass spectrometry peptide and protein identification data generated by the scientific community. Several layers of data are captured to promote understanding of the experiment and analysis of the underlying data.

Using Peptidome

[Browse data](#)

[Submission guidelines](#)

[Submit Data](#)

[FAQ](#)

Other Resources

[Resources](#)

[Downloads \(FTP\)](#)

[Contact Us](#)

Publications and Posters

[Nucleic Acids Res. 2010 Jan](#)

[Nat Biotechnol. 2009 Jul](#)

[ASMS 2010 poster](#)

[ASMS 2010 poster](#)

NCBI Resources (2) How To (2)

Peptidome
NCBI Peptide Data Resource

Search: Peptidome Save search Limits Advanced search Help

Display Settings: Summary [Send To \(2\)](#)

P5E116 **Quantitative mass spectrometry reveals a role for the GTPase Rho in actin organization on the peroxisome membrane** [*Saccharomyces cerevisiae*] [Download](#) [Links](#)

Summary: We have combined classical subcellular fractionation with large-scale quantitative mass spectrometry to identify proteins that enrich specifically with peroxisomes of *Saccharomyces cerevisiae*. In two complementary experiments, isotope-coded affinity tags and tandem mass spectrometry were used to quantify the relative enrichment of proteins during the purification of peroxisomes. Mathematical modeling of the data from 336 quantified proteins led to a prioritized list of 70 candidates whose enrichment scores indicated a high likelihood of them being peroxisomal. more...

Search Engine: SEQUEST

Counts: Proteins: 386 Peptides: 1,119 Spectra: 6,593

Samples: PSM1065: Sample 1: Identification of ICAT labeled peroxisomal yeast proteins with subcellular fractionation
 PSM1066: Sample 2: Identification of ICAT labeled peroxisomal yeast proteins with subcellular fractionation
 PSM1067: Sample 3: Identification of ICAT labeled peroxisomal yeast proteins with subcellular fractionation
 PSM1068: Sample 4: Identification of ICAT labeled peroxisomal yeast proteins with subcellular fractionation

Peptidome ...

Study PSE116

PSE116: Quantitative mass spectrometry reveals a role for the GTPase Rho1p in actin organization on the peroxisome membrane [Jump to downloads](#)

Organism: *Saccharomyces cerevisiae*

Statistics
Samples: View samples (4)
Proteins: 306 **Peptides:** 1119 **Spectra:** 6593

Dates
Release: May 21, 2009 **Deposit:** May 20, 2009
Metadata update: May 20, 2009 **Data update:** May 20, 2009

Contact
Name: Institute for Systems Biology
Organization: Institute for Systems Biology
Email: edutsch@systemsbiology.org
Address: 1441 N 34th St
 Seattle WA 98103-8904

Overall Design
 ICAT labeled four samples of subcellular fractionation of yeast.

Summary
 We have combined classical subcellular fractionation with large-scale quantitative mass spectrometry to identify proteins that enrich specifically with peroxisomes of *Saccharomyces cerevisiae*. In two complementary experiments, isotope-coded affinity tags and tandem mass spectrometry were used to quantify the relative enrichment of proteins during the purification of peroxisomes. Mathematical modeling of the data from 306 quantified proteins led to a prioritized list of 70 candidates whose enrichment scores indicated a high likelihood of them being peroxisomal. Among these proteins, eight novel peroxisome-associated proteins were identified. The top novel peroxisomal candidate was the small GTPase Rho1p. Although Rho1p has been shown to be tethered to membranes of the secretory pathway, we show that it is specifically recruited to peroxisomes upon their induction in a process dependent on its interaction with the peroxisome membrane protein Pex21p. Rho1p regulates the assembly state of actin on the peroxisome membrane, thereby controlling peroxisome membrane dynamics and biogenesis.

Contributors
 Marcello Marelli, Jennifer J Smith, Sunhee Jung, Eugene Yi, Alexey I Nesvizhskii, Rowan H Christmas, Ramsey A Saleem, Yuen-Yi C Tam, Andrei Fagarasanu, David R Goodlett, Rudolf Aebersold, Richard A Rachubinski, John D Atchison

Publications
 Marelli M, Smith JJ, Jung S, Yi E et al. Quantitative mass spectrometry reveals a role for the GTPase Rho1p in actin organization on the peroxisome membrane. *J Cell Biol* 2004 Dec 20;167(4):1099-112. PMID: 15596542

Processing Methods
Search Engine: SEQUEST
Modifications:
Static: ICAT-D (13), ICAT-C (105)
Variable: Oxidation (35)
Platform(s): ESI-ION TRAP

Download Files
[FTP Directory](#)

File	Type	Size
Peptidump.xml	PeptidomeXML	3.76 KB
Marelli_metadata.txt	Metadata	13.46 KB

