

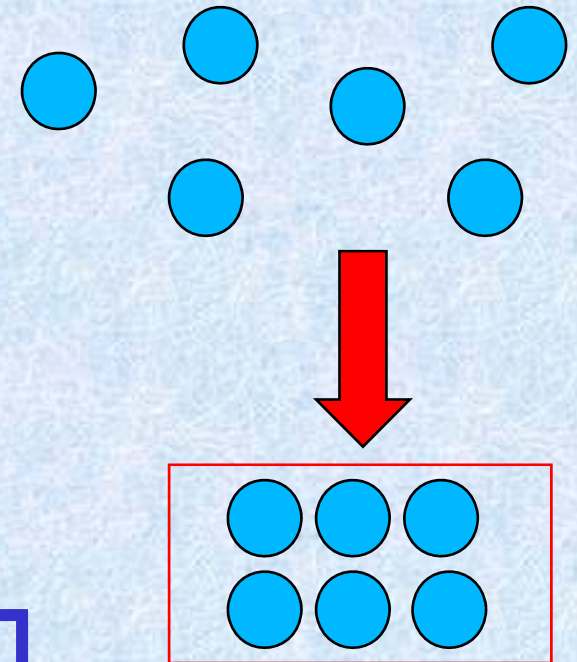
Databases

Biological experiments (**macromolecular sequences, structures, expression profiles, pathways etc**) provide wealth of data.

The data are available randomly in the literature

It is necessary to collect the scattered data and put in proper order in the form of a database

Database is an organized collection of information, in computer-readable form.



Databases: Characteristics

1. The contents
2. The ontology: the list of valid terms and their definitions
3. The logical structure, or the expression of the inter-relationships among the data, called the schema.
4. The format of the data
5. The routes for selective retrieval of data, and presentation of results, or pasting them on to a program for analysis
6. Links to other information resources: other databases, references to original publications of data, tutorial background etc.

Characteristics

NO.	193
***** Sequence and structural information*****	
PROTEIN NAME	Lysozyme
SOURCE	Bacteriophage T4
LENGTH	164
MOLECULAR WEIGHT	18603.94
PIR_ID	LZBPT4
SWISSPROT_ID	LYCV_BPT4 (P00720)
E.C. NUMBER	EC3.2.1.17 Go to BRENDA
PMD NO	A921426
PDB_wild	2LZM Homologous PDB Entries
PDB_mutant	1L56
MUTATION	K 60 P
NO. OF MOLECULE	1
SECONDARY STRUCTURE	Helix
ACCESSIBLE SURFACE AREA	107.7 A**2

```

NO.      1
PROTEIN  staphylococcal nuclease
SOURCE   Staphylococcus aureus
MUTATION G 20 E
MUTATED_CHAIN
PDB_MUTANT
SEC_STR.  ASA
NO. MOLECULE
STATE
DS_H2O    6.3
DS_H2O    -3.2
DS        20
DSR
T
ITR
DHUN
DHCA1
M
4.7
1.3
OCP
PH
3.9
BUFFER_NAME  HEPES
BUFFER_CONC  10 mM
ION_NAME_1    NaCl
ION_CONC_1    100 mM
ADDITIVES
PROTEIN_CONC  50 microg/mL
MEASURE
METHOD        gdnHCl
REVERSIBILITY unknown
ACTIVITY
ACTIVITY_KM
ACTIVITY_KCAT
ACTIVITY_KD
KEY_WORDS
REFERENCE
AUTHOR
REMARKS
//
    
```

Entry: - PDB Code: Start Clear

Protein: Source:

Mol-weight: To

Mutation: To ☒ Single ☐ Double ☐ Multiple ☐ Wild Type

Sec. Structure: ☒ Helix ☒ Sheet ☐ Turn ☐ Coil

Accessibility: ☐ Any ☐ Buried ☐ Partially Buried ☐ Exposed ☐ ASA 0 To 20 %

Measure: ☐ Absorbance ☐ CD ☐ DSC ☐ Fluorescence ☐ NMR ☐ Others

Method: ☒ Thermal ☐ Denaturants ☐ Others

pH: 5 To 9

dTm/dT: dTm 0 To 50 C

dH/dCp/dG/dG H2O: dH To energy unit kcal

ddG/dG H2O: ddG To

State: ☒ 2 ☐ 3 ☐ >3

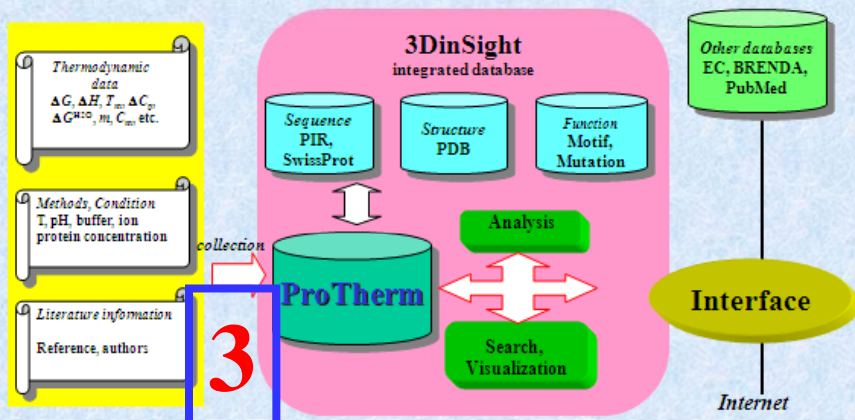
Reversibility: Yes

Keyword: OR

Author: OR

Year: Since Until

Terms	Explanations
No	Entry number. This option can be used for getting data from a particular entry (Eg. 3012) or a range of entries (Eg. 10107-10365) with/without other search conditions
Protein Name	Name of the protein. Multiple words can be entered with spaces. Wild card can also be used: * for a string of characters and ? for a character. If * and ? were to be used as real characters, place "\" (backslash) before them.
E.C.No.	Enzyme Commission number
PMD No.	Protein Mutant Database accession number
PDB_wild	Protein Data Bank code for the native protein



Entry	Protein	PDB_wild	PDB_mutant	Mutation	Sec.Str.	ASA (%)	Tm	dTm	pH	REFERENCE
1459	LYSOZYME	2LZM	1QT6	E 11 H	H	18.89	65.20	0.10	5.40	PROC NATL ACAD SCI U S A 92, 452-456 (1995)
1522	LYSOZYME	2LZM	1L87	F 153 L	H	0.05	65.68	0.80	5.70	J MOL BIOL 229, 747-769 (1993)
199	LYSOZYME	2LZM	1L23	G 77 A	H	4.83	65.60	0.90	6.50	BIOPOLYMERS 32, 1431-1441 (1992)

PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

Search: PubMed Limits [Advanced search](#) [Help](#)

Display Settings: ☒ Abstract

Nucleic Acids Res. 1999 Jan 1;27(1):266-8.

ProTherm: Thermodynamic Database for Proteins and Mutants.

Gromiha MM, An J, Kono H, Oobatake M, Uedaira H, Sarai A.

Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Dr. Michael Gromiha, IFM, BT4010

Organization: Relational database

In 1970, E.F. Codd from IBM described the relational database.

The basic unit of a relational database is a set of correspondence between different features of the database contents, called **tables**.

Relational database is the one in which data are organized as tables, each table comprising a group of records with the same fields (known as attributes). This allows related data to be linked (reassembled) as required without reorganizing the original tables.

The set theoretic operations (union, intersection, difference, Cartesian product) on tables facilitate processing of logically complex queries.

