CHAPTER 2

PROTEINS

Part 2 – protein engineering

N. Ochotny

J. Swaminathan
Proteins with Novel Properties

Random Mutagenesis

- Enhance stability/function under new conditions
  - temperature, pH, organic/aqueous solvent, [salt]
- Alter enzyme substrate specificity
- Enhance enzymatic rate
- Alter epitope binding properties

RATIONAL DESIGN

- Site directed mutagenesis of one or more residues
- Fusion of functional domains from different proteins to create chimaeric (Domain swapping)
- Functional evaluation

Protein engineering: rational design and directed evolution

Nobel prize chemistry 2018
Frances Arnold

„for pioneering the use of directed evolution to engineer enzymes”
Directed evolution

- Mutagenic PCR
- Gene library
- Gene amplification (heredity)
- Mutagenesis (variation)
- Screening (fitness differences)
- Activity assay
- E. coli expressing gene library
- Isolation of desired variants
- Gene isolation
- Fitness
- Sequence space
- Point mutations
- Insertions and deletions
- Shuffling
Mutagenesis

Insertions

Substitutions

Deletions

Assembly (Substitution Mutagenesis Shown)

3’ Chew Back with GA SDM Assembly Mix A (2X)

Extension and Repair with GA SDM Assembly Mix B (2X)

Assembled fragment containing substitution

Transform, Plate, and Screen
Evolution of Green Fluorescent Protein
Green Fluorescent Protein

Maturation of the Enhanced Green Fluorescent Protein Chromophore

Extended Polypeptide (a) 

Imidazolin-5-one Ring Formation (c)

Cyclization, Dehydration and Oxidation (d)

Torsional Rearrangement (b)

Figure 2

Figure 3

N. Shaner

D. Goodsell
Green Fluorescent Protein

GFP (−30) M3H3H3H3GGSKGEEELGDGVPI LVELDVGNVHGFSVREGGEGDATGE
GFP (−25) M3H3H3H3GGSKGEEELGDGVPI LVELDVGNVHGFSVREGGEGDATGE
stGFP M3H3H3H3GGSKGEEELGDGVPI LVELDVGNVHGFSVREGGEGDATNG
GFP (+36) M3H3H3H3GGSKGEEELGDGVPI LVELDVGNVHGFSVREGGEGDATGE
GFP (+48) M3H3H3H3GGSKGEEELGDGVPI LVELDVGNVHGFSVREGGEGDATGE

GFP (−30) E3LKEICTGELPVAPALTLLTTTGYVQFSRYPDHMKHDFFKSAMPE
GFP (−25) E3LKEICTGELPVAPALTLLTTTGYVQFSRYPDHMKHDFFKSAMPE
stGFP K3LKEICTGELPVAPALTLLTTTGYVQFSRYPDHMKHDFFKSAMPE
GFP (+36) K3LKEICTGELPVAPALTLLTTTGYVQFSRYPDHMKHDFFKSAMPE
GFP (+48) K3LKEICTGELPVAPALTLLTTTGYVQFSRYPDHMKHDFFKSAMPE

GFP (−30) G3YVEERTISFKDDGTKTRAEVKFEGETTLVNRILKEDGNIHGN
GFP (−25) G3YVEERTISFKDDGTKTRAEVKFEGETTLVNRILKEDGNIHGN
stGFP G3YVEERTISFKDDGTKTRAEVKFEGETTLVNRILKEDGNIHGN
GFP (+36) G3YVEERTISFKDDGTKTRAEVKFEGETTLVNRILKEDGNIHGN
GFP (+48) G3YVEERTISFKDDGTKTRAEVKFEGETTLVNRILKEDGNIHGN

GFP (−30) L3EYNFSVHYITADKENGIAKFEIRHHEEDGSVQLDHYQYNTPIGD
GFP (−25) L3EYNFSVHYITADKENGIAKFEIRHHEEDGSVQLDHYQYNTPIGD
stGFP L3EYNFSVHYITADKENGIAKFEIRHHEEDGSVQLDHYQYNTPIGD
GFP (+36) L3EYNFSVHYITADKENGIAKFEIRHHEEDGSVQLDHYQYNTPIGD
GFP (+48) L3EYNFSVHYITADKENGIAKFEIRHHEEDGSVQLDHYQYNTPIGD

