Blok 3

Funkce proteinů

C3211 Aplikovaná bioinformatika

Přednášející: Josef Houser

Funkce proteinů

- **Enzymy** – 6 hlavních tříd
- **Strukturní** proteiny – keratin, kolagen
- **Transportní** proteiny – albumin, hemoglobin
- **Obranné** proteiny – protilátky
- **Regulátory** a **receptory** – hormony, transkripční faktory, rhodopsin
- …

Klasifikace enzymů

Dle IUBMB: http://www.chem.qmul.ac.uk/iubmb/enzyme/

	Třída	Charakteristika	Počet Podtříd
1	Oxidoreduktasy	Katalyzují různé redoxní reakce – přenos vodíku, kyslíku, elektronů (obvykle s využitím koenzymů, např. NADH, NADPH, $FADH_2$ nebo hemu)	23
2	Transferasy	Katalyzují přenos skupin: amino-, metyl-, acyl-, glykosyl-, fosforyl-	10
3	Hydrolasy	Katalyzují hydrolytické štěpení vazeb mezi atomem uhlíku a jinými atomy (spotřebování molekuly H_2O)	13
4	Lyasy	Katalyzují adiční reakci na dvojné vazbě nebo eliminační reakci mezi 2 atomy uhlíku za vzniku dvojné vazby	7
5	Isomerasy	Katalyzují racemizaci optických izomerů nebo vytvoření polohových izomerů	6
6	Ligasy	Katalyzují tvorbu vazeb mezi uhlíkem a jinými atomy spojenou se štěpením ATP	6

Transportní proteiny

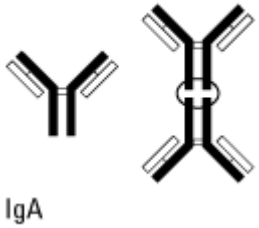
Dle TCDB (transporter classification database):

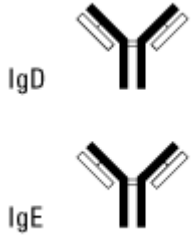
http://www.chem.qmul.ac.uk/iubmb/mtp/

	Třída
1	Póry a kanály
2	Přenašeče řízené elektrochemickým potenciálem
3	Přenašeče řízené chemickou reakcí
4	Skupinové přenašeče
5	Transmembránové elektronové přenašeče
6	Nepřiřazeno
7	Nepřiřazeno
8	Accessory factors involved in transport
9	Nedostatečně charakterizované transportní systémy

Protilátky

	Subtypů (člověk)	Charakteristika
IgA	2	Monomer/dimer, přítomna ve slinách, slzách (15%)
IgD	1	Monomer, funkce neznámá (0,2%)
IgE	1	Monomer, obrana proti parazitům, význam pro alergické reakce (0,002%)
IgG	4	Monomer, hlavní lidská protilátka v sekundární imunitní odpovědi (75%)
IgM	1	Pentamer, hlavní protilátka v primární imunitní odpovědi (10%)





Určení funkce proteinu

- **Experimentální**
  - Izolace proteinu s konkrétní funkcí
  - Stanovení funkce u konkrétního proteinu

- **Predikce** – na základě podobnosti
  - Lokalizace
  - Multiple sequence alignment (BLAST, Pfam)
  - Struktura molekuly (ProFunc, Catalytic site atlas)
  - Textové hledání v publikacích (STRING)

Určení funkce proteinu

- **Nezavrhujte jednoduchá řešení**

Jakou funkci má tento protein?

Predikce funkce proteinu

Nutno znát **sekvenci**

- Databáze
- Sekvenace

Lépe znát **strukturu** (2D, 3D)

- Databáze
- Určení 2D struktury viz. předchozí blok
- Určení 3D struktury viz. následující blok

Databáze strukturních a funkčních motivů

- Neanotované, nerevidované – „slepé“ přebírání dat
- Anotované, revidované – probíhá kontrola vkládaných dat

- Obsahují různé informace – sekvenční, strukturní, odkazy na experimentální data,…
- Slouží jako zdroj informací pro nadstavbové programy.

Databáze strukturních a funkčních motivů

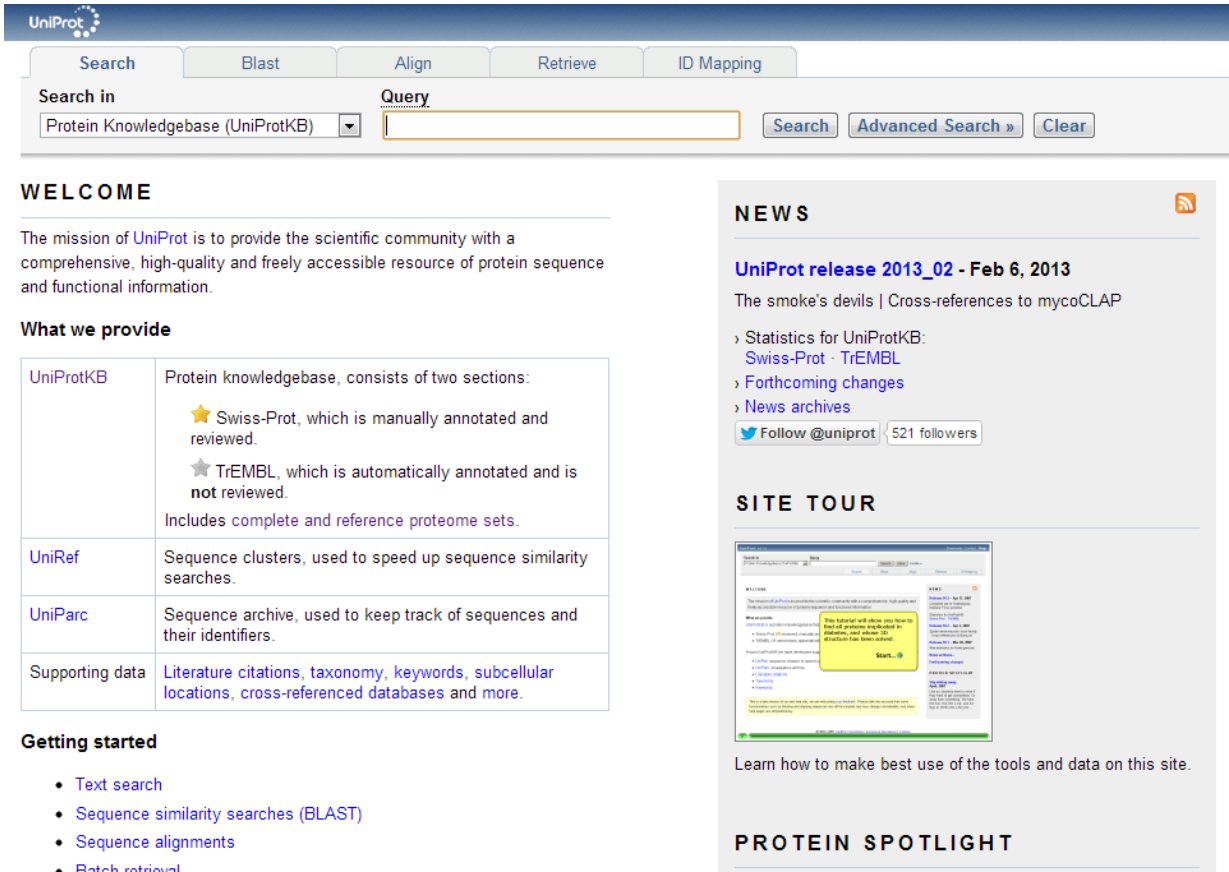
Často navzájem provázané. Např.:

- **UniProtKB** – kombinovaná proteinová databáze, vč. biologických dat
- **Pfam** – odvozená z UniProtKB
- **KEGG** – složená databáze obsahující systémové, genomické a chemické informace
- **CDD** – proteinové domény a další data
- …

UniProt http://www.uniprot.org/

Kombinace několika databází

Vyhledávání pomocí klíčových slov i pomocí sekvence



Úloha

- Vyhledejte homologní proteiny k následující sekvenci pomocí Blastu na serveru **UniProt**. Pokuste se na základě výsledku určit funkci tohoto proteinu.

