

**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
  - directed evolution
  - rational design
  - semi-rational design



# Protein engineering approaches

## RATIONAL DESIGN

1. Computer aided design



2. Site-directed mutagenesis



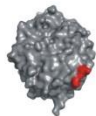
Individual mutated gene

3. Transformation

4. Protein expression

5. Protein purification

6. *not applied*



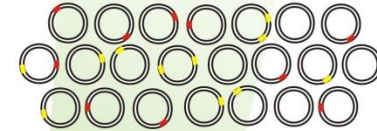
Constructed mutant enzyme

7. Biochemical testing

## DIRECTED EVOLUTION

1. *not applied*

2. Random mutagenesis



Library of mutated genes  
( >10,000 clones )

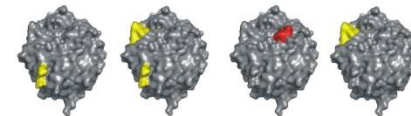
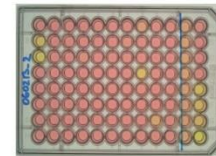
3. Transformation

4. Protein expression

5. *not applied*

6. Screening and selection

- stability
- selectivity
- affinity
- activity



Selected mutant enzymes



**IMPROVED  
ENZYME**

# Protein engineering approaches



	<b>Rational design</b>	<b>Directed evolution</b>	<b>Semi-rational design</b>
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



# Structure information

## □ worldwide Protein Data Bank (wwPDB)

- <http://www.wwpdb.org/>
- central repository of ~180,000 experimental macromolecular structures (April 2021)

## □ RCSB PDB

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



## □ PDBe

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



## □ PDBj

- <https://pdbj.org/>



# Homology modelling

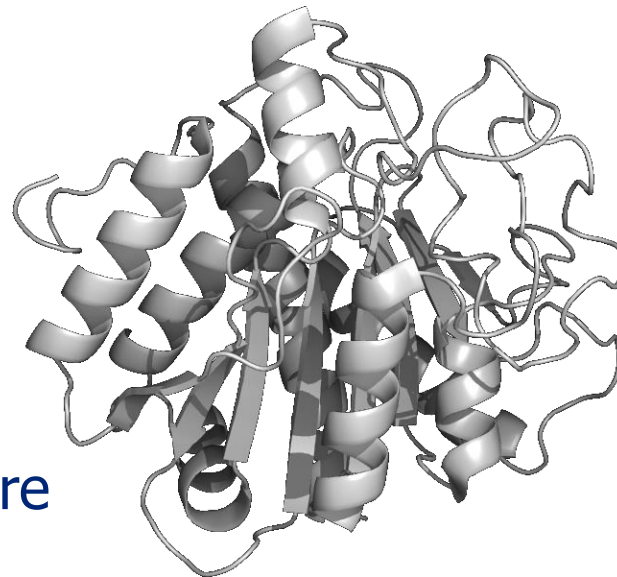
3D structure is determined by the sequence

```
MSLGAKPFGEEKKFIEIKGRRMAYIDEGTGDPILFQHGNTSSYLWRNI  
MPHCAGLGRLIACDLIGMGDSKLDPSGPERYAYAEHRDYLDALWEA  
LDLGDRVVLVHWDWGSALGFDWARRHRERVQGIAYMEAIAMPIEWA  
DFPEQDRDLFQAFRSQAGEELVLQD
```

sequence

**prediction**

structure

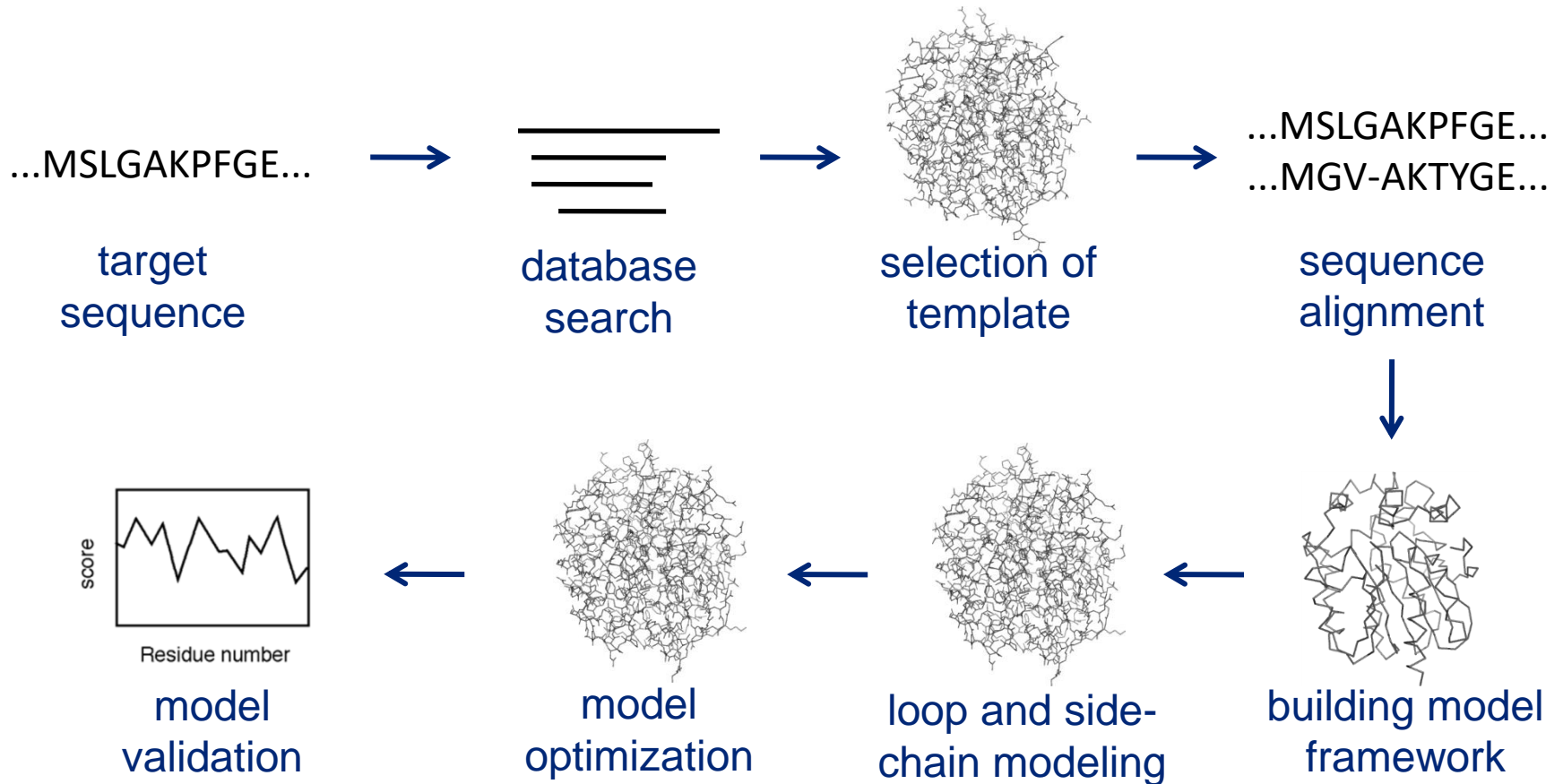


function





# Homology modelling



# Homology modelling - tools



- ❑ **MODELLER**

- <http://salilab.org/modeller/>

- ❑ **SWISS-MODEL**

- <http://swissmodel.expasy.org/>

- ❑ **Robetta**

- <http://rosetta.bakerlab.org/>

- ❑ **I-TASSER**

- <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>



## ❑ 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**

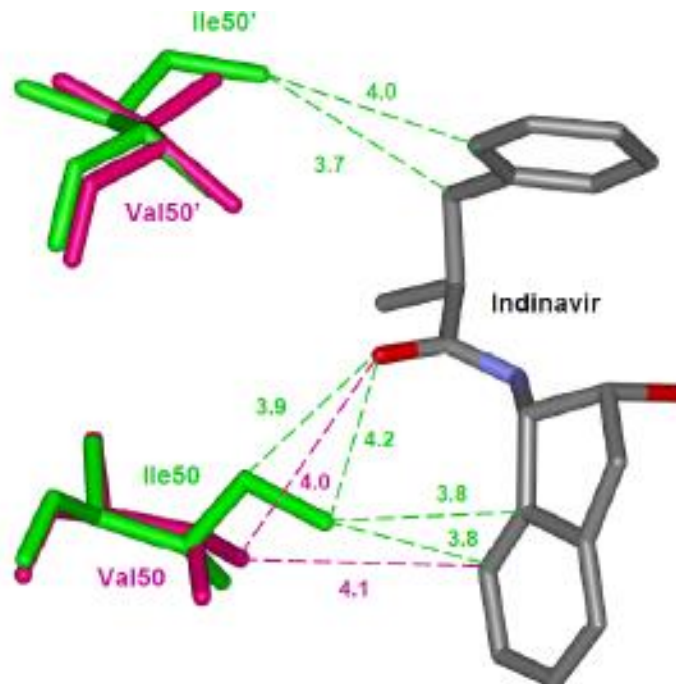
# Hot-spots for engineering catalytic properties

- 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

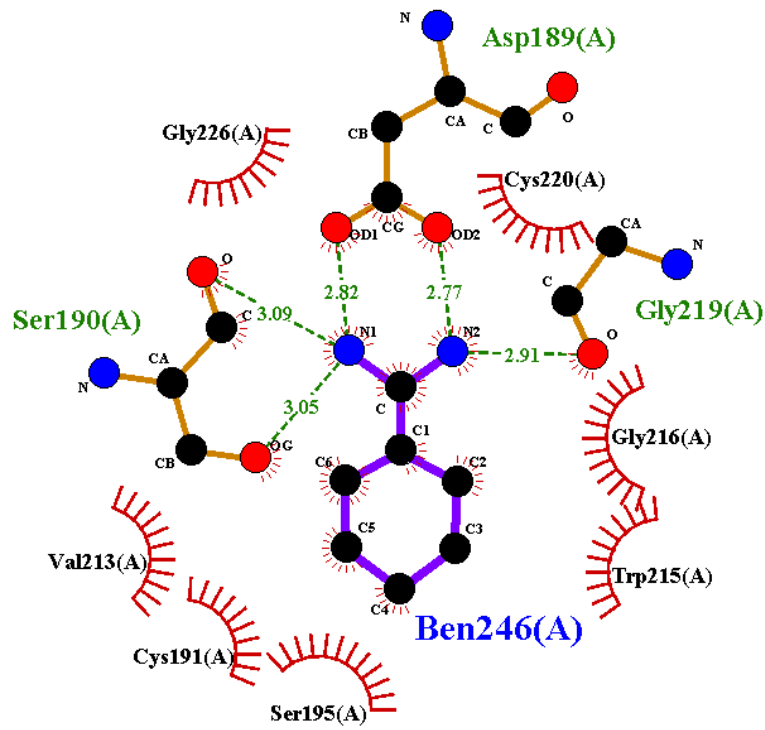
- requires 3D structure of protein-ligand complex
  - experimental structure (wwPDB)
  - theoretical model (molecular docking)



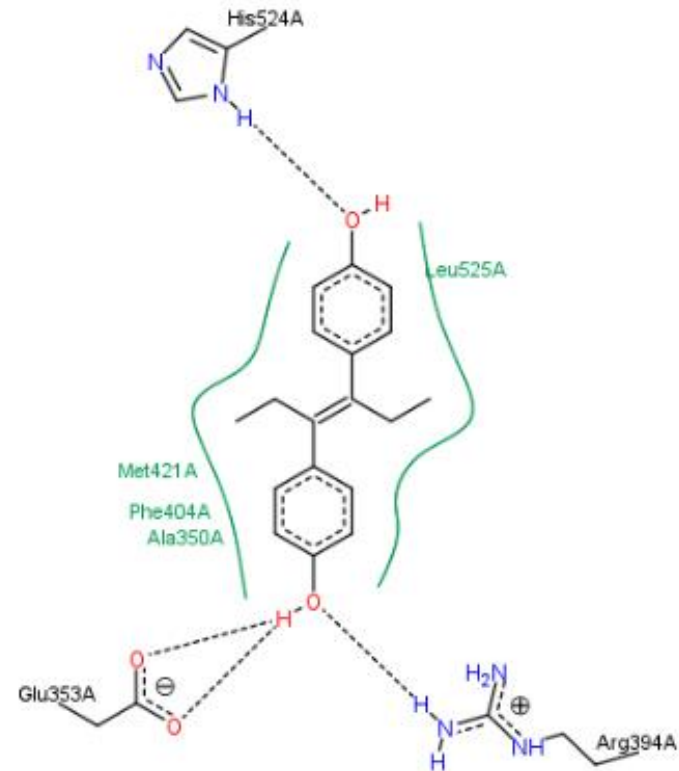


# Analysis of protein-ligand interactions

- schematic diagrams of protein-ligand interactions



LigPlot, LigPlot+



PoseView

# Analysis of protein-ligand interactions



- inter-atomic contacts between protein and bound ligands

Residue		Dist(Å)	Surf(Å <sup>2</sup> )	Number of contacts
ASN	38 A	2.7	22.5	2
ASP	108 A	2.8	35.1	5
ILE	134 A	6.3	0.7	1
PHE	143 A	5.0	6.5	2
PHE	151 A	3.3	26.7	4
PHE	169 A	3.5	6.4	2
VAL	173 A	3.6	23.4	1
LEU	177 A	4.8	8.5	2
ILE	211 A	5.2	3.8	1
LEU	248 A	5.6	10.3	4
HIS	272 A	3.8	33.8	9
PHE	273 A	3.5	2.3	2
BR	901 A	3.8	30.7	2



