

PODATKOVNE ZBIRKE

1. Internet

Nekateri strežniki, EBI, NCBI, KEGG, ExPASy

2. Podatkovne zbirke

Lastnosti: velikost, določenost, presežnost

Primarne, sekundarne; bibliografske

Podatkovne zbirke nukleotidnih zaporedij

Podatkovne zbirka proteinskih zaporedij

Podatkovna zbirka 3D struktur

Bibliografska informacija: PubMed, OMIM

STREŽNIKI Z BIOKEMIJSKO IN MOL BIOL INFORMACIJO

EBI	European Bioinformatics Institute
NCBI	The National Center for Biotechnology Information
ExpASy	Expert Protein Analysis System
KEGG	Kyoto Encyclopedia of Genes and Genomes

In še mnogi drugi!

NCBI- The National Center for Biotechnology Information

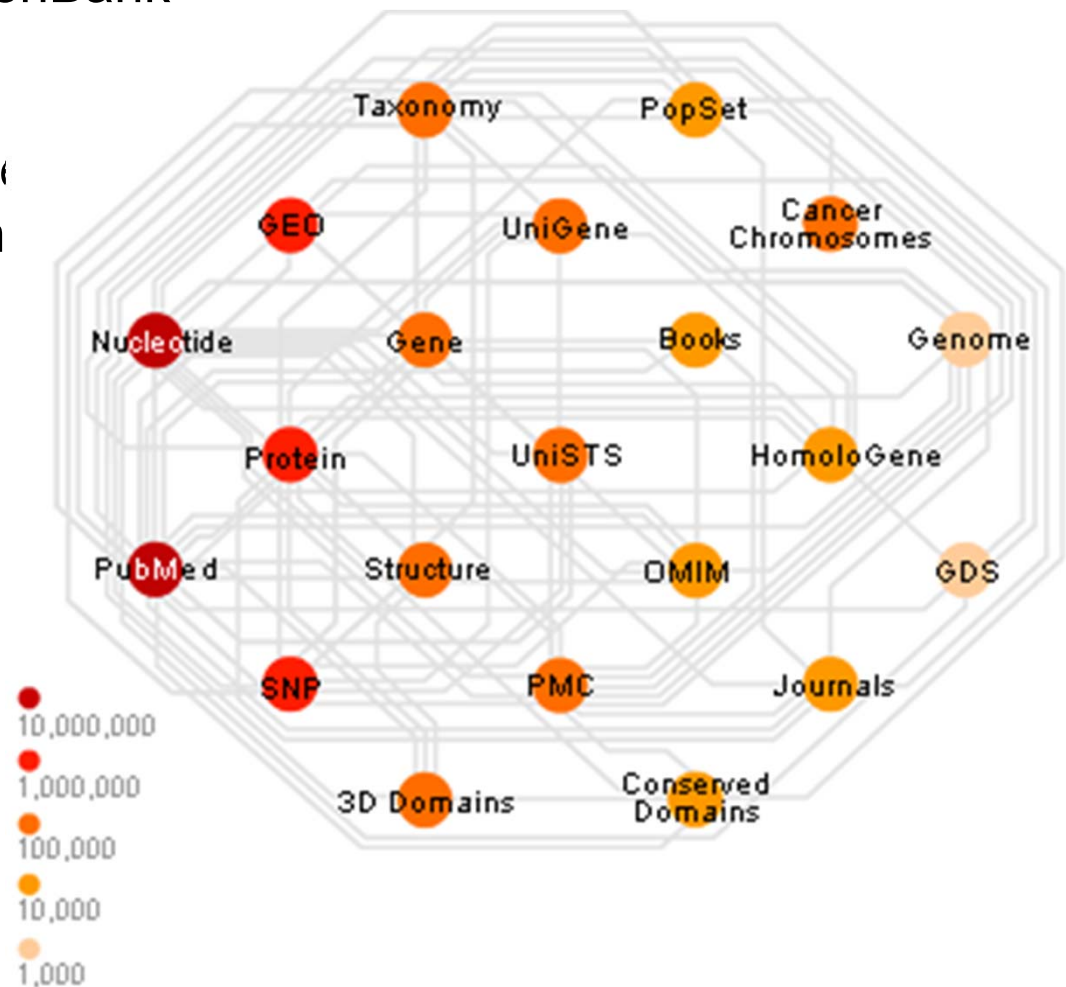
1988 Kot del National Library of Medicine

Namen: *razvoj novih informacijskih tehnologij, ki pomagajo razumeti molekularne in genetske procese*

1992 Upravljanje z GenBank

Entrez

<http://www.ncbi.nlm.nih.gov/Entrez>
Brskalnik za iskanje po bioloških podatkovnih zbirkah, ki jih ureja NCBI (proteinska zaporedja, nukleotidna zaporedja, gensko mapiranje, OMIM, 3D strukture iz PDB, PubMed)



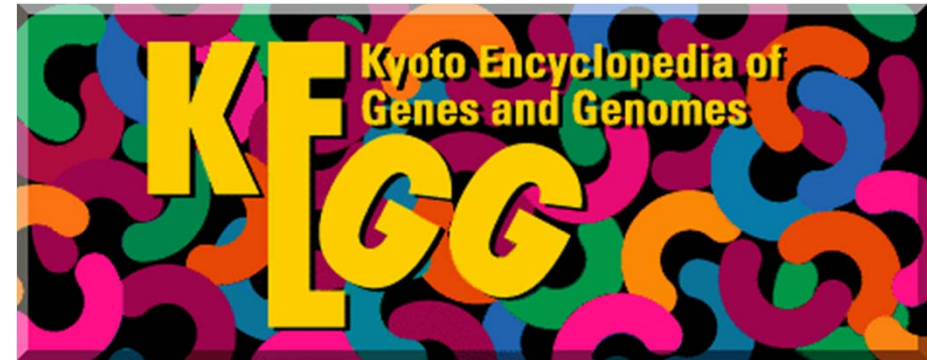
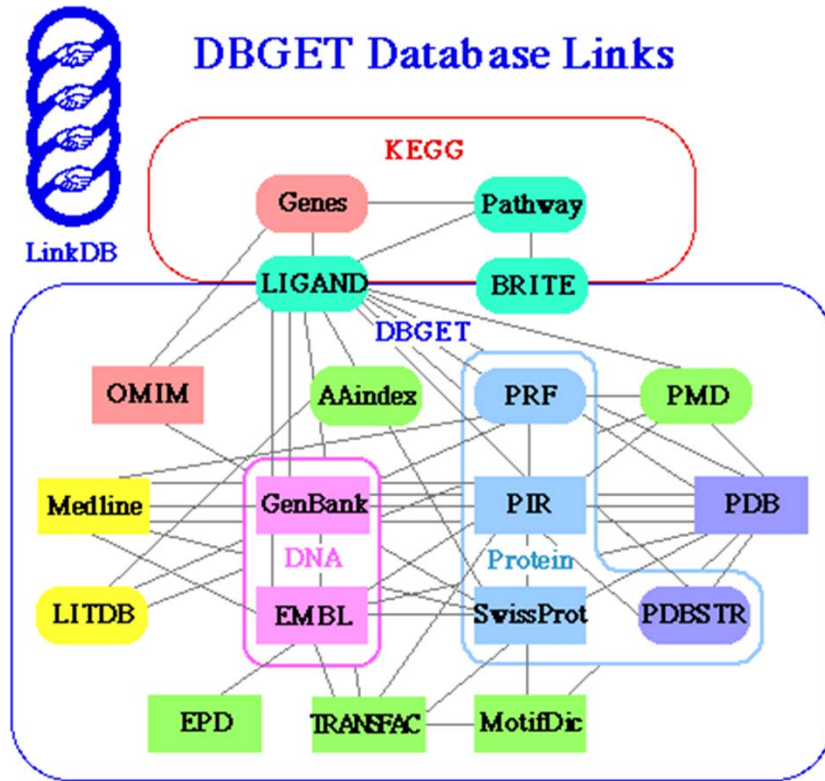


HOME SEARCH SITE MAP PubMed Entrez Human Genome GenBank Map Viewer BLAST

Search across databases [input field] GO CLEAR Help

Welcome to the new Entrez cross-database search page

- PubMed:** biomedical literature citations and abstracts
- PubMed Central:** free, full text journal articles
- Books:** online books
- OMIM:** online Mendelian Inheritance in Man
- Site Search:** NCBI web and FTP sites
- Nucleotide:** sequence database (GenBank)
- Protein:** sequence database
- Genome:** whole genome sequences
- Structure:** three-dimensional macromolecular structures
- Taxonomy:** organisms in GenBank
- SNP:** single nucleotide polymorphism
- Gene:** gene-centered information
- UniGene:** gene-oriented clusters of transcript sequences
- CDD:** conserved protein domain database
- 3D Domains:** domains from Entrez Structure
- UniSTS:** markers and mapping data
- PopSet:** population study data sets
- GEO Profiles:** expression and molecular abundance profiles
- GEO DataSets:** experimental sets of GEO data



KEGG: Kyoto Encyclopedia of Genes and Genomes

Kyoto Encyclopedia of Genes and Genomes (KEGG) is an effort to computerize current knowledge of molecular and cellular biology in terms of the information pathways that consist of interacting molecules or genes and to provide links from the gene catalogs produced by **genome sequencing projects**. The KEGG project is undertaken in the Bioinformatics Center, Institute for Chemical Research, Kyoto University with supports from the Ministry of Education, Culture, Sports, Science and Technology and the Japan Society for the Promotion of Science.

ExPASy Expert Protein Analysis System

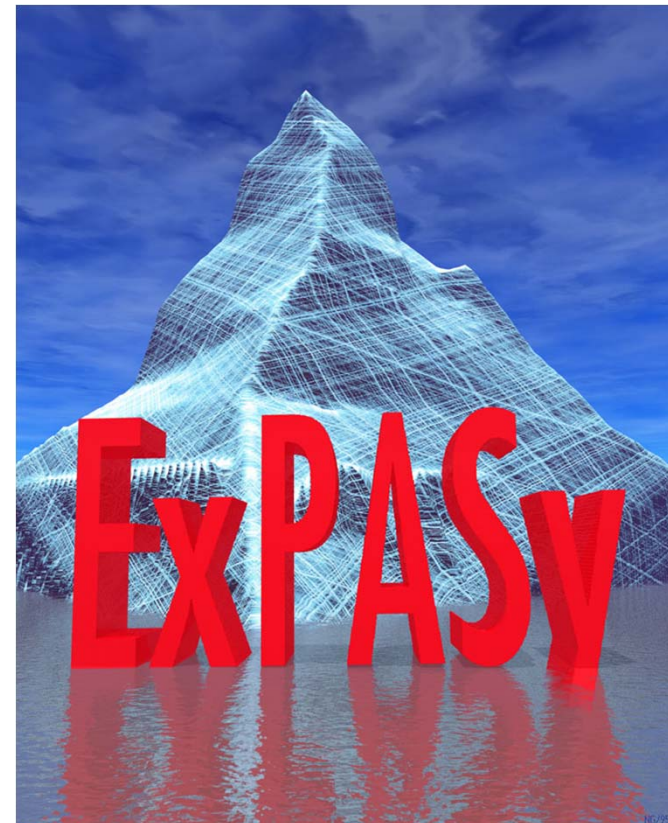
<http://www.expasy.org/>

Swiss Institute of bioinformatics (SIB)

The ExPASy (**Expert Protein Analysis System**) proteomics server from the Swiss Institute of Bioinformatics (SIB) is dedicated to molecular biology with an emphasis on data relevant to **proteins**.

Podatkovne zbirke: **SwissProt**
 /UniProt
 Swiss-Model
 PROSITE

Programi: Orodja za proteomiko



1. VEEEEEELIKOST PODATKOVNIH ZBIRK

Povezano z zajemanjem podatkov – kako težko je podatek dobiti.

Podatkovne zbirke so lahko

VELIKE- npr. GeneBank; Največja podatkovna zbirka, nukleotidna zaporedja.

171.123.749 zaporedjih (februar 2014) iz **>165.000** organizmov

SREDNJE VELIKE- npr. UniProt. proteinska zaporedja.

