Blok 4

Terciární struktura a oligomerizace proteinů

C3211 Aplikovaná bioinformatika

Přednášející: Josef Houser

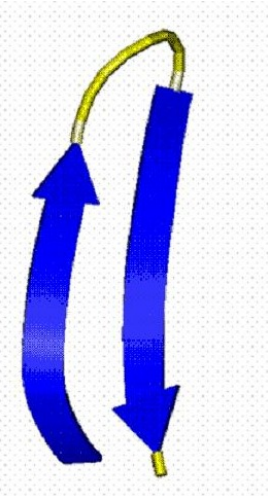
Terciární struktura proteinů

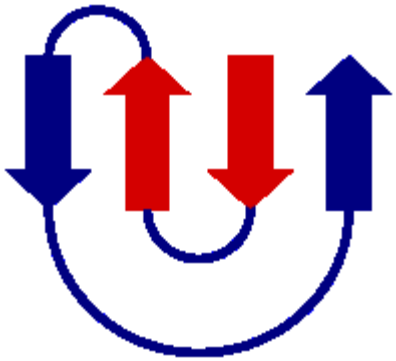
- Komplikovaná hiearchie
- Sekundární – (Supersekundární) – Terciární
- Motivy – Foldy – Domény

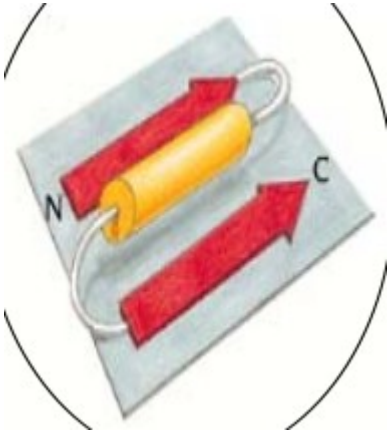
Strukturní motivy

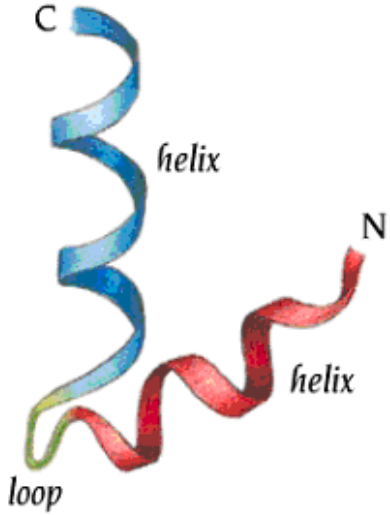
Způsob řazení úseků sekundární struktury, např:

- beta-vlásenka (beta-hairpin)
- řecký klíč (Greek key)
- beta-alfa-beta
- helix-otáčka-helix (helix-turn-helix)





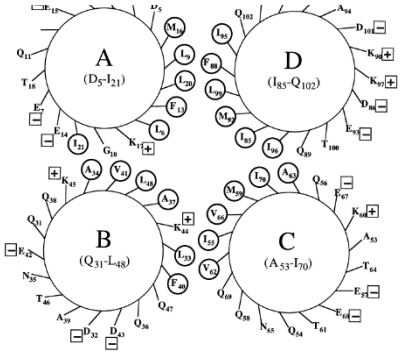


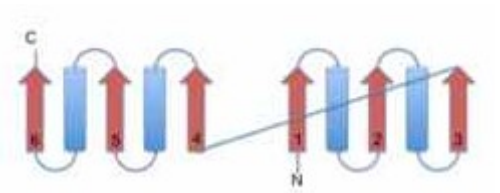


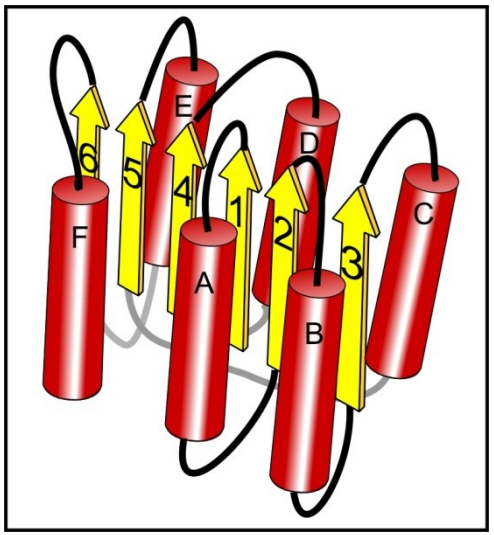
Proteinové foldy

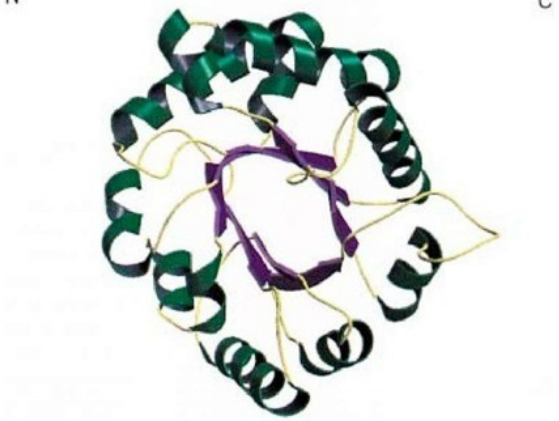
Vznikají kombinací několika motivů, např.:

- Helix-bundle
- Rossmanův fold
- TIM-barrel



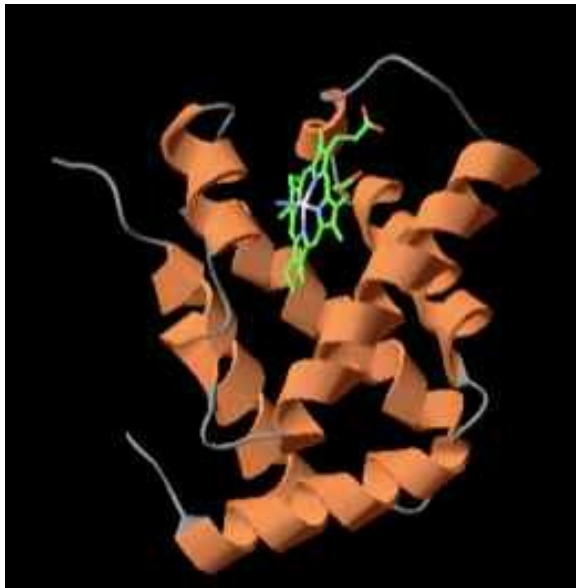






Proteinové domény

- Proteinová doména je prostorově vymezený úsek proteinu, obvykle s vlastní funkcí
- Známe proteiny jednodoménové i vícedoménové



3D struktura

- Konkrétní umístění jednotlivých atomů polypeptidového řetězce v prostoru

- **Absolutní souřadnice** – x, y, z

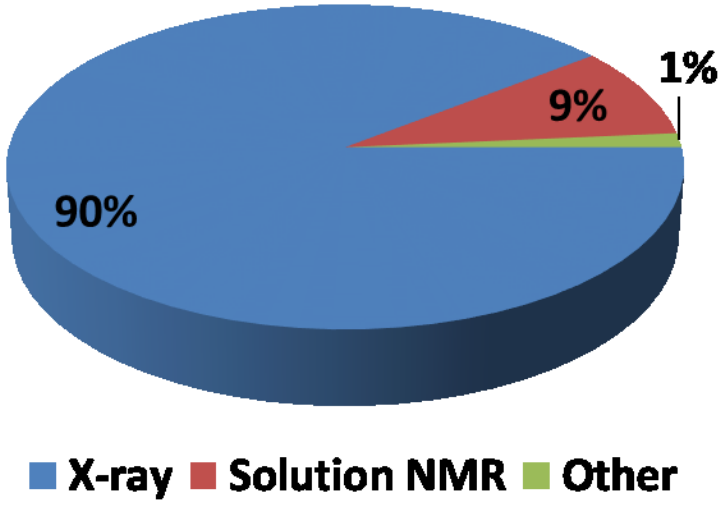
  pro N atomů je třeba 3N souřadnic

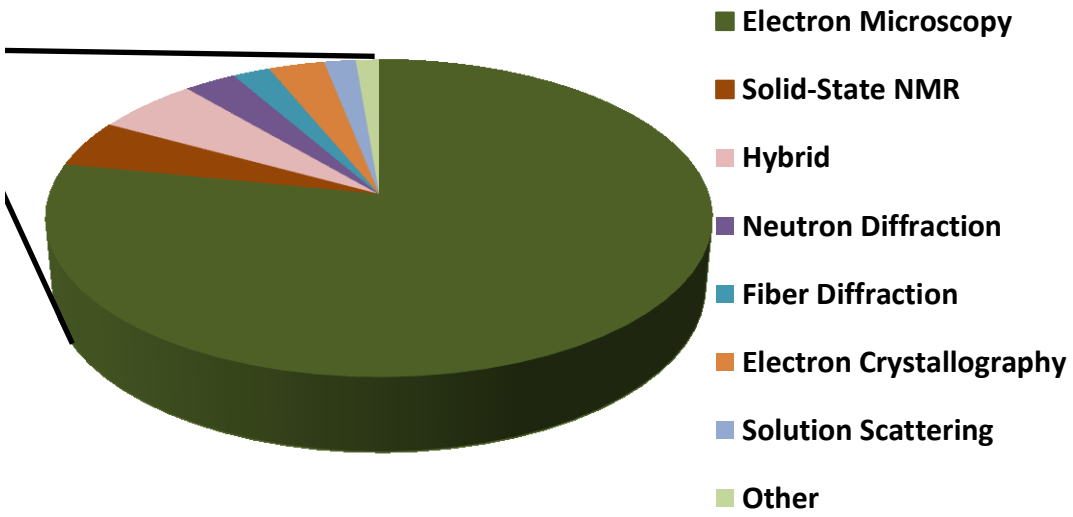
- **Relativní souřadnice** – vzdálenost, úhel, torzní úhel

  pro N atomů je třeba 3N – 3 souřadnic

Metody určení 3D struktury

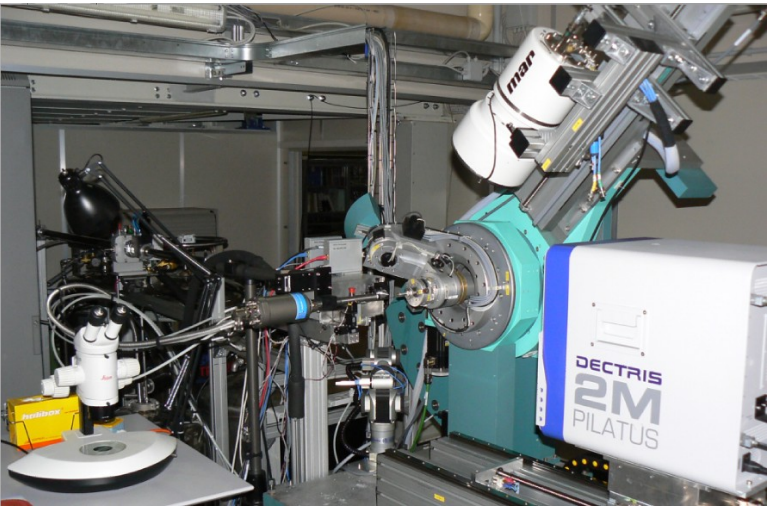
- Rentgenová difrakce - nejvíce struktur v PDB
- NMR
- Ostatní metody ~ 1%





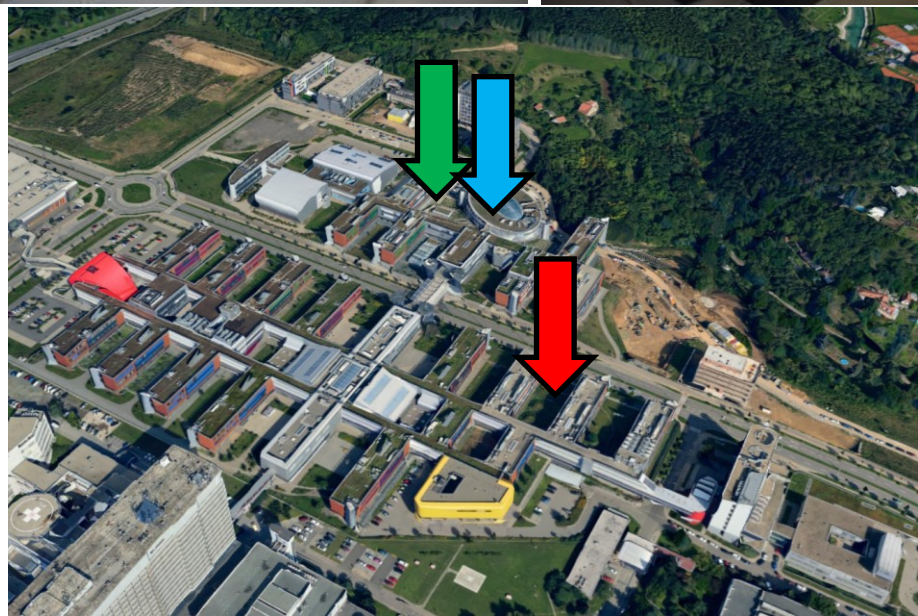
Čas na exkurzi

X-ray



NMR

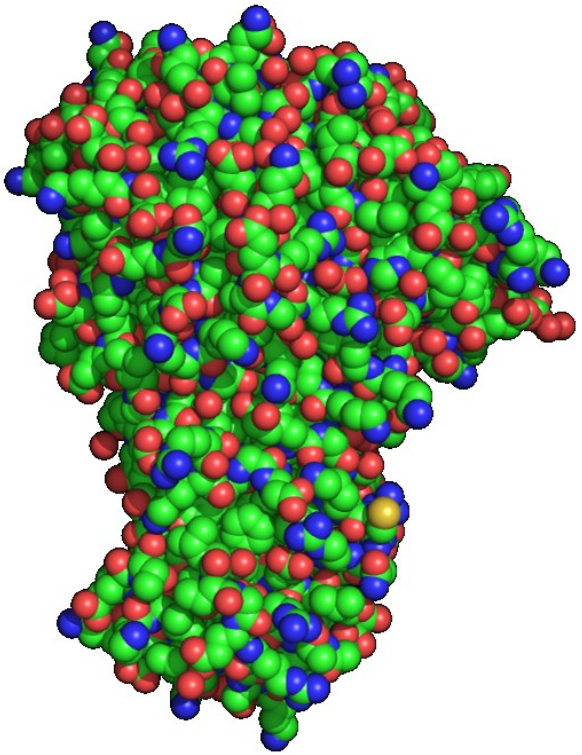
Cryo-EM



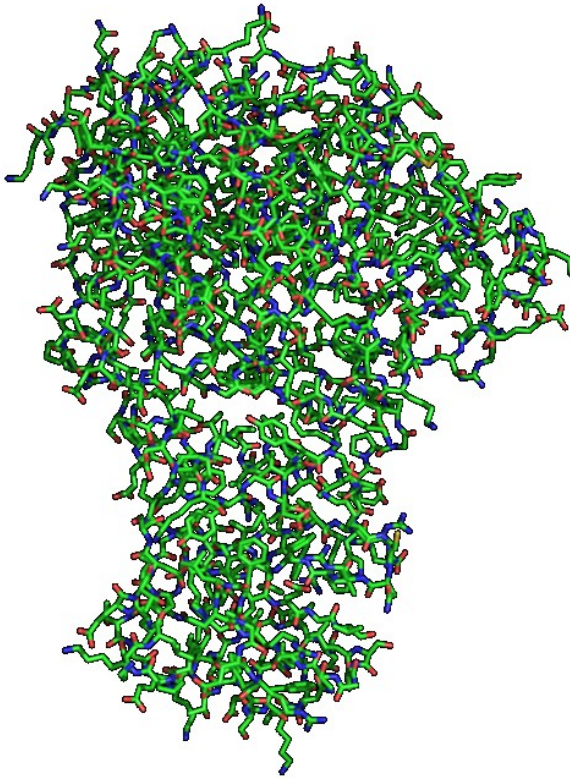
Zobrazení 3D struktury

- Zobrazovací SW: **PyMol**, Jmol, RasMol, VMD, Chimera, Cn3D,…

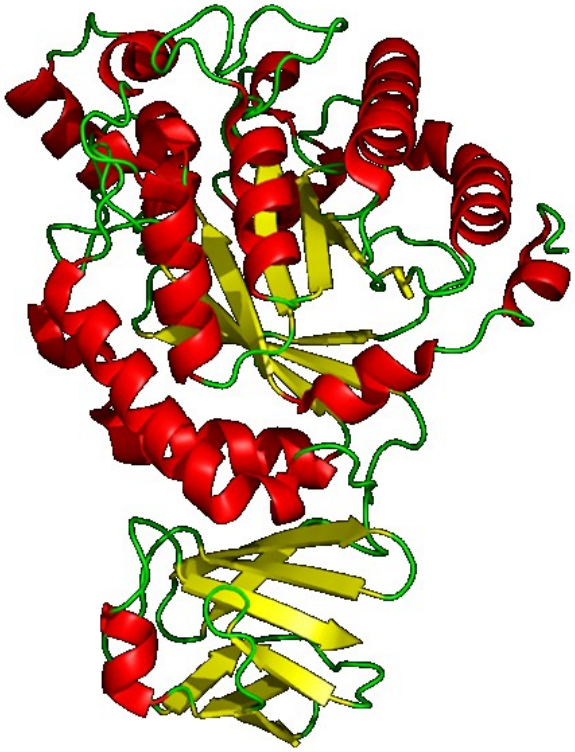
spheres, surface



sticks/balls and sticks



cartoon/ribbon



Databáze 3D struktur

- **wwPDB** (http://www.wwpdb.org)

  RCSB PDB – Research Collaboratory for Structural Bioinformatics Protein Data Bank (http://www.rcsb.org)

  PDBe – Protein Data Bank Europe

  PDBj – Protein Data Bank Japan

  BMRB – Biological Magnetic Resonance Data Bank

- **SCOP** (http://scop.mrc-lmb.cam.ac.uk/scop/) – strukturní klasifikace proteinů

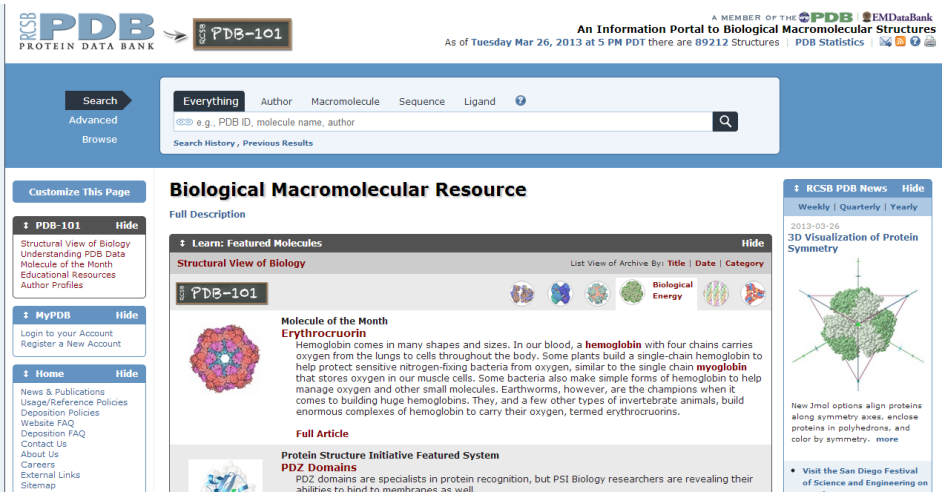
- **CATH** (http://www.cathdb.info/) – klasifikace proteinových domén z PDB

- **EMDataBank** (http://www.emdatabank.org/) – struktury z elektronové mikroskopie

Formáty uložení 3D struktury

PDB (Protein Data Bank)

- PDB File Format
- mmCIF File Format and PDB Exchange Dictionary
- PDBML/XML File Format



