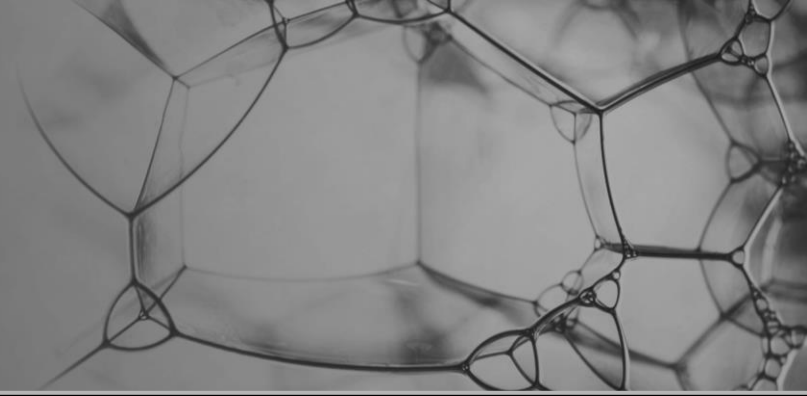


**LOSCHMIDT**  
**LABORATORIES**



# Vývoj software v Loschmidtových laboratořích

Jan Štourač

# Vývoj software



1. Nové algoritmy a metody
2. Komplexní bioinformatická workflow

# Vývoj software



- 1. Nové algoritmy a metody**
- 2. Komplexní bioinformatická workflow**

# Motivace



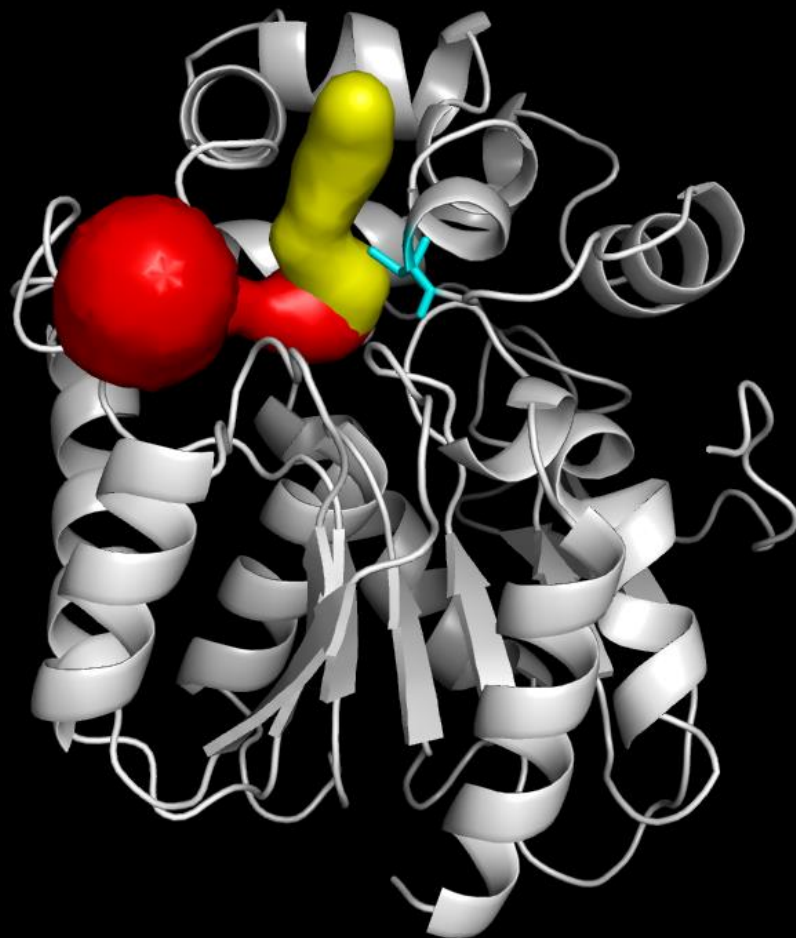
- **Nový** nápad 😊
- **Problémy** existujících programů
  - Nízká/nedostatečná přesnost
  - Nespolehlivost
  - Zastaralost
  - Výpočetní náročnost



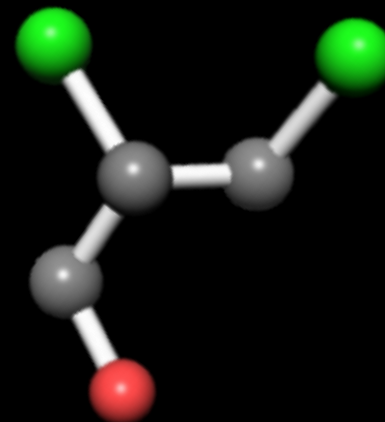
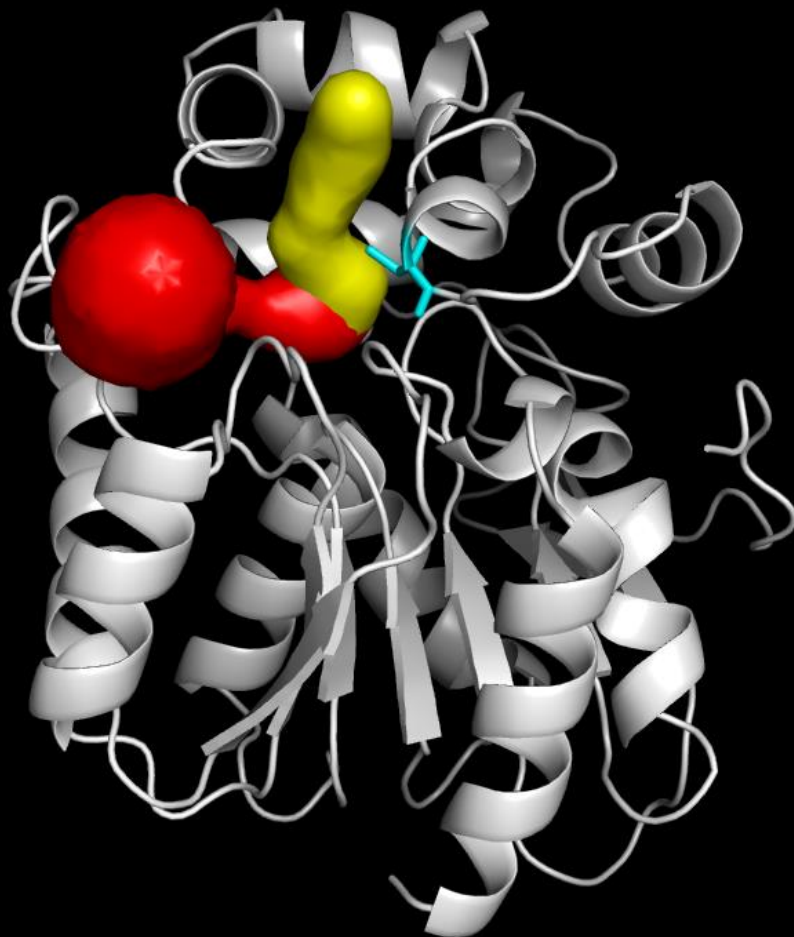
Automatická analýza **transportu**  
**malých molekul** tunely v proteinech s využitím  
**molekulového dokování.**



