

Formatos, portais, bancos de dados e ferramentas on-line

Aula TP2

- **Portais:** locais de acesso a recursos de vários tipos (sequências, estruturas, genomas, reacções, bibliografia, pequenas moléculas,...)
- **Ferramentas on-line:** conversão de formatos, tradução de sequência, alinhamento, pesquisa, visualização, ...
- **Bancos de dados:** repositórios de informação estruturada, interconectada e facilmente acessível

Podem ser serviços de acesso livre, ou sites comerciais com custos de utilização

NCBI Entrez

The screenshot shows a web browser window with the address bar displaying `ncbi.nlm.nih.gov/search/`. The browser's address bar includes navigation icons (back, forward, refresh, home), search, and various extension icons. Below the address bar, there are several tabs: "Search NCBI databases - NLM", "Hugging Face", "DeepLearnAI", "makemore", "FastAI", "ZoteroBib: constant-p...", "Chrome", and "HPCC-WP9-report -...".

The main content area of the browser shows the NCBI website. At the top, there is a blue header with the NIH logo and the text "National Library of Medicine National Center for Biotechnology Information". A "Log in" button is located in the top right corner of this header. Below the header is a search bar with the text "Search NCBI" and a "Search" button.

The main content area is divided into several sections:

- Literature**
 - PubMed**: PubMed® comprises more than 36 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full text content from PubMed Central and publisher web sites. A "PubMed.gov" logo is displayed to the right.
 - Featured Bookshelf titles**:
 - LiverTox**: Clinical and Research Information on Drug-Induced Liver Injury. Includes a "Browse the Bookshelf" link.
 - Drugs and Lactation Database (LactMed®)**: Includes a "Browse the Bookshelf" link.
 - Literature databases**:
 - Bookshelf**: Books and reports
 - MeSH**: Ontology used for PubMed indexing
 - NLM Catalog**: Books, journals and more in the NLM Collections
 - PubMed**: Scientific and medical abstracts/citations
 - PubMed Central**: Full-text journal articles
- Data**
 - Genes**: Gene sequences and annotations used as
 - Proteins**: Protein sequences, 3-D structures, and tools
 - BLAST**: A tool to find regions of similarity between

NCBI Entrez

Gene

Gene sources: Genomic, Categories, Sequence content, Status (Current), Clear all, Show additional filters

Gene dropdown menu:

- All Databases
- Assembly
- Biocollections
- BioProject
- BioSample
- Books
- ClinVar
- Conserved Domains
- dbGaP
- dbVar
- Gene
- Genome
- GEO DataSets
- GEO Profiles
- GTR
- HomoloGene
- Identical Protein Groups
- MedGen
- MeSH
- NLM Catalog

Search results table:

Description	Location	Aliases
lysozyme (renal amyloidosis) [<i>Gallus gallus</i> (chicken)]	Chromosome 1, NC_052532.1 (35550009..35553725)	LYZC, dystrophin
GH25 family lysozyme [<i>Ligilactobacillus animalis</i>]		GSR62_RS06660, GSR62_06660
Possible base plate lysozyme [<i>Campylobacter phage CP220</i>]	NC_027997.1 (66941..67825)	APL47_gp073, CPT_0073
GH25 family lysozyme [<i>Enterococcus gallinarum</i>]		EB54_RS15560, EB54_03104
hypothetical protein [<i>Acinetobacter radioresistens</i>]		DOM24_RS06280, DOM24_06280

Filters: Manage Filters

Results by taxon: Top Organisms (Tree)

- Acinetobacter radioresistens (1)
- Ligilactobacillus animalis (1)
- Enterococcus gallinarum (1)
- Gallus gallus (1)
- Campylobacter phage CP220 (1)

Find related data: Database: [Select] Find items

Search details: `((("Gallus gallus"[Organism] OR Gallus gallus[All Fields]) AND lysozyme[All Fields] AND C[All Fields]) AND alive[prop])` Search See more...

Vantagens da utilização dos serviços online

- Disponíveis em qualquer local
- Custos de manutenção reduzidos
- Custos de licenciamento reduzidos
- Integração de diferentes tipos de software
- Fácil monitorização da utilização
- Computação em *cloud*
- Compatibilidade com múltiplas plataformas informáticas (Win, Mac, Linux, Android, etc)

Bancos de dados primários e secundários

- **Bancos de dados primários:** informação obtida diretamente a partir da determinação experimental da sequência ou estrutura, com pouco processamento.
Exemplos: GeneBank, EMBL, DDBJ, PDB, GEO, Kegg
- **Bancos de dados secundários:** informação manualmente curada ou processada computacionalmente com base em um ou mais bancos de dados primários.
Exemplos; Swiss-Prot, Prosite, PFAM, PDBind

Curação de bancos de dados

- Manualmente curados:
 - Análise, verificação, comentário, conexão e expansão dos dados levado a cabo por operadores humano (ex: Swiss-Prot)
- Automaticamente curados:
 - Informação verificada e processada de modo automático por via de *software* (ex: Uniprot Trembl)

Actualmente, muitas bases de dados combinam curação manual e automática dos dados.