PDB

```
ATOM      7  CD  ARG A   7     -24.396 -12.945  52.578  1.00 59.72           C
ATOM      8  NE  ARG A   7     -25.048 -12.736  53.869  1.00 61.30           N
ATOM      9  CZ  ARG A   7     -24.413 -12.499  55.014  1.00 61.72           C
ATOM     10 NH1 ARG A   7     -23.087 -12.440  55.065  1.00 61.05           N
ATOM     11 NH2 ARG A   7     -25.115 -12.320  56.126  1.00 63.61           N
ATOM     12  N   TYR A   8     -24.055  -9.007  49.545  1.00 50.83           N
ATOM     13  CA  TYR A   8     -23.096  -8.100  48.940  1.00 48.87           C
ATOM     14  C   TYR A   8     -21.680  -8.609  49.201  1.00 47.84           C
ATOM     15  O   TYR A   8     -21.378  -9.123  50.279  1.00 47.98           O
ATOM     16  CB  TYR A   8     -23.287  -6.680  49.481  1.00 47.56           C
ATOM     17  CG  TYR A   8     -24.700  -6.147  49.294  1.00 48.37           C
ATOM     18  CD1 TYR A   8     -25.123  -5.630  48.067  1.00 49.00           C
ATOM     19  CD2 TYR A   8     -25.619  -6.180  50.332  1.00 48.91           C
ATOM     20  CE1 TYR A   8     -26.419  -5.156  47.889  1.00 48.83           C
ATOM     21  CE2 TYR A   8     -26.918  -5.707  50.160  1.00 50.24           C
ATOM     22  CZ  TYR A   8     -27.306  -5.192  48.936  1.00 49.98           C
ATOM     23  OH  TYR A   8     -28.589  -4.719  48.773  1.00 51.15           O
ATOM     24  N   LYS A   9     -20.837  -8.493  48.178  1.00 46.89           N
ATOM     25  CA  LYS A   9     -19.429  -8.849  48.240  1.00 45.81           C
ATOM     26  C   LYS A   9     -18.589  -7.576  48.327  1.00 43.61           C
ATOM     27  O   LYS A   9     -19.052  -6.506  47.921  1.00 42.40           O
ATOM     28  CB  LYS A   9     -19.034  -9.623  46.986  1.00 47.08           C
ATOM     29  CG  LYS A   9     -19.825 -10.894  46.763  1.00 49.32           C
ATOM     30  CD  LYS A   9     -19.594 -11.448  45.365  1.00 51.43           C
ATOM     31  CE  LYS A   9     -20.847 -11.313  44.498  1.00 53.57           C
ATOM     32  NZ  LYS A   9     -21.783 -12.464  44.647  1.00 54.87           N
ATOM     33  N   PRO A  10     -17.364  -7.701  48.849  1.00 42.21           N
ATOM     34  CA  PRO A  10     -16.466  -6.570  49.049  1.00 40.92           C
ATOM     35  C   PRO A  10     -15.637  -6.210  47.808  1.00 40.77           C
ATOM     36  O   PRO A  10     -14.406  -6.346  47.819  1.00 40.13           O
ATOM     37  CB  PRO A  10     -15.542  -7.084  50.158  1.00 40.95           C
ATOM     38  CG  PRO A  10     -15.397  -8.529  49.837  1.00 42.10           C
ATOM     39  CD  PRO A  10     -16.736  -8.954  49.308  1.00 42.99           C
ATOM     40  N   ASP A  11     -16.313  -5.778  46.748  1.00 40.65           N
ATOM     41  CA  ASP A  11     -15.646  -5.150  45.620  1.00 41.05           C
ATOM     42  C   ASP A  11     -16.594  -4.117  45.054  1.00 40.11           C
ATOM     43  O   ASP A  11     -17.804  -4.241  45.194  1.00 40.53           O
ATOM     44  CB  ASP A  11     -15.159  -6.174  44.567  1.00 42.62           C
ATOM     45  CG  ASP A  11     -16.278  -6.799  43.791  1.00 45.99           C
ATOM     46 OD1 ASP A  11     -16.825  -7.814  44.274  1.00 50.77           O
ATOM     47 OD2 ASP A  11     -16.673  -6.360  42.684  1.00 49.06           O
ATOM     48  N   TRP A  12     -16.044  -3.075  44.451  1.00 38.94           N
```

mmCIF

```
ATOM   1    N  N    . GLU A 1 1   ? 7.254   11.020 4.888  1.00 61.38 ? ? ? ? ? ? 546 GLU A N   1
ATOM   2    C  CA   . GLU A 1 1   ? 6.404   12.200 5.071  1.00 67.04 ? ? ? ? ? ? 546 GLU A CA  1
ATOM   3    C  C    . GLU A 1 1   ? 7.111   13.526 4.729  1.00 59.60 ? ? ? ? ? ? 546 GLU A C   1
ATOM   4    O  O    . GLU A 1 1   ? 6.576   14.360 3.999  1.00 64.05 ? ? ? ? ? ? 546 GLU A O   1
ATOM   5    C  CB   . GLU A 1 1   ? 5.842   12.232 6.500  1.00 74.02 ? ? ? ? ? ? 546 GLU A CB  1
ATOM   6    C  CG   . GLU A 1 1   ? 5.625   13.627 7.094  1.00 74.52 ? ? ? ? ? ? 546 GLU A CG  1
ATOM   7    C  CD   . GLU A 1 1   ? 4.448   14.369 6.495  1.00 78.40 ? ? ? ? ? ? 546 GLU A CD  1
ATOM   8    O  OE1  . GLU A 1 1   ? 3.968   13.977 5.409  1.00 81.00 ? ? ? ? ? ? 546 GLU A OE1 1
ATOM   9    O  OE2  . GLU A 1 1   ? 3.997   15.354 7.118  1.00 79.97 ? ? ? ? ? ? 546 GLU A OE2 1
ATOM   10   N  N    . ASP A 1 2   ? 8.299   13.714 5.287  1.00 44.26 ? ? ? ? ? ? 547 ASP A N   1
ATOM   11   C  CA   . ASP A 1 2   ? 9.213   14.768 4.873  1.00 34.80 ? ? ? ? ? ? 547 ASP A CA  1
ATOM   12   C  C    . ASP A 1 2   ? 10.508  14.039 4.527  1.00 30.06 ? ? ? ? ? ? 547 ASP A C   1
ATOM   13   O  O    . ASP A 1 2   ? 11.245  13.650 5.424  1.00 29.92 ? ? ? ? ? ? 547 ASP A O   1
ATOM   14   C  CB   . ASP A 1 2   ? 9.460   15.735 6.039  1.00 34.15 ? ? ? ? ? ? 547 ASP A CB  1
ATOM   15   C  CG   . ASP A 1 2   ? 10.399  16.909 5.672  1.00 36.09 ? ? ? ? ? ? 547 ASP A CG  1
ATOM   16   O  OD1  . ASP A 1 2   ? 11.138  16.835 4.665  1.00 33.05 ? ? ? ? ? ? 547 ASP A OD1 1
ATOM   17   O  OD2  . ASP A 1 2   ? 10.397  17.917 6.418  1.00 36.96 ? ? ? ? ? ? 547 ASP A OD2 1
ATOM   18   N  N    . LEU A 1 3   ? 10.778  13.854 3.239  1.00 32.19 ? ? ? ? ? ? 548 LEU A N   1
ATOM   19   C  CA   . LEU A 1 3   ? 11.922  13.061 2.787  1.00 30.81 ? ? ? ? ? ? 548 LEU A CA  1
ATOM   20   C  C    . LEU A 1 3   ? 13.253  13.688 3.155  1.00 27.21 ? ? ? ? ? ? 548 LEU A C   1
ATOM   21   O  O    . LEU A 1 3   ? 14.229  12.996 3.390  1.00 28.91 ? ? ? ? ? ? 548 LEU A O   1
ATOM   22   C  CB   . LEU A 1 3   ? 11.876  12.852 1.272  1.00 34.20 ? ? ? ? ? ? 548 LEU A CB  1
ATOM   23   C  CG   . LEU A 1 3   ? 10.861  11.859 0.700  1.00 39.97 ? ? ? ? ? ? 548 LEU A CG  1
ATOM   24   C  CD1  . LEU A 1 3   ? 10.990  11.774 -0.804 1.00 38.37 ? ? ? ? ? ? 548 LEU A CD1 1
ATOM   25   C  CD2  . LEU A 1 3   ? 11.016  10.475 1.322  1.00 42.33 ? ? ? ? ? ? 548 LEU A CD2 1
ATOM   26   N  N    . PHE A 1 4   ? 13.292  15.008 3.199  1.00 29.71 ? ? ? ? ? ? 549 PHE A N   1
ATOM   27   C  CA   . PHE A 1 4   ? 14.506  15.722 3.570  1.00 25.93 ? ? ? ? ? ? 549 PHE A CA  1
ATOM   28   C  C    . PHE A 1 4   ? 14.824  15.475 5.044  1.00 26.15 ? ? ? ? ? ? 549 PHE A C   1
ATOM   29   O  O    . PHE A 1 4   ? 15.946  15.130 5.402  1.00 28.52 ? ? ? ? ? ? 549 PHE A O   1
ATOM   30   C  CB   . PHE A 1 4   ? 14.288  17.201 3.306  1.00 27.76 ? ? ? ? ? ? 549 PHE A CB  1
ATOM   31   C  CG   . PHE A 1 4   ? 15.455  18.070 3.638  1.00 24.82 ? ? ? ? ? ? 549 PHE A CG  1
ATOM   32   C  CD1  . PHE A 1 4   ? 15.616  18.586 4.924  1.00 18.06 ? ? ? ? ? ? 549 PHE A CD1 1
ATOM   33   C  CD2  . PHE A 1 4   ? 16.357  18.428 2.658  1.00 25.10 ? ? ? ? ? ? 549 PHE A CD2 1
ATOM   34   C  CE1  . PHE A 1 4   ? 16.678  19.423 5.227  1.00 22.26 ? ? ? ? ? ? 549 PHE A CE1 1
ATOM   35   C  CE2  . PHE A 1 4   ? 17.430  19.281 2.954  1.00 28.10 ? ? ? ? ? ? 549 PHE A CE2 1
ATOM   36   C  CZ   . PHE A 1 4   ? 17.586  19.769 4.244  1.00 21.53 ? ? ? ? ? ? 549 PHE A CZ  1
ATOM   37   N  N    . LYS A 1 5   ? 13.825  15.663 5.896  1.00 24.70 ? ? ? ? ? ? 550 LYS A N   1
ATOM   38   C  CA   . LYS A 1 5   ? 13.979  15.407 7.317  1.00 26.90 ? ? ? ? ? ? 550 LYS A CA  1
ATOM   39   C  C    . LYS A 1 5   ? 14.403  13.966 7.541  1.00 28.64 ? ? ? ? ? ? 550 LYS A C   1
```

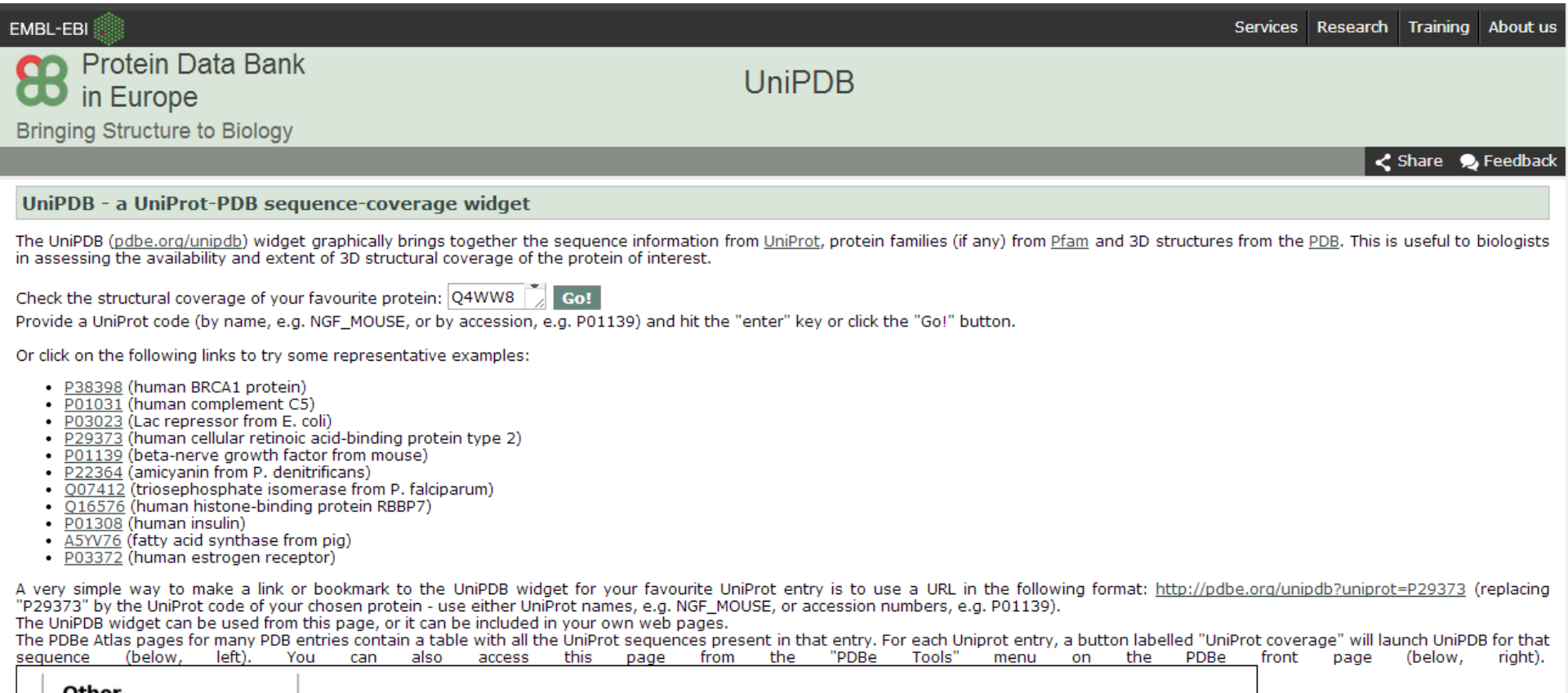
Úloha

Seznamte se s formátem .pdb Otevřete v textovém prohlížeči soubor **1RI6.pdb** a uveďte:

- O jaký protein se jedná?
- Kdo je autorem struktury?
- Jakou technikou byla struktura získána?
- Kolik atomů obsahuje kompletní struktura (včetně ligandů a solventu)?
- Jaké prvky sekundární struktury obsahuje daný protein?

UniPDB www.ebi.ac.uk/pdbe/widgets/unipdb

- Propojení databází UniProt (sekvence) a PDB (struktury)



UniPDB www.ebi.ac.uk/pdbe/widgets/unipdb

- **Vstup** – UniProt kód proteinu

- **Výstup** – přehled struktur se vztahem
  k tomuto proteinu (komplexy, domény,…),
  tj. všechny PDB záznamy, které obsahují
  alespoň část žádaného proteinu

Úloha

- **Pomocí aplikace UniPDB** (https://www.ebi.ac.uk/pdbe/widgets/unipdb/) zjistěte, zda je známa struktura alfa podjednotky A DNA polymerasy ze *Saccharomyces cerevisiae* (DPOA_YEAST) samotné nebo v komplexu
- Kolik dalších struktur obsahuje alespoň některou z domén tohoto enzymu?

EMBL-EBI Services Research Training Abou

Protein Data Bank in Europe
Bringing Structure to Biology

UniPDB

Share Feed

Download results | Related PDB sequences

DNA polymerase alpha catalytic subunit A
P13382 DPOA_YEAST *Saccharomyces cerevisiae...*

Type another UniProt code and hit enter

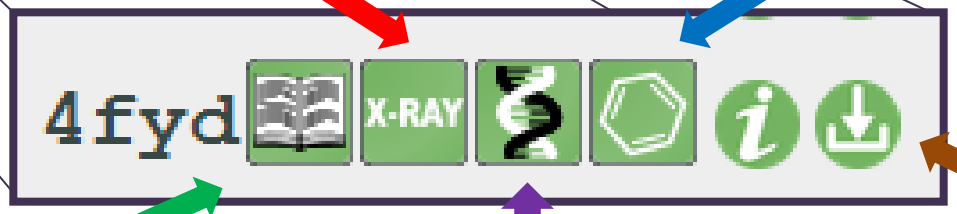
11 chains in 6 PDB entries map to this Uniprot accession.

P13382 — DNA_pol_B_exol | DNA_pol_B | zf-DNA_Pol

3oiq	X-RAY	B
4b08	X-RAY	A
4fyd	X-RAY	A,B
4fxd	X-RAY	A,B
4fvm	X-RAY	A
3flo	X-RAY	B,D,F,H

Strukturní techniky

Ligandy



Publikace

Nukleové kyseliny

PDB soubor

Search results

Reset DPOA_YEAST

View basket (0)

Save search Download

Refine query:

Macromolecules (7)

DNA polymerase alpha catalytic ...	(7)
Cell division control protein 13	(1)
DNA (5'-D(*TP*GP*AP*GP*CP*...	(1)
DNA (5'-D(*TP*TP*TP*TP*CP*G...	(1)
DNA polymerase alpha subunit B	(1)
DNA polymerase alpha-binding p..	(1)
RNA (5'-R(*AP*GP*GP*CP*GP*...	(1)

Molecule type (3)

Protein	(7)

Entries | Macromolecules | Compounds | Protein families

< 1 > Entry 1 to 7 of 7 Sort results

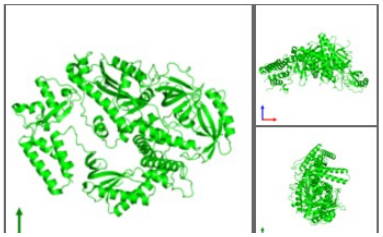
4b08 Yeast DNA polymerase alpha, Selenomethionine protein

Perera RL, Torella R, Klinge S, Kilkenny ML, Maman JD, Pellegrini L
Elife (2013) [PMID: 23599895]

Source organism: *Saccharomyces cerevisiae*
Assembly composition: protein only structure

Add to basket Download files

X-ray diffraction
2.67Å resolution
Released: 27 Feb 2013
Model geom...
Fit model/d...



In silico predikce 3D struktury

- *Ab initio*
- Homologní modelování
- Threading („navlékání“)

1) *Ab initio* - Quark

- Nevyžaduje existenci homologního proteinu
- Predikuje 2D strukturu, modeluje fragmenty a kombinuje je navzájem
- Nízká spolehlivost zejm. pro větší proteiny

QUARK http://zhanglab.ccmb.med.umich.edu/QUARK/

QUARK Ab Initio Results for Job Q12270

Submitted Primary Sequence

```
>Length 71
HVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNEVEKRPFVEEAERLRVQHKKDHPDYKY
12345678901234567890123456789012345678901234567890123456789012345678901
--------10--------20--------30--------40--------50--------60--------70
```

Predicted Secondary Structure

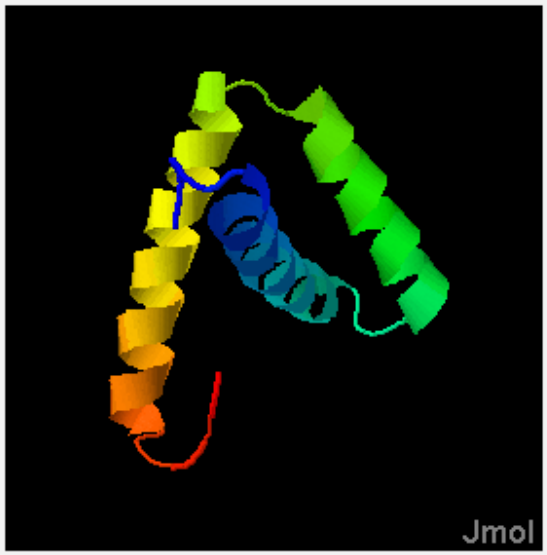
```
>C-coil;H-helix;E-sheet;T-beta turn
HVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNEVEKRPFVEEAERLRVQHKKDHPDYKY
CTTTTCHHHHHHHHHHHHHHHHTTTTCHHHHHHHHHHHHHHCCHHHHHHHHHHHHHHHHHHHHTTTTTT
12345678901234567890123456789012345678901234567890123456789012345678901
--------10--------20--------30--------40--------50--------60--------70
```

Download Predicted 3-state Secondary Structure Types
Download Predicted Starting Beta-turn Position
Download Predicted Real-value Phi-angle
Download Predicted Real-value Psi-angle
Download Distance Profile from Fragments
Download Clustered Torsion Angle Pairs from Fragments

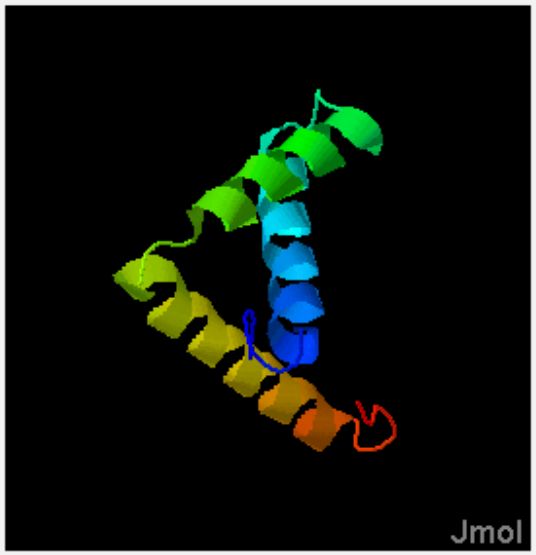
Predicted Solvent Accessibility

```
>0-buried to 9-exposed
HVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNEVEKRPFVEEAERLRVQHKKDHPDYKY
54420110100002101220143124121120020013004312321122013203301220353123143
12345678901234567890123456789012345678901234567890123456789012345678901
--------10--------20--------30--------40--------50--------60--------70
```

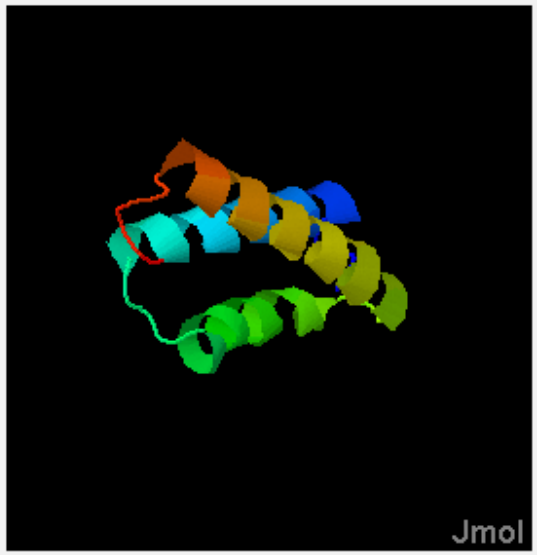
Predicted Tertiary Structure



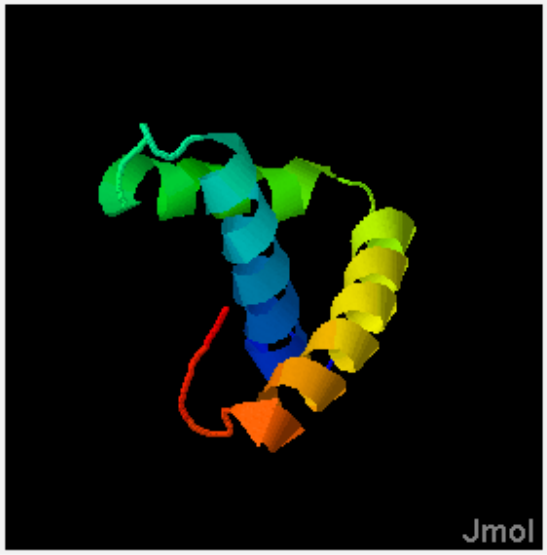
Download Model 1

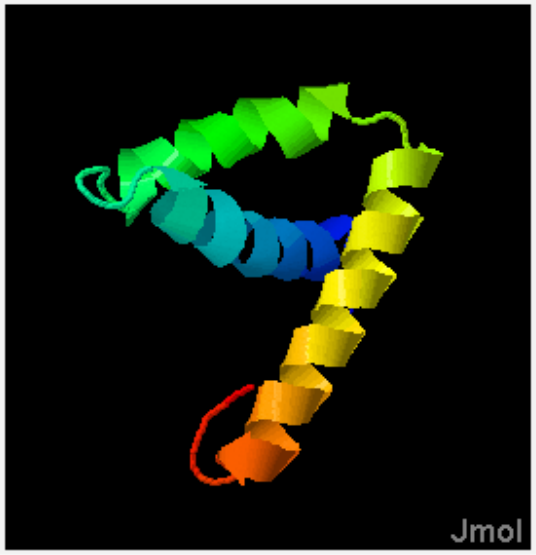


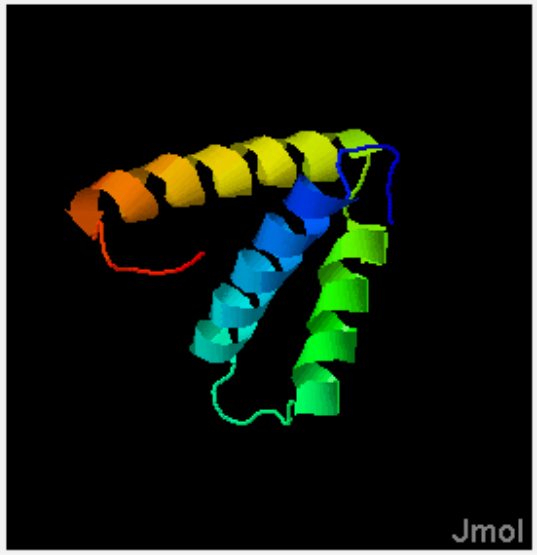
Download Model 2



Download Model 3



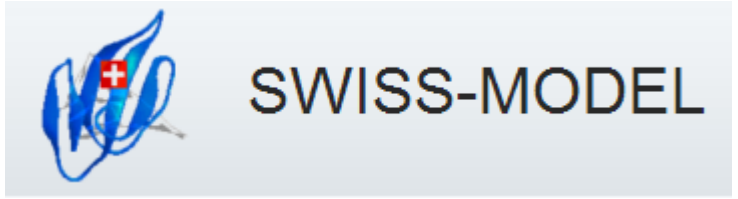




2) Homologní modelování

- Využívá skutečnosti, že dva proteiny ze stejné rodiny a s podobnou sekvencí mají i podobnou 3D strukturu
- Nutno znát strukturu homologního proteinu = templát

SWISS-MODEL http://swissmodel.expasy.org/



Úloha

- Pokuste se vytvořit model struktury pro zadaný protein pomocí serveru **SWISS-MODEL** (http://swissmodel.expasy.org/). Použijte automatický mód.