LPC server

# Essential residues



UniProt (Swiss-Prot) - <https://www.uniprot.org/>

## Sites

Feature key	Position(s)	Description
Binding site <sup>i</sup>	38	Halide
Active site <sup>i</sup>	108	Nucleophile <span>3 Publications</span>
Binding site <sup>i</sup>	109	Halide
Active site <sup>i</sup>	132	Proton donor <span>3 Publications</span>
Active site <sup>i</sup>	272	Proton acceptor <span>3 Publications</span>

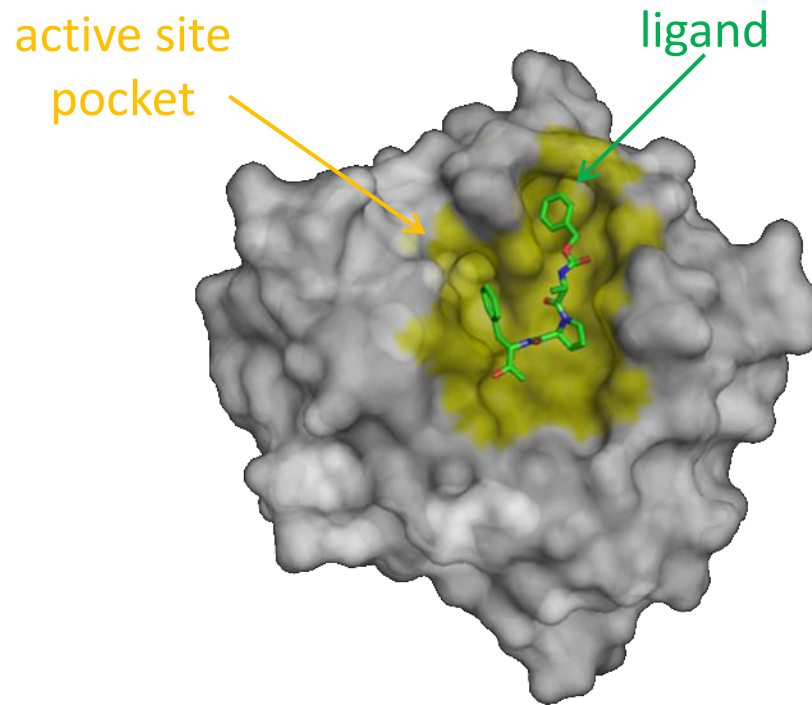
Catalytic Site Atlas - <https://www.ebi.ac.uk/thornton-srv/m-csa/>

## Catalytic Residues Roles

UniProt	PDB*	
His272	His272A	Acts as general acid base to deprotonate water, thus activating water so its lone pair can attack the covalent enzyme intermediate.
Asp108	Asp108A	Acts as nucleophile on the electrophilic carbon atom to form a covalent enzyme intermediate which is hydrolysed to give the product.
Asn38, Trp109	Asn38A, Trp109A	Involved in stabilisation of the halogen, transition-states and product.
Glu132	Glu132A	Acts to modify the pKa of His 272 so that it remains in the correct protonation state for its role in catalysis.

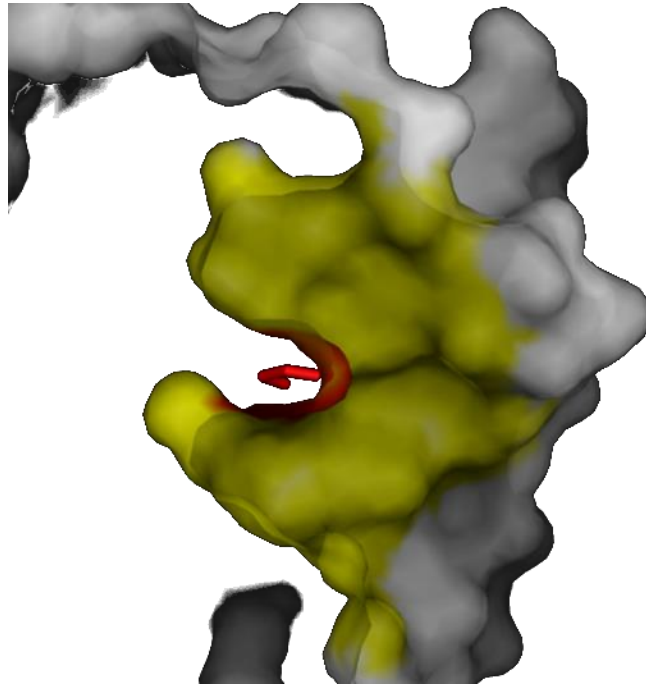
# Analysis of binding pockets

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



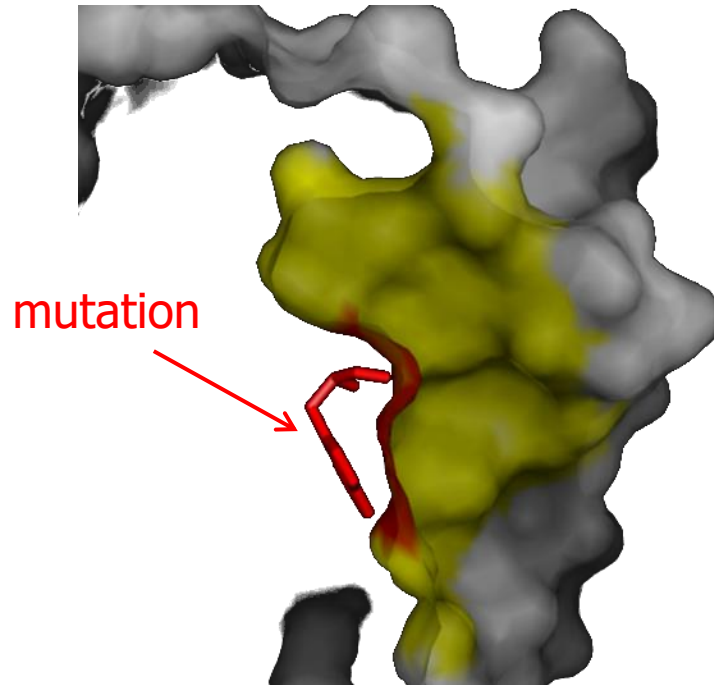
# Analysis of binding pockets

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



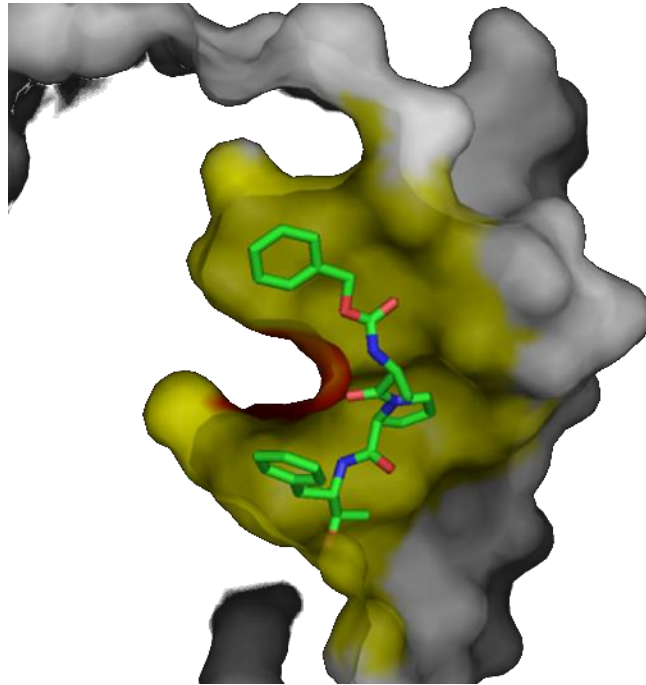
# Analysis of binding pockets

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



# Analysis of binding pockets

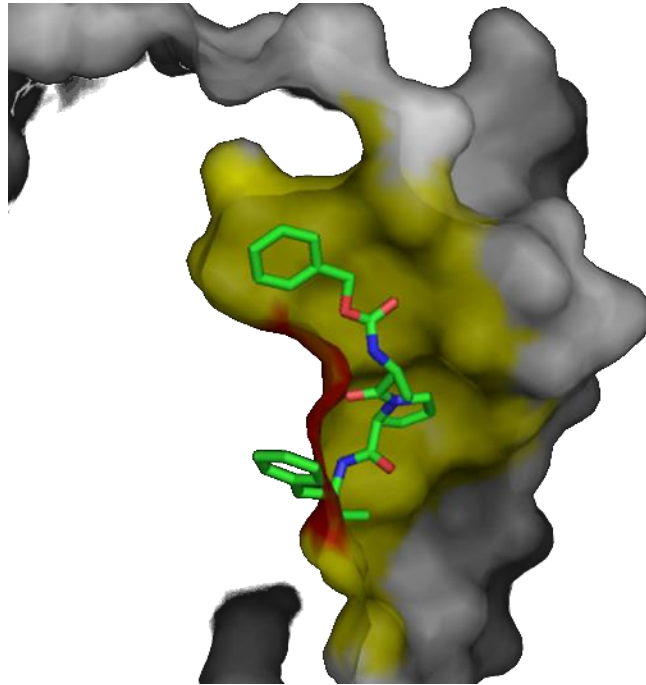
- binding and active sites of enzymes are often associated with structural pockets and cavities



A 3D molecular model illustrating an enzyme binding pocket. The protein structure is shown in a semi-transparent grey surface representation. The binding pocket is highlighted in a yellow-green color. A ligand molecule, consisting of a green benzene ring connected to a red and blue chain, is shown bound within the pocket. The ligand is oriented vertically, with the benzene ring at the top and the chain extending downwards. The pocket's shape is irregular and deep, with a narrow entrance at the top.

# 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 yellow-green color. A ligand molecule, represented by a stick model with green, red, and blue atoms, is shown 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**
  - 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, MetaPocket, Caver Analyst...



# Analysis of binding pockets

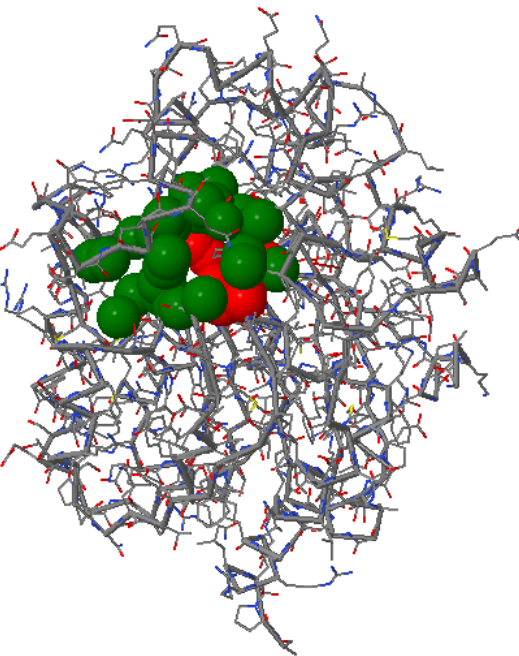
- detailed characterization of all pockets in the structure

**jobID: 1d07** hydrolase  
hydrolytic haloalkane dehalogenase linb from sphingomonas paucimobilis ut26 with 1,3-propanediol, a product of debromidation of dibromopropane, at 2.0a resolution

**Pocket Information**

ID	Area	Vol
<input checked="" type="checkbox"/> 42	356.4	364.2
<input type="checkbox"/> 41	291.3	318.2
<input type="checkbox"/> 40	105.1	80
<input type="checkbox"/> 39	101.6	79.2
<input type="checkbox"/> 38	89.6	67.3
<input type="checkbox"/> 37	68	62.1
<input type="checkbox"/> 36	116	82.5
<input type="checkbox"/> 35	71.9	86.8
<input type="checkbox"/> 34	95.3	65.8
<input type="checkbox"/> 33	65.1	43.9
<input type="checkbox"/> 32	80.8	57.1
<input type="checkbox"/> 31	76.6	74.3
<input type="checkbox"/> 30	38.3	21.6
<input type="checkbox"/> 29	84.6	53.8

42	38	CB	ASN	A
42	38	ND2	ASN	A
42	38	O	ASN	A
42	108	CG	ASP	A
42	108	OD1	ASP	A
42	108	OD2	ASP	A
42	109	CD1	TRP	A
42	109	NE1	TRP	A
42	134	CD1	ILE	A
42	134	CG1	ILE	A



