

Predikce trojrozměrných struktur proteinů



VYSOKÁ ŠKOLA
CHEMICKO-TECHNOLOGICKÁ
V PRAZE



Vojtěch Spiwok
spiwokv@vscht.cz

Ústav biochemie a mikrobiologie
VŠCHT, Praha

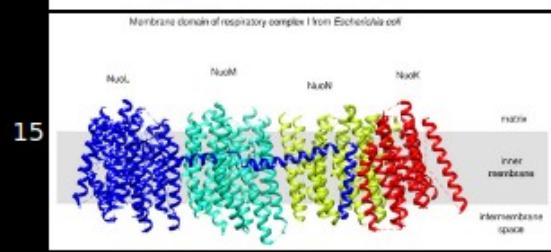
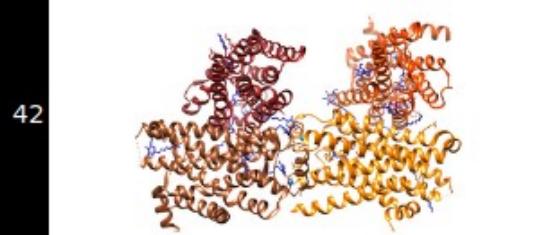
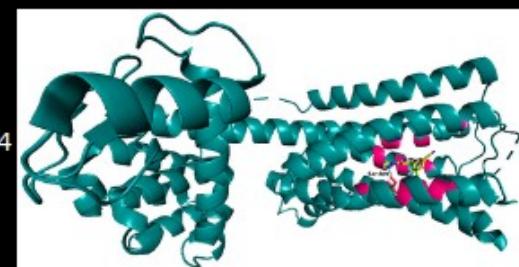
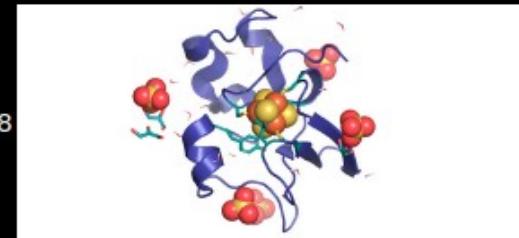
<http://web.vscht.cz/spiwokv/>
<http://www.metadynamics.cz>

Domácí úkol

<https://web.vscht.cz/~spiwokv/modelovani/>

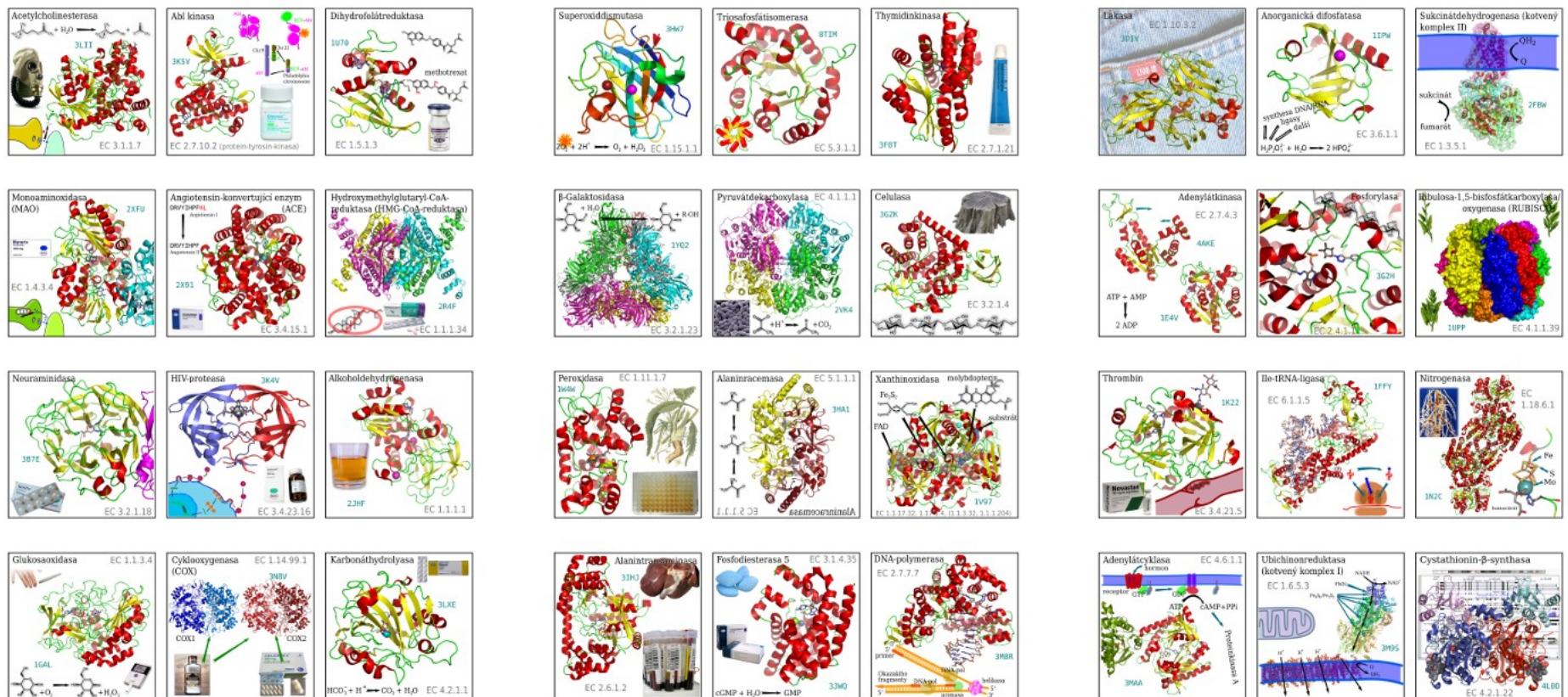


Soutěž o nejlepší ilustraci prostorové struktury
Účastníci studenštva lze pro vytvoření prostorové struktury využít software podle zadání.
Pravidla soutěže



Pexeso

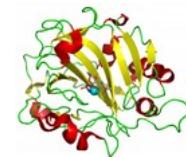
<https://web.vscht.cz/~spiwokv/enzymologie/>



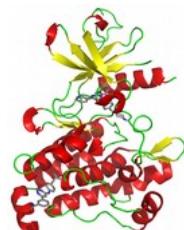
Protein databank (*1971)