Em alguns casos o volume de dados é simplesmente demasiado grande para permitir curação manual.

Bancos de dados

- Macromoléculas:
 - Estrutura (Protein Data Bank, PDB, TTD, ModBase)
 - Sequência (Uniprot, Genbank, ...)

- Moléculas pequenas:
 - (PubChem, Drugbank, Cambridge Database, ZINC, ChEMBL, TCM, WOMBAT,)

Contêm muita informação além da *estrutura/sequência* propriamente dita.

Formatos de representação

- Estrutura:
 - PDB, MDL, SDF, MOL2, CIF, ASN.1, HIN, Trypos, Sybil, Gaussian, XYZ, CML, XML, SMILES
- Sequência:
 - Fasta, SWISSPROT, ASN.1, GCG, GenBank, PIR, Phylip....

Ferramenta de conversão entre formatos:

OpenBabel (<http://openbabel.org>)

E-Babel: conversão de formatos online

The screenshot shows a web browser window with two tabs: "E-BABEL Molecular" and "Open Babel". The address bar shows the URL "www.vcclab.org/lab/babel/start.html". The browser's bookmark bar contains folders for "Apps", "Enzymology", "Piano", "Music Production", "Bioinformatics", "Databases", "Bioinformatics T...", "Misc", "Programming", and "Other bookmarks". The page header features the text "Virtual Computational Chemistry Laboratory" and a link to "http://www.vcclab.org".

The main content area displays a welcome message: "Welcome to the Open Babel Molecular Structure Formats Interconversion program!". Below this is a form for file conversion. The form includes a dropdown menu for "Examples of atropine", an "Input format" dropdown set to "mol2 - Sybyl Mol2 file", and an "Output format" dropdown set to "smiles - SMILES file". A button labeled "upload file and perform conversion" is positioned below the dropdowns. A status message at the bottom of the form reads "Connection to Server http://146.107.217.178/vcc is established".

Below the form, a paragraph of text states: "For more information click on a keyword or a calculated result. If you cannot upload data or see results, enable pop-up windows or/and use Firefox." This is followed by links for "See FAQ" and "How to cite this applet?".

The footer of the page contains a link to "http://www.vcclab.org" and a copyright notice: "Copyright 2001 -- 2011 http://www.vcclab.org. All rights reserved."

Formato de ficheiros de sequências

- FASTA
- FASTA-PEARSON
- NBRF
- GCG
- PIR
- GenBank
- PHYLIP
- ASN.1
- PAUP

Formato FASTA

- É um formato de representação de sequências biológicas (DNA ou proteína)
- Consiste numa linha de cabeçalho, seguida de linhas contendo a sequência de aminoácidos ou nucleótidos representada em códigos de 1 letra
- Contem muito pouca informação para além da sequência

Formato FASTA

Cabeçalho

```
>gi|19151|emb|Z14088.1| L.esculentum mRNA for 108 protein
```

```
AACAATCATGGCATCTGTGAAGTCGTCGTCGTCATCATCATCATCATTTTATTTTCCTTGTT  
GTTGTTGATTTTGCTTGTGATTGTACTGCAAAGCCAAGTTATCGAGTGTCAACCTCAACAGT  
CATGCACCGCGTCACTTACTGGCCTGAACGTCTGCGCCCCATTCTGGTCCCAGGCTCACCTAC  
TGCAAGTACGGAGTGTTGCAA TGCAAGTACAGTCGATTAATCATGACTGTATGTGCAACACT  
ATGCGCATTGCAGCTCAAATTCCAGCTCAG TGCAACCTCCCTCCACTCTCTTGTCTGCAAAT  
TGAGTTTGAGATCAGTGGCCAGCAAGTTTACATCTGC TACATGAGCAAATTAATAATATC  
GTAACAATAAATTAAGTTGTCTTTTTTTTTTTTTTTTGGTTATGCAAC AGACCAAGGGGGTCA  
TGAGAAAAGAGTTTGTACTATCATATGATTATCAATAAAAAAAATTATGAG
```

```
>Q43495|108_SOLLC Protein 108 precursor - Solanum lycopersicum
```

```
MASVKSSSSSSSSSFISLLLLILLVIVLQSQVIECQPQQSCTASLTGLNVCAPFLVPGSP  
TASTECCNAVQSINHDCMCNTMRIA AQIPAQC�LPPLSCSAN
```

Sequência

Formato SWISSPROT

- Representação de sequências de proteína
- Sintaxe complexa com uma variedade de *campos*
- Contem muita informação além da sequência



Formato SWISSPROT (1/7)

