Protein Engineering – Mol Biotech Lecture #4 (16.10.2024)

1

LOSCHMIDT LABORATORIES

Protein Engineering

Molecular Biotechnology Lecture #4

Michal Vašina

16/10/2024

2

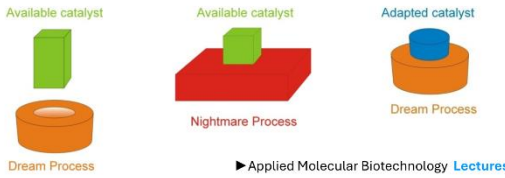
Outline

1. Proteins in biotechnology
2. Aims of protein engineering
3. Main strategies
   - ▶Directed evolution
   - ▶Rational design
   - ▶Machine learning
   - ▶Semi-rational design

3

Proteins in Biotechnology

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



▶Applied Molecular Biotechnology **Lectures 8-12**

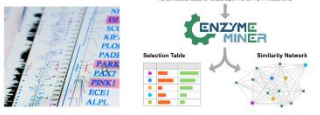
4

How to get new protein?

Classical Screening
- ▶screening culture collections
- ▶polluted and extreme environment

Environmental Gene Libraries
- ▶metagenomic DNA

Database mining
- ▶gene databases
- ▶(meta)genome sequencing projects
- ▶numerous uncharacterized proteins



Hon et al. *Nucleic Acids Research* 2020, link
https://loschmidt.chemi.muni.cz/enzymeminer/

5

How to get new protein?

Classical Screening
- screening culture collections
- pollutes

Environmental Gene Libraries
- metagenomic DNA

Database mining
- gene databases
- (meta)genome sequencing projects
- numerous uncharacterized proteins

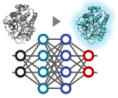
If suitable protein does not exist in nature?

▶ Protein Engineering

6

How to get new protein?

Protein Discovery	Directed Evolution	Rational Design	Machine Learning



7

Protein Engineering at a glance

▶ use of genetic manipulations to alter the coding sequence of a gene and thus **modify the properties of the protein**

AIMS AND APPLICATIONS

▶ **technological** - optimization of the protein to be suitable in particular technology process

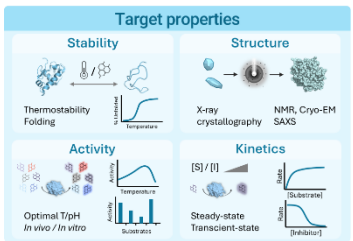
▶ **scientific** - desire to understand what elements of proteins contribute to folding, stability and function

8

What shall we improve?

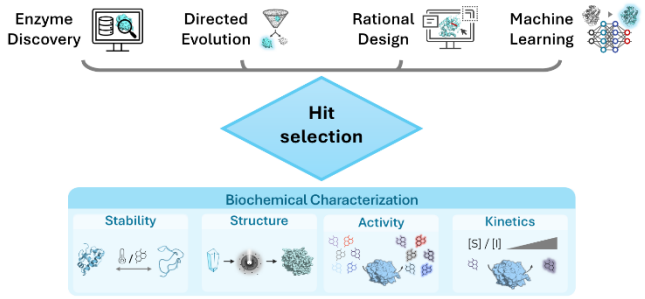
structural properties of proteins
- ▶ **stability** (temperature, solvents)
- ▶ tolerance to pH, salt
- ▶ Resolve the **atomic structure** (to understand function)

functional properties of proteins
- ▶ **substrate specificity** and selectivity
- ▶ **kinetic properties** (e.g., K_m, k_{cat}, K_i)
- ▶ Inhibition by small molecules (drugs)
- ▶ protein-protein or protein-DNA interactions



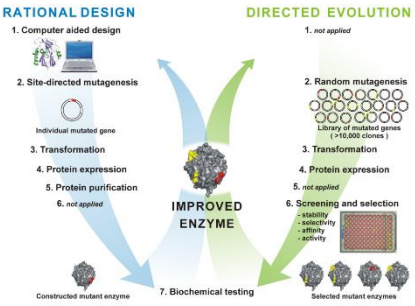
9

Overview of strategies



10

Main strategies



11

Directed Evolution

- ▶ emerged during mid-1990s
- ▶ inspired by natural evolution
- ▶ "laboratory evolution"
  - ▶ requires outside intelligence, not blind chance
  - ▶ does not take millions of years, but happens rapidly