Sekvence:

MYPFFDNPNYTNTYATNEDFVCPYFLDYYNNSQDDYKNFRGENYDFEDTEE
NIENRNIEETEYEGLFRAWNPWNNLGGNITSGLGASSWAANRIDLFARGRG
GELIHNWFDNGKWNYWENLGGILTSSPKAVSWGFNRIDVVCRGTDNAMYHK
WWDGSSWSGFENLGGQLTSAPTICSWAPNRLDCFARGTDNQLHHKWWDGSS
WSQWEALGGSLTSGPGAVSWGPNRIDVFARGRNNTLIHKWWNGTSWSQWED
LGGFLTSAPCASSRGQNRIDVFARGRNNRLMYKYWDGSRWSDWTFLQGYLT
SEPVSVSRNSSSINVFAKGPRENVIERIYS

Swiss-Model

Welcome to SWISS-MODEL

SWISS-MODEL is a fully automated protein structure homology-modelling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists worldwide.

Start Modelling

SWISS-MODEL has recently had a face lift! You can still access the familiar version here.

Protein Structure Bioinformatics Group
c/o Prof. Torsten Schwede
Swiss Institute of Bioinformatics
Biozentrum, University of Basel
Klingelbergstrasse 50/70
CH-4056 Basel / Switzerland
help-swissmodel@unibas.ch

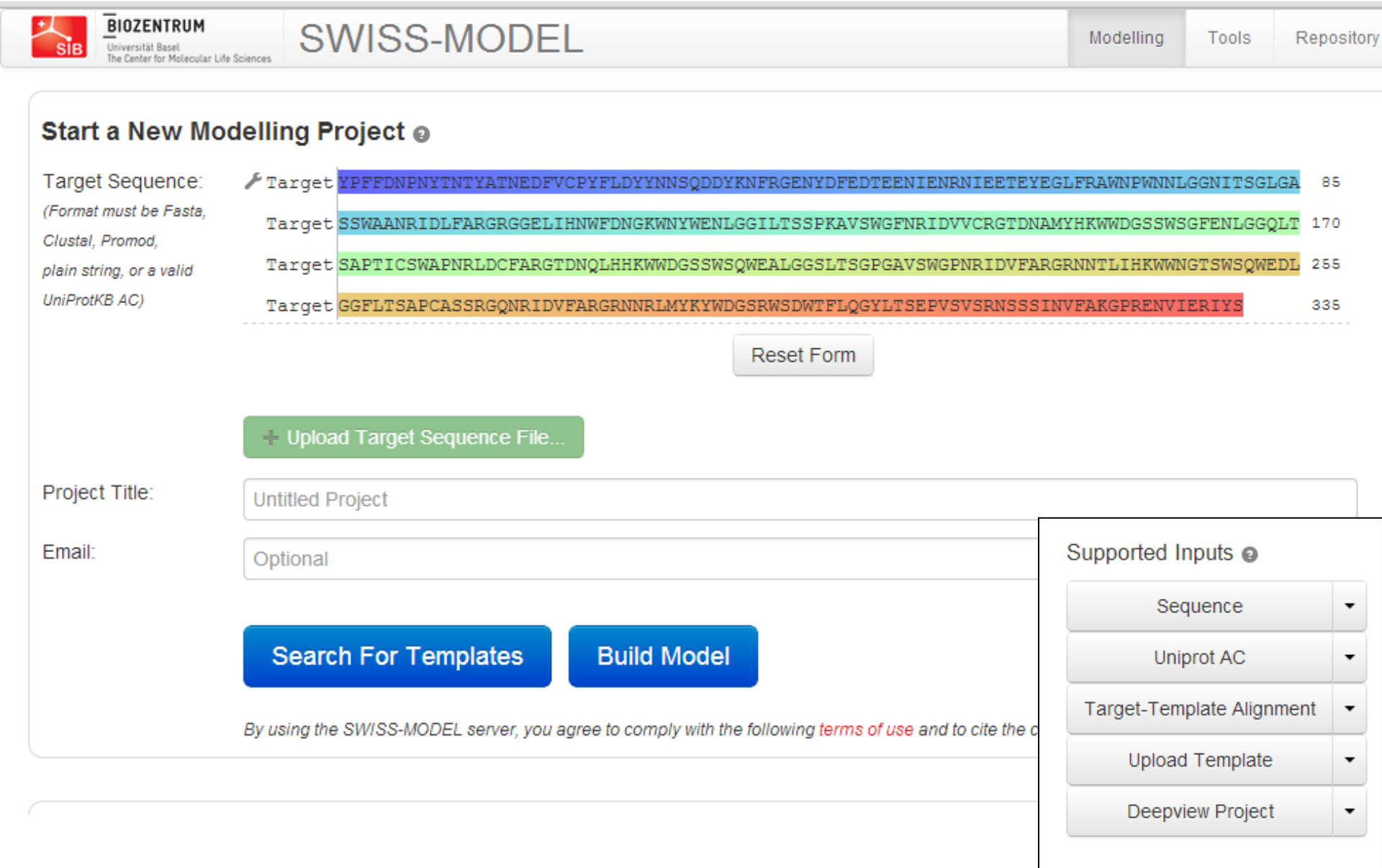
SIB Swiss Institute of Bioinformatics

BIOZENTRUM
Universität Basel
The Center for Molecular Life Sciences

When you publish or report results using SWISS-MODEL, please cite the relevant publications:

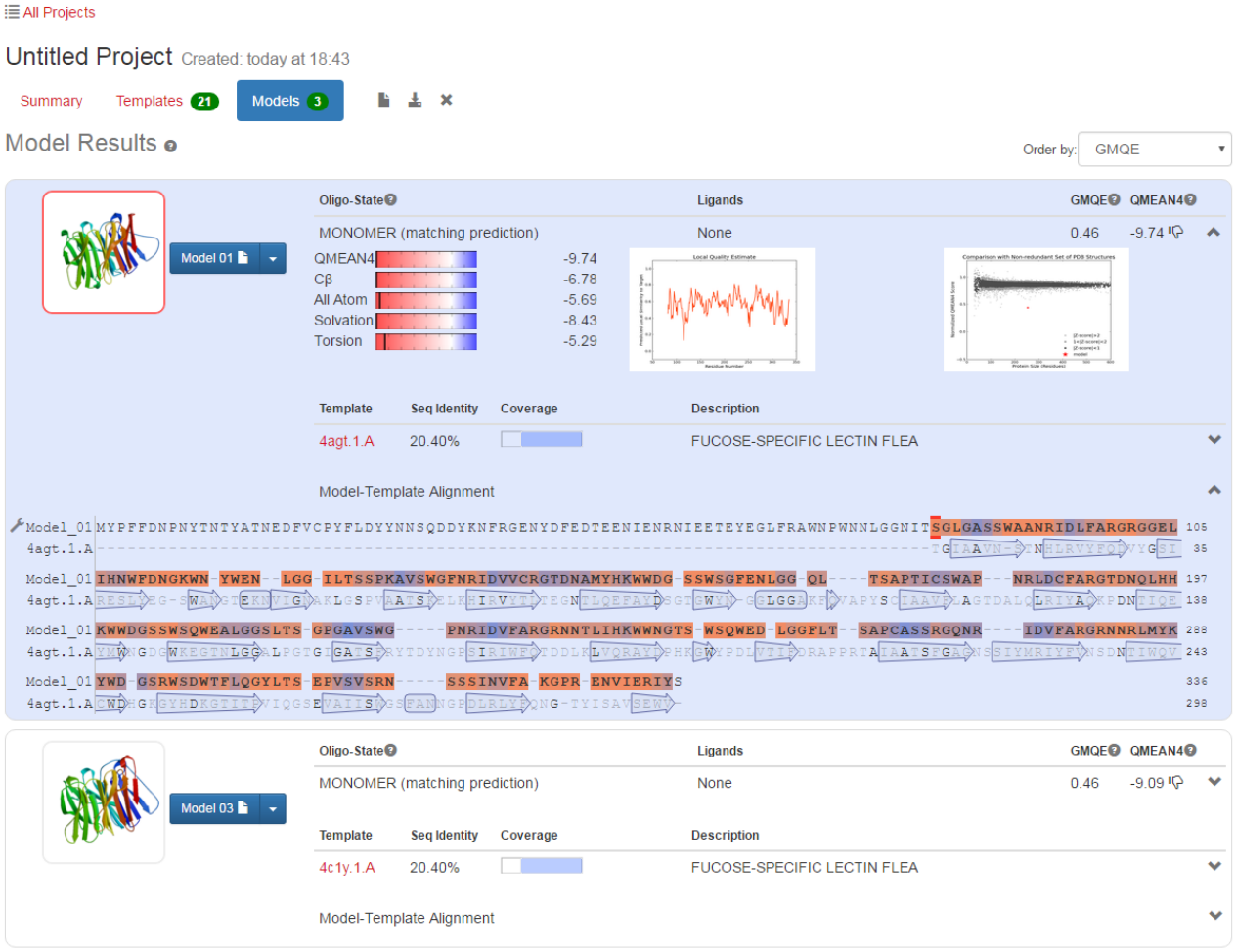
- Arnold K., Bordoli L., Kopp J., and Schwede T. (**2006**). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22,195-201.
- Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (**2009**). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*. 37, D387-D392.
- Guex, N.; Peitsch, M.C.; Schwede, T. **(2009)** SWISS-MODEL and Swiss PdbViewer. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, 30(S1), S162-S173.

Swiss-Model


BIOZENTRUM
Universität Basel
The Center for Molecular Life Sciences
SWISS-MODEL
Modelling
Tools
Repository
Start a New Modelling Project
Target Sequence:
(Format must be Fasta, Clustal, Promod, plain string, or a valid UniProtKB AC)
Target YPFFDNPNYTNTYATNEDFVCPYFLDYYNNSQDDYKNFRGENYDFEDTEENIENRNIEETEYEGLFRAWNPWNNLGGNITSGLGA 85
Target SSWAANRIDLFARGRGGELIHNWFDNGKWNYWENLGGILTSSPKAVSWGFNRIDVVCRGTDNAMYHKWWDGSSWSGFENLGGQLT 170
Target SAPTICSWAPNRLDCFARGTDNQLHHKWWDGSSWSQWEALGGSLTSGPGAVSWGPNRIDVFARGRNNTLIHKWWNGTSWSQWEDL 255
Target GGFLTSAPCASSRGQNRIDVFARGRNNRLMYKYWDGSRWSDWTFLQGYLTSEPVSVSRNSSSINVFAKGPRENVIERIYS 335
Reset Form
Upload Target Sequence File...
Project Title:
Untitled Project
Email:
Optional
Search For Templates
Build Model
By using the SWISS-MODEL server, you agree to comply with the following terms of use and to cite the c
Supported Inputs
Sequence
Uniprot AC
Target-Template Alignment
Upload Template
Deepview Project

Swiss-Model

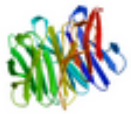
- Několik modelů

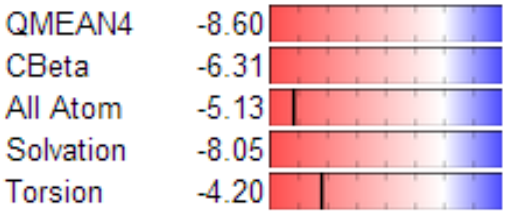


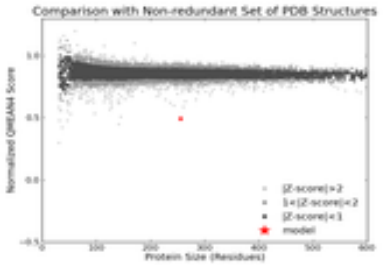
and HHBlits (Remmert, et al., 2011) for evolutionary related structures matching the target sequence in Table T1. For details on the template search, see Materials and Methods. Overall 13 templates were found (Table T2).

Models

The following model was built (see Materials and Methods "Model Building"):

Model #03	File	Built with	Oligo-State	Ligands	GMQE	QMEAN4
	PDB	ProMod Version 3.70.	MONOMER (matching prediction)	None	0.45	-8.60



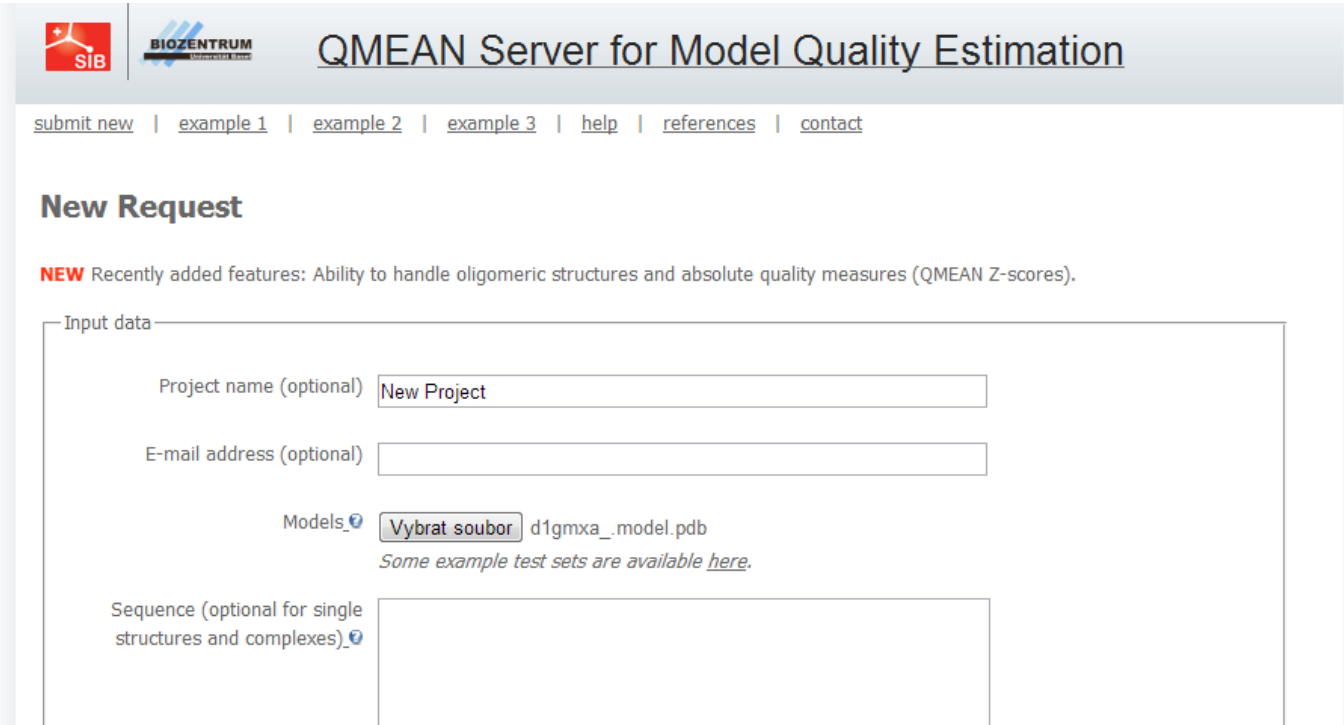


Template	Seq Identity	Oligo-state	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
1ofz.1.A	22.00	homo-dimer	HHblits	X-ray	1.50Å	0.32	80 - 334	0.75	FUCOSE-SPECIFIC LECTIN

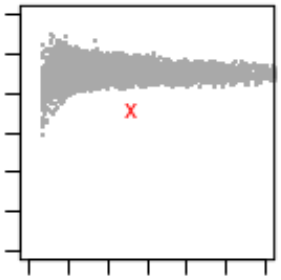
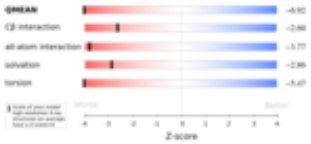
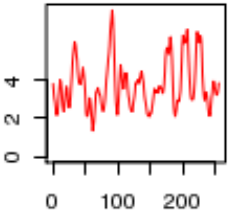
Ligand	Added to Model	Description
FUC	× - Binding site not conserved.	SUGAR (ALPHA-L-FUCOSE)
FUC	× - Binding site not conserved.	SUGAR (ALPHA-L-FUCOSE)
FUC	× - Binding site not conserved.	SUGAR (ALPHA-L-FUCOSE)
FUC	× - Binding site not conserved.	SUGAR (ALPHA-L-FUCOSE)
FUL	× - Binding site not conserved.	SUGAR (BETA-L-FUCOSE)
FUL	× - Binding site not conserved.	SUGAR (BETA-L-FUCOSE)
FUL	× - Binding site not conserved.	SUGAR (BETA-L-FUCOSE)
FUL	× - Binding site not conserved.	SUGAR (BETA-L-FUCOSE)

QMEAN http://swissmodel.expasy.org/qmean/cgi/index.cgi

- Stanovení **kvality** strukturního modelu
- Automaticky provedeno při použití serveru SWISS-MODEL
- Možno analyzovat i vlastní modely



Global Model Quality Estimation [+/-]

QMEAN4 global scores:			Local scores	
QMEANscore4	Estimated absolute model quality	Score components	Coloring by residue error	Residue error plot
0.28	 Z-Score: -6.92 Plot 1: [save png] Plot 2: [save png]	 [save png]		 [save png]
			Coloring (all chains): [save jpg] [save pdb]	Energy profile: [save raw scores]

QMEAN4 global scores:

The QMEAN4 score is a composite score consisting of a linear combination of 4 statistical potential terms (estimated model reliability between 0-1). The pseudo-energies of the contributing terms are given below together with their Z-scores with respect to scores obtained for high-resolution experimental structures of similar size solved by X-ray crystallography:

Scoring function term	Raw score	Z-score
C_beta interaction energy	-15.91	-2.60
All-atom pairwise energy	-398.94	-3.77
Solvation energy	-0.24	-2.86
Torsion angle energy	23.20	-5.47
QMEAN4 score	0.277	-6.92

If you publish results from QMEAN, please cite the following paper:
Benkert P, Biasini M, Schwede T. (2011). "Toward the estimation of the absolute quality of individual protein structure models." Bioinformatics, 27(3):343-50.

3) Threading

- Modelování při nízké homologii s proteiny se známou strukturou
- Porovnává možnost přiložení sekvence na proteiny známých foldů

Phyre2 http://www.sbg.bio.ic.ac.uk/phyre2/

Phyre2

ARDLVIPMIYCGHGY

PSI-Blast

HMM

HMM-HMM matching

Hidden Markov Model DB of **KNOWN STRUCTURES**

Very powerful – able to reliably detect extremely remote homology

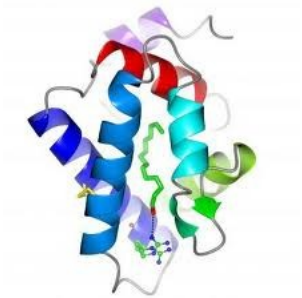
Routinely creates accurate models even when sequence identity is <15%

ARDL--VIPMIYCGHGY

AFDLCDLIPV--CGMAY

Sequence of known structure

3D-Model



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	houser@mail.muni.cz
Optional Job description	CBL
Amino Acid Sequence	MYPFFDNPNYTNTYATNEDFVCPYFLDYYNNSQDDYKNFRGENYDFEDTEENIENRNI EETEYEGLFRAWNPWNNLGGNITSGLGASSWAANRIDLFARGRGGELIHNWFDNGKWN YWENLGGILTSSPKAVSWGFNRIDVVCRGTDNAMYHKWWDGSSWSGFENLGGQLTSAP TICSWAPNRLDCFARGTDNQLHHKWWDGSSWSQWEALGGSLTSGPGAVSWGPNRIDVF ARGRNNTLIHKWWNGTSWSQWEDLGGFLTSAPCASSRGQNRIDVFARGRNNRLMYKYW DGSRWSDWTFLQGYLTSEPVSVSRNSSSINVFAKGPRENVIERIYS
Modelling Mode	Normal ◉ Intensive ○
	Phyre Search Reset

Phyre2

Job Status

Email	houser@mail.muni.cz
Job Description	CBL_____
Unique Job ID	6ae742ede312f99d
Date	Mon Feb 4 14:20:24 GMT 2013

Estimated total processing time: 2.2 hours ± 1.9 hours

Your job has entered the queue....*Waiting for job to start.*

171 jobs running **0** jobs queued

This page auto-refreshes every 30 seconds until job completion ....