ID LYZL4_MOUSE Reviewed; 145 AA.
AC Q9D925;
DT 27-JUN-2006, integrated into UniProtKB/Swiss-Prot.
DT 27-JUN-2006, sequence version 3.
DT 13-SEP-2023, entry version 143.
DE RecName: Full=Lysozyme-like protein 4;
DE Short=Lysozyme-4;
DE Flags: Precursor;
GN Name=Lyzl4; Synonyms=Lyc4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae;
OC Murinae; Mus; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M., Davis M.J.,
RA Wilming L.G., Aidinis V., Allen J.E., Ambesi-Impiombato A., Apweiler R.,
RA Aturaliya R.N., Bailey T.L., Bansal M., Baxter L., Beisel K.W., Bersano T.,
RA Bono H., Chalk A.M., Chiu K.P., Choudhary V., Christoffels A.,
RA Clutterbuck D.R., Crowe M.L., Dalla E., Dalrymple B.P., de Bono B.,
RA Della Gatta G., di Bernardo D., Down T., Engstrom P., Fagiolini M.,
RA Faulkner G., Fletcher C.F., Fukushima T., Furuno M., Futaki S.,
RA Gariboldi M., Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustinich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N., Hill D.,
RA Huminiecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T., Jakt M.,
RA Kanapin A., Katoh M., Kawasaki Y., Kelso J., Kitamura H., Kitano H.,
RA Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K., Kurochkin I.V.,
RA Lareau L.F., Lazarevic D., Lipovich L., Liu J., Liuni S., McWilliam S.,
RA Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuzawa S., Miki H.,

Formato SWISSPROT (2/7)

RA Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuzawa S., Miki H.,
RA Mignone F., Miyake S., Morris K., Mottagui-Tabar S., Mulder N., Nakano N.,
RA Nakauchi H., Ng P., Nilsson R., Nishiguchi S., Nishikawa S., Nori F.,
RA Ohara O., Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G.,
RA Pesole G., Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z.,
RA Ringwald M., Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA Schoenbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B., Sperling S.,
RA Stupka E., Sugiura K., Sultana R., Takenaka Y., Taki K., Tammoja K.,
RA Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A., Ueda H.R.,
RA van Nimwegen E., Verardo R., Wei C.L., Yagi K., Yamanishi H.,
RA Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C., Grimmond S.M.,
RA Teasdale R.D., Liu E.T., Brusic V., Quackenbush J., Wahlestedt C.,
RA Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y., Fukuda S.,
RA Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T., Iida J., Imamura K.,
RA Itoh M., Kato T., Kawaji H., Kawagashira N., Kawashima T., Kojima M.,
RA Kondo S., Konno H., Nakano K., Ninomiya N., Nishio T., Okada M., Plessy C.,
RA Shibata K., Shiraki T., Suzuki S., Tagami M., Waki K., Watahiki A.,
RA Okamura-Oho Y., Suzuki H., Kawai J., Hayashizaki Y.;
RT ["The transcriptional landscape of the mammalian genome.";](#)
RL [Science 309:1559-1563\(2005\).](#)
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY [LARGE SCALE ANALYSIS].
RC TISSUE=Testis;
RX PubMed=21183079; DOI=10.1016/j.cell.2010.12.001;
RA Huttlin E.L., Jedrychowski M.P., Elias J.E., Goswami T., Rad R.,
RA Beausoleil S.A., Villen J., Haas W., Sowa M.E., Gygi S.P.;
RT ["A tissue-specific atlas of mouse protein phosphorylation and expression.";](#)
RL [Cell 143:1174-1189\(2010\).](#)
RN [3]

Formato SWISSPROT (3/7)

RP FUNCTION, TISSUE SPECIFICITY, DEVELOPMENTAL STAGE, SUBCELLULAR LOCATION,
RP AND ABSENCE OF BACTERIOLYTIC ACTIVITY.
RX PubMed=21444326; DOI=10.1093/abbs/gmr017;
RA Sun R., Shen R., Li J., Xu G., Chi J., Li L., Ren J., Wang Z., Fei J.;
RT "Lyzl4, a novel mouse sperm-related protein, is involved in
RT fertilization."
RL Acta Biochim. Biophys. Sin. 43:346-353(2011).
RN [4]
RP TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX PubMed=24013621; DOI=10.1038/aja.2013.93;
RA Wei J., Li S.J., Shi H., Wang H.Y., Rong C.T., Zhu P., Jin S.H., Liu J.,
RA Li J.Y.;
RT "Characterisation of Lyzls in mice and antibacterial properties of human
RT LYZL6."
RL Asian J. Androl. 15:824-830(2013).
CC [-!- FUNCTION: May be involved in fertilization \(PubMed:21444326\). Has no
CC detectable bacteriolytic in vitro \(PubMed:21444326\). Has no lysozyme
CC activity in vitro \(By similarity\). {ECO:0000250|UniProtKB:D4ABW7,
CC ECO:0000269|PubMed:21444326}.](#)
CC [-!- SUBUNIT: Monomer. {ECO:0000305}.](#)
CC [-!- SUBCELLULAR LOCATION: Secreted {ECO:0000269|PubMed:21444326}.](#)
CC [Cytoplasmic vesicle, secretory vesicle, acrosome](#)
CC [{ECO:0000269|PubMed:21444326}. Cell projection, cilium, flagellum](#)
CC [{ECO:0000269|PubMed:21444326}. Note=Found in the principal piece of](#)
CC [sperm tail \(PubMed:21444326\). {ECO:0000269|PubMed:21444326}.](#)
CC [-!- TISSUE SPECIFICITY: Expressed strongly in testis and in epididymis, and](#)
CC [weakly in brain and lung \(PubMed:21444326, PubMed:24013621\). Detected](#)
CC [in sperm \(at protein level\) \(PubMed:21444326\).](#)
CC [{ECO:0000269|PubMed:21444326, ECO:0000269|PubMed:24013621}.](#)

