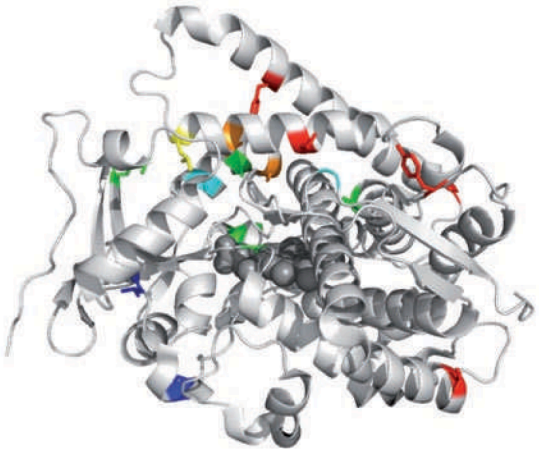
Career of Frances H. Arnold



Chris Prier

MacMillan Group Meeting

June 5, 2013

Frances Arnold

- B.S., 1979, Princeton University, Mechanical and Aerospace Engineering
- Ph.D., 1985, University of California, Berkeley, Chemical Engineering

  Advisor: Harvey W. Blanch

  Theoretical treatment of affinity chromatography
- Postdoctoral, 1985-1986, University of California, Berkeley

  Advisor: Ignacio Tinoco, Jr.

  Structural studies of mismatched DNA oligonucleotide duplexes
- Postdoctoral, 1986, California Institute of Technology
- Professor, 1987 to present, California Institute of Technology

  Dick and Barbara Dickinson Professor of Chemical Engineering, Bioengineering and Biochemistry
- Co-founder of Gevo, Inc. (2005)

Frances Arnold

- Awards and recognition

  2011, Charles Stark Draper Prize

  2011, National Medal of Technology and Innovation

  2011, American Academy of Arts and Sciences

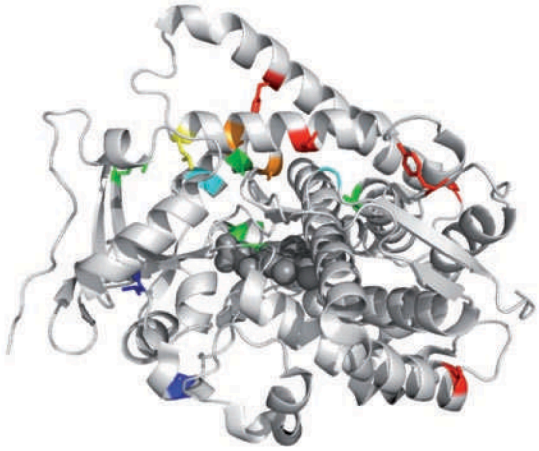
  2008, National Academy of Sciences

  2004, Institute of Medicine

  2000, National Academy of Engineering

- 37 former students and co-workers in academia

The Arnold Lab: Directed Evolution and Biocatalysis



Natural enzyme

- Highly efficient catalyst for a specific biological reaction
- Functions in an aqueous environment at biological pH and temperature
- Activity often finely tuned for metabolism
- Evolved over millennia

Desired industrial catalyst

- Accepts unnatural substrates or carries out reactions unknown in biology
- Catalyzes reactions with high yield and selectivity
- Stable in organic solvents, over broad pH range, and at elevated temperatures
- Easily expressed in host organism

Can we carry out evolution in the lab to create enzymes with desired properties?

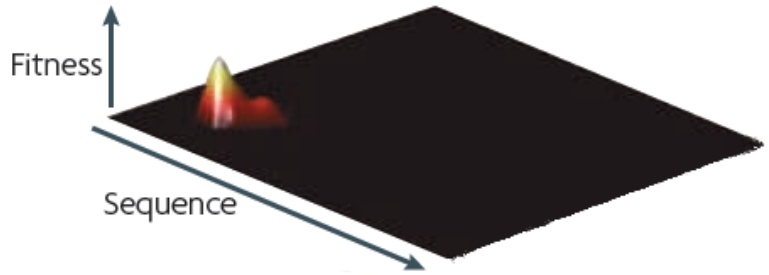
Arnold, F. H. *Chem. Eng. Sci.* **1996**, *51*, 5091.

Challenges of Evolving New Enzymes

- Sequence space is extremely large and cannot be evaluated exhaustively

	no. of amino acid changes	no. of possible variants
For a 300 amino acid enzyme:	0	1
	1	5,700
	2	16,190,850
	3	30,557,530,900
	4	43,109,036,717,100
	5	48,489,044,499,400,000

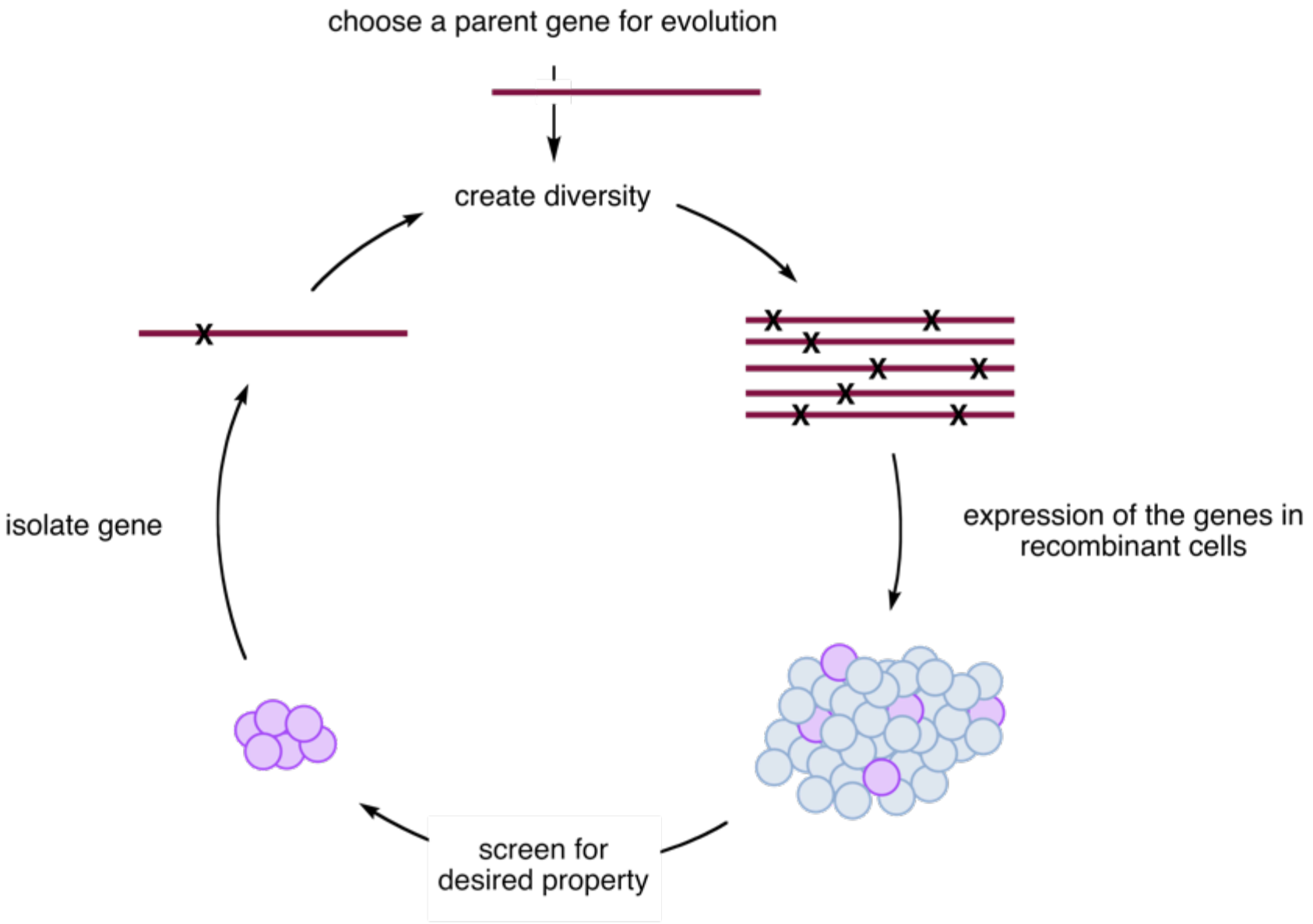
- Most mutations are deleterious or neutral; very few are beneficial
- Directed evolution takes the form of "random walks" across a fitness landscape



- Beneficial mutations, once discovered, are incorporated into the next round of evolution; improved properties accrue slowly over time

Arnold, F. H. *Chem. Eng. Sci.* **1996**, *51*, 5091.
Romero, P. A.; Arnold, F. H. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 866.

Workflow of Directed Evolution



May, O.; Voigt, C. A.; Arnold, F. H. In: *Enzyme Catalysis in Organic Synthesis*; Drauz, K. H.; Waldmann, H., Eds. Wiley-VCH. **2000**, *1*, 95.

Creating Diversity: Tools for Directed Evolution

site-directed mutagenesis

random mutagenesis

random recombination

SCHEMA recombination

Creating Diversity: Tools for Directed Evolution

site-directed mutagenesis

random mutagenesis

random recombination

SCHEMA recombination

Specific amino acids are mutated based on a rational design.

- Requires structural information.
- Requires a hypothesis about how a given amino acid contributes to structure, stability, catalysis, etc.

Creating Diversity: Tools for Directed Evolution

site-directed mutagenesis

Specific amino acids are mutated based on a rational design.

random mutagenesis

Single amino acid mutations are introduced at random throughout a gene.