Peptidome

Accession	Sample Title	Organism	Platform	Prote...	Peptl...	Spec...	Release Date
PSM1065	Sample 1: identification of ICAT labeled peroxisomal yeas...	<i>Saccharomyces cerevisiae</i>	ESHON-TRAP	152	465	3,063	May 21, 2009
PSM1066	Sample 2: identification of ICAT labeled peroxisomal yeas...	<i>Saccharomyces cerevisiae</i>	ESHON-TRAP	240	605	3,698	May 21, 2009
PSM1067	Sample 3: identification of ICAT labeled peroxisomal yeas...	<i>Saccharomyces cerevisiae</i>	ESHON-TRAP	127	195	314	May 21, 2009
PSM1068	Sample 4: identification of ICAT labeled peroxisomal yeas...	<i>Saccharomyces cerevisiae</i>	ESHON-TRAP	195	440	1,510	May 21, 2009

File	Type	Size
Peptidump.xml	PeptidomeXML	1.67 MB
PAe000157_pepXML	Ident	64.47 MB
PAe000157_pro-pep	Results	32.79 KB
012403_04.mzXML.bz2	Spectra	25.12 MB
012403_05.mzXML.bz2	Spectra	19.50 MB
012403_06.mzXML.bz2	Spectra	19.55 MB
012403_07.mzXML.bz2	Spectra	18.20 MB
012403_08.mzXML.bz2	Spectra	19.06 MB
012403_09.mzXML.bz2	Spectra	19.34 MB
012403_10.mzXML.bz2	Spectra	19.09 MB
012403_12.mzXML.bz2	Spectra	23.14 MB
012403_13.mzXML.bz2	Spectra	23.98 MB
012403_14.mzXML.bz2	Spectra	28.71 MB
012403_15.mzXML.bz2	Spectra	28.07 MB
012403_16.mzXML.bz2	Spectra	29.72 MB
012403_17.mzXML.bz2	Spectra	30.36 MB
012403_18.mzXML.bz2	Spectra	30.44 MB

Peptidome ...

Gene	Accession	Organism	Gene	Length	Mass	Cons.	Prot.	Search	Display
IL13L2	FK00011	Homo sapiens	IL13L2	845	92905.05	1%	1	Protein: Full-length: Insert 1 (accession FK00011)	
IL13L1	FK00010	Homo sapiens	IL13L1	844	92917.72	2%	1	Protein: Full-length: Insert 2 (accession FK00010)	
IL13L2C	FK00011	Saccharomyces cerevisiae	IL13L2C	845	93827.24	1%	1	Protein: Full-length: short protein (233) - M24969.1	

Proteins

Protein Name	Length	Mass	Protein	Search	Display
IL13L2	845	92905.05	1	1	1
IL13L1	844	92917.72	1	1	1
IL13L2C	845	93827.24	1	1	1
IL13L2D	845	93827.24	1	1	1
IL13L2E	845	93827.24	1	1	1
IL13L2F	845	93827.24	1	1	1
IL13L2G	845	93827.24	1	1	1
IL13L2H	845	93827.24	1	1	1
IL13L2I	845	93827.24	1	1	1
IL13L2J	845	93827.24	1	1	1
IL13L2K	845	93827.24	1	1	1
IL13L2L	845	93827.24	1	1	1
IL13L2M	845	93827.24	1	1	1
IL13L2N	845	93827.24	1	1	1
IL13L2O	845	93827.24	1	1	1
IL13L2P	845	93827.24	1	1	1
IL13L2Q	845	93827.24	1	1	1
IL13L2R	845	93827.24	1	1	1
IL13L2S	845	93827.24	1	1	1
IL13L2T	845	93827.24	1	1	1
IL13L2U	845	93827.24	1	1	1
IL13L2V	845	93827.24	1	1	1
IL13L2W	845	93827.24	1	1	1
IL13L2X	845	93827.24	1	1	1
IL13L2Y	845	93827.24	1	1	1
IL13L2Z	845	93827.24	1	1	1

Peptides

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

Ensembl Home | Login | Register | BLAST/BLAT | BioMart | Tools | Downloads | Help | More... | Search using Google | isocitrate dehydrogenase

Search Homo sapiens for isocitrate dehydrogenase
 Go

e.g. Human gene BRCA2 or rat X:100000_200000 or coronary heart disease

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 (Log in to customize this list)

- Human** (HGNC07)
- Mouse** (NCBI037)
- Zebrafish** (ZFIN)

All genomes
 - Select a species -
[View full list of all Ensembl species](#)
 Other species are available in [Ensembl FISH](#) and [Ensembl Genomes](#)

Ensembl is a joint project between **Sanger**, **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.