Example

Table 1

Two tables from a relational database of properties of amino acids

Amino acid	3-letter code	1-letter code	Volume (\AA^3)	Surface area (\AA^2)	Distal group
Alanine	Ala	A	88.6	115	Methyl
Arginine	Arg	R	173.4	225	Guanidinium
Asparagine	Asn	N	111.1	150	Amide
Aspartic acid	Asp	D	114.1	160	Carboxyl
Cysteine	Cys	C	108.5	135	Sulphydryl
Glutamic acid	Glu	E	138.4	190	Carboxyl
Glutamine	Gln	Q	143.8	180	Amide
Glycine	Gly	G	60.1	75	Hydrogen
Histidine	His	H	153.2	195	Imidazole
Isoleucine	Ile	I	166.7	175	Methyl
Leucine	Leu	L	166.7	170	Methyl
Lysine	Lys	K	168.6	200	Amino
Methionine	Met	M	162.9	185	Methyl
Phenylalanine	Phe	F	189.9	210	Phenyl
Proline	Pro	P	112.7	145	Pyrrolidine
Serine	Ser	S	89.0	115	Hydroxyl
Threonine	Thr	T	116.1	140	Hydroxyl
Tryptophan	Trp	W	227.8	255	Indole
Tyrosine	Tyr	Y	193.6	230	Phenol
Valine	Val	V	140.0	155	Methyl

Two tables from a relational database of properties of amino acids (continued)

Distal group	H-bond donor	H-bond acceptor
Amide	yes	yes
Amino	yes	no
Carboxyl	no	yes
Guanidinium	yes	yes
Hydrogen	no	no
Hydroxyl	yes	yes
Indole	yes	yes
Methyl	no	no
Phenol	yes	yes
Phenyl	no	no
Pyrrolidine	yes	no
Sulphydryl	yes	no

Simple: What are the three letter codes of the amino acids, which have distal carboxyl group?

View

Compound: What are the three letter codes of the amino acids with volume more than 125 \AA^3 and have distal carboxyl group?

What are the three letter codes of the amino acids, which can serve as hydrogen bond donors?

Join

Lesk, 2008

M. Michael Gromiha, IITM, BT4010

Example

General form of joining is the Cartesian product of the two tables. If the set contains n and m elements the product will contain nm elements. Here, 20 amino acids and 12 distal groups and the total will be 240 rows.

From Table 1				From Table 2			
AA	3	1	V	A Group	Group	Donar	Acceptor
<hr/>							
Alanine	Ala	A	88.6	115 Methyl	Amide	Yes	Yes
Alanine	Ala	A	88.6	115 Methyl	Amide	Yes	No
Alanine	Ala	A	88.6	115 Methyl	Methyl	No	No
Aspartic acid	Asp	D	114.1	160 Carboxyl	Carboxyl	No	Yes

Three letter codes of amino acids that have side chains that could serve as hydrogen bond acceptors:

Natural join

Complex queries

What are the **three letter codes** of amino acids with **volumes** between 100 and 150 AND [(that can serve as **hydrogen bond donors** AND NOT serve as **hydrogen bond acceptors**) OR (that have **surface areas** greater than 120 A2 AND have **distal** methyl groups)].

The **structured Query Language (SQL)** is fairly well standardized syntax for probing relational databases with complex queries.

Complex queries containing logical connectivities are translatable into Codd's set of operations on tables.

Syntax

```
SELECT <3_letter_code> from  
    <amino_acid_table>
```

```
WHERE (sidechain_volume between  
    100 and 150)
```

```
AND
```

```
(H-bond_donor = "yes" AND H-  
    hond_acceptor = "no")
```

```
OR
```

```
(surface_area > 120 AND  
    distal_group = "methyl"))
```

Database collections

Nucleic acid research Database issue (First issue in every year). It is available for free access.

<http://nar.oupjournals.org/>

Listing of databases

<http://www.oxfordjournals.org/nar/database/a/>

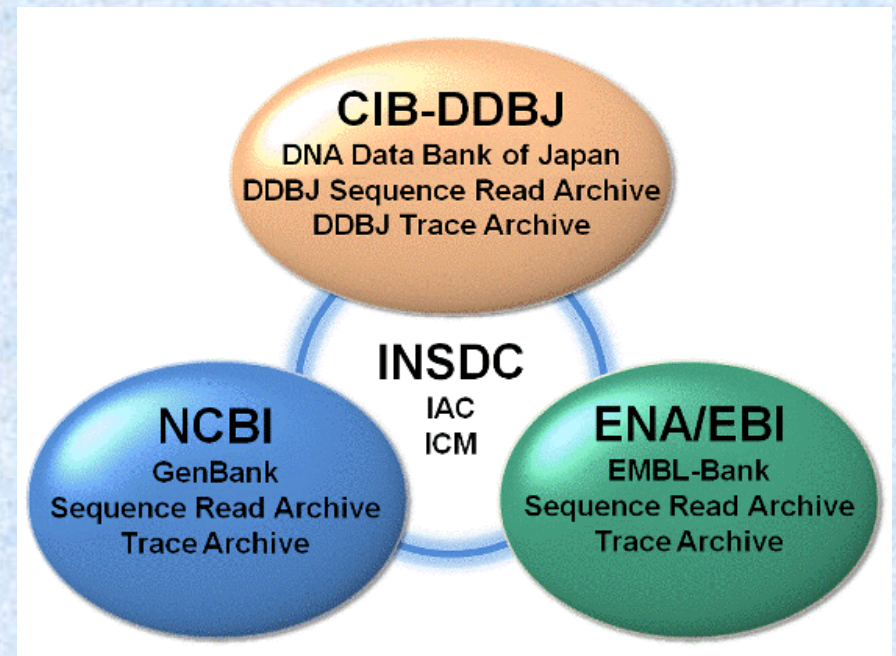
Database categories

- Nucleotide Sequence Databases
- RNA sequence databases
- Protein sequence databases
- Structure Databases
- Thermodynamic databases
- Genomics Databases (non-vertebrate)
- Metabolic and Signaling Pathways
- Human and other Vertebrate Genomes
- Human Genes and Diseases
- Microarray Data and other Gene Expression Databases
- Proteomics Resources
- Other Molecular Biology Databases (PUBMED)
- Organelle databases
- Plant databases
- Immunological databases

Nucleotide sequence databases

International collaboration

1. DDBJ (**D**N**A** **D**ata **B**ank of **J**apan)
2. EMBL (**E**uropean **M**olecular **B**iology **L**aboratory)
3. Genbank (USA)



They exchange sequences via SINET3 Computer network
SINET: Science Information Network

