Artificial Enzyme
-Computational Design of Catalytic Functions-

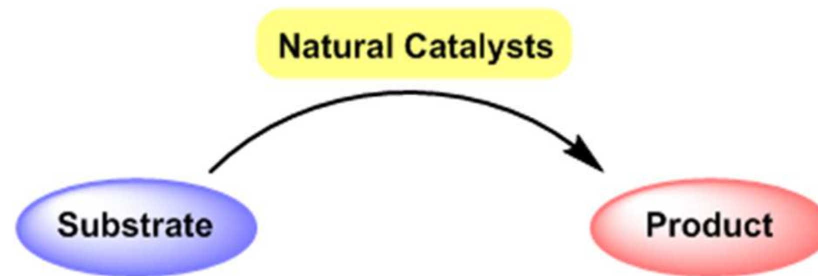
Literature Seminar

2014.2.3 (Mon.)

Yusuke Shimizu (B4)

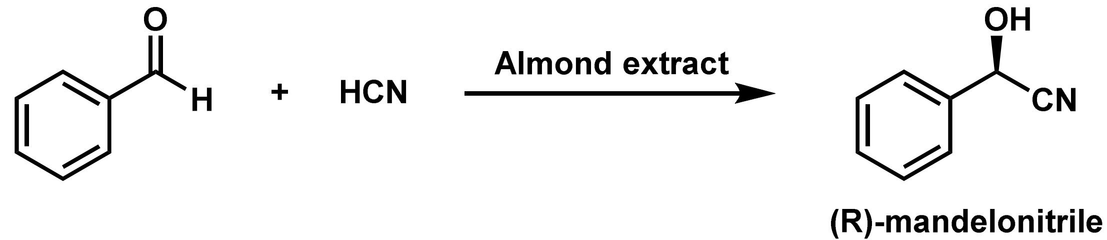
Biocatalysis

Biocatalysis = Application of enzymes and microbes to chemical transformations



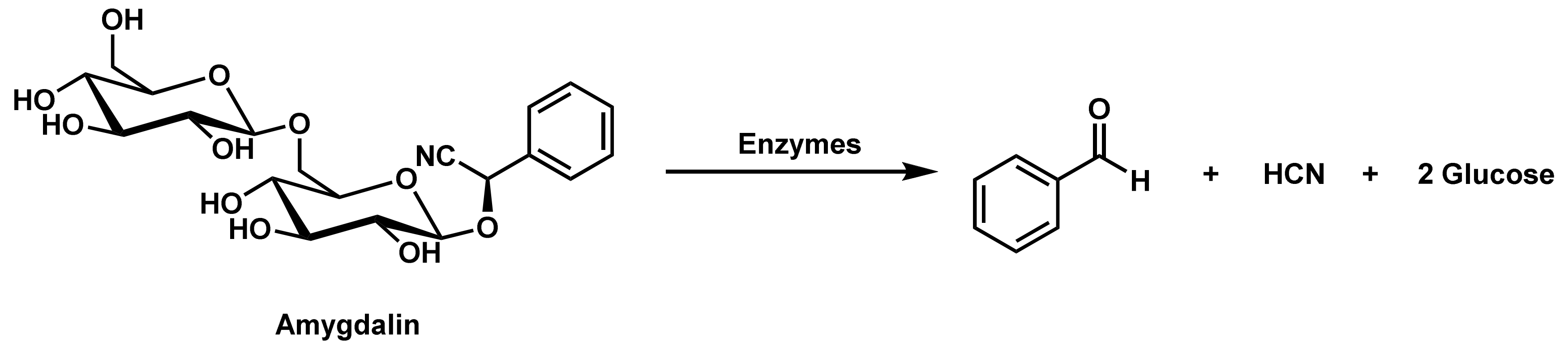
Humans have utilized biocatalysis in the fermentation processes for millennia.

Biocatalysis



Rosenthaler, L. *Biochem. Z.* **1908**, 14, 238

Actual “biocatalyst” is an enzyme related to cyanogenesis



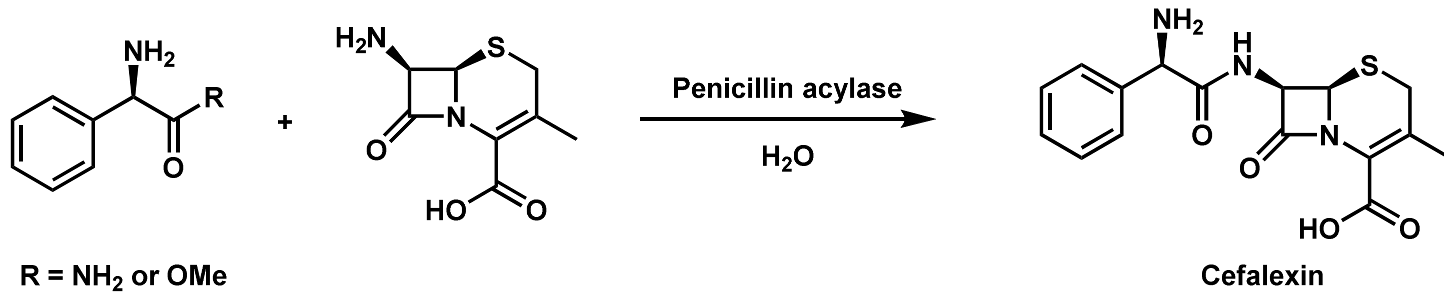
What is Enzyme?

- First isolation in 1833 by Payen. He discovered and isolated amylase.
- Kühne coined term *enzyme* means “in yeast” in 1876
- Sumner crystallized urease, and showed enzymes are proteins in 1926.
- The first crystal structure was obtained in 1965 via X-ray crystallography

Enzyme is a protein catalyst , which is...

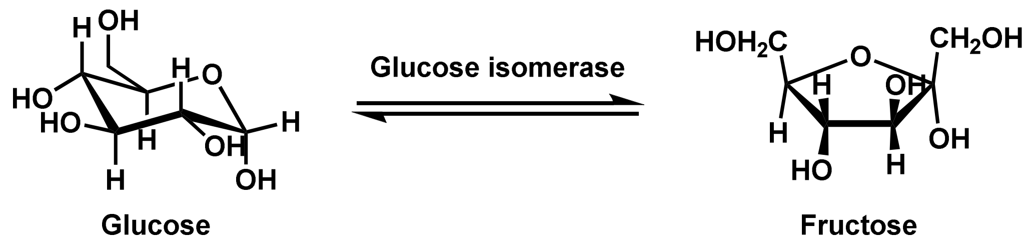
- Highly selective (chemo-, regio-, diastereo-, and enatnio-selective)
- Activated in mild condition (typically pH 5–8 and 20–40ºC)
- Environment friendly (completely degraded in the environment)
- Generally very efficient

Enzymes in Industry



Bruggink, A. *Chimia*, **1996**, 50, 431

Semi-synthetic antibiotics



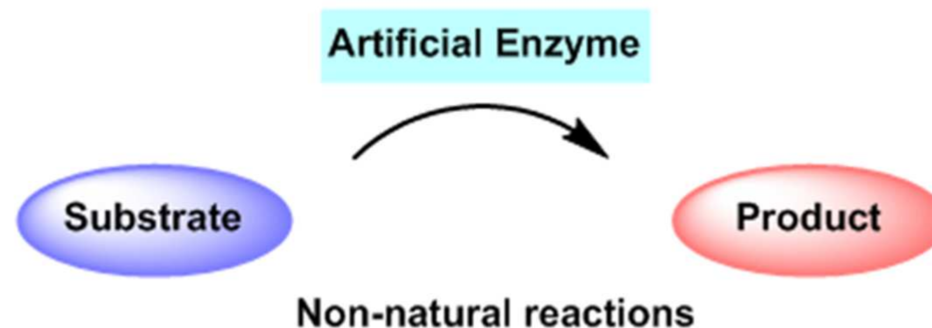
Jensen, V.J., Rough, S. *Methods Enzymol.* **1987**, 136, 356

Convert glucose to the sweeter-tasting fructose

Laundry-detergent contain enzymes

Artificial Enzyme

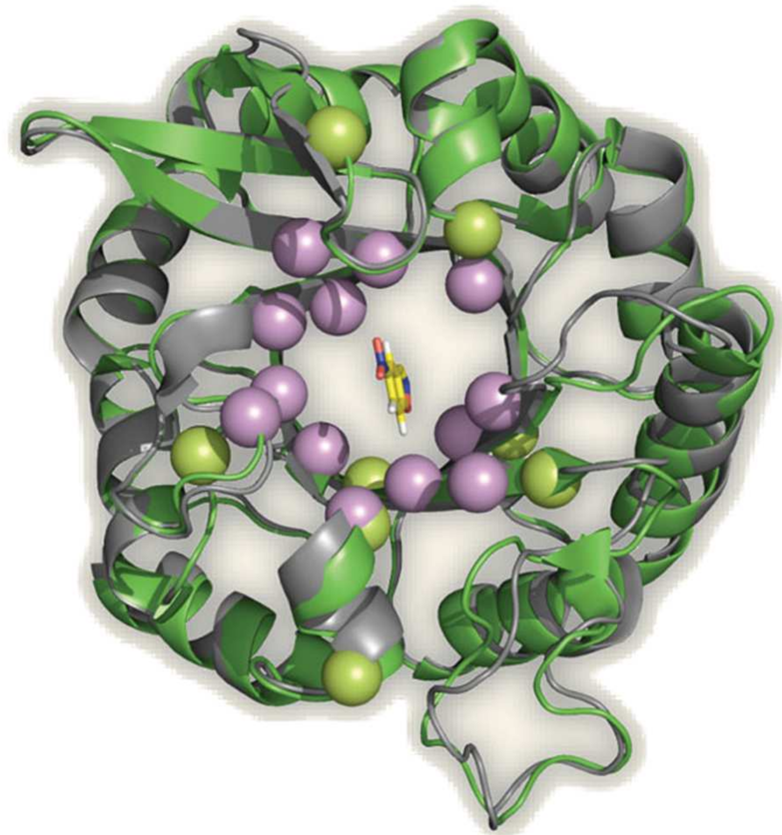
Protein engineering technologies has developed to design and synthesize molecules with the attributes of enzymes (selective, proficient, green, nontoxic, and biodegradable) for non-natural reactions



Today, I introduce development of approach to enzyme design and finally, *de novo* Computational Enzyme Design.

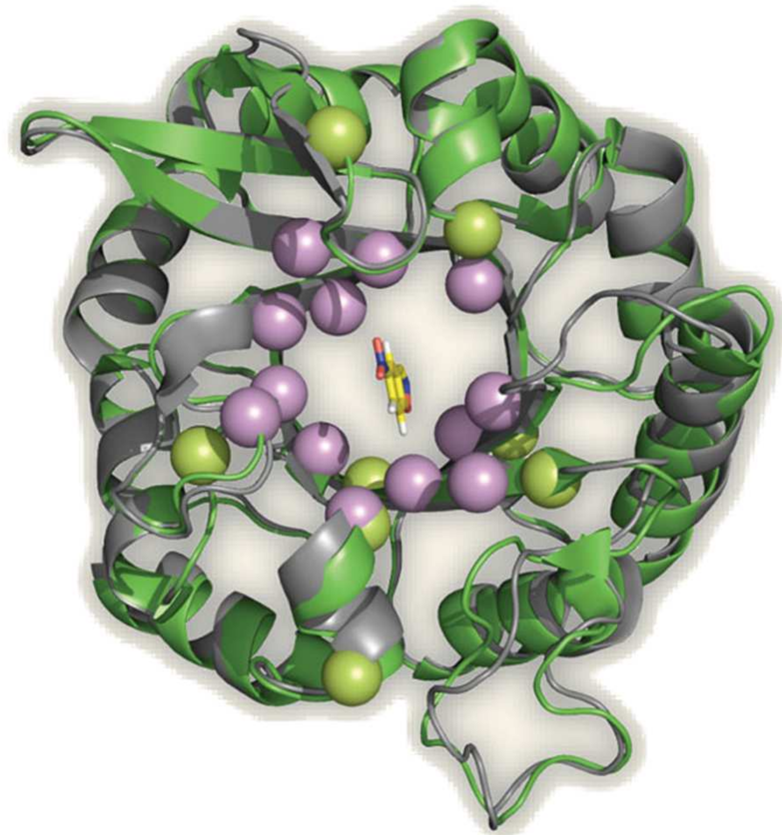
Contents

1. Introduction
2. Catalytic Antibodies
3. Directed Evolution
4. Computational Enzyme Design
5. Summary & Future Outlook



Contents

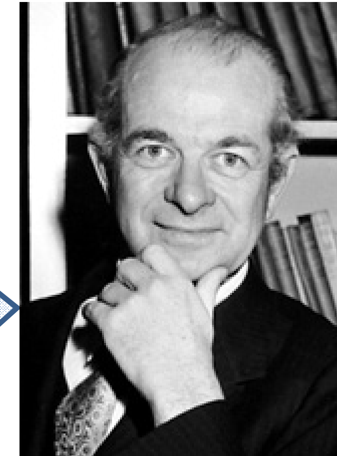
1. Introduction
2. Catalytic Antibodies
3. Directed Evolution
4. Computational Enzyme Design
5. Summary & Future Outlook



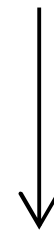
Pauling's hypothesis

I think that enzymes are molecules that are complementary in structure to the activated complexes of the reaction that they catalyze, that is, to the molecular configuration that is intermediate between reacting substances and the products of reaction for these catalyzed process

Pauling, L. *Nature*, **1948**, 161, 707



Linus Pauling

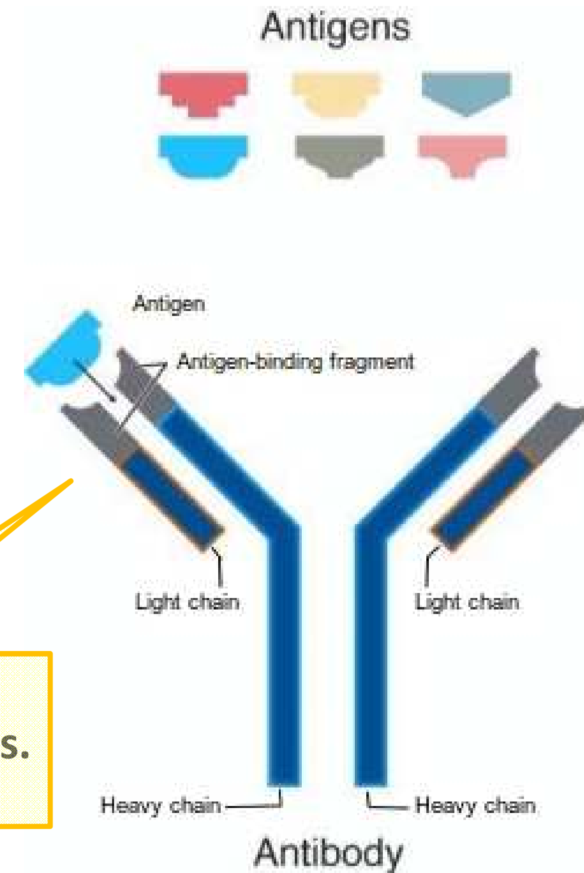


The transition state analog of a particular enzymatic reaction would be bound tightly to enzymes. (Concept of enzyme inhibitors)

