

Search of Biological Databases and Literature

Jianlin Cheng, PhD

School of Electrical Engineering and Computer Science
University of Central Florida



2006

Free for academic use. Copyright @ Jianlin Cheng & original sources for some materials

Databases and Literature

1. NCBI and Entrez
2. Genbank (nucleotide (DNA and RNA))
3. SwissProt and PIR (protein sequences)
4. PDB (protein structure)
5. GO (gene ontology, gene/protein function)
6. Ensembl
7. ExPASy (Expert Protein Analysis System)
8. BioMedical Literature (PubMed)

1. NCBI - Main Portal

The screenshot shows the NCBI Home Page in a Mozilla Firefox browser window. The browser's address bar displays the URL <http://www.ncbi.nlm.nih.gov/>. The page header includes the NCBI logo and the text "National Center for Biotechnology Information", "National Library of Medicine", and "National Institutes of Health". Below the header is a navigation menu with links to "PubMed", "All Databases", "BLAST", "OMIM", "Books", "TaxBrowser", and "Structure". A search bar is present with a dropdown menu set to "All Databases" and a "Go" button.

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
Data mining

Research at NCBI
People, projects, and seminars

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...](#)

Whole Genome Association
The NCBI Whole Genome Association (WGA) resource provides researchers with access to genotype and associated phenotype information that will help elucidate the link between genes and disease. For more information, click here to see the the [WGA](#) resource page and click here to read the [press release](#).

1 Billion Live Traces
The Trace Archive of sequencing traces has reached 1 billion live traces from over 480 organisms. For more information about the Trace Archive database [click here](#).

PubMed Central
An archive of life sciences journals

- Free fulltext
- Over 500,000 articles from over 200 journals
- Linked to PubMed and fully searchable

Use of PubMed Central requires no registration or fee. Access it from any computer with an Internet connection.

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
- dbMHC
- Mouse genome resources
- My NCBI
- ORF finder
- Rat genome



National Center for Biotechnology Information

National Library of Medicine

National Institutes of Health

PubMed

All Databases

BLAST

OMIM

Books

TaxBrowser

Structure

Search

PubMed



for

Go

PubMed is...

- National Library of Medicine's search service
- 16 million citations in MEDLINE
- links to participating online journals
- PubMed tutorial (via “Education” on side bar)

Adapted from J. Pevsner, 2005

Search PubMed for [Save Search](#)

Display Show Sort by Send to

Items 1 - 4 of 4

About Entrez

Text Version

Entrez PubMed

Overview
Help | FAQ
Tutorials
New/Noteworthy 
E-Utilities

PubMed Services

Journals Database
MeSH Database
Single Citation
Matcher
Batch Citation Matcher
Clinical Queries
Special Queries
LinkOut
My NCBI

Related Resources

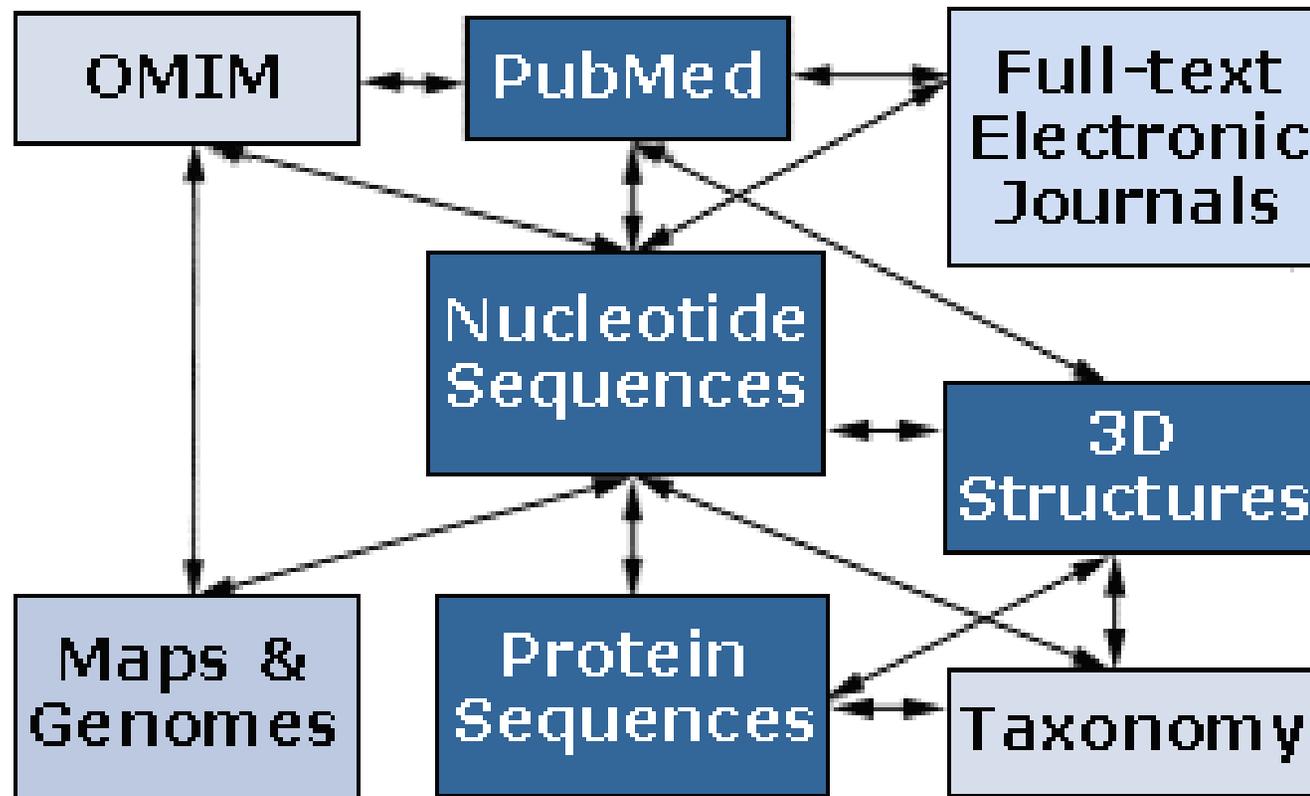
Order Documents
NLM Mobile
NLM Catalog
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

- 1:** [Cheng J, Baldi P.](#)
 A machine learning information retrieval approach to protein fold recognition.
 Bioinformatics. 2006 Jun 15;22(12):1456-63. Epub 2006 Mar 17.
 PMID: 16547073 [PubMed - indexed for MEDLINE]
- 2:** [Cheng J, Randall A, Baldi P.](#)
 Prediction of protein stability changes for single-site mutations using support vector machines.
 Proteins. 2006 Mar 1;62(4):1125-32.
 PMID: 16372356 [PubMed - indexed for MEDLINE]
- 3:** [Cheng J, Saigo H, Baldi P.](#)
 Large-scale prediction of disulphide bridges using kernel methods, two-dimensional recursive neural networks, and weighted graph matching.
 Proteins. 2006 Mar 15;62(3):617-29.
 PMID: 16320312 [PubMed - indexed for MEDLINE]
- 4:** [Cheng J, Baldi P.](#)
 Three-stage prediction of protein beta-sheets by neural networks, alignments and graph algorithms.
 Bioinformatics. 2005 Jun;21 Suppl 1:i75-84.
 PMID: 15961501 [PubMed - indexed for MEDLINE]

Entrez

A robust and flexible database retrieval system that covers over 20 biological databases containing DNA and protein sequence data, genome mapping data, population sets, phylogenetic sets, environmental sample sets, gene expression data, the NCBI taxonomy, protein domain, protein structure, MEDLINE references via PubMed.

Entrez is a search and retrieval system that integrates NCBI databases



Entrez

Entrez cross-database search - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi

Getting Started Latest Headlines

GenBank -- Benson et al. 33 (Supplement 1)... NCBI Taxonomy Homepage The DDBJ/EMBL/GenBank Feature Table: De... Entrez cross-database search

NCBI

Entrez, The Life Sciences Search Engine

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

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
Journals: detailed information about the journals indexed in PubMed and other Entrez databases	MeSH: detailed information about NLM's controlled vocabulary
NLM Catalog: catalog of books, journals, and audiovisuals in the NLM collections	

Enter terms and click 'GO' to run the search against ALL the databases, OR
Click Database Name or Icon to go directly to the Search Page for that database, OR
Click Question Mark for a short explanation of that database.

Done

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



National Center for Biotechnology Information

National Library of Medicine

National Institutes of Health

PubMed

All Databases

BLAST

OMIM

Books

TaxBrowser

Structure

Search

PubMed



for

Go

BLAST is...

- Basic Local Alignment Search Tool
- NCBI's sequence similarity search tool
- supports analysis of DNA and protein databases
- 80,000 searches per day



National Center for Biotechnology Information

National Library of Medicine

National Institutes of Health

PubMed

All Databases

BLAST

OMIM

Books

TaxBrowser

Structure

Search

PubMed



for

Go

OMIM is...

- Online Mendelian Inheritance in Man
- catalog of human genes and genetic disorders
- edited by Dr. Victor McKusick, others at JHU



National Center for Biotechnology Information

National Library of Medicine

National Institutes of Health

PubMed

All Databases

BLAST

OMIM

Books

TaxBrowser

Structure

Search

PubMed

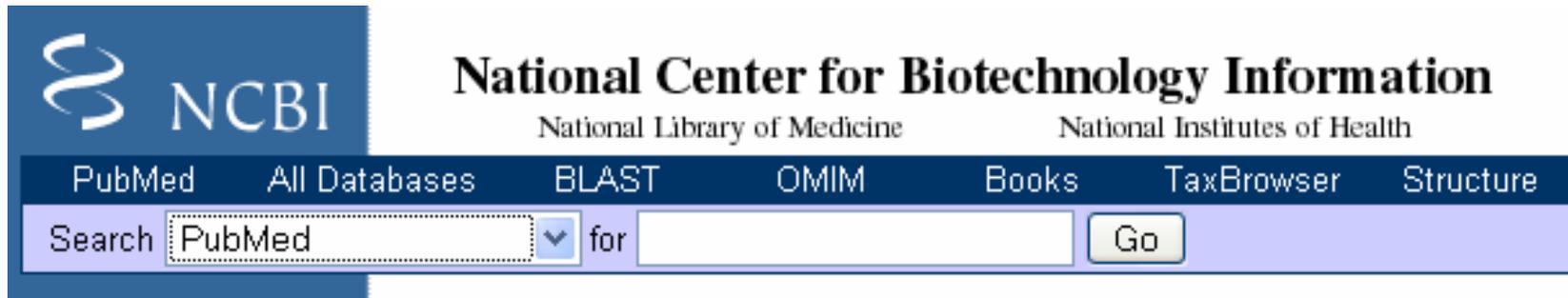


for

Go

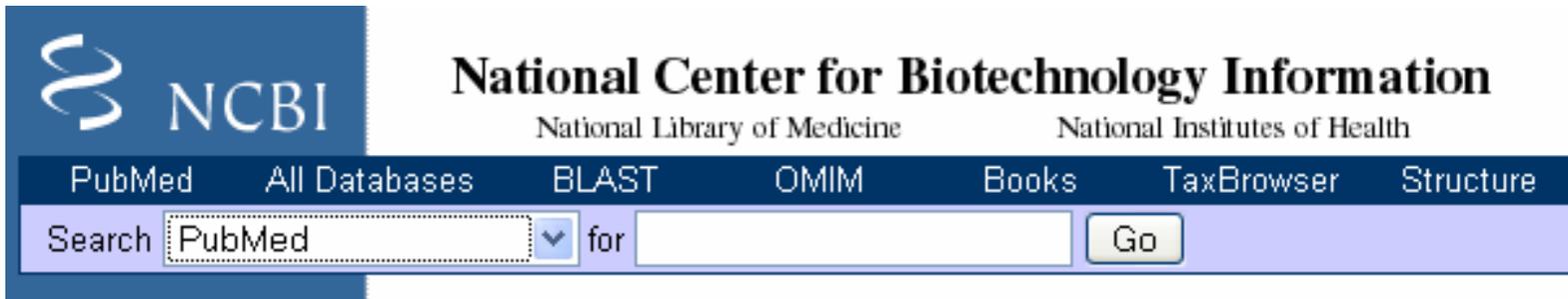
Books is...

- searchable resource of on-line books



TaxBrowser is...

- browser for the major divisions of living organisms (archaea, bacteria, eukaryota, viruses)
- taxonomy information such as genetic codes
- molecular data on extinct organisms



Structure site includes...

- Molecular Modelling Database (MMDB)
- biopolymer structures obtained from the Protein Data Bank (PDB)
- Cn3D (a 3D-structure viewer)
- vector alignment search tool (VAST)

Gene Search

The screenshot shows the NCBI Entrez Gene search interface. At the top, the NCBI logo is on the left, and the 'Entrez Gene' title is in the center. Below the title, there are tabs for 'All Databases', 'PubMed', 'Nucleotide', 'Protein', 'Genome', and 'Structure'. A search bar contains the text 'Gene' in a dropdown menu, followed by 'for rbp4'. To the right of the search bar are buttons for 'Go', 'Clear', and 'Save Search'. Below the search bar, there are buttons for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. A 'Display' section shows 'Summary' selected in a dropdown, 'Show 20' in another dropdown, and a 'Send to' dropdown. Below this, a summary bar shows 'All: 19', 'Current Only: 19', 'Genes Genomes: 17', and 'SNP GeneView: 8'. The main content area displays 'Items 1 - 19 of 19' and a list of five search results, each with a checkbox and a link to the gene entry. The results include species names like *Homo sapiens*, *Sus scrofa*, *Mus musculus*, and *Rattus norvegicus*.

