- 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 Gonzalez-Mendez, University of Puerto Rico Medical Sciences Campus.
- Dr. Alade Tokuta, North Carolina Central University.
- Dr. Jaime Seguel and Dr. Bienvenido Velez, University of Puerto Rico at Mayaguez.
- Dr. Satish Bhalla, Johnson C. Smith University.

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

- This material is targeted towards students with a general background in Biology. It was developed to introduce biology students to the computational mathematical and biological issues surrounding bioinformatics. This specific lesson deals with the following fundamental topics:
  - Computing for biologists
  - Computer Science track

  - This material has been developed by:

  Dr. Hugh B. Nicholas, Jr.

  National Center for Biomedical Supercomputing

  Pittsburgh Supercomputing Center

  Carnegie Mellon University

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

Bioinformatics Data Management

Lecture 1

Course Overview

The Need for Biological Information

Bienvenido Vélez

UPR Mayaguez

Reference: BioInformatics for Dummies

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

- Categories of Information Needs and Their Supporting Databases
  - Reference vs. Discovery Needs
  - General versus Domain Specific Databases
- Overview of Current Biological Databases
- The Future of Biological Databases and Tools:
  - Integration of Biological Information
  - Computer Assisted Bioinformatics (CAB)

Reference and discovery are two fundamentally different information needs

- Reference:
  - find something that I have seen before
  - Example:
    - find out who discovered a DNA sequence or protein
    - Find some characteristic of a known sequence or protein
- Discovery:
  - find something new. Infer new knowledge.
  - Examples:
    - Find new sequences that evolved from known common ancestor
    - Find sequences that may have similar function in other organisms

No single information system can support both information needs effectively

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

Finding Reference Information

- Reference information searches can be accomplished:
  - By key
    - Find a DNA sequence by its accession number
  - By attribute (exact)
    - Find sequences belonging to C. Elegans
  - By attribute (inexact)
    - Find proteins related to some type of cancer

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

Discovering Information

- By Association (similarity) vs. by Fr..?? ss by structure
- Discovery searches can be accomplished:
  - By similarity of:
    - Structure
    - Function
    - Combination of the above

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

General Databases

- Contain information on virtually any subject
- Information exists in large variety of formats and styles:
  - Images, web pages, emails, PDF’s, blog entries, forum entries, WIKI pages, etc
- Provide a generic query model often based on term occurrence
  - Find me everything that contains the terms “aldehyde dehydrogenase”
- Pros: One stop shopping for information
- Cons: Hard to exploit the nature of information in order to speed up the search. May yield lots of irrelevant information

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

Domain-specific Databases

- Contain information specific to a relatively small knowledge domain (e.g. DNA sequences)
- Information appears in somewhat homogeneous form
- Provide a specific query model that can exploit the particularities of the information
- Pros: Specific questions can be answered quickly
- Cons: User must often integrate results from multiple specific databases in order to answer a more general question

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

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

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

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

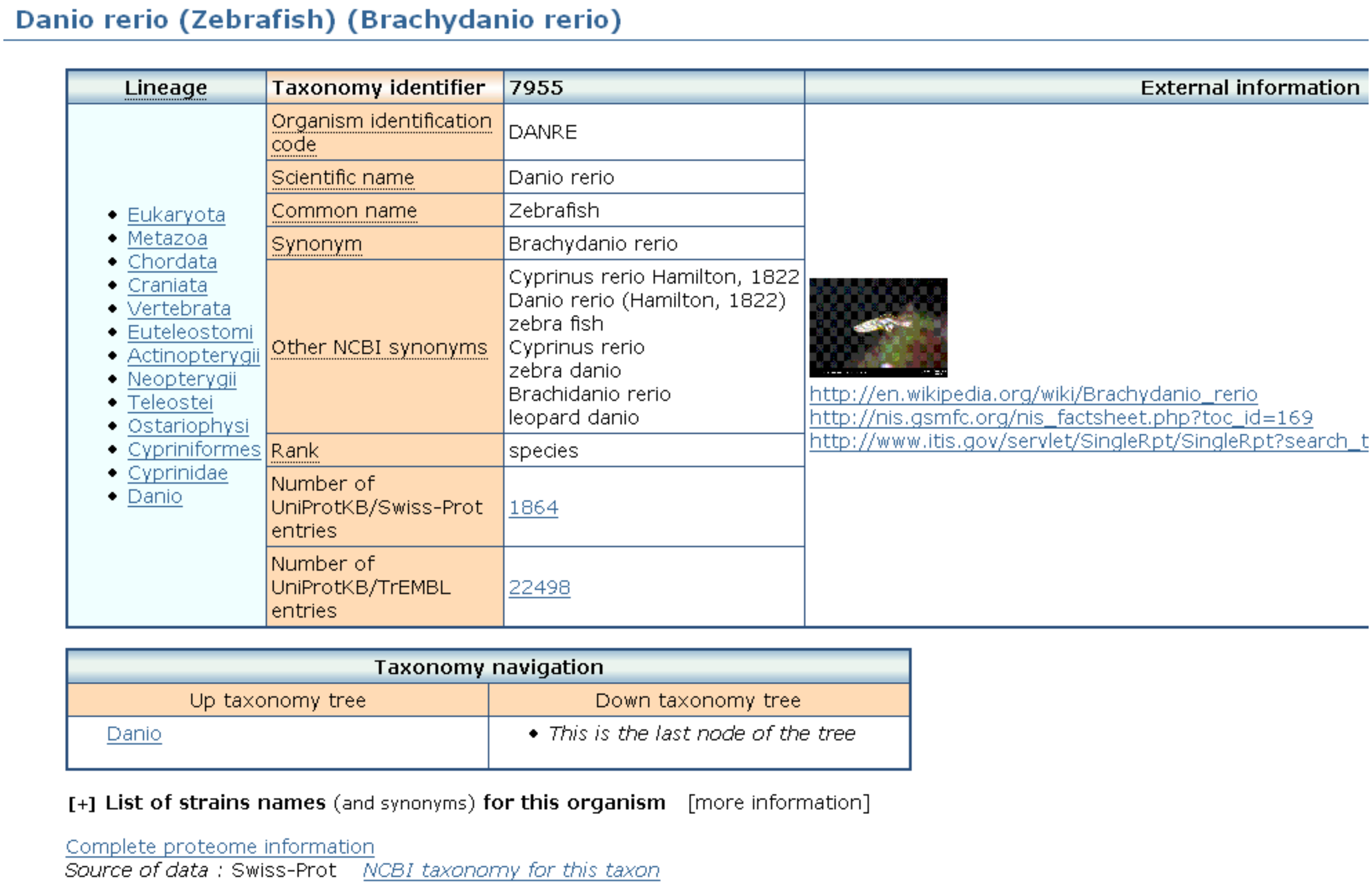
Species Libraries

- 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
• Eukaryota • Metazoa • Chordata • Craniata • Vertebrata • Euteleostomi • Actinopterygii • Neopterygii • Teleostei • Ostariophysi • Cypriniformes • Cyprinidae • Danio	Organism identification code	DANRE	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
	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	

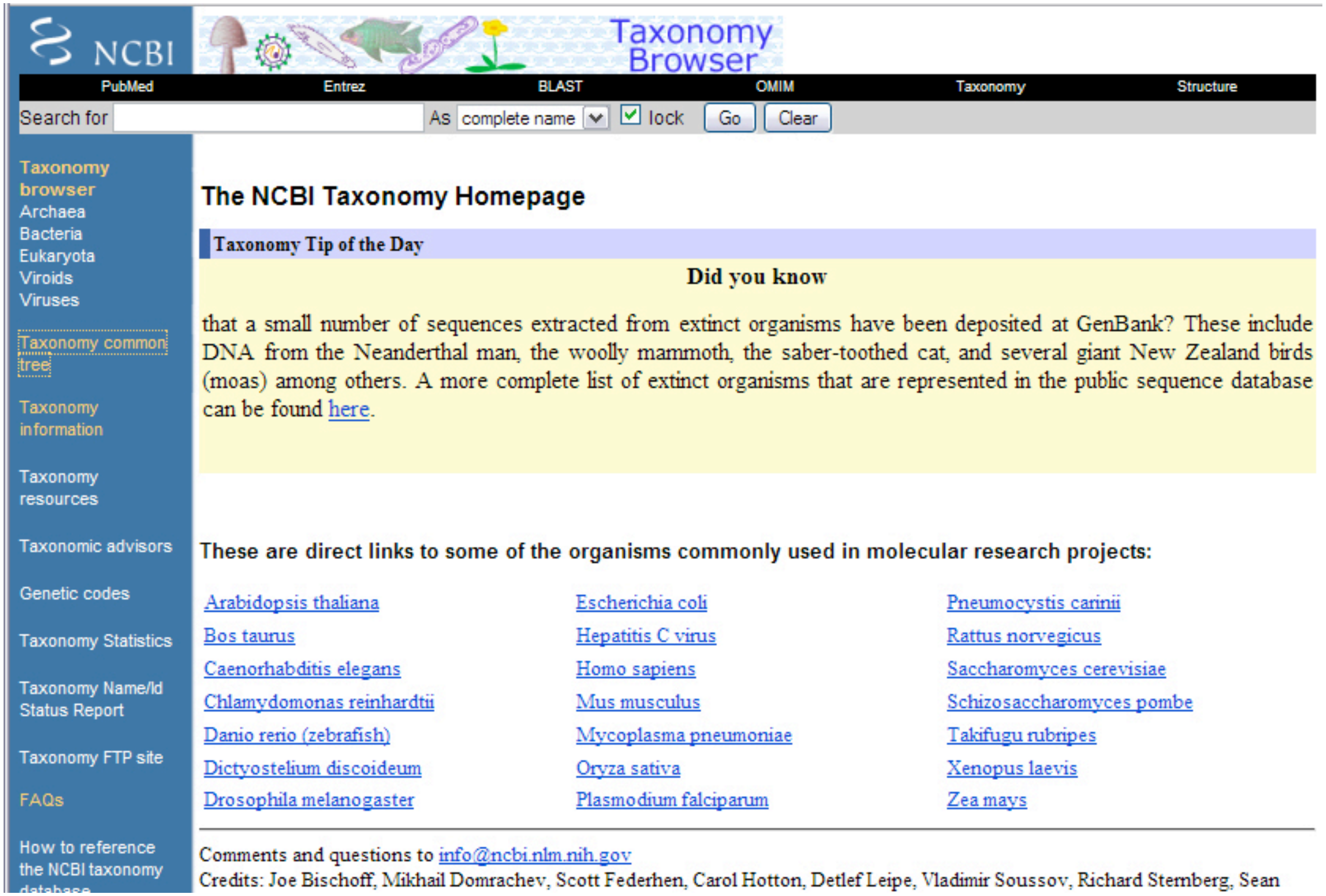
Taxonomy navigation	
Up taxonomy tree	Down taxonomy tree
Danio	• *This is the last node of the tree*

[+] 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 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
Bos taurus
Caenorhabditis elegans
Chlamydomonas reinhardtii
Danio rerio (zebrafish)
Dictyostelium discoideum
Drosophila melanogaster
Escherichia coli
Hepatitis C virus
Homo sapiens
Mus musculus
Mycoplasma pneumoniae
Oryza sativa
Plasmodium falciparum
Pneumocystis carinii
Rattus norvegicus
Saccharomyces cerevisiae
Schizosaccharomyces pombe
Takifugu rubripes
Xenopus laevis
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 Sternberg, 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.

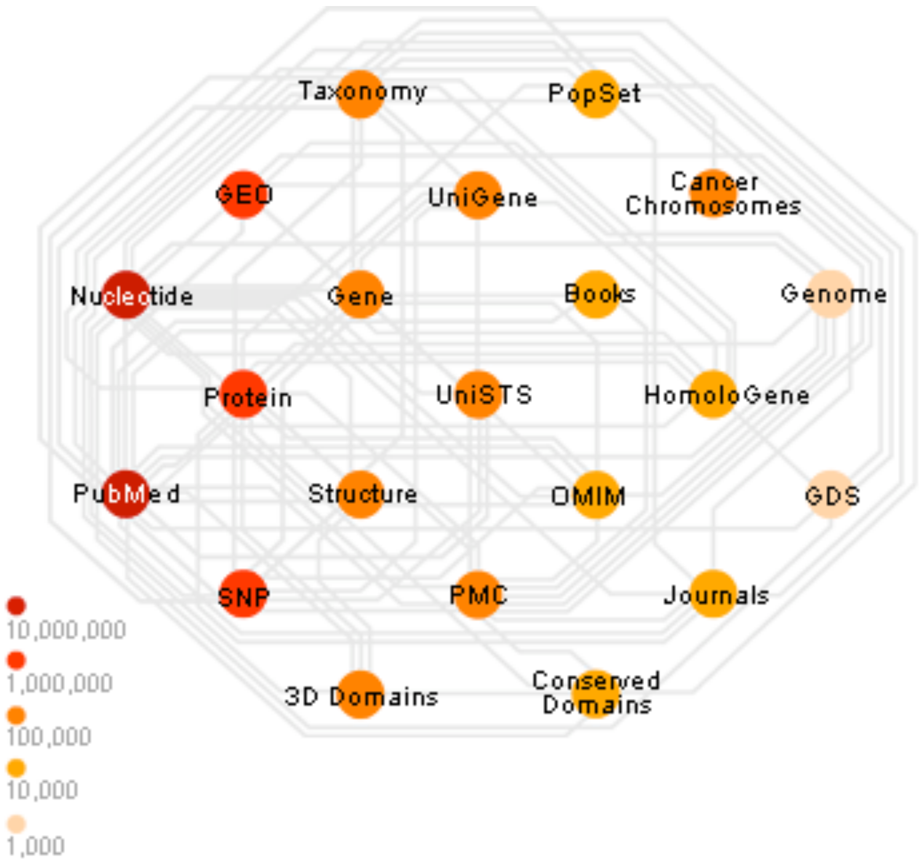
What has been done about this?

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

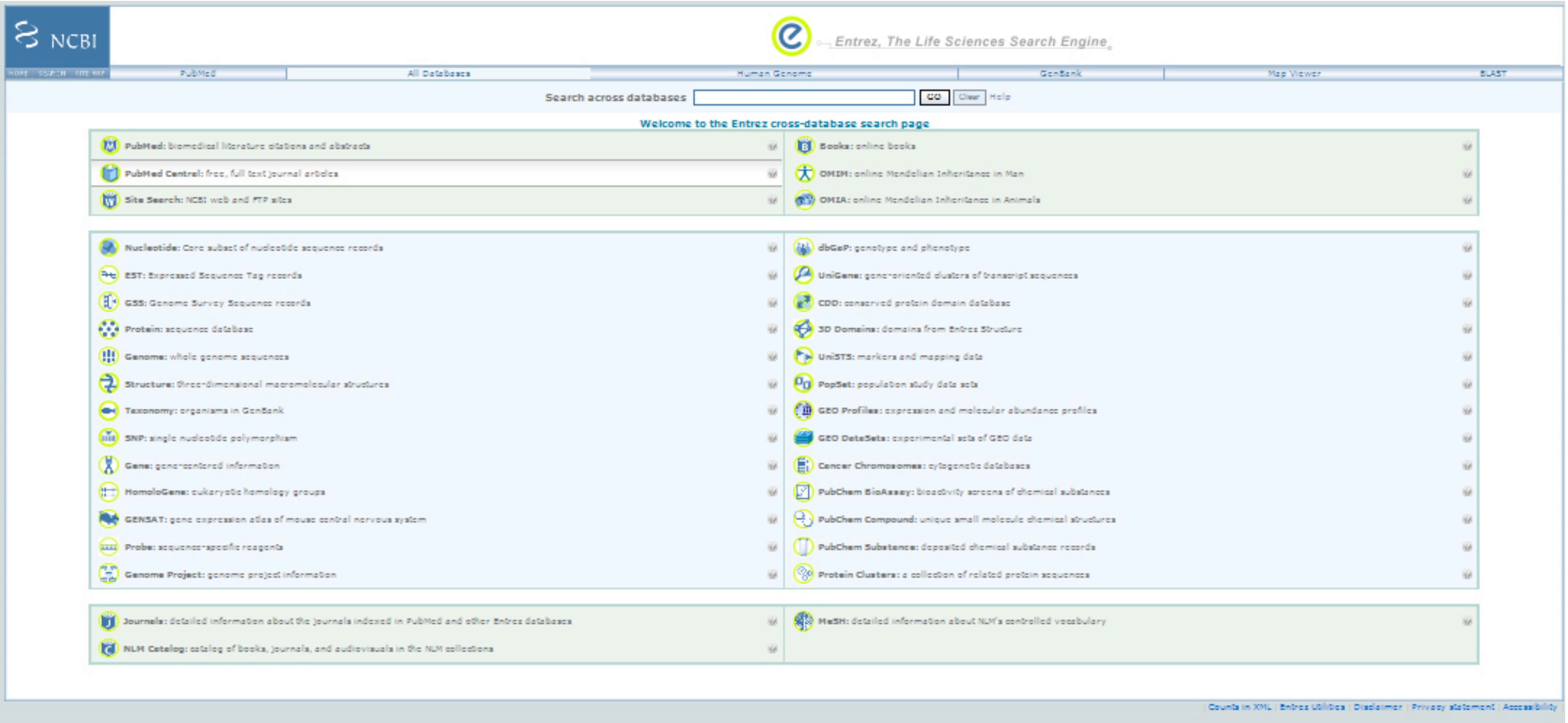
Integrated Information Resources

- 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



Welcome to the Entrez cross-database search page

PubMed: biomedical literature citations and abstracts

PubMed Central: free, full text journal articles

Site Search: NCBI web and FTP sites

Books: online books

OMIM: online Mendelian Inheritance in Man

OMIA: online Mendelian Inheritance in Animals

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

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 screens of chemical substances

