

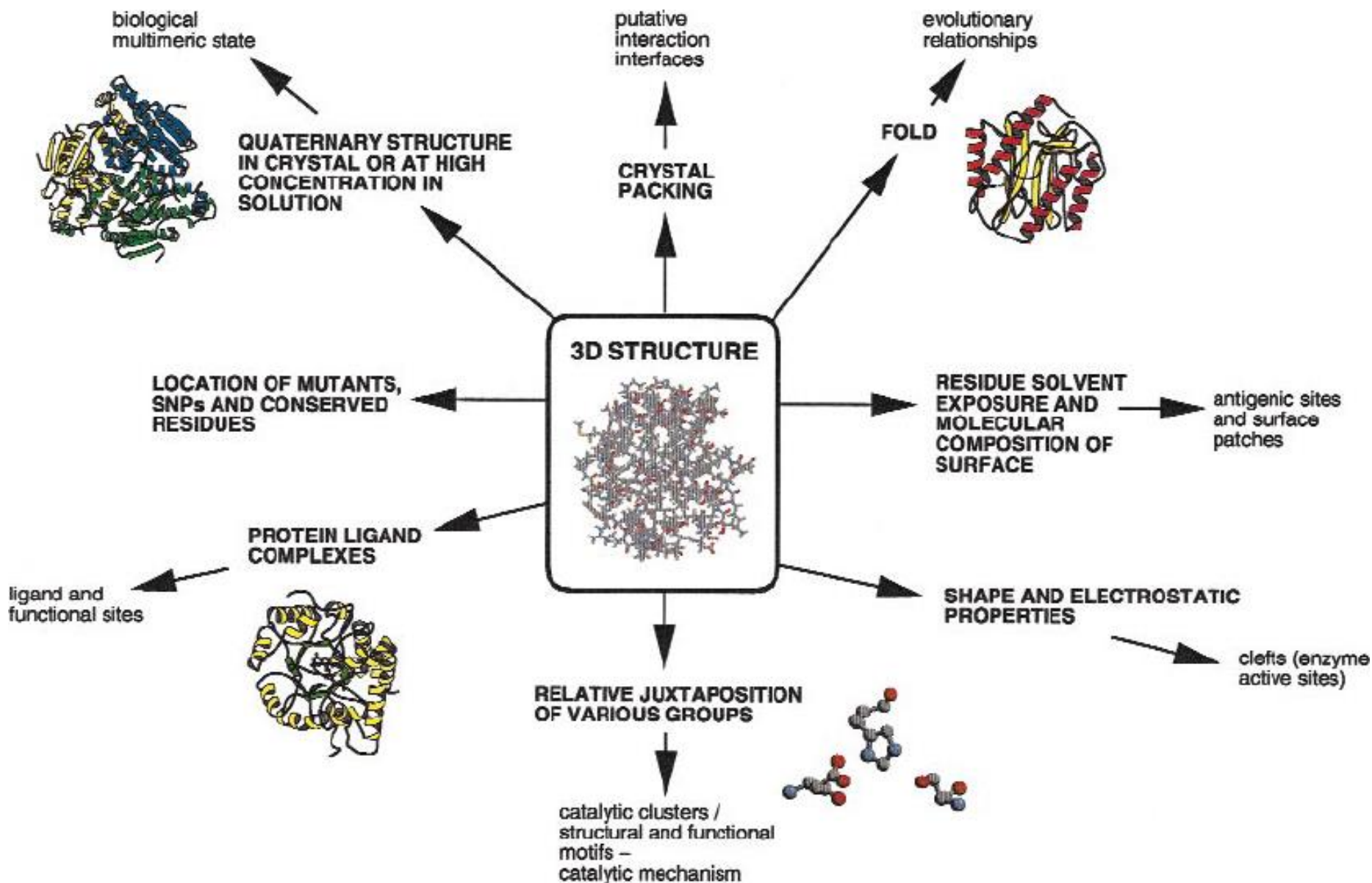
KFC/STBI

Strukturní bioinformatika

07_predikce vlastností
a
výzvy strukturní bioinformatiky

Karel Berka

Co vše lze zjistit z 3D struktury?



Syllabus

- Výpočty vlastností ze struktur
 - funkce (ProFunc, Gene3D)
 - 3D motivy
 - identifikace aktivního místa (QSiteFinder, Pocket Finder, SURFNET, ConSurf)
 - analýza tunelů (CAVER, MOLE, MoleOnline)
 - analýza mutability (Hotspot Wizard)
- Výzvy strukturní bioinformatiky
 - analýza pohybů (Molmovdb, PMG)
 - membránové proteiny (OPM)
 - neuspořádanost (DISPRED)
 - RNA (mfold)

Funkce

Definice funkce

EC – enzymatická klasifikace

(čtyři úrovně popisu)

- Oxidoreduktázy (EC 1)
 - substrát je oxidován jde o donor buď vodíku, nebo elektronů
 - jaký substrát
 - typ akceptoru
- jen pro enzymy
- 1 enzym – více funkcí (lze)
 - methylenetetrahydrofolate dehydrogenase/cyclohydrolase (EC 1.5.1.5 a 3.5.4.9)

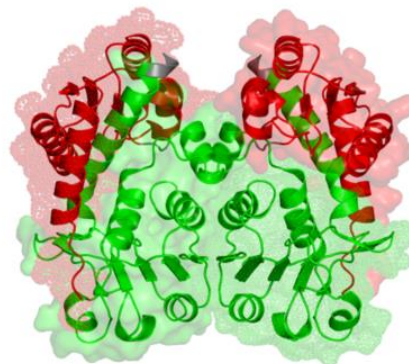
3ngx

GO - Gene Ontology

Anotace genomů

- Biologický proces
 - obecný (cell growth)
 - specifický (glykolýza)
- Molekulární funkce
 - obecný (enzym)
 - specifický (hexokináza)
- Buněčná kompartmentace
 - kde je gen aktivní

