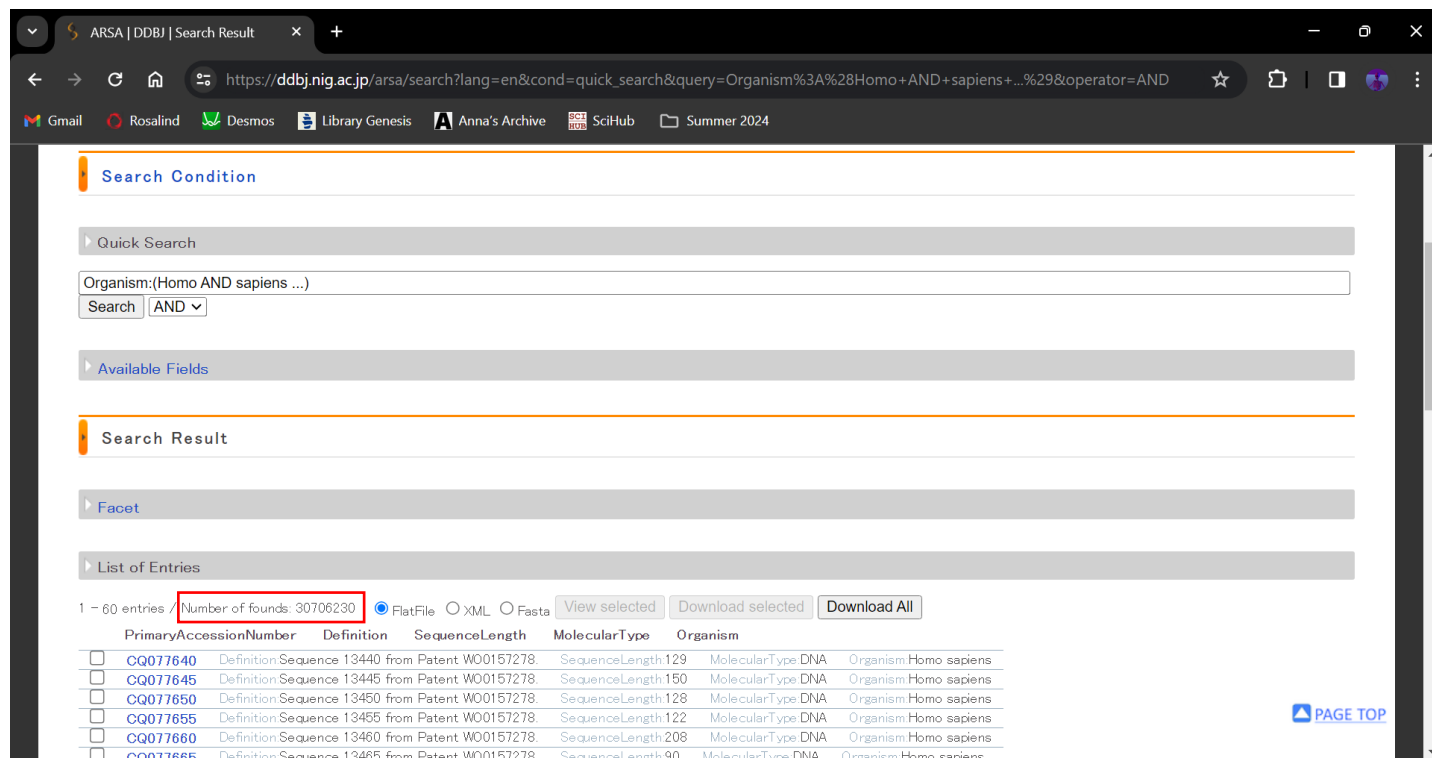


## BT3040 – Bioinformatics

## Practical 2

1

DDBJ has 30706230 sequences from *Homo sapiens*.



Search Condition

Quick Search

Organism:(Homo AND sapiens ...)

