Protein Engineering

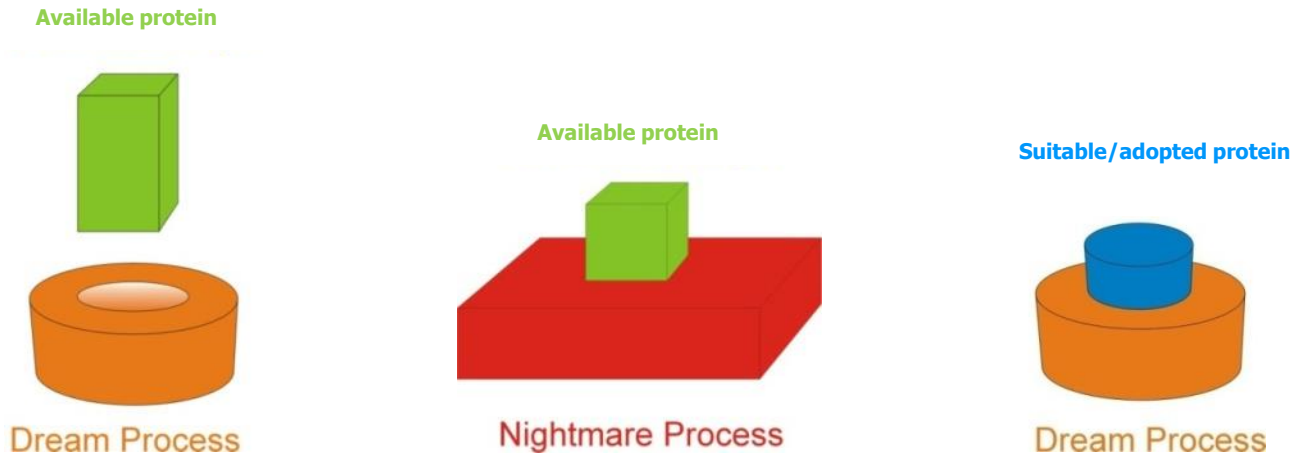
Outline

- Limitations of proteins in biotechnology processes
- Definition and aim of protein engineering
- Targeted properties of proteins
- Basic approaches in protein engineering
  - **DIRECTED EVOLUTION**
  - **RATIONAL DESIGN**
  - **SEMI-RATIONAL DESIGN**
- Examples, application of artificial inteligence

Proteins in biotechnology

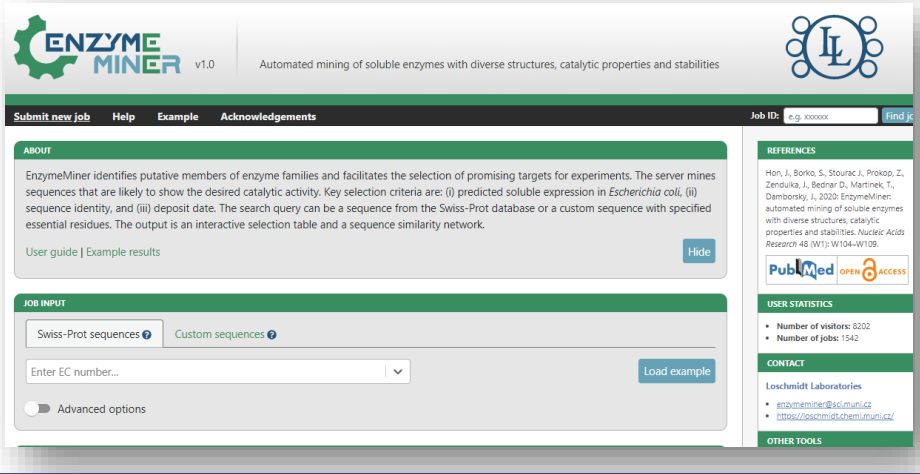
- **key problem** -availability of optimal protein for specific process
- **traditional biotechnology** - adapt process
- **modern biotechnology** - adapt protein

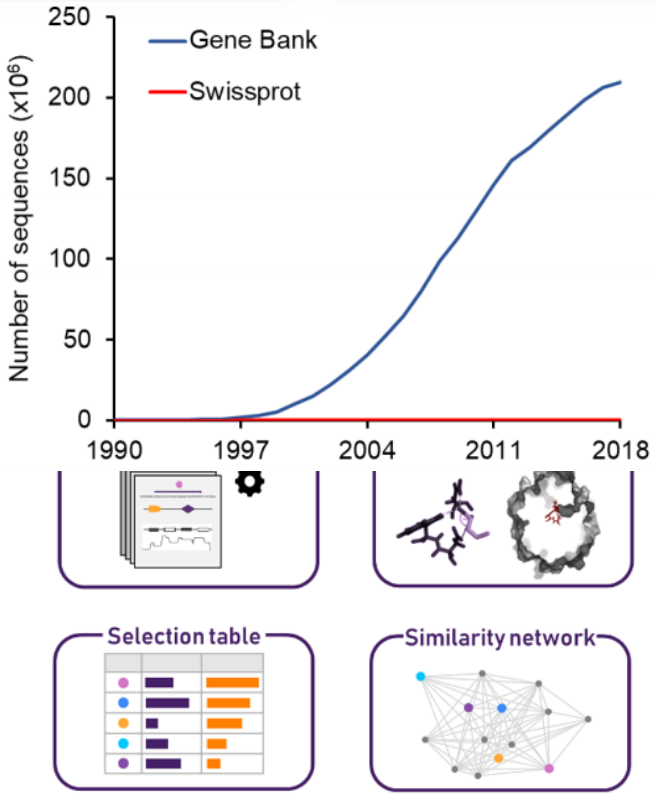
HOW TO OBTAIN OPTIMAL PROTEIN?



Proteins in biotechnology

- **classical screening**
  - screening culture collections
  - polluted and extreme environment
- **environmental gene libraries**
  - metagenomic DNA
- **data-base mining**
  - gene databases
  - (meta)genome sequencing projects
  - numerous uncharacterised proteins





Proteins in biotechnology

- **classical screening**
  - screening culture collections
  - polluted and extreme environment
- **environmental gene libraries**
  - metagenomic DNA
- **data-base mining**
  - gene databases
  - (meta)genome sequencing projects
  - numerous uncharacterised proteins

IF SUITABLE PROTEIN DOES NOT EXIST IN NATURE?

- **PROTEIN ENGINEERING**

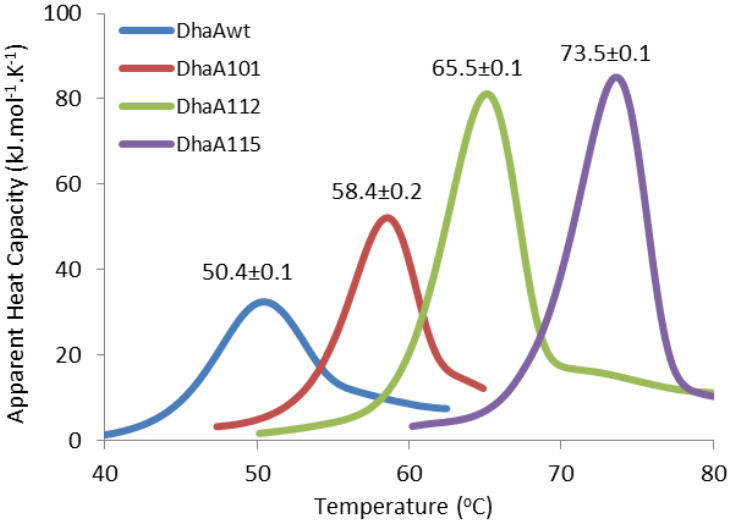
Aims of protein engineering

- the process of **constructing novel protein** molecules by design first principles or altering existing structure
- use of genetic manipulations to alter the coding sequence of a gene and thus **modify the properties of the protein**

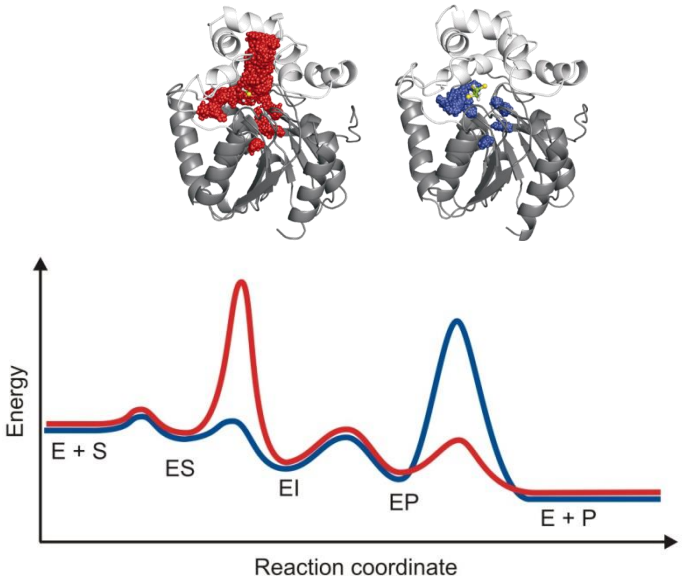
- AIMS AND APPLICATIONS
  - **technological** - optimisation of the protein to be suitable in particular technology purpose
  - **scientific** - desire to understand what elements of proteins contribute to folding, stability and function

Targeted properties of proteins

- **structural** properties of proteins
  - **stability** (temperature, solvents)
  - tolerance to pH, salt
  - resistance to oxidative stress



- **functional** properties of proteins
  - **substrate specificity** and selectivity
  - **kinetic properties** (e.g., K_m, k_{cat}, K_i)
  - cofactor selectivity
  - protein-protein or protein-DNA interactions



Strategies in protein engineering

RATIONAL DESIGN

1. Computer aided design

2. Site-directed mutagenesis

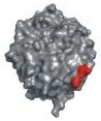
Individual mutated gene

3. Transformation

4. Protein expression

5. Protein purification

6. *not applied*

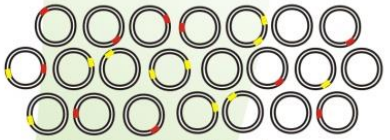


Constructed mutant enzyme

DIRECTED EVOLUTION

1. *not applied*

2. Random mutagenesis



Library of mutated genes
(>10,000 clones)

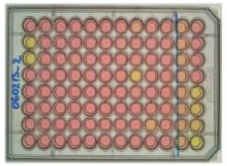
3. Transformation

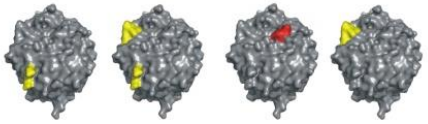
4. Protein expression

5. *not applied*

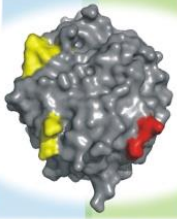
6. Screening and selection

- stability
- selectivity
- affinity
- activity





Selected mutant enzymes



Improved protein

7. Biochemical testing

Directed evolution

- directed evolution techniques emerged during mid-1990s
- **inspired by natural evolution**
- this form of "evolution" does not match what Darwin had envisioned
  - requires **outside intelligence**, not blind chance
  - does not take millions of years, but **happens rapidly**

