Bioinformatika és genomanalízis az orvostudományban

Szekvenciaelemzés

Cserző Miklós

2020

https://semmelweis.zoom.us/j/96102872458?pwd=Rk1PL2tqS21sdlUwc3B4eDFCZkNKQT09

A mai előadás

- Szekvencia analízis statisztikus szempontból
- Annotálás homológia alapján
- Az annotálás szempontjai
- Annotálás térszerkezet szerint

Megfontolandó szempontok

DNS

- 4 betűs ABC
- Hosszú szekvenciák
- Szekezet nem annyira fontos
  - RNS kivétel
- Elsődleges adatok

fehérje

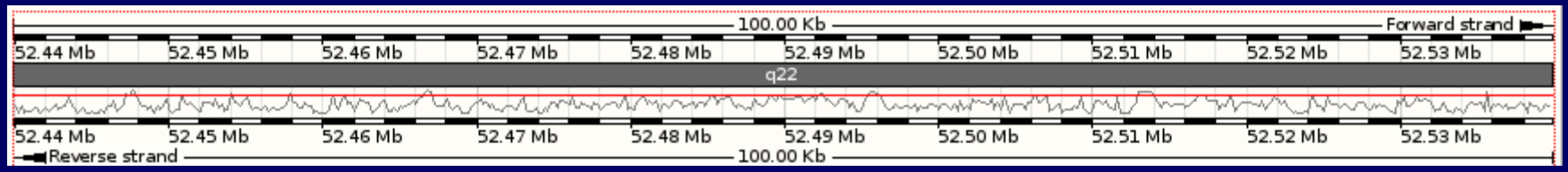
- 20 betűs ABC
- Tömör információ
- A szerkezet nagyon fontos
- Származtatott adatok

Informatikai vs. biológiai elemzés

- A szekvencia betűsorozat
- Statisztikus módszerekkel vizsgálható
- Nem használunk biológiai információt
- Az eredményt utólag kell értelmezni
- Egyszrű, viszonylag hatékony, de pontatlan

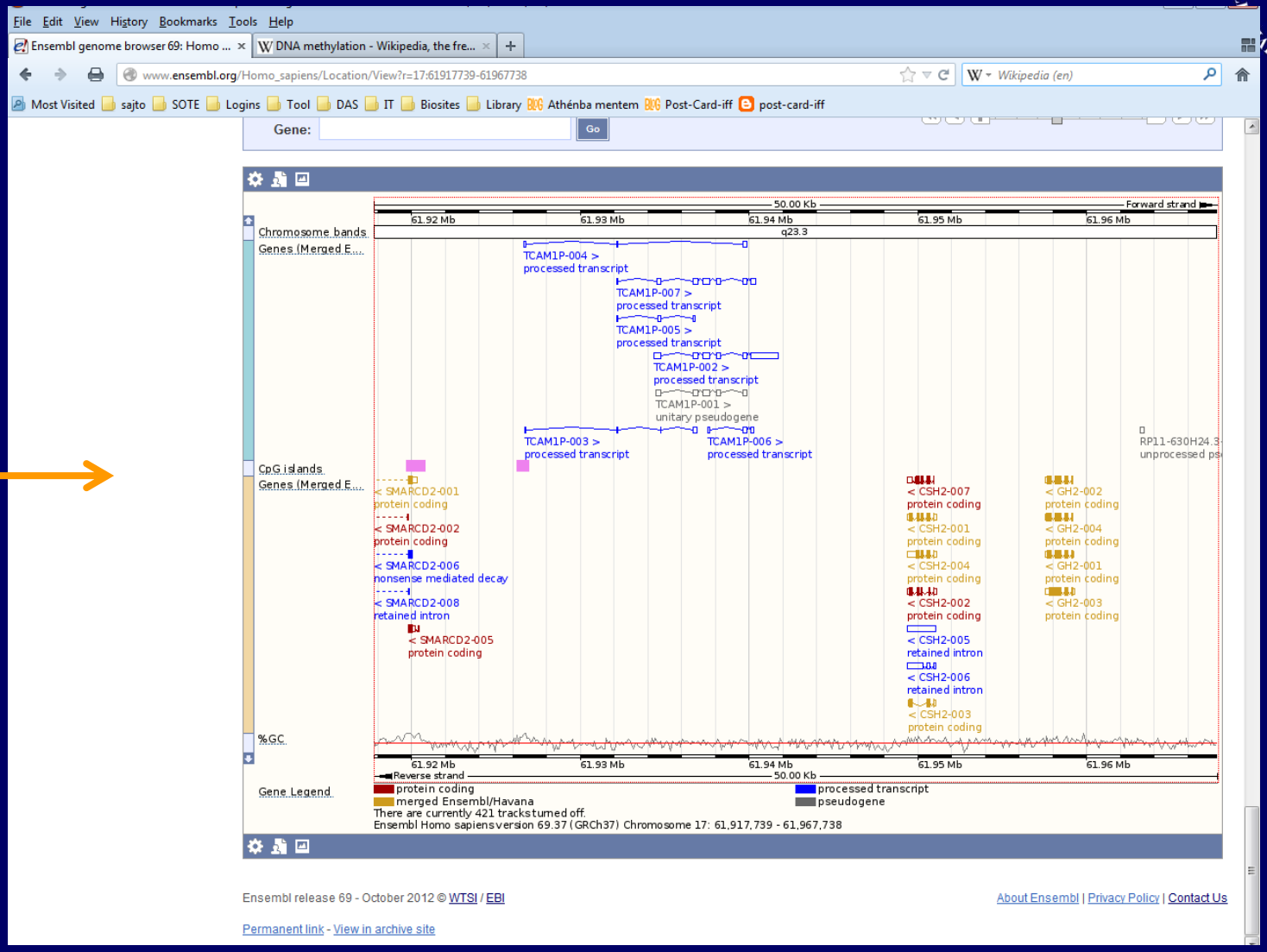
Az összetétel elemzése

- C/G vs. A/T tartalom
- A gének promóter szakasza gyakran C/G-ben gazdag
- Csúszó-ablakban megjelenített összetétel



CpG szigetek

- A DNS képes metilálódni a “CG” dinukleotidok “C”-jén
- A metilált állapot öröklődik osztódás után
- A metilált DNS –szakaszon a gének kifejeződése tartósan megváltozik
- Epigenetikus génszabályozás


File Edit View History Bookmarks Tools Help
Ensembl genome browser 69: Homo ...
DNA methylation - Wikipedia, the fre...
www.ensembl.org/Homo_sapiens/Location/View?r=17:61917739-61967738
Most Visited sajto SOTE Logins Tool DAS IT Biosites Library Athénba mentem Post-Card-iff post-card-iff
Gene:
Go
50.00 Kb
Forward strand
61.92 Mb
61.93 Mb
61.94 Mb
61.95 Mb
61.96 Mb
Chromosome bands
q23.3
Genes (Merged E...
TCAM1P-004 > processed transcript
TCAM1P-007 > processed transcript
TCAM1P-005 > processed transcript
TCAM1P-002 > processed transcript
TCAM1P-001 > unitary pseudogene
TCAM1P-003 > processed transcript
TCAM1P-006 > processed transcript
RP11-630H24.3 unprocessed ps
CpG islands
Genes (Merged E...
< SMARCD2-001 protein coding
< SMARCD2-002 protein coding
< SMARCD2-006 nonsense mediated decay
< SMARCD2-008 retained intron
< SMARCD2-005 protein coding
< CSH2-007 protein coding
< CSH2-001 protein coding
< CSH2-004 protein coding
< CSH2-002 protein coding
< CSH2-005 retained intron
< CSH2-006 retained intron
< CSH2-003 protein coding
< GH2-002 protein coding
< GH2-004 protein coding
< GH2-001 protein coding
< GH2-003 protein coding
%GC
Reverse strand
Gene Legend
protein coding
merged Ensembl/Havana
processed transcript
pseudogene
There are currently 421 tracks turned off.
Ensembl Homo sapiens version 69.37 (GRCh37) Chromosome 17: 61,917,739 - 61,967,738
Ensembl release 69 - October 2012 © WTSI / EBI
About Ensembl | Privacy Policy | Contact Us
Permanent link - View in archive site

CpG – bioinformatikai megközelítés

- A szekvenciát csúszóablakban vizsgáljuk
- A C/G tartalom legyen magas
- A tartalom alapján megbecsüljük a “CG” dinukleotid tartalmat
- Ezt összevetjük a talált “CG” tartalommal
- Ha ez magasabb a vártnál – találtunk egy CpG szigetet
- Az ablakméret 200 – 500 bázis

És mennyire jó ez?

- Vegyünk egy 350 bázis hosszú darabot
- Ebben legyen 35 “CG” (10%)
- A maradék szekvenciában egyenlően van a négy bázisból
- Rögzitjük az eredeti szekvencia “CG”-it
- A maradékot összekeverjük
- Visszatesszük a “CG”-et a helyükre
- Mindegyik szekvencia CpG-sziget lesz

Tanulság

- Összesen 10^{165} lehetséges varians van
- Ezek mind legalabb olyan jó CpG-szigetek, mint az eredeti
- Teljesen kizárt, hogy funkcióra nézve is egyenértékűek legyenek
- A modell mégsem tud különbséget tenni köztük
- Egyszerű modelltől nem várhatunk sokat

Szekvenciák komplexitása

- A szekvenciát rövid motívumokra bontjuk
- A motívumok hosszával 4 hatványai szerint nő a lehetőségek száma
- Egy szekvencia komplex, ha jól kihasználja a lehetőségeket
- Viszont az alacsony komplexitású szekvencia kevés motívumot használ nagy számban

Pl.

- TACAATGGCCTTCCCCATCCTACCTCTGCCCTGGCTTG
TAACTTGGGGAGACCTTCACTTTGGGGGCGTCGGCCCT
TTCCAAGTCAGGAGTGGAAATGGAAGGAGAGGCTGGGA
ATCCCCTCCCACAAACATGAAGTGGTCTCCTGGTACTG
TACGAACGAACGAACGTAGCCTTGGGCTTGGAGCTCAG
AGCCCCCACGTTTCCCGTTGCCTCTGTGGTTTTCTTTC
CCACCACTACCCCCACCCTGCACCTCCCCACCAAAGAA
TTCTCAACTGGAAAAGCCAGGAGGCGGTTCTGACAAAA
GGCAGGGGCTCCAGGGGAGACTCCGCCCGTCCCTGGGT
GGCTGGCTGTATCGCAGAGCTGGCTTTGCGATTGCGTG
TCCGCAATTGTGCCCATCAGAGTGTGAATGTATTGATA
TTTCTTTAAGGATGCTCTTTCGTTCTTCCAAGCCCGAG
GTACCTTAGGGGAGGGACTTAGAACTTATTGGCATTGC
ATCACTTTAGTTTTCAACCTGCTTGCATAAGAATTAAG
AGCGAATAAATATTAGTGTGGGGGGAGGGGAAGCTAAG
CAAAATATGAATTCCTCTCTCTCTCCCCACCTCCTTTG
AGATTTCTGAGCTGCCAATCTCCCAGCCAATTCTAGAC
TTTCTGAAACTCCATGCACGTATAACTGAAGCCAGAAA
TGGGTTTCCTTGCAAATATAGGTCAACATCCTTTTTAT
TGCCCTATTAAAATATTCAAGTCCTACCTTTAGGGCTA
GGTGCGTACAGCGGCTGATGGAGTGGCGCTGGTGGGGC
GCAAGTGCAGGGGGAGGGTACTGACGGCAGAGAGAGAG
GAGCTACCTCCGTGCCGCCCTGCTTCCCGACCCGATTC
CCAGGCTTGCTTGAGGCCGAGAAAGGCGAGGGGCAGGC
- ATTGCGTAGGATTGCGTAGGATTGCGTAGGATTGCGTA
GGATTGCGTAGGATTGCGTAGGATTGCGTAGGATTGCG
TAGGATTGCGTAGGATTGCGTAGGATTGCGTAGGATTG
CGTAGGATTGCGTAGGATTGCGTAGGATTGCGTAGGAT
TGCGTAGGATTGCGTAGGATTGCGTAGGATTGCGTAGG
ATTGCGTAGGATTGCGTAGGATTGCGTAGGATTGCGTA
GGATTGCGTAGGATTGCGTAGGATTGCGTAGGATTGCG
TAGGATTGCGTAGGATTGCGTAGGATTGCGTAGGATTG
CGTAGGATTGCGTAGGATTGCGTAGGATTGCGTAGGAT
TGCGTAGGATTGCGTAGGATTGCGTAGGATTGCGTAGG
ATTGCGTAGGATTGCGTAGGATTGCGTAGGATTGCGTA
GGATTGCGTAGGATTGCGTAGGATTGCGTAGGATTGCG
TAGGATTGCGTAGGATTGCGTAGGATTGCGTAGGATTG
CGTAGGATTGCGTAGGATTGCGTAGGATTGCGTAGGAT
TGCGTAGGATTGCGTAGGATTGCGTAGGATTGCGTAGG
ATTGCGTAGGATTGCGTAGGATTGCGTAGGATTGCGTA
GGATTGCGTAGGATTGCGTAGGATTGCGTAGGATTGCG
TAGGATTGCGTAGGATTGCGTAGGATTGCGTAGGATTG
CGTAGGATTGCGTAGGATTGCGTAGGATTGCGTAGGAT
TGCGTAGGATTGCGTAGGATTGCGTAGGATTGCGTAGG
ATTGCGTAGGATTGCGTAGGATTGCGTAGGATTGCGTA
GGATTGCGTAGGATTGCGTAGGATTGCGTAGGATTGCG
TAGGATTGCGTAGGATTGCGTAGGATTGCGTAGGATTG
CGTAGGATTGCGTAGGATTGCGTAGGATTGCGTAGGAT

Komplex szekvencia

- AAA 16; AAC 11; AAG 16; AAT 16; ACA 6; ACC 13; ACG 7; ACT 13; AGA 18; AGC 14; AGG 24; AGT 9; ATA 9; ATC 7; ATG 9; ATT 16; CAA 16; CAC 11; CAG 14; CAT 8; CCA 20; CCC 30; CCG 10; CCT 25; CGA 10; CGC 6; CGG 4; CGT 10; CTA 8; CTC 19; CTG 22; CTT 26; GAA 19; GAC 7; GAG 26; GAT 6; GCA 14; GCC 17; GCG 10; GCT 20; GGA 18; GGC 23; GGG 31; GGT 11; GTA 10; GTC 7; GTG 15; GTT 7; TAA 8; TAC 10; TAG 9; TAT 11; TCA 9; TCC 25; TCG 3; TCT 17; TGA 12; TGC 19; TGG 24; TGT 9; TTA 10; TTC 21; TTG 18; TTT 21;

Repetitív szekvencia

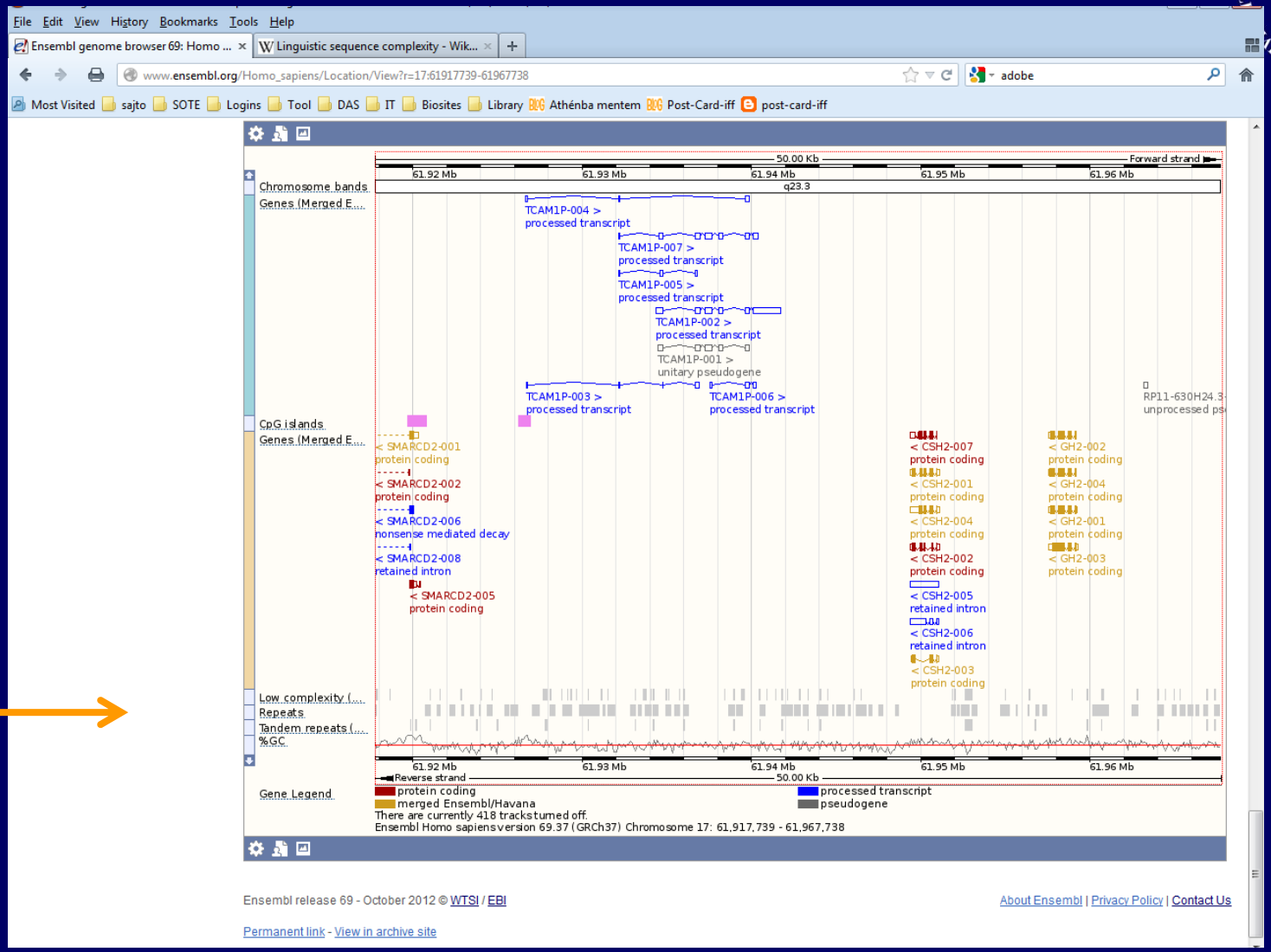
- AGG 76; ATT 76; CGT 76; GAT 75; GCG 76; GGA 75; GTA 76; TAG 76; TGC 76; TTG 76;

"Repetitív DNS"

- "tandem repeat" DNS – a kis komplexitású DNS alesete
- Ugyan az a motívum ismétlődik egymás után sokszor
- Jellemzés:
  - elemi motívum
  - ismétlésszám
  - konzerváltság

Repetitív DNS – a másik típus

- Ez komplex DNS-ből áll
- A genomban sok példányban fordul elő
- Nem egymás után helyezkednek el a példányok
- A példányok töredékesek
- Millió éves virusfertőzések maradványai
- Igazából nem szerencsés az elnevezése


File Edit View History Bookmarks Tools Help
Ensembl genome browser 69: Homo ...
Linguistic sequence complexity - Wik...
www.ensembl.org/Homo_sapiens/Location/View?r=17:61917739-61967738
Most Visited sajto SOTE Logins Tool DAS IT Biosites Library Athénba mentem Post-Card-iff post-card-iff
50.00 Kb
Forward strand
61.92 Mb
61.93 Mb
61.94 Mb
61.95 Mb
61.96 Mb
Chromosome bands
q23.3
Genes (Merged E...
TCAM1P-004 > processed transcript
TCAM1P-007 > processed transcript
TCAM1P-005 > processed transcript
TCAM1P-002 > processed transcript
TCAM1P-001 > unitary pseudogene
TCAM1P-003 > processed transcript
TCAM1P-006 > processed transcript
RP11-630H24.3 unprocessed ps
CpG islands
Genes (Merged E...
< SMARCD2-001 protein coding
< SMARCD2-002 protein coding
< SMARCD2-006 nonsense mediated decay
< SMARCD2-008 retained intron
< SMARCD2-005 protein coding
< CSH2-007 protein coding
< CSH2-001 protein coding
< CSH2-004 protein coding
< CSH2-002 protein coding
< CSH2-005 retained intron
< CSH2-006 retained intron
< CSH2-003 protein coding
< GH2-002 protein coding
< GH2-004 protein coding
< GH2-001 protein coding
< GH2-003 protein coding
Low complexity (...
Repeats
Tandem repeats (...
%GC
Reverse strand
Gene Legend
protein coding
merged Ensembl/Havana
processed transcript
pseudogene
There are currently 418 tracks turned off.
Ensembl Homo sapiens version 69.37 (GRCh37) Chromosome 17: 61,917,739 - 61,967,738
Ensembl release 69 - October 2012 © WTSI / EBI
About Ensembl | Privacy Policy | Contact Us
Permanent link - View in archive site

Fehérje kódoló szakaszok

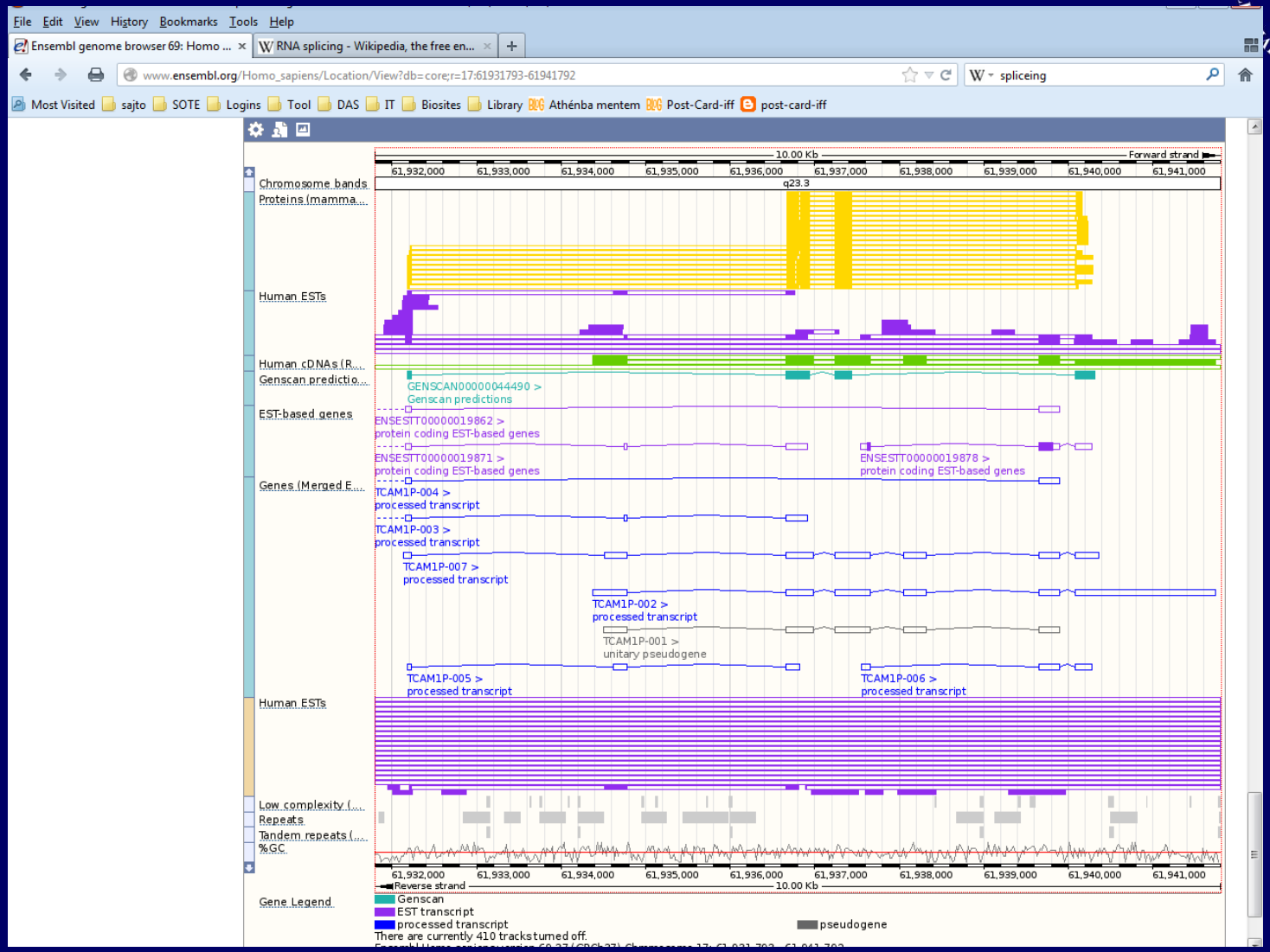
- A lehetséges 64 kódonból 3 stop-jel
- Véletlen esetben kb. minden 21.
- "Open Reading Frame" – ORF: hosszú, stop-jel nélküli szakasz
- 3 lehetséges olvasási keretben és mindkét szálon lehet
- Rövid szakaszokra bizonytalan

Prokarióták és eukarióták

- Prokariótákban:
  - A kódoló szakaszok egy darabban vannak
  - Kapcsolódó gének operonokba rendeződnek
  - A gének átfedhetnek
- Eukariótákban:
  - A kódoló szakasz nem folytonos (exon/intron)
  - Az mRNS a “splicing” soran alakul ki
  - A gének átfedése ritkább

Génpredikció – az ensembl példája

- A tisztán statisztikai alapú predikció gyenge
- Fel kell használni külső adatokat is
  - EST adatbázis
  - cDNS adatbázis
  - Humán és egyébb (gerinces) adatok
  - Fehérje adatbázis
  - Humán és egyébb (gerinces) adatok


File Edit View History Bookmarks Tools Help
Ensembl genome browser 69: Homo ...
RNA splicing - Wikipedia, the free en...
www.ensembl.org/Homo_sapiens/Location/View?db=core;r=17:61931793-61941792
spliceing
Most Visited sajto SOTE Logins Tool DAS IT Biosites Library Athénba mentem Post-Card-iff post-card-iff
10.00 Kb
Forward strand
61,932,000 61,933,000 61,934,000 61,935,000 61,936,000 61,937,000 61,938,000 61,939,000 61,940,000 61,941,000
Chromosome bands
q23.3
Proteins (mamma...
Human ESTs
Human cDNAs (R...
Genscan predictio...
GENSCAN00000044490 >
Genscan predictions
EST-based genes
ENSESTT00000019862 >
protein coding EST-based genes
ENSESTT00000019871 >
protein coding EST-based genes
ENSESTT00000019878 >
protein coding EST-based genes
Genes (Merged E...
TCAM1P-004 >
processed transcript
TCAM1P-003 >
processed transcript
TCAM1P-007 >
processed transcript
TCAM1P-002 >
processed transcript
TCAM1P-001 >
unitary pseudogene
TCAM1P-005 >
processed transcript
TCAM1P-006 >
processed transcript
Human ESTs
Low complexity (...
Repeats
Tandem repeats (...
%GC
Reverse strand
10.00 Kb
Gene Legend
Genscan
EST transcript
processed transcript
pseudogene
There are currently 410 tracks turned off.

Statisztikus megközelítés – fehérjék

- A szekvenciák komplexitása alapján a fehérjék is elemezhetők
- Transzmembrán fehérjék:
  - Sok hidrofób aminósav egymás után
  - Torzított összetétel
  - Alacsony szekvencia-komplexitás
- Szerkezet nélküli fehérjék:
  - Hidrofil aminósavak csoportjai
  - Torzított összetétel, stb….

