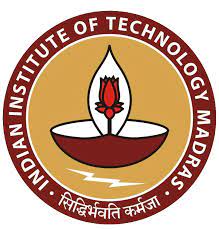
..

**BT 3040: Bioinformatics**

**Assignment 2**



Indian Institute of Technology Madras

Atharva Mandar Phatak | BE21B009

Department of Biotechnology

**Q1)** **How many “Homo sapiens” sequences deposited in DDBJ? Compare with Genbank and**

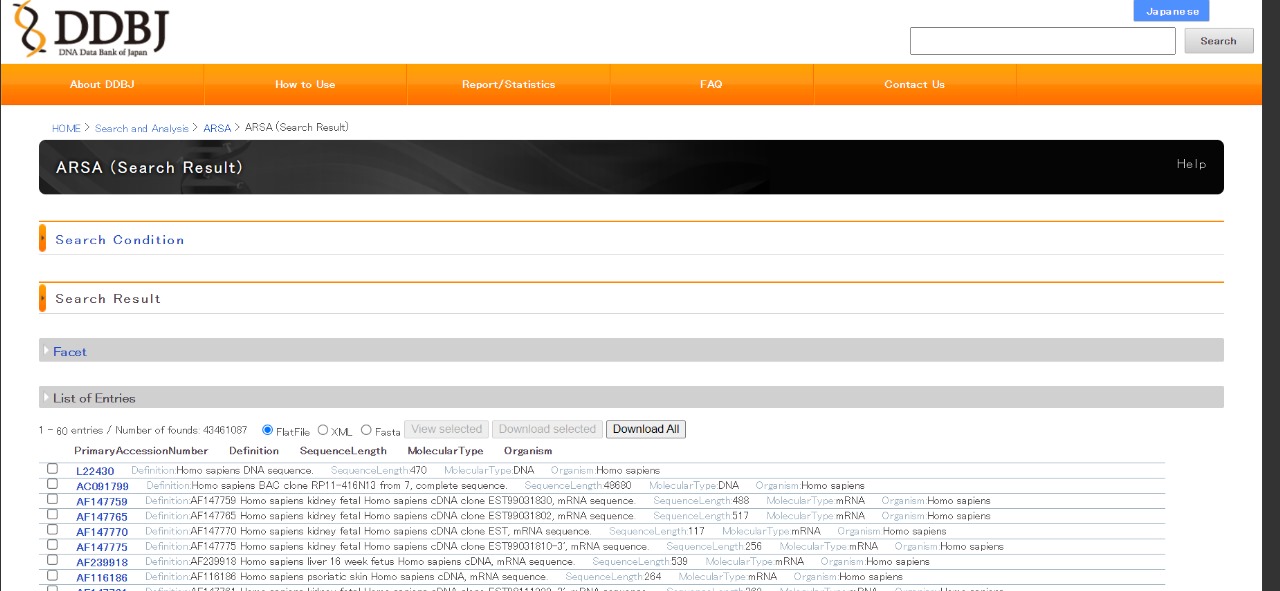
**EMBL**

**http://www.ddbj.nig.ac.jp/**

**http://www.ncbi.nlm.nih.gov/genbank/**

[**http://www.ebi.ac.uk/embl/**](http://www.ebi.ac.uk/embl/)