GFP (−30) G3PVLLDPHYLSTESALSKDPEFRHMLLLEFFVTAAGIDJGMDREDYK
GFP (−25) G3PVLLDPHYLSTESALSKDPEFRHMLLLEFFVTAAGIDJGMDREDYK
stGFP G3PVLLDPHYLSTESALSKDPEFRHMLLLEFFVTAAGIDJGMDREDYK
GFP (+36) G3PVLLDPHYLSTESALSKDPEFRHMLLLEFFVTAAGIDJGMDREDYK
GFP (+48) G3PVLLDPHYLSTESALSKDPEFRHMLLLEFFVTAAGIDJGMDREDYK

stGFP (−7) GFP (+36) GFP (−30)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP (+36)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>stGFP</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>GFP (−30)</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>DNA</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>tRNA</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>NaCl</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Example – protein capsids

- Ferritin: 12-14 nm
- Heat shock Protein cage: 12-14 nm
- Adenoassociated virus-2: 25 nm
- Brome mosaic virus: 26 nm
- MS2 phage: 26 nm
- Lumazine synthase complex: 16 nm
- Tobacco mosaic virus 20S intermediate: (18 nm x 4 nm)
- Cowpea chlorotic mosaic virus: 26 nm
- Papillomavirus L1 capsid: 27 nm
- Cowpea mosaic virus: 27 nm
- Murine polyomavirus: 48.6 nm
- Turnip yellow mosaic virus: 28 nm
- Cucumber mosaic virus: 28 nm
- Human hepatitis B virus: 31 nm
Evolvability of proteins from thermophiles

Aquifex aeolicus (Aa) – thermophilic bacteria

Protein capsids - lumazine synthase

wt AaLS:

MEIYEGKLTA EGLRFGIVAS RFNHALVDRL VEGAIDCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAIERAGT KHGNKGWEAA LSAIEMANLF KSLRLEHHHH HH***

162 Aminoacids
Engineering of lumazine synthase

AaLS-pos co-assembles with RNA in vivo and protects its cargo from nuclease digestion. The capsid retains the same 16 nm diameter as its parent AaLS-wt and encapsulates RNAs of up to 350 nt in length.

Engineering of lumazine synthase

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


**Error prone PCR library construction**

Taq DNA polymerase - Rates of error in PCR can be increased in the following ways:

* Increase concentration of magnesium chloride, which stabilizes non complementary base pairing.
* Add MnCl₂ to reduce base pair specificity.
* Increased and unbalanced addition of dNTPs.
* Addition of base analogs dITP, 8 oxo-dGTP,dPTP.
* Increase concentration of Taq polymerase, extension time, cycle time.
* Use less accurate Taq polymerase.
**DNA Shuffling**

(a) Two DNA double helices sharing high sequence similarity approach each other.

(b) One strand from each helix swaps with a strand from the other helix.

(c) All strands go through one round of normal semi-conservative DNA replication, resulting in the regeneration of the two original double helices, together with two new hybrid DNA double helices.
Engineering of lumazine synthase

Engineering of lumazine synthase

Engineering of lumazine synthase

PDB IDs: 1HQK (AaLS-wt), 5MQ3 (AaLS-neg), and 5MQ7 (AaLS-13)

Engineering of lumazine synthase

Engineering of lumazine synthase

(a) GFP-R10 + AaLS-neg → in vivo assembly

(b) HIVp-R10 + AaLS-13 → in vivo assembly

(c) in vitro assembly → internalization


Engineering of lumazine synthase

Engineering of lumazine synthase

**Lumazine synthase-based nanoreactors**

Rapid and quantitative encapsulation of GFP(+36)-fusion enzymes by empty AaLS-13 cages in vitro (RA, retroaldolase)

A synthetic mimic of the carboxysome: ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and carbonic anhydrase (CA), were co-packaged in the lumen to study the reaction cascade.