SHLSQPWPITCFADRPTPRRSSPDASGQTMHSVFVVHVPYPVVFLKPAH
LTPQWYRHPIPVNPVVRQPHLPVLYPAPNAGHTPAHSRQGDAALQPLF
SVPQTVNPTGPVIHGDVAKQKPDTGQSWALNPYCTENWRRILRISRNS
HGQRMPLTTLLQKTSGRNATLITKNSDQNTTTSIVSESSMTISACCHSAIL
RNN

Graphical overview

Color code for identity 0-100% =

Accession	Entry name	0Query hit199	0Match hit (sqrt scale)3105	Name (Organism)
Query 2013031826IMB16GAX				
Q8CLU5	Q8CLU5_YERPE			Uncharacterized protein (Yersinia pestis)
J2E0Y8	J2E0Y8_KLEPN			Uncharacterized protein (Klebsiella pneumoniae subsp. pneumoni...)
I1BJU5	I1BJU5_RHIO9			Uncharacterized protein (Rhizopus delemar (strain RA 99-880 / ...)
G6HG35	G6HG35_9ACTO			Putative uncharacterized protein (Frankia sp. CN3)
F9VC24	F9VC24_LACGL			Putative uncharacterized protein (Lactococcus garvieae (strain Lg2))
F9V737	F9V737_LACGT			Putative uncharacterized protein (Lactococcus garvieae (strain ATCC 491...)
B4NC70	B4NC70_DROWI			GK25804 (Drosophila willistoni)
I1RMT3	I1RMT3_GIBZE			Uncharacterized protein (Gibberella zeae (strain PH-1 / ATCC M...)
E7FCT7	E7FCT7_DANRE			Uncharacterized protein (Danio rerio)
Q9RS36	Q9RS36_DEIRA			Cell wall glycyl-glycine endopeptidas... (Deinococcus radiodurans (strain ATCC ...)
G7L8Y9	G7L8Y9_MEDTR			Extensin-like protein (Medicago truncatula)
A9XEB8	A9XEB8_SACOF			27kD gamma canein (Saccharum officinarum)
G2R972	G2R972_THITE			Putative uncharacterized protein (Thielavia terrestris (strain ATCC 380...)
I8SZ67	I8SZ67_9LACT			Uncharacterized protein (Lactococcus garvieae IPLA 31405)
K0SQJ7	K0SQJ7_THAOC			Uncharacterized protein (Thalassiosira oceanica)
G7JXY3	G7JXY3_MEDTR			Putative uncharacterized protein (Medicago truncatula)
A7S8E1	A7S8E1_NEMVE			Predicted protein (Nematostella vectensis)
Q2GXL2	Q2GXL2_CHAGB			Putative uncharacterized protein (Chaetomium globosum (strain ATCC 6205...)
Q3ZB05	Q3ZB05_MOUSE			Acr protein (Mus musculus)
Q3ZB06	Q3ZB06_MOUSE			Acr protein (Mus musculus)
P23578	ACRO_MOUSE			Acrosin (Mus musculus)
Q6PAA4	Q6PAA4_XENLA			MGC68472 protein (Xenopus laevis)
Q3KQB9	Q3KQB9_XENLA			MGC68472 protein (Xenopus laevis)
A7XYK4	A7XYK4_TETNG			Jxc1-B (Tetraodon nigroviridis)
H3C1M2	H3C1M2_TETNG			Uncharacterized protein (Tetraodon nigroviridis)
H3D438	H3D438_TETNG			Uncharacterized protein (Tetraodon nigroviridis)

Detailed BLAST results Customize

CDD (conserved domain database)

http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml

- Doména – část proteinu s vlastní aktivitou nebo strukturní funkcí (více v bloku o 3D a 4D struktuře proteinů)
- Domény často obsahují sekvenční motiv, který můžeme nalézt u více proteinů s určitou funkcí – **konzervované domény**

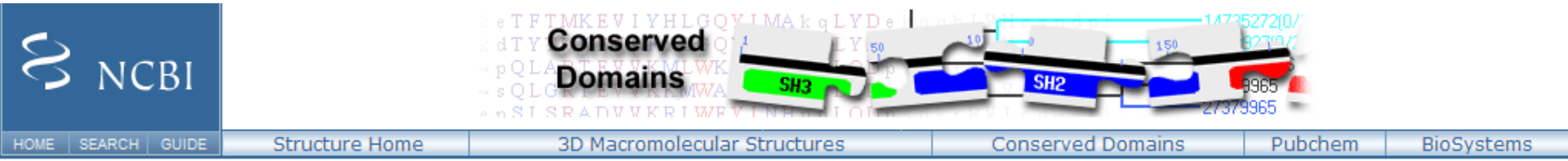
CDD zdroje

Abbreviation	Database Name	Description
SMART	**S**imple **M**odular**A**rchitecture **R**esearch**T**ool	SMART is a web tool for the identification and annotation of protein domains, and provides a platform for the comparative study of complex domain architectures in genes and proteins. SMART is maintained by Chris Ponting, Peer Bork and colleagues, mainly at the EMBL Heidelberg. CDD contains a large fraction of the SMART collection.
Pfam	**P**rotein **fam**ilies	Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families. Pfam is maintained by Alex Bateman and colleagues, mainly at the Wellcome Trust Sanger Institute. CDD contains a large fraction of the Pfam collection.
COGs	**C**lusters of **O**rthologous**G**roups of proteins	COGs is an NCBI-curated protein classification resource. Sequence alignments corresponding to COGs are created automatically from constituent sequences and have not been validated manually when imported into CDD.
TIGRFAM	**T**he **I**nstitute for**G**enomic **R**esearch's database of **p**rotein**fam**ilies	TIGRFAM, a research project of the J. Craig Venter Institute, is a collection of manually curated protein families from The Institute for Genomic Research and consists of hidden Markov models (HMMs), multiple sequence alignments, Gene Ontology (GO) terminology, cross-references to related models in TIGRFAM and other databases, and pointers to literature.
PRK	**PR**otein **K**(c)lusters	Protein Clusters is an NCBI collection of related protein sequences (clusters) consisting of Reference Sequence proteins encoded by complete prokaryotic and chloroplast plasmids and genomes. It includes both curated and non-curated (automatically generated) clusters.

CD search - NCBI

http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

Možnost hledání záznamů v CDD dle klíčového slova nebo identifikace konzervované domény v zadané sekvenci (CDS)



Search for Conserved Domains within a protein or coding nucleotide sequence

NEW! Use Batch CD-search to submit multiple query proteins at once!

