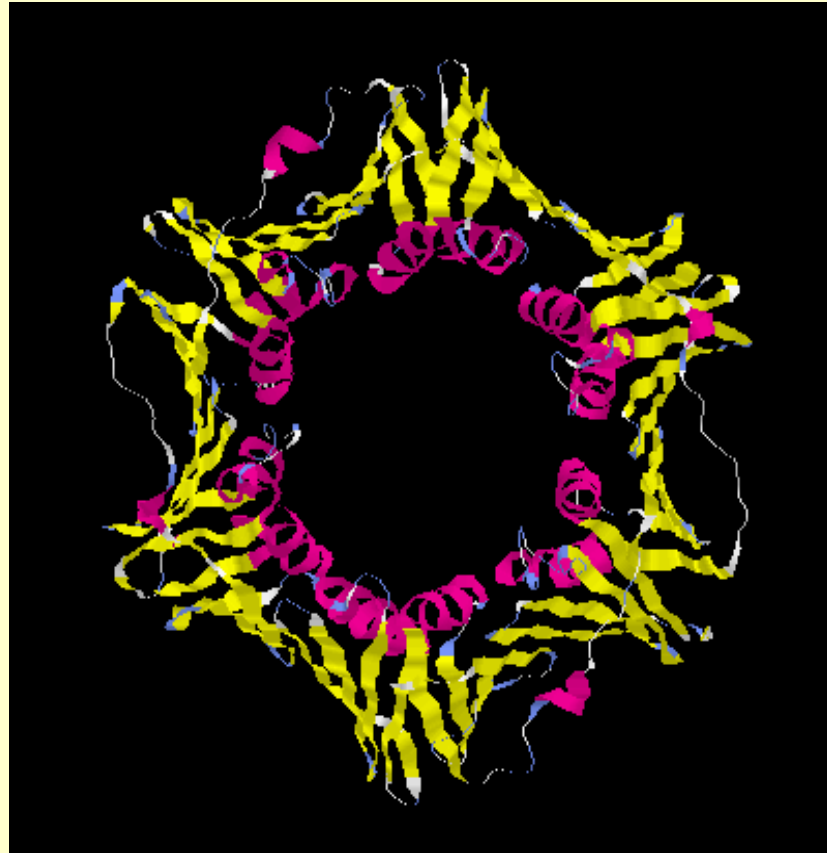


Protein Sequence Databases



Doug Brutlag
Professor Emeritus
Biochemistry & Medicine (by courtesy)

2nd Homework

- Transcripts and splice variants
 - NCBI - Evidence Viewer
 - Full Length RefSeq cDNAs
 - UniGene EST Assemblies
 - dbEST Assemblies
 - ENSEMBL Genes
 - Transcripts
 - Splice Variants
 - UCSC Genome Browser
 - RefSeq Tracts
 - Human EST tracks
 - Spliced EST tracks
 - UniGene Track

Protein Sequence and Classification Databases

- Evolution
 - Sequence Similarity or Distance
 - Clustered into protein Superfamilies and Families
 - Multiple sequence alignment profiles or Hidden Markov Models
 - Phylogenies
- Function
 - Sequence motifs (consensus sequences), Positions Specific Scoring Matrices (PSSMs), Profiles, Hidden Markov Models (HMMs)
 - Active sites
 - Binding sites for ligands, substrates, DNA
 - Protein-protein interaction sites
 - Enzymes, Structures, Receptors, Pores, Signals, Regulators
- Structure
 - Secondary Structure
 - Tertiary Folds – Hidden Markov Models
 - Clustered into protein domains, superfamilies and families
 - Quaternary Complexes - Interaction Databases

Protein Sequence Databases (Historical Order)

- National Biomedical Research Foundation (NBRF) => [Protein Identification Resource \(PIR\)](#)
 - Margaret Dayhoff
 - Atlas of Protein Sequences
 - Phylogenies, evolution, amino acid substitution matrices (PAM) and discovering active sites in enzymes
 - [PIR SF Evolutionary Family](#) , [iProClass](#) Functional site analysis and ontologies , [iProLink](#) to literature
 - [UniProt](#) - Universal Protein Resource
- [SwissProt \(EXPASY site\)](#)
 - Amos Bairoch
 - Manual curation and annotation
 - Highly cross-referenced
 - Many useful analytical tools ([EXPASY Tools](#))
 - [2D-PAGE](#) and Mass Spectrometry databases
 - [Prosites](#) functional motif database
 - [UniProt](#) - Universal Protein Resource
- TREMBL
 - Translation of mRNAs (RefSeq), UniGene, open reading frames (ORFs) and predicted genes from genomes
 - Automatic annotations
- EMBL => [EBI Protein databases](#)
 - [Clusters](#)
 - [Interpro](#) linked to domain and motif databases (CATH , PANTHER , PRINTs, PROSITE, pFAM, PIRSF, PRODOM, SCOP, SMART, SUPERFAMILY)
 - Intron-exon structure and links to ORFs, coding regions
 - [UniProt](#) - Universal Protein Resource
- [NCBI Protein Database](#)
 - Protein and nrPRO database SwissProt, PIR and translated genes/genomes
 - [Protein Clusters Database](#) (prokaryotic) and [COGS](#) and [KOGS](#)
 - Linked to coding regions and intron/exon structure
 - Linked to coding [SNPs](#) and variations databases
 - Linked to [MMDSB](#) structure database
 - Linked to [3D domains](#)
 - Linked to [CDD](#) Conserved Domain Database
- [UCSC Proteome Browser](#)



UniProt Sequence Databases

<http://www.uniprot.org/>

- UniProt Archive (UniParc)
 - Stable, comprehensive, non-redundant collection of all protein sequences ever published
 - Merged from PIR, SwissProt, TrEMBL, DDBJ/EMBL/GenBank proteins and proteomes, PDB, International Protein Index, RefSeq translations and other organism proteomes not yet in DDBJ/EMBL/GenBank
- UniProt Reference (UniRef)
 - Three non-redundant collections based on sequence similarity clusters
 - UniRef100 has all identical and identical overlapping subsequences merged into one entry in UniRef100
 - UniRef90 merges all protein sequence clusters with 90% sequence identity into a single entry.
 - UniRef50 merges all protein sequence clusters with 50% sequence identity into a single entry

UniProt Sequence Databases (cont.)

<http://www.uniprot.org/>

- UniProt Archive (UniProt)
 - UniProt/SwissProt
 - Manually curated highly-annotated sequences from SwissProt & PIRSF including descriptions, taxonomy, citations, GO terms, motifs, functional and structural classifications, residue specific annotations including variations.
 - Some automatic rule-based annotations including InterPro domains and motifs, PROSITE, PRINTS, Prodom, SMART, PFAM, PIRSF, Superfamily and TIGRFAMS classifications.
 - UniProt/TREMBL
 - Automatically translated from genomes including predicted as well as RefSeq genes.
 - Automated rule-based annotations.

RuleBase Protein Annotations

Fleischmann et al. 1998

- Learn rules from SwissProt
 - Find all SwissProt sequences with a specific motif, profile or HMM (from InterPro)
 - Examine all annotations (keywords, taxonomy, GO Terms, etc.) of SwissProt proteins that share a motif or a domain profile
 - Discover annotations common to all these proteins
 - Record rule that motif, profile or HMM=> annotation
- Apply rules to TREMBL
 - Find TREMBL sequences which have the same domain or motif as in RuleBase
 - Apply common annotations to TREMBL
 - Flag annotations as “BY SIMILARITY”
- Generates many false annotations
 - Due to low specificity domain or motif profiles
 - Large size of TREMBL compared to SwissProt training set
 - Used *e*MOTIFs (higher specificity motifs < 1 error per 10⁹ amino acids) to confirm patterns
 - Build rules based on **multiple motifs (Spearmint)**
 - Introduced confidence thresholds

Serine Protease Catalytic Triad



Three conserved motifs containing the three residues of the catalytic triad in the serine protease subtilisin BPN of *Bacillus amyloliquefaciens*. The first conserved motif is shown in red (residues 135–146; Prosite PS00136) and contains the active site aspartate (red spheres). The second conserved motif is shown in blue (residues 171–181; Prosite PS00137) and contains the active site histidine (blue spheres). The third motif is shown in yellow (residues 326–336; Prosite PS00138) and contains the active site serine (yellow spheres).

Filtering erroneous annotations with negative rules (Xanthippe) Wieser et al. 2004

- Examine taxonomic and other “core” annotations that lack a more specific annotation in SwissProt
 - All bacterial proteins lack “nuclear protein” annotation
 - If RuleBase assigns “nuclear protein” to a bacterial protein in TrEMBL then remove it
 - >4,000 exclusion rules based just on taxonomic class
 - Build exclusion rules based on decision trees involving multiple motif annotations
- Some exclusion rules in error due to database bias
 - All venomous snakes lacked ATP binding proteins!
 - The only SwissProt proteins from venomous snakes came from venom, not from the rest of the organism!

Performance of Xanthippe (Tested by Crossvalidation on SwissProt)

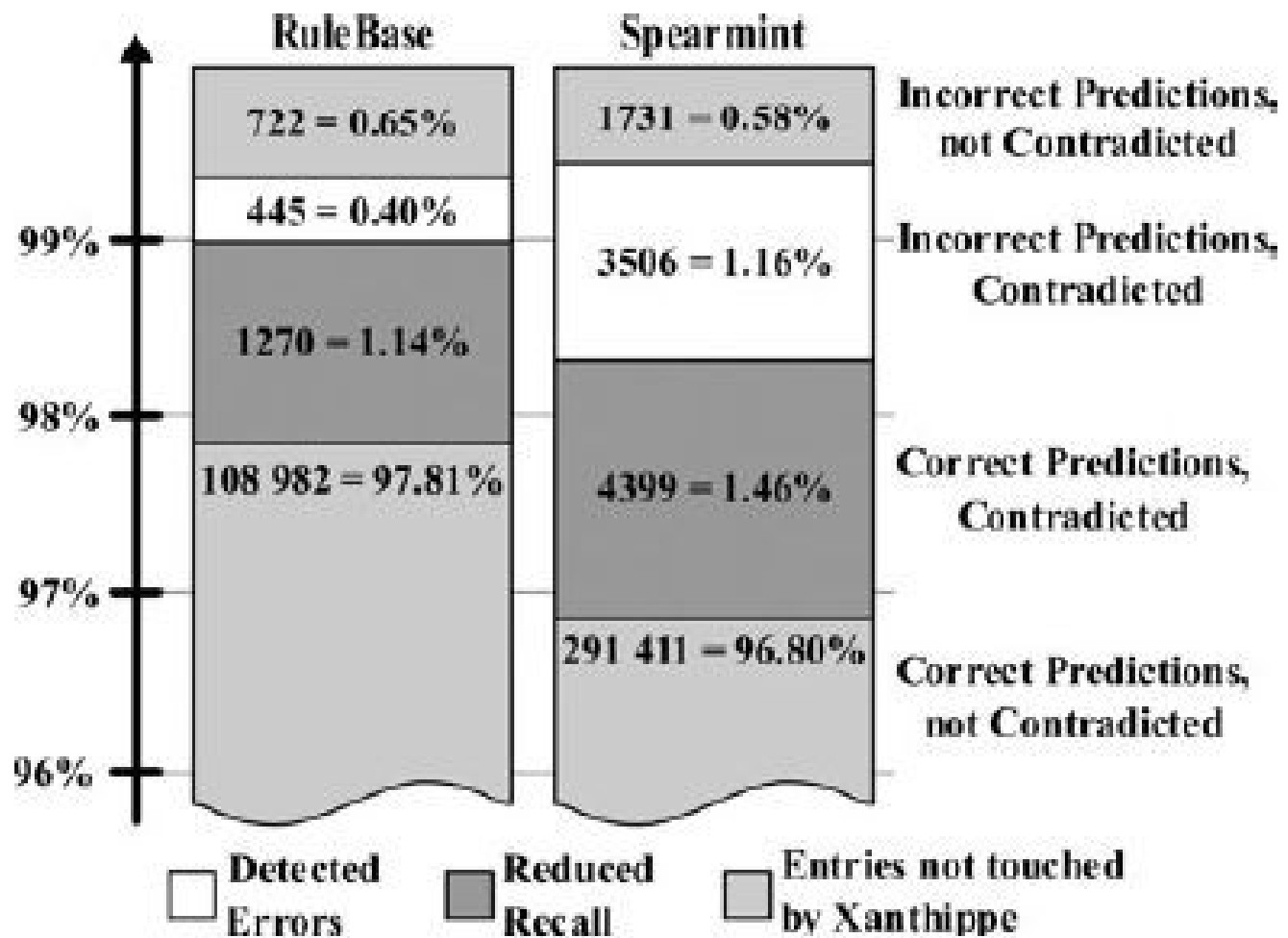


Fig. 3. Performance of Xanthippe exclusion trees on keyword predictions from RuleBase and Spearmint.



ExPASy Proteomics Server

<http://www.expasy.org/>



Swiss Institute of Bioinformatics



Search for

ExPASy Proteomics Server

[Databases](#) [Tools](#) [Services](#) [Mirrors](#) [About](#) [Contact](#)

You are here: [ExPASy CH](#)

The ExPASy (**Expert Protein Analysis System**) proteomics server of the [Swiss Institute of Bioinformatics](#) (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE ([Disclaimer](#) / [References](#) / [Linking to ExPASy](#)).

Databases

[UniProtKB](#), [PROSITE](#), [HAMAP](#), [SwissVar](#),
[ViralZone](#), [SWISS-MODEL Repository](#),
[SWISS-2DPAGE](#), [World-2DPAGE](#)
[Repository](#), [MIAPEGelDB](#), [ENZYME](#),
[GlycoSuiteDB](#), [UniPathway](#)
[details] [full list]

Education & services

[Downloads](#), [Protein Spotlight](#),
[Protéines à la «Une»](#), [e-proxemis](#),
[Bioinformatics core facility for Proteomics](#)
[full list]

Tools & Software

[Proteomics tools](#), [Blast](#), [ScanProsite](#),
[Melanie](#), [MSight](#), [Make2D-DB](#), [SWISS-](#)
[MODEL](#), [Swiss-PdbViewer](#)
[full list]

Documentation

[What's New?](#), [E-mail alerts](#), [UniProtKB](#)
[documentation](#), [How to link to ExPASy](#),
[Advanced search](#)
[full list]

Latest News

Protein Spotlight - Dec 21, 2009 **String of intrusion**

When I was little, I used to wear small cotton shirts that were knitted by my grandmother. So? Well, onto them she sewed tiny nacre buttons you could never get hold of and which mesmerized me because of the different colours that shone off them depending on how you oriented them in the light. [more]

World-2DPAGE - Oct 23, 2009

New data uploaded into the [World-2DPAGE Repository](#). Currently, 113 maps for 16 species are available from the [World-2DPAGE Portal](#).

[more news] [SIB news]





ExPASy Proteomic Tools

<http://www.expasy.org/tools/>



Swiss Institute of
Bioinformatics




ExPASy Proteomics Server

Search for

[Databases](#) [Tools](#) [Services](#) [Mirrors](#) [About](#) [Contact](#)

You are here: [ExPASy CH](#) > [Tools](#)





ExPASy Proteomics tools

The tools marked by  are local to the ExPASy server. The remaining tools are developed and hosted on other servers.



[\[Protein identification and characterization\]](#) [\[Other proteomics tools\]](#) [\[DNA -> Protein\]](#) [\[Similarity searches\]](#) [\[Pattern and profile searches\]](#) [\[Post-translational modification prediction\]](#) [\[Topology prediction\]](#)
[\[Primary structure analysis\]](#) [\[Secondary structure prediction\]](#) [\[Tertiary structure\]](#) [\[Sequence alignment\]](#) [\[Phylogenetic analysis\]](#) [\[Biological text analysis\]](#)

Protein identification and characterization

Identification and characterization with peptide mass fingerprinting data

- [Aldente](#)  - Identify proteins with peptide mass fingerprinting data. A new, fast and powerful tool that takes advantage of Hough transformation for spectra recalibration and outlier exclusion. [Download the stand-alone version](#)
- [FindMod](#)  - Predict potential protein post-translational modifications and potential single amino acid substitutions in peptides. Experimentally measured peptide masses are compared with the theoretical peptides calculated from a specified Swiss-Prot entry or from a user-entered sequence, and mass differences are used to better characterize the protein of interest.
- [FindPept](#)  - Identify peptides that result from unspecific cleavage of proteins from their experimental masses, taking into account artefactual chemical modifications, post-translational modifications (PTM) and protease autolytic cleavage
- [GlycoMod](#)  - Predict possible oligosaccharide structures that occur on proteins from their experimentally determined masses (can be used for free or derivatized oligosaccharides and for glycopeptides)
- [Mascot](#) - Peptide mass fingerprint from Matrix Science Ltd., London
- [PepMAPPER](#) - Peptide mass fingerprinting tool from UMIST, UK
- [ProFound](#) - Search known protein sequences with peptide mass information from Rockefeller and NY Universities [or from [Genomic Solutions](#)]
- [ProteinProspector](#) - UCSF tools for peptide masses data (MS-Fit, MS-Pattern, MS-Digest, etc.)

Identification and characterization with MS/MS data

- [Popitam](#)  - Identification and characterization tool for peptides with unexpected modifications (e.g. post-translational modifications or mutations) by tandem mass spectrometry
- [Phenyx](#)  - Protein and peptide identification/characterization from MS/MS data from GeneBio, Switzerland
- [Mascot](#) - Sequence query and MS/MS ion search from Matrix Science Ltd., London



UniProt Database

<http://www.uniprot.org/>



The screenshot shows the UniProt search interface. At the top left is the UniProt logo. Below it, there's a search bar with a dropdown menu set to 'Protein Knowledgebase (UniProtKB)'. The search query is 'PCNA'. To the right of the search bar are buttons for 'Search', 'Clear', and 'Fields »'. Below the search bar is a navigation menu with buttons for 'ID Mapping', 'Retrieve', 'Align', 'Blast', and 'Search'.

WELCOME

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

What we provide

UniProtKB	<p>Protein knowledgebase, consists of two sections:</p> <ul style="list-style-type: none"> ★ Swiss-Prot, which is manually annotated and reviewed. ★ TrEMBL, which is automatically annotated and is not reviewed. <p>Includes Complete Proteome Sets.</p>
UniRef	Sequence clusters, used to speed up similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations , taxonomy , keywords and more .



NEWS

UniProt release 15.12 – Dec 15, 2009

Through the Looking-Glass - Cross-references to ArachnoServer and InParanoid

- > [Statistics for UniProtKB: Swiss-Prot · TrEMBL](#)
- > [Forthcoming changes](#)
- > [News archives](#)

SITE TOUR



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

PROTEIN SPOTLIGHT

String of intrusion January 2010

When I was little, I used to wear little cotton shirts that were knitted by my grandmother...

UniProt PCNA Search

<http://www.uniprot.org/uniprot/?query=PCNA&sort=score>

UniProt > UniProtKB Downloads · Contact · Documentation/Help

Search Blast Align Retrieve ID Mapping *

Search in: Protein Knowledgebase (UniProtKB) Query: PCNA Search Clear Fields »

1 - 25 of 1,035 results for PCNA in UniProtKB sorted by score descending

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50% | Customize display Download...

› Show only reviewed (UniProtKB/Swiss-Prot) or unreviewed (UniProtKB/TrEMBL) entries

› Restrict term "pcna" to protein family, gene name, gene ontology, protein name, strain, taxonomy, web resource

Page 1 of 42 | Next »

Accession	Entry name	Status	Protein names	Gene names	Organism	Length
P12004	PCNA_HUMAN	★	Proliferating cell nuclear antigen	PCNA	Homo sapiens (Human)	261
P17918	PCNA_MOUSE	★	Proliferating cell nuclear antigen	Pcna	Mus musculus (Mouse)	261
Q9DEA3	PCNA_CHICK	★	Proliferating cell nuclear antigen	PCNA	Gallus gallus (Chicken)	262
Q9PTP1	PCNA_DANRE	★	Proliferating cell nuclear antigen	pcna	Danio rerio (Zebrafish) (Brachydanio rerio)	260
P04961	PCNA_RAT	★	Proliferating cell nuclear antigen	Pcna	Rattus norvegicus (Rat)	261
P17917	PCNA_DROME	★	Proliferating cell nuclear antigen	mus209 (PCNA) (CG9193)	Drosophila melanogaster (Fruit fly)	260
Q9UWR9	PCNA_THEFM	★	DNA polymerase sliding clamp	pcn (pcnA)	Thermococcus fumicolans	249
Q9M7Q7	PCNA1_ARATH	★	Proliferating cellular nuclear antigen 1	PCNA (PCNA1) (At1g07370) (F22G5.29)	Arabidopsis thaliana (Mouse-ear cress)	263
P11038	PCNA_NPVAC	★	Probable DNA polymerase sliding clamp	PCNA (ETL)	Autographa californica nuclear polyhedrosis virus (AcMNPV)	256
		★	DNA polymerase sliding	pcnA (pcnA-1)		

UniProt PCNA Search

<http://www.uniprot.org/uniprot/?query=name%3Apcna&sort=score>

UniProt > UniProtKB Downloads · Contact · Documentation/Help

Search Blast Align Retrieve ID Mapping *

Search in Query

Protein Knowledgebase (UniProtKB) [Fields »](#)

1 - 25 of 161 results for **name:pcna** in UniProtKB sorted by **score** descending

[Browse by taxonomy, keyword, gene ontology, enzyme class or pathway](#) | [Reduce sequence redundancy to 100%, 90% or 50%](#) | [Customize display](#) [Download...](#)

• Show only **reviewed** (UniProtKB/Swiss-Prot) or **unreviewed** entries

Page 1 of 7 | [Next »](#)

<input type="checkbox"/>	Accession	Entry name	Status	Protein names	Gene names	Organism	Length
<input type="checkbox"/>	P15873	PCNA_YEAST	★	Proliferating cell nuclear antigen	POL30 (YBR088C) (YBR0811)	Saccharomyces cerevisiae (Baker's yeast)	258
<input type="checkbox"/>	P12004	PCNA_HUMAN	★	Proliferating cell nuclear antigen	PCNA	Homo sapiens (Human)	261
<input type="checkbox"/>	Q9M7Q7	PCNA1_ARATH	★	Proliferating cellular nuclear antigen 1	PCNA (PCNA1) (At1g07370) (F22G5.29)	Arabidopsis thaliana (Mouse-ear cress)	263
<input type="checkbox"/>	Q9DEA3	PCNA_CHICK	★	Proliferating cell nuclear antigen	PCNA	Gallus gallus (Chicken)	262
<input type="checkbox"/>	Q9PTP1	PCNA_DANRE	★	Proliferating cell nuclear antigen	pcna	Danio rerio (Zebrafish) (Brachydanio rerio)	260
<input type="checkbox"/>	P17918	PCNA_MOUSE	★	Proliferating cell nuclear antigen	Pcna	Mus musculus (Mouse)	261
<input type="checkbox"/>	P04961	PCNA_RAT	★	Proliferating cell nuclear antigen	Pcna	Rattus norvegicus (Rat)	261
<input type="checkbox"/>	Q03392	PCNA_SCHPO	★	Proliferating cell nuclear antigen	pcn1 (pcn) (SPBC16D10.09)	Schizosaccharomyces pombe (Fission yeast)	260
<input type="checkbox"/>	Q9ZW35	PCNA2_ARATH	★	Proliferating cell nuclear antigen 2	At2g29570 (F16P2.5)	Arabidopsis thaliana (Mouse-ear cress)	264
<input type="checkbox"/>	O02115	PCNA_CAEEL	★	Proliferating cell nuclear antigen	pcn-1 (W03D2.4)	Caenorhabditis elegans	229

UniProt PCNA Reviewed Search

<http://www.uniprot.org/uniprot/?query=name%3Apcna+AND+reviewed%3Ayes&sort=score>

UniProtKB Downloads · Contact · Documentation/Help

Search Blast Align Retrieve ID Mapping *

Search in Query

Protein Knowledgebase (UniProtKB) [Fields »](#)

1 - 25 of 102 results for name:pcna AND reviewed:yes in UniProtKB sorted by score descending

[Browse by taxonomy, keyword, gene ontology, enzyme class or pathway](#) | [Reduce sequence redundancy to 100%, 90% or 50%](#) | [Customize display](#)

Page 1 of 5 | [Next »](#)

Accession	Entry name	Status	Protein names	Gene names	Organism	Length
<input type="checkbox"/> P15873	PCNA_YEAST	★	Proliferating cell nuclear antigen	POL30 (YBR088C) (YBR0811)	Saccharomyces cerevisiae (Baker's yeast)	258
<input type="checkbox"/> P12004	PCNA_HUMAN	★	Proliferating cell nuclear antigen	PCNA	Homo sapiens (Human)	261
<input type="checkbox"/> Q9M7Q7	PCNA1_ARATH	★	Proliferating cellular nuclear antigen 1	PCNA (PCNA1) (At1g07370) (F22G5.29)	Arabidopsis thaliana (Mouse-ear cress)	263
<input type="checkbox"/> Q9DEA3	PCNA_CHICK	★	Proliferating cell nuclear antigen	PCNA	Gallus gallus (Chicken)	262
<input type="checkbox"/> Q9PTP1	PCNA_DANRE	★	Proliferating cell nuclear antigen	pcna	Danio rerio (Zebrafish) (Brachydanio rerio)	260
<input type="checkbox"/> P17918	PCNA_MOUSE	★	Proliferating cell nuclear antigen	Pcna	Mus musculus (Mouse)	261
<input type="checkbox"/> P04961	PCNA_RAT	★	Proliferating cell nuclear antigen	Pcna	Rattus norvegicus (Rat)	261
<input type="checkbox"/> Q03392	PCNA_SCHPO	★	Proliferating cell nuclear antigen	pcn1 (pcn) (SPBC16D10.09)	Schizosaccharomyces pombe (Fission yeast)	260
<input type="checkbox"/> Q9ZW35	PCNA2_ARATH	★	Proliferating cell nuclear antigen 2	At2g29570 (F16P2.5)	Arabidopsis thaliana (Mouse-ear cress)	264
<input type="checkbox"/> O02115	PCNA_CAEEL	★	Proliferating cell nuclear antigen	pcn-1 (W03D2.4)	Caenorhabditis elegans	229
<input type="checkbox"/> Q54K47	PCNA_DICDI	★	Proliferating cell nuclear antigen	pcna (DDB_G0287607)	Dictyostelium discoideum (Slime mold)	258

Sequence

```
>P12004
MFEARLVQGSILKKVLEALKDLINAECDWISSGVNLQSMDSHVSVLVQLTRSEGFDTY
RCDRLNLAGVNLTSMSKILKCAGNEDIITLRAEDNADTLALVFEAPNQEKVSDYEMKLM
LDVEQLGIPEQEYSCVVKMPSGFEFARICRDLSHIGDAVVISCAKDGVKFSASGELGNGI
KLSQTSNVDKKEEAVTIEMNEPQLTFALRYLNFFTKATPLSSTVTLMSADVPLVVEYK
IADMGHLYYLAPKIEDEEGS
```

Blast

Clear

« Options

Blast tips

You can enter:

- a protein or nucleotide sequence
- a valid UniProt identifier, for example:
P00750 or A4_HUMAN or UPI00000000

Database

UniProtKB

Threshold

10

Matrix

BLOSUM-62

Filtering

None

Gapped

yes

Hits

250

ID Mapping

Retrieve

Align

Blast

Search

1 - 25 of 250 results for **Blast Search** of in **UniRef90**

[Graphical Overview](#) | [Customize display](#)

[TEXT](#) [XML](#) [TA](#)

[Job information](#) · [Blast parameters](#)

Page 1 of 10

UniRef Id	Representative Member SHOW ALL MEMBERS						
	Accession	Status	Protein names [^]	Organism [^]	Local alignment	Length	
<input type="checkbox"/> UniRef90_P12004	<input type="checkbox"/> P12004	★	Proliferating cell nuclear antigen	Homo sapiens (Human)			
<input type="checkbox"/> UniRef90_UPI00005A449C	<input type="checkbox"/> UPI00005A449C		PREDICTED: similar to proliferating cell nuclear antigen	Canis familiaris			
<input type="checkbox"/> UniRef90_Q9DEA3	<input type="checkbox"/> Q9DEA3	★	Proliferating cell nuclear antigen	Gallus gallus (Chicken)			
<input type="checkbox"/> UniRef90_Q9PTP1	<input type="checkbox"/> Q9PTP1	★	Proliferating cell nuclear antigen	Danio rerio (Zebrafish) (Brachydanio rerio)			
<input type="checkbox"/> UniRef90_P18248	<input type="checkbox"/> P18248	★	Proliferating cell nuclear antigen	Xenopus laevis (African clawed frog)			
<input type="checkbox"/> UniRef90_C3ZRF2	<input type="checkbox"/>	★					
<input type="checkbox"/> UniRef90_UPI00006A37AA	<input type="checkbox"/> UPI00006A37AA		PREDICTED: similar to proliferating cell nuclear antigen	Ciona intestinalis			



ExPASy Proteomics Server

<http://www.expasy.org/>



Swiss Institute of Bioinformatics



Search for

ExPASy Proteomics Server

[Databases](#) [Tools](#) [Services](#) [Mirrors](#) [About](#) [Contact](#)

You are here: [ExPASy CH](#)

The ExPASy (**Expert Protein Analysis System**) proteomics server of the [Swiss Institute of Bioinformatics](#) (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE ([Disclaimer](#) / [References](#) / [Linking to ExPASy](#)).

Databases

[UniProtKB](#), [PROSITE](#), [HAMAP](#), [SwissVar](#),
[ViralZone](#), [SWISS-MODEL Repository](#),
[SWISS-2DPAGE](#), [World-2DPAGE](#)
[Repository](#), [MIAPEGelDB](#), [ENZYME](#),
[GlycoSuiteDB](#), [UniPathway](#)
[details] [full list]

Education & services

[Downloads](#), [Protein Spotlight](#),
[Protéines à la «Une»](#), [e-proxemis](#),
[Bioinformatics core facility for Proteomics](#)
[full list]

Tools & Software

[Proteomics tools](#), [Blast](#), [ScanProsite](#),
[Melanie](#), [MSight](#), [Make2D-DB](#), [SWISS-](#)
[MODEL](#), [Swiss-PdbViewer](#)
[full list]

Documentation

[What's New?](#), [E-mail alerts](#), [UniProtKB](#)
[documentation](#), [How to link to ExPASy](#),
[Advanced search](#)
[full list]

Latest News

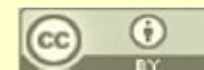
Protein Spotlight - Dec 21, 2009 **String of intrusion**

When I was little, I used to wear small cotton shirts that were knitted by my grandmother. So? Well, onto them she sewed tiny nacre buttons you could never get hold of and which mesmerized me because of the different colours that shone off them depending on how you oriented them in the light. [more]

World-2DPAGE - Oct 23, 2009

New data uploaded into the [World-2DPAGE Repository](#). Currently, 113 maps for 16 species are available from the [World-2DPAGE Portal](#).

[more news] [SIB news]





Database of protein domains, families and functional sites

PROSITE consists of [documentation entries](#) describing protein domains, families and functional sites as well as associated [patterns](#) and [profiles](#) to identify them [[More details](#) / [References](#) / [Disclaimer](#) / [Commercial users](#)].

PROSITE is complemented by **ProRule**, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [[More details](#)].

