Web-based Bioinformatics Applications

Joint Lecture GBS 722 and BMG 744

Chiquito Crasto

Department of Genetics, UAB

chiquito@uab.edu

February 1, 2012

Philosophical underpinnings …

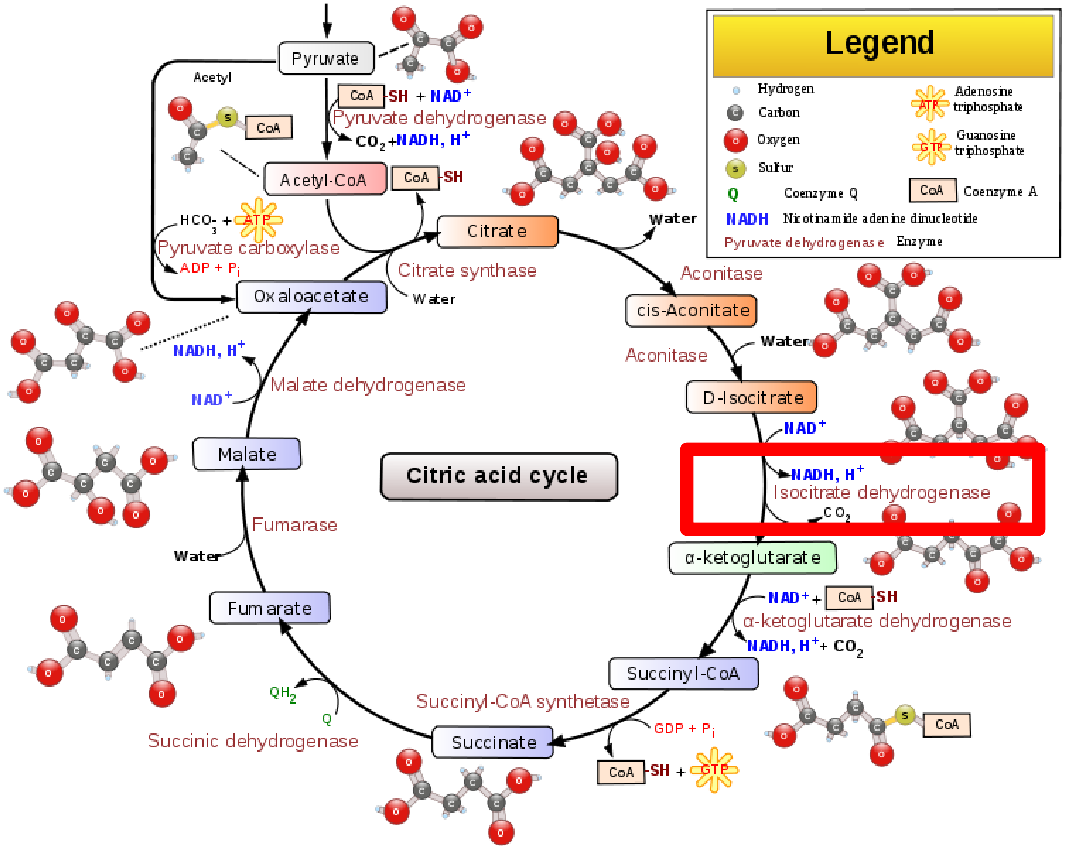
- Bioinformatics is here to stay—simply because computers are part of everyday life. This is not going to change in the near or distant future
- Students, researchers, etc., will be better served embracing bioinformatics ideas even if they do not necessarily want to pursue bioinformatics-driven careers, 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

Interoperability & Database Accessibility

- Interoperability: the ability of systems to interoperate, that is exchange information in meaningful ways without having to reproduce information
- Database Accessibility: Access relevant information that is stored within the database architecture ONCE, but accessed and presented from different sources

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

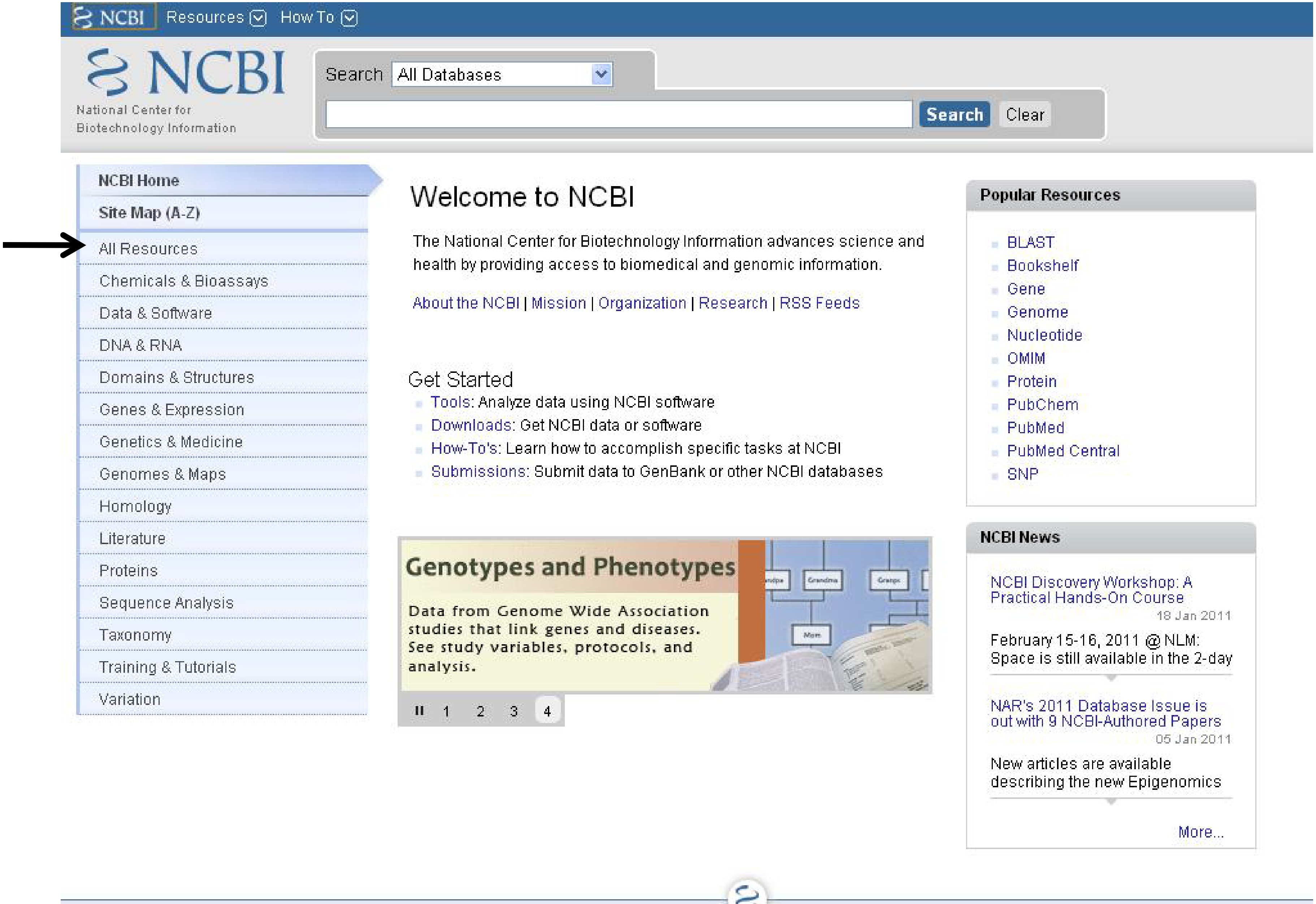
- **Isocitrate dehydrogenase** (EC 1.1.1.42) and (EC 1.1.1.41), also known as **IDH**, is an enzyme that participates in the citric acid cycle. It catalyzes the third step of the cycle: the oxidative decarboxylation of isocitrate, producing alpha-ketoglutarate (α-ketoglutarate) and CO_2 while converting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



Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

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

Genotypes and Phenotypes

Data from Genome Wide Association studies that link genes and diseases. See study variables, protocols, and analysis.

NCBI News

NCBI Discovery Workshop: A Practical Hands-On Course
18 Jan 2011

February 15-16, 2011 @ NLM: Space is still available in the 2-day

NAR's 2011 Database Issue is out with 9 NCBI-Authored Papers
05 Jan 2011

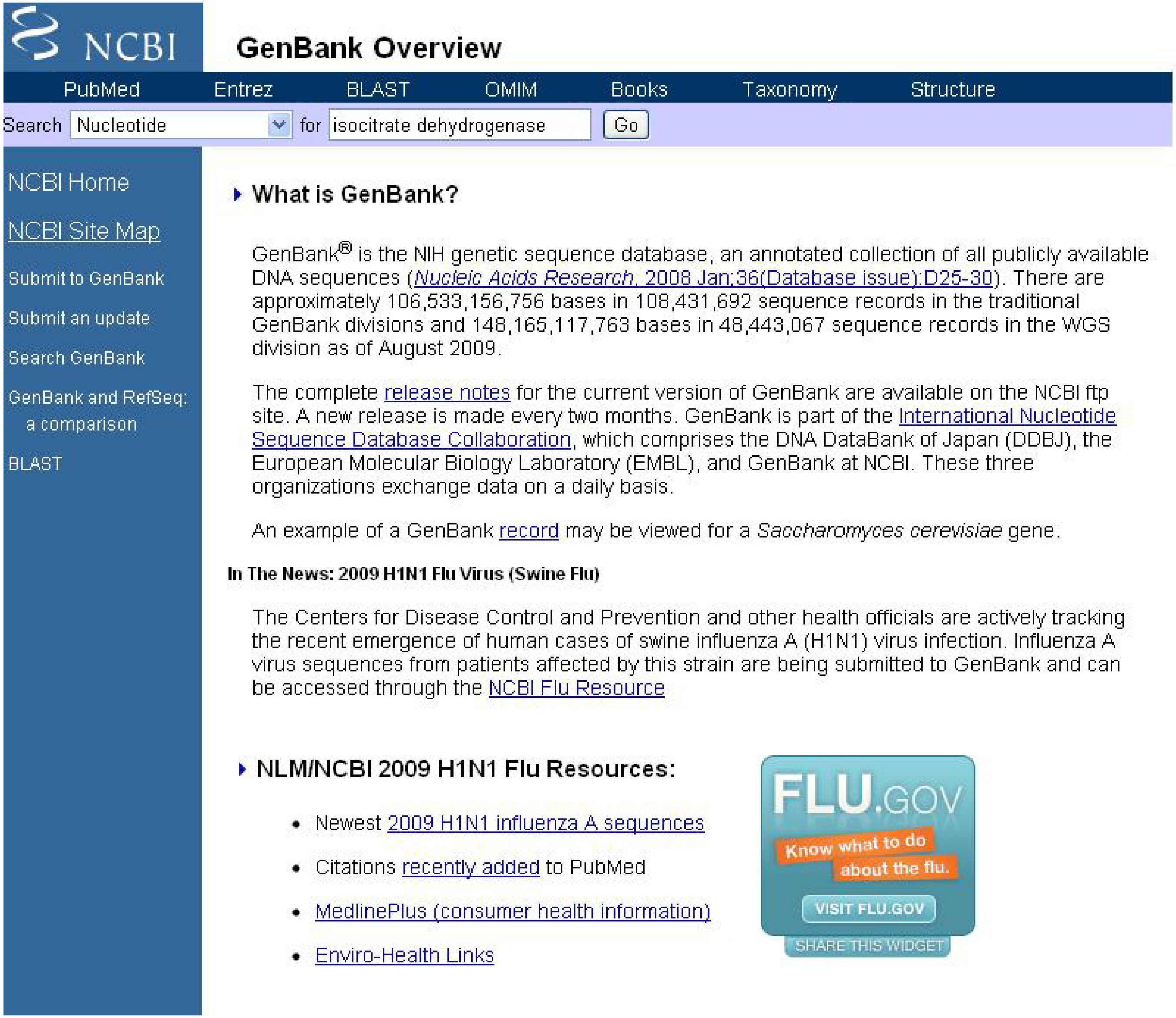
New articles are available describing the new Epigenomics

More...

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 Overview

Search Nucleotide for isocitrate dehydrogenase Go

What is GenBank?

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.

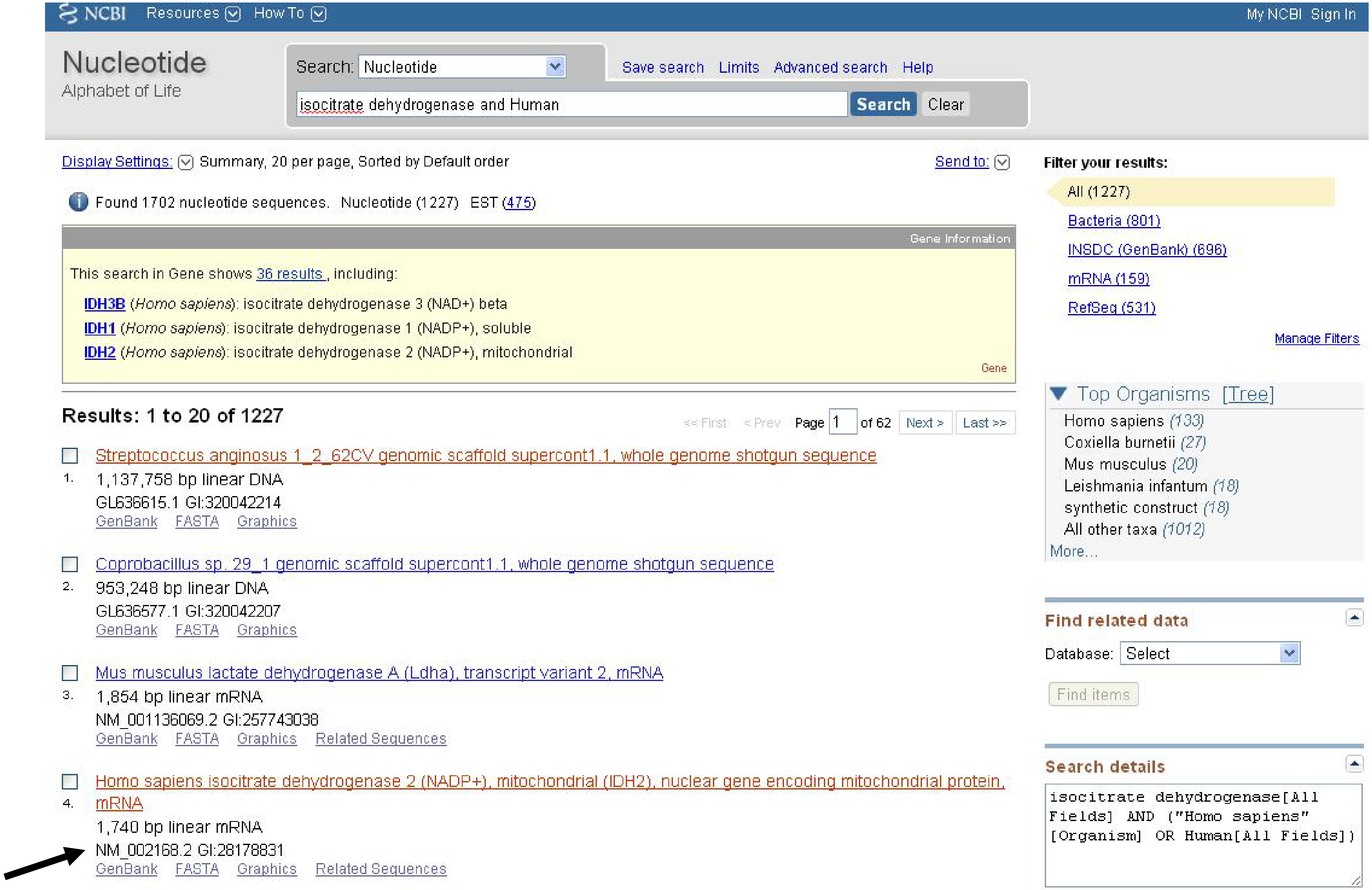
In The News: 2009 H1N1 Flu Virus (Swine Flu)

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

Nucleotide-Genbank’s gene repository



Search: Nucleotide

isocitrate dehydrogenase and Human

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

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

This search in Gene shows 36 results , including:

IDH3B (*Homo sapiens*): isocitrate dehydrogenase 3 (NAD+) beta

IDH1 (*Homo sapiens*): isocitrate dehydrogenase 1 (NADP+), soluble

IDH2 (*Homo sapiens*): isocitrate dehydrogenase 2 (NADP+), mitochondrial

Results: 1 to 20 of 1227

Page 1 of 62

1. ☐ Streptococcus anginosus 1_2_62CV genomic scaffold supercont1.1, whole genome shotgun sequence
1,137,758 bp linear DNA
GL636615.1 GI:320042214
GenBank FASTA Graphics

2. ☐ Coprobacillus sp. 29_1 genomic scaffold supercont1.1, whole genome shotgun sequence
953,248 bp linear DNA
GL636577.1 GI:320042207
GenBank FASTA Graphics

3. ☐ Mus musculus lactate dehydrogenase A (Ldha), transcript variant 2, mRNA
1,854 bp linear mRNA
NM_001136069.2 GI:257743038
GenBank FASTA Graphics 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 Graphics Related Sequences

Filter your results:

- All (1227)
- Bacteria (801)
- INSDC (GenBank) (696)
- mRNA (159)
- RefSeq (531)

Top Organisms [Tree]

Homo sapiens	*(133)*
Coxiella burnetii	*(27)*
Mus musculus	*(20)*
Leishmania infantum	*(18)*
synthetic construct	*(18)*
All other taxa	*(1012)*

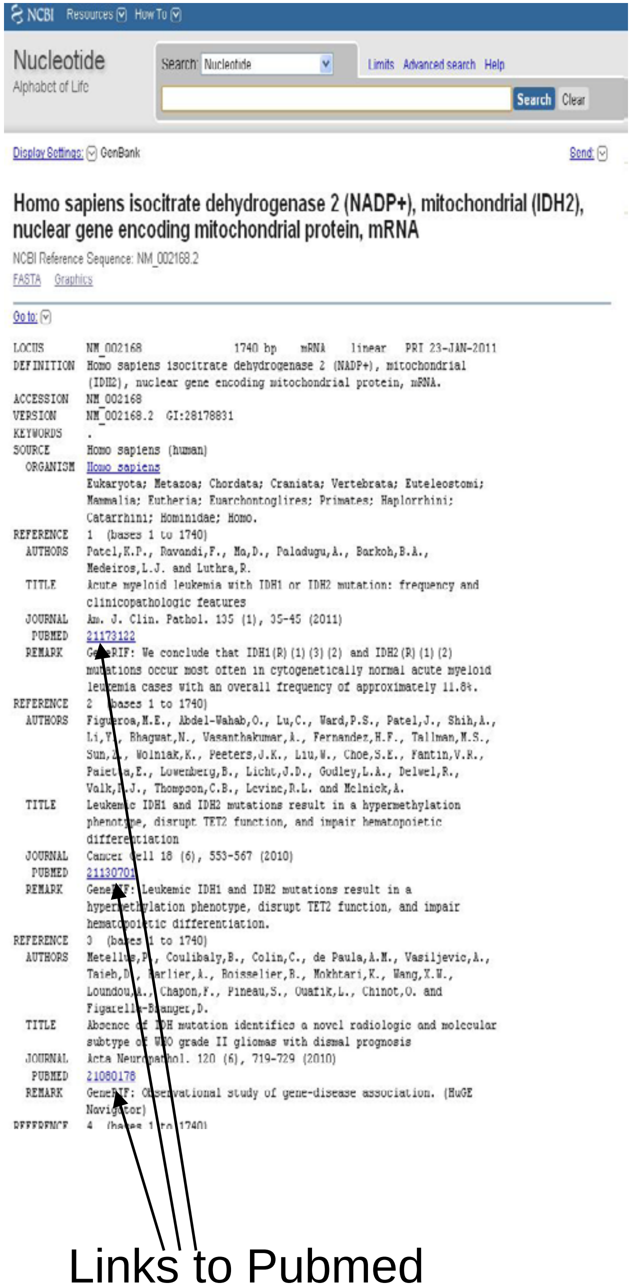
More...

Database: Select

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

Accession Number

A Nucleotide Entry in Genbank



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

NCBI Reference Sequence: NM_002168.2

```
LOCUS       NM_002168                1740 bp    mRNA    linear   PRI 23-JAN-2011
DEFINITION  Homo sapiens isocitrate dehydrogenase 2 (NADP+), mitochondrial
            (IDH2), nuclear gene encoding mitochondrial protein, mRNA.
ACCESSION   NM_002168
VERSION     NM_002168.2  GI:28178831
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1740)
  AUTHORS   Patel,K.P., Ravandi,F., Ma,D., Paladugu,A., Barkoh,B.A.,
            Medeiros,L.J. and Luthra,R.
  TITLE     Acute myeloid leukemia with IDH1 or IDH2 mutation: frequency and
            clinicopathologic features
  JOURNAL   Am. J. Clin. Pathol. 135 (1), 35-45 (2011)
   PUBMED   21173122
  REMARK    GeneRIF: We conclude that IDH1(R)(1)(3)(2) and IDH2(R)(1)(2)
            mutations occur most often in cytogenetically normal acute myeloid
            leukemia cases with an overall frequency of approximately 11.8%.
REFERENCE   2  (bases 1 to 1740)
  AUTHORS   Figueroa,M.E., Abdel-Wahab,O., Lu,C., Ward,P.S., Patel,J., Shih,A.,
            Li,Y., Bhagwat,N., Vasanthakumar,A., Fernandez,H.F., Tallman,M.S.,
            Sun,Z., Wolniak,K., Peeters,J.K., Liu,W., Choe,S.E., Fantin,V.R.,
            Paietta,E., Lowenberg,B., Licht,J.D., Godley,L.A., Delwel,R.,
            Valk,P.J., Thompson,C.B., Levine,R.L. and Melnick,A.
  TITLE     Leukemic IDH1 and IDH2 mutations result in a hypermethylation
            phenotype, disrupt TET2 function, and impair hematopoietic
            differentiation
  JOURNAL   Cancer Cell 18 (6), 553-567 (2010)
   PUBMED   21130701
  REMARK    GeneRIF: Leukemic IDH1 and IDH2 mutations result in a
            hypermethylation phenotype, disrupt TET2 function, and impair
            hematopoietic differentiation.
REFERENCE   3  (bases 1 to 1740)
  AUTHORS   Metellus,P., Coulibaly,B., Colin,C., de Paula,A.M., Vasiljevic,A.,
            Taieb,D., Barlier,A., Boisselier,B., Mokhtari,K., Wang,X.W.,
            Loundou,A., Chapon,F., Pineau,S., Ouafik,L., Chinot,O. and
            Figarella-Branger,D.
  TITLE     Absence of IDH mutation identifies a novel radiologic and molecular
            subtype of WHO grade II gliomas with dismal prognosis
  JOURNAL   Acta Neuropathol. 120 (6), 719-729 (2010)
   PUBMED   21080178
  REMARK    GeneRIF: Observational study of gene-disease association. (HuGE
            Navigator)
REFERENCE   4  (bases 1 to 1740)
```

```
                     /inference="alignment:Splign"
                     /number=1
     CDS             87..1445
                     /gene="IDH2"
                     /gene_synonym="D2HGA2; ICD-M; IDH; IDHM; IDP; IDPM;
                     mNADP-IDH"
                     /EC_number="1.1.1.42"
                     /note="oxalosuccinate decarboxylase; NADP(+)-specific
                     ICDH"
                     /codon_start=1
                     /product="isocitrate dehydrogenase [NADP], mitochondrial
                     precursor"
                     /protein_id="NP_002159.2"
                     /db_xref="GI:28178832"
                     /db_xref="CCDS:CCDS10359.1"
                     /db_xref="GeneID:3418"
                     /db_xref="HGNC:5383"
                     /db_xref="HPRD:00973"
                     /db_xref="MIM:147650"
                     /translation="MAGYLRVVRSLCRASGSRPAWAPAALTAPTSQEQPRRHYADKRI
                     KVAKPVVEMDGDEMTRIIWQFIKEKLILPHVDIQLKYFDLGLPNRDQTDDQVTIDSAL
                     ATQKYSVAVKCATITPDEARVEEFKLKKMWKSPNGTIRNILGGTVFREPIICKNIPRL
                     VPGWTKPITIGRHAHGDQYKATDFVADRAGTFKMVFTPKDGSGVKEWEVYNFPAGGVG
                     MGMYNTDESISGFAHSCFQYAIQKKWPLYMSTKNTILKAYDGRFKDIFQEIFDKHYKT
                     DFDKNKIWYEHRLIDDMVAQVLKSSGGFVWACKNYDGDVQSDILAQGFGSLGLMTSVL
                     VCPDGKTIEAEAAHGTVTRHYREHQKGRPTSTNPIASIFAWTRGLEHRGKLDGNQDLI
                     RFAQMLEKVCVETVESGAMTKDLAGCIHGLSNVKLNEHFLNTTDFLDTIKSNLDRALG
                     RQ"
```