542.503 zaporedij (februar 2014) iz različica **13.056** vrst

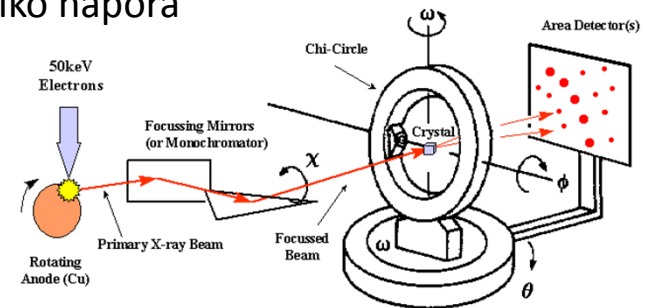
MAJHNE- npr. PDB. 3D strukture makromolekul.

76.349 struktur (februar 2014)

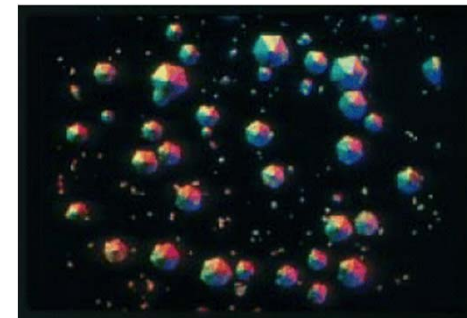
nukleotidna zaporedja možno dobiti
sorazmerno enostavno



3D strukture je tudi možno dobiti, a z
veliko napora



4-Circle Goniometer (Eulerian or Kappa Geometry)



RAZKORAK MED ŠTEVILOM ZAPOREDIJ IN ŠTEVILOM 3D STRUKTUR

>170 000 000	GeneBank
>500 000	UniProt
>70 000	Protein Data Bank

Problemi pri pripravi vzorca (količina, izražanje), lastnostih (netopni, membranski), kristalizaciji

Npr. MsbA iz *E. coli* ABC transporter
>20 homologov iz 12 bakterijskih vrst izražajo
>20 detergentov z vsemi, skupaj 96 000 testov
35 kristalnih oblik
1 kristal, ki dobro sipa

....

1 SCIENCE članek

(Chang G and Roth CB (2001) Science 293, 1793-1800)

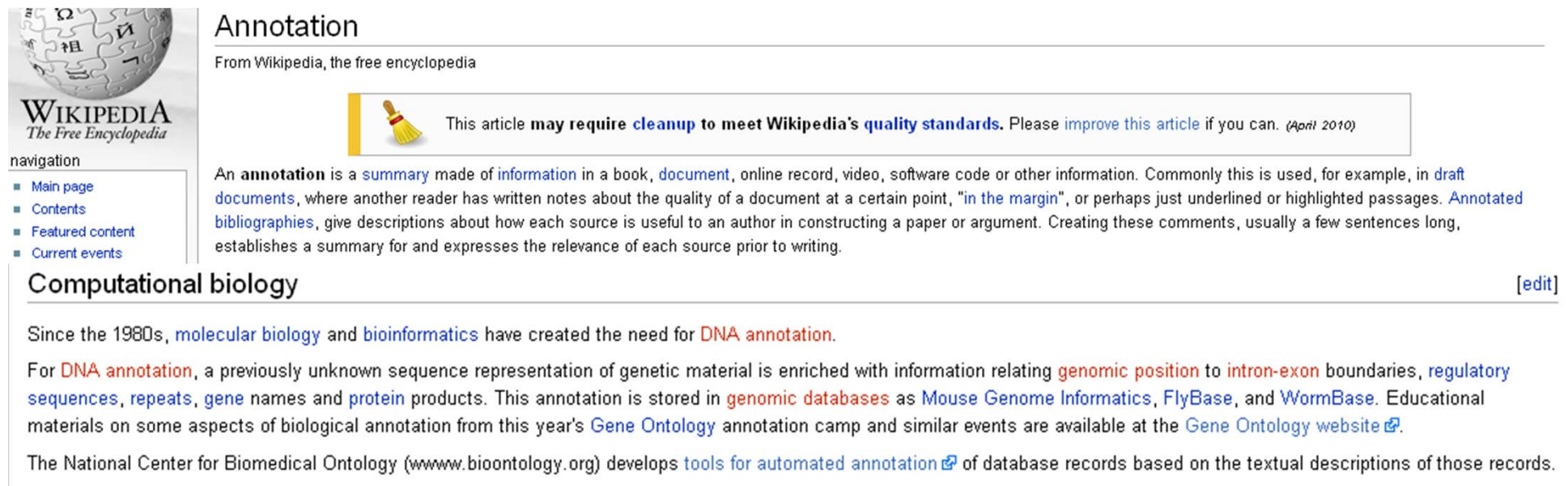
2. Anotacija- določitev

Biološke, strukturne, funkcionalne, etc. lastnosti pripišemo biološkemu podatku, npr. zaporedju.

Podatkovne zbirke so lahko

DOBRO anotirane- npr. UniProt; kuratorji vpisujejo "na roke" različne podatke iz primarnih publikacij (glej naslednji slide). Zahteva veliko napora za urejanje. Zapisov navadno malo.

SLABO anotirane- npr. GeneBank; posledica avtomatskega vnašanja podatkov v podatkovne zbirke, količine podatkov. Manj napora za urejanje. Zapisov veliko.



The screenshot shows a Wikipedia article page for "Annotation". At the top left is the Wikipedia logo with the text "WIKIPEDIA The Free Encyclopedia". Below it is a navigation menu with links for "Main page", "Contents", "Featured content", and "Current events". The main heading is "Annotation" with a sub-heading "From Wikipedia, the free encyclopedia". A yellow banner with a broom icon states: "This article may require cleanup to meet Wikipedia's quality standards. Please improve this article if you can. (April 2010)". The main text defines an annotation as a summary made of information in a book, document, online record, video, software code or other information. It mentions that annotations are commonly used in draft documents, in the margin, or as underlined/highlighted passages. It also notes that annotated bibliographies give descriptions of source usefulness. At the bottom, there is a section for "Computational biology" with an [edit] link. The text in this section discusses the need for DNA annotation since the 1980s, mentioning molecular biology and bioinformatics. It explains that DNA annotation enriches genetic material with information about genomic position, intron-exon boundaries, regulatory sequences, repeats, gene names, and protein products. It lists genomic databases like Mouse Genome Informatics, FlyBase, and WormBase. It also mentions educational materials from the Gene Ontology annotation camp and the Gene Ontology website. Finally, it notes that the National Center for Biomedical Ontology develops tools for automated annotation of database records.

Annotation
From Wikipedia, the free encyclopedia

This article may require cleanup to meet Wikipedia's quality standards. Please improve this article if you can. (April 2010)

An **annotation** is a **summary** made of **information** in a book, **document**, online record, video, software code or other information. Commonly this is used, for example, in **draft documents**, where another reader has written notes about the quality of a document at a certain point, "**in the margin**", or perhaps just underlined or highlighted passages. **Annotated bibliographies**, give descriptions about how each source is useful to an author in constructing a paper or argument. Creating these comments, usually a few sentences long, establishes a summary for and expresses the relevance of each source prior to writing.

Computational biology [edit]

Since the 1980s, **molecular biology** and **bioinformatics** have created the need for **DNA annotation**.

For **DNA annotation**, a previously unknown sequence representation of genetic material is enriched with information relating **genomic position** to **intron-exon** boundaries, **regulatory sequences**, **repeats**, **gene** names and **protein** products. This annotation is stored in **genomic databases** as **Mouse Genome Informatics**, **FlyBase**, and **WormBase**. Educational materials on some aspects of biological annotation from this year's **Gene Ontology** annotation camp and similar events are available at the **Gene Ontology website**.

The National Center for Biomedical Ontology (www.bioontology.org) develops **tools for automated annotation** of database records based on the textual descriptions of those records.

Core data aminokislinsko zaporedje
ime proteina
taksonomski podatki
citati

Določitev (“annotation”)

funkcija proteina
podatki o encimski aktivnosti (katalitska aktivnost,
kofaktorji, metabolne poti...)
biološko pomembna mesta na molekuli (domene),
posttranslacijske modifikacije
M določena z masno spektroskopijo
tkivno specifična izražanje
razvojno specifično izražanje
sekundarna struktura
kvartarna struktura
polimorfizmi
podobnost z drugimi proteini
uporaba v biotehnoloških procesih
napake (konflikti) v zaporedjih

Vpis “na roke” iz primarnih publikacij, preglednih člankov.

Pomoč zunanjih ekspertov za posamezne družine proteinov.

UniProt > UniProtKB Downloads · Contact · Documentation/Help

Search Blast* Align Retrieve ID Mapping*

Search in: Protein Knowledgebase (UniProtKB) [Fields »](#)

★ Reviewed, UniProtKB/Swiss-Prot **P61626** (LYSC_HUMAN)
 Last modified March 2, 2010. Version 80. [History...](#)

Contribute
[Send feedback](#)
[Read comments \(?\) or add your own](#)

Clusters with 100%, 90%, 50% identity | Documents (7) | Third-party data | Customize display [TEXT](#) [XML](#) [RDF/XML](#) [GFF](#) [FASTA](#)

[Names and origin](#) · [Protein attributes](#) · [General annotation \(Comments\)](#) · [Ontologies](#) · [Binary interactions](#) · [Sequence annotation \(Features\)](#) · [Sequences](#) · [References](#) · [Web resources](#) · [Cross-references](#) · [Entry information](#) · [Relevant documents](#)

Names and origin Hide | Top

Protein names	<i>Recommended name:</i> Lysozyme C EC=3.2.1.17 <i>Alternative name(s):</i> 1,4-beta-N-acetylmuramidase C
Gene names	Name: LYZ Synonyms: LZM
Organism	Homo sapiens (Human) [Complete proteome]
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo

Protein attributes Hide | Top

Sequence length	148 AA.
Sequence status	Complete.
Sequence processing	The displayed sequence is further processed into a mature form.
Protein existence	Evidence at protein level.

General annotation (Comments) Hide | Top

Function	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.
Catalytic activity	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.
Subunit structure	Monomer.
Involvement in disease	Defects in LYZ are a cause of amyloidosis type 8 (AMYL8) [MIM:105200]; also known as systemic non-neuropathic amyloidosis or Ostertag-type amyloidosis. AMYL8 is a hereditary generalized amyloidosis due to deposition of apolipoprotein A1, fibrinogen and lysozyme amyloids. Viscera are particularly affected. There is no involvement of the nervous system. Clinical features include renal amyloidosis resulting in nephrotic syndrome, arterial hypertension, hepatosplenomegaly, cholestasis, petechial skin rash. (Ref.19)
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



Anotacije

Strukturne anotacije (iz 3D zgradb)



Sequence annotation (Features) Hide Top					
Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Molecule processing					
<input type="checkbox"/> Signal peptide	1 – 18	18	Ref.7 Ref.8 Ref.9 Ref.10		
<input type="checkbox"/> Chain	19 – 148	130	Lysozyme C		PRO_0000018467
Sites					
<input type="checkbox"/> Active site	53	1			
<input type="checkbox"/> Active site	71	1			
Amino acid modifications					
<input type="checkbox"/> Disulfide bond	24 ↔ 146				
<input type="checkbox"/> Disulfide bond	48 ↔ 134				
<input type="checkbox"/> Disulfide bond	83 ↔ 99				
<input type="checkbox"/> Disulfide bond	95 ↔ 113				
Natural variations					
<input type="checkbox"/> Natural variant	74	1	I → T in AMYL8. Ref.19		VAR_004280
<input type="checkbox"/> Natural variant	85	1	D → H in AMYL8. Ref.19		VAR_004281
<input type="checkbox"/> Natural variant	88	1	T → N: dbSNP rs1800973.		VAR_012050
Experimental info					
<input type="checkbox"/> Sequence conflict	10	1	V → A in AAC63078. Ref.5		
<input type="checkbox"/> Sequence conflict	41	1	I → M in AAA36188. Ref.1		
<input type="checkbox"/> Sequence conflict	111	1	V → A in AAC63078. Ref.5		
<input type="checkbox"/> Sequence conflict	124	1	I → V in AAC63078. Ref.5		
<input type="checkbox"/> Sequence conflict	128	1	V → A in AAC63078. Ref.5		
<input type="checkbox"/> Sequence conflict	136	1	N → D in AAC63078. Ref.5		
Secondary structure					
1 148					
■ Helix ■ Strand ■ Turn					
Details...					