Release 20.59, of 19-Jan-2010 (1567 documentation entries, 1308 patterns, 873 profiles and 874 ProRule)

PROSITE access

e.g: PDOC00022, PS50089, SH3, zinc finger

add wildcard ^{!*1}

Browse:

- [by documentation entry](#)
- [by ProRule description](#)
- [by taxonomic scope](#)
- [by number of positive hit](#)

PROSITE tools

Scan a sequence against PROSITE patterns and profiles - quick scan

(Output includes graphical view and feature detection)



Enter your sequence or a [UniProtKB \(Swiss-Prot or TrEMBL\)](#) ID or AC
[[help](#)]:

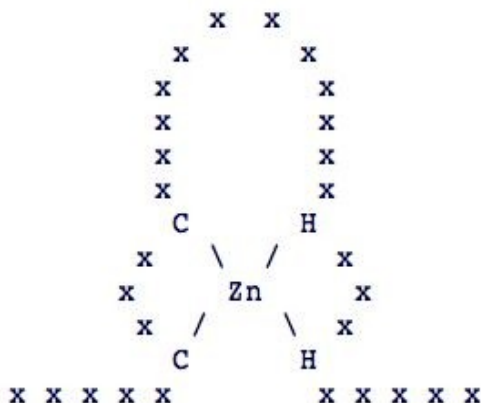
- **ScanProsite** - advanced scan
- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** ^{new} - allows to generate custom domain figures.



Zinc finger C2H2-type domain signature and profile

Description:

'Zinc finger' domains [1,2,3,4,5] are nucleic acid-binding protein structures first identified in the *Xenopus* transcription factor TFIIIA. These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino-acid residues. There are two cysteine or histidine residues at both extremities of the domain, which are involved in the tetrahedral coordination of a zinc atom. It has been proposed that such a domain interacts with about five nucleotides. A schematic representation of a zinc finger domain is shown below:



Many classes of zinc fingers are characterized according to the number and positions of the histidine and cysteine residues involved in the zinc atom coordination. In the first class to be characterized, called C2H2, the first pair of zinc coordinating residues are cysteines, while the second pair are histidines. A number of experimental reports have demonstrated the zinc-dependent DNA or RNA binding property of some members of this class.

Some of the proteins known to include C2H2-type zinc fingers are listed below. We have indicated, between brackets, the number of zinc finger regions found in each of these proteins; a '+' symbol indicates that only partial sequence data is available and that additional finger domains may be present.

- *Saccharomyces cerevisiae*: ACE2 (3), ADR1 (2), AZF1 (4), FZF1 (5), MIG1 (2), MSN2 (2), MSN4 (2), RGM1 (2), RIM1 (3), RME1 (3), SFP1 (2), SSL1 (1), STP1 (3), SWI5 (3), VAC1 (1) and ZMS1 (2).
- *Emericella nidulans*: brlA (2), creA (2).
- *Drosophila*: AEF-1 (4), Cf2 (7), ci-D (5), Disconnected (2), Escargot (5), Glass (5), Hunchback (6), Kruppel (5), Kruppel-H (4+), Odd-skipped (4), Odd-paired (4), Pep (3), Snail (5), Spalt-major (7), Serendipity locus β (6), delta (7), h-1 (8), Suppressor of hairy wing su(Hw) (12), Suppressor of variegation suvar(3)7 (5), Teashirt (3) and Tramtrack (2).
- *Xenopus*: transcription factor TFIIIA (9), p43 from RNP particle (9), Xfin (37 !!), Xsna (5), gastrula XlCGF5.1 to XlCGF71.1 (from 4+ to 11+), Oocyte XlCOF2 to XlCOF22 (from 7 to 12).
- Mammalian: hoesenuelin (6), BCL-6 (4), AZP (6), onchroid kruppel like transcription factor (2), transcription factors Sn1 (2), Sn2 (2), Sn3 (2), Sn4 (2), Sn5 (2), Sn6 (2), Sn7 (2), Sn8 (2), Sn9 (2), Sn10 (2), Sn11 (2), Sn12 (2), Sn13 (2), Sn14 (2), Sn15 (2), Sn16 (2), Sn17 (2), Sn18 (2), Sn19 (2), Sn20 (2), Sn21 (2), Sn22 (2), Sn23 (2), Sn24 (2), Sn25 (2), Sn26 (2), Sn27 (2), Sn28 (2), Sn29 (2), Sn30 (2), Sn31 (2), Sn32 (2), Sn33 (2), Sn34 (2), Sn35 (2), Sn36 (2), Sn37 (2), Sn38 (2), Sn39 (2), Sn40 (2), Sn41 (2), Sn42 (2), Sn43 (2), Sn44 (2), Sn45 (2), Sn46 (2), Sn47 (2), Sn48 (2), Sn49 (2), Sn50 (2), Sn51 (2), Sn52 (2), Sn53 (2), Sn54 (2), Sn55 (2), Sn56 (2), Sn57 (2), Sn58 (2), Sn59 (2), Sn60 (2), Sn61 (2), Sn62 (2), Sn63 (2), Sn64 (2), Sn65 (2), Sn66 (2), Sn67 (2), Sn68 (2), Sn69 (2), Sn70 (2), Sn71 (2), Sn72 (2), Sn73 (2), Sn74 (2), Sn75 (2), Sn76 (2), Sn77 (2), Sn78 (2), Sn79 (2), Sn80 (2), Sn81 (2), Sn82 (2), Sn83 (2), Sn84 (2), Sn85 (2), Sn86 (2), Sn87 (2), Sn88 (2), Sn89 (2), Sn90 (2), Sn91 (2), Sn92 (2), Sn93 (2), Sn94 (2), Sn95 (2), Sn96 (2), Sn97 (2), Sn98 (2), Sn99 (2), Sn100 (2).

Prosite Zinc Finger C2H2 Profile Entry

<http://www.expasy.org/cgi-bin/nicedoc.pl?PDOC00028>

Technical section:

PROSITE methods (with tools and information) covered by this documentation:

ZINC_FINGER_C2H2_2, PS50157; Zinc finger C2H2 type domain profile (MATRIX)

Sequences known to belong to this class detected by the profile: ALL

Other sequence(s) detected in Swiss-Prot: 2.

- Domain architecture view of Swiss-Prot proteins matching PS50157



- Retrieve an alignment of Swiss-Prot true positive hits:
Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic tree view of all Swiss-Prot/TrEMBL entries matching PS50157
- Retrieve a list of all Swiss-Prot/TrEMBL entries matching PS50157
- Scan Swiss-Prot/TrEMBL entries against PS50157
- view ligand binding statistics

Matching PDB structures: 1A1F 1A1G 1A1H 1A1I ... [ALL]

Prosite Zinc Finger C2H2 Profile Entry

<http://www.expasy.org/prosite/PS50157>



Entry: **PS50157**

[Home](#) [ScanProsite](#) [ProRule](#) [Documents](#) [Downloads](#) [Links](#) [Fu](#)

General information about the entry

Entry name	ZINC_FINGER_C2H2_2
Accession number	PS50157
Entry type	MATRIX
Date	DEC-2001 (CREATED); DEC-2001 (DATA UPDATE); DEC-2009 (INFO UPDATE).
PROSITE Documentation	PDOC00028
Associated ProRule	PRU00042

Name and characterization of the entry

Description	Zinc finger C2H2 type domain profile.
Matrix / Profile	<pre> /GENERAL_SPEC: ALPHABET='ABCDEFGHIJKLMNPQRSTUVWXYZ'; LENGTH=28; /DISJOINT: DEFINITION=PROTECT; N1=3; N2=26; /NORMALIZATION: MODE=1; FUNCTION=LINEAR; R1=-0.6689; R2=0.02078310; TEXT='-LogE'; /CUT_OFF: LEVEL=0; SCORE=441; N_SCORE=8.5; MODE=1; TEXT='!'; /CUT_OFF: LEVEL=-1; SCORE=344; N_SCORE=6.5; MODE=1; TEXT='?'; /DEFAULT: D=-20; I=-20; B1=-50; E1=-50; MI=-105; MD=-105; IM=-105; DM=-105; A B C D E F G H I K L M N P Q R S T V W Y Z /I: B1=0; BI=-105; BD=-105; /M: SY='Y'; M=-19,-21,-24,-25,-21, 39,-28, 10, -2,-17, 2, 0,-17,-28,-19,-13,-18,-10, -7, 12, 48,-21; /M: SY='K'; M= -4, -5,-23, -5, 6,-20,-18, -9,-14, 7,-15, -7, -5, -9, 3, 2, -2, -2, -9,-25,-12, 4; /M: SY='C'; M=-10,-20,118,-30,-30,-20,-30,-30,-30,-30,-20,-20,-20,-40,-30,-30,-10,-10,-10,-50,-30,-30; /M: SY='E'; M= -5, 3,-24, 3, 6,-22,-11, -6,-20, 1,-21,-14, 4, -1, 1, -3, 5, 2,-18,-29,-15, 3; /I: I=-12; MI=0; MD=-30; IM=0; DM=-30; /M: SY='E'; M= -9, -2,-26, 1, 14,-18,-17, -4,-13, -1,-11, -8, -5,-12, 4, -5, -5, -8,-12,-24, -9, 8; /M: SY='C'; M=-10,-20,119,-30,-30,-20,-30,-30,-30,-20,-20,-20,-40,-30,-30,-10,-10,-10,-50,-29,-30; /M: SY='G'; M= -3, -1,-28, -1, -7,-28, 36,-11,-33,-11,-27,-18, 4,-15,-10,-12, 1,-13,-27,-24,-23, -9; /M: SY='K'; M=-10, -2,-28, -3, 8,-25,-19, -7,-26, 36,-24, -8, -1,-12, 10, 27, -9, -9,-18,-19, -8, 8; /M: SY='A'; M= 8, -7, -9,-11, -7,-17, -7,-14,-16, -6,-16,-11, -4,-15, -6, -5, 8, 4, -7,-27,-15, -7; /M: SY='F'; M=-19,-29,-19,-37,-28, 71,-29,-17, 0,-28, 9, 0,-20,-30,-36,-19,-19, -9, -1, 9, 31,-28; /M: SY='S'; M= 0, -5,-17, -9, -6,-16,-11,-10,-14, -3,-16,-10, 0,-12, -4, 0, 8, 7, -8,-27,-12, -6; /M: SY='R'; M=-10, -3,-20, -4, 0,-18,-17, 2,-19, 3,-16, -8, 0,-17, 8, 9, -1, -3,-17,-19, -5, 3; /M: SY='R'; M= -4, -4,-22, -6, 0,-19,-13, -5,-18, 7,-18, -9, 1,-10, 2, 8, 2, -2,-14,-25,-11, 0; /M: SY='S'; M= 2, -1,-16, -1, -1,-18, -4, -6,-19, -7,-22,-14, 4,-12, -2, -7, 16, 7,-13,-29,-12, -2; /M: SY='N'; M= -5, 5,-20, 1, 0,-18,-10, 8,-18, -5,-18,-11, 9,-16, 1, -4, 4, -1,-17,-27, -7, -1; /M: SY='L'; M=-11,-29,-20,-30,-20, 12,-29,-19, 17,-27, 43, 18,-28,-29,-20,-18,-28,-10, 9,-18, 2,-20; /M: SY='R'; M= -6, -6,-22,-10, -4,-15,-20, -8, -7, 2, -9, -3, -2,-16, 0, 3, -3, 0, -6,-24, -8, -3; </pre>

Prosite Zinc Finger Pattern

<http://www.expasy.org/prosite/PS00028>

ZINC_FINGER_C2H2_1, PS00028; Zinc finger C2H2 type domain signature (PATTERN)

Consensus pattern: C - x(2,4) - C - x(3) - [LIVMFYWC] - x(8) - H - x(3,5) - H
The 2 C's and the 2 H's are zinc ligands

Sequences known to belong to this class detected by the pattern: ALL

Other sequence(s) detected in Swiss-Prot: 42.

- Retrieve an alignment of Swiss-Prot true positive hits:
[Clustal format, color, condensed view](#) / [Clustal format, color](#) / [Clustal format, plain text](#) / [Fasta format](#)
- Retrieve the sequence logo from the alignment
- Taxonomic tree view of all Swiss-Prot/TrEMBL entries matching PS00028
- Retrieve a list of all Swiss-Prot/TrEMBL entries matching PS00028
- Scan Swiss-Prot/TrEMBL entries against PS00028
- view ligand binding statistics

Matching PDB structures: 1A1F 1A1G 1A1H 1A1I ... [ALL]

Prosite Zinc Finger Pattern

<http://www.expasy.org/prosite/PS00028>

You are here: [ExPASy CH](#) > [Databases](#) > [PROSITE](#)



Entry: **PS00028**

[Home](#) [ScanProsite](#) [ProRule](#) [Documents](#) [Downloads](#) [Links](#) [Funding](#)

General information about the entry

Entry name	ZINC_FINGER_C2H2_1
Accession number	PS00028
Entry type	PATTERN
Date	APR-1990 (CREATED); JUN-1994 (DATA UPDATE); DEC-2009 (INFO UPDATE).
PROSITE Documentation	PDOC00028

Name and characterization of the entry

Description	Zinc finger C2H2 type domain signature.
Pattern	C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H.

Numerical results

- UniProtKB/Swiss-Prot release number: **57.13**, total number of sequence entries in that release: **514212**.
- Total number of hits in UniProtKB/Swiss-Prot: **12503 hits in 2081 different sequences**
- Number of hits on proteins that are known to belong to the set under consideration: **12197 hits in 1858 different sequences**
- Number of hits on proteins that could potentially belong to the set under consideration: **24 hits in 10 different sequences**
- Number of false hits (on unrelated proteins): **282 hits in 213 different sequences**
- Number of known missed hits: **80**
- Number of partial sequences which belong to the set under consideration, but which are not hit by the pattern or profile because they are partial (fragment) sequences: **1**
- Precision (true hits / (true hits + false positives)): **97.74 %**
- Recall (true hits / (true hits + false negatives)): **99.35 %**

Comments

- Taxonomic range: **Eukaryotes, Eukaryotic viruses**
- Maximum known number of repetitions of the pattern in a single protein: **35**
- `Interesting' site in the pattern: **1,zinc**
- `Interesting' site in the pattern: **3,zinc**
- `Interesting' site in the pattern: **7,zinc**
- `Interesting' site in the pattern: **9,zinc**
- VERSION: **1**



ScanProsite Patterns and Profiles

<http://www.expasy.org/tools/scanprosite/>



ExPASy Proteomics Server

Search for

[Databases](#) [Tools](#) [Services](#) [Mirrors](#) [About](#) [Contact](#)

You are here: [ExPASy CH](#) > [Tools](#) > [Pattern and profile searches](#) > [ScanProsite](#)



[Home](#) [ScanProsite](#) [ProRule](#) [Documents](#) [Downloads](#) [Links](#) [Funding](#)

Sequence(s) to be scanned:	Motif(s) to scan for:
<p>Enter:</p> <ul style="list-style-type: none"> UniProtKB(Swiss-Prot and TrEMBL) AC and/or ID (e.g. P00747, ENTK_HUMAN) PDB identifier(s) your own protein sequence(s) <div style="border: 1px solid black; height: 100px; width: 100%;"></div> <p><input type="button" value="Clear"/></p> <p><input checked="" type="checkbox"/> Exclude motifs with a high probability of occurrence</p> <p><input type="checkbox"/> Do not scan profiles</p>	<p>Enter:</p> <ul style="list-style-type: none"> PROSITE AC and/or ID (e.g. PS50808, CHEB) your own pattern(s) <div style="border: 1px solid black; height: 50px; width: 100%;"></div> <p><input type="button" value="Clear"/></p> <p>Protein database(s):</p> <p><input checked="" type="checkbox"/> UniProtKB/Swiss-Prot <input checked="" type="checkbox"/> including splice variants</p> <p><input type="checkbox"/> UniProtKB/TrEMBL <input type="checkbox"/> PDB</p> <p>randomize databases <input type="text" value="no"/></p> <p><input type="checkbox"/> excluding fragments</p> <p>Filter(s):</p> <p>• On taxonomy: <input type="text"/> (e.g. Eukaryota; Escherichia coli;)</p> <p>• On description: <input type="text"/> (e.g. protease)</p> <p>Pattern option(s):</p> <p>• Allow at most <input type="text" value="0"/> X sequence characters to match a conserved position in the pattern</p> <p>• Match mode <input type="text" value="greedy, overlaps, no includes"/></p>
<p>• Format <input type="text" value="Graphical rich view"/></p> <p>• Show only sequences with at least <input type="checkbox"/> hit(s)</p> <p>Maximum of matched sequences</p> <p><input type="text" value="1000"/></p>	<p>Output:</p> <p><input type="checkbox"/> Show low level score</p> <p><input type="checkbox"/> Retrieve complete sequences</p> <p>Your e-mail: <input type="text"/></p>



ScanProsite Results



Swiss Institute of Bioinformatics



ExPASy Proteomics Server

Search for

[Databases](#) [Tools](#) [Services](#) [Mirrors](#) [About](#) [Contact](#)

You are here: [ExPASy CH](#) > [Databases](#) > [PROSITE](#)



[Home](#) [ScanProsite](#) [ProRule](#) [Documents](#) [Downloads](#) [Links](#)

ScanProsite Results Viewer

This view shows ScanProsite results together with ProRule-based predicted intra-domain features ([help](#)).

[show hits of frequently occurring signatures](#)

Hits for all PROSITE (release 20.58) motifs on sequence P12004 [UniProtKB/Swiss-Pro]

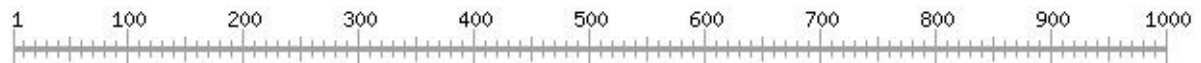
found: 10 hits in 1 sequence

P12004 PCNA_HUMAN (261 aa)

RecName: Full=Proliferating cell nuclear antigen; Short=PCNA; AltName: Full=Cyclin; *Homo sapiens (Human)*

```
MFEARLVQGSILKKVLEALKDLINACWDISSSGVNLQSMDSHVSLVQLTLRSEGFDTYRCDRNL
AMGVNLTSMKILKCAGNEDIITLRAEDNADTLALVFEAPNQEKVSDYEMKLMDDLVEQLGIPEQE
YSCVVKMPSGEFARICRDLSHIGDAVVISCAKDGVKFSASGELGNGNIKLSQTSNVDKEEEAVTIE
MNEPVQLTFALRYLNFFTKATPLSSTVTLSMSADVPLVVEYKIADMGHLKYYLAPKIEDEEGS
```

ruler:



hits by patterns: [2 hits (by 2 distinct patterns) on 1 sequence]





ScanProsite Results

hits by patterns: [2 hits (by 2 distinct patterns) on 1 sequence]

P12004
(PCNA_HUMAN)  (261 aa)

RecName: Full=Proliferating cell nuclear antigen; Short=PCNA; AltName: Full=Cyclin;. *Homo sapiens* (Human)

PS01251 **PCNA_1** *Proliferating cell nuclear antigen signature 1*: Hits on PDB 3D structures: [1AXC-A, 1AXC-C, 1AXC-E, 1U76-A, 1U76-C, 1U76-E, 1U7B-A, 1UL1-A, 1UL1-B, 1UL1-C, 1VYJ-A, 1VYJ-C, 1VYJ-E, 1VYJ-G, 1VYJ-I, 1VYJ-K, 1VYM-A, 1VYM-B, 1VYM-C, 1W60-A, 1W60-B, 2ZVK-A, 2ZVK-B, 2ZVK-C, 2ZVL-A, 2ZVL-B, 2ZVL-C, 2ZVL-D, 2ZVL-E, 2ZVL-F, 2ZVM-A, 2ZVM-B, 2ZVM-C]

34 - 57: GVnLqSMDsSHVsLVqLtLrsegF

PS00293 **PCNA_2** *Proliferating cell nuclear antigen signature 2*: Hits on PDB 3D structures: [1AXC-A, 1AXC-C, 1AXC-E, 1U76-A, 1U76-C, 1U76-E, 1U7B-A, 1UL1-A, 1UL1-B, 1UL1-C, 1VYJ-A, 1VYJ-C, 1VYJ-E, 1VYJ-G, 1VYJ-I, 1VYJ-K, 1VYM-A, 1VYM-B, 1VYM-C, 1W60-A, 1W60-B, 2ZVK-A, 2ZVK-B, 2ZVK-C, 2ZVL-A, 2ZVL-B, 2ZVL-C, 2ZVL-D, 2ZVL-E, 2ZVL-F, 2ZVM-A, 2ZVM-B, 2ZVM-C]

61 - 79: RCDRnlaMgvnLtSMsKIL

hits by patterns with a high probability of occurrence or by user-defined patterns: [8 hits (by 4 distinct patterns) on 1 sequence]





ScanProsite Results

hits by patterns with a high probability of occurrence or by user-defined patterns: [8 hits (by 4 distinct patterns) on 1 sequence]

P12004
(PCNA_HUMAN)  (261 aa)

RecName: Full=Proliferating cell nuclear antigen; Short=PCNA; AltName: Full=Cyclin;. *Homo sapiens (Human)*

PS00005 PKC_PHOSPHO_SITE *Protein kinase C phosphorylation site* :

51 - 53: TLR

59 - 61: TyR

89 - 91: TLR

PS00008 MYRISTYL *N-myristoylation site* :

69 - 74: GVnITS

166 - 171: GVkfSA

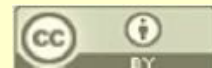
PS00001 ASN_GLYCOSYLATION *N-glycosylation site* :

71 - 74: NLTS

PS00006 CK2_PHOSPHO_SITE *Casein kinase II phosphorylation site* :

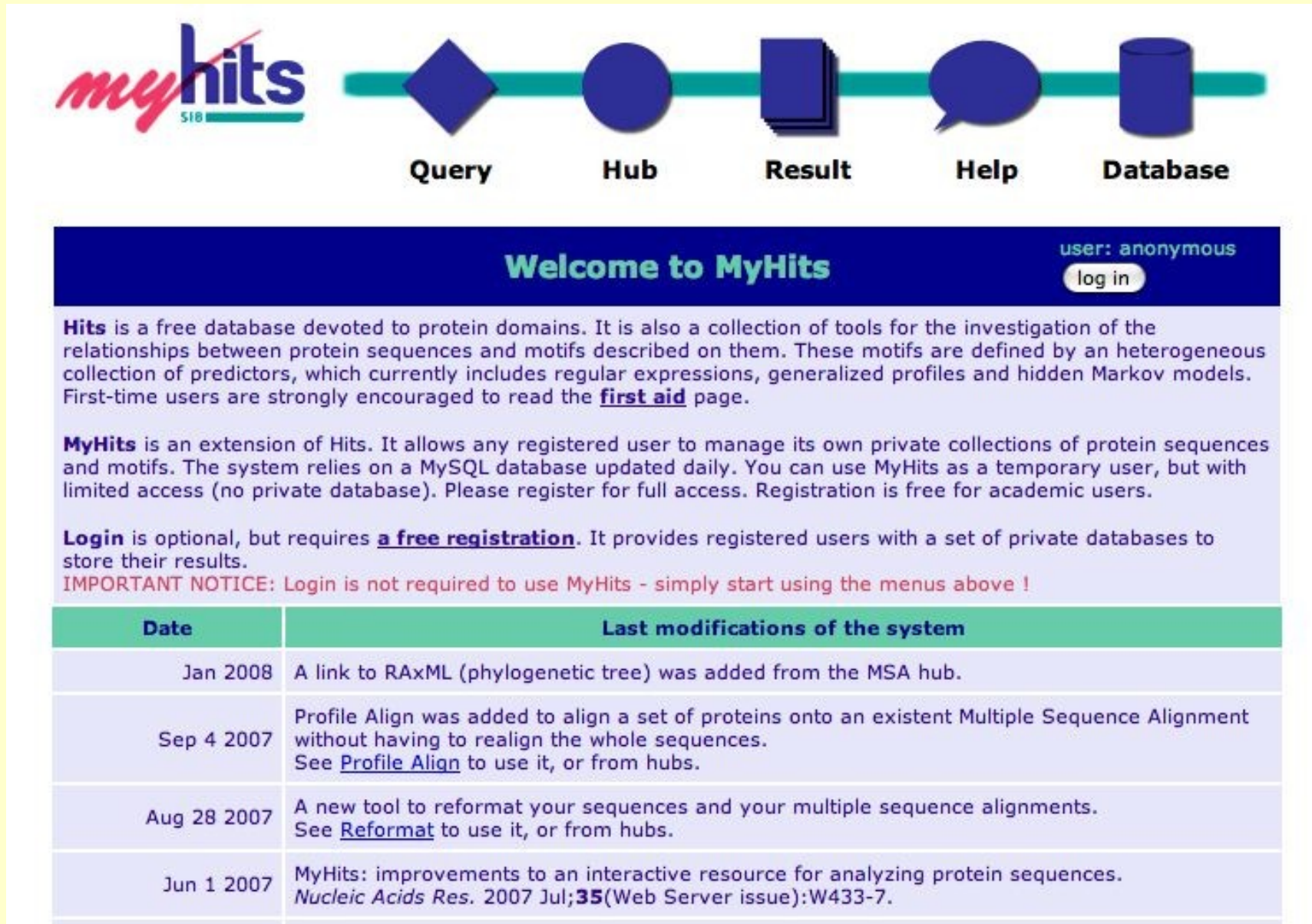
112 - 115: SdyE

186 - 189: SnvD



MyHits Local Motifs Search

<http://myhits.isb-sib.ch/>

The screenshot shows the MyHits website interface. At the top left is the myhits SIB logo. To its right is a navigation bar with five icons: a diamond for 'Query', a circle for 'Hub', a stack of papers for 'Result', a speech bubble for 'Help', and a cylinder for 'Database'. Below the navigation bar is a dark blue header with the text 'Welcome to MyHits' and 'user: anonymous' with a 'log in' button. The main content area contains three paragraphs of text describing the database and its features. Below the text is a table with two columns: 'Date' and 'Last modifications of the system'. The table lists several updates from 2007 and 2008.

Welcome to MyHits user: anonymous
log in

Hits is a free database devoted to protein domains. It is also a collection of tools for the investigation of the relationships between protein sequences and motifs described on them. These motifs are defined by a heterogeneous collection of predictors, which currently includes regular expressions, generalized profiles and hidden Markov models. First-time users are strongly encouraged to read the [first aid](#) page.

MyHits is an extension of Hits. It allows any registered user to manage its own private collections of protein sequences and motifs. The system relies on a MySQL database updated daily. You can use MyHits as a temporary user, but with limited access (no private database). Please register for full access. Registration is free for academic users.

Login is optional, but requires [a free registration](#). It provides registered users with a set of private databases to store their results.

IMPORTANT NOTICE: Login is not required to use MyHits - simply start using the menus above !

Date	Last modifications of the system
Jan 2008	A link to RAXML (phylogenetic tree) was added from the MSA hub.
Sep 4 2007	Profile Align was added to align a set of proteins onto an existent Multiple Sequence Alignment without having to realign the whole sequences. See Profile Align to use it, or from hubs.
Aug 28 2007	A new tool to reformat your sequences and your multiple sequence alignments. See Reformat to use it, or from hubs.
Jun 1 2007	MyHits: improvements to an interactive resource for analyzing protein sequences. <i>Nucleic Acids Res.</i> 2007 Jul; 35 (Web Server issue):W433-7.

MyHits Local Motifs Query

<http://myhits.isb-sib.ch/>



user: anonymous
[log in](#)

Motif Scan

Protein Sequence Input
Enter a protein sequence in RAW or FASTA or Swiss-Prot format or a db:AC or db:ID identifier

```
>gi|4503965|ref|NP_000504.1| opsin 1 (cone pigments),
MAQQWSLQRLAGRHPQDSYEDSTQSSIFTYTNNSNSTRGPFEGPNYHIAPRWVY
NGLVLAATMKFKLRHPLNWLNLAVADLAETVIASISVVNQVYGYFVLGF
LWSLAIISWERWMVVKPFGNVRFDKLAIVGIAFSWIWAAVWTAPPPIFGWSE
SSYPGVQSYMIVLMVTCCITPLSIIIVLCYLQVWLAIKRAVAKQKKESESTQKAE
WGPYAFFACFAAANPGYPFHPLMAALPAFFAKSATIYNPVIYVFMNRQFRNCI
SKTEVSSVSSVSPA
```

[clear input](#)
[reset page](#)

Motif scanning means finding all known motifs that occur in a sequence. This form lets you paste a protein sequence, select the collections of motifs to scan for, and launch the search. Some general [documentation](#) is available about the Prosite and Pfam collections of motifs. Another [document](#) deals with the interpretation of the match scores. You should consult the home pages of [Prosite](#) on ExPASy, [Pfam](#) and [InterPro](#) for additional information.

Warning: The scan might take a few minutes, thus if your proteins of interest are already in the sequence databases (see [list](#)), the [Query by Protein](#) form is much faster, and the [Protein Hub](#) provides a collection of tools that you might find useful.

Parameters

Database of motifs (db description)	<input type="checkbox"/> PeroxiBase profiles <input checked="" type="checkbox"/> PROSITE patterns <input checked="" type="checkbox"/> PROSITE patterns (frequent match producers) <input checked="" type="checkbox"/> PROSITE profiles <input checked="" type="checkbox"/> Prefile (more profiles) <input type="checkbox"/> HAMAP profiles <input type="checkbox"/> Na-channel profiles <input checked="" type="checkbox"/> Pfam HMMs (local models) <input type="checkbox"/> Pfam HMMs (global models)	<input type="button" value="search"/>
---	---	---------------------------------------

[Question or comment about this page.](#)

MyHits Local Motifs Search

<http://myhits.isb-sib.ch/>



Motif Scan Results

user: anonymous

[log in](#)

Query Protein temporarily stored [here](#).

Database of motifs PROSITE patterns, PROSITE patterns (frequent match producers), PROSITE profiles, Profile (more profiles), Pfam HMMs (local models).