Substrate sorting for encapsulated tobacco etch virus protease (TEVp) based on electrostatic interactions with the lumenal surface of the AaLS-13 cage

AaLS-13 as templates for monodisperse polymer synthesis. An engineered variant of ascorbate peroxidase (APEX2) enables confined polymerization of compound (7) in the lumen of the cage.
A statistically significant kinetic effect of coencapsulated CA on RuBisCO activity was not observed, however the capsid shell protected the enzyme from proteolytic damage.

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.

Progress of the polymerization reaction monitored by electron microscopy

Design of circularly permuted AaLS (cpAaLS). The left structure shows a pentameric capsomer of AaLS-wt, where one monomer unit is colored: residues 1–119, orange; residues 120–156, blue. Upon circular permutation, the native termini are connected with a flexible peptide linker, and new sequence termini are generated between residues 119 and 120 in a loop region facing the interior of the assembly.

Y. Azuma, M. Herger, D. Hilvert

Design of circularly permuted AaLS (cpAaLS). A pentameric capsomer of AaLS-wt, where one monomer unit is colored: residues 1–119, orange; residues 120–156, blue. Upon circular permutation, the native termini are connected with a flexible peptide linker, and new sequence termini are generated between residues 119 and 120 in a loop region facing the interior of the assembly.

Y. Azuma, M. Herger, D. Hilvert

Linker length controls the assembly state of circularly permuted AaLS.

(a) Scheme illustrating the hypothetical relationship between linker length, capsomer shape, and assembly of higher-order structures.

(b–d) TEM images of the assemblies produced by different cpAaLS(LxHy) variants. The rod-shaped cpAaLS(L8H4) structures were obtained by self-assembly of isolated capsid fragments. Scale bar = 100 nm

Y. Azuma, M. Herger, D. Hilvert
Engineering of lumazine synthase

Diversification of Protein Cage Structure Using Circularly Permuted Subunits

Incorporation of cpAaLS into other AaLS assemblies.

(a) Scheme illustrating the formation of patchwork assemblies in E. coli.

\[ P_{tet} \text{, tetracycline promoter; } \]

\[ tetO, \text{ tetracycline operon; } \]

\[ P_{T7}, T7 \text{ promoter; } \]

\[ lacO, \text{ lactose operon.} \]

(b) Size-exclusion chromatogram of cpAaLS(L8)-GFP coassembled with AaLS-wt (black), AaLS-neg (blue), and AaLS-13 (red). Continuous and dashed lines respectively indicate absorbance at 280 nm (A280) and fluorescence (F500) (ex, 470 nm; em, 500 nm) for each fraction.

(c) TEM images of cpAaLS(L8)-GFP coassembled with AaLS-wt, AaLS-neg, and AaLS-13. Scale bar = 100 nm.

Y. Azuma, M. Herger, D. Hilvert

**Diversification of Protein Cage Structure Using Circularly Permutated Subunits**

Encapsulation of a protein guest by fusion to cpAaLS and patchwork formation with other AaLS variants.

Capture of small RNAs in AaLS patchwork assemblies using cpAaLS variants with an appended oligoarginine tag (cpAaLS-RX).

**Diversification of Protein Cage Structure Using Circularly Permutated Subunits**

Several different NBDs were tested to study the effect of charge and size on RNA encapsulation.

![Diagram](image)

Co-expression of cpAaLS-NBD and AaLS-wt proteins provides patchwork capsids that encapsidate RNA upon assembly in vivo.

Patchwork capsids with 1:1 ratios of the two proteins have the same 16 nm external diameter. However, increasing the steric bulk of the NBDs decreases the capsid cavity volume and favors the packaging of shorter RNAs.

An artificial virus

Design of λcpAaLS for specific RNA recognition

(a) Outside

(b) λN+ peptide

Inside

C’

N’

His6

The λN+ peptide, which recognizes the BoxB RNA sequence, was appended to the N terminus and a hexahistidine tag inserted in the exterior loop for purification.


