

Web-based Bioinformatics (Proteomics) Applications

Chiquito Crasto
Department of Genetics, UAB
chiquito@uab.edu
January 30, 2013

Philosophical underpinnings ...

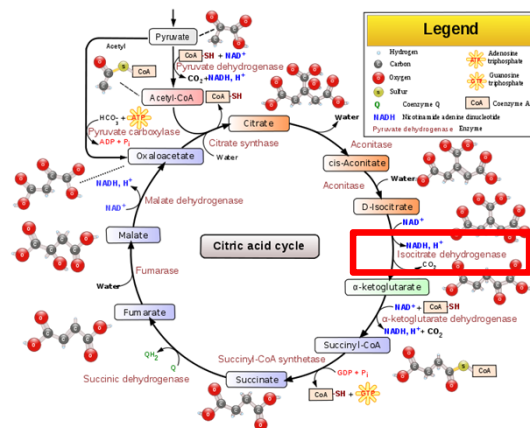
- Bioinformatics is here to stay—simply because computers are part of everyday life. This is not going to change in the near or distant future
- Students, researchers, etc., will be better served embracing bioinformatics ideas even if they do not necessarily want to pursue bioinformatics-driven careers, and opt to be “bench” scientists
 - By bioinformatics-driven, one means developmental aspects, e.g., developing software to do sequence-similarity searches
- 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

Interoperability & Database Accessibility

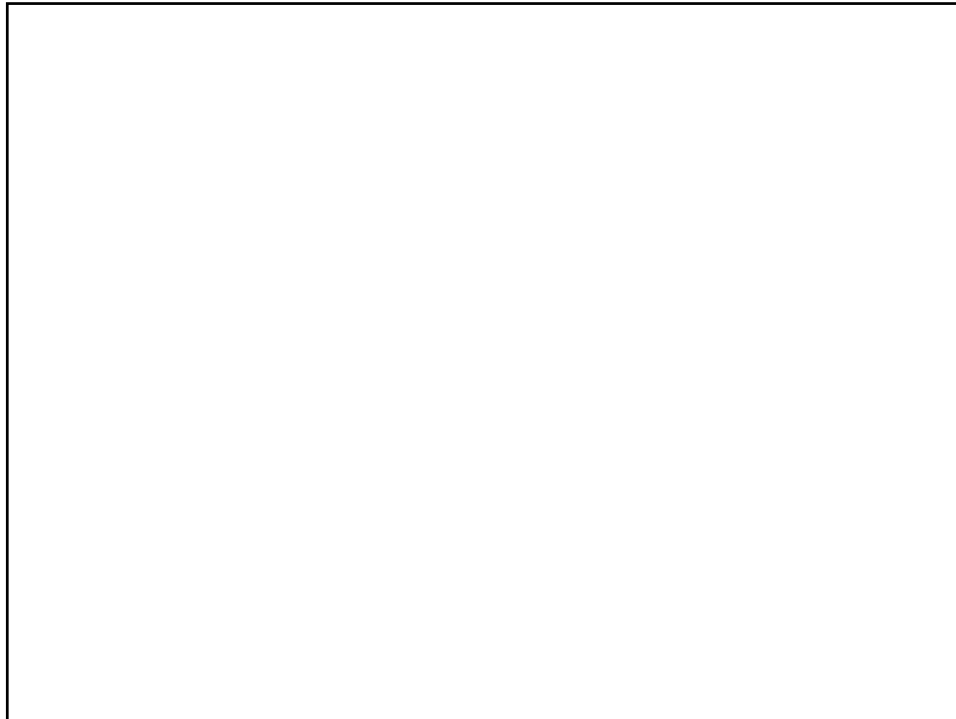
- Interoperability: the ability of systems to interoperate, that is exchange information in meaningful ways without having to reproduce information
- Integration: accessing and presenting information that is stored in different resources
 - This precludes the need to store the same information in different resources
 - Examples, how information is stored in the NCBI databases

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_aconitase_2.svg



NCBI (National Center for Biotechnology Information)

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


The screenshot shows the NCBI homepage with the following elements:

- Navigation:** 'NCBI Home', 'Resource List (A-Z)', 'All Resources', 'Chemicals & Bioassays', 'Data & Software', 'DNA & RNA', 'Domains & Structures', 'Genes & Expression', 'Genetics & Medicine', 'Genomes & Maps', 'Homology', 'Literature', 'Proteins', 'Sequence Analysis', 'Taxonomy', 'Training & Tutorials', 'Variation'.
- Welcome to NCBI:** 'The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.' Links for 'About the NCBI | Mission | Organization | Research | RSS Feeds'.
- Get Started:**
 - Tools: Analyze data using NCBI software
 - Downloads: Get NCBI data or software
 - How-To's: Learn how to accomplish specific tasks at NCBI
 - Submissions: Submit data to GenBank or other NCBI databases
- NCBI Twitter feed:** A section with a Twitter logo and a 'GO' button.
- Popular Resources:** PubMed, Bookshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome, SNP, Gene, Protein, PubChem.
- NCBI Announcements:** 'Now Available: NCBI Insights Blog!' (dated 28 Jan 2013), 'NCBI has just released a new blog called NCBI Insights. Blog posts will provide an...', 'Come to the NCBI Discovery Workshops on February 4&5!' (dated 16 Jan 2013), 'Spaces are still available for the free...', 'New version of Genome Workbench available' (dated 09 Sep 2012), 'An integrated, downloadable application...'

Selected Applications through NCBI

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

Genbank (Search Nucleotide)



GenBank Nucleotide Search Help

GenBank Submit Genomes WGS HTGs EST/GSS Metagenomes TPA TSA INSDC

GenBank Overview

What is GenBank?

GenBank[®] is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences ([Nucleic Acids Research](#), 2011, Jan. 39(Database issue):D327). There are approximately 126,551,501,141 bases in 135,440,924 sequence records in the traditional GenBank divisions and 191,401,393,188 bases in 62,715,288 sequence records in the WGS division as of April 2011.

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.

Access to GenBank

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with [Entrez Nucleotide](#), which is divided into three divisions: [CoreNucleotide](#) (the main collection), [dbEST](#) (Expressed Sequence Tags), and [dbGSS](#) (Genome Survey Sequences).
- Search and align GenBank sequences to a query sequence using [BLAST](#) (Basic Local Alignment Search Tool); BLAST searches CoreNucleotide, dbEST, and dbGSS independently; see [BLAST info](#) for more information about the numerous BLAST databases.
- Search, link, and download sequences programmatically using [NCBI e-utils](#).

GenBank Data Usage

The GenBank database is designed to provide and encourage access within the scientific community to the most up to date and comprehensive DNA sequence information. Therefore, NCBI places no restrictions on the use or distribution of the GenBank data. However, some submitters may claim patent, copyright, or other intellectual property rights in all or a portion of the data they have submitted. NCBI is not in a position to assess the validity of such claims, and therefore cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in GenBank.