Reference Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJ, Hofmann K & Bairoch A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res.* **30**:235-238

searching PROSITE patterns
 searching PROSITE patterns (frequent match producers)
 searching PROSITE profiles
 searching Profile (more profiles)
 searching Pfam HMMs (local models)
 postprocessing

MyHits Local Motifs Summary

<http://myhits.isb-sib.ch/>

Summary																																																																																	
Original output	<u>pat</u> , <u>prf</u> , <u>pre</u> , <u>pfam fs</u> , <u>pfam ls</u> .																																																																																
Matches map (features from query are above the ruler, matches of the motif scan are below the ruler)	<p>Legends: 1, pat:EGF_2 [!]; 2, pat:G_PROTEIN_RECEP_F1_1 [!]; 3, pat:OPSIN [!]; 4, prf:ALPHA_BOX [?].</p>																																																																																
List of matches	<table border="1"> <tbody> <tr><td>FT</td><td>MYHIT</td><td>278</td><td>289</td><td>pat:EGF_2 [!]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>139</td><td>155</td><td>pat:G_PROTEIN_RECEP_F1_1 [!]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>306</td><td>322</td><td>pat:OPSIN [!]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>1</td><td>9</td><td>prf:ALPHA_BOX [?]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>70</td><td>322</td><td>prf:G_PROTEIN_RECEP_F1_2 [!]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>70</td><td>322</td><td>pfam_fs:7tm_1 [!]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>56</td><td>331</td><td>pfam_ls:7TM_GPCR_Srd [?]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>64</td><td>337</td><td>pfam_ls:7TM_GPCR_Srsx [?]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>26</td><td>345</td><td>pfam_ls:7TM_GPCR_Srt [?]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>48</td><td>339</td><td>pfam_ls:7TM_GPCR_Srv [?]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>51</td><td>340</td><td>pfam_ls:7TM_GPCR_Srw [?]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>61</td><td>339</td><td>pfam_ls:7TM_GPCR_Srx [?]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>56</td><td>316</td><td>pfam_ls:7TM_GPCR_Str [?]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>70</td><td>322</td><td>pfam_ls:7tm_1 [!]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>223</td><td>348</td><td>pfam_ls:EBP [?]</td></tr> <tr><td>FT</td><td>MYHIT</td><td>56</td><td>333</td><td>pfam_ls:Srg [?]</td></tr> </tbody> </table>	FT	MYHIT	278	289	pat:EGF_2 [!]	FT	MYHIT	139	155	pat:G_PROTEIN_RECEP_F1_1 [!]	FT	MYHIT	306	322	pat:OPSIN [!]	FT	MYHIT	1	9	prf:ALPHA_BOX [?]	FT	MYHIT	70	322	prf:G_PROTEIN_RECEP_F1_2 [!]	FT	MYHIT	70	322	pfam_fs:7tm_1 [!]	FT	MYHIT	56	331	pfam_ls:7TM_GPCR_Srd [?]	FT	MYHIT	64	337	pfam_ls:7TM_GPCR_Srsx [?]	FT	MYHIT	26	345	pfam_ls:7TM_GPCR_Srt [?]	FT	MYHIT	48	339	pfam_ls:7TM_GPCR_Srv [?]	FT	MYHIT	51	340	pfam_ls:7TM_GPCR_Srw [?]	FT	MYHIT	61	339	pfam_ls:7TM_GPCR_Srx [?]	FT	MYHIT	56	316	pfam_ls:7TM_GPCR_Str [?]	FT	MYHIT	70	322	pfam_ls:7tm_1 [!]	FT	MYHIT	223	348	pfam_ls:EBP [?]	FT	MYHIT	56	333	pfam_ls:Srg [?]
FT	MYHIT	278	289	pat:EGF_2 [!]																																																																													
FT	MYHIT	139	155	pat:G_PROTEIN_RECEP_F1_1 [!]																																																																													
FT	MYHIT	306	322	pat:OPSIN [!]																																																																													
FT	MYHIT	1	9	prf:ALPHA_BOX [?]																																																																													
FT	MYHIT	70	322	prf:G_PROTEIN_RECEP_F1_2 [!]																																																																													
FT	MYHIT	70	322	pfam_fs:7tm_1 [!]																																																																													
FT	MYHIT	56	331	pfam_ls:7TM_GPCR_Srd [?]																																																																													
FT	MYHIT	64	337	pfam_ls:7TM_GPCR_Srsx [?]																																																																													
FT	MYHIT	26	345	pfam_ls:7TM_GPCR_Srt [?]																																																																													
FT	MYHIT	48	339	pfam_ls:7TM_GPCR_Srv [?]																																																																													
FT	MYHIT	51	340	pfam_ls:7TM_GPCR_Srw [?]																																																																													
FT	MYHIT	61	339	pfam_ls:7TM_GPCR_Srx [?]																																																																													
FT	MYHIT	56	316	pfam_ls:7TM_GPCR_Str [?]																																																																													
FT	MYHIT	70	322	pfam_ls:7tm_1 [!]																																																																													
FT	MYHIT	223	348	pfam_ls:EBP [?]																																																																													
FT	MYHIT	56	333	pfam_ls:Srg [?]																																																																													

MyHits Local Motif Hits

<http://myhits.isb-sib.ch/>








Match details		
match detail	match score	motif information
<p>1</p> <pre> KK XGRR I:: FGKK </pre> <p>^ image ^</p>	<p>Status: ?</p> <p>pos.: 337-340</p>	<p>freq_pat:AMIDATION</p> <p><i>Amidation site.</i></p> <p>[entry]</p> <p>Legends: 1, amidation.</p>
<p>1</p> <pre> Y Y W W V V T T S S R R Q Q N N M M L L K K I I H H G G F F E E D D C T C NASA I::: NSTR </pre> <p>^ image ^</p>	<p>Status: ?</p> <p>pos.: 34-37</p>	<p>freq_pat:ASN_GLYCOSYLATION</p> <p><i>N-glycosylation site.</i></p> <p>[entry]</p> <p>Legends: 1, carbohydrate.</p>
<p>1</p> <pre> T E SXXD ::: SYED </pre>	<p>Status: ?</p> <p>pos.: 18-21</p>	<p>freq_pat:CK2_PHOSPHO_SITE</p> <p><i>Casein kinase II phosphorylation site.</i></p> <p>[entry]</p>
<p>1</p> <pre> T E SXXD ::: SKTE </pre> <p>^ image ^</p>	<p>Status: ?</p> <p>pos.: 351-354</p>	<p>Legends: 1, phosphorylation.</p>

MyHits Local Motifs Hits (Cont.)

<http://myhits.isb-sib.ch/>



 <pre> 1 T K S R : : STR </pre>	Status: ? pos.: 35-37	freq_pat:PKC_PHOSPHO_SITE <i>Protein kinase C phosphorylation site.</i> [entry]
 <pre> 1 T K SXR : : TMK </pre>	Status: ? pos.: 78-80	Legends: 1, phosphorylation.
 <pre> 1 T K SXR : : TQK </pre>	Status: ? pos.: 259-261	Legends: 1, phosphorylation.
 <pre> 1 K E RXXXDXXX Y : : RGPfEGPhY </pre>	Status: ? pos.: 37-45	freq_pat:TYR_PHOSPHO_SITE <i>Tyrosine kinase phosphorylation site.</i> [entry] Legends: 1, phosphorylation.
 <pre> disulfide disulfide disulfide PY CXCXXGFXXXXC : CFCWGPYAFFAC </pre>	Status: ! pos.: 278-289	pat:EGF_2 <i>EGF-like domain signature 2.</i> [entry]



MyHits Local Motifs Hits (Cont.)



[^ image ^](#)

<pre> Y YY Y W WW W V VV V T TT T S SS S W R QQ W C V N PP C W T M NN C Y S L MM A FEQ K LL T MDN IAI I S VPM HGHH W ICL GQGGTCY HH LNI FNFFNF SG AAG EMEEMAMH CF M TTF DVDDVTYN WD V SSC CICCISIE YC I GGAXALAAALGLDRFAXL ::: ::::: ::: TGLWSLAIISWERWMVW </pre>	<p>Status: ! pos.: 139-155</p>	<p>pat:G_PROTEIN_RECEP_F1_1 <i>G-protein coupled receptors family 1 signature.</i> [entry]</p>
---	---	--

[^ image ^](#)

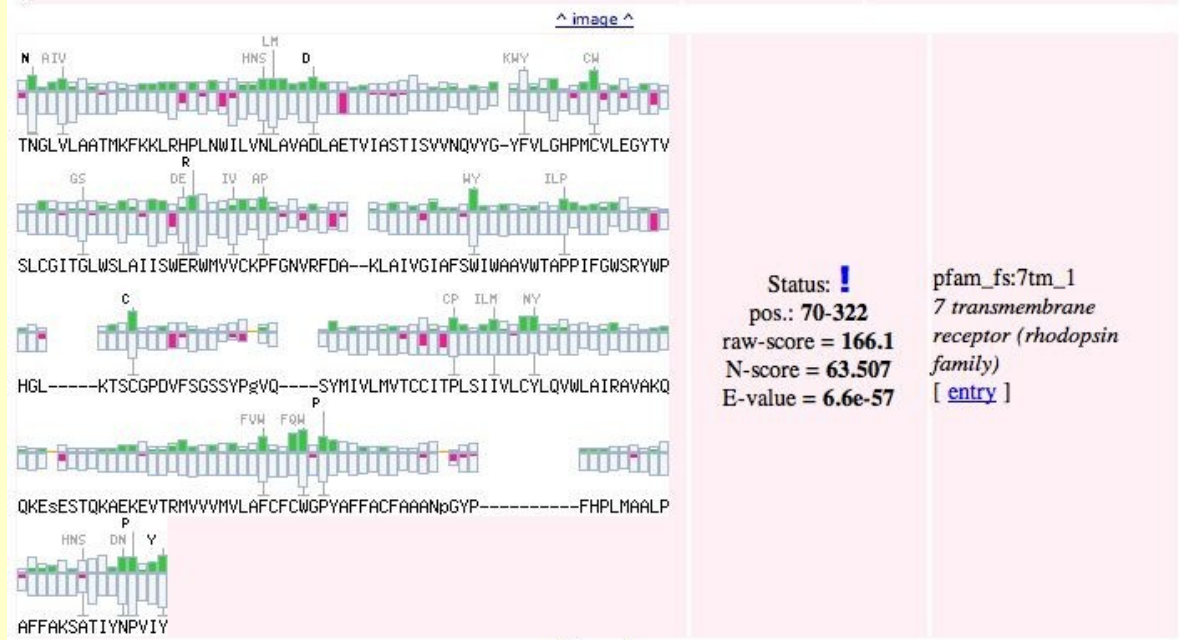
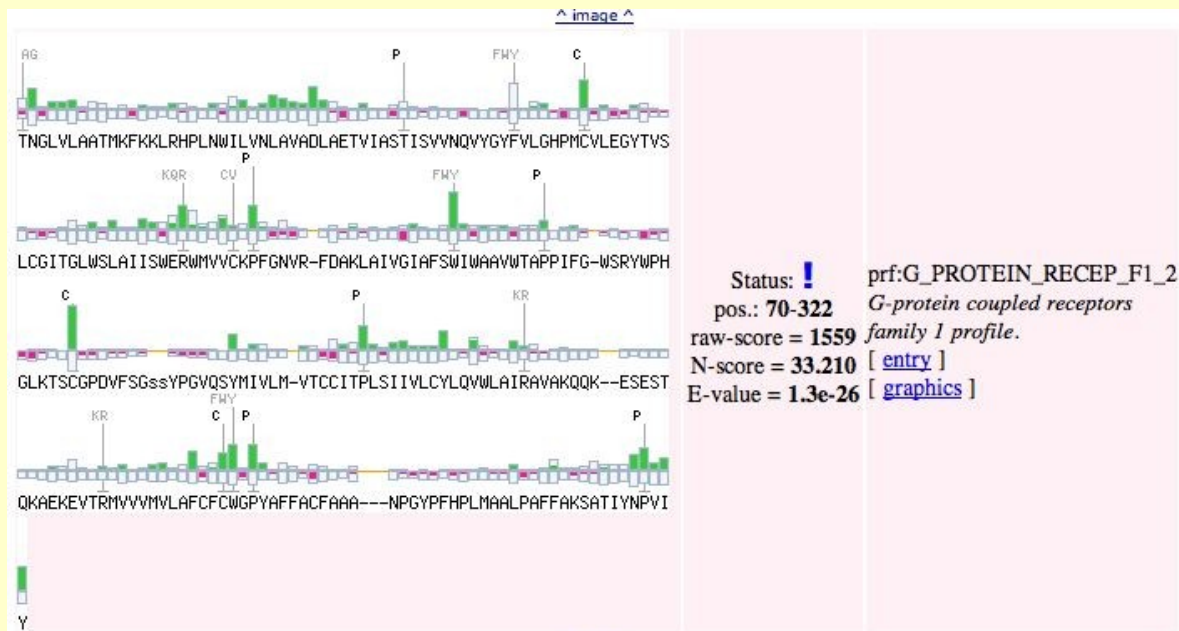
<pre> retinal Y W V T S R R R Q P N M L K I H C A W F C M A E V G D C I S C A LPXAXSKSGSXXDAXXI ::: : ::: : LPAFFAKSATIYNPVIY </pre>	<p>Status: ! pos.: 306-322</p>	<p>pat:OPSIN <i>Visual pigments (opsins) retinal binding site.</i> [entry]</p>
--	---	--

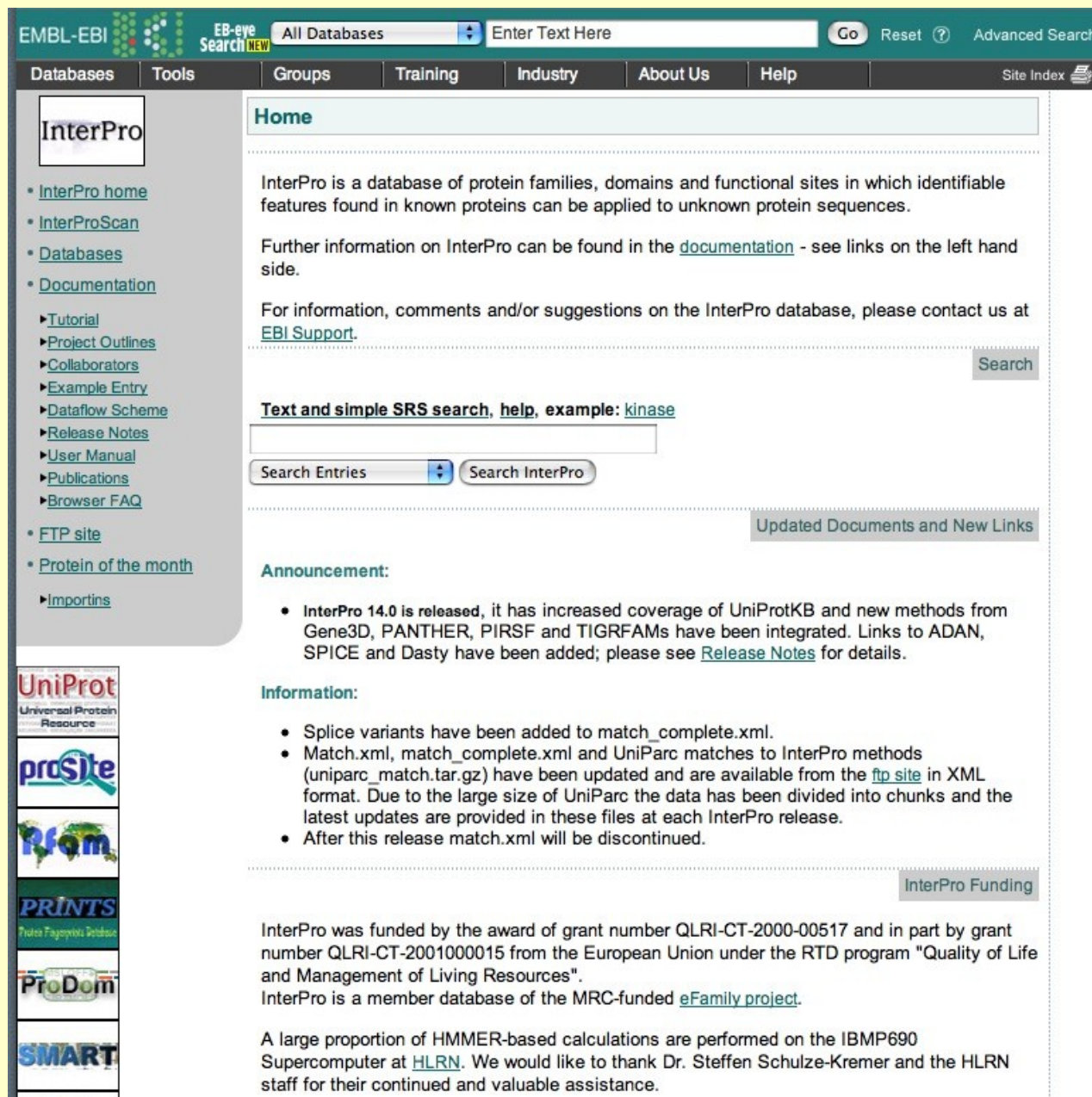
[^ image ^](#)

 <p>MAQQWSLQR</p>	<p>Status: ? pos.: 1-9 raw-score = 191 N-score = 7.485 E-value = 0.69</p>	<p>prf:ALPHA_BOX <i>Alpha box DNA-binding domain profile.</i> [entry] [graphics]</p>
--	--	--

[^ image ^](#)

MyHits Local Motifs Hits (Cont.)





EMBL-EBI EB-eye Search **NEW** All Databases Enter Text Here Go Reset ? Advanced Search

Databases Tools Groups Training Industry About Us Help Site Index

InterPro

- [InterPro home](#)
- [InterProScan](#)
- [Databases](#)
- [Documentation](#)
 - ▶ [Tutorial](#)
 - ▶ [Project Outlines](#)
 - ▶ [Collaborators](#)
 - ▶ [Example Entry](#)
 - ▶ [Dataflow Scheme](#)
 - ▶ [Release Notes](#)
 - ▶ [User Manual](#)
 - ▶ [Publications](#)
 - ▶ [Browser FAQ](#)
- [FTP site](#)
- [Protein of the month](#)
 - ▶ [Importins](#)

Home

InterPro is a database of protein families, domains and functional sites in which identifiable features found in known proteins can be applied to unknown protein sequences.

Further information on InterPro can be found in the [documentation](#) - see links on the left hand side.

For information, comments and/or suggestions on the InterPro database, please contact us at [EBI Support](#).

Search

Text and simple SRS search, help, example: kinase

Search Entries Search InterPro

Updated Documents and New Links

Announcement:

- **InterPro 14.0 is released**, it has increased coverage of UniProtKB and new methods from Gene3D, PANTHER, PIRSF and TIGRFAMs have been integrated. Links to ADAN, SPICE and Dasty have been added; please see [Release Notes](#) for details.

Information:

- Splice variants have been added to match_complete.xml.
- Match.xml, match_complete.xml and UniParc matches to InterPro methods (uniparc_match.tar.gz) have been updated and are available from the [ftp site](#) in XML format. Due to the large size of UniParc the data has been divided into chunks and the latest updates are provided in these files at each InterPro release.
- After this release match.xml will be discontinued.

InterPro Funding

InterPro was funded by the award of grant number QLRI-CT-2000-00517 and in part by grant number QLRI-CT-2001000015 from the European Union under the RTD program "Quality of Life and Management of Living Resources".

InterPro is a member database of the MRC-funded [eFamily project](#).

A large proportion of HMMER-based calculations are performed on the IBMP690 Supercomputer at [HLRN](#). We would like to thank Dr. Steffen Schulze-Kremer and the HLRN staff for their continued and valuable assistance.

UniProt Universal Protein Resource

proSite

ProDom

PRINTS Protein Family Signatures

ProDom

SMART

EMBL-EBI EB-eye Search **NEW** All Databases Enter Text Here Go Reset ? Advanced Search

Databases Tools Groups Training Industry About Us Help Site Index

- InterPro home
- Text Search
- InterProScan
- Databases
- Documentation
- FTP Site

InterProScan Help

- Help
- FAQ
- README

InterProScan Programmatic Access

Database Information

- UniProt
- UniParc

EBI > Tools > Protein Functional Analysis

InterProScan Sequence Search

This form allows you to query your sequence against InterPro. For more detailed information see the documentation for the perl stand-alone InterProScan package ([Readme file](#) or [FAQ's](#)), or the InterPro [user manual](#) or [help pages](#).

Please Note: InterProScan job submissions should be limited to one sequence only. The system will no longer process 6 protein sequences simultaneously as of Monday Feb 13, 2006. Please contact [support](#) for help in submitting multiple sequences.

[Download Software](#)

YOUR EMAIL:

RESULTS:

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		

TRANSLATION TABLE (DNARNA only):

MIN. OPEN READING FRAME SIZE:

Enter or Paste a Sequence in any format:

```
SFTLTNKNVIFVAGLGGIGLDTSKELLKRDCLKNLVILDRIENPAAIAELK
AINPKVTVTFYPYDVTVPAAETTKLLKTIFAQLKTVLDVINGAGILDDHQ
IERTIAVNYTGLVNTTTAILDFWDKRKGPGGIIICNIGSVTFNAIYQVP
VYSGTKAAVVNFTSSLAKLAPITGVTAYTVNPGITRTLVLVHKFNSWLDV
E
PQVAEKLLAHTQPQLACAENFVKAIELNQNGAIWKLDLGTLEAIQWTK
H
WDSGI
```

Upload a file: no file selected

InterPro Scan Hourglass

<http://www.ebi.ac.uk/InterProScan/>

EMBL-EBI



Your job is currently running...
...please be patient

The results of your job will appear in this browser window.

Your Job output: <http://www.ebi.ac.uk/cgi-bin/iprscan/iprscan?tool=iprscan&jobid=iprscan-20070124-18021178>

Please Note the Following:

- You may press Shift+Refresh or Reload on your browser at any time to check if results are ready. Should this window go blank please press the Shift+Refresh or Reload button on your browser.
- You may bookmark this page to view your results later if you wish.
Netscape users: Use Bookmark - Add Bookmark or CTRL-D | Alt-K to bookmark this page.
IE users: Click -> [BookMark](#) to bookmark this page.
- Results are stored for 24 hours. Some big files will be deleted after ca. 15 minutes.

EMBL-EBI EB-eye Search All Databases Enter Text Here Go Reset ? Give us feedback
 Advanced Search

Databases Tools EBI Groups Training Industry About Us Help Site Index

- Help
- General Help
- Formats
- Gaps
- Matrix
- References
- InterProScan Help

InterProScan Results

Table View Raw Output XML Output Original Sequences SUBMIT ANOTHER JOB

SEQUENCE: NP_872590.1 CRC64: E6F08E7EDBC48B00 LENGTH: 261 aa

InterPro IPR000730 Family InterPro 	Proliferating cell nuclear antigen, PCNA	
	PR00339	PCNACYCLIN
	PTHR11352	PROLIFERATING CELL NUCLEAR ANTIGEN
	PF00705	PCNA_N
	PF02747	PCNA_C
	TIGR00590	pcna: proliferating cell nuclear antigen (p)
	PS00293	PCNA_2
	PS01251	PCNA_1
noIPR unintegrated	unintegrated	
	G3DSA:3.70.10.10	no description
	SSF55979	DNA clamp

Table View Raw Output XML Output Original Sequences SUBMIT ANOTHER JOB

InterPro Scan PCNA Table View

<http://www.ebi.ac.uk/InterProScan/>

EMBL-EBI All Databases

Databases Tools EBI Groups Training Industry About Us Help

- Help
- General Help
- Formats
- Gaps
- Matrix
- References
- InterProScan Help

InterProScan Results

SEQUENCE: NP_872590.1 CRC64: E6F08E7EDBC48B00 LENGTH: 261 aa

InterPro IPR000730 Family InterPro 	Proliferating cell nuclear antigen, PCNA			
	PRINTS	PR00339	<i>PCNACYCLIN</i>	3e-68 [10-29]T 3e-68 [34-52]T 3e-68 [56-80]T 3e-68 [110-132]T 3e-68 [203-217]T 3e-68 [241-258]T
	PANTHER	PTHR11352	<i>PROLIFERATING CELL NUCLEAR ANTIGEN</i>	8.8e-200 [2-259]T
	PFAM	PF00705	<i>PCNA_N</i>	1.2e-93 [1-125]T
	PFAM	PF02747	<i>PCNA_C</i>	6e-97 [127-254]T
	TIGRFAMs	TIGR00590	<i>pcna: proliferating cell nuclear antigen (p</i>	5.5e-161 [1-259]T
	PROSITE	PS00293	<i>PCNA_2</i>	NA [61-79]T
	PROSITE	PS01251	<i>PCNA_1</i>	NA [34-57]T
Parent	no parent			
Children	no children			
Found in	no entries			
Contains	no entries			
GO terms	Molecular Function: DNA binding (GO:0003677) Biological Process: regulation of DNA replication (GO:0006275) Molecular Function: DNA polymerase processivity factor activity (GO:0030337) Cellular Component: PCNA complex (GO:0043626)			
noIPR unintegrated	unintegrated			
	GENE3D	G3DSA:3.70.10.10	<i>no description</i>	1.7e-100 [1-261]T
	SUPERFAMILY	SSF55979	<i>DNA clamp</i>	1.1e-49 [127-259]T 2.7e-48 [1-126]T
Parent	no parent			
Children	no children			
Found in	no entries			
Contains	no entries			
GO terms	none			

Family: PCNA_N (PF00705)



4
architectures



510
sequences



2
interactions



264 species



103
structures

Summary

[Domain organisation](#)
[Alignments](#)
[HMM logo](#)
[Trees](#)
[Curation & models](#)
[Species](#)
[Interactions](#)
[Structures](#)
[Jump to...](#)

Summary

Proliferating cell nuclear antigen, N-terminal domain [Add annotation](#)

N-terminal and C-terminal domains of PCNA are topologically identical. Three PCNA molecules are tightly associated to form a closed ring encircling duplex DNA.

Literature references

1. Krishna TS, Kong XP, Gary S, Burgers PM, Kuriyan J; , Cell 1994;79:1233-1243.: Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. [PUBMED:8001157](#)

InterPro entry [IPR000730](#)

Proliferating cell nuclear antigen (PCNA), or cyclin, is a non-histone acidic nuclear protein [PUBMED:2884104](#) that plays a key role in the control of eukaryotic DNA replication [PUBMED:1346518](#). It acts as a co-factor for DNA polymerase delta, which is responsible for leading strand DNA replication [PUBMED:2565339](#). The sequence of PCNA is well conserved between plants and animals, indicating a strong selective pressure for structure conservation, and suggesting that this type of DNA replication mechanism is conserved throughout eukaryotes [PUBMED:1671766](#). In *Saccharomyces cerevisiae* (Baker's yeast), POL30, is associated with polymerase III, the yeast analog of polymerase delta.

Homologues of PCNA have also been identified in the archaea (Euryarchaeota and Crenarchaeota) and in *Paramecium bursaria* [Chlorella virus 1](#) (PBCV-1) and in nuclear polyhedrosis viruses.

Clan

This family is a member of clan [DNA clamp](#) ([CL0060](#)), which contains the following 10 members:

[DNA pol3 beta](#)
[Herpes_UL42](#)
[Rad1](#)

[DNA pol3 beta 2](#)
[Hus1](#)
[Rad9](#)

[DNA pol3 beta 3](#)
[PCNA_C](#)

[DNA_PPF](#)
[PCNA_N](#)



Example structure

PDB entry 2nti: Crystal structure of PCNA123 heterotrimer.

View a different structure:

2nti

InterPro PRINTS Result for PCNA

http://www.bioinf.manchester.ac.uk/cgi-bin/dbbrowser/sprint/searchprintss.cgi?display_opts=Prints&category=N

EM
D
H
C
F
C
M
F
It

==SPRINT==> Query Results

[SPRINT Home](#) [UMBER Home](#) [Contents](#) [Standard Search](#)

==SPRINT==> PRINTS View

view selected as

PR00339

Identifier	PCNACYCLIN [View Relations] [View Alignment] [View Structure]
Accession	PR00339
No. of Motifs	6
Creation Date	07-OCT-1994 (UPDATE 10-JUN-1999)
Title	Proliferating cell nuclear antigen (cyclin) signature
Database References	PROSITE; PS01251 PCNA_1 ; PS00293 PCNA_2 BLOCKS; PR00339 PFAM; PF00705 PCNA INTERPRO; IPR000730
Literature References	<ol style="list-style-type: none">1. MATSUMOTO, K., MORIUCHI, T., KOJI, T. AND NAKANE, P.K. Molecular cloning of cDNA coding for rat proliferating cell nuclear antigen EMBO J. 6 637-642 (1987).2. HATA, S., KOUCHI, H., TANAKA, Y., MINAMI, E., MATSUMOTO, T., SUZUKA, I. AND HASHIMOTO, J. Identification of carrot cDNA clones encoding a second putative proliferating cell-nuclear antigen, DNA polymerase delta auxiliary protein. EUR.J.BIOCHEMISTRY 203 367-371 (1991).3. TRAVALI, S., KU, D.H., RIZZO, M.G., OTTAVIO, L., BASERGA, R. AND CALABRETTA, B. Structure of the human gene for the proliferating cell nuclear antigen. J.BIOL.CHEM. 264 7466-7472 (1989).4. SUZUKA, I., HATA, S., MATSUOKA, M., KOSUGI, S. AND HASHIMOTO, J. Highly conserved structure of proliferating cell nuclear antigen (DNA polymerase delta auxiliary protein) gene in plants. EUR.J.BIOCHEMISTRY 195 571-575 (1991).

I

InterPro TIGRFams Result for PCNA

http://cmr.jcvi.org/tigr-scripts/CMR/HmmReport.cgi?hmm_acc=TIGR00590



Comprehensive Microbial Resource

Search Locus for Go

[Home](#) [Genome Tools](#) [Searches](#) [Comparative Tools](#) [Lists](#) [Downloads](#) [Carts](#)

[CMR Manual](#) | [Home](#) > [HMM Summary Page](#)

HMM Summary Page: TIGR00590

[Download](#)

Accession: TIGR00590 | Name: pcna | Function: proliferating cell nuclear antigen (pcna)

[HMM Summary](#)

[HMM NIAA Members](#)

[HMM CMR Members](#)

[HMM seed alignment in FASTA](#)

[HMM seed alignment in MSF](#)

[HMM seed alignment in Belvu](#)

[HMM CMR Search](#)

[HMM BLAST](#)

[HMM Text Search](#)

Accession	TIGR00590
Name	pcna
Function	proliferating cell nuclear antigen (pcna)
Trusted Cutoff	50.00
Domain Trusted Cutoff	50.00
Noise Cutoff	0.00
Domain Noise Cutoff	0.00
Isology Type	equivalog
HMM Length	269
Mainrole Category	DNA metabolism
Subrole Category	DNA replication, recombination, and repair
Gene Ontology Term	GO:0006260 : DNA replication (biological_process)
Gene Ontology Term	GO:0030337 : DNA polymerase processivity factor activity (molecular_function)
Relationship	InterPro assignment: IPR000730
Author	Loftus BJ, Eisen JA
Entry Date	Oct 19 1999 6:19PM
Last Modified	Sep 23 2003 4:26PM
Comment	All proteins in this family for which functions are known form sliding DNA clamps that are used in DNA replication processes. This family is based on the phylogenomic analysis of JA Eisen (1999, Ph.D. Thesis, Stanford University).
References	GA hmmsearch DR EGAD; 95043; DR SWISSPROT; P12004; DR HAMAP; MF_00317; 35 of 48

EMBL-EBI EB-eye Search All Databases Enter Text Here Go Reset Advanced Search Give us feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

- Databases Home
- Databases A-Z
- :: Database Browsing
- :: Biological Ontologies
- :: Literature
- :: Microarray
- :: Nucleotide
- :: Pathways & Networks
- Protein
 - CluSTR
 - CSA
 - GOA
 - HPI
 - IntAct
 - IntEnz
 - InterPro
 - IPI
 - LGICdb
 - PANDIT
 - UniProt
 - UniProtKB/Swiss-Prot
 - UniProtKB/TrEMBL
- :: Proteomic
- :: Structure
- Submissions
- Downloads

UniProt

UniProt is a central database of protein sequence and function created by joining the information contained in UniProtKB/Swiss-Prot, UniProtKB/TrEMBL, and PIR.

UniProt Databases:

UniProt Q? SRS ↓
 UniRef Q? SRS ↓
 UniParc Q? SRS

EBI > Databases > Proteins

Protein Databases

The EBI has developed and maintained a number of protein related databases. Below is a list of the projects and databases provided.