PDB



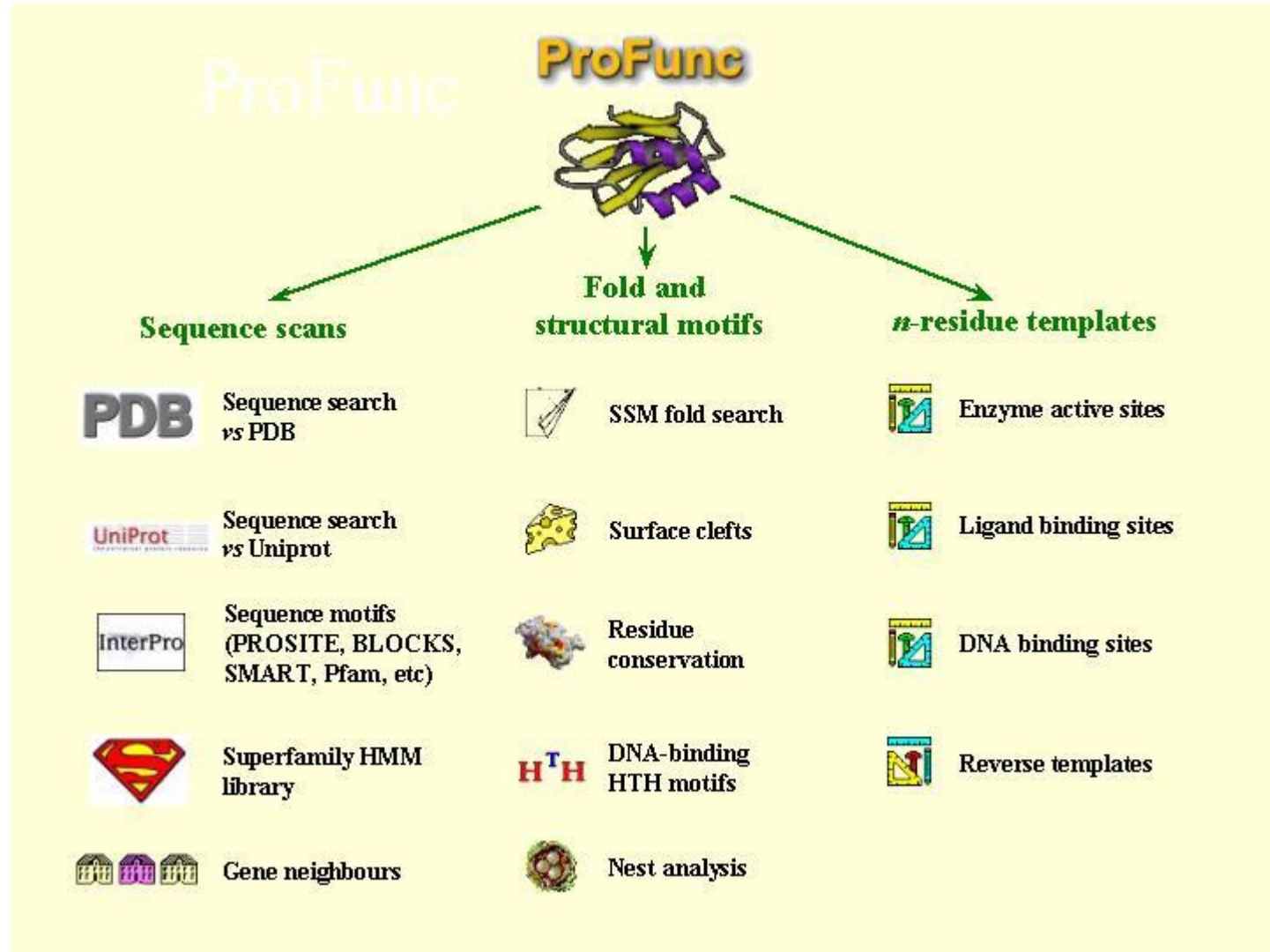
katalytická doména

NADPH vazebná
doména

3ngx

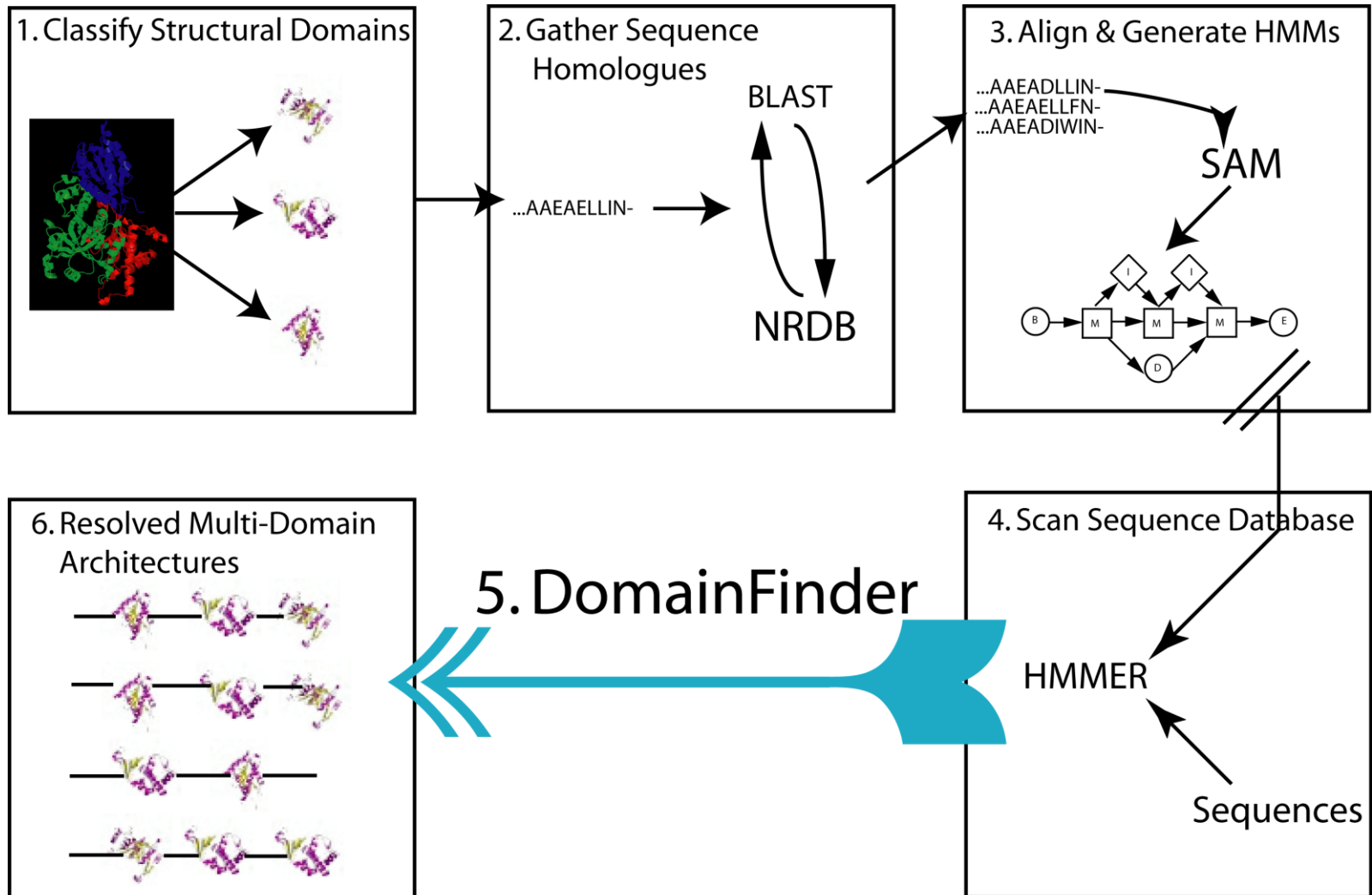
PDB

Hledání funkce - ProFunc



Gene3D

- anotace genomů pomocí informace z CATH



Hledání funkce - střípky

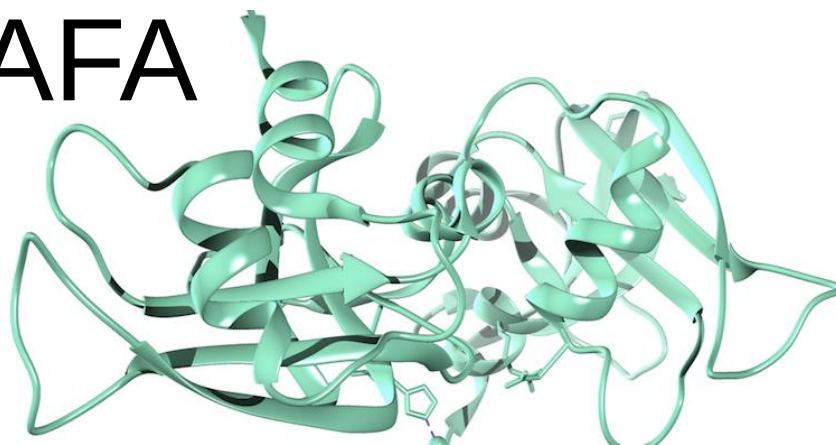
- Hledání v sekvencích
 - pod 40% identity bývá funkce zachována málo – jsou výjimky (CYP)
- Fold matching
 - porovnávání struktury proti databázi (CATH, SCOP) pomocí strukturního alignmentu (DALI, FATCAT)
- 3D motivy residuí
 - podobně jako v PDBe, (TESS, PDBSiteScan)
- Aktivní místa
 - geometrické podobnosti (SURFNET)
 - fyzikální vlastnosti (SURF'S UP!, SiteEngine)
 - ale stejný substrát může mít mnoho míst – ATP

Hledání funkce – střípky II

- Fylogenetická analýza
 - hledání konzervovaných residuí multiple alignmentem a zobrazení konzervovanosti na struktuře (ConSurf)
- DNA binding
 - Helix-turn-helix motiv (HTHquery)
- Ligandy
 - porovnávání pozic se známými pozicemi v krystalech v PDB (MSDsite)
- Anotace tunelů

**Protein
Function
Prediction**

CAFA



- Critical Assessment of Function Annotation
 - large-scale assessment of computational methods dedicated to [predicting protein function](http://biofunctionprediction.org/).
 - <http://biofunctionprediction.org/>

P. Radivojac et al A large-scale evaluation of computational protein function prediction. Nat Methods, 10(3):221–227, 2013.

3D motivy

hledání strukturních podobností

3D Motifs

- **Sekvenční motiv** je popis relativně konzervovaného lokálního **sekvenčního rysu**, který lze korelovat s funkcí

analogicky

- **Strukturní motiv** je popis relativně konzervovaného **strukturního rysu**, který lze korelovat s funkcí

Příklad:

- Aspartátové proteázy
 - Třída enzymů které štěpí polypeptidickou (hlavní řetězec) vazbu
 - Sekvenční analýza ukazuje, že tyto enzymy sdílí společný sekvenční rys - Asp,Thr,Gly aminokyseliny v aktivním místě
 - Kromě tohoto sekvenčního znaku sdílejí tyto proteázy množství fyzikálně chemických vlastností, které jim umožňují plnit stejnou funkci (záporný náboj v aktivním místě, uspořádání aktivního místa, apod.)

3D MOTIFS definice

V 1D případě provedeme alignment sekvencí a vytvoříme pravděpodobnostní model pro danou aminokyselinu v daném místě

Analogicky lze totéž provést pro strukturu. Určíme, jaká aminokyselina se vyskytuje v daném strukturním kontextu

- Aminokyseliny nemusí být nejvhodnější úroveň popisu (sekundární struktura může být lepší, je konzervovanější)

Možné deskripty

- Typ atomu (C,N,O,H,S...)
- Typ funkční skupiny (OH,CH₃,NH,C=O..)
- Typ aminokyseliny (Val,Tyr,Gly..) – nabitá, polární, nepolární, aromatická...
- Typ sekundární struktury nebo obecně geometrického parametru
- Fyzikální parametry (mobilita, elektronegativita, náboj, hydrofobicita)

Příklad 3D motivů

Sada struktur má více negativního náboje v místě kolem pozice x_1, y_1, z_1 , ale méně kladného náboje v okolí x_2, y_2, z_2

nebo

Sada struktur má výrazný nedostatek hydrofobních aminokyselin v této oblasti (x_3, y_3, z_3) ale je bohatá na polární aminokyseliny, významně Ser v oblasti x_4, y_4, z_4

PatternQuery

Find atom patterns in molecular databases

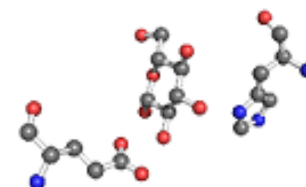


```
Rings(5*["C"] + ["O"]).ConnectedResidues(0).AmbientResidues(3)
```

Find a pyran ring
(5 carbons and an oxygen)

Identify the entire residue the
ring is in

Add all residues within 3Å



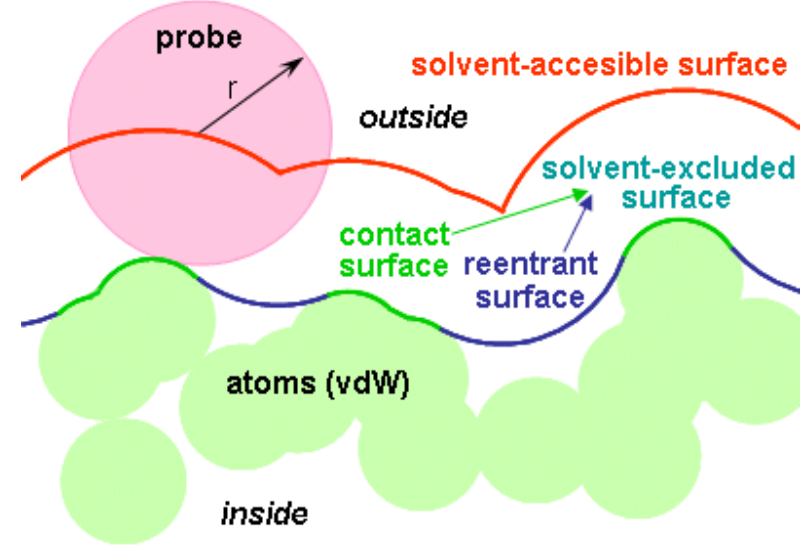
<https://webchem.ncbr.muni.cz/Platform/PatternQuery/Index>