PubChem Compound: unique small molecule chemical structures

PubChem Substance: deposited chemical substance records

Protein Clusters: a collection of related protein sequences

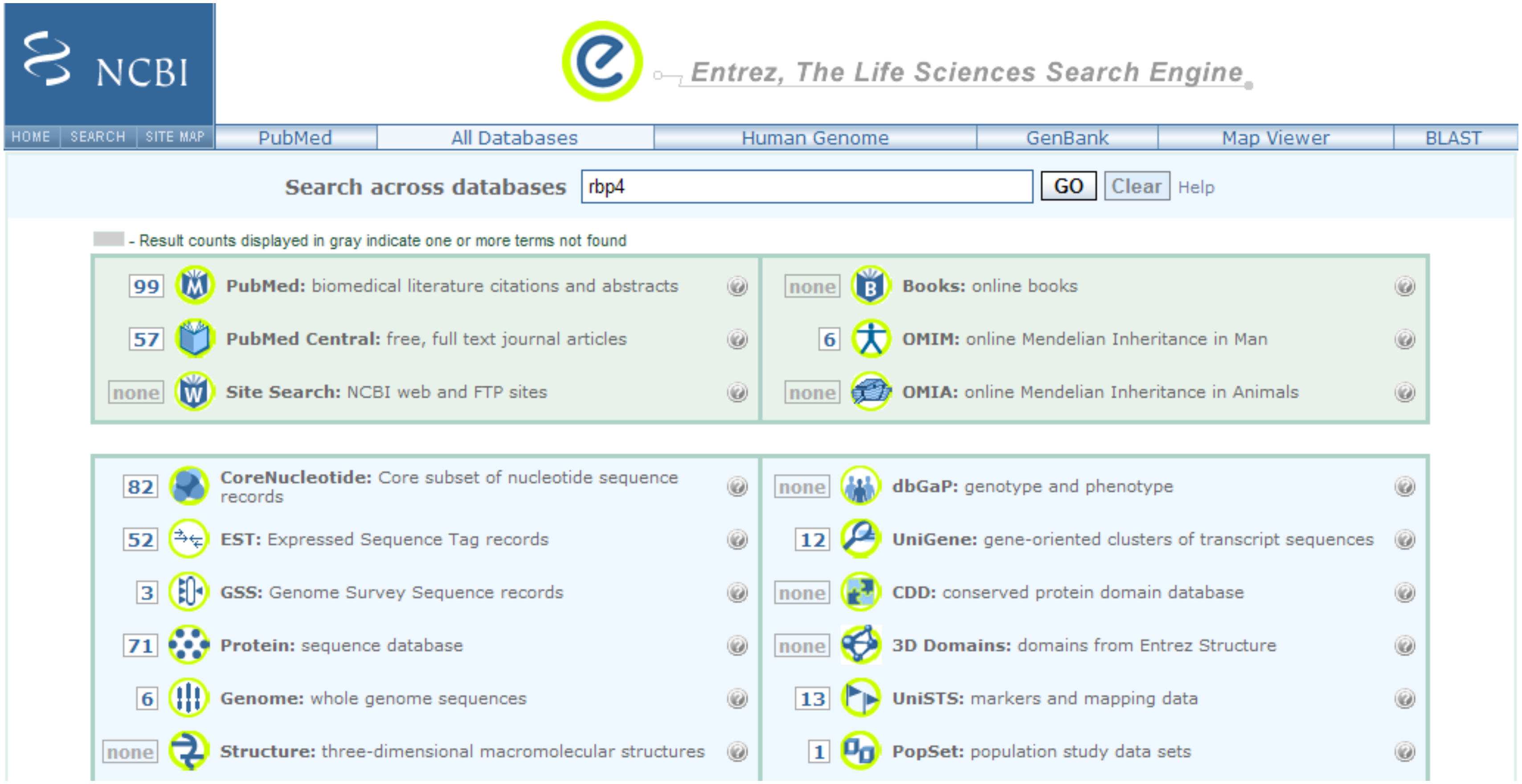
Journals: detailed information about the journals indexed in PubMed and other Entrez databases

NLM Catalog: catalog of books, journals, and audiovisuals in the NLM collections

MeSH: detailed information about NLM's controlled vocabulary

| Counts in XML | Entrez Utilities | Disclaimer | Privacy statement | Accessibility

NCBI Entrez Results



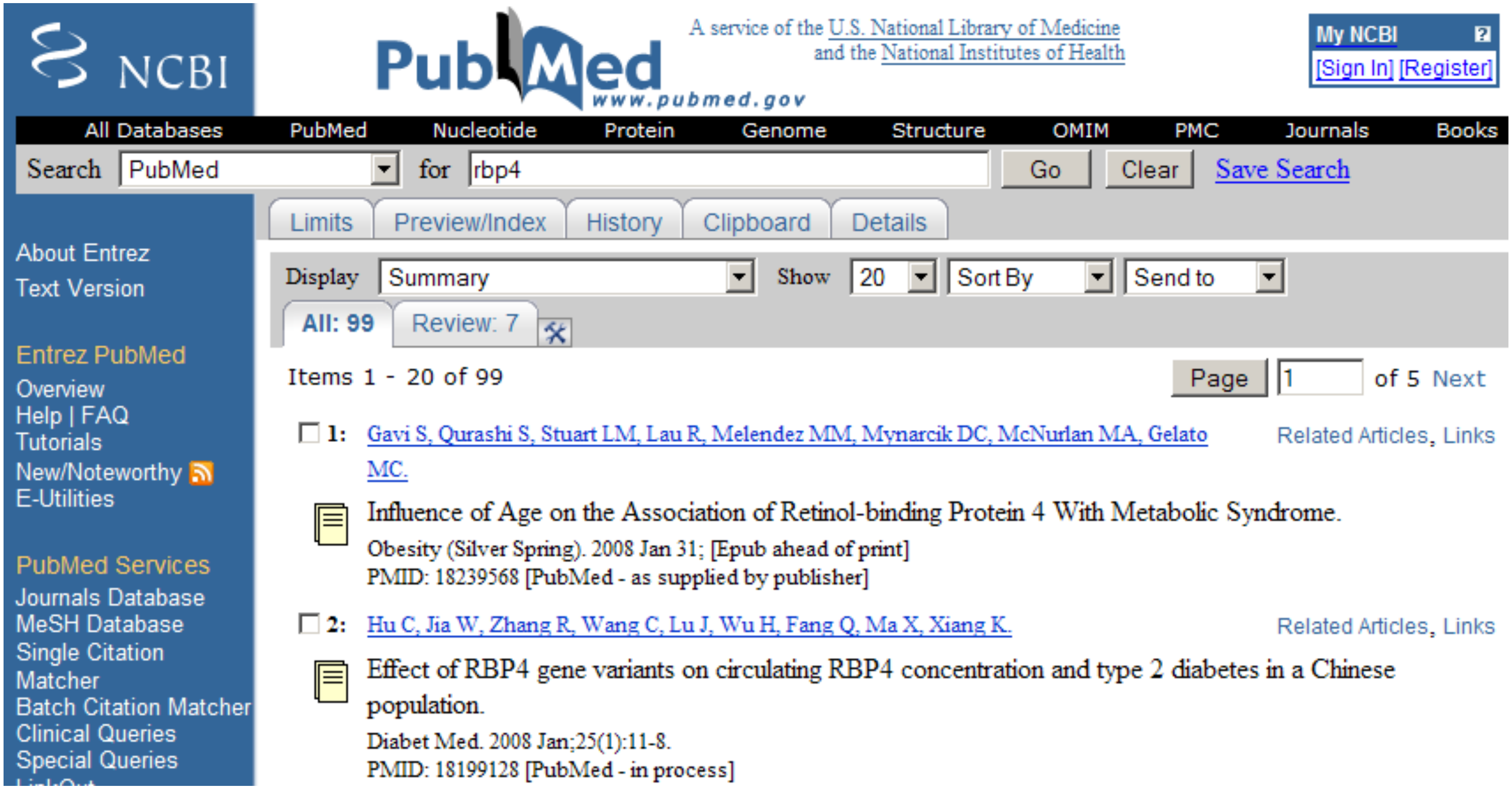
Search across databases: rbp4

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

99	**PubMed:** biomedical literature citations and abstracts
57	**PubMed Central:** free, full text journal articles
none	**Site Search:** NCBI web and FTP sites
none	**Books:** online books
6	**OMIM:** online Mendelian Inheritance in Man
none	**OMIA:** online Mendelian Inheritance in Animals

82	**CoreNucleotide:** Core subset of nucleotide sequence records
52	**EST:** Expressed Sequence Tag records
3	**GSS:** Genome Survey Sequence records
71	**Protein:** sequence database
6	**Genome:** whole genome sequences
none	**Structure:** three-dimensional macromolecular structures
none	**dbGaP:** genotype and phenotype
12	**UniGene:** gene-oriented clusters of transcript sequences
none	**CDD:** conserved protein domain database
none	**3D Domains:** domains from Entrez Structure
13	**UniSTS:** markers and mapping data
1	**PopSet:** population study data sets

NCBI Entrez PubMed Results



Search PubMed for rbp4

Display Summary Show 20

Items 1 - 20 of 99 Page 1 of 5

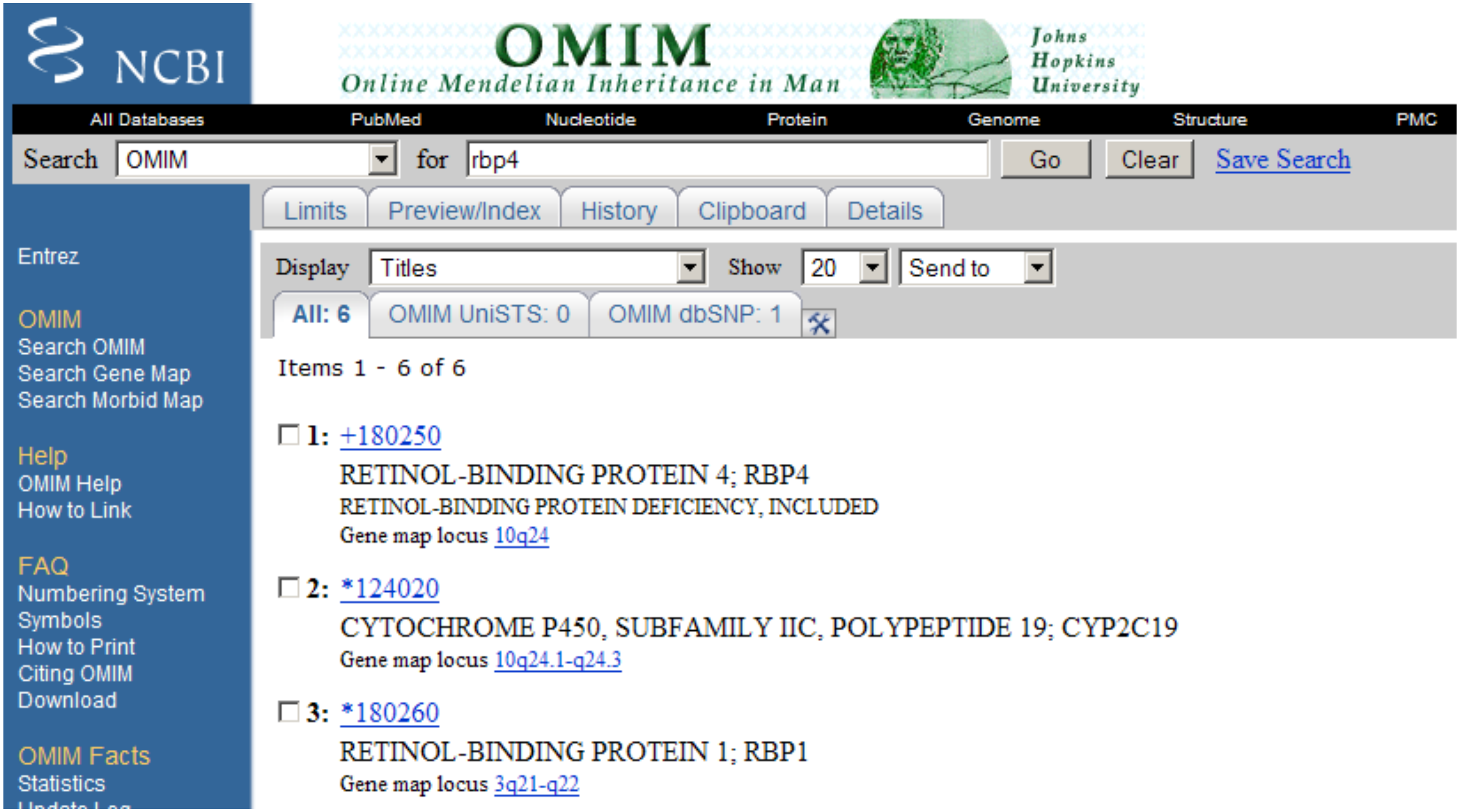
□ 1: Gavi S, Qurashi S, Stuart LM, Lau R, Melendez MM, Mynarcik DC, McNurlan MA, Gelato MC. Related Articles, Links

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



Search OMIM for rbp4

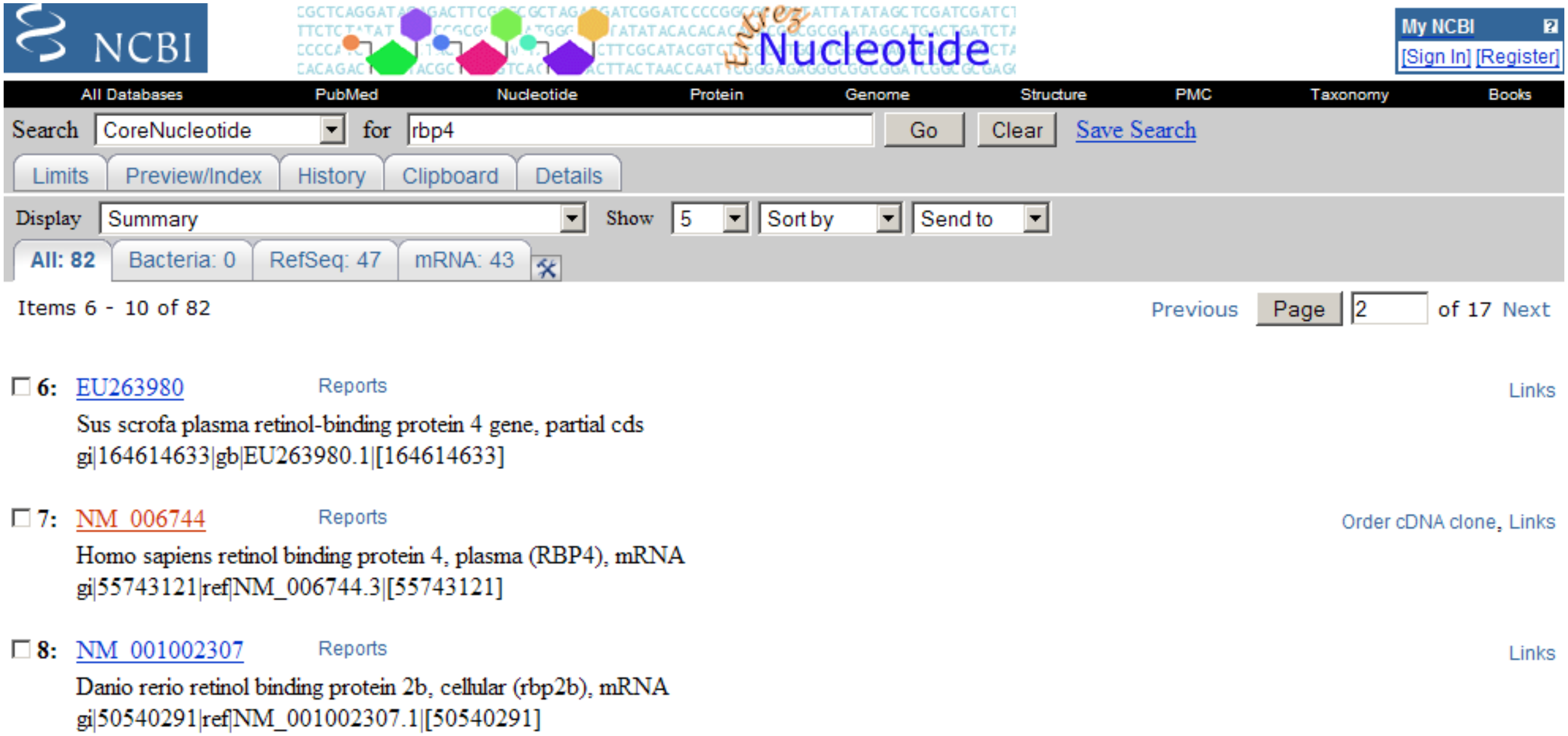
Items 1 - 6 of 6

□ 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



Search CoreNucleotide for rbp4

Display Summary Show 5

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

Items 6 - 10 of 82

Page 2 of 17

□ 6: EU263980 Reports Links

Sus scrofa plasma retinol-binding protein 4 gene, partial cds
gi|164614633|gb|EU263980.1|[164614633]