Enter **protein** or **nucleotide** query as accession, gi, or sequence in FASTA format

Submit Reset

OPTIONS

Search against database: CDD v3.08 - 43334 PSSMs

Expect Value threshold: 0.01

Apply low-complexity filter

Force live search

Maximum number of hits: 500

Result mode Concise Full

Retrieve previous CD-search result

Request ID: Retrieve

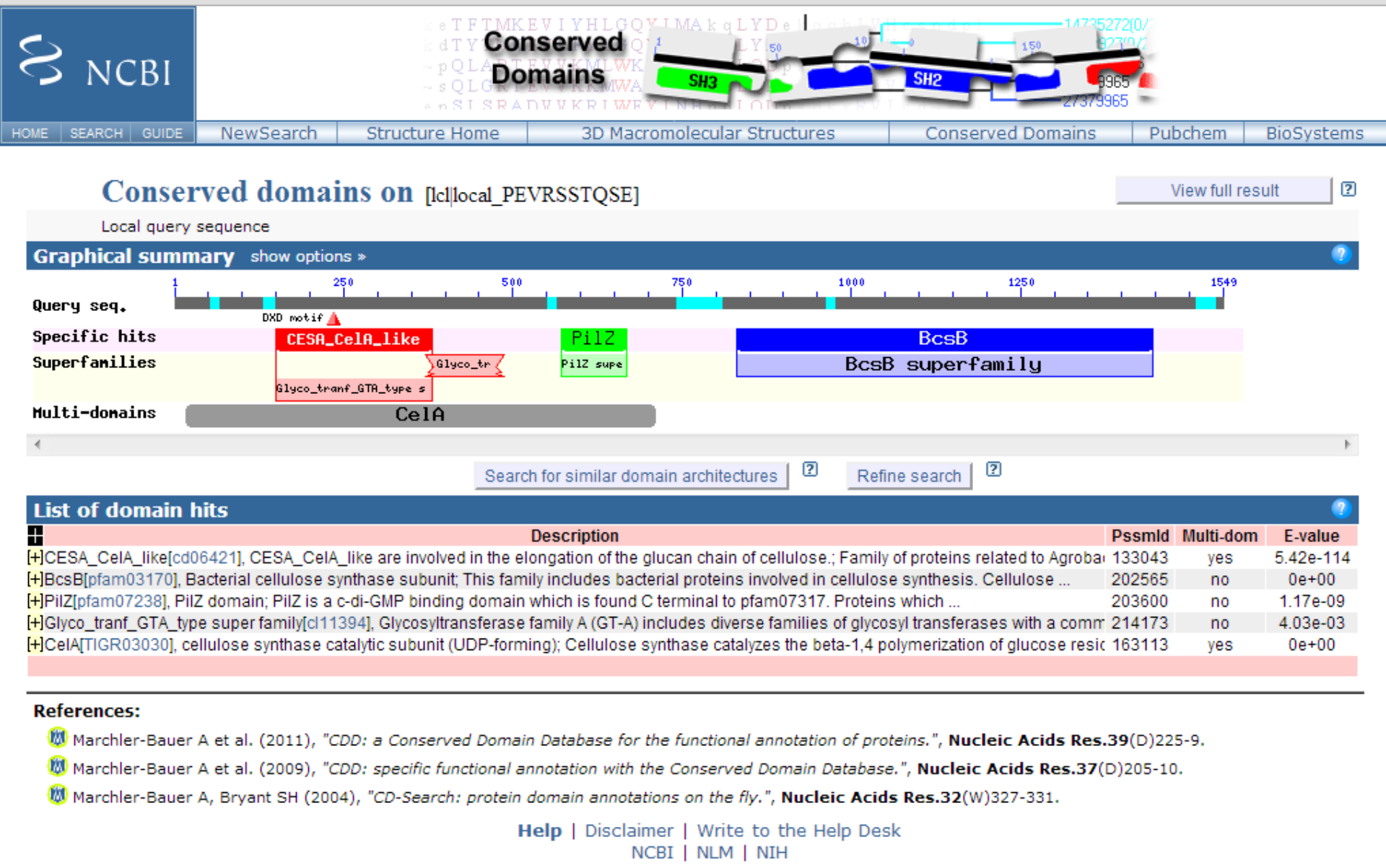
Úloha

- Vyhledejte pomocí nástroje CD search (NCBI) konzervované domény následujícího proteinu.

Sekvence:

PEVRSSTQSESGMSQWMGKILSIRGAGLIIGVFGLCALIAATSVTLPPEQQLIVAFVCVVIFFIVGHKPSRRSQIFLEVLSGLVSLRYLTWRLT
ETLSFDTWLQGLLGTMLLVAELYALMMLFLSYFQTIAPLHRAPLPLPPNPDEWPTVDIFVPTYNEELSIVRLTVLGSLGIDWPPEKVRVHIL
DDGRRPEFAAFAAECGANYIARPTNEHAKAGNLNYAIGHTDGDYILIFDCDHVPTRAFLQLTMGWMVEDPKIALMQTPHHFYSPDPF
QRNLSAGYRTPPEGNLFYGVVQDGNDFWDATFFCGSCAILRRTAIEQIGGFATQTVTEDAHTALKMQRLGWSTAYLRIPLAGGLATERLI
LHIGQRVRWARGMLQIFRIDNPLFGRGLSWGQRLCYLSAMTSFLFAVPRVIFLSSPLAFLFFGQNIIAASPLALLAYAIPHMFHAVGTASKI
NKGWRYSFWSEVYETTMALFLVRVTIVTLLSPSRGKFNVTDKGGLLEKGYFDLGAVYPNIILGLIMFGGLARGVYELSFGHLDQIAERAYL
LNSAWAMLSLIIILAAIAVGRETQQKRNSHRIPATIPVEVANADGSIIVTGVTEDLSMGGAAVKMSWPAKLSGPTPVYIRTVLDGEELILPA
RIIRAGNGRGIFIWTIDNLQQEFSVIRLVFGRADAWVDWGNYKADRPLLSLMDMVLSVKGLFRSSGDIVHRSSPTKPSAGNALSDDTNN
PSRKERVLKGTVKMVSLLALLTFASSAQAASAPRAVAAKAPAHQPEASDLPPLPALLPATSGAAQAGSGDAGADGPGSPTGQPLAADSA
DALVENAENTSDTATVHNYTLKDLGAAGSITMRGLAPLQGIEFGIPSDQLVTSARLVLSGSMSPNLRPETNSVTMTLNEQYIGTLRPDPA
HPTFGPMSFEINPIFFVSGNRLNFNFASGSKGCSDITNDTLWATISQNSQLQITTIALPPRRLLSRLPQPFYDKNVRQHVTVPMVLAQTYD
PQILKSAGILASWFGKQTDFLGVTFPVSSTIPQSGNAILIGVADELPTSFGRPQVNGPAVLELPNPSDANATILVVTGRDRDEVITASKGIAF
ASAPLPTDSHMDVAPVDIAPRKPNDAPSFIAMDHPVRFGDLVTASKLQGTGFTSGVLSVPFRIPPDLYTWRNRPYKMQVRFRSPAGEA
KDVEKSRLDVGINEVYLHSYPLRETHGLIGAVLQGVGLARPASGMQVHDLDVPPWTVFGQDQLNFYFDAMPLARGICQSGAANNAF
HLGLDPDSTIDFSRAHHIAQMPNLAYMATVGFPFTTYADLSQTAVVLPEHPNAATVGAYLDLMGFMGAATWYPVAGVDIVSADHVSD
VADRNLLVISTLATSGEIAPLLSRSSYEVADGHLRTVSHASALDNAIKAVDDPLTAFRDRDSKPQDVDTPLTGGVGAMIEAESPLTAGRTVL
ALLSSDGAGLNNLLQMLGERKKQANIQGDLVVAHGEDLSSYRTSPVYTIGTLPLWLWPDWYMHNRPVRVLLVGLLGCILIVSVLARALA
RHAARRFKQLEDERRKS

- Conserved domain search (CDS)



CDART

(Conserved domain architecture retrieval tool)