Search AND

Available Fields

Search Result

Facet

List of Entries

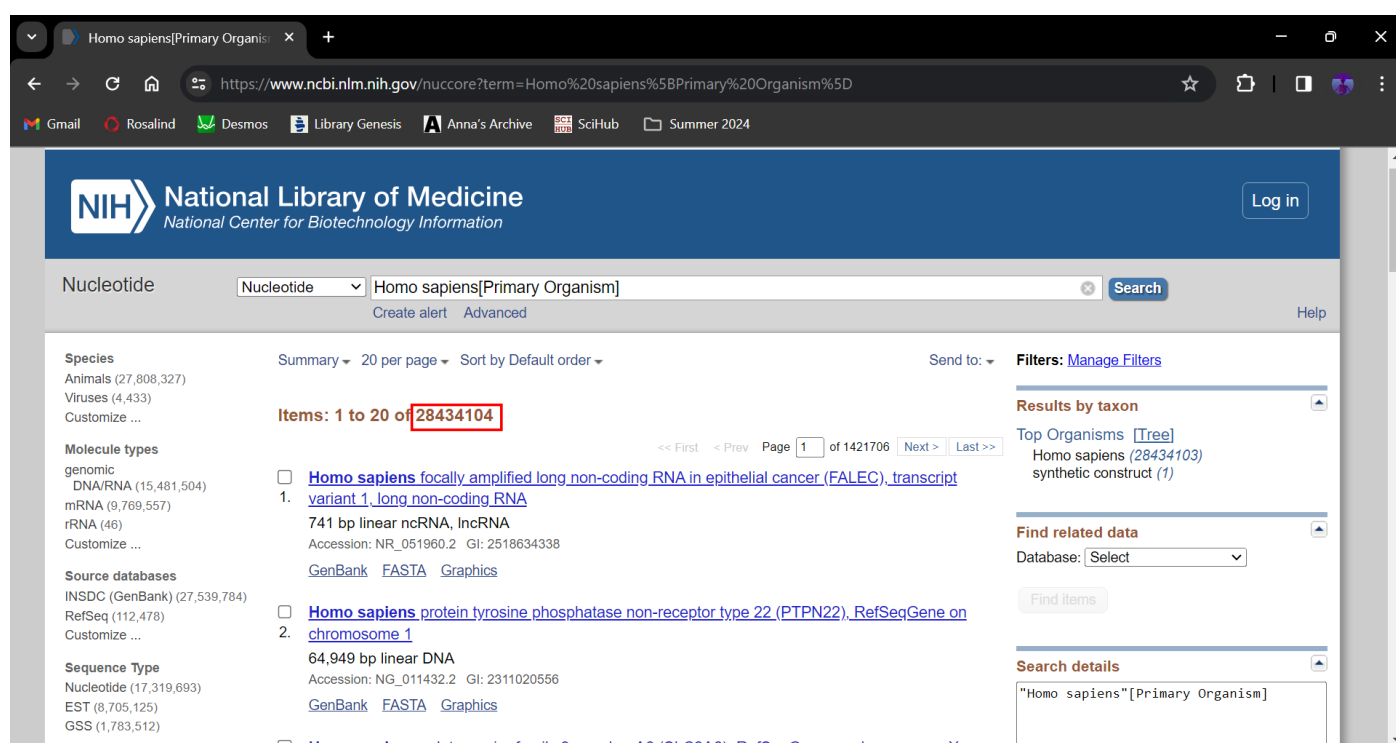
1 - 60 entries / Number of founds: 30706230

FlatFile XML Fasta View selected Download selected Download All

PrimaryAccessionNumber	Definition	SequenceLength	MolecularType	Organism
CQ077640	Definition:Sequence 13440 from Patent W00157278.	SequenceLength:129	MolecularType:DNA	Organism:Homo sapiens
CQ077645	Definition:Sequence 13445 from Patent W00157278.	SequenceLength:150	MolecularType:DNA	Organism:Homo sapiens
CQ077650	Definition:Sequence 13450 from Patent W00157278.	SequenceLength:128	MolecularType:DNA	Organism:Homo sapiens
CQ077655	Definition:Sequence 13455 from Patent W00157278.	SequenceLength:122	MolecularType:DNA	Organism:Homo sapiens
CQ077660	Definition:Sequence 13460 from Patent W00157278.	SequenceLength:208	MolecularType:DNA	Organism:Homo sapiens
CQ077665	Definition:Sequence 13465 from Patent W00157278.	SequenceLength:90	MolecularType:DNA	Organism:Homo sapiens

PAGE TOP

GenBank has 28434104 sequences from *Homo sapiens*.



Nucleotide Nucleotide Homo sapiens[Primary Organism] Search

Create alert Advanced Help

Species Summary 20 per page Sort by Default order

Animals (27,808,327)  
Viruses (4,433)  
Customize ...

Molecule types  
genomic  
DNA/RNA (15,481,504)  
mRNA (9,769,557)  
rRNA (46)  
Customize ...

Source databases  
INSDC (GenBank) (27,539,784)  
RefSeq (112,478)  
Customize ...