# CaverDock

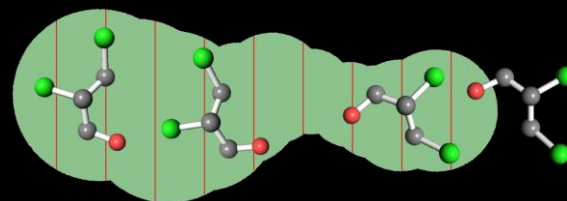
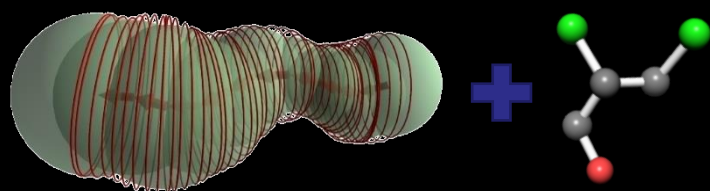
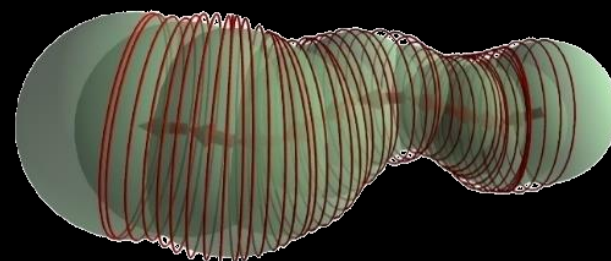
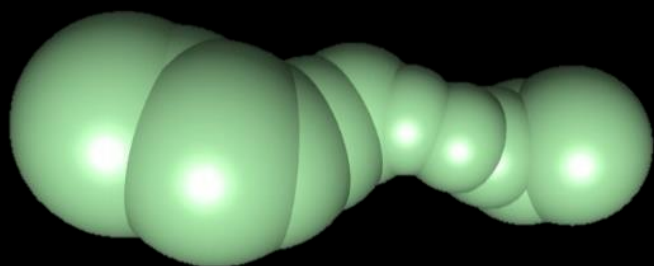


# CaverDock



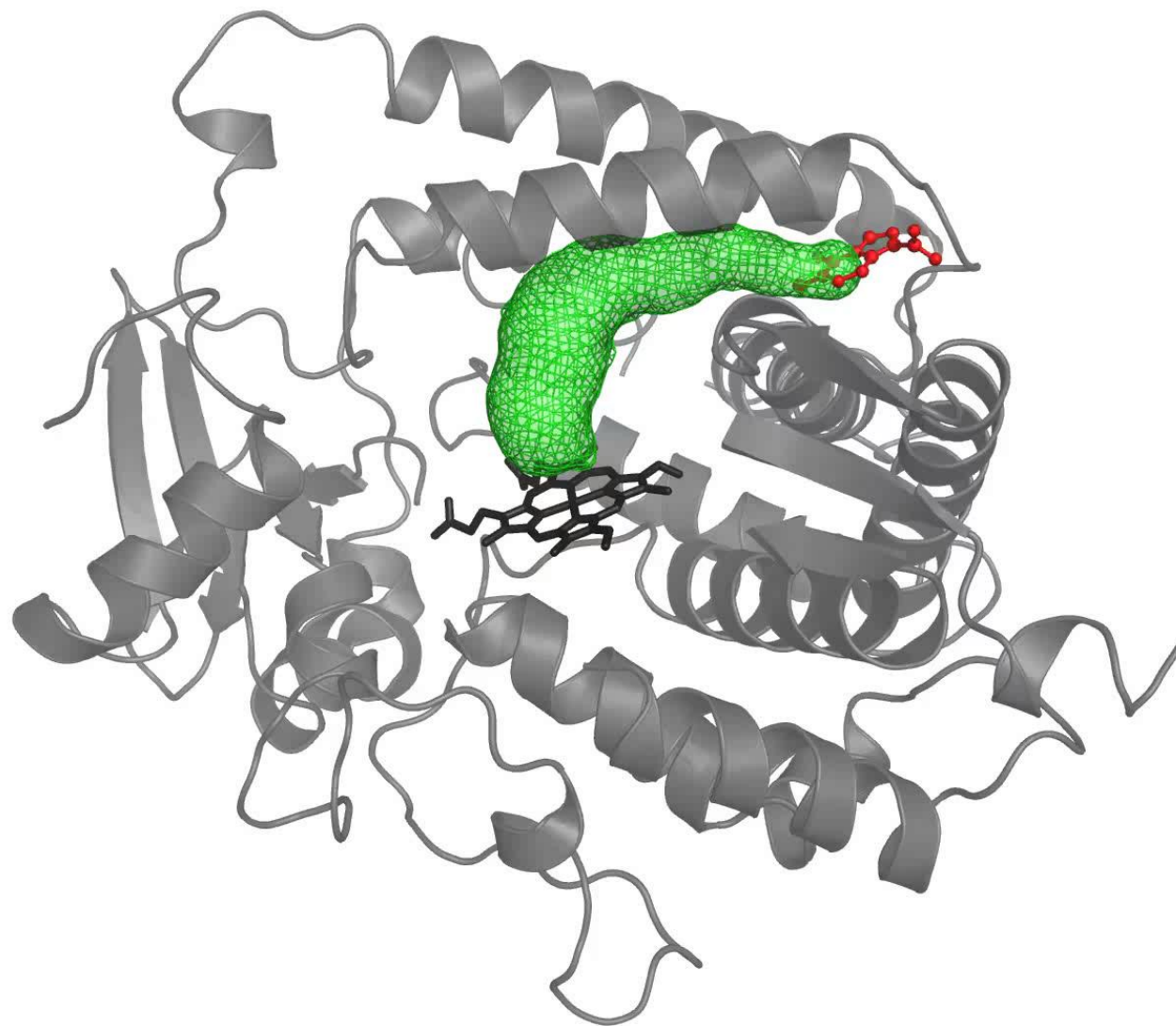
?

