



**IOCB
PRAGUE**



Pluskal Group
IOCB Prague



Faculty of Science

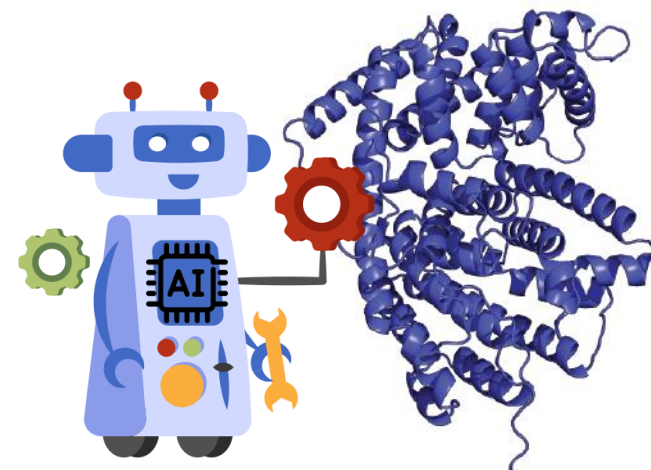


AI pro enzymy: Jak pochopit chemii přírody - a překonat ji

Matouš Soldát | Letní škola Czech AI Factory | 10.9.2026

Introduction

2



Protein engineering

- Co je to **enzym**?
- **Hledání** vhodných přírodních enzymů
- **Vylepšování** přírodních enzymů
- **Generování** nových enzymů

Co je to enzym?

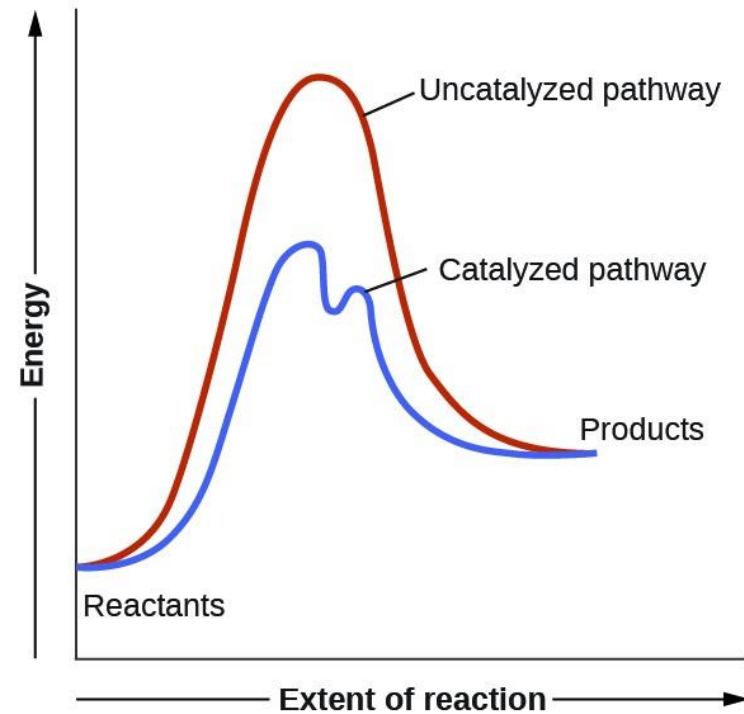
Co je to **enzym**?

5

Katalyzátor = látka nebo zařízení, které urychluje chemickou reakci, aniž by se samo během reakce spotřebovalo.



Enzym = **biologický** katalyzátor

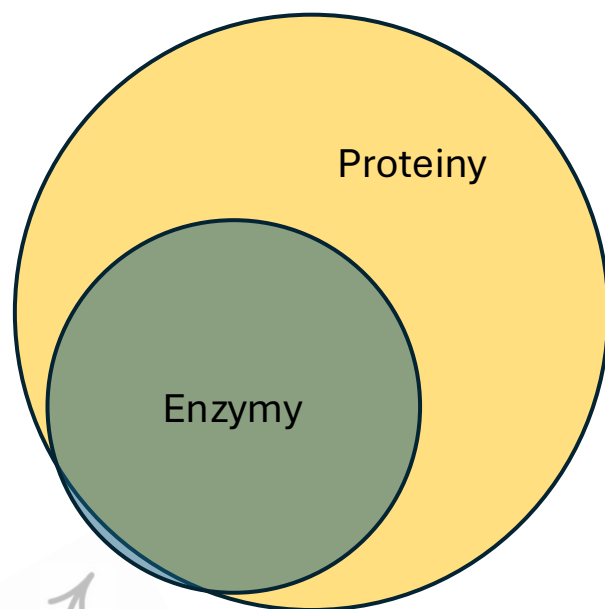


Potential energy diagram

Jaké biomolekuly mohou být enzymy?

6

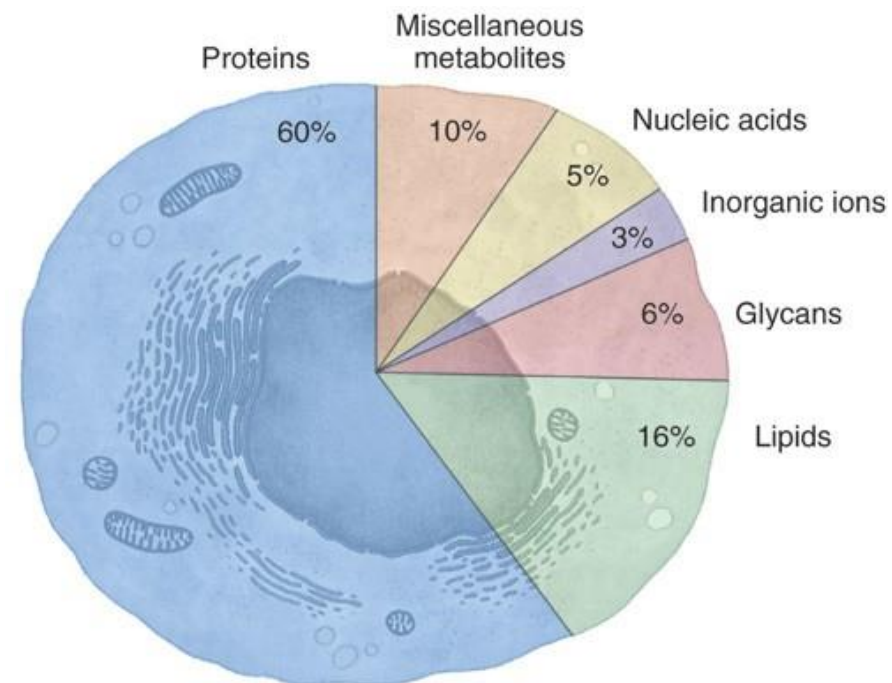
Proteiny = polypeptidy = bílkoviny



Enzymy $\wedge$ $\neg$ Proteiny



- Ribozymes (RNA-enzymes)
- Deoxyribozymes (DNA-enzymes)
- ...



Composition of a typical mammalian cell (dry mass)

Prescher, Jennifer A., and Carolyn R. Bertozzi. "Chemistry in living systems." Nature chemical biology 1.1 (2005): 13-21.

7

Proteins are “mini robots” that make us alive



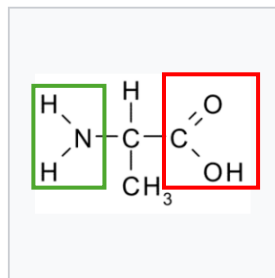
Image: D. S. Goodsell, Chloroplast (2011), RCSB PDB (CC BY 4.0)

7

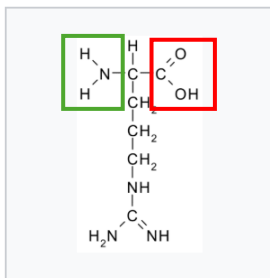
Z čeho se skládají enzymy (a ostatní proteiny)?

8

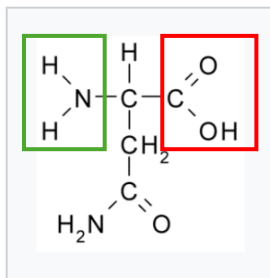
Aminokyseliny = základní stavební bloky proteinů:



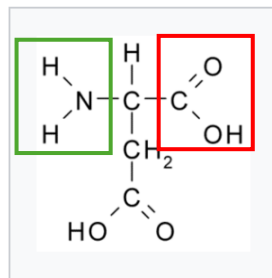
Alanin



Arginin



Asparagin

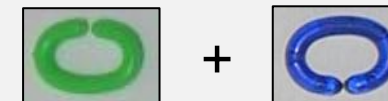
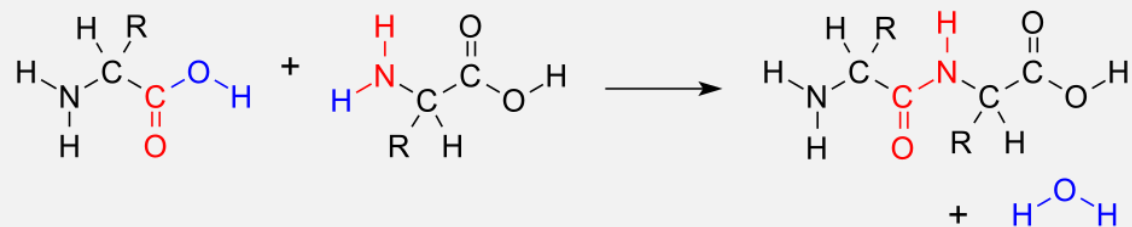


