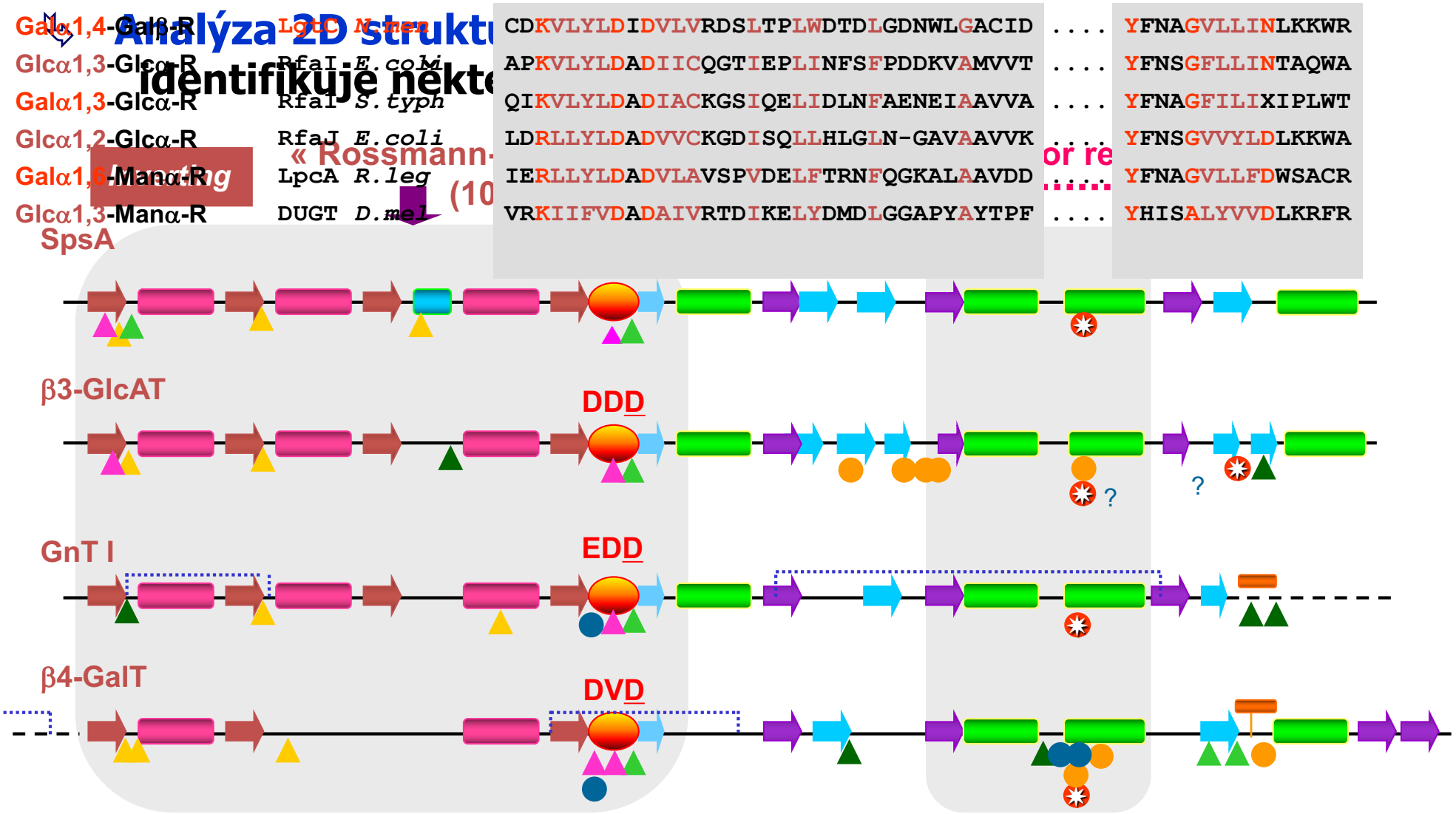
Předpověď 3D-struktury a topologie bílkovin, strukturní a funkční klasifikace

Předpověď 3D-struktury/ foldu

- Klasifikace proteinů
- Předpověď funkce
- Vytvoření modelu pro další studium

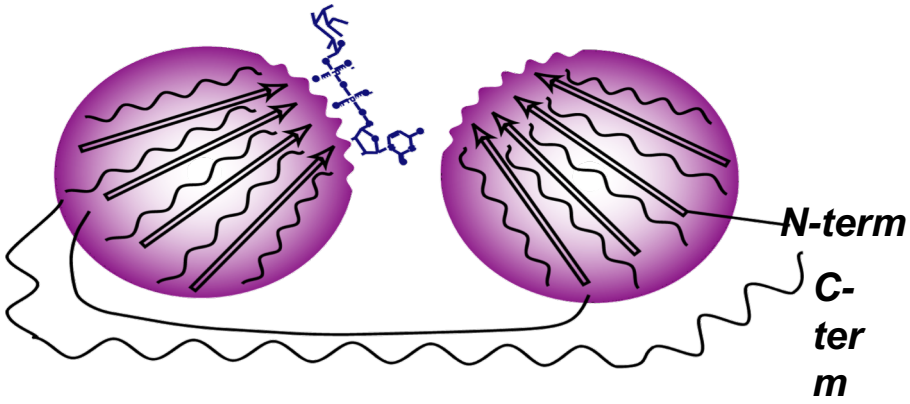
Metody pro predikci funkce

- **„klasické“ metody: vícenásobné aminokyselinové přiložení**
  pozitivní alignment pouze mezi sekvencemi stejné rodiny



Dvě pozorované topologie 3D struktur glykosyltransferas

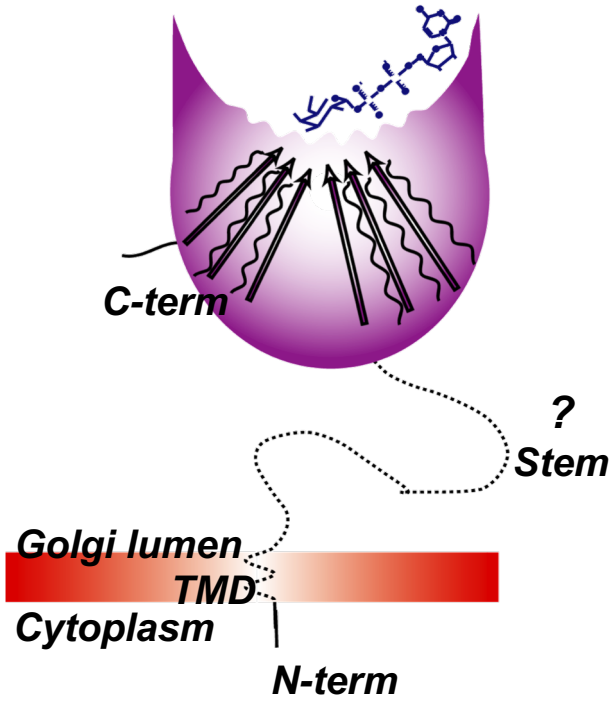
BGT-fold



(Procaryotes/Phage)

β-GlcT (**BGT**, phage T4)	**n.c.**	**inv**
β4-GlcNAcT (**MurG**, *E.coli*)	**GT28**	**inv**
β-GlcT (**GtfB**, *M. orientalis*)	**GT1**	**inv**

SpsA-fold



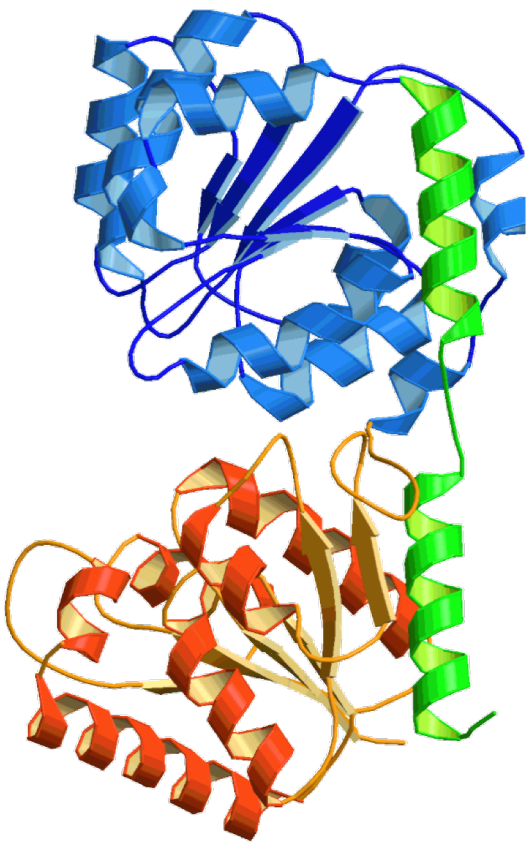
(Procaryotes)

SpsA (B. subtilis)	**GT2**	**inv**
α4-GalT (**LgtC**, N.meningitis)	**GT8**	**ret**

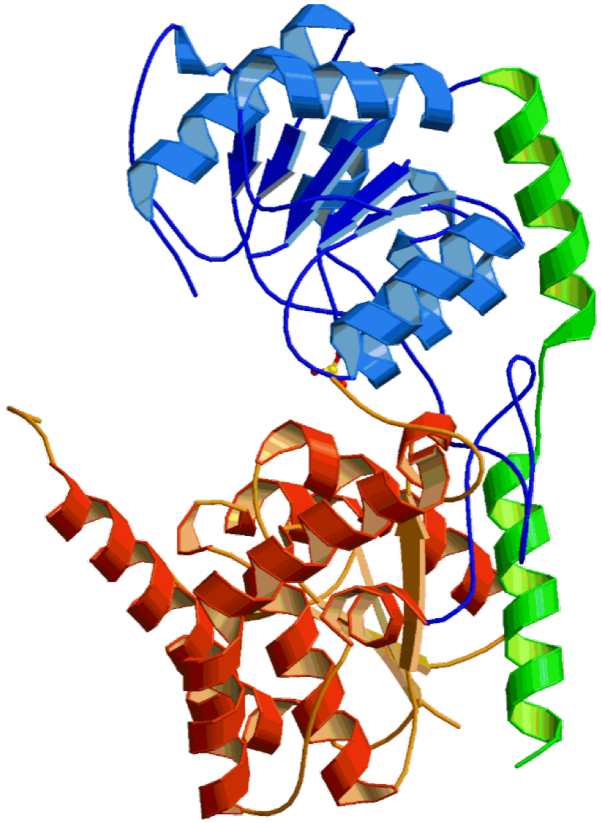
(Eucaryotes)

β4-GalT1 (bovine)	**GT7**	**inv**
β2-GlcNAcT (**GnT I**, rabbit)	**GT13**	**inv**
β3-GlcAT I (human)	**GT43**	**inv**
α3-GalT (bovine)	**GT6**	**ret**
Glycogenin (rabbit)	**GT8**	**ret**
α3-GalNacT (**GTA**, human)	**GT6**	**ret**
α3-GalT (**GTB**, human)	**GT6**	**ret**

