Document Title

# Engineered Cytochrome c-Catalyzed Lactone-Carbene B-H Insertion.

2019

Synlett : accounts and rapid communications in synthetic organic chemistry

Chen, Kai

Huang, Xiongyi

Zhang, Shuo-Qing

Zhou, Andrew Z

Kan, S B Jennifer

Hong, Xin

Arnold, Frances H

Previous work has demonstrated that variants of a heme protein, Rhodothermus marinus cytochrome c (Rma cyt c), catalyze abiological carbene boron-hydrogen (B-H) bond insertion with high efficiency and selectivity. Here we investigated this carbon-boron bondforming chemistry with cyclic, lactone-based carbenes. Using directed evolution, we obtained a Rma cyt c variant BOR (LAC) that shows high selectivity and efficiency for B-H insertion of 5- and 6-membered lactone carbenes (up to 24,500 total turnovers and 97.1:2.9 enantiomeric ratio). The enzyme shows low activity with a 7-membered lactone carbene. Computational studies revealed a highly twisted geometry of the 7membered lactone carbene intermediate relative to 5- and 6-membered ones. Directed evolution of cytochrome c together with computational characterization of key iron-carbene intermediates has allowed us to expand the scope of enzymatic carbene B-H insertion to produce new lactone-based organoborons.

# A Biocatalytic Platform for Synthesis of Chiral alpha-Trifluoromethylated Organoborons.

2019

ACS central science

Huang, Xiongyi

Garcia-Borras, Marc

Miao, Kun

Kan, S B Jennifer

Zutshi, Arjun

Houk, K N

Arnold, Frances H

There are few biocatalytic transformations that produce fluorine-containing molecules prevalent in modern pharmaceuticals. To expand the scope of biocatalysis for organofluorine synthesis, we have developed an enzymatic platform for highly enantioselective carbene B-H bond insertion to yield versatile alpha-trifluoromethylated (alpha-CF3) organoborons, an important class of organofluorine molecules that contain stereogenic centers bearing both CF3 and boron groups. In contrast to current "carbene transferase" enzymes that use a limited set of simple diazo compounds as carbene precursors, this system based on Rhodothermus marinus cytochrome c (Rma cyt c) can accept a broad range of trifluorodiazo alkanes and deliver versatile chiral alpha-CF3 organoborons with total turnovers up to 2870 and enantiomeric ratios up to 98.5:1.5. Computational modeling reveals that this broad diazo scope is enabled by an active-site environment that directs the alkyl substituent on the heme CF3-carbene intermediate toward the solvent-exposed face, thereby allowing the protein to accommodate diazo compounds with diverse structural features.

# Enantioselective Aminohydroxylation of Styrenyl Olefins Catalyzed by an Engineered Hemoprotein.

2019

Angewandte Chemie (International ed. in English)

Cho, Inha

Prier, Christopher K

Jia, Zhi-Jun

Zhang, Ruijie K

Gorbe, Tamas

Arnold, Frances H

Chiral 1,2-amino alcohols are widely represented in biologically active compounds from neurotransmitters to antivirals. While many synthetic methods have been developed for accessing amino alcohols, the direct aminohydroxylation of alkenes to unprotected, enantioenriched amino alcohols remains a challenge. Using directed evolution, we have engineered a hemoprotein biocatalyst based on a thermostable cytochrome c that directly transforms alkenes to amino alcohols with high enantioselectivity (up to 2500 TTN and 90 % ee) under anaerobic conditions with O-pivaloylhydroxylamine as an aminating reagent. The reaction is proposed to proceed via a reactive iron-nitrogen species generated in the enzyme active site, enabling tuning of the catalyst's activity and selectivity by protein engineering.

# Enzymatic assembly of carbon-carbon bonds via iron-catalysed sp(3) C-H functionalization.

2018

Nature

Zhang, Ruijie K

Chen, Kai

Huang, Xiongyi

Wohlschlager, Lena

Renata, Hans

Arnold, Frances H

Although abundant in organic molecules, carbon-hydrogen (C-H) bonds are typically considered unreactive and unavailable for chemical manipulation. Recent advances in C-H functionalization technology have begun to transform this logic, while emphasizing the importance of and challenges associated with selective alkylation at a sp(3) carbon(1,2). Here we describe iron-based catalysts for the enantio-, regio- and chemoselective intermolecular alkylation of sp(3) C-H bonds through carbene C-H insertion. The catalysts, derived from a cytochrome P450 enzyme in which the native cysteine axial ligand has been substituted for serine (cytochrome P411), are fully genetically encoded and produced in bacteria, where they can be tuned by directed evolution for activity and selectivity. That these proteins activate iron, the most abundant transition metal, to perform this chemistry provides a desirable alternative to noble-metal catalysts, which have dominated the field of C-H functionalization(1,2). The laboratory-evolved enzymes functionalize diverse substrates containing benzylic, allylic or alpha-amino C-H bonds with high turnover and excellent selectivity. Furthermore, they have enabled the development of concise routes to several natural products. The use of the native iron-haem cofactor of these enzymes to mediate sp(3) C-H alkylation suggests that diverse haem proteins could serve as potential catalysts for this abiological transformation, and will facilitate the development of new enzymatic C-H functionalization reactions for applications in chemistry and synthetic biology.

# Alternate Heme Ligation Steers Activity and Selectivity in Engineered Cytochrome P450-Catalyzed Carbene-Transfer Reactions.

2018

Journal of the American Chemical Society

Chen, Kai

Zhang, Shuo-Qing

Brandenberg, Oliver F

Hong, Xin

Arnold, Frances H

We report a biocatalytic platform of engineered cytochrome P450 enzymes to carry out carbene-transfer reactions using a lactone-based carbene precursor. By simply altering the heme-ligating residue, we obtained two enzymes that catalyze olefin cyclopropanation (Ser) or S-H bond insertion (Cys). Both enzymes exhibit high catalytic efficiency and stereoselectivity, thus enabling facile access to structurally diverse spiro[2.4]lactones and alpha-thio-gamma-lactones. Computational studies revealed the mechanism of carbene S-H insertion and explain how the axial ligand controls reactivity and selectivity. This work expands the catalytic repertoire of hemeproteins and offers insights into how these enzymes can be tuned for new chemistry.