Transzmembrán fehérjék

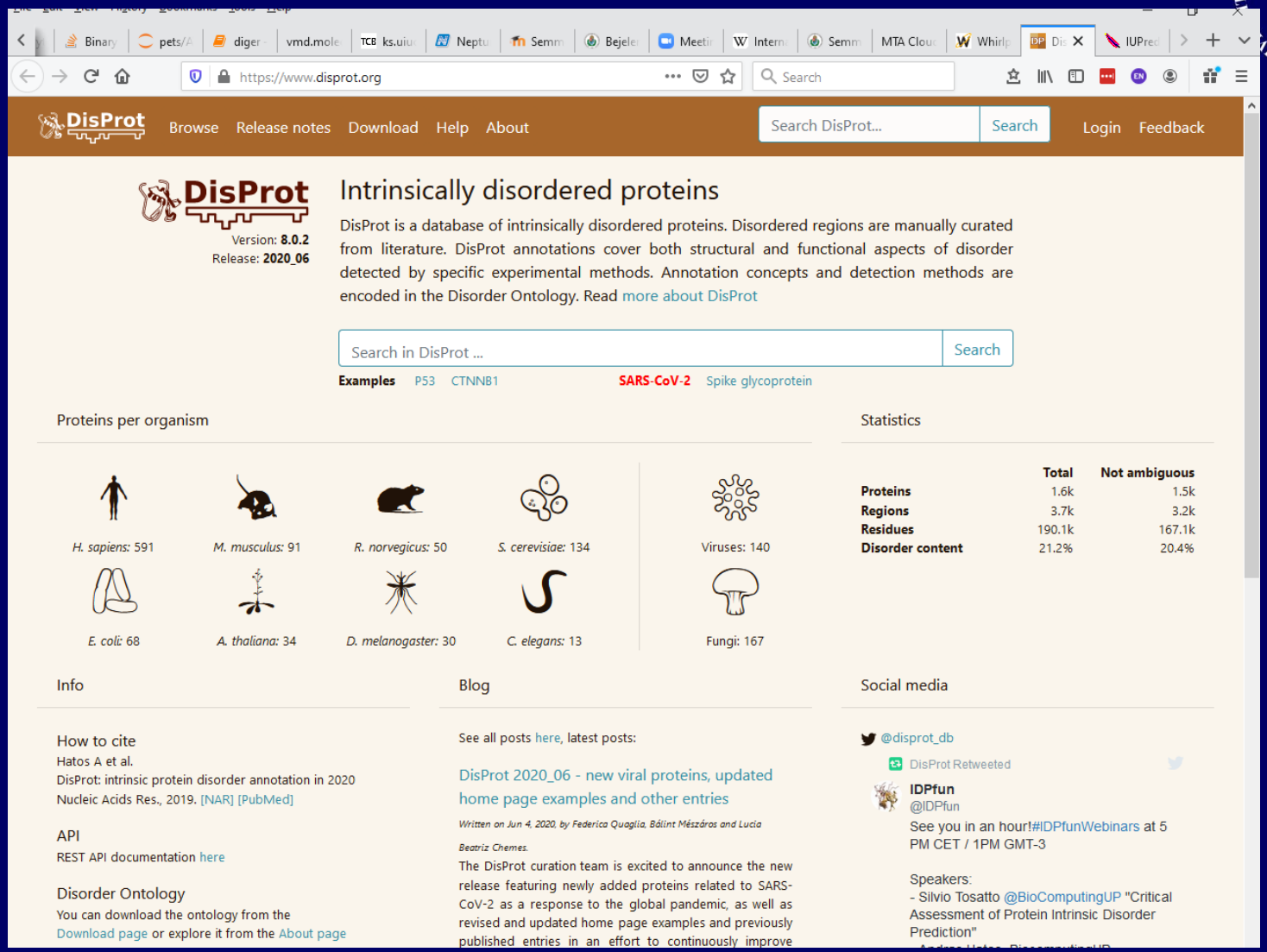
- Kapcsolat a külvilággal:
  - Anyagforgalom, jelátvitel
- Érdekes kérdések:
  - A fehérje membrán kötött vagy nem?
  - Hol vannak benne a TM szakaszok?
  - Milyen a topológiája?
- Szignál peptid:
  - Teljes értékű TM szakasz, csak nem állandó

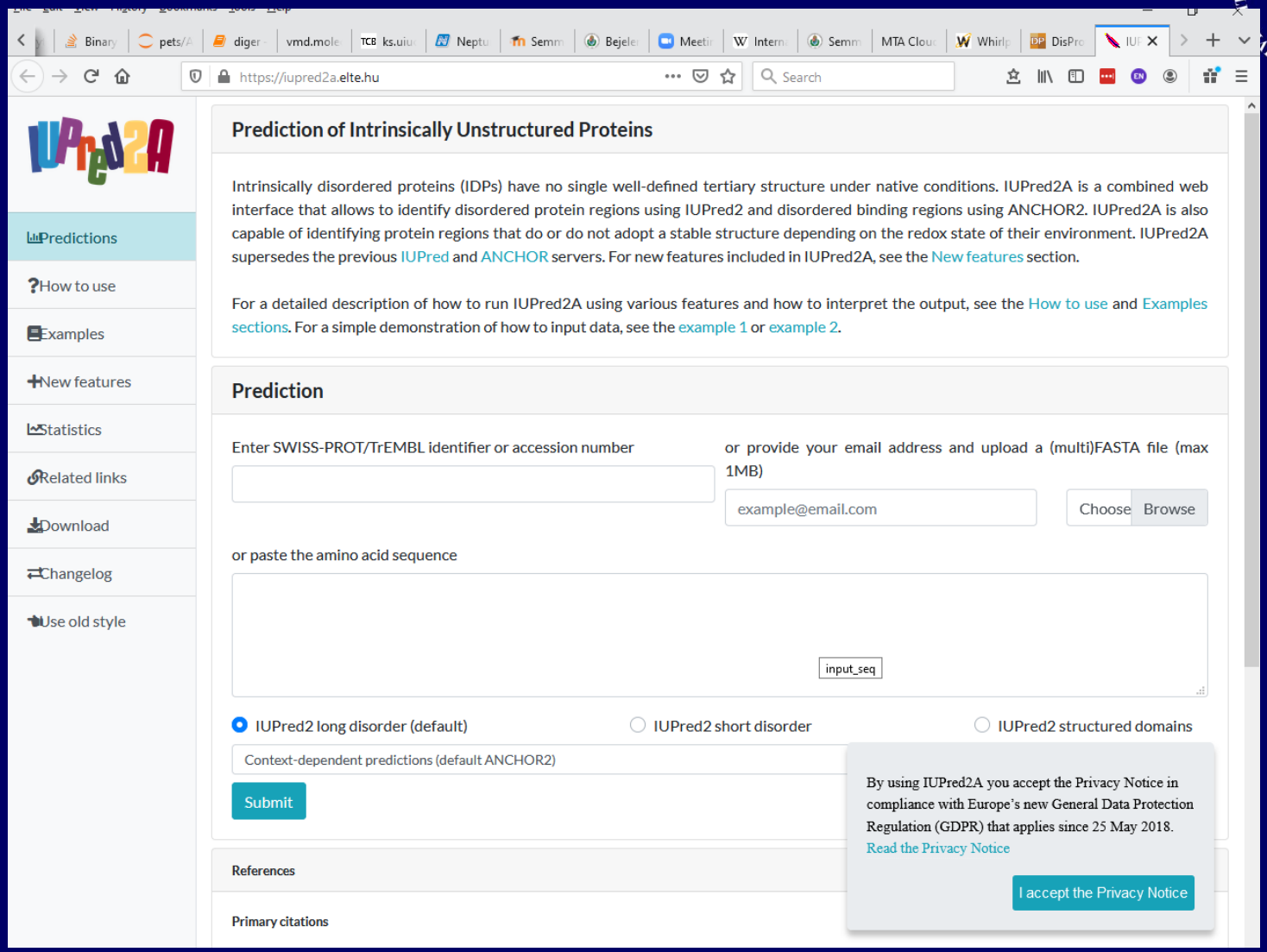
Szerkezet nélküli fehérjék (IUP)

- Sok hidrofíl aminósav jellemzi
- Nagyon jól oldódik vízben, ezért nincs stabil szerkezete
- Stabil szerkezetű fehérjével képes kölcsönhatásba lépni
- Rugalmas kötőelemek
- Fehérje-fehérje kölcsönhatások résztvevői
- Kölcsönhatási hálózatok csomópontjai

Kapcsolódó web-helyek

- DisProt adatbázis – részben vagy egészben szerkezet nélküli fehérjék
- Web: http://www.disprot.org/
- IUPred: online szolgáltatás szerkezet nélküli fehérjék elemzéséhez
- Web: http://iupred.enzim.hu/
- Dosztányi, Tompa, Simon


https://www.disprot.org
DisProt
Browse
Release notes
Download
Help
About
Search DisProt...
Search
Login
Feedback
DisProt
Version: 8.0.2
Release: 2020_06
Intrinsically disordered proteins
DisProt is a database of intrinsically disordered proteins. Disordered regions are manually curated from literature. DisProt annotations cover both structural and functional aspects of disorder detected by specific experimental methods. Annotation concepts and detection methods are encoded in the Disorder Ontology. Read more about DisProt
Search in DisProt ...
Search
Examples P53 CTNNB1 SARS-CoV-2 Spike glycoprotein
Proteins per organism
H. sapiens: 591
M. musculus: 91
R. norvegicus: 50
S. cerevisiae: 134
Viruses: 140
E. coli: 68
A. thaliana: 34
D. melanogaster: 30
C. elegans: 13
Fungi: 167
Statistics
Total
Not ambiguous
Proteins 1.6k 1.5k
Regions 3.7k 3.2k
Residues 190.1k 167.1k
Disorder content 21.2% 20.4%
Info
How to cite
Hatos A et al.
DisProt: intrinsic protein disorder annotation in 2020
Nucleic Acids Res., 2019. [NAR] [PubMed]
API
REST API documentation here
Disorder Ontology
You can download the ontology from the
Download page or explore it from the About page
Blog
See all posts here, latest posts:
DisProt 2020_06 - new viral proteins, updated home page examples and other entries
Written on Jun 4, 2020, by Federica Quaglia, Bálint Mészáros and Lucia Beatriz Chemes.
The DisProt curation team is excited to announce the new release featuring newly added proteins related to SARS-CoV-2 as a response to the global pandemic, as well as revised and updated home page examples and previously published entries in an effort to continuously improve
Social media
@disprot_db
DisProt Retweeted
IDPfun
@IDPfun
See you in an hour!#IDPfunWebinars at 5 PM CET / 1PM GMT-3
Speakers:
- Silvio Tosatto @BioComputingUP "Critical Assessment of Protein Intrinsic Disorder Prediction"

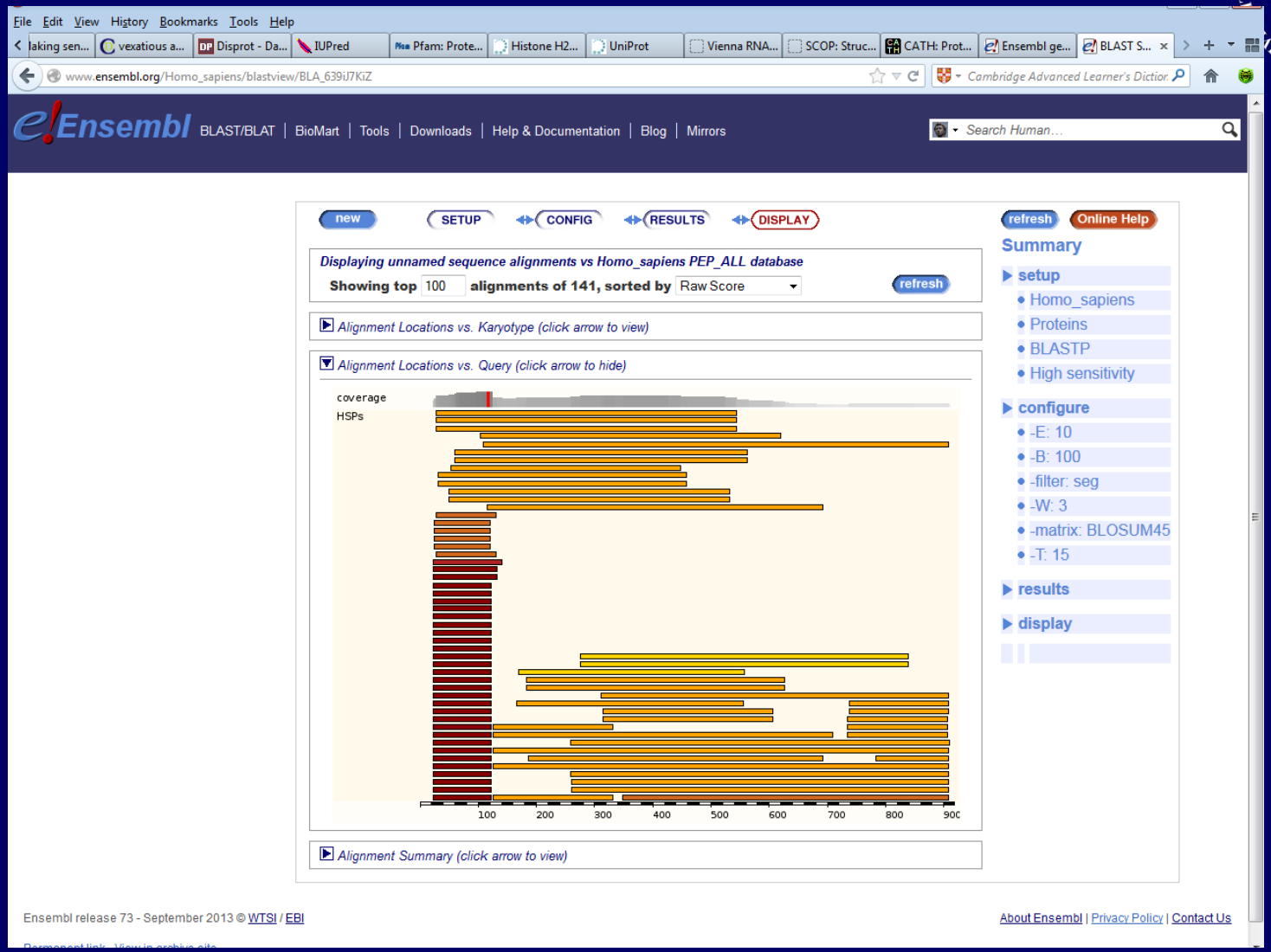

https://iupred2a.elte.hu
Search
Prediction of Intrinsically Unstructured Proteins
Intrinsically disordered proteins (IDPs) have no single well-defined tertiary structure under native conditions. IUPred2A is a combined web interface that allows to identify disordered protein regions using IUPred2 and disordered binding regions using ANCHOR2. IUPred2A is also capable of identifying protein regions that do or do not adopt a stable structure depending on the redox state of their environment. IUPred2A supersedes the previous IUPred and ANCHOR servers. For new features included in IUPred2A, see the New features section.
For a detailed description of how to run IUPred2A using various features and how to interpret the output, see the How to use and Examples sections. For a simple demonstration of how to input data, see the example 1 or example 2.
Predictions
How to use
Examples
New features
Statistics
Related links
Download
Changelog
Use old style
Prediction
Enter SWISS-PROT/TrEMBL identifier or accession number
or provide your email address and upload a (multi)FASTA file (max 1MB)
example@email.com
Choose
Browse
or paste the amino acid sequence
input_seq
IUPred2 long disorder (default)
IUPred2 short disorder
IUPred2 structured domains
Context-dependent predictions (default ANCHOR2)
Submit
By using IUPred2A you accept the Privacy Notice in compliance with Europe's new General Data Protection Regulation (GDPR) that applies since 25 May 2018.
Read the Privacy Notice
I accept the Privacy Notice
References
Primary citations

Annotálás homológia alapján

- Alapelv: ami fontos – az konzervált
- Már ismert szekvenciák alapján annotáljuk az újakat
- Szerkezeti szempontból
- Funkció szempontjából
- Szabályozás szempontjából
- Kölcsönhatás szempontjából

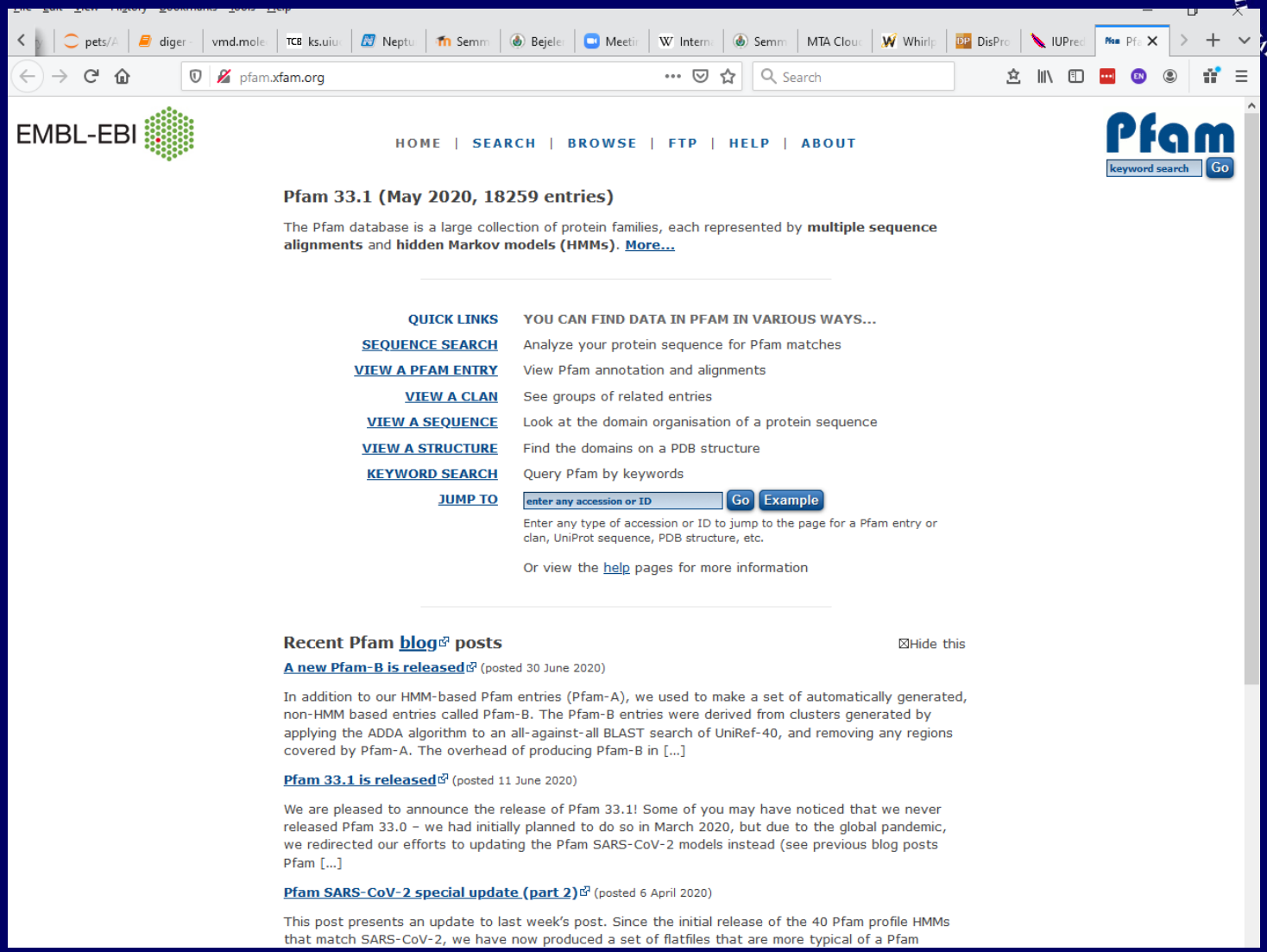
Moduláris fehérjék

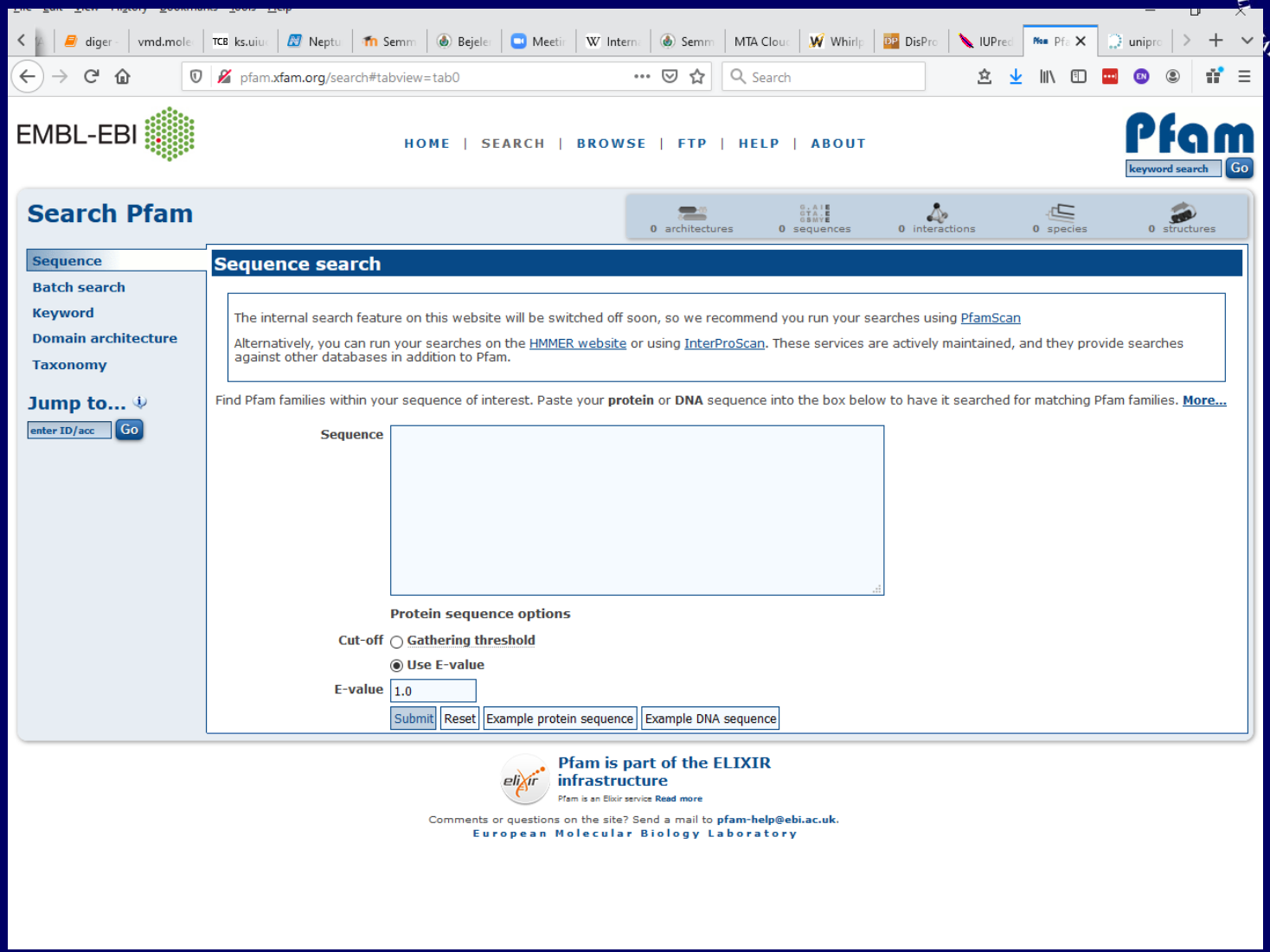
- A BLAST-találatok nem egyenletesen oszlanak el a szekvencia mentén
- Egyes szakaszokból sok van egy genomon belül
- Fehérje domén egységek:
  - Génszerkezeti
  - Fehérje térszerkezeti
  - Funkcionális értelemben

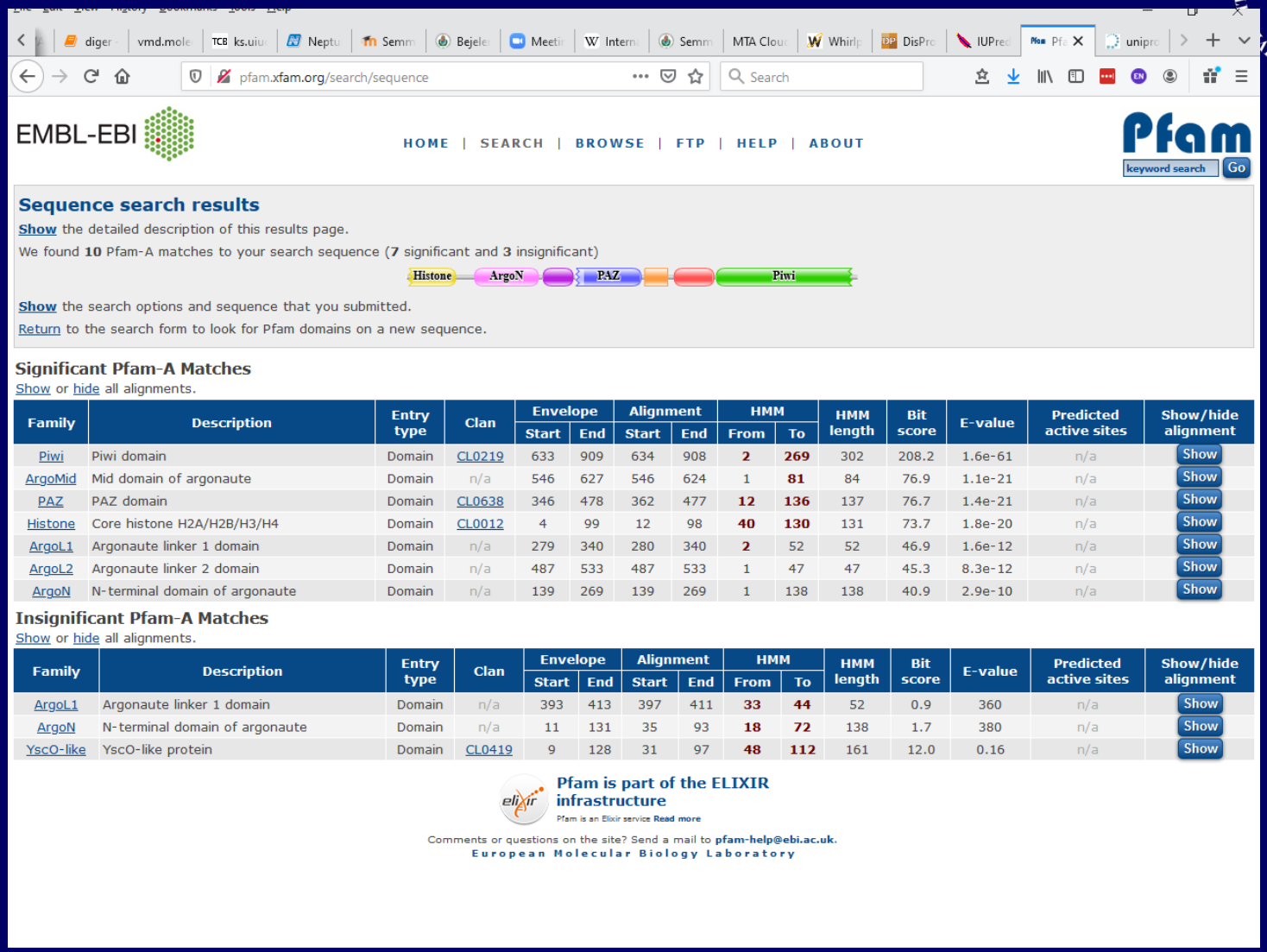

File Edit View History Bookmarks Tools Help
www.ensembl.org/Homo_sapiens/blastview/BLA_639iJ7KiZ
Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Search Human...
new
SETUP
CONFIG
RESULTS
DISPLAY
refresh
Online Help
Displaying unnamed sequence alignments vs Homo_sapiens PEP_ALL database
Showing top 100 alignments of 141, sorted by Raw Score
refresh
Alignment Locations vs. Karyotype (click arrow to view)
Alignment Locations vs. Query (click arrow to hide)
coverage
HSPs
100
200
300
400
500
600
700
800
Alignment Summary (click arrow to view)
Summary
setup
Homo_sapiens
Proteins
BLASTP
High sensitivity
configure
-E: 10
-B: 100
-filter: seg
-W: 3
-matrix: BLOSUM45
-T: 15
results
display
Ensembl release 73 - September 2013 © WTSI / EBI
About Ensembl | Privacy Policy | Contact Us

