

# BT 3040: BIOINFORMATICS

## Assignment 2



Atharva Mandar Phatak | BE21B009  
Department of Biotechnology

Indian Institute of Technology  
Madras

# 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/>

a) DDBJ: 43461087 sequences

1 - 80 entries / Number of founds: 43461087 ☒ FlatFile ☐ XML ☐ Fasta

| PrimaryAccessionNumber | Definition   | SequenceLength        | MolecularType       | Organism               |
|------------------------|--|-----------------------|---------------------|------------------------|
| L22430                 | Definition: Homo sapiens DNA sequence.   | SequenceLength: 470   | MolecularType: DNA  | Organism: Homo sapiens |
| AC081798               | Definition: Homo sapiens BAC clone RP11-416N13 from 7, complete sequence.                            | SequenceLength: 49680 | MolecularType: DNA  | Organism: Homo sapiens |
| AF147758               | Definition: AF147758 Homo sapiens kidney fetal Homo sapiens cDNA clone EST9001800, mRNA sequence.    | SequenceLength: 488   | MolecularType: mRNA | Organism: Homo sapiens |
| AF147765               | Definition: AF147765 Homo sapiens kidney fetal Homo sapiens cDNA clone EST9001802, mRNA sequence.    | SequenceLength: 517   | MolecularType: mRNA | Organism: Homo sapiens |
| AF147770               | Definition: AF147770 Homo sapiens kidney fetal Homo sapiens cDNA clone EST, mRNA sequence.           | SequenceLength: 117   | MolecularType: mRNA | Organism: Homo sapiens |
| AF147775               | Definition: AF147775 Homo sapiens kidney fetal Homo sapiens cDNA clone EST9001810-3', mRNA sequence. | SequenceLength: 256   | MolecularType: mRNA | Organism: Homo sapiens |
| AF239818               | Definition: AF239818 Homo sapiens liver 18 week fetus Homo sapiens cDNA, mRNA sequence.              | SequenceLength: 539   | MolecularType: mRNA | Organism: Homo sapiens |
| AF116186               | Definition: AF116186 Homo sapiens psoriatic skin Homo sapiens cDNA, mRNA sequence.                   | SequenceLength: 284   | MolecularType: mRNA | Organism: Homo sapiens |

b) Gen bank: 64311072 sequences

Species: Animals (28,130,854), Plants (227,300), Fungi (318,204), Protists (175,583), Bacteria (18,728,620), Archaea (3,171), Viruses (4,018,632), Customize ...

Molecular types: genomic, DNA/RNA (38,818,220), mRNA (10,233,993), rRNA (1,306), Customize ...

Source databases: INSDC (GenBank) (42,314,621), RefSeq (9,384,625), Customize ...

Sequence Type: Nucleotide (41,244,935), EST (8,864,018), GSS (1,795,947)

Genetic compartments: Chromosome (1)

Summary: 20 per page, Sort by Default order

Items: 1 to 20 of 64311072

1. [Lynx pardinus genome assembly, contig: lp23s36493, whole genome shotgun sequence](#)  
Accession: CAAGRJ010006848.1 GI: 1603550370  
[BioProject](#) [BioSample](#) [Protein](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

2. [Homo sapiens Kidd blood group protein \(SLC14A1\) gene, exons 4, 5 and partial cds](#)  
Accession: KJ946236.1 GI: 683576128  
[Protein](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

3. [Homo sapiens syntaxin 16 isoform C-like protein \(STX16\) mRNA, 3' UTR](#)  
Accession: DQ268529.1 GI: 82780403  
[Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

4. [Homo sapiens hydroxymethylbilane synthase \(HMBS\), transcript variant 14, mRNA](#)  
Accession: DQ268529.1 GI: 82780403  
[Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

Filters: [Manage Filters](#)

Results by taxon

Top Organisms [Tree]

Homo sapiens (28472631)  
Severe acute respiratory syndrome-related coronavirus (8723757)  
Escherichia coli (4569390)  
Klebsiella pneumoniae (2334738)  
Acinetobacter baumannii (1596327)  
All other taxa (18614229)  
More...

