Applications of structural biology and bioinformatics
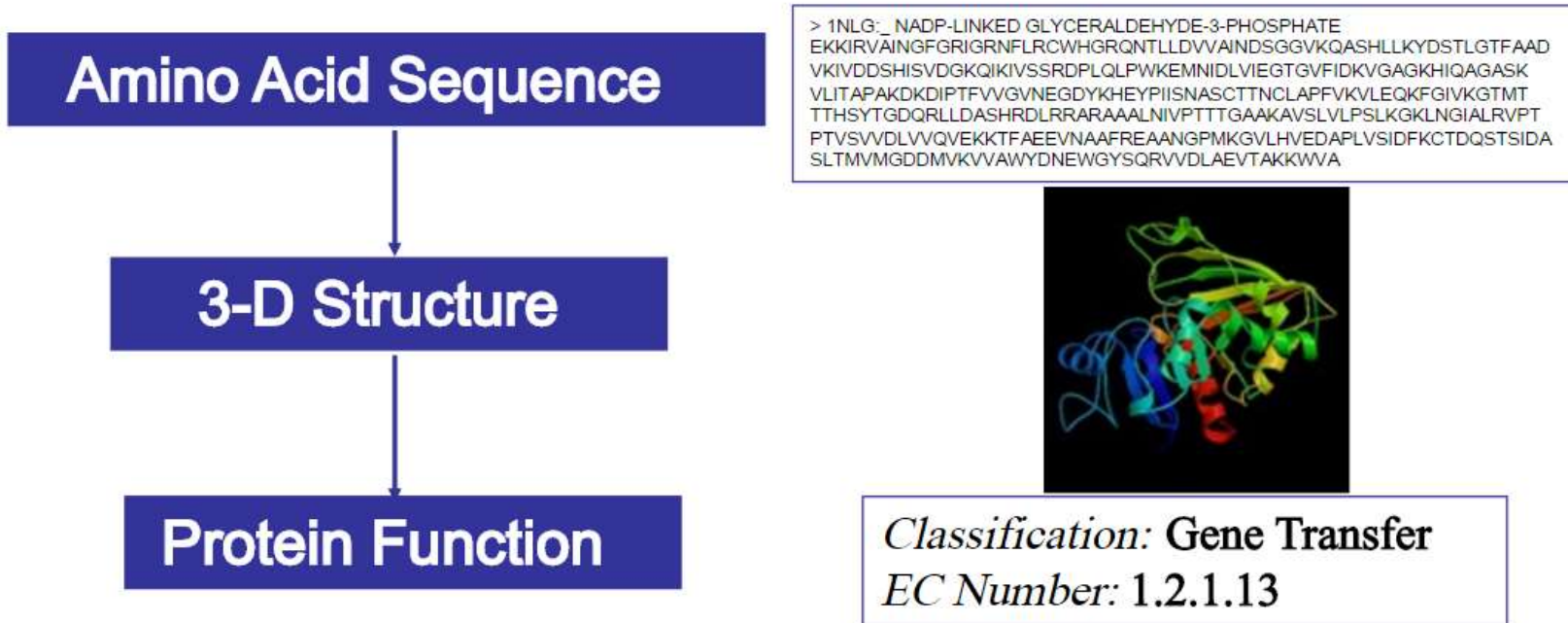
Outline

- Structural biology paradigm

- Applications of structural biology and bioinformatics
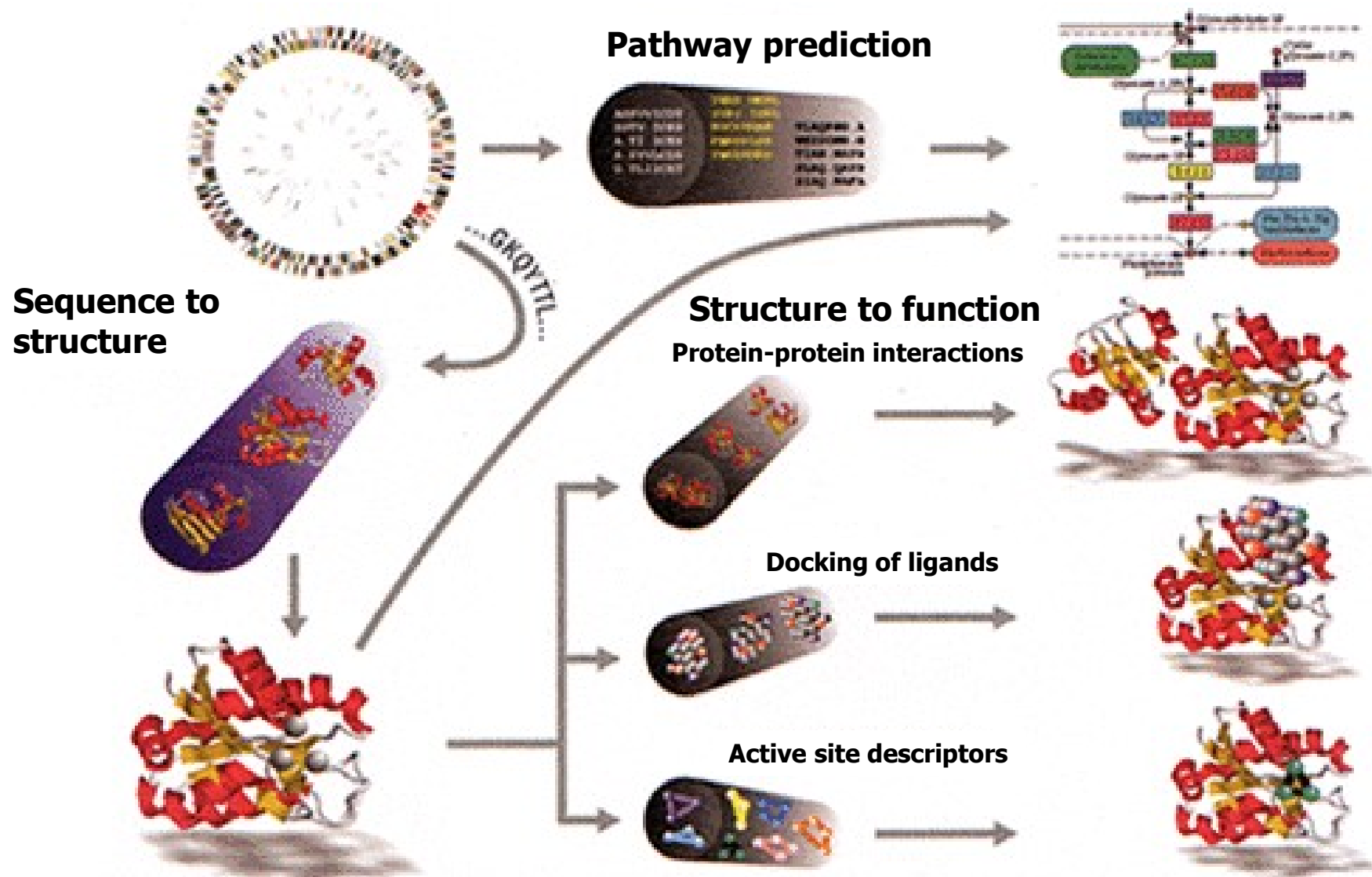
- Summary

- Final remarks on the course

A structural biology paradigm...

- Sequence-to-Structure-to-Function

Amino Acid Sequence

> 1NLG:_ NADP-LINKED GLYCERALDEHYDE-3-PHOSPHATE
EKKIRVAINGFGRIGRNFLRCWHGRQNTLLDVVAINDSGGVKQASHLLKYDSTLGTFAAD
VKIVDDSHISVDGKQIKIVSSRDPLQLPWKEMNIDLVIEGTGVFIDKVGAGKHIQAGASK
VLITAPAKDKDIPTFVVGVNEGDYKHEYPIISNASCTTNCLAPFVKVLEQKFGIVKGTMT
TTHSYTGDQRLLDASHRDLRRARAAALNIVPTTTGAAKAVSLVLPSLKGKLNGIALRVPT
PTVSVVDLVVQVEKKTFAEEVNAAFREAANGPMKGVLHVEDAPLVSIDFKCTDQSTSIDA
SLTMVMGDDMVKVVAWYDNEWGYSQRVVDLAEVTAKKWVA

↓

3-D Structure

↓

Protein Function

Classification: **Gene Transfer**
EC Number: 1.2.1.13

- Computational challenges:
  - Determine structure from sequence
  - Determine function from sequence/3D structure
  - Modify function

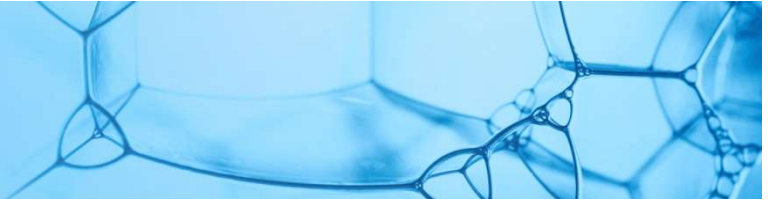# Sequence-to-Structure-to-Function

Pathway prediction

Sequence to structure

...GKQYTL...

Structure to function

Protein-protein interactions

Docking of ligands

Active site descriptors

Applications of structural biology and bioinformatics

- ❑ Biological research

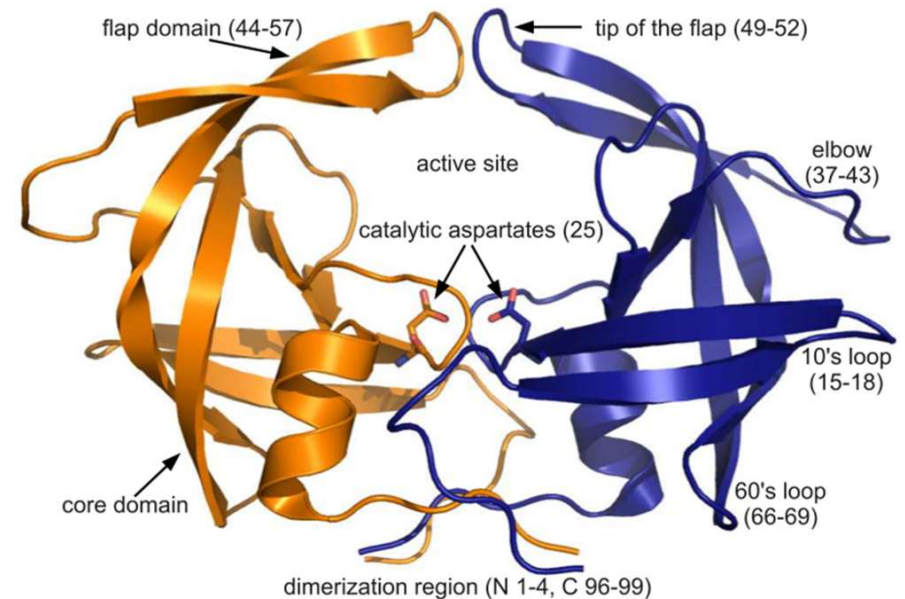- ❑ Drug design

- ❑ Protein engineering

Biological research

- ❑ Drug resistance of HIV protease
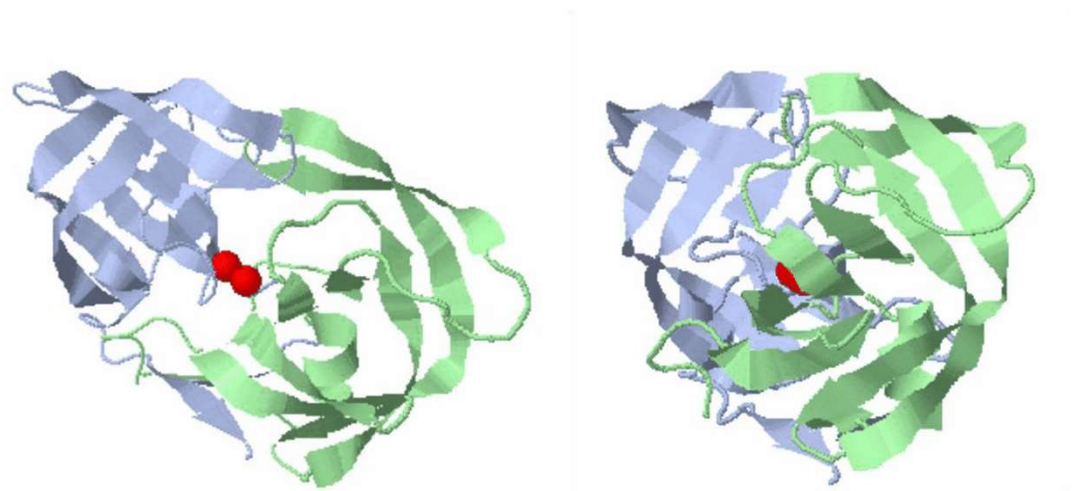
Drug resistance of HIV protease

❑ HIV-1 protease

- plays critical role in viral maturation for producing viral particles

- aspartic protease with characteristic triad Asp-Thr-Gly

- symmetric homodimer, 99 amino acids per monomer

- 3 functionally important regions in the protease structure

  - active site cavity

  - flexible flaps

  - dimer interface

flap domain (44-57) → tip of the flap (49-52)

active site

elbow (37-43)

catalytic aspartates (25)

10's loop (15-18)

60's loop (66-69)

core domain

dimerization region (N 1-4, C 96-99)

Drug resistance of HIV protease

- HIV-1 protease

  - plays critical role in viral maturation for producing viral particles

  - aspartic protease with characteristic triad Asp-Thr-Gly

  - symmetric homodimer, 99 amino acids per monomer

  - 3 functionally important regions in the protease structure

    - active site cavity

    - flexible flaps

    - dimer interface

  - flap opening/closing

    is crucial for catalysis

By comparing 2 crystal structures
(PDBs: 1HXW and 1TW7)

Drug resistance of HIV protease

- ❏ HIV-1 protease

  - ■ plays critical role in viral maturation for producing viral particles

  - ■ aspartic protease with characteristic triad Asp-Thr-Gly

  - ■ symmetric homodimer, 99 amino acids per monomer

  - ■ 3 functionally important regions in the protease structure

    - ■ active site cavity

    - ■ flexible flaps

    - ■ dimer interface

  - ■ flap opening/closing

    is crucial for catalysis

Molecular dynamics
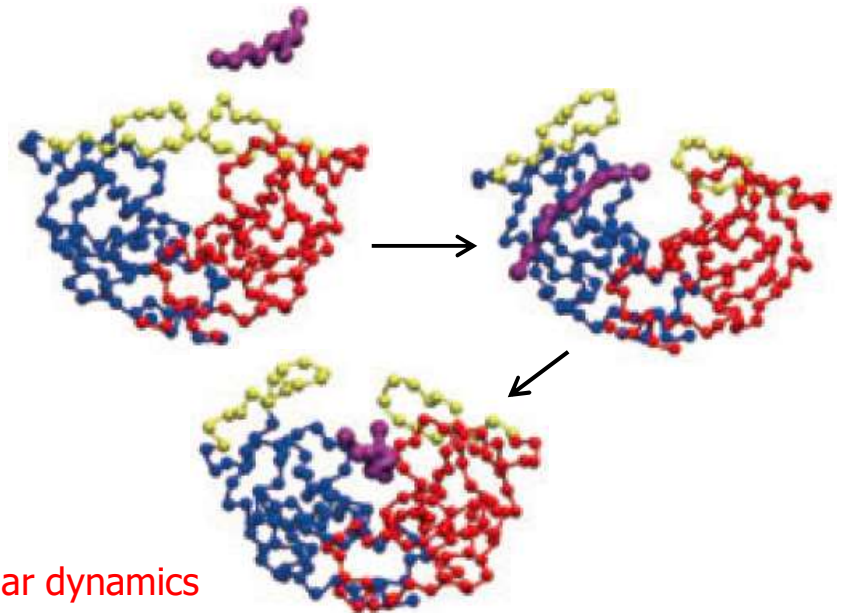
Drug resistance of HIV protease

❑ Protease inhibitors (PIs)

- ▪ introduced into clinical practice in 1995

- ▪ competitive inhibitors, designed to mimic the transition state
  of the substrate-enzyme complex

- ▪ binding affinity of PIs in nanomolar to picomolar range

- ▪ currently about 10 different inhibitors available

  - ▪ ritonavir, lopinavir
  - ▪ fosamprenavir, saquinavir
  - ▪ tipranavir, darunavir
  - ▪ amprenavir, indinavir
  - ▪ nelfinavir, atazanavir
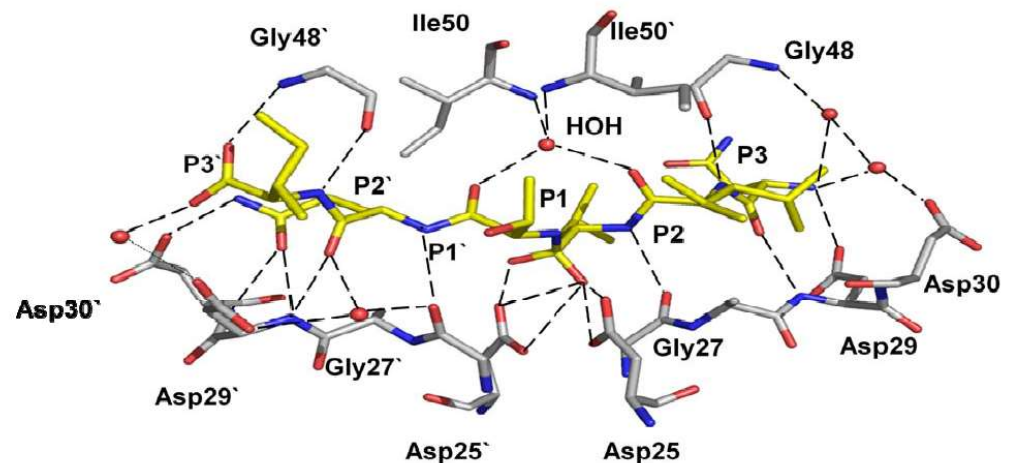
Drug resistance of HIV protease

- ❑ **Drug resistance to PIs**

  - ■ drug resistance emerged against all clinically available PIs

  - ■ resistant mutations in HIV-1 protease reduced susceptibility to inhibitors while maintaining protease function

- ❑ **Important factors in development of drug resistance**

  - ■ rapid mutation

    - ■ high rate of viral replication (10^8-10^9 virions/day)

    - ■ high error rate of HIV reverse transcriptase (≈1 in 10,000 bases)

