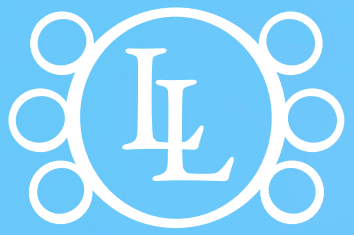


LOSCHMIDT
LABORATORIES

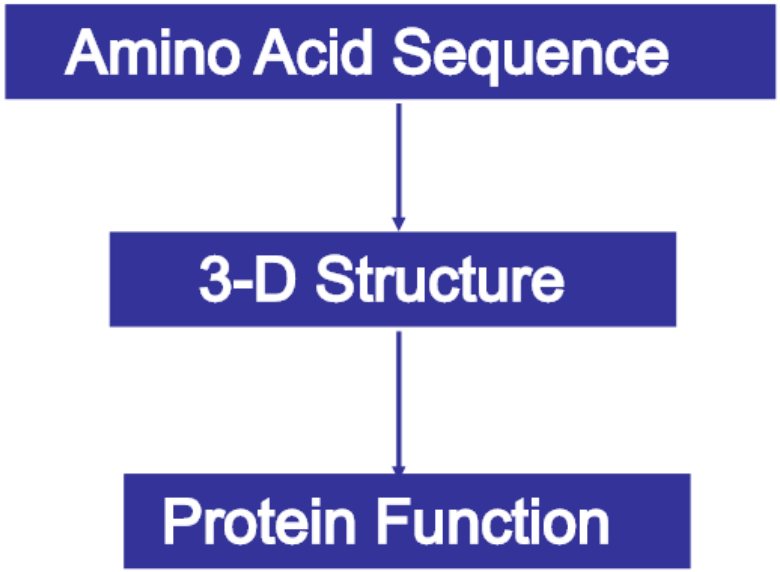
Applications of structural biology and bioinformatics

Outline

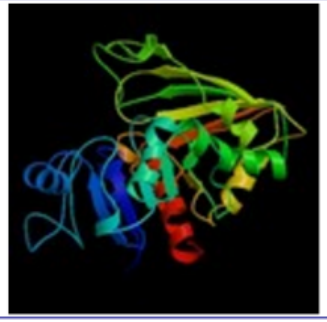
- Structural biology paradigm
- Applications of structural biology and bioinformatics
  - Biological research
  - Drug design
  - Protein engineering
- Summary
- Final remarks on the course

A structural biology paradigm…

- Sequence-Structure-Function



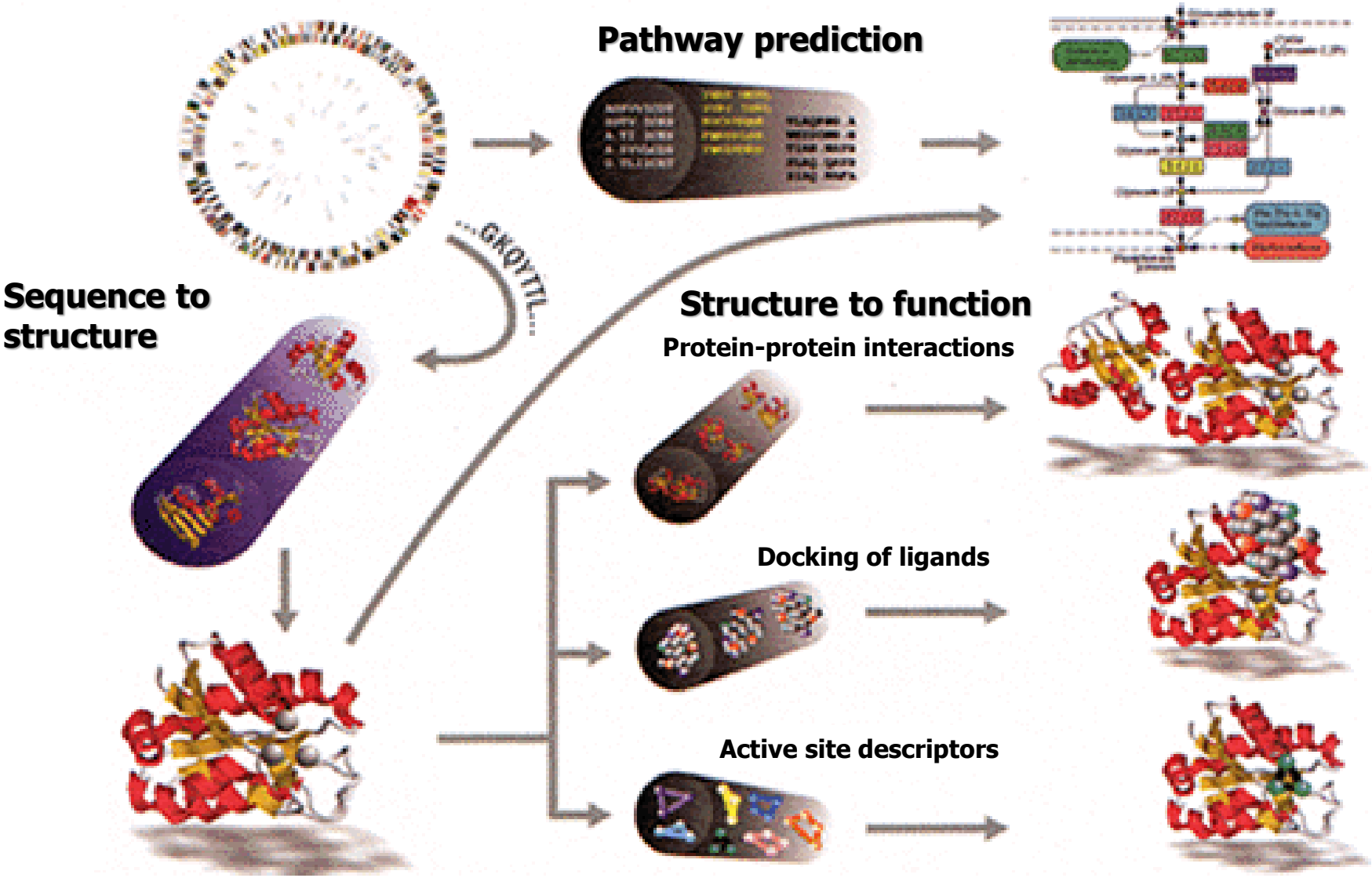
> 1NLG:_ NADP-LINKED GLYCERALDEHYDE-3-PHOSPHATE
EKKIRVAINGFGRIGRNFLRCWHGRQNTLLDVVAINDSGGVKQASHLLKYDSTLGTFAAD
VKIVDDSHISVDGKQIKIVSSRDPLQLPWKEMNIDLVIEGTGVFIDKVGAGKHIQAGASK
VLITAPAKDKDIPTFVVGVNEGDYKHEYPIISNASCTTNCLAPFVKVLEQKFGIVKGTMT
TTHSYTGDQRLLDASHRDLRRARAAALNIVPTTTGAAKAVSLVLPSLKGKLNGIALRVPT
PTVSVVDLVVQVEKKTFAEEVNAAFREAANGPMKGVLHVEDAPLVSIDFKCTDQSTSIDA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



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

- Challenges:
  - Determine structure from sequence
  - Determine function from sequence/3D structure
  - Modify function (by modifying sequence or external molecules)

Sequence-Structure-Function

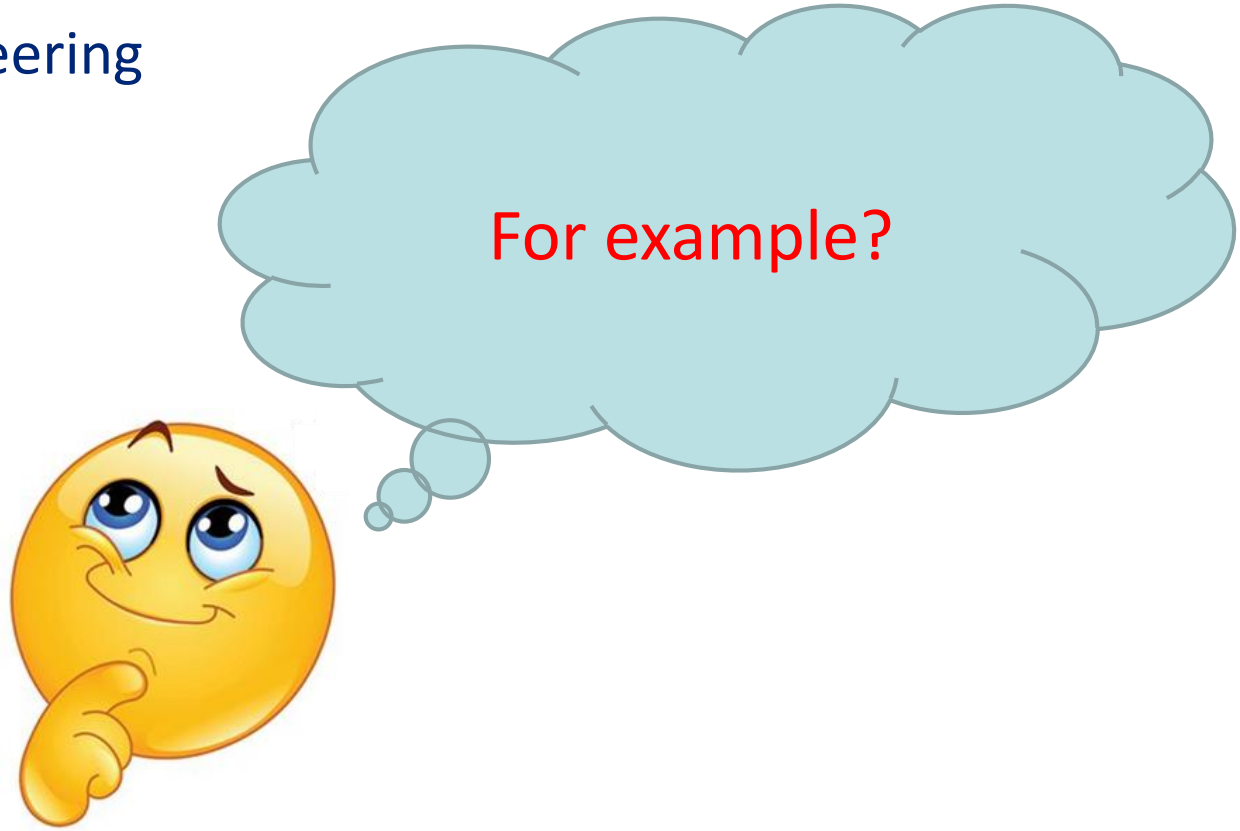


Applications of structural biology and bioinformatics

- Biological research
- Drug design
- Protein engineering

Applications of structural biology and bioinformatics

- Biological research
- Drug design
- Protein engineering

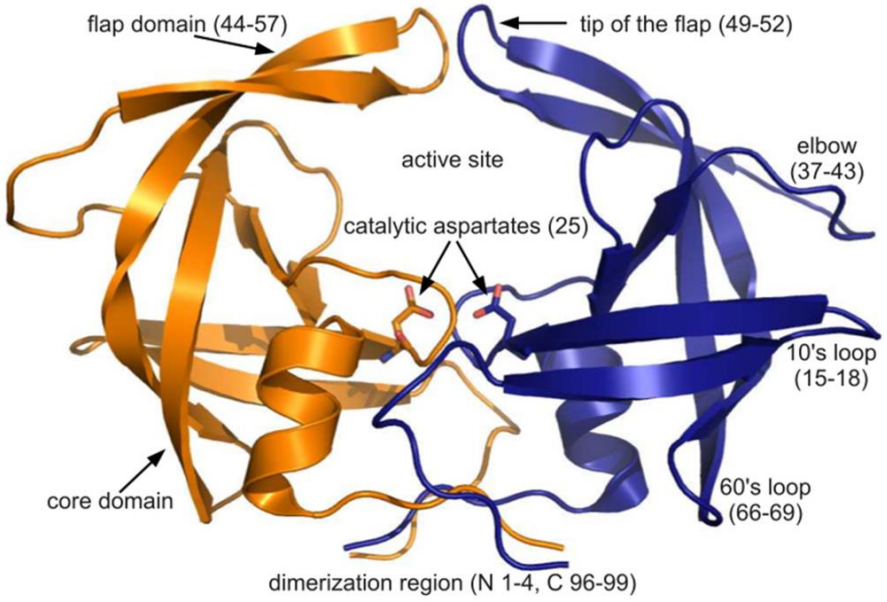


Biological research

- Drug resistance of HIV protease

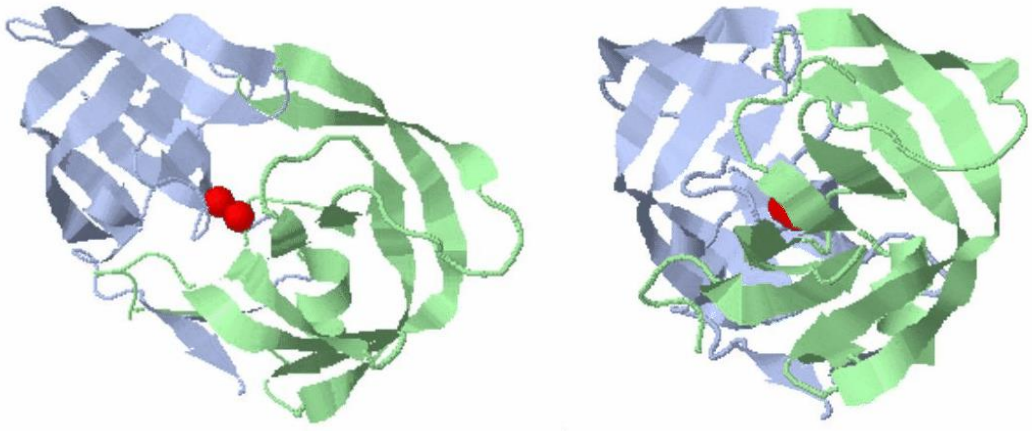
Drug resistance of HIV protease

- 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



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

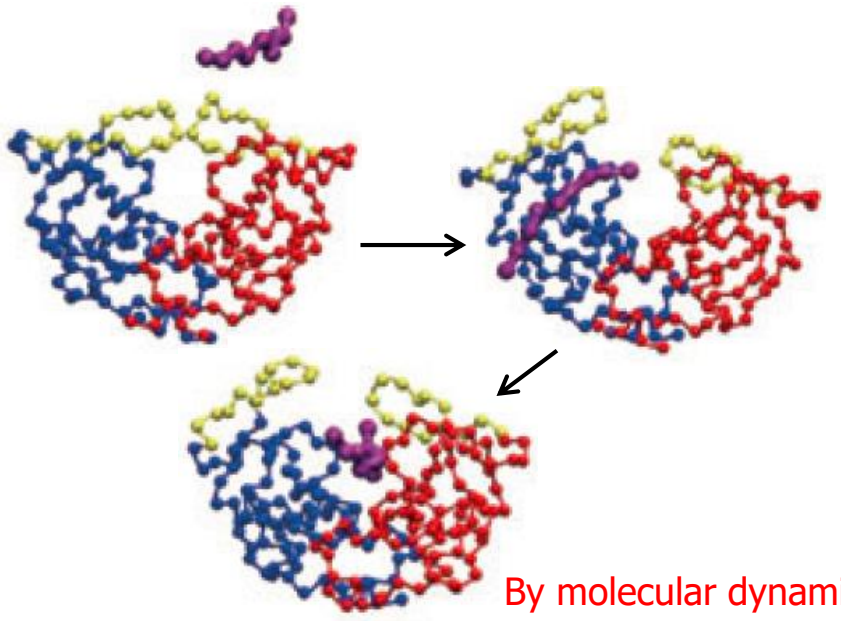


By comparing 2 crystal structures
(PDBs: 1HXW and 1TW7)

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

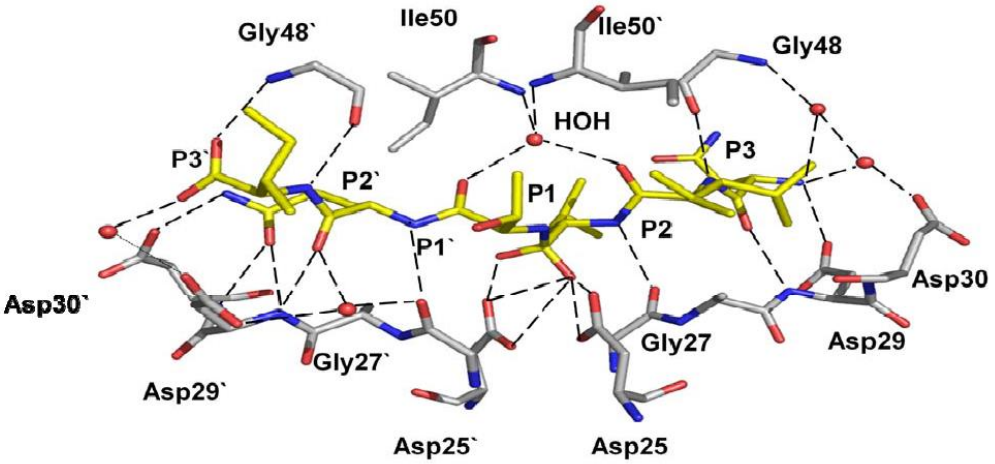
    is crucial for catalysis



By molecular dynamics

Drug resistance of HIV protease

- Protease inhibitors (PIs)
  - Introduced into clinical practice in 1995 known as antiretrovirals
  - Competitive inhibitors, designed to mimic the transition state of the substrate-enzyme complex
  - Binding affinity in nanomolar to picomolar range (very high)
  - Currently ~ 10 different inhibitors available
    - Darunavir, indinavir
    - Fosamprenavir, saquinavir
    - Tipranavir, ritonavir
    - Amprenavir, lopinavir
    - Nelfinavir, atazanavir



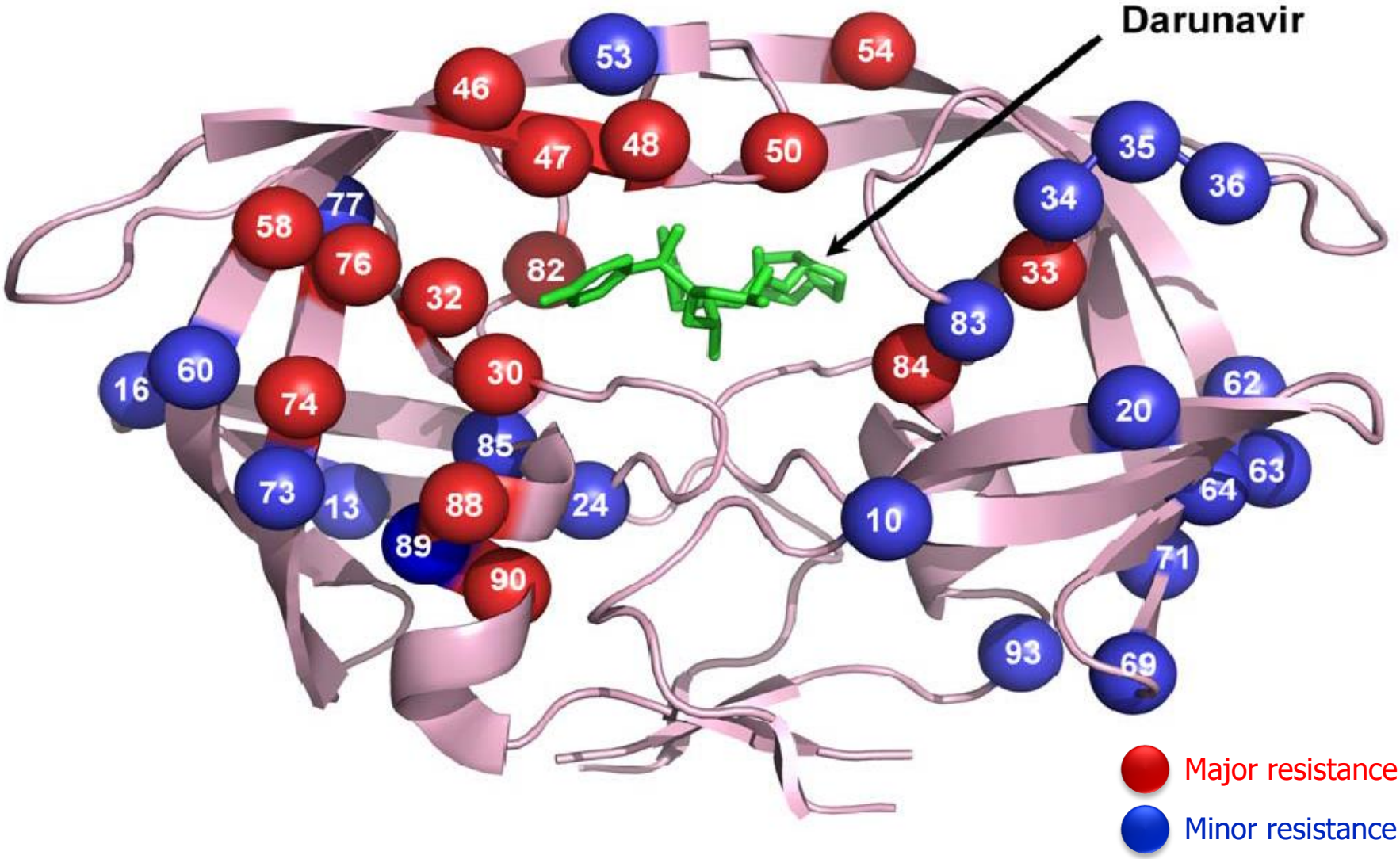
Drug resistance of HIV protease

- 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

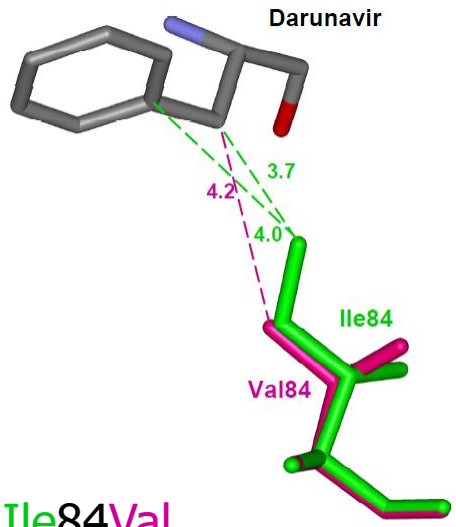


Drug resistance of HIV protease

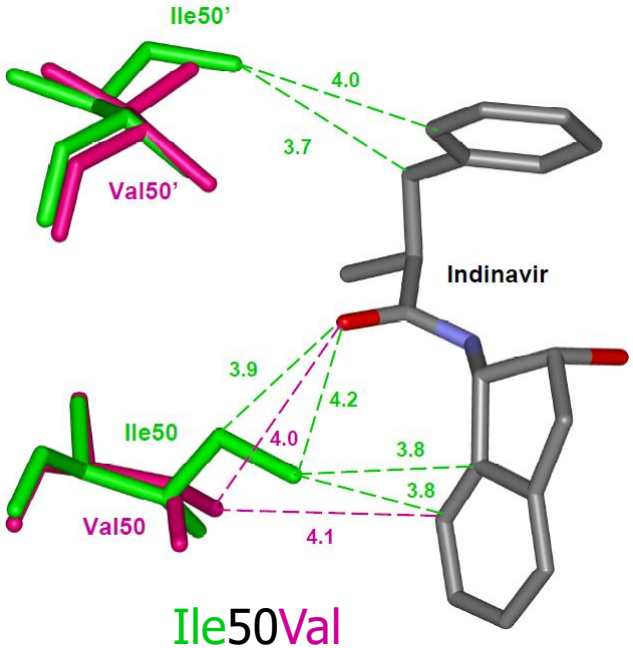
- Molecular mechanisms of drug resistance
  - Deduced from comparison of structures and activities of native and mutant proteases
- Several distinct mechanisms
  - Active site mutations
  - Mutations at dimer interface
  - Mutations at distal positions

