

**LOSCHMIDT
LABORATORIES**



PROTEIN ENGINEERING

7. Rational and semi-rational design

Loschmidt Laboratories

Department of Experimental Biology

Masaryk University, Brno

Outline



- ❑ Protein engineering approaches
- ❑ Semi-rational design
 - identification of hot-spots
 - evaluation of hot-spots
 - selection of substitutions
 - design of library
 - mutagenesis and screening
- ❑ Rational design
 - molecular modeling

Outline



- ❑ Protein engineering approaches
- ❑ Semi-rational design
 - identification of hot-spots
 - evaluation of hot-spots
 - selection of substitutions
 - design of library
 - mutagenesis and screening
- ❑ Rational design
 - molecular modeling

Protein engineering

- altering protein structure to improve its properties
- three main approaches
 - rational design
 - directed evolution
 - semi-rational design

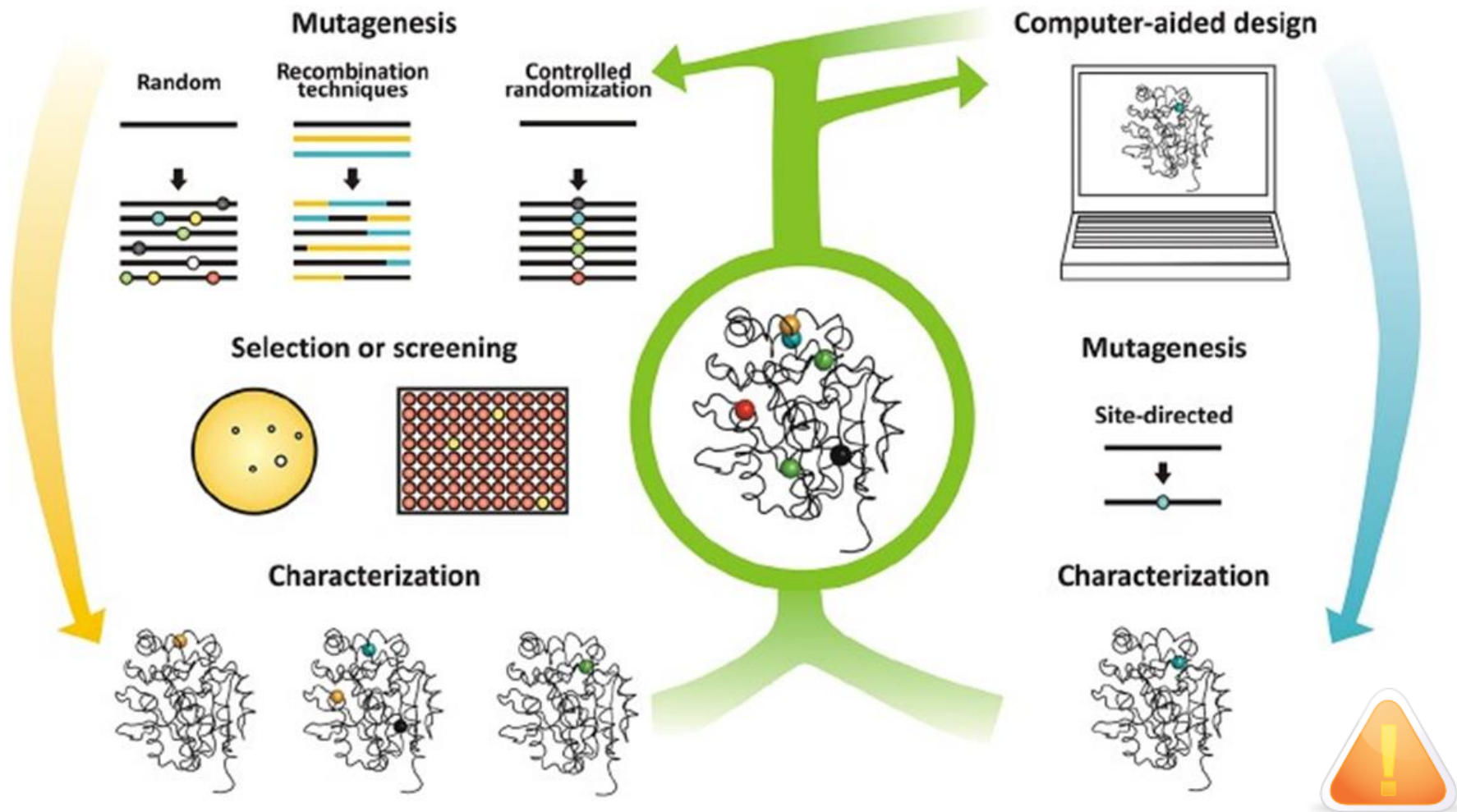


Protein engineering approaches

DIRECTED EVOLUTION

SEMI-RATIONAL DESIGN

RATIONAL DESIGN



Protein engineering approaches



| | Rational design | Directed evolution | Semi-rational design |
|---|-----------------|--------------------|--------------------------------|
| high-throughput screening/selection | not essential | essential | advantageous but not essential |
| structural and/or functional information | both essential | neither essential | either is sufficient |
| sequence space exploration | low | high, random | moderate, targeted |
| probability to obtain synergistic mutations | moderate | low | high |



Structural information

❑ worldwide Protein Data Bank (wwPDB)

- <http://www.wwpdb.org/>
- central repository of ~220,000 experimental macromolecular structures (April 2024)

❑ RCSB PDB



- <https://www.rcsb.org>

❑ PDBe



- <https://www.ebi.ac.uk/pdbe>

❑ PDBj



- <https://pdbj.org>

Structure Summary | Structure | Annotations | Experiment | Sequence | Genome | Versions

Biological Assembly 1

4E46

Structure of Rhodococcus rhodochrous haloalkane dehalogenase DhaA in complex with 2-propanol

PDB DOI: <https://doi.org/10.2210/pdb4E46/pdb>

Classification: HYDROLASE

Organism(s): Rhodococcus rhodochrous

Expression System: Escherichia coli BL21

Mutation(s): No

Deposited: 2012-03-12 Released: 2013-03-13

Deposition Author(s): Stsiapanava, A., Chaloupkova, R., Damborsky, J., Kula Smatanova, I.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.26 Å

R-Value Work: 0.113

R-Value Observed: 0.113

Global Symmetry: Asymmetric - C1

Global Stoichiometry: Monomer - A1

wwPDB Validation

| Metric | Percentile Ranks | Value |
|-----------------------|------------------|-------|
| Clashscore | 4 | 4 |
| Ramachandran outliers | 0 | 0 |
| Sidechain outliers | 1.2% | 1.2% |
| RSRZ outliers | 0.3% | 0.3% |

Structural prediction

□ Alpha Fold 2

- Galaxy: https://usegalaxy.eu/?tool_id=alphafold, Colab: <https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>
- structure prediction directly from sequence using deep learning, evolutionary information (MSA), and structure optimization
- Multimer mode – lower accuracy
- Not precise in sidechain orientations prediction (not appropriate for protein-ligand interaction - molecular docking)
- Rare folds, alternative conformations, and co-factors not predicted



❑ Protein engineering approaches

❑ **Semi-rational design**

- identification of hot-spots
- evaluation of hot-spots
- selection of substitutions
- design of library
- mutagenesis and screening

❑ Rational design

- molecular modeling

Semi-rational design

- ❑ combine advantages of **rational and random** approaches
- ❑ selection of promising target sites (hot-spots) → mutagenesis
→ creation of small **“smart” libraries**
- ❑ based on **knowledge** of protein structure and function
- ❑ 😊 high-throughput screening usually not needed
- ❑ 😊 increased chance of obtaining variants with desired properties
- ❑ ☹️ certain knowledge of protein structure-function relationships is still required, 😊 but not that much





- Protein engineering approaches

- **Semi-rational design**

- identification of hot-spots
- evaluation of hot-spots
- selection of substitutions
- design of library
- mutagenesis and screening

- Rational design

- molecular modeling

Identification of hot-spots



- ❑ hot-spots for engineering **catalytic properties**
- ❑ hot-spots for engineering **thermostability**

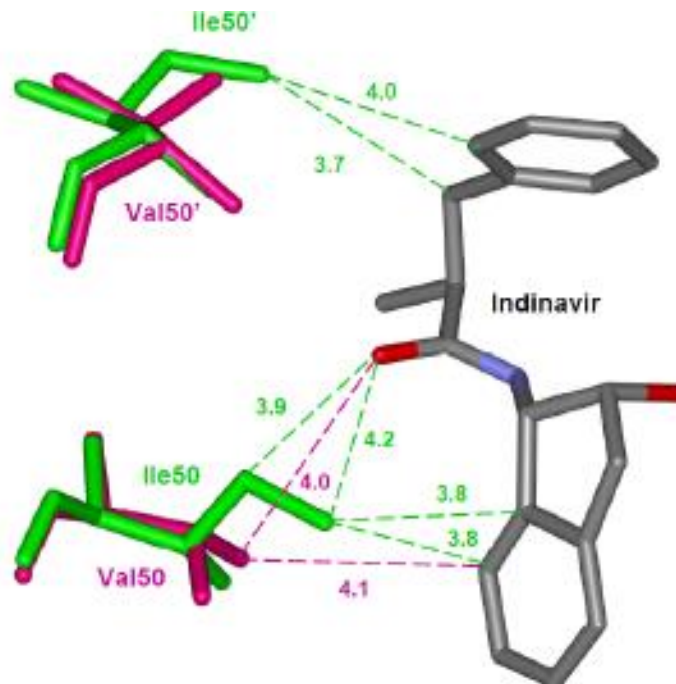
Hot-spots for engineering catalytic properties

- residues mediating substrate binding, transition-state stabilization or product release → mutations can improve or disrupt binding, catalysis or ligand transport
 - residues involved in **protein-ligand interactions**
 - residues located in **binding pockets**
 - residues located in **access tunnels**
- these residues also include **catalytic** or other **essential** residues which generally should not be mutated!



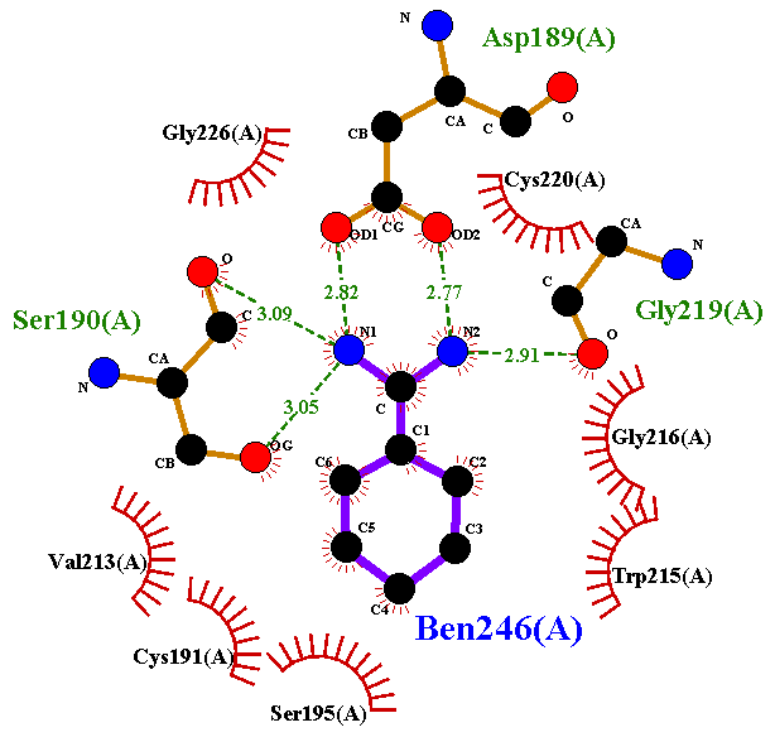
Analysis of protein-ligand interactions

- requires 3D structure of protein-ligand complex
 - experimental structure (wwPDB, PDBbind)
 - theoretical model (molecular docking)

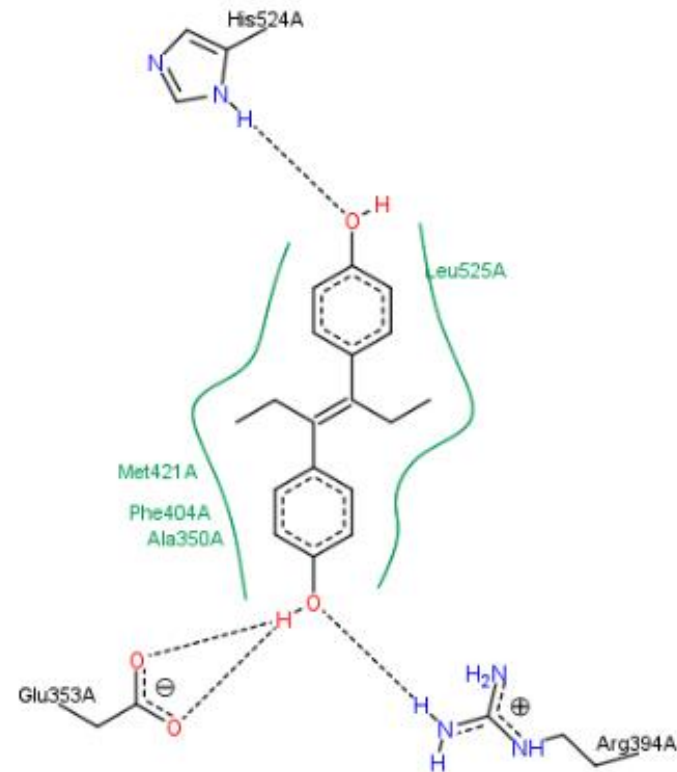


Analysis of protein-ligand interactions

- schematic diagrams of protein-ligand interactions



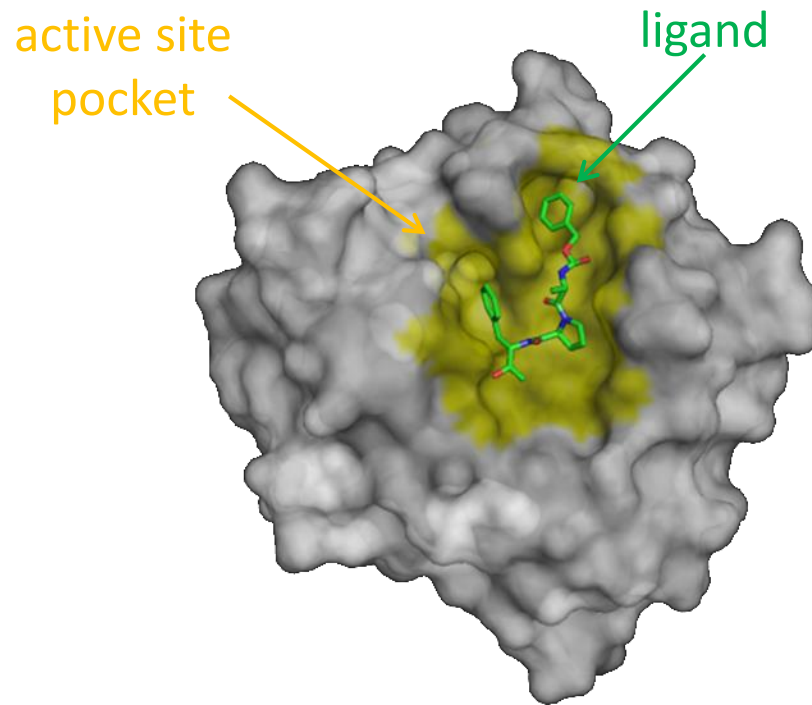
LigPlot, LigPlot+



PoseView

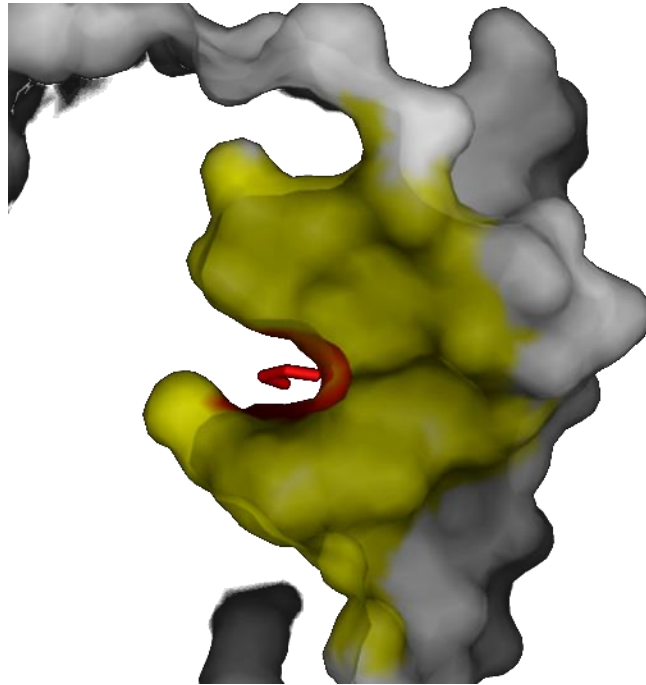
Analysis of binding pockets

- binding and active sites of enzymes are often associated with structural pockets and cavities