GenBank Resources

- [GenBank Home](#)
- [Submission Types](#)
- [Submission Tools](#)
- [Search GenBank](#)
- [Update GenBank Records](#)

Nucleotide-Genbank's gene repository

Nucleotide (isocitrate dehydrogenase and Human) AND "Homo sapiens"[porgn_bcid9606]

Save search Limits Advanced

Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filter your results:

Found 582 nucleotide sequences. Nucleotide (181) EST (61)

Results: 1 to 20 of 181

- Homo sapiens mannoseidase, alpha, class 2C, member 1 (MAN2C1), transcript variant 4, mRNA**
2,994 bp linear mRNA
Accession: NM_001254936.1 GI: 374532778
GenBank FASTA Graphics Related Sequences
- Homo sapiens mannoseidase, alpha, class 2C, member 1 (MAN2C1), transcript variant 2, mRNA**
3,342 bp linear mRNA
Accession: NM_001254934.1 GI: 374532774
GenBank FASTA Graphics Related Sequences
- Homo sapiens mannoseidase, alpha, class 2C, member 1 (MAN2C1), transcript variant 1, mRNA**
3,291 bp linear mRNA
Accession: NM_006715.3 GI: 374532773
GenBank FASTA Graphics Related Sequences
- Homo sapiens mannoseidase, alpha, class 2C, member 1 (MAN2C1), transcript variant 3, mRNA**
3,222 bp linear mRNA
Accession: NM_001254935.1 GI: 374532776
GenBank FASTA Graphics Related Sequences
- Homo sapiens succinate dehydrogenase complex assembly factor 2 (SDHAF2), nuclear gene encoding mitochondrial protein, mRNA**
1,227 bp linear mRNA
Accession: U017841.2 GI: 300795354
GenBank FASTA Graphics Related Sequences

Find related data Database: Select Find name

Search details (isocitrate dehydrogenase[All Fields] AND ("Homo sapiens"[Organism]) OR Human[All Fields]) AND "Homo sapiens"[porgn]

Recent activity Turn Off Clear

Q (isocitrate dehydrogenase and Human) [All Fields] [All Fields]

Accession Number

A Nucleotide Entry in Genbank

Nucleotide

Links to Pubmed

Homo sapiens isocitrate dehydrogenase 2 (NADP+), mitochondrial (IDH2), nuclear gene encoding mitochondrial protein, mRNA

NCBI Reference Sequence: NC_021082

FASTA Graphics

LOCUS: NC_021082 1740 bp linear DNA 01-21-2011

DEFINITION: Homo sapiens isocitrate dehydrogenase 2 (NADP+), mitochondrial (IDH2), nuclear gene encoding mitochondrial protein, mRNA.

ACCESSION: NC_021082

VERSION: NC_021082.1 GI:2117051

KEYWORDS: DNA; Mitochondrion

SOURCE: Homo sapiens (Human)

ORGANISM: Hominidae; Eukaryota; Chordata; Chiroptera; Vertebrata; Eumetazoa; Metazoa; Mammalia; Eucaryota; Eukaryota; Primate; Primate; Hominidae; Hominidae; Homo

REFERENCE 1: Wang L, et al. (2011) *Genes* 2:1-10

REFERENCE 2: ...

REFERENCE 3: ...

REFERENCE 4: ...

REFERENCE 5: ...

REFERENCE 6: ...

REFERENCE 7: ...

REFERENCE 8: ...

REFERENCE 9: ...

REFERENCE 10: ...

ORIGIN

```

//
1   ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
61  cccctccctc cccctccctc cccctccctc cccctccctc cccctccctc cccctccctc
121 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
181 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
241 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
301 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
361 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
421 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
481 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
541 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
601 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
661 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
721 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
781 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
841 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
901 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
961 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1021 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1081 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1141 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1201 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1261 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1321 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1381 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1441 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1501 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1561 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1621 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
1681 ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac ccaagctgac
//

```

Protein Sequence

```

//
1  MGLTLPVYSLCRASGSRPAAAPALPTDQSPRRRTADERT
2  SYADPVEKSGDRTIIKDFLEELIKRFDGKQVLSGDFRSTQDPTDIAL
3  ATQKTVAVKCAITTFDEAFVVEFKLRSKSPFTIISLGGVTFPFIICNIFPL
4  VYSHFTPTITDRAADQVTAIFADVAATDFRFTFTFDSDVSGSEVTFPAQGV
5  MGRVYTDSDISFARICQVALGKQPLVTHSTNTLKVYGRFFDFCFEIVKRYTK
6  DFYDHTVTFEELIDYRPAQLVDSKDFPVAQNTSDVQVPSLQVPSLGLTDFV
7  VCFQSTTEAAARVYVTEFRFGRVSTPFLAIPATKGLRGLRGLRGLRGLR
8  RFAQLKRVVVEVSGATKGLAGLISNVLKLRFLNTDFLDTKSNLRALG
9  RQ

```

Gene Sequence

Protein Sequence in Genbank (isocitrate dehydrogenase)

The search in Genbank shows 31 results, including:

- isocitrate dehydrogenase 2 (NAD+), beta (Homo sapiens)
- isocitrate dehydrogenase 1 (NADP+), cytosolic (Homo sapiens)
- isocitrate dehydrogenase 2 (NADP+), mitochondrial (Homo sapiens)

Results: 1 to 20 of 603

1. 419 aa protein
AAC0465.1 GI:1277203
GenFasta FASTA Graphics Related Sequences Identical Proteins

2. 742 aa protein
VP_051481.1 GI:9636793
GenFasta FASTA Graphics Related Sequences Identical Proteins

3. isocitrate dehydrogenase [Corynebacterium jeikeium H411]
VP_051481.1 GI:9636793
GenFasta FASTA Graphics Related Sequences Identical Proteins

4. 742 aa protein
CAD2870.1 GI:6264362
GenFasta FASTA Graphics Related Sequences Identical Proteins

Protein
Translators of Life

Display Settings: Abstract Send to:

isocitrate dehydrogenase [Homo sapiens]
GenBank AAC0465.1
FASTA GenBank

LOCUS AAC0465 isocitrate dehydrogenase [Homo sapiens], linear, PRI 25-APR-1996
DEFINITION AAC0465 isocitrate dehydrogenase [Homo sapiens].
VERSION AAC0465.1 GI:1277203
FEATURES
location_map locus MAP02144 accession B51443.1
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Subphylum: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eulipotyphla; Eumarchontia; Primates; Haplorhina; Cercopithecoidea; Hominoidea; Hominidae; Homininae; Hominini; Homo; Homo sapiens.
REFERENCE 1 (residues 1 to 419)
Lee J, Shim J, and Wu J
Expression of human mitochondrial NADP-dependent isocitrate dehydrogenase during embryonic development
JOURNAL J. Cell. Biochem. 60 (4), 495-507 (1996)
PMID 8833802
REFERENCE 2 (residues 1 to 419)
Wu J
TITLE Direct Purification
JOURNAL Submittal (21-048-1994) Jieping Wu, Nite-Sum Hospital Research Center, University of Montreal, 1560 Sherbrooke East, Montreal, Quebec H3J 3W4, Canada
METHOD conceptual translation.
COMMENT
FEATURES
source 1..419
format="Homo sapiens"
db_xref="taxon:9606"
EcoRI 1..119
genbank="isocitrate dehydrogenase"
EcoRV 42..418
Fosmid clone="p20_04"
/note="Zinc-finger/magnesium-dependent dehydrogenase; c300445"
db_xref="taxon:9606"
1..119
Fosmid clone="M2144.1144_1103"
ORIGIN
1 aacitcaga aagcaggcag gacacaaagc palahcdhvt dtdqkydy lqsnqcdg
41 dptidnha aggrvrvh aalidnaw evklkkmw ppprlahv agvrvrvl
111 dksatcpg wvcllqgh wqphnqlq dhwvclvt vllahqwy vsvvrvvz
181 agyphvpt dtdaagph rtdqglldh qvqvnghv lrvdqvrl dqrldlsh
241 wvcdnask wvcltldh wvclhbnw pvaawvdy wvclqldg fdrqyqpw
301 wvqyghst wvavqrvv vrvrvrvvq vrvrvrvvq vrvrvrvvq vrvrvrvvq
361 fapvlvov vrvrvrvvq dlapglhiv nvlkshlv tsvrlshlv aldvqlqv
//

Note that the protein sequence and the rest of the entries are formatted similar to that of the nucleotide sequences in Genbank.