Frances H. Arnold

Frances H. Arnold
The Nobel Prize in Chemistry 2018

Born: 25 July 1956, Pittsburgh, PA, USA

Affiliation at the time of the award: California Institute of Technology (Caltech), Pasadena, CA, USA

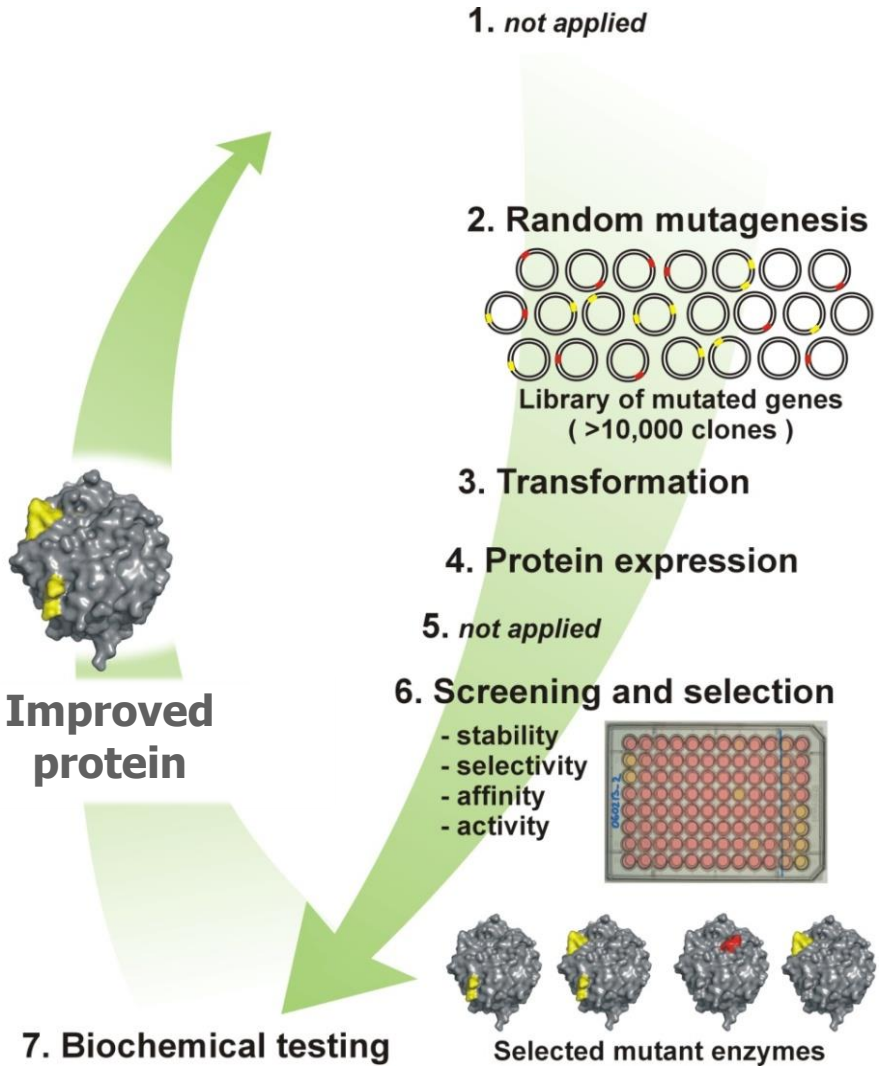
Prize motivation: "for the directed evolution of enzymes."

Prize share: 1/2

Directed evolution

- **evolution in test tube** comprises two steps
  - **random mutagenesis**
    building mutant library (diversity)
  - **screening and selection**
    identification of desired biocatalyst

- **prerequisites** for directed evolution
  - gene encoding protein of interest
  - method to create mutant library
  - suitable expression system
  - screening or selection system

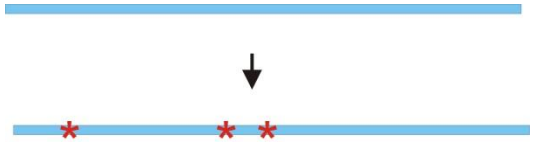


Methods to create mutant libraries

- technology to **generate large diversity**

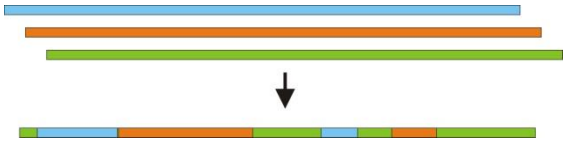
  - **NON-RECOMBINING**

    one parent gene -> variants with point mutations



  - **RECOMBINING**

    several parental homologous genes -> chimeras



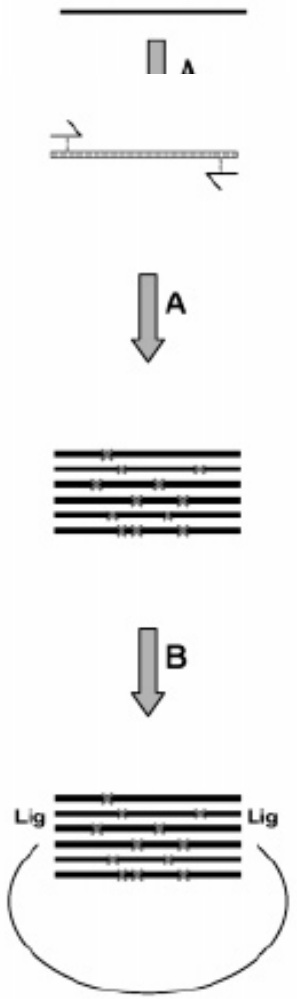
Non-recombining mutagenesis

- **UV irradiation** or **chemical mutagens** (traditional)
- **mutator strains** - lacks DNA repair mechanism

  mutations during replication (e.g., *Epicurian coli* XL1-Red)
- **error-prone polymerase chain reaction** (ep-PCR)
  - gene amplified in imperfect copying process

    (e.g., unbalanced deoxyribonucleotides concentrations,

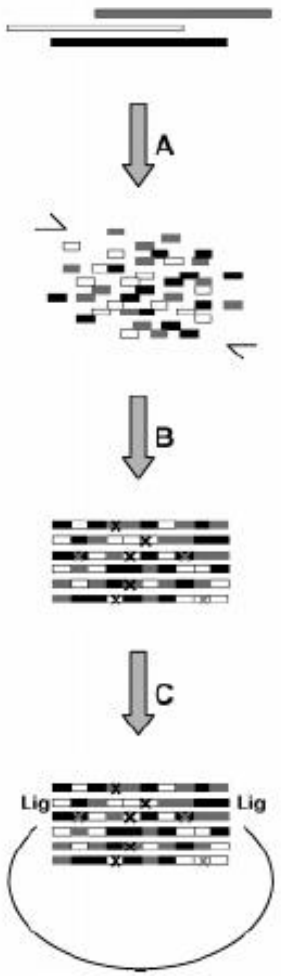
    high Mg^{2+} concentration, Mn^{2+}, low annealing temperatures)
  - 1 to 20 mutation per 1000 base pairs
- **saturation mutagenesis**
  - randomization of single or multiple codons
  - gene site saturation mutagenesis
- **other methods**
  - insertion/deletions (InDel)
  - cassette mutagenesis (region mutagenesis)



Recombining mutagenesis

- also refered to as **„sexual mutagenesis“**
- **DNA shuffling**
  - fragmentation step
  - random reassembly of segments
- **StEP** - staggered extension process
  - simpler then shuffling
  - random reannealing combined with limited primer extension
- **other methods**

  shuffling of genes with lower homology down to 70% (e.g., RACHITT, ITCHY, SCRATCHY)

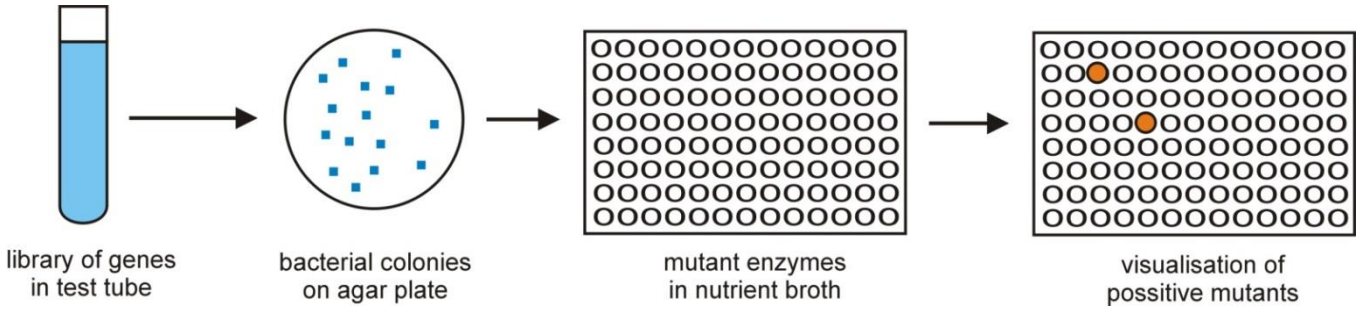


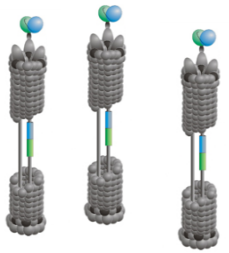
Screening and selection

- most **critical step** of direct evolution
- isolation of positive mutants hiding in library
  - **HIGH THROUGHPUT SCREENING**

    individual assays of variants one by one
  - **DIRECT SELECTION**

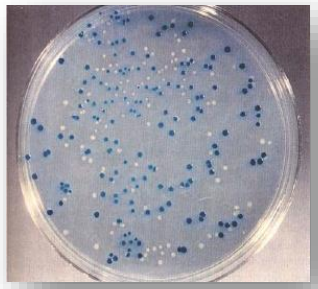
    display techniques (link between genotype and phenotype)