Catalytic Antibodies

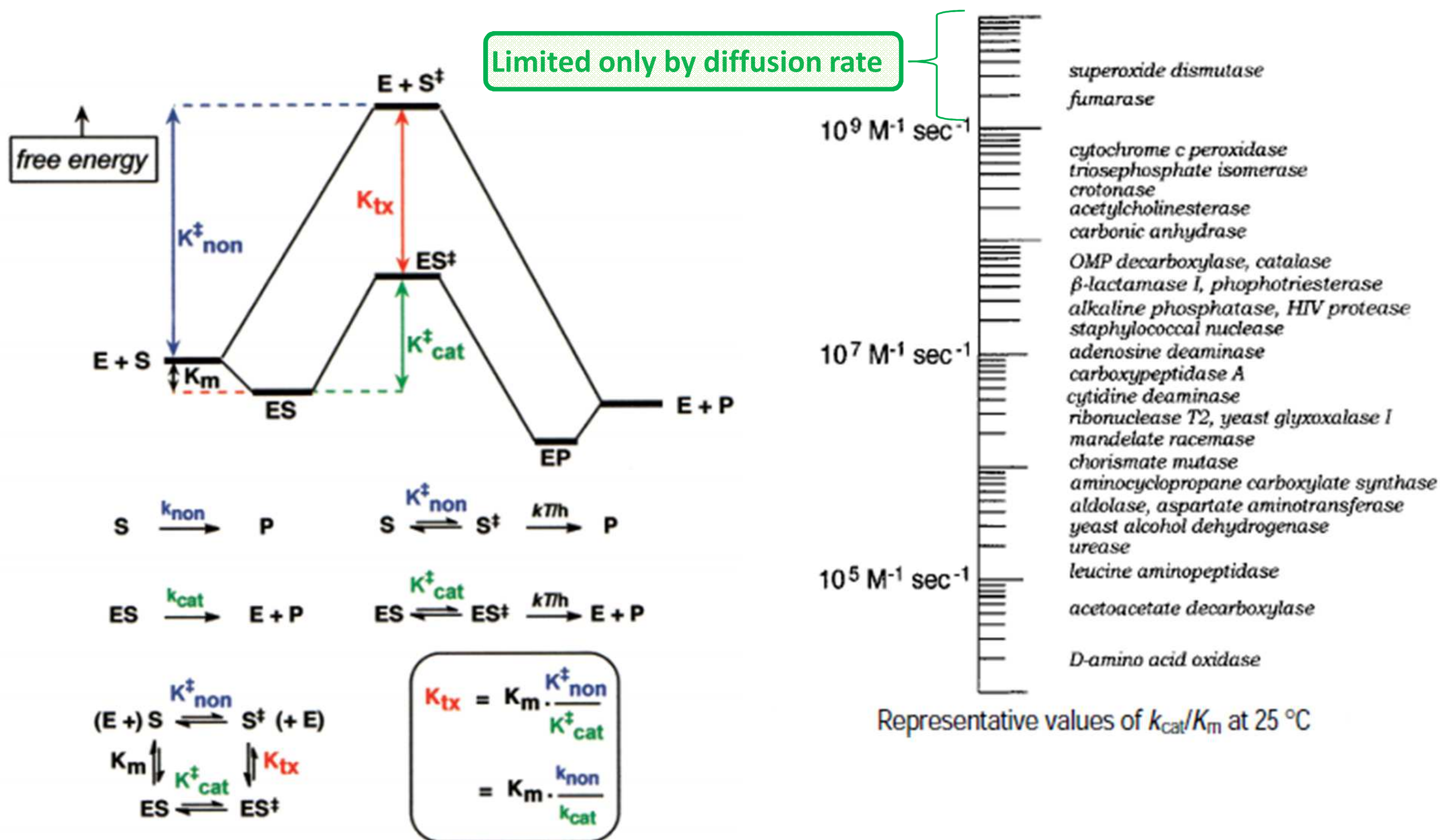
According to Pauling's hypothesis, conversely, a receptor designed to optimally bind a suitable analog of a transition state would achieve the catalytic function of an enzyme.

The immunological reservoir provides the variety of receptor sites with the required specificities for such a study. The combining sites of antibodies have been considered as useful templates for simulating the environment of an enzyme active site



Antibodies specifically bind to corresponding antigens.

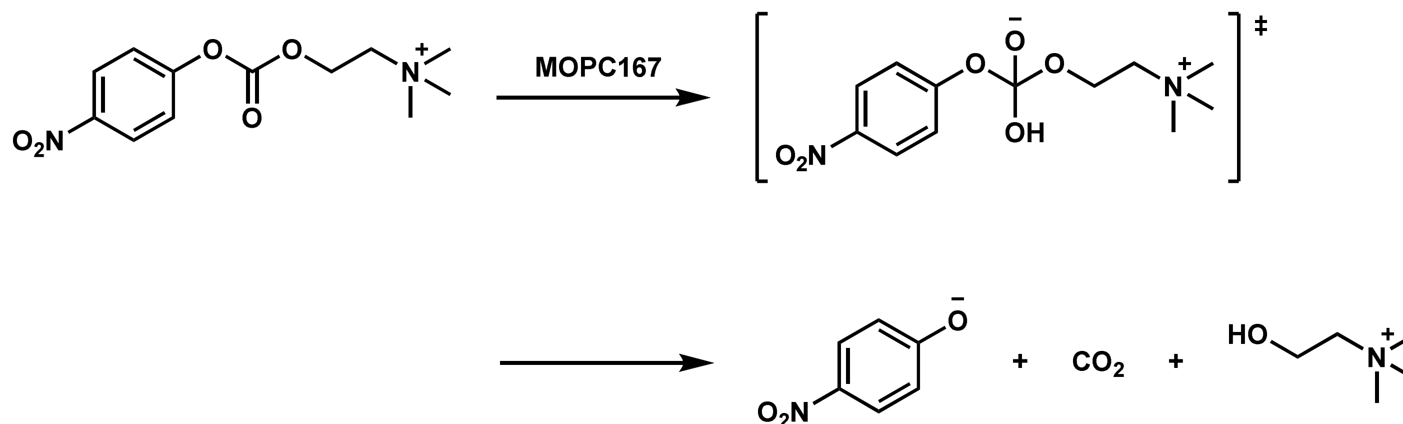
Kinetics

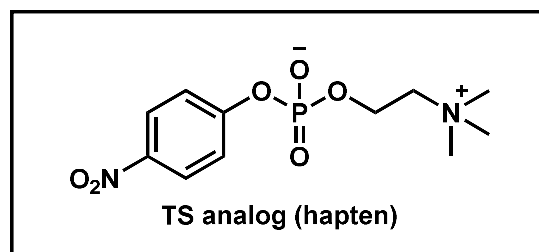


Wolfenden, R., Snider M. J. *Acc. Chem. Res.* **2001**, 34, 938

Pioneering work by Schulz

Pioneering work of catalytic antibody using IgA MOPC167



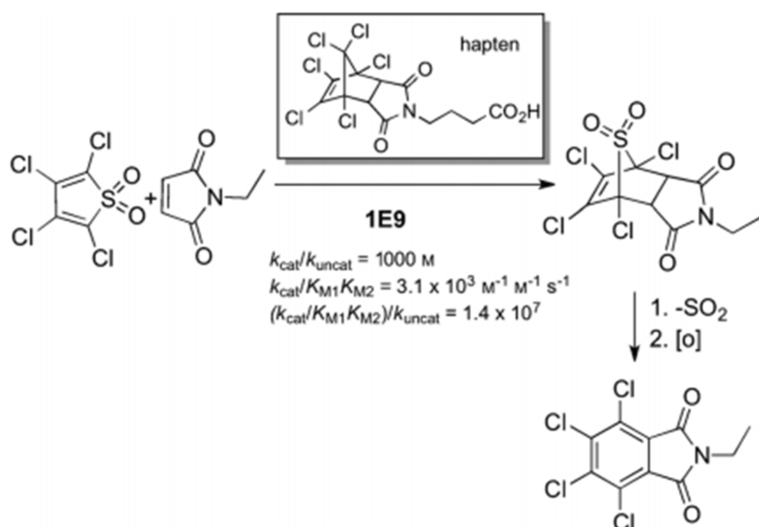


K_M = 208 µM
K_{cat} = 6.0*10^{-3} sec^{-1}
k_{cat}/K_M = 28.8 M^{-1}s^{-1}
k_{cat} / k_{uncat} = 770

Schulz, P.G., *et al*. *Science*, **1986**, 234, 1570

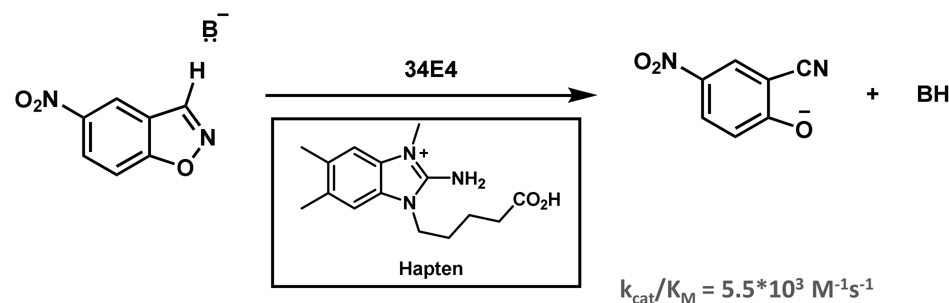
Achievement

A number of catalytic antibodies catalyze natural and non-natual reaction were generated.



Diels-Alder reaction

Hilvert, D., *et al. J. Am. Chem. Soc*. **1989**, 111, 9262



$k_{cat}/K_M = 5.5*10^3\ M^{-1}s^{-1}$

Kemp elimination

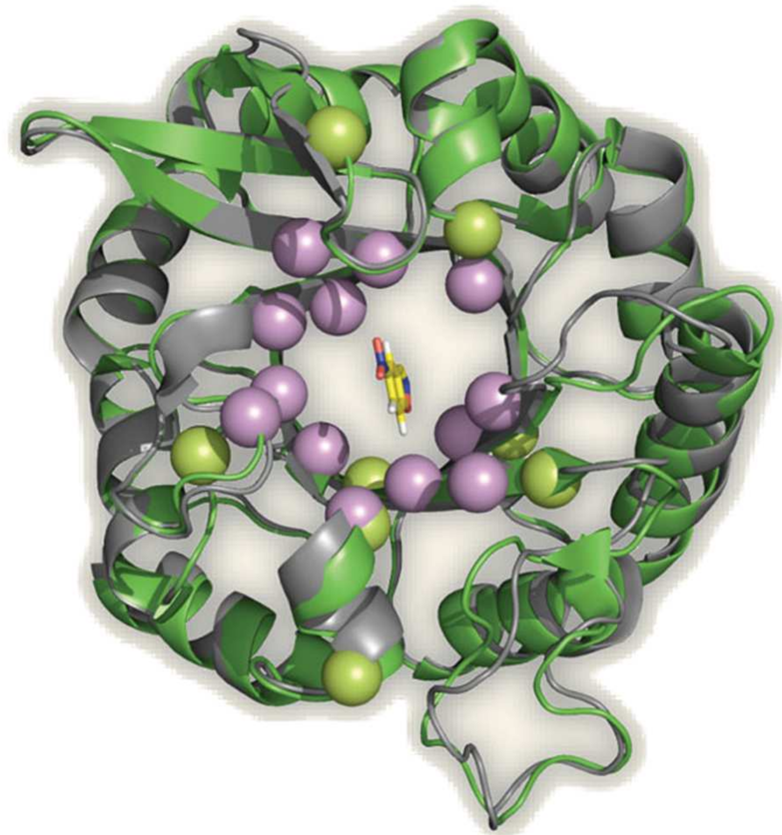
Hilvert, D., *et al*. *Nature*, **1995**, 373, 228

Limitations

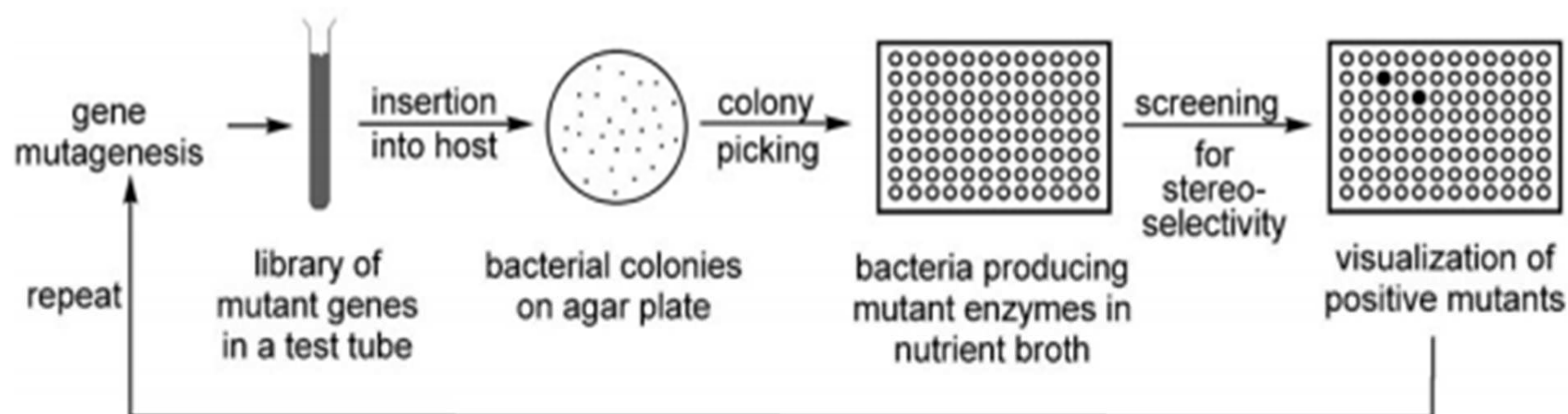
- Lower binding constants
- Lack of covalent binding and catalysis
- Smaller buried surface area
- Inadequacies of the immunoglobulin fold
- Product inhibition
- The comparatively low stability of the immunoglobulin fold
- High cost of producing antibody catalysts

Contents

1. Introduction
2. Catalytic Antibodies
3. Directed Evolution
4. Computational Enzyme Design
5. Summary & Future Outlook



Directed Evolution

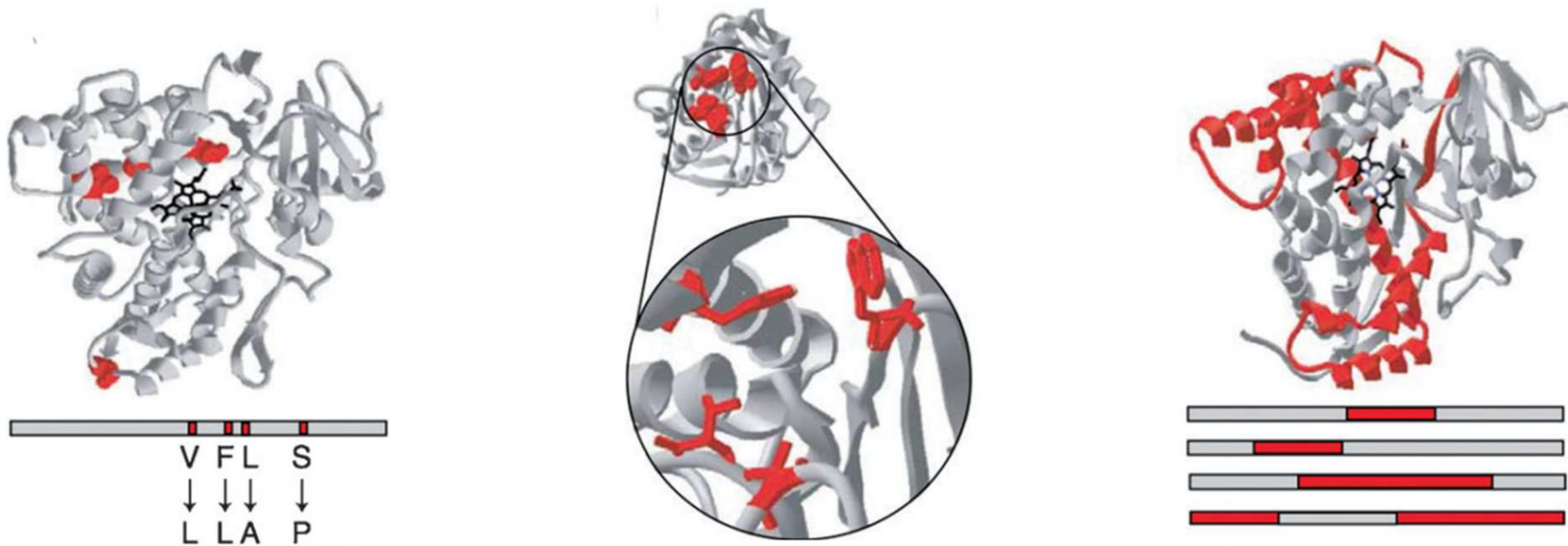


Directed evolution is a molecular biology methods to modify biocatalysts via *in vitro* version of "Darwinian evolution"

Directed evolution provide improved enzymatic activity, thermostability, tolerance to organic solvent, substrate specificity, enantioselectivity and so on.