Summary

Top model

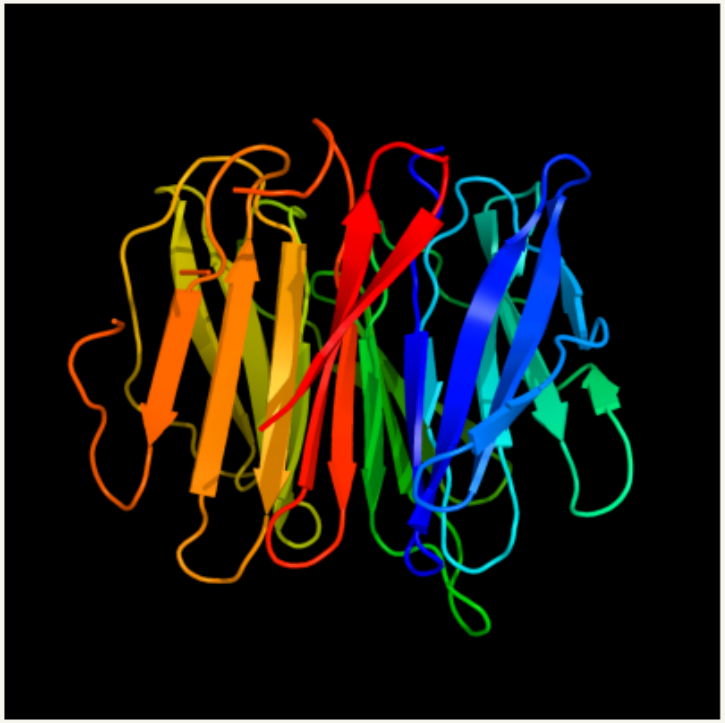


Image coloured by rainbow N → C terminus

Model (left) based on template d1ofza_

Top template information

Fold:6-bladed beta-propeller
Superfamily: Fucose-specific lectin
Family: Fucose-specific lectin

Confidence and coverage

Confidence:	100.0%	Coverage:	76%

255 residues (76% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

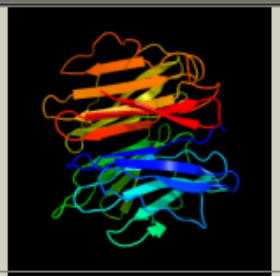
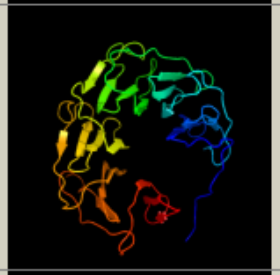
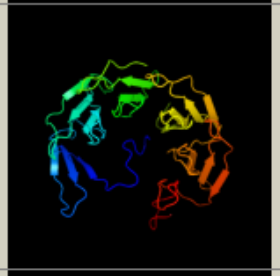
3D viewing

Interactive 3D view in Jmol

Sequence analysis

View PSI-Blast Pseudo-Multiple Sequence Alignment

Secondary structure and disorder prediction

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Info
1	d1ofza_	Alignment		100.0	20	**Fold:**6-bladed beta-propeller **Superfamily:**Fucose-specific lectin **Family:**Fucose-specific lectin
2	c2xbgA_	Alignment		97.9	13	**PDB header:**photosynthesis **Chain:** A: **PDB Molecule:**ycf48-like prote **PDBTitle:** crystal structure of ycf48 from
3	c2c4dA_	Alignment		95.6	16	**PDB header:**lectin **Chain:** A: **PDB Molecule:**psathyrella velu **PDBTitle:** 2.6a crystal structure of psathy with n-acetylglucosamine
4	c2xcyA_	Alignment		95.3	14	**PDB header:**hydrolase **Chain:** A: **PDB Molecule:**extracellular sia **PDBTitle:** crystal structure of aspergillus f

TASSER

- Několikrát označen jako nejlepší predikční server
- Nutná registrace

http://cssb.biology.gatech.edu/skolnick/webservice/TASSER/index.html

- Možnost práce online nebo stažení a instalace na lokální počítač (Linux)

TASSER online

TASSER online

Sequence is submitted. The result will be emailed to the given email ID within 100 hours.

Services

Super PSiFR

Protein Structure Prediction

chunk-TASSER
MetaTASSER
pro-sp3-TASSER
TASSER-VMT
TASSER

Protein Structure Refinement

BSR

Protein Structure Alignment

Fr-TM-align

Submitted Sequence

```
>protein
YPFFDNPNYTNTYATNEDFVCPYFLDYYNNSQDDYKNFRGENYDFEDTEENIENRNIEET
EYEGLFRAWNPWNNLGGNITSGLGASSWAANRIDLFARGRGGELIHNWFDNGKWNYWENL
GGILTSSPKAVSWGFNRIDVVCRGTDNAMYHKWWDGSSWSGFENLGGQLTSAPTICSWAP
NRLDCFARGTDNQLHHKWWDGSSWSQWEALGGSLTSGPGAVSWGPNRIDVFARGRNNTLI
HKWWNGTSWSQWEDLGGFLTSAPCASSRGQNRIDVFARGRNNRLMYKYWDGSRWSDWTFL
QGYLTSEPVSVSRNSSSINVFAKGPRENVIERIYS
```

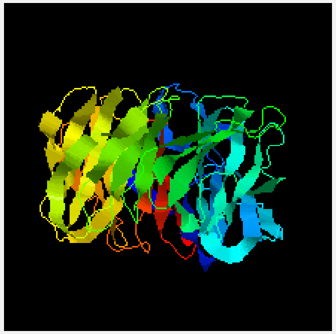
Predicted Secondary Structure

	20 40 60 80 100 120 140
Sequence	YPFFDNPNYTNTYATNEDFVCPYFLDYYNNSQDDYKNFRGENYDFEDTEENIENRNIEETEYEGLFRAWNPWNNLGGNITSGLGASSWAANRIDLFARGRGGELIHNWFDNGKWNYWENLGGILTSSPKAVSWGFNRIDVVCRGT
Prediction	CCCCCCCCCCCCCCCCCCSSCCCHHHHHHCCCCCCCCCCCCSSCCSSSSSSSSSCCCSSSSSCCCCCCCCCCSSCCCCCCCCCSSSSCCCCCSSSSSSCCCCCSSSSSSCCCCCCCCSCCCCCCCCCCSSSSSCCCCSSSSSSCC
Conf.Score	9956778667888866536775235655126000458983036011147999825725887516989888771189977898747980899689999959995899982599876664068976588678993899489999989

Predicted Solvent Accessibility

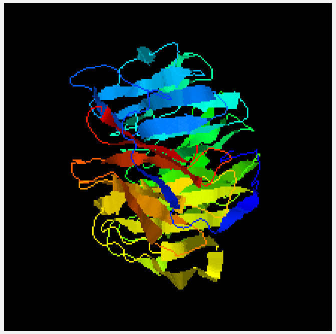
	20 40 60 80 100 120 140
Sequence	YPFFDNPNYTNTYATNEDFVCPYFLDYYNNSQDDYKNFRGENYDFEDTEENIENRNIEETEYEGLFRAWNPWNNLGGNITSGLGASSWAANRIDLFARGRGGELIHNWFDNGKWNYWENLGGILTSSPKAVSWGFNRIDVVCRGT
Prediction	40103001223323343301011105412234741731433313143323103233122321323323123123031322120000023721000000034110101223323032133122322220000013722000000103
	Values range from 0 (buried residue) to 9 (highly exposed residue)

Top 5 Models predicted by I-TASSER



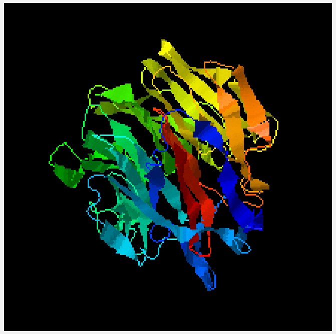
Download Model 1

C-score=-2.09



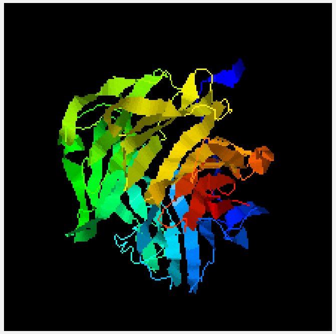
Download Model 2

C-score=-2.42



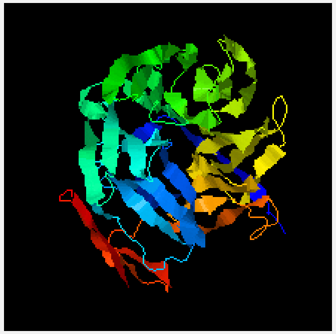
Download Model 3

C-score=-3.44



Download Model 4

C-score=-3.46



Download Model 5

C-score=-3.53

Estimated accuracy of Model1: 0.47±0.15 (TM-score) 11.3±4.5Å (RMSD) (Read more about C-score of generated models)

Proteins with highly similar structure in PDB (as identified by TM-align)

Spin On/Off

Top 10 Identified stuctural analogs in PDB

	Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov.	Download Alignment
◉	1	1rn0A	0.773	3.51	0.066	0.952	Download
◉	2	3vi3A	0.772	3.56	0.085	0.952	Download
◉	3	2xbgA	0.772	2.92	0.094	0.892	Download
◉	4	3ijeA	0.770	3.55	0.063	0.949	Download
◉	5	3cikB	0.752	3.08	0.060	0.890	Download
◉	6	4a7zA	0.747	3.74	0.074	0.928	Download
◉	7	2pbiD	0.747	3.19	0.054	0.890	Download
◉	8	3iz6a	0.744	3.41	0.059	0.904	Download
◉	9	3k71G	0.744	3.70	0.074	0.931	Download
◉	10	1trjA	0.743	3.15	0.068	0.884	Download

(a) Query structure is shown in cartoon, while the structural analog is
(b) Ranking of proteins is based on TM-score of the structural alignme
(c) RMSD[a] is the RMSD between residues that are structurally aligne
(d) IDEN[a] is the percentage sequence identity in the structurally aligne
(e) Cov. represents the coverage of the alignment by TM-align and is e

Navíc predikce funkce a vazebných míst – na základě struktury homologních komplexů

CASP

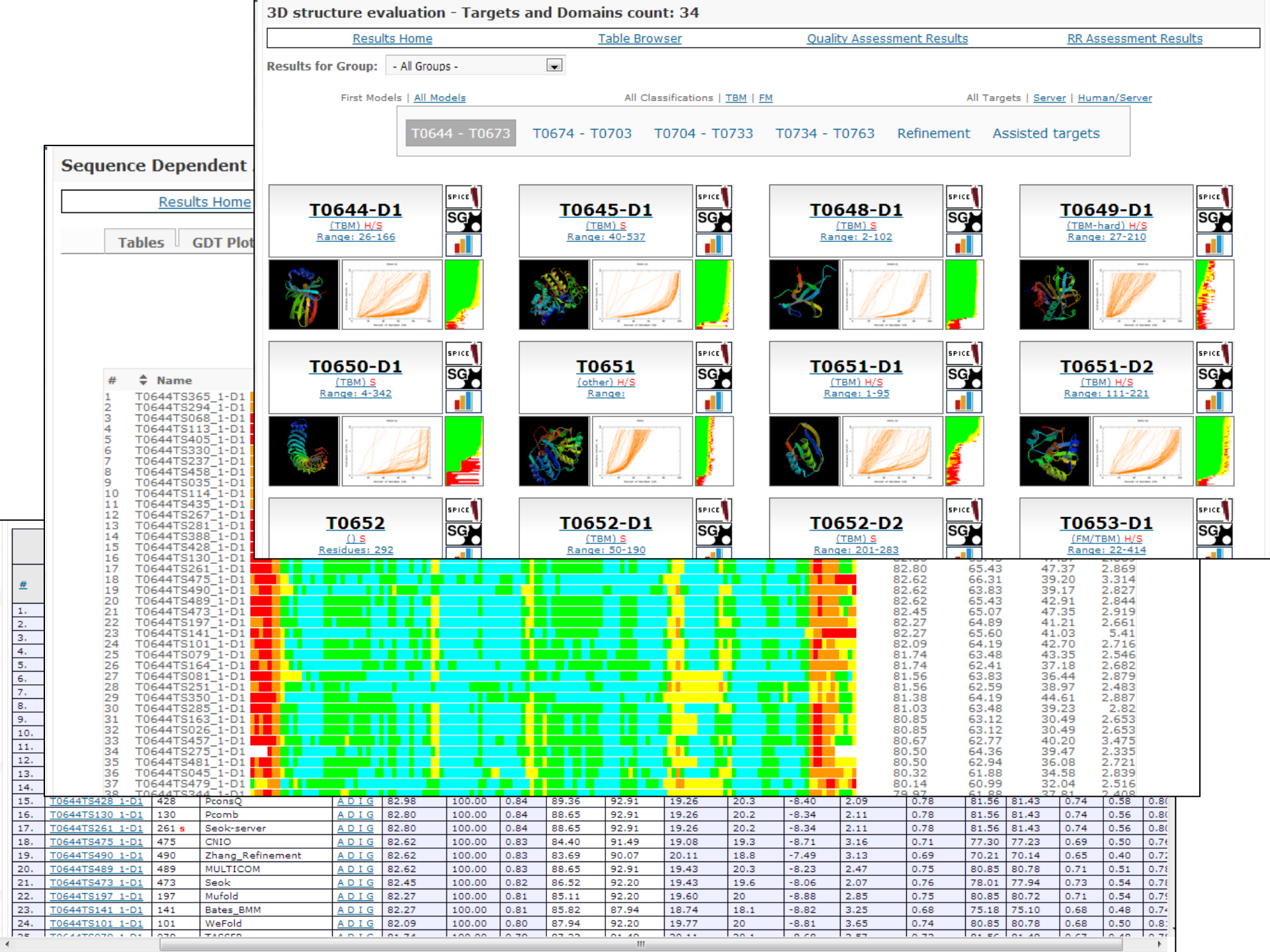
Critical Assessment of Techniques for Protein Structure Prediction

http://predictioncenter.org/casp10/index.cgi

Rozsáhlá analýza predikčních programů

Poslední kolo 2012 (CASP10). Mj. zahrnuje:

- ***Predikce terciárních struktur:***
  - *Template Based Modeling* - je-li dostupný templát
  - *Template free modeling* - bez templátu
  - *Contact-assisted structure modeling* - při znalosti několika dlouhodosahových interakcí
- ***Další kategorie:***
  - Identifikace neuspořádaných oblastí (disordered regions)
  - Funkční predikce (predikce vazebných míst)


3D structure evaluation - Targets and Domains count: 34
Results Home
Table Browser
Quality Assessment Results
RR Assessment Results
Results for Group: - All Groups -
First Models | All Models
All Classifications | TBM | FM
All Targets | Server | Human/Server
T0644 - T0673
T0674 - T0703
T0704 - T0733
T0734 - T0763
Refinement
Assisted targets
T0644-D1
(TBM) H/S
Range: 26-166
T0645-D1
(TBM) S
Range: 40-537
T0648-D1
(TBM) S
Range: 2-102
T0649-D1
(TBM-hard) H/S
Range: 27-210
T0650-D1
(TBM) S
Range: 4-342
T0651
(other) H/S
Range:
T0651-D1
(TBM) H/S
Range: 1-95
T0651-D2
(TBM) H/S
Range: 111-221
T0652
() S
Residues: 292
T0652-D1
(TBM) S
Range: 50-190
T0652-D2
(TBM) S
Range: 201-283
T0653-D1
(FM/TBM) H/S
Range: 22-414
Sequence Dependent
Results Home
Tables
GDT Plot
Name
1 T0644TS365_1-D1
2 T0644TS294_1-D1
3 T0644TS068_1-D1
4 T0644TS113_1-D1
5 T0644TS405_1-D1
6 T0644TS330_1-D1
7 T0644TS237_1-D1
8 T0644TS458_1-D1
9 T0644TS035_1-D1
10 T0644TS114_1-D1
11 T0644TS435_1-D1
12 T0644TS267_1-D1
13 T0644TS281_1-D1
14 T0644TS388_1-D1
15 T0644TS428_1-D1
16 T0644TS130_1-D1
17 T0644TS261_1-D1
18 T0644TS475_1-D1
19 T0644TS490_1-D1
20 T0644TS489_1-D1
21 T0644TS473_1-D1
22 T0644TS197_1-D1
23 T0644TS141_1-D1
24 T0644TS101_1-D1
25 T0644TS079_1-D1
26 T0644TS164_1-D1
27 T0644TS081_1-D1
28 T0644TS251_1-D1
29 T0644TS350_1-D1
30 T0644TS285_1-D1
31 T0644TS163_1-D1
32 T0644TS026_1-D1
33 T0644TS457_1-D1
34 T0644TS275_1-D1
35 T0644TS481_1-D1
36 T0644TS045_1-D1
37 T0644TS479_1-D1
82.80 65.43 47.37 2.869
82.62 66.31 39.20 3.314
82.62 63.83 39.17 2.827
82.62 65.43 42.91 2.844
82.45 65.07 47.35 2.919
82.27 64.89 41.21 2.661
82.27 65.60 41.03 5.41
82.09 64.19 42.70 2.716
81.74 63.48 43.35 2.546
81.74 62.41 37.18 2.682
81.56 63.83 36.44 2.879
81.56 62.59 38.97 2.483
81.38 64.19 44.61 2.887
81.03 63.48 39.23 2.82
80.85 63.12 30.49 2.653
80.85 63.12 30.49 2.653
80.67 62.77 40.20 3.475
80.50 64.36 39.47 2.335
80.50 62.94 36.08 2.721
80.32 61.88 34.58 2.839
80.14 60.99 32.04 2.516
15. T0644TS428_1-D1 428 PconsQ A D I G 82.98 100.00 0.84 89.36 92.91 19.26 20.3 -8.40 2.09 0.78 81.56 81.43 0.74 0.58
16. T0644TS130_1-D1 130 Pcomb A D I G 82.80 100.00 0.84 88.65 92.91 19.26 20.2 -8.34 2.11 0.78 81.56 81.43 0.74 0.56
17. T0644TS261_1-D1 261 s Seok-server A D I G 82.80 100.00 0.84 88.65 92.91 19.26 20.2 -8.34 2.11 0.78 81.56 81.43 0.74 0.56
18. T0644TS475_1-D1 475 CNIO A D I G 82.62 100.00 0.83 84.40 91.49 19.08 19.3 -8.71 3.16 0.71 77.30 77.23 0.69 0.50
19. T0644TS490_1-D1 490 Zhang_Refinement A D I G 82.62 100.00 0.83 83.69 90.07 20.11 18.8 -7.49 3.13 0.69 79.21 70.14 0.65 0.40
20. T0644TS489_1-D1 489 MULTICOM A D I G 82.62 100.00 0.83 88.65 92.91 19.43 20.3 -8.23 2.47 0.75 80.85 80.78 0.71 0.51
21. T0644TS473_1-D1 473 Seok A D I G 82.45 100.00 0.82 86.52 92.20 19.43 19.6 -8.06 2.07 0.76 78.01 77.94 0.73 0.54
22. T0644TS197_1-D1 197 Mufold A D I G 82.27 100.00 0.81 85.11 92.20 19.60 20 -8.88 2.85 0.75 80.85 80.72 0.71 0.54
23. T0644TS141_1-D1 141 Bates_BMM A D I G 82.27 100.00 0.81 85.82 87.94 18.74 18.1 -8.82 3.25 0.68 75.18 75.10 0.68 0.48
24. T0644TS101_1-D1 101 WeFold A D I G 82.09 100.00 0.80 87.94 92.20 19.77 20 -8.81 3.65 0.74 80.85 80.78 0.68 0.50

Analýza 3D struktur

- Určení **strukturních prvků** (sekundární struktura, motivy, foldy) a zařazení do příslušných nadrodin
- **Povrchy** – přístupnost pro solvent, hydrofobicita, analýza kavit a tunelů
- **Vazebná místa** – predikce funkce
- **Interakce** (protein-protein – 4D, protein-DNA, protein-ligand) – plochy, energie, vazby
- **Homology** – hledání, porovnávání

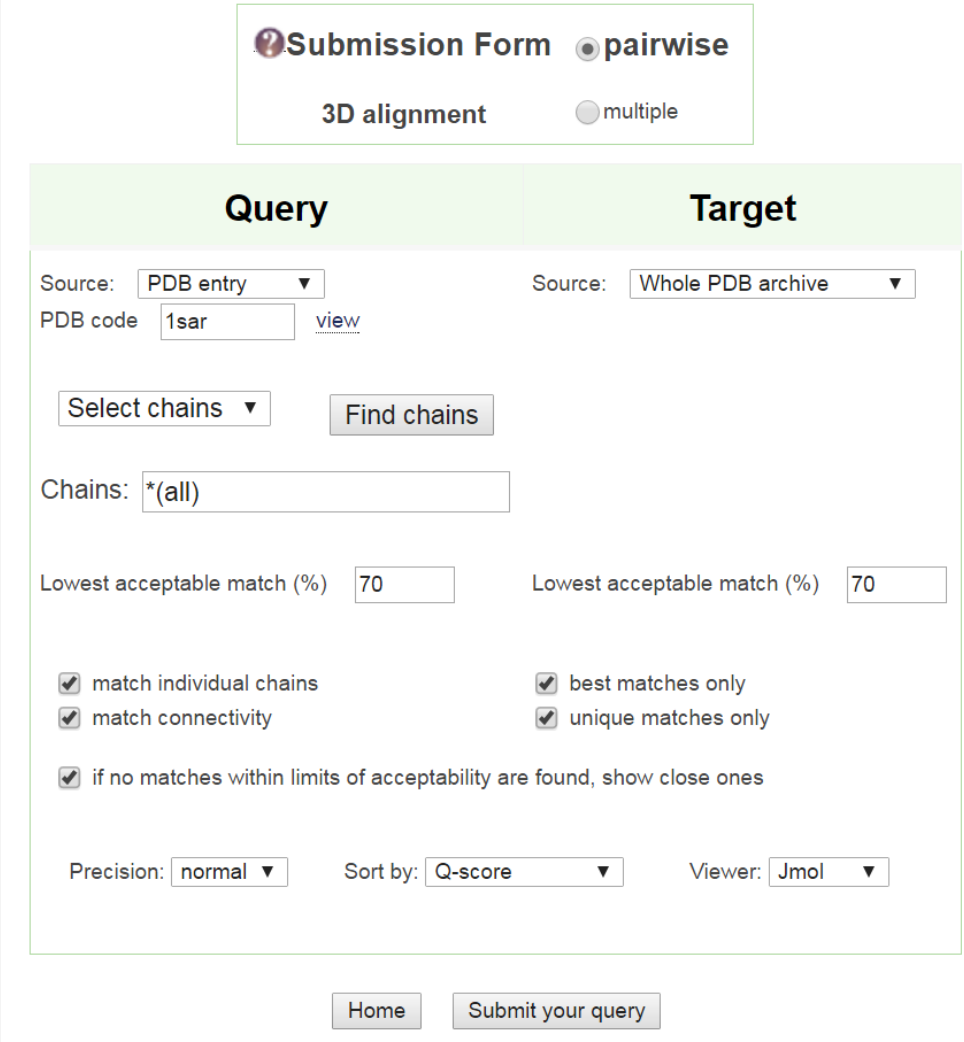
Hledání 3D-homologních proteinů

Hledání na úrovni terciárních struktur může být časově (i výpočetně) náročné.