Hledání aktivního místa

Aktivní místo

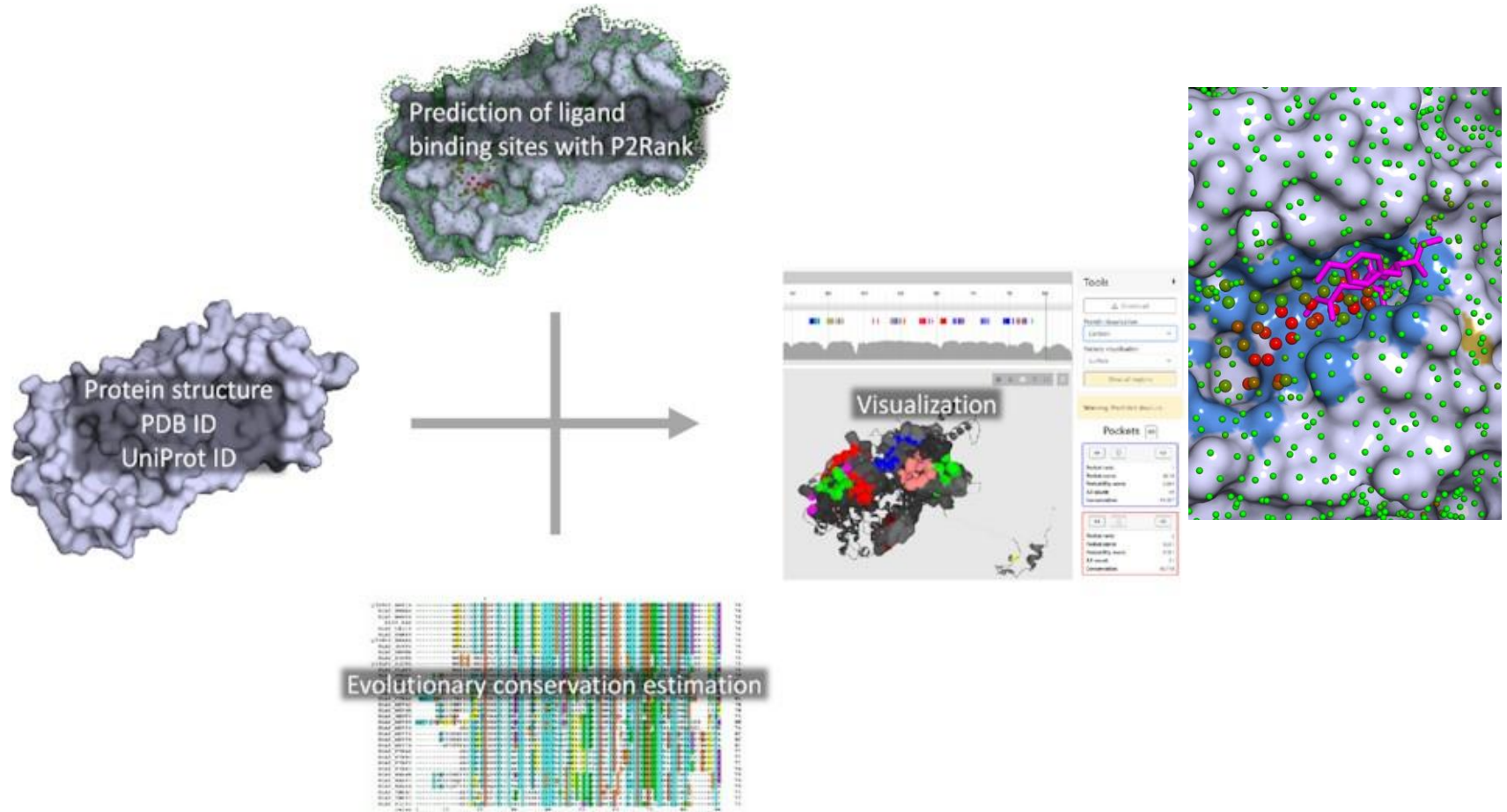
- tvar a velikost
 - velké (protein-protein, ev. protein-DNA)
 - PPI-PRED
 - malé (záhyby, ligandy)
 - na povrchu, ev. hluboko a spojené kanály
 - SURFNET, PocketFinder, Q-SiteFinder
- fyz chem vlastnosti
 - hydrofobic patches
 - nabité povrchy
- konzervace residuí – evolution trace method
 - ConSurf
- kombinace metod – meta searches

Hledání kapes



- SAS a SES
 - solvent accessible surface
 - solvent excluded surface
- hledání kapes
 - Pocket-Finder
 - na základě geometrie receptoru na řezu na 7 liniích v mřížce
 - uzavřené body – kavity
 - částečně uzavřené – kapsy (pockets)
 - popisuje je jen, když je počet bodů vedle sebe > threshold
 - QSiteFinder
 - na gridu
 - místa s výhodnými interakčními energiemi k probe představující generický ligand (methan)

P2Rank - PrankWeb



Each point represents its local chemical neighborhood –

- predicted ligandability score (0 = green to 1 = red)
- clustered to form predicted binding sites

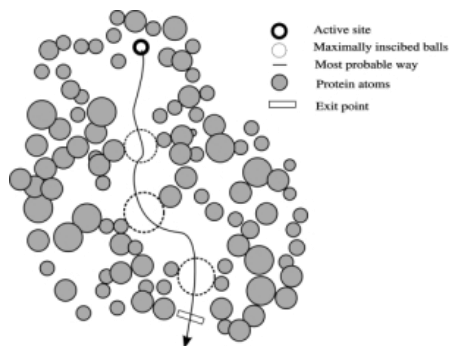
- <https://prankweb.cz/>
- <https://github.com/cusbg/p2rank-framework>

Hledání cesty do aktivního místa

tunely, kanály a póry

Hledání tunelů

- Caver



Dijkstrův algoritmus
na vyhledání optimální cesty
na gridu
nověji na Voronoi mesh

CAVER

Manual :: FAQ :: Contact

News
21/10/2010 15:33:28
CAVER in Nature Methods

Main
Online
Download
Concept
Screenshots
Examples
Authors
Links

Dijkstra's algorithm

CAVER - secure caving in the world of biomolecules ...

CAVER provides rapid, accurate and fully automated calculation of pathways leading from buried cavities to outside solvent in protein structures. Study of these pathways is important in drug design and molecular enzymology.

CAVER facilitates analysis of any molecular structure including proteins, nucleic acids, inorganic materials, and is available as online version or PyMOL plugin.

CAVER Viewer enables efficient visualization of calculated pathways and can be run online using Java Web Start technology.

www.combinatorica.com

© 2008 Petr Haimb

<http://loschmidt.sci.muni.cz/caver/index.php>

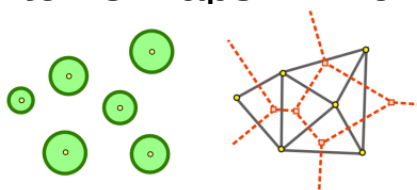
Petrek, M. et al - **CAVER: a new tool to explore routes from protein clefts, pockets and cavities**, *BMC Bioinformatics* 2006, 7:316

Beneš, P. et al. - **CAVER 2.1 software**, 2010.

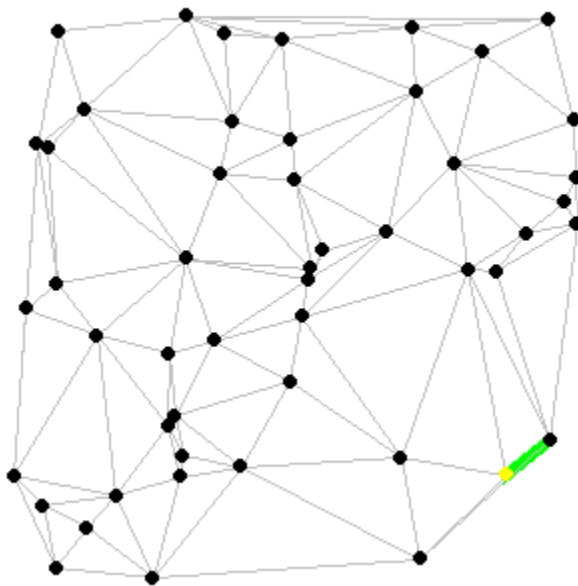
Hledání tunelů

MOLE

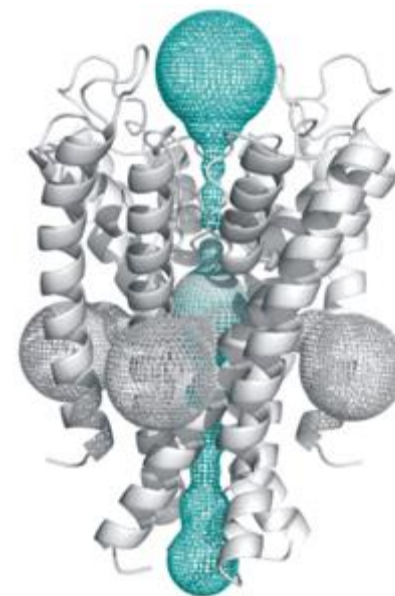
Atoms maps => Voronoi diagram



Dijkstra's algorithm



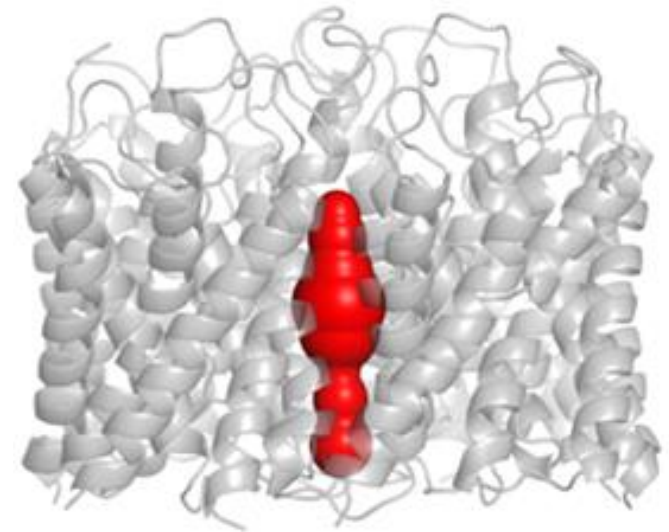
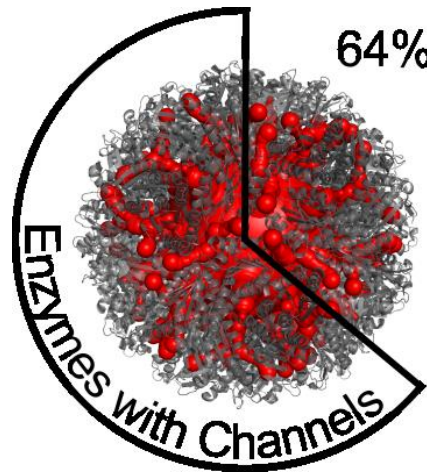
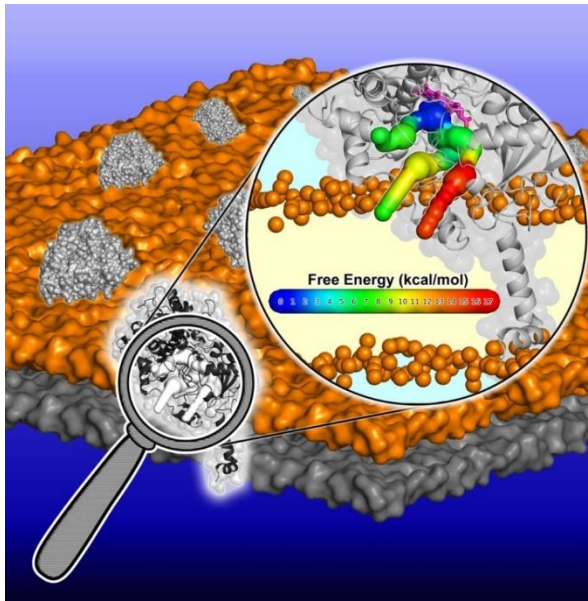
www.combinatorica.com



<http://mole.chemi.muni.cz/online/current/>
<http://mole.upol.cz>

Porous Pathways in Proteins

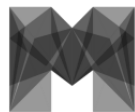
- Channels/Tunnels
 - Connect active site with exterior
- Pores
 - Running through the structure



- Pathways have important function of gate-keeping

- Palonciová M, Navrátilová V, Berka K, Laio A, Otyepka M: Role of Enzyme Flexibility in Ligand Access and Egress to Active Site – Bias-Exchange Metadynamics Study of 1,3,7-Trimethyluric Acid in Cytochrome P450 3A4 *J. Chem. Theory Comput.*, 12(4), 2101–2109, 2016.