Kapcsolódó adatbázisok

- Pfam – "Protein Family"
  - Többszörös illesztés alapján készült
  - Web: http://pfam.xfam.org/
- UniProt
  - Fehérjék általános adatbázisa
  - Web: http://www.uniprot.org/
- Nemzetközi konzorciumok üzemeltetik
- Nagyon megbízhatóak


pfam.xfam.org
Search
EMBL-EBI
HOME | SEARCH | BROWSE | FTP | HELP | ABOUT
Pfam
keyword search
Go
Pfam 33.1 (May 2020, 18259 entries)
The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs). More...
QUICK LINKS
YOU CAN FIND DATA IN PFAM IN VARIOUS WAYS...
SEQUENCE SEARCH
Analyze your protein sequence for Pfam matches
VIEW A PFAM ENTRY
View Pfam annotation and alignments
VIEW A CLAN
See groups of related entries
VIEW A SEQUENCE
Look at the domain organisation of a protein sequence
VIEW A STRUCTURE
Find the domains on a PDB structure
KEYWORD SEARCH
Query Pfam by keywords
JUMP TO
enter any accession or ID
Go
Example
Enter any type of accession or ID to jump to the page for a Pfam entry or clan, UniProt sequence, PDB structure, etc.
Or view the help pages for more information
Recent Pfam blog posts
Hide this
A new Pfam-B is released (posted 30 June 2020)
In addition to our HMM-based Pfam entries (Pfam-A), we used to make a set of automatically generated, non-HMM based entries called Pfam-B. The Pfam-B entries were derived from clusters generated by applying the ADDA algorithm to an all-against-all BLAST search of UniRef-40, and removing any regions covered by Pfam-A. The overhead of producing Pfam-B in [...]
Pfam 33.1 is released (posted 11 June 2020)
We are pleased to announce the release of Pfam 33.1! Some of you may have noticed that we never released Pfam 33.0 – we had initially planned to do so in March 2020, but due to the global pandemic, we redirected our efforts to updating the Pfam SARS-CoV-2 models instead (see previous blog posts Pfam [...]
Pfam SARS-CoV-2 special update (part 2) (posted 6 April 2020)
This post presents an update to last week's post. Since the initial release of the 40 Pfam profile HMMs that match SARS-CoV-2, we have now produced a set of flatfiles that are more typical of a Pfam


pfam.xfam.org/search#tabview=tab0
EMBL-EBI
HOME | SEARCH | BROWSE | FTP | HELP | ABOUT
Pfam
keyword search
Go
Search Pfam
0 architectures
0 sequences
0 interactions
0 species
0 structures
Sequence
Batch search
Keyword
Domain architecture
Taxonomy
Jump to...
enter ID/acc
Go
Sequence search
The internal search feature on this website will be switched off soon, so we recommend you run your searches using PfamScan
Alternatively, you can run your searches on the HMMER website or using InterProScan. These services are actively maintained, and they provide searches against other databases in addition to Pfam.
Find Pfam families within your sequence of interest. Paste your protein or DNA sequence into the box below to have it searched for matching Pfam families. More...
Sequence
Protein sequence options
Cut-off
Gathering threshold
Use E-value
E-value
1.0
Submit
Reset
Example protein sequence
Example DNA sequence
Pfam is part of the ELIXIR infrastructure
Pfam is an Elixir service Read more
Comments or questions on the site? Send a mail to pfam-help@ebi.ac.uk.
European Molecular Biology Laboratory



pfam.xfam.org/search/sequence

EMBL-EBI

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Pfam keyword search Go

Sequence search results

Show the detailed description of this results page.

We found **10** Pfam-A matches to your search sequence (**7** significant and **3** insignificant)

Histone — ArgoN — PAZ — Piwi

Show the search options and sequence that you submitted.

Return to the search form to look for Pfam domains on a new sequence.

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
Piwi	Piwi domain	Domain	CL0219	633	909	634	908	2	269	302	208.2	1.6e-61	n/a	Show
ArgoMid	Mid domain of argonaute	Domain	n/a	546	627	546	624	1	81	84	76.9	1.1e-21	n/a	Show
PAZ	PAZ domain	Domain	CL0638	346	478	362	477	12	136	137	76.7	1.4e-21	n/a	Show
Histone	Core histone H2A/H2B/H3/H4	Domain	CL0012	4	99	12	98	40	130	131	73.7	1.8e-20	n/a	Show
ArgoL1	Argonaute linker 1 domain	Domain	n/a	279	340	280	340	2	52	52	46.9	1.6e-12	n/a	Show
ArgoL2	Argonaute linker 2 domain	Domain	n/a	487	533	487	533	1	47	47	45.3	8.3e-12	n/a	Show
ArgoN	N-terminal domain of argonaute	Domain	n/a	139	269	139	269	1	138	138	40.9	2.9e-10	n/a	Show

Insignificant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
ArgoL1	Argonaute linker 1 domain	Domain	n/a	393	413	397	411	33	44	52	0.9	360	n/a	Show
ArgoN	N-terminal domain of argonaute	Domain	n/a	11	131	35	93	18	72	138	1.7	380	n/a	Show
YscO-like	YscO-like protein	Domain	CL0419	9	128	31	97	48	112	161	12.0	0.16	n/a	Show

Pfam is part of the ELIXIR infrastructure

Pfam is an Elixir service Read more

Comments or questions on the site? Send a mail to pfam-help@ebi.ac.uk.

European Molecular Biology Laboratory

pfam.xfam.org/family/PF02170.22#tabview=tab0

EMBL-EBI

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Pfam keyword search Go

Family: *PAZ* (PF02170)

289 architectures | 8493 sequences | 4 interactions | 1033 species | 87 structures

- Summary
- Domain organisation
- Clan
- Alignments
- HMM logo
- Trees
- Curation & model
- Species
- Interactions
- Structures

Jump to... enter ID/acc Go

Summary: PAZ domain

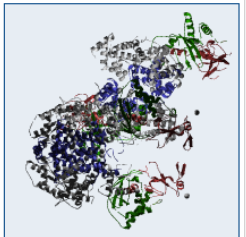
Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

No Wikipedia article | Pfam | InterPro

This tab holds the annotation information that is stored in the Pfam database. As we move to using Wikipedia as our main source of annotation, the contents of this tab will be gradually replaced by the Wikipedia tab.

PAZ domain Provide feedback

This domain is named PAZ after the proteins Piwi Argonaut and Zwille. This domain is found in two families of proteins that are involved in post-transcriptional gene silencing. These are the Piwi family and the Dicer family, that includes the Carpel factory protein. The function of the domains is unknown but has been suggested to mediate complex formation between proteins of the Piwi and Dicer families by hetero-dimerisation. The three-dimensional structure of this domain has been solved [2-4]. The PAZ domain is composed of two subdomains. One subdomain is similar to the OB fold, albeit with a different topology. The OB-fold is well known as a single-stranded nucleic acid binding fold. The second subdomain is composed of a beta-hairpin followed by an alpha-helix. The PAZ domains shows low-affinity nucleic acid binding and appears to interact with the 3' ends of single-stranded regions of RNA in the cleft between the two subdomains. PAZ can bind the characteristic two-base 3' overhangs of siRNAs, indicating that although PAZ may not be a primary nucleic acid binding site in Dicer or RISC, it may contribute to the specific and productive incorporation of siRNAs and miRNAs into the RNAi pathway.



Example structure

PDB entry 2QVW: Structure of Giardia Dicer refined against twinned data

View a different structure: 2QVW

Literature references

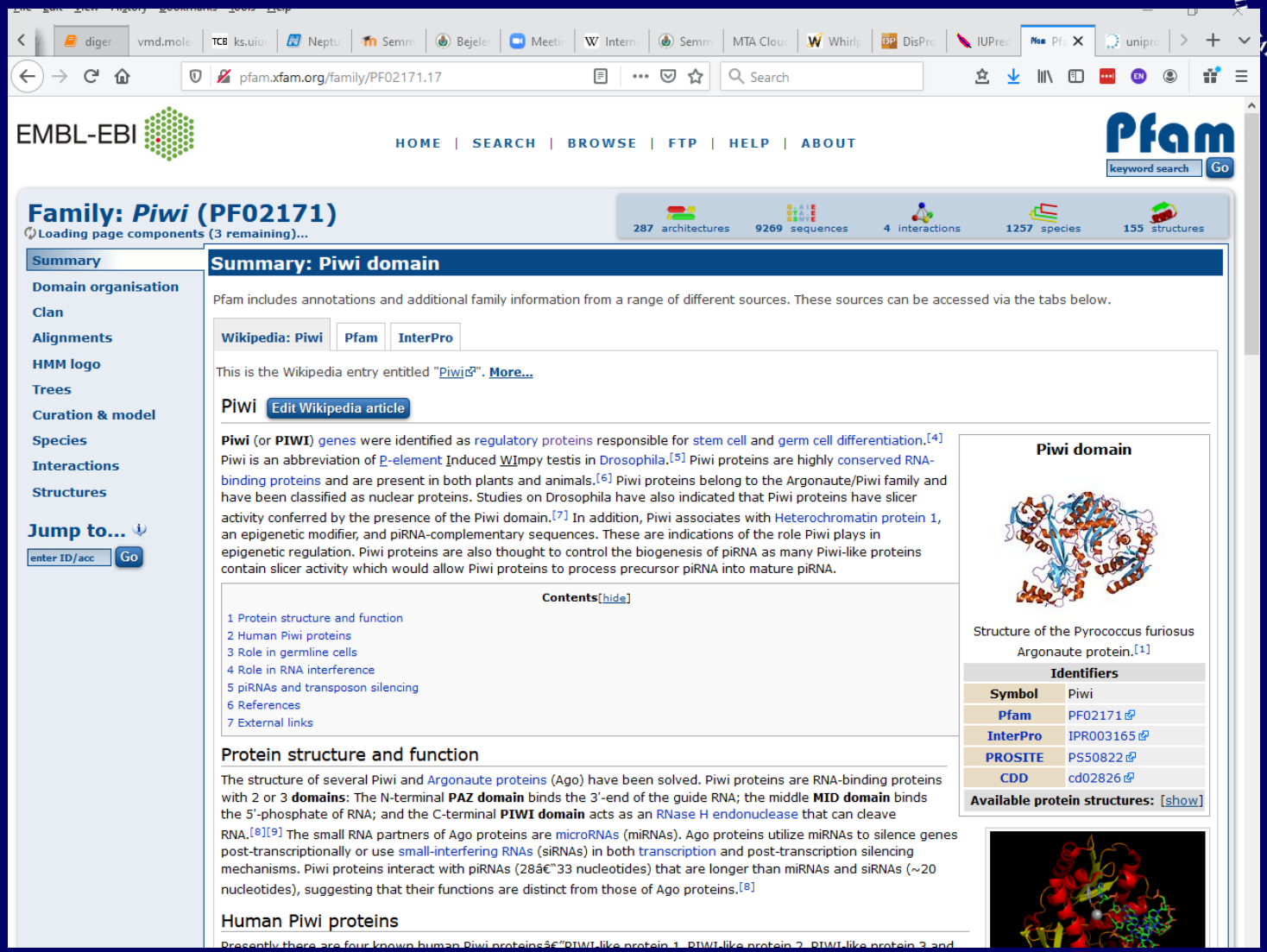
1. Cerutti L, Mian N, Bateman A; , Trends Biochem Sci 2000;25:481-482.: Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. PUBMED:11050429 EPMC:11050429
2. Song JJ, Liu J, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, Hannon GJ, Joshua-Tor L; , Nat Struct Biol 2003;10:1026-1032.: The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. PUBMED:14625589 EPMC:14625589
3. Yan KS, Yan S, Farooq A, Han A, Zeng L, Zhou MM; , Nature 2003;426:468-474.: Structure and conserved RNA binding of the PAZ domain. PUBMED:14615802 EPMC:14615802
4. Lingel A, Simon B, Izaurralde E, Sattler M; , Nature 2003;426:465-469.: Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. PUBMED:14615801 EPMC:14615801

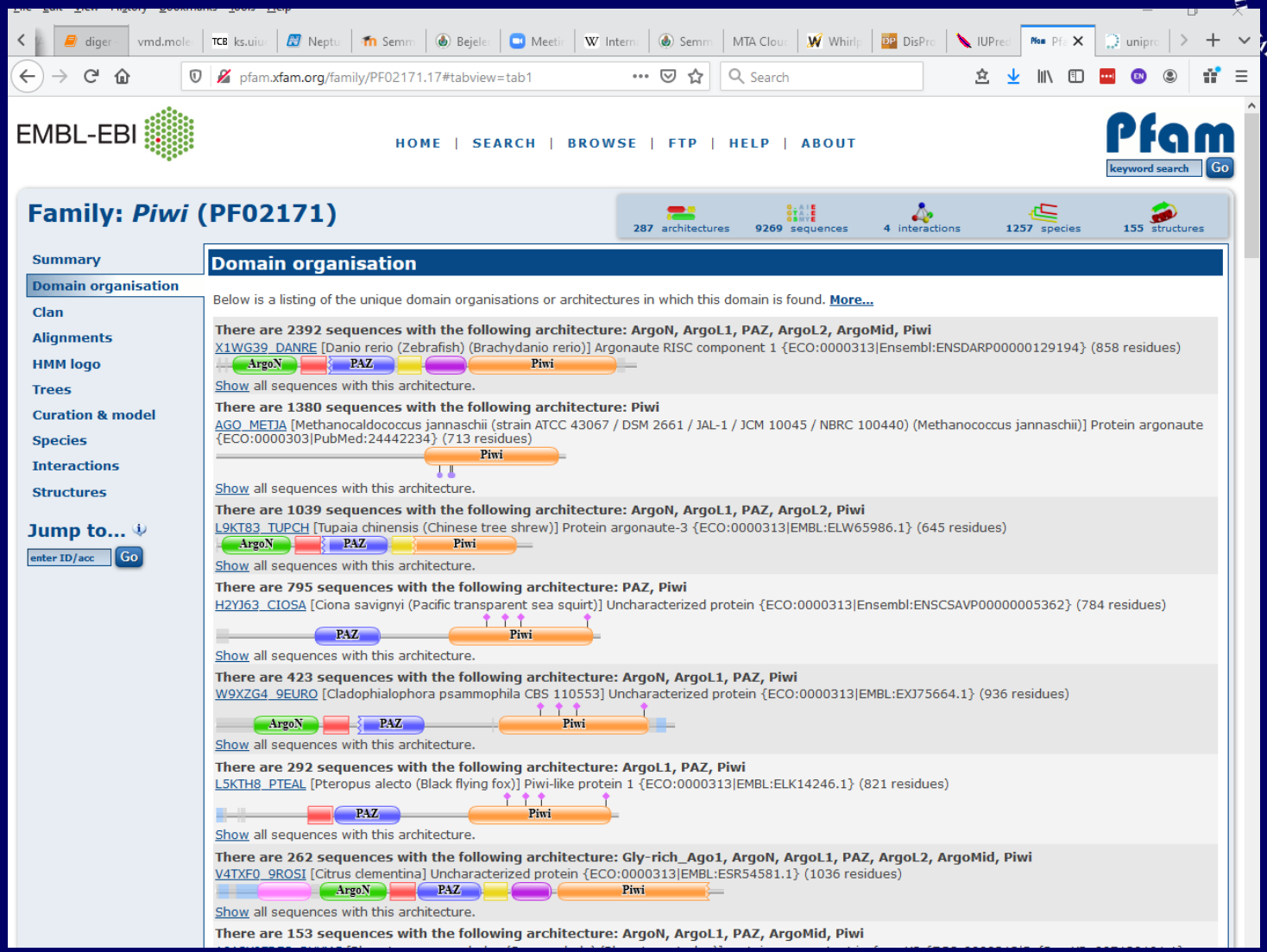
Internal database links

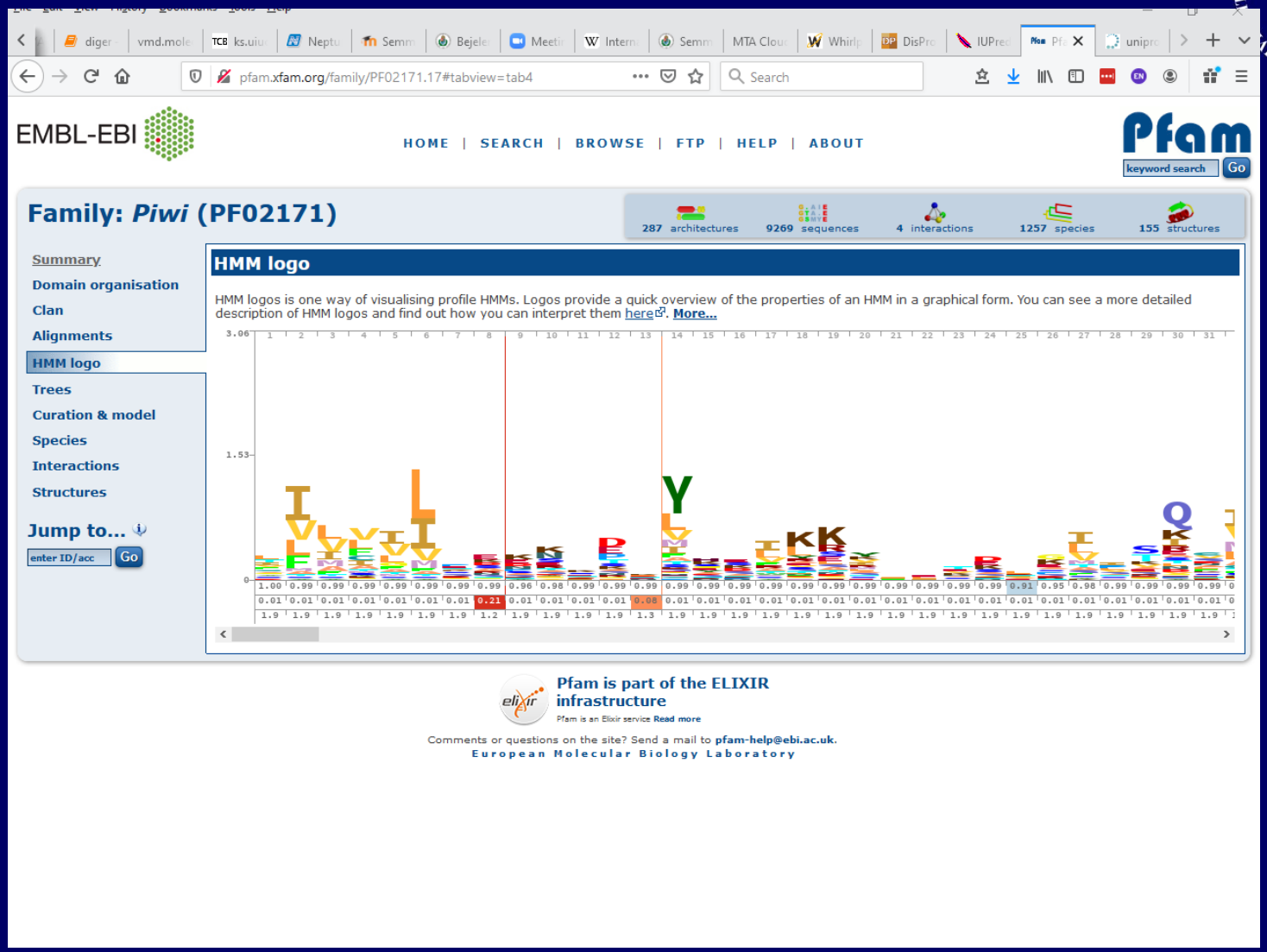
SCOOP: ArgoL2

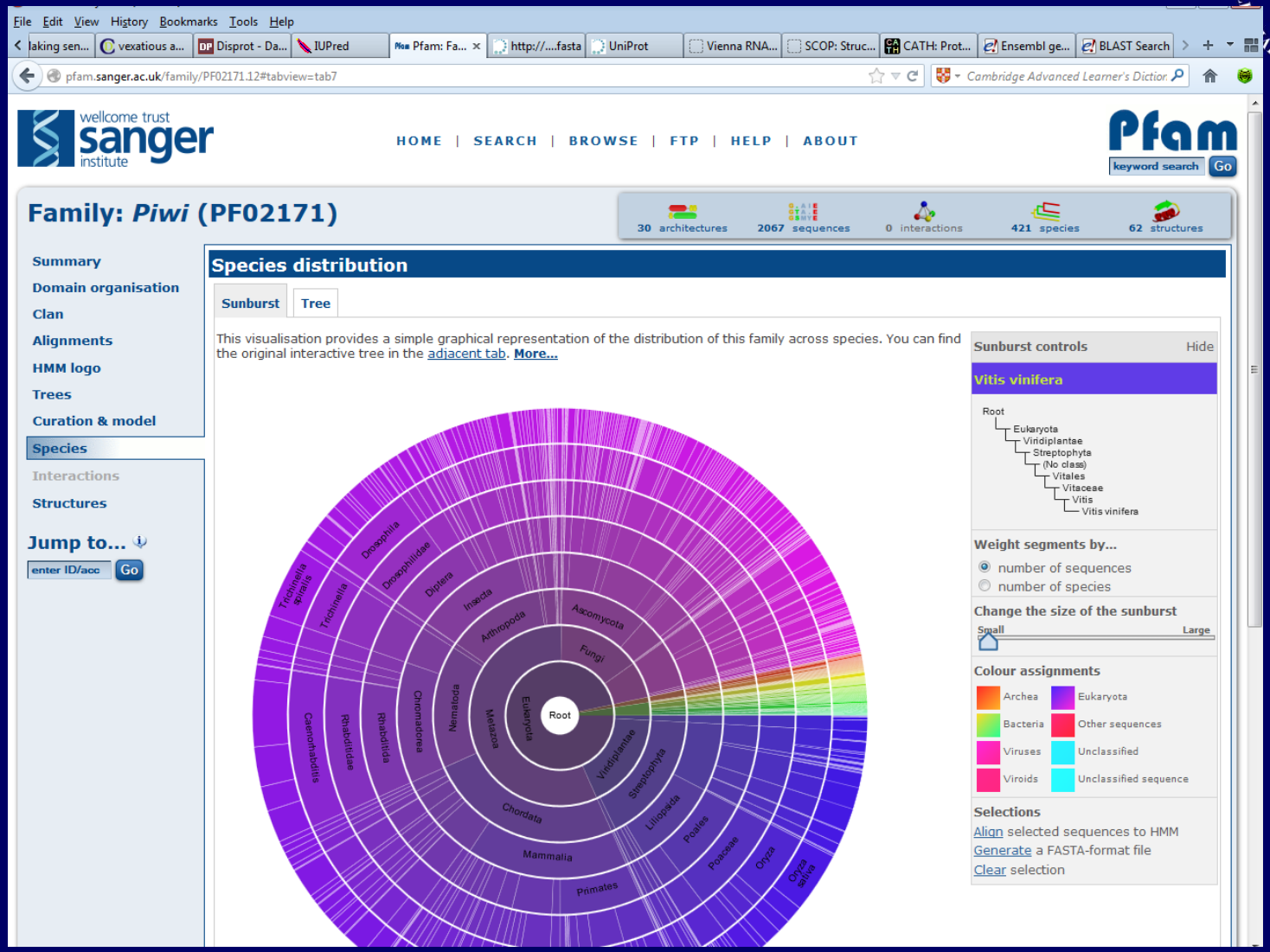
External database links

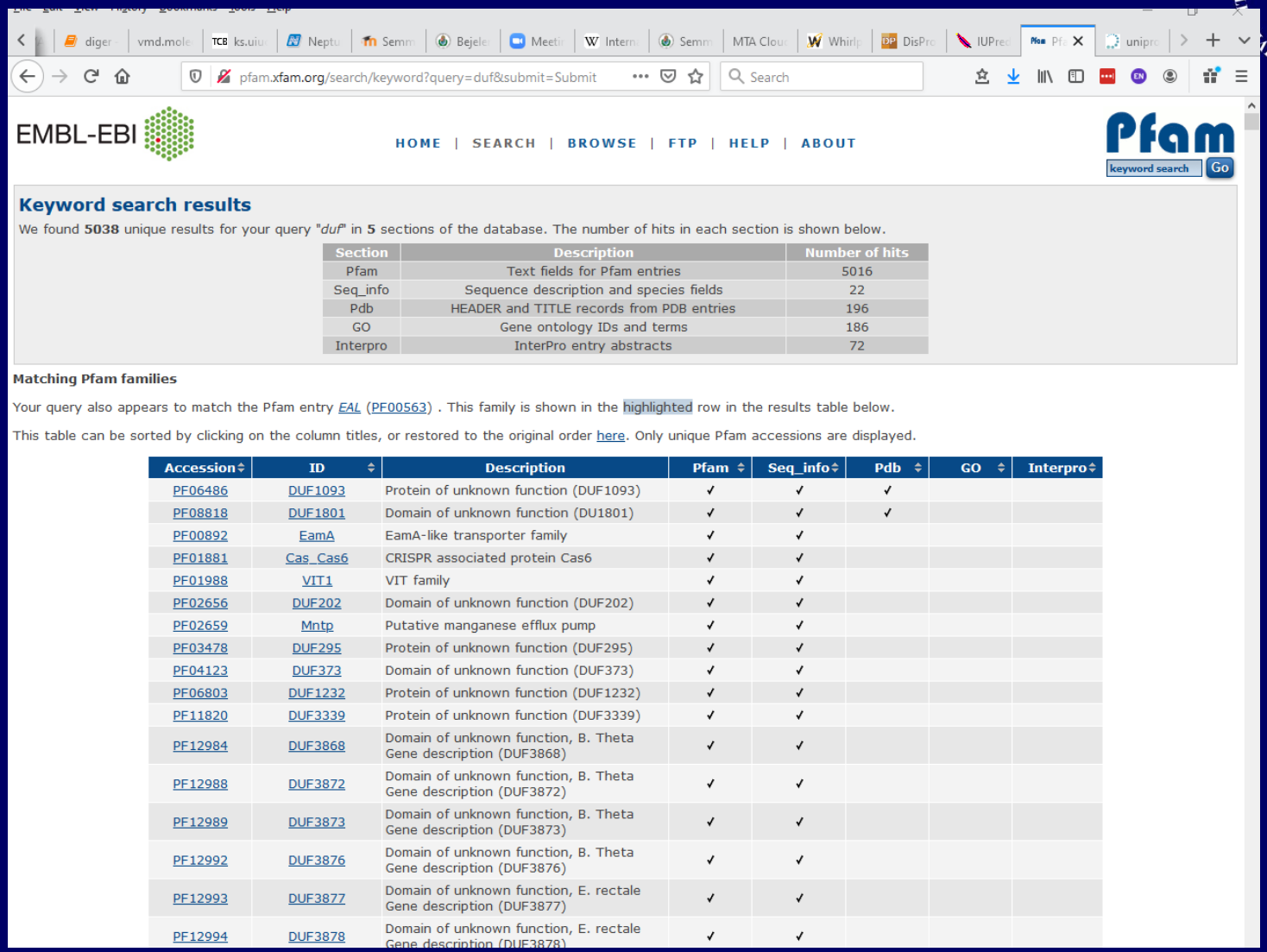
SCOOP: 1r4k


pfam.xfam.org/family/PF02171.17
EMBL-EBI
HOME | SEARCH | BROWSE | FTP | HELP | ABOUT
Pfam
keyword search
Go
Family: Piwi (PF02171)
Loading page components (3 remaining)...
287 architectures
9269 sequences
4 interactions
1257 species
155 structures
Summary
Domain organisation
Clan
Alignments
HMM logo
Trees
Curation & model
Species
Interactions
Structures
Jump to...
enter ID/acc
Go
Summary: Piwi domain
Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.
Wikipedia: Piwi
Pfam
InterPro
This is the Wikipedia entry entitled "Piwi". More...
Piwi
Edit Wikipedia article
Piwi (or PIWI) genes were identified as regulatory proteins responsible for stem cell and germ cell differentiation.[4] Piwi is an abbreviation of P-element Induced WImpy testis in Drosophila.[5] Piwi proteins are highly conserved RNA-binding proteins and are present in both plants and animals.[6] Piwi proteins belong to the Argonaute/Piwi family and have been classified as nuclear proteins. Studies on Drosophila have also indicated that Piwi proteins have slicer activity conferred by the presence of the Piwi domain.[7] In addition, Piwi associates with Heterochromatin protein 1, an epigenetic modifier, and piRNA-complementary sequences. These are indications of the role Piwi plays in epigenetic regulation. Piwi proteins are also thought to control the biogenesis of piRNA as many Piwi-like proteins contain slicer activity which would allow Piwi proteins to process precursor piRNA into mature piRNA.
Contents[hide]
1 Protein structure and function
2 Human Piwi proteins
3 Role in germline cells
4 Role in RNA interference
5 piRNAs and transposon silencing
6 References
7 External links
Protein structure and function
The structure of several Piwi and Argonaute proteins (Ago) have been solved. Piwi proteins are RNA-binding proteins with 2 or 3 domains: The N-terminal PAZ domain binds the 3'-end of the guide RNA; the middle MID domain binds the 5'-phosphate of RNA; and the C-terminal PIWI domain acts as an RNase H endonuclease that can cleave RNA.[8][9] The small RNA partners of Ago proteins are microRNAs (miRNAs). Ago proteins utilize miRNAs to silence genes post-transcriptionally or use small-interfering RNAs (siRNAs) in both transcription and post-transcription silencing mechanisms. Piwi proteins interact with piRNAs (28â€“33 nucleotides) that are longer than miRNAs and siRNAs (~20 nucleotides), suggesting that their functions are distinct from those of Ago proteins.[8]
Human Piwi proteins
Piwi domain
Structure of the Pyrococcus furiosus Argonaute protein.[1]
Identifiers
Symbol Piwi
Pfam PF02171
InterPro IPR003165
PROSITE PS50822
CDD cd02826
Available protein structures: [show]