# Selective CH bond functionalization with engineered heme proteins: new tools to generate complexity.

2018

Current opinion in chemical biology

Zhang, Ruijie K

Huang, Xiongyi

Arnold, Frances H

CH functionalization is an attractive strategy to construct and diversify molecules. Heme proteins, predominantly cytochromes P450, are responsible for an array of CH oxidations in biology. Recent work has coupled concepts from synthetic chemistry, computation, and natural product biosynthesis to engineer heme protein systems to deliver products with tailored oxidation patterns. Heme protein catalysis has been shown to go well beyond these native reactions and now accesses new-to-nature CH transformations, including CN and CC bond forming processes. Emerging work with these systems moves us along the ambitious path of building complexity from the ubiquitous CH bond.

# Engineering enzymes for noncanonical amino acid synthesis.

2018

Chemical Society reviews

Almhjell, Patrick J

Boville, Christina E

Arnold, Frances H

The standard proteinogenic amino acids grant access to a myriad of chemistries that harmonize to create life. Outside of these twenty canonical protein building blocks are countless noncanonical amino acids (ncAAs), either found in nature or created by man. Interest in ncAAs has grown as research has unveiled their importance as precursors to natural products and pharmaceuticals, biological probes, and more. Despite their broad applications, synthesis of ncAAs remains a challenge, as poor stereoselectivity and low functional-group compatibility stymie effective preparative routes. The use of enzymes has emerged as a versatile approach to prepare ncAAs, and nature's enzymes can be engineered to synthesize ncAAs more efficiently and expand the amino acid alphabet. In this tutorial review, we briefly outline different enzyme engineering strategies and then discuss examples where engineering has generated new 'ncAA synthases' for efficient, environmentally benign production of a wide and growing collection of valuable ncAAs.

# Engineered Biosynthesis of beta-Alkyl Tryptophan Analogues.

2018

Angewandte Chemie (International ed. in English)

Boville, Christina E

Scheele, Remkes A

Koch, Philipp

Brinkmann-Chen, Sabine

Buller, Andrew R

Arnold, Frances H

Noncanonical amino acids (ncAAs) with dual stereocenters at the alpha and beta positions are valuable precursors to natural products and therapeutics. Despite the potential applications of such bioactive beta-branched ncAAs, their availability is limited due to the inefficiency of the multistep methods used to prepare them. Herein we report a stereoselective biocatalytic synthesis of beta-branched tryptophan analogues using an engineered variant of Pyrococcus furiosus tryptophan synthase (PfTrpB), PfTrpB(7E6) . PfTrpB(7E6) is the first biocatalyst to synthesize bulky beta-branched tryptophan analogues in a single step, with demonstrated access to 27 ncAAs. The molecular basis for the efficient catalysis and broad substrate tolerance of PfTrpB(7E6) was explored through X-ray crystallography and UV/Vis spectroscopy, which revealed that a combination of active-site and remote mutations increase the abundance and persistence of a key reactive intermediate. PfTrpB(7E6) provides an operationally simple and environmentally benign platform for the preparation of beta-branched tryptophan building blocks.

# Chemistry Takes a Bath: Reactions in Aqueous Media.

2018

The Journal of organic chemistry

Romney, David K

Arnold, Frances H

Lipshutz, Bruce H

Li, Chao-Jun

# Catalytic iron-carbene intermediate revealed in a cytochrome c carbene transferase.

2018

Proceedings of the National Academy of Sciences of the United States of America

Lewis, Russell D

Garcia-Borras, Marc

Chalkley, Matthew J

Buller, Andrew R

Houk, K N

Kan, S B Jennifer

Arnold, Frances H

Recently, heme proteins have been discovered and engineered by directed evolution to catalyze chemical transformations that are biochemically unprecedented. Many of these nonnatural enzyme-catalyzed reactions are assumed to proceed through a catalytic iron porphyrin carbene (IPC) intermediate, although this intermediate has never been observed in a protein. Using crystallographic, spectroscopic, and computational methods, we have captured and studied a catalytic IPC intermediate in the active site of an enzyme derived from thermostable Rhodothermus marinus (Rma) cytochrome c High-resolution crystal structures and computational methods reveal how directed evolution created an active site for carbene transfer in an electron transfer protein and how the laboratory-evolved enzyme achieves perfect carbene transfer stereoselectivity by holding the catalytic IPC in a single orientation. We also discovered that the IPC in Rma cytochrome c has a singlet ground electronic state and that the protein environment uses geometrical constraints and noncovalent interactions to influence different IPC electronic states. This information helps us to understand the impressive reactivity and selectivity of carbene transfer enzymes and offers insights that will guide and inspire future engineering efforts.

# Learned protein embeddings for machine learning.

2018

Bioinformatics (Oxford, England)

Yang, Kevin K

Wu, Zachary

Bedbrook, Claire N

Arnold, Frances H

# Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble.

2018

Journal of the American Chemical Society

Buller, Andrew R

van Roye, Paul

Cahn, Jackson K B

Scheele, Remkes A

Herger, Michael

Arnold, Frances H

Allosteric enzymes contain a wealth of catalytic diversity that remains distinctly underutilized for biocatalysis. Tryptophan synthase is a model allosteric system and a valuable enzyme for the synthesis of noncanonical amino acids (ncAA). Previously, we evolved the beta-subunit from Pyrococcus furiosus, PfTrpB, for ncAA synthase activity in the absence of its native partner protein PfTrpA. However, the precise mechanism by which mutation activated TrpB to afford a stand-alone catalyst remained enigmatic. Here, we show that directed evolution caused a gradual change in the rate-limiting step of the catalytic cycle. Concomitantly, the steady-state distribution of the intermediates shifts to favor covalently bound Trp adducts, which have increased thermodynamic stability. The biochemical properties of these evolved, stand-alone TrpBs converge on those induced in the native system by allosteric activation. High-resolution crystal structures of the wild-type enzyme, an intermediate in the lineage, and the final variant, encompassing five distinct chemical states, show that activating mutations have only minor structural effects on their immediate environment. Instead, mutation stabilizes the large-scale motion of a subdomain to favor an otherwise transiently populated closed conformational state. This increase in stability enabled the first structural description of Trp covalently bound in a catalytically active TrpB, confirming key features of catalysis. These data combine to show that sophisticated models of allostery are not a prerequisite to recapitulating its complex effects via directed evolution, opening the way to engineering stand-alone versions of diverse allosteric enzymes.