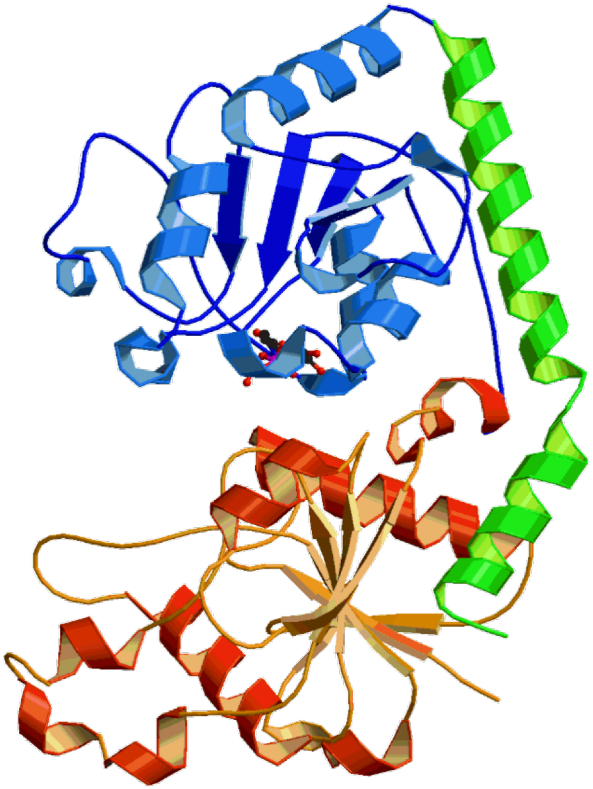
Nadrodina s BGT foldem



MurG (β-GlcNAcT)
GT28
E. coli
Ha *et al.*, 2000



GtfB (β-GlcT)
GT1
A. orientalis
Mulichak *et al*, 2001

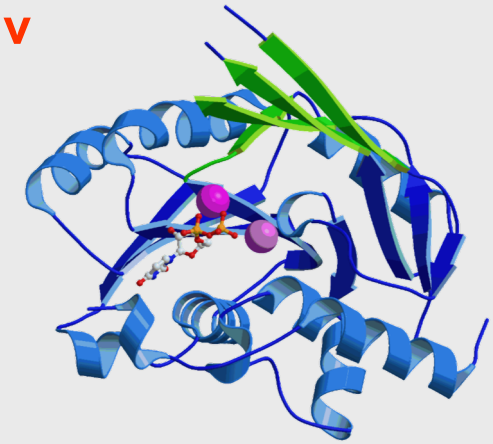


BGT (β-GlcT)
n.c.
Phage T4
Vrielink *et al*, 1994

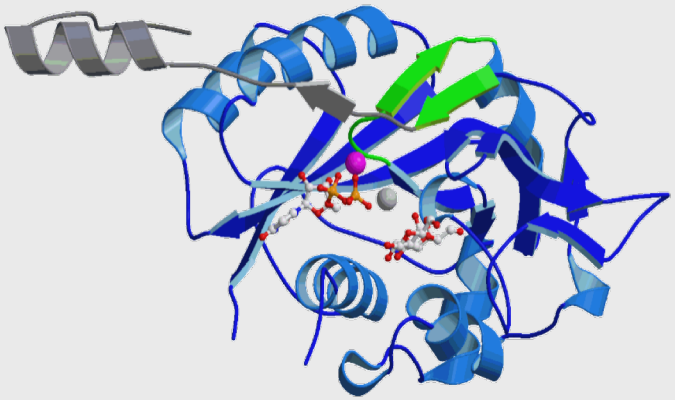
Nadrodina s SpsA foldem

Společná NBD

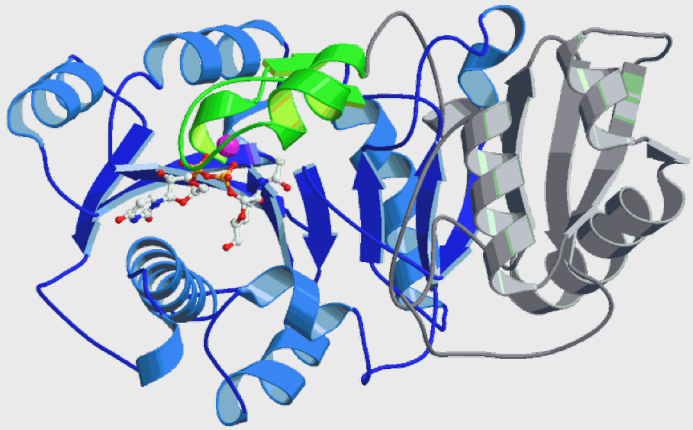
Inv



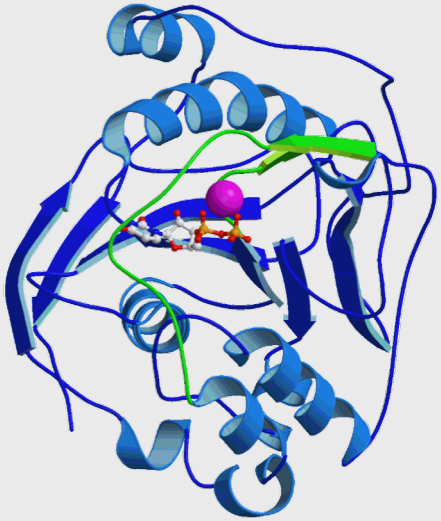
SpsA [GT2]
Charnok *et al*, 1999, 2001



Hum β3-GlcAT [GT43]
Pedersen *et al*, 2000

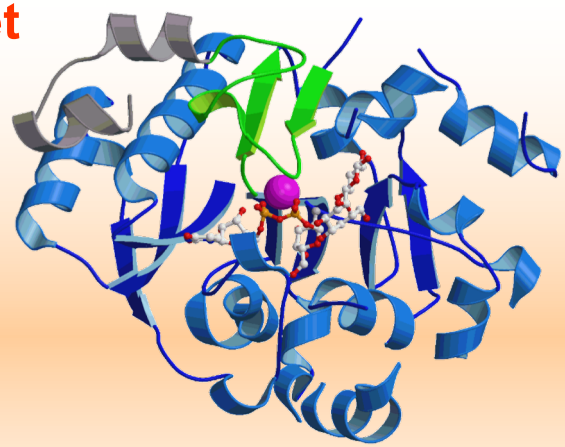


Rabbit GnT I [GT13]
Ünligil *et al*, 2000

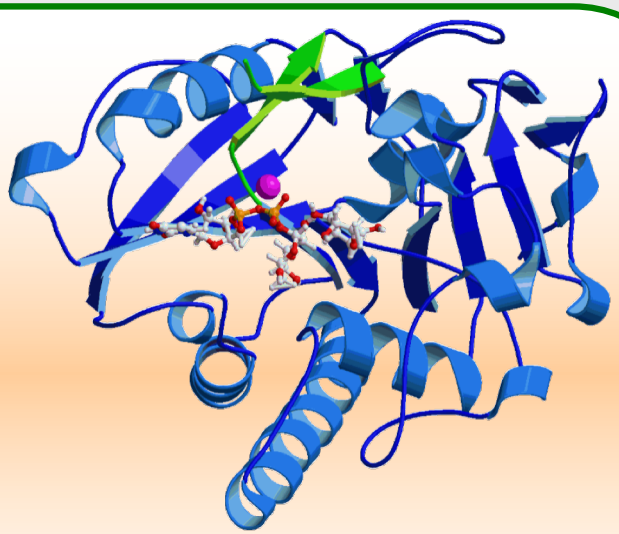


Bovine β4-GalT [GT7]
Gastinel *et al*, 1999
Ramakrishnan *et al*, 2001, 2002

Ret



LgtC (α4-GalT) [GT8]
Neisseria meningitidis
Persson *et al*, 2001



Bovine α3-GalT [GT6]
Gastinel *et al*, 2001
Boix *et al*, 2001, 2002

Předpověď 3D-struktury/ foldu

- Klasifikace proteinů
- Předpověď funkce
- Vytvoření modelu pro další studium

- Threading - „navlékání“
- Homology modeling
- *Ab inicio* metody

Threading

- „navlékání“ = rozpoznání a přiřazení proteinového foldu aminokyselinové sekvenci

- sekvence je porovnávána s databází existujících foldů (3D profilů) a na jejich základě jsou konstruovány 3D- modely

- 3D profil - každému reziduu v 3D struktuře je přiřazena environmentální proměnná (obsah polárních atomů v postranním řetězci, skrytá plocha, sekundární elementy, apod.) vycházející z předpokladu, že okolí rezidua je více konzervováno než aminokyselina samotná.

- Reziduum může být také popsáno pomocí svých interakcí

- Výsledná kvalita modelu shoda je popsána pomocí Z-skóre nebo energie

- U multidoménových struktur je potřeba aminokyselinovou sekvenci rozdělit na jednotlivé domény a analyzovat je separátně

PHYRE2 (3D-PSSM)

http://www.sbg.bio.ic.ac.uk/phyre2

Threading at 2D level and scoring at 3D level :
matching of secondary structure elements, and propensities of the residues in the query sequence to occupy varying levels of solvent accessibility

The PSIPRED Protein Sequence Analysis Workbench

http://bioinf.cs.ucl.ac.uk/psipred/

GenTHREADER Rapid fold recognition, matching your sequence against a library of whole PDB chains.
pGenTHREADER Highly sensitive fold recognition using profile-profile comparison (whole chain library).
pDomTHREADER Highly sensitive homologous domain recognition using profile-profile comparison (domain library).

I-TASSER

https://zhanglab.ccmb.med.umich.edu/I-TASSER/

a hierarchical approach to protein structure and function prediction. It first identifies structural templates from the PDB by multiple threading approach LOMETS, with full-length atomic models constructed by iterative template fragment assembly simulations. Function insights of the target are then derived by threading the 3D models through protein function database BioLiP.

Threading

Protein **H**omology/analog**Y** **R**ecognition **E**ngine

(nástupce 3D-PSSM)

•sekvenční „alignment“ s porovnávanou strukturou
•Využívá PSSMs (position-specific scoring matrix) generovanou metodou PSI-Blast jak pro cílovou sekvenci tak sekvencemi ze známých struktur.

•Kopírování 3D souřadnic a přepis jednotlivých reziduí podle zkoumané sekvence

•Následně porovná shodu profilů cílové sekvence a porovnáváné struktury společně se shodou jejich sekundárních struktur.

•Jediné zásahy do aminokyselinové páteře templátu jsou při modelování inzercí a delecí v sekvenci oproti porovnávané struktuře.

Phyre2

Search the 10 million known sequences for homologues using PSI-Blast.

Phyre2

User sequence

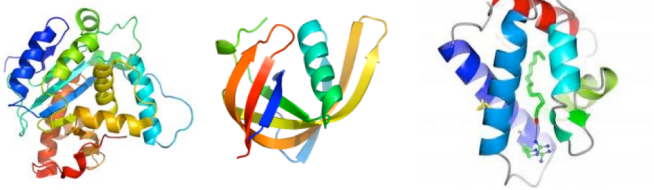
PSI-Blast

Hidden Markov model

Capture the mutational propensities at each position in the protein

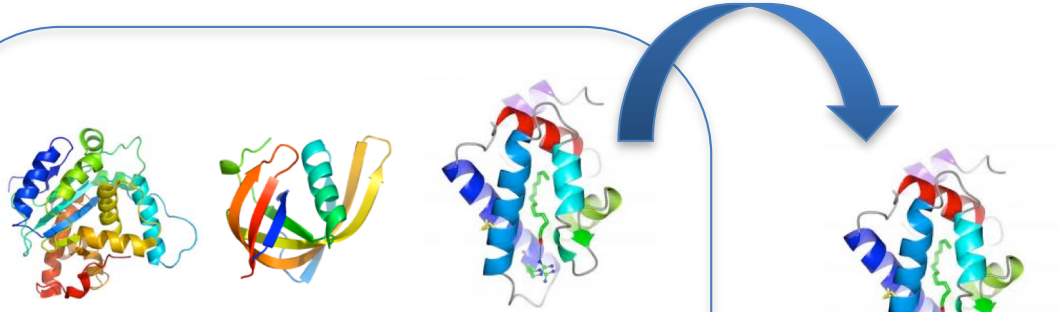
An evolutionary fingerprint

Phyre2



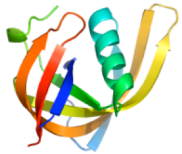
~ 65,000 known 3D structures

Phyre2



~ 65,000 known 3D structures

Phyre2


~ 65,000 known 3D structures

Extract sequence

HAPTLVRDC…….

Phyre2

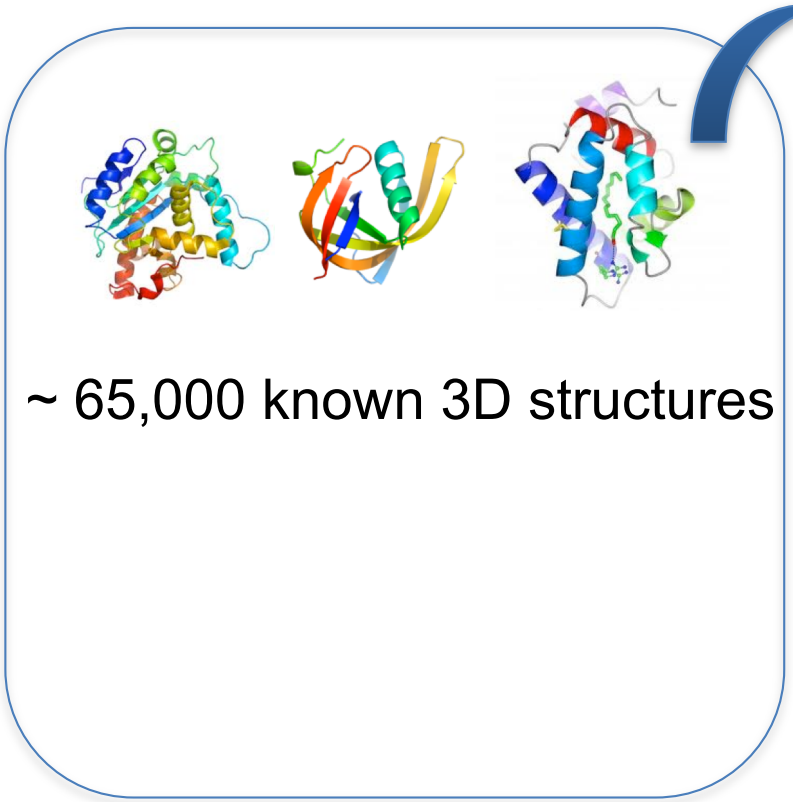
~ 65,000 known 3D structures

Extract sequence

HAPTLVRDC…….