# CaverDock

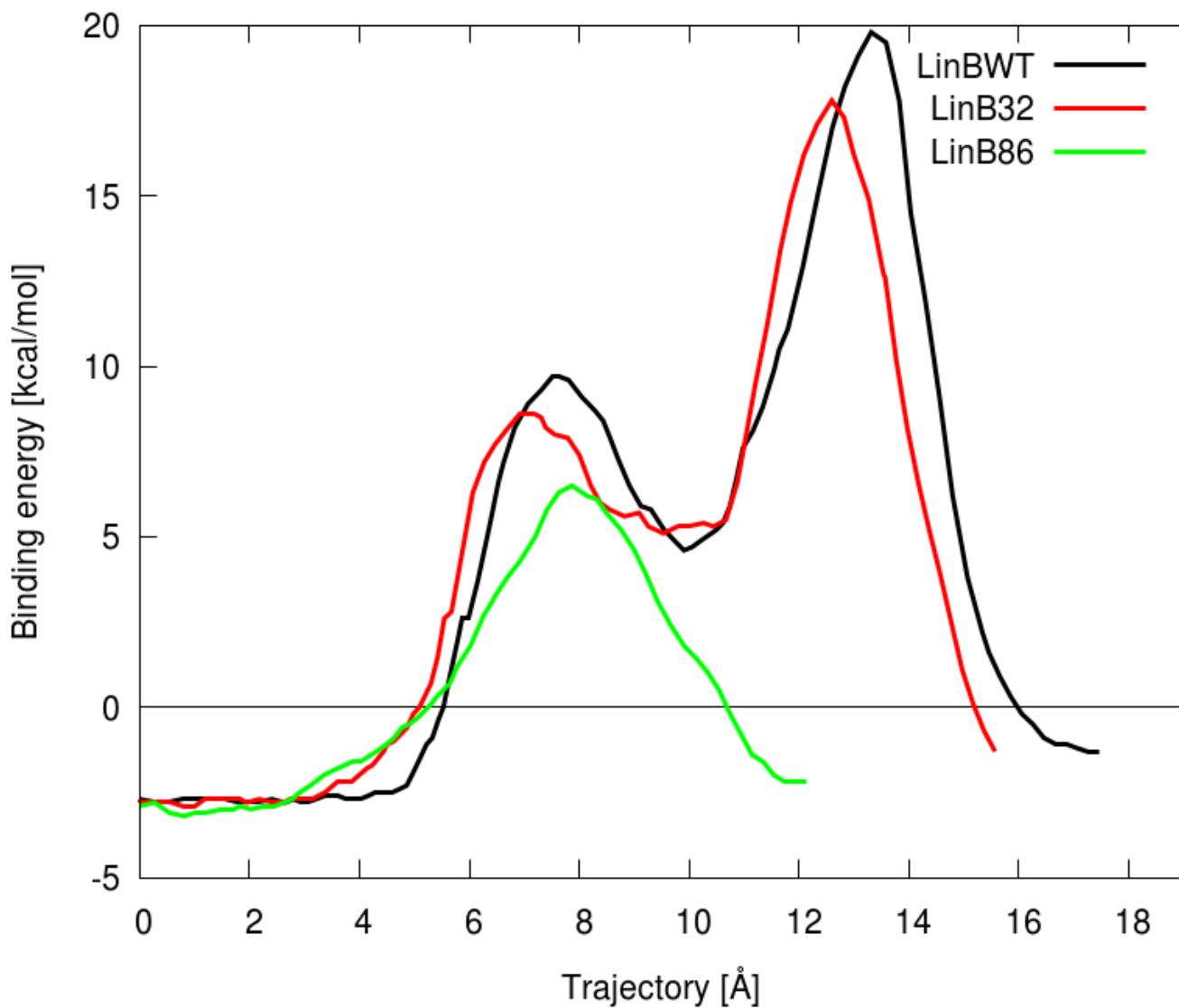




# CaverDock



# CaverDock

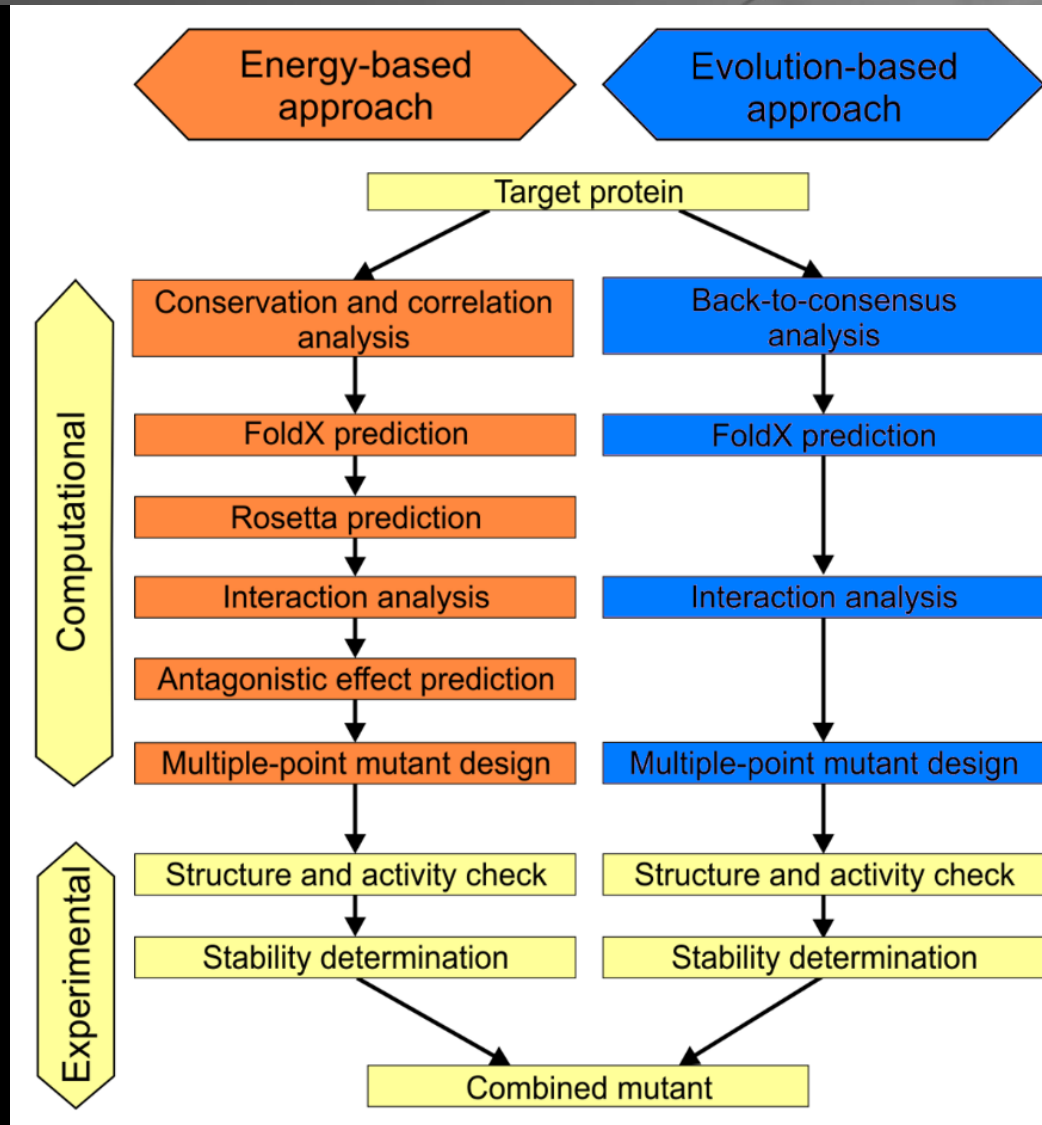


# Vývoj software

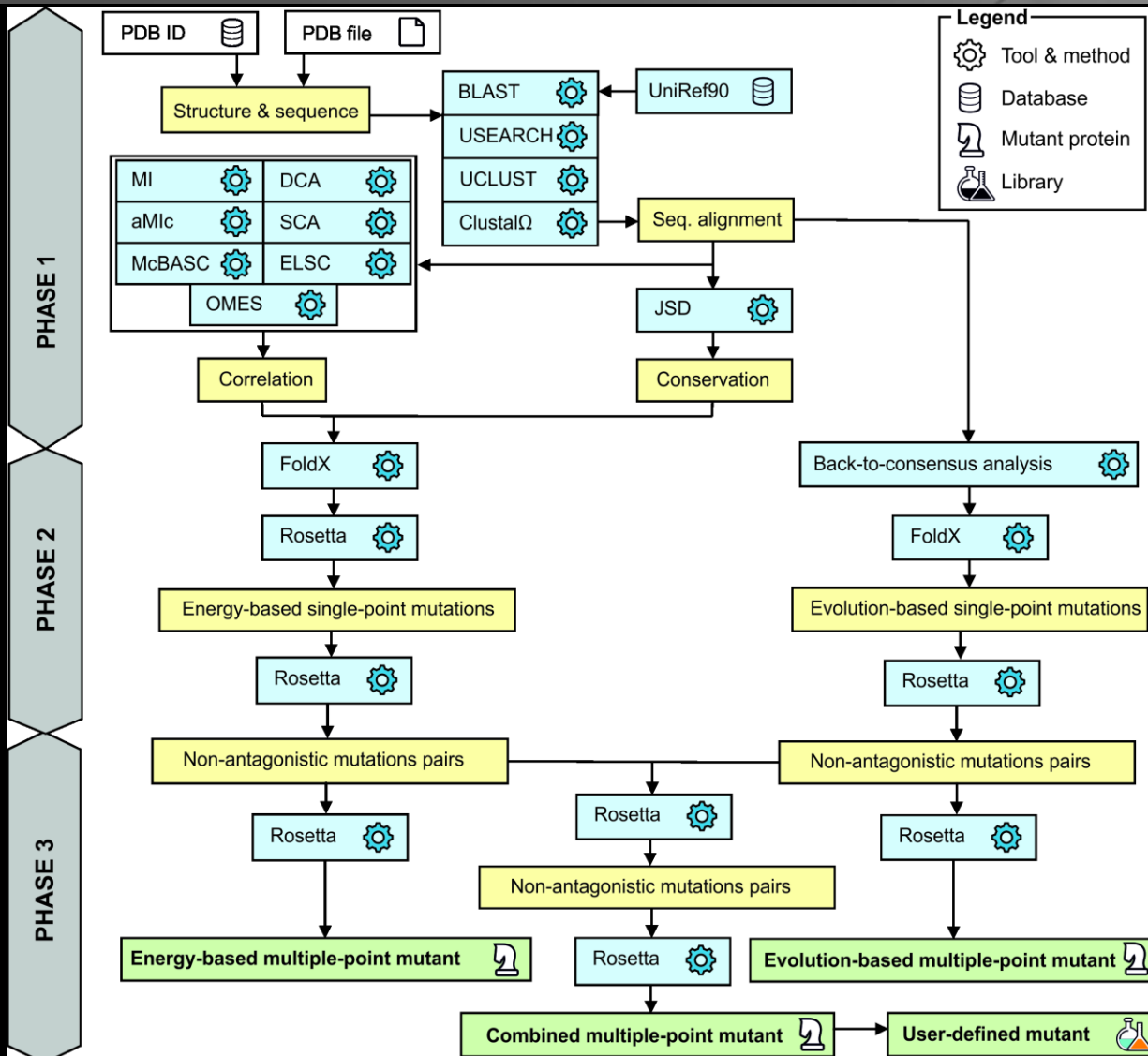


1. Nové algoritmy a metody
2. **Komplexní bioinformatická workflow**

# Motivace



# Motivace



15 nástrojů  
2 databáze

- Náročná **instalace**
  - Většinou podpora pouze pro Linux
  - Nedostupnost balíčků v repozitářích → nutnost kompilace
  - Zastaralý zdrojový kód
  - Obskurní knihovny
  - Velké množství (desítky) nástrojů

- Náročné **používání**
  - Často žádné uživatelské rozhraní
  - Velké množství konfiguračních parametrů (i stovky)
  - Každý nástroj má své “háčky”
  - Výstupy většinou textové nebo binární soubory

- Náročné **propojení** a **analýza výsledků**
  - Výstupy a vstupy často nejsou kompatibilní
  - Výsledky je nutné manuálně kombinovat a vizualizovat



- Nedostatek **výkonu**
  - Málokdo má k dispozici velké výpočetní zdroje
  - Náročné na správu a používání
- 1 výpočet FireProt = ~3 000 CPU hodin
  - → na běžném notebooku cca 31 dní

# Motivace



Computational Design of Thermostable Multiple-Point Mutants



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

Job ID:

[Find job](#)

## INPUT STRUCTURE

[Load example](#)

Source :  Enter PDB code  
 Upload PDB file

PDB ID :

Biological unit : 

molecule	chains	quaternary structure