Find related data

Database:

Find items

Search details

"Homo sapiens"[Organism] OR ("Homo sapiens"[Organism] OR Homo sapiens[All Fields])

c) EMBL: 40620498

The screenshot shows the EMBL-EBI website interface. At the top, there's a navigation bar with links to Home, Submit, Search, Rulespace, About, and Support. The main header features the ENA logo and a search bar with 'Homo sapiens' entered. Below the header, the 'Text Search' section is active, displaying search results for 'Homo sapiens'. The results are categorized into Assembly, Sequence, and Coding. The Sequence category is highlighted, showing 40,620,498 results. A specific result is shown: 'Homo sapiens Kidd blood group protein (SLC14A1) gene, exons 4, 5 and partial cds.' with accession number KJ946236.

Comparison (based on number of sequences) :

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

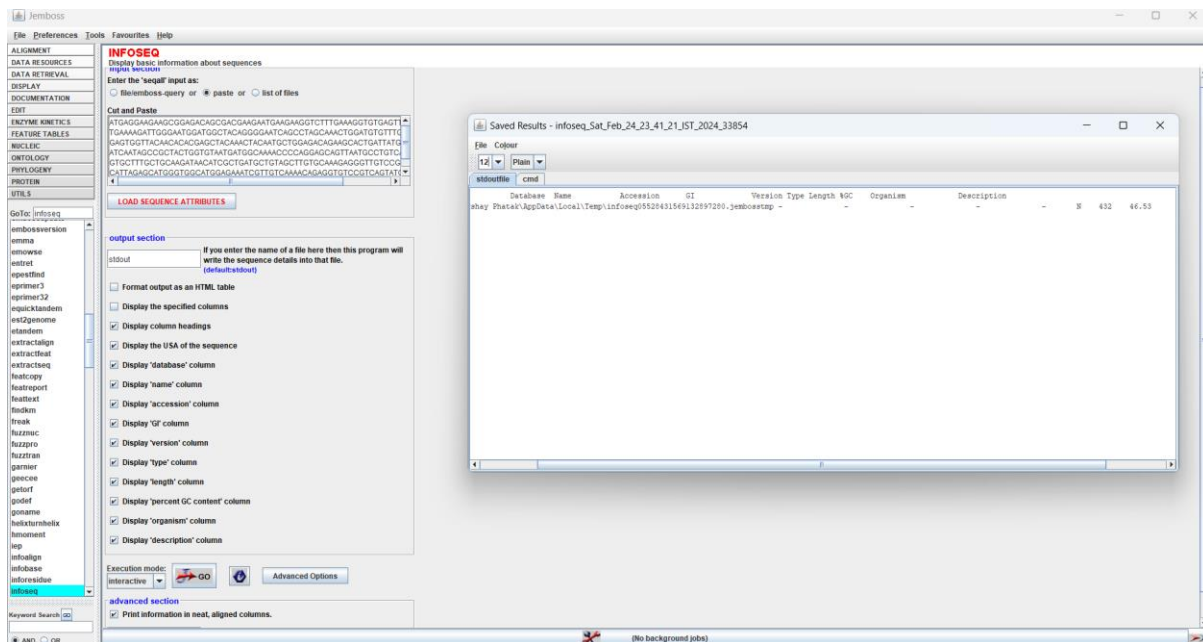
Thus GenBank has the highest number of deposited “Homo Sapiens” sequences

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

| DDBJ   | GenBank   | EMBL  |
|--|---|---|
| I. LOCUS<br>II. DEFINITION<br>III. ACCESSION<br>IV. VERSION<br>V. DBLINK<br>VI. KEYWORDS<br>VII. SOURCE<br>a. ORGANISM<br>VIII. REFERENCE<br>a. AUTHORS<br>b. TITLE<br>c. JOURNAL<br>IX. COMMENT<br>X. FEATURES<br>XI. BASE COUNT<br>XII. ORIGIN | I. LOCUS<br>II. DEFINITION<br>III. ACCESSION<br>IV. VERSION<br>V. KEYWORDS<br>VI. SOURCE<br>a) ORGANISM<br>VII. REFERENCE<br>a) AUTHORS<br>b) TITLE<br>c) JOURNAL | I. Submit entries<br>II. Search options –<br>i. Text search<br>ii. Advanced search<br>iii. Sequence search<br>iv. Xref search<br>v. Sequence versions<br>III. Rulespace<br>IV. General info<br>a. EMBL ENA<br>b. Research<br>c. Industry<br>d. Training |