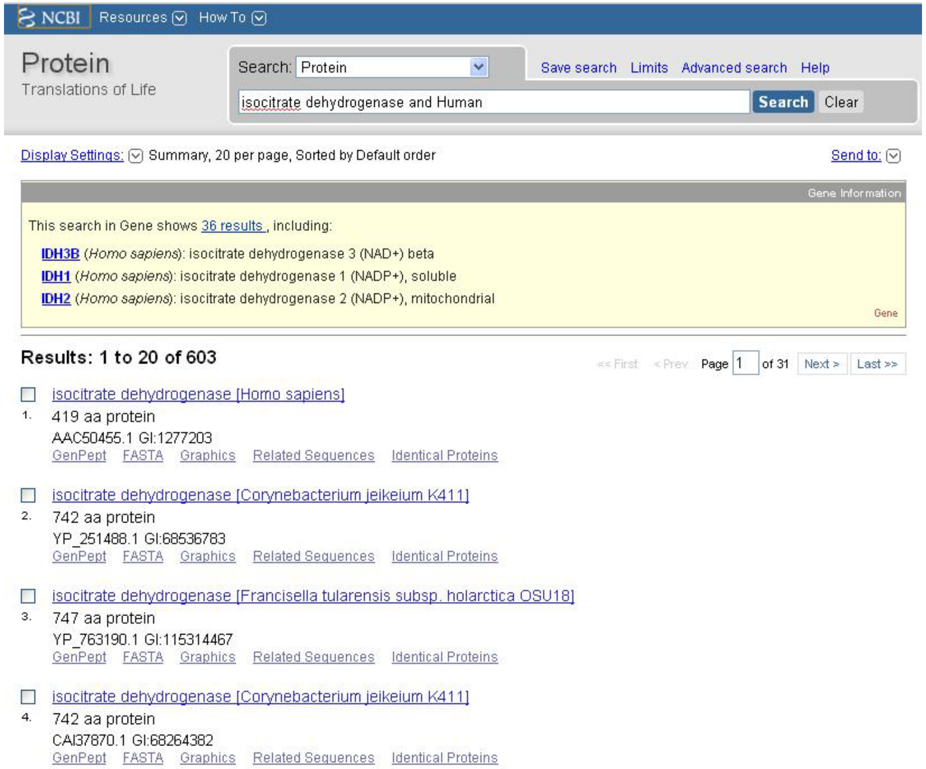
Protein Sequence

```
                     mNADP-IDH"
ORIGIN
        1 ccagcgttag cccgcggcca ggcagccggg aggagcggcg cgcgctcgga cctctcccgc
       61 cctgctcgtt cgctctccag cttgggatgg ccggctacct gcgggtcgtg cgctcgctct
      121 gcagagcctc aggctcgcgg ccggcctggg cgccggcggc cctgacagcc cccacctcgc
      181 aagagcagcc gcggcgccac tatgccgaca aaaggatcaa ggtggcgaag cccgtggtgg
      241 agatggatgg tgatgagatg acccgtatta tctggcagtt catcaaggag aagctcatcc
      301 tgccccacgt ggacatccag ctaaagtatt ttgacctcgg gctcccaaac cgtgaccaga
      361 ctgatgacca ggtcaccatt gactctgcac tggccaccca gaagtacagt gtggctgtca
      421 agtgtgccac catcacccct gatgaggccc gtgtggaaga gttcaagctg aagaagatgt
      481 ggaaaagtcc caatggaact atccggaaca tcctgggggg gactgtcttc cgggagccca
      541 tcatctgcaa aaacatccca cgcctagtcc ctggctggac caagcccatc accattggca
      601 ggcacgccca tggcgaccag tacaaggcca cagactttgt ggcagaccgg gccggcactt
      661 tcaaaatggt cttcacccca aaagatggca gtggtgtcaa ggagtgggaa gtgtacaact
      721 tccccgcagg cggcgtgggc atgggcatgt acaacaccga cgagtccatc tcaggttttg
      781 cgcacagctg cttccagtat gccatccaga agaaatggcc gctgtacatg agcaccaaga
      841 acaccatact gaaagcctac gatgggcgtt tcaaggacat cttccaggag atctttgaca
      901 agcactataa gaccgacttc gacaagaata agatctggta tgagcaccgg ctcattgatg
      961 acatggtggc tcaggtcctc aagtcttcgg gtggctttgt gtgggcctgc aagaactatg
     1021 acggagatgt gcagtcagac atcctggccc agggctttgg ctcccttggc ctgatgacgt
     1081 ccgtcctggt ctgccctgat gggaagacga ttgaggctga ggccgctcat gggaccgtca
     1141 cccgccacta tcgggagcac cagaagggcc ggcccaccag caccaacccc atcgccagca
     1201 tctttgcctg gacacgtggc ctggagcacc gggggaagct ggatgggaac caagacctca
     1261 tcaggtttgc ccagatgctg gagaaggtgt gcgtggagac ggtggagagt ggagccatga
     1321 ccaaggacct ggcgggctgc attcacggcc tcagcaatgt gaagctgaac gagcacttcc
     1381 tgaacaccac ggacttcctc gacaccatca agagcaacct ggacagagcc ctgggcaggc
     1441 agtaggggga ggcgccaccc atggctgcag tggaggggcc agggctgagc cggcgggtcc
     1501 tcctgagcgc ggcagagggt gagcctcaca gcccctctct ggaggccttt ctaggggatg
     1561 tttttttata agccagatgt tttaaaagc atatgtgtgt ttcccctcat ggtgacgtga
     1621 ggcaggagca gtgcgtttta cctcagccag tcagtatgtt ttgcatactg taatttatat
     1681 tgcccttgga acacatggtg ccatatttag ctactaaaaa gctcttcaca aaaaaaaaaa
//
```

Links to Pubmed

Gene Sequence

Protein Sequence in Genbank (isocitrate dehydrogenase)



Summary, 20 per page, Sorted by Default order

This search in Gene shows 36 results , including:

- **IDH3B** (*Homo sapiens*): isocitrate dehydrogenase 3 (NAD+) beta
- **IDH1** (*Homo sapiens*): isocitrate dehydrogenase 1 (NADP+), soluble
- **IDH2** (*Homo sapiens*): isocitrate dehydrogenase 2 (NADP+), mitochondrial

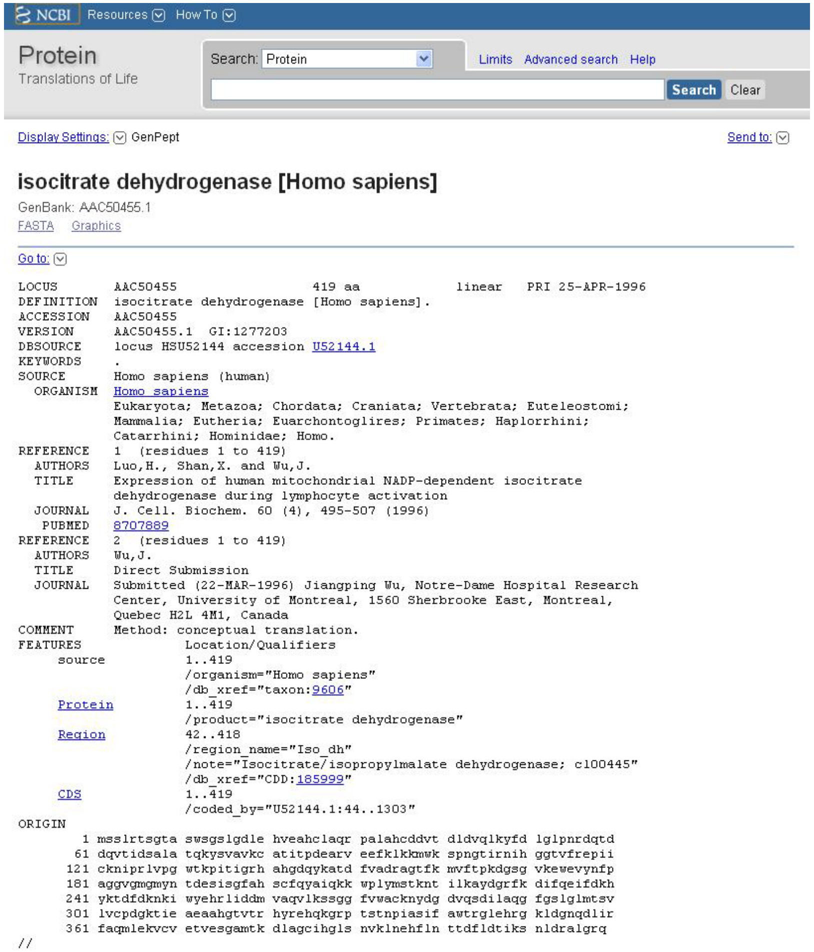
Results: 1 to 20 of 603

☐ 1. isocitrate dehydrogenase [Homo sapiens]
419 aa protein
AAC50455.1 GI:1277203
GenPept FASTA Graphics Related Sequences Identical Proteins

☐ 2. isocitrate dehydrogenase [Corynebacterium jeikeium K411]
742 aa protein
YP_251488.1 GI:68536783
GenPept FASTA Graphics Related Sequences Identical Proteins

☐ 3. isocitrate dehydrogenase [Francisella tularensis subsp. holarctica OSU18]
747 aa protein
YP_763190.1 GI:115314467
GenPept FASTA Graphics Related Sequences Identical Proteins

☐ 4. isocitrate dehydrogenase [Corynebacterium jeikeium K411]
742 aa protein
CAI37870.1 GI:68264382
GenPept FASTA Graphics Related Sequences Identical Proteins



GenPept

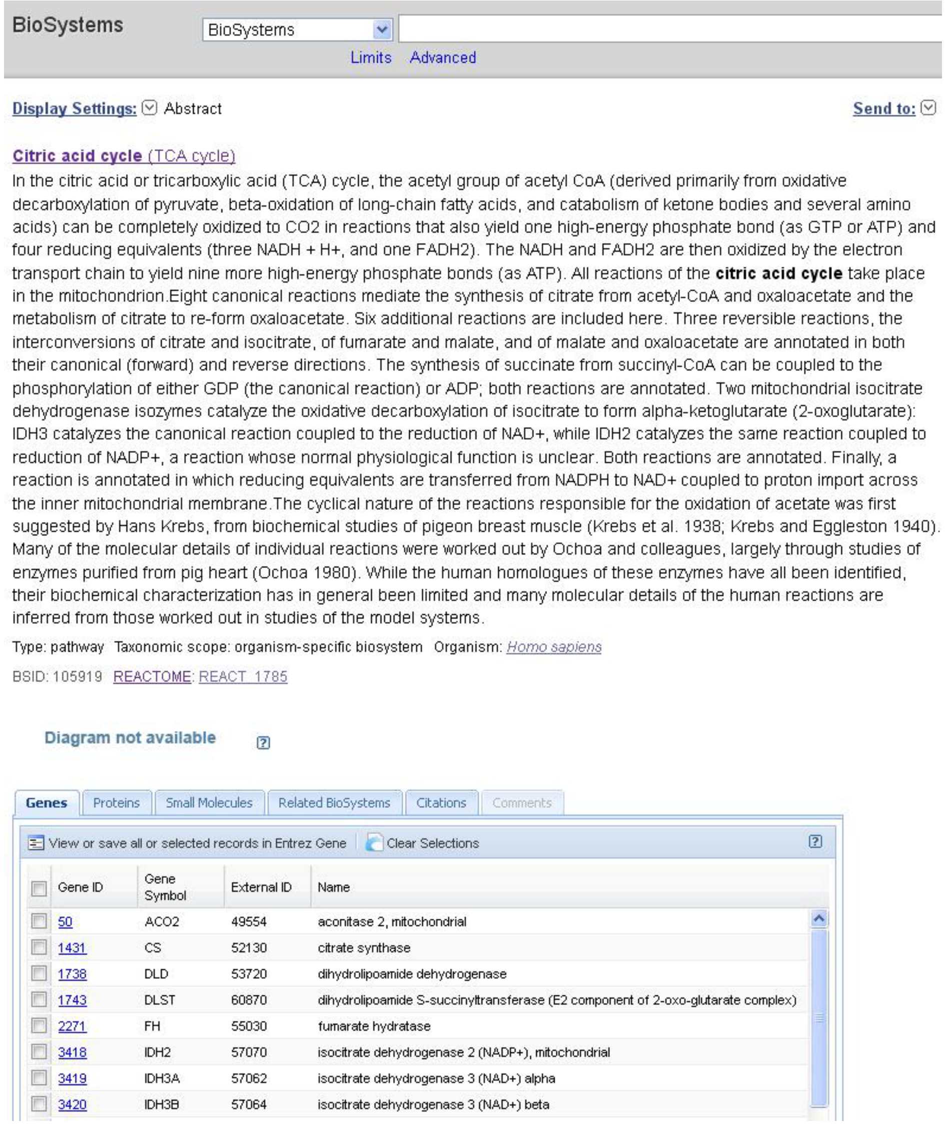
isocitrate dehydrogenase [Homo sapiens]

GenBank: AAC50455.1

```
LOCUS       AAC50455                 419 aa            linear   PRI 25-APR-1996
DEFINITION  isocitrate dehydrogenase [Homo sapiens].
ACCESSION   AAC50455
VERSION     AAC50455.1  GI:1277203
DBSOURCE    locus HSU52144 accession U52144.1
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 419)
  AUTHORS   Luo,H., Shan,X. and Wu,J.
  TITLE     Expression of human mitochondrial NADP-dependent isocitrate
            dehydrogenase during lymphocyte activation
  JOURNAL   J. Cell. Biochem. 60 (4), 495-507 (1996)
   PUBMED   8707889
REFERENCE   2  (residues 1 to 419)
  AUTHORS   Wu,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) Jiangping Wu, Notre-Dame Hospital Research
            Center, University of Montreal, 1560 Sherbrooke East, Montreal,
            Quebec H2L 4M1, Canada
COMMENT     Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..419
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     Protein         1..419
                     /product="isocitrate dehydrogenase"
     Region          42..418
                     /region_name="Iso_dh"
                     /note="Isocitrate/isopropylmalate dehydrogenase; cl00445"
                     /db_xref="CDD:185999"
     CDS             1..419
                     /coded_by="U52144.1:44..1303"
ORIGIN
        1 msslrtsgta swsgslgdle hveahclaqr palahcddvt dldvqlkyfd lglpnrdqtd
       61 dqvtidsala tqkysvavkc atitpdearv eefklkkmwk spngtirnih ggtvfrepii
      121 ckniprlvpg wtkpitigrh ahgdqykatd fvadragtfk mvftpkdgsg vkewevynfp
      181 aggvgmgmyn tdesisgfah scfqyaiqkk wplymstknt ilkaydgrfk difqeifdkh
      241 yktdfdknki wyehrliddm vaqvlkssgg fvwacknydg dvqsdilaqg fgslglmtsv
      301 lvcpdgktie aeaahgtvtr hyrehqkgrp tstnpiasif awtrglehrg kldgnqdlir
      361 faqmlekvcv etvesgamtk dlagcihgls nvklnehfln ttdfldtiks nldralgrq
//
```

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.

BioSystems



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 cyclical nature of the reactions responsible for the oxidation of acetate was first suggested by Hans Krebs, from biochemical studies of pigeon breast muscle (Krebs et al. 1938; Krebs and Eggleston 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.

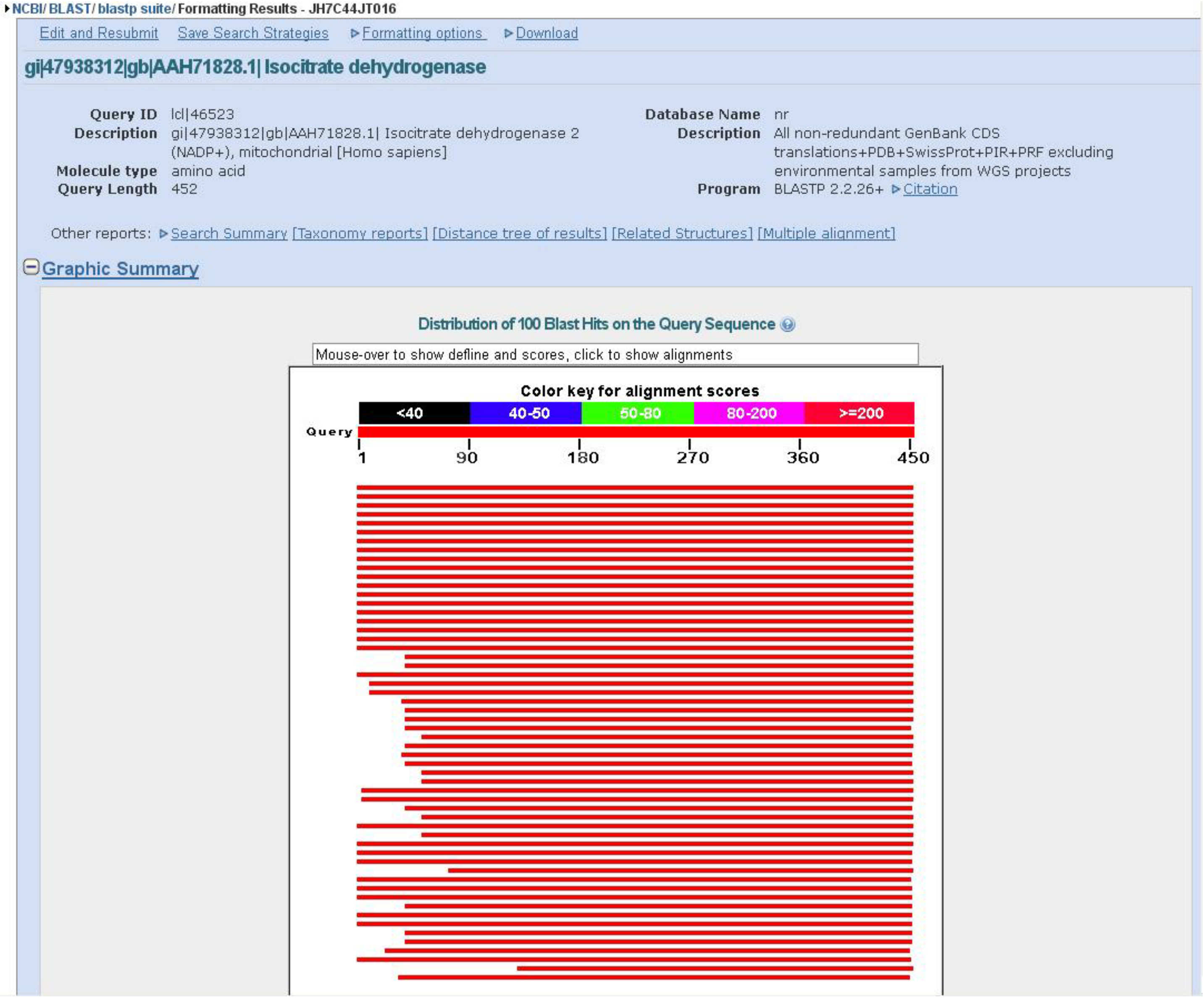
Type: pathway   Taxonomic scope: organism-specific biosystem   Organism: *Homo sapiens*

BSID: 105919   REACTOME: REACT_1785

Diagram not available

☐	Gene ID	Gene Symbol	External ID	Name
☐	50	ACO2	49554	aconitase 2, mitochondrial
☐	1431	CS	52130	citrate synthase
☐	1738	DLD	53720	dihydrolipoamide dehydrogenase
☐	1743	DLST	60870	dihydrolipoamide S-succinyltransferase (E2 component of 2-oxo-glutarate complex)
☐	2271	FH	55030	fumarate hydratase
☐	3418	IDH2	57070	isocitrate dehydrogenase 2 (NADP+), mitochondrial
☐	3419	IDH3A	57062	isocitrate dehydrogenase 3 (NAD+) alpha
☐	3420	IDH3B	57064	isocitrate dehydrogenase 3 (NAD+) beta

BLAST Results



NCBI/ BLAST/ blastp suite/ Formatting Results - JH7C44JT016

gi|47938312|gb|AAH71828.1| Isocitrate dehydrogenase

Query ID lcl|46523
Description gi|47938312|gb|AAH71828.1| Isocitrate dehydrogenase 2 (NADP+), mitochondrial [Homo sapiens]
Molecule type amino acid
Query Length 452

Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTP 2.2.26+ Citation

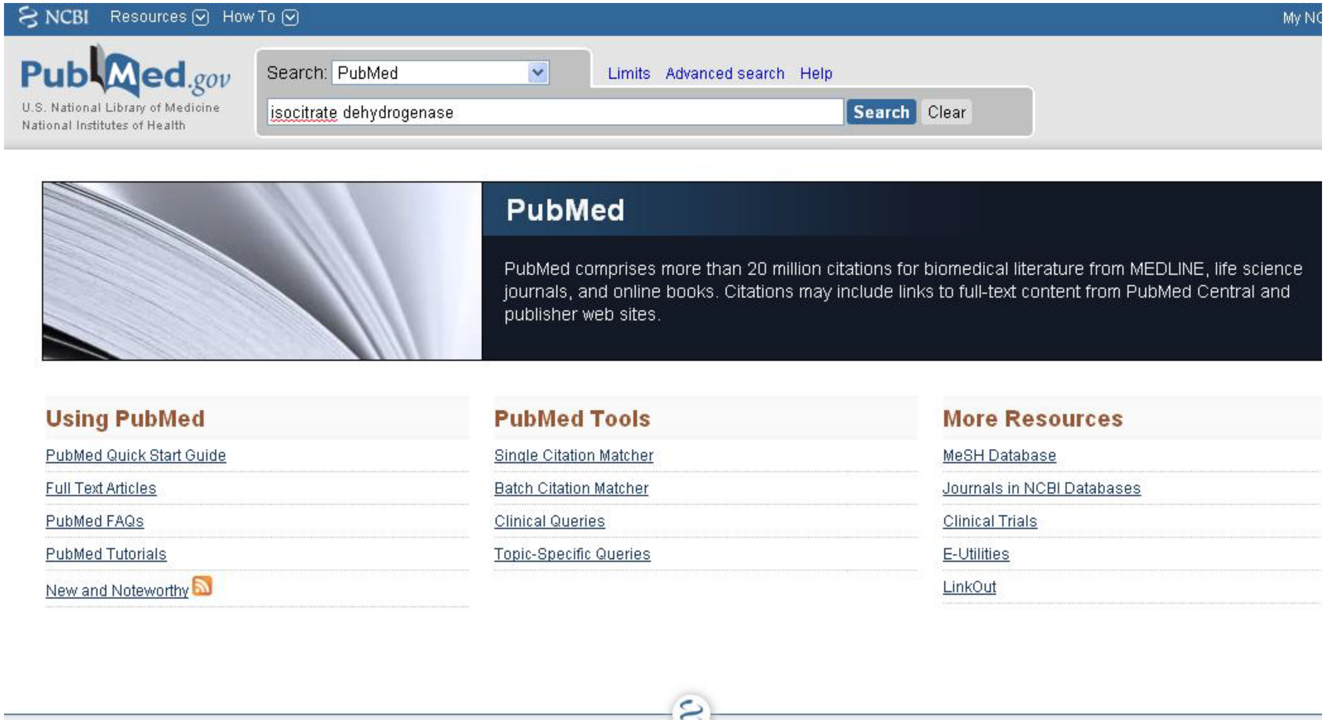
Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Related Structures] [Multiple alignment]