- PDBeFold – prohledává PDB databázi
- Dali – prohledává PDB databázi
- FATCAT
- BackPhyre – umožňuje prohledání konkrétního genomu proti zadané 3D struktuře

PDBeFold

Porovnání 3D struktur z PDB / SCOP pomocí 2D elementů



Úloha

- Pomocí serveru **PDBeFold** (http://www.ebi.ac.uk/msd-srv/ssm/) analyzujte N-terminální doménu proteinu BC2L-C. (PDB kód: 2WQ4). Nalezněte proteiny s nejvyšší strukturní homologií k tomuto proteinu.

Protein Data Bank in Europe

Bringing Structure to Biology

PDBeFold

Structure Alignment Results.

Query: pdb entry 2wq4, chain A : 134 residues.

N-TERMINAL DOMAIN OF BC2L-C LECTIN FROM BURKHOLDERIA CENOCEPACIA

Examined 86833 entries, (217743 chains). **Displaying Matches 1-17 of 17.**

Back to query | Sort by Q-score | arrange by SCOP family

##	Scoring			RMSD	N_{align}	N_g	$\%_{seq}$	Query	Target (PDB entry)				
	Q	P	Z					$\%_{sse}$	Match	$\%_{sse}$	N_{res}	×	Title
1	1.00	45.6	20.3	0.00	134	0	100	100	2wq4:A	100	134		N-TERMINAL DOMAIN OF BC2L-C LECTIN FROM BURKHOLDERIA CENOCEPACIA
2	0.94	29.5	16.3	0.44	130	1	100	100	2wq4:C	100	131		N-TERMINAL DOMAIN OF BC2L-C LECTIN FROM BURKHOLDERIA CENOCEPACIA
3	0.93	26.2	15.3	0.60	130	0	100	91	2wq4:B	100	130		N-TERMINAL DOMAIN OF BC2L-C LECTIN FROM BURKHOLDERIA CENOCEPACIA
4	0.23	0.0	3.1	3.86	89	8	8	73	3t30:C	100	95		HUMAN NUCLEOPLASMIN (NPM2): A HISTONE CHAPERONE IN OOCYTES AND EARLY EMBRYOS
5	0.23	0.0	3.2	3.78	86	9	8	73	3t30:B	100	92		HUMAN NUCLEOPLASMIN (NPM2): A HISTONE CHAPERONE IN OOCYTES AND EARLY EMBRYOS
6	0.23	0.0	3.3	3.76	85	11	8	73	3t30:I	100	92		HUMAN NUCLEOPLASMIN (NPM2): A HISTONE CHAPERONE IN OOCYTES AND EARLY EMBRYOS
7	0.22	0.0	3.2	3.82	84	8	8	73	3t30:J	100	90		HUMAN NUCLEOPLASMIN (NPM2): A HISTONE CHAPERONE IN OOCYTES AND EARLY EMBRYOS

DALI http://ekhidna.biocenter.helsinki.fi/dali_server/start

- **Dali server** - Prohledává PDB databázi za účelem nalezení 3D homologů k zadané struktuře
- **Dali Database** – databáze homologů známých struktur
- **Pairwise Dali Light** - Umožňuje porovnat dvě 3D struktury navzájem

Hasegawa H, Holm L (2009) Advances and pitfalls of protein structural alignment. Curr. Opin. Struct. Biol. 19, 341-348.

Dali server

Institute of Biotechnology

SERVICES & TOOLS | GROUP MEMBERS | NEWS & VACANCIES | RESEARCH | PUBLICATIONS

Protein Structure Database Searching by DaliLite v. 3

The Dali server is a network service for comparing protein structures in 3D. You submit the coordinates of a query protein structure and Dali compares them against those in the Protein Data Bank (PDB). You receive an email notification when the search has finished. In favourable cases, comparing 3D structures may reveal biologically interesting similarities that are not detectable by comparing sequences.

Requests can also be submitted by e-mail to *dali-server at helsinki dot fi*. The body of the e-mail message must contain atomic coordinates in PDB format.

If you want to know the structural neighbours of a protein already in the Protein Data Bank (PDB), you can find them in the Dali Database.

If you want to superimpose two particular structures, you can do it in the pairwise DaliLite server.

Upload a structure:

Vybrat soubor Soubor nevybrán

Or enter PDB identifier: **chain:** (optional)

(Keyword search for PDB identifiers)

Job name:

(optional)

Enter email address for notification:

(recommended)

lower priority queue

submit clear

Most jobs finish within an hour, but if a queue builds up, then it takes longer.

Example

Úloha

- Použijte server **Dali** (http://ekhidna.biocenter.helsinki.fi/dali_server/start) a najděte blízké strukturní homology k některému z proteinů v PDB databázi (http://pdb.rcsb.org). Využijte **Dali Database**.

Query: mol1A

MOLECULE: BETA-FRUCTOFURANOSIDASE;

Select neighbours (check boxes) for viewing as multiple structural alignment or 3D superimposition. The list of neighbours is sorted by Z-score. Similarities with a Z-score lower than 2 are spurious. Each neighbour has links to pairwise structural alignment with the query structure, to pre-computed structural neighbours in the Dali Database, and to the PDB format coordinate file where the neighbour is superimposed onto the query structure.

Structural Alignment ☐ Expand gaps 3D Superimposition (Jmol Applet) Reset Selection

Summary

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	3piq-A	69.3	0.0	523	523	100	PDB	MOLECULE: BETA-FRUCTOFURANOSIDASE;
2:	3pij-A	67.7	0.2	523	523	100	PDB	MOLECULE: BETA-FRUCTOFURANOSIDASE;
3:	3piq-B	65.1	0.6	520	526	100	PDB	MOLECULE: BETA-FRUCTOFURANOSIDASE;
4:	3pij-B	65.0	0.6	520	526	100	PDB	MOLECULE: BETA-FRUCTOFURANOSIDASE;
5:	1w2t-A	42.9	2.1	423	432	28	PDB	MOLECULE: BETA FRUCTOSIDASE;
6:	1uyp-B	42.9	2.1	423	432	29	PDB	MOLECULE: BETA-FRUCTOSIDASE;
7:	1uyp-A	42.9	2.1	423	432	29	PDB	MOLECULE: BETA-FRUCTOSIDASE;
8:	1w2t-E	42.7	2.1	422	432	29	PDB	MOLECULE: BETA FRUCTOSIDASE;
9:	1uyp-F	42.7	2.1	422	432	29	PDB	MOLECULE: BETA-FRUCTOSIDASE;
10:	1uyp-E	42.7	2.1	421	432	29	PDB	MOLECULE: BETA-FRUCTOSIDASE;
11:	1w2t-D	42.7	2.2	422	432	29	PDB	MOLECULE: BETA FRUCTOSIDASE;
12:	1uyp-D	42.7	2.2	423	432	29	PDB	MOLECULE: BETA-FRUCTOSIDASE;
13:	1w2t-F	42.6	2.2	422	432	29	PDB	MOLECULE: BETA FRUCTOSIDASE;
14:	1w2t-B	42.5	2.1	423	432	28	PDB	MOLECULE: BETA FRUCTOSIDASE;
15:	1uyp-C	42.5	2.1	422	432	29	PDB	MOLECULE: BETA-FRUCTOSIDASE;
16:	1w2t-C	42.1	2.1	421	432	29	PDB	MOLECULE: BETA FRUCTOSIDASE;

Pairwise Structural Alignments

Notation: three-state secondary structure definitions by DSSP (reduced to H=helix, E=sheet, L=coil) are shown above the amino acid sequence. Structurally equivalent residues are in uppercase, structurally non-equivalent residues (e.g. in loops) are in lowercase. Amino acid identities are marked by vertical bars.

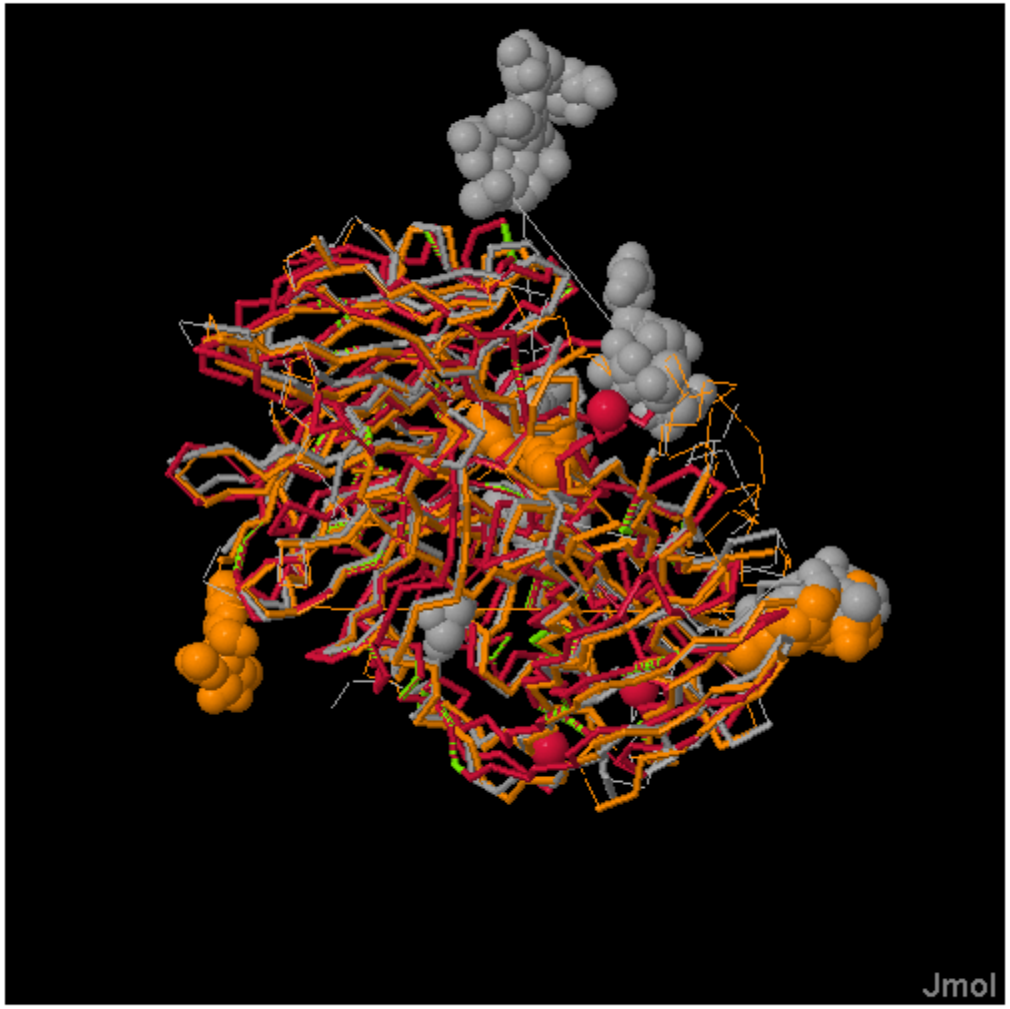
No 1: Query=mol1A Sbjct=3pigA Z-score=69.3

back to top

```
DSSP  LLLLLLLLLLLLLLLHHHHHHHHHHHHHHHHHLLLLLLLLLLLLLLLEEEEEEEEEEE
Query MTDFTPETPVLTPIRDHAAELAKAEAGVAEMAAKRNNRWYPKYHIASNGGWINDPNGLCF   60
ident ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct MTDFTPETPVLTPIRDHAAELAKAEAGVAEMAAKRNNRWYPKYHIASNGGWINDPNGLCF   60
DSSP  LLLLLLLLLLLLLLLHHHHHHHHHHHHHHHHHLLLLLLLLLLLLLLLEEEEEEEEEEE


DSSP  ELLEEEEEEEELLLLLLLLEEEEEELLLLEEELLLLLLLLHHHLLEEEEEEEEL
Query YKGRWHVFYQLHPYGTQWGPMHWGHVSSTDMLNWKREPIMFAPSLEQEKDGVFSGSAVID  120
ident ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct YKGRWHVFYQLHPYGTQWGPMHWGHVSSTDMLNWKREPIMFAPSLEQEKDGVFSGSAVID  120
DSSP  ELLEEEEEEEELLLLLLLLEEEEEELLLLEEELLLLLLLLHHHLLEEEEEEEEL


DSSP  LLLLEEEEEEEEELLLLLHHHLEEEEEEEEELLLLLLEEEEEEELLLHHHEEEEEE
Query DNGDLRFYYTGHRWANGHDNTGGDWQVQMTALPDNDELTSATKQGMIIDCPTDKVDHHYR  180
ident ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct DNGDLRFYYTGHRWANGHDNTGGDWQVQMTALPDNDELTSATKQGMIIDCPTDKVDHHYR  180
DSSP  LLLLEEEEEEEEELLLLLHHHLEEEEEEEEELLLLLLEEEEEEELLLHHHEEEEEE


DSSP  EEEEEELLEEEEEEEEELLLLEEEEEEELLLLLEEEEEEEELLLLLLEEEEEEE
Query DPKVWKTGDTWYMTFGVSSADKRGQMWLFSSKDMVRWEYERVLFQHPDPDVFMLECPDFS  240
ident ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct DPKVWKTGDTWYMTFGVSSADKRGQMWLFSSKDMVRWEYERVLFQHPDPDVFMLECPDFS  240
```

DaliLite Results: Superimposed structures

Starting a Jmol applet; it may take a few seconds. If you are loading many structures, you can monitor progress from Jmol's control panel (right-click Jmol wind
(see About Jmol -> Java memory usage), then close all Jmol applets and other Java applications, go back to the summary page and (i) try again, or (ii) select fe



Toggle: ☑ spinning ☑ superimpose all ligands Clear labels

First structure's backbone: ◉ CA trace ○ Cartoon Rockets. Matched structures' backbone: ○ hide all ◉ show all

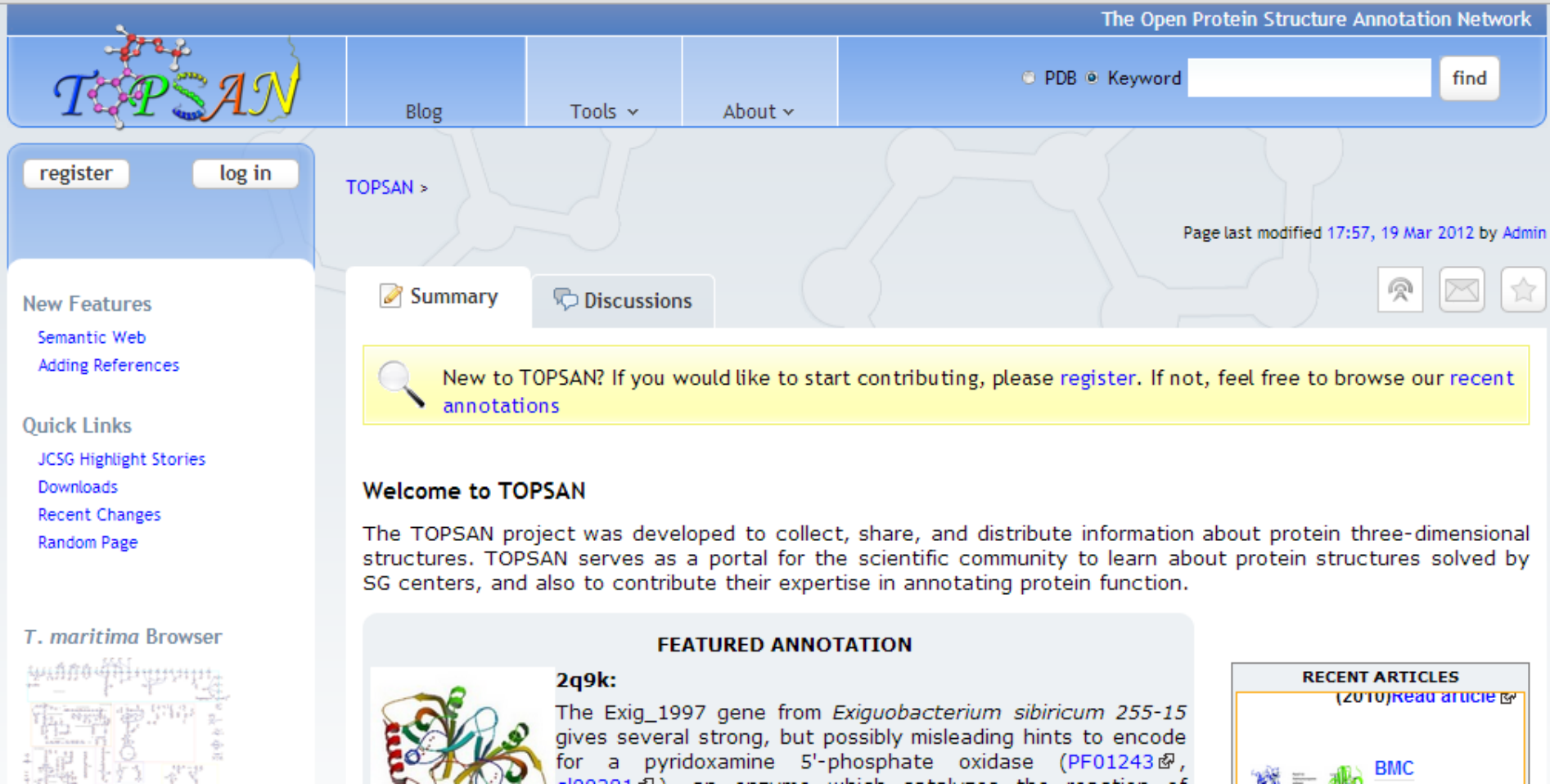
DaliLite Results: Superimposed structures

Starting a Jmol applet; it may take a few seconds. If you are loading many structures, you can moni
(see About Jmol -> Java memory usage), then close all Jmol applets and other Java applications, g

TOPSAN

http://proteins.burnham.org/

Kombinace analýzy a kontextového vyhledávání



Úloha

- Pomocí serveru **Topsan** (http://proteins.burnham.org/) zjistěte, do jaké strukturní rodiny patří protein z *Xanthomonas campestris* (PDB ID: 2qjw) a zda jsou známy nějaké jeho strukturní homology.
- Pokud možno, porovnejte strukturu 2qjw a nejbližšího homologu.

Page last modified 21:51, 8 May 2012 by Admin

Summary | Discussions

Title	Crystal structure of uncharacterized protein XCC1541 (NP_636912.1) from Xanthomonas campestris at 1.35 A resolution. To be published
Site	JCSG
PDB Id	2qjw
Target Id	372467

Molecular Characteristics

Source	*Xanthomonas campestris pv. campestris str. atcc 33913*		
Alias Ids	TPS1610,NP_636912.1, BIG_556	Molecular Weight	18712.45 Da.
Residues	175	Isoelectric Point	6.37
Sequence	msrghcilahgfesgpdalkvtalaevaerlgwtherpdftdldarrdlgqlgdvrgrlqrlleiaraa tekgpvvlagsslgsyiaaqvslqvptralflmvpptkmgplpaldaaavpisivhawhdelipaadvi awaqarsarlllvddghrlgahvqaasrafaellqsl BLAST FFAS		

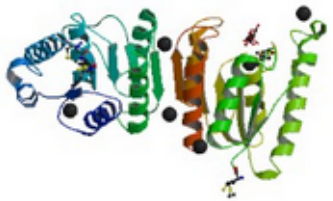
Structure Determination

Method	XRAY	Chains	4
Resolution (Å)	1.35	Rfree	0.216
Matthews' coefficent	2.14	Rfactor	0.182
Waters	767	Solvent Content	42.51

Ligand Information

Ligands

Metals



Jmol

RCSB PDB Comparison Tool

Compare the following two proteins:

PDB1: 2qjw Chain1:

PDB2: Chain2:

-- Select Method --

compare

New Features

Semantic Web

Adding References

Quick Links

JCSG Highlight Stories

Downloads

Recent Changes

Random Page

Page Authors

tinab (3 edits)

pascual (3 edits)

T. maritima Browser

RESOURCES

Google Scholar output for 2qjw

1. CMASA: an accurate algorithm for detecting local protein structural similarity and its application to enzyme catalytic site annotation
GH Li, JF Huang - BMC bioinformatics, 2010 - biomedcentral.com

2. Distributed structure determination at the JCSG
H van den Bedem, G Wolf, Q Xu - Section D: Biological , 2011 - scripts.iucr.org

Protein Summary

Gene XCC1541 from *Xanthomonas campestris* encodes the NP_636912 protein from the COG2945 group (e-val=1e-4). Its genomic neighbor, XCC1542, is annotated as acetyltransferase.

pre-SCOP classifies 2qjw in the alpha/beta class, alpha/beta hydrolases superfamily, acetyl xylan esterase-like family. According to DALI, 2qjw shows significant structural similarity to PDB entries 2i3d (1.8 Å rmsd, 158 aligned residues, 10% sequence id; Z=21), 1uxo (2.3 Å rmsd, 155 aligned residues, 14% sequence id; Z=17) and 3bdv (2.3 Å, 153 aligned residues, 20% seq.id.; Z=18). Similar values are obtained for lipases (PDB ids: 3d2c, 1t2n, 1t4m and 2qxt, rmsd 2.2 Å, 147 aligned residues, 11% sequence id; Z=17) and for an a/b hydrolase (PDB id: 2fuk, rmsd 2.0 Å, 157 aligned residues, 18% sequence id; Z=20). 1uxo structure has been described as an a/b hydrolase (1).

To do: check if 2qjw ligands (P6G or TLA) might be mimetics of biological substrate.

Ligand Summary

References

References | Images | Files | Tags

Tag page add tag

BackPhyre

BackPhyre

Use Phyre in reverse. Submit a PDB structure and search that structure against a wide range of genomes

Backphyre options

E-mail address

Optional job description

Upload PDB format file — Vybrat soubor — Soubor nevybrán — Single chain only

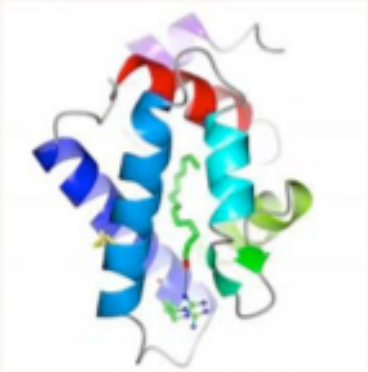
Choose 1 or more genomes to search (ctrl-click or cmd-click)

- Arabidopsis_thaliana
- Bdellovibrio_bacteriovorus
- Caenorhabditis_elegans.WS220.66
- Drosophila_melanogaster
- Homo_sapiens
- Mus_musculus
- Mycobacterium_tuberculosis_CDC1551
- Plasmodium_falciparum
- Saccharomyces_cerevisiae
- Sulfolobus_solfataricus_P2
- Thermoplasma_acidophilum
- Agrobacterium_tumefaciens_C58_Cereon
- Bacillus_subtilis
- Bartonella_henselae_Houston-1
- Corynebacterium_diphtheriae

Submit BackPhyre job

Phyre2: BackPhyre

User structure



Extract sequence and Secondary structure information

SVYDAAAQLTADVKKDLRDSW
KVIGSDKKGNGVALMTTLFAD
NQETIGYFKRLGNVSQGMAND
KLRGHSITLMYALQNFIDQLD
NPDSLDLVCS.....