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/>

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

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

#### Q5) The papers published by Dr. Sathyanarayana N Gummadi Sir.

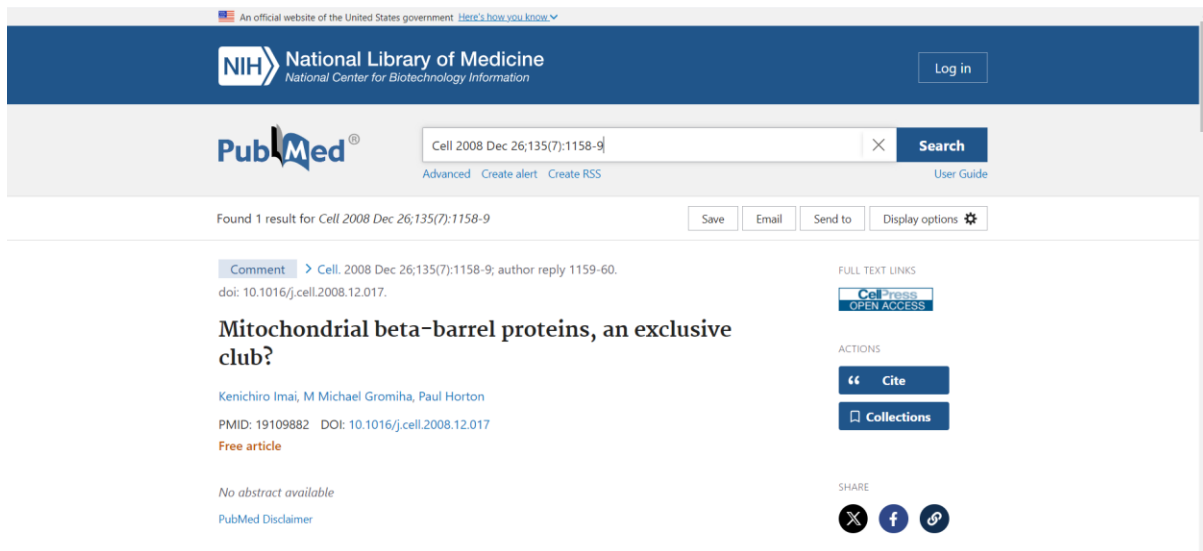
#### Q6) How many related articles are listed in PUBMED for the paper, Cell 2008 Dec

**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.

| 2007   | 2004  |  | Page 1 |
|--|-------|--|--------|
| TEXT AVAILABILITY                                    |       | <input type="checkbox"/> Evolution. Tinkering inside the organelle.  |        |
| <input type="checkbox"/> Abstract                    | 2     | Alcock F, Clements A, Webb C, Lithgow T.   |        |
| <input type="checkbox"/> Free full text              | Cite  | Science. 2010 Feb 5;327(5966):649-50. doi: 10.1126/science.1182129.  |        |
| <input type="checkbox"/> Full text                   | Share | PMID: 20133559 No abstract available.  |        |
| ARTICLE ATTRIBUTE                                    |       | <input type="checkbox"/> Transport proteins (carriers) of mitochondria.  |        |
| <input type="checkbox"/> Associated data             | 3     | Wohlrab H.   |        |
| ARTICLE TYPE   | Cite  | IUBMB Life. 2009 Jan;61(1):40-6. doi: 10.1002/iub.139.   |        |
| <input type="checkbox"/> Books and Documents         | Share | PMID: 18816452 Free article. Review.   |        |
| <input type="checkbox"/> Clinical Trial              |       | <input type="checkbox"/> Systematic analysis of the twin c(9)c protein family.   |        |
| <input type="checkbox"/> Meta-Analysis               | 4     | Longen S, Bien M, Bihlmaier K, Kloeppel C, Kauff F, Hammermeister M, Westermann B, Herrmann JM, Riemer J.                          |        |
| <input type="checkbox"/> Randomized Controlled Trial | Cite  | J Mol Biol. 2009 Oct 23;393(2):356-68. doi: 10.1016/j.jmb.2009.08.041. Epub 2009 Aug 21.   |        |
| <input type="checkbox"/> Review                      | Share | PMID: 19703468   |        |
| <input type="checkbox"/> Systematic Review           |       | <input type="checkbox"/> Mitochondrial permeability transition pore opening as a promising therapeutic target in cardiac diseases. |        |
| PUBLICATION DATE                                     | 5     | Javadov S, Karmazyn M, Escobales N.  |        |
| <input type="radio"/> 1 year                         | Cite  | J Pharmacol Exp Ther. 2009 Sep;330(3):670-8. doi: 10.1124/jpet.109.153213. Epub 2009 Jun 9.  |        |
| <input type="radio"/> 5 years                        | Share | PMID: 19509316   |        |
| <input type="radio"/> 10 years                       |       | <input type="checkbox"/> Vitamin E protects against the mitochondrial damage caused by cyclosporin A in LLC-PK1 cells.             |        |
| <input type="radio"/> Custom Range                   | 6     | de Arriba G, de Hornedo JP, Rubio SR, Fernández MC, Martínez SB, Camarero MM, Cid TP.  |        |
|  | Cite  | Toxicol Appl Pharmacol. 2009 Sep 15;239(3):241-50. doi: 10.1016/j.taap.2009.05.028. Epub 2009 Jun 11.                              |        |
|  | Share | PMID: 19523970   |        |