New to Ensembl?
 Did you know you can:

- [Learn how to use Ensembl](#) with our video tutorials and walk-throughs
- [Add custom tracks](#) using our new Control Panel
- [Upload and analyse your data](#) and save it to your Ensembl account
- [Search for a DNA or protein sequence](#) using BLAST or BLAT
- [Fetch only the data you want](#) from our public database, using the 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](#) or [glossary](#)

What's New in Release 60 (8 November 2010)

- New species - Giant Panda** (Giant panda)
- New zebrafish assembly** (Zebrafish)
- New rabbit assembly** (Rabbit)
- New Gene Ontology displays** (all species)
- New navigation options on Region in Detail** (all species)

[More news...](#)

Latest blog posts

- 2011-01-18: [Ensembl Events in February 2011](#)
- 2011-01-19: [New Search - Luceres](#)
- 2011-01-13: [New Ensembl mirror in Asia](#)

[Go to Ensembl blog](#)

Ensembl release 60 - Nov 2010 © (RHS) / EBI | [About Ensembl](#) | [Contact Us](#) | [Help](#)

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

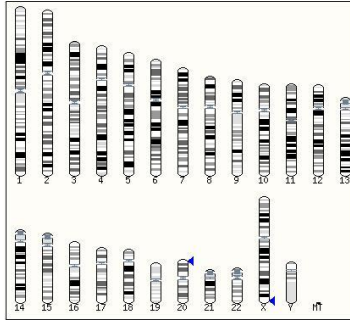
Gene: **ENSM0056000770977**

ISOCITRATE DEHYDROGENASE [NAD] SUBUNIT BETA MITOCHONDRIAL PRECURSOR EC_1.1.1.41 ISOCITRIC DEHYDROGENASE SUBUNIT BETA, NAD + SPECIFIC IC DH SUBUNIT BETA.

« Protein families

HUMAN genes in this family

Ensembl genes containing proteins in family ENSFM0056000770977



Gene ID and Location	Gene Name	Description(if known)
ENSG00000101365 Chromosome 20: 2,944m	IDH3B	isocitrate dehydrogenase 3 (NAD+) beta [Source:HGNC Symbol;Acc:5385]
ENSG00000067829 Chromosome X: 159,05m	IDH3G	isocitrate dehydrogenase 3 (NAD+) gamma [Source:HGNC Symbol;Acc:5386]

Ensembl ...



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>

UniProt has replaced 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](#)

UNIPROT SWISSPROT

The screenshot shows the UniProt website interface. At the top, there is a search bar and navigation links. Below that, a 'WELCOME' message is displayed. A 'NEWS' section highlights the 'UniProt release 2011_01' from January 11, 2011. A 'SITE TOUR' section is also visible, along with a 'PROTEIN SPOTLIGHT' section for January 2011. The UniProt logo is prominently displayed at the bottom left of the page.

The screenshot shows the ExPASy Proteomics Server homepage. At the top, there is a search bar and navigation links. Below that, the 'Swiss-Prot Protein knowledgebase' and 'TrEMBL' are highlighted as a 'Computer-annotated supplement to Swiss-Prot'. A section titled 'The UniProt Knowledgebase consists of' lists the databases. An 'Access to the UniProt Knowledgebase' section provides links to the UniProt website, UniProtKB/TrEMBL, UniProtKB/Swiss-Prot, and UniProtKB/Swiss-Prot entries. The UniProt logo is prominently displayed at the bottom left of the page.

SwissProt—search for Proteins

Search in **Protein Knowledgebase (UniProtKB)** Query **isocitrate dehydrogenase** [Advanced Search »](#)

-25 of 8,133 results for **isocitrate AND dehydrogenase** in **UniProtKB** sorted by **score descending**
 ⚙️ [Browse by taxonomy, keyword, gene ontology, enzyme class or pathway](#) | 📄 [Reduce sequence redundancy to 100%, 90% or 50%](#) | [Download](#)
 Page 1 of 326 | [Next](#) >

Results [Customize](#)

- > Show only reviewed (655) (UniProtKB/Swiss-Prot) or unreviewed (7,576) (UniProtKB/TrEMBL) entries
- > Quote terms: "isocitrate dehydrogenase"
- > Restrict term "isocitrate" to protein family (6,747), gene ontology (4,102), protein name (4,818), web resource (1)
- > Restrict term "dehydrogenase" to protein family (2), gene ontology (6,887), protein name (7,752), web resource (1)
- > Restrict term "isocitrate" to pathway
- > Restrict term "dehydrogenase" to pathway