<http://www.pdb.org>

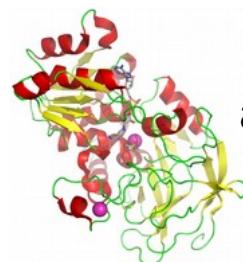
<http://www.rcsb.org>



karbonáthydrolyasa



Abl



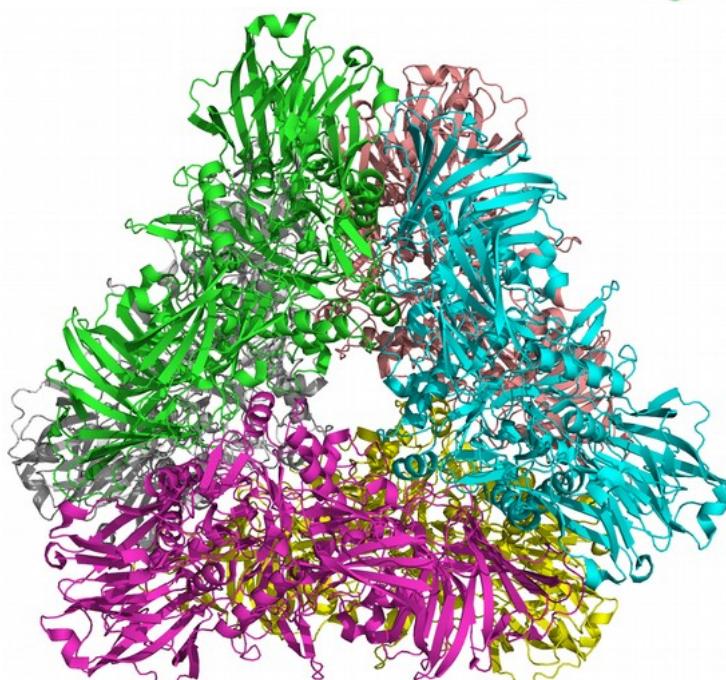
alkoholdehydrogenasa



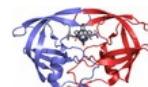
nitrogenasa



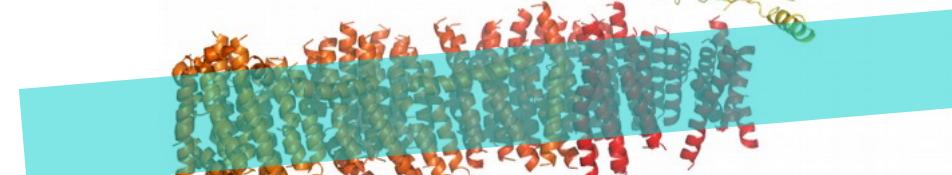
adenylátkinasa



β -galaktosidasa



HIV proteasa

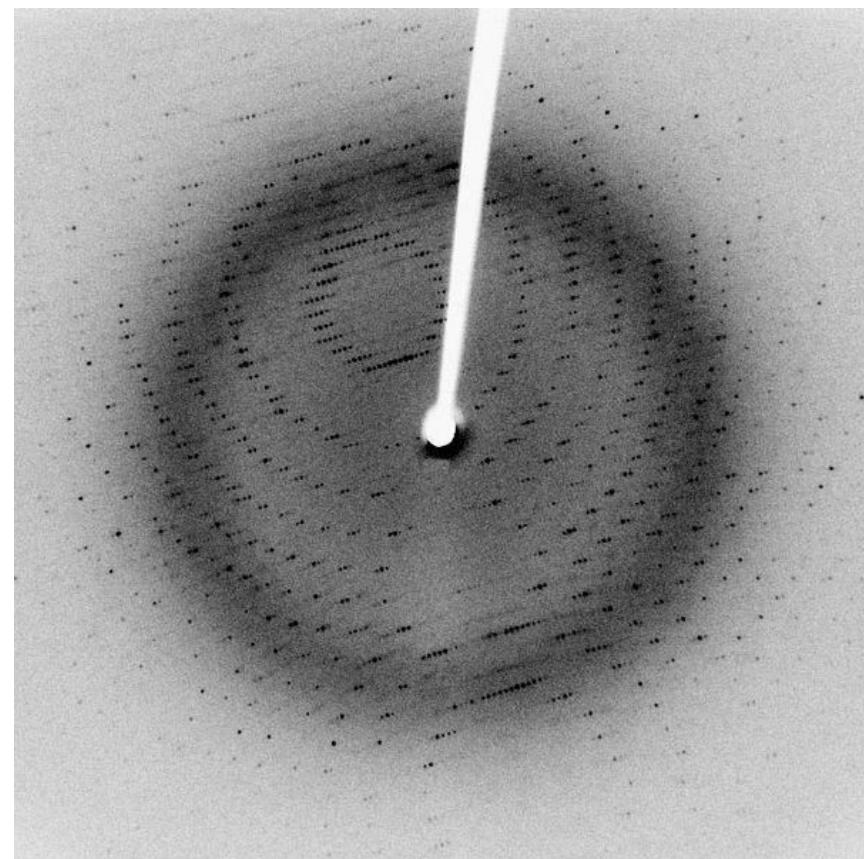
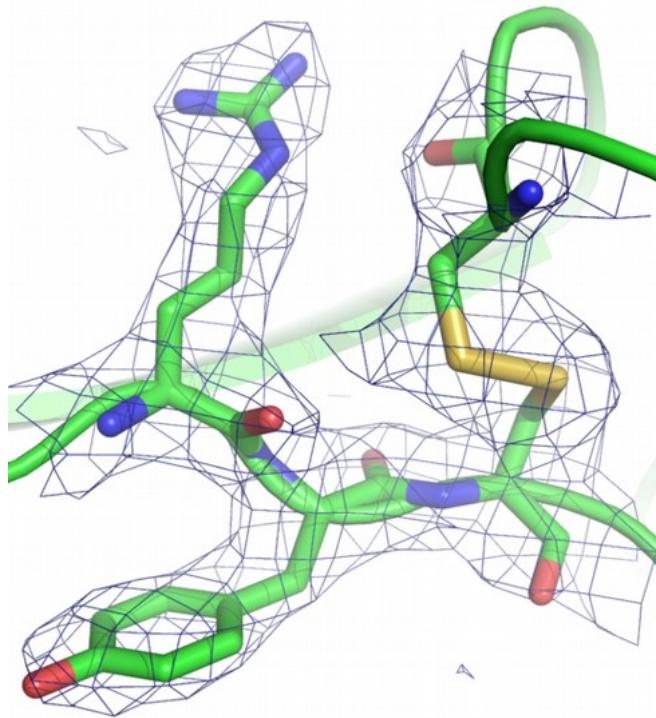
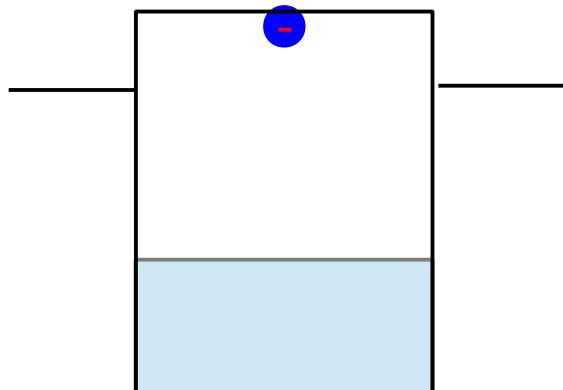


ubichinonreduktasa (kotvený komplex I)

K čemu je strukturní biologie?

- vědecká zvědavost
- vývoj nových léčiv
 - fragmentový design
 - lead optimization
- proteinové/enzymové inženýrství
- vysvětlení patologických nebo rezistence způsobujících mutací
- ...

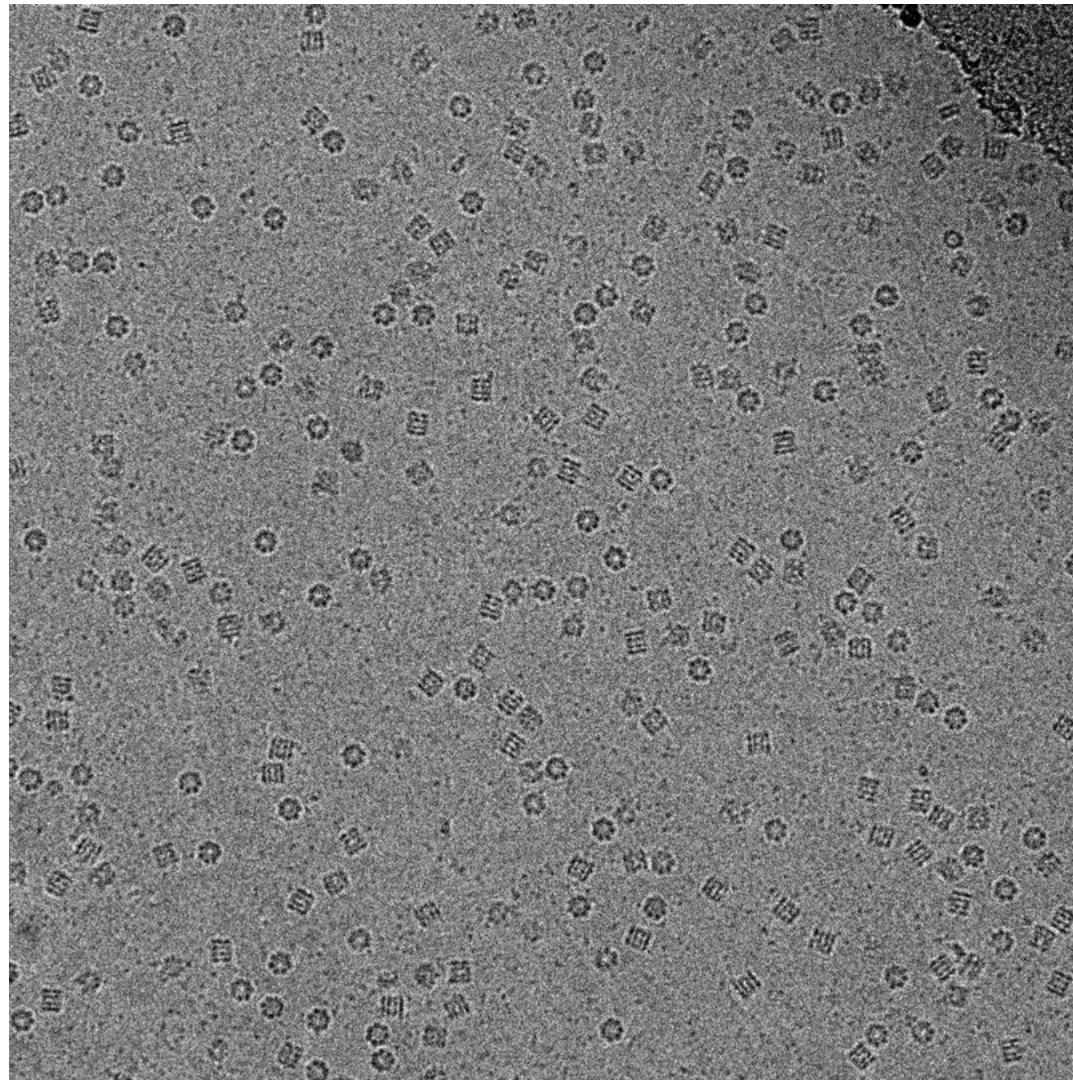
Strukturní krystalografie



NMR



CryoEM



Predikce struktury biomolekul:

Swiss model (<https://swissmodel.expasy.org/>)

 BIOZENTRUM
University of Basel
The Center for Molecular Life Sciences

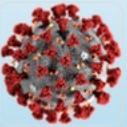
SWISS-MODEL

Modelling Repository Tools Documentation Log In Create Account

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 life science researchers worldwide.

[Start Modelling](#)



Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a positive-sense, single-stranded RNA coronavirus. It is a contagious virus that causes coronavirus disease 2019 (COVID-19).

We modelled the full SARS-CoV-2 proteome based on the NCBI reference sequence [NC_045512](#) and annotations from [UniProt](#).

The results are available [here](#).

Every week we model all the sequences for thirteen core species based on the latest UniProtKB proteome. Is your protein already modelled and up to date in [SWISS-MODEL Repository](#)?

Search SWISS-MODEL Repository



Predikce struktury biomolekul:

Swiss model (<https://swissmodel.expasy.org/>)

P0DTD1 · R1AB_SARS2

PRO_0000449621

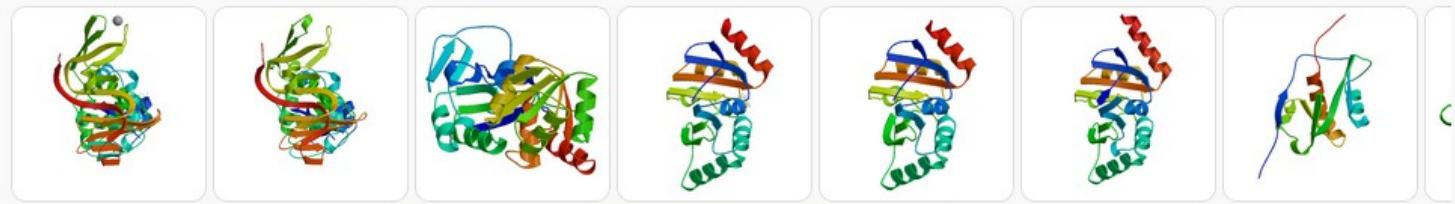


(1945 residues)

Papain-like proteinase (PL-PRO)

Responsible for the cleavages located at the N-terminus of the replicase polyprotein. In addition, PL-PRO possesses a deubiquitinating/deISGylating activity and processes both 'Lys-48'- and 'Lys-63'-linked polyubiquitin chains from cellular substrates. Participates together with nsp4 in the assembly of virally induced cytoplasmic double-membrane vesicles necessary for viral replication. Antagonizes innate immune induction of type I interferon by blocking the phosphorylation, dimerization and subsequent nuclear translocation of host IRF3. Prevents also host NF-kappa-B signaling.

