CHAPTER 2

PROTEINS
Part 2 – protein engineering

Protein engineering
RATIONAL DESIGN
- Site directed mutagenesis of one or more residues
- Fusion of functional domains from different proteins to create chimaeric (Domain swapping)

Functional evaluation

A protein library having the mass of our galaxy could only cover the combinatorial possibilities for a peptide with 50 residues

Therefore even genetic selection approaches for designing novel functional proteins will not generally build on fully random sequences, but will be based on existing protein scaffold that serve as template.

Proteins with Novel Properties
Random Mutagenesis

Nature
Rational Protein Design

- Enhance stability/function under new conditions
  - temperature, pH, organic/aqueous solvent, [salt]
- Alter enzyme substrate specificity
- Enhance enzymatic rate
- Alter epitope binding properties
In order to consider the rational design of a target enzyme, one needs to have several pieces of information:

1. A cloned gene coding for the enzyme.
2. The sequence of the gene.
3. Information on the chemistry of the active site, ideally one would know which amino acids in the sequence are involved in activity.
4. Either a crystal/NMR structure for of the enzyme, or the structure of another protein displaying a high degree of structural homology.

The above information is needed in order to have a clear idea of which amino acids one should mutate to which likely effect.

Typically, protein engineering is a cyclic activity involving many scientists with different skills:

Engineering of lumazine synthase

wt AaLS:
MEIYEGKLTA EGLRFGIVAS RFNHALVDRL VEGAIDCIVR HGGREDITL VRVPGSWEIP
VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT
LEQAIERAGT KHGNKGWEAA LSAIEMANLF KSIBLEHHHH HH***

162 Amino acids

The A. Aeolicus lumazine synthase – origin and mode of action

Electron microscopy wtAaLS

A. Aeolicus - warme Quellen (80-95 °C)
(Ane K.G. Jenny & R. Rachel, University of Regensburg; springs: Octopus Spring, Yellowstone National Park, Allen Treiman, UPH)

Engineering of lumazine synthase

wt AaLS:
MEIYEGKLTA EGLRFGIVAS RFNHALVDRL VEGAIDCIVR HGGREDITL VRVPGSWEIP
VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT
LEQAIERAGT KHGNKGWEAA LSAIEMANLF KSIBLEHHHH HH***

Arg83Glu/Thr86Glu/Thr120Glu/Gln123Glu
300 extra charges per 60-meric capsid

The Overlap-Extension PCR

Recombinant lumazine synthase from *Aquifex aeolicus*. The lumazine synthase gene from *A. aeolicus* was amplified from plasmid pNCO-AA-ribH. by PCR using the primers 200304AQS (GATATACGTAATAACGAACTAACTAATTAACTA) and 200304AQA (GATACCTCAGATCCGAGACGAGCTGGAATAAGT). The restriction sites are underlined. The NcoI/XhoI double digested fragment was ligated into a modified version of the pMG209 vector.

Mutations R83E, T86E, Q123E and T120E were introduced by standard overlap-extension PCR using Pfu-polymerase and the following additional primers:

Arg83 Glu/Thr86 Glu
200304NEG1 (GTTCTCA
GAA
GGGGCA
GAG
CCACA
TTTC)
200304NEG2 (GAAA
TGTTG
CTC
TGCCCC
TATGGAGAAC).

Thr120 Glu/Gln123 Glu
203004NEG3 (CAGCTGAC
GAA
TTGGAA
GAG
GCTA
TCGAG)
203004NEG4 (CTCGA
TAGC
CTC
TTCCAA
TATGGCTG).

The altered codons are shown in bold. All coding portions of the constructed plasmids were confirmed by DNA sequencing.

Recombinant GFP-Rs. The plasmid pMG-GFP encodes GFP from *Aequorea victoria* with an inframe, C-terminal His-tag. We replaced the His-tag with a deca-arginine tag by ligating a duplex of the primers 220304polyRs (TCGAG
CGT
AGA
CGA
CGC
CGT
CGG
CGA
CGT
CGA
CGT
GAA
GGGGCA
GAG
CCACA
TTTC) and 220304polyRa (CTAGTTAACGTCGACGTCGCCGACGGCGTCGTCTACGC) into pMG-GFP that was linearized by double digestion with XhoI and SpeI. The resulting low copy number plasmid (pAC4C-GFP-R) encodes GFP-Rs (under control of the T7 and salicylate promoters) and also chloramphenicol resistance.