Nucleotide - Human lysozyme mRNA, complete cds - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.ncbi.nlm.nih.gov/nuccore/187245?report=genbank

Novo na ARRS Kvakadabra - časopis...

Wikipedia go random watchlist talk/article edit + history watch unwatch move WLH NR

NCBI Nucleotide

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Nucleotide for Go Clear

Limits Preview/Index History Clipboard Details

Format: GenBank FASTA Graphics More Formats

GenBank: M21119.1

Human lysozyme mRNA, complete cds

[Comment](#) [Features](#) [Sequence](#)

LOCUS M21119.1 746 bp mRNA linear PRI 11-JUN-1993

DEFINITION Human lysozyme mRNA, complete cds.

ACCESSION M21119

VERSION M21119.1 GI:187245

KEYWORDS lysozyme.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 746)

AUTHORS Castanon,M.J., Spevak,W., Adolf,G.R., Chlebowski-Sledziewska,E. and Sledziewski,A.

TITLE Cloning of human lysozyme gene and expression in the yeast *Saccharomyces cerevisiae*

JOURNAL Gene 66 (2), 223-234 (1988)

PUBMED [2971532](#)

COMMENT Original source text: Human placenta, cDNA to mRNA, clone HL-51. Draft entry and computer-readable sequence for [1] kindly provided by M.J.Castanon, 26-OCT-1988.

FEATURES

Location/Qualifiers

source 1..746
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"

CDS 14..460
 /note="lysozyme precursor (EC 3.2.1.17)"
 /codon_start=1
 /protein_id="AAA36188.1"
 /db_xref="GI:307141"
 /translation="MGALIVLGLVLLSVTVQKVFERCELAKLKLKLGMDYRGSLSA
 MMSCLAWDSYVYTRATMYSAGDRSTIQGIFQMSKQCNMGKTPGAVNAHLSLCSAL
 LQDMADAWACAKKQVSDPQIERMAMWREKQKRDVRFVQGGCV"

sig_peptide 14..67
 /note="lysozyme signal peptide"

mat_peptide 68..457
 /product="lysozyme"

ORIGIN 52 bp upstream of HmlI site.

```

1 cctagcagtc aacatgaagg ctctcattgt ctctgggctt gctctcttt ctgttarggt
61 ccagggaag gttcttgaaa ggtgtgagtt ggcagaact ctgaaaagt tgggaatgga
121 tggctaragg ggaatgagct tagcaaatg gatgtgttt gcaaaatgg agagtgggta
181 caacacarga gctacaaact acaatgctgg agacagaag actgattatg ggaatcttca
241 gatcaatag cgtactcgtt gtaatgatgg caaaaccca ggaagagttt atgctgttca
201 tttatctctc agtgctttgc tgcagaata cctcgtgat gctgtagctt gtcgaaag
261 ggttgctctg gatccacaag gattagagc atgggtggca tggagaattt gctgtcaaaa
421 cagagatgct cgtcagatg ttaaggttg tgaagtgtta ctccagaatt tctcttcttc
481 agtctatctt gctctctca caattaagg agtaggtta gtagaagtc acataccatt
541 atctctcttt caaacaaata atattcttca agaaagagga gcaaaatag cctctctctc
601 taagagatat aatgttact aatgtggct attttatat taagcttca acatttttca
661 gtttcaaat agaartaata ctggtgaaa tttactaaa acrttgggta ttaaatatc
721 ctccagtaca tctcgtctt tttttttg

```

V bistvu samo zaporedje



3. Presežnost/nepresežnost (redundancy)

Podatkovne zbirke so lahko

PRESEŽNE- npr. GeneBank; vsebujejo več identičnih podatkov.

NEPRESEŽNE- npr. UniProt. Bolj pregledne. Lažje in hitrejše iskanje podatkov.

- [Homo sapiens lactotransferrin \(LTF\), mRNA](#)
14. 2,390 bp linear mRNA
NM_002343.2 GI:54607119

- [Homo sapiens arylsulfatase B \(ARSB\), transcript variant 1, mRNA](#)
15. 6,076 bp linear mRNA
NM_000046.3 GI:158634485

- [Human lysozyme mRNA, complete cds](#)
16. 1,483 bp linear mRNA
M19045.1 GI:187247

- [Human lysozyme mRNA, complete cds with an Alu repeat in the 3' flank](#)
17. 1,487 bp linear mRNA
J03801.1 GI:187243

- [Homo sapiens alpha-lactalbumin \(LALBA\) mRNA, complete cds](#)
18. 701 bp linear mRNA
J00270.1 GI:186830

- [Homo sapiens lactalbumin, alpha-, mRNA \(cDNA clone MGC:138521 IMAGE:8327784\), complete cds](#)
19. 581 bp linear mRNA
BC112316.1 GI:85567732

- [Homo sapiens lysozyme precursor, mRNA, complete cds](#)
20. 456 bp linear mRNA
U25677.1 GI:847819

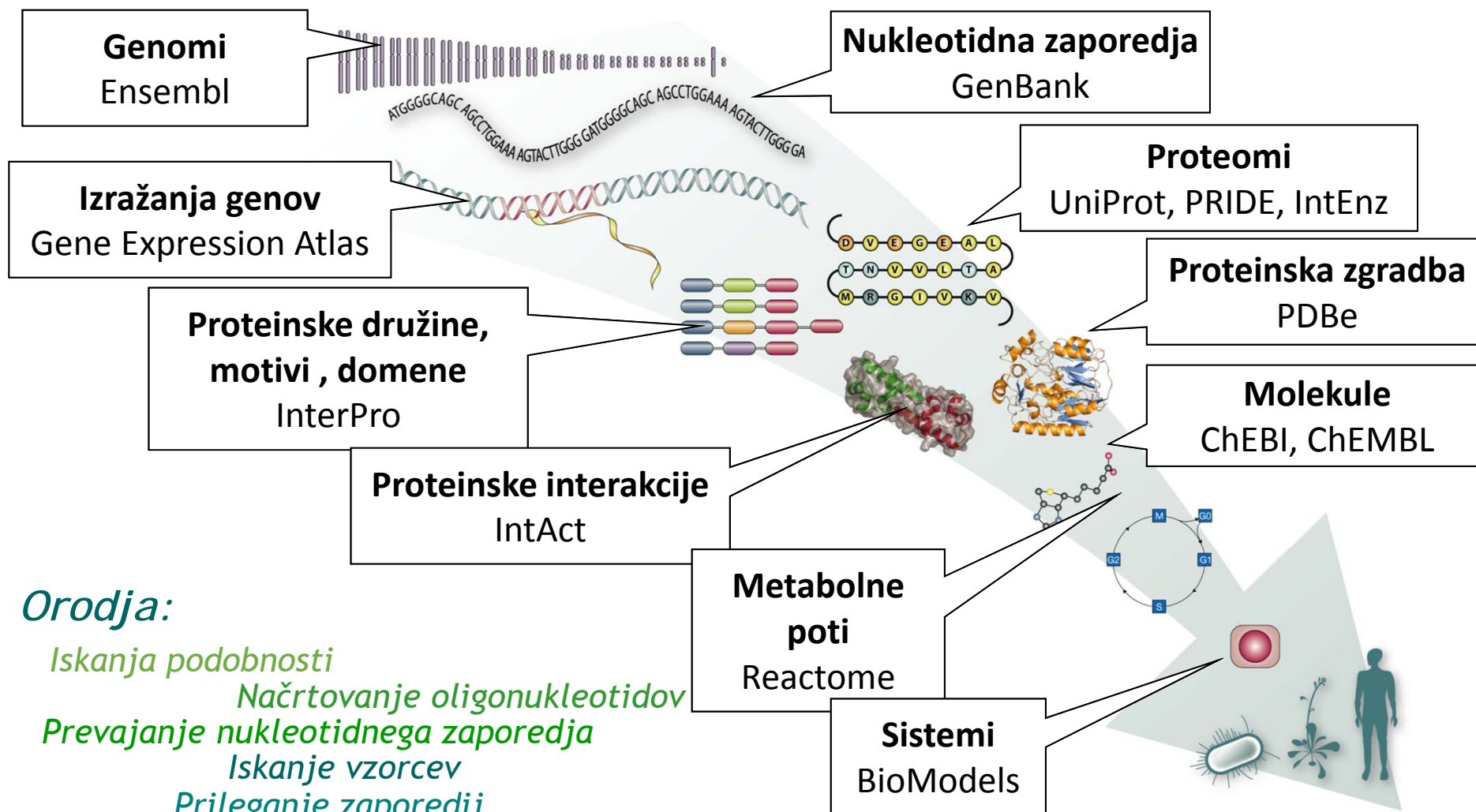


V bistvu isto
zaporedje



Podatkovne zbirke

od molekul do sistemov



Orodja:

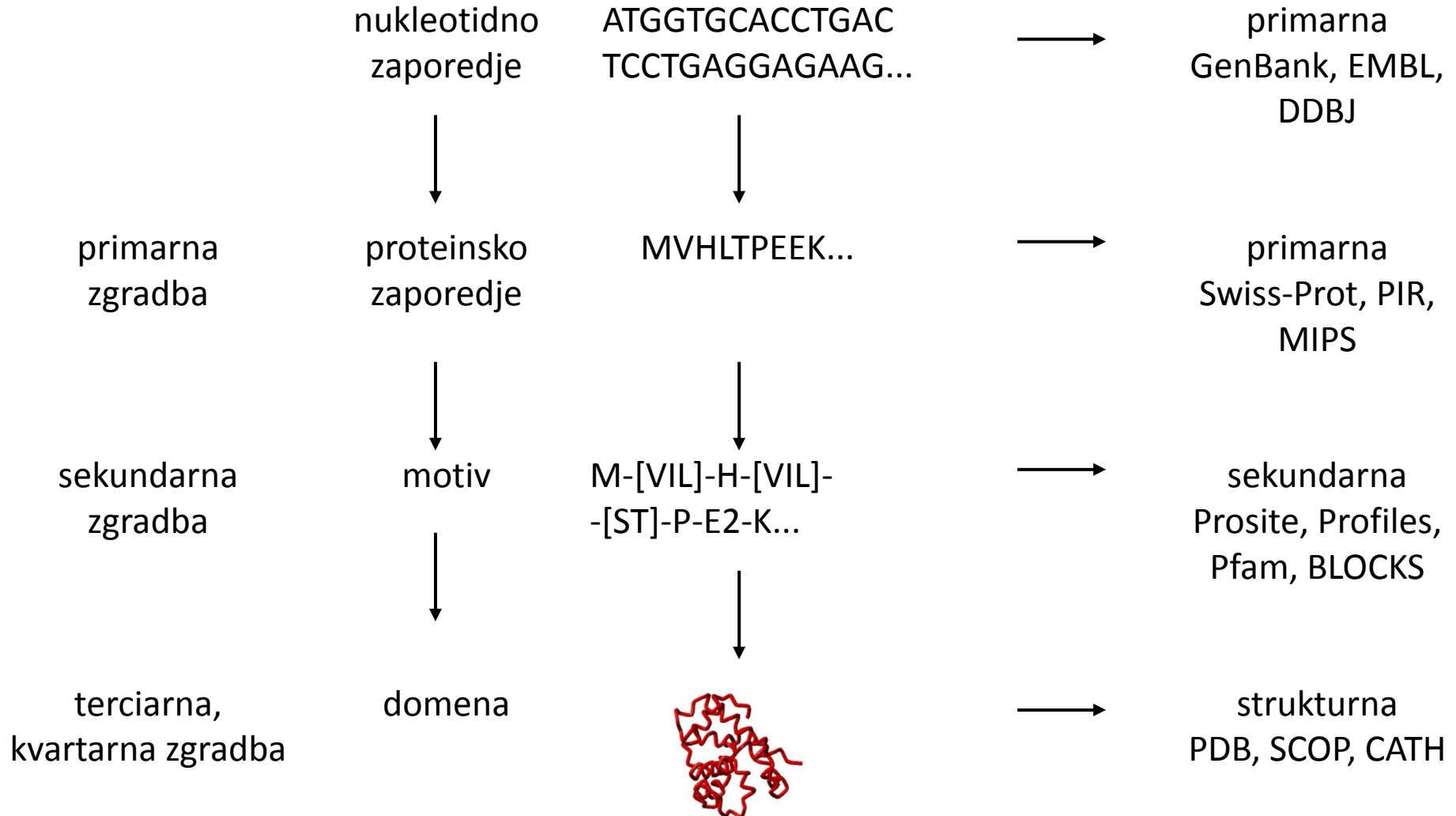
- Iskanja podobnosti
- Načrtovanje oligonukleotidov
- Prevajanje nukleotidnega zaporedja
- Iskanje vzorcev
- Prileganje zaporedij
- Napoved genov
- Napoved zgradbe

PODATKOVNE ZBIRKE

PRIMARNE
SEKUNDARNE

(“archival”, arhivske)
(“curated”, izpeljane)

Podatkovna zbirka



NUKLEOTIDNE PODATKOVNE ZBIRKE

Dnevna izmenjava zaporedij!