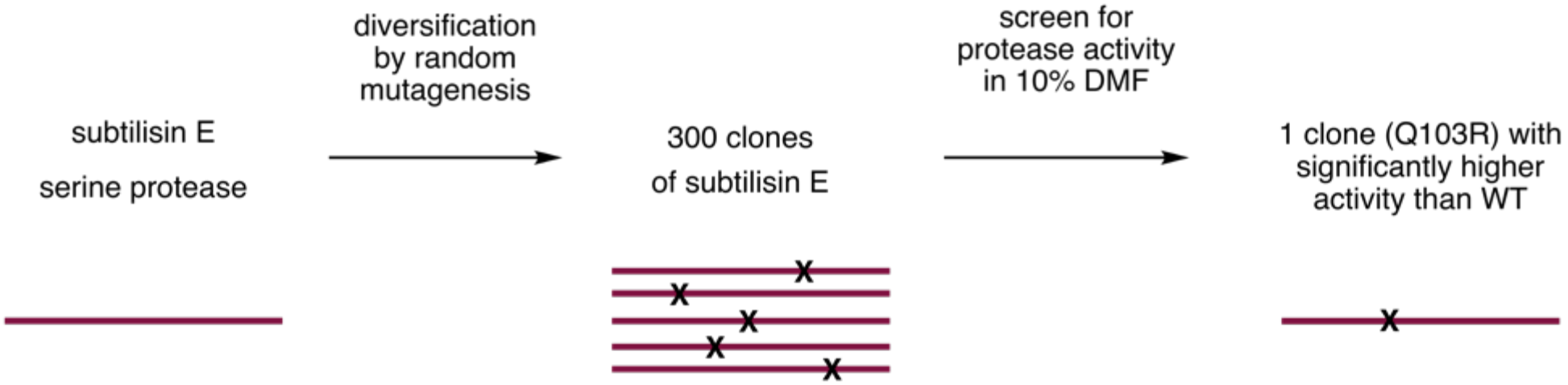
- Error-prone PCR introduces base substitutions at random throughout a DNA sequence at a controllable rate.
- No structural or mechanistic knowledge required.

random recombination

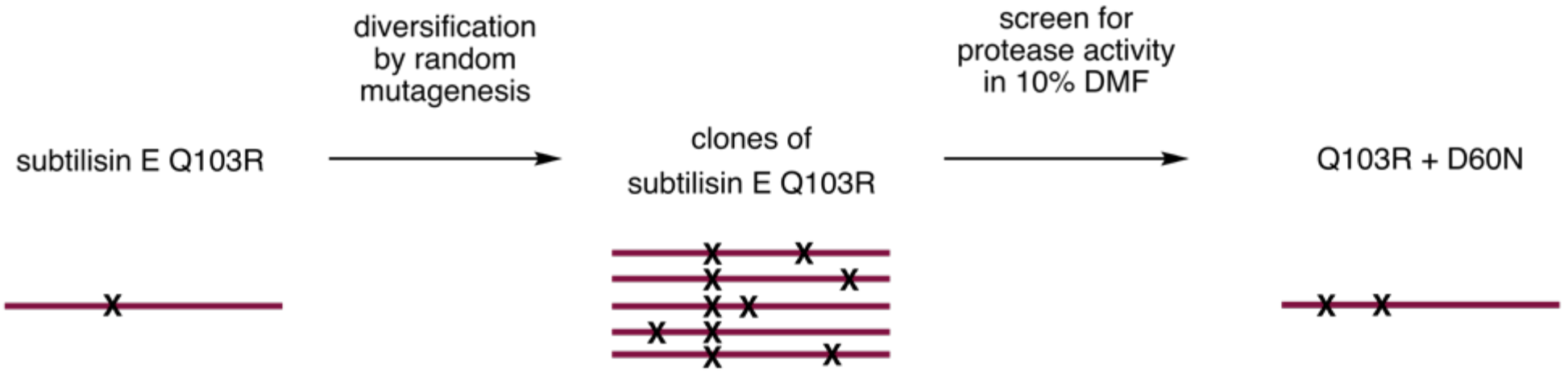
SCHEMA recombination

Improving Enzyme Activity in Non-Aqueous Solvents

- Can directed evolution be used to improve the activity of an enzyme in non-aqueous solvents?



- Mutant identified in first round becomes the starting point for the next round of mutagenesis and screening

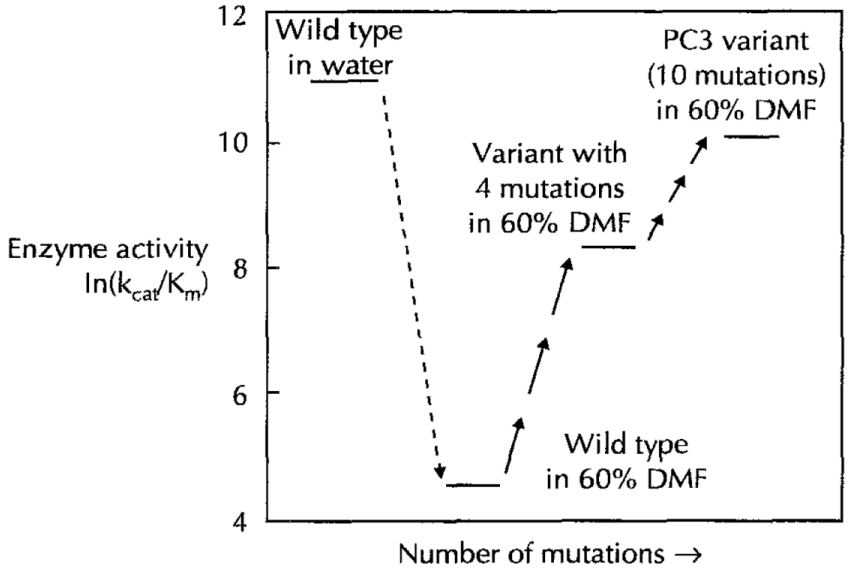


Chen, K.; Arnold, F. H. *Bio/Technology.* **1991**, *9*, 1073.

Improving Enzyme Activity in Non-Aqueous Solvents

- Multiple rounds of screening with progressively higher concentrations of DMF ultimately identify variant PC3
- PC3 hydrolyzes a peptide substrate 256 times more efficiently than wild-type subtilisin in 60% DMF

	k_{cat}/K_M (M^{-1} s^{-1} x 10^{-3})		
	0% DMF	20% DMF	40% DMF
WT	38	1.4	0.16
3M	360	44	6.8
PC3	274	99	21



evolutionary pathway from wild type subtilisin E to PC3

- PC3 differs from wild type by 10 amino acids
  - all ten substitutions clustered on one side of the enzyme, near active site and substrate binding pocket
  - substitutions are located in loops that connect secondary structure elements
- Related studies have evolved subtilisin variants having improved stability at high temperatures

Chen, K.; Arnold, F. H. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5618.

Creating Diversity: Tools for Directed Evolution

site-directed mutagenesis

Specific amino acids are mutated based on a rational design.

random mutagenesis

Single amino acid mutations are introduced at random throughout a gene.

random recombination

SCHEMA recombination

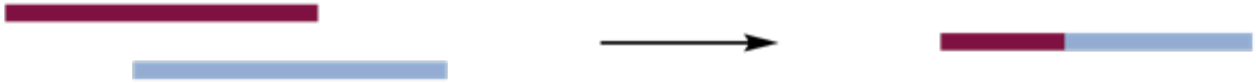
Creating Diversity: Tools for Directed Evolution

site-directed mutagenesis	Specific amino acids are mutated based on a rational design.
random mutagenesis	Single amino acid mutations are introduced at random throughout a gene.
random recombination	Homologous genes are shuffled to give sequences which differ by many amino acids.
SCHEMA recombination	

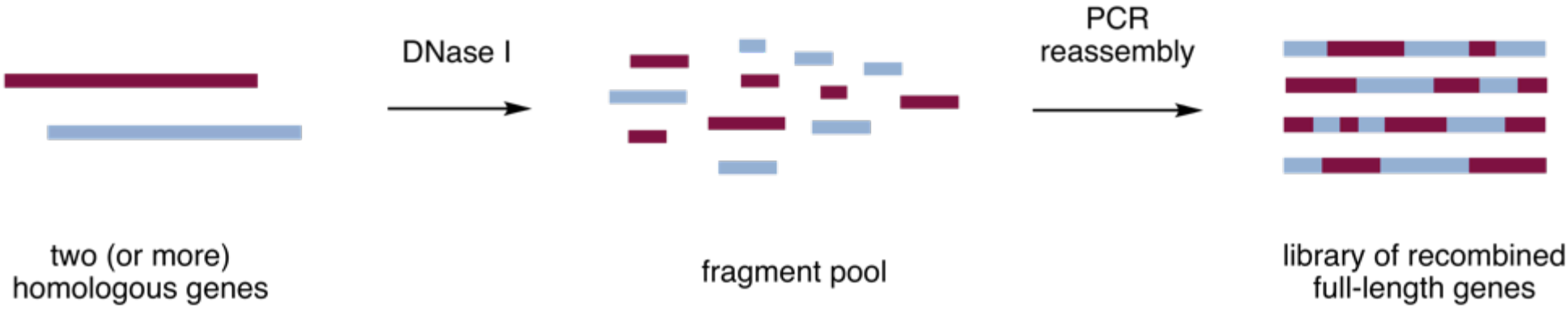
Random Recombination

- Mutagenesis is "asexual" – one gene is altered at a single specific location

- Recombination is "sexual" – two homologous genes are mixed



- Random recombination may be achieved by DNA shuffling

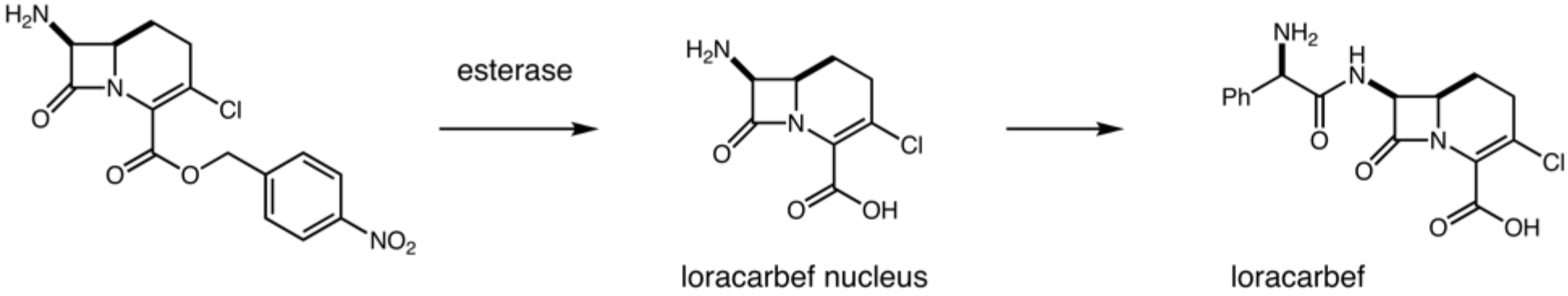


- Crossover occurs at points of high sequence fidelity; approach generates highly diverse sequence libraries but conserves function

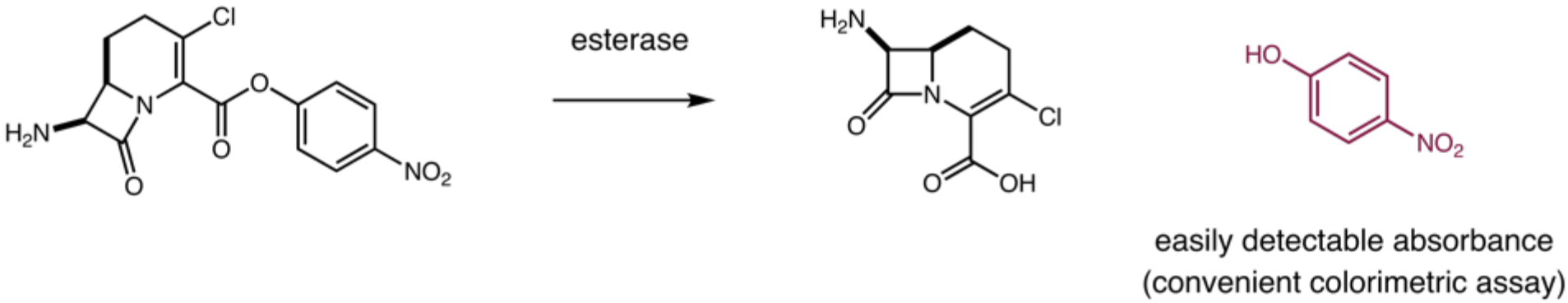
Stemmer, W. P. C. *Nature* **1994**, *340*, 389.