The Nobel Prize in Chemistry 2018

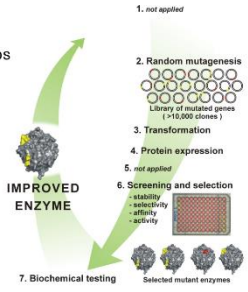
Frances H. Arnold
Prize share: 1/2

The Nobel Prize in Chemistry 2018 was divided, one half awarded to Frances H. Arnold "for the directed evolution of enzymes",

12

Directed Evolution

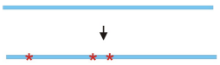
- ▶ **evolution in test tube** comprises two steps
  - ▶ **random mutagenesis**
    building mutant library (diversity)
  - ▶ **screening and selection**
    identification of desired biocatalyst
- ▶ **prerequisites** for directed evolution
  - ▶ gene encoding protein of interest
  - ▶ method to create mutant library
  - ▶ suitable expression system
  - ▶ screening or selection system



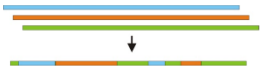
13

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

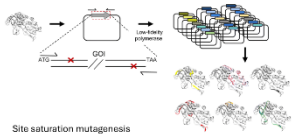


14

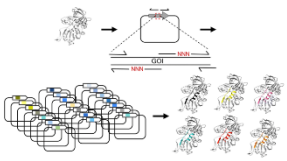
Non-recombining mutagenesis

- 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 mutations per 1,000 base pairs
- **site-saturation mutagenesis**
  - randomization of single or multiple codons
  - degenerate primers (NNN for complete randomization)
- **other methods**
  - insertion/deletions (InDel)
  - cassette mutagenesis (region mutagenesis)

Error-prone mutagenesis



Site saturation mutagenesis

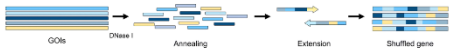


15

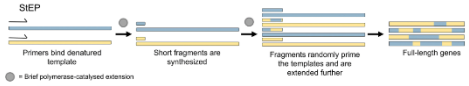
Recombining mutagenesis

- **DNA shuffling**
  - fragmentation step
  - random reassembly of segments
- **StEP - staggered extension process**
  - simpler then shuffling, no fragmentation
  - random reannealing combined with limited primer extension
- **other methods**
  shuffling of genes with lower homology down to 70% (e.g., RACHITT, ITCHY, SCRATCHY)

DNA-shuffling



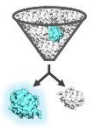
StEP



16

Screening and selection

- most critical step of direct evolution
- isolation of positive mutants hiding in library
- genotype to phenotype linkage is crucial

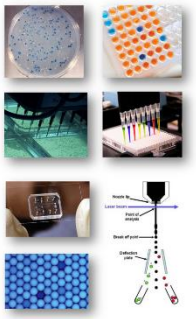


- **High-throughput screening**
  experimental testing of variants one by one
- **Direct selection**
  applying selective pressure to the library

17

(Ultra)-High throughput screening

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



18

Experimental throughput is critical

STANDARD DESIGN

- ▶Random mutagenesis (2-3 positions)
- ▶Library of **10^4 clones**

ADVANCED DESIGN

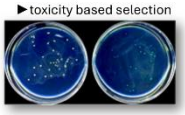
- ▶Random mutagenesis (5-7 positions)
- ▶Library of **> 10^6 clones**

▶Microfluidics **Lecture 7**

19

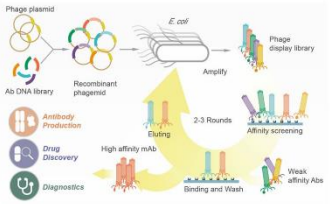
Direct selection

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



The Nobel Prize in Chemistry 2018

The Nobel Prize in Chemistry 2018 was divided, the other half jointly to George P. Smith and Sir Gregory P. Winter "for the phage display of peptides and antibodies"