DDBJ *DNA Data Bank of Japan*

Zaporedja javno dostopnih virov in večjih genomskih projektov

EMBL

Podatkovna zbirka EBI. Vsebuje direktno vpisana zaporedja, rezultate določevanj zaporedij genomov, zaporedja iz literature in patentov.

Iskanje in primerjave zaporedij preko vmesnikov

GenBank

Zaporedja javno dostopnih virov in večjih genomskih projektov.

Dnevna izmenjava podatkov z DDBJ in EMBL.

Razdeljena na posamezne odseke (nivo organizmov, EST, PAT, STS, GSS, HTG).

157.943.793.171 baz v **171.123.749** zaporedjih (14. 2. 2014)

iz več kot **165 000** organizmov

Dostop preko Entreza (Medline ali BLAST)

dbEST podatkovna zbirka oznak izraženih zaporedij

Ostale specializirane podatkovne zbirke

GSDB (The Genome Sequence DataBase), SGD (*Saccharomyces* Genome Database), UniGene, TDB (TIGR podatkovna zbirka), ACeDB (*A. C. elegans* DataBase)

STRUKTURA GenBank ZAPISA

GBFF- GenBank flat file oblika zapisa v GenBank. Zapis, ki si ga izmenjujejo podatkovne zbirke.

Glava (*header*)

Lastnosti

Zaporedje

```
LOCUS          NM_000239                1487 bp    mRNA    linear    PRI 18-DEC-2001
DEFINITION    Homo sapiens lysozyme (renal amyloidosis) (LYZ), mRNA.
ACCESSION     NM_000239
VERSION       NM_000239.1  GI:4557893
KEYWORDS      .
SOURCE        human.
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 1487)
  AUTHORS     Reitamo,S., Klockars,M., Adinolfi,M. and Osserman,E.F.
  TITLE       Human lysozyme (origin and distribution in health and disease)
  JOURNAL     Ric Clin Lab 8 (4), 211-231 (1978)
  MEDLINE     79097291
  PUBMED     366724
```

Glava

LOCUS

NM_000239

1487 bp

mRNA

linear

PRI 18-DEC-2001

kratka oznaka zapisa
včasih ime lokusa
Npr. HUMHBB človeški
β-globinski lokus

dolžina
Manj kot 50 bp
ne sprejmejo, več
kot 350 kb tudi
ne

tip zaporedja
DNA, tRNA, rRNA,
mRNA, uRNA

Odsek GenBank

PRI primati
ROD glodalci
MAM ostali sesalci
VRT ostali vretenčarji
INV nevretenčarji
PLN rastline
BCT bakterije
VRL virusi
PHG bakteriofag
SYN sintetične
EST oznaka izraženega zaporedja
PAT patent
STS sequence tagged sites
GSS genome survey sequences
HTG "high throughput" genomska zaporedja
HTC nedokončane "high throughput" EST zap.

datum, ko je zapis
postal javen

DEFINITION Homo sapiens lysozyme (renal amyloidosis) (LYZ), mRNA.

Biološki opis zaporedja (izpisan v FASTA formatu)

ACCESSION NM_000239 NT_123456 constructed genomic contigs
NM_123456 mRNAs
Edinstvena koda, se ne spreminja, podana v publikacijah NP_123456 proteins
NC_123456 chromosomes

VERSION NM_000239.1 GI:4557893

Katera verzija zaporedja je. Se spreminja. GI (*geneInfo identifier*) edinstveno določena koda za zaporedje (enaka med različnimi podatkovnimi zbirkami). Tudi PID (*protein identifier*) v primeru prevedenih proteinskih zaporedij.

KEYWORDS .

SOURCE human.

ORGANISM [Homo sapiens](#)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Izvor zaporedja.

REFERENCE 1 (bases 1 to 1487)

AUTHORS Reitamo,S., Klockars,M., Adinolfi,M. and Osserman,E.F.

TITLE Human lysozyme (origin and distribution in health and disease)

JOURNAL Ric Clin Lab 8 (4), 211-231 (1978)

MEDLINE [79097291](#)

PUBMED [366724](#)

Bibliografske enote povezane z zaporedjem (kdaj sklonirano, kdaj določena sturktura gena...)


```

FEATURES             Location/Qualifiers
     source           1..1487
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="12"
                       /map="12q15"
                       /clone="pHL-2"
                       /cell_line="PMA treated U937"
     gene              1..1487
                       /gene="LYZ"
                       /db_xref="LocusID:4069"
                       /db_xref="MIM:153450"
     sig\_peptide     26..469
     CDS              26..472
                       /gene="LYZ"
                       /EC_number="3.2.1.17"
                       /codon_start=1
                       /db_xref="LocusID:4069"
                       /db_xref="MIM:153450"
                       /product="lysozyme precursor"
                       /protein_id="NP_000230.1"
                       /db_xref="GI:4557894"
                       /translation="MKALIVLGLVLLSVTVQGVKVFERCELARTLKRLGMDGYRGISLA
NWMCLAKWESGYNTRATNYNAGDRSTDYGIFQINSRYWCNDGKTPGAVNACHLSCSAL
LQDNIADAVACAKRVVRDPQGIRAWVAWRNRCQNRDVRQYVQGCYV"
     mat\_peptide     26..469
                       /product="lysozyme"
     misc\_feature     80..463
                       /note="lys; Region: C-type lysozyme/alpha-lactalbunin
family. Alpha-lactalbunin is the regulatory subunit of
lactose synthase"
     misc\_feature     80..466
                       /note="LYZ1; Region: Alpha-lactalbunin / lysozyme C"
     variation        288
                       /allele="C"
                       /allele="A"
                       /db_xref="dbSNP:1800973"
     repeat\_region    740..1054
                       /note="Alu repeat"

```

Lastnosti zaporedja.

Svojo lastno strukturo in povezave
z drugimi podatkovnimi zbirkami

Zaporedje

BASE COUNT 435 a 306 c 308 g 438 t

ORIGIN

```
1  ctagcactct gacctagcag tcaacatgaa ggctctcatt gttctggggc ttgtcctcct
61 ttctgttacg gtccagggca aggtctttga aagggtgtgag ttggccagaa ctctgaaaag
121 attgggaatg gatggctaca ggggaatcag cctagcaaac tggatgtggt tggccaaatg
181 ggagagtggg tacaacacac gagctacaaa ctacaatgct ggagacagaa gcactgatta
241 tgggatattt cagatcaata gccgctactg gtgtaatgat ggcaaaacc caggagcagt
301 taatgcctgt catttatcct gcagtgcttt gctgcaagat aacatcgctg atgctgtagc
361 ttgtgcaaag agggttgtcc gtgatccaca aggcattaga gcatgggtgg catggagaaa
421 tcgttgtcaa aacagagatg tccgtcagta tgttcaagg tgtggagtgt aactccagaa
481 ttttccttct tcagctcatt ttgtctctct cacattaagg gagtaggaat taagtgaaag
541 gtcacactac cattatttcc ccttcaaaca aataatattt ttacagaagc aggagcaaaa
601 tatggccttt cttctaagag atataatggt cactaatgtg gttattttac attaagccta
661 caacatTTTT cagtttgcaa atagaactaa tactggtgaa aatttaccta aaaccttggg
721 tatcaaatac atctccagta cattccgctc tttttttttt ttgagacagt ctcgctctgt
781 cgcccaggct ggagtgcagt ggcgcaatct cggctcactg caacctccac ctcccggggt
841 cacgccattc tctgcctca gcctcccag tagctgggat tacgggcgcc cgccaccacg
901 cccggcta at tttttgtatt tttagtagag acagggtttc accgtgttag ccaggatggg
961 ctgatctcc tgacctgtg atccaccac ctggcctcc caaagtgtg ggattacagg
1021 cgtgagccac tgcgcccggc cacattcagt tcttatcaaa gaaataacc agacttaatc
1081 ttgaatgata cgattatgcc caatattaag taaaaaatat aagaaaaggg tatcttaaat
1141 agatcttagg caaaatacca gctgatgaag gcatctgatg cttcatctg ttcagtcatc
1201 tccaaaaaca gtaaaaataa ccactttttg ttgggcaata tgaaattttt aaaggagtag
1261 aataccaaat gatagaaaca gactgcctga attgagaatt ttgatttctt aaagtgtgtt
1321 tctttctaaa ttgctgttcc ttaatttgat taatttaatt catgtattat gattaaatct
1381 gaggcagatg agcttacaag tattgaaata attactaatt aatcacaat gtgaagttat
1441 gcatgatgta aaaaatacaa acattcta at taaaggcttt gcaacac
```

//

VSMGLDAVDE SSMTGSFGGS NAQTSTEEVS QOSTDIMALL DNNMLGSMGD
T L A S I T F E T K R L M E P V E E L R D E I Q I M N V P G A G P L P A G P F A Q M N L
K I I H E S E V I N G N S I L N V T G Y S V E E I Q D I F L
N P P F A T E R Y I T N L R L F T K F E T V G H A H I A G S K F A P N P N
Q S L S L M F S S S S E S N L R S R L E A E L R R E E E A E N D E A Q X Q X M
UniProt
the universal protein resource



<http://www.expasy.uniprot.org/index.shtml>

UniProt (Universal Protein Resource) is the world's most comprehensive catalog of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR.

UniProt is comprised of three components, each optimized for different uses. The **UniProt Knowledgebase (UniProt)** is the central access point for extensive curated protein information, including function, classification, and cross-reference. The **UniProt Non-redundant Reference (UniRef)** databases combine closely related sequences into a single record to speed searches. The **UniProt Archive (UniParc)** is a comprehensive repository, reflecting the history of all protein sequences.

KAJ POTREBUJE NEKDO, KI SE UKVARJA S PROTEINI?

Kvalitetno podatkovno zbirko: nepresežno zbirko, čimboljša pokritost (izoforme, alternativno izrezovanje, variante pri boleznih).

Dobro identifikacijo proteinov: dobri označevalci, konsistentna nomenklatura.

Dobro določitev proteinov: podrobna informacija o proteinski funkciji, biološkem procesu, molekularnih interakcijah in udeležbi v različnih poteh in omrežjih. Reference na zunanje vire.

Swiss-Prot & TrEMBL



PIR-PSD





UniProtKB/TrEMBL

Presežna zbirka, avtomatsko anotirana, nepregledana.

V UniProtKB/TrEMBL zapis predstavlja prevod enega ENA (nukleotidnega) zaporedja.

Več zapisov za isti protein je lahko rezultat:

- Napačnih napovedi genov
- Napak v zaporedjih (npr. premikov bralnega okvirja)
- Polimorfizmov
- Alternativnih startnih kodonov
- Izoform

ALI

- Ker so zaporedja vpisali različni raziskovalci

UniProtKB/Swiss-Prot

Nepresežna zbirka, zelo kvalitetna ročna anotacija, pregledana.

V UniProtKB/Swiss-Prot zapis predstavlja prevod enega proteina.

Vsa identična zaporedja so združena in dodatno analizirana s kuratorji, ki jim tudi pripišejo anotacije.

