

Blok 3

Funkce proteinů

C3211 Aplikovaná bioinformatika
Přednášející: Josef Houser



Funkce proteinů

- Jako funkci označujeme obvykle činnost, kterou je protein schopen provádět v organismu
- Protein může mít jen jednu funkci, ale také více funkcí
- Část proteinu, která má vlastní funkci, je obvykle označovaná jako **doména**



Funkce proteinů

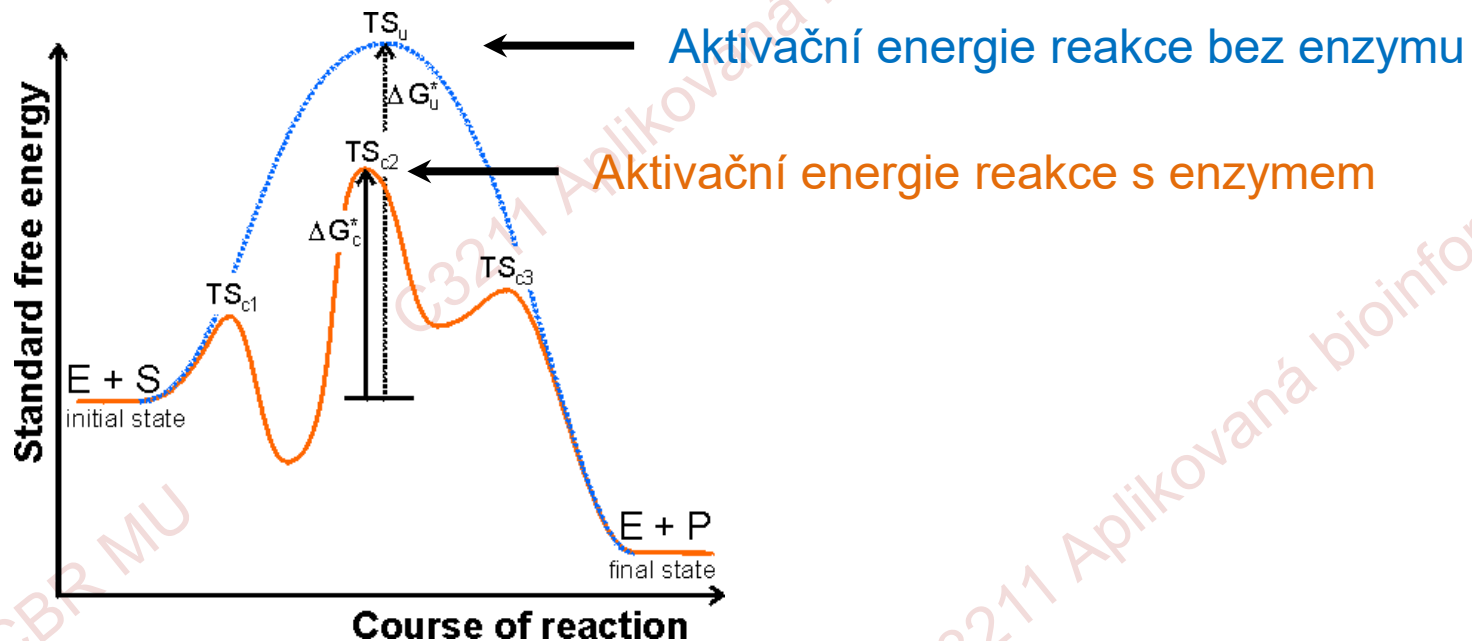
Podle hlavních funkcí rozeznáváme mezi proteiny:

- **Enzymy** – 7 hlavních tříd
- **Strukturní proteiny** – keratin, kolagen
- **Transportní proteiny** – přenašeče, kanály
- **Obranné proteiny** – protilátky
- **Regulátory a receptory** – hormony, transkripční faktory, rhodopsin
- ...