NCBI

Entrez Gene

All Databases PubMed Nucleotide Protein Genome Structure

Search Gene for rbp4 Go Clear Save Search

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Send to

All: 19 Current Only: 19 Genes Genomes: 17 SNP GeneView: 8

Items 1 - 19 of 19

- 1: [RBP4](#)
Official Symbol: RBP4 **and Name:** retinol binding protein 4, plasma [*Homo sapiens*]
Other Designations: retinol-binding protein 4, interstitial; retinol-binding protein 4, plasma
Chromosome: 10; **Location:** 10q23-q24
MIIM: 180250
GeneID: 5950
- 2: [RBP4](#)
retinol-binding protein [*Sus scrofa*]
GeneID: 397124
- 3: [Rbp4](#)
Official Symbol: Rbp4 **and Name:** retinol binding protein 4, plasma [*Mus musculus*]
Other Aliases: Rbp-4
Other Designations: retinol binding protein 4, cellular
Chromosome: 19; **Location:** 19 38.0 cM
GeneID: 19662
- 4: [Rbp4](#)
Official Symbol: Rbp4 **and Name:** retinol binding protein 4, plasma [*Rattus norvegicus*]
Other Aliases: RBPA
Other Designations: retinol-binding protein
Chromosome: 1; **Location:** 1q53
GeneID: 25703
- 5: [RBP4](#)
retinol binding protein 4, plasma [*Pan troglodytes*]
Other Designations: retinol-binding protein 4, plasma

Entrez Gene

- Home
- About
- FAQ
- Help
- Gene Handbook
- Statistics
- Downloads (FTP)

Mailing Lists

- Gene
- RefSeq

Feedback

- Help Desk
- Corrections
- About GeneRIFs

Related Sites

- BLAST
- Entrez Genome
- Genome Projects
- Genomic Biology
- GEO
- HomoloGene
- Map Viewer
- OMIM
- Probe
- RefSeq
- UniGene
- UniSTS

Resources

Gene Name

Narrow down choices using Limit

The screenshot shows the Entrez Gene website interface. At the top, there is a search bar with "Gene" selected and a "for" field. Below the search bar are navigation tabs for "Limits", "Preview/Index", "History", "Clipboard", and "Details". The "Limits" tab is active, displaying a yellow header "Limits: Homo sapiens".

The "Limits" section includes the following instructions and options:

- To apply your query to all fields, leave this page unchanged.
- [Boolean operators](#) AND, OR, NOT must be in upper case.
- To limit your search to a specific field, select:
 - All Fields (selected in a dropdown menu)
 - or use [search field tags](#) enclosed in square brackets, e.g. aaa[title].
 - You may also limit your search by checking boxes below.

Below the instructions are several sections of checkboxes:

- Exclude:**
 - Mitochondria
 - Plasmids
 - Plastids
 - Pseudogenes
 - RefSeqs
 - NEWENTRY
- Include Only:**
 - Current Records
 - Genomic
 - Mitochondria
 - Plasmids
 - Plastids
 - RefSeqs
 - NEWENTRY
- Limit by RefSeq Status:**
 - Inferred
 - Known
 - Model
 - Predicted
 - Provisional
 - Reviewed
 - Validated
- Limit by Taxonomy:**
 - Mammalia
 - Bos taurus
 - Canis familiaris
 - Homo sapiens
 - Mus musculus
 - Pan troglodytes
 - Ratus norvegicus
 - Sus scrofa
 - Non-mammalian vertebrates
 - Danio rerio
 - Gallus gallus
 - Xenopus laevis
 - Xenopus tropicalis
 - Invertebrates
 - Anopheles gambiae
 - Caenorhabditis elegans
 - Drosophila melanogaster
 - Plasmodium falciparum
 - Plants
 - Arabidopsis thaliana
 - Oryza sativa
 - Fungi



Search Gene for rbp4 [Save Search](#)

Limits

Limits: **Homo sapiens**

Display Summary 20

All: 5

Items 1 - 5 of 5

1: [RBP4](#)

Official Symbol: RBP4 **and Name:** retinol binding protein 4, plasma [*Homo sapiens*]
Other Designations: retinol-binding protein 4, interstitial; retinol-binding protein 4, plasma
Chromosome: 10; **Location:** 10q23-q24
MIM: 180250
GeneID: 5950

2: [POLR2D](#)

Official Symbol: POLR2D **and Name:** polymerase (RNA) II (DNA directed) polypeptide D [*Homo sapiens*]
Other Aliases: HSRBP4, HSRPB4, RBP4
Other Designations: DNA directed RNA polymerase II polypeptide D; RNA polymerase II subunit hsRBP4
Chromosome: 2; **Location:** 2q21
MIM: 606017
GeneID: 5433

3: [IRF1](#)

Official Symbol: IRF1 **and Name:** interferon regulatory factor 1 [*Homo sapiens*]
Other Aliases: IRF-1, MAR
Chromosome: 5; **Location:** 5q31.1
MIM: 147575
GeneID: 3659

4: [TTR](#)

Official Symbol: TTR **and Name:** transthyretin (prealbumin, amyloidosis type I) [*Homo sapiens*]
Other Aliases: HsT2651, PALB, TBPA

Entrez Gene

[Home](#)
[About](#)
[FAQ](#)
[Help](#)
[Gene Handbook](#)
[Statistics](#)
[Downloads \(FTP\)](#)

Mailing Lists

[Gene](#)
[RefSeq](#)

Feedback

[Help Desk](#)
[Corrections](#)
[About GeneRIFs](#)

Related Sites

[BLAST](#)
[Entrez Genome](#)
[Genome Projects](#)
[Genomic Biology](#)
[GEO](#)
[HomoloGene](#)
[Map Viewer](#)
[OMIM](#)
[Probe](#)
[RefSeq](#)
[UniGene](#)
[UniSTS](#)

Search Gene for Go Clear

Limits Preview/Index History Clipboard Details

Display Full Report Show 5 Send to

All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 1

1: RBP4 retinol binding protein 4, plasma [*Homo sapiens*]

GeneID: 5950 Primary source: [HGNC:9922](#)

updated 07-Jul-2006

Summary

Official Symbol: RBP4 and **Name:** retinol binding protein 4, plasma provided by [HUGO Gene Nomenclature Committee](#)

See related: [HPRD:01580](#), [MIM:180250](#)

Gene type: protein coding

Gene name: RBP4

Gene description: retinol binding protein 4, plasma

RefSeq status: Reviewed

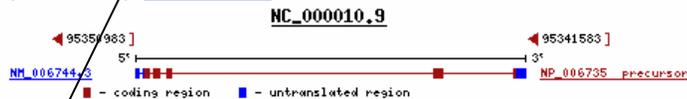
Organism: [Homo sapiens](#)

Lineage: *Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo*

Summary: This protein belongs to the lipocalin family and is the specific carrier for retinol (vitamin A alcohol) in the blood. It delivers retinol from the liver stores to the peripheral tissues. In plasma, the RBP-retinol complex interacts with transthyretin which prevents its loss by filtration through the kidney glomeruli. A deficiency of vitamin A blocks secretion of the binding protein posttranslationally and results in defective delivery and supply to the epidermal cells.

Genomic regions, transcripts, and products

(minus strand) [RefSeq below](#)

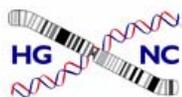


Genomic context

[See RBP4 in MapViewer](#)

chromosome: 10; **Location:** 10q23-q24

click



Symbol Report: **RBP4**

[About HGNC](#)[Gene Search](#)[Guidelines](#)[Gene Submission](#)[Downloads](#)

Giving unique and meaningful names to every human gene

 [Quick Gene Search](#)

Core Data		Database Links		
Approved Symbol +	RBP4	Accession Numbers +		
Approved Name +	retinol binding protein 4, plasma	BC020633	GenBank	UCSC Browser
HGNC ID +	HGNC:9922	RefSeq IDs +		
Status +	Approved	NM_006744	GenBank	UCSC Browser
Chromosome +	10q23-q24	Entrez Gene ID +		
Previous Symbols +		5950	Gene	Map Viewer
Previous Names +	"retinol-binding protein 4, plasma"	OMIM ID (mapped data) +		
Aliases +		180250	OMIM	
Gene Symbol Links		UniProt ID (mapped data) +		
		P02753	SwissProt	UniProt
Ensembl GeneView	GENATLAS	GeneCards		
GeneClinics/GeneTests	Vega			

See [Column definitions](#) for descriptions of the various data fields. Also see our [custom downloads](#) page for bulk access to our data

Click

HTML format of the gene

LOCUS BC020633 950 bp mRNA linear PRI 15-JUL-2006

DEFINITION Homo sapiens retinol binding protein 4, plasma, mRNA (cDNA clone MGC:22269 IMAGE:4767481), complete cds.

ACCESSION BC020633

VERSION BC020633.1 GI:18088325

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 950)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shemmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

CONSRMT Mammalian Gene Collection Program Team

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

PUBMED [12477932](#)

REFERENCE 2 (bases 1 to 950)
CONSRMT NIH MGC Project
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov

source 1..950
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:22269 IMAGE:4767481"
/tissue_type="Liver"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"

gene 1..950
/gene="RBP4"
/db_xref="GeneID:5950"
/db_xref="HGNC:9922"
/db_xref="MIM:180250"

CDS 85..690
/gene="RBP4"
/codon_start=1
/product="retinol binding protein 4, plasma"
/protein_id="AAH20633.1"
/db_xref="GI:18088325"
/db_xref="GeneID:5950"
/db_xref="HGNC:9922"
/db_xref="MIM:180250"
/translation="MKVWVWVLLAALGSGRAERDCRVSSFRVKNFKARFSGTWYA
MAKKDPEGLFLQDNIVAIEFSVDETGMSATLKRVRLLNNDVADIVGTFTDTPA
FKMKYVGVASFLQKGNDDHIVDVTDTYAVQVSCRLLNLDTGTCADSVSVFSDP
GLPPEAQKIVRQREELCLARQYRLVHWGVCYDGRSERLL"

misc difference 108
/gene="RBP4"
/note="'T' in cDNA is 'G' in the human genome; amino acid difference: 'F' in cDNA, 'L' in the human genome. The chimpanzee genome agrees with the human genomic sequence and not the cDNA."

misc difference 920..950
/gene="RBP4"
/note="polyA tail: 31 bases do not align to the human genome."

ORIGIN
1 cgcctccctc gctccacagc cgcctggact cggcggccag gcttgccgc ggttccctc
61 cggtyggcgy gattctctgg caagatgaag tgggtgtggc cgtctctct gttggcggc
121 ctgggcagcy cgcgcgggga gcgcgactgc cagtgagca gcttccagat caaggagaac
181 ttcgacaagc ctgcctcttc tgggcaactg tacgccatg ccaagaagga ccccgaggc
241 ctctttctgc aggacaacat cgtcggagc tctccgtgg acgagaccgg ccagatgagc
301 gccacagcca agggccgagc cgtctctttg aataactgg acgtgtgccc agacatggtg
361 ggcaccttca caaacaccga ggaacctgac aagttcaagc tgaagctctg gggctagcc
421 tctctttccc agaaaggaac tgatgaccac tggatgctg acacagcta cgcacagtat
481 gcctgagcgt actctgccc cctctgaac ctgatggca cctgtgctga cagctactc
541 tctgttttt cccgggaccc caacggctg ccccccagaag cgcagaagat tgaaggcag
601 cggcaggagc agctgtgact ggcacggcag tacagctga tegtccaaa cgttactgc
661 gatgcagat cagaaagaaa cctttttag caatatcaag aatctagtt catctgagaa
721 ctcttgatta gctctcagc ttcagctcta tttacttag gaatttaatt tgcctcttc
781 tcccactctt cctctcagtc ccataaacc ttcattacc ataaagata acgtgggggt
841 cagtgaaatc gcttgcttt cctgaaagt tctgggctt caagttcc acctctgattc
901 attaaacatc agtcaccgcc aaaaaaaaa aaaaaaaaa aaaaaaaaa

Fasta Format

NCBI Nucleotide

PubMed Nucleotide Protein Genome Structure

Search Nucleotide for [] Go Clear

Limits Preview/Index History

Display FASTA Show 5 Send to []