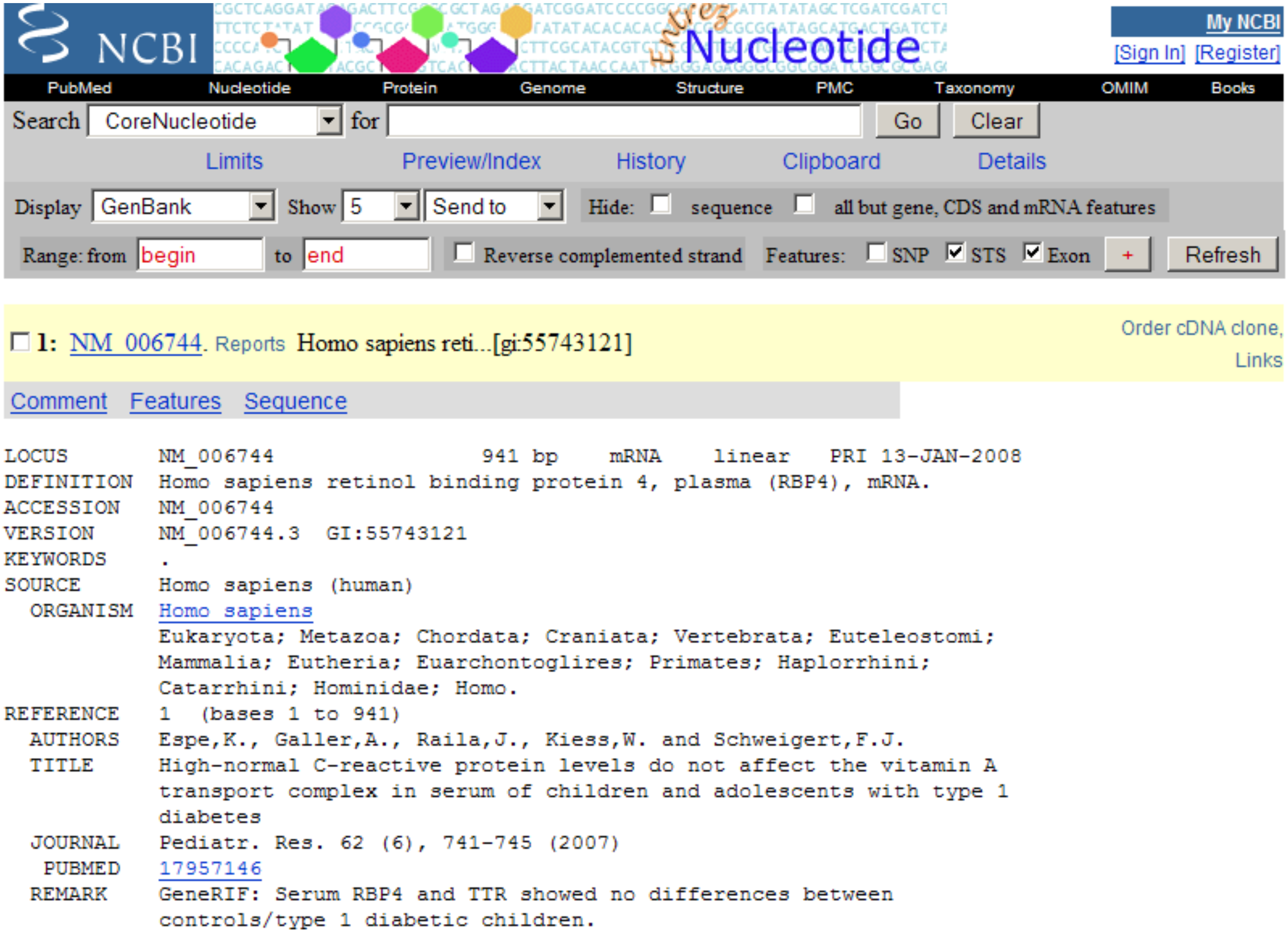
□ 7: NM_006744 Reports Order cDNA clone, Links

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]

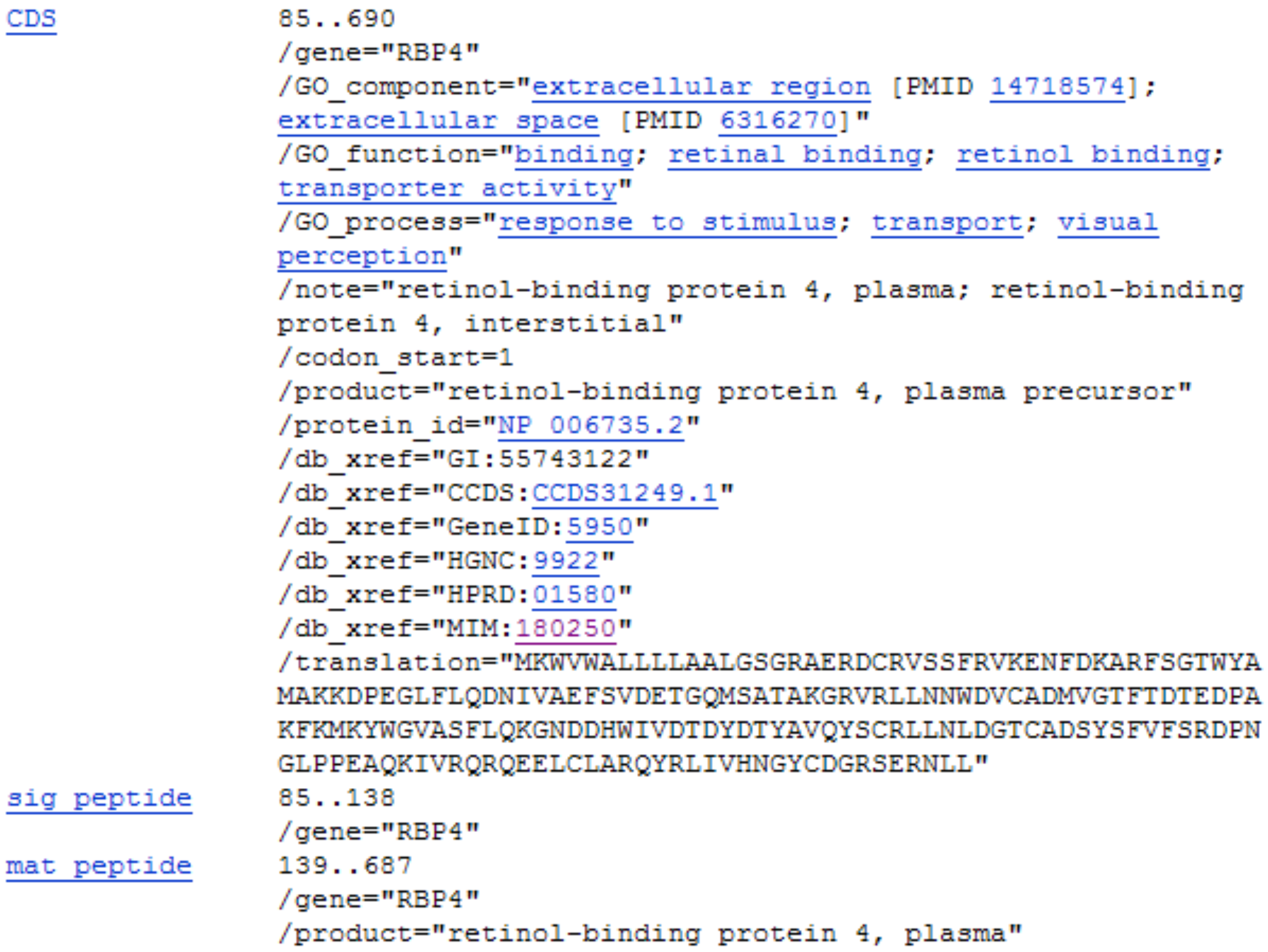
NCBI Entrez Core Nucleotide Results



□ 1: NM_006744. Reports Homo sapiens reti...[gi:55743121]

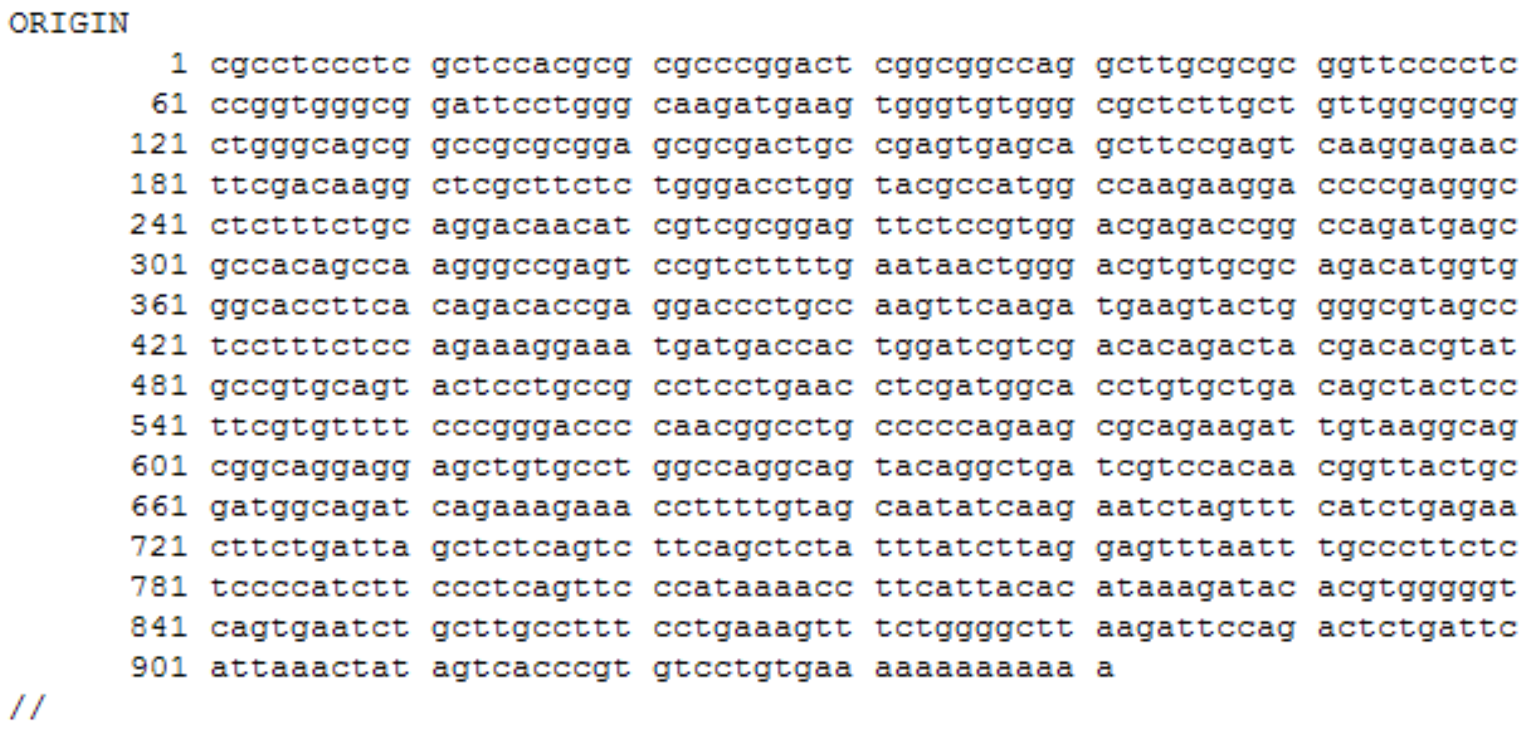
```
LOCUS       NM_006744                941 bp    mRNA    linear   PRI 13-JAN-2008
DEFINITION  Homo sapiens retinol binding protein 4, plasma (RBP4), mRNA.
ACCESSION   NM_006744
VERSION     NM_006744.3  GI:55743121
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; 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



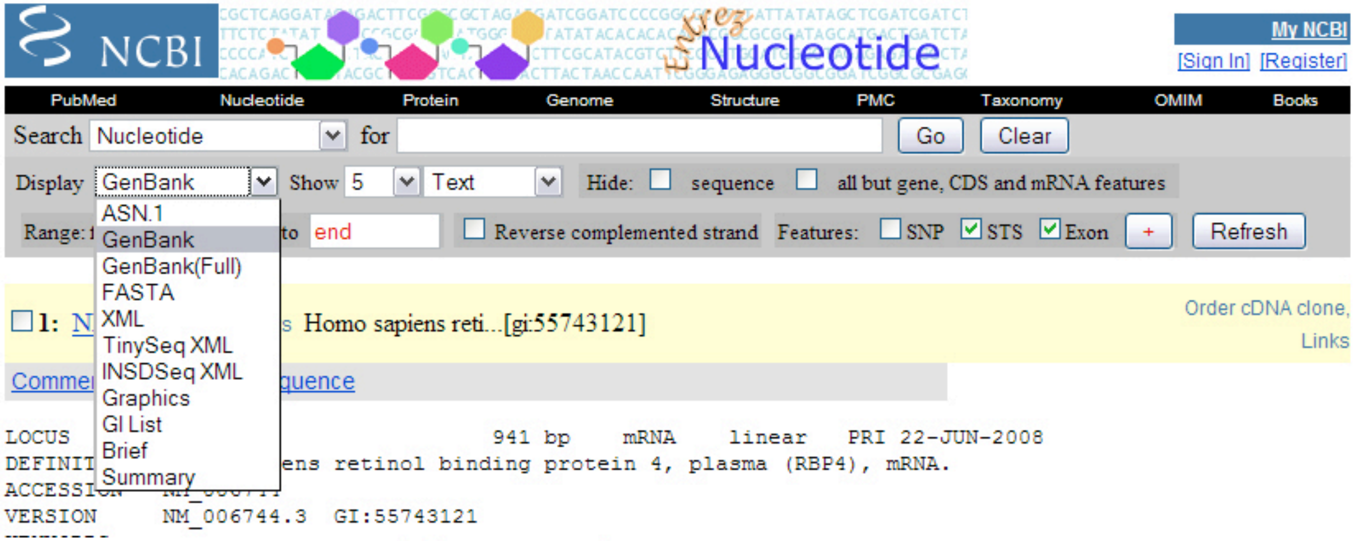
```
CDS                  85..690
                     /gene="RBP4"
                     /GO_component="extracellular region [PMID 14718574];
                     extracellular space [PMID 6316270]"
                     /GO_function="binding; retinal binding; retinol binding;
                     transporter 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
                     MAKKDPEGLFLQDNIVAEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPA
                     KFKMKYWGVASFLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPN
                     GLPPEAQKIVRQRQEELCLARQYRLIVHNGYCDGRSERNLL"
sig peptide          85..138
                     /gene="RBP4"
mat peptide          139..687
                     /gene="RBP4"
                     /product="retinol-binding protein 4, plasma"
```

NCBI Entrez Core Nucleotide Results



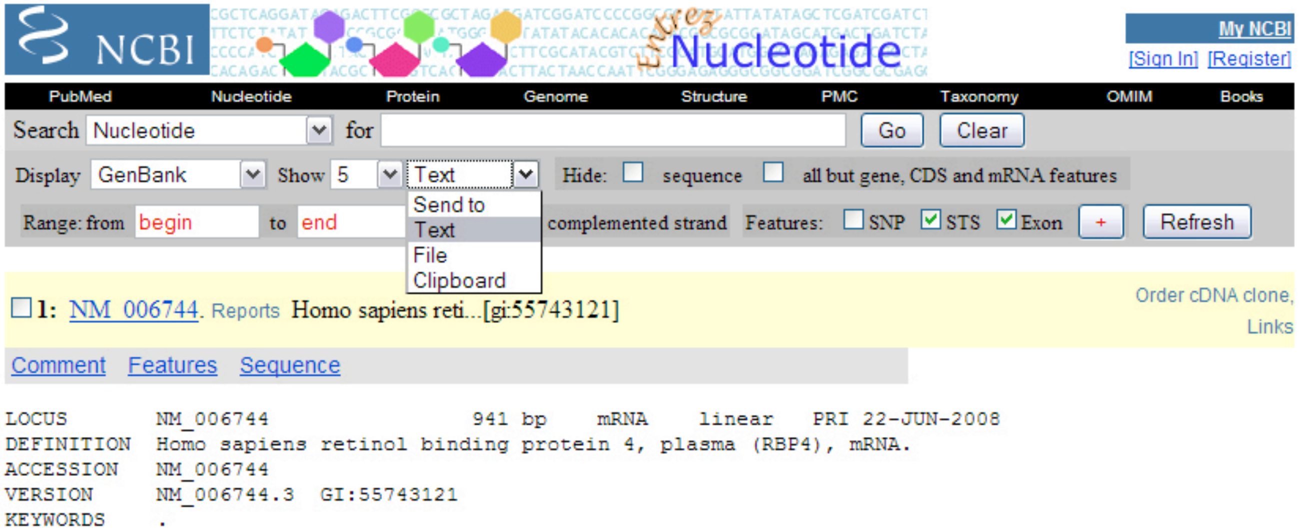
```
ORIGIN
        1 cgcctccctc gctccacgcg cgcccggact cggcggccag gcttgcgcgc ggttcccctc
       61 ccggtgggcg gattcctggg caagatgaag tgggtgtggg cgctcttgct gttggcggcg
      121 ctgggcagcg gccgcgcgga gcgcgactgc cgagtgagca gcttccgagt caaggagaac
      181 ttcgacaagg ctcgcttctc tgggacctgg tacgccatgg ccaagaagga 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 cccgggaccc caacggcctg cccccagaag cgcagaagat tgtaaggcag
      601 cggcaggagg agctgtgcct ggccaggcag tacaggctga tcgtccacaa cggttactgc
      661 gatggcagat cagaaagaaa ccttttgtag caatatcaag aatctagttt catctgagaa
      721 cttctgatta gctctcagtc ttcagctcta tttatcttag gagtttaatt tgcccttctc
      781 tccccatctt ccctcagttc ccataaaacc ttcattacac ataaagatac acgtgggggt
      841 cagtgaatct gcttgccttt cctgaaagtt tctggggctt aagattccag actctgattc
      901 attaaactat agtcacccgt gtcctgtgaa aaaaaaaaaa a
//
```

NCBI Entrez Saving Sequences



□ 1: N s Homo sapiens reti...[gi:55743121]

