Genome Informatics

Systems Biology and the Omics Cascade (Course 2143)
Day 3, June 11th, 2008

Kiyoko F. Aoki-Kinoshita

Introduction

- Genome informatics covers the computer-based modeling and data processing of genome-based data.
- This includes databases and resources for genomic analysis.
- You were introduced to KEGG on Day 2.
- Some other useful databases and resources will be covered today.

But first! Data formats

- It is usually not enough to simply look at the data provided by databases
- To actually use the data for analysis, one often needs to save the retrieved data
- This requires knowledge about the data formats used by each database
- So we will cover the major data formats used in bioinformatics

Data Formats

- Major data formats:
  - GenBank
  - EMBL/UniProt
  - FASTA
  - PDB
- Formats suited for programming:
  - ASN.1 (Abstract Syntax Notation One)
  - XML (eXtensible Markup Language)

GenBank format

- Each line starts with a keyword in capital letters.
- Each keyword is followed by a tab and the information corresponding to it.
- Some keywords are hierarchical:

```
LOCUS       SCU49845     5028 bp    DNA             PLN       21-JUN-1999
DEFINITION  Saccharomyces cerevisiae TCP1-beta gene, partial cds, and Axl2p
            (AXL2) and Rev7p (REV7) genes, complete cds.
ACCESSION   U49845
VERSION     U49845.1  GI:1293613
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 5028)
  AUTHORS   Torpey,L.E., Gibbs,P.E., Nelson,J. and Lawrence,C.W.
  TITLE     Cloning and sequence of REV7, a gene whose function is required for
            DNA damage-induced mutagenesis in Saccharomyces cerevisiae
  JOURNAL   Yeast 10 (11), 1503-1509 (1994)
  PUBMED    7871890
REFERENCE   2  (bases 1 to 5028)
  AUTHORS   Roemer,T., Madden,K., Chang,J. and Snyder,M.
  TITLE     Selection of axial growth sites in yeast requires Axl2p, a novel
            plasma membrane glycoprotein
  JOURNAL   Genes Dev. 10 (7), 777-793 (1996)
  PUBMED    8846915
REFERENCE   3  (bases 1 to 5028)
  AUTHORS   Roemer,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Terry Roemer, Biology, Yale University, New
            Haven, CT, USA
FEATURES             Location/Qualifiers
     source          1..5028
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="IX"
                     /map="9"
     CDS             <1..206
                     /codon_start=3
                     /product="TCP1-beta"
                     /protein_id="AAA98665.1"
                     /db_xref="GI:1293614"
                     /translation="SSIYNGISTSGLDLNNGTIADMRQLGIVESYKLKRAVVSSASEA
                     AEVLLRVDNIIRARPRTANRQHM"
     gene            687..3158
                     /gene="AXL2"
     CDS             687..3158
                     /gene="AXL2"
```

EMBL format

- Similar to GenBank, except that keywords are two-letter IDs.
- UniProt's format is similar to this format.

```
ID   SC10H5 standard; DNA; PRO; 4870 BP.
XX
AC   AL031232;
XX
DE   Streptomyces coelicolor cosmid 10H5.
XX
OS   Streptomyces coelicolor
OC   Eubacteria; Firmicutes; Actinomycetes; Streptomycetes;
OC   Streptomycetaceae; Streptomyces.
XX
RN   [1]
RP   1-4870
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A.,
RA   Cullum J., Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical
RT   map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21(1):77-96(1996).
XX
CC   Notes:
CC
CC   Streptomyces coelicolor sequencing at The Sanger Centre is funded
CC   by the BBSRC.
CC
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..4870
FT                   /organism="Streptomyces coelicolor"
FT                   /strain="A3(2)"
FT                   /clone="cosmid 10H5"
FT   CDS             complement(350..805)
FT                   /note="SC10H5.02c, probable integral membrane protein, len:
FT                   151 aa; similar to S. coelicolor hypothetical protein
FT                   TR:O54194 (EMBL:AL021411) SC7H1.35 (155 aa), fasta scores;
FT                   opt: 431 z-score: 749.8 E(): 0, 53.5% identity in 114 aa
FT                   overlap."
FT                   /product="putative integral membrane protein"
FT                   /gene="SC10H5.02c"
FT   RBS             complement(812..815)
FT                   /note="possible RBS upstream of SC10H5.02c"
FT   CDS             complement(837..1301)
FT                   /note="SC10H5.03c, probable integral membrane protein, len:
FT                   154 aa"
FT                   /product="putative integral membrane protein"
FT                   /gene="SC10H5.03c"
FT   RBS             complement(1308..1312)
FT                   /note="possible RBS upstream of SC10H5.03c"
```

FASTA sequence format

- > Randseq1 first randomly generated seq
  GGTGGTTACTAACCGTAAGAGATGATGTCGCCGTGGTCGCGTGGC
  GCCGCGGACCCAGATTGTACTTCTCTGAGTCGTTCTAGATCGACC
  AGTCTTCTAGCTTGCCCGTGAGGTATGGGG
  AGCCGCATATTGCCCACAAT
  > Randseq2 second randomly generated seq
  GCGACGCGTCTCTACACCAGACGCTTCTGTTGAGGAAGAGTGCCT
  GAGTGCAGGTCCTCGAGAACCCACTGGAACTTGAAGGGCGCGTCT
  CACTGGTCGTGAGAAGGCTCCGTCGATACG
  AAAGTCCATGCCAAGGACAT
  > Randseq3 third randomly generated seq
  GGCGAGTCTGAACTCACAAATATTGCACGAGAGTTTAGTGTATGT
  TCCTCTTAGGCTGATAACAATAGTTTAGTGAGCGGAAATGCAACC
  GCGAGGCGGTCCCCTGCGCTTGTAATGGCC
  ACCTGTTGCCCGTCGGATAT

Nucleic acid code for FASTA

- A → adenosine
- C → cytidine
- G → guanine
- T → thymidine
- U → uridine
- R → G A (purine)
- Y → T C (pyrimidine)
- K → G T (keto)
- M → A C (amino)
- S → G C (strong)
- W → A T (weak)
- B → G T C
- D → G A T
- H → A C T
- V → G C A
- N → A G C T (any)
- - → gap of indeterminate length

Amino acid code for FASTA

- A alanine
- B aspartate or asparagine
- C cystine
- D aspartate
- E glutamate
- F phenylalanine
- G glycine
- H histidine
- I isoleucine
- K lysine
- L leucine
- M methionine
- N asparagine
- P proline
- Q glutamine
- R arginine
- S serine
- T threonine
- U selenocysteine
- V valine
- W tryptophan
- Y tyrosine
- Z glutamate or glutamine
- X any
- * translation stop
- - gap of indeterminate length

PDB format

- Similar to GenBank, using different keywords
- Includes 3-dimensional coordinates of amino acids

```
HEADER    HYDROLASE(ACID PROTEINASE ZYMOGEN)      03-SEP-91   3PSG      3PSG   2
COMPND    PEPSINOGEN                                                    3PSG   3
SOURCE    PORCINE (SUS ¥$SCROFA)                                         3PSG   4
AUTHOR    J.A.HARTSUCK,G.KOELSCH,S.J.REMINGTON                          3PSG   5
REVDAT   1   15-JAN-93 3PSG    0                                        3PSG   6
SPRSDE     15-JAN-93 3PSG      1PSG                                     3PSG   7
JRNL        AUTH   J.A.HARTSUCK,G.KOELSCH,S.J.REMINGTON                3PSG   8
JRNL        TITL   THE HIGH RESOLUTION CRYSTAL STRUCTURE OF PORCINE     3PSG   9
JRNL        TITL 2 PEPSINOGEN                                          3PSG  10
JRNL        REF    PROTEINS.STRUCT.,FUNCT.,      V.  13     1 1992     3PSG  11
JRNL        REF  2 GENET.                                              3PSG  12
JRNL        REFN   ASTM PSFGEY  US ISSN 0887-3585                  867  3PSG  13
REMARK   1                                                            3PSG  14
REMARK   2                                                            3PSG  15
REMARK   2 RESOLUTION. 1.65 ANGSTROMS.                                3PSG  16
REMARK   3                                                            3PSG  17
REMARK   3 REFINEMENT.                                                3PSG  18
REMARK   3    PROGRAM                     TNT                         3PSG  19
REMARK   3    AUTHORS                     TRONRUD,TEN EYCK & MATHEWS  3PSG  20
REMARK   3    R VALUE                     0.173                       3PSG  21
REMARK   3    RMSD BOND DISTANCES         0.014 ANGSTROMS             3PSG  22
REMARK   3    RMSD BOND ANGLES            2.60  DEGREES               3PSG  23
SEQRES   1    370  LEU VAL LYS VAL PRO LEU VAL ARG LYS LYS SER LEU ARG  3PSG  76
SEQRES   2    370  GLN ASN LEU ILE LYS ASP GLY LYS LEU LYS ASP PHE LEU  3PSG  77
SEQRES   3    370  LYS THR HIS LYS HIS ASN PRO ALA SER LYS TYR PHE PRO  3PSG  78
SEQRES   4    370  GLU ALA ALA ALA LEU ILE GLY ASP GLU PRO LEU GLU ASN  3PSG  79
SEQRES   5    370  TYR LEU ASP THR GLU TYR PHE GLY THR ILE GLY ILE GLY  3PSG  80
SEQRES   6    370  THR PRO ALA GLN ASP PHE THR VAL ILE PHE ASP THR GLY  3PSG  81
FTNOTE   1                                                            3PSG 105
FTNOTE   1 RESIDUE PRO 23 IS A CIS PROLINE.                           3PSG 106
FORMUL   2  HOH   *180(H2 O1)                                         3PSG 107
CRYST1  106.100   43.700   88.900  90.00  91.40  90.00 C 2          4  3PSG 108
ORIGX1      1.000000  0.000000  0.000000        0.00000               3PSG 109
ORIGX2      0.000000  1.000000  0.000000        0.00000               3PSG 110
ORIGX3      0.000000  0.000000  1.000000        0.00000               3PSG 111
SCALE1      0.009425  0.000000  0.000230        0.00000               3PSG 112
SCALE2      0.000000  0.022883  0.000000        0.00000               3PSG 113
SCALE3      0.000000  0.000000  0.011252        0.00000               3PSG 114
ATOM      1  N   LEU     1P     57.364  -9.595   2.554  1.00 21.58     3PSG 115
ATOM      2  CA  LEU     1P     56.687  -8.586   3.371  1.00 22.68     3PSG 116
ATOM      3  C   LEU     1P     57.052  -8.758   4.847  1.00 16.32     3PSG 117
ATOM      4  O   LEU     1P     57.270  -9.875   5.293  1.00 26.09     3PSG 118
ATOM      5  CB  LEU     1P     55.129  -8.641   3.226  1.00 36.56     3PSG 119
ATOM      6  CG  LEU     1P     54.503  -8.303   1.850  1.00 32.34     3PSG 120
ATOM      7  CD1 LEU     1P     52.977  -8.461   1.892  1.00 37.73     3PSG 121
ATOM      8  CD2 LEU     1P     54.798  -6.861   1.453  1.00 30.91     3PSG 122
ATOM      9  N   VAL     2P     57.149  -7.672   5.627  1.00 17.11     3PSG 123
ATOM     10  CA  VAL     2P     57.472  -7.872   7.047  1.00 15.32     3PSG 124
```

ASN.1 format

```
{     name              {givenName "John", initial "P", familyName "Smith"},
      title             "Director",
      number            51,
      dateOfHire        "19710917",
      nameOfSpouse      {givenName "Mary", initial "T", familyName "Smith"},
      children
      {  {name  {givenName "Ralph", initial "T", familyName "Smith"} ,
          dateOfBirth "19571111"},
         {name  {givenName "Susan", initial "B", familyName "Jones"} ,
          dateOfBirth "19590717" }
      }
}
```

- Hierarchical data format
- Groups are delineated by curly brackets
- Data type names precede the brackets
- Data within a group are separated by commas

Name:	John P Smith
Title:	Director
Employee Number:	51
Date of Hire:	17 September 1971
Name of Spouse:	Mary T Smith
Number of Children:	2
Child Information	
Name:	Ralph T Smith
Date of Birth	11 November 1957
Child Information	
Name:	Susan B Jones
Date of Birth	17 July 1959

XML format

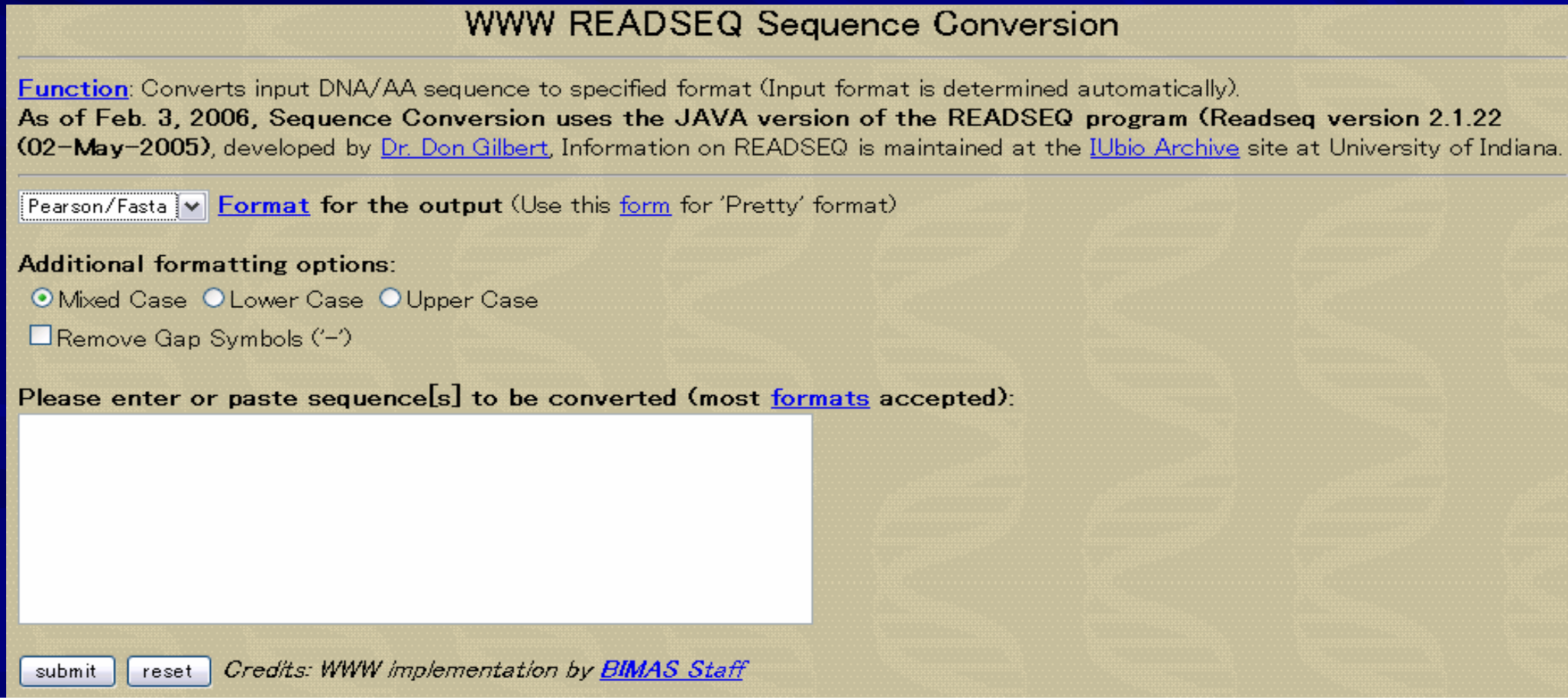
```
person ::=
      <PersonnelRecord>
         <name>
            <givenName>John</givenName>
            <initial>P</initial>
            <familyName>Smith</familyName>
         </name>
         <title>Director</title>
         <number>51</number>
         <dateOfHire>19710917</dateOfHire>
         <nameOfSpouse>
            <givenName>Mary</givenName>
            <initial>T</initial>
            <familyName>Smith</familyName>
         </nameOfSpouse>
         <children>
            <ChildInformation>
               <name>
                  <givenName>Ralph</givenName>
                  <initial>T</initial>
                  <familyName>Smith</familyName>
               </name>
               <dateOfBirth>19571111</dateOfBirth>
            </ChildInformation>
            <ChildInformation>
               <name>
                  <givenName>Susan</givenName>
                  <initial>B</initial>
                  <familyName>Jones</familyName>
               </name>
               <dateOfBirth>19590717</dateOfBirth>
            </ChildInformation>
         </children>
      </PersonnelRecord>
```