Sequence Type  
Nucleotide (17,319,693)  
EST (8,705,125)  
GSS (1,783,512)

Items: 1 to 20 of 28434104

<< First < Prev Page 1 of 1421706 Next > Last >>

1. [Homo sapiens focally amplified long non-coding RNA in epithelial cancer \(FALEC\), transcript variant 1, long non-coding RNA](#)  
741 bp linear ncRNA, lncRNA  
Accession: NR\_051960.2 GI: 2518634338  
[GenBank](#) [FASTA](#) [Graphics](#)

2. [Homo sapiens protein tyrosine phosphatase non-receptor type 22 \(PTPN22\), RefSeqGene on chromosome 1](#)  
64,949 bp linear DNA  
Accession: NG\_011432.2 GI: 2311020556  
[GenBank](#) [FASTA](#) [Graphics](#)

3. [Homo sapiens solute carrier family 9 member A6 \(SLC9A6\), RefSeqGene on chromosome X](#)

Send to: Filters: Manage Filters

Results by taxon  
Top Organisms [Tree](#)  
Homo sapiens (28434103)  
synthetic construct (1)

Find related data  
Database: Select  
Find items

Search details  
"Homo sapiens"[Primary Organism]

EMBL has 40620498 sequences from *Homo sapiens*.

ENA Browser

https://www.ebi.ac.uk/ena/browser/text-search?query=homo%20sapiens

Gmail

Rosalind

Desmos

Library Genesis

Anna's Archive

SciHub

Summer 2024

ENA

European Nucleotide Archive

homo sapiens

Search

Enter accession

View

Examples: histone, BN000065

Examples: Taxon:9606, BN000065, PRJEB402

Home

Submit

Search

Rulespace

About

Support

Text Search

Uses EBI Search to perform a free text search across ENA data. For more detailed usage please refer to the help & documentation section.

Search term:

homo sapiens

Search

Search results for homo sapiens

Assembly

Assembly (1,821)

Sequence

Sequence (40,620,498)

Sequence (CON) (565,646)

Sequence (Standard) (40,054,852)

Contig set

Genome assembly contig set (337,585)

Transcriptome assembly contig set (9)

Sequence

(showing first 1,000,000 of 40,620,498 results)

Accession	Description/Title
KJ946236	Homo sapiens Kidd blood group protein (SLC14A1) gene, exons 4, 5 and partial cds.
DQ268529	Homo sapiens syntaxin 16 isoform C-like protein (STX16) mRNA, 3' UTR.
AY441560	Homo sapiens SINE element Ind1.

Download ENA records:

FASTA

TEXT

TSV

2

The GC content of AY330867 is  $\frac{129+72}{432} \times 100 = 46.5278\%$ .

AY330867

https://getentry.ddbj.nig.ac.jp/getentry/na/AY330867?filetype=html

Gmail

Rosalind

Desmos

Library Genesis

Anna's Archive

SciHub

Summer 2024

LOCUS

AY330867

432 bp

mRNA

linear

SYN 29-JUL-2004

DEFINITION

Synthetic construct human lysozyme mRNA, complete cds.

ACCESSION

AY330867

VERSION

AY330867.1

KEYWORDS

.

SOURCE

synthetic construct

ORGANISM

synthetic construct

other sequences; artificial sequences.

REFERENCE

1 (bases 1 to 432)

AUTHORS

Xie,X.D., Chen,Z.H. and Zhou,C.S.

TITLE

Direct Submission

JOURNAL

Submitted (26-JUN-2003) Postdoctoral Scientific Research Station of Gansu Yasheng Industrial (Group) Co., Ltd., Zhangye Road, Lanzhou, Gansu 730030, P.R.China

FEATURES

source

Location/Qualifiers

1..432

/organism="synthetic construct"

/mol\_type="mRNA"

/db\_xref="taxon:32630"

CDS

1..429

/codon\_start=1

/transl\_table=11

/product="human lysozyme"

/protein\_id="AAP93336.1"

/translation="MRKKRRQRRMKKVFRCCELARTLKRLGMDGYRGISLANMMCLA KWESGYNTRATNYNAGDRSTDYGFIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIA DAVACAKRIARDDPGTRAWAWBNRCQNRGVRQVVGCGV"