```
LOCUS                                941 bp    mRNA    linear   PRI 22-JUN-2008
DEFINIT                 ens retinol binding protein 4, plasma (RBP4), mRNA.
ACCESSI
VERSION     NM_006744.3  GI:55743121
```



□ 1: NM_006744. Reports Homo sapiens reti...[gi:55743121]

```
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

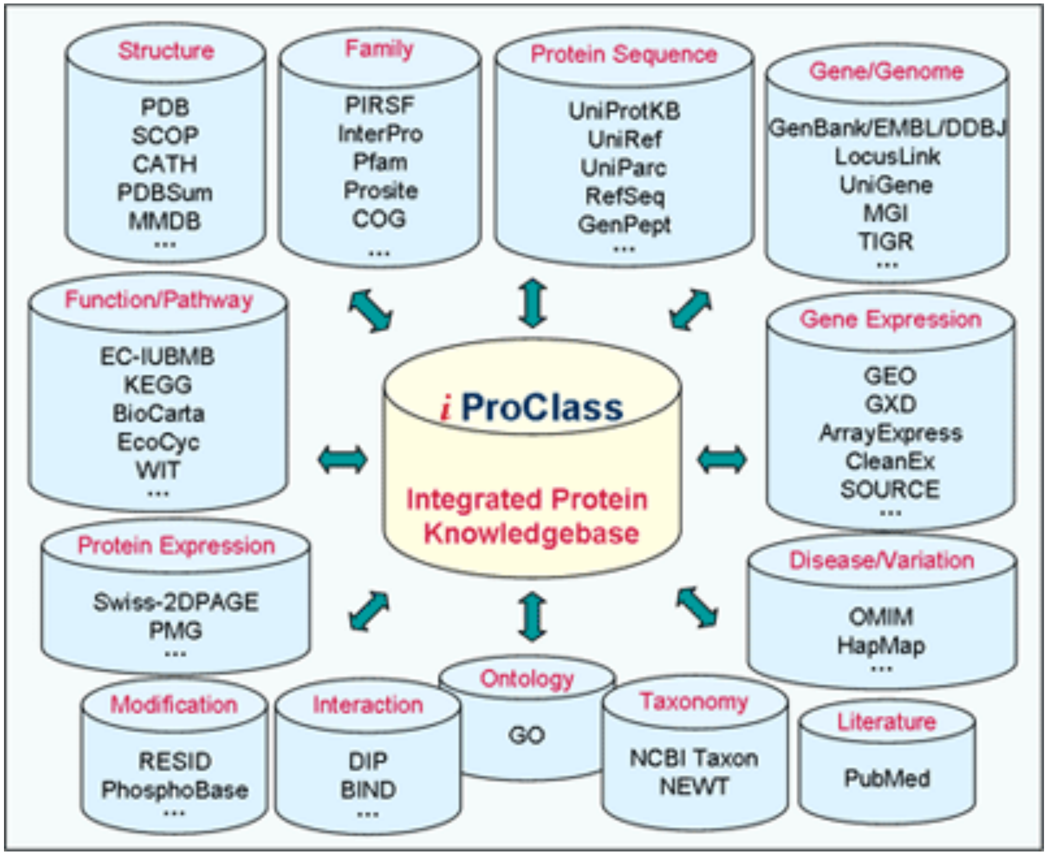
- **Accession Number:** unique identifier given to a sequence when it is submitted to one of the DNA repositories (GenBank, EMBL, DDBJ). These identifiers follow an accession.version format. Updates increment the version, while the accession remains constant.

- **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.

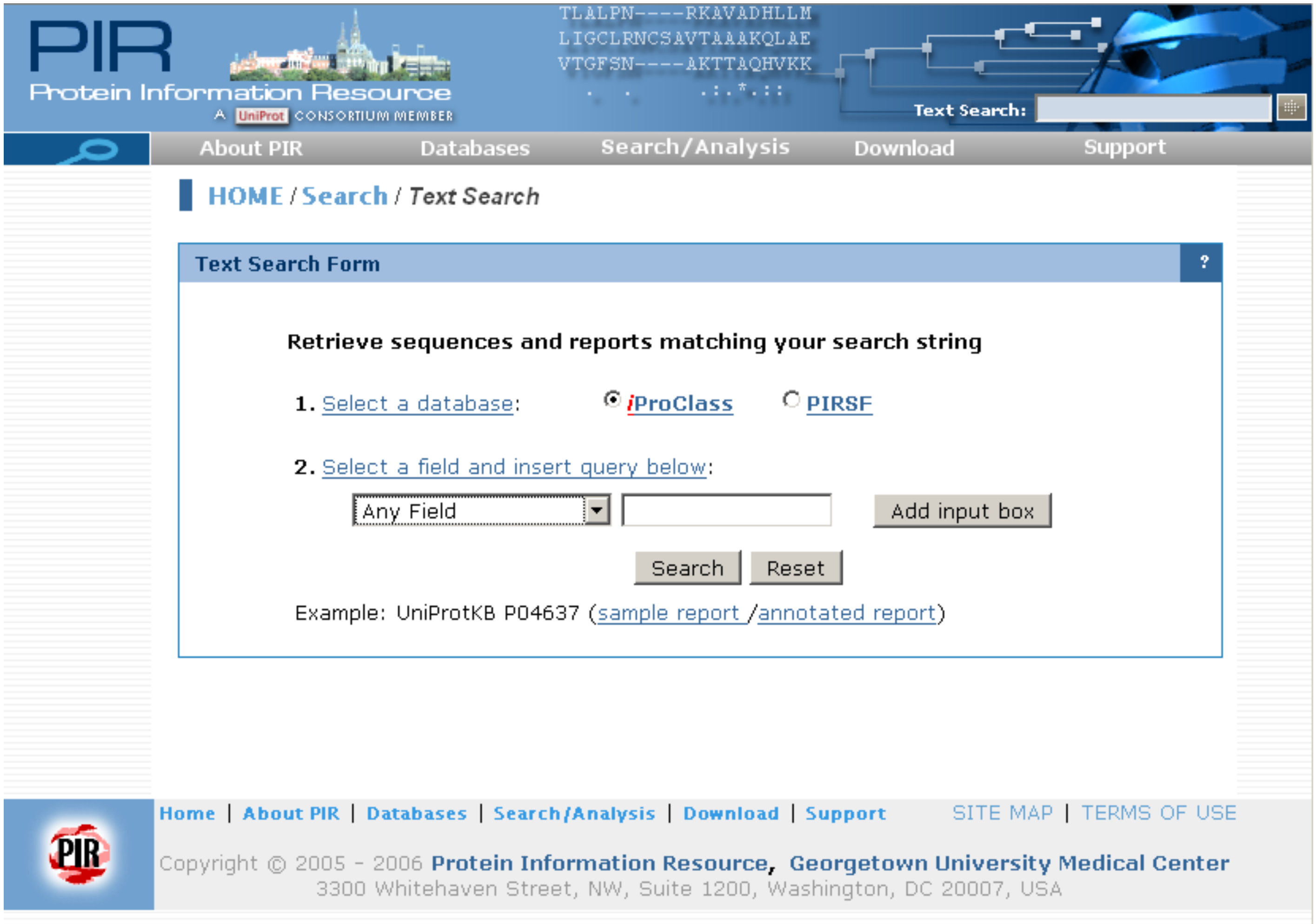
iProClass Protein Knowledgebase

- 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



PIR
Protein Information Resource
A UniProt CONSORTIUM MEMBER
TLALPN----RKAVADHLLM
LIGCLRNCSAVTAAAKQLAE
VTGFSN----AKTTAQHVKK
Text Search:
About PIR
Databases
Search/Analysis
Download
Support

HOME / Search / *Text Search*

Text Search Form ?

Retrieve sequences and reports matching your search string

1. Select a database: ● *i*ProClass ○ PIRSF

2. Select a field and insert query below:

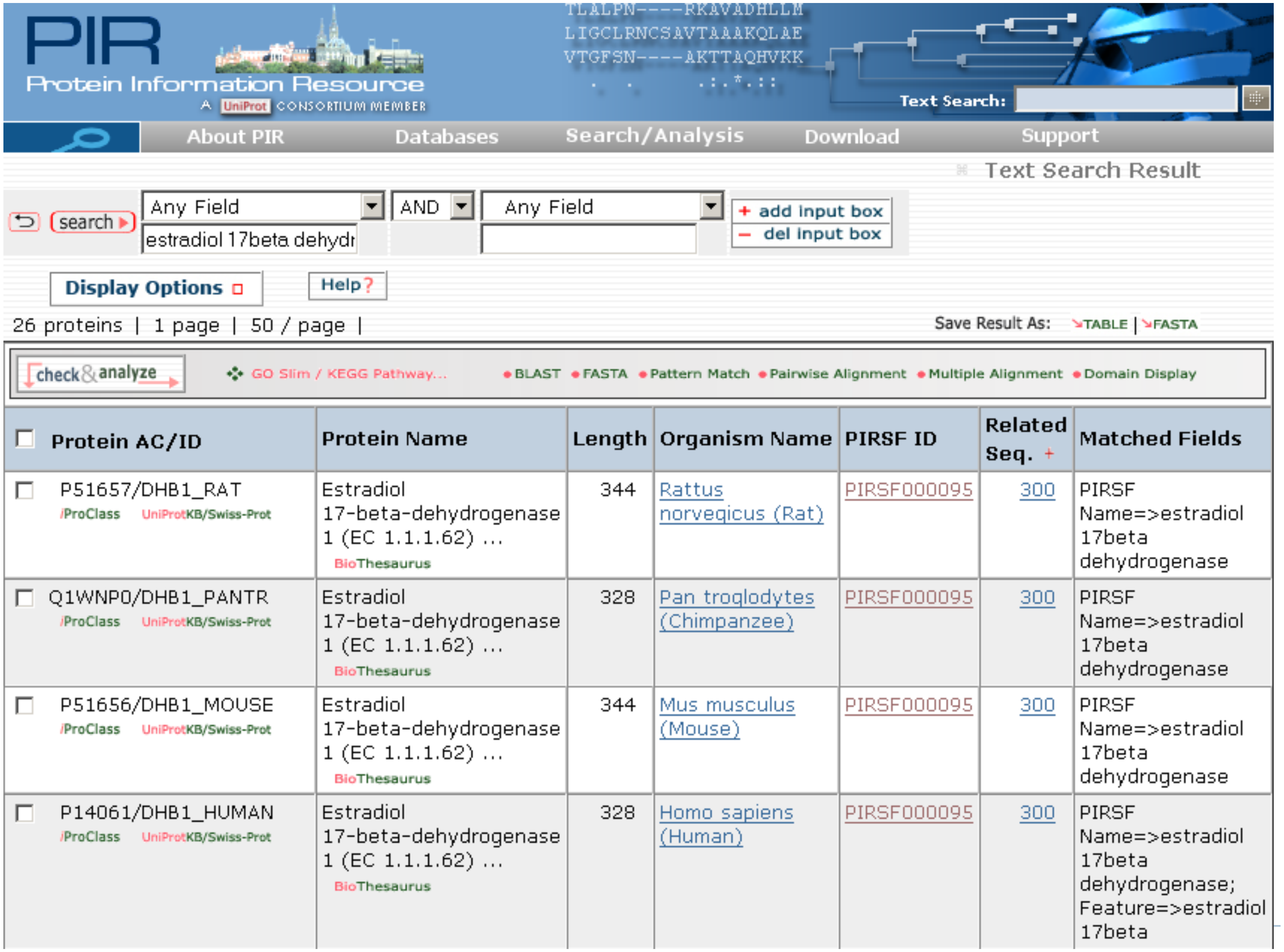
Any Field ▾ [ ] Add input box

Search Reset

Example: UniProtKB P04637 (sample report /annotated report)

Home | About PIR | Databases | Search/Analysis | Download | Support SITE MAP | TERMS OF USE
PIR

iProclass Results



Text Search Result

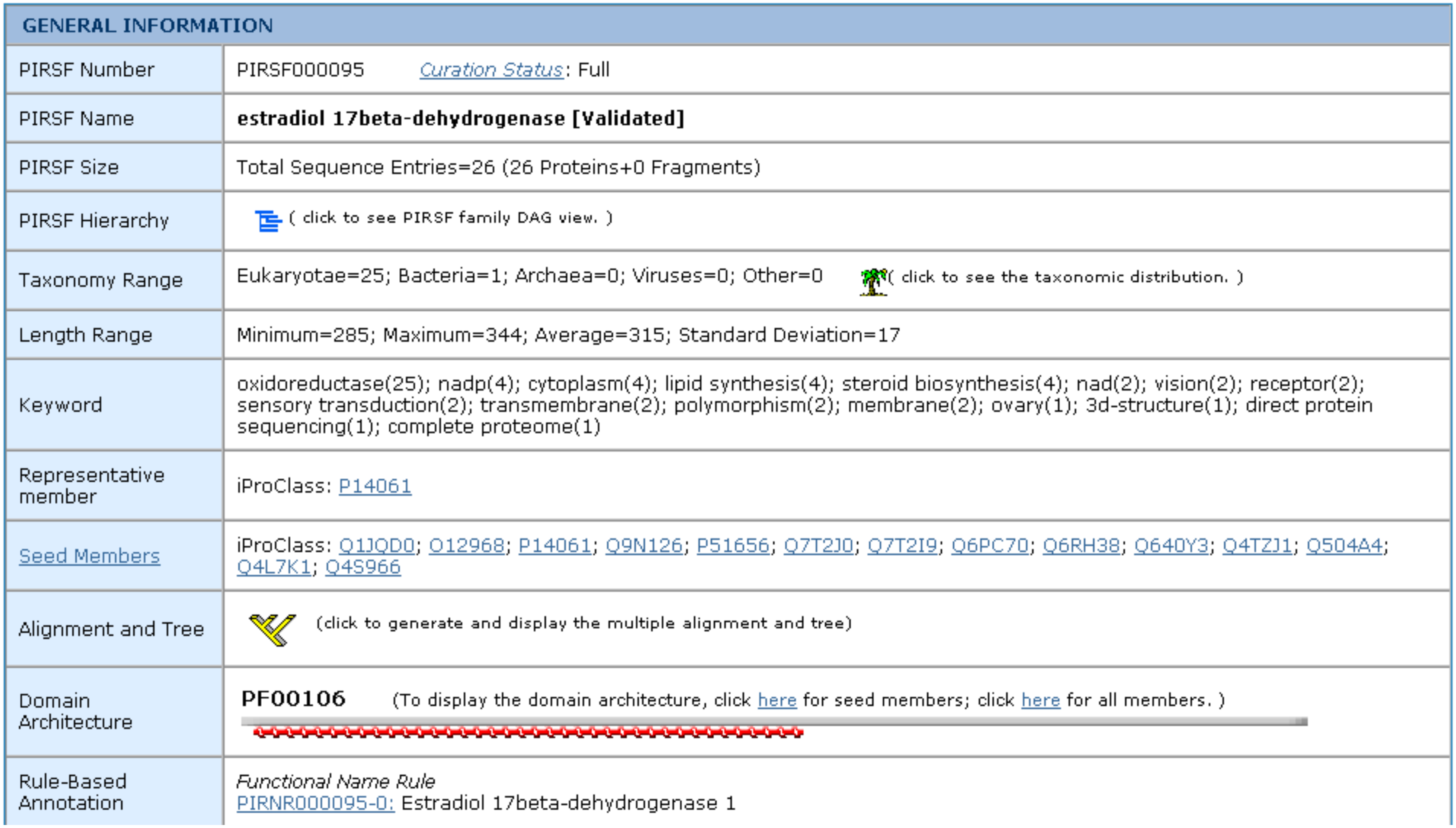
Any Field: estradiol 17beta dehydr
AND
Any Field:

26 proteins | 1 page | 50 / page |

Save Result As: TABLE | FASTA

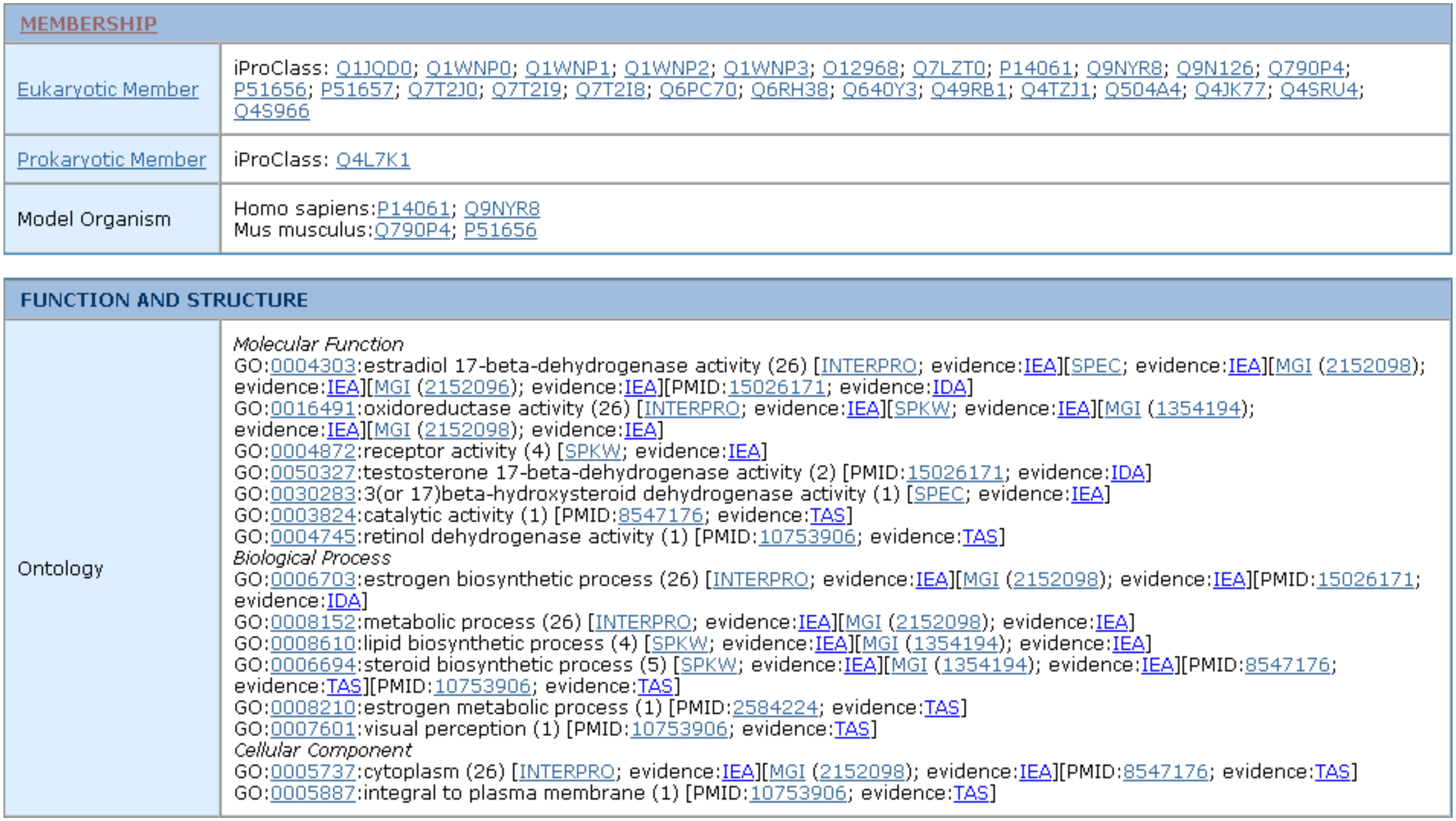
□ Protein AC/ID	Protein Name	Length	Organism Name	PIRSF ID	Related Seq. +	Matched Fields
□ P51657/DHB1_RAT iProClass UniProtKB/Swiss-Prot	Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) ... BioThesaurus	344	Rattus norvegicus (Rat)	PIRSF000095	300	PIRSF Name=>estradiol 17beta dehydrogenase
□ Q1WNP0/DHB1_PANTR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

iProClass SuperFamily Summary



GENERAL INFORMATION	
PIRSF Number	PIRSF000095 *Curation Status*: 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: Q1JQD0; 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	*Functional Name Rule* PIRNR000095-0: Estradiol 17beta-dehydrogenase 1

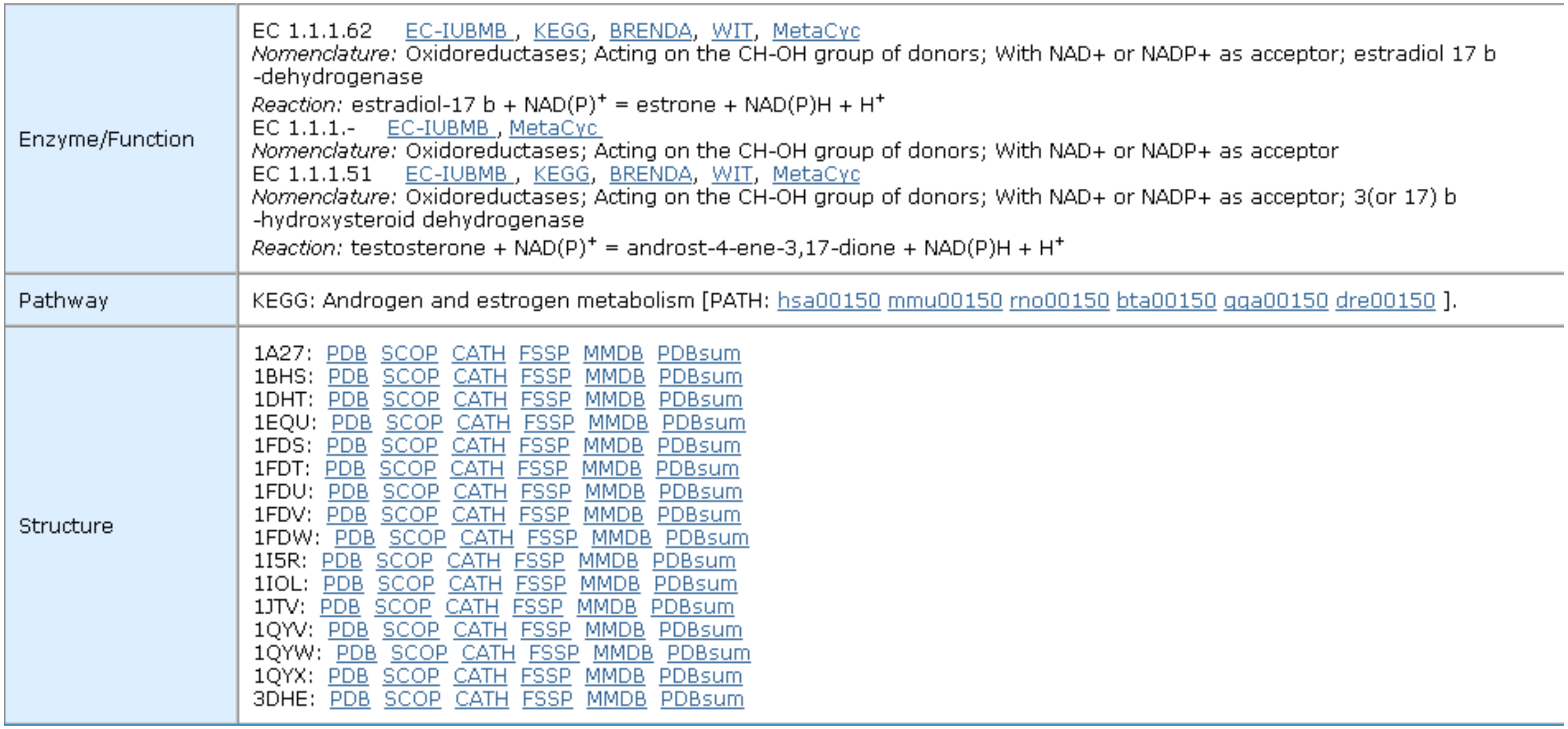
iProClass SuperFamily Summary



MEMBERSHIP	
Eukaryotic Member	iProClass: Q1JQD0; Q1WNP0; Q1WNP1; Q1WNP2; Q1WNP3; O12968; Q7LZT0; P14061; Q9NYR8; Q9N126; Q790P4; P51656; P51657; Q7T2J0; Q7T2I9; Q7T2I8; Q6PC70; Q6RH38; Q640Y3; Q49RB1; Q4TZJ1; Q504A4; Q4JK77; Q4SRU4; Q4S966
Prokaryotic Member	iProClass: Q4L7K1
Model Organism	Homo sapiens:P14061; Q9NYR8 Mus musculus:Q790P4; P51656

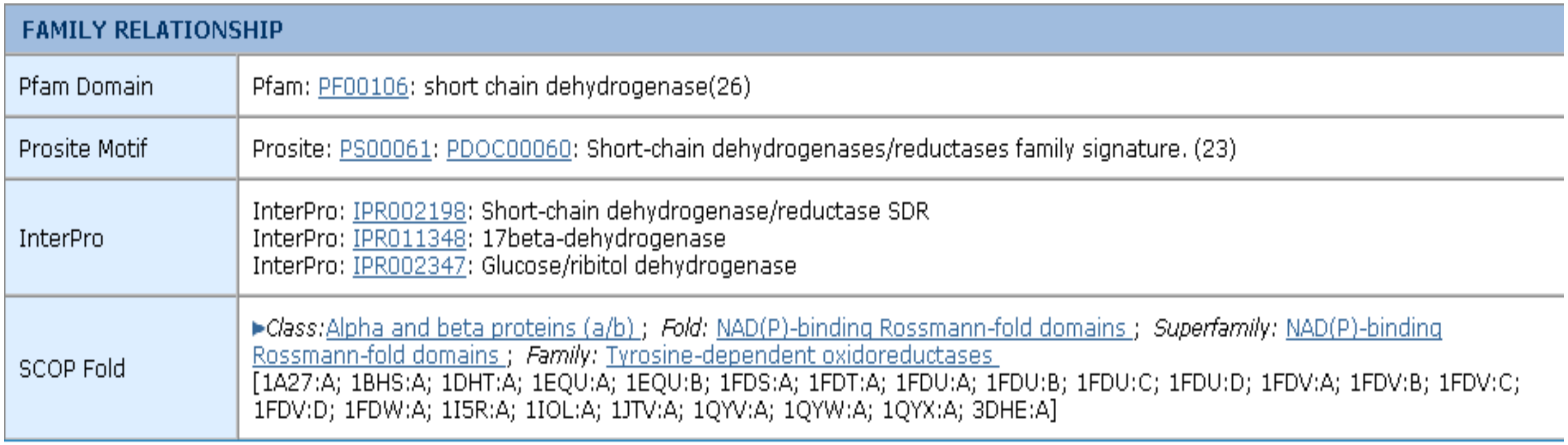
FUNCTION AND STRUCTURE	
Ontology	*Molecular Function* GO:0004303:estradiol 17-beta-dehydrogenase activity (26) [INTERPRO; evidence:IEA][SPEC; evidence:IEA][MGI (2152098); evidence:IEA][MGI (2152096); evidence:IEA][PMID:15026171; evidence:IDA] GO:0016491:oxidoreductase activity (26) [INTERPRO; evidence:IEA][SPKW; evidence:IEA][MGI (1354194); evidence:IEA][MGI (2152098); evidence:IEA] GO:0004872:receptor activity (4) [SPKW; evidence:IEA] GO:0050327:testosterone 17-beta-dehydrogenase activity (2) [PMID:15026171; evidence:IDA] GO:0030283:3(or 17)beta-hydroxysteroid dehydrogenase activity (1) [SPEC; evidence:IEA] GO:0003824:catalytic activity (1) [PMID:8547176; evidence:TAS] GO:0004745:retinol dehydrogenase activity (1) [PMID:10753906; evidence:TAS] *Biological Process* GO:0006703:estrogen biosynthetic process (26) [INTERPRO; evidence:IEA][MGI (2152098); evidence:IEA][PMID:15026171; evidence:IDA] GO:0008152:metabolic process (26) [INTERPRO; evidence:IEA][MGI (2152098); evidence:IEA] GO:0008610:lipid biosynthetic process (4) [SPKW; evidence:IEA][MGI (1354194); evidence:IEA] GO:0006694:steroid biosynthetic process (5) [SPKW; evidence:IEA][MGI (1354194); evidence:IEA][PMID:8547176; evidence:TAS][PMID:10753906; evidence:TAS] GO:0008210:estrogen metabolic process (1) [PMID:2584224; evidence:TAS] GO:0007601:visual perception (1) [PMID:10753906; evidence:TAS] *Cellular Component* GO:0005737:cytoplasm (26) [INTERPRO; evidence:IEA][MGI (2152098); evidence:IEA][PMID:8547176; evidence:TAS] GO:0005887:integral to plasma membrane (1) [PMID:10753906; evidence:TAS]

iProClass SuperFamily Summary



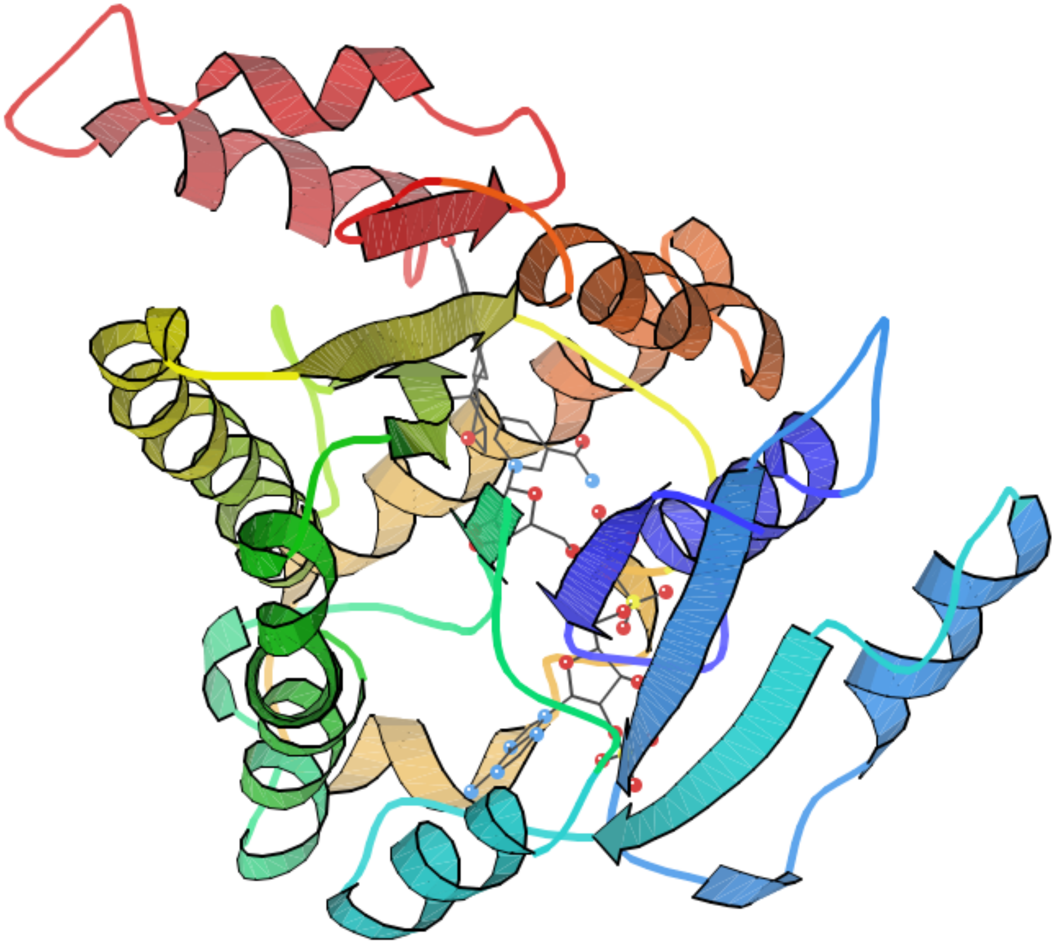
Enzyme/Function	EC 1.1.1.62 EC-IUBMB , KEGG, BRENDA, WIT, MetaCyc *Nomenclature:* Oxidoreductases; Acting on the CH-OH group of donors; With NAD+ or NADP+ as acceptor; estradiol 17 b -dehydrogenase *Reaction:* estradiol-17 b + NAD(P)$^+$ = estrone + NAD(P)H + H$^+$ EC 1.1.1.- EC-IUBMB , MetaCyc *Nomenclature:* Oxidoreductases; Acting on the CH-OH group of donors; With NAD+ or NADP+ as acceptor EC 1.1.1.51 EC-IUBMB , KEGG, BRENDA, WIT, MetaCyc *Nomenclature:* Oxidoreductases; Acting on the CH-OH group of donors; With NAD+ or NADP+ as acceptor; 3(or 17) b -hydroxysteroid dehydrogenase *Reaction:* testosterone + NAD(P)$^+$ = androst-4-ene-3,17-dione + NAD(P)H + H$^+$
Pathway	KEGG: Androgen and estrogen metabolism [PATH: hsa00150 mmu00150 rno00150 bta00150 gga00150 dre00150].
Structure	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

iProClass SuperFamily Summary

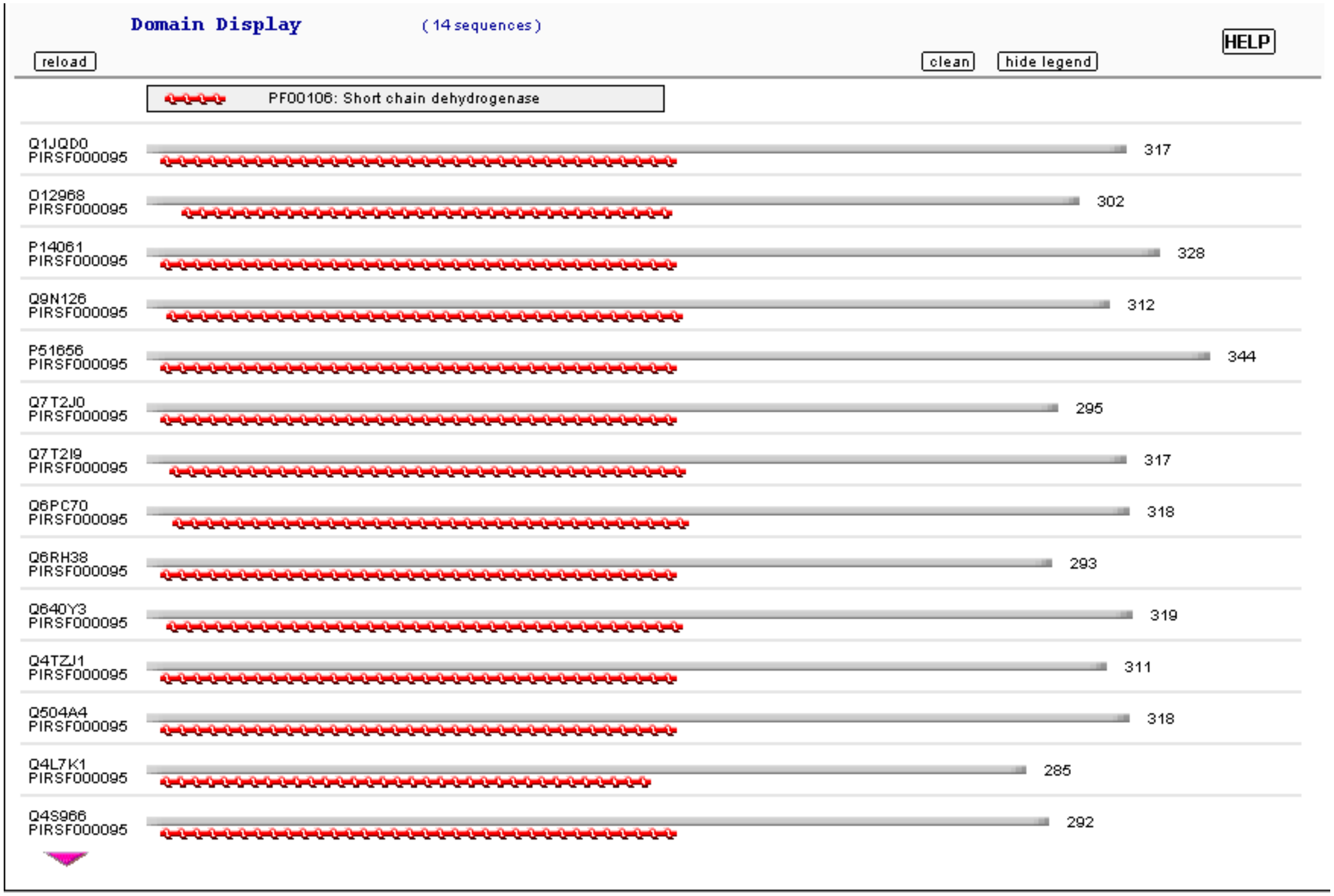


FAMILY RELATIONSHIP	
Pfam Domain	Pfam: PF00106: short chain dehydrogenase(26)
Prosite Motif	Prosite: PS00061: PDOC00060: Short-chain dehydrogenases/reductases family signature. (23)
InterPro	InterPro: IPR002198: Short-chain dehydrogenase/reductase SDR InterPro: IPR011348: 17beta-dehydrogenase InterPro: 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