Accession	Entry name	Status	Protein names	Gene names	Organism	Length
<input type="checkbox"/> O75874	IDHC_HUMAN	★	Isocitrate dehydrogenase [NADP] cytoplasmic	IDH1 PICD	Homo sapiens (Human)	414
<input type="checkbox"/> P08200	IDH_ECOLI	★	Isocitrate dehydrogenase [NADP]	icd icdA icdE b1136 JW1122	Escherichia coli (strain K12)	416
<input type="checkbox"/> P0A9G6	ACEA_ECOLI	★	Isocitrate lyase	aceA icl b4015 JW3975	Escherichia coli (strain K12)	434
<input type="checkbox"/> P28241	IDH2_YEAST	★	Isocitrate dehydrogenase [NAD] subunit 2, mit...	IDH2 YOR136W Q3326 YOR3326W	Saccharomyces cerevisiae (Baker's yeast)	369
<input type="checkbox"/> P11071	ACEK_ECOLI	★	Isocitrate dehydrogenase kinase phosphatase	aceK b4016 JW3976	Escherichia coli (strain K12)	576
<input type="checkbox"/> P33198	IDHP_PIG	★	Isocitrate dehydrogenase [NADP], mitochondria...	IDH2	Sus scrofa (Pig)	421
<input type="checkbox"/> P28834	IDH1_YEAST	★	Isocitrate dehydrogenase [NAD] subunit 1, mit...	IDH1 YNL037C N2690	Saccharomyces cerevisiae (Baker's yeast)	360
<input type="checkbox"/> P39126	IDH_BACSU	★	Isocitrate dehydrogenase [NADP]	icd citC BSU29130	Bacillus subtilis	423
<input type="checkbox"/> Q8LFC0	IDH1_ARATH	★	Isocitrate dehydrogenase [NAD] regulatory sub...	IDH1 A14g35260 F23E12.180	Arabidopsis thaliana (Mouse-ear cress)	367
<input type="checkbox"/> Q43837	IDH3B_HUMAN	★	Isocitrate dehydrogenase [NAD] subunit beta, ...	IDH3B	Homo sapiens (Human)	385


Swissprot-Actin (Results in a Single Page)

The screenshot shows the UniProt entry for Actin (P0A922) from Escherichia coli. The interface is organized into several main sections:

- Search and Navigation:** Top left with search bar and filters.
- General annotation (Consensus):** Middle left, containing protein name, accession number (P0A922), and basic properties like length (375) and molecular weight (41,820).
- Sequence:** Below general annotation, showing the amino acid sequence with a color-coded secondary structure diagram above it.
- Domain Diagrams:** Middle right, showing domain architecture with arrows pointing to detailed domain information.
- Secondary structure:** Bottom left, showing a detailed alpha-helical and beta-strand structure.
- Database Cross-links:** Multiple panels on the right (e.g., InterPro, Pfam, PROSITE) linking to other resources.
- References:** Bottom left, listing scientific publications.

Arrows in the image highlight the flow from the general annotation to the sequence viewer, from domain diagrams to detailed domain data, and from the secondary structure to the domain diagram.

SwissProt-Databases and Feature



Swiss Institute of Bioinformatics

EXPASY Proteomics Server

Databases Tools Services Mirrors About Contact

Search for

You are here: EXPASY.CH > Databases

Proteins

- UniProt Knowledgebase (Swiss-Prot and TrEMBL) - Protein knowledgebase
- Around UniProtKB - links to related databases and portals
 - ENZYME - Enzyme nomenclature database
 - HAMAP - Portal to microbial UniProtKB/Swiss-Prot entries
 - SwissVar - Portal to human diseases and variant information in UniProtKB/Swiss-Prot
 - UniPathway - Metabolic pathways database
 - ViralZone - Portal to viral UniProtKB/Swiss-Prot entries
 - HPI - Human Proteomics Initiative
 - PPAP - Plant Proteome Annotation Project
 - Tox-Prot - Toxin Annotation Project
- PROSITE - Protein families and domains
- Swiss Human Plasma protein dataset - Novartis/Geneprot MicroProt2 dataset from Human Plasma samples [see](#)
- SWISS-MODEL Repository - Automatically generated protein models
- neXtProt - A knowledge platform dedicated to human proteins

Proteomics gel-based

- SWISS-2DPAGE - Two-dimensional polyacrylamide gel electrophoresis
- World-2DPAGE Repository - A public standards-compliant repository for gel-based proteomics data published in the literature
- MIAPEGelDB - A public repository for MIAPE Gel electrophoresis documents
- WORLD-2DPAGE List - Links to other proteomics gel-based databases

Others

- GlycoSuiteDB - a curated and annotated glycan database
- GPSDB - GeneProtein Synonyms Database
- Links to many other molecular biology databases


Latest News

neXtProt - January 13, 2011

GENEVA/LAUSANNE, Switzerland - The first public version of neXtProt, an innovative knowledge platform dedicated to human proteins, is now available. This new resource contains a wealth of high-quality data on all the human proteins that are produced by the 20,000 protein-coding genes found in the human genome. The content of neXtProt will be continuously extended so as to provide many more carefully selected data sets and analysis tools. [\[Read more\]](#)

[\[more news\]](#) [\[SIB news\]](#)

Swiss 2DPAGE



Swiss Institute of Bioinformatics

EXPASY Proteomics Server

Home (search engine)

Search for

You are here: EXPASY.CH > Databases > SWISS-2DPAGE

Search by

- Accession number
- Description, ID, or gene
- Author's name
- Spot ID - serial number
- Identification method
- pI (Mr range)
- Combined fields

Maps

Experimental info

Protein list

Graphical interface

Open Database Interfaces:

Web-2DPAGE Portal

Exclude local DBs

Exclude entries if a mirror

Metadata is visible

SWISS-2DPAGE

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

Enter search keywords:

Limit to: All sites OE ID GN KW

Include external UniProtKB data in search

Sort by: Accession number Protein ID Gene name

Please enter a keyword. This may be any word or partial word appearing in the entry identifier (ID), the description (DE), the gene names (GN) or a UniProtKB/Swiss-Prot keyword (KW). For example, you may type apoA1_human, or just apoA, or APOA or APOA1_HUMAN.