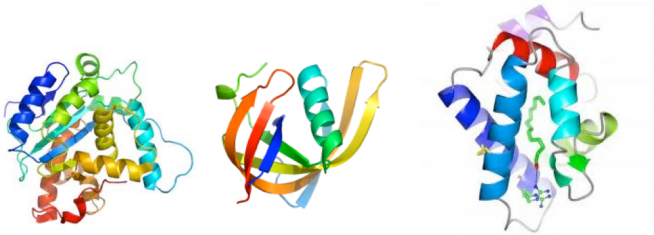
PSI-
Blast

Phyre2



Hidden Markov model
for sequence of KNOWN structure

Phyre2

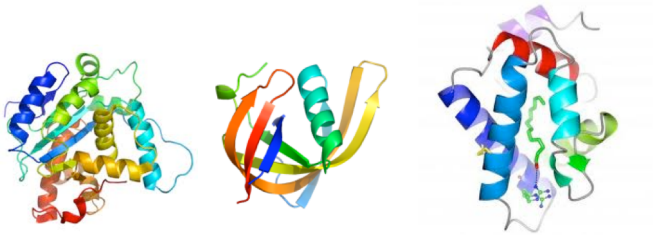


~ 65,000 known 3D structures

HMM HMM HMM

~ 65,000 hidden Markov models

Phyre2



~ 65,000 known 3D structures

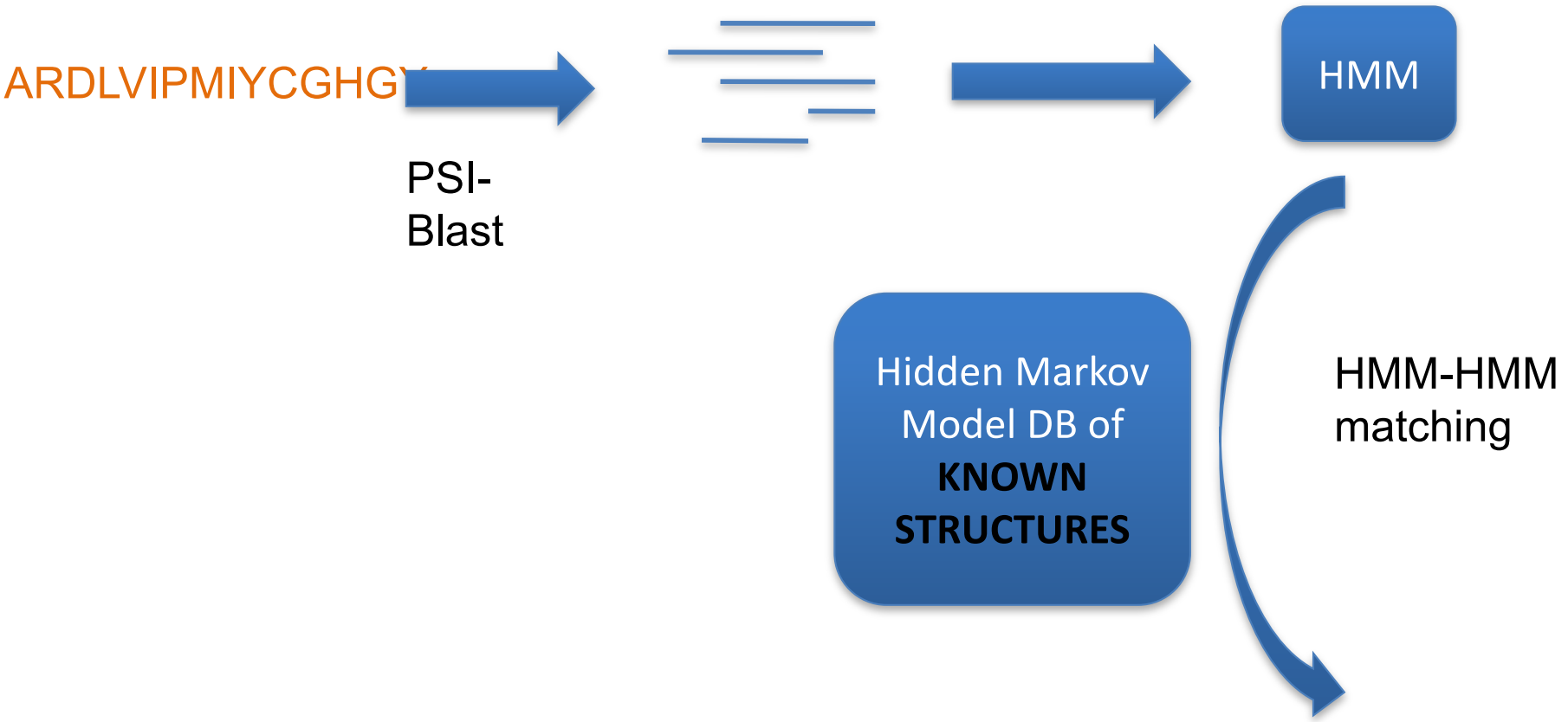
Hidden Markov Model Database of **KNOWN STRUCTURES**

Phyre2

Capture the mutational propensities at each position in the protein

An evolutionary fingerprint

Phyre2



Alignments of user sequence to known structures ranked by confidence.

ARDL--VIPMIYCGHGY

AFDLCD**LIPV--CGMAY**

Sequence of known structure

Phyre2

ARDLVIPMIYCGHGY

PSI-
Blast

HMM

HMM-HMM
matching

Hidden Markov
Model DB of
**KNOWN
STRUCTURES**

ARDL--VIPMIYCGHGY
AFDLCDLIPV--CGMAY

Sequence of known structure

3D-Model

Phyre2

ARDLVIPMIYCGHGY

PSI-
Blast

HMM

Hidden Markov Model DB of **KNOWN STRUCTURES**

HMM-HMM matching

**Very powerful –
able to reliably detect extremely remote homology**

Routinely creates accurate models even when sequence identity is <15%

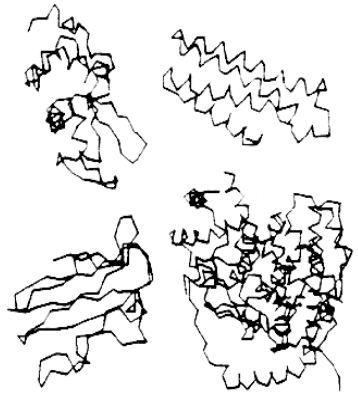
3D-Model

ARDL--VIPMIYCGHGY
AFDLCDLIPV--CGMAY

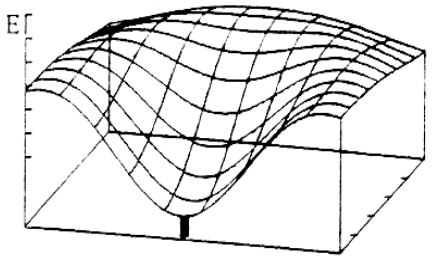
Sequence of known structure

Knowledge-based Potentials used by Fold Recognition methods

Calculation of Mean Force Potentials



Databank of 3D structures

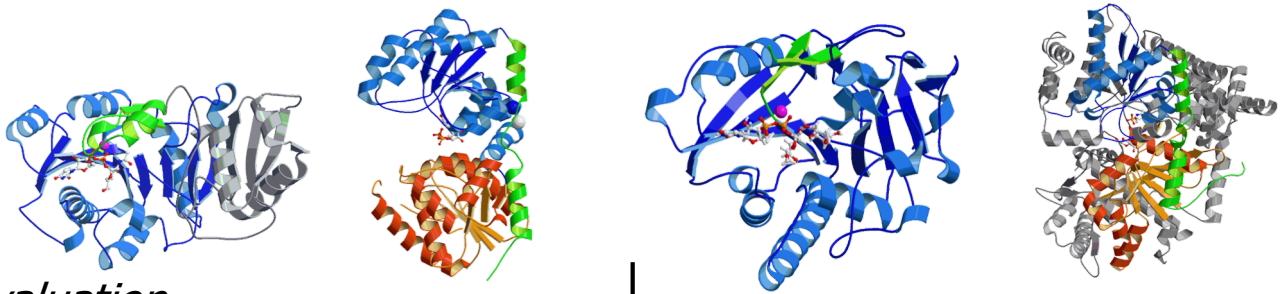


Mean force potential derived from the databank

Building up a proposed model from amino acid sequence that is based on a real protein structure

SDVDIEAGQTLVQVVNISNGETWVAIQLPAQYRSFDLVFENVSPSTSGSVLVAQMA
PQSGGVYGSNYSGSGWGNDLGGGGFYGYSEAKWMCLWPANRSGPNSKTGIYG
TCKLMNLNQSNAVPSVTSNLFAPTAYKNEPGYANVGGCCQKIRGLASSIQFAFALH
GGNVPQNTDTFSGGTIKVYGWN

3D-fold calculation based on known structures



Model quality evaluation

pair
residue-residue interactions

surface
residue-solvent interactions

pair/surface
residue-residue and residue-solvent interaction

"Quality" scores

Phyre2

Protein Homology/analogY Recognition Engine V 2.0

Subscribe to Phyre at Google Groups
Email: Subscribe
Visit Phyre at Google Groups

What's New in Phyre2

E-mail Address	
Optional Job description	
Amino Acid Sequence	
Modelling Mode	Normal ◉ Intensive ○
	Phyre Search Reset

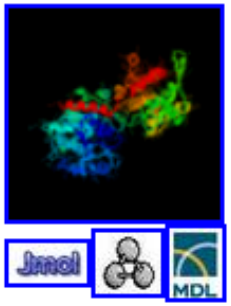
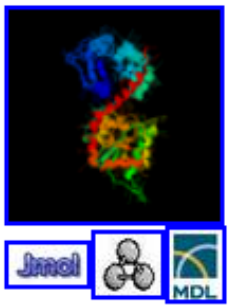
553492 submissions since Feb 14 2011

Glykogensynthasa – rodina GT3 (v rodině v době analýzy nebyla vyřešena 3D-struktura)

http://www.sbg.bio.ic.ac.uk/phyre/qphyre_output/95cbaa7600a9bfff/summary.html

To predict functional residues and GO classification, try ConFunc

ognition

nments	SCOP Code	View Model	E-value	Estimated Precision	BioText	Fold/PDB descriptor	Superfamily
	d2bisa1 (length:437) **18% i.d.**		3.9e-36	100 %	n/a	UDP-Glycosyltransferase/glycogen phosphorylase	UDP-Glycosyltransferase/g phosphorylase
	d1rzua (length:477) **14% i.d.**		6.1e-36	100 %	n/a	UDP-Glycosyltransferase/glycogen phosphorylase	UDP-Glycosyltransferase/g phosphorylase
	c3c48A (length:438) **11% i.d.**		6.1e-31	100 %	n/a	**PDB header:**transferase	**Chain:** A: **PDB Molecule:**predicted glycosyltransferases;

A co protein, který nemá v sekvenčních databázích žádný homolog?

RS-20L
No sequence homology
in databases !

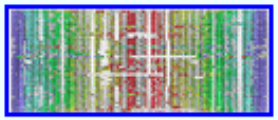
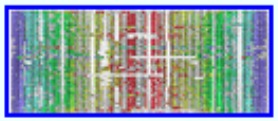
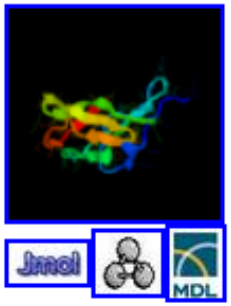
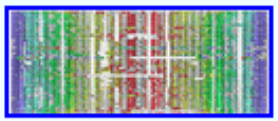
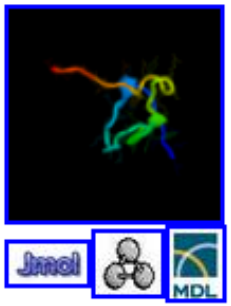
Quickphyre results for job rs20____ - Mozilla Firefox

Soubor Úpravy Zobrazit Historie Záložky Nástroje Nápověda

http://www.sbg.bio.ic.ac.uk/phyre/qphyre_output/964f0704319f5953/summary.html

Nejnavštěvovanější Jak začít Přehled zpráv http://www.ncbi.nlm.... http://www.glycoscie... CHMI Radar Departme...

Fold Recognition

View Alignments	SCOP Code	View Model	E-value	Estimated Precision	BioText	Fold/PDB descriptor	Superfamily	Fa
	d1eh9a2 (length:67) **24% i.d.**		50	0 %	n/a	Glycosyl hydrolase domain	Glycosyl hydrolase domain	alpha-Amylas C-termi beta-sh domain
	c2fsdA (length:142) **19% i.d.**		50	0 %	n/a	**PDB header:**virus/viral protein	**Chain:** A: **PDB Molecule:**putative baseplate protein;	**PDB Tit** commo the rece binding domain lactoco phages crystal s of the h domain phage b
	c2ct4A (length:70) **11% i.d.**		56	0 %	n/a	**PDB header:**signaling protein	**Chain:** A: **PDB Molecule:**cdc42-interacting protein 4;	**PDB Tit** solution strutcur sh3 dor the cdc interact protein

Homology modeling

•přiložení cílové sekvence se sekvencí homologního proteinu se známou 3D strukturou

• extrakce uhlíkové páteře ze struktury templátu a umístění postranních řetězců

•modelování otoček a smyček

• minimalizace energie

•validace modelované struktury

MODELLER

Mostly used program in academic environment for serious homology modeling

SWISS-MODEL

An automated knowledge-based protein modelling server

```
- Start SMR-Pipeline in automated mode on BC2-cluster at Thu May  2 08:51:47 2013


- Start BLAST for highly similar template structure identification
- No suitable templates found!

- Run HHSearch to detect remotely related template structures
- Unfortunately, we could not identify useful template structures

- For troubleshooting, please see our article in Nature Protocols:

- Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. and Schwede, T. (2009).
Protein structure homology modelling using SWISS-MODEL Workspace.  Nature Protocols, 4, 1.
```

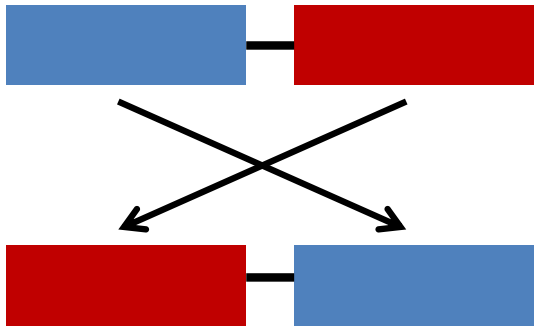
What are protein domains?

Since the first protein structures were solved, it was apparent that the polypeptide chain could often fold into one or more distinct regions of structure. Such substructures, or domains, are considered as the basic units of folding, function and evolution and often have similar chain topologies (Holm & Sander, 1994). Protein domains are often considered as independent or, at the least, semi-independent units, able to fold and in some cases retain function if separated from the parent chain. The independent, modular nature of many domains means that they can often be found in proteins with the same domain content, but in different orders, or in different proteins in combination with entirely different domain structures.

The concept of the protein domain is just as valid at the sequence level as the structural level. This can be shown by the fact that the alignment of sequences containing similar domains, but in different orders can result in poor and possibly misleading alignments.
However alignment of the shared domains if extracted from the parent sequence may reveal a high level of sequence similarity, demonstrating an evolutionary link between the domain sequences.