1. DNA Data Bank of Japan

www.ddbj.nig.ac.jp

DDBJ
DNA Data Bank of Japan


Japanese


Google™ カスタム検索 Search


About DDBJ How to Use Report/Statistics FAQ Contact Us


RSS DDBJ Twitter

DDBJ Service

 Data Submission

 Search / Analysis

 Super Computer

 ftp.ddbj.nig.ac.jp

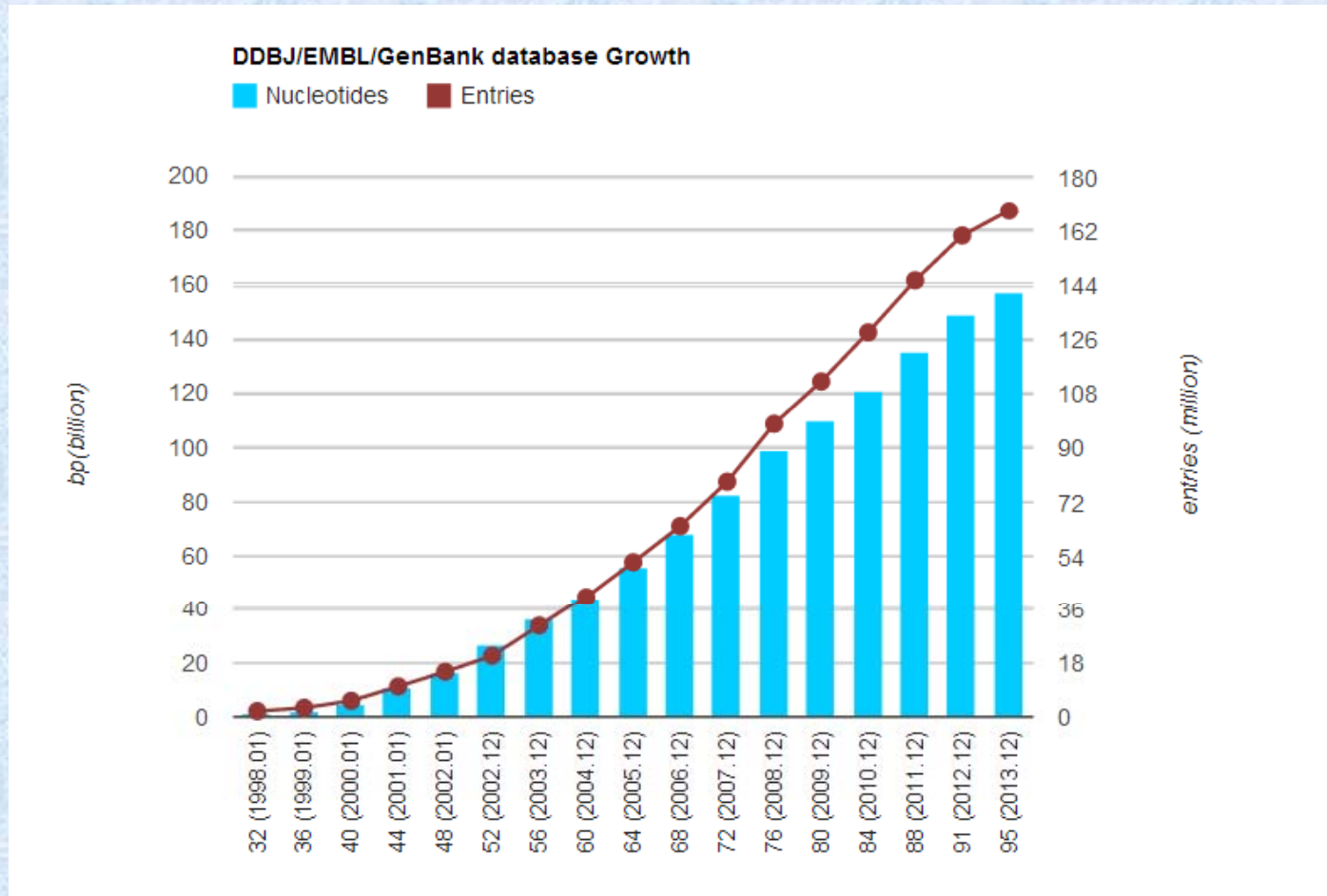
Hot Topics News Archive

News Release Super Computer Maintenance Operation All

INSDC
DDBJ NCBI ENA/EBI
International Nucleotide Sequence Database Collaboration

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

DDBJ: Growth



Aug 2014: > 156.5 billion nucleotides

Number of sequences: > 169 millions

DDBJ: Topmost 10 organisms

No. Organism	Nucleotides	Entries
001 Homo sapiens	17382318654 bp	20522483
002 Mus musculus	9988986985 bp	9727377
003 Rattus norvegicus	6525616319 bp	2197920
004 Bos taurus	5390887315 bp	2202851
005 Zea mays	5076664404 bp	3963159
006 Sus scrofa	4890614452 bp	3289376
007 Danio rerio	3120659920 bp	1726789
008 Marine metagenome	2482805950 bp	3173890
009 Vitis vinifera	1555395119 bp	810152
010 Hordeum vulgare	1455338435 bp	1009480

Search

Search Condition

Quick Search

Homo sapiens mRNA Glyceraldehyde-3-phosphate-dehydrogenase

Search

AND ▾

Available Fields

Search Result

Facet

List of Entries

1 - 1 entries / Number of founds: 1 ☒ FlatFile ☐ XML ☐ Fasta [View selected](#) [Download selected](#) [Download All](#)

PrimaryAccessionNumber ◆ Definition ◆ SequenceLength ◆ MolecularType ◆ Organism ◆

☐ [Z36833](#) Definition:H.sapiens (xs4) mRNA, 315bp. SequenceLength:315 MolecularType:mRNA Organism:Homo sapiens

End of search results

 [PAGE TO](#)

Contents

- Name
- Source
- Accession number
- Keywords
- Authors
- Reference
- PUBMED index
- Nucleotide sequence
- Number of bases
- A, T, C and G
- Protein sequence (translated)

```

LOCUS       Z36833               315 bp    mRNA    linear    HUM 10-APR-1997
DEFINITION  H.sapiens (xs4) mRNA, 315bp.
ACCESSION   Z36833
VERSION     Z36833.1
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   Mueller-Pillasch,F., Gress,T., Lehrach,H. and Adler,G.
  TITLE     Differential gene expression in pancreatic cancer. Use of an
            automated approach for the large scale isolation and
            characterisation of cDNA clones containing differentially expressed
            sequences.
  JOURNAL   Unpublished.
REFERENCE   2  (bases 1 to 315)
  AUTHORS   Gress,T.
  JOURNAL   Submitted (16-AUG-1994) to the INSDC. Gress T., University of Ulm,
            Department of Internal Medicine I, Robert Koch Str.8, 89081 Ulm,
            Germany, 89081
REFERENCE   4  (bases 1 to 315)
  AUTHORS   Gress,T.M., Muller-Pillasch,F., Geng,M., Zimmerhackl,F.,
            Zehetner,G., Friess,H., Buchler,M., Adler,G. and Lehrach,H.
  TITLE     A pancreatic cancer-specific expression profile
  JOURNAL   Oncogene 13(8), 1819-1830(1996).
  PUBMED    8895530

FEF BASE COUNT      80 a      74 c      82 g      76 t
ORIGIN
      1 atcagcgggat ttgcgtcgta ttgggcgcct ggatcaccag ggctgctttt atctctggta
     61 aagtggatat tgttgacatc actgaccccc acattgacca catatacatg gtttacatgt
    121 tccaatatga ctccacccat gagatatton atgacacoga caggggtgag aacgggnagc
    181 ttgacatcaa tggaaatccc acaccatctn cgaggagaga catcctccaa catcatgtgg
    241 cgagatgttg cgctgagtac gtctgtggagt cactgtgtct cacacatgag aggtgtgctc
    301 atggagggggg agcaa
  
```

//

Advanced search

Advanced Search

Quick Search

Field

Show examples

Primary Accession Number

Search Result

Facet

List of Entries

1 - 30 entries / Number of founds: 26506 ● FlatFile

PrimaryAccessionNumber	Definition	Se
<input type="checkbox"/> AF217656	Definition:Human papillomavirus is Organism:Human papillomavirus	
<input type="checkbox"/> AF217657	Definition:Human papillomavirus is Organism:Human papillomavirus	
<input type="checkbox"/> AF217659	Definition:Human papillomavirus is Organism:Human papillomavirus	
<input type="checkbox"/> AF217658	Definition:Human papillomavirus Organism:Human papillomavirus	
<input type="checkbox"/> AF217660	Definition:Human papillomavirus Organism:Human papillomavirus	
<input type="checkbox"/> AF217661	Definition:Human papillomavirus Organism:Human papillomavirus	
<input type="checkbox"/> AF455142	Definition:Human papillomavirus Organism:Human papillomavirus	
<input type="checkbox"/> AF455144	Definition:Human papillomavirus Organism:Human papillomavirus	
<input type="checkbox"/> AF455146	Definition:Human papillomavirus Organism:Human papillomavirus	

Reference PubMedID

LOCUS AF217656 434 bp DNA linear VRL 30-NOV-2000
DEFINITION Human papillomavirus isolate FA14 major capsid protein L1 gene,
partial cds.
ACCESSION [AF217656](#)
VERSION AF217656.1
KEYWORDS .
SOURCE Human papillomavirus
ORGANISM [Human papillomavirus](#)
Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
unclassified Papillomaviridae.
REFERENCE 1 (bases 1 to 434)
AUTHORS Antonsson,A., Forslund,O., Ekberg,H., Sterner,G. and Hansson,B.G.
TITLE The ubiquity and impressive genomic diversity of human skin
papillomaviruses suggest a commensal nature of these viruses
/translation="NIYNNQGTRELVPKVSGNQHRVFRLKLPDPNRFALADMSVYNPD
KERLVWGLKGIEIGRGQPLGIGSSGHPLFNKVNNTENGNTYRNSSKDDRQNI SFDPKQ
LQMFIIIGCTPCIGEHWDRA PACVNDQAGRCPPIELINSYIQ"
BASE COUNT 153 a 71 c 93 g 117 t
ORIGIN
1 tcaatatttta taacaatcaa ggcacacgat tggaggttcc taaagtatca ggaaatcaac
61 acagggtatt tagattaaag ctaccagatc ctaatagggt tgcgttagct gacatgtcag
121 tatataaccc tgacaaagaa agattagtat ggggtttgaa aggcatagaa ataggcaggg
181 gccaaccttt aggaataggc agcagtggtc atccactgtt taataagggt aatgatacag
241 aaaatggcaa tacatatagg aactcctcta aggatgatag acaaaatatt tcatttgacc
301 ccaagcagtt gcaaattgtt attattggct gtactccatg tataggagaa cattgggaca
361 gagcaccagc atgtgttaat gatgatcaag ctggtagatg tcctcctata gagttaataa
421 actcatatat acag

//

DDBJ Data Submission

- Single sequence
- Multiple sequences
- Updates

Sequence, address, contact details, status of publications.

