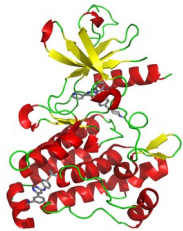


Protein databank (*1971)

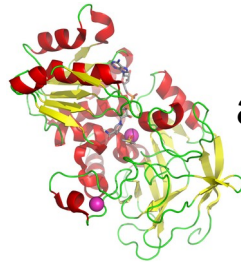
<http://www.pdb.org>



karbonáthydrolyasa



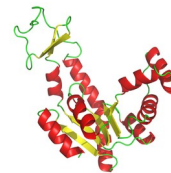
Abl



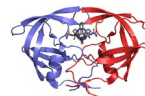
alkoholdehydrogenasa



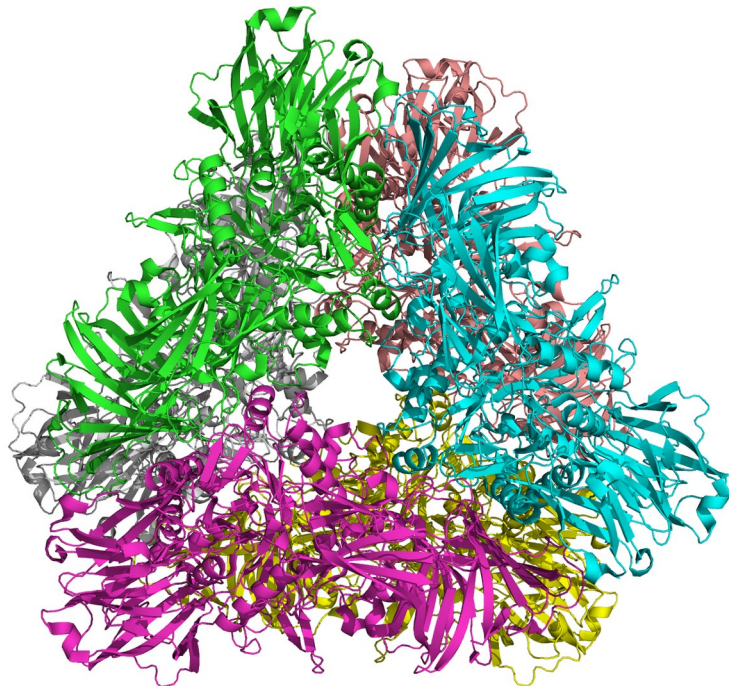
nitrogenasa



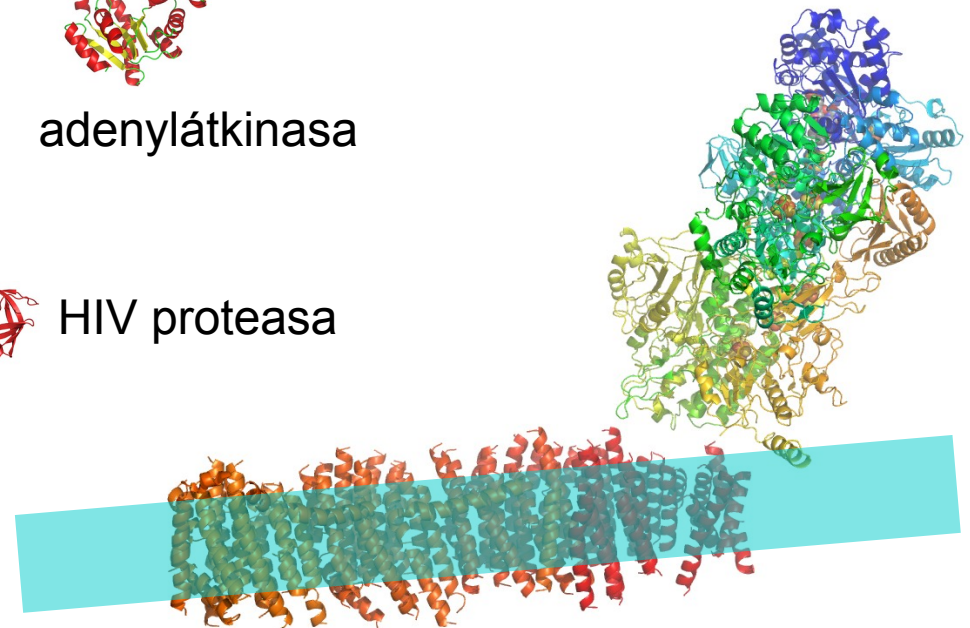
adenylátkinasa



HIV proteasa

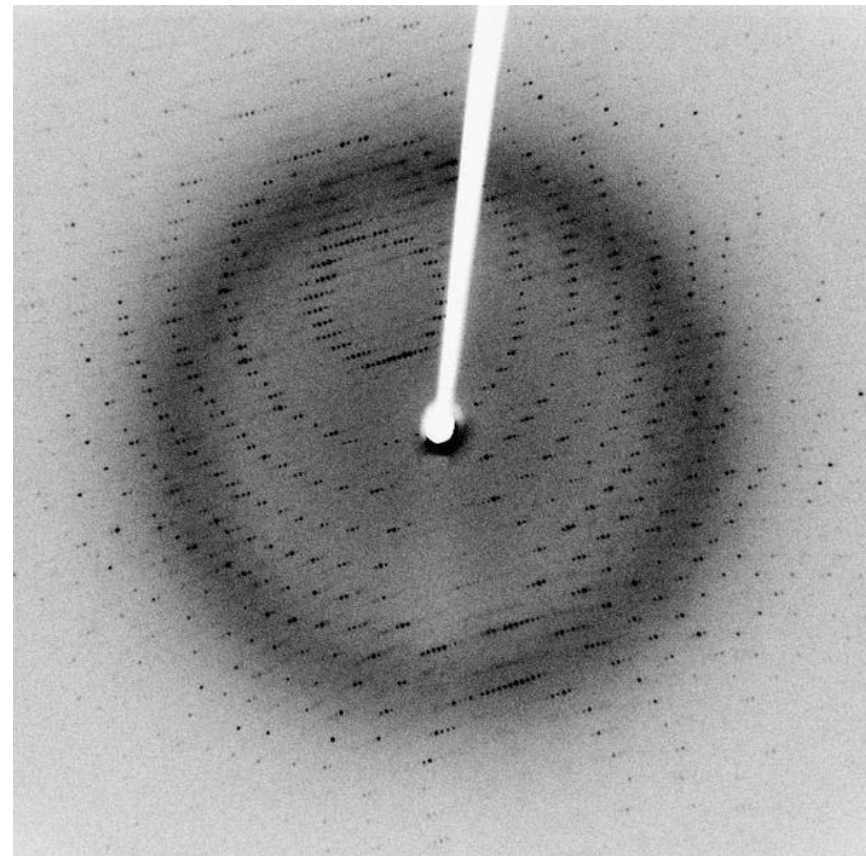
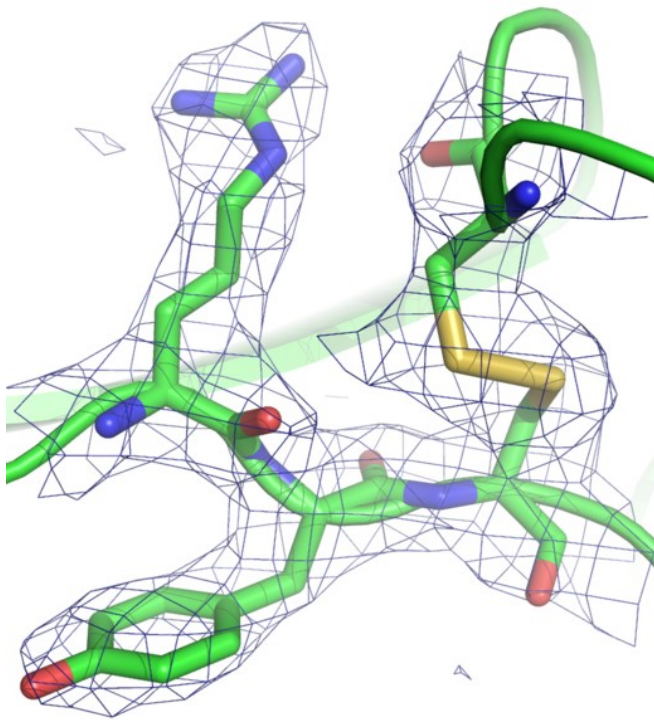
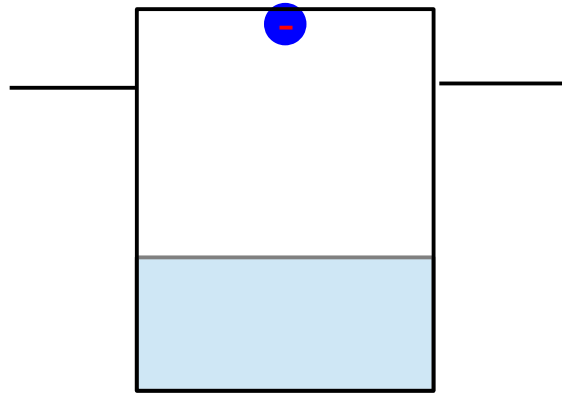


β -galaktosidasa



ubichinonreduktasa (kotvený komplex I)

