
UniProt
Universal Protein Resource
prosite
Pfam
PRINTS
Protein Fingerprints Database
ProDom
SMART
TIGR
THE INSTITUTE FOR GENOMIC RESEARCH
tigr fams
PANTHER
Classification System
PIRSF
Superfamily
Gene3D
Domain Architecture Classification
MSD
CATH
Protein Structure Classification
SCOP
SWISS-MODEL
MODBASE

National Center for Biotechnology Information (NCBI)

www.ncbi.nlm.nih.gov

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

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

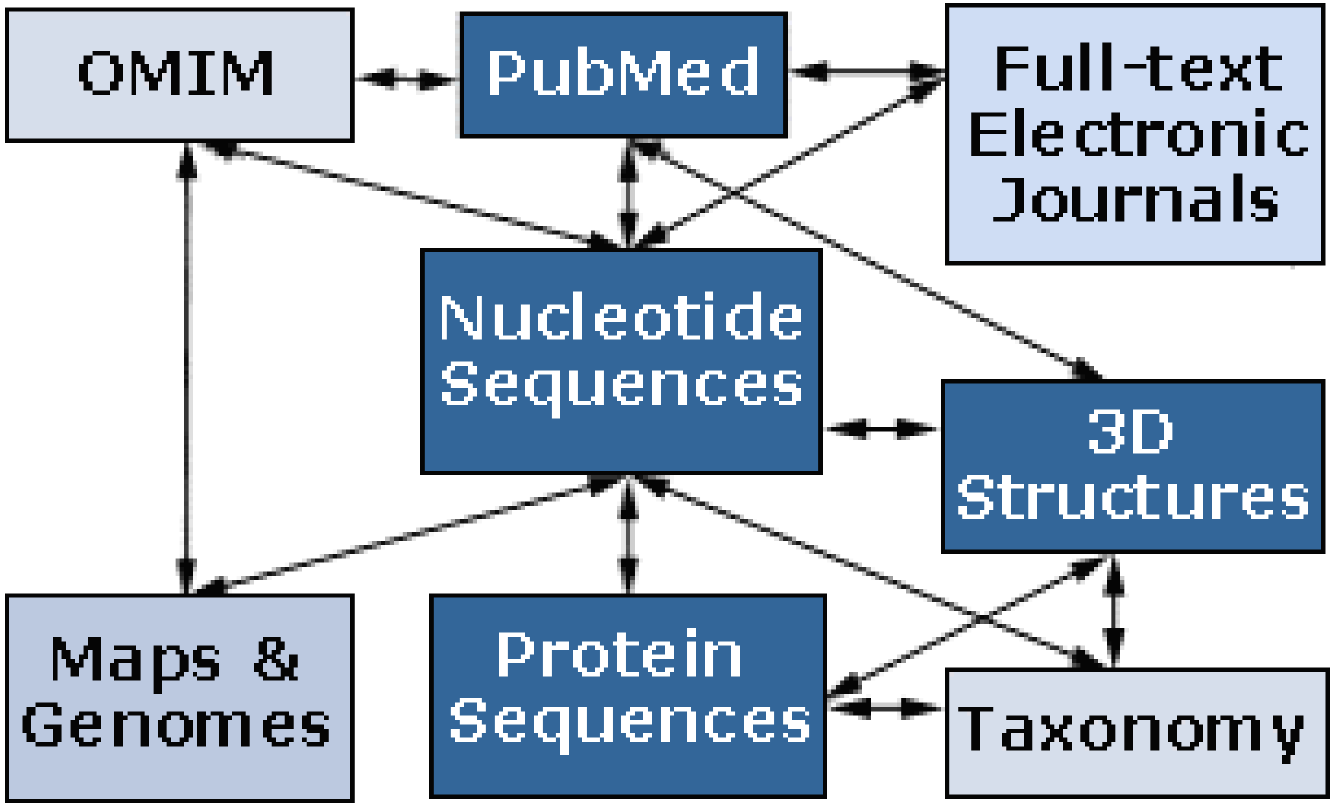
NCBI's dbGaP Genome Wide Association Database

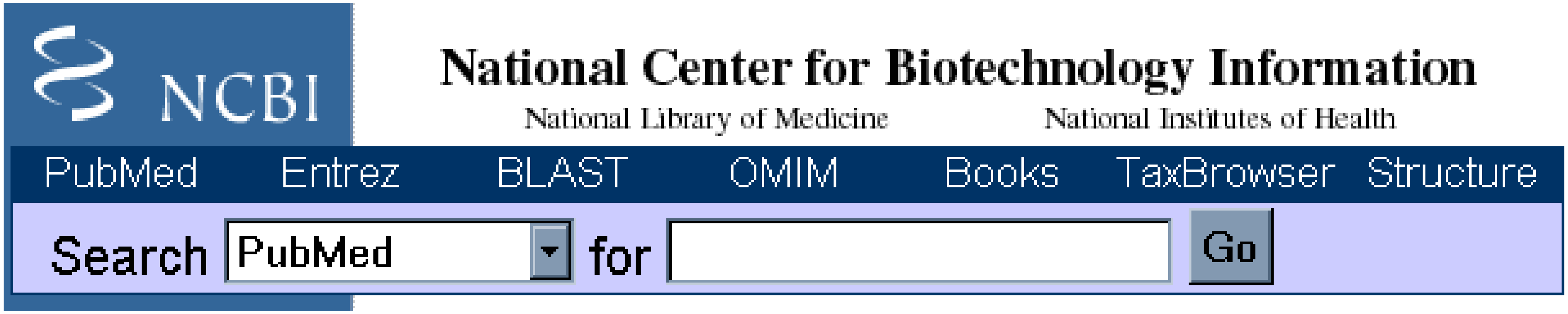
NCBI's dbGaP (database of Genotype and Phenotype) provides data from Genome Wide Association (GWA) studies. The resource is intended to help elucidate the link between genes and disease. For each study, users have access to detailed information about the phenotypic variables measured and pre-computed 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

Entrez is a search and retrieval system that integrates NCBI databases





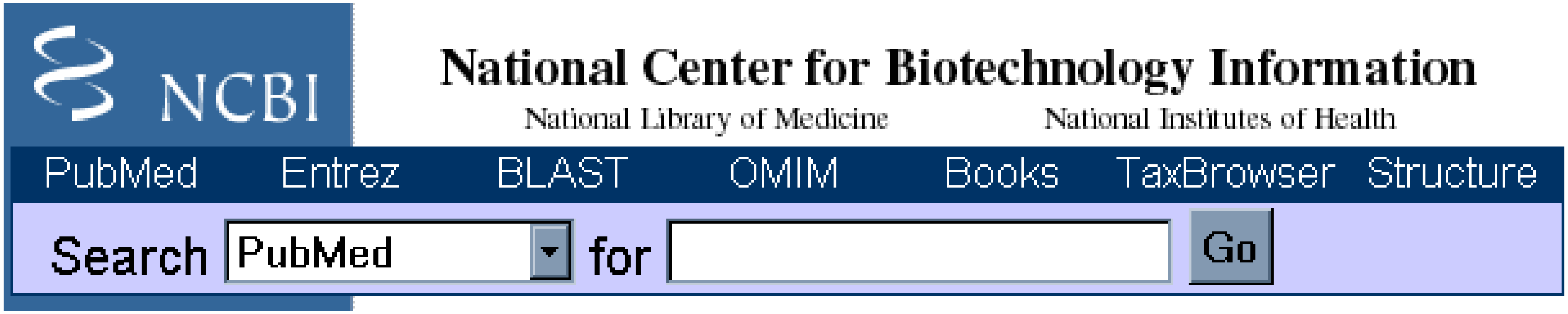
BLAST is…

Basic Local Alignment Search Tool

NCBI's sequence similarity search tool

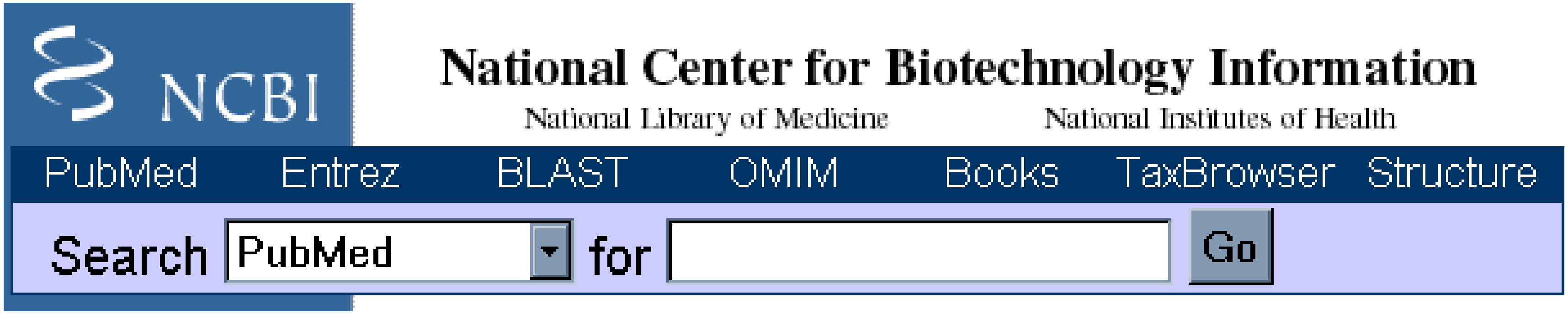
supports analysis of DNA and protein databases