Database	Description
CluSTR	Offers an automatic classification of UniProtKB/Swiss-Prot + UniProtKB/TrEMBL.
CSA	CSA - The Catalytic Site Atlas is a resource of catalytic sites and residues identified in enzymes using structural data.
GOA	Provides assignments of proteins in UniProtKB/Swiss-Prot, UniProtKB/TrEMBL and IPI to the Gene Ontology resource.
HPI	Human Proteomics Initiative (HPI) is an initiative, by SIB and the EBI, to annotate all known human sequences according to the quality standards of UniProtKB/Swiss-Prot.
IntAct	IntAct is a protein interaction database and analysis system. It provides a query interface and modules to analyse interaction data.
IntEnz	The Integrated relational Enzyme database (IntEnz) will contain enzyme data approved by the Nomenclature Committee. The goal is to create a single relational enzyme database.
InterPro	The InterPro database is an integrated documentation resource for protein families, domains and functional sites.
IPI	International Protein Index contains a number of non-redundant proteome sets of higher eukaryotic organisms constructed from UniProtKB/Swiss-Prot, UniProtKB/TrEMBL, Ensembl and RefSeq.
LGICdb	The Ligand Gated Ion Channel Database.
PANDIT	PANDIT - Protein and Associated Nucleotide Domains with Inferred Trees. PANDIT is a collection of multiple sequence alignments and phylogenetic trees covering many common protein domains.
UniProt	The Universal Protein Resource for protein sequences and is the central hub for the collection of functional information on proteins, with accurate, consistent, and rich annotation, the amino acid sequence, protein name or description, taxonomic data and citation information.
UniProt Archive	A non-redundant archive of protein sequences extracted from public databases and contains only protein sequences.
UniProt/UniRef	Features clustering of similar sequences to yield a representative subset of sequences. This produces very fast search times.
UniProt/UniMES	A repository specifically developed for metagenomic and environmental data.
UniProtKB/Swiss-Prot	An annotated protein sequence database. Part of the UniProtKB.
UniProtKB/TrEMBL	A computer generated protein database enriched with automated classification and annotation. Part of the UniProtKB.

NIH funds major universal protein resource at EBI

Dec 15th 2003. NIH funds major universal protein resource at EBI. National Human Genome Research

EBI UniProt

<http://www.ebi.ac.uk/uniprot/>



EMBL-EBI **EB-eye Search** All Databases **Go** **Reset** **Advanced Search** **Give us feedback**

Databases Tools EBI Groups Training Industry About Us Help Site Index


- UniProt Home
- UniProt Website
- About UniProt
- Searches/Tools
- UniSave
- Submissions
- Downloads
- Documentation
- Publications
- People
- Help

EBI > Databases > Protein > UniProt


UniProt - Welcome to UniProt

The mission of UniProt is to provide the scientific community with a comprehensive, high quality and freely accessible resource of protein sequence and functional information. UniProt is comprised of four components, each optimised for different uses. The **UniProt Knowledgebase (UniProtKB)** is the central access point for extensive curated protein information, including function, classification, and cross-reference. It consists of two sections: **UniProtKB/Swiss-Prot** which is manually annotated and is reviewed and **UniProtKB/TrEMBL** which is automatically annotated and is not reviewed. The **UniProt Reference Clusters (UniRef)** databases provide clustered sets of sequences from the UniProtKB and selected UniProt Archive records to obtain complete coverage of sequence space at several resolutions while hiding redundant sequences. The **UniProt Archive (UniParc)** is a comprehensive repository, used to keep track of sequences and their identifiers. The **UniProt Metagenomic and Environmental Sequences (UniMES)** database is a repository specifically developed for metagenomic and environmental data.


The sequences and information in UniProt is accessible via [text search](#), [BLAST similarity search](#), and [FTP](#).



[European Bioinformatics Institute](#)



[Swiss Institute of Bioinformatics](#)



[Georgetown University](#)

Terms of Use | EBI Funding | Contact EBI | © European Bioinformatics Institute 2010. EBI is an Outstation of the European Molecular Biology Laboratory.

EBI UniProt Search Tools

<http://www.ebi.ac.uk/uniprot/search/SearchTools.html>

The screenshot shows the EBI UniProt Search Tools website. At the top, there is a navigation bar with the EMBL-EBI logo, an 'EB-eye Search' button, a dropdown menu for 'All Databases', a search input field with the placeholder 'Enter Text Here', and buttons for 'Go', 'Reset', and 'Advanced Search'. A 'Give us feedback' button is also present. Below the navigation bar is a menu with links for 'Databases', 'Tools', 'EBI Groups', 'Training', 'Industry', 'About Us', and 'Help'. On the right side of this menu, there are links for 'Site Index' and a RSS feed icon. A left sidebar contains a list of navigation links: 'UniProt Home', 'UniProt Website', 'About UniProt', 'Searches/Tools' (highlighted), 'UniSave', 'Submissions', 'Downloads', 'Documentation', 'Publications', 'People', and 'Help'. The main content area displays the breadcrumb 'EBI > Databases > Protein > UniProt' and a section titled 'UniProt Searches/Tools'. This section lists several search and analysis tools: 'Text Search' (with a sub-section 'Search the UniProt databases using:' and a bulleted list of links: 'UniProt text query [preferred]', 'SRS text query', and ''EB-eye' search for UniProt'), 'Blast' (described as 'Search UniProt databases using a protein sequence query.'), 'ClustalW2' (described as 'Align multiple protein sequences in UniProt.'), 'ID mapping' (described as 'Map external database identifiers to UniProt identifiers.'), 'UniProt Java API' (described as 'A Java library providing a remote API for accessing the UniProt data.'), 'UniProt BioMart' (described as 'Query and link the UniProt Knowledgebase with Ensembl and/or PRIDE data sets.'), and 'Useful Tools/Links' (described as 'Various tools and analyses available from the individual UniProt consortium member web sites and other sites that complement the UniProt Databases, including proteomics tools and multiple sequence alignments.').


 A  CONSORTIUM MEMBER
 Protein Information Resource

TLALPN----RKAVADHLLM
 LIGCLRNC SAVTAAAKQLAE
 VTGFSN----AKTTAQHVKK

Protein Search Site Search

About PIR Databases Search/Analysis Download Support

INTEGRATED PROTEIN INFORMATICS RESOURCE FOR GENOMIC, PROTEOMIC AND SYSTEMS BIOLOGY RESEARCH

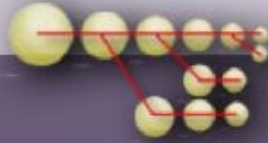


The Universal Protein Resource (UniProt) provides the scientific community with a single, centralized, authoritative resource for protein sequences and functional information.

UniProtKB | UniRef | UniParc

Current release: 15.12


PRO Protein Ontology



- Representation of protein objects with descriptions and relationships
- [Browse PRO](#)
- Annotate with [RACE-PRO](#)

[*Sample PRO report*](#)

iProClass Integrated Protein Knowledgebase



- Value-added reports for [UniProtKB](#) and unique [UniParc](#) proteins
- Functional analysis and [protein ID mapping](#)

[*Sample protein report*](#)

iProLINK Literature Information & Knowledge



- Source for text mining and ontology development
- [RLIMS-P](#) text mining tool, [BioThesaurus](#)
- [Bibliography mapping](#)

[*Sample Biblio. report*](#)

O OTHER RESOURCE

- [Proteomics](#): NIAID Biodefense Proteomics Admin. Center
- [PIR Grid-Enablement](#): Data node on NCI's [caBIG](#)

P PEPTIDE SEARCH

DATABASE: UniProtKB

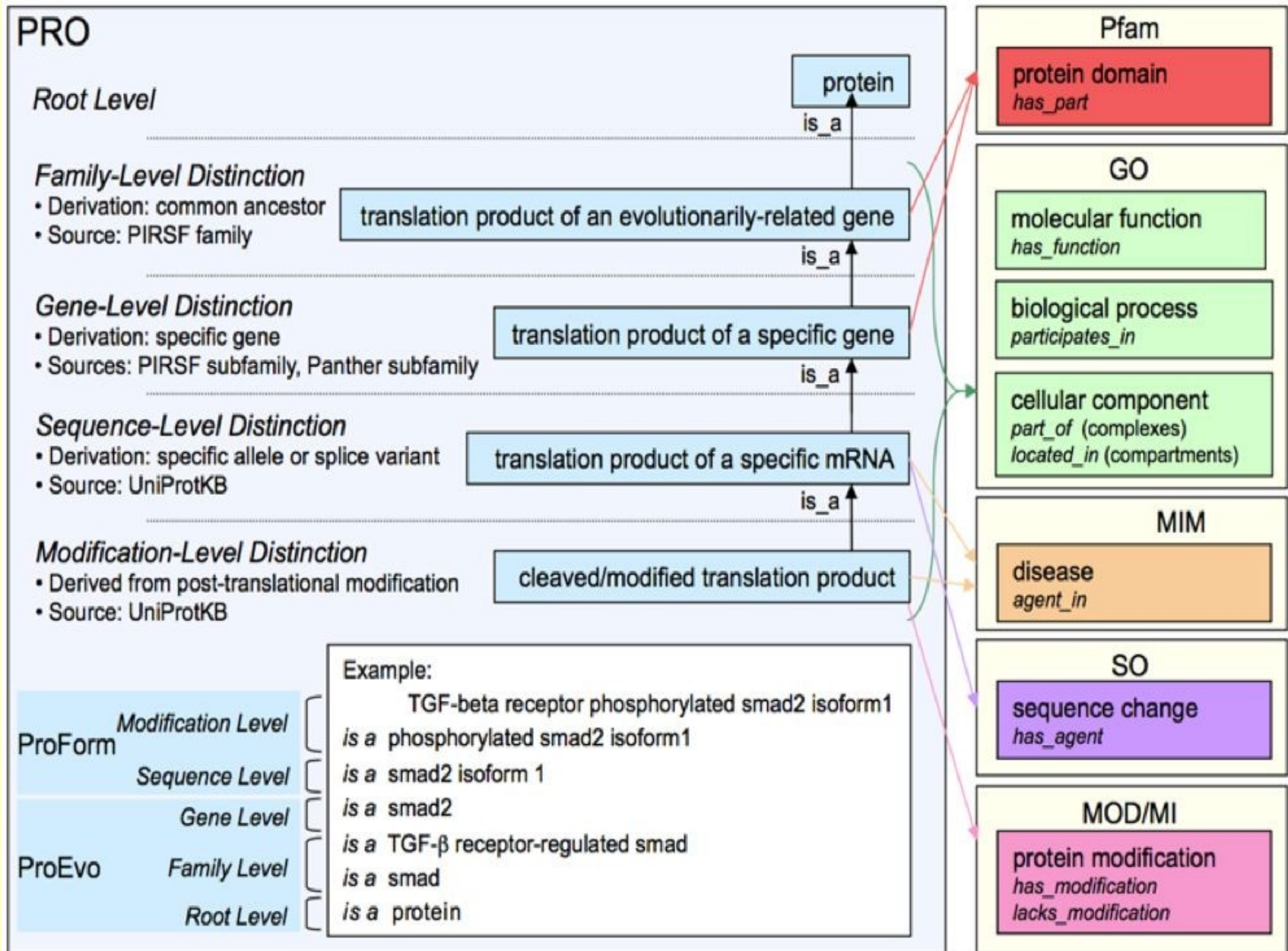
Use single letter amino acid code

T TEXT SEARCH

DATABASE: iProClass

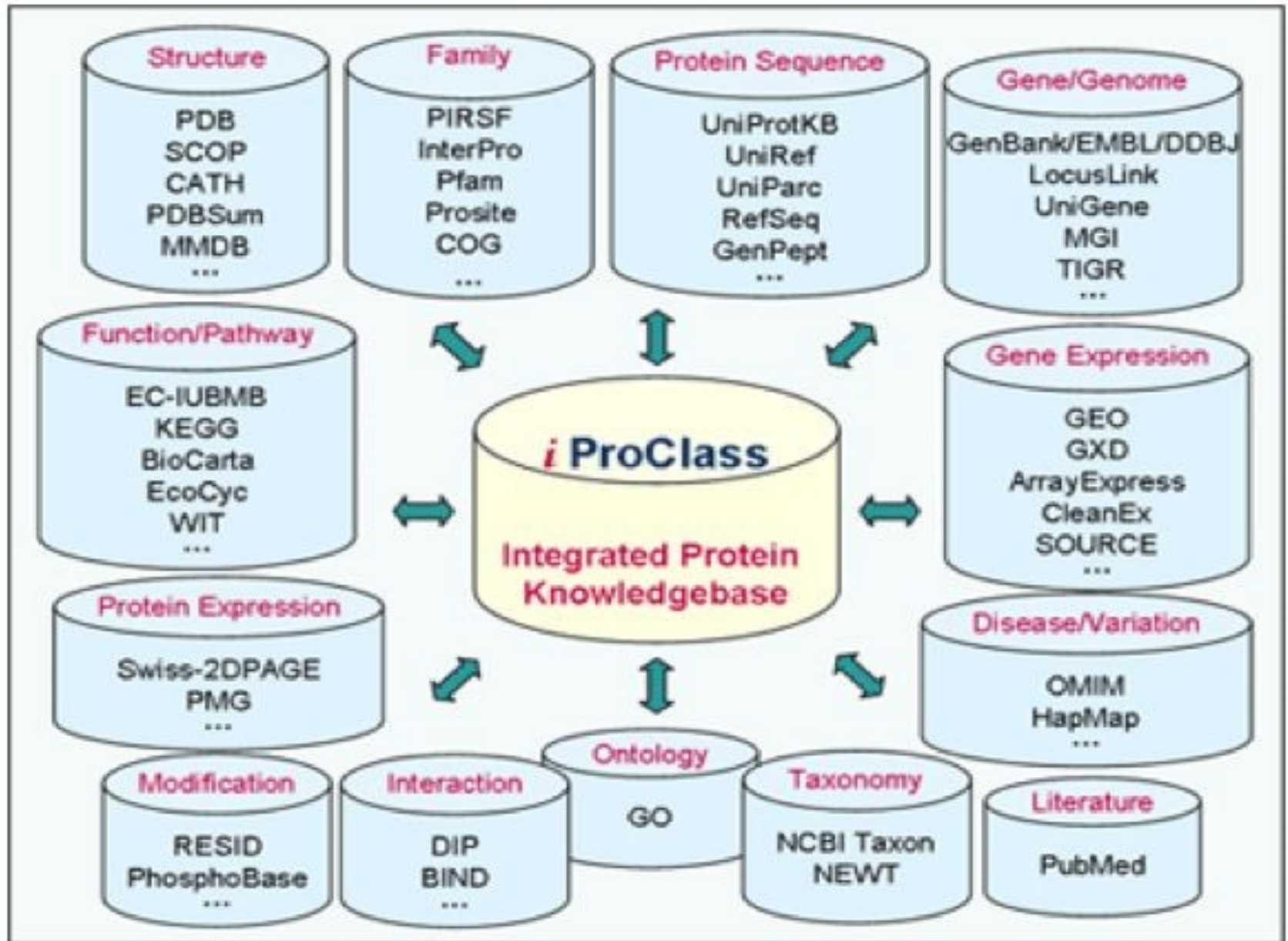
PIR PRO Database

<http://pir.georgetown.edu/pro/pro.shtml>



PIR iProClass Database

<http://pir.georgetown.edu/pirwww/dbinfo/iproclass.shtml>



NCBI Protein Database

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein>

NCBI

All Databases PubMed Nucleotide Protein Genome

Search Protein for Go

Limits Preview/Index History Clipboard Details

About Entrez

Entrez Protein
Help | FAQ

Entrez Tools

Check sequence revision history

LinkOut

My NCBI

Related resources
BLAST

Reference sequence project

Search for Genes

Clusters of orthologous groups

Protein reviews on the web

Search for full length cDNAs

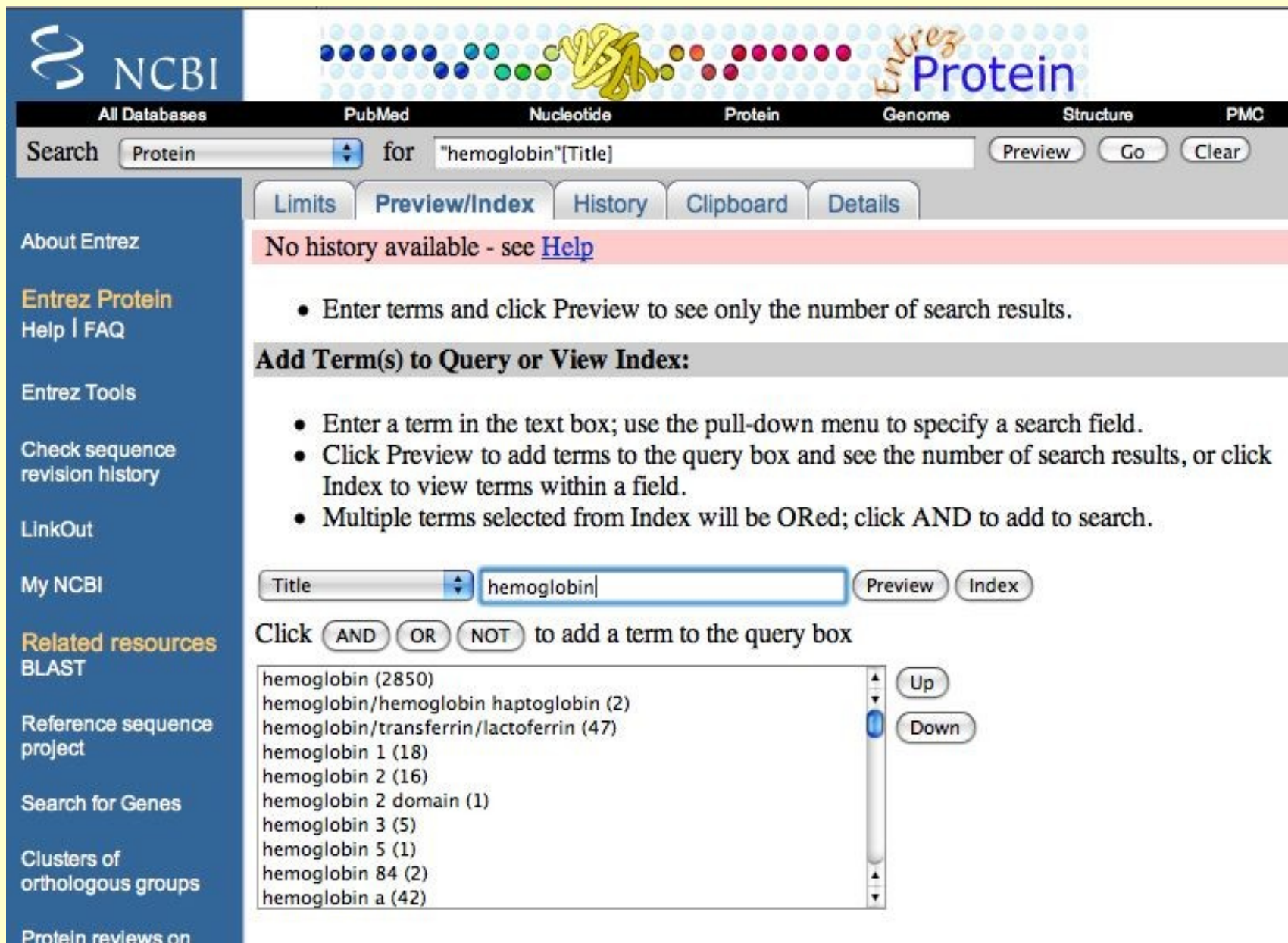
The protein entries in the Entrez search and retrieval system have been compiled from a variety of sources, including SwissProt, PIR, PRF, PDB, and translations from annotated coding regions in GenBank and RefSeq.

Human Genome
Explore [human genome resources](#) or browse the human genome sequence using the [Map Viewer](#).

Additional protein information
In addition to Protein sequences, other protein-related information is available via Entrez. Search the [Structure](#) database by choosing, "Structure" from the Entrez pull down menu, [Conserved Domains Database](#) (CDD) by choosing, "Domains", and [3D Domains](#) by choosing, the "3D Domains" option.

Retrieve taxonomy information
The Entrez protein database is cross-linked to the [Entrez taxonomy database](#). This allows you to find taxonomy information for the species from which a protein sequence was derived. First, look up a protein in Entrez. A "Taxonomy" link appears to the right of each entry that is linked to the Entrez taxonomy database. To view all non-redundant taxonomy links for a search result, select "Taxonomy Links" from the drop-down menu above the search results and click on the "Display" button to the left of that menu.

Hemoglobin in Title Index



NCBI Entrez Protein

All Databases PubMed Nucleotide Protein Genome Structure PMC

Search Protein for "hemoglobin"[Title] Preview Go Clear

Limits Preview/Index History Clipboard Details

No history available - see [Help](#)

- Enter terms and click Preview to see only the number of search results.

Add Term(s) to Query or View Index:

- Enter a term in the text box; use the pull-down menu to specify a search field.
- Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.
- Multiple terms selected from Index will be ORed; click AND to add to search.

Title hemoglobin Preview Index

Click AND OR NOT to add a term to the query box

hemoglobin (2850)	Up
hemoglobin/hemoglobin haptoglobin (2)	Down
hemoglobin/transferrin/lactoferrin (47)	
hemoglobin 1 (18)	
hemoglobin 2 (16)	
hemoglobin 2 domain (1)	
hemoglobin 3 (5)	
hemoglobin 5 (1)	
hemoglobin 84 (2)	
hemoglobin a (42)	

Protein reviews on

Hemoglobin Title & Human Organism

NCBI Entrez Protein

All Databases PubMed Nucleotide Protein Genome Structure PMC

Search Protein for "hemoglobin"[Title] AND "human"[Organism] Preview Go Clear

Limits Preview/Index History Clipboard Details

No history available - see [Help](#)

- Enter terms and click Preview to see only the number of search results.

Add Term(s) to Query or View Index:

- Enter a term in the text box; use the pull-down menu to specify a search field.
- Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.
- Multiple terms selected from Index will be ORed; click AND to add to search.


Organism Preview Index

Click AND OR NOT to add a term to the query box

human (418152)	Up
human adenovirus 1 (128)	Down
human adenovirus 1 isolate from a cat (3)	
human adenovirus 10 (17)	
human adenovirus 11 (152)	
human adenovirus 11a (4)	
human adenovirus 11p (6)	
human adenovirus 12 (139)	
human adenovirus 13 (14)	
human adenovirus 14 (63)	

Protein reviews on

Hemoglobin Title & Human Organism Results



NCBI Entrez Protein

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search Protein for "hemoglobin"[Title] AND "human"[Organism] Go Clear Save Search

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort by Relevance Send to

All: 645 Bacteria: 0 RefSeq: 1 Related Structures: 642

Items 1 - 20 of 645 Page 1 of 33 Next

- 1: [P69892](#) Reports BLink, Conserved Domains, Links
Hemoglobin subunit gamma-2 (Hemoglobin gamma-2 chain) (Gamma-2-globin) (Hemoglobin gamma-G chain) (Hb F Ggamma)
gil56749861|splP69892.2|HBG2_HUMAN[56749861]
- 2: [P69891](#) Reports BLink, Conserved Domains, Links
Hemoglobin subunit gamma-1 (Hemoglobin gamma-1 chain) (Gamma-1-globin) (Hemoglobin gamma-A chain) (Hb F Agamma)
gil56749860|splP69891.2|HBG1_HUMAN[56749860]
- 3: [P69905](#) Reports BLink, Conserved Domains, Links
Hemoglobin subunit alpha (Hemoglobin alpha chain) (Alpha-globin)
gil57013850|splP69905.2|HBA_HUMAN[57013850]
- 4: [P68871](#) Reports BLink, Conserved Domains, Links
Hemoglobin subunit beta (Hemoglobin beta chain) (Beta-globin) [Contains: LVV-hemorphin-7]
gil56749856|splP68871.2|HBB_HUMAN[56749856]
- 5: [P02100](#) Reports BLink, Conserved Domains, Links
Hemoglobin subunit epsilon (Hemoglobin epsilon chain) (Epsilon-globin)
gil122726|splP02100.2|HBE_HUMAN[122726]
- 6: [P02042](#) Reports BLink, Conserved Domains, Links
Hemoglobin subunit delta (Hemoglobin delta chain) (Delta-globin)
gil122713|splP02042.2|HBD_HUMAN[122713]
- 7: [P02008](#) Reports BLink, Conserved Domains, Links
Hemoglobin subunit zeta (Hemoglobin zeta chain) (Zeta-globin) (HBAZ)
gil122335|splP02008.2|HBAZ_HUMAN[122335]



UCSC Proteome Browser

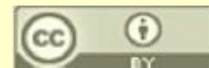
<http://genome.ucsc.edu/cgi-bin/pbGateway>

[Home](#) [Genomes](#) [Blat](#) [Tables](#) [Gene Sorter](#) [PCR](#) [Session](#) [FAQ](#) [Help](#)

UCSC Proteome Browser Gateway

The UCSC Proteome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).
Software Copyright (c) The Regents of the University of California. All rights reserved.

Enter a gene symbol or a Swiss-Prot/TrEMBL protein ID:



UCSC Hemoglobin Protein

<http://genome.ucsc.edu/cgi-bin/pbGateway>

[Home](#)

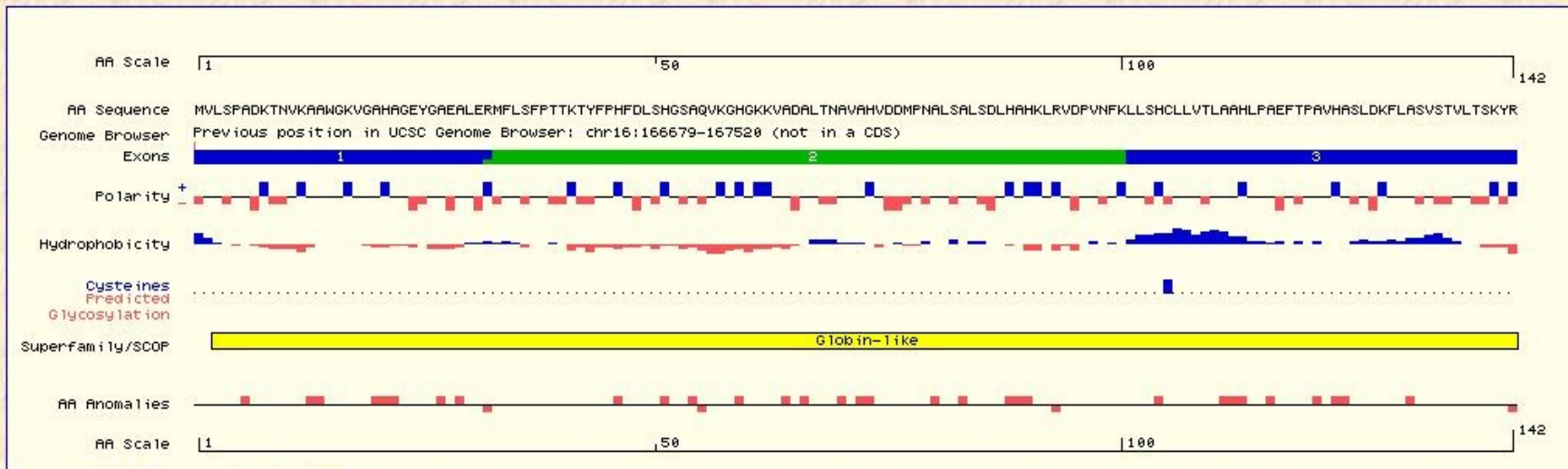
UCSC Proteome Browser

[PDF/PS](#)

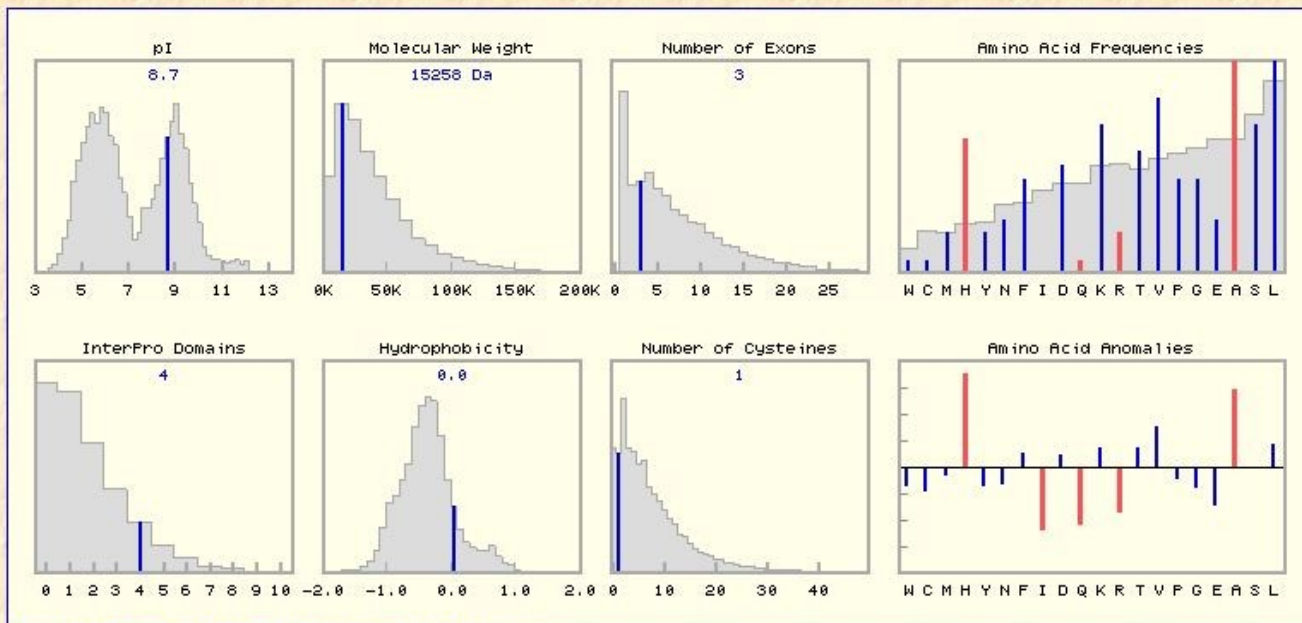
[New Query](#)

Protein **P69905** (aka HBA_HUMAN) Hemoglobin subunit alpha (Hemoglobin alpha chain) (Alpha-globin).
Organism: Homo sapiens (human)