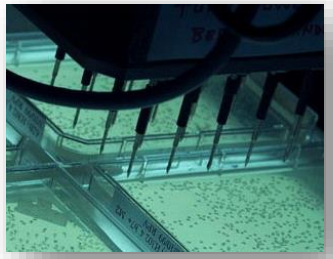
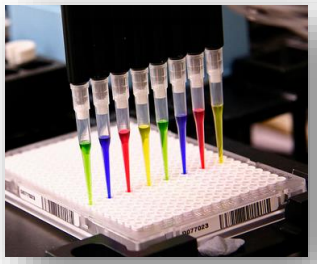
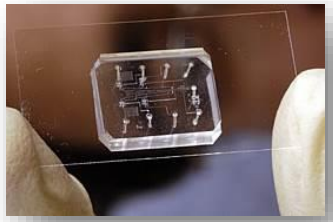
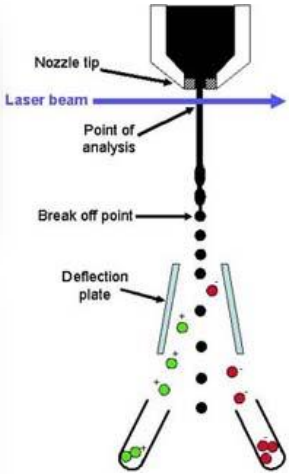


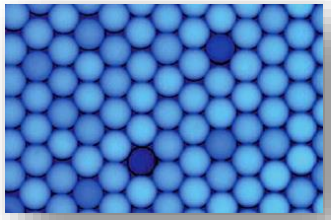


(Utra)High throughput screening

- common methods not applicable
- **agar plate (pre)screening**
- **microtiter plates screening**
  - 96-, 384- or 1536-well formate
  - robot assistance
    (colony picker, liquid handler)
  - 10^4 libraries
  - volume 10 – 100 uL
- **microfluidic systems *(Lesson 6)***
  - water in oil emulsions (up to 10 kHz)
  - FACS sorting (10^8 events/hour)
  - 10^9 libraries
  - volume 1 – 10 pL

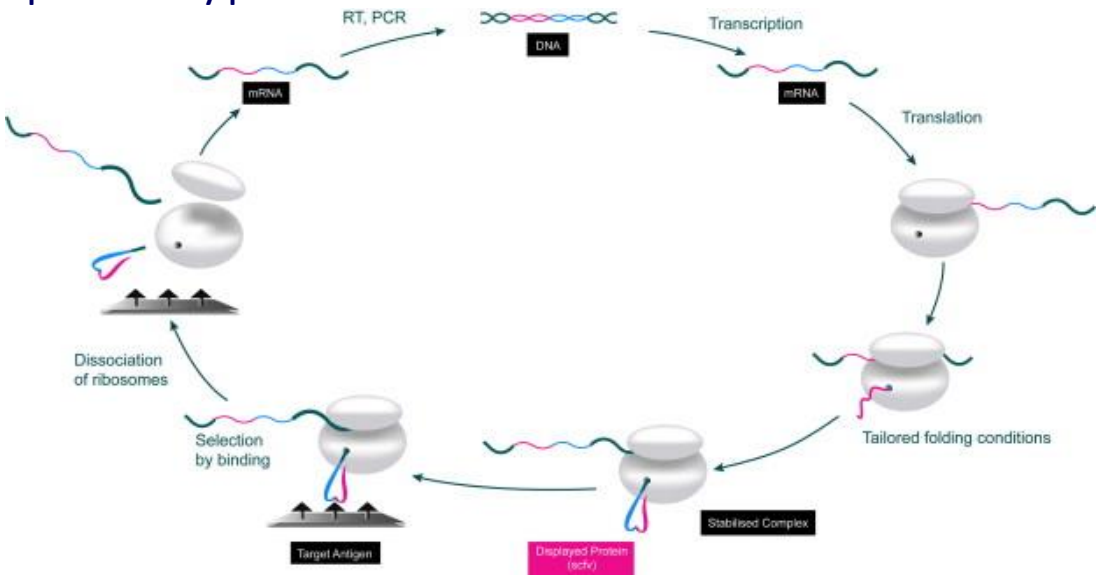


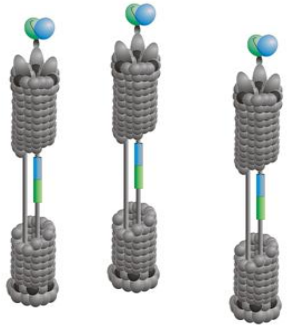







Direct selection

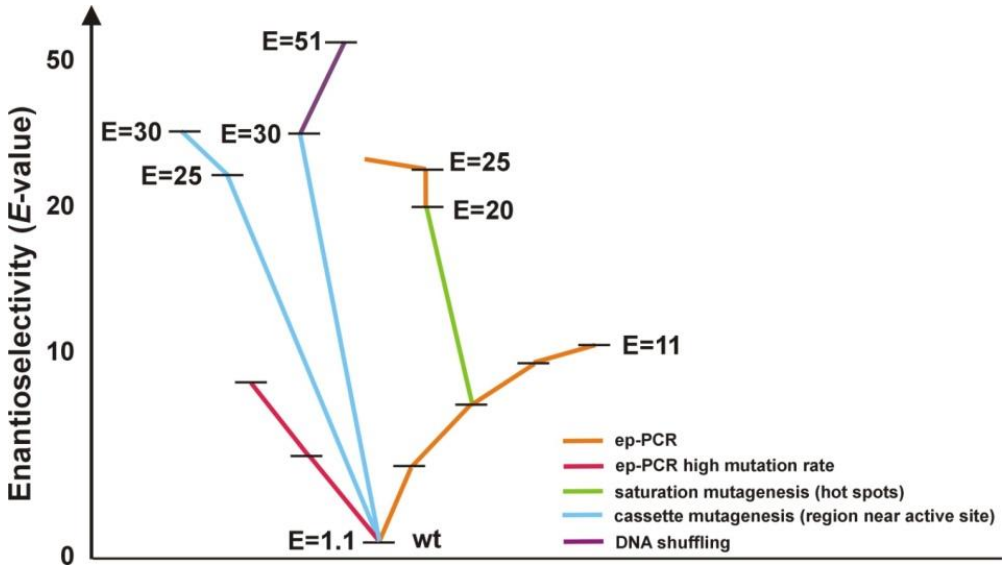
- not generally applicable (mutant libraries >10^6 variants)
- link between genotype and phenotype
- **display technologies**
  - ribosome display
  - phage display
- **life-or-death assay**
  - auxotrophic strain
  - toxicity based selection

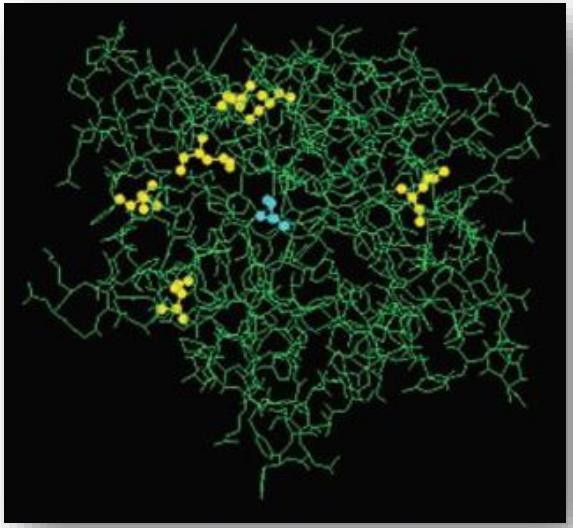




Example of Directed evolution

- directed evolution of **enantioselectivity**
  - lipase from *P. aeruginosa* (E-value improved from 1.1 into 51)
  - **spectrophotometric screening** of (*R*)- and (*S*)-nitrophenyl esters
  - **40 000 variants** screened
  - the best mutant contains six amino acid substitutions





Reetz, M., et al. 2001. Angew. Chem. Int. Ed. 40: 3589-91

Strategies in protein engineering

RATIONAL DESIGN

1. Computer aided design

2. Site-directed mutagenesis

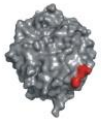
Individual mutated gene

3. Transformation

4. Protein expression

5. Protein purification

6. *not applied*

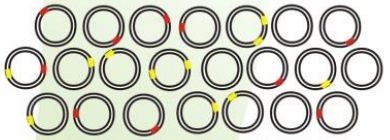


Constructed mutant enzyme

DIRECTED EVOLUTION

1. *not applied*

2. Random mutagenesis



Library of mutated genes
(>10,000 clones)

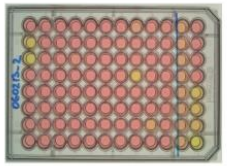
3. Transformation

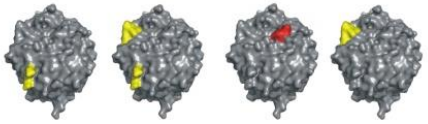
4. Protein expression

5. *not applied*

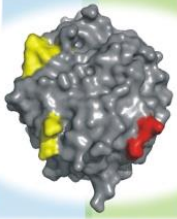
6. Screening and selection

- stability
- selectivity
- affinity
- activity





Selected mutant enzymes



Improved protein

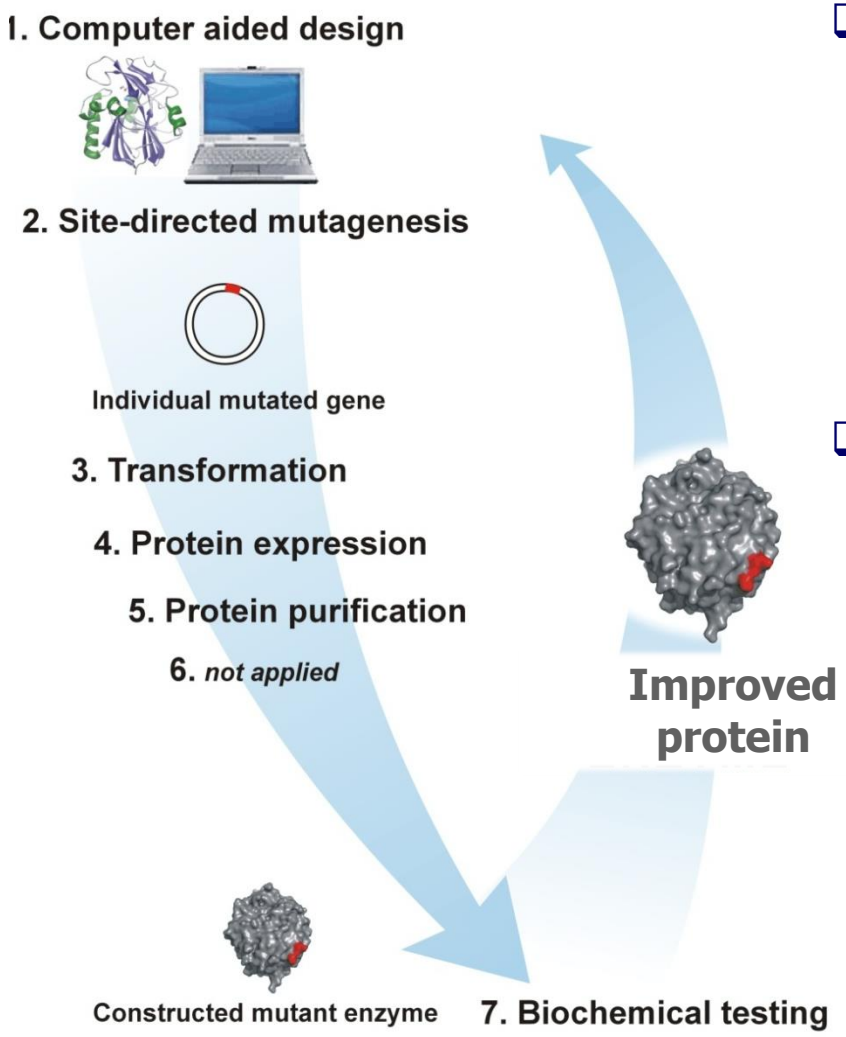
7. Biochemical testing

Rational design

- emerged around 1980s as the original protein engineering approach
- **knowledge based** - combining theory and experiment
- protein engineering cycle:

  „structure-theory-design-mutation-purification-analysis“
- **difficulty in prediction** of mutation effects on protein property
- **de novo design** most challenging

Principal of rational design



- rational design comprises:
  - **design** - understanding of protein functionality
  - **experiment** - construction and testing of mutants
- **prerequisites** for rational design:
  - **gene** encoding protein of interest
  - **3D structure** (e.g., X-ray, NMR) or sequence alignment
  - structure-function relationship
  - computational methods and capacity
  - side directed mutagenesis techniques
  - efficient expression system
  - biochemical tests

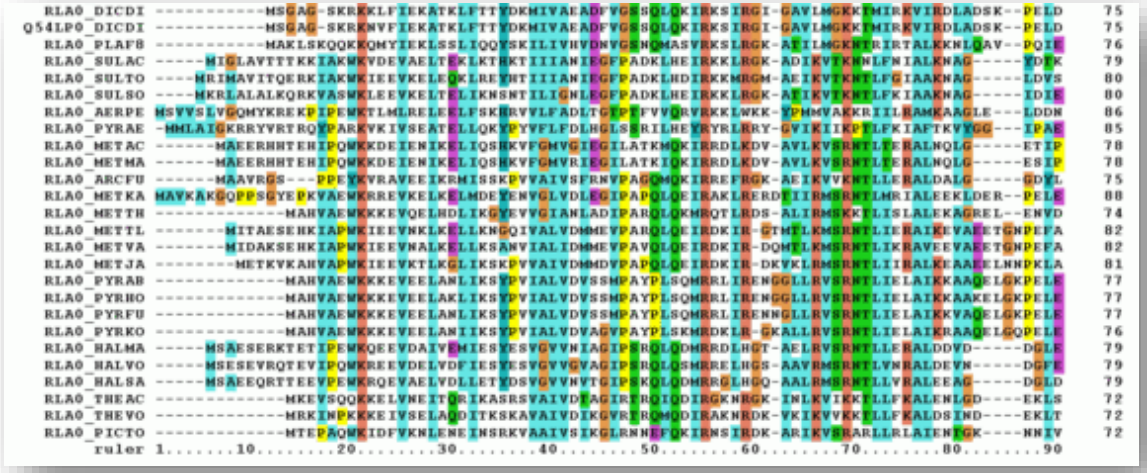
Design

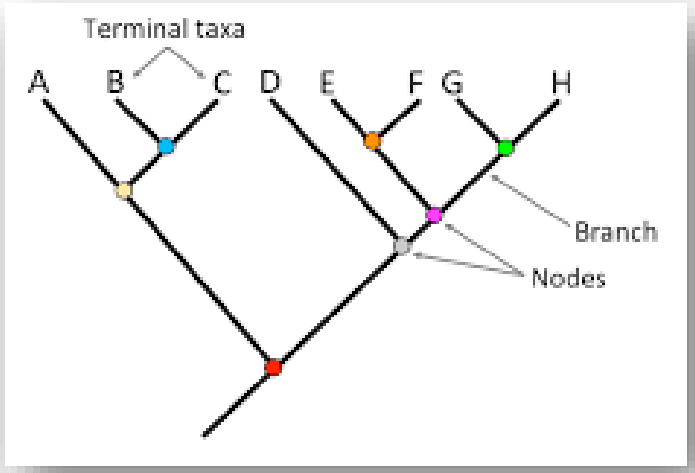
❑ SEQUENCE HOMOLOGY APPROACH

- homologous wild-type **sequences alignment**
- **identifying** amino acid residues responsible for differences
- **design** - combination of possitive mutation from all parental proteins

❑ ANCESTRAL RECONSTRUCTION

- construction of **phylogenetic tree**
- **design -** nods prediction by **consensus approach**





Design

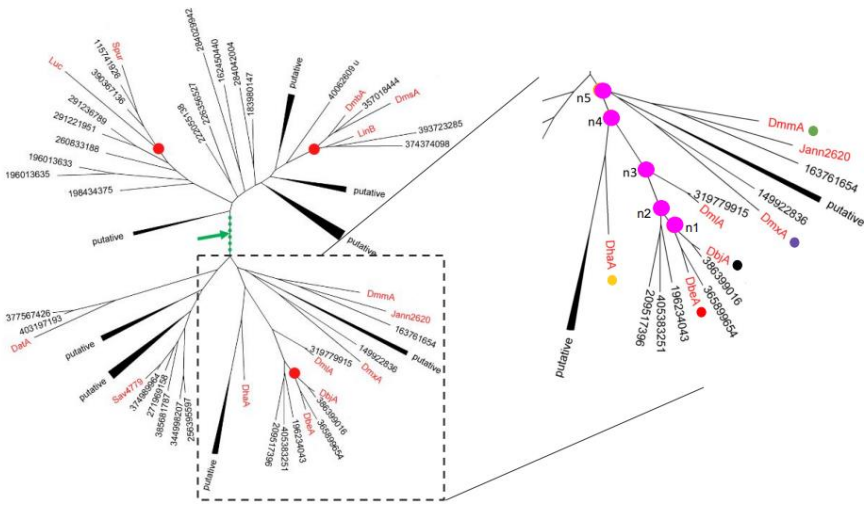
❑ SEQUENCE HOMOLOGY APPROACH

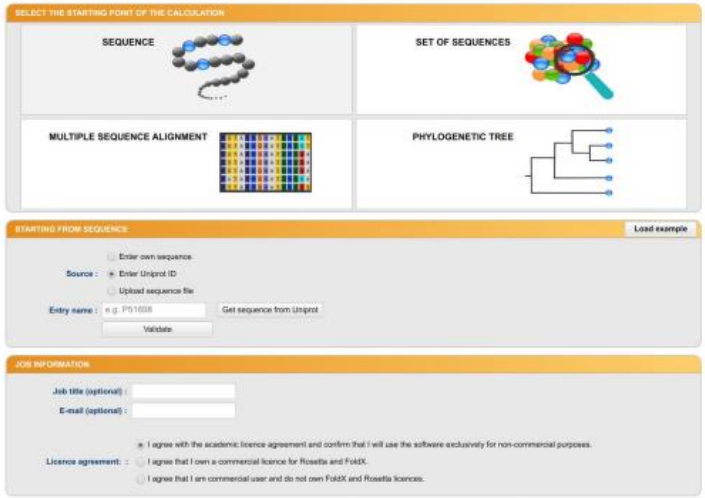
- homologous wild-type **sequences alignment**
- **identifying** amino acid residues responsible for differences
- **design** - combination of possitive mutation from all parental proteins

❑ ANCESTRAL RECONSTRUCTION

- construction of **phylogenetic tree**
- **design -** nods prediction by **consensus approach**

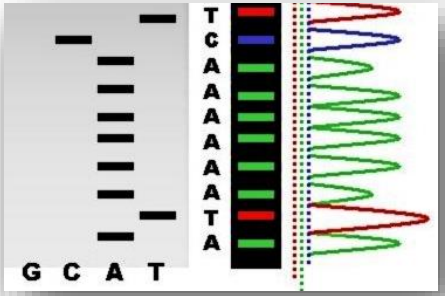
FIREPROT ASR

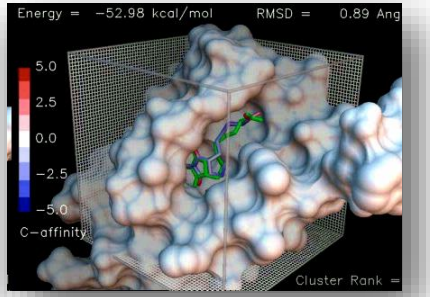




Bioinformatika Bi5000

- Období: podzim
- Rozsah: přednáška 2 hodiny/týden, cvičení 2 hodiny/týden
- Vyučující: prof. Mgr. Jiří Damborský, Dr., doc. RNDr. Roman Pantůček, Ph.D.,
- Osnova:
  - bioinformatické databáze a jejich prohledávání
  - analýza nukleotidových a proteinových sekvencí
  - hledání a identifikace genů
  - analýza a předpověď struktury proteinů

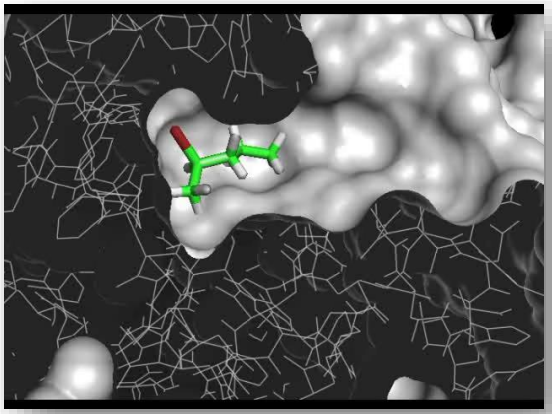


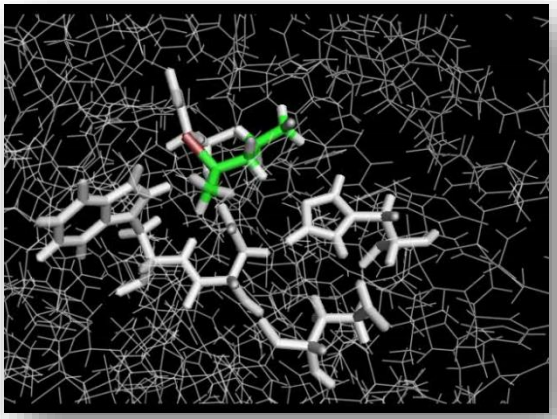


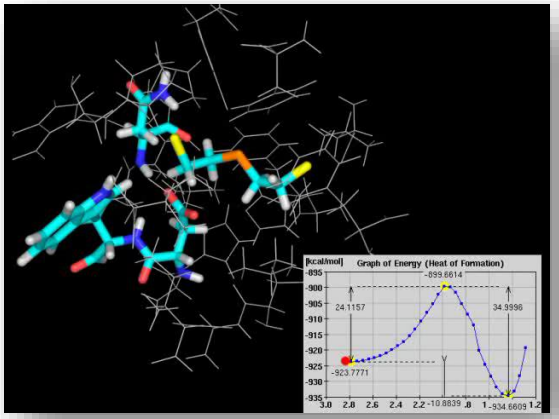
Design

❑ STRUCTURE-BASED APPROACH

- **prediction** of enzyme function from structure alone is challenging
- **protein structure** (X-ray crystallography, NMR, *homology models!*)
- **molecular modelling**
  - molecular docking
  - molecular dynamics
  - quantum mechanics/molecular mechanics (QM/MM)