Graphic Summary

Distribution of 100 Blast Hits on the Query Sequence

Mouse-over to show defline and scores, click to show alignments

Pubmed—repository of biomedical abstracts



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

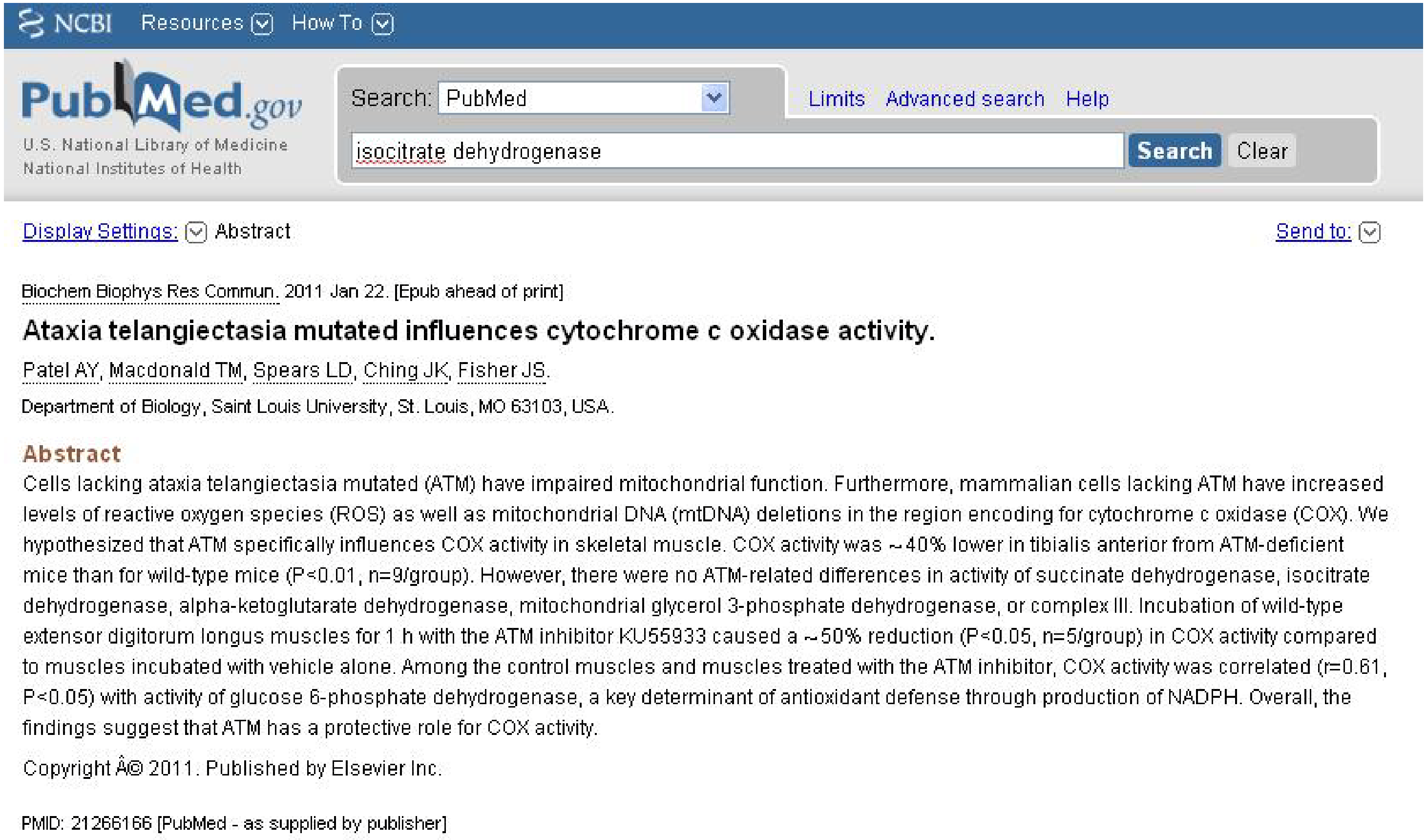
- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources

- MeSH Database
- Journals in NCBI Databases
- Clinical Trials
- E-Utilities
- 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



isocitrate dehydrogenase

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.

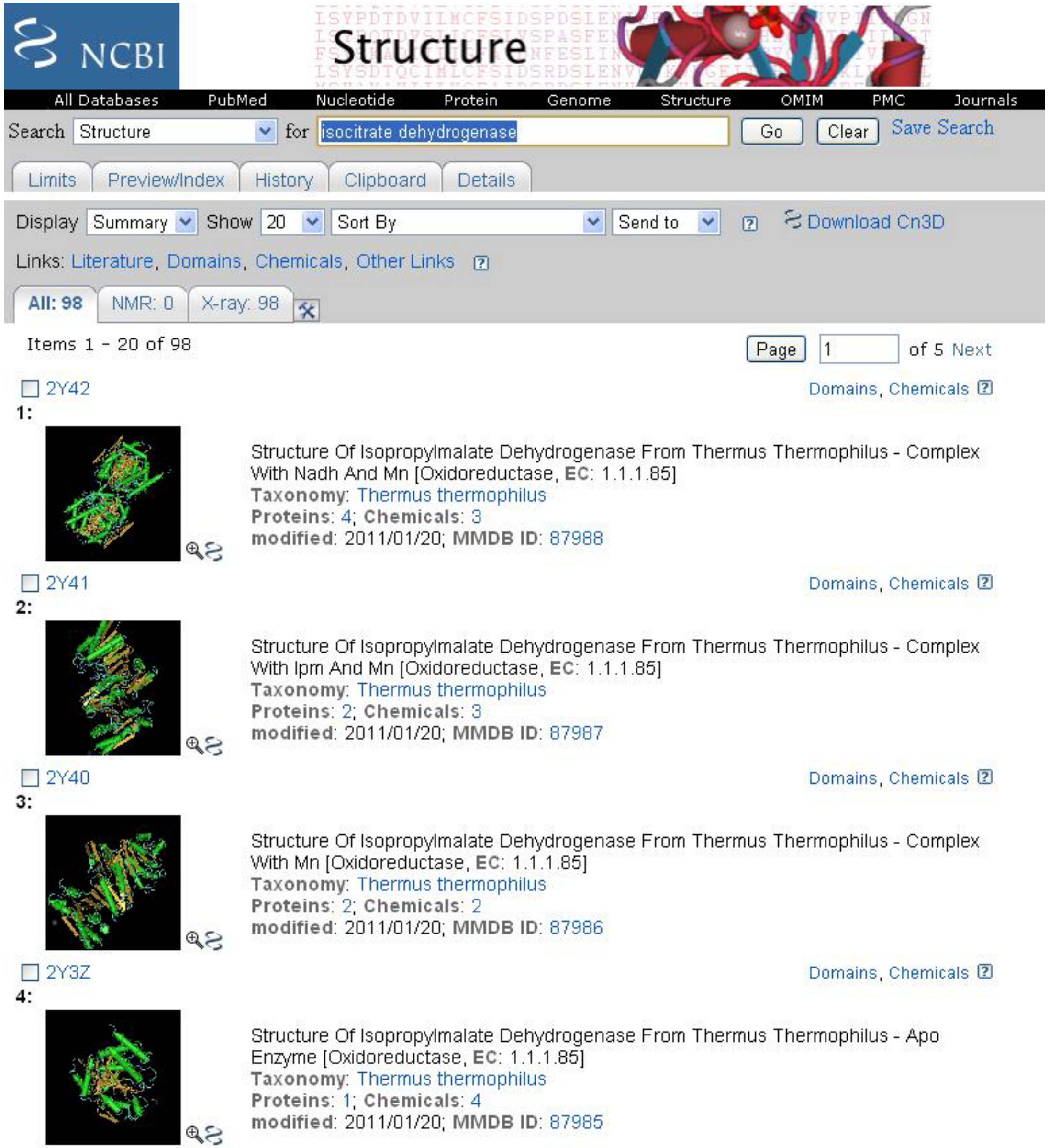
Copyright Â© 2011. Published by Elsevier Inc.

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



Search Structure for isocitrate dehydrogenase

Display Summary Show 20

Links: Literature, Domains, Chemicals, Other Links

All: 98 NMR: 0 X-ray: 98

Items 1 - 20 of 98

Page 1 of 5

☐ 2Y42 Domains, Chemicals

1: 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

☐ 2Y41 Domains, Chemicals

2: 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

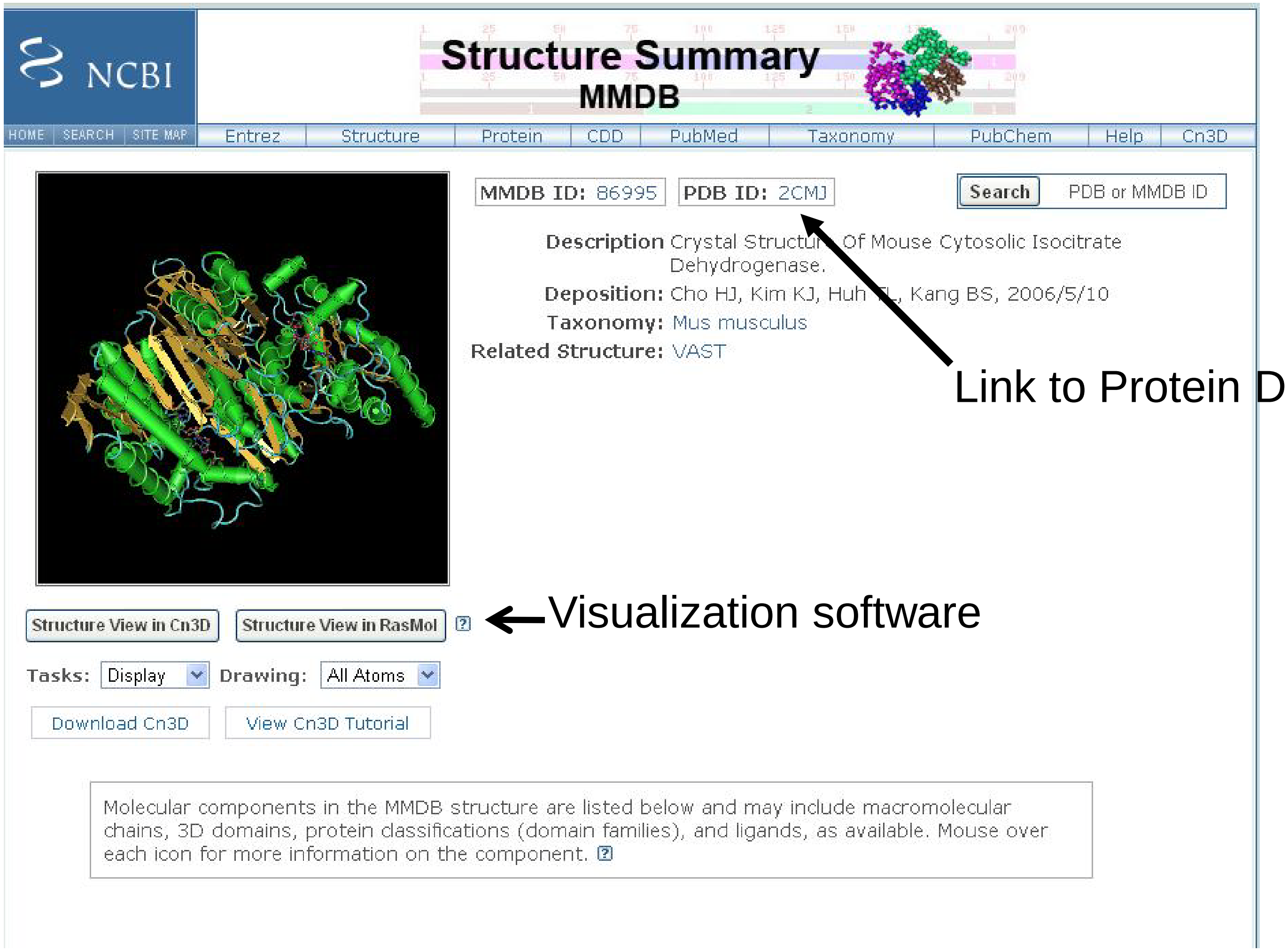
☐ 2Y40 Domains, Chemicals

3: 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

☐ 2Y3Z Domains, Chemicals

4: 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



MMDB ID: 86995 **PDB ID:** 2CMJ

Description Crystal Structure Of Mouse Cytosolic Isocitrate Dehydrogenase.

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

Taxonomy: Mus musculus

Related Structure: VAST

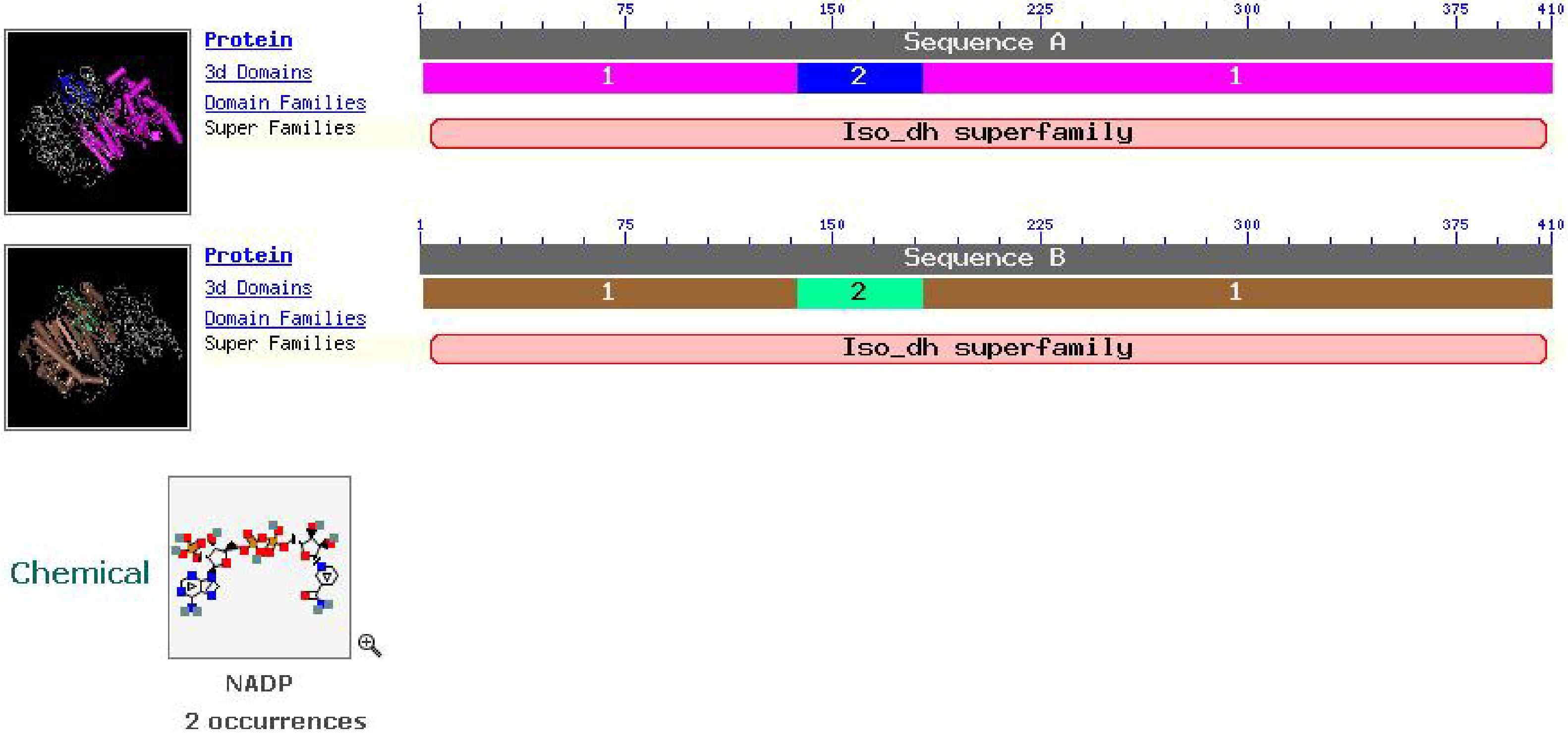
Tasks: Display **Drawing:** All Atoms

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.

Link to Protein Databank

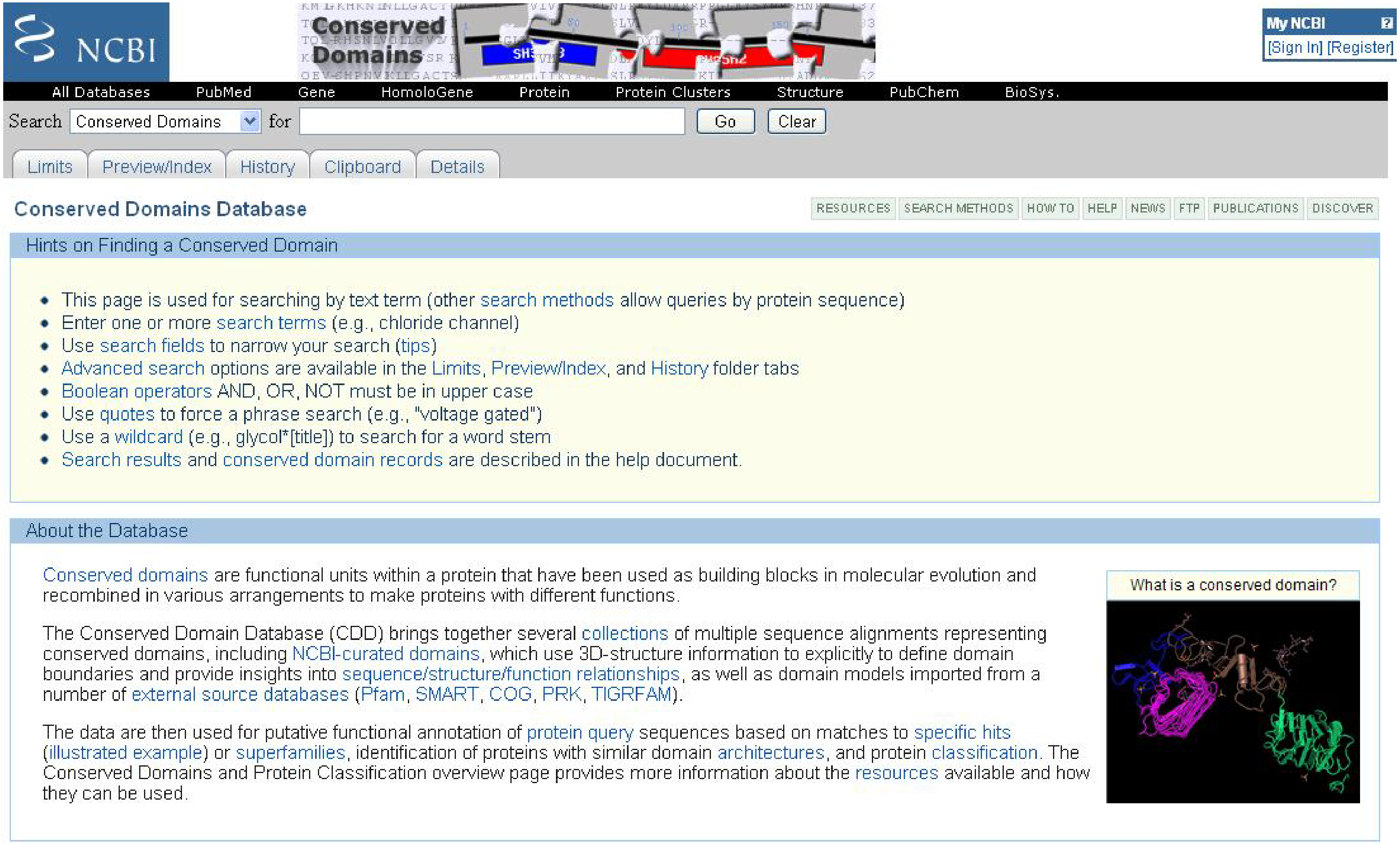
Visualization software

Structure of Domains in Genbank



List of domains related to
or associated with
Isocitrate Dehydrogenase

Conserved domain database (CDD) in Genbank



Conserved Domains Database

Hints on Finding a Conserved Domain

- 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., glycol*[title]) to search for a word stem
- Search results and conserved domain records are described in the help document.

About the Database

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.

What is a conserved domain?

CDD …

Search Conserved Domains for isocitrate dehydrogenase

Display Summary Show 20 Sort By Send to

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

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

Items 1 - 20 of 24

Page 1 of 2 Next

1: ☐ cl01891 — Related CDs, Literature, Sequence, Structure, Other Links

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 — Related CDs, Literature, Sequence, Structure, Other Links

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 — Literature, Sequence, Structure, Other Links

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 — Sequence, Structure, Other Links

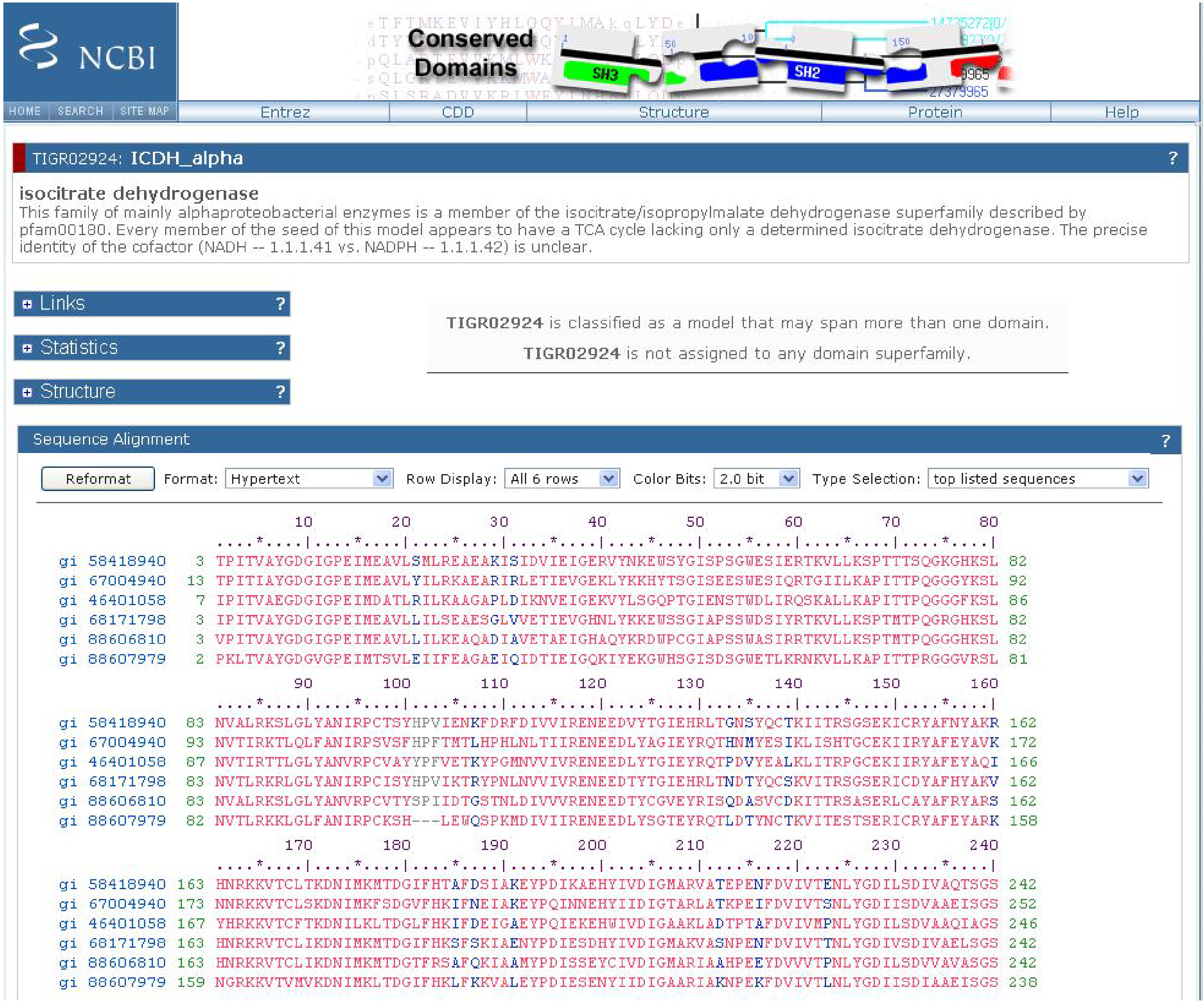
ICDH_alpha: isocitrate dehydrogenase
This family of mainly alphaproteobacterial enzymes is a member of the isocitrate/isopropylmalate dehydrogenase superfamily described by pfam00180. Every member of the seed of this model appears to have a TCA cycle lacking only a determined isocitrate dehydrogenase. The precise identity of the cofactor (NADH -- 1.1.1.41 vs. NADPH -- 1.1.1.42) is unclear. [163075]

5: ☐ TIGR00178 — Literature, Sequence, Structure, Other Links

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, Rhodomicrobium vannielii, and Neisseria meningitidis. It is NADP-specific. [129282]

☐ TIGR00183 — Related CDs, Sequence, Structure, Other Links

CDD…



TIGR02924: **ICDH_alpha**

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.