80,000 searches per day



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

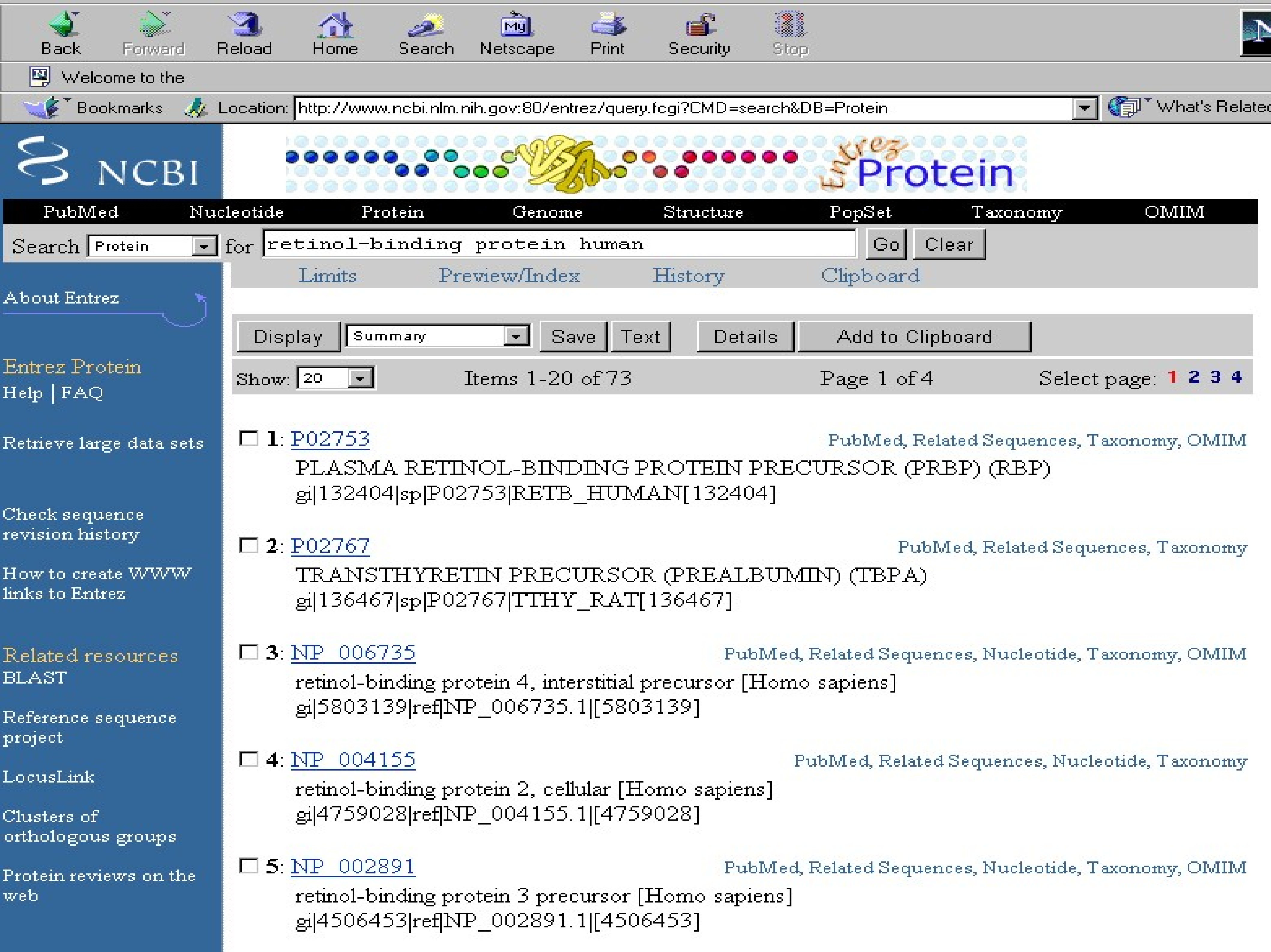
National Library of Medicine National Institutes of Health

What is an accession number?

An accession number is label that used to identify a sequence. It is a string of letters and/or numbers that corresponds to a molecular sequence.

Examples (all for retinol-binding protein, RBP4):

X02775	GenBank genomic DNA sequence	DNA
NT_030059	Genomic contig	
Rs7079946	dbSNP (single nucleotide polymorphism)	
N91759.1	An expressed sequence tag (1 of 170)	RNA
NM_006744	RefSeq DNA sequence (from a transcript)	
NP_007635	RefSeq protein	protein
AAC02945	GenBank protein	
Q28369	SwissProt protein	
1KT7	Protein Data Bank structure record	



Location: http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?CMD=search&DB=Protein

Search Protein for retinol-binding protein human

Display Summary

Show: 20 Items 1-20 of 73 Page 1 of 4 Select page: 1 2 3 4

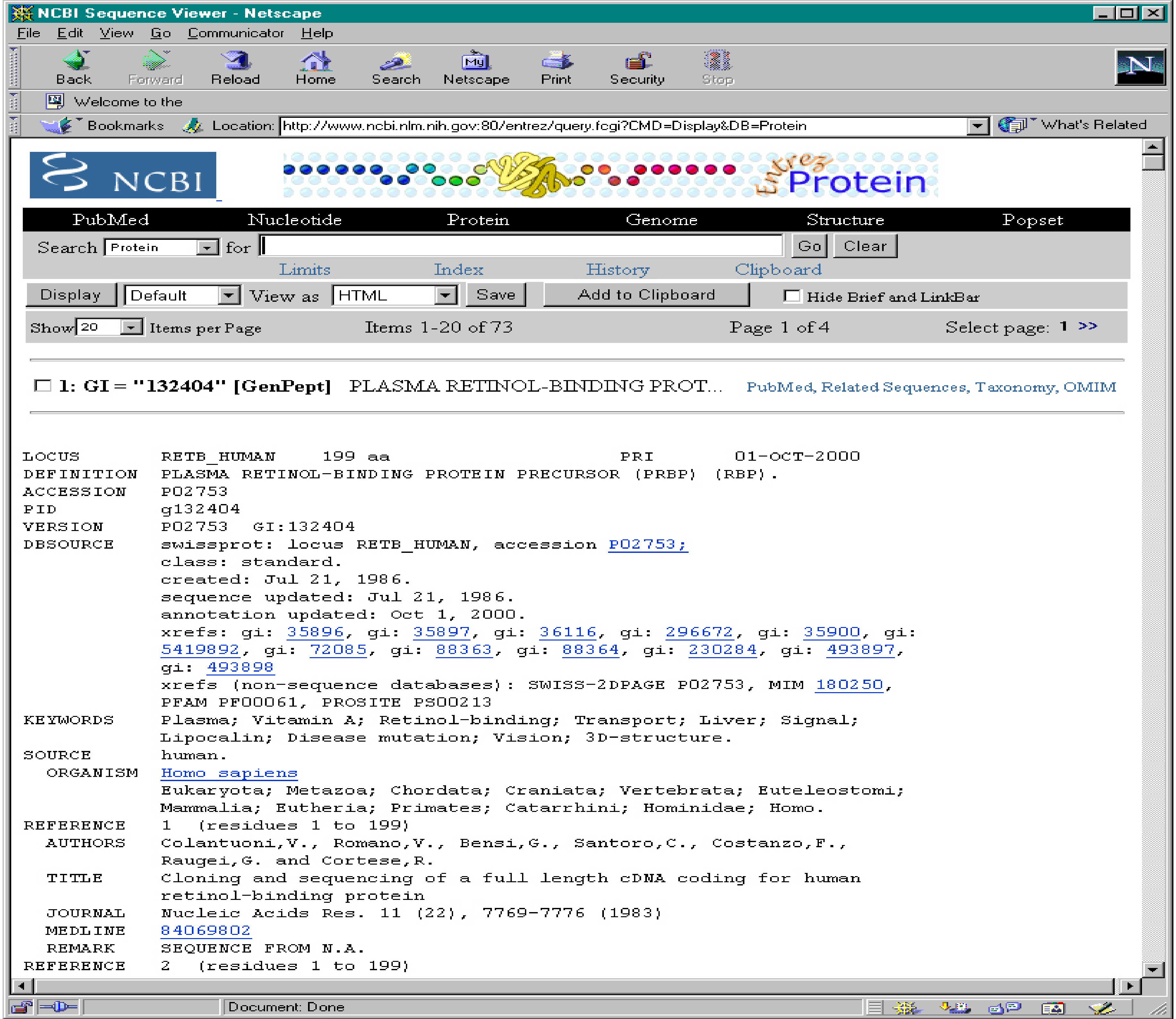
☐ 1: P02753 PubMed, Related Sequences, Taxonomy, OMIM
PLASMA RETINOL-BINDING PROTEIN PRECURSOR (PRBP) (RBP)
gi|132404|sp|P02753|RETB_HUMAN[132404]