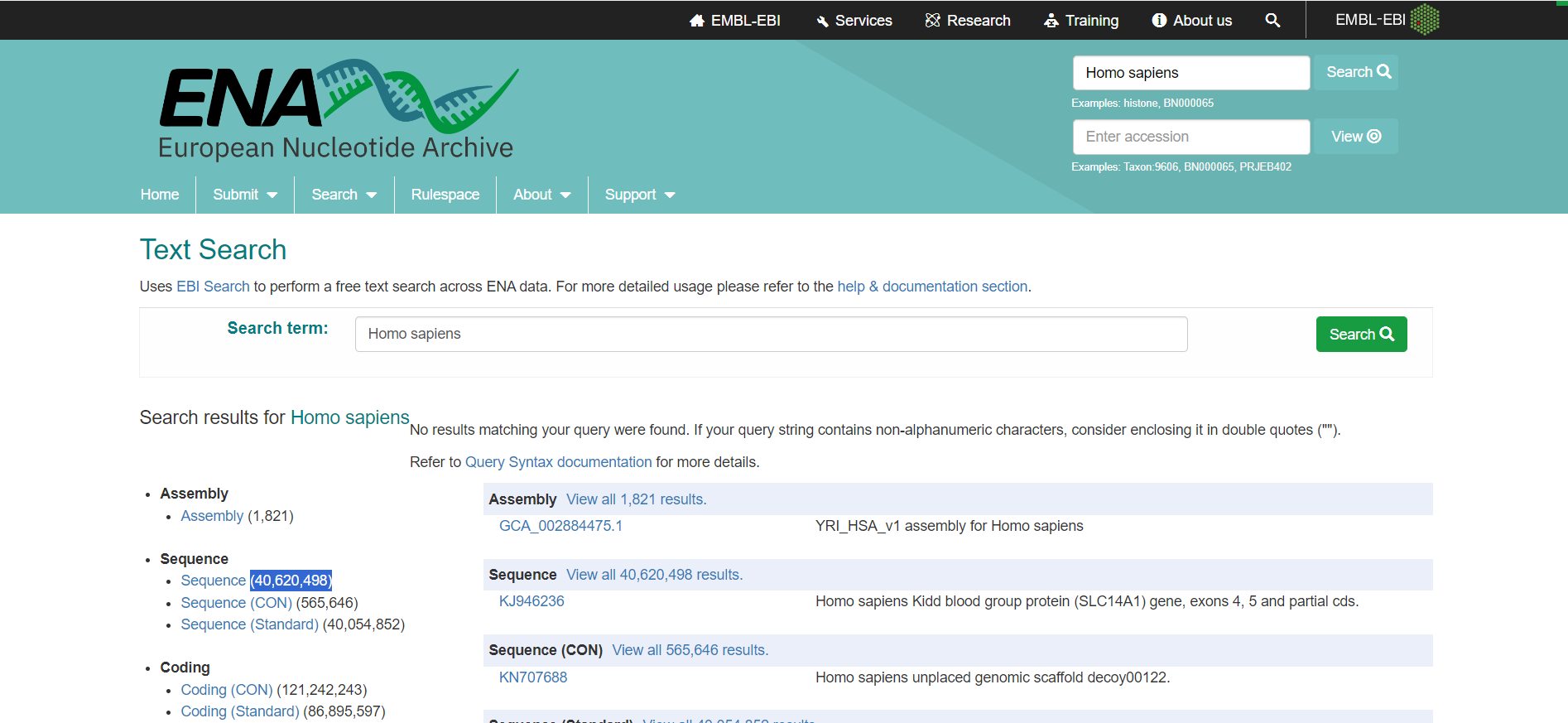
a) DDBJ: **43461087 sequences**



b) Gen bank: **64311072 sequences**



c) EMBL: **40620498**



**Comparison (based on number of sequences) :**

|  |  |  |
| --- | --- | --- |
| DDBJ | GenBank | EMBL |
| **43461087** | **64311072** | **40620498** |

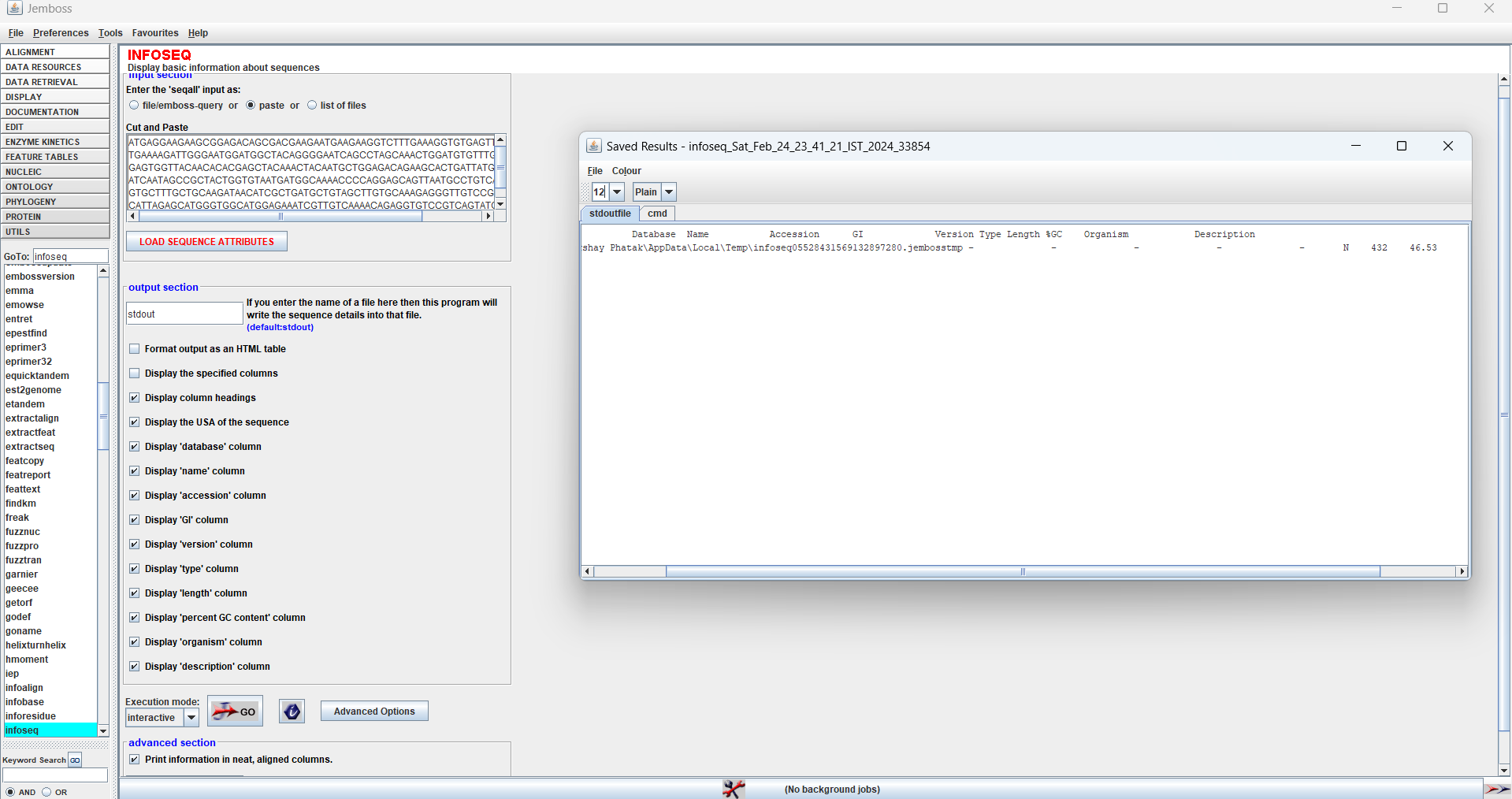
Thus GenBank has the highest number of deposited “Homo Sapiens” sequences