Range: from begin to end Reverse complemented strand Refresh

Select
Here

```
 1: BC020633. Reports Homo sapiens reti...[gi:18088325]
>gi|18088325|gb|BC020633.1| Homo sapiens retinol binding protein 4, plasma, mRNA (cDNA clone
CGCCTCCCTCGCTCCACGCGCGCCCGGACTCGGCGGCCAGGCTTGCGCGCGGTTCCCCTCCCGGTGGGCG
GATTCCTGGGCAAGATGAAAGTGGGTGTGGGCGCTCTTTCTGTTGGCGGGCGCTGGGCAGCGGCCGCGCGGA
GCGCGACTGCCGAGTGAGCAGCTTCCGAGTCAAGGAGAACTTCGACAAGGCTCGCTTCTCTGGGACCTGG
TACGCCATGGCCAAGAAGGACCCCGAGGGCCCTCTTTCTGCAGGACAACATCGTTCGCGGAGTTCTCCGTGG
ACGAGACCGGCCAGATGAGCGCCACAGCCAAGGGCCGAGTCCGTCTTTTGAATAACTGGGACGTGTGCGC
AGACATGGTGGGCACCTTCACAGACACCGAGGACCCTGCCAAGTTCAAGATGAAAGTACTGGGGCGTAGCC
TCCTTTCTCCAGAAAAGGAAATGATGACCCTGGATCGTTCGACACAGACTACGACACGTATGCCGTGCAGT
ACTCCTGCCGCCTCCTGAACTCGATGGCACCTGTGCTGACAGCTACTCCTTCGTGTTTTCCCGGGACCC
CAACGGCCTGCCCCCAGAAGCGCAGAAGATTGTAAGGCAGCGGCAGGAGGAGCTGTGCCTGGCCAGGCAG
TACAGGCTGATCGTCCACAACGGTTACTGCGATGGCAGATCAGAAAAGAAACCTTTTGTAGCAATATCAAG
AATCTAGTTTCATCTGAGAACTTCTGATTAGCTCTCAGTCTTCAGCTCTATTTATCTTAGGAGTTTAATT
TGCCCTTCTCTCCCCATCTCCCTCAGTTCCCATAAAAACCTTCATTACACATAAAGATACACGTGGGGGT
CAGTGAATCTGCTTGCCCTTTCCTGAAAAGTTTCTGGGGCTTAAGATTCCAGACTCTGATTCATTAACATAT
AGTCACCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

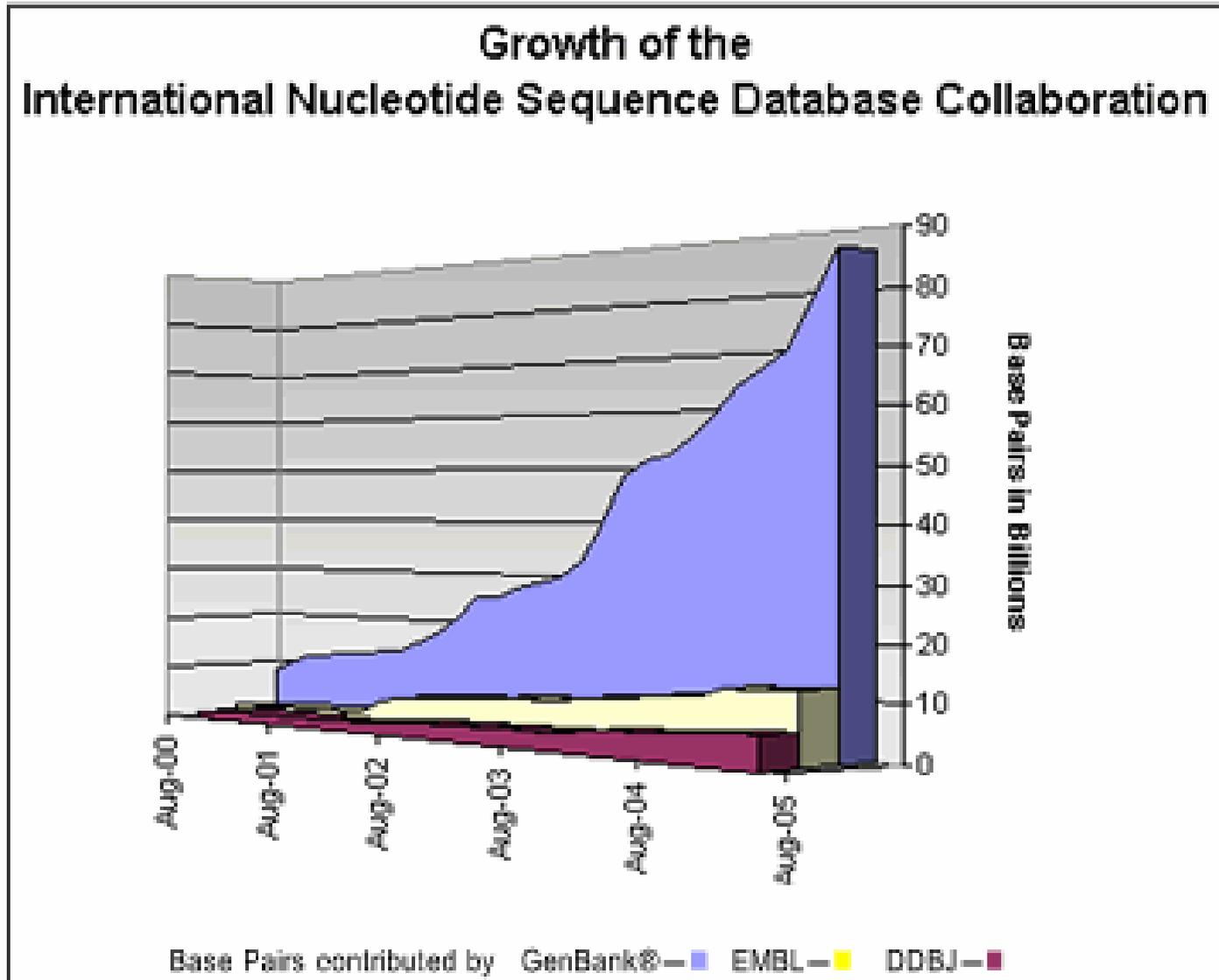
2. GenBank

- GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences
- Web: <http://www.ncbi.nlm.nih.gov/Genbank/index.html>
- Contain publicly available DNA sequences for more than 165000 named organisms
- Accessible through NCBI's retrieval system, Entrez
- BLAST provides sequence similarity searches of GenBank
- Complete releases are also available by FTP.

Data Sources of GenBank

- Sequence submission from authors
- Bulk submission of expressed sequence tag (EST), genome survey sequence from sequencing centers
- Sequences of issued patents from US Office of Patents and Trademarks (USPTO)
- EMBL Data Library in UK and DNA Databank of Japan (DDBJ)

Growth of GenBank



<http://www.ncbi.nlm.nih.gov/Genbank/index.html>

Description of Entry

- Each record has a unique identifier (accession number) (never change)
- Each sequence has a unique id: gi + unique number
- Version number. When a change is made to a sequence given in a GenBank record, a new gi number is issued to the sequence and the version is incremented. The older sequence remains available under the old “accession version” identifier and gi.
- Description of sequence
- Scientific name
- Taxonomy of the source organism
- Bibliographic reference
- Regions of biological significance such as coding regions and their protein translations, transcription units, repeat regions, and sites of mutations.

(<http://www.ncbi.nlm.nih.gov/collab/FT/index.html>)

Entrez to GenBank

Entrez cross-database search - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.ncbi.nlm.nih.gov/query/query.fcgi

Getting Started Latest Headlines

GenBank -- Benson et al. 33 (Supplement 1)... NCBI Taxonomy Homepage The DDBJ/EMBL/GenBank Feature Table: De... Entrez cross-database search

NCBI

Entrez, The Life Sciences Search Engine

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

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
Journals: detailed information about the journals indexed in PubMed and other Entrez databases	MeSH: detailed information about NLM's controlled vocabulary
NLM Catalog: catalog of books, journals, and audiovisuals in the NLM collections	

Enter terms and click 'GO' to run the search against ALL the databases, OR
Click Database Name or Icon to go directly to the Search Page for that database, OR
Click Question Mark for a short explanation of that database.

Done

Click
here

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

Accession num

Length
Type

Type keyword: hiv-1

Version

Sequence id

NCBI Nucleotide search results for 'hiv-1'. The interface shows search filters, a list of 6 results, and a detailed view of the first result (DQ863896). The detailed view includes fields for LOCUS, DEFINITION, ACCESSION, VERSION, KEYWORDS, SOURCE, ORGANISM, REFERENCE, AUTHORS, TITLE, JOURNAL, PUBMED, and FEATURES. The FEATURES section shows source information and a partial protein translation.

LOCUS DQ863895 519 bp RNA linear VRL 21-AUG-2006
 DEFINITION HIV-1 isolate patient2.egg41.w48 clone 5 from France envelope glycoprotein (env) gene, partial cds.
 ACCESSION DQ863895
 VERSION DQ863895.1 GI:112379288

SOURCE Human immunodeficiency virus 1 (HIV-1)
 ORGANISM [Human immunodeficiency virus 1](#)
 Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae; Lentivirus; Primate lentivirus group.
 REFERENCE 1 (bases 1 to 519)
 AUTHORS Labrosse,B., Morand-Joubert,L., Goubard,A., Rochas,S., Labernardiere,J.L., Pacanowski,J., Meynard,J.L., Hance,A.J., Clavel,F. and Mammano,F.
 TITLE Role of the envelope genetic context in the development of enfuvirtide resistance in human immunodeficiency virus type 1-infected patients
 JOURNAL J. Virol. 80 (17), 8807-8819 (2006)
 PUBMED [16912327](#)
 REFERENCE 2 (bases 1 to 519)
 AUTHORS Labrosse,B., Morand-Joubert,L., Goubard,A., Rochas,S., Labernardiere,J.-L., Pacanowski,J., Meynard,J.-L., Hance,A.J., Clavel,F. and Mammano,F.
 TITLE Direct Submission
 JOURNAL Submitted (18-JUL-2006) Unite de Recherche Antivirale Inserm U552, 46 rue Henri Huchard, Paris 75018, France

FEATURES
 source
 1..519
 /organism="Human immunodeficiency virus 1"
 /virion
 /mol_type="genomic RNA"
 /isolate="patient2.egg41.w48"
 /isolation_source="patient 2 at week 48"
 /db_xref="taxon:11676"
 /clone="5"
 /country="France"
 <1..>519
 /gene="env"
 <1..>519
 /gene="env"
 /note="ectodomain of the transmembrane envelope glycoprotein"
 /codon_start=1
 /product="envelope glycoprotein"
 /protein_id="ABI16427.1"
 /db_xref="GI:112379289"
 /translation="AIGALGANFLGFLGAAGSTMGAAALTLTVQARQLLSGIVQQNLLRAIEAQQHLLQLTVNGIKQLQARVLAVERYLQDQHLGLGNGCSGKLICTTTVPNNWSNKSLEEIWDNHTWQNDREISNYTFSIYTLIEKAQQQEKNEQELLALDKWASLWNWFDISKWLWYIK"

Protein Sequence

ORIGIN
 1 gcaattggag cactaggagc tatgttcctt gggttcttgg gacgacgag aagcactatg
 61 ggcgacgag cactgacgt gacggtacag gccagacaac tgttgtctgg tatagtgcaa
 121 cagcagaaca atctgtgag agctattgag gcgcaacaac atctgttgca actcacagtc

Nucleotide Sequence

Search GenBank by Accession Number

The screenshot displays the NCBI Entrez Nucleotide search interface. At the top, the NCBI logo is on the left, and a decorative graphic with the text 'Entrez Nucleotide' is on the right. Below the logo, a navigation bar includes 'All Databases', 'PubMed', 'Nucleotide', 'Protein', 'Genome', and 'Structure'. The search bar contains the text 'Search Nucleotide for DQ863895' with 'Go', 'Clear', and 'Save Search' buttons. Below the search bar are buttons for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The 'Display' dropdown is set to 'Summary', and 'Show' is set to '20'. A summary bar shows 'All: 1' with sub-counts for 'bacteria: 0', 'mRNA: 0', and 'RefSeq: 0'. A yellow banner below the summary bar reads 'Show only records from: CoreNucleotide (1), EST (0), GSS (0). [What's this?]'. The search results list one entry: 1: [DQ863895](#) Reports. The description for this entry is 'HIV-1 isolate patient2.egp41.w48 clone 5 from France envelope glycoprotein (env) gene, partial cds gi|112379288|gb|DQ863895.1|[112379288]'. A blue sidebar on the left contains navigation links: 'About Entrez', 'Entrez Nucleotide Help | FAQ', 'Entrez Tools', 'Check sequence revision history', 'LinkOut', 'My NCBI (Cubby)', and 'Related resources BLAST'. A black arrow points from the text 'Type: accession number' at the bottom right to the accession number 'DQ863895' in the search bar.

Type: accession number

BLAST Search GenBank and other DNA data

NCBI → BLAST Latest news: 7 May 2006 : BLAST 2.2.14 released

About

- Getting started
- News
- FAQs

More info

- NAR 2004
- NCBI Handbook
- The Statistics of Sequence Similarity Scores

Software

- Downloads
- Developer info

Other resources

- References
- NCBI Contributors
- Mailing list
- Contact us

The **Basic Local Alignment Search Tool (BLAST)** finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

Nucleotide <ul style="list-style-type: none">• Quickly search for highly similar sequences (megablast)• Quickly search for divergent sequences (discontiguous megablast)• Nucleotide-nucleotide BLAST (blastn)• Search for short, nearly exact matches• Search trace archives with megablast or discontiguous megablast	Protein <ul style="list-style-type: none">• Protein-protein BLAST (blastp)• Position-specific iterated and pattern-hit initiated BLAST (PSI- and PHI-BLAST)• Search for short, nearly exact matches• Search the conserved domain database (rpsblast)• Protein homology by domain architecture (cdart)
Translated <ul style="list-style-type: none">• Translated query vs. protein database (blastx)• Protein query vs. translated database (tblastn)• Translated query vs. translated database (tblastx)	Genomes <ul style="list-style-type: none">• Human, mouse, rat, chimp, cow, pig, dog, sheep, cat• Chicken, puffer fish, zebrafish• Fly, honey bee, other insects• Microbes, environmental samples• Plants, nematodes• Fungi, protozoa, other eukaryotes
Special <ul style="list-style-type: none">• Search for gene expression data (GEO BLAST)• Align two sequences (bl2seq)• Screen for vector contamination (VecScreen)• Immunoglobulin BLAST (IgBlast)• SNP BLAST	Meta <ul style="list-style-type: none">• Retrieve results

Click

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