- Pravda L, Berka K, Svobodová Vařeková R, Sehnal D, Banáš P, Laskowski RA, Koča J, Otyepka M: Anatomy of enzyme channels. *BMC Bioinf.*, 15(1), 379, 2014.



OLEonline



<https://mole.upol.cz>

- Detection and analysis of channels, tunnels and pores

Home ChannelsDB LiteMol viewer
PDBe Reports Help form

Current selection
List of Channels
Interactive origins
Surface, Cavities
Membrane position
Sequence
Submissions details
Mode switch
Parameters

Channel profile
Channel properties
Layer properties
Submit
Submissions and ChannelsDB data

Selection
Path (P1C0), Length: 50.9 Å
Channels
- Paths (1) All None
Path (P1C0), Length: 50 Å
Origins
Computed
Cavities
+ Surface (1)
+ Cavities (1)
Membrane
Membrane

Protein Sequence
Chain A
SEFKKKLFWRAVVAEFLATTLFVFISIGSALGFKYPVGNNGTAVQDNVVKVSLAFGLSIATLAQSVGHI SGA
HLNPAVTLGLLLSCQISIFRALMYIIAGCVGAIVATAILSGITSSLTGNSLGRNDLADGVNSQGGGLGIEII

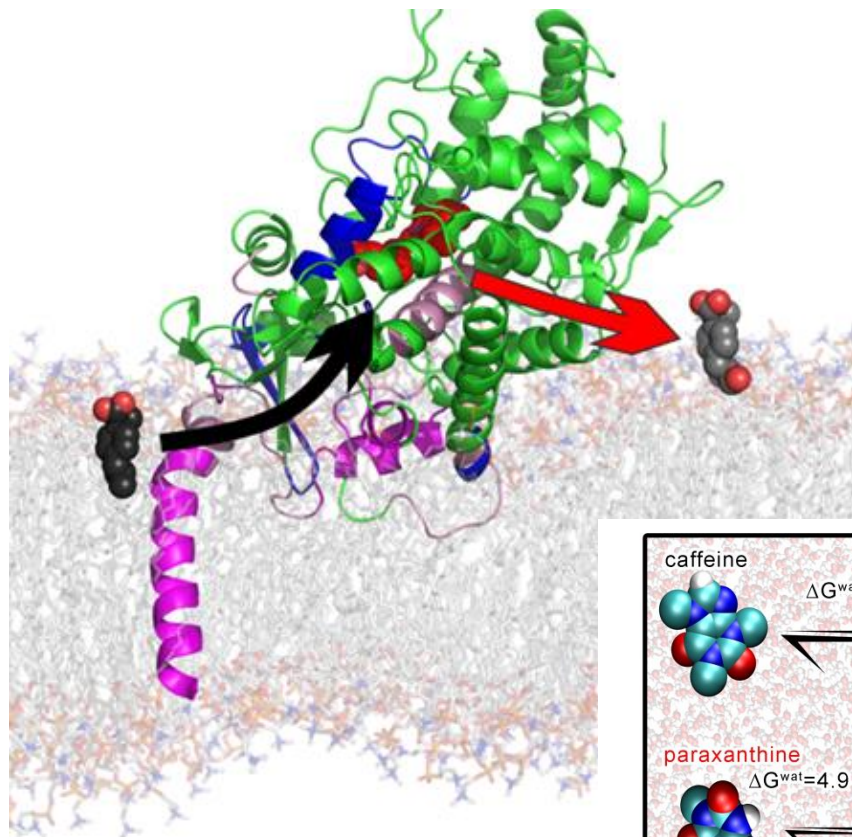
Channel profile (P1C0) Channels properties
Tunnel radius: Radius Color by: Charge Color bounds: Absolute Export
11 Å 0 Å 50.9 Å
Tunnel radius: Radius Color by: Hydrophobicity
11 Å 0 Å 50.9 Å

Layer Lining residues
Phys. Chem. Properties
Residue
VAL 80 Å
LEU 54 Å
Property Value
Hydropathy 4.12
Polarity 0.13
Hydrophobicity 1.13
Mutability 89

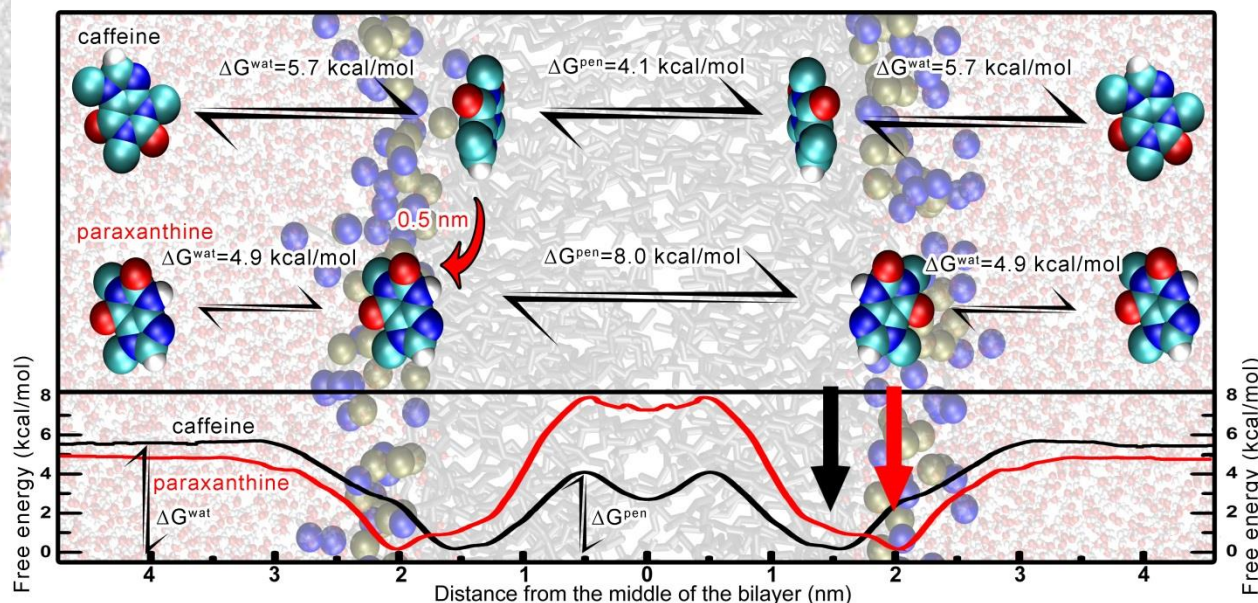
Submission settings Submissions
Pores
Beta Structure x Off
Membrane Region ✓ On
Specific Chains A, B, ...
Probe Radius 13
Interior Threshold 0.8
Submit Kill Delete Clear < 2 >

Application: Drugs with ochromes P450

- Deeper in the membrane
- Higher lipids affinity
- Metabolites
- Easier membrane escape



- Drugs
 - Access tunnel
- Metabolites
 - Egress tunnel



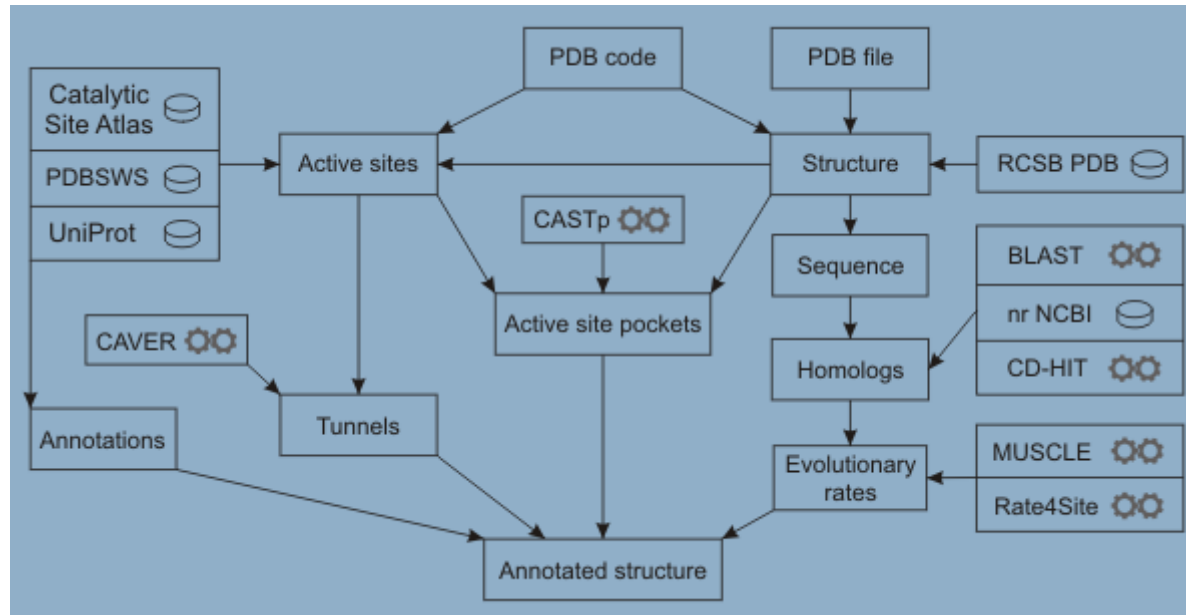
Paloncýová M, Berka K, Otyepka M: *J. Phys. Chem. B* **2013**, 117, 2403–10.

Berka K, Paloncýová M, Anzenbacher P, Otyepka* M: *J. Phys. Chem. B*, **2013**, 117(39), 11556–11564.

Berka K, Hendrychová T, Anzenbacher P, Otyepka M: *J. Phys. Chem. A* **2011**, 115, 11248–11255.