Name:	John P Smith
Title:	Director
Employee Number:	51
Date of Hire:	17 September 1971
Name of Spouse:	Mary T Smith
Number of Children:	2
Child Information	
Name:	Ralph T Smith
Date of Birth	11 November 1957
Child Information	
Name:	Susan B Jones
Date of Birth	17 July 1959

- Hierarchical data format
- Tags define the type of data:
  - Opening tag: <name>
  - Closing tag: </name>
- Data are delineated by opening and closing tags

Data format converter

- READSEQ
  - URL : http://thr.cit.nih.gov/molbio/readseq/



WWW READSEQ Sequence Conversion

Function: Converts input DNA/AA sequence to specified format (Input format is determined automatically).
As of Feb. 3, 2006, Sequence Conversion uses the JAVA version of the READSEQ program (Readseq version 2.1.22 (02-May-2005), developed by Dr. Don Gilbert, Information on READSEQ is maintained at the IUbio Archive site at University of Indiana.

Pearson/Fasta **Format for the output** (Use this form for 'Pretty' format)

Additional formatting options:

Mixed Case Lower Case Upper Case

Remove Gap Symbols ('-')

Please enter or paste sequence[s] to be converted (most formats accepted):

submit reset *Credits: WWW implementation by BIMAS Staff*

Types of databases

- Data resources of multiple types of data
  - EBI (European Bioinformatics Institute)
  - NCBI (National Center for Biotechnology Information)
  - KEGG (Kyoto Encyclopedia of Genes and Genomes)
- Gene and protein information
  - GenBank, UniProt, and PDB
  - Species specific: FlyBase, dictyBase, etc.
- Ontological data
  - Gene Ontology
- Pathway data
  - KEGG PATHWAY, Reactome, BRENDA, etc.
- Protein-protein interaction data
  - IntAct, BioGRID, etc.

EBI

- http://www.ebi.ac.uk/
- European base of molecular biology information, including genomic, gene expression, and literature information

EMBL-EBI

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

Databases | Tools | EBI Groups | Training | Industry | About Us | Help | Site Index

Data Resources &Tools

- EMBL-BANK
- UniProt
- ArrayExpress
- Ensembl
- InterPro
- PDB-EBI
- Genomes
- Nucleotide Sequences
- Protein Sequences
- Macromolecular Struc
- Small Molecules
- Gene Expression
- Molecular Interactions
- Literature
- Taxonomy
- Sequence Similarity & Analysis
- Pattern & Motif Searches

About the EBI

- Research
- PhD Studies
- Training
- Industry Support
- Group & Team Leaders
- EBI Funders
- User Su
- EBI Miss
- People
- Events a
- How to F

EMBL-EBI

EB-eye Search | All Databases | HIV | Go | Reset | Advanced Search | Give us feedback

Databases | Tools | EBI Groups | Training | Industry | About Us | Help | Site Index

Search for *HIV* in *All the EBI*

Expand all | Collapse all

Genomes	87	Molecular Interactions	3
Nucleotide Sequences	250,495	Reactions & Pathways	36
Protein Sequences	51,222	Protein Families	41
Macromolecular Structures	628	Enzymes	2
Small molecules	3	Literature	183,676
Gene Expression	23	Ontologies	6
		EBI Web Site	34

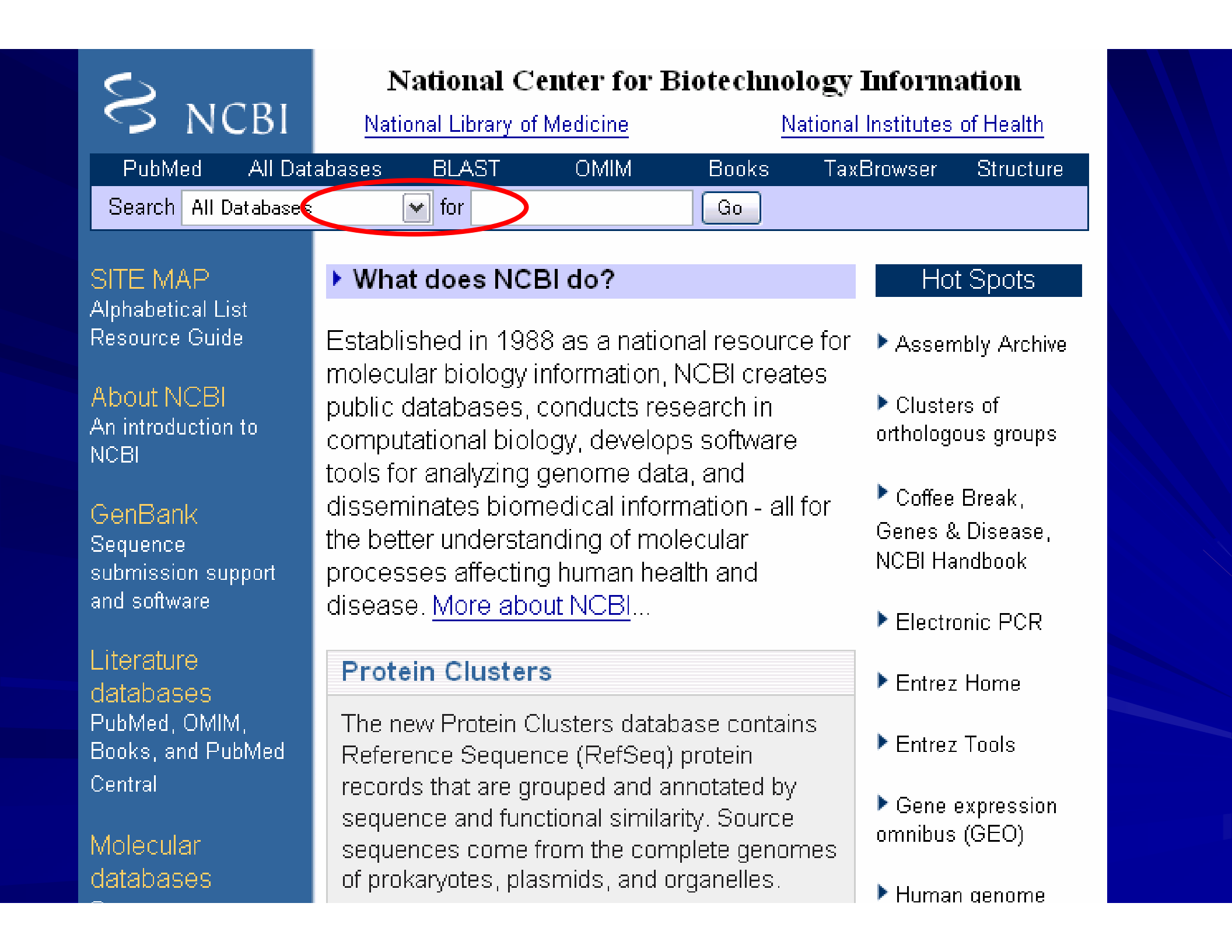
Refine your search:

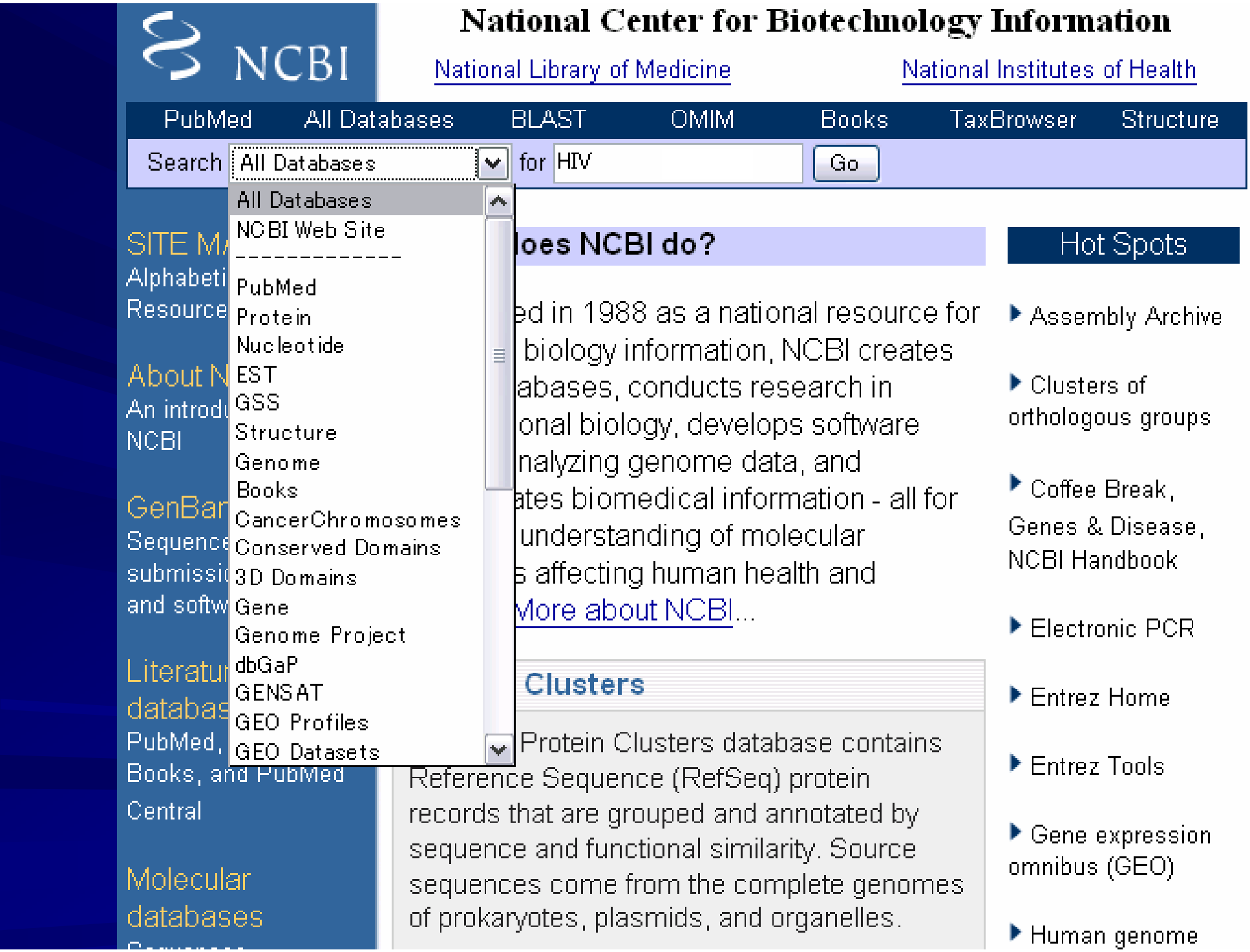
Search for *HIV* in *All the EBI*

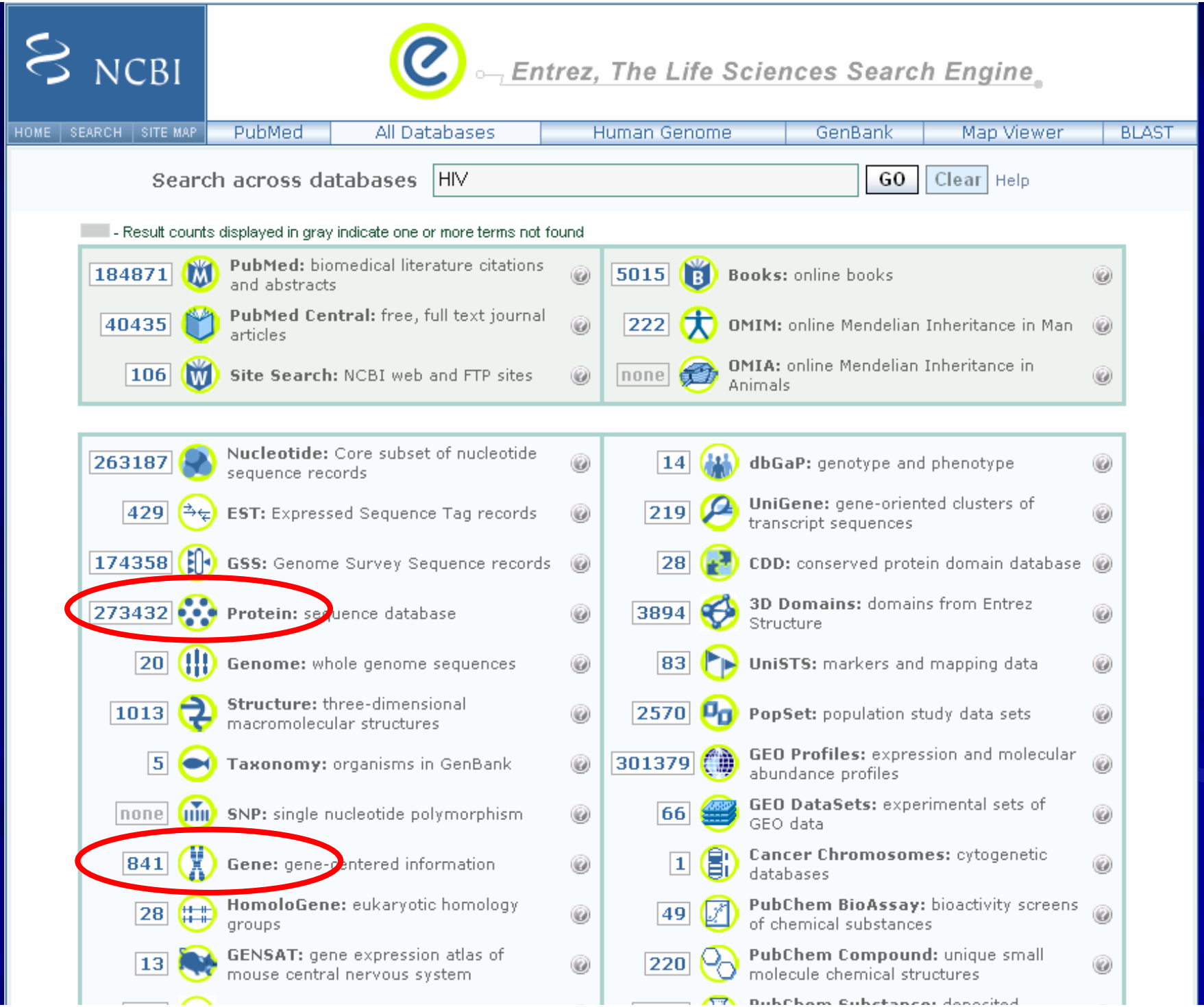
with the following keywords | Refine

NCBI

- http://www.ncbi.nlm.nih.gov/
- Contains public databases of molecular biology information including genomes, microarray gene expression, protein sequence domains, etc.
- Develops software for analyzing genome data, including BLAST
- Provides PubMed, an archive of biomedical and life science journals


