



Essential Computing for Bioinformatics

Lecture 2

Using Bioinformatics Data Sources

MARC: Developing Bioinformatics
Programs
July 2009

Alex Ropelewski
PSC-NRBSC
Bienvenido Vélez
UPR Mayaguez



Essential Computing for Bioinformatics

- The following material is the result of a curriculum development effort to provide a set of courses to support bioinformatics efforts involving students from the biological sciences, computer science, and mathematics departments. They have been developed as a part of the NIH funded project “Assisting Bioinformatics Efforts at Minority Schools” (2T36 GM008789). The people involved with the curriculum development effort include:
 - Dr. Hugh B. Nicholas, Dr. Troy Wymore, Mr. Alexander Ropelewski and Dr. David Deerfield II, National Resource for Biomedical Supercomputing, Pittsburgh Supercomputing Center, Carnegie Mellon University.
 - Dr. Ricardo González Méndez, University of Puerto Rico Medical Sciences Campus.
 - Dr. Alade Tokuta, North Carolina Central University.
 - Dr. Jaime Seguel and Dr. Bienvenido Vélez, University of Puerto Rico at Mayagüez.
 - Dr. Satish Bhalla, Johnson C. Smith University.
- Unless otherwise specified, all the information contained within is Copyrighted © by Carnegie Mellon University. Permission is granted for use, modify, and reproduce these materials for teaching purposes.
- Most recent versions of these presentations can be found at <http://marc.psc.edu/>

Course Outline

- Course Overview
- Introduction to Information Needs and Databases
- Unstructured Data Repositories
 - Query models and implementation issues
- Structured Data Repositories
 - Query models and implementation issues
- Biology-specific Repositories
 - Query models and implementation issues



Outline



Definition: Biological Database

- Any repository containing Biological information which can be used to:
 - assess the current state of knowledge
 - Formulate new scientific hypotheses
 - Validate these hypotheses
- Some Examples of Biological Databases
 - Sequence
 - Structure
 - Family/Domain
 - Species
 - Taxonomy
 - Function/Pathway
 - Disease/Variation
 - Publication Journal
 - And many other ways

How is Biological Information Stored?

- From a computer-science perspective, there are several ways that data can be organized and stored:
 - In a flat text file
 - In a spreadsheet
 - In an image
 - In an video animation
 - In a relational database
 - In a networked (hyperlinked) model
 - In any combination of the above
 - Others

Sequence Data Libraries

- Organized according to sequence
- When one talks about “searching sequence databases” these are the libraries that they are searching
- Main sources for sequence libraries are direct submissions from individual researchers, genome sequencing projects, patent applications and other public resources.
 - Genbank, EMBL, and the DNA Database of Japan (DDBJ) are examples of annotated collections publicly available DNA sequences.
 - The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data

Structural Data Libraries

- Contain information about the (3-dimensional) structure of the molecule
- Main sources of structural data are direct submissions from researchers. Data can be submitted via a variety of experimental techniques including
 - X-ray crystallography
 - NMR structure depositions.
 - EM structure depositions.
 - Other methods (including Electron diffraction, Fiber diffraction).
- The Protein Data Bank and the Cambridge Structural Database are two well-known repositories of structural information

Family and Domain Libraries

- Typically built from sets of related sequences and contain information about the residues that are essential to the structure/function of the sequences
- Used to:
 - Generate a hypothesis that the query sequence has the same structure/function as the matching group of sequences.
 - Quickly identify a good group of sequences known to share a biological relationship.
- Some examples:
 - PFAM, Prosite, BLOCKS, PRINTS

Species Libraries

- Goal is to collect and organize a variety of information concerning the genome of a particular species
- Usually each species has its own portal to access information such as genomic-scale datasets for the species.
- Examples:
 - EuPathDB - Eukaryotic Pathogens Database (Cryptosporidium, Giardia, Plasmodium, Toxoplasma and Trichomonas)
 - Saccharomyces Genome Database
 - Rat Genome Database
 - Candida Genome Database

Taxonomy Libraries

- The science of naming and classifying organisms
- Taxonomy is organized in a tree structure, which represents the taxonomic lineage.
- Bottom level leafs represents species or sub-species
- Top level nodes represent higher ranks like phylum, order and family
- Examples:
 - NEWT
 - NCBI Taxonomy

Taxonomy Libraries - NEWT

Danio rerio (Zebrafish) (Brachydanio rerio)

Lineage	Taxonomy identifier	7955	External information
<ul style="list-style-type: none"> ◆ Eukaryota ◆ Metazoa ◆ Chordata ◆ Craniata ◆ Vertebrata ◆ Euteleostomi ◆ Actinopterygii ◆ Neopterygii ◆ Teleostei ◆ Ostariophysi ◆ Cypriniformes ◆ Cyprinidae ◆ Danio 	Organism identification code	DANRE	 <p> http://en.wikipedia.org/wiki/Brachydanio_rerio http://nis.gsmfc.org/nis_factsheet.php?toc_id=169 http://www.itis.gov/servlet/SingleRpt/SingleRpt?search_t </p>
	Scientific name	Danio rerio	
	Common name	Zebrafish	
	Synonym	Brachydanio rerio	
	Other NCBI synonyms	Cyprinus rerio Hamilton, 1822 Danio rerio (Hamilton, 1822) zebra fish Cyprinus rerio zebra danio Brachidanio rerio leopard danio	
	Rank	species	
	Number of UniProtKB/Swiss-Prot entries	1864	
Number of UniProtKB/TrEMBL entries	22498		

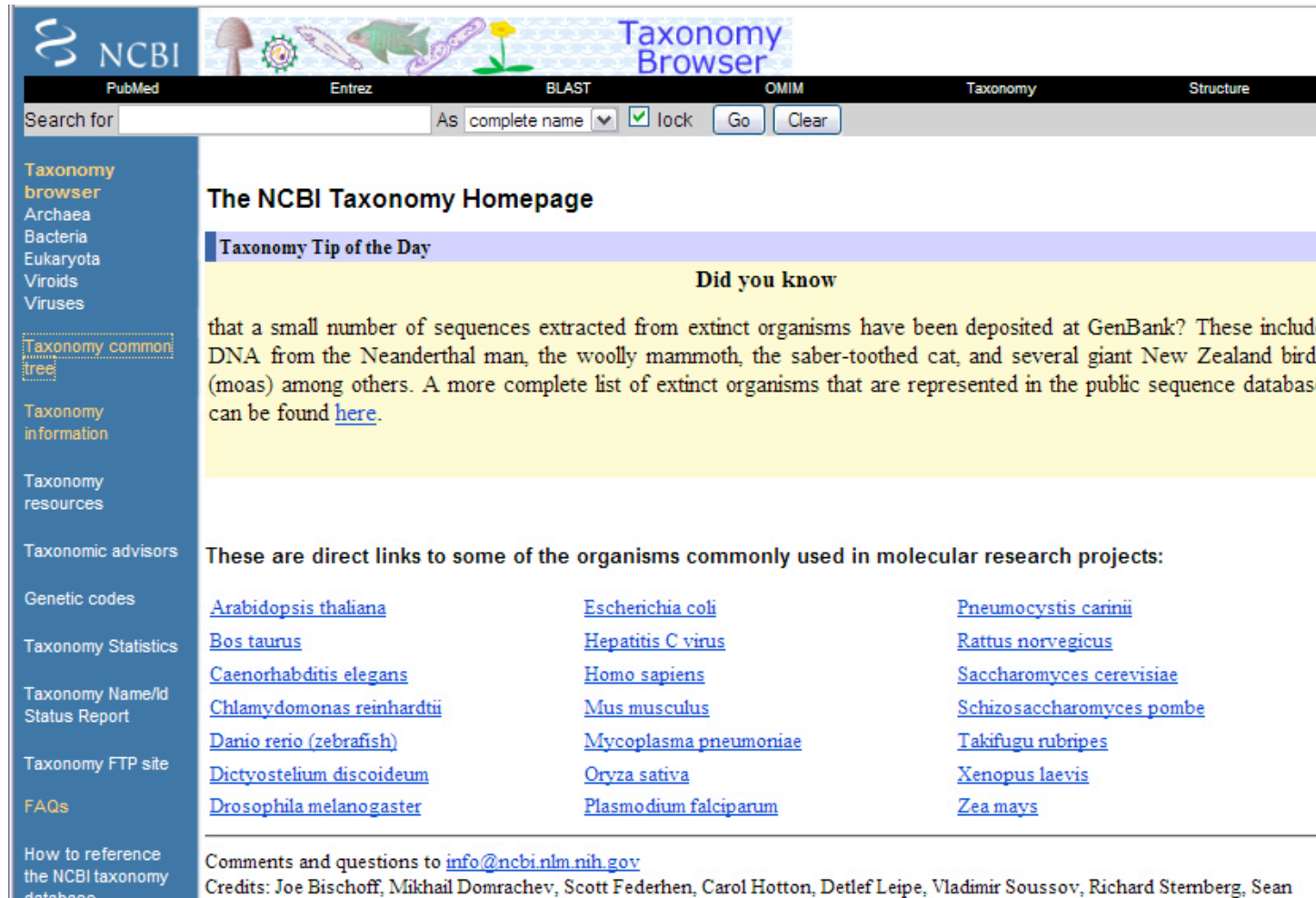
Taxonomy navigation	
Up taxonomy tree	Down taxonomy tree
Danio	◆ <i>This is the last node of the tree</i>

[+] **List of strains names** (and synonyms) **for this organism** [more information]

[Complete proteome information](#)

Source of data : Swiss-Prot [NCBI taxonomy for this taxon](#)

NCBI Taxonomy Browser



The screenshot shows the NCBI Taxonomy Browser interface. At the top, there is a navigation bar with links to PubMed, Entrez, BLAST, OMIM, Taxonomy, and Structure. Below this is a search bar with a dropdown menu set to 'As complete name', a 'lock' checkbox, and 'Go' and 'Clear' buttons. The main content area is titled 'The NCBI Taxonomy Homepage' and features a 'Taxonomy Tip of the Day' section with a 'Did you know' header. The tip discusses sequences from extinct organisms in GenBank. Below this is a section titled 'These are direct links to some of the organisms commonly used in molecular research projects:' followed by a grid of organism names with blue underlined links. A footer section contains contact information and credits.

Search for As complete name lock

NCBI Taxonomy Browser

PubMed Entrez BLAST OMIM Taxonomy Structure

The NCBI Taxonomy Homepage

Taxonomy Tip of the Day

Did you know

that a small number of sequences extracted from extinct organisms have been deposited at GenBank? These include DNA from the Neanderthal man, the woolly mammoth, the saber-toothed cat, and several giant New Zealand birds (moas) among others. A more complete list of extinct organisms that are represented in the public sequence database can be found [here](#).

These are direct links to some of the organisms commonly used in molecular research projects:

Arabidopsis thaliana	Escherichia coli	Pneumocystis carinii
Bos taurus	Hepatitis C virus	Rattus norvegicus
Caenorhabditis elegans	Homo sapiens	Saccharomyces cerevisiae
Chlamydomonas reinhardtii	Mus musculus	Schizosaccharomyces pombe
Danio rerio (zebrafish)	Mycoplasm pneumoniae	Takifugu rubripes
Dictyostelium discoideum	Oryza sativa	Xenopus laevis
Drosophila melanogaster	Plasmodium falciparum	Zea mays

Comments and questions to info@ncbi.nlm.nih.gov
 Credits: Joe Bischoff, Mikhail Domrachev, Scott Federhen, Carol Hotton, Detlef Leipe, Vladimir Soussov, Richard Stenberg, Sean

Function/Pathway

- Collection of pathway maps representing our knowledge on the molecular interaction and reaction networks for:
 - Metabolism
 - Genetic Information Processing
 - Environmental Information Processing
 - Cellular Processes
 - Human Diseases
 - Drug Development
- Examples:
 - KEGG Pathway Database
 - NCI-Nature Pathway Interaction Database

Disease/Variation

- Catalogs of genes involving variations including within populations and among populations in different parts of the world as well as genetic disorders and other diseases.
- Examples:
 - OMIM, Online Mendelian Inheritance in Man - focuses primarily on inherited, or heritable, genetic diseases in humans
 - HapMap - a catalog of common genetic variants that occur in humans.