If you give more than one keyword, entries having any keyword will be listed. Please do NOT use any boolean operators (and, or, etc.), nor quotes (!).

[\[Search\]](#) [\[Documents\]](#) [\[Services\]](#) [\[Software\]](#) [\[Related services\]](#) [\[Other databases\]](#) [\[Job openings\]](#)

Access to SWISS-2DPAGE

- by description (any word in the ID, DE, GN and KW lines)
- by accession number (AC lines)
- 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 pI/Mw range
- by experimental identification methods
- by full text search
- retrieve all the protein entries identified on a given reference map
- complex queries (SRS like)
- compute estimated location on reference maps for a user-entered sequence

Services

- Downloading SWISS-2DPAGE by FTP
- Proteomics Core Facility - Get your 2-D Gels performed according to Swiss standards
- 2-D PAGE museum - gels run by trainees during the 2-D PAGE courses

SWISS-2DPAGE documents

- User manual
- Release notes (September 26, 2006)
- FAQ (Frequently Asked Questions about SWISS-2DPAGE)
- **Protocols:**
 - Technical information about 2-D PAGE (IPG's, silver staining, protocols, etc)
 - High performance 2-D gel comparison
- **Figure captions of SWISS-2DPAGE maps available from publications:**
 - Human CSF, ELC, HEPG2, HEPG2SP, LIVER, LYMPHOMA, PLASMA, PLATELET, RBC, U937, CEC, KIDNEY.
 - *Dictyostelium discoideum*, *Escherichia coli*, *Saccharomyces cerevisiae*.

Software

- ImageMaster / Melanie - Software package for 2-D PAGE analysis
- Make2D-DB package ver 2.50 - A package preparing the data and the programs necessary to build a federated 2-DE database on one's own web site.

SwissModel Repository ...

SWISS-MODEL Repository

Modeling Tools Repository Documents

[Repository Query] [Full Text Query]

SWISS-MODEL Repository - Model Details

Model Overview [-]
Click on the bars to get more details about individual Models or experimental structures.

Sequence [-]
UniProt P08200 Isocitrate dehydrogenase [NADP] (IDP)
Escherichia coli (strain K12)
Database: Swiss-Prot (Reviewed) ★

Domain [-]
Model 3D Structure [-]

Based on structure: 1pb1 [SMTL] [RCSB] [PDBE] [SCOP] [CATH]
Sequence identity: 100%
Residue range: 1 to 416

Primary Citation
A new model for protein stereospecificity.
Hevener, A.D., Kochand J., D.E.
Journal: 2009 Nature 463: 614-625
PubMed: 20081887
DOI: 10.1038/3308114a
Search related articles in PubMed
PubMed Abstract:
No abstract available... [Read More & Search PubMed Abstracts]

Molecular Descriptions
Classification: Cofactor-dependent
Molecular Weight: 46274.67
Molecule: Isocitrate dehydrogenase [NADP]
Polymer: 1 Type: polypeptide(s)
Chain: A
EC: 1.1.1.42
Other details: wildtype

Source
Polymer: 1
Scientific Name: Escherichia coli Taxonomy: Escherichia coli
System: Escherichia coli

Related PDB Entries
ID: Details
PDB: 1PBF

Liquid Chemical Component
Identifier: Formula: Name: Interactions: Hide

Biological Assembly
View in Jmol | SpinViewer
Other viewers: Protein Workshop
Biological assembly assigned by authors

Hydrophobic Annotations
To view hydrophobic annotations, please login to your HYDRUS account.

Description Summary
Authors: Hevener, A.D., Kochand J., D.E.
Deposition: 2003-09-14
Release: 2003-06-17
Last Modified: 2009-02-24 (PDB)

Uniref—Clustering of Proteins

Cluster: Isocitrate dehydrogenase [NADP] (50%) ★

Published January 11, 2011

Build on [seed sequence A1WZE1](#) | List component clusters with 100% or 90% identity

[xml](#) | [rdf/xml](#) | [fasta](#) | [tab](#)

Filter Members Sequence Customize order Page 1 of 46 | Next

Filter

1 - 25 of 1,137 members from 837 organisms

Dataset

UniProt (1137)

Taxonomy

...