Kyselina asparagová

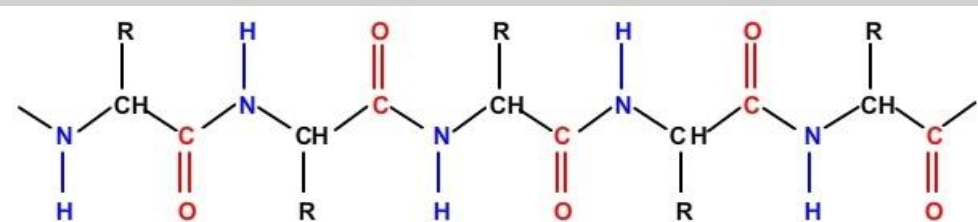
Celkem
20 aminokyselin



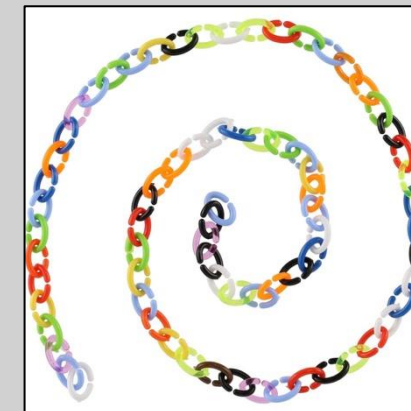
Peptidová vazba:



Peptidový řetězec:



https://www.studiumbiochemie.cz/prirodni_latky_bilkoviny.html



3D struktura enzymů (a ostatních proteinů)

9

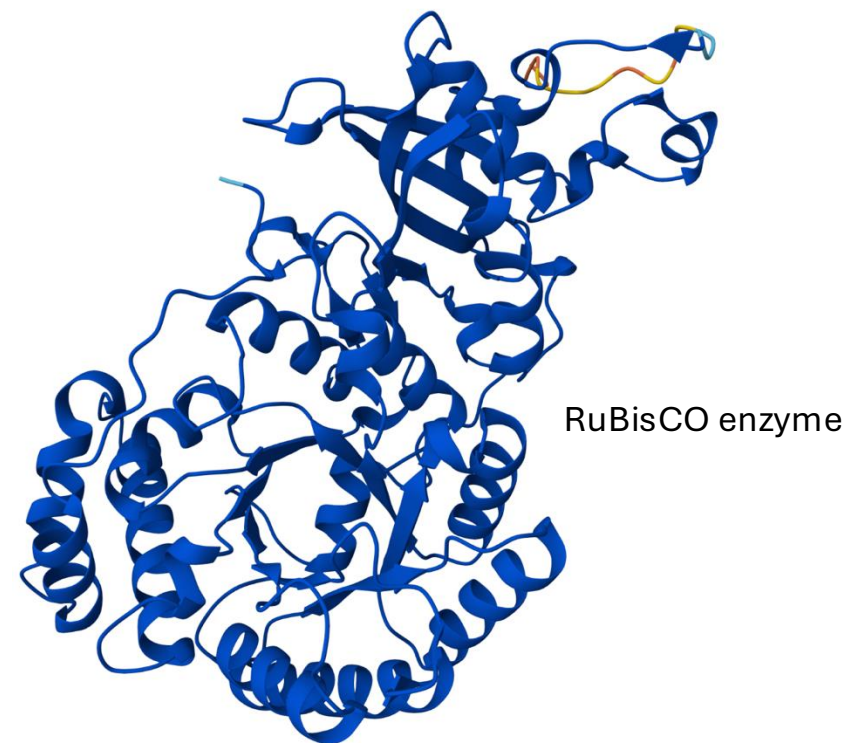
Sekvence

MAPTVMASSATSVAPFQGLKSTAGLPV
SRRSTNSGFGNVSNNGGRIKCMQVWPI
EGIKKFETLSYLPPLTVEDLLKQIEYLLRS
KWVPCLEFSKVG FVYRENHRSPGYD
GRYWTMWKLPMFGCTDATQVLKELEE
AKKAYPDAFVRIIGFDNVRQVQLISFIAY
KPPGCEESGGN