Formato SWISSPROT (4/7)

```
CC      -!- DEVELOPMENTAL STAGE: No expression in the testis of 2-weeks-old
CC      neonates, the expression reaches a peak level at 12 weeks. After that,
CC      the level gradually decreases as the age increases (PubMed:21444326,
CC      PubMed:24013621). {ECO:0000269|PubMed:21444326,
CC      ECO:0000269|PubMed:24013621}.
CC      -!- SIMILARITY: Belongs to the glycosyl hydrolase 22 family.
CC      {ECO:0000255|PROSITE-ProRule:PRU00680}.
CC      -!- CAUTION: Although it belongs to the glycosyl hydrolase 22 family, Gly-
CC      72 is present instead of the conserved Asp which is an active site
CC      residue. It is therefore expected that this protein lacks hydrolase
CC      activity. {ECO:0000305}.
CC      -!- SEQUENCE CAUTION:
CC      Sequence=BAB25023.2; Type=Erroneous initiation; Evidence={ECO:0000305};
CC      -----
CC      Copyrighted by the UniProt Consortium, see https://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution (CC BY 4.0) License
CC      -----
DR      EMBL; AK007412; BAB25023.2; ALT_INIT; mRNA.
DR      CCDS; CCDS23632.1; -.
DR      RefSeq; NP_081191.1; NM_026915.2.
DR      RefSeq; XP_006512311.1; XM_006512248.2.
DR      AlphaFoldDB; Q9D925; -.
DR      SMR; Q9D925; -.
DR      STRING; 10090.ENSMUSP00000076887; -.
DR      CAZy; GH22; Glycoside Hydrolase Family 22.
DR      PhosphoSitePlus; Q9D925; -.
DR      PaxDb; Q9D925; -.
DR      ProteomicsDB; 292157; -.
DR      Antibodypedia; 29191; 111 antibodies from 20 providers.
DR      DNASU; 69032; -.
DR      Ensembl; ENSMUST00000077706; ENSMUSP00000076887; ENSMUSG00000032530.
DR      Ensembl; ENSMUST00000120918; ENSMUSP00000113034; ENSMUSG00000032530.
```

Formato SWISSPROT (5/7)

DR GeneID; 69032; -.
DR KEGG; mmu:69032; -.
DR UCSC; uc009sdi.1; mouse.
DR AGR; MGI:1916282; -.
DR CTD; 131375; -.
DR MGI; MGI:1916282; Lyz14.
DR VEuPathDB; HostDB:ENSMUSG00000032530; -.
DR eggNOG; ENOG502SSER; Eukaryota.
DR GeneTree; ENSGT00940000162293; -.
DR HOGENOM; CLU_111620_1_1_1; -.
DR InParanoid; Q9D925; -.
DR OMA; AWPSWSL; -.
DR OrthoDB; 5344399at2759; -.
DR PhylomeDB; Q9D925; -.
DR TreeFam; TF324882; -.
DR BioGRID-ORCS; 69032; 0 hits in 76 CRISPR screens.
DR ChiTaRS; Lyz14; mouse.
DR PRO; PR:Q9D925; -.
DR Proteomes; UP000000589; Chromosome 9.
DR RNAct; Q9D925; Protein.
DR Bgee; ENSMUSG00000032530; Expressed in spermatid and 35 other tissues.
DR ExpressionAtlas; Q9D925; baseline and differential.
DR Genevisible; Q9D925; MM.
DR GO; GO:0001669; C:acrosomal vesicle; IDA:UniProtKB.
DR GO; GO:0005615; C:extracellular space; IDA:UniProtKB.
DR GO; GO:0036126; C:sperm flagellum; IDA:UniProtKB.
DR GO; GO:0003796; F:lysozyme activity; IEA:InterPro.
DR GO; GO:0009566; P:fertilization; IMP:UniProtKB.
DR GO; GO:0007342; P:fusion of sperm to egg plasma membrane involved in single

Formato SWISSPROT (6/7)

```
DR CDD; cd16897; LYZ_C; 1.
DR Gene3D; 1.10.530.10; -; 1.
DR InterPro; IPR001916; Glyco_hydro_22.
DR InterPro; IPR019799; Glyco_hydro_22_CS.
DR InterPro; IPR000974; Glyco_hydro_22_lys.
DR InterPro; IPR023346; Lysozyme-like_dom_sf.
DR PANTHER; PTHR11407; LYSOZYME C; 1.
DR PANTHER; PTHR11407:SF21; LYSOZYME-LIKE PROTEIN 4; 1.
DR Pfam; PF00062; Lys; 1.
DR PRINTS; PR00137; LYSOZYME.
DR PRINTS; PR00135; LYZLACT.
DR SMART; SM00263; LYZ1; 1.
DR SUPFAM; SSF53955; Lysozyme-like; 1.
DR PROSITE; PS00128; GLYCOSYL_HYDROL_F22_1; 1.
DR PROSITE; PS51348; GLYCOSYL_HYDROL_F22_2; 1.
PE 1: Evidence at protein level;
KW Cell projection; Cilium; Cytoplasmic vesicle; Disulfide bond;
KW Fertilization; Flagellum; Reference proteome; Secreted; Signal.
FT SIGNAL 1..19
FT /evidence="ECO:0000255"
FT CHAIN 20..145
FT /note="Lysozyme-like protein 4"
FT /id="PRO_0000240639"
FT DOMAIN 20..145
FT /note="C-type lysozyme"
FT /evidence="ECO:0000255|PROSITE-ProRule:PRU00680"
FT ACT_SITE 54
```