Klasifikace enzymů

- **Enzymy** jsou proteiny, které katalyzují chemickou reakci – snižují aktivační energii nutnou pro přeměnu substrátu na produkt
- Dělíme je podle typu reakce, kterou katalyzují



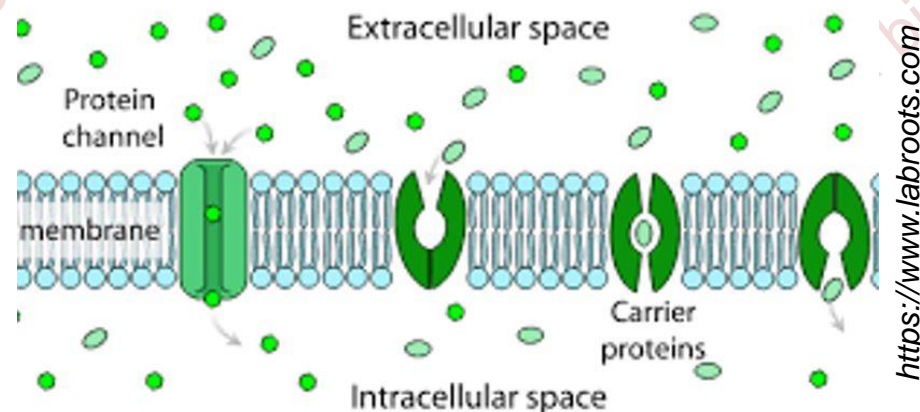
Klasifikace enzymů

Dle IUBMB: <https://www.qmul.ac.uk/sbcs/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)	24
2	Transferasy	Katalyzují přenos skupin: amino-, metyl-, acyl-, glykosyl-, fosforyl-	10
3	Hydrolasy	Katalyzují hydrolytické štěpení vazeb mezi atomem uhlíku a jinými atomy (spotřeba molekuly H ₂ O)	13
4	Lyasy	Katalyzují adiční reakci na dvojnou vazbu nebo eliminační reakci mezi 2 atomy uhlíku za vzniku dvojnou vazby	8
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
7	Translokasy	Membránové enzymy, které zajišťují aktivní transport látek s využitím energie ATP	6

Transportní proteiny

- Jejich funkcí je **přenos** látek přes membránu. Přenášena může být molekula (např. glukosa), iont (např. K^+) nebo elektron.
- Přenos může probíhat **pasivně** (ve směru koncentrace) nebo **aktivně** (proti směru koncentrace). U aktivního transportu je typicky přenášena víc než jedna částice (přenos jedné částice ve směru koncentračního gradientu umožňuje přenos druhé částice proti směru jejího koncentračního gradientu) nebo je přenos spojený s chemickou reakcí



Transportní proteiny

Dle TCDB (transporter classification database):

<http://gmul.ac.uk/sbcs/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

- Protilátky slouží jako **obrana** organismu proti patogenům
- U člověka existuje několik typů protilátek, které se liší funkcí, strukturou i relativním zastoupením v krevní plazmě
- Protilátky jednotlivých typů mají velmi konzervovanou celkovou strukturu, liší se zejména v sekvenci krátkých variabilních úseků
- U jiných živočichů se vyskytují i jiné typy protilátek

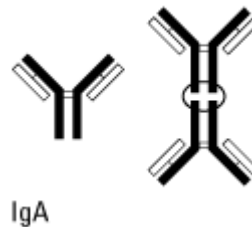
Lidské protilátky

	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%)