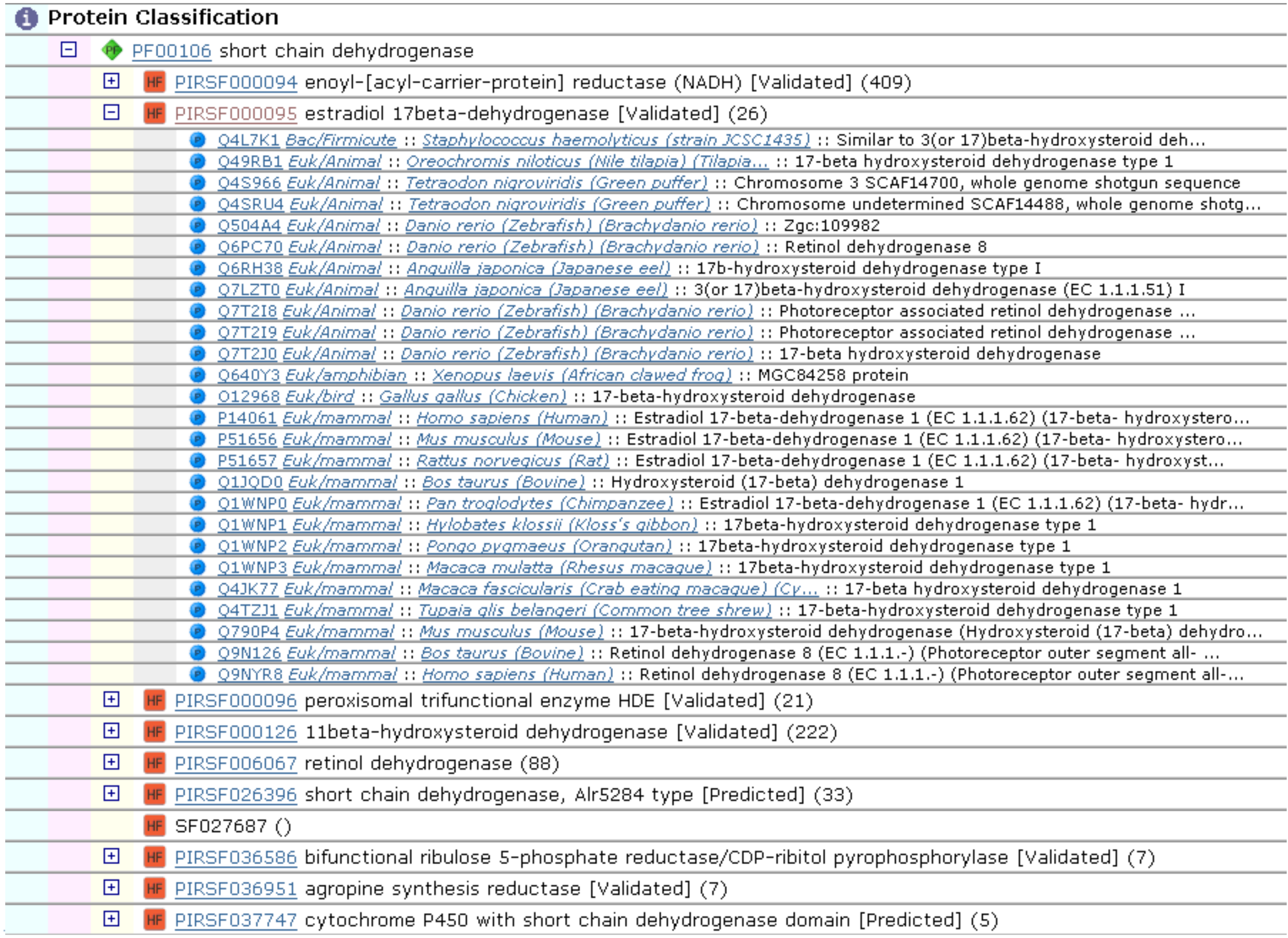


Domain Display (14 sequences)

PF00106: Short chain dehydrogenase

Q1JQD0	PIRSF000095	317
O12968	PIRSF000095	302
P14061	PIRSF000095	328
Q9N126	PIRSF000095	312
P51656	PIRSF000095	344
Q7T2J0	PIRSF000095	295
Q7T2I9	PIRSF000095	317
Q6PC70	PIRSF000095	318
Q6RH38	PIRSF000095	293
Q640Y3	PIRSF000095	319
Q4TZJ1	PIRSF000095	311
Q504A4	PIRSF000095	318
Q4L7K1	PIRSF000095	285
Q4S966	PIRSF000095	292

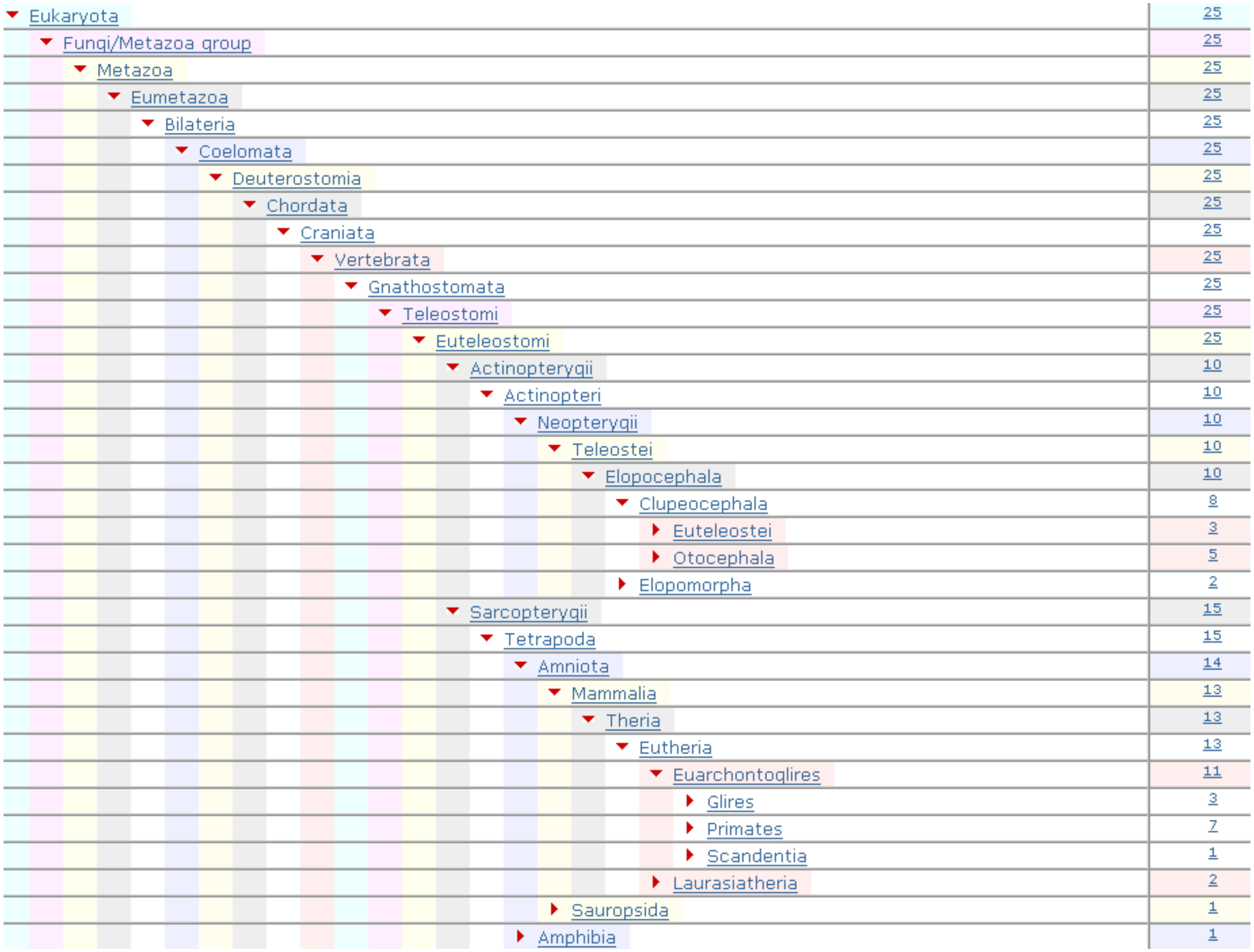
PIRSF Family Hierarchy



Protein Classification

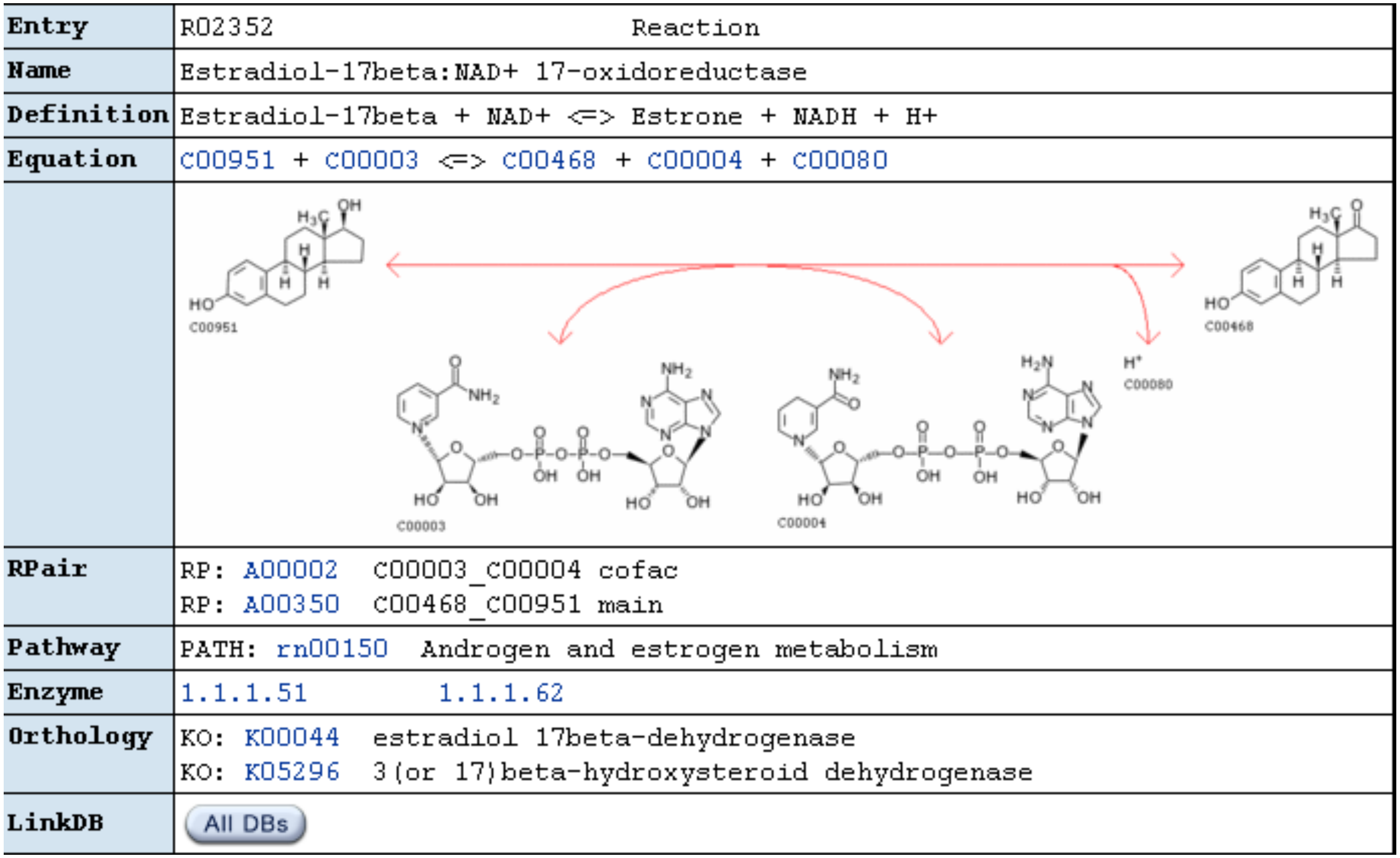
- ⊟ PF00106 short chain dehydrogenase
  - ⊞ HF PIRSF000094 enoyl-[acyl-carrier-protein] reductase (NADH) [Validated] (409)
  - ⊟ HF PIRSF000095 estradiol 17beta-dehydrogenase [Validated] (26)
    - Q4L7K1 *Bac/Firmicute* :: *Staphylococcus haemolyticus (strain JCSC1435)* :: Similar to 3(or 17)beta-hydroxysteroid deh...
    - Q49RB1 *Euk/Animal* :: *Oreochromis niloticus (Nile tilapia) (Tilapia...* :: 17-beta hydroxysteroid dehydrogenase type 1
    - Q4S966 *Euk/Animal* :: *Tetraodon nigroviridis (Green puffer)* :: Chromosome 3 SCAF14700, whole genome shotgun sequence
    - Q4SRU4 *Euk/Animal* :: *Tetraodon nigroviridis (Green puffer)* :: Chromosome undetermined SCAF14488, whole genome shotg...
    - Q504A4 *Euk/Animal* :: *Danio rerio (Zebrafish) (Brachydanio rerio)* :: Zgc:109982
    - Q6PC70 *Euk/Animal* :: *Danio rerio (Zebrafish) (Brachydanio rerio)* :: Retinol dehydrogenase 8
    - Q6RH38 *Euk/Animal* :: *Anguilla japonica (Japanese eel)* :: 17b-hydroxysteroid dehydrogenase type I
    - Q7LZT0 *Euk/Animal* :: *Anguilla japonica (Japanese eel)* :: 3(or 17)beta-hydroxysteroid dehydrogenase (EC 1.1.1.51) I
    - Q7T2I8 *Euk/Animal* :: *Danio rerio (Zebrafish) (Brachydanio rerio)* :: Photoreceptor associated retinol dehydrogenase ...
    - Q7T2I9 *Euk/Animal* :: *Danio rerio (Zebrafish) (Brachydanio rerio)* :: Photoreceptor associated retinol dehydrogenase ...
    - Q7T2J0 *Euk/Animal* :: *Danio rerio (Zebrafish) (Brachydanio rerio)* :: 17-beta hydroxysteroid dehydrogenase
    - Q640Y3 *Euk/amphibian* :: *Xenopus laevis (African clawed frog)* :: MGC84258 protein
    - O12968 *Euk/bird* :: *Gallus gallus (Chicken)* :: 17-beta-hydroxysteroid dehydrogenase
    - P14061 *Euk/mammal* :: *Homo sapiens (Human)* :: Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta- hydroxystero...
    - P51656 *Euk/mammal* :: *Mus musculus (Mouse)* :: Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta- hydroxystero...
    - P51657 *Euk/mammal* :: *Rattus norvegicus (Rat)* :: Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta- hydroxyst...
    - Q1JQD0 *Euk/mammal* :: *Bos taurus (Bovine)* :: Hydroxysteroid (17-beta) dehydrogenase 1
    - Q1WNP0 *Euk/mammal* :: *Pan troglodytes (Chimpanzee)* :: Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta- hydr...
    - Q1WNP1 *Euk/mammal* :: *Hylobates klossii (Kloss's gibbon)* :: 17beta-hydroxysteroid dehydrogenase type 1
    - Q1WNP2 *Euk/mammal* :: *Pongo pygmaeus (Orangutan)* :: 17beta-hydroxysteroid dehydrogenase type 1
    - Q1WNP3 *Euk/mammal* :: *Macaca mulatta (Rhesus macaque)* :: 17beta-hydroxysteroid dehydrogenase type 1
    - Q4JK77 *Euk/mammal* :: *Macaca fascicularis (Crab eating macaque) (Cy...* :: 17-beta hydroxysteroid dehydrogenase 1
    - Q4TZJ1 *Euk/mammal* :: *Tupaia glis belangeri (Common tree shrew)* :: 17-beta-hydroxysteroid dehydrogenase type 1
    - Q790P4 *Euk/mammal* :: *Mus musculus (Mouse)* :: 17-beta-hydroxysteroid dehydrogenase (Hydroxysteroid (17-beta) dehydro...
    - Q9N126 *Euk/mammal* :: *Bos taurus (Bovine)* :: Retinol dehydrogenase 8 (EC 1.1.1.-) (Photoreceptor outer segment all- ...
    - Q9NYR8 *Euk/mammal* :: *Homo sapiens (Human)* :: Retinol dehydrogenase 8 (EC 1.1.1.-) (Photoreceptor outer segment all-...
  - ⊞ HF PIRSF000096 peroxisomal trifunctional enzyme HDE [Validated] (21)
  - ⊞ HF PIRSF000126 11beta-hydroxysteroid dehydrogenase [Validated] (222)
  - ⊞ HF PIRSF006067 retinol dehydrogenase (88)
  - ⊞ HF PIRSF026396 short chain dehydrogenase, Alr5284 type [Predicted] (33)
  - HF SF027687 ()
  - ⊞ HF PIRSF036586 bifunctional ribulose 5-phosphate reductase/CDP-ribitol pyrophosphorylase [Validated] (7)
  - ⊞ HF PIRSF036951 agropine synthesis reductase [Validated] (7)
  - ⊞ HF PIRSF037747 cytochrome P450 with short chain dehydrogenase domain [Predicted] (5)