SAKURA

SAKURA is a nucleotide sequence data submission system through the WWW server at DDBJ.
SAKURA has been open to public and continuously refined since 1995. Using this system, you can interactively enter and submit nucleotide and translated amino acid sequences, functions and features of the sequences, and references as well as your name, affiliation and address.

Mass Submission System (MSS)



[Japanese](#)

Go

We recommend using Mass Submission System (MSS) when:

1. The submission consists of large number of entries.
2. The submission involves long nucleotide, complex submission resulting in a many features such as genome data.
3. The submission is unsuitable for SAKURA

Genbank

The screenshot shows the top navigation bar of the NCBI website with links for 'NCBI', 'Resources', and 'How To'. Below this is the 'GenBank' section header, followed by a search bar with a dropdown menu set to 'Nucleotide'. A horizontal menu contains links for 'GenBank', 'Submit', 'Genomes', 'WGS', 'HTGs', 'EST/GSS', and 'Metagen'. The main content area is titled 'GenBank Overview' and includes a section 'What is GenBank?' which describes the database as an annotated collection of publicly available DNA sequences. It mentions the International Nucleotide Sequence Collaboration (INSDC) and provides links to release notes, growth statistics, and an example record for *Saccharomyces cerevisiae*. Another section, 'Access to GenBank', lists search methods like Entrez Nucleotide, CoreNucleotide, dbEST, and dbGSS.

NCBI Resources How To

GenBank Nucleotide

GenBank Submit Genomes WGS HTGs EST/GSS Metagen

GenBank Overview

What is GenBank?

GenBank[®] is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2013 Jan;41(Database issue):D36-42). GenBank is part of the [International Nucleotide Sequence Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

The complete [release notes](#) for the current version of GenBank are available on the NCBI ftp site. A new release is issued every two months. GenBank growth [statistics](#) for both the traditional GenBank divisions and the WGS divisions are available from each release.

An example of a GenBank [record](#) may be viewed for a *Saccharomyces cerevisiae* gene.

Access to GenBank

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with [Entrez Nucleotide](#), which is divided into three divisions: [CoreNucleotide](#) (the main collection), [dbEST](#) (Expressed Sequence Tags), and [dbGSS](#) (Genomic Sequences).

GenBank[®] is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences

***Nucleic Acids Research*, 2013 Jan;41(Database issue):D36-42.**

Genbank: contents

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
 TITLE Cloning and sequence of REV7, a gene whose function is in
 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 a 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, Haven, CT, USA

FEATURES
 source Location/Qualifiers
 1..5028
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /chromosome="IX"
 /map="9"

gene complement (3300..4037)
 /gene="REV7"
 CDS complement (3300..4037)
 /gene="REV7"
 /codon_start=1
 /product="Rev7p"
 /protein_id="AAA98667.1"
 /db_xref="GI:1293616"
 /translation="MNRWVEKWLRVYLKCYINLILFYRNVYPPQSFDDYTTYQSFNLPO
 FVPINRHPALIDYIEELILDVLSKLTHVYRFSICINKKNDLCEKYVLDVDFSELQHV
 KDDQIITETEVFDEFSSLSNLSIMHLEKLPKVNDTITFEAVINAIIELELGHKLDRNR
 RVDSLEEKAEIERDSNWWKCQEDENLPDNNGFQPPKIKLTSLVGSDVGPLIIHQFSEK
 LISGDDKILNGVYSQYEEGESIFGSLF"

ORIGIN

```

1 gatcctccat atacaacggt atctccacct cagggttaga tctcaacaac ggaaccattg
61 ccgacatgag acagtttaggt atcgtcgaga gttacaagct aaaacgagca gtagtcagct
121 ctgcatttga agccgctgaa gttctactaa ggggtggataa catcatccgt gcaagaccaa
181 gaaccgcaa tagacaacat atgtaacata tttaggatat acctcgaaaa taataaacgg
241 ccacactgtc attattataa ttagaagacg aacgcaaaaa ttatccacta tataattcaa
301 agacgcgaaa aaaaaagaac aacgcgtcat agaacttttg gcaattcgcg tcacaaataa
361 attttggcaa cttatgtttc ctcttcgagc agtactcgag ccctgtctca agaatgtaat
421 aatacccatc gtaggtatgg ttaaagatag catctccaca acctcaaaagc tccttgccga
481 gagtcgccct cttttgtcga gtaattttca cttttcatat gagaacttat tttcttattc

```

Genbank: contents

The **LOCUS** field contains locus name, sequence length, molecule type, GenBank division, and modification date

Definition: Brief description of sequence; includes information such as source organism, gene name/protein name, or some description of the sequence's function

Accession: The unique identifier for a sequence record

Version: A nucleotide sequence identification number that represents a single, specific sequence in the GenBank database. GI: GenInfo identifier.

Keywords: Word or phrase describing the sequence

Source: organism name

Reference: Publications by the authors of the sequence that discuss the data reported in the record.

The GenBank database is divided into 18 divisions:

1. PRI - primate sequences
2. ROD - rodent sequences
3. MAM - other mammalian sequences
4. VRT - other vertebrate sequences
5. INV - invertebrate sequences
6. PLN - plant, fungal, and algal sequences
7. BCT - bacterial sequences
8. VRL - viral sequences
9. PHG - bacteriophage sequences
10. SYN - synthetic sequences
11. UNA - unannotated sequences
12. EST - EST sequences (expressed sequence tags)
13. PAT - patent sequences
14. STS - STS sequences (sequence tagged sites)
15. GSS - GSS sequences (genome survey sequences)
16. HTG - HTG sequences (high-throughput genomic sequences)
17. HTC - unfinished high-throughput cDNA sequencing
18. ENV - environmental sampling sequences

Genbank: contents

Title: TMFunction: database for functional residues in membrane proteins.

Authors: Gromiha MM, Yabuki Y, Suresh MX, Thangakani AM, Suwa M, Fukui K.

Journal: Nucleic Acids Res. 2009 Jan;37(Database issue):D201-4.

PMID: 18842639

Features: Information about genes and gene products, as well as regions of biological significance reported in the sequence.

CDs: Coding sequence; region of nucleotides that corresponds with the sequence of amino acids in a protein (location includes start and stop codons).

complete feature is simply written as *n..m*
Example: 687..3158

< indicates **partial on the 5' end**

Example: <1..206

> indicates **partial on the 3' end**

Example: 4821..5028>

(complement) indicates that the feature is on the complementary strand

Example: complement(3300..4037)

The feature extends from base 3300 through base 4037 but is actually on the complementary strand.

Translation: The amino acid translation corresponding to the nucleotide coding sequence

Base count: Frequency of occurrence of the different base types, A, C, G and T in the sequence.

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

Genbank: search

NCBI

GenBank Overview

PubMed Entrez BLAST OMIM Books Taxonomy Structure

Search for

NCBI Home

NCBI Site Map

GenBank Submissions Handbook

Submit to GenBank

Submit an update

Search GenBank

What is GenBank?

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2011 Jan 39(Database issue):D32-7). There are approximately 126,551,501,141 bases in 135,440,924 sequence records in the traditional GenBank divisions and 191,401,393,188 bases in 62,715,288 sequence records in the WGS division as of April 2011.

The complete [release notes](#) for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the

Nucleotide
Alphabet of Life

Search:

[Display Settings:](#) ☒ Summary, 20 per page, Sorted by Default order

Are you looking for gene information?