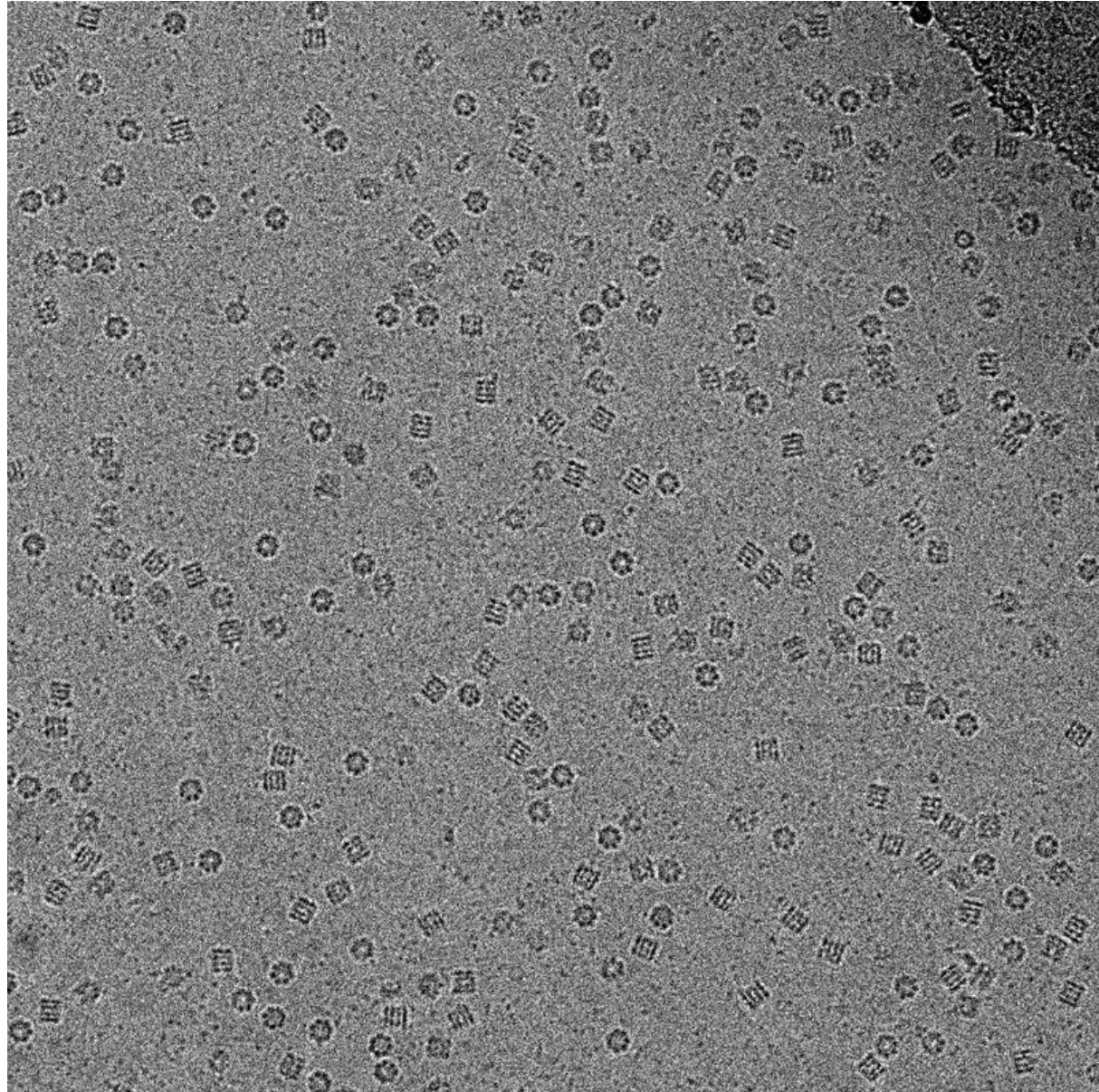
Struktura proteinů: Rentgenová strukturní krystalografie



Struktura proteinů:
NMR



Struktura proteinů:
CryoEM



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

- je možné naučit se „jazyk“ proteinových sekvencí a překládat je do 3D struktur

Simulace sbalování

- protein se dokáže sbalit do nativní struktury

Předpověď struktur proteinů:

Homologní modelování

- proteiny s podobnou sekvencí mají podobnou strukturu

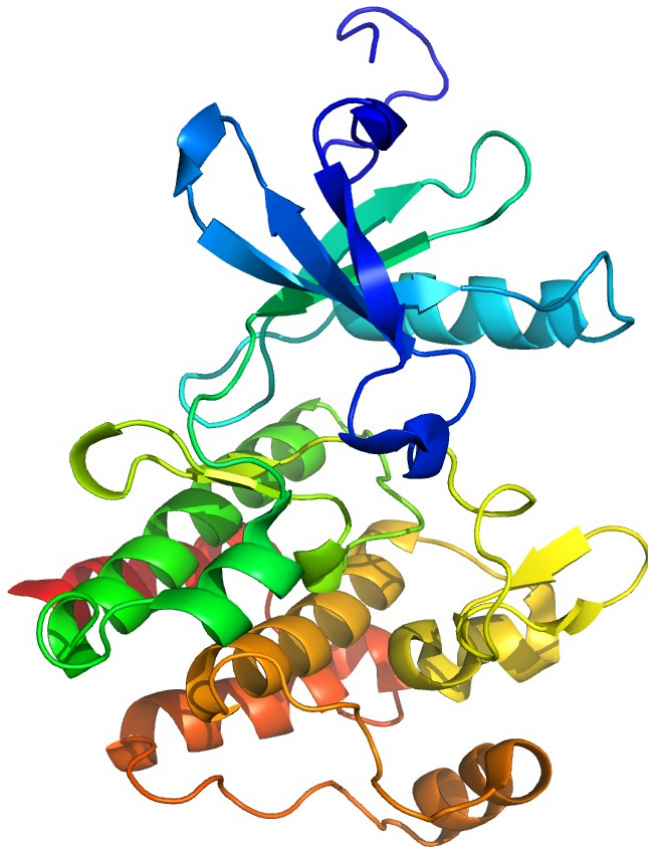
Postup:

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

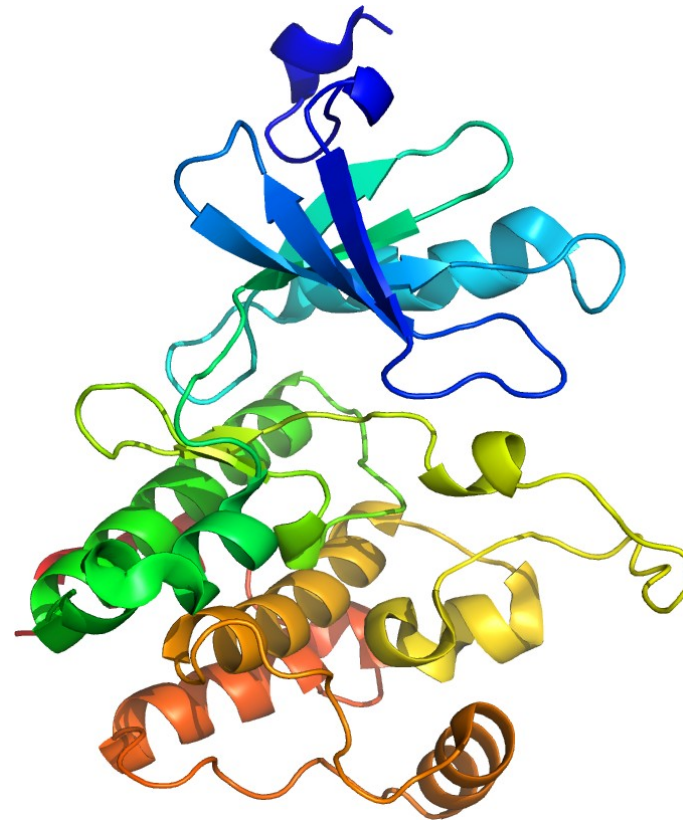
Předpověď struktur proteinů:

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

```
GSHMQTQKPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWVMGYNGHTKVAVKSLKQGSMSPD AFLAEANLMKQLQHQR L
VRLYAVVTQEPIYIITEYMENGLVDFLKTPSGIKLTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILVSDTLSCKI
ADFG LARLIEDNEYTAREGAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTEIVTHGRIPYPGMTNPEVIQNLERGYRMV
RPDNCPEELYQLMRLCWKERPEDRPTFDYLRSVLEDDFF TATEGQYQPQP
```

Identita 48 %

Query	16	DEWEVPRETLKLVERLGAGQFGEVWVMGYNGHT-KVAVKSLKQGSMSPD AFLAEANLMKQ	74
		D+WE+ R + + +LG GQ+GEV+ G + ++ VAVK+LK+ +M + FL EA +MK+	
Sbjct	6	DKWEMERTDITMKHKLGGGQYGEVYEGVWKKYSLTVAVKTLKEDTMEVEEFLKEAAVMKE	65
Query	75	LQHQR L VRLYAVVTQEP-IYIITEYMENGLVDFLKTPSGIKLTINKLLDMAAQIAEGMA	133
		++H LV+L V T+EP YIITE+M G+L+D+L+ + ++ LL MA QI+ M	
Sbjct	66	IKHPNLVQLLGVCTREPPFYIITEFMTYGNLLDYLRECNRQEVNAVVL LYMATQISSAME	125
Query	134	FIEERNYIHRDLRAANILVSDTLSCKIADFG LARLIEDNEYTAREGAKFPIKWTAPEAIN	193
		++E++N+IHRDL A N LV + K+ADFG L+RL+ + YTA GAKFPIKWTAPE++	
Sbjct	126	YLEKKNFIHRDLAARNCLVGENHLVKVADFG LSR LMTGDTYTAHAGAKFPIKWTAPESLA	185
Query	194	YGTFTIKSDVWSFGILLTEIVTHGRIPYPGMTNPEVIQNLERGYRMV RPDNCPEELYQLM	253
		Y F+IKSDVW+FG+LL EI T+G PYPG+ +V + LE+ YRM RP+ CPE++Y+LM	
Sbjct	186	YNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLSQVYELLEKDYRMERPEGCPEKVYELM	245
Query	254	RLCWKERPEDRPTFDYLRSVLEDDFF	278
		R CW+ P DRP+F + E F	
Sbjct	246	RACQWNP SDRPSFAEIHQAFETMF	270

Předpověď struktur proteinů:

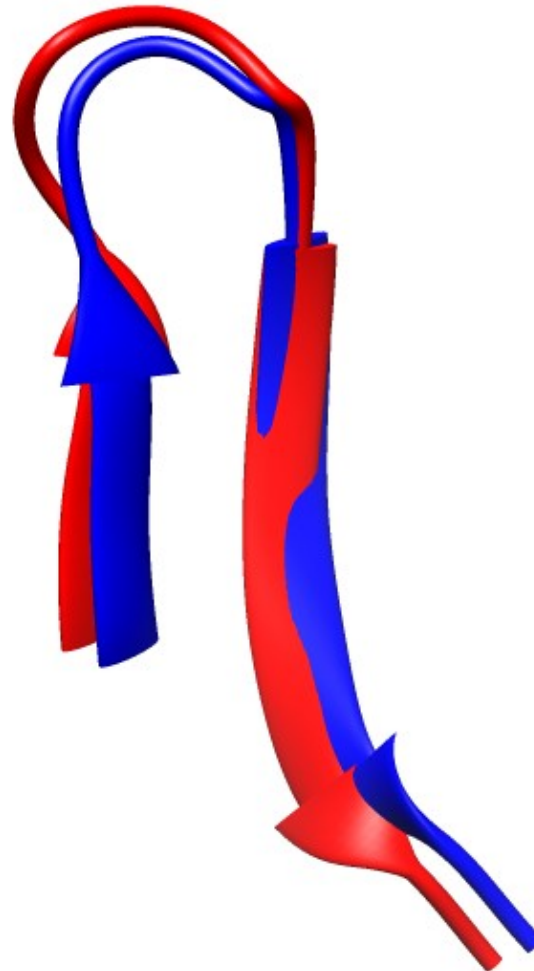
Homologní modelování

- 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

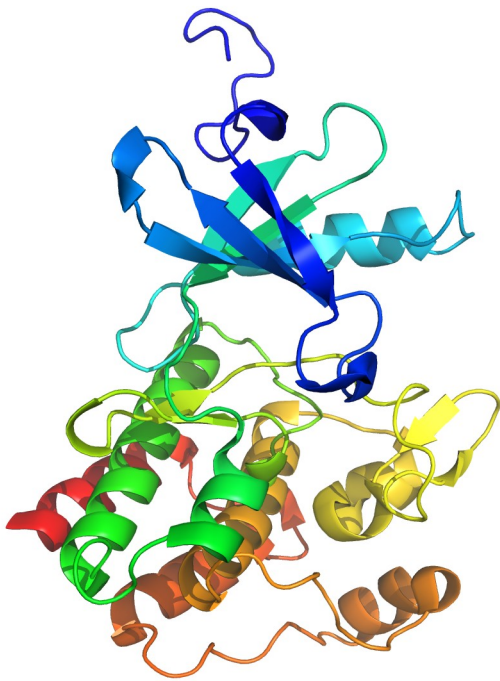
```
>P1;1IEP
structureX:1IEP:233 :A:498 :A::::
-----DKWEMERTDITMKHKLGGGQYGEVYEGVWKKYSLTVAVKTLKEDT
MEVEEFLKEAAMKEIKHPNLVQLLGVCTREPPFYIITEFMTYGNLLDYLRECNRQEVSA
VLLLYMATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAHA
GAKFPIKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLSQVYELLEKDYR
MERPEGCPEKVYELMRACWQWNPSDRPSFAEIHQAFETMFQ*
```

```
>P1;lck
sequence:lck:1 :@ : ::::
-----DEWEVPRETLKLVERLGAGQFGEVWVGYYNGHT-KVAVKSLKQGS
MSPDAFLAEANLMKQLQHQRLLVRLYAVVTQEP-IYIITEYMEENGSLVDFLKTTPSGIKLTI
NKLLDMAAQIAEGMAFIEERNYIHRDLRAANILVSDTLSCKIADFGGLARLIEDNEYTARE
GAKFPIKWTAPEAINYGTFTIKSDVWSFGILLTEIVTHGRIPYPGMTNPEVIQNLERGYR
MVRPDNCPEELYQLMRLCWKERPEDRPTFDYLRVLEDDFFT*
```

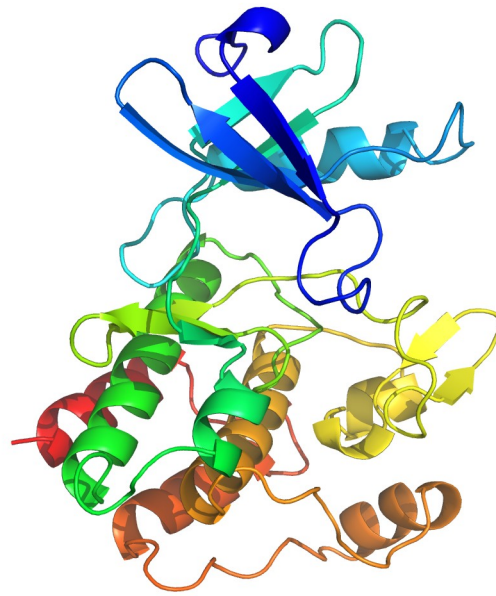
Předpověď struktur proteinů:

Homologní modelování

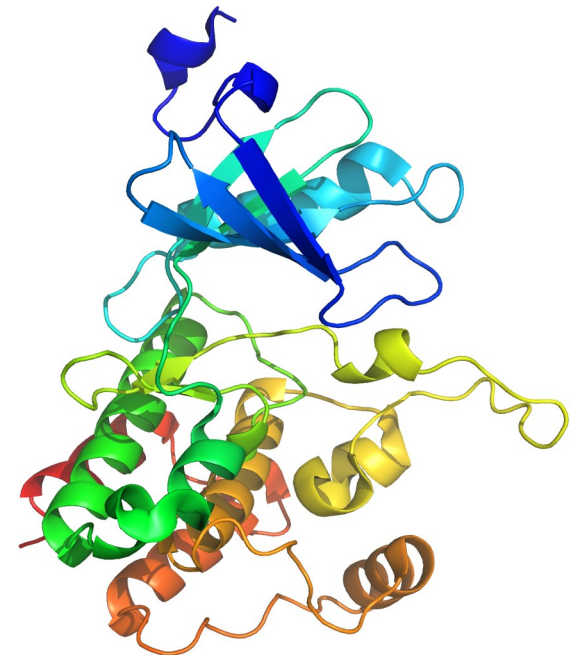
- proteiny s podobnou sekvencí mají podobnou strukturu



Abl (1IEP)



Lck (model)



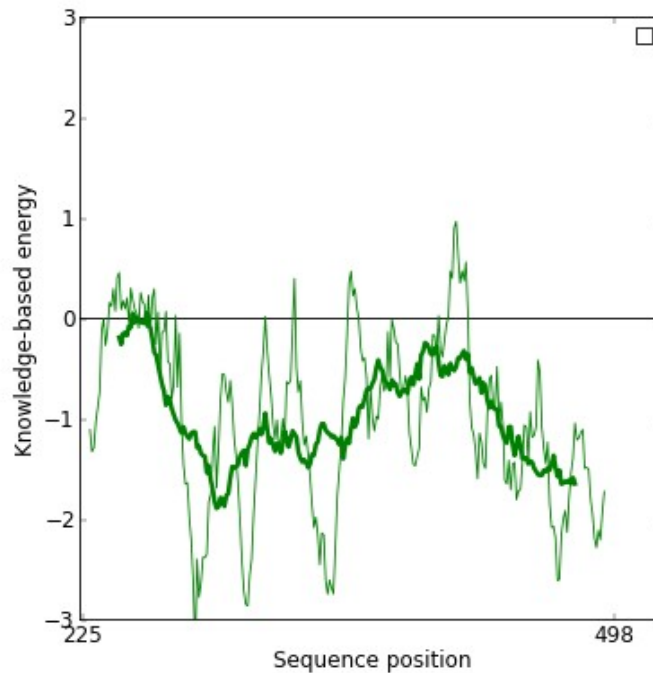
Lck (2PL0)

Předpověď struktur proteinů:

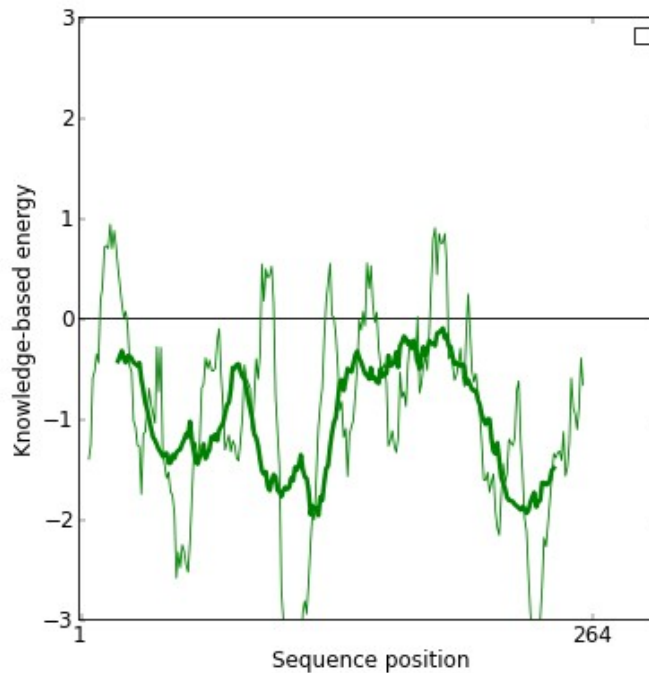
Homologní modelování

- proteiny s podobnou sekvencí mají podobnou strukturu

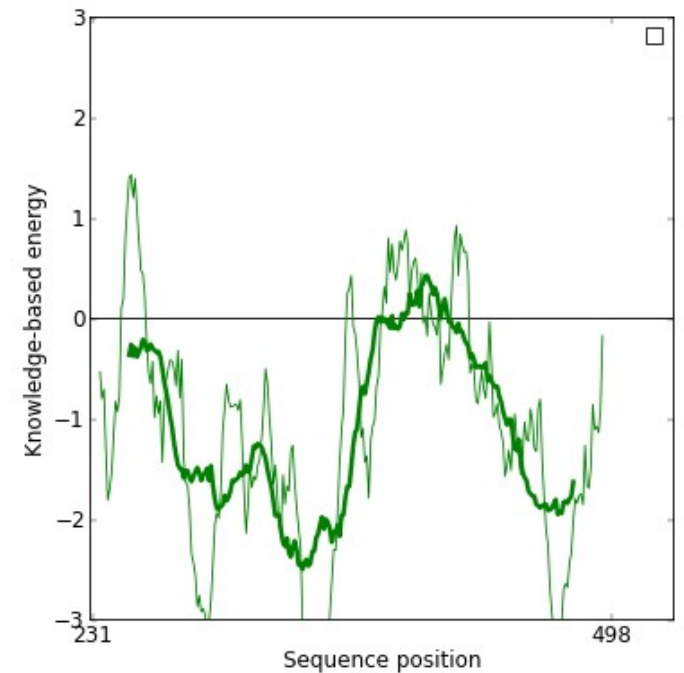
Prosall



Abl (1IEP)



Lck (model)



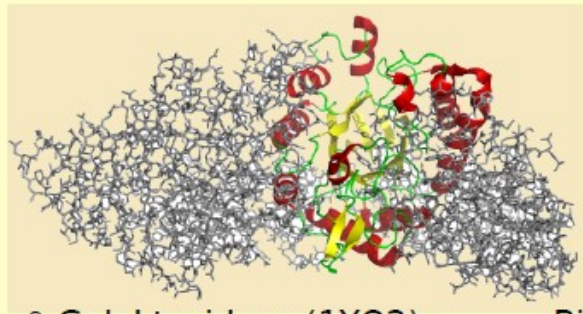
Lck (2PL0)