Jmol

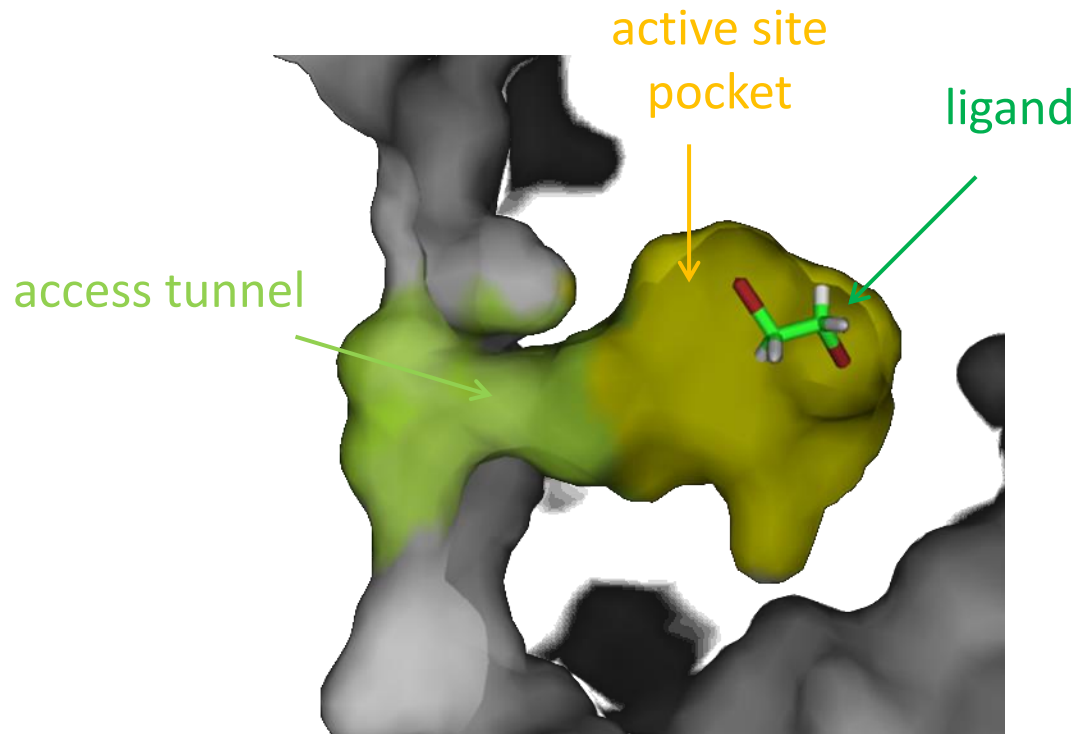
**Annotated Sites**

Residue:	ASN
Residue #Chain:	38:A
Pocket/Pockets:	30,42
Databases	
SWP	BINDING Halide.
Residue:	ASP
Residue #Chain:	108:A
Pocket/Pockets:	30,42
Databases	
SWP	ACT_SITE Nucleophile.
Residue:	TRP
Residue #Chain:	109:A
Pocket/Pockets:	42,32
Databases	
SWP	BINDING Halide.
Residue:	GLU
Residue #Chain:	132:A
Pocket/Pockets:	24,34
Databases	
SWP	ACT_SITE Proton_donor.
Residue:	HIS
Residue #Chain:	272:A
Pocket/Pockets:	42
Databases	
SWP	ACT_SITE Proton_acceptor.

CASTp

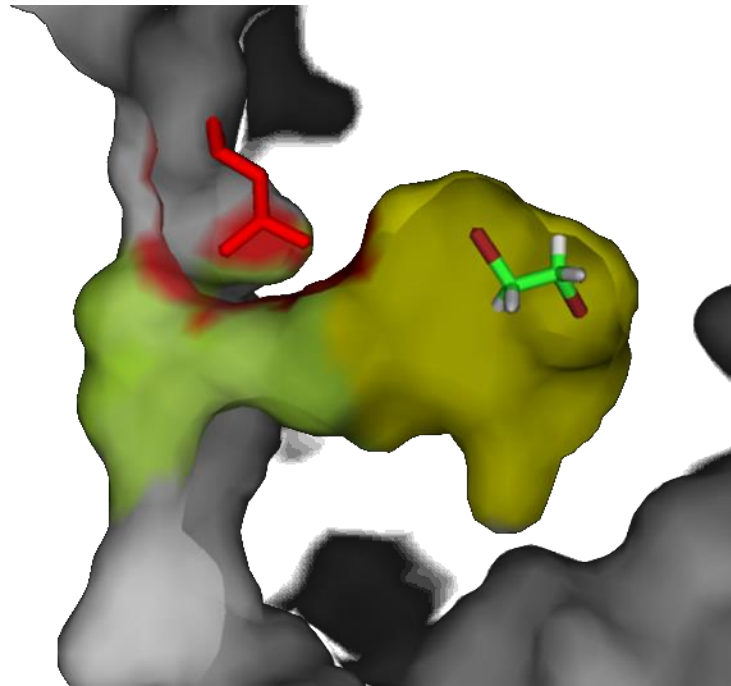
# Analysis of access tunnels

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



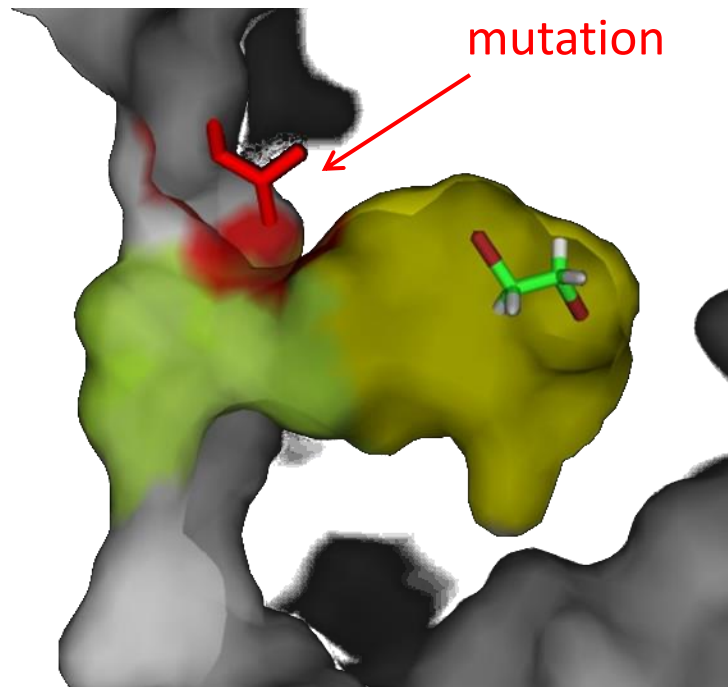
# Analysis of access tunnels

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



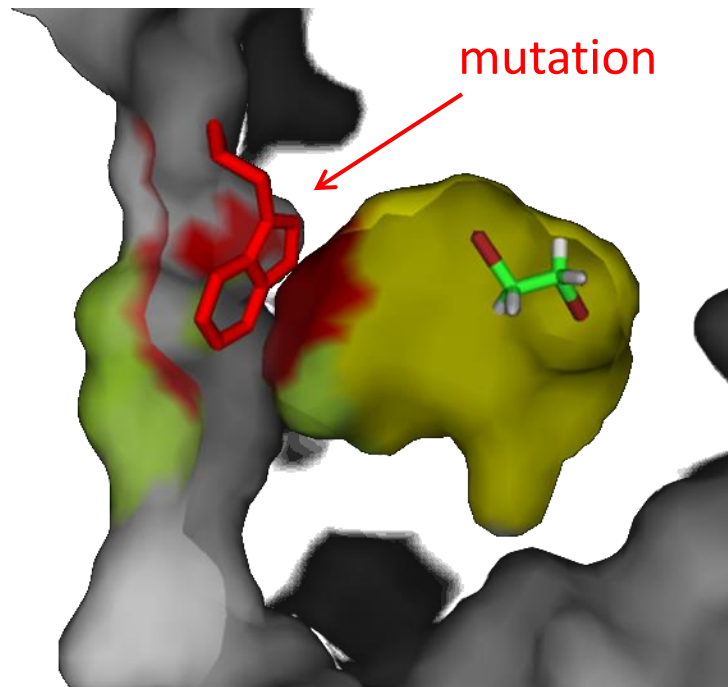
# Analysis of access tunnels

- 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



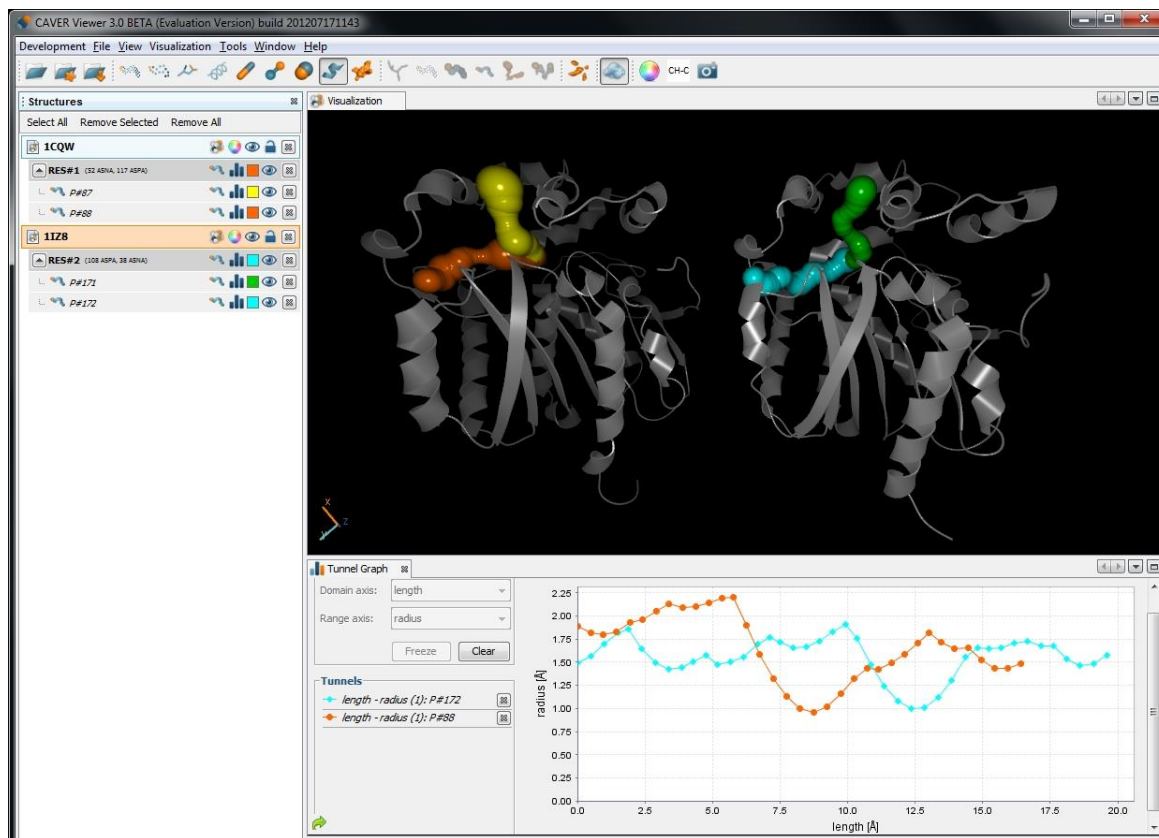
# Analysis of access tunnels

- ❑ 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

- Detection and detailed characteristics of access tunnels



CAVER Analyst 2.0



# Hot-spots for engineering thermostability

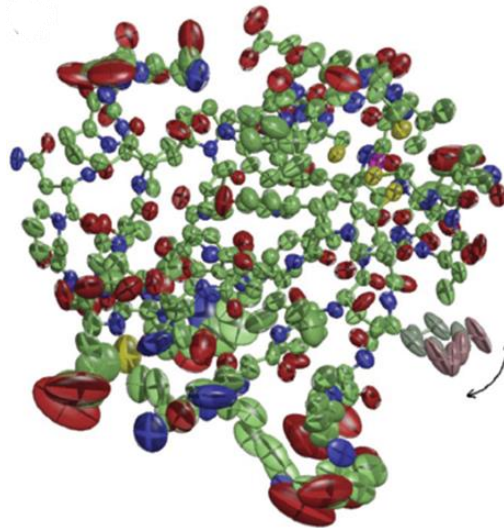
- ❑ highly **flexible** residues – introduction of rigidifying mutations
- ❑ residues located in **access tunnels**
- ❑ residues predicted by systematic ***in silico* saturation** mutagenesis

→ these residues may also include **catalytic** or other **essential** residues which generally should not be mutated!