☐ 2: P02767 PubMed, Related Sequences, Taxonomy
TRANSTHYRETIN PRECURSOR (PREALBUMIN) (TBPA)
gi|136467|sp|P02767|TTHY_RAT[136467]

☐ 3: NP_006735 PubMed, Related Sequences, Nucleotide, Taxonomy, OMIM
retinol-binding protein 4, interstitial precursor [Homo sapiens]
gi|5803139|ref|NP_006735.1|[5803139]

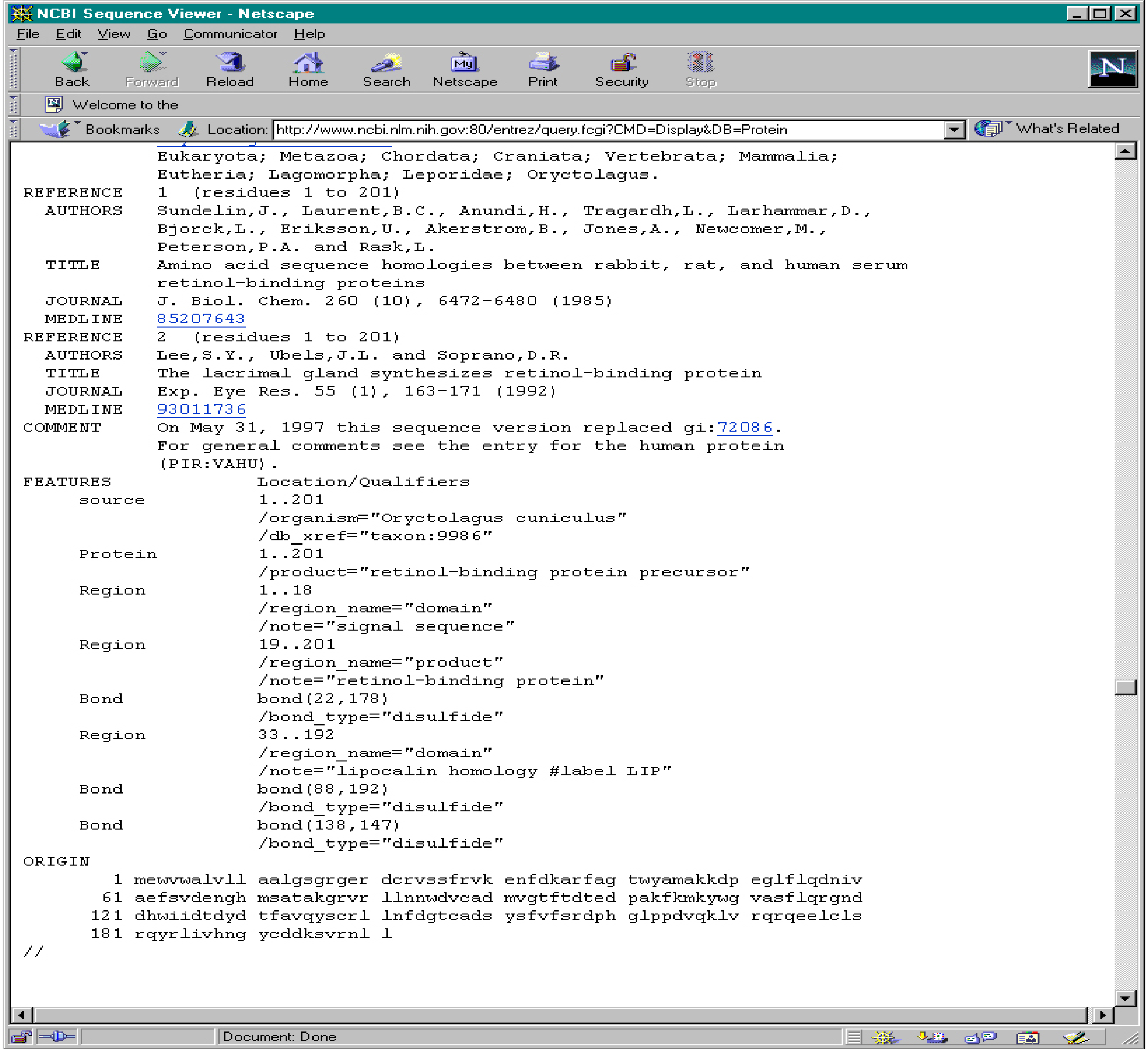
☐ 4: NP_004155 PubMed, Related Sequences, Nucleotide, Taxonomy
retinol-binding protein 2, cellular [Homo sapiens]
gi|4759028|ref|NP_004155.1|[4759028]

☐ 5: NP_002891 PubMed, Related Sequences, Nucleotide, Taxonomy, OMIM
retinol-binding protein 3 precursor [Homo sapiens]
gi|4506453|ref|NP_002891.1|[4506453]