NCBI
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
PubMed
All Databases
BLAST
OMIM
Books
TaxBrowser
Structure
Search
All Databases
for
Go
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
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 about NCBI...
Protein Clusters
The new Protein Clusters database contains Reference Sequence (RefSeq) protein records that are grouped and annotated by sequence and functional similarity. Source sequences come from the complete genomes of prokaryotes, plasmids, and organelles.
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


NCBI
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
PubMed
All Databases
BLAST
OMIM
Books
TaxBrowser
Structure
Search
All Databases
for
HIV
Go
All Databases
NCBI Web Site
PubMed
Protein
Nucleotide
EST
GSS
Structure
Genome
Books
CancerChromosomes
Conserved Domains
3D Domains
Gene
Genome Project
dbGaP
GENSAT
GEO Profiles
GEO Datasets
SITE MA
Alphabeti
Resource
About N
An introdu
NCBI
GenBar
Sequence
submissio
and softw
Literatur
databas
PubMed,
Books, and PubMed
Central
Molecular
databases
oes NCBI do?
ed in 1988 as a national resource for
biology information, NCBI creates
abases, conducts research in
onal biology, develops software
nalyzing genome data, and
ates biomedical information - all for
understanding of molecular
s affecting human health and
More about NCBI...
Clusters
Protein Clusters database contains
Reference Sequence (RefSeq) protein
records that are grouped and annotated by
sequence and functional similarity. Source
sequences come from the complete genomes
of prokaryotes, plasmids, and organelles.
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


NCBI
Entrez, The Life Sciences Search Engine
HOME
SEARCH
SITE MAP
PubMed
All Databases
Human Genome
GenBank
Map Viewer
BLAST
Search across databases
HIV
GO
Clear
Help
- Result counts displayed in gray indicate one or more terms not found
184871 PubMed: biomedical literature citations and abstracts
40435 PubMed Central: free, full text journal articles
106 Site Search: NCBI web and FTP sites
5015 Books: online books
222 OMIM: online Mendelian Inheritance in Man
none OMIA: online Mendelian Inheritance in Animals
263187 Nucleotide: Core subset of nucleotide sequence records
429 EST: Expressed Sequence Tag records
174358 GSS: Genome Survey Sequence records
273432 Protein: sequence database
20 Genome: whole genome sequences
1013 Structure: three-dimensional macromolecular structures
5 Taxonomy: organisms in GenBank
none SNP: single nucleotide polymorphism
841 Gene: gene-centered information
28 HomoloGene: eukaryotic homology groups
13 GENSAT: gene expression atlas of mouse central nervous system
14 dbGaP: genotype and phenotype
219 UniGene: gene-oriented clusters of transcript sequences
28 CDD: conserved protein domain database
3894 3D Domains: domains from Entrez Structure
83 UniSTS: markers and mapping data
2570 PopSet: population study data sets
301379 GEO Profiles: expression and molecular abundance profiles
66 GEO DataSets: experimental sets of GEO data
1 Cancer Chromosomes: cytogenetic databases
49 PubChem BioAssay: bioactivity screens of chemical substances
220 PubChem Compound: unique small molecule chemical structures

Gene and protein databases

- UniProt
  - Universal Protein Resource
  - http://beta.uniprot.org/
  - Consists of three components:
    - 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.

UniProt
the universal protein resource

Home | About UniProt | Getting Started | Searches/T

Notice: This site will be replaced with beta.u

Text Search
BLAST
FAQ
Help Desk
Download

Welcome to UniProt

UniProt (Universal Protein R
catalog of information on p
sequence and function cre
Swiss-Prot, TrEMBL, and PIR

UniProt has three compone
UniProt Knowledgebase (
extensive curated protein
and cross-reference. The
databases combine closely r
searches. The **UniProt Arch**
reflecting the history of all p

The sequences and informat
BLAST similarity search, a

EBI
European
Bioinformatics Institute

About UniProt | Getting Started | Searches/Tools | Databas
HOME | HELP | SITE MAP © 2006 UniProt Consortium | License

UniProt

Downloads · Contact · **Help**

Search in Protein Knowledgebase (UniProtKB)

Query Search | Clear | Fields »

Search | Blast | Align | Retrieve | ID Mapping

WELCOME

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

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: ★ Swiss-Prot, which is manually annotated and reviewed. ★ TrEMBL, which is automatically annotated and is **not** reviewed.
UniRef	Sequence clusters, used to speed up similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations, taxonomy, keywords and more.

UniProt

This is a beta version of our new web site, we are welcoming your feedback. Please take into account that some functions are incomplete and many help pages are still preliminary.

NEWS

Release 13.4 – May 20, 2008

Swiss-Prot in the Wonderland of protein names · Cross-references to CGD

› Statistics for UniProtKB: Swiss-Prot · TrEMBL
› Forthcoming changes
› News archives

SITE TOUR

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

PROTEIN SPOTLIGHT

The selfish smell
May 2008

We are surrounded by smells. Pleasant ones and not so pleasant ones, hard to distinguish ones, mild ones and strong ones. Smells are not part of our everyday life for the simple sake of pleasure. They are there for a purpose…

UniProt > UniProtKB

Downloads · Contact · Help

Search in Protein Knowledgebase (UniProtKB) **Query** HIV Search Clear Fields »

Search | Blast | Align | Retrieve | ID Mapping *

1 – 25 of **151,248** results for **HIV** in **UniProtKB** sorted by **score** descending

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

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

> Restrict term "hiv" to author, gene name, protein name, organism, strain, taxonomy, web resource

Page 1 of 6,050 | Next »

All	Accession	Entry Name	Status	Protein Names	Genes	Organism	Length
☐	P04585	POL_HV1H2	★	**Gag-Pol polyprotein** (Pr160Gag-Pol) [Cleaved into: Matrix protein p17 (MA); Capsid protein p24 (CA); Spacer peptide p2; Nucleocapsid protein p7 (NC); Transframe peptide (TF); p6-pol (p6*); Protease (EC 3.4.23.16) (Retropepsin) (PR); Reverse transcriptase/ribonuclease H (EC 2.7.7.49) (EC 2.7.7.7) (EC 3.1.26.4) (p66 RT); p51 RT; p15; Integrase (IN)]	**gag-pol**	Human immunodeficiency virus type 1 (isolate HXB2 group M subtype B) (HIV-1)	1,435
☐	P03366	POL_HV1B1	★	**Gag-Pol polyprotein** (Pr160Gag-Pol) [Cleaved into: Matrix protein p17 (MA); Capsid protein p24 (CA); Spacer peptide p2; Nucleocapsid protein p7 (NC); Transframe peptide (TF); p6-pol (p6*); Protease (EC 3.4.23.16) (Retropepsin) (PR); Reverse transcriptase/ribonuclease H (EC 2.7.7.49) (EC 2.7.7.7) (EC 3.1.26.4) (p66 RT); p51 RT; p15; Integrase (IN)]	**gag-pol**	Human immunodeficiency virus type 1 (isolate BH10 group M subtype B) (HIV-1)	1,447
☐	P03367	POL_HV1BR	★	**Gag-Pol polyprotein** (Pr160Gag-Pol) [Cleaved into: Matrix protein p17 (MA); Capsid protein p24 (CA); Spacer peptide p2; Nucleocapsid protein p7 (NC); Transframe peptide (TF); p6-pol (p6*); Protease (EC 3.4.23.16) (Retropepsin) (PR); Reverse transcriptase/ribonuclease H (EC 2.7.7.49) (EC 2.7.7.7) (EC 3.1.26.4) (p66 RT); p51 RT; p15; Integrase (IN)]	**gag-pol**	Human immunodeficiency virus type 1 (isolate BRU/LAI group M subtype B) (HIV-1)	1,447

EMBL-EBI

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

Databases | Tools | EBI Groups | Training | Industry | About Us | Help | Site Index

Data Resources &Tools

- EMBL-BANK
- UniProt
- ArrayExpress
- Ensembl
- InterPro
- PDB-EBI
- Genomes
- Nucleotide Sequences
- Protein Sequences
- Macromolecular Struc
- Small Molecules
- Gene Expression
- Molecular Interactions
- Literature
- Taxonomy
- Sequence Similarity & Analysis
- Pattern & Motif Searches

About the EBI

- Research
- PhD Studies
- Training
- Industry Support
- Group & Team Leaders
- EBI Funders
- User Su
- EBI Miss
- People
- Events a
- How to F

EMBL-EBI

EB-eye Search | All Databases | HIV | Go | Reset | Advanced Search | Give us feedback

Databases | Tools | EBI Groups | Training | Industry | About Us | Help | Site Index

Search for *HIV* in *All the EBI*

Expand all | Collapse all

Genomes	87	Molecular Interactions	3
Nucleotide Sequences	250,495	Reactions & Pathways	36
Protein Sequences	51,222	Protein Families	41
Macromolecular Structures	628	Enzymes	2
Small molecules	3	Literature	183,676
Gene Expression	23	Ontologies	6
		EBI Web Site	34

Refine your search:

Search for *HIV* in *All the EBI*

with the following keywords | Refine

51,108 results found in UniProt KB

HTSF1_HUMAN
O43719, Q59G06, Q99730
HIV Tat-specific factor 1 (Tat-SF1).
View: in UniProt format | in SRS | in UniSave | in Interpro matches
References: EMBL-Bank | Ensembl | InterPro | Medline | Taxonomy | MSD/PDB | EMBL-Bank (Coding Sequence) | GO

A2R057_ASPNG
A2R057
Function: N. ellipsosporum CV-N has anti-HIV activity.
View: in UniProt format | in SRS | in UniSave | in Interpro matches
References: Taxonomy | EMBL-Bank | EMBL-Bank (Coding Sequence) | InterPro | Medline

Q9UDI3_HUMAN
Q9UDI3
26 S protease subunit 7, MSS1=MODULATOR of HIV TAT-mediatedtransactivation (Fragments).
View: in UniProt format | in SRS | in UniSave | in Interpro matches
References: Taxonomy | Medline

more...

51 results found in UniRef100

UNIREF100_O43719
Cluster: HIV Tat-specific factor 1
View: in UniRef format | in SRS
References: Taxonomy | UniParc | UniRef50 | UniRef90 | UniProt KB

UNIREF100_UPI0000DBF22B
Cluster: HIV TAT specific factor 1
View: in UniRef format | in SRS
References: Taxonomy | UniParc | UniRef50 | UniRef90

UNIREF100_UPI00015A4CD3
Cluster: HIV TAT specific factor 1
View: in UniRef format | in SRS
References: Taxonomy | UniParc | UniRef50 | UniRef90

more...

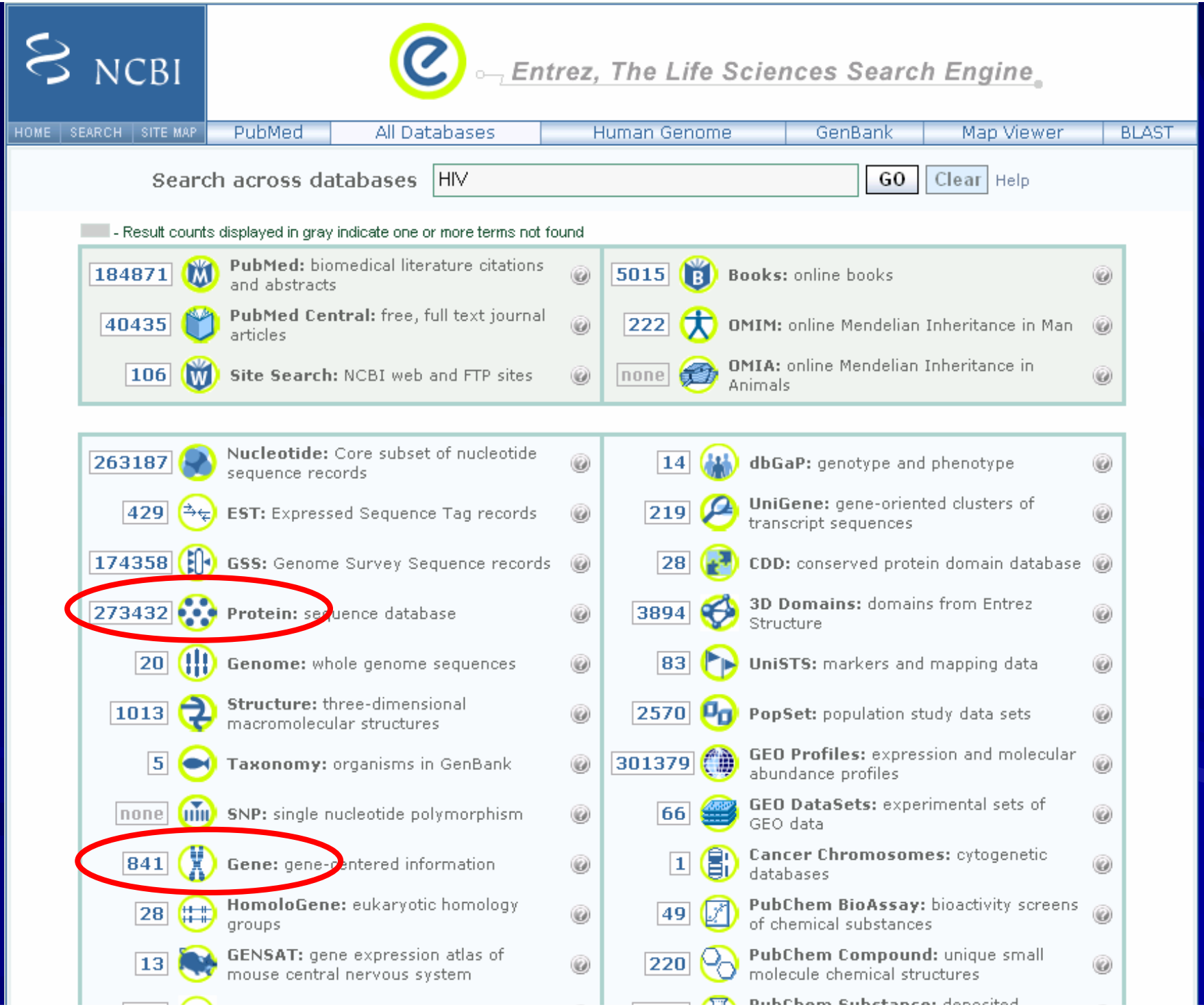
The UniRef databases provide clustered sets of sequences from UniProt (including splice variants and isoforms) and selected UniParc records, in order to obtain complete coverage of sequence space at several resolutions while hiding redundant sequences (but not their descriptions) from view. The UniRef100 database combines identical sequences and sub-fragments with 11 or more residues (from any organism) into a single UniRef entry, displaying the sequence of a representative protein, the accession numbers of all the merged entries, and links to the corresponding UniProtKB and UniParc records.