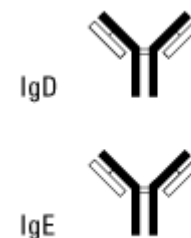
IgG



IgM



IgA

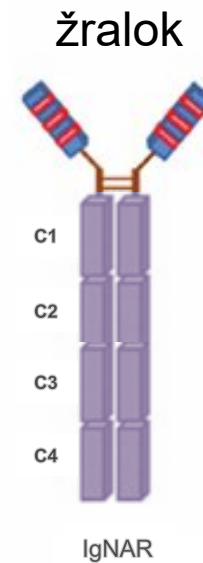
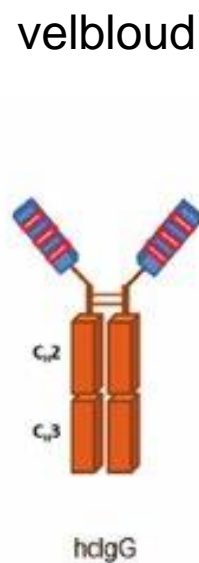
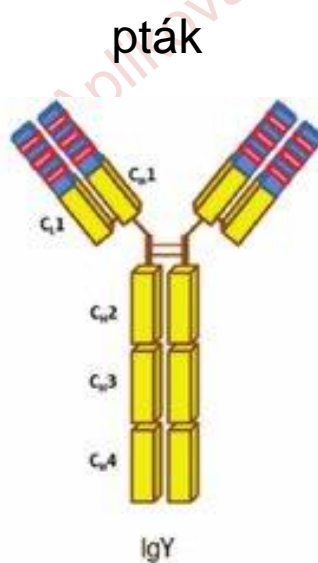
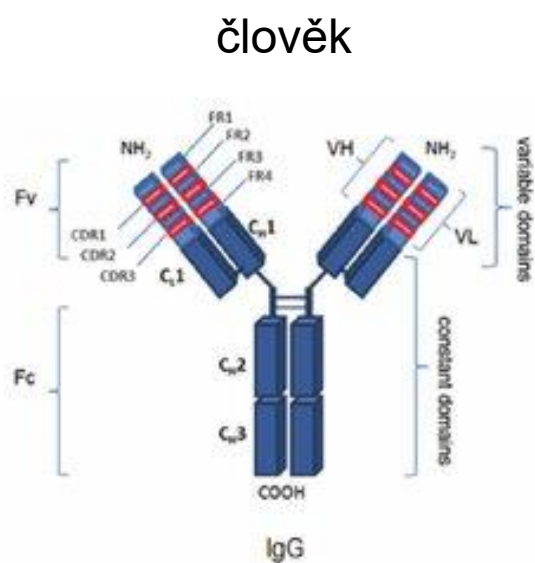


IgD

IgE

Protilátky obratlovců

- Většinu protilátek máme společnou s dalšími obratlovcí
- Příklady dalších protilátek některých obratlovců:
 - IgY – ptáci (např. kur domácí)
 - hdgG – velbloudi, lamy
 - IgNAR – žraloci



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)



Experimentální určení funkce

- **Izolace proteinu s konkrétní funkcí**

- Typický **vstupní materiál** je lyzát (rozemletá rostlina, lyzované krvinky, rozdrcená tkáň,...)
- Proveďte se **separační** krok (např. filtrace) a určí, ve které části zůstala zachována sledovaná funkce
- Část s funkcí se podrobí dalšímu separačnímu kroku a opět se určí, ve které frakci zůstala aktivita
- Postupně se odstraní všechny části bez aktivity až zůstane čistý aktivní protein
- Tento protein je **identifikován** (např. hmotnostní spektrometrií)

Experimentální určení funkce

- **Stanovení funkce u konkrétního proteinu**

- U čistého proteinu je možno **stanovit experimentálně aktivitu** v různých testech (tzv. functional assays)

- Je nutno mít odhad (např. na základě příbuznosti), o jakou aktivitu se může jednat, jinak je její určení otázkou náhody

- Existují stovky různých možných aktivit zahrnující tisíce (ne-li miliony) možných substrátů.

Pro mnohé reakce navíc neexistuje vhodný laboratorní test

→ **náhodné určení proteinové funkce je téměř vyloučeno**

Určení funkce proteinu

Situace – Přichází za Vámi méně zkušený kolega, např. mladší student, s roztokem proteinu ve zkumavce a otázkou:

Jakou funkci má tento protein?



- Nabízí se několik řešení
- Nezavrhujte ta jednoduchá

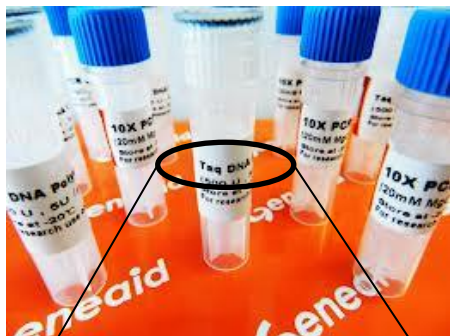
Určení funkce proteinu

1. U komerčně dodávaného proteinu bývá funkce uvedena (většinou je známa)

2. Ten kdo protein připravil má často reálnou představu o jeho funkci (většinou ví, proč se jím zabývá)



Jakou funkci má tento protein?