Formato SWISSPROT (7/7)

```
FT          /evidence="ECO:0000255|PROSITE-ProRule:PRU00680"  
FT  DISULFID 25..143  
FT          /evidence="ECO:0000255|PROSITE-ProRule:PRU00680"  
FT  DISULFID 49..130  
FT          /evidence="ECO:0000255|PROSITE-ProRule:PRU00680"  
FT  DISULFID 84..95  
FT          /evidence="ECO:0000255|PROSITE-ProRule:PRU00680"  
FT  DISULFID 91..109  
FT          /evidence="ECO:0000255|PROSITE-ProRule:PRU00680"  
SQ  SEQUENCE 145 AA; 16198 MW; 2AC71B4AB7EE96BD CRC64;  
      MQLYLVLLLI SYLLTPIGAS ILGRCTVAKM LYDGGLNYFE GYSLENWVCL AYFESKFNPS  
      AVYEDPQDGS TGFGLFQIRD NEWCGHGKNL CSVSCTALLN PNLKDTIQCA KKIVKGKHGM  
      GAWPIWSKNC QLSDVLDLDRWL DGCDL  
//
```

Formato Genbank:Cabeçalho

Header

RefSeq Id
↓

```
LOCUS      NC_000854                1669695 bp    DNA    circular BCT 03-DEC-2005
DEFINITION Aeropyrum pernix K1, complete genome.
ACCESSION  NC_000854
VERSION    NC_000854.1  GI:14600379
KEYWORDS   .
SOURCE     Aeropyrum pernix K1 K1
  ORGANISM Aeropyrum pernix
           Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
           Desulfurococcaceae; Aeropyrum.
REFERENCE  1
  AUTHORS  Kawarabayasi,Y., Hino,Y., Horikawa,H., Yamazaki,S., Haikawa,Y.,
           Jin-no,K., Takahashi,M., Sekine,M., Baba,S., Ankai,A., Kosugi,H.,
           Hosoyama,A., Fukui,S., Nagai,Y., Nishijima,K., Nakazawa,H.,
           Takamiya,M., Masuda,S., Funahashi,T., Tanaka,T., Kudoh,Y.,
           Yamazaki,J., Kushida,N., Oguchi,A., Aoki,K., Kubota,K.,
           Nakamura,Y., Nomura,N., Sako,Y. and Kikuchi,H.
  TITLE    Complete genome sequence of an aerobic hyper-thermophilic
           crenarchaeon, Aeropyrum pernix K1
  JOURNAL  DNA Res. 6 (2), 83-101 (1999) PUBMED 10382966
REFERENCE  2 (bases 1 to 1669695)
  AUTHORS  Direct submission
  TITLE    Submitted (05-JUL-2001) National Center for Biotechnology
           Information, NIH, Bethesda, MD 20894, USA
REFERENCE  3 (bases 1 to 1669695)
  AUTHORS  Tanaka,T., Hino,Y., Kawarabayasi,Y. and Kikuchi,H.
  TITLE    Direct submission
  JOURNAL  Submitted (14-DEC-1998) National Institute of Technology and
           Evaluation, Biotechnology Center, 2-49-10 Nishihara, Shibuyaku,
           Tokyo 151-0066, Japan
COMMENT   PROVISIONAL REFSEQ: This record has not yet been subject to final
           NCBI review. The reference sequence was derived from BA000002.
           COMPLETENESS: full length.
```

Formato Genbank:Anotações

Features	FEATURES	Location/Qualifiers
	source	1..1669695 /mol_type="genomic DNA" /db_xref="taxon:272557" /organism="Aeropyrum pernix K1"
	gene	complement(213..938) <u>/locus_tag="APE0001"</u> ← Locus tag /db_xref="GeneID:1445602"
	CDS	complement(213..938) /locus_tag="APE0001" /protein_id="NP_146894.1" /transl_table=11 /db_xref="GI:14600380" /db_xref="GeneID:1445602" /codon_start=1 /product="hypothetical protein" /translation="MVDILSSLLLSLPGFVIGFLLVLSPGSIWTPVKESIGYVYVSR VTVKASKLLGSLTLLASLISFVGAAYGITIQASTLALLLALITVVTVEYSMRLAEIE SLNQPVLEGFEPVGSIKLKYLTIIILLVYLISIVFSIEGSLKLYSIGAYGTLASHLSIE ILAGYTVFLSVKRPEAYVIPGLSRETIELLQFFMPTSLSLIAIGVYMILAGFHMWII LLAGVTTLFVVTMLIMINKEGKY"
	gene	complement(938..1276) /locus_tag="APE0002"
	CDS	complement(938..1276) <u>/locus_tag="APE0002"</u> ← Locus tag /protein_id="NP_146895.1" /transl_table=11 /note="similar to PIR:C69525 percent identity:39.583 in 96aa." /db_xref="GI:14600381" /db_xref="GeneID:1445577" /codon_start=1 /product="hypothetical protein" /translation="MDPADKLMKDARTGVLALAVLHVLVNHGALHGYWLRKILGNLMG WTPPETSLYDALKRLEKLGLIKGRWVRSRGRPLRKYIEITDAGRETYEVVVKDFSKMV GWLICRKGRE"
	misc_feature	complement(1001..>1180) /locus_tag="APE0002" /db_xref="CDD:43477" /note="Transcriptional regulator PadR-like family; Region: PadR"