```
gcaattggagcactaggagctatgttccttgggttcttgggagcagcagg
aagcactatggcgagcggcactgacgctgacggtacagccacacacac
tggtgtctggtatagtgcaacagcagaacaatctgctgagagctattgag
ggcaacaacatctgttgcaactcacagtctggggcattaaagcagctcca
ggcaagagtctgtgtggaagatcactacaggatcagcacctccttag
gaattggggttgcctctggaaaactcatctgcaccactactgtgcttgg
```

Set subsequence From: To:

Choose database:

Now: or

Rank list of similar DNA

Your request has been successfully submitted and put into the Blast Queue.

Query = (519 letters)

The request ID is

or

The results are estimated to be ready in 13 seconds but may be done sooner.

Please press "FORMAT!" when you wish to check your results. You may change the formatting options for your result other valid request ID to see other recent jobs.

RID=115677820-29076-193742948320.BLASTQ4, - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.ncbi.nlm.nih.gov/BLAST/blast.cgi

Distance tree of results **NEW**

Sequences producing significant alignments:

	Score	E
	(Bits)	Value
gi 112379288 gb DQ863895.1	HIV-1 isolate patient2.egg41.w48 ...	1029 0.0
gi 112379254 gb DQ863878.1	HIV-1 isolate patient2.egg41.w31 ...	989 0.0
gi 112379160 gb DQ863830.1	HIV-1 isolate patient2.egg41.w-13 ...	989 0.0
gi 112379178 gb DQ863839.1	HIV-1 isolate patient2.egg41.w-13 ...	973 0.0
gi 112379162 gb DQ863831.1	HIV-1 isolate patient2.egg41.w-13 ...	973 0.0
gi 112379284 gb DQ863893.1	HIV-1 isolate patient2.egg41.w48 ...	965 0.0
gi 112379266 gb DQ863894.1	HIV-1 isolate patient2.egg41.w31 ...	965 0.0
gi 112379182 gb DQ863841.1	HIV-1 isolate patient2.egg41.w-13 ...	965 0.0
gi 112379168 gb DQ863834.1	HIV-1 isolate patient2.egg41.w-13 ...	965 0.0
gi 112379164 gb DQ863832.1	HIV-1 isolate patient2.egg41.w-13 ...	965 0.0
gi 112379156 gb DQ863828.1	HIV-1 isolate patient2.egg41.w-13 ...	965 0.0
gi 112379190 gb DQ863845.1	HIV-1 isolate patient2.egg41.w-13 ...	957 0.0
gi 112379188 gb DQ863844.1	HIV-1 isolate patient2.egg41.w-13 ...	957 0.0
gi 112379186 gb DQ863843.1	HIV-1 isolate patient2.egg41.w-13 ...	957 0.0
gi 112379184 gb DQ863842.1	HIV-1 isolate patient2.egg41.w-13 ...	957 0.0
gi 112379174 gb DQ863837.1	HIV-1 isolate patient2.egg41.w-13 ...	957 0.0
gi 112379166 gb DQ863833.1	HIV-1 isolate patient2.egg41.w-13 ...	957 0.0
gi 112379158 gb DQ863829.1	HIV-1 isolate patient2.egg41.w-13 ...	957 0.0
gi 112379154 gb DQ863827.1	HIV-1 isolate patient2.egg41.w-13 ...	957 0.0
gi 112379286 gb DQ863894.1	HIV-1 isolate patient2.egg41.w48 ...	950 0.0
gi 112379180 gb DQ863840.1	HIV-1 isolate patient2.egg41.w-13 ...	950 0.0
gi 112379172 gb DQ863836.1	HIV-1 isolate patient2.egg41.w-13 ...	950 0.0
gi 112379170 gb DQ863835.1	HIV-1 isolate patient2.egg41.w-13 ...	950 0.0
gi 112379152 gb DQ863826.1	HIV-1 isolate patient2.egg41.w-22 ...	950 0.0
gi 112379264 gb DQ863883.1	HIV-1 isolate patient2.egg41.w31 ...	942 0.0
gi 112379272 gb DQ863887.1	HIV-1 isolate patient2.egg41.w31 ...	934 0.0
gi 112379270 gb DQ863886.1	HIV-1 isolate patient2.egg41.w31 ...	934 0.0
gi 112379280 gb DQ863891.1	HIV-1 isolate patient2.egg41.w48 ...	926 0.0
gi 112379278 gb DQ863890.1	HIV-1 isolate patient2.egg41.w31 ...	926 0.0
gi 112379260 gb DQ863881.1	HIV-1 isolate patient2.egg41.w31 ...	926 0.0
gi 112379256 gb DQ863879.1	HIV-1 isolate patient2.egg41.w31 ...	926 0.0
gi 112379229 gb DQ863865.1	HIV-1 isolate patient2.egg41.w13 ...	926 0.0
gi 112379199 gb DQ863850.1	HIV-1 isolate patient2.egg41.w9 c ...	926 0.0
gi 112379290 gb DQ863896.1	HIV-1 isolate patient2.egg41.w48 ...	918 0.0
gi 112379282 gb DQ863892.1	HIV-1 isolate patient2.egg41.w48 ...	918 0.0
gi 112379274 gb DQ863888.1	HIV-1 isolate patient2.egg41.w31 ...	918 0.0
gi 112379176 gb DQ863838.1	HIV-1 isolate patient2.egg41.w-13 ...	918 0.0
gi 112379276 gb DQ863889.1	HIV-1 isolate patient2.egg41.w31 ...	910 0.0
gi 112379268 gb DQ863885.1	HIV-1 isolate patient2.egg41.w31 ...	910 0.0
gi 112379258 gb DQ863880.1	HIV-1 isolate patient2.egg41.w31 ...	910 0.0
gi 112379207 gb DQ863854.1	HIV-1 isolate patient2.egg41.w9 c ...	910 0.0
gi 112379203 gb DQ863852.1	HIV-1 isolate patient2.egg41.w9 c ...	910 0.0
gi 112379248 gb DQ863875.1	HIV-1 isolate patient2.egg41.w13 ...	902 0.0
gi 112379233 gb DQ863867.1	HIV-1 isolate patient2.egg41.w13 ...	902 0.0
gi 112379225 gb DQ863863.1	HIV-1 isolate patient2.egg41.w13 ...	902 0.0
gi 112379217 gb DQ863859.1	HIV-1 isolate patient2.egg41.w9 c ...	902 0.0
gi 112379252 gb DQ863877.1	HIV-1 isolate patient2.egg41.w13 ...	894 0.0
gi 112379250 gb DQ863876.1	HIV-1 isolate patient2.egg41.w13 ...	894 0.0
gi 112379242 gb DQ863872.1	HIV-1 isolate patient2.egg41.w13 ...	894 0.0
gi 112379238 gb DQ863870.1	HIV-1 isolate patient2.egg41.w13 ...	894 0.0
gi 112379236 gb DQ863869.1	HIV-1 isolate patient2.egg41.w13 ...	894 0.0

Done

3. Protein Sequence Database: UniProt, SwissProt and PIR

SwissProt:

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

PIR (Protein Information
Resource)

UniProt is a combination of
SwissProt and PIR

The screenshot shows the UniProt website interface. At the top, there is a navigation bar with links for 'ExPASy Home page', 'Site Map', 'Search ExPASy', 'Contact us', 'PROSITE', and 'Proteomics tools'. Below this is a search bar with a dropdown menu set to 'Swiss-Prot/TrEMBL' and a search input field containing 'hiv-1'. The main content area features the 'Swiss-Prot Protein knowledgebase TrEMBL Computer-annotated supplement to Swiss-Prot' logo and the 'UniProt the universal protein resource' logo. A section titled 'The UniProt Knowledgebase consists of:' lists two databases: 'UniProtKB/Swiss-Prot' and 'UniProtKB/TrEMBL'. Below this, it states 'These databases are developed by the Swiss-Prot groups at SIB and at EBI.' A yellow box highlights 'Swiss-Prot headlines' with the text '10'000 species in UniProtKB/Swiss-Prot! (Read more ...)'. A blue box titled 'Access to the UniProt Knowledgebase' lists various search methods: 'SRS - Access to UniProtKB/Swiss-Prot, UniProtKB/TrEMBL and other databases using the Sequence Retrieval System', 'Full text search in the UniProt Knowledgebase', 'Advanced search in the UniProt Knowledgebase by description, gene name and organism (can be used to create html links to UniProt Knowledgebase queries)', 'Taxonomy browser (NEWT)', 'BLAST similarity search', 'by description or identification (any word in the DE, OS, OG, GN and ID lines)', 'by citation (RL line; UniProtKB/Swiss-Prot only)', 'Retrieve a list of UniProtKB entries', 'Randomly retrieve a UniProtKB entry', 'UniProtKB Sequence/Annotation Version Database', and 'Swiss-Prot ID tracker'.

Database search by fields

Free Text Search

Fields search by SRS

[Top Page](#) [Query Form](#) [Query Manager](#) [View Manager](#) [Databanks](#) [Help](#)

Search [SWISS PROT](#) [TREMBL](#)

 Combine searches with Append wildcard "*" to words.

Info	AccNumber	
Info	SubmissionDate	
Info	Organism	
Info	Keywords	

Include fields in output	<input type="checkbox"/> ID <input type="checkbox"/> AccNumber <input type="checkbox"/> Date <input type="checkbox"/> SubmissionDate <input type="checkbox"/> Description <input type="checkbox"/> GeneName <input type="checkbox"/> Keywords	Entry List in chunks of <input type="button" value="100"/>
Display in	<input checked="" type="radio"/> list <input type="radio"/> table	Sequence Format <input type="button" value="* default *"/>
		Use view <input type="button" value="Short Description"/>
		Retrieve set of <input type="button" value="entry"/>

Separate multiple values by & (and), | (or), ! (and not)

Free Text Search

UniProt Knowledgebase (Swiss-Prot/TREMBL)

Full text search

The search is performed on the current biweekly UniProt Knowledgebase release (Swiss-Prot and TrEMBL).
(UniProtKB/Swiss-Prot Release 50.6 of 22-Aug-2006, 230160 entries; UniProtKB/TrEMBL Release 33.6 of 22-Aug-2006, 3081935 entries)
The index has been created by the [Glimpse](#) search engine.

Enter search terms:

 Prefix and append wildcard "*" to words.
Search in: Swiss-Prot TrEMBL

For more directed searches, you can use the [Advanced search form](#), or the Sequence Retrieval System [SRS](#).

Blast Search

SIB BLAST Network Service

This NCBI BLAST2 service is maintained by the [Swiss Institute of Bioinformatics](#).

Click on the  icons to access the [online BLAST help](#).

Accession number or sequence

Enter a Swiss-Prot/TREMBL accession number or a **PROTEIN** sequence in **RAW** format

Output format:

or

Choose the appropriate BLAST  program and  database:

blastp - query against the UniProt Knowledgebase (Swiss-Prot + TrEMBL)

SwissProt Blast Search Results

SIB BLAST Network Service

This NCBI BLAST2 service is maintained by the [Swiss Institute of Bioinformatics](#).

Click on the icons to access the [online BLAST help](#).

● Accession number or sequence

Enter a Swiss-Prot/TrEMBL accession number or a **PROTEIN** sequence in **RAW** format.