**GenBank>DDBJ>EMBL**

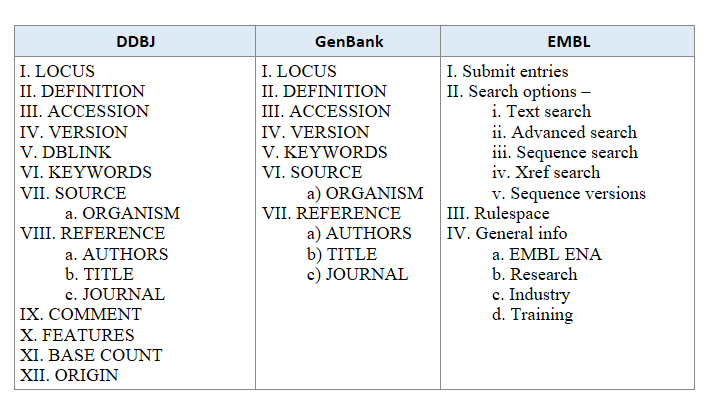
\*Based on number of sequences

## **Q2) What is the GC-content of the AY330867?**

GC content = **46.53%**



## **Q3** **Compare the contents in DDBJ, Genbank and EMBL**



The contents of EMBL will be nearly identical, although the main sources of nucleoid sequences are direct submissions from individual researchers, genome sequencing projects and patent application.   
Another notable distinction between the databases is that GeneBank doesn’t show base count while DDBJ, EMBL show it.