# Identification of highly flexible residues

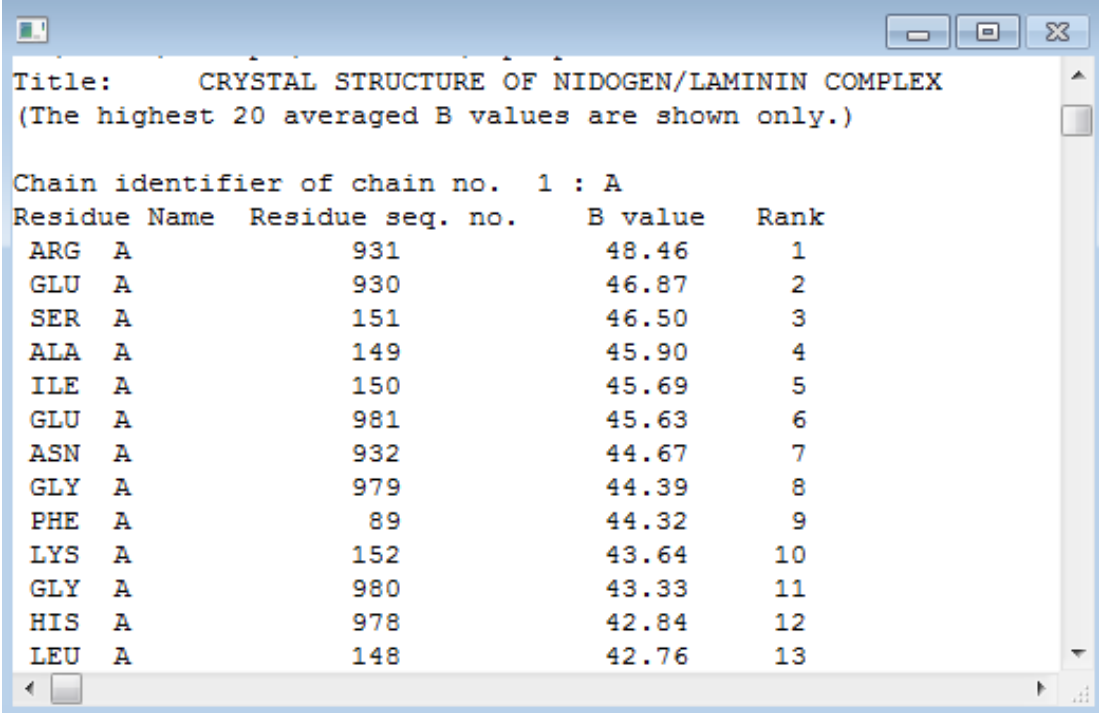
- prediction based on **crystallographic B-factors**
  - reflect the degree of thermal motion, and thus the flexibility of individual residues



- 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.)

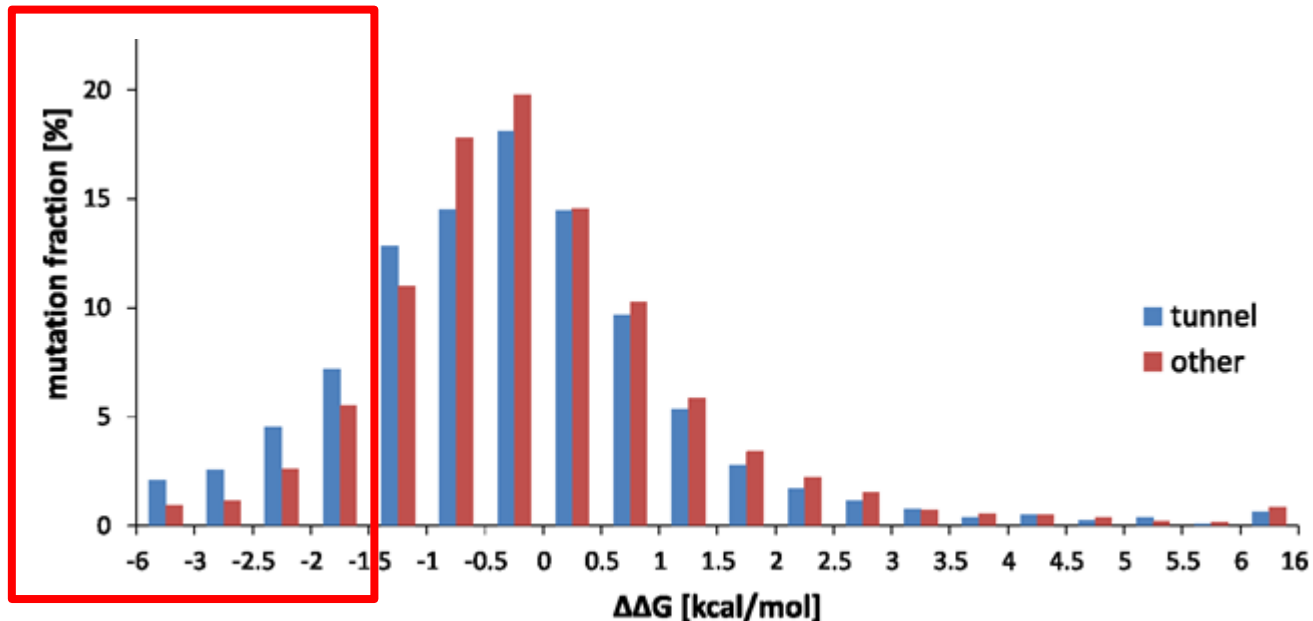
Chain identifier of chain no. 1 : A

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)



# Analysis of access tunnels

- Detection of tunnels in proteins and analysis of ligand transport

The screenshot displays the CAVER web interface, which is a tool for analyzing tunnels and channels in protein structures. The interface is divided into several panels:

- Tunnels info:** A table listing detected tunnels. Tunnel 1 has a bottleneck radius of 1.5 Å, a length of 10.8 Å, and a throughput of 0.73. Tunnel 2 has a bottleneck radius of 1.3 Å, a length of 16.1 Å, and a throughput of 0.00.
- Job information:** Shows the job ID (logrvv), title (Untitled), and structure (4q60). It includes buttons for downloading PyMOL sessions, results, and configuration files.
- Visualization settings:** Allows users to choose between 'Spheres' and 'Line' for tunnel visualization, and 'Wireframe' or 'Carbon' for structure visualization. It also includes options for 'Sticks', 'Trace', 'Balls & sticks', and 'Backbone'.
- Tunnel profile:** A graph showing the radius of the tunnel (in Å) versus its length (in Å). A specific point is highlighted with a tooltip: 'Point #2 of tunnel 1', with a length of 0.5 Å, a radius of 2.1 Å, and coordinates [0.6, -0.7, 9.3].
- Details for tunnel 1:** Provides a detailed view of the selected tunnel, including its radius (1.5 Å), distance from the starting point (0.5 Å), and coordinates (5.2, 2.6, 11.9). It also lists the bottleneck residues: Tip (141, A), Phe (144, A), Ala (145, A), and Thr (148, A).

CAVER Web

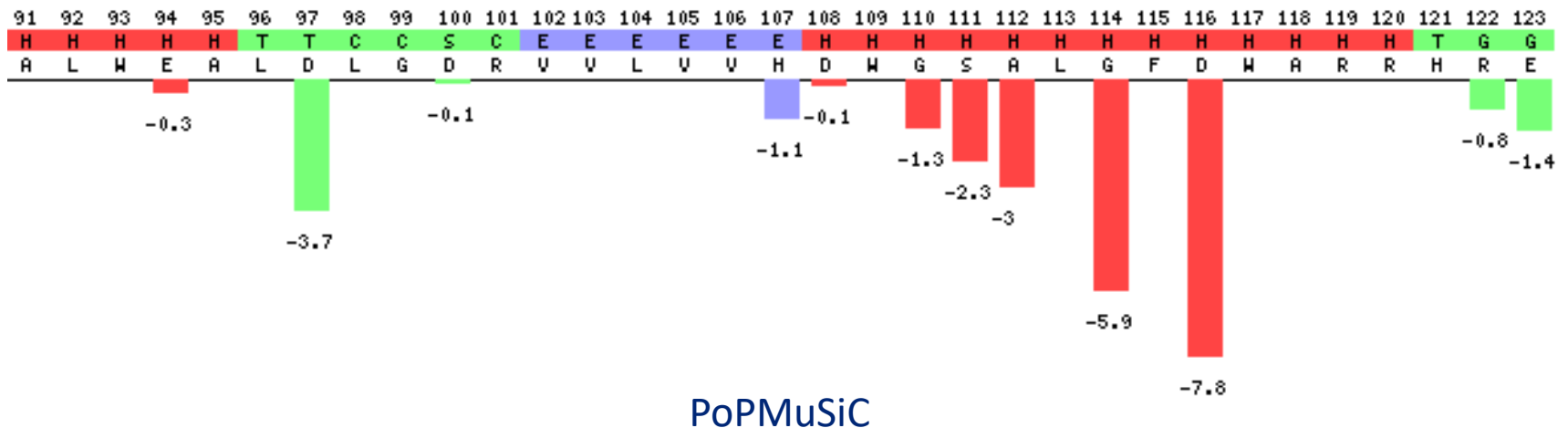
# Systematic *in silico* saturation mutagenesis

- computational tools for the prediction of effect of amino acid substitutions on protein stability
  - each residue in the protein structure is replaced by all other possible amino acids and the **change in folding free energy ( $\Delta\Delta G$ )** upon mutation is estimated
  - positions with a **high proportion of stabilizing** mutations and/or **low proportion of destabilizing** mutations are good candidates for randomization by experimental saturation mutagenesis
- usually requires 3D structure of protein
  - experimental structure (wwPDB)
  - theoretical model (homology modeling)



# Systematic *in silico* saturation mutagenesis

- fast systematic scan of all possible single-point mutations – prediction of stability changes upon mutation
- sequence optimality score (the sum of all negative  $\Delta\Delta G$ s at a given position) – indicates poorly optimized positions





- 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

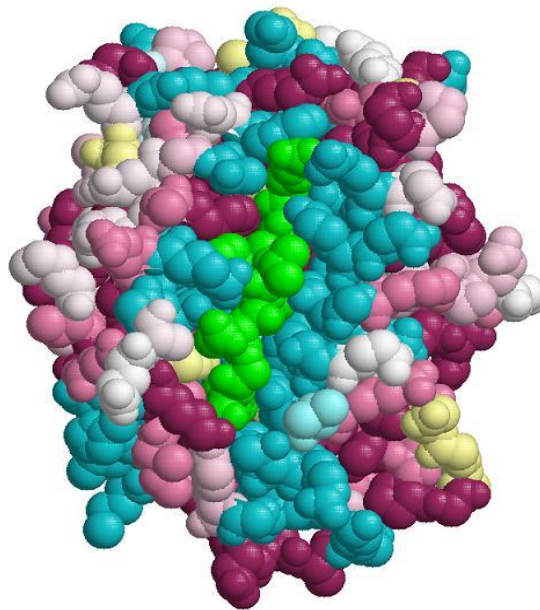
- residue conservation can be derived from a **multiple alignment** of a set of related proteins (3D structure not required)

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 (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

Amino Acid Mutations			
Amino acid	Overall Stability	Torsion	Predicted $\Delta\Delta G$ (kcal/mol)
GLY	Stabilising	Unfavourable	1.48
ALA	Destabilising	Unfavourable	-0.9
VAL	Destabilising	Unfavourable	-2.23
ILE	Destabilising	Unfavourable	-2.12
MET	Stabilising	Unfavourable	1.89
PRO	Stabilising	Unfavourable	1.55
TRP	Stabilising	Favourable	2.73
SER	Stabilising	Unfavourable	1.2
THR	Destabilising	Unfavourable	-0.44
PHE	Stabilising	Favourable	3.64
GLN	Destabilising	Unfavourable	-0.69
LYS	Stabilising	Unfavourable	9.91
TYR	Stabilising	Favourable	0.96
ASN	Stabilising	Favourable	4.14
CYS	Destabilising	Favourable	-6.73
GLU	Stabilising	Unfavourable	4.98
ASP	Stabilising	Favourable	1.31
ARG	Stabilising	Unfavourable	2.94
HIS	Stabilising	Favourable	1.38

CUPSAT

# 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...

# Prediction of mutation effects

- prediction of effect of substitutions on protein function

**PREDICTSNP<sup>2</sup>** Unified platform for prediction of SNP effect in distinct genomic regions

Home Help Job ID:  Find job

**JOB INFORMATION**

ID: o0tpgx  
Title: Example  
Calculation successful, showing results.