Předpověď struktur proteinů:

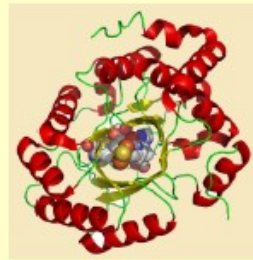
Fold recognition, threading

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

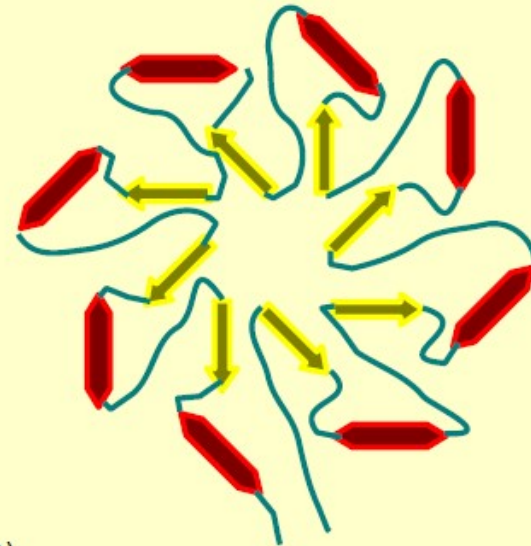
TIM barel



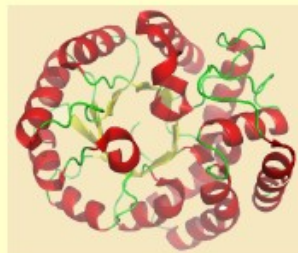
β -Galaktosidasa (1YQ2)



Biotin synthasa (1R30)



Triosafofátisomerasa (8TIM)



Dihydropikolinátsynth. (2ATS)



Luciferasa (1LUC)

Předpověď struktur proteinů:

Ab initio, de novo

- nativní struktura má určité vlastnosti

19:47:27 GMT

fold.it BETA
Solve Puzzles for Science

PUZZLES BLOG CATEGORIES FEEDBACK GROUPS FORUM PLAYERS WIKI FAQ RECIPES ABOUT CONTESTS CREDITS

Click to learn how you contribute to science by playing Foldit.

What's New

Developer Preview Release Soon

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.

- * 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.

GET STARTED: DOWNLOAD

Win Beta
Windows (XP/Vista/7)

Mac Beta
OSX (Intel 10.5 or later)

Linux Beta
Linux (64-bit)

Are you new to Foldit? [Click here.](#)

Are you an educator? [Click here.](#)

SEARCH

Google Search Only search fold.it

RECOMMEND FOLDIT

Send

USER LOGIN

Username: *

Password: *

Log in

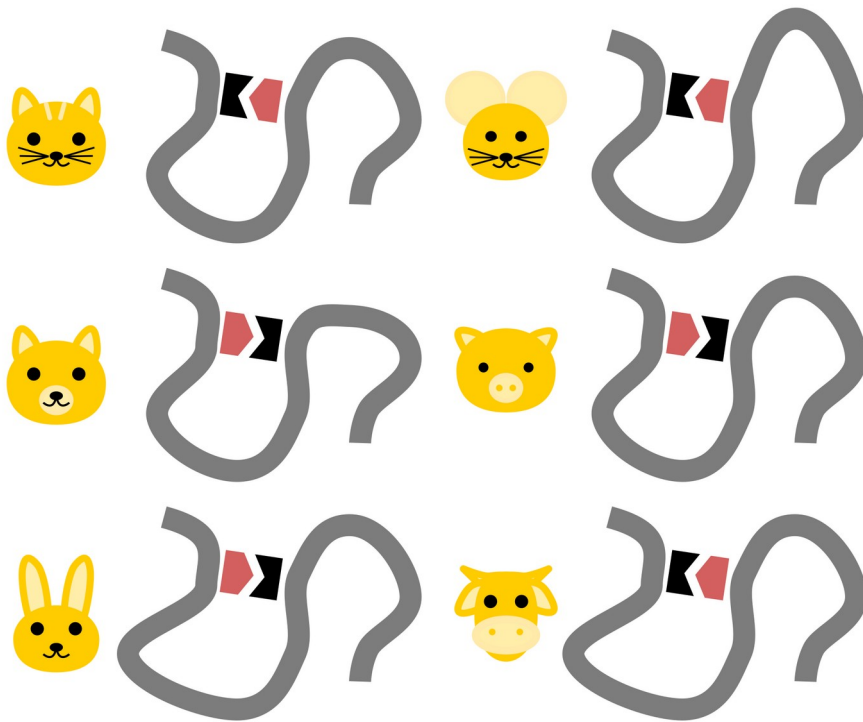
- [Create new account](#)
- [Request new password](#)

<http://fold.it>

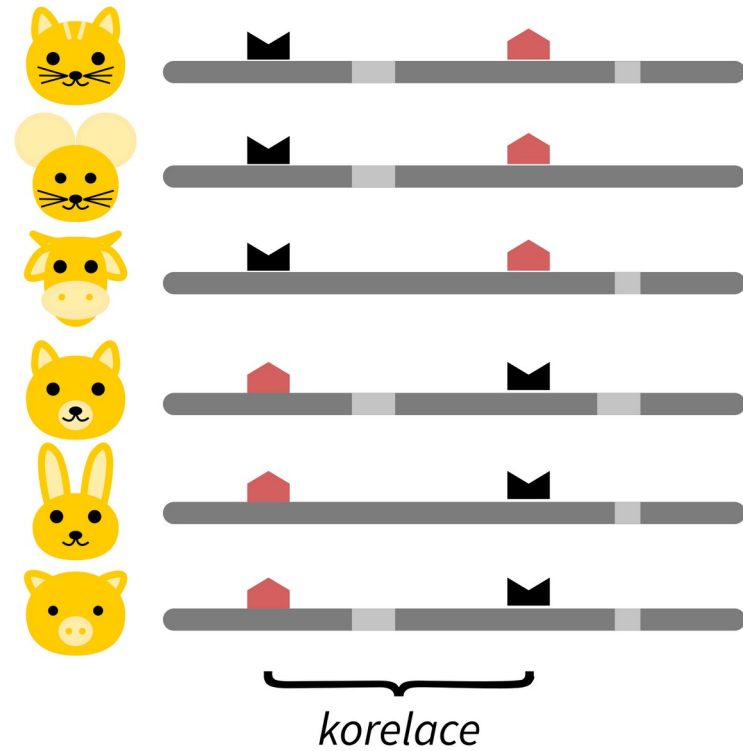
Alphafold:

- koevoluce

3D struktura

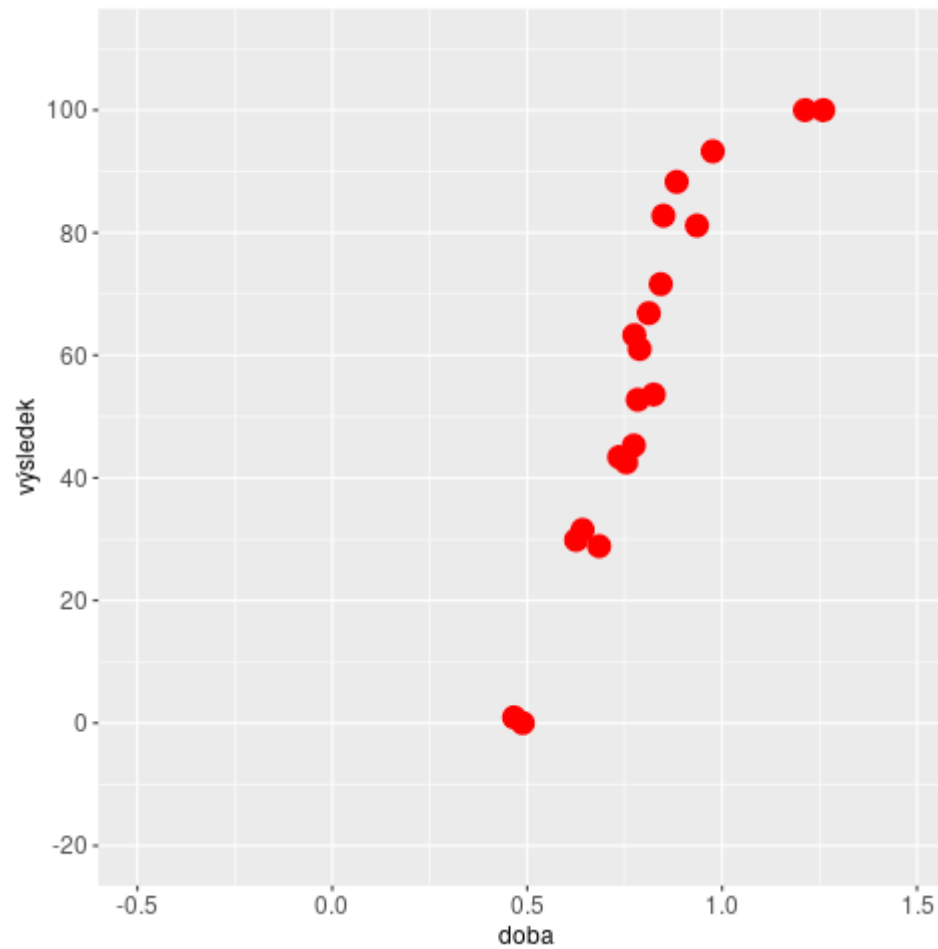


zarovnání sekvencí



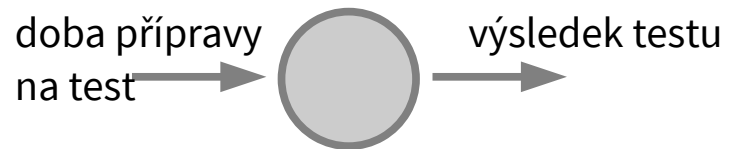
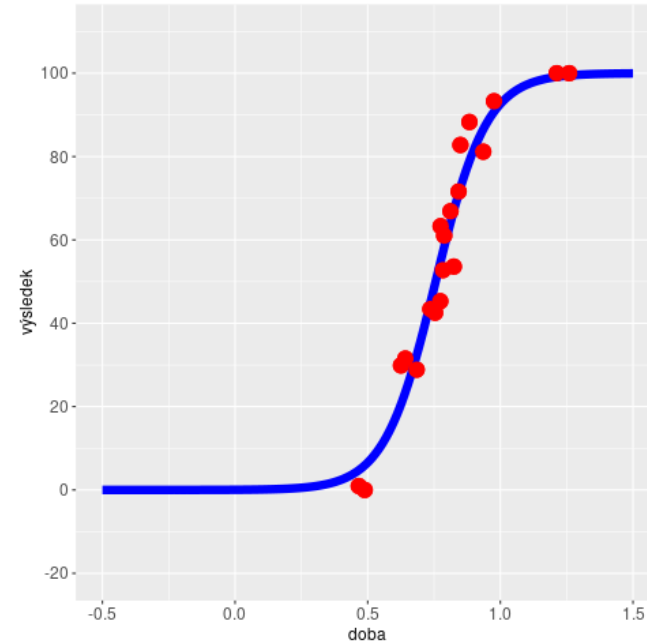
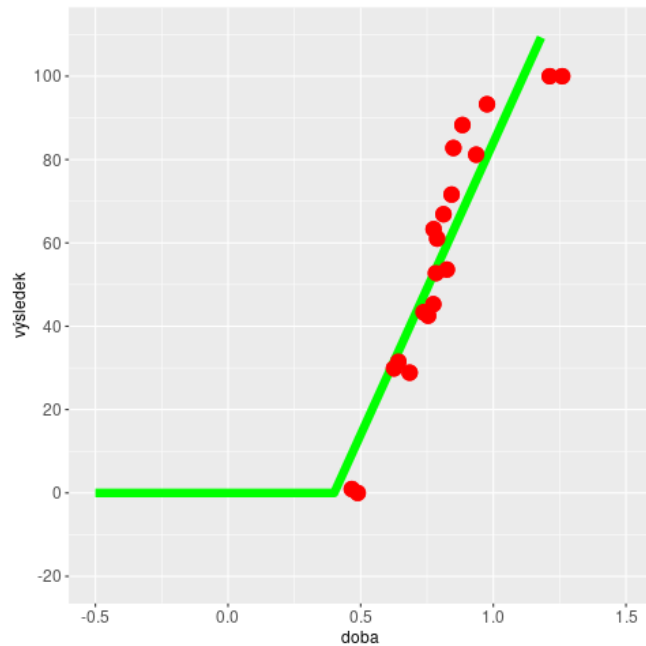
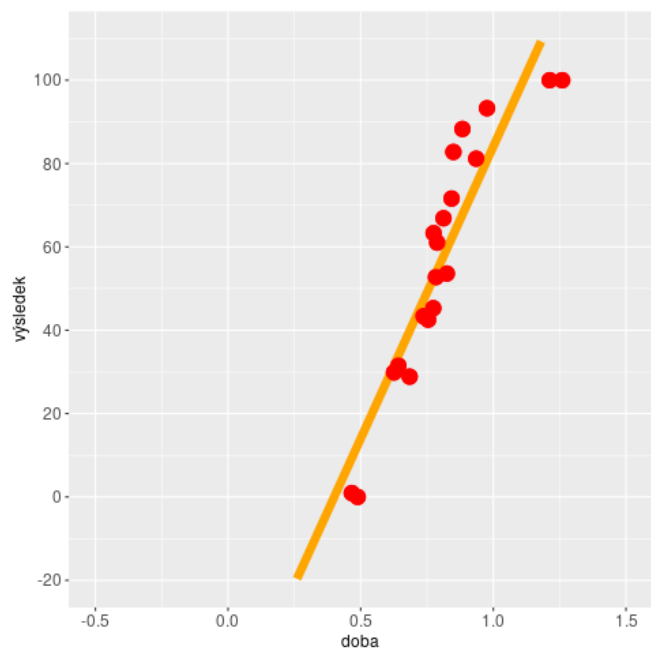
Alphafold:

- neuronové sítě



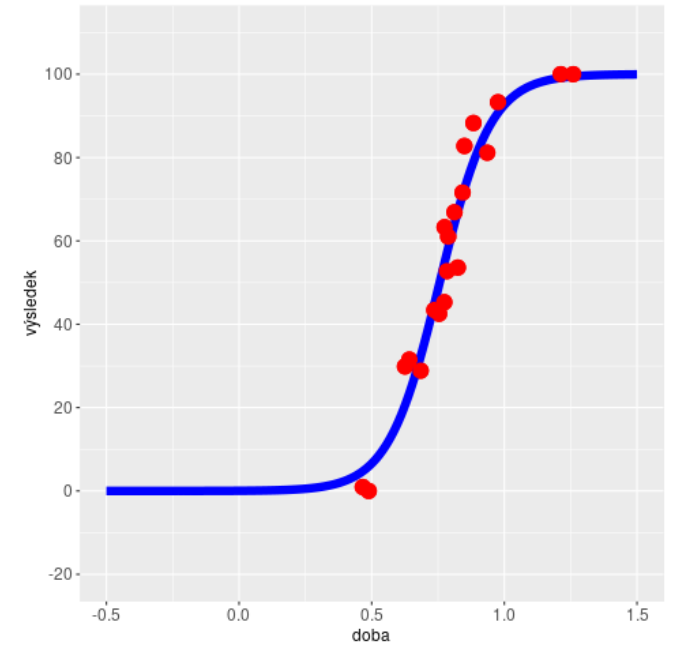
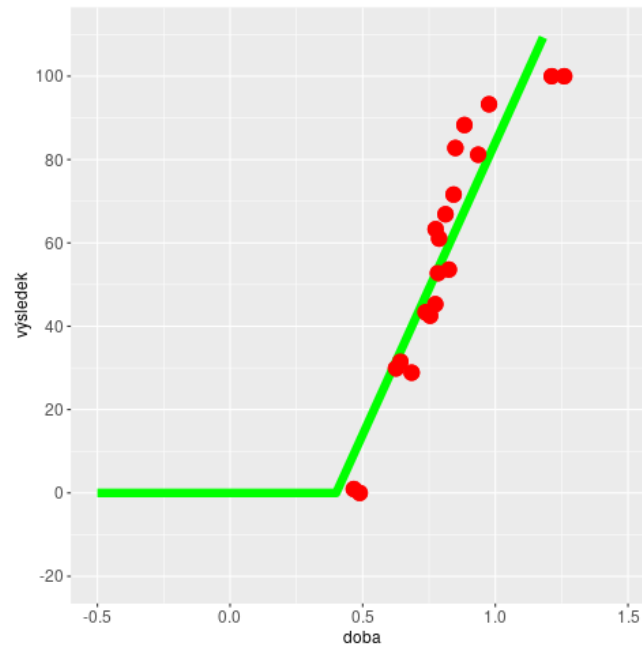
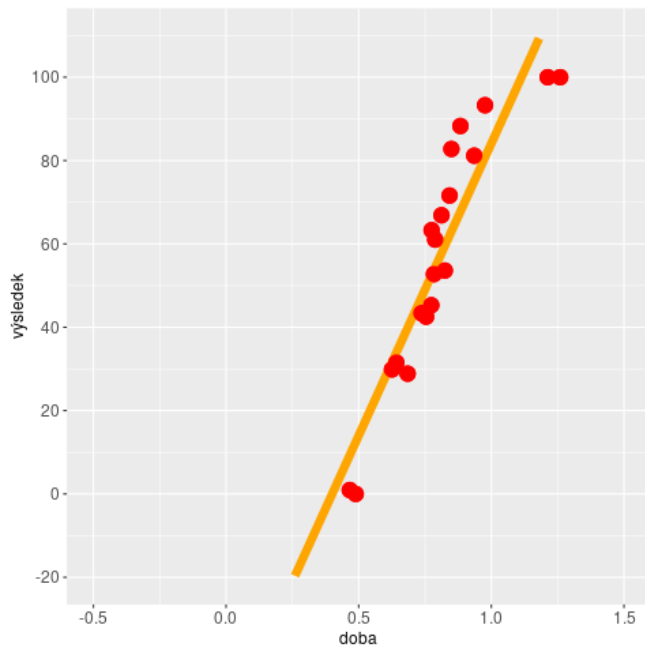
Alphafold:

- neuronové sítě

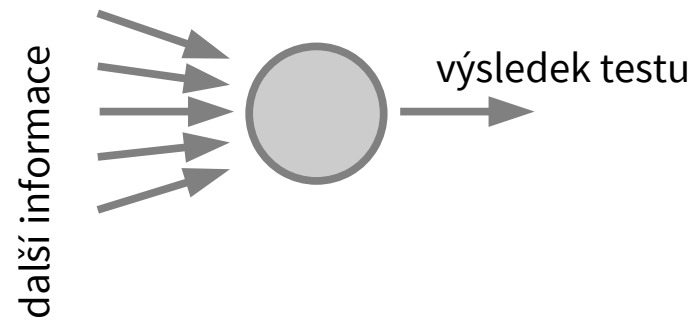


Alphafold:

- neuronové sítě

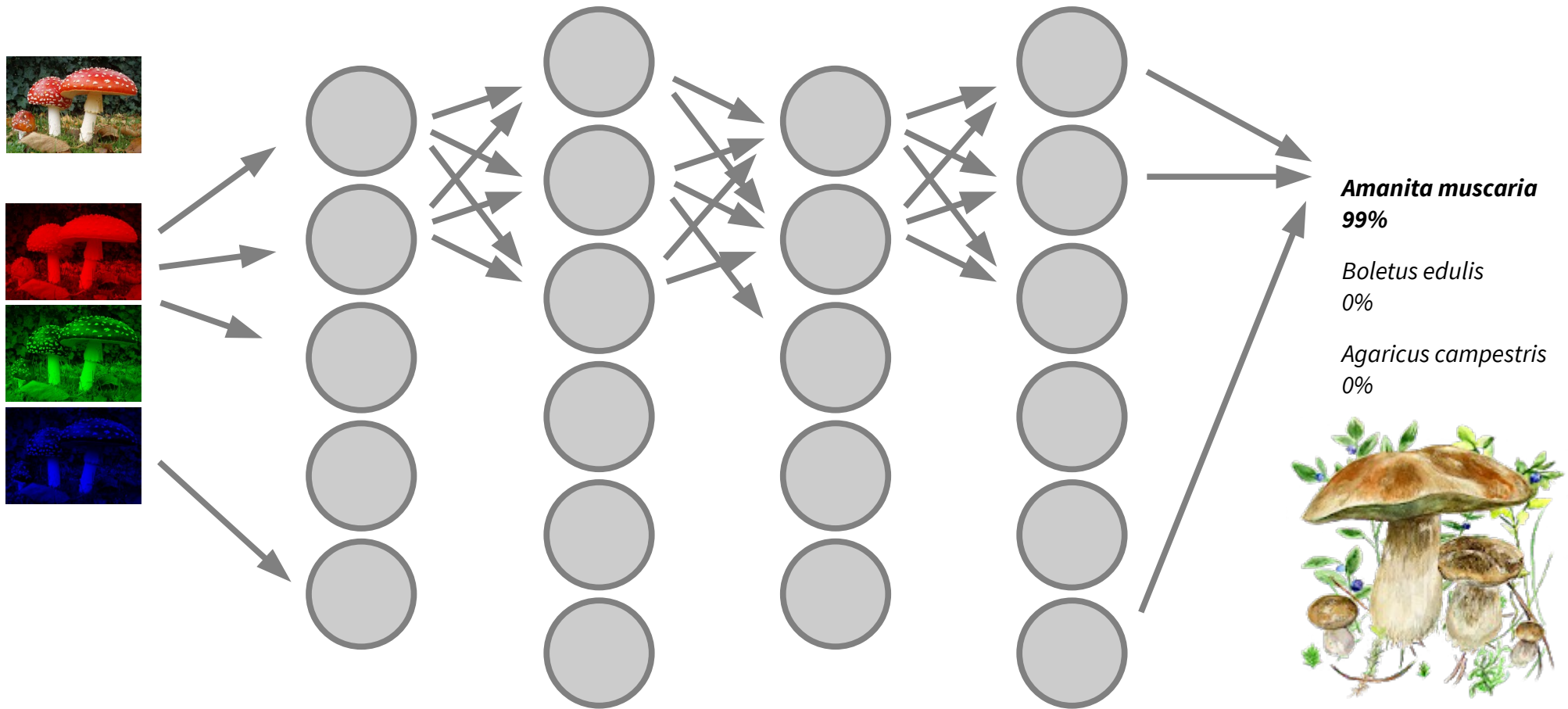


doba přípravy
na test



Alphafold:

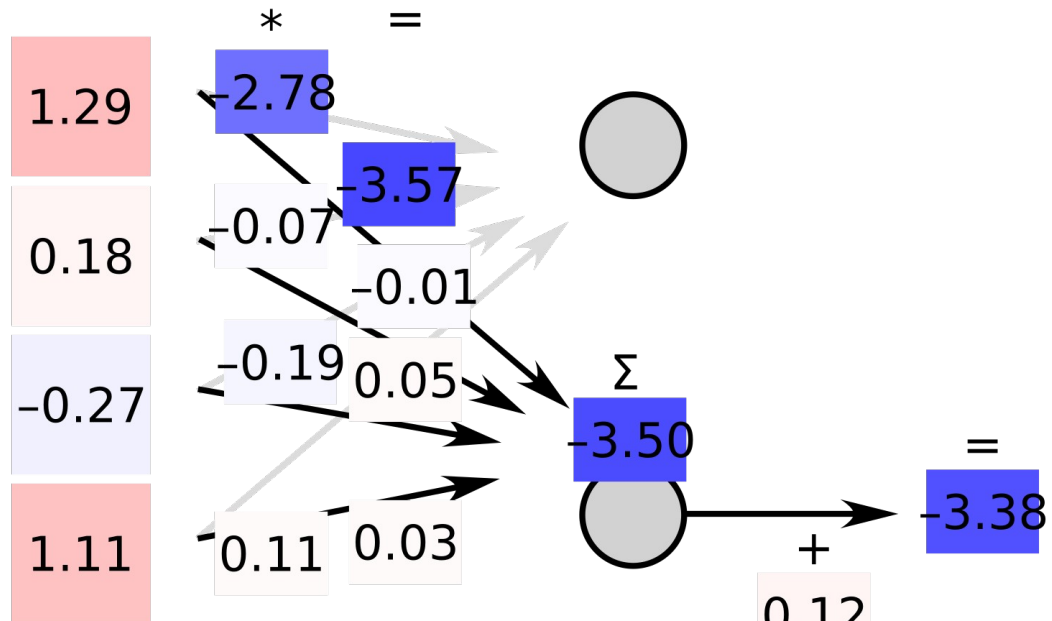
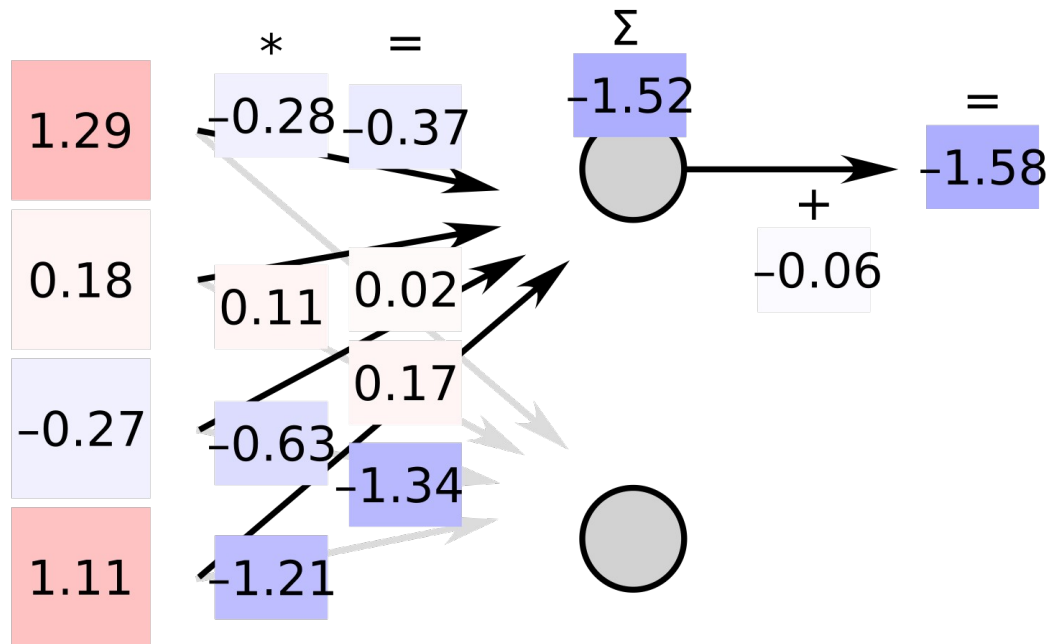
- neuronové sítě



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

Alphafold:

- neuronové sítě

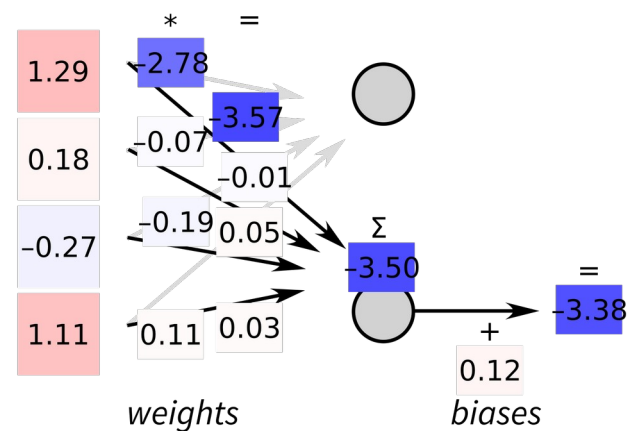
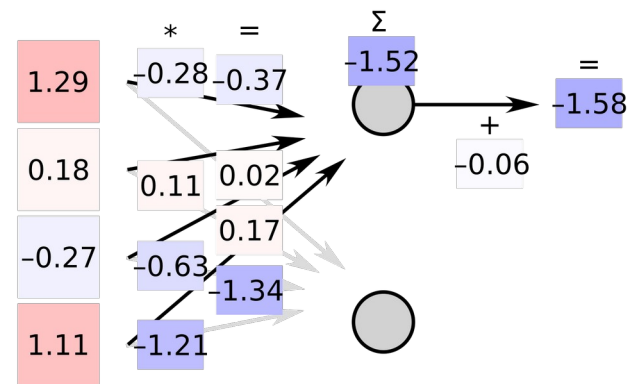
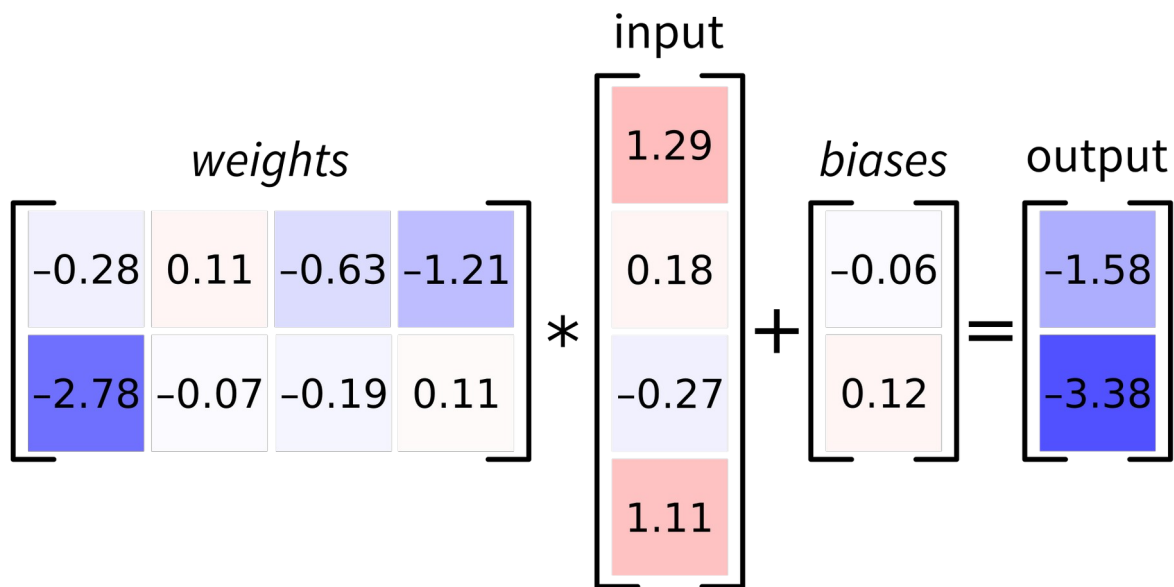