pfam.xfam.org/family/PF02171.17#tabview=tab1
EMBL-EBI
HOME | SEARCH | BROWSE | FTP | HELP | ABOUT
Pfam
keyword search
Go
Family: Piwi (PF02171)
287 architectures
9269 sequences
4 interactions
1257 species
155 structures
Summary
Domain organisation
Clan
Alignments
HMM logo
Trees
Curation & model
Species
Interactions
Structures
Jump to...
enter ID/acc
Go
Domain organisation
Below is a listing of the unique domain organisations or architectures in which this domain is found. More...
There are 2392 sequences with the following architecture: ArgoN, ArgoL1, PAZ, ArgoL2, ArgoMid, Piwi
X1WG39_DANRE [Danio rerio (Zebrafish) (Brachydanio rerio)] Argonaute RISC component 1 {ECO:0000313|Ensembl:ENSDARP00000129194} (858 residues)
ArgoN PAZ Piwi
Show all sequences with this architecture.
There are 1380 sequences with the following architecture: Piwi
AGO_METJA [Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) (Methanococcus jannaschii)] Protein argonaute {ECO:0000303|PubMed:24442234} (713 residues)
Piwi
Show all sequences with this architecture.
There are 1039 sequences with the following architecture: ArgoN, ArgoL1, PAZ, ArgoL2, Piwi
L9KT83_TUPCH [Tupaia chinensis (Chinese tree shrew)] Protein argonaute-3 {ECO:0000313|EMBL:ELW65986.1} (645 residues)
ArgoN PAZ Piwi
Show all sequences with this architecture.
There are 795 sequences with the following architecture: PAZ, Piwi
H2YJ63_CIOSA [Ciona savignyi (Pacific transparent sea squirt)] Uncharacterized protein {ECO:0000313|Ensembl:ENSCSAVP00000005362} (784 residues)
PAZ Piwi
Show all sequences with this architecture.
There are 423 sequences with the following architecture: ArgoN, ArgoL1, PAZ, Piwi
W9XZG4_9EURO [Cladophialophora psammophila CBS 110553] Uncharacterized protein {ECO:0000313|EMBL:EXJ75664.1} (936 residues)
ArgoN PAZ Piwi
Show all sequences with this architecture.
There are 292 sequences with the following architecture: ArgoL1, PAZ, Piwi
L5KTH8_PTEAL [Pteropus alecto (Black flying fox)] Piwi-like protein 1 {ECO:0000313|EMBL:ELK14246.1} (821 residues)
PAZ Piwi
Show all sequences with this architecture.
There are 262 sequences with the following architecture: Gly-rich_Ago1, ArgoN, ArgoL1, PAZ, ArgoL2, ArgoMid, Piwi
V4TXF0_9ROSI [Citrus clementina] Uncharacterized protein {ECO:0000313|EMBL:ESR54581.1} (1036 residues)
ArgoN PAZ Piwi
Show all sequences with this architecture.
There are 153 sequences with the following architecture: ArgoN, ArgoL1, PAZ, ArgoMid, Piwi


pfam.xfam.org/family/PF02171.17#tabview=tab4
EMBL-EBI
HOME | SEARCH | BROWSE | FTP | HELP | ABOUT
Pfam
keyword search
Go
Family: Piwi (PF02171)
287 architectures
9269 sequences
4 interactions
1257 species
155 structures
Summary
Domain organisation
Clan
Alignments
HMM logo
Trees
Curation & model
Species
Interactions
Structures
Jump to...
enter ID/acc
Go
HMM logo
HMM logos is one way of visualising profile HMMs. Logos provide a quick overview of the properties of an HMM in a graphical form. You can see a more detailed description of HMM logos and find out how you can interpret them here. More...
Pfam is part of the ELIXIR infrastructure
Pfam is an Elixir service Read more
Comments or questions on the site? Send a mail to pfam-help@ebi.ac.uk.
European Molecular Biology Laboratory


File Edit View History Bookmarks Tools Help
pfam.sanger.ac.uk/family/PF02171.12#tabview=tab7
wellcome trust sanger institute
HOME | SEARCH | BROWSE | FTP | HELP | ABOUT
Pfam
keyword search Go
Family: Piwi (PF02171)
30 architectures
2067 sequences
0 interactions
421 species
62 structures
Summary
Domain organisation
Clan
Alignments
HMM logo
Trees
Curation & model
Species
Interactions
Structures
Jump to...
enter ID/acc Go
Species distribution
Sunburst
Tree
This visualisation provides a simple graphical representation of the distribution of this family across species. You can find the original interactive tree in the adjacent tab. More...
Sunburst controls
Hide
Vitis vinifera
Root
Eukaryota
Viridiplantae
Streptophyta
(No class)
Vitales
Vitaceae
Vitis
Vitis vinifera
Weight segments by...
number of sequences
number of species
Change the size of the sunburst
Small
Large
Colour assignments
Archea
Eukaryota
Bacteria
Other sequences
Viruses
Unclassified
Viroids
Unclassified sequence
Selections
Align selected sequences to HMM
Generate a FASTA-format file
Clear selection
Root
Eukaryota
Metazoa
Nematoda
Chromadorea
Rhabditida
Rhabditidae
Caenorhabditis
Chordata
Mammalia
Primates
Viridiplantae
Streptophyta
Liliopsida
Poales
Poaceae
Oryza
Oryza sativa
Arthropoda
Insecta
Diptera
Drosophilidae
Drosophila
Trichinella
Trichinella spiralis
Fungi
Ascomycota



pfam.xfam.org/search/keyword?query=duf&submit=Submit

EMBL-EBI

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Pfam keyword search Go

Keyword search results

We found **5038** unique results for your query "*duf*" in **5** sections of the database. The number of hits in each section is shown below.

Section	Description	Number of hits
Pfam	Text fields for Pfam entries	5016
Seq_info	Sequence description and species fields	22
Pdb	HEADER and TITLE records from PDB entries	196
GO	Gene ontology IDs and terms	186
Interpro	InterPro entry abstracts	72

Matching Pfam families

Your query also appears to match the Pfam entry *EAL* (PF00563) . This family is shown in the highlighted row in the results table below.

This table can be sorted by clicking on the column titles, or restored to the original order here. Only unique Pfam accessions are displayed.

Accession	ID	Description	Pfam	Seq_info	Pdb	GO	Interpro
PF06486	DUF1093	Protein of unknown function (DUF1093)	✓	✓	✓		
PF08818	DUF1801	Domain of unknown function (DU1801)	✓	✓	✓		
PF00892	EamA	EamA-like transporter family	✓	✓			
PF01881	Cas_Cas6	CRISPR associated protein Cas6	✓	✓			
PF01988	VIT1	VIT family	✓	✓			
PF02656	DUF202	Domain of unknown function (DUF202)	✓	✓			
PF02659	Mntp	Putative manganese efflux pump	✓	✓			
PF03478	DUF295	Protein of unknown function (DUF295)	✓	✓			
PF04123	DUF373	Domain of unknown function (DUF373)	✓	✓			
PF06803	DUF1232	Protein of unknown function (DUF1232)	✓	✓			
PF11820	DUF3339	Protein of unknown function (DUF3339)	✓	✓			
PF12984	DUF3868	Domain of unknown function, B. Theta Gene description (DUF3868)	✓	✓			
PF12988	DUF3872	Domain of unknown function, B. Theta Gene description (DUF3872)	✓	✓			
PF12989	DUF3873	Domain of unknown function, B. Theta Gene description (DUF3873)	✓	✓			
PF12992	DUF3876	Domain of unknown function, B. Theta Gene description (DUF3876)	✓	✓			
PF12993	DUF3877	Domain of unknown function, E. rectale Gene description (DUF3877)	✓	✓			
PF12994	DUF3878	Domain of unknown function, E. rectale Gene description (DUF3878)	✓	✓			

DisProt - Database of prote... | IUPred | http://www.u...8WWP9.fasta | Pfam: Family: DUF1775 (PF...

pfam.xfam.org/family/PF07987

EMBL-EBI

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Pfam keyword search Go

Family: *DUF1775* (PF07987)

6 architectures | 394 sequences | 0 interactions | 304 species | 1 structure

- Summary
- Domain organisation
- Clan
- Alignments
- HMM logo
- Trees
- Curation & model
- Species
- Interactions
- Structures

Jump to... enter ID/acc Go

Summary: Domain of unkown function (DUF1775)

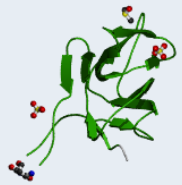
Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

Wikipedia: Domain of unknown function | Pfam | InterPro

"DUF" families are annotated with the Domain of unknown function Wikipedia article. This is a general article, with no specific information about individual Pfam DUFs. If you have information about this particular DUF, please let us know using the "Add annotation" button below.

Domain of unkown function (DUF1775) Provide feedback

Domain found in bacteria with undetermined function. Its structure has been determined and is an immunoglobulin-like fold.



PDB entry 3ESM: Crystal structure of an uncharacterized protein from Nocardia farcinica reveals an immunoglobulin-like fold

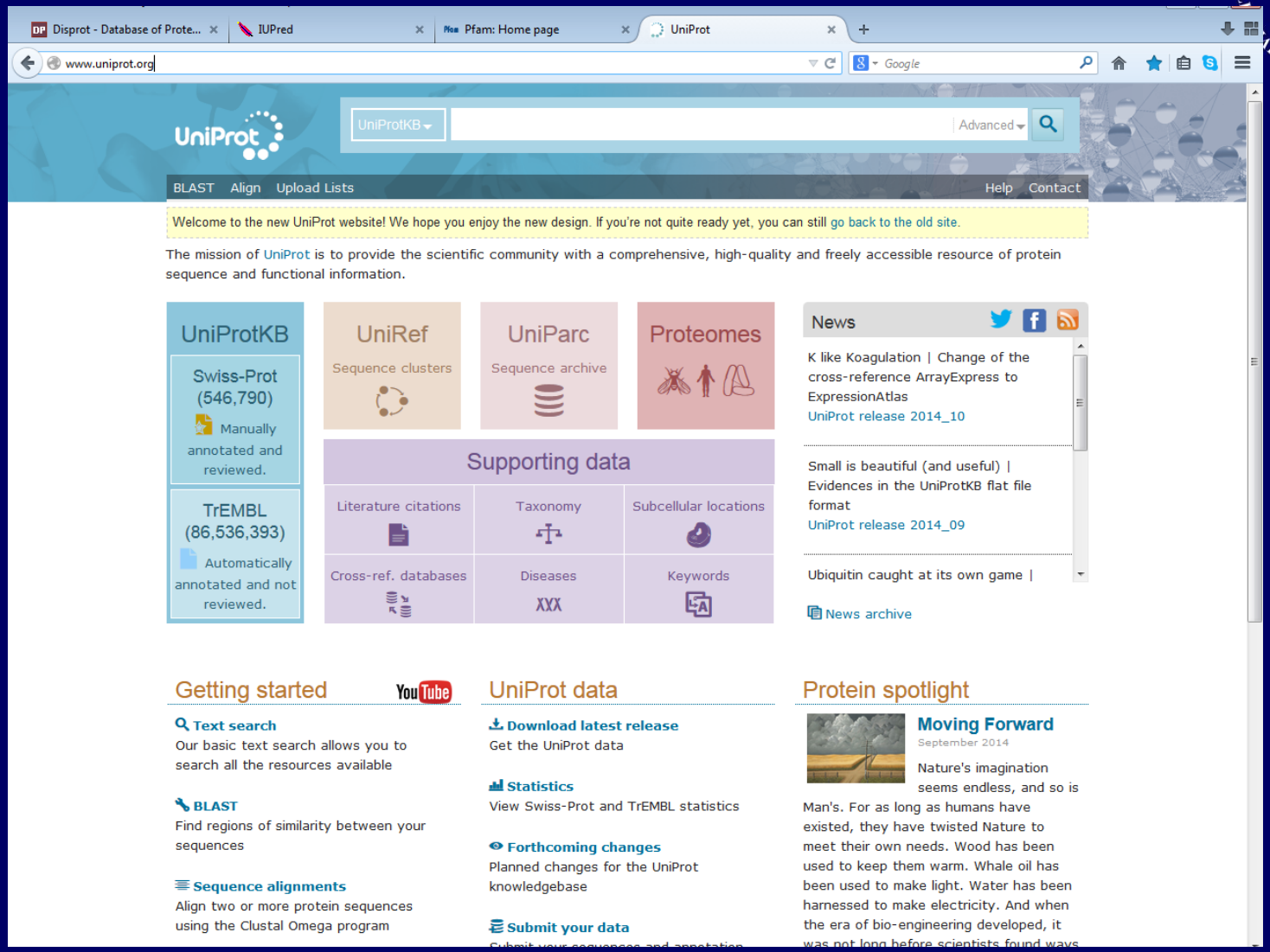
Comments or questions on the site? Send a mail to pfam-help@ebi.ac.uk.
European Molecular Biology Laboratory

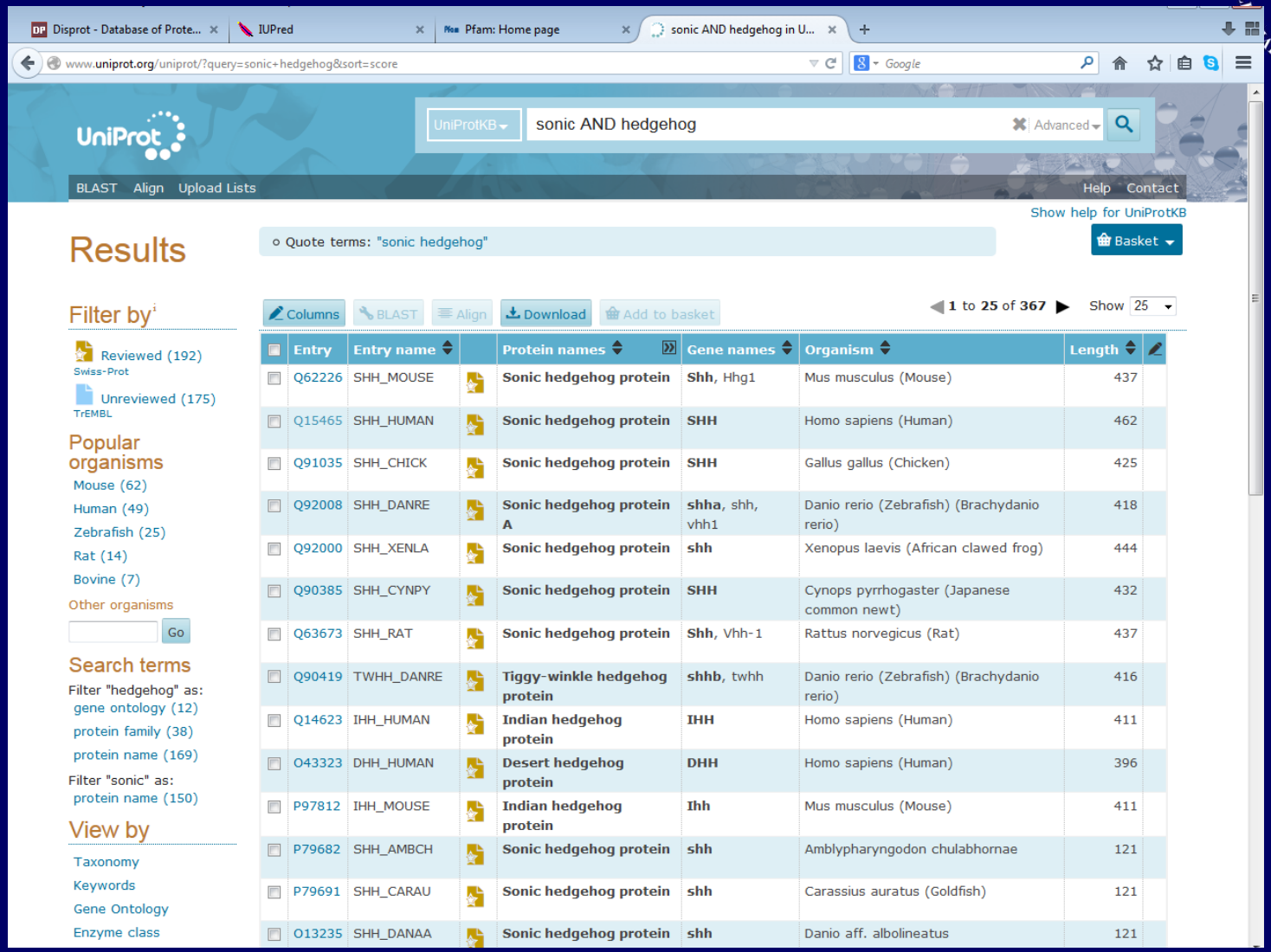
Mit kereshetünk még homológia alapján

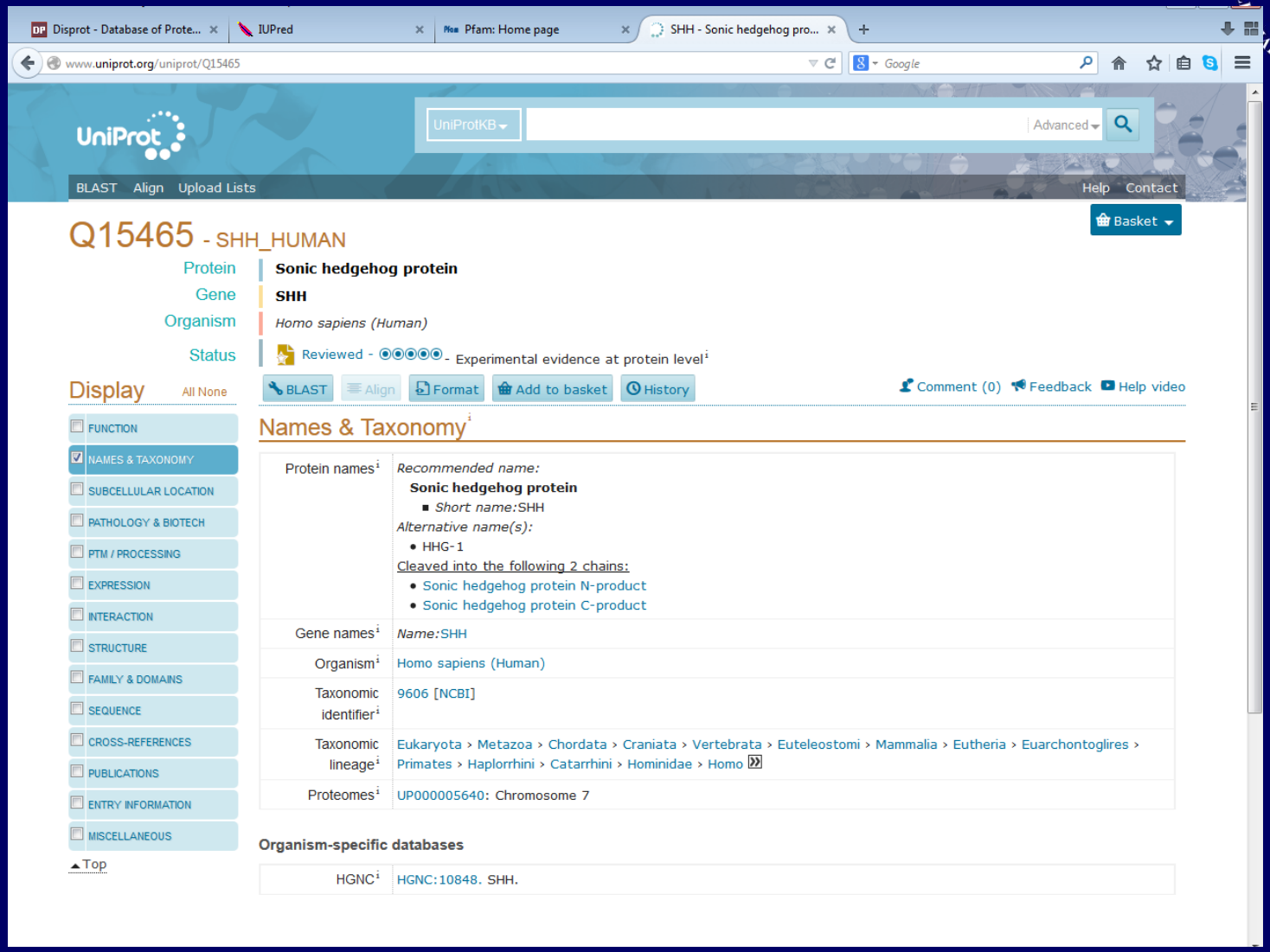
- Poszt-transzlációs módosítások
  - Glikozilálás
  - Palmitálás, mirisztoilálás
  - Egyébb módosítások
- Foszforilálás
- Ligandok kötése
- Aktív helyek azonosítása
- Hasítóhelyek azonosítása