Drug resistance of HIV protease

- Active site mutations
  - Mutation of single residue in the active site cavity eliminating direct interactions with inhibitor
  - Mutations are very conservative – ex: substitutions of hydrophobic amino acids



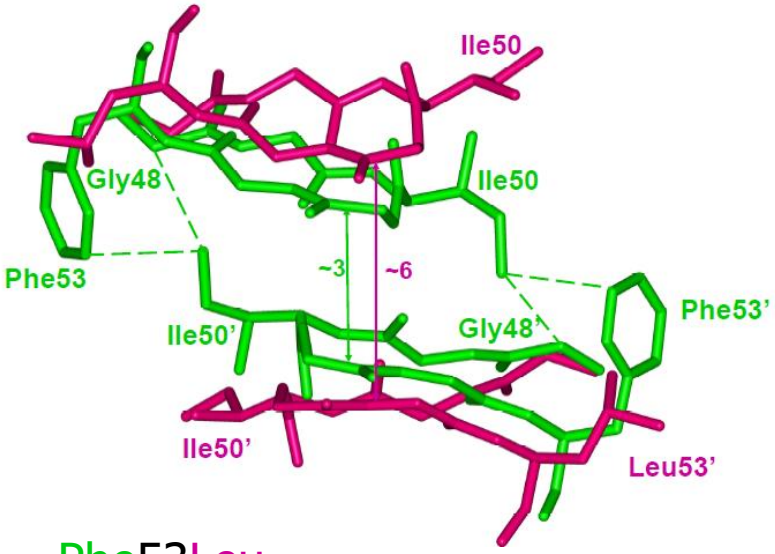
Ile84Val



Ile50Val

Drug resistance of HIV protease

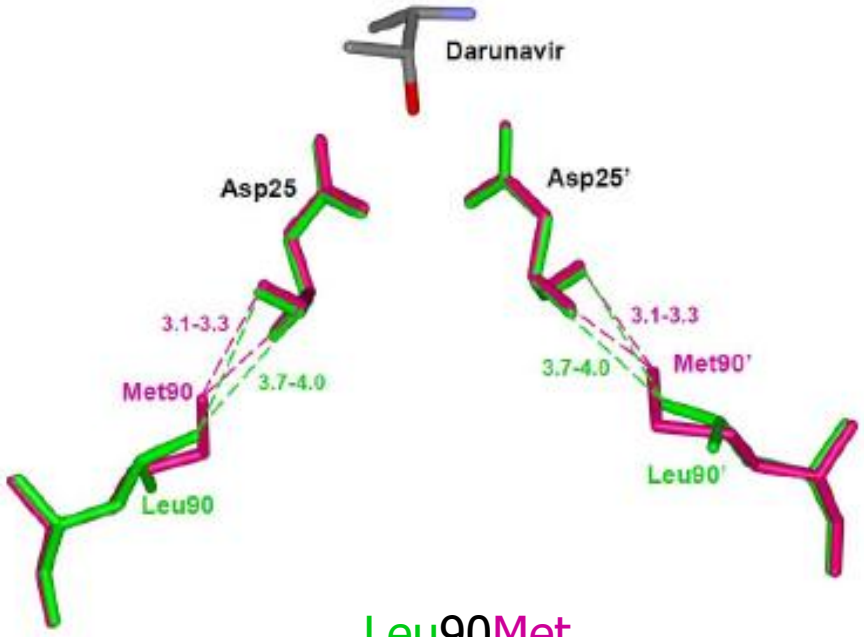
- Mutations at dimer interface
  - For example: Phe53Leu
    - Wider separation of the two flaps
    - Reduced stabilization of bound inhibitor



Phe53Leu

Drug resistance of HIV protease

- Mutations at distal positions
  - For example: Leu90Met
    - Promoted contacts with catalytic Asp25
    - Reduced interaction with inhibitor



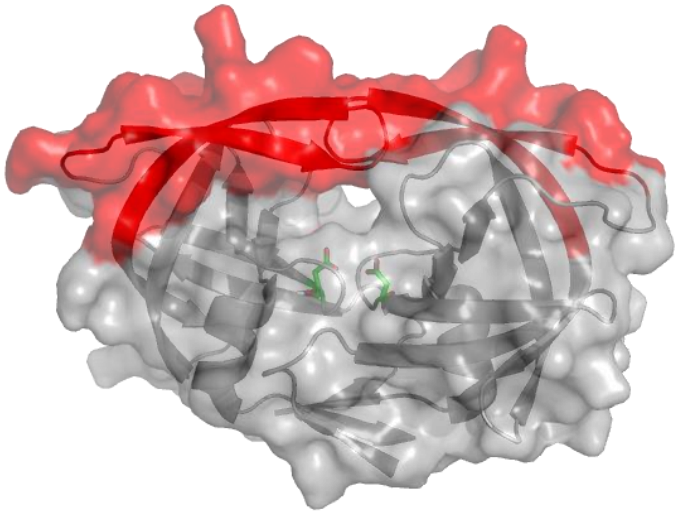
Leu90Met

Drug resistance of HIV protease

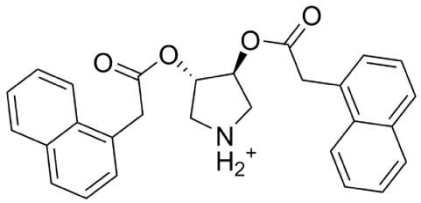
- Novel PIs for resistant HIV-1 protease
  - Inhibitors fitting within envelope formed by bound substrate
  - Inhibitors binding flaps or the dimer interface
  - Inhibitors targeting main chain and conserved regions of active site
  - Inhibitors targeting the gating mechanism

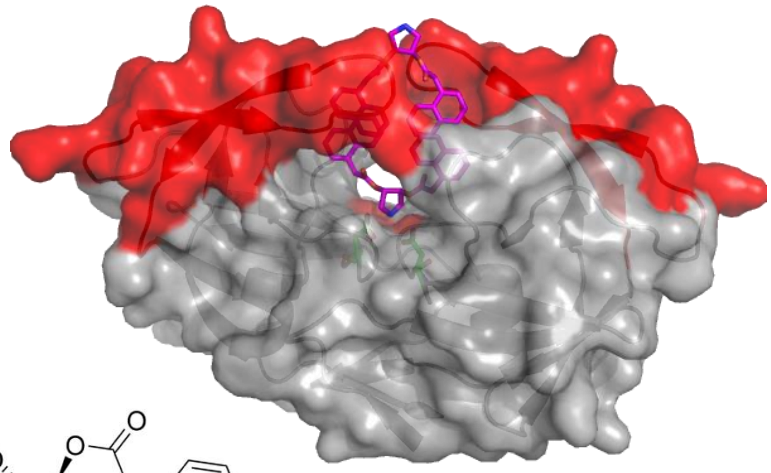
Drug resistance of HIV protease

- Novel PIs for resistant HIV-1 protease
  - Inhibitors targeting the gating mechanism
    - Stabilize the closed state
    - Stabilize the open state
    - Mixed interactions (AS and gating elements)



Closed conformation
(PDB ID: 1HVR)





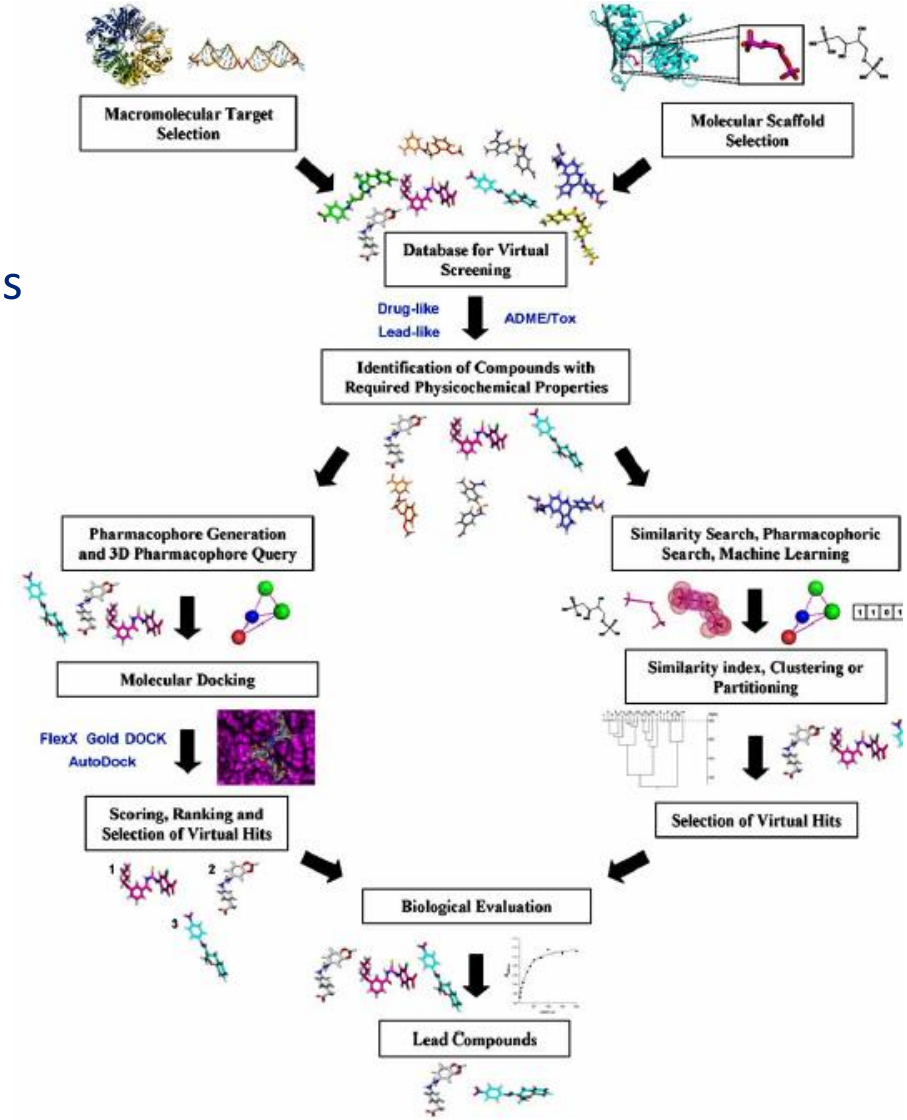
Open conformation
(PDB ID: 3BC4)

Drug design

- Virtual screening of inhibitors of endonuclease MUS81
- Selective inhibitor of LTA4H

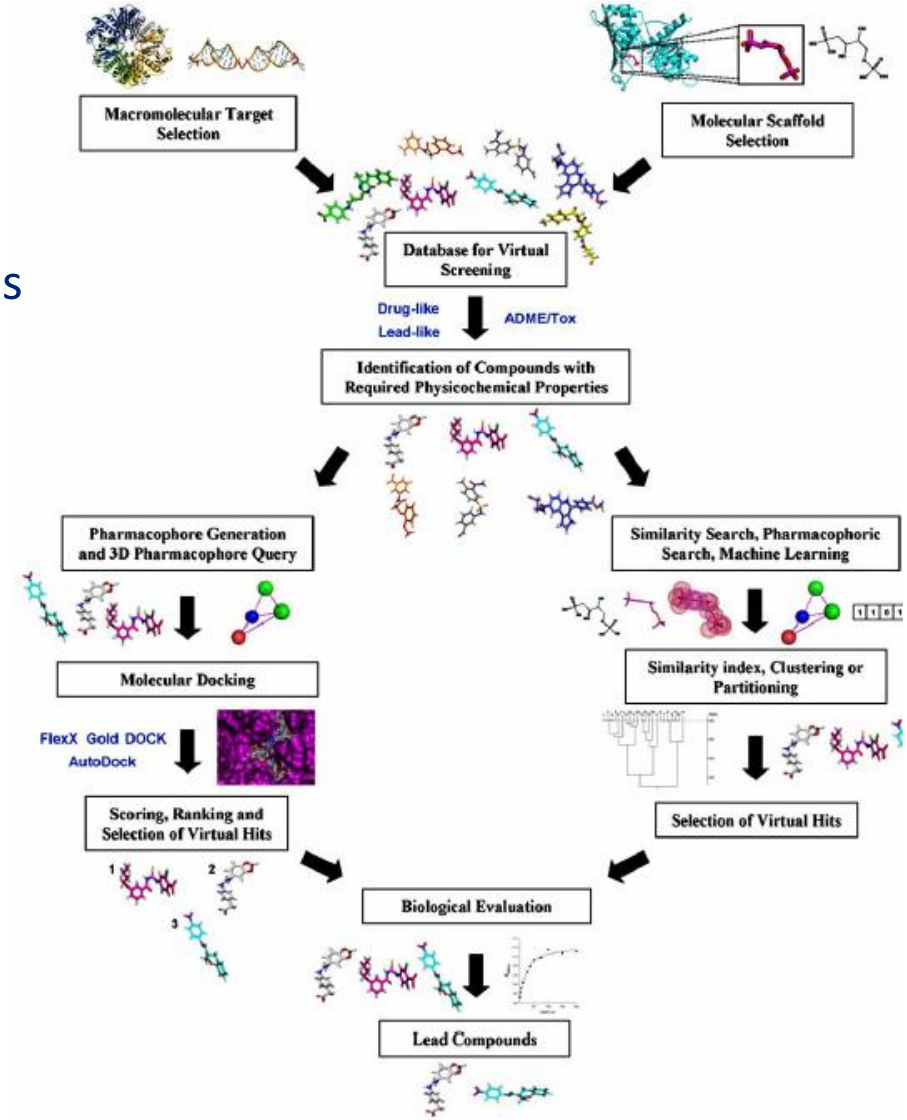
Drug design

- Methods of drug discovery
  - Ligand-based
    - Knowledge of active ligands
    - Search for similar ones



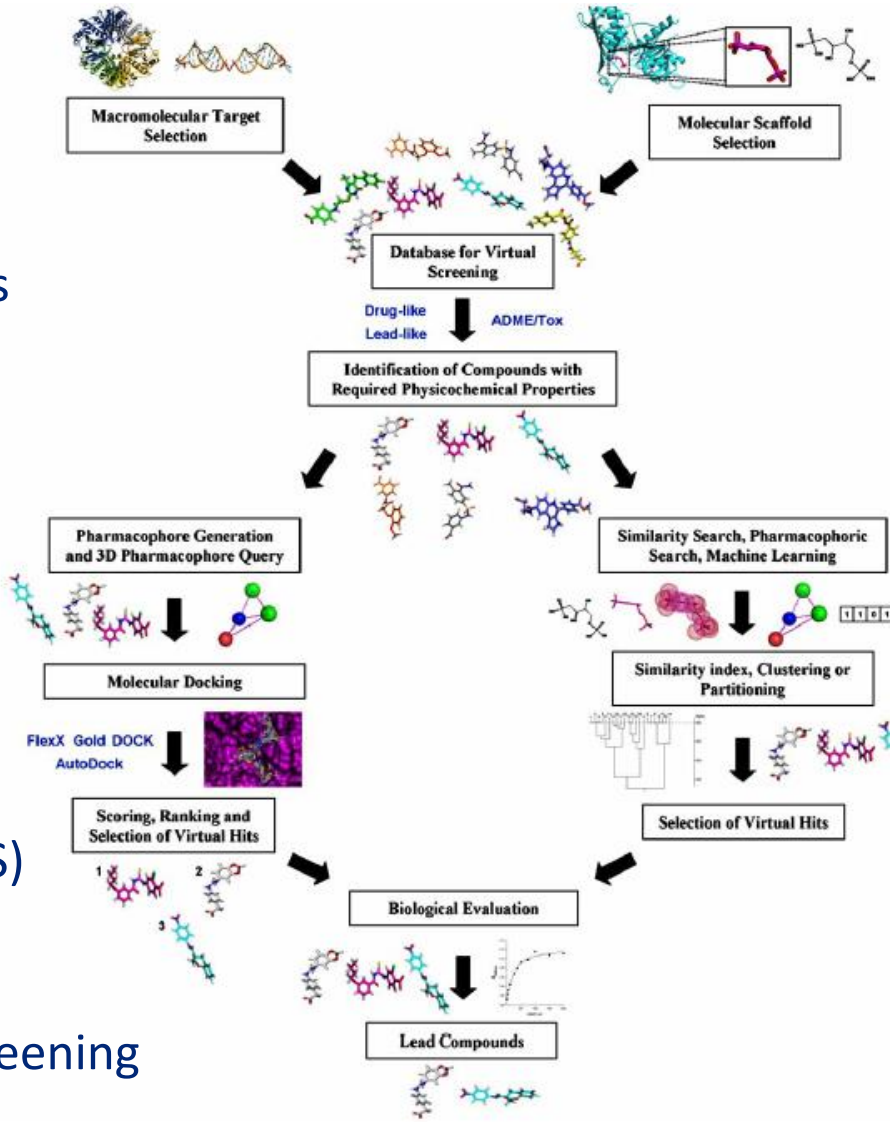
Drug design

- Methods of drug discovery
  - Ligand-based
    - Knowledge of active ligands
    - Search for similar ones
  - Structure-based
    - Knowledge of receptor
    - Search for strong binders
    - Molecular docking