Zanesljivost zaporedij

- 1 Dokaz na nivoju proteina
- 2 Dokaz na nivoju transkripta
- 3 Določen na podlagi homologije
- 4 Napovedan
- 5 Negotov

UniProtKB/TrEMBL

Protein existence (PE):	entries	%
1: Evidence at protein level	12724	0.08%
2: Evidence at transcript level	534559	3.17%
3: Inferred from homology	3847758	22.79%
4: Predicted	12491797	73.97%
5: Uncertain	0	0.00%

UniProtKB/Swiss-Prot

Protein existence (PE):	entries	%
1: Evidence at protein level	72765	13.7%
2: Evidence at transcript level	69863	13.1%
3: Inferred from homology	373177	70.1%
4: Predicted	14474	2.7%
5: Uncertain	1867	0.4%

Anotirane proteine lahko uporabimo kot vodila za proteine, ki ji še niso podrobno študirali:

- 149 zapisov za ime „rhodopsin“ manualno anotiranih v UniProtKB/Swiss-Prot
- 7424 zapisov za ime „rhodopsin“ računalniško anotiranih v UniProtKB/TrEMBL

Informacije, ki jih dodajo UniProt/Swiss-Prot zapisu

- [1] "The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins."
Ebersole T.A., Chen Q., Justice M.J., Artzt K.
Nat. Genet. 12:260-265 (1996) [PubMed: 8589716] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 3), INVOLVEMENT IN QKV, TISSUE SPECIFICITY, MUTAGENESIS [GLUTAMINE]
- [2] "Genomic organization and expression analysis of the mouse qki locus."
Kondo T., Furuta T., Mitsunaga K., Ebersole T.A., Shichiri M., Wu J., Artzt K., Yamamura K., Abe K.
Mamm. Genome 10:662-669 (1999) [PubMed: 10384037] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA / MRNA] (ISOFORMS 2; 3; 4 AND 7), ALTERNATIVE SPLICING (ISOFORM 1).
Strain: 129/J.

Reference

General annotation (Comments)	Links
<p>Function</p> <p>RNA-binding protein that plays a central role in myelination. Also required for visceral endoderm function and blood vessel development. Binds to the 5'-UNCUUAAUAG-3' UAAU-3' RNA core sequence. Acts by regulating pre-mRNA splicing, mRNA export, mRNA stability and protein translation, as well as cellular processes including apoptosis, cell cycle, glial cell fate and development. Required to protect and promote stability of mRNAs such as MBP and CCHNB3 to promote oligodendrocyte differentiation. Participates in mRNA transport by regulating the nuclear export of MBP mRNA. Isoform 1 is involved in regulation of mRNA splicing of MAG pre-mRNA by acting as a negative regulator of MAG exon 12 alternative splicing. Isoform 3 can induce apoptosis, while heterodimerization with other isoforms result in nuclear translocation of isoform 3 and suppression of apoptosis. Isoform 4 acts as a translational repressor for GLUT1. May also play a role in smooth muscle development.</p>	
<p>Subunit structure</p> <p>Homodimer. Does not require RNA to homodimerize. Able to heterodimerize with BIC1.</p>	
<p>Subcellular location</p> <p>Cytoplasm. Also in nucleus. 1 localizes primarily to nucleus and is liver level 1 cytoplasmic. It shuttles between the cytoplasm and nucleus. Isoform 2 is cytoplasmic. Isoform 3 is nuclear. Isoform 4 localizes both in the cytoplasm and nucleus.</p>	
<p>Tissue specificity</p> <p>Highly expressed in myelinating oligodendrocytes and oligocytes in the peripheral nervous system as well as Schwann cells in the peripheral nervous system. Also expressed in the yolk sac, endoderm, adjacent to the mesodermal site of developing blood islands, where the differentiation of blood and endothelial cells first occurs (at protein level). Expressed in brain, lung, heart and testes.</p>	
<p>Developmental stage</p> <p>Expressed in ventral progenitors of the ventricular zone (vz) during CNS development, but that expression is down-regulated during neuronal maturation. Expression in the vz maintains expression as they differentiate and migrate. Expression in the vz ceases with the acquisition of a glial rather than neuronal phenotype. Expression in the vz is maintained in the adult brain. Expressed in the developing myelin sheath. Expressed in E17.5. Expression is strongly present ventrally in the rat brain. Expressed in the developing brain and heart. Expressed in early embryos, while isoform 3 and isoform 4 are found in late development when myelination begins.</p>	
<p>Post-translational modification</p> <p>Methylated by PRMT1.</p> <p>Tyrosine phosphorylated at its C-terminus, probably by FYN. Phosphorylation leads to decreased mRNA-binding affinity, affecting transport and/or stabilization of MBP mRNA. The level of Ty phosphorylation in the developing myelin is highest in the first postnatal week (P7). During the vigorous accumulation of MBP mRNA between P7 and P20, phosphorylation in the developing myelin drastically declined. By the end of the fourth postnatal week (P28), phosphorylation is reduced approximately 90%.</p>	
<p>Involvement in disease</p> <p>Defects in Qki are the cause of quaking/shake (qkv). Qkv is a spontaneous mutation resulting in hypomyelination of the central and peripheral nervous systems. Mutant mice develop normally until postnatal day 10 when they display rapid tremors or 'quaking' that is especially pronounced in hindlimbs, and experience convulsive tonic-clonic seizures as they mature. Mice with qkv specifically lack isoform 3.</p>	

Določitve iz literature

Cell cycle	Regulation of cell proliferation
DNA damage	Traceable author statement. Source: UniProtKB
DNA repair	Regulation of transcription from RNA polymerase II promoter
Fatty acid biosynthesis	Traceable author statement. Source: Protinc
Lipid synthesis	Regulation of transcription from RNA polymerase III promoter
Nucleus	Traceable author statement. Source: UniProtKB
Polymorphism	See also: QKV in humans
Disease mutation	Inferred from direct assay. Source: UniProtKB
Repeat	BRCA1-BARD1 complex
Zinc-finger	Inferred from direct assay. Source: UniProtKB
DNA-binding	Gamma-tubulin ring complex
Metal-binding	Non-traceable author statement. Source: UniProtKB
Zinc	DNA binding
Anti-oncogene	Traceable author statement. Source: Protinc
Phosphorylation	Androgen receptor binding
3D-structure	Non-traceable author statement. Source: UniProtKB
	Enzyme binding
	Inferred from physical interaction. Source: UniProtKB

Ontologije

KURATORJI



```

10      20      30      40      50      60
MVGEMETKEK PKPTPDYLMQ LMNDKMLSS LNFFCGIFNH LERLLDEEIS RVRKIMYNDT

70      80      90      100     110     120
LNGSTEKRSA ELPDAVGPIV QLQEKLYVIV KEYDFNFVVG RILGFRGLTA KQLEATGCKK

130     140     150     160     170     180
IMVVRGKSMR DKKKLEKLEK EAVVHVNINR DAVYKQTEFAV RAEIKLKR AVEEVKLLIV

190     200     210     220     230     240
PAABGEDSLK RMQLMELAIL NGTYRDANIK SPALAFSLAA TAQAAPRIIT GPAPVLPFAA

250     260     270     280     290     300
LRTPTPAGPT IMPLIRQIQI AVMPNGTPEP TAAIVPPGPE AGLIYTFPEY PYTLAPATSI

310     320     330     340
LEYPIEPSGV LGAVATKVRK HDMRVHPYQR IVTADRAATG N
    
```

Zaporedje

Isoform 1 (identifier: Q9QYS9-1)
Also known as QKI-5;
This isoform has been chosen as the 'canonical' sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.

Isoform 2 (identifier: Q9QYS9-2)
Also known as QKI-7B;
The sequence of this isoform differs from the canonical sequence as follows:
312-341 GAVATKVRKDMRVHPYQRIVTADRAATG → VMVQRKAKNSRVLTPESDDNLNLTA

Isoform 3 (identifier: Q9QYS9-3)
Also known as QKI-7;
The sequence of this isoform differs from the canonical sequence as follows:
312-341 GAVATKVRHDMRVHPYQRIVTADRAATG → EWIEMPVMPDISAH

Različice zaporedja

Protein names

Protein quaking

Also known as:

Mqkl
Qki

Nomenklatura

Gene names

Name: **Qki**
Synonyms: Qk, Qk1, Qka1

Molecule processing

Chain 1–341 341 Protein quaking

Regions

Domain 87–153 67 KH

Motif 276–279 4 SH3-binding

Motif 324–330 7 Nuclear localization signal

Lastnosti zaporedja

Search Blast* Align Retrieve ID Mapping*
Search in Query
Protein Knowledgebase (UniProtKB) Search Clear Fields »

★ Reviewed, UniProtKB/Swiss-Prot **P61626** (LYSC_HUMAN)
Last modified January 19, 2010. Version 78. History...

Clusters with 100%, 90%, 50% identity | Documents (7) | Third-party data | Customize display
TEXT XML RDF/XML GFF FASTA

Names and origin Protein attributes General annotation (Comments) Ontologies Binary interactions Sequence annotation (Features) Sequences References Web resources Cross-references Entry information Relevant documents

Names and origin Hide | Top

Protein names	Recommended name: Lysozyme C EC=3.2.1.17 Alternative name(s): 1,4-beta-N-acetylmuramidase C
Gene names	Name: LYZ Synonyms: LZM
Organism	Homo sapiens (Human) [Complete proteome]
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominoidea > Homo

Protein attributes Hide | Top

Sequence length	148 AA.
Sequence status	Complete.
Sequence processing	The displayed sequence is further processed into a mature form.
Protein existence	Evidence at protein level.

General annotation (Comments) Hide | Top

Function	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.
Catalytic activity	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.
Subunit structure	Monomer.
Involvement in disease	Defects in LYZ are a cause of amyloidosis type 8 (AMYL8) [MIM:105200]; also known as systemic non-neuropathic amyloidosis or Ostertag-type amyloidosis. AMYL8 is a hereditary generalized amyloidosis due to deposition of apolipoprotein A1, fibrinogen and lysozyme amyloids. Viscera are particularly affected. There is no involvement of the nervous system. Clinical features include renal amyloidosis resulting in nephrotic syndrome, arterial hypertension, hepatosplenomegaly, cholestasis, petechial skin rash. [Ref.19]
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.
Sequence similarities	Belongs to the glycosyl hydrolase 22 family .

STRUKTURA SWISS-PROT ZAPISA

SWISS-PROT: P00695

NiceProt - a user-friendly view of this SWISS-PROT entry

ID	LYC_HUMAN	STANDARD;	PRT;	148 AA.	PROTEIN_IZVOR	se lahko spreminjna	
AC	P00695;	Q13170;			SwissProt koda	se ne spreminja	
DT	21-JUL-1986	(Rel. 01,	Created)		DT	datum vnosa	
DT	01-OCT-1989	(Rel. 12,	Last sequence update)				
DT	16-OCT-2001	(Rel. 40,	Last annotation update)				
DE	Lysozyme C precursor	(EC 3.2.1.17)	(1,4-beta-N-acetylmuramidase C).	DE	opis proteina		
GN	LYZ OR LZM.			GN	ime gena		
OS	Homo sapiens (Human) ,			OS	organizem		
OS	Pan troglodytes (Chimpanzee) ,	and					
OS	Pan paniscus (Pygmy chimpanzee)	(Bonobo).					
OC	Eukaryota ;	Metazoa ;	Chordata ;	Craniata ;	Vertebrata ;	Euteleostomi ;	
OC	Mammalia ;	Eutheria ;	Primates ;	Catarrhini ;	Hominidae ;	Homo .	
OX	NCBI_TaxID= 9606 ,	9598 ,	9597 ;				
RN	[1]						
RP	SEQUENCE FROM N.A.			RN	bibliografski	podatki	
RC	SPECIES=Human;						
RX	MEDLINE=89325294;	PubMed=2546758;	[NCBI ,	ExPASy ,	EBI ,	Israel ,	Japan]
RA	Peters C.W.B. ,	Kruse U. ,	Pollwein R. ,	Grzeschik K.H. ,	Sippel A.E. ;		
RT	"The human lysozyme gene. Sequence organization and chromosomal						
RT	localization."						
RL	Eur. J. Biochem. 182:507-516(1989).						
RN	[2]						
RP	SEQUENCE FROM N.A.						
RC	SPECIES=Human;						
RX	MEDLINE=89006264;	PubMed=2971592;	[NCBI ,	ExPASy ,	EBI ,	Israel ,	Japan]
RA	Peters C.W.B. ,	Kruse U. ,	Pollwein R. ,	Grzeschik K.H. ,	Sippel A.E. ;		
RT	"The human lysozyme gene. Sequence organization and chromosomal						
RT	localization."						
RL	Eur. J. Biochem. 182:507-516(1989).						
RN	[2]						
RP	SEQUENCE FROM N.A.						
RC	SPECIES=Human;						
RX	MEDLINE=89006264;	PubMed=2971592;	[NCBI ,	ExPASy ,	EBI ,	Israel ,	Japan]
RA	Castanon M.J. ,	Spevak W. ,	Adolf G.R. ,	Chlebowicz-Sledziewska E. ,			
RA	Sledziewski A. ;						
RT	"Cloning of human lysozyme gene and expression in the yeast						
RT	<i>Saccharomyces cerevisiae</i> ."						
RL	Gene 66:223-234(1988).						