Reetz, M. T. *Angew. Chem. Int. Ed*. **2011**, 50, 138

Mutagenesis



Random mutagenesis | Site saturation mutagenesis | Sequence recombination

Random mutagenesis : No structural or machanistic information is required

Site saturation mutagenesis : Require prior structural or biochemical knowledge

Dramatic functional alteration can be sought

Sequence recombination : Diversity can be further expanded

Challenges

- **Identification of variants that have desired improvements out of large set of sequences. (e.g. two mutations anywhere in 200 amino-acid protein have 7,183,900 possibilities)**
- **Set up of high-throughput assays which can detect slight improvements**

Application – Sitagliptin Manufacture

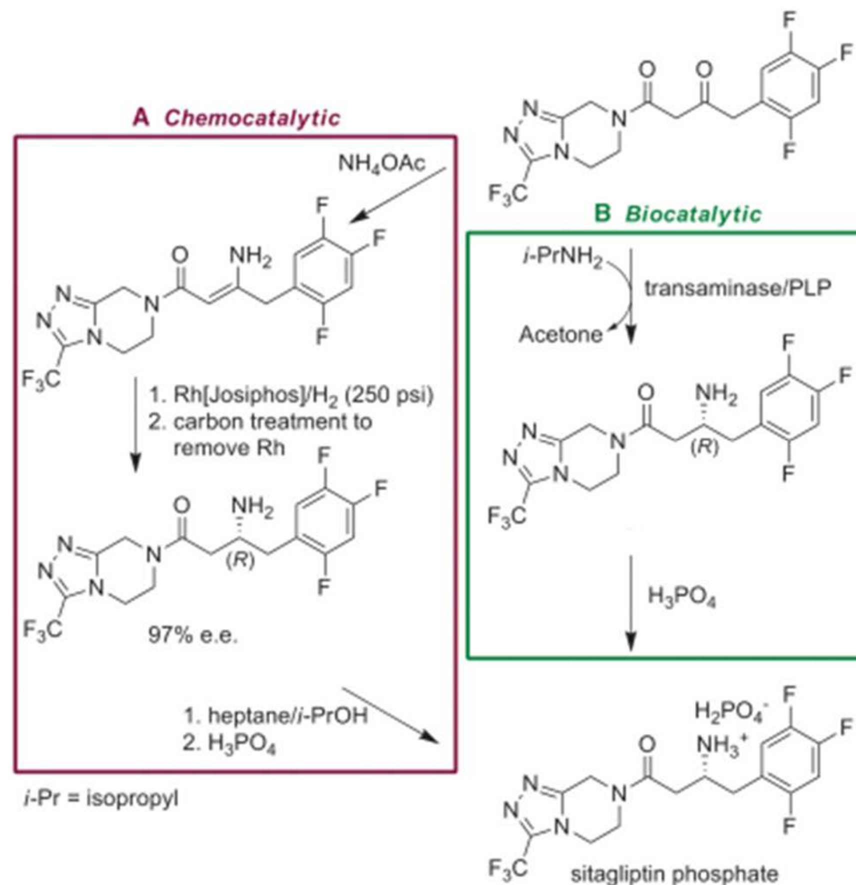
Route A

- Need for high pressure H2
- Use and removal of precious and toxic Rh
- Ligand screening and synthesis
- Insufficient stereoselectivity

Route B

- Limited substrate range
- Low turn over numbers
- Stability to chemical processes conditions

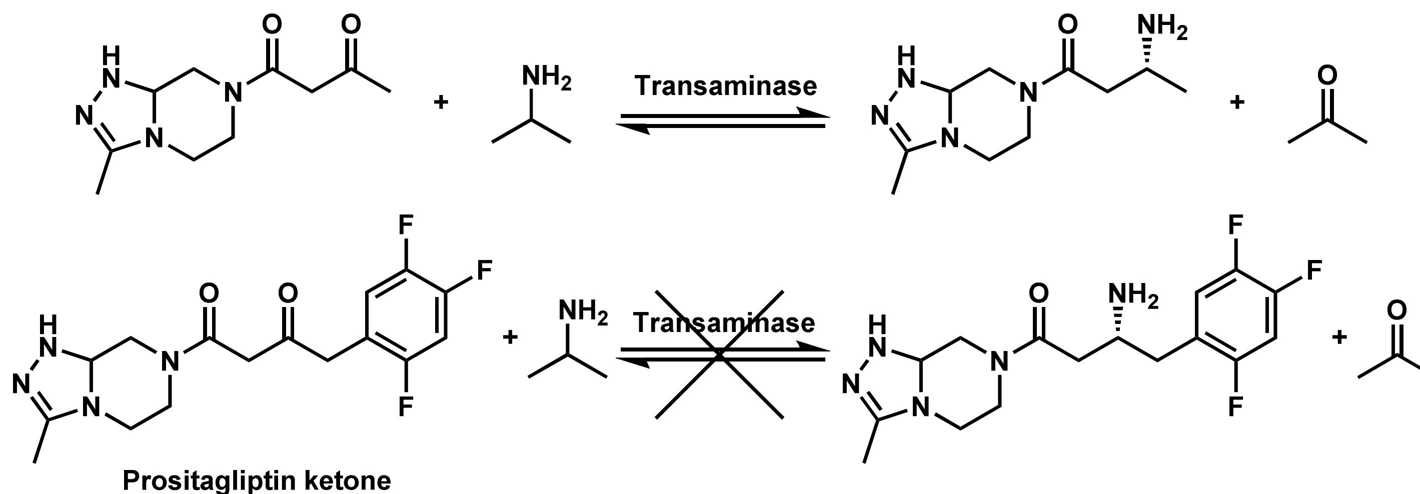
Directed Evolution



Savile, C. K., *et al. Science*, **2010**, 329, 305

Substrate range of transamirase

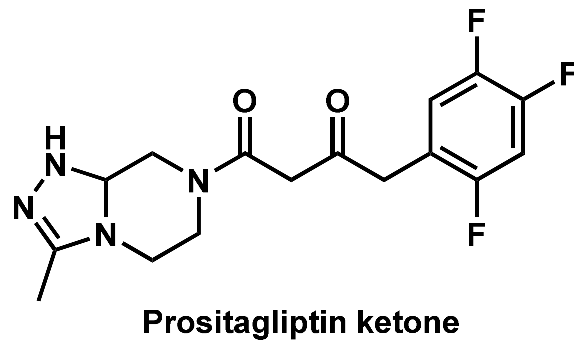
Transaminases have a limited substrate range;
Most of them accept only substrates with a substituent
no longer than methyl group at the position adjacent to the ketone

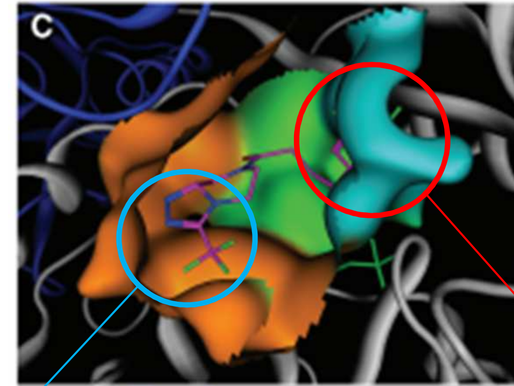


Screening a variety of transaminases provided no detectable activity.

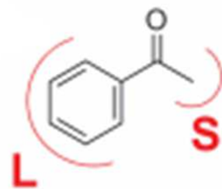
Docking Study

Using (R)-selective transaminase ATA-117, homology model was generated.

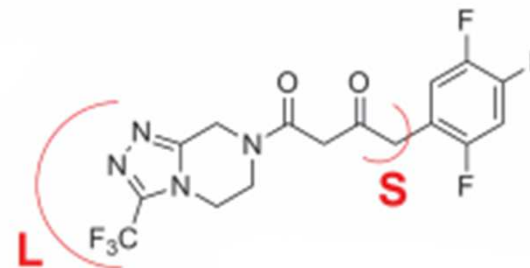




Prositagliptin ketone has problems in binding both with small pocket and large pocket.



substrate of ATA-117

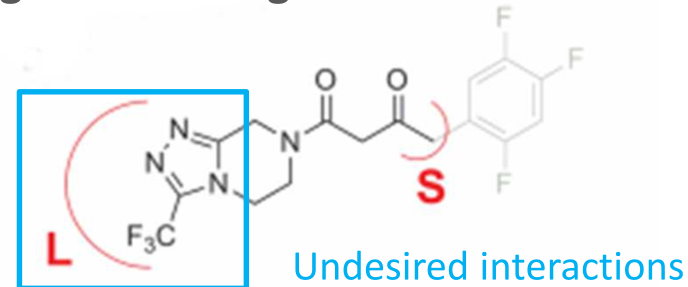


Savile, C. K., *et al. Science*, **2010**, 329, 305

Round 1

Firstly, the large binding pocket was engineered using truncated substrate

Consistent with the model,
this substrate was poorly active.
(4% conversion of 2g/L substrate)



Site saturation mutagenesis around the large pocket.

The best variant containing S^{223}→P^{223} showed 11-fold activity

A small library for potential activity on prositagliptin ketone

Analysis of the model suggested four residues (V^{69}, F^{122}, T^{283}, and A^{284}) potentially interact with trifluorophenyl group. Savile, C. K., *et al. Science*, **2010**, 329, 305

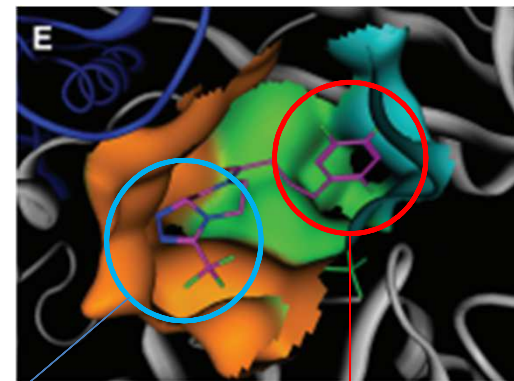
Round 1

ATA-117: $S^{223} \rightarrow P^{223}$(S223P) variant

Saturation mutagenesis on V^{69}, F^{122}, T^{283}, and A^{284} individually
&
Combinatorial library based on structural considerations.
(V69GA, F122AVLIG, T283GAS, and A284GF; 216 variants)

Single site saturation library provided no active variant.
Y26H, V65A, V69G, F122I, A284G provided initial active variant.
(0.7% conversion of 2g/L ketone)

Round 2



Savile, C. K., *et al. Science*, **2010**, 329, 305

Round 2

Initial active variant

Library screening

Variant containing 12 mutations showed 75-fold activity

Despite these enhancement, this catalyst was not yet practical...

- **Organic co-solvent (low water-solubility of ketone)**
- **Higher temperature (rate and solubility enhancement)**
- **Large excess of iPrNH2(transamination is equilibrium controlled)**

Transaminase have to withstand these harsh conditions.

Further evolution

Round 3-11

Transaminase with enhanced activity

Library generated by a variety of methods (total 36480 variants)
Rendering condition more stringent with the rising tolerance

Final screening conditions
Substrate : 2 → 100g/L
iPrNH2 : 0.5 → 1M
Co-solvent : 5 → 50% DMSO
pH : 7.5 to 8.5
Temp. : 22°C to 45°C

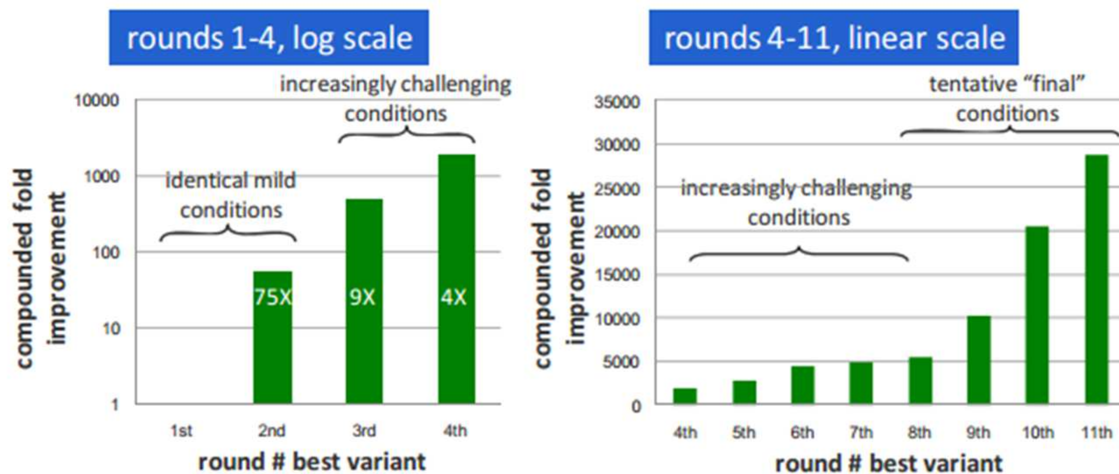
Process tolerant transaminase containing 27 mutations

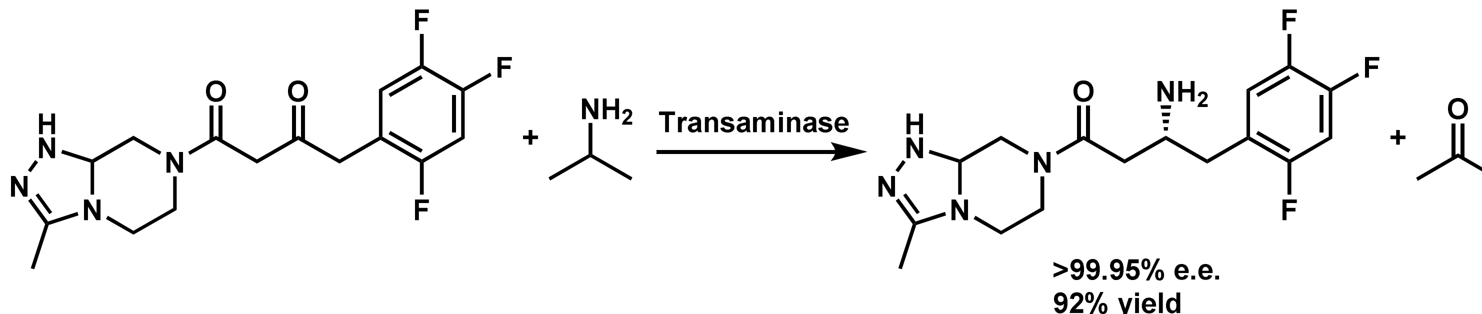
Summary of evolution

Substrate	Added Mutations*	[Substrate] in g/l	Assay changes	Round identified	Improvement over parent†
1	ATA-117	2		–	N/A
1	G136Y	2		1a	6
1	S223P	2		1a	11
2	S223P	2		1a	Not active
2	Y26H;‡ V65A;‡ V69G; F122I; A284G	2		1b	first active
2	H62T; G136Y; E137I; V199I; A209L; T282S	2		2	75
2	S8P; H26Y; G69C; M94I; I137T; G215C	5	5% DMSO to 5% MeOH; RT to 30°C	3	9
2	L61Y; C69T; Y136F; T137E	10	0.5 to 1 M $iPrNH_2$; pH 7.5 to pH 8.5	4	4
2	D81G; I94L; I96L; T178S; L269P; P297S; S321P	40	5 to 10% MeOH; 30 to 45°C	5	1.4
2	Y60F; L94I; A169L; S178T; G217N; L273Y	100	10 to 20% MeOH	6	1.6
2	S124H	100	20% MeOH to 25% DMSO	7	1.1
2	I122M; H124N	100		8	1.1
2	Q329H	100		9	1.9
2	N124T; Y150S; V152C; H329Q	50	25 to 50% DMSO; 0.5% acetone	10	2
2	S126T	50		11	1.4