```
AIGALGANFLGFLGAAGSTNGAALTLTVQARQLLSGIVQQNN  
LLRAIEAQHLLQLTVGKIQARVLAVERYLDQHLGIGWCSGLICTTVPWNN  
SWSNKSLEEIWDNHTMQMDREISNYTTFIYTLIEKAQSQEKEQELLALDKWASLW  
NUPDISKWLWYIK
```

Output format:

or

Choose the appropriate BLAST program and database:

blastp - query against the UniProt Knowledgebase (Swiss-Prot + TrEMBL)

List of potentially matching sequences

Send selected sequences to:

Include query sequence

Db	AC	Description	Score	E-value
<input type="checkbox"/>	tr Q1AIG7	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	275	6e-73
<input type="checkbox"/>	tr Q1AJ13	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	271	5e-72
<input type="checkbox"/>	tr Q1AIW4	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	271	5e-72
<input type="checkbox"/>	tr Q2NN45	_SHIV1 Env protein [env] [Human immunodeficiency virus 1]	271	5e-72
<input type="checkbox"/>	tr Q4ADE1	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	271	7e-72
<input type="checkbox"/>	tr Q9E1R6	_SHIV1 Envelope protein [env] [Human immunodeficiency v...]	271	7e-72
<input type="checkbox"/>	tr Q1AID9	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	271	9e-72
<input type="checkbox"/>	tr Q1AJ39	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	271	9e-72
<input type="checkbox"/>	tr Q72993	_SHIV1 Env protein [env] [Human immunodeficiency virus 1]	270	1e-71
<input type="checkbox"/>	tr Q1AJE0	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	270	2e-71
<input type="checkbox"/>	tr Q1AJ11	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	270	2e-71
<input type="checkbox"/>	tr Q1AJE1	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	270	2e-71
<input type="checkbox"/>	tr Q1AJ12	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	270	2e-71
<input type="checkbox"/>	tr Q1AJ14	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	270	2e-71
<input type="checkbox"/>	tr Q2NN07	_SHIV1 Envelope glycoprotein [env] [Human immunodeficie...]	270	2e-71
<input type="checkbox"/>	tr Q6E7N0	_SHIV1 Envelope glycoprotein [env] [Human immunodeficie...]	270	2e-71
<input type="checkbox"/>	tr Q9DL90	_SHIV1 Envelope glycoprotein [env] [Human immunodeficie...]	270	2e-71
<input type="checkbox"/>	tr Q6BC04	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	270	2e-71
<input type="checkbox"/>	tr Q73336	_SHIV1 Envelope polyprotein [env] [Human immunodeficien...]	269	3e-71
<input type="checkbox"/>	tr Q1AID8	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	269	3e-71
<input type="checkbox"/>	tr Q1AJE2	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	269	3e-71
<input type="checkbox"/>	tr Q1AIU7	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	269	3e-71
<input type="checkbox"/>	tr Q2NNK8	_SHIV1 Env protein [env] [Human immunodeficiency virus 1]	269	3e-71
<input type="checkbox"/>	tr Q92903	_SHIV1 Envelope glycoprotein [env] [Human immunodeficie...]	269	3e-71
<input type="checkbox"/>	sp P35961	ENV_HV1Y2 Envelope glycoprotein gp160 precursor [Env p...]	269	3e-71
<input type="checkbox"/>	tr Q1AID5	_SHIV1 Envelope glycoprotein (Fragment) [env] [Human im...]	269	3e-71

UniProtKB/TrEMBL entry Q1AIG7



[\[Entry info\]](#) [\[Name and origin\]](#) [\[References\]](#) [\[Comments\]](#) [\[Cross-references\]](#) [\[Keywords\]](#) [\[Features\]](#) [\[Sequence\]](#) [\[Tools\]](#)

Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.

Entry information

Entry name	Q1AIG7_SHIV1
Primary accession number	Q1AIG7
Secondary accession numbers	None
Integrated into TrEMBL on	July 11, 2006
Sequence was last modified on	July 11, 2006 (Sequence version 1)
Annotations were last modified on	July 11, 2006 (Entry version 1)

Name and origin of the protein

Protein name	Envelope glycoprotein [Fragment]
Synonyms	None
Gene name	Name: env
From	Human immunodeficiency virus 1 [TaxID: 11676]
Taxonomy	Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae; Lentivirus; Primate lentivirus group.

References

[1] NUCLEOTIDE SEQUENCE:

Loutfy M.R., Walmsley S.L., Raboud J.M., Montaner J.S.G., Wynhoven B., Brumme Z.L., Mo T., Smail F., Rouleau D., Gill J., Schlech W., Gough K., Rachlis A., Cameron B., Lapointe N., Trottier B., Shafran S., Harrigan P.R.;
"Assay of HIV gp-41 amino acid sequence to identify baseline variation and development of mutations in patients failing enfuvirtide."
Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

Comments

Copyright

Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>. Distributed under the Creative Commons Attribution-NoDerivs License.

Cross-references

Sequence databases

EMBL [DQ076989](#); [AAZ84395.1](#); -; Genomic_RNA.[EMBL / GenBank / DDBJ] [CoDingSequence]

3D structure databases

ModBase [Q1AIG7](#).

2D gel databases

SWISS-2DPAGE: Get region on 2D PAGE.

PIR

Welcome to PIR [Protein Information Resource] - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://pir.georgetown.edu/

GenBank Overview Welcome to PIR [Protein Information...]

PIR Protein Information Resource UniProt CONSORTIUM MEMBER

TLALPN---RRVADHLLH
LIGCLRNCSAVTAARQLAE
VTGFSN---ARTTAQHVKK

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
Current release: 8.5

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

PIRSF
Protein Family Classification System

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

ProClass
Integrated Protein Knowledgebase

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

ProLINK
Literature, Information & Knowledge

- Source for text mining and ontology development
- Biothesaurus of protein/gene names and RLIMS-P text mining tool
- Bibliography mapping

OTHER RESOURCE

- Proteomics: NIAID Biodefense Proteomics Admin. Center
- PIR Grid-Enablement: Data node on NC's caBL3

P PEPTIDE SEARCH ?
DATABASE: UniProtKB
Use single letter amino acid code

T TEXT SEARCH ?
DATABASE: ProClass

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

Copyright © 2005 - 2006 Protein Information Resource, Georgetown University Medical Center
3300 Whitehaven Street, NW, Suite 1200, Washington, DC 20007, USA

Done

Sequence Search

Text Search

BLAST Search Form

Retrieve sequences similar to your query

1. Select a database: [UniProtKB](#) (or restricted by organism)
 [UniRef100](#)

2. Insert the query sequence below (single letter [amino acid code](#))
or enter ">" followed by a [UniProtKB identifier](#):