[GAPDH \(GAPD\)](#) glyceraldehyde-3-phosphate dehydrogenase [Homo sapiens]

[gpad](#) reference sequences [Genomic \(1\)](#) | [Transcript \(1\)](#) | [All reference sequences](#)

Results: 19

☐ [Homo sapiens glyceraldehyde-3-phosphate dehydrogenase \(GAPDH\), mRNA](#)

NCBI

HOME SEARCH SITE MAP

PubMed All Databases Human Genome

Search across databases

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

<input checked="" type="checkbox"/> 656 PubMed: biomedical literature citations and abstracts	<input type="checkbox"/> none Bo
<input checked="" type="checkbox"/> 245 PubMed Central: free, full text journal articles	<input type="checkbox"/> none OM
<input checked="" type="checkbox"/> 1 Site Search: NCBI web and FTP sites	

<input checked="" type="checkbox"/> 19 Nucleotide: Core subset of nucleotide sequence records	<input type="checkbox"/> none dbGa
<input type="checkbox"/> none EST: Expressed Sequence Tag records	<input checked="" type="checkbox"/> 2 UniGe
<input type="checkbox"/> none GSS: Genome Survey Sequence records	<input type="checkbox"/> none CDD:
<input checked="" type="checkbox"/> 3 Protein: sequence database	<input type="checkbox"/> none UniST
<input type="checkbox"/> none Genome: whole genome sequences	<input type="checkbox"/> none PopSe
<input type="checkbox"/> none Structure: three-dimensional macromolecular structures	<input type="checkbox"/> none GEO P
<input type="checkbox"/> none Taxonomy: organisms in GenBank	<input checked="" type="checkbox"/> 2 GEO D

Nucleotide
Alphabet of Life

Search:

[Display Settings:](#) ☒ GenBank [Send:](#)

Human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA, complete cds

GenBank: M33197.1

[FASTA](#) [Graphics](#)

[Go to:](#) ☒

LOCUS HUMGAPDH 1268 bp mRNA linear PRI 08-NOV-1994

DEFINITION Human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA, complete cds.

ACCESSION M33197

VERSION M33197.1 GI:182976

KEYWORDS glyceraldehyde-3-phosphate dehydrogenase.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1268)

AUTHORS Tokunaga,K., Nakamura,Y., Sakata,K., Fujimori,K., Ohkubo,M., Sawada,K. and Sakiyama,S.

TITLE Enhanced expression of a glyceraldehyde-3-phosphate dehydrogenase gene in human lung cancers

JOURNAL Cancer Res. 47 (21), 5616-5619 (1987)

PUBMED [3664468](#)

COMMENT Original source text: Human lung cancer cell, cDNA to mRNA.

FEATURES Location/Qualifiers

source 1..1268

/organism="Homo sapiens"

EMBL: Nucleotide sequence database

The EMBL Nucleotide Sequence Database (also known as EMBL-Bank) constitutes **Europe's primary nucleotide sequence resource**. Main sources for DNA and RNA sequences are direct submissions from individual researchers, genome sequencing projects and patent applications.

The screenshot shows the EMBL-EBI website interface. At the top, there's a search bar with the text 'Enter Text Here' and a 'Find' button. Below the search bar is a navigation menu with links: Databases, Tools, Research, Training, Industry, About Us, Help, and Site Index. The main content area is titled 'EMBL Nucleotide Sequence Database' and contains a paragraph describing the database as Europe's primary nucleotide sequence resource. It mentions that main sources for DNA and RNA sequences are direct submissions from individual researchers, genome sequencing projects, and patent applications. A sidebar on the left contains links to ENA Home, EMBL-Bank Home, Access, Documentation, News, Submission, Publications, People, and Contact. Below these links are sections for 'EMBL Fetch' (with a search box and 'Go' button), 'News' (with a recent update from January 2010), and 'Collaborations' (listing INSDC and NCBJ). A table at the bottom right provides links and explanations for various database features.

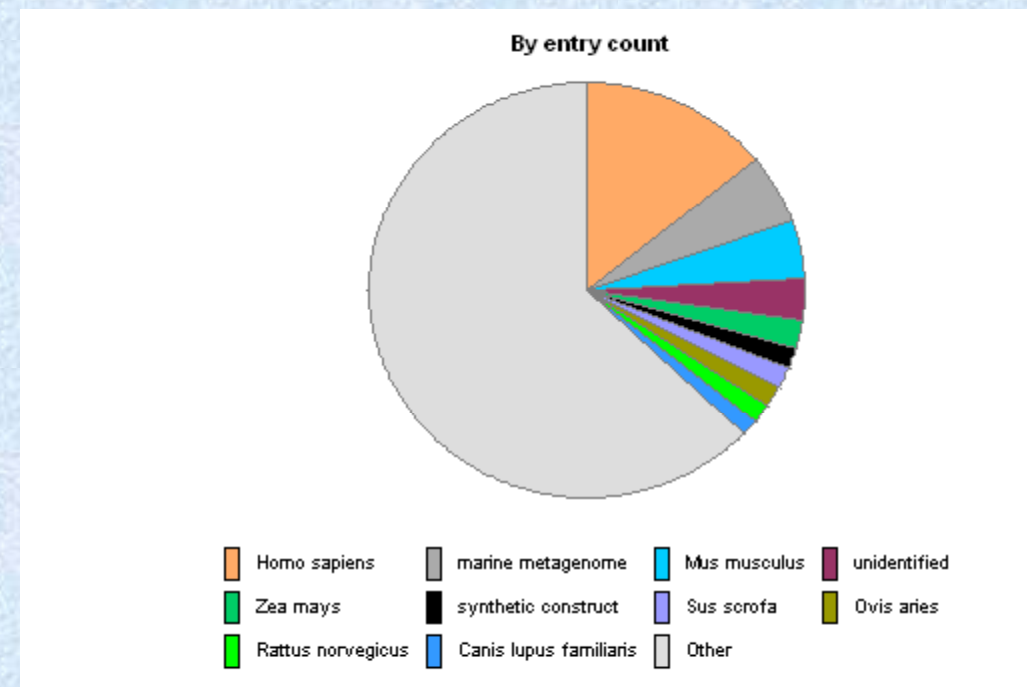
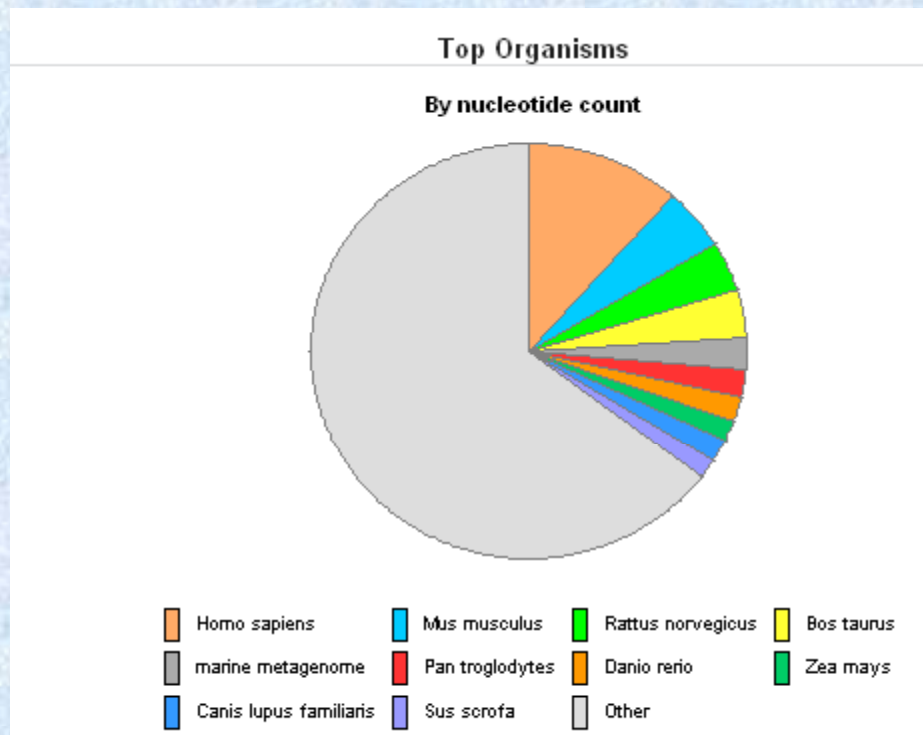
Link	Explanation
Access	Database queries, Completed genomes webserver, FTP archives (EMBL release, alignments etc), EMBL sequence version archive (SVA), Browse by geography .
Submission	Primary sequence submissions, third party annotation, updates.
Documentation	Release notes user manual , Information for Submitters , FAQ , Release information , Forthcoming Changes , EMBL database statistics , Feature table , XML documentation , Sample entry , Examples of annotation , EMBL Features & Qualifiers , DE line standards , Database Policies
Publications	Group publications
People	Group members
Contact	How to contact the EMBL Nucleotide Sequence Database
News	List of recent changes on this site

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

EMBL: Statistics

Release on 18-AUG-2014.