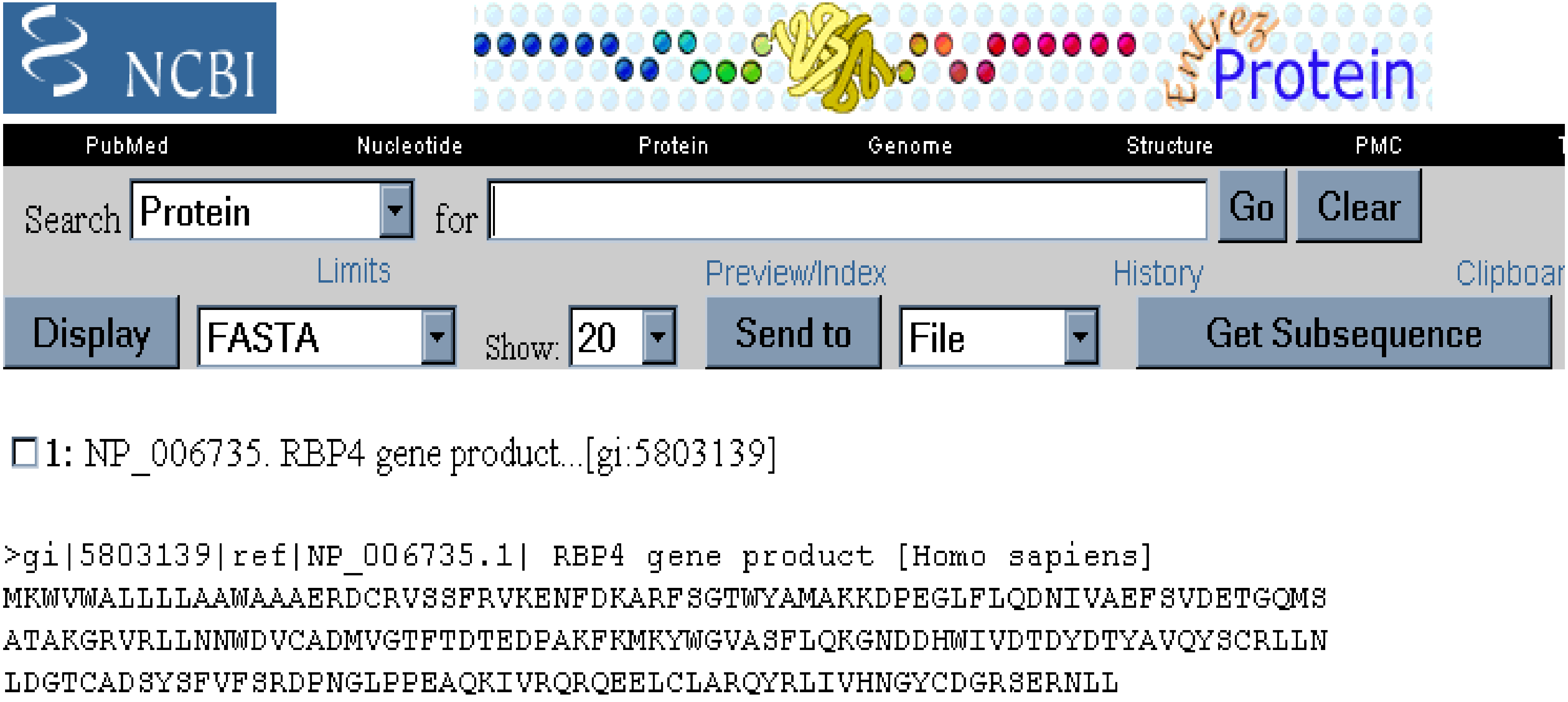
☐ **1: GI = "132404" [GenPept]** PLASMA RETINOL-BINDING PROT... PubMed, Related Sequences, Taxonomy, OMIM

```
LOCUS       RETB_HUMAN     199 aa            PRI       01-OCT-2000
DEFINITION  PLASMA RETINOL-BINDING PROTEIN PRECURSOR (PRBP) (RBP).
ACCESSION   P02753
PID         g132404
VERSION     P02753  GI:132404
DBSOURCE    swissprot: locus RETB_HUMAN, accession P02753;
            class: standard.
            created: Jul 21, 1986.
            sequence updated: Jul 21, 1986.
            annotation updated: Oct 1, 2000.
            xrefs: gi: 35896, gi: 35897, gi: 36116, gi: 296672, gi: 35900, gi:
            5419892, gi: 72085, gi: 88363, gi: 88364, gi: 230284, gi: 493897,
            gi: 493898
            xrefs (non-sequence databases): SWISS-2DPAGE P02753, MIM 180250,
            PFAM PF00061, PROSITE PS00213
KEYWORDS    Plasma; Vitamin A; Retinol-binding; Transport; Liver; Signal;
            Lipocalin; Disease mutation; Vision; 3D-structure.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 199)
  AUTHORS   Colantuoni,V., Romano,V., Bensi,G., Santoro,C., Costanzo,F.,
            Raugei,G. and Cortese,R.
  TITLE     Cloning and sequencing of a full length cDNA coding for human
            retinol-binding protein
  JOURNAL   Nucleic Acids Res. 11 (22), 7769-7776 (1983)
  MEDLINE   84069802
  REMARK    SEQUENCE FROM N.A.
REFERENCE   2  (residues 1 to 199)
```



```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE   1  (residues 1 to 201)
  AUTHORS   Sundelin,J., Laurent,B.C., Anundi,H., Tragardh,L., Larhammar,D.,
            Bjorck,L., Eriksson,U., Akerstrom,B., Jones,A., Newcomer,M.,
            Peterson,P.A. and Rask,L.
  TITLE     Amino acid sequence homologies between rabbit, rat, and human serum
            retinol-binding proteins
  JOURNAL   J. Biol. Chem. 260 (10), 6472-6480 (1985)
  MEDLINE   85207643
REFERENCE   2  (residues 1 to 201)
  AUTHORS   Lee,S.Y., Ubels,J.L. and Soprano,D.R.
  TITLE     The lacrimal gland synthesizes retinol-binding protein
  JOURNAL   Exp. Eye Res. 55 (1), 163-171 (1992)
  MEDLINE   93011736
COMMENT     On May 31, 1997 this sequence version replaced gi:72086.
            For general comments see the entry for the human protein
            (PIR:VAHU).
FEATURES             Location/Qualifiers
     source          1..201
                     /organism="Oryctolagus cuniculus"
                     /db_xref="taxon:9986"
     Protein         1..201
                     /product="retinol-binding protein precursor"
     Region          1..18
                     /region_name="domain"
                     /note="signal sequence"
     Region          19..201
                     /region_name="product"
                     /note="retinol-binding protein"
     Bond            bond(22,178)
                     /bond_type="disulfide"
     Region          33..192
                     /region_name="domain"
                     /note="lipocalin homology #label LIP"
     Bond            bond(88,192)
                     /bond_type="disulfide"
     Bond            bond(138,147)
                     /bond_type="disulfide"
ORIGIN
        1 mewvwalvll aalgsgrger dcrvssfrvk enfdkarfag twyamakkdp eglflqdniv
       61 aefsvdengh msatakgrvr llnnwdvcad mvgtftdted pakfkmkywg vasflqrgnd
      121 dhwiidtdyd tfavqyscrl lnfdgtcads ysfvfsrdph glppdvqklv rqrqeelcls
      181 rqyrlivhng ycddksvrnl l
//
```

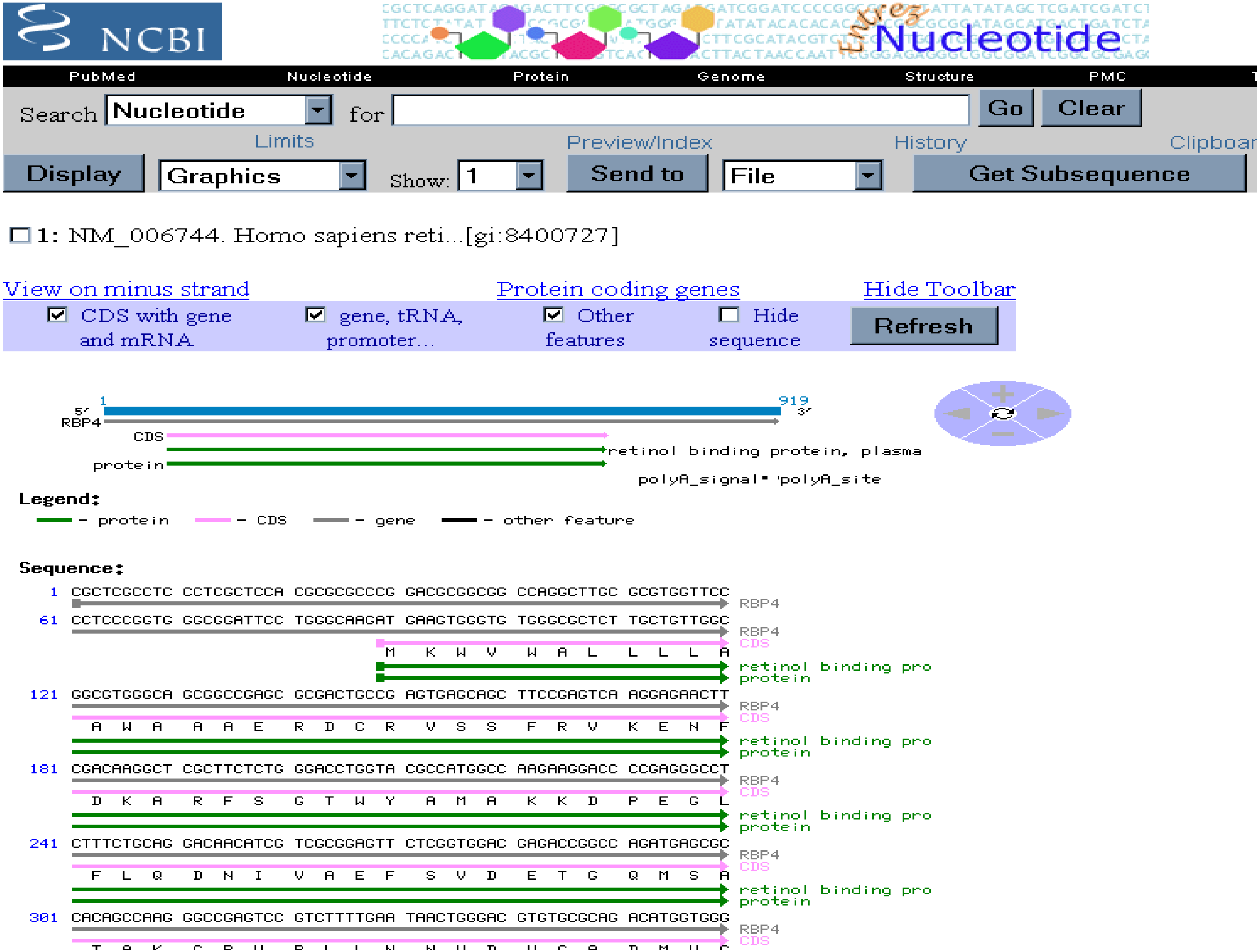
FASTA format



☐ 1: NP_006735. RBP4 gene product...[gi:5803139]

```
>gi|5803139|ref|NP_006735.1| RBP4 gene product [Homo sapiens]
MKWVWALLLLAAWAAAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEFSVDETGQMS
ATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIVDTDYDTYAVQYSCRLLN
LDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNGYCDGRSERNLL
```

Graphical format



Search Nucleotide for

Display Graphics Show: 1 Send to File

☐ **1:** NM_006744. Homo sapiens reti...[gi:8400727]

☑ CDS with gene and mRNA ☑ gene, tRNA, promoter... ☑ Other features ☐ Hide sequence

Legend:

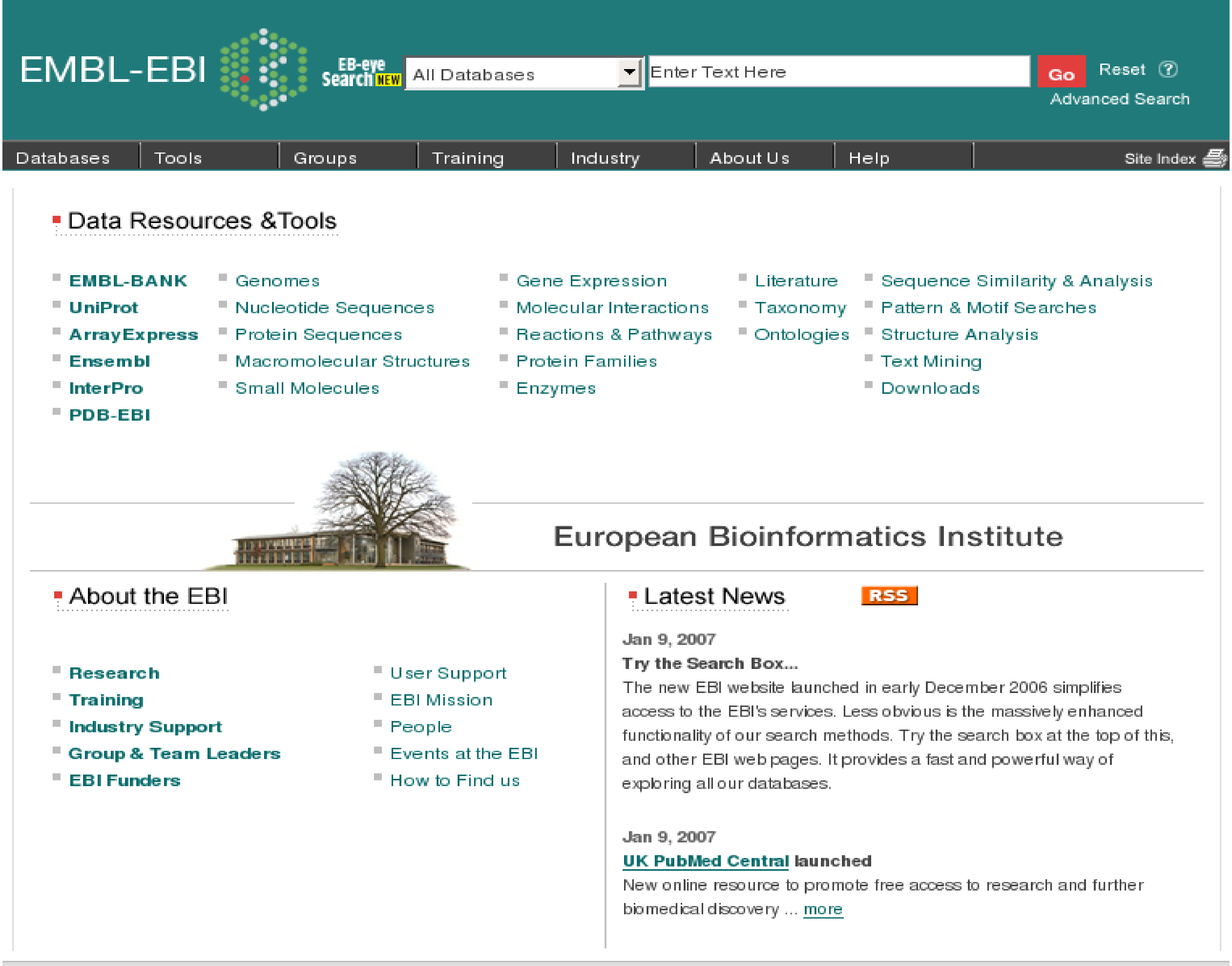
- protein
- CDS
- gene
- other feature

Sequence:

```
  1 CGCTCGCCTC CCTCGCTCCA CGCGCGCCCG GACGCGGCGG CCAGGCTTGC GCGTGGTTCC
                                                                      RBP4
 61 CCTCCCGGTG GGCGGATTCC TGGGCAAGAT GAAGTGGGTG TGGGCGCTCT TGCTGTTGGC
                                                                      RBP4
                                                                      CDS
                                     M   K  W  V   W  A  L   L  L  L  A
                                                                      retinol binding pro
                                                                      protein
121 GGCGTGGGCA GCGGCCGAGC GCGACTGCCG AGTGAGCAGC TTCCGAGTCA AGGAGAACTT
                                                                      RBP4
                                                                      CDS
      A  W  A   A  A  E   R  D  C  R   V  S  S   F  R  V   K  E  N  F
                                                                      retinol binding pro
                                                                      protein
181 CGACAAGGCT CGCTTCTCTG GGACCTGGTA CGCCATGGCC AAGAAGGACC CCGAGGGCCT
                                                                      RBP4
                                                                      CDS
      D  K  A   R  F  S   G  T  W  Y   A  M  A   K  K  D   P  E  G  L
                                                                      retinol binding pro
                                                                      protein
241 CTTTCTGCAG GACAACATCG TCGCGGAGTT CTCGGTGGAC GAGACCGGCC AGATGAGCGC
                                                                      RBP4
                                                                      CDS
      F  L  Q   D  N  I   V  A  E  F   S  V  D   E  T  G   Q  M  S  A
                                                                      retinol binding pro
                                                                      protein
301 CACAGCCAAG GGCCGAGTCC GTCTTTTGAA TAACTGGGAC GTGTGCGCAG ACATGGTGGG
                                                                      RBP4
                                                                      CDS
```

European Molecular Biology Laboratory
European Bioinformatics Institute
(EMBL-EBI)

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



Data Resources &Tools

- **EMBL-BANK**
- **UniProt**
- **ArrayExpress**
- **Ensembl**
- **InterPro**
- **PDB-EBI**

- Genomes
- Nucleotide Sequences
- Protein Sequences
- Macromolecular Structures
- Small Molecules

- Gene Expression
- Molecular Interactions
- Reactions & Pathways
- Protein Families
- Enzymes

- Literature
- Taxonomy
- Ontologies

- Sequence Similarity & Analysis
- Pattern & Motif Searches
- Structure Analysis
- Text Mining
- Downloads

European Bioinformatics Institute

About the EBI

- **Research**
- **Training**
- **Industry Support**
- **Group & Team Leaders**
- **EBI Funders**

- User Support
- EBI Mission
- People
- Events at the EBI
- How to Find us

Latest News

Jan 9, 2007

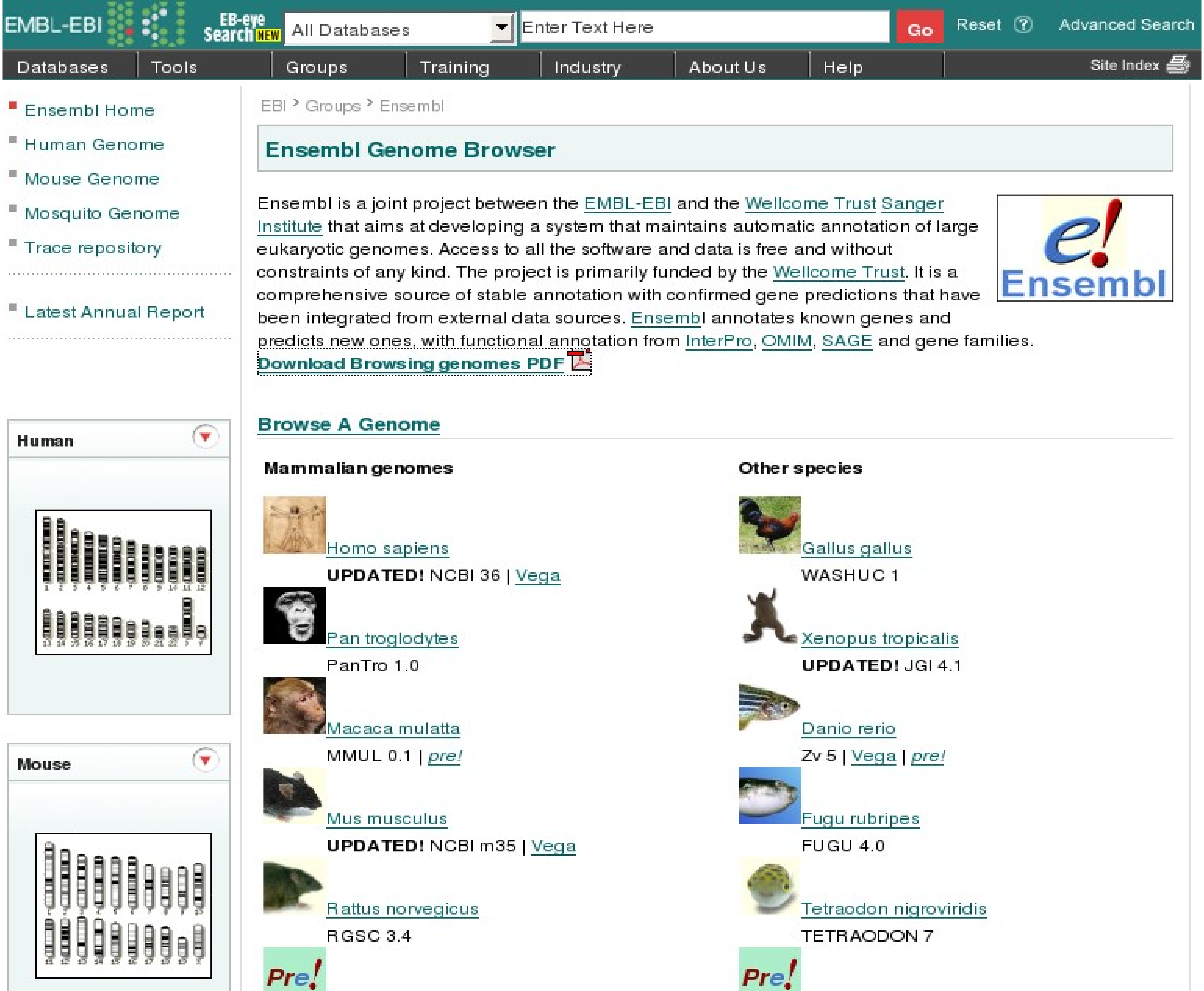
Try the Search Box...