PLLSASIVSAPVVTSETYVDIPGLYLDVAKAGIRDGKLQVILNVPTPYATGNNFPGIYFAIATNQGVVADGCFTYSSKV
PESTGRMPFTLVATIDVGSGVTFVKGQWKSVRGSAMHIDSYASLSAIWGTAAPSSQGSGNQGAETGGTGAGNIG
GGGERDGTFNLPPHIKFGVTALTHAANDQTIDIYIDDDPKPAATFKGAGAQDQNLGTKVLDSGNGRVRVIVMANGR
PSRLGSRQVDIFKKSYFGIIGSEDGADDDYNDGIVFLNWPLG

ERDGTFNLPPHIKFGVTALTHAANDQTIDIYIDDDPKPAATFKGAGAQDQNLGTKVLDSGNGRVRVIVMANGRPSR
LGSRQVDIFKKSYFGIIGSEDGADDDYNDGIVFLNWPLGPLLSASIVSAPVVTSQTYVDIPGLYLDVAKAGIRDGKLQ
VILNVPTPYATGNNFPGIYFAIATNQGVVADGCFTYSSKVPESTGRMPFTLVATIDVGSGVTFVKGQWKSVRGSAM
HIDSYASLSAIWGTAAPSSQGSGNQGAETGGTGAGNIGGGGKLAAALEIKRASQPELAPEDPEDVEHHHHHH



```
#
#=======================================

EMBOSS_001         1 --------------------------------------------------      0

EMBOSS_001         1 ERDGTFNLPPHIKFGVTALTHAANDQTIDIYIDDDPKPAATFKGAGAQDQ     50

EMBOSS_001         1 --------------------------------------------------      0

EMBOSS_001        51 NLGTKVLDSGNGRVRVIVMANGRPSRLGSRQVDIFKKSYFGIIGSEDGAD    100

EMBOSS_001         1 ---------------PLLSASIVSAPVVTSETYVDIPGLYLDVAKAGIRD     35
                                    |||||||||||||||:||||||||||||||||||
EMBOSS_001       101 DDYNDGIVFLNWPLGPLLSASIVSAPVVTSQTYVDIPGLYLDVAKAGIRD    150

EMBOSS_001        36 GKLQVILNVPTPYATGNNFPGIYFAIATNQGVVADGCFTYSSKVPESTGR     85
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       151 GKLQVILNVPTPYATGNNFPGIYFAIATNQGVVADGCFTYSSKVPESTGR    200

EMBOSS_001        86 MPFTLVATIDVGSGVTFVKGQWKSVRGSAMHIDSYASLSAIWGTAAPSSQ    135
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       201 MPFTLVATIDVGSGVTFVKGQWKSVRGSAMHIDSYASLSAIWGTAAPSSQ    250

EMBOSS_001       136 GSGNQGAETGGTGAGNIGGGGERDGTFNLPPHIKFGVTALTHAANDQTID    185
                     ||||||||||||||||||||
EMBOSS_001       251 GSGNQGAETGGTGAGNIGGGG----------------------------    271

EMBOSS_001       186 IYIDDDPKPAATFKGAGAQDQNLGTKVLDSGNGRVRVIVMANGRPSRLGS    235
                            |.||:          |..|                 |:|
EMBOSS_001       272 -------KLAAA----------LEIK----------------RAS----    283

EMBOSS_001       236 RQVDIFKKSYFGIIGSEDGADDDYNDGIVFLNWPLG    271
                      |.:         ::.||..|.:.:.       ...
EMBOSS_001       284 -QPE---------LAPEDPEDVEHHH-------HHH    302
```

domain boundary/disorder/globularity prediction:

LinkPred (at NIMR)

SnapDRAGON domain boundary prediction (at NIMR)

PASS (at RIKEN)

Domain Guess by Size (DGS) (at NCBI)

UMA (Udwary-Merski Algorithm) (at Johns Hopkins Univ.)

DomPred (at UCL)

Domain boundary prediction based on entropy profile (at IPR, Moscow]

GlobPlot (at EMBL) Prediction of protein disorder/order/globularity

DisEMBL (at EMBL) Protein disorder prediction

Příklad: Předpověď spojovacích úseků mezi doménami - program DomCut

Předpovídání doménových a spojovacích oblastí v sekvencích proteinů

Domény = funkční jednotky, z nichž jsou bílkoviny složeny

Linker = spojovací úsek aminokyselinového řetězce spojujícího dvě sousední domény

DomCut

- Metoda programu DomCut vychází ze statisticky potvrzeného předpokladu odlišného složení doménových a linkerových úseku v řetězcích aminokyselin.

- Jestliže známe relativní frekvence výskytu jednotlivých AK v linkerových a doménových úsecích, můžeme u neznámé sekvence odhadnout zda je ten či onen úsek spíše linker nebo doména, podle toho, zda v něm převládají AK vyskytující se více v linkerech nebo v doménách.

- Pro vyjádření přednosti AK v linkerech je definován tzv. „linker index“ *S* (f_i^{linker} a f_i^{domain} je frekvence zastoupení aminokyseliny *i* v úsecích linkeru a domény)
- :

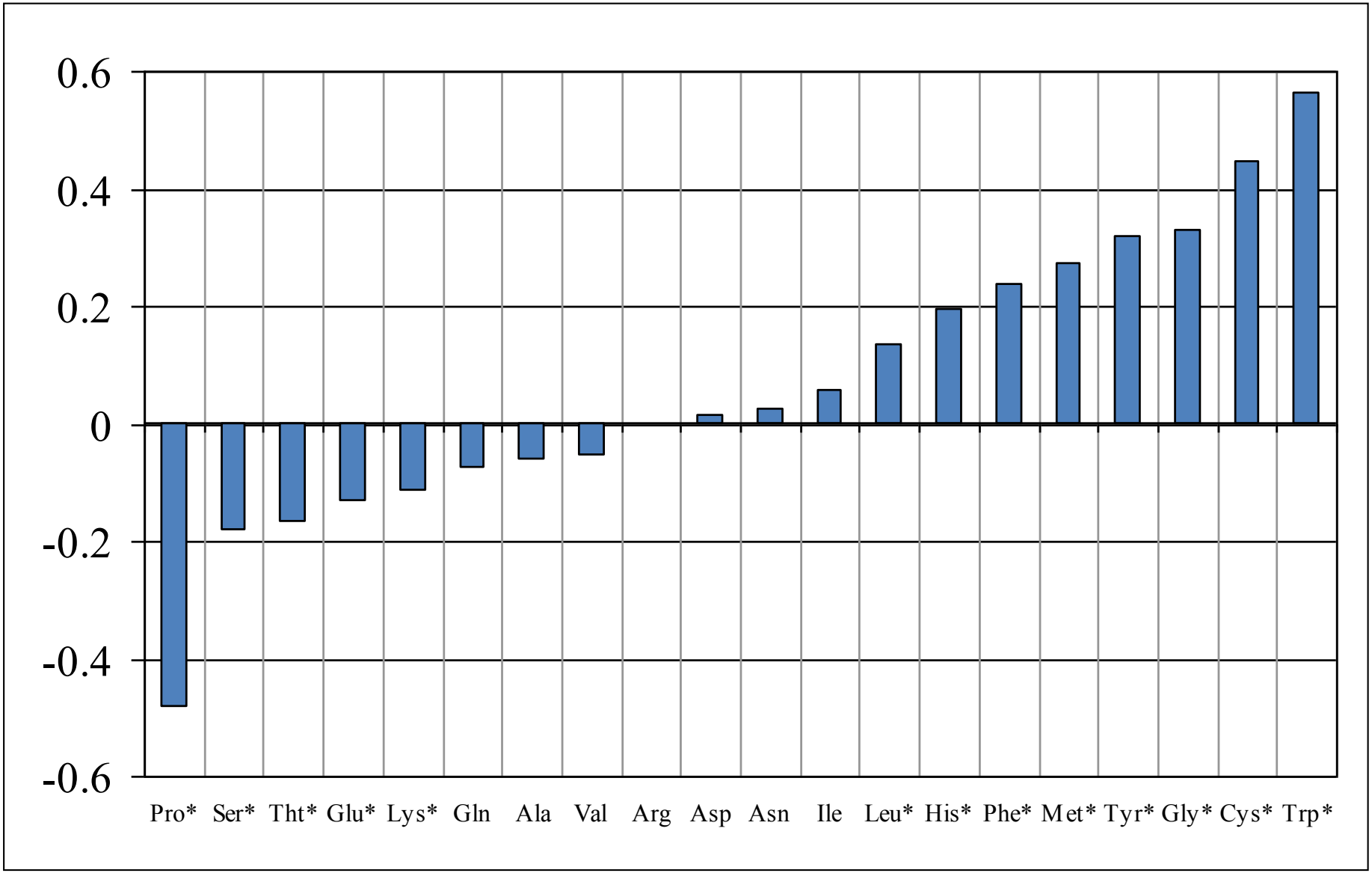
$$S_i = - \ln \frac{f_i^{\text{linker}}}{f_i^{\text{domain}}}$$

Četnost výskytu jednotlivých aminokyselin v doménách a linkerech

- Záporná hodnota znamená, že daná AK se častěji vyskytuje v linkerových úsecích
- Výjimku tvoří Gly, který je hojně zastoupený v doménách, ale je častým prvkem v linkerových oblastech – zajišťuje „ohebnost“

Aminokyselina		f_i^{linker} **(%)**	f_i^{domain} **(%)**	S_i	**Aminokyselina**		f_i^{linker} **(%)**	f_i^{domain} **(%)**	S_i
Proline	Pro*	7.95	4.93	-0.478	Asparagine	Asn	4.29	4.41	0.027
Serine	Ser*	8.32	6.97	-0.177	Isoleucine	Ile	4.86	5.16	0.060
Threonine	Thr*	6.68	5.67	-0.163	Leucine	Leu*	7.62	8.75	0.138
Glutamic acid	Glu*	7.53	6.62	-0.128	Histidine	His*	2.13	2.59	0.195
Lysine	Lys*	6.30	5.64	-0.112	Phenylalanine	Phe*	2.92	3.71	0.240
Glutamine	Gln	4.35	4.04	-0.073	Methionine	Met*	1.47	1.94	0.275
Alanine	Ala	7.03	6.64	-0.058	Tyrosine	Tyr*	2.49	3.44	0.322
Valine	Val	7.33	6.96	-0.052	Glycine	Gly*	5.46	7.60	0.331
Arginine	Arg	5.39	5.39	0.000	Cysteine	Cys*	1.62	2.53	0.447
Aspartic acid	Asp	5.39	5.47	0.016	Thryptophan	Trp*	0.89	1.56	0.564

DomCut - grafické znázornění S_i faktoru



DomCut – příklad predikce spojovacích úseků

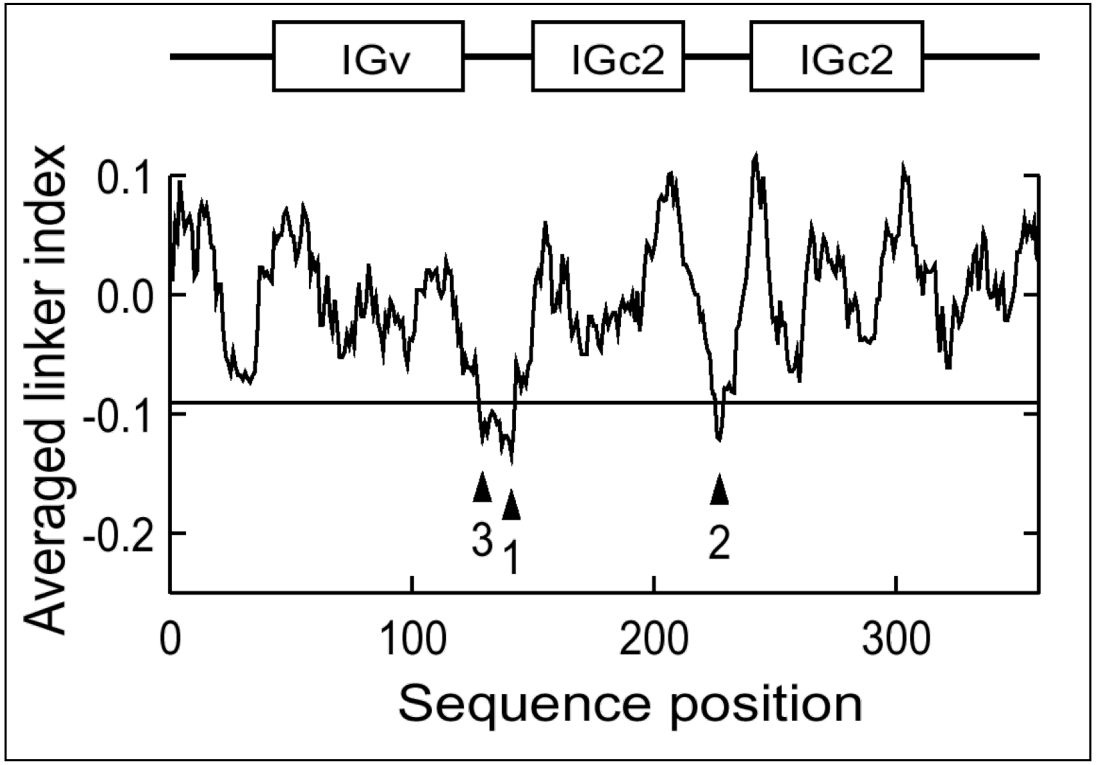
Aminokyselinová sekvence Q24372

- není podobná s žádnou z referenční množiny (podobnost <40%)

- úseky linkerů mezi doménami odpovídají jasně odhadům (prohlubně pod prahovou hodnotou –0,09)

Záznam trEMBL:

Lachesin, Contains 2 Ig-like C2-type domains, 1 Ig-like V-type domain.