# Improved Synthesis of 4-Cyanotryptophan and Other Tryptophan Analogues in Aqueous Solvent Using Variants of TrpB from Thermotoga maritima.

2018

The Journal of organic chemistry

Boville, Christina E

Romney, David K

Almhjell, Patrick J

Sieben, Michaela

Arnold, Frances H

The use of enzymes has become increasingly widespread in synthesis as chemists strive to reduce their reliance on organic solvents in favor of more environmentally benign aqueous media. With this in mind, we previously endeavored to engineer the tryptophan synthase beta-subunit (TrpB) for production of noncanonical amino acids that had previously been synthesized through multistep routes involving water-sensitive reagents. This enzymatic platform proved effective for the synthesis of analogues of the amino acid tryptophan (Trp), which are frequently used in pharmaceutical synthesis as well as chemical biology. However, certain valuable compounds, such as the blue fluorescent amino acid 4-cyanotryptophan (4-CN-Trp), could only be made in low yield, even at elevated temperature (75 degrees C). Here, we describe the engineering of TrpB from Thermotoga maritima that improved synthesis of 4-CN-Trp from 24% to 78% yield. Remarkably, although the final enzyme maintains high thermostability ( T50 = 93 degrees C), its temperature profile is shifted such that high reactivity is observed at approximately 37 degrees C (76% yield), creating the possibility for in vivo 4-CN-Trp production. The improvements are not specific to 4-CN-Trp; a boost in activity at lower temperature is also demonstrated for other Trp analogues.

# Diverse Engineered Heme Proteins Enable Stereodivergent Cyclopropanation of Unactivated Alkenes.

2018

ACS central science

Knight, Anders M

Kan, S B Jennifer

Lewis, Russell D

Brandenberg, Oliver F

Chen, Kai

Arnold, Frances H

Developing catalysts that produce each stereoisomer of a desired product selectively is a longstanding synthetic challenge. Biochemists have addressed this challenge by screening nature's diversity to discover enzymes that catalyze the formation of complementary stereoisomers. We show here that the same approach can be applied to a new-to-nature enzymatic reaction, alkene cyclopropanation via carbene transfer. By screening diverse native and engineered heme proteins, we identified globins and serine-ligated "P411" variants of cytochromes P450 with promiscuous activity for cyclopropanation of unactivated alkene substrates. We then enhanced their activities and stereoselectivities by directed evolution: just 1-3 rounds of site-saturation mutagenesis and screening generated enzymes that transform unactivated alkenes and electron-deficient alkenes into each of the four stereoisomeric cyclopropanes with up to 5,400 total turnovers and 98% enantiomeric excess. These fully genetically encoded biocatalysts function in whole Escherichia coli cells in mild, aqueous conditions and provide the first example of enantioselective, intermolecular iron-catalyzed cyclopropanation of unactivated alkenes.

# Enzymatic construction of highly strained carbocycles.

2018

Science (New York, N.Y.)

Chen, Kai

Huang, Xiongyi

Kan, S B Jennifer

Zhang, Ruijie K

Arnold, Frances H

Small carbocycles are structurally rigid and possess high intrinsic energy due to their ring strain. These features lead to broad applications but also create challenges for their construction. We report the engineering of hemeproteins that catalyze the formation of chiral bicyclobutanes, one of the most strained four-membered systems, via successive carbene addition to unsaturated carbon-carbon bonds. Enzymes that produce cyclopropenes, putative intermediates to the bicyclobutanes, were also identified. These genetically encoded proteins are readily optimized by directed evolution, function in Escherichia coli, and act on structurally diverse substrates with high efficiency and selectivity, providing an effective route to many chiral strained structures. This biotransformation is easily performed at preparative scale, and the resulting strained carbocycles can be derivatized, opening myriad potential applications.

# Learned protein embeddings for machine learning.

2018

Bioinformatics (Oxford, England)

Yang, Kevin K

Wu, Zachary

Bedbrook, Claire N

Arnold, Frances H

Motivation: Machine-learning models trained on protein sequences and their measured functions can infer biological properties of unseen sequences without requiring an understanding of the underlying physical or biological mechanisms. Such models enable the prediction and discovery of sequences with optimal properties. Machine-learning models generally require that their inputs be vectors, and the conversion from a protein sequence to a vector representation affects the model's ability to learn. We propose to learn embedded representations of protein sequences that take advantage of the vast quantity of unmeasured protein sequence data available. These embeddings are low-dimensional and can greatly simplify downstream modeling. Results: The predictive power of Gaussian process models trained using embeddings is comparable to those trained on existing representations, which suggests that embeddings enable accurate predictions despite having orders of magnitude fewer dimensions. Moreover, embeddings are simpler to obtain because they do not require alignments, structural data, or selection of informative amino-acid properties. Visualizing the embedding vectors shows meaningful relationships between the embedded proteins are captured. Availability and implementation: The embedding vectors and code to reproduce the results are available at https://github.com/fhalab/embeddings\_reproduction/. Supplementary information: Supplementary data are available at Bioinformatics online.

# Genetically programmed chiral organoborane synthesis.

2017

Nature

Kan, S B Jennifer

Huang, Xiongyi

Gumulya, Yosephine

Chen, Kai

Arnold, Frances H

Recent advances in enzyme engineering and design have expanded nature's catalytic repertoire to functions that are new to biology. However, only a subset of these engineered enzymes can function in living systems. Finding enzymatic pathways that form chemical bonds that are not found in biology is particularly difficult in the cellular environment, as this depends on the discovery not only of new enzyme activities, but also of reagents that are both sufficiently reactive for the desired transformation and stable in vivo. Here we report the discovery, evolution and generalization of a fully genetically encoded platform for producing chiral organoboranes in bacteria. Escherichia coli cells harbouring wild-type cytochrome c from Rhodothermus marinus (Rma cyt c) were found to form carbon-boron bonds in the presence of borane-Lewis base complexes, through carbene insertion into boron-hydrogen bonds. Directed evolution of Rma cyt c in the bacterial catalyst provided access to 16 novel chiral organoboranes. The catalyst is suitable for gram-scale biosynthesis, providing up to 15,300 turnovers, a turnover frequency of 6,100 h(-1), a 99:1 enantiomeric ratio and 100% chemoselectivity. The enantiopreference of the biocatalyst could also be tuned to provide either enantiomer of the organoborane products. Evolved in the context of whole-cell catalysts, the proteins were more active in the whole-cell system than in purified forms. This study establishes a DNA-encoded and readily engineered bacterial platform for borylation; engineering can be accomplished at a pace that rivals the development of chemical synthetic methods, with the ability to achieve turnovers that are two orders of magnitude (over 400-fold) greater than those of known chiral catalysts for the same class of transformation. This tunable method for manipulating boron in cells could expand the scope of boron chemistry in living systems.

