

Practical 2

Question 1

a) DDBJ : 37,759,253

List of Entries					
1 - 210 entries / Number of founds: 37759253 <input checked="" type="radio"/> FlatFile <input type="radio"/> XML <input type="radio"/> Fasta <input type="button" value="View selected"/> <input type="button" value="Download selected"/> <input type="button" value="Download All"/>					
PrimaryAccessionNumber	Definition	SequenceLength	MolecularType	Organism	
<input type="checkbox"/> L22430	Definition:Homo sapiens DNA sequence.	SequenceLength:470	MolecularType:DNA	Organism:Homo sapiens	
<input type="checkbox"/> AC091799	Definition:Homo sapiens BAC clone RP11-416N13 from 7, complete sequence.	SequenceLength:48680	MolecularType:DNA	Organism:Homo sapiens	
<input type="checkbox"/> AF116186	Definition:AF116186 Homo sapiens psoriatic skin Homo sapiens cDNA, mRNA sequence.	SequenceLength:264	MolecularType:mRNA	Organism:Homo sapiens	
<input type="checkbox"/> AF147761	Definition:AF147761 Homo sapiens kidney fetal Homo sapiens cDNA clone EST98111222-3', mRNA sequence.	SequenceLength:264	MolecularType:mRNA	Organism:Homo sapiens	
<input type="checkbox"/> AF147767	Definition:AF147767 Homo sapiens kidney fetal Homo sapiens cDNA clone EST98111229-3', mRNA sequence.	SequenceLength:264	MolecularType:mRNA	Organism:Homo sapiens	
<input type="checkbox"/> AF147772	Definition:AF147772 Homo sapiens kidney fetal Homo sapiens cDNA clone EST99031860, mRNA sequence.	SequenceLength:264	MolecularType:mRNA	Organism:Homo sapiens	
<input type="checkbox"/> AF147777	Definition:AF147777 Homo sapiens kidney fetal Homo sapiens cDNA clone EST99031825, mRNA sequence.	SequenceLength:264	MolecularType:mRNA	Organism:Homo sapiens	
<input type="checkbox"/> AF005417	Definition:AF005417 Homo sapiens placenta Homo sapiens cDNA, mRNA sequence.	SequenceLength:265	MolecularType:mRNA	Organism:Homo sapiens	
<input type="checkbox"/> AF486819	Definition:AF486819 Homo sapiens thyroid adenoma Homo sapiens cDNA clone 199, mRNA sequence.	SequenceLength:264	MolecularType:mRNA	Organism:Homo sapiens	
<input type="checkbox"/> AY032902	Definition:AY032902 Homo sapiens library (Geerts D) Homo sapiens cDNA clone IMAGE:1847576 similar to ACF7; n	SequenceLength:264	MolecularType:mRNA	Organism:Homo sapiens	

b) Genbank : 27,979,315

Species	Summary ▾ 20 per page ▾ Sort by Default order ▾	Send to: ▾
Animals (27,922,038)		
Viruses (4,545)		
Customize ...		
Items: 1 to 20 of 27979315		
Molecule types	<< First < Prev Page 1 of 1398966 Next > Last >>	
genomic		
DNA/RNA (15,509,785)	<input type="checkbox"/> JP 2021525076-A/26: SHARED ANTIGENS	
mRNA (9,823,382)	1. 2,339 bp linear DNA	
rRNA (46)	Accession: PA344065.1 GI: 2184295280	
Customize ...	GenBank FASTA Graphics	
Source databases	<input type="checkbox"/> JP 2021525076-A/24: SHARED ANTIGENS	
INSDC (GenBank)	2. 3,178 bp linear DNA	
(27,628,585)	Accession: PA344063.1 GI: 2184295278	
RefSeq (194,637)		