Journal

- U.S. National Library of Medicine
- PubMed is the premiere resources for scientific literature relevant to the biomedical sciences.
- Includes over 18 million citations from MEDLINE and other life science journals for articles back to the 1950s.
- PubMed includes links to full text articles and other related resources.
- Common uses of PubMed:
 - Find journal articles that describe the structure/function/evolution of sequences that you are interested in
 - Find out if anyone has already done the work that you are proposing

Current databases are loosely integrated

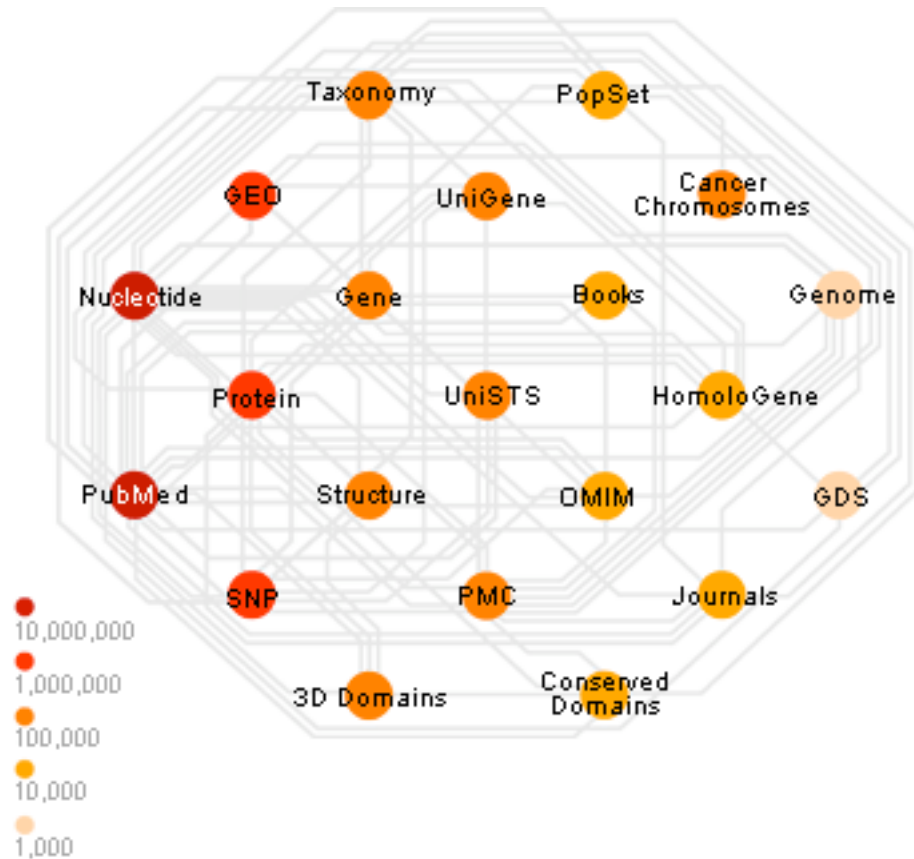
- In order to prove a hypothesis one must often collect information from several independent databases and tools
- Lots of time are spent converting data back and forth among the multiple specific formats required by the various tools and databases
- Discovery process may take a long time, weeks or even months, to complete and tools do not effectively assist the scientist in saving intermediate results in order to continue the search from that point at a later time.

What has been done about this?

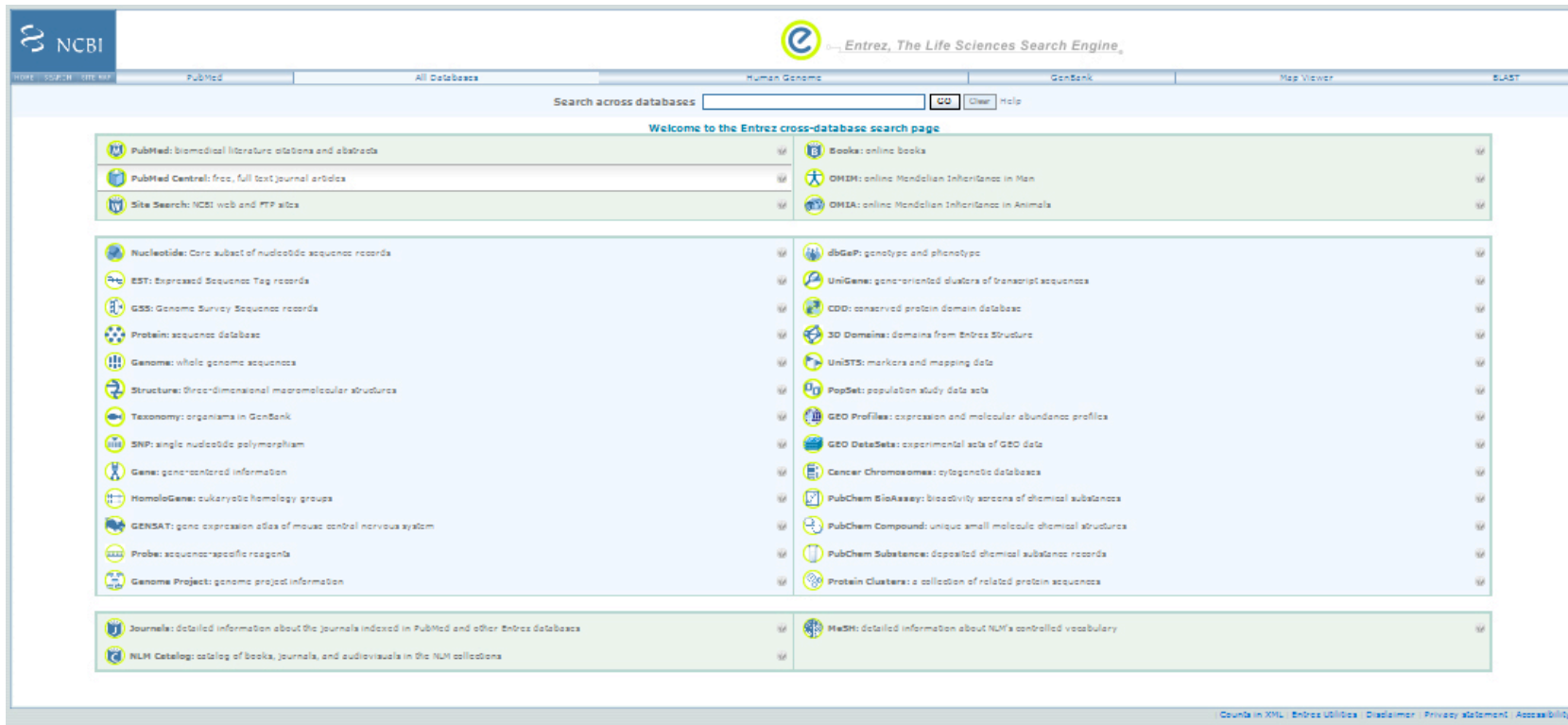
Integrated Information Resources

- Integrated resources typically use a combination of relational databases and hyperlinks to databases maintained by others to provide more information than any single data source can provide
- Many Examples:
 - NCBI Entrez – NCBI’s cross-database tool
 - iProClass - proteins with links to over 90 biological databases. including databases for protein families, functions and pathways, interactions, structures and structural classifications, genes and genomes, ontologies, literature, and taxonomy
 - InterPro - Integrated Resource Of Protein Domains And Functional Sites.

NCBI Entrez Data Integration



NCBI Entrez

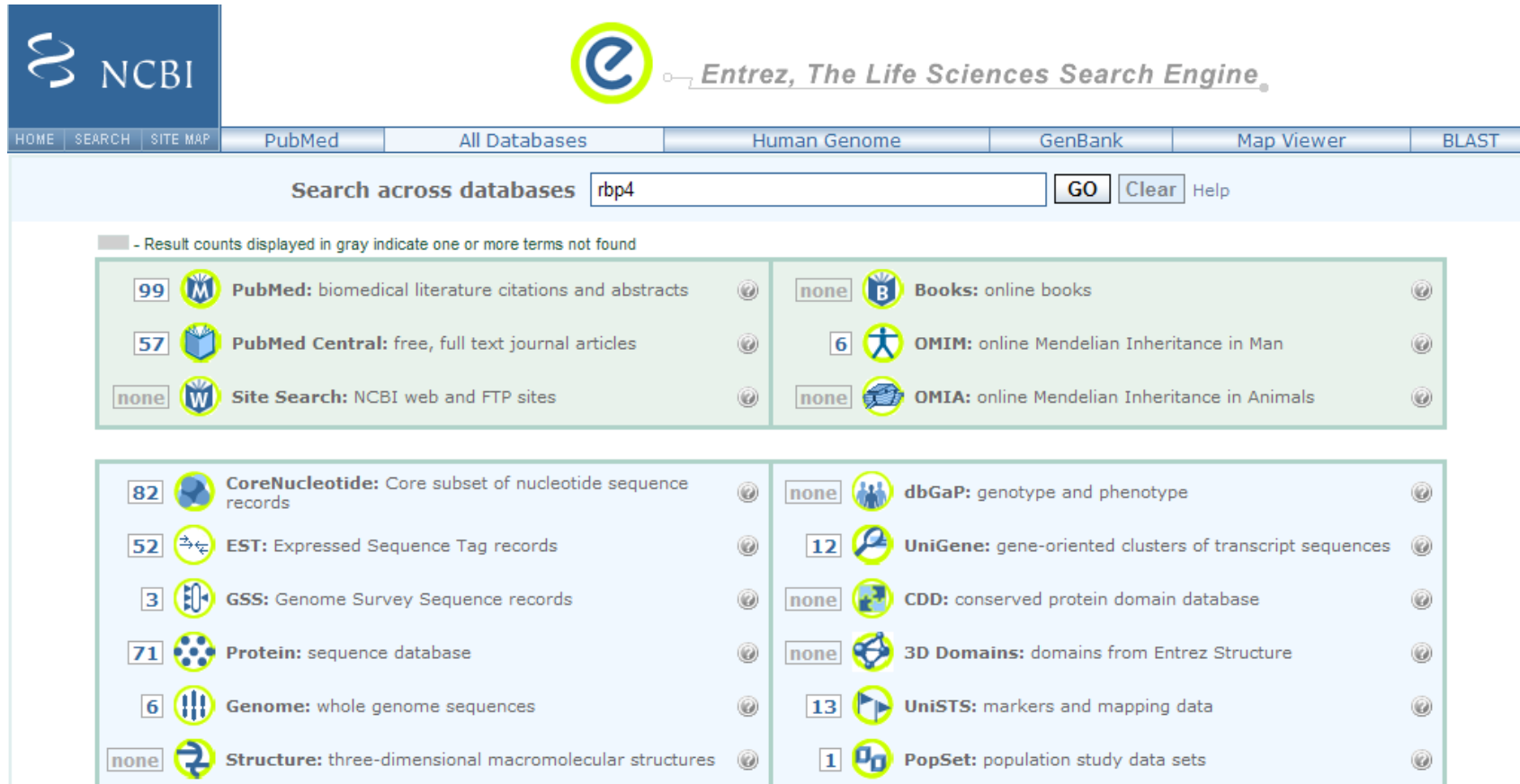


The screenshot shows the NCBI Entrez search engine interface. At the top, there is a navigation bar with tabs for PubMed, All Databases, Human Genome, GenBank, and Map Viewer. Below this is a search bar with a search button and a 'GO' button. The main content area is titled 'Welcome to the Entrez cross-database search page' and displays a grid of database categories, each with an icon and a brief description. The categories include:

- PubMed: biomedical literature citations and abstracts
- PubMed Central: free, full text journal articles
- Site Search: NCBI web and FTP sites
- Nucleotide: Core subset of nucleotide sequence records
- EST: Expressed Sequence Tag records
- GSS: Genome Survey Sequence records
- Protein: sequence database
- Genome: whole genome sequences
- Structure: three-dimensional macromolecular structures
- Taxonomy: organisms in GenBank
- SNP: single nucleotide polymorphism
- Gene: gene-centered information
- HomoloGene: eukaryotic homology groups
- GENSAT: gene expression atlas of mouse central nervous system
- Probe: sequence-specific reagents
- Genome Project: genome project information
- Books: online books
- OMIM: online Mendelian Inheritance in Man
- OMIA: online Mendelian Inheritance in Animals
- dbGAP: genotype and phenotype
- UniGene: gene-oriented clusters of transcript sequences
- CDD: conserved protein domain database
- 3D Domains: domains from Entrez Structure
- UniSTS: markers and mapping data
- PopSet: population study data sets
- GEO Profiles: expression and molecular abundance profiles
- GEO DataSets: experimental sets of GEO data
- Cancer Chromosomes: cytogenetic databases
- PubChem BioAssay: bioactivity assays of chemical substances
- PubChem Compound: unique small molecule chemical structures
- PubChem Substance: deposited chemical substance records
- Protein Clusters: a collection of related protein sequences
- Journals: detailed information about the journals indexed in PubMed and other Entrez databases
- NLM Catalog: catalog of books, journals, and audiovisuals in the NLM collections
- MeSH: detailed information about NLM's controlled vocabulary

At the bottom right of the page, there are links for 'Search in XML', 'Entrez Utilities', 'Disclaimer', 'Privacy Statement', and 'Accessibility'.