<input checked="" type="radio"/> 1	<input type="radio"/> A	<input type="radio"/> Monomeric

Specify chains manually

## REFERENCE

Musil M, Stourac J, Bendi J, Brezovsky J, Prokop Z, Zendulka J, Martinek T, Bednar D, Damborsky J: FireProt: Web Server for Automated Design of Thermostable Proteins. Nucleic Acids Research - Web Issue (submitted).

## USER STATISTICS

• Number of visitors: 177

## JOB INFORMATION

Job title (optional) :

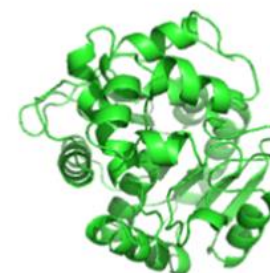
E-mail (optional) :

[Previous](#)

### Details about biological unit 0

Biomolecule:	1
Burried surface area [Å <sup>2</sup> ]:	-
Surface area [Å <sup>2</sup> ]:	-
Change in solvent free energy [kcal/mol]:	-

chain	original chain	residues	atoms
A	A	4 - 2542	1 - 3073



# Motivace

The screenshot displays a molecular viewer interface with the following components:

- Viewer:** A 3D ribbon representation of a protein structure. Mutations are highlighted in orange and blue.
- Visualization settings:** A panel on the right with the following options:
  - Structure visualization style: Wireframe, **Cartoon**, Sticks, Trace, Balls & sticks, Backbone, Balls.
  - Buttons: Hide all visualized residues, Save image, Reset view.
  - Visualization quality: A slider set to 1, ranging from 1 to 8.
- FireProt protocol design:** A table with the following data:

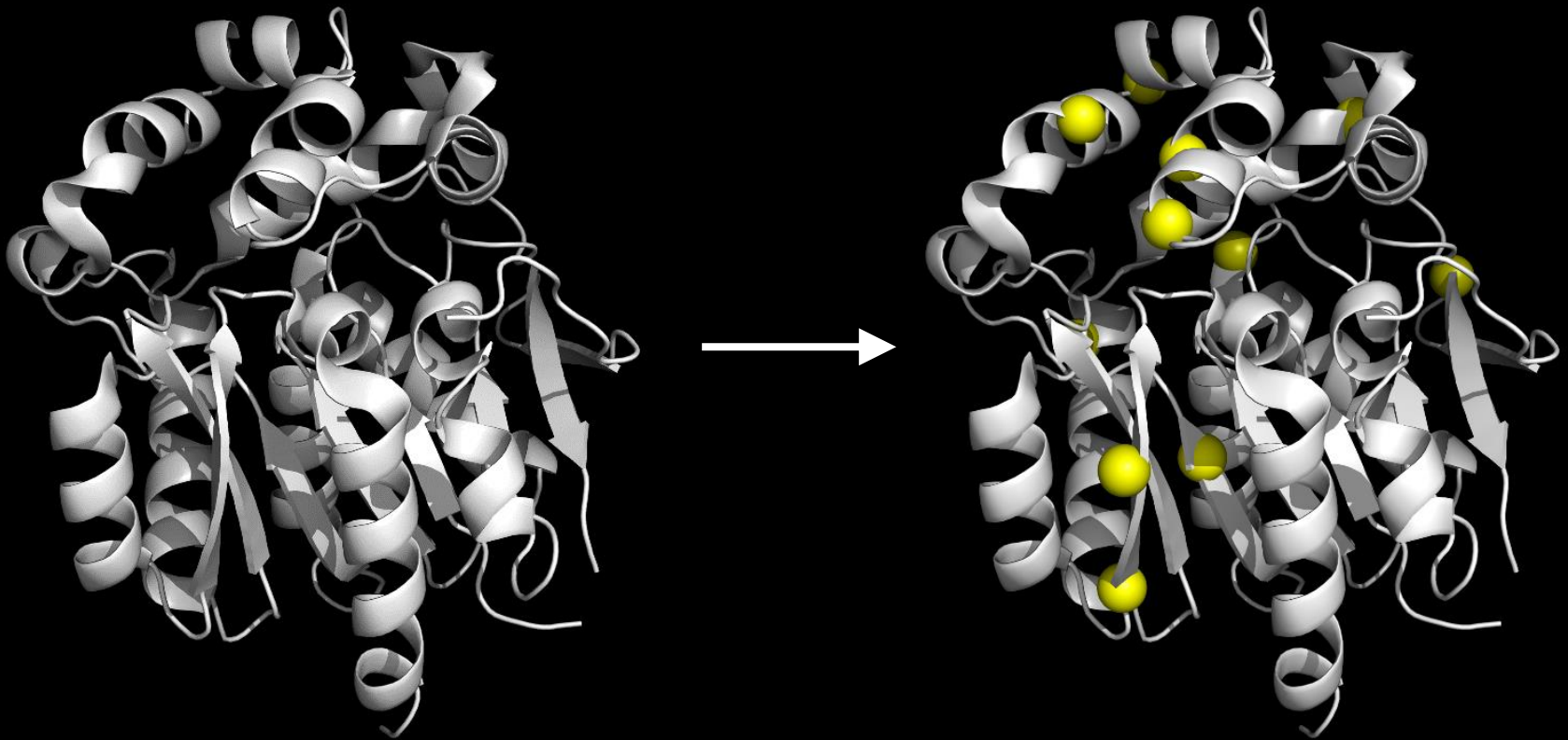
PDB ID:	4e46
Length:	292
Evolution mutant:	-3.7 kcal/mol (6 mutations)
Energy mutant:	-20.85 kcal/mol (8 mutations)
- Mutant designer:** A panel with buttons for "Original selection", "Save mutant", and "Download all designs (.zip)". Below is a table for the energy mutant:

ENERGY MUTANT		DDG: -20.85 KCAL/MOL				
	chain	position	ref	alt	foldx	rosetta
[-] A	A	11	D	P	-1.39	-1.89
[-] A	A	33	T	I	-1.31	-1.94
[-] A	A	145	A	L	-2.77	-1.71

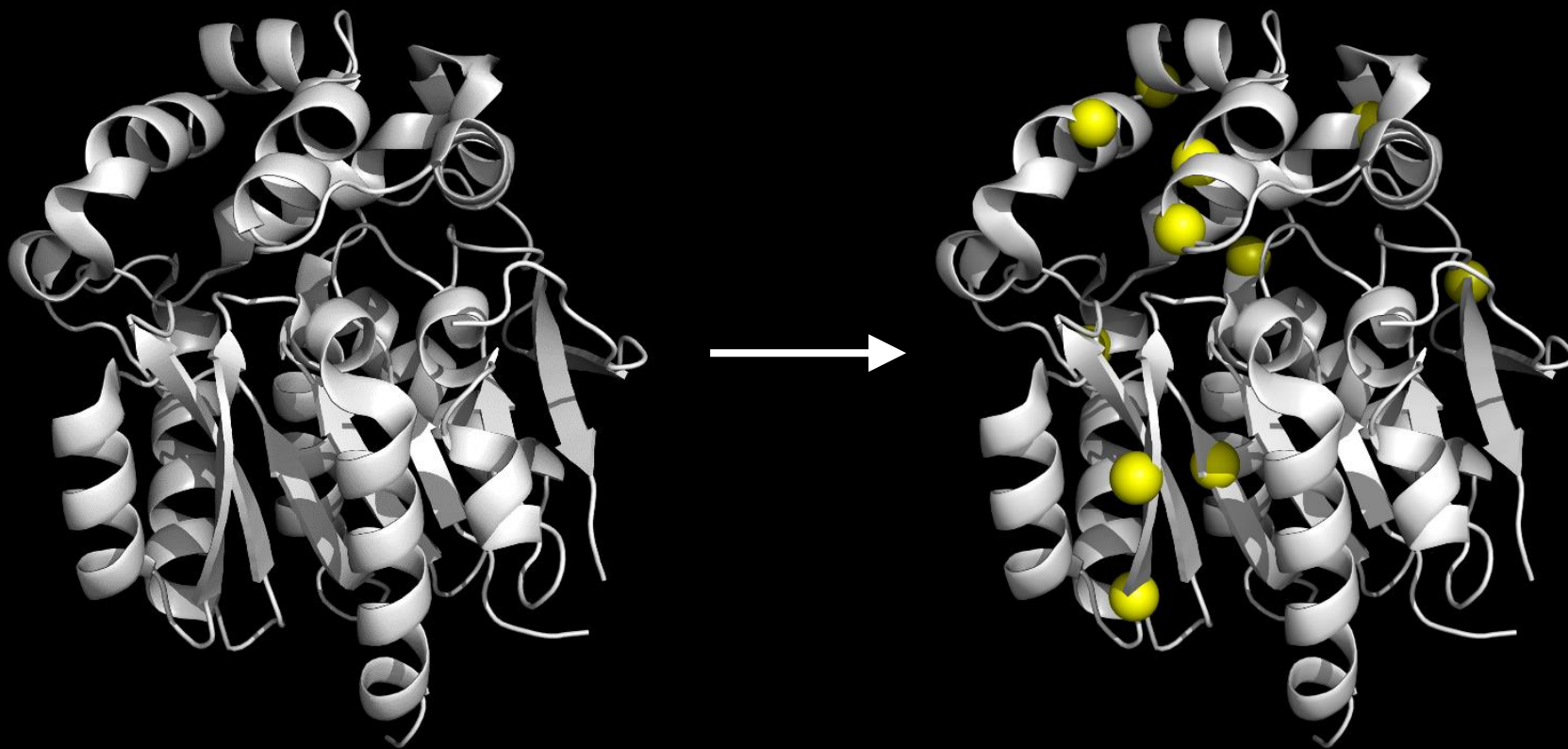
Automatická identifikace **reziduí vhodných k mutagenezi** a návrh „**chytrých knihoven**“ pro proteinové inženýrství.