BASE COUNT

131 a 72 c 129 g 100 t

ORIGIN

1 atgaggaga agcggagaca gcgacgaaga atgaagaagg tctttgaaag gtgtgagttg

61 gccagaactc tgaagaagatt gggaaatgat ggctacaggg gaatcagcct agcaaaactg

121 atgtgttttg ccaaatggga gagggtttac aacacacgag ctacaaacta caatgctgga

181 gcagaaacca ctgattatgg gatatttcag atcaatagcc gctactggtg taatgatggc

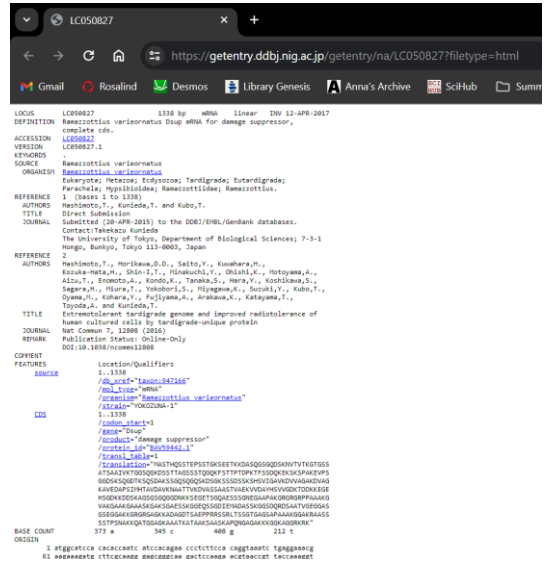
241 aaaacccagg gagcagttaa tgcctgtcat ttatcctgca gtgctttgct gcaagataac

301 atcgctgatg ctgtagcttg tgcaagagg gttgtccgtg atccacaagg cattagagca

361 tcaacacacg gctgctgctg gctgctgctg gctgctgctg gctgctgctg gctgctgctg

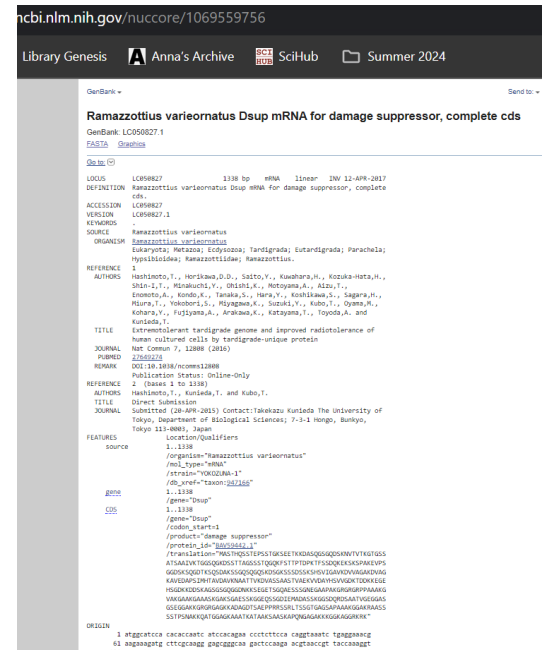
The contents in DDBJ include:

- Locus
- Definition
- Accession
- Version
- Keywords
- Source
- Reference
- Features
- Translation
- Base count
- Sequence



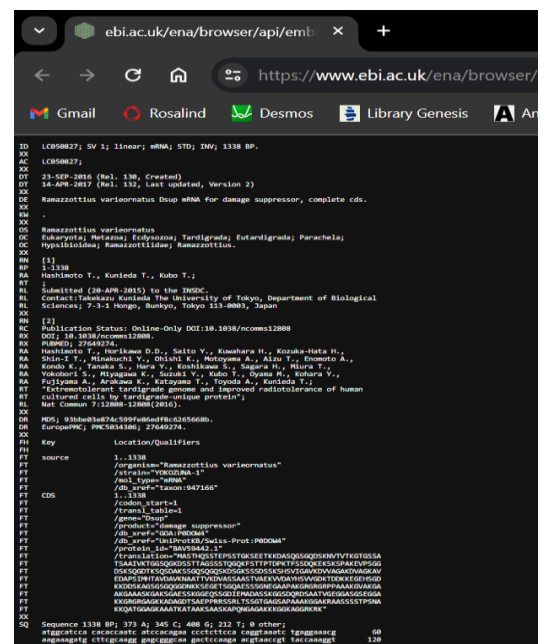
The contents in GenBank include:

- Locus
- Definition
- Accession
- Version
- Keywords
- Source
- Reference
- Features
- Translation
- Sequence



The contents in EMBL include:

- ID
- Accession
- Date of creating, date of updating
- Description
- Source
- Reference
- Features
- Translation
- Base count
- Sequence



discrimination of beta barrel me x +

https://pubmed.ncbi.nlm.nih.gov/?term=discrimination+of+beta+barrel+membrane+proteins