  - ■ long term exposure to drugs

Drug resistance of HIV protease

❑ Molecular mechanisms of drug resistance

- deduced from comparison of structures and activities of

native and mutant proteases

Drug resistance of HIV protease
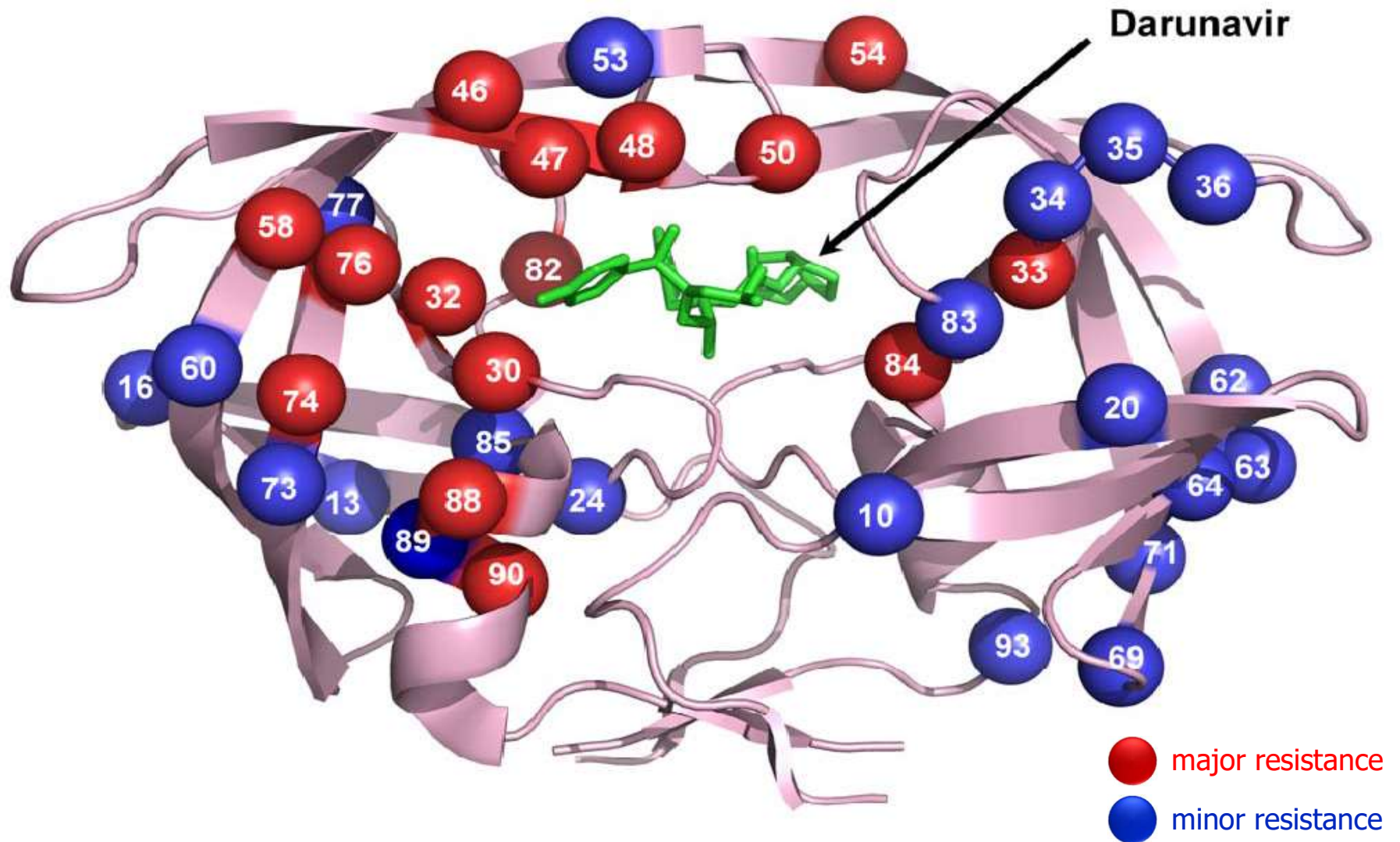
Drug resistance of HIV protease

- Molecular mechanisms of drug resistance

  - deduced from comparison of structures and activities of native and mutant proteases

- Several distinct mechanisms

  - active site mutations

  - mutations at dimer interface

  - mutations at distal positions

Drug resistance of HIV protease

❑ Active site mutations

- mutation of single residue in active site cavity eliminating direct interactions with inhibitor

- mutations are very conservative – e.g. substitutions of hydrophobic amino acids
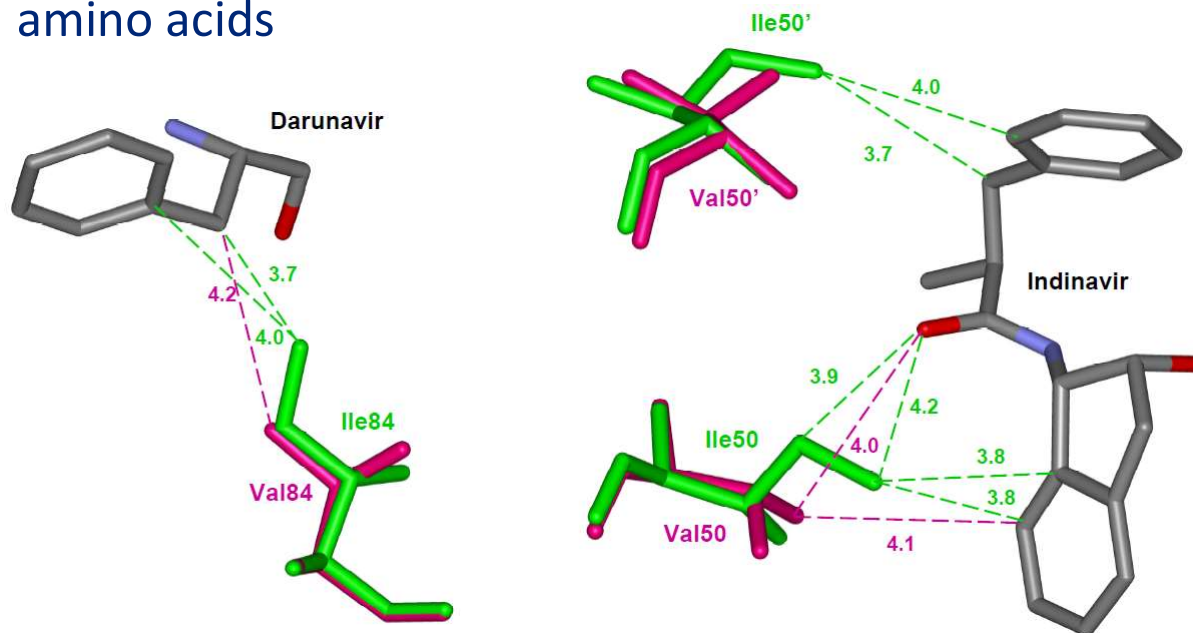
Drug resistance of HIV protease

- ❑ Mutations at dimer interface

  - ▪ for example Phe53Leu

    - ▪ wider separation of the two flaps

    - ▪ reduced stabilization of bound inhibitor
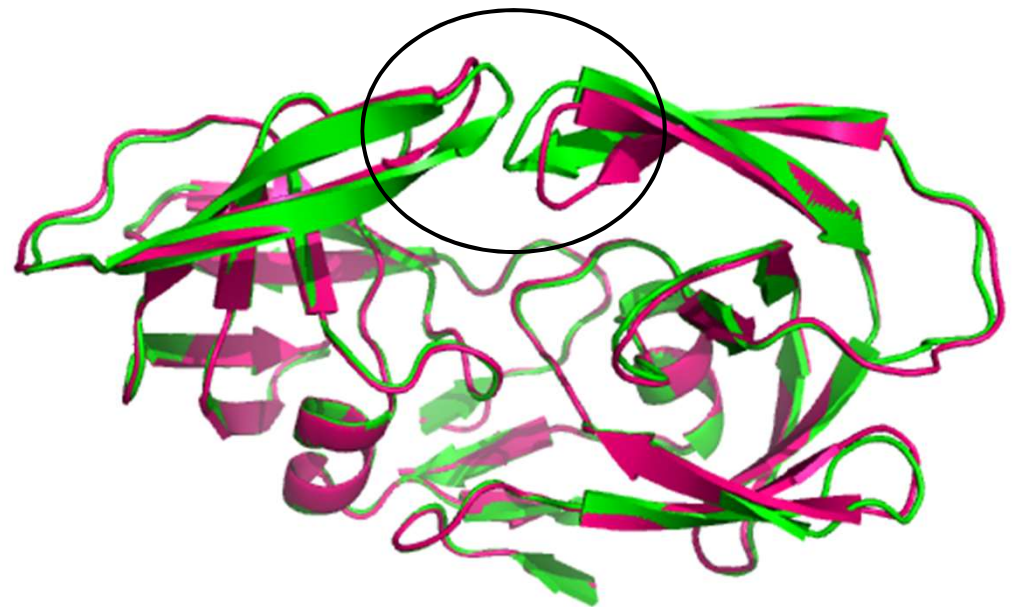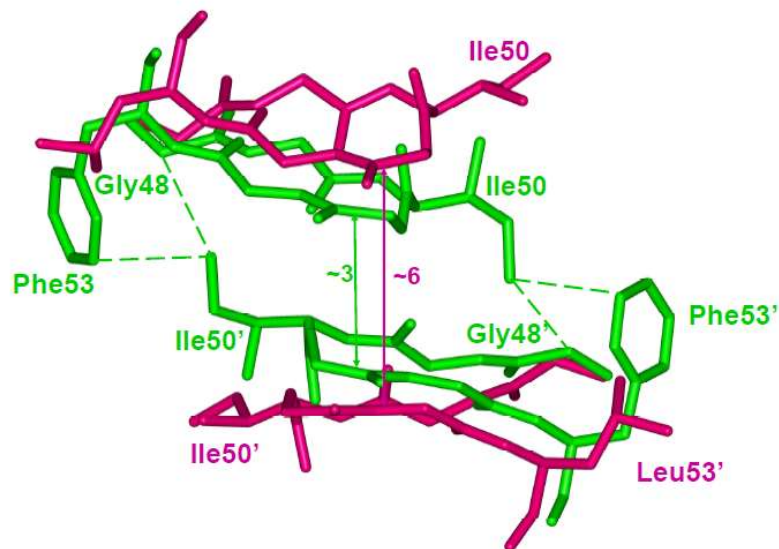
Drug resistance of HIV protease

- Mutations at distal positions

  - for example Leu90Met

    - promoted contacts with catalytic Asp25

    - reduced interaction with inhibitor
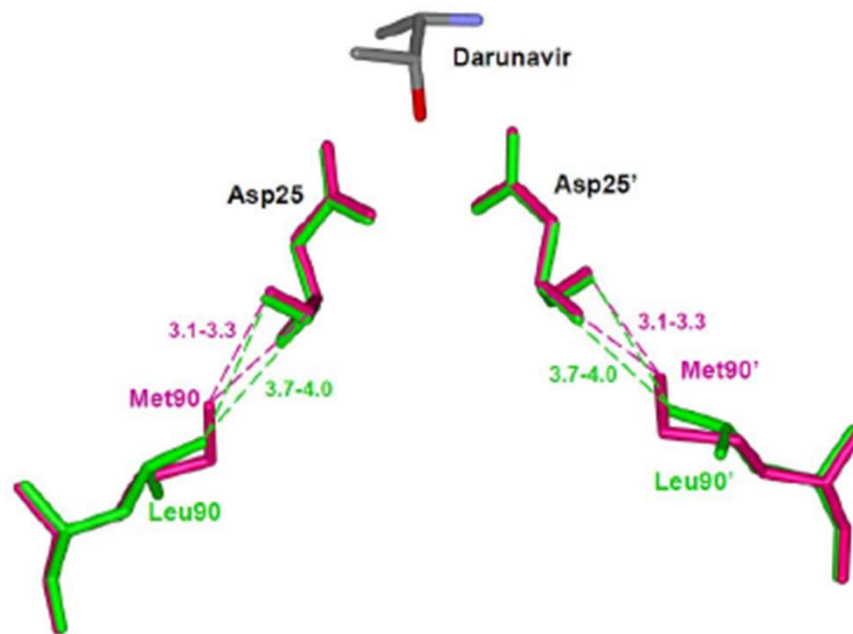
Drug resistance of HIV protease

- Novel PIs for resistant HIV-1 protease

  - inhibitors fitting within envelope formed by bound substrate

  - inhibitors binding flaps or the dimer interface

  - inhibitors targeting main chain and conserved regions of active site

Drug resistance of HIV protease

- ❑ Novel PIs for resistant HIV-1 protease

  - ▪ inhibitors targeting the gating mechanism

    - ○ stabilize the closed state

    - ○ stabilize the open state

    - ○ mixed interactions
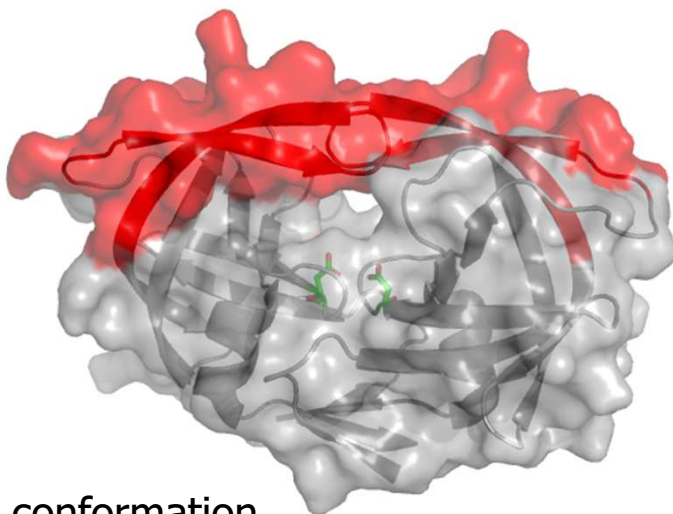
Closed conformation
(PDB ID: 1HVR)
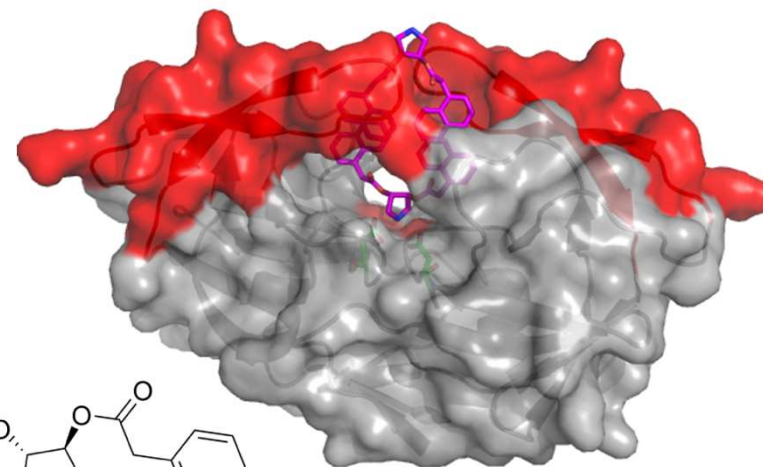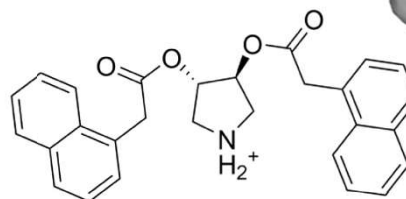
Open conformation
(PDB ID: 3BC4)