Formato Genbank: Sequência

Sequence

```
BASE COUNT      360022 a 473378 c 466849 g 369446 t
ORIGIN
1 aaataaat  aaaaattaag  tgactcatgc  attatcctac  gaggtaaaaa  tatggtataa
61 attgtcccag  actaccatca  atttagggac  aatagtgttt  aagggatggc  cttcggagct
121 ggcagctcgc  gggttcaaac  tcgcgtaggg  cccgagttct  agttatagtt  gcggtgattt
181 agataaattg  agtatgatct  ctcaactttt  tatcaatact  tacctcttt  ataatcata
241 attaacattg  tfacaacgaa  tagagtggtc  actcccgcca  acaggattat  ccaccacata
301 tggaatcctg  ctaaaatcat  atatacacct  atagctatga  gagataagga  ggttggcatg
361 aaaaattgta  atagctcgat  cgttcccga  cttagtctg  gtattacata  tgctccggc
421 ctttttacag  atagaaaaac  ggtatatcct  gctaataatt  caatagataa  atgtgaagct
481 aacggtccgt  atgcaccaat  actatatagt  ttaagagaac  cttcaattga  gaatacaatc
541 gagattagat  aaactagtaa  tataatagtc  aaatatttta  attaataga  acctactggc
601 tcgaatcctt  caaggactgg  ttggtttaaa  gactctattt  cggcgcgact  catagaatat
661 tccacggtaa  ctacgggtat  caatgcaagt  aatagggcta  gtgtagatgc  ctgaatagta
721 ataccatatt  ctgcaccaac  cacaaaagat  attaaactcg  ctacaaaagt  aagcgaacct
781 aatagcttgc  tcgctttaa  cgtgacgcgc  ctagaaacat  aaacatagcc  tatgctctcc
841 tttacaggcg  tccatattga  cccaggagac  aataccaaga  gaaatccaat  tacaccaaat
901 ggaagcgaca  gcaggagtga  agatagtata  tctaccatta  ctctctccc  tttctgcaa
961 taagccagcc  aaccatcttt  gagaaatcct  ttactacaac  ctcatatgtc  tctctaccag
1021 catcggttat  ttcattagat  ttccttaaag  gccccctacc  gctcctaacc  catcggccct
1081 ttattagccc  cagcttttct  aacctcttca  aagcatcata  aagactcgtc  tctggaggcg
1141 tccatcccat  tagattgcca  agaattttcc  tcaaccaata  cccatgtaga  gctccatgat
1201 tgacaagtac  gtgtaatact  gccaatgcaa  gcacaccagt  ccttgcatcc  ttcattagct
1261 tatctgctgg  atccacgtga  caccaccat  tttattagga  agcctactat  tagcatggag
1321 accacgacag  agataccggc  tggaggggca  acaagcctgt  taccgatagt  tagggctgca
1381 aaaaactcct  caataccatt  aaccgttcca  tgcgctattg  ctggagtaat  gatggagtgt
1441 gaatgtctcc  taagaggtaa  aaggatgctt  gtaaatgcta  tgggtataaa  tgtgaagact
1501 actatagcgg  gccaccctcg  ggaatagctt  ccacaatctc  ctatagatga  tacgttgtaa
1561 ttataaccag  cataaattaa  gggagcatgc  cagacactcc  agataagacc  tataataatg
1621 accttaccga  gatcgttaac  tttcttatcg  agtatggtga  agagatatcc  tctccagccg
1681 agttcttctc  caagtgcaac  aagtgcgctc  atggtaaact  ctgctataag  accaagtaat
1741 attagtatta  taaccgttgt  aattagcaga  gtagtattag  atacctcctt  gaagtatccg
1801 catggtccaa  tactaacgcc  taaagcctta  gcgattggt  atgacatcac  atatgaggct
1861 aatggcgcta  ccgctgataa  tatggtccat  ttcaaggatg  gaatattaat  tctcaagatt
1921 tctttatttt  tctccatggt  acgatatcct  tcgaccata  aggctgcat  aacgcctgta
1981 gcaggtagcc  acatcctaaa  taggaggacg  atcgtgagga  ggagtttatt  tcggggtaag
2041 gttgtggtgg  gctcttgcat  tgacgttagt  aactttatgg  ctattgtata  gtctaggagg
2101 tatgctggga  cgaatgatac  tgttaggaag  actgctaaac  ctatgtaatg  gcgtttatcg
2161 atttctatac  gcatactacg  tccaccgggg  tatttatcat  gttatagatg  tatttaagac
2221 caaagctgat  ttaagaacct  aacattgtat  atatagtttg  gtgttaccgt  tggcggtaga
2281 gcaattaacg  attgctggaa  gcgagctact  aaaacatgag  ctacaagca  agctagttat
2341 cggcgtatta  ttgtccggtt  ggatagtcac  agttgcagtc  atcaagctca  ggaagctac
2401 gaggaatagg  cagatagctg  gtctaattgt  agcggctgta  gccacagctc  tggcttagg
2461 tacaatagcc  tafatattta  acccgctcca  aacctatggc  gcttatctcg  agagtagaac
2521 attacaaatt  agattctaca  tgaatgatga  agtggctgtt  gacttatgta  acgctcagtt
2581 gtcattgcta  tcgaggagca  acgcaataaa  cttactctac  attagaacta  acggtattgc
2641 tgatcctttc  tcaggtatta  ctgcgggata  ttacaaaact  gtggatgaac  ggggaagccta
2701 tgtacttatc  gctggtaagg  acattgatga  tgtccttgca  atcgaattcg  acagtaaaat
2761 tatcctccta  ggacttaaa  gagcgaatga  cttttaccaa  aaactcataa  tttacaaaag
```