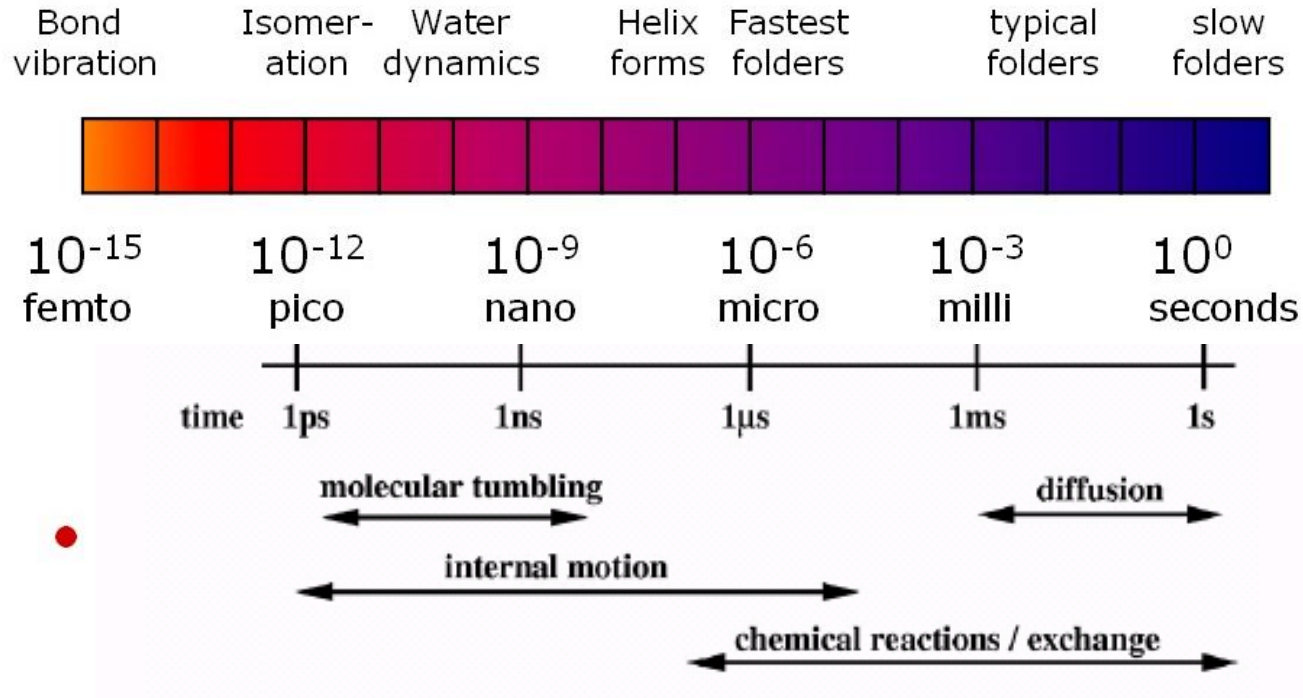
Hledání mutability

- Hotspot Wizard
 - hledání aktivních míst pro případné enzymové inženýrství
 - metasever



Analýza pohybů

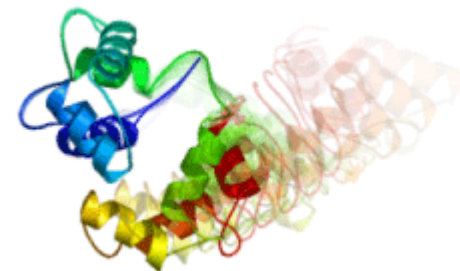
Časová škála biochemických pochodů



- **16 order of magnitude range**
 - Femtosecond timesteps
 - Need to simulate micro to milliseconds

Analýza pohybů

- MolMovDB
 - Flexibility and Geometric Analysis
- PMG
 - Protein Movie Generator



Intrinsically Disordered Proteins

- DisprotCentral



MobiDB



PED³: Protein Ensemble Database

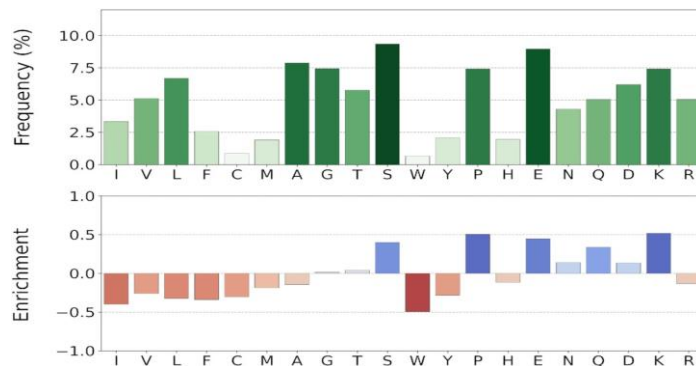
ELM

Disorder

- neviditelný – nejsou k němu strukturní data
- odhady – teplotní B-faktor v Xray, NMR ensemble, DISOPRED
- intristically disordered proteins – nepotřebují strukturu k tomu, aby fungovali
- často regulační funkce
 - molecular recognition (promiskuitní)
 - molecular assembly (virální kapsidy)
 - protein modification
 - entropic chain activities (pružiny, entropické hodinky)

Disprot

- Database of disorder in proteins
- Amino acid composition of DisProt disordered regions.
- sorted by the Kyte-Doolittle hydrophobicity scale.
- enrichment is calculated and normalized over the TrEMBL database frequencies (2021_03).



DisProt Database of protein disorder

Created: 9 Aug 2016 Last Update: 22 Sep 2016 [Download](#)

DP00086 - Cellular tumor antigen p53

Organism: Human

Taxonomy: Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorhini > Catarrhini > Hominoidea > Homo

Synonyms: P53_HUMAN; Antigen NY-CO-13; Phosphoprotein p53; Tumor suppressor p53

Cross-references: UniProt (P04637) MobiDB (P04637)

WARNING! This entry contains ambiguous evidences [AMBLIT](#) [AMBEXP](#) [AMBSEQ](#)

Functional Annotation

Molecular function of disorder (6) Activator (6) cis-regulatory elements (inhibitory modules) (1) Molecular recognition - assembler (1) Ubiquitination (1)

Molecular transition (10) Disordered state (7) Disorder to order transition (3)

Molecular partner (2) Protein-protein binding (1) Protein-DNA binding (1)

Disorder Overview

Sequence

DisProt confident DisProt analog MobiDB PDB MobiDB prot Plan

Legend: DisProt Disorder (red), DisProt Context-dependent (blue), PDB Disorder (orange), Predicted Disorder (yellow), PDB Ambiguous (grey), Predicted Structure (light blue)

Disorder Region Details

Color by: Evidences Molecular function Type of molecular transitions Molecular partner ☐ Hide ambiguous evidences

Sequence

Legend: Primary detection method (red), Secondary detection method (blue), N/A (grey)

Region Evidences

Showing 11/11 regions [Show all regions](#)

Region position: 320 - 393 Detection method: X-Ray Crystallography Curator: Marco Necci

Region Sequence: 320 KKPLDGEYPTLQIGRERFEMFELNEALELKAQACQEPGOSRAHSHLKKKGQSTSRHKKLMFKTEGPDSD

Molecular Function: cis-regulatory elements (inhibitory modules)

Molecular Transition: Disorder to order transition

Molecular Partner: Protein-protein binding

PubMed: Kannan S, Lane DP, Verma CS Long range recognition and selection in IDPs: the interactions of the C-terminus of p53. Sci Rep. vol. 6, pp. 23750, 2016. PMID: 27030593

Region position: 1 - 93 Detection method: Small-Angle X-Ray Scattering (Saxs) Curator: Marco Necci

Region Sequence: 1 MEEQSDPSVEFPLSQETFSDLKLLFENNVLPLFSQMDLMLSPDIEQWTFPGDEAPRMEAPFPVAPAPAP 81 TPAAPAPAPSWFL

Molecular Function: Activator

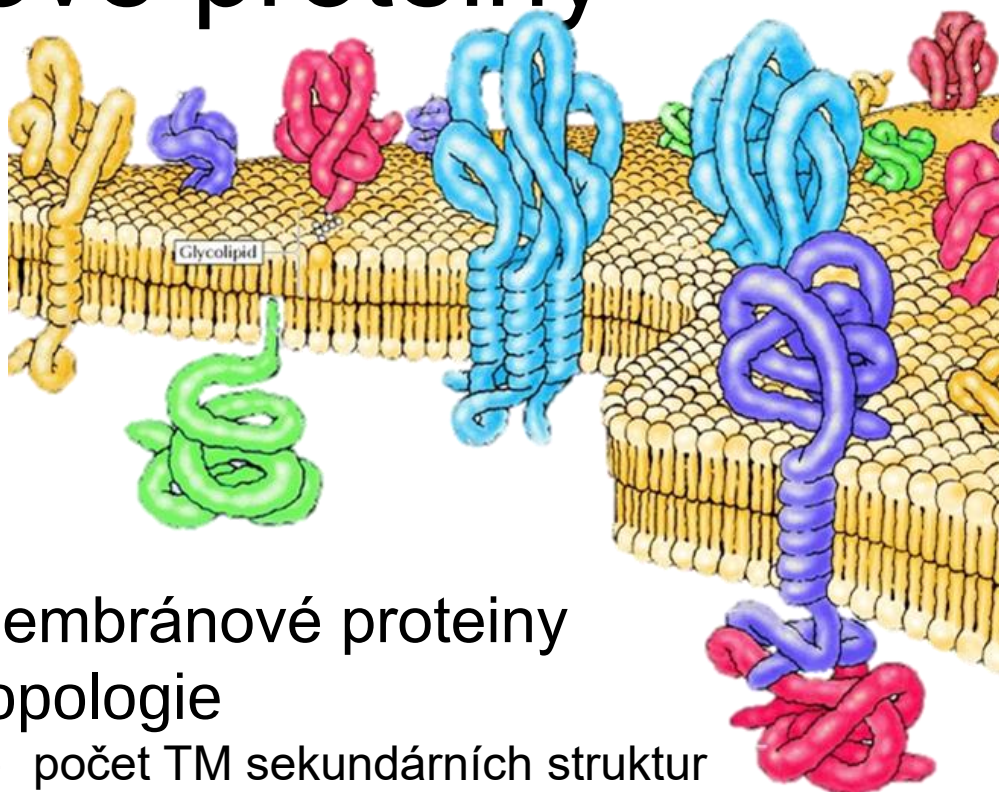
Molecular Transition: Disordered state

PubMed: Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, Svergun DI, Blackledge M, Fersht AR Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transcription domain. Proc. Natl. Acad. Sci. 113:4, vol. 115, pp. 1154-1159, 2016. PMID: 26800000

- *Nucleic Acids Res*, Volume 50, Issue D1, 7 January 2022, Pages D480–D487, <https://doi.org/10.1093/nar/gkab1082>
- *Nucleic Acids Res*, Volume 45, Issue D1, January 2017, Pages D219–D227, <https://doi.org/10.1093/nar/gkw1056>

Membránové proteiny

Membránové proteiny

- Buněčná membrána
 - lipidická dvojvrstva + proteiny
 - Buněčná stěna
 - rigidní vrstva vně buněčné membrány chránící buňku
 - Vnější membrána
 - další membrána – mitochondrie, chloroplasty, Gram-negativní bakterie
 - Membrane core
 - hydrofobní část membrány
 - Membrane interface
 - polární oblasti u povrchu membrány
 - Rovina x Osa membrány
- 
- Membránové proteiny
 - Topologie
 - počet TM sekundárních struktur
 - lokalizace N-terminálního konce (IN x OUT)
 - re-entrant loops (neprojdou skrz membránu)
 - Positive inside rule
 - Arg a Lys jsou až 4x častější v cytosolu oproti periplasmě
 - Lily-pad effect
 - interakce s jádrem membrány – častější aromatické residua – Tyr, Trp