Evolution of a p-Nitrobenzyl Esterase

- Aimed to evolve a *p*-nitrobenzyl esterase for the synthesis of the antibiotic loracarbef
- Substrate is only sparingly soluble in water; enzyme must function in mixed aqueous/organic solvent



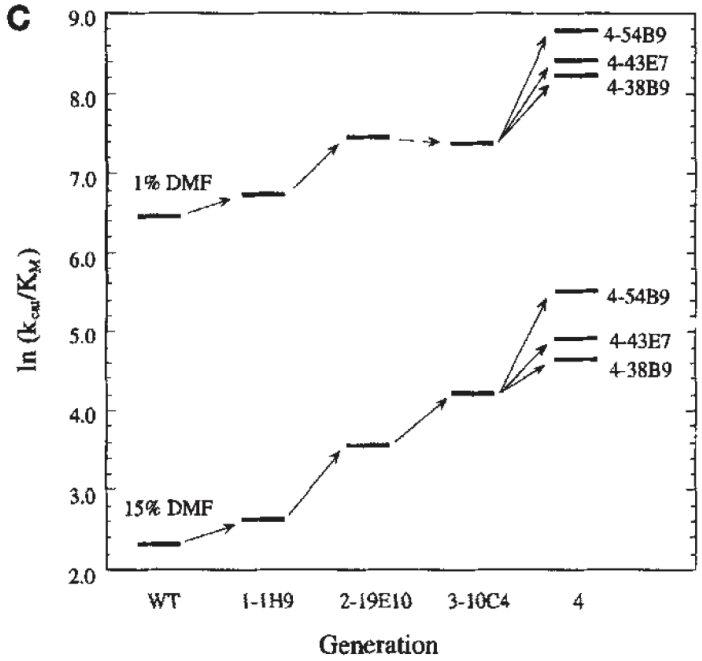
- *p*-Nitrophenyl substrate analog was used for screening to allow for a rapid colorimetric assay

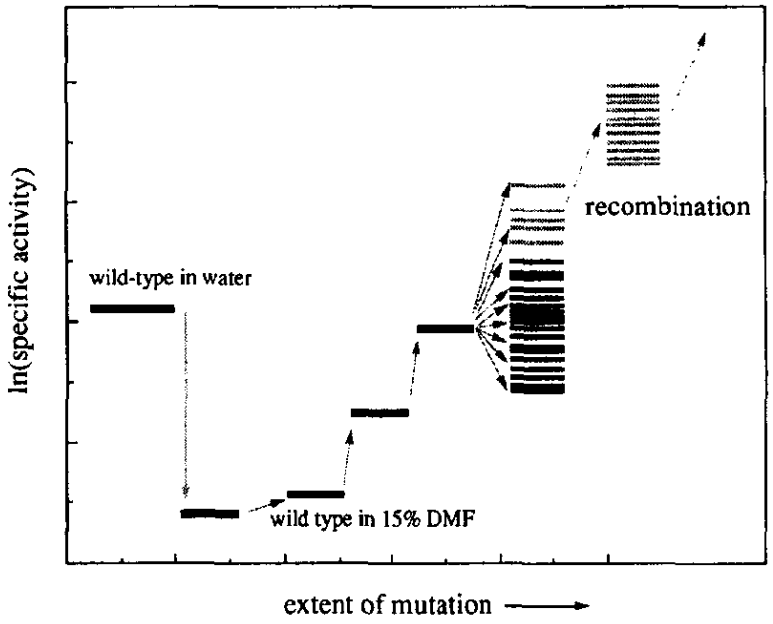


Moore, J. C.; Arnold, F. H. *Nat. Biotech.* **1996**, *14*, 458.

Evolution of a p-Nitrobenzyl Esterase

- An enzyme with activity as a *p*-nitrobenzyl esterase had been identified from screening of bacterial cultures
- Directed evolution strategy involved four generations of random mutagenesis and one recombination step
- Recombination step generated a library from the five best mutants from generation 4





enzyme activities on *p*-nitrobenzyl substrate

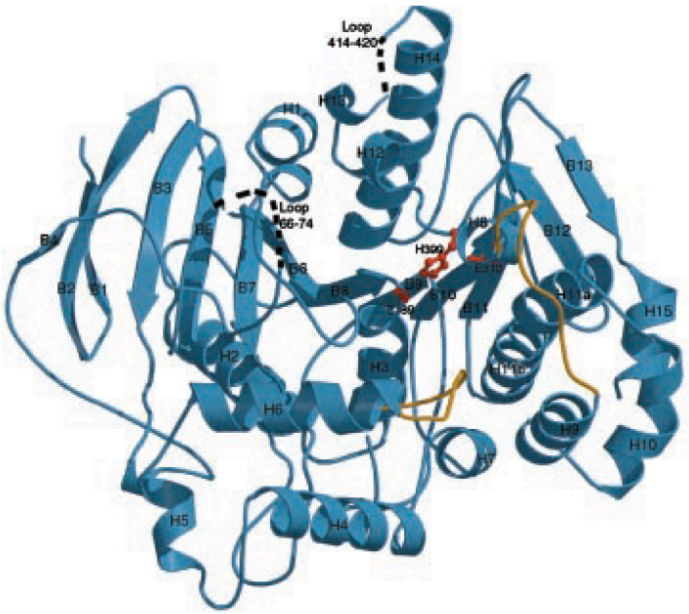
- Approach identified a variant with greater than 30-fold improved *p*-nitrobenzyl esterase activity in 15% DMF

Moore, J. C.; Arnold, F. H. *Nat. Biotech.* **1996**, *14*, 458.

Arnold, F. H. *Chem. Eng. Sci.* **1996**, *51*, 5091.

Structural Basis for Improved Esterase Activity

- Crystal structures obtained for wild type *p*-nitrobenzyl esterase as well as the evolved 5-6c8 mutant
- None of the seven amino acid mutations are in the active site (range from 11.4 to 21.8 Å distant from the catalytic serine 189)
- Mutations instead reorient loops and helices which make up the active site cavity
- Impact of these subtitutions would have been very difficult to predict!



wild type structure

5-6c8 mutant structure

Spiller, B.; Gershenson, A.; Arnold, F. H.; Stevens, R. C. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12305.

Creating Diversity: Tools for Directed Evolution

site-directed mutagenesis	Specific amino acids are mutated based on a rational design.
random mutagenesis	Single amino acid mutations are introduced at random throughout a gene.
random recombination	Homologous genes are shuffled to give sequences which differ by many amino acids.
SCHEMA recombination	

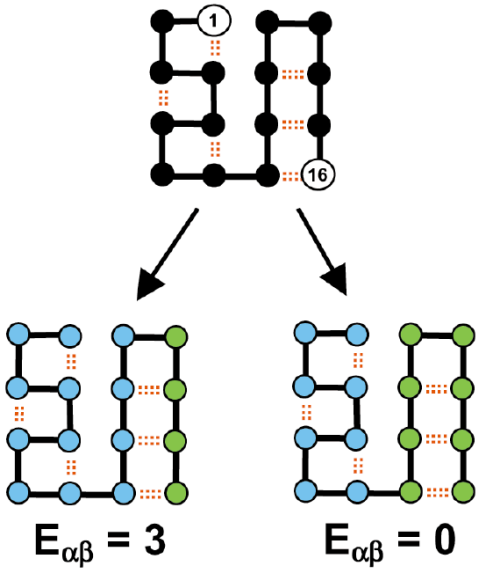
Tools for Directed Evolution

site-directed mutagenesis	Specific amino acids are mutated based on a mechanistic hypothesis.
random mutagenesis	Single amino acid mutations are introduced at random throughout a gene.
random recombination	Homologous genes are shuffled to give sequences which differ by many amino acids.
SCHEMA recombination	Computation predicts sites for recombination likely to retain overall protein structure.

SCHEMA Recombination

- Random recombination is not ideal for combining distantly related proteins, can lead to a high fraction of unfolded or incorrectly folded chimeras

- SCHEMA is an algorithm that identifies the number of structural contacts that are disrupted when portions of a sequence are inherited from different parents



E = the number of structural contacts disrupted upon recombination

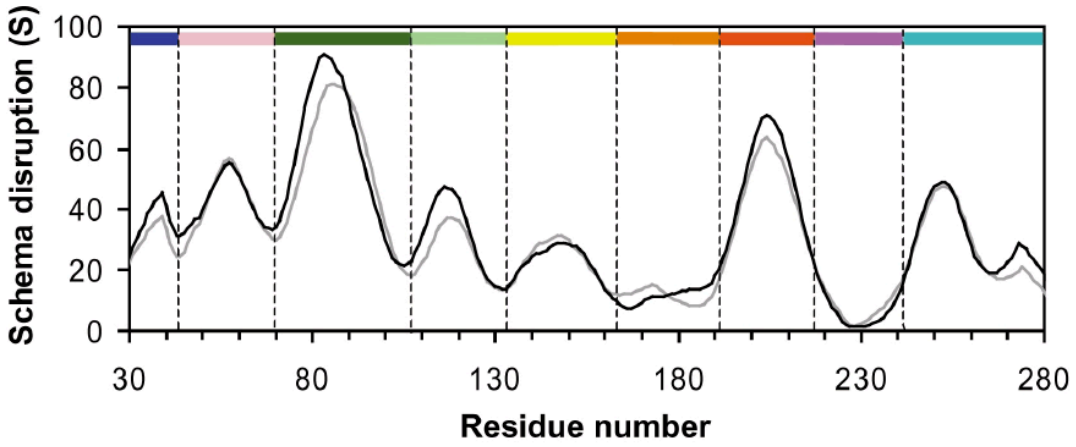
- two residues are defined as interacting if the distance between any of their (non-hydrogen) atoms is 4.5 Å

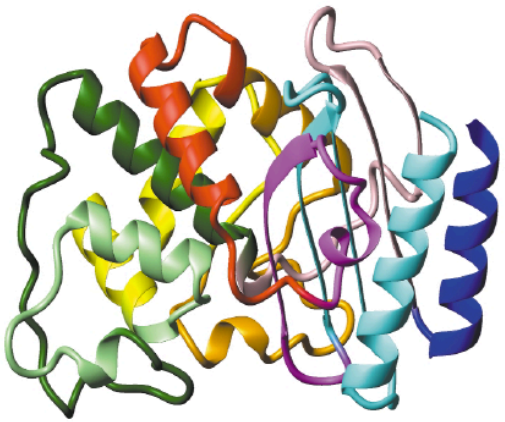
- Aim is creation of chimeras that are highly mutated but also have a high likelihood of folding into the parent structure

Voigt, C. A.; Martinez, C.; Wang, Z.-G.; Mayo, S. L.; Arnold, F. H. *Nat. Struct. Biol.* **2002**, *9*, 553.