Taq DNA polymerase



3. Nemáme-li jiný zdroj informací, je možno funkci predikovat pomocí SW nástrojů

Drobné nápovědy

- **Některé základní informace o proteinu mohou pomoci:**
 - **Lokalizace** – predikce lokalizace (např. hledání signálních peptidů) může napovědět KDE se daný protein realizuje
 - **Hydrofobicita** – může napovědět, že protein se nachází v membráně nebo uvnitř proteinového komplexu
 - **pI** – některé proteiny mají typické hodnoty izoelektrického bodu (např. histony jsou bazické → $pI > 7$)

Predikce funkce proteinu

Je nutno znát **sekvenci**:

- **Databáze** – pracujeme-li se známým proteinem
- **Sekvenace** – pracujeme-li s neznámým proteinem nebo pro ověření identity proteinu

Ještě lepší je znát **strukturu** (2D, 3D)

- **Databáze** – u známých proteinů
- Určení **2D struktury** (viz. minulá přednáška)
- Určení **3D struktury** (viz. příští přednáška)

Proteinové domény

- **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 se stejnou nebo podobnou funkcí – **konzervované domény**

Vícedoménové proteiny

- Mnoho proteinů obsahuje více než jednu funkční doménu
- V jednom proteinu se mohou zároveň vyskytovat enzymatické, strukturní i interakční domény
- Vzájemné uspořádání domén může být charakteristické pro některé **proteinové rodiny**

Proteinové rodiny

- Proteiny se na základě své struktury (a funkce) řadí do hierarchických skupin. Většinou rozlišujeme:
 - Proteinová nadrodina
 - Proteinová rodina
 - Proteinová podrodina
- Pro charakterizaci proteinových rodin se často používá Multiple sequence alignment (MSA) s aplikací skrytých Markovových modelů (HMM)
- Dělení na základě doménového uspořádání a proteinových rodin může ale nemusí být totožné

Databáze strukturních a funkčních motivů

- **Neannotované**, 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** – databáze proteinových rodin odvozená z UniProtKB
- **KEGG** – složená databáze obsahující systémové, genomické a chemické informace
- **CDD** – databáze proteinových domén
- ...

Hledání v proteinových databázích

- Nejčastěji využíváme vyhledávání pomocí **sekvence** (prohledávání pomocí párového alignmentu, např. BLAST)
- Méně časté je vyhledávání pomocí **klíčových slov** (např. lokalizace, známé vlastnosti, ...)
- Známe-li konkrétní záznam, je možno hledat rovněž pomocí **identifikačního čísla**

Úloha 1

- Vyhledejte homologní proteiny k následující sekvenci pomocí nástroje BLAST na serveru **UniProt** (<http://www.uniprot.org/>). Lze na základě výsledku určit funkci tohoto proteinu?



SHLSQPWPITCFADRPTPRRSSPDASGQTMHSVFVVHVPYPVVFLKPAH
LTPQWYRHPIPVNPVVRQPHLPVLYPAPNAGHTPAHSRQGDAALQPLF
SVPQTVNPTGPVIHGDVAKQKPDTGQSWALNPYCTENWRRILRISRNS
HGQRMPLTTLLQKTSGRNATLITKNSDQNTTTSIVSESSMTISACCHSAIL
RNN

PFAM

(Protein Families)



<http://pfam.xfam.org/>

- Spravuje ji EBI (Evropský institut pro bioinformatiku)
- Databáze proteinových rodin, vytvořená na základě Multiple sequence alignmentů (MSA) a Skrytých Markovových modelů (HMM)
- Na základě výsledků je možné určovat typické oblasti – proteinové domény

Pozor funkcionality databáze Pfam byly přemístěny do Interpro

Úloha 2

- Pokuste se určit funkci následujícího proteinu pomocí databáze **Pfam** (<https://pfam.xfam.org/>).