Move <<< << < > >> >>> Current scale: FULL Rescale to



Explanation of Protein Tracks



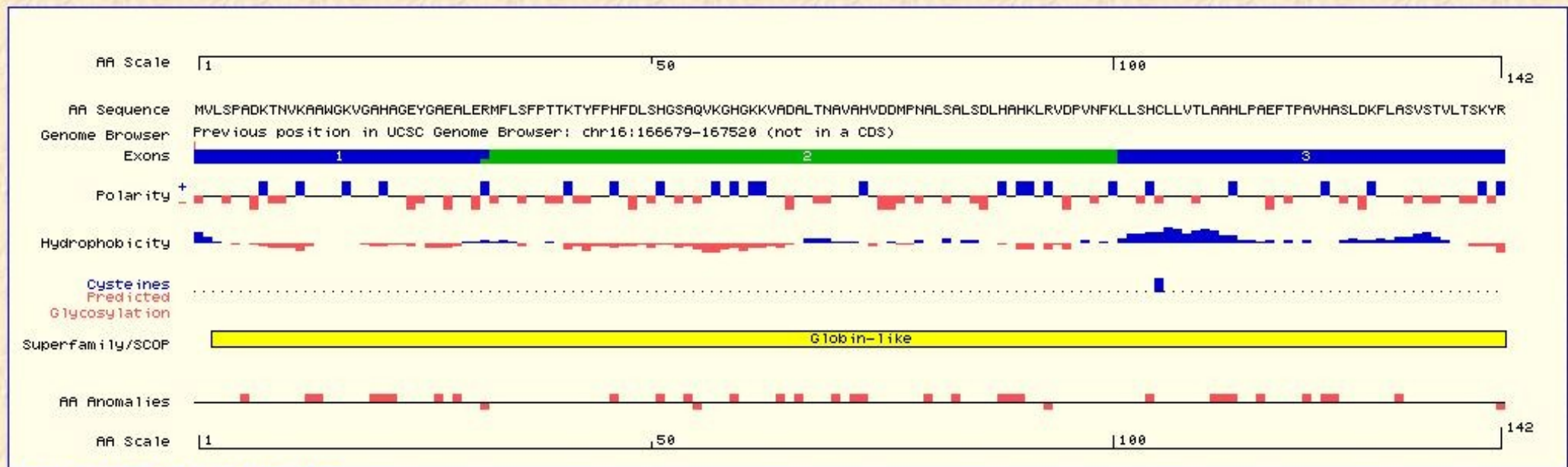
Explanation of Protein Property Histograms

UCSC Hemoglobin Coding Region

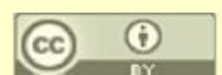
<http://genome.ucsc.edu/cgi-bin/pbGateway>

Protein **P69905** (aka HBA_HUMAN) Hemoglobin subunit alpha (Hemoglobin alpha chain) (Alpha-globin).
Organism: Homo sapiens (human)

Move <<< << < > >> >>> Current scale: FULL Rescale to 1/6 1/2 FULL DNA

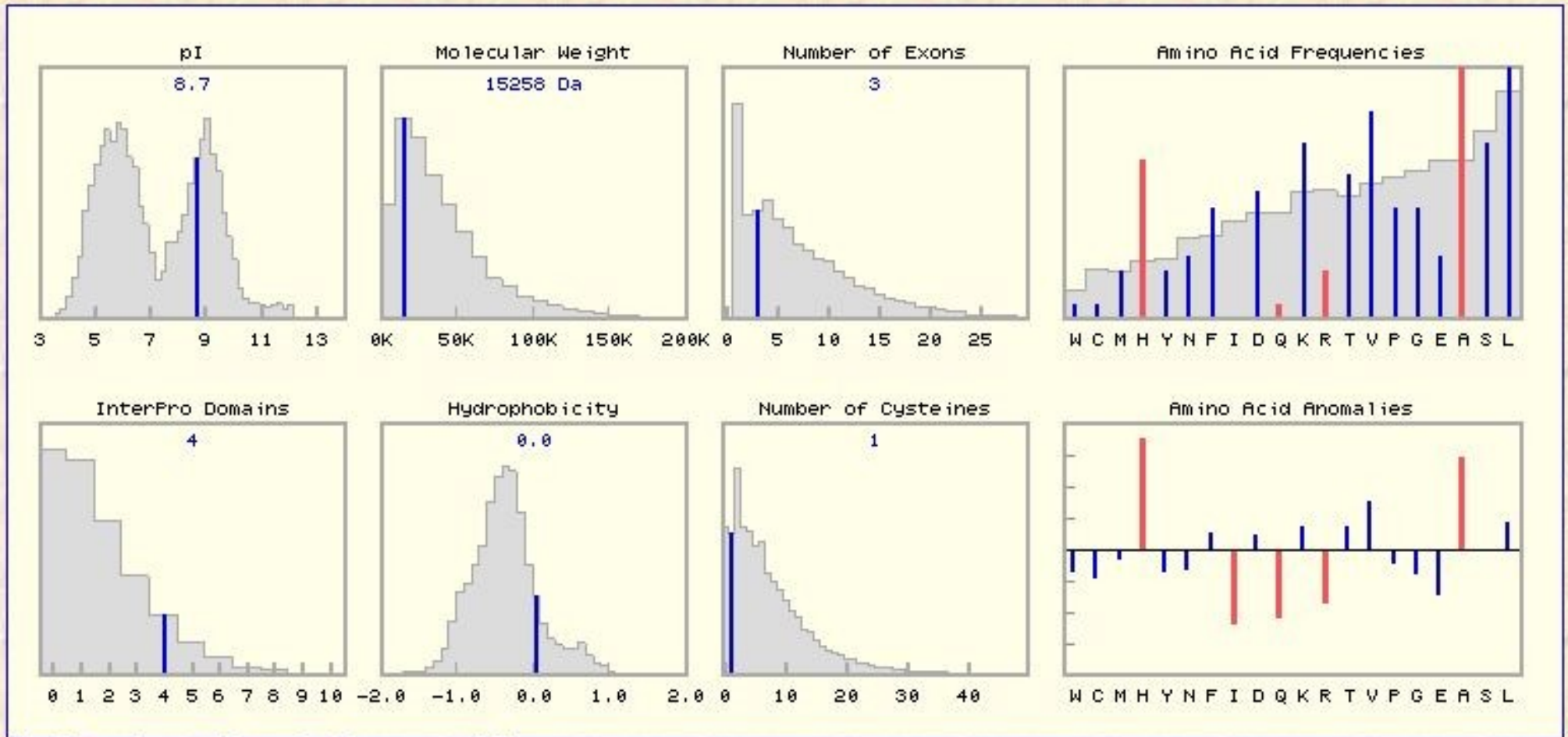


[Explanation of Protein Tracks](#)



UCSC Hemoglobin Protein Properties

<http://genome.ucsc.edu/cgi-bin/pbGateway>



Explanation of Protein Property Histograms

UCSC Hemoglobin Gene Region

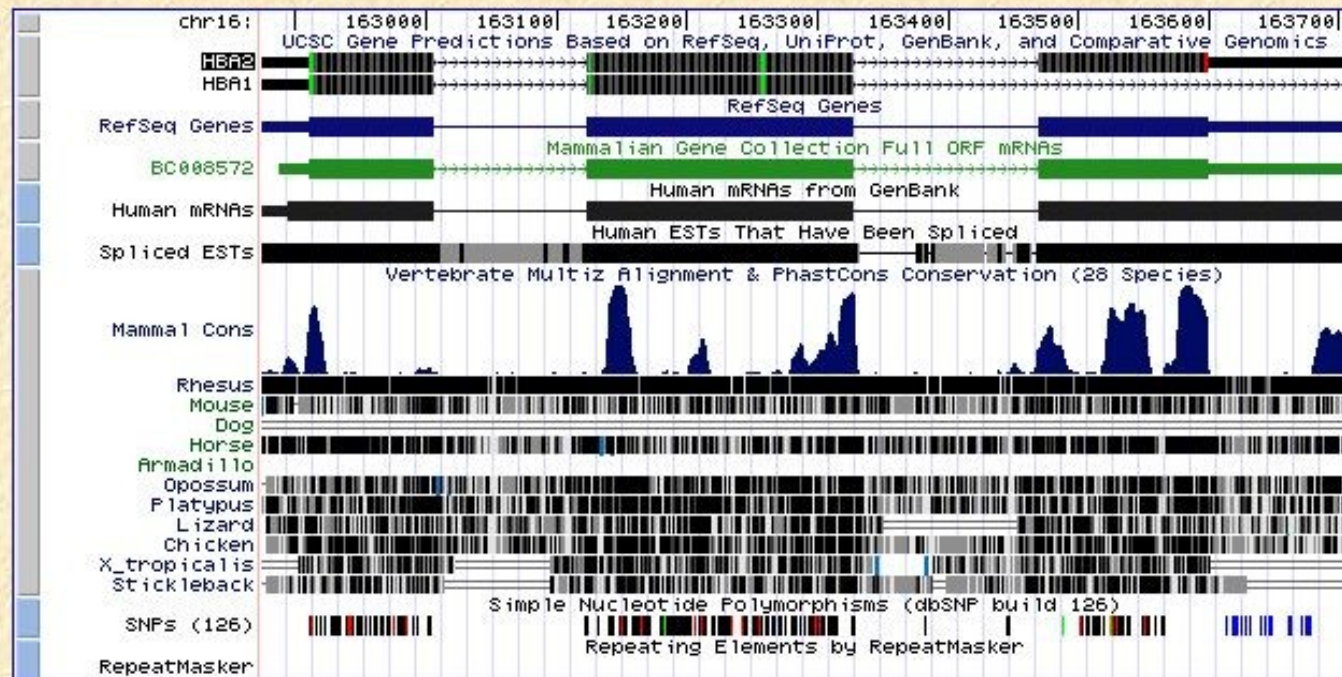
<http://genome.ucsc.edu/cgi-bin/hgGateway?org=Human&db=hg18>

UCSC Genome Browser on Human Mar. 2006 Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chr16:162,875-163,705 jump clear size 831 bp. configure

chr16 (p13.3) 13.3 12.1 p11.2 12.1 16q21



move start
< 2.0 >

Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.

move end
< 2.0 >

default tracks hide all add custom tracks configure refresh

Use drop down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes.



UCSC Hemoglobin Links

<http://genome.ucsc.edu/cgi-bin/hgGateway?org=Human&db=hg18>

[Home](#) [Genomes](#) [Genome Browser](#) [Blat](#) [Tables](#) [Gene Sorter](#) [PCR](#) [Session](#) [FAQ](#) [Help](#)

Human Gene HBA2 (uc002cfv.1) Description and Page Index

Description: alpha 2 globin

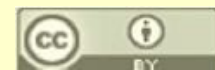
RefSeq Summary (NM_000517): The human alpha globin gene cluster located on chromosome 16 spans about 30 kb and includes seven loci: 5'- zeta - pseudozeta - mu - pseudoalpha-1 - alpha-2 - alpha-1 - theta - 3'. The alpha-2 (HBA2) and alpha-1 (HBA1) coding sequences are identical. These genes differ slightly over the 5' untranslated regions and the introns, but they differ significantly over the 3' untranslated regions. Two alpha chains plus two beta chains constitute HbA, which in normal adult life comprises about 97% of the total hemoglobin; alpha chains combine with delta chains to constitute HbA-2, which with HbF (fetal hemoglobin) makes up the remaining 3% of adult hemoglobin. Alpha thalassemias result from deletions of each of the alpha genes as well as deletions of both HBA2 and HBA1; some nondeletion alpha thalassemias have also been reported.

Strand: + **Genomic Size:** 831 **Exon Count:** 3 **Coding Exon Count:** 3

Page Index	Sequence and Links	UniProt Comments	Microarray	RNA Structure	Protein Structure
Other Species	GO Annotations	mRNA Descriptions	Other Names	Model Information	Methods

- Sequence and Links to Tools and Databases

Genomic Sequence (chr16:162,875-163,705)		mRNA (may differ from genome)		Protein (142 aa)	
Gene Sorter	Genome Browser	Proteome Browser	Table Schema	CGAP	Ensembl
Entrez Gene	ExonPrimer	GeneCards	GeneTests	Gepis Tissue	H-INV
HGNC	HPRD	Jackson Labs	OMIM	PubMed	Stanford SOURCE
Treefam	UniProt				



Resources

[NCBI Home](#)

[All Resources \(A-Z\)](#)

[Literature](#)

[DNA & RNA](#)

[Proteins](#)

[Sequence Analysis](#)

[Genes & Expression](#)

[Genomes](#)

[Maps & Markers](#)

[Domains & Structures](#)

[Genetics & Medicine](#)

[Taxonomy](#)

[Data & Software](#)

[Training & Tutorials](#)

[Homology](#)

[Small Molecules](#)

[Variation](#)

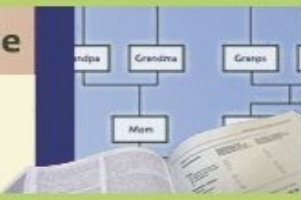
Welcome to NCBI

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

[More about the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [RSS](#)

Genotype and Phenotype

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



▶ 1 2 3 4

How To...

- Obtain the full text of an article
- Retrieve all sequences for an organism or taxon
- Find a homolog for a gene in another organism
- Find genes associated with a phenotype or disease
- Design PCR primers and check them for specificity
- Find the function of a gene or gene product
- Determine conserved synteny between the genomes of two organisms

[See all ...](#)

NLM/NCBI H1N1 Flu Resources

Popular Resources

- [PubMed](#)
- [PubMed Central](#)
- [Bookshelf](#)
- [BLAST](#)
- [Gene](#)
- [Nucleotide](#)
- [Protein](#)
- [GEO](#)
- [Conserved Domains](#)
- [Structure](#)
- [PubChem](#)

NCBI News

[November and October News](#) 02 Dec 2009

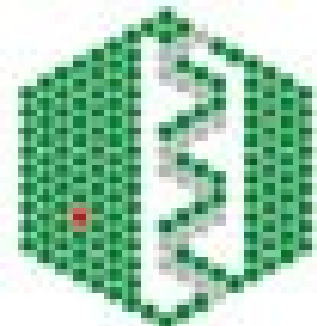
Featured: New Discovery-oriented PubMed and NCBI Homepage. T...

[NCBI News - September 2009](#) 05 Oct 2009

The September 2009 issue of the NCBI News is available ...

[NCBI News - August 2009](#) 19 Aug 2009

The August 2009 issue of the NCBI News is available online. ...



European Bioinformatics Institute

<http://www.ebi.ac.uk/>

EMBL-EBI  EB-eye Search

Databases Tools EBI Groups Training Industry About Us Help Site Index  

Data Resources & Tools

- EMBL-BANK
- UniProt
- ArrayExpress
- Ensembl
- InterPro
- PDBe
- Genomes
- Nucleotide Sequences
- Protein Sequences
- Macromolecular Structures
- Small Molecules
- Gene Expression
- Molecular Interactions
- Reactions & Pathways
- Protein Families
- Enzymes
- Literature
- Taxonomy
- Ontologies
- Patent Resources
- Sequence Similarity & Analysis
- Pattern & Motif Searches
- Structure Analysis
- Text Mining
- Downloads
- Web Services



European Bioinformatics Institute

About the EBI

- Research
- PhD Studies
- Training
- Industry Support
- Group & Team Leaders
- EBI Funders
- User Support
- EBI Mission
- People
- Events at the EBI
- Genome Campus Events
- How to Find us
- Jobs

EBI Hosted Project Websites

- 1000 Genomes
- BioCatalogue
- BioSapiens
- E-MeP
- EGA
- ELIXIR
- EMBRACE
- EMERALD
- ENFIN
- FELICS
- IMPACT
- INSDC
- LRG
- SPINE
- SYMBIOmatics

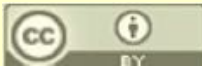
Latest News

▪ **Open access drug discovery database launches with half a million compounds**
 18 January 2010
[ChEMBLdb](#), a vast online database of information on the properties and activities of drugs and drug-like small molecules and their targets, launches today with information on over half a million compounds. The data lie at the heart of translating information from the human genome into successful new drugs in the clinic... [more](#)

Research Highlights

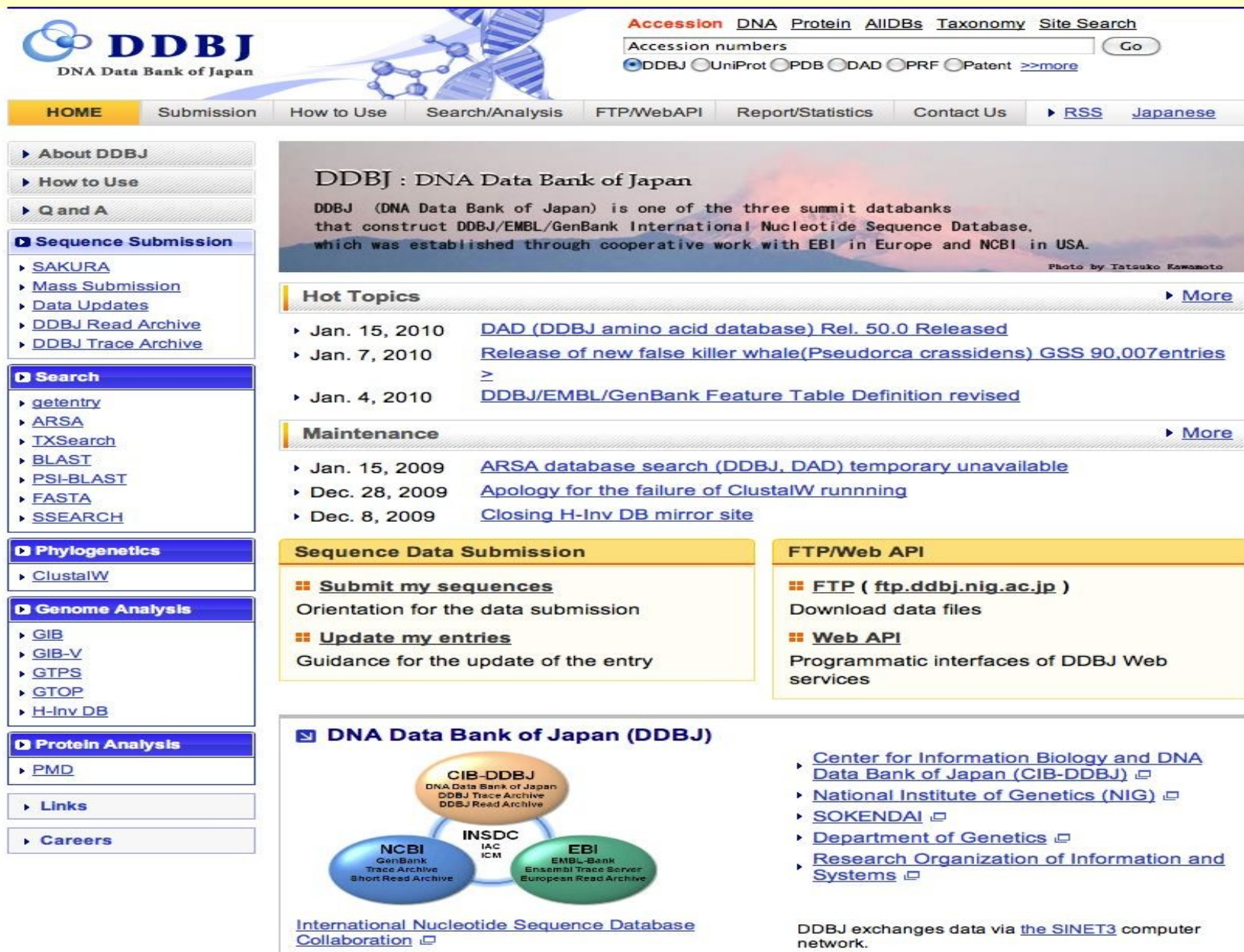
▪ **EMBL-EBI articles are top of the list**
 20 November 2009
 Articles on three resources hosted by EMBL-EBI ([PDBe](#), [Ensembl Genomes](#) and [Gene Expression Atlas](#)) are highlighted as featured articles in the latest Database issue of *Nucleic Acids Research*. Featured articles are selected by the journal's Executive Editors based upon their originality, significance and scientific excellence ... [more](#)

Events



DNA Database of Japan

<http://www.ddbj.nig.ac.jp/>



DDBJ
DNA Data Bank of Japan

Accession DNA Protein AIDBs Taxonomy Site Search
 Accession numbers
 DDBJ UniProt PDB DAD PRF Patent >>more

HOME Submission How to Use Search/Analysis FTP/WebAPI Report/Statistics Contact Us [RSS](#) [Japanese](#)

- About DDBJ
- How to Use
- Q and A
- Sequence Submission**
 - SAKURA
 - Mass Submission
 - Data Updates
 - DDBJ Read Archive
 - DDBJ Trace Archive
- Search**
 - getentry
 - ARSA
 - TXSearch
 - BLAST
 - PSI-BLAST
 - FASTA
 - SSEARCH
- Phylogenetics**
 - ClustalW
- Genome Analysis**
 - GIB
 - GIB-V
 - GTPS
 - GTOP
 - H-Inv DB
- Protein Analysis**
 - PMD
- Links
- Careers

DDBJ : DNA Data Bank of Japan
 DDBJ (DNA Data Bank of Japan) is one of the three summit databanks that construct DDBJ/EMBL/GenBank International Nucleotide Sequence Database, which was established through cooperative work with EBI in Europe and NCBI in USA.
Photo by Tatsuko Kawamoto

Hot Topics [More](#)

- Jan. 15, 2010 [DAD \(DDBJ amino acid database\) Rel. 50.0 Released](#)
- Jan. 7, 2010 [Release of new false killer whale\(Pseudorca crassidens\) GSS 90,007 entries](#)
- Jan. 4, 2010 [DDBJ/EMBL/GenBank Feature Table Definition revised](#)

Maintenance [More](#)

- Jan. 15, 2009 [ARSA database search \(DDBJ, DAD\) temporary unavailable](#)
- Dec. 28, 2009 [Apology for the failure of ClustalW running](#)
- Dec. 8, 2009 [Closing H-Inv DB mirror site](#)


Sequence Data Submission

- Submit my sequences**
Orientation for the data submission
- Update my entries**
Guidance for the update of the entry

FTP/Web API

- FTP (<ftp.ddbj.nig.ac.jp>)**
Download data files
- Web API**
Programmatic interfaces of DDBJ Web services

DNA Data Bank of Japan (DDBJ)



The diagram shows the International Nucleotide Sequence Database Collaboration (INSDC) structure. At the top is CIB-DDBJ (DNA Data Bank of Japan, DDBJ Trace Archive, DDBJ Read Archive). Below it are three main partners: NCBI (GenBank, Trace Archive, Short Read Archive), EBI (EMBL-Bank, Ensembl Trace Server, European Read Archive), and INSDC (IAC, ICM). Arrows indicate data exchange between these entities.

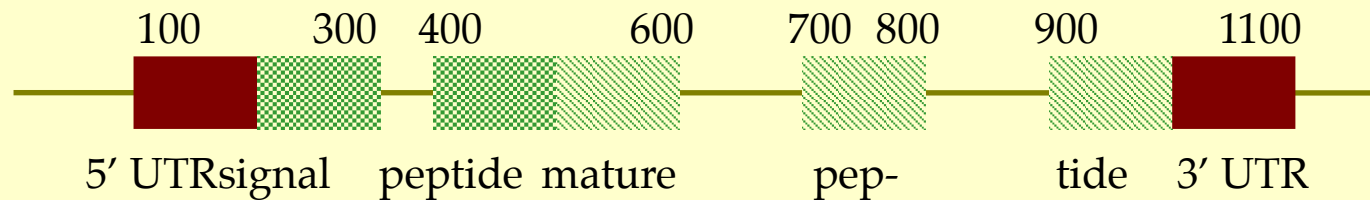
- [Center for Information Biology and DNA Data Bank of Japan \(CIB-DDBJ\)](#)
- [National Institute of Genetics \(NIG\)](#)
- [SOKENDAI](#)
- [Department of Genetics](#)
- [Research Organization of Information and Systems](#)

International Nucleotide Sequence Database Collaboration

DDBJ exchanges data via the SINET3 computer network.

DDBJ / EMBL / GenBank Feature Table

Example: Protein-coding region Prototypical eukaryotic gene



FEATURES	Location/Qualifiers
	100..200
exon	100..300 /number
intron	301..400 /number=1
exon	401..600 /number=2
intron	601..700 /number=2
exon	701..800 /number=3
intron	801..900 /number=3
3'UTR	1002..1100
exon	901..1100
	/number=4
signal-peptide	join(201..300,401..501)
mature-peptide	join(502..600,701..800,901..1001)
	/product="prototypical protein"
CDS	join(201..300,401..600,701..800,901..1001)
	/product="prototypical protein"
mRNA	join(100..300,401..600,701..800,901..1100)
prim_transcript	100..1100

International Common Fields

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

International Nucleotide Sequence Database Collaboration

The collaboration that exists among the International Nucleotide Sequence Databases has led to many beneficial projects that promise to proliferate in the molecular biology community.

Currently, the following projects are part of the collaborative effort among the three databases:

The Taxonomy Project

One of the goals of the collaborators is to use a unified taxonomy across all databases, largely one based on sequence information. The [taxonomy project](#) was set up as a tool for biologists worldwide, and also as a shared instrument for the collaborators. This is one of the important resources used for the maintenance of [Genetic Codes](#), important for the correct translation of coding sequences.

The Feature Table

The [Feature Table](#) documentation represents the shared rules that allow the three databases to exchange data on a daily basis. The Feature Table represent the vocabulary that is used to describe the DNA sequence annotations as well as that of the protein sequence(s) they encode. The copy we present here is a mirror of the original document prepared by [EBI](#).

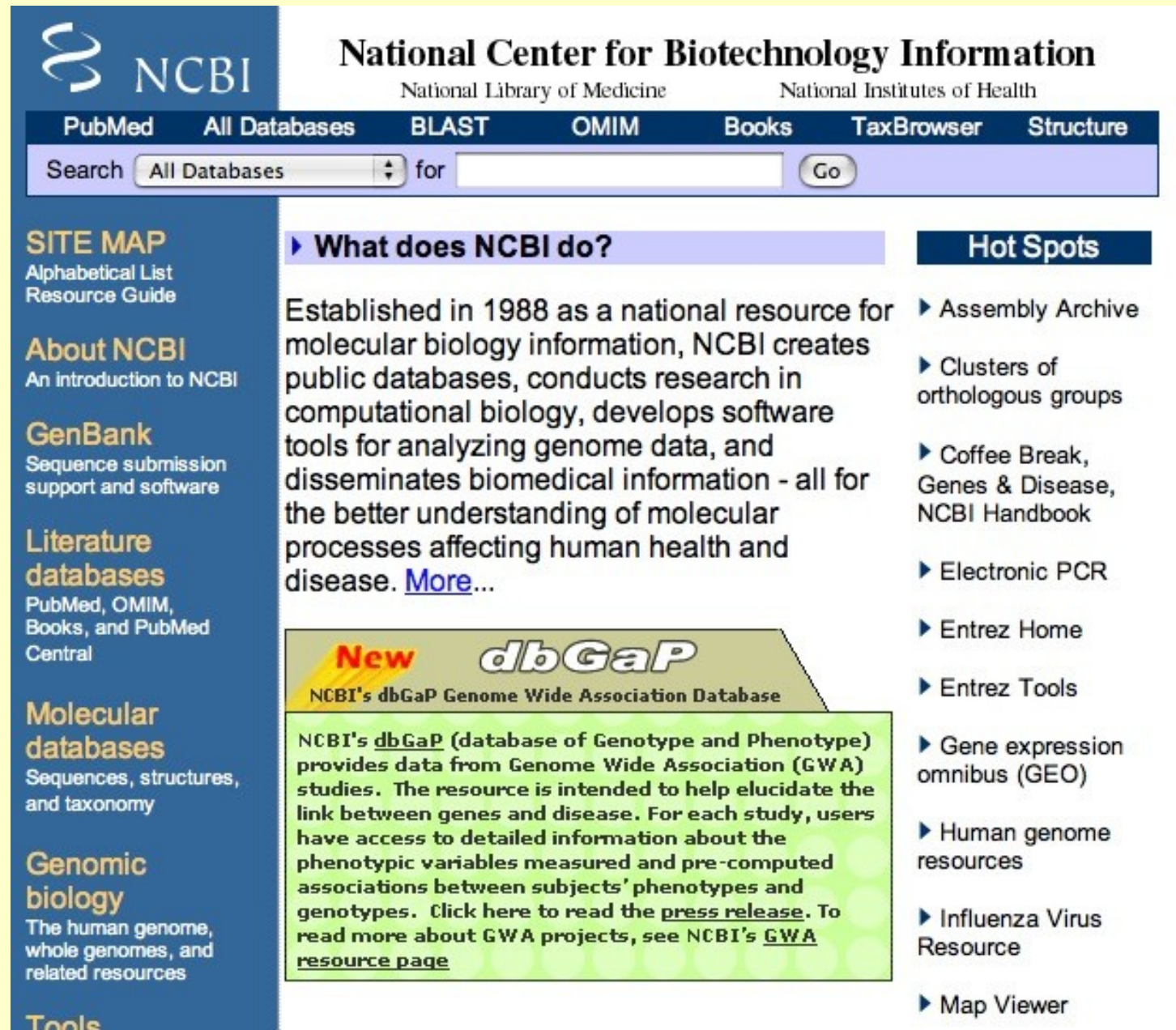
This documentation is also available as a series of postscript files from [EBI](#) (in Europe) or [NCBI](#) (in the USA).

The db_xref Qualifier

A new qualifier was recently added to the Feature Table definition ([/db_xref](#)) that allows the nucleotide databases to explicitly reference specific sequences (protein sequences) or other identifiers within other databases.