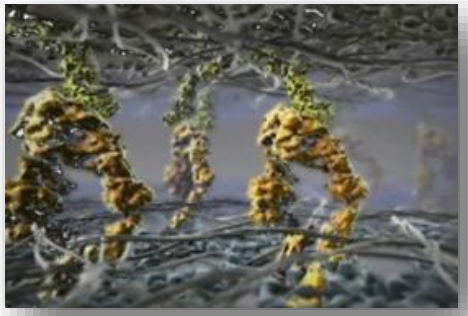


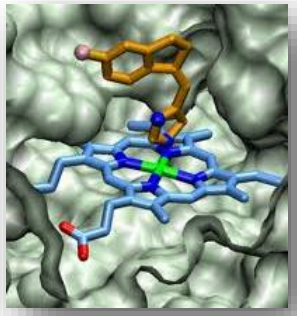


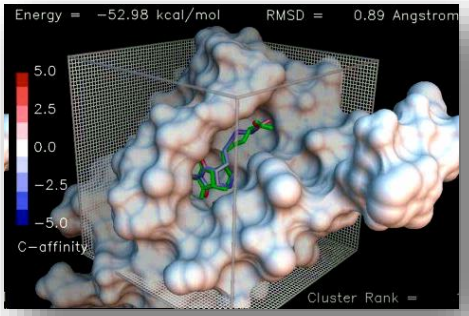


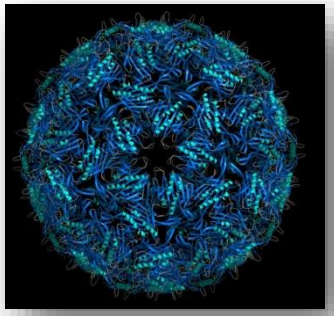
Strukturní biologie Bi9410

- Období: podzim
- Rozsah: přednáška 2 hodiny/týden, cvičení 2 hodiny/týden
- Vyučující: Mgr. David Bednář
- Osnova:
  - struktura, stabilita a dynamika biologických makromolekul
  - makromolekulární interakce a komplexy
  - stanovení a předpověď struktury, identifikace důležitých oblastí
  - stanovení vlivu mutace na strukturu a funkci proteinu
  - aplikace v biologickém výzkumu, návrhu léčiv a biokatalyzátorů



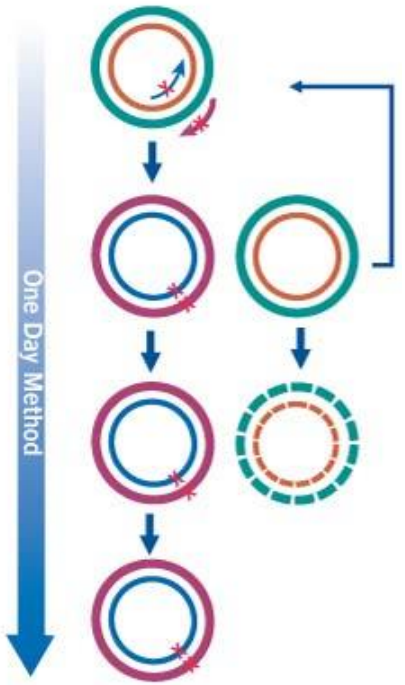






Construction

- **site-directed mutagenesis**
  - introducing point mutations
- **multi site-directed mutagenesis**
- **gene synthesis**
  - commercial service
  - codone optimisation

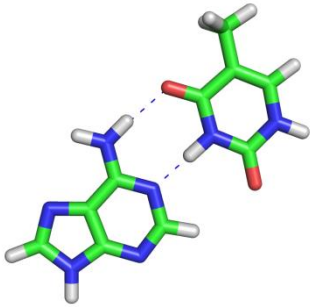


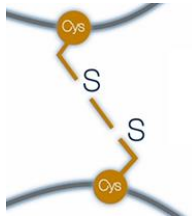
GENEART
THE GENE OF YOUR CHOICE

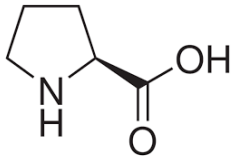
GenScript
Make Research Easy

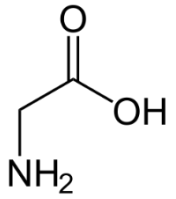
Example of rational design

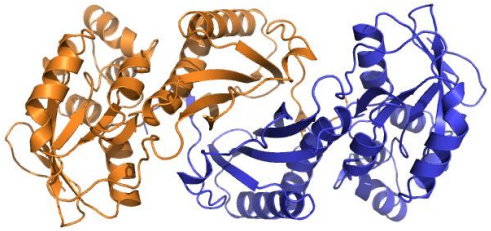
- rational design of protein **stability**
  - stability to high temperature, extreme pH, proteases etc.
  - **stabilizing mutations** increase strength of weak interactions
    - **salt bridges and H-bonds**
      Eijsink et al., Biochem. J. 285: 625-628, 1992
    - **S-S bonds**
      Matsumura et al., Nature 342: 291-293, 1989
    - **addition of prolines**
      Watanabe et al., Eur. J. Biochem. 226: 277-283, 1994
    - **less glycines**
      Margarit et al., Protein Eng. 5: 543-550, 1992
    - **oligomerisation**
      Dalhus et al., J. Mol. Biol. 318: 707-721, 2002





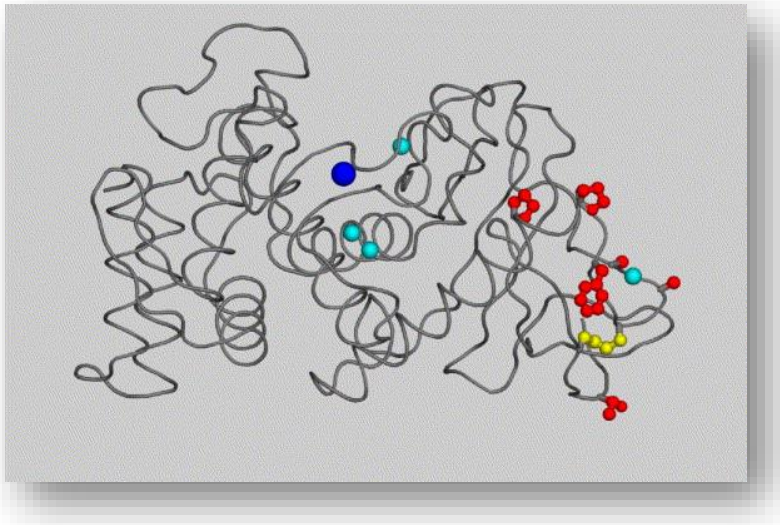






Example of rational design

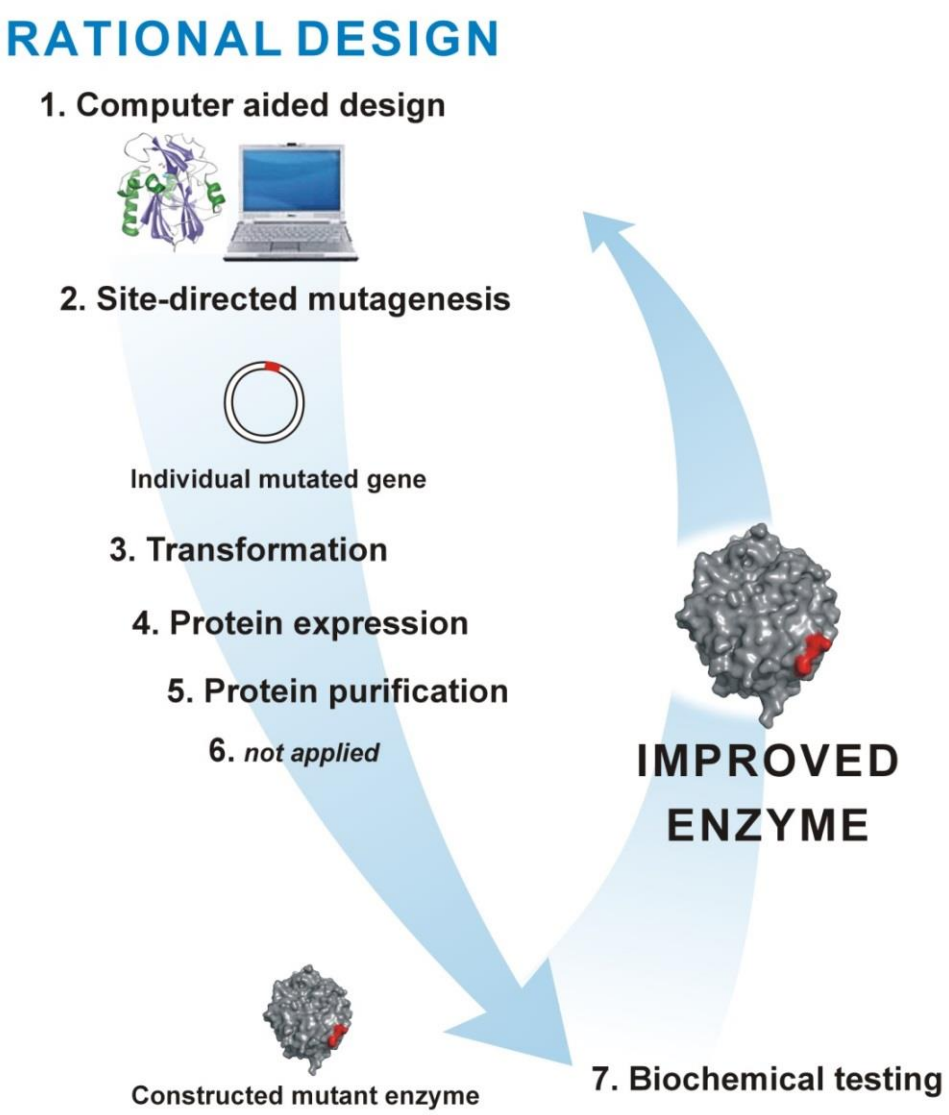
- **engineering protein to resist boiling**
  - **reduced rotational freedom**
    Ser65Pro, Ala96Pro
  - **introduction of disulfide bridge**
    Gly8Cys + Asn60Cys
  - **improved internal hydrogen bond**
    Ala4Thr
  - **filling cavity**
    Tyr63Phe