The release contains 468.5 million sequence entries comprising 995.2 billion nucleotides.



Database entry

EMBL-EBI

Enter

Databases

Tools

Research

Training

Indust

EBI > Databases > Database Browsing > Dbfetch > EMBL-Bank: TRBG361

EBI Dbfetch

ID X56734; SV 1; linear; mRNA; STD; PLN; 1859 BP.
XX
AC X56734; S46826;
XX
DT 12-SEP-1991 (Rel. 29, Created)
DT 25-NOV-2005 (Rel. 85, Last updated, Version 11)
XX
DE Trifolium repens mRNA for non-cyanogenic beta-glucosidase.
XX
KW beta-glucosidase.
XX
OS Trifolium repens (white clover)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta
OC Spermatophyta; Magnoliophyta; eudicotyledons; core
OC fabids; Fabales; Fabaceae; Papilionoideae; Trifolium
XX
RN [5]
RP 1-1859
RX DOI; [10.1007/BF00039495](https://doi.org/10.1007/BF00039495)
RX PUBMED; [1907511](https://pubmed.ncbi.nlm.nih.gov/1907511/).
RA Oxtoby E., Dunn M.A., Pancoro A., Hughes M.A.;
RT "Nucleotide and derived amino acid sequence of the cDNA for
RT beta-glucosidase (linamarase) from white clover (Trifolium repens L.)"
RL Plant Mol. Biol. 17(2):209-219(1991).
XX
RN [6]
RP 1-1859
RA Hughes M.A.;
RT ;
RL Submitted (19-NOV-1990) to the INSDC.

XX		
FH	Key	Location/Qualifiers
FH		
FT	source	1..1859
FT		/organism="Trifolium repens"
FT		/mol_type="mRNA"
FT		/clone_lib="lambda gt10"
FT		/clone="TRE361"
FT		/tissue_type="leaves"
FT		/db_xref="taxon:3899"
FT	CDS	14..1495
FT		/product="beta-glucosidase"
FT		/EC_number="3.2.1.21"
FT		/note="non-cyanogenic"
FT		/db_xref="GOA:P26204"
FT		/db_xref="InterPro:IPR001360"
FT		/db_xref="InterPro:IPR013781"
FT		/db_xref="InterPro:IPR017853"
FT		/db_xref="InterPro:IPR018120"
FT		/db_xref="UniProtKB/Swiss-Prot:P26204"
FT		/protein_id="CAA40058.1"
FT		/translation="MDFIVAIFALFVISSFTITSTNAVEASTLLDIGNLSRSSFPGRGFI
FT		FGAGSSAYQFEGAVNEGGRPSIWDTFTHKYPEKIRDGSMADITVDQYHRYKEDVGIMK
FT		DQNMDSYRFSISWPRILPKGKLSGGINHEGIKYNNLINELLANGIQPFVTLFHWDLPO
FT		VLEDEYGGFLNSGVINDFRDYTDLCFKEFGDRVRYWSTLNEPWVFSNSGYALGTNAPGR
FT		CSASNVAKPGDSGTGPIVTHNQILAHAEAVHVYKTKYQAYQKGKIGITLVSNWLMPLD
FT		DNSIPDIKAAERSLDFQGLFMEQLTTGDYSKSMRRIVKNRLPKFSKFESSLVNGSFD
FT		IGINYYSSSYISNAPSHGNAPSYSTNPMTNISFEKHGIPLPRAASIWIVVYPYMFIO
FT		EDFEIFCYILKINITILQFSITENGMEFNDAITLPVEEALLNTYRIDYYRHLYYIRSA
FT		IRAGSNVKGFPYAWSFLDCNEWFAGFTVRFGLNFVD"
FT	mRNA	1..1859
FT		/experiment="experimental evidence, no additional details
FT		recorded"
XX		
SQ	Sequence 1859 BP; 609 A; 314 C; 355 G; 581 T; 0 other;	
	aaacaaacca aatatggatt ttattgtagc catatttgct ctgtttgtta ttagctcatt	60
	cacaattact tccacaaatg cagttgaagc ttctactctt cttgacatag gtaacctgag	120
	tcggagcagt ttctctcgtg gcttcatctt tgggtcgtga tcttcagcat accaatttga	180
	aagtcacata aacgaagacc atagagagacc aagattattcc atacccttca cccataaata	240

EMBL: Search

ID	M33197; SV 1; linear; mRNA; STD; HU	FT	mRNA	<1..1268
XX		FT		/note="GAPDH mRNA"
AC	M33197;	FT	CDS	61..1068
XX		FT		/codon_start=1
DT	10-JUL-1990 (Rel. 24, Created)	FT		/gene="GAPD"
DT	17-APR-2005 (Rel. 83, Last updated)	FT		/note="glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) "
XX		FT		/db_xref="GDB:119249"
DE	Human glyceraldehyde-3-phosphate de	FT		/db_xref="GOA:PO4406"
XX		FT		/db_xref="H-InvDB:HIT000195501"
KW	glyceraldehyde-3-phosphate dehydro	FT		/db_xref="HGNC:4141"
XX		FT		/db_xref="InterPro:IPR006424"
OS	Homo sapiens (human)	FT		/db_xref="InterPro:IPR016040"
OC	Eukaryota; Metazoa; Chordata; Crani	FT		/db_xref="InterPro:IPR020828"
OC	Eutheria; Euarchontoglires; Primat	FT		/db_xref="InterPro:IPR020829"
OC	Homo.	FT		/db_xref="InterPro:IPR020830"
XX		FT		/db_xref="InterPro:IPR020831"
RN	[1]	FT		/db_xref="PDB:1U8F"
RP	1-1268	FT		/db_xref="PDB:1ZNQ"
RX	PUBMED; 3664468.	FT		/db_xref="PDB:2FEH"
RA	Tokunaga K., Nakamura Y., Sakata K.	FT		/db_xref="PDB:3GPD"
RA	Sakiyama S.;	FT		/db_xref="UniProtKB/Swiss-Prot:PO4406"
RT	"Enhanced expression of a glyceral	FT		/protein_id="AAA52518.1"
RT	human lung cancers";	FT		/translation="MGKVKVGVNGFGRIGRLVTRAAFNSGKVDIVAINDPFIDLNYMVY
RL	Cancer Res. 47(21):5616-5619(1987).	FT		MFQYDSTHGKFHGTGTVKAENGKLVINGNPITIFQERDPSKIKWGDAGAEYVVESTGVFTT
XX		FT		MEKAGAHLQGGAKRVIISAPSAAPMFVMGVNHEKYDNSLKIISNASCTTNCLAPLAKV
DR	Ensembl-Gn; ENSG00000111640; Homo_s	FT		IHDNFGIVEGLMTTVHAITATQKTVDGPSGKLWRDGRGALQNIIPASTGAAKAVGKVIP
DR	Ensembl-Tr; ENST00000229239; Homo_s	FT		ELNGKLTGMAFRVPTANVSVVDLTCRLEKPAKYDDIKKVKQASEGPLKGILGYTEHQV
DR	Ensembl-Tr; ENST00000396859; Homo_s	FT		VSSDFNSDTHSSTFDAGAGIALNDHFVKLISWYDNEFGYSNRVVDLMAHMASKE"
DR	Ensembl-Tr; ENST00000396861; Homo_s	XX		
XX		SQ		
FH	Key	Location/Qualifiers:	Sequence 1268 BP; 295 A; 385 C; 326 G; 262 T; 0 other;	
FH			gttcgacagt cagccgcctc ttctttttgcg tcgccagccg agccacatcg ctcagacacc	60
FT	source	1..1268	atgggggaagg tgaaggtcgg agtcaacgga tttgggtcgta ttgggcgcct ggtcaccagg	120
FT		/organism="Homo sa	gctgcttttta actctggttaa agtggatatt gttgccatca atgacccctt cattgacctc	180
FT		/map="12p13"	aactacatgg ttacatggt ccaatatgat tccacccatg gcaaattcca tggcaccgtc	240
FT		/mol_type="mRNA"	aaggctgaga acgggaagct tgtcatcaat ggaaatccca tcaccatctt ccaggagcga	300
FT				---

Literature database

Chemical abstracts

Physical abstracts

Science citation index

PUBMED

Mainly for biological sciences and are freely available

Useful for getting the references of any work and related papers

Search with the keyword “DDBJ Nucleic acid sequence database”



Abstract

NCBI Resources ☒ How To ☒ My NCBI Sign In

PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

Search: PubMed Limits Advanced search Help

Search Clear

[Display Settings:](#) ☒ Abstract [Send to:](#) ☐

OPEN ACCESS **FREE full text article**
OXFORD JOURNALS in PubMed Central

Nucleic Acids Res. 2011 Jan;39(Database issue):D22-7. Epub 2010 Nov 9.