Gmail Rosalind Desmos Library Genesis Anna's Archive SciHub Summer 2024

**PubMed**

discrimination of beta barrel membrane proteins

Advanced Create alert Create RSS User Guide

Save Email Send to Sort by: Best match Display options

MY NCBI FILTERS 53 results << < Page 1 of 6 > >>

RESULTS BY YEAR

1992 2024

TEXT AVAILABILITY

☐ Abstract

☐ Free full text

☐ Full text

**i** 2 articles found by citation matching

**TMBETADISC-RBF: Discrimination of beta-barrel membrane proteins using RBF networks and PSSM profiles.**  
Ou YY, et al. Comput Biol Chem. 2008. PMID: 18434251

**Current developments on beta-barrel membrane proteins: sequence and structure analysis, discrimination and prediction.**  
Gromiha MM, et al. Curr Protein Pept Sci. 2007. PMID: 18220845 Review.

Show all

Gromiha, M Michael[Author] x +

https://pubmed.ncbi.nlm.nih.gov/?term=Gromiha%2C+M+Michael%5BAuthor%5D&sort=

Gmail Rosalind Desmos Library Genesis Anna's Archive SciHub Summer 2024

**PubMed**

Gromiha, M Michael[Author]

Advanced Create alert Create RSS User Guide

Save Email Send to Sort by: Best match Display options

MY NCBI FILTERS 223 results << < Page 1 of 23 > >>

RESULTS BY YEAR

2002 2024

TEXT AVAILABILITY

☐ Abstract

☐ Free full text

☐ Full text

☐ COVID-19 outbreak: history, mechanism, transmission, structural studies and therapeutics.

1

Cite Yesudhas D, Srivastava A, **Gromiha MM**.  
Infection. 2021 Apr;49(2):199-213. doi: 10.1007/s15010-020-01516-2. Epub 2020 Sep 4.  
Share PMID: 32886331 Free PMC article.

☐ Editorial: Protein recognition and associated diseases.

2

Cite **Gromiha MM**, Kundrotas P, Marti MA, Venclovas Č, Li M.  
Front Bioinform. 2023 May 22;3:1215141. doi: 10.3389/fbinf.2023.1215141. eCollection 2023.  
Share PMID: 37283696 Free PMC article. No abstract available.

☐ Drug-Target Interactions: Prediction Methods and Applications

There are 27 related articles listed in PubMed for the given article.

The screenshot shows the PubMed website with the search results for PMID: 19109882. The browser address bar displays the URL: [https://pubmed.ncbi.nlm.nih.gov/?linkname=pubmed\\_pubmed&from\\_uid=19109882](https://pubmed.ncbi.nlm.nih.gov/?linkname=pubmed_pubmed&from_uid=19109882). The PubMed logo is visible at the top left. The search bar contains the text "Similar articles for PMID: 19109882". Below the search bar, there are buttons for "Save", "Email", and "Send to". The "Sort by" dropdown is set to "Best match". The "Display options" button is also visible. On the left side, there is a "MY NCBI FILTERS" section and a "RESULTS BY YEAR" chart showing a peak in 2007. The main content area displays "Similar articles for PMID: 19109882" with a red box highlighting "27 results". The results are listed in a table with columns for "Cite", "Share", and "Text Availability". The first result is "Mitochondrial beta-barrel proteins, an exclusive club?" by Imai K, Gromiha MM, Horton P. The second result is "Evolution. Tinkering inside the organelle." by Alcock F, Clements A, Webb C, Lithgow T.

There are 593 articles published in Nature in 2024 according to PubMed.