```
AIGALGAMFLGFLGAAGSTMGAAALTLTVQARQLLSGIVQQONN  
LLRAIEAQHLLQLTVWGIKQLQARVLAVERYLQDQHLLGIWGCSGKLICTTVPWNN  
SWSNKSLEEIWDNMTWMQWDREISNYTSF IYTLIEKAQNQQEKNEQELLALDKWASLW  
NWFDISKWLWYIK
```

Options

Example: P53039 ([annotated output](#))

Note: DO NOT repeat search within a short period without waiting for results.
Delays may be experienced due to heavy loads on our server or network traffic.

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

250 proteins | 13 pages | 20 / page | [K<<<](#) | [1](#) | [2](#) | [3](#) | [4](#) | [5](#) | [>>>](#)

Save Result As: [TABLE](#) | [FASTA](#)

BLAST FASTA Pattern Match Multiple Alignment Domain Display

Protein AC/ID	Protein Name	Length	Organism Name	PIRSF ID	SSearch		BLAST Search		
					Overlap	%iden	E-Value	Score	Alignment
<input type="checkbox"/> Q9E1R6/Q9E1R6_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope protein <small>BioThesaurus</small>	851	Human immunodeficiency virus 1	PIRSF003843	173	<u>89</u>	1e-86	320	
<input type="checkbox"/> Q1HS90/Q1HS90_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope glycoprotein <small>BioThesaurus</small>	851	Human immunodeficiency virus 1		173	<u>88</u>	7e-86	318	
<input type="checkbox"/> Q1HS80/Q1HS80_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope glycoprotein <small>BioThesaurus</small>	851	Human immunodeficiency virus 1		173	<u>88</u>	5e-86	318	
<input type="checkbox"/> Q1HS86/Q1HS86_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope glycoprotein <small>BioThesaurus</small>	851	Human immunodeficiency virus 1		173	<u>88</u>	6e-86	318	
<input type="checkbox"/> Q1HS89/Q1HS89_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope glycoprotein <small>BioThesaurus</small>	852	Human immunodeficiency virus 1		173	<u>88</u>	7e-86	318	
<input type="checkbox"/> Q4A400/Q4A400_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope glycoprotein ... <small>BioThesaurus</small>	677	Human immunodeficiency virus 1		173	<u>87</u>	1e-85	317	
<input type="checkbox"/> Q1HS91/Q1HS91_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope glycoprotein <small>BioThesaurus</small>	851	Human immunodeficiency virus 1		173	<u>88</u>	8e-86	317	
<input type="checkbox"/> Q1HS84/Q1HS84_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope glycoprotein <small>BioThesaurus</small>	851	Human immunodeficiency virus 1		173	<u>88</u>	8e-86	317	
<input type="checkbox"/> Q1HS85/Q1HS85_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope glycoprotein <small>BioThesaurus</small>	851	Human immunodeficiency virus 1		173	<u>88</u>	8e-86	317	
<input type="checkbox"/> Q1HS88/Q1HS88_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope glycoprotein <small>BioThesaurus</small>	851	Human immunodeficiency virus 1		173	<u>88</u>	8e-86	317	
<input type="checkbox"/> Q1HS93/Q1HS93_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope glycoprotein <small>BioThesaurus</small>	852	Human immunodeficiency virus 1		173	<u>88</u>	8e-86	317	
<input type="checkbox"/> Q1HS92/Q1HS92_9HIV1 <small>/ProClass UniProtKB/TrEMBL</small>	Envelope glycoprotein <small>BioThesaurus</small>	852	Human immunodeficiency virus 1		173	<u>88</u>	8e-86	317	

Click

iProClass Summary Report for UniProtKB Entry: Q9E1R6

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

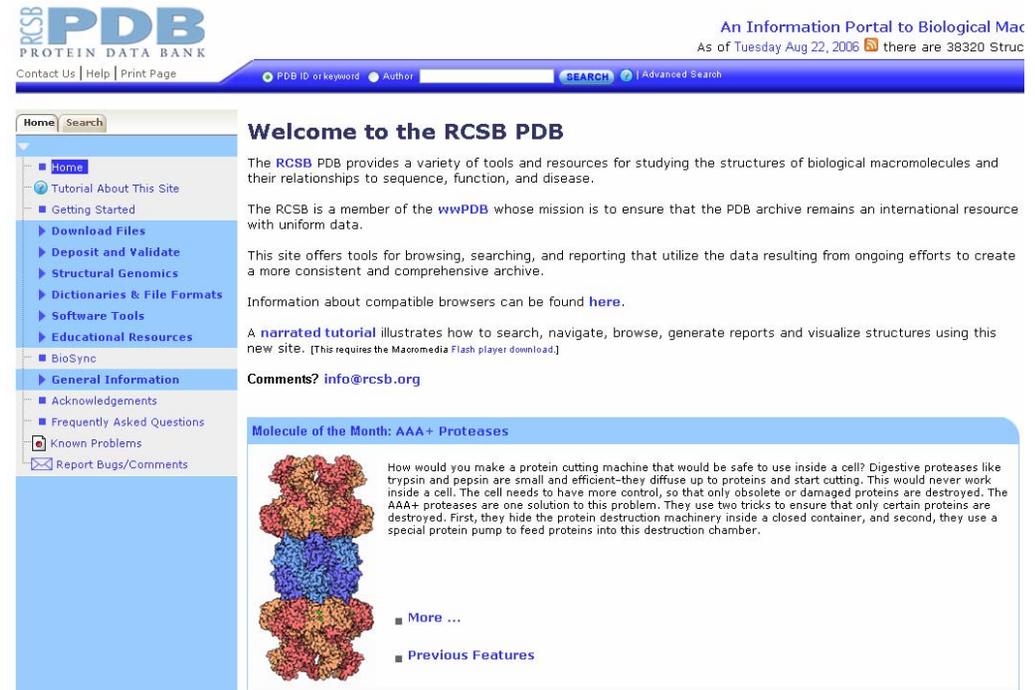
GENERAL INFORMATION			
Protein Name and ID	UniProtKB ID	UniProtKB Accession	Protein Name
	Q9E1R6_9HIV1	Q9E1R6	Envelope protein
GenPept: AAG22515.1			
Taxonomy	<i>Source Organism:</i> Human immunodeficiency virus 1 <i>Taxon Group:</i> Virus <i>NCBI Taxon:</i> 11676 <i>Lineage:</i> Viruses; Retro-transcribingviruses; Retroviridae; Orthoretrovirinae; Lentivirus; PrimateLentivirusgroup.		
Gene Name	env		
Keywords	aids; envelope protein; transmembrane		

CROSS-REFERENCES	
Bibliography	►View Bibliography Information ►Submit Bibliography PubMed: PMID: 10984542
DNA Sequence	GenBank/EMBL/DDBJ: AF277072
Ontology	<i>Molecular Function</i> GO: 0005198 :structural molecule activity [INTERPRO ; evidence: IEA][SPKW ; evidence: IEA] <i>Cellular Component</i> GO: 0016021 :integral to membrane [SPKW ; evidence: IEA] GO: 0019031 :viral envelope [INTERPRO ; evidence: IEA][SPKW ; evidence: IEA] GO: 0019028 :viral capsid [SPKW ; evidence: IEA]
Structure	Modbase: Q9E1R6 Target Region: (82-486), Sequence Identity: 86.00%, E-Value: 0, Model Score: 1.00, Template PDB Code: 1q9m:G

FAMILY CLASSIFICATION	
UniRef	UniRef100_Q9E1R6 ; UniRef90_Q9E1S8 ; UniRef50_P17281
PIRSF	PIRSF003843 type E retrovirus env polyprotein
Pfam Domain	Pfam: PF00516 : Envelope glycoprotein GP120 (33-505) Pfam: PF00517 : Envelope Polyprotein GP41 (514-851)

4. Protein Data Bank

- Web address:
<http://www.rcsb.org/>
- Central repository of protein and RNA structures
- Support search by id, author, or keyword



The screenshot shows the RCSB Protein Data Bank homepage. At the top left is the logo for RCSB PDB (Protein Data Bank). To the right, it says "An Information Portal to Biological Macromolecules" and "As of Tuesday Aug 22, 2006 there are 38320 Structures". Below the logo is a search bar with options for "PDB ID or keyword", "Author", and "Advanced Search". A navigation menu on the left includes links for Home, Tutorial, Getting Started, Download Files, Deposit and Validate, Structural Genomics, Dictionaries & File Formats, Software Tools, Educational Resources, BioSync, General Information, Acknowledgements, Frequently Asked Questions, Known Problems, and Report Bugs/Comments. The main content area features a "Welcome to the RCSB PDB" section with introductory text and a "Molecule of the Month: AAA+ Proteases" section with a 3D protein structure visualization and a brief description of its function.

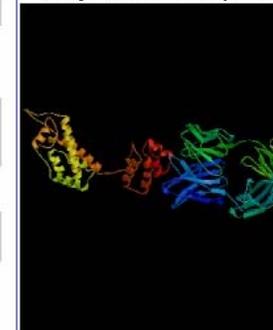
Search by ID (most common)

- 1E6J
- Download Files
- FASTA Sequence
- Display Files
- Display Molecule
- Structural Reports
- Structure Analysis
- Help

1E6J

Images and Visualization

Biological Molecule / Asymmetric



Display Options

- KiNG
- Jmol
- WebMol
- Protein Workshop
- QuickPDB
- All Images

Title CRYSTAL STRUCTURE OF HIV-1 CAPSID PROTEIN (P24) IN COMPLEX WITH FAB13B5

Authors Berthet-Colominas, C., Monaco, S., Novelli, A., Sibai, G., Mallet, F., Cusack, S.

Primary Citation Monaco-Malbet, S., Berthet-Colominas, C., Novelli, A., Battai, N., Piga, N., Cheynet, V., Mallet, F., Cusack, S. Mutual Conformational Adaptations in Antigen and Antibody Upon Complex Formation between an Fab and HIV-1 Capsid Protein P24' *Structure Fold.Des.* v8 pp.1069, 2000

History Deposition 2000-08-18 Release 2000-11-27

Experimental Method Type X-RAY DIFFRACTION Data [EDS]

Parameters	Resolution[Å]	R-Value	R-Free	Space Group
	3.00	0.210 (obs.)	0.284	C 2 (C 1 2 1)

Unit Cell	Length [Å]	a	193.08	b	45.60	c	132.47
	Angles [°]	alpha	90.00	beta	132.41	gamma	90.00

Molecular Description Asymmetric Unit
 Polymer: 1 Molecule: IMMUNOGLOBULIN Fragment: HEAVY CHAIN 1-219 Chains: H Other Details: OBTAINED BY PAPAINE CLEAVAGE (FAB)
 Polymer: 2 Molecule: IMMUNOGLOBULIN Fragment: LIGHT CHAIN 1-210 Chains: L Other Details: OBTAINED BY PAPAINE CLEAVAGE (FAB)
 Polymer: 3 Molecule: CAPSID PROTEIN P24 Fragment: GAG POLYPROTEIN RESIDUES 142-351 Mutation: YES Chains: P Other Details: HIS6 TAG AT N-TERM, PRO1 AND ILE2 DELETED, C-TERMINUS MODIFIED

Functional Class Hiv Capsid Protein (p24)

Source Polymer: 1 Scientific Name: **Mus musculus** Common Name: **House mouse** House mouse Polymer: 2 Scientific Name: **Mus musculus** Common Name: **House mouse** House mouse Polymer: 3 Scientific Name: **Human immunodeficiency virus (clone 12)** Common Name: **Hiv-1** Expression system: **Human immunodeficiency virus type 1 (clone 12)**

SCOP Classification Domain Info Class Fold Superfamily Family Domain Species

Sequence and Details

Home Search **Structure** Queries Structure Summary Biology & Chemistry Materials & Methods **Sequence Details** Geometry

1E6J

Download Files

- PDB File
- PDB gz
- PDB File (Header)
- mmCIF File
- mmCIF gz
- mmCIF File (Header)
- PDBML/XML File
- PDBML/XML gz
- PDBML/XML File (Header)
- Structure Factors File
- Structure Factors gz
- Biological Unit Coordinates
- 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

Sequence Details

Chain H, representative of identical chains Chain H

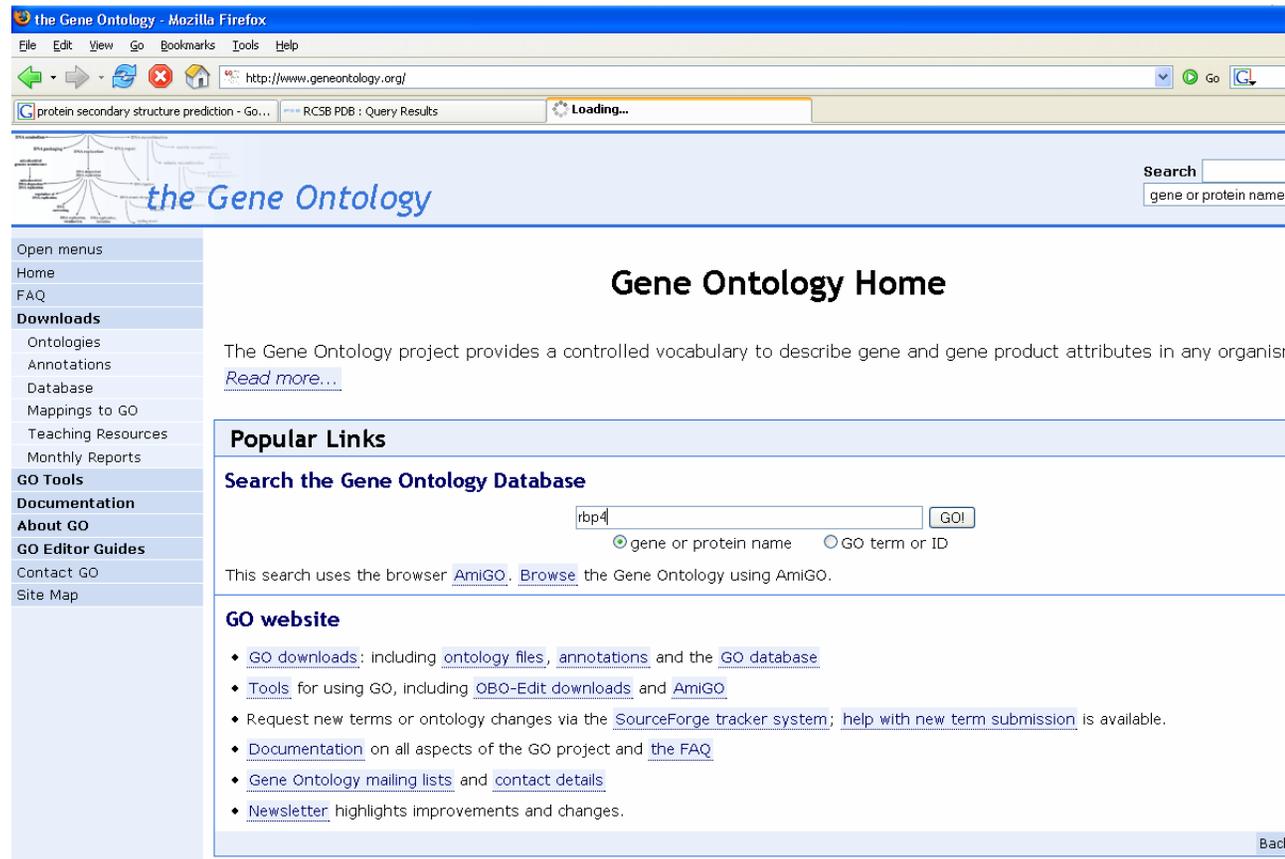
Description	IMMUNOGLOBULIN
Type	polypeptide(L)
Polymer Id	1
Number of residues	219
Domains	d1e6jh1: Immunoglobulin heavy chain variable domain, VH d1e6jh2: Immunoglobulin heavy chain gamma constant domain 1, CH1-gamma

Sequence and Secondary Structure

Key: = extended strand, = turn, = disulfide bond
 = alpha helix, = 310 helix, = pi helix,
Greyed out residues have no structural information

EVOLQOSGAELARPGASVKMSKASGYTFTSYTMHWVKORPGQGLEWIGYINPSSGYSNY
10 20 30 40 50 60
NOKFRDKATLTADKSSSTAYMOLSSLTSEDSAVYYCSRPVVRLGYNFDYWGGSTLTVSS
70 80 90 100 110 120
AKTTPPSVYPLAPGSAQAQTNMVTLGCLVKGYPPEPVTVTWNSGSLSSGVHTFPAVLQSD
130 140 150 160 170 180
LYTLSSSVTPSSWTPSETVTCNVAHPASSTKVDKIVP

5. Gene and Protein Function Database: Gene Ontology (GO)



The screenshot shows the Gene Ontology homepage in a Mozilla Firefox browser. The browser's address bar displays <http://www.geneontology.org/>. The page features a navigation menu on the left with categories like "Open menus", "Downloads", "GO Tools", and "Documentation". The main content area is titled "Gene Ontology Home" and includes a search bar, a "Popular Links" section, and a "GO website" section with various resource links.

the Gene Ontology

Search

Gene Ontology Home

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

Popular Links

Search the Gene Ontology Database

gene or protein name GO term or ID

This search uses the browser [AmiGO](#). [Browse](#) the Gene Ontology using AmiGO.

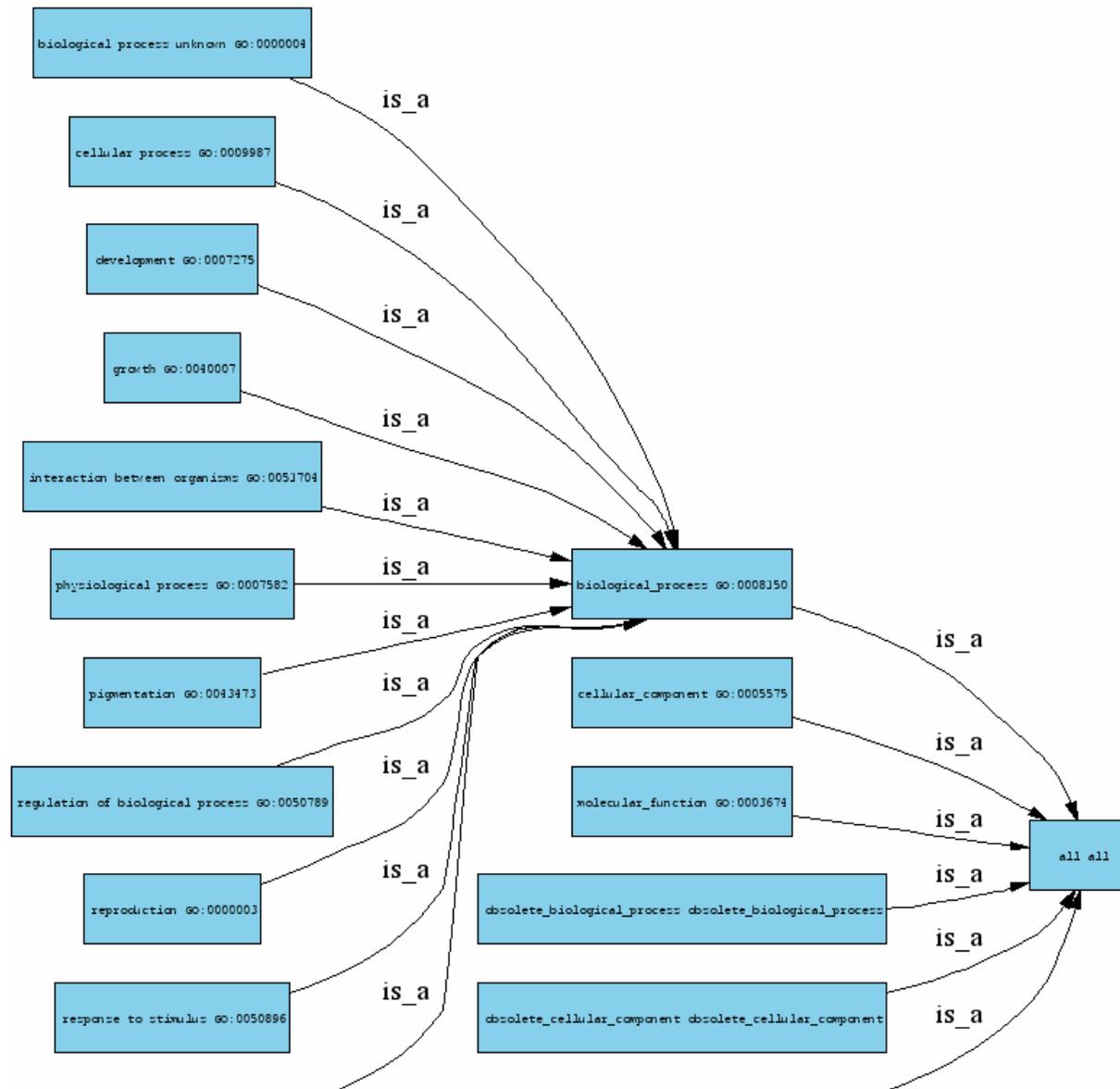
GO website

- [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 [SourceForge tracker system](#); [help with new term submission](#) is available.
- [Documentation](#) on all aspects of the GO project and [the FAQ](#)
- [Gene Ontology mailing lists](#) and [contact details](#)
- [Newsletter](#) highlights improvements and changes.

Bad

www.geneontology.org

Gene Function Hierarchy



6. Ensembl

Portal of European Bioinformatics Institute

Web: www.ensembl.org

Particularly useful for genome / gene search (e.g. human genome web browser)

click

Use Ensembl to...

- Run a BLAST search
- Search Ensembl
- Data mining [BioMart]
- Export data
- Download data

Docs and downloads

- Information
- What's New
- About Ensembl
- Ensembl data
- Software

Other links

- Home
- Sitemap
- Vega
- Pre Ensembl
- View previous release of page in Archive!
- Stable Archive! link for this page
- Archive! sites
- Trace server



Macaca mulatta 1.0 assembly and genebuild

What's New in Ensembl 40

- New low-coverage genomes (*L. africana*, *D. novemcinctus*, *E. telfairi*, *O. cuniculus*)
- Stickleback assembly and genebuild (*Gasterosteus aculeatus*)
- New species - *Aedes aegypti* (*Aedes aegypti*)
- New Macaque assembly and genebuild (*Macaca mulatta*)
- New genebuild on Rat assembly (*Rattus norvegicus*)

More news...

About Ensembl

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