NCBI Entrez Results





















NCBI Entrez, The Life Sciences Search Engine

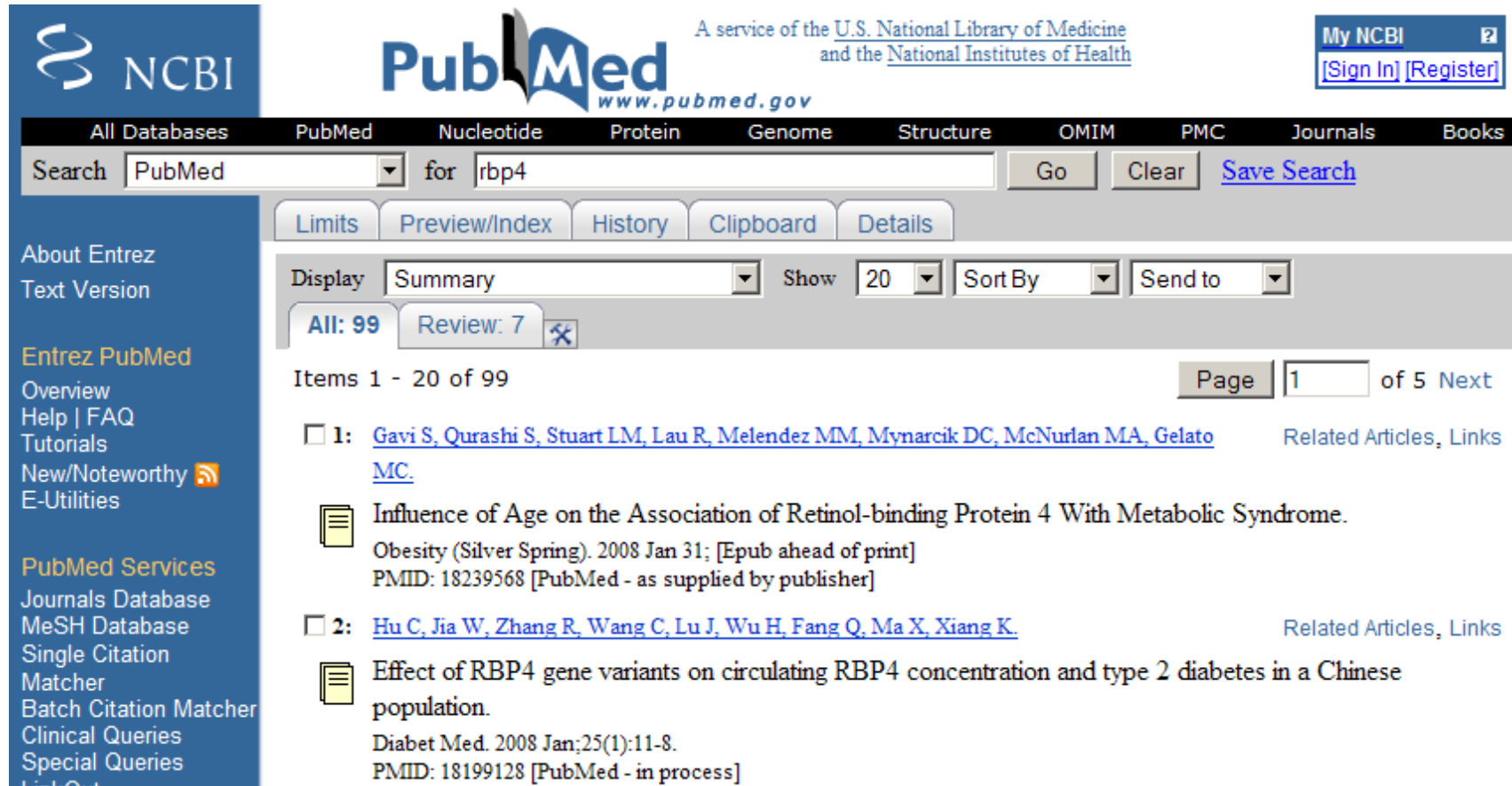
HOME | SEARCH | SITE MAP | PubMed | All Databases | Human Genome | GenBank | Map Viewer | BLAST

Search across databases [Help](#)

- Result counts displayed in gray indicate one or more terms not found

99		PubMed: biomedical literature citations and abstracts	<input type="checkbox"/>	none		Books: online books	<input type="checkbox"/>
57		PubMed Central: free, full text journal articles	<input type="checkbox"/>	6		OMIM: online Mendelian Inheritance in Man	<input type="checkbox"/>
none		Site Search: NCBI web and FTP sites	<input type="checkbox"/>	none		OMIA: online Mendelian Inheritance in Animals	<input type="checkbox"/>
82		CoreNucleotide: Core subset of nucleotide sequence records	<input type="checkbox"/>	none		dbGaP: genotype and phenotype	<input type="checkbox"/>
52		EST: Expressed Sequence Tag records	<input type="checkbox"/>	12		UniGene: gene-oriented clusters of transcript sequences	<input type="checkbox"/>
3		GSS: Genome Survey Sequence records	<input type="checkbox"/>	none		CDD: conserved protein domain database	<input type="checkbox"/>
71		Protein: sequence database	<input type="checkbox"/>	none		3D Domains: domains from Entrez Structure	<input type="checkbox"/>
6		Genome: whole genome sequences	<input type="checkbox"/>	13		UniSTS: markers and mapping data	<input type="checkbox"/>
none		Structure: three-dimensional macromolecular structures	<input type="checkbox"/>	1		PopSet: population study data sets	<input type="checkbox"/>

NCBI Entrez PubMed Results



The screenshot shows the NCBI Entrez PubMed search results for the query 'rbp4'. The search was performed in the PubMed database. The results are displayed in a list format, showing the first two items. The first item is a review article titled 'Influence of Age on the Association of Retinol-binding Protein 4 With Metabolic Syndrome' by Gavi S, et al. The second item is an article titled 'Effect of RBP4 gene variants on circulating RBP4 concentration and type 2 diabetes in a Chinese population' by Hu C, et al. The search results are displayed on page 1 of 5.

NCBI **PubMed** A service of the U.S. National Library of Medicine and the National Institutes of Health [www.pubmed.gov](#) [My NCBI](#) [\[Sign In\]](#) [\[Register\]](#)

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for rbp4 [Go](#) [Clear](#) [Save Search](#)

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort By Send to

All: 99 Review: 7

Items 1 - 20 of 99 Page 1 of 5 Next

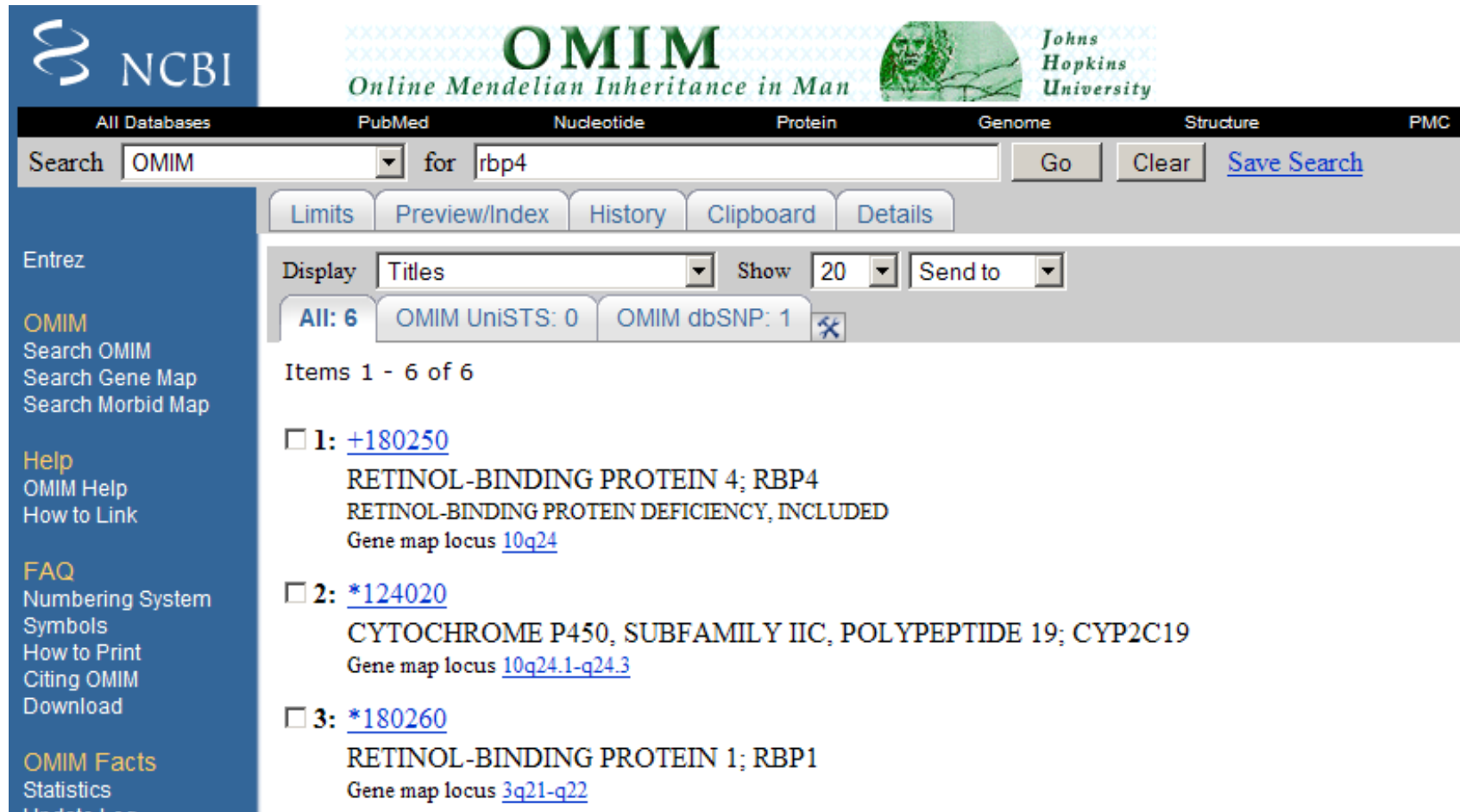
1: [Gavi S, Qurashi S, Stuart LM, Lau R, Melendez MM, Mynarcik DC, McNurlan MA, Gelato MC.](#) Related Articles, Links

Influence of Age on the Association of Retinol-binding Protein 4 With Metabolic Syndrome. Obesity (Silver Spring). 2008 Jan 31; [Epub ahead of print] PMID: 18239568 [PubMed - as supplied by publisher]


2: [Hu C, Jia W, Zhang R, Wang C, Lu J, Wu H, Fang Q, Ma X, Xiang K.](#) Related Articles, Links

Effect of RBP4 gene variants on circulating RBP4 concentration and type 2 diabetes in a Chinese population. Diabet Med. 2008 Jan;25(1):11-8. PMID: 18199128 [PubMed - in process]

NCBI Entrez OMIM Results



The screenshot shows the NCBI Entrez OMIM search interface. The search was performed for 'rbp4' in the OMIM database. The results are displayed in a list format, showing the first three items. The interface includes a search bar, navigation tabs (All Databases, PubMed, Nucleotide, Protein, Genome, Structure, PMC), and a sidebar with navigation links.

NCBI **OMIM** Online Mendelian Inheritance in Man  Johns Hopkins University

All Databases PubMed Nucleotide Protein Genome Structure PMC

Search OMIM for rbp4 Go Clear Save Search

Limits Preview/Index History Clipboard Details

Entrez

OMIM
Search OMIM
Search Gene Map
Search Morbid Map

Help
OMIM Help
How to Link

FAQ
Numbering System
Symbols
How to Print
Citing OMIM
Download

OMIM Facts
Statistics
Update Log

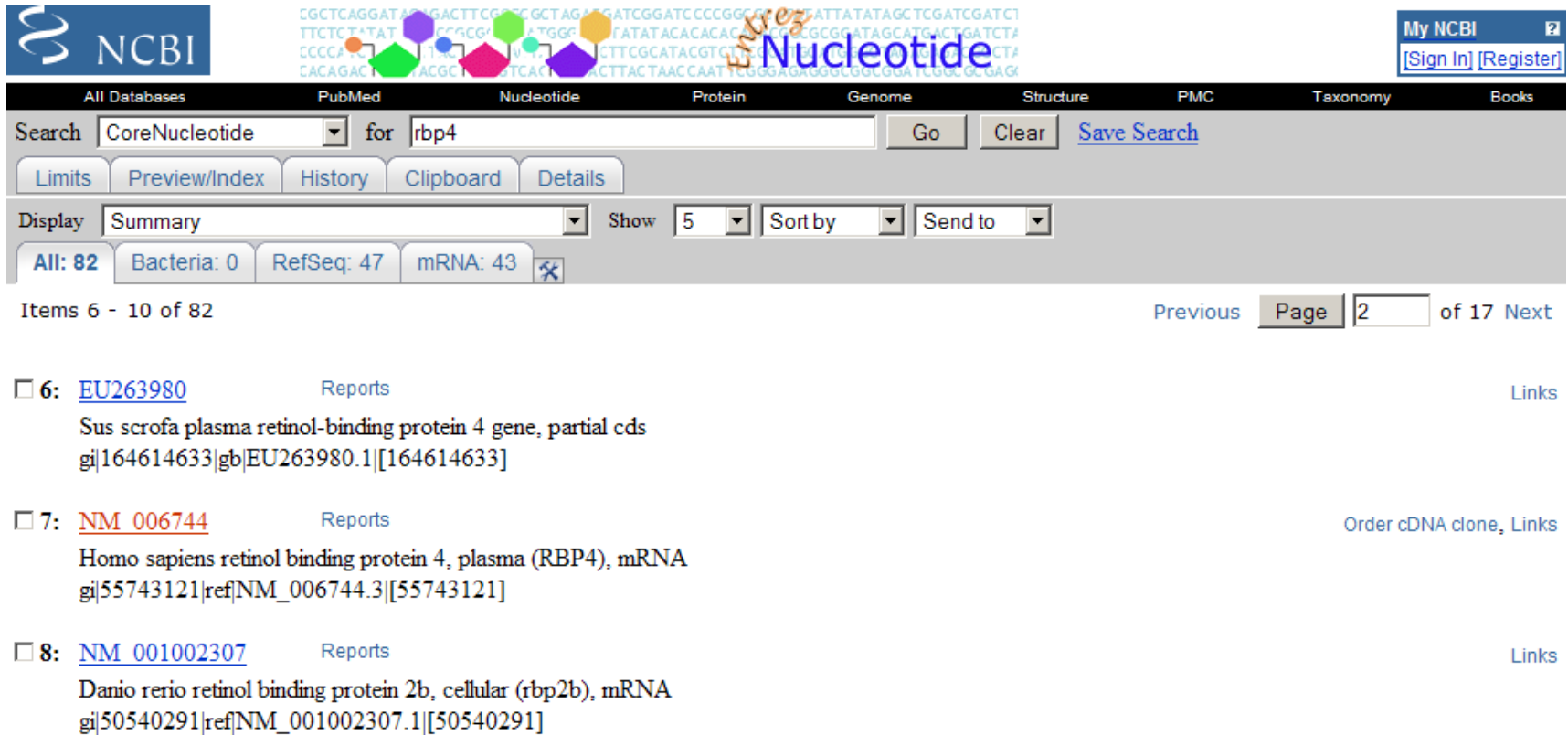
Display Titles Show 20 Send to

All: 6 OMIM UniSTS: 0 OMIM dbSNP: 1

Items 1 - 6 of 6

- 1: [+180250](#)
RETINOL-BINDING PROTEIN 4; RBP4
RETINOL-BINDING PROTEIN DEFICIENCY, INCLUDED
Gene map locus [10q24](#)
- 2: [*124020](#)
CYTOCHROME P450, SUBFAMILY IIC, POLYPEPTIDE 19; CYP2C19
Gene map locus [10q24.1-q24.3](#)
- 3: [*180260](#)
RETINOL-BINDING PROTEIN 1; RBP1
Gene map locus [3q21-q22](#)