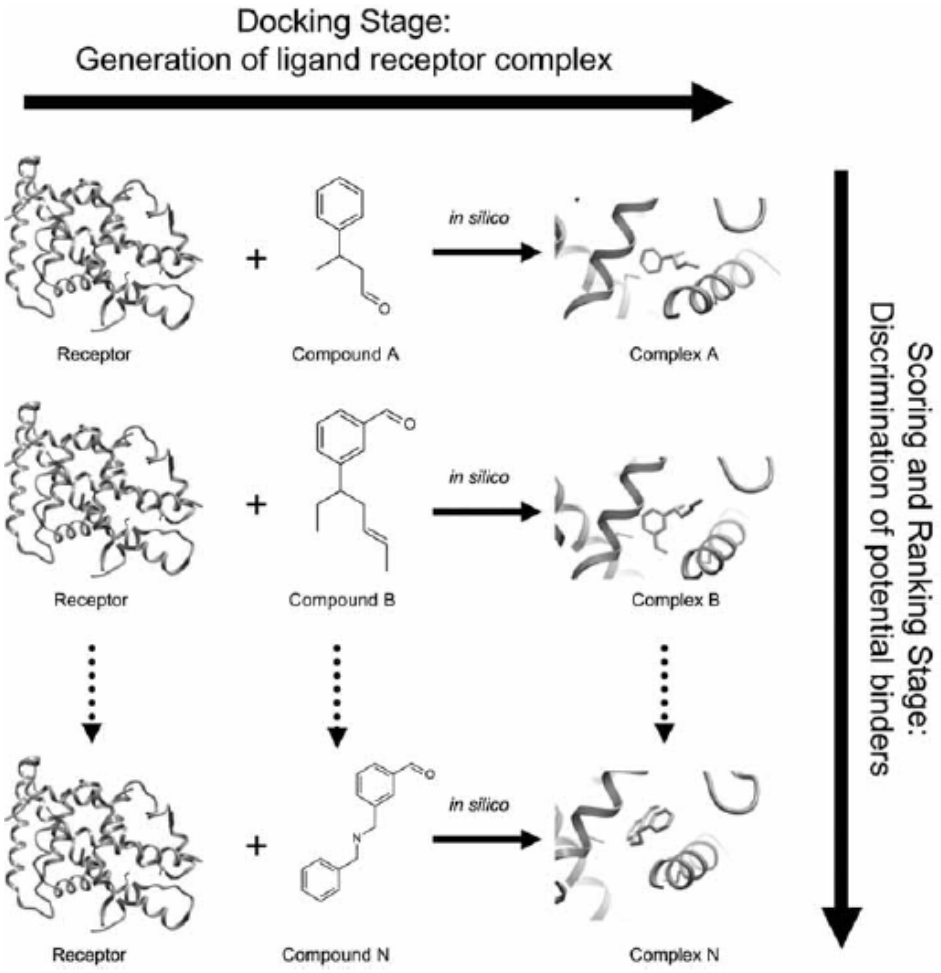
Drug design

- Methods of drug discovery
  - Ligand-based
    - Knowledge of active ligands
    - Search for similar ones
  - Structure-based
    - Knowledge of receptor
    - Search for strong binders
    - Molecular docking
  - High-throughput screening (HTS)
    - Large library of compounds
    - *Experimental* or *in silico* screening



Virtual screening

- Structure-based VS
  - Receptor-ligand docking
  - Often combined with HTS
  - Followed by hit optimization
  - Many success stories
  - Speed-up drug discovery
  - Lower the costs

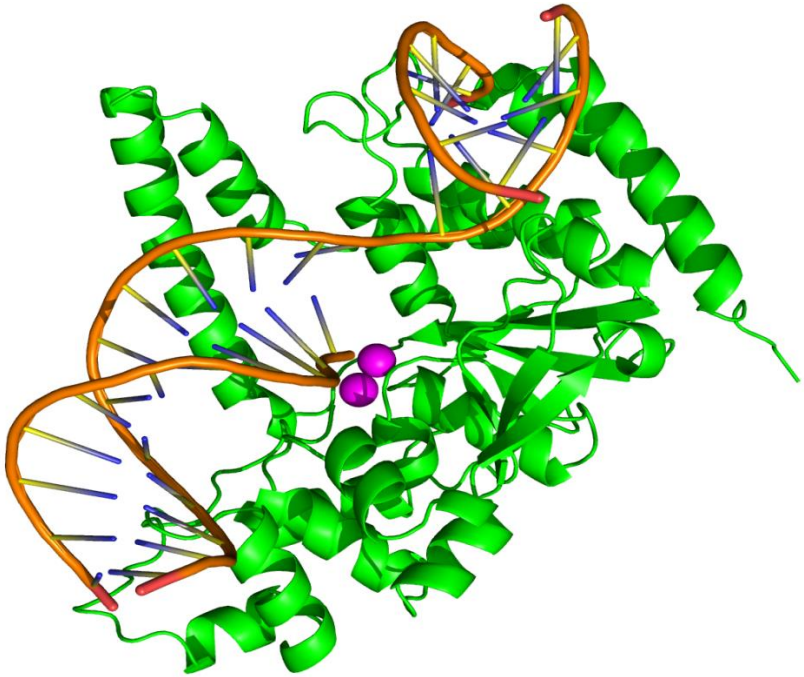


Virtual screening

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
  - Endonucleases are involved in DNA reparation
  - Help maintaining genomic stability
  - Cancer cells often have higher replication rates
  - MUS81 is a target for anti-cancer drug development

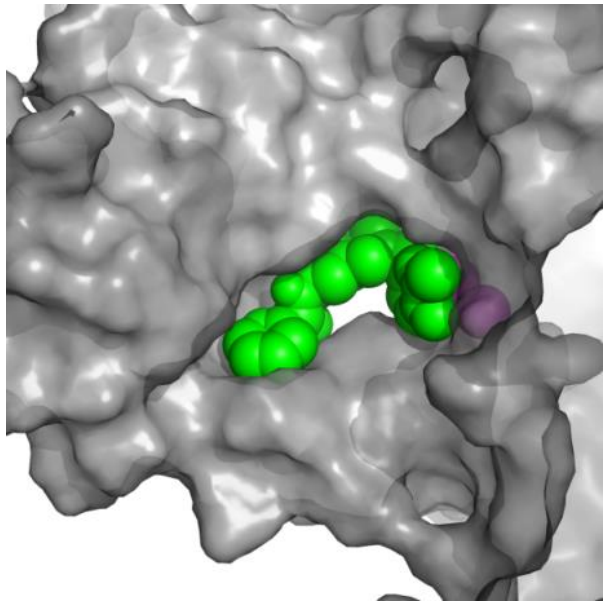


Inhibitors of endonuclease MUS81

- 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



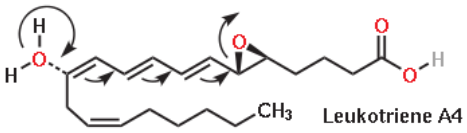
Inhibitors of endonuclease MUS81

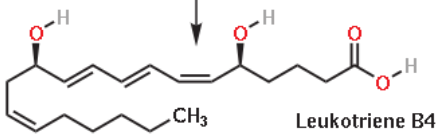
- Comparison

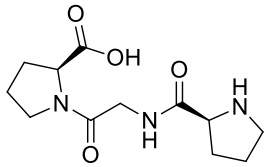
	HTS	VS
Equipment (Kč)	50,000,000	500,000
Testing		
Computational	-	140,000
Experimental	23,000	19
Costs (Kč)	2,000,000	40,000
Time	Weeks	Days
Results		
# of inhibitors	1	6
Best: IC_{50} (µM)	50	5

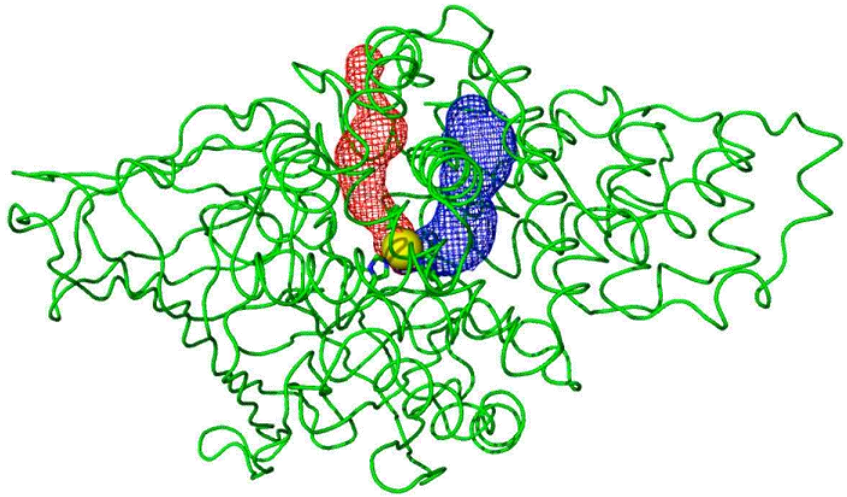
Selective inhibitor of LTA4H

- Leukotriene A4 hydrolase/aminopeptidase (LTA4H)
  - Involved in chronic inflammatory and immunological 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



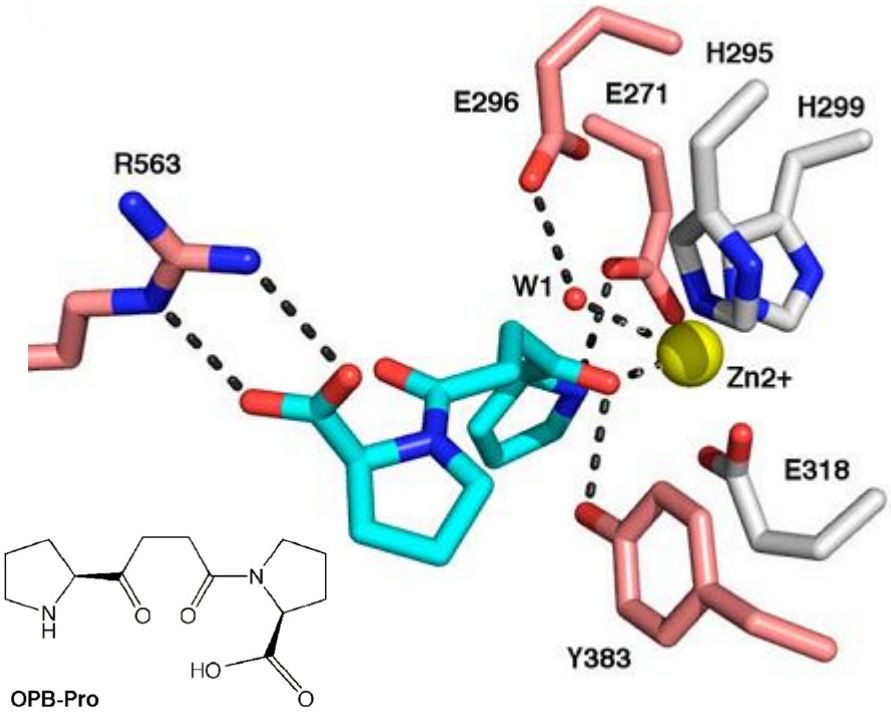


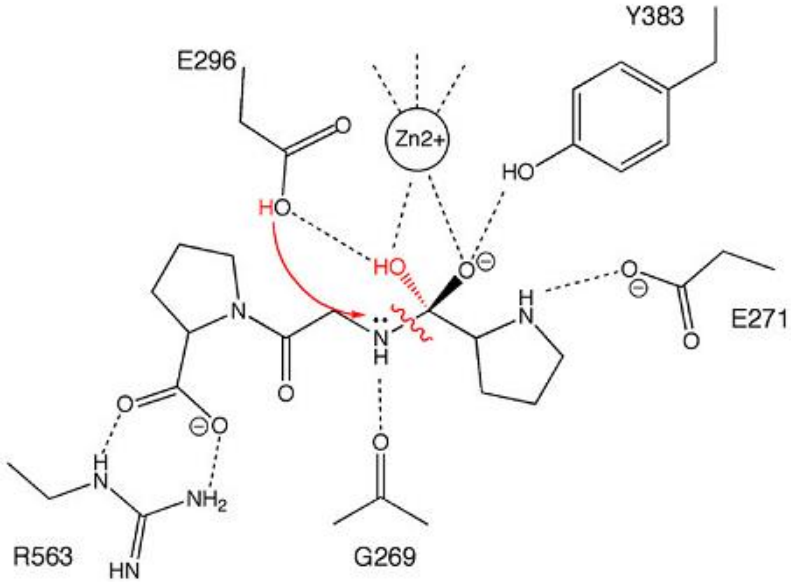




Selective inhibitor of LTA4H

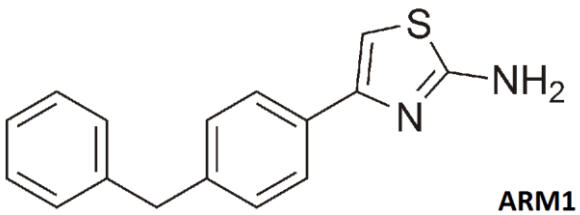
- Leukotriene A4 hydrolase/aminopeptidase (LTA4H)
  - Structural studies (crystallography) with a tripeptide analogue revealed the aminopeptidase mechanism



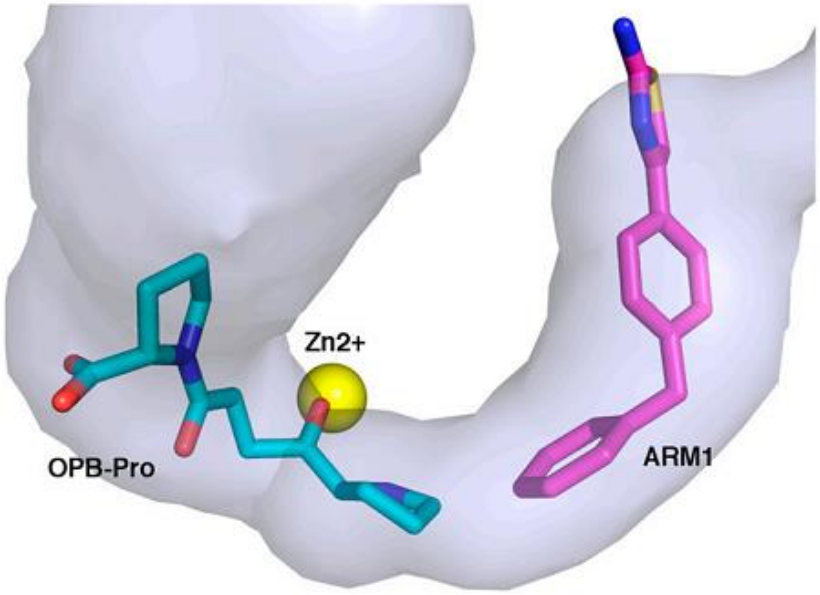


Selective inhibitor of LTA4H

- 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



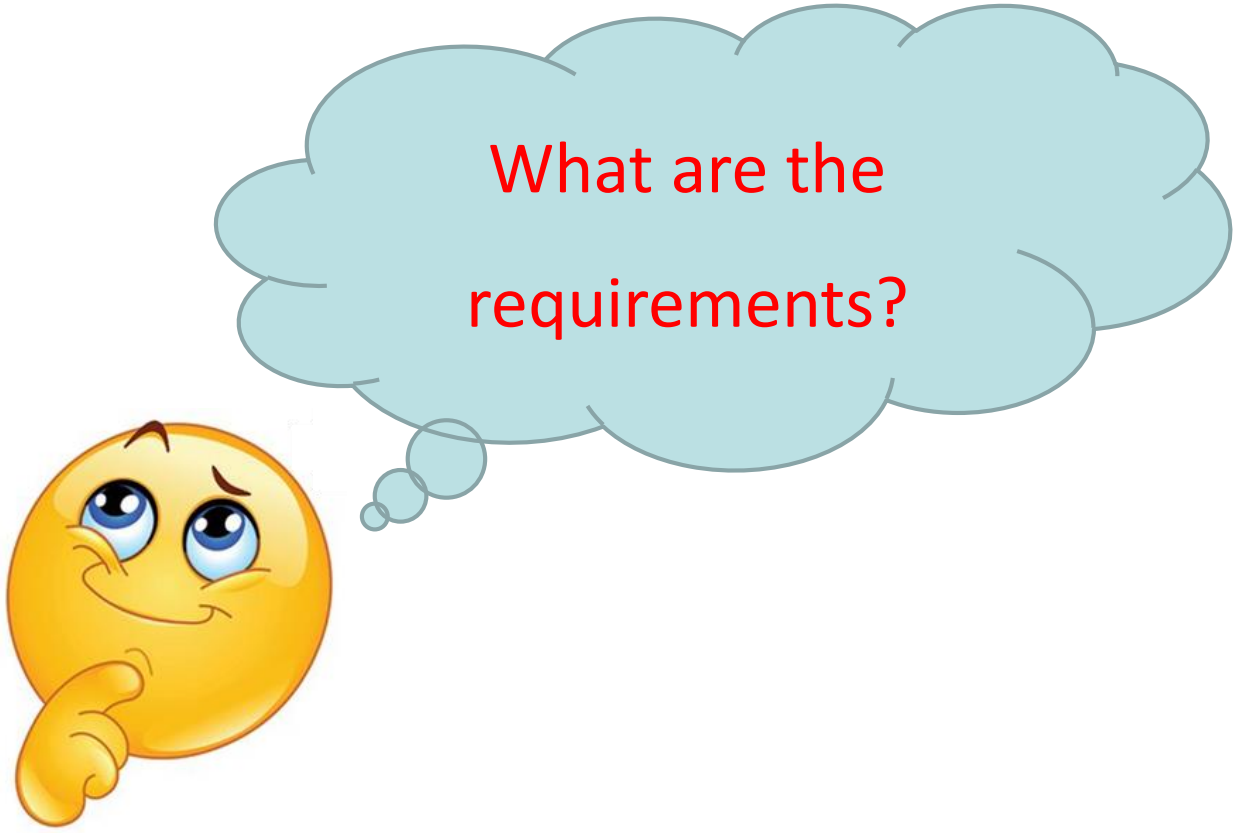
ARM1



Protein engineering

- Stabilization of dehalogenase
- Dehalogenase activity
- Lipase enantioselectivity
- *De novo* design of a Diels-Alderase

Enzymes: practical applications?



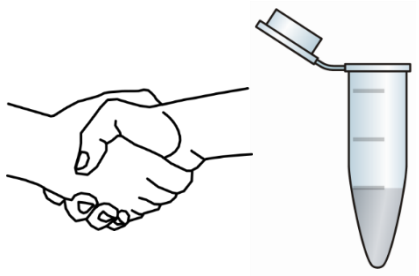
Enzymes: practical applications?

- Ability to catalyse a desirable reaction
- Stable under operating conditions
- Soluble expression

Enzymes: practical applications?

