Slide 1

BIOINFORMATIKA- UVOD

Podatkovne zbirke v biokemiji in molekularni biologiji

31.3.-1.4.2005

1. **Internet, strežniki z biokemijsko in molekularno-biološko vsebino**
2. **Nacionalni center za biotehnološko informacijo (NCBI), Expert Protein Analysis System (ExPASy)**
3. **Proteinske podatkovne zbirke, UniProt, nukleotidne podatkovne zbirke, GenBank**
4. **Bibliografske podatkovne zbirke, Pubmed, OMIM, Bookshelf**

VIRI

http://rcul.uni-lj.si/~bfbiochem/bioinfo/bioinfo.htm; Spletna stran o bioinformatiki
Literatura- Popoln seznam na zgornjih spletnih straneh
Attwood TK, Parry-Smith DJ (1999) Introduction to Bioinformatics. Prentice Hall, Harlow, United Kingdom

Slide 2

BIOINFORMATIKA

Vmesnik med biologijo in računalništvom.

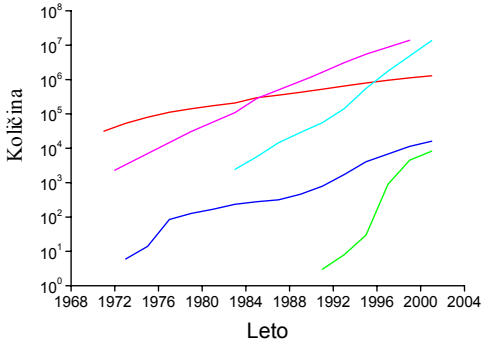
Matematične, statistične in računalniške metode za reševanje bioloških problemov z uporabo DNA in proteinskih zaporedij in povezane informacije.

Organiziranje, analiza in distribucija biološke informacije pri opisovanju in reševanju bioloških problemov.

Zbiranje, organiziranje, shranjevanje in iskanje biološke informacije v podatkovnih zbirkah.

NI ISKANJE ČLANKOV IN BRSKANJE ZA REVIJAMI!!!

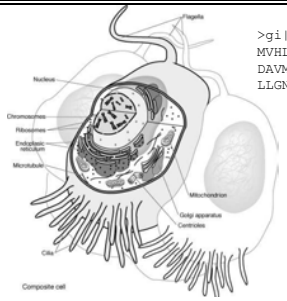
Slide 3



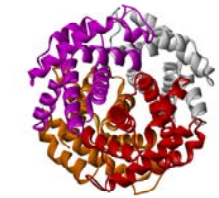
Ninety-nine percent of people don't have an inkling about how fast this revolution is coming.

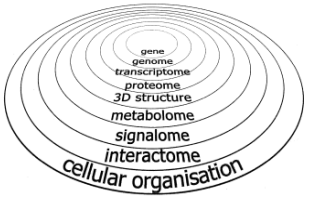
Steve Fodor, president of Affimetrix

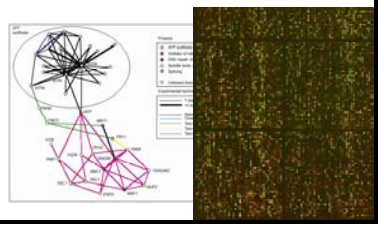
Slide 4



```
>gi|122615|sp|P02023|HBB_HUMAN HEMOGLOBIN BETA CHAIN
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTP
DAVMGNPVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFR
LLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH
```







Vihinen M (2001) *Biomol. Eng.* 18:241-248

Slide 5

STREŽNIKI Z BIOKEMIJSKO IN MOL.-BIOL. INFORMACIJO

SRS	Sequence Retrieval System
NCBI	The National Center for Biotechnology Information
ExPASy	Expert Protein Analysis System
KEGG	Kyoto Encyclopedia of Genes and Genomes

Slide 6

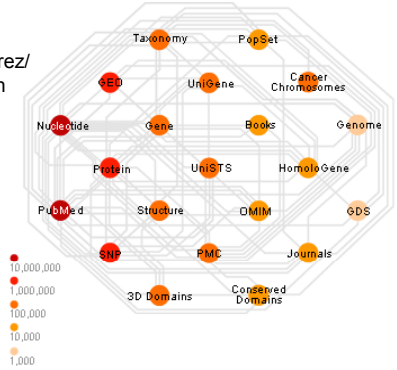
NCBI- The National Center for Biotechnology Information