Molecular Transport on Membranes

- Majority of drugs has target on membrane

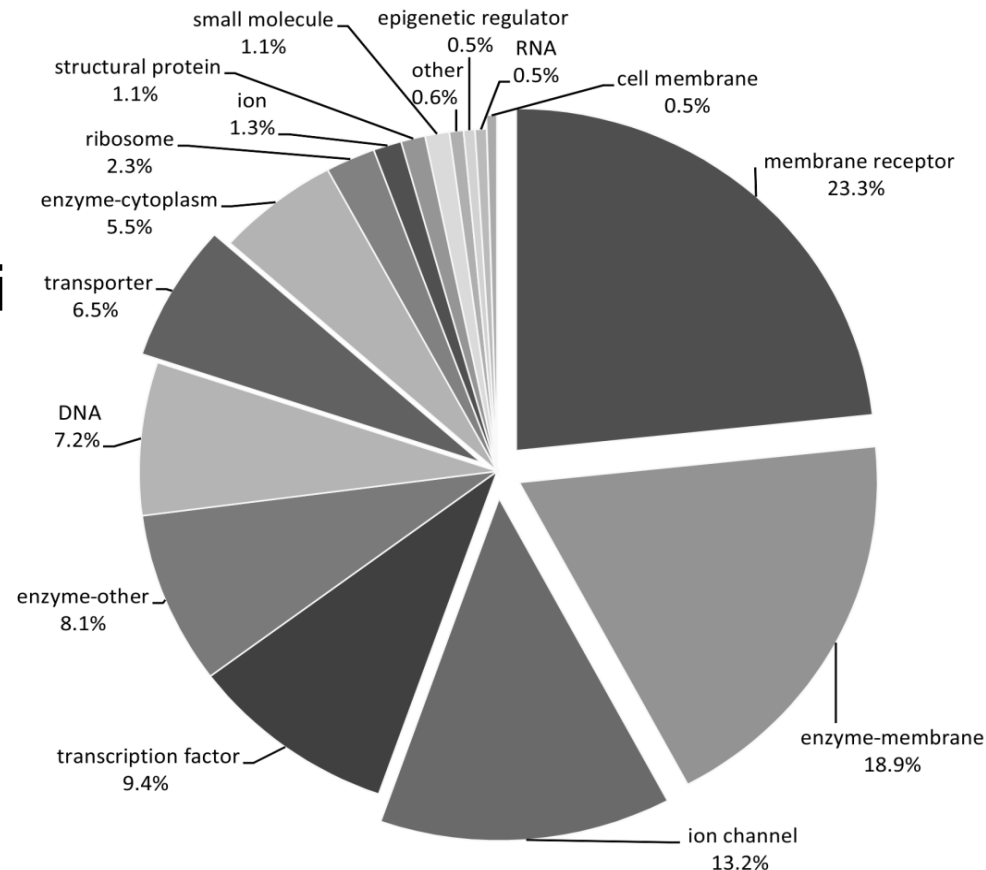
- Permeation

- **Passive transport**
 - **Protein-assisted passive transport**

- Active Transport

- Accumulation

- Metabolism
 - Activity
 - Toxicity



Di Meo F, et al *Pharmacol. Res.*, 111, 471–486, 2016.

Membránové proteiny

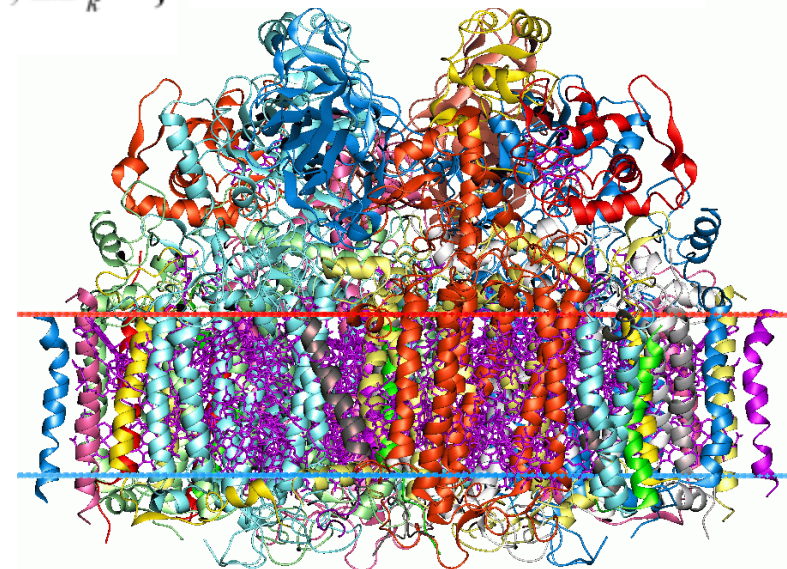
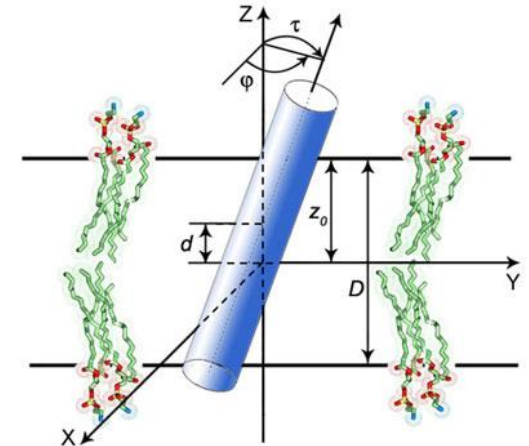
- málo proteinů v PDB skutečně v membráně (300 v MPDB.org) (doména EM), většinou solubilizované (10000 v PDB)
- extrémně zajímavé jako receptory pro léčiva (GPCRs)
- predikce membránových částí ze sekvence (TMpred)
- predikce struktury málokdy (Swiss-Model-7TM)

Zanoření v membránách

- OPM server
- protein = rigid body

$$\Delta G_{transf} = \sum_{i=1}^N \sigma_i^{wat \rightarrow bil}(z_i) ASA_i + \sum_{j=1}^M \eta_j^{wat \rightarrow bil}(z_j) \mu_j + \sum_{k=1}^L \min \{ \Delta E_k^{ion}, \Delta E_k^{neutr} \}$$

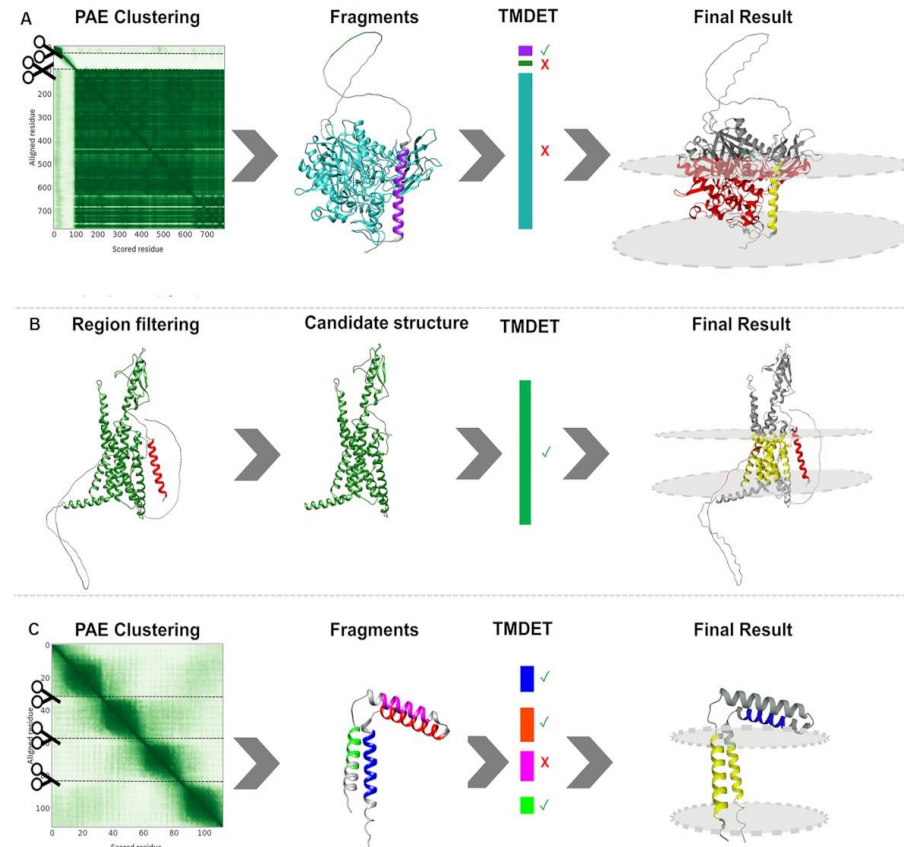
- optimalizace v
anisotropickém
implicitním modelu
membrány



fotosystém II

TmAlphaFold

- **ASAH_HUMAN (Q9NR71)**
 - sliced into domains based on the Predicted Alignment Error matrix;
 - 2 domains, and flexible linker
 - TMDet detected the membrane plane in the helical domain
 - => TM helix, not compatible with membrane
- **PTH2R_HUMAN (P49190)**
 - signal peptide and low reliability regions masked out
 - TMDet detects the membrane plane in the structure
 - => TM domain, masked out highlighted
- **TM14C_HUMAN (Q9P0S9)**
 - sliced by Predicted Alignment Error matrix
 - 4 helical fragments
 - TMDet detected 3 membrane helices
 - extra TM segment is highlighted.
 - Exp. (2los_A): one TM segment is erroneously predicted to lie parallelly to the membrane plane.