## **Q4) Get the papers about “discrimination of beta barrel membrane proteins”.**

## [**https://pubmed.ncbi.nlm.nih.gov/**](https://pubmed.ncbi.nlm.nih.gov/)

These two articles (out of 53) were found by **citation matching**

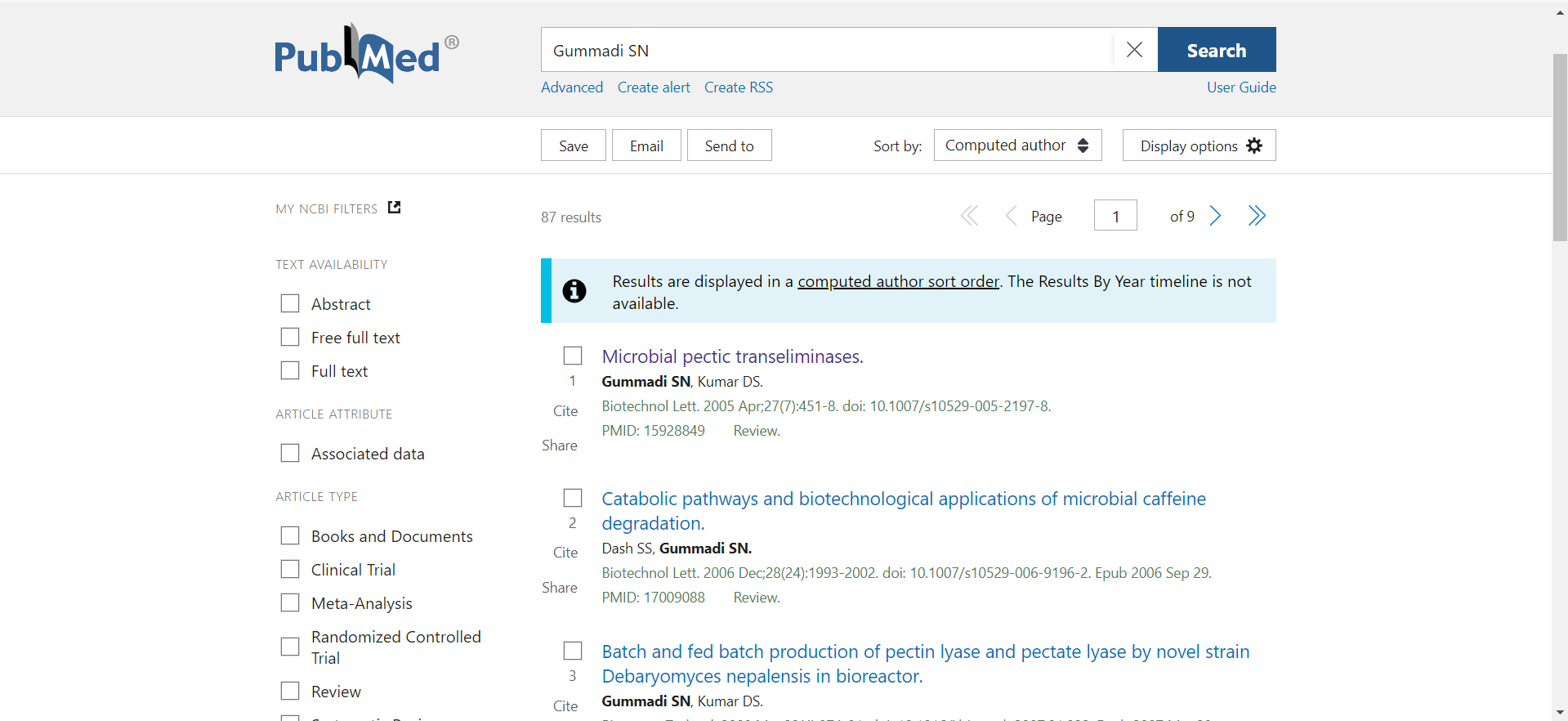
1. [TMBETADISC-RBF: Discrimination of beta-barrel membrane proteins using RBF networks and PSSM profiles.](https://pubmed.ncbi.nlm.nih.gov/18434251/)

Ou YY, et al. Comput Biol Chem. 2008. PMID: 18434251

1. [Current developments on beta-barrel membrane proteins: sequence and structure analysis, discrimination and prediction.](https://pubmed.ncbi.nlm.nih.gov/18220845/)

Gromiha MM, et al. Curr Protein Pept Sci. 2007. PMID: 18220845 Review.

## **Q5) The papers published by Dr.** [**Sathyanarayana N Gummadi**](https://pubmed.ncbi.nlm.nih.gov/?term=Gummadi+SN&cauthor_id=15928849) **Sir.**