iProClass Taxonomy Nodes



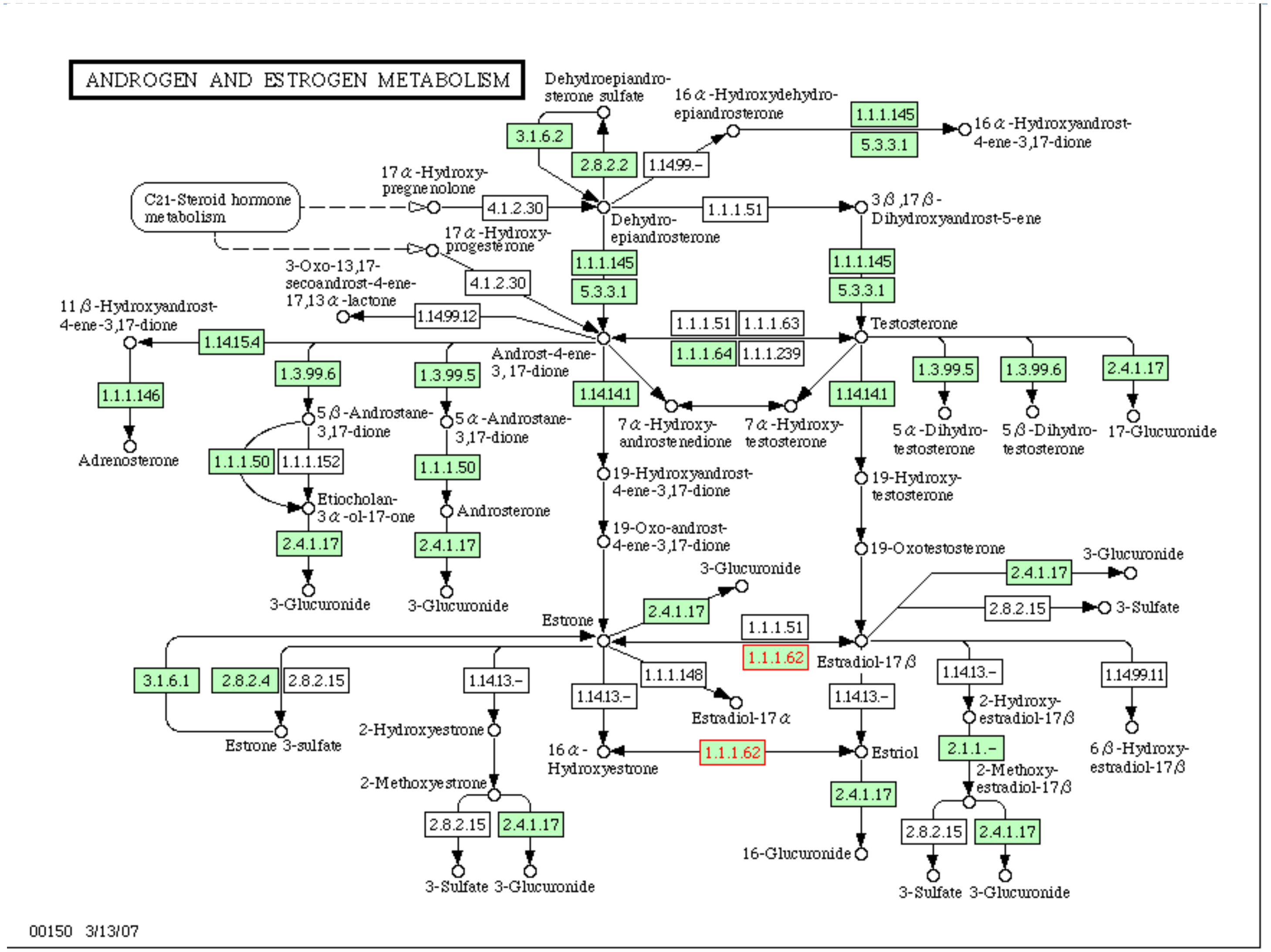
```
▼ Eukaryota                                                   25
  ▼ Fungi/Metazoa group                                       25
    ▼ Metazoa                                                 25
      ▼ Eumetazoa                                             25
        ▼ Bilateria                                           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
```

iProClass Enzyme Function: KEGG

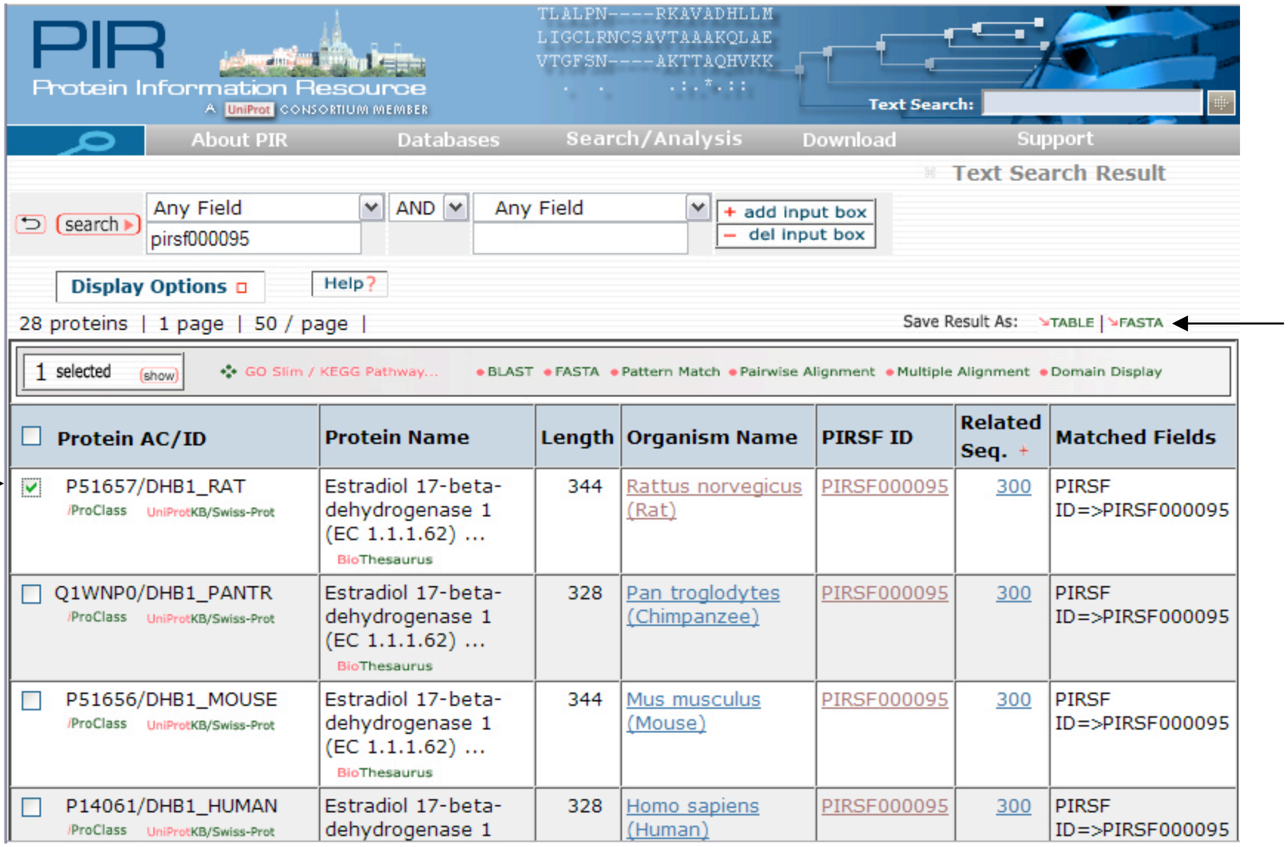


Entry	R02352 Reaction
Name	Estradiol-17beta:NAD+ 17-oxidoreductase
Definition	Estradiol-17beta + NAD+ <=> Estrone + NADH + H+
Equation	C00951 + C00003 <=> C00468 + C00004 + C00080
RPair	RP: A00002 C00003_C00004 cofac RP: A00350 C00468_C00951 main
Pathway	PATH: rn00150 Androgen and estrogen metabolism
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 Pathway: KEGG



iProClass: Saving Sequences



PIR Protein Information Resource

A UniProt CONSORTIUM MEMBER

```
TLALPN----RKAVADHLLM
LIGCLRNCSAVTAAAKQLAE
VTGFSN----AKTTAQHVKK
 .  .     .:.*.::
```

Text Search:

About PIR | Databases | Search/Analysis | Download | Support

Text Search Result

search ▸ Any Field: pirsf000095 | AND | Any Field

+ add input box
− del input box

Display Options □ | Help ?

28 proteins | 1 page | 50 / page |

Save Result As: TABLE | FASTA

1 selected (show) | GO Slim / KEGG Pathway... | BLAST | FASTA | Pattern Match | Pairwise Alignment | Multiple Alignment | Domain Display

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

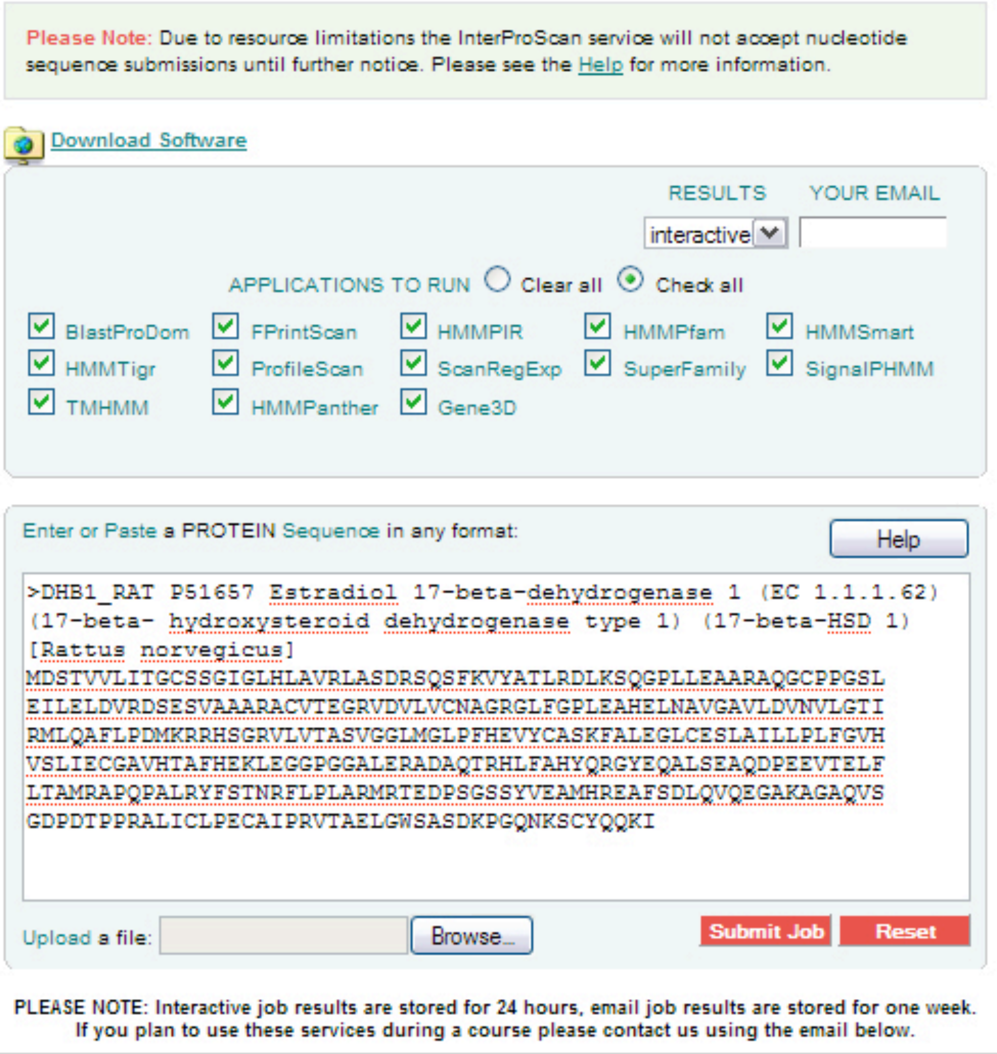
Save Format

Check Entries

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 meaticies, 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: interactive

YOUR EMAIL:

APPLICATIONS TO RUN ○ Clear all ● Check all

☑ BlastProDom ☑ FPrintScan ☑ HMMPIR ☑ HMMPfam ☑ HMMSmart
☑ HMMTigr ☑ ProfileScan ☑ ScanRegExp ☑ SuperFamily ☑ SignalPHMM
☑ TMHMM ☑ HMMPanther ☑ Gene3D