NCBI Entrez Core Nucleotide Results



NCBI

CGCTCAGGATAGGACTTCCTCCGCTAGAGGATCGGATCCCCGGGATGATTATATAGCTCGATCGATCT
 TTCTCTATATCCGCGGATATGGGATATACACACACATCCGCGGATAGCATGACTGATCTA
 CCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
 CACAGACTCTACGCTCTCACTCTTACTTAACCAATTCGGGAGAGGGCCGCGGATCGGCGGAG

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search CoreNucleotide for rbp4 Go Clear Save Search

Limits Preview/Index History Clipboard Details

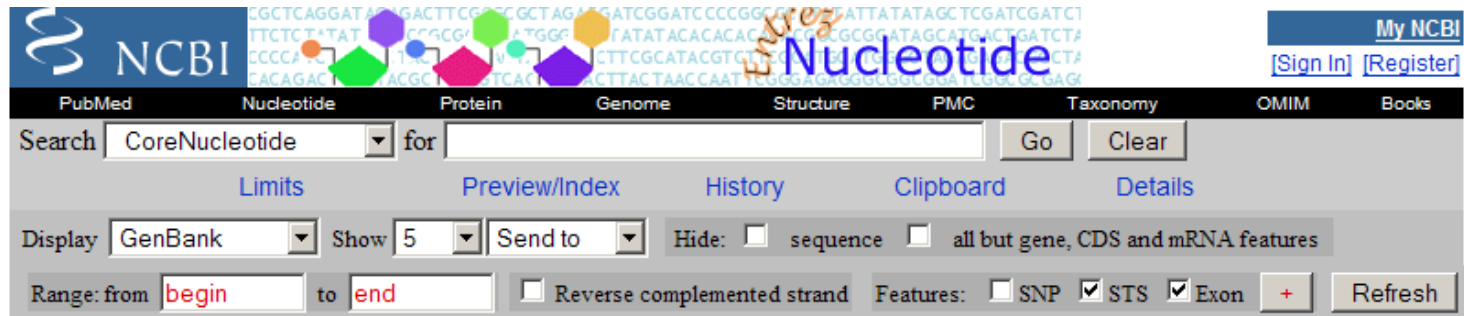
Display Summary Show 5 Sort by Send to

All: 82 Bacteria: 0 RefSeq: 47 mRNA: 43

Items 6 - 10 of 82 Previous Page 2 of 17 Next

- 6: [EU263980](#) Reports Links
 Sus scrofa plasma retinol-binding protein 4 gene, partial cds
 gi|164614633|gb|EU263980.1|[164614633]
- 7: [NM_006744](#) Reports Order cDNA clone, Links
 Homo sapiens retinol binding protein 4, plasma (RBP4), mRNA
 gi|55743121|ref|NM_006744.3|[55743121]
- 8: [NM_001002307](#) Reports Links
 Danio rerio retinol binding protein 2b, cellular (rbp2b), mRNA
 gi|50540291|ref|NM_001002307.1|[50540291]

NCBI Entrez Core Nucleotide Results



The screenshot shows the NCBI Entrez Nucleotide search interface. The search term 'CoreNucleotide' is entered in the search box. The interface includes navigation tabs for PubMed, Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, OMIM, and Books. Below the search box are options for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The 'Display' dropdown is set to 'GenBank', and 'Show' is set to '5'. There are checkboxes for 'Reverse complemented strand', 'SNP', 'STS', and 'Exon'. A 'Refresh' button is also present.

1: [NM_006744](#). Reports Homo sapiens reti...[gi:55743121] Order cDNA clone, Links

[Comment](#) [Features](#) [Sequence](#)

LOCUS NM_006744 941 bp mRNA linear PRI 13-JAN-2008
 DEFINITION Homo sapiens retinol binding protein 4, plasma (RBP4), mRNA.
 ACCESSION NM_006744
 VERSION NM_006744.3 GI:55743121
 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 941)
 AUTHORS Espe,K., Galler,A., Raila,J., Kiess,W. and Schweigert,F.J.
 TITLE High-normal C-reactive protein levels do not affect the vitamin A
 transport complex in serum of children and adolescents with type 1
 diabetes
 JOURNAL Pediatr. Res. 62 (6), 741-745 (2007)
 PUBMED [17957146](#)
 REMARK GeneRIF: Serum RBP4 and TTR showed no differences between
 controls/type 1 diabetic children.

NCBI Entrez Core Nucleotide Results

CDS

```

85..690
/gene="RBP4"
/GO_component="extracellular region [PMID 14718574];
extracellular space [PMID 6316270]"
/GO_function="binding; retinal binding; retinol binding;
transporter activity"
/GO_process="response to stimulus; transport; visual
perception"
/note="retinol-binding protein 4, plasma; retinol-binding
protein 4, interstitial"
/codon_start=1
/product="retinol-binding protein 4, plasma precursor"
/protein_id="NP_006735.2"
/db_xref="GI:55743122"
/db_xref="CCDS:CCDS31249.1"
/db_xref="GeneID:5950"
/db_xref="HGNC:9922"
/db_xref="HPRD:01580"
/db_xref="MIM:180250"
/translation="MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYA
MAKKDPEGLFLQDNIVAEFSVDETGQMSATAKGRVRLNNDVDCADMVGTFTDTEPA
KFKMKYWGVASFQKGNDDHWIVDTDYDTYAVQYSCRLNLDGTCADSYSFVFSRDPN
GLPPEAQKIVRQRQEELCLARQYRLIVHNGYCDGRSERLL"

```

sig peptide

85..138

```
/gene="RBP4"
```

mat peptide

139..687

```
/gene="RBP4"
```

```
/product="retinol-binding protein 4, plasma"
```

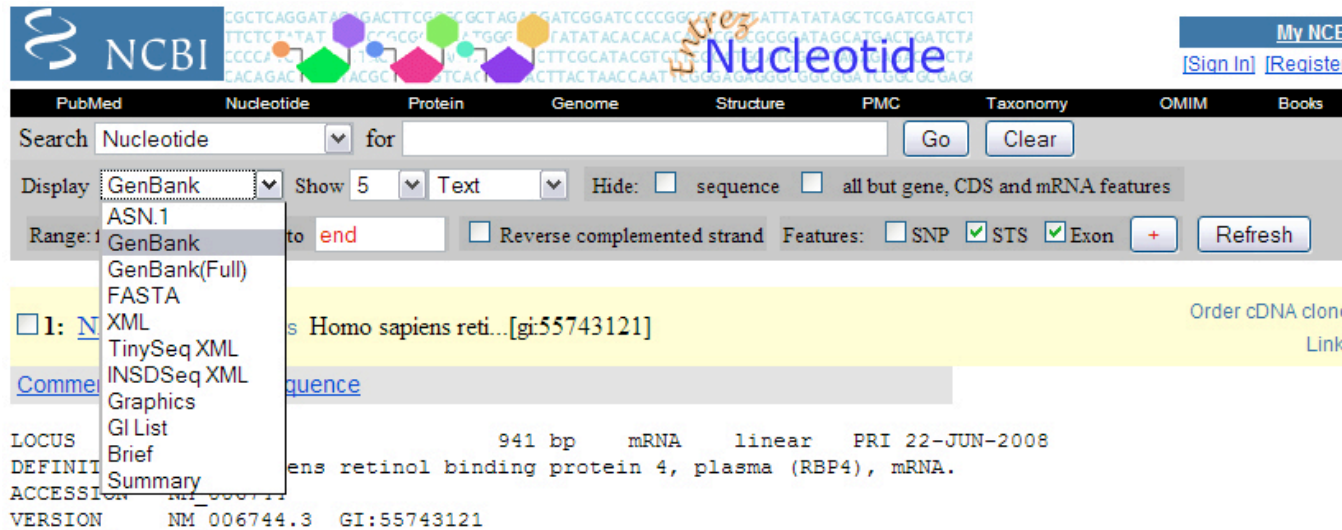
NCBI Entrez Core Nucleotide Results

ORIGIN

```
1 cgctccctc gctccacgcg cgcccggact cggcggccag gcttgcgcgc ggttcccctc
61 ccggtgggcg gattcctggg caagatgaag tgggtgtggg cgctcttgct gttggcggcg
121 ctgggcagcg gccgcgcgga gcgcgactgc cgagtgaaca gcttccgagt caaggagaac
181 ttcgacaagg ctgccttctc tgggacctgg tacgccatgg ccaagaagga ccccgagggc
241 ctctttctgc aggacaacat cgtcgcggag ttctccgtgg acgagaccgg ccagatgagc
301 gccacagcca agggccgagt ccgtcttttg aataactggg acgtgtgcgc agacatggtg
361 ggcaccttca cagacaccga ggaccctgcc aagttcaaga tgaagtactg gggcgtagcc
421 tcctttctcc agaaaggaaa tgatgaccac tggatcgtcg acacagacta cgacacgtat
481 gccgtgcagt actcctgccg cctcctgaac ctcgatggca cctgtgctga cagctactcc
541 ttcgtgtttt cccgggaccg caacggcctg ccccagaag cgcagaagat tgtaaggcag
601 cggcaggagg agctgtgcct ggccaggcag tacaggctga tcgtccaca cggttactgc
661 gatggcagat cagaaagaaa cctttttag caatatcaag aatctagtt catctgagaa
721 cttctgatta gctctcagtc ttcagctcta tttatcttag gagtttaatt tgccttctc
781 tccccatctt ccctcagttc ccataaaacc ttcattacac ataaagatac acgtgggggt
841 cagtgaatct gcttgcctt cctgaaagt tctggggctt aagattccag actctgattc
901 attaaactat agtcaccogt gtctgtgtaa aaaaaaaaa a
```

//

NCBI Entrez Saving Sequences



NCBI Nucleotide

Search: Nucleotide for [] Go Clear

Display: GenBank Show 5 Text Hide: sequence all but gene, CDS and mRNA features

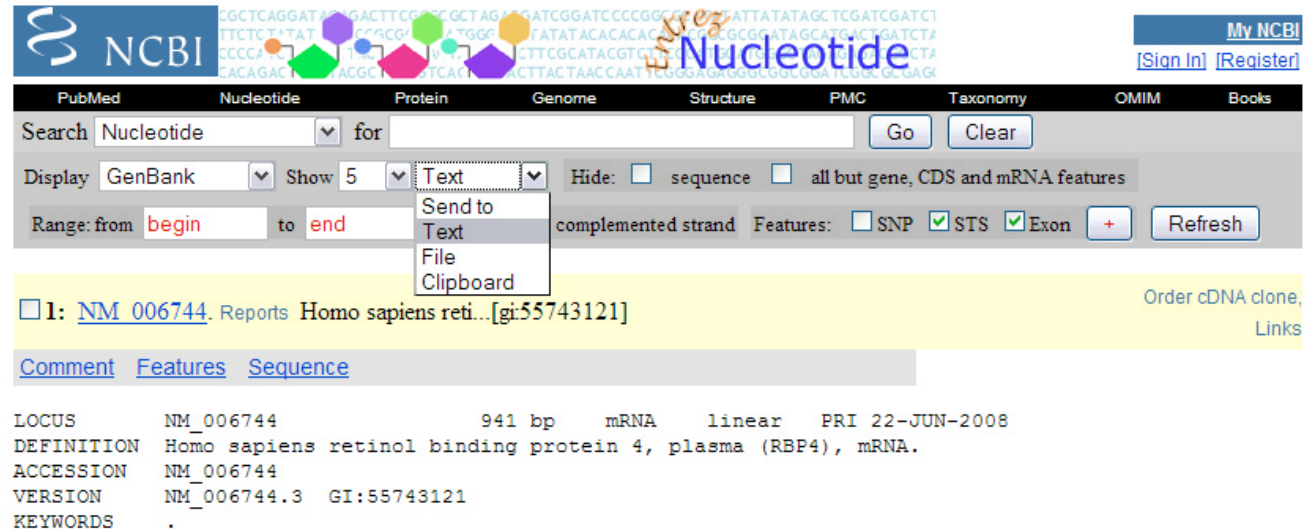
Range: 1 to end Reverse complemented strand Features: SNP STS Exon Refresh

1: [NM_006744.3](#) Homo sapiens reti...[gi:55743121] Order cDNA clone, Links

941 bp mRNA linear PRI 22-JUN-2008

Definition: Homo sapiens retinol binding protein 4, plasma (RBP4), mRNA.