Folding



3D struktura

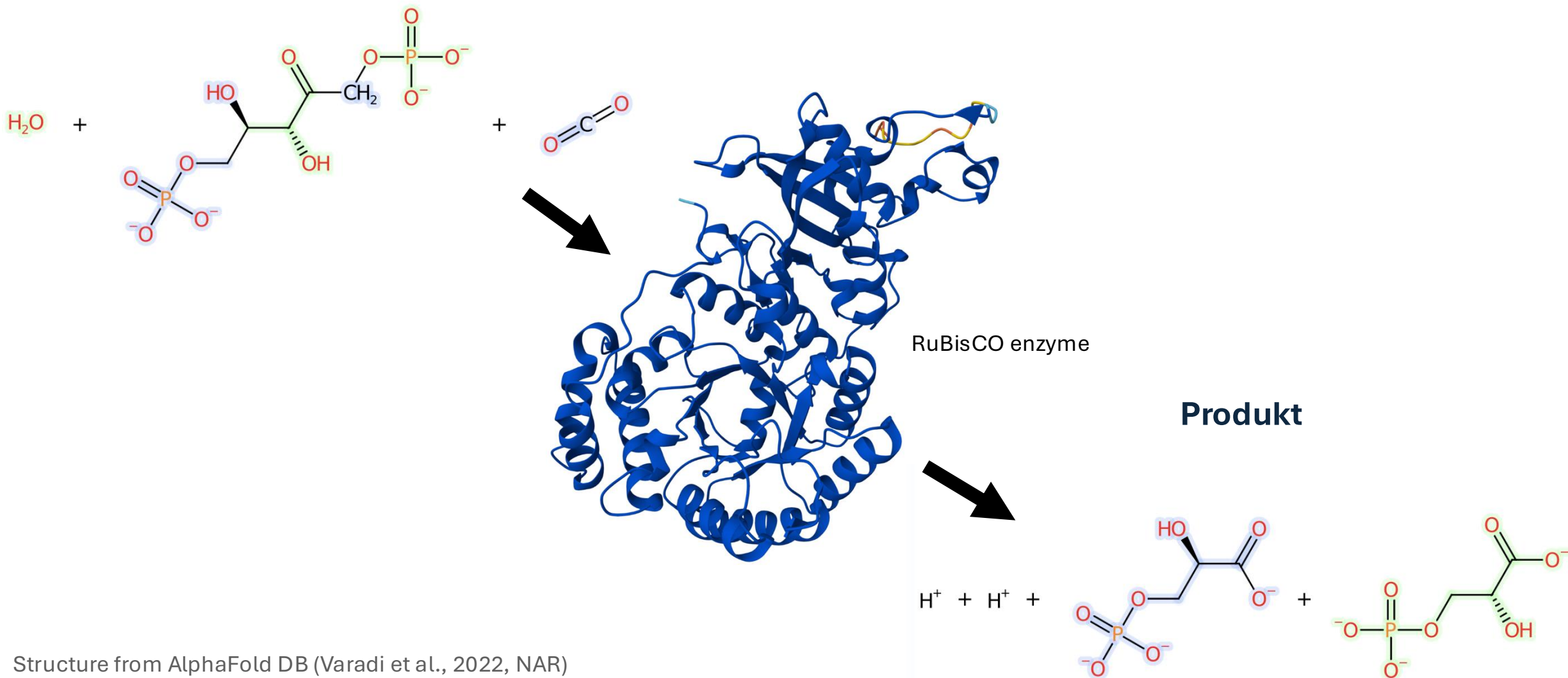


1 písmeno = 1 aminokyselina
Např. M = Methionine, A = Alanine, P = Proline

3D struktura určuje funkci

10

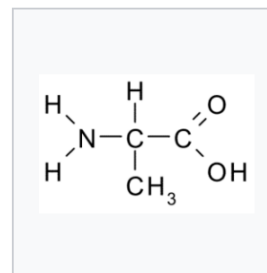
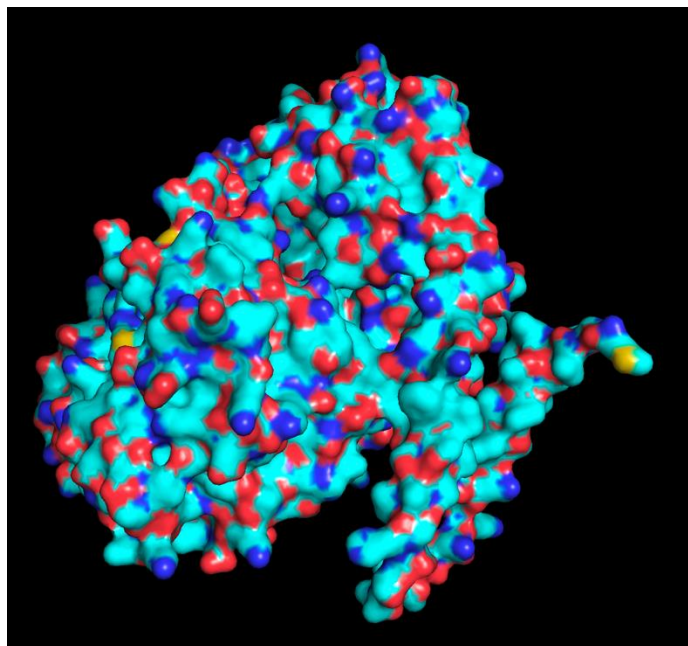
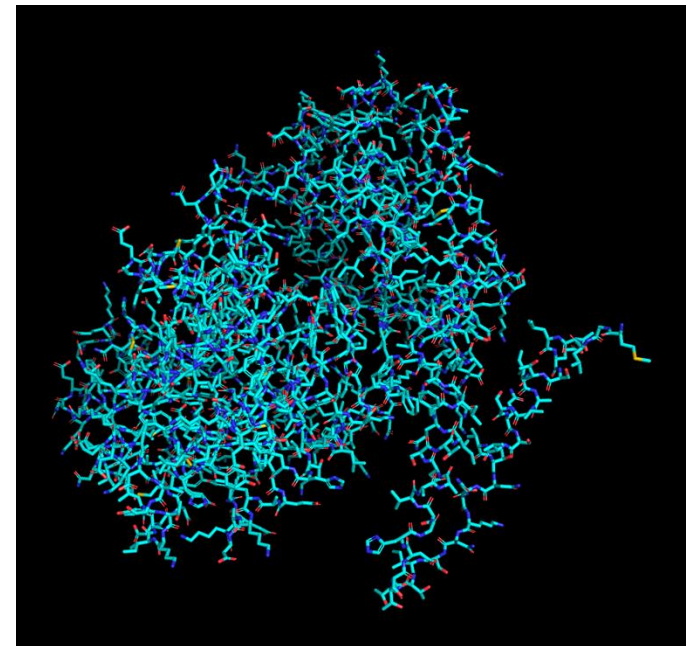
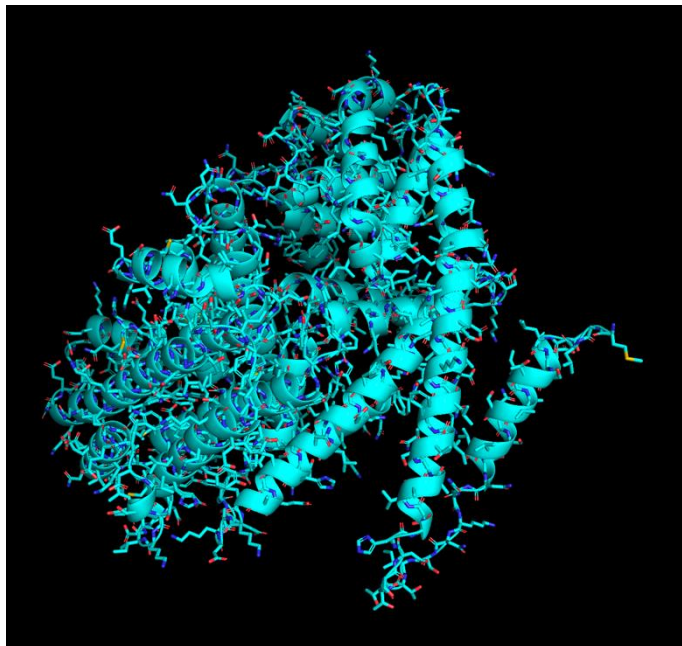
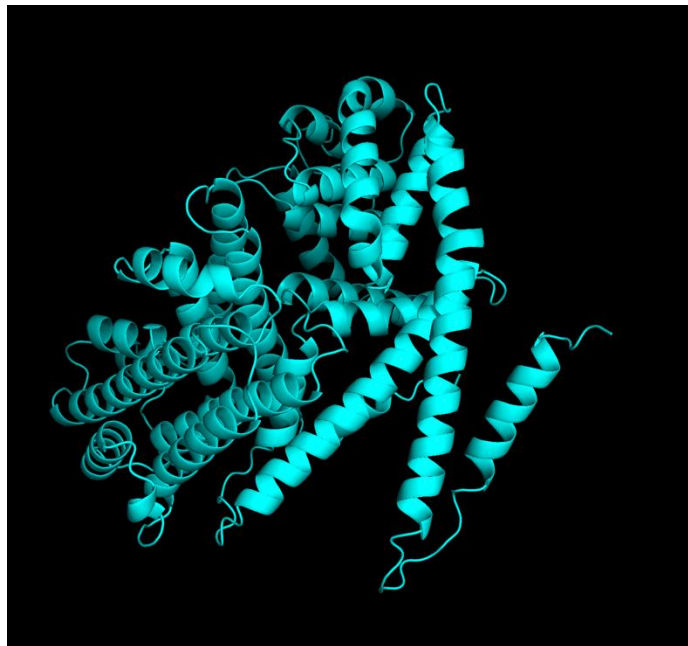
Substrát



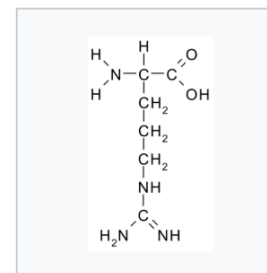
Structure from AlphaFold DB (Varadi et al., 2022, NAR)

Visualizace 3D struktury proteinů

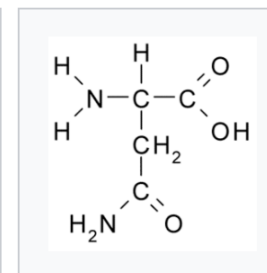
11



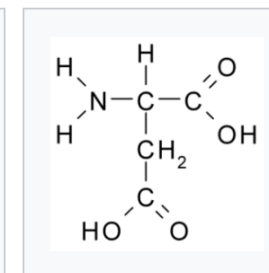
Alanin



Arginin



Asparagin



Kyselina asparagová

atd.

Využití AI k návrhu enzymů

Příklad 1: Hledání vhodných přírodních enzymů

Příroda už vyrobila spoustu enzymů pro nás

13

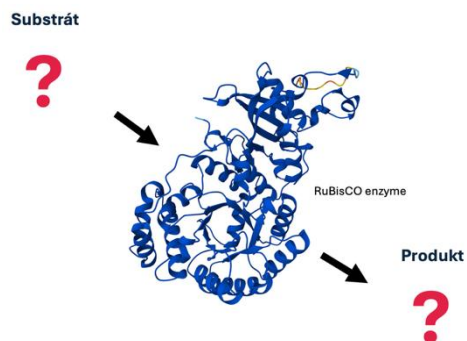
Sekvence

MAPTVMASSATSVAPFQGLKSTAGLPVSR
RSTNSGFGNVSNGGRIKCMQVWP
KFETLSYLPPLTVEDLLKQIEYLLRSKWVP
CLEFSKVG FVYRENHRSPGYDGRYWTM
WKLPMFGCTDATQVLKELEEAKKAYPDA
FVRIIGFDNVRQVQLISFIAYKPPGCEESG
GN

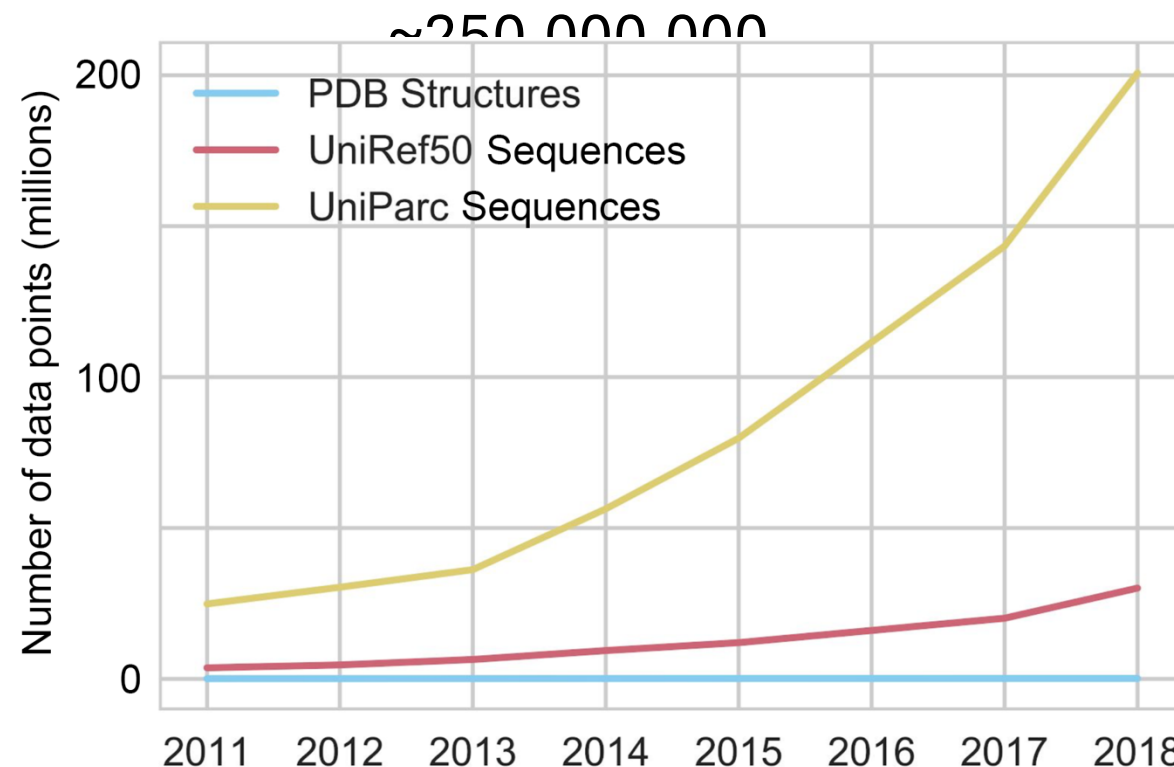
3D struktura



Funkce



Experimentálně ověřeno:



~100,000 ← 0.04 %

Mohammed AlQuraishi. <https://moalquraishi.wordpress.com/2019/04/01/the-future-of-protein-science-will-not-be-supervised/>