## **Q6) How many related articles are listed in PUBMED for the paper, Cell 2008 Dec**

## **26;135(7):1158-9?**

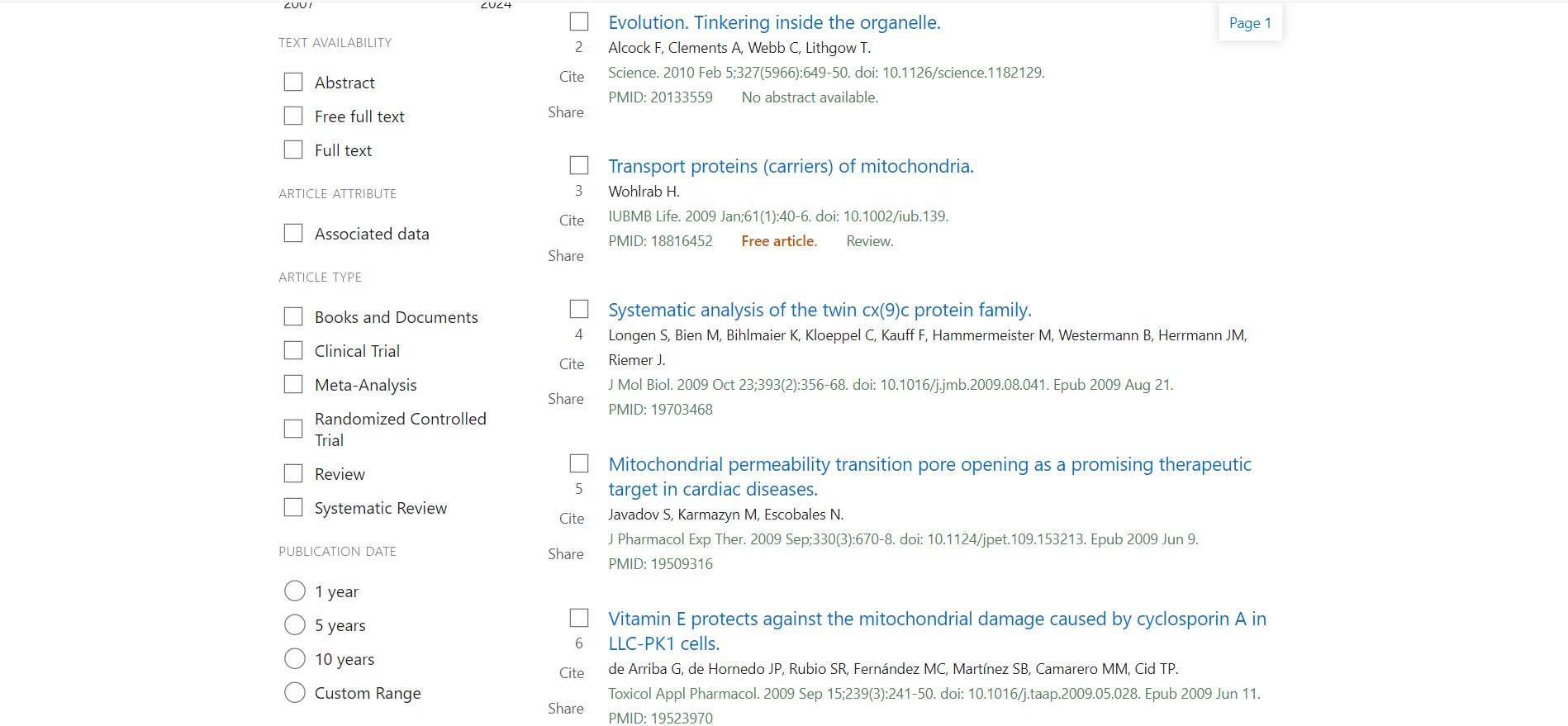
Following paper correspond to the given code.

https://pubmed.ncbi.nlm.nih.gov/19109882/

Authors: Kenichiro Imai, M Michael Gromiha, Paul Horton

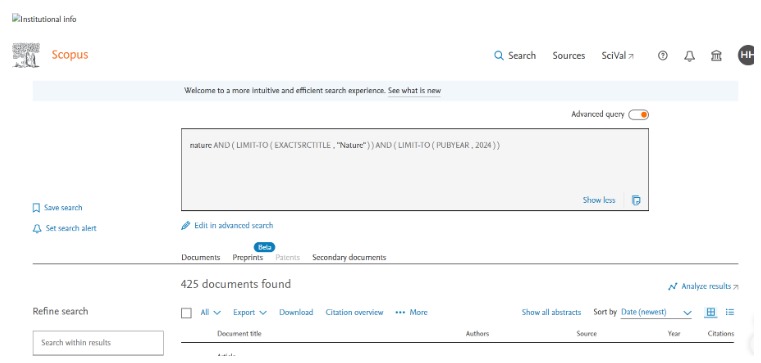


On scrolling down to the ‘Similar Articles’ section, total of **27** articles re found. The screenshot shows the **top 5** similar articles.