Savile, C. K., *et al. Science*, **2010**, 329, 305

Improvements





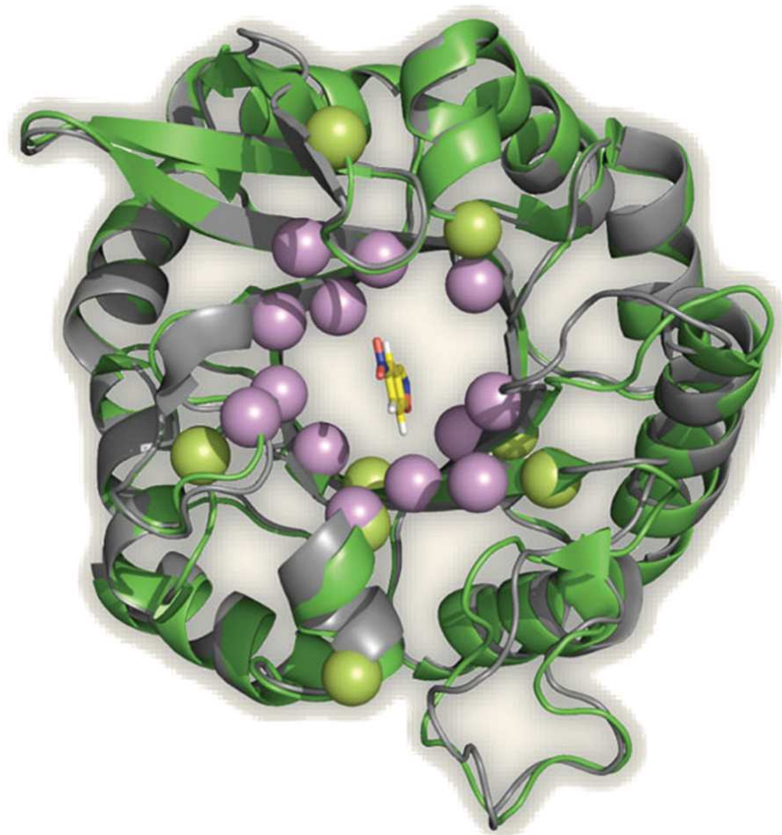
In comparison with Rh-catalyzed process,

- 10-13% increase in overall yield
- 53% increase in productivity (kg/L per day)
- 19% reduction in total waste
- No need for high-pressure H2 equipment
- Reduction in total manufacturing cost
- Elimination of all heavy metals

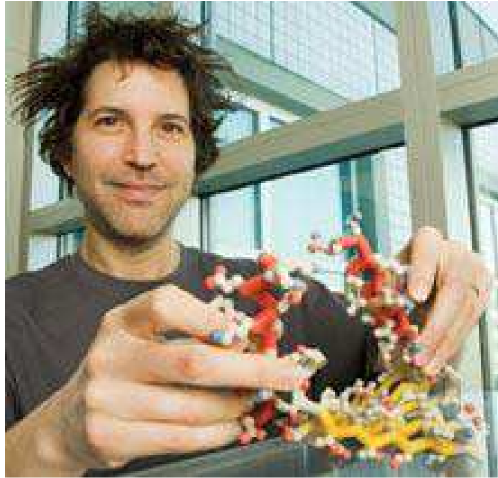
Savile, C. K., *et al. Science*, **2010**, 329, 305

Contents

1. Introduction
2. Catalytic Antibodies
3. Directed Evolution
4. Computational Enzyme Design
5. Summary & Future Outlook



Computational enzyme design



1984	B.A. at Harvard University
1989	Ph.D. at University of California, Berkeley
1990-1993	Postdoctoral work at University of California, San Francisco
199X	Professor of Biochemistry, University of Washington

David Baker

Investigator, Howard Hughes Medical Institute
Professor of Biochemistry, University of Washington
Adjunct Professor of Genome Sciences
Adjunct Professor of Physics
Adjunct Professor of Computer Science
Adjunct Professor of Chemical Engineering
Adjunct Professor of Bioengineering

The Baker laboratory developed the Rosetta algorithm for *de novo* protein structure prediction.
He aims to produce structural models for protein complexes as well as individual polypeptide chains.
His group is recognized as the first group to have designed a protein
Here, I'll focus on his works.

De novo computational Enzyme Design

1. QM caliculation to generate “Theozyme”,
   functional groups stabilizing the TS.
   A number of theozyme motif are usually generated.

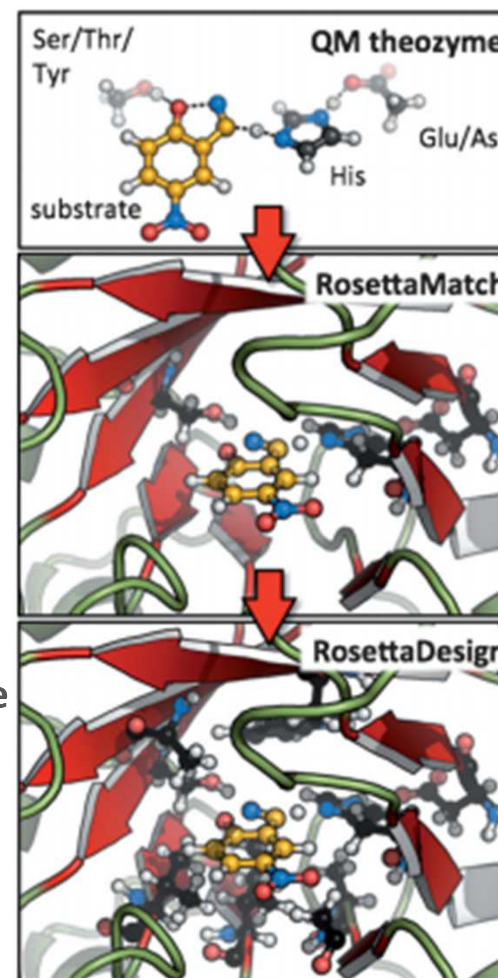
2. Run RosettaMatch to search active sites of existing proteins
   for backbone position that can accommodate
   three-dimensional side chain rearrangement of theozyme.

3. RossetaDesign attempts to generate the best possible
   stabilization for a geometry .
   Normally, even the highest ranked design difffer quite
   considerably from the original theozyme geometry.

4. Final designs are assessed towards their capability to stabilize
   the key catalytic residues on the basis of criteria such as
   Rosetta energy, hydrogen bond, active site geometry etc...

5. Enzyme assay

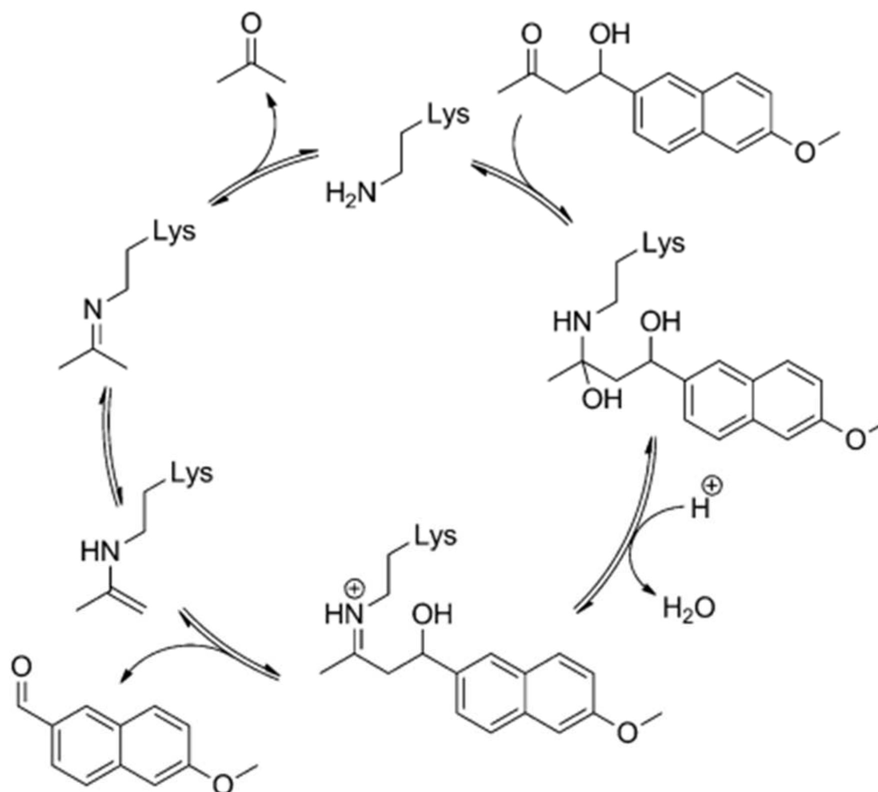
6. Directed evolution



Houk, K. N., *et al*. *Angew. Chem. Int. Ed*. **2013**, 52, 5700

Achievement – Retro-Aldolases

The first example of *de novo* computational enzyme design is retro-aldolase

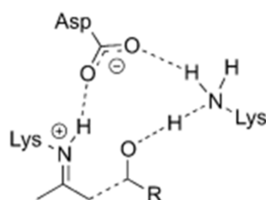


Jiang, L. *Science*, **2008,** 319, 1387

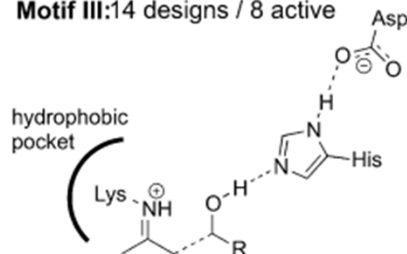
Achievement – Retro-Aldolases

Theozyme design

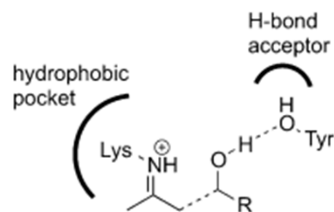
Motif I: 12 designs / 0 active



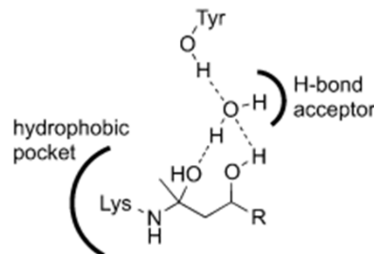
Motif III: 14 designs / 8 active



Motif II: 10 designs / 0 active



Motif IV: 38 designs / 22 active



30 design showed detectable activity

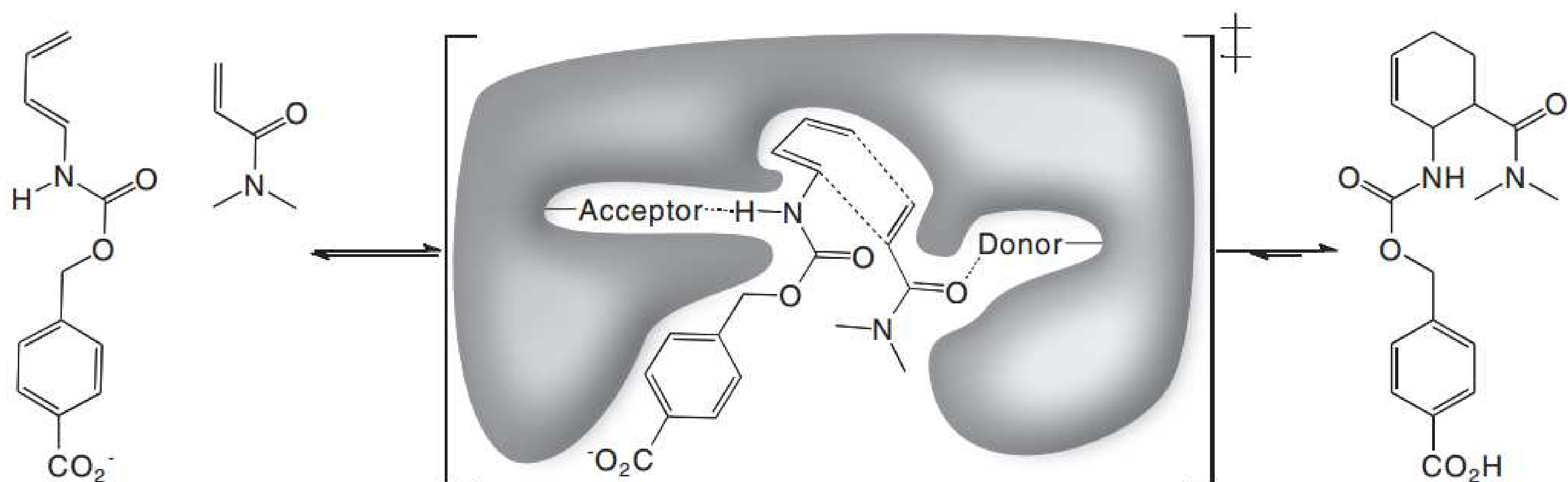
k_{cat}/K_M = 0.02 – 0.74$M^{-1}s^{-1}$

Enhancement of rate is inferior to that of catalytic antibodies.

Low activity can be attributed to dynamic distortion such as solvation, conformational flexibility.