The screenshot shows the PubMed website with the search results for "Nature" journal in 2024. The browser address bar displays the URL: <https://pubmed.ncbi.nlm.nih.gov/?term=%28Nature%5BJournal%5D%29&filter=years:2024-2024>. The PubMed logo is visible at the top left. The search bar contains the text "Nature". Below the search bar, there are buttons for "Save", "Email", and "Send to". The "Sort by" dropdown is set to "Best match". The "Display options" button is also visible. On the left side, there is a "MY NCBI FILTERS" section and a "RESULTS BY YEAR" chart showing a peak in 2024. The main content area displays "593 results" with a red box highlighting the count. The results are listed in a table with columns for "Cite", "Share", and "Text Availability". The first result is "A novel antibiotic class targeting the lipopolysaccharide transporter." by Zampaloni C, Mattei P, Bleicher K, Winther L, Thäte C, Bucher C, Adam JM, Alanine A, Amrein KE, Baidin V, Bieniossek C, Bissantz C, Boess F, Cantrill C, Clairfeuille T, Dey F, Di Giorgio P, du Castel P, Dylus D, Dzygiel P, Felici A, García-Alcalde F, Haldimann A, Leipner M, Leyn S, Louvel S, Misson P, Osterman A, Pahil K, Rigo S, Schäublin A, Scharf S, Schmitz P, Stoll T, Trauner A, Zoffmann S, Kahne D, Young JAT, Lobritz MA, Bradley KA. The second result is "An IL-4 signalling axis in bone marrow drives pro-tumorigenic myelopoiesis." by LaMarche NM, Hegde S, Park MD, Maier BB, Troncoso L, Le Berichel J, Hamon P, Belabed M, Mattiuz R, Hennequin C, Chin T, Reid AM, Reyes-Torres I, Nemeth E, Zhang R, Olson OC, Doroshov DB, Rohs NC, Gomez JE, Velurugany R, Hall N, Venturini N, Giboux E, Liu Z, Buckup M, Figueiredo J, Roudko V, Mirak.

There are 420 articles published in Nature in 2024 according to Scopus.

Scopus - Document search results

https://www.scopus.com/results/results.uri?sort=plf-f&src=s&sid=dace223f11d39362b21df285b9a36cf28&sot=a&sdt=cl&sl=14&s=PUBYE...

GmailRosalindDesmosLibrary GenesisAnna's ArchiveSciHubSummer 2024

PUBYEAR = 2024 AND ( LIMIT-TO ( EXACTSRCTITLE , "Nature" ) )

Show less

Save search

Set search alert

Edit in advanced search

Documents

Preprints

Patents

Secondary documents

420 documents found

Analyze results

Refine search

Search within results

Filters

Clear all

Document title

Authors

Source

Year

Citations

1

The mystery of Feynman's sprinkler is solved at last

[No Authors Found]

Nature, 626(7998), pp. 2024 237

0

Year

Show abstract

Full Text

View at Publisher

Related documents

Burkhard Rost has a h-index of 106 and 54103 citations.

Burkhard Rost - Google Scholar


https://scholar.google.com/citations?user=BP3ofxcAAAAJ&hl=en&oi=ao

GmailRosalindDesmosLibrary GenesisAnna's ArchiveSciHubSummer 2024

Google Scholar

Search

Profile icon



Burkhard Rost

Professor of Computer Science & Computational Biology & Bioinformatics, TUM, Munich

Verified email at tum.de - Homepage

machine learning artificial intelligence bioinformatics computer science protein prediction

FOLLOW

GET MY OWN PROFILE

Cited by

VIEW ALL

All

Since 2019

Citations

h-index

i10-index

54103

106

271

13719

57

182

Bar chart showing citation trends from 2017 to 2024

Public access

VIEW ALL

EC 1.7.2.3 is an oxidoreductase. Its function is to reduce trimethylamine-N-oxide.

The screenshot shows the BRENDA database interface for EC 1.7.2.3 - trimethylamine-N-oxide reductase. The page includes a sidebar with navigation links such as Enzyme Nomenclature (43), Enzyme-Ligand Interactions (212), Diseases (2), Functional Parameters (145), Organism related Information (50), General Information (4), Enzyme Structure (4667), Molecular Properties (54), Applications (1), References (44), and External Links. The main content area displays the enzyme's name, a brief description, and an EC Tree showing its classification: 1 Oxidoreductases > 1.7 Acting on other nitrogenous compounds as donors > 1.7.2 With a cytochrome as acceptor > 1.7.2.3 trimethylamine-N-oxide reductase. IUBMB Comments state: 'Contains bis(molybdopterin guanine dinucleotide)molybdenum cofactor. The reductant is a membrane-bound multiheme cytochrome c. Also reduces dimethyl sulfoxide to dimethyl sulfide.' A search section allows marking special words or phrases, and a Word Map shows terms like 'cofactor-containing', 'molybdoenzyme', '1.7.2.3', and 'twin-arginine'.

From BRENDA, we obtain the EC number of asparagine synthetase to be EC 6.3.5.4.