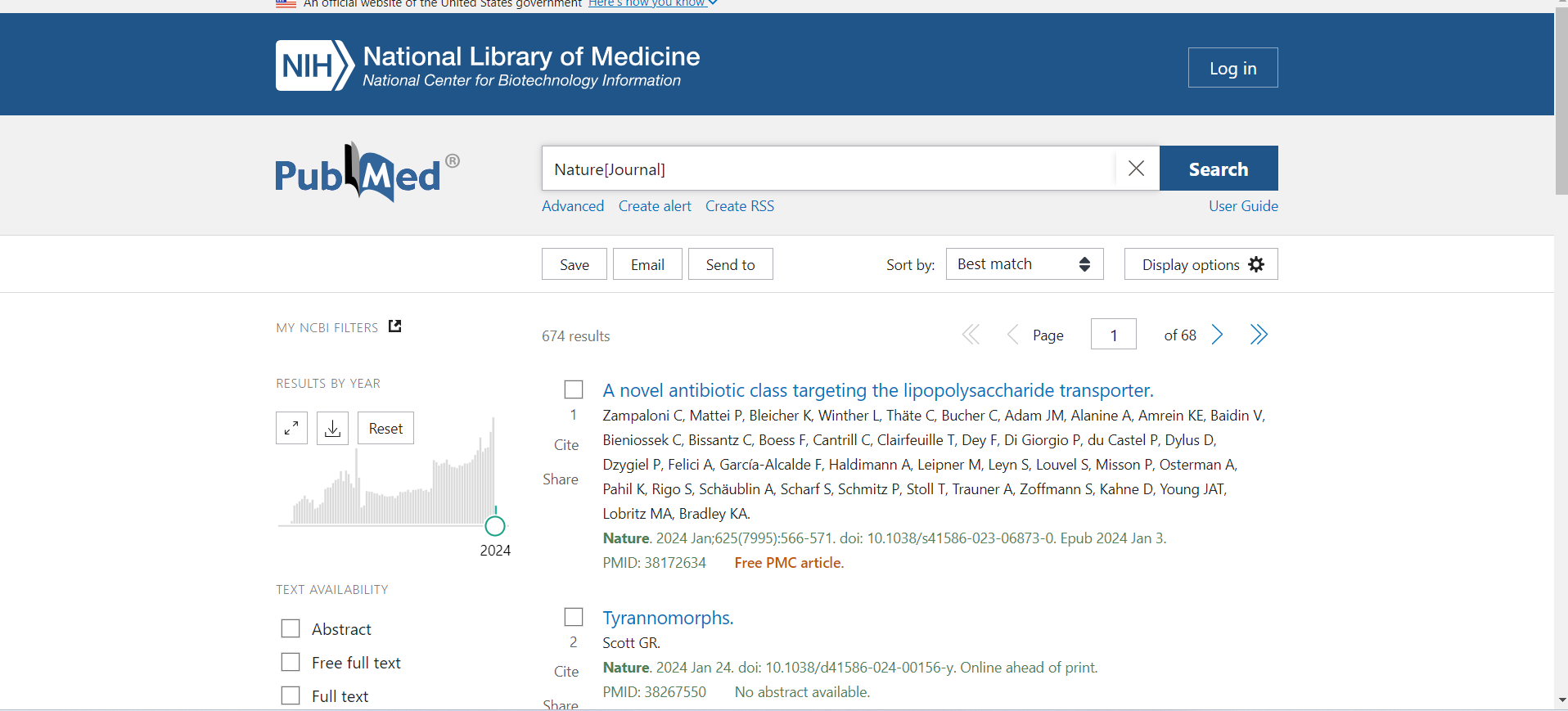


## **Q7) List the papers published in the journal “Nature” for the year 2024. Check the list in SCOPUS and PUBMED**

Total of articles have been published in the year 2024 till now in “Nature”.   
a) Scopus: 425 papers