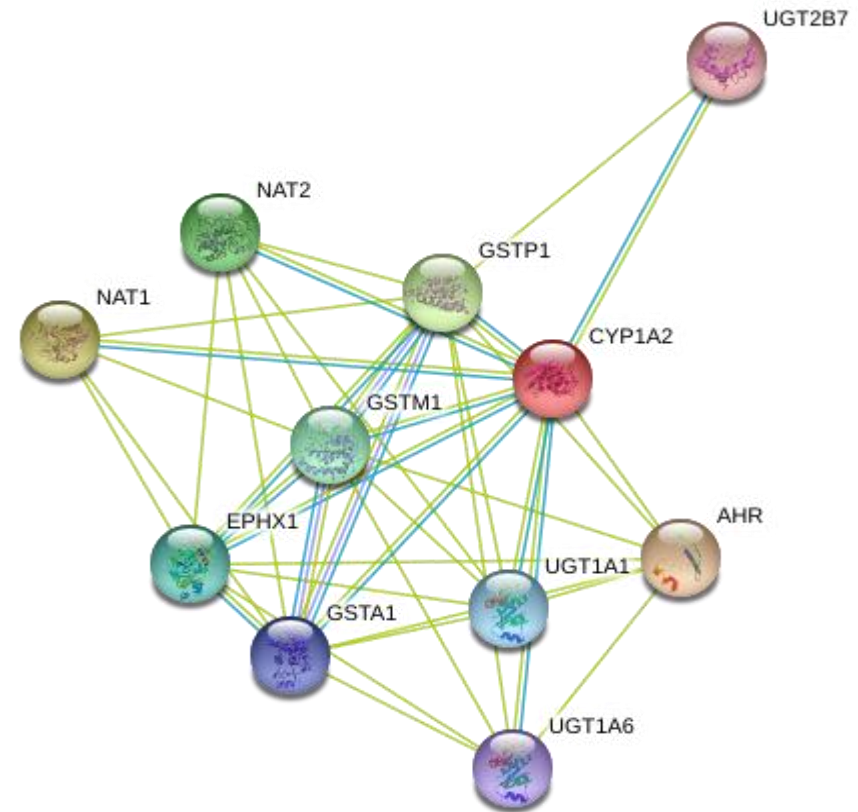
• *Nucleic Acids Res*, gkac928, <https://doi.org/10.1093/nar/gkac928>

PPI

Protein-protein interakce

Hledání protein-protein interakcí

- bouřlivě se rozvíjející oblast
 - hledání léčiv blokuujících interakce
- založená hlavně na evolučních analýzách sekvencí a porovnávání dat ze známých komplexů a dalších experimentů
- na pomezí systémové biologie



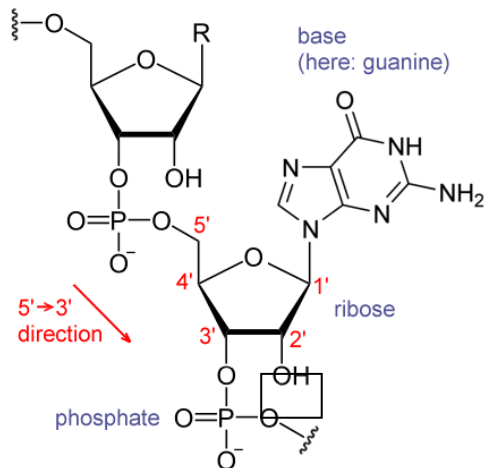
STRING DB – interakce mezi proteiny
<http://string-db.org>

Nukleové kyseliny

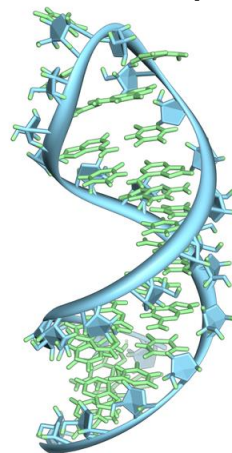
aneb na co se v mezičase
nedostalo

RNA/DNA

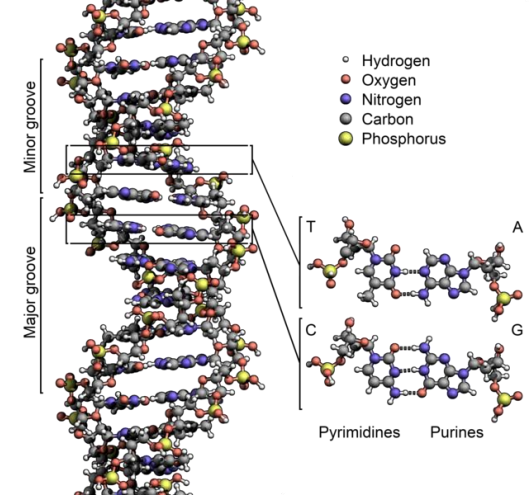
- RNA world <http://www.imb-jena.de/RNA.html>
 - hypotéza prvotního RNA světa
- Rfam 10.0
 - collection of RNA families
 - multiple sequence alignments, consensus secondary structures and covariance models (CMs)



RNA hairpin

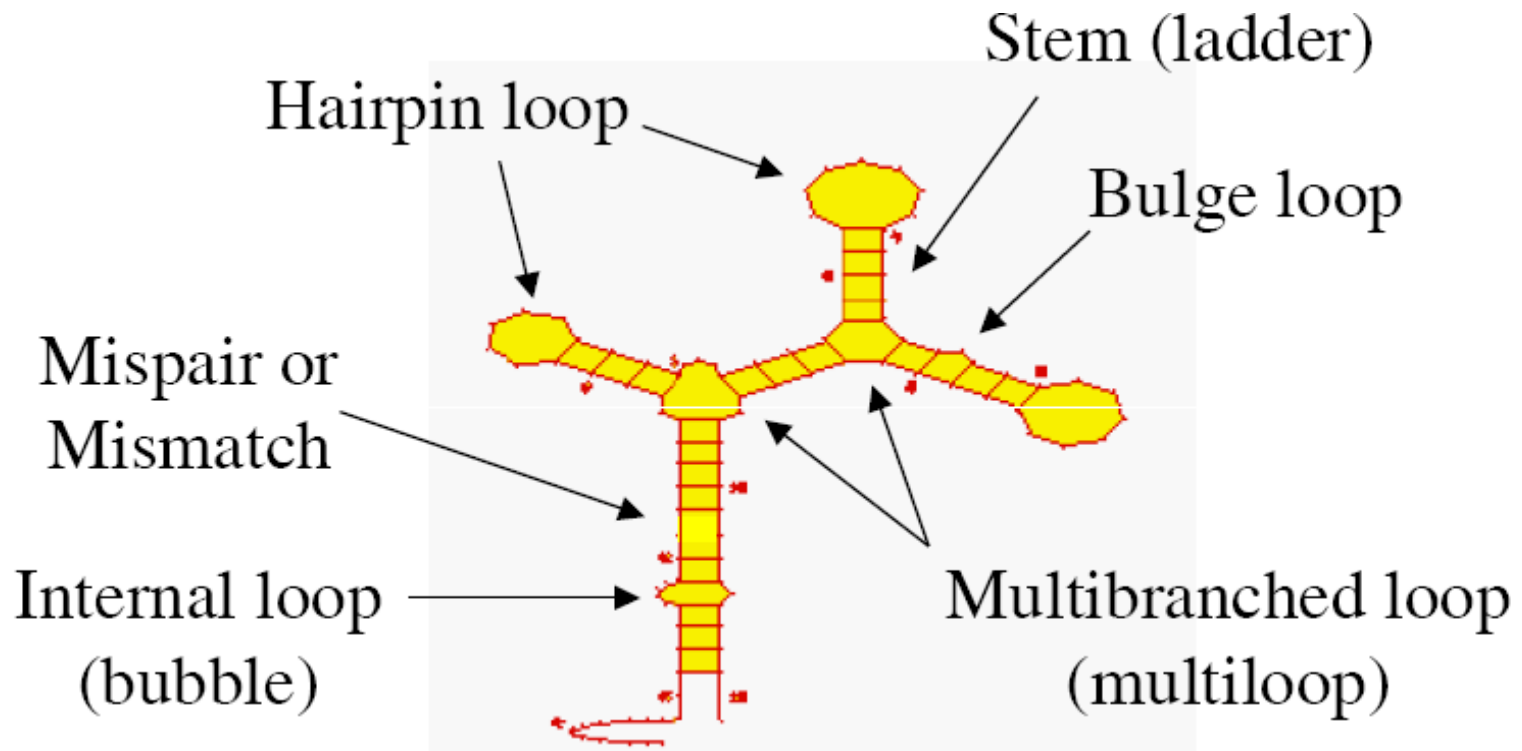


DNA šroubovice



RNA struktura

- Terminologie



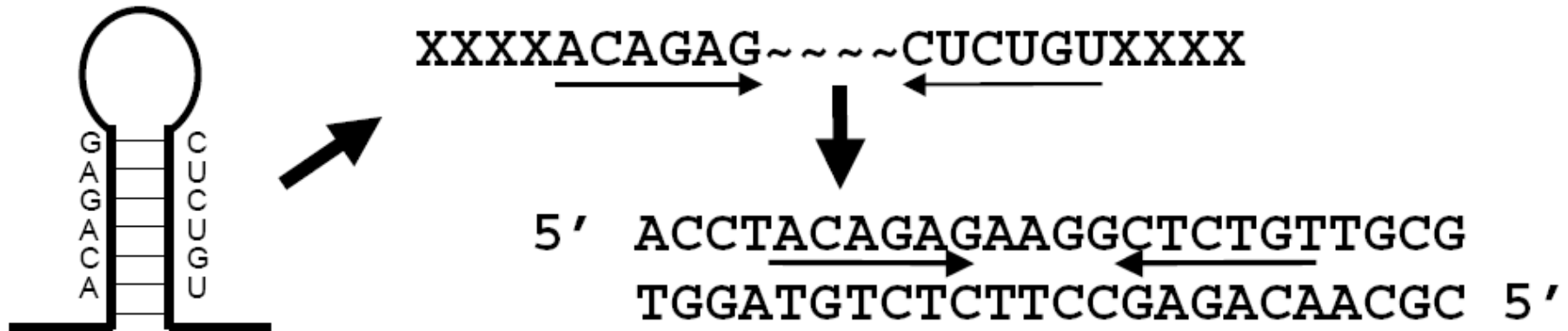
Dotplot

- **Paired regions can be seen in dotplots**

- Use scores for complementary bases or plot vs reverse-complement sequence

- **Use a windowing method to scan for stems**

- Window about the length of shortest stem
- Can include crude energetics, GC=3 AU=2 GU=1



Způsoby RNA predikce

- Naivní (stemloop)
- Zuker algoritmus
- s pseudoknoty
- s suboptimálními řešeními

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>

<http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form>

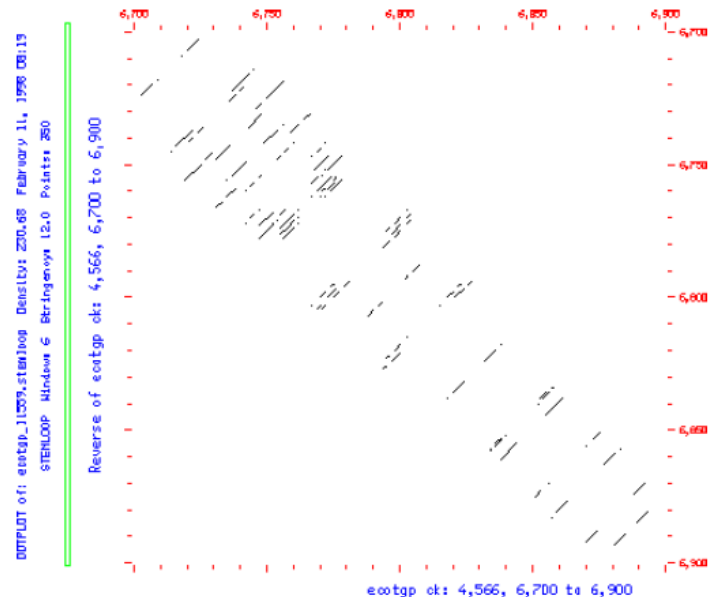
Naivní (STEMLOOP)