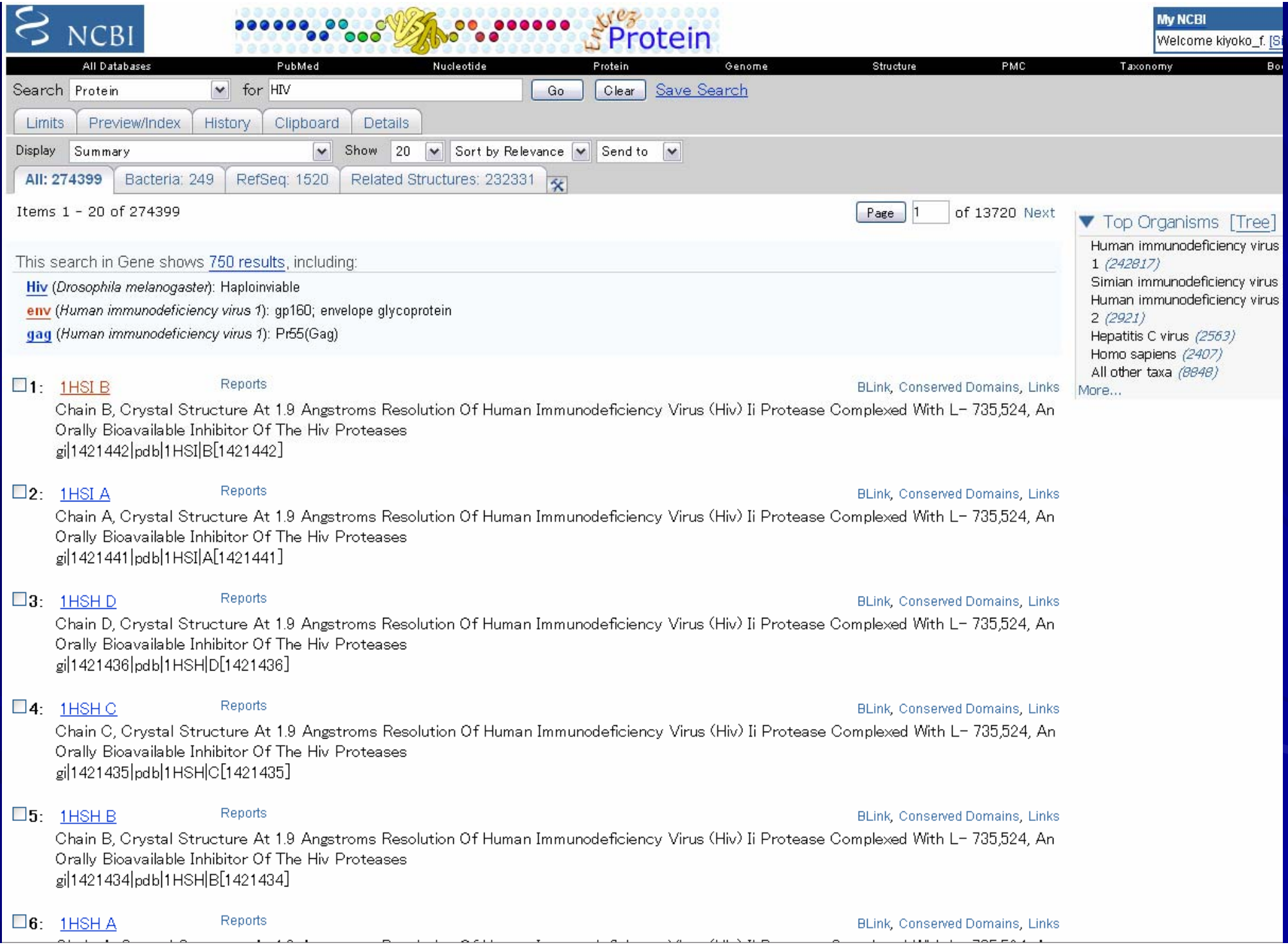
Types of databases

- Data resources of multiple types of data
  - EBI (European Bioinformatics Institute)
  - NCBI (National Center for Biotechnology Information)
  - KEGG (Kyoto Encyclopedia of Genes and Genomes)
- Gene and protein information
  - GenBank, UniProt, and PDB
  - Species specific: FlyBase, dictyBase, etc.
- Ontological data
  - Gene Ontology
- Pathway data
  - KEGG PATHWAY, Reactome, BRENDA, etc.
- Protein-protein interaction data
  - IntAct, BioGRID, etc.

Gene and protein databases

- GenBank
  - A part of NCBI
  - http://www.ncbi.nlm.nih.gov/
  - Search can be performed through the Entrez interface, which searches for the query in all databases available at the NCBI


NCBI
Entrez, The Life Sciences Search Engine
HOME
SEARCH
SITE MAP
PubMed
All Databases
Human Genome
GenBank
Map Viewer
BLAST
Search across databases
HIV
GO
Clear
Help
- Result counts displayed in gray indicate one or more terms not found
184871 PubMed: biomedical literature citations and abstracts
40435 PubMed Central: free, full text journal articles
106 Site Search: NCBI web and FTP sites
5015 Books: online books
222 OMIM: online Mendelian Inheritance in Man
none OMIA: online Mendelian Inheritance in Animals
263187 Nucleotide: Core subset of nucleotide sequence records
429 EST: Expressed Sequence Tag records
174358 GSS: Genome Survey Sequence records
273432 Protein: sequence database
20 Genome: whole genome sequences
1013 Structure: three-dimensional macromolecular structures
5 Taxonomy: organisms in GenBank
none SNP: single nucleotide polymorphism
841 Gene: gene-centered information
28 HomoloGene: eukaryotic homology groups
13 GENSAT: gene expression atlas of mouse central nervous system
14 dbGaP: genotype and phenotype
219 UniGene: gene-oriented clusters of transcript sequences
28 CDD: conserved protein domain database
3894 3D Domains: domains from Entrez Structure
83 UniSTS: markers and mapping data
2570 PopSet: population study data sets
301379 GEO Profiles: expression and molecular abundance profiles
66 GEO DataSets: experimental sets of GEO data
1 Cancer Chromosomes: cytogenetic databases
49 PubChem BioAssay: bioactivity screens of chemical substances
220 PubChem Compound: unique small molecule chemical structures


NCBI
Entrez Protein
My NCBI
Welcome kiyoko_f. [S
All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Bo
Search Protein for HIV Go Clear Save Search
Limits Preview/Index History Clipboard Details
Display Summary Show 20 Sort by Relevance Send to
All: 274399 Bacteria: 249 RefSeq: 1520 Related Structures: 232331
Items 1 - 20 of 274399
Page 1 of 13720 Next
This search in Gene shows 750 results, including:
Hiv (Drosophila melanogaster): Haploinviable
env (Human immunodeficiency virus 1): gp160; envelope glycoprotein
gag (Human immunodeficiency virus 1): Pr55(Gag)
1: 1HSI B Reports BLink, Conserved Domains, Links
Chain B, Crystal Structure At 1.9 Angstroms Resolution Of Human Immunodeficiency Virus (Hiv) Ii Protease Complexed With L- 735,524, An Orally Bioavailable Inhibitor Of The Hiv Proteases
gi|1421442|pdb|1HSI|B[1421442]
2: 1HSI A Reports BLink, Conserved Domains, Links
Chain A, Crystal Structure At 1.9 Angstroms Resolution Of Human Immunodeficiency Virus (Hiv) Ii Protease Complexed With L- 735,524, An Orally Bioavailable Inhibitor Of The Hiv Proteases
gi|1421441|pdb|1HSI|A[1421441]
3: 1HSH D Reports BLink, Conserved Domains, Links
Chain D, Crystal Structure At 1.9 Angstroms Resolution Of Human Immunodeficiency Virus (Hiv) Ii Protease Complexed With L- 735,524, An Orally Bioavailable Inhibitor Of The Hiv Proteases
gi|1421436|pdb|1HSH|D[1421436]
4: 1HSH C Reports BLink, Conserved Domains, Links
Chain C, Crystal Structure At 1.9 Angstroms Resolution Of Human Immunodeficiency Virus (Hiv) Ii Protease Complexed With L- 735,524, An Orally Bioavailable Inhibitor Of The Hiv Proteases
gi|1421435|pdb|1HSH|C[1421435]
5: 1HSH B Reports BLink, Conserved Domains, Links
Chain B, Crystal Structure At 1.9 Angstroms Resolution Of Human Immunodeficiency Virus (Hiv) Ii Protease Complexed With L- 735,524, An Orally Bioavailable Inhibitor Of The Hiv Proteases
gi|1421434|pdb|1HSH|B[1421434]
6: 1HSH A Reports BLink, Conserved Domains, Links
Top Organisms [Tree]
Human immunodeficiency virus 1 (242817)
Simian immunodeficiency virus
Human immunodeficiency virus 2 (2921)
Hepatitis C virus (2563)
Homo sapiens (2407)
All other taxa (8848)
More...

PubMed Nucleotide Protein Genome Structure PMC

Search Protein for Go Clear

Display GenPept Show 5 Send to

Range: from begin to end Features: CDD + Refresh

1: 1HSI B. Reports Chain B, Crystal ...[gi:1421442]

Comment Features Sequence

```
LOCUS       1HSI_B                     99 aa            linear   VRL 01-OCT-2007
DEFINITION  Chain B, Crystal Structure At 1.9 Angstroms Resolution Of Human
            Immunodeficiency Virus (Hiv) Ii Protease Complexed With L- 735,524,
            An Orally Bioavailable Inhibitor Of The Hiv Proteases.
ACCESSION   1HSI_B
VERSION     1HSI_B  GI:1421442
DBSOURCE    pdb: molecule 1HSI, chain 66, release Aug 27, 2007;
            deposition: Aug 27, 2007;
            class: Hydrolase (Acid Proteinase);
            source: Mol_id: 1; Organism_scientific: Human Immunodeficiency
            Virus 2; Gene: Hiv-2 Protease From The Rod Isolate;
            Expression_system: Escherichia Coli;
            Exp. method: X-Ray Diffraction.
KEYWORDS    .
SOURCE      Human immunodeficiency virus 2 (HIV-2)
  ORGANISM  Human immunodeficiency virus 2
            Viruses; Retro-transcribing viruses; Retroviridae;
            Orthoretrovirinae; Lentivirus; Primate lentivirus group.
REFERENCE   1  (residues 1 to 99)
  AUTHORS   Chen,Z., Li,Y., Chen,E., Hall,D.L., Darke,P.L., Culberson,C.,
            Shafer,J.A. and Kuo,L.C.
  TITLE     Crystal structure at 1.9-A resolution of human immunodeficiency
            virus (HIV) II protease complexed with L-735,524, an orally
            bioavailable inhibitor of the HIV proteases
```

Gene and protein databases

- PDB: Protein DataBank
  - Contains 3-dimensional protein structures
  - http://www.rcsb.org
  - Data submitted by individual researchers

RCSB PDB
PROTEIN DATA BANK

An Informat
As of Tuesday May 2

CONTACT US | HELP | PRINT PAGE PDB ID or keyword Author

Home | Search

- Home
- Getting Started
- Download Files
- Deposit and Validate
- Structural Genomics
- Dictionaries & File Formats
- Software Tools
- General Education
- Site Tutorials
- BioSync
- General Information
- Acknowledgements
- Frequently Asked Questions
- Report Bugs/Comments

Quick Tips :

Having trouble with the web site? Try the tutorial: click here

Are you missing da
ftp://ftp.wwpdb.org.
For more information clic

Welcome to th

The RCSB PDB provides a
for studying the structures
and their relationships to s
disease.

The RCSB is a member of t
to ensure that the PDB arc
resource with uniform data

This site offers tools for br
reporting that utilize the d
efforts to create a more c
archive.

Information about compati
here.

A narrated tutorial illu
navigate, browse, generat
structures using this new
player download.]

Comments? info@rcsb.org

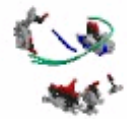
Molecule of the Month: Pri

Help | 458 Structure Hits | 51 Unreleased Structures | 183 Ligand Hits | 229 Citations
44 Web Page Hits | GO Hits | SCOP Hits | CATH Hits

Advanced Keyword Query for: HIV

1 2 3 4 5 .. 46

1j9v



Solution structure of a lactam analogue (DabD) of HIV gp41 600-612 loop.

Characteristics Classification: Release Date: 01-Jul-2003 Exp. Method: NMR 49 Structures
Viral Protein

Compound: Polymer: 1 Molecule: DabD (Ace)IWG(DAB)SGKLIDTTA ANALOGUE OF HIV GP41 Chains: A

Authors: Phan Chan Du, A., Limal, D., Semetey, V., Dali, H., Jolivet, M., Desgranges, C., Cung, M.T., Briand, J.P., Petit, M.C., Muller, S.

1jaa

Solution structure of lactam analogue (DapE) of HIV gp41 600-612 loop.

Characteristics Classification: Release Date: 01-Jul-2003 Exp. Method: NMR 50 Structures
Viral Protein

Compound: Polymer: 1 Molecule: DapE : (Ace)IWG(Dap)SGKLIETTA ANALOGUE OF HIV GP41 Chains: A

Authors: Phan Chan Du, A., Limal, D., Semetey, V., Dali, H., Jolivet, M., Desgranges, C., Cung, M.T., Briand, J.P., Petit, M.C., Muller, S.

1jar

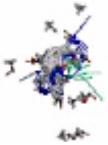
Solution structure of lactam analogue (DDab)of HIV gp41 600-612 loop.