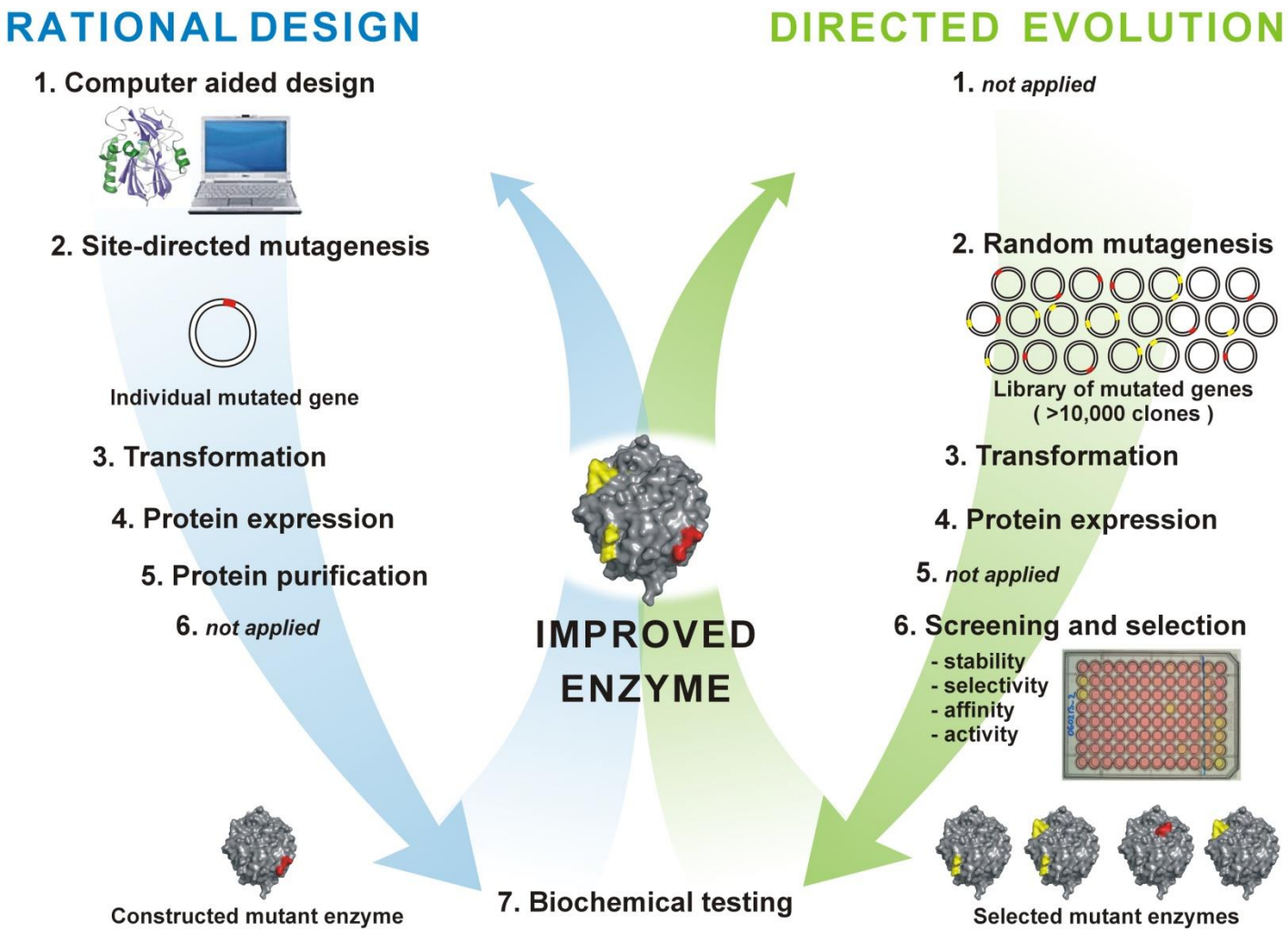
- Ability to catalyse a desirable reaction
- Stable under operating conditions
- Soluble expression

Protein engineering process



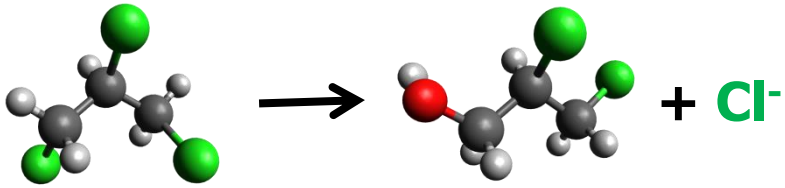
- Improvement of activity or selectivity
- Robust stabilization of proteins
- Design of more soluble proteins

Different approaches



Stabilization of dehalogenase

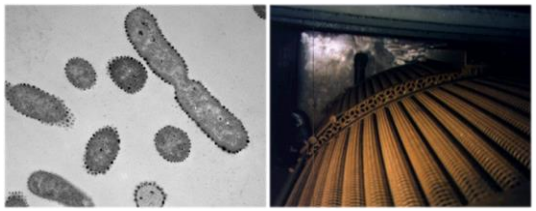
- Dehalogenase DhaA
  - Bacterial origin
  - Hydrolytic cleavage of C-X bond
  - Multiple biotechnological applications



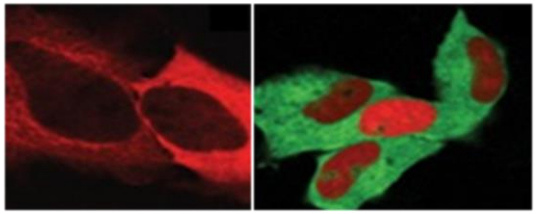
By-product recycling

Biosensing

Bioremediation



Cell imaging & protein analysis

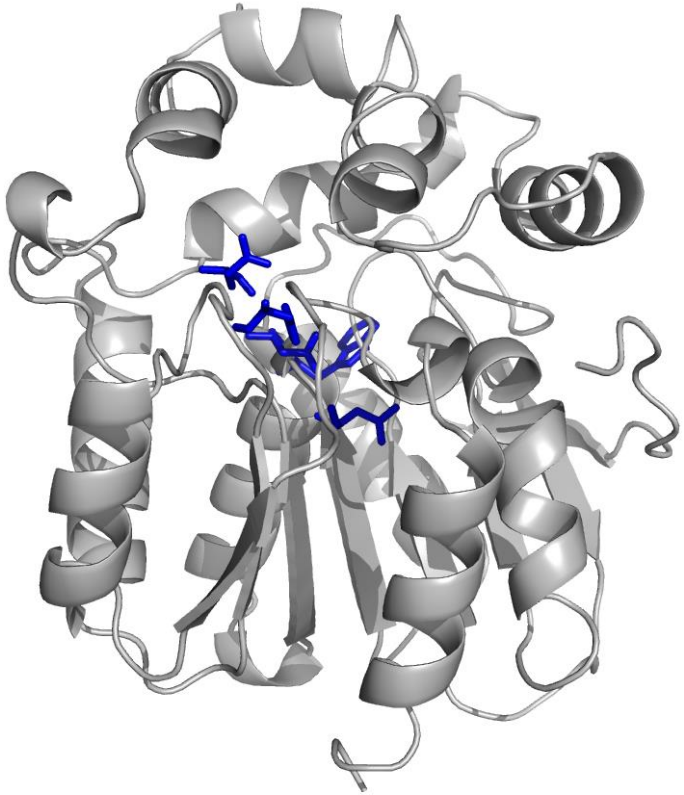


Biocatalysis

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

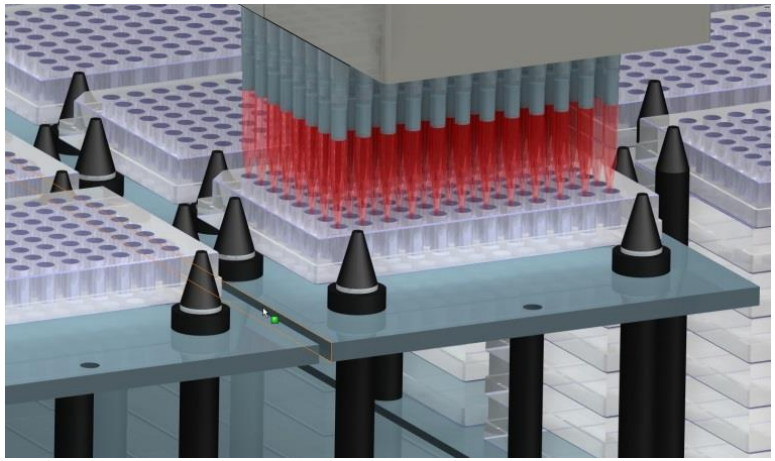


Stabilization of dehalogenase

- 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:

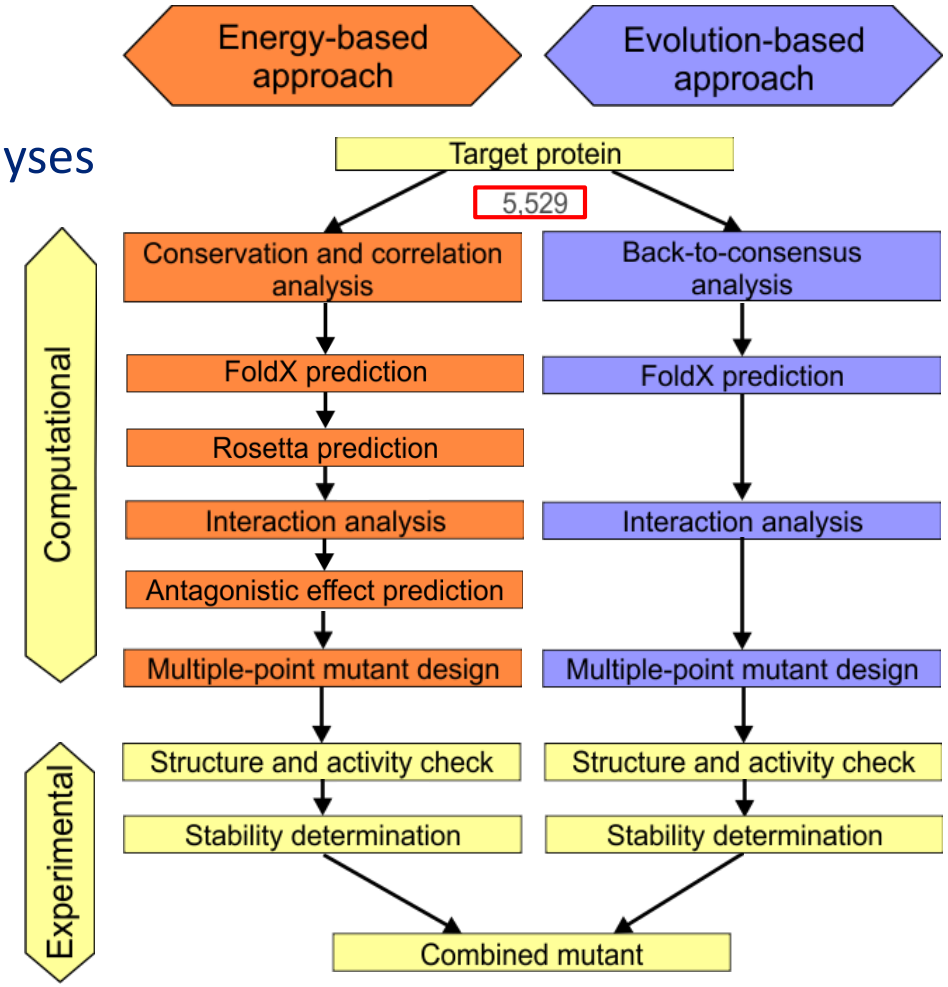
    T_m = 67 °C (18 °C ↑)

    $\tau_{1/2}$ = 36 h (ca. 36 h ↑)



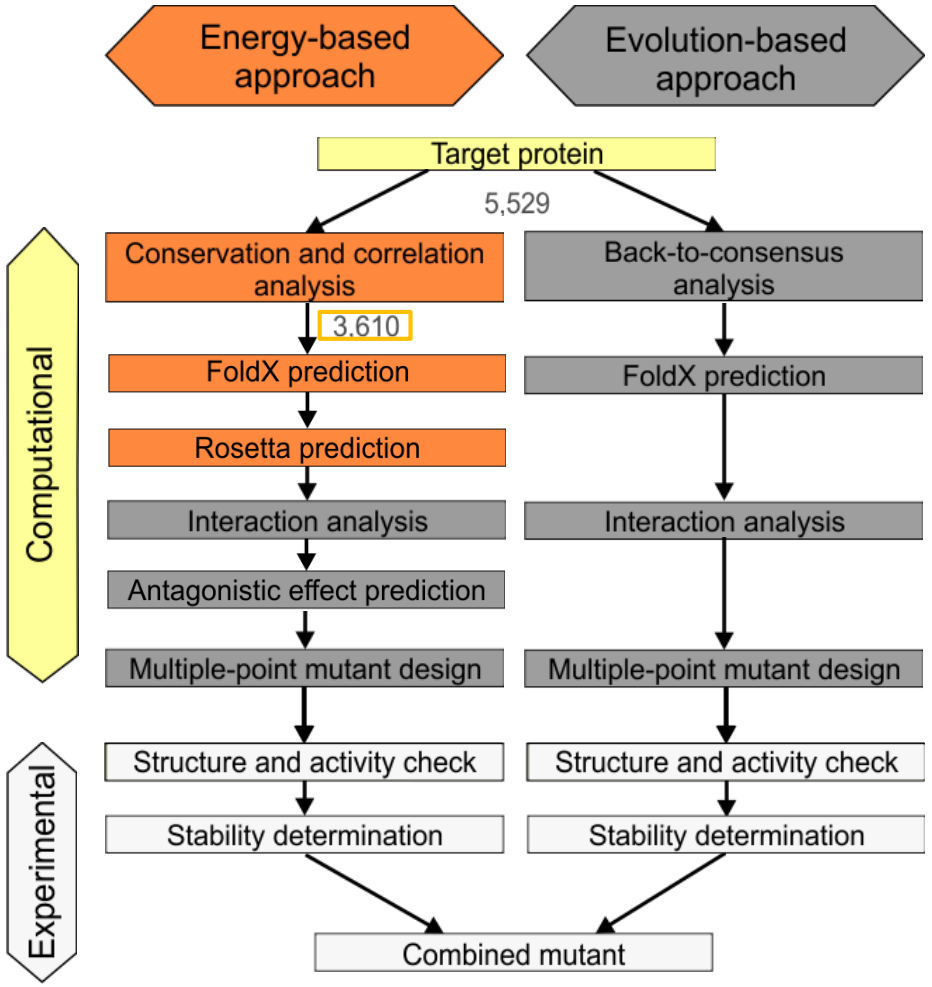
Stabilization of dehalogenase

- Rational design
  - FIREPROT method
  - Structure and sequence analyses
  - ~5,500 possible mutants