1. Calculates score over a window
 2. Finds stems over a threshold score
 3. Minimum/maximum loopsize
 4. Sort by position or score
- + Shows all stems not just lowest energy
- Energetics are very crude
 - No bulges or bubbles
 - Complex optimization problem

Minimum Stem: 6 Minimum bonds/stem: 8.00 Maximum loop size: 20
 Stems found: 72 Stems shown: 25
 Average Match: 1.80 Average Mismatch: 0.00 Nibbling Threshold: 0.50

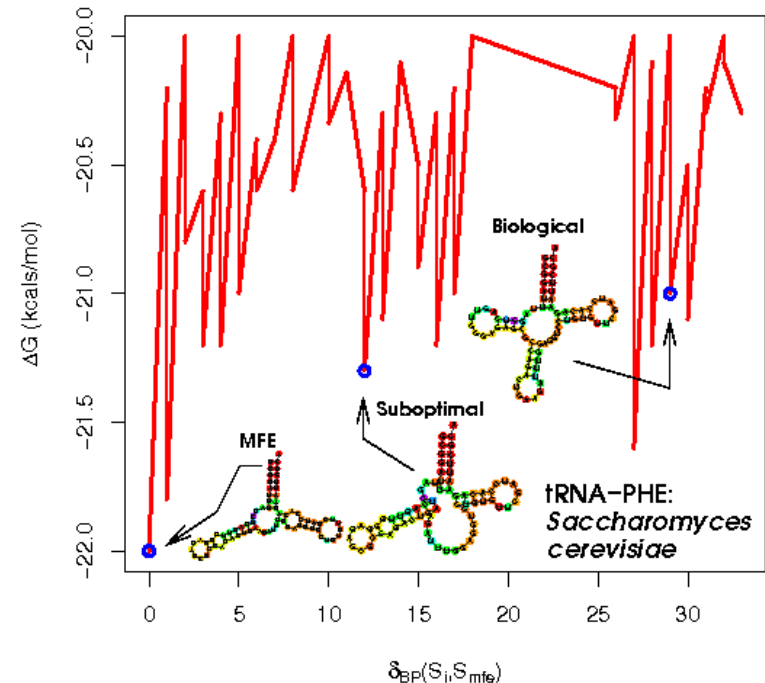
```

6850 CCGCCCCGATCAGGTAAAAG CAG      21, 30.0
      | | | | | | | |
6897 GCGGGCGTCGAGGACGTAGTT AGC      6
      | | | | | | | |
6823 CACCTCCATTGCAGGGATTGG TATT     22, 29.0
      | | | | | | | | T
6875 GCGACGAAAATGGACTAGGCCCG CC      9
      | | | | | | | |
6841 TTGGTATTTCCGCCCCGATC AGGTAA    21, 28.0
      | | | | | | | | A
6895 GGGCGTCGAGGACGTAGTTAG CGACGA   13
      | | | | | | | |
6823 CACCTCCATTGCAGGGAT TT          18, 26.0
      | | | | | | | | G
6863 GACTAGGCCCGCCTTTA TG           5
      | | | | | | | |
6829 CATTGCAGGGATTGTGTAT TTCC       19, 25.0
      | | | | | | | | G
6875 GCGACGAAAATGGACTAGG CCCC       9
  
```



RNA energetika

- RNA je dynamická, hodně struktur s podobnou energií
- citlivé na podmínky (sůl, teplota, proteiny)
- pseudoknots
- biologická struktura nemusí být ta nejnižší, ale musí být jedna z hlavních
- hodně napoví fylogenetická analýza



$$\Delta G = \Delta G_{stack} + \Delta G_{bulge} + \Delta G_{hairpin} + \Delta G_{internal} + \Delta G_{multibranch}$$

- ΔG_{stack} je energie párování a stakování ve stemech, vše ostatní je pro loopy
- Jen ΔG_{stack} stabilizuje sekundární strukturu, ostatní ji destabilizují

RNA energetika

- stacking
(kcal/mol)

	A: U	C: G	G: C	U: A	G: U	U: G
A: U	-0.9	-2.1	-1.7	-0.9	-0.5	-1.0
C: G	-1.8	-2.9	-2.0	-1.7	-1.2	-1.9
G: C	-2.3	-3.4	-2.9	-2.1	-1.4	-2.1
U: A	-1.1	-2.3	-1.8	-0.9	-0.8	-1.1
G: U	-1.1	-2.1	-1.9	-1.0	-0.4	-1.5
U: G	-0.8	-1.4	-1.2	-0.5	-0.2	-0.4

G:U base pairs are OK!

- Tetraloops

Exceptionally common 4 base long loops

>60% of loops in rRNA are

AUUU

CUUG

GAAA GAGA GCAA GCGA GGAA GUGA GUAA

UACG UCCG UUCG UUUA

Clearly more stable but exact energy unknown

Zuker gives -2 kcal/mol

Zuker algorithmus

- Calculation proceeds from center towards edges
- Includes stacking, bulge, internal, and hairpin loop terms
- Start from center because the center line is location of hairpin loops

mfold

nearest neighbor energy rule:

free energies are assigned to loops rather than to base pairs.
These have also been called *loop dependent* energy rules.

A secondary structure, **S** on an RNA sequence, $R=r_1, r_2, r_3, \dots, r_n$,
is a set of **base pairs**.

A base pair between nucleotides r_i and r_j ($i < j$) is denoted by $i.j$.

A few constraints are imposed.

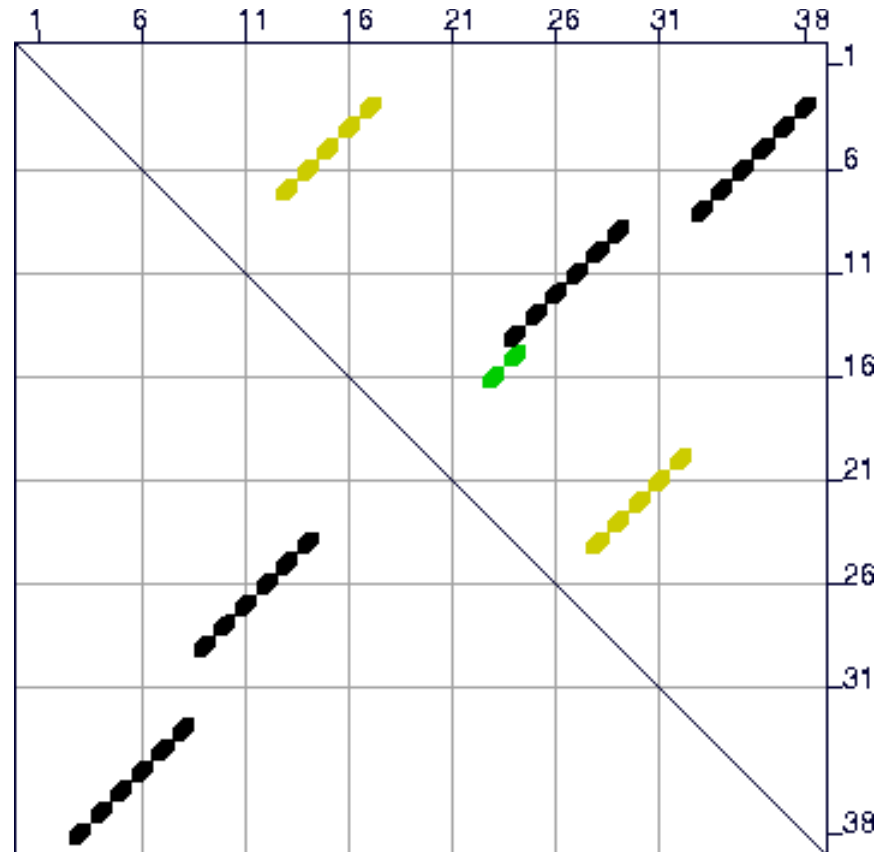
- Two base pairs, $i.j$ and $i'.j'$ are either identical, or else $i \neq i'$, and $j \neq j'$. Thus base triples are deliberately excluded from the definition of secondary structure.
- Sharp U-turns are prohibited. A U-turn, called a hairpin loop, must contain at least 3 bases.
- Pseudoknots are prohibited. That is, if $i.j$ and $i'.j' \in S$, then, assuming $i < i'$,
• either $i < i' < j' < j$ or $i < j < i' < j'$.

EXAMPLE 1

The *energy dot plot* is an integral part of the folding prediction. Consider the folding of a short RNA sequence:

AAGGGGUUGG UCGCCUCGAC UAAGCGGCUU GGAAUUCC

The *energy dot plot* for the ``Example 1'' sequence. Surrounding annotation, which would not be legible at this scale, has been removed. The yellow dots indicate base pairs in foldings within 0.3 kcal/mole of the optimal folding free energy of -9.8 kcal/mole

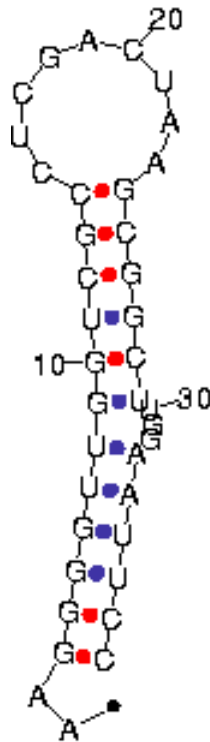


The 2 predicted foldings for the ``Example 1'' sequence.

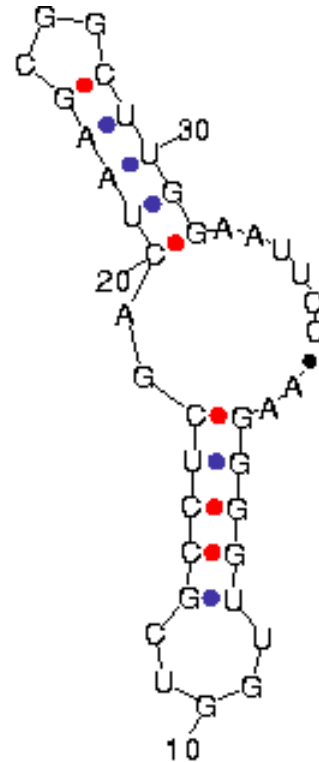
(a) The optimal folding with $\Delta G = -9.8\text{kcal/mole}$.

(b) (b) The suboptimal fold ($\Delta G = -9.5\text{kcal/mole}$) found

(c) after refolding with $W=0$.



(c)



(f)

mfold

- RNA secondary structure with suboptimal folding
- Display results as mountains, domes, circles, squiggles Zuker's web site (includes server)
- <http://bioinfo.math.rpi.edu/~zukerm/rna/>
- Calculate energies for specified structure (efn server)
- ssDNA structure prediction

Eterna

- predikce stabilní 2D struktury RNA