The new EBI website launched in early December 2006 simplifies access to the EBI's services. Less obvious is the massively enhanced functionality of our search methods. Try the search box at the top of this, and other EBI web pages. It provides a fast and powerful way of exploring all our databases.

Jan 9, 2007

UK PubMed Central launched

New online resource to promote free access to research and further biomedical discovery ... more



Ensembl Genome Browser

Ensembl is a joint project between the EMBL-EBI and the Wellcome Trust Sanger Institute that aims at developing a system that maintains automatic annotation of large eukaryotic genomes. Access to all the software and data is free and without constraints of any kind. The project is primarily funded by the Wellcome Trust. It is a comprehensive source of stable annotation with confirmed gene predictions that have been integrated from external data sources. Ensembl annotates known genes and predicts new ones, with functional annotation from InterPro, OMIM, SAGE and gene families.

Download Browsing genomes PDF

Browse A Genome

Mammalian genomes

Homo sapiens
UPDATED! NCBI 36 | Vega

Pan troglodytes
PanTro 1.0

Macaca mulatta
MMUL 0.1 | *pre!*

Mus musculus
UPDATED! NCBI m35 | Vega

Rattus norvegicus
RGSC 3.4

Other species

Gallus gallus
WASHUC 1

Xenopus tropicalis
UPDATED! JGI 4.1

Danio rerio
Zv 5 | Vega | *pre!*

Fugu rubripes
FUGU 4.0

Tetraodon nigroviridis
TETRAODON 7

Uniprot

http://www.ebi.uniprot.org

Welcome to UniProt

UniProt (Universal Protein Resource) is the world's most comprehensive catalog of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR.

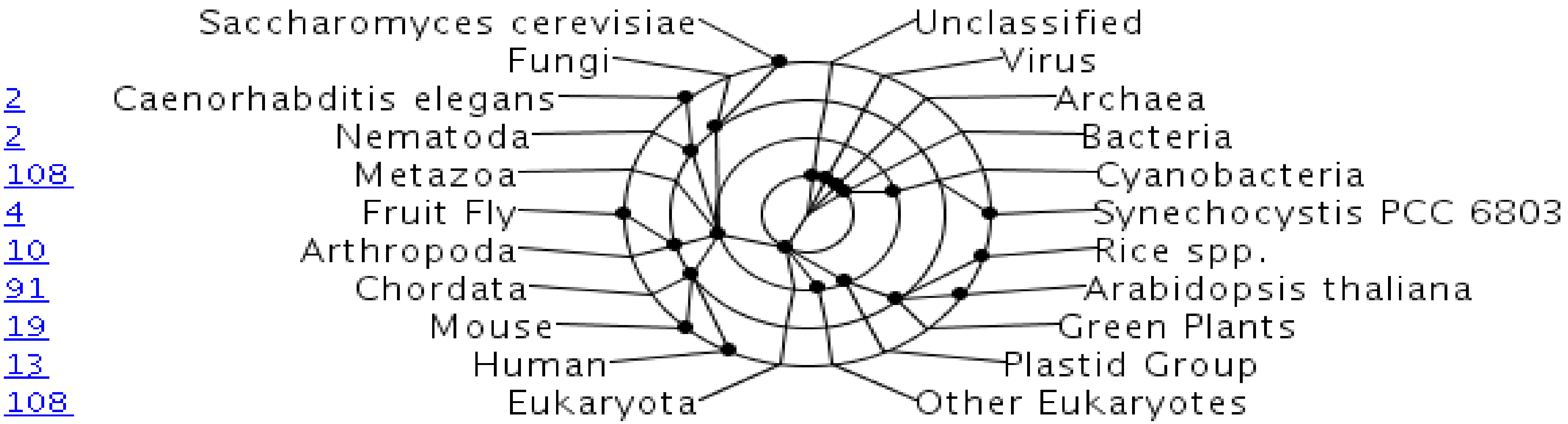
UniProt has three components, each optimized for different uses. The **UniProt Knowledgebase (UniProtKB)** is the central access point for extensive curated protein information, including function, classification, and cross-reference. The **UniProt Reference Clusters (UniRef)** databases combine closely related sequences into a single record to speed searches. The **UniProt Archive (UniParc)** is a comprehensive repository, reflecting the history of all protein sequences.

The sequences and information in UniProt are accessible via **text search**, **BLAST similarity search**, and **FTP**.

Interpro

http://www.ebi.ac.uk/interpro

Taxonomic coverage

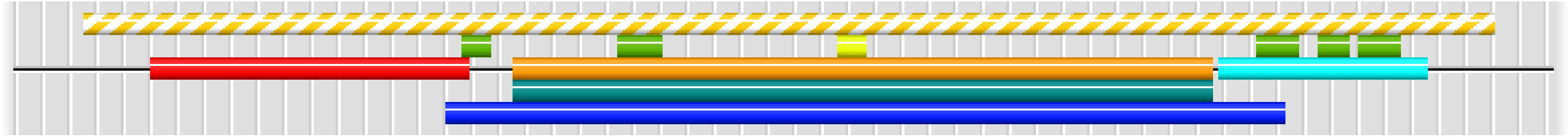


Overlapping InterPro entries

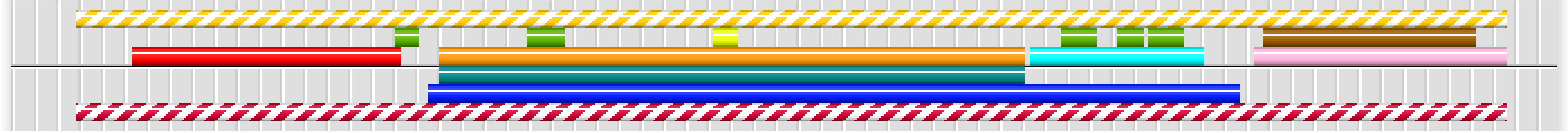
IPR000239	Numbers of overlapping proteins			Average numbers of overlapping amino acids
IPR011009 % Overlap: 99	1	107	36685	N/A
IPR000719 % Overlap: 97	3	105	30592	N/A
IPR000342 % Overlap: 94	6	102	576	N/A
IPR000961 % Overlap: 93	7	101	1315	N/A
IPR002290 % Overlap: 90	10	98	9194	N/A

Example proteins

O70291 G protein-coupled receptor kinase 4 (EC 2.7.11.16) (G protein-coupled receptor kinase GRK4)