TIGR02924 is classified as a model that may span more than one domain.

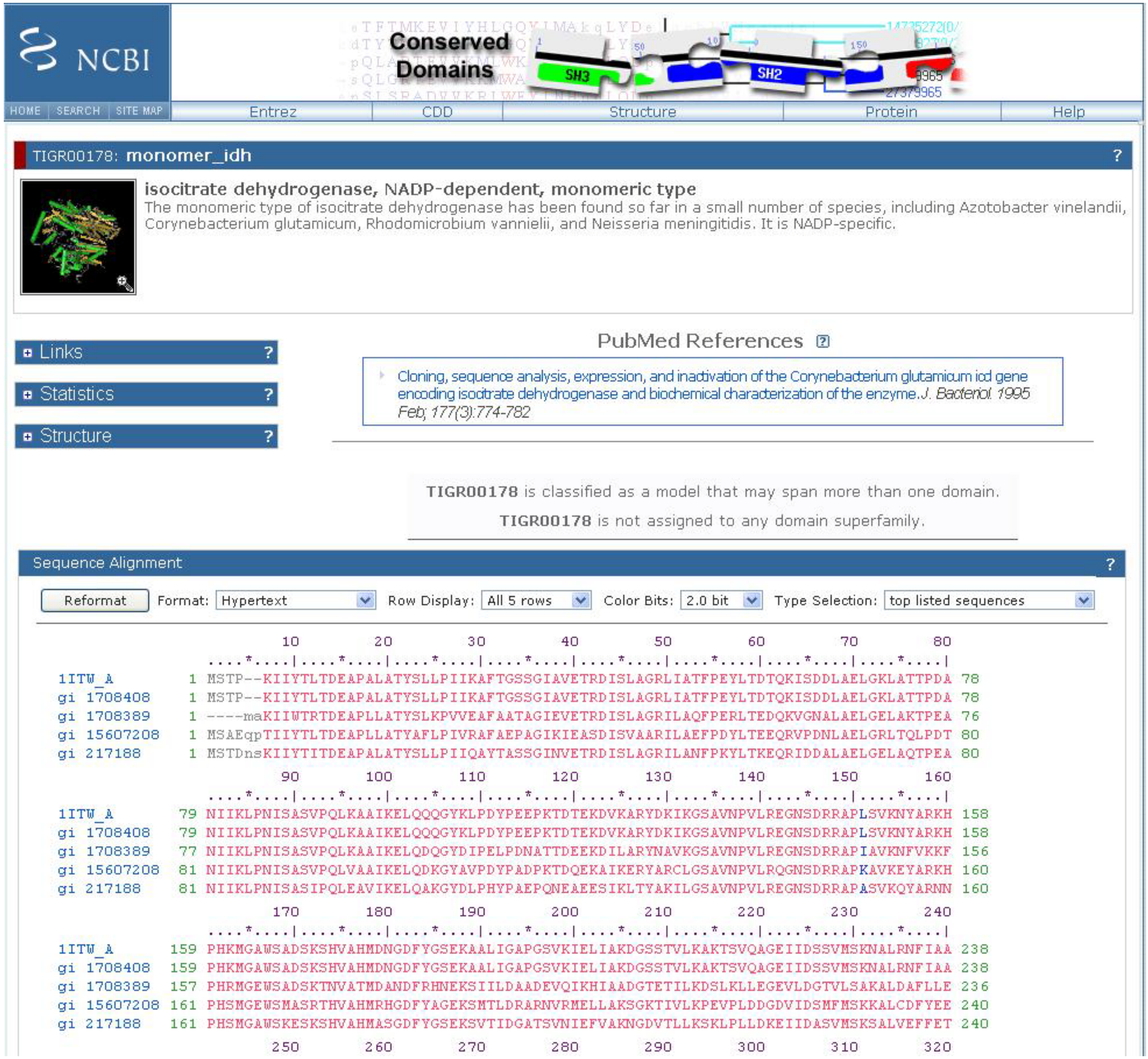
TIGR02924 is not assigned to any domain superfamily.

Sequence Alignment

Format: Hypertext
Row Display: All 6 rows
Color Bits: 2.0 bit
Type Selection: top listed sequences

```
                        10        20        30        40        50        60        70        80
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 58418940   3 TPITVAYGDGIGPEIMEAVLSMLREAEAKISIDVIEIGERVYNKEWSYGISPSGWESIERTKVLLKSPTTTSQGKGHKSL 82
gi 67004940  13 TPITIAYGDGIGPEIMEAVLYILRKAEARIRLETIEVGEKLYKKHYTSGISEESWESIQRTGIILKAPITTPQGGGYKSL 92
gi 46401058   7 IPITVAEGDGIGPEIMDATLRILKAAGAPLDIKNVEIGEKVYLSGQPTGIENSTWDLIRQSKALLKAPITTPQGGGFKSL 86
gi 68171798   3 IPITVAYGDGIGPEIMEAVLLILSEAESGLVVETIEVGHNLYKKEWSSGIAPSSWDSIYRTKVLLKSPTMTPQGRGHKSL 82
gi 88606810   3 VPITVAYGDGIGPEIMEAVLLILKEAQADIAVETAEIGHAQYKRDWPCGIAPSSWASIRRTKVLLKSPTMTPQGGGHKSL 82
gi 88607979   2 PKLTVAYGDGVGPEIMTSVLEIIFEAGAEIQIDTIEIGQKIYEKGWHSGISDSGWETLKRNKVLLKAPITTPRGGVRSL 81
                        90       100       110       120       130       140       150       160
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 58418940  83 NVALRKSLGLYANIRPCTSYHPVIENKFDRFDIVVIRENEEDVYTGIEHRLTGNSYQCTKIITRSGSEKICRYAFNYAKR 162
gi 67004940  93 NVTIRKTLQLFANIRPSVSFHPFTMTLHPHLNLTIIRENEEDLYAGIEYRQTHNMYESIKLISHTGCEKIIRYAFEYAVK 172
gi 46401058  87 NVTIRTTLGLYANVRPCVAYYPFVETKYPGMNVVIVRENEEDLYTGIEYRQTPDVYEALKLITRPGCEKIIRYAFEYAQI 166
gi 68171798  83 NVTLRKRLGLYANIRPCISYHPVIKTRYPNLNVVIVRENEEDTYTGIEHRLTNDTYQCSKVITRSGSERICDYAFHYAKV 162
gi 88606810  83 NVALRKSLGLYANVRPCVTYSPIIDTGSTNLDIVVVRENEEDTYCGVEYRISQDASVCDKITTRSASERLCAYAFRYARS 162
gi 88607979  82 NVTLRKKLGLFANIRPCKSH---LEWQSPKMDIVIIRENEEDLYSGTEYRQTLDTYNCTKVITESTSERICRYAFEYARK 158
                       170       180       190       200       210       220       230       240
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 58418940 163 HNRKKVTCLTKDNIMKMTDGIFHTAFDSIAKEYPDIKAEHYIVDIGMARVATEPENFDVIVTENLYGDILSDIVAQTSGS 242
gi 67004940 173 NNRKKVTCLSKDNIMKFSDGVFHKIFNEIAKEYPQINNEHYIIDIGTARLATKPEIFDVIVTSNLYGDIISDVAAEISGS 252
gi 46401058 167 YHRKKVTCFTKDNILKLTDGLFHKIFDEIGAEYPQIEKEHWIVDIGAAKLADTPTAFDVIVMPNLYGDILSDVAAQIAGS 246
gi 68171798 163 HNRKRVTCLIKDNIMKMTDGIFHKSFSKIAENYPDIESDHYIVDIGMAKVASNPENFDVIVTTNLYGDIVSDIVAELSGS 242
gi 88606810 163 HNRKRVTCLIKDNIMKMTDGTFRSAFQKIAAMYPDISSEYCIVDIGMARIAAHPEEYDVVVTPNLYGDILSDVVAVASGS 242
gi 88607979 159 NGRKKVTVMVKDNIMKLTDGIFHKLFKKVALEYPDIESENYIIDIGAARIAKNPEKFDVIVTLNLYGDIISDIAAEISGS 238
```

CDD … if structure is available



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, Rhodomicrobium vannielii, and Neisseria meningitidis. It is NADP-specific.

PubMed References

Cloning, sequence analysis, expression, and inactivation of the Corynebacterium 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.

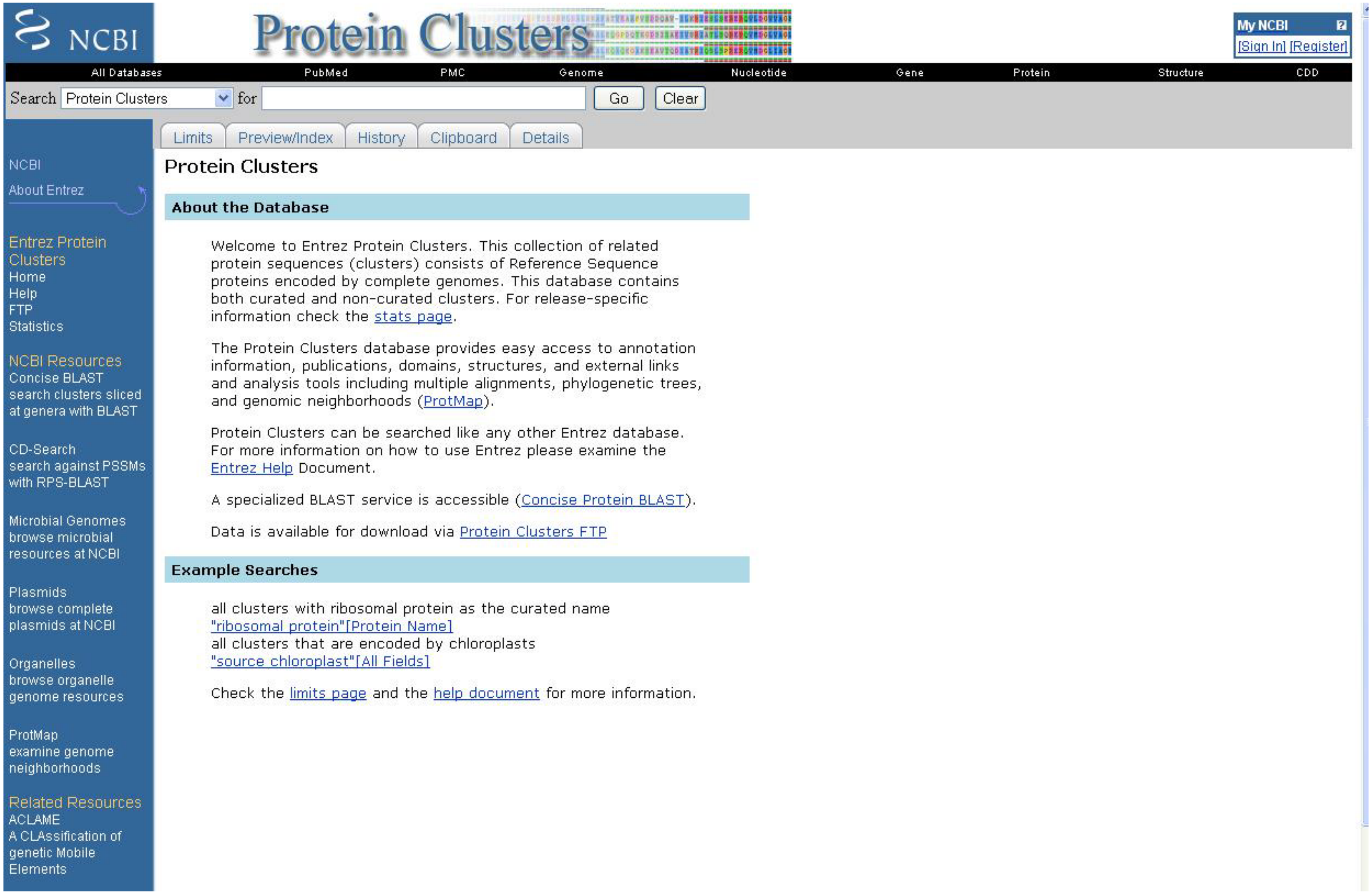
TIGR00178 is not assigned to any domain superfamily.

