

**LOSCHMIDT**  
LABORATORIES



# Výpočetní návrh stabilních proteinů

*Miloš Musil*



# Loschmidtovy laboratoře



Prof. Jiří Damborský



1994

MUNI

FAKULTNÍ  
NEMOCNICE  
U SV. ANNY  
V BRNĚ



Z. Prokop



D. Bednář



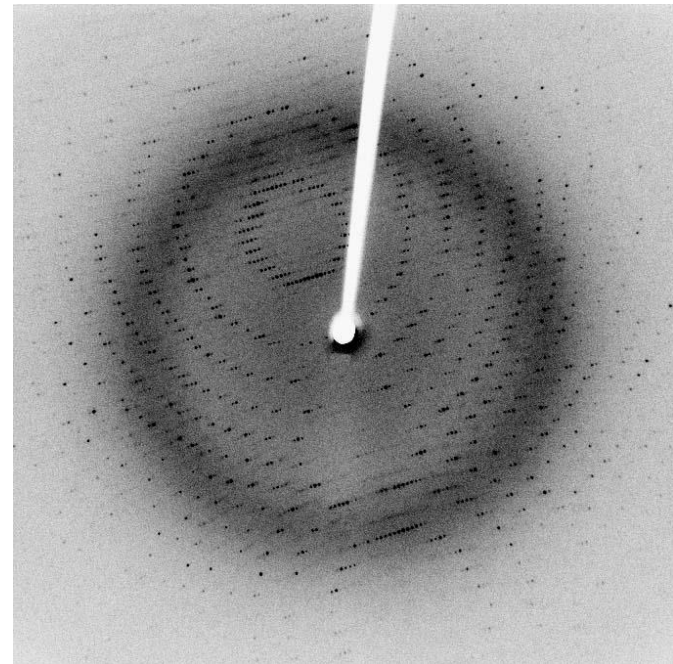
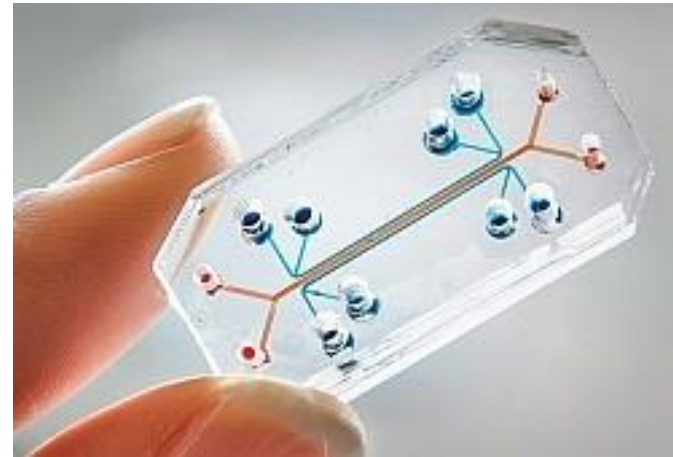
M. Marek



S. Mazurenko

# Co děláme...?

- Experimentální tým
  - Charakterizace proteinů
  - Řešení struktur
  - Mikrofluidika
- Teoretický tým
  - Molekulární dynamiky
  - **Bioinformatika**
  - Strojové učení



# Naše nástroje



# Proteiny



# Stabilní proteiny

- Stabilní proteiny jsou **odolnější**



Teplota



Nepříznivé pH



Rozpouštědla

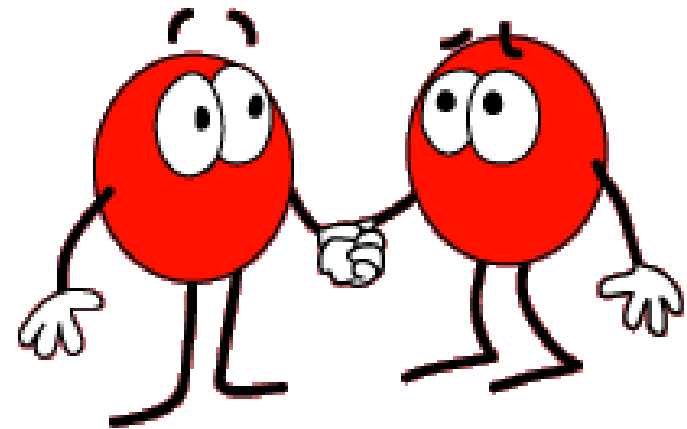


Životnost

- **Odolnost proti mutacím**
  - Důležité během evoluce

# Stabilita proteinů

- Silně spojena se **strukturoou proteinů**
  - **Náhodné skládání není možné**
  - Struktura proteinu závisí pouze na sekvenci
  - Způsobeno interakcemi mezi aminokyselinami
- Fyzikálně-chemické interakce
  - Kovalentní vazby
  - Nekovalentní interakce



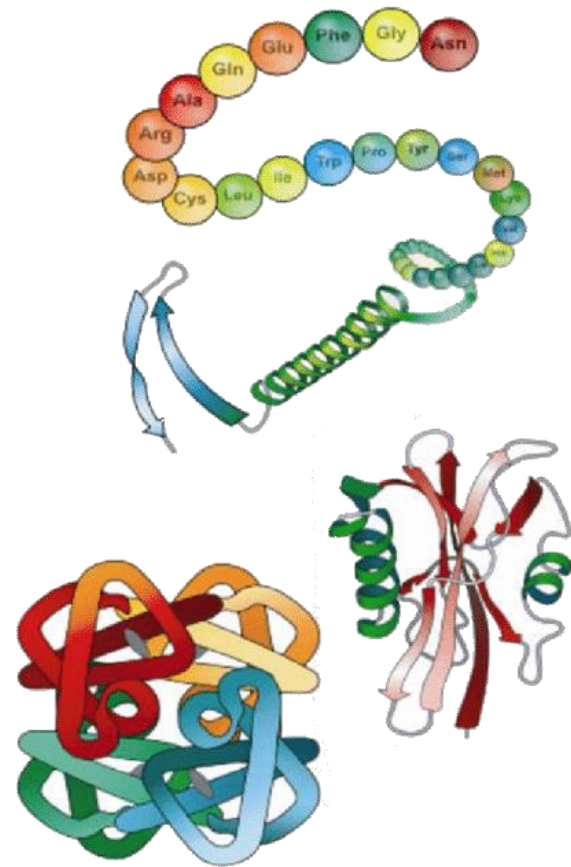
# Molekulární interakce

- Kovalentní interakce

- Sdílení elektronu
- Spojeno s primární strukturou

- **Nekovalentní interakce**

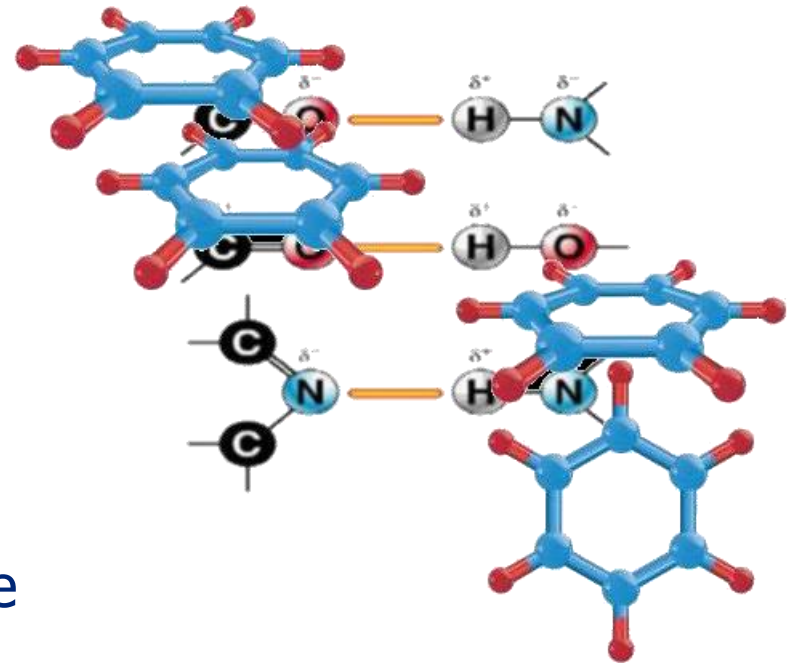
- Elektrostatické interakce
- Polární interakce
- Nepolární interakce
- Spojeno se sekundární a terciální strukturou