**RESULTS** ■ neutral ■ deleterious ■ unknown XX % expected accuracy

Variant	Region	Region function	Prediction tools					Databases									
			PredictSNP2	CADD	DANN	FATHMM	FunSeq2	GWAVA	dbSNP	GenBank	Clinvar	OMIM	Regulome	HaploReg	UCSC	Ensembl	PredictSNP1
6:1613076,A-T	UTR3		97 %	79 %	75 %	91 %	76 %	78 %									
13:84452863,C-T	UTR3		97 %	83 %	89 %	94 %	?	86 %									
1:45480678,G-A	exonic	synonymous	93 %	87 %	87 %	97 %	93 %	54 %									
18:48575659,A-G	intronic		91 %	79 %	62 %	92 %	67 %	80 %									
11:5248388,G-A	upstream		91 %	86 %	66 %	91 %	64 %	86 %									
20:35532559,C-A	splicing		89 %	69 %	72 %	69 %	65 %	?									
18:21118528,G-C	exonic	nonsynonymous	87 %	80 %	62 %	83 %	61 %	?									
16:31202373,C-T	exonic	stopgain	57 %	53 %	54 %	69 %	65 %	68 %									
9:6534707,C-T	splicing		58 %	69 %	72 %	69 %	64 %	77 %									
5:131934564,C-T	exonic	stopgain	58 %	53 %	51 %	64 %	77 %	76 %									

Filter by category:  Prioritize by tool:  Genome assembly:

**DOWNLOAD (ASSEMBLY GCHR37/HG19)**



- 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. <sup>[a]</sup>	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	$> 1.0 \times 10^9$	$> 3.2 \times 10^9$	$> 2.9 \times 10^8$	$> 8.9 \times 10^8$
7	$> 3.4 \times 10^{10}$	$> 1.0 \times 10^{11}$	$> 3.5 \times 10^9$	$> 1.1 \times 10^{10}$
8	$> 1.0 \times 10^{12}$	$> 3.3 \times 10^{12}$	$> 4.2 \times 10^{10}$	$> 1.3 \times 10^{11}$
9	$> 3.5 \times 10^{13}$	$> 1.0 \times 10^{14}$	$> 5.1 \times 10^{11}$	$> 1.5 \times 10^{12}$
10	$> 1.1 \times 10^{15}$	$> 3.4 \times 10^{15}$	$> 6.1 \times 10^{12}$	$> 1.9 \times 10^{13}$

[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



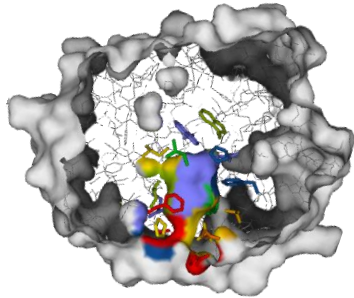
# HotSpot Wizard



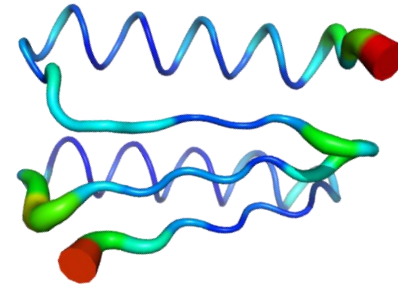
- ❑ meta-server combining several tools
  - automatic **identification of hot-spots** for engineering of enzyme catalytic properties
  - prioritization of hot-spots by their mutability
  - distribution of amino acids at individual positions

# HotSpot Wizard

## Functional hot-spots



## Stability hot-spots (flexibility)



## Stability hot-spots (evolution)

T	S	S	Y	L	W	Y	N	I	M	P	N	H	C	A	G	L
-	-	S	W	L	W	R	N	I	M	-	-	H	C	A	G	L
T	S	S	Y	L	W	Y	N	I	M	P	N	H	C	A	G	L
T	S	S	Y	L	W	R	N	I	M	P	N	H	C	A	G	L
T	S	S	Y	L	W	R	N	I	M	P	P	P	P	A	G	L
T	S	S	Y	L	W	R	N	I	M	P	P	P	P	A	G	L
T	S	S	Y	L	W	R	N	I	M	P	P	P	P	A	G	L
T	S	S	Y	L	W	R	N	I	M	P	N	H	C	A	G	L
T	S	S	Y	L	W	R	N	I	M	P	N	H	C	A	G	L
T	S	S	Y	L	W	R	N	I	M	P	N	H	C	A	G	L

Y ⇒ R

## Correlated hot-spots

T	S	S	R	L	W	Y	N	I	D	P	N	H	C	A	G	L
-	-	S	R	L	W	R	N	I	D	-	-	H	C	A	G	L
T	S	S	R	L	W	R	N	I	D	P	N	H	C	A	G	L
T	S	S	K	L	W	R	N	I	E	P	N	H	C	A	G	L
T	S	S	K	L	W	R	N	I	E	P	P	P	P	A	G	L
T	S	S	K	L	W	R	N	I	E	P	P	P	P	A	G	L
T	S	S	K	L	W	R	N	I	E	P	P	P	P	A	G	L
T	S	S	K	L	W	R	N	I	E	P	N	H	C	A	G	L
T	S	S	W	L	W	R	N	I	V	P	N	H	C	A	G	L
T	S	S	W	L	W	R	N	I	V	P	N	H	C	A	G	L

↑ ↑

# HotSpot Wizard

## Functional hot spots of 1CV2

Viewer



Residue features

- Exclude correlated positions  
  Exclude catalytic pockets  
  Exclude tunnels  
  Exclude  $\alpha$ -helices and  $\beta$ -sheets  
 Exclude buried residues  
  Include residues with moderate mutability

Show all residues

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

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

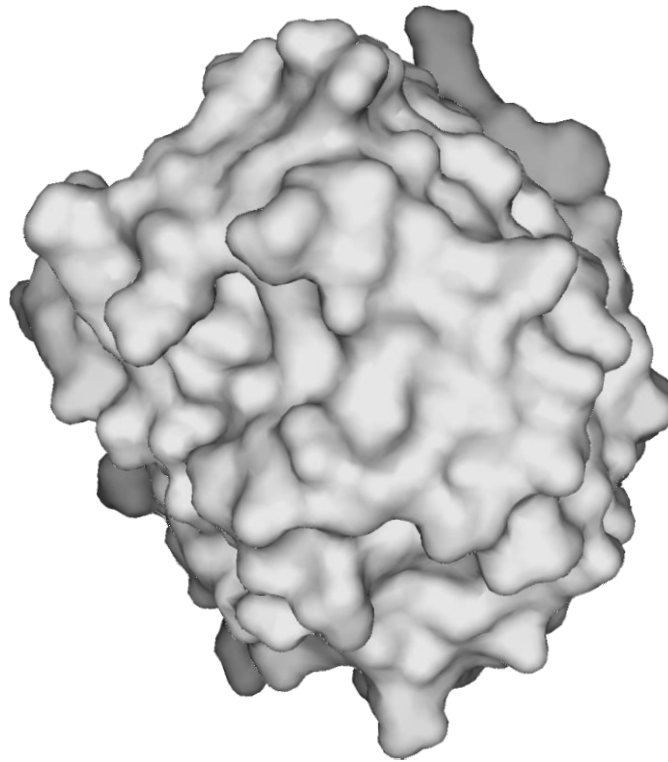
Zoom residues

Design library

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

# HotSpot Wizard

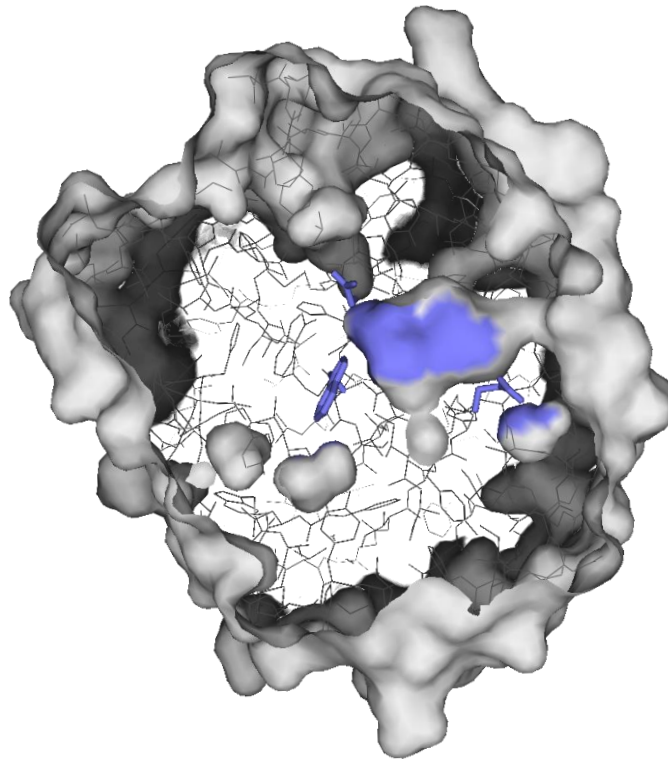
## 1. protein structure





# HotSpot Wizard

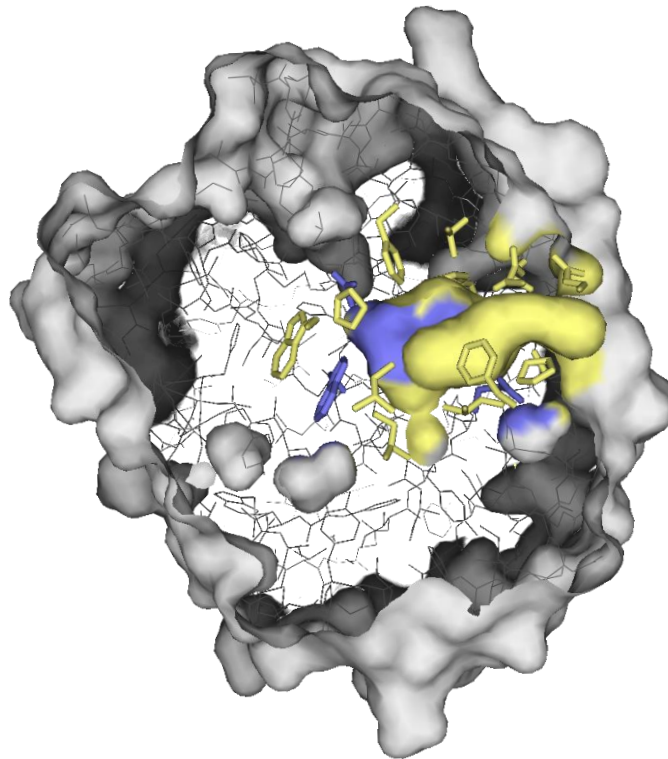
2. residues indispensable for protein function: **catalytic and binding** residues



# HotSpot Wizard



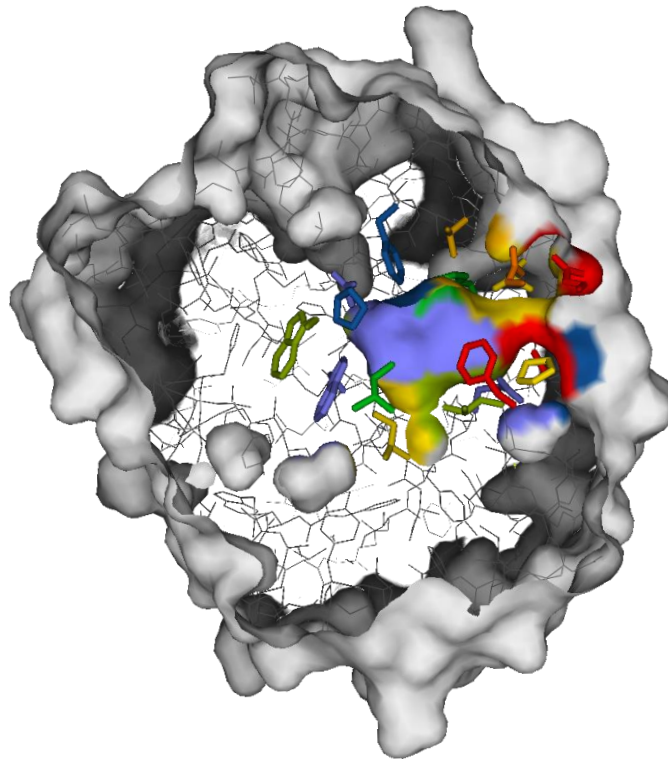
3. functional residues: active site **pocket and tunnels**



# HotSpot Wizard



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





- 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

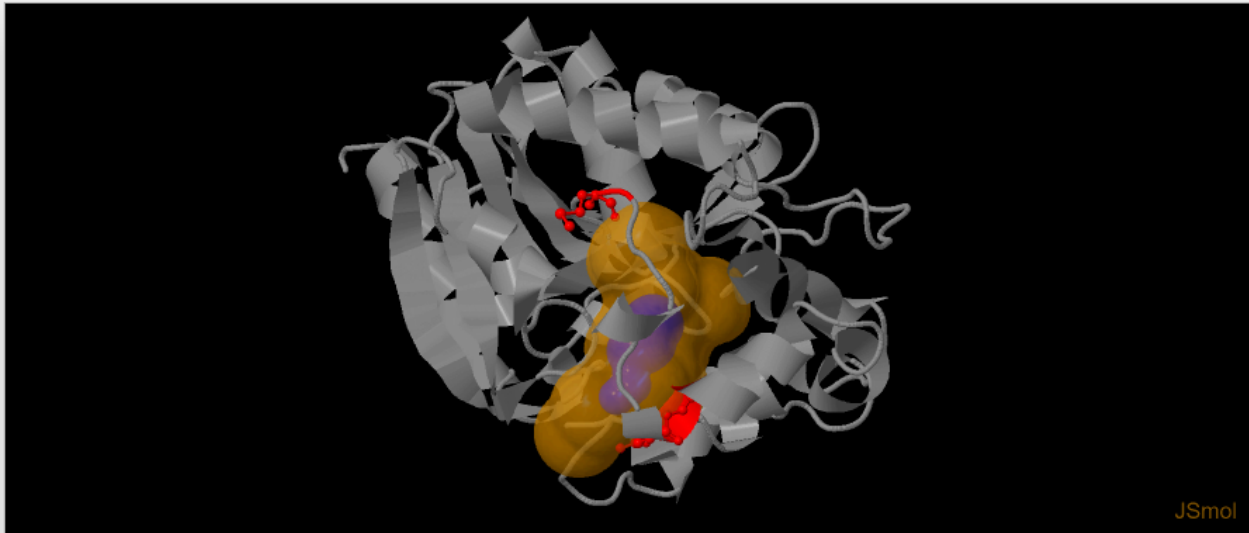
- decisions to be made after evaluation and prioritization of hot-spots:
    - how many and which positions to target?
    - should the positions be randomized simultaneously or separately?
    - should all or only a reduced set of amino acids be introduced at individual positions?
- dramatic effect 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  $\alpha$ -helices and  $\beta$ -sheets  
 Exclude buried residues  
  Include residues with moderate mutability

[Show all residues](#)