Filter

Reset

Members Customize

Member	Entry name	Status	Protein names	Organism	Component clusters	Length
<input type="checkbox"/>	09ZN36	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori J99 (Campylobacter pylori J99)	UniRef100_09ZN36 UniRef90_09ZN36	425
<input type="checkbox"/>	B2ZP83	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (Campylobacter pylori)	UniRef100_B2ZP83 UniRef90_G9ZN36	425
<input type="checkbox"/>	B6JPC2	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (strain P12)	UniRef100_B6JPC2 UniRef90_G9ZN36	425
<input type="checkbox"/>	B9XWT5	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori 98-10	UniRef100_B9XWT5 UniRef90_G9ZN36	425
<input type="checkbox"/>	E1Q800	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (strain Cuz20)	UniRef100_E1Q800 UniRef90_G9ZN36	425
<input type="checkbox"/>	Q1CVDD	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (strain HPAG1)	UniRef100_Q1CVDD UniRef90_G9ZN36	425

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 SSDB LIGAND DIGEST

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 newly developed viewers and a cancer map with the traditional KEGG map viewer.

Metabolism map (version 0.1, to be phased out)

Metabolism map viewer

Plant secondary metabolism map

Cancer map

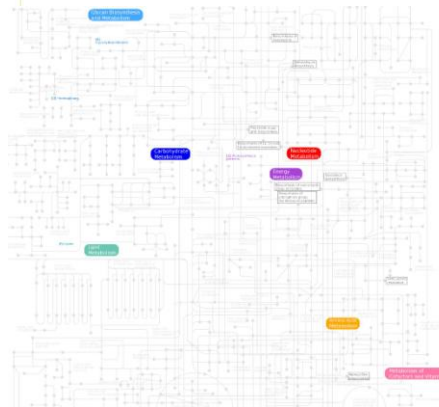
Other global maps are being developed or planned including:

Cell map

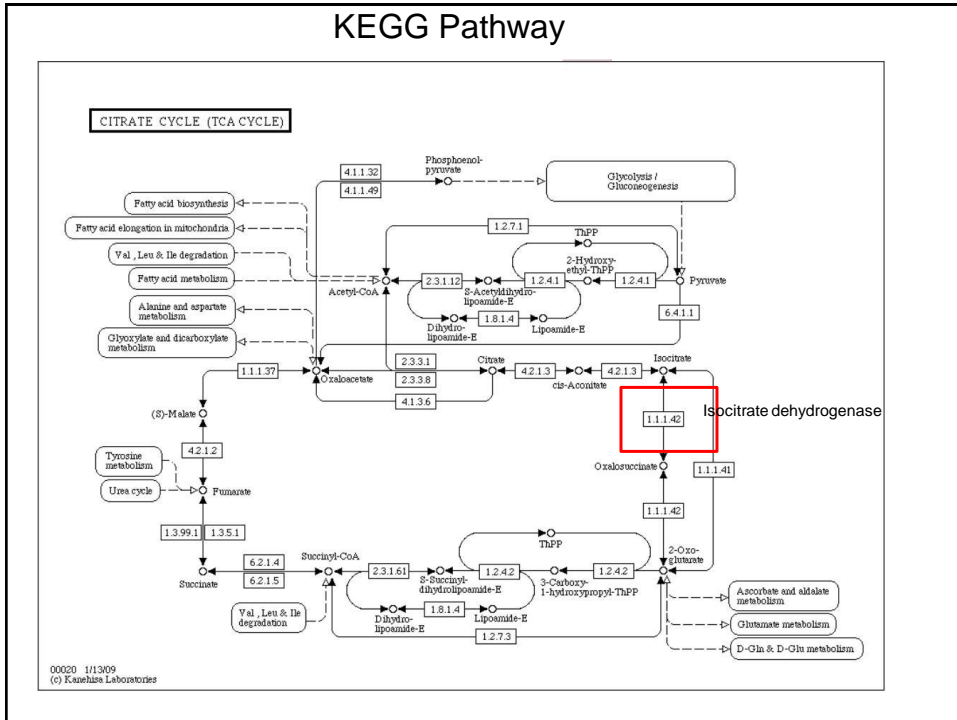
Body map

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, listing branches in the existing maps, named NetElement, and identified by the N number. Each NetElement corresponds to one to several KOs (such as this) in the reference pathway view, or one to several genes (such as this) in an organism-specific view.