Drug design

- Virtual screening of inhibitors of endonuclease MUS81

- Selective inhibitor of LTA4H

Drug design

- ❑ Methods of drug discovery

  - ▪ ligand-based

    - ▪ knowledge of active ligands

    - ▪ search for similar ones
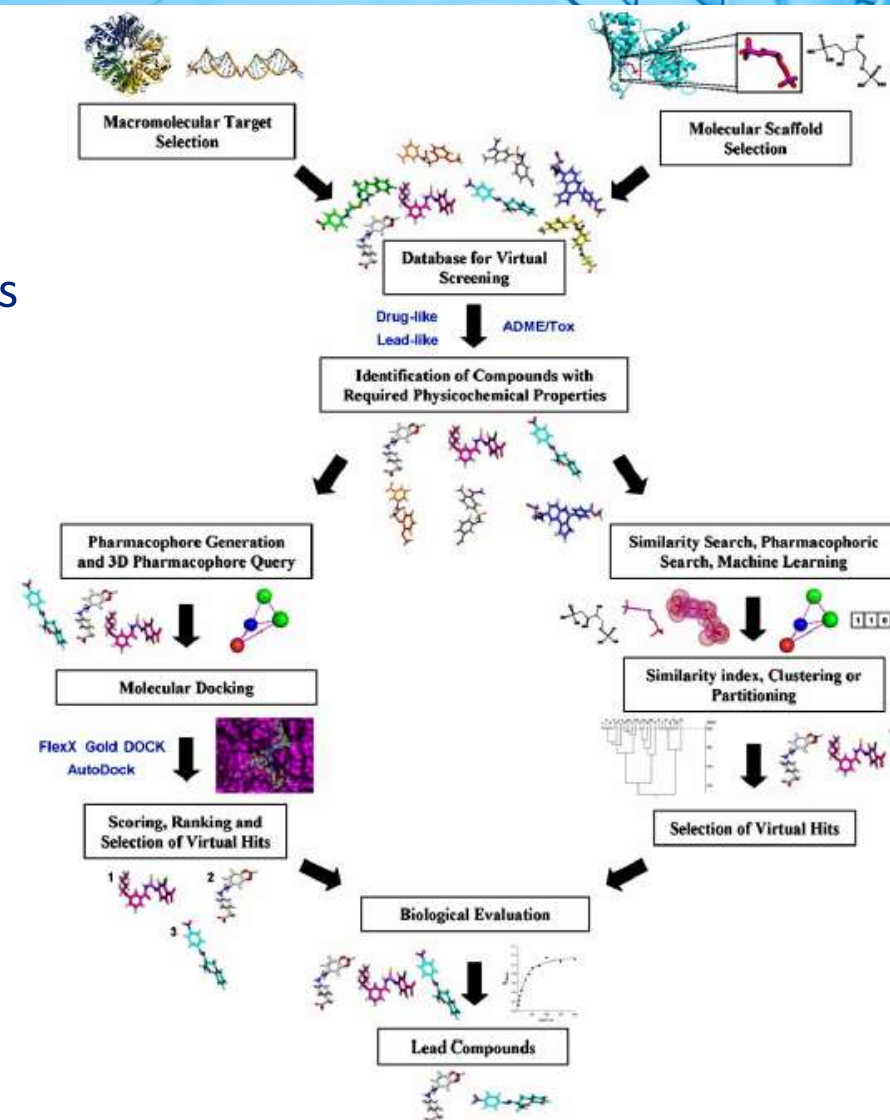
Drug design

❑ Methods of drug discovery

- ligand-based
  - knowledge of active ligands
  - search for similar ones
- structure-based
  - knowledge of receptor
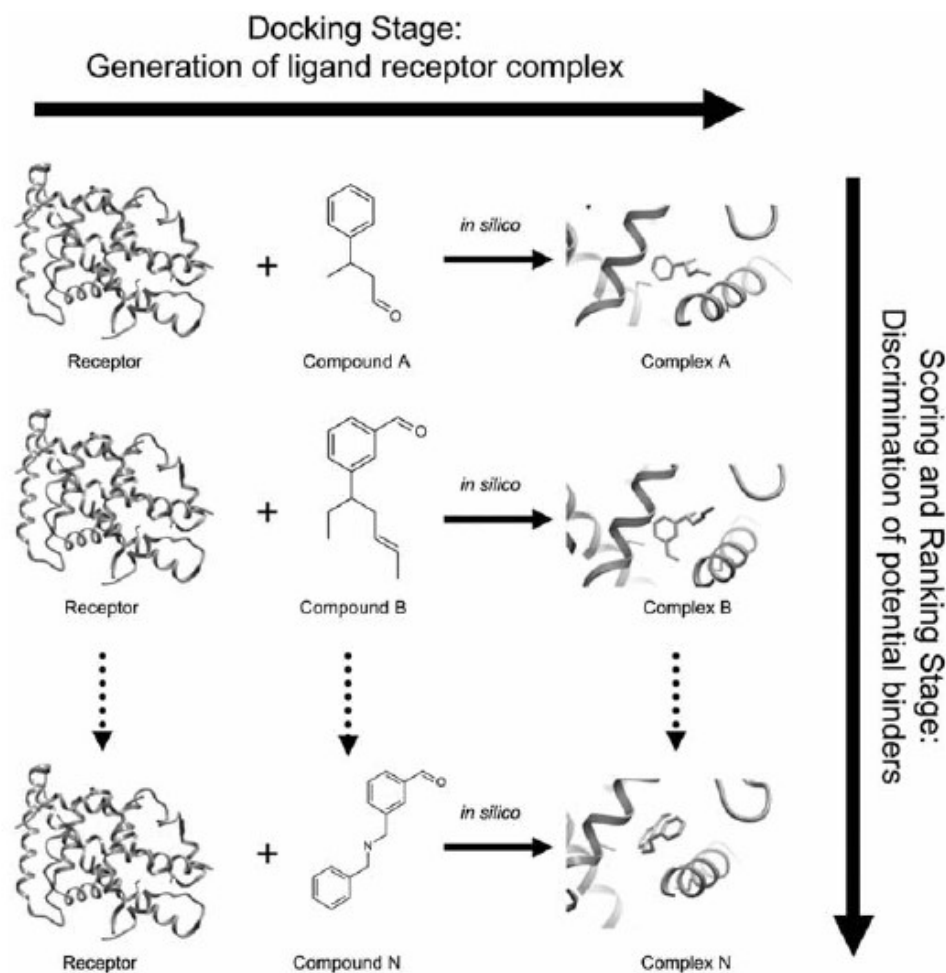 - search for tight binders
  - molecular docking
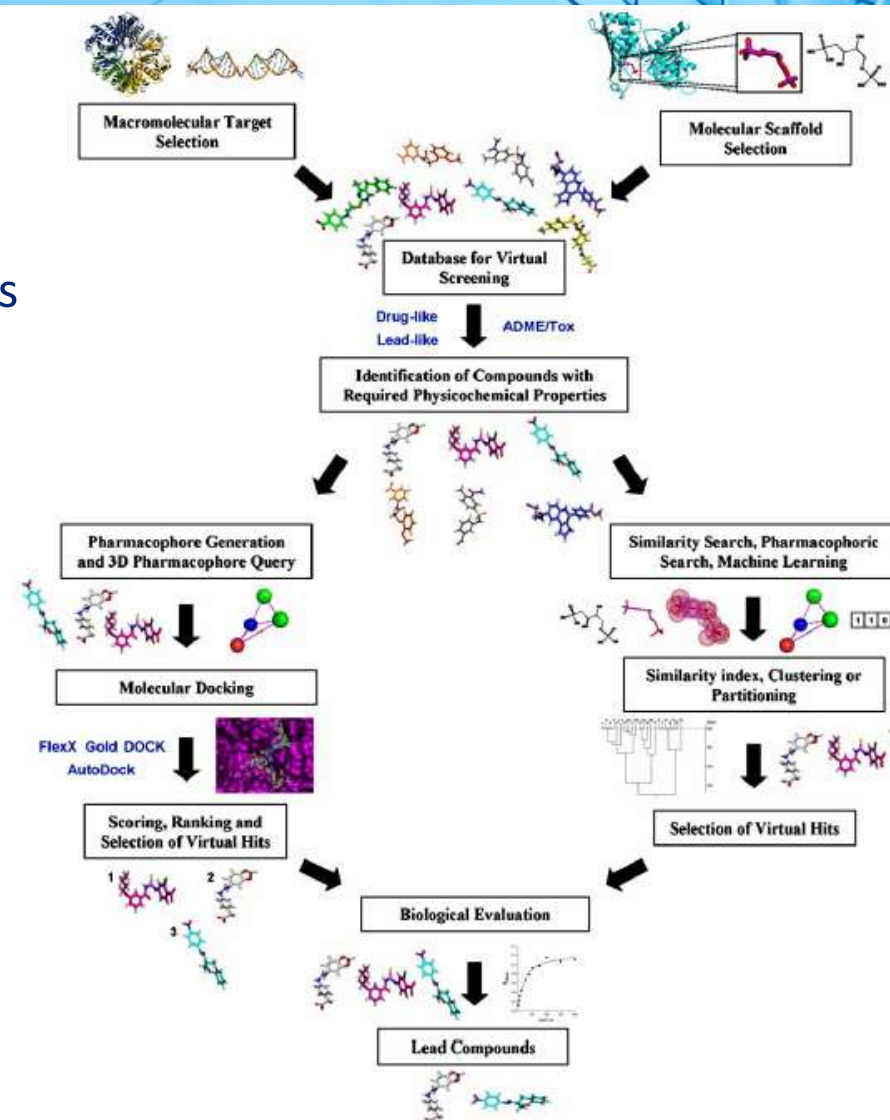
Drug design

- ❑ Methods of drug discovery

  - ▪ ligand-based

    - ▪ knowledge of active ligands
    - ▪ search for similar ones

  - ▪ structure-based

    - ▪ knowledge of receptor
    - ▪ search for tight binders
    - ▪ molecular docking

  - ▪ high-throughput screening (HTS)

    - ▪ large library of compounds
    - ▪ *in silico* or experimental screening
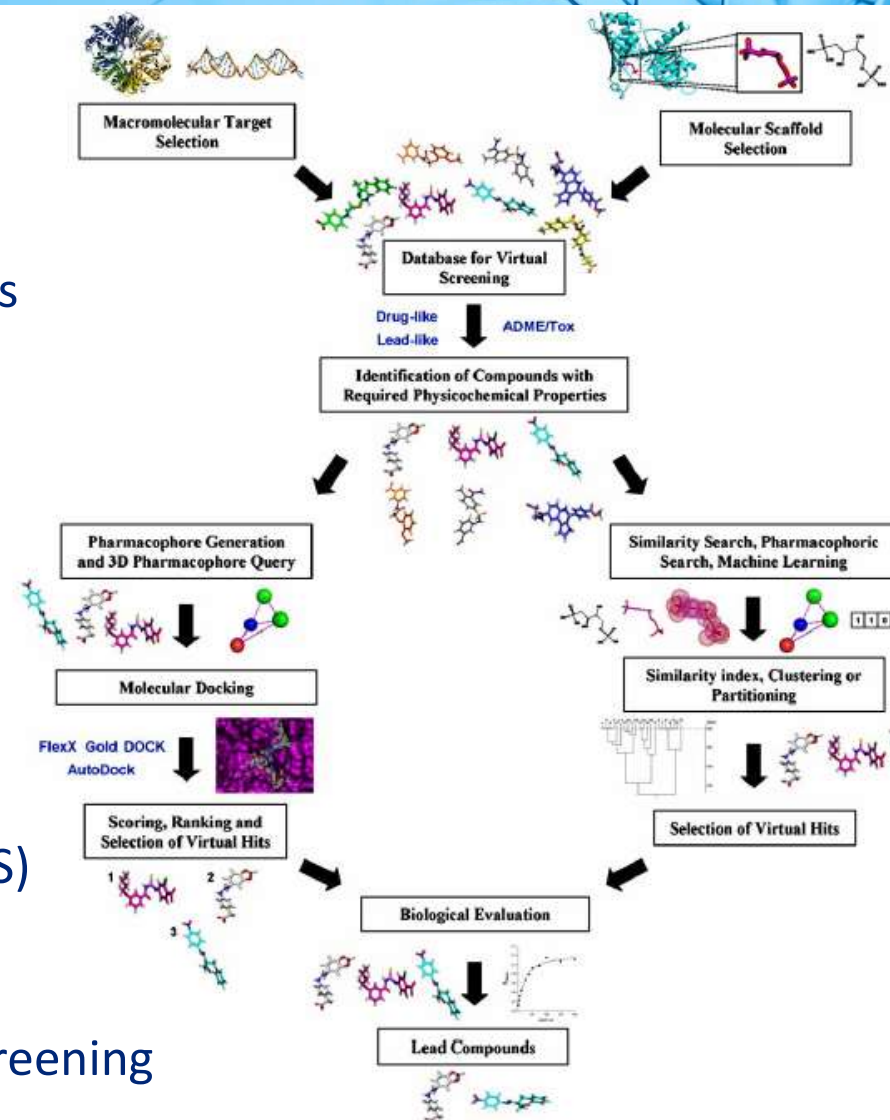
Virtual screening

□ Structure-based VS

- receptor-ligand docking

- often combined with HTS

- followed by hit optimization

- many success stories

- speed-up drug discovery

- lowering expenses

Docking Stage:
Generation of ligand receptor complex

Receptor + Compound A → *in silico* → Complex A

Receptor + Compound B → *in silico* → Complex B

Receptor + Compound N → *in silico* → Complex N

Scoring and Ranking Stage:
Discrimination of potential binders

Virtual screening

Drug Target	Disease Target or Function	Receptor	SBVS Method	Comment	Potency of Lead Scaffold (IC$_{50}$)
EGFR	Cancer	X-Ray	ICM	First SBVS to EGFR crystal structure	10 μM
Casein Kinase 2	Prostate Cancer	X-Ray	MOE, GLIDE, FRED and GOLD	Multiple docking algorithms and consensus scoring	20 nM
β-Secretase	Alzheimers	X-Ray	SEED	Fragment-based	10 μM
DPP-IV	Diabetes	X-Ray	FlexX	Fragment-based	3-70 μM
SARS-CoV	SARS	X-Ray	EUDOC	Receptor Ensemble Docking approach	23 μM
SHBG	Endometrial cancer, ovarian dysfunction, male and female infertility osteoporosis and diabetes	X-Ray	GLIDE	Ligand-based and structure-based	13-124 μM
SARS-CoV	SARS	Model	DOCK 4.01	Screened NCI, ACD, MDDR + consensus scoring	K_i = 61-178 μM
L-xylulose reductase	Diabetes	X-Ray	DOCK 4.01	Screened NCI database	29-100 μM
HSP 90	Cancer	X-Ray	RDOCK	Post VS crystal structure provides rationale to docking results	0.6-26 μM
ER-β	Alzheimers	X-Ray	GOLD 2.0	25000 plant based ligands	680 nM

Inhibitors of endonuclease MUS81

- DNA structure-specific endonuclease MUS81

  - involved in DNA reparation

  - helps maintaining genomic stability

  - target for anti-cancer drug development
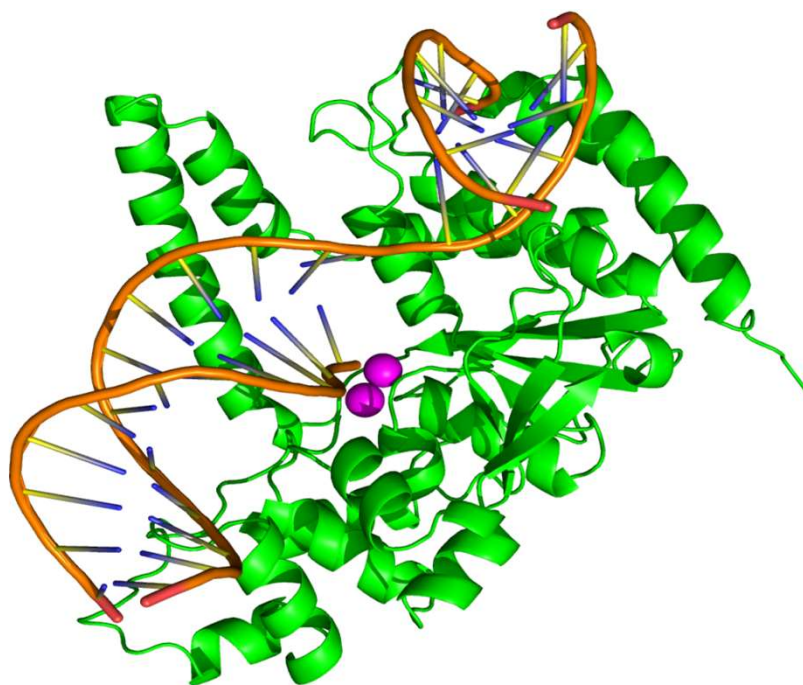
Inhibitors of endonuclease MUS81

❑ High-throughput screening (HTS)

- robotic platform at Center of Chemical Genetics, ASCR, Prague

- about 23,000 compounds experimentally tested

- identified 1 effective inhibitor: $IC_{50} = 50$ µM

Inhibitors of endonuclease MUS81

- Structure-based VS

  - molecular docking + rescoring of binding interaction

  - binding of more than 140,000 compounds predicted

  - experimental verification on 19 potential inhibitors
  - identified 6 effective inhibitors with $IC_{50} \leq 50$ µM
  - best inhibitor: $IC_{50} = 5$ µM
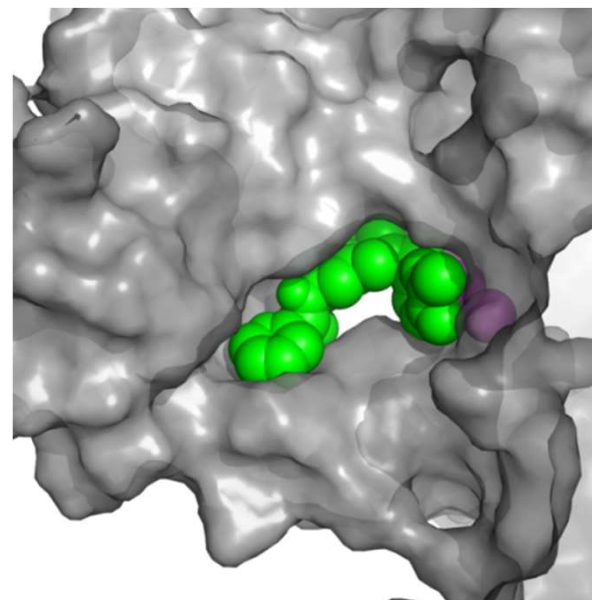
Inhibitors of endonuclease MUS81

❑ Comparison

	HTS	VS
Equipment (Kč)	50,000,000	500,000
Testing		
Computational	-	140,000
Experimental	23,000	19
Costs (Kč)	2,000,000	40,000
Time	Weeks	Days
Results		
# of inhibitors	1	6
Best: IC_{50} (μM)	50	5

Selective inhibitor of LTA4H

❑ Leukotriene A4 hydrolase/aminopeptidase (LTA4H)

- Involved in chronic inflammatory diseases

- Bifunctional metalloenzyme

  - Catalyzes hydrolysis of the leukotriene A4 (LTA4) into the pro-inflammatory mediator LTB4

  - Also hydrolyses the pro-inflammatory Pro-Gly-Pro

  - Distinct but overlapping binding sites and 2 tunnels
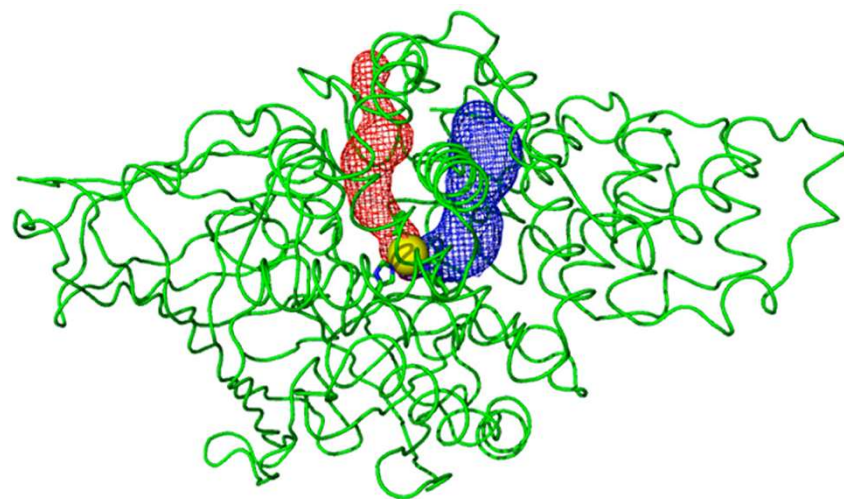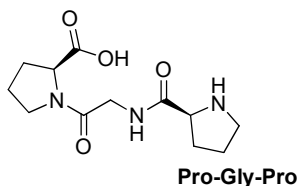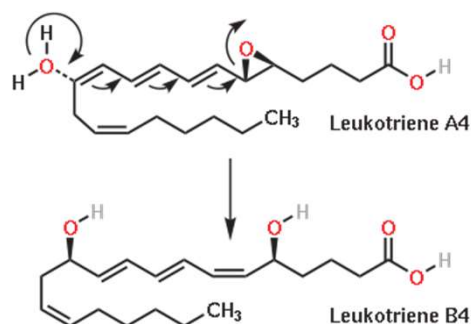
Leukotriene A4

Leukotriene B4

Pro-Gly-Pro

Selective inhibitor of LTA4H

- ❑ Leukotriene A4 hydrolase/aminopeptidase (LTA4H)

  - ▪ Structural studies (crystallography) with a tripeptide analogue revealed the aminopeptidase mechanism
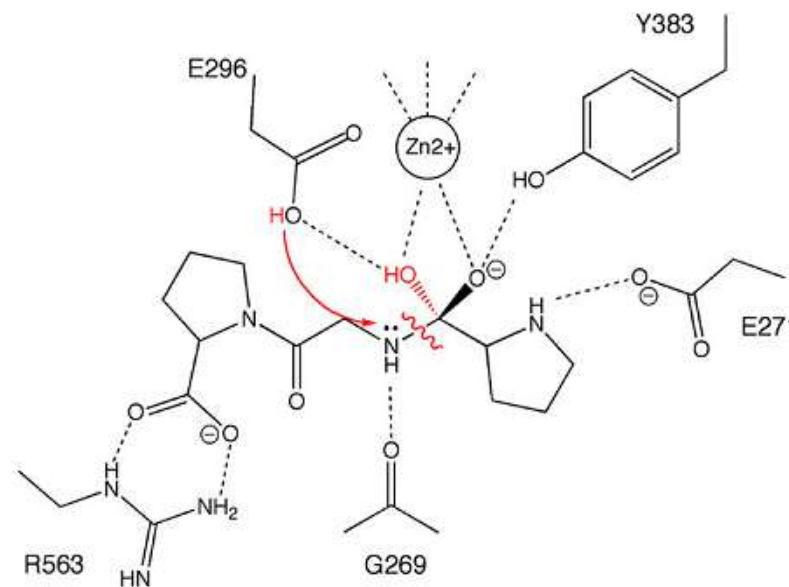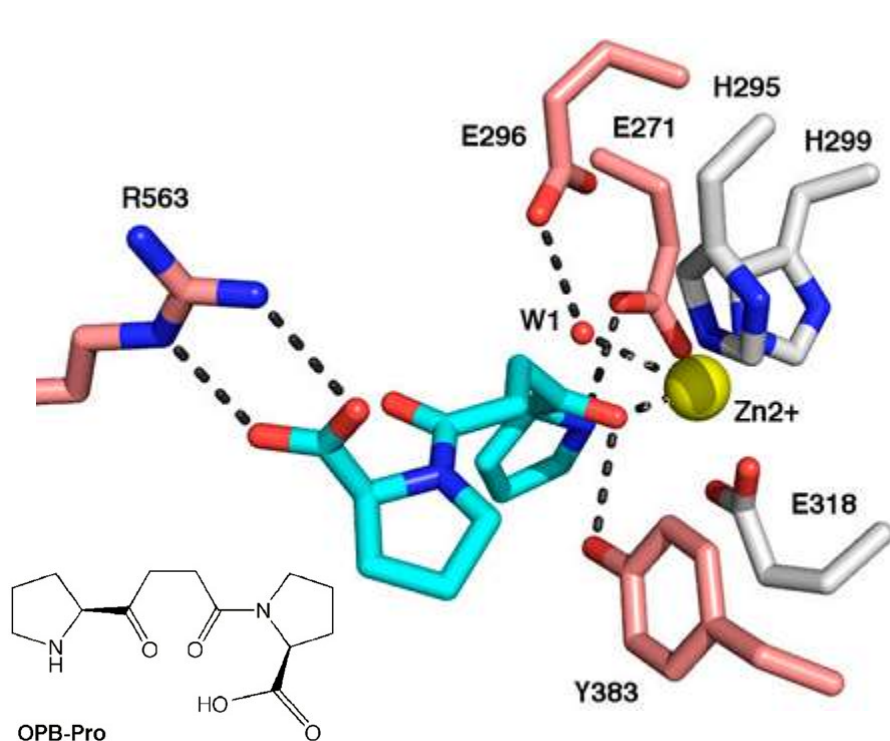
Selective inhibitor of LTA4H