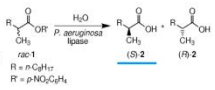


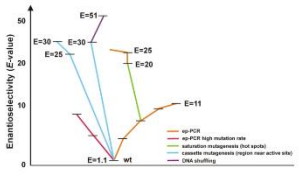
20

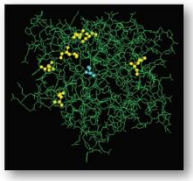
Success story #1

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

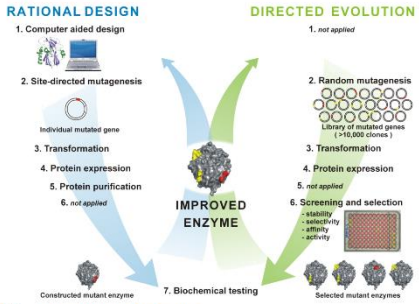






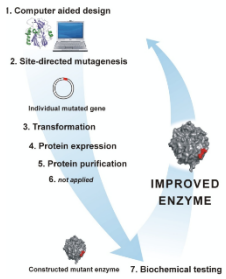
21

Main strategies



22

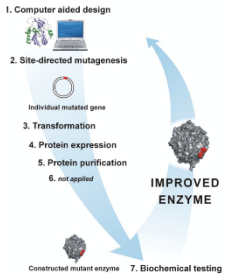
Rational design introduction



- ▶ emerged around 1980s as the original protein engineering approach
- ▶ knowledge based - combining theory and experiment
- ▶ protein engineering cycle: „learn-design-build-test-learn“
- ▶ difficulty in prediction of mutation effects on protein property
- ▶ *de novo* design most challenging

23

Principals of rational design

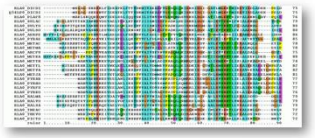


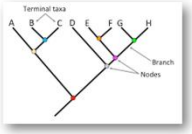
- ▶ rational design comprises:
  - ▶ **design** - understanding of protein functionality
  - ▶ **experiment** - construction and testing of mutants
- ▶ **prerequisites** for rational design:
  - ▶ **gene** encoding protein of interest
  - ▶ **3D structure** (e.g., X-ray, NMR) or sequence alignment
 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

24

Design

- ▶ **Sequence homology approach**
  - ▶ homologous wild-type sequences alignment
  - ▶ identifying amino acid residues responsible for differences
  - ▶ design - combination of positive mutation from all parental proteins
- ▶ **Ancestral reconstruction**
  - ▶ construction of phylogenetic tree
  - ▶ design - nods prediction by consensus approach

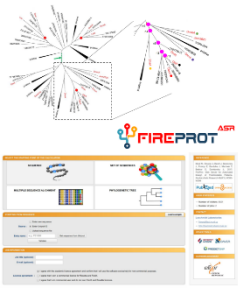




25

Design

- **Sequence homology approach**
  - homologous wild-type sequences alignment
  - identifying amino acid residues responsible for differences
  - design - combination of positive mutation from all parental proteins
- **Ancestral reconstruction**
  - construction of phylogenetic tree
  - design - nods prediction by consensus approach



Musil et al. *Brief Bioinform* 2020, link

https://loschmidt.chemi.muni.cz/fireprotasr/

26

Bioinformatika Bi5000/Bi5000c

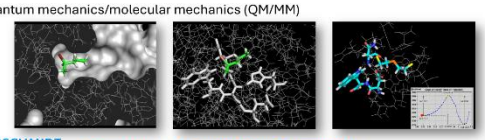
- Období: podzim
- Rozsah: přednáška 2 hodiny/týden, cvičení 2 hodiny/týden
- Vyučující: prof. Mgr. Jiří Damborský, Dr., prof. RNDr. Roman Pantůček, Ph.D.,
- Osnova:
  - bioinformatické databáze a jejich prohledávání
  - analýza nukleotidových a proteinových sekvencí
  - hledání a identifikace genů
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

27

Design

Structure-based approach

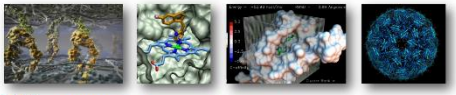
- **prediction** of enzyme function from structure alone is challenging
- protein structure: experimental (X-ray crystallography, NMR), computational (AlfaFold models, homology models!)
- molecular modelling
  - molecular docking
  - molecular dynamics
  - quantum mechanics/molecular mechanics (QM/MM)