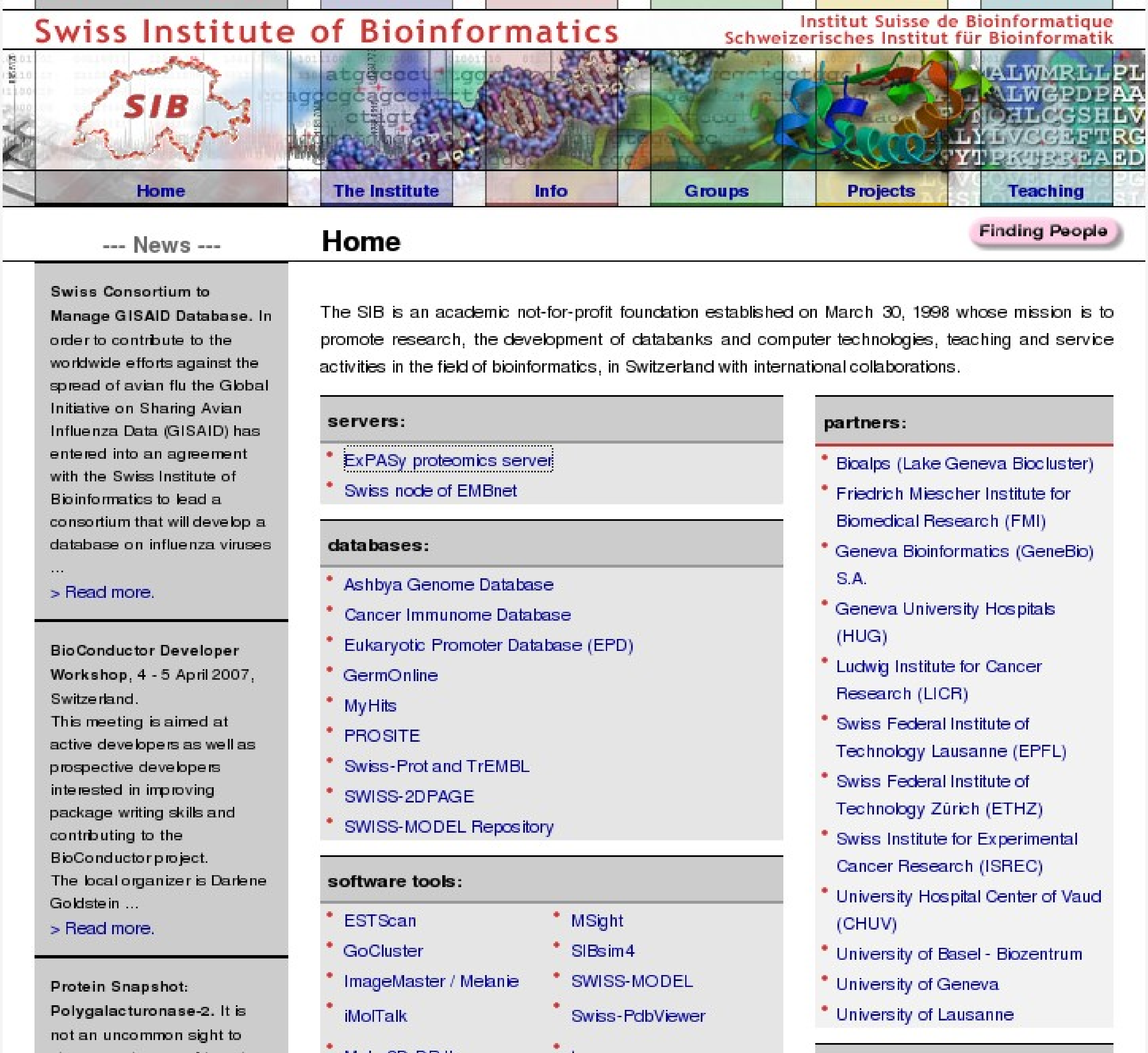


O97627 Beta-adrenergic receptor kinase 1 (EC 2.7.11.15) (Beta-ARK-1) (G- protein coupled receptor kinase 2)



Swiss Institute of Bioinformatics (SIB)

http://www.isb-sib.ch



Swiss Institute of Bioinformatics

Institut Suisse de Bioinformatique
Schweizerisches Institut für Bioinformatik

--- News ---

Swiss Consortium to Manage GISAID Database. In order to contribute to the worldwide efforts against the spread of avian flu the Global Initiative on Sharing Avian Influenza Data (GISAID) has entered into an agreement with the Swiss Institute of Bioinformatics to lead a consortium that will develop a database on influenza viruses ...
> Read more.

BioConductor Developer Workshop, 4 - 5 April 2007, Switzerland.
This meeting is aimed at active developers as well as prospective developers interested in improving package writing skills and contributing to the BioConductor project.
The local organizer is Darlene Goldstein ...
> Read more.

Protein Snapshot: Polygalacturonase-2. It is not an uncommon sight to

Home

The SIB is an academic not-for-profit foundation established on March 30, 1998 whose mission is to promote research, the development of databanks and computer technologies, teaching and service activities in the field of bioinformatics, in Switzerland with international collaborations.

servers:

- ExPASy proteomics server
- Swiss node of EMBnet

databases:

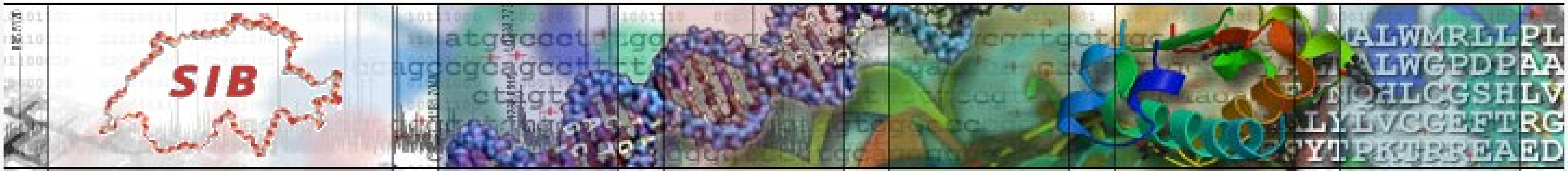
- Ashbya Genome Database
- Cancer Immunome Database
- Eukaryotic Promoter Database (EPD)
- GermOnline
- MyHits
- PROSITE
- Swiss-Prot and TrEMBL
- SWISS-2DPAGE
- SWISS-MODEL Repository

software tools:

- ESTScan
- GoCluster
- ImageMaster / Melanie
- iMolTalk
- MSight
- SIBsim4
- SWISS-MODEL
- Swiss-PdbViewer

partners:

- Bioalps (Lake Geneva Biocluster)
- Friedrich Miescher Institute for Biomedical Research (FMI)
- Geneva Bioinformatics (GeneBio) S.A.
- Geneva University Hospitals (HUG)
- Ludwig Institute for Cancer Research (LICR)
- Swiss Federal Institute of Technology Lausanne (EPFL)
- Swiss Federal Institute of Technology Zürich (ETHZ)
- Swiss Institute for Experimental Cancer Research (ISREC)
- University Hospital Center of Vaud (CHUV)
- University of Basel - Biozentrum
- University of Geneva
- University of Lausanne



Global Info

Job Openings - Positions Available

Finding People

> Addresses
> Careers
> Conferences
> Finding People
> Maps
> News
> Publications

PhD / PostDoc Position in Protein Structure Modelling

Description

The Protein Structure Bioinformatics Group at the Biozentrum, University of Basel & Swiss Institute of Bioinformatics is looking for a highly-motivated PhD student or PostDoc in protein structure modelling and prediction. Successful candidates will have a degree in computational biology, biophysics, structural biology, or computational sciences. Good programming skills and motivation to adopt modern software engineering principles using C++/Python are essential. A strong background in protein structure modelling, prediction or experimental structure determination is highly desirable.

The research project will focus on the development of automated computational methods for the SWISS-MODEL expert system for comparative protein structure modelling. The project aims at the development, validation and implementation of novel computational approaches in all aspects of template-based protein structure modelling, including multiple template modelling, modelling of loop regions, model refinement, and model quality assessment. Collaborations with experimental groups on practical applications of comparative modelling in protein engineering or structure based drug design are encouraged.

