# Practical 2

## Question 1

### a) DDBJ : 37,759,253



▶ **List of Entries**

1 - 210 entries / Number of founds: `37759253`  ⦿FlatFile ○XML ○Fasta  [View selected]  [Download selected]  [Download All]

| PrimaryAccessionNumber | Definition | SequenceLength | MolecularType | Organism |
|---|---|---|---|---|
| L22430 | Definition:Homo sapiens DNA sequence. | SequenceLength:470 | MolecularType:DNA | Organism:Homo sapiens |
| AC091799 | Definition:Homo sapiens BAC clone RP11-416N13 from 7, complete sequence. | SequenceLength:48680 | Molecu... | |
| AF116186 | Definition:AF116186 Homo sapiens psoriatic skin Homo sapiens cDNA, mRNA sequence. | SequenceLength:264 | | |
| AF147761 | Definition:AF147761 Homo sapiens kidney fetal Homo sapiens cDNA clone EST98111222-3', mRNA sequence. | Se... | | |
| AF147767 | Definition:AF147767 Homo sapiens kidney fetal Homo sapiens cDNA clone EST98111229-3', mRNA sequence. | Se... | | |
| AF147772 | Definition:AF147772 Homo sapiens kidney fetal Homo sapiens cDNA clone EST99031860, mRNA sequence. | Sequ... | | |
| AF147777 | Definition:AF147777 Homo sapiens kidney fetal Homo sapiens cDNA clone EST99031825, mRNA sequence. | Sequ... | | |
| AF005417 | Definition:AF005417 Homo sapiens placenta Homo sapiens cDNA, mRNA sequence. | SequenceLength:265 | Mole... | |
| AF486819 | Definition:AF486819 Homo sapiens thyroid adenoma Homo sapiens cDNA clone 199, mRNA sequence. | Sequence... | | |
| AY032902 | Definition:AY032902 Homo sapiens library (Geerts D) Homo sapiens cDNA clone IMAGE:1847576 similar to ACF7; n... | | | |

Organism:Homo sapiens

### b) Genbank : 27,979,315



**Species**
Animals (27,922,038)
Viruses (4,545)
Customize ...

**Molecule types**
genomic
  DNA/RNA (15,509,785)
mRNA (9,823,382)
rRNA (46)
Customize ...

**Source databases**
INSDC (GenBank)
  (27,628,585)
RefSeq (194,637)

Summary ⏷   20 per page ⏷   Sort by Default order ⏷                    Send to: ⏷

**Items: 1 to 20 of 27979315**

<< First   < Prev   Page 1   of 1398966   Next >   Last >>

☐ JP 2021525076-A/26: SHARED ANTIGENS
1.   2,339 bp linear DNA
     Accession: PA344065.1  GI: 2184295280

     GenBank   FASTA   Graphics

☐ JP 2021525076-A/24: SHARED ANTIGENS
2.   3,178 bp linear DNA
     Accession: PA344063.1  GI: 2184295278

### c) EMBL: 37,723,166



**Sequence**
(showing first 1,000,000 of 37,723,166 results)

**Download ENA records:**   FASTA   TEXT   TSV

| Accession | Description/Title |
|---|---|
| JX420610 | Homo sapiens clone HGNC-38N-2 immunoglobulin kappa light chain variable region mRNA, partial cds. |
| JX420611 | Homo sapiens clone HGNC-38N-3 immunoglobulin kappa light chain variable region mRNA, partial cds. |
| JX420609 | Homo sapiens clone HGNC-38N-1 immunoglobulin kappa light chain variable region mRNA, partial cds. |

## Question 2

I found out the sequence of AY330867 from
https://www.ncbi.nlm.nih.gov/nuccore/AY330867

I then used "geecee" program of emboss to calculate the GC content,
which is **0.47**

```
[(base) Ajeet@MacBook-touch Downloads % geecee
Calculate fractional GC content of nucleic acid sequences
Input nucleotide sequence(s): seq.fasta
Output file [ay330867.geecee]: ans
[(base) Ajeet@MacBook-touch Downloads % more ans
#Sequence    GC content
AY330867.1    0.47
ans (END)
```