Characteristics Classification: Release Date: 01-Jul-2003 Exp. Method: NMR 49 Structures
Viral Protein

Compound: Polymer: 1 Molecule: DDab: (ACE)IWGDSGKLI(DAB)TTA ANALOGUE OF HIV GP41 Chains: A

Authors: Phan Chan Du, A., Limal, D., Semetey, V., Dali, H., Jolivet, M., Desgranges, C., Cung, M.T., Briand, J.P., Petit, M.C., Muller, S.

1jc8



Solution structure of lactam analogue (DDap) of gp41 600-612 loop of HIV

Characteristics Classification: Release Date: 01-Jul-2003 Exp. Method: NMR 20 Structures
Viral Protein

Compound: Polymer: 1 Molecule: DDap: (ACE)IWGDSGKLI(DNP)TTA ANALOGUE OF HIV GP41 Chains: A Other Details: ATOM OD2

RCSB PDB PROTEIN DATA BANK

A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

As of Tuesday May 27, 2008 there are 51079 Structures | PDB Statistics

CONTACT US | HELP | PRINT PAGE PDB ID or keyword Author Site Search | Advanced Search

Home | Search | Structure | Results | Queries

Are you missing data updates? The PDB archive has moved to ftp://ftp.wwpdb.org. For more information click here.

Help | Structure Summary | Biology & Chemistry | Materials & Methods | Sequence Details | Geometry

- 2DIT
- Download Files
- FASTA Sequence
- Download Original Files
- Display Files
- Display Molecule
  - Image Gallery
  - KiNG Viewer
  - Jmol Viewer
  - WebMol Viewer
  - Protein Workshop
  - FirstGlance
  - Rasmol Viewer (Plugin required)
  - Swiss-PDB Viewer (Plugin required)
  - Molecular Viewers Help
  - KiNG Help
  - Jmol Help
  - WebMol Help
  - Protein Workshop Help
  - QuickPDB
  - Asymmetric Unit
  - Assumed Biological Molecule 1
- Structural Reports
- External Links
- Structure Analysis
- Help

Quick Tips :

When exploring a

2dit

DOI 10.2210/pdb2dit/pdb

Red - Derived Information

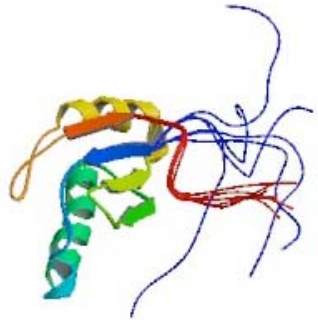
Title	Solution structure of the RRM_1 domain of HIV TAT specific factor 1 variant
Authors	**Dang, W., Muto, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Yokoyama, S., RIKEN Structural Genomics/Proteomics Initiative (RSGI)**
Primary Citation	**Dang, W., Muto, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Yokoyama, S.** Solution structure of the RRM_1 domain of HIV TAT specific factor 1 variant *To be Published*
History	Deposition 2006-03-30 Release 2006-09-30
Experimental Method	Type NMR, 20 STRUCTURES Data N/A
NMR Ensemble	Conformers Calculated 100 Conformers Submitted 20 Selection Criteria target function, structures with the lowest energy, structures with the least restraint violations
NMR Refine	Method NMR, 20 STRUCTURES
Molecular Description Asymmetric Unit	Polymer: 1 Molecule: HIV TAT specific factor 1 variant Fragment: RRM_1 domain Chains: A
Classification	**RNA Binding Protein**
Source	Polymer: 1 Scientific Name: **Homo sapiens** Common Name: **Human** Expression system: **Cell free synthesis**

GO Terms

Polymer	Molecular Function	Biological Process	Cellular Component
HIV TAT specific factor 1 variant (2DIT:A)	• none	• none	• none

Images and Visualization

Asymmetric Unit



Display Options

KiNG
Jmol
WebMol
MBT Protein Workshop
QuickPDB
All Images

RCSB PDB PROTEIN DATA BANK

A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

As of Tuesday May 27, 2008 there are 51079 Structures | PDB Statistics

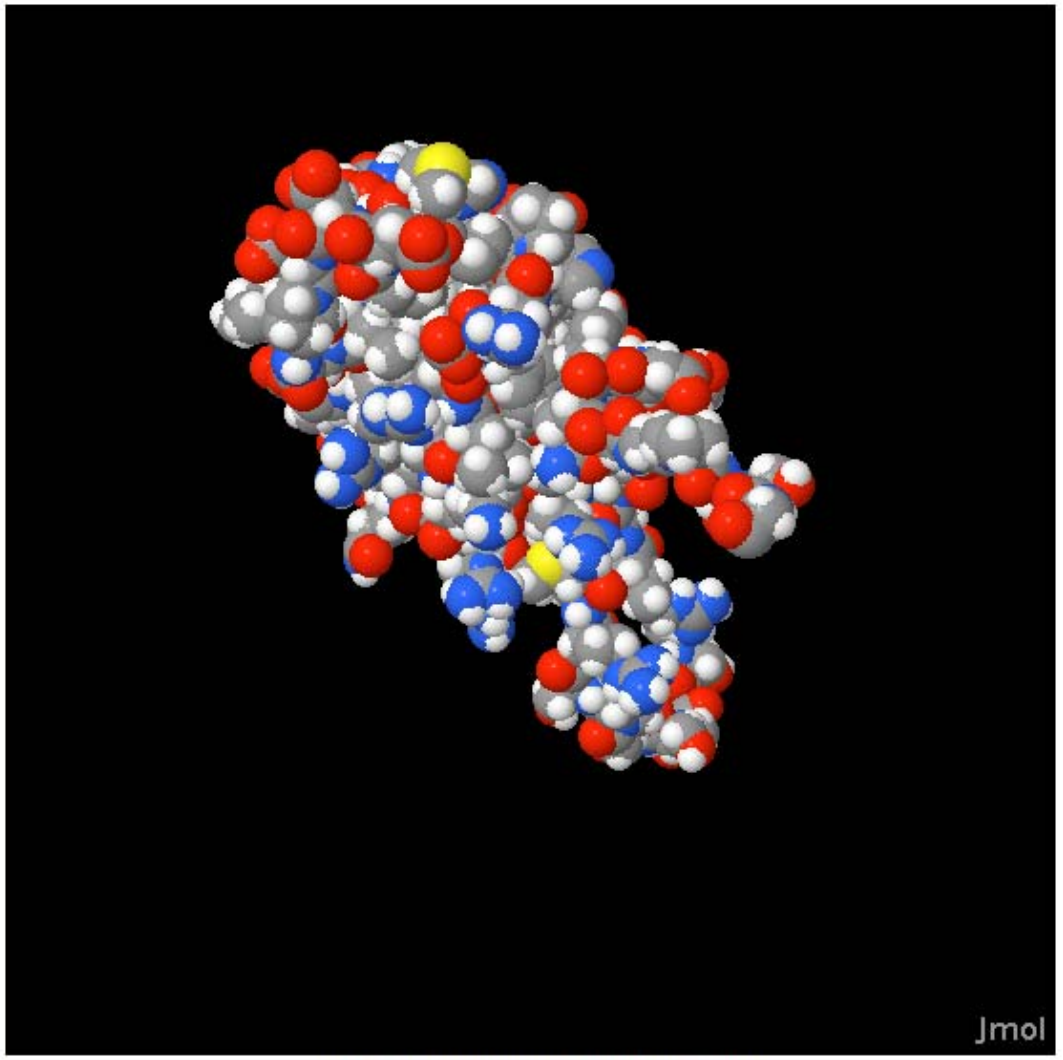
CONTACT US | HELP | PRINT PAGE | PDB ID or keyword | Author | Site Search | Advanced Search

Home | Search | Structure | Results | Queries

Are you missing data updates? The PDB archive has moved to ftp://ftp.wwpdb.org. For more information click here.

- 2DIT
- **Download Files**
- FASTA Sequence
- **Download Original Files**
- **Display Files**
- **Display Molecule**
  - Image Gallery
  - KiNG Viewer
  - **Jmol Viewer**
  - WebMol Viewer
  - Protein Workshop
  - FirstGlance
  - Rasmol Viewer **(Plugin required)**
  - Swiss-PDB Viewer **(Plugin required)**
  - Molecular Viewers Help
  - KiNG Help
  - Jmol Help
  - WebMol Help
  - Protein Workshop Help
  - QuickPDB
  - Asymmetric Unit
  - Assumed Biological Molecule 1
- **Structural Reports**
- **External Links**
- **Structure Analysis**
- **Help**

Jmol 2dit



For help select one of the options below:

Databases of Model Organisms

- Model organisms are those whose genome has been extensively studied such that the workings of biological phenomena for more complex organisms can be inferred.
- For example, the mouse has been most studied to understand other mammalian systems.
- For plant species, *Arabidopsis thaliana* is most often used as a model organism.

Databases of model organisms

- *Arabidobsis* (mustard plant)
  - TAIR: The Arabidopsis Information Resource
    - http://www.arabidopsis.org/
  - The Carnegie Institution of Washington, the National Center for Genome Resources
- *C. elegans*
  - WormBase
    - http://www.wormbase.org/
  - Cold Spring Harbor Laboratory
- *Dictyostelium* (slime mold)
  - dictyBase
    - http://dictybase.org/
  - Northwestern University

Databases of model organisms

- *Drosophila* (Fruit fly)
  - FlyBase
    - http://www.flybase.org/
  - Indiana University
- Mouse
  - Mouse Genome Informatics (MGI)
    - http://www.informatics.jax.org/
  - the Jackson Laboratory
- Rat
  - Rat Genome Database (RGD)
    - http://rgd.mcw.edu
  - The Medical College of Wisconsin

Databases of model organisms

- *Saccharomyces* (Yeast)
  - *Saccharomyces* Genome Database (SGD)
    - http://www.yeastgenome.org/
  - Stanford University
- *Xenopus* (African clawed frog)
  - Xenbase
    - http://www.xenbase.org/
  - The University of Calgary, Alberta
- Zebrafish
  - ZFIN
    - http://zfin.org/
  - The University of Oregon

Types of databases

- Data resources of multiple types of data
  - EBI (European Bioinformatics Institute)
  - NCBI (National Center for Biotechnology Information)
  - KEGG (Kyoto Encyclopedia of Genes and Genomes)
- Gene and protein information
  - GenBank, UniProt, and PDB
  - Species specific: FlyBase, dictyBase, etc.
- Ontological data
  - Gene Ontology
- Pathway data
  - KEGG PATHWAY, Reactome, BRENDA, etc.
- Protein-protein interaction data
  - IntAct, BioGRID, etc.

Break time!

The Gene Ontology

- The Gene Ontology (GO) project is a collaborative effort to address the need for consistent descriptions of gene products in different databases.
- http://www.geneontology.org/
- In annotating gene and protein function, scientists from different backgrounds tend to use different terminology to describe the same event.
  - Eg. Looking for all the gene products that are involved in bacterial protein synthesis, if one database describes these molecules as being involved in 'translation', whereas another uses the phrase 'protein synthesis', it will be difficult for you - and even harder for a computer - to find functionally equivalent terms.

The GO Hierarchy

- The GO project has developed three structured controlled vocabularies (ontologies) that describe gene products in terms of their associated *biological processes, cellular components and molecular functions* in a species-independent manner.
- There are three separate aspects to this effort:

1. the development and maintenance of the ontologies themselves
2. the annotation of gene products, which entails making associations between the ontologies and the genes and gene products in the collaborating databases
3. the development of tools that facilitate the creation, maintenance and use of ontologies.

Terms in the GO

- GO entries, or terms, are given IDs in the form GO:nnnnnnn and a term name, such as “signal transduction”
- Each term is assigned to one of the three ontologies (molecular function, cellular component, or biological process)

Term-Term relationships

- GO terms can be linked by five types of relationships: is_a, part_of, regulates, positively_regulates and negatively_regulates.
- A is_a B: A is a subclass of B
  - E.g. nuclear chromosome is_a chromosome

```
GO:0043232 : intracellular non-membrane-bound organelle
[i] GO:0005694 : chromosome
---[i] GO:0000228 : nuclear chromosome
```

Term-Term relationships

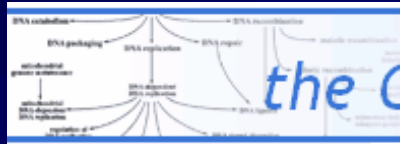
- C part_of D: C is always a part of D if C is present
  - E.g. periplasmic flagellum part_of periplasmic space

```
GO:0044464 : cell part
[i] GO:0042995 : cell projection
---[i] GO:0019861 : flagellum
------[i] GO:0009288 : flagellin-based flagellum
---------[i] GO:0055040 : periplasmic flagellum
[i] GO:0042597 : periplasmic space
---[p] GO:0055040 : periplasmic flagellum
```

Term-Term relationships

- The regulates, positively_regulates and negatively_regulates relationships describe interactions between biological processes and other biological processes, molecular functions or biological qualities.
- E regulates F: E modulates the occurrence of F. If F is a biological quality, then E modulates the value of F.
  - The term regulation of transcription. When regulation of transcription occurs, it always alters the rate, extent or frequency at which a gene is transcribed.

```
GO:0009987 : cellular process
[i] GO:0010467 : gene expression
---[r] GO:0010468 : regulation of gene expression
------[i] GO:0045449 : regulation of transcription
---[p] GO:0006350 : transcription
------[r] GO:0045449 : regulation of transcription
```



the Gene Ontology

Search

gene or protein name

go!

- Open menus
- Home
- FAQ
- **Downloads**
- **Tools**
- **Documentation**
- **About GO**
- **Projects**
- Contact GO
- Site Map

Gene Ontology Home

The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism. *Read more about the Gene Ontology...*

Search the Gene Ontology Database

Search for genes, proteins or GO terms using AmiGO:

lectin GO!

gene or protein name GO term or ID

AmiGO is the official GO browser and search engine. Browse the Gene Ontology with AmiGO.