BioSystems

BioSystems BioSystems Limits Advanced

Display Settings: Abstract Send to:

Citric acid cycle (TCA cycle)

In the citric acid or tricarboxylic acid (TCA) cycle, the acetyl group of acetyl CoA (derived primarily from oxidative decarboxylation of pyruvate, beta-oxidation of long-chain fatty acids, and catabolism of ketone bodies and several amino acids) can be completely oxidized to CO₂ in reactions that also yield one high-energy phosphate bond (as GTP or ATP) and four reducing equivalents (three NADH + H⁺ and one FADH₂). The NADH and FADH₂ are then oxidized by the electron transport chain to yield nine more high-energy phosphate bonds (as ATP). All reactions of the **citric acid cycle** take place in the mitochondrion. Eight canonical reactions mediate the synthesis of citrate from acetyl-CoA and oxaloacetate and the metabolism of citrate to re-form oxaloacetate. Six additional reactions are included here. Three reversible reactions, the interconversions of citrate and isocitrate, of fumarate and malate, and of malate and oxaloacetate are annotated in both their canonical (forward) and reverse directions. The synthesis of succinate from succinyl-CoA can be coupled to the phosphorylation of either GDP (the canonical reaction) or ADP, both reactions are annotated. Two mitochondrial isocitrate dehydrogenase isozymes catalyze the oxidative decarboxylation of isocitrate to form alpha-ketoglutarate (2-oxoglutarate). IDH3 catalyzes the canonical reaction coupled to the reduction of NAD⁺, while IDH2 catalyzes the same reaction coupled to reduction of NADP⁺, a reaction whose normal physiological function is unclear. Both reactions are annotated. Finally, a reaction is annotated in which reducing equivalents are transferred from NADPH to NAD⁺ coupled to proton import across the inner mitochondrial membrane. The cyclic nature of the reactions responsible for the oxidation of acetate was first suggested by Hans Krebs, from biochemical studies of pigon breast muscle (Krebs et al. 1936; Krebs and Eggensten 1940). Many of the molecular details of individual reactions were worked out by Ochoa and colleagues, largely through studies of enzymes purified from pig heart (Ochoa 1980). While the human homologues of these enzymes have all been identified, their biochemical characterization has in general been limited and many molecular details of the human reactions are inferred from those worked out in studies of the model systems.

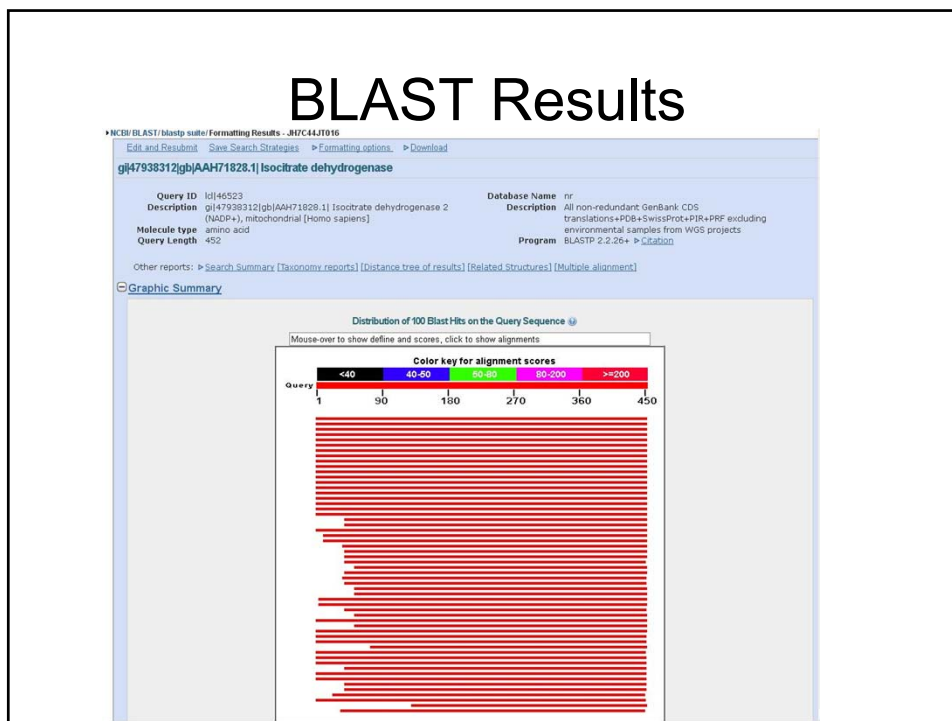
Type: pathway Taxonomic scope: organism-specific biosystem Organism: [Homo sapiens](#)

BSD: 105919 REACTOME REACT_1745

Diagram not available

Gene ID	Gene Symbol	External ID	Name
<input type="checkbox"/> 50	AC02	49054	isocitrate 2, mitochondrial
<input type="checkbox"/> 112	CS	52320	citrate synthase
<input type="checkbox"/> 1728	DLD	53720	dl-hydroxyacid dehydrogenase
<input type="checkbox"/> 1761	DLS1	60870	dl-hydroxyacid S-succinyltransferase (E2 component of 2-oxoglutarate complex)
<input type="checkbox"/> 2221	FM	55030	fumarate hydratase
<input type="checkbox"/> 3418	IKH2	57070	isocitrate dehydrogenase 2 (NADP+), mitochondrial
<input type="checkbox"/> 3419	IKH3A	57082	isocitrate dehydrogenase 3 (NAD+), alpha
<input type="checkbox"/> 3522	IKH3B	57084	isocitrate dehydrogenase 3 (NAD+), beta

BLAST Results



Pubmed—repository of biomedical abstracts

NCBI Resources (0) How To (0) My NCBI

PubMed.gov Search PubMed Limits Advanced search Help

isocitrate dehydrogenase Search Clear

PubMed

PubMed comprises more than 20 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Using PubMed

- PubMed Quick Start Guide
- Full Text Articles
- PubMed FAQs
- PubMed Tutorials
- New and Noteworthy

PubMed Tools

- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources

- MESH Database
- Journals in NCBI Databases
- Clinical Trials
- E-Books
- 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.

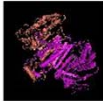
PMID: 21266166 [PubMed - as supplied by publisher]

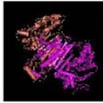
Computational Resources from NCBI's Structure Group

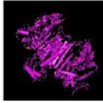
<http://www.ncbi.nlm.nih.gov/Structure/index.shtml>

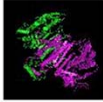
Three-dimensional structure views in Genbank-- STRUCTURE

Results: 1 to 20 of 133 « First < Prev Page 1 of 7 Next > Last »»

- 

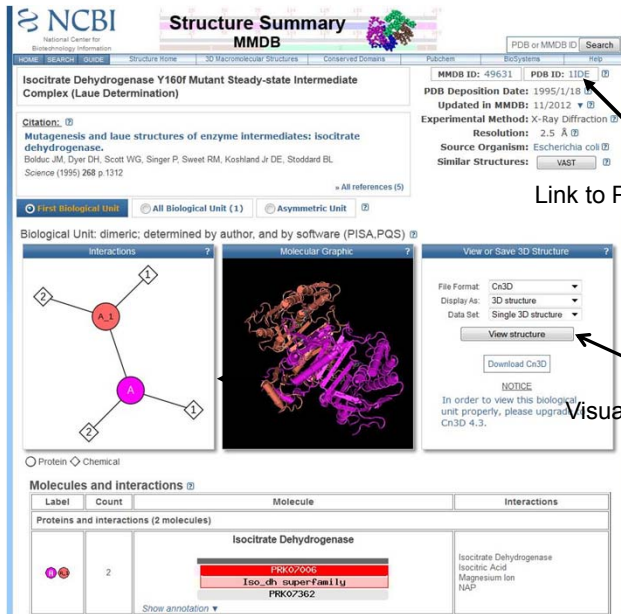
1. Isocitrate Dehydrogenase Y160f Mutant Steady-State Intermediate Complex (Laue Determination)[Oxidoreductase (Nad(A)-Choh(D)), EC: 1.1.1.42]
 Taxonomy: Escherichia coli
 Proteins: 2 Chemicals: 3 modified: 2012/11/01 00:00
 MMDB ID: 49631 PDB ID: 1IDE
[View in Cn3D](#) [PubMed](#) [Protein](#) [Similar Structures](#) [Conserved Domains](#) [PubChem Compound](#)
- 

2. Crystal Structure Of Isocitrate Dehydrogenase Mutant K230m Bound To Isocitrate And Mn2+Oxidoreductase, EC: 1.1.1.42]
 Taxonomy: Escherichia coli
 Proteins: 2 Chemicals: 3 modified: 2012/10/19 00:00
 MMDB ID: 10849 PDB ID: 1CW1
[View in Cn3D](#) [PubMed](#) [Protein](#) [Similar Structures](#) [Conserved Domains](#) [PubChem Compound](#)
- 