## Question 3

The screenshots of contents of the 3 databases for AY330867 can be
found below in the order - DDBJ, Genbank, EMBL. All are the same.

```
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             Location/Qualifiers
     source          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="MRKKRRQRRRMKKVFERCELARTLKRLGMDGYRGISLANWMCLA
                     KWESGYNTRATNYNAGDRSTDYGIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIA
                     DAVACAKRVVRDPQGIRAWVAWRNRCQNRGVRQYVQGCGV"
BASE COUNT      131 a          72 c          129 g          100 t
ORIGIN
        1 atgaggaaga agcggagaca gcgacgaaga atgaagaagg tctttgaaag gtgtgagttg
       61 gccagaactc tgaaaagatt gggaatggat ggctacaggg gaatcagcct agcaaactgg
      121 atgtgtttgg ccaaatggga gagtggttac aacacacgag ctacaaacta caatgctgga
      181 gacagaagca ctgattatgg gatatttcag atcaataagc gctactggtg taatgatggc
      241 aaaaccccag gagcagttaa tgcctgtcat ttatcctgca gtgctttgct gcaagataac
      301 atcgctgatg ctgtagcttg tgcaaagagg gttgtccgtg atccacaagg cattagagca
      361 tgggtggcat ggagaaatcg ttgtcaaaac agaggtgtcc gtcagtatgt tcaaggttgt
      421 ggagtgtaat ga
//
```

# Synthetic construct human lysozyme mRNA, complete cds
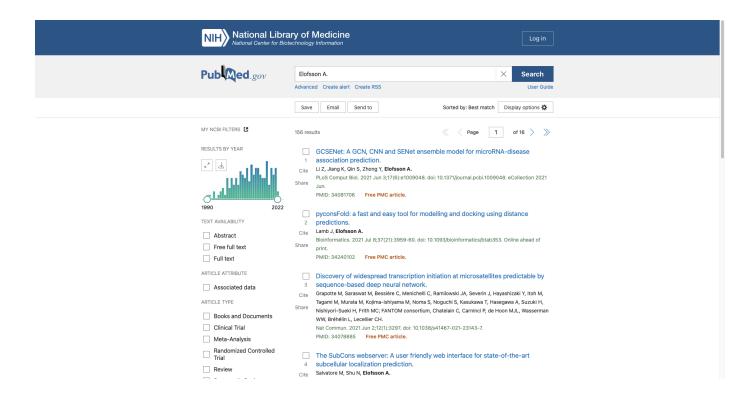
GenBank: AY330867.1

FASTA   Graphics

Go to: ⌄

```
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             Location/Qualifiers
     source          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="MRKKRRQRRRMKKVFERCELARTLKRLGMDGYRGISLANWMCLA
                     KWESGYNTRATNYNAGDRSTDYGIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIA
                     DAVACAKRVVRDPQGIRAWVAWRNRCQNRGVRQYVQGCGV"
ORIGIN
        1 atgaggaaga agcggagaca gcgacgaaga atgaagaagg tctttgaaag gtgtgagttg
       61 gccagaactc tgaaaagatt gggaatggat ggctacaggg gaatcagcct agcaaactgg
      121 atgtgtttgg ccaaatggga gagtggttac aacacacgag ctacaaacta caatgctgga
      181 gacagaagca ctgattatgg gatatttcag atcaatagcc gctactggtg taatgatggc
      241 aaaaccccag gagcagttaa tgcctgtcat ttatcctgca gtgctttgct gcaagataac
      301 atcgctgatg ctgtagcttg tgcaaagagg gttgtccgtg atccacaagg cattagagca
      361 tgggtggcat ggagaaatcg ttgtcaaaac agaggtgtcc gtcagtatgt tcaaggttgt
      421 ggagtgtaat ga
//
```

## Question 4

[Papers](#) on discrimination of beta barrel membrane proteins.



## Question 5

[Papers](#) by Effson A.

Publ**Med**.gov

Elofsson A.    ✕    Search

Advanced    Create alert    Create RSS                    User Guide

Save    Email    Send to                    Sorted by: Best match    Display options ⚙

MY NCBI FILTERS ⧉

156 results                    ≪  <  Page  1  of 16  >  ≫

RESULTS BY YEAR

1990                    2022

TEXT AVAILABILITY
☐ Abstract
☐ Free full text
☐ Full text

ARTICLE ATTRIBUTE
☐ Associated data

ARTICLE TYPE
☐ Books and Documents
☐ Clinical Trial
☐ Meta-Analysis
☐ Randomized Controlled Trial
☐ Review

1 ☐ GCSENet: A GCN, CNN and SENet ensemble model for microRNA-disease association prediction.
Cite  Li Z, Jiang K, Qin S, Zhong Y, **Elofsson A.**
Share  PLoS Comput Biol. 2021 Jun 3;17(6):e1009048. doi: 10.1371/journal.pcbi.1009048. eCollection 2021 Jun.
PMID: 34081706    Free PMC article.

2 ☐ pyconsFold: a fast and easy tool for modelling and docking using distance predictions.
Cite  Lamb J, **Elofsson A.**
Share  Bioinformatics. 2021 Jul 8;37(21):3959-60. doi: 10.1093/bioinformatics/btab353. Online ahead of print.
PMID: 34240102    Free PMC article.