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

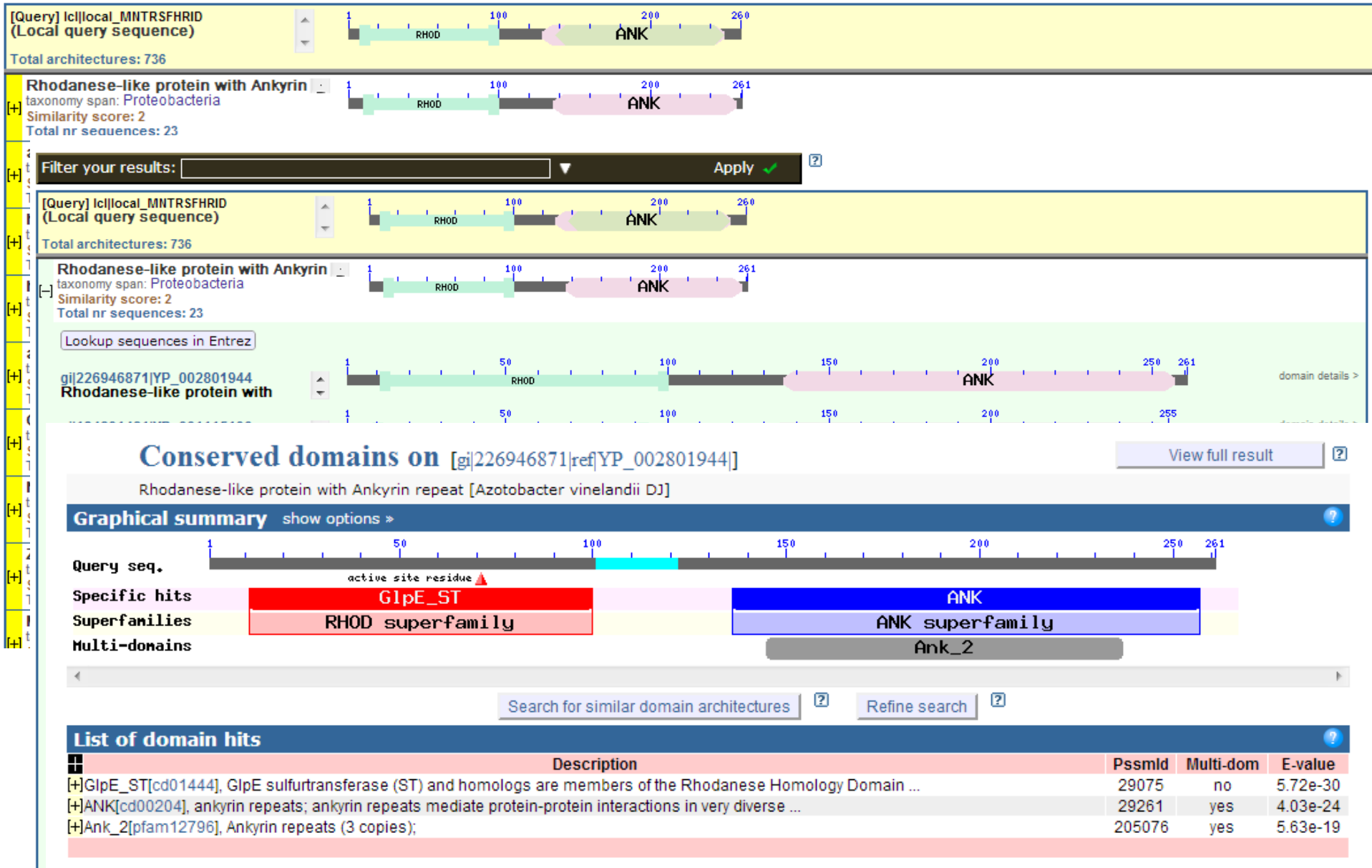
- Nevyhledává pouze izolované domény, ale zohledňuje jejich kombinace a vzájemná umístění v jednom proteinovém řetězci.

Úloha

- V následujícím proteinu byla zjištěna kombinace rhodanasové a ankyrinové domény. Zjistěte, zda je tento případ unikátní a v jakých jiných kombinacích se tyto domény v přírodě vyskytují. Použijte aplikaci **CDART**.

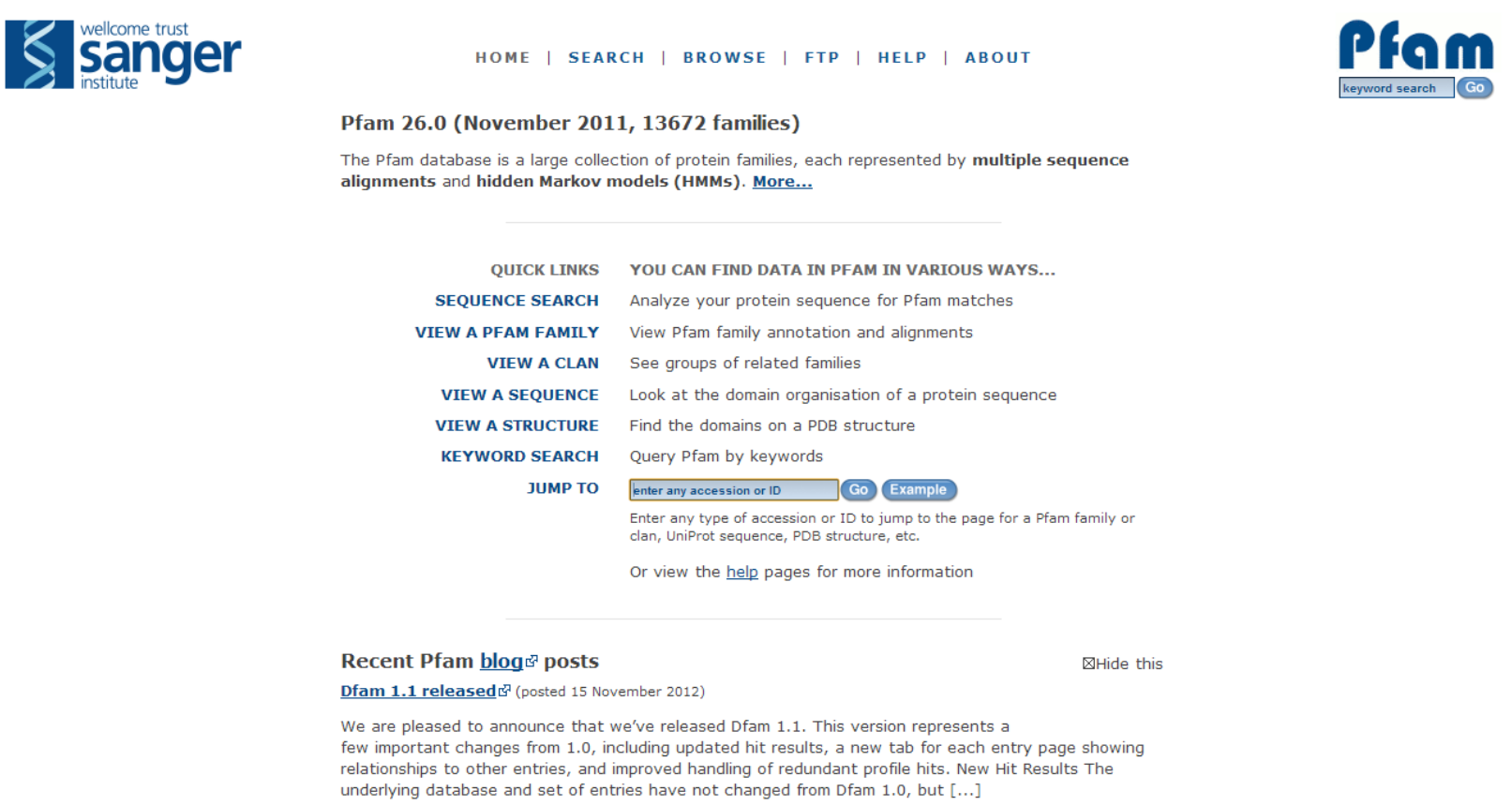
Sekvence:
MNTRSFHRIDVHKARELLQRPDTVLLDCRHPSDFRAGHIAGASPLGDYNADDHVLNIAKHRPVLIYCYHG
NASQMRAQLFADFGFAEVYSLDGGYEAWRKVHTPANSQLTEALQCWLMAQEFPAADIHARTRDGVTPL
MRAAGEGDPARVAELLAAGADPHQRNNDGNQALWFACVSENLDTLDLLVAVGAHLNHQNDNGATCL
MYAASA GKTAVVERLLAFGADRSLLSLDDFTALDMAANLECLNLLRETPRRIKAVT