Sequence Alignment

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
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
1ITW_A        1 MSTP--KIIYTLTDEAPALATYSLLPIIKAFTGSSGIAVETRDISLAGRLIATFPEYLTDTQKISDDLAELGKLATTPDA 78
gi 1708408    1 MSTP--KIIYTLTDEAPALATYSLLPIIKAFTGSSGIAVETRDISLAGRLIATFPEYLTDTQKISDDLAELGKLATTPDA 78
gi 1708389    1 ----maKIIWTRTDEAPLLATYSLKPVVEAFAATAGIEVETRDISLAGRILAQFPERLTEDQKVGNALAELGELAKTPEA 76
gi 15607208   1 MSAEqpTIIYTLTDEAPLLATYAFLPIVRAFAEPAGIKIEASDISVAARILAEFPDYLTEEQRVPDNLAELGRLTQLPDT 80
gi 217188     1 MSTDnsKIIYTITDEAPALATYSLLPIIQAYTASSGINVETRDISLAGRILANFPKYLTKEQRIDDALAELGELAQTPEA 80
                       90       100       110       120       130       140       150       160
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
1ITW_A       79 NIIKLPNISASVPQLKAAIKELQQQGYKLPDYPEEPKTDTEKDVKARYDKIKGSAVNPVLREGNSDRRAPLSVKNYARKH 158
gi 1708408   79 NIIKLPNISASVPQLKAAIKELQQQGYKLPDYPEEPKTDTEKDVKARYDKIKGSAVNPVLREGNSDRRAPLSVKNYARKH 158
gi 1708389   77 NIIKLPNISASVPQLKAAIKELQDQGYDIPELPDNATTDEEKDILARYNAVKGSAVNPVLREGNSDRRAPIAVKNFVKKF 156
gi 15607208  81 NIIKLPNISASVPQLVAAIKELQDKGYAVPDYPADPKTDQEKAIKERYARCLGSAVNPVLRQGNSDRRAPKAVKEYARKH 160
gi 217188    81 NIIKLPNISASIPQLEAVIKELQAKGYDLPHYPAEPQNEAEESIKLTYAKILGSAVNPVLREGNSDRRAPASVKQYARNN 160
                      170       180       190       200       210       220       230       240
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
1ITW_A      159 PHKMGAWSADSKSHVAHMDNGDFYGSEKAALIGAPGSVKIELIAKDGSSTVLKAKTSVQAGEIIDSSVMSKNALRNFIAA 238
gi 1708408  159 PHKMGAWSADSKSHVAHMDNGDFYGSEKAALIGAPGSVKIELIAKDGSSTVLKAKTSVQAGEIIDSSVMSKNALRNFIAA 238
gi 1708389  157 PHRMGEWSADSKTNVATMDANDFRHNEKSIILDAADEVQIKHIAADGTETILKDSLKLLEGEVLDGTVLSAKALDAFLLE 236
gi 15607208 161 PHSMGEWSMASRTHVAHMRHGDFYAGEKSMTLDRARNVRMELLAKSGKTIVLKPEVPLDDGDVIDSMFMSKKALCDFYEE 240
gi 217188   161 PHSMGAWSKESKSHVAHMASGDFYGSEKSVTIDGATSVNIEFVAKNGDVTLLKSKLPLLDKEIIDASVMSKSALVEFFET 240
                      250       260       270       280       290       300       310       320
```

Clustering Proteins in terms of Sequence Similarities--Genbank



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. This 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 (ProtMap).

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).

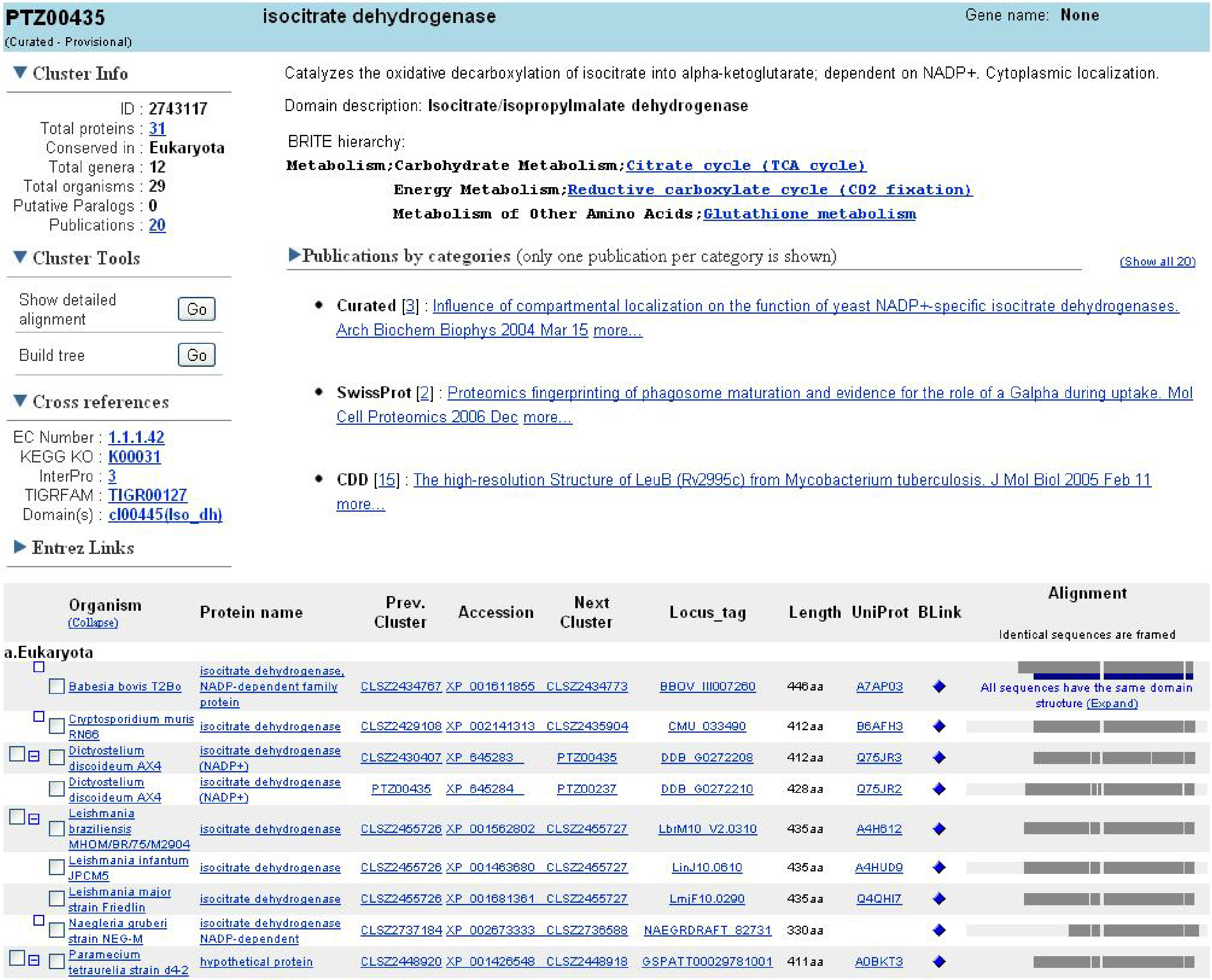
Data is available for download via Protein Clusters FTP

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 - Provisional)

▼ Cluster Info

ID : 2743117
Total proteins : 31
Conserved in : Eukaryota
Total genera : 12
Total organisms : 29
Putative Paralogs : 0
Publications : 20

▼ Cluster Tools

▼ Cross references

EC Number : 1.1.1.42
KEGG KO : K00031
InterPro : 3
TIGRFAM : TIGR00127
Domain(s) : cl00445(Iso_dh)

▶ Entrez Links

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

Domain description: **Isocitrate/isopropylmalate dehydrogenase**

BRITE hierarchy:

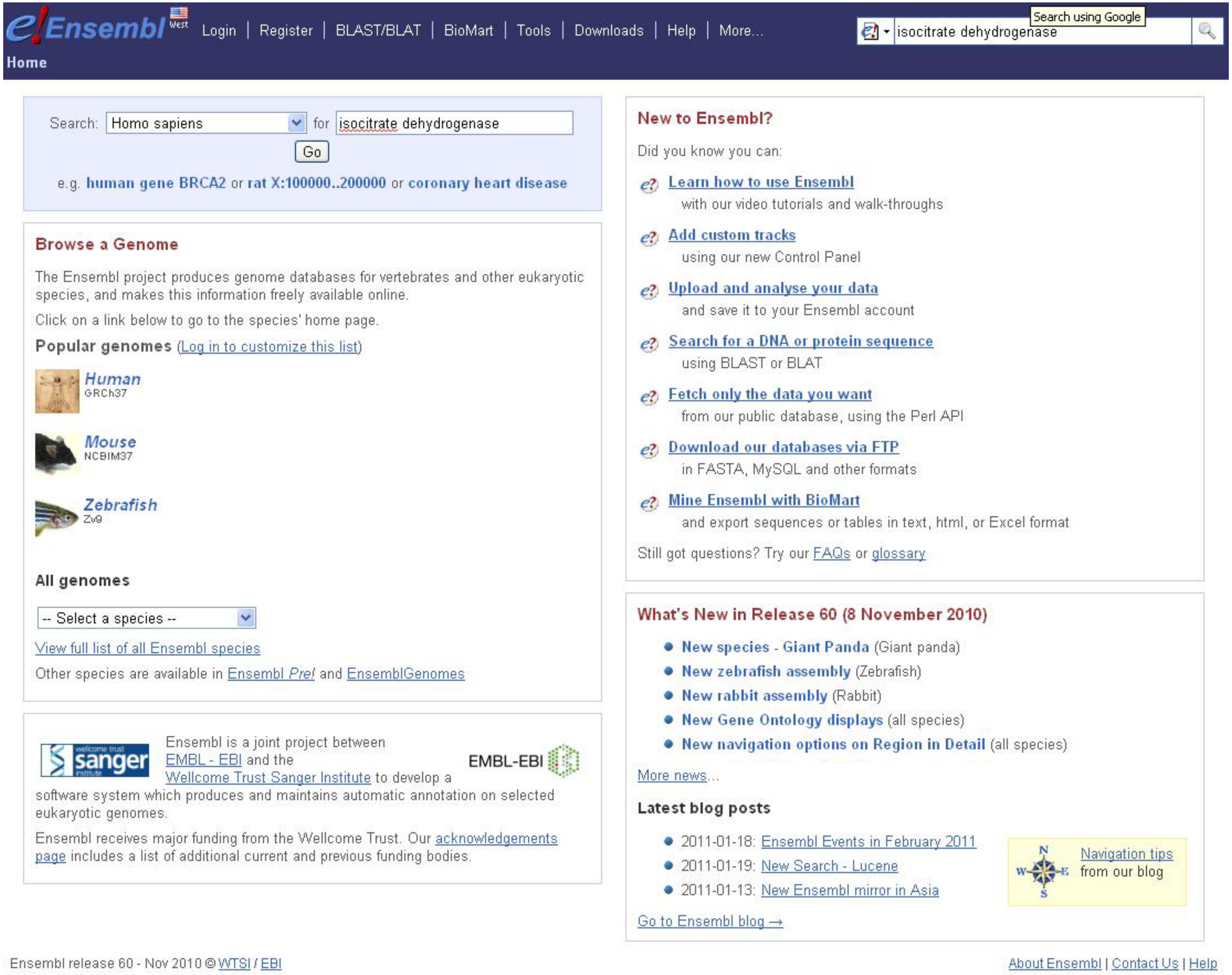
Metabolism;Carbohydrate Metabolism;Citrate cycle (TCA cycle)
Energy Metabolism;Reductive carboxylate cycle (CO2 fixation)
Metabolism of Other Amino Acids;Glutathione metabolism

▶ 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 LeuB (Rv2995c) from Mycobacterium tuberculosis. J Mol Biol 2005 Feb 11 more...

Organism (Collapse)	Protein name	Prev. Cluster	Accession	Next Cluster	Locus_tag	Length	UniProt	BLink	Alignment Identical sequences are framed
a.Eukaryota									
☐ Babesia bovis T2Bo	isocitrate dehydrogenase, NADP-dependent family protein	CLSZ2434767	XP_001611855	CLSZ2434773	BBOV_III007260	446aa	A7AP03	◆	All sequences have the same domain structure (Expand)
☐ Cryptosporidium muris RN66	isocitrate dehydrogenase	CLSZ2429108	XP_002141313	CLSZ2435904	CMU_033490	412aa	B6AFH3	◆	
☐ ⊟ ☐ Dictyostelium discoideum AX4	isocitrate dehydrogenase (NADP+)	CLSZ2430407	XP_645283	PTZ00435	DDB_G0272208	412aa	Q75JR3	◆	
☐ Dictyostelium discoideum AX4	isocitrate dehydrogenase (NADP+)	PTZ00435	XP_645284	PTZ00237	DDB_G0272210	428aa	Q75JR2	◆	
☐ ⊟ ☐ Leishmania braziliensis MHOM/BR/75/M2904	isocitrate dehydrogenase	CLSZ2455726	XP_001562802	CLSZ2455727	LbrM10_V2.0310	435aa	A4H612	◆	
☐ Leishmania infantum JPCM5	isocitrate dehydrogenase	CLSZ2455726	XP_001463680	CLSZ2455727	LinJ10.0610	435aa	A4HUD9	◆	
☐ Leishmania major strain Friedlin	isocitrate dehydrogenase	CLSZ2455726	XP_001681361	CLSZ2455727	LmjF10.0290	435aa	Q4QHI7	◆	
☐ Naegleria gruberi strain NEG-M	isocitrate dehydrogenase NADP-dependent	CLSZ2737184	XP_002673333	CLSZ2736588	NAEGRDRAFT_82731	330aa		◆	
☐ ⊟ ☐ Paramecium tetraurelia strain d4-2	hypothetical protein	CLSZ2448920	XP_001426548	CLSZ2448918	GSPATT00029781001	411aa	A0BKT3	◆	

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



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 GRCh37

Mouse NCBIM37

Zebrafish Zv9

All genomes

-- Select a species --

View full list of all Ensembl species

Other species are available in Ensembl *Pre!* and EnsemblGenomes

Ensembl is a joint project between EMBL - EBI and the Wellcome Trust Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes.

Ensembl receives major funding from the Wellcome Trust. Our acknowledgements page includes a list of additional current and previous funding bodies.

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 - Lucene
- 2011-01-13: New Ensembl mirror in Asia

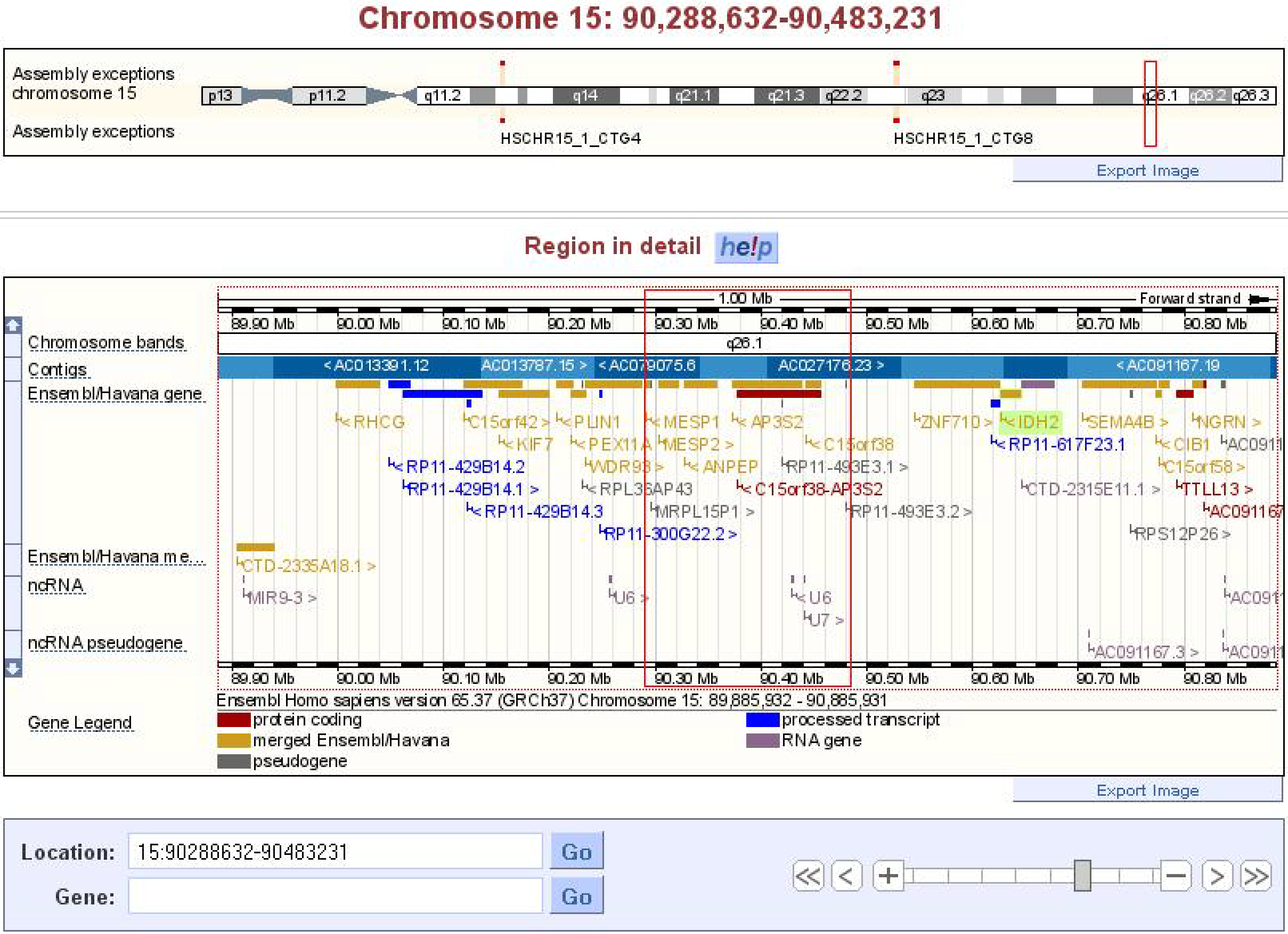
Navigation tips from our blog

Go to Ensembl blog →

Ensembl release 60 - Nov 2010 © WTSI / EBI

About Ensembl | Contact Us | Help

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



Ensembl Homo sapiens version 65.37 (GRCh37) Chromosome 15: 89,885,932 - 90,885,931

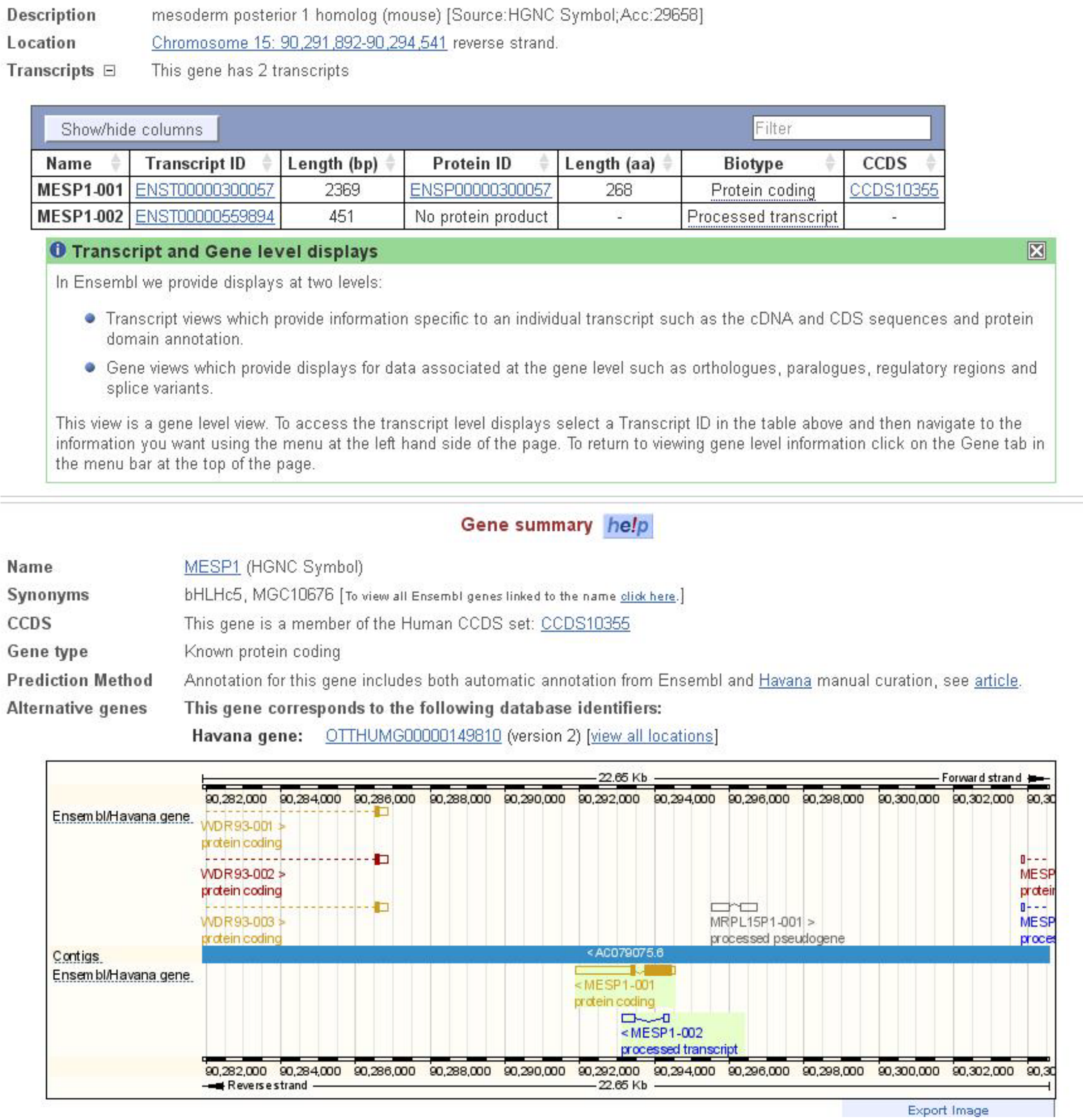
Gene Legend

🟥 protein coding	🟦 processed transcript
🟧 merged Ensembl/Havana	🟪 RNA gene
⬛ pseudogene	

Location: 15:90288632-90483231

Gene:

ENSEMBL—Gene Summary



Description mesoderm posterior 1 homolog (mouse) [Source:HGNC Symbol;Acc:29658]

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.

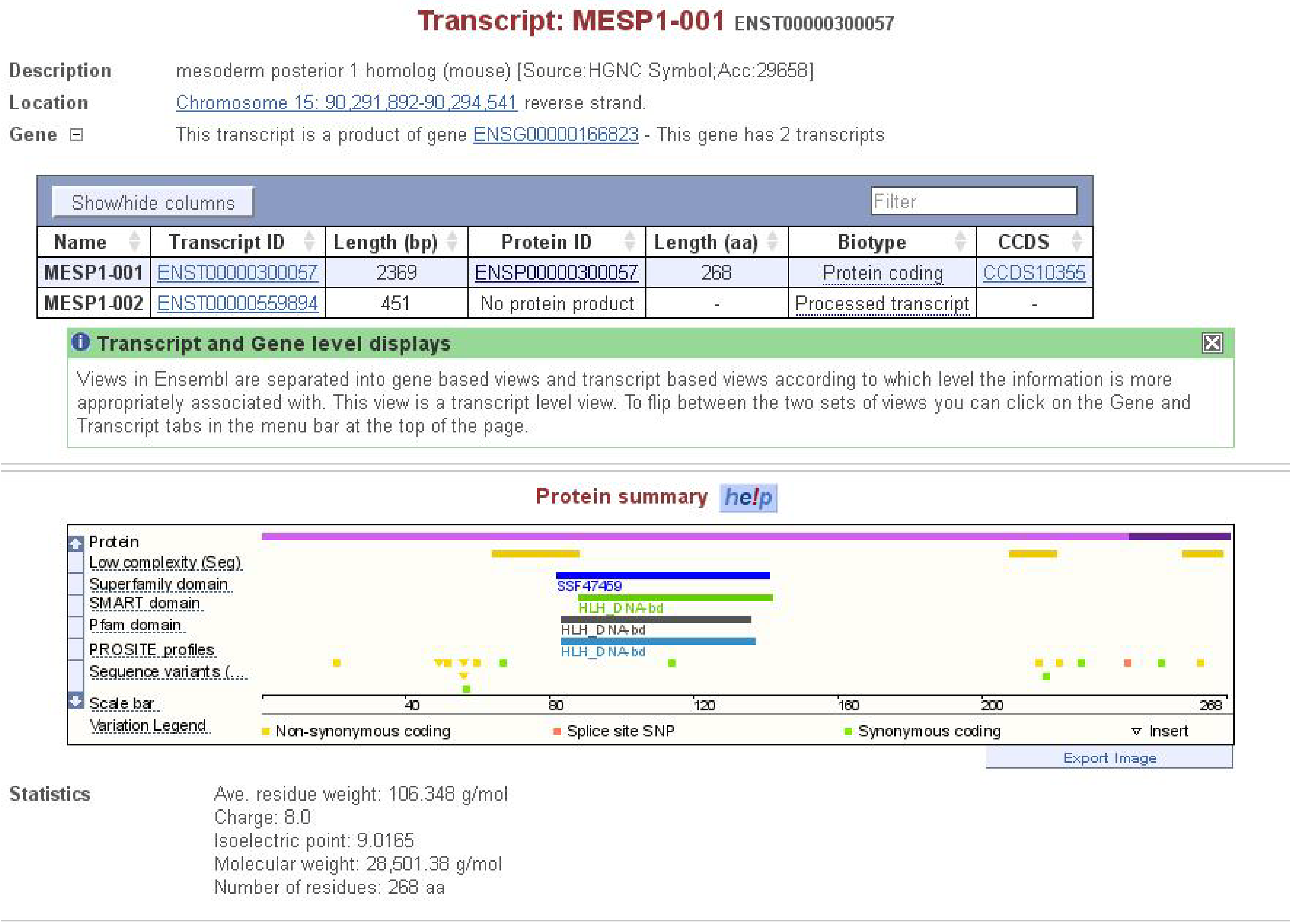
This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want using the menu at the left hand side of the page. To return to viewing gene level information click on the Gene tab in the menu bar at the top of the page.

Gene summary *help*

Name	MESP1 (HGNC Symbol)
Synonyms	bHLHc5, 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: OTTHUMG00000149810 (version 2) [view all locations]

ENSEMBL—Protein



Transcript: MESP1-001 ENST00000300057

Description mesoderm posterior 1 homolog (mouse) [Source:HGNC Symbol;Acc:29658]

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

Show/hide columns   Filter

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: 8.0
Isoelectric point: 9.0165
Molecular weight: 28,501.38 g/mol
Number of residues: 268 aa

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

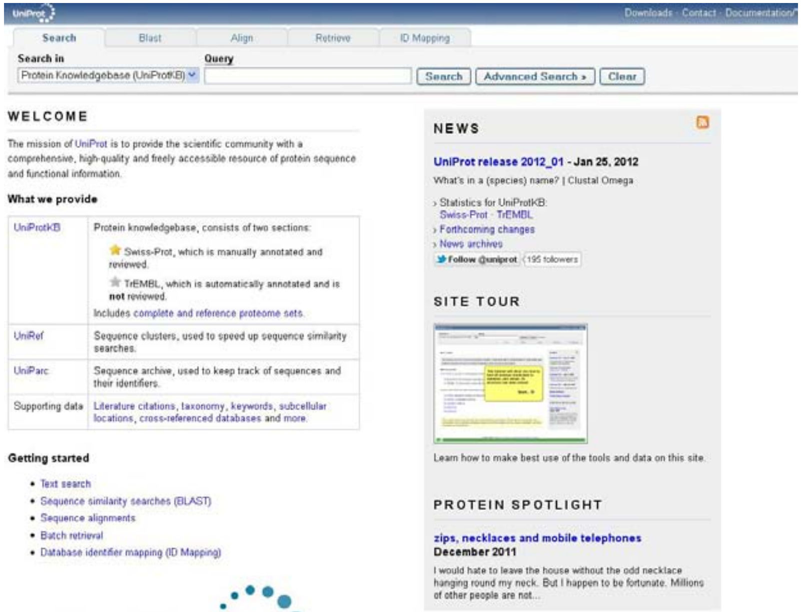
– UniProt combines SwissProt and TrEMBl

“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

Mirro 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



WELCOME

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: Swiss-Prot, which is manually annotated and reviewed. TrEMBL, which is automatically annotated and is **not** reviewed. Includes complete and reference proteome sets.
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations, taxonomy, keywords, subcellular locations, cross-referenced databases and more.

Getting started

- Text search
- Sequence similarity searches (BLAST)
- Sequence alignments
- Batch retrieval
- Database identifier mapping (ID Mapping)

NEWS

UniProt release 2012_01 - Jan 25, 2012