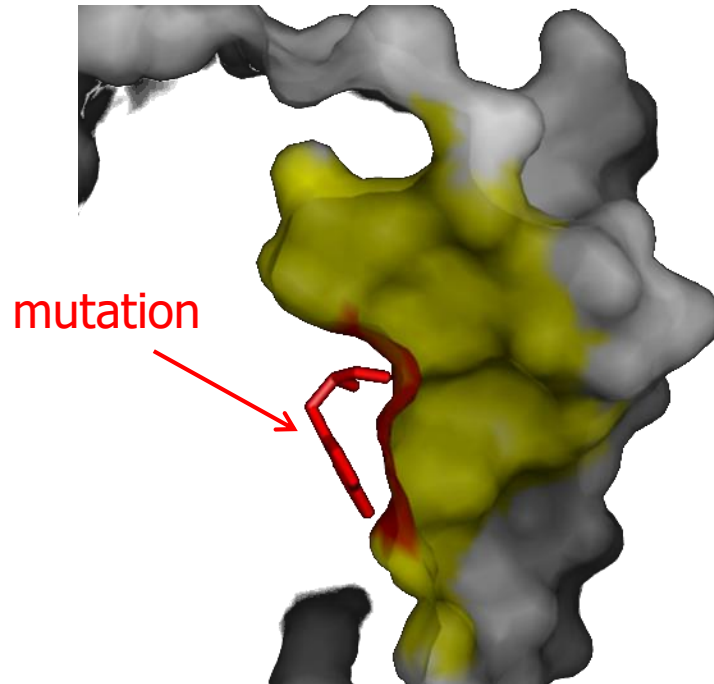
Analysis of binding pockets

- binding and active sites of enzymes are often associated with structural pockets and cavities



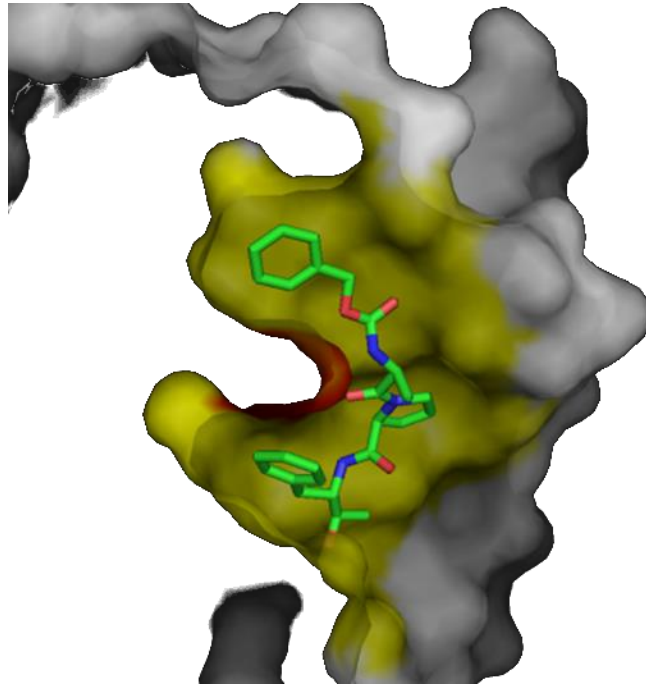
Analysis of binding pockets

- binding and active sites of enzymes are often associated with structural pockets and cavities



Analysis of binding pockets

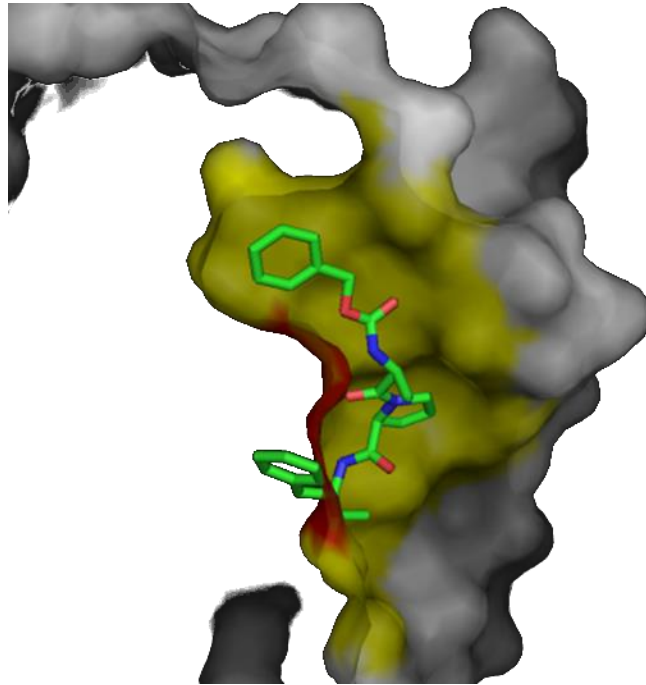
- binding and active sites of enzymes are often associated with structural pockets and cavities



A 3D molecular model illustrating a binding pocket. The protein structure is shown as a grey surface, with a specific region highlighted in yellow. A ligand molecule, represented by a stick model with green, red, and blue atoms, is bound within the pocket. The ligand consists of a central ring system connected to a chain of atoms, including a nitrogen atom (blue) and a carbonyl group (red and green).

Analysis of binding pockets

- binding and active sites of enzymes are often associated with structural pockets and cavities



A 3D molecular model illustrating an enzyme binding pocket. The enzyme's surface is shown in a semi-transparent grey, revealing a deep, irregular cavity. The binding pocket is highlighted in a yellowish-green color. A ligand molecule is bound within this pocket, represented by a stick model with green carbon atoms, red oxygen atoms, and blue nitrogen atoms. The ligand is oriented vertically, fitting snugly into the pocket's structure.

Analysis of binding pockets


- binding and active sites of enzymes are often associated with **structural pockets** and **cavities**
 - most amino acid residues located in these pockets may come into contact with the ligands during the catalytic cycle
 - one can accurately predict which residues may interact with the ligand even without precise knowledge of ligand orientation in the active site
- requires 3D structure of protein
- software for detection of pockets
 - CASTp, fPocket, CavityPlus, etc.



Analysis of binding pockets

- detailed characterization of all pockets in the structure

CavityPlus INTRODUCTION COMPUTING RESULTS TOOLBOX TUTORIAL REFERENCE RESOURCE CONTACT US



Cavity ▼

CavPharmer ▲

Step 1: Select a cavity

Select a cavity

Cavity_1 Cavity_2 Cavity_3
 Cavity_4 Cavity_5 Cavity_6
 Cavity_7

Step 2: Choose mode

Mode ? Use Ligand Mode

[Submit](#)

Cavity Results ▲

[Download results](#)

| # 1↓ | Pred Max pKd 1↓ | Pred Ave pKd 1↓ | DrugScore 1↓ | Druggability 1↓ | Surface | More |
|------|-----------------|-----------------|--------------|-----------------|-------------------------------------|------|
| 1 | 10.19 | 6.11 | 493.00 | Medium | <input checked="" type="checkbox"/> | ... |
| 2 | 8.87 | 5.66 | -745.00 | Weak | <input type="checkbox"/> | ... |
| 3 | 8.16 | 5.42 | -420.00 | Weak | <input type="checkbox"/> | ... |

CorrSite ▲

Step 1: Select/Upload an orthosteric pocket

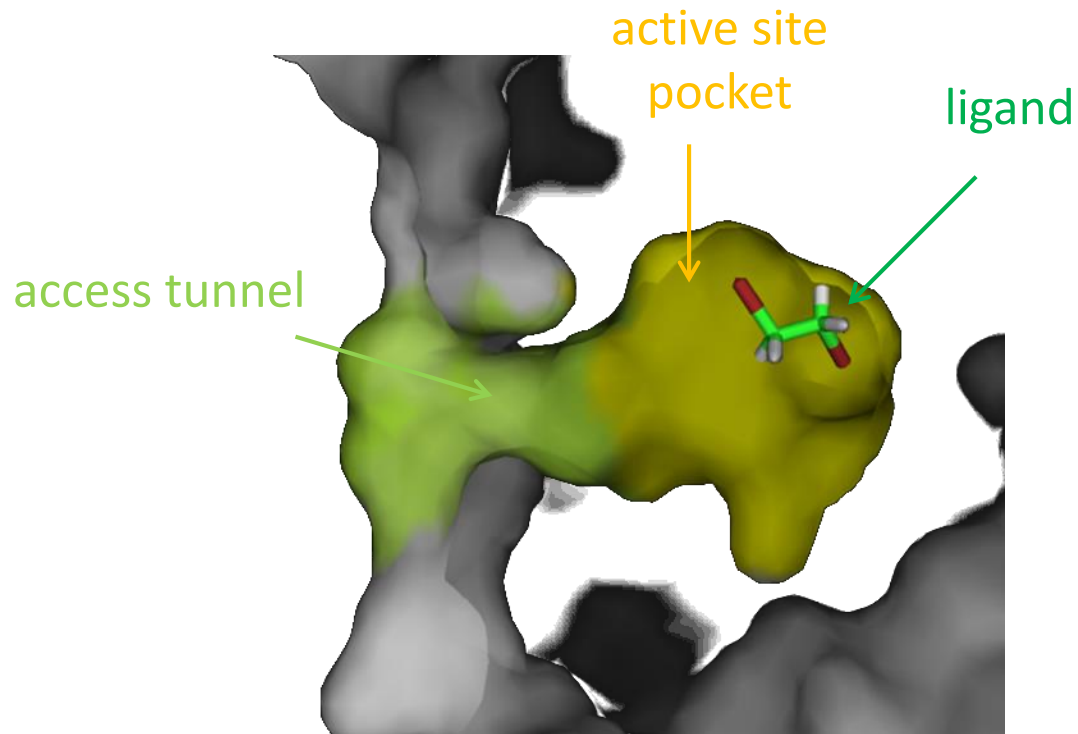
InputType CAVITY result ▼

Cavity_1 Cavity_2 Cavity_3
 Cavity_4 Cavity_5 Cavity_6
 Cavity_7

CavityPlus

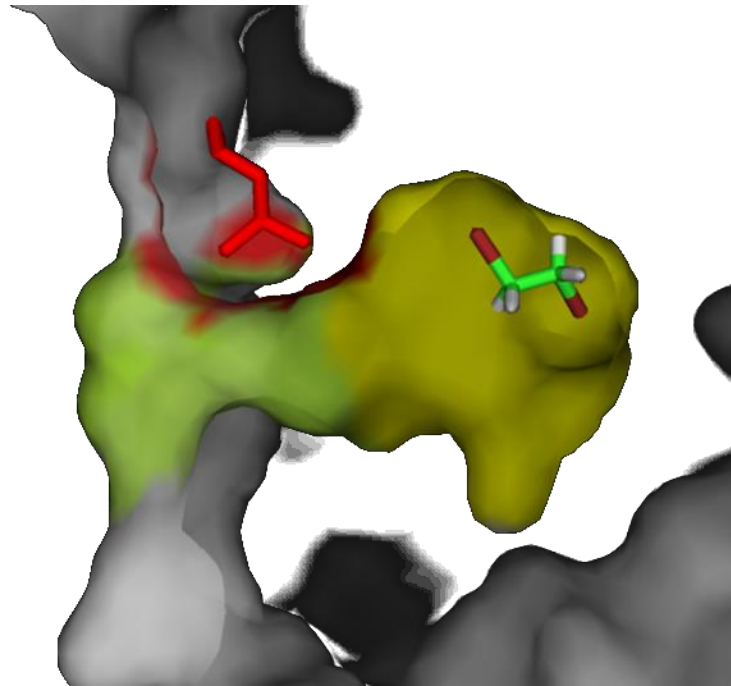
Analysis of access tunnels

- buried binding or active sites are connected with bulk solvent by access tunnels



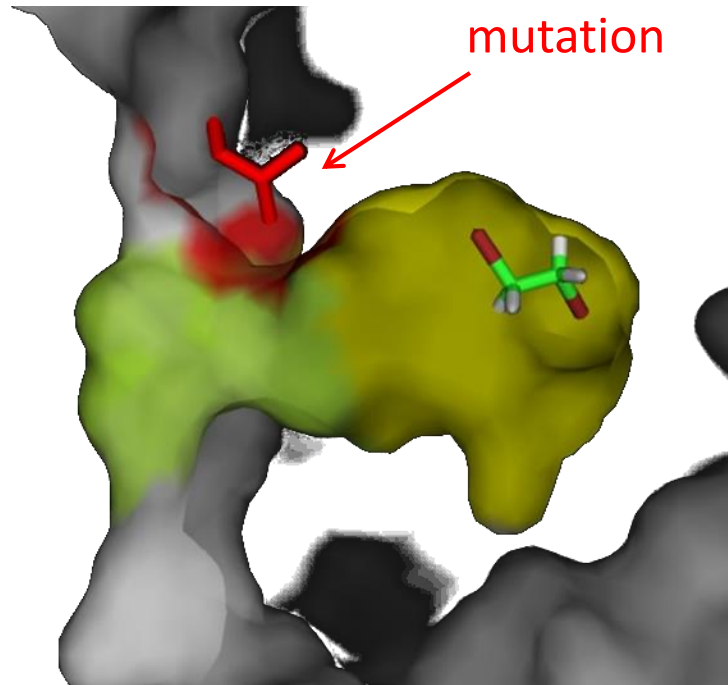
Analysis of access tunnels

- buried binding or active sites are connected with bulk solvent by access tunnels



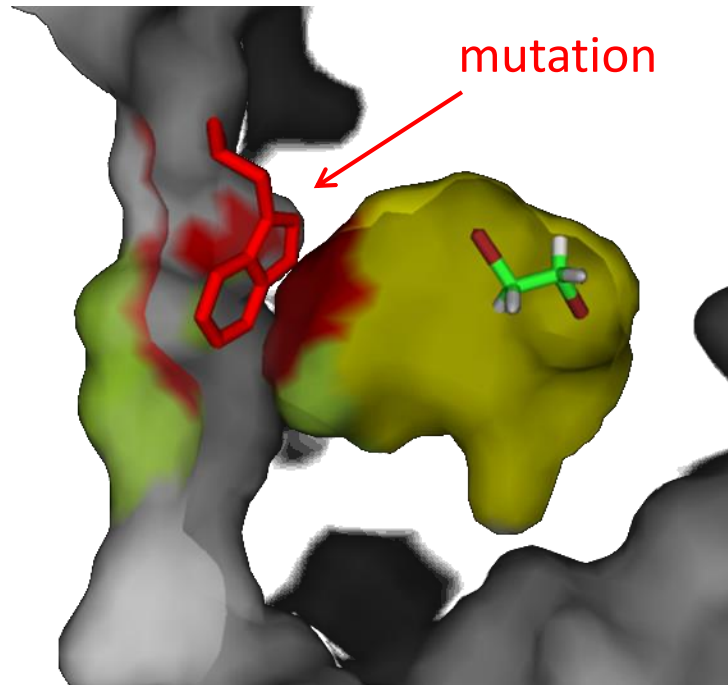
Analysis of access tunnels

- buried binding or active sites are connected with bulk solvent by access tunnels



Analysis of access tunnels

- buried binding or active sites are connected with bulk solvent by access tunnels



Analysis of access tunnels

- buried binding or active sites are connected with bulk solvent by **access tunnels**
 - adjusted to permit transport of specific molecules
 - mutations can speed-up or hinder transport of molecules as well as allow transport of other molecules
- requires 3D structure of protein
- software for detection of tunnels
 - Caver, Mole, HOLE, PoreWalker