weights

biases

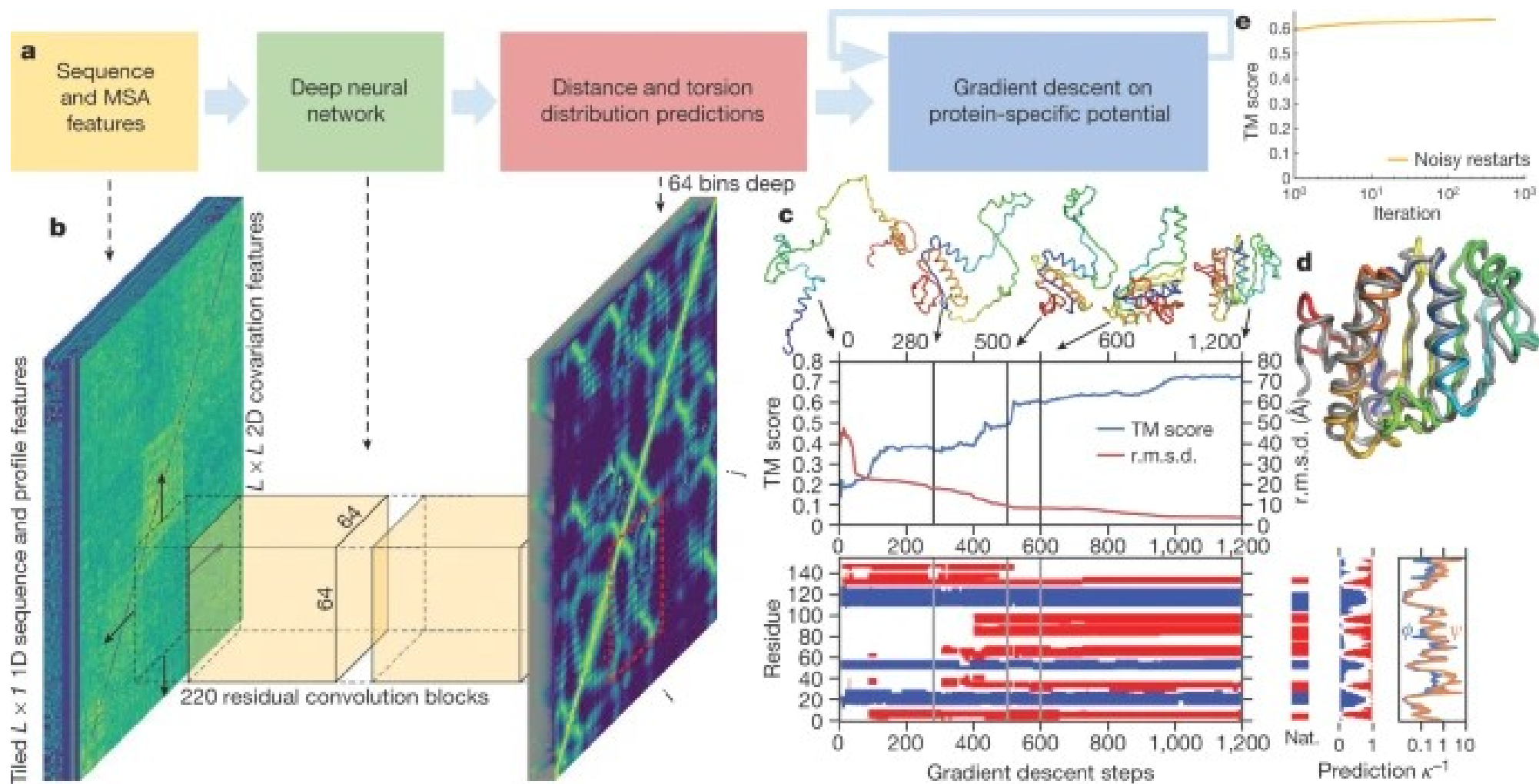
Alphafold:

- neuronové sítě



AlphaFold:

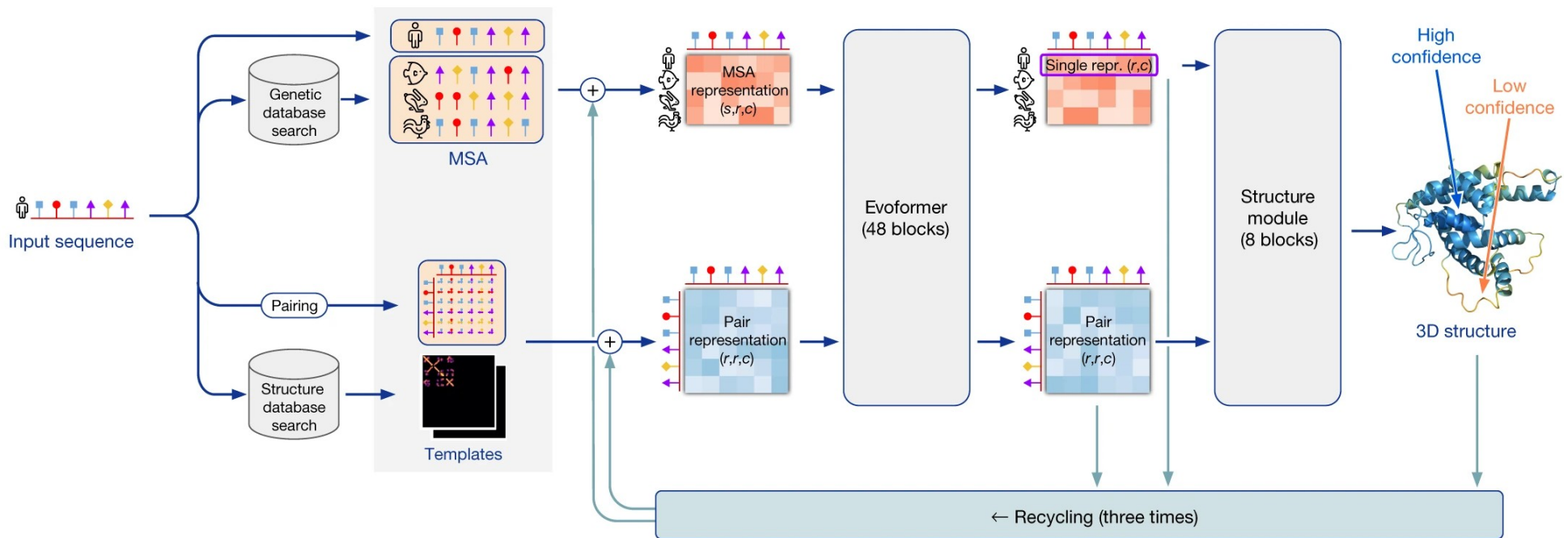
- AF1



Senior, A.W., Evans, R., Jumper, J. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020). <https://doi.org/10.1038/s41586-019-1923-7>

AlphaFold:

- AF2



Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). <https://doi.org/10.1038/s41586-021-03819-2>

Jazykové modely:



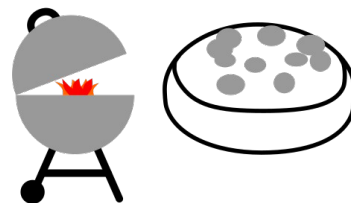
grilovaný hermelín



attention



grilled cheese



královský hermelín



attention



royal ermine fur

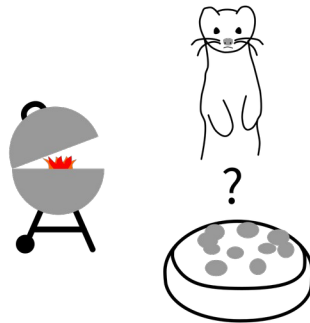


K



Jazykové modely:

 grilovaný hermelín



$L = 2$

grilovaný
hermelín

0.79	0.29
0.34	-0.57
0.8	0.46
0.29	-0.74
...	...
0.67	-0.54
-0.38	-0.36

query

-0.28	-0.06
-0.14	0.04
-0.51	-0.68
0.41	-0.25
...	...
0.25	-0.25
-0.45	-0.56

query weights 512

-0.39	0.01	0.13	-0.74	...	0.13	0.01
0.24	-0.12	-0.1	0.58	...	0.43	-0.09
-0.36	0.15	-0.07	0.14	...	-0.2	-0.1
0.18	0.15	0.5	0.2	...	-0.21	0.21
...
0.02	0.33	0.16	-0.18	...	-0.3	-0.26
0.01	0.28	-0.39	0.42	...	-0.08	0.15

query

-0.28	-0.06
-0.14	0.04
-0.51	-0.68
0.41	-0.25
...	...
0.25	-0.25
-0.45	-0.56

key weights 512

-0.35	-0.07	0.12	-0.36	...	-0.55	-0.1
0.11	0.09	-0.34	0.23	...	-0.04	0.15
-0.24	-0.06	0.3	-0.35	...	0.14	0.14
0.1	-0.06	0.3	0.26	...	0	0.06
...
-0.11	0.04	-0.39	0.42	...	-0.08	-0.45
0.14	-0.11	0.13	0.26	...	-0.8	-0.02

key

-0.19	0.16
-0.08	0.26
-0.23	-0.9
0	0.31
...	...
-0.02	-0.03
-0.53	-0.34

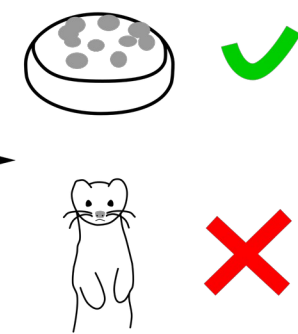
key^T

-0.19	-0.08	-0.23	0	...	-0.02	-0.53
0.16	0.26	-0.9	0.31	...	-0.03	-0.34

attention weights

value weights → values → FFNN

0.65	0.73
0.41	0.47




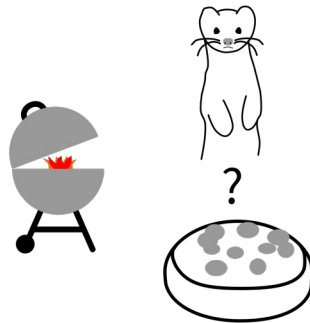
 grilled cheese

Jazykové modely:

a	b	product	a	b	product
-0.05	0.07	0	-0.9	-0.88	0.79
-1.01	0.64	-0.65	-0.17	-0.2	0.03
-1.15	1.67	-1.91	0.75	0.74	0.55
-0.11	-1.08	0.12	0.86	0.8	0.69
0.81	1.11	0.9	0.4	0.45	0.18
0.62	-0.62	-0.38	0.52	0.54	0.28
-1.13	-2.05	2.31	-1.02	-1	1.02
0.95	-0.49	-0.47	0.23	0.21	0.05
		Σ			Σ
dot product		-0.08	dot product		3.6