3 ☐ Discovery of widespread transcription initiation at microsatellites predictable by sequence-based deep neural network.
Cite  Grapotte M, Saraswat M, Bessière C, Menichelli C, Ramilowski JA, Severin J, Hayashizaki Y, Itoh M,
Share  Tagami M, Murata M, Kojima-Ishiyama M, Noma S, Noguchi S, Kasukawa T, Hasegawa A, Suzuki H, Nishiyori-Sueki H, Frith MC; FANTOM consortium, Chatelain C, Carninci P, de Hoon MJL, Wasserman WW, Bréhélin L, Lecellier CH.
Nat Commun. 2021 Jun 2;12(1):3297. doi: 10.1038/s41467-021-23143-7.
PMID: 34078885    Free PMC article.

4 ☐ The SubCons webserver: A user friendly web interface for state-of-the-art subcellular localization prediction.
Cite  Salvatore M, Shu N, **Elofsson A.**

# Question 6

There are [28 related articles](#) of the paper.

Publ**Med**.gov

Search

Advanced                    User Guide

Save    Email    Send to                    Sorted by: Best match    Display options ⚙

MY NCBI FILTERS ⧉

**Similar articles for PMID: 19109882**

28 results                    ≪  <  Page  1  of 3  >  ≫

RESULTS BY YEAR

2007                    2022

TEXT AVAILABILITY
☐ Abstract
☐ Free full text
☐ Full text

ARTICLE ATTRIBUTE
☐ Associated data

ARTICLE TYPE
☐ Books and Documents
☐ Clinical Trial
☐ Meta-Analysis
☐ Randomized Controlled Trial
☐ Review

1 ☐ Mitochondrial beta-barrel proteins, an exclusive club?
Cite  Imai K, Gromiha MM, Horton P.
Share  Cell. 2008 Dec 26;135(7):1158-9; author reply 1159-60. doi: 10.1016/j.cell.2008.12.017.
PMID: 19109882    Free article.    No abstract available.

2 ☐ Evolution. Tinkering inside the organelle.
Cite  Alcock F, Clements A, Webb C, Lithgow T.
Share  Science. 2010 Feb 5;327(5966):649-50. doi: 10.1126/science.1182129.
PMID: 20133559    No abstract available.

3 ☐ Transport proteins (carriers) of mitochondria.
Cite  Wohlrab H.
Share  IUBMB Life. 2009 Jan;61(1):40-6. doi: 10.1002/iub.139.
PMID: 18816452    Free article.    Review.

4 ☐ Systematic analysis of the twin cx(9)c protein family.
Longen S, Bien M, Bihlmaier K, Kloeppel C, Kauff F, Hammermeister M, Westermann B, Herrmann
Cite  JM, Riemer J.
Share  J Mol Biol. 2009 Oct 23;393(2):356-68. doi: 10.1016/j.jmb.2009.08.041. Epub 2009 Aug 21.
PMID: 19703468

5 ☐ Betulinic acid induces cytochrome c release and apoptosis in a Bax/Bak-independent, permeability transition pore dependent fashion.

**Question 7**

According to Pubmed : 326 and according to scopus : 124



**Question 8**

H-index : 103
Citations : 47150

## Question 9

Class : Hydrolase
Function : Release of the N-terminal residue from a tripeptide.
Source

## Question 10

EC number : 6.3.5.4

## 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 |

## Question 11

<u>Human</u>
Scientific Name : Homo Sapiens
Taxonomy ID : 9606
Chromosomes : 46

<u>Cat</u>
Scientific Name : *Felis catus*
Taxonomy ID : 9685
Chromosomes : 38

<u>Dog</u>
Scientific Name : Canis lupus familiaris
Taxonomy ID : 9615
Chromosomes : 78

<u>House mouse</u>
Scientific Name : Mus musculus
Taxonomy ID : 10090
Chromosomes : 40

Onion
Scientific Name : Allium Cepa
Taxonomy ID : 4679
Chromosomes : 16

Thale Cress
Scientific Name : Arabidopsis thaliana
Taxonomy ID : 3702
Chromosomes : 5


**Question 12**

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. [Source]

Syntax for fetching a record in FASTA format:

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

&retmode=<retrieval_mode>
```

retrieval_type = fasta for fasta format.

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

**Question 13**

Databases:
    a) Protein properties : AAindex, BindingDB
    b) Small molecules (Structure Related) : BitterDB, ChemBank
    c) Cancer gene: ArrayMap, BCCTBbp