- Leukotriene A4 hydrolase/aminopeptidase (LTA4H)

    - Structural studies (crystallography) with a tripeptide analogue revealed the aminopeptidase mechanism

    - This knowledge allowed designing a selective inhibitor that blocks the hydrolysis of LTA4 but NOT the hydrolysis of Pro-Gly-Pro

    - New promising lead compound against chronic inflammation

ARM1
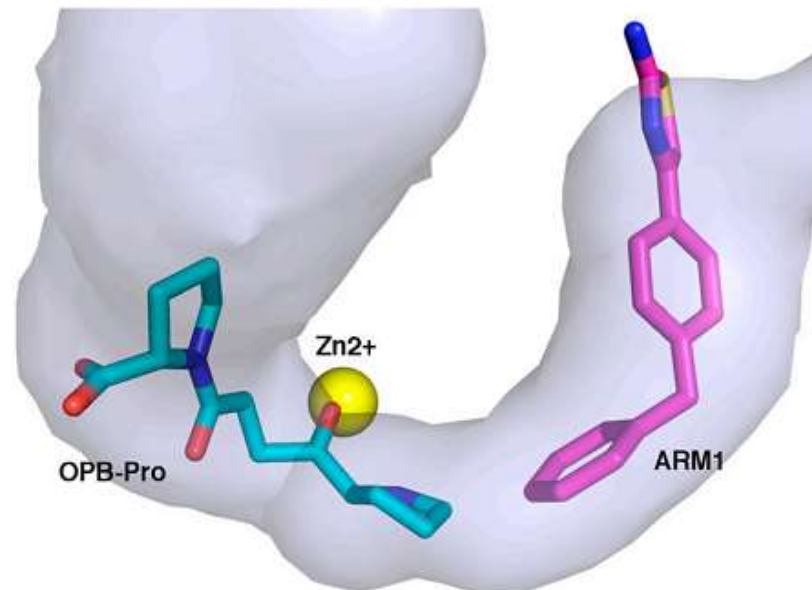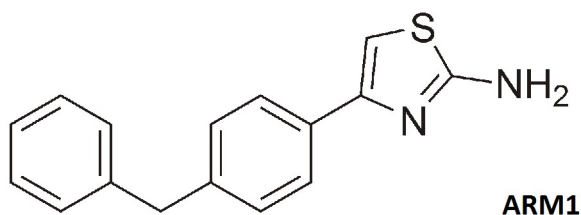
Protein engineering

- Stabilization of dehalogenase

- Dehalogenase activity

- Lipase enantioselectivity

- *De novo* design of a Diels-Alderase

Enzymes: practical applications?

- ❑ Ability to catalyze a desirable reaction

- ❑ Stable under process conditions

- ❑ Soluble expression

Enzymes: practical applications?

- ❑ Ability to catalyse a desirable reaction

- ❑ Stable under process conditions

- ❑ Soluble expression

Protein engineering process
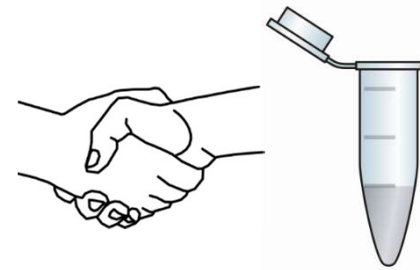
- ❑ Improvement of activity and/or selectivity

- ❑ Robust stabilization of proteins

- ❑ Design of more soluble proteins

Different 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

IMPROVED ENZYME

7. Biochemical testing

DIRECTED EVOLUTION

1. *not applied*

2. Random mutagenesis
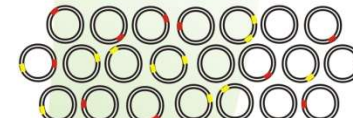
Library of mutated genes
(>10,000 clones)

3. Transformation

4. Protein expression

5. *not applied*

6. Screening and selection
- stability
- selectivity
- affinity
- activity
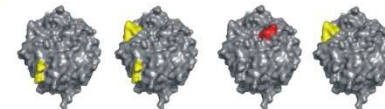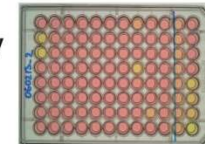
Selected mutant enzymes

Stabilization of dehalogenase

- ❑ Dehalogenase DhaA
  - ■ bacterial origin
  - ■ hydrolytic cleavage of C-X bond
  - ■ applications converting halogenated compounds to alcohols
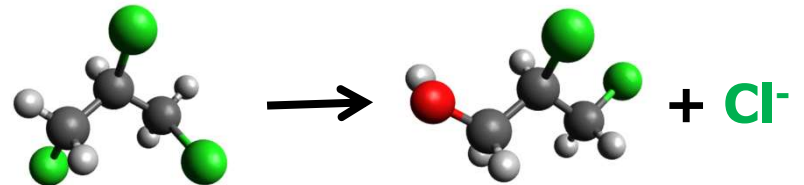
By-product recycling
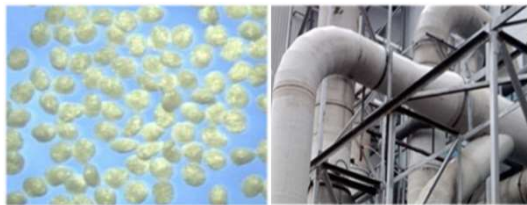
Biosensing

Bioremediation
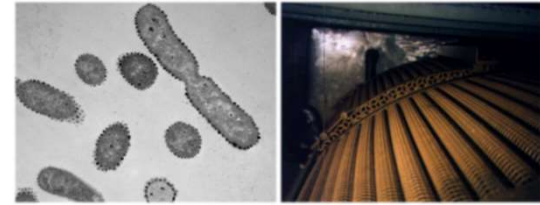
Cell imaging & protein analysis
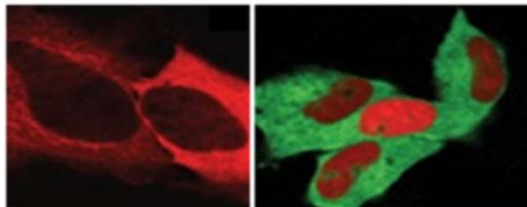
Biocatalysis

Decontamination

Stabilization of dehalogenase

❑ Dehalogenase DhaA

▪ melting temperature T_m = 49°C

▪ unstable at high temperatures

▪ activity half live at 60 °C $\tau_{1/2}$ ~ 5 min
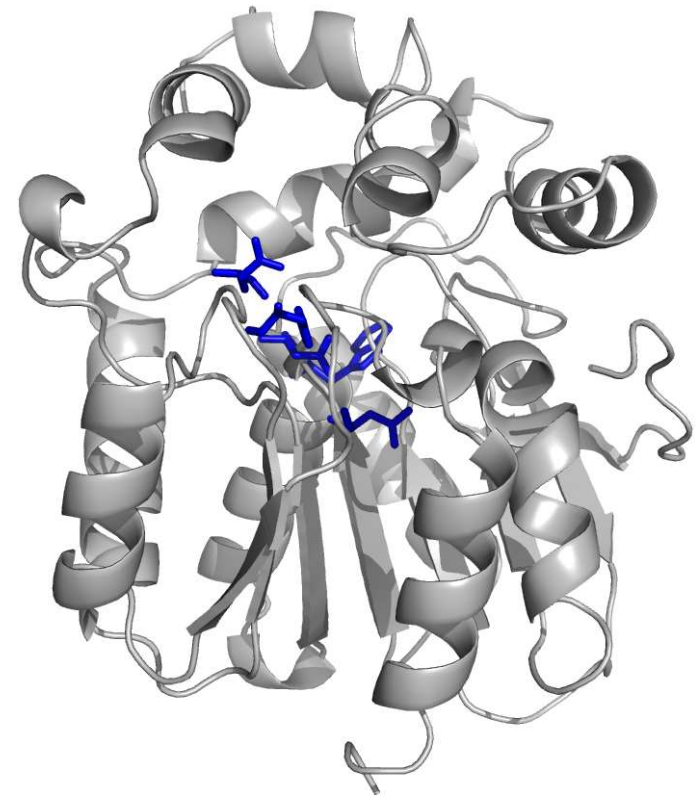
Stabilization of dehalogenase

- Gene Site Saturation Mutagenesis

    - joint project of Diversa and DOW Chemical

    - all **19** possible mutations at **315** positions tested experimentally

    - → **120,000** measurements

    - **10** single-point mutants more stable

    - cumulative mutant:

      T_m = 67 °C (18 °C ↑)

      and $\tau_{1/2}$ = 36 h (ca. 36 h ↑)
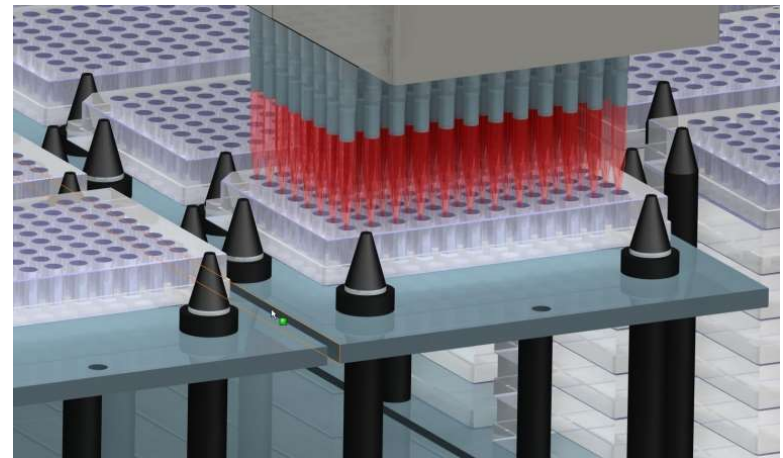
Stabilization of dehalogenase

❑ Rational design

- ▪ FIREPROT method

- ▪ structure and sequence analyses

- ▪ 5,500 possible mutants
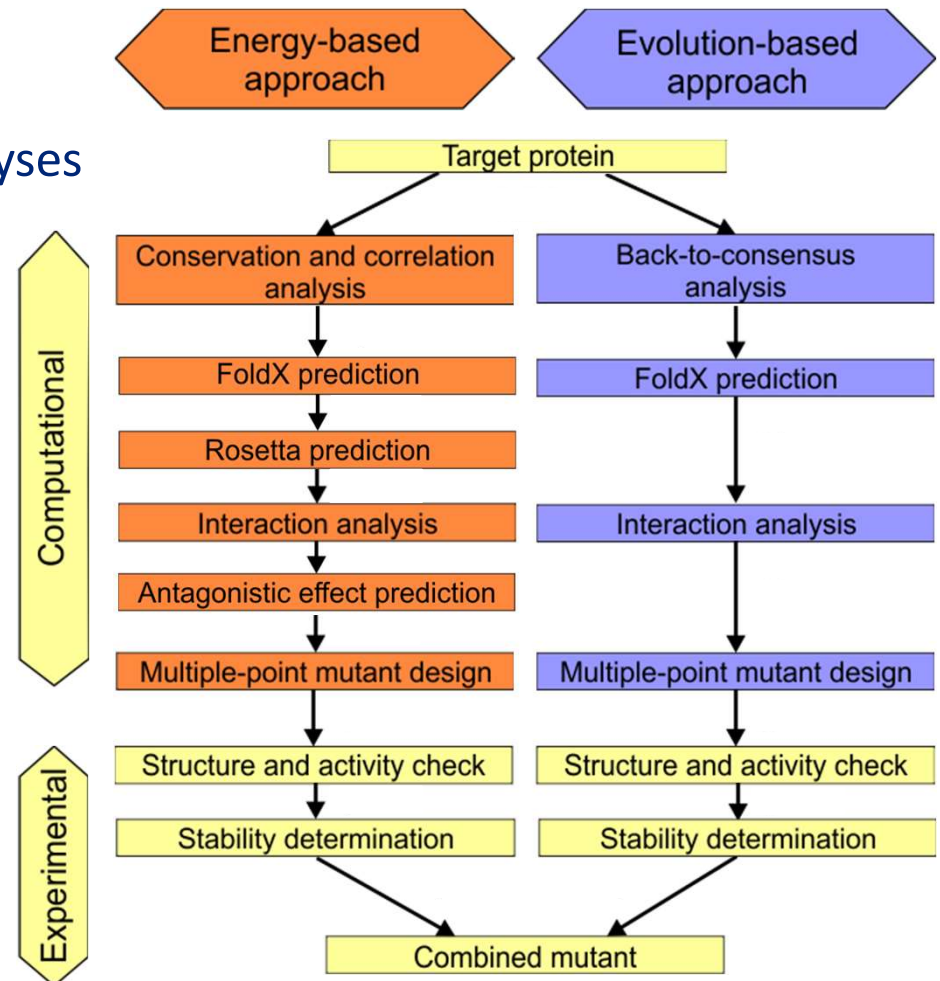
Stabilization of dehalogenase
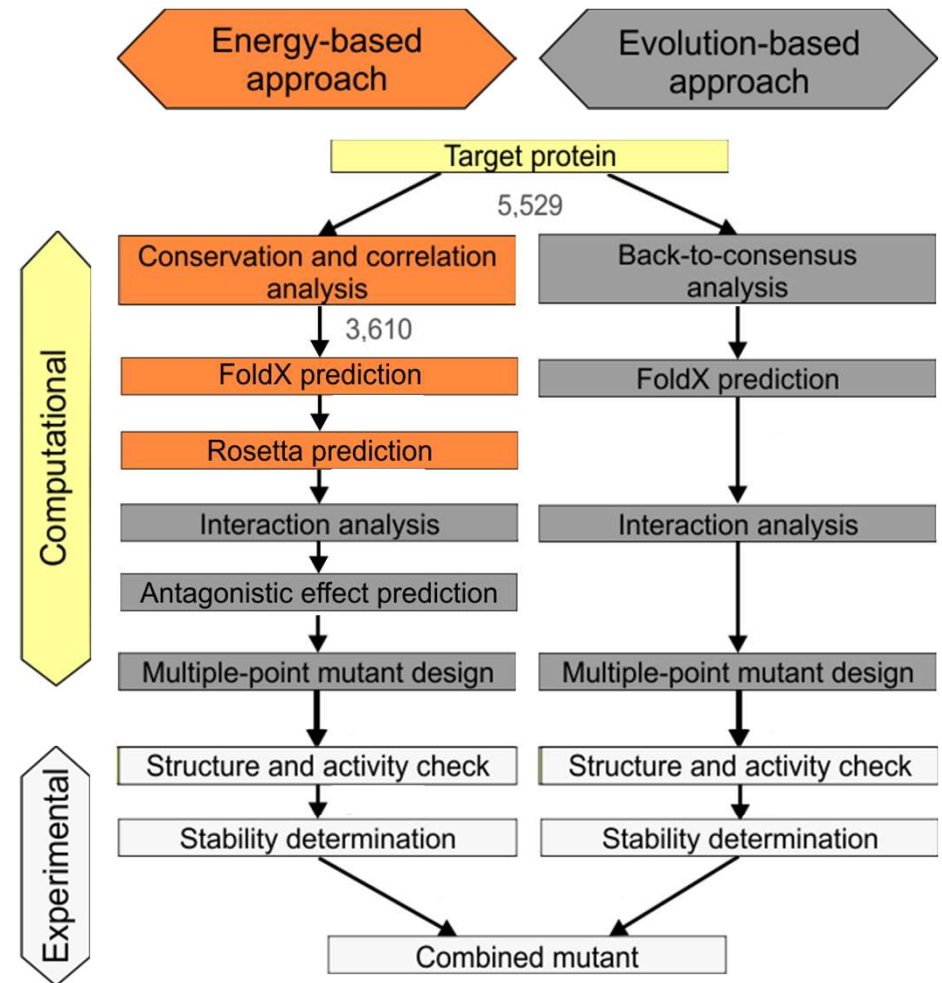
- ❑ Rational design
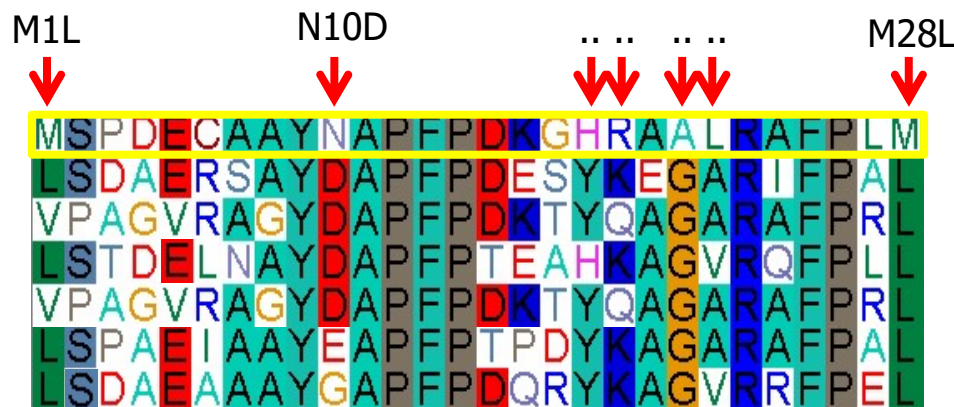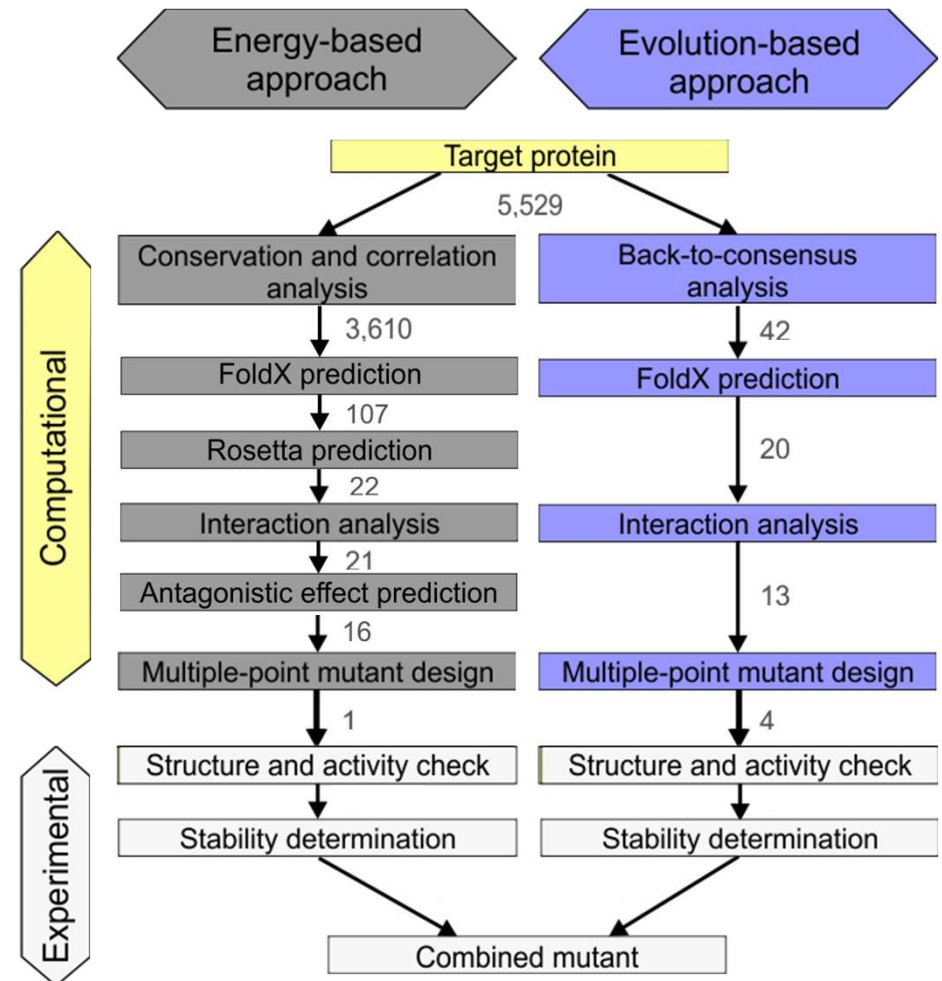 - ▪ FIREPROT method