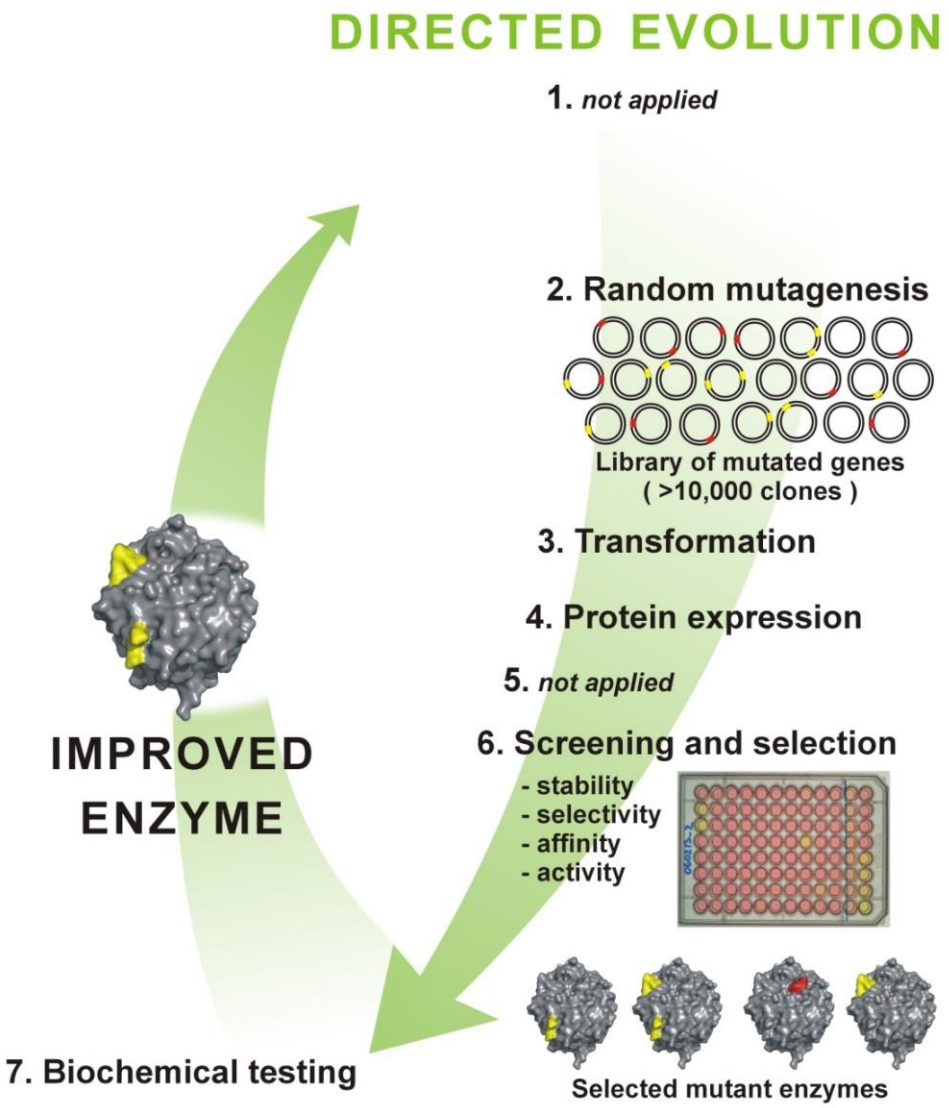
Half-lifes (min.)	80°C	100°C
wild type	17.5	>0.5
mutant	stable	170

Burg, B., et al., 1998. PNAS 95: 2056-2060

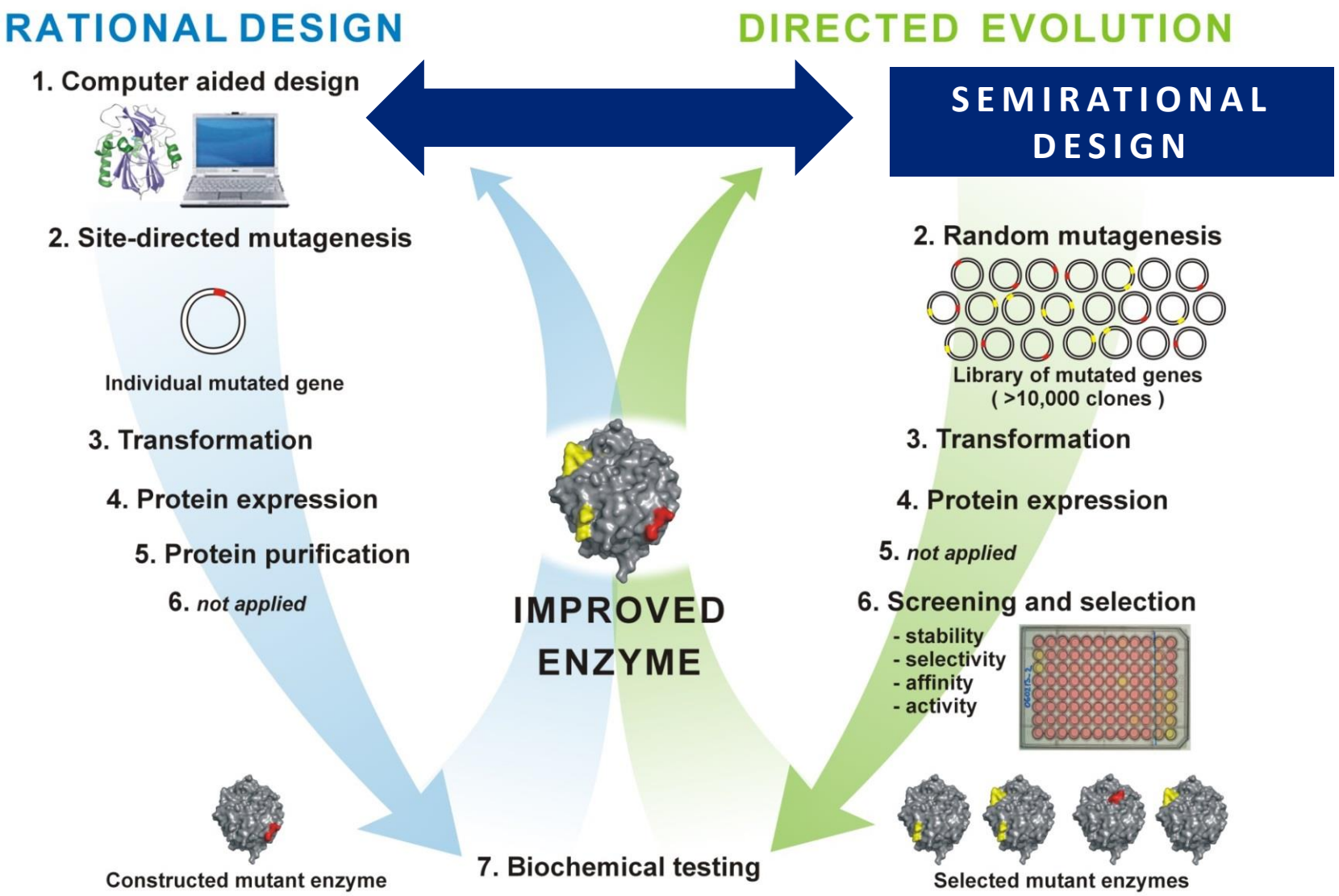
Strategies in protein engineering



Strategies in protein engineering



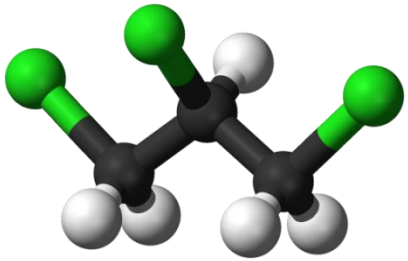
Strategies in protein engineering



Example of semi-rational design

- conversion of 1,2,3-trichloropropane

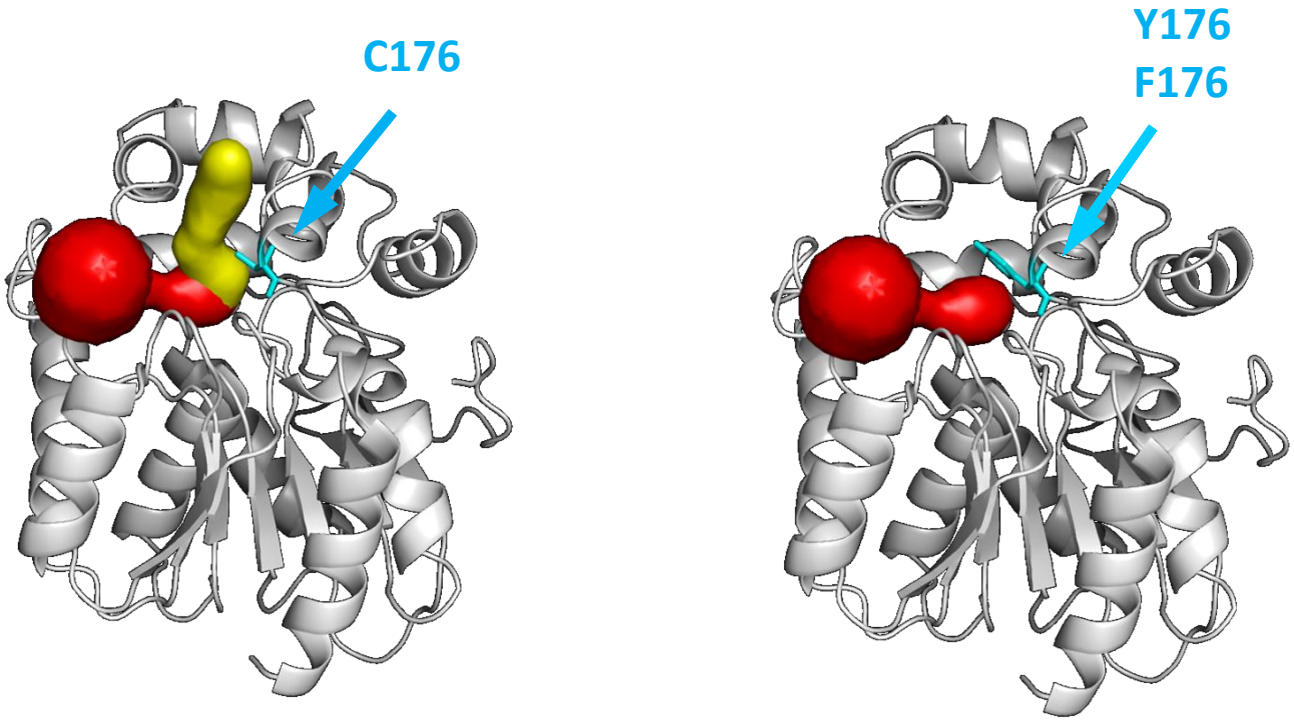
  by DhaA from *Rhodococcus erythropolis* Y2