Analysis of access tunnels

□ Caver Web

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

Single structure

| Tunnels info | | | | | |
|-------------------------------------|-----------------------|------------|-----------|------------|------|
| id | bottleneck radius [Å] | length [Å] | curvature | throughput | |
| <input checked="" type="checkbox"/> | 1 | 1.9 | 10.3 | 1.4 | 0.80 |
| <input checked="" type="checkbox"/> | 2 | 1.8 | 11.2 | 1.2 | 0.78 |
| <input checked="" type="checkbox"/> | 3 | 1.8 | 23.8 | 1.3 | 0.66 |
| <input checked="" type="checkbox"/> | 4 | 1.2 | 16.7 | 1.2 | 0.63 |
| <input type="checkbox"/> | 5 | 1.8 | 27.4 | 1.3 | 0.62 |
| <input type="checkbox"/> | 6 | 1.1 | 19.0 | 1.4 | 0.45 |

Return to Results browser

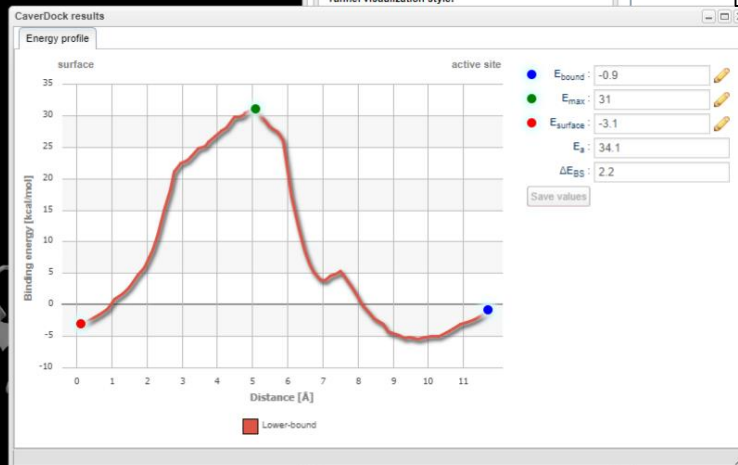
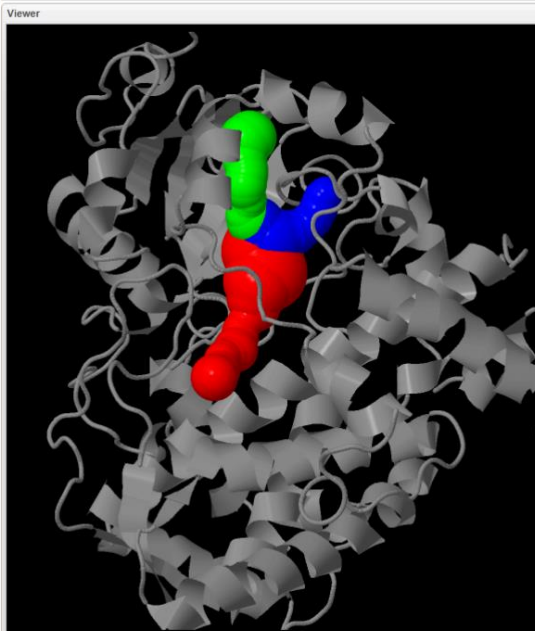
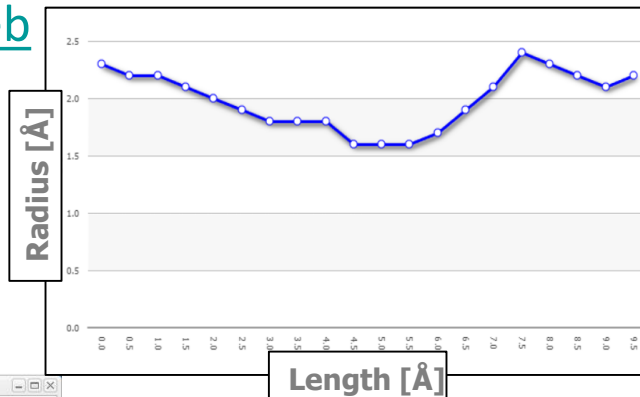
Job information

Job ID: rokjh
Title: Untitled
Structure: 4NY4

Download PyMOL session
Download results in single zip
View CAVER configuration
View CAVER log

Visualization settings

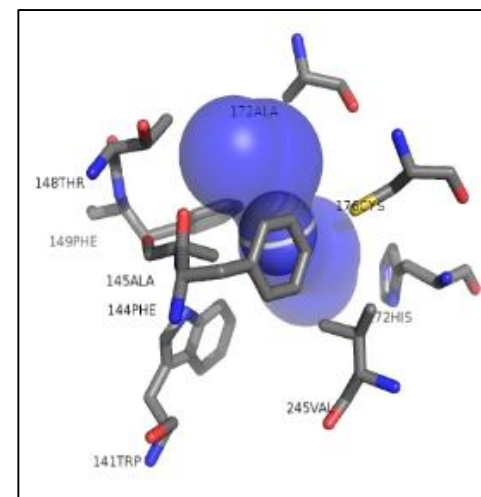
Tunnel visualization style:



Reset view

Visualization quality:

1 8



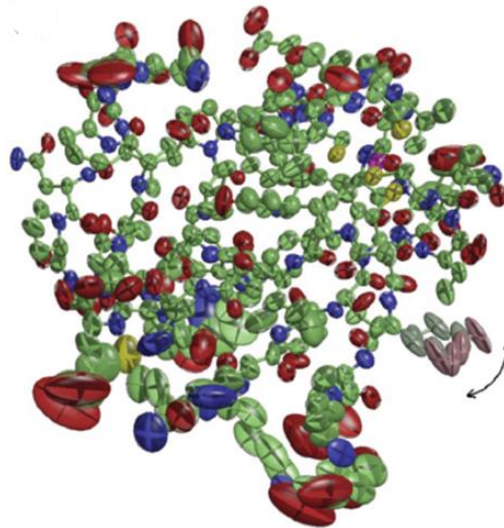
Hot-spots for engineering thermostability

- ❑ highly **flexible** residues – introduction of rigidifying mutations
 - ❑ residues located in **access tunnels**
- these residues may also include **catalytic** or other **essential** residues which generally should not be mutated!



Identification of highly flexible residues

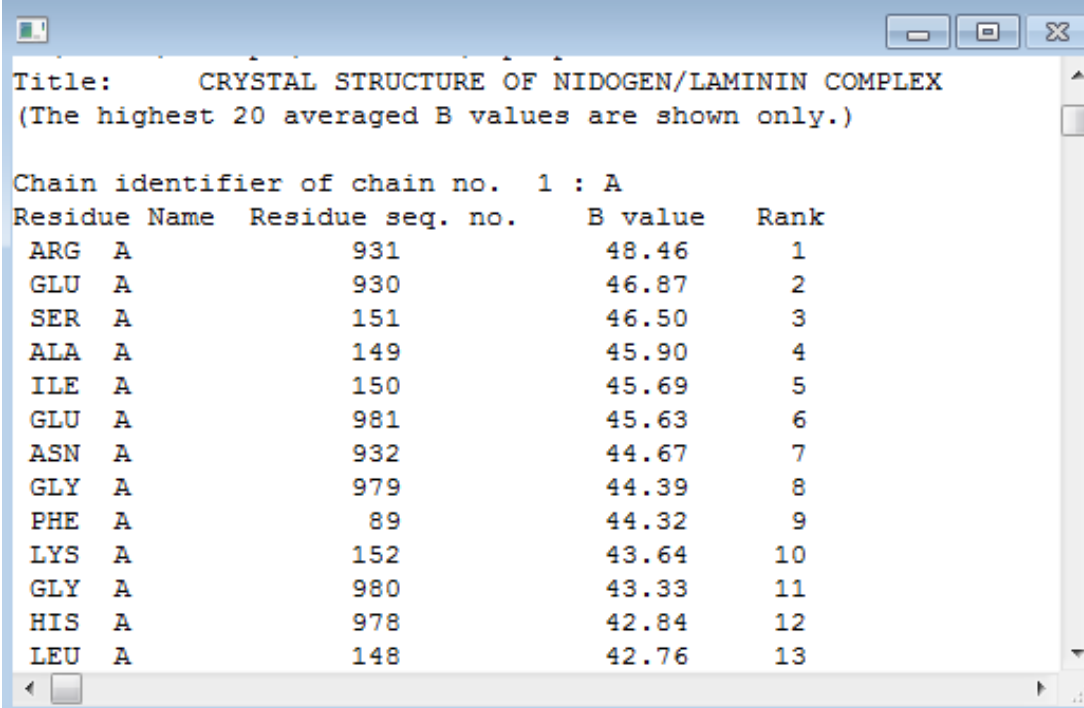
- prediction based on **crystallographic B-factors**
 - reflect the degree of thermal motion, and thus the flexibility of individual residues



- requires 3D structure of protein
 - experimental structure determined by X-ray crystallography (wwPDB)

Identification of highly flexible residues

- average B-factor of each residue in the target protein



Title: CRYSTAL STRUCTURE OF NIDOGEN/LAMININ COMPLEX
(The highest 20 averaged B values are shown only.)

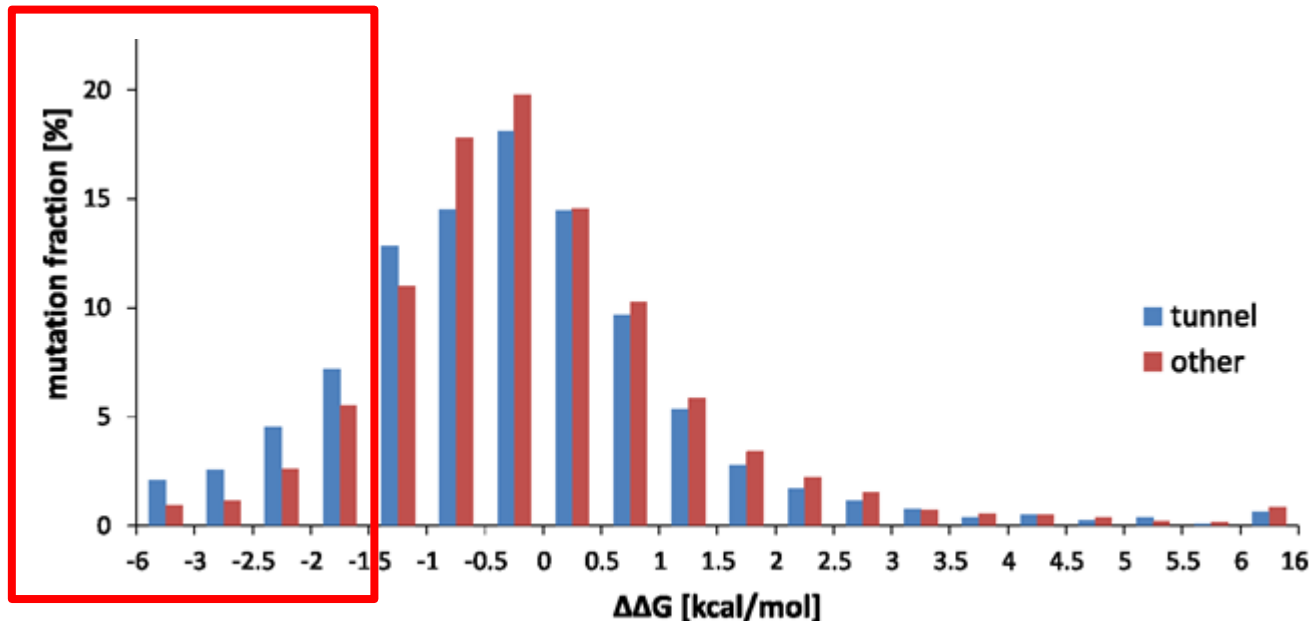
Chain identifier of chain no. 1 : A

| Residue Name | Residue seq. no. | B value | Rank |
|--------------|------------------|---------|------|
| ARG A | 931 | 48.46 | 1 |
| GLU A | 930 | 46.87 | 2 |
| SER A | 151 | 46.50 | 3 |
| ALA A | 149 | 45.90 | 4 |
| ILE A | 150 | 45.69 | 5 |
| GLU A | 981 | 45.63 | 6 |
| ASN A | 932 | 44.67 | 7 |
| GLY A | 979 | 44.39 | 8 |
| PHE A | 89 | 44.32 | 9 |
| LYS A | 152 | 43.64 | 10 |
| GLY A | 980 | 43.33 | 11 |
| HIS A | 978 | 42.84 | 12 |
| LEU A | 148 | 42.76 | 13 |

B-FITTER

Analysis of access tunnels

- saturation mutagenesis in **tunnel residues** has 2× better chance to significantly improve stability than mutagenesis in other protein regions (based on computational predictions)





- Protein engineering approaches
- **Semi-rational design**
 - identification of hot-spots
 - evaluation of hot-spots
 - selection of substitutions
 - design of library
 - mutagenesis and screening
- Rational design
 - molecular modeling

Evaluation of hot-spots

- ❑ hot-spots identified by computational tools can be further **evaluated** to prevent replacing indispensable amino acid residues and to **prioritize** the hot-spots (i.e., order the hot-spots based on their suitability for mutagenesis)
- ❑ analysis of evolutionary conservation
- ❑ prediction of effects of mutations on protein stability or function



Analysis of evolutionary conservation

- residues essential for maintaining structural or functional properties of a protein tend to be conserved during evolution
 - **conserved residues** are generally **not** recommended as **suitable** targets for mutagenesis - their replacement often leads to the loss of protein function
 - mutagenesis targeting **highly mutable** positions provides a significantly higher proportion of **viable variants** than random mutagenesis
 - targeting **moderately or highly variable positions**, which are expected to be tolerant to a wide range of substitutions, represents a good approach for producing **efficient smart libraries** (i.e., libraries with a high proportion of correctly folded and active variants)



Analysis of evolutionary conservation



- residue conservation can be derived from a **multiple alignment** of a set of related proteins (3D structure not required)

1 | **I T L V V H D W G G M I G M G Y A A R Y P E R I K**

Analysis of evolutionary conservation

- residue conservation can be derived from a **multiple alignment** of a set of related proteins (3D structure not required)

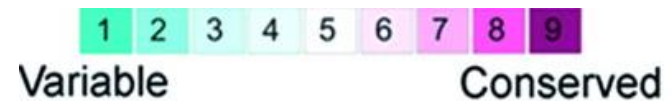
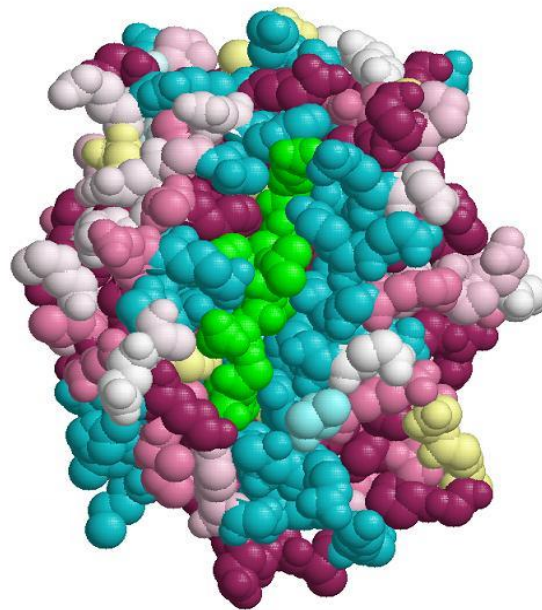
| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | I | T | L | V | V | H | D | W | G | G | M | I | G | M | G | Y | A | A | R | Y | P | E | R | I | K |
| 2 | I | T | L | V | V | H | D | W | G | G | M | I | G | M | A | Y | A | V | A | H | P | R | A | I | R |
| 3 | L | T | L | A | V | H | D | W | G | G | M | I | G | F | G | W | A | L | A | H | A | V | O | V | R |
| 4 | I | T | L | V | M | O | D | W | G | G | P | I | G | L | G | M | A | A | R | H | P | A | R | I | K |
| 5 | V | T | L | V | C | O | N | W | G | S | L | L | G | L | R | L | A | A | E | H | H | R | R | F | S |
| 6 | I | T | L | F | C | O | D | W | G | G | L | I | G | L | R | L | V | A | E | N | P | D | R | F | A |
| 7 | V | T | L | V | L | O | D | Y | G | A | A | F | G | L | N | W | A | S | R | N | P | D | R | V | R |



Analysis of evolutionary conservation



- evolutionary conservation of individual positions in protein mapped on protein 3D structure