Stabilization of dehalogenase

- ❑ Rational design
  - ■ FIREPROT method

Stabilization of dehalogenase
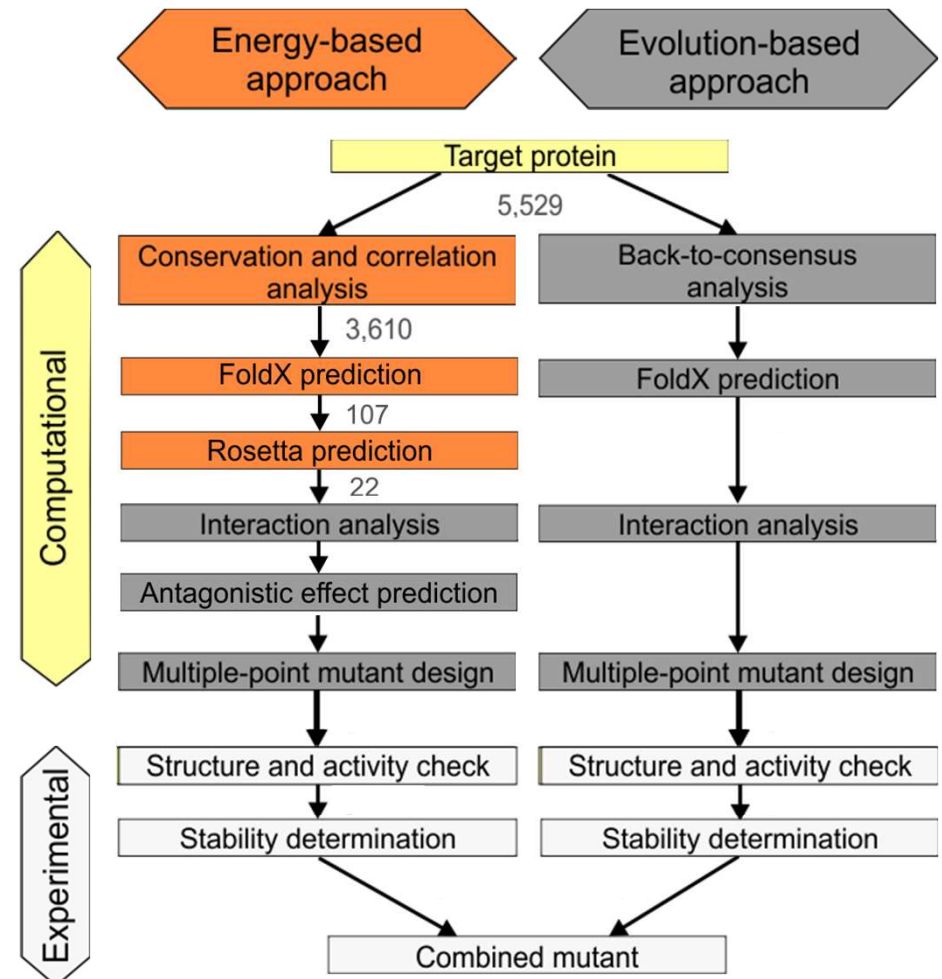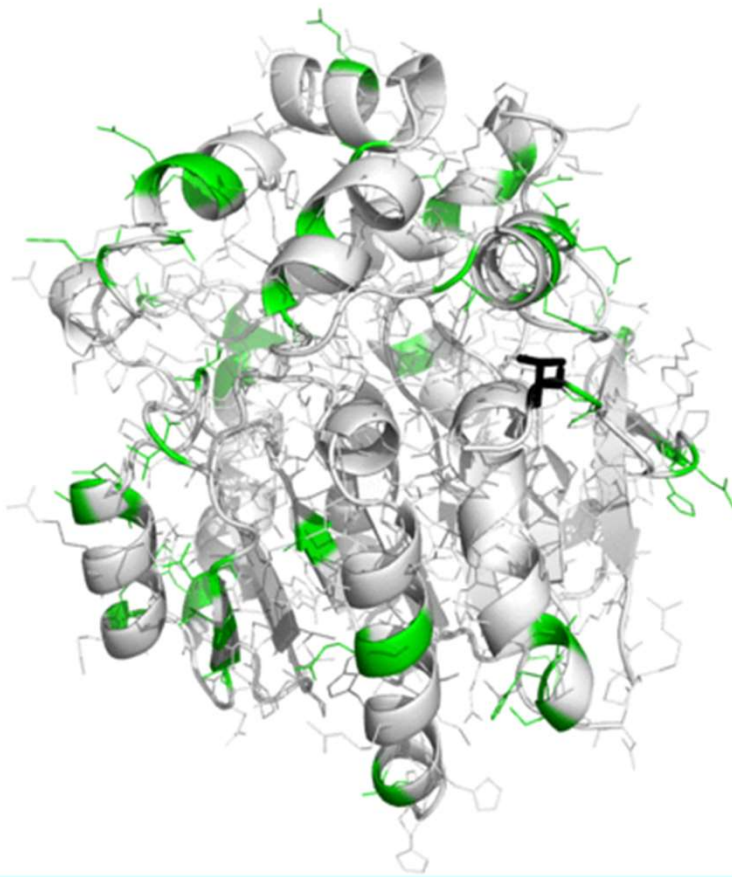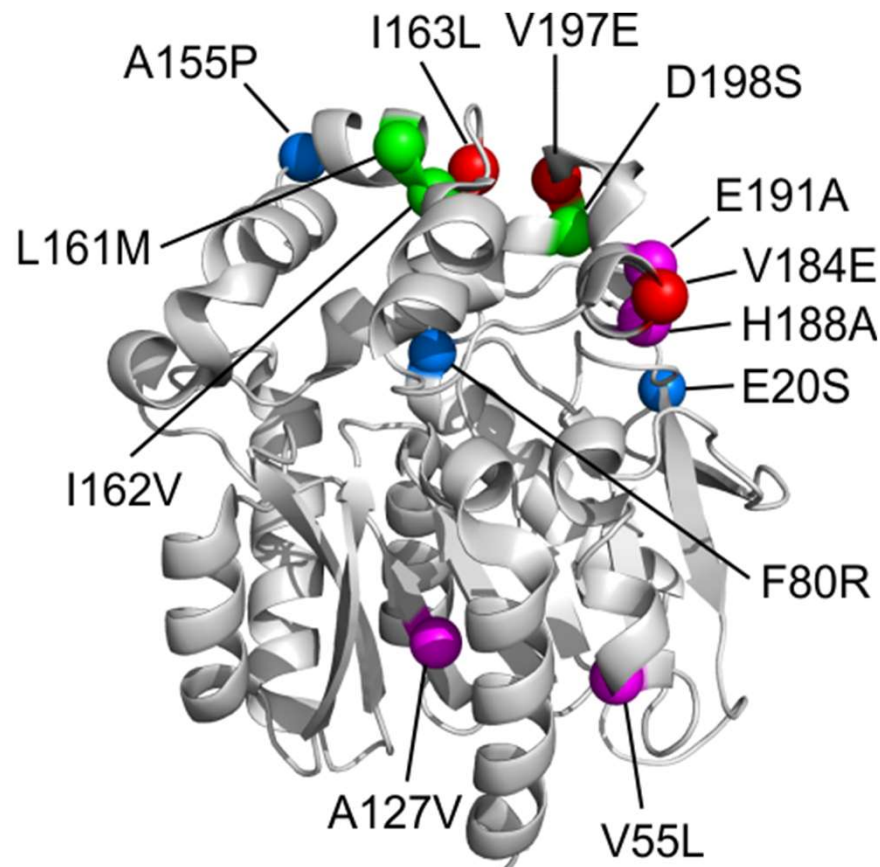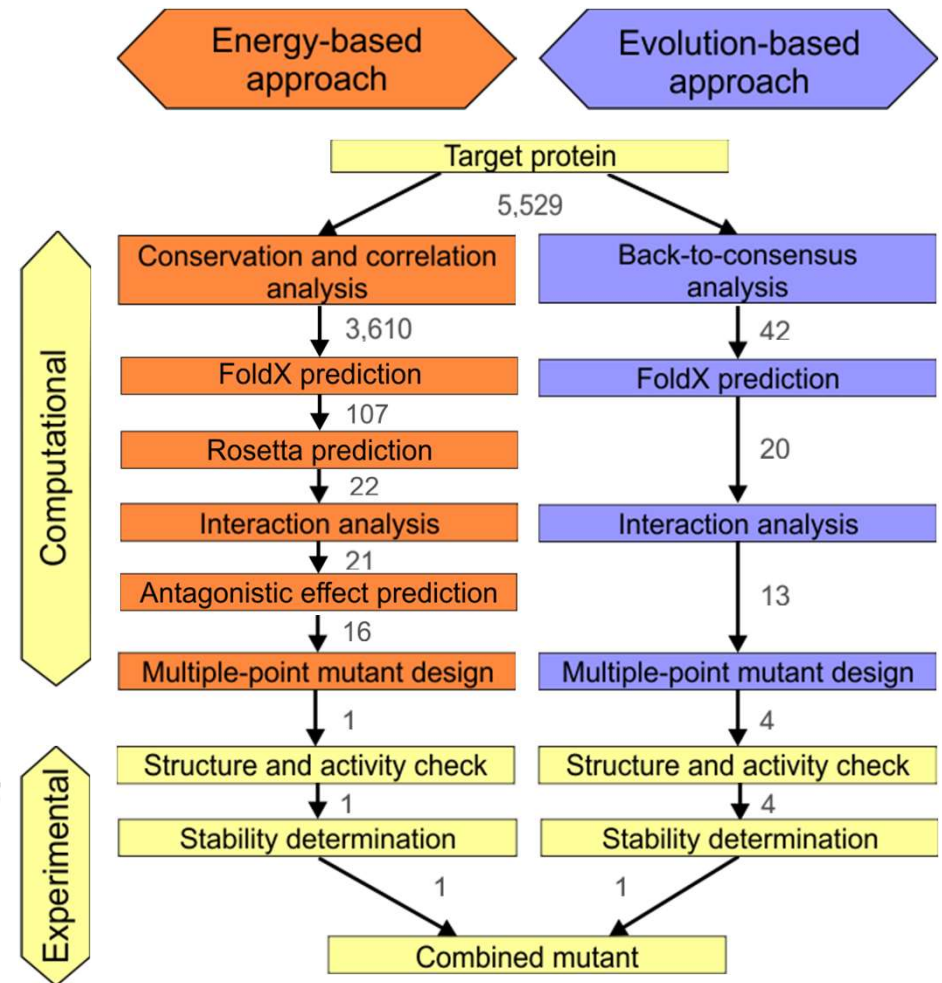
❑ Rational design

- ■ FIREPROT method

❑ Rational design
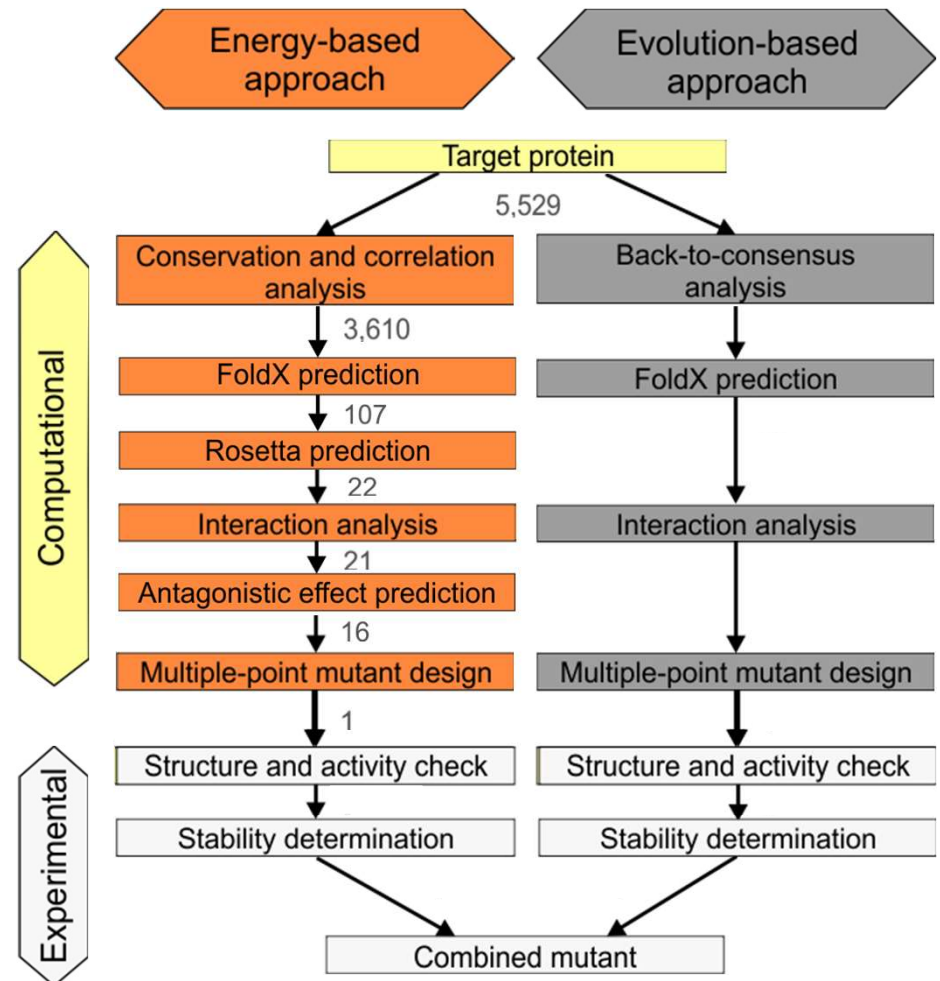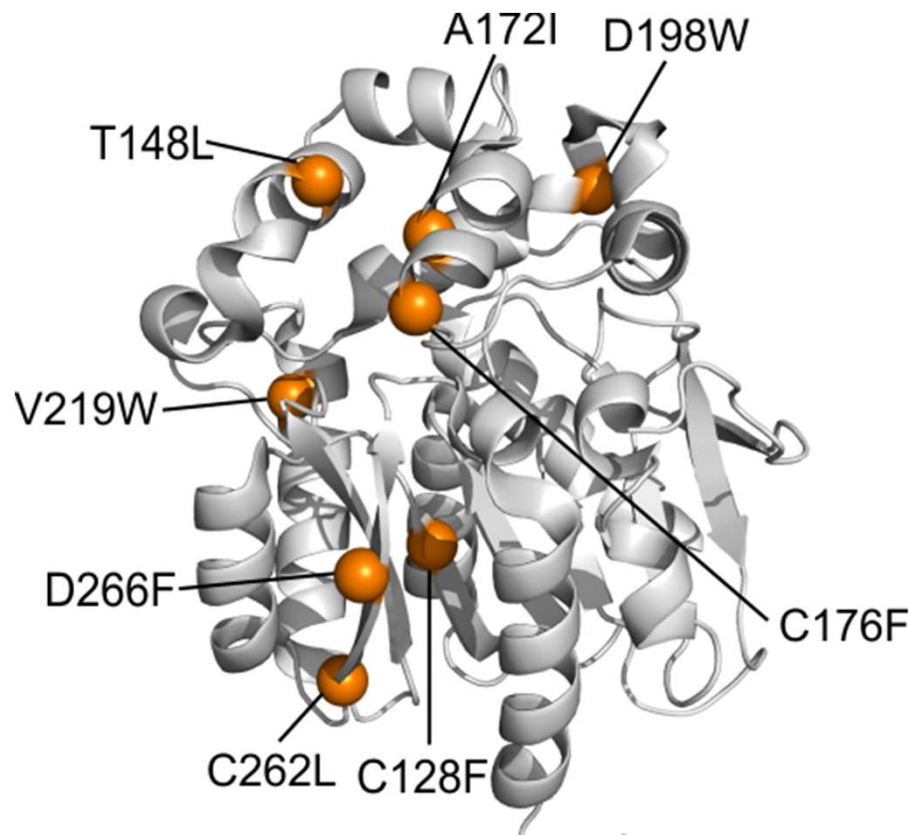
  ▪ FIREPROT method