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

[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

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

# Design of library – HotSpot Wizard

Library design

Standard SwiftLib

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

<input checked="" type="checkbox"/>	chain	position	residue	desired amino acids	codon	desired ratio (%)	stop ratio (%)
<input checked="" type="checkbox"/>	A	136	Met	Ala, Lys, Pro, Gln, Arg, Thr	VVR	77.8	0.0
<input checked="" type="checkbox"/>	A	146	Gln	Ala, Asp, Glu, Gly, Pro, Gln, Ser	BVV	63.0	11.1
<input checked="" type="checkbox"/>	A	147	Asp	Ala, Phe, Gly, Leu, Met, Thr, Val	DBS	61.1	0.0

Library size : 7315 Codon usage : Escherichia coli K12

Expected coverage : 0.95 Generate report

Probability of full coverage : 0

# Design of library – HotSpot Wizard

Library design

Standard SwiftLib

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

<input checked="" type="checkbox"/>	chain	position	residue	desired amino acids	codon	desired ratio (%)	stop ratio (%)
<input checked="" type="checkbox"/>	A	136	Met	Ala, Lys, Pro, Gln, Arg, Thr	VVR	77.8	0.0
<input checked="" type="checkbox"/>	A	146	Gln	Ala, Asp, Glu, Gly, Pro, Gln, Ser	BVV	63.0	11.1
<input checked="" type="checkbox"/>	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*

- Applications
  - Protein stabilization
  - prediction of protein dynamics
  - prediction of protein-ligand interactions
  - prediction of reaction barriers and reaction mechanisms



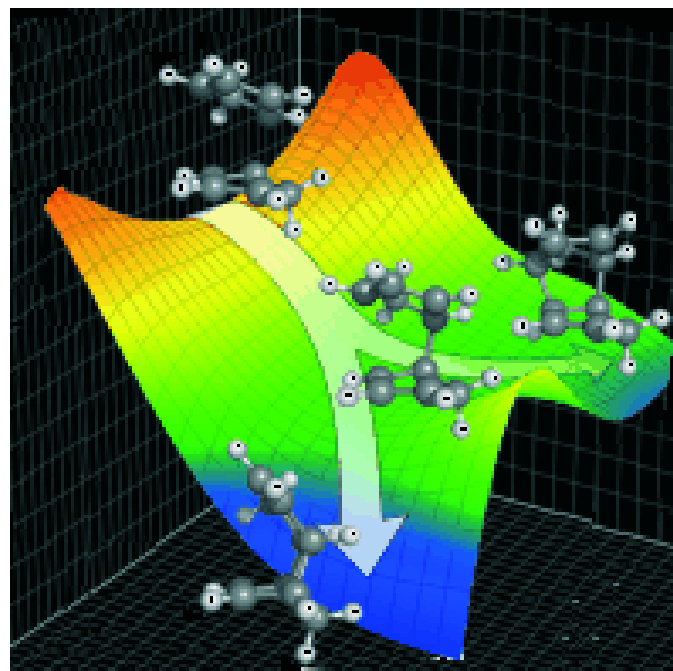
# Molecular modeling

- relationship between energy and 3D-structure

- potential energy surface

- basic methods

- molecular mechanics
- molecular dynamics
- quantum chemistry
- molecular docking



# Design of stability

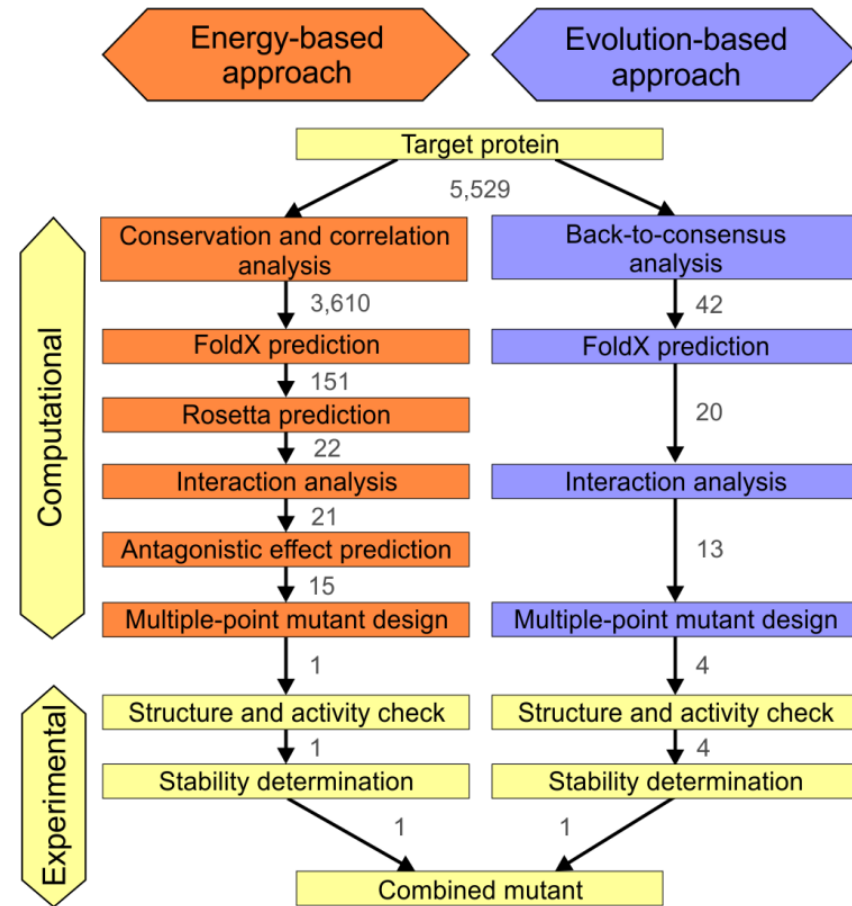


- Enzymes as biocatalysts
  - good activity and selectivity in water solution and standard temperature
  - for many biotechnological applications, high temperature or addition of organic solvents are necessary
  - this conditions can lead to denaturation > importance of stable proteins



# Design of stability

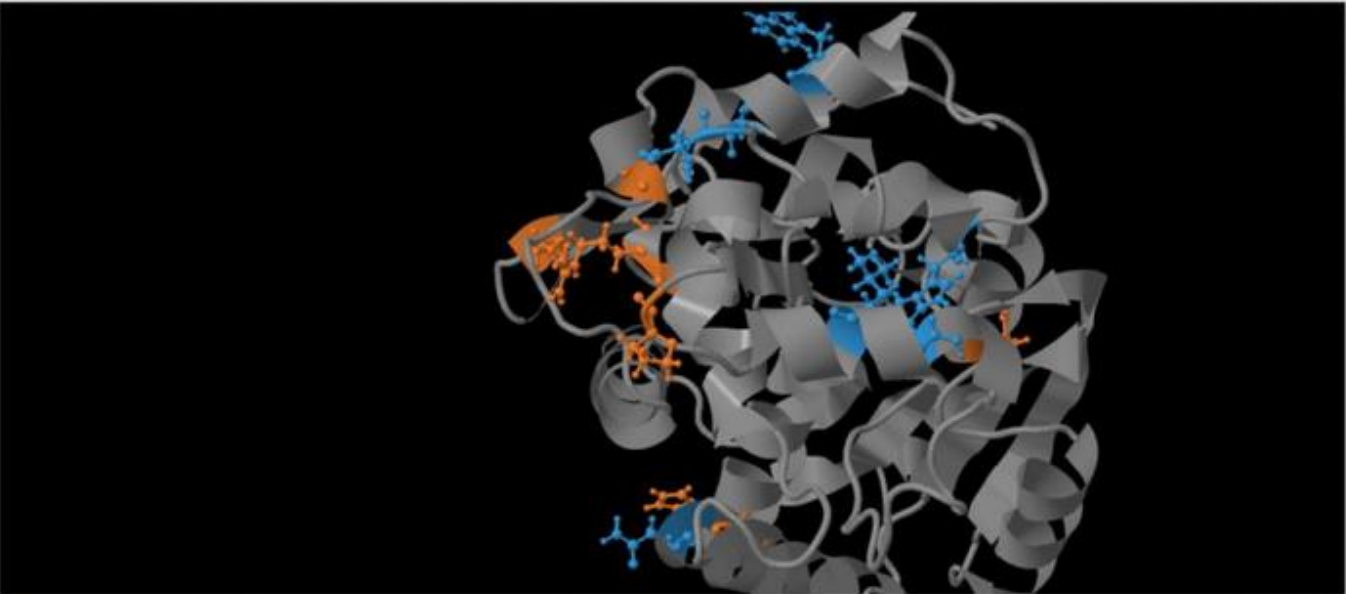
- Computational method FireProt  
<https://loschmidt.chemi.muni.cz/fireprot/>
  - prediction of all single-point mutants by FoldX, Rosetta, and back-to-consensus
  - smart filtering based on conservation, correlation, electrostatic interactions, and antagonistic effect
  - final prediction of multiple-point mutants for gene synthesis



# Design of stability



Viewer



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

FireProt protocol design

PDB ID:	4e46
Length:	292
Evolution mutant:	-3.7 kcal/mol (6 mutations)

## Mutations

Combined mutant Energy mutant Evolution mutant Wild-type

Mutation info					Energy information			Evolution information		
visualize	chain	position	ref	alt	not conserved	not correlated	rosetta	mutable by majority	mutable by ratio	foldx
<input type="checkbox"/>	A	11	D	P	✓	✓	-1.89	✗	✗	-1.39
<input type="checkbox"/>	A	20	E	S	✓	✓	-	✓	✓	0.08
<input type="checkbox"/>	A	33	T	I	✓	✓	-1.94	✗	✗	-1.31

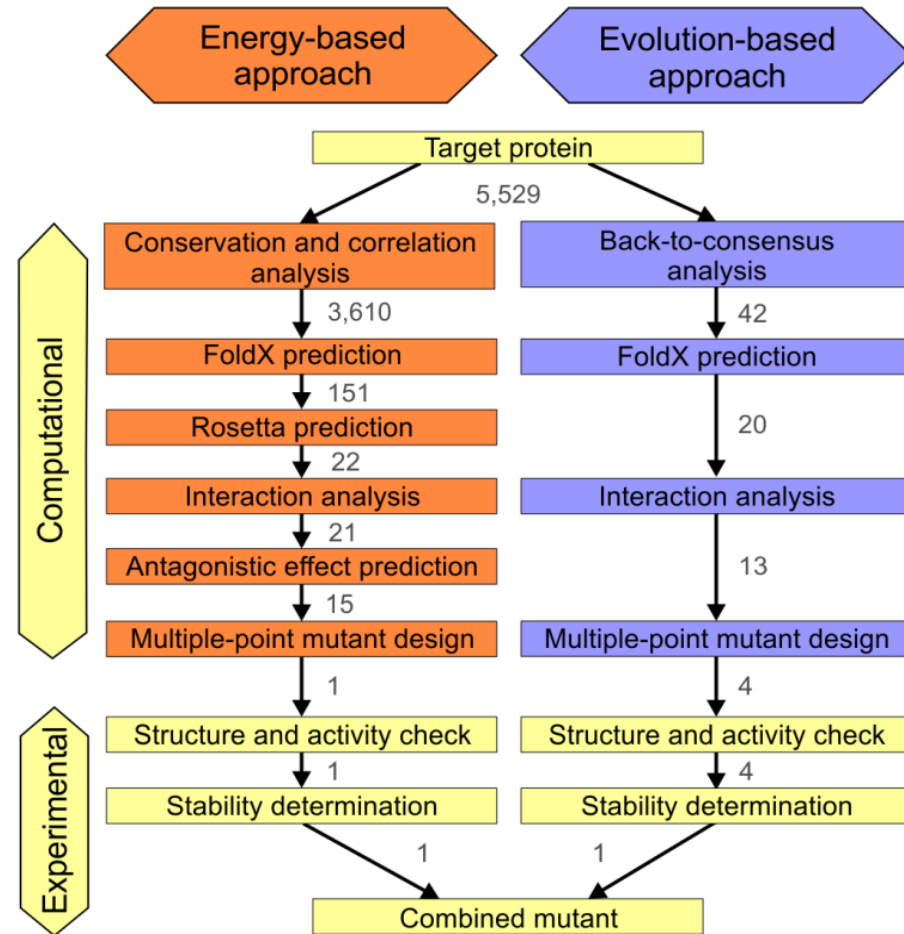
# Design of stability – use case

## □ Stabilization of haloalkane dehalogenase DhaA

- *In silico* prediction of 5,500 mutants
- Experimental testing of 5 mutants

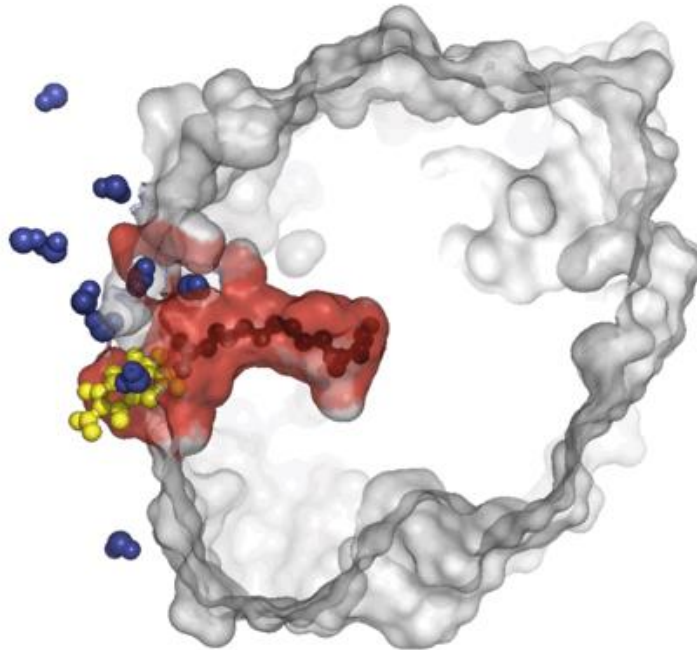
## □ Output

- 3 more stable mutants
- Combined mutant  $\Delta T_m = 24^\circ\text{C}$



# Molecular dynamics

- ❑ successive configurations of system in time
- ❑ provides information on energetics, amplitudes and time scales of local motions on atomic level