ConSurf

Prediction of mutation effects

- computational tools for the prediction of **effect** of amino acid substitutions on protein **stability** or protein **function**
 - *in silico* site-saturation **mutagenesis** of identified hot-spots – check if mutations at a given site are likely to be tolerated
 - many highly **destabilizing/deleterious** mutations predicted for a certain position – given site is **not** a very **good** target for mutagenesis
 - sites with only a **few** highly **destabilizing /deleterious** mutations predicted can still represent promising hot-spots (the amino acids with potentially destabilizing/deleterious effects can be discarded from the library by the appropriate selection of degenerate codons)



Prediction of mutation effects

- effects on protein stability – usually requires 3D structure of protein
 - experimental structure (wwPDB)
 - theoretical model (AlphaFold, homology modeling)
- effects on protein function – sequence information often sufficient



Prediction of mutation effects



- prediction of effect of substitutions on protein stability
 - Evaluation of the change of protein free energy upon mutation
 - Evaluation of contributions of individual interactions to total energy
 - Usually requires structural information
- software for prediction of effect of mutation on stability
 - Rosetta, FoldX, CUPSAT, ERIS

Prediction of mutation effects

- prediction of effect of substitutions on protein stability

Singlepoint stability results

Stabilizing mutations Destabilizing mutations
Energy is in kcal/mol

| chain | position | residue | Ala | Arg | Asn | Asp | Cys | Gln | Glu | Gly | His | Ile | Lys |
|-------|----------|---------|------|------|------|-----|-----|------|------|-----|------|------|------|
| A | 169 | Phe | 3.5 | 7.1 | 5.4 | 6.5 | 6.0 | 4.2 | 5.1 | 5.9 | 5.7 | 45.7 | 5.8 |
| A | 135 | Ala | - | 35.9 | 4.1 | 4.5 | 2.4 | 17.2 | 15.5 | 2.1 | 25.7 | 12.5 | 17.9 |
| A | 11 | Lys | -0.2 | 2.3 | 0.1 | 1.7 | 2.0 | 0.2 | 1.3 | 1.3 | -1.9 | -1.1 | - |
| A | 264 | Thr | -0.1 | 0.9 | 0.6 | 1.8 | 1.9 | 0.0 | 0.7 | 1.3 | -0.7 | -0.3 | 0.8 |
| A | 126 | Gln | 1.3 | 1.3 | 3.1 | 3.4 | 5.5 | - | -0.5 | 3.3 | 1.4 | 28.7 | 1.6 |
| A | 259 | Thr | -0.2 | -0.3 | -0.5 | 0.3 | 1.7 | -1.1 | -0.1 | 0.6 | -0.4 | 4.9 | -1.0 |

Export table to CSV

Evaluate multiple point stability

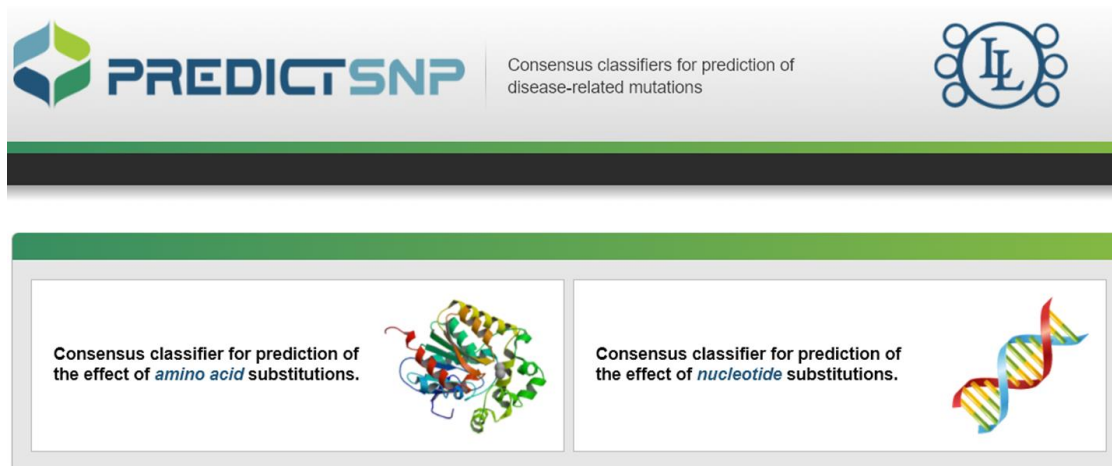
Codon usage : Escherichia coli K12

Generate report

Rosetta in HotSpot Wizard

Prediction of mutation effects

- prediction of effect of substitutions on protein function
 - Evaluation if a mutation would impair protein function
 - Hard to describe by physico-chemical properties > machine learning
 - Usually sequence based calculation
- software for prediction of effect of mutation on function
 - PredictSNP, SIFT, MAPP, PhD-SNP...



The image shows a screenshot of the PredictSNP web interface. At the top, the PredictSNP logo is displayed on the left, with the text "Consensus classifiers for prediction of disease-related mutations" in the center, and a circular logo with the letters "IL" on the right. Below this, there are two main prediction options presented in a grid-like layout:

- Left Option:** "Consensus classifier for prediction of the effect of *amino acid* substitutions." This option is accompanied by a 3D ribbon diagram of a protein structure.
- Right Option:** "Consensus classifier for prediction of the effect of *nucleotide* substitutions." This option is accompanied by a 3D model of a DNA double helix.

Prediction of mutation effects

- prediction of effect of substitutions on protein function

INPUT Load example

Insert protein sequence in FASTA format:

```
>HEA_HUMAN
MVLSPADKTNVKAAGKVGARHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
KKVADALTNVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAETFP
AVHASLDKFLASVSTLTSEYR
```

MUTATIONS Manual input

Select positions:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | M | V | L | S | P | A | D | K | T | N | V | K | A | A | W | G | K | V | G | A | H | A | G | E | Y | G | A | E | A | L | E | R | M | F | L | S | F | P | T | T |
| 41 | K | T | Y | F | P | H | F | D | L | S | H | G | S | A | Q | V | K | G | H | G | K | K | V | A | D | A | L | T | N | A | V | A | H | V | D | D | M | P | N | A |
| 81 | L | S | A | L | S | D | L | H | A | H | K | L | R | V | D | P | V | N | F | K | L | L | S | H | C | L | L | V | T | L | A | A | H | L | P | A | E | F | T | P |
| 121 | A | V | H | A | S | L | D | K | F | L | A | S | V | S | T | V | L | T | S | K | Y | R | | | | | | | | | | | | | | | | | | |

| Pos | Wild-type | Mutations | Clear |
|-----|-----------|------------------|----------------------------------|
| 59 | H | Y - Tyr | <input type="button" value="X"/> |
| 60 | G | D - Asp, V - Val | <input type="button" value="X"/> |
| 63 | V | T - Thr | <input type="button" value="X"/> |
| 68 | T | V - Val | <input type="button" value="X"/> |
| 72 | A | E - Glu, V - Val | <input type="button" value="X"/> |

TOOLS FOR EVALUATION

| Tool name | Time demands | Expected accuracy |
|--|--------------|-------------------|
| <input checked="" type="checkbox"/> PredictSNP | 32 min | 73.4% |
| <input checked="" type="checkbox"/> MAPP | 10 min | 70.7% |
| <input checked="" type="checkbox"/> PhD-SNP | 32 min | 71.5% |
| <input checked="" type="checkbox"/> PolyPhen-1 | 15 min | 68.1% |
| <input checked="" type="checkbox"/> PolyPhen-2 | 15 min | 69.2% |
| <input checked="" type="checkbox"/> SIFT | 15 min | 70.3% |
| <input checked="" type="checkbox"/> SNAP | 30 min | 67.6% |

JOB INFORMATION

ID: vb17zolzbrteajgmdyd5fv13u7cg4nzsqr2z6logfjbgz3pht

RESULTS neutral deleterious XX % expected accuracy Expand all annotations

| Annotation | Mutation | PredictSNP | MAPP | PhD-SNP | PolyPhen-1 | PolyPhen-2 | SIFT | SNAP |
|------------|----------|------------|------|---------|------------|------------|------|------|
| ▶ | H59Y | 87 % | 63 % | 82 % | 74 % | 65 % | 79 % | 85 % |
| ▶ | G60D | 87 % | 88 % | 68 % | 59 % | 55 % | 79 % | 85 % |
| ▶ | G60V | 87 % | 91 % | 82 % | 74 % | 59 % | 53 % | 72 % |
| ▶ | V63T | 87 % | 76 % | 61 % | 74 % | 63 % | 79 % | 62 % |
| ▶ | T68V | 71 % | 41 % | 72 % | 67 % | 75 % | 46 % | 67 % |
| ▶ | A72E | 74 % | 70 % | 58 % | 67 % | 87 % | 65 % | 77 % |
| ▶ | A72V | 60 % | 59 % | 73 % | 67 % | 76 % | 53 % | 71 % |
| ▶ | N79H | 74 % | 72 % | 55 % | 67 % | 87 % | 46 % | 67 % |
| ▶ | V97W | 61 % | 46 % | 45 % | 74 % | 81 % | 79 % | 58 % |
| ▶ | L110R | 87 % | 88 % | 88 % | 74 % | 81 % | 79 % | 62 % |
| ▶ | A112T | 83 % | 75 % | 58 % | 67 % | 74 % | 67 % | 83 % |
| ▶ | P115S | 65 % | 72 % | 59 % | 67 % | 75 % | 46 % | 77 % |
| ▶ | F117A | 68 % | 46 % | 55 % | 67 % | 87 % | 43 % | 67 % |

DOWNLOAD

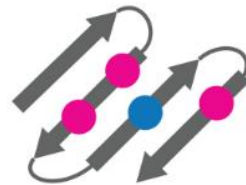
PredictSNP

Prediction of mutation effects

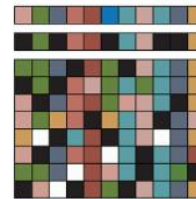
□ AlphaMissense

- <https://github.com/google-deepmind/alphamissense>
- deep learning predictor based on AlphaFold
- analysis of human and primate missense mutations
- trained on population frequency data and uses sequence and predicted structural context
- all single–amino acid substitutions in the human proteome are provided

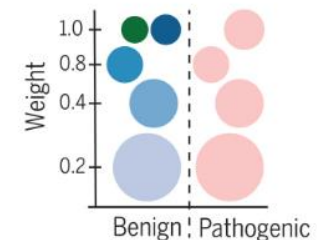
① Structure context



② Protein language modeling



③ Training variants





- Protein engineering approaches
- Semi-rational design
 - identification of hot-spots
 - evaluation of hot-spots
 - selection of substitutions
 - design of library
 - mutagenesis and screening
- Rational design
 - molecular modeling

Selection of substitutions

- substitutions introduced using **degenerate codons**
 - e.g., NNK (N = A/T/G/C; K = T/G)

IUPAC Nucleotide Nomenclature Table

| symbol | base | symbol | base |
|--------|------------------|--------|---------------|
| A | adenosine | M | A C (amino) |
| C | cytidine | S | G C (strong) |
| G | guanine | W | A T (weak) |
| T | thymidine | B | G T C |
| U | uridine | D | G A T |
| R | G A (purine) | H | A C T |
| Y | T C (pyrimidine) | V | G C A |
| K | G T (keto) | N | A G C T (any) |

Selection of substitutions

- all possible substitutions - **NNK** or NNS degenerate codons
 - 😊 encode all 20 amino acids with the lowest redundancy and price (mixture of 32 codons)
 - ☹️ redundancy is not completely eliminated (3× Arg, Leu, Ser, 2× Ala, Gly, Pro, Thr and Val)



Selection of substitutions

- all possible substitutions - NNK or NNS degenerate codons
- introduction of only selected substitutions using degenerate codons encoding **reduced amino acid alphabets**
 - ☹ do not encode all 20 amino acids
 - ☺ decreased library size → improved screening efficiency
 - **NDT** – balanced set of 12 amino acids (12 codons)



Selection of substitutions

- all possible substitutions - NNK or NNS degenerate codons
- introduction of only selected substitutions using degenerate codons encoding **reduced amino acid alphabets**

Table 1. Oversampling necessary for 95% coverage as a function of NNK and NDT codon degeneracy.

| No. ^[a] | NNK | | NDT | |
|--------------------|------------------------|------------------------|------------------------|------------------------|
| | Codons | Transformants needed | Codons | Transformants needed |
| 1 | 32 | 94 | 12 | 34 |
| 2 | 1 028 | 3 066 | 144 | 430 |
| 3 | 32 768 | 98 163 | 1 728 | 5 175 |
| 4 | 1 048 576 | 3 141 251 | 20 736 | 62 118 |
| 5 | 33 554 432 | 100 520 093 | 248 832 | 745 433 |
| 6 | $> 1.0 \times 10^9$ | $> 3.2 \times 10^9$ | $> 2.9 \times 10^8$ | $> 8.9 \times 10^6$ |
| 7 | $> 3.4 \times 10^{10}$ | $> 1.0 \times 10^{11}$ | $> 3.5 \times 10^9$ | $> 1.1 \times 10^8$ |
| 8 | $> 1.0 \times 10^{12}$ | $> 3.3 \times 10^{12}$ | $> 4.2 \times 10^9$ | $> 1.3 \times 10^9$ |
| 9 | $> 3.5 \times 10^{13}$ | $> 1.0 \times 10^{14}$ | $> 5.1 \times 10^9$ | $> 1.5 \times 10^{10}$ |
| 10 | $> 1.1 \times 10^{15}$ | $> 3.4 \times 10^{15}$ | $> 6.1 \times 10^{10}$ | $> 1.9 \times 10^{11}$ |

[a] Number of aa positions at one site.

Selection of reduced amino acid alphabets

- introduction of amino acids exhibiting **certain properties**
 - VRK – 8 hydrophilic amino acids (12 codons)
 - NYC – 8 hydrophobic amino acids (8 codons)
 - KST – 4 small amino acids (4 codons)
 - ...



Selection of reduced amino acid alphabets

- introduction of amino acids exhibiting certain properties
- introduction of a **balanced set** of amino acids
 - NDT – balanced set of 12 amino acids (12 codons)



Selection of reduced amino acid alphabets

- introduction of amino acids exhibiting certain properties
- introduction of a balanced set of amino acids
- introduction of substitutions existing (at a given site) in known **natural proteins**
 - likely increasing the proportion of viable variants in the resulting library
 - can be obtained by analysis of multiple sequence alignment



Selection of reduced amino acid alphabets

- ❑ introduction of amino acids exhibiting certain properties
- ❑ introduction of a balanced set of amino acids
- ❑ introduction of substitutions existing (at a given site) in known natural proteins
- ❑ discarding amino acids with potentially **destabilizing/ deleterious effects**
 - can be obtained by prediction of effects of mutations on protein stability or function



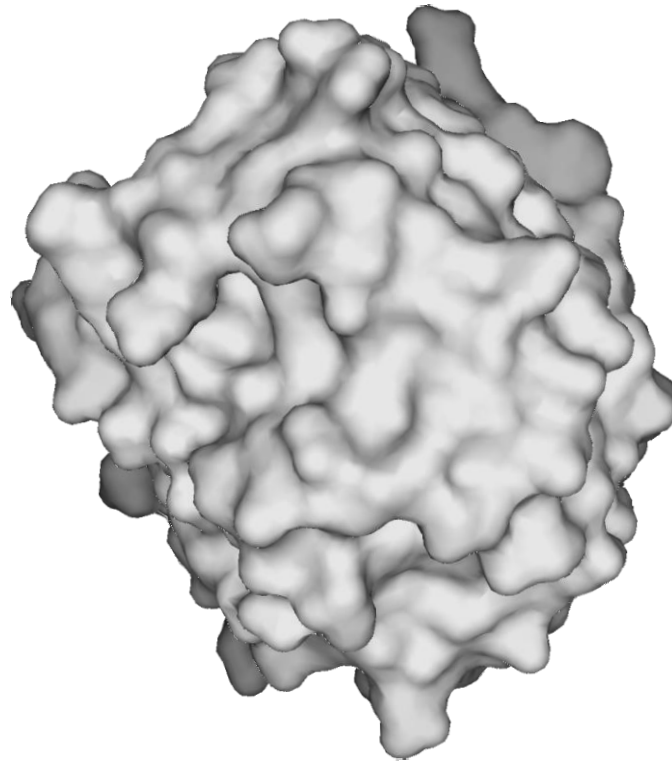
HotSpot Wizard



- ❑ meta-server combining several tools
 - automatic **identification of hot-spots** for engineering of enzyme catalytic properties
 - prioritization of hot-spots by their mutability
 - distribution of amino acids at individual positions
 - prediction of stability
 - molecular docking
 - design of smart libraries