c) EMBL: 37,723,166

Sequence		Download ENA records:	FASTA	TEXT	TSV
(showing first 1,000,000 of 37,723,166 results)					
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="MRKKRRQRRRMKKVFERCELARTLKRRLGMDGYRGISLANWMCIA
KWESGYNTRATNYNAGDRSTDYGIQINSRYWCNDGKTPGAVNACHLSCSALLQDNIA
DAVACAKRVVRDPQGIRAWVAWRNRQNRGVRQYVQCGV"
BASE COUNT        131 a           72 c           129 g           100 t
ORIGIN
1 atgaggaaga agcggagaca ggcacgaaga atgaagaagg tctttgaaag gtgtgagttg
61 gccagaaactc tgaaaagatt gggaaatgat ggctacaggg gaatcagcct agcaaaactgg
121 atgtgtgttgg ccaaatggga gagtgggttac aacacacgag ctacaaacta caatgctgga
181 gacagaagca ctgattatgg gatatttcag atcaatagcc gctaactggtg taatgatggc
241 aaaacccccag gacgagttaa tgcctgtcat ttatcctgca gtgctttgct gcaagataac
301 atcgtgatg ctgtagcttg tgcaaaagagg 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="MRKKRRQRRMKVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGYNTRATNYNAGDRSTDYGFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIA
DAVACAKRVVRDPQGIRAWVAWRNRCQNRGVRQYVQGC GV"

ORIGIN
1 atgaggaaga agcggagaca gcgacgaaga atgaagaagg tctttgaaag gtgtgagttg
61 gccagaactc tgaaaaagatt gggaatggat ggctacaggg gaatcagcct agcaaaactgg
121 atgtgtttgg ccaaatggga gagtgggttac aacacacgag ctacaaacta caatgctgga
181 gacagaagca ctgattatgg gatatttcag atcaatagcc gctactgggt taatgatggc
241 aaaacccag gagcagttaa tgcctgtcat ttatcctgca gtgctttgct gcaagataac
301 atcgctgatg ctgtagcttg tgcaaaagagg gttgtccgtg atccacaagg cattagagca
361 tgggtggcat ggagaaatcg ttgtcaaac agaggtgtcc gtcagtatgt tcaaggttgt
421 ggagtgtaat ga
//

ID AY330867; SV 1; linear; mRNA; STD; SYN; 432 BP.
XX
AC AY330867;
XX
DT 26-JUL-2003 (Rel. 76, Created)
DT 30-JUL-2004 (Rel. 80, Last updated, Version 2)
XX
DE Synthetic construct human lysozyme mRNA, complete cds.
XX
KW .
XX
OS synthetic construct
OC other sequences; artificial sequences.
XX
RN [1]
RP 1-432
RA Xie X.D., Chen Z.H., Zhou C.S.;
RT ;
RL Submitted (26-JUN-2003) to the INSDC.
RL Postdoctoral Scientific Research Station of Gansu Yasheng Industrial
RL (Group) Co., Ltd., Zhangye Road, Lanzhou, Gansu 730030, P.R.China
XX
DR MD5: 06195aec74e85195c05c9ba6525a4d83.
XX
FH Key Location/Qualifiers
FH
FT source 1..432
FT /organism="synthetic construct"
FT /mol_type="mRNA"
FT /db_xref="taxon:32630"
FT CDS 1..429
FT /codon_start=1
FT /transl_table=11
FT /product="human lysozyme"
FT /protein_id="AAP93336.1"
FT /translation="MRKKRRQRRMKVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGYNTRATNYNAGDRSTDYGFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIA
FT VACAKRVVRDPQGIRAWVAWRNRCQNRGVRQYVQGC GV"
XX
SQ Sequence 432 BP; 131 A; 72 C; 129 G; 100 T; 0 other;
atgaggaaga agcggagaca gcgacgaaga atgaagaagg tctttgaaag gtgtgagttg 60
gccagaactc tgaaaaagatt gggaatggat ggctacaggg gaatcagcct agcaaaactgg 120
atgtgtttgg ccaaatggga gagtgggttac aacacacgag ctacaaacta caatgctgga 180
gacagaagca ctgattatgg gatatttcag atcaatagcc gctactgggt taatgatggc 240
aaaacccag gagcagttaa tgcctgtcat ttatcctgca gtgctttgct gcaagataac 300
atcgctgatg ctgtagcttg tgcaaaagagg gttgtccgtg atccacaagg cattagagca 360
tgggtggcat ggagaaatcg ttgtcaaac agaggtgtcc gtcagtatgt tcaaggttgt 420
ggagtgtaat ga 432
//