What's in a (species) name? | Clustal Omega

› Statistics for UniProtKB:
Swiss-Prot · TrEMBL
› Forthcoming changes
› News archives

Follow @uniprot 195 followers

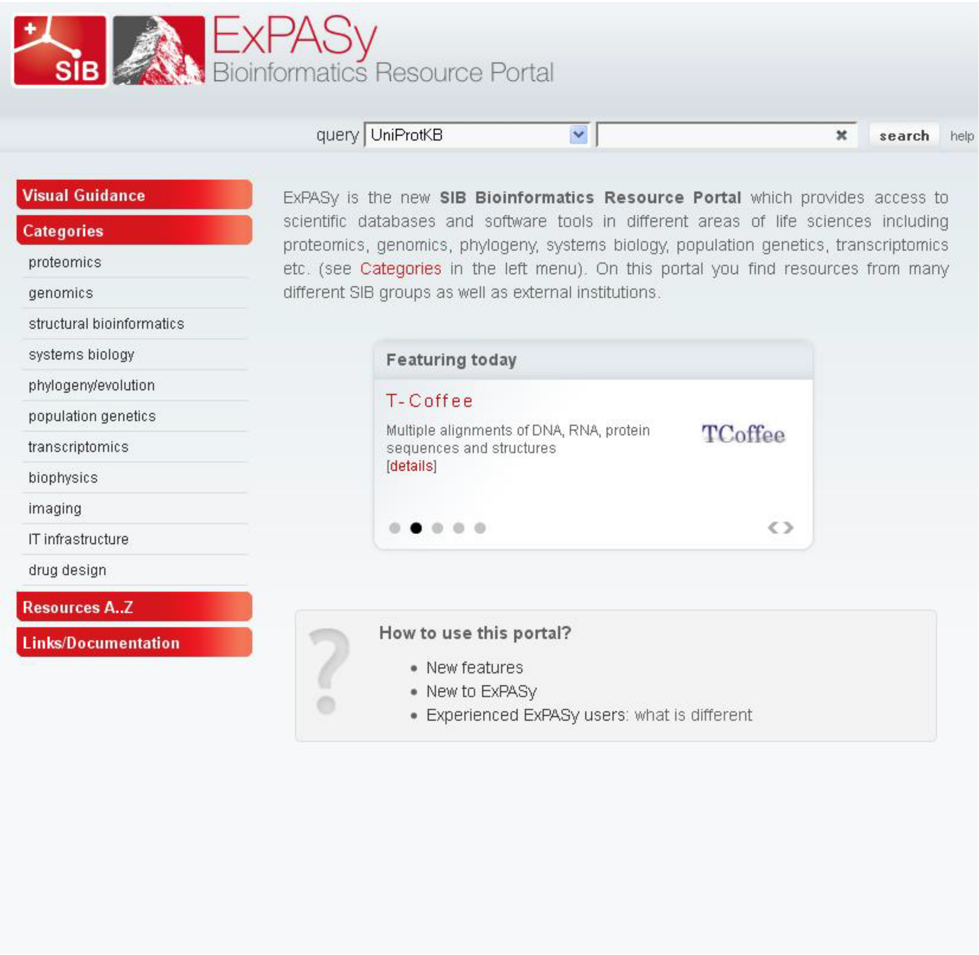
SITE TOUR

Learn how to make best use of the tools and data on this site.

PROTEIN SPOTLIGHT

zips, necklaces and mobile telephones
December 2011

I would hate to leave the house without the odd necklace hanging round my neck. But I happen to be fortunate. Millions of other people are not...



ExPASy is the new **SIB Bioinformatics Resource Portal** which provides access to scientific databases and software tools in different areas of life sciences including proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc. (see Categories in the left menu). On this portal you find resources from many different SIB groups as well as external institutions.

Featuring today

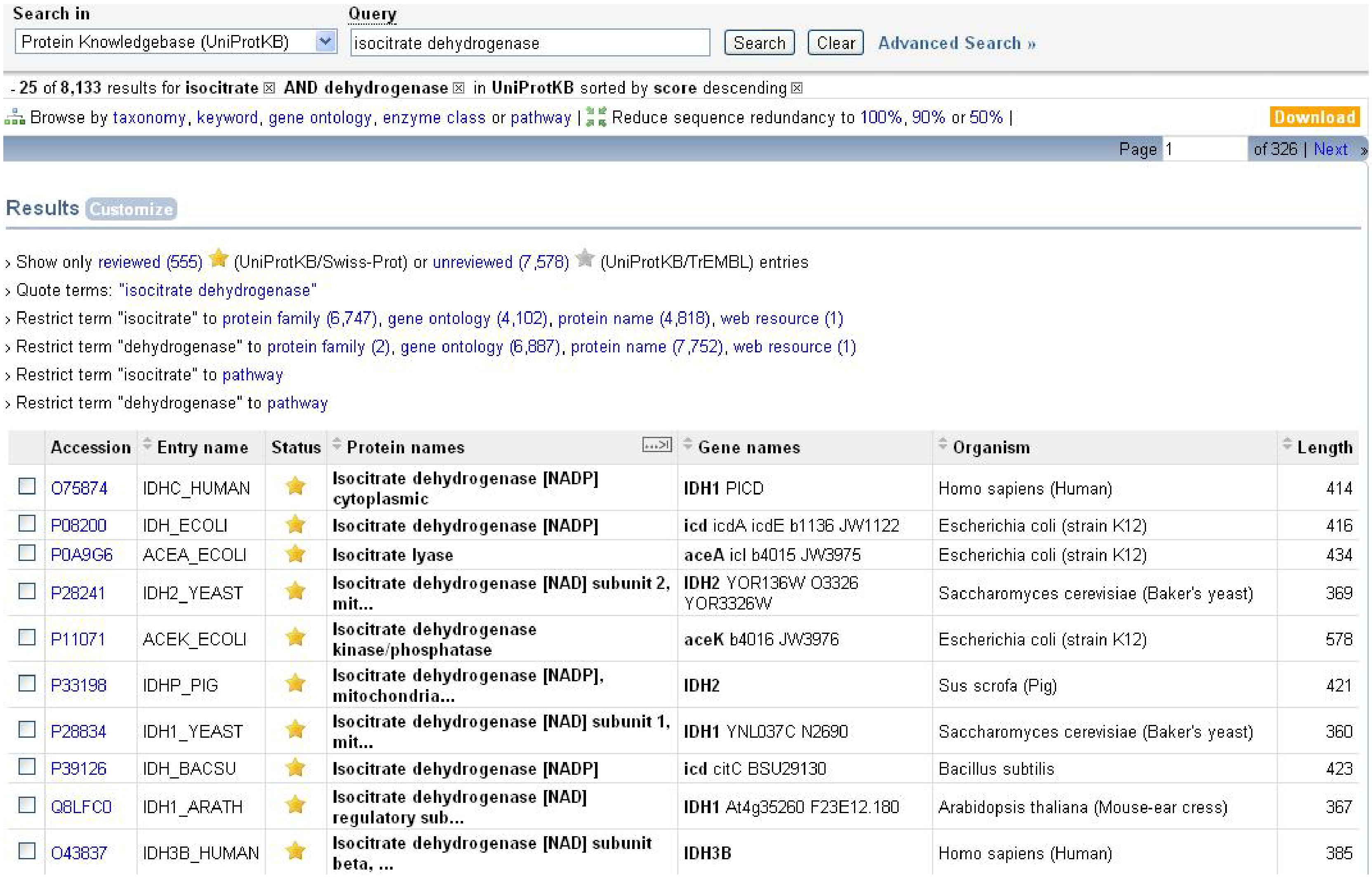
T-Coffee

Multiple alignments of DNA, RNA, protein sequences and structures
[details]

How to use this portal?

- New features
- New to ExPASy
- Experienced ExPASy users: what is different

SwissProt—search for Proteins



Search in: Protein Knowledgebase (UniProtKB)

Query: isocitrate dehydrogenase

- 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% |

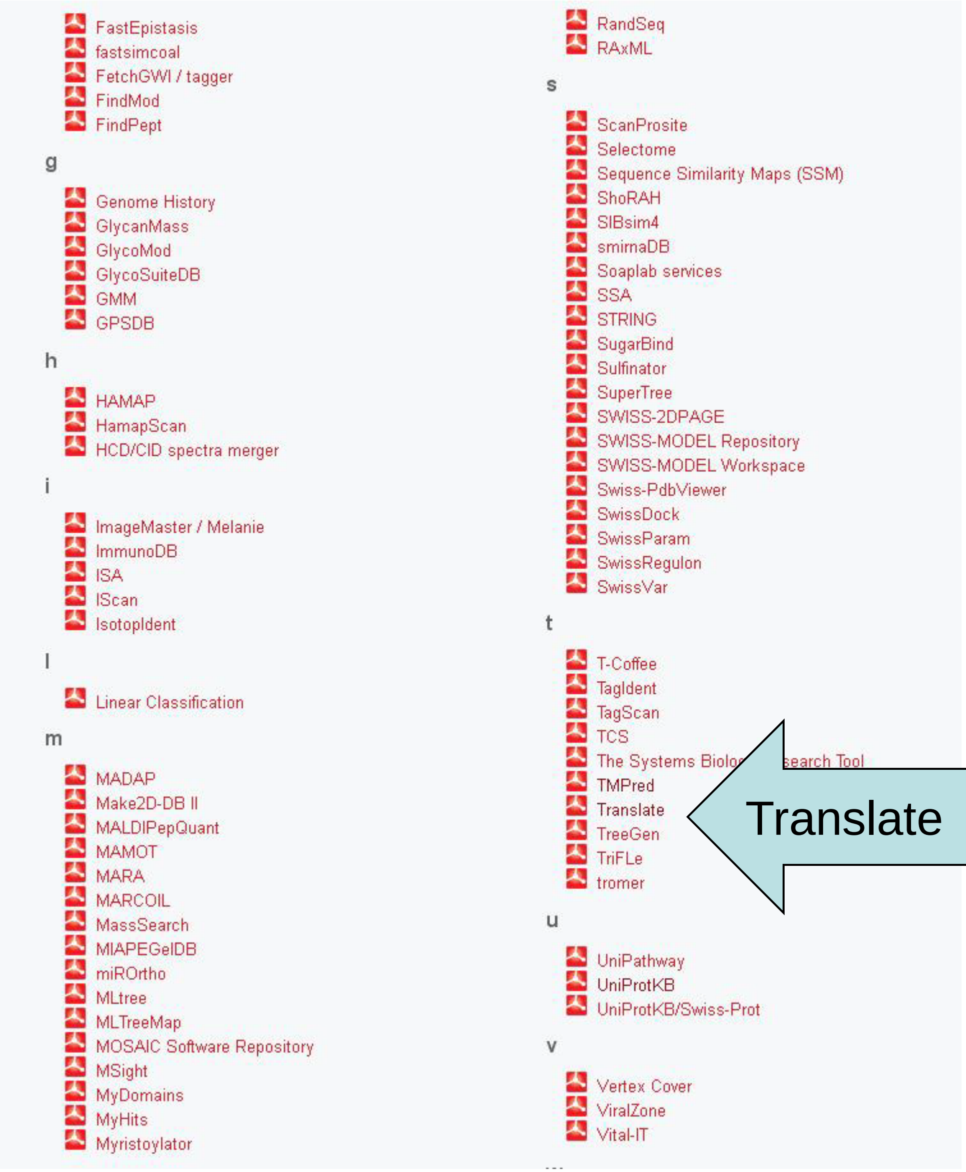
Page 1 of 326 | Next »

Results

› Show only reviewed (555) ★ (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
☐	O75874	IDHC_HUMAN	★	**Isocitrate dehydrogenase [NADP] cytoplasmic**	**IDH1** PICD	Homo sapiens (Human)	414
☐	P08200	IDH_ECOLI	★	**Isocitrate dehydrogenase [NADP]**	**icd** icdA icdE b1136 JW1122	Escherichia coli (strain K12)	416
☐	P0A9G6	ACEA_ECOLI	★	**Isocitrate lyase**	**aceA** icl b4015 JW3975	Escherichia coli (strain K12)	434
☐	P28241	IDH2_YEAST	★	**Isocitrate dehydrogenase [NAD] subunit 2, mit...**	**IDH2** YOR136W O3326 YOR3326W	Saccharomyces cerevisiae (Baker's yeast)	369
☐	P11071	ACEK_ECOLI	★	**Isocitrate dehydrogenase kinase/phosphatase**	**aceK** b4016 JW3976	Escherichia coli (strain K12)	578
☐	P33198	IDHP_PIG	★	**Isocitrate dehydrogenase [NADP], mitochondria...**	**IDH2**	Sus scrofa (Pig)	421
☐	P28834	IDH1_YEAST	★	**Isocitrate dehydrogenase [NAD] subunit 1, mit...**	**IDH1** YNL037C N2690	Saccharomyces cerevisiae (Baker's yeast)	360
☐	P39126	IDH_BACSU	★	**Isocitrate dehydrogenase [NADP]**	**icd** citC BSU29130	Bacillus subtilis	423
☐	Q8LFC0	IDH1_ARATH	★	**Isocitrate dehydrogenase [NAD] regulatory sub...**	**IDH1** At4g35260 F23E12.180	Arabidopsis thaliana (Mouse-ear cress)	367
☐	O43837	IDH3B_HUMAN	★	**Isocitrate dehydrogenase [NAD] subunit beta, ...**	**IDH3B**	Homo sapiens (Human)	385

EXPASY-Databases and Features



FastEpistasis
fastsimcoal
FetchGWI / tagger
FindMod
FindPept

g

Genome History
GlycanMass
GlycoMod
GlycoSuiteDB
GMM
GPSDB

h

HAMAP
HamapScan
HCD/CID spectra merger

i

ImageMaster / Melanie
ImmunoDB
ISA
IScan
IsotopIdent

l

Linear Classification

m

MADAP
Make2D-DB II
MALDIPepQuant
MAMOT
MARA
MARCOIL
MassSearch
MIAPEGelDB
miROrtho
MLtree
MLTreeMap
MOSAIC Software Repository
MSight
MyDomains
MyHits
Myristoylator

RandSeq
RAxML

s

ScanProsite
Selectome
Sequence Similarity Maps (SSM)
ShoRAH
SIBsim4
smirnaDB
Soaplab services
SSA
STRING
SugarBind
Sulfinator
SuperTree
SWISS-2DPAGE
SWISS-MODEL Repository
SWISS-MODEL Workspace
Swiss-PdbViewer
SwissDock
SwissParam
SwissRegulon
SwissVar

t

T-Coffee
TagIdent
TagScan
TCS
The Systems Biolo search Tool
TMPred
Translate
TreeGen
TriFLe
tromer

u

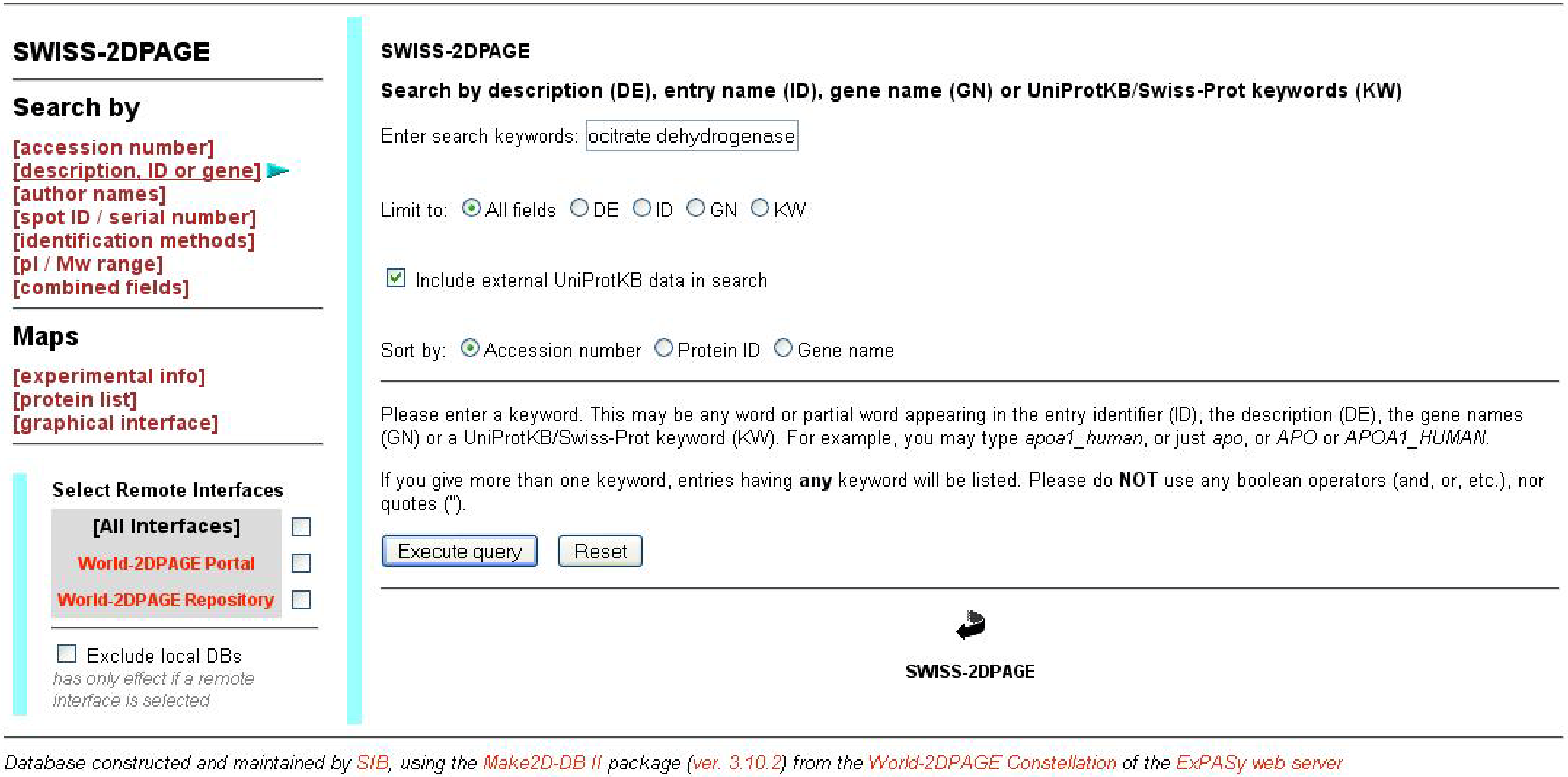
UniPathway
UniProtKB
UniProtKB/Swiss-Prot

v

Vertex Cover
ViralZone
Vital-IT

Translate

Swiss 2D-PAGE



SWISS-2DPAGE

Search by

[accession number]
[description, ID or gene]
[author names]
[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: ocitrate dehydrogenase

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 *APO* or *APOA1_HUMAN*.

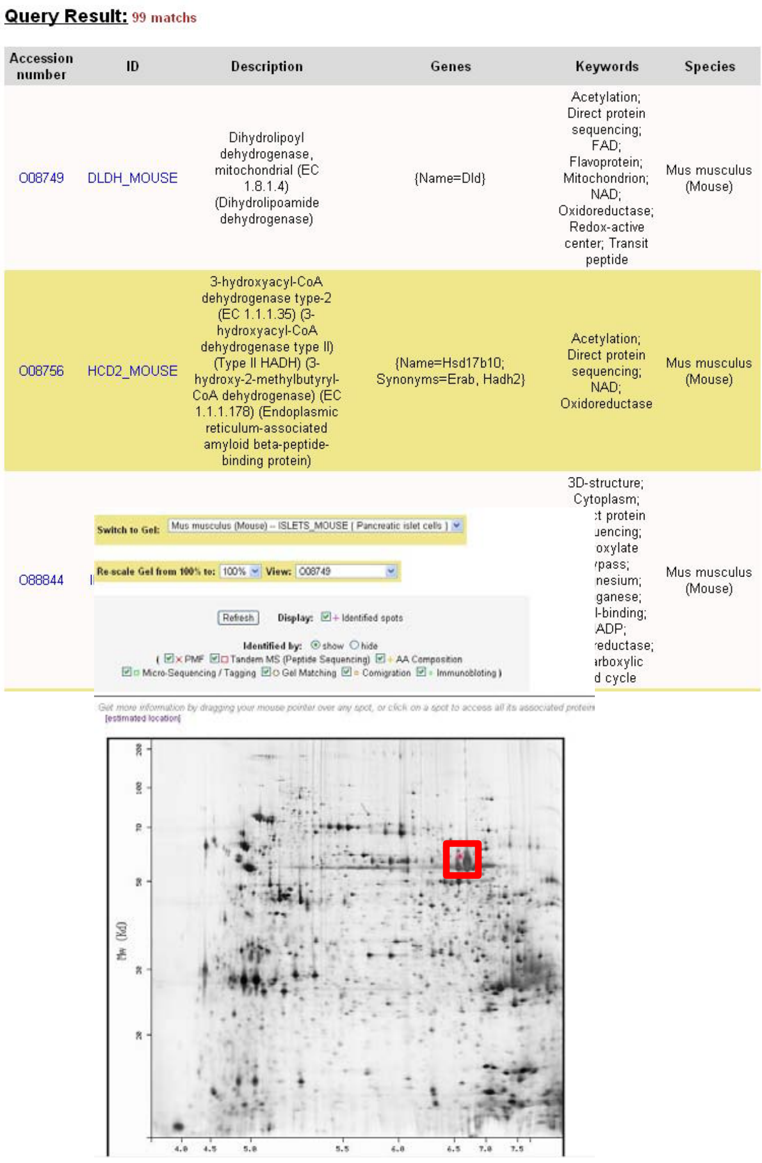
If you give more than one keyword, entries having **any** keyword will be listed. Please do **NOT** use any boolean operators (and, or, etc.), nor quotes (").

SWISS-2DPAGE

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 matching any of the keywords:
'isocitrate dehydrogenase'
in their description (DE), entry name (ID), gene names (GN) or UniProtKB/Swiss-Prot keywords (KW).



Query Result: 99 matchs

Accession number	ID	Description	Genes	Keywords	Species
O08749	DLDH_MOUSE	Dihydrolipoyl dehydrogenase, mitochondrial (EC 1.8.1.4) (Dihydrolipoamide dehydrogenase)	{Name=Dld}	Acetylation; Direct protein sequencing; FAD; Flavoprotein; Mitochondrion; NAD; Oxidoreductase; Redox-active center; Transit peptide	Mus musculus (Mouse)
O08756	HCD2_MOUSE	3-hydroxyacyl-CoA dehydrogenase type-2 (EC 1.1.1.35) (3-hydroxyacyl-CoA dehydrogenase type II) (Type II HADH) (3-hydroxy-2-methylbutyryl-CoA dehydrogenase) (EC 1.1.1.178) (Endoplasmic reticulum-associated amyloid beta-peptide-binding protein)	{Name=Hsd17b10; Synonyms=Erab, Hadh2}	Acetylation; Direct protein sequencing; NAD; Oxidoreductase	Mus musculus (Mouse)
O88844	I			3D-structure; Cytoplasm; t protein uencing; oxylate ypass; nesium; ganese; l-binding; ADP; reductase; arboxylic d cycle	Mus musculus (Mouse)

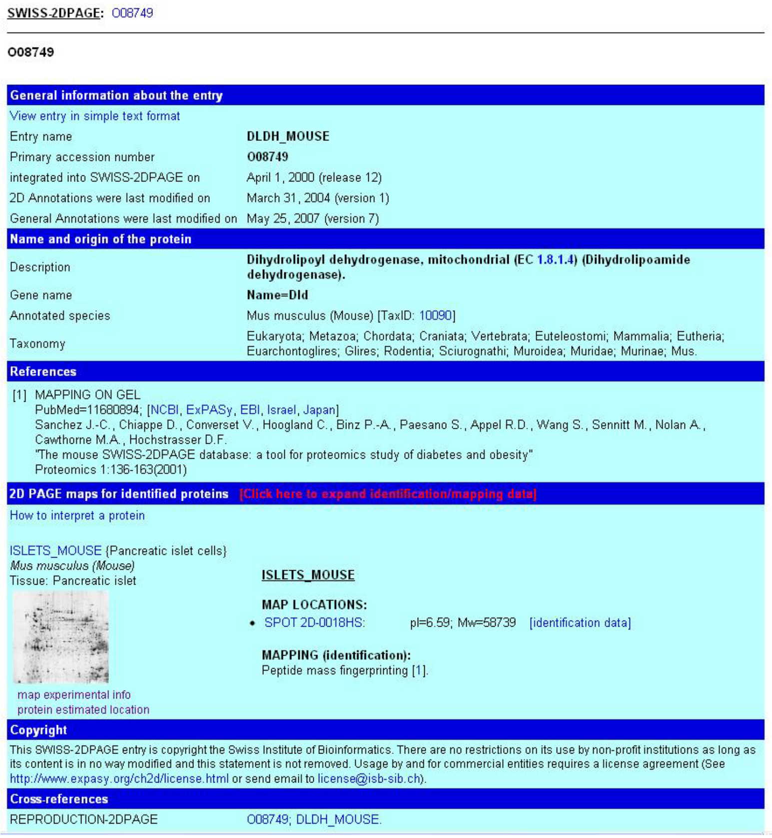
Switch to Gel: Mus musculus (Mouse) -- ISLETS_MOUSE (Pancreatic islet cells)

Re-scale Gel from 100% to: 100% **View:** O08749

Display: ☑ + Identified spots

Identified by: ● show ○ hide
(☑ × PMF ☑ □ Tandem MS (Peptide Sequencing) ☑ + AA Composition
☑ □ Micro-Sequencing / Tagging ☑ ○ Gel Matching ☑ = Comigration ☑ + Immunobloting)

Get more information by dragging your mouse pointer over any spot, or click on a spot to access all its associated protein
[estimated location]



SWISS-2DPAGE: O08749

O08749

General information about the entry

View entry in simple text format

Entry name	**DLDH_MOUSE**
Primary accession number	**O08749**
integrated into SWISS-2DPAGE on	April 1, 2000 (release 12)
2D Annotations were last modified on	March 31, 2004 (version 1)
General Annotations were last modified on	May 25, 2007 (version 7)

Name and origin of the protein

Description	**Dihydrolipoyl dehydrogenase, mitochondrial (EC 1.8.1.4) (Dihydrolipoamide dehydrogenase).**
Gene name	**Name=Dld**
Annotated species	Mus musculus (Mouse) [TaxID: 10090]
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

References

[1] MAPPING ON GEL
PubMed=11680894; [NCBI, ExPASy, EBI, Israel, Japan]
Sanchez J.-C., Chiappe D., Converset V., Hoogland C., Binz P.-A., Paesano S., Appel R.D., Wang S., Sennitt M., Nolan A., Cawthorne M.A., Hochstrasser D.F.
"The mouse SWISS-2DPAGE database: a tool for proteomics study of diabetes and obesity"
Proteomics 1:136-163(2001)

2D PAGE maps for identified proteins [Click here to expand identification/mapping data]

How to interpret a protein

ISLETS_MOUSE {Pancreatic islet cells}
Mus musculus (Mouse)
Tissue: Pancreatic islet

map experimental info
protein estimated location

ISLETS_MOUSE

MAP LOCATIONS:

- SPOT 2D-0018HS: pI=6.59; Mw=58739 [identification data]

MAPPING (identification):
Peptide mass fingerprinting [1].

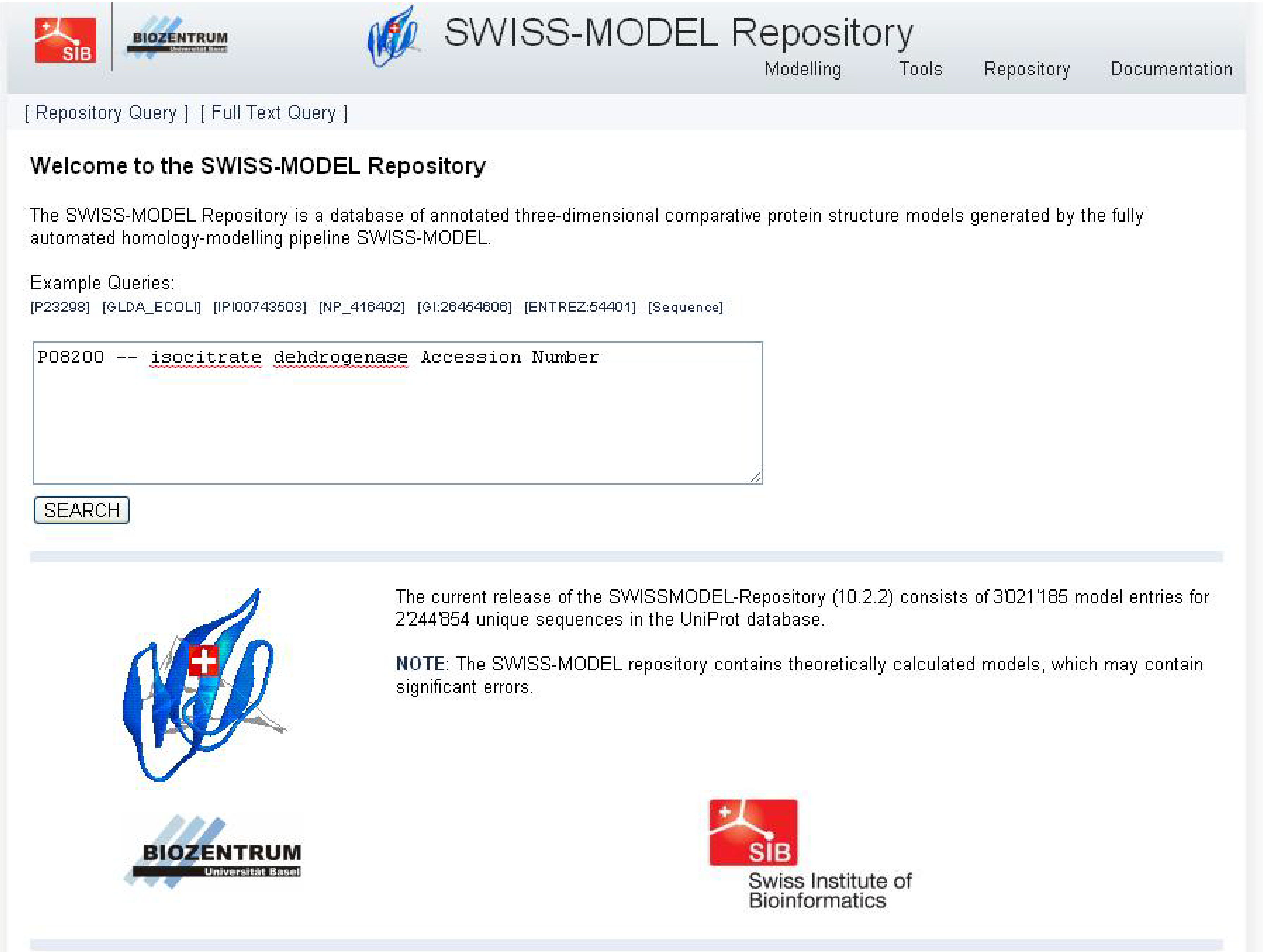
Copyright

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 in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.expasy.org/ch2d/license.html or send email to license@isb-sib.ch).

Cross-references

REPRODUCTION-2DPAGE	O08749; DLDH_MOUSE.

Swiss PDB Model Respository



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.

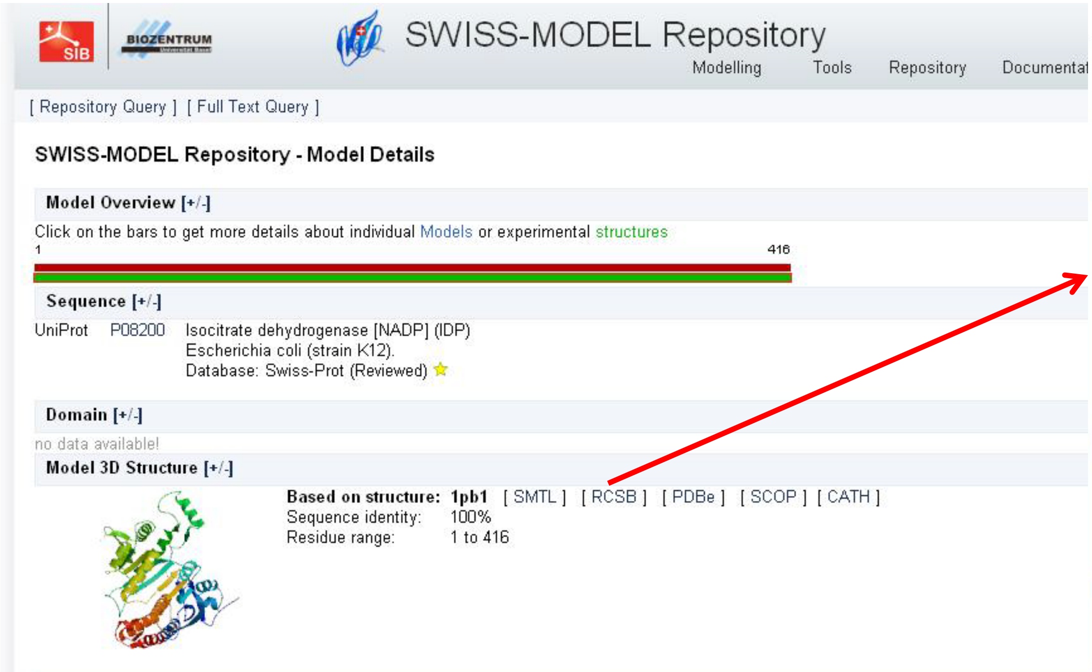
Example Queries:
[P23298] [GLDA_ECOLI] [IPI00743503] [NP_416402] [GI:26454606] [ENTREZ:54401] [Sequence]