PSI-Blast vs sequence database

HMM

Hidden Markov model of user structure

Hidden Markov Model DB of **Genomes**

hmm-hmm matching

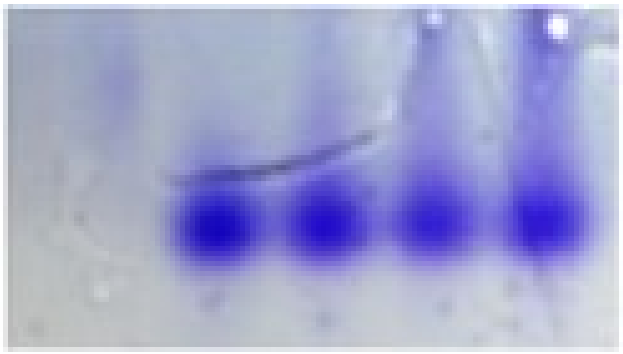
Rank	Hit	Confid-ence
1	Gi...	
2	Gi..	
3	Gi..	
.	.	
.	.	

Ranked list of genome hits

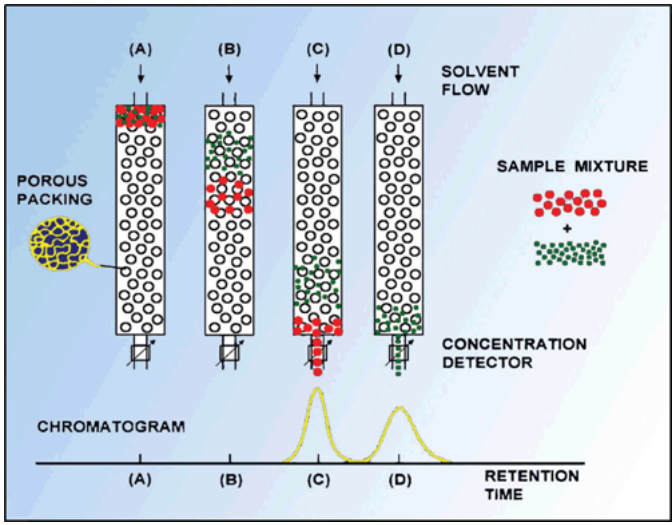
Kvartérní struktura

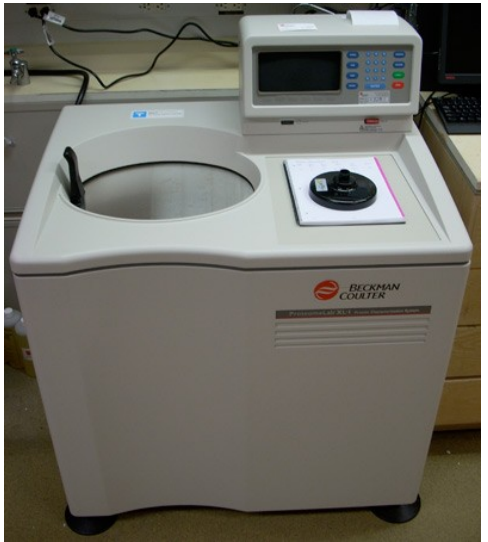
- Spojování několika řetězců (monomerů) do funkčních jednotek
- Homo-oligomery, hetero-oligomery
- Komplexy proteinů s dalšími makromolekulami – ribozomy, viry
- Nadmolekulární komplexy – buněčná membrána

Určení kvartérní struktury



- Nativní gelová elektroforéza
- Gelová permeační chromatografie (GPC)
- Analytická ultracentrifugace (AUC)
- Light scattering (MALS)
- Analýza 3D struktury

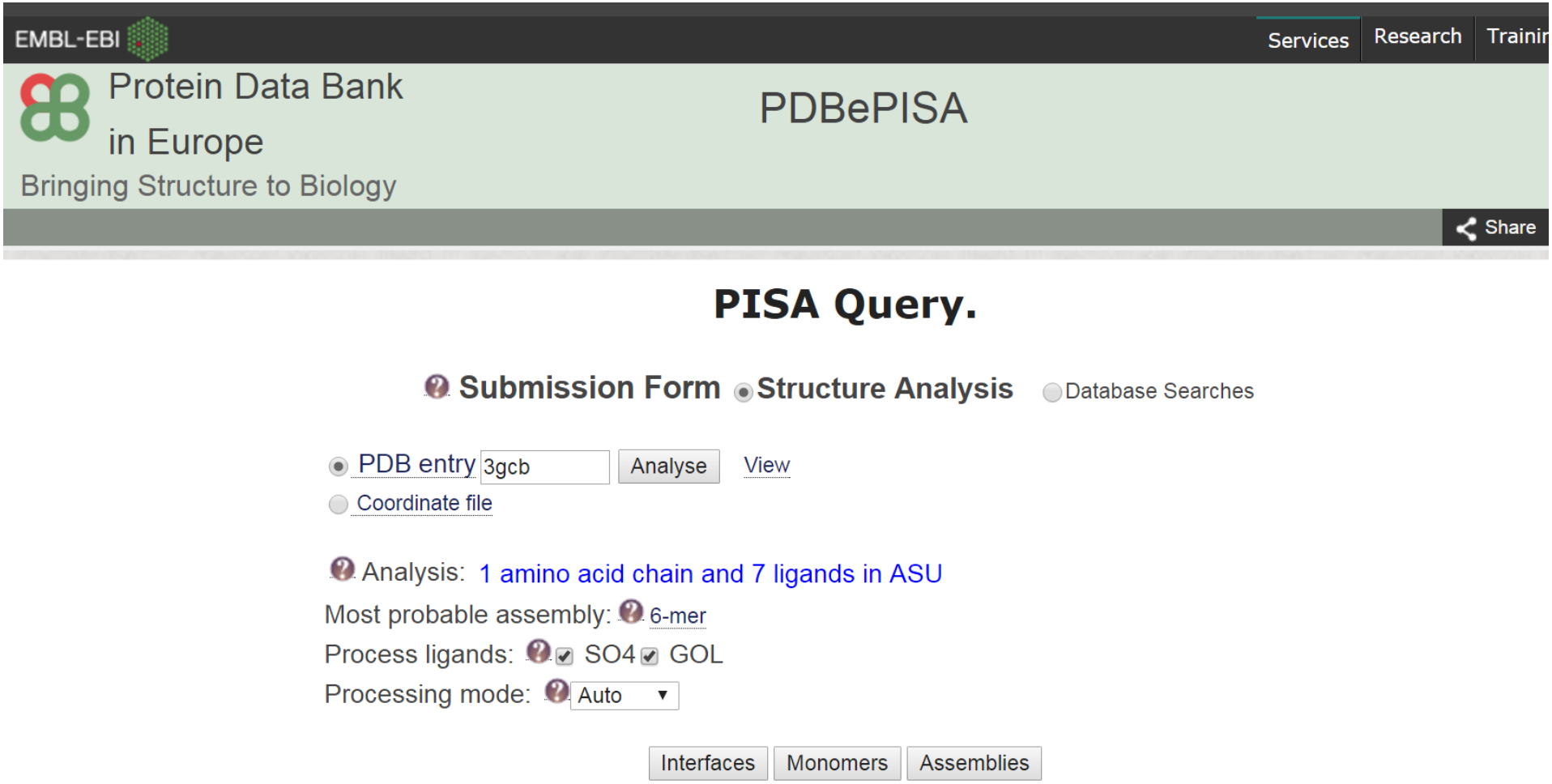




Běžná je kombinace více metod

Analýza kvartérní struktury - PDBePISA

http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html



Úloha

Analyzujte pomocí serveru **PDBePISA** (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) lidskou glutamátdehydrogenasu:

- Tvoří lidská glutamátdehydrogenasa oligomer?
- S kolika dalšími molekulami tvoří každá molekula dehydrogenasy stabilní vazbu?

PISA Assembly List.

Session Map (id=113-7J-1AA)

Start | Interfaces | Interface Search | Monomers | Assemblies

PQS sets 1 to 3 of total 3

Probable Assemblies in PDB 1l1f crystal.

Space symmetry group: P 1. Resolution: 2.70 Å

STRUCTURE OF HUMAN GLUTAMATE DEHYDROGENASE-APO FORM

Complex Analysis of the complex represented by the coordinate section only of the PDB entry.

Analysis of protein interfaces suggests that the following quaternary structures are stable in solution.

Details | Download | View | XML

PQS set NN	«»	mm Size	Formula	Composition	Id	Biomol R350	Stable	Surface area, sq. Å	Buried area, sq. Å	ΔG^{int}, kcal/mol	ΔG^{diss}, kcal/mol
1	◉	6	A_6	ABCDEF	1	1	yes	103940	30360	-146.2	52.1
2	○	3	A_3	ABF	2	–	yes	57610	9550	-54.0	54.2
	○	3	A_3	CDE	2	–	yes	57540	9610	-53.6	53.5
3	○	2	A_2	AE	3	–	yes	41790	2950	-13.3	7.7
	○	2	A_2	CF	3	–	yes	41990	2930	-13.1	7.7
	○	2	A_2	BD	3	–	yes	41690	2950	-12.6	7.6

Details | Download | View | XML

l 3

/09/2014]

Engaged interfaces

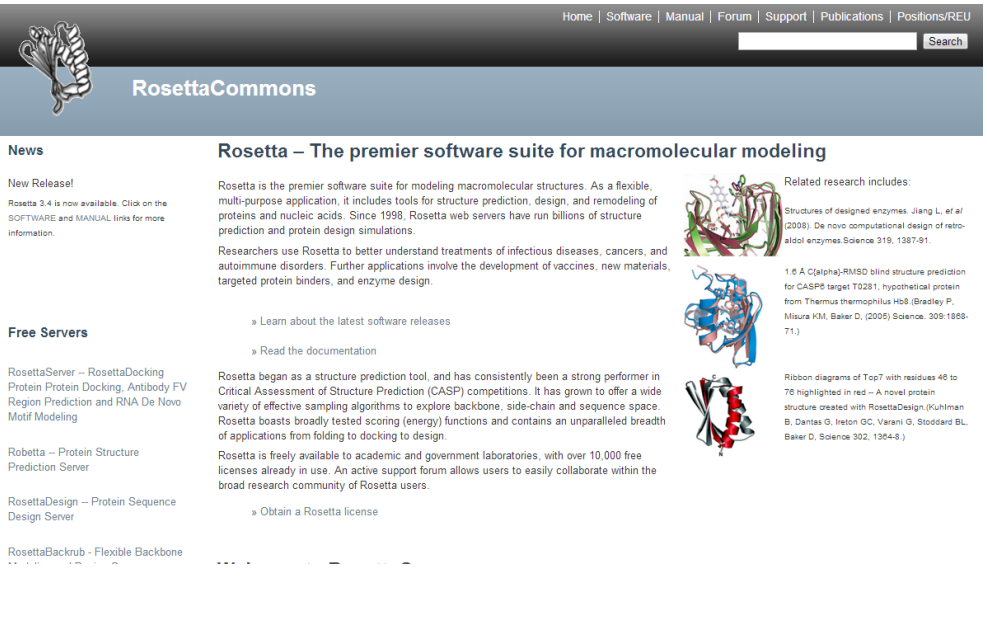
Id	##	Interfacing structures	N_{occ}	Diss.	Sym.ID	Buried area, Å^2	Δ^iG, kcal/mol	N_{HB}	N_{SB}	N_{DS}	CSS
1	1	E + D	1		1_555	1624.6 (5%)	-17.5 (12%)	20 (10%)	4 (12%)	0	1.000
	2	F + A	1		1_555	1593.7 (5%)	-17.8 (12%)	24 (12%)	4 (12%)	0	1.000
	3	F + B	1		1_555	1591.8 (5%)	-18.0 (12%)	21 (11%)	4 (12%)	0	1.000
	4	D + C	1		1_555	1589.6 (5%)	-17.9 (12%)	25 (13%)	5 (15%)	0	1.000
	5	E + C	1		1_555	1589.0 (5%)	-18.1 (12%)	20 (10%)	4 (12%)	0	1.000
	6	B + A	1		1_555	1588.1 (5%)	-18.2 (12%)	21 (11%)	4 (12%)	0	1.000
					Average:	1596.1 (5%)	-17.9 (12%)	22 (11%)	4 (12%)	0	1.000
2	7	E + A	1	×	1_555	1475.7 (5%)	-13.3 (9%)	20 (10%)	2 (6%)	0	1.000
	8	D + B	1	×	1_555	1472.8 (5%)	-12.6 (9%)	21 (11%)	3 (9%)	0	1.000
	9	F + C	1	×	1_555	1464.0 (5%)	-13.1 (9%)	20 (10%)	3 (9%)	0	1.000
					Average:	1470.8 (5%)	-13.0 (9%)	20 (10%)	3 (9%)	0	1.000
3	10	C + A	1	×	1_555	386.9 (1%)	0.4 (0%)	2 (1%)	0 (0%)	0	0.009
	11	E + B	1	×	1_555	384.9 (1%)	0.0 (0%)	2 (1%)	0 (0%)	0	0.009
	12	F + D	1	×	1_555	380.8 (1%)	0.1 (0%)	2 (1%)	0 (0%)	0	0.009
					Average:	384.2 (1%)	0.2 (0%)	2 (1%)	0 (0%)	0	0.009
12	23	F + E	1	×	1_555	13.1 (0%)	-0.0 (0%)	0 (0%)	0 (0%)	0	0.000
	24	C + B	1	×	1_555	12.9 (0%)	-0.0 (0%)	0 (0%)	0 (0%)	0	0.000
	25	D + A	1	×	1_555	11.7 (0%)	-0.0 (0%)	0 (0%)	0 (0%)	0	0.000
					Average:	12.6 (0%)	-0.0 (0%)	0 (0%)	0 (0%)	0	0.000

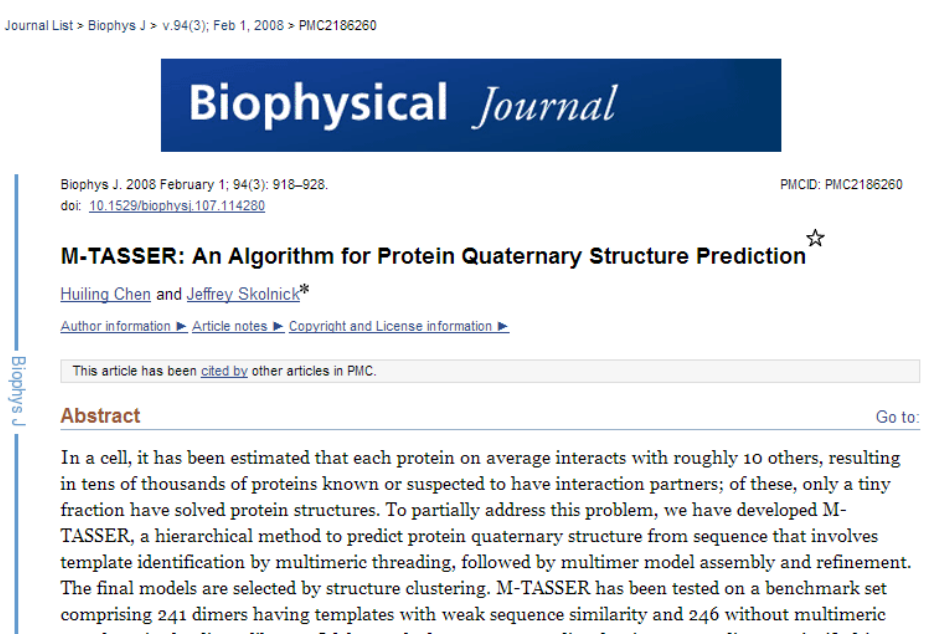
View Details XML

Predikce kvartérní struktury

SW dosud nedokonalý, často nedostupný online

- Rosetta
- M-tasser
- Protein-protein docking

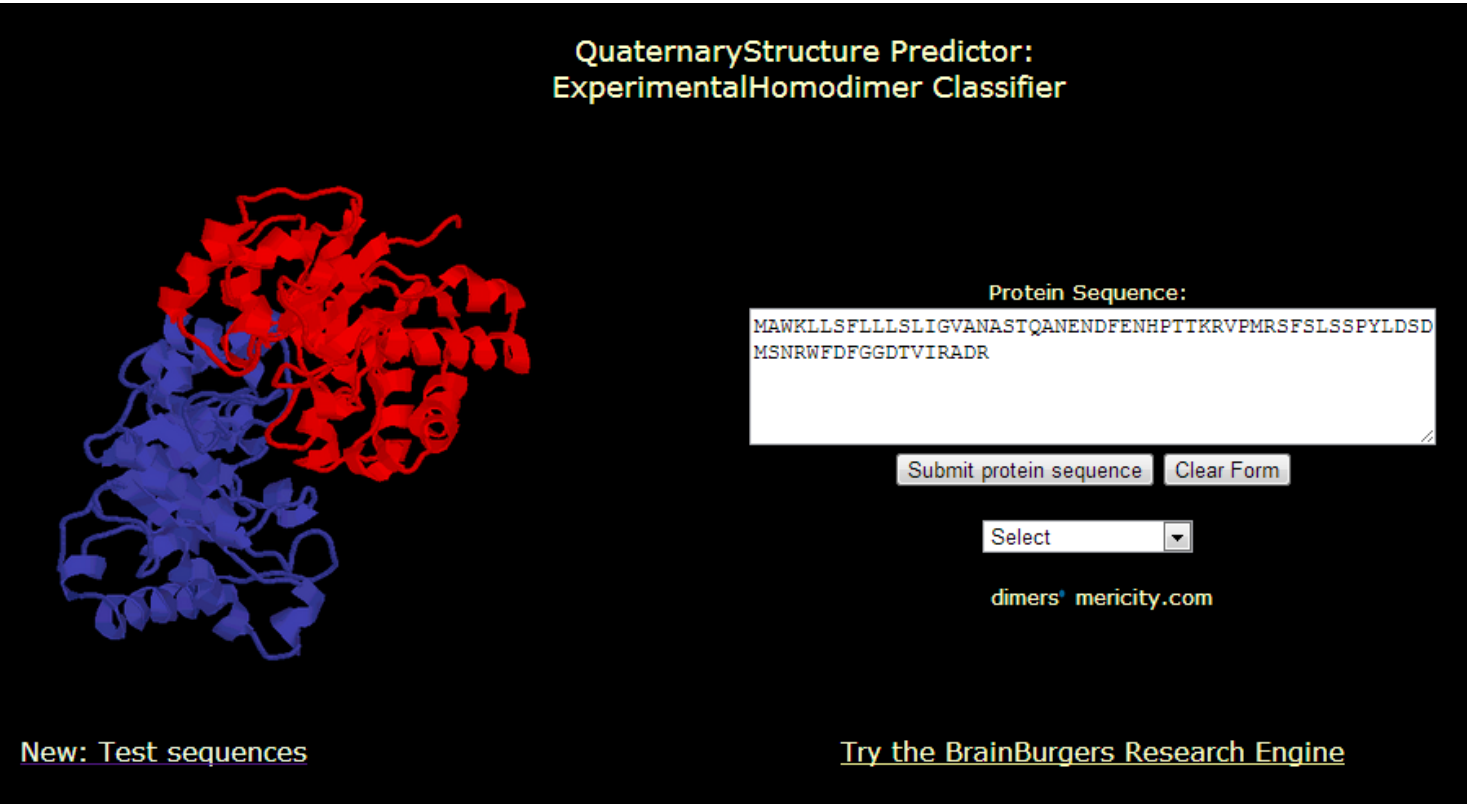




Quaternary structure predictor

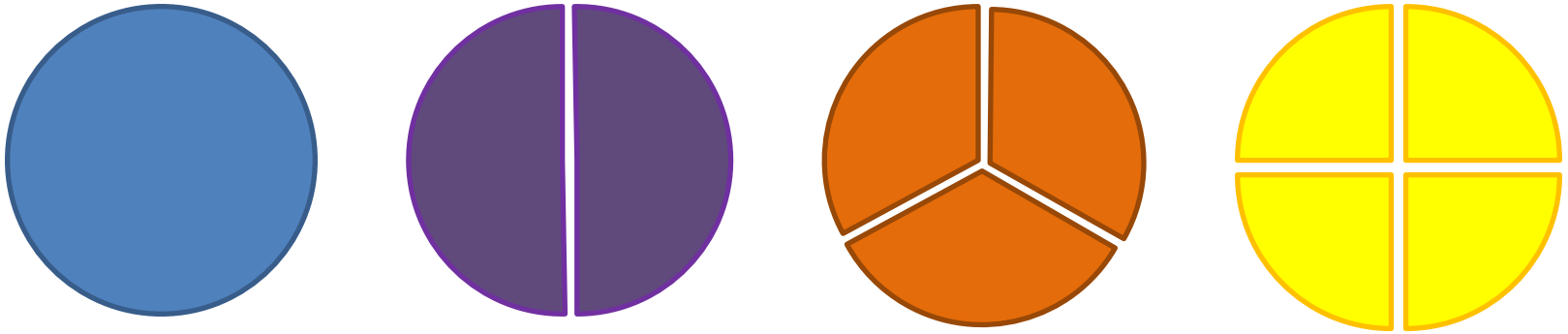
Homodimer classifier http://www.mericity.com/
(již delší dobu nefunkční ?!)

- Predikoval schopnost proteinu vytvářet homodimery na základě sekvence.



Literatura: *Robert Garian: Prediction of quaternary structure from primary structure, Bioinformatics 17 (6) 2001, 551–556*

Oligomer nebo repetice?