# Enzyme Nicotinamide Cofactor Specificity Reversal Guided by Automated Structural Analysis and Library Design.

2017

Methods in molecular biology (Clifton, N.J.)

Cahn, Jackson K B

Brinkmann-Chen, Sabine

Arnold, Frances H

The specificity of enzymes for nicotinamide adenine dinucleotide (NAD) or nicotinamide adenine dinucleotide phosphate (NADP) as redox carriers can pose a significant hurdle for metabolic engineering and synthetic biology applications, where switching the specificity might be beneficial. We have developed an easy-to-use computational tool (CSR-SALAD) for the design of mutant libraries to simplify the process of reversing the cofactor specificity of an enzyme. Here, we describe the optimal use of this tool and present methods for its application in a laboratory setting.

# Directed Evolution: Bringing New Chemistry to Life.

2017

Angewandte Chemie (International ed. in English)

Arnold, Frances H

Tailor-made: Discussed herein is the ability to adapt biology's mechanisms for innovation and optimization to solving problems in chemistry and engineering. The evolution of nature's enzymes can lead to the discovery of new reactivity, transformations not known in biology, and reactivity inaccessible by small-molecule catalysts.

# Machine learning to design integral membrane channelrhodopsins for efficient eukaryotic expression and plasma membrane localization.

2017

PLoS computational biology

Bedbrook, Claire N

Yang, Kevin K

Rice, Austin J

Gradinaru, Viviana

Arnold, Frances H

There is growing interest in studying and engineering integral membrane proteins (MPs) that play key roles in sensing and regulating cellular response to diverse external signals. A MP must be expressed, correctly inserted and folded in a lipid bilayer, and trafficked to the proper cellular location in order to function. The sequence and structural determinants of these processes are complex and highly constrained. Here we describe a predictive, machine-learning approach that captures this complexity to facilitate successful MP engineering and design. Machine learning on carefully-chosen training sequences made by structure-guided SCHEMA recombination has enabled us to accurately predict the rare sequences in a diverse library of channelrhodopsins (ChRs) that express and localize to the plasma membrane of mammalian cells. These light-gated channel proteins of microbial origin are of interest for neuroscience applications, where expression and localization to the plasma membrane is a prerequisite for function. We trained Gaussian process (GP) classification and regression models with expression and localization data from 218 ChR chimeras chosen from a 118,098-variant library designed by SCHEMA recombination of three parent ChRs. We use these GP models to identify ChRs that express and localize well and show that our models can elucidate sequence and structure elements important for these processes. We also used the predictive models to convert a naturally occurring ChR incapable of mammalian localization into one that localizes well.

# Anti-Markovnikov alkene oxidation by metal-oxo-mediated enzyme catalysis.

2017

Science (New York, N.Y.)

Hammer, Stephan C

Kubik, Grzegorz

Watkins, Ella

Huang, Shan

Minges, Hannah

Arnold, Frances H

Catalytic anti-Markovnikov oxidation of alkene feedstocks could simplify synthetic routes to many important molecules and solve a long-standing challenge in chemistry. Here we report the engineering of a cytochrome P450 enzyme by directed evolution to catalyze metal-oxo-mediated anti-Markovnikov oxidation of styrenes with high efficiency. The enzyme uses dioxygen as the terminal oxidant and achieves selectivity for anti-Markovnikov oxidation over the kinetically favored alkene epoxidation by trapping high-energy intermediates and catalyzing an oxo transfer, including an enantioselective 1,2-hydride migration. The anti-Markovnikov oxygenase can be combined with other catalysts in synthetic metabolic pathways to access a variety of challenging anti-Markovnikov functionalization reactions.

# Enantioselective Total Synthesis of Nigelladine A via Late-Stage C-H Oxidation Enabled by an Engineered P450 Enzyme.

2017

Journal of the American Chemical Society

Loskot, Steven A

Romney, David K

Arnold, Frances H

Stoltz, Brian M

An enantioselective total synthesis of the norditerpenoid alkaloid nigelladine A is described. Strategically, the synthesis relies on a late-stage C-H oxidation of an advanced intermediate. While traditional chemical methods failed to deliver the desired outcome, an engineered cytochrome P450 enzyme was employed to effect a chemo- and regioselective allylic C-H oxidation in the presence of four oxidizable positions. The enzyme variant was readily identified from a focused library of three enzymes, allowing for completion of the synthesis without the need for extensive screening.

# Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions.

2017

Current opinion in biotechnology

Brandenberg, Oliver F

Fasan, Rudi

Arnold, Frances H

The surge in reports of heme-dependent proteins as catalysts for abiotic, synthetically valuable carbene and nitrene transfer reactions dramatically illustrates the evolvability of the protein world and our nascent ability to exploit that for new enzyme chemistry. We highlight the latest additions to the hemoprotein-catalyzed reaction repertoire (including carbene Si-H and C-H insertions, Doyle-Kirmse reactions, aldehyde olefinations, azide-to-aldehyde conversions, and intermolecular nitrene C-H insertion) and show how different hemoprotein scaffolds offer varied reactivity and selectivity. Preparative-scale syntheses of pharmaceutically relevant compounds accomplished with these new catalysts are beginning to demonstrate their biotechnological relevance. Insights into the determinants of enzyme lifetime and product yield are providing generalizable cues for engineering heme-dependent proteins to further broaden the scope and utility of these non-natural activities.

# Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues.