Sekvence:

MRYIRLCIISLLATLPLAVHASPQPLEQIKQSESQLSGRVGMIEMDLASGRTLTAWRADERFPMM**S**TFKVVLCGAMLA
RVDAGDKQLERKIHQRQDLVDYSPVSEKHLADGMTVGELCAAITMSDNSAANLLLATVGGPAGLTAFLRQIGDNV
TRLDRWETELN**E**ALPGDARDTTTPASMAATLRKLLTSQRLSARSQRQLLQWMVDDR**V**AGPLIRSVLPAGWFIADKTG
ASKRGARGIVALLGPNKAERIVVIYLRDTPASMAERNQQIAGIGAAL IEHWQR

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

SMART

(Simple Modular Architecture Research Tool)

<http://smart.embl-heidelberg.de/>



- Rozpoznává přes 500 rodin **proteinových domén**
- Prohledávání dle sekvencí, domén, druhů i vyšších taxonů
- Umožňuje také vyhledávání podle zvolených kombinací domén – vhodné pro vícedoménové proteiny

Úloha 3

- Pomocí serveru **SMART** (<http://smart.embl-heidelberg.de/>) identifikujte domény v následující sekvenci.

Sekvence:

MSIEHILIIDDDPHILALLSEILGARNFSVSSAPGVKQAIKQISNCPFDLIISDMNMPDGSGLDII
QYTKQHRPQTPILVITAFGTIQNAVEAMRFGAFNYLTKPFSPDALFTLIAKAEELQALQQDNLF
LQSQGSSISHPLIAESPSMKQLLDKARRAANSSANIFVHGESGCGKENLSFFIHKHSPRSTKPYI
KVNCAAIPDTLLESEFFGHEKGAFTGATTKKVGRFELAHQGTLLLDEITEIPIHLQAKLLRAIQE
QEFEHIGGIKTLPVNIRFLATSNRDLEEAIETKVLQRDLYYRLSVISLHIPPLRDRKEDILPLAHYYL
EKFCKMNNKPPKTLSEAQRNLLDYSWPGNVR ELSNVLERTVILENDPAITPSMLALL

CDD

Databáze konzervovaných domén – conserved domain database

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

- Spravovaná Národním centrem pro biotechnologické informace (NCBI)
- Obsahuje přehledné informace o známých proteinových doménách
- Shromažďuje data z různých zdrojů

CDD zdroje

Abbreviation	Database Name	Description
SMART	Simple Modular Architecture Research Tool	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	Protein families	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	Clusters of Orthologous Groups 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	The Institute for Genomic Research's database of protein families	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	Protein 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.

Úloha 4

- Vyhledejte pomocí nástroje **CD search** (NCBI) (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) konzervované domény následujícího proteinu.

Sekvence:

PEVRSSTQSESGMSQWMGKILSIRGAGLIIGVFGLCALIAATSVTLPEQQQLIVAFVVCVVIFFIVGHKPSRRSQIFLEVLVSLVSLRYLTWRLT
ETLSFDTWLQGLLGTMLLVAELYALMMLFLSYFQTIAPLHRAPLPLPPNPDEWPTVDIFVPTYNEELSIVRLTVLGSGLIDWPPEKVRVHIL
DDGRRPEFAAFAAECGANIARPTNEHAKAGNLNYAIGHTDGDYILIFDCDHVPTRAFLQLTMGWMVEDPKIALMQTPHHFYSPPDF
QRNLSAGYRTPPEGNLFYGVVQDGNDFWDATFFCGSCAILRRTAIEQIGGFATQTVTEDAHTALKMQRLGWSTAYLRIPLAGGLATERLI
LHIGQRVRWARGMLQIFRIDNPLFGRGLSWGQRLCYLSAMTSFLFAVPRVIFLSSPLAFLFFGQNIIAASPLALLAYAIPHMFHAVGTASKI
NKGWRYSEFWSEVYETTMALFLVRVTIVTLLSPSRGKFNVTDKGGLEKGYFDLGAVYPNIILGLIMFGGLARGVYELSFHGLDQIAERAYL
LNSAWAMLSLIIILAAIAVGRETQQKRNSHRIPATIPVEVANADGSIIVTGVTEDLSMGGAAVKMSWPAKLSGPTPVYIRTVLDGEELILPA
RIIRAGNGRGIFIWTIDNLQQEFSVIRLVFGRADAWVDWGNKYADRPLLSLMDMVLSVKGLFRSSGDIVHRSSPTKPSAGNALSDDTNN
PSRKERV LKGTVMVSL LALLTFASSAQAASAPRAVA AKAPAHQPEASDL PPLPALLPATSGAAQAGSGDAGADGPGSPTGQPLAADSA
DALVENAENTS DTATVHNYTLKDLGAAGSITMRGLAPLQGIEFGIPSDQLVTSARLVLSGSMSPNLRPETNSVTMTLNEQYIGTLRPDPA
HPTFGPMSFEINPIFFVSGNRLNFNFASGSKGCS DITNDTLWATISQNSQLQITTIALPPRRLSRLPQPFYDKNVRQHVTVPMVLAQTYD
PQILKSAGILASWFGKQTD FLGVTFPVSSTIPQSGNAILIGVADELPTS FGRPQVNGPAVLELPNPS DANATILVVTGRDRDEVITASKGIAF
ASAPLPTDSHMDVAPVDIAPRKPN DAPSFIAMDHPVRF GDLVTASKLQGTGFTSGVLSVPFRIPPDLYTWRNRPYKMQVRF RSPAGEA
KDVEKSRLDVGINEVYLHSYPLRETHGLIGAVLQGVGLARPASGMQVHDL DVPPWTVFGQDQLNFYFDAMPLARGICQSGAANNAF
HLGLDPDSTIDFSRAHHIAQMPNLAYMATVGF PFTTYADLSQTAVVLPEHPNAATVGAYLDLMGFMGAATWYPVAGVDIVSADHVSD
VADRNLLVISTLATS GEIAPLLSRSSYEADGHLRTVSHASALDNAIKAVDDPLTAFRDRD SKPQDQDVTPLTGGVGAMIEAESPLTAGRTVL
ALLSSDGAGLNNLLQMLGERKKQANIQQD LVAHGEDLSSYRTSPVYTIGTLPLWLWPDWYMHNRPV RVLVGLLGCILIVSVLARALA
RHAARRFKQLEDERRKS