3. Low Temperature Structure Of Wild-Type Idh Complexed With Mg-Isocitrate[Oxidoreductase, EC: 1.1.1.42]
 Taxonomy: Escherichia coli
 Proteins: 2 Chemicals: 3 modified: 2012/11/13 00:00
 MMDB ID: 11279 PDB ID: 1CW7
[View in Cn3D](#) [PubMed](#) [Protein](#) [Similar Structures](#) [Conserved Domains](#) [PubChem Compound](#)
- 

4. Regulation Of An Enzyme By Phosphorylation At The Active Site[Oxidoreductase (Nad(A)-Choh(D)), EC: 1.1.1.42]
 Taxonomy: Escherichia coli
 Proteins: 2 Chemicals: 2 modified: 2013/01/13 00:00
 MMDB ID: 58498 PDB ID: 5ICD
[View in Cn3D](#) [PubMed](#) [Protein](#) [Similar Structures](#) [Conserved Domains](#) [PubChem Compound](#)

Structure of Actin—Genbank Structure View

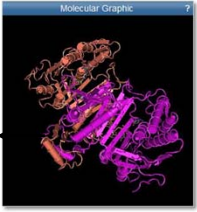


Structure Summary
MMDB

Isocitrate Dehydrogenase Y160f Mutant Steady-state Intermediate Complex (Laue Determination)


Citation:
 Mutagenesis and laue structures of enzyme intermediates: isocitrate dehydrogenase.
 Bolduc JM, Dyer DH, Scott WG, Singer P, Sweet RM, Koshland Jr DE, Stoddard BL
 Science (1995) 268 p 1312

Biological Unit: dimeric; determined by author, and by software (PISA,PQS)

Molecular Graphic: 

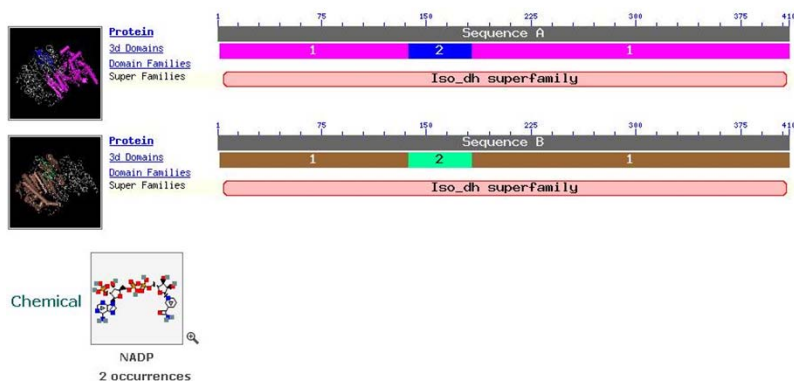
View or Save 3D Structure:
 File Format: Cn3D
 Display As: 3D structure
 Data Set: Single 3D structure
View structure
 Download Cn3D

Molecules and Interactions

Label	Count	Molecule	Interactions
Proteins and interactions (2 molecules)			
	2	Isocitrate Dehydrogenase Iso_dh_isoperFam11y PRK07362	Isocitrate Dehydrogenase Isocitrate Acid Magnesium Ion NAP

Annotations:
 - Link to Protein Databank (points to PDB ID: 1IDE)
 - Visualization software (points to View structure button)

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 (CDD) interface. The top navigation bar includes links for All Databases, PubMed, Gene, HomoloGene, Protein, Protein Clusters, Structure, PubChem, and BioSys. The search bar is set to 'Conserved Domains' and includes 'Go' and 'Clear' buttons. Below the search bar are tabs for Limits, Preview/Index, History, Clipboard, and Details. The main content area is titled 'Conserved Domains Database' and includes a 'Hints on Finding a Conserved Domain' section with the following bullet points:

- This page is used for searching by text term (other search methods allow queries by protein sequence)
- Enter one or more search terms (e.g., chloride channel)
- Use search fields to narrow your search (tips)
- Advanced search options are available in the Limits, Preview/Index, and History folder tabs
- Boolean operators AND, OR, NOT must be in upper case
- Use quotes to force a phrase search (e.g., "voltage gated")
- Use a wildcard (e.g., glyco[title]) to search for a word stem
- Search results and conserved domain records are described in the help document.

Below the hints is an 'About the Database' section. It states: 'Conserved domains are functional units within a protein that have been used as building blocks in molecular evolution and recombined in various arrangements to make proteins with different functions. The Conserved Domain Database (CDD) brings together several collections of multiple sequence alignments representing conserved domains, including NCBI-curated domains, which use 3D-structure information to explicitly to define domain boundaries and provide insights into sequence/structure/function relationships, as well as domain models imported from a number of external source databases (Pfam, SMART, COG, PRK, TIGRFAM). The data are then 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.'

In the bottom right corner, there is a small 3D protein structure model with the text 'What is a conserved domain?' next to it.

CDD ...

NCBI Conserved Domains

Search: Conserved Domains for isocitrate dehydrogenase

Display: Summary Show 20 Sort By Send to

Links: Related CDs, Literature, Sequence, Structure, BioSystems, Other Links

Alt: 24 NCBI-curated: 0 families: 22 imported: 17 superfamilies: 3

Items 1 - 20 of 24

1: **cd1891** AceK Isocitrate dehydrogenase kinase/phosphatase (AceK)
This family consists of several bacterial isocitrate dehydrogenase kinase/phosphatase (AceK) proteins (EC 2.7.1.116). [186492]

2: **pfam06315** AceK Isocitrate dehydrogenase kinase/phosphatase (AceK)
This family consists of several bacterial isocitrate dehydrogenase kinase/phosphatase (AceK) proteins (EC 2.7.1.116). [148119]

3: **pfam03971** IDH Monomeric isocitrate dehydrogenase
NADP(+) dependent isocitrate dehydrogenase (ICD) is an important enzyme of the intermediary metabolism, as it controls the carbon flux within the citric acid cycle and supplies the cell with 2-oxoglutarate EC:1.1.1.42 and NADPH for biosynthetic purposes. [112770]

4: **TIGR02924** ICDH_alpha isocitrate dehydrogenase
This family of mainly alpha-proteobacterial enzymes is a member of the isocitrate/isopropylate 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. [163075]

5: **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*, *Corynebacterium glutamicum*, *Rhodocrobium vanielii*, and *Neisseria meningitidis*. It is NADP-specific. [129282]

6: **TIGR00193**

CDD...

pfam03971: IDH

Monomeric isocitrate dehydrogenase
NADP(+) dependent isocitrate dehydrogenase (ICD) is an important enzyme of the intermediary metabolism, as it controls the carbon flux within the citric acid cycle and supplies the cell with 2-oxoglutarate EC:1.1.1.42 and NADPH for biosynthetic purposes.

Links
Statistics
Structure

PubMed References

- Cloning, sequence analysis, expression, and inactivation of the *Corynebacterium glutamicum* *idh* gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme. *J. Bacteriol.* 1995 Feb; 177(3):774-782
- Genes encoding two isocitrate dehydrogenase isozymes of a psychrophilic bacterium, *Vibrio* sp. strain ABE-1. *J. Bacteriol.* 1993 Nov; 175(21):6873-6880

pfam03971 is a member of the superfamily cl15383.