**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

Institutional info

Scopus

Welcome to a more intuitive and efficient search experience. See what is new

Advanced query ☐

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

Show less

Save search

Set search alert

Edit in advanced search

Documents Preprints Patents Secondary documents

425 documents found

Analyze results

Refine search

Search within results

Document title Authors Source Year Citations

b) PubMed: 674 papers

NIH National Library of Medicine National Center for Biotechnology Information

Log in

PubMed

Nature[Journal]

Search

Advanced Create alert Create RSS User Guide

Save Email Send to Sort by: Best match Display options

MY NCBI FILTERS

674 results

Page 1 of 68

RESULTS BY YEAR

2024

TEXT AVAILABILITY

Abstract

Free full text

Full text

1 A novel antibiotic class targeting the lipopolysaccharide transporter.

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, Haldemann 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.

Nature. 2024 Jan;625(7995):566-571. doi: 10.1038/s41586-023-06873-0. Epub 2024 Jan 3. PMID: 38172634 Free PMC article.

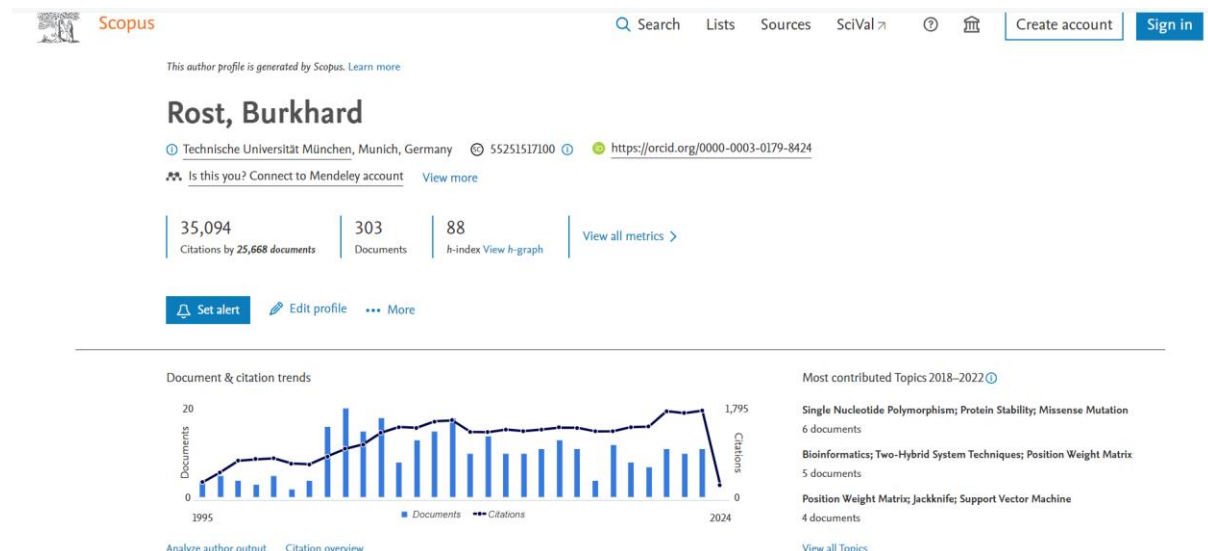
2 Tyrannomorphs.

Scott GR.

Nature. 2024 Jan 24. doi: 10.1038/d41586-024-00156-y. Online ahead of print. PMID: 38267550 No abstract available.

## 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/>

- 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.