CDART

(Conserved domain architecture retrieval tool)

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

- Využívá anotované informace z databází Pfam a SMART
- Nevyhledává pouze izolované domény, ale zohledňuje jejich kombinace a vzájemná umístění v jednom proteinovém řetězci.

Úloha 5

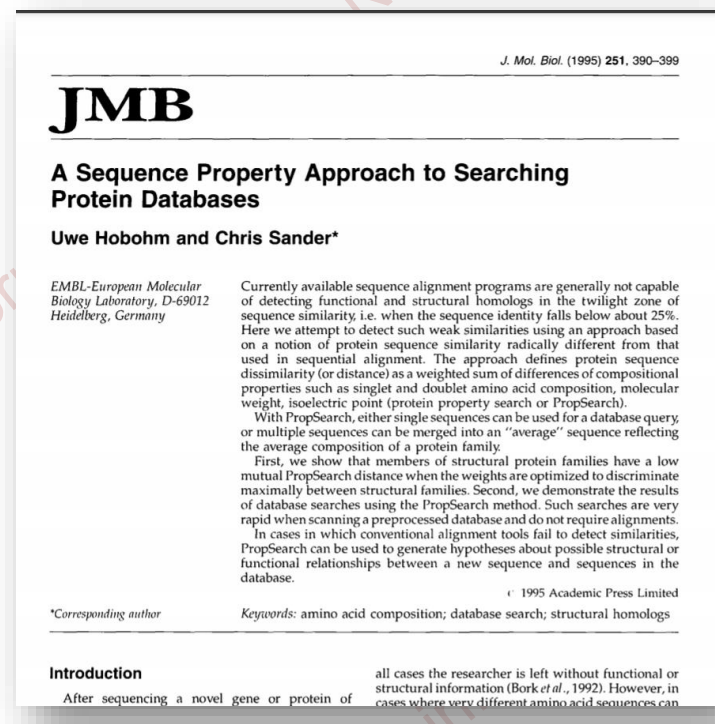
- V následujícím proteinu byla zjištěna kombinace rhodanasové a ankyrinové domény. Zjistěte, zda je tento případ unikátní a v jakých jiných kombinacích se tyto domény v přírodě vyskytují. Použijte aplikaci **CDART** (<http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi>).