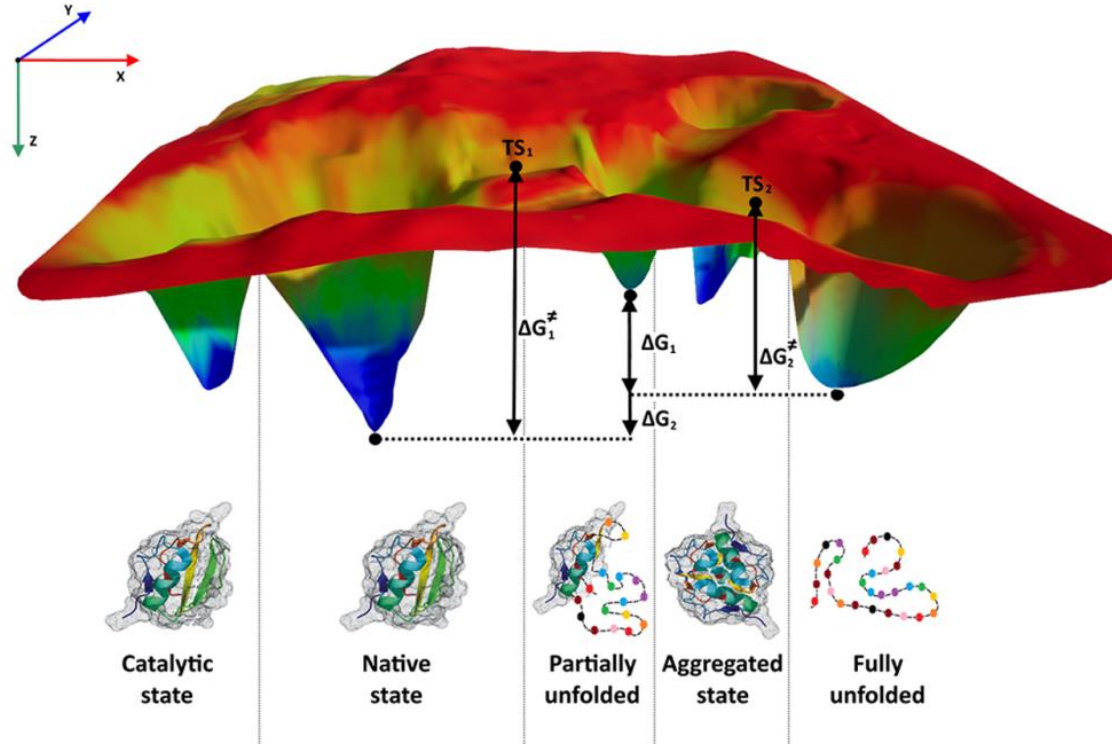
# Nekovalentní interakce

- Interakce mezi náboji
- Polární interakce
  - Vodíkové můstky
  - Aromatické interakce
- Nepochární interakce
  - Van der Waalsovy interakce
  - Hydrofobicita



# Stabilita proteinů

- Obecně vzato reprezentováno změnou **Gibbsovy volné energie při skládání ( $\Delta G$ )**



# Návrh stabilních mutantů

- Měřeno v  $\Delta\Delta G$  nebo  $\Delta T_m$

$$\Delta\Delta G = \Delta G_{mutant} - \Delta G_{wt}$$

- Záporné  $\Delta\Delta G$  značí vyšší stabilitu
- Jednobodové x vícebodové mutace
  - Efekt na stabilitu
  - Synergické, aditivní, **antagonické efekty**

# Návrh stabilních mutantů

- **Saturační mutagenese není možná**
  - Průměrný protein o 300 aminokyselinách dává téměř 6000 jednobodových mutací
  - Nespočet potenciálních kombinací
- Nutnost výpočetních nástrojů
  - Solidní přesnost ( $\sim 70-75\%$ )
  - Použitelné pro **prioritizaci**