The screenshot shows the BRENDA database interface for EC 6.3.5.4 - asparagine synthase (glutamine-hydrolysing). The sidebar lists various categories with entry counts, including Enzyme Nomenclature (132), Enzyme-Ligand Interactions (484), Diseases (1834), Functional Parameters (367), Organism related Information (368), General Information (47), Enzyme Structure (20k), Molecular Properties (113), Applications (12), and References (113). The main content area provides the enzyme's name, a description, and an EC Tree: 6 Ligases > 6.3 Forming carbon-nitrogen bonds > 6.3.5 Carbon-nitrogen ligases with glutamine as amido-N-donor > 6.3.5.4 asparagine synthase (glutamine-hydrolysing). IUBMB Comments describe the enzyme from Escherichia coli, noting its two active sites and the chemical reactions it catalyzes. A search section and a Word Map (showing terms like 'amycin', 'dioxins', 'hard-rock', 'toddler', 'rubicon') are also visible.



Using this in the Catalytic Site Atlas, we obtain the catalytic residues.

HomeBrowseSearchStatisticsDownload / APIDocumentationAboutContact UsLog in

Catalytic Residues Roles

UniProt	PDB* (1ct9)		
Cys2 (N-term)	Ala1A (N-term)	Acts as a general acid/base to activate the cysteine nucleophile.	proton acceptor, proton donor
Leu51 (main-C)	Leu50A (main-C)	Helps stabilise the reactive intermediates formed.	hydrogen bond acceptor, electrostatic stabiliser
Thr322, Arg325	Thr321A, Arg324A	Bind and stabilise the phosphate groups of the ATP and reactive intermediates formed.	hydrogen bond donor, electrostatic stabiliser
Cys2	Ala1A	Acts as a catalytic nucleophile in the glutaminase domain reaction.	covalently attached, hydrogen bond acceptor, nucleofuge, nucleophile, proton acceptor, proton donor
Gly76 (main-N), Asn75	Gly75A (main-N), Asn74A	Forms the oxyanion hole that stabilises the reactive intermediates and transition states formed.	hydrogen bond donor, electrostatic stabiliser

\*PDB label guide - RESx(y)B(C) - RES: Residue Name; x: Residue ID in PDB file; y: Residue ID in PDB sequence if different from PDB file; B: PDB Chain; C: Biological Assembly Chain if different from PDB. If label is "Not Found" it means this residue is not found in the reference PDB.

Catalytic Residues Roles

proton transfer, bimolecular nucleophilic addition, proton relay, enzyme-substrate complex formation, overall reactant used, intermediate formation, unimolecular elimination by the conjugate base, enzyme-substrate complex cleavage, deamination, intermediate collapse, overall product formed, native

Organism	Scientific Name	Taxonomy ID	No. of Chromosomes
Human	<i>Homo sapiens</i>	9606	46 (22 autosomal pairs, X, Y)
Cat	<i>Felis catus</i>	9685	38 (18 autosomal pairs, X, Y)
Dog	<i>Canis lupus familiaris</i>	9615	78 (38 autosomal pairs, X, Y)
Domestic guinea pig	<i>Cavia porcellus</i>	10141	64 (31 autosomal pairs, X, Y)
Thale cress	<i>Arabidopsis thaliana</i>	3702	5

Search for [ ] as [ ] complete name [ ] lock [ ] Go [ ] Clear [ ]

Display [ 3 ] levels using filter: none [ ]