GO website

- The latest news and views in the GO newsletter
- GO downloads, including ontology files, annotations and the GO database
- Tools for using GO, including OBO-Edit downloads, AmiGO, and the GO Online SQL Environment.
- Request new terms or ontology changes or get help with new term submission
- Documentation on all aspects of the GO project and the GO FAQ
- Projects within the GO consortium, including Reference Genomes and immune system annotation
- Gene Ontology mailing lists and contact details

14 results for lectin in terms fields **term accession, term name, synonyms**

Filter search results

Ontology

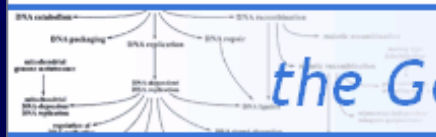
All
biological process
cellular component
molecular function

Set filters

Remove all filters

Select all | Clear all | Perform an action with the selected terms... | Go!

rel	Accession , Term		Ontology
☐	GO:0001862 : collectin binding [show def]	0 gene products view in tree	molecular function
☐	GO:0043698 : iridosome [show def] Query matches synonym "reflectin g platelet" [exact synonym]	0 gene products view in tree	cellular component
☐	GO:0030246 : carbohydrate binding [show def] Query matches synonym "selectin" [related synonym]	692 gene products view in tree	molecular function
☐	GO:0001863 : collectin receptor activity [show def]	0 gene products view in tree	molecular function
☐	GO:0046703 : natural killer cell lectin-like receptor binding [show def] Query matches synonym "NK cell lectin-like receptor binding" [exact synonym]	10 gene products view in tree	molecular function
☐	GO:0001867 : complement activation, lectin pathway [show def]	16 gene products view in tree	biological process
☐	GO:0001868 : regulation of complement activation, lectin pathway [show def]	8 gene products view in tree	biological process
☐	GO:0002772 : inhibitory C-type lectin receptor signaling pathway [show def]	0 gene products view in tree	biological process
☐	GO:0002223 : stimulatory C-type lectin receptor signaling pathway [show def]	4 gene products	biological process



the Gene Ontology

AmiGO

Search Browse BLAST Term Enrichment GO Slimmer GOOSE Help

Search GO [ ] terms genes or proteins exact match 実行

Tree Browser

Filter tree view

Filter by ontology

Ontology: All / biological process / cellular component / molecular function

Filter Gene Product Counts

Data source: All / CGD / dictyBase / FlyBase

Species: All / Anaplasma phagocy... / Arabidopsis thaliana / Bacillus anthraci...

View Options

Tree view: Full Compact

Set filters

Remove all filters

- all : all [250413 gene products]
  - GO:0003674 : molecular_function [168550 gene products]
    - GO:0005488 : binding [46693 gene products]
      - GO:0030246 : carbohydrate binding [692 gene products]

Actions...
Last action: Reset the tree
Graphical View
Permalink
Download...
OBO
RDF-XML
GraphViz dot

GO database release 2008-06-01

Cite this data • Terms of use • GO helpdesk

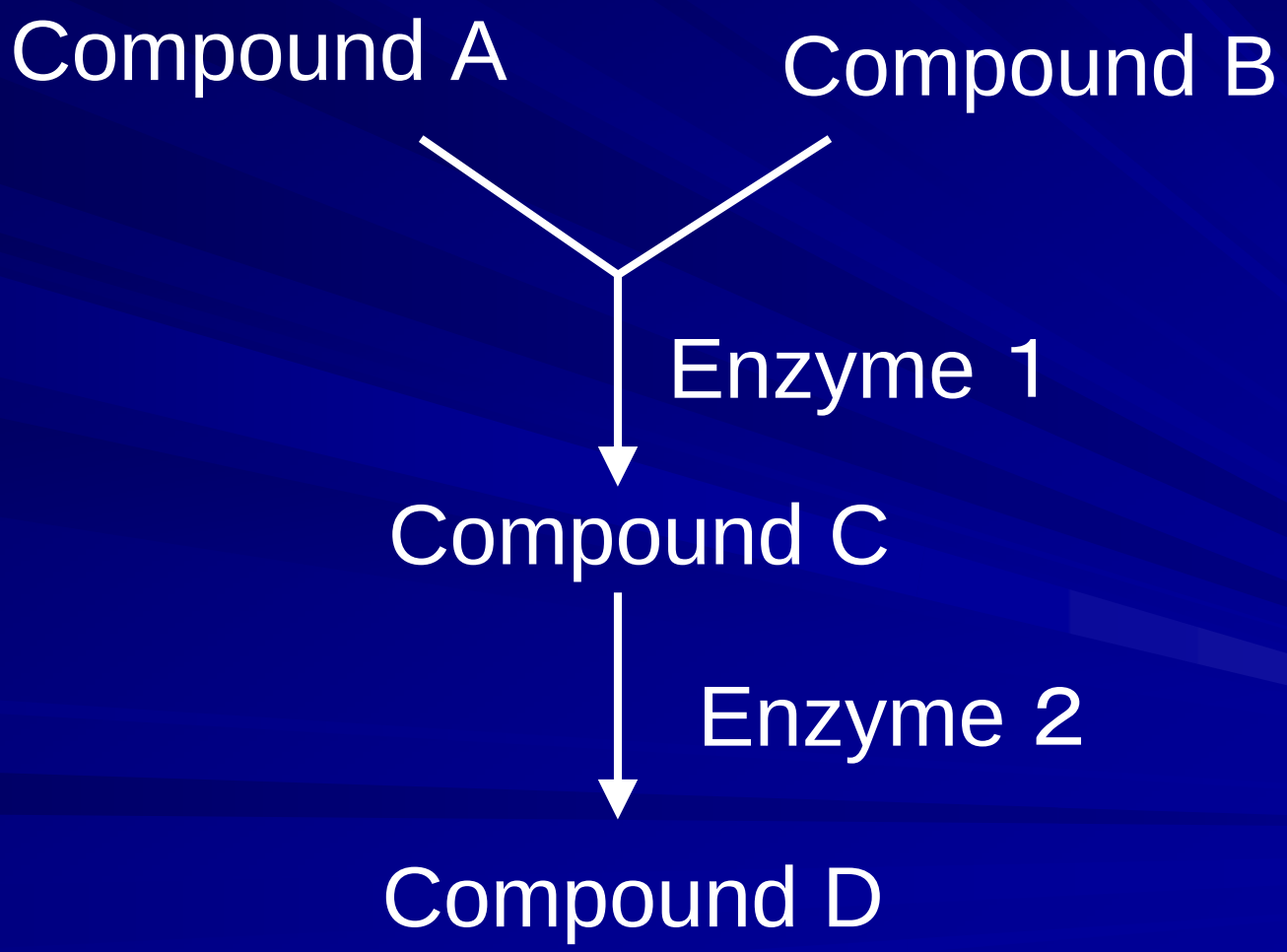
Types of databases

- Data resources of multiple types of data
  - EBI (European Bioinformatics Institute)
  - NCBI (National Center for Biotechnology Information)
  - KEGG (Kyoto Encyclopedia of Genes and Genomes)
- Gene and protein information
  - GenBank, UniProt, and PDB
  - Species specific: FlyBase, dictyBase, etc.
- Ontological data
  - Gene Ontology
- Pathway data
  - KEGG PATHWAY, Reactome, BRENDA, etc.
- Protein-protein interaction data
  - IntAct, BioGRID, etc.

What are pathways?

- Within the body, pathways consist of signal transduction, metabolism and transcription regulation, where information of biological entities interact with one another, forming a map.
- Since the genomic era began, an abundant amount of such pathway information has accumulated, and pathway analysis has become important for the understanding of biological systems.
  - For example, given a particular protein known to express in a certain disease, we want to find the other proteins with which it interacts.
  - As another example, we can build a network of multiple proteins known to interact with one another to glean insights into how it affects the system.

A simple pathway



Pathway Databases

- KEGG
  - Bioinformatics Center, Kyoto University
  - http://www.genome.jp
- BRENDA
  - Braunschweig Dept. of Bioinformatics
  - http://www.biocyc.org
- Reactome
  - EBI (European Bioinformatics Institute), Cold Spring Harbor Laboratory and the Gene Ontology Consortium
  - http://www.reactome.org

BRENDA

- Database of enzyme information, linked with KEGG pathways
- Data on enzyme function are extracted directly from the primary literature by scientists holding a degree in Biology or Chemistry.
- Formal and consistency checks are done by computer programs, each data set on a classified enzyme is checked manually by at least one biologist and one chemist.

BRENDA home
login
history
All enzymes

BRENDA

The Comprehensive Enzyme Information System

TU Braunschweig Dept. of Bioinformatics

SEARCH-Navigator
close all open all
Nomenclature
Reaction & Specificity
Functional Parameters
Organism related Information
Enzyme Structure
Isolation & Preparation
Stability
Disease & References
Application & Engineering

Quick search
Fulltext search
Advanced search
Substructure search
TaxTree Explorer
EC Explorer
Sequence Search
Genome Explorer
Ontology Explorer
Download

Introduction/References
News
Contact/Team/Errors
Jobs
Copyright
Related Links
Help

EC-Number | Enzyme Name | Organism | Protein | Full text | Advanced Search

Search Display 10 entries

Latest BRENDA update 12/2007

Nomenclature	Reaction & Specificity	Functional Parameters
Enzyme Names EC Number Common/ Recommended Name Systematic Name Synonyms CAS Registry Number	Pathway Catalysed Reaction Reaction Type Natural Substrates and Products Substrates and Products Substrates Natural Substrate Products Natural Product Inhibitors Cofactors Metals/Ions Activating Compounds Ligands	Km Value Ki Value pI Value Turnover Number Specific Activity pH Optimum pH Range Temperature Optimum Temperature Range
Isolation & Preparation		**Organism-related information**
Purification Cloned Renatured Crystallization		Organism Source Tissue Localization Protein-Specific Search
Stability	**Enzyme Structure**	**Disease & References**
pH Stability Temperature Stability General Stability Organic Solvent Stability Oxidation Stability Storage Stability	Sequence/ SwissProt link 3D-Structure/ PDB link Molecular Weight Subunits Posttranslational Modification	Disease References
		Application & Engineering
		Engineering Application

Webmaster: **Maurice Scheer**
m.scheer@tu-bs.de

BRENDA

The Comprehensive Enzyme Information System

TU Braunschweig
Dept. of Bioinformatics

SEARCH-Navigator
close all open all
Nomenclature
Reaction & Specificity
Functional Parameters
Organism related Information
Enzyme Structure
Isolation & Preparation
Stability
Disease & References
Application & Engineering

Quick search
Fulltext search
Advanced search
Substructure search
TaxTree Explorer
EC Explorer
Sequence Search
Genome Explorer
Ontology Explorer
Download

Introduction/References
News
Contact/Team/Errors
Jobs
Copyright
Related Links
Help

Any question? -> Use the BRENDA Discussion groups

Search Pathway

citrate cycle | contains | Search | show 10 results | clear

Recommended Name: | contains
EC Number: | contains
KEGG Link: | contains

:= amino acid sequences := comprehensive online version := show the catalyzed reaction

Results **1** - **10** of **23**

download this result as tab stop separated values (Excel,OpenOffice) format

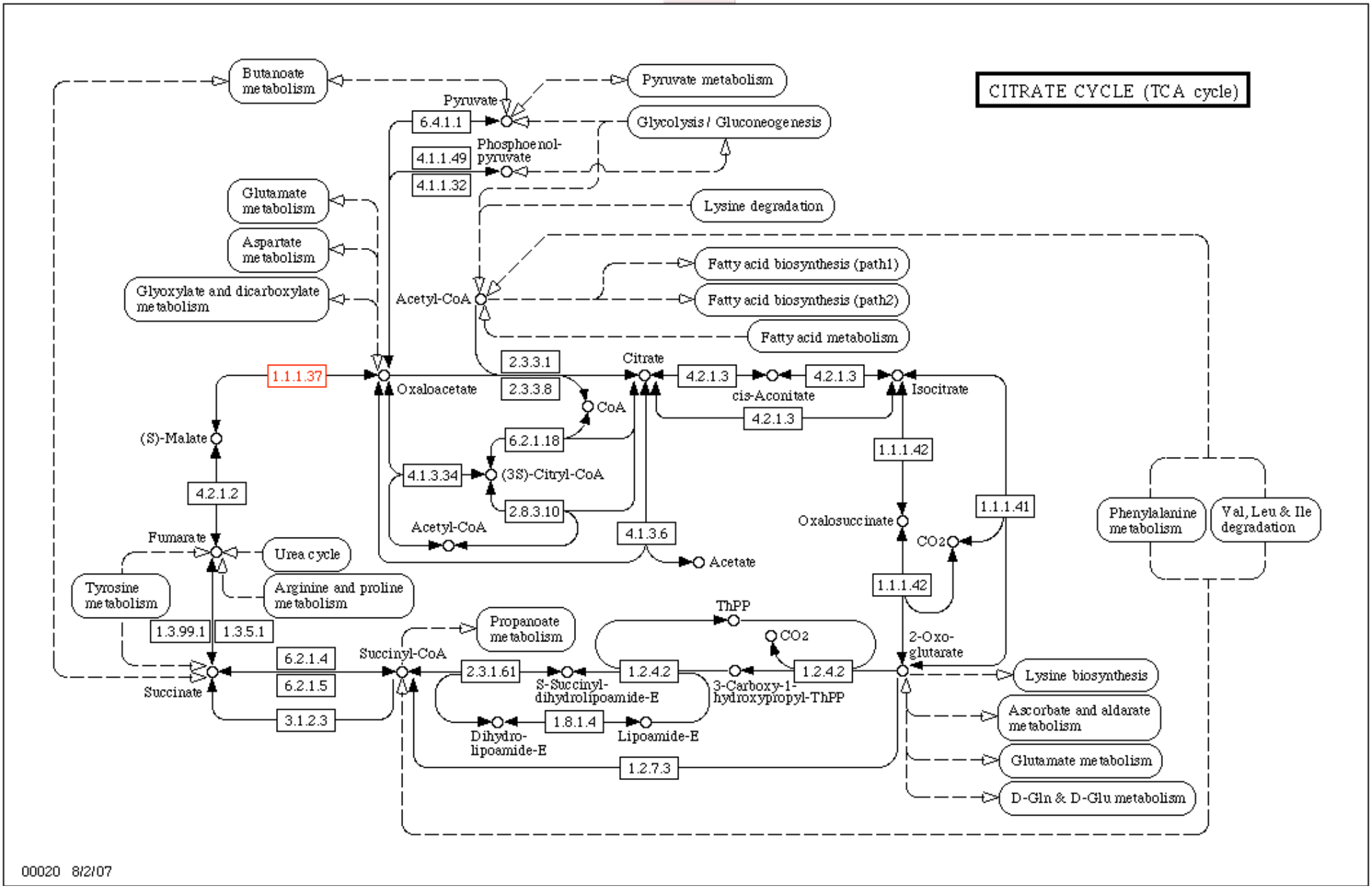
EC Number	Recommended Name	Pathway	KEGG Link
1.1.1.37	malate dehydrogenase	Citrate cycle (TCA cycle)	✓
1.1.1.41	isocitrate dehydrogenase (NAD+)	Citrate cycle (TCA cycle)	✓
1.1.1.42	isocitrate dehydrogenase (NADP+)	Citrate cycle (TCA cycle)	✓
1.2.4.2	oxoglutarate dehydrogenase (succinyl-transferring)	Citrate cycle (TCA cycle)	✓
1.2.7.3	2-oxoglutarate synthase	Citrate cycle (TCA cycle)	✓
1.3.5.1	succinate dehydrogenase (ubiquinone)	Citrate cycle (TCA cycle)	✓
1.3.99.1	succinate dehydrogenase	Citrate cycle (TCA cycle)	✓
1.8.1.4	dihydrolipoyl dehydrogenase	Citrate cycle (TCA cycle)	✓
2.3.1.61	dihydrolipoyllysine-residue succinyltransferase	Citrate cycle (TCA cycle)	✓
2.3.3.1	citrate (Si)-synthase	Citrate cycle (TCA cycle)	✓

KEGG **Citrate cycle (TCA cycle) - Reference pathway**

Help

[Pathway menu | Ortholog table]

Reference pathway | Go | Current selection | Select



00020 8/2/07

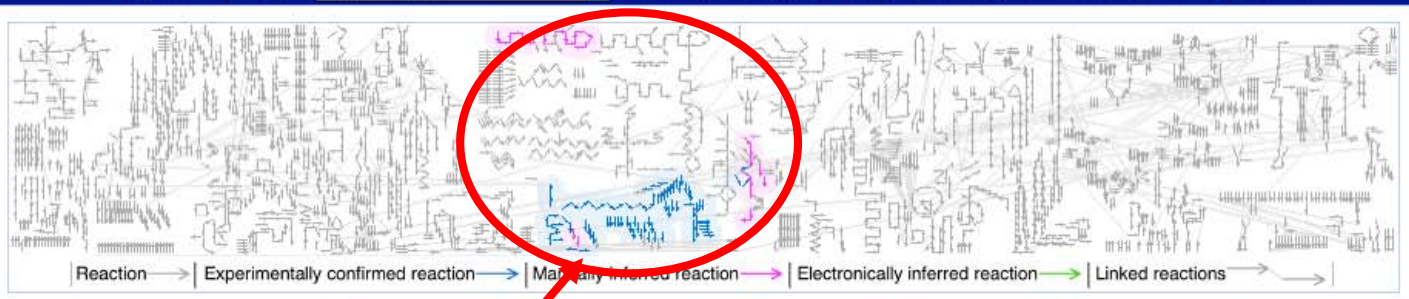
Reactome

- A curated resource of core pathways and reactions in human biology.
- The information in this database is authored by biological researchers with expertise in their fields, and cross-referenced with NCBI, UniProt, KEGG (Gene and Compound), PubMed, GO and others.
- In addition to curated human events, inferred orthologous events in 22 non-human species including mouse, rat, chicken, puffer fish, worm, fly, yeast, and *E.coli* are also available.