Partial coverage by experimental structures (PDB: 6vxs, 6w02, 6w6y, 6w9c, 6wcf, 6wen, 6wey, 6woj, 6wrh, 6wuu, 6wx4, 6wzu, 6xg3, 6ywk, 6ywl, 6ywm, 7cjd, 7cjm, 7cmd, 7jir, 7jlt, 7jiv, 7jiw, 7jme, 7jn2, 7jrn) and high quality models are available. Partial coverage by experimental structures of hetero-oligomeric complexes exist with ISG15_HUMAN (PDB: 6xa9), ISG15_MOUSE (PDB: 6yva), UBB_HUMAN (PDB: 6xaa).



Předpověď struktur proteinů:

Homologní modelování

- proteiny s podobnou sekvencí mají podobnou strukturu

Fold recognition, threading

- proteiny mohou mít podobnou strukturu a nemusí mít (moc) podobnou sekvenci

Ab initio, de novo

- nativní struktura má určité vlastnosti

Strojové učení - AlphaFold

Simulace sbalování

- protein se dokáže sbalit do nativní struktury

Předpověď struktur proteinů:

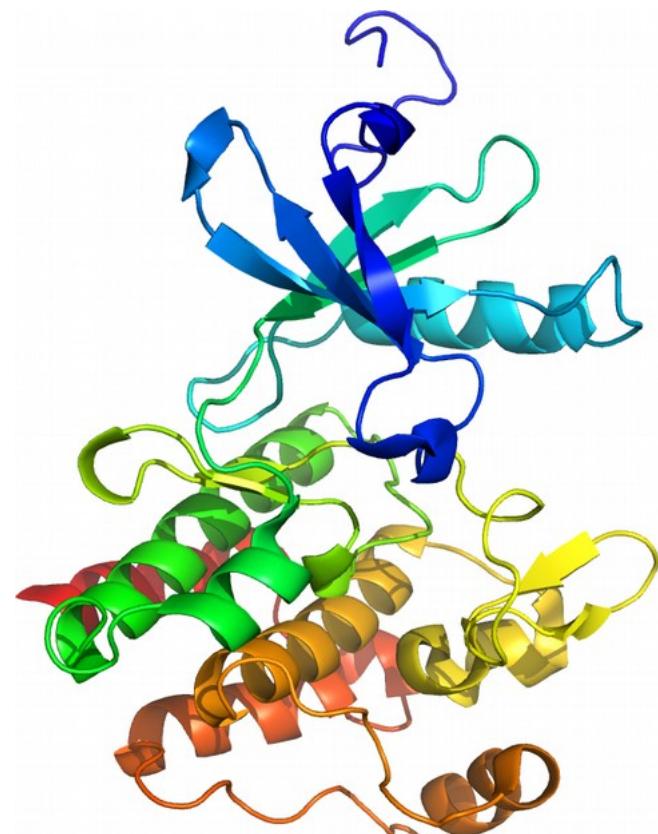
Homologní modelování

- proteiny s podobnou sekvencí mají podobnou strukturu

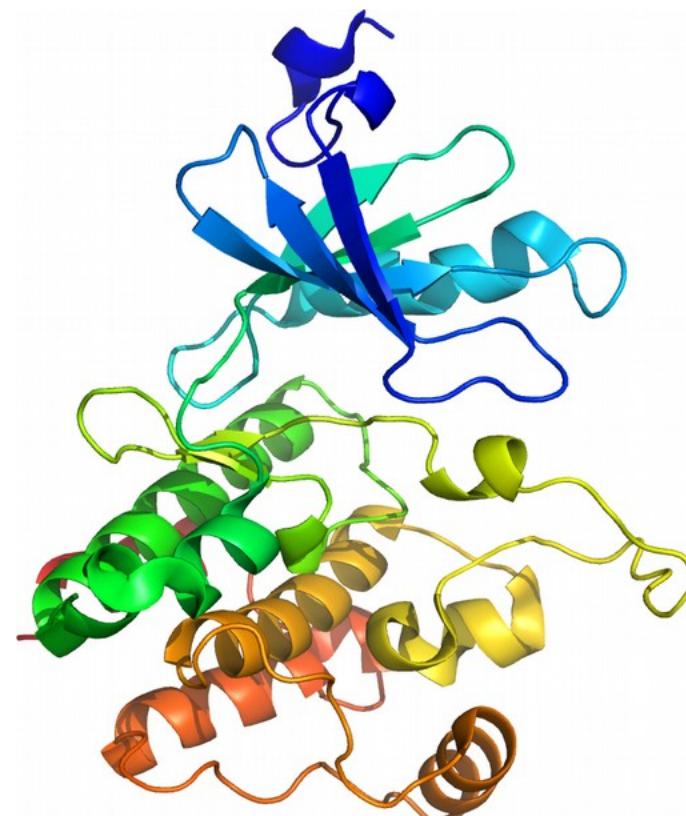
Postup:

1. nalezneme protein(y) se známou prostorovou strukturou a s podobnou sekvencí našemu proteinu
2. vytvoříme zarovnání sekvencí
3. vytvoříme model struktury našeho proteinu

Předpověď struktur proteinů:
Homologní modelování
- proteiny s podobnou sekvencí mají podobnou strukturu



Abl (1IEP)



Lck (2PL0)

Předpověď struktur proteinů:

Homologní modelování

- proteiny s podobnou sekvencí mají podobnou strukturu

>Lck

GSHMQTQKPQKPPWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTKAVKSLKGSMSPDAFLAEANLMKQLQHQRL
VRLYAVVTQEPIYYIITEYMEENGSLVDFLKTPSGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILVSDTLSCKI
ADFGGLARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTEIVTHGRIPYPGMTNPEVIQNLERGYRMV
RPDNCPEELYQLMRLCWKERPEDRPTFDYLRSVLEDFFTATEGQYQPQP

Identita 48 %

Query	16	DEWEVPRETLKLVERLGAGQFGEVWMGYYNGHT-KVAVKSLKGSMSPDAFLAEANLMKQ	74
		D+WE+ R + + +LG GQ+GEV+ G + ++ VAVK+LK+ +M + FL EA +MK+	
Sbjct	6	DKWEMERTDITMKHKLGGGQYGEVYEGVWKYSLTAVKTLKEDTMEVEEFLKEAAVMKE	65
Query	75	LQHQRLVRLYAVVTQEPIYYIITEYMEENGSLVDFLKTPSGIKLTINKLLDMAAQIAEGMA	133
		++H LV+L V T+EP YIITE+M G+L+D+L+ + ++ LL MA QI+ M	
Sbjct	66	IKHPNLVQLLGVC TREPPFYIITEFMTYGNL LDYLRECNRQEVA VVLLYMATQISSAME	125
Query	134	FIEERNYIHRDLRAANILVSDTLSCKIADFGGLARLIEDNEYTAREGAKFPIKWTAPEAIN	193
		++E++N+IHRDL A N LV + K+ADFG+RL+ + YTA GAKFPIKWTAPE++	
Sbjct	126	YLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAHAGAKFPIKWTAPESLA	185
Query	194	YGTFTIKSDVWSFGILLTEIVTHGRIPYPGMTNPEVIQNLERGYRMVRPDNCPEELYQLM	253
		Y F+IKSDVW+FG+LL EI T+G PYPG+ +V + LE+ YRM RP+ CPE++Y+LM	
Sbjct	186	YNKFSI KSDVWA FGVLLWEIATYGMSPY PGIDLSQVYELLEKD YRMERPEGCPEK VYELM	245
Query	254	RLCWKERPEDRPTFDYLRSVLEDFF 278	
		R CW+ P DRP+F + E F	
Sbjct	246	RACWQWNPSDRPSFAEIHQAFETMF 270	

Předpověď struktur proteinů:

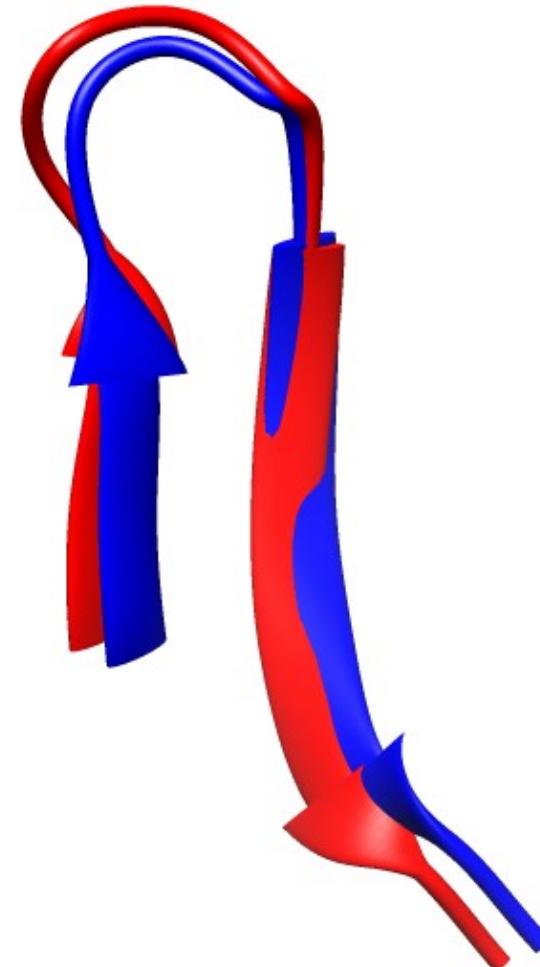
Homologní modelování

- proteiny s podobnou sekvencí mají podobnou strukturu

Lck AVVTQEP-IYIITEY

V T+EP YIITE+

Abl GVCTREPPFYIITEF



Předpověď struktur proteinů: Homologní modelování

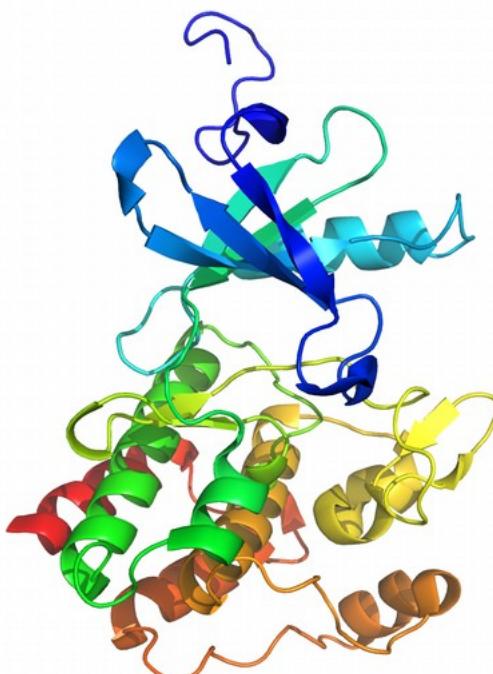
- proteiny s podobnou sekvencí mají podobnou strukturu

C; A sample alignment in the PIR format; used in tutorial

```
>P1;1IEP
structureX:1IEP:233  :A:498  :A::::
-----DKWEMERTDITMKHLGGGQYGEVYEGVWKYSLTAVKTLKEDT
MEVEEFLKEAAVMKEIKHPNLVQLLGVCTREPPFYIITEFMTYGNLLDYLRECNRQEVS
VVLLYMATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAHA
GAKFPIKWTAPESLAYNKFSIKSDVWAFFVLLWEIATYGMSPYPGIDLSQVYELLEKDYR
MERPEGCPEKVYELMRACWQWNPSDRPSFAEIHQAFETMFQ*
```

```
>P1;lck
sequence:lck:1  : :@  : ::::
-----DEWEVPRETLKLVERLGAGQFGEVWMGYYNGHT-KVAVKSLKQGS
MSPDAFLAEANLMKQLQHQRLVRLYAVVTQEP-IYIITEYMEMGSLVDFLKTSGIKLTI
NKLLDMAAQIAEGMAFIEERNYIHRDLRAANILVSDTLSCKIADFGLARLIEDNEYTARE
GAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTEIVTHGRIPYPGNTNPEVIQNLERGYR
MVRPDNCPEELYQLMRLCWKERPEDRPTFDYLRSVLEDFFT*
```

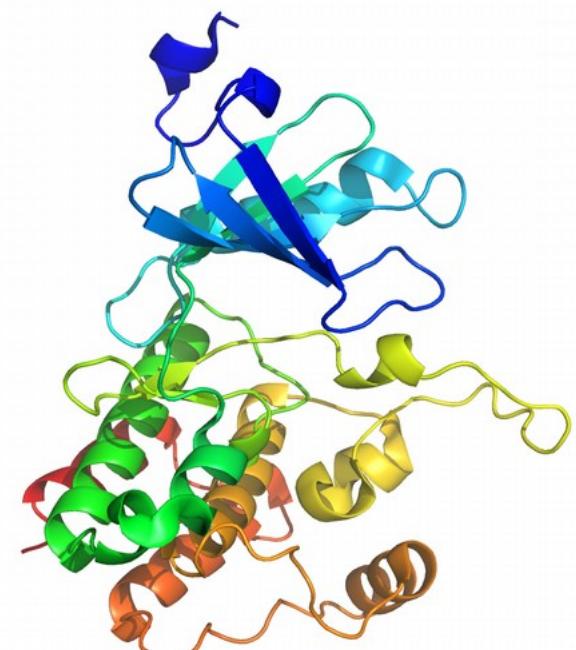
Předpověď struktur proteinů:
Homologní modelování
- proteiny s podobnou sekvencí mají podobnou strukturu



Abl (1IEP)



Lck (model)



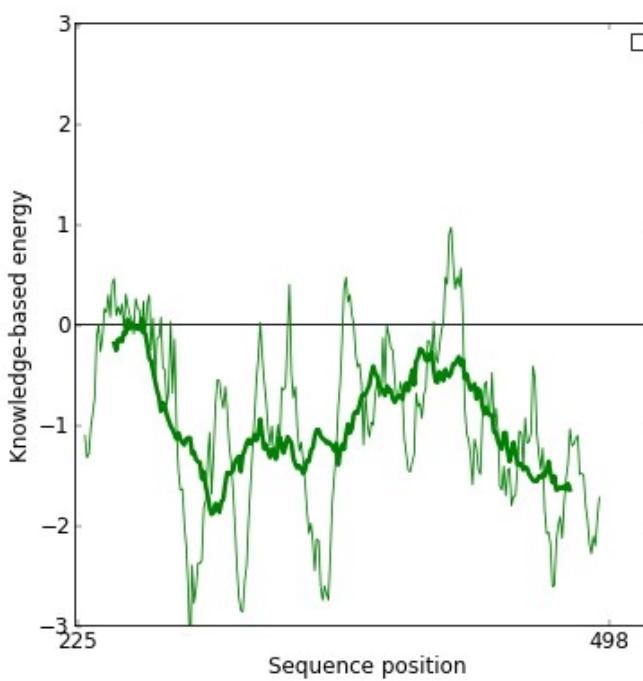
Lck (2PL0)

Předpověď struktur proteinů:

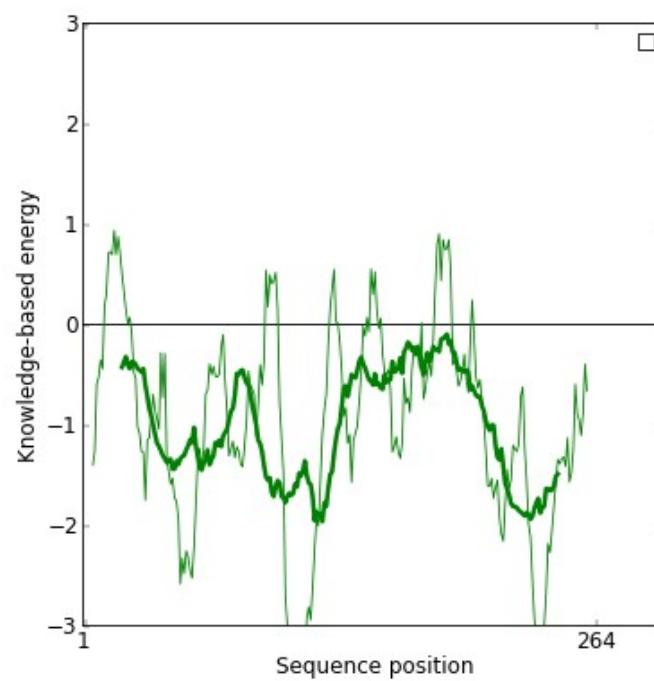
Homologní modelování

- proteiny s podobnou sekvencí mají podobnou strukturu

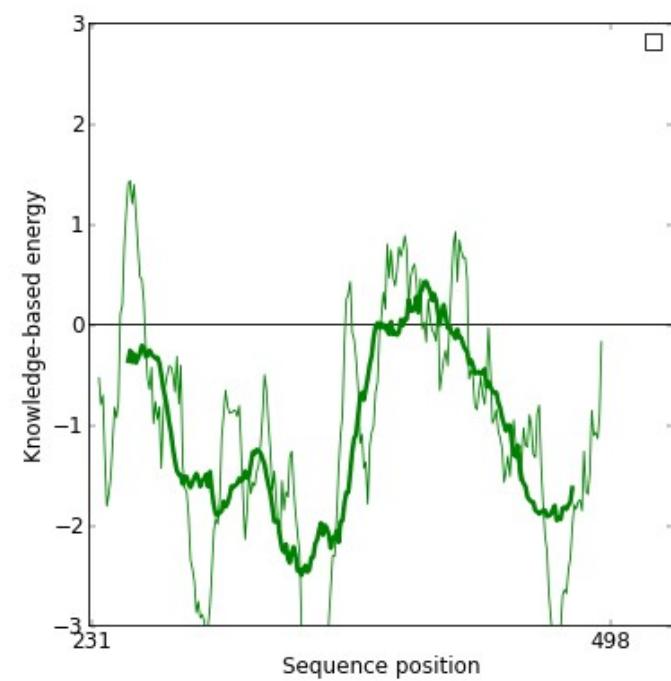
Prosall



Abl (1IEP)



Lck (model)