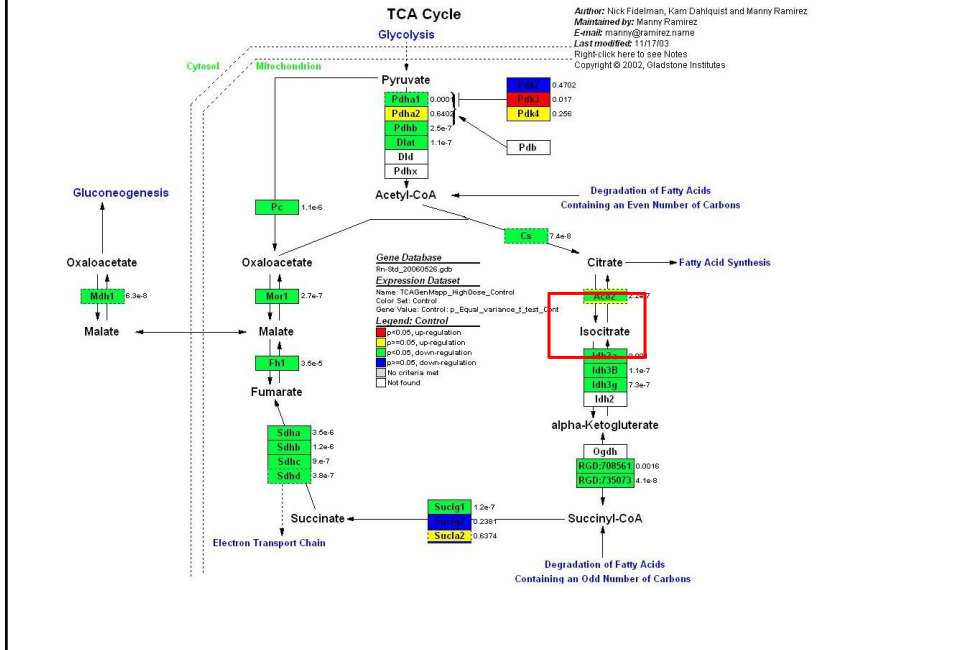
KEGG Pathway



Isocitrate Dehydrogenase in KEGG

KEGG		Homo sapiens (human): 3417		Help
Entry	3417	CDS	H.sapiens	
Gene name	IDH1, IDCD, IDH, IDP, IDPC, PICD			
Definition	isocitrate dehydrogenase I (NADP+), soluble (EC:1.1.1.42)			
Orthology	K00031 isocitrate dehydrogenase [EC:1.1.1.42]			
Pathway	hsa00020 Citrate cycle (TCA cycle) [PATH:hsa00020] hsa00480 Glutathione metabolism hsa01100 Metabolic pathways hsa04146 Peroxisome			
Class	Metabolism; Carbohydrate Metabolism; Citrate cycle (TCA cycle) [PATH:hsa00020] Metabolism; Metabolism of Other Amino Acids; Glutathione metabolism [PATH:hsa00480] Cellular Processes; Transport and Catabolism; Peroxisome [PATH:hsa04146] (BRTE hierarchy)			
SSDB	Ortholog Paralog GFIT			
Motif	Pfam: Iso_dh_PQ2 DUF505 PROSITE: IDH_IMDH Motif			
Other DBs	NCBI-GI: 28178825 NCBI-GeneID: 3417 OMIM: 147700 HGNC: 5382 HPRD: 00984 Ensembl: ENSG00000138413 UniProt: Q75874 Q5P805			
Structure	PDB: 3TMM 1TDL 1T09 3MAP 3MAS 3MAR Thumbnails  Jmol			
Position	2q33.3			

Center for Gene Nutrition Interactions in Cancer --UAB



MASCOT—Protein Identification from Mass Spectrometry Data

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

(MATRIX) INCORPORATED

HOME WHAT'S NEW MASCOT HELP PRODUCTS SUPPORT TRAINING CONTACT 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 [small print](#). If you include results from Mascot in a publication, please cite either [www.matrixscience.com](#) or [Electronshock, OUP \(9\) 353-451 \(1999\) \(abstract\)](#).

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 Novartis, developed by Daniel Pappas and David Perkins when working at the former Imperial Cancer Research Fund, and licensed from its technology transfer subsidiary, Cancer Research Technology.

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

swiss-prot
We are grateful to the Swiss Institute of Bioinformatics for permission to make Swiss-Prot available on this web site for searching with Mascot.

MASCOT Search Results

(MATRIX SCIENCE) Mascot Search Results

Peptide View

MS/MS Fragmentation of **QAGLSYIR**

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

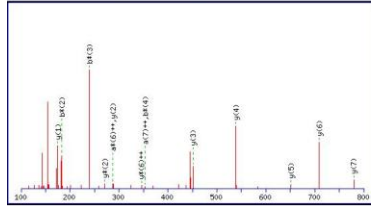
Match to Query 39_906.598884 from(454.306718.24)

File: NanoSET1.wf1_Sample: Complex V Lane-1-1 (sample number 1), Etkon: 23 07 min, Period: 1, Cycle(s): 2586 (Experiment 4)

From data file C:\DOCTUME-1\MASSEP-1\LOCALS-1\Temp\mas2c5.msp

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

Or, [Plot from] 100 to 800 Da



Monoisotopic mass of neutral peptide M(calc): 906.49

Ions Score: 47 Expect: 0.043

Hatches (Bold Red): 14/94 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	288.20	144.61	271.18	136.09	2
8									R	175.12	88.06	158.09	79.55	1

MRM-Path

MRMPath

- Analysis of Protein Mass Fragments from Pathways: Metabolism, Genetic, Environmental, Cellular, Organismal, Human

Presented here is a methodology that allows the user to select individual proteins from the pathways associated with a disease process in a given species. Once the sequence of the protein is recovered, it is subjected to *in silico* digestion with trypsin to determine peptides that are suited to multiple reaction on monitoring. For each fragment, the m/z values of the b- and y- ions are presented (only those with values greater than the doubly charged parent ion are included). For each tryptic peptide, an automated BLAST search is deployed, which results in a list of the highest similarity hits, each with the hits to GENBANK. The resulting data can be exported to a comma-delimited file.