UniProt, a referência universal para sequências de proteínas

- A fusão das bases de dados PIR, TrEMBL e Swiss-Prot numa única base de dados vem constituir uma referência definitiva para a pesquisa de sequências de proteína.
- Uniprot contem as seguintes subsecções:
 - **UniProtKB:** contem SwissProt e TrEMBL (translated EMBL)
 - **UniParc:** contem sequências não-annotadas de várias fontes
 - **UniRef:** contem sequências agrupadas por similaridade
- UniprotKB (Uniprot Knowledge base):
 - **Swiss-Prot** (manualmente curada) – 570 157 sequências
 - **Unreviewd** (tradução automático do EMBL) – 251 600 768 sequências



Find your protein

UniProtKB Advanced | List Search


Examples: Insulin, APP, Human, P05067, organism_id:9606

Feedback

Help


UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. [Cite UniProt](#)

Proteins UniProt Knowledgebase




Reviewed (Swiss-Prot) 570,157
Unreviewed (TrEMBL) 251,600,768

Species Proteomes



Protein sets for species with sequenced genomes from across the tree of life

Protein Clusters UniRef



Clusters of protein sequences at 100%, 90% & 50% identity

Sequence Archive UniParc



Non-redundant archive of publicly available protein sequences seen across different databases

Find your protein

UniProtKB lysozyme Advanced | List Search

Examples: Insulin, APP, Human, P05067, organism_id:9606

Feedback

Help

UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. [Cite UniProt](#)

Proteins

UniProt Knowledgebase

Reviewed (Swiss-Prot)	Unreviewed (TrEMBL)
570,157	251,600,768

Species

Proteomes

Protein sets for species with sequenced genomes from across the tree of life

Protein Clusters

UniRef

Clusters of protein sequences at 100%, 90% & 50% identity

Sequence Archive

UniParc

Non-redundant archive of publicly available protein sequences seen across different databases

Status

- Reviewed (Swiss-Prot) (52)
- Unreviewed (TrEMBL) (10,634)

Popular organisms

- Human (38)
- Zebrafish (10)
- Mouse (3)
- Bovine (1)

Taxonomy

[Filter by taxonomy](#)

Group by

- Taxonomy
- Keywords
- Gene Ontology
- Enzyme Class

Proteins with

- 3D structure (24)
- Active site (221)
- Activity regulation (6)
- Allergen (9)
- Alternative products

UniProtKB 10,686 results

BLAST Align Map IDs Download Add View: Cards Table **Customize columns** Share

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
<input type="checkbox"/> P61626	LYSC_HUMAN	Lysozyme C[...]	LYZ, LZM	Homo sapiens (Human)	148 AA
<input type="checkbox"/> P00709	LALBA_HUMAN	Alpha-lactalbumin[...]	LALBA, LYZL7	Homo sapiens (Human)	142 AA
<input type="checkbox"/> O75951	LYZL6_HUMAN	Lysozyme-like protein 6[...]	LYZL6, LYC1, UNQ754/PRO1485	Homo sapiens (Human)	148 AA
<input type="checkbox"/> Q8IXA5	SACA3_HUMAN	Sperm acrosome membrane-associated protein 3[...]	SPACA3, LYC3, LYZL3, SLLP1, SPRASA, UNQ424/PRO862	Homo sapiens (Human)	215 AA
<input type="checkbox"/> P29590	PML_HUMAN	Protein PML[...]	PML, MYL, PP8675, RNF71, TRIM19	Homo sapiens (Human)	882 AA
<input type="checkbox"/> Q7Z4W2	LYZL2_HUMAN	Lysozyme-like protein 2[...]	LYZL2	Homo sapiens (Human)	148 AA
<input type="checkbox"/> Q96KX0	LYZL4_HUMAN	Lysozyme-like protein 4[...]	LYZL4, LYC4	Homo sapiens (Human)	146 AA
<input type="checkbox"/> Q86SG7	LYG2_HUMAN	Lysozyme g-like protein 2[...]	LYG2, LYGH	Homo sapiens (Human)	212 AA
<input type="checkbox"/> P02788	TRFL_HUMAN	Lactotransferrin[...]	LTF, GIG12, LF	Homo sapiens (Human)	710 AA
<input type="checkbox"/> P02489	CRYAA_HUMAN	Alpha-crystallin A	CRYAA, CRYA1, HSPB4	Homo sapiens	173 AA