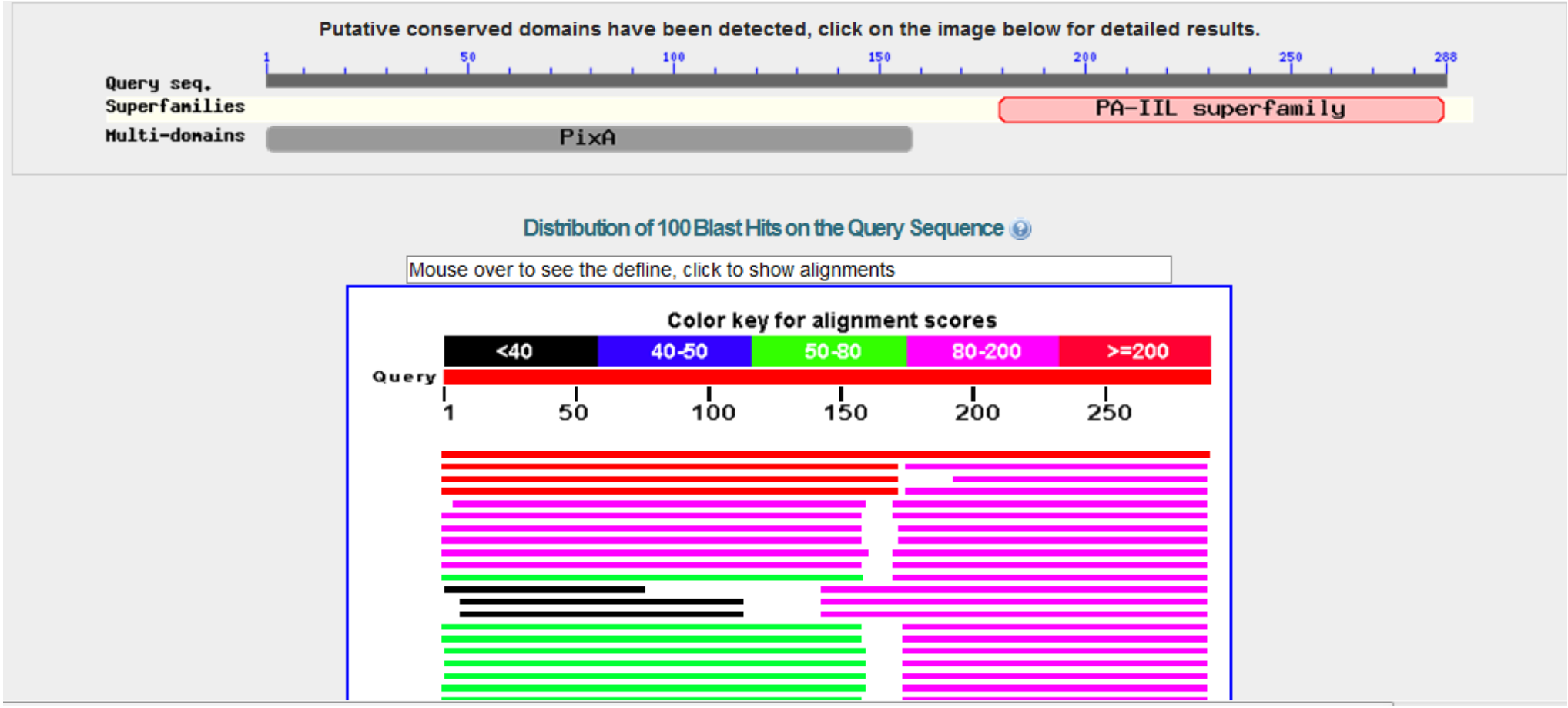


Domény předpovídají i programy používané primárně pro jiné účely na základě podobnosti s dosud identifikovanými doménami/funkčními jednotkami

NCBI – Blast (Basic Local Alignment Search Tool) (National Centre for Biotechnology Information)