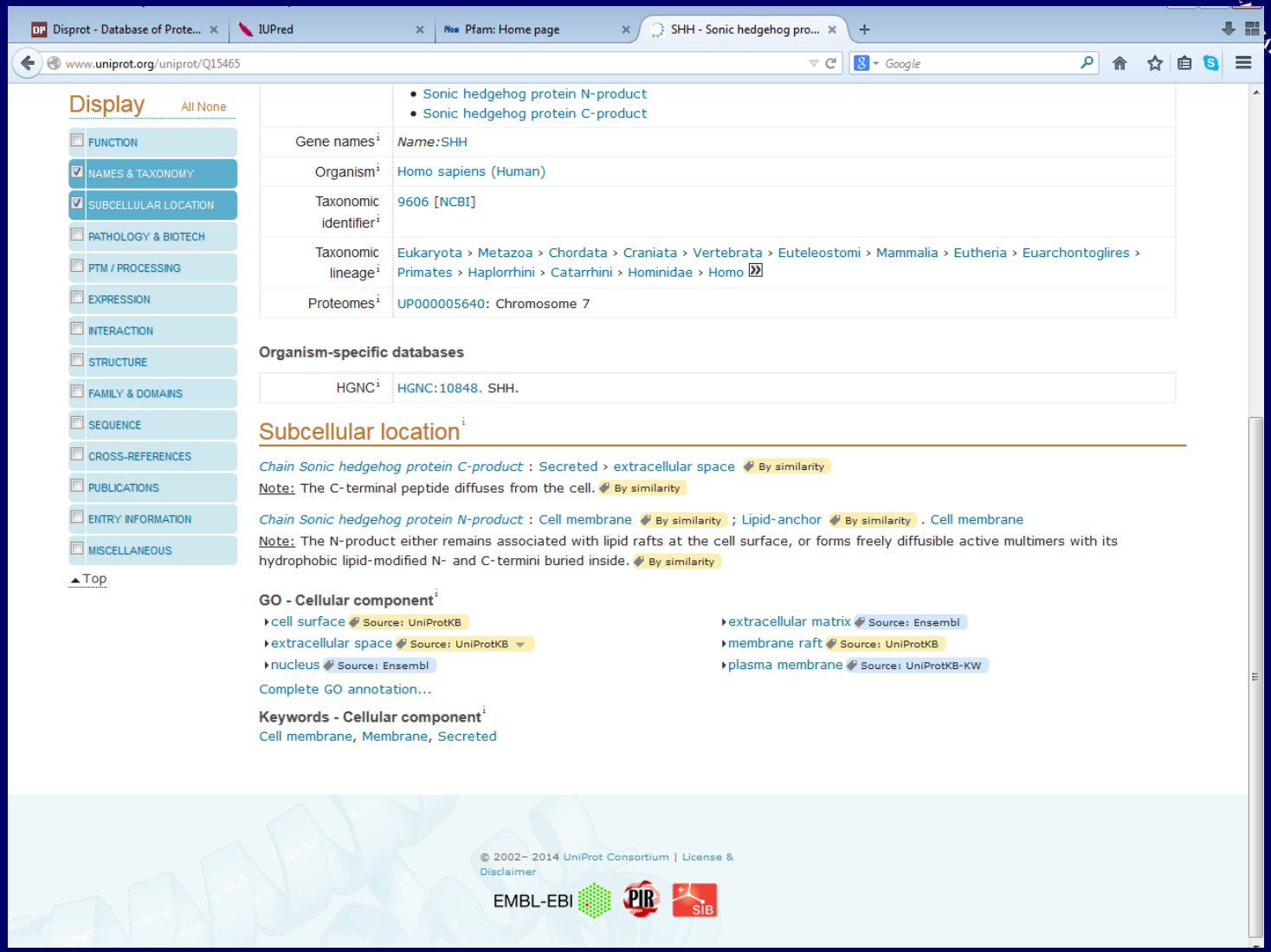
A módszerek

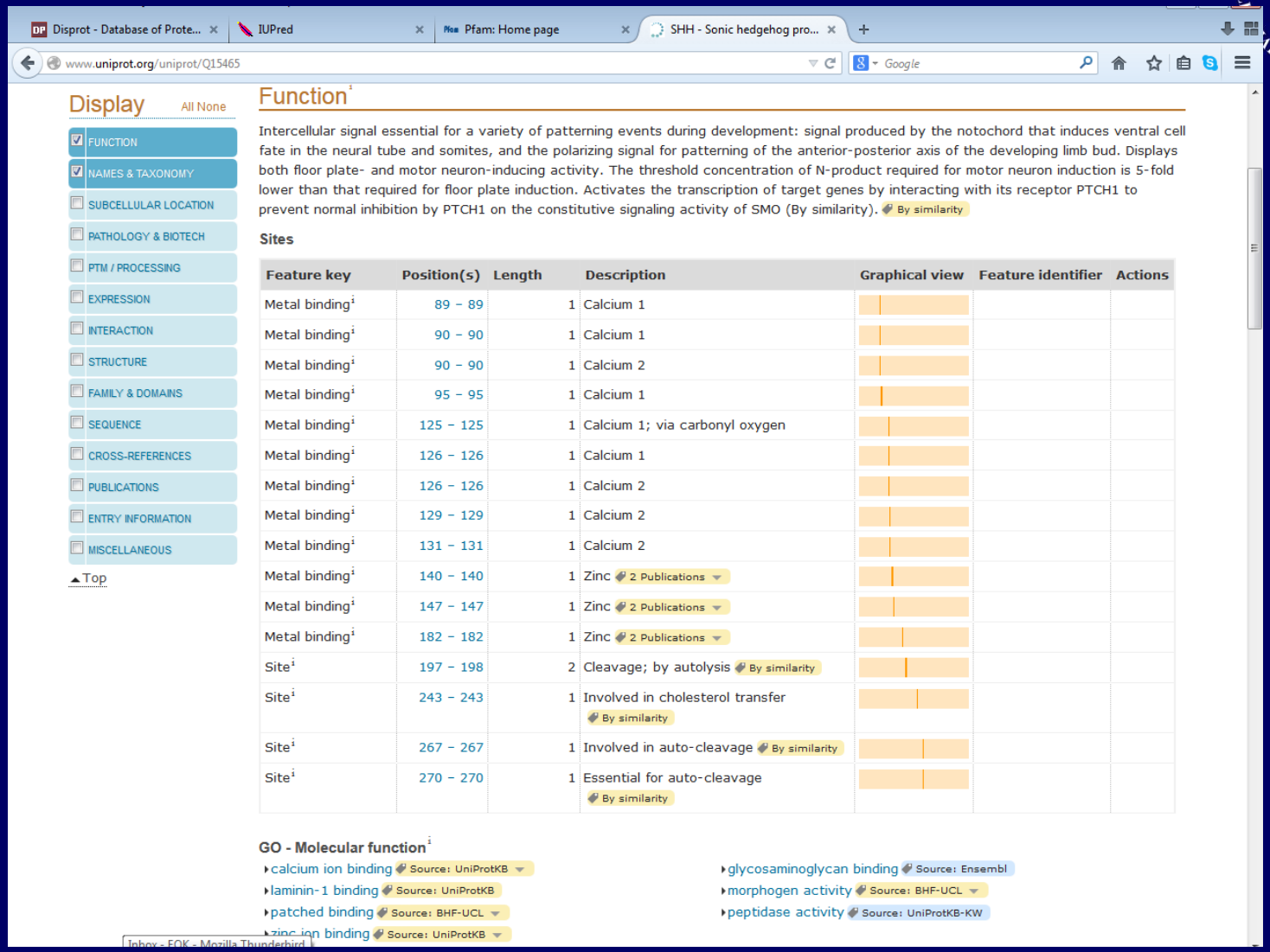
- Kisérletes adatok gyüjtése
- Statisztikus modell építése
- A modell tanítása a kisérletes adatbázis alapján
- A modell alkalmazása az adott szekvenciára
- Másik lehetőség: szekvenciaillesztés az ismert szekvenciák ellenében

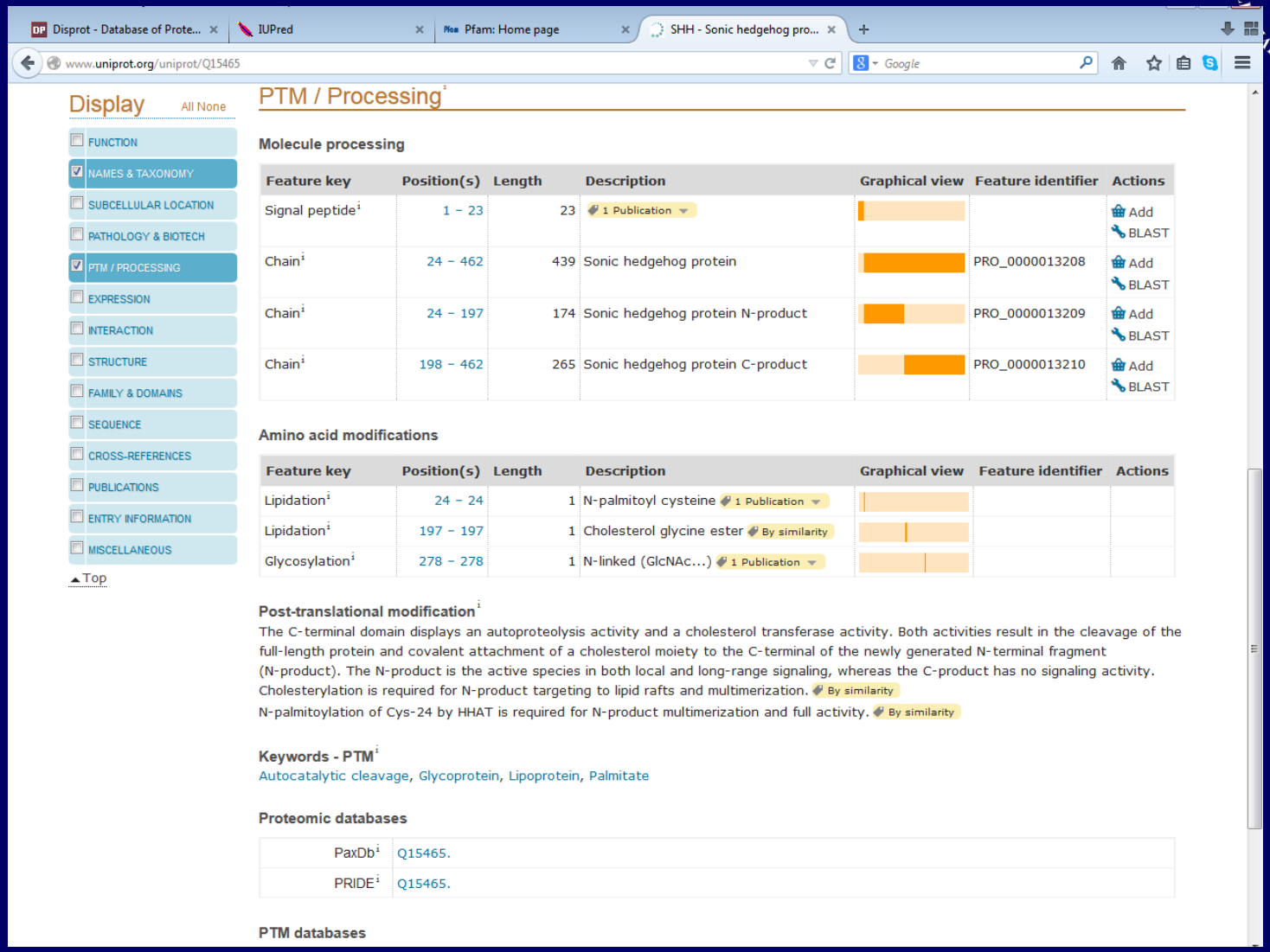

Disprot - Database of Prote...
IUPred
Pfam: Home page
UniProt
www.uniprot.org
UniProt
UniProtKB
Advanced
BLAST Align Upload Lists
Help Contact
Welcome to the new UniProt website! We hope you enjoy the new design. If you're not quite ready yet, you can still go back to the old site.
The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.
UniProtKB
Swiss-Prot (546,790)
Manually annotated and reviewed.
TrEMBL (86,536,393)
Automatically annotated and not reviewed.
UniRef
Sequence clusters
UniParc
Sequence archive
Proteomes
Supporting data
Literature citations
Taxonomy
Subcellular locations
Cross-ref. databases
Diseases
XXX
Keywords
News
K like Koagulation | Change of the cross-reference ArrayExpress to ExpressionAtlas
UniProt release 2014_10
Small is beautiful (and useful) | Evidences in the UniProtKB flat file format
UniProt release 2014_09
Ubiquitin caught at its own game |
News archive
Getting started
YouTube
Text search
Our basic text search allows you to search all the resources available
BLAST
Find regions of similarity between your sequences
Sequence alignments
Align two or more protein sequences using the Clustal Omega program
UniProt data
Download latest release
Get the UniProt data
Statistics
View Swiss-Prot and TrEMBL statistics
Forthcoming changes
Planned changes for the UniProt knowledgebase
Submit your data
Protein spotlight
Moving Forward
September 2014
Nature's imagination seems endless, and so is Man's. For as long as humans have existed, they have twisted Nature to meet their own needs. Wood has been used to keep them warm. Whale oil has been used to make light. Water has been harnessed to make electricity. And when the era of bio-engineering developed, it

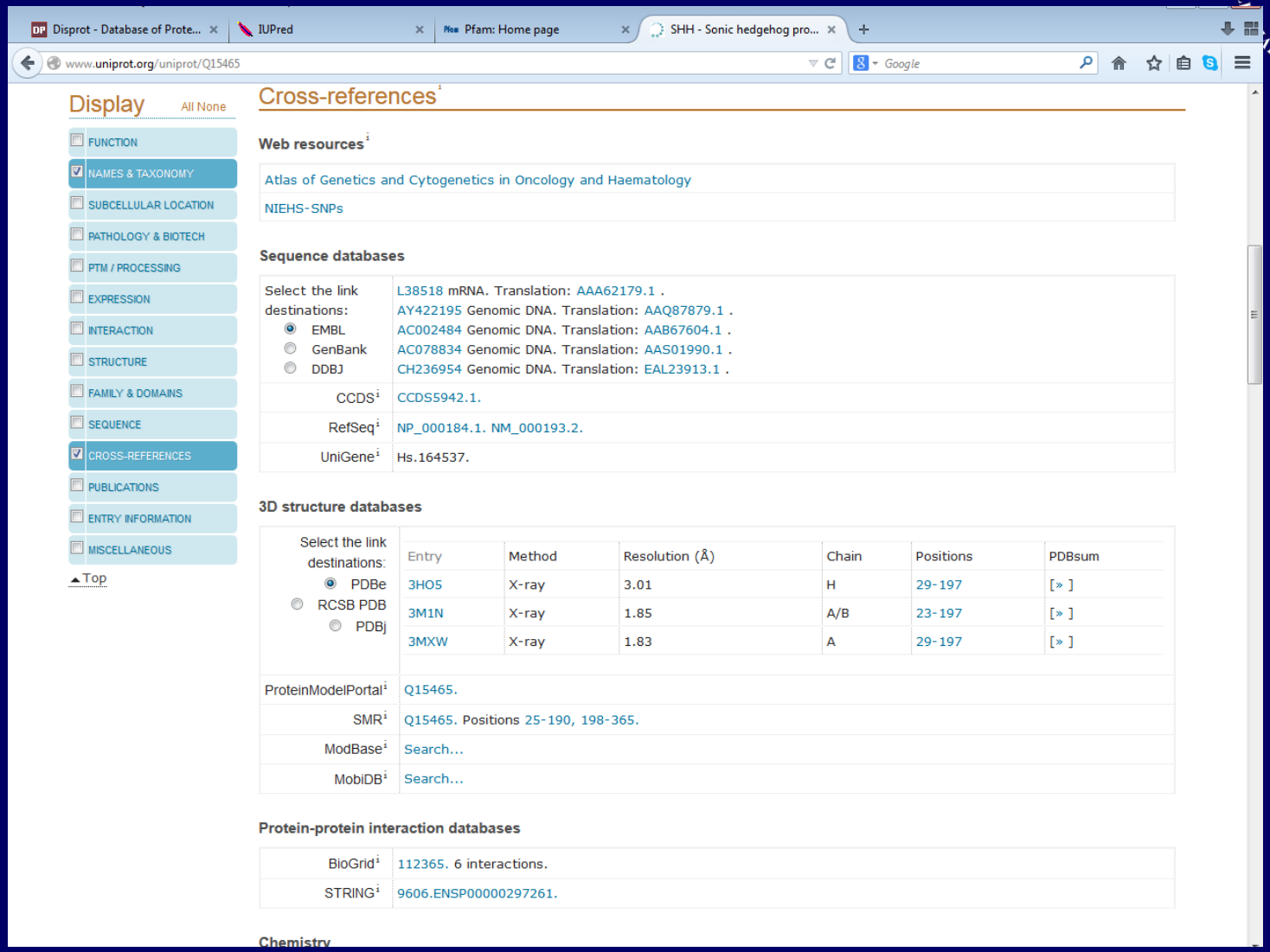

Disprot - Database of Prote...
IUPred
Pfam: Home page
sonic AND hedgehog in U...
www.uniprot.org/uniprot/?query=sonic+hedgehog&sort=score
UniProt
UniProtKB
sonic AND hedgehog
Advanced
BLAST Align Upload Lists
Help Contact
Show help for UniProtKB
Basket
Results
Quote terms: "sonic hedgehog"
Filter by
Reviewed (192)
Swiss-Prot
Unreviewed (175)
TrEMBL
Popular organisms
Mouse (62)
Human (49)
Zebrafish (25)
Rat (14)
Bovine (7)
Other organisms
Go
Search terms
Filter "hedgehog" as:
gene ontology (12)
protein family (38)
protein name (169)
Filter "sonic" as:
protein name (150)
View by
Taxonomy
Keywords
Gene Ontology
Enzyme class
Columns
BLAST
Align
Download
Add to basket
1 to 25 of 367
Show 25
Entry | Entry name | Protein names | Gene names | Organism | Length
Q62226 SHH_MOUSE Sonic hedgehog protein Shh, Hhg1 Mus musculus (Mouse) 437
Q15465 SHH_HUMAN Sonic hedgehog protein SHH Homo sapiens (Human) 462
Q91035 SHH_CHICK Sonic hedgehog protein SHH Gallus gallus (Chicken) 425
Q92008 SHH_DANRE Sonic hedgehog protein A shha, shh, vhh1 Danio rerio (Zebrafish) (Brachydanio rerio) 418
Q92000 SHH_XENLA Sonic hedgehog protein shh Xenopus laevis (African clawed frog) 444
Q90385 SHH_CYNPY Sonic hedgehog protein SHH Cynops pyrrhogaster (Japanese common newt) 432
Q63673 SHH_RAT Sonic hedgehog protein Shh, Vhh-1 Rattus norvegicus (Rat) 437
Q90419 TWHH_DANRE Tiggy-winkle hedgehog protein shhb, twhh Danio rerio (Zebrafish) (Brachydanio rerio) 416
Q14623 IHH_HUMAN Indian hedgehog protein IHH Homo sapiens (Human) 411
O43323 DHH_HUMAN Desert hedgehog protein DHH Homo sapiens (Human) 396
P97812 IHH_MOUSE Indian hedgehog protein Ihh Mus musculus (Mouse) 411
P79682 SHH_AMBCH Sonic hedgehog protein shh Amblypharyngodon chulabhornae 121
P79691 SHH_CARAU Sonic hedgehog protein shh Carassius auratus (Goldfish) 121
O13235 SHH_DANAA Sonic hedgehog protein shh Danio aff. albolineatus 121

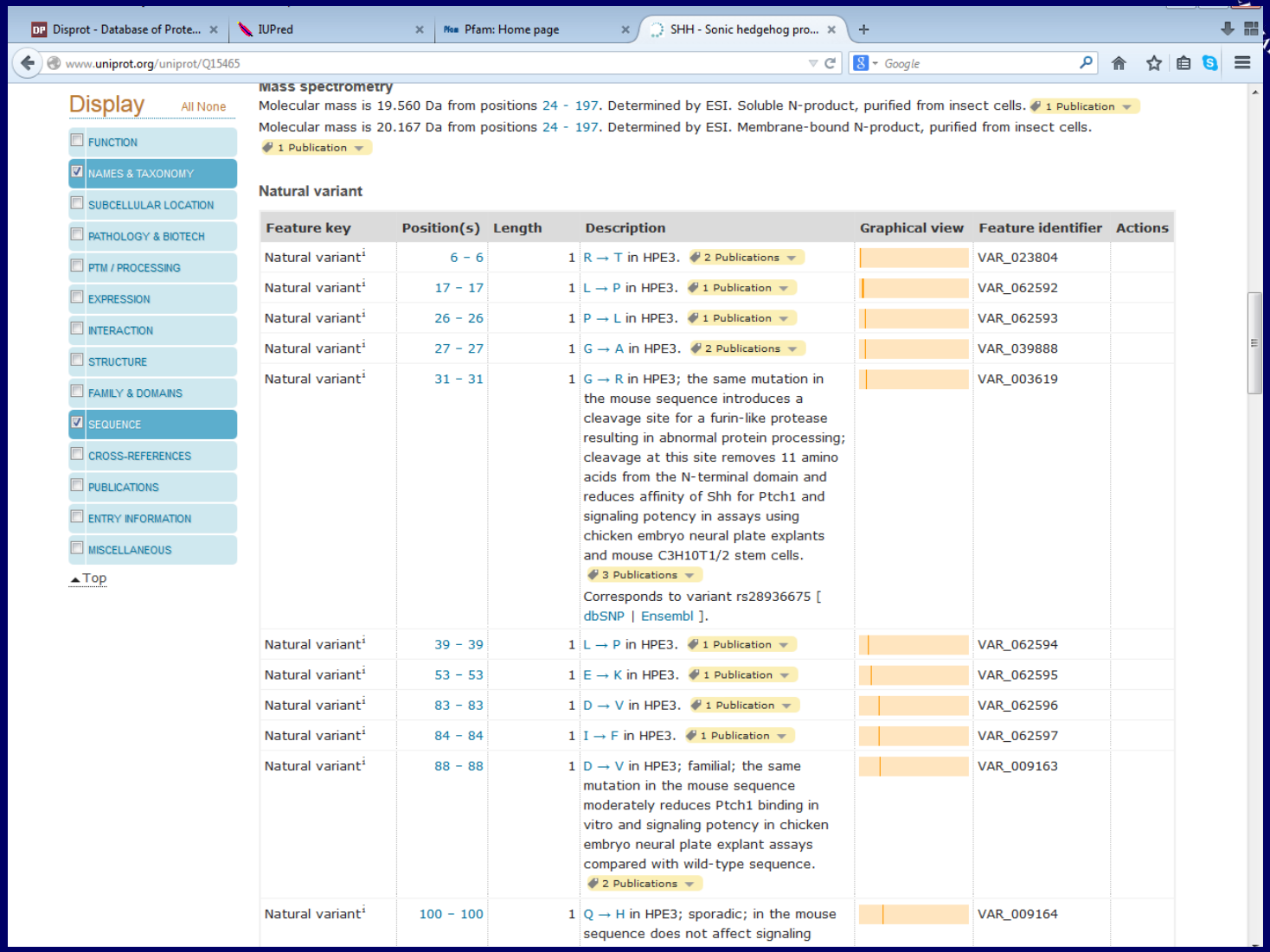

Disprot - Database of Prote...
IUPred
Pfam: Home page
SHH - Sonic hedgehog pro...
www.uniprot.org/uniprot/Q15465
UniProt
UniProtKB
Advanced
BLAST Align Upload Lists
Help Contact
Basket
Q15465 - SHH_HUMAN
Protein Sonic hedgehog protein
Gene SHH
Organism Homo sapiens (Human)
Status Reviewed - ●●●●● - Experimental evidence at protein level
Display All None
FUNCTION
NAMES & TAXONOMY
SUBCELLULAR LOCATION
PATHOLOGY & BIOTECH
PTM / PROCESSING
EXPRESSION
INTERACTION
STRUCTURE
FAMILY & DOMAINS
SEQUENCE
CROSS-REFERENCES
PUBLICATIONS
ENTRY INFORMATION
MISCELLANEOUS
Top
BLAST Align Format Add to basket History
Comment (0) Feedback Help video
Names & Taxonomy
Protein names
Recommended name:
Sonic hedgehog protein
Short name:SHH
Alternative name(s):
HHG-1
Cleaved into the following 2 chains:
Sonic hedgehog protein N-product
Sonic hedgehog protein C-product
Gene names
Name:SHH
Organism
Homo sapiens (Human)
Taxonomic identifier
9606 [NCBI]
Taxonomic lineage
Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo
Proteomes
UP000005640: Chromosome 7
Organism-specific databases
HGNC
HGNC:10848. SHH.


Disprot - Database of Prote...
IUPred
Pfam: Home page
SHH - Sonic hedgehog pro...
www.uniprot.org/uniprot/Q15465
Display
All None
FUNCTION
NAMES & TAXONOMY
SUBCELLULAR LOCATION
PATHOLOGY & BIOTECH
PTM / PROCESSING
EXPRESSION
INTERACTION
STRUCTURE
FAMILY & DOMAINS
SEQUENCE
CROSS-REFERENCES
PUBLICATIONS
ENTRY INFORMATION
MISCELLANEOUS
Top
Sonic hedgehog protein N-product
Sonic hedgehog protein C-product
Gene names
Name:SHH
Organism
Homo sapiens (Human)
Taxonomic identifier
9606 [NCBI]
Taxonomic lineage
Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo
Proteomes
UP000005640: Chromosome 7
Organism-specific databases
HGNC
HGNC:10848. SHH.
Subcellular location
Chain Sonic hedgehog protein C-product : Secreted › extracellular space By similarity
Note: The C-terminal peptide diffuses from the cell. By similarity
Chain Sonic hedgehog protein N-product : Cell membrane By similarity ; Lipid-anchor By similarity . Cell membrane
Note: The N-product either remains associated with lipid rafts at the cell surface, or forms freely diffusible active multimers with its hydrophobic lipid-modified N- and C-termini buried inside. By similarity
GO - Cellular component
cell surface Source: UniProtKB
extracellular space Source: UniProtKB
nucleus Source: Ensembl
extracellular matrix Source: Ensembl
membrane raft Source: UniProtKB
plasma membrane Source: UniProtKB-KW
Complete GO annotation...
Keywords - Cellular component
Cell membrane, Membrane, Secreted
© 2002– 2014 UniProt Consortium | License & Disclaimer
EMBL-EBI
PIR
SIB


Disprot - Database of Prote...
IUPred
Pfam: Home page
SHH - Sonic hedgehog pro...
www.uniprot.org/uniprot/Q15465
Google
Display
All None
FUNCTION
NAMES & TAXONOMY
SUBCELLULAR LOCATION
PATHOLOGY & BIOTECH
PTM / PROCESSING
EXPRESSION
INTERACTION
STRUCTURE
FAMILY & DOMAINS
SEQUENCE
CROSS-REFERENCES
PUBLICATIONS
ENTRY INFORMATION
MISCELLANEOUS
Top
Function
Intercellular signal essential for a variety of patterning events during development: signal produced by the notochord that induces ventral cell fate in the neural tube and somites, and the polarizing signal for patterning of the anterior-posterior axis of the developing limb bud. Displays both floor plate- and motor neuron-inducing activity. The threshold concentration of N-product required for motor neuron induction is 5-fold lower than that required for floor plate induction. Activates the transcription of target genes by interacting with its receptor PTCH1 to prevent normal inhibition by PTCH1 on the constitutive signaling activity of SMO (By similarity). By similarity
Sites
Feature key | Position(s) | Length | Description | Graphical view | Feature identifier | Actions
Metal binding 89 – 89 1 Calcium 1
Metal binding 90 – 90 1 Calcium 1
Metal binding 90 – 90 1 Calcium 2
Metal binding 95 – 95 1 Calcium 1
Metal binding 125 – 125 1 Calcium 1; via carbonyl oxygen
Metal binding 126 – 126 1 Calcium 1
Metal binding 126 – 126 1 Calcium 2
Metal binding 129 – 129 1 Calcium 2
Metal binding 131 – 131 1 Calcium 2
Metal binding 140 – 140 1 Zinc 2 Publications
Metal binding 147 – 147 1 Zinc 2 Publications
Metal binding 182 – 182 1 Zinc 2 Publications
Site 197 – 198 2 Cleavage; by autolysis By similarity
Site 243 – 243 1 Involved in cholesterol transfer By similarity
Site 267 – 267 1 Involved in auto-cleavage By similarity
Site 270 – 270 1 Essential for auto-cleavage By similarity
GO - Molecular function
calcium ion binding Source: UniProtKB
laminin-1 binding Source: UniProtKB
patched binding Source: BHF-UCL
zinc ion binding Source: UniProtKB
glycosaminoglycan binding Source: Ensembl
morphogen activity Source: BHF-UCL
peptidase activity Source: UniProtKB-KW
Inbox - EOK - Mozilla Thunderbird


Disprot - Database of Prote...
IUPred
Pfam: Home page
SHH - Sonic hedgehog pro...
www.uniprot.org/uniprot/Q15465
Google
Display All None
FUNCTION
NAMES & TAXONOMY
SUBCELLULAR LOCATION
PATHOLOGY & BIOTECH
PTM / PROCESSING
EXPRESSION
INTERACTION
STRUCTURE
FAMILY & DOMAINS
SEQUENCE
CROSS-REFERENCES
PUBLICATIONS
ENTRY INFORMATION
MISCELLANEOUS
Top
PTM / Processing
Molecule processing
Feature key | Position(s) | Length | Description | Graphical view | Feature identifier | Actions
Signal peptide | 1 – 23 | 23 | 1 Publication | Add BLAST
Chain | 24 – 462 | 439 | Sonic hedgehog protein | PRO_0000013208 | Add BLAST
Chain | 24 – 197 | 174 | Sonic hedgehog protein N-product | PRO_0000013209 | Add BLAST
Chain | 198 – 462 | 265 | Sonic hedgehog protein C-product | PRO_0000013210 | Add BLAST
Amino acid modifications
Feature key | Position(s) | Length | Description | Graphical view | Feature identifier | Actions
Lipidation | 24 – 24 | 1 | N-palmitoyl cysteine 1 Publication
Lipidation | 197 – 197 | 1 | Cholesterol glycine ester By similarity
Glycosylation | 278 – 278 | 1 | N-linked (GlcNAc...) 1 Publication
Post-translational modification
The C-terminal domain displays an autoproteolysis activity and a cholesterol transferase activity. Both activities result in the cleavage of the full-length protein and covalent attachment of a cholesterol moiety to the C-terminal of the newly generated N-terminal fragment (N-product). The N-product is the active species in both local and long-range signaling, whereas the C-product has no signaling activity. Cholesterylation is required for N-product targeting to lipid rafts and multimerization. By similarity
N-palmitoylation of Cys-24 by HHAT is required for N-product multimerization and full activity. By similarity
Keywords - PTM
Autocatalytic cleavage, Glycoprotein, Lipoprotein, Palmitate
Proteomic databases
PaxDb Q15465.
PRIDE Q15465.
PTM databases