# Výpočetní metody

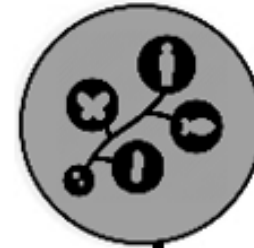
Strojové učení



Silová pole



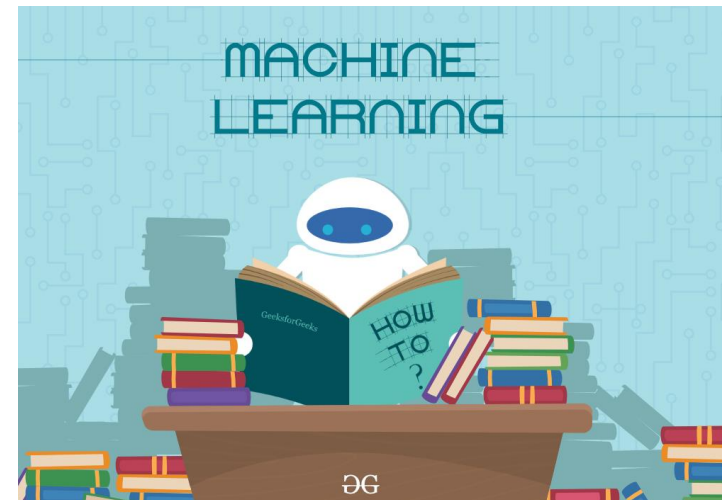
Evoluce



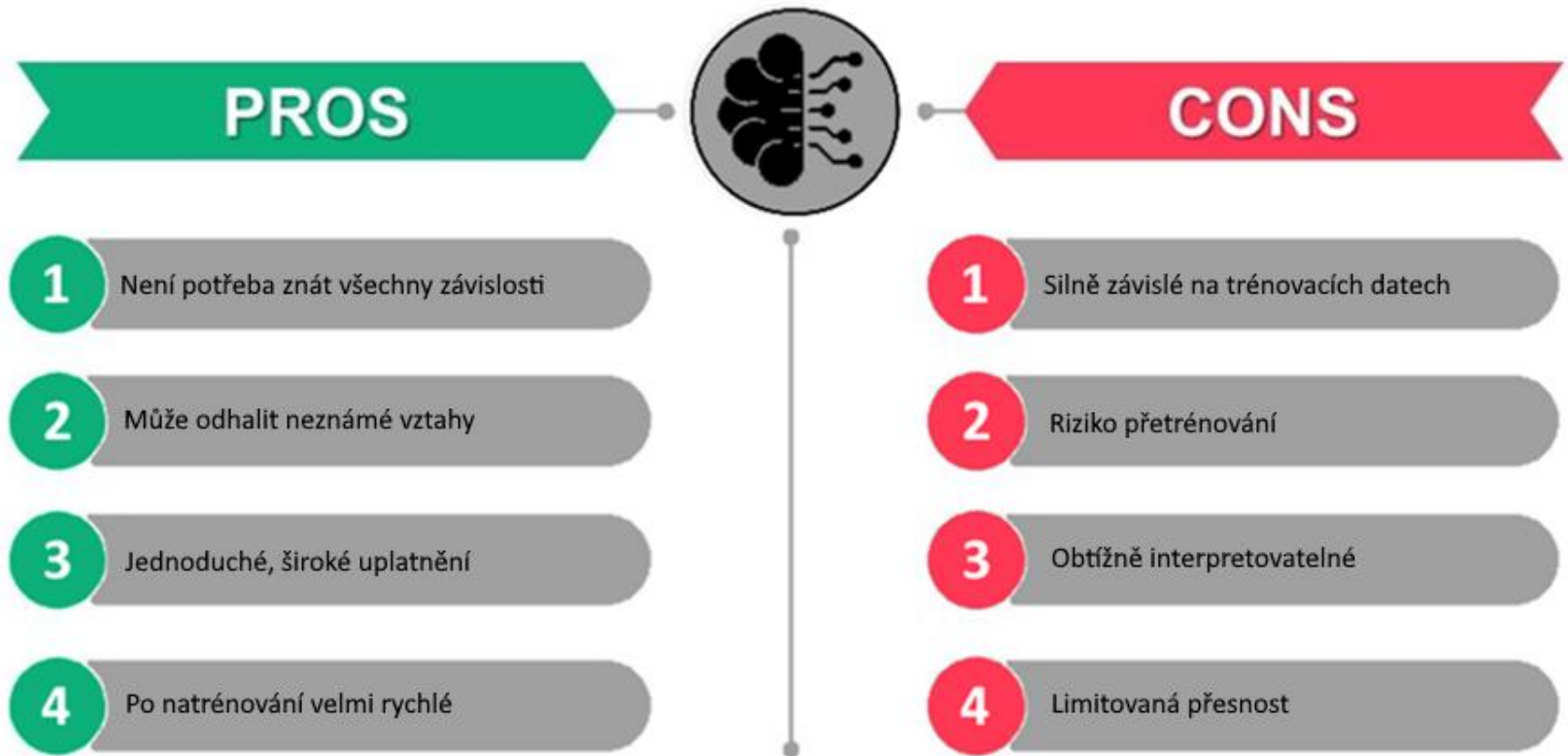
Hybridní metody

# Strojové učení

- Nejběžnější přístup
- Založeno na statistice a počítačových vědách
- **Rozpoznání vzorů nad trénovacími daty**
- S učením  $x$  bez učitele
  - Klasifikace
  - Shlukování
  - Rozpoznání vzorů

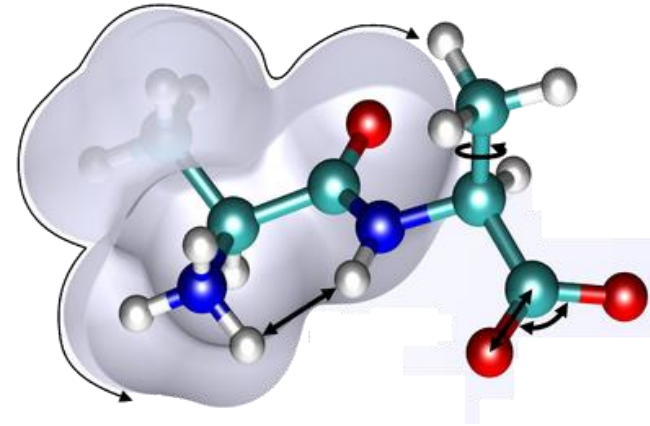


# Strojové učení



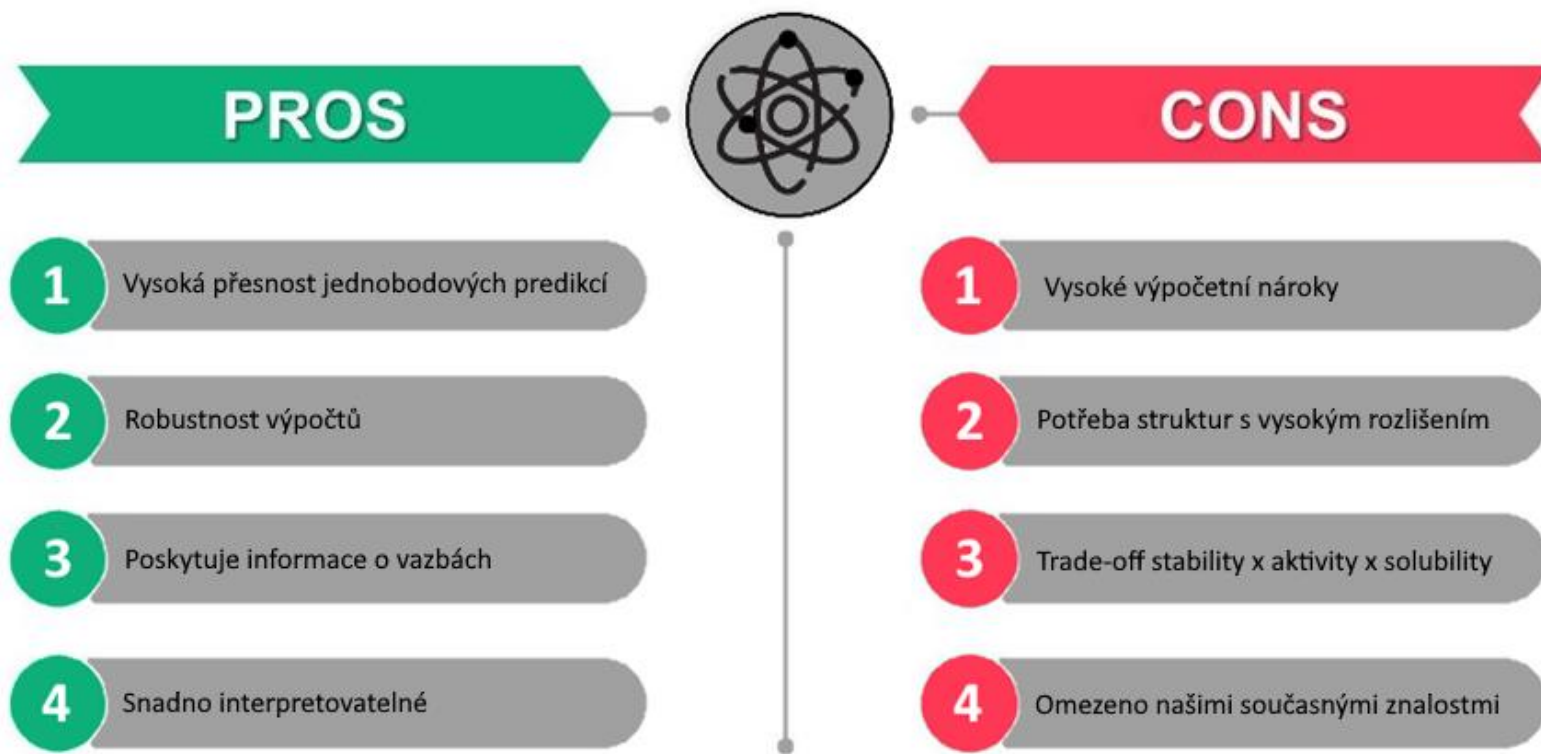
# Silová pole

- Výpočet **volné energie**
- Založeno na (zjednodušených) zákonech fyziky
- **Zvažuje různé fyzikální a biochemické parametry**
- Fyzické x statistické potenciály
  - Přesnost vs výpočetní nároky
  - Statistické jsou založené na datech
  - (Empirické potenciály)