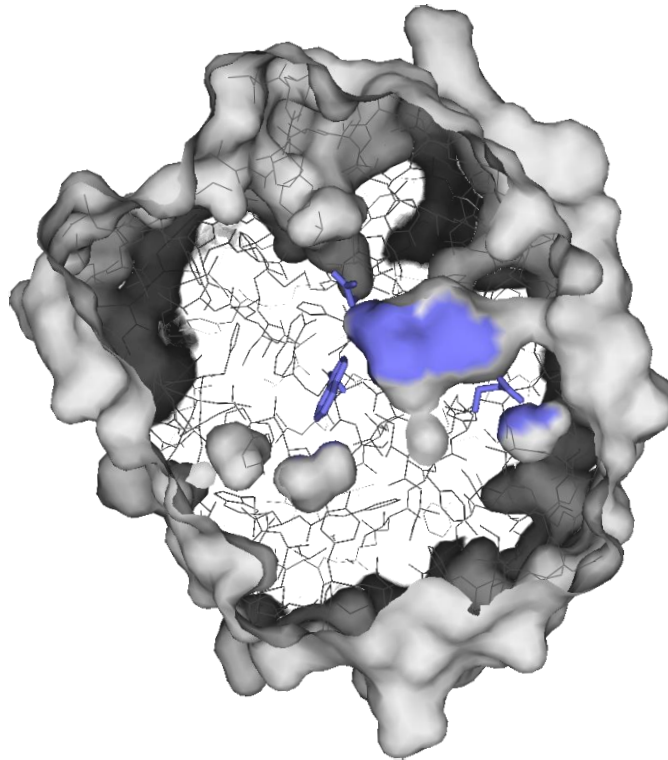
HotSpot Wizard

1. protein structure



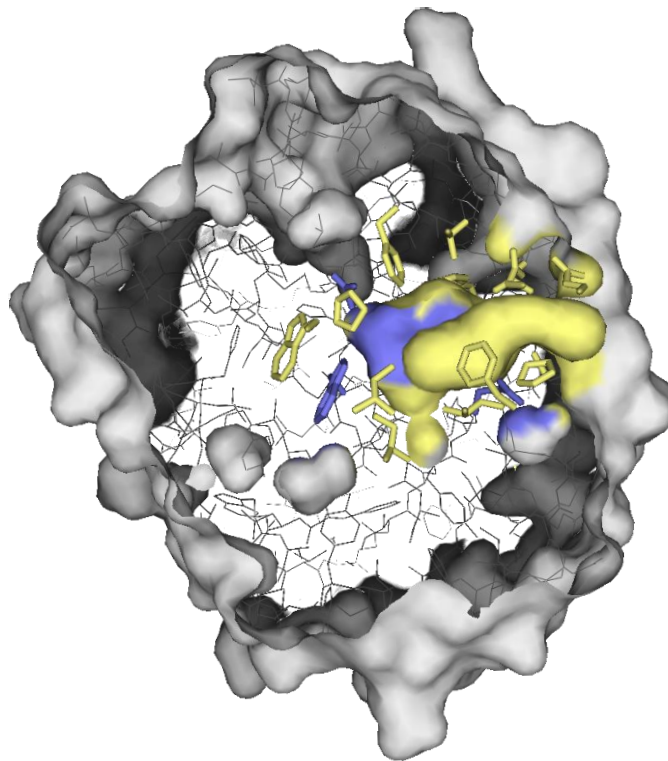
HotSpot Wizard

2. residues indispensable for protein function: **catalytic and binding** residues



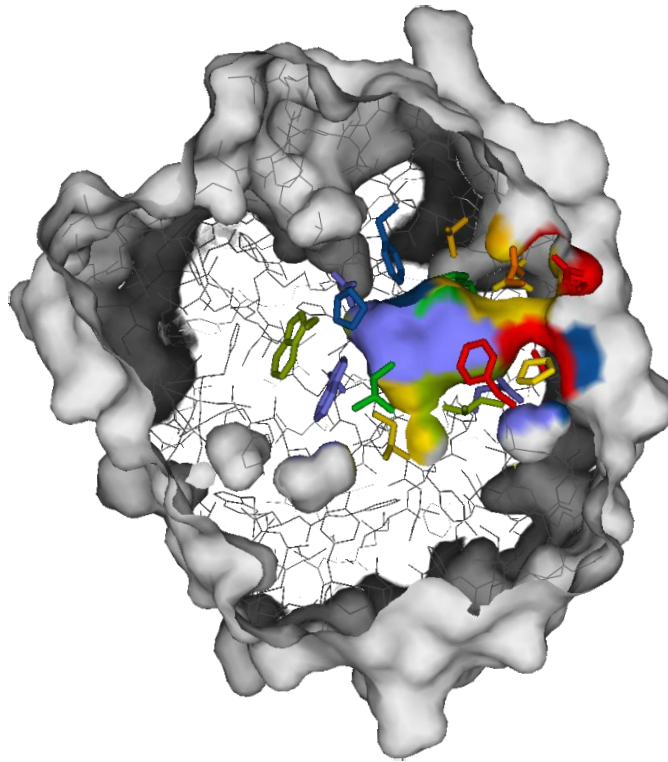
HotSpot Wizard

3. functional residues: active site **pocket and tunnels**



HotSpot Wizard

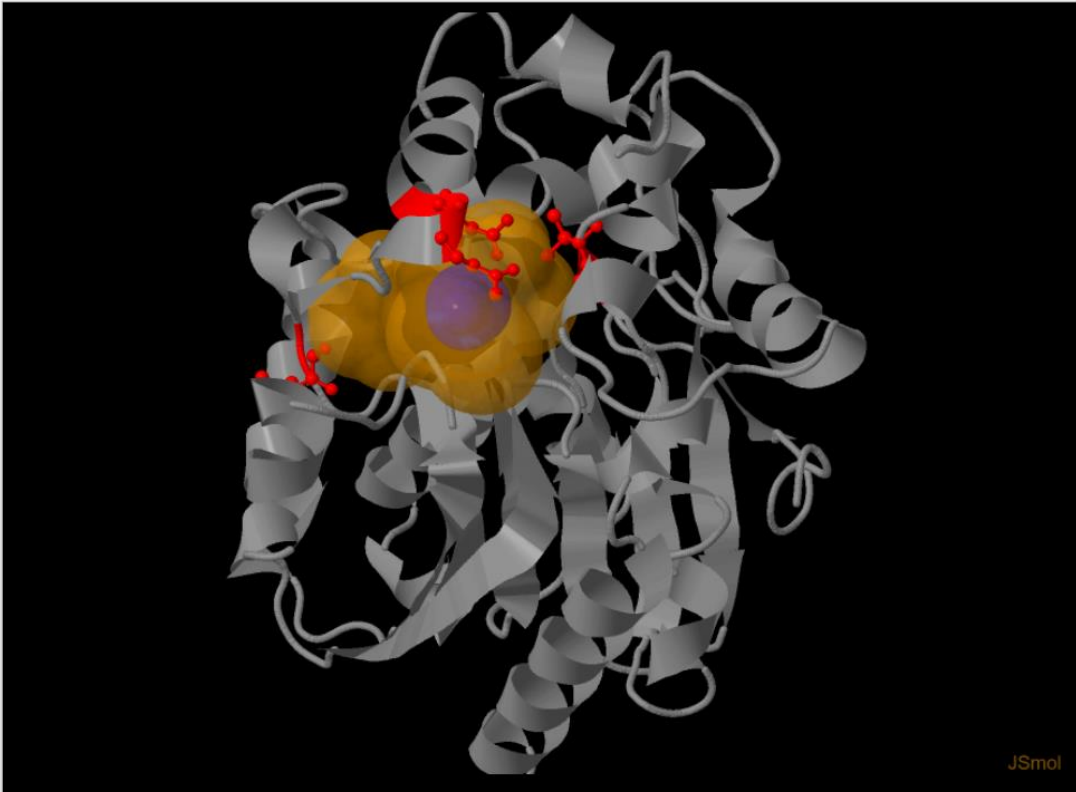
4. **mutability** of individual positions of protein



HotSpot Wizard

Functional hot spots of 1CV2

Viewer



JSmol

Return to Results browser

Visualization settings

Structure visualization style:

| | |
|----------------|----------|
| Wireframe | Cartoon |
| Sticks | Trace |
| Balls & sticks | Backbone |
| Balls | |

Hide all visualized residues

Save image

Reset view

Visualization quality:

1 8

Tunnels

| id | length (Å) | bottleneck radius (Å) |
|-------------------------|------------|-----------------------|
| Starting from pocket: 1 | | |

Pockets

| id | chain... | relevance (...) | volume (Å ³) | |
|-------------------------------------|----------|-----------------|--------------------------|-----|
| <input checked="" type="checkbox"/> | 1 | A | 100 | 514 |
| <input type="checkbox"/> | 2 | A | 82 | 907 |
| <input type="checkbox"/> | 3 | A | 62 | 245 |

Molecular docking

Run Docking

| | job id | ligand | pose |
|--------------------------|--------|--------|-----------------|
| <input type="checkbox"/> | ✓ | ztapvt | UDP-galactose 1 |
| <input type="checkbox"/> | ✓ | pxlsok | yperite 1 |
| <input type="checkbox"/> | ✓ | dymxo8 | ligand5 1 |
| <input type="checkbox"/> | ✓ | vquhab | yperite 5 |
| <input type="checkbox"/> | ✓ | nkwl1t | yperite 1 |
| <input type="checkbox"/> | ✓ | x5yh9u | ligand10 3 |

Residues selected for mutagenesis

Zoom residues Reset view

Stability evaluation Design library

Evaluated stabilities Designed libraries

| chain | position | residue | HotSpot |
|-------|----------|---------|---------|
| A | 136 | Met | ✓ |
| A | 146 | Gln | ✓ |
| A | 147 | Asp | ✓ |
| A | 173 | Val | ✓ |
| A | 249 | Thr | ✓ |
| A | 253 | Met | ✓ |
| A | 271 | Ala | ✓ |

Residue features

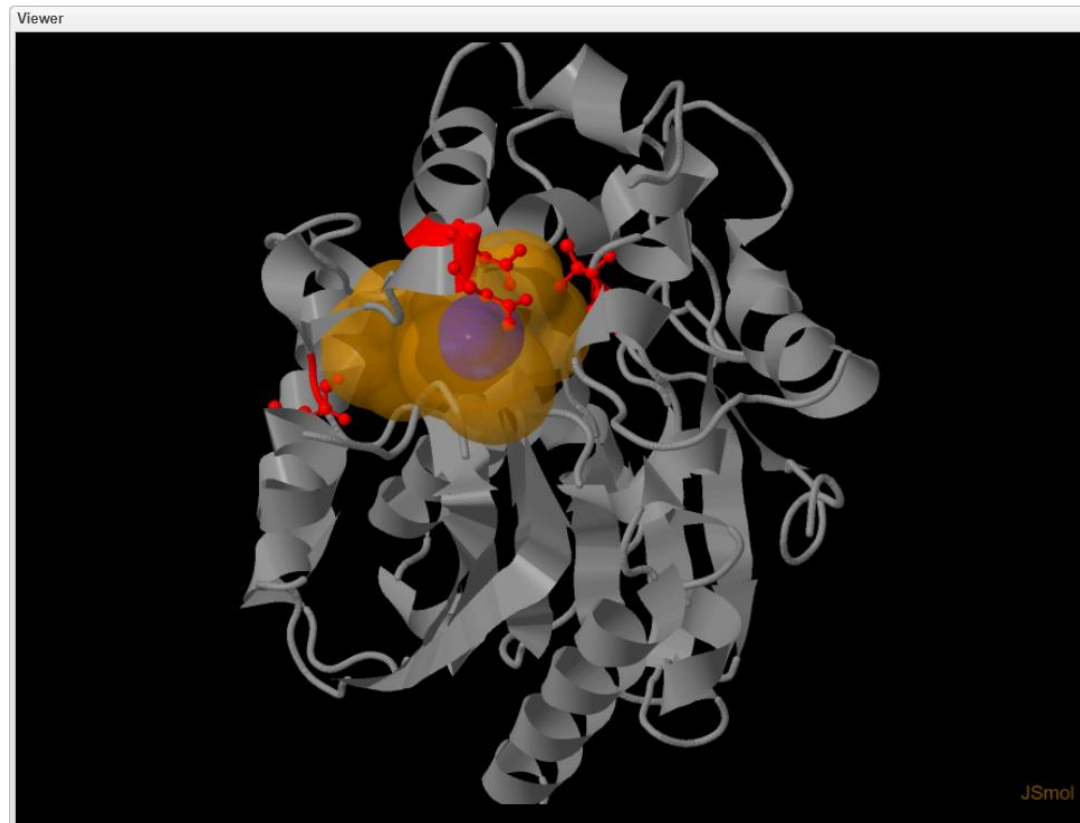
Exclude correlated positions
 Exclude catalytic pockets
 Exclude tunnels
 Exclude α -helices and β -sheets
 Show all residues

Exclude buried residues
 Include residues with moderate mutability

| chain | position | residue | mutable | non-essential | in tunnel | in catalytic pocket | HotSpot |
|-------------------------------------|----------|---------|---------|---------------|-----------|---------------------|---------|
| Chain A | | | | | | | |
| <input checked="" type="checkbox"/> | A | 136 | Met | ✓ | ✓ | X | ✓ |
| <input checked="" type="checkbox"/> | A | 146 | Gln | ✓ | ✓ | ✓ | ✓ |
| <input checked="" type="checkbox"/> | A | 147 | Asp | ✓ | ✓ | ✓ | ✓ |
| <input checked="" type="checkbox"/> | A | 173 | Val | ✓ | ✓ | ✓ | ✓ |
| <input type="checkbox"/> | A | 249 | Thr | ✓ | ✓ | ✓ | X |
| <input type="checkbox"/> | A | 253 | Met | ✓ | ✓ | X | ✓ |
| <input type="checkbox"/> | A | 271 | Ala | ✓ | ✓ | ✓ | ✓ |

HotSpot Wizard

Functional hot spots of 1CV2



Residue features

- Exclude correlated positions Exclude catalytic pockets Exclude tunnels Exclude α -helices and β -sheets
- Exclude buried residues Include residues with moderate mutability

| | chain | position | residue | mutable | non-essential | in tunnel | in catalytic pocket | HotSpot |
|-------------------------------------|-------|----------|---------|---------|---------------|-----------|---------------------|---------|
| <input checked="" type="checkbox"/> | A | 136 | Met | ✓ | ✓ | ✗ | ✓ | ✓ |
| <input checked="" type="checkbox"/> | A | 146 | Gln | ✓ | ✓ | ✓ | ✓ | ✓ |
| <input checked="" type="checkbox"/> | A | 147 | Asp | ✓ | ✓ | ✓ | ✓ | ✓ |
| <input checked="" type="checkbox"/> | A | 173 | Val | ✓ | ✓ | ✓ | ✓ | ✓ |
| <input type="checkbox"/> | A | 249 | Thr | ✓ | ✓ | ✓ | ✗ | ✓ |
| <input type="checkbox"/> | A | 253 | Met | ✓ | ✓ | ✗ | ✓ | ✓ |

Hot spots

Visualization settings

Structure visualization style:

| | |
|----------------|----------|
| Wireframe | Cartoon |
| Sticks | Trace |
| Balls & sticks | Backbone |
| Balls | |

Visualization quality:

1 8

Tunnels

| id | length (Å) | bottleneck radius (Å) |
|-------------------------|------------|-----------------------|
| Starting from pocket: 1 | | |

Pockets

| id | chain... | relevance (...) | volume (Å ³) | |
|-------------------------------------|----------|-----------------|--------------------------|-----|
| <input checked="" type="checkbox"/> | 1 | A | 100 | 514 |
| <input type="checkbox"/> | 2 | A | 82 | 907 |
| <input type="checkbox"/> | 3 | A | 62 | 245 |

Molecular docking

| | job id | ligand | pose |
|--------------------------|--------|--------|-----------------|
| <input type="checkbox"/> | ✓ | ztapvt | UDP-galactose 1 |
| <input type="checkbox"/> | ✓ | pxlsok | yperite 1 |
| <input type="checkbox"/> | ✓ | dymxo8 | ligand5 1 |
| <input type="checkbox"/> | ✓ | vquhab | yperite 5 |
| <input type="checkbox"/> | ✓ | nkwl1t | yperite 1 |
| <input type="checkbox"/> | ✓ | x5yh9u | ligand10 3 |