28

Strukturní biologie Bi9410/Bi9410c

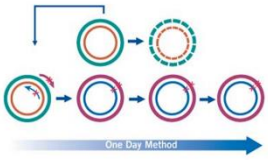
- Období: podzim
- Rozsah: přednáška 2 hodiny/týden, cvičení 2 hodiny/týden
- Vyučující: doc. Mgr. David Bednář, Ph.D.
- Osnova:
  - struktura, stabilita a dynamika biologických makromolekul
  - makromolekulární interakce a komplexy
  - stanovení a předpověď struktury, identifikace důležitých oblastí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



29

Gene of interest construction

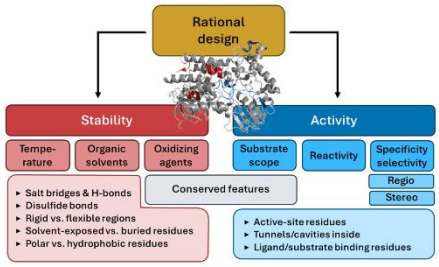
- ▶site-directed mutagenesis
  - ▶introducing point mutations
- ▶multi site-directed mutagenesis
- ▶gene synthesis
  - ▶commercial service
  - ▶codon optimization



GENEART THE GENE OF YOUR CHOICE GenScript Make Research Easy

30

Rational design targets



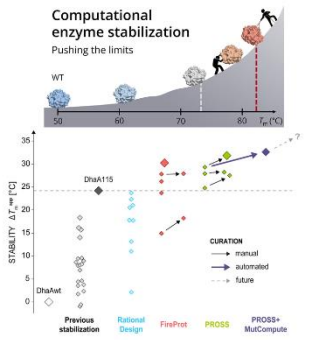
31

Success story #2

Stabilizing already stabilized enzyme

DhaA115: T_m = 73.3 °C (previously by FireProt)

1. Introduction of disulfide bridges
   - ▶no increase in T_m
2. Automated platforms FireProt and PROSS
   - ▶FireProt : best T_m = 77.0 °C
   - ▶PROSS : best T_m = 78.4 °C
3. Further stability increase by **manual curation**
   - ▶FireProt : best T_m = 79.3 °C
   - ▶PROSS : best T_m = 80.9 °C
4. Automated curation by **machine learning**
   - ▶MutCompute : T_m = 81.7 °C



32

Targets for Machine Learning

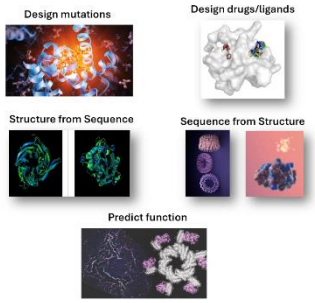
Designing molecules

- ▶Design mutations/protein variants
- ▶Design drugs/ligands to bind proteins

Predictions

- ▶Structure prediction (AlphaFold, ...)
- ▶Sequence from structure (find binding proteins)
- ▶Function from sequence

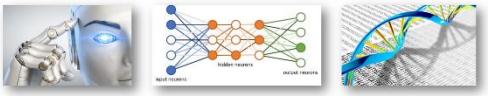
- ▶AI in Life Sciences **Lecture 6**



33

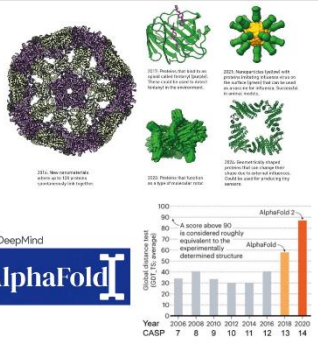
AI in Biology, Chemistry, and Bioengineering Bi9680En