1988 Kot del National Library of Medicine

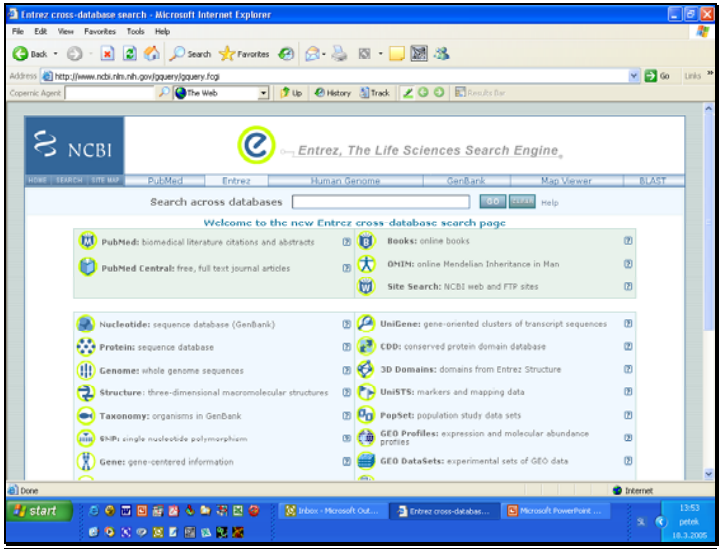
Namen: *razvoj novih informacijskih tehnologij, ki pomagajo razumeti molekularne in genetske procese*

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)



Slide 7



Slide 8

ExPASy Expert Protein Analysis System
http://www.expasy.org/
Swiss Institute of bioinformatics (SIB)

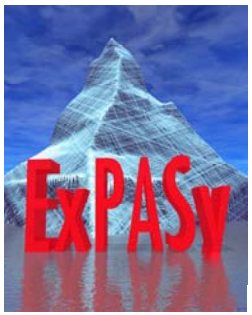
The ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) proteomics server from the Swiss Institute of Bioinformatics (SIB) is dedicated to molecular biology with an emphasis on data relevant to

proteins.

Appel R.D., Bairoch A., Hochstrasser D.F.
A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server.
Trends Biochem. Sci. 19:258-260(1994).

Podatkovne zbirke: **UniProt**
"SwissProt"
TrEMBL
PROSITE
SWISS 2D-PAGE

Programi: Orodja za proteomiko
Melanie (analiza 2D gelov)
Biochemical Pathways



Slide 9

PODATKOVNE ZBIRKE

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

Nucleic Acids Research (2003) 31, number 1

Primarne nukleotidne (GenBank, EMBL, DDBJ), proteinske (UniProt), strukturne (Protein Data Bank)

Bibliografske PubMed, OMIM

```
ATGGTGCACCTGAC
TCCTGAGGAGAAG...
```
GenBank, EMBL, DDBJ

```
MVHLTPEEK...
```
UniProt


Protein Data Bank

Slide 10

PROTEINSKE PODATKOVNE ZBIRKE

UniProt
the universal protein resource

UniProt 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.
Upravljata jo SIB (Swiss Institute of Bioinformatics) in EBI/EMBL (The European Bioinformatics Institute, UK). Vsebuje proteinska zaporedja
UniProt Različica 4.3 vsebuje: Swiss-Prot različico 46.3 of 15-Mar-2005: 176.469 zapisov iz 8.998 vrst
Zelo kvalitetna **določitev** biološke funkcije (**ANOTACIJA, “annotation”**) → zelo uporabna podatkovna zbirka
1996 TrEMBL (SP-TrEMBL, REM-TrEMBL). Prevedena nukleotidna zaporedja vseh kodirajočih zaporedij (*coding sequences, CDS*) iz EMBL. TrEMBL različica 29.3 15-Mar-2005 vsebuje 1.640.768 zapisov