Accession: NM_006744.3 GI:55743121



NCBI Nucleotide

Search: Nucleotide for [] Go Clear

Display: GenBank Show 5 Text Hide: sequence all but gene, CDS and mRNA features

Range: from begin to end Reverse complemented strand Features: SNP STS Exon Refresh

1: [NM_006744](#). Reports Homo sapiens reti...[gi:55743121] Order cDNA clone, Links

Comment Features Sequence

LOCUS NM_006744 941 bp mRNA linear PRI 22-JUN-2008

DEFINITION Homo sapiens retinol binding protein 4, plasma (RBP4), mRNA.

ACCESSION NM_006744

VERSION NM_006744.3 GI:55743121

KEYWORDS .

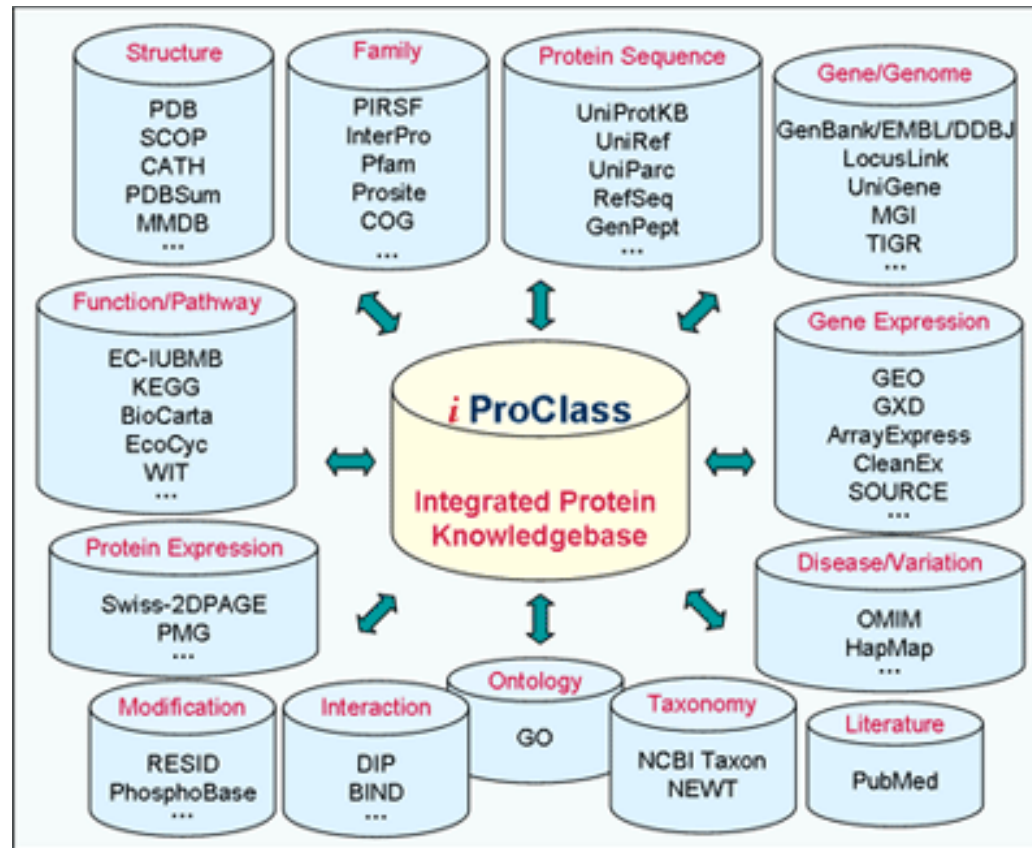
NCBI Sequence Identifiers

- **Accession Number:** unique identifier given to a sequence when it is submitted to one of the DNA repositories (GenBank, EMBL, DDBJ). These identifiers follow an accession.version format. Updates increment the version, while the accession remains constant.
- **GI:** GenInfo Identifier. If a sequence changes a new GI number will be assigned. A separate GI number is also assigned to each protein translation.

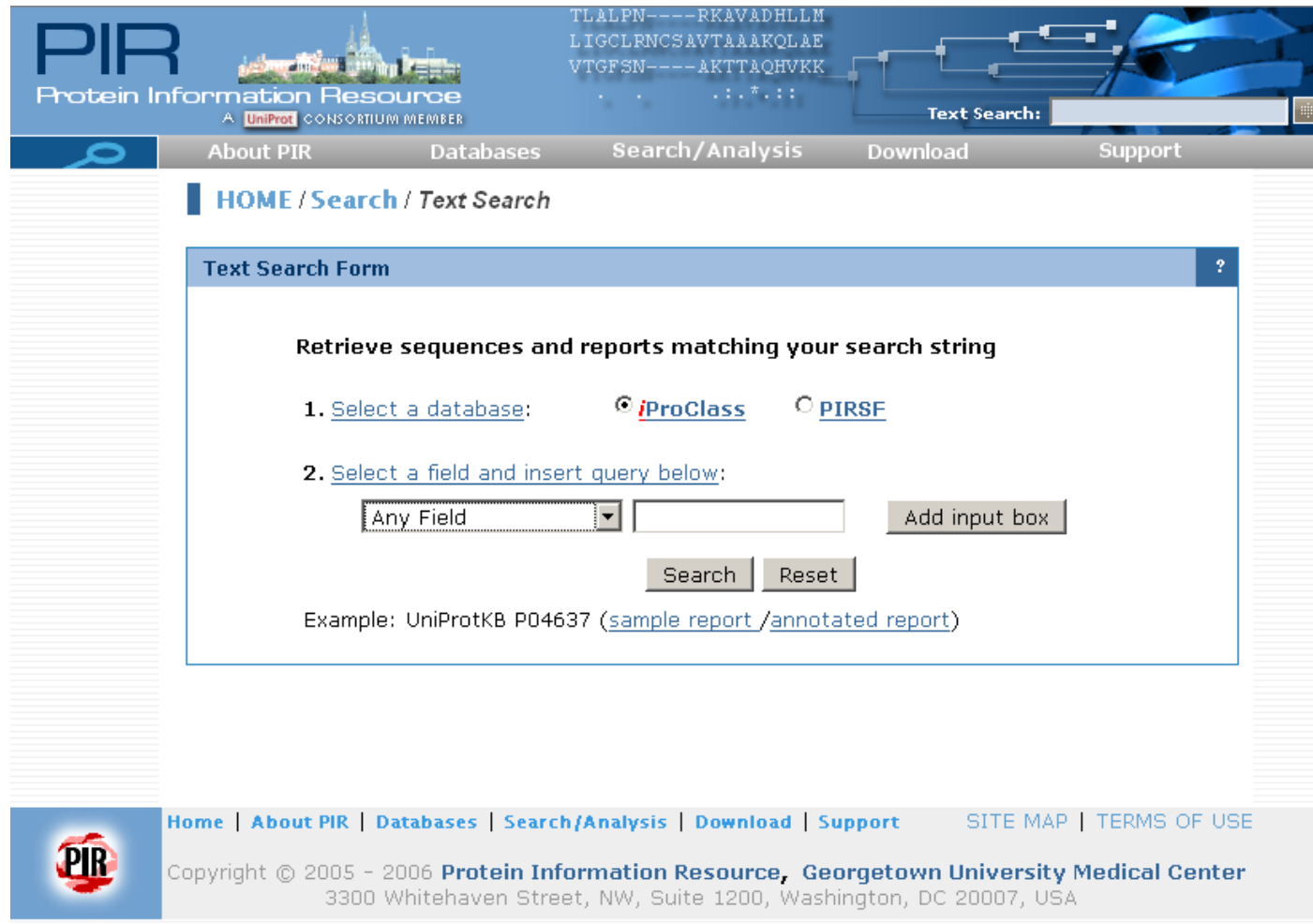
iProClass Protein Knowledgebase

- Protein centric
- Links to over 90 biological data libraries
- Goal is to provide a comprehensive picture of protein properties that may lead to functional inference for previously uncharacterized "hypothetical" proteins and protein groups.
- Uses both data warehousing in relational databases as well as hypertext links to outside data sources

iProclass Integration



iProclass Search Form




The screenshot shows the PIR (Protein Information Resource) website interface. At the top, there is a navigation bar with links for "About PIR", "Databases", "Search/Analysis", "Download", and "Support". The main content area is titled "Text Search Form" and contains the following elements:

- A header: "Retrieve sequences and reports matching your search string"
- Step 1: "Select a database:" with radio buttons for "iProClass" (selected) and "PIRSF".
- Step 2: "Select a field and insert query below:" with a dropdown menu set to "Any Field", an empty text input box, and an "Add input box" button.
- Buttons for "Search" and "Reset".
- An example: "Example: UniProtKB P04637 ([sample report](#) / [annotated report](#))"

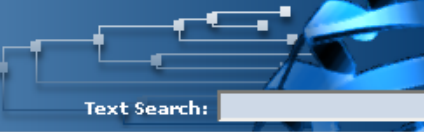
The footer of the page includes a PIR logo, navigation links (Home, About PIR, Databases, Search/Analysis, Download, Support), "SITE MAP" and "TERMS OF USE" links, and copyright information: "Copyright © 2005 - 2006 Protein Information Resource, Georgetown University Medical Center, 3300 Whitehaven Street, NW, Suite 1200, Washington, DC 20007, USA".

iProclass Results



PIR
Protein Information Resource
A UniProt Consortium Member

TLALPN----RKAVADHLLM
 LIGCLRNCSAVTAQAKQLAE
 VTGFPSN----AKTTAQHVKK
 ...*...



Text Search:

About PIR
Databases
Search/Analysis
Download
Support

Text Search Result





26 proteins | 1 page | 50 / page |
Save Result As:

<input type="checkbox"/> Protein AC/ID	Protein Name	Length	Organism Name	PIRSF ID	Related Seq. +	Matched Fields
<input type="checkbox"/> P51657/DHB1_RAT <small>ProClass UniProtKB/Swiss-Prot</small>	Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) ... <small>BioThesaurus</small>	344	Rattus norvegicus (Rat)	PIRSF000095	300	PIRSF Name=>estradiol 17beta dehydrogenase
<input type="checkbox"/> Q1WNP0/DHB1_PANTR <small>ProClass UniProtKB/Swiss-Prot</small>	Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) ... <small>BioThesaurus</small>	328	Pan troglodytes (Chimpanzee)	PIRSF000095	300	PIRSF Name=>estradiol 17beta dehydrogenase
<input type="checkbox"/> P51656/DHB1_MOUSE <small>ProClass UniProtKB/Swiss-Prot</small>	Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) ... <small>BioThesaurus</small>	344	Mus musculus (Mouse)	PIRSF000095	300	PIRSF Name=>estradiol 17beta dehydrogenase
<input type="checkbox"/> P14061/DHB1_HUMAN <small>ProClass UniProtKB/Swiss-Prot</small>	Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) ... <small>BioThesaurus</small>	328	Homo sapiens (Human)	PIRSF000095	300	PIRSF Name=>estradiol 17beta dehydrogenase; Feature=>estradiol 17beta

These images were developed with funding from the US National Institutes of Health grant #2150-CE000070 to the Pittsburgh Supercomputing Center

33

iProClass SuperFamily Summary

GENERAL INFORMATION	
PIRSF Number	PIRSF000095 <i>Curation Status</i> : Full
PIRSF Name	estradiol 17beta-dehydrogenase [Validated]
PIRSF Size	Total Sequence Entries=26 (26 Proteins+0 Fragments)
PIRSF Hierarchy	 (click to see PIRSF family DAG view.)
Taxonomy Range	Eukaryotae=25; Bacteria=1; Archaea=0; Viruses=0; Other=0  (click to see the taxonomic distribution.)
Length Range	Minimum=285; Maximum=344; Average=315; Standard Deviation=17
Keyword	oxidoreductase(25); nadp(4); cytoplasm(4); lipid synthesis(4); steroid biosynthesis(4); nad(2); vision(2); receptor(2); sensory transduction(2); transmembrane(2); polymorphism(2); membrane(2); ovary(1); 3d-structure(1); direct protein sequencing(1); complete proteome(1)
Representative member	iProClass: P14061
Seed Members	iProClass: Q11QD0 ; O12968 ; P14061 ; Q9N126 ; P51656 ; Q7T2J0 ; Q7T2I9 ; Q6PC70 ; Q6RH38 ; Q640Y3 ; Q4TZJ1 ; Q504A4 ; Q4L7K1 ; Q4S966
Alignment and Tree	 (click to generate and display the multiple alignment and tree)
Domain Architecture	PF00106 (To display the domain architecture, click here for seed members; click here for all members.) 
Rule-Based Annotation	<i>Functional Name Rule</i> PIRNR000095-0 : Estradiol 17beta-dehydrogenase 1