```
P08200 -- isocitrate dehdrogenase Accession Number
```

The current release of the SWISSMODEL-Repository (10.2.2) consists of 3'021'185 model entries for 2'244'854 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 - Model Details

Model Overview [+/-]

Click on the bars to get more details about individual Models or experimental structures

1 416

Sequence [+/-]

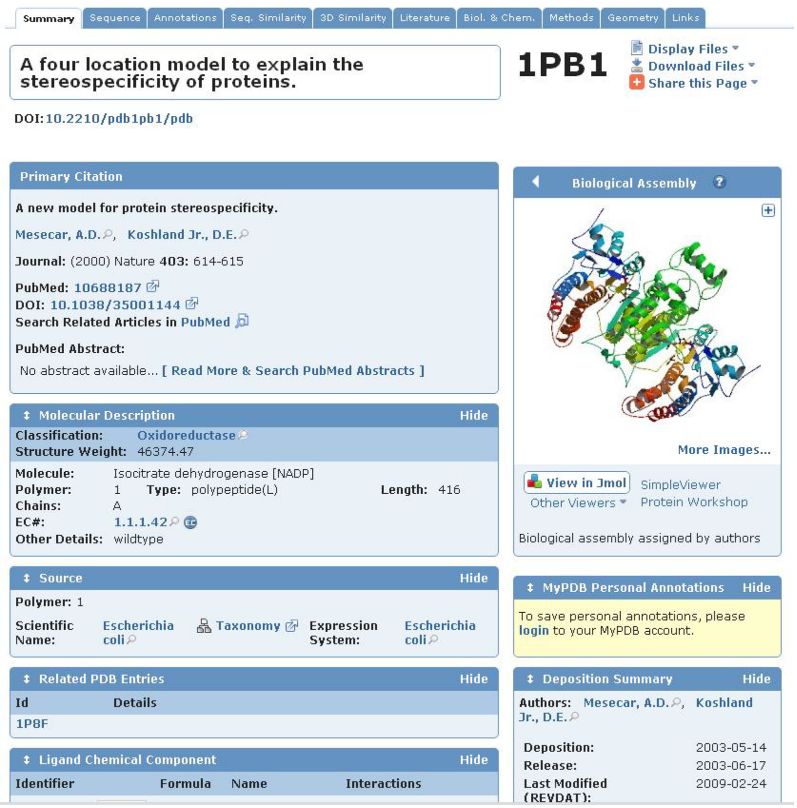
UniProt P08200 Isocitrate dehydrogenase [NADP] (IDP)
Escherichia coli (strain K12).
Database: Swiss-Prot (Reviewed) ★

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.

1PB1

DOI: 10.2210/pdb1pb1/pdb

Primary Citation

A new model for protein stereospecificity.

Mesecar, A.D., Koshland Jr., D.E.

Journal: (2000) Nature **403:** 614-615

PubMed: 10688187
DOI: 10.1038/35001144
Search Related Articles in PubMed

PubMed Abstract:

No abstract available... [Read More & Search PubMed Abstracts]

Biological Assembly

More Images...

Biological assembly assigned by authors

Molecular Description Hide

Classification: Oxidoreductase
Structure Weight: 46374.47

Molecule: Isocitrate dehydrogenase [NADP]
Polymer: 1 **Type:** polypeptide(L) **Length:** 416
Chains: A
EC#: 1.1.1.42
Other Details: wildtype

Source Hide

Polymer: 1

Scientific Name: Escherichia coli Taxonomy **Expression System:** Escherichia coli

MyPDB Personal Annotations Hide

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

Related PDB Entries Hide

Id	Details
1P8F	

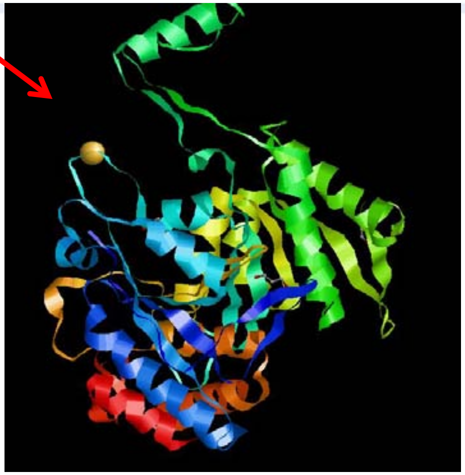
Deposition Summary Hide

Authors: Mesecar, A.D., Koshland Jr., D.E.

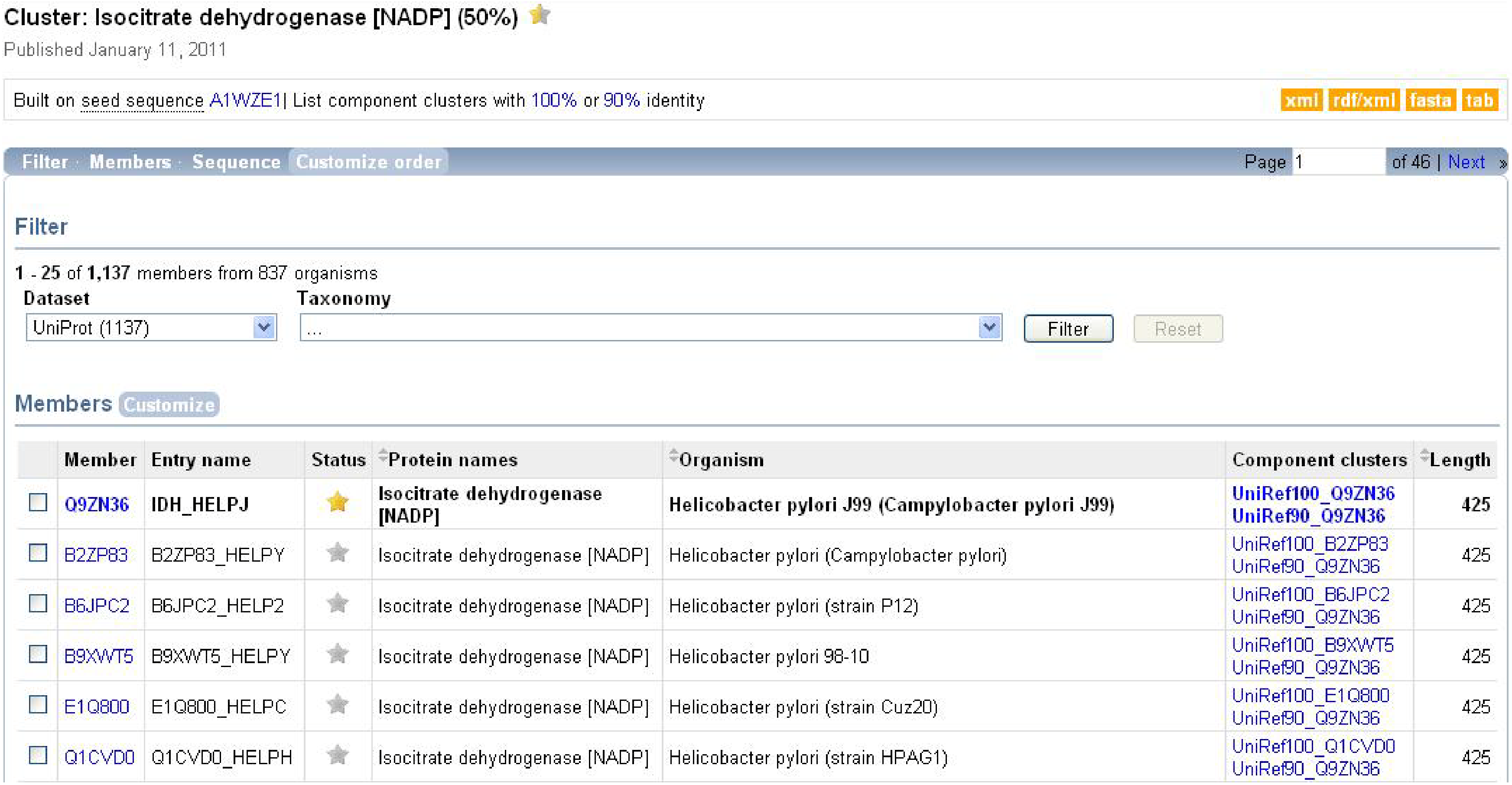
Deposition: 2003-05-14
Release: 2003-06-17
Last Modified (REVDAT): 2009-02-24

Ligand Chemical Component Hide

Identifier	Formula	Name	Interactions



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

Filter

1 - **25** of **1,137** members from 837 organisms

Dataset: UniProt (1137)

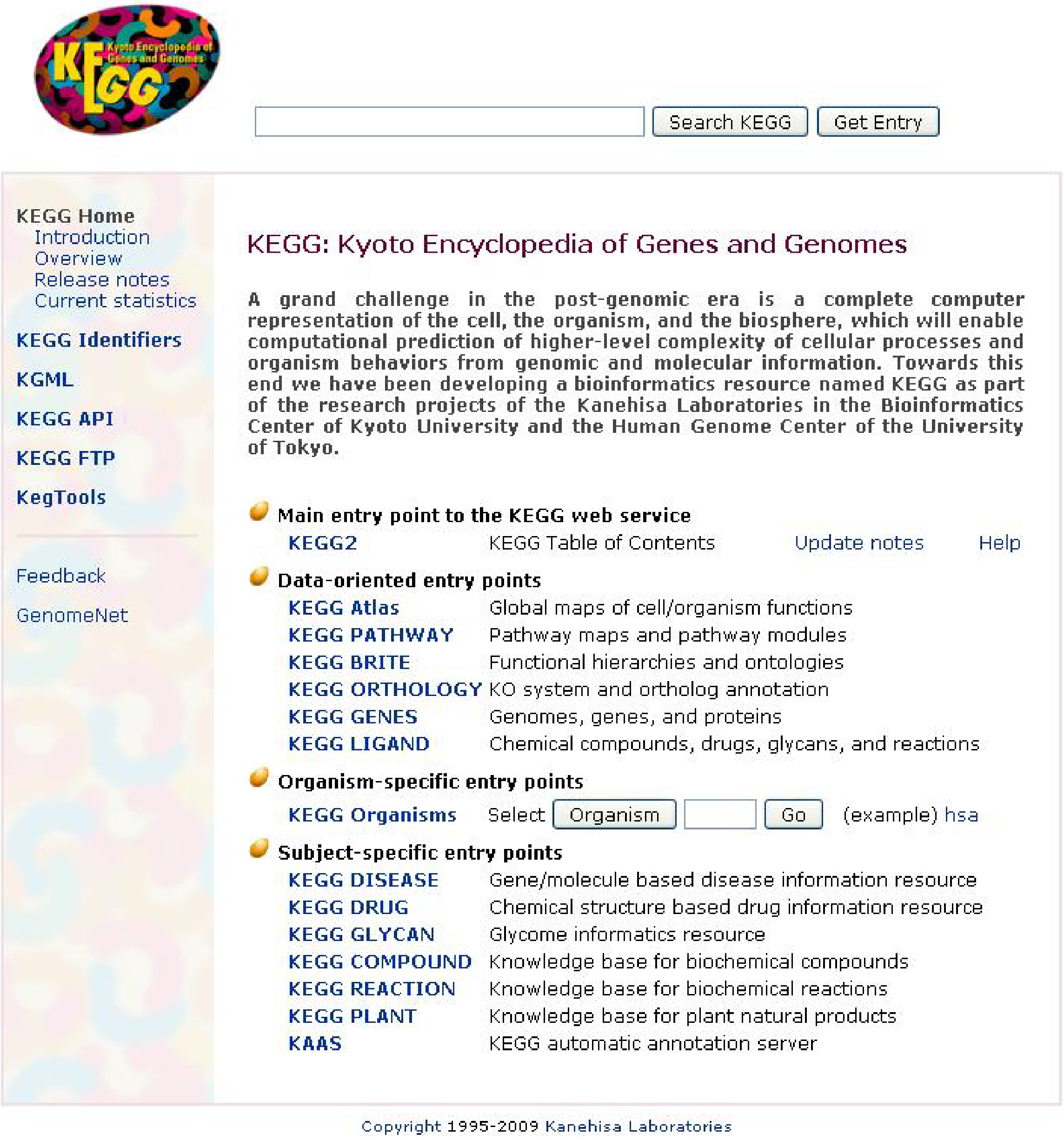
Taxonomy: ...

Members

	Member	Entry name	Status	Protein names	Organism	Component clusters	Length
☐	**Q9ZN36**	**IDH_HELPJ**	★	**Isocitrate dehydrogenase [NADP]**	**Helicobacter pylori J99 (Campylobacter pylori J99)**	**UniRef100_Q9ZN36** **UniRef90_Q9ZN36**	**425**
☐	B2ZP83	B2ZP83_HELPY	☆	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (Campylobacter pylori)	UniRef100_B2ZP83 UniRef90_Q9ZN36	425
☐	B6JPC2	B6JPC2_HELP2	☆	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (strain P12)	UniRef100_B6JPC2 UniRef90_Q9ZN36	425
☐	B9XWT5	B9XWT5_HELPY	☆	Isocitrate dehydrogenase [NADP]	Helicobacter pylori 98-10	UniRef100_B9XWT5 UniRef90_Q9ZN36	425
☐	E1Q800	E1Q800_HELPC	☆	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (strain Cuz20)	UniRef100_E1Q800 UniRef90_Q9ZN36	425
☐	Q1CVD0	Q1CVD0_HELPH	☆	Isocitrate dehydrogenase [NADP]	Helicobacter pylori (strain HPAG1)	UniRef100_Q1CVD0 UniRef90_Q9ZN36	425

KEGG (Kyoto Encyclopedia of Genes and Genomes)

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



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

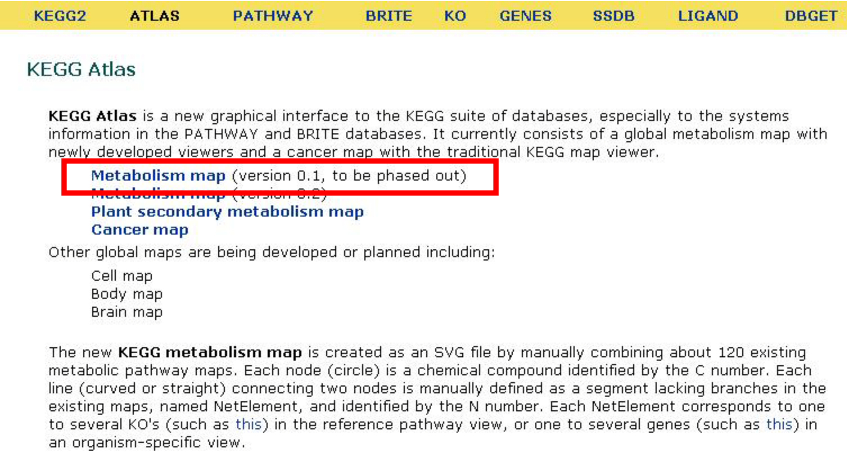
Organism-specific entry points

KEGG Organisms Select [Organism] [ ] [Go] (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

Kegg Atlas



KEGG Atlas

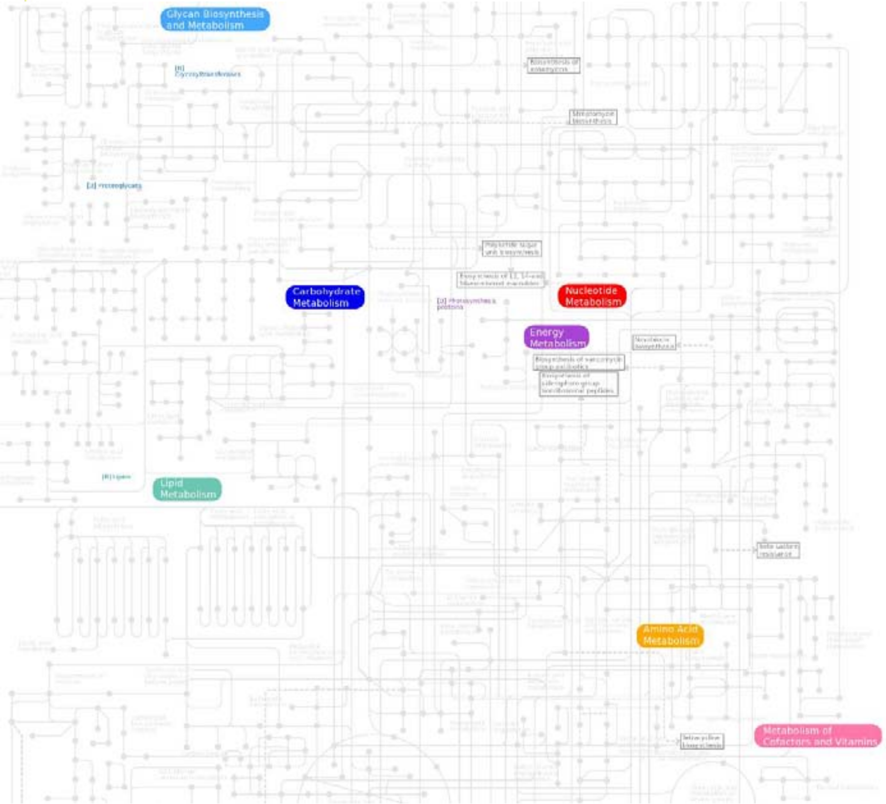
KEGG Atlas is a new graphical interface to the KEGG suite of databases, especially to the systems information in the PATHWAY and BRITE databases. It currently consists of a global metabolism map with newly developed viewers and a cancer map with the traditional KEGG map viewer.

- **Metabolism map** (version 0.1, to be phased out)