This site provides free access to all the data and software from the Ensembl project. Click on a species name to browse the data.

Access to all the data produced by the project, and to the software used to analyse and present it, is provided free and without constraints. Some data and software may be subject to third-party constraints.

For all enquiries, please contact the Ensembl HelpDesk (helpdesk@ensembl.org).

Other sites using the Ensembl system

- EBI Genome Reviews database - mainly archaea and bacteria.
- VEGA - Vertebrate Genome Annotation

More...

Mammalian genomes

- Homo sapiens** NCBI 36 | Vega
- Pan troglodytes** PanTro 1.0 | **NEW!** pre!
- Macaca mulatta** **UPDATED!** MMUL 1.0
- Mus musculus** NCBI m36 | Vega
- Rattus norvegicus** **UPDATED!** R6SC 3.4
- Oryctolagus cuniculus** **NEW!** RABBIT
- Canis familiaris** CanFam 1.0 | Vega | **UPDATED!** pre!
- Bos taurus** Btau 2.0
- Sus scrofa** **Pre!** **NEW!** (clone status map)
- Dasyus novemcinctus** **NEW!** ARMA
- Loxodonta africana** **NEW!** BROAD E1
- Echinops telfairi** **NEW!** TENREC
- Monodelphis domestica** MonDom 4
- Ornithorhynchus anatinus** **Pre!** **NEW!** DANA 5

Other species

- Gallus gallus** WASHUC 1
- Xenopus tropicalis** JGI 4.1
- Danio rerio** Zv8 | Vega
- Takifugu rubripes** FUGU 4.0
- Tetraodon nigroviridis** TETRAODON 7
- Gasterosteus aculeatus** **NEW!** BROAD S1
- Oryzias latipes** **Pre!** MEDAKA 1
- Ciona intestinalis** JGI2
- Ciona savignyi** CSAV 2.0
- Drosophila melanogaster** **UPDATED!** BDGP 4
- Anopheles gambiae** AgamP3
- Aedes aegypti** **NEW!** AaegL 1
- Caenorhabditis elegans** WS 150
- Saccharomyces cerevisiae** SGD 1

Type in the gene name

Use Ensembl to...

- Run a BLAST search
- Search Ensembl
- Data mining [BioMart]
- Upload and view data on chromosome
- Export data
- Download data

Docs and downloads

- Information
- What's New
- About Ensembl
- Ensembl data
- Software

Select a species

- Mammals
- Other chordates
- Other eukaryotes

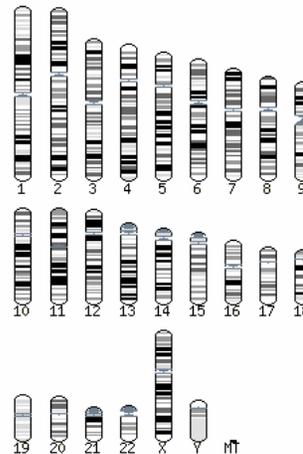
Other links

- Home
- Sitemap
- Vega
- Pre! Pre Ensembl
- e! View previous release of page in Archive!

Explore the *Homo sapiens* genome

Karyotype

Click on a chromosome for a closer view



Jump directly to sequence position

Chromosome: or region

From (bp):

To (bp):

About the Human genome

Assembly



This release is based on the [NCBI 36 assembly](#) of the human genome [November 2005].

The International Human Genome Sequencing Consortium have published their scientific analysis of the finished human genome.

- [Nature 431, 931 - 945 \(21 October 2004\)](#)
- [WT Sanger Institute Press Release](#)

Annotation

The human genome sequence is now considered sufficiently stable that the three major genome browsers have come together to produce a common set of gene IDs for their annotations. This Consensus CDS ID set has been incorporated into the Ensembl database alongside the existing identifiers.

- More information about the [CCDS project](#).

The [ENCODE](#) (ENCyclopedia Of DNA Elements) project aims to find functional elements in the human genome.

- More information about the [ENCODE resources](#) at Ensembl.

Ensembl text search

Your query matched 3 entries in the search database

Ensembl Gene: ENSG00000138207 [\[ContigView\]](#)

Ensembl gene ENSG00000138207 has 5 transcripts: ENST00000260780, ENST00000371463, ENST00000371464, ENST00000371467, ENST00000371469 and associated peptides: ENSP00000260780, ENSP00000360518, ENSP00000360519, ENSP00000360522, ENSP00000360524

Plasma retinol-binding protein precursor (PRBP) (RBP) [Contains: Plasma retinol-binding protein(1-182); Plasma retinol-binding protein(1-181); Plasma retinol-binding protein(1-176)]. [Source:Uniprot/SWISSPROT;Acc:P02753]

The gene has the following external identifiers mapped to it:

Affymx Microarray Focus: 219140_s_at

Affymx Microarray HuGeneFL: X00129_at

Affymx Microarray U133: g8400727_3p_a_at, 219140_s_at

Affymx Microarray U95: 61253_at, 32552_at

Agilent Probe: A_23_P75283, A_32_P56183

EMBL: AF119868, AL356214, X02824, BC020633, X00129, AF025335, AF025334, X02775

Entrez Gene: 5950

GO: GO:0006810, GO:0005615, GO:0016918, GO:0050896, GO:0005576, GO:0007601, GO:0005488, GO:0005501, GO:0019841, GO:0005215

Havana transcripts: OTTHUMT00000049431, OTTHUMT00000049429, OTTHUMT00000049430

HGNC Symbol: **RBP4**, 9922

Illumina: GL_8400727

IPI: IPI00480192.1, IPI00744715, IPI00022420.3, IPI00480192, IPI00744715.1, IPI00022420

MIM disease: 180250

MIM gene: 180250

PDB: 1RBP, 1JYJ, 1JYD, 1QAB, 1BRP, 1BRQ

Protein ID: CAA24959.1, CAB46489, AAC02946, AAF69622.1, CAB46489.1, CAA26553.1, CAH72329, CAH72329.1, AAH20633, AAF69622, AAC02945, CAA24959, AAC02946.1, CAH72328.1, CAA26553, AAC02945.1, AAH20633.1

RefSeq DNA: NM_006744, NM_006744.3

RefSeq peptide: NP_006735, NP_006735.2

UniProtKB/Swiss-Prot: Q9P178, Q5VY24, P02753, Q8WWA3, RETBP_HUMAN, O43478, O43479

UniProtKB/TrEMBL: Q5VY30, Q5VY30_HUMAN

Feature type: Gene; Homo sapiens; **Species:** Homo sapiens; Gene;

Ensembl Gene: ENSG00000144231 [\[ContigView\]](#)

Ensembl gene ENSG00000144231 has 1 transcript: ENST00000272645 and associated peptide: ENSP00000272645

DNA-directed RNA polymerase II 16 kDa polypeptide (EC 2.7.7.6) (RPB4). [Source:Uniprot/SWISSPROT;Acc:O15514]

The gene has the following external identifiers mapped to it:

Affymx Microarray Focus: 203664_s_at

Affymx Microarray U133: 214144_at, Hs.194638.1.A1_3p_at, 203664_s_at, g4758573_3p_s_at

Affymx Microarray U95: 33726_at, 56881_at, 63651_at

Agilent Probe: A_32_P192506, A_24_P687085, A_23_P209778, A_32_P118512, A_32_P219753

click

Ensembl Gene Report for ENSG00000138207

Gene	RBP4 (HGNC Symbol ID) . To view all Ensembl genes linked to the name click here .					
Ensembl Gene ID	ENSG00000138207					
Genomic Location	This gene can be found on Chromosome 10 at location 95,341,434-95,351,491 . The start of this gene is located in Contig AL356214.20.1.163964 .					
Description	Plasma retinol-binding protein precursor (PRBP) (RBP) [Contains: Plasma retinol-binding protein(1-182); Plasma retinol-binding protein(1-181); Plasma retinol-binding protein(1-179); Plasma retinol-binding protein(1-176)]. Source: Uniprot/SwissProt P02753					
Prediction Method	gene containing both ensembl predicted transcripts and havana manual annotation					
Transcripts	ENST00000260780	ENSP00000260780	RBP4	[Transcript info]	[Exon info]	[Peptide info]
	ENST00000371463	ENSP00000360518	novel transcript	[Transcript info]	[Exon info]	[Peptide info]
	ENST00000371464	ENSP00000360519	RETBP_HUMAN	[Transcript info]	[Exon info]	[Peptide info]
	ENST00000371467	ENSP00000360522	RETBP_HUMAN	[Transcript info]	[Exon info]	[Peptide info]
	ENST00000371469	ENSP00000360524	Q5VY30_HUMAN	[Transcript info]	[Exon info]	[Peptide info]

Features ▾

Chr. 10
Length
Forward strand
95.33 Mb 95.34 Mb 95.34 Mb 95.35 Mb 95.36 Mb 95.36 Mb
30.06 Kb

Ensembl trans.
Q5NUL3-2 >
Ensembl Known Protein Coding
GPR120 >
Ensembl Known Protein Coding
Q5NUL3-2 >
Havana Known Protein coding
GPR120 >
Havana Known Protein coding

DNA(contigs)
AL356214.20.1.163964 >

Ensembl trans.
< Q5VY30_HUMAN
Havana Known Protein coding
< RETBP_HUMAN
Havana Known Protein coding
< RETBP_HUMAN
Havana Known Protein coding

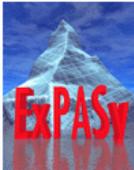
7. ExPASy

ExPASy = Expert Protein Analysis System

Web: <http://www.expasy.ch/>

Main portal of SwissProt

Provide links to many proteomics tools



ExPASy Proteomics Server

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](#)).

[\[Announcements\]](#) [\[Job opening\]](#) [\[Mirror Sites\]](#)

Databases

- [UniProt Knowledgebase \(Swiss-Prot and TrEMBL\)](#) - Protein knowledgebase
- [PROSITE](#) - Protein families and domains
- [SWISS-2DPAGE](#) - Two-dimensional polyacrylamide gel electrophoresis
- [ENZYME](#) - Enzyme nomenclature
- [SWISS-MODEL Repository](#) - Automatically generated protein models

- [Links to many other molecular biology databases](#)

Tools and software packages

- [Proteomics and sequence analysis tools](#)
 - ◊ [Proteomics](#)
 - ◊ [DNA -> Protein](#)
 - ◊ [Similarity searches \(BLAST...\)](#)
 - ◊ [Pattern and profile searches \(ScanProsite...\)](#)
 - ◊ [Post-translational modification and topology prediction](#)
 - ◊ [Primary structure analysis](#)
 - ◊ [Secondary and tertiary structure tools \(Swiss PdbViewer...\)](#)
 - ◊ [Alignment and Phylogenetic analysis](#)
- [ImageMaster / Melanie](#) - Software for 2-D PAGE analysis
- [MSight](#) - Mass Spectrometry Imager
- [Roche Applied Science's Biochemical Pathways](#)

Education and services

- [The ExPASy FTP server](#)
- [Swiss-Shop](#) - automatically obtain (by email) new sequence entries relevant to your field(s) of interest
- [Vital-IT](#) - The HPC Center for Life Sciences
- [e-Proxemis](#) - Bioinformatics Learning Portal for Proteomics
- [Master's degree in Proteomics and Bioinformatics](#)
- [Proteomics courses](#) - two courses covering Separation Sciences & Mass Spectrometry for Proteomics
- [Proteomics Core Facility](#) (previously SWISS-2DSERVICE) - get your 2-D Gels performed according to Swiss standards

Documentation

- [What's New on ExPASy](#)
- [SWISS-FLASH](#) electronic bulletins
- [Swiss-Prot documents](#)
- [How to create HTML links to ExPASy](#)
- [Complete table of available documents](#)

SwissProt

Tools

Example of how to access sequence data:

HIV-1 *pol*

There are many possible approaches. Begin at the main page of NCBI, and type an Entrez query: hiv-1 pol

Search across databases [Help](#)

1444		PubMed: biomedical literature citations and abstracts	?	52		Books: online books	?
1315		PubMed Central: free, full text journal articles	?	4		OMIM: Online Mendelian Inheritance in Man	?
				7		Site Search: NCBI web and FTP sites	?
39110		Nucleotide: sequence database (GenBank)	?	none		UniGene: gene-oriented clusters of transcript sequences	?
37936		Protein: sequence database	?	none		CDD: conserved protein domain database	?
3		Genome: whole genome sequences	?	640		3D Domains: domains from Entrez Structure	?
93		Structure: three-dimensional macromolecular structures	?	none		UniSTS: markers and mapping data	?
none		Taxonomy: organisms in GenBank	?	560		PopSet: population study data sets	?
934		SNP: single nucleotide polymorphism	?	none		GEO Profiles: expression and molecular abundance profiles	?
98		Gene: gene-centered information	?	none		GEO DataSets: experimental sets of GEO data	?
29		HomoloGene: Eukaryotic homology groups	?	113		Cancer Chromosomes: cytogenetic databases	?
none		Journals: detailed information about the journals indexed in PubMed and other Entrez databases	?	3		MeSH: detailed information about NLM's controlled vocabulary	?

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

NCBI Entrez Genome

All Databases PubMed Nucleotide Protein Genome

Search Genome for hiv-1 pol [Go] [Clear] [Save Search]

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Send to

All: 3

Items 1 - 3 of 3

- 1: [NC_001802](#)
Human immunodeficiency virus 1, complete genome
ssRNA, linear; Length: **9,181 nt**
Created: **1998/01/22**
- 2: [NC_003075](#)
Arabidopsis thaliana chromosome 4, complete sequence
DNA, linear; Length: **18,585,042 nt**
Replicon Type: **chromosome**
Replicon Name: **4**
Created: **2001/08/13**
- 3: [NC_001664](#)