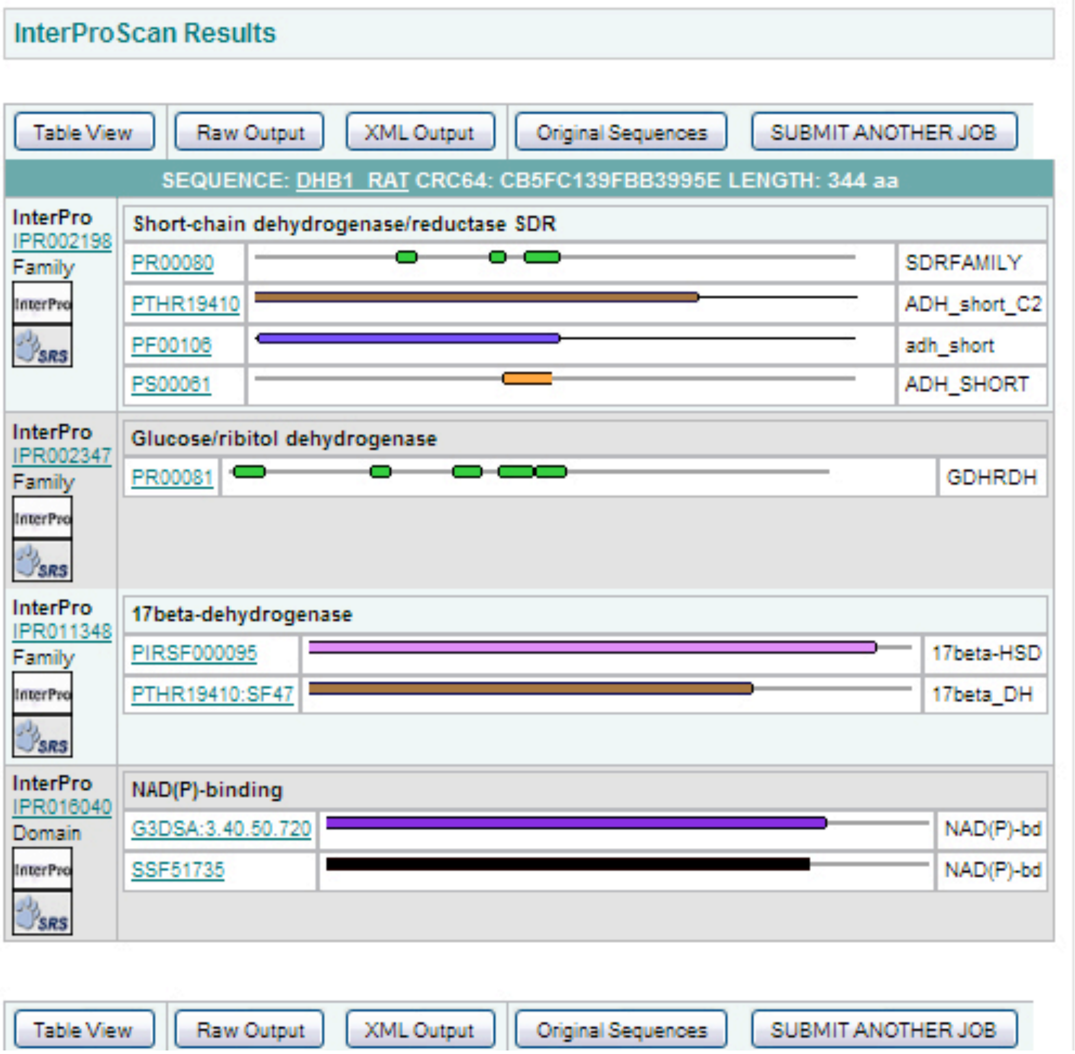
Enter or Paste a PROTEIN Sequence in any format:

```
>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]
MDSTVVLITGCSSGIGLHLAVRLASDRSQSFKVYATLRDLKSQGPLLEAARAQGCPPGSL
EILELDVRDSESVAAARACVTEGRVDVLVCNAGRGLFGPLEAHELNAVGAVLDVNVLGTI
RMLQAFLPDMKRRHSGRVLVTASVGGLMGLPFHEVYCASKFALEGLCESLAILLPLFGVH
VSLIECGAVHTAFHEKLEGGPGGALERADAQTRHLFAHYQRGYEQALSEAQDPEEVTELF
LTAMRAPQPALRYFSTNRFLPLARMRTEDPSGSSYVEAMHREAFSDLQVQEGAKAGAQVS
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

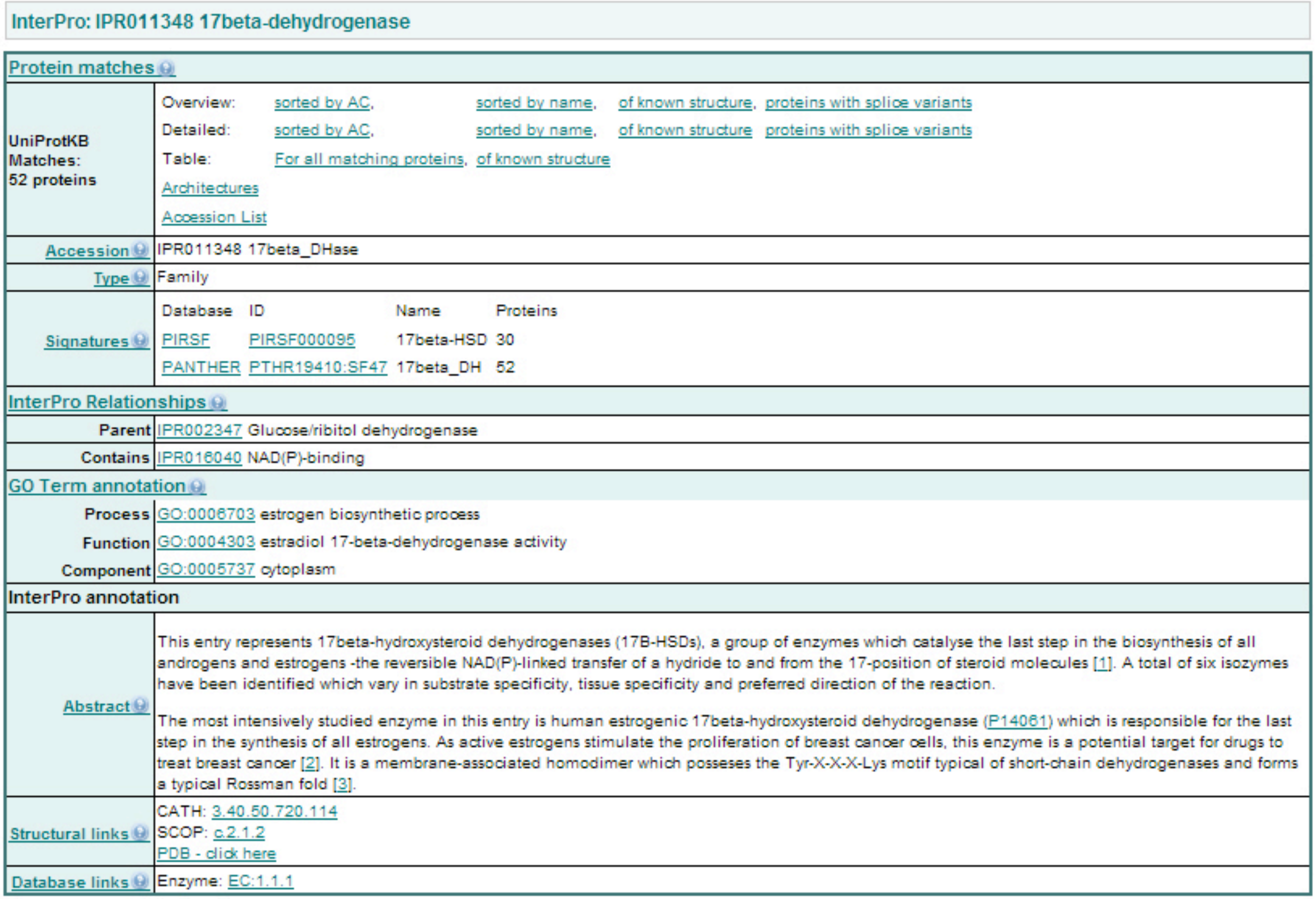


InterProScan Results

SEQUENCE: DHB1_RAT CRC64: CB5FC139FBB3995E LENGTH: 344 aa

InterPro IPR002198 Family	Short-chain dehydrogenase/reductase SDR	
	PR00080	SDRFAMILY
	PTHR19410	ADH_short_C2
	PF00106	adh_short
	PS00061	ADH_SHORT
InterPro IPR002347 Family	Glucose/ribitol dehydrogenase	
	PR00081	GDHRDH
InterPro IPR011348 Family	17beta-dehydrogenase	
	PIRSF000095	17beta-HSD
	PTHR19410:SF47	17beta_DH
InterPro IPR016040 Domain	NAD(P)-binding	
	G3DSA:3.40.50.720	NAD(P)-bd
	SSF51735	NAD(P)-bd

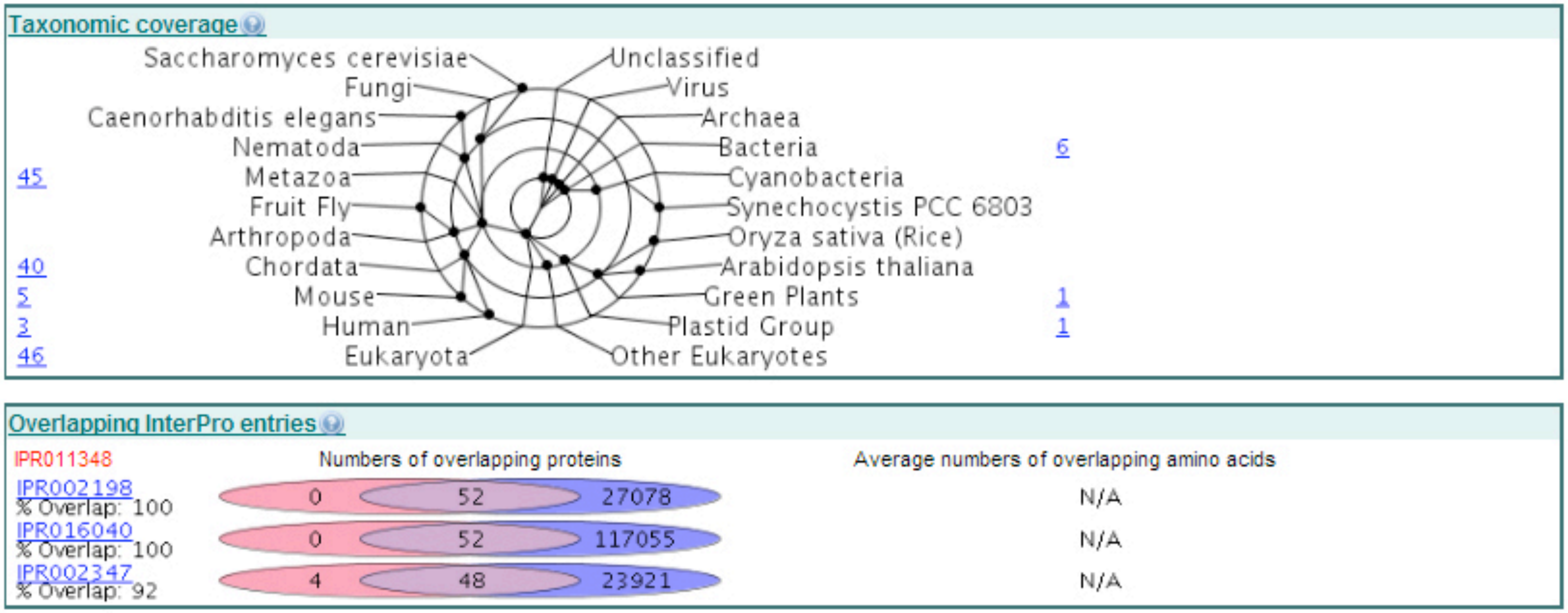
InterPro - InterProScan Results



InterPro: IPR011348 17beta-dehydrogenase

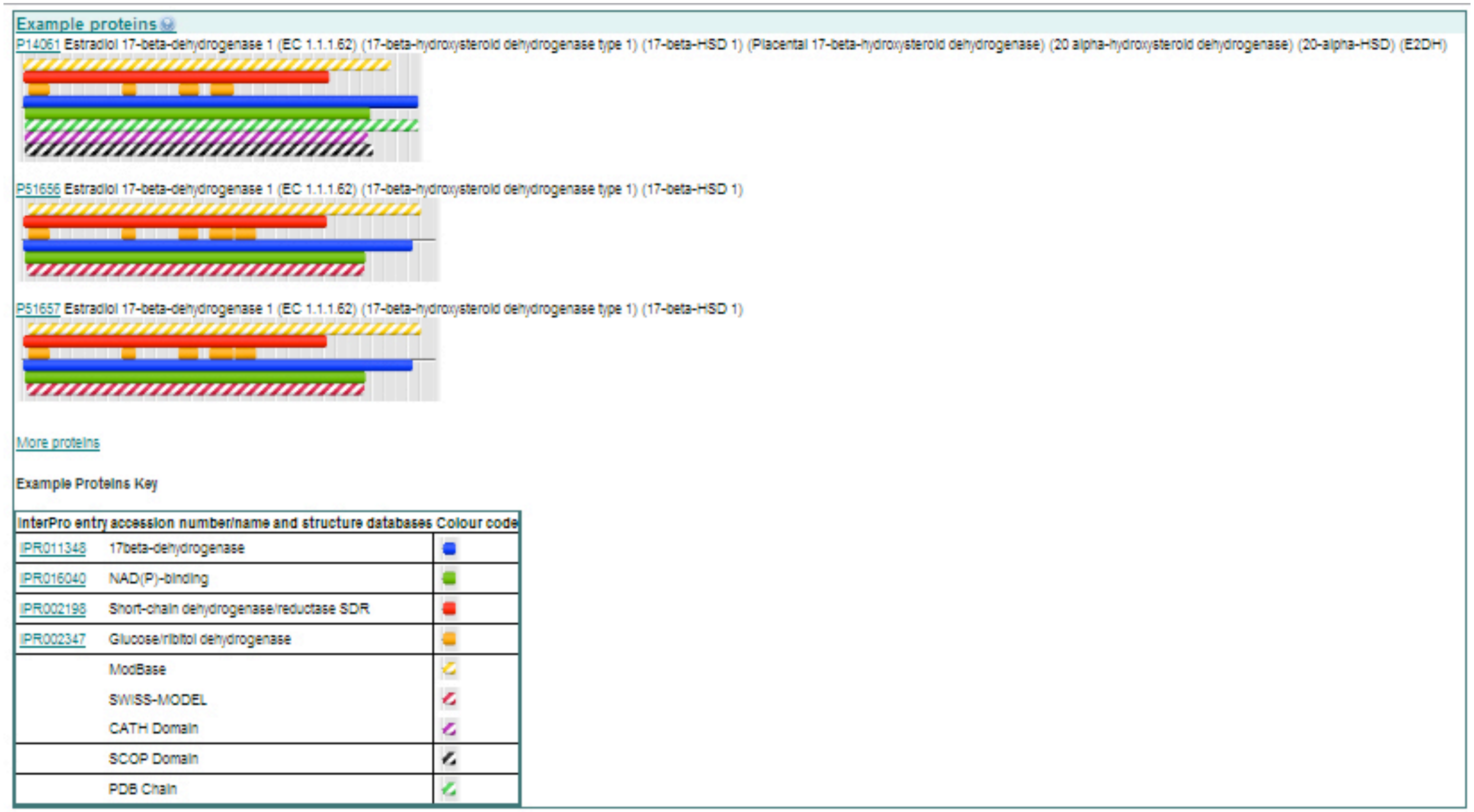
Protein matches	
UniProtKB Matches: 52 proteins	Overview: sorted by AC, sorted by name, of known structure, proteins with splice variants Detailed: sorted by AC, sorted by name, of known structure proteins with splice variants Table: For all matching proteins, of known structure Architectures Accession List
Accession	IPR011348 17beta_DHase
Type	Family
Signatures	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	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. 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 posseses the Tyr-X-X-X-Lys motif typical of short-chain dehydrogenases and forms a typical Rossman fold [3].
Structural links	CATH: 3.40.50.720.114 SCOP: c.2.1.2 PDB - click here
Database links	Enzyme: EC:1.1.1

InterPro - InterProScan Results



IPR011348	Numbers of overlapping proteins	Average numbers of overlapping amino acids
IPR002198 % Overlap: 100	0 52 27078	N/A
IPR016040 % Overlap: 100	0 52 117055	N/A
IPR002347 % Overlap: 92	4 48 23921	N/A

InterPro - InterProScan Results



Example proteins

P14061 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)

P51656 Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta-hydroxysteroid dehydrogenase type 1) (17-beta-HSD 1)

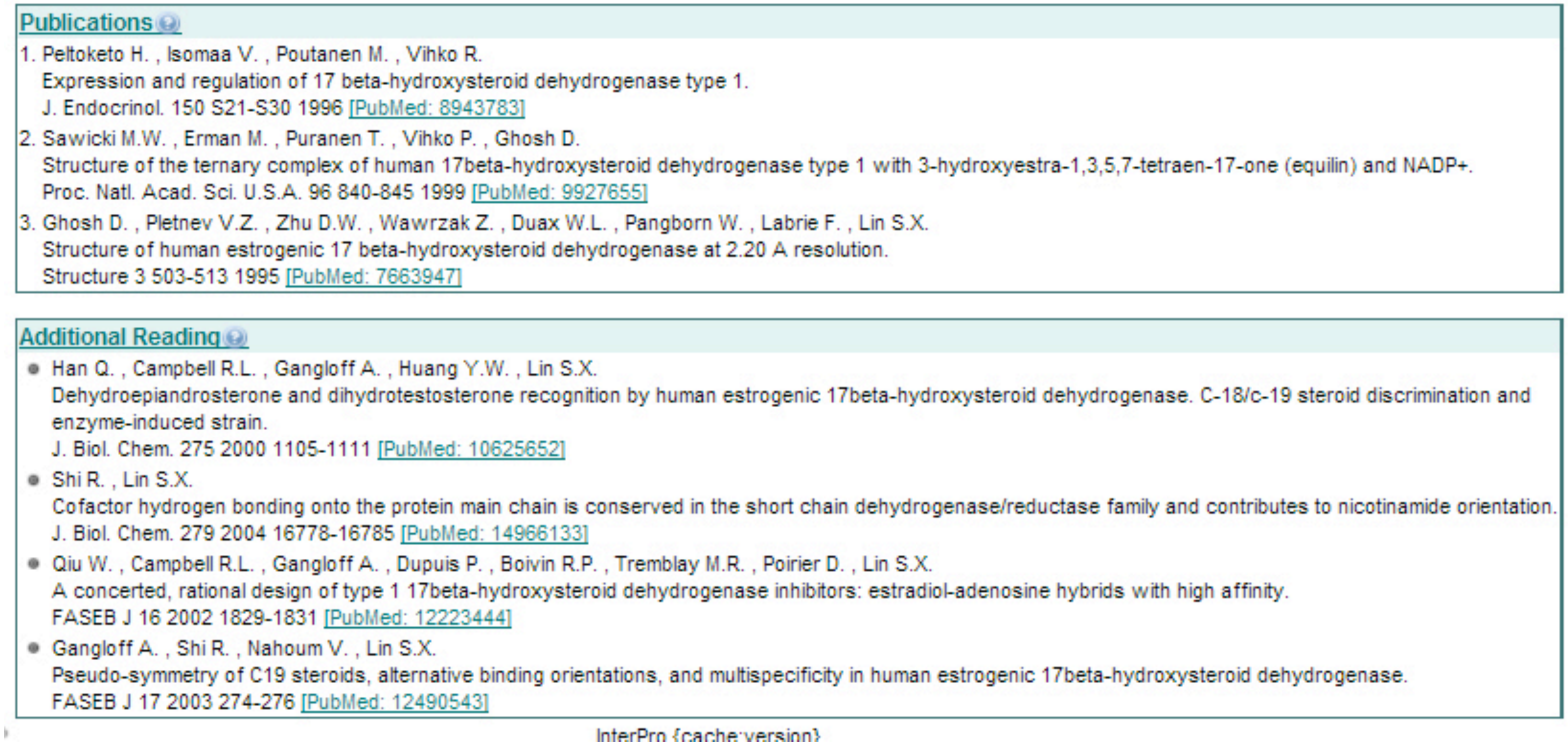
P51657 Estradiol 17-beta-dehydrogenase 1 (EC 1.1.1.62) (17-beta-hydroxysteroid dehydrogenase type 1) (17-beta-HSD 1)

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 A 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

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