Find everything with ALL of the words in Homo sapiens Go!

Reactome - a curated knowledgebase of biological pathways

The data displayed is for Homo sapiens. Use the menu to change the species. Check for cross-species comparison.



Apoptosis	Botulinum neurotoxicity	Cell Cycle Checkpoints	Cell Cycle, Mitotic
DNA Repair	DNA Replication	Electron Transport Chain	Gap junction trafficking and regulation
Gene Expression	HIV Infection	Hemostasis	Influenza Infection
Integration of energy metabolism	Lipid and lipoprotein metabolism	Membrane Trafficking	Metabolism of amino acids
Metabolism of carbohydrates	Metabolism of nitric oxide	Metabolism of non-coding RNA	Metabolism of vitamins and cofactors
Metabolism of xenobiotics	Nucleotide metabolism	Porphyrin metabolism	Pyruvate metabolism and TCA cycle
Post-translational protein modification	Regulatory RNA pathways	Signaling by BMP	Signaling by EGFR
Signaling by FGFR	Signaling in Immune system	Signaling by Insulin receptor	Signalling by NGF
Signaling by Notch	Signaling by Rho GTPases	Signaling by TGF beta	Signaling by VEGF
Signaling by Wnt	Telomere Maintenance	Transcription	Translation
mRNA Processing			

Reactome is seeking expert help for the curation of new modules

The Reactome knowledgebase relies on collaborations with research biologists to construct expert consensus views of key biological processes, and to integrate these with other processes already in Reactome. We are seeking new author-collaborators. If you're interested, or would like more information about our data acquisition process, please contact us at editorial@reactome.org. Click here to view/hide a list of high-priority projects now being developed.

About Reactome

The **Reactome** project is a collaboration among Cold Spring Harbor Laboratory, The European Bioinformatics Institute, and The Gene Ontology Consortium to develop a curated resource of core pathways and reactions in human biology. The information in this database is authored by biological researchers with expertise in their fields, maintained by the Reactome editorial staff, and cross-referenced with the sequence databases at NCBI, Ensembl and UniProt, the UCSC Genome Browser , HapMap, KEGG(Gene and Compound), ChEBI, PubMed and GO. In addition to curated human events, inferred orthologous events in 22 non-human species including mouse, rat, chicken, puffer fish, worm, fly, yeast, two plants and E.coli are also available. A description of Reactome has been published in **Genome Biology**.

Reactome is a free on-line resource, and Reactome software is open-source. However, please take note of our disclaimer.

News and Notes

- **March 12, 2008 Version 24 Released**
  New additions to curated content include the pathway topics: Signaling by VEGF, Metabolism of nitric oxide, and MicroRNA biogenesis. The pathway topics updated with new curated events include: Apoptosis - **Cleavage reactions**, Gene expression - **Generic transcription pathway**, Hemostasis - **Tie2 and PECAM events**, Immune signaling -**TCR cascade**, Energy metabolism - **PDC regulation**, Membrane Trafficking - **COP II mediated events**, Post translational modifications - **Hypusine synthesis**, EGFR signaling - **PLC-gamma and Grb2 mediated downstream events**.
  K Schulze-Osthoff, and L Castognoli are our external expert annotators, and S Ranganathan, LP Freedman, J Trowsdale, J Lippincott-Schwartz, G Enikolopov, F Karginov, GJ Hannon, CH Heldin, and L Claesson-Welsh are our external reviewers, in this release.
  As usual, protein-protein interaction datasets, Statistics and the Editorial Calendar are available. Reactome team thanks the users for their comments on the pathway visualization tool.
  Click on contact to reach us, on editorial to contribute to Reactome content, and on subscribe to receive Reactome announcements.
- More...

The development of Reactome is supported by a grant from the US National Institutes of Health (P41 HG003751), a grant from the European Union Sixth Framework Programme (LSHG-CT-2003-503269), and a subcontract from the EBI Industry Programme. Information on topic-specific funding can be found here.

HIV Infection [Homo sapiens]

Reactionmap

Details

open to selected event | open all | close all

- HIV Infection [Homo sapiens]
  - HIV Life Cycle
  - Host Interactions of HIV factors
  - Host immune responses to HIV infection

HIV Infection	
Stable identifier	REACT_6185.3
Authored	Bukrinsky, M, D'Eustachio, P, Gillespie, ME, Gopinathrao, G, Iordanskiy, S, Morrow, MP, Matthews, L, Rice, AP

The global pandemic of Human Immunodeficiency Virus (HIV) infection has resulted in tens of millions of people infected by the virus and millions more affected. UNAIDS estimates around 40 million HIV/AIDS patients worldwide with 75% of them living in sub-Saharan Africa. The primary method of HIV infection is by sexual exposure while nonsexual HIV transmission also can occur through transfusion with contaminated blood products, injection drug use, occupational exposure,accidental needlesticks or mother-to-child transmission. HIV damages the immune system, leaving the infected person vulnerable to a variety of "opportunistic" infections arising from host immune impairment (Hare, 2004).
HIV-1 and the less common HIV-2 belong to the family of retroviruses. HIV-1 contains a single-stranded RNA genome that is 9 kilobases in length and contains 9 genes that encode 15 different proteins. These proteins are classified as: structural proteins (Gag, Pol, and Env), regulatory proteins (Tat and Rev), and accessory proteins (Vpu, Vpr, Vif, and Nef) (Frankel and Young,1998).
HIV infection cycle can be divided into two phases:
1. An **Early phase** consisting of early events occuring after HIV infection of a susceptible target cell and a 2. **Late phase** comprising the later events in the HIV-infected cell resulting in the assembly of new infectious virions. The section titled **HIV lifecycle** consists of annotations of events in these two phases.
The virus has developed various molecular strategies to suppress the antiviral immune responses (innate, cellular and humoral) of the host. HIV-1 viral auxiliary proteins (Tat, Rev, Nef, Vif, Vpr and Vpu) play important roles in these host-pathogen interactions (Li et al.,2005). The section titled **Host interactions of HIV factors** will highlight these complex post-infection processes and the annotations will be released in near future.
[Hare 2004, Frankel & Young 1998, Li *et al* 2005]

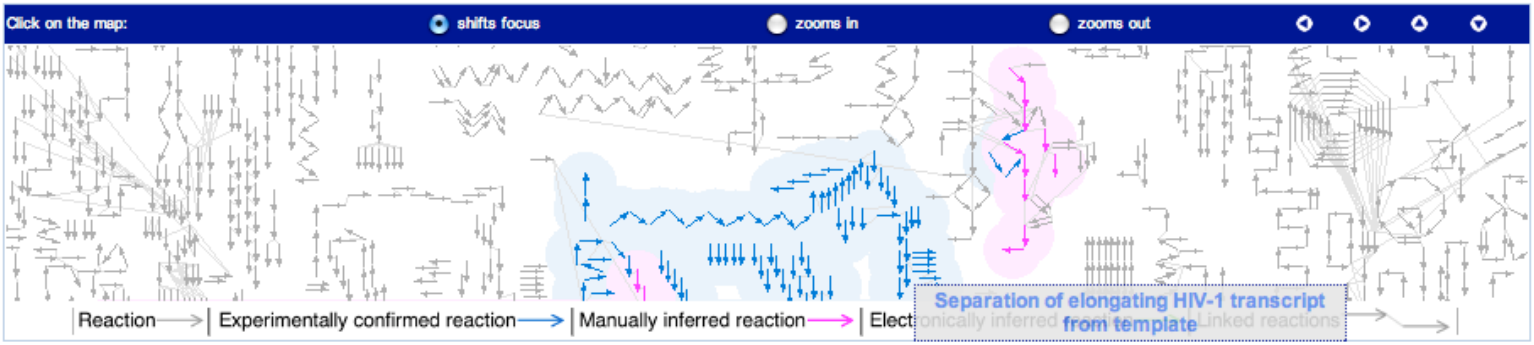
Organism	Homo sapiens Human immunodeficiency virus 1
Cellular compartment	cell GO
Represents GO biological process	viral life cycle GO

Participating molecules

- 26S protease regulatory subunit 4 [cytosol] UKEOGRHUR
- 26S protease regulatory subunit 6A [cytosol] UKEOGRHUR
- 26S protease regulatory subunit 6B [cytosol] UKEOGRHUR
- 26S protease regulatory subunit 7 [cytosol] UKEOGRHUR
- 26S protease regulatory subunit S10B [cytosol] UKEOGRHUR
- 26S proteasome non-ATPase regulatory subunit 1 [cytosol] UKEGRHUR
- 26S proteasome non-ATPase regulatory subunit 10 [cytosol] UKEOGRHUR
- 26S proteasome non-ATPase regulatory subunit 11 [cytosol] UKEOGRHUR
- 26S proteasome non-ATPase regulatory subunit 12 [cytosol] UKEOGRHUR
- 26S proteasome non-ATPase regulatory subunit 13 [cytosol] UKEOGRHUR

HIV Infection [Homo sapiens]

Reactionmap



Details

open to selected event | open all | close all

- HIV Infection [Homo sapiens]
  - HIV Life Cycle
    - Early Phase of HIV Life Cycle
      - Binding and entry of HIV virion
      - Uncoating of the HIV Virion
      - Formation of RTC (Reverse Transcripti
      - Annealing of 3'-end of unwound transfe
      - Reverse Transcription of HIV RNA
      - Removal of plus-strand flap and gap cl
      - Integration of provirus
    - Late Phase of HIV Life Cycle
  - Host Interactions of HIV factors
  - Host immune responses to HIV infection

HIV Infection

Stable identifier	REACT_6185.3
Authored	Bukrinsky, M, D'Eustachio, P, Gillespie, ME, Gopinathrao, G, Iordanskiy, S, Morrow, MP, Matthews, L, Rice, AP

The global pandemic of Human Immunodeficiency Virus (HIV) infection has resulted in tens of millions of people infected by the virus and millions more affected. UNAIDS estimates around 40 million HIV/AIDS patients worldwide with 75% of them living in sub-Saharan Africa. The primary method of HIV infection is by sexual exposure while nonsexual HIV transmission also can occur through transfusion with contaminated blood products, injection drug use, occupational exposure,accidental needlesticks or mother-to-child transmission. HIV damages the immune system, leaving the infected person vulnerable to a variety of "opportunistic" infections arising from host immune impairment (Hare, 2004).
HIV-1 and the less common HIV-2 belong to the family of retroviruses. HIV-1 contains a single-stranded RNA genome that is 9 kilobases in length and contains 9 genes that encode 15 different proteins. These proteins are classified as: structural proteins (Gag, Pol, and Env), regulatory proteins (Tat and Rev), and accessory proteins (Vpu, Vpr, Vif, and Nef) (Frankel and Young,1998).
HIV infection cycle can be divided into two phases:
1. An **Early phase** consisting of early events occuring after HIV infection of a susceptible target cell and a
2. **Late phase** comprising the later events in the HIV-infected cell resulting in the assembly of new infectious virions. The section titled **HIV lifecycle** consists of annotations of events in these two phases.
The virus has developed various molecular strategies to suppress the antiviral immune responses (innate, cellular and humoral) of the host. HIV-1 viral auxiliary proteins (Tat, Rev, Nef, Vif, Vpr and Vpu) play important roles in these host-pathogen interactions (Li et al.,2005). The section titled **Host interactions of HIV factors** will highlight these complex post-infection processes and the annotations will be released in near future.
[Hare 2004, Frankel & Young 1998, Li *et al* 2005]

Organism	Homo sapiens Human immunodeficiency virus 1
Cellular compartment	cell GO
Represents GO biological process	viral life cycle GO

Participating molecules

- 26S protease regulatory subunit 4 [cytosol] UKEOGRHUR
- 26S protease regulatory subunit 6A [cytosol] UKEOGRHUR
- 26S protease regulatory subunit 6B [cytosol] UKEOGRHUR
- 26S protease regulatory subunit 7 [cytosol] UKEOGRHUR
- 26S protease regulatory subunit S10B [cytosol] UKEOGRHUR
- 26S proteasome non-ATPase regulatory subunit 1 [cytosol] UKEGRHUR
- 26S proteasome non-ATPase regulatory subunit 10 [cytosol] UKEOGRHUR
- 26S proteasome non-ATPase regulatory subunit 11 [cytosol] UKEOGRHUR
- 26S proteasome non-ATPase regulatory subunit 12 [cytosol] UKEOGRHUR
- 26S proteasome non-ATPase regulatory subunit 13 [cytosol] UKEOGRHUR
- ...

List all 1191 participating molecules

GO:0016032 viral reproduction

Back

The process by which a virus reproduces. Usually, this is by infection of a host cell, replication of the viral genome, and assembly of progeny virus particles. In some cases the viral genetic material may integrate into the host genome and only subsequently, under particular circumstances, 'complete' its life cycle.

Term Information | Ancestor chart | Ancestor table | Child Terms | Protein Annotation | Statistics

ID	GO:0016032
Name	viral reproduction
Definition	The process by which a virus reproduces. Usually, this is by infection of a host cell, replication of the viral genome, and assembly of progeny virus particles. In some cases the viral genetic material may integrate into the host genome and only subsequently, under particular circumstances, 'complete' its life cycle.
Comment	See also the biological process terms 'viral infectious cycle ; GO:0019058' and 'lysogeny ; GO:0030069'.