SCHEMA Recombination

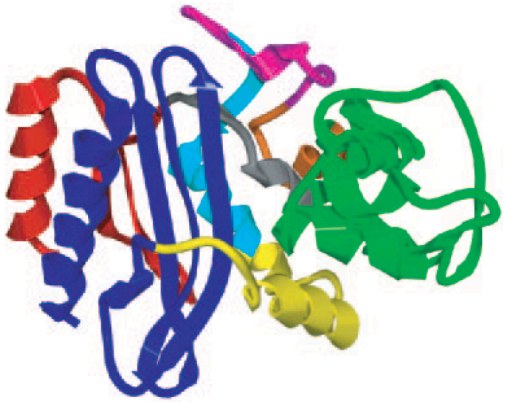
- The schema profile S is generated by calculating how many interactions are disrupted when a crossover is made at a certain position
- If S is high, that residue is involved in a compact structural module with many residue-residue contacts
- Crossovers for recombination should occur at the minima of schema profiles to preserve the maximum number of internal interactions
- Individual elements identified by SCHEMA are not obvious based on secondary or domain structure
  - crossovers are frequently predicted to be optimal in the center of α-helices
  - crossovers in loops can be highly disruptive if they divide interacting units of secondary structure
  - individual blocks, however, combine to form larger structural elements that coincide with protein domains





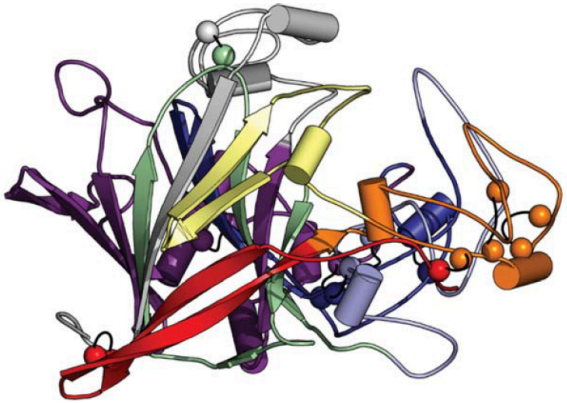
Voigt, C. A.; Martinez, C.; Wang, Z.-G.; Mayo, S. L.; Arnold, F. H. *Nat. Struct. Biol.* **2002**, *9*, 553.

SCHEMA Recombination Applied to Diverse Enzyme Families



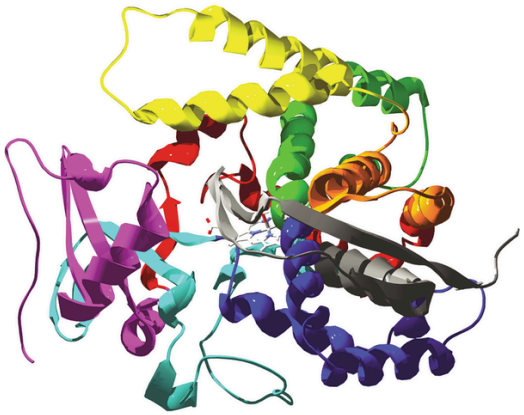
β-lactamases

20% of chimeras retain activity, despite low shared sequence identity of parents (34-42%)



fungal cellulases

applied to generation of thermostable variants



cytochrome P450s

~3,000 variants prepared

differ from any known sequence by an average of 72 residues

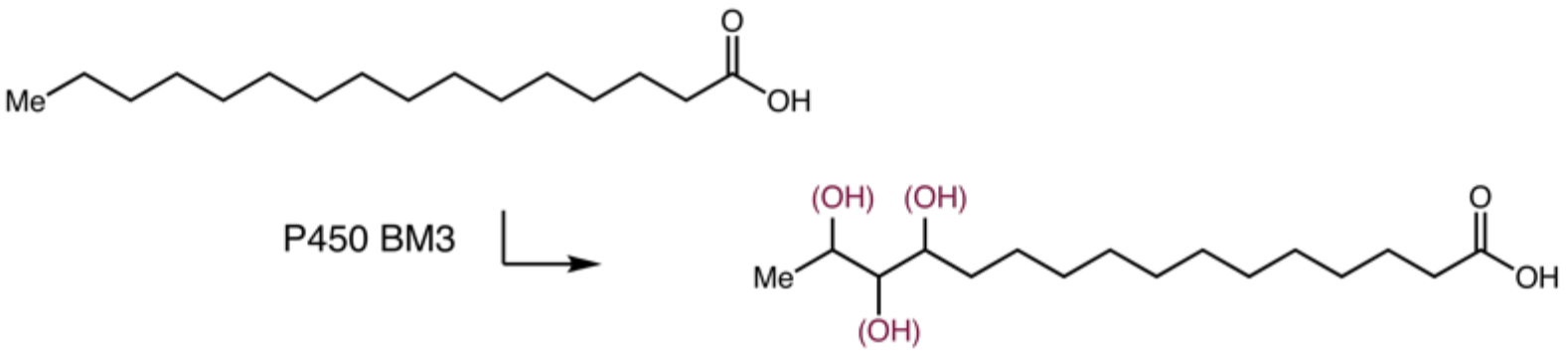
Meyer, M. M.; Hochrein, L.; Arnold, F. H. *Protein Engineering, Design & Selection* **2006**, *19*, 563.

Heinzelman, P.; Komor, R.; Kanaan, A.; Romero, P.; Yu, X.; Mohler, S.; Snow, C.; Arnold, F. H. *Protein Engineering, Design & Selection* **2010**, *23*, 871.

Otey, C. R.; Landwehr, M.; Endelman, J. B.; Hiraga, K.; Bloom, J. D.; Arnold, F. H. *PLoS Biology* **2006**, *4*, 789.

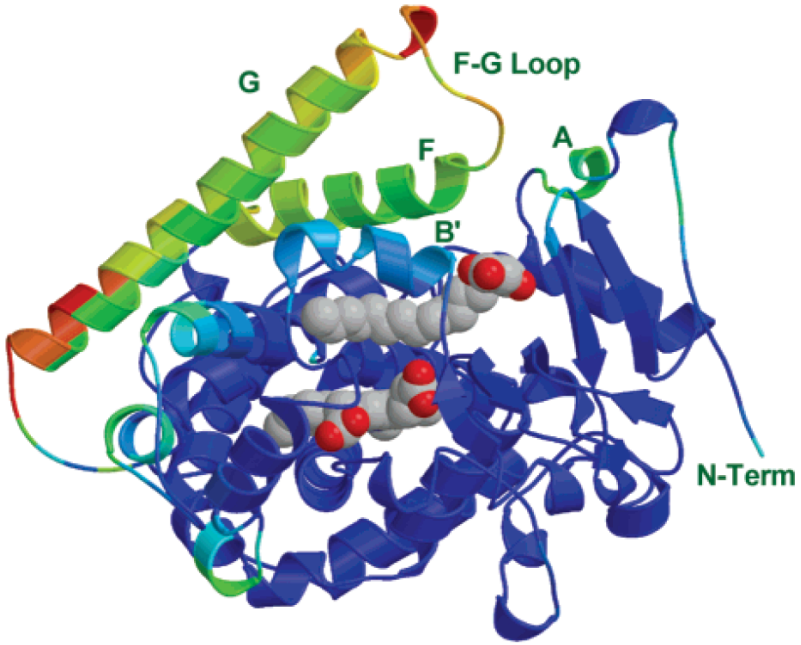
Cytochrome P450 Monooxygenases

- More than 7,500 P450 variants identified in all kingdoms of life
- Catalyze the hydroxylation of C–H bonds by inserting an oxygen atom derived from O_2
- Cytochrome P450 BM3 from *Bacillus megaterium* is widely studied for chemical synthesis
  - Heme domain (64 kDa) and flavin reductase domains (54 kDa) are fused in a single peptide chain
  - Protein is water-soluble and readily overexpressed in host organisms
  - Natural substrates are C12–C18 fatty acids, which undergo subterminal hydroxylation

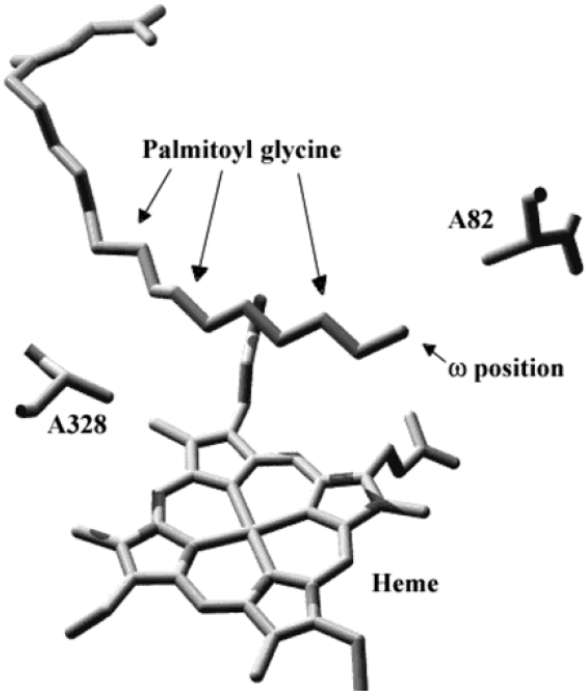


Farinas, E. T.; Schwaneberg, U.; Glieder, A.; Arnold, F. H. *Adv. Synth. Catal.* **2001**, *343*, 601.