2017

Journal of the American Chemical Society

Romney, David K

Murciano-Calles, Javier

Wehrmuller, Jori E

Arnold, Frances H

Derivatives of the amino acid tryptophan (Trp) serve as precursors for the chemical and biological synthesis of complex molecules with a wide range of biological properties. Trp analogues are also valuable as building blocks for medicinal chemistry and as tools for chemical biology. While the enantioselective synthesis of Trp analogues is often lengthy and requires the use of protecting groups, enzymes have the potential to synthesize such products in fewer steps and with the pristine chemo- and stereoselectivity that is a hallmark of biocatalysis. The enzyme TrpB is especially attractive because it can form Trp analogues directly from serine (Ser) and the corresponding indole analogue. However, many potentially useful substrates, including bulky or electron-deficient indoles, are poorly accepted. We have applied directed evolution to TrpB from Pyrococcus furiosus and Thermotoga maritima to generate a suite of catalysts for the synthesis of previously intractable Trp analogues. For the most challenging substrates, such as nitroindoles, the key to improving activity lay in the mutation of a universally conserved and mechanistically important residue, E104. The new catalysts express at high levels (>200 mg/L of Escherichia coli culture) and can be purified by heat treatment; they can operate up to 75 degrees C (where solubility is enhanced) and can synthesize enantiopure Trp analogues substituted at the 4-, 5-, 6-, and 7-positions, using Ser and readily available indole analogues as starting materials. Spectroscopic analysis shows that many of the activating mutations suppress the decomposition of the active electrophilic intermediate, an amino-acrylate, which aids in unlocking the synthetic potential of TrpB.

# Enantioselective, intermolecular benzylic C-H amination catalysed by an engineered iron-haem enzyme.

2017

Nature chemistry

Prier, Christopher K

Zhang, Ruijie K

Buller, Andrew R

Brinkmann-Chen, Sabine

Arnold, Frances H

C-H bonds are ubiquitous structural units of organic molecules. Although these bonds are generally considered to be chemically inert, the recent emergence of methods for C-H functionalization promises to transform the way synthetic chemistry is performed. The intermolecular amination of C-H bonds represents a particularly desirable and challenging transformation for which no efficient, highly selective, and renewable catalysts exist. Here we report the directed evolution of an iron-containing enzymatic catalyst-based on a cytochrome P450 monooxygenase-for the highly enantioselective intermolecular amination of benzylic C-H bonds. The biocatalyst is capable of up to 1,300 turnovers, exhibits excellent enantioselectivities, and provides access to valuable benzylic amines. Iron complexes are generally poor catalysts for C-H amination: in this catalyst, the enzyme's protein framework confers activity on an otherwise unreactive iron-haem cofactor.

# Highly Stereoselective Biocatalytic Synthesis of Key Cyclopropane Intermediate to Ticagrelor.

2017

ACS catalysis

Hernandez, Kari E

Renata, Hans

Lewis, Russell D

Jennifer Kan, S B

Zhang, Chen

Forte, Jared

Rozzell, David

McIntosh, John A

Arnold, Frances H

Extending the scope of biocatalysis to important non-natural reactions such as olefin cyclopropanation will open new opportunities for replacing multi-step chemical syntheses of pharmaceutical intermediates with efficient, clean, and highly selective enzyme-catalyzed processes. In this work, we engineered the truncated globin of Bacillus subtilis for the synthesis of a cyclopropane precursor to the antithrombotic agent ticagrelor. The engineered enzyme catalyzes the cyclopropanation of 3,4-difluorostyrene with ethyl diazoacetate on a preparative scale to give ethyl-(1R, 2R)-2-(3,4-difluorophenyl)-cyclopropanecarboxylate in 79% yield, with very high diastereoselectivity (>99% dr) and enantioselectivity (98% ee), enabling a single-step biocatalytic route to this pharmaceutical intermediate.

# Structure-guided SCHEMA recombination generates diverse chimeric channelrhodopsins.

2017

Proceedings of the National Academy of Sciences of the United States of America

Bedbrook, Claire N

Rice, Austin J

Yang, Kevin K

Ding, Xiaozhe

Chen, Siyuan

LeProust, Emily M

Gradinaru, Viviana

Arnold, Frances H

Integral membrane proteins (MPs) are key engineering targets due to their critical roles in regulating cell function. In engineering MPs, it can be extremely challenging to retain membrane localization capability while changing other desired properties. We have used structure-guided SCHEMA recombination to create a large set of functionally diverse chimeras from three sequence-diverse channelrhodopsins (ChRs). We chose 218 ChR chimeras from two SCHEMA libraries and assayed them for expression and plasma membrane localization in human embryonic kidney cells. The majority of the chimeras express, with 89% of the tested chimeras outperforming the lowest-expressing parent; 12% of the tested chimeras express at even higher levels than any of the parents. A significant fraction (23%) also localize to the membrane better than the lowest-performing parent ChR. Most (93%) of these well-localizing chimeras are also functional light-gated channels. Many chimeras have stronger light-activated inward currents than the three parents, and some have unique off-kinetics and spectral properties relative to the parents. An effective method for generating protein sequence and functional diversity, SCHEMA recombination can be used to gain insights into sequence-function relationships in MPs.

# Directed Evolution of a Bright Near-Infrared Fluorescent Rhodopsin Using a Synthetic Chromophore.

2017

Cell chemical biology

Herwig, Lukas

Rice, Austin J

Bedbrook, Claire N

Zhang, Ruijie K

Lignell, Antti

Cahn, Jackson K B

Renata, Hans

Dodani, Sheel C

Cho, Inha

Cai, Long

Gradinaru, Viviana

Arnold, Frances H

By engineering a microbial rhodopsin, Archaerhodopsin-3 (Arch), to bind a synthetic chromophore, merocyanine retinal, in place of the natural chromophore all-trans-retinal (ATR), we generated a protein with exceptionally bright and unprecedentedly red-shifted near-infrared (NIR) fluorescence. We show that chromophore substitution generates a fluorescent Arch complex with a 200-nm bathochromic excitation shift relative to ATR-bound wild-type Arch and an emission maximum at 772 nm. Directed evolution of this complex produced variants with pH-sensitive NIR fluorescence and molecular brightness 8.5-fold greater than the brightest ATR-bound Arch variant. The resulting proteins are well suited to bacterial imaging; expression and stability have not been optimized for mammalian cell imaging. By targeting both the protein and its chromophore, we overcome inherent challenges associated with engineering bright NIR fluorescence into Archaerhodopsin. This work demonstrates an efficient strategy for engineering non-natural, tailored properties into microbial opsins, properties relevant for imaging and interrogating biological systems.