Stabilization of dehalogenase
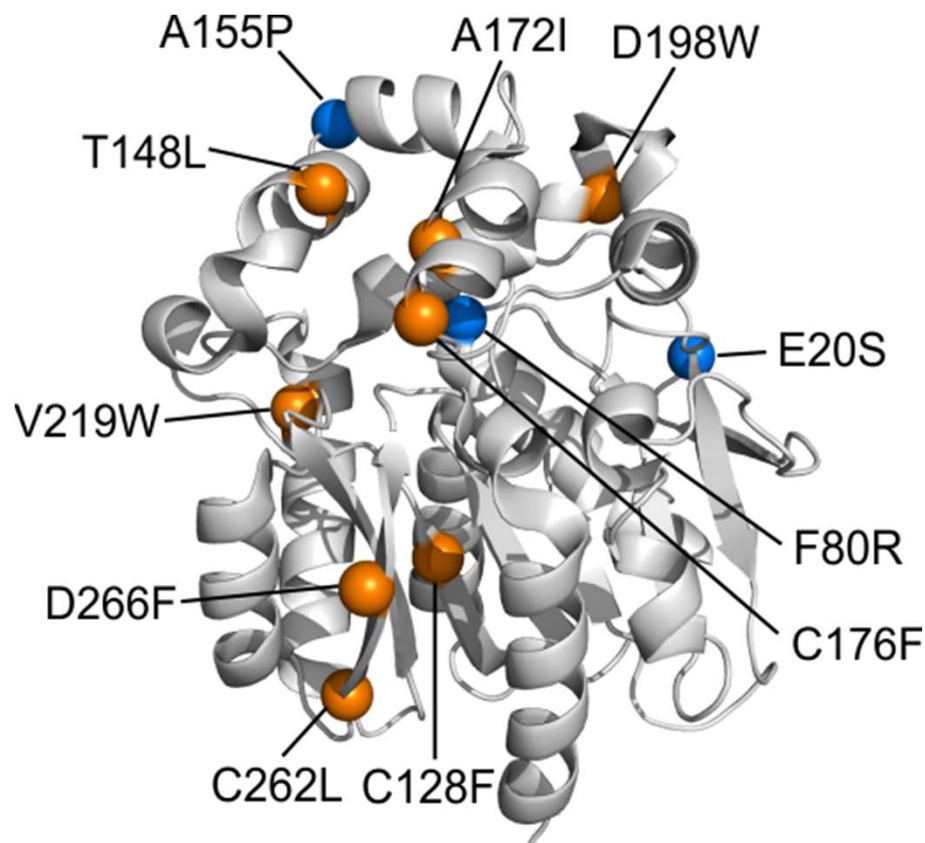
❑ Rational design
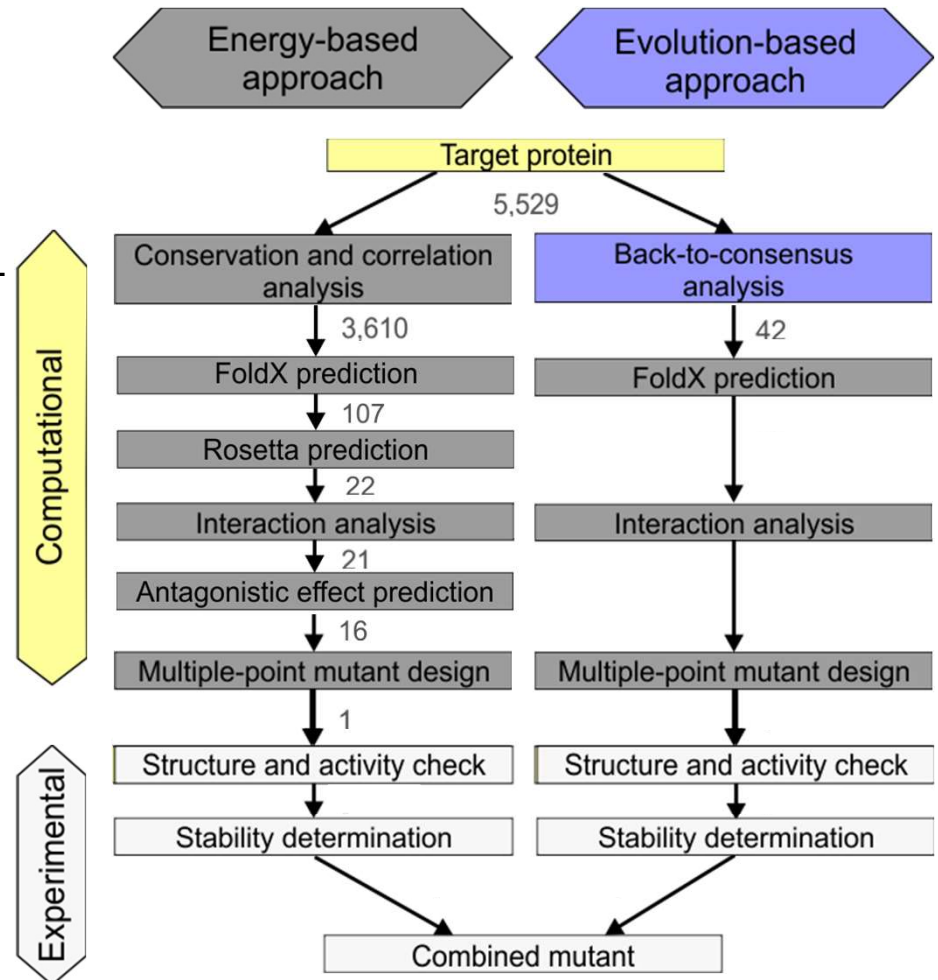
  ▪ FIREPROT method

Stabilization of dehalogenase

❑ Rational design

■ FIREPROT method
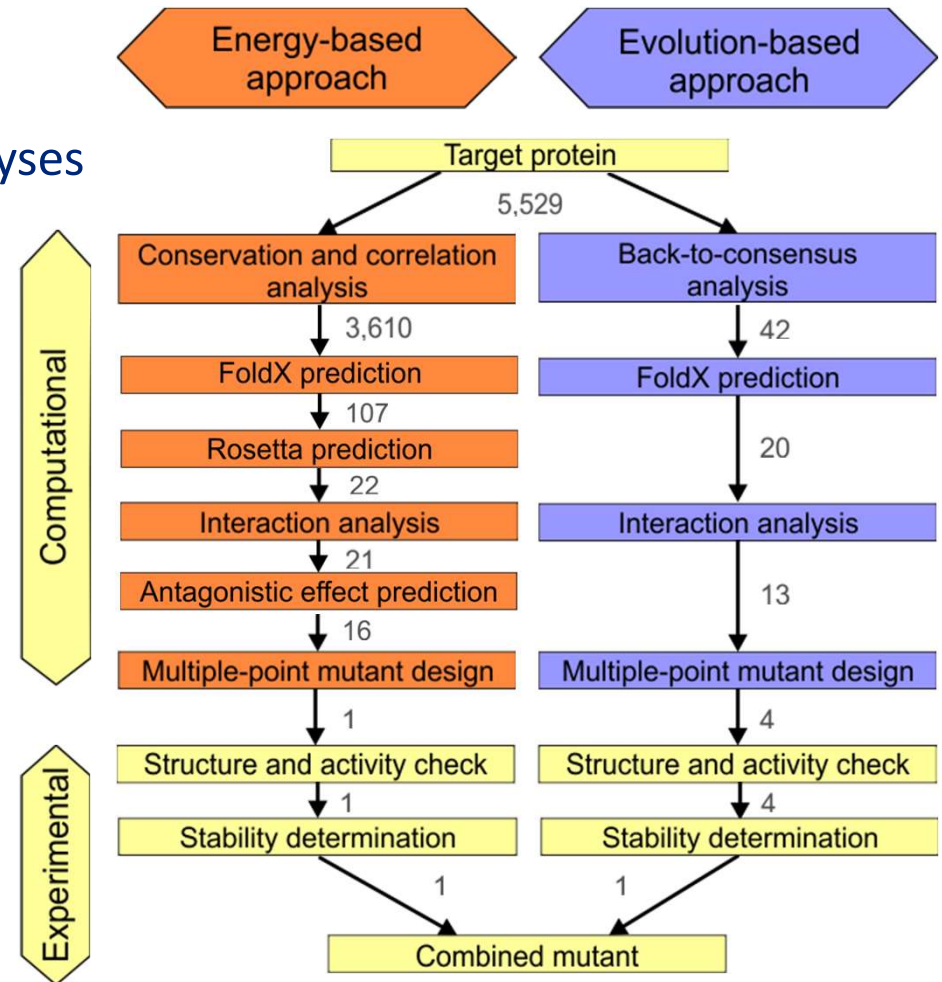
Stabilization of dehalogenase

❑ Rational design

■ FIREPROT method

Stabilization of dehalogenase

❑ Rational design

- FIREPROT method

- structure and sequence analyses

- **5,500 mutants predicted**

- experimental verification on

  5 multiple-point mutants

- **3 mutants more stable**

- best mutant (combined):

  T_m = 74 °C (25 °C ↑)

  and $\tau_{1/2}$ = 72 h (ca. 72 h↑)

Stabilization of dehalogenase

❑ Comparison

	GSSM	Rational design
Equipment (Kč)	20,000,000	500,000
Testing		
Computational	-	5,500
Experimental	120,000	5
Costs (Kč)	1,000,000	80,000
Time	Months	Weeks
Results		
# of stable mutants	11	3
Best: ΔT_m (°C)	18	25
$\tau_{1/2}$ (h)	36	72

Dehalogenase activity

- **TCP**: toxic persistent pollutant

    from industrial sources
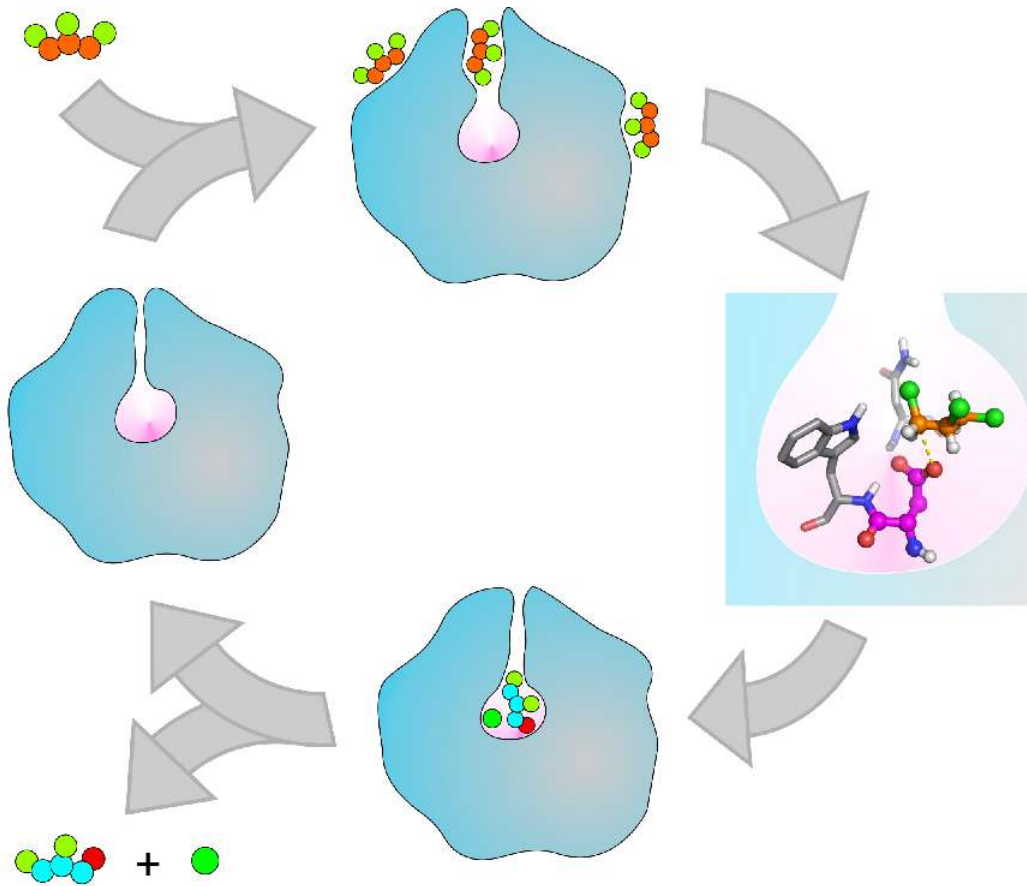

TCP → DCP $+ Cl^-$
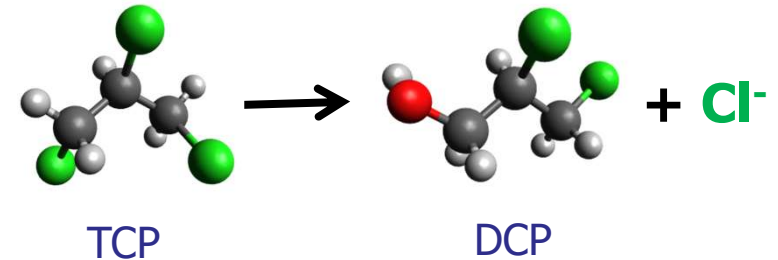
- DhaA dehalogenase (poor catalyst)

    - DhaA31: 5 mutations narrowed the access tunnels

    - Catalytic rate (k_{cat}) with TCP increased 32-fold
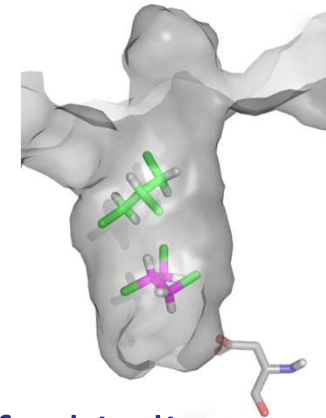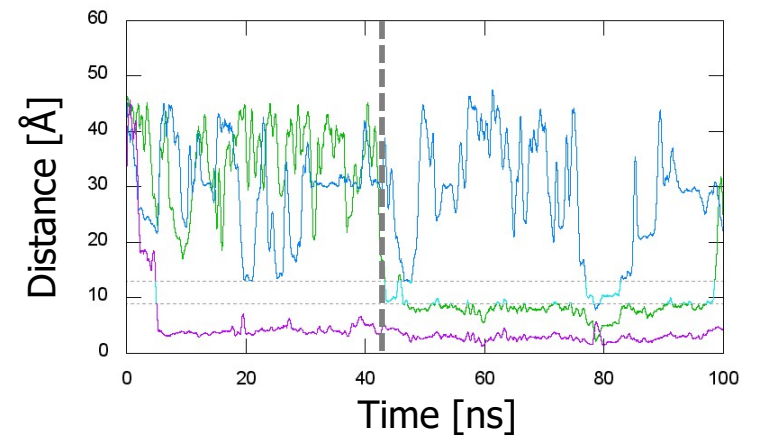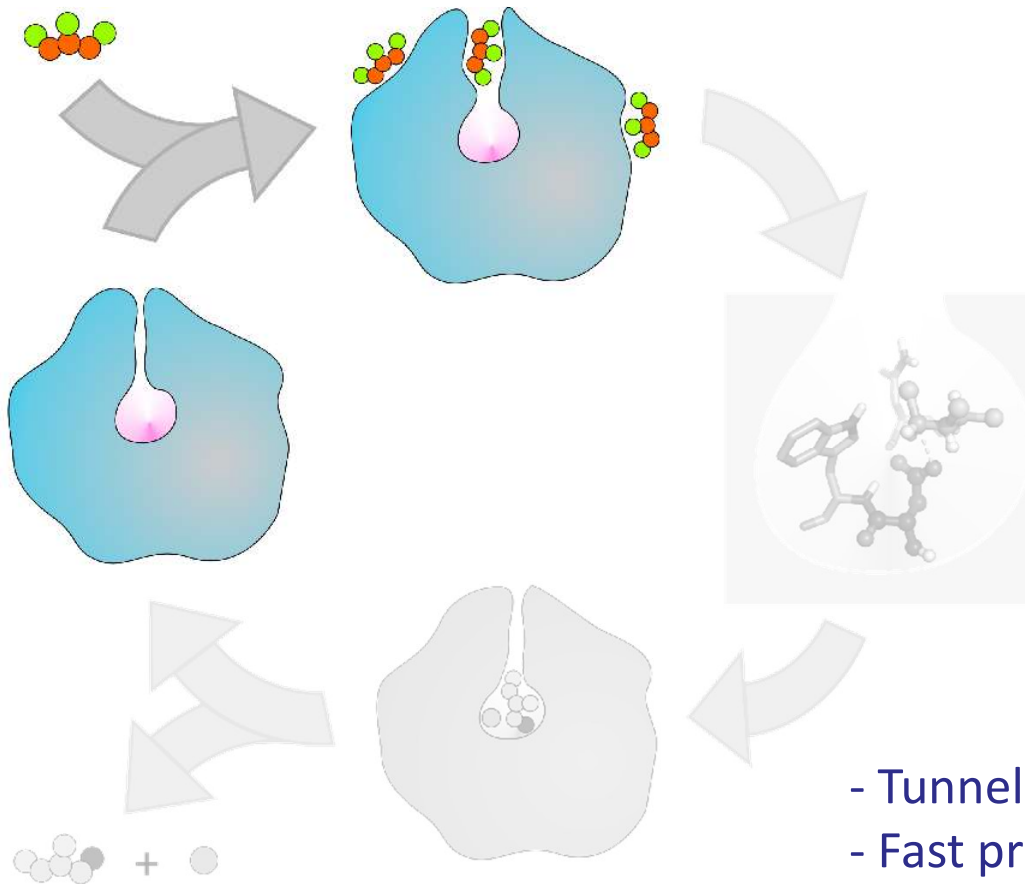