Synonyms

Type	Synonym
related	viral infection
related	virulence
exact	viral replication cycle
exact	viral life cycle

Types of databases

- Data resources of multiple types of data
  - EBI (European Bioinformatics Institute)
  - NCBI (National Center for Biotechnology Information)
  - KEGG (Kyoto Encyclopedia of Genes and Genomes)
- Gene and protein information
  - GenBank, UniProt, and PDB
  - Species specific: FlyBase, dictyBase, etc.
- Ontological data
  - Gene Ontology
- Pathway data
  - KEGG PATHWAY, Reactome, BRENDA, etc.
- Protein-protein interaction data
  - IntAct, BioGRID, etc.

Protein-protein interactions

- Pathways are a type of protein-protein interaction (PPI)
- High-throughput methods of determining PPI has produced a large amount of data
  - Two-hybrid
  - Microarray gene expression

Protein-protein interaction (PPI) data

- IntAct
  - Proteomic database provided by the EBI
  - http://www.ebi.ac.uk/intact/
  - Interaction data is derived from the literature or via direct user submissions
  - Interaction data can be searched, analyzed, and downloaded

IntAct

Search IntAct
Search Clear

- IntAct Home
- Advanced Search
- Tools
- Data Submission
- Downloads
- Documentation
  - FAQ
  - User manual
  - Annotation manual
  - Publications
  - Statistics
- Developer Resources
  - Development Site
- Contact IntAct

- Printer Friendly View

News RSS

2008/02/27
Dataset of the month

You can read more about this dataset in the Sanger Press Release.

2007/12/11
Export of protein

IntAct Home

Search IntAct

To perform a search in the IntAct da
Examples:

- Gene name: BRCA2
- UniProtKB Ac: Q06609
- UniProtKB Id: dmc1
- Pubmed Id: 10831611

Introduction

IntAct provides a freely available, op
for protein interaction data. All intera
direct user submissions and are fr

Dataset of the month: May

- **Protein-protein interactions**
  Myxococcus xanthus.

  Whitworth et al. IntAct

Go to Archive.

License

All software, available on this site, is
Commons Attribution License. This
use of all records from the IntAct dat

Acknowledgements

IntAct is fu
021902 (R

Database Search

Search Results for {ac='%HIV%';shortLabel='%HIV%';description='%HIV%';xref='%HIV%';disjunction='true'},
(short labels of search criteria matches are highlighted)

Please click on any name to view more detail.

Graph | Path | InterPro | Select all | Clear all

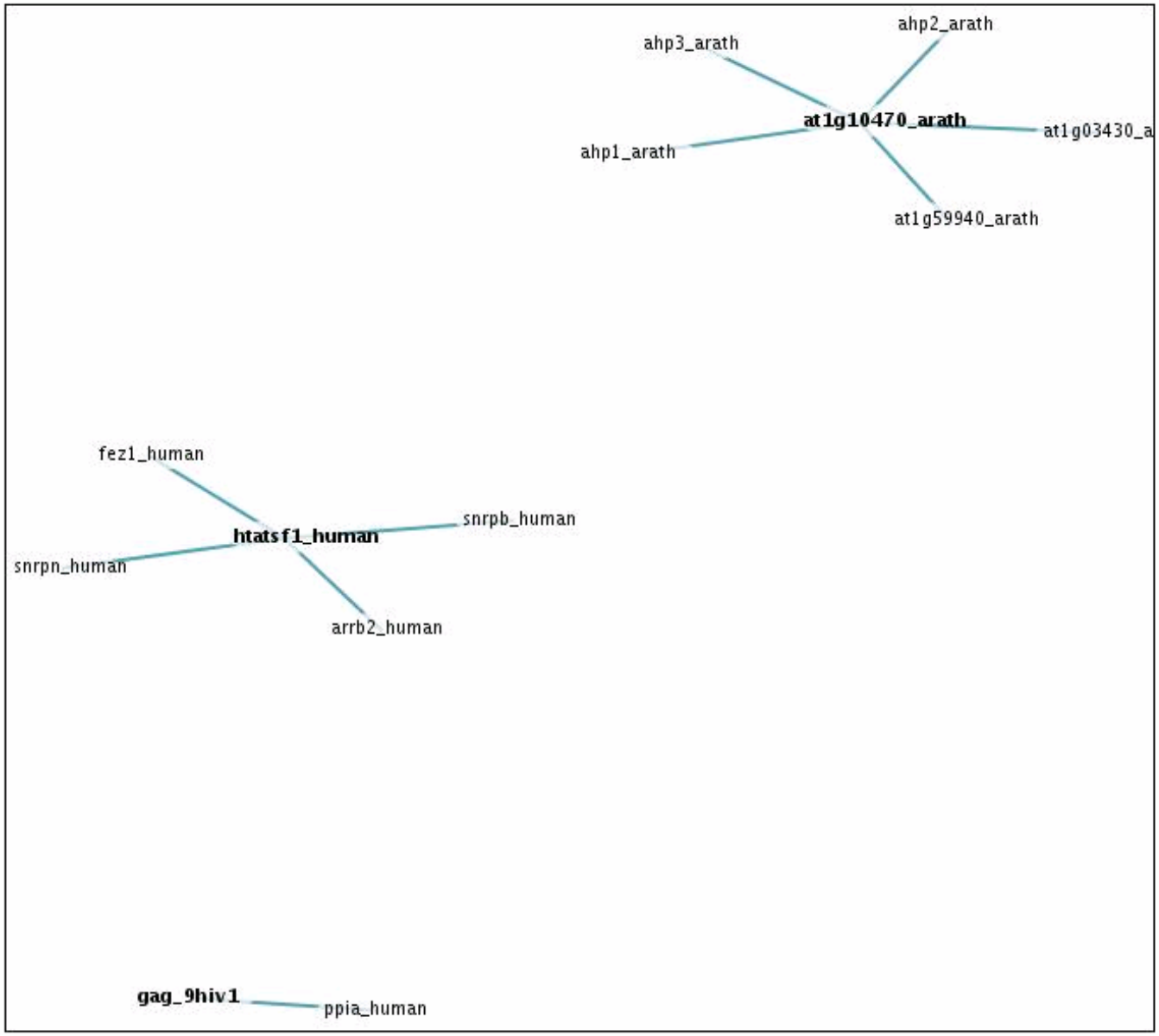
Proteins [2]	Name	Ac	Gene-Name	Description	Interactions	Interactors
☐	q72497_9hiv1	EBI-1036263	gag	Gag polyprotein	6	2
☐	q2hiv3_arath	EBI-1100868	At1g10470	At1g10470	7	5
☐	q97429_9hiv1	EBI-1183896	pol	Reverse transcriptase	0	0
☐	q97705_9hiv1	EBI-1562682	-	Matrix protein P17	0	0
☐	p88053_9hiv1	EBI-352318	env	Envelope glycoprotein, C2-V5 region	0	0
☐	q75888_9hiv1	EBI-519201	gag	Gag protein	0	0
☐	o10752_9hiv1	EBI-538834	env	Envelope glycoprotein	0	0
☐	q97762_9hiv1	EBI-626690	env	Envelope glycoprotein	0	0
☐	q9qjd5_9hiv1	EBI-638685	env	Envelope protein	0	0
☐	htsf1_human	EBI-720468	HTATSF1	HIV Tat-specific factor 1	5	265
☐	q78560_9hiv1	EBI-995253	env	Envelope protein	0	0

Graph | Path | InterPro | Select all | Clear all

HierarchView ?

Query: htatsf1_human, gag_9hiv1, at1g10470_arath

Result: **13** molecules, **10** interactions.



Annotation for molecules in network. Click to highlight molecules. ?

MoleculeType (1) | Confidence (0) | Publication (5)

GO (23) | Interpro (21) | Role (5) | Species (3)

Description	Show	Count	ID
protein binding		10	GO:0005515
RNA splicing		2	GO:0008380
small nuclear ribonucleoprotein complex		2	GO:0030532
nucleus		2	GO:0005634
spliceosome		2	GO:0005681
cytoplasm		2	GO:0005737
transcription elongation regulator activity		1	GO:0003711
nucleic acid binding		1	GO:0003676
plasma membrane		1	GO:0005886
unfolded protein binding		1	GO:0051082
virion binding		1	GO:0046790
identical protein binding		1	GO:0042802
regulation of viral genome replication		1	GO:0045069
structural molecule activity		1	GO:0005198
RNA polymerase II transcription factor activity		1	GO:0003702
cytokinin mediated signaling		1	GO:0009736
cell adhesion		1	GO:0007155
regulation of			

Protein-protein interaction (PPI) data

- BioGRID
  - Biological General Repository for Interaction Datasets
  - Contains protein and genetic interactions from major model organism species
  - http://www.thebiogrid.org/
  - Currently includes a virtually complete set of interactions reported in the literature for *S. cerevisiae* and *S. pombe*.

BioGRID BETA

General Repository for Interaction Datasets

home | support | contribute | downloads | mirrors | about us

Search the BioGRID

Examples: Genbank ID's, Entrez-Gene ID's, SGD ID's, Gene Names [more]

HIV

Organism: All Organisms

Submit Your Search

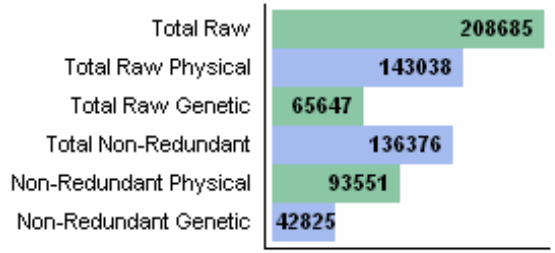
Having Problems Searching?

Download Osprey

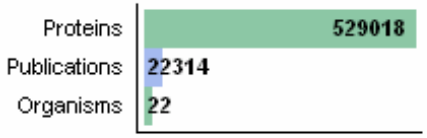
Osprey is a software platform for visualization of complex interaction networks. Osprey builds data-rich graphical representations from Gene Ontology (GO) annotated interaction data maintained by the BioGRID.

http://biodata.mshri.on.ca/osprey

Interaction Statistics



Database Statistics



Latest News

XML

BioGRID version 2.0.40 release (3938 physical and genetic interactions added)

May. 1st, 2008 @ 02:44:20

The BioGRID's curated set of physical and genetic interactions has been updated to include an additional 3938 interactions. These additions bring our total number of non-redundant interactions to 136,376 and raw interactions to 208,685. New interactions will be added in curation updates on a monthly basis. Please let us know if we have missed or incorrectly reported any interactions by sending an e-mail to biogridadmin@gmail.com.

BioGRID version 2.0.39 release (980 physical and genetic interactions added)

Apr. 1st, 2008 @ 03:57:20

The BioGRID's curated set of physical and genetic interactions has been updated to include an additional 980 interactions. These additions bring our total number of non-redundant interactions to 133,595 and raw interactions to 204,747. New interactions will be added in curation updates on a monthly basis. Please let us know if we have missed or incorrectly reported any interactions by sending an e-mail to biogridadmin@gmail.com.

BioGRID version 2.0.38 release (354 physical and genetic interactions added)

Mar. 1st, 2008 @ 02:37:43

Your search for *"HIV"* produced the following 1 results with associations:

Click on your gene of interest for a detailed interaction summary

Homo sapiens (Human)

HIVEP3

Organism: *Homo sapiens*

# of Associations: 3

human immunodeficiency virus type I enhancer-binding protein 3|kappabinding protein-1; human immunodeficiency virus type I enhancer binding protein 3 ***[more]***

Your search for *"HIV"* produced the following 19 results without associations:

Click on your gene of interest for a detailed interaction summary

Canis familiaris (Dog)

HIVEP2

Organism: *Canis familiaris*

# of Associations: 0

human immunodeficiency virus type I enhancer binding protein 2 ***[more]***

Gallus gallus (Chicken)

HIVEP1

Organism: *Gallus gallus*

# of Associations: 0

human immunodeficiency virus type I enhancer binding protein 1; chicken alphaA-CRYBP1 ***[more]***

HIVEP2

Organism: *Gallus gallus*

# of Associations: 0

human immunodeficiency virus type I enhancer binding protein 2 ***[more]***

HIVEP3

Organism: *Gallus gallus*

# of Associations: 0

human immunodeficiency virus type I enhancer binding protein 3 ***[more]***

Homo sapiens (Human)

HIVEP2

Organism: *Homo sapiens*

# of Associations: 0

MHC binding protein-2|OTTHUMP00000040180|Schnurri-2|c-myc intron-binding protein 1|human immunodeficiency virus type I enhancer-binding protein 2; human immunodeficiency virus ***[more]***

[JPG] [PNG] [SVG]

HIVEP3

Homo sapiens

Aliases: KBP-1, KBP1, KIAA1555, KRC, ZAS3, FLJ16752, EG59269

Description: human immunodeficiency virus type I enhancer-binding protein 3|kappabinding protein-1; human immunodeficiency virus type I enhancer binding protein 3

Links: [Entrez-Gene] [OMIM] [HGNC] [HPRD] [Ensembl]

Gene Ontology: 4 total GO mappings for this record. [view list]

Download interactions associated with HIVEP3

HIVEP3 was identified with 5 protein interactions

Name	Aliases	Description	Evidence Code(s)	Role	Source(s)
TRAF2	• *MGC:45012* • *TRAP* • *TRAP3* • *EG7186*	OTTHUMP00000064745|tumor necrosis factor type 2 receptor associated protein 3; TNF receptor-associated factor 2 Gene Ontology: [View List]	Invivo Two-hybrid	hit hit	Oukka M (2002) Oukka M (2002)
TRAF1	• *EBI6* • *MGC:10353* • *EG7185*	Epstein-Bar virus-induced protein 6; TNF receptor-associated factor 1 Gene Ontology: [View List]	Invivo Two-hybrid	hit hit	Oukka M (2002) Oukka M (2002)
TNFRSF1B	• *CD120b* • *TBPII* • *TNF-R-II* • *EG7133*	p75 TNF receptor|tumor necrosis factor beta receptor|tumor necrosis factor binding protein 2|tumor necrosis factor receptor 2; tumor necrosis factor receptor superfamily, member 1B Gene Ontology: [View List]	Invivo	hit	Oukka M (2002)

Summary (1)

- Many major databases, with many useful information
- For a specific purpose, a single database may suffice
- For more comprehensive analyses, computer programs may be developed to access multiple databases and process the data to retrieve the desired information
  - Application Programming Interfaces (APIs), Web services, etc.

Summary (2)

- APIs: programs (libraries) that can be downloaded such that users can develop computer programs that use the provided libraries to retrieve data via the Internet
- Web services: programs that can be executed over the Internet by users' programs to retrieve information
  - Workflows can be developed from these web services, such that the retrieved data from DB 1 can be automatically sent as an input query to DB 2, for example.