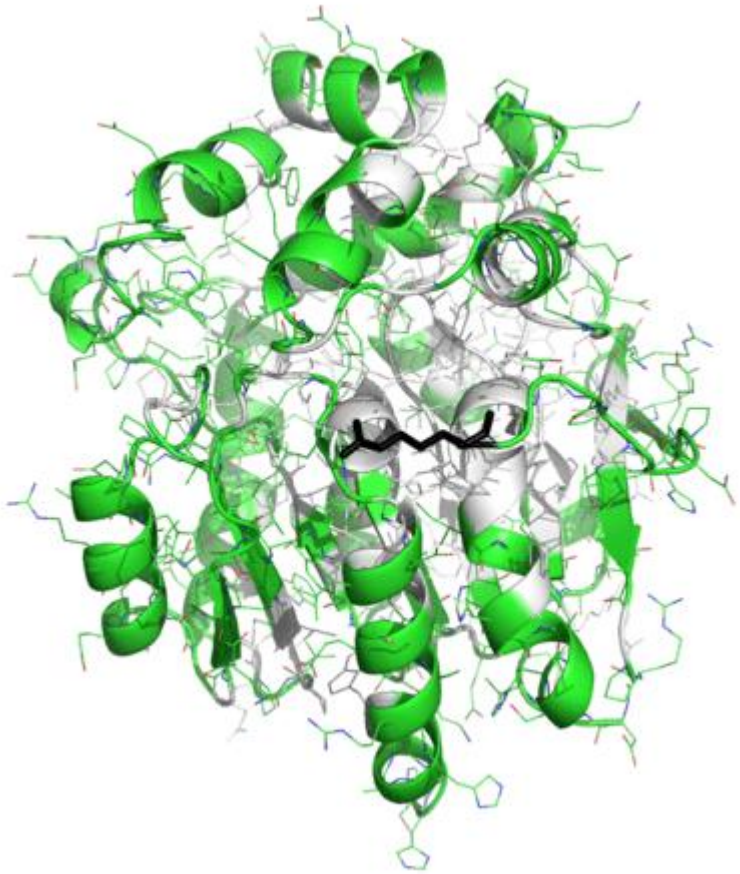
Stabilization of dehalogenase

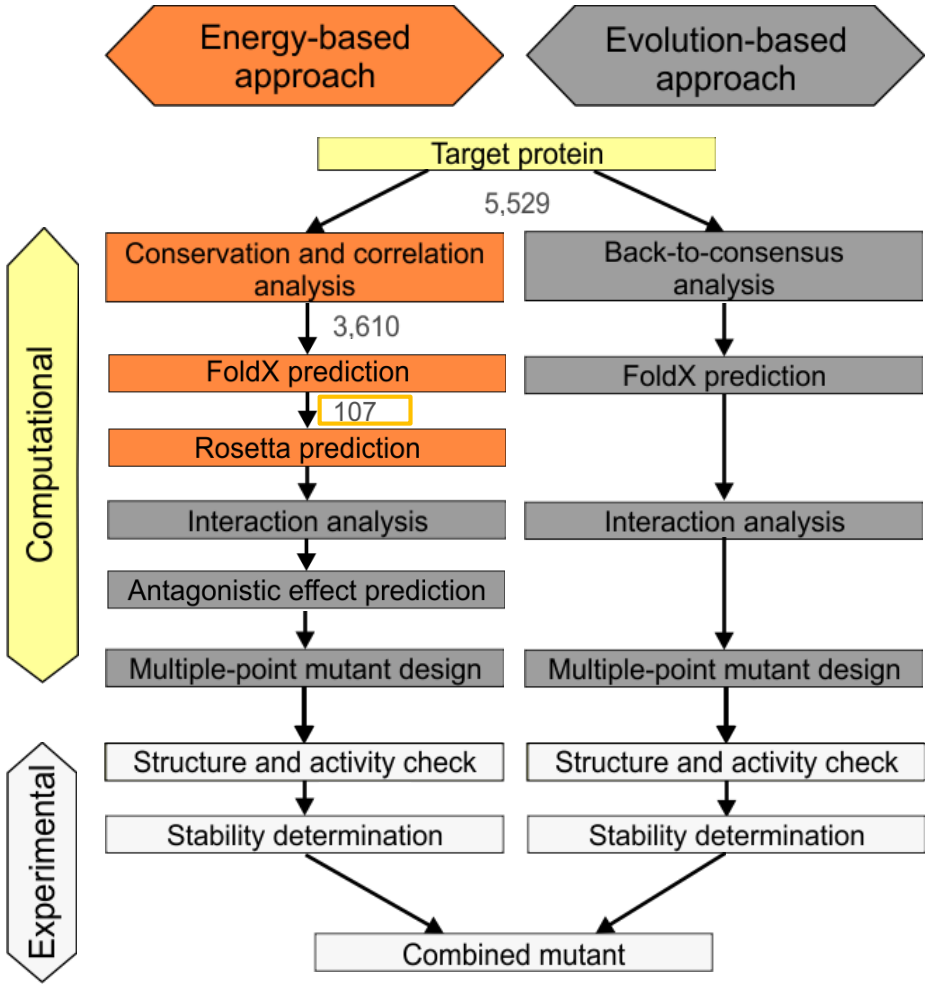
- Rational design
  - FIREPROT method



Stabilization of dehalogenase

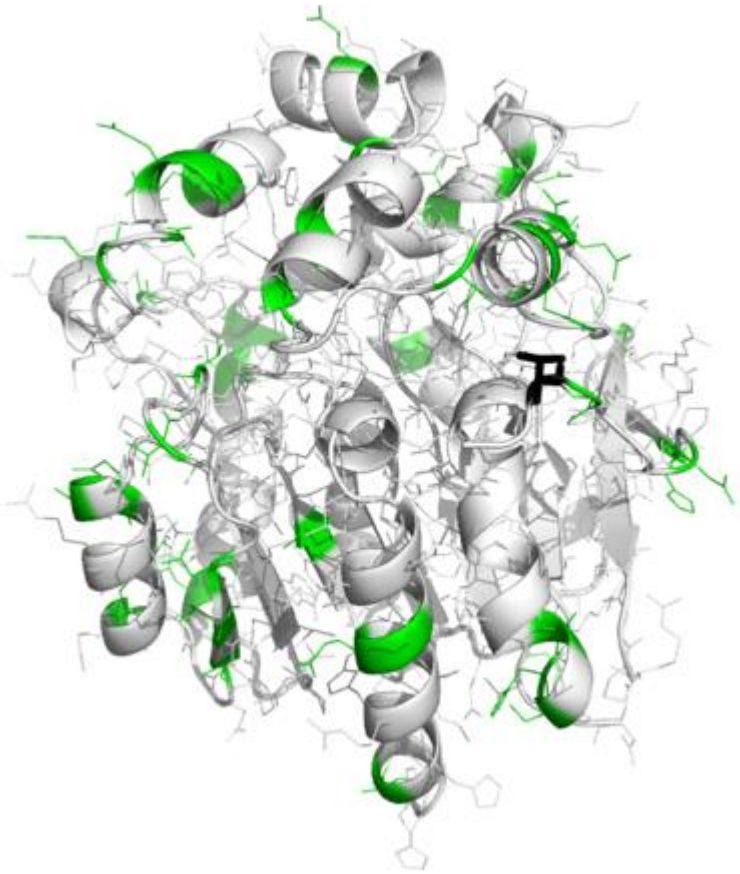
- Rational design
  - FIREPROT method

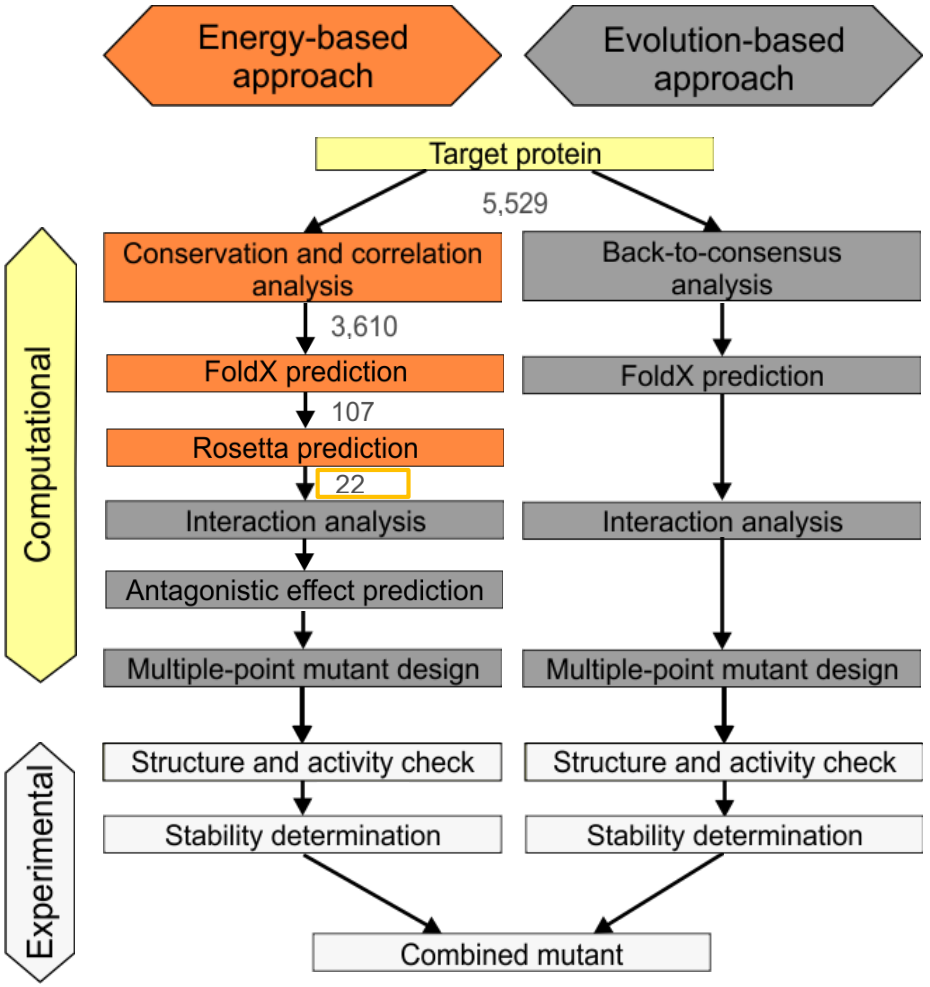




Stabilization of dehalogenase

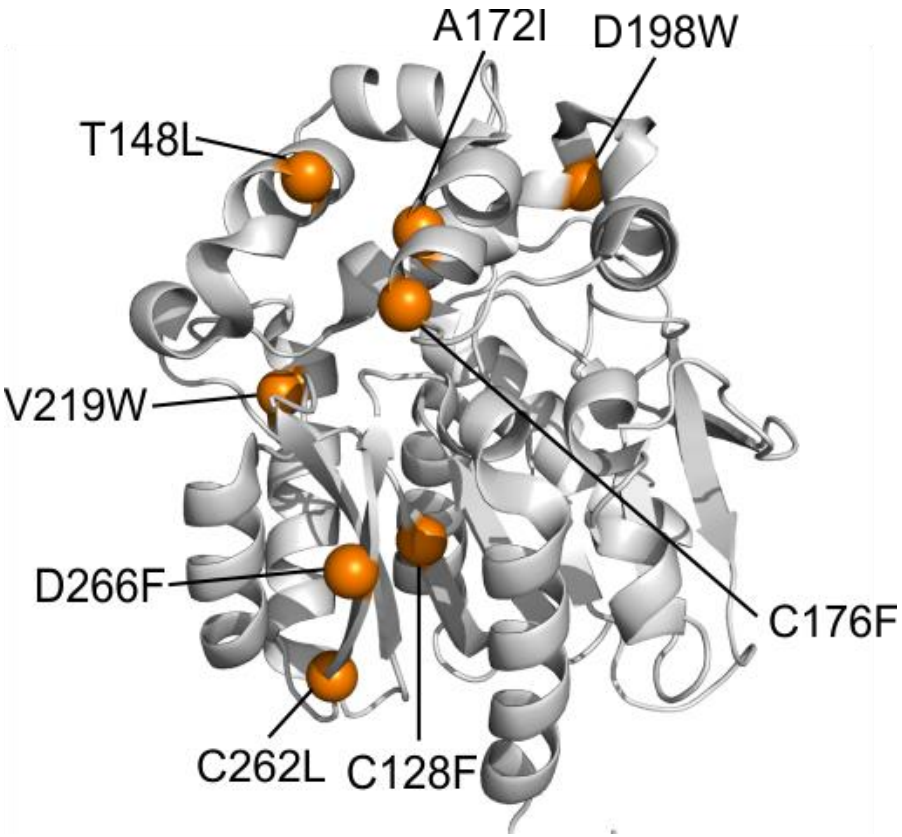
- Rational design
  - FIREPROT method

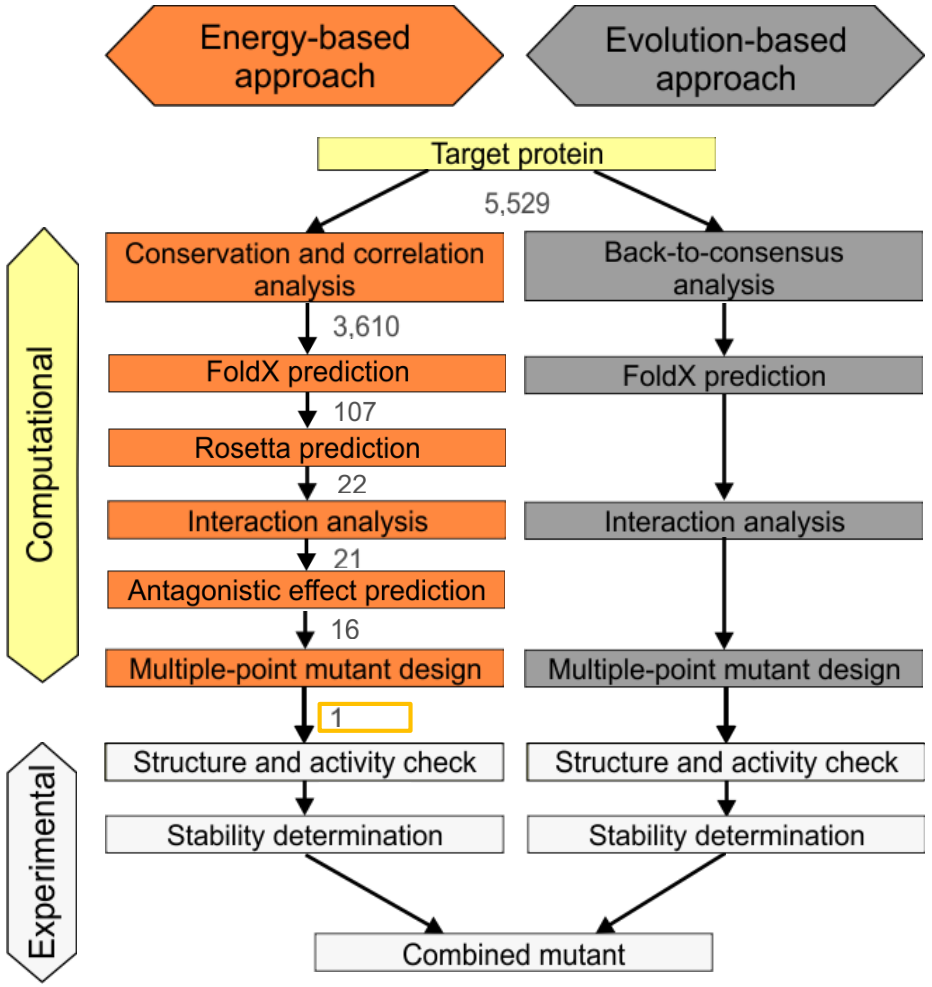




Stabilization of dehalogenase

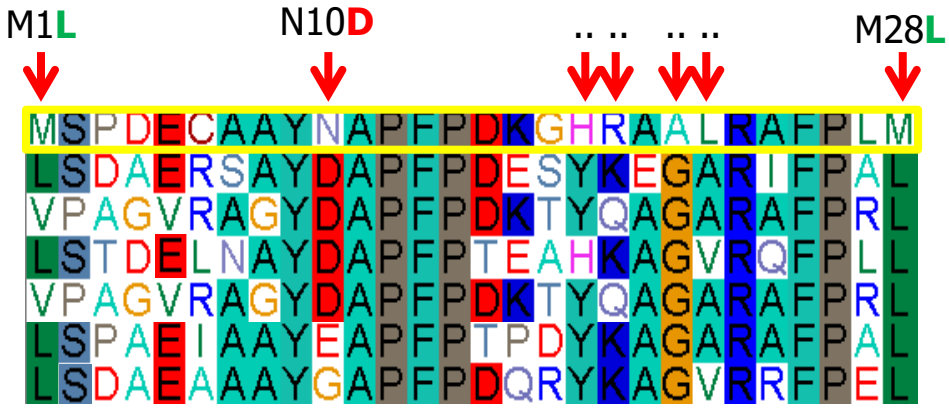
- Rational design
  - FIREPROT method

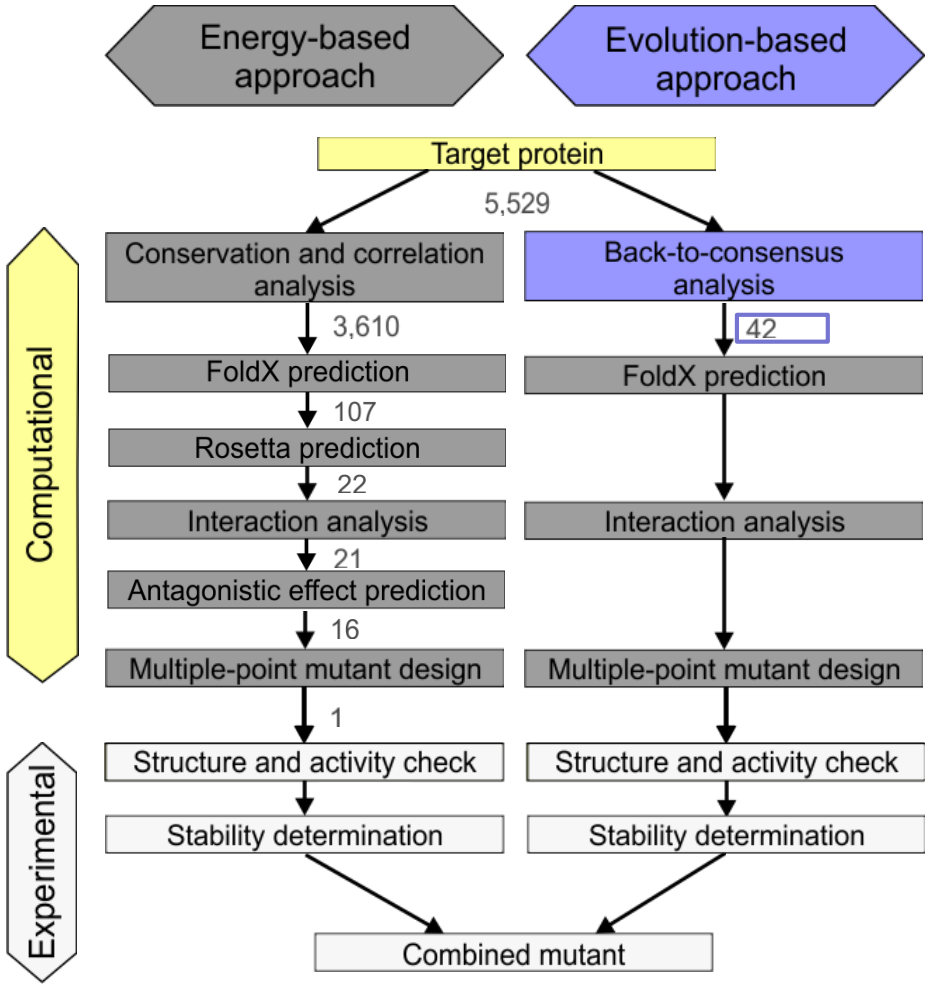




Stabilization of dehalogenase

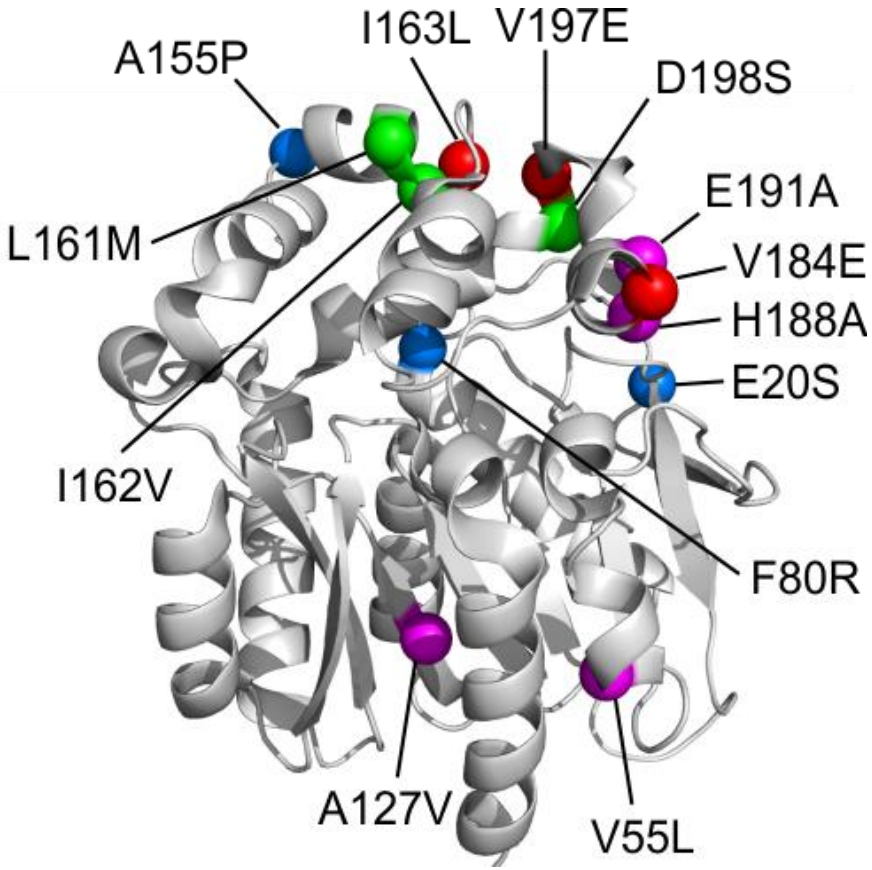
- Rational design
  - FIREPROT method

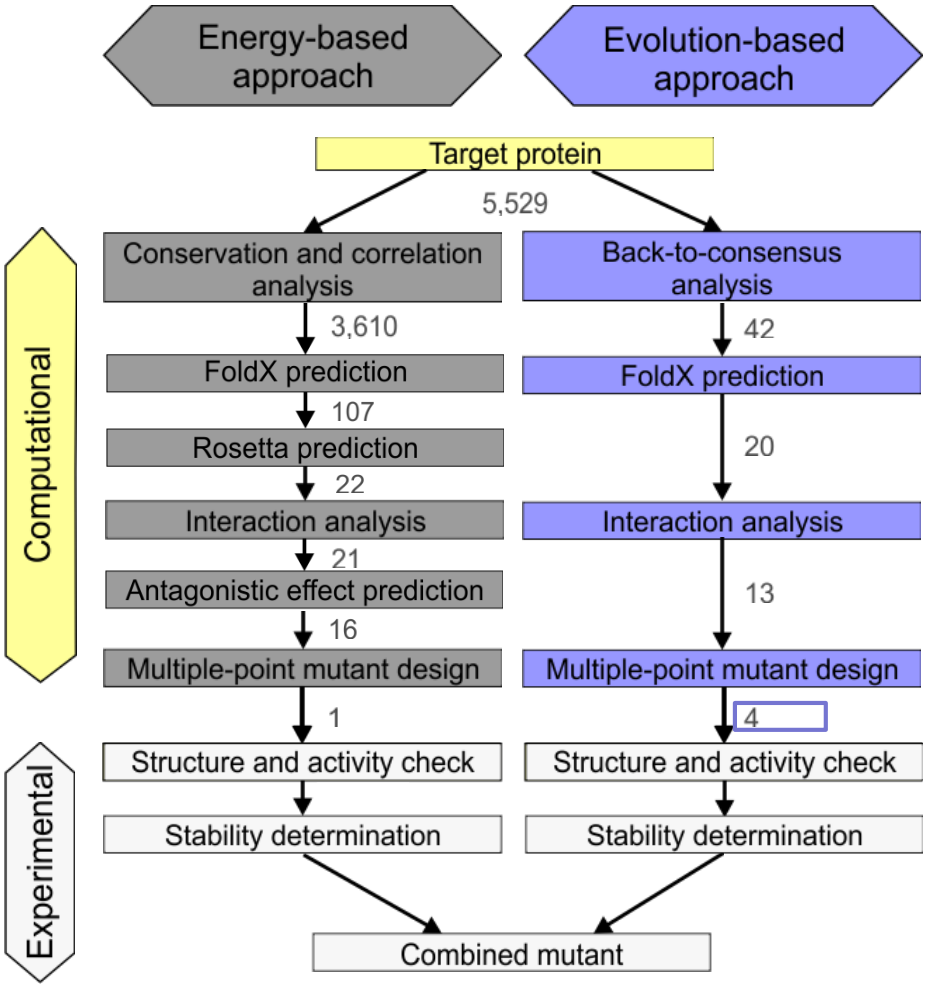




Stabilization of dehalogenase

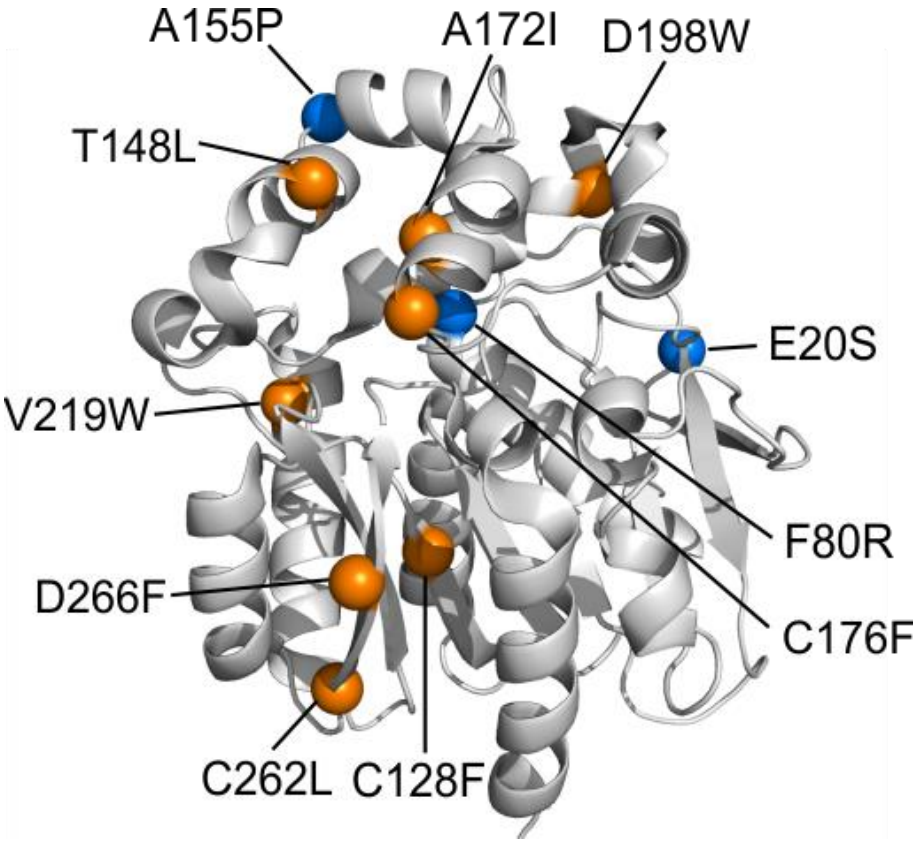
- Rational design
  - FIREPROT method

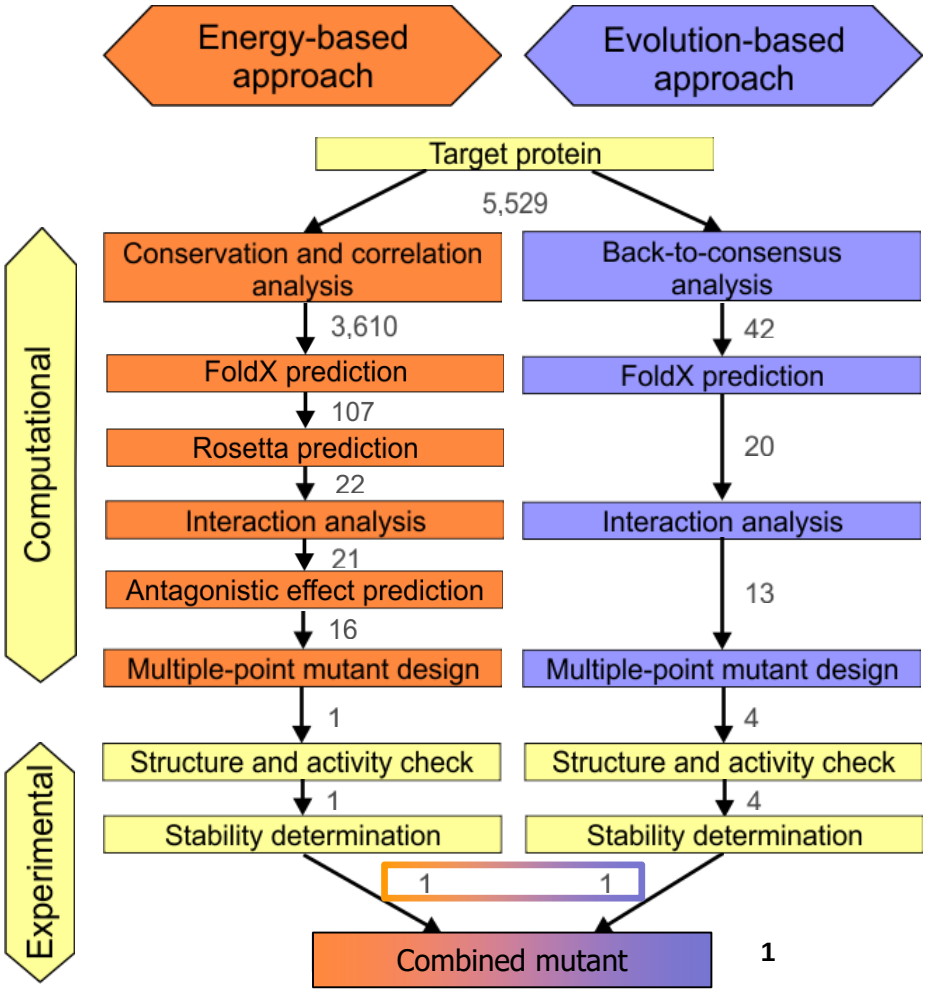




Stabilization of dehalogenase

- Rational design
  - FIREPROT method





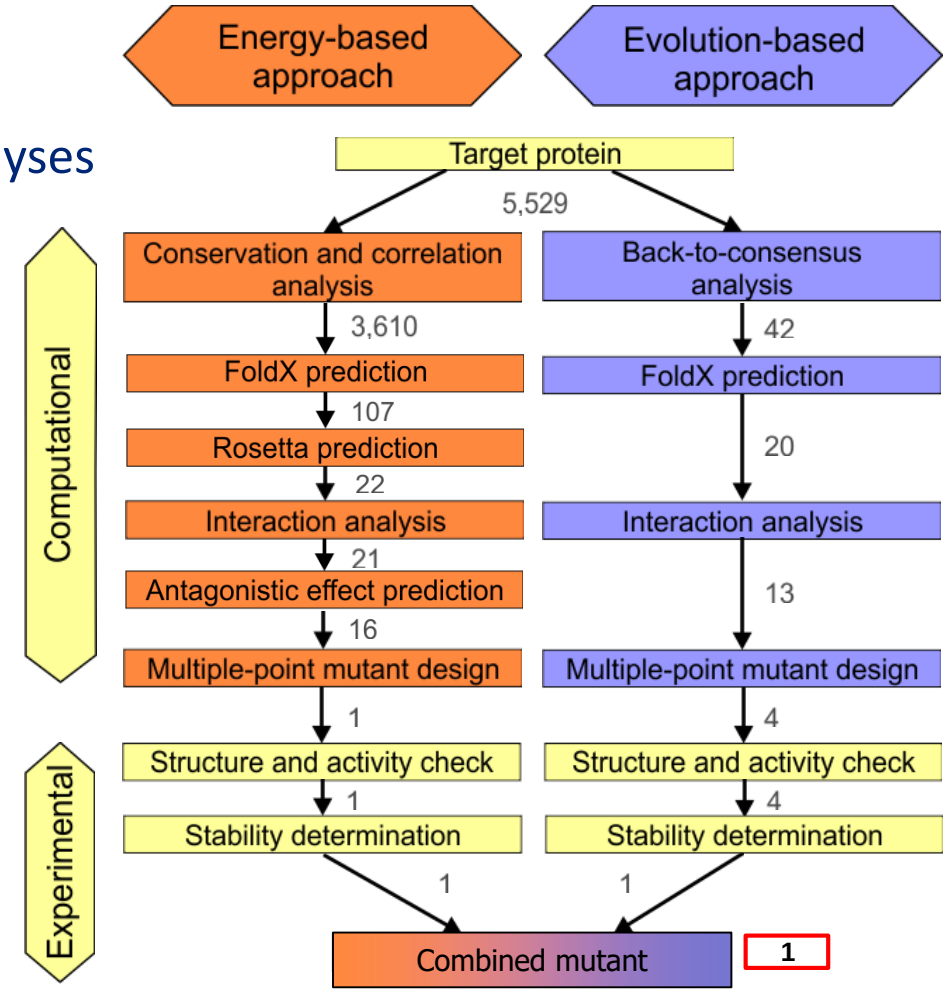
Stabilization of dehalogenase

- 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 ↑)