The country Qualifier

A new qualifier was recently added to the Feature Table definition ([/country](#)) to indicate the country of origin of a DNA sample



NCBI
National Library of Medicine National Institutes of Health

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search for

SITE MAP
Alphabetical List
Resource Guide

About NCBI
An introduction to NCBI

GenBank
Sequence submission
support and software

**Literature
databases**
PubMed, OMIM,
Books, and PubMed
Central

**Molecular
databases**
Sequences, structures,
and taxonomy

**Genomic
biology**
The human genome,
whole genomes, and
related resources

Tools

What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)

New dbGaP
NCBI's dbGaP Genome Wide Association Database

NCBI's **dbGaP** (database of Genotype and Phenotype) provides data from Genome Wide Association (GWA) studies. The resource is intended to help elucidate the link between genes and disease. For each study, users have access to detailed information about the phenotypic variables measured and pre-computed associations between subjects' phenotypes and genotypes. [Click here to read the press release.](#) To read more about GWA projects, see NCBI's [GWA resource page](#)

Hot Spots

- ▶ Assembly Archive
- ▶ Clusters of orthologous groups
- ▶ Coffee Break, Genes & Disease, NCBI Handbook
- ▶ Electronic PCR
- ▶ Entrez Home
- ▶ Entrez Tools
- ▶ Gene expression omnibus (GEO)
- ▶ Human genome resources
- ▶ Influenza Virus Resource
- ▶ Map Viewer

Entrez Databases

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



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

Search across databases

GO

CLEAR

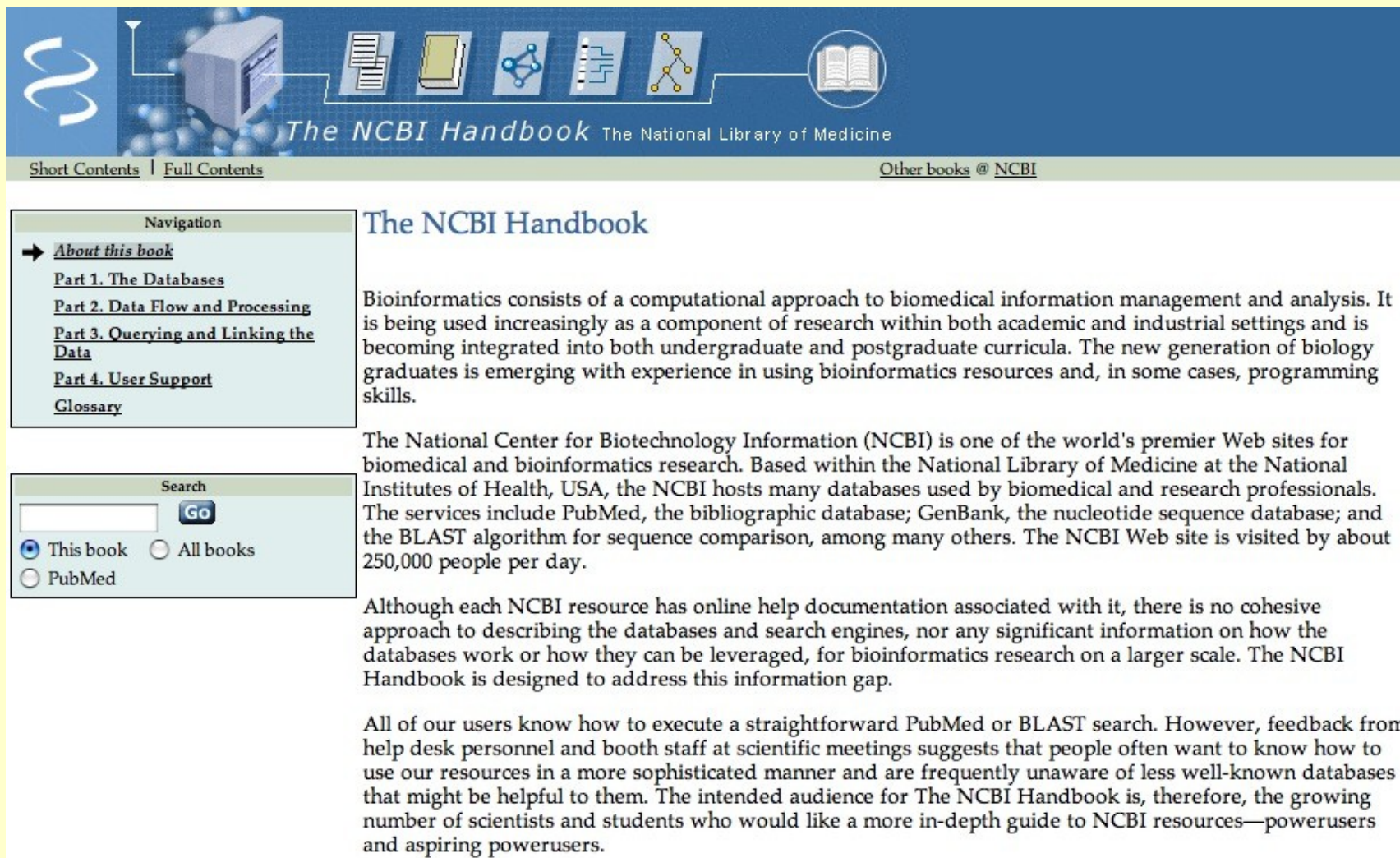
Help

Welcome to the Entrez cross-database search page

PubMed: biomedical literature citations and abstracts	Books: online books
PubMed Central: free, full text journal articles	OMIM: online Mendelian Inheritance in Man
Site Search: NCBI web and FTP sites	OMIA: online Mendelian Inheritance in Animals
Nucleotide: sequence database (includes GenBank)	UniGene: gene-oriented clusters of transcript sequences
Protein: sequence database	CDD: conserved protein domain database
Genome: whole genome sequences	3D Domains: domains from Entrez Structure
Structure: three-dimensional macromolecular structures	UniSTS: markers and mapping data
Taxonomy: organisms in GenBank	PopSet: population study data sets
SNP: single nucleotide polymorphism	GEO Profiles: expression and molecular abundance profiles
Gene: gene-centered information	GEO DataSets: experimental sets of GEO data
HomoloGene: eukaryotic homology groups	Cancer Chromosomes: cytogenetic databases
PubChem Compound: unique small molecule chemical structures	PubChem BioAssay: bioactivity screens of chemical substances
PubChem Substance: deposited chemical substance records	GENSAT: gene expression atlas of mouse central nervous system
Genome Project: genome project information	Probe: sequence-specific reagents
dbGaP: genotype and phenotype	

NCBI Handbook

<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook>



The screenshot shows the NCBI Handbook website. At the top, there is a navigation bar with the NCBI logo, a search bar, and icons for various resources. Below the navigation bar, there are links for "Short Contents" and "Full Contents", and a link for "Other books @ NCBI". The main content area is titled "The NCBI Handbook" and contains a navigation menu on the left, a search box, and a main text area.

Navigation

- [About this book](#)
- [Part 1. The Databases](#)
- [Part 2. Data Flow and Processing](#)
- [Part 3. Querying and Linking the Data](#)
- [Part 4. User Support](#)
- [Glossary](#)

Search

This book All books
 PubMed

The NCBI Handbook

Bioinformatics consists of a computational approach to biomedical information management and analysis. It is being used increasingly as a component of research within both academic and industrial settings and is becoming integrated into both undergraduate and postgraduate curricula. The new generation of biology graduates is emerging with experience in using bioinformatics resources and, in some cases, programming skills.

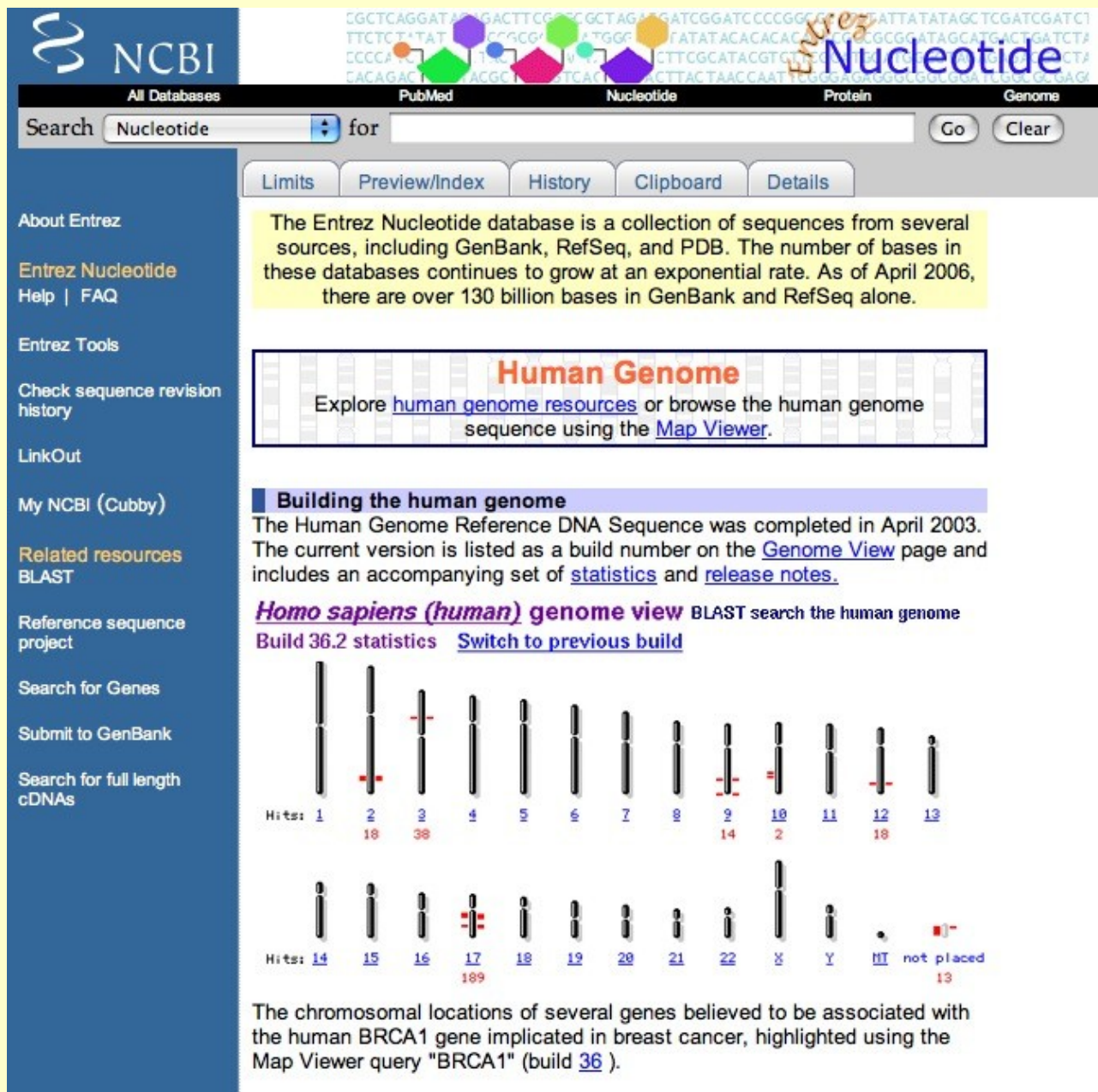
The National Center for Biotechnology Information (NCBI) is one of the world's premier Web sites for biomedical and bioinformatics research. Based within the National Library of Medicine at the National Institutes of Health, USA, the NCBI hosts many databases used by biomedical and research professionals. The services include PubMed, the bibliographic database; GenBank, the nucleotide sequence database; and the BLAST algorithm for sequence comparison, among many others. The NCBI Web site is visited by about 250,000 people per day.

Although each NCBI resource has online help documentation associated with it, there is no cohesive approach to describing the databases and search engines, nor any significant information on how the databases work or how they can be leveraged, for bioinformatics research on a larger scale. The NCBI Handbook is designed to address this information gap.

All of our users know how to execute a straightforward PubMed or BLAST search. However, feedback from help desk personnel and booth staff at scientific meetings suggests that people often want to know how to use our resources in a more sophisticated manner and are frequently unaware of less well-known databases that might be helpful to them. The intended audience for The NCBI Handbook is, therefore, the growing number of scientists and students who would like a more in-depth guide to NCBI resources—powerusers and aspiring powerusers.

NCBI Nucleotide Sequence Database

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide>



The screenshot shows the NCBI Nucleotide database search results for the query "BRCA1". The interface includes a search bar with "Nucleotide" selected, a "Go" button, and a "Clear" button. Below the search bar are tabs for "Limits", "Preview/Index", "History", "Clipboard", and "Details".

The main content area features a yellow box with the following text: "The Entrez Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The number of bases in these databases continues to grow at an exponential rate. As of April 2006, there are over 130 billion bases in GenBank and RefSeq alone."

Below this is a blue box titled "Human Genome" with the text: "Explore [human genome resources](#) or browse the human genome sequence using the [Map Viewer](#)."

The next section is titled "Building the human genome" and contains the text: "The Human Genome Reference DNA Sequence was completed in April 2003. The current version is listed as a build number on the [Genome View](#) page and includes an accompanying set of [statistics](#) and [release notes](#)."

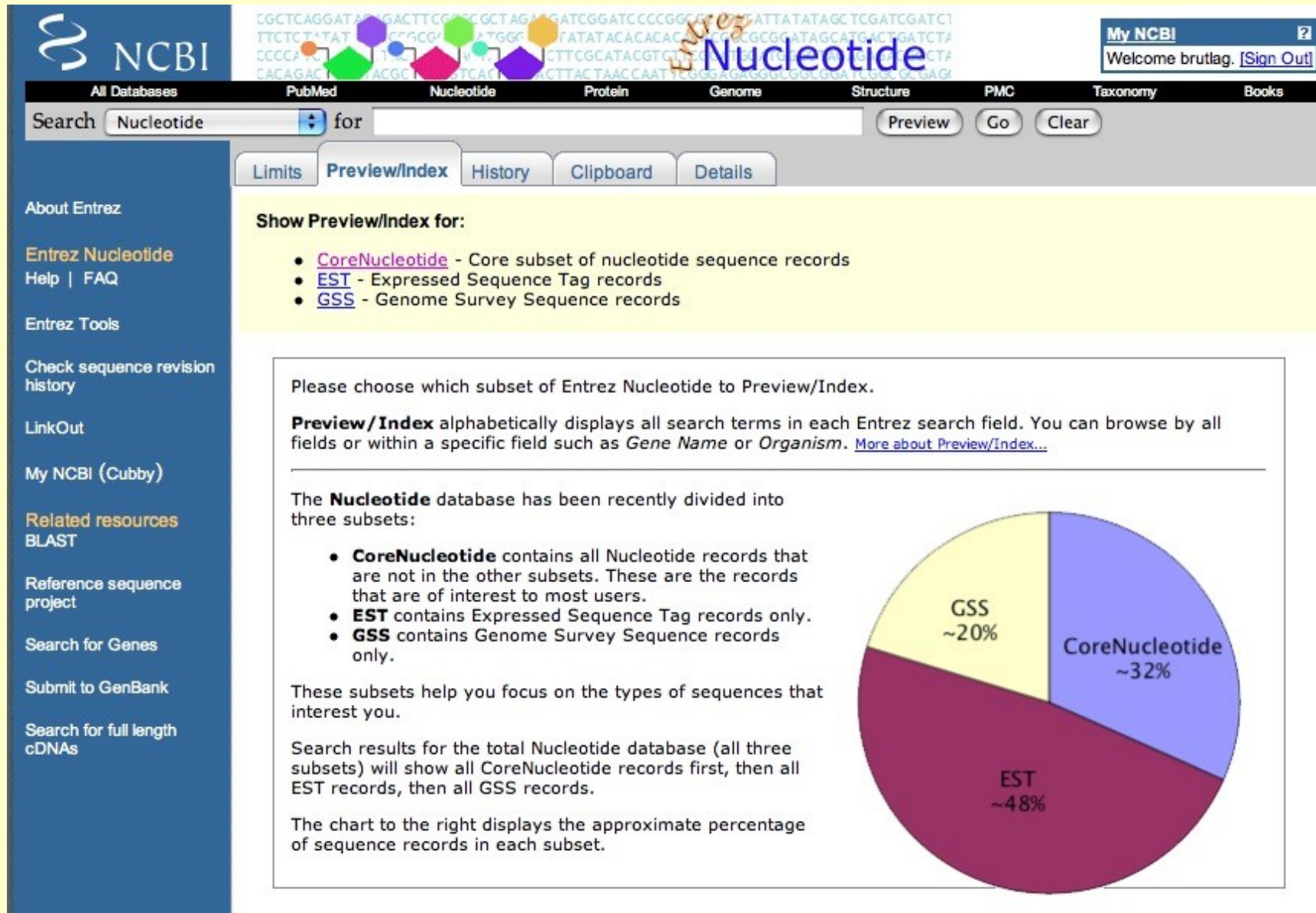
Below this is a link: "[Homo sapiens \(human\) genome view](#) BLAST search the human genome".

The next section is titled "Build 36.2 statistics" and contains the text: "[Switch to previous build](#)".

The main visual element is a chromosome map showing the locations of genes associated with the human BRCA1 gene. The map consists of two rows of chromosomes, labeled "Hits" 1 through 13 in the first row and 14 through 22, X, Y, and "not placed" in the second row. Red bars on the chromosomes indicate the locations of genes. The BRCA1 gene is highlighted in red on chromosome 17.

The text below the map reads: "The chromosomal locations of several genes believed to be associated with the human BRCA1 gene implicated in breast cancer, highlighted using the Map Viewer query 'BRCA1' (build 36)."

NCBI Indexes for Core DNA Sequences



The screenshot shows the NCBI Entrez Nucleotide search page. The search bar contains "Nucleotide" and "for". Below the search bar are buttons for "Limits", "Preview/Index", "History", "Clipboard", and "Details". The "Preview/Index" button is selected. The page displays a list of search options under "Show Preview/Index for:":

- [CoreNucleotide](#) - Core subset of nucleotide sequence records
- [EST](#) - Expressed Sequence Tag records
- [GSS](#) - Genome Survey Sequence records

Please choose which subset of Entrez Nucleotide to Preview/Index.

Preview/Index alphabetically displays all search terms in each Entrez search field. You can browse by all fields or within a specific field such as *Gene Name* or *Organism*. [More about Preview/Index...](#)

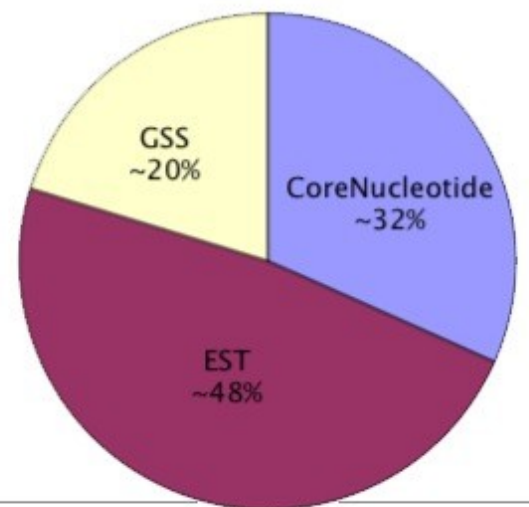
The **Nucleotide** database has been recently divided into three subsets:

- **CoreNucleotide** contains all Nucleotide records that are not in the other subsets. These are the records that are of interest to most users.
- **EST** contains Expressed Sequence Tag records only.
- **GSS** contains Genome Survey Sequence records only.

These subsets help you focus on the types of sequences that interest you.

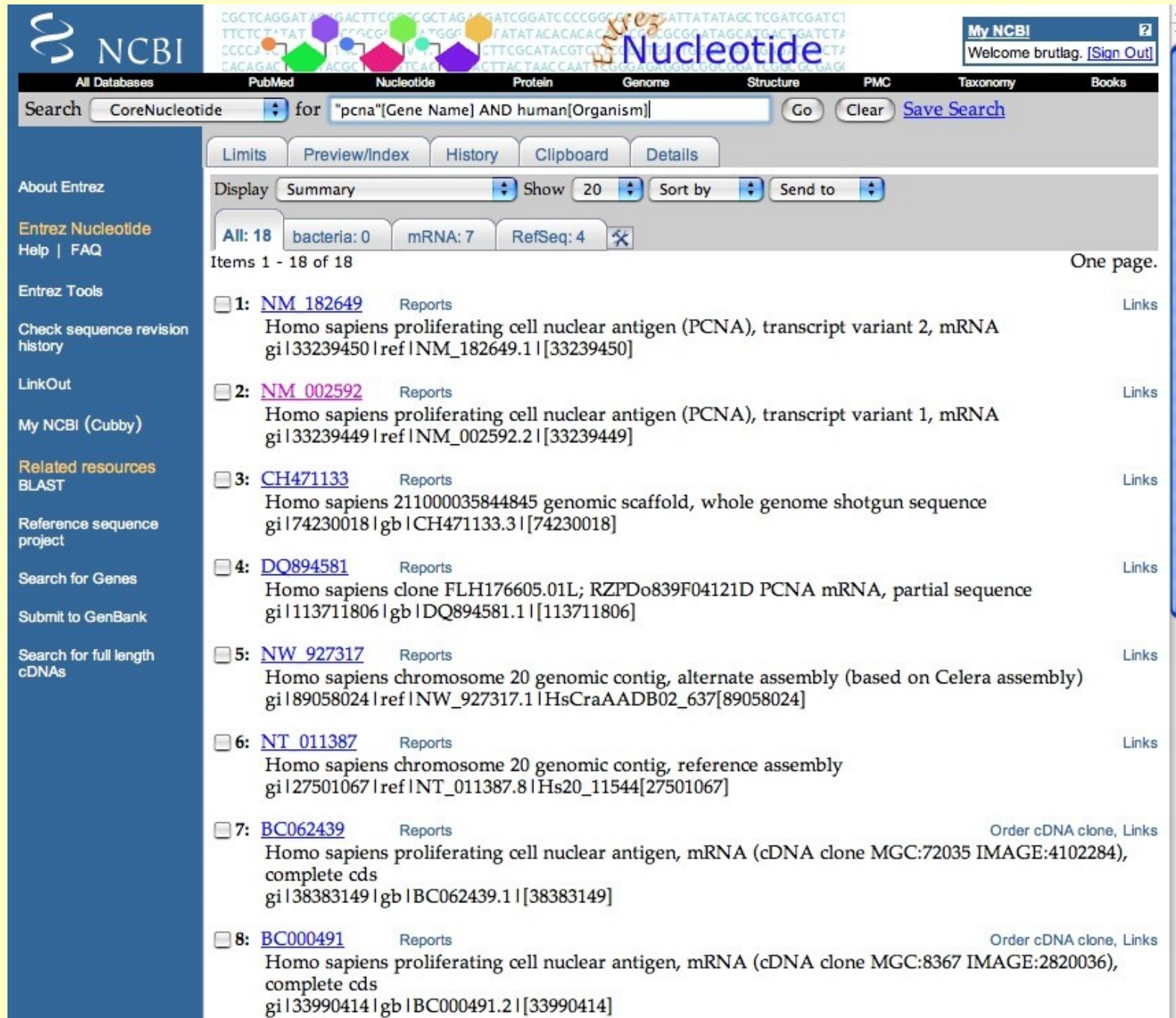
Search results for the total Nucleotide database (all three subsets) will show all CoreNucleotide records first, then all EST records, then all GSS records.

The chart to the right displays the approximate percentage of sequence records in each subset.



Subset	Approximate Percentage
EST	~48%
CoreNucleotide	~32%
GSS	~20%

NCBI Search for PCNA Gene Sequences



The screenshot shows the NCBI Entrez Nucleotide search interface. The search query is "pcna"[Gene Name] AND human[Organism]. The search results display 18 items, with the first 8 items listed below. Each item includes a checkbox, a link to the accession number, the number of reports, and a brief description of the sequence.

Search: CoreNucleotide for "pcna"[Gene Name] AND human[Organism]

Display: Summary Show 20 Sort by Send to

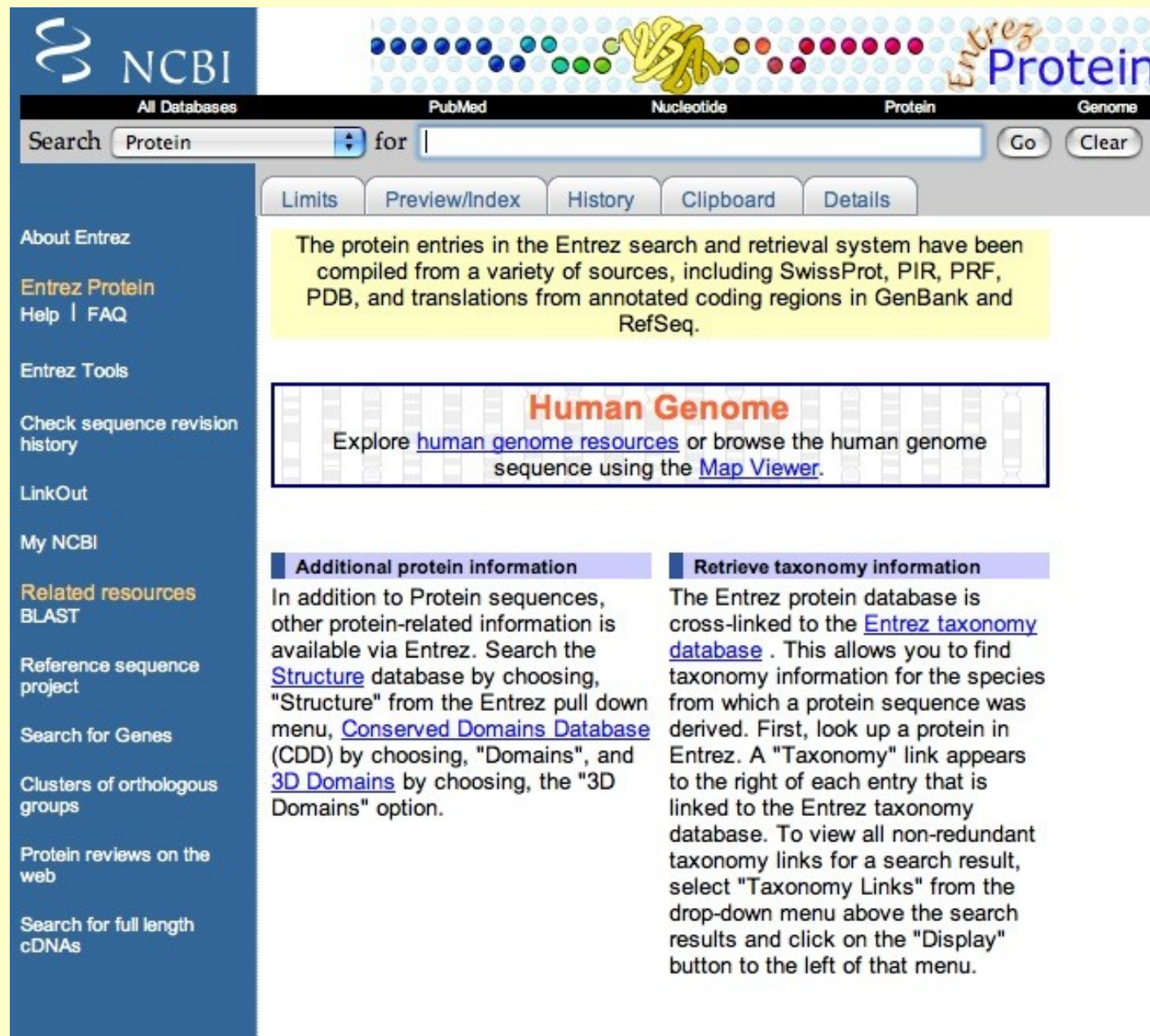
Results: All: 18 bacteria: 0 mRNA: 7 RefSeq: 4

Items 1 - 18 of 18 One page.

- 1: [NM_182649](#) Reports Links
Homo sapiens proliferating cell nuclear antigen (PCNA), transcript variant 2, mRNA
gi|33239450|ref|NM_182649.1|[33239450]
- 2: [NM_002592](#) Reports Links
Homo sapiens proliferating cell nuclear antigen (PCNA), transcript variant 1, mRNA
gi|33239449|ref|NM_002592.2|[33239449]
- 3: [CH471133](#) Reports Links
Homo sapiens 211000035844845 genomic scaffold, whole genome shotgun sequence
gi|74230018|gb|CH471133.3|[74230018]
- 4: [DQ894581](#) Reports Links
Homo sapiens clone FLH176605.01L; RZPDo839F04121D PCNA mRNA, partial sequence
gi|113711806|gb|DQ894581.1|[113711806]
- 5: [NW_927317](#) Reports Links
Homo sapiens chromosome 20 genomic contig, alternate assembly (based on Celera assembly)
gi|89058024|ref|NW_927317.1|HsCraAADB02_637[89058024]
- 6: [NT_011387](#) Reports Links
Homo sapiens chromosome 20 genomic contig, reference assembly
gi|27501067|ref|NT_011387.8|Hs20_11544[27501067]
- 7: [BC062439](#) Reports Order cDNA clone, Links
Homo sapiens proliferating cell nuclear antigen, mRNA (cDNA clone MGC:72035 IMAGE:4102284),
complete cds
gi|38383149|gb|BC062439.1|[38383149]
- 8: [BC000491](#) Reports Order cDNA clone, Links
Homo sapiens proliferating cell nuclear antigen, mRNA (cDNA clone MGC:8367 IMAGE:2820036),
complete cds
gi|33990414|gb|BC000491.2|[33990414]

NCBI Protein Sequence Database

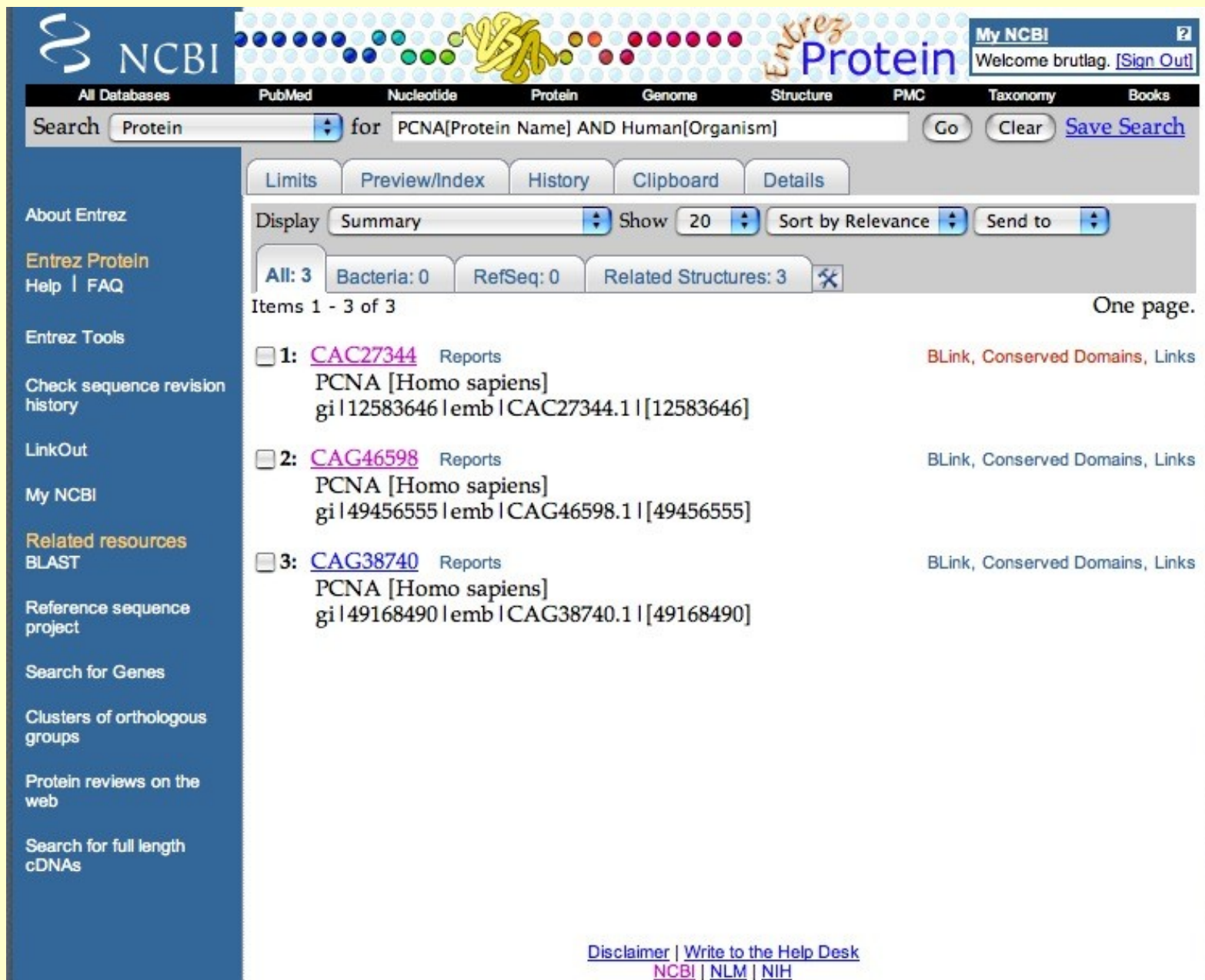
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein>



The screenshot shows the NCBI Protein Sequence Database homepage. At the top, there is a navigation bar with tabs for "All Databases", "PubMed", "Nucleotide", "Protein", and "Genome". Below this is a search bar with the text "Search Protein for" and buttons for "Go" and "Clear". To the left of the search bar is a vertical menu with various links such as "About Entrez", "Entrez Protein Help | FAQ", "Entrez Tools", "Check sequence revision history", "LinkOut", "My NCBI", "Related resources BLAST", "Reference sequence project", "Search for Genes", "Clusters of orthologous groups", "Protein reviews on the web", and "Search for full length cDNAs".