Prohledávání databází známých aminokyselinových sekvencí

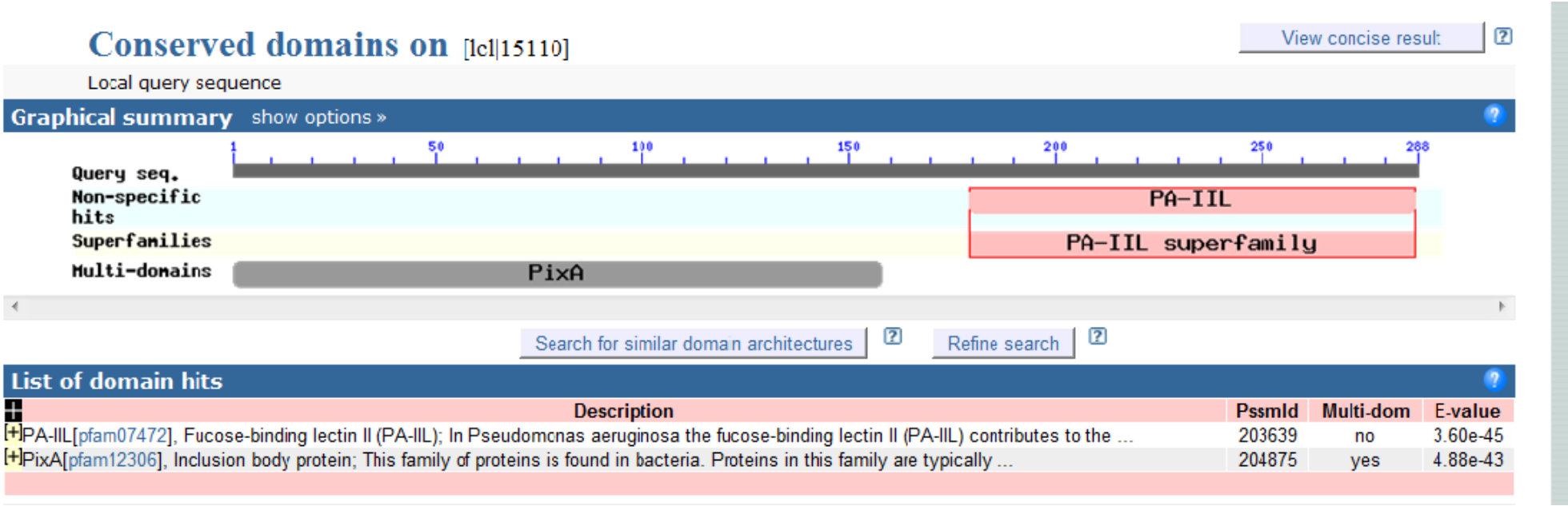
- **celý protein**



NCBI – Blast

Prohledávání databází známých aminokyselinových sekvencí

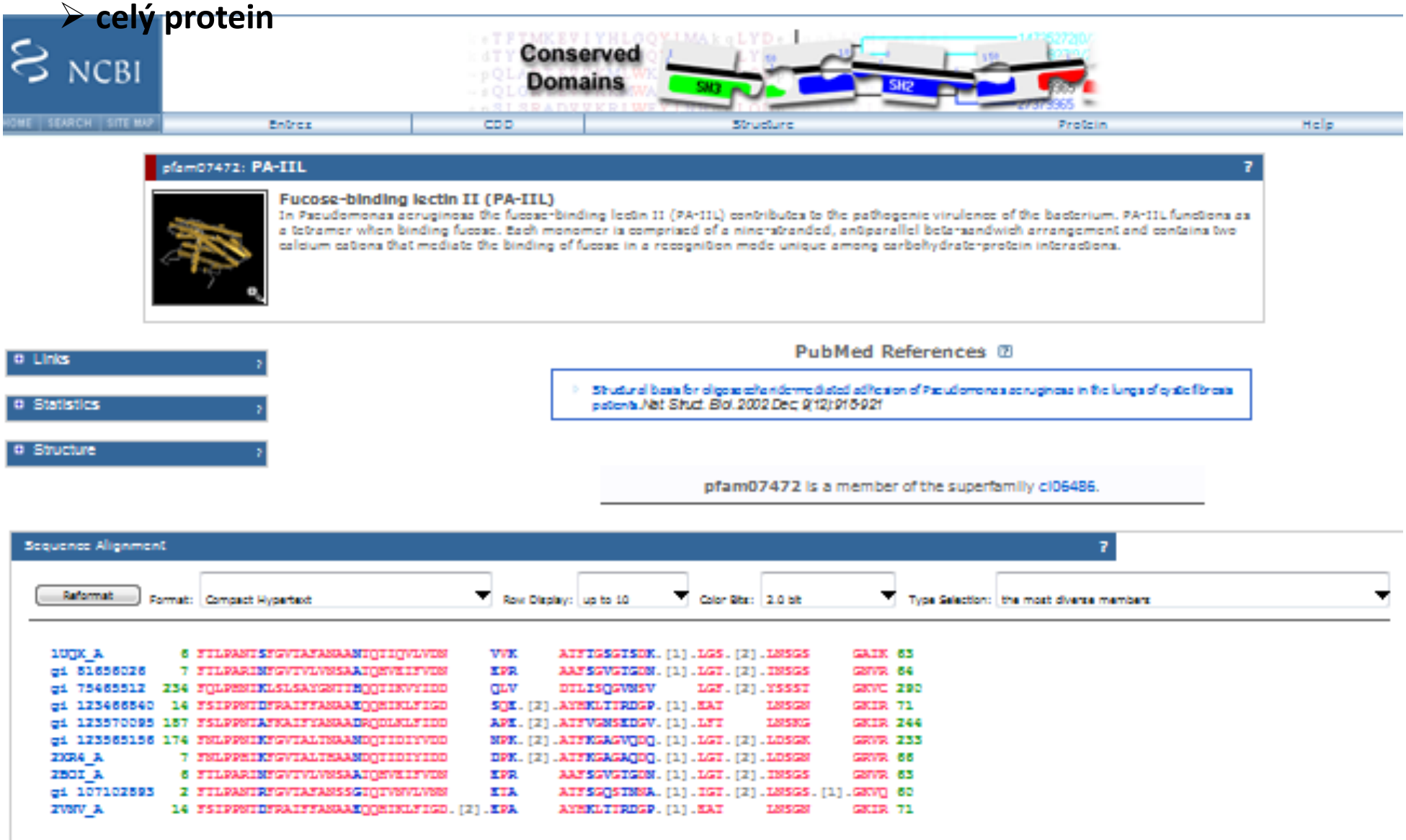
- **celý protein**



NCBI – Blast

Prohledávání databází známých aminokyselinových sekvencí

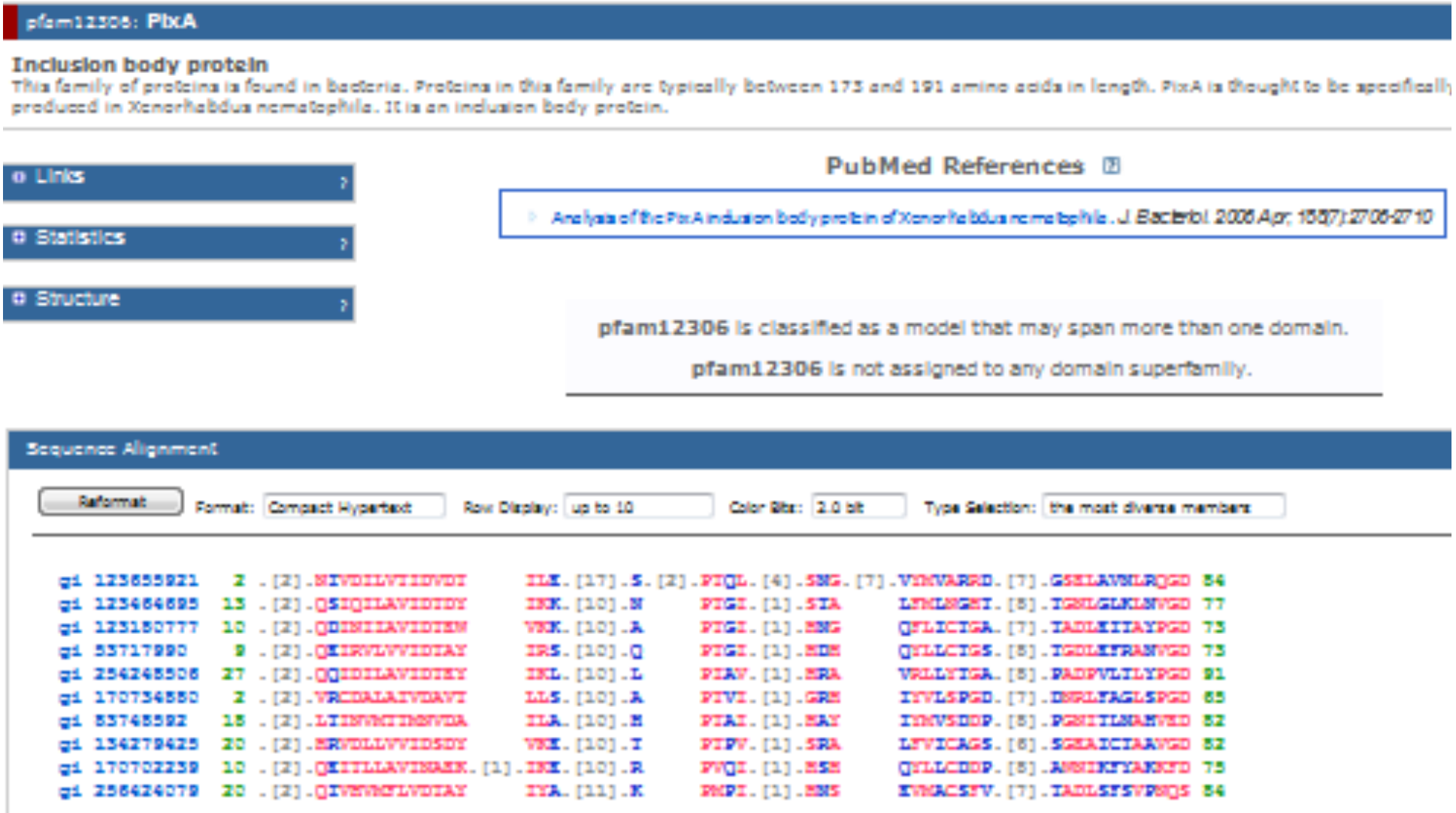
- **celý protein**



NCBI – Blast

Prohledávání databází známých aminokyselinových sekvencí

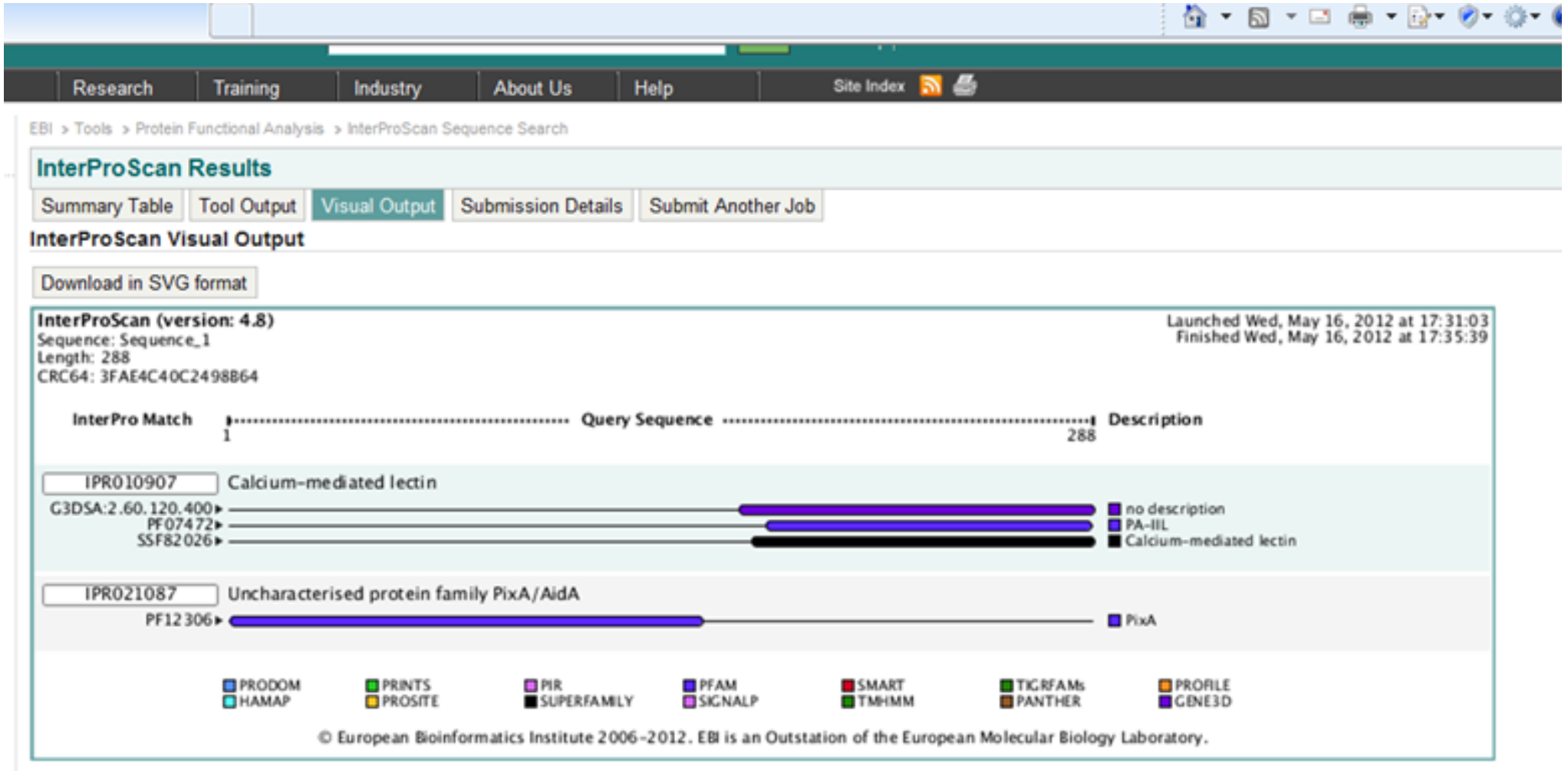
- **celý protein**



InterPro protein sequence analysis & classification

InterPro is an integrated database of predictive protein signatures used for the classification and automatic annotation of proteins and genomes. InterPro classifies sequences at superfamily, family and subfamily levels, predicting the occurrence of functional domains, repeats and important sites. InterPro adds in-depth annotation, including GO terms, to the protein signatures.

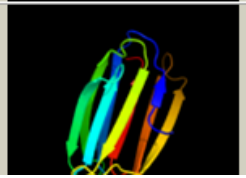
European Bioinformatics Institute - http://www.ebi.ac.uk/



Proč potřebujeme predikci domén?

- Prohledávání sekvenčních databází bez predikce domén může být neúspěšné
- Automatická predikce struktury se zaměří jen na nejlépe „definovanou“ část
- ….

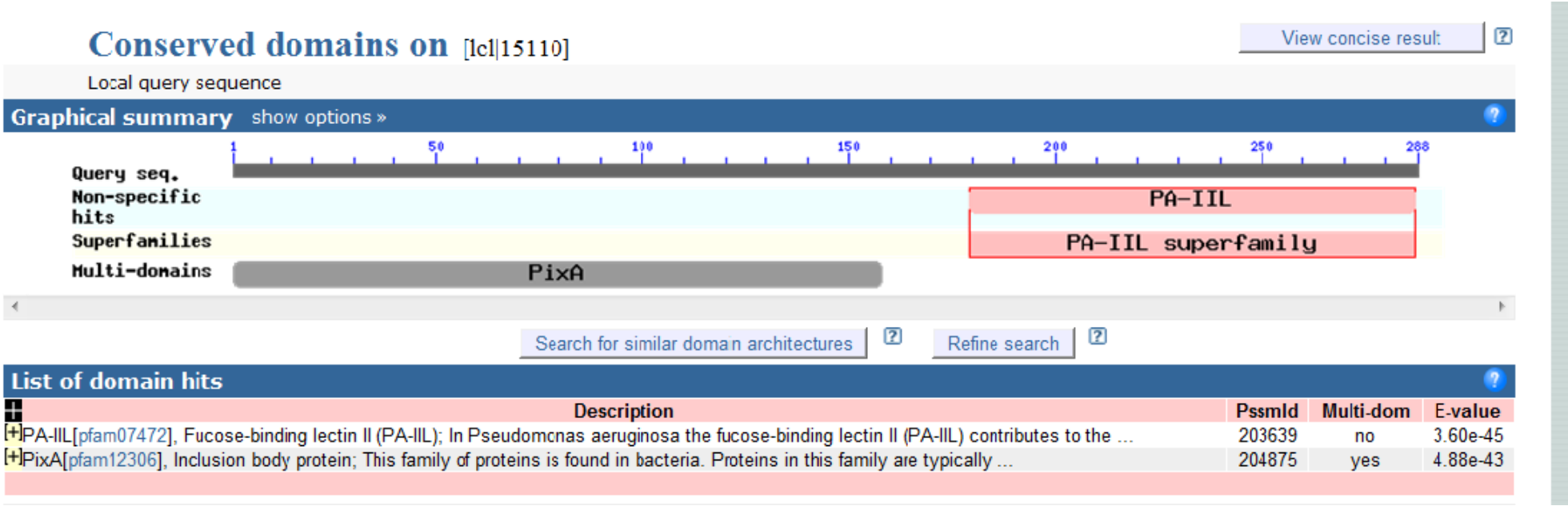
Phyre – whole protein http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/a132b051273537c4/summary.htm

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c2vnvC_	Alignment		100.0	60	**PDB header:**sugar-binding protein **Chain:** C: **PDB Molecule:**bcla; **PDBTitle:** crystal structure of bcla lectin from burkholderia2 cenocepacia in complex with alpha-methyl-mannoside at 1.73 angstrom resolution
2	c2xr4A_	Alignment		100.0	43	**PDB header:**sugar binding protein **Chain:** A: **PDB Molecule:**lectin; **PDBTitle:** c-terminal domain of bc2l-c lectin from burkholderia cenocepacia
3	d2chha1	Alignment		100.0	37	**Fold:**Calcium-mediated lectin **Superfamily:**Calcium-mediated lectin **Family:**Calcium-mediated lectin

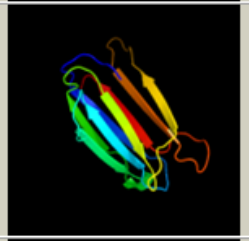
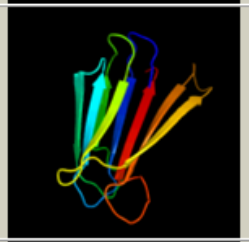
NCBI – Blast

Prohledávání databází známých aminokyselinových sekvencí

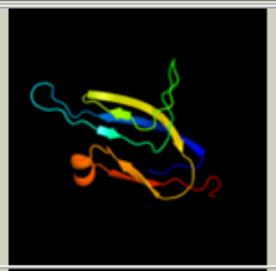
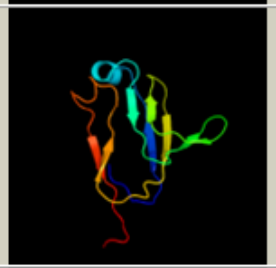
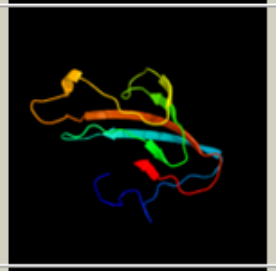
- **celý protein**



Phyre – C-term http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/e332b1ecabb8d0a6/summary.html

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c2xr4A_	Alignment		100.0	44	**PDB header:**sugar binding protein **Chain:** A: **PDB Molecule:**lectin; **PDBTitle:** c-terminal domain of bc2l-c lectin from burkholderia cenocepacia
2	c2vnvC_	Alignment		100.0	62	**PDB header:**sugar-binding protein **Chain:** C: **PDB Molecule:**bcla; **PDBTitle:** crystal structure of bcla lectin from burkholderia2 cenocepacia in complex with alpha-methyl-mannoside at 1.73 angstrom resolution
3	d1uzva_	Alignment		100.0	30	**Fold:**Calcium-mediated lectin **Superfamily:**Calcium-mediated lectin **Family:**Calcium-mediated lectin

Phyre – n-term http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/e332b1ecabb8d0a6/summary.html

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c1sddB_	Alignment		83.7	9	**PDB header:**blood clotting **Chain:** B: **PDB Molecule:**coagulation factor v; **PDBTitle:** crystal structure of bovine factor vai
2	c3cdzB_	Alignment		76.1	6	**PDB header:**blood clotting **Chain:** B: **PDB Molecule:**coagulation factor viii light chair **PDBTitle:** crystal structure of human factor viii
3	d1kbva2	Alignment		68.0	13	**Fold:**Cupredoxin-like **Superfamily:**Cupredoxins **Family:**Multidomain cupredoxins

Swissprot – whole protein

BIOZENTRUM
Universität Basel
The Center for Molecular Life Sciences

SWISS-MODEL Workspace

Modelling Tools Repository Documentation

[myWorkspace] [login]

Workunit: P000007 - Overview

1 288

Print/Save this page as

Model Summary

Model information:
Modelled residue range: 169 to 288
Based on template: [2vnvD] (1.7 Å)
Sequence Identity [%]: 56.35
Evalue: 0.00e-1

Quality information: [details]
QMEAN Z-Score: -0.71

Quaternary structure information: [details]
Template (2vnv): DIMER
Model built: SINGLE CHAIN

Ligand information: [details]
Ligands in the template: CA: 3, MMA: 1, SO4: 1.
Ligands in the model: CA: 2

logs: [Templates] [Alignment] [Modelling]
display model: as [pdb] - as [DeepView project] - in [AstexViewer]
download model: as [pdb] - as [Deepview project] - as [text]

Global Model Quality Estimation [+/-]

http://swissmodel.expasy.org/workspace/index.php?userid=michaw@chemi.muni.cz&key=0f449e99bc0176edfa75fba19b2d96e4&func=workspace_modelling&prjid=P000007