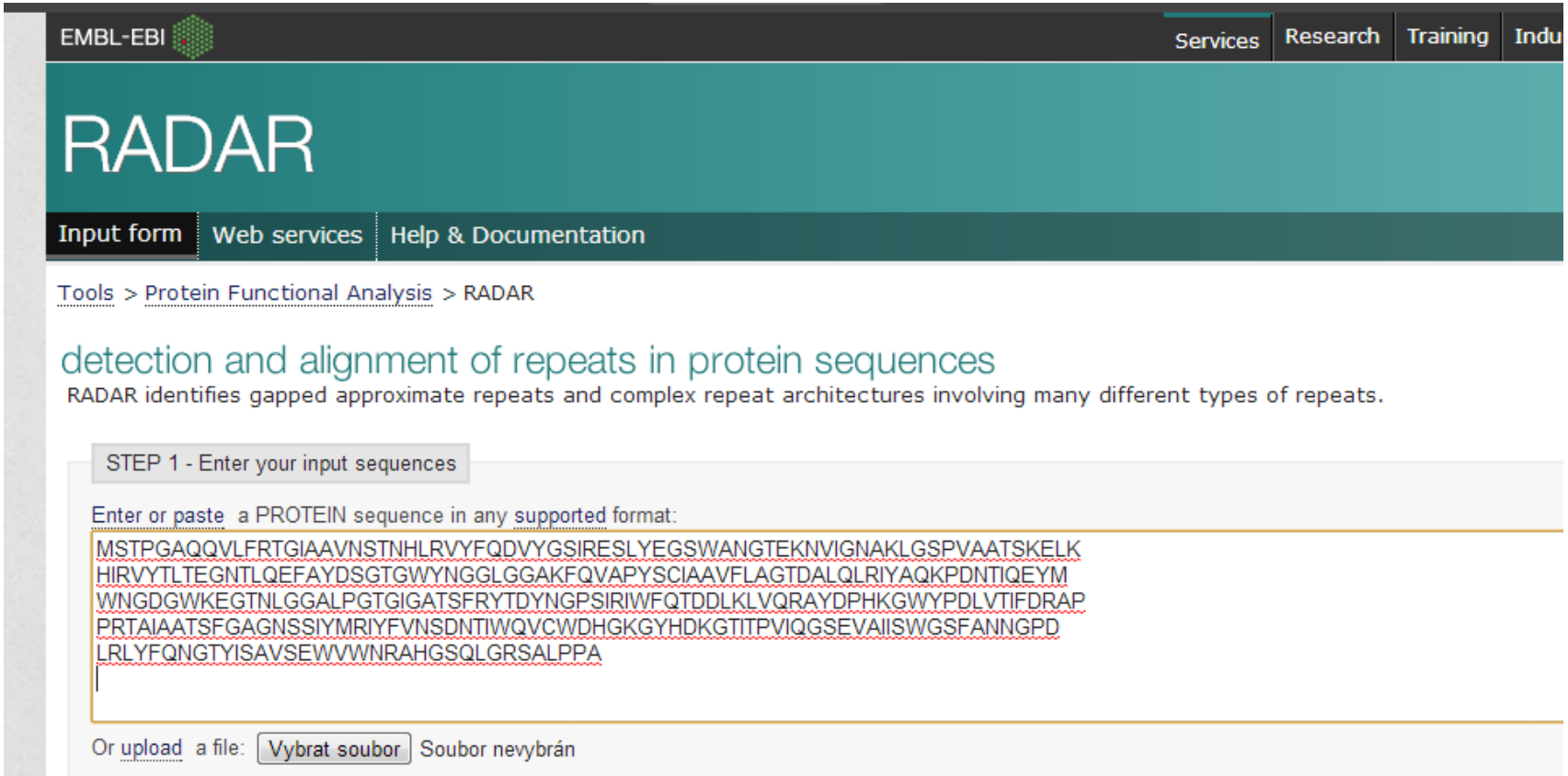
- **Homo-oligomer**: kratší protein = kratší gen = ušetřené místo

- **Repetice**: možnost mutací = větší variabilita

Analýza repetic - RADAR

Na serveru EBI (http://www.ebi.ac.uk/Tools/pfa/radar/)

Analyzuje proteinovou sekvenci (1D struktura) a hledá v ní možné repetice.



Analýza repetic - RADAR

RADAR

Input form | Web services | Help & Documentation

Tools > Protein Functional Analysis > RADAR

Results for job radar-I20130329-171523-0632-74301799-pg

Summary | Submission Details

View/download raw output file

```
--------------------------------------------------------------------------------
No. of Repeats|Total Score|Length  |Diagonal| BW-From|   BW-To|   Level
             6|     377.21|      50|      51|      81|     130|       1
--------------------------------------------------------------------------------
    8-   29 (25.67/ 7.06)        ............................QVL.FRTGIAAV...NS..TN....HLRVYFQ
   33-   77 (57.29/22.58)        GSIRESLYE.G.SWANGTeknvIGNAKLG.S.P...VAA....TS..KE..LKHIRVYTL
   81-  130 (89.97/38.62)        NTLQEFAYDSGTGWYNGG....LGGAKFQVA.PYSCIAAV.FLAG..TD..ALQLRIYAQ
  134-  181 (74.05/30.80)        NTIQEYMWN.GDGWKEGT...nLGGA.L....PGTGIGAT.SFRY..TDynGPSIRIWFQ
  185-  232 (60.06/23.93)        LKLVQRAYDPHKGWYPDL....VTI..FDRApPRTAIAATsFGAG..NS..SIYMRIY..
  238-  286 (70.16/28.89)        NTIWQVCWDHGKGYHDKG...tITPVIQG.....SEVAII.SWGSfaNN..GPDLRLYFQ
--------------------------------------------------------------------------------
```

Úloha

Detekujte repetice v sekvenci domény ARC4 z lidské Tankyrasy 2. Použijte program **RADAR** (http://www.ebi.ac.uk/Tools/pfa/radar/) :

Sekvence ARC4:

GAMGNSEADRQLLEAAKAGDVETVKKLCTVQSVNCR
DIEGRQSTPLHFAAGYNRVSVVEYLLQHGADVHAKD
KGGLVPLHNACSYGHYEVAELLVKHGAVVNVADLWK
FTPLHEAAAKGKYEICKLLLQHGADPTKKNRDGNTP
LDLVKDGDTDIQDLLRGDAAL

Analýza repetic - RADAR

RADAR

Input form | Web services | Help & Documentation

Tools > Protein Functional Analysis > RADAR

Results for job radar-I20140416-090225-0426-73908772-oy

Summary | Submission Details

View/download raw output file

```
--------------------------------------------------------------------------------
No. of Repeats|Total Score|Length  |Diagonal| BW-From|   BW-To|   Level
             2|      59.07|      15|      31|      45|      59|       1
--------------------------------------------------------------------------------
   45-   59 (28.50/17.07)       PLHFAAGYNRVSVVE
   78-   92 (30.57/18.72)       PLHNACSYGHYEVAE
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
No. of Repeats|Total Score|Length  |Diagonal| BW-From|   BW-To|   Level
             3|      80.22|      15|      65|      61|      75|       2
--------------------------------------------------------------------------------
   61-   75 (28.16/18.06)       LLQHGADVHAKDKGG
   94-  108 (24.51/14.88)       LVKHGAVVNVADLWK
  127-  141 (27.55/17.53)       LLQHGADPTKKNRDG
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
No. of Repeats|Total Score|Length  |Diagonal| BW-From|   BW-To|   Level
             2|      51.16|      16|      30|     110|     126|       3
--------------------------------------------------------------------------------
  110-  126 (23.51/21.63)       TPLhEAAAKGKYEICKL
  143-  158 (27.65/19.08)       TPL.DLVKDGDTDIQDL
--------------------------------------------------------------------------------
```

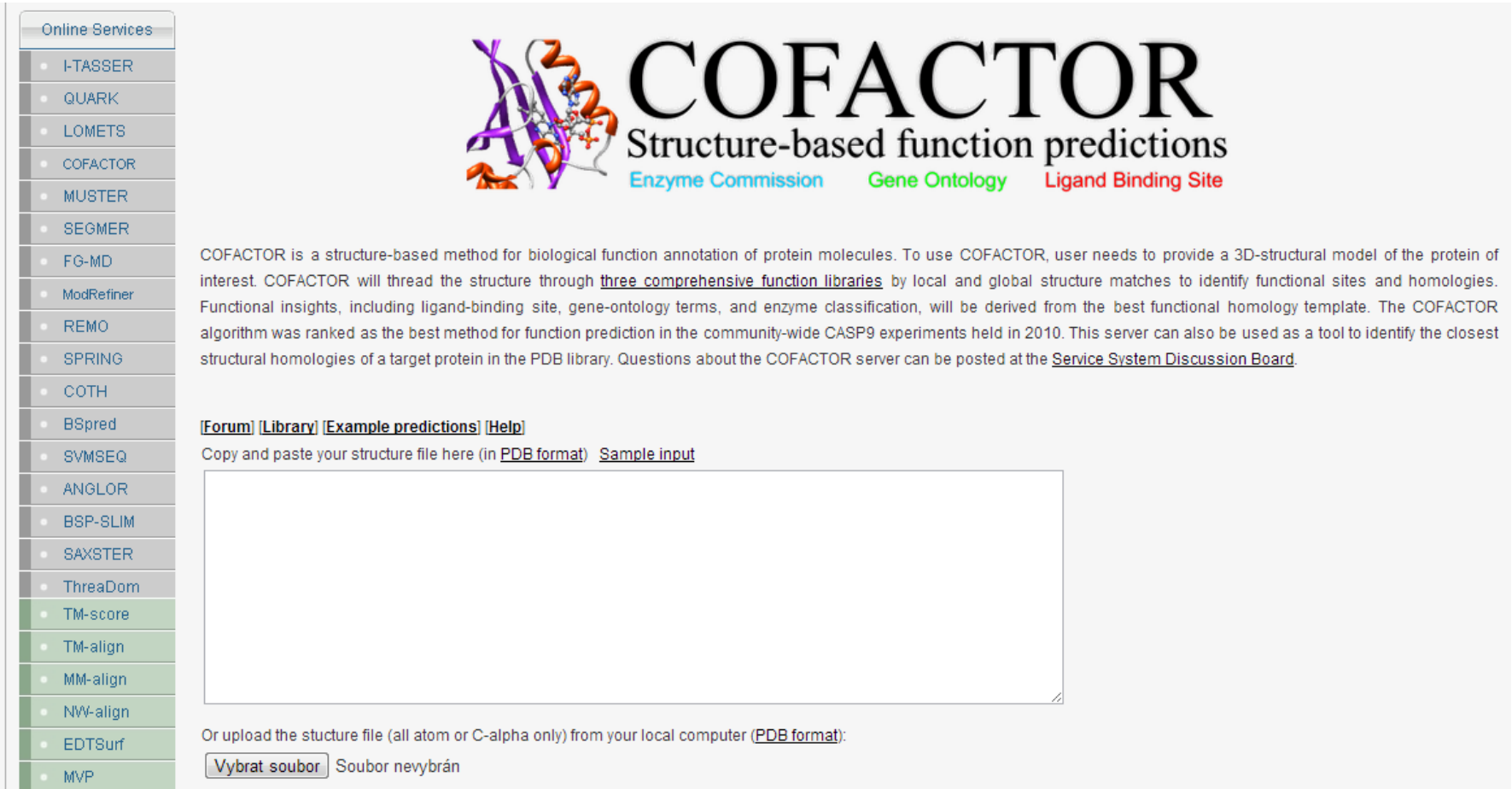
Predikce funkce na základě 3D struktury

- Strukturně podobné enzymy
- Analýza komplexů homologů s ligandy
- Komplexy protein-protein, protein-DNA/RNA
- Přítomnost/absence klíčových residuí
- Provázanost s dalšími databázemi

COFACTOR

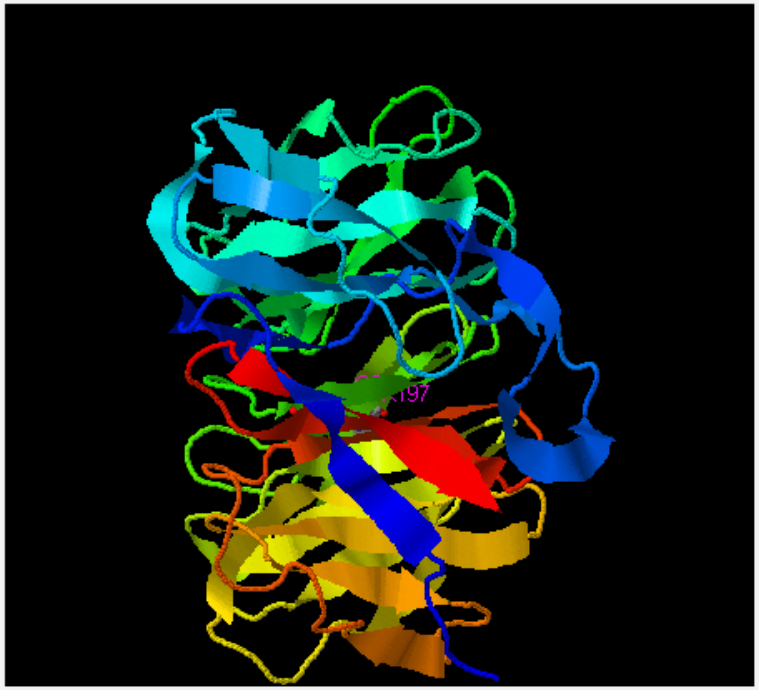
Součást serveru Zhanglab (podobně jako I-TASSER)

http://zhanglab.ccmb.med.umich.edu/COFACTOR/



Homologie se známými enzymy – proteiny se stejnou funkcí mívají stejný fold

Predicted EC Numbers



Spin On/Off

Top 5 enzyme homologs in PDB

Rank	Cscore[EC]	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov.	EC Number	Predicted Active Site Residues
1	0.187	2ebsB	0.694	3.84	0.095	0.878	3.2.1.150	197,206
2	0.183	1k3iA	0.716	3.90	0.059	0.916	1.1.3.9	NA
3	0.178	2madH	0.700	3.35	0.060	0.842	1.4.99.3	226
4	0.177	1fwxA	0.714	3.46	0.072	0.866	1.7.99.6	35,275
5	0.176	2gc7A	0.700	3.43	0.046	0.848	1.4.99.3	200

Click on the radio buttons to visualize predicted active site residues.

(a) Cscore[EC] is the confidence score for the Enzyme Classification (EC) number prediction. Cscore[EC] values range higher score indicates a more reliable EC number prediction.

(b) TM-score is a measure of global structural similarity between query and template protein.

(c) RMSD[a] is the RMSD between residues that are structurally aligned by TM-align.

(d) IDEN[a] is the percentage sequence identity in the structurally aligned region.

(e) Cov. represents the coverage of global structural alignment and is equal to the number of structurally aligned residu query protein.

Gene ontology (GO) terms – popisují biologické zařazení proteinu (funkci, biologický proces, buněčnou lokalizaci)

Predicted GO terms

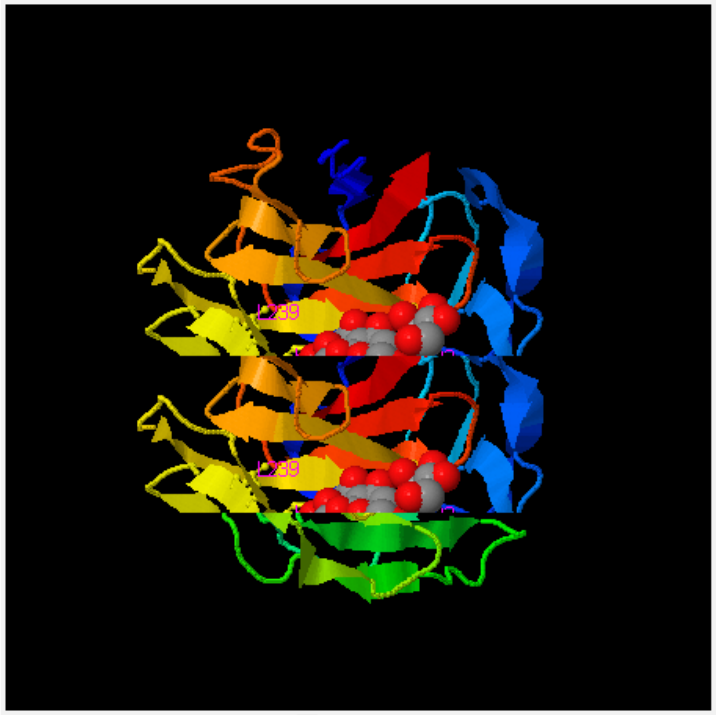
Rank	Cscore	GO TMscore	RMSD[a]	IDEN[a]	Cov.	PDB Hit	Associated GO Terms
1	0.28	0.7701	3.55	0.06	0.95	3iieA	GO:0009897 GO:0044419 GO:0007160 GO:0007596 GO:0008305 GO:0001846 GO:0004872 GO:0005887 GO:0016020 GO:0016021 GO:0043277 GO:0001568 GO:0005515 GO:0045715 GO:0070371 GO:0050748 GO:0052066 GO:0005886 GO:0050900 GO:0043066 GO:0045785 GO:0007155 GO:0001525 GO:0010745 GO:0050764 GO:2000425 GO:0097024 GO:0008284 GO:0010888 GO:0050431 GO:0009986 GO:0031994 GO:0035635 GO:0007411 GO:0032369 GO:0046718 GO:0007229
2	0.28	0.7176	3.34	0.09	0.87	2z2oC	GO:0046872 GO:0016829 GO:0046677 GO:0000287 GO:0016835 GO:0017001
3	0.24	0.6233	4.09	0.05	0.80	1iubA	GO:0005529
4	0.23	0.7251	3.32	0.09	0.88	2qc5A	GO:0000287 GO:0016835 GO:0017001 GO:0046677
5	0.23	0.7400	3.33	0.07	0.89	2h9lA	GO:0005622 GO:0071339 GO:0005634 GO:0006351 GO:0016568 GO:0035097 GO:0051568 GO:0001501 GO:0042800 GO:0005671 GO:0006355 GO:0043966 GO:0005515 GO:0048188 GO:0035064 GO:0034968
6	0.22	0.7438	3.41	0.06	0.90	3iz6A	GO:0005515
7	0.20	0.6509	3.78	0.05	0.82	3k6sE	GO:0005515 GO:0016021 GO:0008305 GO:0050900 GO:0004872 GO:0005886 GO:0007155 GO:0009887 GO:0007596 GO:0007229 GO:0016020
8	0.20	0.7292	3.49	0.06	0.90	3ow8C	GO:0005515
9	0.19	0.7343	3.21	0.07	0.88	2hesX	GO:0005737 GO:0000055 GO:0003674 GO:0005829 GO:0005515 GO:0005634 GO:0016226 GO:0002098
10	0.19	0.7356	3.37	0.06	0.89	3izbA	GO:0001965 GO:0005737 GO:0005515 GO:0010255 GO:0005092 GO:0017148 GO:0007186 GO:0001403 GO:0022627 GO:0004871

Consensus Prediction of Gene Ontology terms

Molecular Function		Biological Process		Cellular Location	
GO term	GO-Score	GO term	GO-Score	GO term	GO-Score
GO:0019955	0.56	GO:0052370	0.56	GO:0043235	0.56
GO:0005520	0.56	GO:0051051	0.56	GO:0005887	0.51
GO:0005080	0.56	GO:0000165	0.56	GO:0016585	0.45
GO:0042393	0.45	GO:0010871	0.56	GO:0034708	0.45
GO:0018024	0.45	GO:0045714	0.56	GO:0000123	0.45
GO:0000287	0.45	GO:0019059	0.56		
GO:0016835	0.45	GO:0048646	0.56		
		GO:0050746	0.56		
		GO:0002376	0.56		

Vazebná místa – predikce umístění vazbených míst na základě struktury homologních komplexů

Predicted Binding Site



Spin On/Off

Template proteins with similar binding site:

	Rank	$Cscore^{LB}$	PDB Hit	TM-score	$RMSD^a$	$IDEN^a$	Cov.	BS-score	Lig. Name	Download Complex	Predicted binding site residues in the model
○	1	0.04	2xl3B	0.739	3.31	0.070	0.890	0.79	PEPTIDE	Download	179,223,229,252,290
◉	2	0.04	3k71E	0.742	3.76	0.070	0.937	0.74	Mul.Part	Download	223,239,241,250
○	3	0.03	2z2pB	0.719	3.32	0.089	0.872	0.82	MG	Download	184,186,229
○	4	0.03	2z2pA	0.719	3.32	0.092	0.872	0.72	PEPTIDE	Download	179,223,225,268,271
○	5	0.02	3eg6A	0.734	3.36	0.074	0.890	0.61	PEPTIDE	Download	18,20,82,83,84,100,128,145,188,309
○	6	0.02	1omw3	0.751	3.09	0.060	0.890	0.64	PEPTIDE	Download	17,19,82,84,100,126,173,306

Click on the radio buttons to visualize predicted binding site and residues.

(a) $Cscore^{LB}$ is the confidence score of predicted binding site. $Cscore^{LB}$ values range in between [0-1]; where a higher score indicates a more reliable ligand-binding site prediction.

(b) BS-score is a measure of local similarity (sequence & structure) between template binding site and predicted binding site in the query structure. Based on large scale benchmarking analysis, we have observed that a BS-score >1 reflects a significant local match between the predicted and template binding site.

(c) TM-score is a measure of global structural similarity between query and template protein.

(d) $RMSD^a$ the RMSD between residues that are structurally aligned by TM-align.

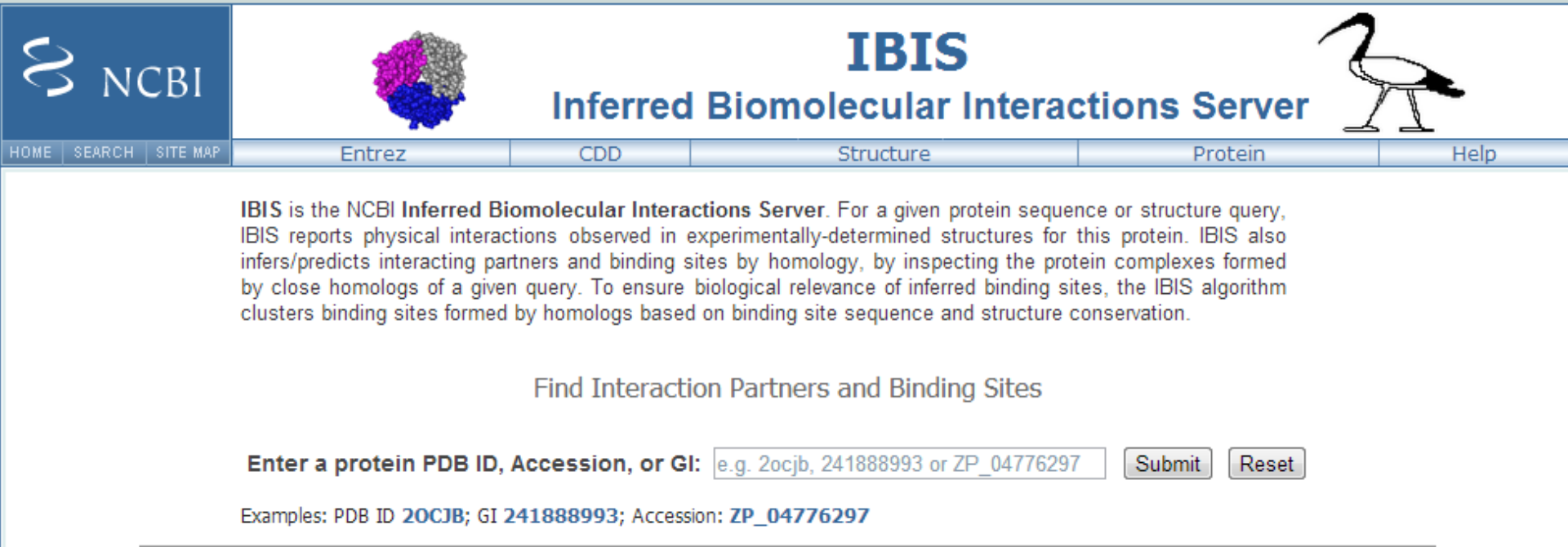
(e) $IDEN^a$ is the percentage sequence identity in the structurally aligned region.