.....
CC -!- FUNCTION: LYSOZYMES HAVE PRIMARILY BACTERIOLYTIC FUNCTION; THOSE
CC IN TISSUES AND BODY FLUIDS ARE ASSOCIATED WITH THE MONOCYTE-
CC MACROPHAGE SYSTEM AND ENHANCE THE ACTIVITY OF IMMUNOAGENTS.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC heteropolymers of the prokaryotes cell walls.
CC -!- SUBUNIT: MONOMER.
CC -!- DISEASE: DEFECTS IN LYZ CAN BE THE CAUSE OF HEREDITARY NON-
CC NEUROPATHIC SYSTEMIC AMYLOIDOSIS (OSTERTAG-TYPE) (AMYLOIDOSIS
CC VIII).
CC -!- MISCELLANEOUS: LYSOZYME C IS CAPABLE OF BOTH HYDROLYSIS AND
CC TRANSGLYCOSYLATION; IT SHOWS ALSO A SLIGHT ESTERASE ACTIVITY. IT
CC ACTS RAPIDLY ON BOTH PEPTIDE-SUBSTITUTED AND UNSUBSTITUTED
CC PEPTIDOGLYCAN &, SLOWLY, ON CHITIN OLIGOSACCHARIDES.
CC -!- MISCELLANEOUS: THE LYSOZYMES ISOLATED FROM HUMAN MILK AND FROM THE
CC URINE OF PATIENTS WITH CHRONIC LEUKEMIA ARE IDENTICAL.
CC -!- SIMILARITY: BELONGS TO FAMILY 22 OF [GLYCOSYL](#) HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03801; AAA59535.1; -. [[EMBL](#) / [GenBank](#) / [DDBJ](#)] [[CoDingSequence](#)]

CC

komentarji

.....
DR PIR; A00849; LZHU.
DR PIR; S04938; S04938.
.....
DR PDB; 1EQ5; 19-APR-00. [[ExPASy](#) / [RCSE](#)]
DR PDB; 1EQE; 19-APR-00. [[ExPASy](#) / [RCSE](#)]
DR [CarbBank](#); [CCSD:42661](#); -.
DR MIM; 153450; -. [[NCBI](#) / [EBI](#)]
DR [GeneCards](#); [LYZ](#).
DR [GeneLynx](#); [LYZ](#).
.....

DR

zapis v
drugih
zbirkah

FT	SIGNAL	<u>1</u>	<u>18</u>	
FT	CHAIN	<u>19</u>	<u>148</u>	LYSOZYME C.
FT	DISULFID	<u>24</u>	<u>146</u>	
FT	DISULFID	<u>48</u>	<u>134</u>	
FT	DISULFID	<u>83</u>	<u>99</u>	
FT	DISULFID	<u>95</u>	<u>113</u>	
FT	ACT_SITE	<u>53</u>	<u>53</u>	
FT	ACT_SITE	<u>71</u>	<u>71</u>	
FT	VARIANT	<u>74</u>	<u>74</u>	I -> T (IN AMYLOIDOSIS VIII).
FT				/FTid=VAR_004280.
.....				
FT	CONFLICT	<u>10</u>	<u>10</u>	V -> A (IN REF. 5).
FT	CONFLICT	<u>41</u>	<u>41</u>	I -> M (IN REF. 2).
.....				
FT	STRAND	<u>20</u>	<u>20</u>	
FT	HELIX	<u>23</u>	<u>32</u>	
FT	TURN	<u>33</u>	<u>34</u>	
FT	TURN	<u>36</u>	<u>37</u>	
FT	STRAND	<u>38</u>	<u>38</u>	
FT	TURN	<u>39</u>	<u>40</u>	
FT	STRAND	<u>41</u>	<u>41</u>	
FT	HELIX	<u>43</u>	<u>54</u>	
FT	TURN	<u>55</u>	<u>56</u>	
FT	STRAND	<u>57</u>	<u>57</u>	
FT	TURN	<u>58</u>	<u>59</u>	
FT	STRAND	<u>61</u>	<u>64</u>	
FT	TURN	<u>65</u>	<u>68</u>	
FT	STRAND	<u>69</u>	<u>72</u>	
FT	TURN	<u>73</u>	<u>76</u>	
FT	STRAND	<u>77</u>	<u>78</u>	
FT	TURN	<u>79</u>	<u>82</u>	
FT	STRAND	<u>84</u>	<u>84</u>	
FT	TURN	<u>89</u>	<u>90</u>	
FT	TURN	<u>94</u>	<u>95</u>	
FT	STRAND	<u>98</u>	<u>98</u>	
FT	HELIX	<u>99</u>	<u>103</u>	
FT	HELIX	<u>108</u>	<u>118</u>	
FT	TURN	<u>119</u>	<u>119</u>	
FT	TURN	<u>121</u>	<u>122</u>	
FT	HELIX	<u>123</u>	<u>126</u>	
FT	HELIX	<u>128</u>	<u>133</u>	
FT	TURN	<u>134</u>	<u>136</u>	
FT	HELIX	<u>140</u>	<u>142</u>	
FT	TURN	<u>143</u>	<u>145</u>	
SQ	SEQUENCE	148 AA; 16537 MW; 8ECFD276BEB2678A CRC64;		
		MKALIVLGLV LLSVTVQGKV FERCELARTL KRLGMDGYRG ISLANWMCLA KWESGYNTRA		
		TNYNAGDRST DYGIFQINSR YWCNDGKTPG AVNACHLSCS ALLQDNIADA VACAQRVVRD		
		PQGIRAWVAW RNRCQNRDVR QYVQCGV		

FT *feature table*
lastnosti proteina

SQ *aminokislinsko zaporedje*

//

OBLIKA ZAPISA

Računalnik vs človek

GBFF

ASN.1

FASTA(Pearson)

najbolj enostaven zapis

>P00695

```
MKALIVLGLVLLSVTVQGKVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGYNTRATNYNAGDRS  
TDYGIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRDPQGIRAWVAWRNRCQNRD  
VRQYVQGCGV
```

>gi|4557894|ref|NP_000230.1| lysozyme precursor [Homo sapiens]

```
MKALIVLGLVLLSVTVQGKVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGYNTRATNYNAGDRS  
TDYGIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRDPQGIRAWVAWRNRCQNRD  
VRQYVQGCGV
```

>gi|4557893|ref|NM_000239.1| Homo sapiens lysozyme (renal amyloidosis)
(LYZ), mRNA

```
CTAGCACTCTGACCTAGCAGTCAACATGAAGGCTCTCATTGTTCTGGGGCTTGTCTCCTTTTCTGTTACG  
GTCCAGGGCAAGGTCTTTGAAAGGTGTGAGTTGGCCAGAACTCTGAAAAGATTGGGAATGGATGGCTACA  
GGGAATCAGCCTAGCAAACCTGGATGTGTTTGGCCAAATGGGAGAGTGGTTACAACACACGAGCTACAAA  
CTACAATGCTGGAGACAGAAGCACTGATTATGGGATATTTTCAGATCAATAGCCGCTACTGGTGTAAATGAT  
GGCAAACCCCAGGAGCAGTTAATGCCTGTCATTTATCCTGCAGTGCTTTGCTGCAAGATAACATCGCTG
```


Brookhaven Protein Data Bank

1973

RCSB Protein Data Bank - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.rcsb.org/pdb/home/home.do

Novo na ARRS Kvakadabra - časopis...

Wikipedia go random watchlist talk/article edit + history watch unwatch move WLH NR purge my + navigate

RCSB PDB PROTEIN DATA BANK

MyPDB Login A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

As of Tuesday Jan 19, 2010 at 4 PM PST there are 62787 Structures | PDB Statistics

WHAT'S NEW | HELP | PRINT

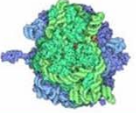
PDB ID or keyword Search Advanced Search

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the wwPDB, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

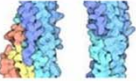
The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Molecule of the Month: 70S Ribosomes



Ribosomes are one of the wonders of the cellular world, and one of the many wonders you can explore yourself at the RCSB PDB. In 2000, structural biologists Venkatesan Ramakrishnan, Thomas A. Steitz and Ada E. Yonath made the first structures of **ribosomal subunits** available in the PDB, and in 2009, they each received the **Nobel Prize** for this work. Structures are also available for many of the other players in protein synthesis, including **transfer RNA** and **elongation factors**. Building on these structures, there are now hundreds of structures of entire ribosomes in the PDB, revealing the atomic details of many important steps in protein synthesis. ■ [Read more ...](#) ■ [Previous Features](#)

PSI Featured Molecule: Hemolysin BL



Researchers at NYSGXRC have solved the structure of the bacterial toxin hemolysin BL, a widespread cause of food poisoning. The pore-forming hydrophobic loop is seen in a soluble, pre-attack conformation. ■ [Read more from the Structural Genomics Knowledgebase](#) ■ [Previous Features](#)


New user? Try the browser [compatibility check](#) and information on [Getting Started](#).

News

- Complete News
- Newsletter
- Discussion Forum
- Job Listings

wwPDB Statement on Retraction of UAB PDB Entries

19-January-2010
NJ Science Olympiad Protein Modeling Results



35 teams across New Jersey expressed their knowledge about hemagglutinin and protein structure at the Science Olympiad. The RCSB PDB sponsored the protein modeling trial event at the three regional competitions. Teams were judged on their knowledge of protein structure and function. [More >>](#)

FTP Archive

The up-to-date PDB archive is available at: <ftp://wwpdb.org>

Time-stamped yearly

Done

start

8:55

100%

četrtek

Queries

- 1J4N
- Download Files
- FASTA Sequence
- Display Files
- Display Molecule
- Structural Reports
- Structure Analysis
- Help

1J4N

Title Crystal Structure of the AQP1 water channel

Authors Sui, H., Han, B.-G., Lee, J.K., Walian, P., Jap, B.K.

Primary Citation Sui, H., Han, B.G., Lee, J.K., Walian, P., Jap, B.K. Structural basis of water-specific transport through the AQP1 water channel. *Nature* v414 pp.872-878, 2001
[Abstract]

History Deposition 2001-10-19 Release 2002-03-27

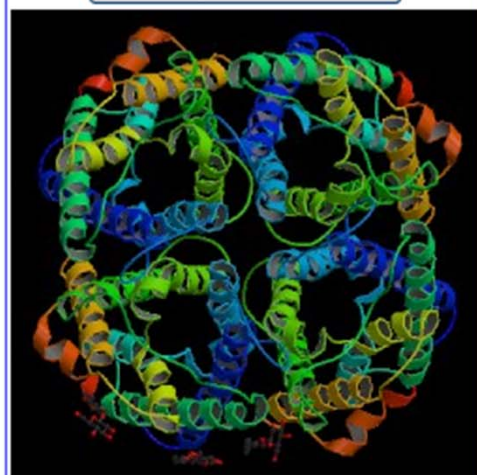
Experimental Method Type x-ray diffraction Data N/A

Parameters	Resolution[Å]	R-Value	R-Free	Space Group
	2.20	0.266 (work)	0.308	I 4 2 2

Unit Cell	Length [Å]	a	93.33	b	93.33	c	180.49
	Angles [°]	alpha	90.00	beta	90.00	gamma	90.00

Images and Visualization

Biological Molecule



Display Options

- KING
- Jmol
- WebMol
- Protein Workshop
- QuickPDB
- All Images

OSTALA BIOLOŠKA INFORMACIJA

OMIM, SPECIALIZIRANE PODATKOVNE ZBIRKE, SPLETNE STRANI
RAZISKOVALCEV...

VIRI INFORMACIJ V SLOVENIJI

Izum, NUK, Lokalne knjižnice (CTK, IJS, biološka...)