Feedback Help

LYZ - Lysozyme C - Homo sapiens

uniprot.org/uniprotkb/P6...

UniProtKB

Advanced | List Search

P61626 · LYSC_HUMAN

Proteinⁱ	Lysozyme C	Amino acids	148 (go to sequence)
Geneⁱ	LYZ	Protein existenceⁱ	Evidence at protein level
Statusⁱ	UniProtKB reviewed (Swiss-Prot)	Annotation scoreⁱ	5/5
Organismⁱ	Homo sapiens (Human)		

Entry Variant viewer Feature viewer Publications External links History

BLAST Download Add Add a publication Entry feedback

Functionⁱ

Lysozymes have primarily a bacteriolytic function; those in tissues and body fluids are associated with the monocyte-macrophage system and enhance the activity of immunoagents.

Miscellaneous

Lysozyme C is capable of both hydrolysis and transglycosylation; it shows also a slight esterase activity. It acts rapidly on both peptide-substituted and unsubstituted peptidoglycan, and slowly on chitin oligosaccharides.

Catalytic activityⁱ

Hydrolysis of (1->4)-beta-linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in a peptidoglycan and between N-acetyl-D-glucosamine residues in chitodextrins.
 EC:3.2.1.17 (UniProtKB | ENZYME | Rhea)

Features

Showing features for active siteⁱ.

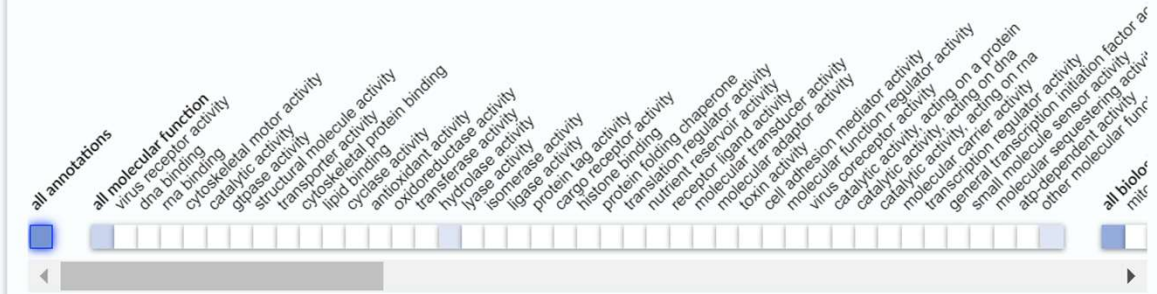
Feedback Help

- Function
- Names & Taxonomy
- Subcellular Location
- Disease & Variants
- PTM/Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Similar Proteins

Entry Variant viewer Feature viewer Publications External links History

GO annotations¹

Slimming set:



Cell color indicative of number of GO terms

ASPECT	TERM	Source	Publications
Cellular Component	azurophil granule lumen	Source:Reactome	
Cellular Component	extracellular exosome	Source:UniProtKB	3 Publications
Cellular Component	extracellular region	Source:Reactome	
Cellular Component	extracellular space	Source:UniProtKB	1 Publication
Cellular Component	specific granule lumen	Source:Reactome	
Cellular Component	tertiary granule lumen	Source:Reactome	
Molecular Function	identical protein binding	Source:IntAct	1 Publication
Molecular Function	lysozyme activity	Source:GO_Central	1 Publication
Biological Process	antimicrobial humoral response	Source:Reactome	

Feedback
Help



Function

Names & Taxonomy

Subcellular Location

Disease & Variants

PTM/Processing

Expression

Interaction

Structure

Family & Domains

Sequence

Similar Proteins

Export table

Keywordsⁱ

Molecular function: #Antimicrobial, #Bacteriolytic enzyme, #Glycosidase, #Hydrolase

Enzyme and pathway databases

BRENDA: 3.2.1.17 2681; PathwayCommons: P61626; Reactome: R-HSA-6798695 Neutrophil degranulation, R-HSA-6803157 Antimicrobial peptides, R-HSA-977225 Amyloid fiber formation; SIGNOR: P61626; SignaLink: P61626; ENZYME: Search...

Protein family/group databases

CAZy: GH22 Glycoside Hydrolase Family 22

Names & Taxonomyⁱ

Protein namesⁱ

Recommended name: Lysozyme C

Feedback Help