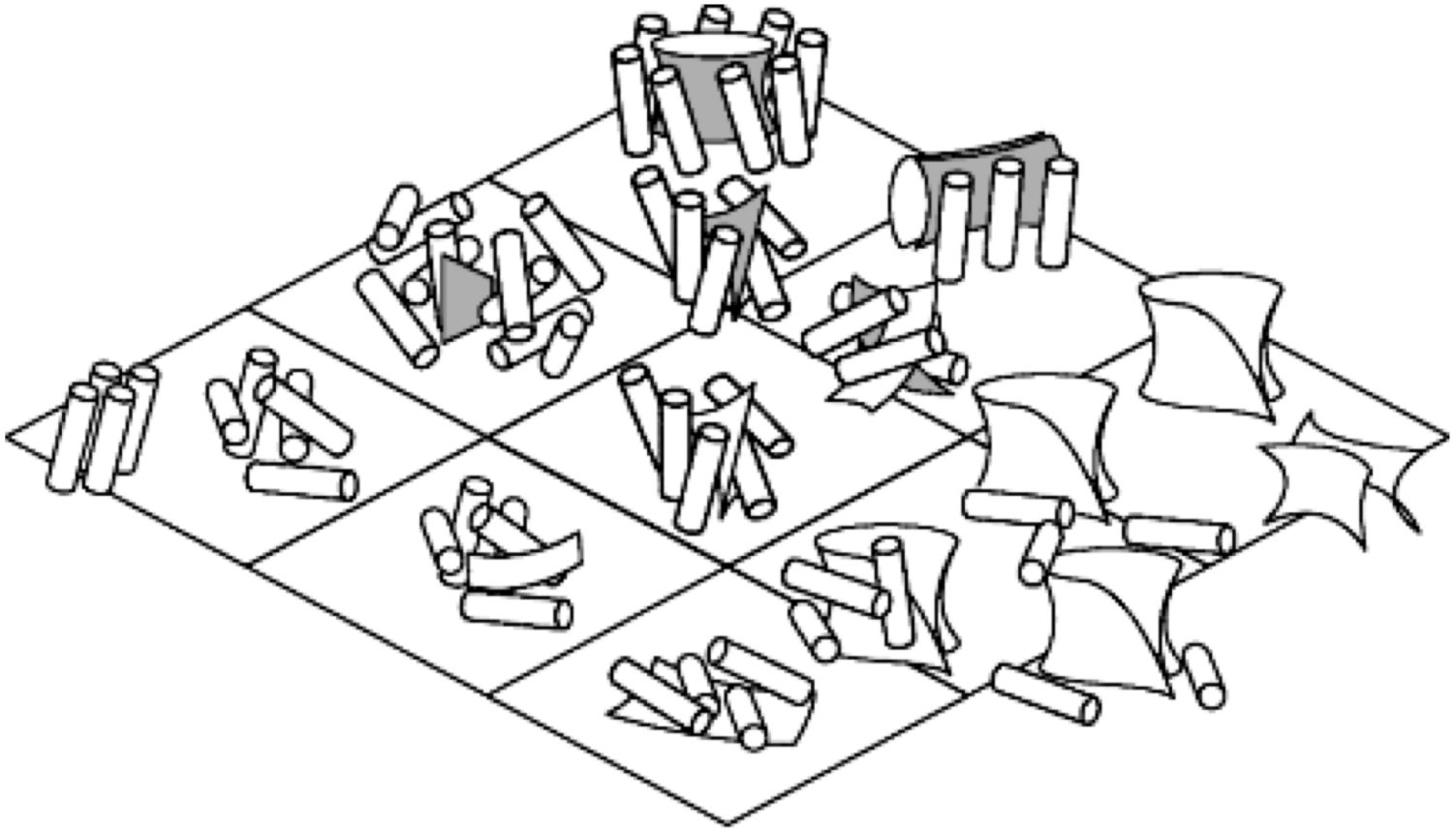
Swissprot - only N terminal part

Computation of this workunit has stopped.

Please see the following log report for details:

```
Started: Thu May 17 15:21:24 2012 (sms_automode_2011)
Reading user input sequence


 - Start SMR-Pipeline in automated mode on BC2-cluster at Thu May 17 13:21:24 2012


 - Start BLAST for highly similar template structure identification
 - No suitable templates found!

 - Run HHSearch to detect remotely related template structures
 - Unfortunately, we could not identify useful template structures

 - For troubleshooting, please see our article in Nature Protocols:

 - Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. and Schwede, T. (2009).
Protein structure homology modelling using SWISS-MODEL Workspace.  Nature Protocols, 4, 1.

 - Workspace Pipeline parameter
   Cut-off parameters to model the target based on a BLAST target-template alignment
    Evalue :                                    0.0001
    Minimum Template size (aa) for ranking :    25
    Minimum Sequence identity :                 60
   Cut-off parameters to model the target based on a HHSearch target-template alignment
    Evalue :                                    0.0001
    Probability :                               50
    MAC :                                       0.3
   Parameters for model selection
    Minimal number of uncovered target
     residues after BLAST to run HHSEARCH :    50
    Minimal number of uncovered target
     residues to model an additional template :  25
 - Finish SMR-Pipeline in automated mode on BC2-cluster at Thu May 17 13:35:44 2012
```

Structural classes of proteins



Others:
Multi-domain, membrane and cell surface, small proteins, peptides and fragments, designed proteins, ..

Databases of Protein Folds

SCOP (http://scop.berkeley.edu/) - **known** domain structure

- Structural Classification of Proteins
- Class-Fold-Superfamily-Family
- Manual assembly by inspection

Superfamily (http://supfam.org/SUPERFAMILY/) - **predicted** domain structures

- HMM models for each SCOP fold
- Fold assignments to all genome ORFs
- Assessment of specificity/sensitivity of structure prediction
- Search by sequence, genome and keywords

CATH + Gene3D (http://www.biochem.ucl.ac.uk/bsm/cath/) - **both**

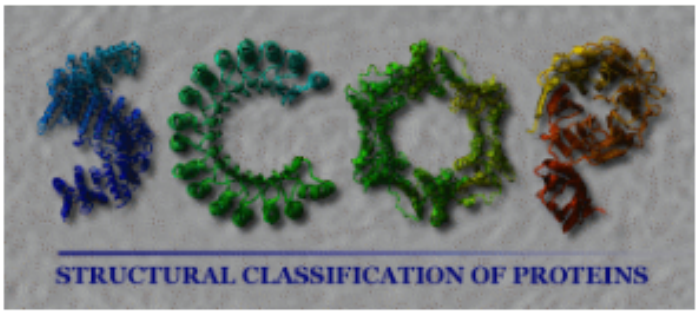
- Class - Architecture - Topology - Homologous Superfamily
- Manual classification at Architecture level
- Automated topology classification using SSAP (Orengo & Taylor)

PDB eFold (http://www.ebi.ac.uk/msd-srv/ssm/)

- Fully automated using the DALI algorithm (Holm & Sander)

Pfam (http://pfam.xfam.org)- domain sequences (MSA, HMM)

SCOP **S**tructural **C**lassification **o**f **P**roteins
(http://scop.mrc-lmb.cam.ac.uk/scop)



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. Acid Res.* advance access, doi:10.1093/nar/gkm993. [PDF].

Access methods

- Enter SCOP at the **top of the hierarchy**
- Keyword search of SCOP entries
- SCOP parseable files (MRC site)
- All SCOP releases and reclassified entry history (MRC site)
- **pre-SCOP - preview of the next release**
- SCOP domain sequences and pdb-style coordinate files (ASTRAL)
- Hidden Markov Model library for SCOP superfamilies (SUPERFAMILY)

The **SCOP** database, created by manual inspection and abetted by a battery of automated methods, aims to provide a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known. http://scop.mrc-lmb.cam.ac.uk/scop

Family: *Clear evolutionarily relationship*

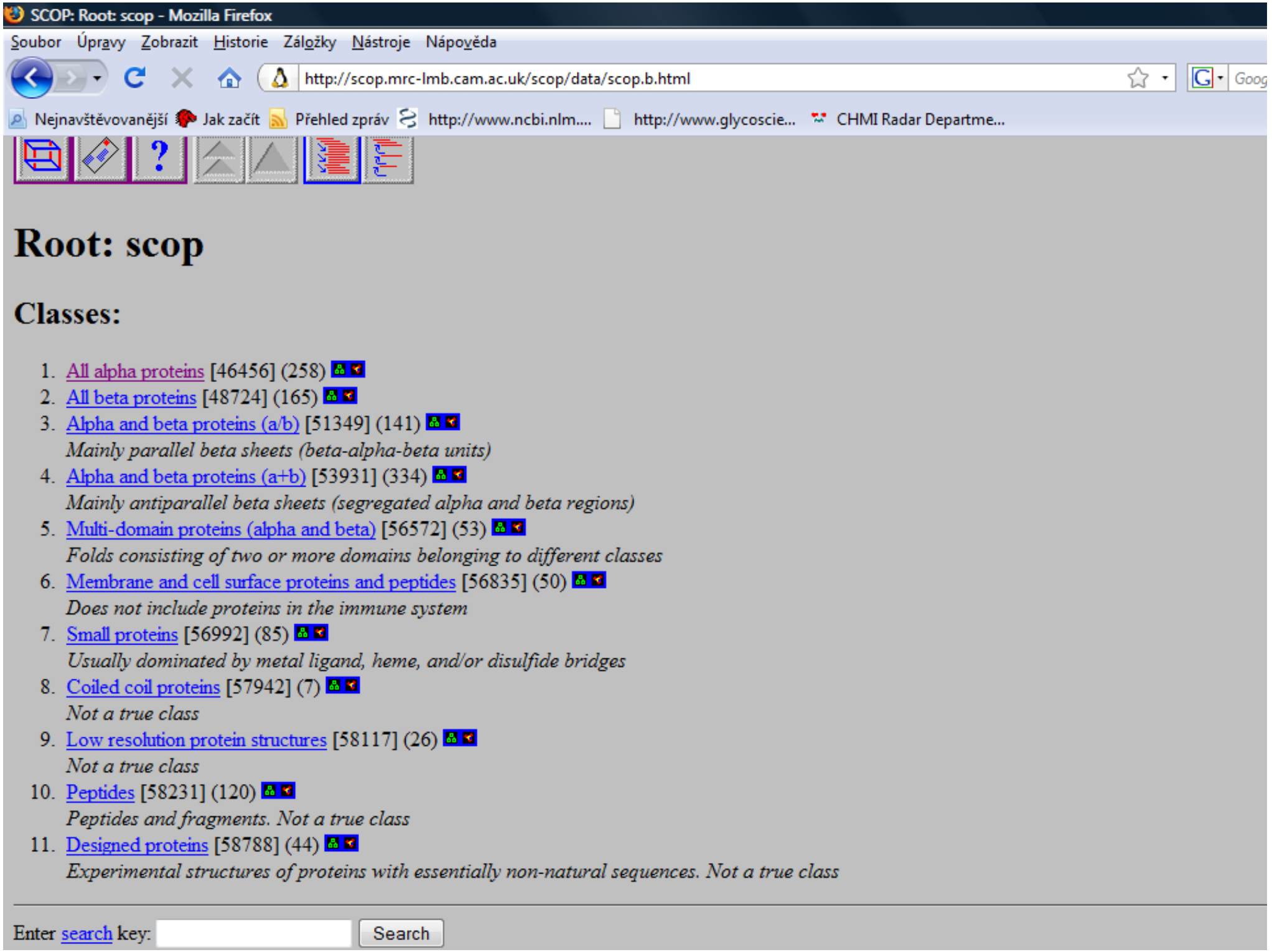
Proteins clustered together into families are clearly evolutionarily related. Generally, this means that pairwise residue identities between the proteins are 30% and greater. *However, in some cases similar functions and structures provide definitive evidence of common descent in the absense of high sequence identity; for example, many globins form a family though some members have sequence identities of only 15%.*

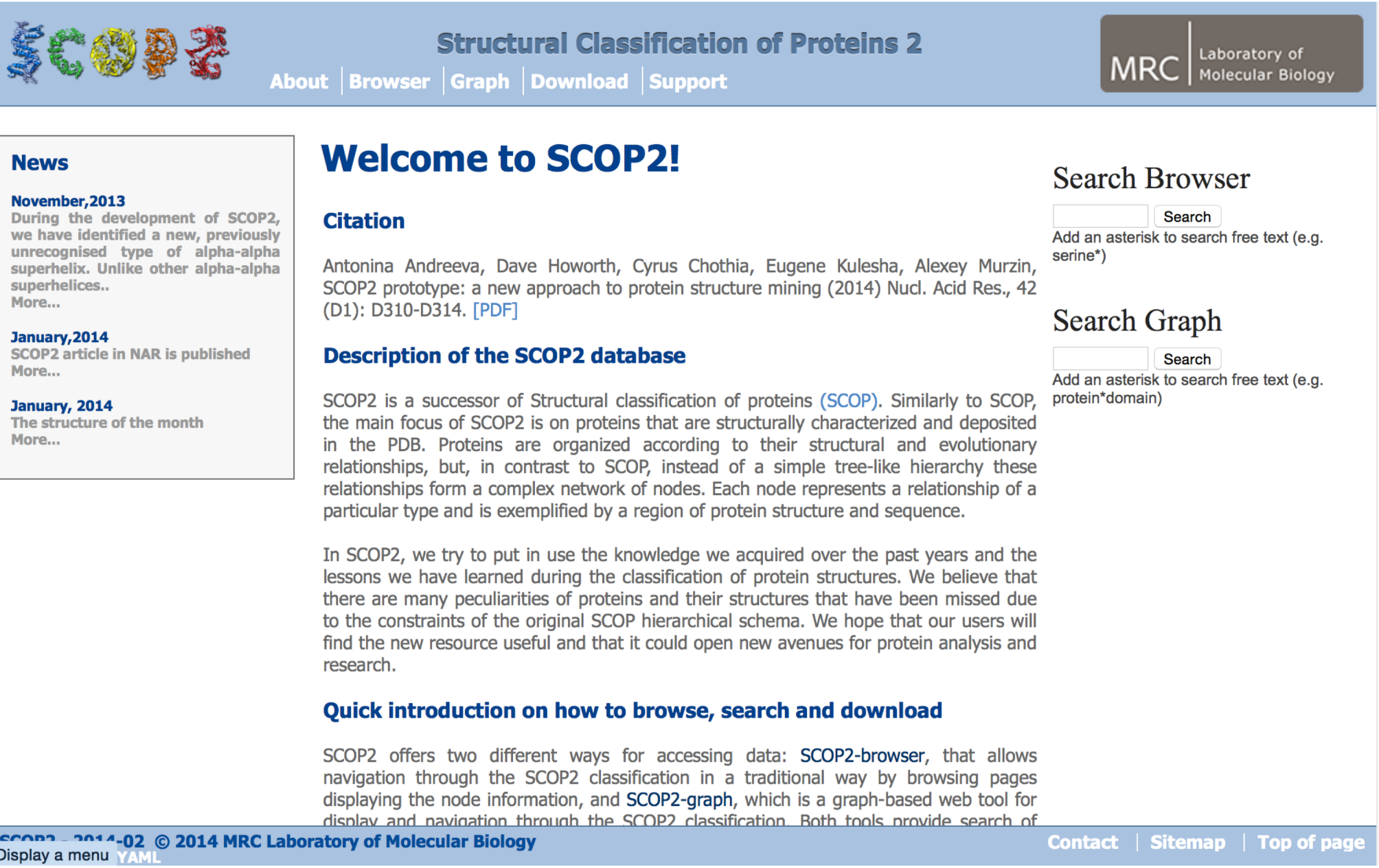
Superfamily: *Probable common evolutionary origin*

Proteins that have low sequence identities, but whose structural and functional features suggest that a common evolutionary origin is probable are placed together in superfamilies. *For example, actin, the ATPase domain of the heat shock protein, and hexakinase together form a superfamily*.