    $\tau_{1/2}$ = 72 h (ca. 72 h↑)



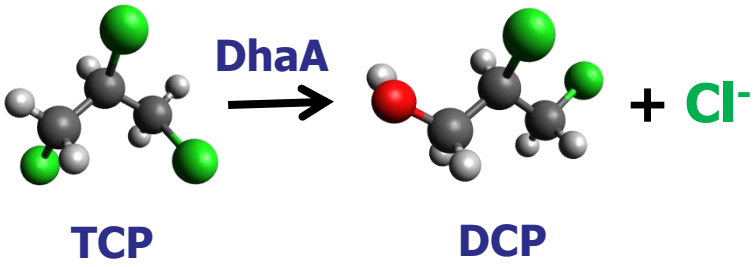
Stabilization of dehalogenase

- Comparison

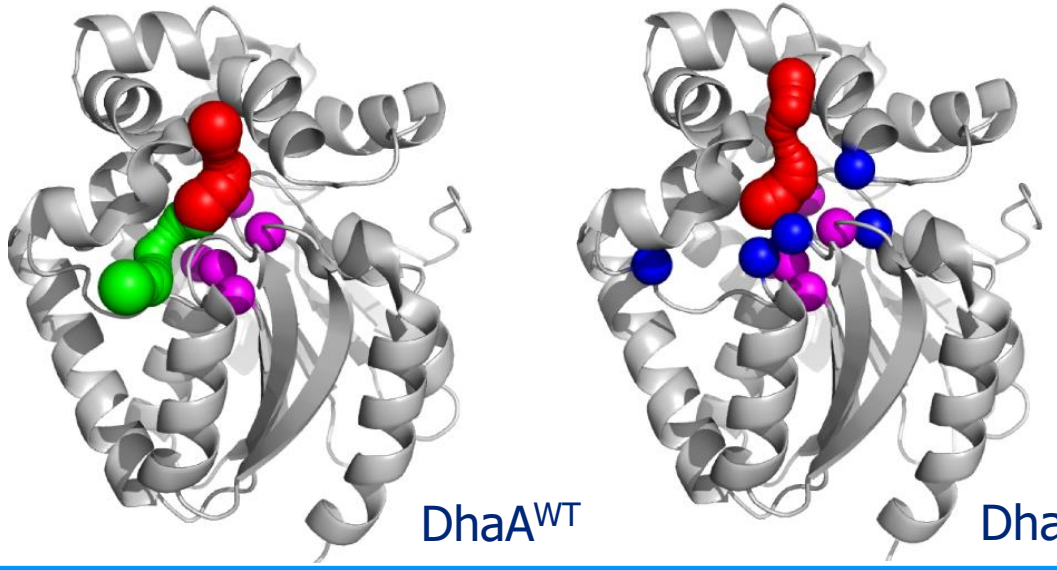
	GSSM	Rational design
Equipment (Kč)	20,000,000	500,000
Testing		
Computational	-	5,500
Experimental	120,000	5
Costs (Kč)	1,000,000	80,000
Time	Months	Weeks
Results		
# of stable mutants	11	3
Best: ΔT_m (°C)	18	25
$\tau_{1/2}$ (h)	36	72

Dehalogenase activity

- **TCP**: toxic persistent pollutant from industrial sources

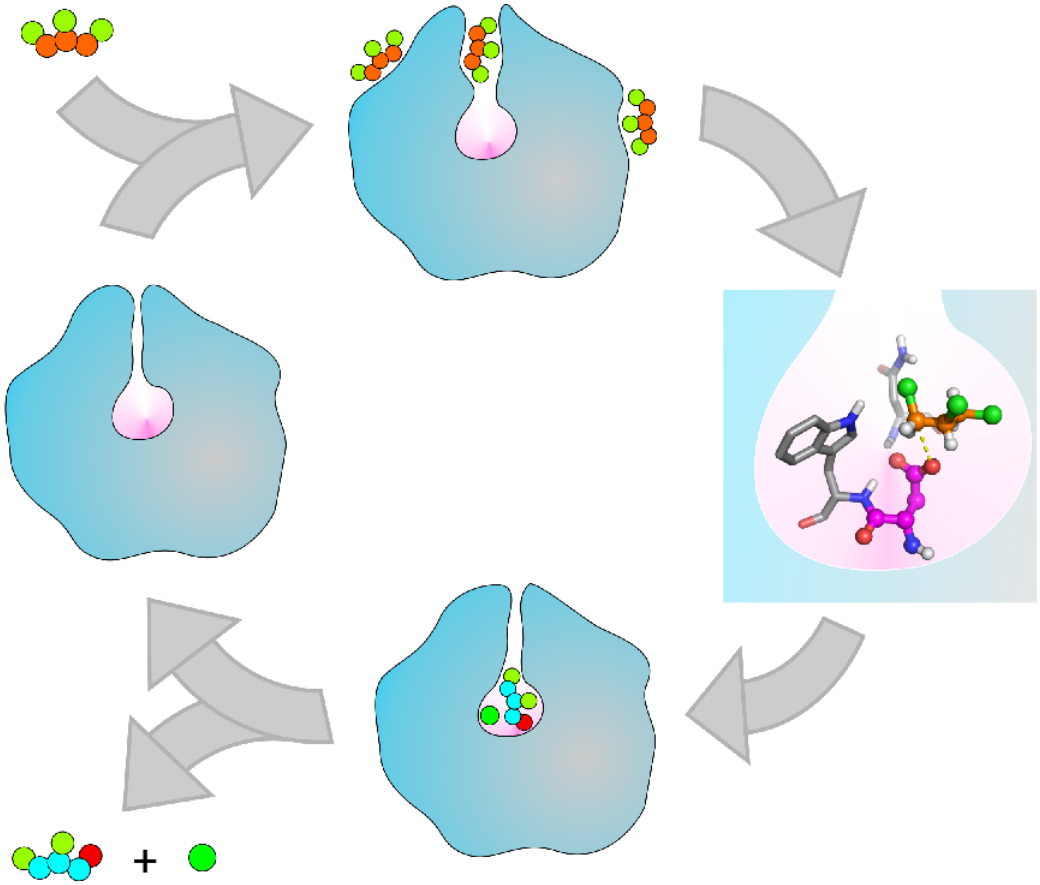


- DhaA dehalogenase (poor catalyst)
  - DhaA31: 5 mutations narrowed the access tunnels
  - 32-fold higher catalytic rate (k_{cat}); release of product became limiting step



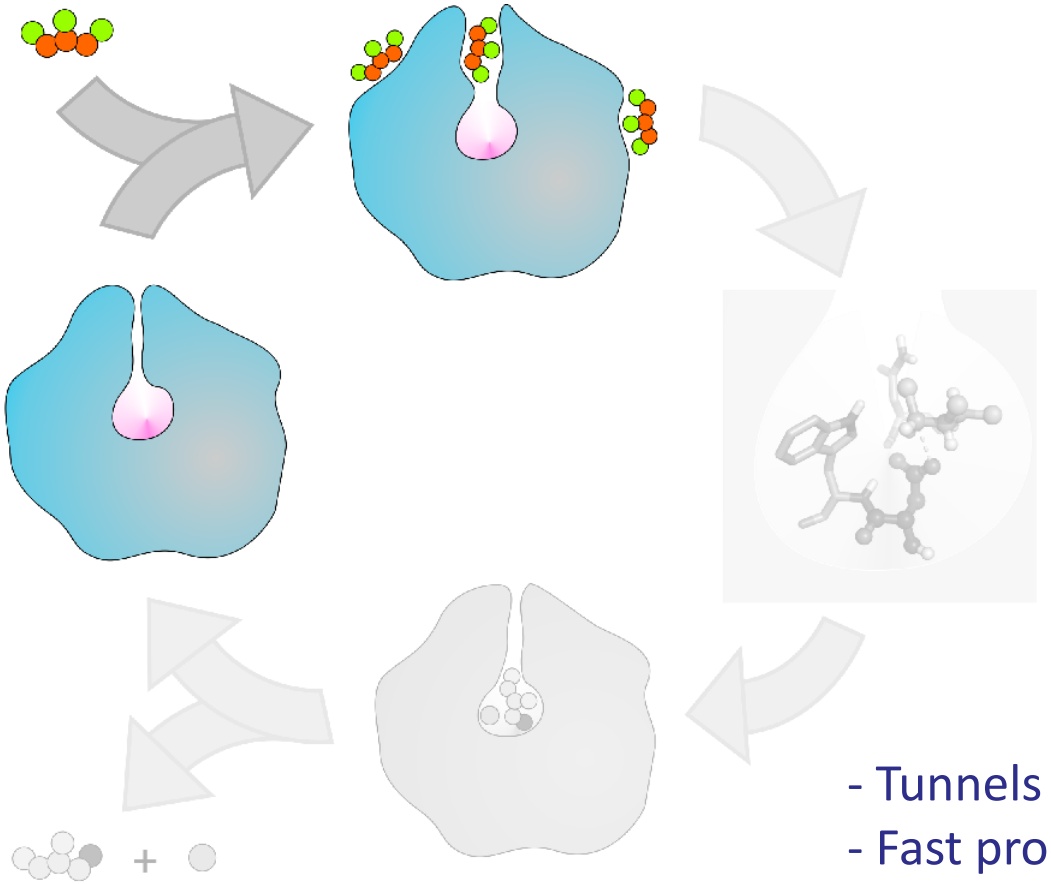
Dehalogenase activity

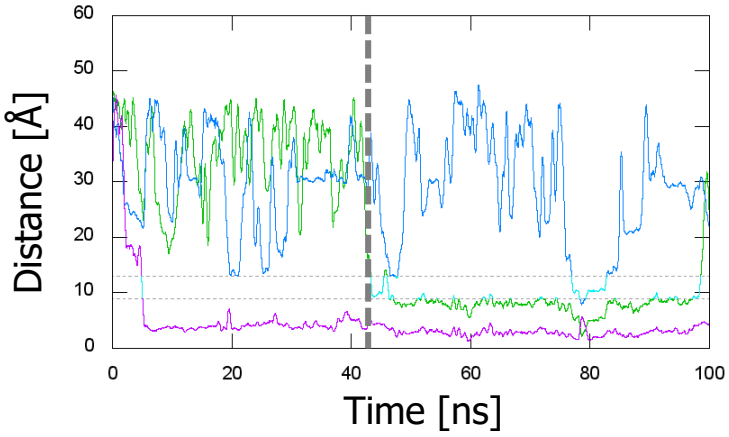
- **Catalytic cycle**: enzymes with buried active site

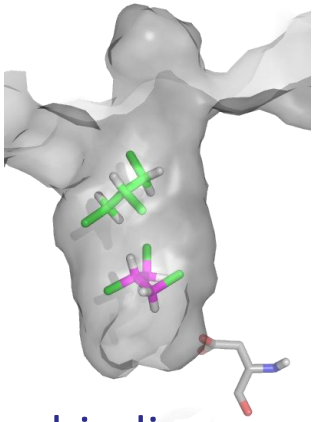


Dehalogenase activity

- Substrate binding: MD simulations



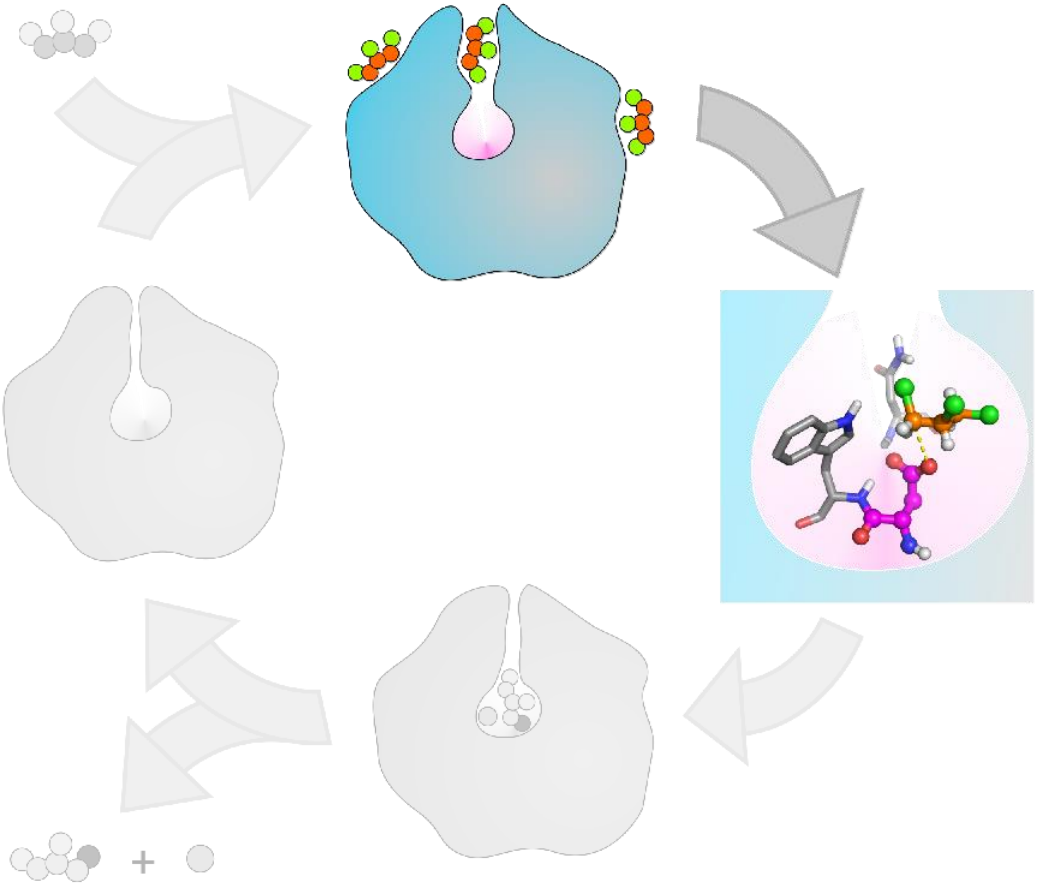




- Tunnels need to open for binding
- Fast process for both DhaA31 and DhaA^{WT}
- Potential substrate inhibition