Sekvence:

```
MNTRSFHRIDVHKARELLQRPDTVLLDCRHPSDFRAGHIAGASPLGDYNADDHVLNIAKHRPVLIYCYHG  
NASQMRAQLFADFGFAEVYSLDGGYEAWRKVHTPANSQLTEALQCWLMAQEFPAADIHARTRDGVTP  
MRAAGEGDPARVAELLAAGADPHQRNNDGNQALWFACVSENLDTLDLLVAVGAHLNHQNDNGATCL  
MYAASA GKTAVVERLLAFGADRSLLSLDDFTALDMAANLECLNLLRETPRRIKAVT
```

PROPSEARCH – Když selže alignment

- Proteiny, které nemají blízké homology je těžké predikovat
- PROPSEARCH neprovádí alignment
- Porovnává složení (zastoupení) aminokyselin, molekulovou hmotnost, izoelektrický bod, atd. > celkem 144 různých parametrů
- Snaha zařadit protein do funkční rodiny
- **Další studium sekvence nutné !**



V kombinaci je síla...

- Máme-li možnost, je vždy lépe **kombinovat informace** z různých zdrojů
- Minimalizujeme tak náhodné chyby a maximalizujeme pravděpodobnost správné predikce



InterProScan

<http://www.ebi.ac.uk/Tools/pfa/iprscan/>

Společné vyhledávací rozhraní, které využívá různé databáze pro maximalizaci výstupu

Slouží pro analýzu proteinové sekvence pomocí 14 různých aplikací v jednom běhu

Spravuje jej EBI (Evropský institut pro bioinformatiku)



InterPro

Classification of protein families

Úloha 6

- Pokuste se určit funkci následujícího proteinu pomocí serveru **InterProScan** (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>).

Sekvence:

MTELKNDRYLRALLRQPVDVTPVWMMRQAGRYLPEYKATRAQAGDFMSLCKNAELACEV
TLQPLRRYPLDAAILFSDILTIPDAMGLGLYFEAGEGPRFTAPVTCKADVDKLPIDPEDELGYV
MNAVRTIRRELKGEVPLIGFSGSPWTLATYMVEGGSSKAFTVIKKMMYADPQALHLLLDKLA
KSVTLYLNAQIKAGAQSVMIFDTWGGVLTGRDYQQFSLYMHKIVDGLLRENDGRRVPVTLF
TKGGGQWLEAMAETGCDALGLDWTTDIADARRRVGHKVALQGNMDPSMLYAPPARIEDE
VATILAGFGQGEGHVFNLGHGIHQDVPPEHAGAFVEAVHRLSAQYHN

STRING



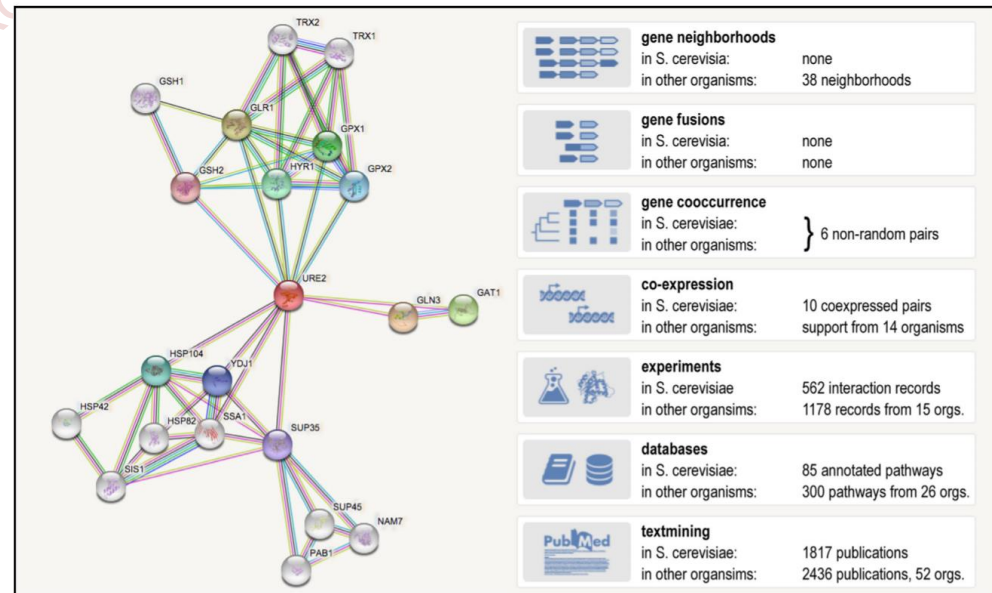
<http://string-db.org>

Vyhledává funkční vazby mezi proteiny

Po zadání zkoumaného proteinu (sekvence, názvu) vyhledá další proteiny ve vztahu k zadanému

Souvislosti hledá na základě:

- databázových informací
- výskytu v genomu
- metabolických drah
- experimentálních interakcí
- textového hledání
- ...