iProClass SuperFamily Summary

MEMBERSHIP	
Eukaryotic Member	iProClass: Q11QD0 ; Q1WNP0 ; Q1WNP1 ; Q1WNP2 ; Q1WNP3 ; Q12968 ; Q7LZT0 ; P14061 ; Q9NYR8 ; Q9N126 ; Q790P4 ; P51656 ; P51657 ; Q7T2J0 ; Q7T2I9 ; Q7T2I8 ; Q6PC70 ; Q6RH38 ; Q640Y3 ; Q49R81 ; Q4TZJ1 ; Q504A4 ; Q4JK77 ; Q4SRU4 ; Q4S966
Prokaryotic Member	iProClass: Q4L7K1
Model Organism	Homo sapiens: P14061 ; Q9NYR8 Mus musculus: Q790P4 ; P51656
FUNCTION AND STRUCTURE	
Ontology	<p><i>Molecular Function</i></p> <p>GO:0004303:estradiol 17-beta-dehydrogenase activity (26) [INTERPRO; evidence:IEA][SPEC; evidence:IEA][MGI (2152098); evidence:IEA][MGI (2152096); evidence:IEA][PMID:15026171; evidence:IDA]</p> <p>GO:0016491:oxidoreductase activity (26) [INTERPRO; evidence:IEA][SPKW; evidence:IEA][MGI (1354194); evidence:IEA][MGI (2152098); evidence:IEA]</p> <p>GO:0004872:receptor activity (4) [SPKW; evidence:IEA]</p> <p>GO:0050327:testosterone 17-beta-dehydrogenase activity (2) [PMID:15026171; evidence:IDA]</p> <p>GO:0030283:3(or 17)beta-hydroxysteroid dehydrogenase activity (1) [SPEC; evidence:IEA]</p> <p>GO:0003824:catalytic activity (1) [PMID:8547176; evidence:TAS]</p> <p>GO:0004745:retinol dehydrogenase activity (1) [PMID:10753906; evidence:TAS]</p> <p><i>Biological Process</i></p> <p>GO:0006703:estrogen biosynthetic process (26) [INTERPRO; evidence:IEA][MGI (2152098); evidence:IEA][PMID:15026171; evidence:IDA]</p> <p>GO:0008152:metabolic process (26) [INTERPRO; evidence:IEA][MGI (2152098); evidence:IEA]</p> <p>GO:0008610:lipid biosynthetic process (4) [SPKW; evidence:IEA][MGI (1354194); evidence:IEA]</p> <p>GO:0006694:steroid biosynthetic process (5) [SPKW; evidence:IEA][MGI (1354194); evidence:IEA][PMID:8547176; evidence:TAS][PMID:10753906; evidence:TAS]</p> <p>GO:0008210:estrogen metabolic process (1) [PMID:2584224; evidence:TAS]</p> <p>GO:0007601:visual perception (1) [PMID:10753906; evidence:TAS]</p> <p><i>Cellular Component</i></p> <p>GO:0005737:cytoplasm (26) [INTERPRO; evidence:IEA][MGI (2152098); evidence:IEA][PMID:8547176; evidence:TAS]</p> <p>GO:0005887:integral to plasma membrane (1) [PMID:10753906; evidence:TAS]</p>

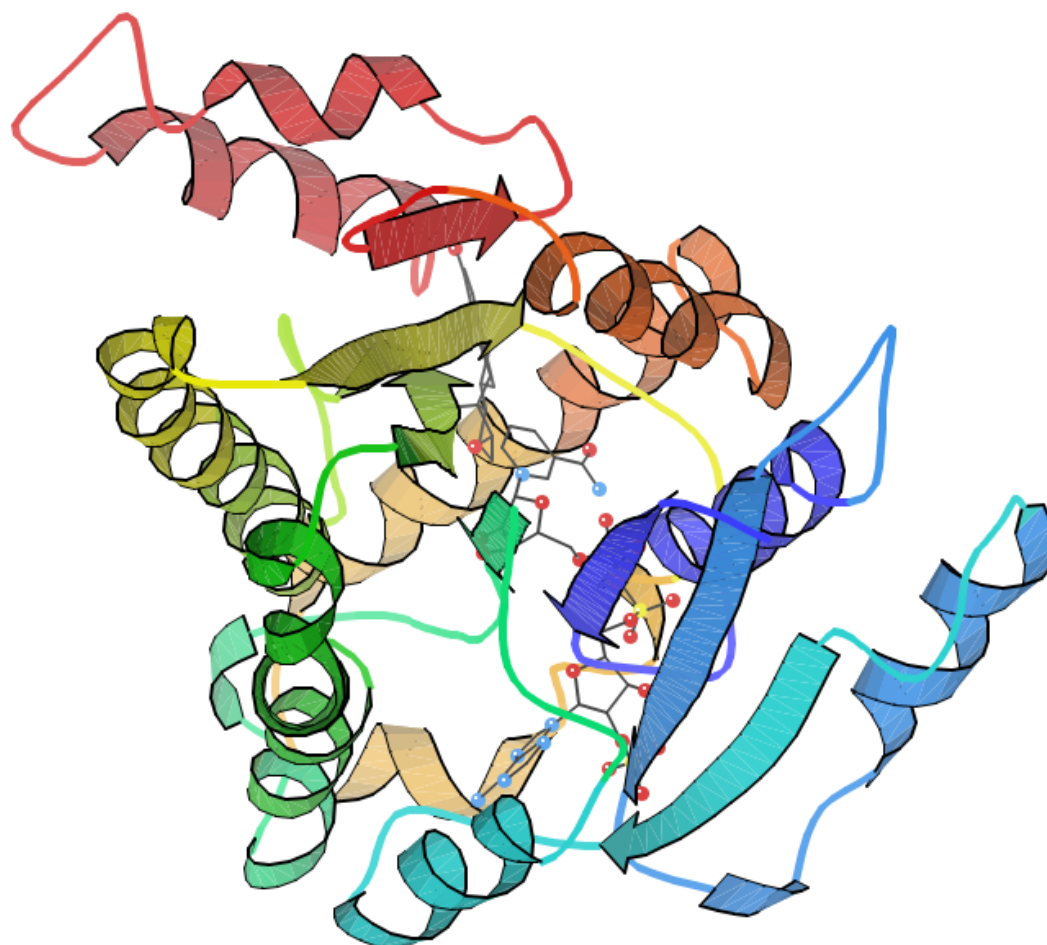
iProClass SuperFamily Summary

Enzyme/Function	<p>EC 1.1.1.62 EC-IUBMB, KEGG, BRENDA, WIT, MetaCyc <i>Nomenclature:</i> Oxidoreductases; Acting on the CH-OH group of donors; With NAD+ or NADP+ as acceptor; estradiol 17 b -dehydrogenase <i>Reaction:</i> estradiol-17 b + NAD(P)⁺ = estrone + NAD(P)H + H⁺</p> <p>EC 1.1.1.- EC-IUBMB, MetaCyc <i>Nomenclature:</i> Oxidoreductases; Acting on the CH-OH group of donors; With NAD+ or NADP+ as acceptor</p> <p>EC 1.1.1.51 EC-IUBMB, KEGG, BRENDA, WIT, MetaCyc <i>Nomenclature:</i> Oxidoreductases; Acting on the CH-OH group of donors; With NAD+ or NADP+ as acceptor; 3(or 17) b -hydroxysteroid dehydrogenase <i>Reaction:</i> testosterone + NAD(P)⁺ = androst-4-ene-3,17-dione + NAD(P)H + H⁺</p>
Pathway	<p>KEGG: Androgen and estrogen metabolism [PATH: hsa00150 mmu00150 rno00150 bta00150 gga00150 dre00150].</p>
Structure	<p>1A27: PDB SCOP CATH FSSP MMDB PDBsum 1BHS: PDB SCOP CATH FSSP MMDB PDBsum 1DHT: PDB SCOP CATH FSSP MMDB PDBsum 1EQU: PDB SCOP CATH FSSP MMDB PDBsum 1FDS: PDB SCOP CATH FSSP MMDB PDBsum 1FDT: PDB SCOP CATH FSSP MMDB PDBsum 1FDU: PDB SCOP CATH FSSP MMDB PDBsum 1FDV: PDB SCOP CATH FSSP MMDB PDBsum 1FDW: PDB SCOP CATH FSSP MMDB PDBsum 1I5R: PDB SCOP CATH FSSP MMDB PDBsum 1IOL: PDB SCOP CATH FSSP MMDB PDBsum 1JTV: PDB SCOP CATH FSSP MMDB PDBsum 1QYV: PDB SCOP CATH FSSP MMDB PDBsum 1QYW: PDB SCOP CATH FSSP MMDB PDBsum 1QYX: PDB SCOP CATH FSSP MMDB PDBsum 3DHE: PDB SCOP CATH FSSP MMDB PDBsum</p>

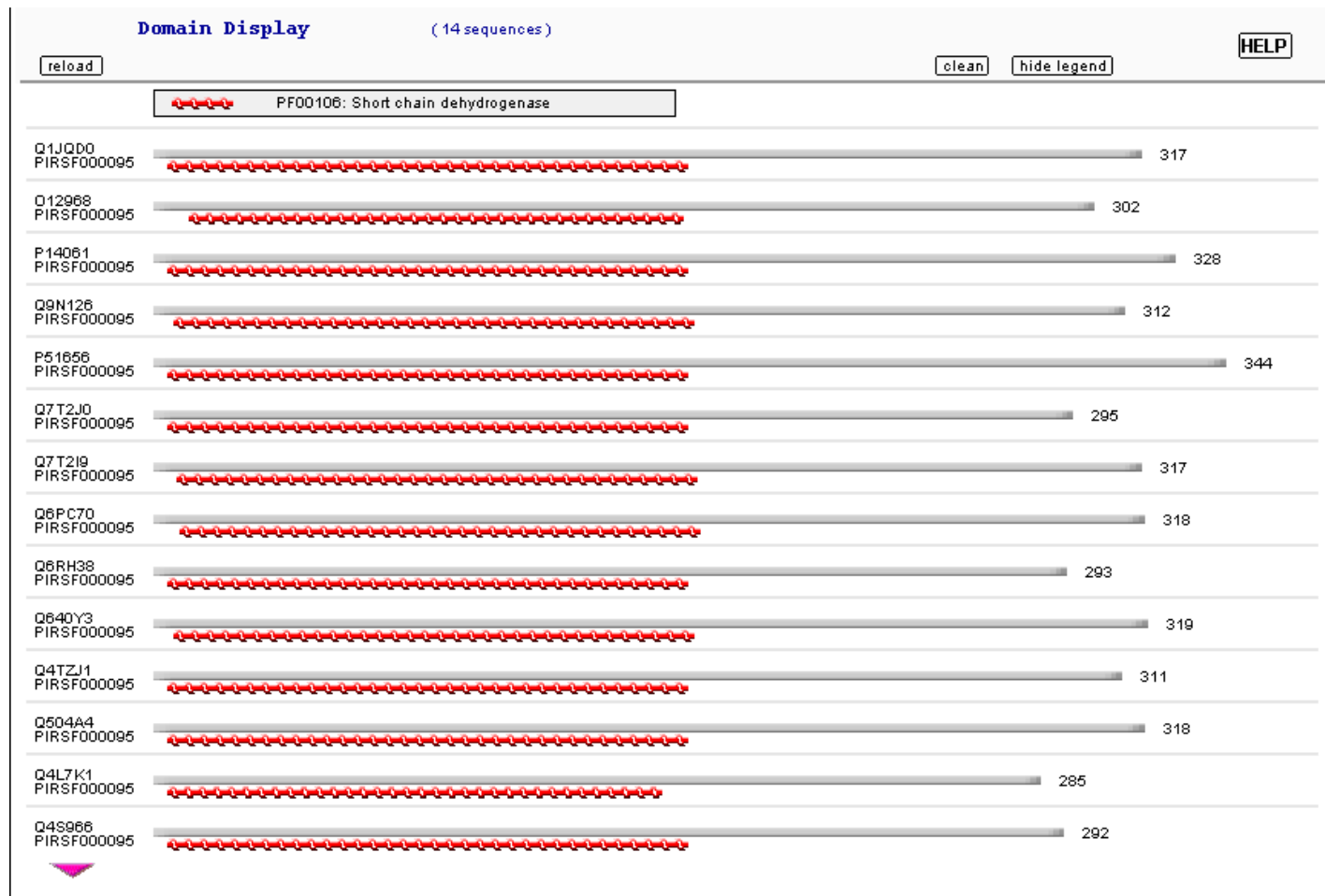
iProClass SuperFamily Summary

FAMILY RELATIONSHIP	
Pfam Domain	Pfam: PF00106 : short chain dehydrogenase(26)
Prosite Motif	Prosite: PS00061 ; PDOC00060 : Short-chain dehydrogenases/reductases family signature. (23)
InterPro	InterPro: IPR002198 : Short-chain dehydrogenase/reductase SDR InterPro: IPR011348 : 17beta-dehydrogenase InterPro: IPR002347 : Glucose/ribitol dehydrogenase
SCOP Fold	►Class: Alpha and beta proteins (a/b) ; Fold: NAD(P)-binding Rossmann-fold domains ; Superfamily: NAD(P)-binding Rossmann-fold domains ; Family: Tyrosine-dependent oxidoreductases [1A27:A; 1BHS:A; 1DHT:A; 1EQU:A; 1EQU:B; 1FDS:A; 1FDT:A; 1FDU:A; 1FDU:B; 1FDU:C; 1FDU:D; 1FDV:A; 1FDV:B; 1FDV:C; 1FDV:D; 1FDW:A; 1I5R:A; 1IOL:A; 1JTV:A; 1QYV:A; 1QYW:A; 1QYX:A; 3DHE:A]