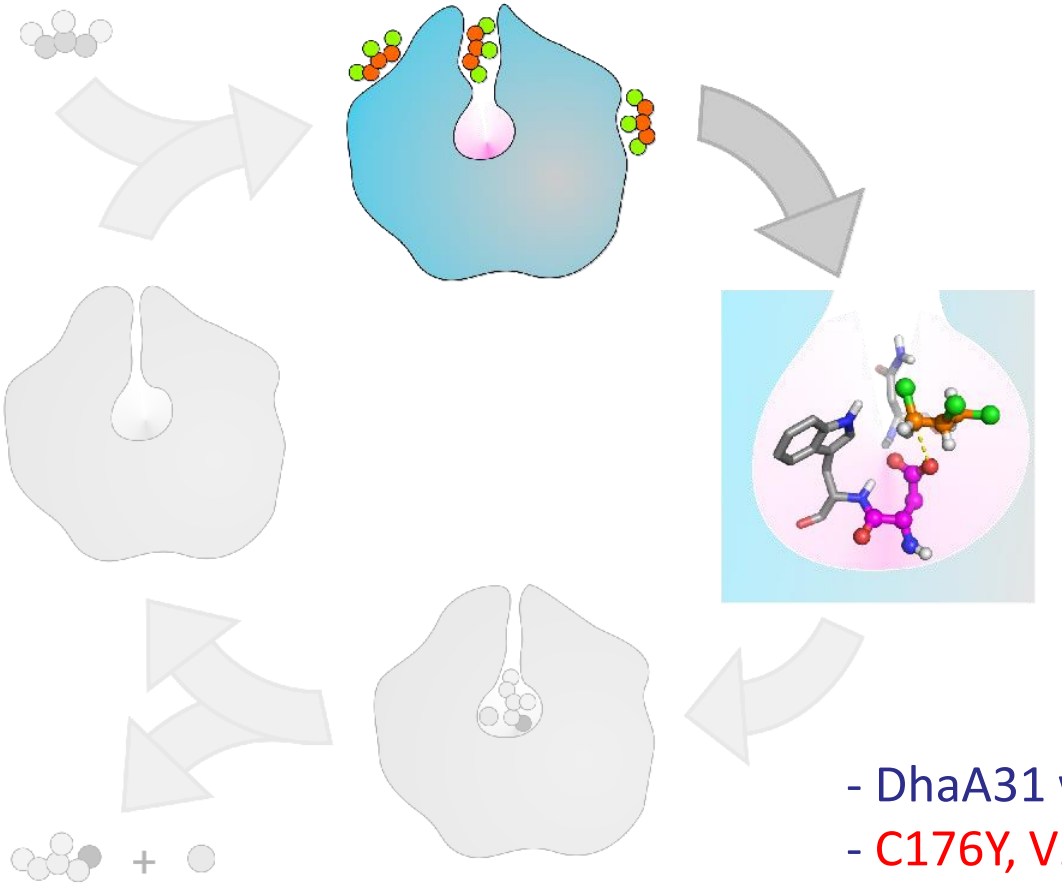
Dehalogenase activity

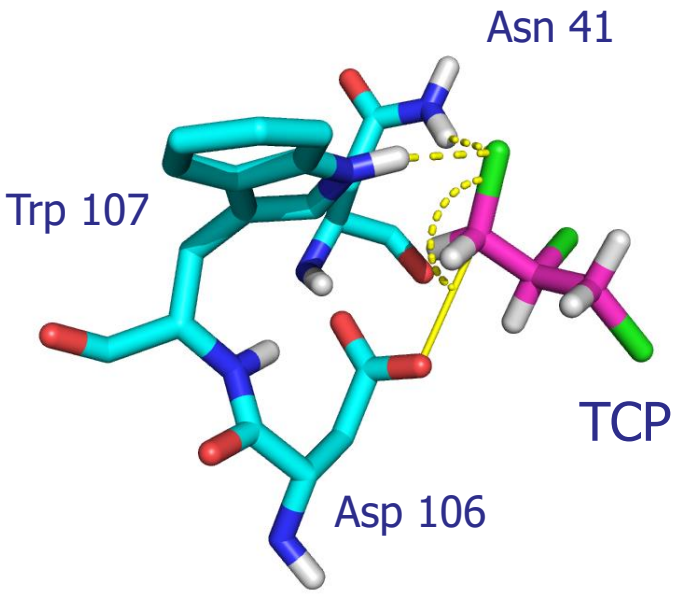
- Reactive binding: MD simulations



Dehalogenase activity

- Reactive binding: MD simulations



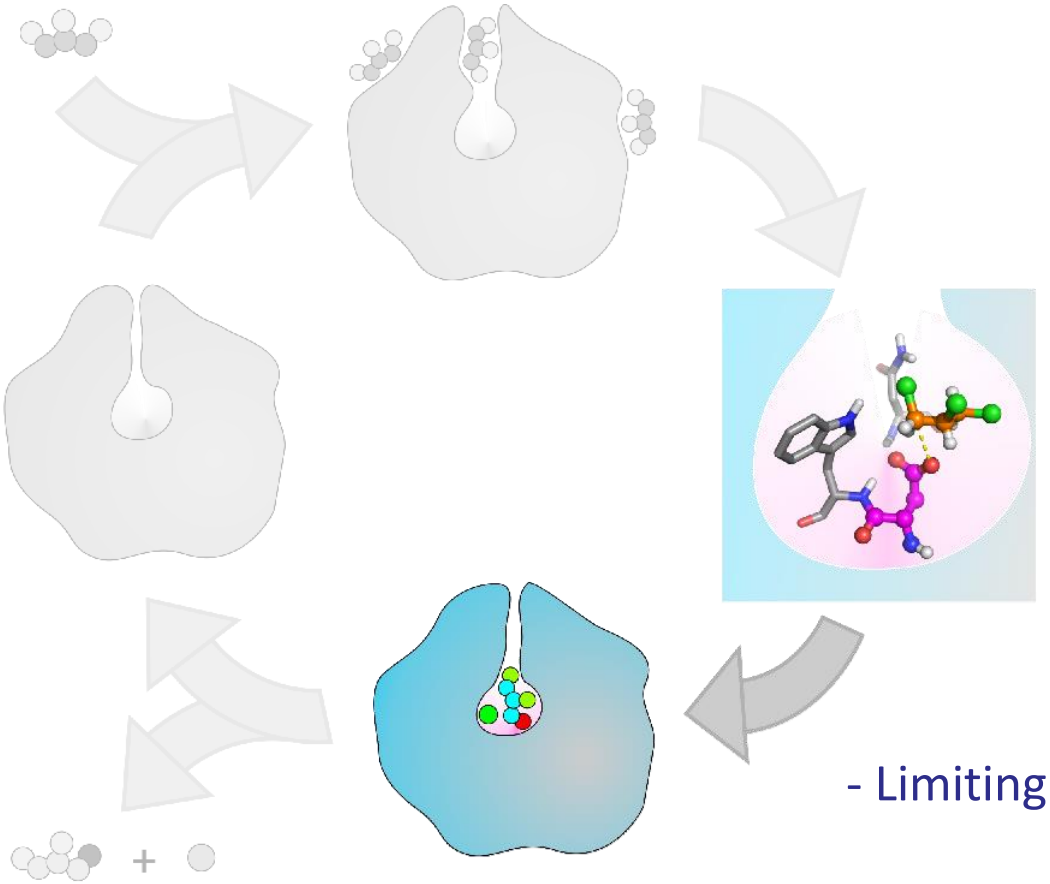


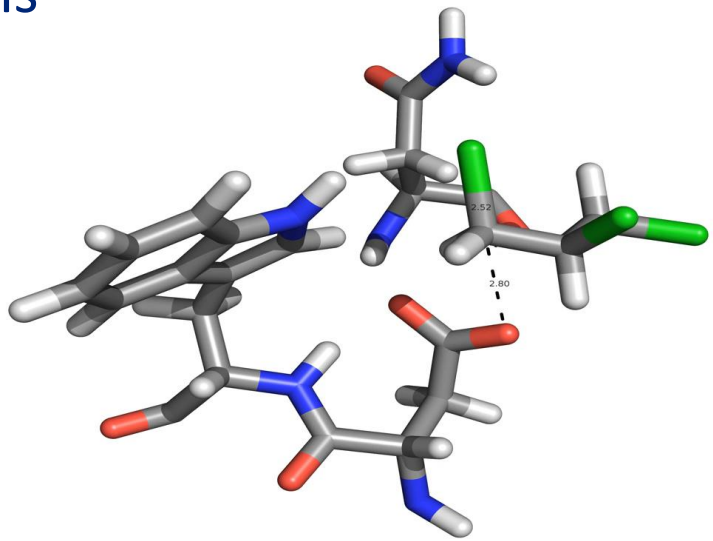
Near-attack conformation (*NAC*)

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

Dehalogenase activity

- Chemical step: QM/MM calculations



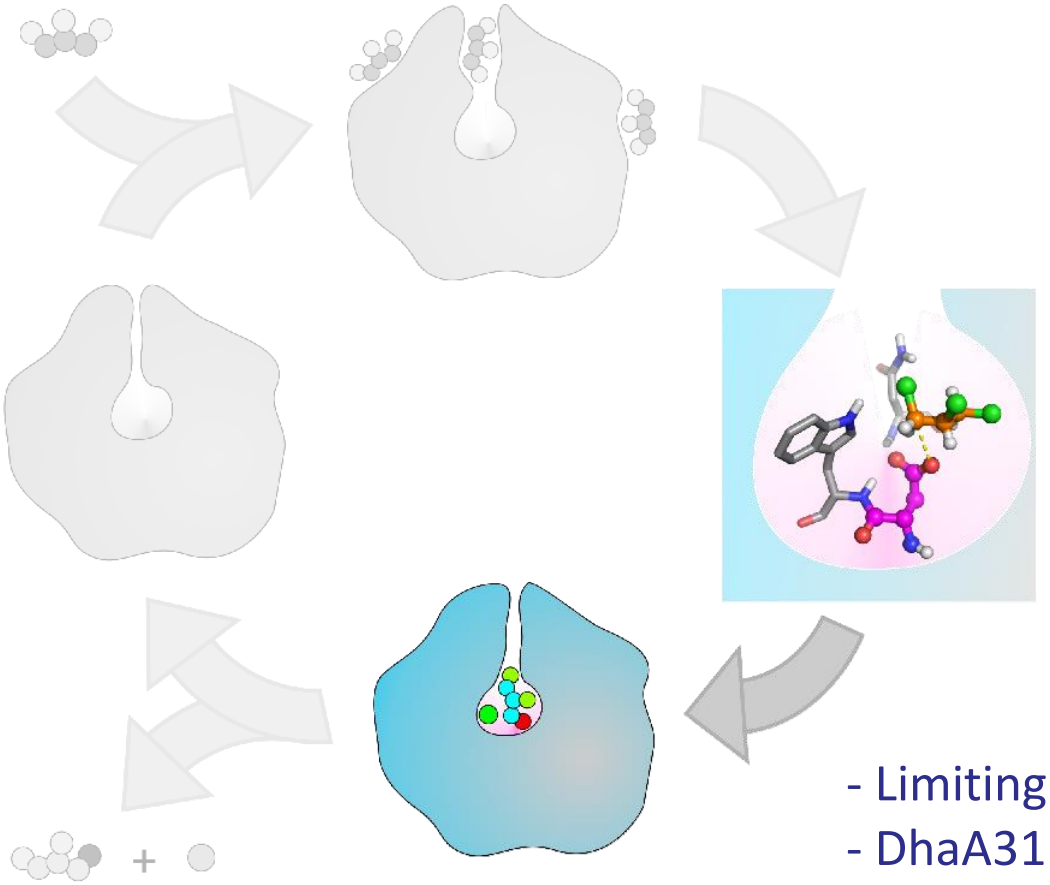


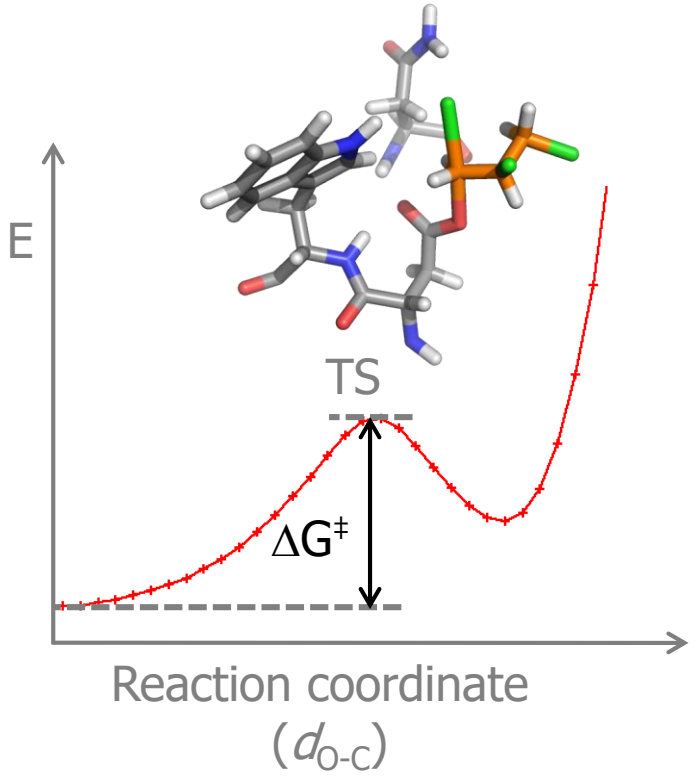
Reaction coordinate
($d_{O\text{-}C}$)

- Limiting step in DhaA^{WT}

Dehalogenase activity

- Chemical step: QM/MM calculations

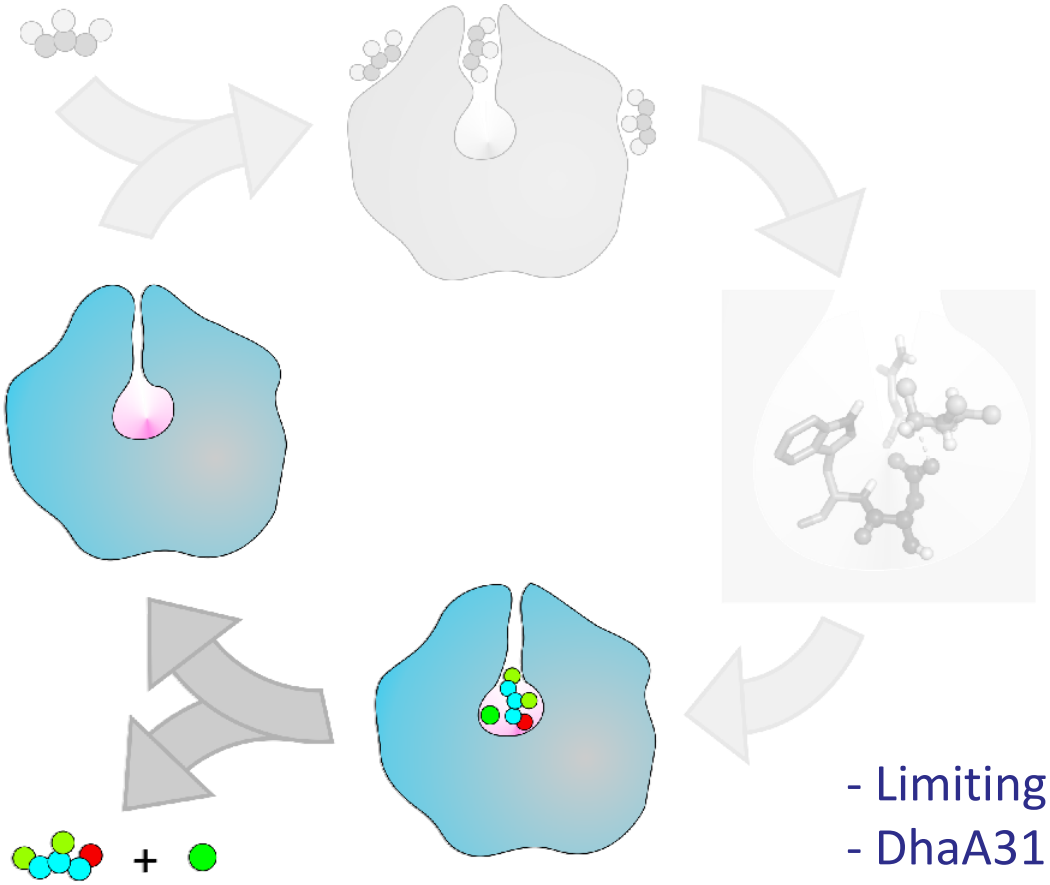


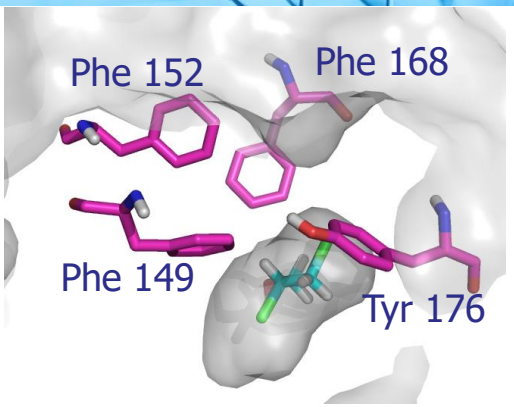


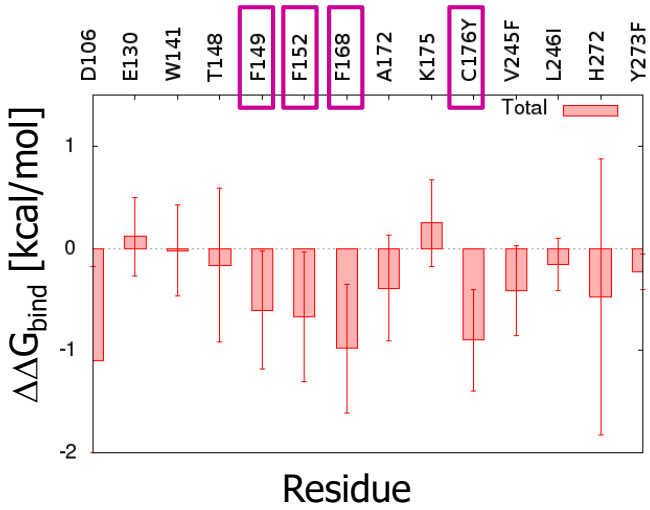
- Limiting step in DhaA^{WT}
- DhaA31 with lower $\Delta G^{‡}$ by 1.6 kcal/mol
  → ~14-fold higher reactivity

Dehalogenase activity

- Product release: MD simulations



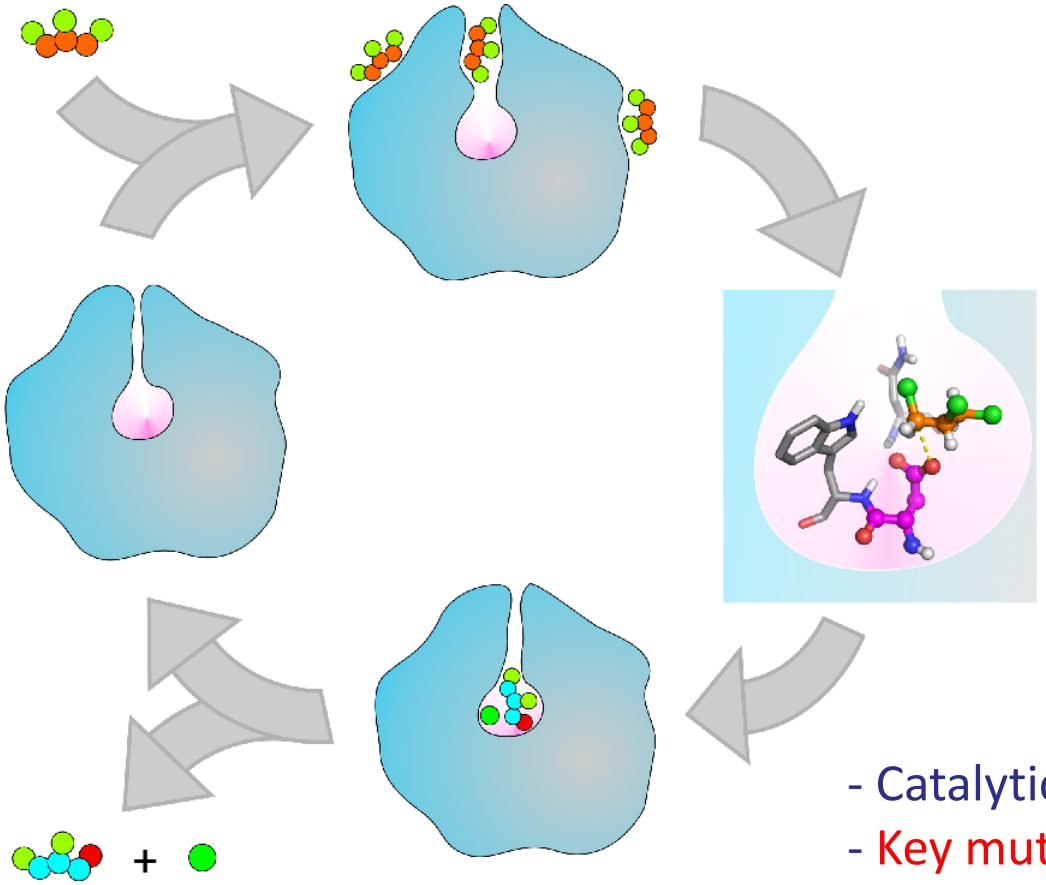


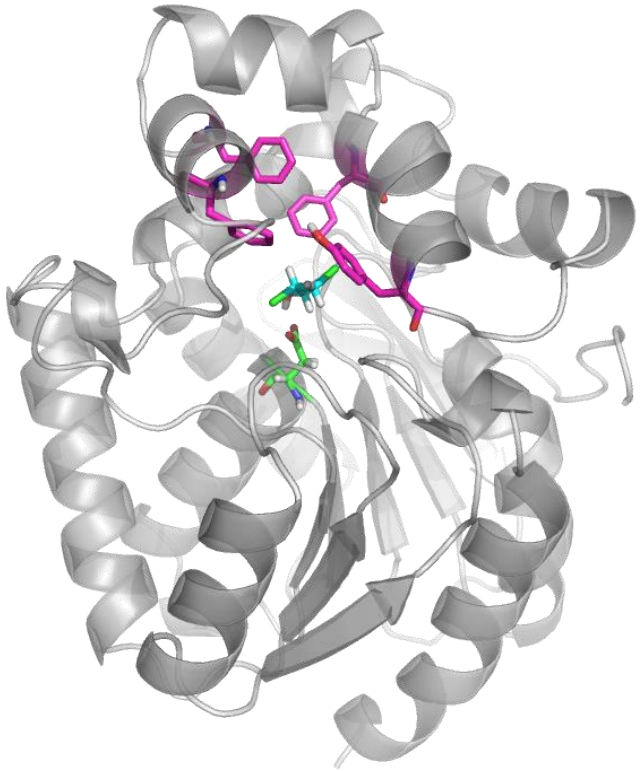


- Limiting step in DhaA31
- DhaA31 with slower release rates
- F149, F152, F168, Y176 prevent release