Jiang, L. *Science*, **2008,** 319, 1387

Achievement – Diels-Alderase

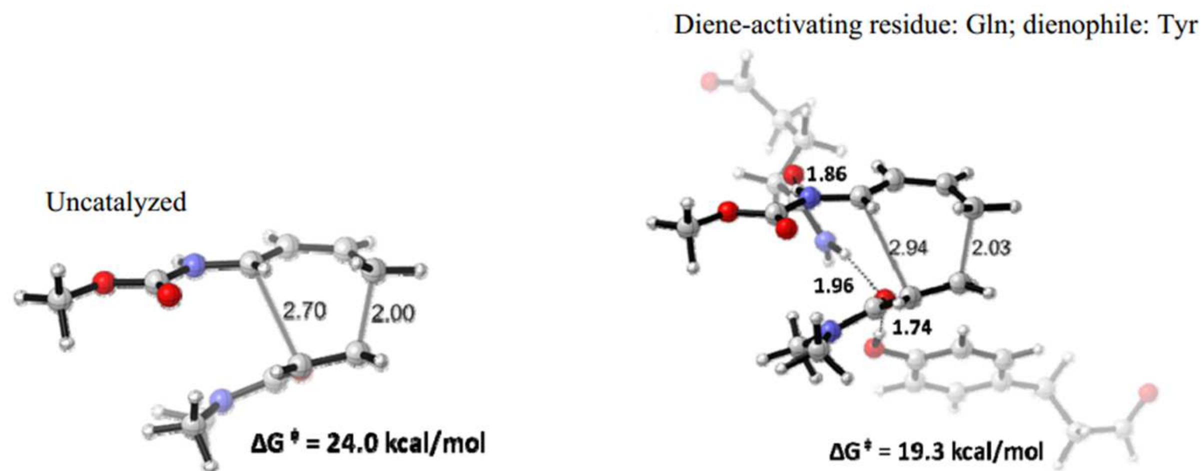


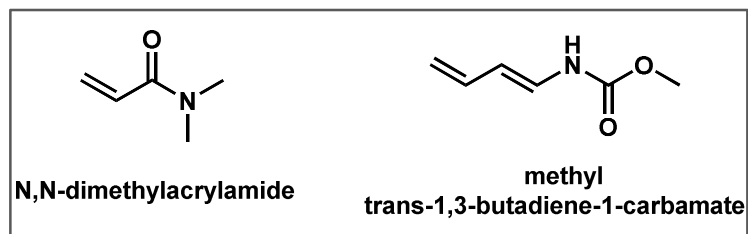
Designing enzyme that catalyze bimolecular bond-forming reaction is challenging. (Both substrate must be bound in the proper relative orientation.)

In this design, hydrogen bond donor & acceptor modulate molecular orbital energies and stabilize TS to accelerate the intermolecular Diels-Alder reaction.

Siegel, J. B., *et al. Science*, **2010**, 329, 309

QM calculations





Model substrates

Quantum mechanical simulation predict that these hydrogen bonds can stabilize the TS by 4.7kcal/mol.

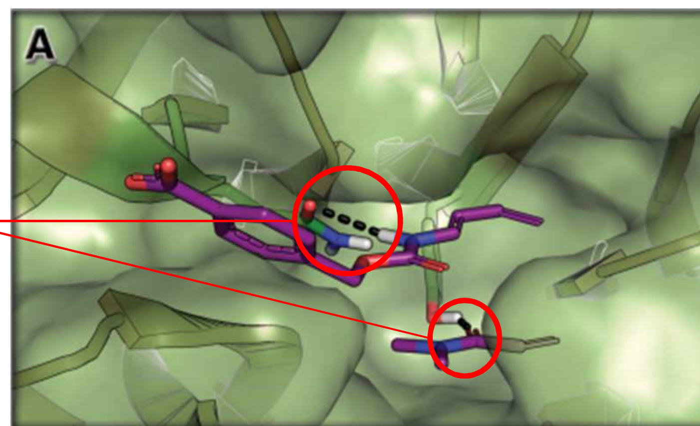
Siegel, J. B., *et al. Science*, **2010**, 329, 309

Design and screening

84 designs were synthesized through QM calculation, RosettaMatch, and RosettaDesign. Two of them were found to have Diels-Alderase activity. (DA_20_10 and DA_42_00)

Then, these Diels-Alderase were evolved to further improved the activity

Since initial activity is promising, directed evolution can be effectively combined with computational enzyme design.



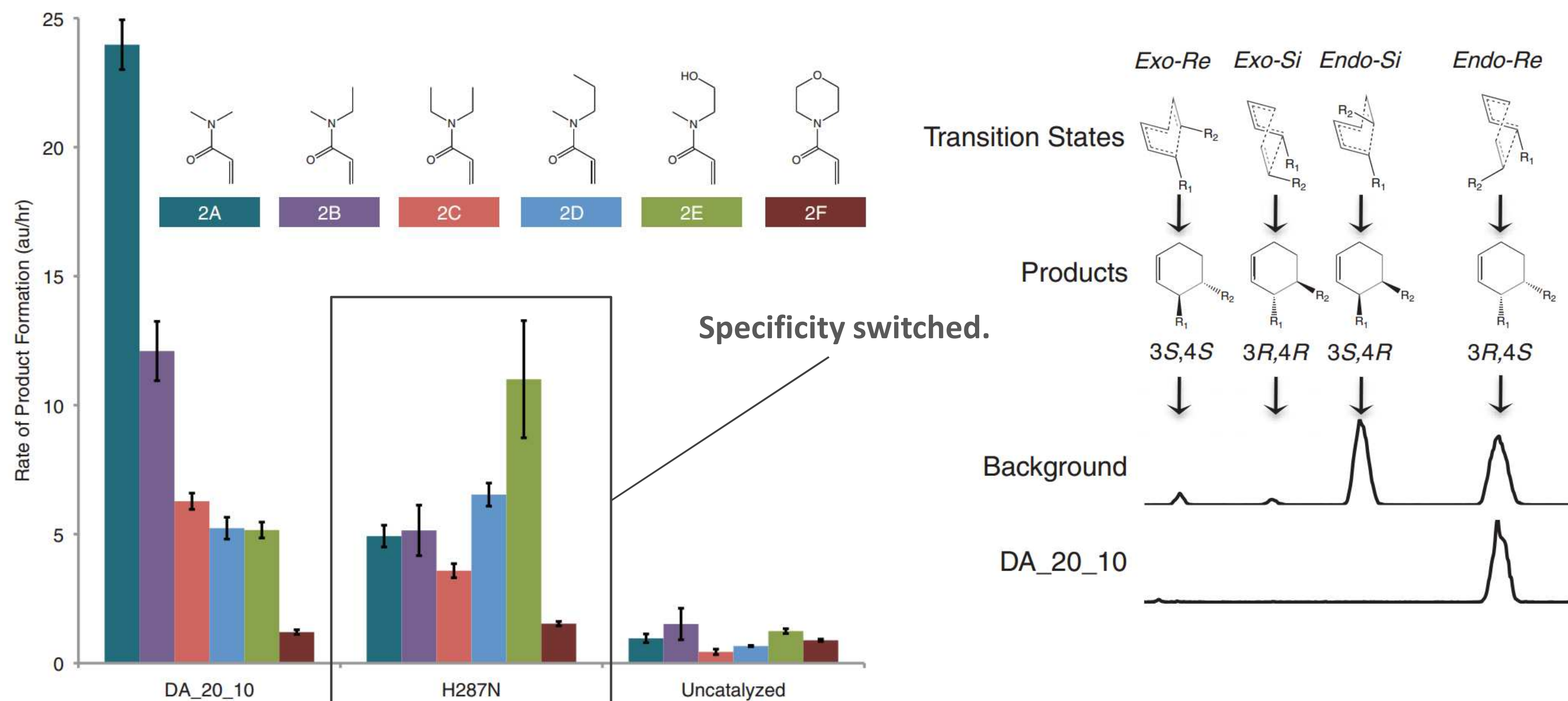
Structure of DA_20_10

DA_20_10(100-fold active), DA_42_04(20-fold active)

Catalyst	k_{cat} (hour^{-1})	$K_{M\text{-diene}}$ (mM)	$K_{M\text{-dienophile}}$ (mM)	$k_{cat}/K_{M\text{-diene}}$ (s^{-1} M^{-1})	$k_{cat}/K_{M\text{-dienophile}}$ (s^{-1} M^{-1})	$k_{cat}/(K_{M\text{-diene}} \times K_{M\text{-dienophile}})$ (s^{-1} M^{-1} M^{-1})
DA_20_00	0.10 ± 0.02	3.5 ± 1.5	146.0 ± 2.5	0.008	0.0002	0.06
DA_20_10	2.13 ± 0.24	1.3 ± 0.1	72.8 ± 5.1	0.455	0.0081	6.23
DA_42_04	0.03 ± 0.01	0.5 ± 0.1	16.2 ± 3.2	0.017	0.0005	1.03
mAb 7D4	0.21	1.0	1.7	0.058	0.0343	20.18
mAb 4D5	0.21	1.6	5.9	0.036	0.0099	6.19

Siegel, J. B., *et al. Science*, **2010**, 329, 309

Specificity and selectivity



In addition to high substrate specificity and stereoselectivity, once an initial active enzyme is engineered, it can be easily modified to catalyze similar reactions with alternate substrate.

Challenges in enzyme design

There is much room for improvement in computational enzyme design.

- Only two of 50 designs exhibited detectable activity (low success rate)
- Low activity relative to natural enzymes
- The resulting microenvironments are still far from “finely tuned”

Design through Crowd Sourcing

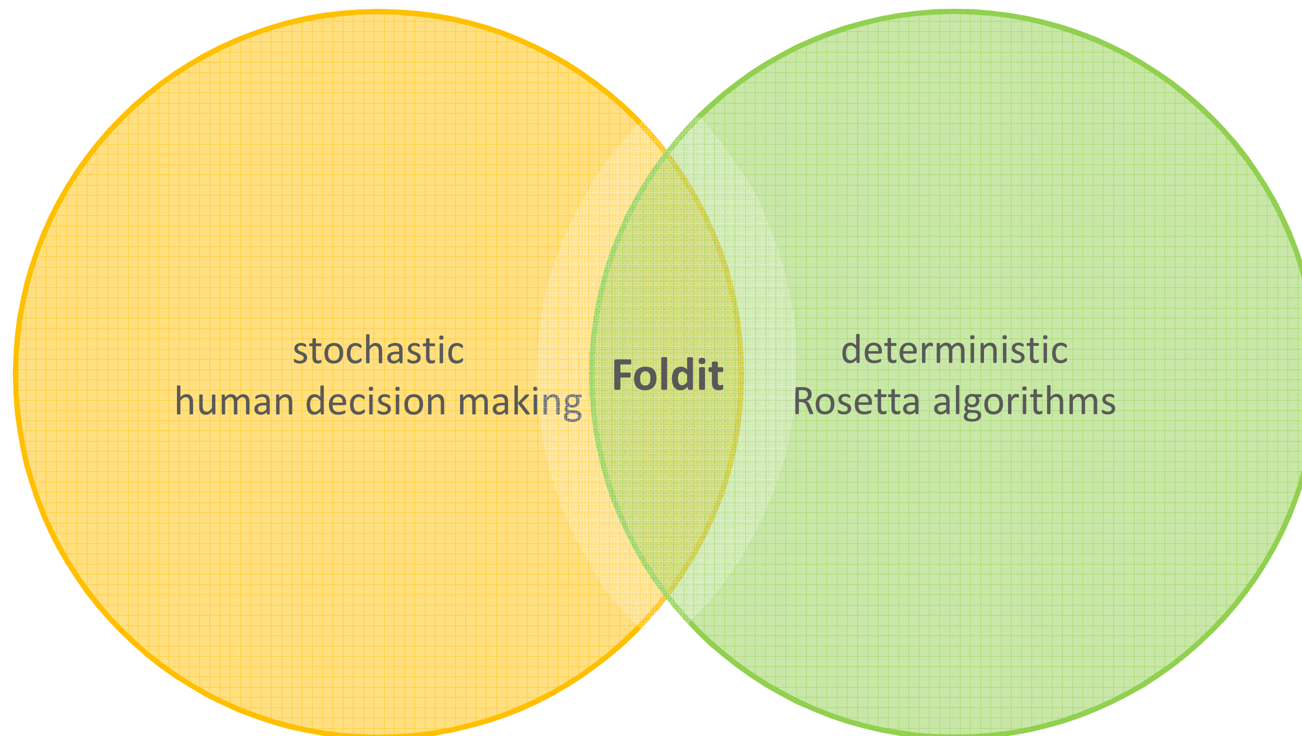
Structure prediction with Foldit

http://fold.it/portal/

About Foldit

Foldit is a multiplayer online game developed on the hypothesis that problem of protein structure prediction can be solved with human directed computing.

The goal of Foldit is producing accurate protein structure models through gameplay.



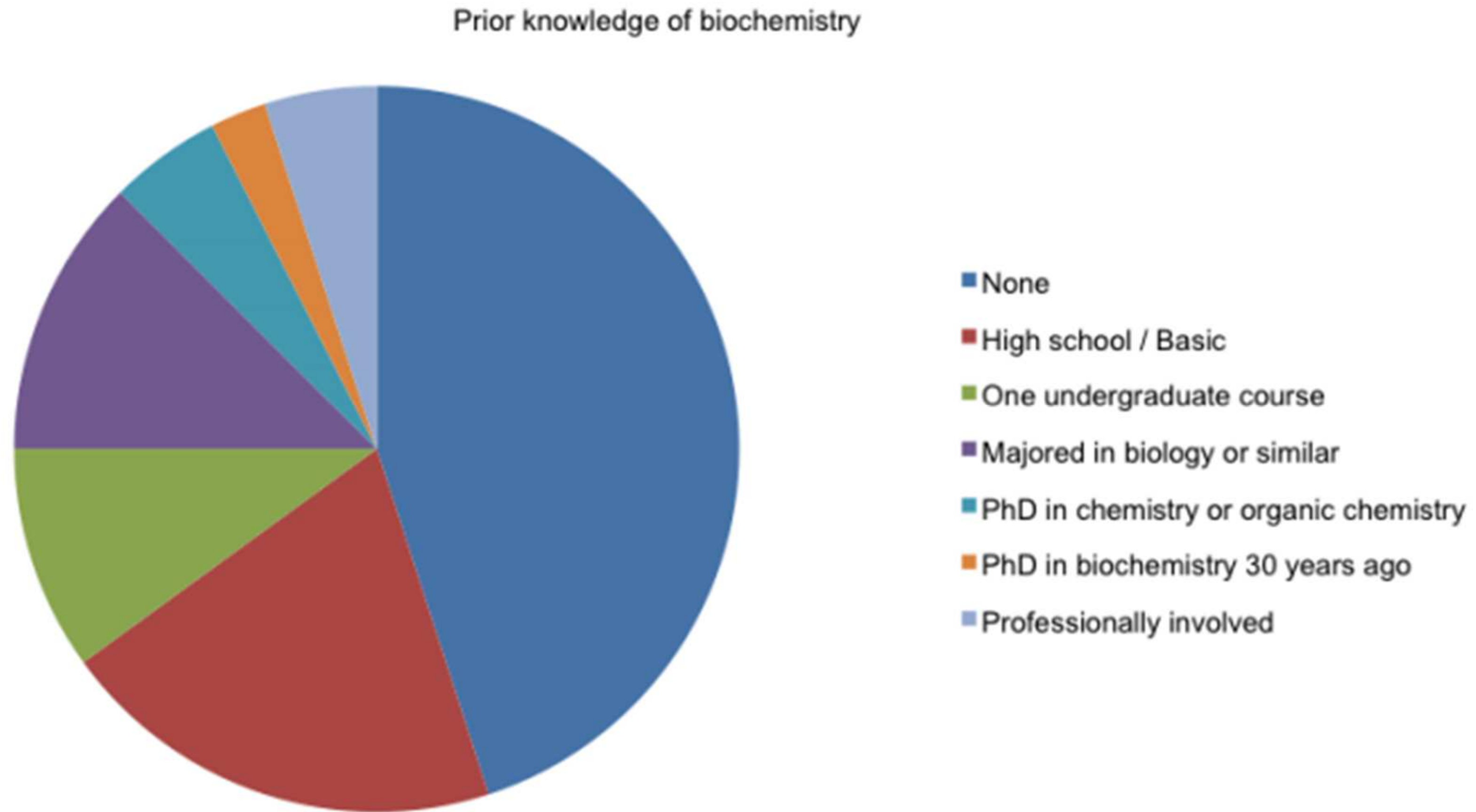
User interface

Tools and visualizations make the game approachable for non-scientists.