Predikce 3D struktury proteinů

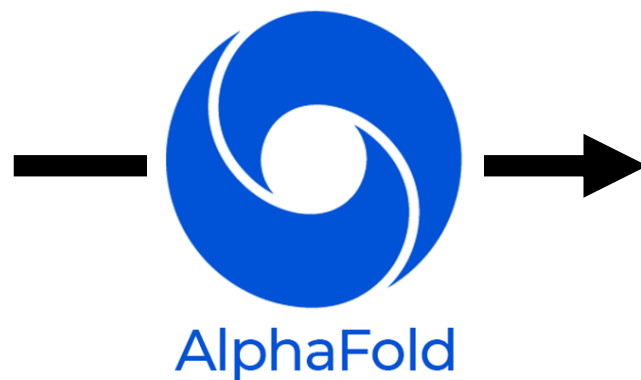
14

Sekvence

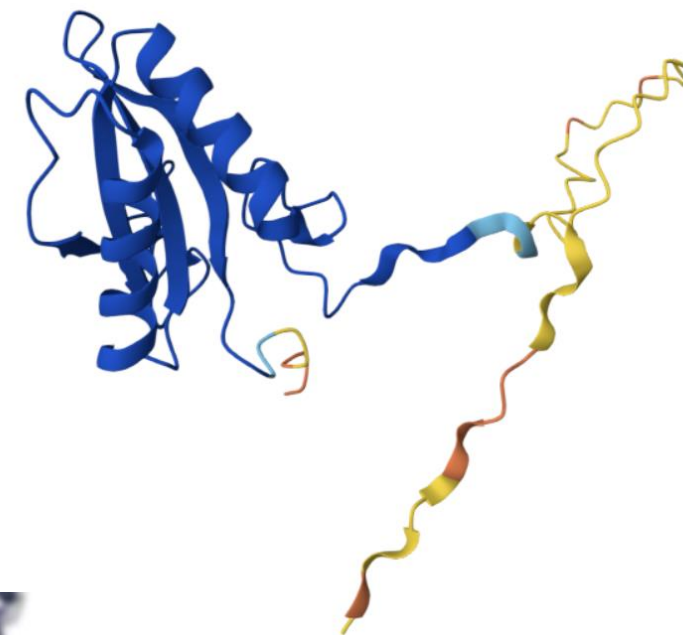
MAPTVMASSATSVAPFQGLKSTAGLPV
SRRSTNSGFGNVSNNGGRIKCMQVWPI
EGIKKFETLSYLPPLTVEDLLKQIEYLLRS
KWVPCLEFSKVG FVYRENHRSPGYD
GRYWTMWKLPMFGCTDATQVLKELEE
AKKAYPDAFVRIIGFDNVRQVQLISFIAY
KPPGCEESGGN



Nobelova cena



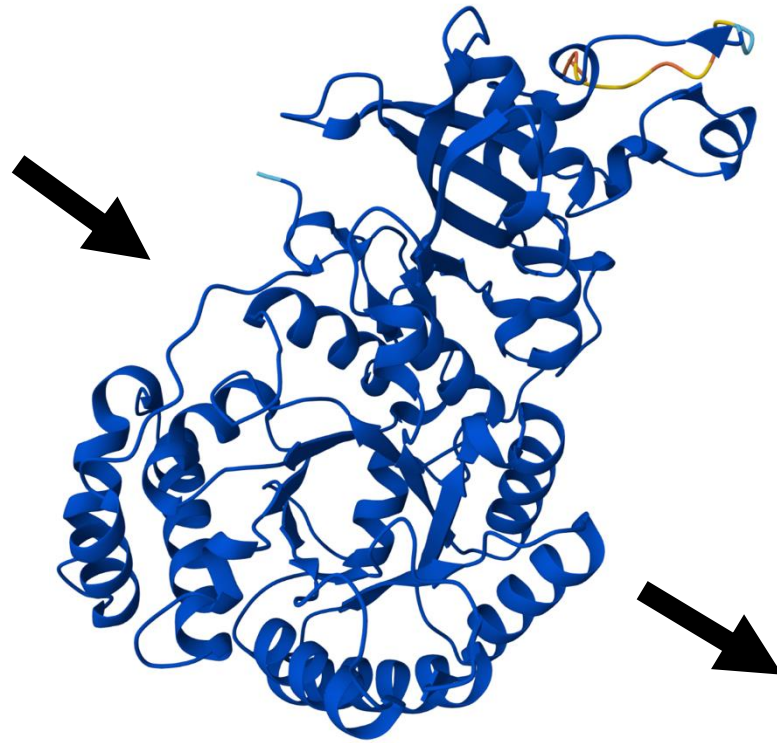
Struktura



Jak predikovat funkci enzymu?