# HotSpot Wizard



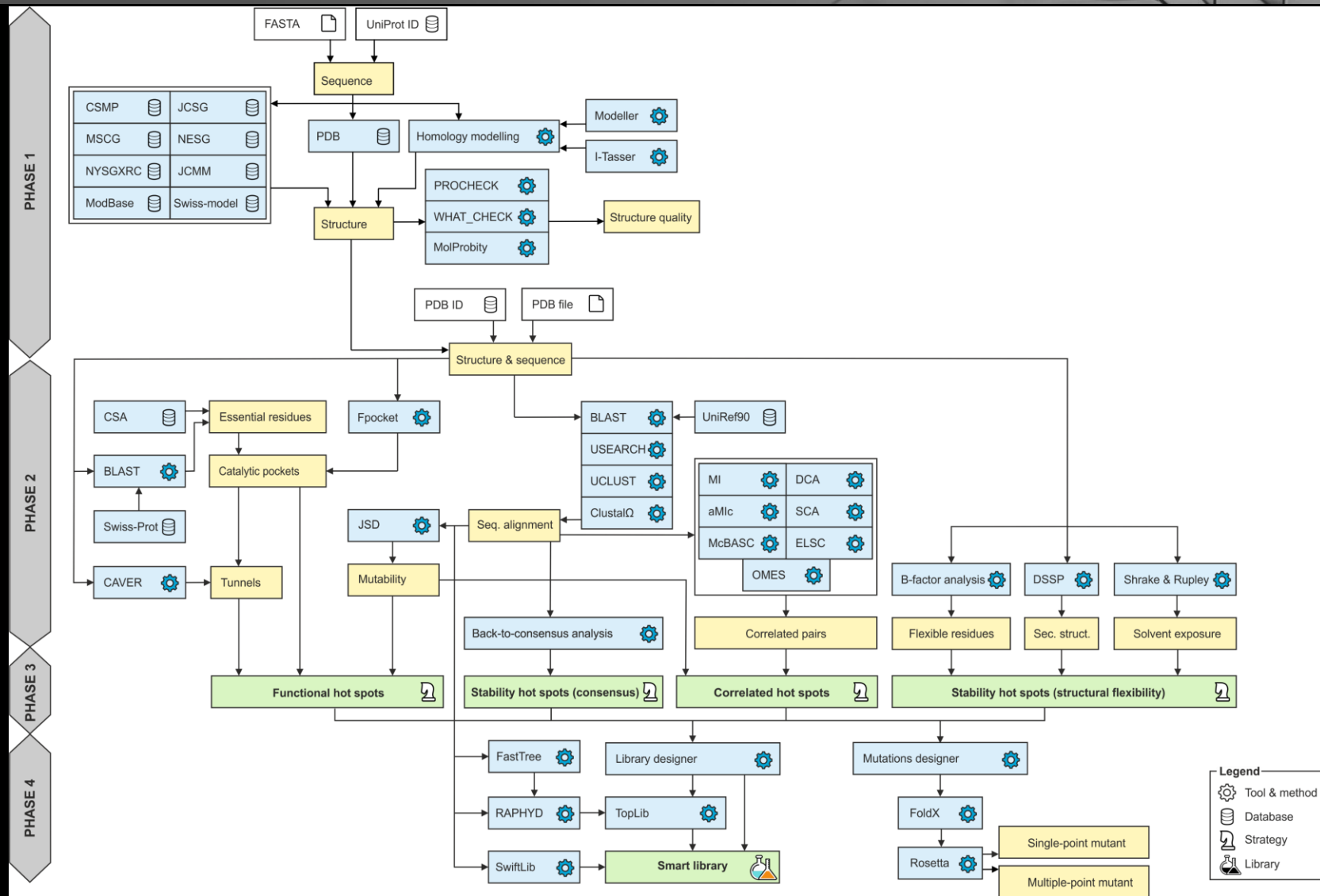
# HotSpot Wizard



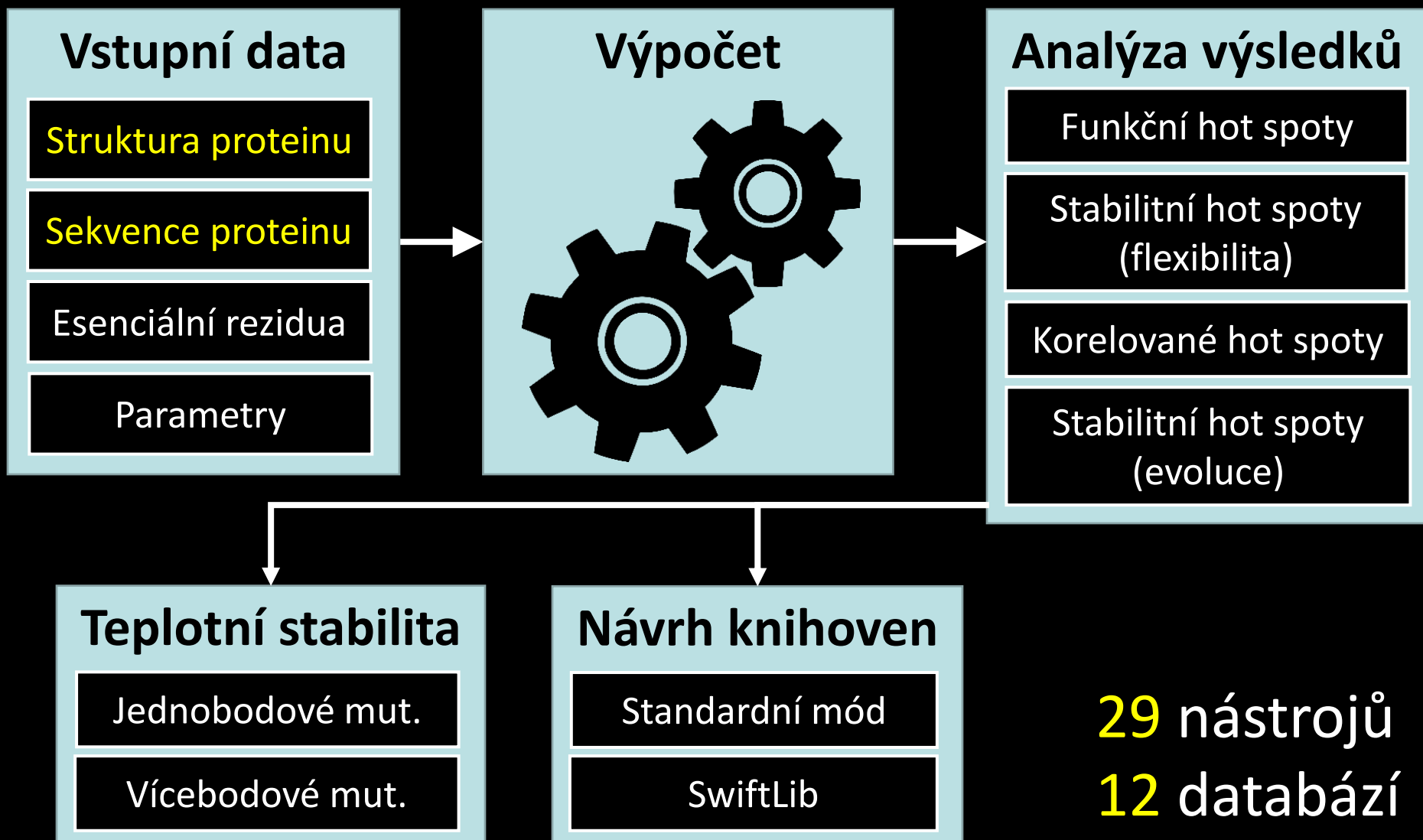
zvýšená termostabilita  
o 24° C



# HotSpot Wizard



# HotSpot Wizard



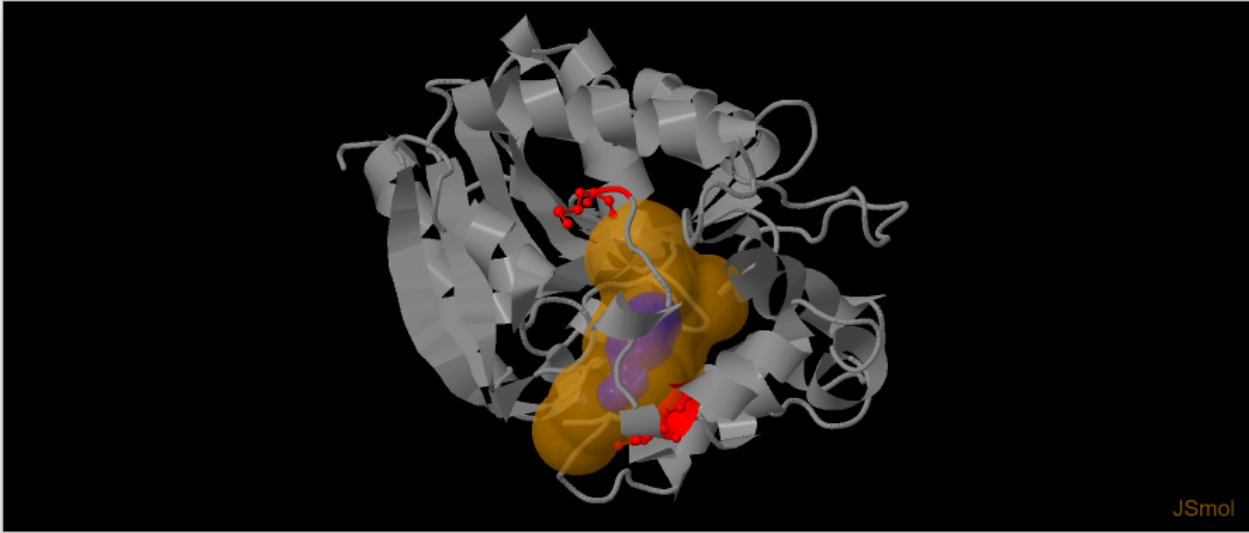
**29** nástrojů  
**12** databází



# HotSpot Wizard

## Functional hot spots of 1CV2

**Viewer**



JSmol

Return to Results browser

**Visualization settings**

**Tunnels**

id	length (Å)	bottleneck radius (Å)
Starting from pocket: 1		
<input checked="" type="checkbox"/> 1	7.7	1.5

**Pockets**

id	chain(s)	relevance (%)	volume (Å <sup>3</sup> )
<input checked="" type="checkbox"/> 1	A	100	576
<input type="checkbox"/> 2	A	82	883
<input type="checkbox"/> 3	A	62	275
<input type="checkbox"/> 4	A	28	753
<input type="checkbox"/> 5	A	25	183
<input type="checkbox"/> 6	A	19	119
<input type="checkbox"/> 7	A	19	632

**Residues selected for mutagenesis**

Zoom residues    **Design library**

	chain	position	residue	HotSpot
<input checked="" type="checkbox"/> <input type="radio"/>	A	146	Gln	✓
<input checked="" type="checkbox"/> <input type="radio"/>	A	136	Met	✓