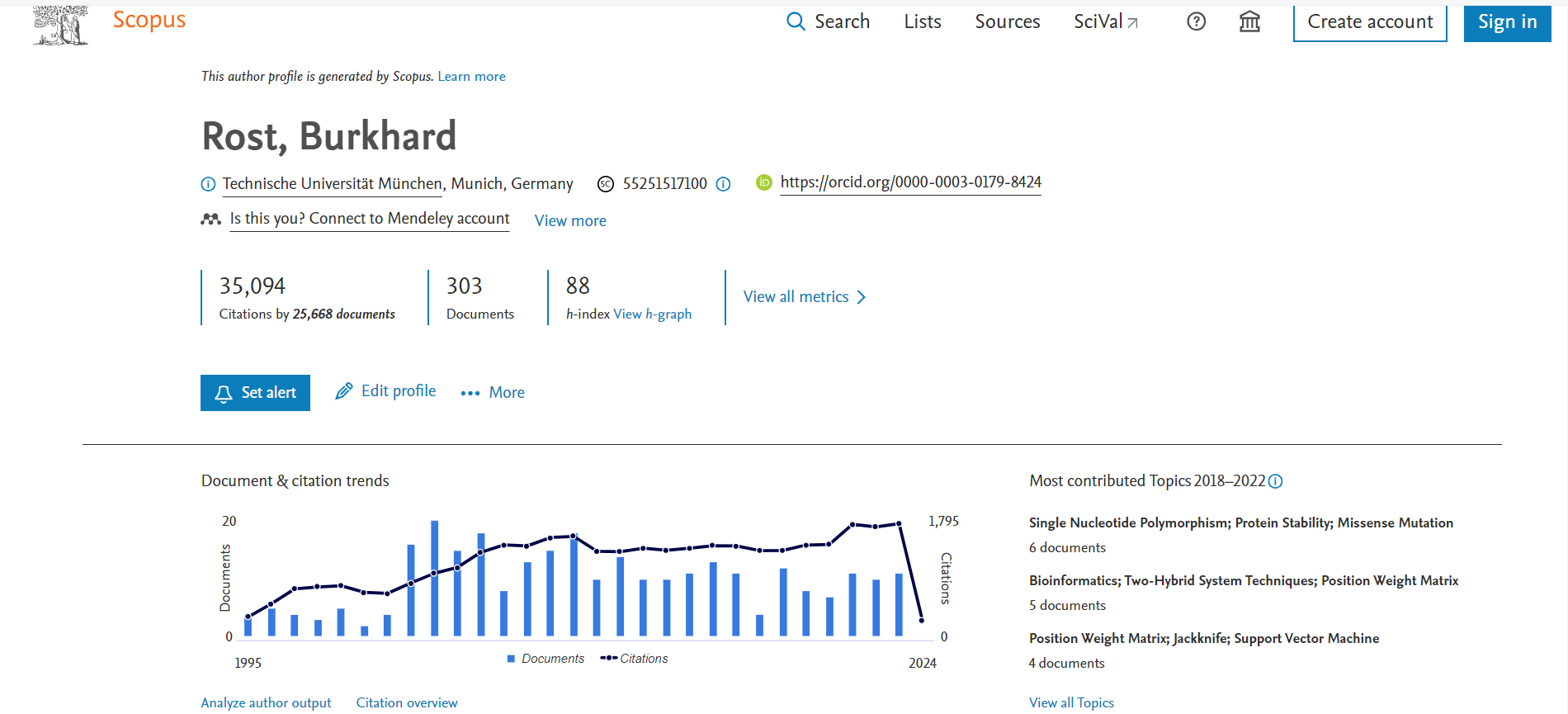


b) PubMed: 674 papers



## **Q8) Find the h-index and number of citations for “Burkhard Rost”.**

The given screenshot shows the profile of “Burkhard Rost” on Scopus. “Burkhard Rost” has a h-index of **88.**

****

## **Q9) Find the class of the enzyme EC 1.7.2.3 and its function**

## [**http://www.brenda-enzymes.org/**](http://www.brenda-enzymes.org/)

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

**Function:** It is the enzyme TMAO reductase, which reduces the cytochrome TorC . Also reduces dimethyl sulfoxide to dimethyl sulfide.



## **Q10) Find the catalytic site residues in Asparagine synthetase**

## **Hint: Find the EC number and search in Catalytic site atlas**

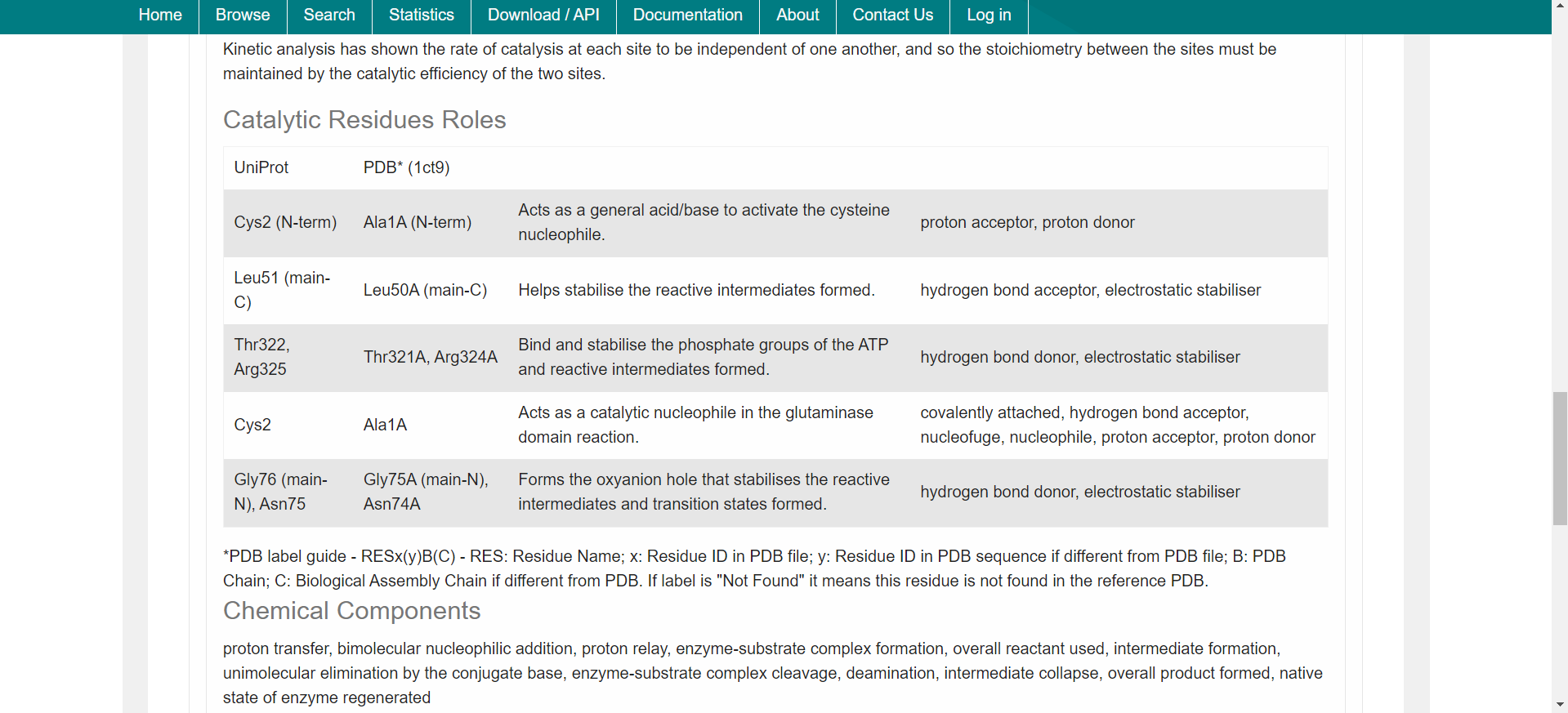
## [**https://www.ebi.ac.uk/thornton-srv/m-csa/**](https://www.ebi.ac.uk/thornton-srv/m-csa/)