This plasmid served as a template for site-directed mutagenesis in order to generate a construct (pAC4C-GFP) encoding GFP that lacks any C-terminal tag. Mutations (in bold) were introduced using the QuickChange® site-directed mutagenesis kit (Stratagene) with the primers 050804killargS (CTCGAGCGT
TAA
CGACGCCGT) and 050804killargA (ACGGCGTCG
TTA
ACGCTCGAG)

UAA – Stop (Ochre)


Engineering of lumazine synthase


HIV protease (toxic)

B. Waenderdorfer, K. J. Woycechowsky, D. Hilvert Science 2011, 321, 580-582
Taq DNA polymerase lacks 3' to 5' exonuclease activity resulting in a low error rate of 0.001-0.002% per nucleotide per replication.

- choosing the gene, or the area within a gene, one wishes to mutate.
- the extent of error required is calculated based upon the type and extent of activity one wishes to generate.
- This extent of error determines the error prone PCR strategy to be employed.
- Following PCR, the genes are cloned into a plasmid and introduced to competent cell systems.
- These cells are then screened for desired traits.
- Plasmids are then isolated for colonies which show improved traits, and are then used as templates the next round of mutagenesis.

Rates of error in PCR can be increased in the following ways:

- Increase concentration of magnesium chloride, which stabilizes non-complementary base pairing.
- Add manganese chloride to reduce base pair specificity.
- Increased and unbalanced addition of dNTPs.
- Addition of base analogs like dITP, 8-oxo-dGTP, and dPTP.
- Increase concentration of Taq polymerase.
- Increase extension time.
- Increase cycle time.
- Use less accurate Taq polymerase.

---

DNA Shuffling

DNA Shuffling is a way to rapidly propagate beneficial mutations in a directed evolution experiment. It is used to rapidly increase DNA library size. First, DNase is used to fragment a set of parent genes into pieces of 50-100 bp in length. This is then followed by a polymerase chain reaction (PCR) without primers - DNA fragments with sufficient overlapping homologous sequence will anneal to each other and are then extended by DNA polymerase. Several rounds of this PCR extension are allowed to occur, after some of the DNA molecules reach the size of the parental genes. These genes can then be amplified with another PCR, this time with the addition of primers that are designed to complement the ends of the strands. It is possible to recombine portions of these genes to generate hybrids or chimeric forms with unique properties, hence the term DNA shuffling.

---

Error prone PCR

Error-prone PCR library construction. The plasmid pMG-AaLS-neg - template in an error-prone PCR (epPCR). The primers AQs (GATATACATGGAAATCTAAGGAAGG) and AQa (GATATCTCGACGGAGAGACTTGAAATAG), flanking the coding region of the capsid gene, were employed for amplification.

The resulting epPCR product (482 bp) was purified by agarose gel electrophoresis. The epPCR product and the acceptor vector pMG-AaLS-neg were digested with NcoI and XhoI. The fragments were ligated with T4 DNA ligase to give a plasmid library containing the mutant capsid genes.

DNA Shuffling library construction. For the second and third rounds of evolution, genes encoding capsid variants selected in the previous round were subjected to DNA shuffling. Capsid genes were amplified (PCR, primers AQs and AQa), then digested with DNaseI (6 ng/μl) to give 50 bp to 100 bp fragments. The fragments were purified and assembled in a PCR-like process, but without primers. The reassembled genes were amplified in a final PCR using primers AQs and AQa, digested with NcoI and XhoI, and ligated into the pMG vector, desalted and concentrated to afford plasmid libraries containing mutant capsid genes.
Selection: Capsid mutants produced under conditions of high protease content
AaLS plasmids + the plasmid pACYC-tetHIV-R10, encoding the R-10-tagged HIV protease
→ electrocompetent E. coli XL1-Blue. Growth in presence of ampicillin (150 μg/ml) and chloramphenicol (30 μg/ml).
Capsid production was induced by salicylate (100 μM).
The HIV protease-R10 gene was induced with 400 to 1600 ng/ml tetracycline.

Internal standards / negative controls: plasmids of AaLS-wt or AaLS-neg, + pACYC-tetHIV-R10
After the 4th round of evolution, 24 surviving capsid variants were isolated, sequenced, and characterized.

For detailed biophysical characterization, AaLS variants were overproduced using the T7 promoter system in E. coli KA13 cells that had been transformed with the appropriate plasmid (pMG-AaLS-wt, pMG-AaLS-neg, or pMG-AaLS-13). For coproduction with HIV protease, the cells were also transformed with either pACYC-tetHIV or pACYC-tetHIV-R10.