# Tryptophan Synthase Uses an Atypical Mechanism To Achieve Substrate Specificity.

2016

Biochemistry

Buller, Andrew R

van Roye, Paul

Murciano-Calles, Javier

Arnold, Frances H

Tryptophan synthase (TrpS) catalyzes the final steps in the biosynthesis of l-tryptophan from l-serine (Ser) and indole-3-glycerol phosphate (IGP). We report that native TrpS can also catalyze a productive reaction with l-threonine (Thr), leading to (2S,3S)-beta-methyltryptophan. Surprisingly, beta-substitution occurs in vitro with a 3.4-fold higher catalytic efficiency for Ser over Thr using saturating indole, despite a >82000-fold preference for Ser in direct competition using IGP. Structural data identify a novel product binding site, and kinetic experiments clarify the atypical mechanism of specificity: Thr binds efficiently but decreases the affinity for indole and disrupts the allosteric signaling that regulates the catalytic cycle.

# Directed evolution of cytochrome c for carbon-silicon bond formation: Bringing silicon to life.

2016

Science (New York, N.Y.)

Kan, S B Jennifer

Lewis, Russell D

Chen, Kai

Arnold, Frances H

Enzymes that catalyze carbon-silicon bond formation are unknown in nature, despite the natural abundance of both elements. Such enzymes would expand the catalytic repertoire of biology, enabling living systems to access chemical space previously only open to synthetic chemistry. We have discovered that heme proteins catalyze the formation of organosilicon compounds under physiological conditions via carbene insertion into silicon-hydrogen bonds. The reaction proceeds both in vitro and in vivo, accommodating a broad range of substrates with high chemo- and enantioselectivity. Using directed evolution, we enhanced the catalytic function of cytochrome c from Rhodothermus marinus to achieve more than 15-fold higher turnover than state-of-the-art synthetic catalysts. This carbon-silicon bond-forming biocatalyst offers an environmentally friendly and highly efficient route to producing enantiopure organosilicon molecules.

# A General Tool for Engineering the NAD/NADP Cofactor Preference of Oxidoreductases.

2016

ACS synthetic biology

Cahn, Jackson K B

Werlang, Caroline A

Baumschlager, Armin

Brinkmann-Chen, Sabine

Mayo, Stephen L

Arnold, Frances H

The ability to control enzymatic nicotinamide cofactor utilization is critical for engineering efficient metabolic pathways. However, the complex interactions that determine cofactor-binding preference render this engineering particularly challenging. Physics-based models have been insufficiently accurate and blind directed evolution methods too inefficient to be widely adopted. Building on a comprehensive survey of previous studies and our own prior engineering successes, we present a structure-guided, semirational strategy for reversing enzymatic nicotinamide cofactor specificity. This heuristic-based approach leverages the diversity and sensitivity of catalytically productive cofactor binding geometries to limit the problem to an experimentally tractable scale. We demonstrate the efficacy of this strategy by inverting the cofactor specificity of four structurally diverse NADP-dependent enzymes: glyoxylate reductase, cinnamyl alcohol dehydrogenase, xylose reductase, and iron-containing alcohol dehydrogenase. The analytical components of this approach have been fully automated and are available in the form of an easy-to-use web tool: Cofactor Specificity Reversal-Structural Analysis and Library Design (CSR-SALAD).

# Identification of Mechanism-Based Inactivation in P450-Catalyzed Cyclopropanation Facilitates Engineering of Improved Enzymes.

2016

Journal of the American Chemical Society

Renata, Hans

Lewis, Russell D

Sweredoski, Michael J

Moradian, Annie

Hess, Sonja

Wang, Z Jane

Arnold, Frances H

Following the recent discovery that heme proteins can catalyze the cyclopropanation of styrenyl olefins with high efficiency and selectivity, interest in developing new enzymes for a variety of non-natural carbene transfer reactions has burgeoned. The fact that diazo compounds and other carbene precursors are known mechanism-based inhibitors of P450s, however, led us to investigate if they also interfere with this new enzyme function. We present evidence for two inactivation pathways that are operative during cytochrome P450-catalyzed cyclopropanation. Using a combination of UV-vis, mass spectrometry, and proteomic analyses, we show that the heme cofactor and several nucleophilic side chains undergo covalent modification by ethyl diazoacetate (EDA). Substitution of two of the affected residues with less-nucleophilic amino acids led to a more than twofold improvement in cyclopropanation performance (total TTN). Elucidating the inactivation pathways of heme protein-based carbene transfer catalysts should aid in the optimization of this new biocatalytic function.

# A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation.

2016

Angewandte Chemie (International ed. in English)

Murciano-Calles, Javier

Romney, David K

Brinkmann-Chen, Sabine

Buller, Andrew R

Arnold, Frances H

Naturally occurring enzyme homologues often display highly divergent activity with non-natural substrates. Exploiting this diversity with enzymes engineered for new or altered function, however, is laborious because the engineering must be replicated for each homologue. A small set of mutations of the tryptophan synthase beta-subunit (TrpB) from Pyrococcus furiosus, which mimics the activation afforded by binding of the alpha-subunit, was demonstrated to have a similar activating effect in different TrpB homologues with as little as 57 % sequence identity. Kinetic and spectroscopic analyses indicate that the mutations function through the same mechanism: mimicry of alpha-subunit binding. From these enzymes, we identified a new TrpB catalyst that displays a remarkably broad activity profile in the synthesis of 5-substituted tryptophans. This demonstrates that allosteric activation can be recapitulated throughout a protein family to explore natural sequence diversity for desirable biocatalytic transformations.