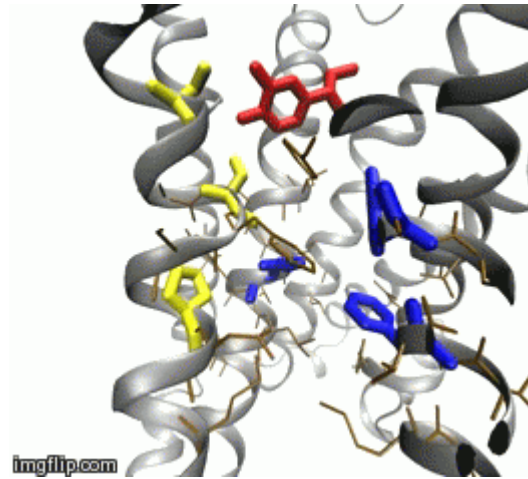
# Molecular dynamics

- generates ensemble of structures
  - more precise calculations of free energies

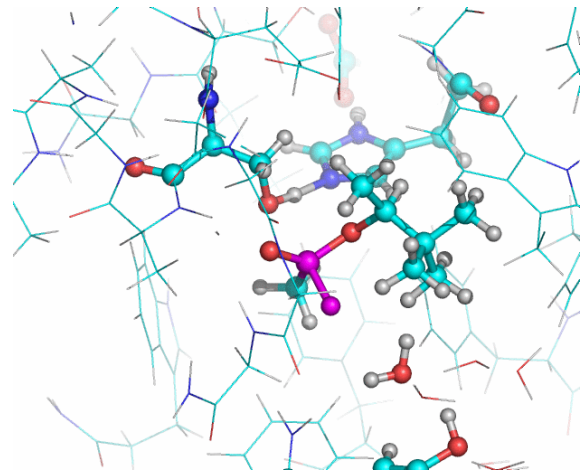




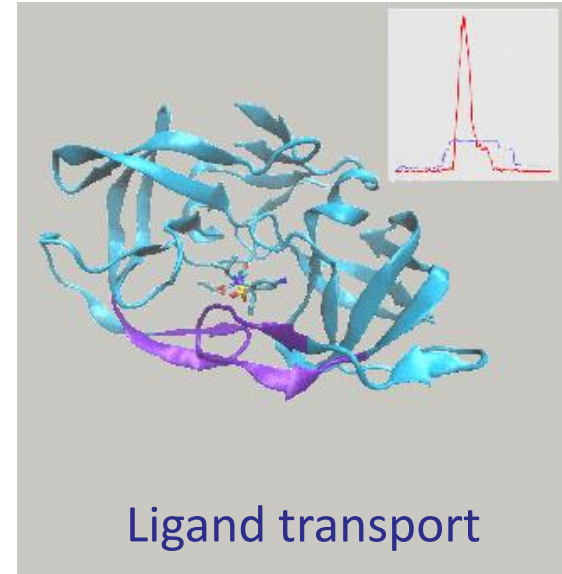
# Molecular dynamics



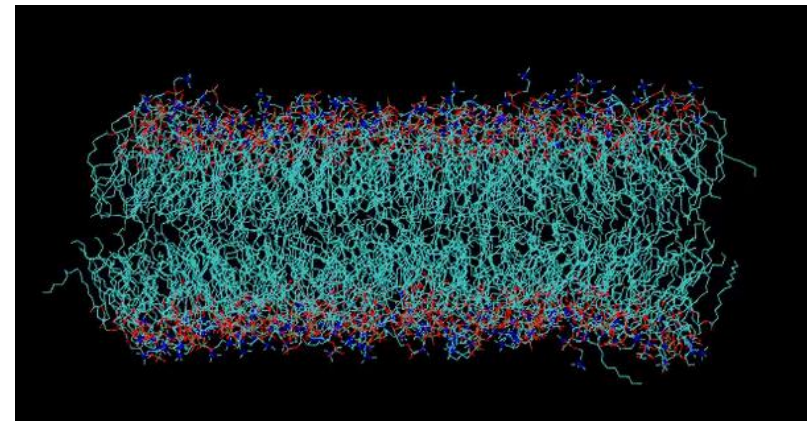
Analysis of interactions



Ligand conversion



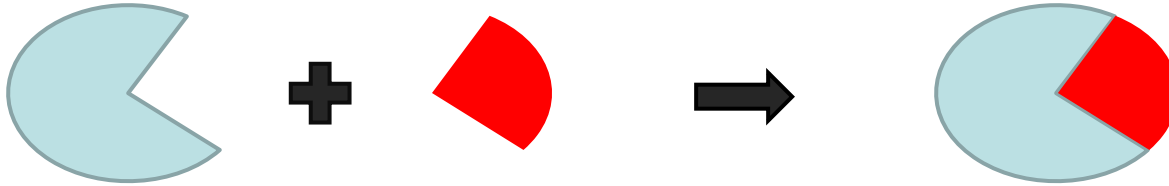
Ligand transport



Dynamical behaviour

# Molecular docking

- predicts structure of receptor (protein) – ligand complex



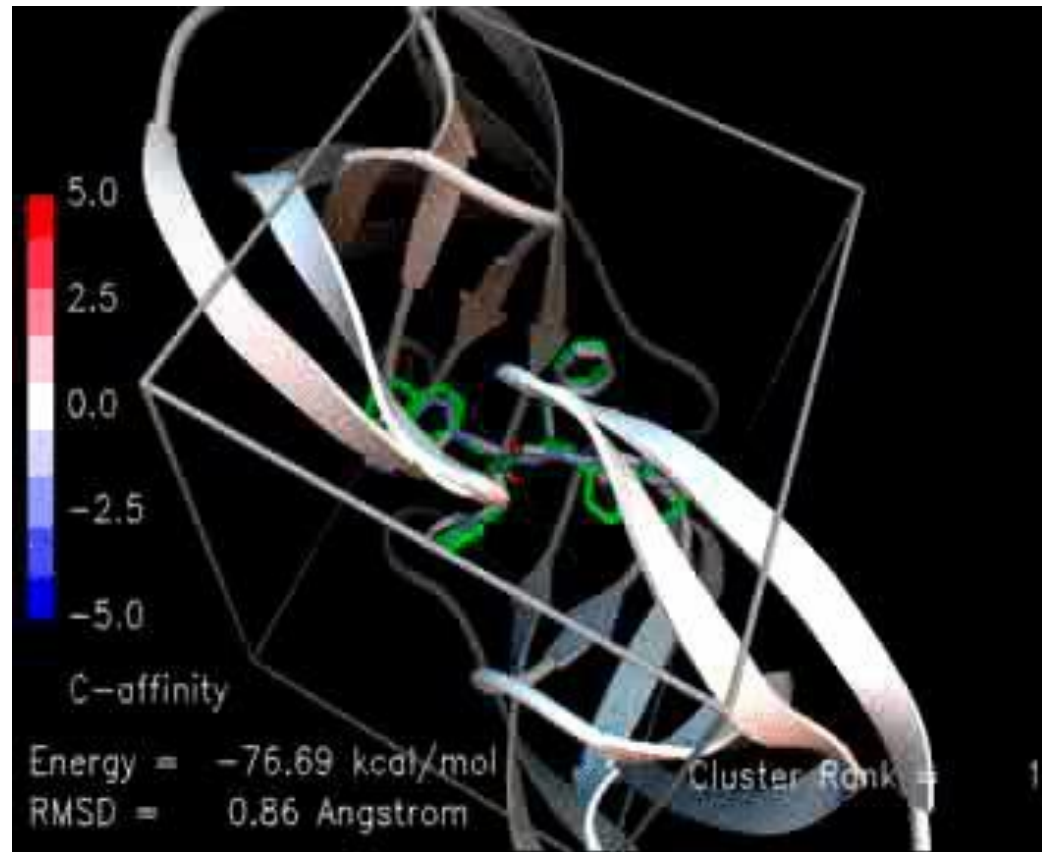
# Molecular docking

## □ 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...

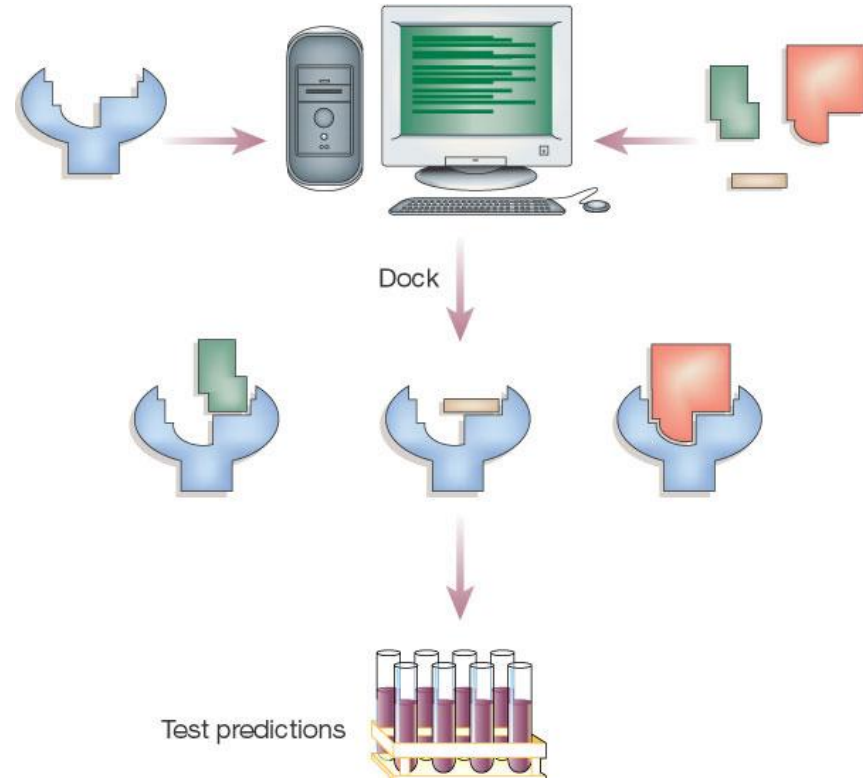




# Molecular docking

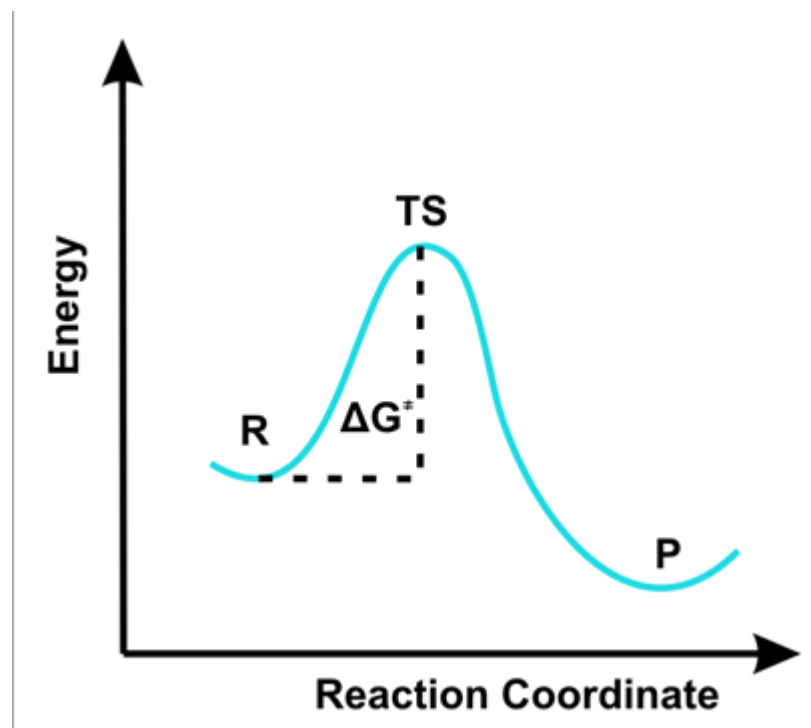
## □ Virtual screening

- many compounds against one enzyme
- one compound against many enzymes



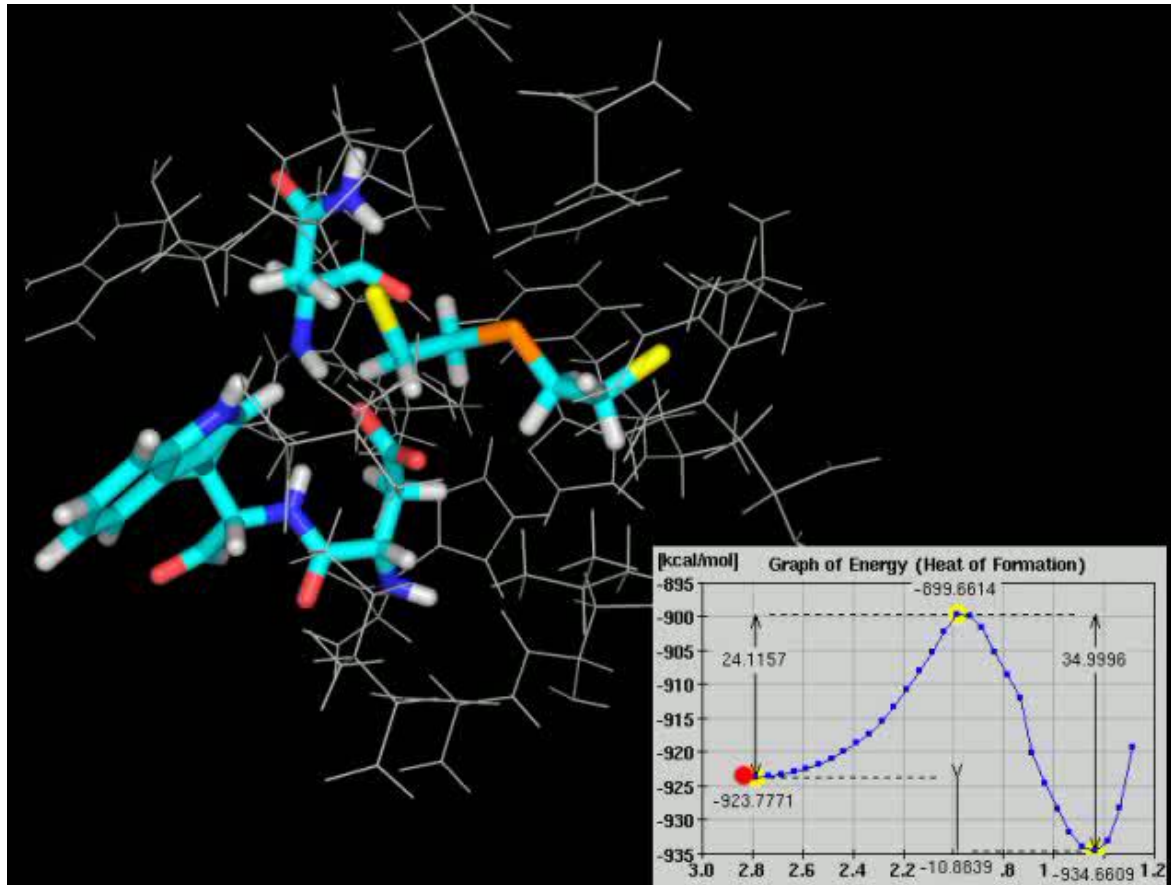
# Quantum chemistry

- modeling of reaction
  - reaction barrier



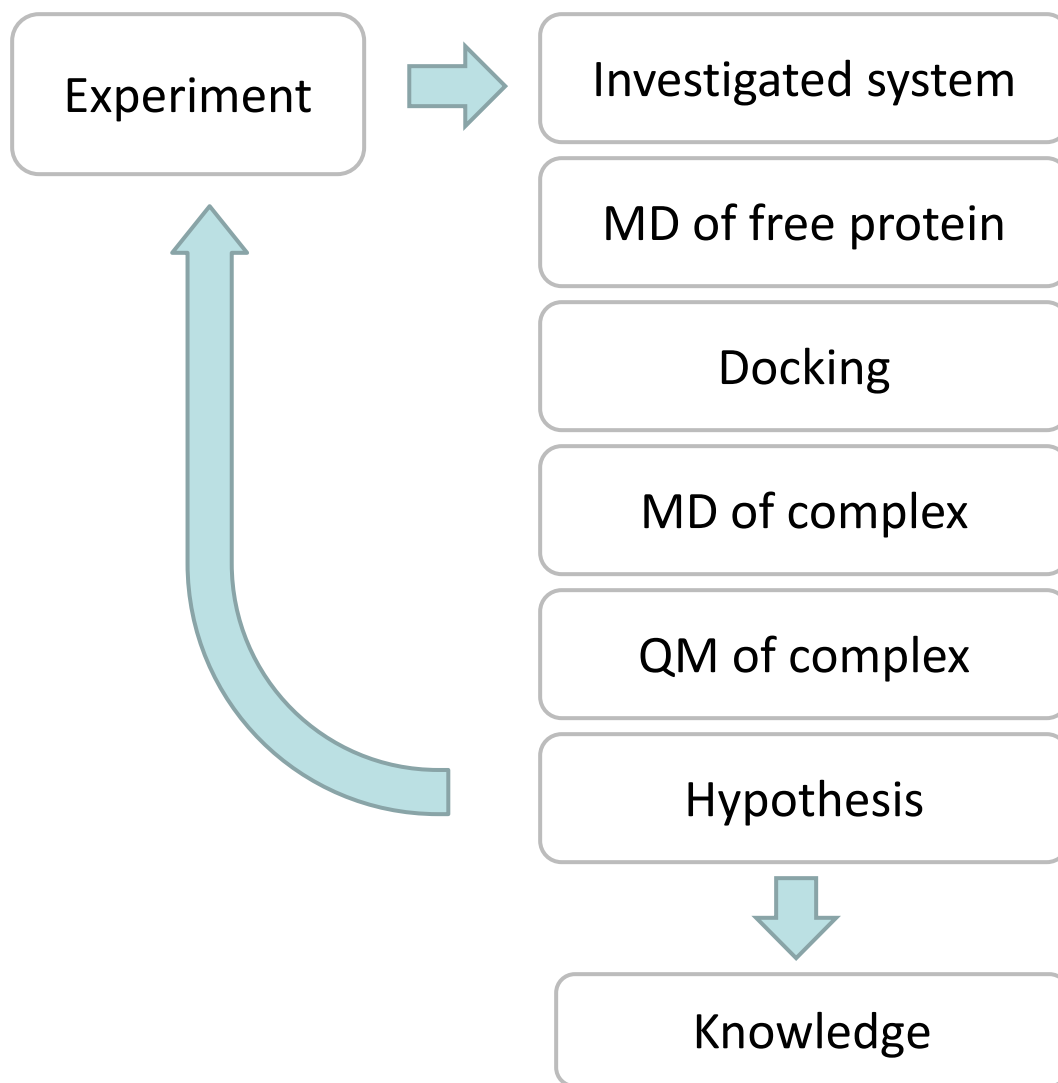