Residues selected for mutagenesis

| chain | position | residue | HotSpot |
|-------|----------|---------|---------|
|-------|----------|---------|---------|

Tunnels

Cavities

Docking

Stability

Design Library



- Protein engineering approaches
- Semi-rational design
 - identification of hot-spots
 - evaluation of hot-spots
 - selection of substitutions
 - design of library
 - mutagenesis and screening
- Rational design
 - molecular modeling

Design of library

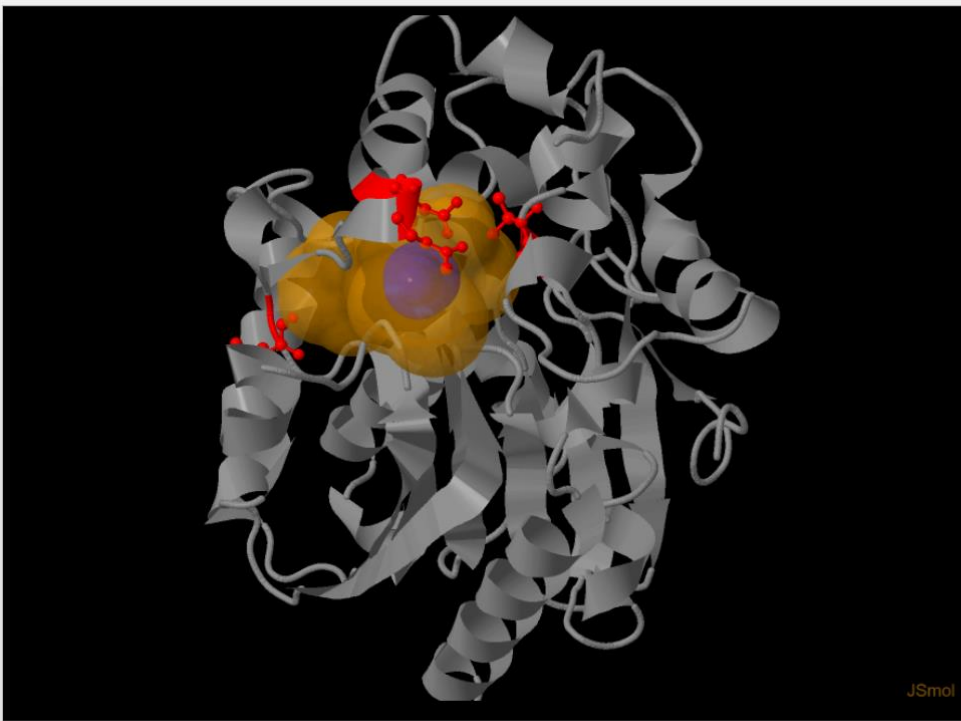
- decisions to be made after evaluation and prioritization of hot-spots:
 - how many and which positions to target?
 - should the positions be randomized simultaneously or separately?
 - should all or only a reduced set of amino acids be introduced at individual positions?
- dramatic effect on the **size** of the resulting **library**



Design of library – HotSpot Wizard

Functional hot spots of 1CV2

Viewer



JSmol

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 |
|---------|-------|----------|---------|---------|---------------|-----------|---------------------|---------|
| Chain A | A | 136 | Met | ✓ | ✓ | ✗ | ✓ | ✓ |
| | A | 146 | Gln | ✓ | ✓ | ✓ | ✓ | ✓ |
| | A | 147 | Asp | ✓ | ✓ | ✓ | ✓ | ✓ |
| | A | 173 | Val | ✓ | ✓ | ✓ | ✓ | ✓ |
| | A | 249 | Thr | ✓ | ✓ | ✓ | ✗ | ✓ |
| | A | 253 | Met | ✓ | ✓ | ✗ | ✓ | ✓ |

Return to Results browser

Visualization settings

Structure visualization style:

| | |
|----------------|----------|
| Wireframe | Cartoon |
| Sticks | Trace |
| Balls & sticks | Backbone |

Balls

Hide all visualized residues

Save image

Reset view

Visualization quality:

1 8

Tunnels

| id | length (Å) | bottleneck radius (Å) |
|-------------------------|------------|-----------------------|
| Starting from pocket: 1 | | |

Pockets

| | id | chain... | relevance (...) | volume (Å ³) |
|-------------------------------------|----|----------|-----------------|--------------------------|
| <input checked="" type="checkbox"/> | 1 | A | 100 | 514 |
| <input type="checkbox"/> | 2 | A | 82 | 907 |
| <input type="checkbox"/> | 3 | A | 62 | 245 |

Molecular docking

Run Docking

| | job id | ligand | pose |
|--------------------------|--------|--------|-----------------|
| <input type="checkbox"/> | ✓ | ztapvt | UDP-galactose 1 |
| <input type="checkbox"/> | ✓ | pxlsok | yperite 1 |
| <input type="checkbox"/> | ✓ | dymxo8 | ligand5 1 |
| <input type="checkbox"/> | ✓ | vquhab | yperite 5 |
| <input type="checkbox"/> | ✓ | nkw1yt | yperite 1 |
| <input type="checkbox"/> | ✓ | x5yh9u | ligand10 3 |

Residues selected for mutagenesis

Zoom residues Reset view

Stability evaluation Design library

Evaluated stabilities Designed libraries

| chain | position | residue | HotSpot |
|-------|----------|---------|---------|
|-------|----------|---------|---------|

Design of library – HotSpot Wizard

Library design

Standard SwiftLib

AA selection mode : Amino acid frequency Minimal frequency (%) : 5 Include wild-type Exclude wild-type

| chain | position | residue | desired amino acids | codon | desired ratio (%) | stop ratio (%) |
|-------|----------|---------|-----------------------------------|-------|-------------------|----------------|
| A | 136 | Met | Ala, Lys, Pro, Gln, Arg, Thr | VVR | 77.8 | 0.0 |
| A | 146 | Gln | Ala, Asp, Glu, Gly, Pro, Gln, Ser | BVV | 63.0 | 11.1 |
| A | 147 | Asp | Ala, Phe, Gly, Leu, Met, Thr, Val | DBS | 61.1 | 0.0 |

| codon | desired ratio (%) | stop ratio (%) | desired amino acids | encoded amino acids |
|-------|-------------------|----------------|---|---|
| DBS | 100.0 | 0.0 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:1 Met:1 Arg:1 Ser:3 Thr:2 Val:2 Trp:1 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:1 Met:1 Arg:1 Ser:3 Thr:2 Val:2 Trp:1 |
| DBK | 100.0 | 0.0 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:1 Met:1 Arg:1 Ser:3 Thr:2 Val:2 Trp:1 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:1 Met:1 Arg:1 Ser:3 Thr:2 Val:2 Trp:1 |
| DBB | 100.0 | 0.0 | Ala:3 Cys:2 Phe:2 Gly:3 Ile:2 Leu:1 Met:1 Arg:1 Ser:5 Thr:3 Val:3 Trp:1 | Ala:3 Cys:2 Phe:2 Gly:3 Ile:2 Leu:1 Met:1 Arg:1 Ser:5 Thr:3 Val:3 Trp:1 |
| DBN | 97.2 | 2.8 | Ala:4 Cys:2 Phe:2 Gly:4 Ile:3 Leu:2 Met:1 Arg:2 Ser:6 Thr:4 Val:4 Trp:1 | Ala:4 Cys:2 Phe:2 Gly:4 Ile:3 Leu:2 Met:1 Arg:2 Ser:6 Thr:4 Val:4 Trp:1 |
| DBV | 96.3 | 3.7 | Ala:3 Cys:1 Phe:1 Gly:3 Ile:2 Leu:2 Met:1 Arg:2 Ser:4 Thr:3 Val:3 Trp:1 | Ala:3 Cys:1 Phe:1 Gly:3 Ile:2 Leu:2 Met:1 Arg:2 Ser:4 Thr:3 Val:3 Trp:1 |
| DBD | 96.3 | 3.7 | Ala:3 Cys:1 Phe:1 Gly:3 Ile:2 Leu:2 Met:1 Arg:2 Ser:4 Thr:3 Val:3 Trp:1 | Ala:3 Cys:1 Phe:1 Gly:3 Ile:2 Leu:2 Met:1 Arg:2 Ser:4 Thr:3 Val:3 Trp:1 |
| NBS | 91.7 | 0.0 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:3 Met:1 Arg:3 Ser:3 Thr:2 Val:2 Trp:1 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:3 Met:1 Pro:2 Arg:3 Ser:3 Thr:2 Val:2 Trp:1 |
| NBK | 91.7 | 0.0 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:3 Met:1 Arg:3 Ser:3 Thr:2 Val:2 Trp:1 | Ala:2 Cys:1 Phe:1 Gly:2 Ile:1 Leu:3 Met:1 Pro:2 Arg:3 Ser:3 Thr:2 Val:2 Trp:1 |
| NBB | 91.7 | 0.0 | Ala:3 Cys:2 Phe:2 Gly:3 Ile:2 Leu:4 Met:1 Arg:4 Ser:5 Thr:3 Val:3 Trp:1 | Ala:3 Cys:2 Phe:2 Gly:3 Ile:2 Leu:4 Met:1 Pro:3 Arg:4 Ser:5 Thr:3 Val:3 Trp:1 |
| NBN | 89.6 | 2.1 | Ala:4 Cys:2 Phe:2 Gly:4 Ile:3 Leu:6 Met:1 Arg:6 Ser:6 Thr:4 Val:4 Trp:1 | Ala:4 Cys:2 Phe:2 Gly:4 Ile:3 Leu:6 Met:1 Pro:4 Arg:6 Ser:6 Thr:4 Val:4 Trp:1 |

Library size : 7315

Expected coverage : 0.95

Probability of full coverage : 0

Codon usage : Escherichia coli K12

Generate report



- Protein engineering approaches
- Semi-rational design
 - identification of hot-spots
 - evaluation of hot-spots
 - selection of substitutions
 - design of library
 - mutagenesis and screening
- Rational design
 - molecular modeling

Mutagenesis and screening



- saturation mutagenesis - next lecture 😊





- ❑ Protein engineering approaches
- ❑ Semi-rational design
 - identification of hot-spots
 - evaluation of hot-spots
 - selection of substitutions
 - design of library
 - mutagenesis and screening
- ❑ **Rational design**
 - molecular modeling → design of mutations

Rational design

- ❑ **site-specific** changes on the target enzyme
- ❑ few amino-acid substitutions that are predicted to elicit desired improvements of enzyme function
- ❑ based on **detailed knowledge** of protein structure, function and catalytic mechanism
- ❑ 😊 relatively simple characterization of constructed variants
- ❑ ☹️ complexity of protein structure-function relationships
- ❑ ☹️ molecular modeling expertise usually required





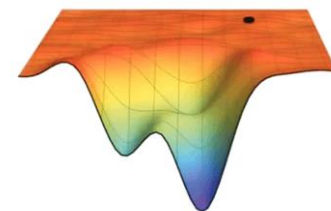
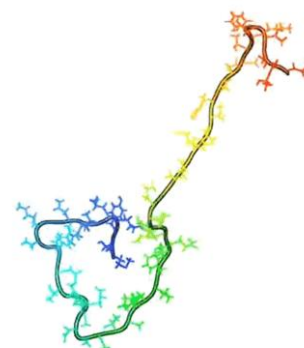
- ❑ Protein engineering approaches
- ❑ Semi-rational design
 - identification of hot-spots
 - evaluation of hot-spots
 - selection of substitutions
 - design of library
 - mutagenesis and screening
- ❑ **Rational design**
 - **molecular modeling → design of mutations**

Molecular modeling

- “Theoretical or computational technique that provides insight into the behavior of molecular system.”

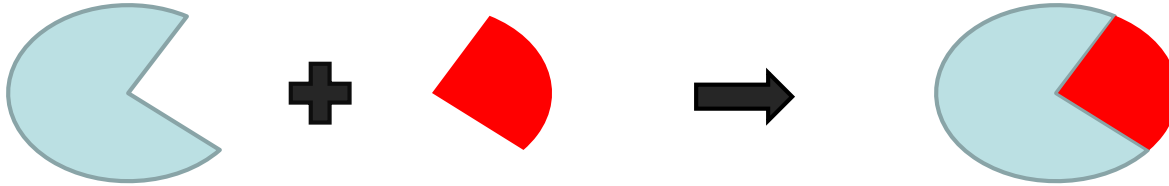
A. R. Leach

- **Ligand binding**
 - Molecular docking
- **Protein dynamics and transport of molecules**
 - Molecular dynamics
- **Reaction barriers and mechanisms**
 - Quantum chemistry or QM/MM
- **Protein design**
 - Molecular mechanics, machine learning



Molecular docking

- predicts structure of receptor (protein) – ligand complex



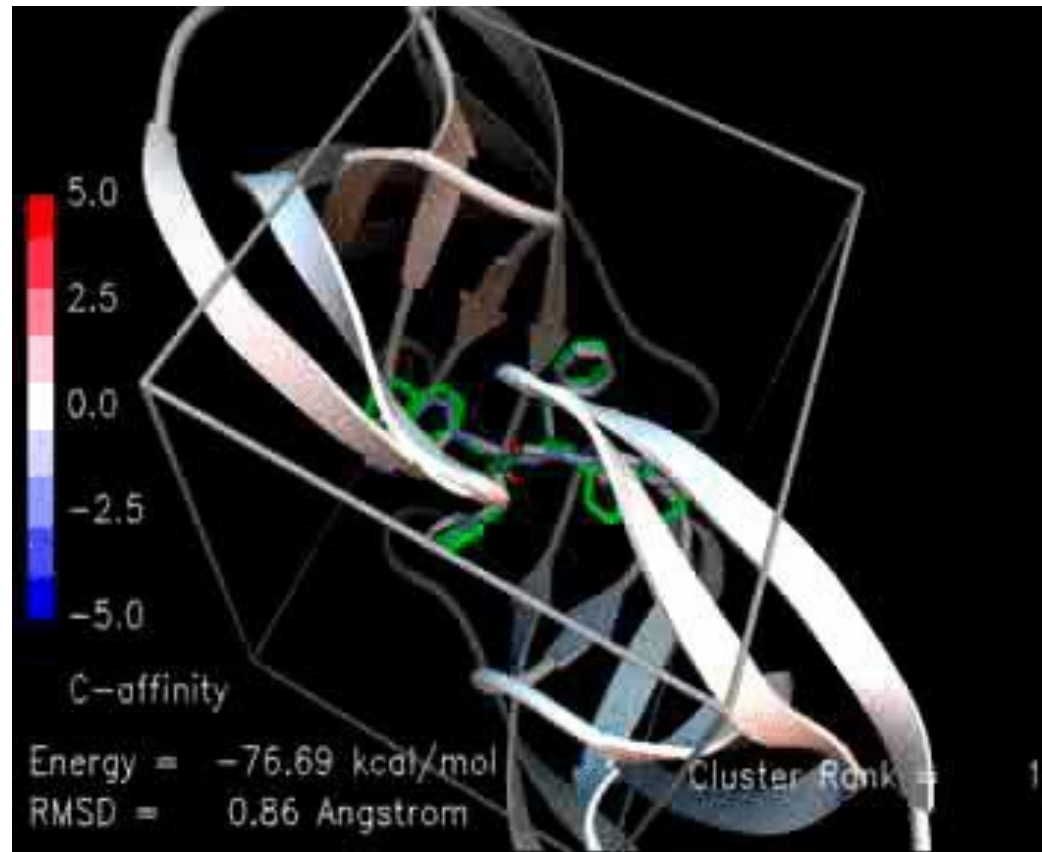
Molecular docking

□ Two components procedure

- searching – finding the conformation of ligand in the active site of the enzyme
- scoring – evaluation of the binding free energy

□ Docking software

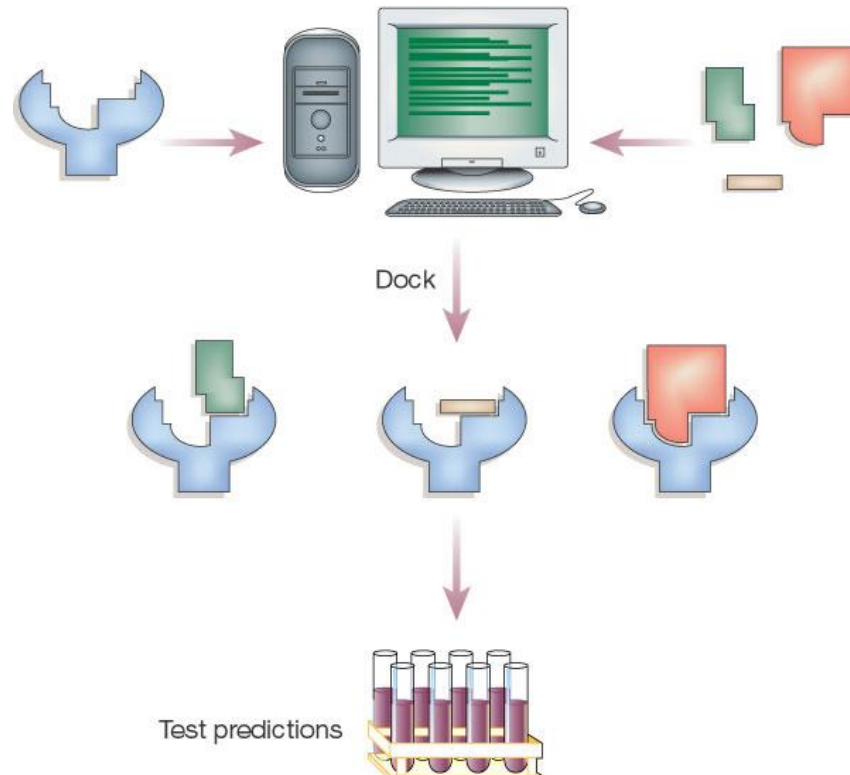
- Autodock, Vina, Gold, Medusa, Rosetta Dock...