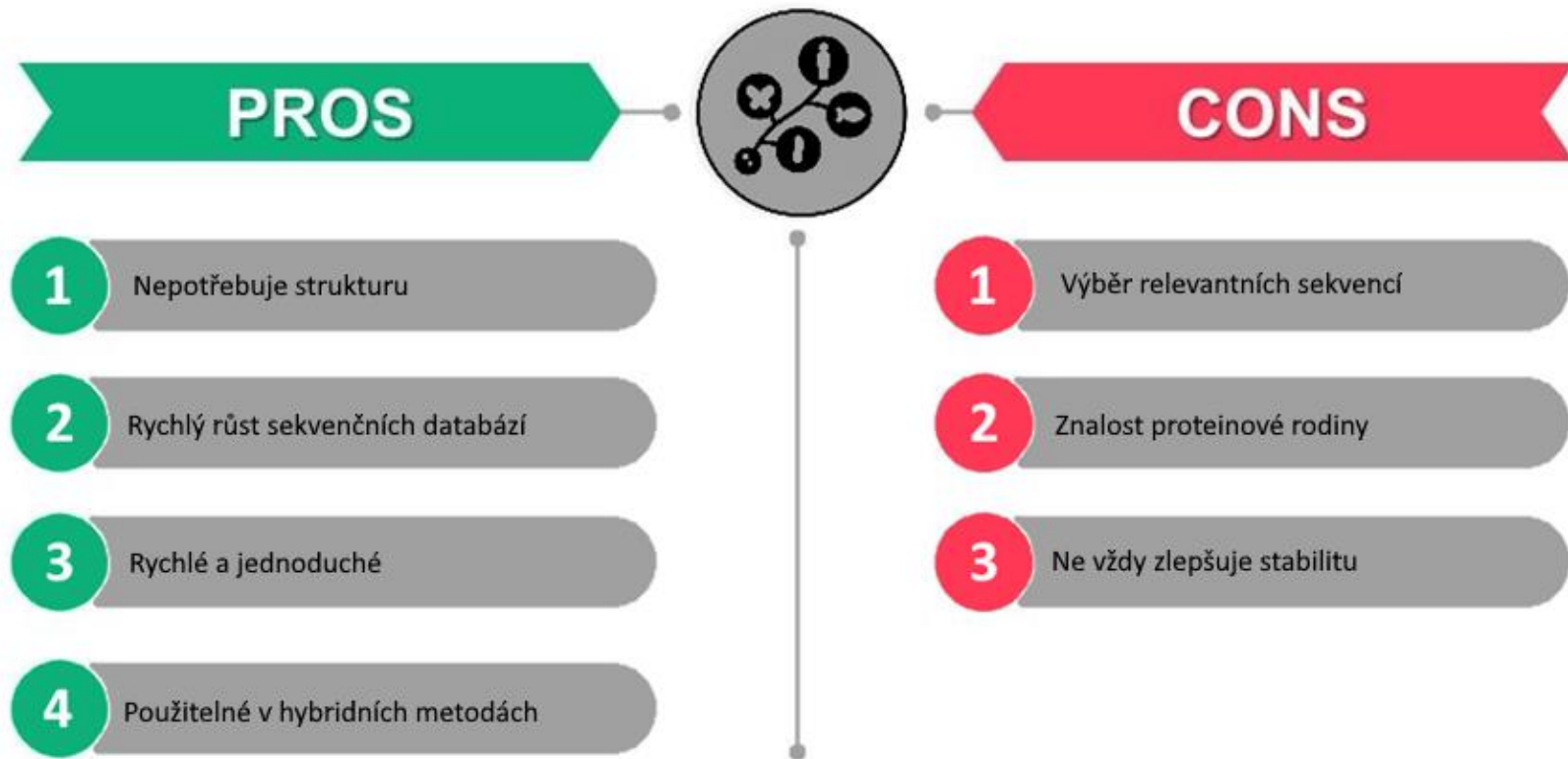
# Force-fields



- **Návrhy založené na evoluci**
- Využívá vícenásobné zarovnání sekvencí
- Analýza konzervovaných regionů
- Dva hlavní přístupy
  - **Ancestrální rekonstrukce**
  - Back-to-consensus

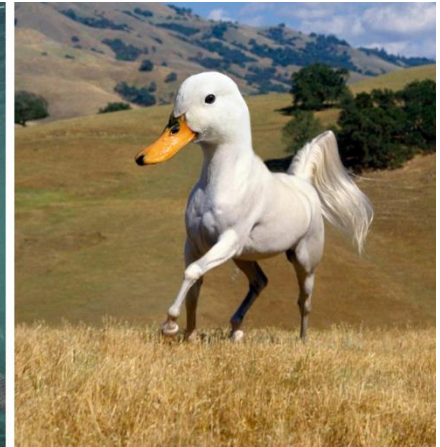


# Evoluce

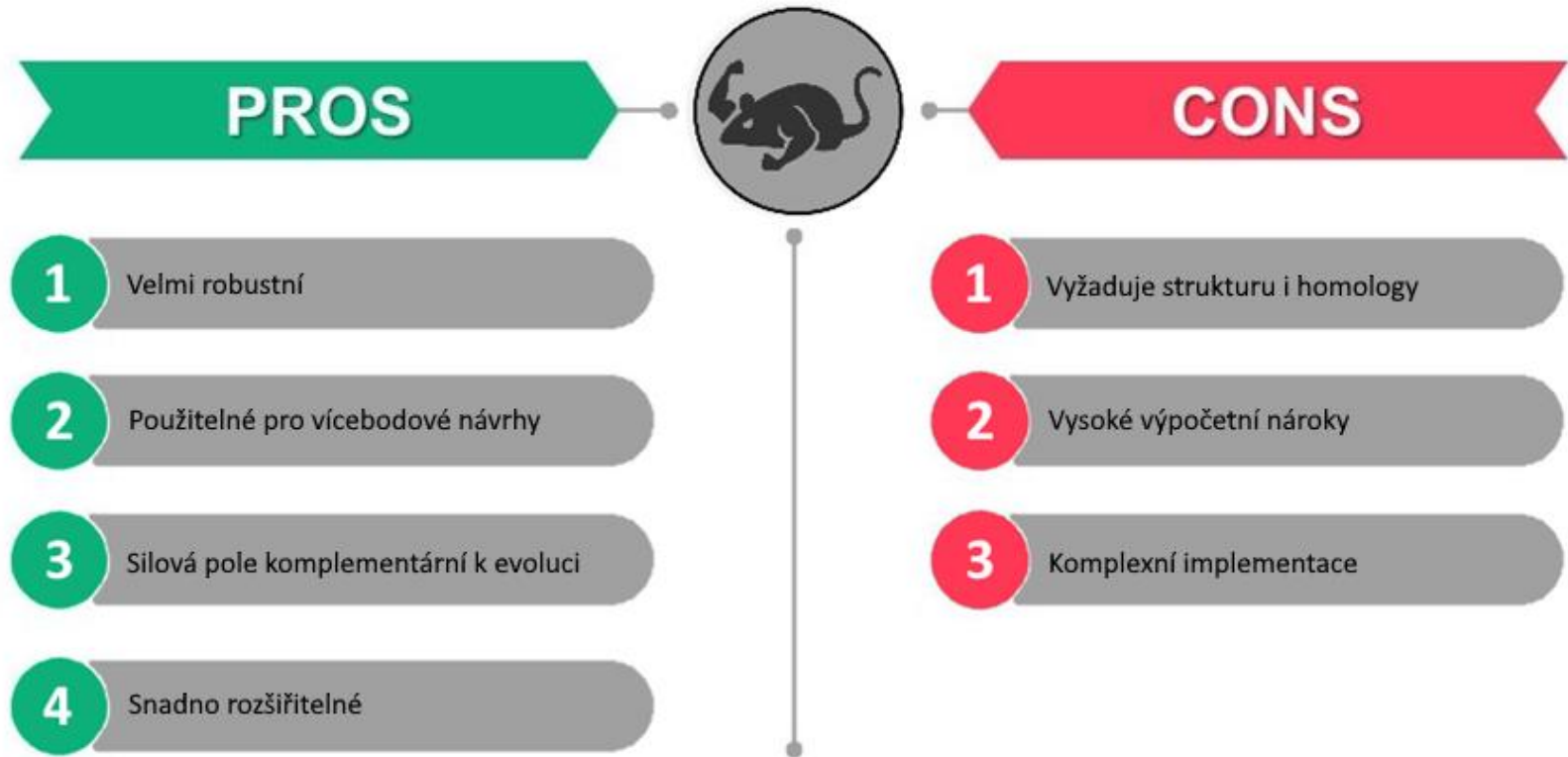


# Hybridní metody

- **Kombinace několika přístupů**
- Evoluce použita jako filtr pro silová pole
- Použitelné pro návrh **vícebodových mutantů**
- Pouze hrstka dostupných nástrojů:
  - FireProt
  - PROSS
  - FRESCO