Sequence Alignment

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

```

10 20 30 40 50 60 70 80
g1 81669476 7 I I I T I L T D E A P L A T A P L P I Y R A F A F A G I K I E A S D I Y A A R I L A A P P D V L T E Q V F O R A E L G R I T Q L R O T H I K I L P 86
g1 81545133 7 I T I T T D E A P A L A T Q E L L P I Q P F A S G I W P T E R D I S L A G R I L A A P P D V L T E Q V F O R A E L G R I T Q L R O T H I K I L P 86
g1 81546035 23 K L I V I L T D E A P L A T Q E L L P I K A V T R V A G I Q V E T R D I S L A G R I L A A P P D V L T E Q V F O R A E L G R I T Q L R O T H I K I L P 102
g1 81784714 6 I V I T H T D E A P A L A T Q E L L P I V Q A F A S G I D I W T S D I S L G R I L A A P P E L I T A Q V F O L A E L G E L V R Q P D A N Y I K L P 85
g1 81540775 6 F I I T T D E A P A L A T E S L F A R F A S A G I W P T E R D I S L A G R I L A A P P D V L T E Q V F O R A E L G R I T Q L R O T H I K I L P 85

90 100 110 120 130 140 150 160
g1 81669476 87 H I S A Y P Q L V A A I K E L Q R G Y A V P D V A P K T D E K A I P E R A C L G S A V N P V L R Q S R R A P P A V F A R R P R N S I G 166
g1 81545133 87 H I S A Y P Q L V A A I K E L Q R G Y A V P D V A P K T D E K A I P E R A C L G S A V N P V L R Q S R R A P P A V F A R R P R N S I G 166
g1 81546035 103 H I S A S P C L A A I K E L Q R G V L L P D V F D E P P S A G Q R V K I R V D V F G S A V N P V L R Q S R R A P P A V F A R R P R N S I G 182
g1 81784714 86 H I S A Y P Q L V A A I K E L Q R G Y A V P D V A P K T D E K A I P E R A C L G S A V N P V L R Q S R R A P P A V F A R R P R N S I G 165
g1 81540775 86 H I S A Y P Q L V A A I K E L Q R G Y A V P D V A P K T D E K A I P E R A C L G S A V N P V L R Q S R R A P P A V F A R R P R N S I G 165

170 180 190 200 210 220 230 240
g1 81669476 167 W N S A S T R V A R B R G D F A R E S N T L D R A R V M E L L A K S R K T I V L K E V F L D C D V I D N P S R G A L C D P Y E Q N C D A P 246
g1 81545133 167 W N S A S T R V A R B R G D F A R E S N T L D R A R V M E L L A K S R K T I V L K E V F L D C D V I D N P S R G A L C D P Y E Q N C D A P 246
g1 81546035 183 W S T T S S H V A R N E G D F P O S E R S A V L Q A G L R V L I A N G D V T V L R Q V A V A G D V D S A V L S G A L A V F A A G S Q A K 262
g1 81784714 166 W T D S K T R V A I N G S D F F R S E Q S V I V E A T S V I V I T T I N Q R R E L R E P V A K A G E I I D A T V N S R G A L A F A L A K Q V R A K 245
g1 81540775 166 W N S A S T R V A R B R G D F A R E S N T L D R A R V M E L L A K S R K T I V L K E V F L D C D V I D N P S R G A L C D P Y E Q N C D A P 246

250 260 270 280 290 300 310 320
g1 81669476 247 E T G V F S L R V A T D N G V S H P I V F G S A V R I P Y G A F A R S G E L F D D G V N P N G L S D L V S E I S P A S Q R E I I E D L R S C H E 326
g1 81545133 247 A G V L F S L S R G A T D N G V S D P I P S G A V F Y T G A L F P R E S G L F D Q G V N P N G L S D L V S E I S P A S Q R E I I E D L R S C H E 326
g1 81546035 248 S Q G L F S L S R G A T D N G V S D P I P S G A V F Y T G A L F P R E S G L F D Q G V N P N G L S D L V S E I S P A S Q R E I I E D L R S C H E 342
g1 81784714 246 A R G V L F S L S R G A T D N G V S D P I I F S G A V F Y F A V F E F F G D L A A G V P N S G N L A N L D L A D D T R T A E A E I A D A V T A 325
g1 81540775 245 E T G V F S L R V A T D N G I S H P I V F G S A V R I P Y G A F A R S G E L F D D G V N P N G I S D V F T K S L L P A S Q E I I D S H E Y T S 324

```

Clustering Proteins in terms of Sequence Similarities--Genbank

NCBI Protein Clusters

Search: Protein Clusters for [] Go Clear

Limits: Preview/Index History Clipboard Details

Display: Overview Show 20 Send to

All: 1 Archaea: 0 Bacteria: 1 Curated: 1

PRK15498 isocitrate dehydrogenase Gene name: None

(Curated - Provisional)

Cluster Info

ID: 293783
 Total proteins: 386
 Conserved in: Bacteria
 Total genes: 159
 Total organisms: 383
 Putative Paralogs: 6
 Publications: 8

Cluster Tools

Show detailed alignment [Go]
 Build tree [Go]
 Genome ProtMap by PRK15498 [Go]
 Genome ProtMap by COG2838C [Go]
 Cluster Patterns [Go]

Cross references

COG(s): COG2838C
 EC Number: 1.1.1.42
 KEGG KO: K00001
 ACLAME: family:all0982
 InterPro: 2
 TIGRFAM: 2
 Domains(s): C15330(DH)
 Structures: 3

Entrez Links

NADP-specific, catalyzes the formation of 2-oxoglutarate from isocitrate or oxalosuccinate

Domain description: **Monomeric isocitrate dehydrogenase**
 COG functional category: **Energy production and conversion**

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)

- GeneRIF [2]**: Structure of a highly NADP+-specific isocitrate dehydrogenase. *Acta Crystallogr D Biol Crystallogr* 2011 [Cite more](#)
- SwissProt [2]**: Substrate-free structure of a monomeric NADP isocitrate dehydrogenase: an open conformation phylogenetic relationship of isocitrate dehydrogenase. *Proteins* 2006 Apr 1 [more](#)
- By Homology [2]**: Genes encoding two isocitrate dehydrogenase isozymes of a psychrophilic bacterium, *Vibrio* sp. strain ABE-1. *J Bacteriol* 1993 Nov [more](#)
- CDD [2]**: Cloning, sequence analysis, expression, and activation of the *Corynebacterium glutamicum* *icd* gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme. *J Bacteriol* 1995 Feb [more](#)
- Structure [1]**: Substrate-free structure of a monomeric NADP isocitrate dehydrogenase: an open conformation phylogenetic relationship of isocitrate dehydrogenase. *Proteins* 2006 Apr 1 [more](#)

Clustering Proteins in terms of Sequence Similarities--Genbank

Top Pattern: <--CLSK967410 PRK15498 CLSK967412 PRK00143

Organism (Click here to search)	Protein name	Prev. Cluster	Accession	Next Cluster	Locus_tag	Length	UniProt	BLink	Alignment
(Limit to paralogs)									
Unknown									
<input type="checkbox"/> <i>Selidibacter</i>	isocitrate dehydrogenase, NADP-dependent	PRK00223	YP_024905038	CLSK2835462	Calni_0402	741aa			Identical sequences are framed
<input type="checkbox"/> <i>Delfinibacter desulfuricans</i> DSM 19672	isocitrate dehydrogenase, NADP-dependent	PRK09238	YP_024961446	PRK005223	DEFDS_0918	746aa			
<input type="checkbox"/> <i>Delfinibacter aestiviphilus</i> DSM 12829	isocitrate dehydrogenase	CLSK2835999	YP_023505006	PRK005223	Daoni_2347	743aa			
<input type="checkbox"/> <i>Desulfosporosillum indicum</i> sp.	isocitrate dehydrogenase, NADP-dependent	CLSK922773	YP_024113410	PRK005223	Selin_2134	742aa			
<input type="checkbox"/> <i>Flavobacterium</i> subsp. <i>suodiosogenes</i> S34	isocitrate dehydrogenase	CLSK2558222	YP_023248769	CLSK942319	Flisu_0674	742aa			
<input type="checkbox"/> <i>Flavobacterium</i> subsp. <i>suodiosogenes</i> S34	isocitrate dehydrogenase	PRK09238	YP_024802930	PRK005223	Flisvi_0687	747aa			
C. Actinobacteria									
<input type="checkbox"/> <i>Amvobacterium</i> subsp. <i>suodiosogenes</i> DS3-9A1	isocitrate dehydrogenase	CLSK730518	YP_024495093	CLSK2991600	AS9A_4400	745aa			
<input type="checkbox"/> <i>Amvobacterium</i> subsp. <i>suodiosogenes</i> DS1-2099	isocitrate dehydrogenase	PRK02813	YP_023097772	CLSK898998	Amv_1481	735aa			
<input type="checkbox"/> <i>Amvobacterium</i> subsp. <i>suodiosogenes</i> Re117	isocitrate dehydrogenase		YP_023915844	CLSK961314	AARI_08450	740aa			
<input type="checkbox"/> <i>Amvobacterium</i> subsp. <i>suodiosogenes</i> TC1	isocitrate dehydrogenase, NADP-dependent	CLSK959579	YP_946987	CLSK777159	AAur_1201	739aa			
<input type="checkbox"/> <i>Amvobacterium</i> subsp. <i>suodiosogenes</i> AB	isocitrate dehydrogenase	CLSK959579	YP_024287260	CLSK2470373	Abi_1170	739aa			
<input type="checkbox"/> <i>Amvobacterium</i> subsp. <i>suodiosogenes</i> Sphe3	isocitrate dehydrogenase, monomeric type	CLSK959579	YP_024240448	CLSK001003	Amch3_11320	739aa			
<input type="checkbox"/> <i>Amvobacterium</i> sp. FB24	isocitrate dehydrogenase	CLSK959579	YP_830588	CLSK778456	Amv_1092	740aa			
<input type="checkbox"/> <i>Brachybacterium</i> subsp. <i>suodiosogenes</i> DSM 4918	isocitrate dehydrogenase, NADP-dependent, monomeric type	PRK00377	YP_023155580	CLSK3067749	Bfae_22050	746aa			