DhaA   DhaA31
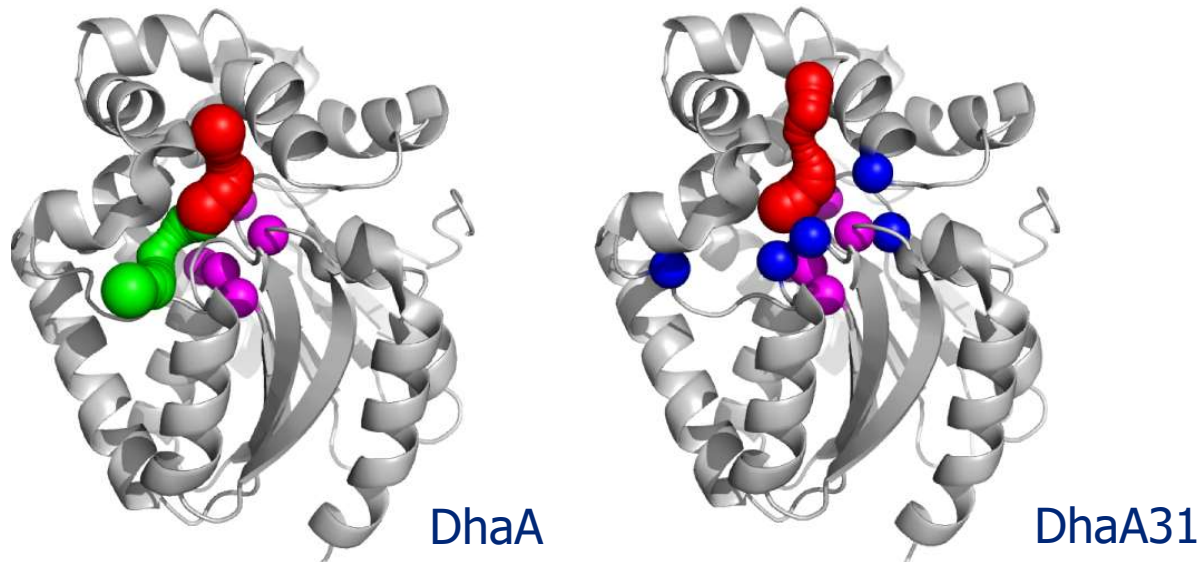
Dehalogenase activity

- **Catalytic cycle**: enzymes with buried active site
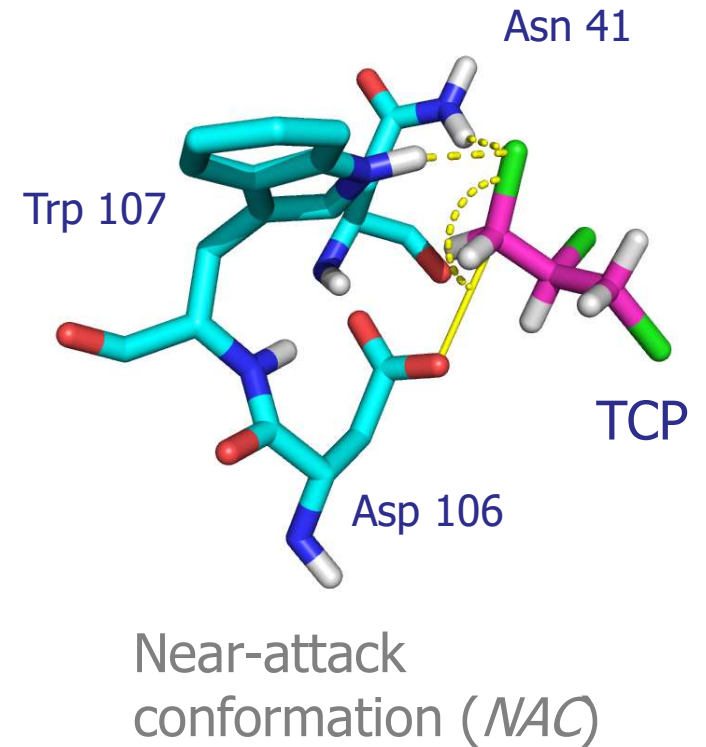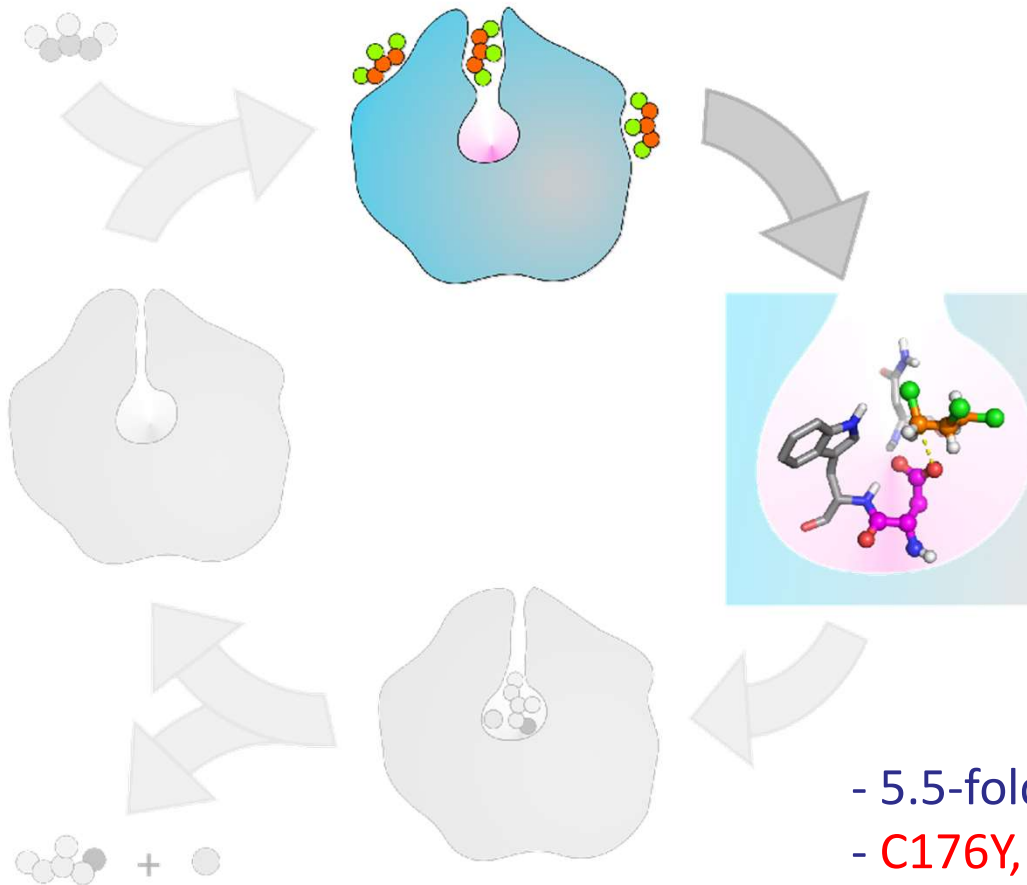
Dehalogenase activity

□ **Substrate binding: MD simulations**

- Tunnels need to open for binding
- Fast process for both DhaA31 and DhaA^{WT}
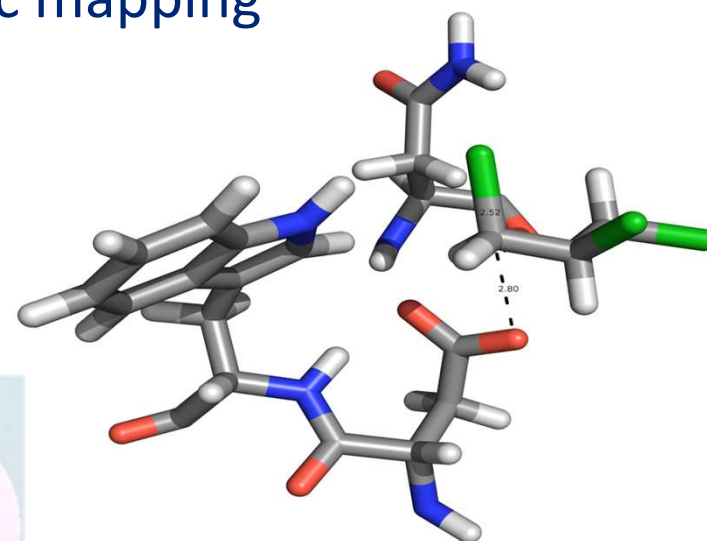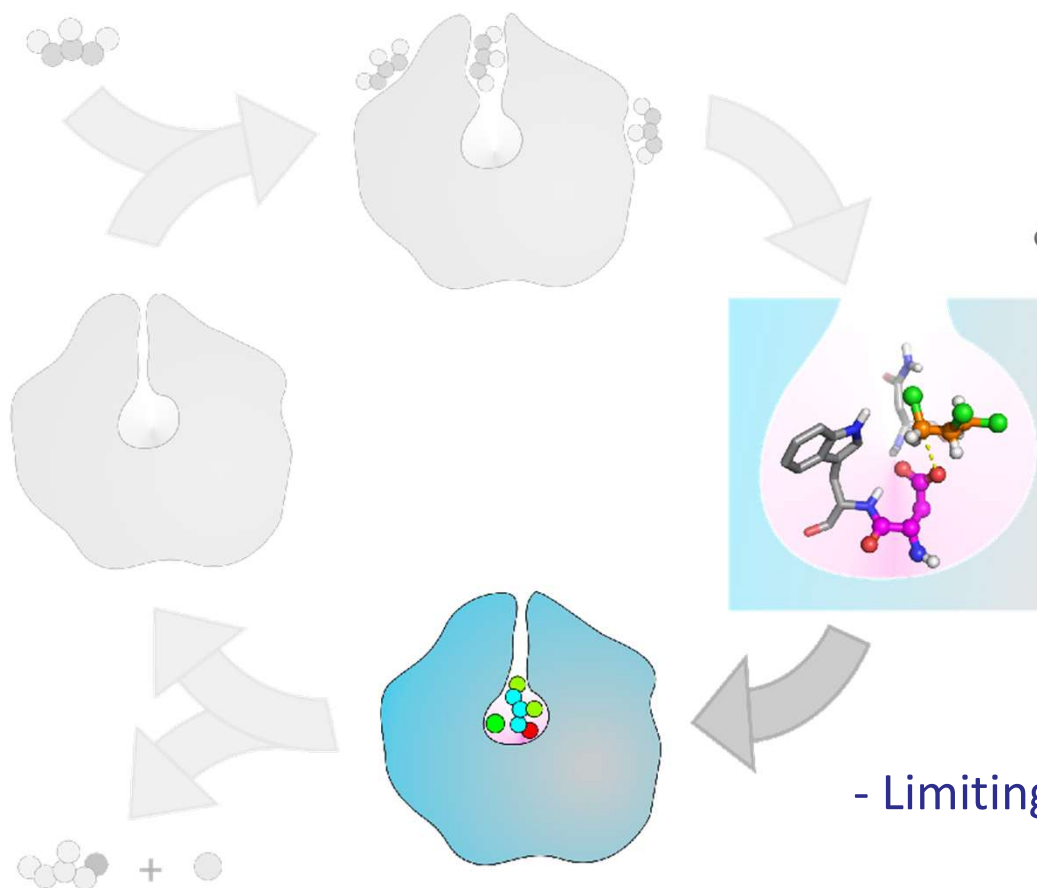- Potential substrate inhibition

Dehalogenase activity

☐ Reactive binding: MD simulations

Near-attack
conformation (*NAC*)

- 5.5-fold higher *NAC* rates in DhaA31
- C176Y, V245F increased interactions
- C176Y induced correct orientation of TCP

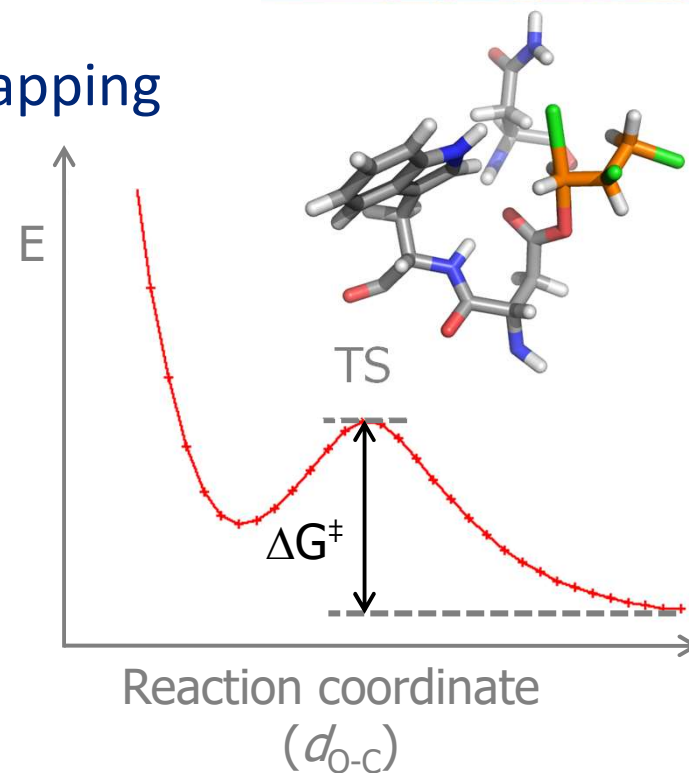# Dehalogenase activity

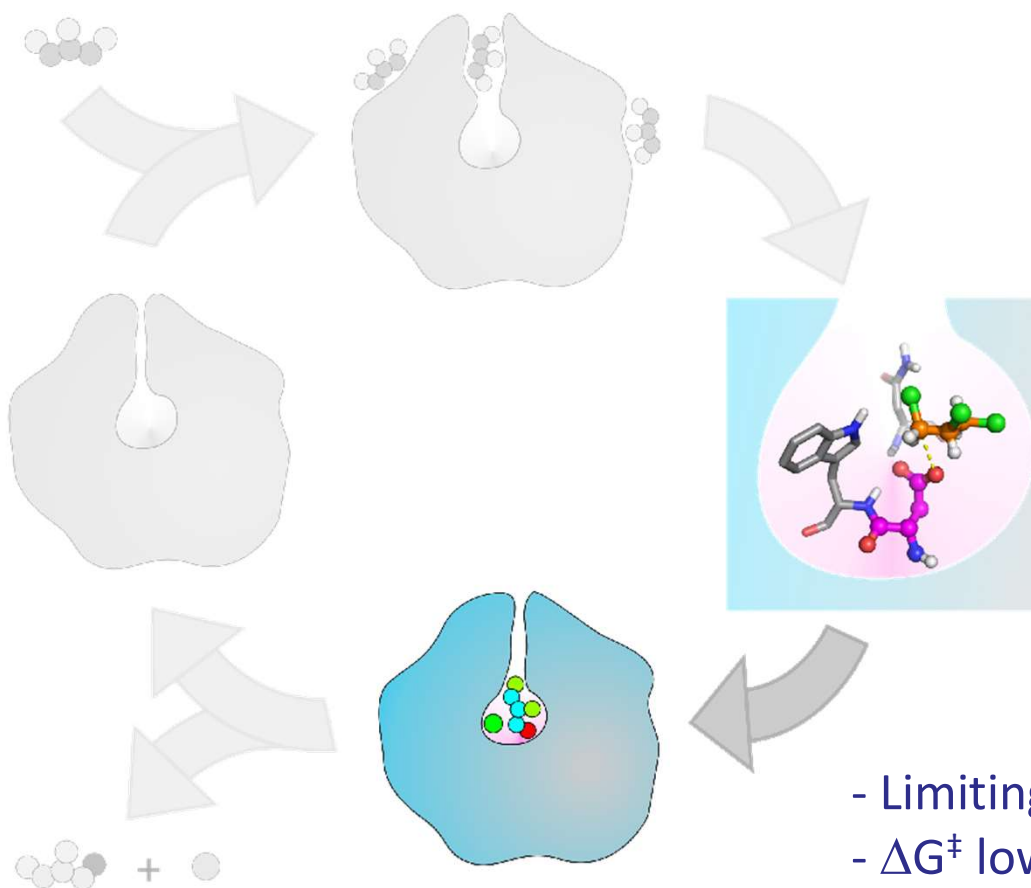- Chemical step: QM/MM adiabatic mapping

Reaction coordinate
(d_{O-C})

- Limiting step in DhaA[WT]

Dehalogenase activity

❑ Chemical step: QM/MM adiabatic mapping

E

TS

$\Delta G^{\ddagger}$

Reaction coordinate
(d_{O-C})

- Limiting step in DhaA^{WT}
- $\Delta G^{\ddagger}$ lower in DhaA31 by 1.6 kcal/mol
- This implies ca. 14-fold higher reactivity

Dehalogenase activity

❑ **Product release: MD simulations**

Phe 152 Phe 168
Phe 149
Tyr 176
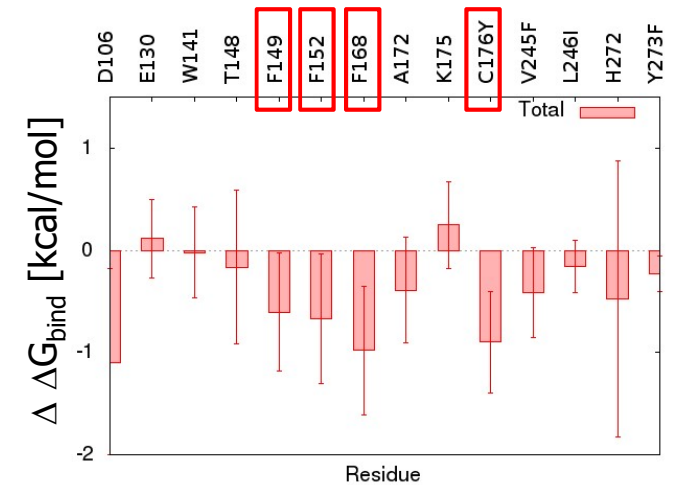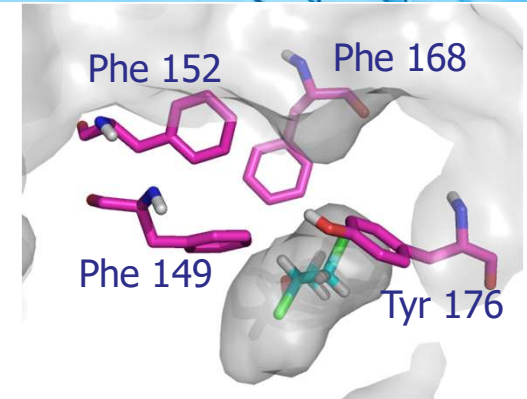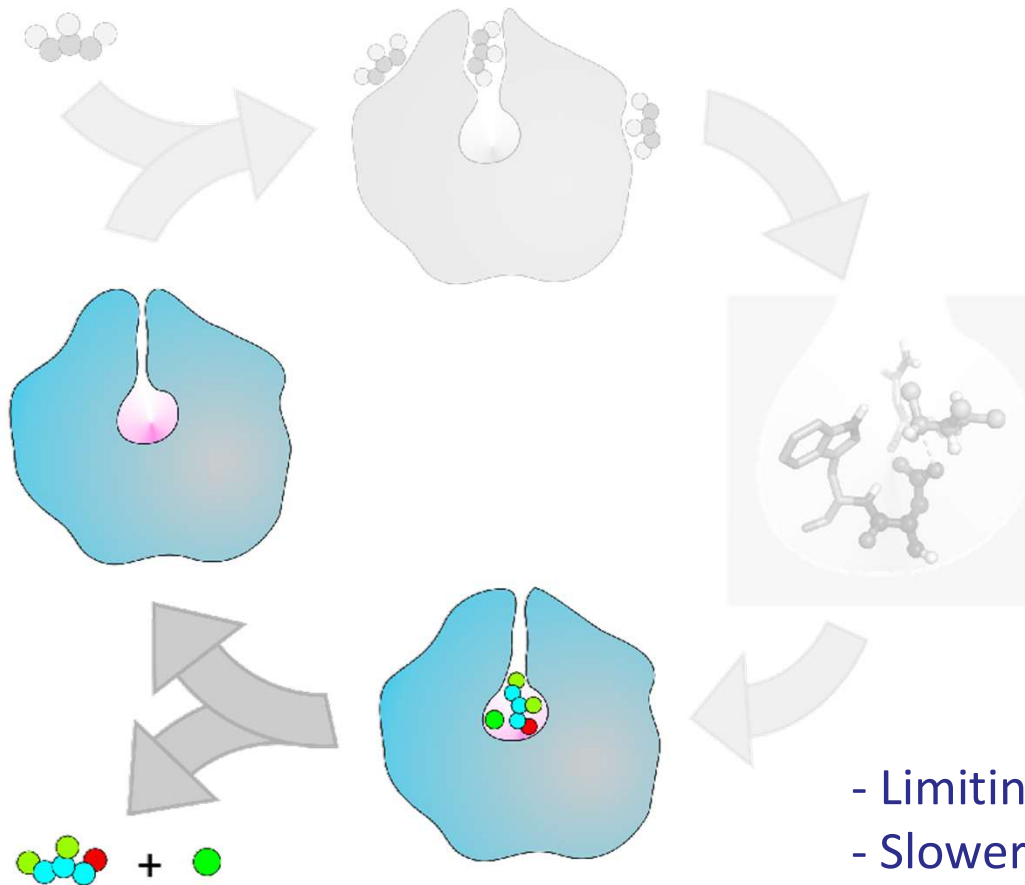
- Limiting step in DhaA31
- Slower release rates in DhaA31
- F149, F152, F168, Y176 prevent release