## The EC number of Asparagine Synthetase is **6.3.5.4**



The Catalytic Site Residues in Asparagine Synthetase as given in Catalytic Site atlas are.



## **Q11) Find the scientific name, the taxonomy ID and the number of chromosomes for the following organisms. http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Root Organisms: Human, Cat, Dog, domestic guinea pig, and Thale cress**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No** | **Common Name** | **Scientific Name** | **Taxonomy ID** | **No. of Chromosomes (Total i.e 2n=)** |
|  | Human | *Homo sapiens* Linnaeus, 1758 | 9606 | 46 |
|  | Cat | *Felis catus* Linnaeus, 1758 | 9685 | 38 |
|  | Dog | *Canis lupus familiaris* Linnaeus, 1758 | 9615 | 78 |
|  | Domestic Guinea Pig | *Cavia porcellus* | 10141 | 64 |
|  | Thale Cress | *Arabidopsis thaliana* (L.) Heynh., 1842 | 3702 | 10 |

## **Q12) What are NCBI E-utilities? Give the syntax for fetching a record in FASTA format using**

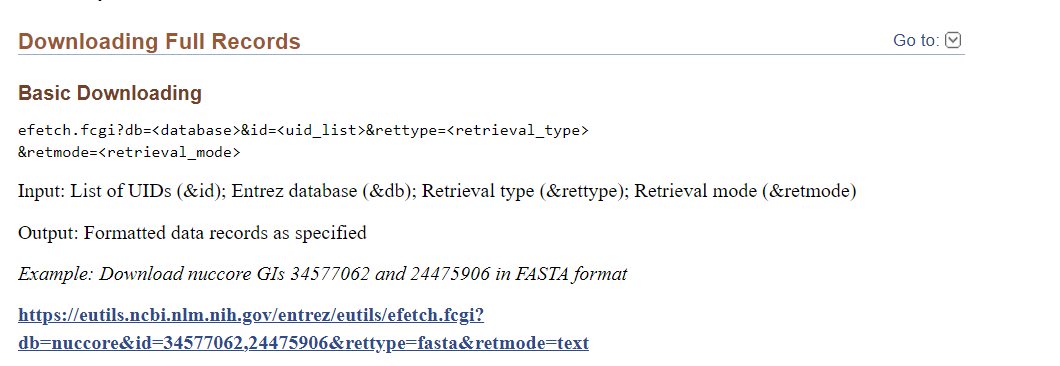
## **E-utilities.** [**http://www.ncbi.nlm.nih.gov/books/NBK25500**](http://www.ncbi.nlm.nih.gov/books/NBK25500)

The Entrez Programming Utilities (E-utilities) are a set of nine server-side programs that provide a stable interface into the Entrez query and database system at the National Center for Biotechnology Information (NCBI). The E-utilities use a fixed URL syntax that translates a standard set of input parameters into the values necessary for various NCBI software components to search for and retrieve the requested data. The E-utilities are therefore the structured interface to the Entrez system, which currently includes 38 databases covering a variety of biomedical data, including nucleotide and protein sequences, gene records, three-dimensional molecular structures, and the biomedical literature

Base URL

<https://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi>

Syntax for fetching a record in FASTA format:



## **Q13) List two databases under each of the following category. a. Protein properties, b. Small molecules (Structure related), c. Cancer gene databases Hint: Use Nucleic Acids Research (NAR) – ‘database category list’** [**https://www.oxfordjournals.org/our\_journals/nar/database/c/**](https://www.oxfordjournals.org/our_journals/nar/database/c/)

Two databases under the given category are as follows:

|  |  |  |
| --- | --- | --- |
| 1. | **Protein Properties** | AAindex |
| TOPPR |
| 2. | **Small molecule (Structure related)** | BitterDB |
| SuperToxic |
| 3. | **Cancer Gene Database** | ArrayMap |
| UMD-BRCA1/BRCA2 databases |