Cytochrome P450 Monooxygenases



crystal structure of P450 BM3 heme domain
in complex with *N*-palmitoylglycine



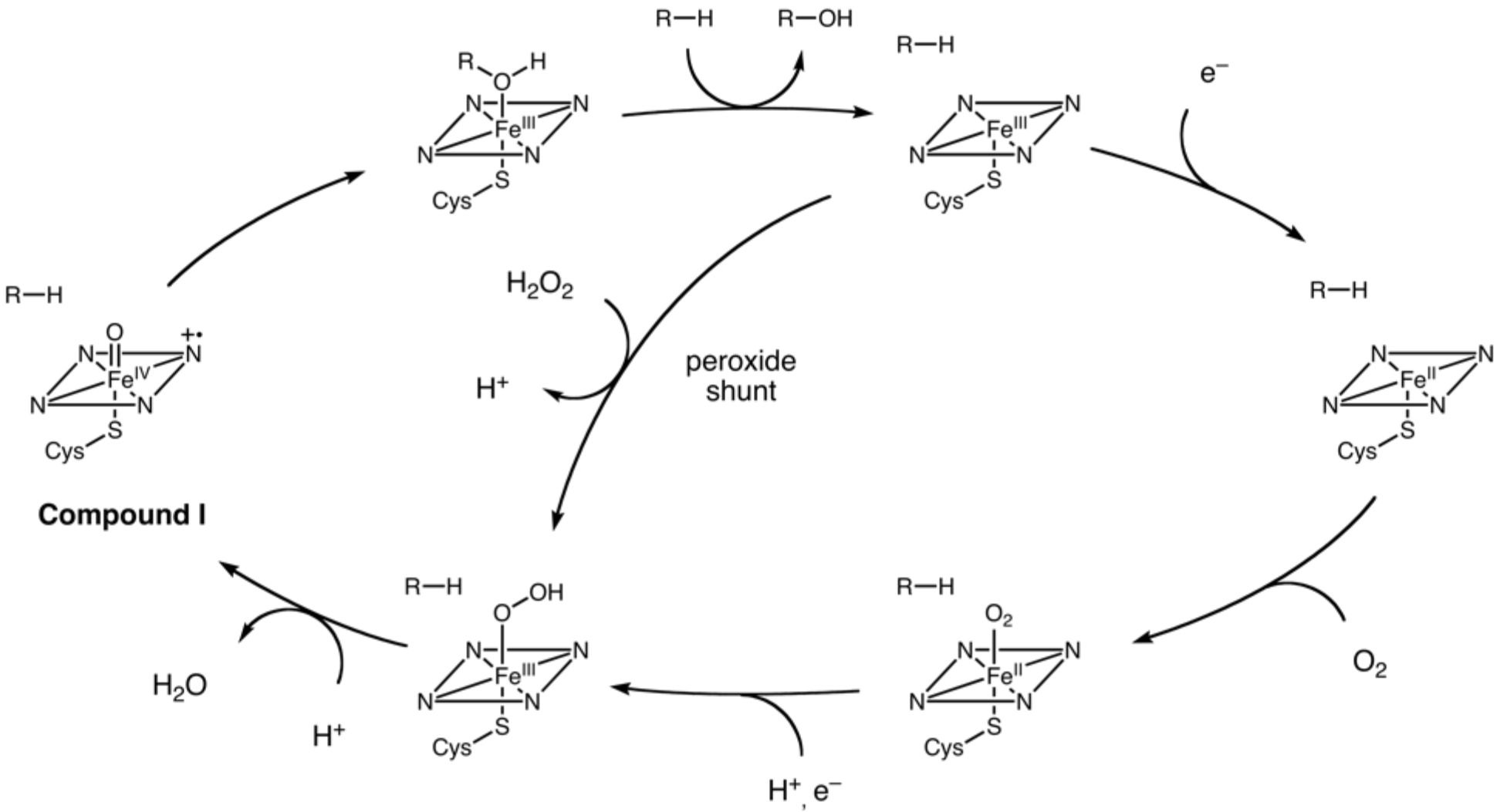
positioning of *N*-palmitoylglycine
in the active site

Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A. *Biochemistry* **2001**, *40,* 13456.

Peters, M. W.; Meinhold, P.; Glieder, A.; Arnold, F. H. *J. Am. Chem. Soc.* **2003**, *125,* 13442.

Mechanism of Cytochrome P450 Hydroxylation

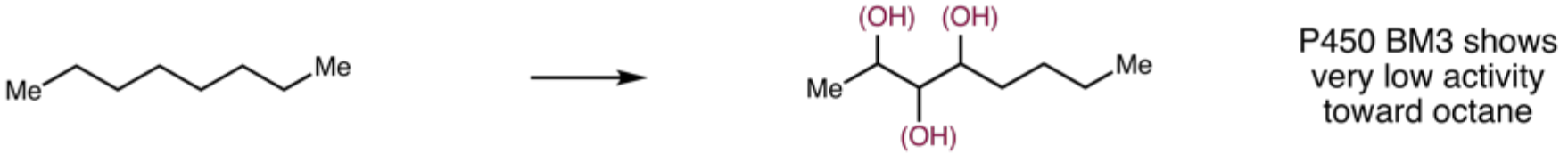
- Formation of an oxoiron(IV)porphyrin cation radical enables insertion into aliphatic C–H bonds



Groves, J. T. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3569.

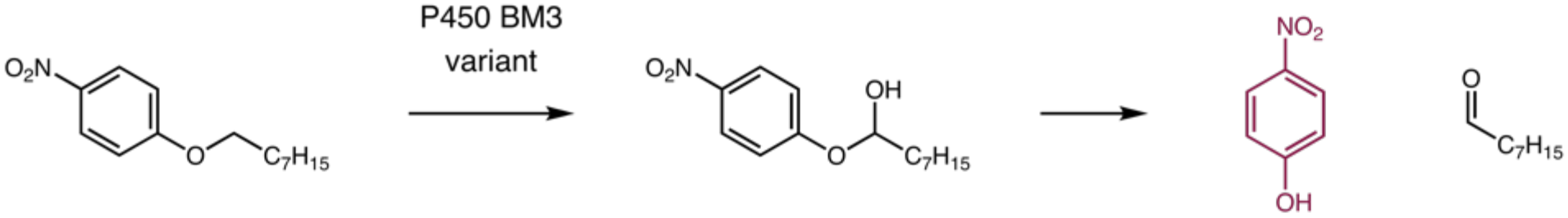
Hydroxylation of Linear Alkanes

- Desired reaction: direct hydroxylation of C–H bonds in unactivated linear alkanes



P450 BM3 shows very low activity toward octane

- Screening approach employs *p*-nitrophenoxyoctane as surrogate for octane



detectable by colorimetric assay

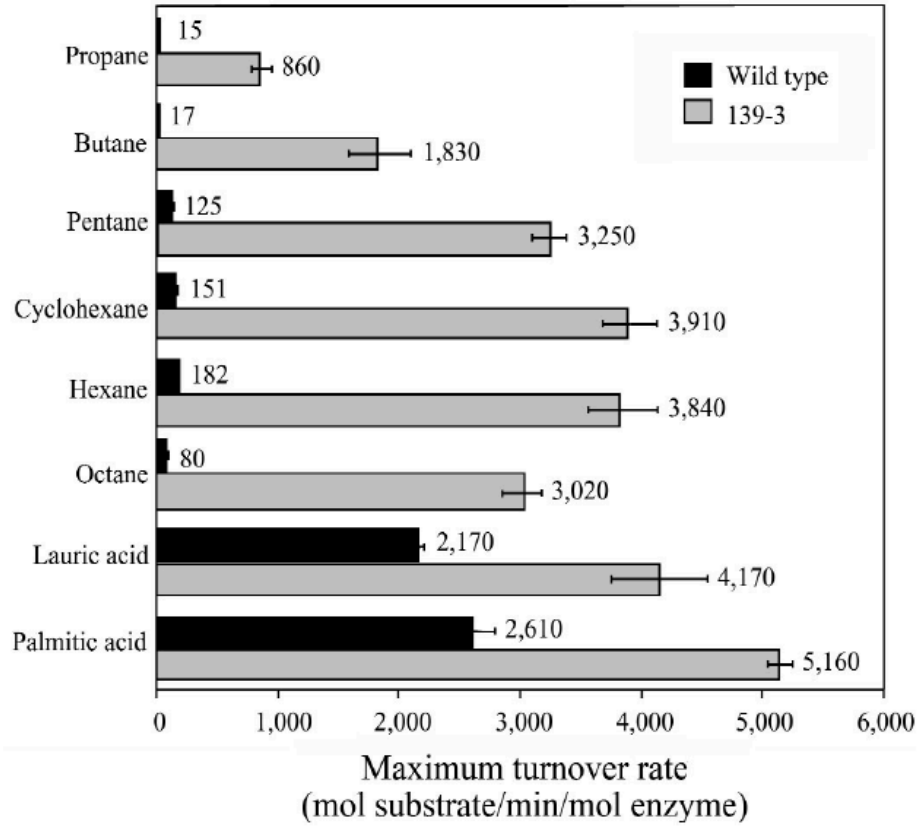
- Variants additionally evaluated on octane using an assay for NADPH consumption

Farinas, E. T.; Schwaneberg, U.; Glieder, A.; Arnold, F. H. *Adv. Synth. Catal.* **2001**, *343*, 601.

Glieder, A.; Farinas, E.; Arnold, F. H. *Nature Biotechnology* **2002**, *20*, 1135.