The main content area features a yellow box with the text: "The protein entries in the Entrez search and retrieval system have been compiled from a variety of sources, including SwissProt, PIR, PRF, PDB, and translations from annotated coding regions in GenBank and RefSeq." Below this is a blue box with the heading "Human Genome" and the text: "Explore [human genome resources](#) or browse the human genome sequence using the [Map Viewer](#)." At the bottom, there are two columns of text. The left column is titled "Additional protein information" and discusses the availability of protein-related information via Entrez, mentioning the "Structure" database, "Conserved Domains Database (CDD)", and "3D Domains". The right column is titled "Retrieve taxonomy information" and explains how the Entrez protein database is cross-linked to the "Entrez taxonomy database", providing instructions on how to find taxonomy information for a protein sequence.

NCBI Protein Search for Human PCNA




The screenshot shows the NCBI Protein search results page. The search query is "Protein" for "PCNA[Protein Name] AND Human[Organism]". The results are displayed in a table with 3 items. Each item includes a checkbox, a link to the protein entry, the protein name, and the organism. The results are:

Item	Protein ID	Protein Name	Organism	Accession	Links
<input type="checkbox"/> 1:	CAC27344	PCNA	[Homo sapiens]	gi 12583646 emb CAC27344.1 [12583646]	BLink, Conserved Domains, Links
<input type="checkbox"/> 2:	CAG46598	PCNA	[Homo sapiens]	gi 49456555 emb CAG46598.1 [49456555]	BLink, Conserved Domains, Links
<input type="checkbox"/> 3:	CAG38740	PCNA	[Homo sapiens]	gi 49168490 emb CAG38740.1 [49168490]	BLink, Conserved Domains, Links

At the bottom of the page, there are links for [Disclaimer](#), [Write to the Help Desk](#), [NCBI](#), [NLM](#), and [NIH](#).

BLAST Link for Human PCNA


[BLAST](#)
[Protein](#)
[Structure](#)
[PubMed](#)
[Taxonomy](#)

[Genome](#)
[Nucleotide](#)
[3D-Domains](#)
[Books](#)
[Help](#)

Query: gi | [12583646](#) PCNA [Homo sapiens]
 Matching gi: [119630835](#), [119630834](#), [119630833](#), [114680824](#), [109092795](#), [109092793](#), [61680063](#), [61680062](#), [61680061](#), [61359492](#), [61358193](#), [60593816](#), [60593815](#), [60593760](#), [60593759](#), [60593758](#), [60593756](#), [60593754](#), [60593752](#), [60593750](#), [60593748](#), [60593746](#), [58176900](#), [58176898](#), [58176896](#), [58176894](#), [55651058](#), [47117084](#), [38383150](#), [33239451](#), [21902516](#), [13540338](#), [12653441](#), [4505641](#), [2914383](#), [387005](#), [181272](#), [129694](#)

COG0592 assigned by Cognitor (7 best hits)

[Show identical](#)
[Best hits](#)
[Common Tree](#)
[Taxonomy Report](#)
[3D structures](#)
[CDD-Search](#)
[GI list](#)
[Run BLAST](#)

200 BLAST hits to 142 unique species [Sort by taxonomy proximity](#)

18 [Archaea](#)
0 [Bacteria](#)
64 [Metazoa](#)
26 [Fungi](#)
27 [Plants](#)
15 [Viruses](#)
48 [Other Eukaryotae](#)

Keep only Cut-Off [Select](#) [Reset](#) New search by GI: [Go](#)

261 aa

	SCORE	P	ACCESSION	GI	PROTEIN DESCRIPTION
Conserved Domain Database hits					
=====	1316	1	AAX43349	61369547	proliferating cell nuclear antigen [synthetic construct]
=====	1313	1	AAX36355	60815711	proliferating cell nuclear antigen [synthetic construct]
=====	1311	31	1AXCE	2914387	Chain E, Human Pona
=====	1311	21	XP_534355	73991708	PREDICTED: similar to proliferating cell nuclear antigen [Canis :
=====	1308	31	CAG38740	49168490	PCNA [Homo sapiens]
=====	1306	21	NP_001...	77735939	hypothetical protein LOC515499 [Bos taurus]
=====	1303	31	CAG46598	49456555	PCNA [Homo sapiens]
=====	1302	22	AAH60570	38014559	Proliferating cell nuclear antigen [Rattus norvegicus]
=====	1300	22	P57761	13124447	Proliferating cell nuclear antigen (PCNA)
=====	1288	22	BAE38082	74222336	unnamed protein product [Mus musculus]
=====	1284	22	BAB28355	15919908	unnamed protein product [Mus musculus]
=====	1284	22	CAA37243	53600	unnamed protein product [Mus musculus]
=====	1254	18	NP_989501	45383776	proliferating cell nuclear antigen [Gallus gallus]
=====	1245	18	Q9DDF1	20139157	Proliferating cell nuclear antigen (PCNA)
=====	1220	17	NP_001...	56118632	pcna protein [Xenopus tropicalis]
=====	1213	17	AAH41549	27371153	MGC53867 protein [Xenopus laevis]
=====	1213	15	ABK51289	117617401	proliferating cellular nuclear antigen [Astatotilapia burtoni]
=====	1212	15	Q9PTP1	66774199	Proliferating cell nuclear antigen (PCNA)
=====	1211	17	AAH57758	35505407	Proliferating cell nuclear antigen [Xenopus laevis]
=====	1203	17	AAA49926	214647	proliferating cell nuclear antigen (PCNA)
=====	1198	15	CAG07797	47226638	unnamed protein product [Tetraodon nigroviridis]
=====	1196	22	BAB28436	12849670	unnamed protein product [Mus musculus]
=====	1191	15	Q9W644	66773944	Proliferating cell nuclear antigen (PCNA)
=====	1189	15	NP_571479	18859223	proliferating cell nuclear antigen [Danio rerio]
=====	1147	17	ABD79020	89892452	proliferating cell nuclear antigen [Pseudacris regilla]
=====	1023	17	AAV73840	56117860	proliferating cell nuclear antigen [Rana catesbeiana]
=====	1015	10	BAE47144	78271817	proliferating cell nuclear antigen [Batrachus proximus]

Conserved Domain Link for PCNA



View PSSM: cd00577

Aligned: 78 rows

PSSM: 248 columns

Status: curated CD

Created: 7-Mar-2002

Updated: 4-May-2005

Structure:

Show Structure

Program: Cn3D

Drawing: All Atoms

Aligned Rows: up to 10

[\[Download Cn3D\]](#)

Hierarchy:

Interactive Display

Program: CDTree v3.0

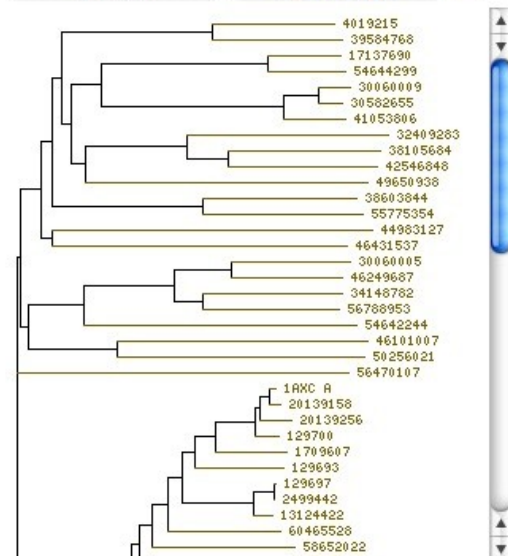
Add Query Sequence:

[\[Download CDTree\]](#)

cd00577 is part of a hierarchy of related CD models.
Use the graphical representation to navigate this hierarchy.

cd00577 Sequence Cluster

Zoom Out Detailed View



Sub-family Hierarchy

Interactive Display with CDTree

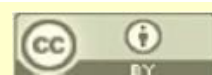
cd00577 PCNA

Other Related Conserved Domains: [pfa00705](#) [pfa02144](#) [pfa02747](#) [pfa04139](#)

Reformat Sequence Alignment Format: Compact Hypert Row Display: up to 10 Color Bits: 2.0 bits Type Selection: the most similar members

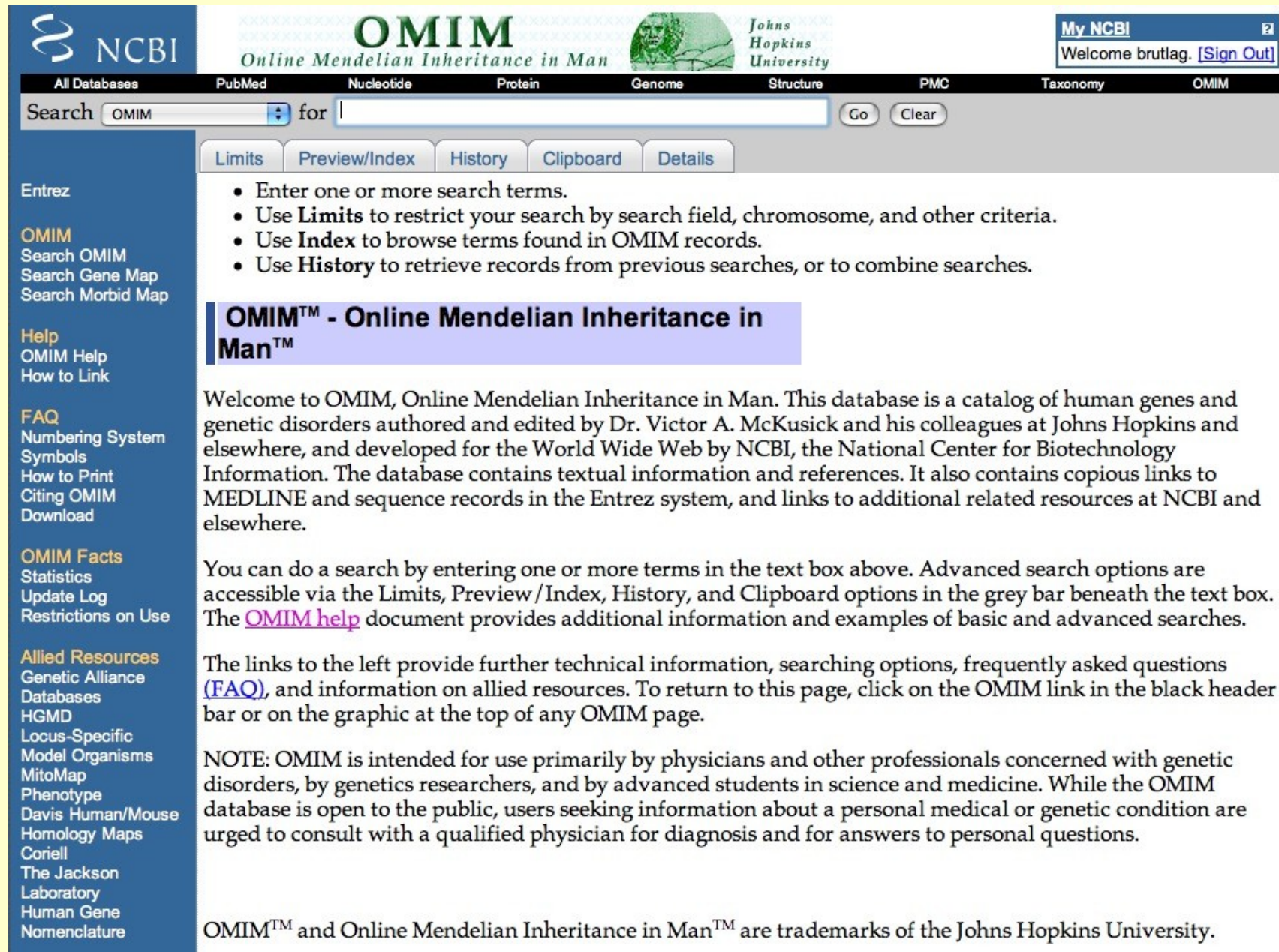
Feature 1

1AXC_A	4	ARLVQGSILKKVLEALKDL	LINEACWDIS	[1]	.SGVN	LQSM	DSS	[3]	.LVQL	TLR	[3]	.FD	TYRC	[1]	.RNL	AMGVN	71
query	4	ARLVQGSILKKVLEALKDL	LINEACWDIS	[1]	.SGVN	LQSM	DSS	[3]	.LVQL	TLR	[3]	.FD	TYRC	[1]	.RNL	AMGVN	71
gi 129693	4	ARLGQATILKKILDAIKDL	NEATFD	[1]	.SGI	LQAM	DNS	[3]	.LVSL	TLR	[3]	.FD	KFRC	[1]	.RNL	SMGMN	71
gi 129697	4	LRLVQGSLLKKVLEAIREL	VTDANF	[1]	.TGF	SLOAM	DSS	[3]	.LVALL	LR	[3]	.FE	HYRC	[1]	.RNL	SMGMN	71
gi 129700	4	ARLVQGSILKKVLEALKDL	IIDEAC	[1]	.SGI	SLOSM	DSS	[3]	.LVQL	TLR	[3]	.FD	TYRC	[1]	.RN	QSIGVK	71
gi 1709607	4	ARLIQGSNLKKVQEA	LKDIVTE	[1]	.SGI	SLOAM	DSS	[3]	.LVQL	TLR	[3]	.FE	NFRC	[1]	.RNL	LAMGIN	71
gi 2499442	4	LRLVQGSLLKKVLEAIREL	VNDANF	[1]	.TGF	SLOAM	DSS	[3]	.LVALL	LR	[3]	.FE	HYRC	[1]	.RNL	SMGMN	71
gi 13124422	4	LRLVQGSLLKKVLESIKDL	VNDANF	[1]	.TGF	SLOAM	DSS	[3]	.LVSL	LLR	[3]	.FE	HYRC	[1]	.RNL	SMGMN	71
gi 20139158	4	ARLVQGSVLKRVLEALKDL	ITEACW	[1]	.GGI	SLOSM	DSS	[3]	.LVQL	TLR	[3]	.FD	TYRC	[1]	.RNI	AMGVN	71
gi 20139256	4	ARLVQGSILKKVLEALKDL	ITEACW	[1]	.SGI	SLOSM	DSS	[3]	.LVQL	TLR	[3]	.FD	SYRC	[1]	.RNL	LAMGVN	71



Online Mendelian Inheritance in Man

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



The screenshot shows the OMIM website interface. At the top, there is a navigation bar with links to various databases: All Databases, PubMed, Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, and OMIM. A search bar is present with the text "Search OMIM for" and buttons for "Go" and "Clear". Below the search bar are tabs for "Limits", "Preview/Index", "History", "Clipboard", and "Details". On the left side, there is a sidebar menu with sections: "Entrez" (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, Restrictions on Use), and "Allied Resources" (Genetic Alliance, Databases, HGMD, Locus-Specific, Model Organisms, MitoMap, Phenotype, Davis Human/Mouse, Homology Maps, Coriell, The Jackson Laboratory, Human Gene Nomenclature). The main content area features a heading "OMIM™ - Online Mendelian Inheritance in Man™" followed by a welcome message and instructions on how to use the search and navigation features. A note at the bottom states that OMIM is intended for use by professionals and that users should consult a physician for personal medical or genetic conditions.

OMIM™ - Online Mendelian Inheritance in Man™

Welcome to OMIM, Online Mendelian Inheritance in Man. This database is a catalog of human genes and genetic disorders authored and edited by Dr. Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere, and developed for the World Wide Web by NCBI, the National Center for Biotechnology Information. The database contains textual information and references. It also contains copious links to MEDLINE and sequence records in the Entrez system, and links to additional related resources at NCBI and elsewhere.

You can do a search by entering one or more terms in the text box above. Advanced search options are accessible via the Limits, Preview/Index, History, and Clipboard options in the grey bar beneath the text box. The [OMIM help](#) document provides additional information and examples of basic and advanced searches.


The links to the left provide further technical information, searching options, frequently asked questions ([FAQ](#)), and information on allied resources. To return to this page, click on the OMIM link in the black header bar or on the graphic at the top of any OMIM page.


NOTE: OMIM is intended for use primarily by physicians and other professionals concerned with genetic disorders, by genetics researchers, and by advanced students in science and medicine. While the OMIM database is open to the public, users seeking information about a personal medical or genetic condition are urged to consult with a qualified physician for diagnosis and for answers to personal questions.

OMIM™ and Online Mendelian Inheritance in Man™ are trademarks of the Johns Hopkins University.


Online Mendelian Inheritance in Man

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





Online Mendelian Inheritance in Man



Johns Hopkins University

All Databases
PubMed
Nucleotide
Protein
Genome
Structure

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
Restrictions on Use

Allied Resources
Genetic


OMIM Statistics for January 15, 2007

Number of Entries

	Autosomal	X-Linked	Y-Linked	Mitochondrial	Total
* Gene with known sequence	10636	495	48	37	11216
+ Gene with known sequence and phenotype	353	32	0	0	385
# Phenotype description, molecular basis known	1846	167	2	26	2041
‰ Mendelian phenotype or locus, molecular basis unknown	1411	135	4	0	1551
Other, mainly phenotypes with suspected mendelian basis	2016	144	2	0	2162
Total	16262	973	56	63	17355

Entrez Structure Database

<http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=Structure>



NCBI

PubMed Nucleotide Protein Genome Structure PMC Taxonomy

Search Structure for

Limits Preview/Index History Clipboard Details

Hints on finding a Structure

- To search by [keyword](#), like "aconitase"
- To search by [protein](#) sequence
- To search by [nucleotide](#) sequence

New structure viewer

[Cn3D](#) is NCBI's 3D structure viewer. As a helper application for your web browser, it allows you to interactively view 3-D structures, sequences, and sequence alignments. Cn3D is available for Windows, MacOS, and Unix. [More...](#)

About the Database

The Molecular Modelling Database (MMDB) contains 3-D macromolecular structures, including proteins and polynucleotides. MMDB contains over 10,000 structures and is linked to the rest of the NCBI databases, including sequences, bibliographic citations, taxonomic classifications, and sequence and structure neighbors.

Entrez Structure
Help | FAQ

Structure Research
The NCBI Structure group

MMDB
About Entrez's structure database

CDD
Conserved Domain Database

PDBeast
Taxonomy in MMDB

Cn3D **NEW**
3D-structure viewer

VAST
Structure comparisons

VAST Search
Submit structure database searches

Research
Structure Group research projects

MMDB Structure for PCNA (1VYM)

<http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=Structure>

NCBI

MMDB Structure Summary

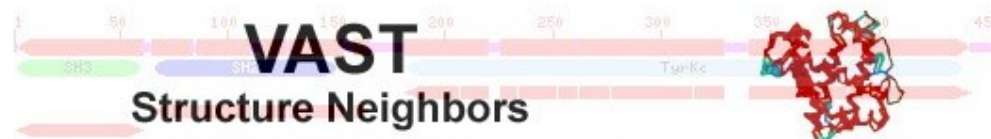
PubMed BLAST Structure Taxonomy OMIM Help? Cn3d

Description: Leghemoglobin (Carbon Monoxy).
Deposition: E.Harutyunyan, T.Safonova & I.Kuranova, 14-Sep-94
Taxonomy: [Lupinus luteus](#)
Reference: [PubMed](#) MMDB: [1014](#) PDB: [1GDI](#)

View 3D Structure of Best Model with Cn3D Display **NEW** [Get Cn3D 4.0!](#)

Protein Chain
CDs globin

VAST Neighbors of PCNA (1VYM)



[PubMed](#)
[BLAST](#)
[Structure](#)
[Taxonomy](#)
[OMIM](#)
[Help?](#)
[Cn3D](#)

VAST neighbors for: **MMDB 31339, 1VYM A.**

Overview: There are two main sections to this page. The first section consists of the alignment view controls, the list controls, and the advanced neighbor search controls. The second section is the VAST neighbor list itself.

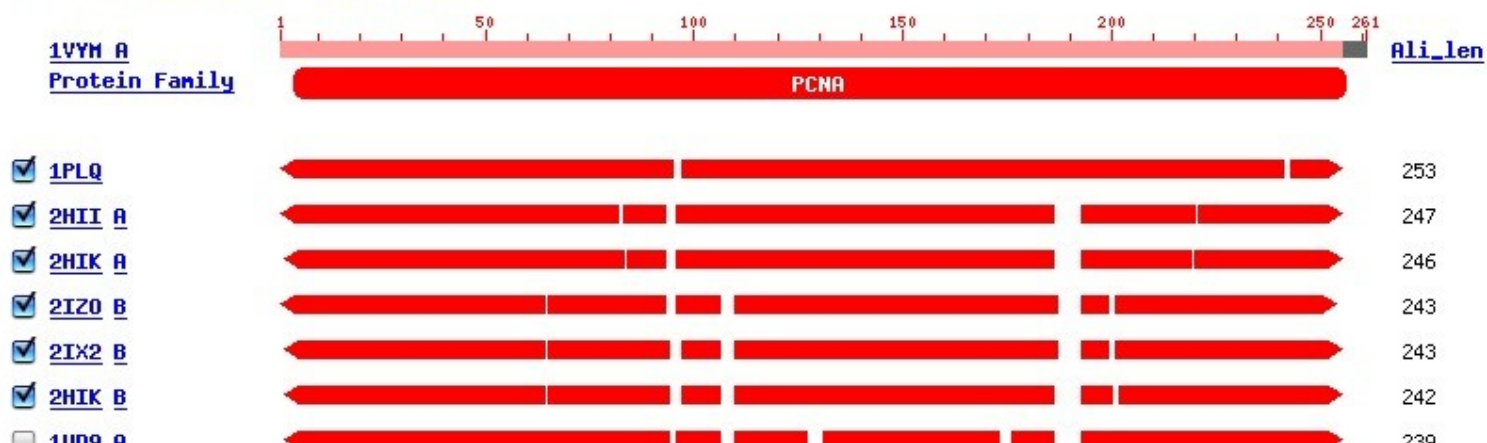
of with [Download Cn3D!](#)
 using for **VAST neighbors**
 subset, sorted by **in**

Advanced neighbor search

Move the mouse over the **red** alignment footprints in the graphics below and click, you will obtain a structure-based sequence alignment.


Total neighbors: 143; 21 representatives from the Medium redundancy subset displayed.

Click to: [Check All](#) [Uncheck All](#)




Protein Data Bank (Structures)

<http://www.pdb.org/>



RCSB
PROTEIN DATA BANK

A MEMBER OF THE  **PDB**

An Information Portal to Biological Macromolecular Structures

As of Tuesday Jan 16, 2007 there are 41136 Structures | [PDB Statistics](#)

Contact Us | Help | Print Page

Home Search

- Home
- Tutorial About This Site
- Getting Started
- ▶ Download Files
- ▶ Deposit and Validate
- ▶ Structural Genomics
- ▶ Dictionaries & File Formats
- ▶ Software Tools
- ▶ General Education
- BioSync
- ▶ General Information
- Acknowledgements
- Frequently Asked Questions
- Known Problems
- ✉ Report Bugs/Comments

Welcome to the RCSB PDB

The **RCSB** PDB provides a variety of tools and resources for studying the structures of biological macromolecules and their relationships to sequence, function, and disease.

The RCSB is a member of the **wwPDB** whose mission is to ensure that the PDB archive remains an international resource with uniform data.


This site offers tools for browsing, searching, and reporting that utilize the data resulting from ongoing efforts to create a more consistent and comprehensive archive.

Information about compatible browsers can be found [here](#).

A **narrated tutorial** illustrates how to search, navigate, browse, generate reports and visualize structures using this new site. [This requires the Macromedia [Flash player download](#).]

Comments? info@rcsb.org

Molecule of the Month: Importins



Inside your cells, the process of protein synthesis is separated into two compartments. The first half of the job, when DNA is transcribed into RNA, is performed in the nucleus. The second half is then performed outside the nucleus, when ribosomes translate the RNA to construct proteins in the cytoplasm. This separation requires a continuous traffic of molecules: new RNA molecules must be transported out of the nucleus and nuclear proteins, such as newly-synthesized histones or polymerases, must be transported back into the nucleus. Huge

NEWS

- [Complete News](#)
- [Newsletter](#)
- [Discussion Forum](#)

16-January-2007
PDB File Formats, Annotation Procedures, and Remediation

wwPDB members work to annotate all data deposited to the PDB archive. Information about data file formats, annotation procedures, and remediation efforts are described in this article.

- [Full Story ...](#)

09-January-2007
Browsing the PDB Using Medical Subject Headings (MeSH)

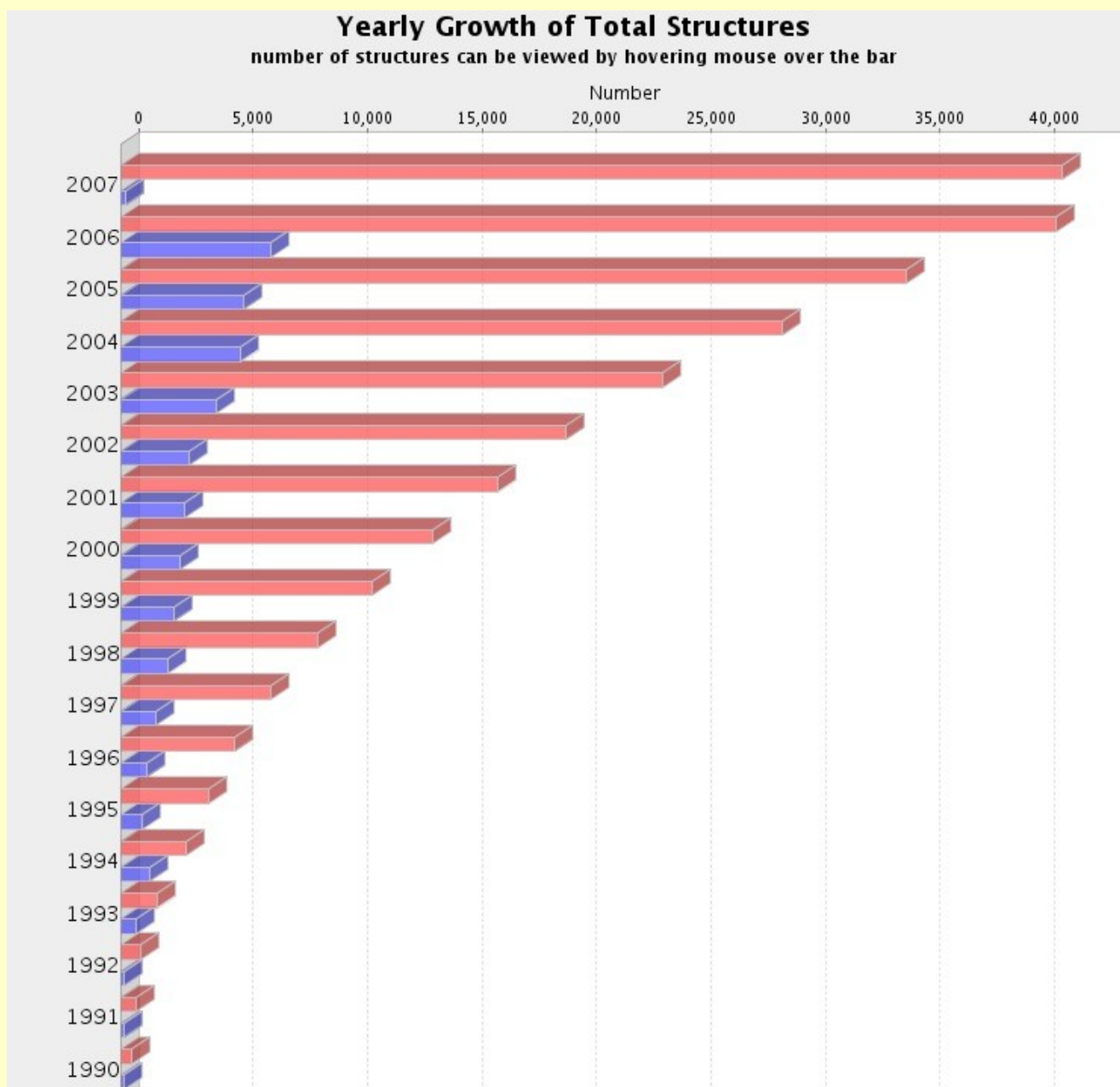
PDB Contents

PDB Current Holdings Breakdown

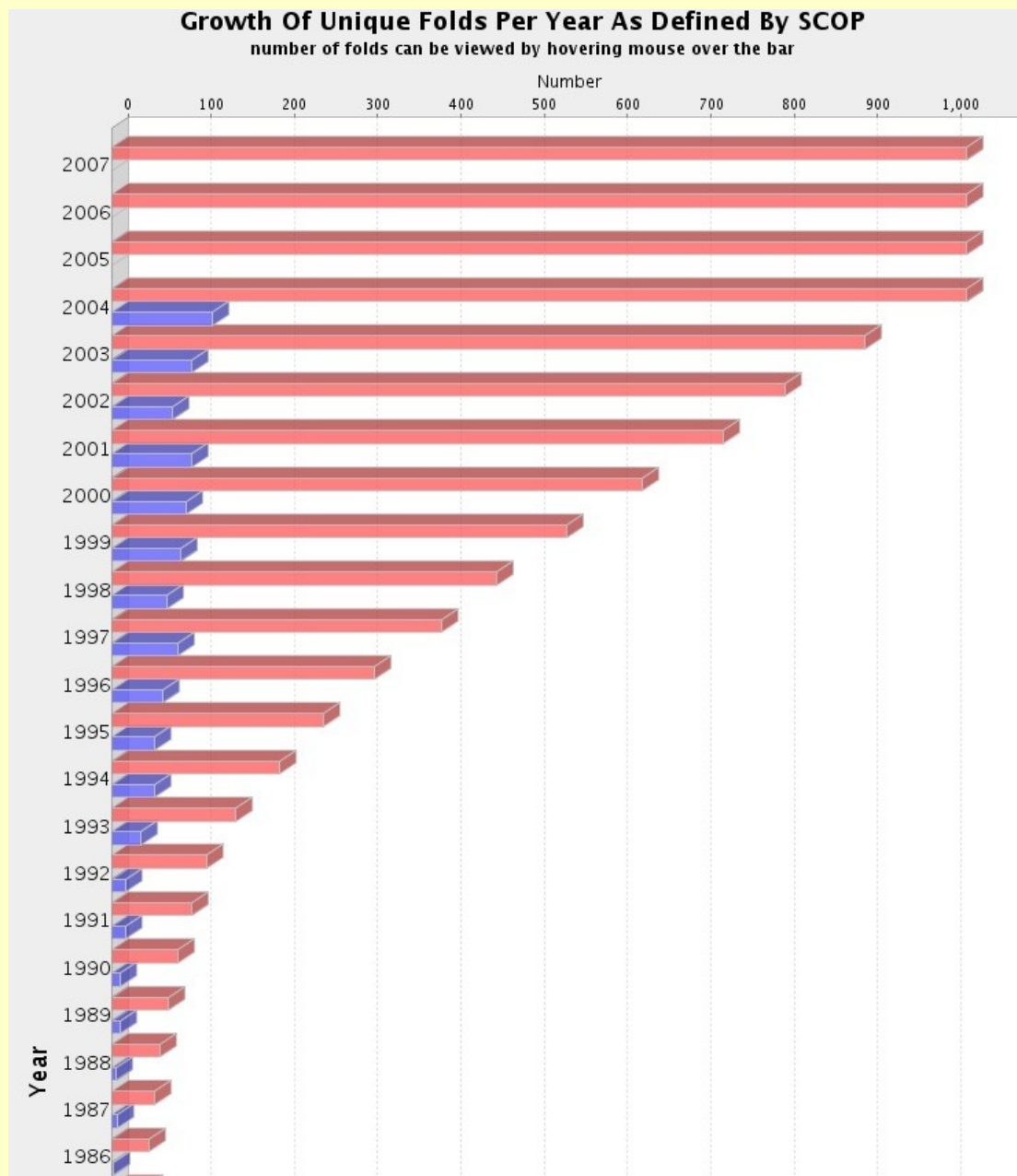
		Molecule Type				
		Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
Exp. Method	X-ray	32371	940	1522	28	34861
	NMR	5180	736	126	7	6049
	Electron Microscopy	95	10	37	0	142
	Other	77	4	3	0	84
	Total	37723	1690	1688	35	41136

(Click on any number to retrieve the results from that category.)