Scheme for the directed evolution of nucleocapsids encapsulating their own genome
Laboratory evolution of virus-like nucleocapsids

Directed evolution strategy.

In the first evolutionary round, BoxBr tags that bind to λN+ peptides were introduced in the 5′- and 3′-UTRs of HIV protease mRNA. Capsid variants were initially selected from large gene libraries for their ability to inhibit production of the toxic protease by sequestration of its BoxBr-tagged mRNA (Upper).

For further optimization (Lower), the BoxBr tags were transferred to the UTRs of mRNA encoding the selected λcpAaLS variants, and the encoded cage proteins were produced in *E. coli*.

After capsid purification and nuclease treatment, encapsulated mRNAs were extracted and converted to cDNA by RT-PCR. The best variant, λcpAaLS-β16, was further diversified by epPCR, and the resulting library was subjected to the procedures shown.

Laboratory evolution of virus-like nucleocapsids

Characterization of evolved λcpAaLS variants.

(A) Backbone traces of capsid subunits (Upper) and negatively stained EM images of assembled cages (Lower) are shown for λcpAaLS, λcpAaLS-β16 (I78F, V80D, E95V, I111V, G114D, and V115I); λcpAaLS-α2 (E2G, A25T, P54T, L81P, and E145G), and λcpAaLS-α9 (M1I, A25T, G64D, I92N, and F150Y).

(B) The amounts of total RNA packaged in λcpAaLS capsids expressed from mRNAs possessing two BoxBr tags (white bars) were determined from the 260/280-nm absorbance ratio, whereas full-length capsid-encoding mRNA (black bars) was determined by quantitative RT-PCR. (D) denaturing PAGE. “mRNA” indicates in vitro-transcribed full-length mRNA (863 nt).

Mapping the electrostatic potential onto the solvent-accessible surface of an AfFtn hexamer highlights the elevated negative charge of the lumenal (left) compared to the exterior (right) surface of the cage.

Computational design of self-assembling protein nanomaterials

Protein building blocks are docked together symmetrically to identify complementary packing arrangements, and low-energy protein-protein interfaces are then designed between the building blocks in order to drive self-assembly.

The building block: trimeric *E.coli* acyl-carrier protein synthase (AcpS)

Computational design of self-assembling protein nanomaterials

(A) First, a target symmetric architecture is chosen. Octahedral point group symmetry is used in this example; the threefold rotational axes are marked here by triangles and shown as black lines throughout. The dashed cube is shown to orient the viewer. A symmetric oligomer which shares an element of symmetry with the target architecture, here a C3 symmetric trimer (green), is selected as a building block.

(B) Multiple copies of the building block are symmetrically arranged in the target architecture by aligning their shared symmetry axes. The preexisting organization of the oligomeric building block fixes several (in this case four) rigid-body degrees of freedom (DOFs). The two remaining DOFs, radial displacement (r) and axial rotation (w), are indicated.

(C) Symmetrical docking is performed by systematically varying the two DOFs (moves are applied symmetrically to all subunits) and computing the suitability of each configuration for interface design (red: more suitable; blue: less suitable). Points corresponding to the docked configurations in (B), in which the building blocks are not in contact, and (D), a highly complementary interface, are indicated.

(E) Closer view of the interface in (D). The interface lies on an octahedral two-fold symmetry axis shown as a gray line. In all steps before interface design, only backbone (shown in cartoon) and carbon beta (shown in sticks) atoms are considered. (F) Sequence design calculations are used to create low-energy protein-protein interfaces that drive self-assembly of the desired material. Designed hydrogen bonds across the interface are indicated by dashed lines.

Computational design of self-assembling protein nanomaterials

Experimental characterization of O3-33, T3-08, and T3-10.

(A) Native PAGE of fluorescently labeled (from left) 3n79-wt, O3-33, 3fttwt, and T3-08 in lysates. Bands corresponding to the designed octahedral (O3-33) and tetrahedral (T3-08) assemblies are indicated with asterisks.