# Enhancement of cellulosome-mediated deconstruction of cellulose by improving enzyme thermostability.

2016

Biotechnology for biofuels

Morais, Sarah

Stern, Johanna

Kahn, Amaranta

Galanopoulou, Anastasia P

Yoav, Shahar

Shamshoum, Melina

Smith, Matthew A

Hatzinikolaou, Dimitris G

Arnold, Frances H

Bayer, Edward A

BACKGROUND: The concerted action of three complementary cellulases from Clostridium thermocellum, engineered to be stable at elevated temperatures, was examined on a cellulosic substrate and compared to that of the wild-type enzymes. Exoglucanase Cel48S and endoglucanase Cel8A, both key elements of the natural cellulosome from this bacterium, were engineered previously for increased thermostability, either by SCHEMA, a structure-guided, site-directed protein recombination method, or by consensus-guided mutagenesis combined with random mutagenesis using error-prone PCR, respectively. A thermostable beta-glucosidase BglA mutant was also selected from a library generated by error-prone PCR that will assist the two cellulases in their methodic deconstruction of crystalline cellulose. The effects of a thermostable scaffoldin versus those of a largely mesophilic scaffoldin were also examined. By improving the stability of the enzyme subunits and the structural component, we aimed to improve cellulosome-mediated deconstruction of cellulosic substrates. RESULTS: The results demonstrate that the combination of thermostable enzymes as free enzymes and a thermostable scaffoldin was more active on the cellulosic substrate than the wild-type enzymes. Significantly, "thermostable" designer cellulosomes exhibited a 1.7-fold enhancement in cellulose degradation compared to the action of conventional designer cellulosomes that contain the respective wild-type enzymes. For designer cellulosome formats, the use of the thermostabilized scaffoldin proved critical for enhanced enzymatic performance under conditions of high temperatures. CONCLUSIONS: Simple improvement in the activity of a given enzyme does not guarantee its suitability for use in an enzyme cocktail or as a designer cellulosome component. The true merit of improvement resides in its ultimate contribution to synergistic action, which can only be determined experimentally. The relevance of the mutated thermostable enzymes employed in this study as components in multienzyme systems has thus been confirmed using designer cellulosome technology. Enzyme integration via a thermostable scaffoldin is critical to the ultimate stability of the complex at higher temperatures. Engineering of thermostable cellulases and additional lignocellulosic enzymes may prove a determinant parameter for development of state-of-the-art designer cellulosomes for their employment in the conversion of cellulosic biomass to soluble sugars.Graphical abstractConversion of conventional designer cellulosomes into thermophilic designer cellulosomes.

# Synthesis of beta-Branched Tryptophan Analogues Using an Engineered Subunit of Tryptophan Synthase.

2016

Journal of the American Chemical Society

Herger, Michael

van Roye, Paul

Romney, David K

Brinkmann-Chen, Sabine

Buller, Andrew R

Arnold, Frances H

We report that l-threonine may substitute for l-serine in the beta-substitution reaction of an engineered subunit of tryptophan synthase from Pyrococcus furiosus, yielding (2S,3S)-beta-methyltryptophan (beta-MeTrp) in a single step. The trace activity of the wild-type beta-subunit on this substrate was enhanced more than 1000-fold by directed evolution. Structural and spectroscopic data indicate that this increase is correlated with stabilization of the electrophilic aminoacrylate intermediate. The engineered biocatalyst also reacts with a variety of indole analogues and thiophenol for diastereoselective C-C, C-N, and C-S bond-forming reactions. This new activity circumvents the 3-enzyme pathway that produces beta-MeTrp in nature and offers a simple and expandable route to preparing derivatives of this valuable building block.

# Discovery of a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models.

2016

Nature chemistry

Dodani, Sheel C

Kiss, Gert

Cahn, Jackson K B

Su, Ye

Pande, Vijay S

Arnold, Frances H

The dynamic motions of protein structural elements, particularly flexible loops, are intimately linked with diverse aspects of enzyme catalysis. Engineering of these loop regions can alter protein stability, substrate binding and even dramatically impact enzyme function. When these flexible regions are unresolvable structurally, computational reconstruction in combination with large-scale molecular dynamics simulations can be used to guide the engineering strategy. Here we present a collaborative approach that consists of both experiment and computation and led to the discovery of a single mutation in the F/G loop of the nitrating cytochrome P450 TxtE that simultaneously controls loop dynamics and completely shifts the enzyme's regioselectivity from the C4 to the C5 position of L-tryptophan. Furthermore, we find that this loop mutation is naturally present in a subset of homologous nitrating P450s and confirm that these uncharacterized enzymes exclusively produce 5-nitro-L-tryptophan, a previously unknown biosynthetic intermediate.

# Exploring the Mechanism Responsible for Cellulase Thermostability by Structure-Guided Recombination.

2016

PloS one

Chang, Chia-Jung

Lee, Cheng-Chung

Chan, Yueh-Te

Trudeau, Devin L

Wu, Mei-Huey

Tsai, Chih-Hsuan

Yu, Su-May

Ho, Tuan-Hua David

Wang, Andrew H-J

Hsiao, Chwan-Deng

Arnold, Frances H

Chao, Yu-Chan

Cellulases from Bacillus and Geobacillus bacteria are potentially useful in the biofuel and animal feed industries. One of the unique characteristics of these enzymes is that they are usually quite thermostable. We previously identified a cellulase, GsCelA, from thermophilic Geobacillus sp. 70PC53, which is much more thermostable than its Bacillus homolog, BsCel5A. Thus, these two cellulases provide a pair of structures ideal for investigating the mechanism regarding how these cellulases can retain activity at high temperature. In the present study, we applied the SCHEMA non-contiguous recombination algorithm as a novel tool, which assigns protein sequences into blocks for domain swapping in a way that lessens structural disruption, to generate a set of chimeric proteins derived from the recombination of GsCelA and BsCel5A. Analyzing the activity and thermostability of this designed library set, which requires only a limited number of chimeras by SCHEMA calculations, revealed that one of the blocks may contribute to the higher thermostability of GsCelA. When tested against swollen Avicel, the highly thermostable chimeric cellulase C10 containing this block showed significantly higher activity (22%-43%) and higher thermostability compared to the parental enzymes. With further structural determinations and mutagenesis analyses, a 310 helix was identified as being responsible for the improved thermostability of this block. Furthermore, in the presence of ionic calcium and crown ether (CR), the chimeric C10 was found to retain 40% residual activity even after heat treatment at 90 degrees C. Combining crystal structure determinations and structure-guided SCHEMA recombination, we have determined the mechanism responsible for the high thermostability of GsCelA, and generated a novel recombinant enzyme with significantly higher activity.

# Asymmetric Enzymatic Synthesis of Allylic Amines: A Sigmatropic Rearrangement Strategy.

2016

Angewandte Chemie (International ed. in English)

Prier, Christopher K

Hyster, Todd K

Farwell, Christopher C

Huang, Audrey

Arnold, Frances H

Sigmatropic rearrangements, while rare in biology, offer opportunities for the efficient and selective synthesis of complex chemical motifs. A "P411" serine-ligated variant of cytochrome P450(BM3) has been engineered to initiate a sulfimidation/[2,3]-sigmatropic rearrangement sequence in whole E. coli cells, a non-natural function for any enzyme, providing access to enantioenriched, protected allylic amines. Five mutations in the enzyme substantially enhance its activity toward this new function, demonstrating the evolvability of the catalyst toward challenging nitrene transfer reactions. The evolved catalyst additionally performs the highly enantioselective imidation of non-allylic sulfides.