Human herpesvirus 6, complete genome
dsDNA, linear; Length: **159,321 nt**
Created: **1995/06/05**

About Entrez

Entrez Genome Help

Submitting Genome Project Genome sequence

Microbial Genome Projects PDB neighbors

Genomic BLAST Microbial Eukaryotic

Archaea Chromosome Plasmid DraftAssembly

Bacteria Chromosome Plasmid DraftAssembly

Eukaryota Chromosome Plasmid Organelles

Manageable three entries

NCBI

All Databases PubMed Nucleotide Protein Genome Structure

Search Nucleotide for hiv-1 Go Clear Save Search

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Send to

All: 170016 bacteria: 3 mRNA: 4267 RefSeq: 1529

Show only records from: CoreNucleotide (168052), EST (296), GSS (1668). [What's this?]

Items 1 - 20 of 170016

1: [DQ866318](#) Reports
HIV-1 clone 17 from South Africa envelope glycoprotein (env) gene, partial cds
gi|112819924|gb|DQ866318.1|[112819924]

2: [DQ866317](#) Reports
HIV-1 clone 6 from South Africa envelope glycoprotein (env) gene, partial cds
gi|112819922|gb|DQ866317.1|[112819922]

3: [DQ866316](#) Reports
HIV-1 clone 1 from South Africa envelope glycoprotein (env) gene, partial cds
gi|112819920|gb|DQ866316.1|[112819920]

4: [DQ866417](#) Reports
HIV-1 isolate 1174 from South Africa envelope glycoprotein (env) gene, partial cds
gi|112819918|gb|DQ866417.1|[112819918]

5: [DQ866416](#) Reports
HIV-1 isolate 1173 from South Africa envelope glycoprotein (env) gene, partial cds
gi|112819916|gb|DQ866416.1|[112819916]

6: [DQ866415](#) Reports
HIV-1 isolate 1172 from South Africa envelope glycoprotein (env) gene, partial cds
gi|112819914|gb|DQ866415.1|[112819914]

7: [DQ866414](#) Reports
HIV-1 isolate 1169 from South Africa envelope glycoprotein (env) gene, partial cds

Search all databases, un-manageable (170016)

NCBI's important RefSeq project: best representative sequences

RefSeq (accessible via the main page of NCBI) provides an expertly curated accession number that corresponds to the most stable, agreed-upon “reference” version of a sequence.

RefSeq identifiers include the following formats:

Complete genome	NC_#####
Complete chromosome	NC_#####
Genomic contig	NT_#####
mRNA (DNA format)	NM_##### e.g. NM_006744
Protein	NP_##### e.g. NP_006735

8. Literature Search

PubMed is the NCBI gateway to MEDLINE.

MEDLINE contains bibliographic citations and author abstracts from over 4,600 journals published in the United States and in 70 foreign countries.

It has >14 million records dating back to 1966.

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC

Search PubMed for retinol-binding protein fish [Save Search](#)

Display Summary Show 20 Sort by Send to

All: 51 Review: 5

Items 1 - 20 of 51

- About Entrez
- Text Version
- Entrez PubMed
 - Overview
 - Help | FAQ
 - Tutorials
 - New/Noteworthy
 - E-Utilities
- PubMed Services
 - Journals Database
 - MeSH Database
 - Single Citation Matcher
 - Batch Citation Matcher
 - Clinical Queries
 - Special Queries
 - LinkOut
 - My NCBI
- Related Resources
 - Order Documents
 - NLM Mobile
 - NLM Catalog
 - NLM Gateway
 - TOXNET
 - Consumer Health
 - Clinical Alerts
 - ClinicalTrials.gov
 - PubMed Central

- 1: [Tingaud-Sequeira A, Forgue J, Andre M, Babin PJ.](#)
 Epidermal transient down-regulation of retinol-binding protein 4 and mirror expression of apolipoprotein Eb and estrogen receptor 2a during zeb development.
 Dev Dyn. 2006 Aug 23; [Epub ahead of print]
 PMID: 16929533 [PubMed - as supplied by publisher]
- 2: [Devirgiliis C, Gaetani S, Apreda M, Bellovino D.](#)
 Glycosylation is essential for translocation of carp retinol-binding protein across the endoplasmic reticulum membrane.
 Biochem Biophys Res Commun. 2005 Jul 1;332(2):504-11.
 PMID: 15910754 [PubMed - indexed for MEDLINE]
- 3: [Sumanas S, Jorniak T, Lin S.](#)
 Identification of novel vascular endothelial-specific genes by the microarray analysis of the zebrafish cloche mutants.
 Blood. 2005 Jul 15;106(2):534-41. Epub 2005 Mar 31.
 PMID: 15802528 [PubMed - indexed for MEDLINE]
- 4: [Sammar M, Levi L, Hurvitz A, Lubzens E.](#)
 Studies on retinol-binding protein during vitellogenesis in the Rainbow Trout (*Oncorhynchus mykiss*).
 Gen Comp Endocrinol. 2005 Apr;141(2):141-51.
 PMID: 15748715 [PubMed - indexed for MEDLINE]
- 5: [Liu RZ, Sun Q, Thisse C, Thisse B, Wright JM, Denovan-Wright EM.](#)
 The cellular retinol-binding protein genes are duplicated and differentially transcribed in the developing and adult zebrafish (*Danio rerio*).
 Mol Biol Evol. 2005 Mar;22(3):469-77. Epub 2004 Oct 27. Erratum in: Mol Biol Evol. 2005 Apr;22(4):1157.
 PMID: 15509725 [PubMed - indexed for MEDLINE]
- 6: [von Schalburg KR, Rise ML, Brown GD, Davidson WS, Koop BF.](#)
 A comprehensive survey of the genes involved in maturation and development of the rainbow trout ovary.

All Databases PubMed Nucleotide Protein Genome Structure OM

Search PubMed for retinol-binding protein fish [Go] [Clear]

Limits Preview/Index History Clipboard Details

- About Entrez
- Text Version
- Entrez PubMed
 - Overview
 - Help | FAQ
 - Tutorials
 - New/Noteworthy
 - E-Utilities
- PubMed Services
 - Journals Database
 - MeSH Database
 - Single Citation Matcher
 - Batch Citation Matcher
 - Clinical Queries
 - Special Queries
 - LinkOut
 - My NCBI
- Related Resources
 - Order Documents
 - NLM Mobile
 - NLM Catalog
 - NLM Gateway
 - TOXNET
 - Consumer Health
 - Clinical Alerts
 - ClinicalTrials.gov
 - PubMed Central

Limit your search by any of the following criteria.

Search by Author [Add Author] [CLEAR]

Search by Journal [Add Journal] [CLEAR]

Full Text, Free Full Text, and Abstracts [CLEAR]

Links to full text Links to free full text Abstracts

Dates [CLEAR]

Published in the Last: Any date [v]

Added to PubMed in the Last: Any date [v]

Humans or Animals [CLEAR]

Humans Animals

Gender [CLEAR]

Male Female

Languages [CLEAR]

English
 French
 German
 Italian
 Japanese
 Russian
 Spanish
[More Languages](#)

Subsets [CLEAR]

Journal Groups

Core clinical journals
 Dental journals
 Nursing journals

Topics

AIDS
 Bioethics
 Cancer

Ultimate Literature Search (my most favorite way)



Web

Proteomics Tools

Secondary structure prediction. AGADIR - An algorithm to predict the helical content of peptides; APSSP - Advanced **Protein Secondary Structure Prediction** ...
www.expasy.org/tools/ - 46k - [Cached](#) - [Similar pages](#)

nnpredict input form

NNPREDICT **Protein Secondary Structure Prediction.** Enter a **protein** sequence and nnpredict will predict the **secondary structure**. Click here for instructions ...
www.cmpharm.ucsf.edu/~nomi/nnpredict.html - 3k - [Cached](#) - [Similar pages](#)

PredictProtein - Structure Prediction and Sequence Analysis

service for **protein structure prediction**, **protein** sequence analysis, **protein** function **prediction**, **protein** sequence alignments, bioinformatics.
www.predictprotein.org/ - 14k - [Cached](#) - [Similar pages](#)

BCM Search Launcher: Protein Secondary Structure Prediction

PSSP / NNSSP - nearest-neighbor **prediction** (Sanger) [H] [O] [P] [E]; SAPS - statistical analysis of **protein** sequences (ISREC) [H] [O] [P] [E] ...
searchlauncher.bcm.tmc.edu/seq-search/struc-predict.html - 16k - [Cached](#) - [Similar pages](#)

Protein Structure Prediction

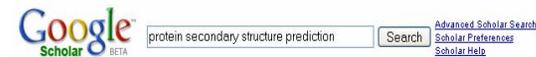
Secondary Structure Prediction; Solvent accessibility **prediction** & Folds ... BCM Search Launcher **Protein Secondary Structure Prediction**; COILS **Prediction** of ...
cmgm.stanford.edu/WWW/www_predict.html - 15k - [Cached](#) - [Similar pages](#)

Jpred - A secondary structure prediction server

Secondary protein structure prediction from amino acid sequence; based on a consensus from several methods including DSC, PHD, NNSSP, PREDATOR, ZPRED, ...
www.compbio.dundee.ac.uk/~www-jpred/ - 4k - [Cached](#) - [Similar pages](#)

Secondary Structure Prediction methods and links

Prediction of Protein Secondary Structure and Active Sites Using the Alignment of Homologous ... Barton, GJ (1995), **Protein Secondary Structure Prediction**, ...
www.sbg.bio.ic.ac.uk/people/rob/CCP11BBS/secstrucpred.html - 18k - [Cached](#) - [Similar pages](#)



Scholar

Results 1 - 10 of about 87,300 for **protein secondary**.

Protein secondary structure prediction based on position-specific scoring matrices - Full Text via SFX @ UCF - group of 19 »

DT Jones - J Mol Biol, 1999 - bioinfo.cs.umn.edu
Page 1. COMMUNICATION **Protein Secondary Structure Prediction** Based on Position-specific Scoring Matrices ... 196 **Protein Secondary Structure Prediction** Page ...
[Cited by 849](#) - [Related Articles](#) - [Web Search](#) - [BL Direct](#)

Prediction of protein secondary structure at better than 70% accuracy - Full Text via SFX @ UCF - group of 6 »

B Rost, C Sander - J Mol Biol, 1993 -.ncbi.nlm.nih.gov
Click here to read **Prediction of protein secondary structure** at better than 70% accuracy. Rost B, Sander C. European Molecular Biology ...
[Cited by 1466](#) - [Related Articles](#) - [Cached](#) - [Web Search](#) - [BL Direct](#)

PHD--an automatic mail server for **protein secondary structure prediction** - group of 3 »

B Rost, C Sander, R Schneider - Bioinformatics - bioinformatics.oupjournals.org
... ARTICLES. PHD--an automatic mail server for **protein secondary structure prediction**. B Rost, C Sander and R Schneider. By the middle ...
[Cited by 369](#) - [Related Articles](#) - [Web Search](#) - [More Info via SFX @ UCF](#) - [BL Direct](#)

Improvements in **protein secondary structure prediction** by an enhanced neural network - Full Text via SFX @ UCF

DG Kneller, FE Cohen, R Langridge - J Mol Biol, 1990 -.ncbi.nlm.nih.gov
Improvements in **protein secondary structure prediction** by an enhanced neural network. Kneller DG, Cohen FE, Langridge R. Department ...
[Cited by 302](#) - [Related Articles](#) - [Cached](#) - [Web Search](#)

... of the concepts important for accurate and reliable **protein secondary structure prediction** - Full Text via SFX @ UCF - group of 7 »

RD KING - Protein Science, 1996 - proteinscience.org
... ARTICLE. Identification and application of the concepts important for accurate and reliable **protein secondary structure prediction**. RD KING and MJE. ...
[Cited by 196](#) - [Related Articles](#) - [Web Search](#) - [BL Direct](#)

SOPM: a self-optimized method for **protein secondary structure prediction** - group of 3 »

C Geourjon, G Deleage - Protein Eng, 1994 - peds.oupjournals.org
... by Oxford University Press. ARTICLES. SOPM: a self-optimized method for **protein secondary structure prediction**. C Geourjon and G Deleage ...
[Cited by 182](#) - [Related Articles](#) - [Web Search](#) - [More Info via SFX @ UCF](#) - [BL Direct](#)

Ten Topics

- 1. Introduction to Molecular Biology and Bioinformatics
- 2. Pairwise Sequence Alignment Using Dynamic Programming
- 3. Practical Sequence/Profile Alignment Using Fast Heuristic Methods (BLAST and PSI-BLAST)
- 4. Multiple Sequence Alignment
- 5. Gene Identification
- 6. Phylogenetic Analysis
- 7. Protein Structure Analysis and Prediction
- 8. RNA Secondary Structure Prediction
- 9. Clustering and Classification of Gene Expression Data
- **10. Search and Mining of Biological Databases, Databanks, and Literature**

Congratulations!