DDBJ progress report.

Kaminuma E, Kosuge T, Kodama Y, Aono H, Mashima J, Gojobori T, Sugawara H, Ogasawara O, Takagi T, Okubo K, Nakamura Y.
Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization for Information and Systems, Yata, Mishima 411-8510, Japan.

Abstract

The DNA Data Bank of Japan (DDBJ, <http://www.ddbj.nig.ac.jp>) provides a nucleotide sequence archive database and accompanying database tools for sequence submission, entry retrieval and annotation analysis. The DDBJ collected and released 3,637,446 entries/2,272,231,889 bases between July 2009 and June 2010. A highlight of the released data was archive datasets from next-generation sequencing reads of Japanese rice cultivar, Koshihikari submitted by the National Institute of Agrobiological Sciences. In this period, we started a new archive for quantitative genomics data, the DDBJ Omics aRchive (DOR). The DOR stores quantitative data both from the microarray and high-throughput new sequencing platforms. Moreover, we improved the content of the DDBJ patent sequence, released a new submission tool of the DDBJ Sequence Read Archive (DRA) which archives massive raw sequencing reads, and enhanced a cloud computing-based analytical system from sequencing reads, the DDBJ Read Annotation Pipeline. In this article, we describe these new functions of the DDBJ databases and support tools.

PMID: 21062814 [PubMed - indexed for MEDLINE] PMCID: PMC3013661 **Free PMC Article**

Related citations

DDBJ launches a new archive database with analytical tools for next-generation sequencing [Nucleic Acids Res. 2010]

Biological databases at DNA Data Bank of Japan in the era of next-generation sequencing [Adv Exp Med Biol. 2010]

The sequence read archive. [Nucleic Acids Res. 2011]

Review The EMBL Nucleotide Sequence Database. Contributing authors [Mol Biotechnol. 1999]

Review From mapping to sequencing, post-sequencing and beyond [Plant Cell Physiol. 2005]

[See reviews...](#)
[See all...](#)

What is Open Access?

Full text



Nucleic Acids Research

Journal List > Nucleic Acids Res > v.39(Database issue); Jan 2011

Nucleic Acids Res. 2011 January; 39(Database issue): D22–D27.
Published online 2010 November 8. doi: [10.1093/nar/gkq1041](https://doi.org/10.1093/nar/gkq1041).

PMCID: PMC3013661

Copyright © The Author(s) 2010. Published by Oxford University Press.

DDBJ progress report

Eli Kaminuma, Takehide Kosuge, Yuichi Kodama, Hideo Aono, Jun Mashima, Takashi Gojobori, Hideaki Sugawara, Osamu Ogasawara, Toshihisa Takagi, Kousaku Okubo, and Yasukazu Nakamura*

Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization for Information and Systems, Yata, Mishima, 411-8510, Japan

*To whom correspondence should be addressed. Tel: Phone: +81 55 981 6859; Fax: +81 55 981 6889; Email: yanakamu@genes.nig.ac.jp

Received September 27, 2010; Revised October 7, 2010; Accepted October 11, 2010.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Other Sections ▼

The DNA Data Bank of Japan (DDBJ, <http://www.ddbj.nig.ac.jp>) provides a nucleotide sequence archive database and accompanying database tools for sequence submission, entry retrieval and annotation analysis. The DDBJ collected and released 3 637 446 entries/2 272 231 889 bases between July 2009 and June 2010. A highlight of the released data was archive datasets from next-generation sequencing

Formats: [Abstract](#) | [Full Text](#) | [PDF \(3.6M\)](#)

PubMed articles by these authors

- ▶ Kaminuma, E.
- ▶ Kosuge, T.
- ▶ Kodama, Y.
- ▶ Nakamura, Y.

PubMed related articles

- ▶ DDBJ launches a new archive database with analytical tools for next-generation sequence data. [Nucleic Acids Res. 2010]
- ▶ Biological databases at DNA Data Bank of Japan in the era of next-generation sequencing technology [Adv Exp Med Biol. 2010]
- ▶ The sequence read archive. [Nucleic Acids Res. 2011]
- ▶ **Review** The EMBL Nucleotide Sequence Database. Contributing and accessing data. [Mol Biotechnol. 1999]
- ▶ **Review** From mapping to sequencing, post-sequencing and beyond. [Plant Cell Physiol. 2005]

» See reviews... | » See all...

Recent Activity

[Turn Off](#) [Clear](#)

DDBJ progress report

Related articles

NCBI Resources How To My NCBI Sign In

PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

Search: PubMed Limits Advanced search Help

Search Clear

Display Settings: Summary, 20 per page, Sorted by Link Ranking Send to:

Results: 1 to 20 of 106

<< First < Prev Page 1 of 6 Next > Last >>

☐ [DDBJ progress report.](#)
1. Kaminuma E, Kosuge T, Kodama Y, Aono H, Mashima J, Gojobori T, Sugawara H, Ogasawara O, Takagi T, Okubo K, Nakamura Y.
Nucleic Acids Res. 2011 Jan;39(Database issue):D22-7. Epub 2010 Nov 9.
PMID: 21062814 [PubMed - indexed for MEDLINE] **Free PMC Article**
[Free full text](#) [Related citations](#)

☐ [DDBJ launches a new archive database with analytical tools for next-generation sequence data.](#)
2. Kaminuma E, Mashima J, Kodama Y, Gojobori T, Ogasawara O, Okubo K, Takagi T, Nakamura Y.
Nucleic Acids Res. 2010 Jan;38(Database issue):D33-8. Epub 2009 Oct 22.
PMID: 19850725 [PubMed - indexed for MEDLINE] **Free PMC Article**
[Free full text](#) [Related citations](#)

☐ [Biological databases at DNA Data Bank of Japan in the era of next-generation sequencing technologies.](#)
3. Kodama Y, Kaminuma E, Saruhashi S, Ikeo K, Sugawara H, Tateno Y, Nakamura Y.
Adv Exp Med Biol. 2010;680:125-35.
PMID: 20865494 [PubMed - indexed for MEDLINE]
[Related citations](#)

☐ [The sequence read archive.](#)
4. Leinonen R, Sugawara H, Shumway M; International Nucleotide Sequence Database Collaboration.
Nucleic Acids Res. 2011 Jan;39(Database issue):D19-21. Epub 2010 Nov 9.

Filter your results:
All (106)
[Free Full Text \(97\)](#)
[Review \(3\)](#)
[Manage Filters](#)

Save Results in Collections Tutorial

My NCBI — Collections

Select: All None 0 items selected Merge Delete

Name	Rating	Last Modified	Type
My Bibliography	0	never	Standard
Other Databases	0	never	Standard
personal cancer	1000	last month	Published
g32 annotation review	95	last month	Published
abstracts from across the database	306	today	Published

0:00 / 2:32

See larger video at YouTube
See all NCBI YouTube video channel videos

PUBMED: Features

Search with keywords display the results from the latest article.

Search with “related entries” show the output from the most relevant article.


One can also use the mesh terms to get a specific article. E.g. Nucleic Acids Res. 34, D204-D206 (2006)

204-206[PAGE] AND 34[VOL] AND 2006[PDAT]

Disease database

Alterations in proteins
cause cancer.

Dr. Akinoi Sarai's
group in Japan



3DinSight

Home 3DinSight ProTherm ProNIT Protein-DNA Recognition Biomolecules Gallery

Advanced search

[Disease database search](#)

Please fill or choose necessary entries below

Display options [?]: ☐ Default ☒ Linking PDB Data

Search Condition

☐ PDB Entry:

☒ Protein Name: [?]

☐ Gene Name: [?]

☐ Disease Name: [?]