Conserved domain architecture retrieval tool (CDART)



PFAM http://pfam.sanger.ac.uk

- Databáze proteinových rodin, vytvořená na základě Multiple sequence alignmentů (MSA) a Skrytých Markovových modelů (HMM)



Úloha

- Pokuste se určit funkci následujícího proteinu pomocí databáze **Pfam**.

Sekvence:
MRYIRLCIISLLATLPLAVHASPQPLEQIKQSESQLSGRVGMIEMDLASGRTLTAWRADERFPMM**S**TFKVVLCGAMLA
RVDAGDKQLERKIHYRQQDLVDYSPVSEKHLADGMTVGELCAAAITMSDNS**A**ANLLLATVGGPAGLTAFLRQIGDNV
TRLDRWETELNEALPGDARDTTTPASMAATLRKLLTSQRLSARSQRQLLQWMVDDR**V**AGPLIRSVLPAGWFIADKTG
ASKRGARGIVALLGPNNKAERIVVIYLRDTPASMAERNQQIAGIGAAL IEHWQR

- Odhadněte, které z vyznačených aminokyselin mají vliv na správnou funkci či strukturu proteinu.

PFAM http://pfam.sanger.ac.uk

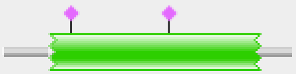
wellcome trust sanger institute

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Sequence search results

Show the detailed description of this results page.

We found **1** Pfam-A match to your search sequence (**all** significant). You did not choose to search for Pfam-B matches.



Show the search options and sequence that you submitted.

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

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment	
				Start	End	Start	End
Beta-lactamase2	Beta-lactamase enzyme family	Family	CL0013	47	257	48	256

```
#HMM    dtgeei.ginadeefpaaStiKvpil....eavlegelslderitvtkedivggsgilgkldgktlsvrdllelmiavSDNtAtnlLidrlg.ldavnawlrelglrdtrlrrklpdle.aldk
#MATCH   +g+++ + +ade+fp++St+Kv++      ++v++g  +l+ +i+++++d+v+ s++ +k+  + ++v +l+++++i+ SDN A+nlL++++g +  ++a+lr++g + trl+r++ +l+ al+
#PP     577888789**************97656699****************************999*****************************99*********************99888689999
#SEQ    ASGRTLtAWRADERFPMMSTFKVVLCgamlARVDAGDKQLERKIHYRQQDLVDYSPVSEKHLADGMTVGELCAAAITMSDNSAANLLLATVGgPAGLTAFLRQIGDNVTRLDRWETELNeALPGI
```