Example of semi-rational design

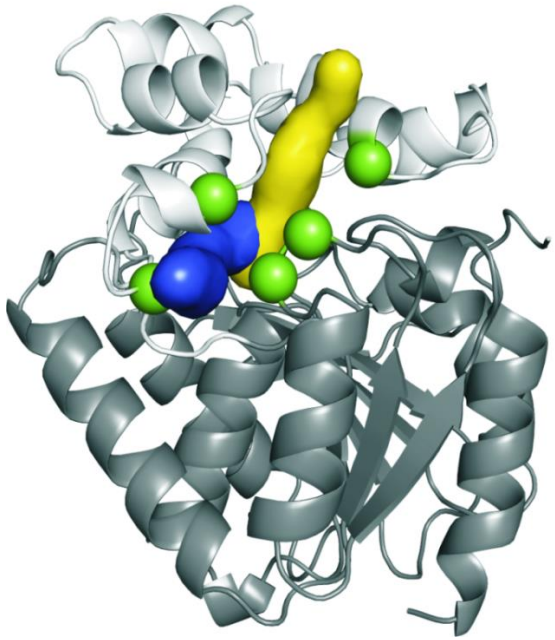
- conversion of 1,2,3-trichloropropane

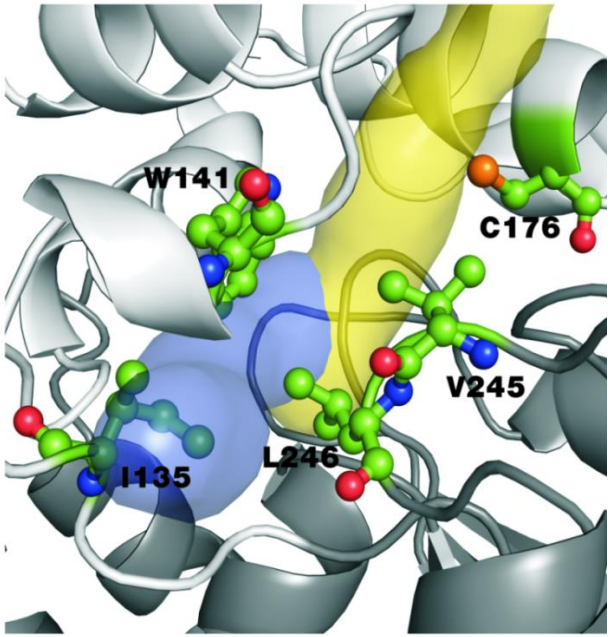
  by DhaA from *Rhodococcus erythropolis* Y2
- **DIRECTED EVOLUTION** - importance of access pathways



Example of semi-rational design

- conversion of 1,2,3-trichloropropane
  by DhaA from *Rhodococcus erythropolis* Y2
- **DIRECTED EVOLUTION -** importance of access pathways
- **SEMI-RATIONAL DESIGN** - hot spots in access tunels
- library of **5,300 clones** screened



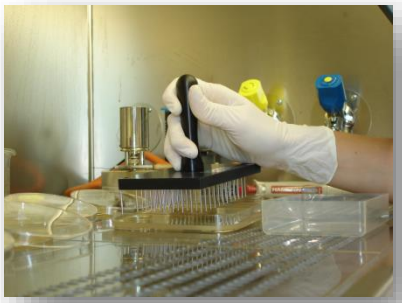


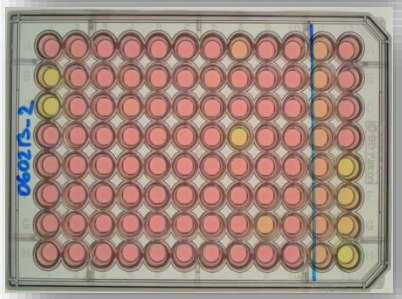
CAVER WEB

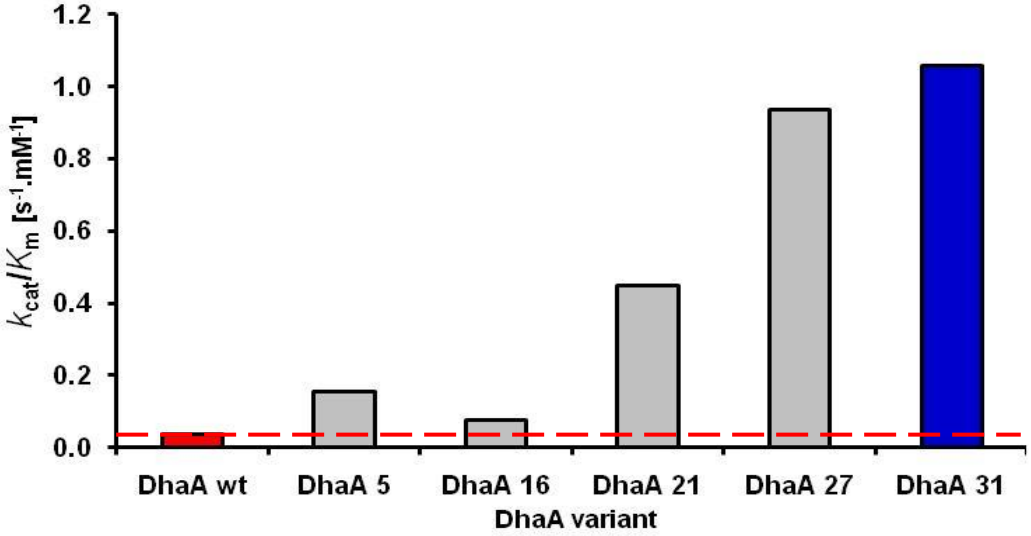
HOTSPOT WIZARD

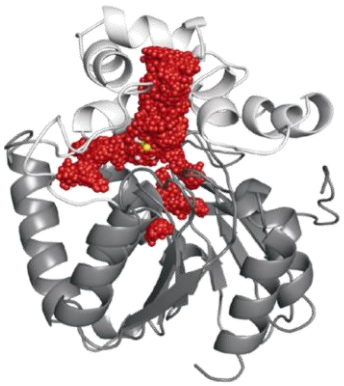
Pavlova, et al. 2009: Nature Chem. Biol. 5: 727-733

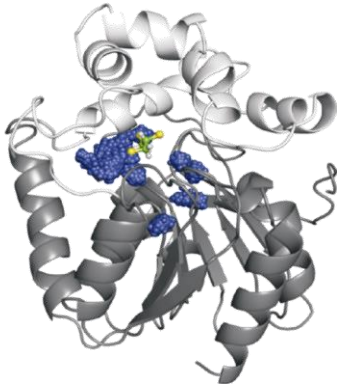
Example of semi-rational design

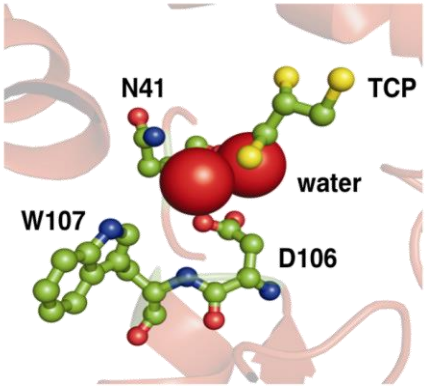


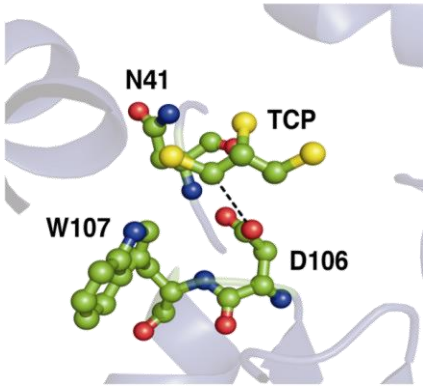












Experimental throughput is critical

STANDARD DESIGN

- random mutagenesis (2-3 positions)
- library of **10^4 clones**

volume: 100´ µL
assays/day: 10^3

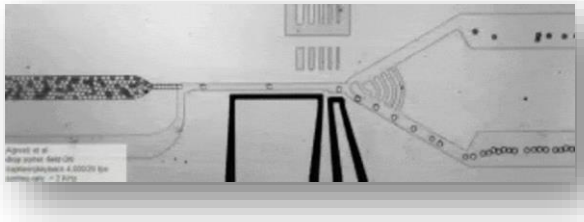
ADVANCED DESIGN

- random mutagenesis (5-7 positions)
- library of **$>10^6$ clones**

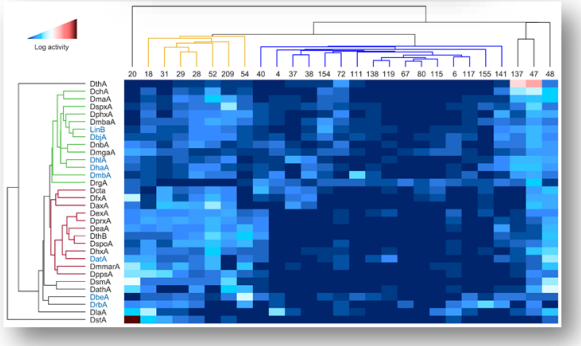
volume: 10´ pL
assays/day: 10^7

AI in Protein Engineering

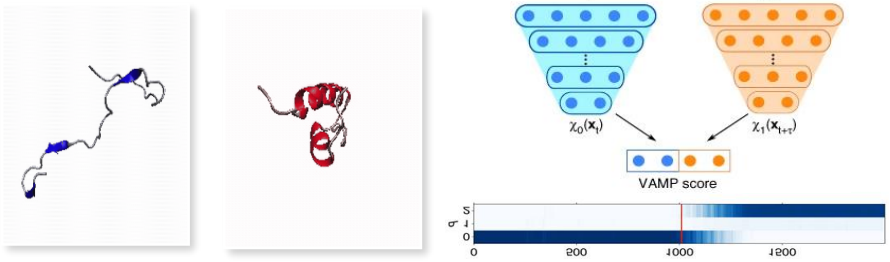
DEEP MUTATIONAL SCANNING
supervised learning



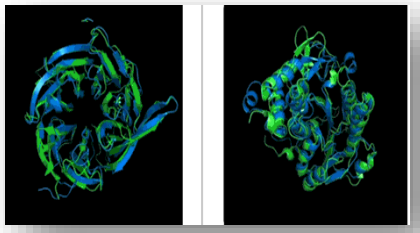
SEQUENCE BASED PREDICTION
supervised learning