Disprot - Database of Prote...
IUPred
Pfam: Home page
SHH - Sonic hedgehog pro...
www.uniprot.org/uniprot/Q15465
Google
Display All None
FUNCTION
NAMES & TAXONOMY
SUBCELLULAR LOCATION
PATHOLOGY & BIOTECH
PTM / PROCESSING
EXPRESSION
INTERACTION
STRUCTURE
FAMILY & DOMAINS
SEQUENCE
CROSS-REFERENCES
PUBLICATIONS
ENTRY INFORMATION
MISCELLANEOUS
Top
Cross-references
Web resources
Atlas of Genetics and Cytogenetics in Oncology and Haematology
NIEHS-SNPs
Sequence databases
Select the link destinations:
EMBL
GenBank
DDBJ
L38518 mRNA. Translation: AAA62179.1 .
AY422195 Genomic DNA. Translation: AAQ87879.1 .
AC002484 Genomic DNA. Translation: AAB67604.1 .
AC078834 Genomic DNA. Translation: AAS01990.1 .
CH236954 Genomic DNA. Translation: EAL23913.1 .
CCDS
CCDS5942.1.
RefSeq
NP_000184.1. NM_000193.2.
UniGene
Hs.164537.
3D structure databases
Select the link destinations:
PDBe
RCSB PDB
PDBj
Entry Method Resolution (Å) Chain Positions PDBsum
3HO5 X-ray 3.01 H 29-197 [»]
3M1N X-ray 1.85 A/B 23-197 [»]
3MXW X-ray 1.83 A 29-197 [»]
ProteinModelPortal
Q15465.
SMR
Q15465. Positions 25-190, 198-365.
ModBase
Search...
MobiDB
Search...
Protein-protein interaction databases
BioGrid
112365. 6 interactions.
STRING
9606.ENSP00000297261.
Chemistry



Display All None

- FUNCTION
- NAMES & TAXONOMY
- SUBCELLULAR LOCATION
- PATHOLOGY & BIOTECH
- PTM / PROCESSING
- EXPRESSION
- INTERACTION
- STRUCTURE
- FAMILY & DOMAINS
- SEQUENCE
- CROSS-REFERENCES
- PUBLICATIONS
- ENTRY INFORMATION
- MISCELLANEOUS

▲Top

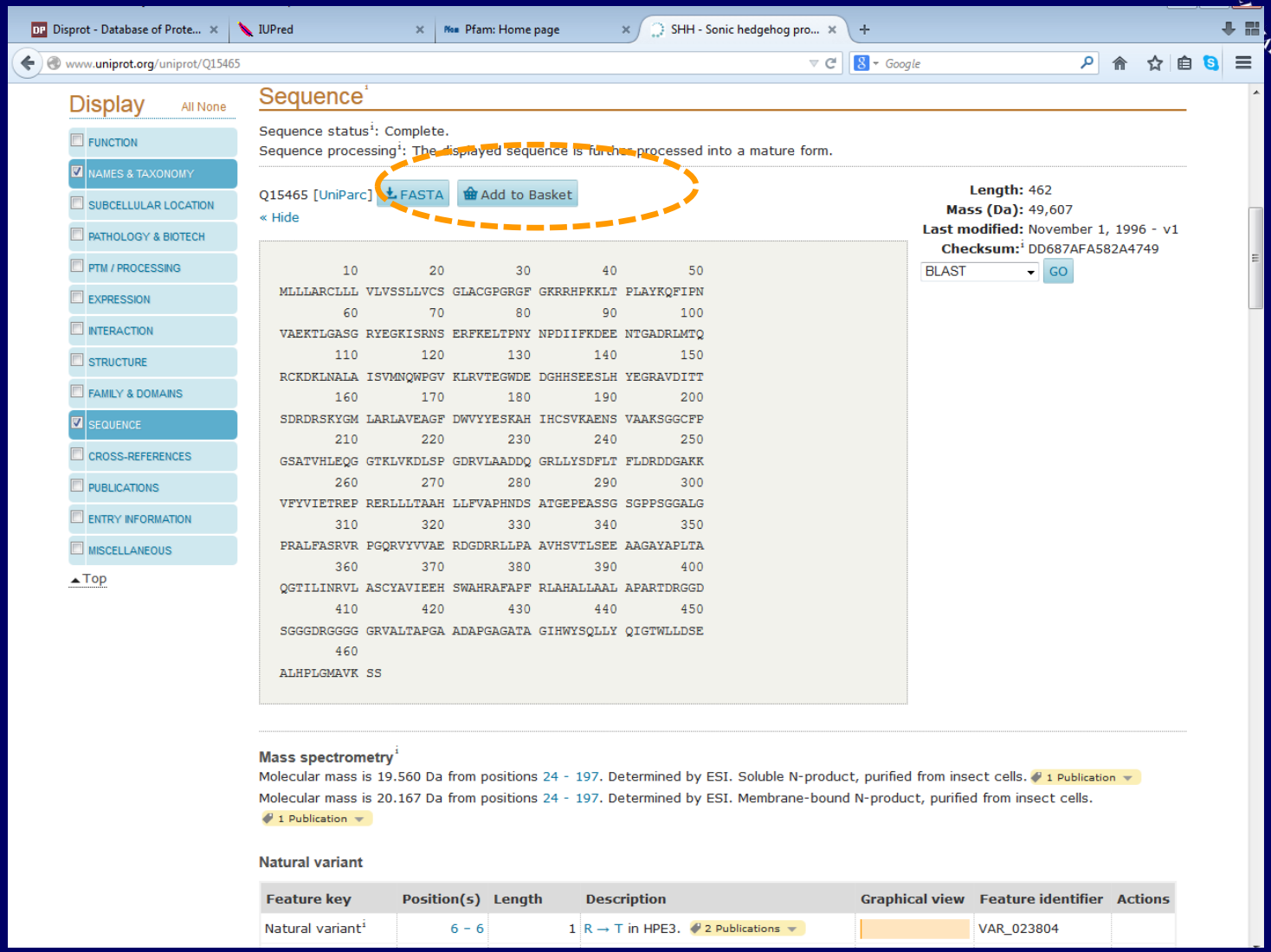
Mass spectrometry

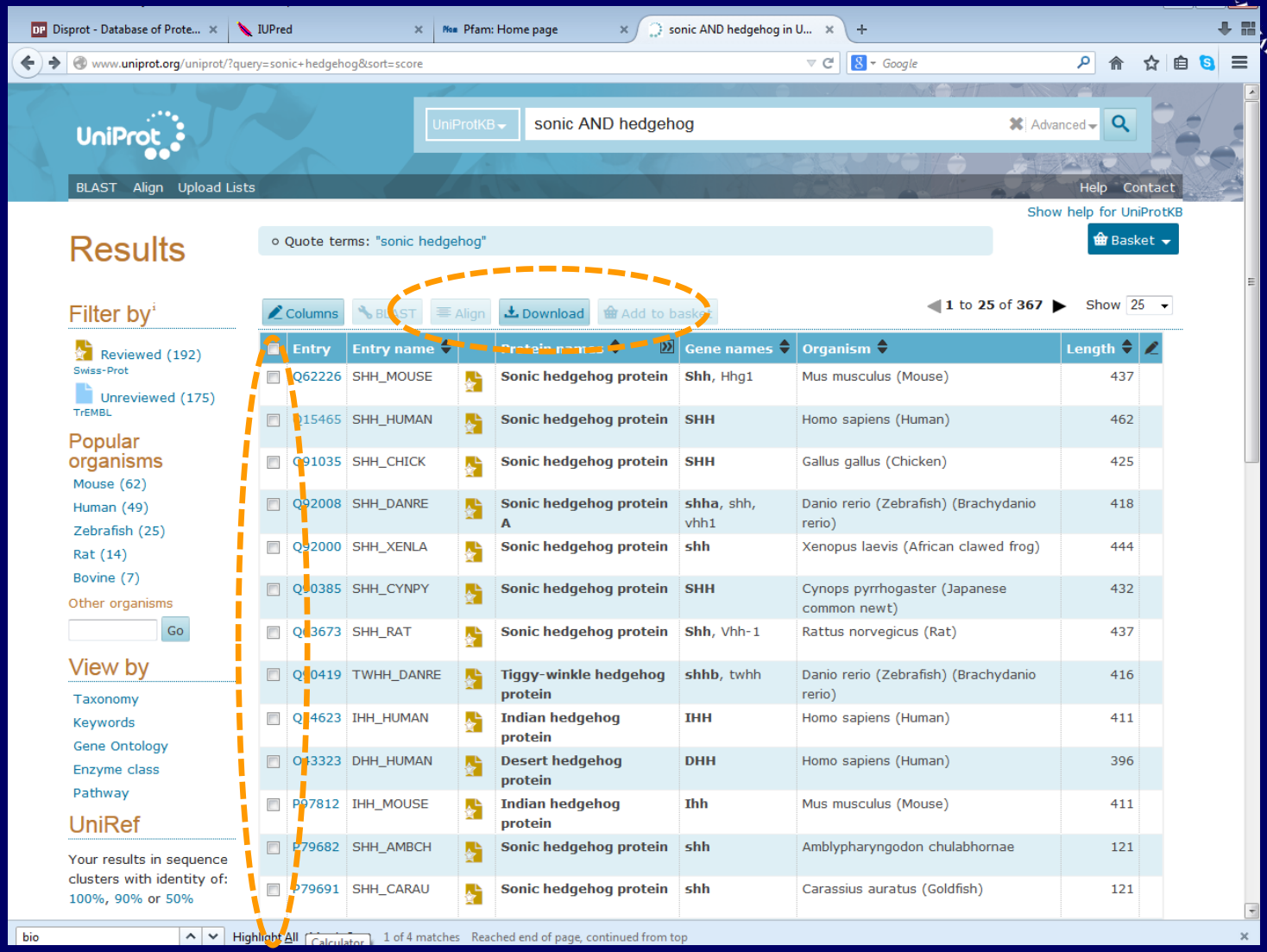
Molecular mass is 19.560 Da from positions 24 - 197. Determined by ESI. Soluble N-product, purified from insect cells. 1 Publication

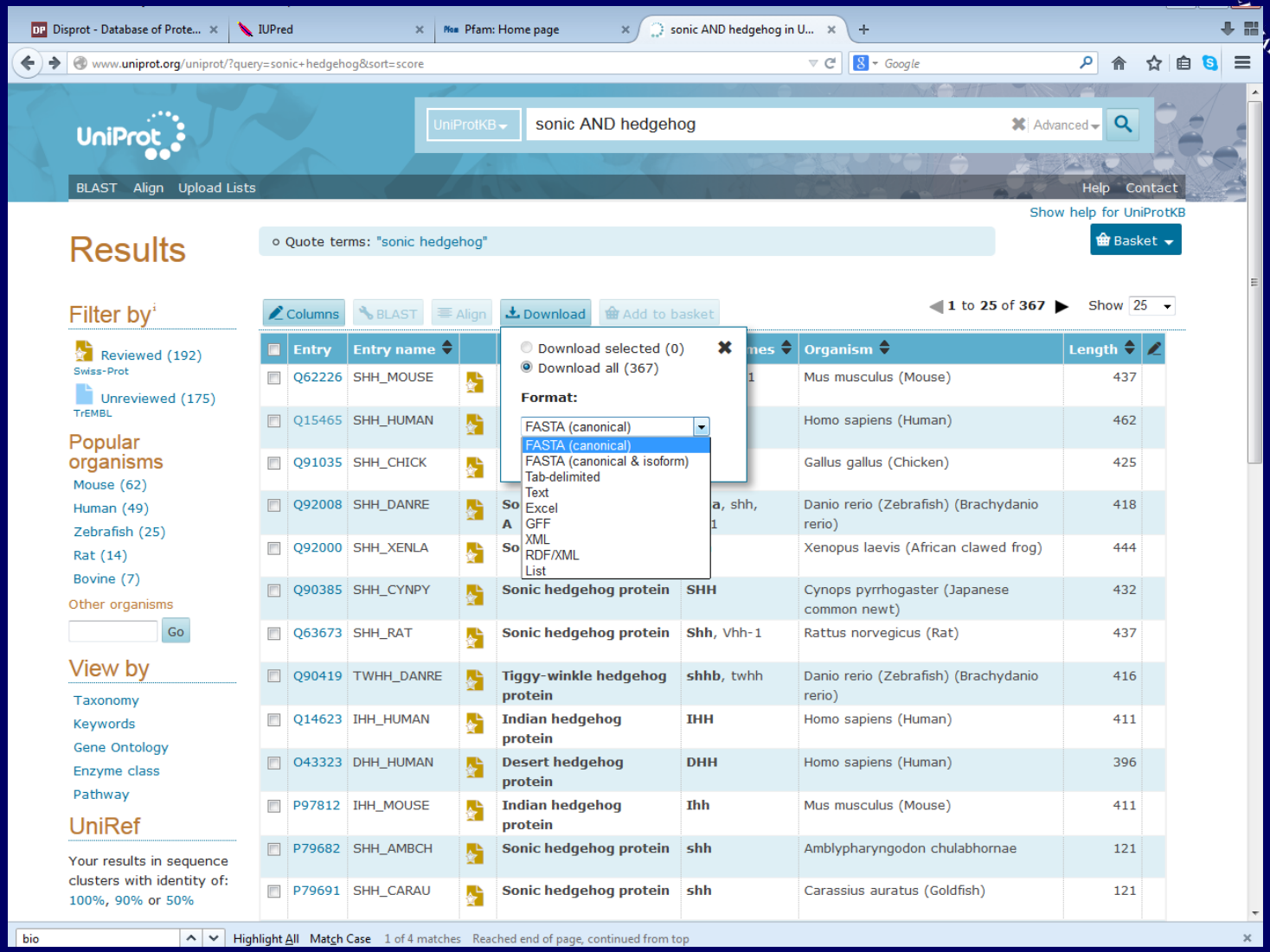
Molecular mass is 20.167 Da from positions 24 - 197. Determined by ESI. Membrane-bound N-product, purified from insect cells. 1 Publication

Natural variant

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Natural variant	6 – 6	1	R → T in HPE3. 2 Publications		VAR_023804	
Natural variant	17 – 17	1	L → P in HPE3. 1 Publication		VAR_062592	
Natural variant	26 – 26	1	P → L in HPE3. 1 Publication		VAR_062593	
Natural variant	27 – 27	1	G → A in HPE3. 2 Publications		VAR_039888	
Natural variant	31 – 31	1	G → R in HPE3; the same mutation in the mouse sequence introduces a cleavage site for a furin-like protease resulting in abnormal protein processing; cleavage at this site removes 11 amino acids from the N-terminal domain and reduces affinity of Shh for Ptch1 and signaling potency in assays using chicken embryo neural plate explants and mouse C3H10T1/2 stem cells. 3 Publications Corresponds to variant rs28936675 [dbSNP | Ensembl].		VAR_003619	
Natural variant	39 – 39	1	L → P in HPE3. 1 Publication		VAR_062594	
Natural variant	53 – 53	1	E → K in HPE3. 1 Publication		VAR_062595	
Natural variant	83 – 83	1	D → V in HPE3. 1 Publication		VAR_062596	
Natural variant	84 – 84	1	I → F in HPE3. 1 Publication		VAR_062597	
Natural variant	88 – 88	1	D → V in HPE3; familial; the same mutation in the mouse sequence moderately reduces Ptch1 binding in vitro and signaling potency in chicken embryo neural plate explant assays compared with wild-type sequence. 2 Publications		VAR_009163	
Natural variant	100 – 100	1	Q → H in HPE3; sporadic; in the mouse sequence does not affect signaling		VAR_009164	


Disprot - Database of Prote...
IUPred
Pfam: Home page
SHH - Sonic hedgehog pro...
www.uniprot.org/uniprot/Q15465
Display
All None
FUNCTION
NAMES & TAXONOMY
SUBCELLULAR LOCATION
PATHOLOGY & BIOTECH
PTM / PROCESSING
EXPRESSION
INTERACTION
STRUCTURE
FAMILY & DOMAINS
SEQUENCE
CROSS-REFERENCES
PUBLICATIONS
ENTRY INFORMATION
MISCELLANEOUS
Top
Sequence
Sequence status: Complete.
Sequence processing: The displayed sequence is further processed into a mature form.
Q15465 [UniParc]
FASTA
Add to Basket
« Hide
Length: 462
Mass (Da): 49,607
Last modified: November 1, 1996 - v1
Checksum: DD687AFA582A4749
BLAST
GO
10 20 30 40 50
MLLLARCLLL VLVSSLLVCS GLACGPGRGF GKRRHPKKLT PLAYKQFIPN
60 70 80 90 100
VAEKTLGASG RYEGKISRNS ERFKELTPNY NPDIIFKDEE NTGADRLMTQ
110 120 130 140 150
RCKDKLNALA ISVMNQWPGV KLRVTEGWDE DGHHSEESLH YEGRAVDITT
160 170 180 190 200
SDRDRSKYGM LARLAVEAGF DWVYYESKAH IHCSVKAENS VAAKSGGCFP
210 220 230 240 250
GSATVHLEQG GTKLVKDLSP GDRVLAADDQ GRLLYSDFLT FLDRDDGAKK
260 270 280 290 300
VFYVIETREP RERLLLTAAH LLFVAPHNDS ATGEPEASSG SGPPSGGALG
310 320 330 340 350
PRALFASRVR PGQRVYVVAE RDGDRRLLPA AVHSVTLSEE AAGAYAPLTA
360 370 380 390 400
QGTILINRVL ASCYAVIEEH SWAHRAFAPF RLAHALLAAL APARTDRGGD
410 420 430 440 450
SGGGDRGGGG GRVALTAPGA ADAPGAGATA GIHWYSQLLY QIGTWLLDSE
460
ALHPLGMAVK SS
Mass spectrometry
Molecular mass is 19.560 Da from positions 24 - 197. Determined by ESI. Soluble N-product, purified from insect cells. 1 Publication
Molecular mass is 20.167 Da from positions 24 - 197. Determined by ESI. Membrane-bound N-product, purified from insect cells.
1 Publication
Natural variant
Feature key | Position(s) | Length | Description | Graphical view | Feature identifier | Actions
Natural variant | 6 – 6 | 1 | R → T in HPE3. 2 Publications | VAR_023804


Disprot - Database of Prote...
IUPred
Pfam: Home page
sonic AND hedgehog in U...
www.uniprot.org/uniprot/?query=sonic+hedgehog&sort=score
UniProt
UniProtKB
sonic AND hedgehog
Advanced
BLAST Align Upload Lists
Help Contact
Show help for UniProtKB
Basket
Results
Quote terms: "sonic hedgehog"
Filter by
Reviewed (192)
Swiss-Prot
Unreviewed (175)
TrEMBL
Popular organisms
Mouse (62)
Human (49)
Zebrafish (25)
Rat (14)
Bovine (7)
Other organisms
Go
View by
Taxonomy
Keywords
Gene Ontology
Enzyme class
Pathway
UniRef
Your results in sequence clusters with identity of:
100%, 90% or 50%
Columns
Download
Add to basket
1 to 25 of 367
Show 25
Entry
Entry name
Protein names
Gene names
Organism
Length
Q62226 SHH_MOUSE Sonic hedgehog protein Shh, Hhg1 Mus musculus (Mouse) 437
Q15465 SHH_HUMAN Sonic hedgehog protein SHH Homo sapiens (Human) 462
Q91035 SHH_CHICK Sonic hedgehog protein SHH Gallus gallus (Chicken) 425
Q92008 SHH_DANRE Sonic hedgehog protein A shha, shh, vhh1 Danio rerio (Zebrafish) (Brachydanio rerio) 418
Q92000 SHH_XENLA Sonic hedgehog protein shh Xenopus laevis (African clawed frog) 444
Q90385 SHH_CYNPY Sonic hedgehog protein SHH Cynops pyrrhogaster (Japanese common newt) 432
Q63673 SHH_RAT Sonic hedgehog protein Shh, Vhh-1 Rattus norvegicus (Rat) 437
Q90419 TWHH_DANRE Tiggy-winkle hedgehog protein shhb, twhh Danio rerio (Zebrafish) (Brachydanio rerio) 416
Q14623 IHH_HUMAN Indian hedgehog protein IHH Homo sapiens (Human) 411
O43323 DHH_HUMAN Desert hedgehog protein DHH Homo sapiens (Human) 396
P97812 IHH_MOUSE Indian hedgehog protein Ihh Mus musculus (Mouse) 411
P79682 SHH_AMBCH Sonic hedgehog protein shh Amblypharyngodon chulabhornae 121
P79691 SHH_CARAU Sonic hedgehog protein shh Carassius auratus (Goldfish) 121
bio
Highlight All
1 of 4 matches
Reached end of page, continued from top

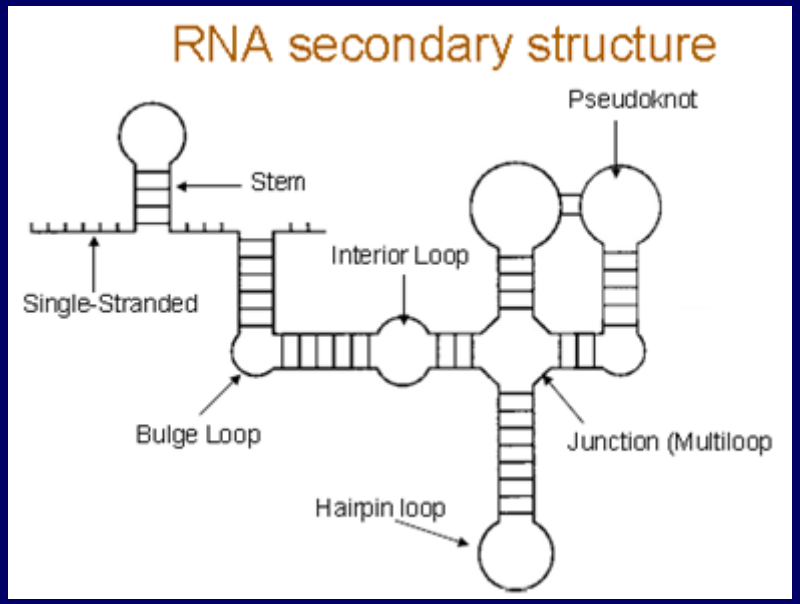

Disprot - Database of Prote...
IUPred
Pfam: Home page
sonic AND hedgehog in U...
www.uniprot.org/uniprot/?query=sonic+hedgehog&sort=score
UniProt
UniProtKB
sonic AND hedgehog
Advanced
BLAST Align Upload Lists
Help Contact
Show help for UniProtKB
Basket
Results
Quote terms: "sonic hedgehog"
Filter by
Reviewed (192)
Swiss-Prot
Unreviewed (175)
TrEMBL
Popular organisms
Mouse (62)
Human (49)
Zebrafish (25)
Rat (14)
Bovine (7)
Other organisms
Go
View by
Taxonomy
Keywords
Gene Ontology
Enzyme class
Pathway
UniRef
Your results in sequence clusters with identity of: 100%, 90% or 50%
Columns
BLAST
Align
Download
Add to basket
1 to 25 of 367
Show 25
Download selected (0)
Download all (367)
Format:
FASTA (canonical)
FASTA (canonical & isoform)
Tab-delimited
Text
Excel
GFF
XML
RDF/XML
List
Entry
Entry name
Organism
Length
Q62226 SHH_MOUSE Mus musculus (Mouse) 437
Q15465 SHH_HUMAN Homo sapiens (Human) 462
Q91035 SHH_CHICK Gallus gallus (Chicken) 425
Q92008 SHH_DANRE Danio rerio (Zebrafish) (Brachydanio rerio) 418
Q92000 SHH_XENLA Xenopus laevis (African clawed frog) 444
Q90385 SHH_CYNPY Sonic hedgehog protein SHH Cynops pyrrhogaster (Japanese common newt) 432
Q63673 SHH_RAT Sonic hedgehog protein Shh, Vhh-1 Rattus norvegicus (Rat) 437
Q90419 TWHH_DANRE Tiggy-winkle hedgehog protein shhb, twhh Danio rerio (Zebrafish) (Brachydanio rerio) 416
Q14623 IHH_HUMAN Indian hedgehog protein IHH Homo sapiens (Human) 411
O43323 DHH_HUMAN Desert hedgehog protein DHH Homo sapiens (Human) 396
P97812 IHH_MOUSE Indian hedgehog protein Ihh Mus musculus (Mouse) 411
P79682 SHH_AMBCH Sonic hedgehog protein shh Amblypharyngodon chulabhornae 121
P79691 SHH_CARAU Sonic hedgehog protein shh Carassius auratus (Goldfish) 121
bio
Highlight All
Match Case
1 of 4 matches
Reached end of page, continued from top