16

Substrát



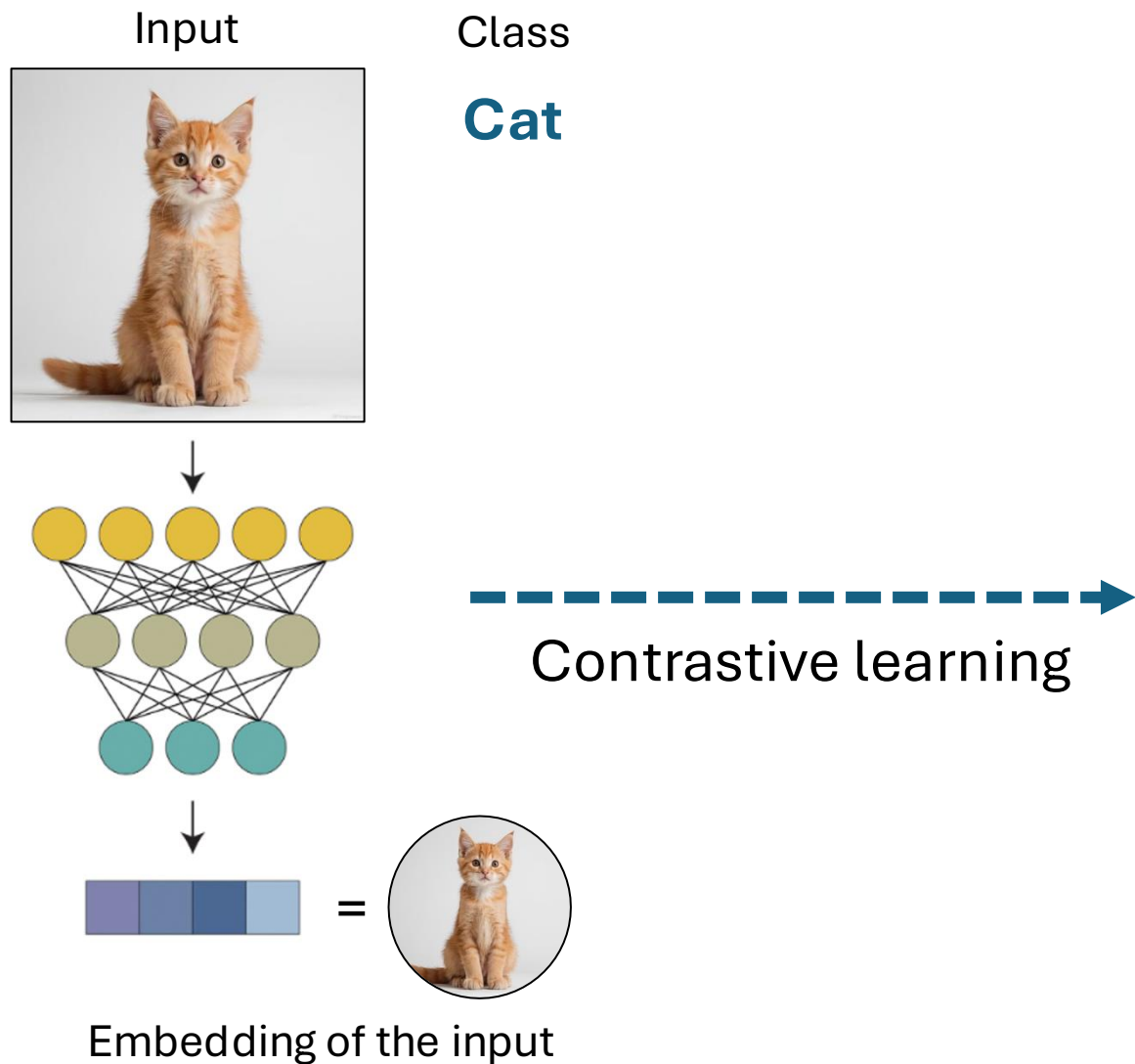
RuBisCO enzyme

Produkt

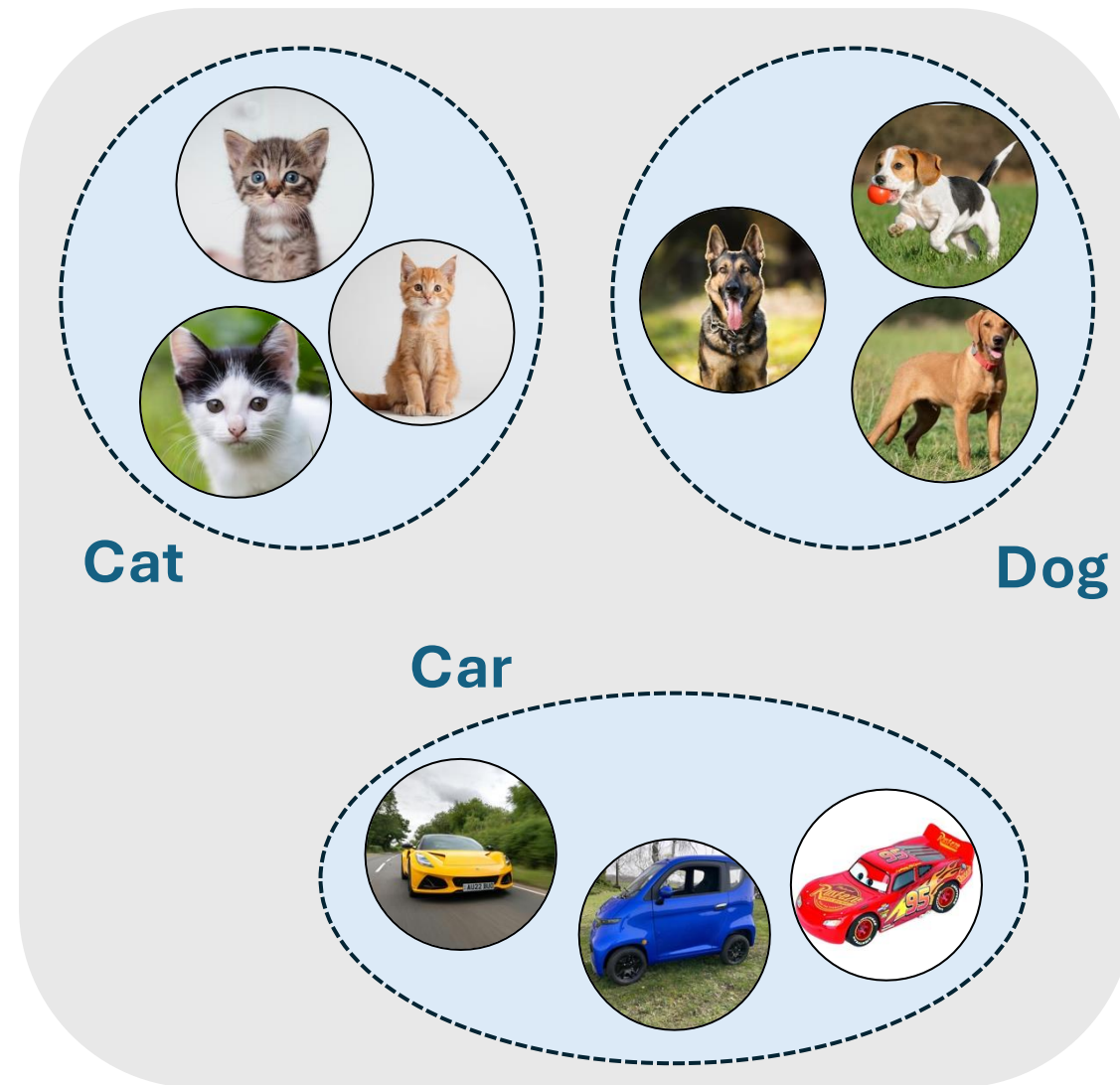


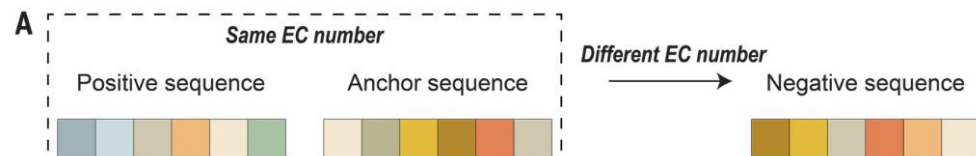
Predikce funkce enzymu – CLEAN

20



Embedding space





Enzyme classification and EC numbers

class (main division)

• EC 1 Oxidoreductases

• EC 2 Transferases

subclass

↑ EC 2.1 Transferring one-carbon groups

sub-subclass

↑ EC 2.1.2 Hydroxymethyl-, formyl- and related transferases

serial number in the sub-subclass

↑ EC 2.1.2.7 D-alanine 2-hydroxymethyltransferase

Reaction: 5,10-methylenetetrahydrofolate + D-alanine + H₂O = tetrahydrofolate + 2-methylserine

• EC 3 Hydrolases

• EC 4 Lyases

• EC 5 Isomerases

• EC 6 Ligases

• EC 7 Translocases

Input

MAPTVMASSATSVAPFQGL
KSTAGLPVSRRTNSGFGN
VSNGGRIKCMQWPIEGIK
KFETLSYLPPLTVEDLLKQIEY
LLRSKWVPCLEFSKVGfVYR
ENHRSPGYDGRYWTMWK
LPMFGCTDATQVLKELEEAK
KAYPDAFVRIIGFDNVRQVQ
LISFIAYKPPGCEESGGN

Class

EC 2.1.39
cat
ribulose-
bisphosphate
carboxylase

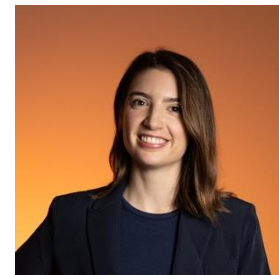
dist (query, EC)

dist (query, EC)

dist (query, EC)

Praktická ukázka 1: Predikce funkce enzymu

22



Ariane Mora



<https://tinyurl.com/czai-clean>

Využití AI k návrhu enzymů

Příklad 2: Vylepšování přírodních enzymů

Jak vylepšit enzym?

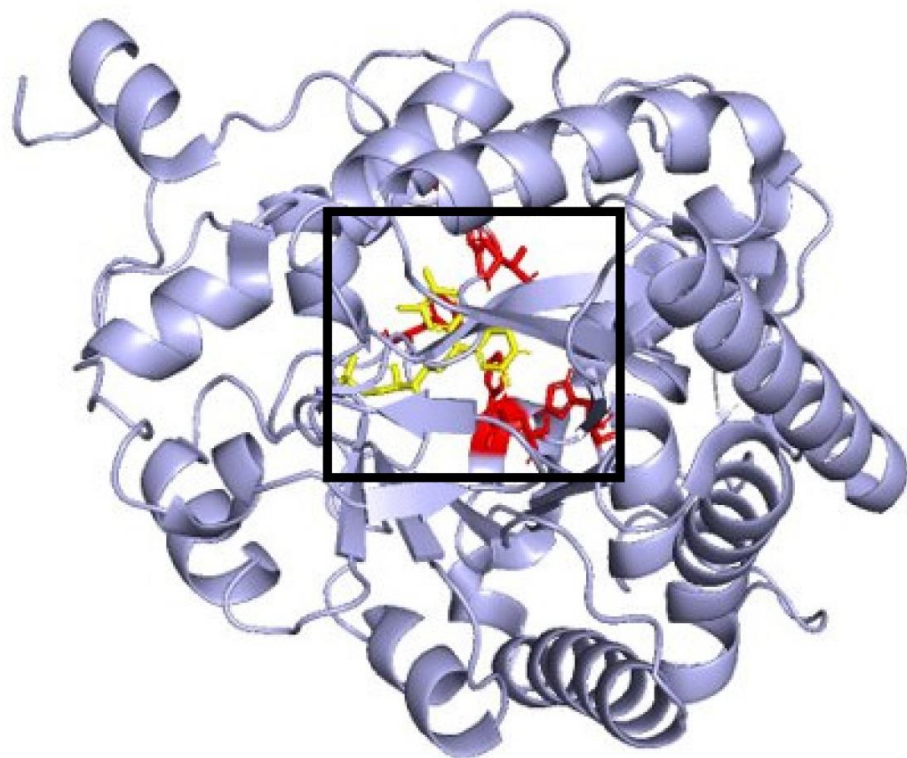
26

MAPTVMASSATSVAPFQGLKSTAGLPVSRRSTNSGF **G**NVSNGGRIKCMQVWPIEGIKKFETLSYLPPLTVEDLLKQIEYLLRSK...

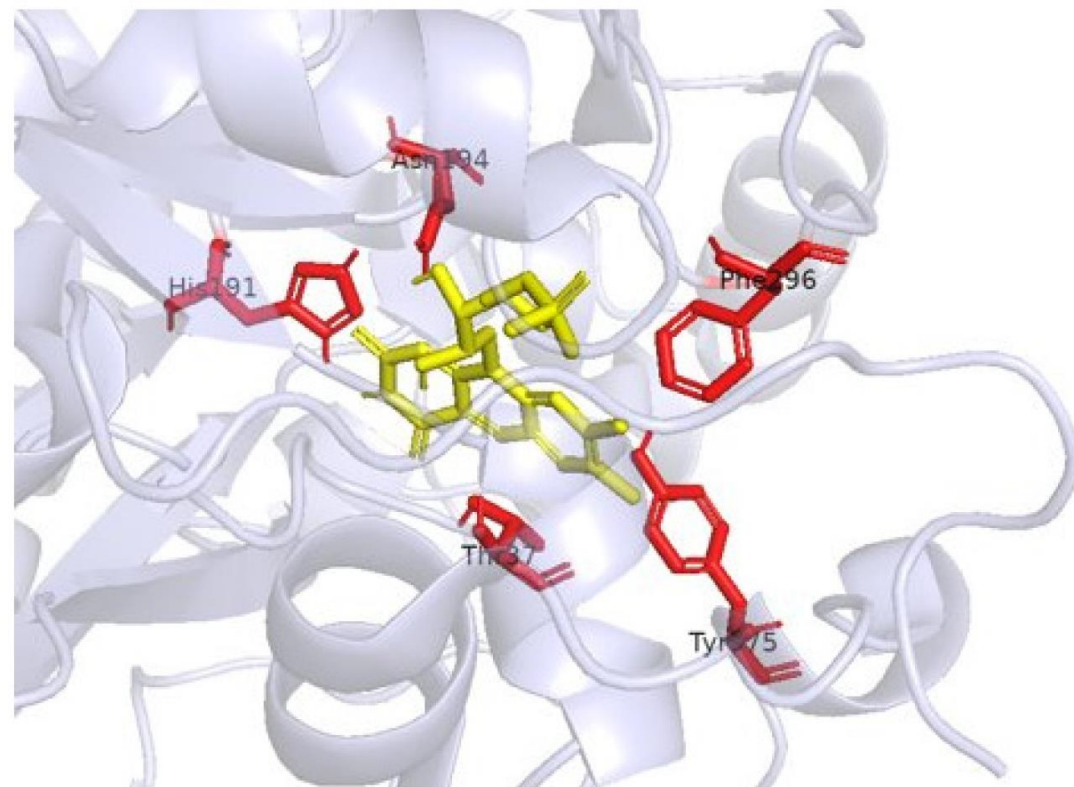


MAPTVMASSATSVAPFQGLKSTAGLPVSRRSTNSGF **K**NVSNGGRIKCMQVWPIEGIKKFETLSYLPPLTVEDLLKQIEYLLRSK...