iProClass PDB Structure 1a27



iProClass Domain Architecture



PIRSF Family Hierarchy

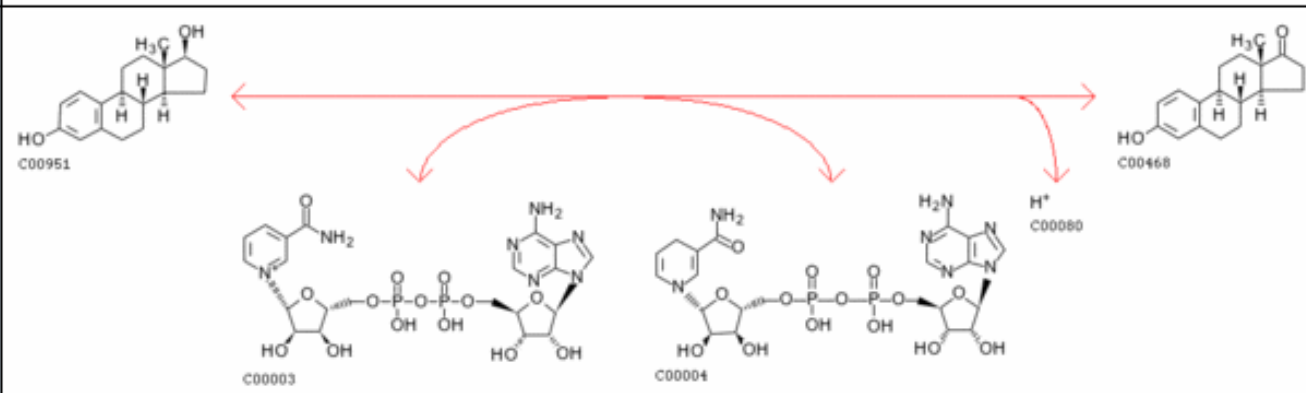
i Protein Classification	
<input type="checkbox"/>	PF00106 short chain dehydrogenase
<input checked="" type="checkbox"/>	PIRSF000094 enoyl-[acyl-carrier-protein] reductase (NADH) [Validated] (409)
<input type="checkbox"/>	PIRSF000095 estradiol 17beta-dehydrogenase [Validated] (26)
<input checked="" type="checkbox"/>	Q4L7K1 <i>Bac/Firmicute</i> :: <i>Staphylococcus haemolyticus (strain JCSG1435)</i> :: Similar to 3(or 17)beta-hydroxysteroid deh...
<input checked="" type="checkbox"/>	Q49RB1 <i>Euk/Animal</i> :: <i>Oreochromis niloticus (Nile tilapia) (Tilapia...)</i> :: 17-beta hydroxysteroid dehydrogenase type 1
<input checked="" type="checkbox"/>	Q4S966 <i>Euk/Animal</i> :: <i>Tetraodon nigroviridis (Green puffer)</i> :: Chromosome 3 SCAF14700, whole genome shotgun sequence
<input checked="" type="checkbox"/>	Q4SRU4 <i>Euk/Animal</i> :: <i>Tetraodon nigroviridis (Green puffer)</i> :: Chromosome undetermined SCAF14488, whole genome shotg...
<input checked="" type="checkbox"/>	Q504A4 <i>Euk/Animal</i> :: <i>Danio rerio (Zebrafish) (Brachydanio rerio)</i> :: Zgc:109982
<input checked="" type="checkbox"/>	Q6PC70 <i>Euk/Animal</i> :: <i>Danio rerio (Zebrafish) (Brachydanio rerio)</i> :: Retinol dehydrogenase 8
<input checked="" type="checkbox"/>	Q6RH38 <i>Euk/Animal</i> :: <i>Anquilla japonica (Japanese eel)</i> :: 17b-hydroxysteroid dehydrogenase type I
<input checked="" type="checkbox"/>	Q7LZT0 <i>Euk/Animal</i> :: <i>Anquilla japonica (Japanese eel)</i> :: 3(or 17)beta-hydroxysteroid dehydrogenase (EC 1.1.1.51) I
<input checked="" type="checkbox"/>	Q7T2I8 <i>Euk/Animal</i> :: <i>Danio rerio (Zebrafish) (Brachydanio rerio)</i> :: Photoreceptor associated retinol dehydrogenase ...
<input checked="" type="checkbox"/>	Q7T2I9 <i>Euk/Animal</i> :: <i>Danio rerio (Zebrafish) (Brachydanio rerio)</i> :: Photoreceptor associated retinol dehydrogenase ...
<input checked="" type="checkbox"/>	Q7T2J0 <i>Euk/Animal</i> :: <i>Danio rerio (Zebrafish) (Brachydanio rerio)</i> :: 17-beta hydroxysteroid dehydrogenase
<input checked="" type="checkbox"/>	Q640Y3 <i>Euk/amphibian</i> :: <i>Xenopus laevis (African clawed frog)</i> :: MGC84258 protein
<input checked="" type="checkbox"/>	O12968 <i>Euk/bird</i> :: <i>Gallus gallus (Chicken)</i> :: 17-beta-hydroxysteroid dehydrogenase
<input checked="" type="checkbox"/>	P14061 <i>Euk/mammal</i> :: <i>Homo sapiens (Human)</i> :: Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta- hydroxystero...
<input checked="" type="checkbox"/>	P51656 <i>Euk/mammal</i> :: <i>Mus musculus (Mouse)</i> :: Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta- hydroxystero...
<input checked="" type="checkbox"/>	P51657 <i>Euk/mammal</i> :: <i>Rattus norvegicus (Rat)</i> :: Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta- hydroxyst...
<input checked="" type="checkbox"/>	Q1JQD0 <i>Euk/mammal</i> :: <i>Bos taurus (Bovine)</i> :: Hydroxysteroid (17-beta) dehydrogenase 1
<input checked="" type="checkbox"/>	Q1WNP0 <i>Euk/mammal</i> :: <i>Pan troglodytes (Chimpanzee)</i> :: Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta- hydr...
<input checked="" type="checkbox"/>	Q1WNP1 <i>Euk/mammal</i> :: <i>Hylobates klossii (Kloss's gibbon)</i> :: 17beta-hydroxysteroid dehydrogenase type 1
<input checked="" type="checkbox"/>	Q1WNP2 <i>Euk/mammal</i> :: <i>Pongo pygmaeus (Oranqutan)</i> :: 17beta-hydroxysteroid dehydrogenase type 1
<input checked="" type="checkbox"/>	Q1WNP3 <i>Euk/mammal</i> :: <i>Macaca mulatta (Rhesus macaque)</i> :: 17beta-hydroxysteroid dehydrogenase type 1
<input checked="" type="checkbox"/>	Q4JK77 <i>Euk/mammal</i> :: <i>Macaca fascicularis (Crab eating macaque) (Cv...)</i> :: 17-beta hydroxysteroid dehydrogenase 1
<input checked="" type="checkbox"/>	Q4TZJ1 <i>Euk/mammal</i> :: <i>Tupaia glis belangeri (Common tree shrew)</i> :: 17-beta-hydroxysteroid dehydrogenase type 1
<input checked="" type="checkbox"/>	Q790P4 <i>Euk/mammal</i> :: <i>Mus musculus (Mouse)</i> :: 17-beta-hydroxysteroid dehydrogenase (Hydroxysteroid (17-beta) dehydro...
<input checked="" type="checkbox"/>	Q9N126 <i>Euk/mammal</i> :: <i>Bos taurus (Bovine)</i> :: Retinol dehydrogenase 8 (EC 1.1.1.-) (Photoreceptor outer segment all- ...
<input checked="" type="checkbox"/>	Q9NYR8 <i>Euk/mammal</i> :: <i>Homo sapiens (Human)</i> :: Retinol dehydrogenase 8 (EC 1.1.1.-) (Photoreceptor outer segment all-...
<input checked="" type="checkbox"/>	PIRSF000096 peroxisomal trifunctional enzyme HDE [Validated] (21)
<input checked="" type="checkbox"/>	PIRSF000126 11beta-hydroxysteroid dehydrogenase [Validated] (222)
<input checked="" type="checkbox"/>	PIRSF006067 retinol dehydrogenase (88)
<input checked="" type="checkbox"/>	PIRSF026396 short chain dehydrogenase, Alr5284 type [Predicted] (33)
<input checked="" type="checkbox"/>	SF027687 ()
<input checked="" type="checkbox"/>	PIRSF036586 bifunctional ribulose 5-phosphate reductase/CDP-ribitol pyrophosphorylase [Validated] (7)
<input checked="" type="checkbox"/>	PIRSF036951 agropine synthesis reductase [Validated] (7)
<input checked="" type="checkbox"/>	PIRSF037747 cytochrome P450 with short chain dehydrogenase domain [Predicted] (5)

iProClass Taxonomy Nodes

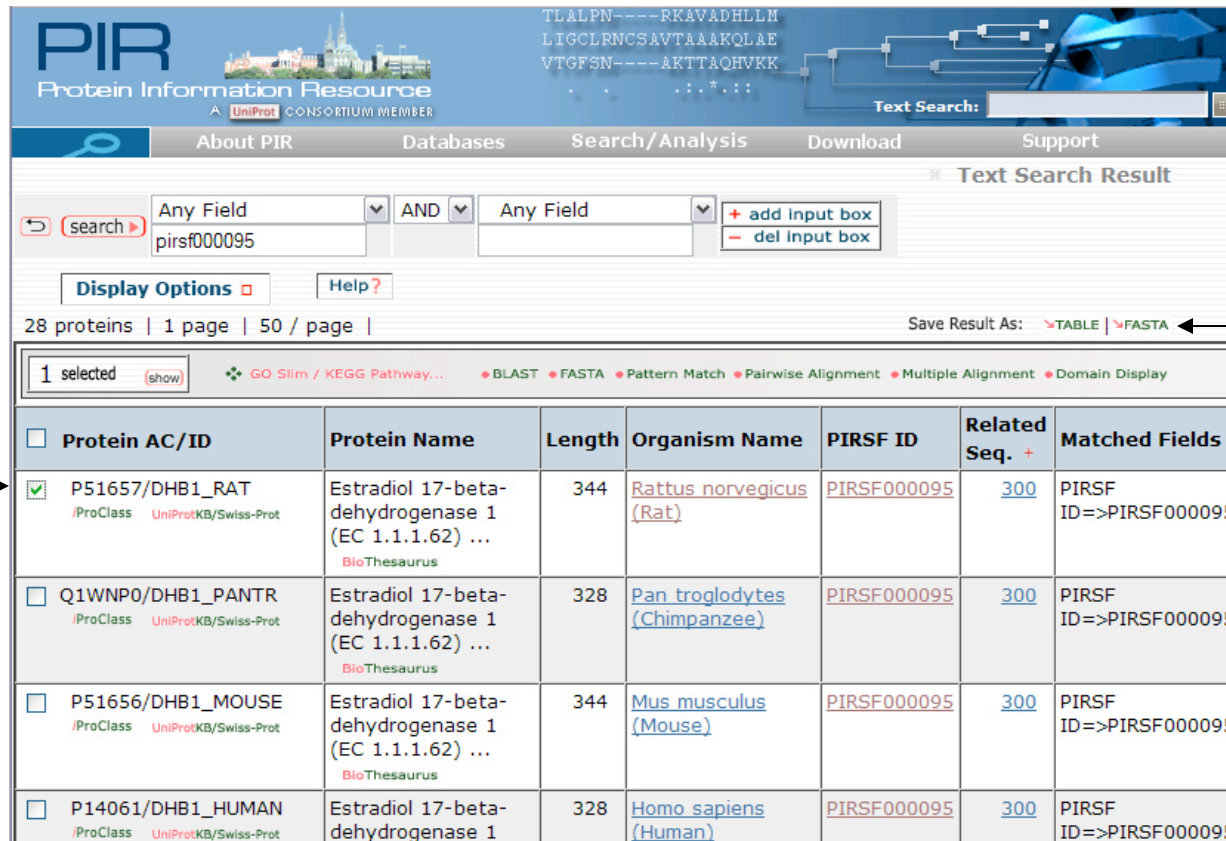
▼ Eukaryota	25
▼ Funqi/Metazoa group	25
▼ Metazoa	25
▼ Eumetazoa	25
▼ Bilateria	25
▼ Coelomata	25
▼ Deuterostomia	25
▼ Chordata	25
▼ Craniata	25
▼ Vertebrata	25
▼ Gnathostomata	25
▼ Teleostomi	25
▼ Euteleostomi	25
▼ Actinopterygii	10
▼ Actinopteri	10
▼ Neopterygii	10
▼ Teleostei	10
▼ Elopocephala	10
▼ Clupeocephala	8
▶ Euteleostei	3
▶ Otocephala	5
▶ Elopomorpha	2
▼ Sarcopterygii	15
▼ Tetrapoda	15
▼ Amniota	14
▼ Mammalia	13
▼ Theria	13
▼ Eutheria	13
▼ Euarchontoglires	11
▶ Glires	3
▶ Primates	7
▶ Scandentia	1
▶ Laurasiatheria	2
▶ Sauropsida	1
▶ Amphibia	1

These materials were developed with funding from the US National Institutes of Health grant #2T36 GM008789 to the Pittsburgh Supercomputing Center

iProClass Enzyme Function: KEGG

Entry	R02352	Reaction
Name	Estradiol-17beta:NAD+ 17-oxidoreductase	
Definition	Estradiol-17beta + NAD+ <=> Estrone + NADH + H+	
Equation	c00951 + c00003 <=> c00468 + c00004 + c00080	
		
RPair	RP: A00002 c00003_c00004 cofac RP: A00350 c00468_c00951 main	
Pathway	PATH: rn00150 Androgen and estrogen metabolism	
Enzyme	1.1.1.51 1.1.1.62	
Orthology	KO: K00044 estradiol 17beta-dehydrogenase KO: K05296 3(or 17)beta-hydroxysteroid dehydrogenase	
LinkDB	All DBs	

iProClass: Saving Sequences



PIR Protein Information Resource
A UniProt Consortium Member

Text Search:

Any Field AND Any Field

Display Options

28 proteins | 1 page | 50 / page | Save Result As:

1 selected

<input type="checkbox"/>	Protein AC/ID	Protein Name	Length	Organism Name	PIRSF ID	Related Seq. +	Matched Fields
<input checked="" type="checkbox"/>	P51657/DHB1_RAT <small>/ProClass UniProtKB/Swiss-Prot</small>	Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) ... <small>BioThesaurus</small>	344	Rattus norvegicus (Rat)	PIRSF000095	300	PIRSF ID=>PIRSF000095
<input type="checkbox"/>	Q1WNP0/DHB1_PANTR <small>/ProClass UniProtKB/Swiss-Prot</small>	Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) ... <small>BioThesaurus</small>	328	Pan troglodytes (Chimpanzee)	PIRSF000095	300	PIRSF ID=>PIRSF000095
<input type="checkbox"/>	P51656/DHB1_MOUSE <small>/ProClass UniProtKB/Swiss-Prot</small>	Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) ... <small>BioThesaurus</small>	344	Mus musculus (Mouse)	PIRSF000095	300	PIRSF ID=>PIRSF000095
<input type="checkbox"/>	P14061/DHB1_HUMAN <small>/ProClass UniProtKB/Swiss-Prot</small>	Estradiol 17-beta-dehydrogenase 1	328	Homo sapiens (Human)	PIRSF000095	300	PIRSF ID=>PIRSF000095

Check Entries →

Save Format

InterPro

- Integrated resource of protein families, domains, repeats and sites from member databases (PROSITE, Pfam, Prints, ProDom, SMART and TIGRFAMs).
- Member databases represent features in different ways: Some use hidden Markov models, some use position specific scoring matrices, some use ambiguous consensus patterns.
- Easy way to search several libraries at once with a query.