Slide 11

SwissProt

Nucleic Acids Research (2003) 31, 365-370

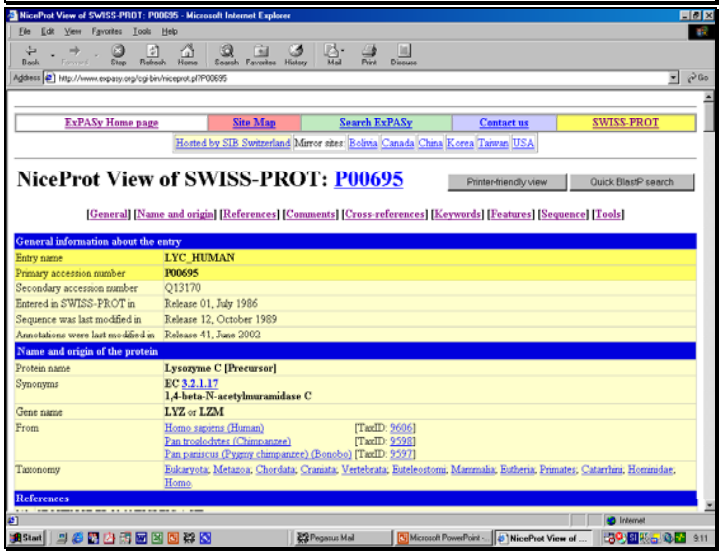
Core data aminokislinsko zaporedje
ime proteina
taksonomski podatki
citati

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.

Slide 12



Slide 13

STRUKTURA SWISS-PROT ZAPISA

SWISS-PROT: P00695

*NicePro*t *- a user-friendly view of this SWISS-PROT entry*

```
ID   LYC_HUMAN      STANDARD;      PRT;   148 AA.
AC   P00695; Q13170;
DT   21-JUL-1986 (Rel. 01, Created)
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).
GN   LYZ OR LZM.
OS   Homo sapiens (Human),
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.
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   Saccharomyces cerevisiae.";
RL   Gene 66:223-234(1988).
```

PROTEIN_IZVOR — se lahko spreminjna
SwissProt koda — se ne spreminja
DT — datum vnosa
DE — opis proteina
GN — ime gena
OS — organizem
RN — bibliografski podatki

Slide 14

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

CC — komentarji
DR — zapis v drugih zbirkah

Slide 15

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

FT — *feature table*
lastnosti proteina
SQ — aminokislinsko zaporedje

Slide 16

NUKLEOTIDNE PODATKOVNE ZBIRKE

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). 37.893.844.733 baz v 32.549.400 zaporedjih.
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)

Slide 17

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
```

Slide 18

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

datum, ko je zapis postal javen

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.

Slide 19

```
DEFINITION  Homo sapiens lysozyme (renal amyloidosis) (LYZ), mRNA.
```
Biološki opis zaporedja (izpisan v FASTA formatu)

```
                                                     NT_123456 constructed genomic contigs
ACCESSION   NM_000239                                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).

```
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...)

Slide 20

```
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="MKALIVLGLVLLSVTVQGKVFERCELARTLKRLGMDGYRGISLA
                     NWMCLAKWESGYNTRATNYNAGDRSTDYGIFQINSRYWCNDGKTPGAVNACHLSCSAL
                     LQDNIADAVACAKRVVRDPQGIRAWVAWRNRCQNRDVRQYVQGCGV"
     mat_peptide     26..469
                     /product="lysozyme"
     misc_feature    80..463
                     /note="lys; Region: C-type lysozyme/alpha-lactalbumin
                     family. Alpha-lactalbumin is the regulatory subunit of
                     lactose synthase"
     misc_feature    80..466
                     /note="LYZ1; Region: Alpha-lactalbumin / lysozyme C"
     variation       288
                     /allele="C"
                     /allele="A"
                     /db_xref="dbSNP:2000977"
     repeat_region   740..1054
                     /note="Alu repeat"
```

Lastnosti zaporedja "feature table".

Svojo lastno strukturo in povezave z drugimi podatkovnimi zbirkami

Slide 21

Zaporedje