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

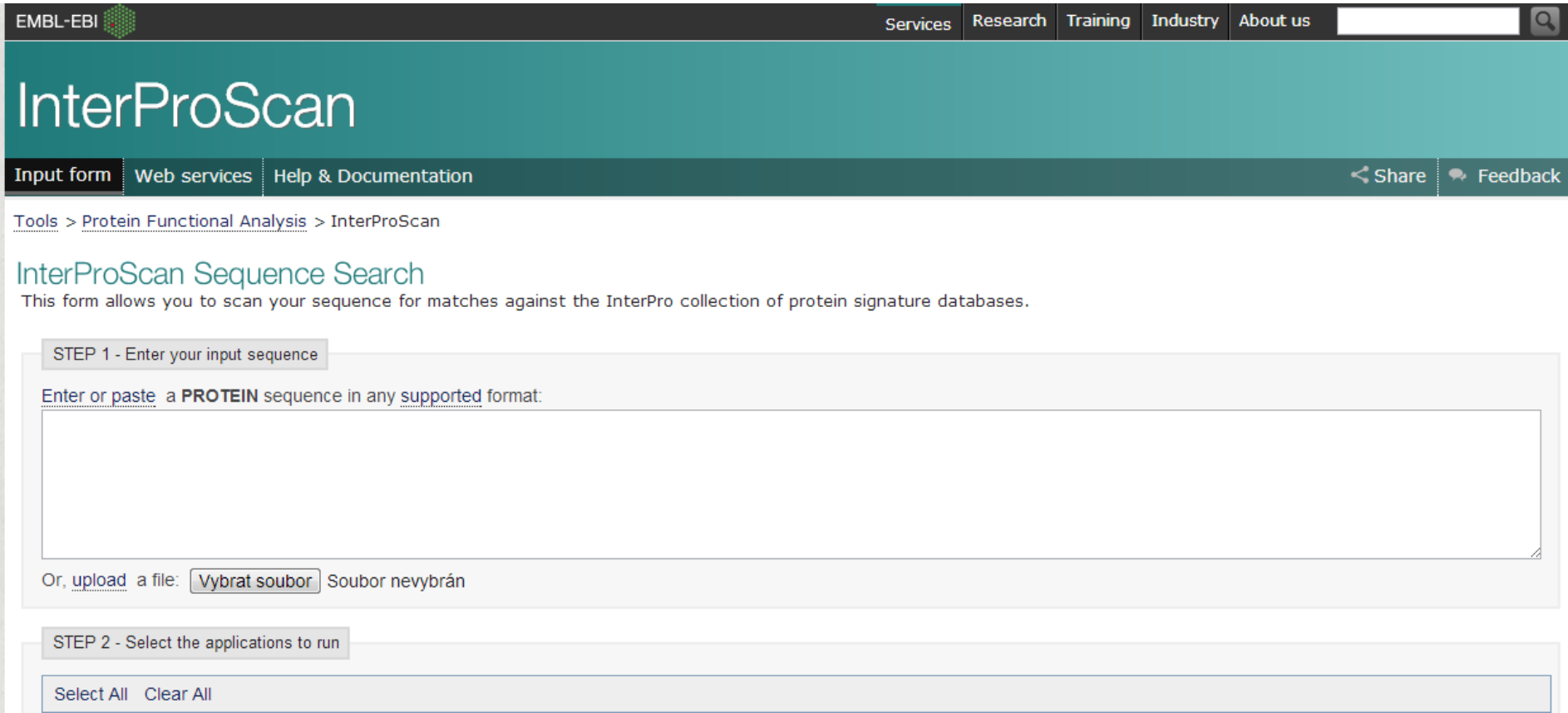
The Wellcome Trust

V kombinaci je síla…

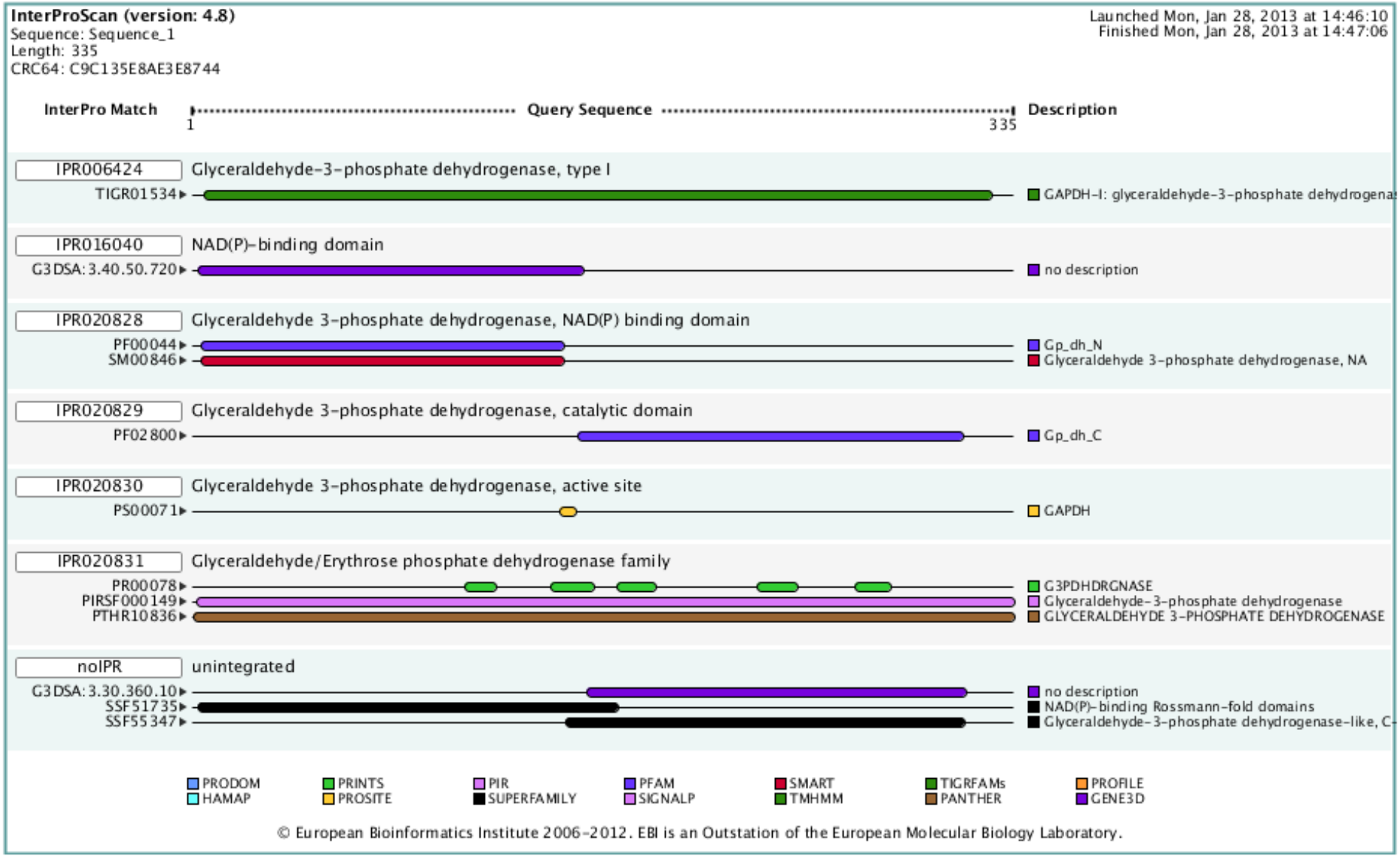
InterProScan http://www.ebi.ac.uk/Tools/pfa/iprscan/

Kombinovaný nástroj pro analýzu proteinové sekvence pomocí různých databází

- 14 aplikací v jednom běhu



InterProScan



Úloha

- Pokuste se určit funkci následujícího proteinu pomocí serveru **InterProScan**.

Sekvence:
MTELKNDRYLRALLRQPVDVTPVWMMRQAGRYLPEYKATRAQAGDFMSLCKNAELACEV
TLQPLRRYPLDAAILFSDILTIPDAMGLGLYFEAGEGPRFTAPVTCKADVDKLPIPDPEDELGYV
MNAVRTIRRELKGEVPLIGFSGSPWTLATYMVEGGSSKAFTVIKKMMYADPQALHLLLDKLA
KSVTLYLNAQIKAGAQSVMIFDTWGGVLTGRDYQQFSLYYMHKIVDGLLRENDGRRVPVTLF
TKGGGQWLEAMAETGCDALGLDWTTDIADARRRVGHKVALQGNMDPSMLYAPPARIEDE
VATILAGFGQGEGHVFNLGHGIHQDVPPEHAGAFVEAVHRLSA QYHN

InterProScan

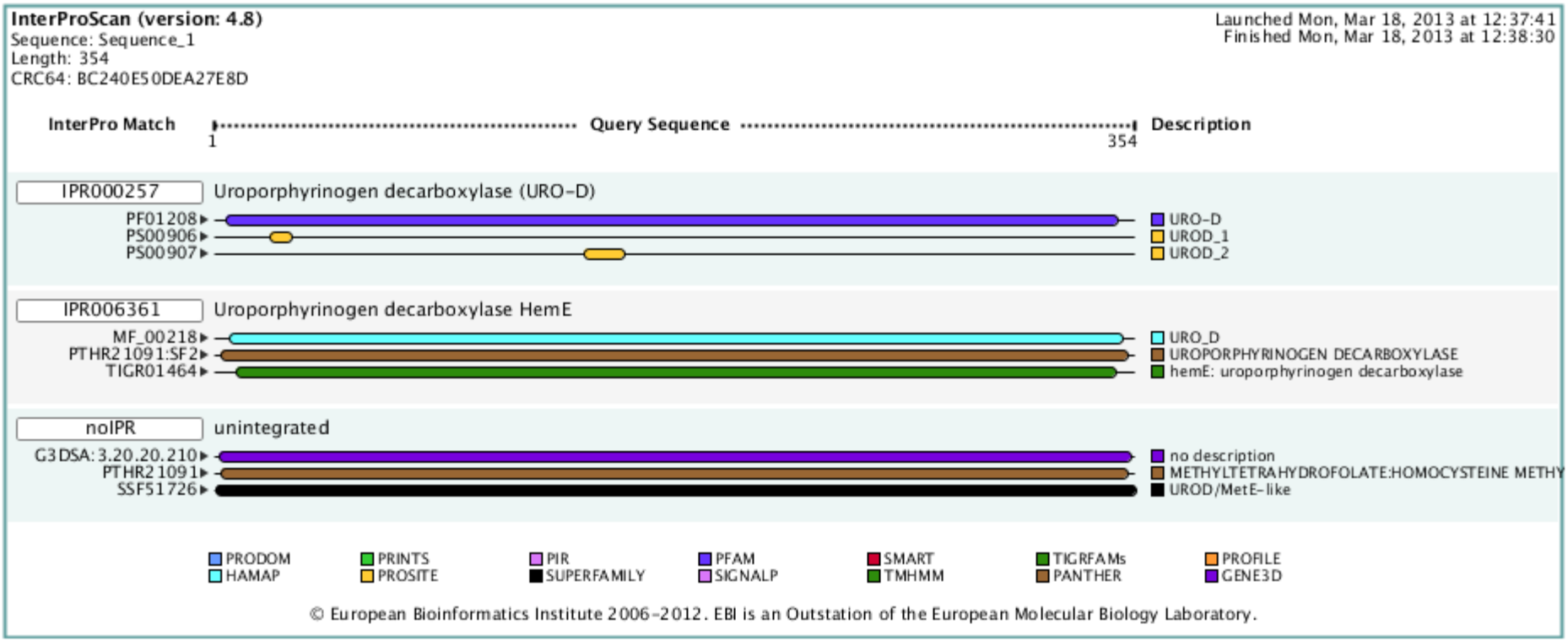
Input form | Web services | Help & Documentation