PROGRAMI ZA ISKANJE IN UREJANJE REFERENC

TEKSTOVNE PODATKOVNE ZBIRKE (NCBI)

- PubMed

- OMIM

On-line Mendelian Inheritance in Man



Katalog človeških genov in bolezni s katerimi so povezani (=fenotipski dodatek človeškemu genomu).

Tekstovna informacija in reference, povezave na podatkovne zbirke znotraj NCBI in ostale.

Skupaj 19967 opisov (april 2010)

Iskanje s ključnimi besedami



- BOOK-SHELF

“on-line” knjig z vsebino mol biologije, biokemije, celične biologije... npr. Molecular Biology of the Cell (Alberts), Molecular Cell Biology (Lodish), Genomes (Brown), Biochemistry (Stryer)

- PubMed Central

Arhiv revij (>100) s področja naravoslovja v polnem tekstu!!!



NCBI- tekstovne podatkovne zbirke

- PubMed
- OMIM
- BOOK-SHELF
- PubMed Central

PubMed

Na NCBI. Prosto dostopna podatkovna zbirka.

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>

Vsebuje citate iz MEDLINE in še dodatno informacijo (nekatero druge revije, ki jih v MEDLINE ni)

MEDLINE: vsebuje bibliografsko informacijo 4300 revij iz 70 držav; vsebuje preko 11 mio citatov od 1960 naprej.

(OLDMEDLINE- 1958-1965)

PubMed Journal Browser- informacija o revijah, iz katerih so članki v PubMed;

JournalLinkOutProvider- povezave na spletne strani založnikov, ki imajo revije na

spletu; MeSH browser (Medical Subject Heading)- pojmi, ki jih PubMed uporablja za indeksiranje člankov

BOOK-SHELF



>50 “on-line” knjig z vsebino mol biologije, biokemije, celične biologije...

Iskanje gesel

Npr. Molecular Biology of the Cell (Alberts), Molecular Cell Biology (Lodish), Genomes (Brown), Biochemistry (Stryer)



PUB-MED CENTRAL

<http://www.pubmedcentral.nih.gov/>

Arhiv revij (>100) s področja naravoslovja v polnem tekstu!!!

Npr. EMBO J

Infection and Immunity

Journal of Bacteriology

The Plant Cell

Plant Physiology....

OMIM

On-line Mendelian Inheritance in Man

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>

Katalog človeških genov in bolezni s katerimi so povezani.

(=fenotipski dodatek človeškemu genomu)

Ureja dr. Victor A McKusick, John Hopkins University

Tekstovna informacija in reference, povezave na podatkovne zbirke znotraj NCBI in ostale.

Skupaj **22.236** opisov (februar 2014)

Iskanje s ključnimi besedami

Abstract

A network of disorders and disease genes linked by known disorder–gene associations offers a platform to explore in a single graph-theoretic framework all known phenotype and disease gene associations, indicating the common genetic origin of many diseases. Genes associated with similar disorders show both higher likelihood of physical interactions between their products and higher expression profiling similarity for their transcripts, supporting the existence of distinct disease-specific functional modules. We find that essential human genes are likely to encode hub proteins and are expressed widely in most tissues. This suggests that disease genes also would play a central role in the human interactome. In contrast, **we find that the vast majority of disease genes are nonessential and show no tendency to encode hub proteins, and their expression pattern indicates that they are localized in the functional periphery of the network.** A selection-based model explains the observed difference between essential and disease genes and **also suggests that diseases caused by somatic mutations should not be peripheral, a prediction we confirm for cancer genes.**

Diseasome

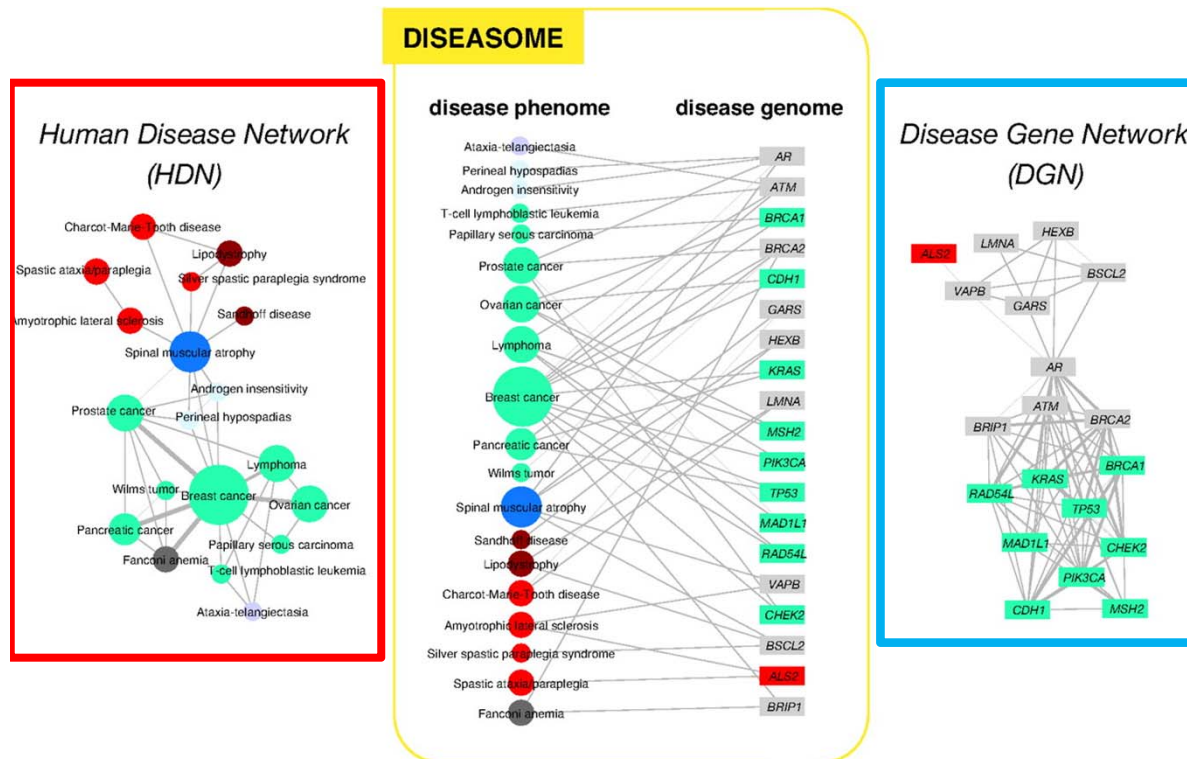


Fig. 1. Construction of the diseasome bipartite network. (Center) A small subset of OMIM-based disorder–disease gene associations (18), where circles and rectangles correspond to disorders and disease genes, respectively. A link is placed between a disorder and a disease gene if mutations in that gene lead to the specific disorder. The size of a circle is proportional to the number of genes participating in the corresponding disorder, and the color corresponds to the disorder class to which the disease belongs. (Left) The HDN projection of the diseasome bipartite graph, in which two disorders are connected if there is a gene that is implicated in both. The width of a link is proportional to the number of genes that are implicated in both diseases. For example, three genes are implicated in both breast cancer and prostate cancer, resulting in a link of weight three between them. (Right) The DGN projection where two genes are connected if they are involved in the same disorder. The width of a link is proportional to the number of diseases with which the two genes are commonly associated.

KAKO POISKATI INFORMACIJO?

Kaj potrebuješ (in a to res potrebuješ???)

INTERNET!!!!!!!!!!!! vs Lokalne knjižnice

Splošni iskalniki (Google, Wikipedia) vs specializirani strežniki (Entrez)

- Pregledovanje kazal (TOC "Table of Contents"- online dostop po mailu).

Večina revij ima na razpolago ta servis.

Staro klasično prelistavanje revij!

- Iskanje člankov (specializirane podatkovne zbirke):

prost dostop- npr. PubMed

licenčne- npr. Science Direct (prost za uporabnike UL)

odprt dostop- npr. PLOS Biology <http://www.plosbiology.org>

Kazala vs polna besedila (pdf format)

- Ostala specializirana informacija, tekstovne podatkovne zbirke (npr. OMIM)

VIRI INFORMACIJ V SLOVENIJI

IZUM- Inštitut informacijskih znanosti

COBISS

Kooperativni online bibliografski sistem in servisi

Co-operative Online Bibliographic System & Services

Podatkovne zbirke, polni članki tujih revij, Podatkovna zbirka citiranosti avtorjev

NUK <http://www.nuk.uni-lj.si>

dLib.si (Digitalna knjižnica Slovenije)

LOKALNE KNJIŽNICE

(CTK, IJS, na posameznih Oddelkih Biotehniške fakultete...)

Science Direct <http://www.sciencedirect.com>

Online dostop do polnih besedil člankov.

World's best Internet resource for Scientific, Technical and Medical Information. Using ScienceDirect you can access over 1,200 peer reviewed academic journals, search over 30 million abstracts from scientific articles and link out to articles from over 80 other publishers

COBISS/OPAC - Windows Internet Explorer

http://www.cobiss.si/scripts/cobiss?ukaz=getid&lani=si

GenBank Home COBISS/OPAC

File Edit View Favorites Tools Help

Google IŠČI Dodajno Prijava Convert Select

I-T... Roc... tag... BBC... Imp... Tabs ARR... Mul... Ori...

COBISS.SI Virtualna knjižnica Slovenije IZUM

Pomoč V ŽIVO Pomoč Nastavitve slovenski

Baze podatkov Iskanje Info Moja knjižnica

Vzajemna bibliografsko-kataložna baza podatkov



zapisov: 4.386.748

Podatke prispeva več kot 430 slovenskih knjižnic (140.000 povezav do e-virov, 2,2 milijona knjig, 1,5 milijona člankov, 100.000 revij, 100.000 CD-jev/DVD-jev ...)

Lokalne baze podatkov (katalogi) knjižnic

- Nacionalna knjižnica
- Univerzitetne in visokošolske knjižnice
- Specialne knjižnice
- Splošne knjižnice
- Šolske knjižnice
- Zamejske slovenske knjižnice

Knjižnica: POIŠČI Vse knjižnice: 432
(vpišite kraj ali akronim ali ime ali siglo)

Druge baze podatkov v sistemu COBISS.SI

- [COLIB.SI](#) - podatki o slovenskih knjižnicah
- [CONOR.SI](#) - normativna datoteka osebnih in korporativnih imen
- [SGC](#) - splošni geslovnik COBISS.SI
- [CORES.SI](#) - serijske publikacije: uredniki idr.
- [ELINKS.SI](#) - dostop do e-publikacij z zapisi v COBIB.SI

Drugi informacijski viri

- [Faktor vpliva revij iz baze podatkov Journal Citation Reports \(JCR\)](#)
- [Faktor vpliva revij iz baze podatkov Source Normalized Impact per Paper \(SNIP\)](#)
- [Kazala tujih znanstvenih in strokovnih revij SwetScan](#)
- [Normativna datoteka imen Kongresne knjižnice - LC/NAF](#)
- [Baze podatkov OCLC - servis FirstSearch](#)
- [Izbrani katalogi drugih tujih knjižnic](#)
- [Druge specializirane baze podatkov](#)

↑ NA VRH Baze podatkov Iskanje Informacije Moja knjižnica Izhod

Mobilna različica Lahki COBISS/OPAC POGOJI UPORABE Piškotki COBISS/OPAC, V6.1 - Novosti Predlogi? Pošijite jih na: cobissuser@izum.si © IZUM, 1997-2014

IZUM - Institut informacijskih znanosti - Windows Internet Explorer

http://www.izum.si/ GenBank Home IZUM - Institut informacijskih...

File Edit View Favorites Tools Help

Google Išč Dodato Prijava Convert Select

I-T... Roc... tag... BBC... Imp... Tabs ARR... Mul... Ori...