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

Ensembl release 60 - Nov 2010 © WTJ | EBI

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

Chromosome 15: 90,288,632-90,483,231

Location:

Gene:

ENSEMBL—Gene Summary

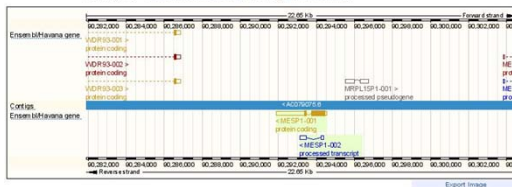
Description mesoderm posterior 1 homolog (mouse) [Source:HGNC Symbol;Acc:29659]
Location [Chromosome 15: 90,291,892-90,294,541](#) reverse strand
Transcripts This gene has 2 transcripts

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
MESP1-001	ENST00000300057	2369	ENSP00000300057	268	Protein coding	CCDS10355
MESP1-002	ENST00000559894	451	No protein product	-	Processed transcript	-

Transcript and Gene level displays
 In Ensembl we provide displays at two levels:
 • Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation.
 • Gene views which provide displays for data associated at the gene level such as orthologues, paralogues, regulatory regions and splice variants.
 This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want using the menu at the left hand side of the page. To return to viewing gene level information click on the Gene tab in the menu bar at the top of the page.

Gene summary [help](#)

Name MESP1 (HGNC Symbol)
Synonyms bHLH5, MGC10676 [to view all Ensembl genes linked to the name [click here](#)]
CCDS This gene is a member of the Human CCDS set: [CCDS10355](#)
Gene type Known protein coding
Prediction Method Annotation for this gene includes both automatic annotation from Ensembl and [Havana](#) manual curation, see [article](#).
Alternative genes This gene corresponds to the following database identifiers:
Havana gene: [GTH.HMS0000149810 \(version 2\)](#) [[view all locations](#)]



ENSEMBL—Protein

Transcript: MESP1-001 ENST00000300057

Description mesoderm posterior 1 homolog (mouse) [Source:HGNC Symbol;Acc:29659]
Location [Chromosome 15: 90,291,892-90,294,541](#) reverse strand
Gene This transcript is a product of gene [ENSG00000166823](#) - This gene has 2 transcripts

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
MESP1-001	ENST00000300057	2369	ENSP00000300057	268	Protein coding	CCDS10355
MESP1-002	ENST00000559894	451	No protein product	-	Processed transcript	-

Transcript and Gene level displays
 Views in Ensembl are separated into gene based views and transcript based views according to which level the information is more appropriately associated with. This view is a transcript level view. To flip between the two sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page.

Protein summary [help](#)



Statistics
 Ave. residue weight: 106.348 g/mol
 Charge: 0.0
 Isoelectric point: 9.0165
 Molecular weight: 28,501.38 g/mol
 Number of residues: 268 aa

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 image displays two side-by-side screenshots of web portals. The left screenshot shows the UniProt website, featuring a search bar at the top, a 'WELCOME' message, and sections for 'What we provide', 'Getting started', and 'PROTEIN SPOTLIGHT'. The right screenshot shows the ExPASy website, titled 'ExPASy Bioinformatics Resource Portal', with a search bar, a 'Visual Guidance' section, a 'Categories' list, and a 'Featuring today' section highlighting 'T-Coffee'.

SwissProt—search for Proteins

Search in: Protein Knowledgebase (UniProtKB) Query: [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 (7,578) (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 O3326 YOR3326W	Saccharomyces cerevisiae (Baker's yeast)	369
<input type="checkbox"/> P11071	ACEK_ECOLI	★	Isocitrate dehydrogenase kinase phosphatase	aceK b4016 JW3976	Escherichia coli (strain K12)	578
<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 ctc BSU29130	Bacillus subtilis	423
<input type="checkbox"/> Q8LFC0	IDH1_ARATH	★	Isocitrate dehydrogenase [NAD] regulatory sub...	IDH1 A4g35260 F23E12.180	Arabidopsis thaliana (Mouse-ear cress)	367
<input type="checkbox"/> Q04837	IDH3B_HUMAN	★	Isocitrate dehydrogenase [NAD] subunit beta, ...	IDH3B	Homo sapiens (Human)	385

EXPASY-Databases and Features

The image shows a grid of icons for various databases and features. A blue arrow points to the 'Translate' icon, which is located in the middle-right section of the grid. The icons are arranged in a grid with letters 'g', 'h', 'i', 'l', 'm', 't', 'u', 'v' on the left side, indicating different categories or groups of tools.

Swiss 2D-PAGE

SWISS-2DPAGE

Search by

- [accession number]
- [description, ID or gene] ▶
- [author name]
- [spot ID / serial number]
- [identification methods]
- [pI / Mw range]
- [combined fields]

Maps

- [experimental info]
- [protein list]
- [graphical interface]

Select Remote Interfaces

- [All Interfaces]
- World-2DPAGE Portal
- World-2DPAGE Repository
- Exclude local DBs
has only effect if a remote interface is selected

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

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 *apo*, or *APO1* 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 ("").



Database constructed and maintained by SIB, using the *Make2D-DB II* package (ver. 3.10.2) from the World-2DPAGE Constellation of the ExPASy web server

Swiss 2DPAGE –Isocitrate dehydrogenase

Searching in SWISS-2DPAGE for entries exhibiting any of the keywords: isocitrate dehydrogenase

in their description (DE), entry name (ID), gene names (GN) or UniProtKB/Swiss-Prot keywords (KW)

Query Result: 09 results

Accession number	ID	Description	Genes	Keywords	Species
008749	DLIH_MOUSE	Dihydrolysoyl dehydrogenase, mitochondrial (EC 1.1.1.41) (Dihydrolysoamide dehydrogenase)	(Name=DIH)	Acetylation, Direct protein acetylation, FAD, Flavoprotein, Mitochondrion, NAD, Oxidoreductase, Hydroxylase center, Transid peptide	Mus musculus (Mouse)
008756	HCID_MOUSE	3-hydroxyacyl-CoA dehydrogenase type-2 (EC 1.1.1.30) (3-hydroxyacyl-CoA dehydrogenase type II) (Type II HADH) (3-hydroxy-2-oxohydroxy-CoA dehydrogenase) (EC 1.1.1.78) (Endoplasmic reticulum-associated acylated beta-oxidation binding proteins)	(Name=Hsd17h10; CytochromeP450; Hsdh2)	Acetylation, Direct protein acetylation, NAD, Oxidoreductase	Mus musculus (Mouse)
008844				2D structure, Cytoplasm, I protein, acetyl, cystine, protein, return, gene, binding, ADP, ubiquitin, thioester, cycle	Mus musculus (Mouse)