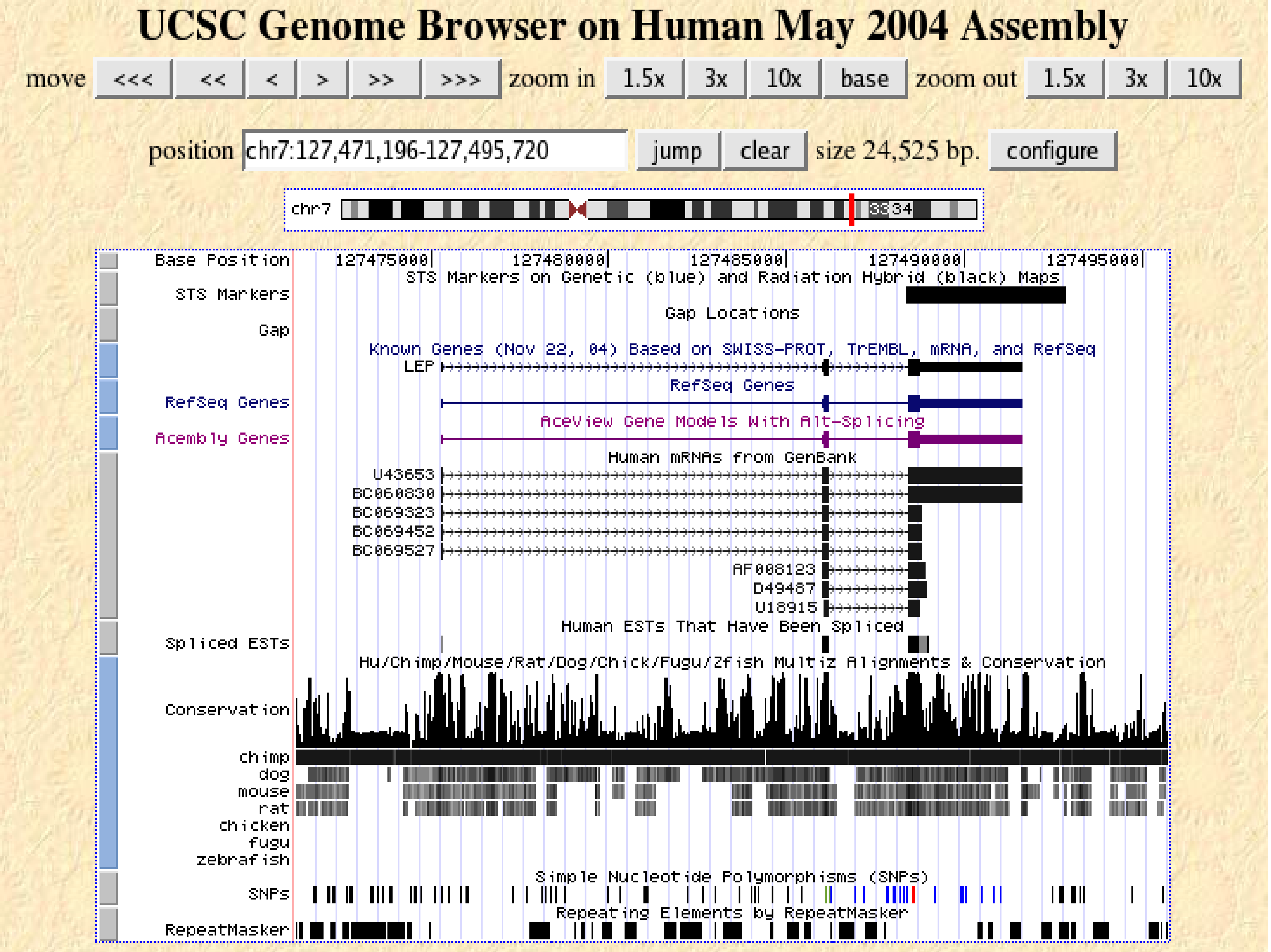
For more information, please see:
Biozentrum Basel http://www.biozentrum.unibas.ch/schwede/
SWISS-MODEL http://swissmodel.expasy.org

Please send your application letter, CV and contacts of three references **before 31. February 2007** to:

Prof. Torsten Schwede
Swiss Institute of Bioinformatics
Biozentrum University of Basel
Klingelbergstrasse 50/70
CH-4056 Basel / Switzerland

UCSC Genome Browser (GB)

http://genome.ucsc.edu

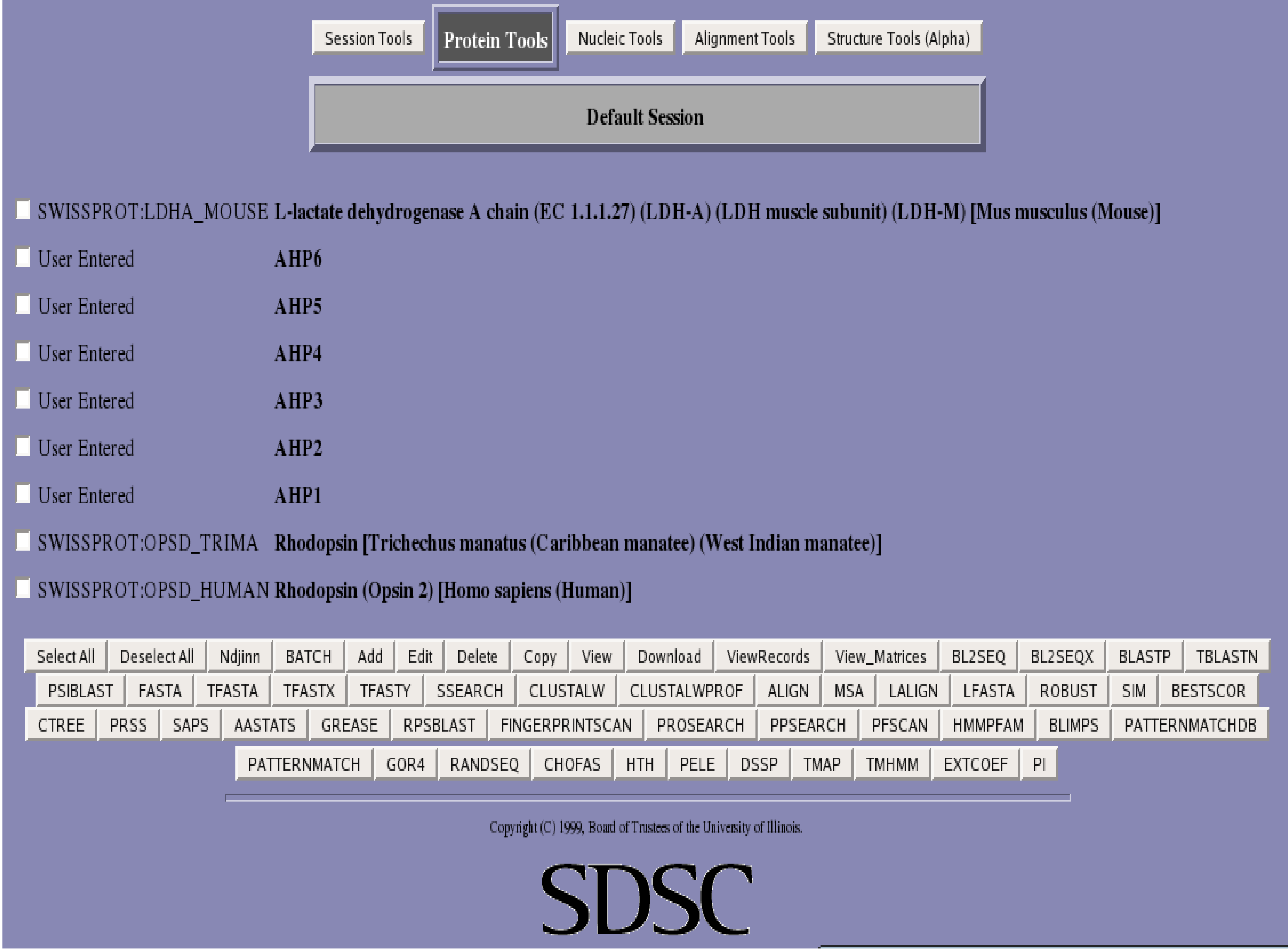


UCSC Genome Browser on Human May 2004 Assembly

position chr7:127,471,196-127,495,720 size 24,525 bp.

Biology Workbench

http://workbench.sdsc.edu



Default Session

☐	SWISSPROT:LDHA_MOUSE	**L-lactate dehydrogenase A chain (EC 1.1.1.27) (LDH-A) (LDH muscle subunit) (LDH-M) [Mus musculus (Mouse)]**
☐	User Entered	**AHP6**
☐	User Entered	**AHP5**
☐	User Entered	**AHP4**
☐	User Entered	**AHP3**
☐	User Entered	**AHP2**
☐	User Entered	**AHP1**
☐	SWISSPROT:OPSD_TRIMA	**Rhodopsin [Trichechus manatus (Caribbean manatee) (West Indian manatee)]**
☐	SWISSPROT:OPSD_HUMAN	**Rhodopsin (Opsin 2) [Homo sapiens (Human)]**

SDSC

Protein Data Bank
(PDB)

http://www.pdb.org

Sequenced genomes	ENSEMBL	~200
DNA sequences	NCBI	30 000 000
PROTEIN sequences	UniProtKB	4 134 341
PROTEIN structures	PDB	42212

Praktické cvičení

- Vyhledejte DNA i proteinovou sekvenci lidského rhodopsinu
  - NCBI Entrez http://www.ncbi.nlm.nih.gov/
- Vyhledejte 3-D strukturu rhodopsinu s organizmu nejbližšího člověku
  - http://pdbbeta.rcsb.org/pdb/
- Srovnejte proteinové sekvence obou proteinů
  - http://clustalw.genome.jp/

Rhodopsin

- NT_005612 (contig, 35742636-35749341)
- NC_000003 (chr03, 130730180-130736885, map 3q21-q24)
- NM_000539 (mRNA)
- NP_000530 (protein RHO, RP4, OPN2)

- PDB bovine RHO: 1F88 1GZM 1HZX 1JFP 1L9H 1U19 1LN6

- 23 mismatches in sequence alignment