<input checked="" type="checkbox"/> <input type="radio"/>	A	147	Asp	✓

**Residue features**

Exclude correlated positions  
  Exclude catalytic pockets  
  Exclude tunnels  
  Exclude α-helices and β-sheets  
 Show all residues

Exclude buried residues  
  Include residues with moderate mutability

	chain	position	residue	mutable	non-essential	in tunnel	in catalytic pocket	HotSpot
<b>Chain A</b>								
<input checked="" type="checkbox"/>	A	146	Gln	✓	✓	✓	✓	✓
<input checked="" type="checkbox"/>	A	136	Met	✓	✓	✗	✓	✓
<input checked="" type="checkbox"/>	A	147	Asp	✓	✓	✓	✓	✓
<input type="checkbox"/> <input checked="" type="radio"/>	A	271	Ala	✓	✓	✓	✓	✓
<input type="checkbox"/> <input checked="" type="radio"/>	A	138	Ile	✓	✓	✗	✓	✓
<input type="checkbox"/> <input checked="" type="radio"/>	A	247	Ala	✓	✓	✓	✓	✓
<input type="checkbox"/> <input checked="" type="radio"/>	A	248	Leu	✓	✓	✓	✓	✓
<input type="checkbox"/> <input checked="" type="radio"/>	A	249	Thr	✓	✓	✓	✗	✓
<input type="checkbox"/> <input checked="" type="radio"/>	A	253	Met	✓	✓	✗	✓	✓