☐ Mutant: From To

☐ Keywords: OR

Display hit list from No to No

Overview

What's New

Start search from

- ▶ Main menu
- ▶ Sequence (PIR)
- ▶ Sequence (SWISS-PROT)
- ▶ Motif (PROSITE)
- ▶ Mutation
- ▶ **Disease**
- ▶ Ligand
- ▶ ASASview

Protein P53								
PDBENTRY	SOURCE	MUTATION	PROTEIN	GENE_NAME	DISEASE	DISEASE_DB	SWISS	
1TUP	Homo sapiens	Gly 245 Arg	TUMOR SUPPRESSOR P53	TP53	ADENOCARCINOMA	OMIM	P53_HUMAN	
	Homo sapiens	Pro 152 Leu	TUMOR SUPPRESSOR P53	TP53	ADRENOCORTICAL CARCINOMA	OMIM	P53_HUMAN	
	Homo sapiens	Arg 158 His	TUMOR SUPPRESSOR P53	TP53	ADRENOCORTICAL CARCINOMA	OMIM	P53_HUMAN	
	Homo sapiens	Arg 196 Ter	TUMOR SUPPRESSOR P53	TP53	ADRENOCORTICAL CARCINOMA	OMIM	P53_HUMAN	
	Homo sapiens	Pro 219 Ser	TUMOR SUPPRESSOR P53	TP53	ADRENOCORTICAL CARCINOMA	OMIM	P53_HUMAN	
	Homo sapiens	Asn 235 Asp	TUMOR SUPPRESSOR P53	TP53	ADRENOCORTICAL CARCINOMA	OMIM	P53_HUMAN	
	Homo sapiens	Glu 286 Ala	TUMOR SUPPRESSOR P53	TP53	ADRENOCORTICAL CARCINOMA	OMIM	P53_HUMAN	
	Homo sapiens	Pro 151 Ser	TUMOR SUPPRESSOR P53	TP53	ASTROCYTOMA	OMIM	P53_HUMAN	
	Homo sapiens	Arg 283 His	TUMOR SUPPRESSOR P53	TP53	ASTROCYTOMA	OMIM	P53_HUMAN	
	Homo sapiens	Arg 181 Cys	TUMOR SUPPRESSOR P53	TP53	BREAST CANCER	OMIM	P53_HUMAN	
	Homo sapiens	Arg 181 His	TUMOR SUPPRESSOR P53	TP53	BREAST CANCER	OMIM	P53_HUMAN	
	Homo sapiens	Arg 267 Gln	TUMOR SUPPRESSOR P53	TP53	BREAST CANCER	OMIM	P53_HUMAN	
	Homo sapiens	Pro 278 Leu	TUMOR SUPPRESSOR P53	TP53	BREAST CANCER	OMIM	P53_HUMAN	

Protein function database

Welcome to TMFunction
Functional Database of Membrane Proteins
Wednesday, June 2, 2010

OVERVIEW

TMFunction is a database of functional residues in alpha-helical and beta-barrel membrane proteins. Each protein is identified with its name and source alongwith the Uniprot code. The protein data bank (PDB) codes are also given for available proteins. Different methods and experimental parameters, for example, affinity, dissociation constant, IC50, activity etc. (details are available at the "help" page) are given in the database. Further, we have provided the numerical experimental value for each residue (or mutant) in a protein. The experimental data are collected from the literature both by searching the journals as well as with the keyword search at PUBMED. In addition, complete reference is given with journal citation and PMID number.

TMFunction is cross-linked with the sequence database, Uniprot, structural database, PDB, and literature database, PubMed. The WWW interface enables users to search data based on various terms with different display options for outputs. To search the database, please click on **SEARCH**.

TMHMM-DISE
Discrimination of Beta-Barrel Membrane Proteins from Amino Acid Sequence.

TMHMM-SVM
Discrimination of Beta-Barrel Membrane Proteins using Support Vector Machines

TMHMM-NET
Prediction of Membrane Spanning h-Strand Segments in Outer Membrane Proteins

TMHMM-GENOME
Annotation of Beta-Barrel Membrane Proteins in Genomic Sequences.

AIST ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)
Computational Biology Research Center, AIST, 2008 All Rights Reserved.

SEARCH

Please fill or choose necessary entries below and set display options

Protein **UniProt ID**

Source ☐ TMHelix ☐ TMStrand

Function

Mutation ☐ Single ☐ Multiple ☐ Wild Type

Keyword **OR**

Author **OR**

Year **From** **TO**

Display Option

<input checked="" type="checkbox"/> Entry	<input checked="" type="checkbox"/> PROTEIN	<input checked="" type="checkbox"/> SOURCE	<input checked="" type="checkbox"/> UniProt ID	<input checked="" type="checkbox"/> PDB code	<input checked="" type="checkbox"/> Type
<input checked="" type="checkbox"/> Mutation	<input checked="" type="checkbox"/> Location	<input checked="" type="checkbox"/> Parameter	<input checked="" type="checkbox"/> Data	<input checked="" type="checkbox"/> Function	<input checked="" type="checkbox"/> Experiment
<input checked="" type="checkbox"/> Conditions	<input checked="" type="checkbox"/> Author	<input checked="" type="checkbox"/> PMID	<input checked="" type="checkbox"/> Journal		

Search **Clear**

Search Results

Important residues for protein function

M.M. Gromiha et al.

CBRC, AIST, Japan

Search Conditions

Function **Drug**
Parameter **All**
Protein **All**
Source **All**
UniProt ID **All**

HIT: 216

No.	Protein	UniProt ID	Mutation	Parameter	Data	Function	Experiment	PubMed ID
187	Multi-drug resistance protein 1; MRP1	MRP1_HUMAN (P33527)	S1233A	Relative resistance factor to vincristine	12.5	Drug resistance	site-directed mutagenesis	11925441
188	Multi-drug resistance protein 1; MRP1	MRP1_HUMAN (P33527)	S1235A	Relative resistance factor to vincristine	9.3	Drug resistance	site-directed mutagenesis	11925441
189	Multi-drug resistance protein 1; MRP1	MRP1_HUMAN (P33527)	Y1236F	Relative resistance factor to vincristine	4.8	Drug resistance (partial)	site-directed mutagenesis	11925441
190	Multi-drug resistance protein 1; MRP1	MRP1_HUMAN (P33527)	S1237A	Relative resistance factor to vincristine	13.7	Drug resistance	site-directed mutagenesis	11925441
191	Multi-drug resistance protein 1; MRP1	MRP1_HUMAN (P33527)	Q1239A	Relative resistance factor to vincristine	17.3	Drug resistance	site-directed mutagenesis	11925441
192	Multi-drug resistance protein 1; MRP1	MRP1_HUMAN (P33527)	T1241A	Relative resistance factor to vincristine	4.4	Drug resistance	site-directed mutagenesis	11925441
193	Multi-drug resistance protein 1; MRP1	MRP1_HUMAN (P33527)	Y1243F	Relative resistance factor to vincristine	4.2	Drug resistance (partial)	site-directed mutagenesis	11925441
194	Multi-drug resistance protein 1; MRP1	MRP1_HUMAN (P33527)	N1245A	Relative resistance factor to vincristine	29.3	Drug resistance (partial)	site-directed mutagenesis	11925441
195	Multi-drug resistance protein 1; MRP1	MRP1_HUMAN (P33527)	null	null	null	Drug resistance	site-directed mutagenesis	11925441
196	Multi-drug resistance protein 1; MRP1	MRP1_HUMAN (P33527)	null	null	null	Drug resistance	site-directed mutagenesis	11925441
198	Multi-drug resistance protein 1; MRP1	MRP1_HUMAN (P33527)	S1233A	Relative resistance factor to VP-16	13.3	Drug resistance	site-directed mutagenesis	11925441

M.M. Gromiha et al. (2009) Nucleic Acids Res. 37, D201-204

TMFunction: database for functional residues in membrane proteins

M. Michael Gromiha^{1,*}, Yukimitsu Yabuki¹, M. Xavier Suresh¹,
A. Mary Thangakani², Makiko Suwa¹ and Kazuhiko Fukui¹

¹Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), AIST Tokyo Waterfront Bio-IT Research Building, 2-42 Aomi, Koto-ku, Tokyo 135-0064 and
²Advanced Technology Inc., Tokyo, Japan

M. Michael Gromiha, IITM, BT4010