# Hybridní metody



# Výpočetní nástroje

NÁVRH STABILNÍCH VÍCEBODOVÝCH MUTANTŮ



ANCESTRÁLNÍ REKONSTRUKCE



STABILNÍ DATA



IDENTIFIKACE HOTSPOTŮ





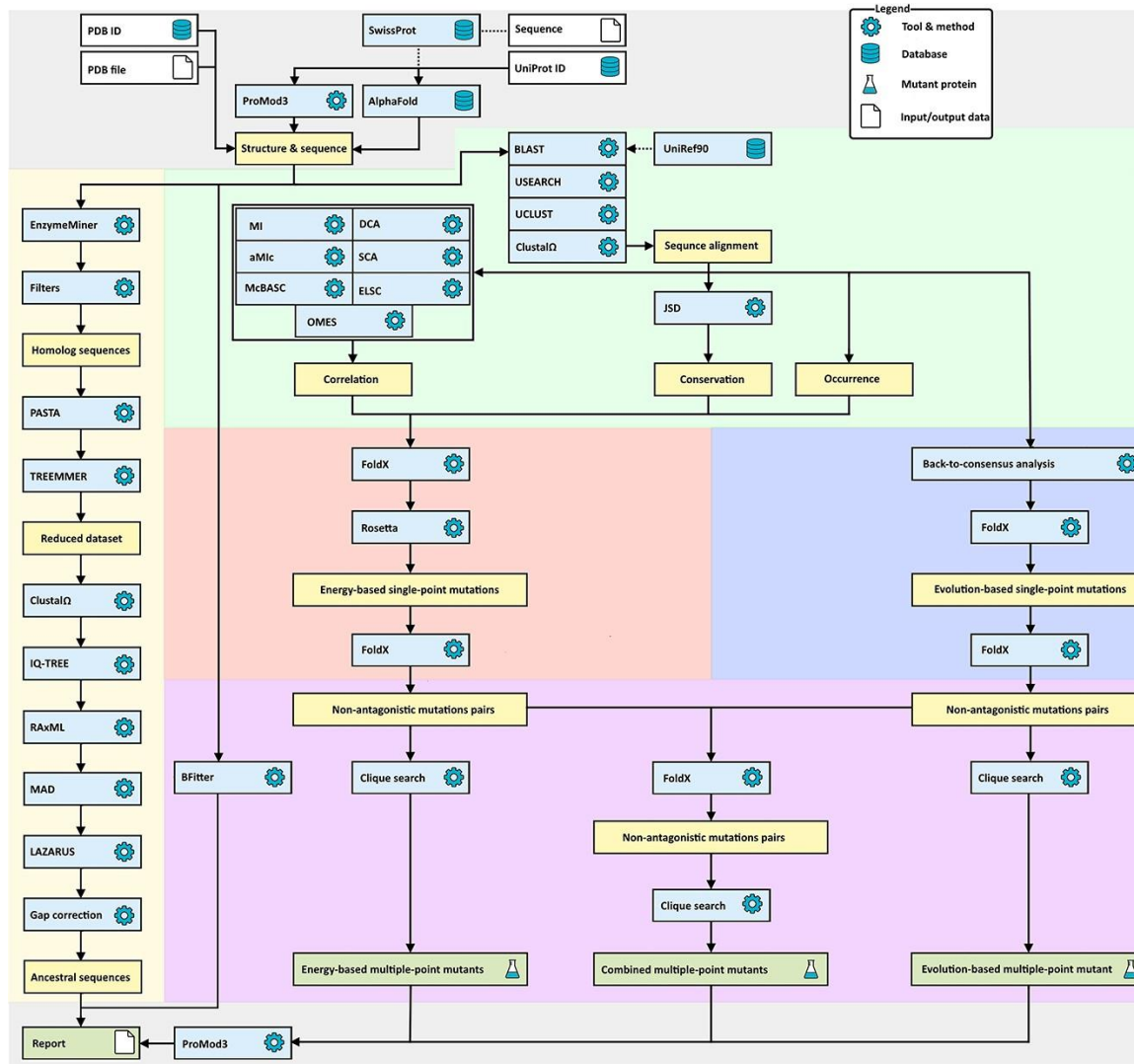
# **FIREPROT**

**AUTOMATED DESIGN OF STABLE MULTIPLE-POINT MUTANTS**

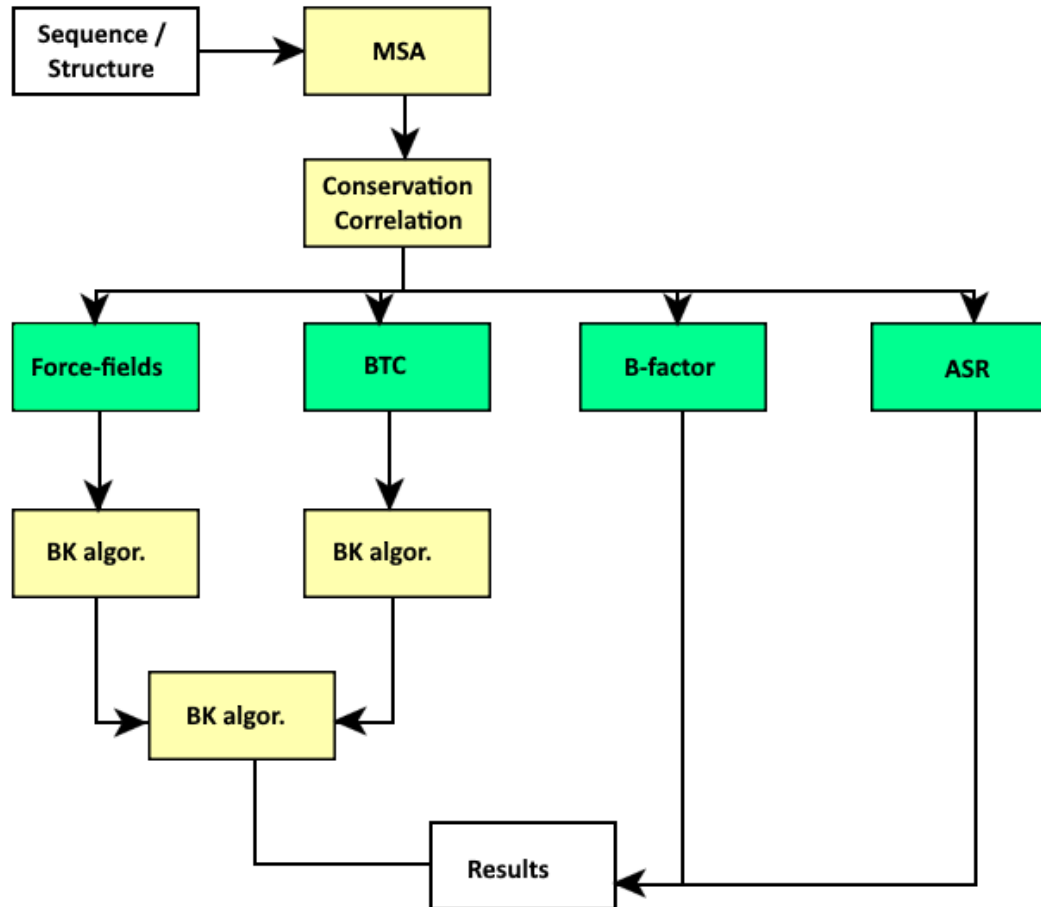
- Plně automatizovaný **návrh stabilních proteinů**
- Dva hlavní přístupy
  - Energetický (silová pole)
  - Evoluční (BTC a ASR)
- Filtrování na základě sekvence
- **Nástroj poskytuje několik vícebodových mutantů**
- Uživatelské rozhraní