Hydroxylation of Linear Alkanes

- Directed evolution via sequential rounds of random mutagenesis and recombination yields variant 139-3
- P450 139-3 differs from WT by 11 amino acid substitutions, possesses very high alkane hydroxylation activities

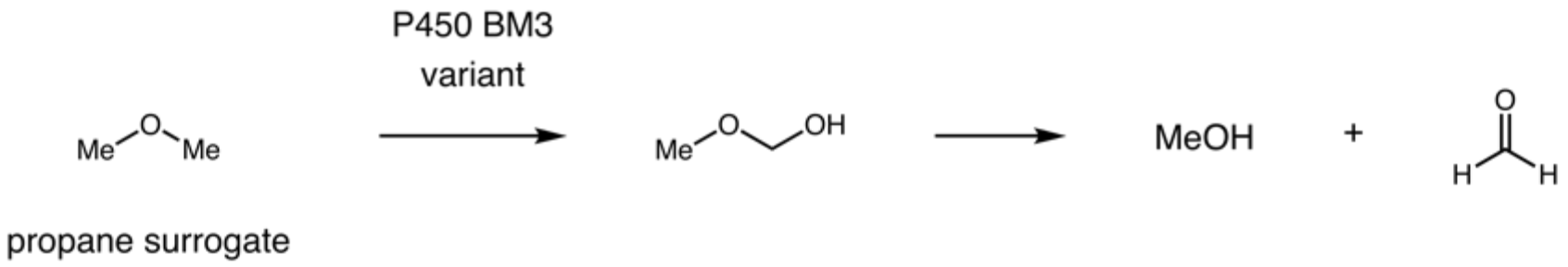


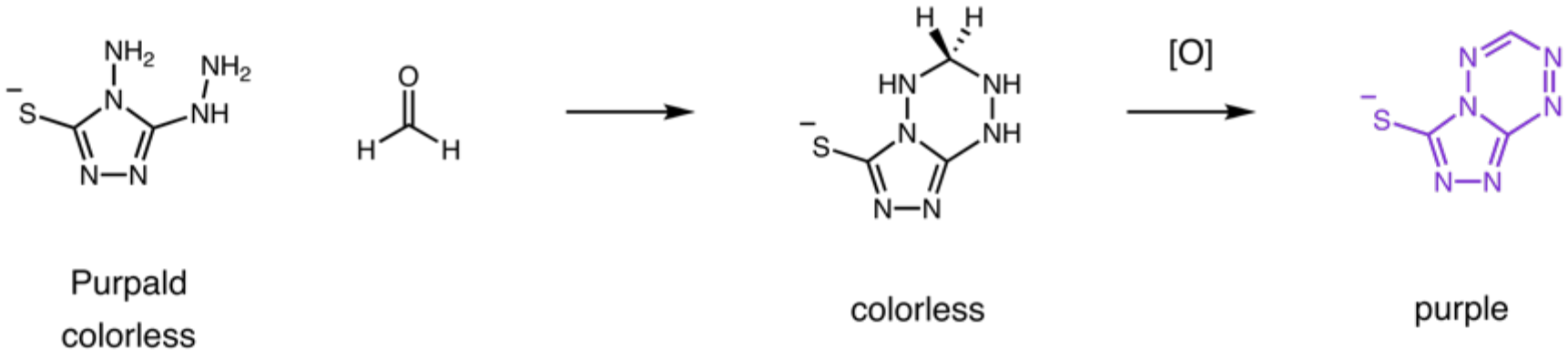
products	product distribution upon 139-3 hydroxylation
2-octanol	66%
3-octanol	32%
4-octanol	2%
2-hexanol	19%
3-hexanol	81%
2-butanol	100%
2-propanol	100%

Glieder, A.; Farinas, E.; Arnold, F. H. *Nature Biotechnology* **2002**, *20*, 1135.

Hydroxylation of Short Chain Alkanes

- Hydroxylation of alkanes found in natural gas (methane, ethane) would provide liquid alcohols for use as fuels
- Screening on progressively shorter alkanes used as strategy to identify active hydroxylases
- Dimethylether is employed in screening as a surrogate for propane





Peters, M. W.; Meinhold, P.; Glieder, A.; Arnold, F. H. *J. Am. Chem. Soc.* **2003**, *125*, 13442.

Hydroxylation of Short Chain Alkanes

- Rounds of directed evolution identify variant 35-E11 capable of hydroxylating propane and ethane

	octane		propane		ethane	
enzyme	rate	TTN	rate	TTN	rate	TTN
BM3 wild type	30	150	--	--	--	--
35-E11	420	8000	210	6000	0.4	250

rate = mol product per min per mol enzyme

TTN = total turnover number, mol product per mol enzyme

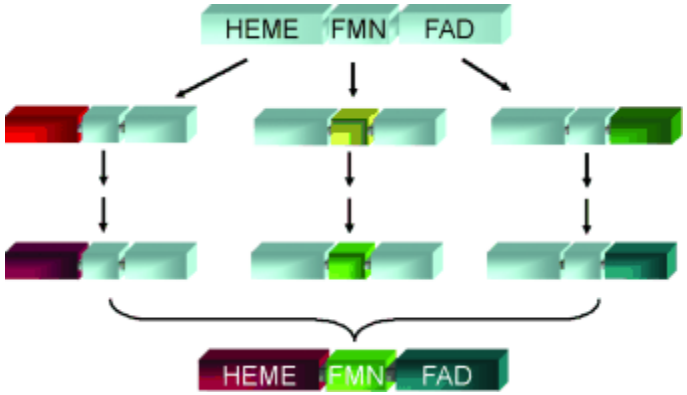
- 35-E11 developed only by mutating the heme domain; result is poor efficiency of coupling with NADPH consumption: 17.4% for propane, 0.01% for ethane – significant limitation for *in vivo* application

Meinhold, P.; Peters, M. W.; Chen, M. M. Y.; Takahashi, K.; Arnold, F. H. *ChemBioChem* **2005**, *6*, 1765.

Fasan, R.; Chen, M. M.; Crook, N. C.; Arnold, F. H. *Angew. Chem. Int. Ed.* **2007**, *46*, 8414.

Hydroxylation of Short Chain Alkanes

- Domain-based directed evolution strategy: independent evolution of heme and reductase domains followed by recombination of beneficial mutations in a final step



– create libraries each having one domain mutated

– evolve and screen for improved variants

– combine beneficial mutations in each domain

- Leads to $P450_{PMO}R2$: provides 45,000 turnovers on propane, produces 2- and 1-propanol in a 9:1 ratio, coupling efficiency is 98.2%

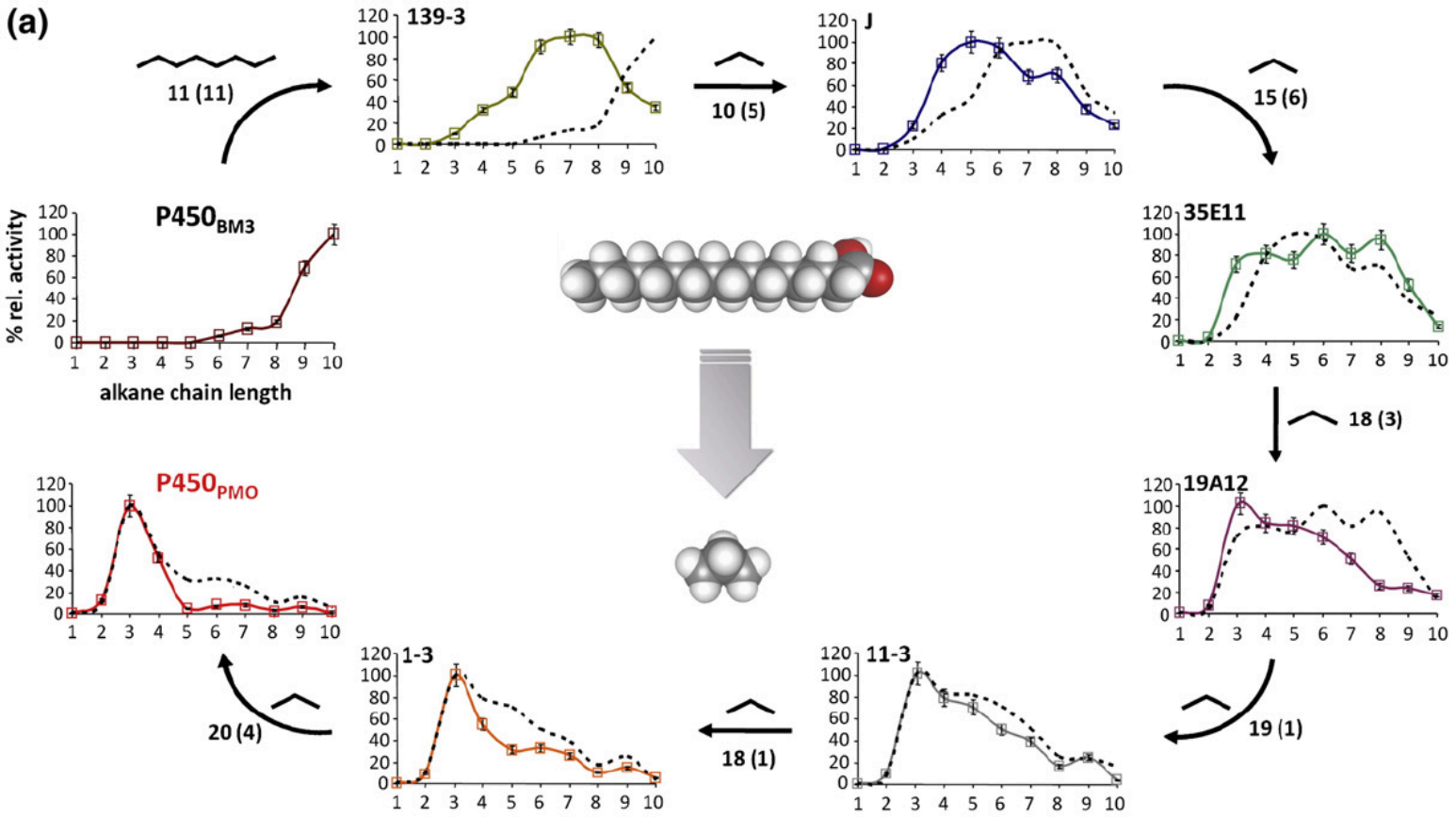
- Overall process evolved an enzyme capable of efficiently hydroxylating a substrate on which the starting enzyme (BM3) has no detectable activity

Fasan, R.; Chen, M. M.; Crook, N. C.; Arnold, F. H. *Angew. Chem. Int. Ed.* **2007**, *46*, 8414.

"You Get What You Screen For"

- Directed evolution provides insights into natural evolution: evolutionary path may be traced through a lineage of known intermediate variants