1. Interatomic repulsion
2. Hydrogen bond
3. Exposed hydrophobic side chain
4. Hydrophilic side chain
5. High residue energy
6. User tool “rubber bands”
7. User tool “freezing”
8. Score
9. Leader board
10. Tools and options
11. Chat window
12. “Cookbook”

Cooper, S., *et al. Nature,* **2010**, 466, 756

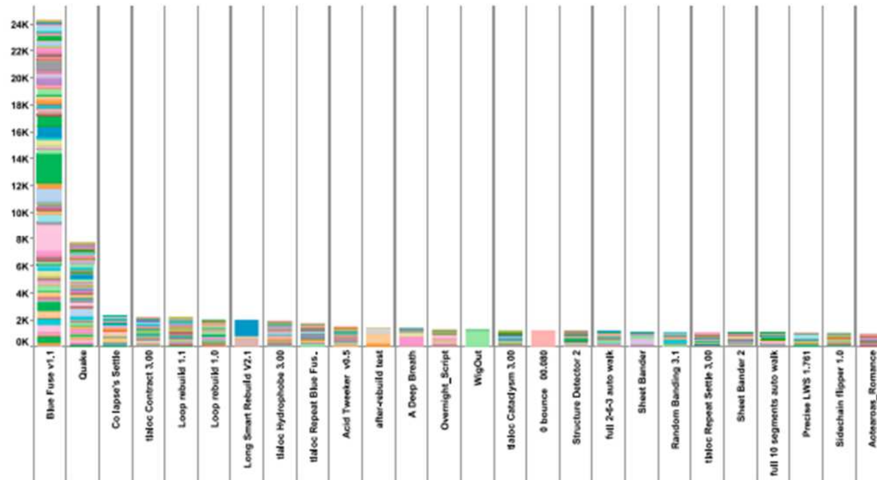
Background of players

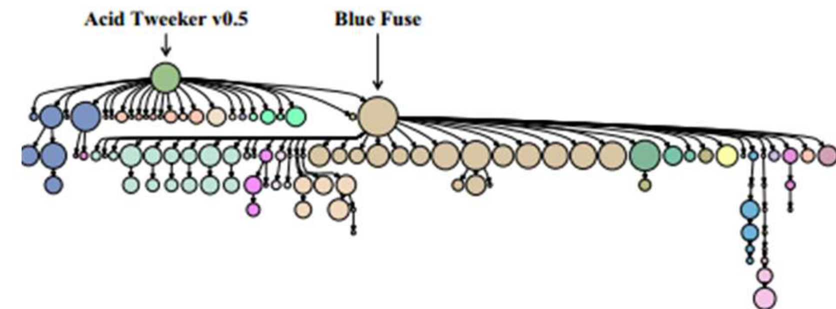


Cooper, S., *et al. Nature,* 2010, **466**, 756

Only a minority have advanced knowledge of biochemistry

Social evolution of “recipes”





Khatib, F., *et al*. *PNAS*, **2011**, 108, 18949

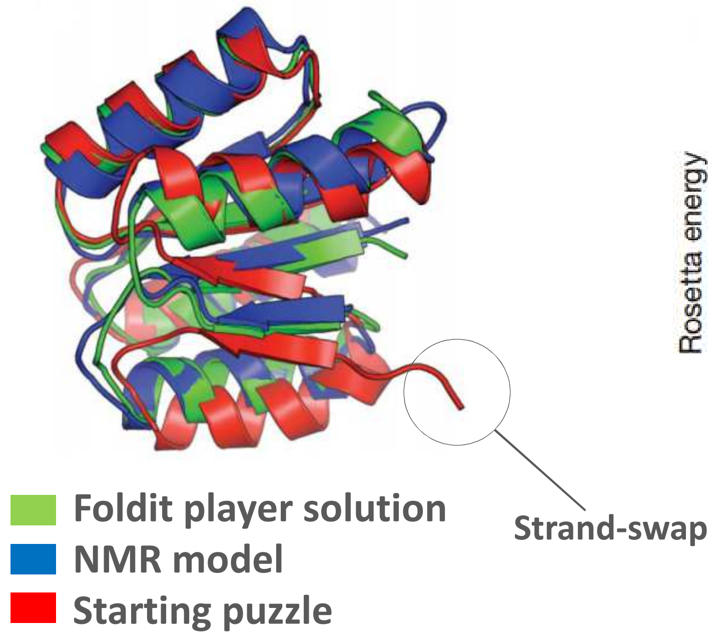
Foldit recipes are used with very different frequencies.

Popular recipes are copied and modified more often to evolve.

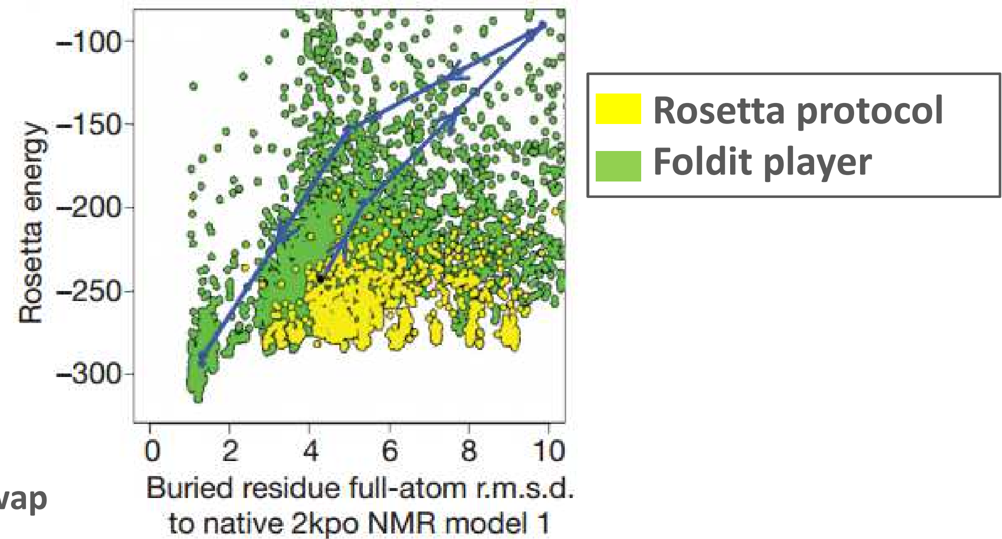
Thus, popular recipes spawn large numbers of descendants,

and there are multiple independent lineages each spanning many generations.

Strand-swap puzzle



Foldit player solution
NMR model
Starting puzzle

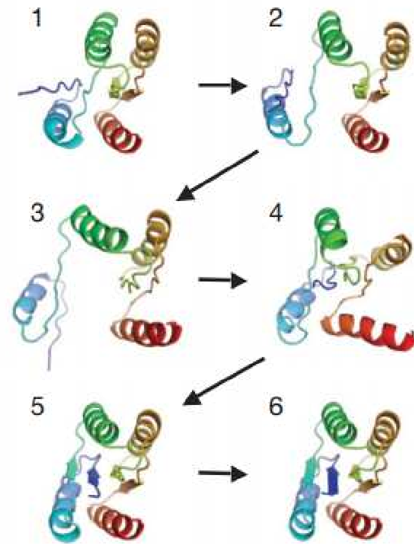


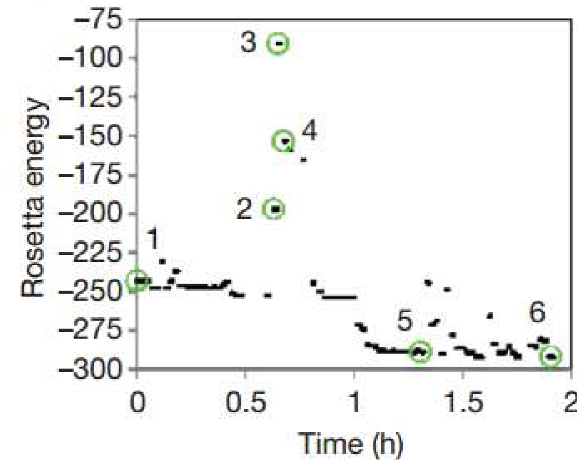
Rosetta couldn't reach swapped conformations

Humans significantly outperformed computers!

Cooper, S., *et al. Nature,* **2010**, 756

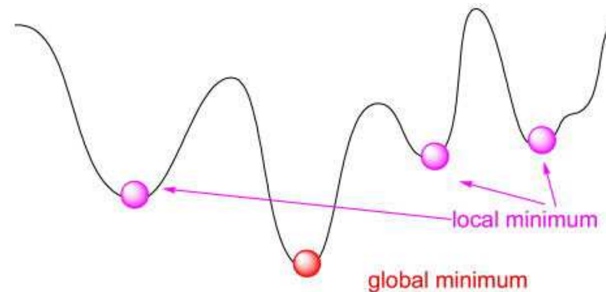
Key difference between humans and computers





Cooper, S., *et al. Nature,* 2010, **466**, 756

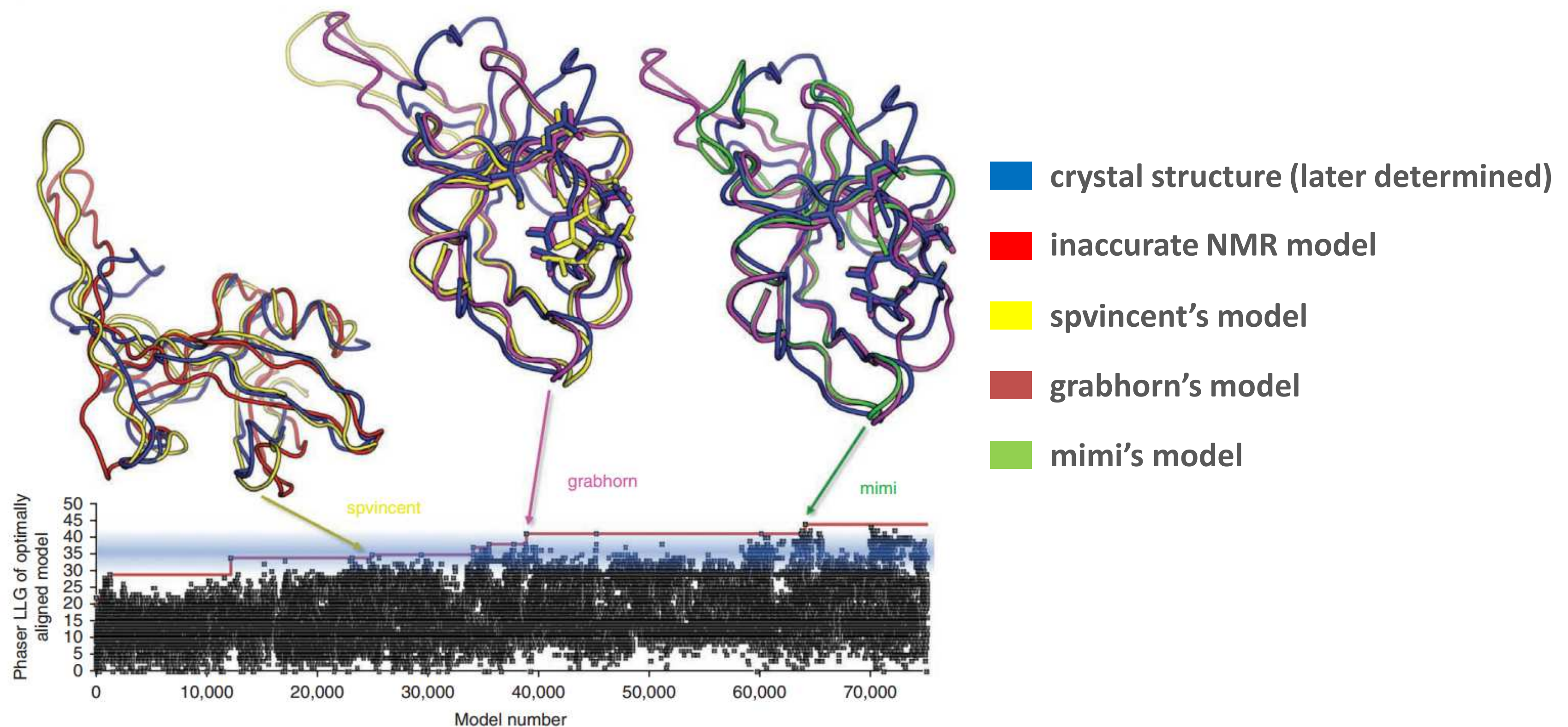
The player went through high energy conformations to reach the native state.



Deterministic minimize function has disadvantage of being trapped at local minima.

Crystal structure solved by Foldit

Crystal structure of M-PMV retroviral protease had been elusive for over a decade
(Despite the availability of crystals, wide range of attempts were unsuccessful.)
Foldit players solved this long-standing protein crystal structure problem

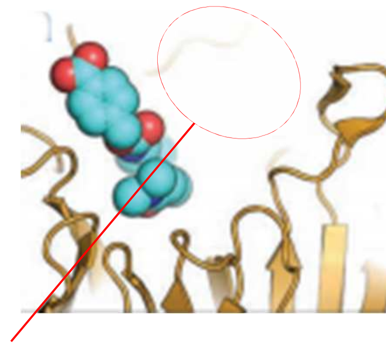


crystal structure (later determined)
inaccurate NMR model
spvincent's model
grabhorn's model
mimi's model

Khatib, F., *et al*. *Nat. Struct. Mol. Biol*. **2011**, 18, 1175

Enzyme remodeling through Foldit

Computationally designed Diels-Alderase, DA2010 has open active site, leaving substrate exposed to solvent.
To improve activity, Foldit players were enlisted to remodel loops.
Players were allowed to insertions and deletions.



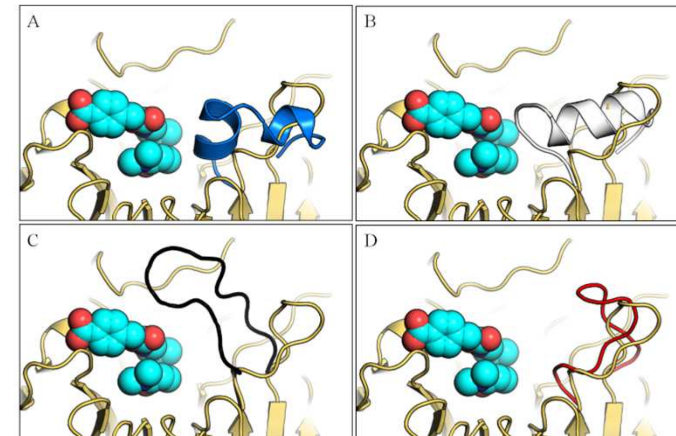
open active site

"Cover the Ligand"
Remodel any of four active site loops

After a week

4 of 69773 designs made particularly favorable interactions

Library screening



Most of variants exhibited no activity.
CE0(based on design A) showed activity 10-fold decrease relative to DA2010

Eiben, C., *et al. Nat. Biotechnol*. **2012**, 30, 190

Enzyme remodeling through Foldit

CE0 (suboptimal interaction with the substrate or transition state)

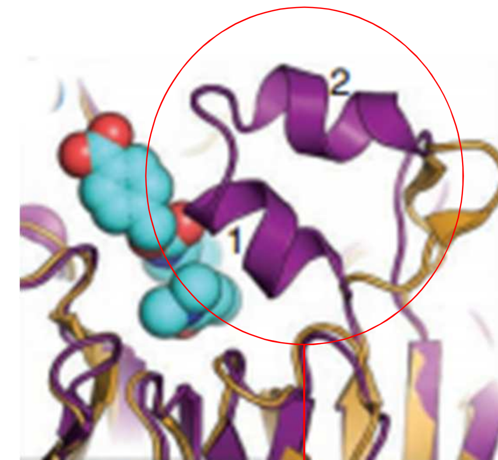
Exploring design further

CE4 was 9-fold more active than DA2010

"Back Me Up"
Transform a loop into additional helix
to stabilize initially designed helix

After a week

Based the top design of 109,421, library was created to provide CE6
CE6 has helix-turn-helix motif, and 18-fold more active than DA2010



Helix-turn-helix motif

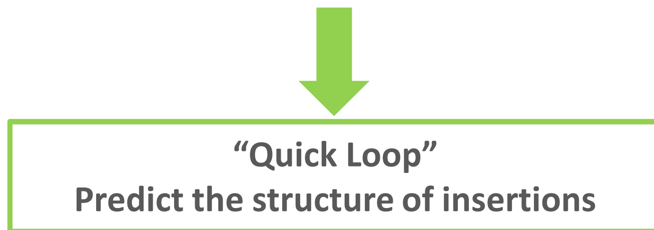
Protein	$K_{M\text{-diene}}$ (mM)	$K_{M\text{-dienophile}}$ (mM)	k_{cat} (h^{-1})	$k_{cat}/(K_{M\text{-diene}} * K_{M\text{-dienophile}})$ ($s^{-1}M^{-1}M^{-1}$)
DA2010	1.2 ± 0.2	101 ± 21	2.1 ± 0.3	4.7 ± 1.5
CE0	n.d.	n.d.	n.d.	0.5 ± 0.05
CE4	0.5 ± 0.03	31 ± 3.0	2.4 ± 0.1	42.4 ± 5.7
CE6	0.2 ± 0.03	35 ± 1.4	2.2 ± 0.1	87.3 ± 13.9

k_{uncat} under these conditions is 2.2×10^{-2} M^{-1} h^{-1}. n.d., not detectable.