Molecular docking

□ Virtual screening

- database of compounds + protein structure > molecular docking > re-scoring > compound prioritization > experimental testing



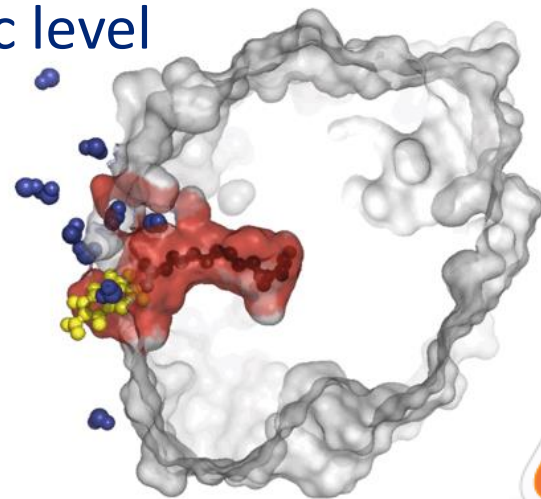
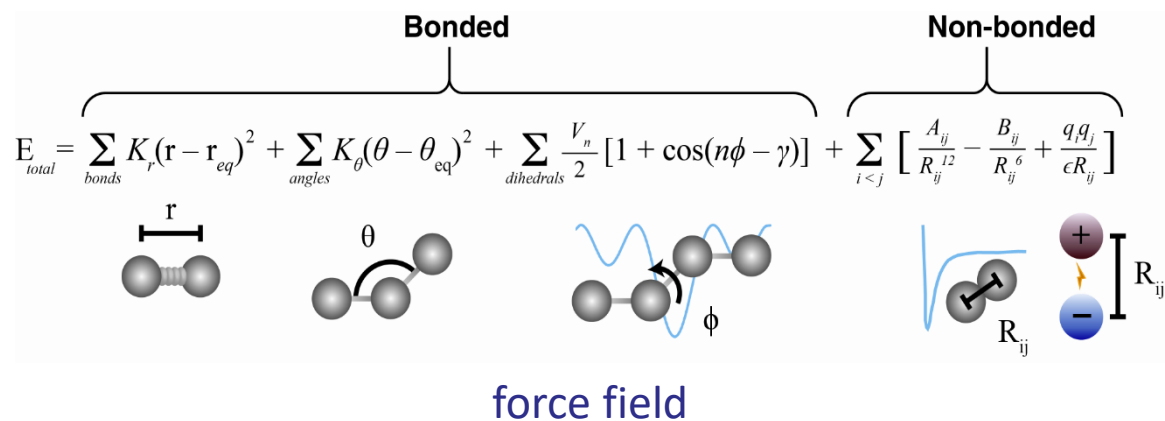
Molecular dynamics



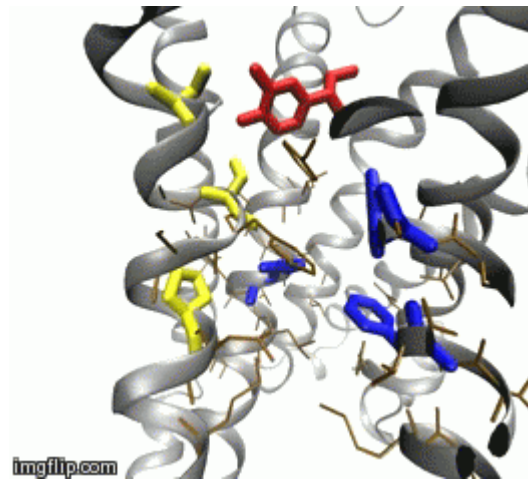
□ Principle

- physical description of interactions within the system (force field)
- Newton's laws of motions
- forces acting on all atoms due to all atoms

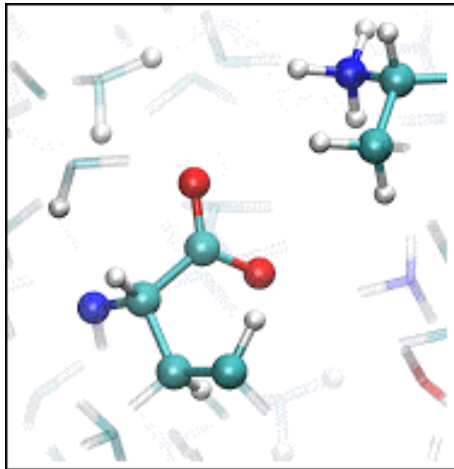
- Provides information on energetics, amplitudes, and time scales of local motions on the atomic level



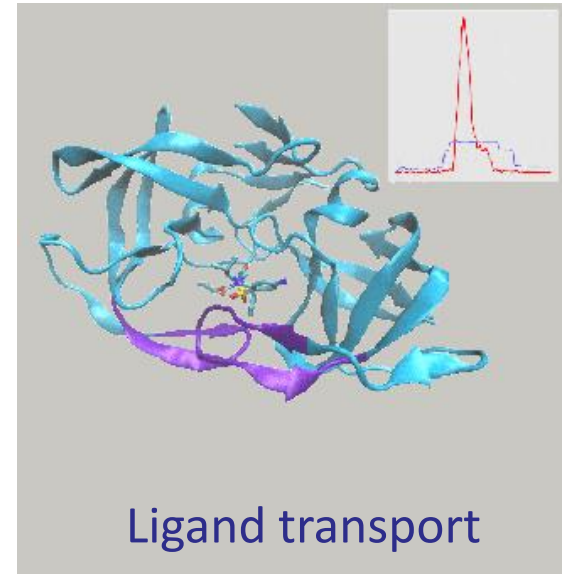
Molecular dynamics



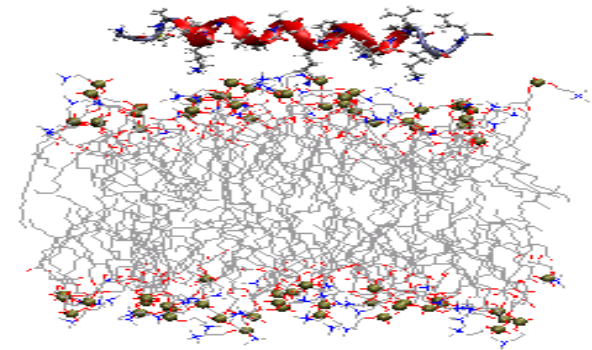
Analysis of interactions



Ligand conversion



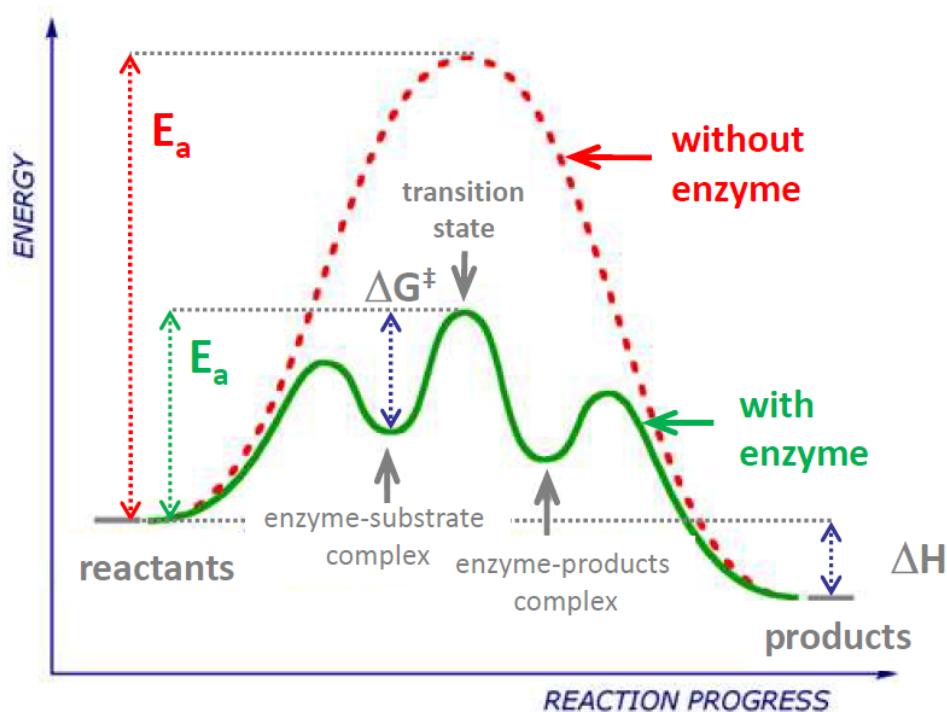
Ligand transport



Interaction with membrane

Quantum chemistry

- Modeling of reaction barriers
- Enzymes increase speed of chemical reactions by decreasing activation barrier



- Kinetic rate:

$$k = Ae^{\frac{-E_a}{RT}}$$

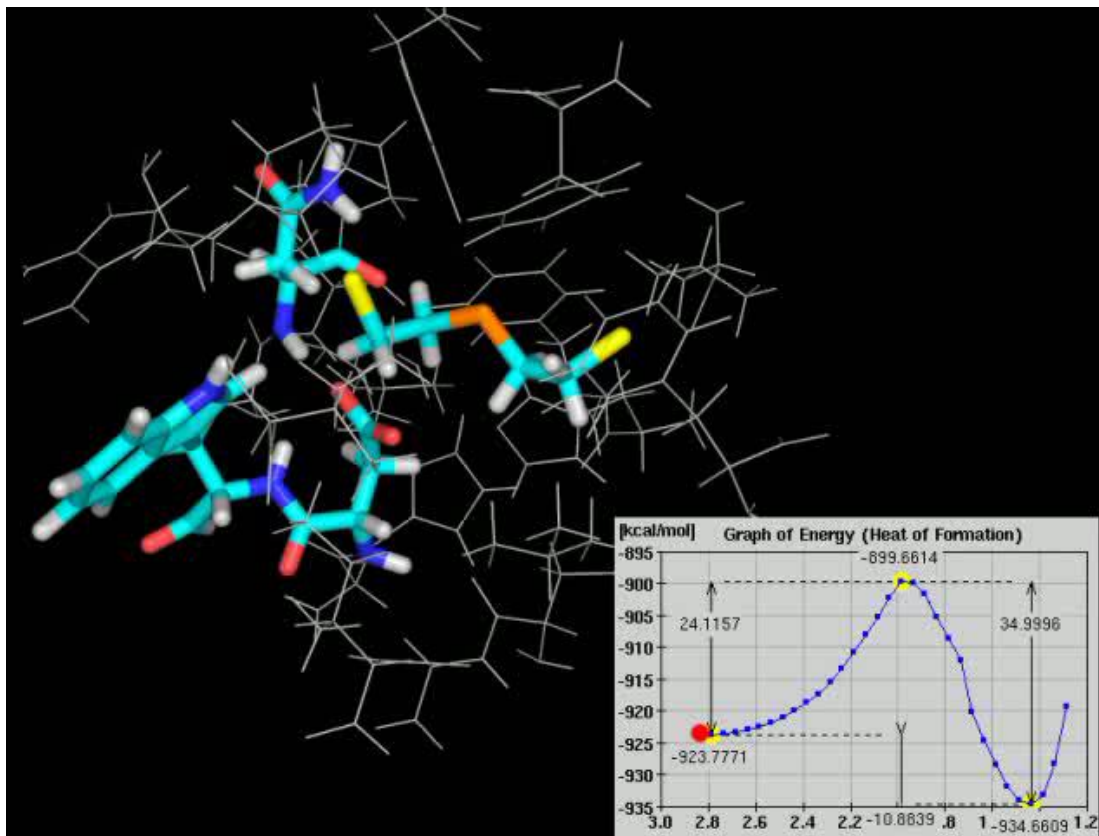
(Arrhenius equation)

- lower $E_a \rightarrow$ higher k
 \Leftrightarrow faster reaction



Quantum chemistry

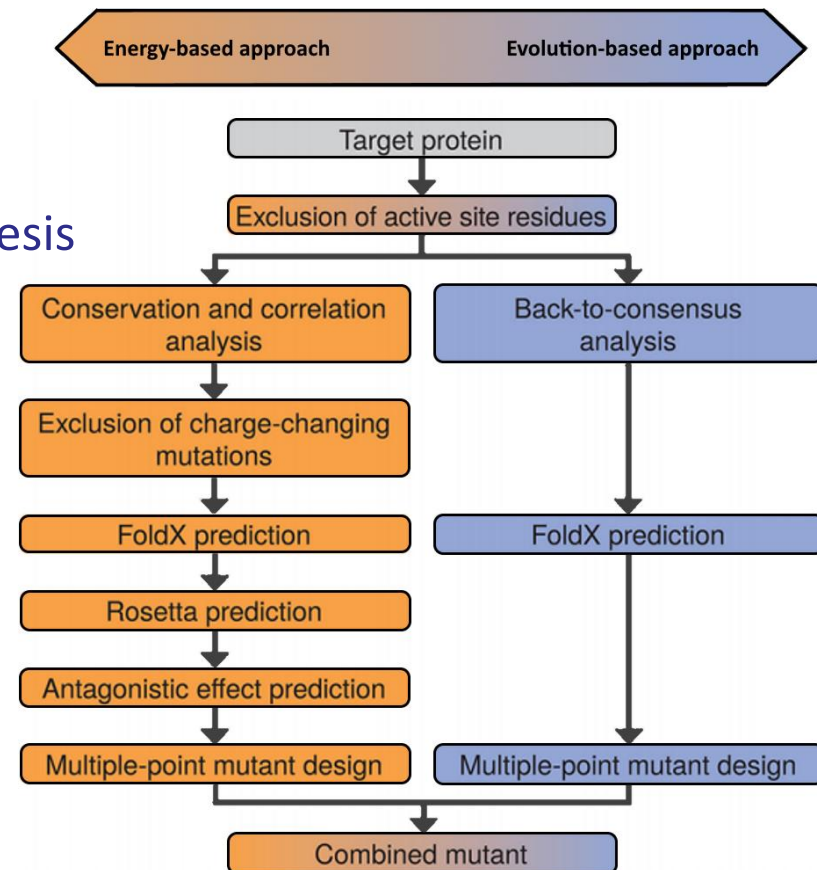
- Using quantum mechanics to create or break bonds (usually hybrid quantum mechanics/ molecular mechanics simulation)



Design of stability

□ FireProt

- <https://loschmidt.chemi.muni.cz/fireprotweb>
- *In silico* analysis of all mutations
- Energy- and evolution-based analyses
- Multiple-point mutants for gene synthesis
- Single-point prediction
- User-defined mutations