(f) Cov. represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

IBIS

http://www.ncbi.nlm.nih.gov/Structure/ibis/ibis.cgi

- Analýza 3D struktur – interakce protein-protein a protein-ligand
- Vyhledává i podobné interakce u homologních proteinů v databázi



Úloha

- Která residua delta-podjednotky DNA polymerasy ze *Saccharomyces cerevisiae* jsou klíčová pro vazbu DNA? Využijte serveru **IBIS** (http://www.ncbi.nlm.nih.gov/Structure/ibis/ibis.cgi).

IBIS

Inferred Biomolecular Interactions Server

IBIS is the NCBI **Inferred Biomolecular Interactions Server**. For a given protein sequence or structure query, IBIS reports physical interactions observed in experimentally-determined structures for this protein. IBIS also infers/predicts interacting partners and binding sites by homology, by inspecting the protein complexes formed by close homologs of a given query. To ensure biological relevance of inferred binding sites, the IBIS algorithm clusters binding sites formed by homologs based on binding site sequence and structure conservation.

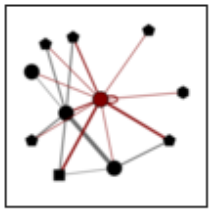
Find Interaction Partners and Binding Sites

Enter a protein PDB ID, Accession, or GI: 3IAYA Submit Reset

Examples: PDB ID **2OCJB**; GI **241888993**; Accession: **ZP_04776297**

Putative interaction sites and partners for query Dna Polymerase Delta Catalytic Subunit (3IAY_A)

Click on a category to view details



All interactions for query sequence

Protein-protein: 6 sites
Interactions with POLBc and other proteins

Protein-chemical: 1 site
Interactions with dCTP

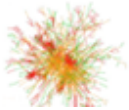
Protein-DNA/RNA: 3 sites
Interactions with nucleotide

Protein-peptide: none

Protein-ion: 8 sites
Interactions with MANGANESE and other ions

Current IBIS Statistics

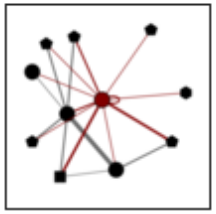
Download IBIS Data



Download Human Interactome

Query 3IAY_A

Dna Polymerase Delta Catalytic Subunit



All interactions for query sequence

Download data

EXCEL ◉ XML ○

Search 3IAY A interactions

Similarity to query

☐ Sequence Identity: %Id list

☐ Structure RMSD: RMSD list

Interaction partner type

☐ PDB Code: PDB code list

☐ Taxonomy: Taxonomies list

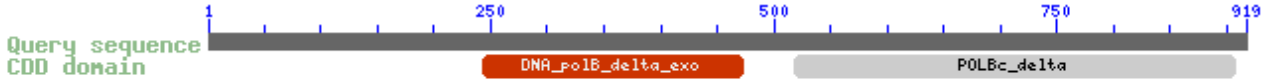
Reset

Biounit Validation: ☑

Protein-Protein (6) | **Protein-Chemical (1)** | **Protein-DNA/RNA (3)** | **Protein-Ion (8)**

Dna Polymerase Delta Catalytic Subunit (range: 243-472)

Domain interactions are listed for each domain of the query chain.
Click on a grey balloon to see the interactions for another domain.



Interaction partner — IBIS annotated binding sites

POLBc

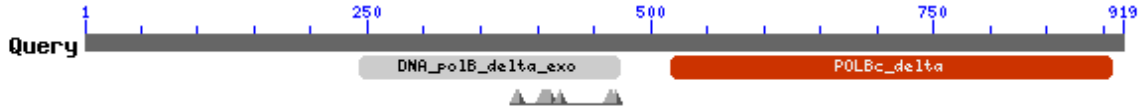
2 singleton clusters

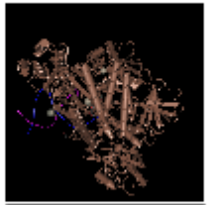
POLBc_delta

POLBc_B3

List of protein interaction partners and binding sites. Similar binding sites of homologs of the query are grouped into clusters. To view the cluster members click on the plus sign. "o" denoted observed interactios. Note: singletons might not provide enough evidence for biological relevance of binding site.

Interaction Partner	Ranking Score	Number of Cluster Members	Average %Identity to Query	Number of Binding Site Residues	Taxonomic Diversity
⊞ POLBc	n/a	2	26	7	root
⊟ o POLBc_delta	singleton	1	100	13	Saccharomyces cerevisiae





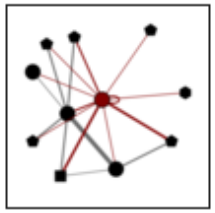
View Binding Sites

Download Cn3D

	Homologous complex	Homolog	Interaction partner	%Identity to query	Binding site													
					382	383	405	406	407	408	409	411	420	464	465	466	470	
Query		-	-	-	T	R	E	Y	K	L	R	Y	F	K	L	M	N	
☑	3IAY	A	A	100	T	R	E	Y	K	L	R	Y	F	K	L	M	N	

Query 3IAY_A

Dna Polymerase Delta Catalytic Subunit



All interactions for query sequence

Download data

EXCEL ◉ XML ○

Search 3IAY A interactions

Similarity to query

☐ Sequence Identity: %Id list

☐ Structure RMSD: RMSD list

Interaction partner type

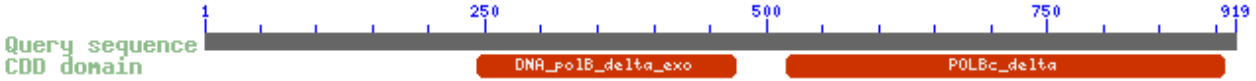
☐ PDB Code: PDB code list

☐ Taxonomy: Taxonomies list

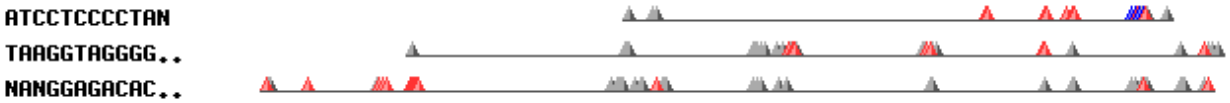
Reset

Protein-Protein (6) | Protein-Chemical (1) | Protein-DNA/RNA (3) | Protein-Ion (8)

Dna Polymerase Delta Catalytic Subunit

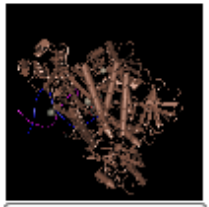


Interaction partner IBIS annotated binding sites



List of DNA/RNA interaction partners and binding sites. Similar binding sites of homologs of the query are grouped into clusters. To view the cluster members click on the plus sign. "o" denoted observed interactios.

	Interaction Partner	Ranking Score	Number of Cluster Members	Average %Identity to Query	Number of Binding Site Residues	Taxonomic Diversity
⊞ o	ATCCTCCCCTAN	2.5	3	50	27	cellular organisms (& synthetic construct)
⊟ o	TAAGGTAGGGGAGGAT	2.3	2	63	41	Saccharomyces cerevisiae (& synthetic construct)



View Binding Sites

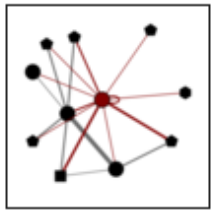
Download Cn3D

Binding site

99	509	514	520	521	522	523	524	526	528	639	640	642	643	646	647	649	651	652	746	747	748	749	773	869	890	894	898	902
F	I	S	Q	Y	E	G	A	V	E	N	S	Y	G	G	A	V	K	L	N	K	K	R	R	L	Y	N	N	S
F	I	S	Q	Y	E	G	A	V	-	N	S	Y	G	G	A	V	K	L	N	K	K	R	R	L	Y	N	N	-
-	-	-	S	Y	T	G	G	V	E	-	-	-	G	G	Y	-	-	-	T	K	K	R	-	I	Y	-	P	R

Query 3IAY_A

Dna Polymerase Delta Catalytic Subunit



All interactions for query sequence

Download data

EXCEL XML

Search 3IAY A interactions

Similarity to query

Sequence Identity: %Id list

Structure RMSD: RMSD list

Interaction partner type

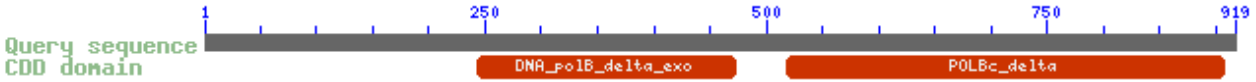
PDB Code: PDB code list

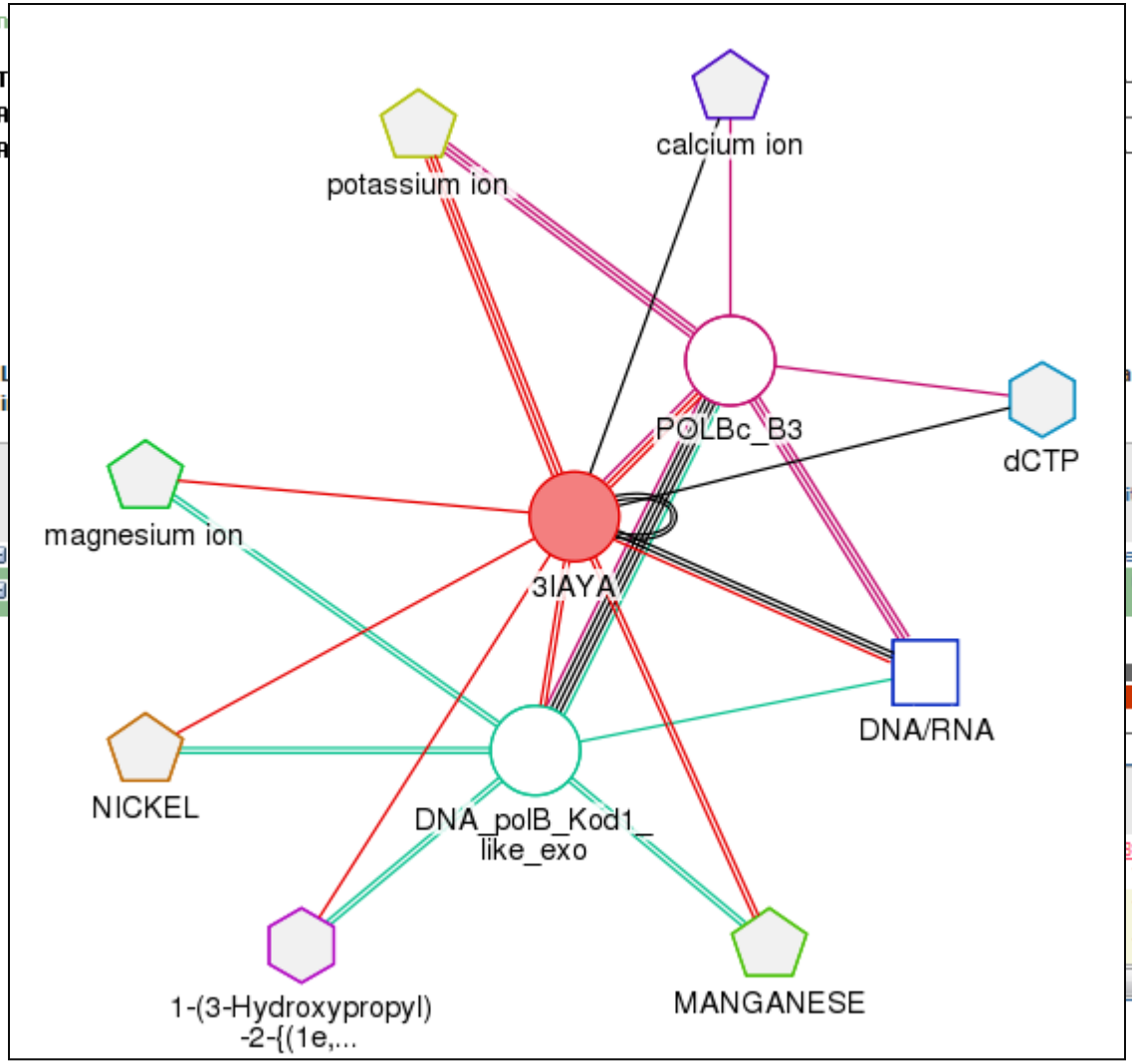
Taxonomy: Taxonomies list

Reset

Protein-Protein (6) | Protein-Chemical (1) | Protein-DNA/RNA (3) | Protein-Ion (8)

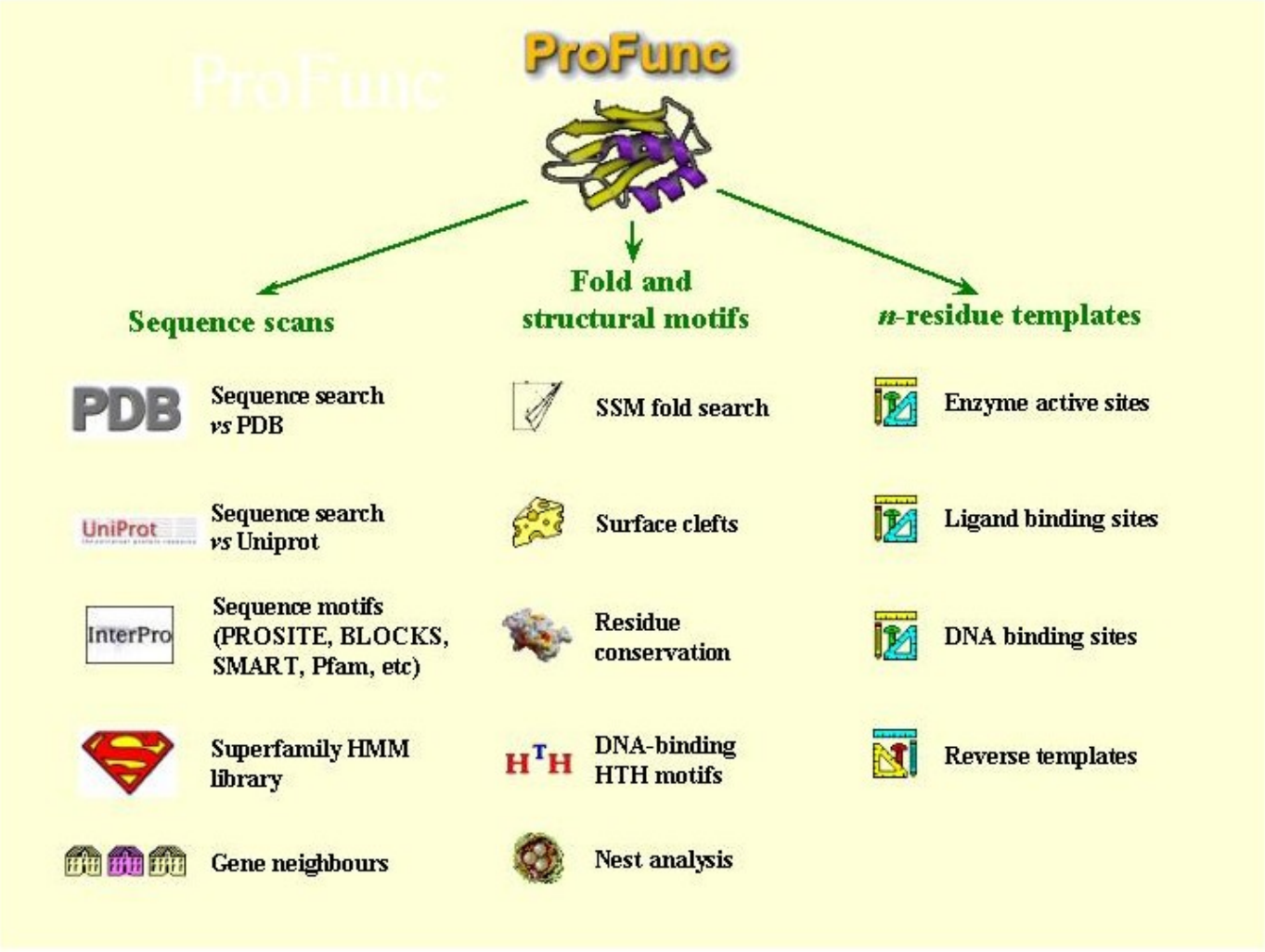
Dna Polymerase Delta Catalytic Subunit



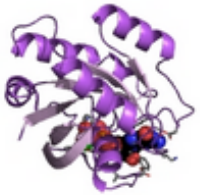


ProFunc

http://www.ebi.ac.uk/thornton-srv/databases/ProFunc/



- Documentation
- Tutorial
- Contact us



Example: 5p21

ProFunc - prediction of protein function from 3D structure

The aim of the ProFunc server is to help identify the likely biochemical function of a protein from its three-dimensional structure. It uses a series of methods, including fold matching, residue conservation, surface cleft analysis, and functional 3D templates, to identify both the protein's likely active site and possible homologues in the PDB.

ProFunc

From this page you can submit your own structure, analyse an existing PDB entry, or retrieve the results of a previously submitted run.

Choose option A, B or C:

A. Upload PDB-format file: Vybrat soubor Soubor nevybrán Upload

B. Use existing PDB file *(4 chars)*: Get Example: "5p21"

Runs ProFunc on an existing PDB entry or, if already done, takes you directly to the results page.

C. Go to previous analysis **Id no.:** **Security code:**

Go

Some of the methods take minutes to run; others take hours. You will be notified by e-mail when the entire process is complete, but can check on preliminary results as they become available.

The files are usually stored for about 6 months before being deleted. However, they are stored on a partition that is not backed up; so, in principle, they could disappear at any time.

Notes

- Please try to limit the number of structures submitted to about 6 per hour to avoid overloading the server. To arrange a large batch run, please contact us.
- If your structure contains any non-standard amino acids (*eg* selenomethionines, phosphotyrosines, *etc*) it

Related databases

PDBsum

Structural analyses of all PDB entries

Tempura

Customized 3D template generation from submitted structure and search *vs* PDB.

ProFunc

ProFunc results for 3kaz

Jmol RasMol

Header details

Structure: Putative uncharacterized protein at2g26040.
Source: Arabidopsis thaliana. Mouse-ear cress,thale-cress. Organism_taxid: 3702. Gene: at2g26040. Escherichia coli. Expression_system_taxid: 562. Pet24a
Date: 19 Oct 09
Author(s): X.E.Zhou,K.Melcher,L.-M.Ng,F.-F.Soon,Y.Xu,K.M.Suino-Powell, A.Kovach,J.Li,H.E.Xu
Resolution: 1.000Å **R-factor:** 0.214 **R-free:** 0.250
Chain(s): [A(175a.a.)&B(175a.a.)&C(176a.a.)]
Waters: 220

Summary of predicted function

Protein name terms

abscisic (20.28) receptor (17.95) acid (16.96) abscisic acid (15.03) uncharacterized (13.85) abscisic acid receptor (12.51) allergen (5.33) phosphatase (4.60)

Gene Ontology (GO) terms

Cellular component: ()
Biological process: ()
Biochemical function: ()

The protein names and GO terms above are the most common terms found in the hits obtained from the analyses below. Each term's score (based on the number of times it occurs independently) is given in brackets. Click on the plus icons for a complete breakdown of which programs, and further, which hits the terms came from.

ProFunc results

Úloha

- Seznamte se se serverem **ProFunc** (EBI, http://www.ebi.ac.uk/thornton-srv/databases/ProFunc/) prostřednictvím modelového příkladu předpokládané glutaminasy z *Bacillus subtilis*. Je možná jiná funkce tohoto enzymu? Vycházejte z predikce na základě 3D struktury.

ProFunc results for 1mki

ProFunc

Jmol RasMol

Header details

Structure: Probable glutaminase ybgj. Ec: 3.5.1.2. . Mutation: yes
Source: Bacillus subtilis. Organism_taxid: 1423. Gene: ybgj. Escherichia coli. Expression_system_taxid: 469008. Bl21(de3)
Date: 29 Aug 02
Author(s): Y.Kim,I.Dementieva,E.Vinokour,A.Joachimiak,Midwest center for structural genomics (mcsg)
Resolution: 2.000Å **R-factor:** 0.212 **R-free:** 0.245
Chain(s): [A&B(312a.a.)]
Waters: 330

View results so far ...

Sequence motifs

InterPro

InterPro scan for sequence motifs. Chains A, B

8 motifs matched in scan against PROSITE, PRINTS, PFam-A, TIGRFAM, PROFILES and PRODOM motifs

	Type	Motif	Name
1.	Gene3D	G3DSA:3.40.710.20	no description
2.	Gene3D	G3DSA:1.10.1500.10	no description
3.	???	MF_00313	Glutaminase
4.	HMMPfam	PF04960	Glutaminase
5.	HMMPanther	PTHR12544:SF3	GLUTAMINASE BACTERIA

Doplňková literatura a další zdroje

- http://www.wikipedia.org ☺
- http://www.proteinstructures.com/
- http://cssb.biology.gatech.edu/resources#services
- http://www.ebi.ac.uk/services/structures
- Odborné články ve studijních materiálech

Co říci závěrem?

- SW nástoje usnadňují analýzu struktur

Predikce není dokonalá.

Nenahradí experiment, ale může významně usnadnit práci

- **Kontroluj, porovnávej, ověřuj,…**