A



B



Návrh prospěšné mutace - Variant effect prediction

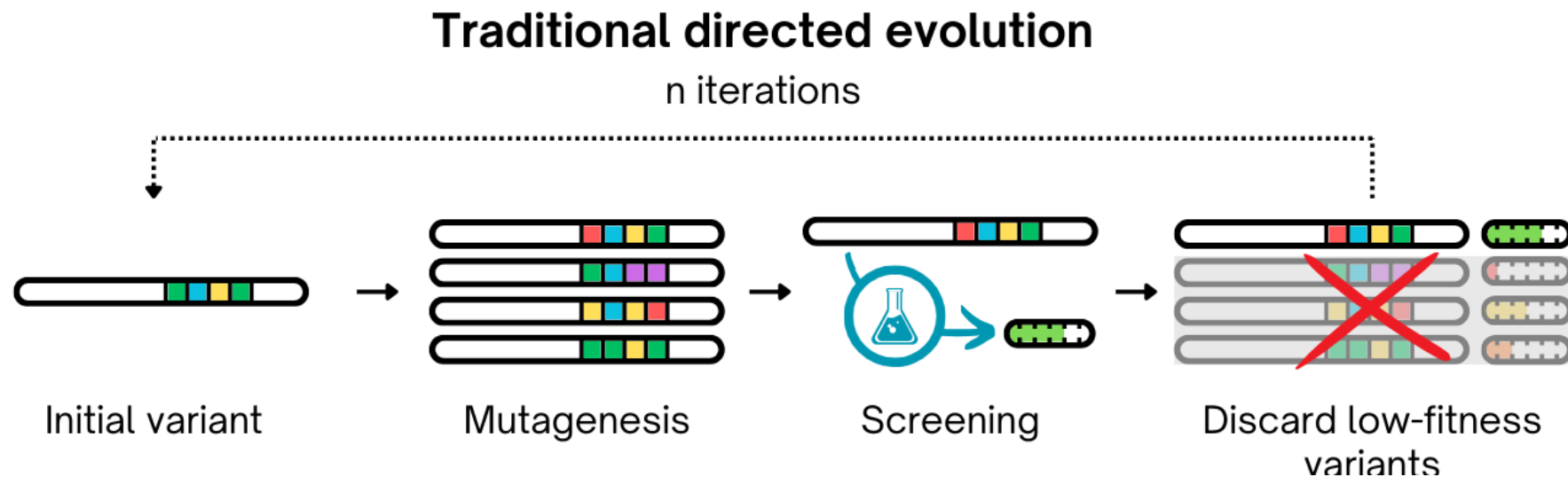
28



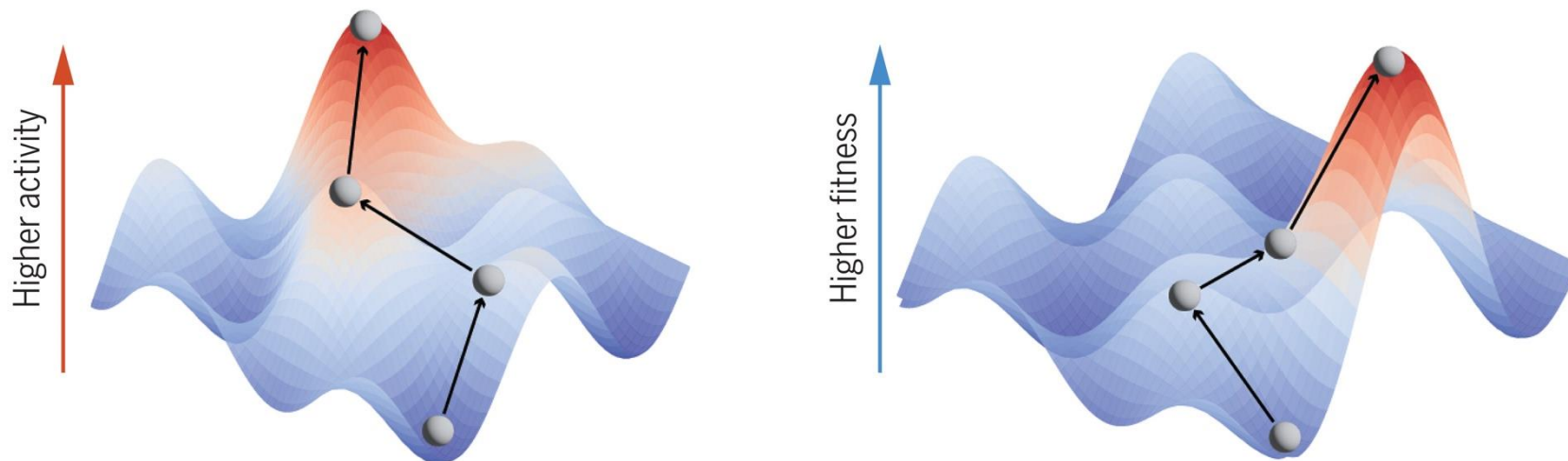
DESIGN		
Find Novel Enzymes 	Improve Stability 	Design Ancestor 
Improve Activity 	Modify Specificity 	Design Point Mutations 
Design Smart Library 	Design Loops 	Modify Dynamics 
Suppress Aggregation 	Design protein-protein interactions 	https://loschmidt.chemi.muni.cz/porta/

Vylepšování proteinů řízenou evolucí

29



Soldát, Matouš, and Jiří Kléma. "Directed evolution of proteins via Bayesian optimization in embedding space." 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2024.



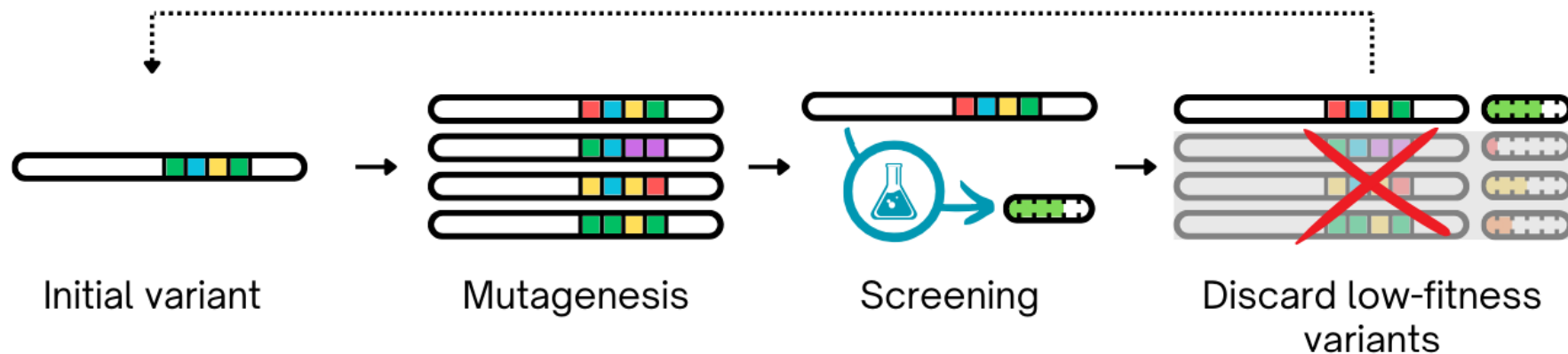
Jiang, Kaiyi, et al. "Rapid in silico directed evolution by a protein language model with EVOLVEpro." Science 387.6732 (2024): eadr6006.

Vylepšování proteinů řízenou evolucí

30

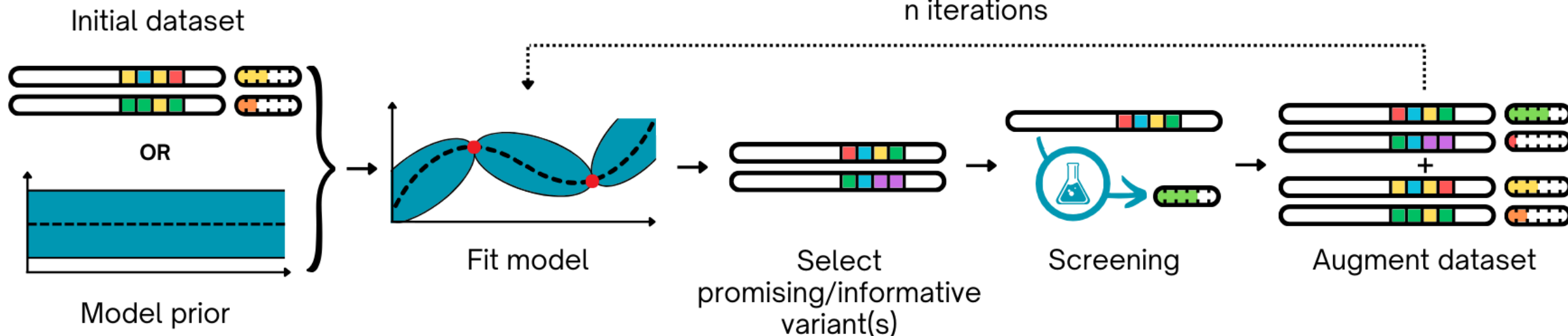
Traditional directed evolution

n iterations



Machine-learning-assisted directed evolution

n iterations



Praktická ukázka 2: Řízená evoluce proteinu

31



<https://tinyurl.com/czai-mlde>

Využití AI k návrhu enzymů

Příklad 3: Generování nových enzymů

Generování nových enzymů

33

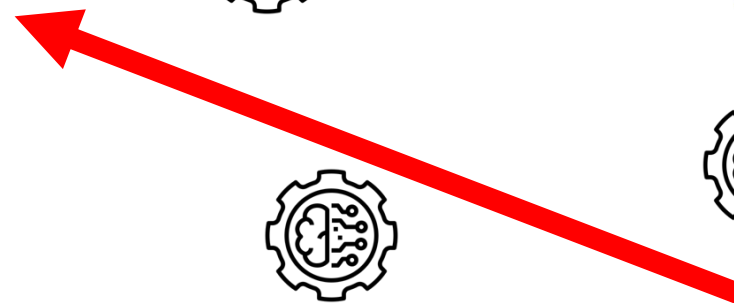
Sekvence

MAPTVMASSATSVAPFQGLKSTAGLPV
SRRSTNSGFGNVSNNGGRIKCMQVWPI
EGIKKFETLSYLPPLTVEDLLKQIEYLLRS
KWVPCLEFSKVGFVYRENHRSPGYD
GRYWTMWKLPMFGCTDATQVLKELEE
AKKAYPDAFVRIIGFDNVRQVQLISFIAY
KPPGCEESGGN