Leibniz Institute DSMZ DSMZ Digital Diversity

Log in

**BRENDA**

Classic view All enzymes Enzyme history BRENDA support

BRENDA Home

Go to enzyme

show all hide all No of entries

- 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
- External Links

### Information on EC 1.7.2.3 - trimethylamine-N-oxide reductase

for references in articles please use BRENDA:EC1.7.2.3

**EC Tree**

- 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**

Contains bis(molybdopterin guanine dinucleotide)molybdenum cofactor. The reductant is a membrane-bound multiheme cytochrome c. Also reduces dimethyl sulfoxide to dimethyl sulfide.

**Specify your search results**

Mark a special word or phrase in this record:  **Mark!**

Search Reference ID:  **Search**

Search UniProt Accession:  **Search**

Select one or more organisms in this record: ☐

All organisms  
Attilavirion ficin

**Word Map** hide

cofactor-containing  
molybdoenzyme  
1.7.2.3  
twin-arginine  
tat-dependent

print visible entries  
print all entries  
show all entries

## 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/>

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

Log in

**BRENDA**

Classic view All enzymes Enzyme history BRENDA support

Refine search

### Search Enzyme Names (Synonyms)

Search term: Asparagine synthetase

Results 1 - 4 of 4

download as CSV  
download all results as CSV

| EC Number | Recommended Name                            | Synonyms  | Commentary |
|-----------|---|---|------------|
| 6.1.1.22  | asparagine-tRNA ligase                      | Asparagine synthetase A   | -          |
| 6.3.1.1   | aspartate-ammonia ligase                    | ammonia-dependent asparagine synthetase, Asparagine synthetase, Asparagine synthetase A, asparagine synthetase, ammonia-dependent, L-Asparagine synthetase, NH4+-dependent asparagine synthetase  | -          |
| 6.3.1.4   | aspartate-ammonia ligase (ADP-forming)      | Asparagine synthetase, Asparagine synthetase (adenosine diphosphate-forming), Asparagine synthetase (ADP-forming)   | -          |
| 6.3.5.4   | asparagine synthase (glutamine-hydrolysing) | Asparagine synthetase, Asparagine synthetase (glutamine hydrolysing), Asparagine synthetase (glutamine), Asparagine synthetase (glutamine-hydrolysing), asparagine synthetase 1, asparagine synthetase 2, Asparagine synthetase A, Asparagine synthetase B, asparagine synthetase, glutamine-dependent, asparagine synthetase1, asparagine synthetase2, bacterial type asparagine synthetase A, Glutamine-dependent asparagine synthetase, L-Asparagine synthetase, type -II asparagine synthetase, type I asparagine synthetase, type II asparagine synthetase | -          |

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

| Home | Browse | Search | Statistics | Download / API | Documentation | About | Contact Us | Log in |
|------|--------|--------|------------|----------------|---------------|-------|------------|--------|
|------|--------|--------|------------|----------------|---------------|-------|------------|--------|

Kinetic analysis has shown the rate of catalysis at each site to be independent of one another, and so the stoichiometry between the sites must be maintained by the catalytic efficiency of the two sites.

### 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.

### Chemical Components

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 state of enzyme regenerated

**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=) |
|--------|---------------------|--|-------------|------------------------------------|
| 1.     | Human               | <i>Homo sapiens</i><br>Linnaeus, 1758            | 9606        | 46                                 |
| 2.     | Cat                 | <i>Felis catus</i><br>Linnaeus, 1758             | 9685        | 38                                 |
| 3.     | Dog                 | <i>Canis lupus familiaris</i><br>Linnaeus, 1758  | 9615        | 78                                 |
| 4.     | Domestic Guinea Pig | <i>Cavia porcellus</i>                           | 10141       | 64                                 |
| 5.     | Thale Cress         | <i>Arabidopsis thaliana</i> (L.)<br>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>

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:

**Downloading Full Records**

Go to: 

**Basic Downloading**

```
efetch.fcgi?db=<database>&id=<uid_list>&rettype=<retrieval_type>
&retmode=<retrieval_mode>
```

Input: List of UIDs (&id); Entrez database (&db); Retrieval type (&rettype); Retrieval mode (&retmode)

Output: Formatted data records as specified

*Example: Download nuccore GIs 34577062 and 24475906 in FASTA format*

<https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=nuccore&id=34577062,24475906&rettype=fasta&retmode=text>

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