Tools > Protein Functional Analysis > InterProScan Sequence Search

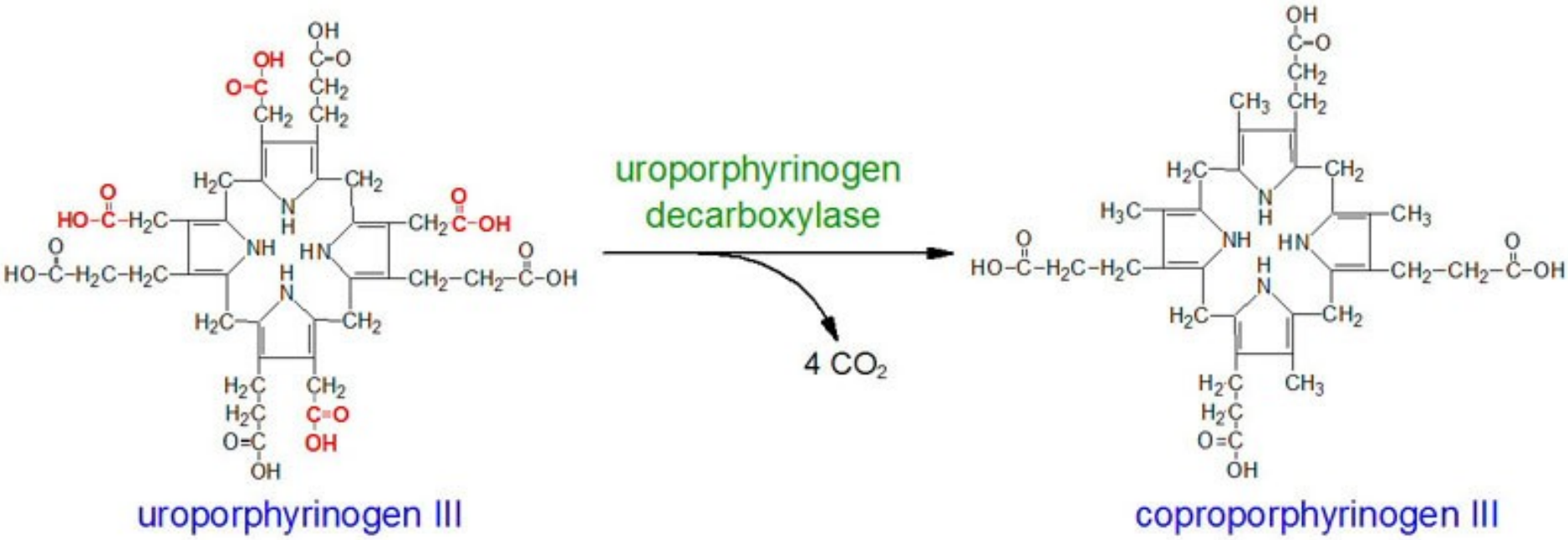
Results for job iprscan-I20130318-123741-0193-32040462-pg

Summary Table | Tool Output | Visual Output | Submission Details

Download in SVG format



Úloha



rofolát

PROPSEARCH – Když selže alignment

http://abcis.cbs.cnrs.fr/propsearch/

- Neprovádí alignment
- Porovnává složení (zastoupení) aminokyselin, molekulovou hmotnost, izoelektrický bod, atd. > celkem 144 parametrů
- Snaha zařadit protein do funkční rodiny
- **Další studium sekvence nutné !**

Úloha

- Následující sekvenci hypotetického proteinu analyzujte pomocí serveru **Propsearch**.

Sekvence:
MASPSILKKYGKYFEYCPLEERMIELAKKGEIADAMLLFEKEKPSEFVYKGDAIEKRLRNIYLSTR
LGVKAKINFNDYVIPRDLRWMLDIYESYLNMGENKVFLILGGELRYLIDFFESYLQFKGFYLLVV
KEAKDLLRFRNTCHYDAIIFSDSSILEYQNVDELKNLFNSLETTLKVHNRKNSVKVLLSPALPKAI
MSSKPYKVLEQFFKEKGIEMEGILPYQLNADDKLLPPHFHNSEMEKSKEYRELESKTKVYIQEF
LKKANMNDENEGNDNQKNTN

Brief introduction to PROPSEARCH ...

Please cite: Uwe Hobohm and Chris Sander: "A sequence property approach to searching protein databases", J.Mol.Biol. 251 (1995) 390-399

For a successful application, please have a look in: Uwe Hobohm and Chris Sander: "Does the HIV Nef protein mimic the MHC ?", FEBS-letters 333(1993)211-213.

Query sequence

Paste your sequence into the text area. The sequence may contain **ONLY amino acid residue characters (one letter code)** and carriage returns but no blanks AND NO position numbers (All characters apart from amino acid letters might now be truncated by PropSearch).

```
MPPGVLNDLNRNGIQCIASVTLLQRFPYHRNHHTIPTMDQNRPTVGGVTAEVLKAVVSEALL
GKFTAVFP
KLFKCQWKTYEGQKKFYANFSTDCTDSNSNNMSRKFCVGGNWKMNGDQKSIAEIAKTLSSAA
LDPNTEVV
IGCPAIYLMYARNLLPCELGLAGQNAYKVAKGAFTGEISPAMLKDIGADWVILGHSERRAIF
GESDALIA
EKAEHALAEGLKVIACIGETLEEREAGKTNEVVARQMCAYAQKIKDWKNVVVAYEPVWAIGT
GQTATPDQ
AQEVHAFLRQWLSDNISKEVSASLRIQYGGSVTAANAKELAKKPDIDGFLVGGASLKPEFVD
IINARQ
```

Clear Form

Rank	ID	DIST	LEN2	POS1	POS2	pI	DE
1	ycx8_euggr	0.00	281	1	281	7.51	Hypothetical 33.1 kDa protein in RBCL-ATPE intergenic region
2	pyrb_aquae	6.69	291	1	291	7.72	Aspartate carbamoyltransferase (EC 2.1.3.2) (Aspartate
3	y800_pyrab	7.35	326	1	326	6.09	Hypothetical ATP-binding protein PAB0800.
4	ynz4_caeel	7.37	297	1	297	9.38	Hypothetical 35.0 kDa protein T09A5.4 in chromosome III.
5	y983_aquae	7.44	292	1	292	8.89	Hypothetical protein AQ_983.
6	ye16_yeast	7.69	306	1	306	7.25	Hypothetical 35.9 kDa protein in ISC10 3'region.
7	cher_camje	7.76	262	1	262	9.38	Chemotaxis protein methyltransferase (EC 2.1.1.80).
8	bra2_chith	7.77	245	1	245	7.26	Balbiani RING A 28 kDa protein precursor.
9	mag_yeast	7.81	296	1	296	8.16	DNA-3-methyladenine glycosylase (EC 3.2.2.21) (3-methyladenine DNA
10	yl51_yeast	7.82	340	1	340	6.59	Hypothetical 39.8 kDa protein in MPT4-ACS2 intergenic region.
11	yj17_aquae	7.91	345	1	345	8.08	Hypothetical protein AQ_1917.
12	yz38_aquae	7.98	320	1	320	9.28	Hypothetical protein AA38.
13	rfbj_salmu	8.02	293	1	293	8.50	CDP-abequose synthase (EC 4.2.1.-).
14	ars2_aquae	8.02	299	1	299	6.73	Putative arsenical pump-driving ATPase 2 (EC 3.6.3.16) (Arsenite-
15	met8_yeast	8.02	274	1	274	6.27	Siroheme biosynthesis protein MET8 [Includes: Precorrin-2 oxidase
16	kkih_lacla	8.03	262	1	262	6.08	Probable aminoglycoside 3'-phosphotransferase (EC 2.7.1.95) (Kanamycin
17	yp76_borbu	8.07	265	1	265	8.61	Hypothetical protein BBA76.
18	rnh_bpt4	8.07	305	1	305	9.00	Ribonuclease H (EC 3.1.26.4) (RNase H).
19	mtm3_metja	8.08	289	1	289	9.46	Modification methylase MjaIII (EC 2.1.1.72) (Adenine-specific