CAVER 3.0, CaverAnalyst 2.0  
analysis of protein tunnels and channels



CAVERDOCK 1.0  
energetic analysis of passage of ligands



HOTSPOT WIZARD 3.0  
safe mutagenesis and smart libraries



FIREPROT 1.0  
design of thermostabilizing mutations



ENZYME MINER 1.0  
mining of novel enzymes in genetic databases



SOLUPROT 1.0  
prediction of protein solubility



CALFITTER 1.0  
analysis of thermal denaturation data

# Technologie

- Frontend

- HTML, CSS, JavaScript (TypeScript)
- React, Redux, d3.js

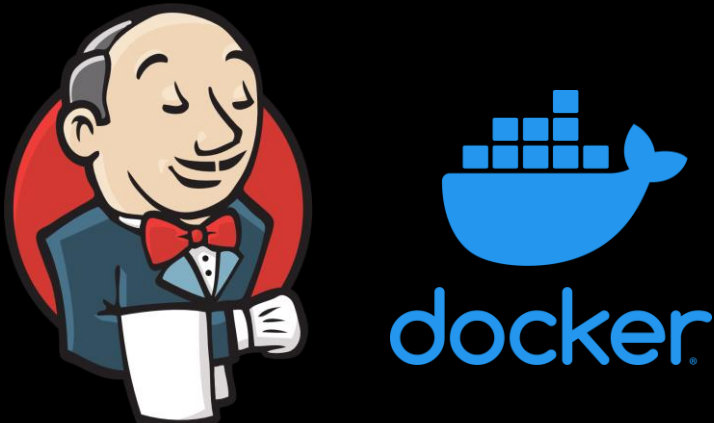
- Backend

- Java, Python
- Spring, Hibernate, Lombok
- MySQL/MariaDB



# Technologie

- Deployment
  - Git, Gitlab
  - Redmine,
  - Jenkins
  - CentOS, Docker

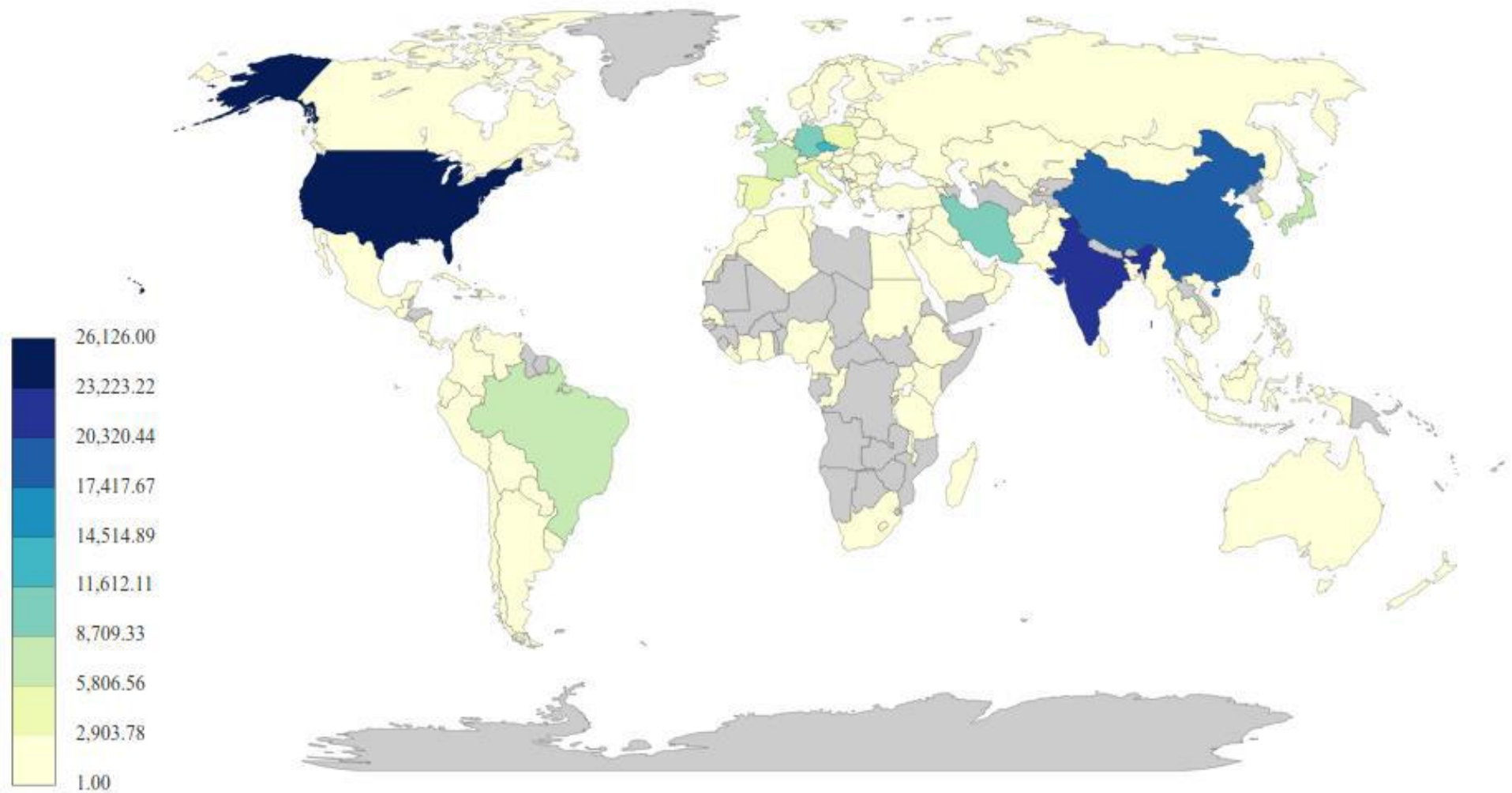


# Infrastruktura

- Vlastní produkční a vývojové prostředí
  - 47 serverů
  - 1500 CPU jader
  - 24 TB RAM
- MetaCentrum
  - > 200 serverů
  - 20 000 CPU jader
  - 19 PB úložiště



# Statistiky používání



>100 000 uživatelů

>111 000 úloh



- Co **nabízíme**
  - Vědecky atraktivní témata na pomezí biologie, chemie a informatiky s celosvětovým dosahem
  - Možnost podílet se na:
    - Návrhu algoritmu nebo workflow a jeho validaci
    - Vývoji a testování software
  - Možnost pracovat s moderními technologiemi
    - Včetně zaškolení
  - Úzká spolupráce s modeláři a experimentátory

- Co **očekáváme**
  - Praktické zkušenosti se:
    - Skriptováním – vývoj a validace workflow a metod
    - Programováním – vývoj software
  - Ochotu se vzdělávat (v informatice i biologii)
  - Ochotu na práci strávit více než poslední semestr



# Kontakt

## Jan Štourač

- UČO: 113869
- Budova A13, místnost 108
- E-mail: [stourac@mail.muni.cz](mailto:stourac@mail.muni.cz)
- Web: [loschmidt.chemi.muni.cz/peg/](http://loschmidt.chemi.muni.cz/peg/)





CAVER 3.0, CaverAnalyst 2.0  
analysis of protein tunnels and channels



CAVERDOCK 1.0  
energetic analysis of passage of ligands



HOTSPOT WIZARD 3.0  
safe mutagenesis and smart libraries



FIREPROT 1.0  
design of thermostabilizing mutations



ENZYME MINER 1.0  
mining of novel enzymes in genetic databases



SOLUPROT 1.0  
prediction of protein solubility



CALFITTER 1.0  
analysis of thermal denaturation data