Jazykové modely:

 grilovaný hermelín



$L = 2$

grilovaný hermelín

0.79	0.29
0.34	-0.57
0.8	0.46
0.29	-0.74
...	...
0.67	-0.54
-0.38	-0.36

query

-0.28	-0.06
-0.14	0.04
-0.51	-0.68
0.41	-0.25
...	...
0.25	-0.25
-0.45	-0.56

query weights 512

-0.39	0.01	0.13	-0.74	...	0.13	0.01
0.24	-0.12	-0.1	0.58	...	0.43	-0.09
-0.36	0.15	-0.07	0.14	...	-0.2	-0.1
0.18	0.15	0.5	0.2	...	-0.21	0.21
...
0.02	0.33	0.16	-0.18	...	-0.3	-0.26
0.01	0.28	-0.39	0.42	...	-0.08	0.15

query

-0.28	-0.06
-0.14	0.04
-0.51	-0.68
0.41	-0.25
...	...
0.25	-0.25
-0.45	-0.56

key weights 512

-0.35	-0.07	0.12	-0.36	...	-0.55	-0.1
0.11	0.09	-0.34	0.23	...	-0.04	0.15
-0.24	-0.06	0.3	-0.35	...	0.14	0.14
0.1	-0.06	0.3	0.26	...	0	0.06
...
-0.11	0.04	-0.39	0.42	...	-0.08	-0.45
0.14	-0.11	0.13	0.26	...	-0.8	-0.02

key

-0.19	0.16
-0.08	0.26
-0.23	-0.9
0	0.31
...	...
-0.02	-0.03
-0.53	-0.34

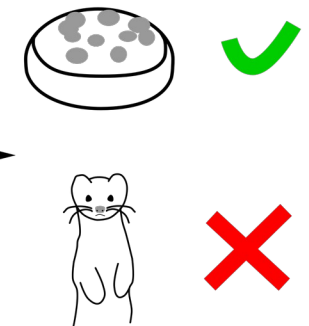
key^T

-0.19	-0.08	-0.23	0	...	-0.02	-0.53
0.16	0.26	-0.9	0.31	...	-0.03	-0.34

attention weights

value weights → values → FFNN

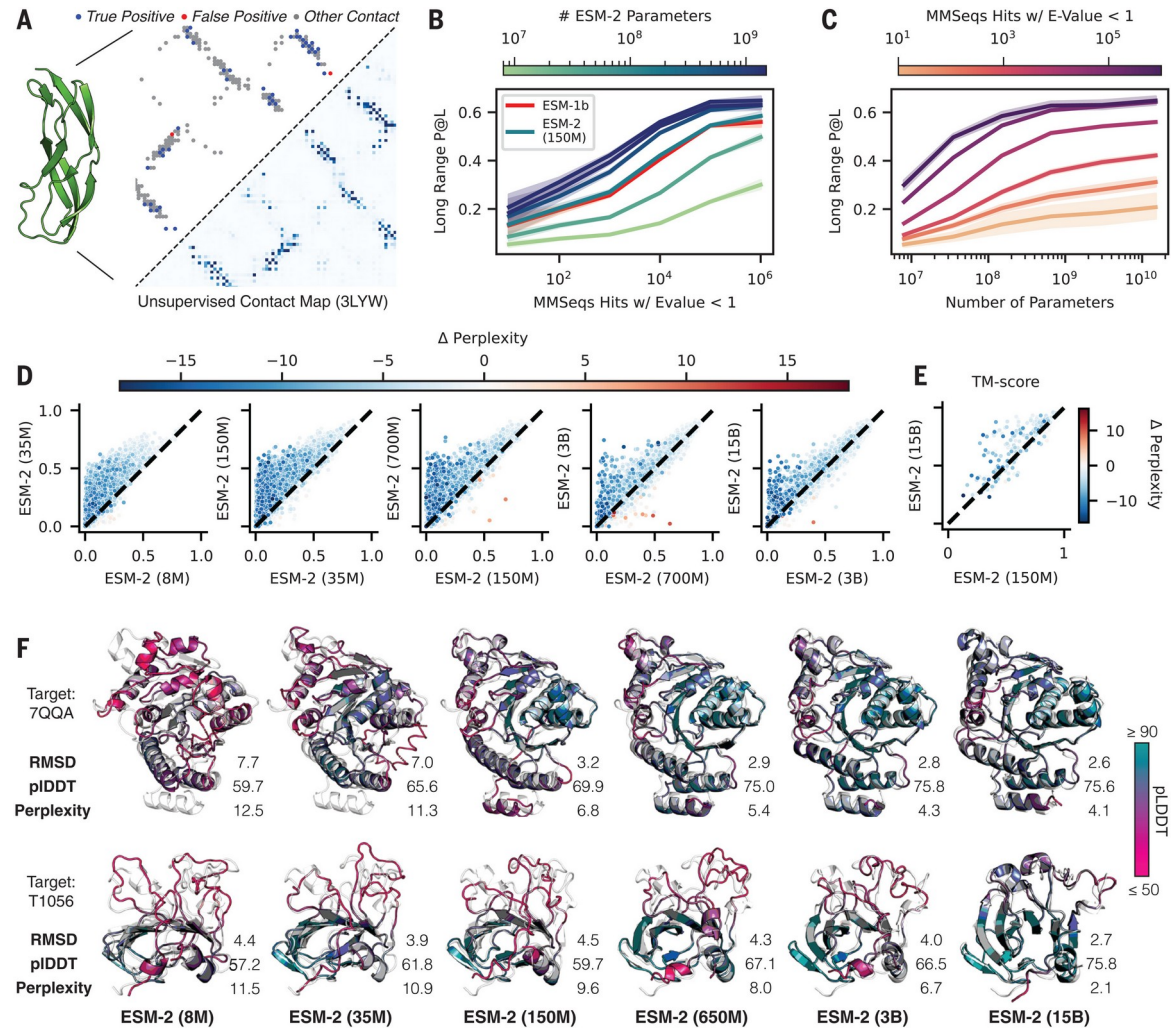
0.65	0.73
0.41	0.47



 grilled cheese

Jazykové modely:

- ESMfold



Zeming Lin et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123-1130 (2023). <https://doi.org/10.1126/science.ade2574>

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



Protein Structure Prediction Center



Menu

- [Home](#)
- [FORCASP Forum](#)
- [PC Login](#)
- [PC Registration](#)
- ▼ [CASP Experiments](#)
 - [CASP ROLL](#)
 - ▼ [CASP11 \(2014\)](#)
 - [Home](#)
 - [Registration](#)
 - [My CASP11 profile](#)
 - ▼ [Targets](#)
 - [Target List](#)
 - [Target Submission](#)
- [CASP10 \(2012\)](#)
- [CASP9 \(2010\)](#)
- [CASP8 \(2008\)](#)
- [CASP7 \(2006\)](#)
- [CASP6 \(2004\)](#)
- [CASP5 \(2002\)](#)
- [CASP4 \(2000\)](#)
- [CASP3 \(1998\)](#)
- [CASP2 \(1996\)](#)
- [CASP1 \(1994\)](#)
- ▶ [Initiatives](#)
- ▶ [Data Archive](#)
- [Local Services](#)
- [Proceedings](#)
- [Feedback](#)
- [Assessors](#)

Welcome to the Protein Structure Prediction Center!

Our goal is to help advance the methods of identifying protein structure from sequence. The Center has been organized to provide the means of objective testing of these methods via the process of blind prediction. The Critical Assessment of protein Structure Prediction (CASP) experiments aim at establishing the current state of the art in protein structure prediction, identifying what progress has been made, and highlighting where future effort may be most productively focused.

There have been ten previous CASP experiments. The eleventh experiment will start in May 2014. Description of these experiments and the full data (targets, predictions, interactive tables with numerical evaluation results, dynamic graphs and prediction visualization tools) can be accessed following the links:

[CASP1 \(1994\)](#) | [CASP2 \(1996\)](#) | [CASP3 \(1998\)](#) | [CASP4 \(2000\)](#) | [CASP5 \(2002\)](#) | [CASP6 \(2004\)](#) | [CASP7 \(2006\)](#) | [CASP8 \(2008\)](#) | [CASP9 \(2010\)](#) | [CASP10 \(2012\)](#) | [CASP11 \(2014\)](#)

Raw data for the experiments held so far are archived and stored in our [data archive](#).

In November 2011 we have opened a new rolling CASP experiment for all-year-round testing of ab initio modeling methods:

[CASP ROLL](#)

Details of the experiments have been published in a scientific journal *Proteins: Structure, Function and Bioinformatics*. [CASP proceedings](#) include papers describing the structure and conduct of the experiments, the numerical evaluation measures, reports from the assessment teams highlighting state of the art in different prediction categories, methods from some of the most successful prediction teams, and progress in various aspects of the modeling.

Prediction methods are assessed on the basis of the analysis of a large number of blind predictions of protein structure. Summary of numerical evaluation of the methods tested in the latest CASP experiment can be found [on this web page](#). The main numerical measures used in evaluations are described in the papers [\[1\]](#), [\[2\]](#). The latter paper also contains explanations of data handling procedures and guidelines for navigating the data presented on this website.

Some of the best performing methods are implemented as [fully automated servers](#) and therefore can be used by public for protein structure modeling.

To proceed to the pages related to the latest CASP experiments click on the logo below:



FORCASP

Message Board

CASP11 dry run for servers

[Dear CASP participants, Registration for CASP11 is currently under way and more than 60 research groups have already registered for the experiment. We are starting checking connectivity and correct ...](#)

CASP11 registration opens March 31

[Dear CASP Participants, Exiting news: new CASP experiment is just around the corner! We hope that you are full of enthusiasm and anxiety \(as we are\) and have your computers greased and warmed up. ...](#)

Resuming CASP ROLL

[Dear CASPs, Best regards for all of you in the New Year! Hoping that you had good rest after the CASP10 experiment and meeting, we are resuming CASP ROLL with two new targets later this week. ...](#)