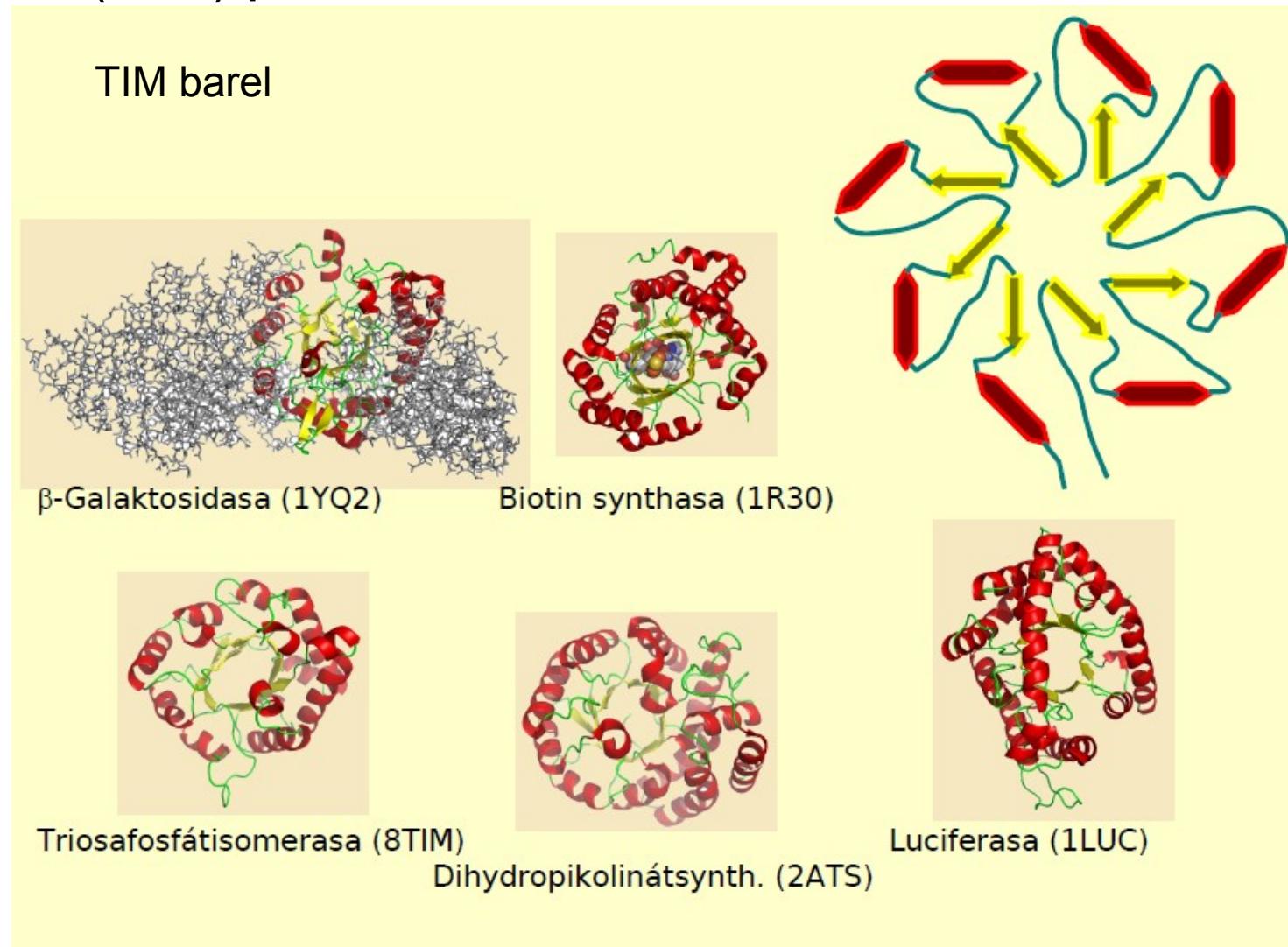


Lck (2PL0)

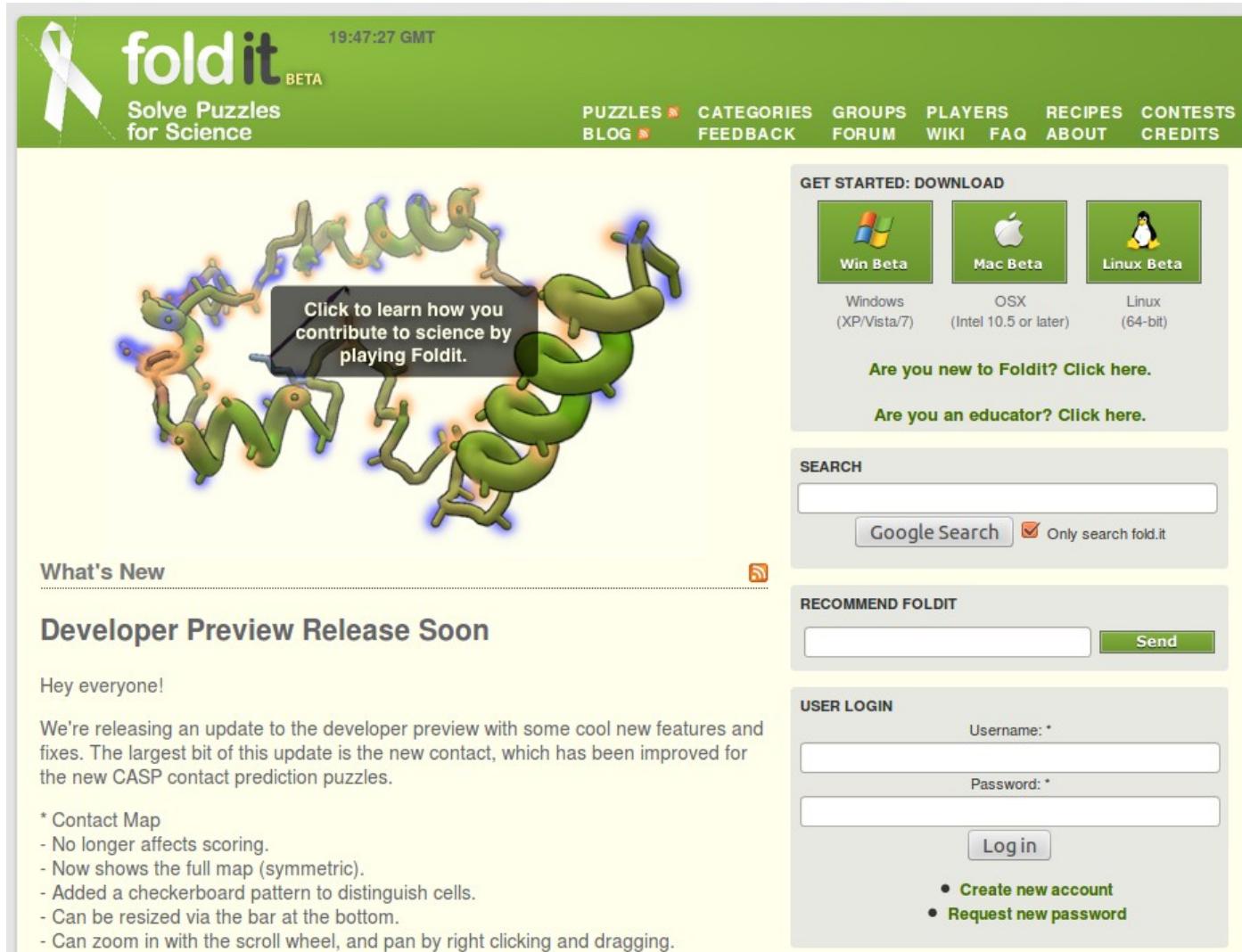
Předpověď struktur proteinů:

Fold recognition, threading

- proteiny mohou mít podobnou strukturu a nemusí mít (moc) podobnou sekvenci



Předpověď struktur proteinů:
Ab initio, de novo
- nátní struktura má určité vlastnosti

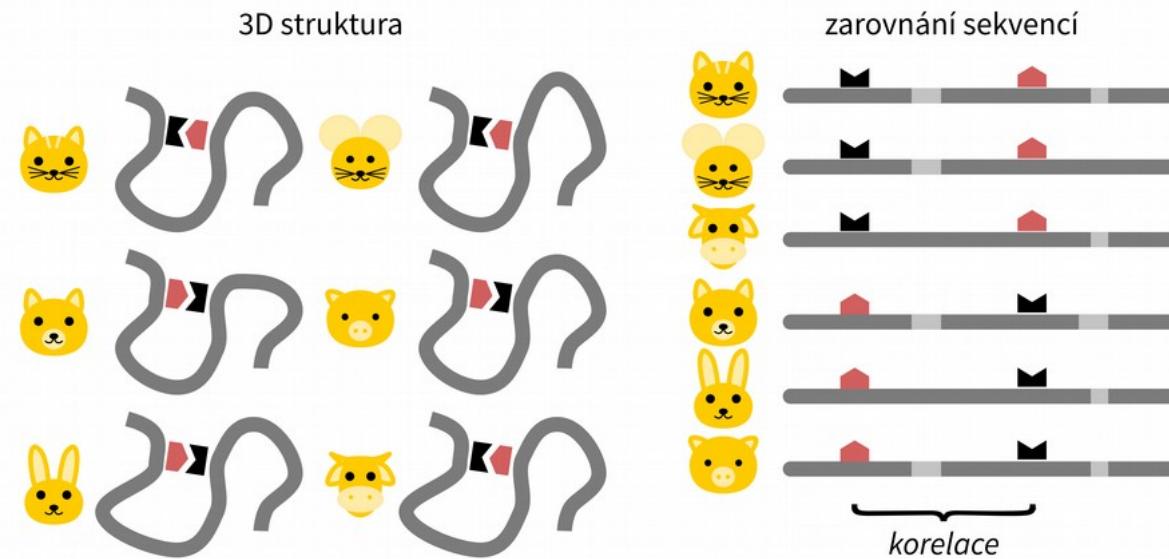


The screenshot shows the homepage of the Foldit website. At the top, there's a green header bar with the Foldit logo (a white ribbon icon) and the text "foldit BETA". Below the logo, it says "Solve Puzzles for Science". The top navigation menu includes links for PUZZLES, CATEGORIES, GROUPS, PLAYERS, RECIPES, CONTESTS, BLOG, FEEDBACK, FORUM, WIKI, FAQ, ABOUT, and CREDITS. The main content area features a large, colorful 3D model of a protein structure composed of green and blue segments. A black callout box with white text is overlaid on the protein, reading: "Click to learn how you contribute to science by playing Foldit.". To the right of the protein model, there's a "GET STARTED: DOWNLOAD" section with icons for Windows (Win Beta), Mac (Mac Beta), and Linux (Linux Beta). Below these icons, text specifies system requirements: "Windows (XP/Vista/7)", "OSX (Intel 10.5 or later)", and "Linux (64-bit)". There are also links for new users ("Are you new to Foldit? Click here.") and educators ("Are you an educator? Click here."). Further down, there's a "SEARCH" field with a "Google Search" button and a checkbox for "Only search fold.it". A "RECOMMEND FOLDIT" section has a text input field and a "Send" button. At the bottom, there's a "USER LOGIN" section with fields for "Username:" and "Password:", a "Log in" button, and links for "Create new account" and "Request new password". On the left side of the main content area, there's a "What's New" section with a link to "Developer Preview Release Soon" and a message from the developer: "Hey everyone! We're releasing an update to the developer preview with some cool new features and fixes. The largest bit of this update is the new contact, which has been improved for the new CASP contact prediction puzzles." Below this message is a list of changes: "* Contact Map", "- No longer affects scoring.", "- Now shows the full map (symmetric).", "- Added a checkerboard pattern to distinguish cells.", "- Can be resized via the bar at the bottom.", "- Can zoom in with the scroll wheel, and pan by right clicking and dragging."