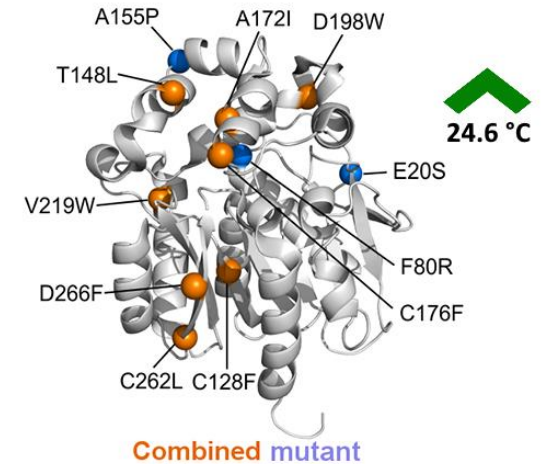
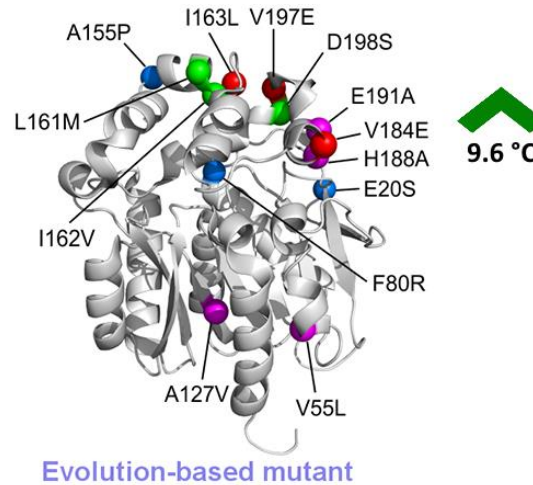
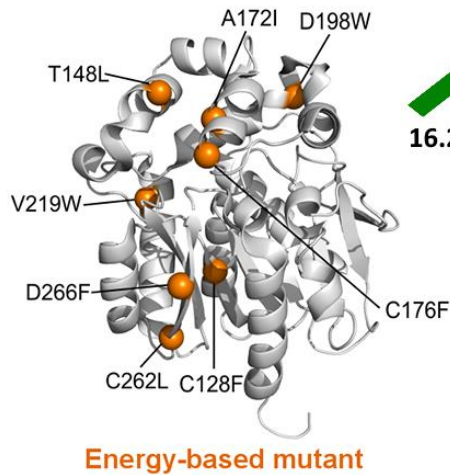
# Schéma



# Schéma



# Výsledky



- Významná stabilizace DhaA, LinB a FGF2
- Další v literatuře

# Uživatelské rozhraní



v2.0

Computational design of stable proteins



[Submit new job](#)

[Example](#)

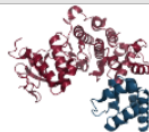
[Acknowledgements](#)

Job ID:

[Find job](#)

## SELECT TYPE OF INPUT DATA

STRUCTURE



SEQUENCE

**IDDQD**  
**MSLGAKPF**  
**GAAIAAFVRAM**  
**VVLVVHDWGSALRGL**

## JOB INFORMATION

Job title

Email (recommended)

*FireProt is free and open to all users and there is no login requirement.*

[Next](#)

## REFERENCES

Musil M, Jezik A, Horackova J, Borko S, Kabourek P, Damborsky J, Bednar D: FireProt 2.0: web-based platform for the fully automated design of thermostable proteins. Briefings in Bioinformatics 25(1), 2024.

## USER STATISTICS

Number of visitors: -  
Number of jobs: 3974

## CONTACT

Loschmidt Laboratories

[fireprot@sci.muni.cz](mailto:fireprot@sci.muni.cz)


<https://loschmidt.chemi.muni.cz/>

## OTHER TOOLS




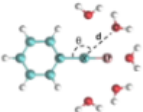

# Uživatelské rozhraní

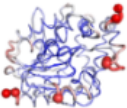

**CALCULATION**



**AUTOMATED** 


**USER-DEFINED** 



**AUTOMATED SINGLE-POINT** 



**Force-Field Mutations**  

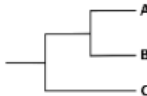

**B-Factor Analysis**  

**Consensus Mutations**  

**AUTOMATED MULTIPLE-POINT** 

**Low-Risk Design**  

**High-Risk Design**  

**Ancestral Design**  

[Back](#) [Next](#)

# Uživatelské rozhraní





# **FIREPROT**

**ASR**

**FULLY AUTOMATED ANCESTRAL SEQUENCE RECONSTRUCTION**

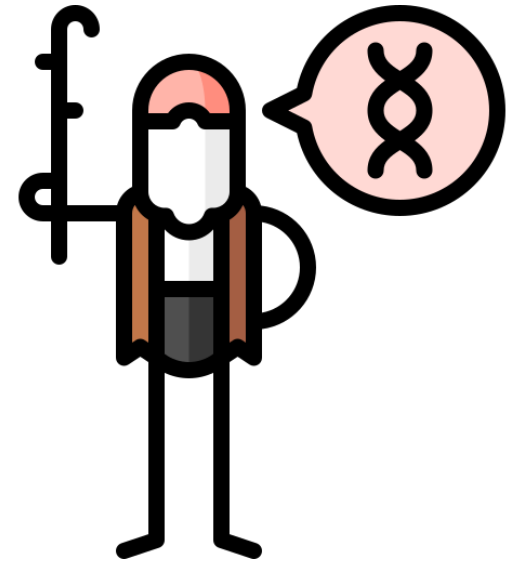
# Ancestrální rekonstrukce

- Ancestrální proteiny -> **stabilnější**
  - Náročné podmínky na Zemi
  - **Náhodné mutace během evoluce**
  - Bias způsobený konsenzem
- Několik kroků výpočtu
  - Vyhledání homologních sekvencí
  - MSA a konstrukce stromu
  - Ancestrální rekonstrukce a mezery
- **Vyžaduje expertní znalosti**



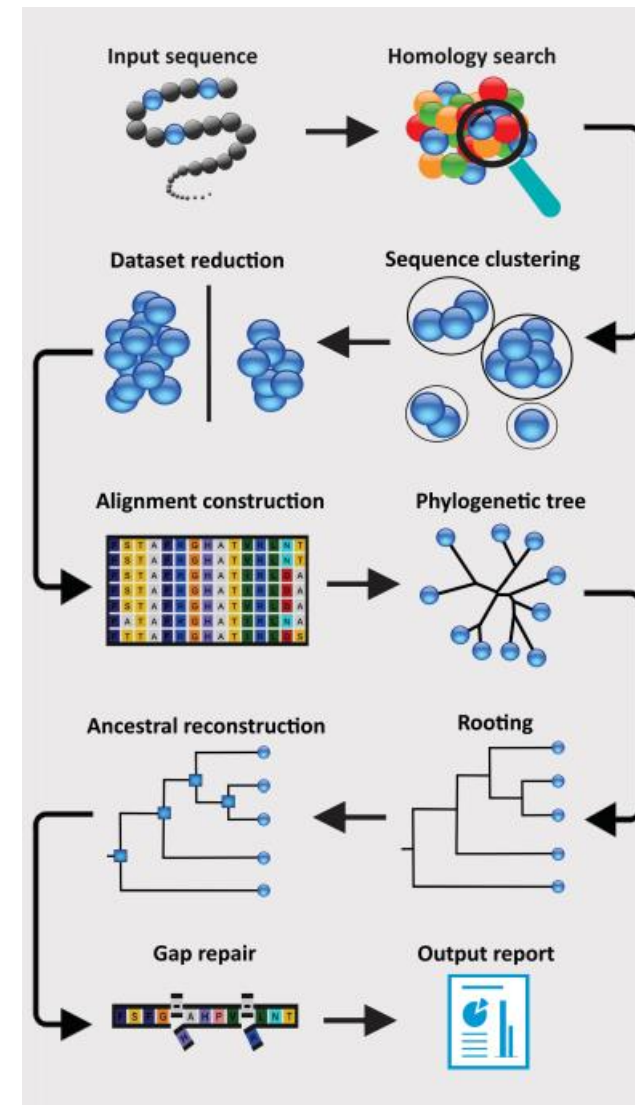


- **Plně automatizovaná metoda pro ASR**
  - Výběr počátečního setu homologů
  - Redukce datasetu
  - MSA a konstrukce stromu
  - Zakořenění a ancestrální rekonstrukce
- Doplněno interaktivním uživatelským rozhraním
  - Vizualizace struktury
  - Návrh nových ancestrálů