```
BASE COUNT      435 a    306 c    308 g    438 t
ORIGIN
        1 ctagcactct gacctagcag tcaacatgaa ggctctcatt gttctggggc ttgtcctcct
       61 ttctgttacg gtccagggca aggtctttga aaggtgtgag ttggccagaa ctctgaaaag
      121 attgggaatg gatggctaca ggggaatcag cctagcaaac tggatgtgtt tggccaaatg
      181 ggagagtggt tacaacacac gagctacaaa ctacaatgct ggagacagaa gcactgatta
      241 tgggatattt cagatcaata gccgctactg gtgtaatgat ggcaaaaccc caggagcagt
      301 taatgcctgt catttatcct gcagtgcttt gctgcaagat aacatcgctg atgctgtagc
      361 ttgtgcaaag agggttgtcc gtgatccaca aggcattaga gcatgggtgg catggagaaa
      421 tcgttgtcaa aacagagatg tccgtcagta tgttcaaggt tgtggagtgt aactccagaa
      481 ttttccttct tcagctcatt ttgtctctct cacattaagg gagtaggaat taagtgaaag
      541 gtcacactac cattatttcc ccttcaaaca aataatattt ttacagaagc aggagcaaaa
      601 tatggccttt cttctaagag atataatgtt cactaatgtg gttattttac attaagccta
      661 caacattttt cagtttgcaa atagaactaa tactggtgaa aatttaccta aaaccttggt
      721 tatcaaatac atctccagta cattccgttc ttttttttt ttggagacagt ctcgctctgt
      781 cgcccaggct ggagtgcagt ggcgcaatct cggctcactg caacctccac ctcccgggtt
      841 cacgccattc tcctgcctca gcctcccgag tagctgggat tacgggcgcc cgccaccacg
      901 cccggctaat tttttgtatt tttagtagag acaggggttc accgtgttag ccaggatggt
      961 ctcgatctcc tgaccttgtg atccacccac ctcggcctcc caaagtgctg ggattacagg
     1021 cgtgagccac tgcgcccggc cacattcagt tcttatcaaa gaaataaccc agacttaatc
     1081 ttgaatgata cgattatgcc caatattaag taaaaaatat aagaaaaggt tatcttaaat
     1141 agatctttagg caaaatacca gctgatgaag gcatctgatg ccttcatctg ttcagtcatc
     1201 tccaaaaaca gtaaaaataa ccacttttttg ttgggcaata tgaaatttt aaaggagtag
     1261 aataccaaat gatagaaaca gactgcctga attgagaatt ttgatttctt aaagtgtgtt
     1321 tctttctaaa ttgctgttcc ttaatttgat taatttaatt catgtattat gattaaatct
     1381 gaggcagatg agcttacaag tattgaaata attactaatt aatcacaaat gtgaagttat
     1441 gcatgatgta aaaaatacaa acattctaat taaaggcttt gcaacac
//
```

Slide 22

OBLIKA ZAPISA ZAPOREDIJ

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
CTAGCACTCTCGACCTAGCAGTCAACATGAAGGCTCTCATTGTTCTGGGGCTTGTCCTCCTTTCTGTTACG
GTCCAGGGCAAGGTCTTTGAAAGGTGTGAGTTGGCCAGAACTCTGAAAAGATTGGGAATGGATGGCTACA
GGGGAATCAGCCTAGCAAACTGGATGTGTTTGGCCAAATGGGAGAGTGGTTACAACACACGAGCTACAAA
CTACAATGCTGGAGACAGAAGCACTGATTATGGGATATTTCAGATCAATAGCCGCTACTGGTGTAATGAT
GGCAAAACCCCAGGAGCAGTTAATGCCTGTCATTTATCCTGCAGTGCTTTGCTGCAAGATAACATCGCTG
```

Slide 23

BIBLIOGRAFSKE PODATKOVNE ZBIRKE

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 (nekatere 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*

Slide 24

OMIM

OMIM Online Mendelian Inheritance in Man — Johns Hopkins University

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 14925 opisov (18.3.2005)

Iskanje s ključnimi besedami

	Autosomal	X-Linked	Y-Linked	Mitochondrial	Total
* Gene with known sequence	9561	430	48	37	10076
+ Gene with known sequence and phenotype	364	37	0	0	401
# Phenotype description, molecular basis known	1527	138	2	27	1694
% Mendelian phenotype or locus, molecular basis unknown	1328	134	4	0	1466
Other, mainly phenotypes with suspected mendelian basis	2145	153	2	0	2300
Total	14925	892	56	64	**15937**

Slide 25

BOOK-SHELF

37 “on-line” knjig z vsebino mol biologije, biokemije, celične biologije...
Iskanje gesel
Npr. Molecular Biology of the Cell (Alberts), Biochemistry (Stryer), Molecular Cell Biology (Lodish)...

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....
PNAS