Page Safety Tools

IZUM

INSTITUT INFORMACIJSKIH ZNANOSTI, MARIBOR, SLOVENIJA

Servisi in storitve COBISS.SI SICRIS COBISS.Net O IZUM-u slovenski



IZUM

Regionalni center za knjižnične informacijske sisteme in informacijske sisteme o raziskovalni dejavnosti

COBISS.SI

Kooperativni online bibliografski sistem in servisi

Virtualna knjižnica Slovenije (COBISS/OPAC)
Na enem mestu dostop do informacij v več kot 400 slovenskih knjižnicah

Tuje baze podatkov in servisi

- Web of Science
- ProQuest
- EIFL Direct - EBSCOhost

SICRIS

Informacijski sistem o raziskovalni dejavnosti v Sloveniji

SICRIS
Podatki za več kot 14.000 raziskovalcev in za več kot 900 raziskovalnih organizacij

REFERENČNI SERVIS VPRAŠAJ KNJIŽNICARJA

Postavite vprašanje knjižničarju. Odgovor boste dobili takoj v [klepetu](#) ali v 24 urah po [e-pošti](#).

metaiskalnik

HKRATI IŠČITE po več elektronskih virih (baze podatkov s polnimi besedili, knjižnični katalogi, časopisi itd.).

o IZUM-u

- Status, dejavnost, organizacija...
- Prosta delovna mesta
- Katalog inf. javnega značaja
- Seje, sestanki, konference...
- Logotipi...
- Knjižnica IZUM

IZUM - Institut informacijskih znanosti
Prešernova ulica 17, 2000 Maribor, Slovenija
telefon: 02 2520-331 | e-pošta: izum@izum.si

© 1997-2014

Varstvo osebnih podatkov
Mapa strani | Kontakt | Piškotki

100%

IZUM- Inštitut informacijskih znanosti

<http://www.izum.si/>

COBISS

Kooperativni online bibliografski sistem in servisi

Co-operative Online Bibliographic System & Services

<http://cobiss.izum.si/>

COBISS povezuje in podpira funkcije in dejavnosti knjižnic, informacijskih centrov in informacijskih servisov v Sloveniji, ki uporabljajo enotno metodologijo distribuirane obdelave podatkov, programsko opremo COBISS ter druge storitve in proizvode Instituta informacijskih znanosti (IZUM), ki ga je ustanovila Vlada Republike Slovenije kot javni neprofitni zavod s funkcijo bibliografskega in informacijskega servisa.

COBIB- vzajemna bibliografsko-kataložna baza podatkov. Vsebuje več kot 2 mio zapisov o knjižnem in neknjižnem gradivu knjižnic v Sloveniji. Vzajemni katalog (preiskuješ kje se kakšna informacija nahaja)

- COBISS/OPAC (Online Public Access Catalogue)

Online dostop do knjižnic in podatkovnih zbirk (tudi tujih)

- Bibliografije raziskovalcev

- Baza podatkov Journal Citation Reports

- Preko COBISSa dostop do nekaterih podatkovnih zbirk v tujini:

OCLC- vodilni bibliografski ponudnik v ZDA, polna besedila člankov revij, naročene v knjižnicah Univerze v Ljubljani (npr. Bioinformatics), ProQuest, ABI/INFORM Global, Social Science Plus (polna besedila člankov!)

- WOS (Web of Science)**

Servis Web of Science omogoča dostop do treh baz podatkov z indeksi citiranosti: **Science Citation Index Expanded®**, **Social Sciences Citation Index®** in **Arts & Humanities Citation Index®**. Vključeni so podatki za obdobje od leta 1970 dalje, baze podatkov pa se dopolnjujejo tedensko. Gradi in vzdržuje jih **Institute for Scientific Information®**, **ISI®**, **Philadelphia, Pennsylvania, USA**, ki je tudi lastnik vseh avtorskih pravic.

- SICRIS

Podatki o raziskovalcij in raziskovalnih inštitucijah v Sloveniji

<http://sicris.izum.si/>

- Povezave do knjižnic in podatkovnih zbirk po svetu

PROGRAMI ZA UREJANJE IN OBDELAVO REFERENC

EndNote

<http://www.isiresearchsoft.com/en/enhome.htm>

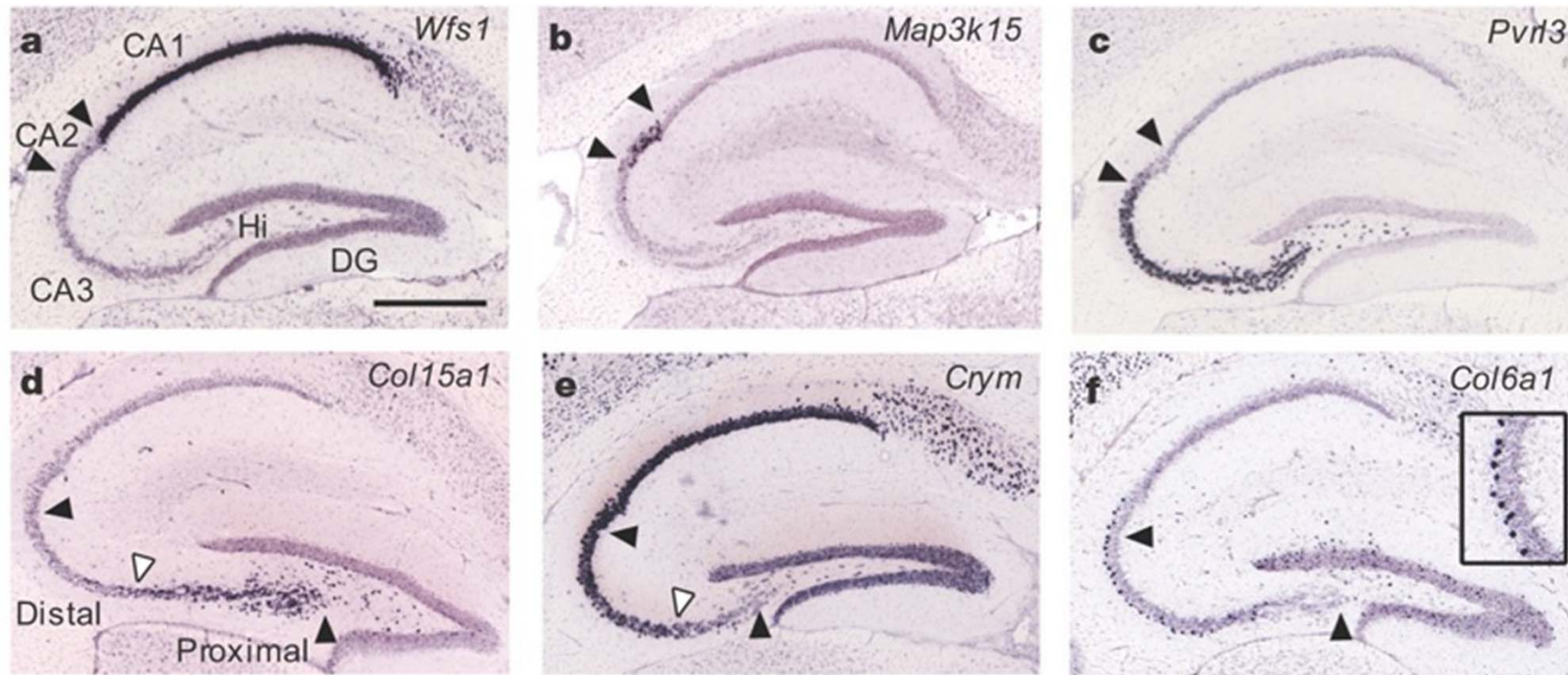
Reference Manager

<http://www.refman.com/>

- **Lastne podatkovne zbirke.**
- **Urejanje člankov in drugih bibliografskih podatkov.**
- **Iskanje (po interni bazi ali po internetu)**
- **Izmenjava z drugimi raziskovalci**
- **Priprava bibliografij v ustreznem formatu- zelo pomembno!!!!**
- **Navezava na programe in sisteme za pregledovanje povzetkov revij (npr. Reference Update)**

Genome-wide atlas of gene expression in the adult mouse brain

Lein ES et al. (2006) Nature 445(7124):168-176.



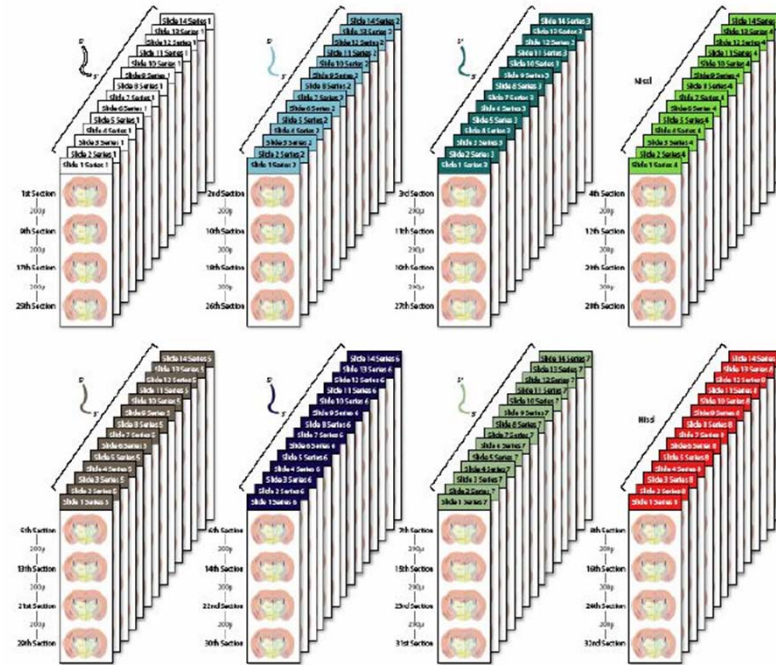
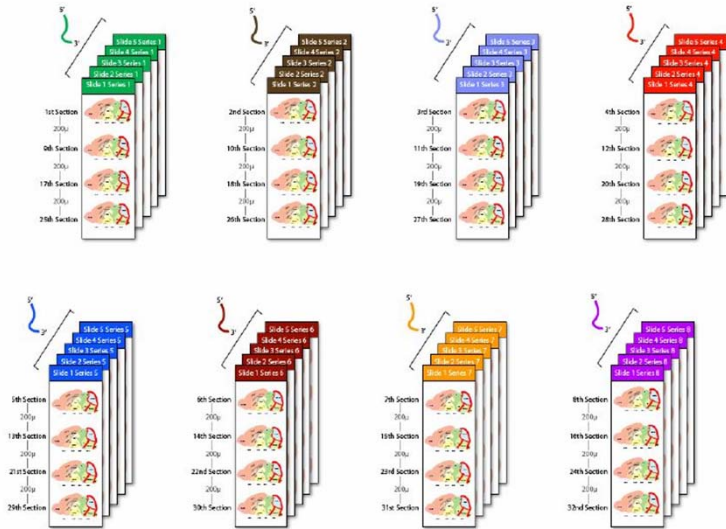
Serijske stekelc



Sagittal Series
8 series x 5 slides x 4 sections/slide
(160 Sagittal Sections - 25µm/Section)



Coronal Series
8 series x 14 slides x 4 sections/slide
(448 Coronal Sections - 25µm)



Postopek

- Fiksirane, acetilirane in dehidrirane rezine tretirajo z različnimi reagenti, ki preprečujejo aktivnost endogenih peroksidaz in povečajo prepustnost tkiva za lažjo penetracijo in hibridizacijo sond
- *In situ* hibridizacija z digoksininom
- Stekelca pokrita s krovnimi stekelci in očiščena za lažjo nadaljnjo obdelavo
- Dve izmed serij (4 in 8) se barvata po Nisslu (obarva se rRNA na grobem ER)
- Sledi zajem in obdelava slike
- 10 stalno delujočih postaj za zajem slike (Image Capturing Systems, ICS)
 - Sestavljen iz mikroskopa, kamere, nalagalca stekelc, čitalca barkode in delovne postaje (računalnik z 2GB rama in 300 ali 400 GB trdim diskom)
 - Naloži stekelce
 - Poskenira barkodo in ustvari lokalno mapo za zbiranje slik
 - Z nekaj praznimi slikami preveri, če so barve pravilno kalibrirane
 - Locira posamezno rezino in pri različnih povečavah slika rezine, delček za delčkom
 - Slike naloži na ustreno lokacijo na mreži
 - Ko poslika vse štiri rezine na stekelcu, stekelce odloži v kaseto
 - Za eno stekelce porabi 15-20 minut
- **Ljudi potrebujejo le za dodajanje novih stekelc**

Rezultat

- 21500 genov
- Žrtvovanih več kot 6000 mišk
- 1000 stekelc oz. 4000 rezin možganov vsak dan
- 85 milijonov slik
- Povprečen dotok podatkov 1 TB (terabyte) /dan, skupno 600 TB
- Javno dostopno: <http://www.brain-map.org>