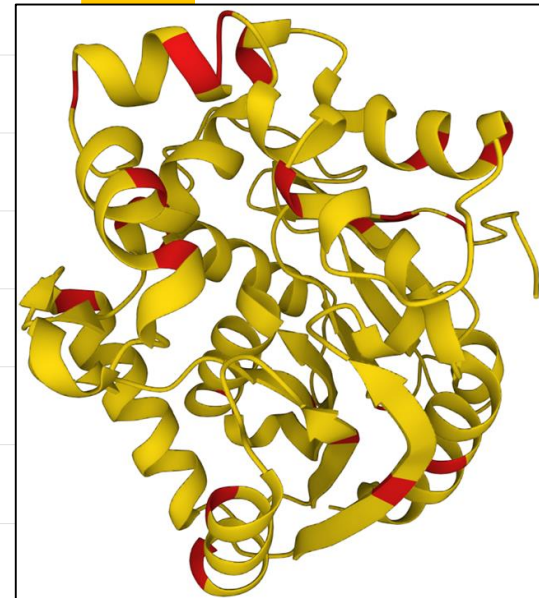


Design of stability

□ FireProt

| Consensus | | Energy low risk | | Energy high risk | | Comb. low risk | | Comb. high risk | | Ancestral design | | All residues | |
|-------------------------------------|-------|-----------------|-----|------------------|--------------|----------------|-------|-----------------|---------|----------------------------------|--|--------------|--|
| <input checked="" type="checkbox"/> | Chain | Position | Ref | Alt | Conservation | Majority | Ratio | Foldx | Rosetta | Add to Design | | | |
| <input checked="" type="checkbox"/> | A | 11 | D | P | 4 | false | false | -1.40276 | -3.423 | <input type="button" value="+"/> | | | |
| <input checked="" type="checkbox"/> | A | 13 | H | F | 5 | false | false | -1.21777 | -2.16 | <input type="button" value="+"/> | | | |
| <input checked="" type="checkbox"/> | A | 23 | H | A | 6 | true | false | 0.451804 | | <input type="button" value="+"/> | | | |
| <input checked="" type="checkbox"/> | A | 33 | T | I | 4 | false | false | -1.55239 | -2.294 | <input type="button" value="+"/> | | | |
| <input checked="" type="checkbox"/> | A | 80 | F | R | 5 | true | true | -0.521484 | | <input type="button" value="+"/> | | | |
| <input checked="" type="checkbox"/> | A | 82 | D | W | 1 | false | false | -1.32168 | -3.595 | <input type="button" value="+"/> | | | |
| <input checked="" type="checkbox"/> | A | 119 | N | H | 6 | true | false | -1.14059 | | <input type="button" value="+"/> | | | |
| <input checked="" type="checkbox"/> | A | 128 | C | F | 6 | true | true | -1.13013 | | <input type="button" value="+"/> | | | |
| <input checked="" type="checkbox"/> | A | 145 | A | L | 2 | false | false | -2.8401 | -2.366 | <input type="button" value="+"/> | | | |
| <input checked="" type="checkbox"/> | A | 148 | T | M | 2 | false | false | -2.13847 | -2.842 | <input type="button" value="+"/> | | | |

Rows per page: 10 ▾ 1-10 of 26 ⏪



Design of stability



□ FireProt^{ASR}

- <https://loschmidt.chemi.muni.cz/fireprotasr>
- ancestral sequence reconstruction
- Analysis of protein evolution, sequence-based protein stabilization
- Ancestrals are highly stable, have broad specificity and good yields

FIREPROT^{ASR} v1.1 Fully automated ancestral sequence reconstruction

Submit new job Help Example Use cases Acknowledgement Job ID: e.g. XXXXXX Find job

SELECT THE STARTING POINT

SEQUENCE  **USER DATA** 



STARTING FROM SEQUENCE Load example

Source : Enter own sequence Upload sequence file

Sequence :

REFERENCE

Musil, M., Khan, R., Beler, 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.

USER STATISTICS

- Number of visitors: 7286
- Number of jobs: 2109

CONTACT

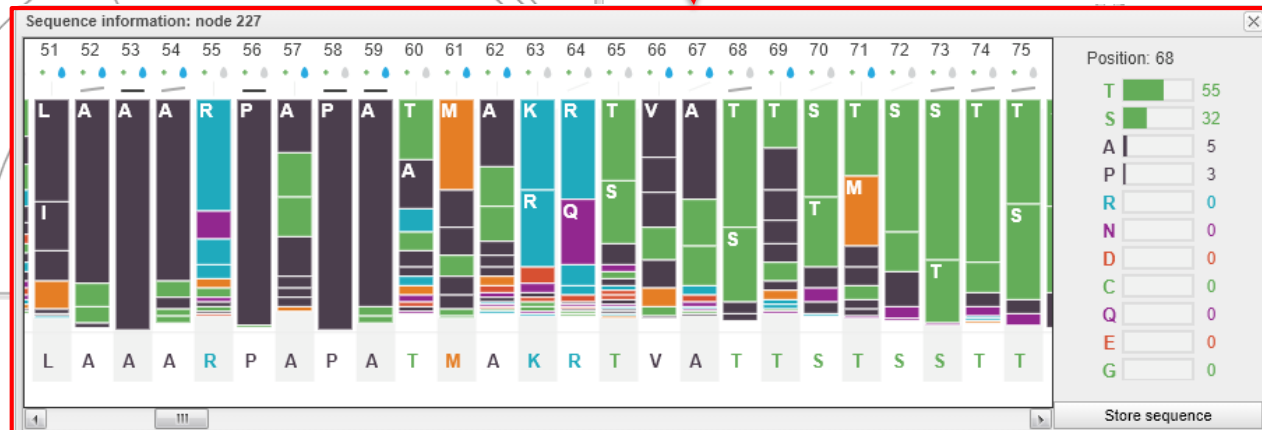
Loschmidt Laboratories

- fireprot@sci.muni.cz
- <https://loschmidt.chemi.muni.cz>

SUPPORTED BY

Design of stability

FireProt^{ASR}



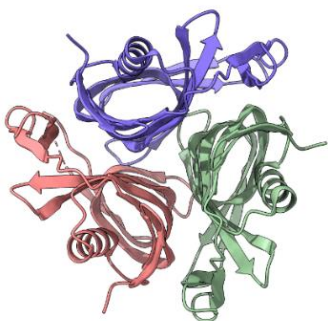
Design of solubility

□ ProteinMPNN

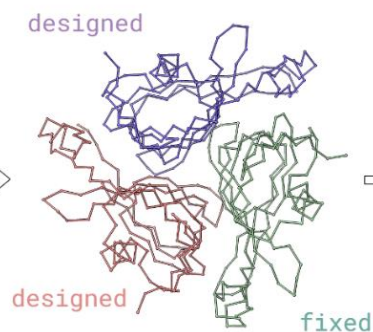
- <https://huggingface.co/spaces/simonduerri/ProteinMPNN>
- deep learning model for protein optimization via mutations
- takes structure on the input and provides optimized sequence folding into the same backbone
- good for improving yields and rescuing folding-compromised designs



Input structure

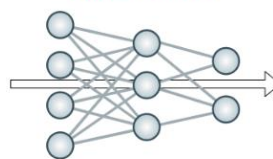


Extract Backbone



Extract backbone and define chains to design

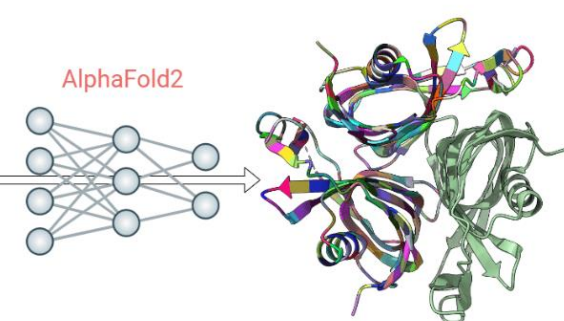
ProteinMPNN



NMYSYKKIGNKYI
VSINNHTEIVKALN
Predicted sequence

From backbone predict diverse sequences using **ProteinMPNN** that fold into the same structure

Verification



Use **AlphaFold2** to predict the structure of the sequence and superimpose with original structure.

From experiment (e.g X-Ray Crystallography) or from design (e.g Rosetta)

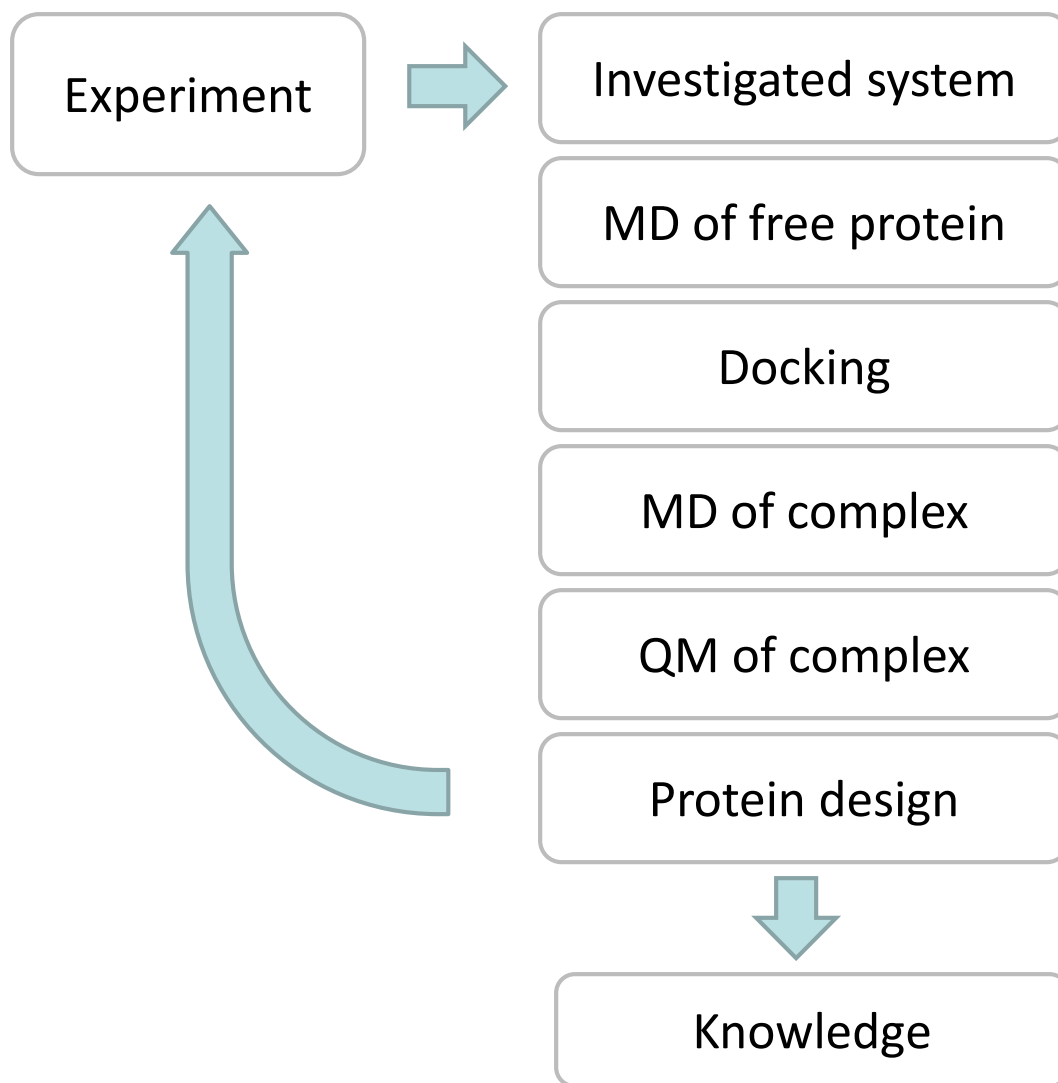
Design of protein-protein interactions

□ AffiLib

- <https://affilib.weizmann.ac.il/bin/steps>
- RosettaDesign and evolution analysis to optimize macromolecular interface
- mutations for improvement of the binding affinity
- up to 50 multiple-point mutants for protein synthesis

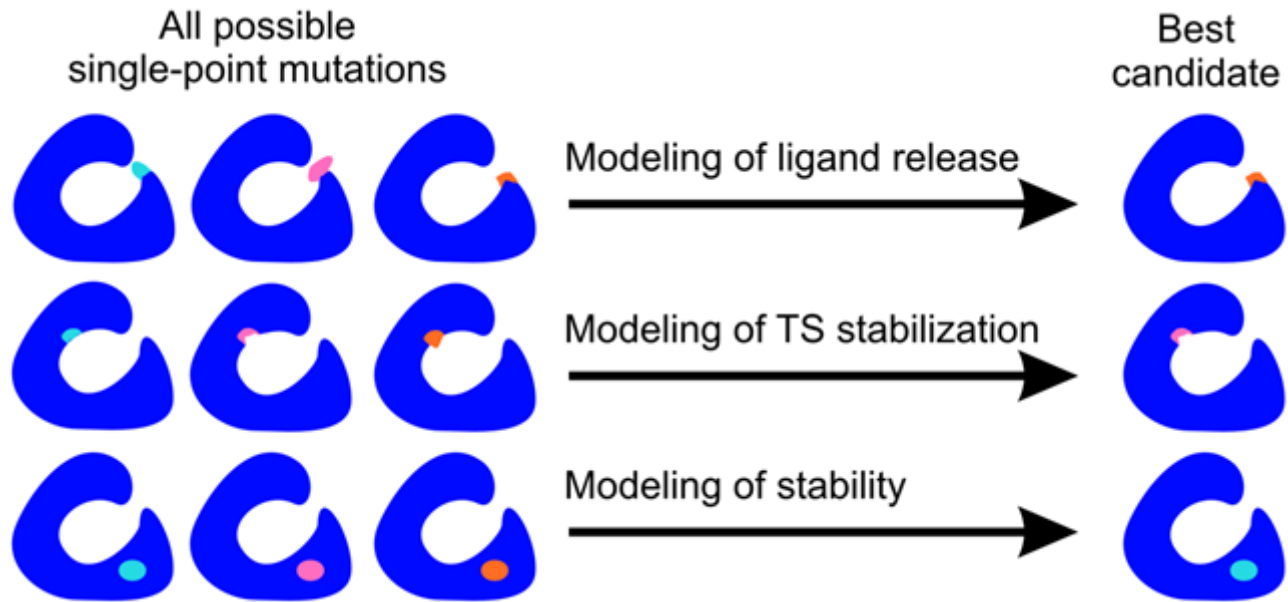
| Parameter | Value | | | | | | | | | | | | | | |
|---|--|------|----|------|---|------|-----|------|--------|------|------|------|---------|------|------|
| Minimal number of mutations per design | <input type="text" value="3"/> | | | | | | | | | | | | | | |
| Maximal number of mutations per design | <input type="text" value="5"/> | | | | | | | | | | | | | | |
| Minimal PSSM threshold | <input type="text" value="-1"/> | | | | | | | | | | | | | | |
| $\Delta\Delta G$ | <input type="text" value="5.5"/> | | | | | | | | | | | | | | |
| Sequence space | <table><tbody><tr><td>143A</td><td>FY</td></tr><tr><td>144A</td><td>P</td></tr><tr><td>151A</td><td>FMY</td></tr><tr><td>177A</td><td>LAGNST</td></tr><tr><td>211A</td><td>ILMV</td></tr><tr><td>247A</td><td>AGMSTVY</td></tr><tr><td>248A</td><td>LIMV</td></tr></tbody></table> | 143A | FY | 144A | P | 151A | FMY | 177A | LAGNST | 211A | ILMV | 247A | AGMSTVY | 248A | LIMV |
| 143A | FY | | | | | | | | | | | | | | |
| 144A | P | | | | | | | | | | | | | | |
| 151A | FMY | | | | | | | | | | | | | | |
| 177A | LAGNST | | | | | | | | | | | | | | |
| 211A | ILMV | | | | | | | | | | | | | | |
| 247A | AGMSTVY | | | | | | | | | | | | | | |
| 248A | LIMV | | | | | | | | | | | | | | |
| Total number of designs in tolerated sequence space | 3,313 | | | | | | | | | | | | | | |

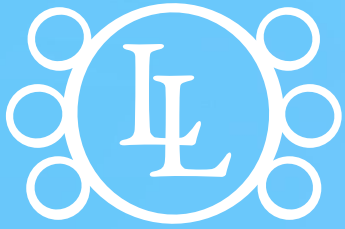
Design of mutations



Design of mutations

- **design** of modified enzymes by *in silico* screening
 - study of effects of all relevant mutations
 - selection and combination of the best mutations





**LOSCHMIDT
LABORATORIES**



PROTEIN ENGINEERING

8. Directed evolution

Loschmidt Laboratories

Department of Experimental Biology

Masaryk University, Brno