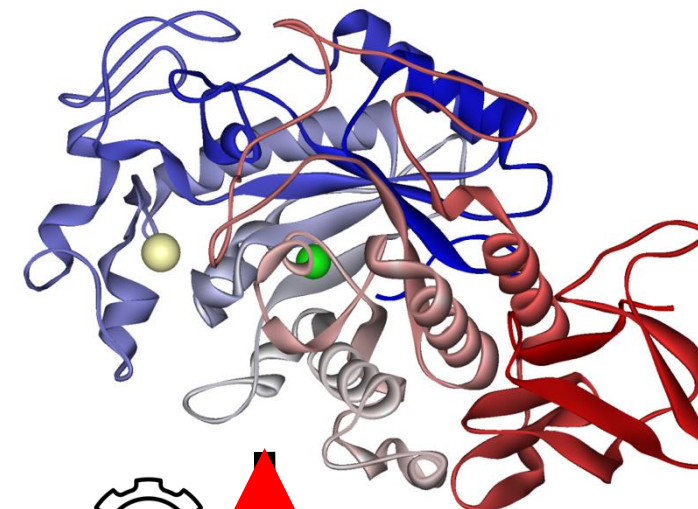


”Svatý grál molekulární biologie”

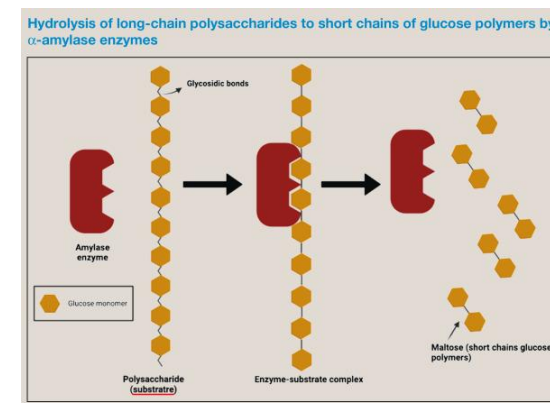
Inverse Folding



Struktura



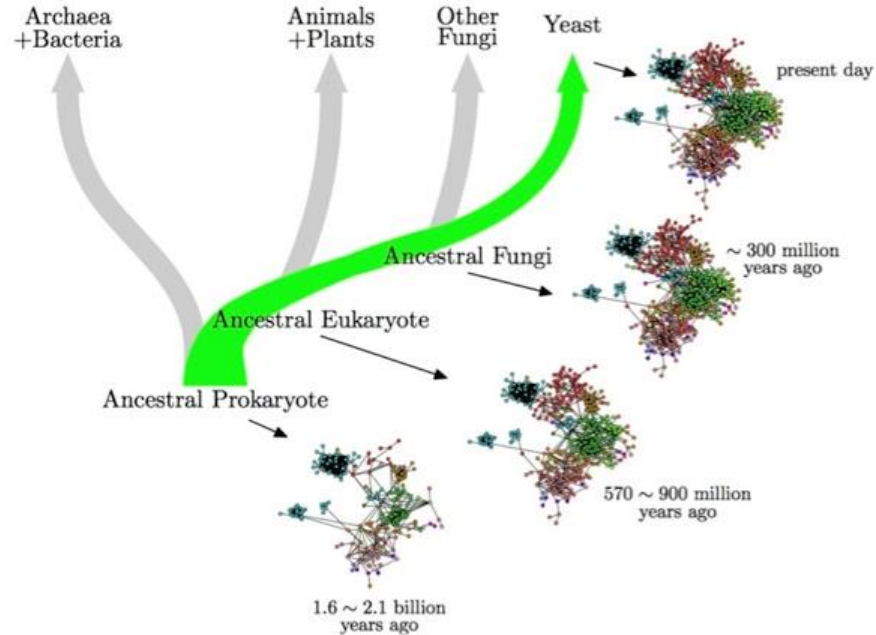
Funkce



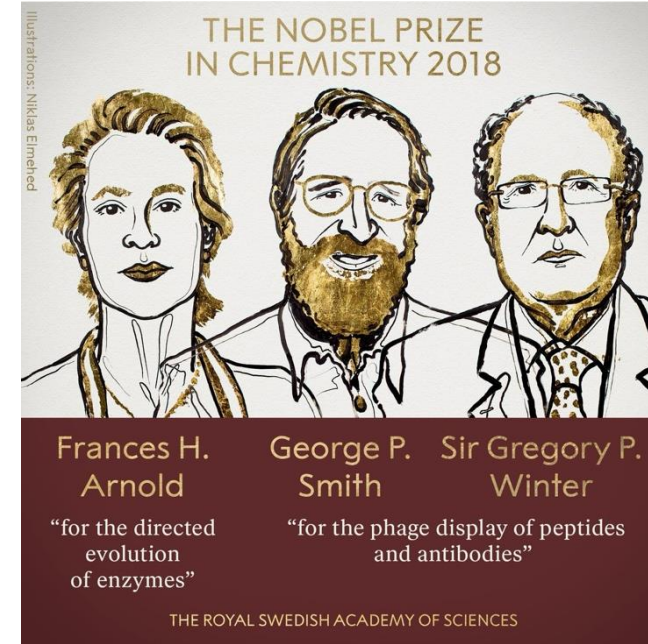
Evolution is limited to a tiny part of the design space

34

Evolution + screening



Directed evolution



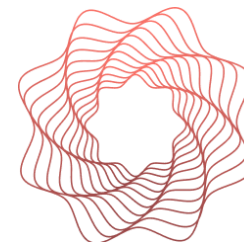
Native proteins

Directed evolution

De novo protein design



ChatGPT



ESM

Read and fill in the missing words from the box.



bunny, only, parents, years, teddy bear,
school, have got, garden, name, birthday, nice

My name is Lisa. I'm seven _____ old.
I go to _____, my class is 2.b. I live in a house with my
_____. I haven't got any brothers or sisters. I'm an
_____ child. My dad is really _____ and
my mum is great. We _____ a cat. It's name is Riki.
He likes to play in the _____ very much.
I have got a lot of toys, but I'd like to have a big _____
and a small _____ for my _____.

____ / 10

...MA__MR_LPLLA...

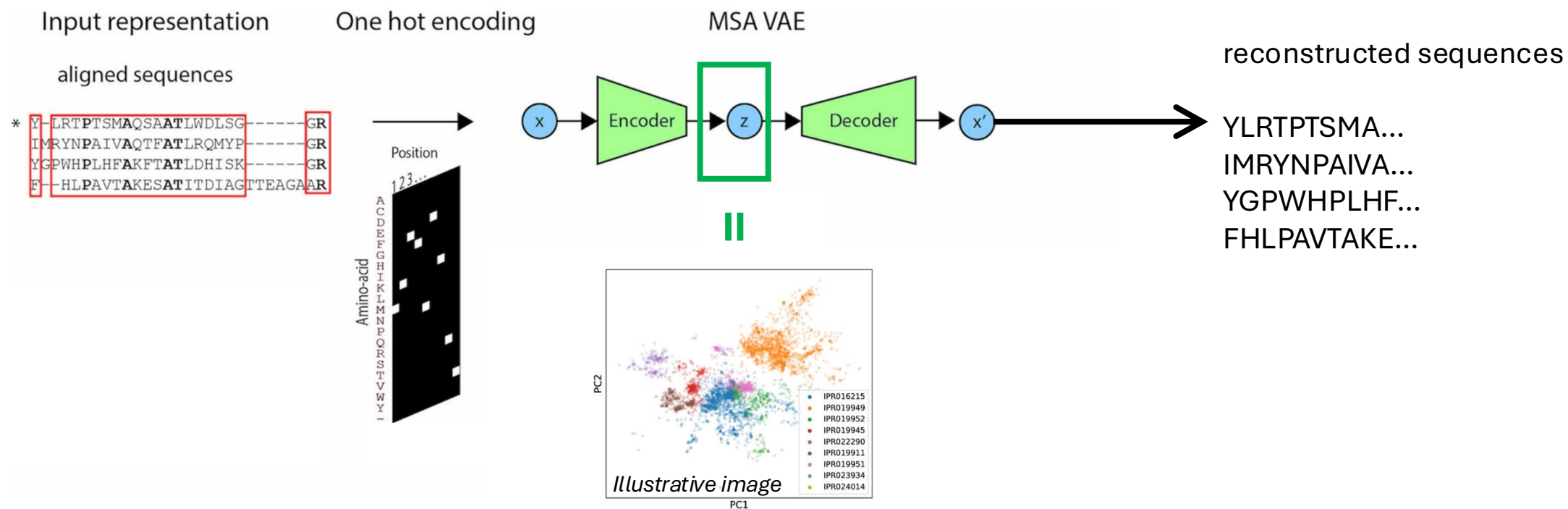


...MALWMRALPLLA...