- **Metabolism map** (version 0.2)
- **Plant secondary metabolism map**
- **Cancer map**

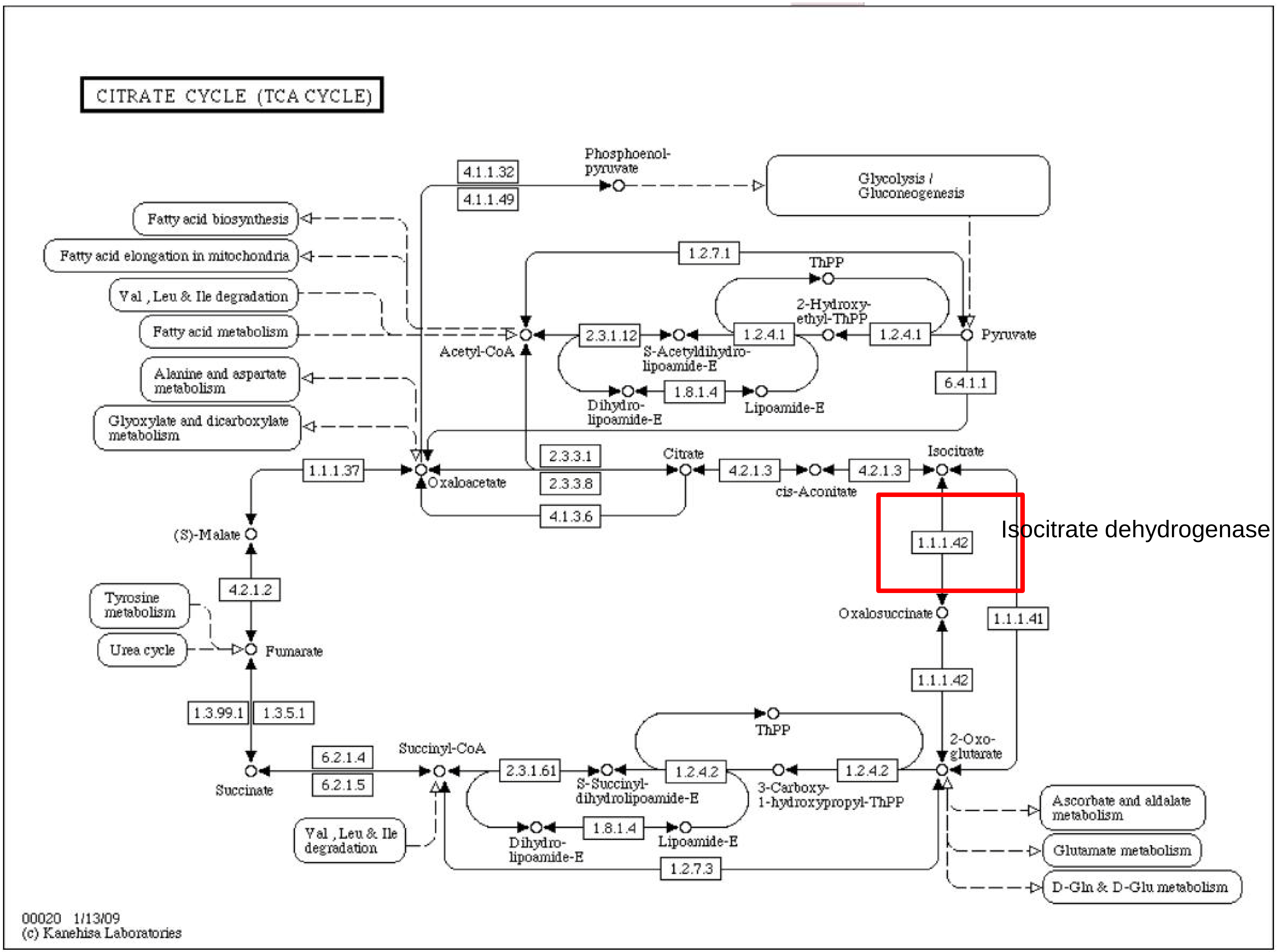
Other global maps are being developed or planned including:

- 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 NetElement, and identified by the N number. Each NetElement 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.

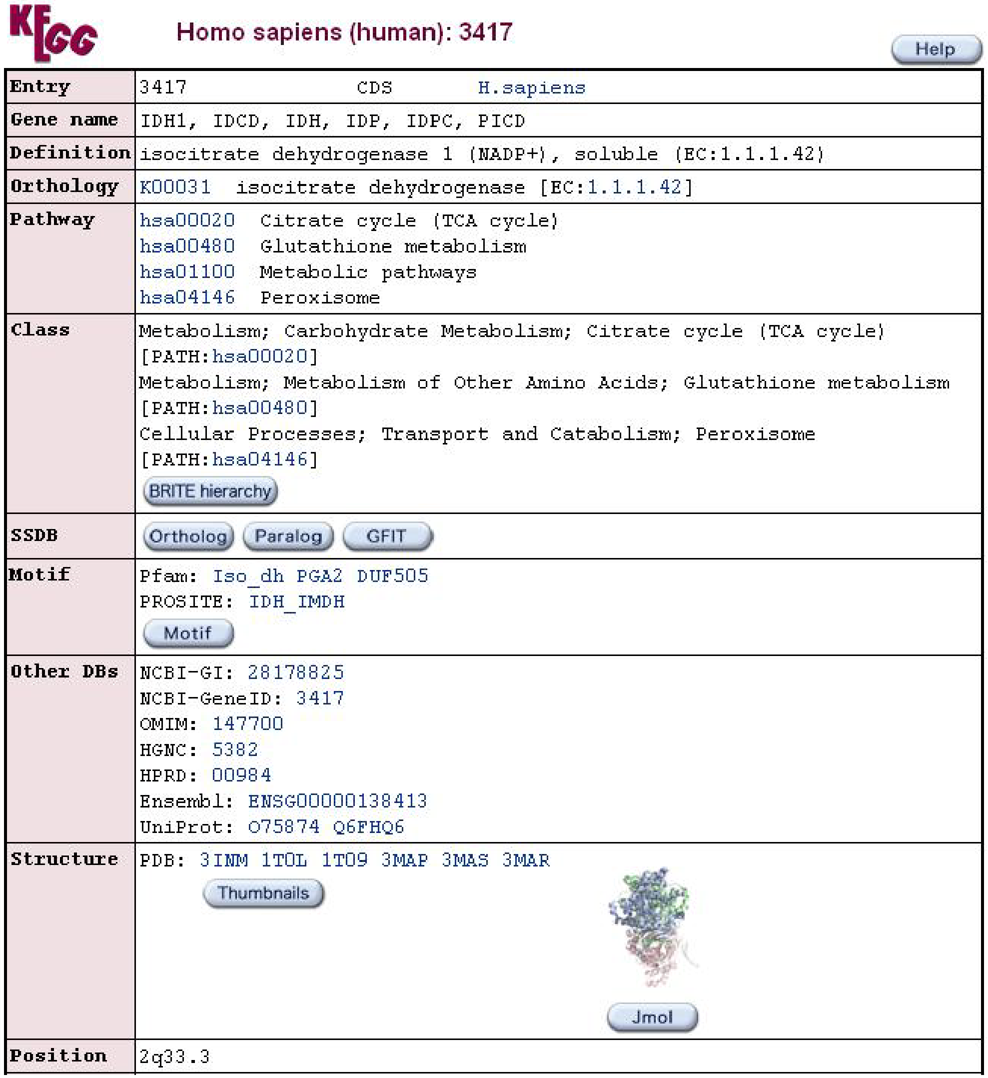


KEGG Pathway



Isocitrate dehydrogenase

Isocitrate Dehydrogenase in KEGG



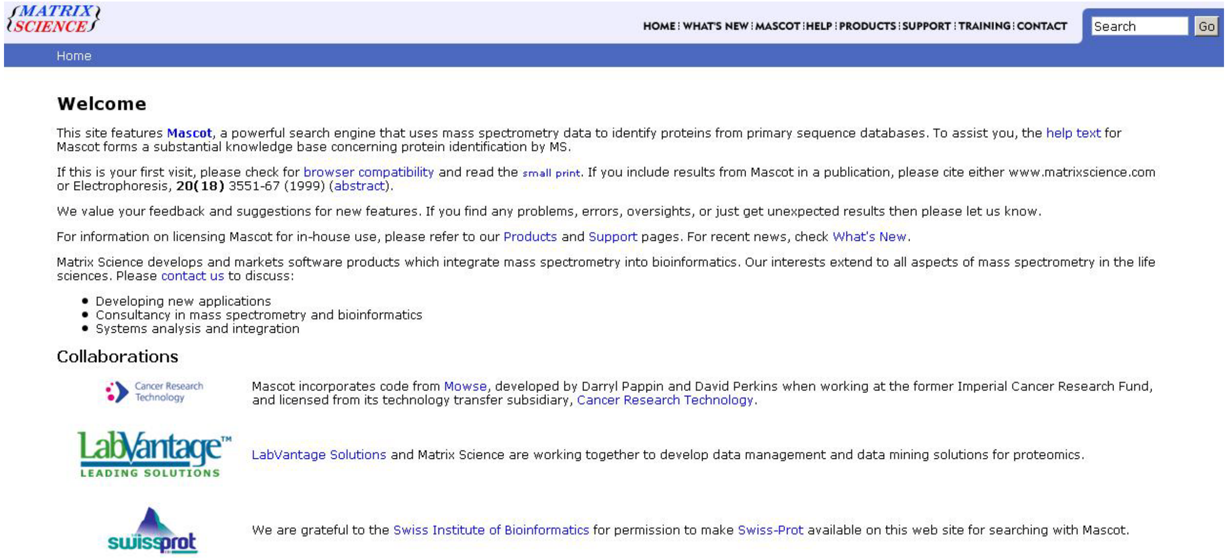
Homo sapiens (human): 3417

Entry	3417 CDS H.sapiens
Gene name	IDH1, IDCD, IDH, IDP, IDPC, PICD
Definition	isocitrate dehydrogenase 1 (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]
SSDB	
Motif	Pfam: Iso_dh PGA2 DUF505 PROSITE: IDH_IMDH
Other DBs	NCBI-GI: 28178825 NCBI-GeneID: 3417 OMIM: 147700 HGNC: 5382 HPRD: 00984 Ensembl: ENSG00000138413 UniProt: O75874 Q6FHQ6
Structure	PDB: 3INM 1T0L 1T09 3MAP 3MAS 3MAR
Position	2q33.3

…

MASCOT—Protein Identification from Mass Spectroscopy Data

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



Welcome

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

If this is your first visit, please check for browser compatibility and read the 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

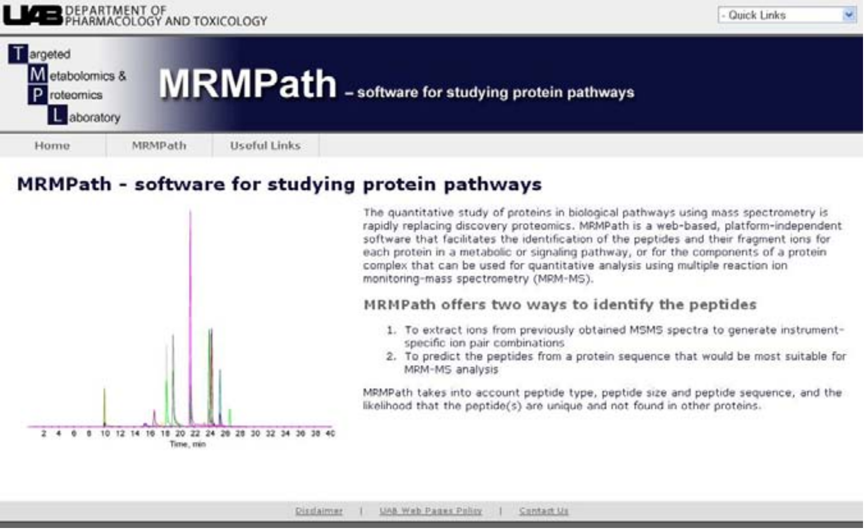
Collaborations

Mascot incorporates code from Mowse, developed by Darryl 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



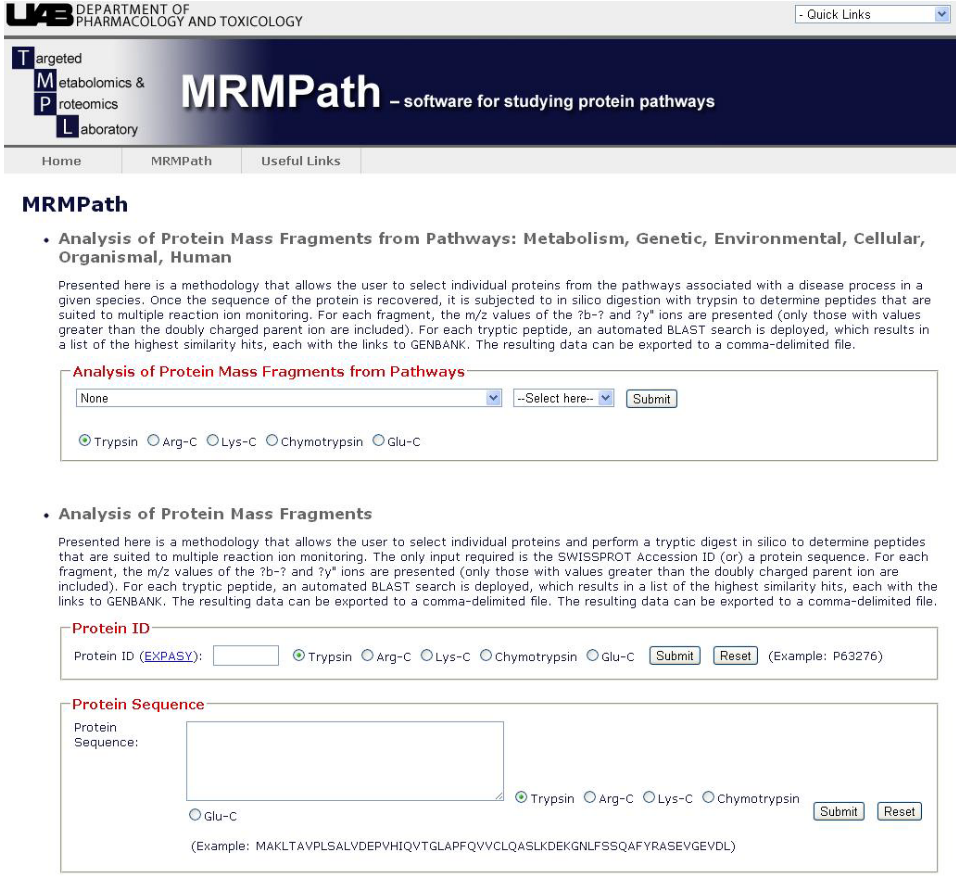
MRMPath - software for studying protein pathways

The quantitative study of proteins in biological pathways using mass spectrometry is rapidly replacing discovery proteomics. 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 MSMS 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

- **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 ion 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 links to GENBANK. The resulting data can be exported to a comma-delimited file.

Analysis of Protein Mass Fragments from Pathways

● 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 SWISSPROT Accession ID (or) 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 links to GENBANK. The resulting data can be exported to a comma-delimited file. The resulting data can be exported to a comma-delimited file.

Protein ID

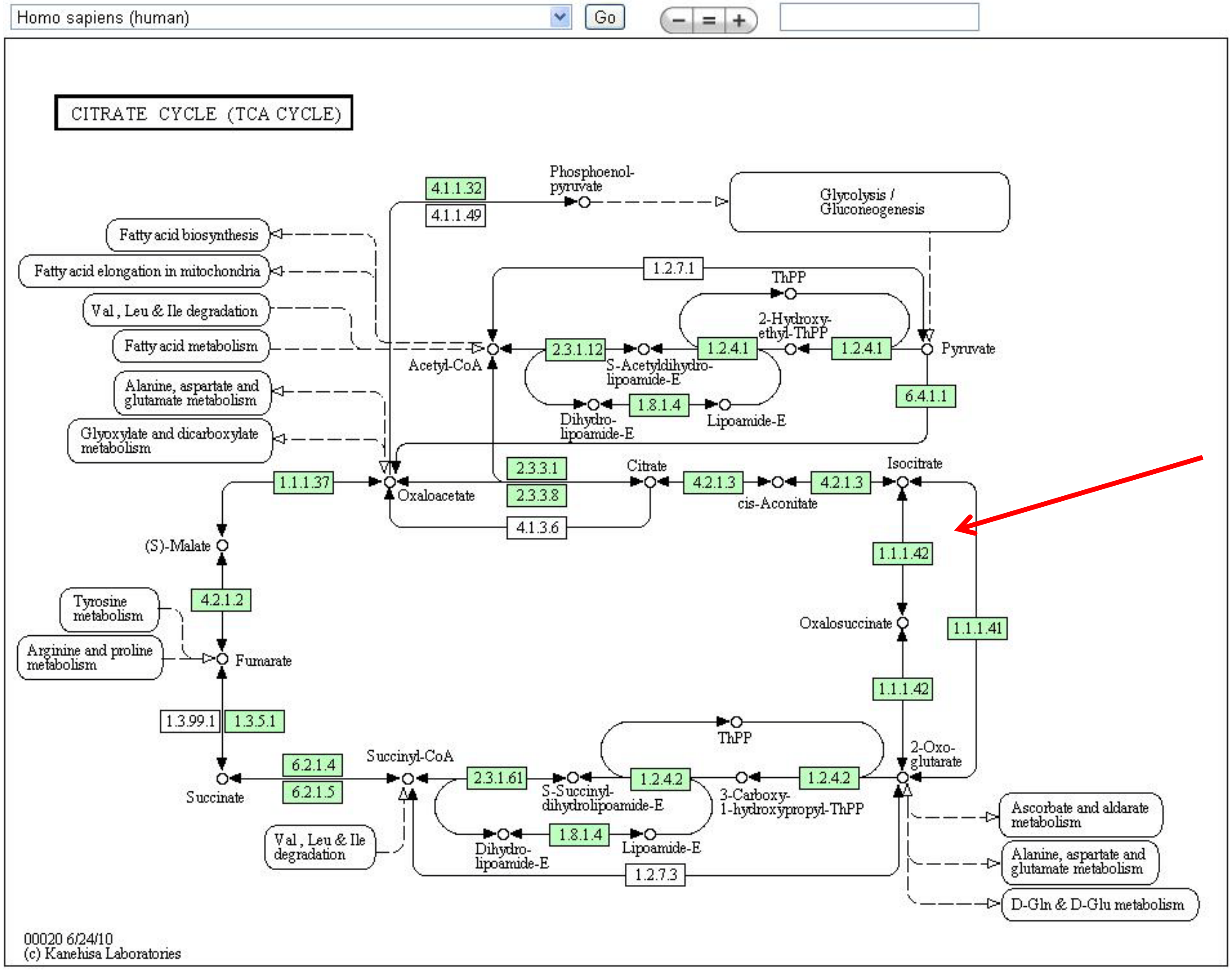
Protein ID (EXPASY): ● Trypsin ○ Arg-C ○ Lys-C ○ Chymotrypsin ○ Glu-C (Example: P63276)

Protein Sequence

Protein Sequence: ● Trypsin ○ Arg-C ○ Lys-C ○ Chymotrypsin ○ Glu-C

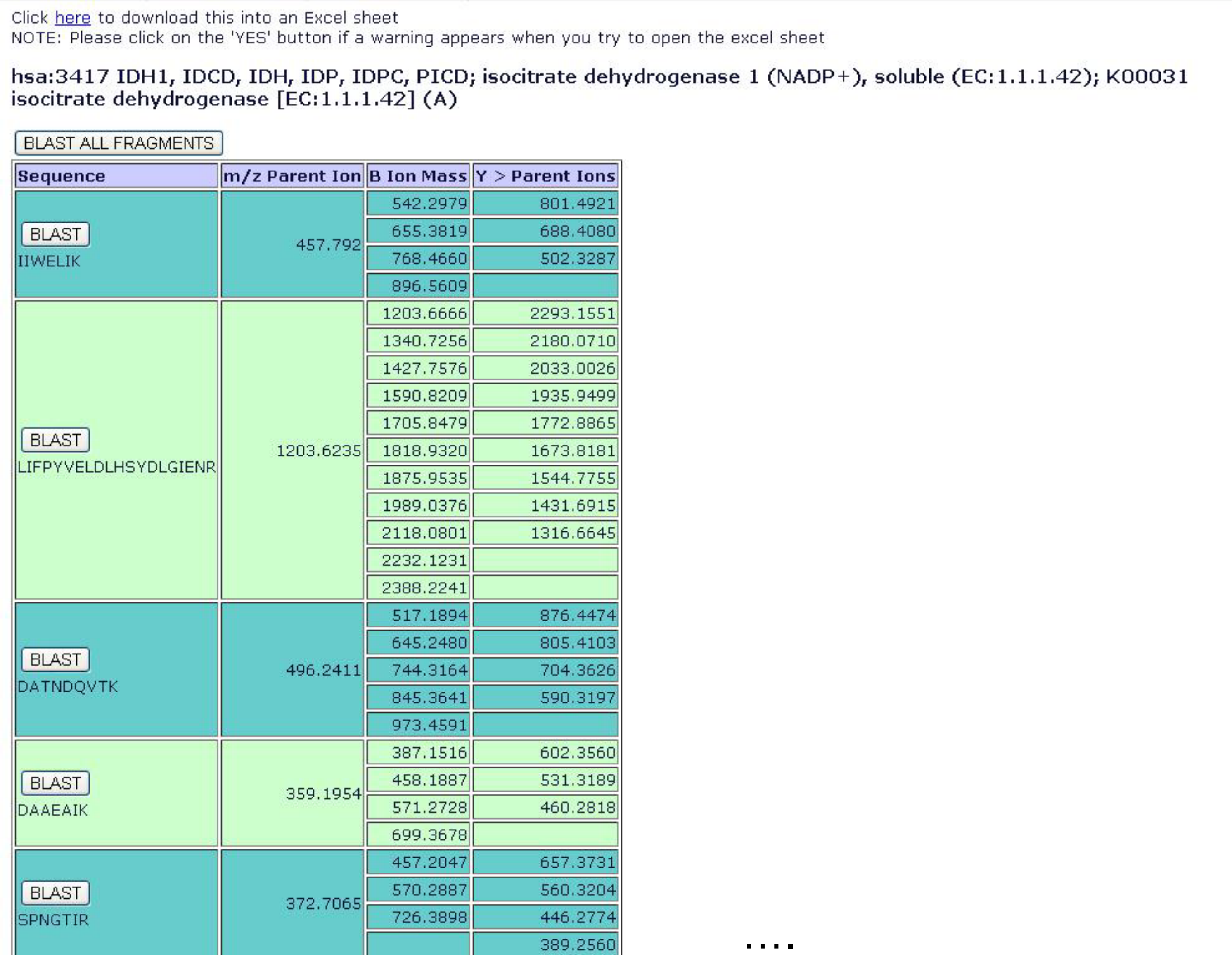
(Example: MAKLTAVPLSALVDEPVHIQVTGLAPFQVVCLQASLKDEKGNLFSSQAFYRASEVGEVDL)

MRMPath …



Isocitrate dehydrogenase

MRMPath results for isocitrate dehydrogenase



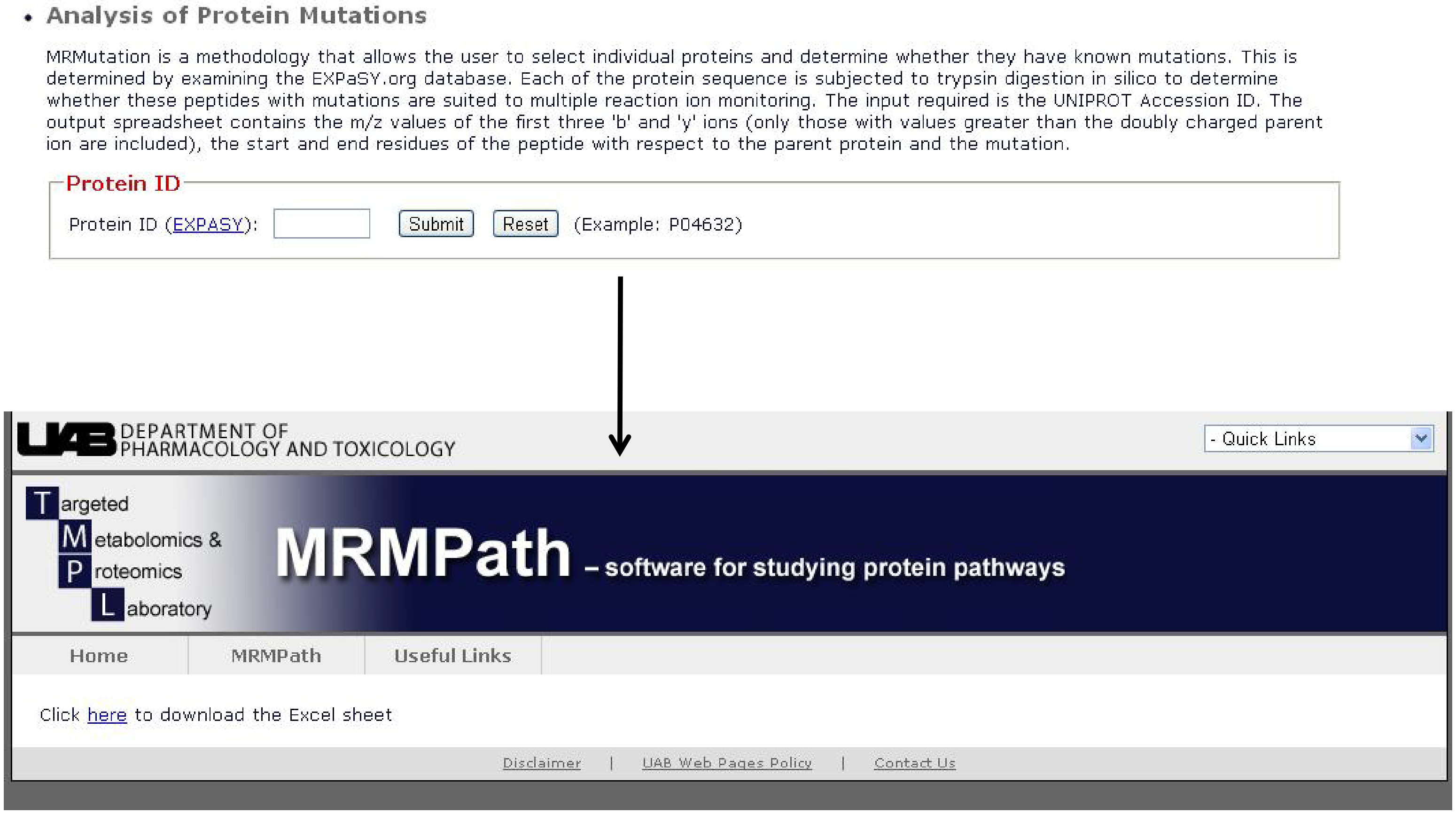
Click here to download this into an Excel sheet
NOTE: Please click on the 'YES' button if a warning appears when you try to open the excel sheet

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

Sequence	m/z Parent Ion	B Ion Mass	Y > Parent Ions
BLAST IIWELIK	457.792	542.2979	801.4921
		655.3819	688.4080
		768.4660	502.3287
		896.5609	
BLAST LIFPYVELDLHSYDLGIENR	1203.6235	1203.6666	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
		973.4591	
BLAST DAAEAIK	359.1954	387.1516	602.3560
		458.1887	531.3189
		571.2728	460.2818
		699.3678	
BLAST SPNGTIR	372.7065	457.2047	657.3731
		570.2887	560.3204
		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)

Click here to download the Excel sheet

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