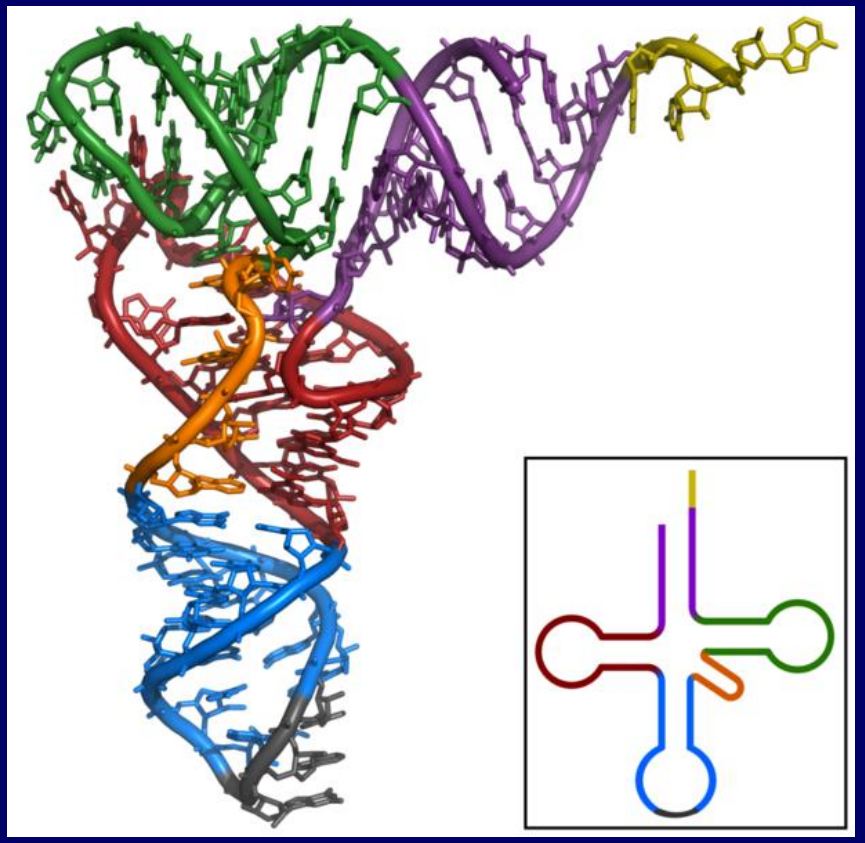
RNS szerkezet

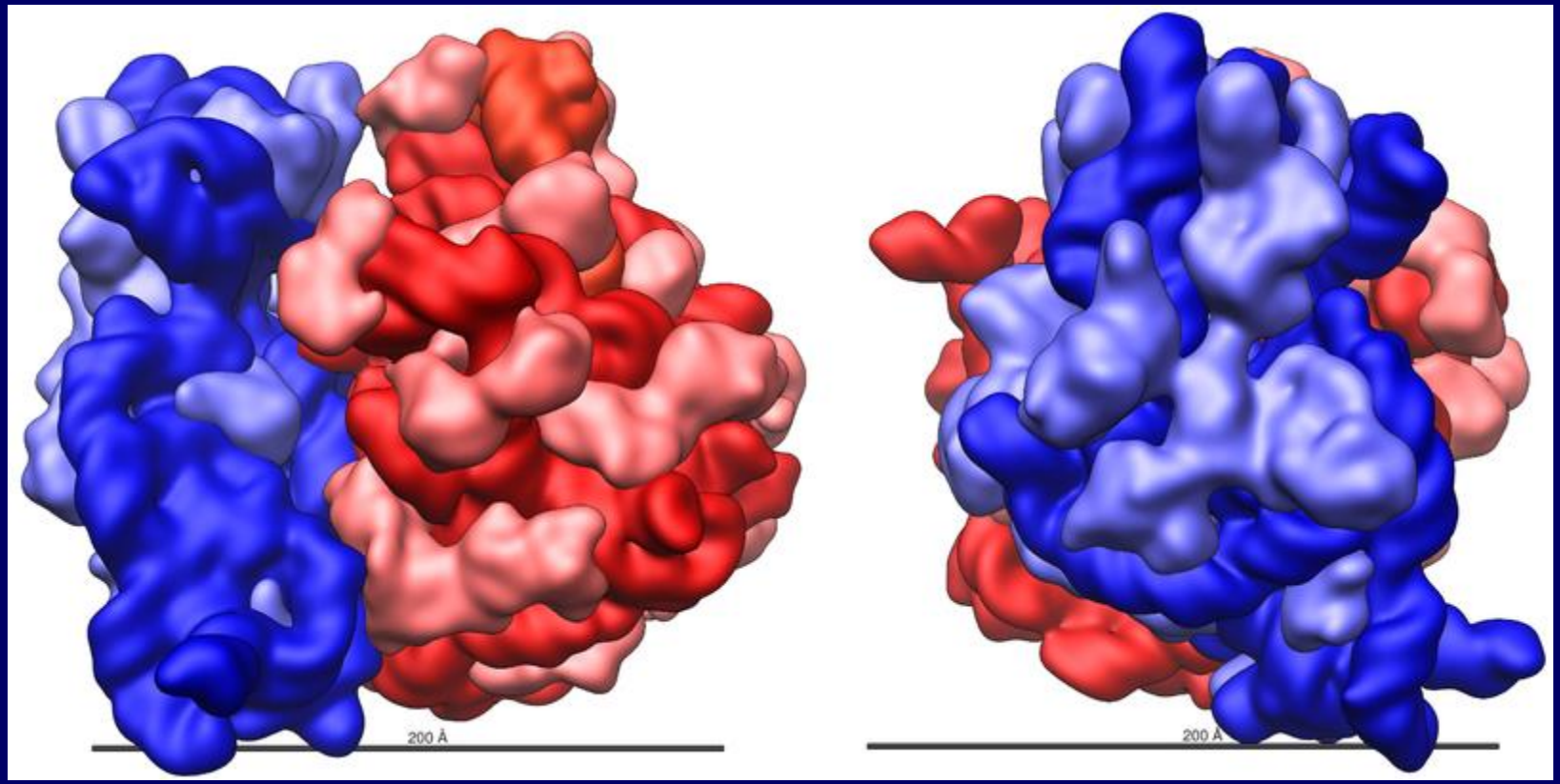
- Az RNS képes meghatározott térszerkezetet felvenni
- A szerkezet léte nagyban befolyásolja az RNS stabilitását
- Az RNS szerkezete miatt enzimatikus aktivitást is képes ellátni – ribozim
- Funkcionális komplexeket tud alkotni fehérjékkel – riboszóma

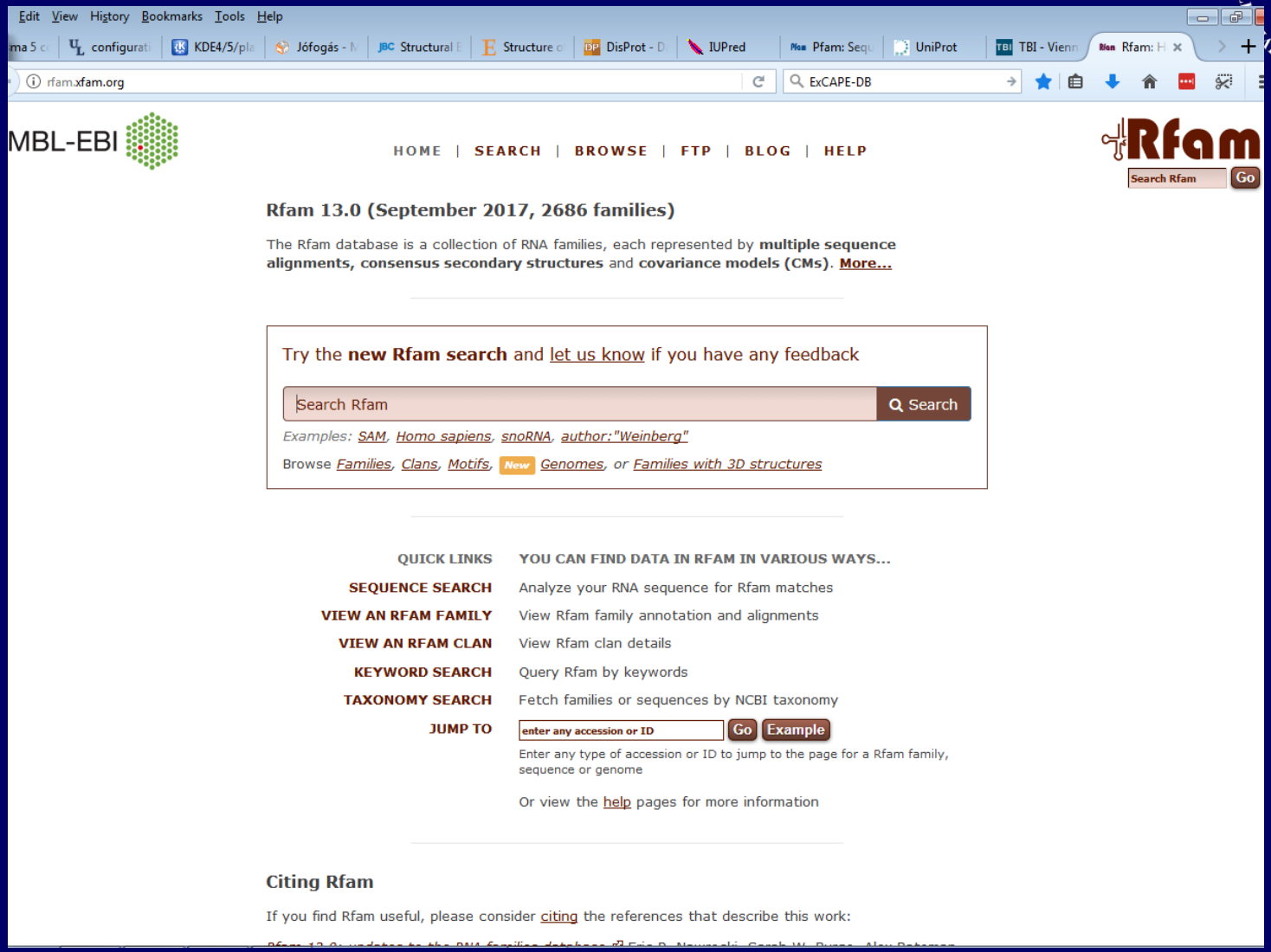
RNS folding

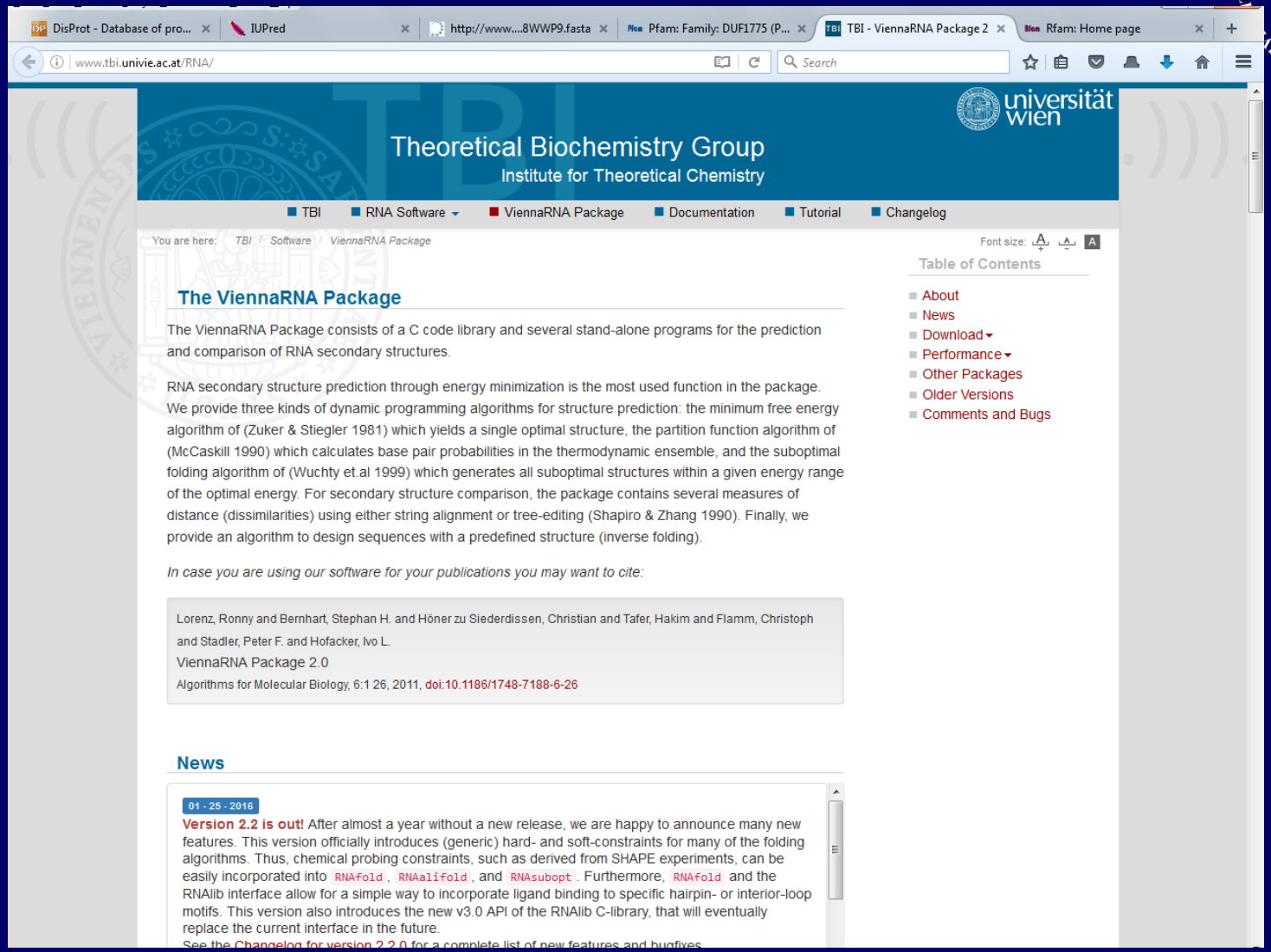
- Az egyszálú RNS saját magával alkot szerkezetet reverz komplement bázispárok alapján
- A lehetséges topológiákhoz energetikai értékek tartoznak
- Ez nagyon bonyolult problémát eredményez
- Web: http://www.tbi.univie.ac.at/~ivo/RNA/


RNA secondary structure
Pseudoknot
Stem
Interior Loop
Single-Stranded
Bulge Loop
Junction (Multiloop
Hairpin loop



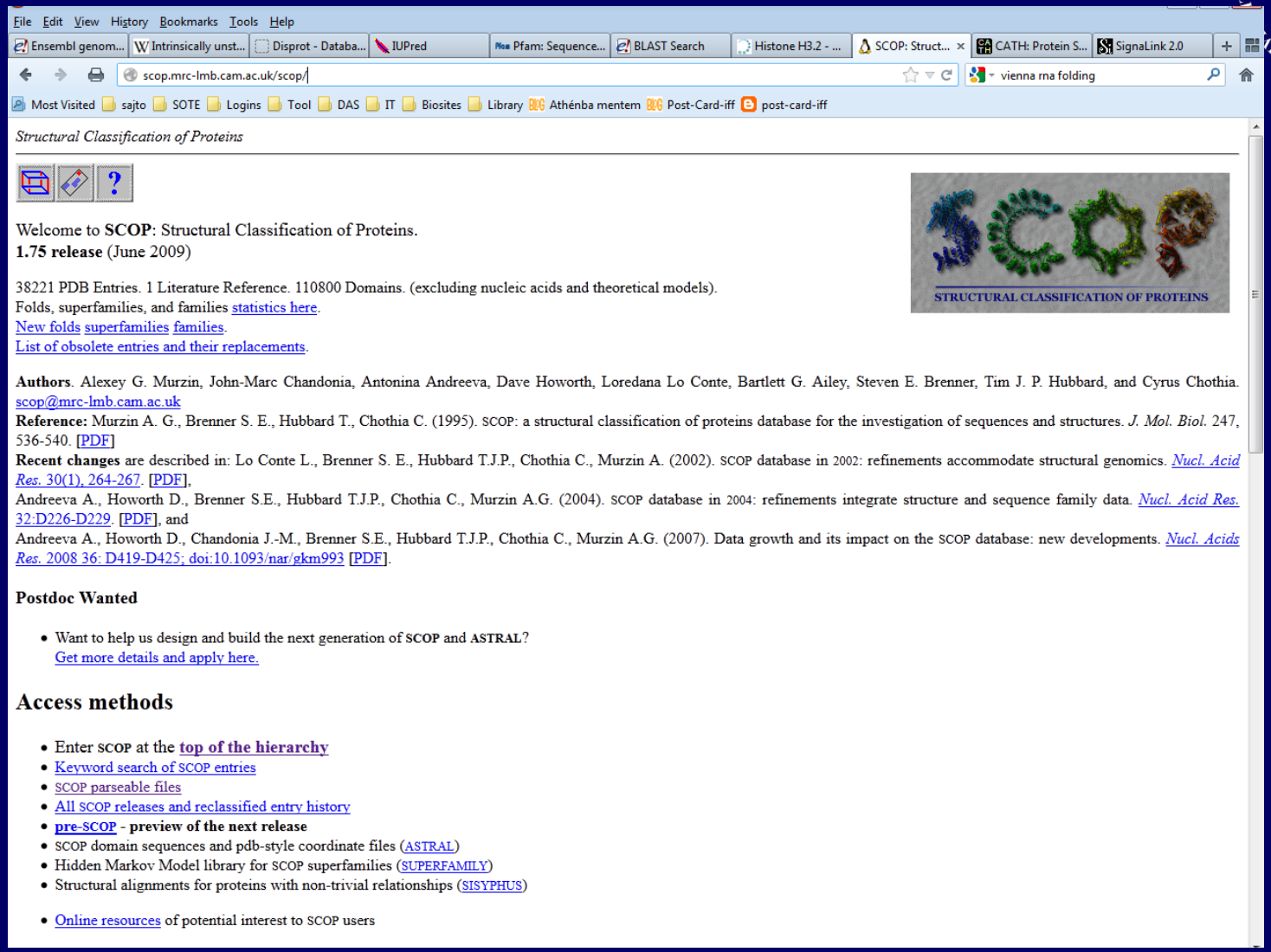

200 Å
200 Å


Edit View History Bookmarks Tools Help
rfam.xfam.org
ExCAPE-DB
MBL-EBI
HOME | SEARCH | BROWSE | FTP | BLOG | HELP
Rfam
Search Rfam
Go
Rfam 13.0 (September 2017, 2686 families)
The Rfam database is a collection of RNA families, each represented by multiple sequence alignments, consensus secondary structures and covariance models (CMs). More...
Try the new Rfam search and let us know if you have any feedback
Search Rfam
Search
Examples: SAM, Homo sapiens, snoRNA, author:"Weinberg"
Browse Families, Clans, Motifs, New Genomes, or Families with 3D structures
QUICK LINKS
YOU CAN FIND DATA IN RFAM IN VARIOUS WAYS...
SEQUENCE SEARCH
Analyze your RNA sequence for Rfam matches
VIEW AN RFAM FAMILY
View Rfam family annotation and alignments
VIEW AN RFAM CLAN
View Rfam clan details
KEYWORD SEARCH
Query Rfam by keywords
TAXONOMY SEARCH
Fetch families or sequences by NCBI taxonomy
JUMP TO
enter any accession or ID
Go
Example
Enter any type of accession or ID to jump to the page for a Rfam family, sequence or genome
Or view the help pages for more information
Citing Rfam
If you find Rfam useful, please consider citing the references that describe this work:


DisProt - Database of pro...
IUPred
http://www....8WWP9.fasta
Pfam: Family: DUF1775 (P...
TBI - ViennaRNA Package 2
Rfam: Home page
www.tbi.univie.ac.at/RNA/
Search
universität wien
Theoretical Biochemistry Group
Institute for Theoretical Chemistry
TBI
RNA Software
ViennaRNA Package
Documentation
Tutorial
Changelog
You are here: TBI / Software / ViennaRNA Package
Font size:
Table of Contents
About
News
Download
Performance
Other Packages
Older Versions
Comments and Bugs
The ViennaRNA Package
The ViennaRNA Package consists of a C code library and several stand-alone programs for the prediction and comparison of RNA secondary structures.
RNA secondary structure prediction through energy minimization is the most used function in the package. We provide three kinds of dynamic programming algorithms for structure prediction: the minimum free energy algorithm of (Zuker & Stiegler 1981) which yields a single optimal structure, the partition function algorithm of (McCaskill 1990) which calculates base pair probabilities in the thermodynamic ensemble, and the suboptimal folding algorithm of (Wuchty et.al 1999) which generates all suboptimal structures within a given energy range of the optimal energy. For secondary structure comparison, the package contains several measures of distance (dissimilarities) using either string alignment or tree-editing (Shapiro & Zhang 1990). Finally, we provide an algorithm to design sequences with a predefined structure (inverse folding).
In case you are using our software for your publications you may want to cite:
Lorenz, Ronny and Bernhart, Stephan H. and Höner zu Siederdissen, Christian and Tafer, Hakim and Flamm, Christoph and Stadler, Peter F. and Hofacker, Ivo L.
ViennaRNA Package 2.0
Algorithms for Molecular Biology, 6:1 26, 2011, doi:10.1186/1748-7188-6-26
News
01 - 25 - 2016
Version 2.2 is out! After almost a year without a new release, we are happy to announce many new features. This version officially introduces (generic) hard- and soft-constraints for many of the folding algorithms. Thus, chemical probing constraints, such as derived from SHAPE experiments, can be easily incorporated into RNAfold , RNAalifold , and RNAsubopt . Furthermore, RNAfold and the RNAlib interface allow for a simple way to incorporate ligand binding to specific hairpin- or interior-loop motifs. This version also introduces the new v3.0 API of the RNAlib C-library, that will eventually replace the current interface in the future.
See the Changelog for version 2.2.0 for a complete list of new features and bugfixes

Fehérjék szerkezeti osztályozása

- A szekvencia alapján a szerkezet megbecsülhető
- A szerkezet és a funkció szorosan összefügg
- Kell egy szerkezet-központú adatbázis
- A szerkezet hierarchhikus megközelítése
- SCOP: http://scop.mrc-lmb.cam.ac.uk/scop/
- CATH: http://www.cathdb.info/



Structural Classification of Proteins

Welcome to **SCOP**: Structural Classification of Proteins.
1.75 release (June 2009)

38221 PDB Entries. 1 Literature Reference. 110800 Domains. (excluding nucleic acids and theoretical models).
Folds, superfamilies, and families statistics here.
New folds superfamilies families.
List of obsolete entries and their replacements.

Authors. Alexey G. Murzin, John-Marc Chandonia, Antonina Andreeva, Dave Howorth, Loredana Lo Conte, Bartlett G. Ailey, Steven E. Brenner, Tim J. P. Hubbard, and Cyrus Chothia. scop@mrc-lmb.cam.ac.uk
Reference: Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536-540. [PDF]
Recent changes are described in: Lo Conte L., Brenner S. E., Hubbard T.J.P., Chothia C., Murzin A. (2002). SCOP database in 2002: refinements accommodate structural genomics. *Nucl. Acid Res.* 30(1), 264-267. [PDF],
Andreeva A., Howorth D., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2004). SCOP database in 2004: refinements integrate structure and sequence family data. *Nucl. Acid Res.* 32:D226-D229. [PDF], and
Andreeva A., Howorth D., Chandonia J.-M., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2007). Data growth and its impact on the SCOP database: new developments. *Nucl. Acids Res.* 2008 36: D419-D425; doi:10.1093/nar/gkm993 [PDF].

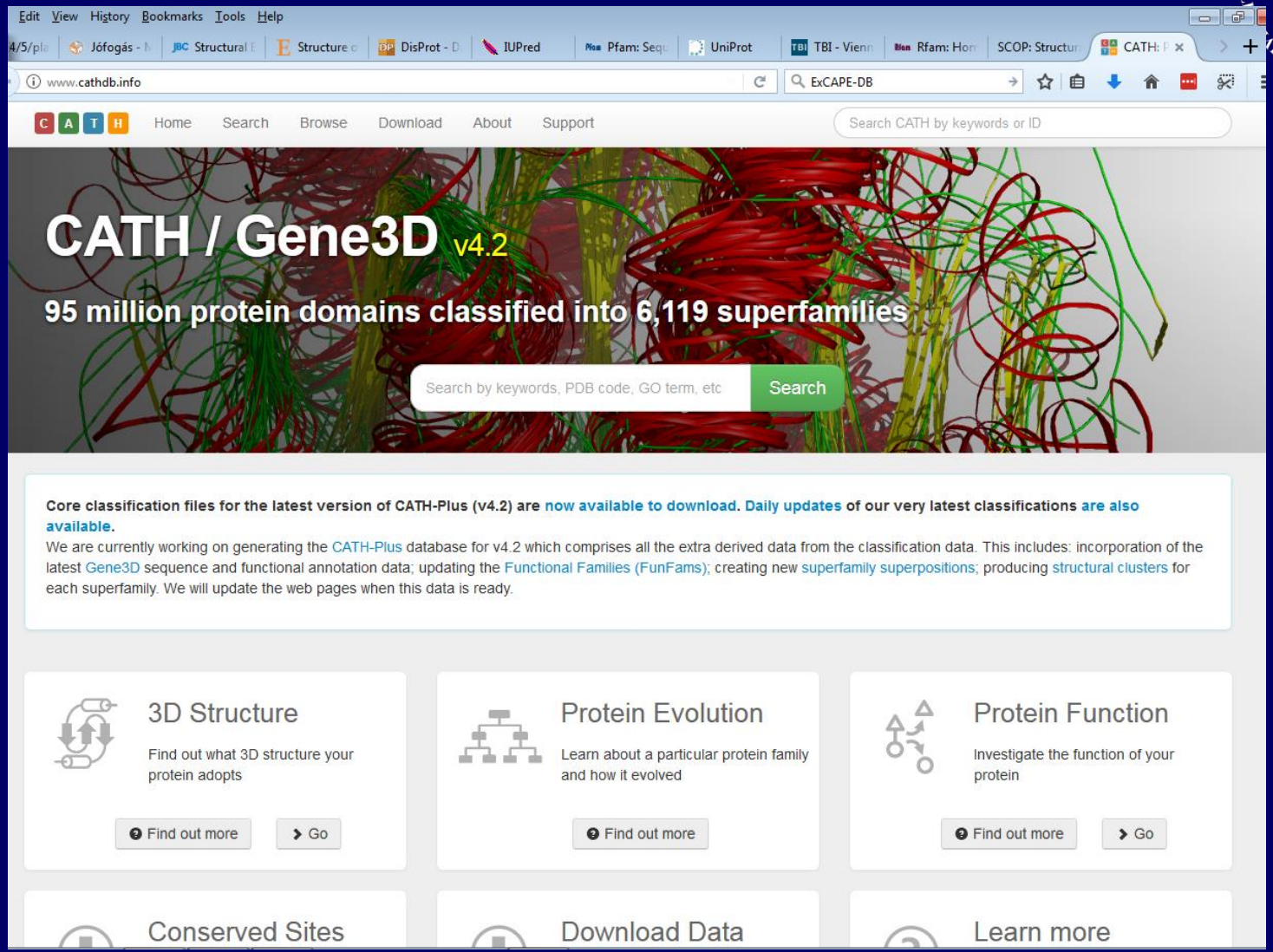
Postdoc Wanted

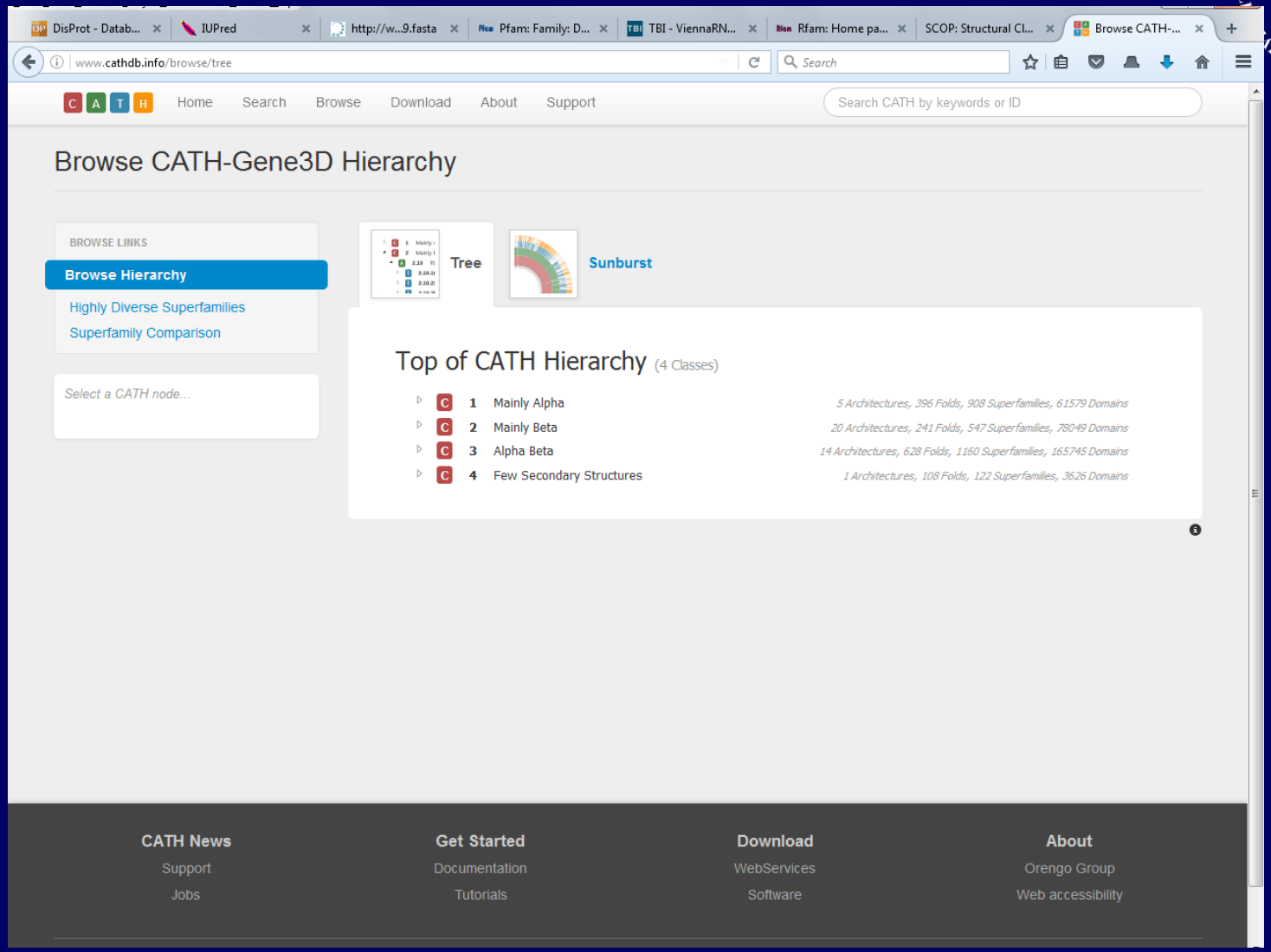
- Want to help us design and build the next generation of **SCOP** and **ASTRAL**?
  Get more details and apply here.