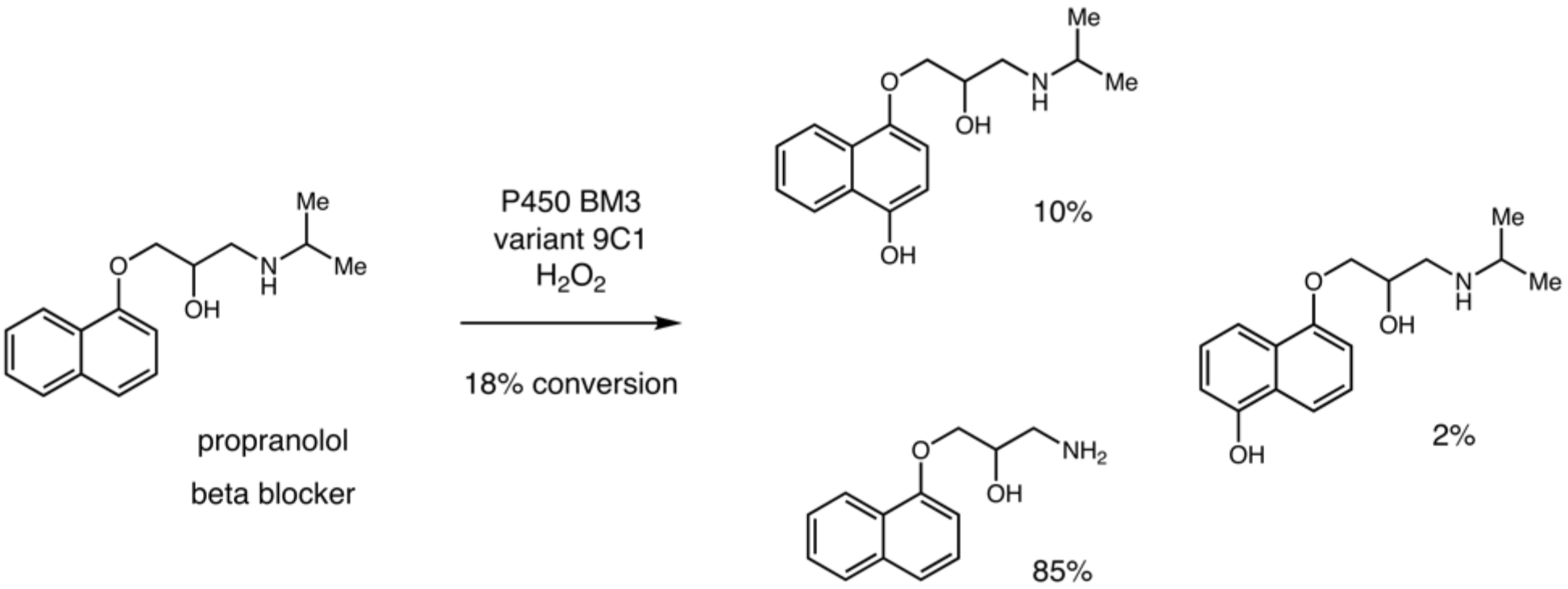
- Evolution of a propane monooxygenase proceeds with a broadening of enzyme specificity followed by a narrowing of activity: enzyme becomes promiscuous before becoming specific



Fasan, R.; Meharenna, Y. T.; Snow, C. D.; Poulos, T. L.; Arnold, F. H. *J. Mol. Biol.* **2008**, *383*, 1069.

Enzymatic Hydroxylation for Synthesis of Drug Metabolites

- Reaction with engineered P450s provides a method for identification and synthesis of likely human metabolites directly from a medicinal compound
- P450 BM3 evolved to react via the "peroxide shunt" is a convenient system for preparative metabolite sythesis

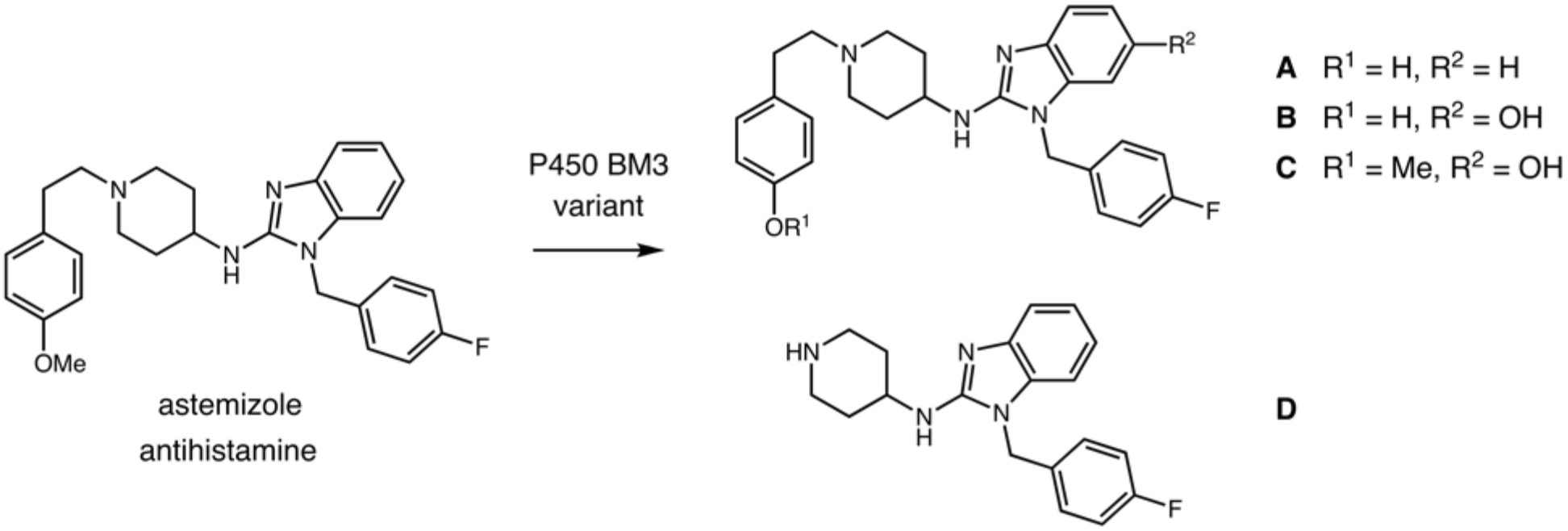


- variant 9C1 produces the three major metabolites produced by human P450s CYP2D6 and CYP1A2

Otey, C. R.; Bandara, G.; Lalonde, J.; Takahashi, K.; Arnold, F. H. *Biotechnology and Bioengineering* **2005**, *93*, 494.

Enzymatic Hydroxylation for Synthesis of Drug Metabolites

- Various evolved P450s provide different selectivities for astemizole metabolites

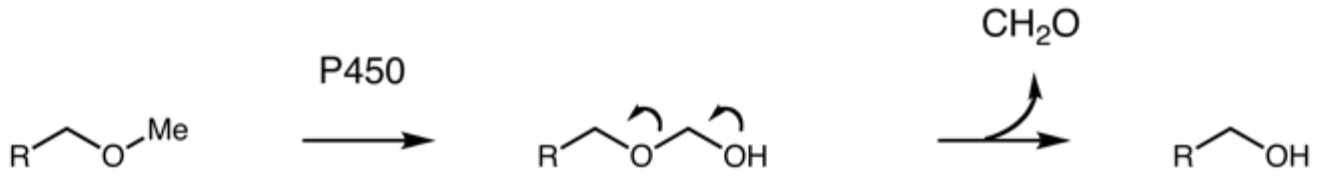


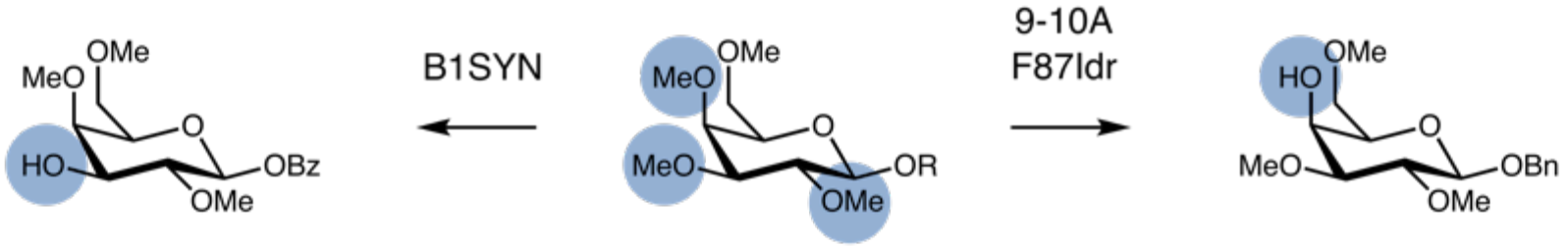
variant	conversion (%)	A	B	C	D
41-5B	5	100	--	--	--
9-10A A78F	21	48	--	38	--
9-10A F87L	36	--	3	88	--
21313311	10	20	--	--	40

Sawayama, A. M.; Chen, M. M. Y.; Kulanthaivel, P.; Kuo, M.-S.; Hemmerle, H.; Arnold, F. H. *Chem. Eur. J.* **2009**, *15*, 11723.

Applications of Enzymatic Hydroxylation: Carbohydrate Synthesis

- Regioselective deprotection of permethylated monosaccharides furnishes carbohydrate derivatives



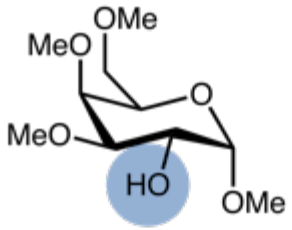


98% yield, 100% selectivity

galactose derivative

64% yield, 80% selectivity

5E5



48% yield, 93% selectivity

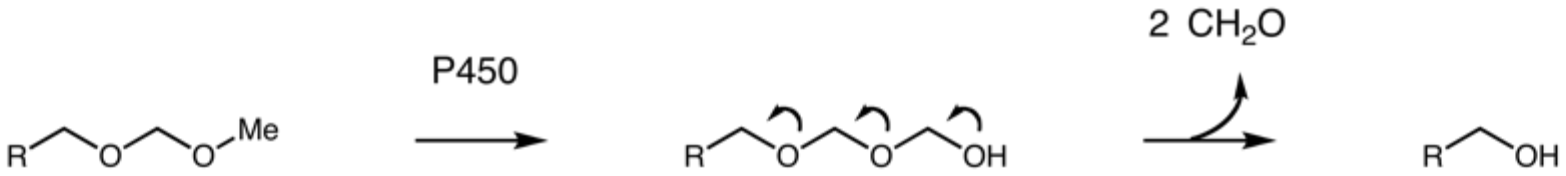
0.05 - 0.1 mol% enzyme loadings

glucose, mannose derivatives
also undergo selective deprotection

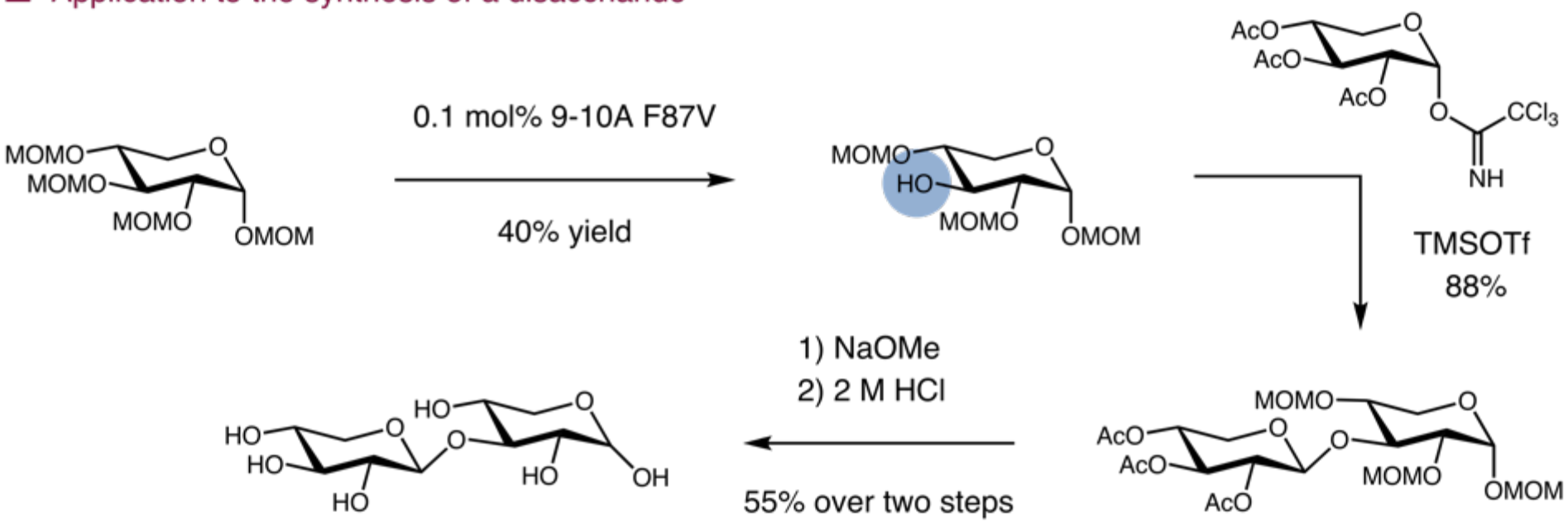
Lewis, J. C.; Bastian, S.; Bennett, C. S.; Fu, Y.; Mitsuda, Y.; Chen, M. M.; Greenberg, W. A.; Wong, C.-H.; Arnold, F. H. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 16550.