Eiben, C., *et al. Nat. Biotechnol.* **2012**, 30, 190

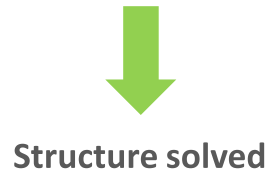
Enzyme remodeling through Foldit

Enhanced Diels-Alderase CE6 with community-designed helix-turn-helix motif

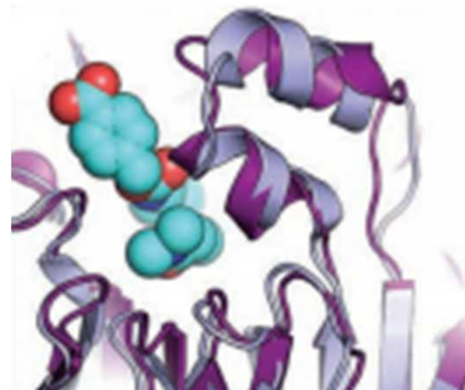


After a week

The lowest energy of 335,697 solutions was selected as predicted structure



Structure solved



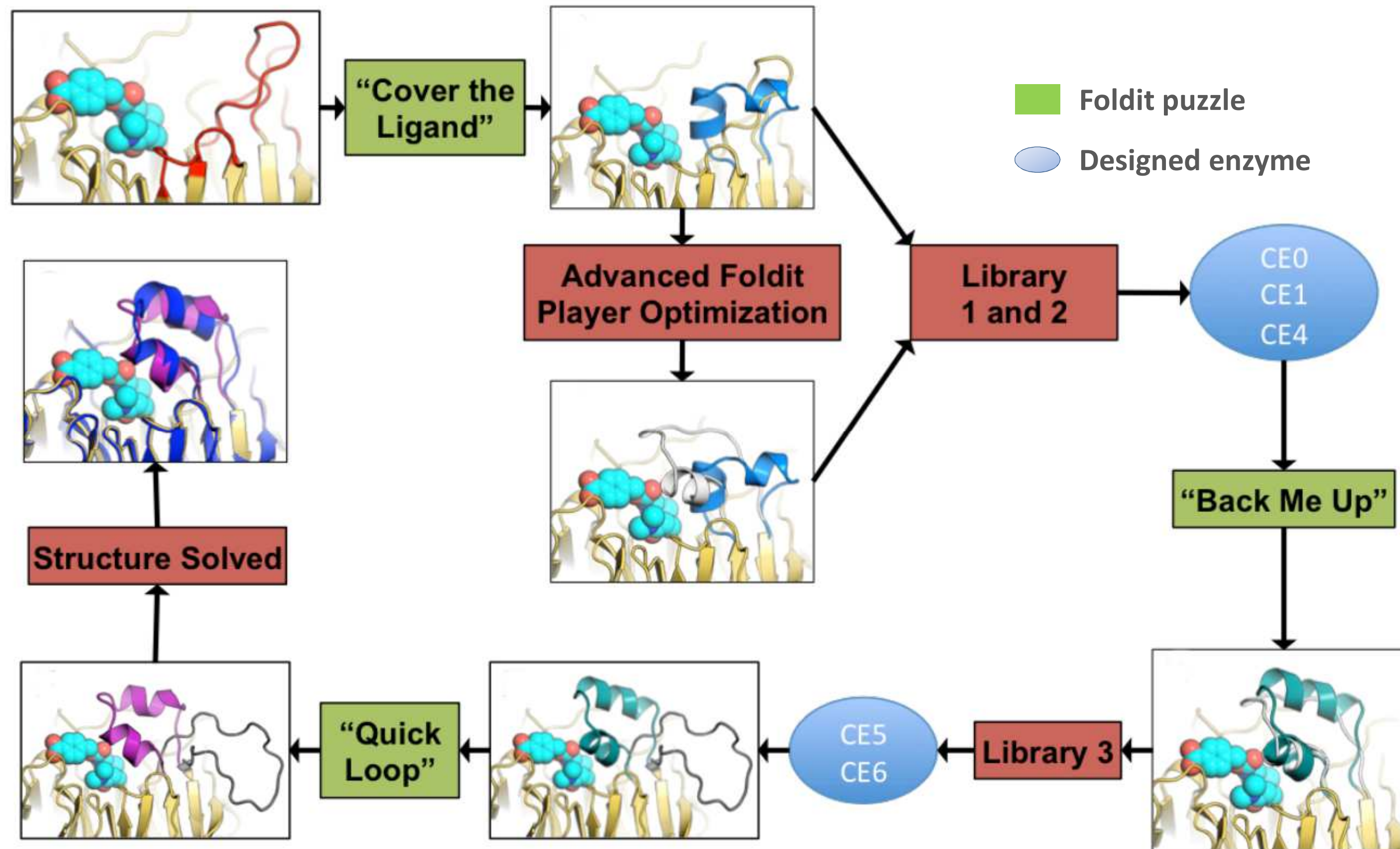
Predicted structure

Crystal structure

Eiben, C., *et al. Nat. Biotechnol*. **2012**, 30, 190

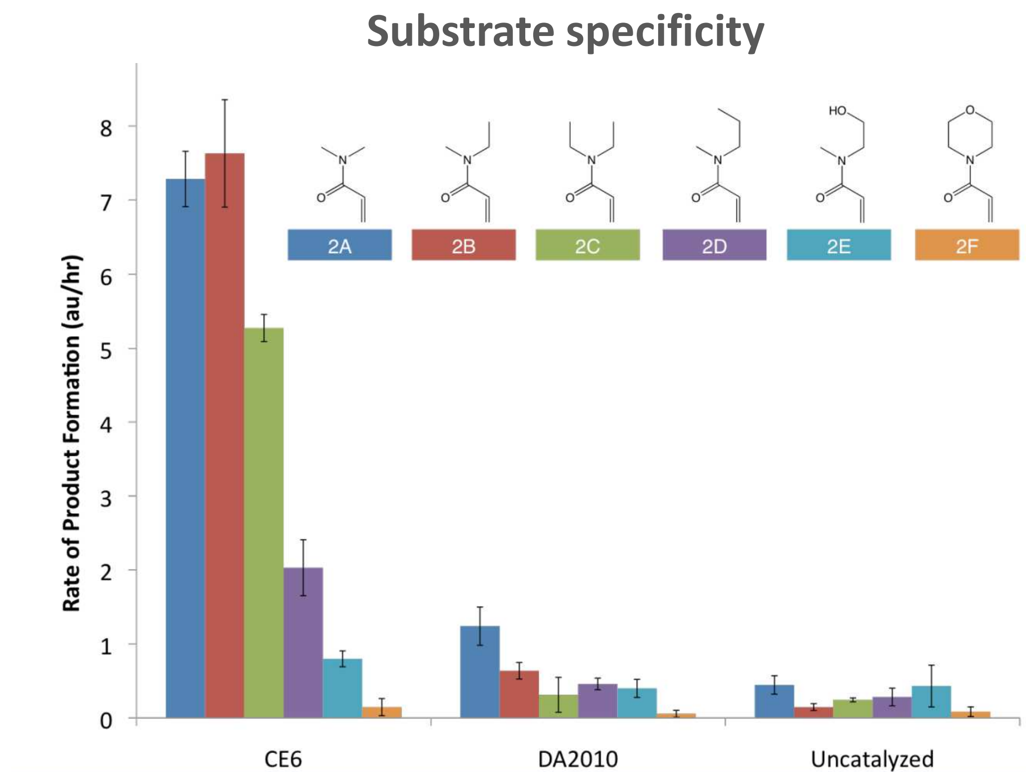
Enzyme remodeling through Foldit

Workflow of design process



Eiben, C., *et al. Nat. Biotechnol*. **2012**, 30, 190

Enzyme remodeling through Foldit



Specificity for hydrophilic dienopiles increased
However, specificity for similar-size hydrophobic dienophiles was lost

Further improvement remain possible

Predicting protein structures with a multiplayer online game

Seth Cooper[1], Firas Khatib[2], Adrien Treuille[1,3], Janos Barbero[1], Jeehyung Lee[3], Michael Beenen[1], Andrew Leaver-Fay[2]†, David Baker[2,4], Zoran Popović[1] & Foldit players

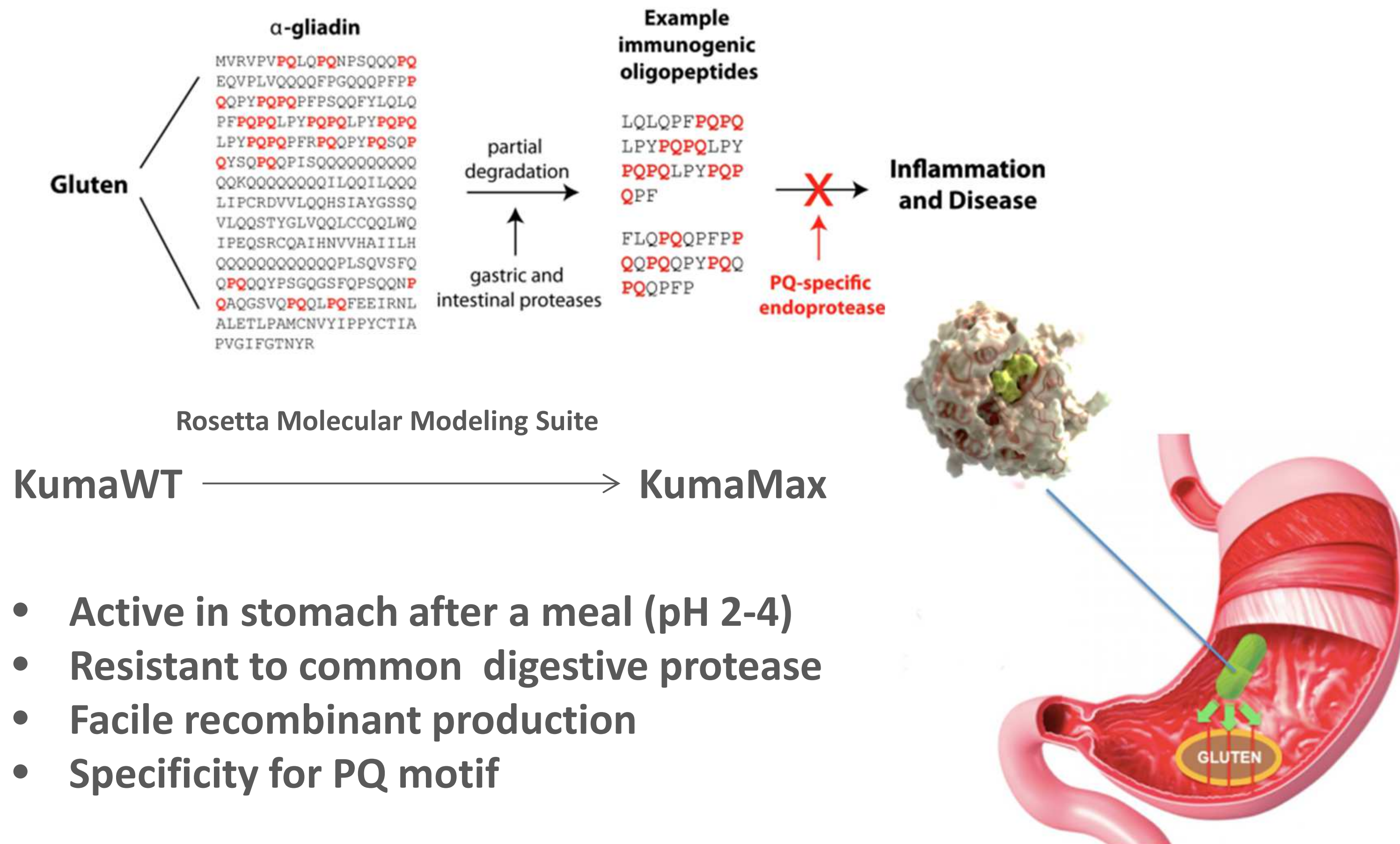
Crystal structure of a monomeric retroviral protease solved by protein folding game players

Firas Khatib[1], Frank DiMaio[1], Foldit Contenders Group, Foldit Void Crushers Group, Seth Cooper[2], Maciej Kazmierczyk[3], Miroslaw Gilski[3,4], Szymon Krzywda[3], Helena Zabranska[5], Iva Pichova[5], James Thompson[1], Zoran Popović[2], Mariusz Jaskolski[3,4] & David Baker[1,6]

Human insight and creativity can be extended to molecular-scale design problems!