Dehalogenase activity

- Conclusions

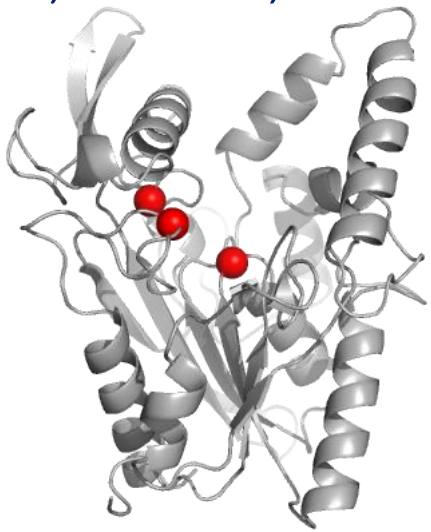


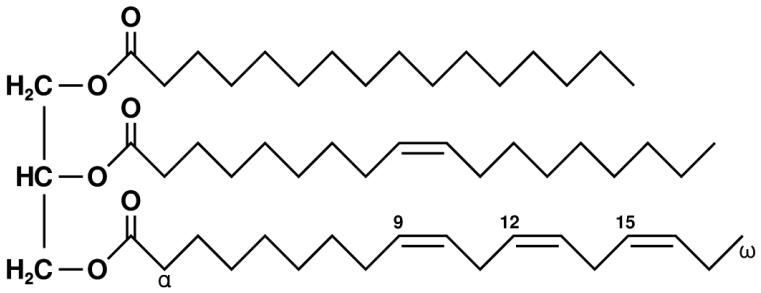


- Catalytic improvements explained
- Key mutations identified
- New hot-spots for mutagenesis

Lipase enantioselectivity

- Lipase (EC 3.1.1.3, bacterial enzyme)
  - Triacylglycerol + H_2O → diacylglycerol + carboxylic acid
  - Versatile biocatalysts: catalyze hydrolysis of carboxylic esters, esterification, transesterification, etc.
  - Many industrial applications
    - Food, detergent, pharmaceutical, leather, textile, cosmetic, paper industries

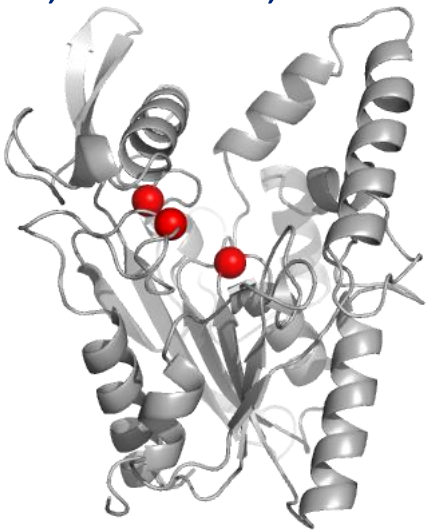




Triglycerides: main constituent of body fat

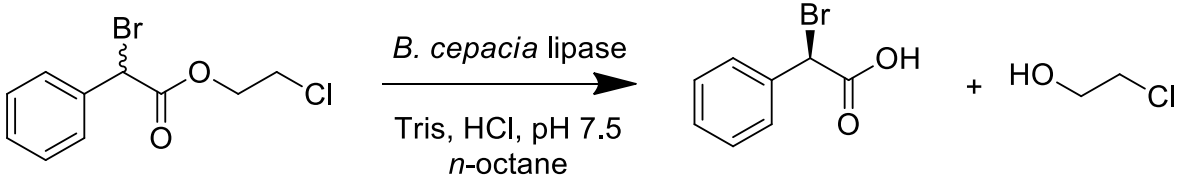
Lipase enantioselectivity

- Lipase (EC 3.1.1.3, bacterial enzyme)
  - Triacylglycerol + H_2O → diacylglycerol + carboxylic acid
  - Versatile biocatalysts: catalyze hydrolysis of carboxylic esters, esterification, transesterification, etc.
  - Many industrial applications
    - Food, detergent, pharmaceutical, leather, textile, cosmetic, paper industries
    - Used to resolve racemic mixtures

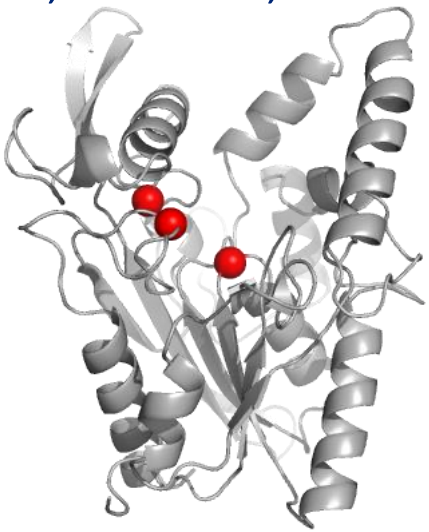


Lipase enantioselectivity

- Lipase (EC 3.1.1.3, bacterial enzyme)
  - Triacylglycerol + H_2O → diacylglycerol + carboxylic acid
  - Versatile biocatalysts: catalyze hydrolysis of carboxylic esters, esterification, transesterification, etc.
  - Many industrial applications
    - Food, detergent, pharmaceutical, leather, textile, cosmetic, paper industries
    - Used to resolve racemic mixtures



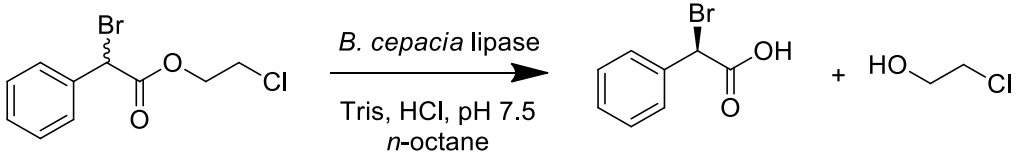
(*R*-enantiomer)

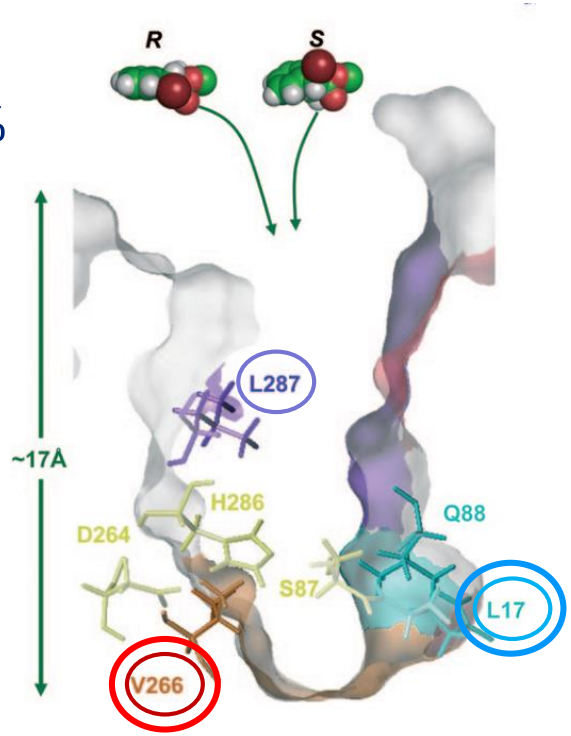


Lipase enantioselectivity

- 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
  - Mutants with 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)

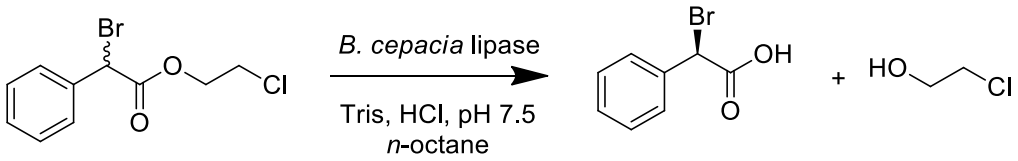


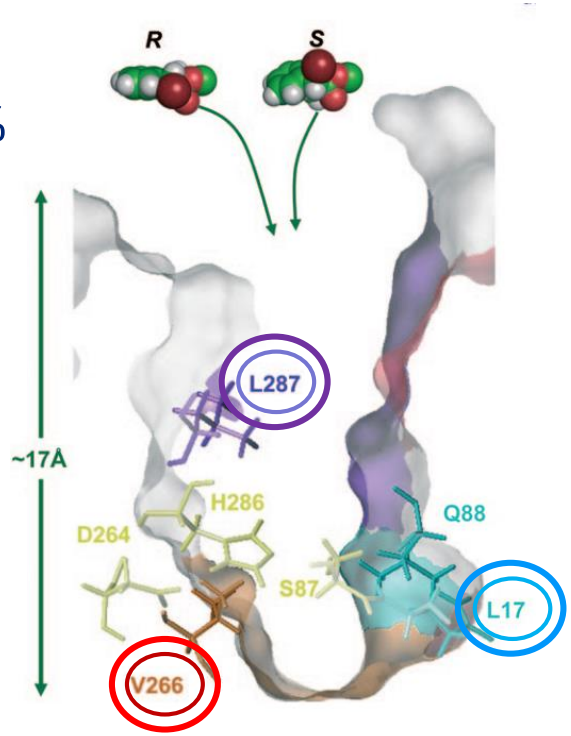


Lipase enantioselectivity

- 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
  - Mutants with 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)

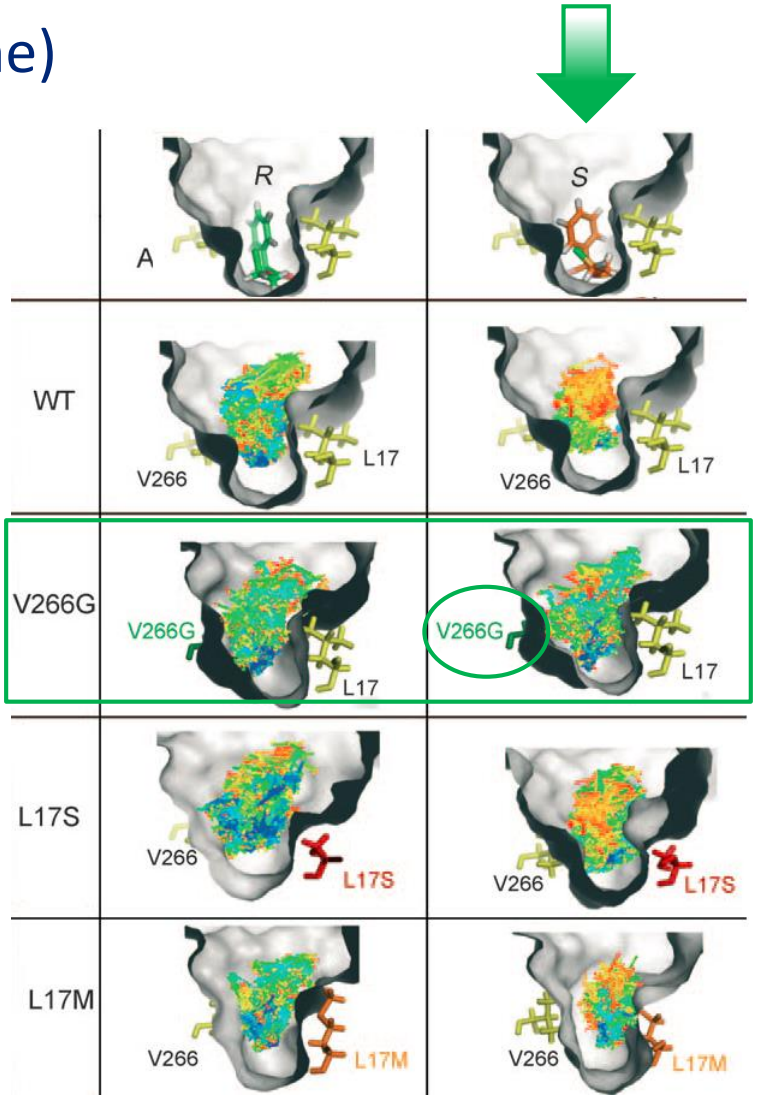




Lipase enantioselectivity

- Lipase (EC 3.1.1.3, bacterial enzyme)
  - Molecular dynamics with substrates

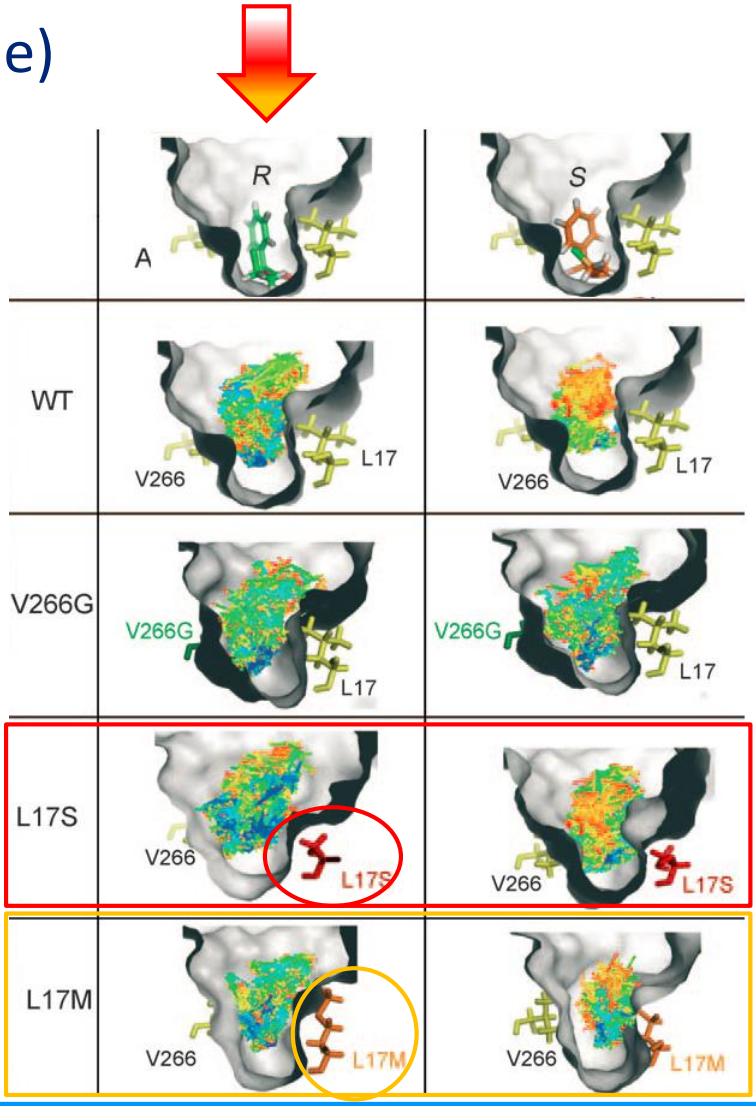
Variants	Enantiopreference	*E* value	Conversion [%]
Wild-type	*R*	13 (±1.8)[b]	6.5 (48 h)
V266G	*S*	20 (±4)[c]	6.6 (51 h)
L17S	*R*	128 (±35)[b]	15.6 (49 h)
L17M	*R*	133 (±31)[b]	15.5 (48 h)



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)

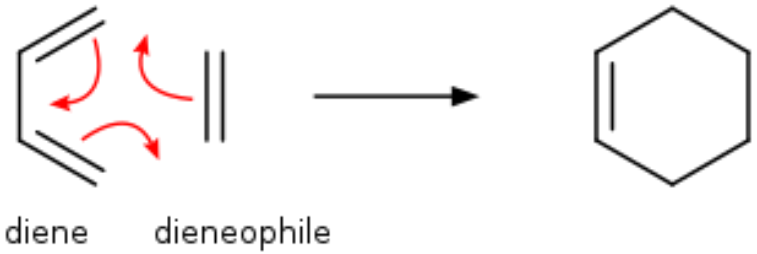


- Steric changes on either side of the active site favor reactive binding of one enantiomer
- Combined mutations favor one enantiomer and disfavor the other

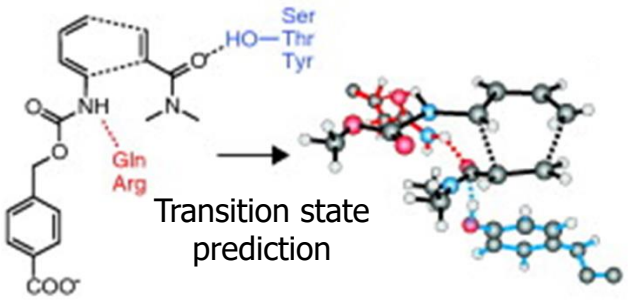
De novo design of a Diels-Alderase

- Non-existing Diels-Alderase
  - Goal: design biocatalyst for intermolecular Diels-Alder reaction

    → very specific geometric and electronic requirements → theozymes



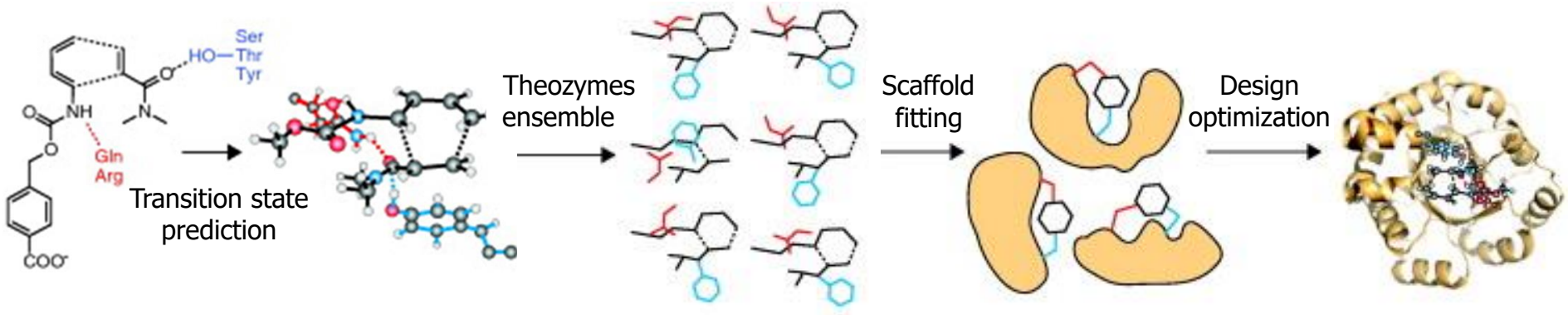
Diels–Alder cycloaddition



De novo design of a Diels-Alderase

- Non-existing Diels-Alderase
  - Goal: design 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
  - Results: creation of functional & stereoselective Diels-Alderase



Summary

- Structural biology methods are important tools to:
  - Explain biological phenomena
  - Increase efficiency of drug discovery
  - Successfully engineer proteins for biotechnological applications
- Often produce better results than experimental brute-force
- Can reduce costs and save time

Summary

- Structural biology methods are important tools to:
  - Explain biological phenomena
  - Increase efficiency of drug discovery
  - Successfully engineer proteins for biotechnological applications
- Often produce better results than experimental brute-force
- Can reduce costs and save time

- **This lesson will not be on the exam!**

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: evaluation

- Teachers' evaluation

  ➔ Evaluation Survey – PLEASE respond!

Final remarks: evaluation

- Exam 1 h, 3 dates
  - 17 Dec. 2024, 10:00 (location: B11/333)
  - 7 Jan 2025, 10:00 (location: B11/333)
  - 28 Jan. 2025, 10:00 (location: B11/333)
- Multiple-choice exam
  - 25 questions
  - 10 points out of 25 needed to pass
  - Multiple correct answers possible
- Only topics with the sign  on the slides will be asked
- Teachers are available for questions. Contact me!

Final remarks: evaluation

- Questions – example 1

Choose the true statements about van der Waals interactions.

1. These are long-range interactions
2. Interaction occurs between any types of atoms
3. These interactions play a role only with charged amino acid residues
4. These are short-range interactions
5. These interactions are entropic in nature

Final remarks: evaluation

- Questions – example 1

Points: Choose the true statements about van der Waals interactions.

-1/3 1. These are long-range interactions
+1/2 2. Interaction occurs between any types of atoms
-1/3 3. These interactions play a role only with charged amino acid residues
+1/2 4. These are short-range interactions
-1/3 5. These interactions are entropic in nature

Final remarks: evaluation

- Questions – example 2

Choose the true statements about homology modeling.

A) It is based on the principle that sequences are much more conserved than 3D structures during evolution
B) The structural model of the target protein is predicted based on the known experimental 3D structure of a related protein
C) A necessary condition for homology modeling is the existence of a suitable template
D) Homology modeling generally provides less accurate results than ab initio predictions

Final remarks: evaluation

- Questions – example 2

Points: Choose the true statements about homology modeling.

-1/2 A) It is based on the principle that sequences are much more conserved than 3D structures during evolution

+1/2 B) The structural model of the target protein is predicted based on the known experimental 3D structure of a related protein

+1/2 C) A necessary condition for homology modeling is the existence of a suitable template

-1/2 D) Homology modeling generally provides less accurate results than ab initio predictions

Final remarks: evaluation

- Bring only one pen and ID card