Fold: *Major structural similarity*

Proteins are defined as having a common fold if they have the same major secondary structures in the same arrangement and with the same topological connections. Different proteins with the same fold often have peripheral elements of secondary structure and turn regions that differ in size and conformation. *Proteins placed together in the same fold category may not have a common evolutionary origin: the structural similarities could arise just from the physics and chemistry of proteins favoring certain packing arrangements and chain topologies.*


SCOP: Root: scop - Mozilla Firefox
Soubor Úpravy Zobrazit Historie Záložky Nástroje Nápověda
http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.html
Nejnavštěvovanější Jak začít Přehled zpráv http://www.ncbi.nlm.... http://www.glycoscie... CHMI Radar Departme...
Root: scop
Classes:
1. All alpha proteins [46456] (258)
2. All beta proteins [48724] (165)
3. Alpha and beta proteins (a/b) [51349] (141)
Mainly parallel beta sheets (beta-alpha-beta units)
4. Alpha and beta proteins (a+b) [53931] (334)
Mainly antiparallel beta sheets (segregated alpha and beta regions)
5. Multi-domain proteins (alpha and beta) [56572] (53)
Folds consisting of two or more domains belonging to different classes
6. Membrane and cell surface proteins and peptides [56835] (50)
Does not include proteins in the immune system
7. Small proteins [56992] (85)
Usually dominated by metal ligand, heme, and/or disulfide bridges
8. Coiled coil proteins [57942] (7)
Not a true class
9. Low resolution protein structures [58117] (26)
Not a true class
10. Peptides [58231] (120)
Peptides and fragments. Not a true class
11. Designed proteins [58788] (44)
Experimental structures of proteins with essentially non-natural sequences. Not a true class
Enter search key:
Search

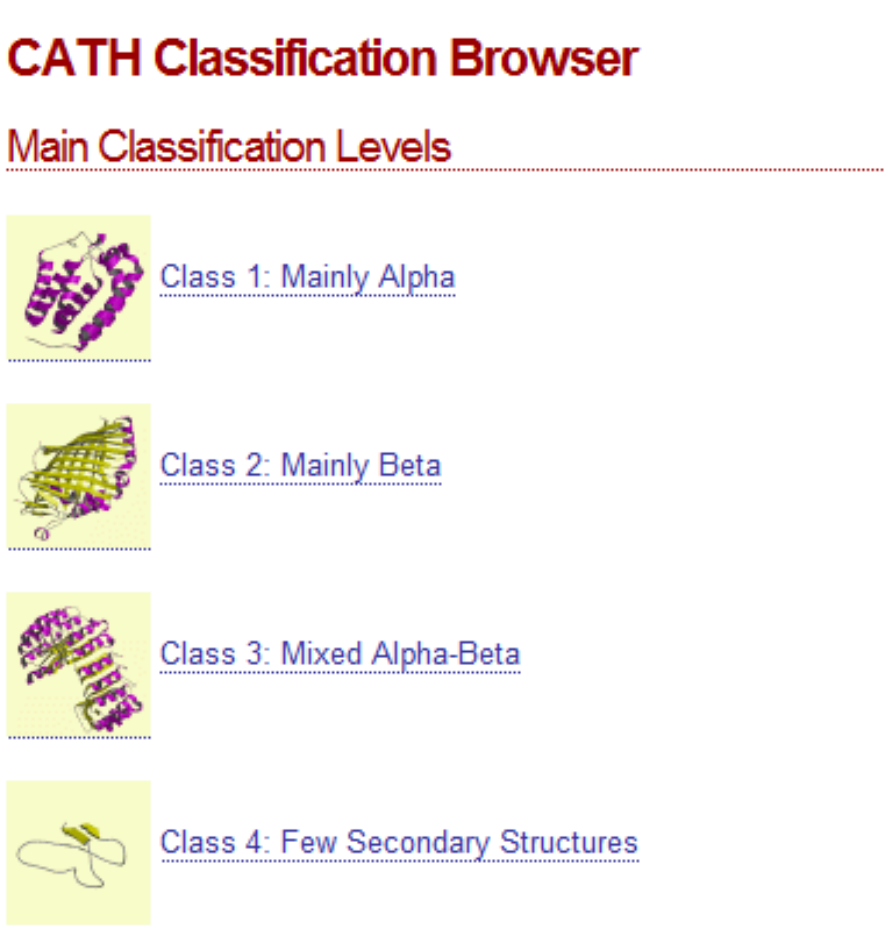


http://scop2.mrc-lmb.cam.ac.uk/front.html

CATH Protein Structure Classification (http:// www.cathdb.info)

CATH is a hierarchical classification of protein **domain** structures, which clusters proteins at four major levels: Class (C), Architecture (A), Topology (T) and Homologous superfamily (H). The boundaries and assignments for each protein domain are determined using a combination of automated and manual procedures which include computational techniques, empirical and statistical evidence, literature review and expert analysis

Class (C), Architecture (A) - the overall shape of the domain structure as determined by the orientations of the secondary structures but ignores the connectivity between the secondary structures., Topology (T) - the same overall shape and connectivity of the secondary structures in the domain core Homologous superfamily (H) - share a common ancestor (Similarities are identified either by high sequence identity or structure comparison)



CATH Protein Structure Classification (http:// www.cathdb.info)

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

16 million protein domains classified into 2,626 superfamilies

Browse » Search » Download » Take the Tou

What's New?

The CATH website has recently undergone a big overhaul. We really hope you find the new pages more useful, easier to use and quicker to load. Please get in touch and let us know what you think.

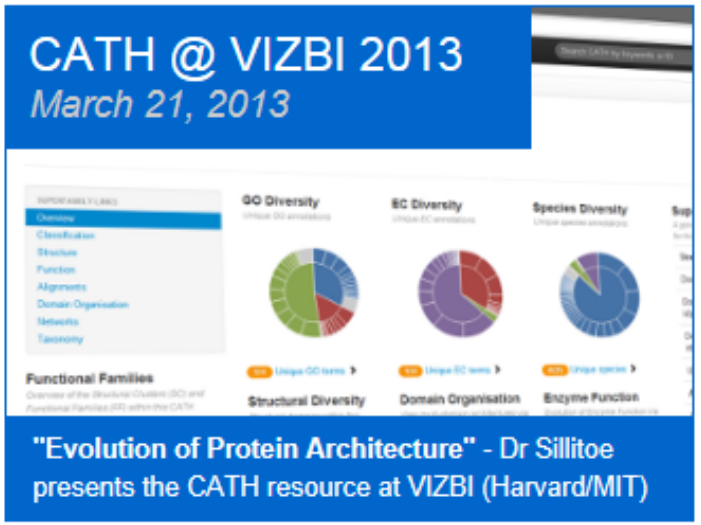
Searching CATH

- Search by ID / keyword
- Search by FASTA sequence
- Search by PDB structure

Example pages

- PDB "2bop"
- Domain "1cukA01"
- Relatives of "1cukA01"
- Superfamily "HUPs"
- Functional Family
- FunFam Alignment
- Search for "enolase"
- Superfamily Comparison

Latest News



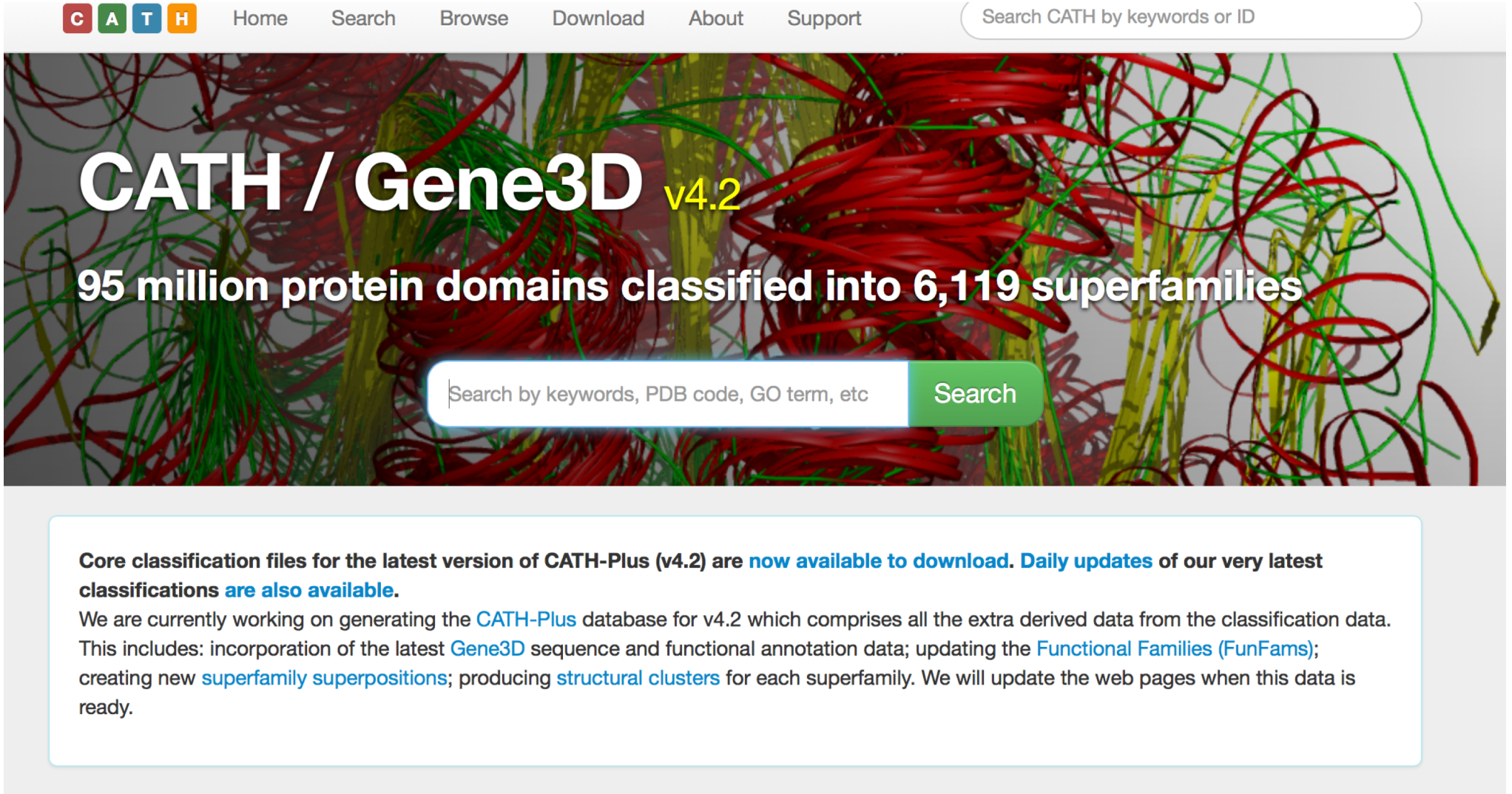
"Evolution of Protein Architecture" - Dr Sillitoe presents the CATH resource at VIZBI (Harvard/MIT)

Latest Release

CATH v3.5	based on PDB dated Septemb
173,536	CATH Domains
2,626	CATH Superfamilies
51,334	PDBs

Gene3D v11	released March 18, 2012
1,639	Cellular Genomes
1,016	Viral Genomes
14,963,305	Protein Sequences
16,297,076	CATH Domain Predicti

CATH Protein Structure Classification (http:// www.cathdb.info)



Fold Databases
SCOP Structural Classification of Proteins (http://scop.mrc-lmb.cam.ac.uk/scop/)
Dali/FSSP (http://www.ebi.ac.uk/dali/)
CATH Protein Structure Classification (http:// www.cathdb.info)

Structural Alignment Tools
Vast (http://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html)
CE (http://cl.sdsc.edu/ce.html)
DALI (http://www.ebi.ac.uk/dali)

Fold Prediction
3D-PSSM and PHYRE Protein Fold Recognition (http://www.sbg.bio.ic.ac.uk/~phyre/)
CPHmodels homology modeling (http://www.cbs.dtu.dk/services/CPHmodels/)
Geno3D (http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html)
3D-JIGSAW (http://www.bmm.icnet.uk/~3djigsaw/)
ESyPred3D (http://www.fundp.ac.be/urbm/bioinfo/esypred/)

Fully Automatic Homology Modelling
Robetta full-chain protein structure prediction server (http://robetta.bakerlab.org/)
Swiss-Model (http://www.expasy.org/swissmod/SWISS-MODEL.html)