SMART http://smart.embl-heidelberg.de/

Analýza zastoupení **proteinových domén**

Prohledávání dle sekvencí, domén, taxonů

STRING http://string-db.org

Funkční vazby různých proteinů

Na základě výskytu v genomu, zapojení do metabolických drah, textového hledání,…

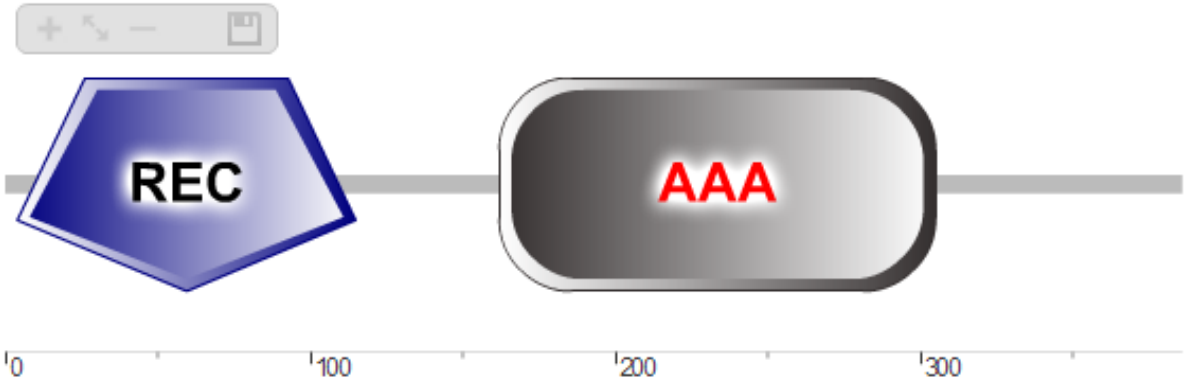
Úloha

- Pomocí serveru **SMART** analyzujte domény v následující sekvenci a následně pomocí serveru **STRING** prozkoumejte pravděpodobné zapojení v metabolismu.

Sekvence:
MSIEHILIIDDDPHILALLSEILGARNFSVSSAPGVKQAIKQISNCPFDLIISDMNMPDGSGLDII
QYTKQHRPQTPILVITAFGTIQNAVEAMRFGAFNYLTKPFSPDALFTLIAKAEELQALQQDNLF
LQSQGSSISHPLIAESPSMKQLLDKARRAANSSANIFVHGESGCGKENLSFFIHKHSPRSTKPYI
KVNCAAIPDTLLESEFFGHEKGAFTGATTKKVGRFELAHQGTLLLDEITEIPIHLQAKLLRAIQE
QEFEHIGGIKTLPVNIRFLATSNRDLEEAIETKVLRQDLYYRLSVISLHIPPLRDRKEDILPLAHYYL
EKFCKMNNKPPKTLSLEAQRNLLDYSWPGNVR ELSNVLERTVILENDPAITPSMLALL

Domains within *Chlamydia trachomatis 434/Bu* protein B0B841_CHLT2 (B0B841)

Two component system response regulator



Information | Architecture | Interactions | Orthology

Length	386 aa
Source database	UniProt
Identifiers	B0B841_CHLT2, B0B841

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the the same piece of sequence; the priority for display is given by **SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > Unstruct** shown in the above diagram are marked as '**overlap**' in the right side table below.

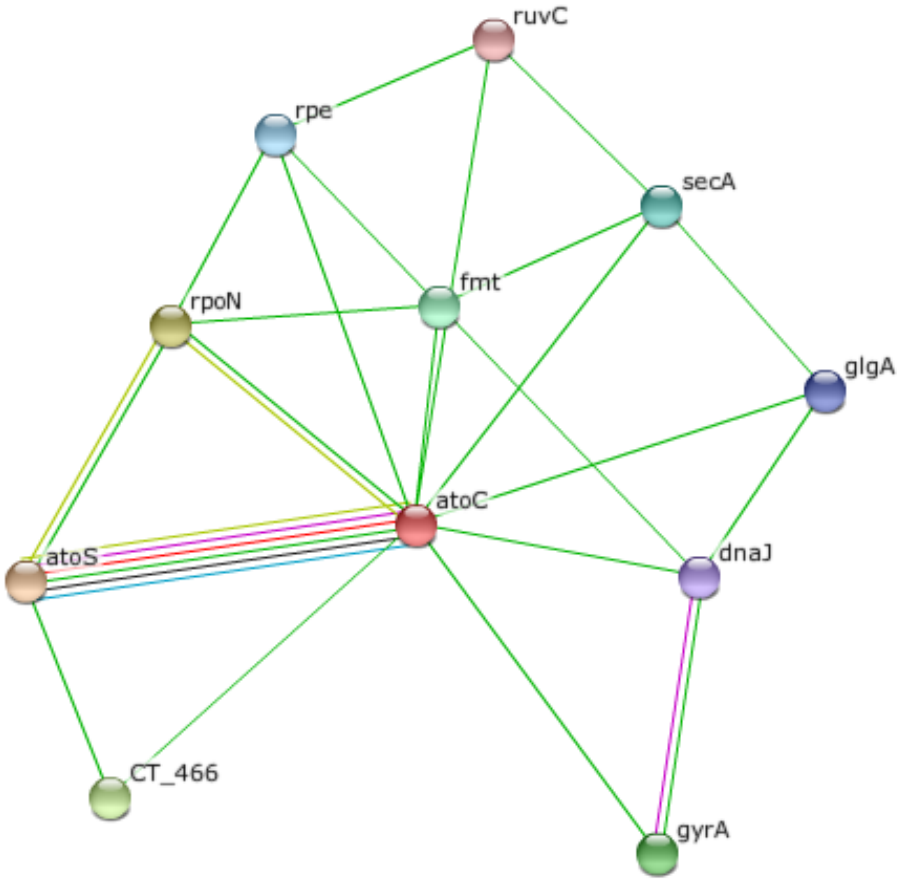
Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
REC	4	115	4.25e-32
AAA	162	305	6.98e-7

Features NO

Name	Start ▲	E
low complexity	6	

Click on a row to highlight the feature in the diagra



This is the **evidence view**. Different line colors represent the types of evidence for the association.

confidence evidence actions interactive advanced more less save

(requires Flash player 10 or better)

Porovnání predikce a experimentu

- **Predikce**:
  - \+ Rychlá (sekundy-hodiny), levná/dostupná (Freeware)
  - \+ Spolehlivá pro známé (!) proteiny a pro proteiny s vysokou homologií
  - – Pouze kvalitativní
  - – Málo spolehlivá pro neznámé proteiny
  - – Nepoužitelná pro unikátní případy
- **Experiment**:
  - \+ Teoreticky použitelný pro libovolný protein
  - – Finančně (i miliony Kč) a časově náročný (minuty-hodiny + příprava vzorku = týdny až roky)

Rady do života

O daném proteinu získej maximum informací

Kombinuj různé predikční programy a přístupy

Kriticky kontroluj SW výstupy