MOLECULAR DYNAMICS
unsupervised learning



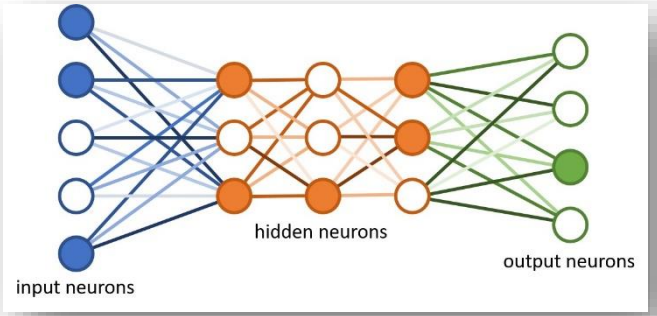
STRUCTURE PREDICTION
deep learning



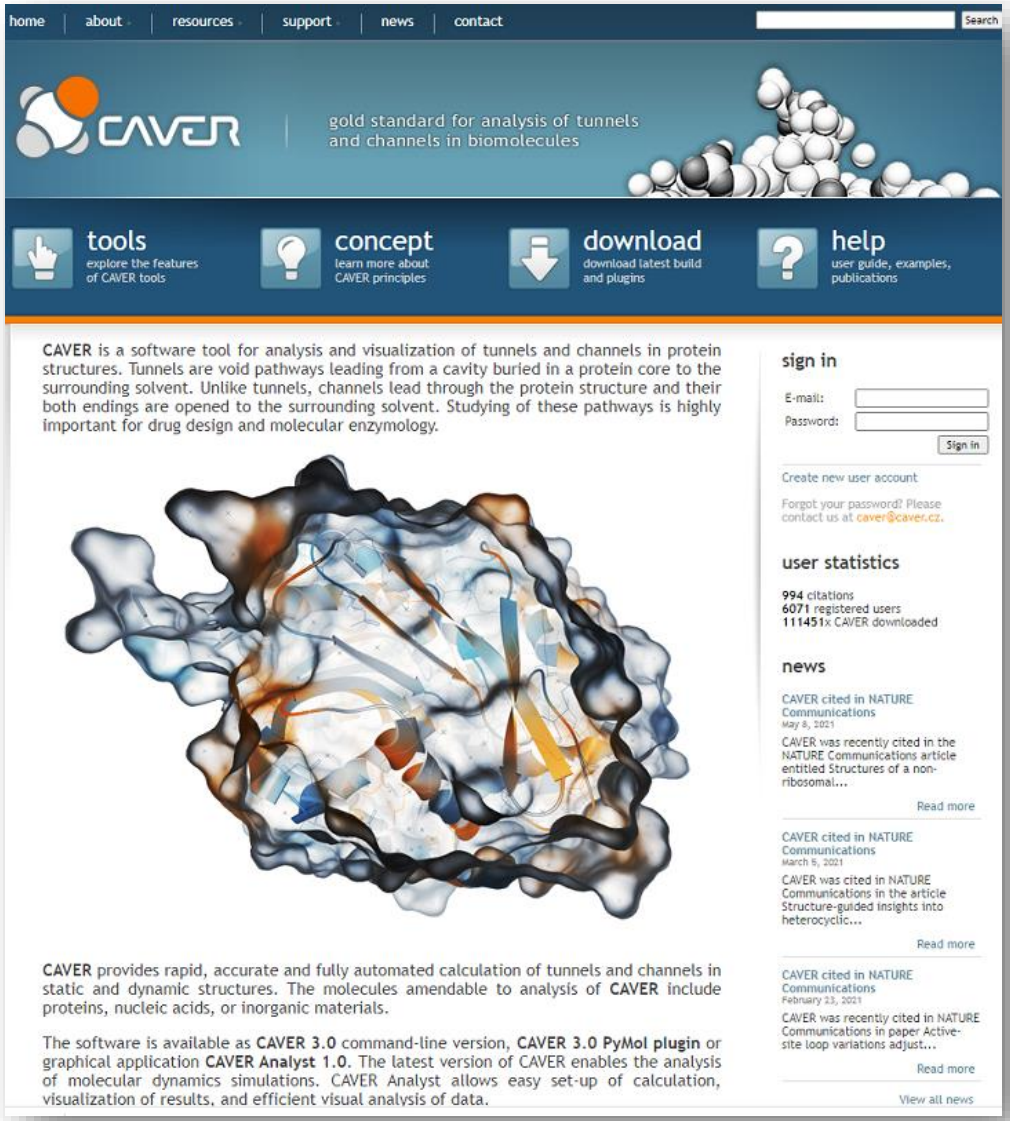
❑ *… next week (Lesson 7)*

AI in Biology, Chemistry, and Bioengineering Bi9680En

- Období: podzim
- Rozsah: přednáška 2 hodiny/týden
- Vyučující: Dr. Stanislav Mazurenko
- Osnova:
  - modern bio-challenges: drug design, DNA interpretation, protein engineering
  - types of AI algorithms and workflow for designing predictors
  - clustering algorithms, random forests, artificial neural networks
  - features, databases, and predictors used in applications



Tools for protein engineering



Bioinformatics 34: 3586-3588 (2018)

CAVERDOCK

CAVER WEB

Bioinformatics 35: 4986-4993 (2019)
Nucleic Acids Res. 47: W414-W422 (2019)

HOTSPOT WIZARD

Nucleic Acids Res. 48, W356-W362 (2018)

FIREPROT

Nucleic Acids Res. 45, W393-W399 (2017)

FIREPROT ASR

Brief. Bioinform., bbaa337 (2020)

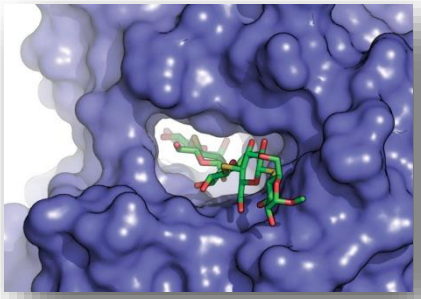
SOLUPROT

Bioinformatics 37, 23–28 (2021)

Proteinové inženýrství Bi7410

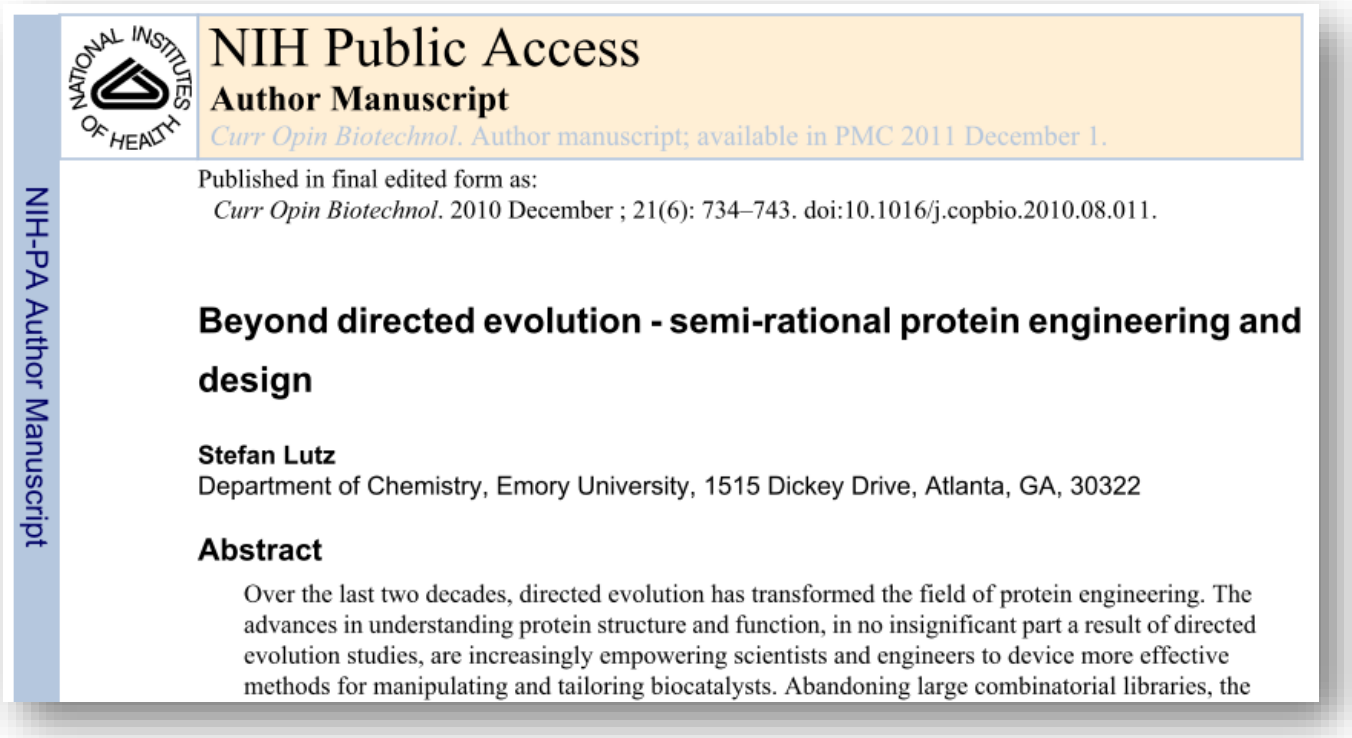
- Období: jaro
- Rozsah: přednáška 1 hodina/týden
- Vyučující: doc. Radka Chaloupková, Ph.D.
- Osnova:
  - strukturně-funkční vztahy proteinů
  - metody exprese a purifikace rekombinantních proteinů
  - metody strukturní a funkční analýzy proteinů
  - racionální design, semi-racionální design a řízená evoluce
  - příklady využití proteinového inženýrství






Reading

- Lutz, S. 2010: **Beyond directed evolution - semi-rational protein engineering and design**. *Curr Opin Biotechnol*. 21(6): 734–743
- *Computational enzyme redesign* and *Computational de novo enzyme design* (**page 5-7**)



NIH-PA Author Manuscript

NIH Public Access
Author Manuscript
Curr Opin Biotechnol. Author manuscript; available in PMC 2011 December 1.

Published in final edited form as:
Curr Opin Biotechnol. 2010 December ; 21(6): 734–743. doi:10.1016/j.copbio.2010.08.011.

Beyond directed evolution - semi-rational protein engineering and design

Stefan Lutz
Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, GA, 30322

Abstract

Over the last two decades, directed evolution has transformed the field of protein engineering. The advances in understanding protein structure and function, in no insignificant part a result of directed evolution studies, are increasingly empowering scientists and engineers to device more effective methods for manipulating and tailoring biocatalysts. Abandoning large combinatorial libraries, the