SEC chromatograms of nickel-purified (B) O3-33, (C) 3n79-wt, (D) O3-33(Ala167Arg), (E) T3-08, (F) T3-10, (G) 3ftt-wt, and (H) T3-08(Ala52Gln) collectively demonstrate that the assembly of the designed proteins is a result of the designed interfaces

Structural characterization of O3-33.

(A) negative-stain electron micrographs of O3-33.

(B) The O3-33 design model, depicted in ribbon format. Each trimeric building block is shown in a different color.

(C) The density map from a 20 Å resolution cryo-EM reconstruction of O3-33.

(D) The crystal structure of O3-33 (R32 crystal form).

(E) The designed interface in O3-33, highlighting the close agreement between the crystal structure (green and magenta) and the design model (white).

Science 2012, 336, 1171
Computational design of self-assembling protein nanomaterials

(A) A representative negative stain electron micrograph of T3-10.
(B) Backbone representation T3-08–T3-10 design model.
(C) The T3-10 crystal structure.
(D) The designed interface in T3-10, revealing the close agreement of the crystal structure (green and magenta) to the design model (white). Alanine 52 is labeled; when mutated to glutamine in T3-08, it disrupts assembly of the designed material.

Structural characterization of T3-10.

Science 2012, 336, 1171
Rational Engineering of a Designed Protein Cage for siRNA Delivery

Surface models of the OP cage – one trimer shown as a colored ribbon.

HeLa cells treated with Atto488-DNA or OP:Atto488-DNA. Blue, Nuclei (Hoechst stained); Green, Atto488.

Superimposed monomers of O3-33 (gray) and OP (orange)

Rational Engineering of a Designed Protein Cage for siRNA Delivery

(a) Surface structure of the OP cage assembly, with blue highlighting the positively charged lumen.

(b) OP monomer showing the six arginine mutations introduced with respect to the original O3-33 scaffold.

(c) OP cages carrying GFP-targeting siRNA (OP:siRNA) knock down protein expression to the same degree as the commercial reagent lipofectamine (Lipo:siRNA). PBS buffer, naked siRNA, and OP carrying a scrambled sequence siRNA (OP:scramble) have no effect on GFP signal. OP cages lacking His6 tags (No His6) are less active.

(d) While stable in the absence of other oligonucleotides, high concentrations of tRNA can displace the siRNA cargo from OP cages, freeing them to induce RNA interference.

PROTEINS

Enzyme engineering

(a few) random mutations → Parent gene (= parent protein) → Evolved gene (= evolved protein)

select/screen → repeat

NO → YES → repeat

**Enzyme engineering**

**A:** A starting point with no activity for the intended reaction is useless since no sequence variations (red arrows) create the new reactivity.

**B:** A promiscuous enzyme with at least low activity for the intended reaction is a suitable staring point. Some combinations of random mutations may improve the new reactivity (black arrow). The first variant (1) serves as a starting state for sequential rounds of variation and screening → (2) → (3) → (4) for improved variants. Only a small number of cycles and are typically needed to boost up the new reactivity

*F. Arnold Nobel lecture 2018; Angew. Chem. Int. Ed. 2019, 58, 14420–14426*
The cytochrome P450 family, whose members were presumably created by gene duplication and natural selection of promiscuous functions, comprises enzymes that use reactive oxygen intermediates to catalyze a wide range of reactions. We reasoned that we could expand the scope of P450 chemistry by using synthetic carbene and nitrene precursors to drive formation of new reactive intermediates. Directed evolution would be used to mold the enzyme, controlling and enhancing new-to-nature activities.

Expanding the scope of P450 chemistry.

An evolved biocatalyst for cyclopropanation.

The cytochrome P411 variant of cytochrome P450 with the protein backbone shown as ribbon representation and side-chains as sticks. Side-chains that were mutated in engineered variants are shown in red.