SWISS2DPAGE: 008749

008749

General information about the entry

View entry in complete text format

Entry name: **DLIH_MOUSE**

Primary accession number: **008749**

Integrated into SWISS-2DPAGE on: April 1, 2002 (release 12)

2D Annotations were last modified on: March 31, 2004 (version 1)

General Annotations were last modified on: May 28, 2007 (version 7)

Name and origin of the protein

Description: **Dihydrolysoyl dehydrogenase, mitochondrial (EC 1.1.1.41) [Dihydrolysoamide dehydrogenase]**

Gene name: **Name=DIH**

Associated species: **Mus musculus (Mouse) [TaxID: 10090]**

Taxonomy: **Eukaryota, Mammalia, Chordata, Carnivora, Vertebrata, Eulacostoma, Mammalia, Eutheria, Euarchontoglires, Glires, Rodentia, Sciuromorpha, Muridae, Murinae, Mus**

References

[1] **MAPPING ON GEL:** Pridmore IR, BIRD, J.C.D., G.P.H.D., G.D., Sirek, Aghajani, Sanchez J, C, Chigara D, Comand V, Hougland C, Eric P, A, Passaro S, Appel R, D, Wang S, Seidlt M, Nolin A., Coulombe RA., Hochstrasser DF. "The mouse SWISS-2DPAGE database: a tool for proteomics study of diabetes and obesity" *Proteomics* 13(9):1830(2013)

2D PAGE maps for identified proteins [Click here to expand details for this protein](#)

How to interpret a protein

ISLETS_MOUSE (Pancreatic islet cells)

Main isoforms (MusMol):

Tissue: Pancreatic islet

ISLETS_MOUSE

MAP LOCATIONS:

- SPOT 20020006

MAPPING: identification: Peptide mass fingerprinting [1]




This experimental entry is not yet annotated.

Copyright: This SWISS-2DPAGE entry is copyright the Swiss Institute of Bioinformatics. There are no restrictions on its use by non-profit institutions as long as its content is not re-modified and the abstract is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.expasy.org/2dp/faq.html> for details).

Cross references:

REFSEQ: SWISS2DPAGE: 008749, DLIH_MOUSE

Swiss PDB Model Repository

SWISS-MODEL Repository

Modelling Tools Repository Documentation

[Repository Query] [Full Text Query]


Welcome to the SWISS-MODEL Repository

The SWISS-MODEL Repository is a database of annotated three-dimensional comparative protein structure models generated by the fully automated homology-modelling pipeline SWISS-MODEL.

Example Queries:
 [P23298] [SLDA_ECOLI] [PII00743503] [NP_416402] [G1:20454808] [ENTREZ:54401] [Sequence]



P08200 -- isocitrate dehydrogenase Accession Number

SEARCH






The current release of the SWISSMODEL-Repository (10.2.2) consists of 3021185 model entries for 2244854 unique sequences in the UniProt database.

NOTE: The SWISS-MODEL repository contains theoretically calculated models, which may contain significant errors.

SwissModel Repository ...

SWISS-MODEL Repository

Modelling Tools Repository Documenta

[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) ★


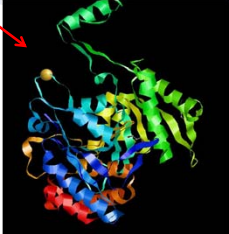
Domain [-]

no data available!

Model 3D Structure [-]

Based on structure: **1pb1** [SMTL] [RCSB] [PDBe] [SCOP] [CATH]

Sequence identity: 100%
 Residue range: 1 to 416

A four location model to explain the stereospecificity of proteins.

DOI: 10.2218/pdb1pb1/1pb1

Primary Citation

A new model for protein stereospecificity.
 Meserac, A.D., Koshland Jr, D.E.
 Journal: (2000) Nature 403: 654-655
 PubMed: 10568187
 DOI: 10.1038/35051144
 Search Related Articles in PubMed
 PubMed Abstracts:
 No abstract available... [Read More & Search PubMed Abstracts]

Molecular Description

Classification: Oxidoreductase
 Structure Weight: 46374.47
 Molecule: Isocitrate dehydrogenase (NADP)
 Polymer: 1 Type: polypeptide(s) Length: 416
 Chain: A 1,1,1,42
 EC#: [Other Details:]

Source

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


Related PDB Entries

1L
 Details

Ligand Chemical Component

Identifier Formula Name Interactions

Biological Assembly



View in Jmol | Show/Hide
 Other options: Protein Workshop
 Biological assembly assigned by authors

MyPDB Personal Annotations

To save personal annotations, please login to your MyPDB account.

Deposition Summary

Authors: Meserac, A.D., Koshland Jr, D.E.
 Deposition: 2003-05-14
 Release: 2003-08-17
 Last Modified: 2009-02-24 (PDBv41)

Uniref—Clustering of Proteins

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

Published January 11, 2011

Built 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 637 organisms

Dataset

UniProt (1137)

Taxonomy

Filter Reset

Members Customize

Member	Entry name	Status	Protein names	Organism	Component clusters	Length
<input type="checkbox"/>	Q9ZN36	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori J99 (Campylobacter pylori J99)	UniRef100_Q9ZN36 UniRef90_Q9ZN36	425
<input type="checkbox"/>	B2ZP83	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (Campylobacter pylori)	UniRef100_B2ZP83 UniRef90_Q9ZN36	425
<input type="checkbox"/>	B6JPC2	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (strain P12)	UniRef100_B6JPC2 UniRef90_Q9ZN36	425
<input type="checkbox"/>	B9XWT5	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori 98-10	UniRef100_B9XWT5 UniRef90_Q9ZN36	425
<input type="checkbox"/>	E1Q800	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (strain Cuz20)	UniRef100_E1Q800 UniRef90_Q9ZN36	425
<input type="checkbox"/>	Q1CVDD	★	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (strain HPAG1)	UniRef100_Q1CVDD UniRef90_Q9ZN36	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 BRITTE KO GENES SSDB LIGAND DBGET

KEGG Atlas

KEGG Atlas is a new graphical interface to the KEGG suite of databases, especially to the systems information in the PATHWAY and BRTE databases. It currently consists of a global metabolism map with newly developed views and a cancer map with the traditional KEGG map view.

Metabolism map (version 0.1, to be phased out)

Metabolism map

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 **linkset**, and identified by the **li** number. Each linkset corresponds to one to several KO's (such as this) in the reference pathway view, or one to several genes (such as this) in an organism-specific view.

KEGG Pathway


CITRATE CYCLE (TCA CYCLE)

Isocitrate dehydrogenase

00020 11/3/09
© Kanehisa Laboratories

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) 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 PGA2 DUF505 PROSITE: IDH_IMDH Motif
Other DBs	NCBI-GI: 28178825 NCBI-GeneID: 3417 OMIM: 147700 HGNC: 5282 HFPD: 00984 Ensembl: ENSG00000138413 UniProt: Q75874 Q6F9Q6
Structure	PDB: 3INM 1TOL 1T09 3MAP 3MAS 3MAR Thumbnails  Jmol
Position	2q33.3

...

MASCOT—Protein Identification from Mass Spectrometry Data

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

MATRIX SCIENCE HOME WHAT'S NEW MASCOT HELP PRODUCTS SUPPORT TRAINING CONTACT [Log in](#) [CS](#)

Welcome

This site features **Matrix**, 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 *Electrophoresis*, 20(18):3551-67 (1999) (abstract).


We value your feedback and suggestions for new features. If you find any problems, errors, oversights, 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 Daryl 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 Swiss-Prot available on this web site for searching with Mascot.

MRM-Path

MRMPATH - software for studying protein pathways

The quantitative study of proteins in biological pathways using mass spectrometry is rapidly becoming a primary method. MRMPATH is a web-based, platform-independent software that facilitates the identification of the peptides and their fragment ions for each protein in a metabolic or signaling pathway, or for the components of a protein complex that can be used for quantitative analysis using multiple reaction ion monitoring mass spectrometry (MRM-MS).

MRMPATH offers two ways to identify the peptides

1. To extract ions from previously obtained MS/MS spectra to generate instrument-specific ion pair combinations.
2. To predict the peptides from a protein sequence that would be most suitable for MRM-MS analysis.