Access methods

- Enter SCOP at the **top of the hierarchy**
- Keyword search of SCOP entries
- SCOP parseable files
- All SCOP releases and reclassified entry history
- **pre-SCOP** - **preview of the next release**
- SCOP domain sequences and pdb-style coordinate files (ASTRAL)
- Hidden Markov Model library for SCOP superfamilies (SUPERFAMILY)
- Structural alignments for proteins with non-trivial relationships (SISYPHUS)

- Online resources of potential interest to SCOP users


Edit View History Bookmarks Tools Help
www.cathdb.info
ExCAPE-DB
CATH
Home
Search
Browse
Download
About
Support
Search CATH by keywords or ID
CATH / Gene3D v4.2
95 million protein domains classified into 6,119 superfamilies
Search by keywords, PDB code, GO term, etc
Search
Core classification files for the latest version of CATH-Plus (v4.2) are now available to download. Daily updates of our very latest classifications are also available.
We are currently working on generating the CATH-Plus database for v4.2 which comprises all the extra derived data from the classification data. This includes: incorporation of the latest Gene3D sequence and functional annotation data; updating the Functional Families (FunFams); creating new superfamily superpositions; producing structural clusters for each superfamily. We will update the web pages when this data is ready.
3D Structure
Find out what 3D structure your protein adopts
Find out more
Go
Protein Evolution
Learn about a particular protein family and how it evolved
Find out more
Protein Function
Investigate the function of your protein
Find out more
Go
Conserved Sites
Download Data
Learn more


DisProt - Datab...
IUPred
http://w...9.fasta
Pfam: Family: D...
TBI - ViennaRN...
Rfam: Home pa...
SCOP: Structural Cl...
Browse CATH-...
www.cathdb.info/browse/tree
Search
CATH
Home
Search
Browse
Download
About
Support
Search CATH by keywords or ID
Browse CATH-Gene3D Hierarchy
BROWSE LINKS
Browse Hierarchy
Highly Diverse Superfamilies
Superfamily Comparison
Select a CATH node...
Tree
Sunburst
Top of CATH Hierarchy (4 Classes)
1 Mainly Alpha
5 Architectures, 396 Folds, 908 Superfamilies, 61579 Domains
2 Mainly Beta
20 Architectures, 241 Folds, 547 Superfamilies, 78049 Domains
3 Alpha Beta
14 Architectures, 628 Folds, 1160 Superfamilies, 165745 Domains
4 Few Secondary Structures
1 Architectures, 108 Folds, 122 Superfamilies, 3626 Domains
CATH News
Support
Jobs
Get Started
Documentation
Tutorials
Download
WebServices
Software
About
Orengo Group
Web accessibility

DisProt - Datab... | IUPred | http://w...9.fasta | Pfam: Family: D... | TBI - ViennaRN... | Rfam: Home pa... | SCOP: Structural Cl... | Browse CATH-...

www.cathdb.info/browse/tree

CATH Home Search Browse Download About Support — Search CATH by keywords or ID

Browse CATH-Gene3D Hierarchy

BROWSE LINKS

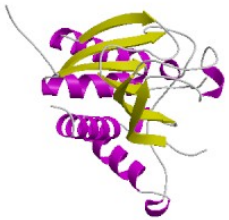
- **Browse Hierarchy**
- Highly Diverse Superfamilies
- Superfamily Comparison

Select a CATH node...

C Alpha Beta

3

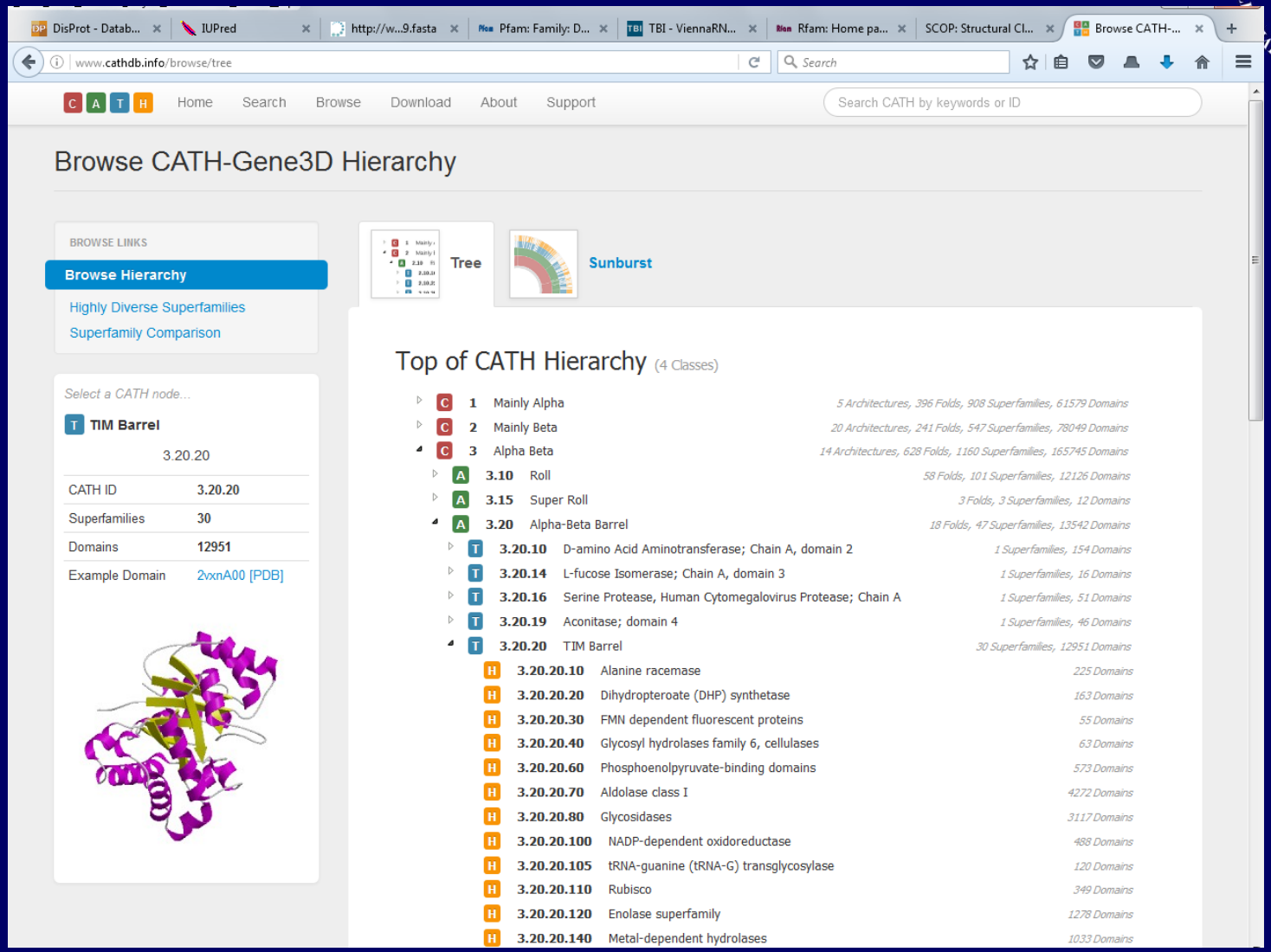
CATH ID	3
Architectures	14
Topologies	628
Superfamilies	1160
Domains	165745
Example Domain	1c0pA01 [PDB]

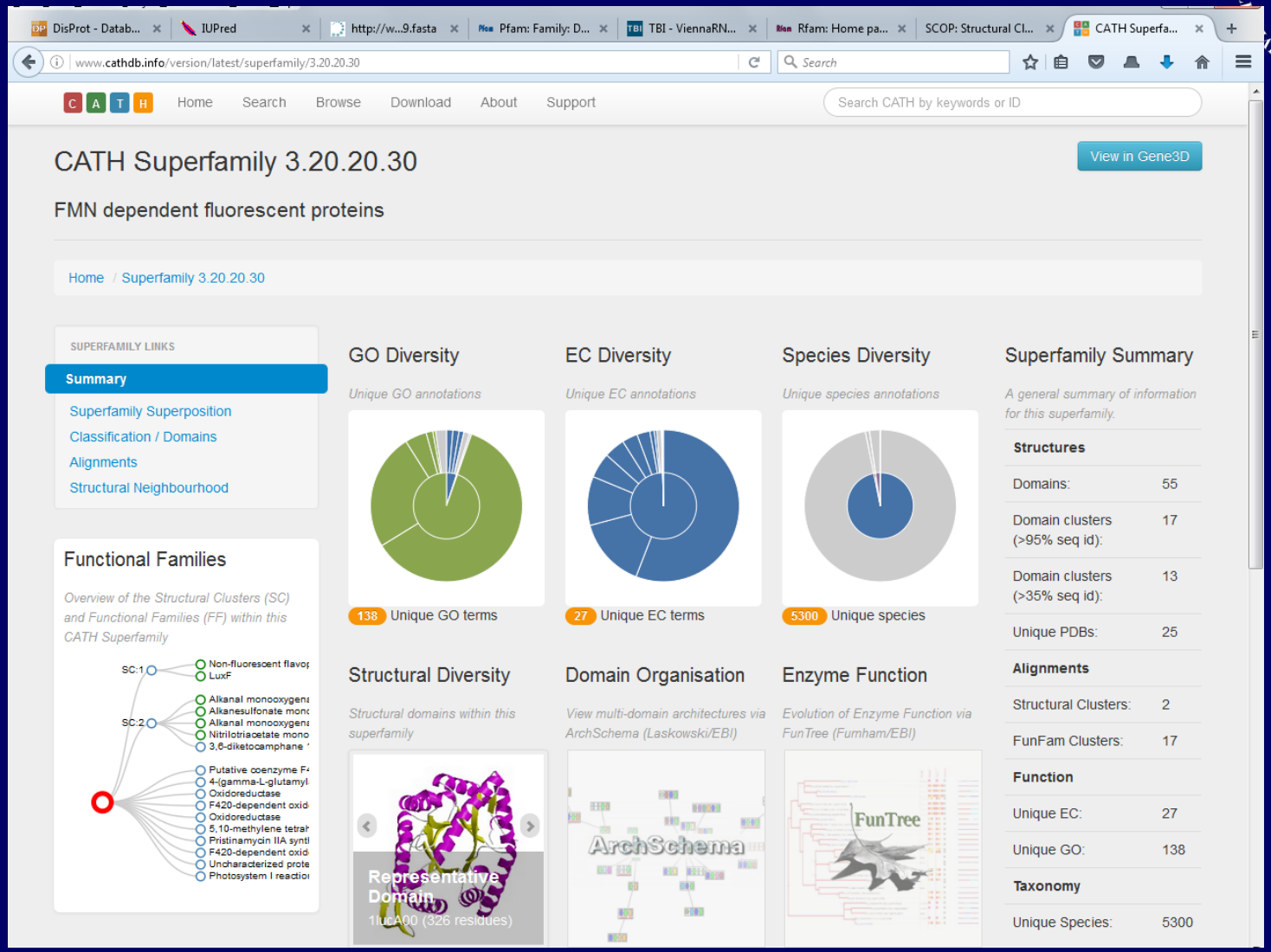


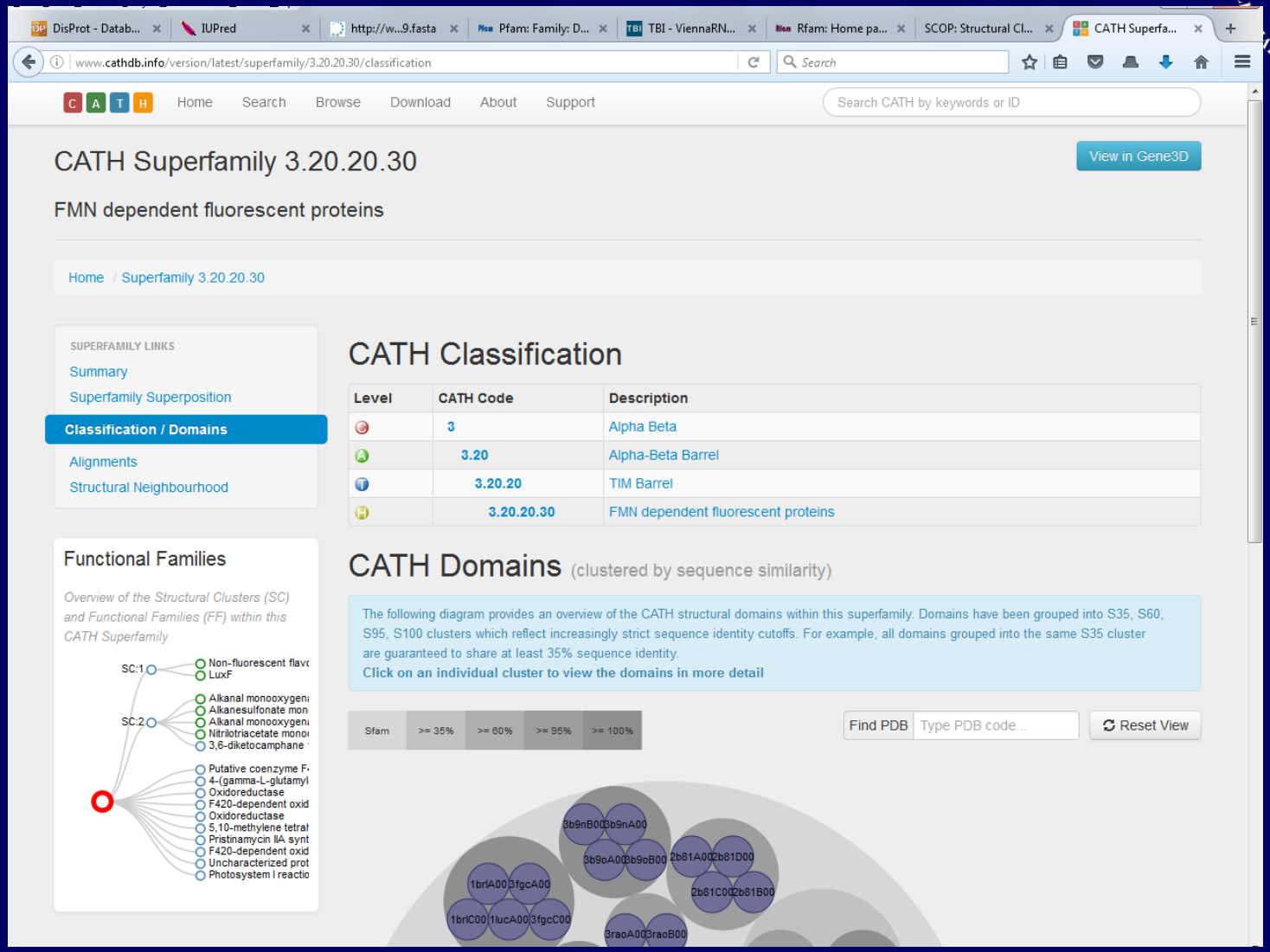
Tree | Sunburst

Top of CATH Hierarchy (4 Classes)

- C 1 Mainly Alpha — *5 Architectures, 396 Folds, 908 Superfamilies, 61579 Domains*
- C 2 Mainly Beta — *20 Architectures, 241 Folds, 547 Superfamilies, 78049 Domains*
- C 3 Alpha Beta — *14 Architectures, 628 Folds, 1160 Superfamilies, 165745 Domains*
  - A 3.10 Roll — *58 Folds, 101 Superfamilies, 12126 Domains*
  - A 3.15 Super Roll — *3 Folds, 3 Superfamilies, 12 Domains*
  - A 3.20 Alpha-Beta Barrel — *18 Folds, 47 Superfamilies, 13542 Domains*
  - A 3.30 2-Layer Sandwich — *224 Folds, 495 Superfamilies, 46678 Domains*
  - A 3.40 3-Layer(aba) Sandwich — *126 Folds, 287 Superfamilies, 65246 Domains*
  - A 3.50 3-Layer(bba) Sandwich — *11 Folds, 17 Superfamilies, 3194 Domains*
  - A 3.55 3-Layer(bab) Sandwich — *6 Folds, 6 Superfamilies, 28 Domains*
  - A 3.60 4-Layer Sandwich — *16 Folds, 18 Superfamilies, 4785 Domains*
  - A 3.65 Alpha-beta prism — *1 Folds, 2 Superfamilies, 436 Domains*
  - A 3.70 Box — *1 Folds, 1 Superfamilies, 186 Domains*
  - A 3.75 5-stranded Propeller — *1 Folds, 2 Superfamilies, 143 Domains*
  - A 3.80 Alpha-Beta Horseshoe — *3 Folds, 4 Superfamilies, 386 Domains*
  - A 3.90 Alpha-Beta Complex — *159 Folds, 176 Superfamilies, 18797 Domains*
  - A 3.100 Ribosomal Protein L15; Chain: K; domain 2 — *1 Folds, 1 Superfamilies, 186 Domains*
- C 4 Few Secondary Structures — *1 Architectures, 108 Folds, 122 Superfamilies, 3626 Domains*


DisProt - Datab...
IUPred
http://w...9.fasta
Pfam: Family: D...
TBI - ViennaRN...
Rfam: Home pa...
SCOP: Structural Cl...
Browse CATH-...
www.cathdb.info/browse/tree
Search
C A T H
Home
Search
Browse
Download
About
Support
Search CATH by keywords or ID
Browse CATH-Gene3D Hierarchy
BROWSE LINKS
Browse Hierarchy
Highly Diverse Superfamilies
Superfamily Comparison
Tree
Sunburst
Select a CATH node...
TIM Barrel
3.20.20
CATH ID 3.20.20
Superfamilies 30
Domains 12951
Example Domain 2vxnA00 [PDB]
Top of CATH Hierarchy (4 Classes)
1 Mainly Alpha 5 Architectures, 396 Folds, 908 Superfamilies, 61579 Domains
2 Mainly Beta 20 Architectures, 241 Folds, 547 Superfamilies, 78049 Domains
3 Alpha Beta 14 Architectures, 628 Folds, 1160 Superfamilies, 165745 Domains
3.10 Roll 58 Folds, 101 Superfamilies, 12126 Domains
3.15 Super Roll 3 Folds, 3 Superfamilies, 12 Domains
3.20 Alpha-Beta Barrel 18 Folds, 47 Superfamilies, 13542 Domains
3.20.10 D-amino Acid Aminotransferase; Chain A, domain 2 1 Superfamilies, 154 Domains
3.20.14 L-fucose Isomerase; Chain A, domain 3 1 Superfamilies, 16 Domains
3.20.16 Serine Protease, Human Cytomegalovirus Protease; Chain A 1 Superfamilies, 51 Domains
3.20.19 Aconitase; domain 4 1 Superfamilies, 46 Domains
3.20.20 TIM Barrel 30 Superfamilies, 12951 Domains
3.20.20.10 Alanine racemase 225 Domains
3.20.20.20 Dihydropteroate (DHP) synthetase 163 Domains
3.20.20.30 FMN dependent fluorescent proteins 55 Domains
3.20.20.40 Glycosyl hydrolases family 6, cellulases 63 Domains
3.20.20.60 Phosphoenolpyruvate-binding domains 573 Domains
3.20.20.70 Aldolase class I 4272 Domains
3.20.20.80 Glycosidases 3117 Domains
3.20.20.100 NADP-dependent oxidoreductase 488 Domains
3.20.20.105 tRNA-guanine (tRNA-G) transglycosylase 120 Domains
3.20.20.110 Rubisco 349 Domains
3.20.20.120 Enolase superfamily 1278 Domains
3.20.20.140 Metal-dependent hydrolases 1033 Domains


DisProt - Datab...
IUPred
http://w...9.fasta
Pfam: Family: D...
TBI - ViennaRN...
Rfam: Home pa...
SCOP: Structural Cl...
CATH Superfa...
www.cathdb.info/version/latest/superfamily/3.20.20.30
Search
CATH
Home
Search
Browse
Download
About
Support
Search CATH by keywords or ID
CATH Superfamily 3.20.20.30
View in Gene3D
FMN dependent fluorescent proteins
Home / Superfamily 3.20.20.30
SUPERFAMILY LINKS
Summary
Superfamily Superposition
Classification / Domains
Alignments
Structural Neighbourhood
Functional Families
Overview of the Structural Clusters (SC) and Functional Families (FF) within this CATH Superfamily
SC:1
SC:2
Non-fluorescent flavop
LuxF
Alkanal monooxygena
Alkanesulfonate mono
Alkanal monooxygena
Nitrilotriacetate mono
3,6-diketocamphane
Putative coenzyme F4
4-(gamma-L-glutamyl
Oxidoreductase
F420-dependent oxid
Oxidoreductase
5,10-methylene tetrah
Pristinamycin IIA synth
F420-dependent oxid
Uncharacterized prote
Photosystem I reactio
GO Diversity
Unique GO annotations
138 Unique GO terms
EC Diversity
Unique EC annotations
27 Unique EC terms
Species Diversity
Unique species annotations
5300 Unique species
Superfamily Summary
A general summary of information for this superfamily.
Structures
Domains: 55
Domain clusters (>95% seq id): 17
Domain clusters (>35% seq id): 13
Unique PDBs: 25
Alignments
Structural Clusters: 2
FunFam Clusters: 17
Function
Unique EC: 27
Unique GO: 138
Taxonomy
Unique Species: 5300
Structural Diversity
Structural domains within this superfamily
Representative Domain
1luca00 (326 residues)
Domain Organisation
View multi-domain architectures via ArchSchema (Laskowski/EBI)
ArchSchema
Enzyme Function
Evolution of Enzyme Function via FunTree (Furnham/EBI)
FunTree


DisProt - Datab...
IUPred
http://w...9.fasta
Pfam: Family: D...
TBI - ViennaRN...
Rfam: Home pa...
SCOP: Structural Cl...
CATH Superfa...
www.cathdb.info/version/latest/superfamily/3.20.20.30/classification
Search
CATH
Home
Search
Browse
Download
About
Support
Search CATH by keywords or ID
CATH Superfamily 3.20.20.30
View in Gene3D
FMN dependent fluorescent proteins
Home / Superfamily 3.20.20.30
SUPERFAMILY LINKS
Summary
Superfamily Superposition
Classification / Domains
Alignments
Structural Neighbourhood
CATH Classification
Level
CATH Code
Description
3
Alpha Beta
3.20
Alpha-Beta Barrel
3.20.20
TIM Barrel
3.20.20.30
FMN dependent fluorescent proteins
Functional Families
Overview of the Structural Clusters (SC) and Functional Families (FF) within this CATH Superfamily
SC:1
SC:2
CATH Domains (clustered by sequence similarity)
The following diagram provides an overview of the CATH structural domains within this superfamily. Domains have been grouped into S35, S60, S95, S100 clusters which reflect increasingly strict sequence identity cutoffs. For example, all domains grouped into the same S35 cluster are guaranteed to share at least 35% sequence identity.
Click on an individual cluster to view the domains in more detail
Sfam
>= 35%
>= 60%
>= 95%
>= 100%
Find PDB
Type PDB code...
Reset View


DisProt - Datab...
IUPred
http://w...9.fasta
Pfam: Family: D...
TBI - ViennaRN...
Rfam: Home pa...
SCOP: Structural Cl...
CATH Superfa...
www.cathdb.info/version/latest/superfamily/3.20.20.30/superposition
Search
CATH
Home
Search
Browse
Download
About
Support
Search CATH by keywords or ID
Overview of the Structural Clusters (SC) and Functional Families (FF) within this CATH Superfamily
SC:1
SC:2
Non-fluorescent flavo
LuxF
Alkanal monooxygena
Alkanesulfonate mono
Alkanal monooxygena
Nitrilotriacetate mono
3,6-diketocamphane
Putative coenzyme F4
4-(gamma-L-glutamyl
Oxidoreductase
F420-dependent oxid
Oxidoreductase
5,10-methylene tetrah
Pristinamycin IIA synt
F420-dependent oxid
Uncharacterized prote
Photosystem I reactio

Mit tanultunk ma?

- A szekvenciák elemzése önmagukban lehetséges, de pontatlan eredményt ad
- "a fontos dolgok evolúciósan konzerváltak"
- Szekvenciák annotálása homológia alapján
- Gyors, pontos, hatékony
- De azért csak óvatosan….

File Edit View Window Help

OPEN ACCESS Freely available online

PLoS BIOLOGY

Deletion of Ultraconserved Elements Yields Viable Mice

Nadav Ahituv[1,2¤], Yiwen Zhu[1], Axel Visel[1], Amy Holt[1], Veena Afzal[1], Len A. Pennacchio[1,2], Edward M. Rubin[1,2*]

1 Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California, United States of America, **2** United States Department of Energy Joint Genome Institute, Walnut Creek, California, United States of America

Ultraconserved elements have been suggested to retain extended perfect sequence identity between the human, mouse, and rat genomes due to essential functional properties. To investigate the necessities of these elements in vivo, we removed four noncoding ultraconserved elements (ranging in length from 222 to 731 base pairs) from the mouse genome. To maximize the likelihood of observing a phenotype, we chose to delete elements that function as enhancers in a mouse transgenic assay and that are near genes that exhibit marked phenotypes both when completely inactivated in the mouse and when their expression is altered due to other genomic modifications. Remarkably, all four resulting lines of mice lacking these ultraconserved elements were viable and fertile, and failed to reveal any critical abnormalities when assayed for a variety of phenotypes including growth, longevity, pathology, and metabolism. In addition, more targeted screens, informed by the abnormalities observed in mice in which genes in proximity to the investigated elements had been altered, also failed to reveal notable abnormalities. These results, while not inclusive of all the possible phenotypic impact of the deleted sequences, indicate that extreme sequence constraint does not necessarily reflect crucial functions required for viability.

Citation: Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, et al. (2007) Deletion of ultraconserved elements yields viable mice. PLoS Biol 5(9): e234. doi:10.1371/journal.pbio.0050234

Introduction

Evolutionary conservation has become a powerful means for identifying functionally important genomic sequences [1,2]. Ultraconserved elements have been defined as a group of extremely conserved sequences that show 100% identity over 200 bp or greater between the human, mouse, and rat genomes [3]. This category of extreme evolutionary sequence conservation is represented by 481 sequences in the human genome, of which over half show no evidence of tran-

Results

Generation and General Characterization of Ultraconserved Knockout Mice

To increase the probability of observing an associated phenotype in the ultraconserved null mice, we employed a variety of criteria in selecting the noncoding ultraconserved elements for deletion. We chose elements that showed tissue-specific in vivo enhancer activity in a mouse transgenic reporter assay that tended to recapitulate aspects of the

Feladat 8.

- A 6. feladatban használt “kedvenc” fehérjédet elemezd funkció és szerkezet alapján.
- A következő órán gyakorlat! Aki tud, hozzon magának laptopot.