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


Directed evolution – bringing new chemistry to life

A) Chiral Si–C bond formation catalyzed by a laboratory-evolved variant of *Rhodothermus marinus* cytochrome c. The three amino acid residues that were mutated to increase this abiological activity include the methionine axial ligand (M100). B) The enzyme catalyzes formation of different organosilane products with high enantiomeric excess from silane and diazo substrates.

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 B–H insertion. The bioconversion can be conducted readily on gram scale, and the enantio-preference of borylation was switched to give either enantiomer of the organoborane products.

A cytochrome P450 anti-Markovnikov oxygenase. A) Competing reaction pathways for P450-catalyzed oxo transfer to alkenes. The concerted epoxidation pathway is favored over the stepwise anti-Markovnikov oxidation consisting of oxo transfer followed by an (enantioselective) 1,2-hydride migration. B) Ten rounds of directed evolution accumulated 12 amino acid mutations, many of which are distant from the active site. C) aMOx can be combined with established (bio)catalysts for various challenging anti-Markovnikov alkene functionalization reactions.

Engineering enzymes for noncanonical amino acid synthesis

Expanding the substrate scope of TPL (tyrosine phenyl lyase) from *Citrobacter freundii* (*CfTPL*) through enzyme engineering. The wild-type enzyme shows activity only with phenol, catechol (3-hydroxyphenol), and various fluorophenols to synthesize the corresponding tyrosine (Tyr) analogue.
Protein engineering – de novo enzyme design

<table>
<thead>
<tr>
<th>Semisynthetic metalloenzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) ( \text{H}<em>2\text{N-} \overset{\text{OPiv}}{\text{O}}\text{OPiv} \text{Rh(III)} \rightarrow \text{O}</em>{\text{N}}\text{NH} \text{O} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Computationally designed enzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b) ( \text{O}_2\text{N} \text{O} \text{O} \text{O} \text{O} \text{His-Cys or His} \rightarrow \text{O}_2\text{N} \text{O} \text{O} \text{O} \text{R} \text{O}_2\text{OH} )</td>
</tr>
<tr>
<td>(c) ( \text{O}_2\text{N} \text{O} \text{O} \text{O} \text{Lys} \rightarrow \text{O}_2\text{N} \text{O} \text{O} \text{O} )</td>
</tr>
<tr>
<td>(d) ( \text{O}_2\text{N} \text{N} \text{O} \text{O} \text{His, Acp, Asp, or Glu} \rightarrow \text{O}_2\text{N} \text{N} \text{O} \text{O} \text{O} )</td>
</tr>
<tr>
<td>(e) ( \text{NH} \text{N} \text{O} \text{O} \text{O} \text{Ar} \rightarrow \text{NH} \text{N} \text{O} \text{O} \text{O} \text{Ar} \text{N} \text{N} )</td>
</tr>
<tr>
<td>(f) ( \text{O}_2\text{O} \text{O} \text{P} \text{O}_2\text{OH} \text{Zn}^{2+} \rightarrow \text{O}_2\text{O} \text{O} \text{P} \text{O}_2\text{OH} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Challenges for computational design</th>
</tr>
</thead>
<tbody>
<tr>
<td>(g) ( \text{O} \text{O} \text{O} \text{O} \text{O} \rightarrow \text{O} \text{O} \text{O} \text{O} \text{O} )</td>
</tr>
<tr>
<td>(h) ( \text{O} \text{O} \text{O} \text{O} \text{O} \rightarrow \text{O} \text{O} \text{O} \text{O} \text{O} )</td>
</tr>
</tbody>
</table>

Protein engineering – de novo enzyme design

(1) define catalytic groups
(2) optimize geometry

(1) characterize
(2) optimize

tailor-made enzyme

dock into scaffold

improve packing

EΨ=HΨ

Protein engineering – de novo enzyme design

De novo retro-aldolases that use four different catalytic motifs to catalyze the breaking of a carbon-carbon bond in a nonnatural substrate were designed. Of the 72 designs that were experimentally characterized, 32, spanning a range of protein folds, had detectable retro-aldolase activity. Designs that used an explicit water molecule to mediate proton shuffling were significantly more successful, with rate accelerations of up to four orders of magnitude and multiple turnovers, than those involving charged side-chain networks. The atomic accuracy of the design process was confirmed by the x-ray crystal structure of active designs embedded in two protein scaffolds, both of which were nearly superimposable on the design model.

