

BIOINFORMATICS PRACTICAL 2

G SHIVA RAM RAJ

BE18B023

1. DDBJ contains 32472146 Homo sapiens sequences on 15:11 pm, March 1st 2021. Whereas at the same point of time, Genbank has 48606814 Homo sapiens sequences and EMBL has 31035202 Homo sapiens sequences.
2. AK311779 has 129 A's, 92 C's, 133 G's and 121 T's.
Therefore the GC content is 47.368%.
3. DDBJ has the contents
 - I. LOCUS
 - II. DEFINITION
 - III. ACCESSION
 - IV. VERSION
 - V. DBLINK
 - VI. KEYWORDS
 - VII. SOURCE
 - a) ORGANISM
 - VIII. REFERENCE
 - a) AUTHORS
 - b) TITLE
 - c) JOURNAL
 - IX. COMMENT
 - X. FEATURES
 - XI. BASE COUNT
 - XII. ORIGIN

GenBank has the contents

- I. LOCUS
- II. DEFINITION
- III. ACCESSION
- IV. VERSION
- V. KEYWORDS
- VI. SOURCE
 - a) ORGANISM
- VII. REFERENCE
 - a) AUTHORS
 - b) TITLE
 - c) JOURNAL
- VIII. FEATURES
- IX. ORIGIN

EMBL will almost have the same contents. For EMBL, main sources for DNA and RNA sequences are direct submissions from individual researchers, genome sequencing projects and patent applications.

Most significant difference is that GenBank won't show the base count while DDBJ and EMBL will show the base count.

4. [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.

5. List of papers published by M. Michael Gromiha can be found in the following link, <https://pubmed.ncbi.nlm.nih.gov/?term=Gromiha+MM>

It has 220 papers with the name of Gromiha MM.

6. There are 337 similar articles to the given paper and it was cited by 102 articles.

7. In Pubmed there are 529 research papers published by Nature in 2021, whereas scopus has 346 research papers published by Nature in 2021.

8. The enzyme is amylo-alpha-1,6-glucosidase.

In 3.2.1.33, 3 refers to hydrolases, 3.2 refers to glycosylases, 3.2.1 refers to glycosidases i.e., enzymes that hydrolyze O- and S- glycosyl compounds and 3.2.1.33 refers to amylo-alpha-1,6-glucosidase.

This enzyme hydrolyses an unsubstituted glucose unit linked by an alpha(1->6) bond to an alpha(1->4) glucose chain. The enzyme activity found in mammals and yeast is in a polypeptide chain containing two active centres. The other activity is similar to that of EC 2.4.1.25 (4-alpha-glucanotransferase), which acts on the glycogen phosphorylase limit dextrin chains to expose the single glucose residues, which the 6-alpha-glucosidase activity can then hydrolyse. Together, these two activities constitute the glycogen debranching system.

9. The EC number of glutamate racemase is EC 5.1.1.3.

The catalytic residue roles as given in catalytic site atlas are

UniProt PDB*

Asp7	Asp7A	Acts as the general acid/base for Cys70 activation.	hydrogen bond acceptor, hydrogen bond donor, proton acceptor, activator, electrostatic stabiliser
Ser8	Ser8A	Activates Asp7	increase basicity, hydrogen bond donor, electrostatic stabiliser
Cys178	Cys178A	The catalytic general acid/base that re-protonates the substrate to produce the D-product. In the reverse reaction it deprotonates the D-substrate.	activator, hydrogen bond acceptor, proton acceptor, proton donor
Glu147	Glu147A (AA)	Re-protonates Cys178 to return the active site to a favourable state. In the reverse reaction it accepts a proton from Cys178 to initiate the reaction.	hydrogen bond acceptor, hydrogen bond donor, proton donor, activator, electrostatic stabiliser, increase basicity
Cys70	Cys70A	The catalytic general acid/base that deprotonates the L-substrate. In the reverse reaction, Cys70 re-protonates the substrate to produce the L-product.	hydrogen bond acceptor, hydrogen bond donor, proton acceptor, proton donor, activator, electrostatic stabiliser
His180	His180A	Activates Glu147	hydrogen bond donor, electrostatic stabiliser

10.

S No.	Organism	Scientific Name	Taxonomy ID	No. of chromosomes
1	Human	Homo sapiens	9606	46
2	Cat	Felis catus	9685	38
3	Dog	Canis lupus familiaris	9615	78
4	House mouse	Mus musculus	10090	40
5	Onion	Allium cepa	4679	16
6	Thale cress	Arabidopsis thaliana	3702	10

11. 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.

To fetch a record in FASTA format using E-utilities, we should use the syntax

```
efetch.fcgi?db=database&id=uid1,uid2,uid3&rettype=report_type&retmode=
data_mode
```

12. a) For Protein properties: MALISAM, ProTherm
- b) For Small molecules (Structure related): Hemolytik, SCIPDB
- c) For Cancer gene: CanSAR, MoKCa