Úloha 7

- Pomocí serveru **STRING** (<http://string-db.org>) prozkoumejte zapojení tohoto lidského proteinu v metabolismu.

Sekvence:

MQRSPLEKASVVSCLFFSWTRPILRKGYSRLELSDIYQIPSVDSADNLSEKLEREWRELSKKNPKLINALRRCFFWRFMFYGFILYLGE
VTKAVQPLLLGRIIASYDPDNKEERSIAIYLGIGLCLLFIVRTLLLHPAIFGLHHIGMQMRIAMFSLIYKTKLSSRVLDKISIGQLVSLSSNNL
NKFDEGLALAHFVWIAPLQVALLMGLIWELLQASAFGLGFLIVLALFQAGLGRMMMKYRDQRAGKISERLVITSEMIENIQSVKAYCW
EEAMEKMIENLRQTEKLTRKAAYVRYFNSSAFFSFGFFVFLSVLPYALIKGIILRKIFTTISFCIVLRMAVTRQFPWAVQWYDSLGAINKI
QDFLQKQEKTYLNTTTEVVMENVTAFWEEGFGELFEKAKQNNNNRKTSSNGDDSLFFSNFSLGTPVLKDINFKIERGQLLAVAGSTG
AGKTSLLMMIMGELEPSEGGKIKHSGRISFCSQFSWIMPGTIKENIIFGVSYDEYRYSVIKACQLEEDISKFAEKDNIVLGEGGITLSGGQR
ARISLARAVYKDADLYLLDSPFGYLDVLTKEIFESCCKLMANKTRILVTSKMEHLKADKILILHEGSSYFYGTFSSELQNLQPDFSSKLMGC
DSFDQFSAERRNSILTETLHRFSLEGDAPVSWTETKKQSFQKQTEGFEKRNKNSILNPINSIRKFSIVQKTPLQMNGIEEDSDEPLERRLSLVP
DSEQGEAILPRISVISTGPTLQARRRQSVLNLMTHSVNOGQNIHRKTTASTRKVSLAPQANLTEDIYSRRLSQETGLEISEEINEEDLKECF
FDDMESIPAVTTWNTYLRVITVHKSLIFVLIWCLVIFLAEVAASLVVLWLLGNTPLQDKGNSTHSRNNSYAVIITSTSSYYVFYIYVGVADTLL
AMGFFRGLPLVHTLITVSKILHHKMLHSVLQAPMSTLNTLKAGGILNRFSDIAILDLLPLTIFDFIQLLLIVIGAIYVAVLQPYIFVATVPVI
VAFIMLRAYFLQTSQQLKQLESEGRSIFTHLVTSKGLWTLRAFGRQPYFETLFHKALNLHTANWFLYLSTLRWFQMRIEMIFVIFIAVTF
ISILTTGEGEGRVGIILTLAMNIMSTLQWAVNSSIDVDSLMSVSRVFKFIDMPTEGKPTKSTKPYKNGQLSKVMIIENSHVKKDDIWPSPG
GQMTVKDLTAKYTEGGNAILENISFSISPGQRVGLLGRGSGKSTLLSAFLRLLNTEGEIQIDGVSWDSITLQQWRKAFGVIPQKVFIFSGT
FRKNLDPYEQWSDQEIWKVADEVGLRSVIEQFPGKLDVFLVDGGCVLSHGKQLMCLARSVLSKAKILLLDEPSAHLDPVTYQIIRRTLK
QAFADCTVILCEHRIEAMLECCQFLVIEENKVRQYDSIQKLLNERSLFRQAISPSDRVKLPHRNSKCKSKPQIAALKEETEEVQDTRL

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