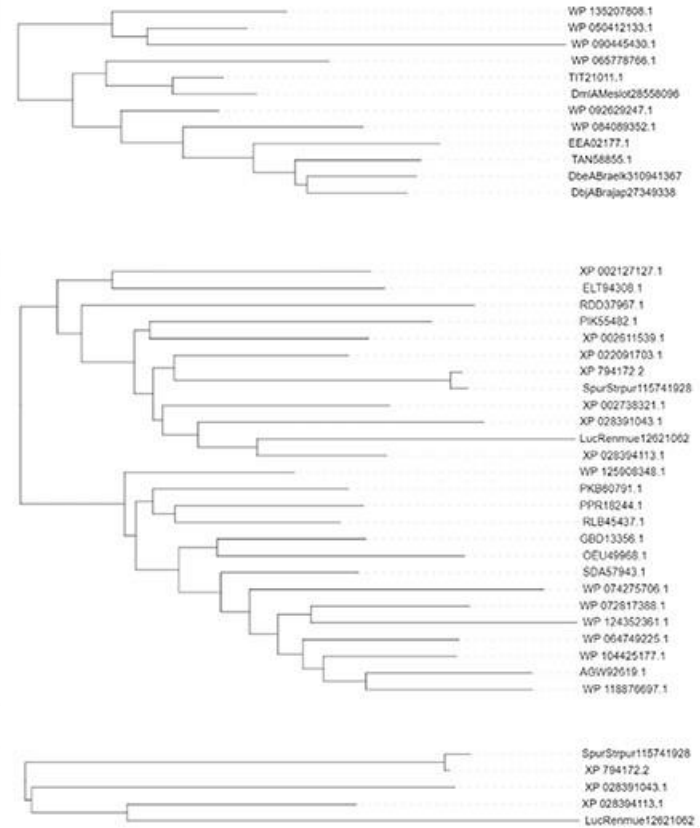
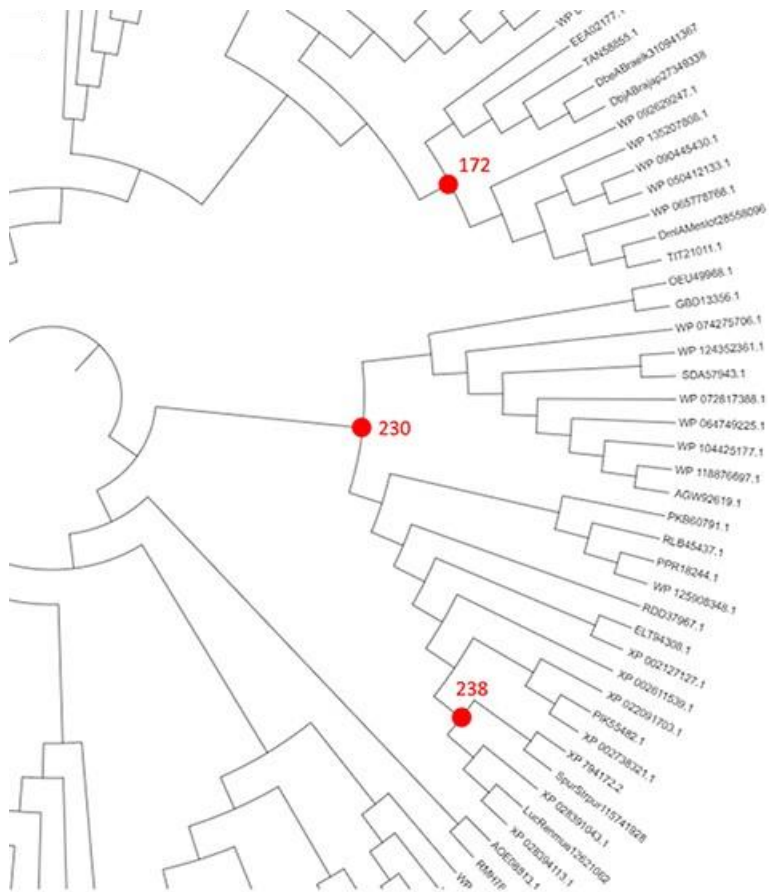


# Schéma

- Hledání homologních sekvencí
- Redukce výchozích dat
- Konstrukce MSA a stromu
- Zakořenění použitím MAD
- Ancestrální rekonstrukce s ML
- Rekonstrukce mezer



# Výsledky




# Results


| Protein code | Expression (%) | Solubility (%) | Yield (mg/l) | $T_m$ (°C)      | HLD act. ( $\mu\text{mol}/\text{mg}\cdot\text{s}$ ) | LUC act. (RLU/mg·s) |
|--------------|----------------|----------------|--------------|-----------------|---|---------------------|
| DhaA wt      | 17             | 83.1           | 91.1         | $50.56 \pm 2.4$ | $0.032 \pm 0.006$                                   | n.a.                |
| DhaA 172Loc  | 23             | 85.5           | 74.9         | $71.60 \pm 0.7$ | $0.038 \pm 0.001$                                   | $1.41 \pm 0.26$     |
| DhaA 172Glob | 21             | 65.2           | 88.2         | $70.04 \pm 1.5$ | $0.061 \pm 0.005$                                   | n.a.                |
| DhaA 230Loc  | 20             | n.d.           | n.d.         | n.d.            | n.d.  | n.d.                |
| DhaA 230Glob | 23             | 84.8           | 108.5        | $72.14 \pm 0.4$ | $0.061 \pm 0.012$                                   | n.a.                |
| DhaA 238Loc  | 23             | 63.2           | 74.9         | $70.36 \pm 0.6$ | $0.014 \pm 0.002$                                   | $353.5 \pm 14.58$   |
| DhaA 238Glob | 19             | 83.3           | 94.4         | $76.19 \pm 0.2$ | $0.030 \pm 0.002$                                   | $3.18 \pm 0.33$     |

- Zlepšení stability o **20 – 26 °C**
- Zachování výtěžků a aktivity

# Uživatelské rozhraní



ASR  
**FIREPROT** v1.1

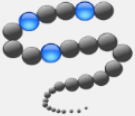



Fully automated ancestral sequence reconstruction

[Submit new job](#) [Help](#) [Example](#) [Use cases](#) [Acknowledgement](#)

Job ID:  [Find job](#)

### SELECT THE STARTING POINT

**SEQUENCE** 

**USER DATA** 

### STARTING FROM SEQUENCE

[Load example](#)

Source :  Enter own sequence  
 Upload sequence file

Sequence :

### JOB INFORMATION

Job title (optional) :



E-mail (optional) :

I agree with the academic license agreement and confirm that I will use the software exclusively for non-commercial purposes.

[Previous](#) [Next](#)


### REFERENCE

Musil, M., Khan, R., Beier, A., Stourac, J., Konegger, H., Damborsky, J., Bednar, D. 2020: FireProt-ASR: A Web Server for Fully Automated Ancestral Sequence Reconstruction. *Briefings in Bioinformatics*, 2020, bbaa337.