InterPro – Searching with InterProScan

Please Note: Due to resource limitations the InterProScan service will not accept nucleotide sequence submissions until further notice. Please see the [Help](#) for more information.

 [Download Software](#)

RESULTS YOUR EMAIL

interactive

APPLICATIONS TO RUN Clear all Check all

<input checked="" type="checkbox"/> BlastProDom	<input checked="" type="checkbox"/> FPrintScan	<input checked="" type="checkbox"/> HMMPIR	<input checked="" type="checkbox"/> HMMPFam	<input checked="" type="checkbox"/> HMMSmart
<input checked="" type="checkbox"/> HMMTigr	<input checked="" type="checkbox"/> ProfileScan	<input checked="" type="checkbox"/> ScanRegExp	<input checked="" type="checkbox"/> SuperFamily	<input checked="" type="checkbox"/> SignalPHMM
<input checked="" type="checkbox"/> TMHMM	<input checked="" type="checkbox"/> HMMPanther	<input checked="" type="checkbox"/> Gene3D		

Enter or Paste a PROTEIN Sequence in any format: [Help](#)

```
>DHB1_RAT P51657 Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62)
(17-beta- hydroxysteroid dehydrogenase type 1) (17-beta-HSD 1)
[Rattus norvegicus]
MDSTVVLITGCSGGIGLHLAVRLASDRSQSFKVYATLRLDKSQGPLLEAARAQGCPPGSL
EILELDVRDSESVAAARACVTEGRVDVLVCNAGRGLFGPLEAHELNAVGAVLVDVNVLGTI
RMLQAFLLPDMKRHSGRVLVTASVGGIMGLPFHEVYCASKFALEGLCESLAILLPLFGVH
VSLIECGAVHTAFHEKLEGGPGGALERADAQTRHLFAHYQRCYEQALSEAQDPPEEVELEF
LTAMRAPQPALRYFSTNRFPLARMRTEDPSGSSYVEAMHREAFSDLQVQEGAKAGAQVS
GDPDTPPRALICLPECAIPRVTAELGWSASDKPGQNKSCYQQKI
```

Upload a file:





PLEASE NOTE: Interactive job results are stored for 24 hours, email job results are stored for one week. If you plan to use these services during a course please contact us using the email below.

InterPro - InterProScan Results

InterProScan Results

[Table View](#)
[Raw Output](#)
[XML Output](#)
[Original Sequences](#)
[SUBMIT ANOTHER JOB](#)

SEQUENCE: DHB1 RAT CRC64: CB5FC139FBB3995E LENGTH: 344 aa

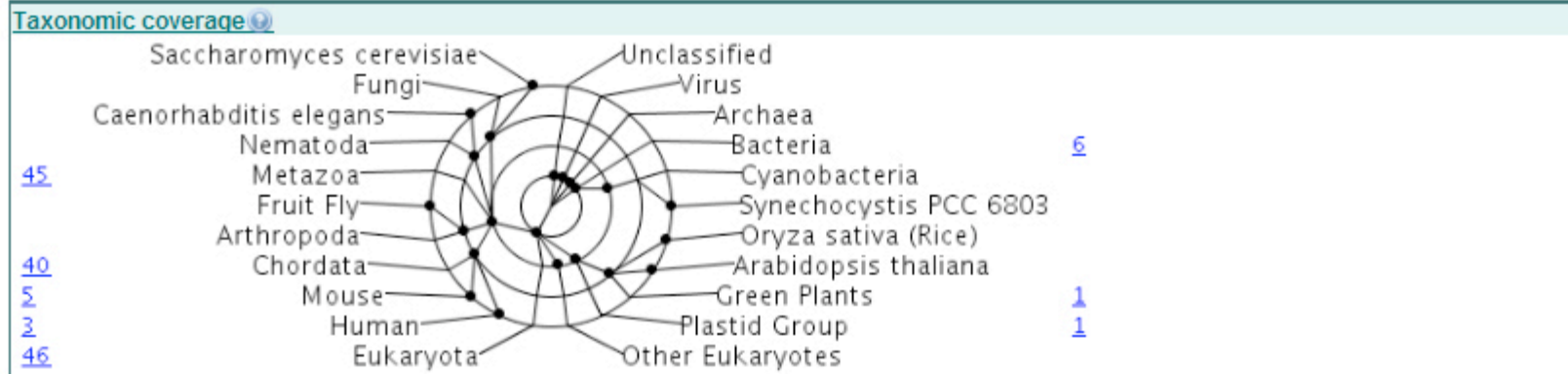
InterPro IPR002198 Family InterPro 	Short-chain dehydrogenase/reductase SDR	
	PR00080	SDRFAMILY
	PTHR19410	ADH_short_C2
	PF00108	adh_short
	PS00061	ADH_SHORT
InterPro IPR002347 Family InterPro 	Glucose/ribitol dehydrogenase	
	PR00081	GDHRDH
InterPro IPR011348 Family InterPro 	17beta-dehydrogenase	
	PIRSF000095	17beta-HSD
	PTHR19410:SF47	17beta_DH
InterPro IPR018040 Domain InterPro 	NAD(P)-binding	
	G3DSA:3.40.50.720	NAD(P)-bd
	SSF51735	NAD(P)-bd

[Table View](#)
[Raw Output](#)
[XML Output](#)
[Original Sequences](#)
[SUBMIT ANOTHER JOB](#)

InterPro - InterProScan Results

InterPro: IPR011348 17beta-dehydrogenase													
Protein matches													
UniProtKB Matches: 52 proteins	Overview: sorted by AC , sorted by name , of known structure , proteins with splice variants												
	Detailed: sorted by AC , sorted by name , of known structure , proteins with splice variants												
	Table: For all matching proteins , of known structure												
	Architectures Accession List												
Accession	IPR011348 17beta_DHase												
Type	Family												
Signatures	<table border="1"> <thead> <tr> <th>Database</th> <th>ID</th> <th>Name</th> <th>Proteins</th> </tr> </thead> <tbody> <tr> <td>PIRSF</td> <td>PIRSF000095</td> <td>17beta-HSD</td> <td>30</td> </tr> <tr> <td>PANTHER</td> <td>PTHR19410:SF47</td> <td>17beta_DH</td> <td>52</td> </tr> </tbody> </table>	Database	ID	Name	Proteins	PIRSF	PIRSF000095	17beta-HSD	30	PANTHER	PTHR19410:SF47	17beta_DH	52
Database	ID	Name	Proteins										
PIRSF	PIRSF000095	17beta-HSD	30										
PANTHER	PTHR19410:SF47	17beta_DH	52										
InterPro Relationships													
Parent	IPR002347 Glucose/ribitol dehydrogenase												
Contains	IPR016040 NAD(P)-binding												
GO Term annotation													
Process	GO:0006703 estrogen biosynthetic process												
Function	GO:0004303 estradiol 17-beta-dehydrogenase activity												
Component	GO:0005737 cytoplasm												
InterPro annotation													
Abstract	<p>This entry represents 17beta-hydroxysteroid dehydrogenases (17B-HSDs), a group of enzymes which catalyse the last step in the biosynthesis of all androgens and estrogens -the reversible NAD(P)-linked transfer of a hydride to and from the 17-position of steroid molecules [1]. A total of six isozymes have been identified which vary in substrate specificity, tissue specificity and preferred direction of the reaction.</p> <p>The most intensively studied enzyme in this entry is human estrogenic 17beta-hydroxysteroid dehydrogenase (P14061) which is responsible for the last step in the synthesis of all estrogens. As active estrogens stimulate the proliferation of breast cancer cells, this enzyme is a potential target for drugs to treat breast cancer [2]. It is a membrane-associated homodimer which possesses the Tyr-X-X-X-Lys motif typical of short-chain dehydrogenases and forms a typical Rossmann fold [3].</p>												
Structural links	CATH: 3.40.50.720.114 SCOP: c2.1.2 PDB - click here												
Database links	Enzyme: EC:1.1.1												

InterPro - InterProScan Results




Overlapping InterPro entries

InterPro ID	Numbers of overlapping proteins	Average numbers of overlapping amino acids
IPR011348		
IPR002198	0 (red) 52 (purple) 27078 (blue)	N/A
% Overlap: 100		
IPR016040	0 (red) 52 (purple) 117055 (blue)	N/A
% Overlap: 100		
IPR002347	4 (red) 48 (purple) 23921 (blue)	N/A
% Overlap: 92		


InterPro - InterProScan Results

[Example proteins](#)


P14051 Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta-hydroxysteroid dehydrogenase type 1) (17-beta-HSD 1) (Placental 17-beta-hydroxysteroid dehydrogenase) (20 alpha-hydroxysteroid dehydrogenase) (20-alpha-HSD) (E2DH)



PS1656 Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta-hydroxysteroid dehydrogenase type 1) (17-beta-HSD 1)












PS1657 Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta-hydroxysteroid dehydrogenase type 1) (17-beta-HSD 1)



[More proteins](#)

Example Proteins Key

InterPro entry accession number/name and structure databases	Colour code
IPR011348 17beta-dehydrogenase	
IPR016040 NAD(P)-binding	
IPR002198 Short-chain dehydrogenase/reductase SDR	
IPR002347 Glucose/ribitol dehydrogenase	
ModBase	
SWISS-MODEL	
CATH Domain	
SCOP Domain	
PDB Chain	

InterPro - InterProScan Results

Publications

1. Peltoketo H. , Isomaa V. , Poutanen M. , Vihko R.
Expression and regulation of 17 beta-hydroxysteroid dehydrogenase type 1.
J. Endocrinol. 150 S21-S30 1996 [[PubMed: 8943783](#)]
2. Sawicki M.W. , Erman M. , Puranen T. , Vihko P. , Ghosh D.
Structure of the ternary complex of human 17beta-hydroxysteroid dehydrogenase type 1 with 3-hydroxyestra-1,3,5,7-tetraen-17-one (equilin) and NADP+.
Proc. Natl. Acad. Sci. U.S.A. 96 840-845 1999 [[PubMed: 9927655](#)]
3. Ghosh D. , Pletnev V.Z. , Zhu D.W. , Wawrzak Z. , Duax W.L. , Pangborn W. , Labrie F. , Lin S.X.
Structure of human estrogenic 17 beta-hydroxysteroid dehydrogenase at 2.20 Å resolution.
Structure 3 503-513 1995 [[PubMed: 7663947](#)]

Additional Reading

- Han Q. , Campbell R.L. , Gangloff A. , Huang Y.W. , Lin S.X.
Dehydroepiandrosterone and dihydrotestosterone recognition by human estrogenic 17beta-hydroxysteroid dehydrogenase. C-18/c-19 steroid discrimination and enzyme-induced strain.
J. Biol. Chem. 275 2000 1105-1111 [[PubMed: 10625652](#)]
- Shi R. , Lin S.X.
Cofactor hydrogen bonding onto the protein main chain is conserved in the short chain dehydrogenase/reductase family and contributes to nicotinamide orientation.
J. Biol. Chem. 279 2004 16778-16785 [[PubMed: 14966133](#)]
- Qiu W. , Campbell R.L. , Gangloff A. , Dupuis P. , Boivin R.P. , Tremblay M.R. , Poirier D. , Lin S.X.
A concerted, rational design of type 1 17beta-hydroxysteroid dehydrogenase inhibitors: estradiol-adenosine hybrids with high affinity.
FASEB J 16 2002 1829-1831 [[PubMed: 12223444](#)]
- Gangloff A. , Shi R. , Nahoum V. , Lin S.X.
Pseudo-symmetry of C19 steroids, alternative binding orientations, and multispecificity in human estrogenic 17beta-hydroxysteroid dehydrogenase.
FASEB J 17 2003 274-276 [[PubMed: 12490543](#)]

InterPro {cache:version}

A Vision: Computer Assisted Bioinformatics

- Goal
 - The computer assists the scientist in the collection of all bioinformatics information relevant to the hypothesis at hand
- A single software application that can:
 - Understand multiple data formats specifically devised to represent structure, function, metabolism, evolution, etc.
 - Assist scientists in creating and maintaining relationships among different types of information collected from multiple sources
 - Support simultaneous searches across multiple data sources of a similar nature (e.g. multiple sequence databases)

Remains an Open Research Problem