MRMPATH takes into account peptide type, peptide size and peptide sequence, and the likelihood that the peptide(s) are unique and not found in other proteins.

MRMPATH - software for studying protein pathways

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 and perform a tryptic digest in silico to determine peptides that are suited to multiple reaction ion monitoring. The only input required is the UniProt Accession ID (or a protein sequence). For each fragment, the m/z values of the bⁿ and yⁿ 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 links to GeneDB. The resulting data can be exported to a comma-delimited file.

Analysis of Protein Mass Fragments from Pathways

None

Trypsin 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 ion monitoring. The only input required is the UniProt Accession ID (or a protein sequence). For each fragment, the m/z values of the bⁿ and yⁿ 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 links to GeneDB. The resulting data can be exported to a comma-delimited file.

Protein ID

Protein ID (GeneDB): (Example: P6276)

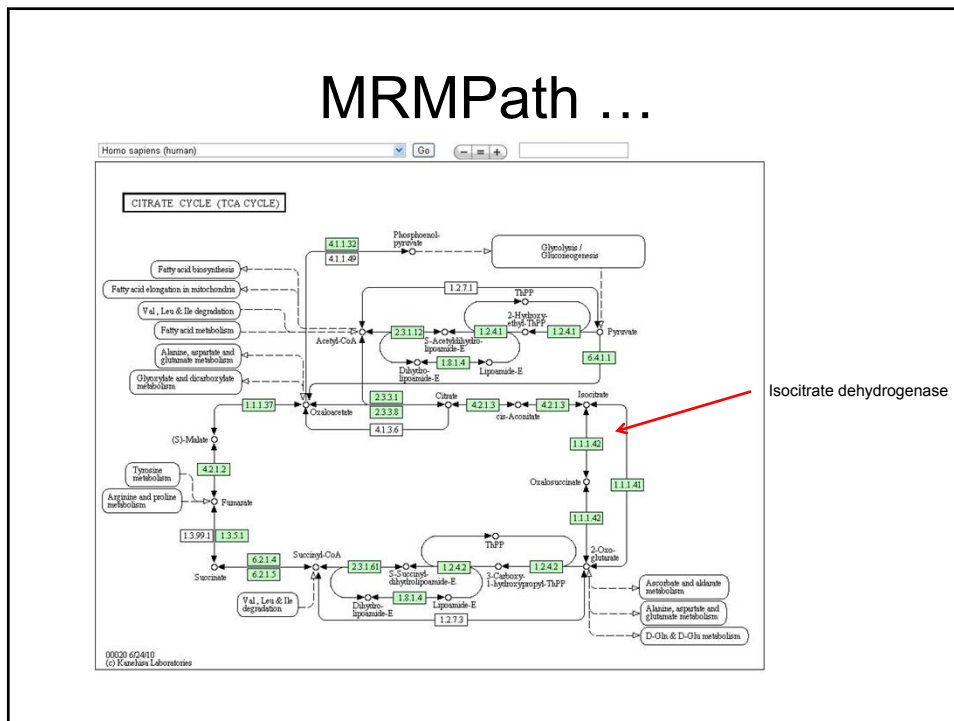
Protein Sequence

Protein Sequence:

Glu-C Trypsin Arg-C Lys-C Chymotrypsin

(Example: MALTAVPLSALVDEPHQIVTGLRFQVCLGSLDEKGLFSSQAFYRSEVGEVGL)

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
 hsa:3417 IDH1, IDGD, 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	B Ion Mass	Y > Parent Ions
BLAST IIWELIK	457.792	542.2979	801.4921
		555.3819	689.4080
		768.4660	502.3287
		896.5609	
BLAST LIFPVVELDLHSYDLGIENR	1203.6235	1203.6660	2293.1551
		1340.7256	2180.0710
		1427.7576	2033.0026
		1590.8209	1935.9499
		1705.8479	1772.8865
		1818.9320	1673.8181
		1875.9535	1544.7755
		1989.0376	1431.6915
		2118.0801	1316.6645
		2232.1231	
		2388.2241	
BLAST DATNDQVTK	496.2411	517.1894	876.4474
		645.2480	805.4103
		744.3164	704.3626
		845.3641	590.3197
BLAST DAAEAIK	359.1954	973.4591	
		387.1516	602.3560
		458.1887	531.3189
		571.2728	460.2918
BLAST SPVGTIR	372.7065	699.3678	
		457.2047	657.3731
		570.2887	550.3204
		726.3898	446.2774
			389.2560

....

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

Protein ID (EXPASy): (Example: P04632)

↓

Targeted Metabolomics & Proteomics Laboratory

Home MRMPath Useful Links

Click [here](#) to download the Excel sheet

Disclaimer | Web Page Policy | Contact Us

Mass Spectrometry Tools— EXPASY

<http://www.expasy.org/resources/search/keywords:mass%20spectrometry>

The screenshot displays a list of five mass spectrometry tools from the EXPASY database. Each tool entry includes a title, categories, a brief description, and keywords. The tools are: FindPept, GlycoMod, GlycoSuiteDB, HCD/CID spectra merger, and IsotopIdent. Each entry is marked with a green checkmark in the top right corner.

- FindPept** [SIB: Swiss-Prot group]
 - Categories: proteomics, (protein sequences and identification, **mass spectrometry** and 2-DE data, protein characterisation and function) - Software type(s), website - tool
 - Identify peptides that result from unspecific cleavage of proteins from their experimental masses, taking into account artifactual chemical modifications, post-translational modifications (PTM) and protease autolytic cleavage.
 - keywords: **mass spectrometry**, molecular weight (MW), peptide mass fingerprinting, post-translational modification, protein identification
- GlycoMod** [SIB: Swiss-Prot group]
 - Categories: proteomics, (protein sequences and identification, **mass spectrometry** and 2-DE data, post-translational modification) - Software type(s), website - tool
 - Predict possible oligosaccharide structures that occur on proteins from their experimentally determined masses. The program can be used for free or derivatized oligosaccharides and for glycopeptides.
 - keywords: glycomics, glycosylation, glycotool, **mass spectrometry**, molecular weight (MW), oligosaccharide, peptide mass fingerprinting, post-translational modification, protein characterization, PTM prediction, sequence analysis, sequence characterisation, sugar epitope
- GlycoSuiteDB** [SIB: Proteome Informatics group]
 - Categories: proteomics, (post-translational modification) - Software type(s), website - database
 - GlycoSuiteDB is a curated and annotated glycan database.
 - keywords: database searching, disease, glycomics, glycosylation, knowledge resource, **mass spectrometry**, oligosaccharide, post-translational modification, sugar epitope
- HCD/CID spectra merger** [SIB: Swiss-Prot group] [download]
 - Categories: proteomics, (**mass spectrometry** and 2-DE data) - Software type(s), website - tool
 - A tool to merge the peptide sequence-ion m/z range from CID spectra and the reporter-ion m/z range from HCD spectra into the appropriate single file, to be further used in identification and quantification search engines
 - keywords: **mass spectrometry**, protein quantitation
- IsotopIdent** [SIB: Swiss-Prot group]
 - Categories: proteomics, (**mass spectrometry** and 2-DE data), genomics, (characterisation/annotation) - Software type(s), website - tool
 - IsotopIdent can estimate the theoretical isotopic distribution of a peptide or protein, a polynucleotide and a chemical compound from its composition (sequence of amino acids expressed in either 3-letter code, sequence of amino acids expressed in 3-letter code, sequence of nucleotides or its chemical formula). IsotopIdent can also compute its monoisotopic mass, and predict the most likely isotope combination and the exact mass of the given input.
 - keywords: mass search, **mass spectrometry**, molecular weight (MW), physico-chemical property

Interesting Papers—Mass Spectrometry and Bioinformatics

- http://masspec.scripps.edu/publications/public_pdf/72_art.pdf
- <http://www.sciencedirect.com/science/article/pii/S0014579309002208>
- <http://www.ingentaconnect.com/content/ben/cbio/2012/00000007/0000001/art00010>

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