- Carlos F. Barbas III
- Donald Hilvert
- Kendall N. Houk
- Barry L. Stoddard
- David Baker

Retro-aldolase - de novo enzyme design

Examples of design models for active designs highlighting groups important for catalysis. The nucleophilic imine-forming lysine is in orange, the TS model is in yellow, the hydrogen-bonding groups are in light green, and the catalytic water is shown explicitly. The designed hydrophobic binding site for the aromatic portion of the TS model is indicated by the gray mesh.

**Kemp eliminase - de novo enzyme design**

The Kemp elimination proceeds by means of a single transition state, which can be stabilized by a base deprotonating the carbon and the dispersion of the resulting negative charge; a hydrogen bond donor can also be used to stabilize the partial negative charge on the phenolic oxygen.

Examples of active site motifs highlighting the two choices for the catalytic base (a carboxylate (left) or a His–Asp dyad (right)) used for deprotonation, and a p-stacking aromatic residue for transition state stabilization. For each catalytic base, all combinations of hydrogen bond donor groups (Lys, Arg, Ser, Tyr, His, water or none) and p-stacking interactions (Phe, Tyr, Trp) were input as active site motifs into RosettaMatch.

**Kemp eliminase - de novo enzyme design**

**Computational design models of the two most active catalysts.**

**a,** KE59 uses indole-3-glycerolphosphate synthase from Sulfolobus solfataricus as a scaffold. The transition state model is almost completely buried, with loops covering the active site. The mostly hydrophobic residues in the active site pocket pack the transition state model tightly, providing high shape complementarity. The polar residue Ser211 interacts with the nitro group of the transition state to promote binding. The key catalytic residues (Glu231 and Trp110) are depicted in cyan.

**b,** The deoxyribose-phosphate aldolase from E. coli is the scaffold for KE70. The shorter loops leave the active-site pocket freely accessible for the substrate. The transition state is surrounded by hydrophobic residues that provide high shape complementarity. His16 and Asp44 (in cyan) constitute the catalytic dyad whereas Tyr47 (in cyan) provides π-stacking interactions.  

Kemp eliminase - de novo enzyme design

The crystal structure (cyan) was solved in the unbound state and shows only modest rearrangement of active site side chains compared to the designed structure (grey) modelled in the presence of the transition state (yellow, transparent). KE07 contains 13 mutations compared to the starting template scaffold (PDB code 1thf).

**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.

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

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.

---

**Theoenzyme**

**Design model**

**Catalytic activity**

Most active variant: \( k_{cat}/K_M \) of 400 M\(^{-1}\) 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.

The tyrosyl ester 1 served as the target substrate for computational design; the fluorogenic coumarin ester 2 and the chromogenic p-nitrophenyl ester 3 were used for screening purposes.

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

Snapshots of the computational design process.

Representation of the calculated theozyme of the ester substrate framed by the catalytic dyad (Cys-His) and the backbone NH-oxyanion contact.

To increase the number of matches, both the histidine sidechain and the substrate could rotate with respect to the cysteine (not shown). Note that in this case, the backbone NH contact is made by the cysteine itself.

Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis

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.

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.

Evolution of a metalloenzyme from short peptides

Crystal structure of MID1sc10
zinc ion - orange sphere, coordinating histidines - green sticks
linkage of two polypeptides – green sticks
beneficial mutations - magenta spheres, residues replaced to prevent competitive zinc
binding modes - cyan spheres).

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

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.

Evolution of a metalloenzyme from short peptides

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(II)His$_3$ complex (orange sphere and green sticks). Arg68 and Gln58 form mechanistically relevant hydrogen bonds to phosphonate 4 and the backside nitrogen of His61, respectively.