Analysis of Protein Mass Fragments from Pathways

New

@Trypan @Arg-C @Lys-C @Chymotrypsin @Glu-C

• Analysis of Protein Mass Fragments

Presented here is a methodology that allows the user to select individual proteins and perform a tryptic digest *in silico* to determine peptides that are suited to multiple reaction on monitoring. The only input required is the UniProt accession ID for a protein sequence. For each fragment, the m/z values of the b- and y- ions are presented (only those with values greater than the doubly charged parent ion are included). For each tryptic peptide, an automated BLAST search is deployed, which results in a list of the highest similarity hits, each with the hits to GENBANK. The resulting data can be exported to a comma-delimited file.

Protein ID

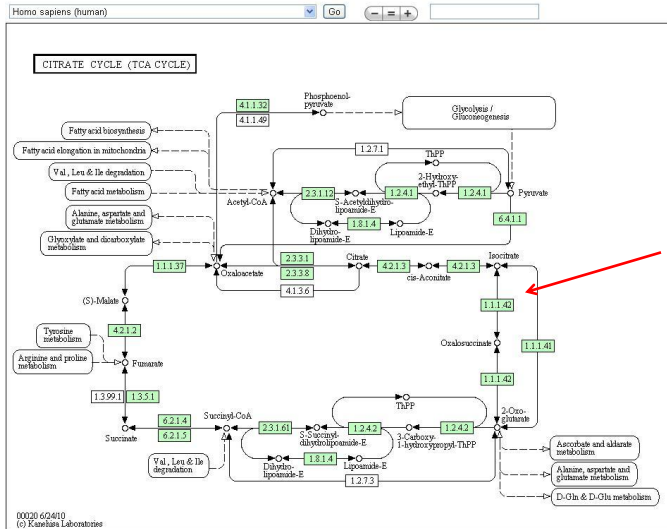
Protein ID: @Trypan @Arg-C @Lys-C @Chymotrypsin @Glu-C (Example: P43276)

Protein Sequence

Protein Sequence:

(Example: MALTARLALCEPHVDYGLAPFQVCLALDELSDMPSQFFRASEVDEL)

MRMPath ...



MRMPath results for isocitrate dehydrogenase

Click [here](#) to download this into an Excel sheet

NOTE: Please click on the 'YES' button if a warning appears when you try to open the excel sheet

hwa:3417 IDH1, IDCD, IDH, IDP, IDPC, PICD; isocitrate dehydrogenase 1 (NADP+), soluble (EC:1.1.1.42); K00031 isocitrate dehydrogenase [EC:1.1.1.42] (A)

BLAST ALL FRAGMENTS

Sequence	m/z	Parent Ion	β Ion Mass	γ > Parent Ions
BLAST IHWELIK	457.792	542.2975	801.4921	
		655.3819	688.4080	
		768.4660	802.3287	
		896.5609		
		1203.6666	2293.1551	
BLAST LIFPYVELDHSYDLGIENR	1203.6235	1340.7256	2180.0710	
		1427.7576	2033.0026	
		1590.8209	1935.9499	
		1705.8479	1772.8865	
		1818.9320	1673.0181	
		1875.9535	1544.7755	
		1989.0376	1431.6915	
		2115.0801	1316.6645	
BLAST DATNDQVTK	496.2411	517.1894	876.4474	
		645.2480	805.4103	
		744.3164	704.3626	
		845.3641	590.3197	
		973.4591	602.3560	
BLAST DAAEAIK	359.1954	458.1887	531.3189	
		571.2728	460.2818	
		699.3678		
		857.2047	657.3731	
BLAST SPNGTIR	372.7065	570.2887	560.3204	
		726.3898	446.2774	
			389.2360	

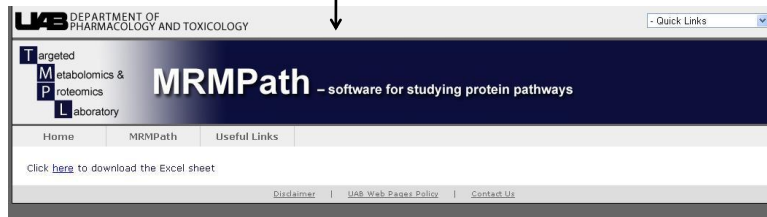
....

MRM-Mutation

• Analysis of Protein Mutations

MRMutation is a methodology that allows the user to select individual proteins and determine whether they have known mutations. This is determined by examining the EXPASY.org database. Each of the protein sequence is subjected to trypsin digestion *in silico* to determine whether these peptides with mutations are suited to multiple reaction ion monitoring. The input required is the UNIPROT Accession ID. The output spreadsheet contains the m/z values of the first three 'b' and 'y' ions (only those with values greater than the doubly charged parent ion are included), the start and end residues of the peptide with respect to the parent protein and the mutation.

Protein ID: (Example: P04632)



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](#), 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.”

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
 -