<http://fold.it>

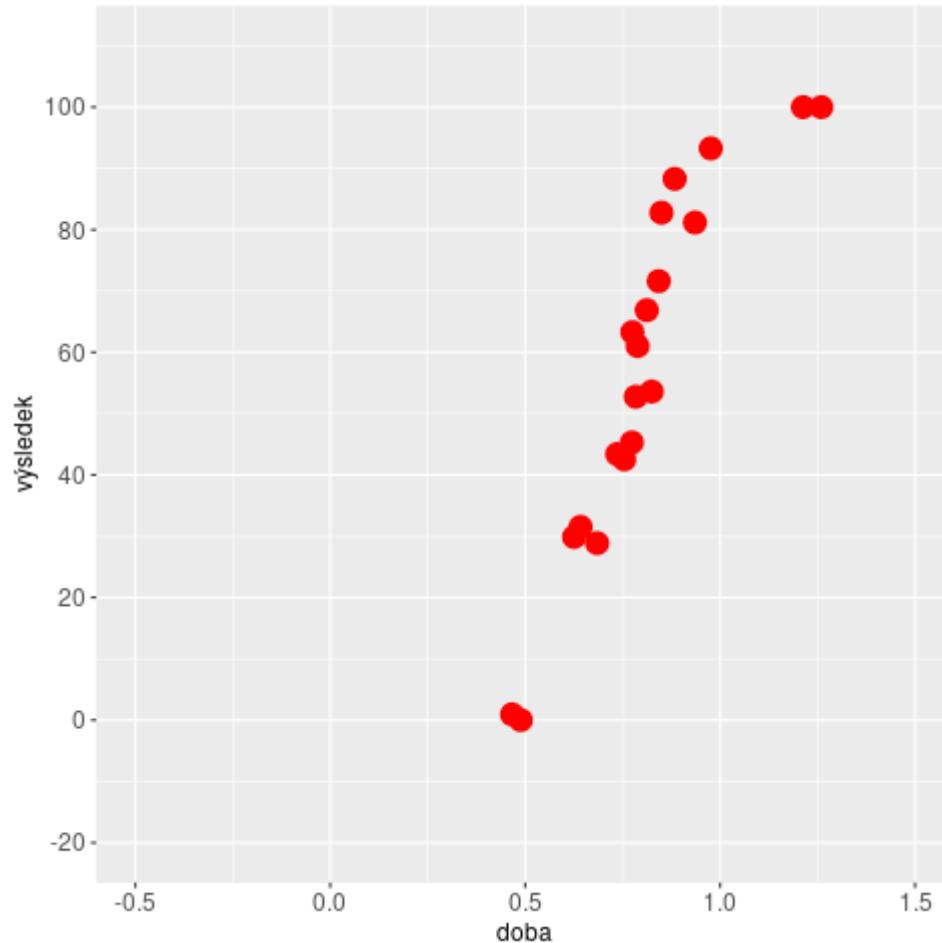
AlphaFold:

- koevoluce



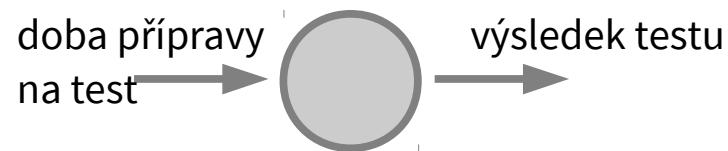
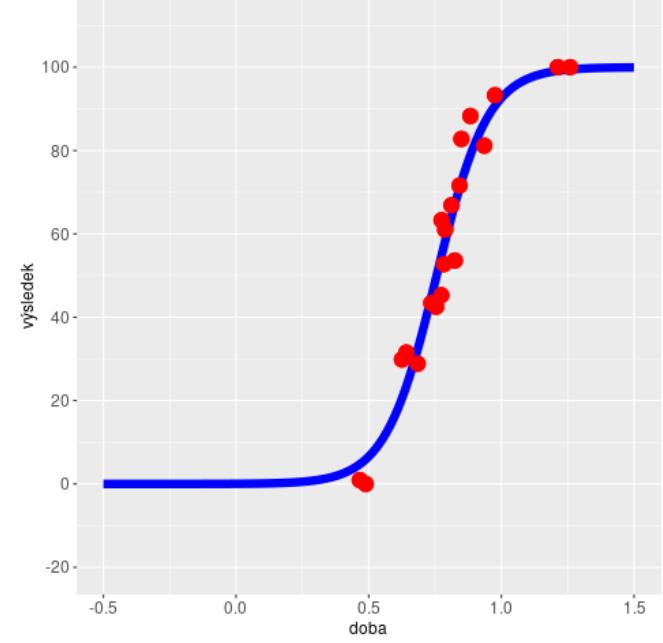
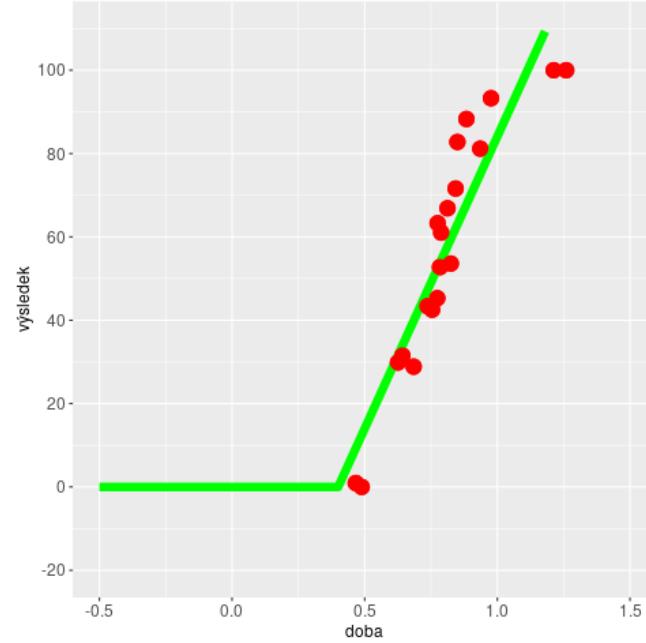
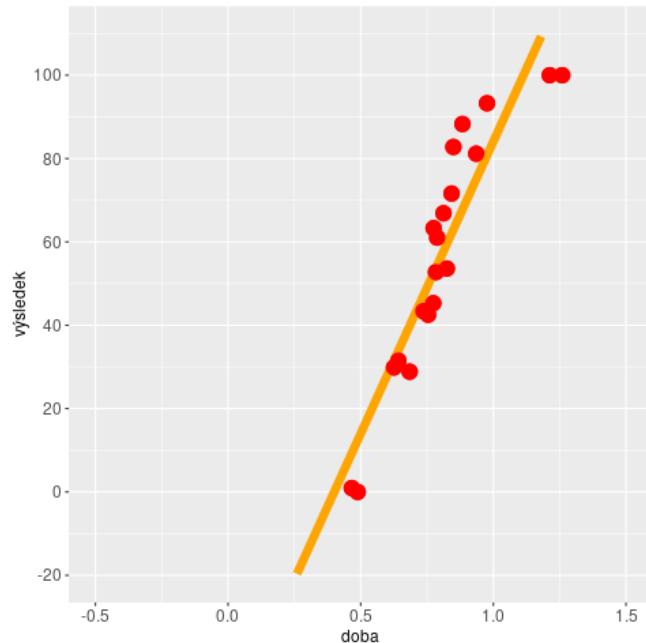
AlphaFold:

- neuronové sítě



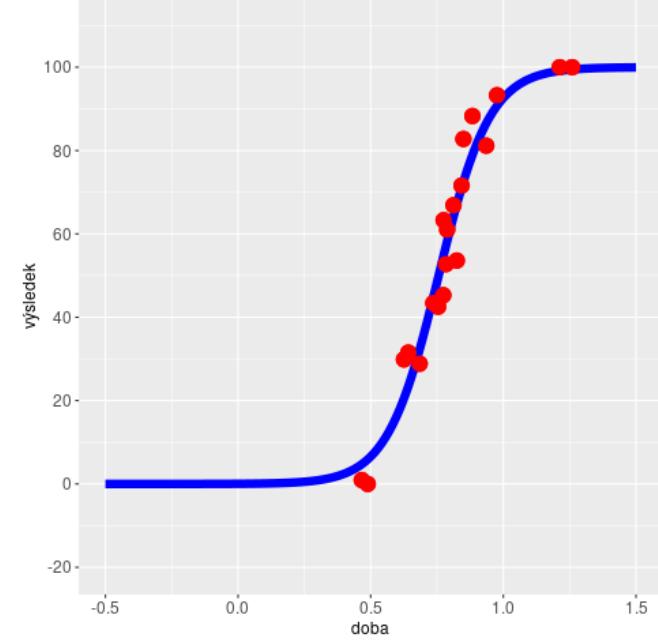
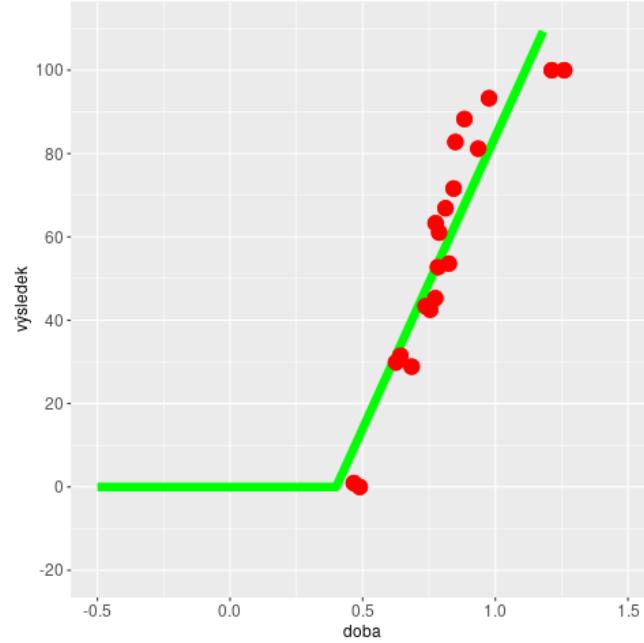
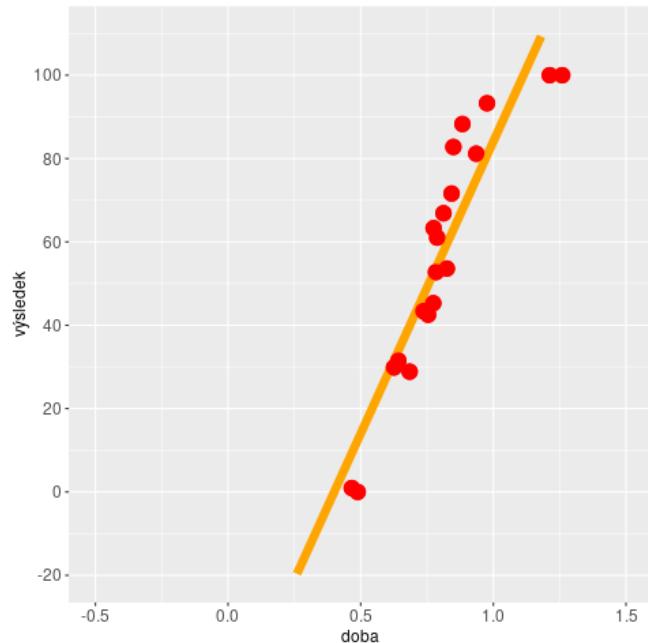
AlphaFold:

- neuronové sítě

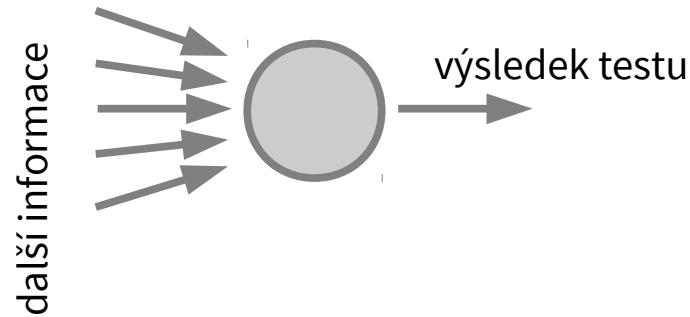


AlphaFold:

- neuronové sítě

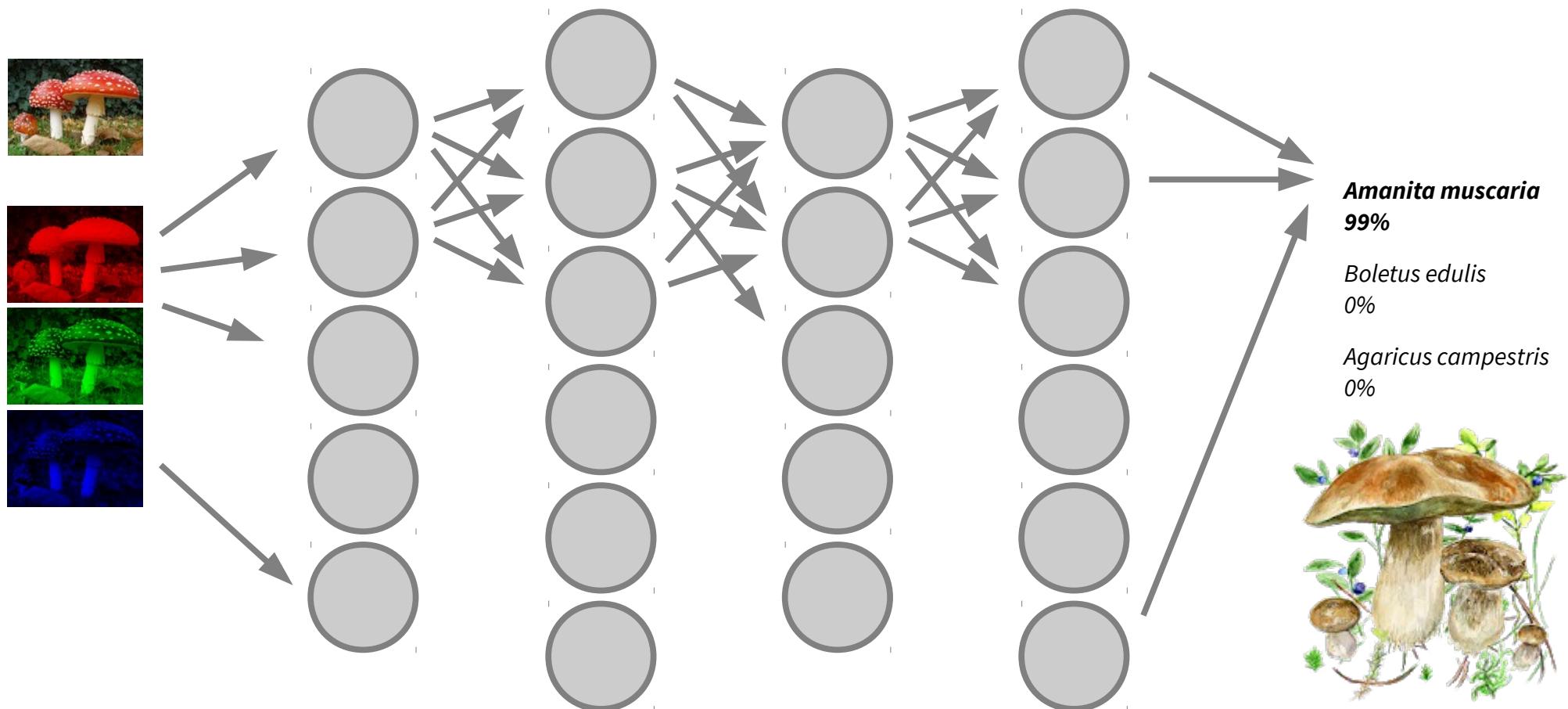


doba přípravy
na test



AlphaFold:

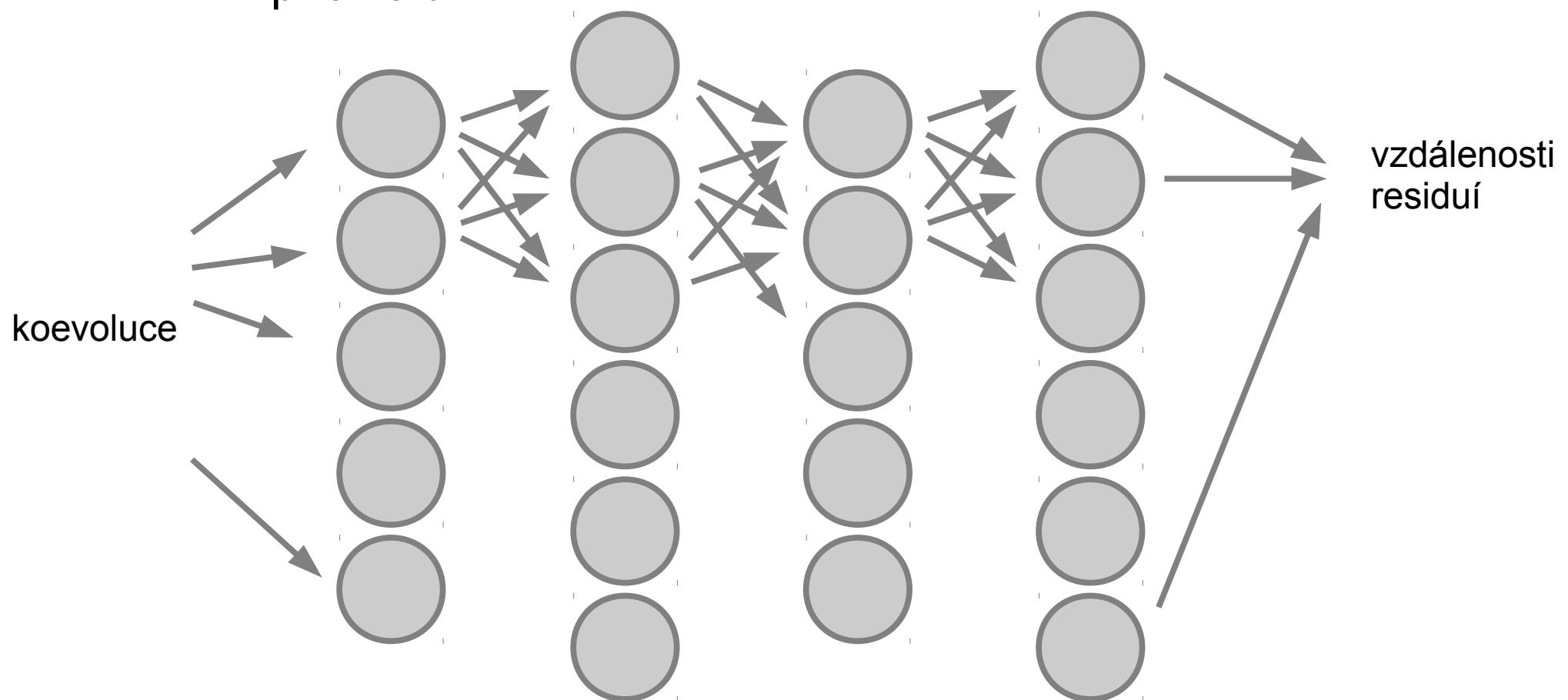
- neuronové sítě



<https://play.google.com/store/apps/details?id=bazinac.aplikacenahouby>

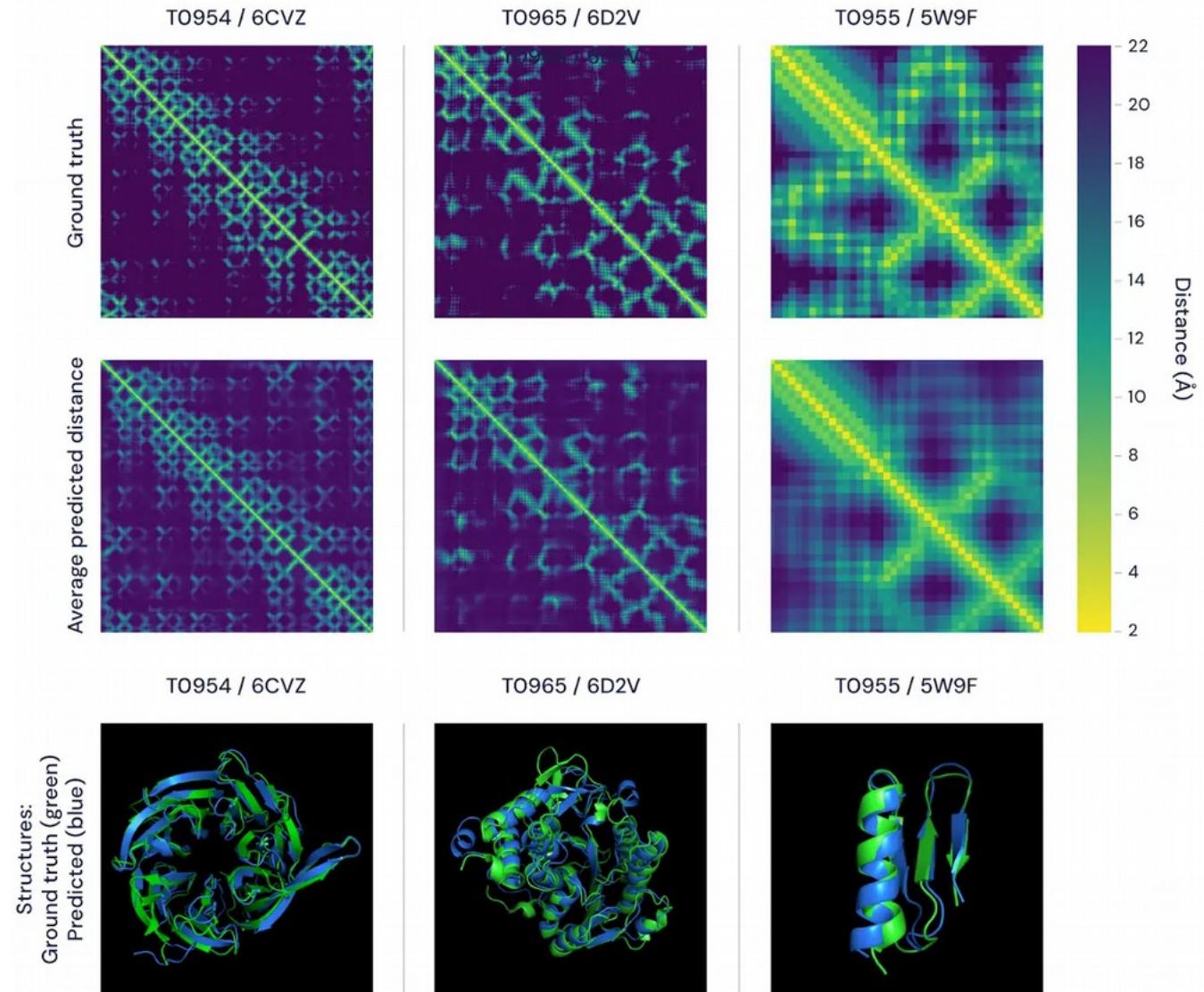
AlphaFold:

- neuronové sítě
AlphaFold 1



AlphaFold:

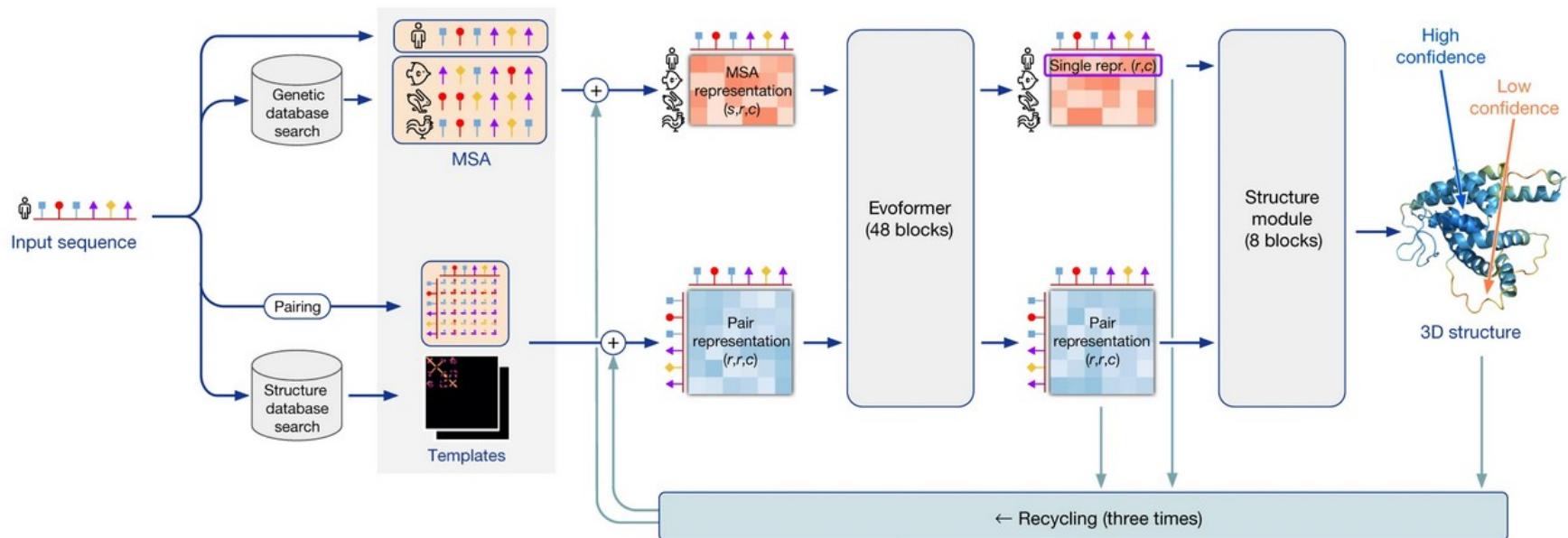
- neuronové sítě
AlphaFold 1



Senior *et al.* *Nature* 2020, **577**, 706-710

AlphaFold:

- neuronové sítě
- ## AlphaFold 2



Jumper et al. *Nature* 2021, **596**, 583-589

Předpověď struktur proteinů: Critical Assessment of protein Structure Prediction (CASP)

The screenshot displays the 'Target List' page for the 15th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP15). The page features a navigation menu on the left and a main content area with a table of targets.

Navigation Menu:

- Home
- PC Login
- PC Registration
- CASP Experiments
- CASP15 (2022)
 - Home
 - Registration
 - My CASP15 profile
- CASP Commons (COVID-19, 2020)
- CASP14 (2020)
- CASP13 (2018)
- CASP12 (2016)
- CASP11 (2014)
- CASP10 (2012)
- CASP9 (2010)
- CASP8 (2008)
- CASP7 (2006)
- CASP6 (2004)
- CASP5 (2002)
- CASP4 (2000)
- CASP3 (1998)
- CASP2 (1996)
- CASP1 (1994)

Main Content Area:

Target List [CSV](#)

Targets expire on the specified date at noon (12:00) local time in California (GMT - 7 hours).

Green color - active target; Yellow color - target expires within 48 hours; Orange color - target expires within 24 hours
RNA targets are highlighted with the light grey background.

* CAPRI target

#	Tar-id	Type	Res	Stoichiom.	Entry Date	Server Expiration	Human Expiration	QA Prediction	Description
1.	T1104	All groups	117	A1	2022-05-02	2022-05-05	2022-05-23	-	EntV136
2.	T1105	All groups	331	A1	2022-05-03	2022-05-06	2022-05-24	-	Bacterial enzyme
3.	H1106 *	All groups	236	A1B1	2022-05-04	2022-05-07	2022-05-25	from: 2022-05-26 to: 2022-05-28	YscY-YscX PDB code 7qih
4.	T1106s1	All groups	122	A1	2022-05-04	2022-05-07	2022-05-25	-	YscY-YscX PDB code 7qih
5.	T1106s2	All groups	114	A1	2022-05-04	2022-05-07	2022-05-25	-	YscY-YscX PDB code 7qih
6.	T1109 *	All groups	227	A2	2022-05-05	2022-05-08	2022-05-26	from: 2022-05-27 to: 2022-05-29	D180A isocyanide hydratase
7.	T1110 *	All groups	227	A2	2022-05-05	2022-05-08	2022-05-26	from: 2022-05-27 to: 2022-05-29	wild-type isocyanide hydratase
8.	H1111	All groups	940	A9B9C9	2022-05-06	2022-05-09	2022-05-27	from: 2022-05-28	YscY-YscX-LcrD