# Quantum chemistry

- modeling of reaction



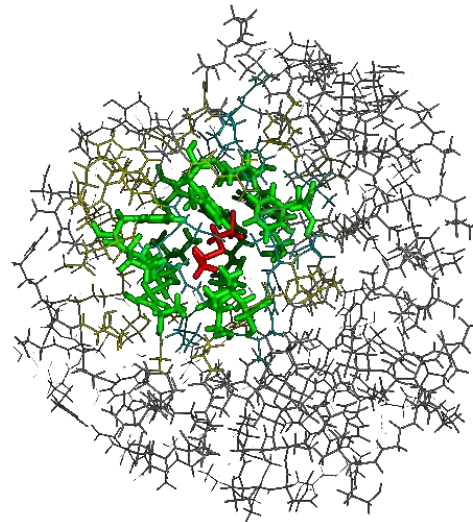
TRITON

# Design of mutations



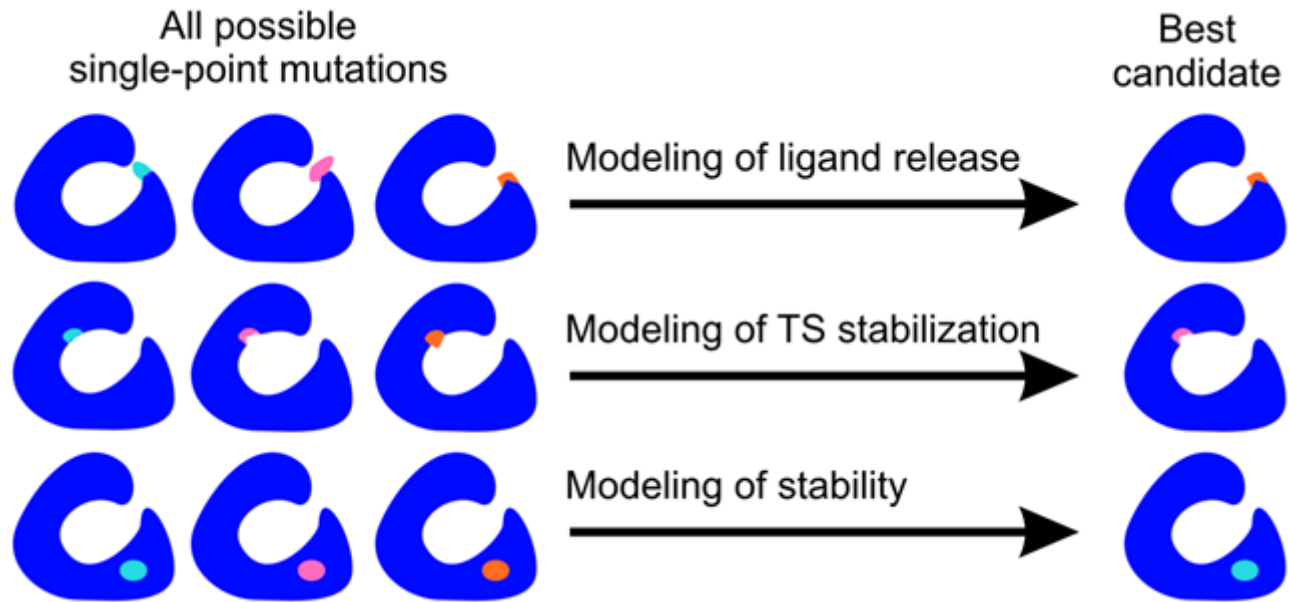
# Design of mutations

- identification of **functionally important residues**
    - decomposition of energies to individual contribution
    - flexible residues – functionally important dynamics
    - residues in contact with ligand
- further molecular modeling
- semi-rational design



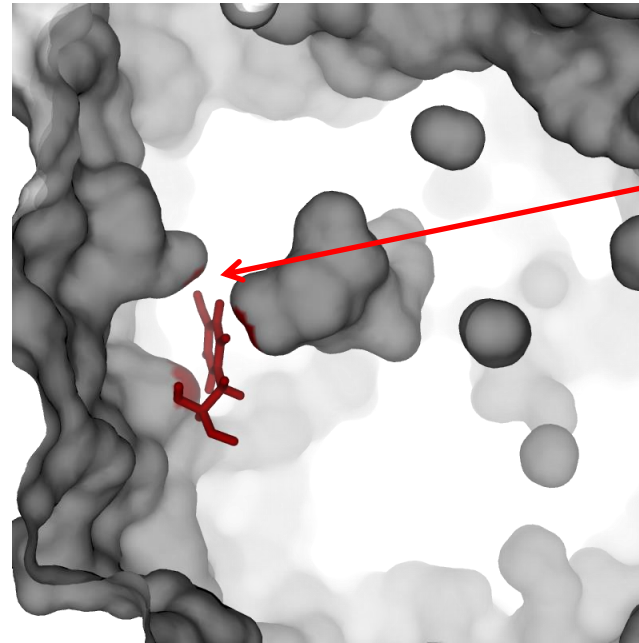
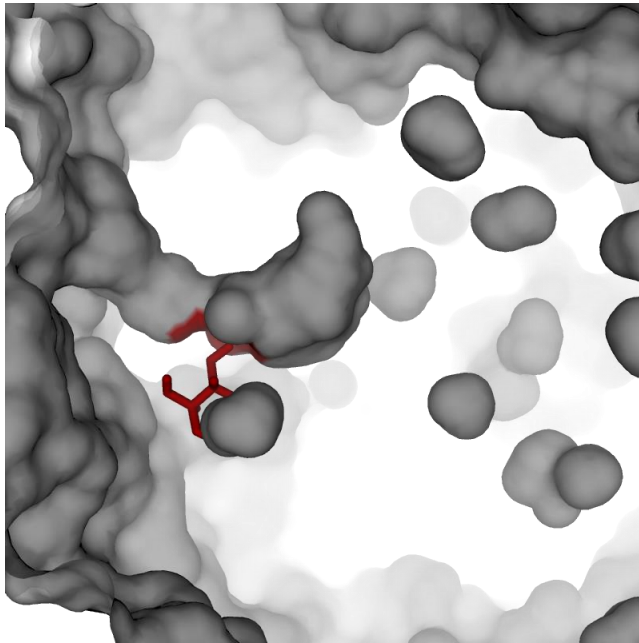
# Design of mutations

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



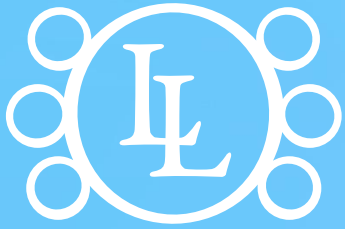
# Design of mutations

- **effect of mutations** at molecular level
  - example: improved activity of tunnel mutant



closed  
tunnel  
+  
improved  
activity





LOSCHMIDT  
LABORATORIES



# PROTEIN ENGINEERING

## 8. Directed evolution

Loschmidt Laboratories

Department of Experimental Biology

Masaryk University, Brno