❑ Conclusions
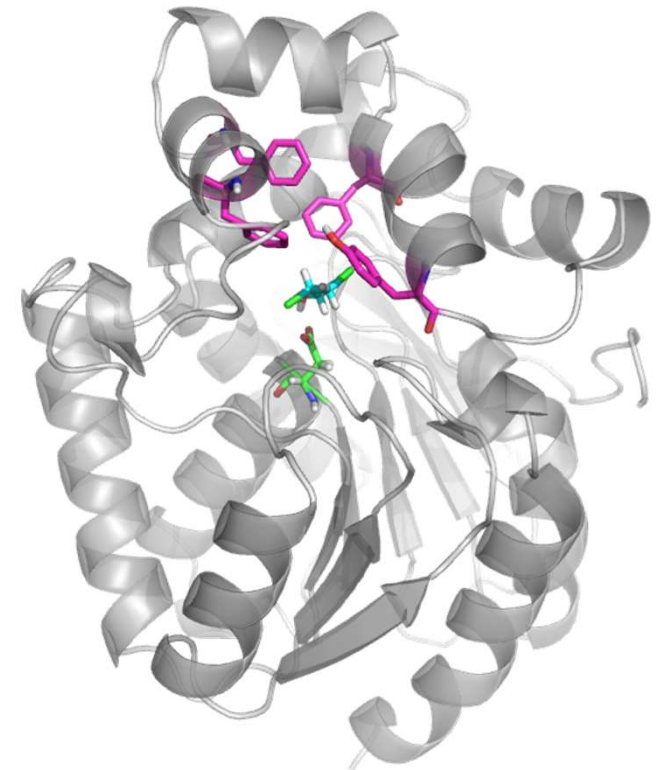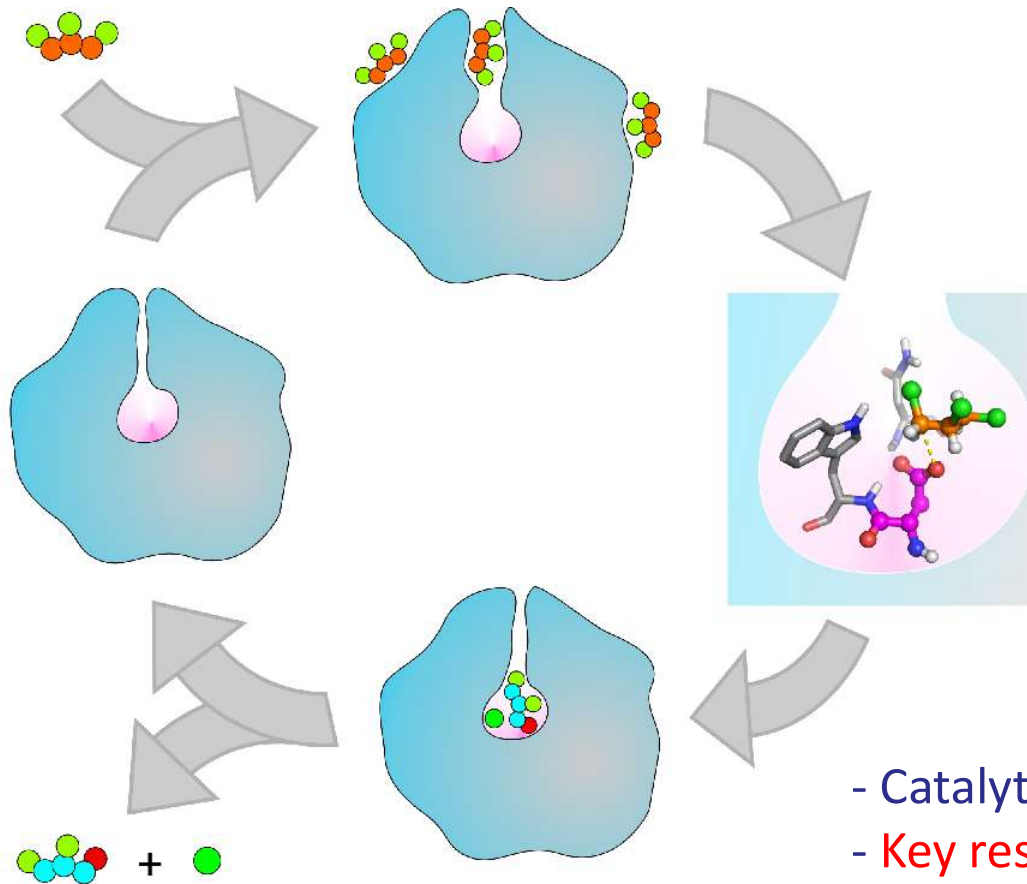
- Catalytic differences explained
- Key residues identified
- New hot-spots for mutagenesis identified

Lipase enantioselectivity

❑ Lipase (EC 3.1.1.3, bacterial enzyme)

- Triacylglycerol + H_2O → diacylglycerol + carboxylic acid

- **Versatile biocatalysts**: catalyze hydrolysis of carboxylic esters, esterification, interesterification, transesterification

- Industrial applications:

    - food, detergent, pharmaceutical, leather, textile, cosmetic, paper industries

    - used to resolve racemic mixtures

B. cepacia lipase

Tris, HCl, pH 7.5
n-octane

(*R*-enantiomer)
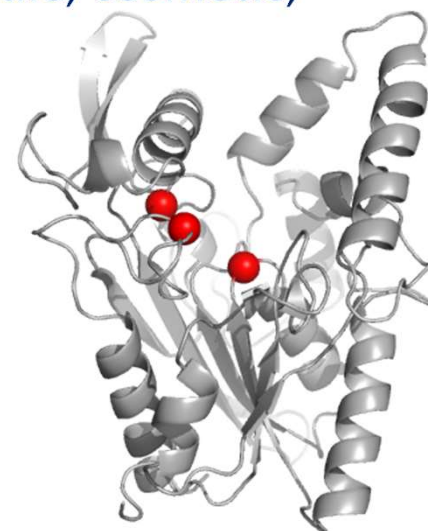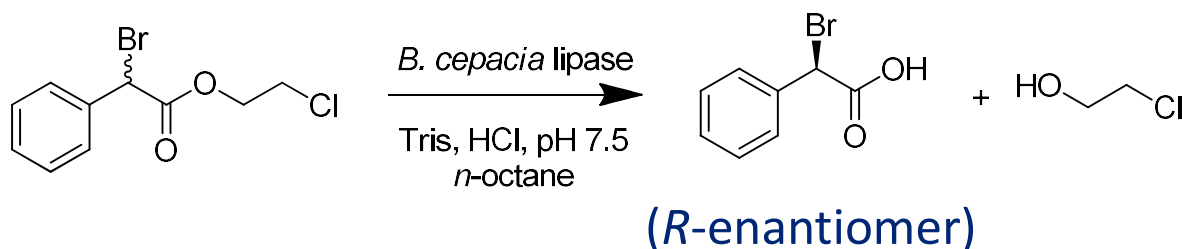
Lipase enantioselectivity

❑ Lipase (EC 3.1.1.3, bacterial enzyme)

- Molecular modeling suggested residues in the tunnel controlling

  substrate access are key to enantioselectivity

- Saturated mutagenesis at 3 positions

- Higher E-value and conversion %

Variants	Enantiopreference	E value	Conversion [%]
Wild-type	R	13 (±1.8)[b]	6.5 (48 h)
V266G	S	20 (±4)[c]	6.6 (51 h)
L17S	R	128 (±35)[b]	15.6 (49 h)
L17M	R	133 (±31)[b]	15.5 (48 h)
L17M/V266M	R	166	9 (19 h)
L17S/L287A	R	22.5	15.6 (20 h)
L17S/L287I	R	178	15.5 (20 h)
L17S/L287W	R	55	6 (20 h)
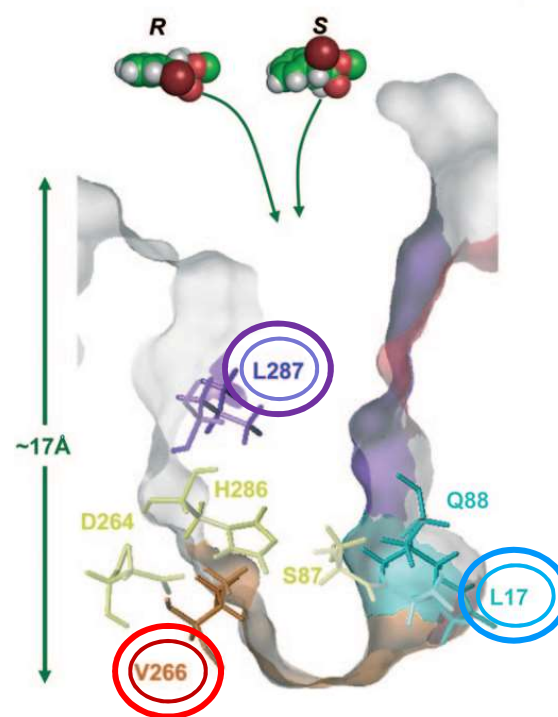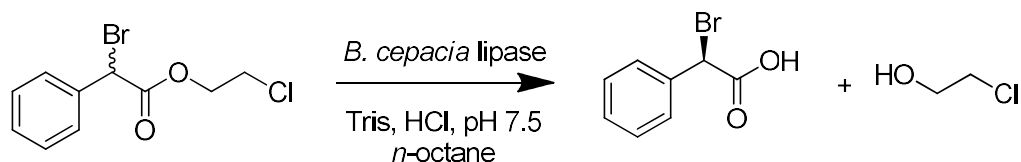
Lipase enantioselectivity

- ❑ Lipase (EC 3.1.1.3, bacterial enzyme)

  - ▪ Molecular dynamics with substrates

Variants	Enantiopreference	E value	Conversion [%]
Wild-type	R	13 ($\pm$1.8)[b]	6.5 (48 h)
V266G	S	20 ($\pm$4)[c]	6.6 (51 h)
L17S	R	128 ($\pm$35)[b]	15.6 (49 h)
L17M	R	133 ($\pm$31)[b]	15.5 (48 h)
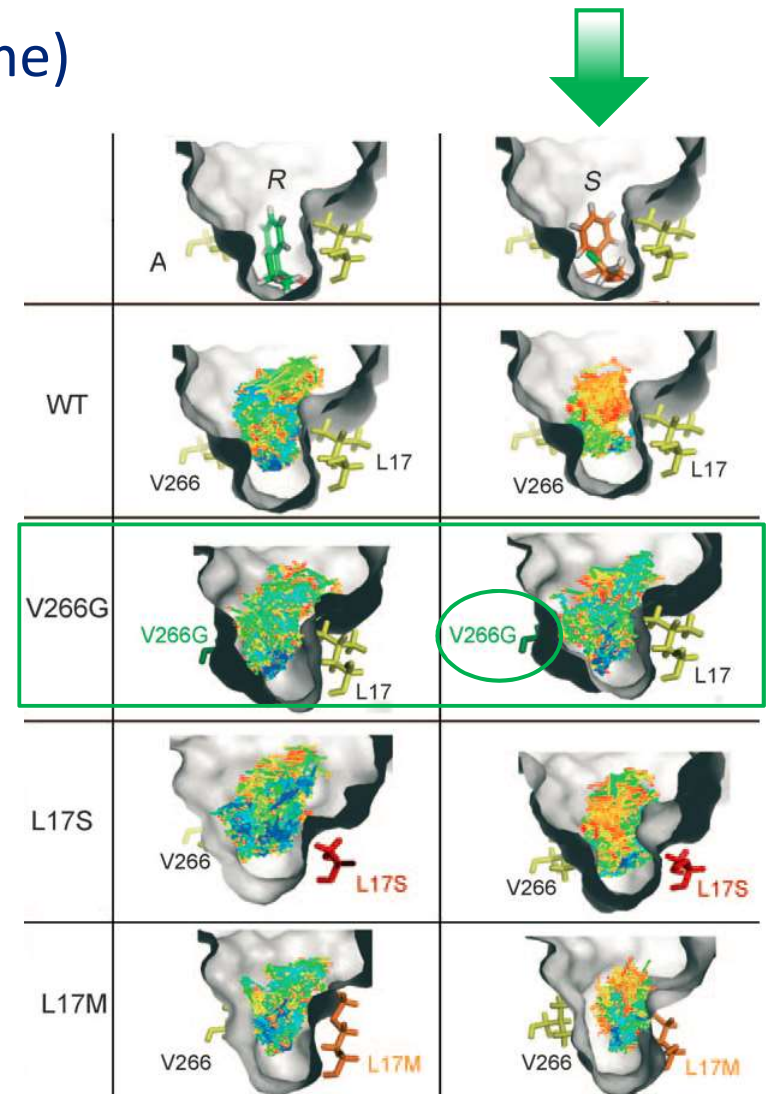
Lipase enantioselectivity

- ❑ Lipase (EC 3.1.1.3, bacterial enzyme)

  - ▪ Molecular dynamics with substrates

Variants	Enantiopreference	E value	Conversion [%]
Wild-type	R	13 (±1.8)[b]	6.5 (48 h)
V266G	S	20 (±4)[c]	6.6 (51 h)
L17S	R	128 (±35)[b]	15.6 (49 h)
L17M	R	133 (±31)[b]	15.5 (48 h)

  - ▪ Steric changes on either side of the active site favor reactive binding of only one enantiomer

  - ▪ Combining mutations, they favor one enantiomer and disfavor the other
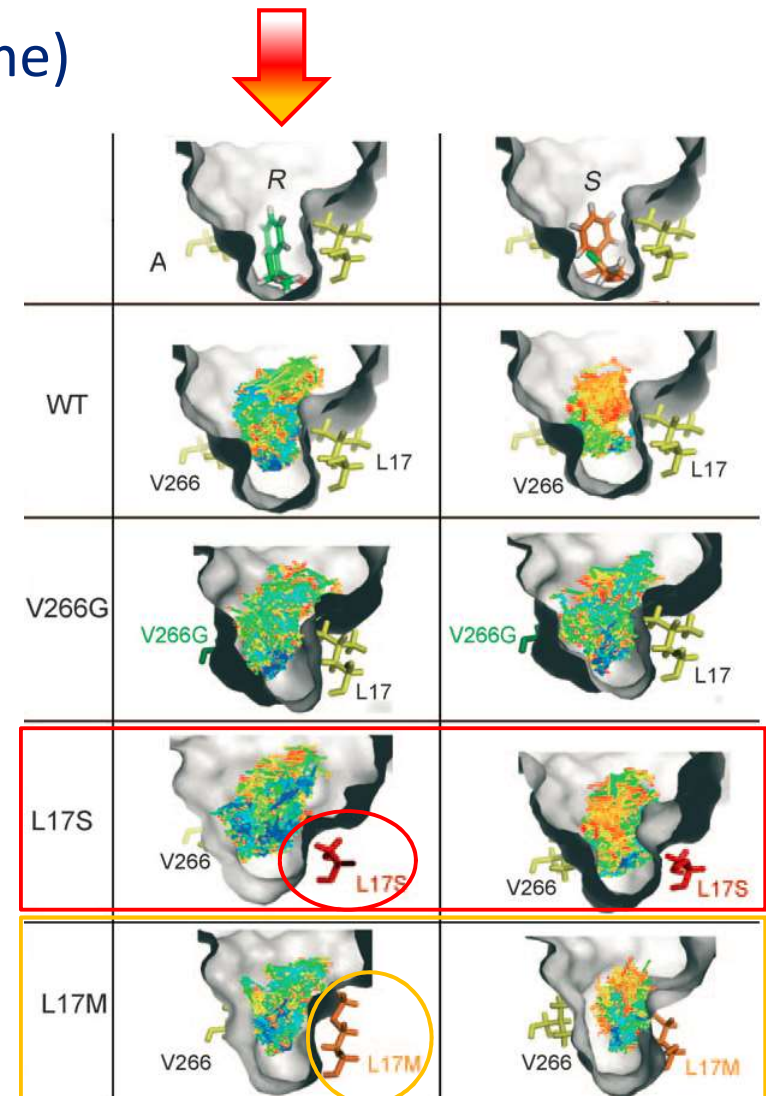
De novo design of a Diels-Alderase

❑ Non-existing Diels-Alderase

  ▪ **goal** – biocatalyst for intermolecular Diels-Alder reaction – very
  specific geometric and electronic requirements – **theozymes**

  ▪ **design** – computational match with protein scaffolds and refinement

  ▪ **mutagenesis** – site-directed to design active site

  ▪ **evaluated library** – < 100

  ▪ **result** – creation of functional & stereoselective Diels-Alderase
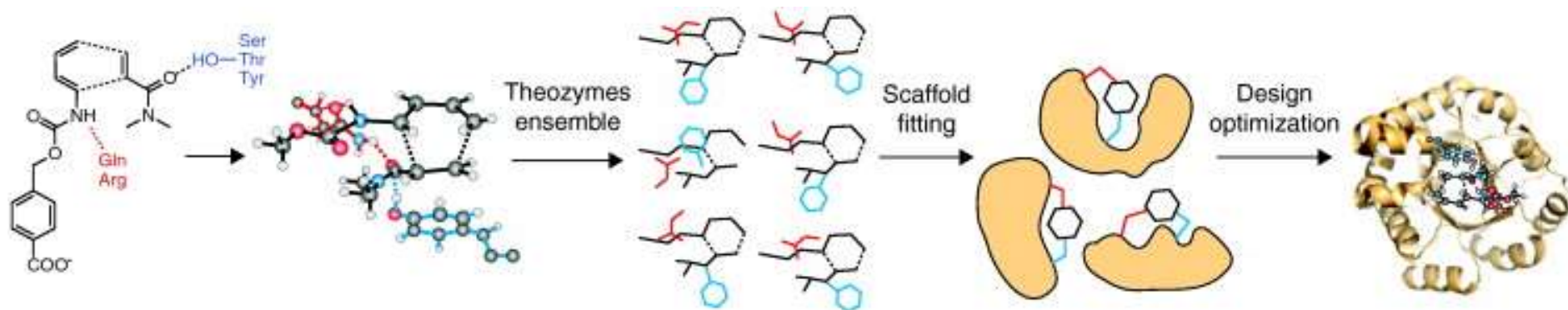
Summary

- Structural biology methods are invaluable tools to:

  - Explain biological phenomena

  - Increase efficiency in Drug Design

  - Successful Protein Engineering aiming at biotechnological applications

- Often produce better results than experimental brute-force

- Can reduce costs and save time

References

- Congreve, M. *et al.* (2005) Structural biology and drug discovery. *Drug Discov Today* **10**: 895-907.

- Lee, D. *et al.* (2007) Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol*. **8**: 995-1005

- Lutz, S. (2010) Beyond directed evolution — semi-rational protein engineering and design. *Curr Opinion Biotechnol* **21**: 734–743.

- Weber, I. T. & Agniswamy, J. (2009) HIV-1 protease: structural perspectives on drug resistance. *Viruses* **1**: 1110-1136.

- Marques, S.M. *et al.* (2017) Catalytic cycle of haloalkane dehalogenases toward unnatural substrates. *J Chem Inf Model*. **57**: 1970-1989.

- Jessop, T.C. *et al.* (2009) Lead optimization and structure-based design of potent and bioavailable deoxycytidine kinase inhibitors. Bioorganic & Medicinal Chemistry Letters 19 6784–6787

Final remarks on the course

- 3 exam dates (location:B11/305):

  - 12 Jan. 2021; 13:00

  - 25 Jan 2021; 13:00

  - 10 Feb. 2021; 09:00

- Multiple-choice exam

  - 25 questions

  - 10 correct answers out of needed to pass the test

  - multiple correct answers possible

- Only the slides with the sign ⚠ will be on the exam

- We will be available for questions. Contact me