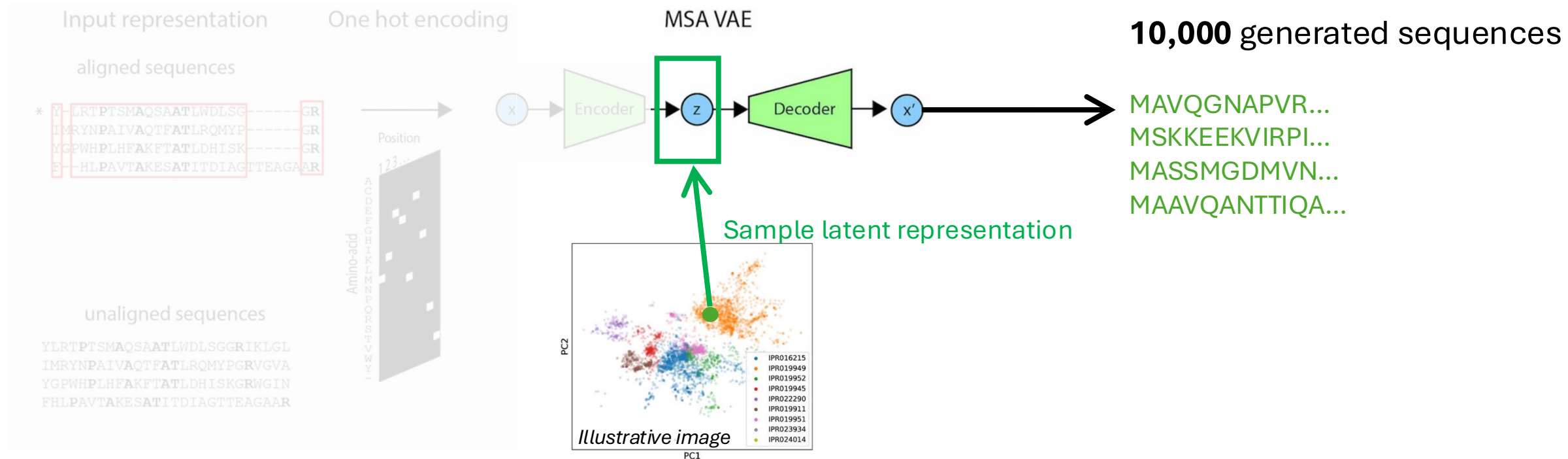
Generativní modely - Variational autoencoder

37

Training



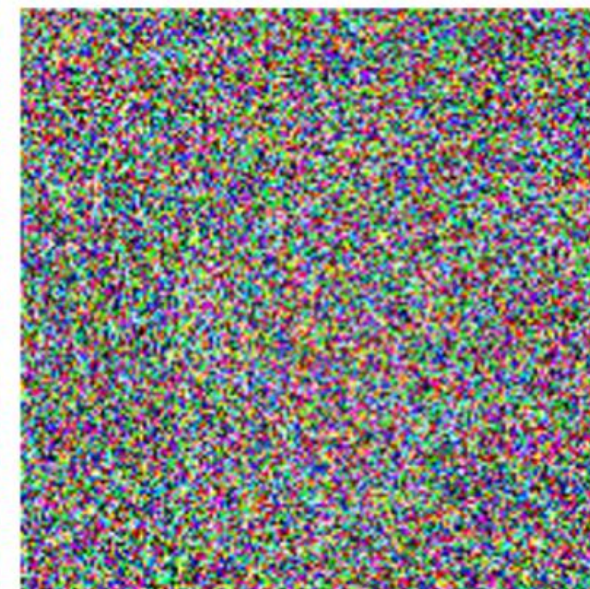
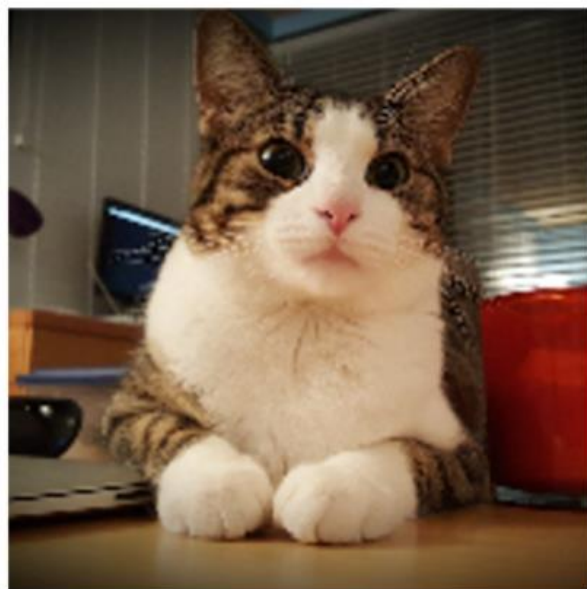
Generation



Generativní modely - Diffusion process

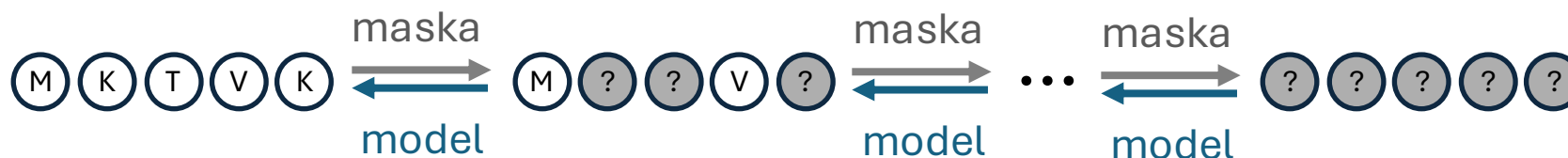
39

Přidávání šumu



<https://developer.nvidia.com/>

Opravování šumu ML modelem

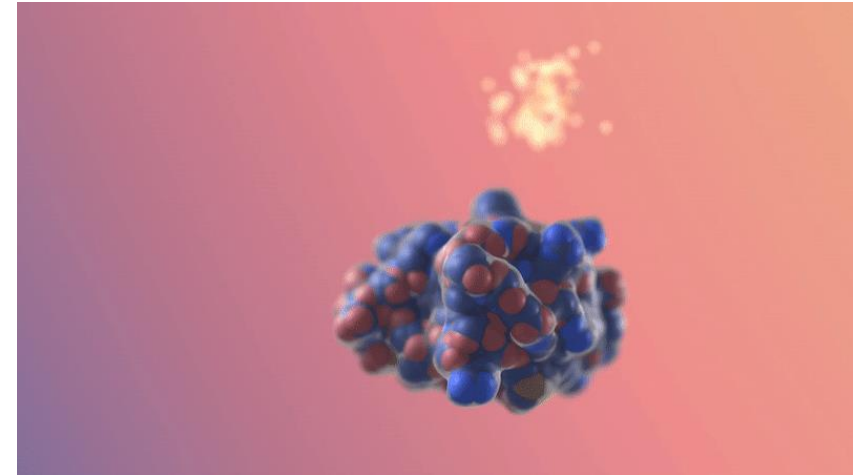


Generativní modely - Diffusion process

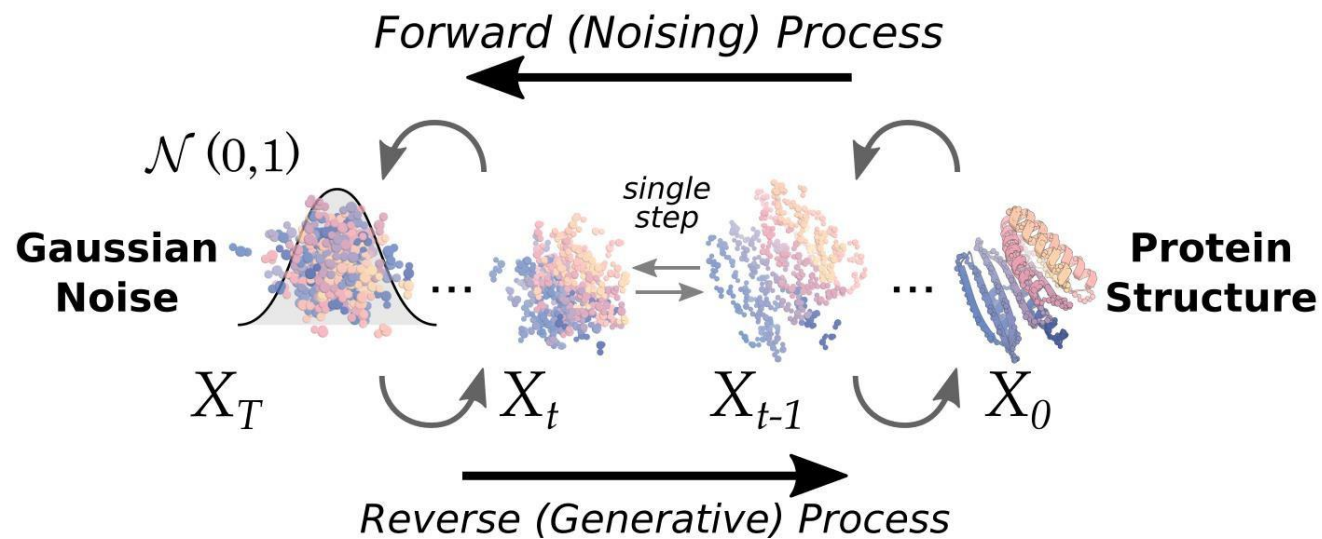
40



Baker Lab. Rfdiffusion.



Baker Lab. Rfdiffusion.



Praktická ukázka 3: Generování nového enzymu

41



<https://tinyurl.com/czai-rfdiffusion>

Thank you!

45



IOCB
PRAGUE



Pluskal Group
IOCB Prague



www.pluskal-lab.org



matous.soldat@uochb.cas.cz



[matous-soldat](#)