B. Woehl, K.J. Woycechowsky, D. Hilvert
Science 2011, 331, 589-592
Engineering of lumazine synthase

B. Woersdorfer, K.J.Woycechowski, D.Hilvert Science 2011, 331, 589-592

Engineering of lumazine synthase


Engineering of lumazine synthase

Bottom-up Construction of a Primordial Carboxysome Mimic

**Engineering of lumazine synthase**

Modular Protein Cages for Size-Selective RNA Packaging in Vivo

Y. Azuma, T. Edwardson, N. Terasaka, D. Hilvert


Enzyme-mediated polymerization inside engineered protein cages

Y. Azuma, M. Herger, D. Hilvert


Oxidation of 3,3-diaminobenzidine (DAB) by the engineered ascorbate peroxidase APEX2 encapsulated in AaLS capsids resulted in templated formation of polyDAB–capsid nanoparticles of homogeneous size and shape.

R. Frey, T. Hayashi, D. Hilvert

*Chem. Commun.*, 2016, 52, 10423-10426

**Protein engineering – de novo enzyme design**

H. Kries, R. Blomberg, D. Hilvert

*Curr. Opin. Chem. Biol.*, 2013, 17, 1-8

<table>
<thead>
<tr>
<th>Protein engineering – de novo enzyme design</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Diagram" /></td>
</tr>
<tr>
<td><img src="image2" alt="Diagram" /></td>
</tr>
</tbody>
</table>

H. Kries, R. Blomberg, D. Hilvert

*Curr. Opin. Chem. Biol.*, 2013, 17, 1-8

![Diagram](image3)
Four active designs in different scaffolds have been obtained by combining the oxyanion hole motif with a Cys-His dyad. Rapid acylation of active site cysteines followed by slow hydrolysis of the acyl-enzyme intermediate limits overall catalytic efficiency.

Minimal requirements for esterase activity have been explored by computationally designing artificial catalysts using catalytic dyads and oxyanion holes.

The evolutionary origin is difficult to track back.

Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis

**Nucleophilic catalysis** is a general strategy for accelerating ester and amide hydrolysis.

In natural active sites, nucleophilic elements such as catalytic dyads and triads are usually paired with oxyanion holes for substrate activation. The evolutionary origin is difficult to track back.

The most active variant:

\[
\frac{k_{\text{cat}}}{K_M} = 400 \text{ M}^{-1} \text{s}^{-1}
\]

(p-nitrophenyl ester cleavage).

A Cys-His dyad, in combination with an oxyanion binder, were used to hydrolyze activated esters via covalent catalysis.

**Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis**

Programmed mechanism and model substrates of the de novo designed esterases.

A tyrosyl ester served as the target substrate for computational design; the fluorogenic coumarin ester and the chromogenic p-nitrophenyl ester were used for screening purposes.

Crystal structures of the four active designs. In each case, the design model is shown in purple with the ligand in cyan, and the crystal structure in green. The theozyme residues and the ligand are shown in stick representation, and selected other active site residues in line representation.
A B. subtilis globin variant, engineered by directed evolution, catalyzes the cyclopropanation of 3,4-difluorostyrene to make the desired stereoisomer of a ticagrelor precursor with high selectivity and yield.

Directed evolution – bringing new chemistry to life

Production of chiral organoboranes by E. coli expressing *Rhodothermus marinus* cytochrome c. The bacterial catalyst uses borane-Lewis base complexes and diazo reagents to construct boron-containing carbon stereocenters efficiently and selectively in cells by carbene insertion. The bioconversion can be conducted readily on gram scale, and the enatio-preference of borylation was switched to give either enantiomer of the organoborane products.
Evolution of a metalloenzyme from short peptides

Zinc-mediated assembly of helix-turn-helix fragments, followed by fusion and asymmetric diversification, afforded MID1sc10, an efficient metalloesterase.

Michaelis-Menten plots for MID1sc (yellow and inset) and MID1sc10 (green) show a 70,000-fold improvement in hydrolysis efficiency for (S)-configured 1 after optimization.

The evolved variant MID1sc10 is highly enantioselective as a consequence of a 2200-fold specificity switch from the modestly (R)-selective starting catalyst MID1sc.

Crystal structure of MID1sc10
- Zinc ion - orange sphere
- Coordinating histidines - green sticks
- Linkage of two polypeptides – orange sticks
- Beneficial mutations - magenta spheres
- Residues replaced to prevent competitive zinc binding modes - cyan spheres.

The observed structural changes transformed the shallow binding site of MID1 (gray) into a deep, hydrophobic pocket in MID1sc10 (green).

Cut-away view of the active site, showing the snug fit of phosphonate 4 in the binding pocket, (zinc ion – orange)

View of the MID1sc10 active site with phosphonate 4 (yellow) coordinating to the Zn(His)₃ complex (orange sphere and green sticks). Arg68 and Gin68 form mechanically relevant hydrogen bonds to phosphonate 4 and the backside nitrogen of His61, respectively.