Enzyme therapy

Computational design of α-gliadin peptidase as a therapeutic for celiac disease



- Active in stomach after a meal (pH 2-4)
- Resistant to common digestive protease
- Facile recombinant production
- Specificity for PQ motif

Gordon, S. R. *J. Am. Chem. Soc.* **2012**, 134, 20513

Contents

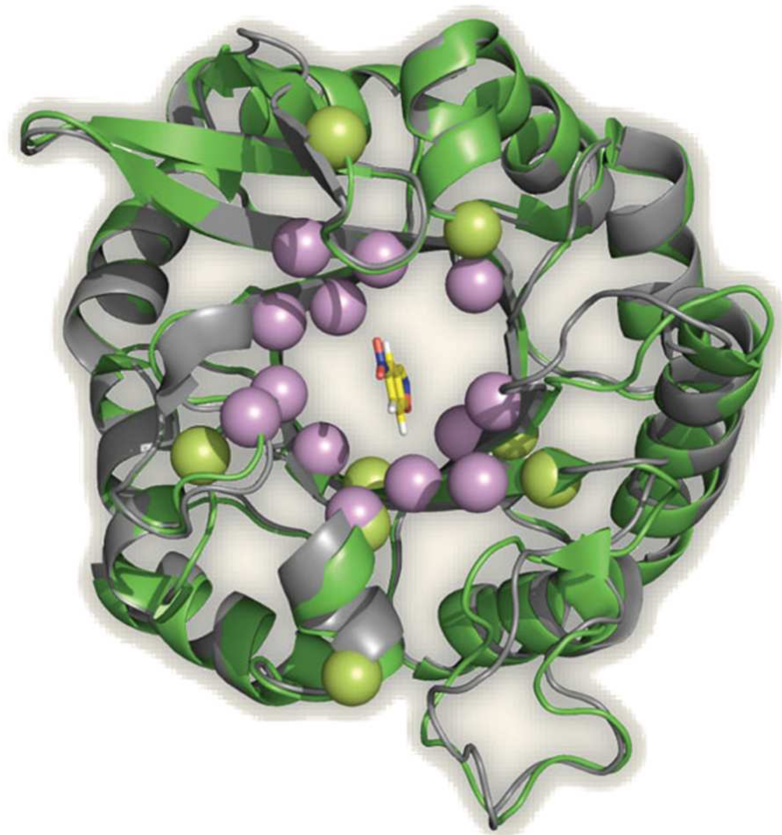
1. Introduction

2. Catalytic Antibodies

3. Directed Evolution

4. Computational Enzyme Design

5. Summary & Future Outlook



Summary

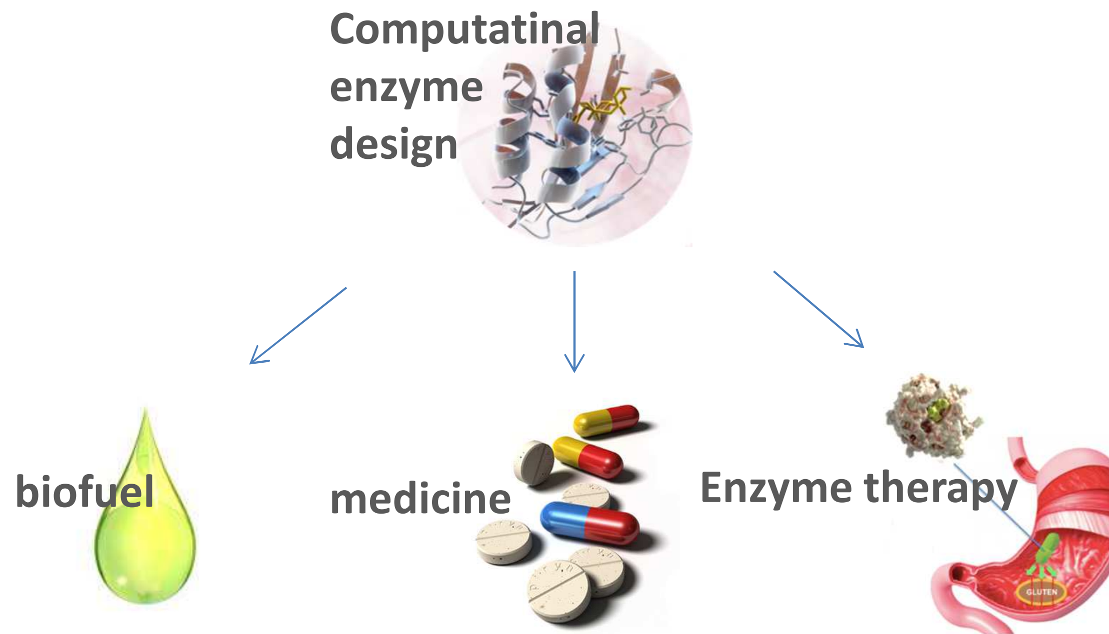
	k_{cat}[a]	K_M[b]	k_{cat}/K_M[c]	k_{cat}/k_{uncat}	$[k_{cat}/K_M]/k_{uncat}$
nat. enzymes	av 10^5	av 10^{-4}	10^6–10^9	10^6–10^{17}	10^8–10^{29}
cat. antibodies	10^{-2}–1	av 10^{-4}	10^2–10^4	10^3–10^6	10^5–10^9
Kemp elim.					
cat. antibodies	10^{-1}–1	10^{-3}–10^{-4}	10^2–10^3	10^3–10^6	10^7–10^9
comp. designs	10^{-2}–1	av 10^{-3}	10–10^2	10^3–10^6	10^7–10^9
evolved designs	1–20	10^{-3}–10^{-5}	10^3–10^6	10^6–10^7	10^7–10^{11}
Retro-Aldol					
cat. antibodies	10^{-3}–10^{-1}	10^{-4}–10^{-5}	10–10^3	10^5–10^6	10^7–10^9
comp. designs	10^{-2}–10^{-1}	av 10^{-4}	10^{-2}–10^{-1}	10^3–10^4	10^6–10^7
Diels–Alder[d]					
cat. antibodies	av 10^{-5}	av 10^{-3}	av 10	av 10^3	10^9
comp. designs	10^{-5}–10^{-4}	10^{-1}–10^{-4}	1–10^2	10^3–10^4	10^7–10^{11}

[a] In units of s^{-1}. [b] In units of M. [c] In units of $M^{-1}s^{-1}$. [d] $k_{cat}/(K_{M\text{-diene}} \times K_{M\text{-dienophile}})$ instead of k_{cat}/K_M in units of $s^{-1}M^{-1}M^{-1}$ and $(k_{cat}/(K_{M\text{-diene}} \times K_{M\text{-dienophile}})/k_{uncat})$ instead of $(k_{cat}/K_M)/k_{uncat}$ in units of $M^{-1}M^{-1}$

Computational design and directed evolution enanced K_M value comparable to natural enzymes. However, in terms of k_{cat}/k_{uncat} value, even the most active designed enzyme have rates comparable only with the least proficient of natural enzyme. Further enhancement is still required.

Future Outlook

Computatinal enzyme design holds promise for the production of renewable fuels,drugs, therapeutics.

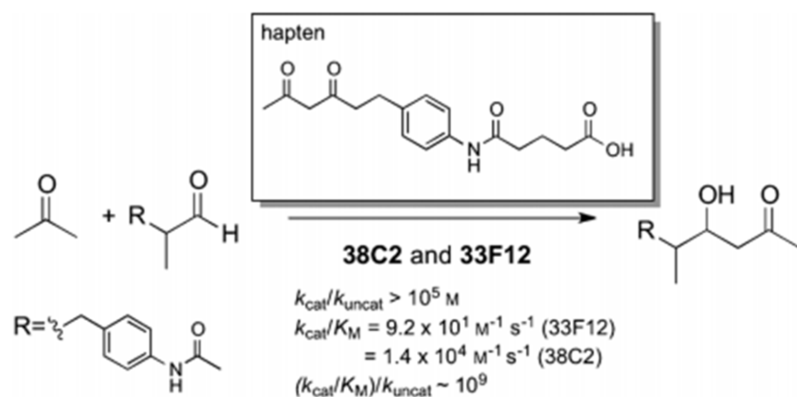


Protein engineering is one of the most dynamically developing scientific fields.

Efforts to engineer truly enzyme-like proteins are just getting started!

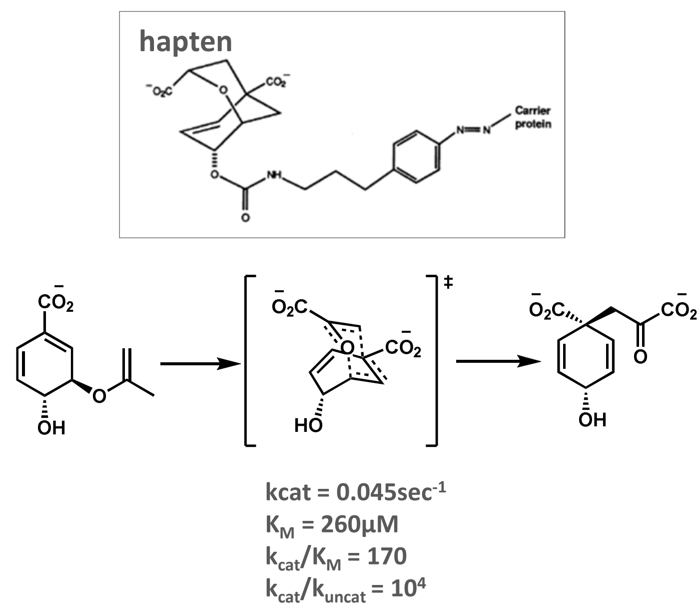
Catalytic Antibodies

Aldol reaction



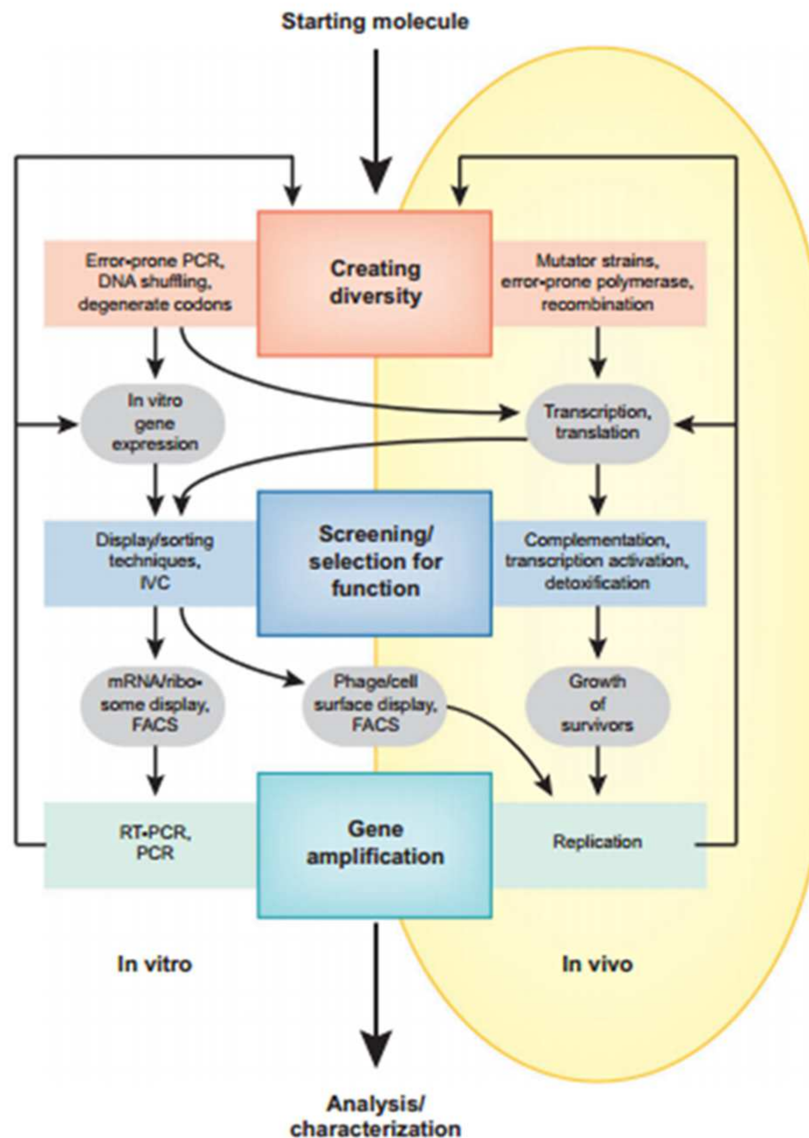
Wagner, J. *Science*, **1995**, 270, 1797

Claisen rearrangement



Schultz, P. *Science*, **1988**, 240, 426

Directed evolution

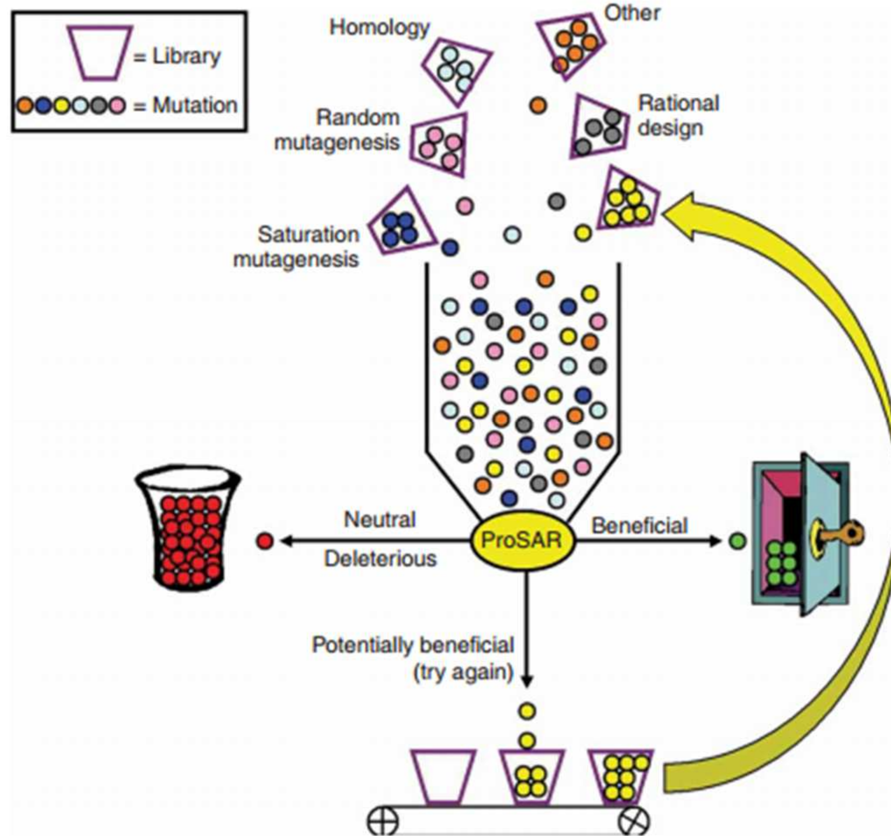


Today, numerous experimental methods are available in directed evolution process.

Hilvert, D. *Annu. Rev. Biophys*. **2008**, 37, 153

Directed evolution

Protein sequence activity relationships (ProSAR)



Variants are categorized into four classes.

Neutral and Deleterous are discarded.

Beneficial variants will be parental enzyme in the next round.

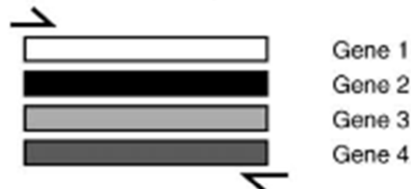
Potentially beneficial variants are retested.

Fox, R. J. *Nat. Biotechnol*. **2007**, 25, 338

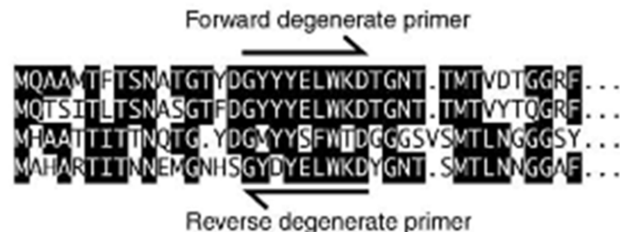
Directed evolution

Gene shuffling
degenerate oligonucleotide gene shuffling(DOGS)

1. Design oligonucleotide primers with 3' ends specific for the N- or C-terminus of each candidate gene. Incorporate common nested 5' ends with suitable restriction sites for directional cloning of PCR products. PCR amplify each gene for use as PCR template.

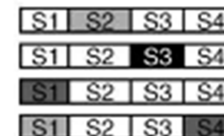


2. Design complementary degenerate primer pairs based upon one or more conserved motifs found in candidate genes.



3. Amplify each of the individual segments (S1-S4) for each gene using the degenerate primers and the common nested primers.

4. Mix segments from each gene to give desired levels of chimerisation. Regenerate full length chimeric genes by overlap extension of segments followed by PCR with primers specific for the common nested ends.



5. Digest and ligate full length fragments into an appropriate cloning vector, transform into expression host and screen individual recombinants for desired properties.

Bergquist, P.L. *Biomol. Eng*. **2005**, 22, 63