### COMMERCIAL VERSION

License for commercial use  
Contact us: [enantis@enantis.com](mailto:enantis@enantis.com)

  
**Enantis**

### USER STATISTICS

- Number of visitors: 25917
- Number of jobs: 11083

### CONTACT

**Loschmidt Laboratories**

- [fireprot@sci.muni.cz](mailto:fireprot@sci.muni.cz)
- <https://loschmidt.chemi.muni.cz>

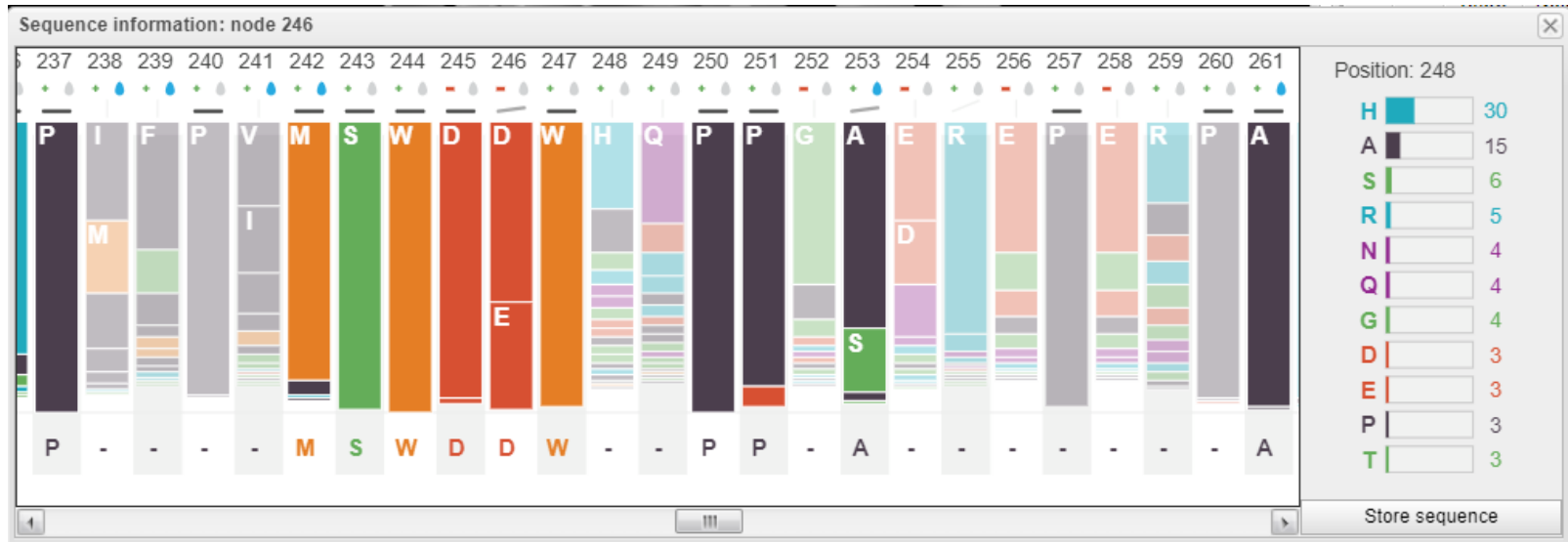
### SUPPORTED BY

# Uživatelské rozhraní

The screenshot displays a software interface for protein structure visualization and analysis. It is divided into several main sections:

- Viewer:** A central window showing a 3D ribbon model of a protein structure in grey on a black background. The label "JSmol" is visible in the bottom right corner of this window.
- Visualization settings:** A panel on the right side containing:
  - Structure visualization style:** A grid of buttons for "Wireframe", "Cartoon", "Sticks", "Tracs", "Balls & sticks", and "Backbone". The "Cartoon" button is currently selected.
  - A "Balls" button.
  - Buttons for "Hide all visualized residues", "Save image", and "Reset view".
  - Visualization quality:** A slider control ranging from 1 to 8.
- Sequence designer:** A panel on the right side containing:
  - Buttons for "Remove selected" and "Remove all".
  - A button for "Download all designs (.zip)".
  - A section titled "DESIRED SEQUENCES" with a table with columns "Show", "Nodes", and "Name". The table is currently empty, displaying "No items to show."
- Mutations:** A panel at the bottom showing a phylogenetic tree. The tree is labeled "Phylogenetic tree" and "Multiple-sequence alignment". It includes options for "Show substitutions" and "Clustal", and a "Show all ancestors" button. The tree branches are labeled with various protein IDs and residue numbers, such as "WP\_154172489.1", "WP\_160497705.1", "WP\_050784989.1", "WP\_054746225.1", "906381.1", "108794117.1", "PPR36569.1", and "KPK22892.1".

# Uživatelské rozhraní







- Databáze pro uchování **stabilitních dat**
- Přes **15,000 experimentů**
- 242 proteinů
- Sekvenční a strukturní predikce
- Interaktivní uživatelské rozhraní
  - Pokročilé vyhledávání, statistiky, vizualizace, ...

# Uživatelské rozhraní

NATURAL VARIANTS

|                   |   |   |   |   |   |   |   |   |   |   |   |
|-------------------|---|---|---|---|---|---|---|---|---|---|---|
| AMINO ACID CHARGE | ● | ● | ▲ | ● | ▼ | ● | ● | ● | ● | ● | ▼ |
| MUTATIONS         |   |   |   |   |   |   |   |   |   |   |   |
| CATALYTIC POCKET  |   |   |   |   |   |   |   |   |   |   |   |
| TUNNELS           |   |   |   |   |   |   |   |   |   |   |   |
| B-FACTORS         | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| CONSERVATION      | ● | ● | ● | ● | ● | ● | ● | ▲ | ▲ | ▼ |   |
| BTC MUTATIONS     |   |   |   |   |   |   |   |   |   |   |   |

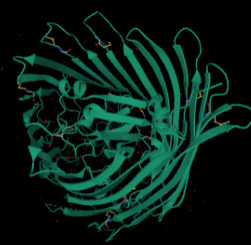
STRUCTURAL FEATURES

Select structure: 1FEP (reference)

Structure: 1FEP (reference)

Method: X-ray

Resolution: 2.4 Å



MUTATIONS LIST

Select slider property:  $\Delta\Delta G$

Legend: ■ Stabilizing ■ Destabilizing ■ Neutral

| Mutation | Curated | $\Delta\Delta G$ (kcal/mol) ↑ | $\Delta T_m$ (°C) | Publication  |
|----------|---------|-------------------------------|-------------------|--------------|
| E272C    | ★       | -9.79                         | -                 | Quan L, 2016 |



# HOTSPOT WIZARD

DESIGN OF MUTATIONS AND SMART LIBRARIES

# HotSpot Wizard

- Identifikace reziduí pro **mutagenezi**
- **Čtyři strategie proteinového inženýrství**
  - Funkcionální hotspoty
  - Analýza flexibility
  - Back-to-consensus
  - Analýza korelovaných pozic
- Predikce vlivu mutací na funkci proteinu

# Uživatelské rozhraní

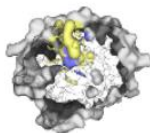
## JOB INFO

Structure: [1CV2](#)  
Title: Example  
ID: w6izlk

✓ Calculation successfully finished. You can now explore results of individual protocols in the section below.

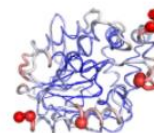
## RESULTS OF PROTEIN-ENGINEERING STRATEGIES

### FUNCTIONAL HOT SPOTS



### STABILITY HOT SPOTS

STRUCTURAL FLEXIBILITY

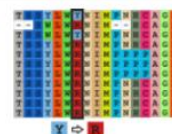


### CORRELATED HOT SPOTS



### STABILITY HOT SPOTS












SEQUENCE CONSENSUS



## DOWNLOAD

Raw data (zip archive)

## REPORT

- ✓ Assigning secondary structure 
- ✓ Identification of flexible residues 
- ✓ Calculation of accessible surface area 
- ✓ Calculation of pockets 
- ✓ Calculation of tunnels 
- ✓ Construction of multiple sequence alignment 
- ✓ Calculation of conservation scores 
- ✓ Identification of sequence consensus mutations 
- ✓ Calculation of deleteriousness of mutations 
- ✓ Calculation of correlated positions 
- ✓ Quality assessment of structure 

# Závěry

- Nutnost zlepšení stability proteinů
- Nákladné a zdlouhavé laboratorní experimenty
- **FireProt:** návrh stabilních mutantů
- **FireProt-ASR:** ancestrální rekonstrukce
- **FireProt-DB:** uchování stabilitních dat
- **HotSpot Wizard:** identifikace hotspotů

# Poděkování



MUNI