Applications of Enzymatic Hydroxylation: Carbohydrate Synthesis

- Regioselective deprotection approach is also amenable to the removal of MOM ethers



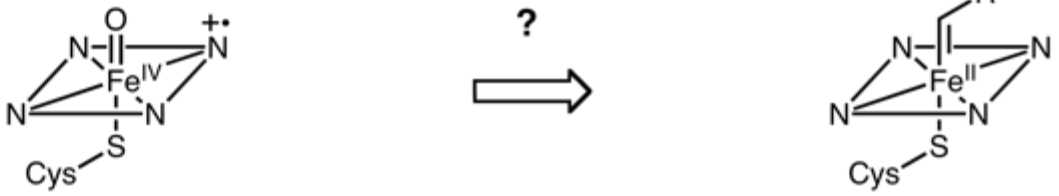
- Application to the synthesis of a disaccharide



Lewis, J. C.; Bastian, S.; Bennett, C. S.; Fu, Y.; Mitsuda, Y.; Chen, M. M.; Greenberg, W. A.; Wong, C.-H.; Arnold, F. H. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 16550.

Olefin Cyclopropanation with Cyctochrome P450s

Can an iron carbene be generated
at the active site of a cytochrome P450?



Compound I: iron oxene

intermediate in cytochrome
P450 hydroxylation

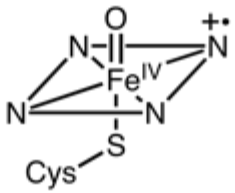
iron carbene

no biological counterpart

Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. *Science* **2013**, *339*, 307.

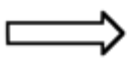
Olefin Cyclopropanation with Cyctochrome P450s

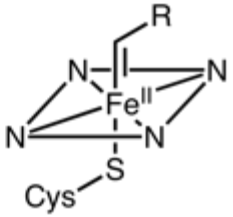
Can an iron carbene be generated at the active site of a cytochrome P450?

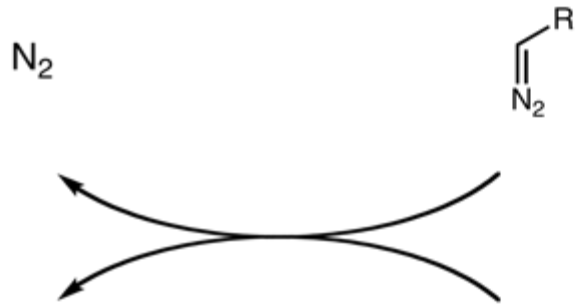


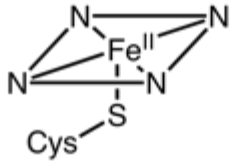
Compound I: iron oxene

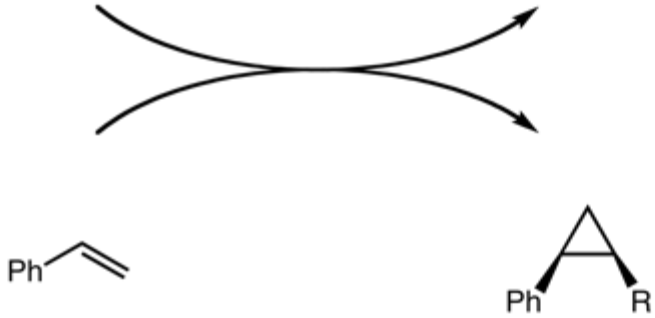
intermediate in cytochrome P450 hydroxylation







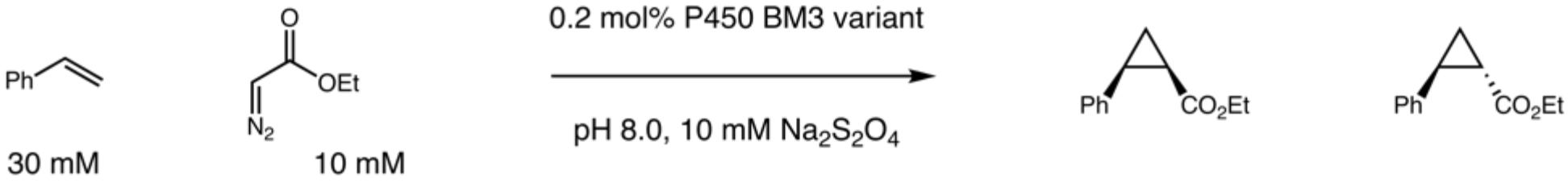




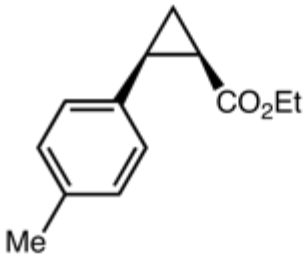
Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. *Science* **2013**, *339*, 307.

Olefin Cyclopropanation with Cytochrome P450s

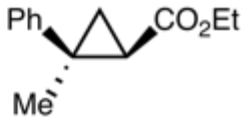
- P450 BM3 variants catalyze the cyclopropanation of styrene with ethyl diazoacetate



enzyme	% yield	TTN	cis:trans	% ee_{cis}	% ee_{trans}
$P450_{BM3}$	1	5	37:63	27	2
$P450_{BM3}$-T268A	65	323	1:99	15	96
BM3-CIS-T438S	59	293	92:8	97	66



BM3-CIS, 25% yield, 228 TTN
78:22 cis:trans, 81% ee_{cis}

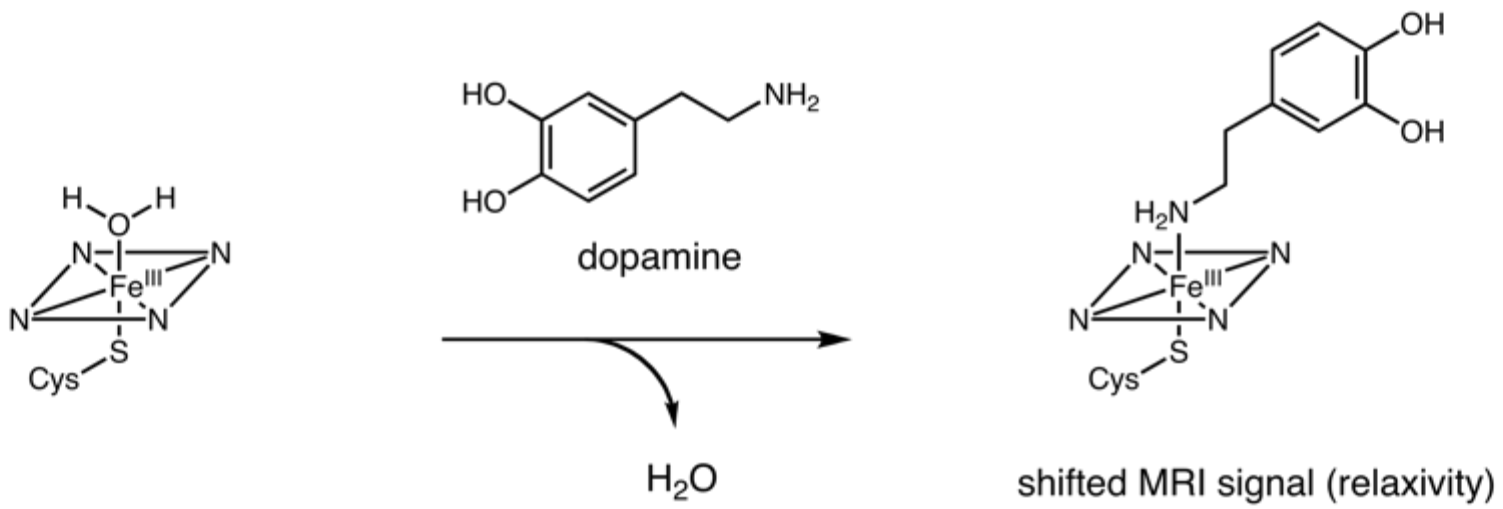


7-11D, 31% yield, 157 TTN
41:49 cis:trans, 42% ee_{cis}

Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. *Science* **2013**, *339*, 307.

Directed Evolution of MRI Contrast Agents

- Cytochrome P450 heme domain used to develop magnetic resonance imaging (MRI) contrast agents



- Paramagnetic iron center produces a signal that is ligand-dependent
- Modulation of signal by varying concentrations of a biological analyte enables *in vivo* imaging
- Directed evolution performed to introduce specificity for binding the neurotransmitters dopamine and serotonin
  - evolved variant BM3h-9D7 with K_d = 1.3 ± 0.1 μM for dopamine and variant BM3h-2G9C6 with K_d = 0.7 ± 0.2 μM for serotonin
  - applied to imaging of dopamine in brains of living rats

Shapiro, M. G.; Westmeyer, G. G.; Romero, P. A.; Szablowski, J. O.; Küster, B.; Shah, A.; Otey, C. R.; Langer, R.; Arnold, F. H.; Jasanoff, A. *Nat. Biotech.* **2010**, *28*, 264.

Brustad, E. M.; Lelyveld, V. S.; Snow, C. D.; Crook, N.; Jung, S. T.; Martinez, F. M.; Scholl, T. J.; Jasanoff, A.; Arnold, F. H. *J. Mol. Biol.* **2012**, *422*, 245.