- ▶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



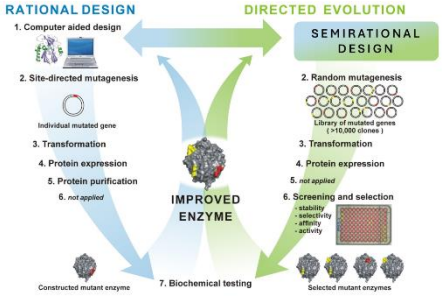
34

Nobel Prize in Chemistry 2024



35

Main strategies



36

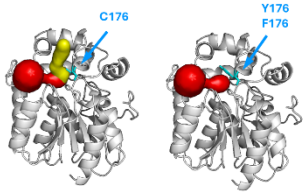
Success story #3: Degrading a toxic pollutant

- ▶conversion of 1,2,3-trichloropropane
  by DhaA from *Rhodococcus erythropolis* Y2

37

First round: Directed evolution

- ▶conversion of 1,2,3-trichloropropane by DhaA from *Rhodococcus erythropolis* Y2
- ▶**Directed Evolution** - importance of access pathways



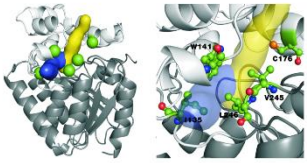
Variant	k_{cat} (s^{-1})
wt	0.08
C176Y+Y273F[1]	0.28
G3D+C176F[2]	0.32

[1] Bosma, et al. ***AEM*** 2002, link,
[2] Gray et al. ***Adv. Synth. Catal.*** 2001, link

38

Second round guided by structural insights

- ▶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

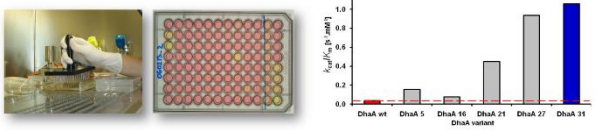


CAVER WEB
HOTSPOT WIZARD

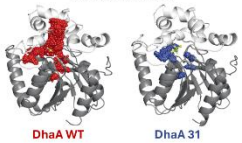
Pavlova, et al. ***Nat Chem Biol*** 2009, link

39

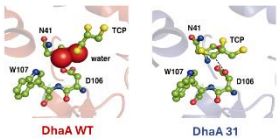
Results



Accessible solvent



Active site



DhaA WT DhaA 31 DhaA WT DhaA 31

Pavlova, et al. ***Nat Chem Biol*** 2009, link

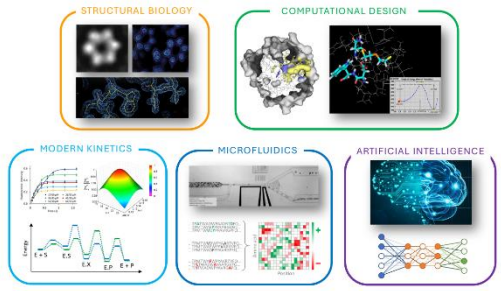
40

Proteinové inženýrství Bi7410

- ▶Období: jaro
- ▶Rozsah: přednáška 2 hodiny/týden
- ▶Vyučující: Mgr. Michal Vašina, Ph.D. , doc. Mgr. David Bednář, 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

41

Multidisciplinary in protein research

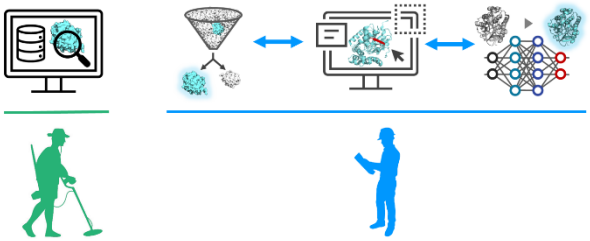


LOSCHMIDT LABORATORIES Summary 42

42

Combine multiple strategies

Protein Discovery | Directed Evolution | Rational Design | Machine Learning

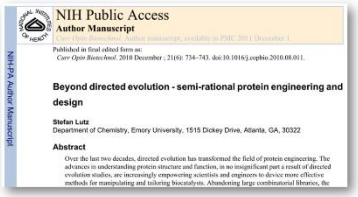


LOSCHMIDT LABORATORIES Summary 43

43

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 Public Access
Author Manuscript

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 devise more effective methods for manipulating and tailoring biocatalysts. Abandoning large combinatorial libraries, the

LOSCHMIDT LABORATORIES 45