Growth of PDB





Growth of PDB by Fold



Protein Structure Initiative

<http://www.nigms.nih.gov/Initiatives/PSI.htm>





 National Institute of
General Medical Sciences
One of the National Institutes of Health

Search NIGMS


[NIGMS Home](#) | [Site Map](#) | [Staff Contacts](#)

[Featured Areas](#) | [Research Funding](#) | [Training & Careers](#) | [Minority Programs](#) | [News & Events](#) | [Publications](#) | [About NIGMS](#)

Back to: [NIGMS Home](#) > [Featured Areas](#)  Print this page  E-mail this page

Protein Structure Initiative

Better Tools and Better Knowledge for Structural Genomics



The **Protein Structure Initiative (PSI)** is a federal, university, and industry effort aimed at dramatically reducing the costs and lessening the time it takes to determine a three-dimensional protein structure. The long-range goal of the PSI is to make the three-dimensional atomic-level structures of most proteins easily obtainable from knowledge of their corresponding DNA sequences.

Expected benefits from the PSI include: structural descriptions to help researchers discover the functions of proteins, design experiments, and solve other key biomedical problems; faster identification of promising new structure-based medicines; better therapeutics for treating both genetic and infectious diseases; and development of technology and methodology for protein production and crystallography.

For more details about the PSI's goals, organization, and benefits, see the PSI [Mission Statement](#). Additional information about the PSI can be found under the links listed below.

[Background Information](#)
Learn about the PSI through answers to frequently asked questions, fact sheets, and other background material.

[Centers](#)
Discover more about the PSI-funded centers, their missions, and principal investigators.

[Funding Opportunities](#)
View PSI-related listings from the NIH Guide for Grants and Contracts, the official publication for NIH research grant policies, guidelines, and funding opportunities.

[News](#)
Read about PSI announcements and results.

[Media Mentions](#)
Articles and other news coverage featuring PSI-supported advances.

[Meetings](#)
Find out about PSI meetings.

[Reports](#)
Read PSI meeting reports.

[Images](#)
View a collection of images featuring PSI-related structures, machines and methods, and people.

[Contact Us](#)
Reach PSI program staff.

Related Information

- [Protein Structure Initiative-
Materials Repository](#)
- [Protein Structure Initiative
Annual Meeting](#)
- [Supplements for Functional
Studies Based on High
Resolution Structures
Obtained in the Protein
Structure Initiative](#)

Structure Summary for 1VYM


Structure Summary
Biology & Chemistry
Materials & Methods
Sequence Details
Geometry

1VYM

Title	NATIVE HUMAN PCNA						
Authors	Kontopidis, G., Wu, S., Zheleva, D., Taylor, P., Mcinnes, C., Lane, D., Fischer, P., Walkinshaw, M.						
Primary Citation	Kontopidis, G., Wu, S., Zheleva, D., Taylor, P., Mcinnes, C., Lane, D., Fischer, P., Walkinshaw, M. Structural and Biochemical Studies of Human Proliferating Cell Nuclear Antigen Complexes Provide a Rationale for Cyclin Association and Inhibitor Design <i>Proc.Natl.Acad.Sci.USA</i> v102 pp.1871, 2005						
History	Deposition	2004-05-03	Release	2005-01-13			
Experimental Method	Type	X-RAY DIFFRACTION				Data	[EDS]
Parameters	Resolution[Å]	R-Value	R-Free	Space Group			
	2.30	0.191 (obs.)	0.279	C 2 (C 1 2 1)			
Unit Cell	Length [Å]	a	136.65	b	83.26	c	71.63
	Angles [°]	alpha	90.00	beta	117.49	gamma	90.00
Molecular Description Asymmetric Unit	Polymer: 1	Molecule: PROLIFERATING CELL NUCLEAR ANTIGEN	Chains: A,B,C				
Classification	DNA Binding Protein						

Images and Visualization

Biological Molecule / Asymmetric Unit



Display Options

[KiNG](#)
[Jmol](#)
[WebMol](#)
[Protein Workshop](#)
[QuickPDB](#)
[All Images](#)

Source Polymer: 1 Scientific Name: **Homo sapiens** Common Name: **Human** Expression system: **Escherichia coli**

PFAM Classification	Chain	PFAM Accession	PFAM ID	Description	Type	Clan ID
	A	PF00705	PCNA_N	Proliferating cell nuclear antigen, N-terminal domain	Domain	DNA_clamp
	A	PF02747	PCNA_C	Proliferating cell nuclear antigen, C-terminal domain	Domain	DNA_clamp
	B	PF00705	PCNA_N	Proliferating cell nuclear antigen, N-terminal domain	Domain	DNA_clamp
	B	PF02747	PCNA_C	Proliferating cell nuclear antigen, C-terminal domain	Domain	DNA_clamp
	C	PF00705	PCNA_N	Proliferating cell nuclear antigen, N-terminal domain	Domain	DNA_clamp
	C	PF02747	PCNA_C	Proliferating cell nuclear antigen, C-terminal domain	Domain	DNA_clamp

GO Terms	Polymer PROLIFERATING CELL NUCLEAR ANTIGEN (1VYM:A,B,C)	Molecular Function • none	Biological Process • none	Cellular Component • none
-----------------	--	------------------------------	------------------------------	------------------------------



Jmol View of 1VYM

RCSB **PDB** PROTEIN DATA BANK

A MEMBER OF THE **PDB**

An Information Portal to Biological Macromolecular Structures

As of Tuesday Jan 16, 2007 there are 41136 Structures | PDB Statistics

Contact Us | Help | Print Page

PDB ID or keyword Author SEARCH | Advanced Search

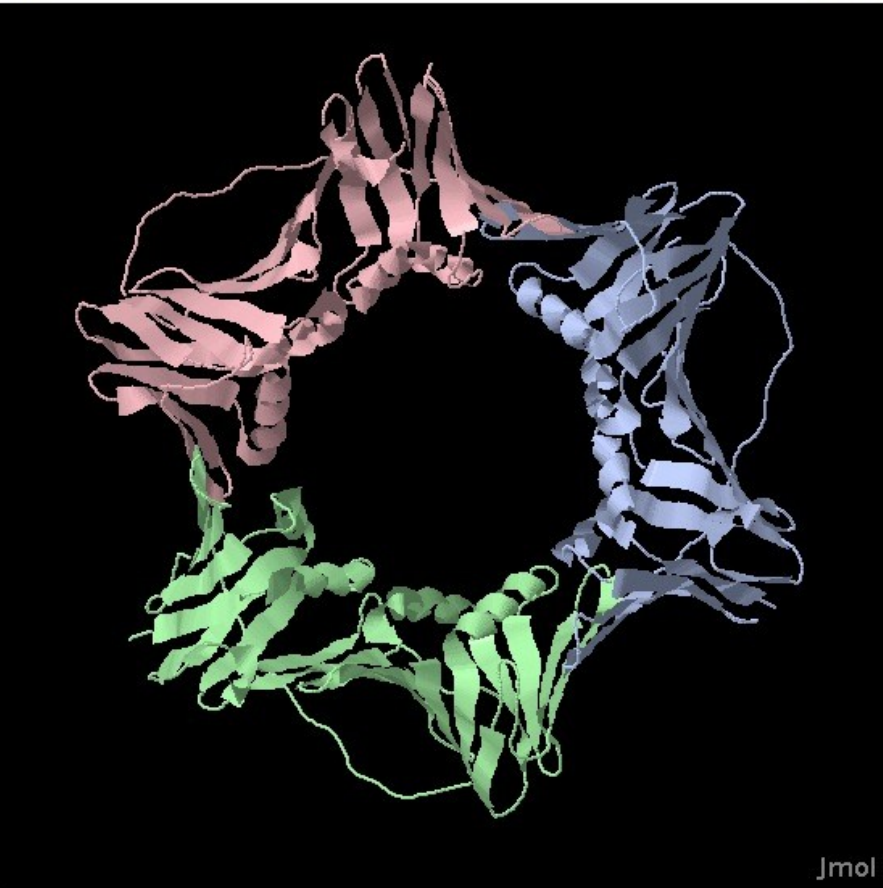
Home Search Structure Results

Queries

- 1VYM
- Download Files
- FASTA Sequence
- Display Files
- Display Molecule
 - Image Gallery
 - KiNG Viewer
 - Jmol Viewer**
 - WebMol Viewer
 - Protein Workshop
 - Rasmol Viewer (Plugin required)
 - Swiss-PDB Viewer (Plugin required)
 - KiNG Help
 - Jmol Help
 - WebMol Help
 - Protein Workshop Help
 - QuickPDB
 - Asymmetric Unit / Biological Molecule
- Structural Reports
- External Links
- Structure Analysis
- Help
 - Structure Explorer Intro
 - Molecular Viewers
 - Structure Summary
 - Biological Molecule
 - Biology & Chemistry

Jmol

1VYM



Jmol

Jmol

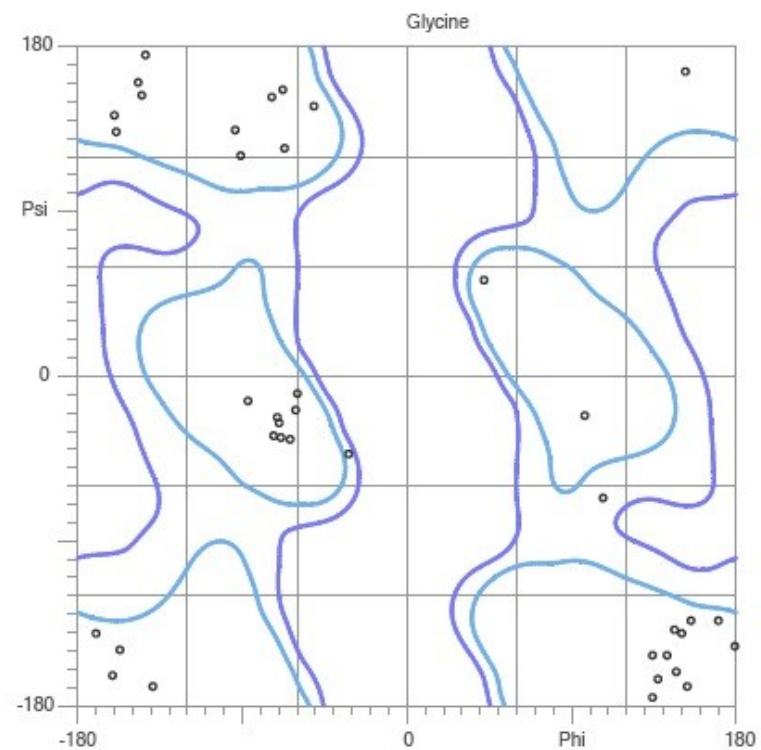
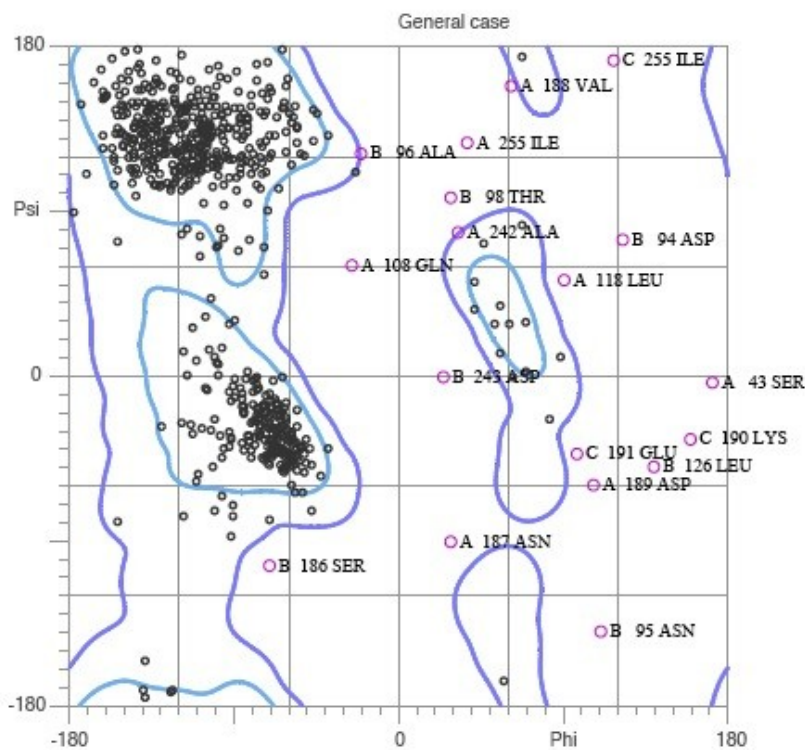
- Help interacting with Jmol
- Simple Interaction Guide (requires flash)
- Advanced Jmol Help

Ramachandran Plot of 1VYM



MolProbity Ramachandran analysis

1VYM, model 1





Swiss Model: Comparative Modeling Server

<http://swissmodel.expasy.org/>

MENU

Modeling requests:

- [First Approach mode](#)
- [Alignment Interface](#)
- [Project \(optimise\) mode](#)
- [Oligomer modeling](#)
- [GPCR mode](#)

Model Database

- [SWISS-MODEL Repository](#), a database for theoretical protein models.

Interactive tools

- [SWISS-MODEL Workspace](#), an interactive working environment for protein structure modelling and assessment.
- [DeepView - Swiss-PdbViewer](#), a tool for viewing and

HELP

- [Frequently Asked Questions.](#)
- [Visualising 3D models.](#)
- [Reliability of models.](#)
- [How SWISS-MODEL works.](#)
- [How ProModII works.](#)
- [Modelling of oligomeric](#)



SWISS-MODEL

An Automated Comparative Protein Modelling Server

SIB - Biozentrum Basel site provided by:

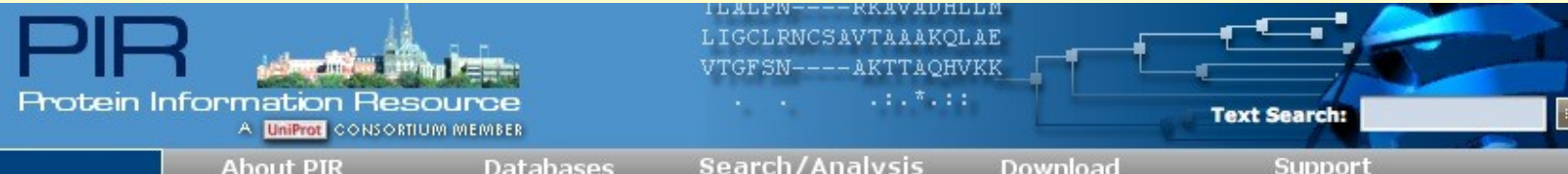


[SWISS-MODEL](#) is a fully automated protein structure homology-modeling server, accessible via the [ExpASY](#) web server, or from the program [DeepView](#) (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists World Wide.

The present version of the server is 3.5 and is under constant improvement and debugging. In order to help us refine the sequence analysis and modelling algorithms, please [report](#) of possible bugs and problems with the modelling procedure.

SWISS-MODEL was initiated in 1993 by Manuel Peitsch, and is now being further developed within the [SIB - Swiss Institute of Bioinformatics](#) in collaboration between Torsten Schwede at the [Structural Bioinformatics Group](#), Biozentrum (University of Basel) and Nicolas Guex at [GlaxoSmithKline](#). The computational resources for the SWISS-MODEL server are provided in collaboration by the Biozentrum (University Basel) and the [Advanced Biomedical Computing Center](#) (NCI Frederick, USA).



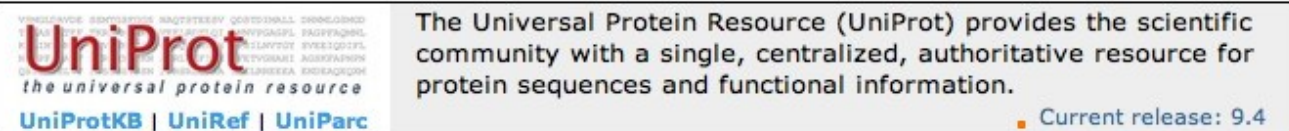


PIR Protein Information Resource
A UniProt CONSORTIUM MEMBER

Text Search:

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

INTEGRATED PROTEIN INFORMATICS RESOURCE FOR GENOMIC AND PROTEOMIC RESEARCH



UniProt
the universal protein resource

UniProtKB | UniRef | UniParc


The Universal Protein Resource (UniProt) provides the scientific community with a single, centralized, authoritative resource for protein sequences and functional information.

Current release: 9.4



PIRSF
Protein Family Classification System

- Classification reflecting evolutionary relationships of full-length proteins
- Functional site and protein name rules
- *Sample family report*



iProClass
Integrated Protein Knowledgebase

- Value-added reports for UniProtKB and unique UniParc proteins
- Functional analysis and protein ID mapping
- *Sample protein report*



iProLINK
Literature, Information & Knowledge

- Source for text mining and ontology development
- RLIMS-P text mining tool, BioThesaurus, and PProtein Ontology
- Bibliography mapping

OTHER RESOURCE

- Proteomics: NIAID Biodefense Proteomics Admin. Center
- PIR Grid-Enablement: Data node on NCI's caBIG

P PEPTIDE SEARCH ?

DATABASE: UniProtKB

Use single letter amino acid code

T TEXT SEARCH ?

DATABASE: iProClass

Controlled Vocabulary for p53

BioThesaurus Report for UniProtKB Entry: **P04637**

PROTEIN INFORMATION	
UniProtKB Name and ID	Protein Name: Cellular tumor antigen p53 (Tumor suppressor p53) (Phosphoprotein p53) (Antigen NY-CO-13) Gene Name: TP53; Synonyms: P53 ID: P53_HUMAN <small>UniProtKB/Swiss-Prot</small> <small>/ProClass</small> Accession: P04637; Q15086; Q15087; Q15088; Q16535; Q16807; Q16808; Q16809; Q16810; Q16811; Q16848; Q86UG1; Q8J016; Q99659; Q9BTM4; Q9HAQ8; Q9NP68; Q9NPJ2; Q9NZD0; Q9UBI2; Q9UQ61
Taxonomy	Source Organism: Homo sapiens (Human) Taxon Group: Euk/mammal NCBI Taxon: 9606
Classification	UniRef: UniRef90 P04637 ; UniRef50 P04637 PIRSF: SF002089 - cellular tumor antigen p53 [Unknown] Related Sequences: P04637

BIOTHESAURUS INFORMATION				
Number of Names	Synonyms	Textual Variants	Source Databases	
		36	41	8
	Synonym	Source Count	Textual Variant	Source Attribution
	P53	7	P53	UniProtKB: P04637 OMIM: 191170 HGNC: 11998 EMBL: BAC16799.1 UMLS:C0079419 (T028) UMLS:C0080055 (T116+T123)
			p53	Entrez Gene: 7157 GenAtlas: 1691 UMLS:C0079419 (T028) UMLS:C0080055 (T116+T123)
	TP53	6	TP53	UniProtKB: P04637 Entrez Gene: 7157 OMIM: 191170 HGNC: 11998 GenAtlas: 1691 UMLS:C0079419 (T028) UMLS:C0080055 (T116+T123)
	Tumor suppressor p53	4	Tumor suppressor p53	UniProtKB: P04637 UMLS:C0080055 (T116+T123)
			tumor suppressor p53	PIR-PSD: A25224 EMBL: AAR10356.1 UMLS:C0080055 (T116+T123)

ProClass Summary for Human p53

ProClass Summary Report for UniProtKB Entry: P04637

[Related Sequences](#) [BioThesaurus](#) [ID Mapping](#)

GENERAL INFORMATION		
Protein Name and ID	UniProtKB ID	UniProtKB Accession
	P53_HUMAN	P04637 ; Q15086 ; Q15087 ; Q15088 ; Q16535 ; Q16807 ; Q16808 ; Q16809 ; Q16810 ; Q16811 ; Q16848 ; Q86UG1 ; Q8J016 ; Q99659 ; Q9BTM4 ; Q9HAQ8 ; Q9NP68 ; Q9NPJ2 ; Q9NZD0 ; Q9UBI2 ; Q9UQ61
	Protein Name	
	Cellular tumor antigen p53 (Tumor suppressor p53) (Phosphoprotein p53) (Antigen NY-CO-13)	
	PIR-PSD: DNHU53 RefSeq: NP_000537.3 GenPept: AAK76359.1 ; AAF36357.1 ; CAA42629.1 ; AAC12971.1 ; AAD28628.1 ; AAA61211.1 ; AAF36360.1 ; CAA42631.1 ; AAF36379.1 ; AAV80424.1 ; AAD28535.1 ; AAB39322.1 ; AAA61212.1 ; CAA42635.1 ; AAF36380.1 ; AAH03596.1 ; CAA42632.1 ; BAC16799.1 ; CAA38095.1 ; CAA26306.1 ; AAF36374.1 ; AAR13239.1 ; CAA42626.1 ; AAA59989.1 ; AAF36381.1 ; AAF36376.1 ; AAG28785.1 ; AAF36359.1 ; AAF36377.1 ; AAF36355.1 ; AAF63442.1 ; AAP30003.1 ; AAA59988.1 ; AAQ90158.1 ; AAF36358.1 ; AAF36361.1 ; AAA59987.1 ; AAF36354.1 ; AAF36382.1 ; CAC22427.1 ; CAA25652.1 ; CAA42634.1 ; CAA42628.1 ; AAF63443.1 ; AAR10356.1 ; AAF36356.1 ; CAA42633.1 ; AAF36378.1 ; CAA42627.1 ; AAF36362.1 ; AAF36375.1 ; CAA42630.1 ; CAD99919.1 ; AAK76358.1 IPI: IPI00025087 ; IPI00375319	
Taxonomy	<i>Source Organism:</i> Homo sapiens (Human) <i>Taxon Group:</i> Euk/mammal <i>NCBI Taxon:</i> 9606 <i>Lineage:</i> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.	
Gene Name	TP53; P53	
Keywords	3d-structure; DNA binding; acetylation; activator; alternative splicing; anti-oncogene; apoptosis; cell cycle; cell division control; disease mutation; dna-binding; glycoprotein; homotetramer; host-virus interaction; li-fraumeni syndrome; metal-binding; nuclear protein; phosphoprotein; phosphorylation; polymorphism; transcription; transcription regulation; tumor suppressor; zinc	



Sequences Related to p53

Related Sequences Search Result

Show Query Sequence Help?

297 hits | 15 pages | 20 entries/page | 1 | 2 | 3 | 4 | 5 | >>>

Save Result As: TABLE FASTA

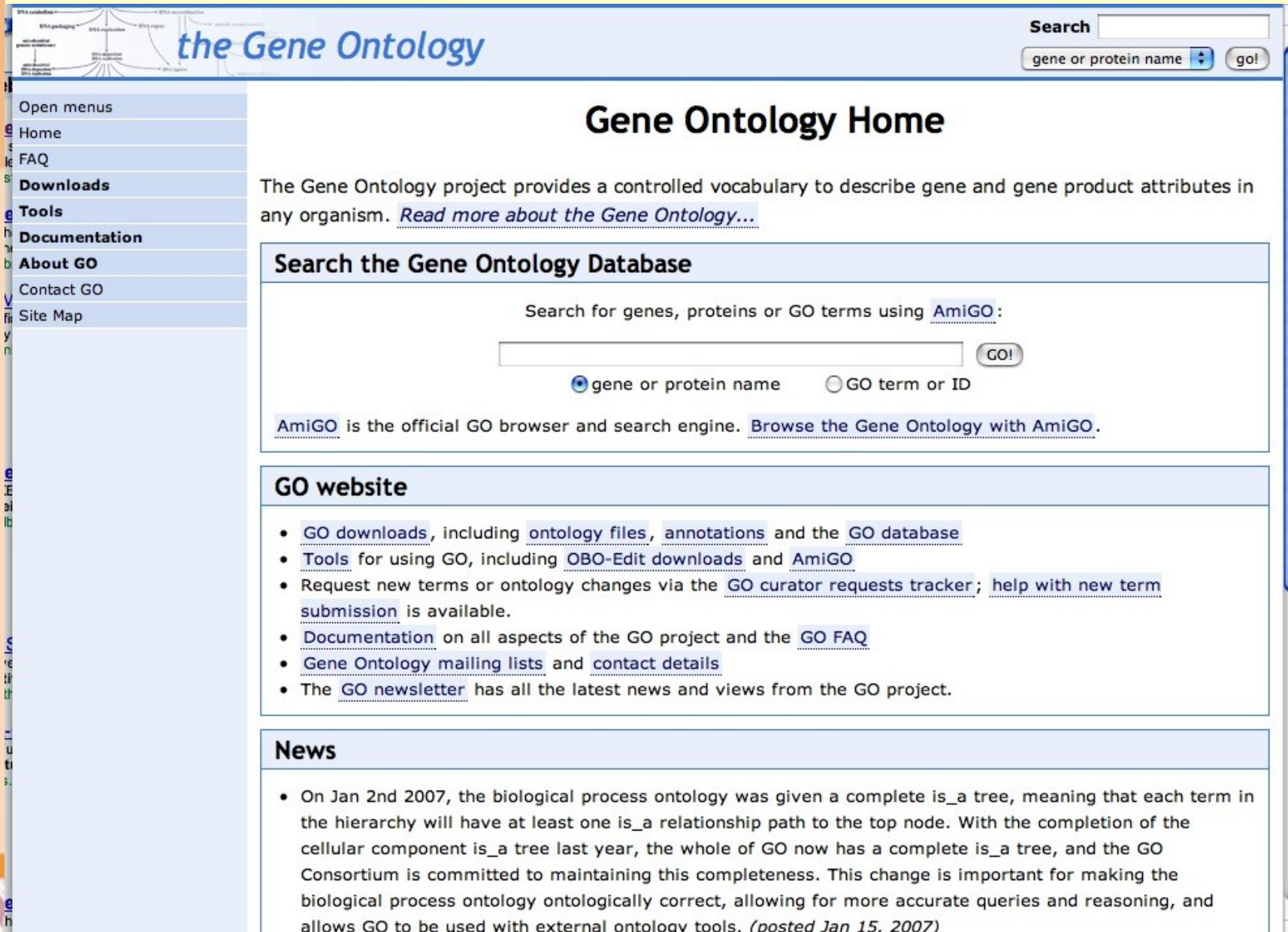
check&analyze

BLAST
 FASTA
 Pattern Match
 Pairwise Alignment
 Multiple Alignment
 Domain Display

Protein AC/ID	Protein Name	Length	Organism Name	PIRSF ID	BLAST Search				
					Overlap	%iden	E-Value	Score	Alignment
<input type="checkbox"/> P04637/P53_HUMAN <small>/ProClass UniProtKB/Swiss-Prot</small>	Cellular tumor antigen p53 (Tumor suppressor p53) ... <small>BioThesaurus</small>	393	Homo sapiens (Human)	PIRSF002089	393	100.00	0.0	821	
<input type="checkbox"/> Q2XN98/Q2XN98_HUMAN <small>/ProClass UniProtKB/TrEMBL</small>	P53 protein <small>BioThesaurus</small>	393	Homo sapiens (Human)	PIRSF002089	393	99.75	0.0	818	
<input type="checkbox"/> Q5U0E4/Q5U0E4_HUMAN <small>/ProClass UniProtKB/TrEMBL</small>	Tumor protein p53 (Li-Fraumeni syndrome) <small>BioThesaurus</small>	393	Homo sapiens (Human)	PIRSF002089	393	99.49	0.0	817	
<input type="checkbox"/> Q2XSC7/Q2XSC7_HUMAN <small>/ProClass UniProtKB/TrEMBL</small>	Tumor protein p53 <small>BioThesaurus</small>	393	Homo sapiens (Human)	PIRSF002089	393	98.98	0.0	811	
<input type="checkbox"/> P56423/P53_MACFA <small>/ProClass UniProtKB/Swiss-Prot</small>	Cellular tumor antigen p53 (Tumor suppressor p53) <small>BioThesaurus</small>	393	Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey)	PIRSF002089	393	95.67	0.0	789	
<input type="checkbox"/> P56424/P53_MACMU <small>/ProClass UniProtKB/Swiss-Prot</small>	Cellular tumor antigen p53 (Tumor suppressor p53) <small>BioThesaurus</small>	393	Macaca mulatta (Rhesus macaque)	PIRSF002089	393	95.67	0.0	789	
<input type="checkbox"/> P61260/P53_MACFU <small>/ProClass UniProtKB/Swiss-Prot</small>	Cellular tumor antigen p53 (Tumor suppressor p53) <small>BioThesaurus</small>	393	Macaca fuscata fuscata (Japanese macaque)	PIRSF002089	393	95.67	0.0	789	
<input type="checkbox"/> Q1MSX0/Q1MSX0_HUMAN <small>/ProClass UniProtKB/TrEMBL</small>	Tumor protein p53 mutant form (Fragment) <small>BioThesaurus</small>	382	Homo sapiens (Human)	PIRSF002089	393	96.95	0.0	788	
<input type="checkbox"/> P13481/P53_CERAE <small>/ProClass UniProtKB/Swiss-Prot</small>	Cellular tumor antigen p53 (Tumor suppressor p53) <small>BioThesaurus</small>	393	Cercopithecus aethiops (Green monkey) (Grivet)	PIRSF002089	393	95.67	0.0	787	

Genome Ontology Consortium

<http://www.geneontology.org/>



the Gene Ontology

Search go!

Open menus

- Home
- FAQ
- Downloads
- Tools
- Documentation
- About GO
- Contact GO
- Site Map

Gene Ontology Home

The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism. [Read more about the Gene Ontology...](#)

Search the Gene Ontology Database

Search for genes, proteins or GO terms using [AmiGO](#):

gene or protein name GO term or ID

[AmiGO](#) is the official GO browser and search engine. [Browse the Gene Ontology with AmiGO.](#)

GO website

- [GO downloads](#), including [ontology files](#), [annotations](#) and the [GO database](#)
- [Tools](#) for using GO, including [OBO-Edit downloads](#) and [AmiGO](#)
- Request new terms or ontology changes via the [GO curator requests tracker](#); [help with new term submission](#) is available.
- [Documentation](#) on all aspects of the GO project and the [GO FAQ](#)
- [Gene Ontology mailing lists](#) and [contact details](#)
- The [GO newsletter](#) has all the latest news and views from the GO project.

News

- On Jan 2nd 2007, the biological process ontology was given a complete is_a tree, meaning that each term in the hierarchy will have at least one is_a relationship path to the top node. With the completion of the cellular component is_a tree last year, the whole of GO now has a complete is_a tree, and the GO Consortium is committed to maintaining this completeness. This change is important for making the biological process ontology ontologically correct, allowing for more accurate queries and reasoning, and allows GO to be used with external ontology tools. (posted Jan 15, 2007)