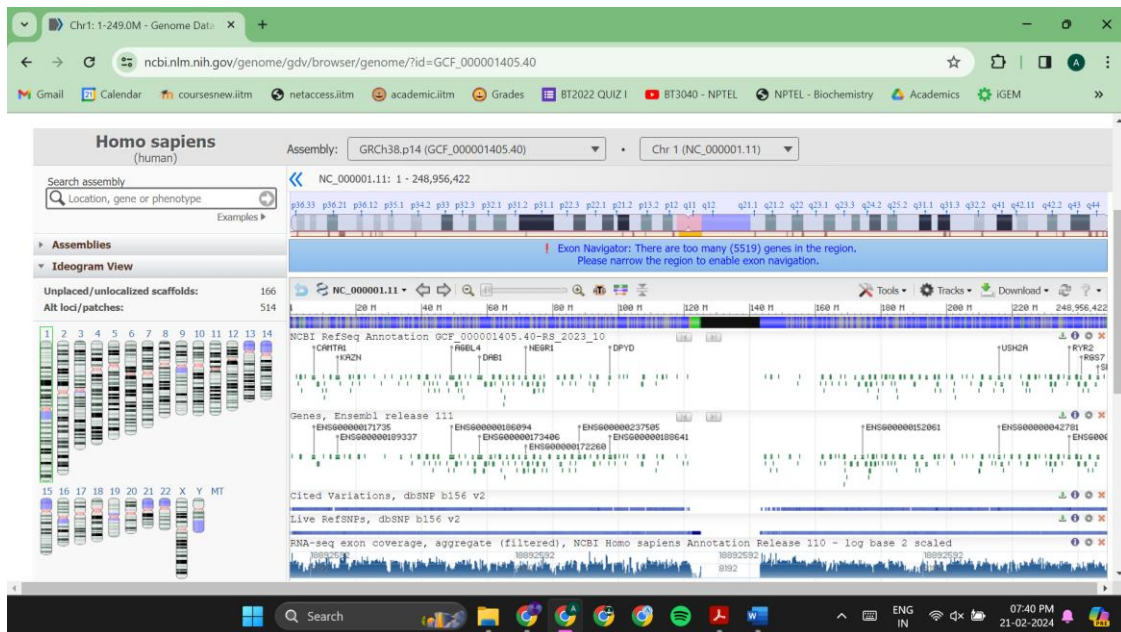
Entrez records

Database name	Subtree links	Direct links
Nucleotide	28,493,843	28,492,441
Protein	1,941,929	1,941,370
Structure	65,278	65,274
Genome	1	1
Popset	25,622	25,621
SNP	1,116,839,409	1,116,839,409
Conserved Domains	83	83
GEO Datasets	3,834,111	3,834,111
PubMed Central	59,726	59,645
Gene	359,883	359,810
HomoloGene	18,713	18,713
SRA Experiments	4,864,780	4,863,676
GEO Profiles	61,958,910	61,958,910
Protein Clusters	13	13
Identical Protein Groups	1,894,887	1,894,794
BioProject	108,074	108,036
BioSamples	9,154,013	9,153,753

Entrez records

Database name	Subtree links	Direct links
Nucleotide	28,493,843	28,492,441
Protein	1,941,929	1,941,370
Structure	65,278	65,274
Genome	1	1
Popset	25,622	25,621
SNP	1,116,839,409	1,116,839,409
Conserved Domains	83	83
GEO Datasets	3,834,111	3,834,111
PubMed Central	59,726	59,645
Gene	359,883	359,810
HomoloGene	18,713	18,713
SRA Experiments	4,864,780	4,863,676
GEO Profiles	61,958,910	61,958,910
Protein Clusters	13	13
Identical Protein Groups	1,894,887	1,894,794
BioProject	108,074	108,036
BioSamples	9,154,013	9,153,753





12

NCBI offers eight server-side programs called the Entrez Programming Utilities, or E-utilities. These provide a stable interface to the query and database system of NCBI, which consists of 38 databases. They use a fixed syntax to search and retrieve data.

The syntax for fetching a record in FASTA format is:

[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=\[database\]&id=\[identifier\]&rettype=fasta&retmode=text](https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=[database]&id=[identifier]&rettype=fasta&retmode=text)

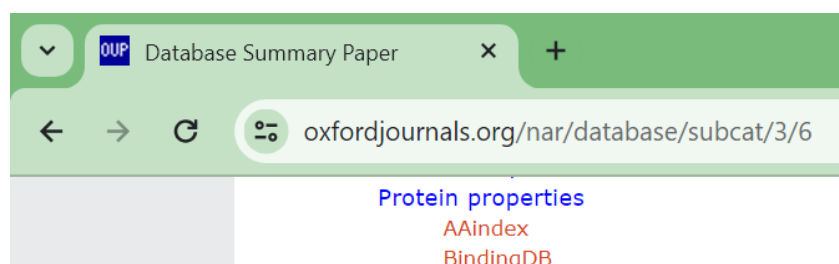
e.g.,

<https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=nuccore&id=34577062&rettype=fasta&retmode=text>

13

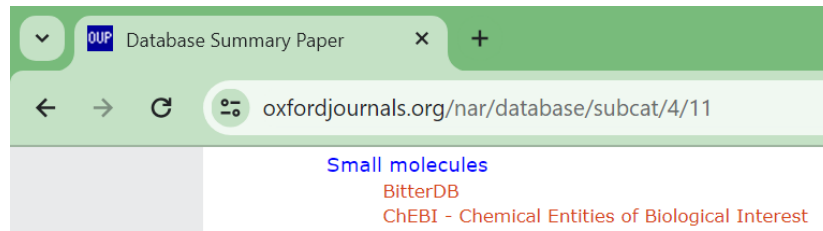
#### (a) Protein properties

- (i) AAindex
- (ii) BindingDB



(b) Small molecules (structure-related)

- (i) BitterDB
- (ii) ChEBI



(c) Cancer gene databases

- (i) ArrayMap
- (ii) BCCTBbp