Question 4

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

NIH National Library of Medicine
National Center for Biotechnology Information

PubMed.gov

discrimination of beta barrel membrane proteins

Search

Advanced Create alert Create RSS User Guide

Save Email Send to Sorted by: Best match Display options

MY NCBI FILTERS

RESULTS BY YEAR

1992 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

50 results

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

☐ Landscape of Eukaryotic Transmembrane Beta Barrel Proteins.
1 Roumia AF, Theodoropoulou MC, Tsirogas KD, Nielsen H, Bagos PG.
J Proteome Res. 2020 Mar 6;19(3):1209-1221. doi: 10.1021/acs.jproteome.9b00740. Epub 2020 Feb 11.
Share PMID: 32008325
Even though in the last few years several families of eukaryotic **beta-barrel** outer **membrane proteins** have been discovered, their computational characterization and their annotation in public databases are far from complete. ...Twelve characteristic pHM ...

☐ Predicting Beta Barrel Transmembrane Proteins Using HMMs.
2 Tsaousis GN, Hamodrakas SJ, Bagos PG.
Methods Mol Biol. 2017;1552:43-61. doi: 10.1007/978-1-4939-6753-7_4.
Share PMID: 28224490
Transmembrane **beta-barrels** (TMBBs) constitute an important structural class of **membrane proteins** located in the outer **membrane** of gram-negative bacteria, and in the outer **membrane** of

Question 5

[Papers](#) by Effson A.

National Library of Medicine
National Center for Biotechnology Information

Log in

[Advanced](#)
[Create alert](#)
[Create RSS](#)
[User Guide](#)

Sorted by: Best match

MY NCBI FILTERS

RESULTS BY YEAR

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

156 results

☐ 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.**
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.**
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, Ramiłowski JA, Severin J, Hayashizaki Y, Itoh M, 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.

National Library of Medicine
National Center for Biotechnology Information

Log in

[Advanced](#)
[User Guide](#)

Sorted by: Best match

MY NCBI FILTERS

RESULTS BY YEAR

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

Similar articles for PMID: 19109882

28 results

☐ 1
Mitochondrial beta-barrel proteins, an exclusive club?
Cite Imai K, Gromiha MM, Horton P.
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.
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.
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.
Cite Longen S, Bien M, Bihmaier K, Kloeppel C, Kauff F, Hammermeister M, Westermann B, Herrmann JM, Riemer J.
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



PubMed.gov

("nature"[Journal]) AND (("2022"[Date - Publication] : "3000"[Date - Publication])

Search

Advanced Create alert Create RSS User Guide

Save Email Send to

Sorted by: Best match Display options

MY NCBI FILTERS 326 results

Scopus

124 document results

ALL (nature) AND (LIMIT-TO (PUBYEAR , 2022)) AND (LIMIT-TO (EXACTSRCTITLE , "Nature"))

Edit Save Set alert

Question 8

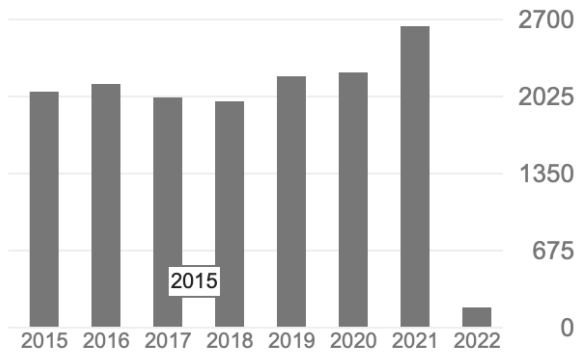
H-index : 103

Citations : 47150

Cited by

[VIEW ALL](#)

	All	Since 2017
Citations	47150	11272
h-index	103	49
i10-index	236	156



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

Human

Scientific Name : Homo Sapiens

Taxonomy ID : 9606

Chromosomes : 46

Cat

Scientific Name : *Felis catus*

Taxonomy ID : 9685

Chromosomes : 38

Dog

Scientific Name : Canis lupus familiaris

Taxonomy ID : 9615

Chromosomes : 78

House mouse

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](#)