# Artificial domain duplication replicates evolutionary history of ketol-acid reductoisomerases.

2015

Protein science : a publication of the Protein Society

Cahn, Jackson K B

Brinkmann-Chen, Sabine

Buller, Andrew R

Arnold, Frances H

The duplication of protein structural domains has been proposed as a common mechanism for the generation of new protein folds. A particularly interesting case is the class II ketol-acid reductoisomerase (KARI), which putatively arose from an ancestral class I KARI by duplication of the C-terminal domain and corresponding loss of obligate dimerization. As a result, the class II enzymes acquired a deeply embedded figure-of-eight knot. To test this evolutionary hypothesis we constructed a novel class II KARI by duplicating the C-terminal domain of a hyperthermostable class I KARI. The new protein is monomeric, as confirmed by gel filtration and X-ray crystallography, and has the deeply knotted class II KARI fold. Surprisingly, its catalytic activity is nearly unchanged from the parent KARI. This provides strong evidence in support of domain duplication as the mechanism for the evolution of the class II KARI fold and demonstrates the ability of domain duplication to generate topological novelty in a function-neutral manner.

# Directed evolution of the tryptophan synthase beta-subunit for stand-alone function recapitulates allosteric activation.

2015

Proceedings of the National Academy of Sciences of the United States of America

Buller, Andrew R

Brinkmann-Chen, Sabine

Romney, David K

Herger, Michael

Murciano-Calles, Javier

Arnold, Frances H

Enzymes in heteromeric, allosterically regulated complexes catalyze a rich array of chemical reactions. Separating the subunits of such complexes, however, often severely attenuates their catalytic activities, because they can no longer be activated by their protein partners. We used directed evolution to explore allosteric regulation as a source of latent catalytic potential using the beta-subunit of tryptophan synthase from Pyrococcus furiosus (PfTrpB). As part of its native alphabetabetaalpha complex, TrpB efficiently produces tryptophan and tryptophan analogs; activity drops considerably when it is used as a stand-alone catalyst without the alpha-subunit. Kinetic, spectroscopic, and X-ray crystallographic data show that this lost activity can be recovered by mutations that reproduce the effects of complexation with the alpha-subunit. The engineered PfTrpB is a powerful platform for production of Trp analogs and for further directed evolution to expand substrate and reaction scope.

# The nature of chemical innovation: new enzymes by evolution.

2015

Quarterly reviews of biophysics

Arnold, Frances H

I describe how we direct the evolution of non-natural enzyme activities, using chemical intuition and information on structure and mechanism to guide us to the most promising reaction/enzyme systems. With synthetic reagents to generate new reactive intermediates and just a few amino acid substitutions to tune the active site, a cytochrome P450 can catalyze a variety of carbene and nitrene transfer reactions. The cyclopropanation, N-H insertion, C-H amination, sulfimidation, and aziridination reactions now demonstrated are all well known in chemical catalysis but have no counterparts in nature. The new enzymes are fully genetically encoded, assemble and function inside of cells, and can be optimized for different substrates, activities, and selectivities. We are learning how to use nature's innovation mechanisms to marry some of the synthetic chemists' favorite transformations with the exquisite selectivity and tunability of enzymes.

# Chemomimetic biocatalysis: exploiting the synthetic potential of cofactor-dependent enzymes to create new catalysts.

2015

Journal of the American Chemical Society

Prier, Christopher K

Arnold, Frances H

Despite the astonishing breadth of enzymes in nature, no enzymes are known for many of the valuable catalytic transformations discovered by chemists. Recent work in enzyme design and evolution, however, gives us good reason to think that this will change. We describe a chemomimetic biocatalysis approach that draws from small-molecule catalysis and synthetic chemistry, enzymology, and molecular evolution to discover or create enzymes with non-natural reactivities. We illustrate how cofactor-dependent enzymes can be exploited to promote reactions first established with related chemical catalysts. The cofactors can be biological, or they can be non-biological to further expand catalytic possibilities. The ability of enzymes to amplify and precisely control the reactivity of their cofactors together with the ability to optimize non-natural reactivity by directed evolution promises to yield exceptional catalysts for challenging transformations that have no biological counterparts.

# Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450.

2015

ACS central science

Farwell, Christopher C

Zhang, Ruijie K

McIntosh, John A

Hyster, Todd K

Arnold, Frances H

One of the greatest challenges in protein design is creating new enzymes, something evolution does all the time, starting from existing ones. Borrowing from nature's evolutionary strategy, we have engineered a bacterial cytochrome P450 to catalyze highly enantioselective intermolecular aziridination, a synthetically useful reaction that has no natural biological counterpart. The new enzyme is fully genetically encoded, functions in vitro or in whole cells, and can be optimized rapidly to exhibit high enantioselectivity (up to 99% ee) and productivity (up to 1,000 catalytic turnovers) for intermolecular aziridination, demonstrated here with tosyl azide and substituted styrenes. This new aziridination activity highlights the remarkable ability of a natural enzyme to adapt and take on new functions. Once discovered in an evolvable enzyme, this non-natural activity was improved and its selectivity tuned through an evolutionary process of accumulating beneficial mutations.

# Structural Adaptability Facilitates Histidine Heme Ligation in a Cytochrome P450.

2015

Journal of the American Chemical Society

McIntosh, John A

Heel, Thomas

Buller, Andrew R

Chio, Linda

Arnold, Frances H

Almost all known members of the cytochrome P450 (CYP) superfamily conserve a key cysteine residue that coordinates the heme iron. Although mutation of this residue abolishes monooxygenase activity, recent work has shown that mutation to either serine or histidine unlocks non-natural carbene- and nitrene-transfer activities. Here we present the first crystal structure of a histidine-ligated P450. The T213A/C317H variant of the thermostable CYP119 from Sulfolobus acidocaldarius maintains heme iron coordination through the introduced ligand, an interaction that is accompanied by large changes in the overall protein structure. We also find that the axial cysteine C317 may be substituted with any other amino acid without abrogating folding and heme cofactor incorporation. Several of the axial mutants display unusual spectral features, suggesting that they have active sites with unique steric and electronic properties. These novel, highly stable enzyme active sites will be fruitful starting points for investigations of non-natural P450 catalysis and mechanisms.

# Genetically Encoded Spy Peptide Fusion System to Detect Plasma Membrane-Localized Proteins In Vivo.

2015

Chemistry & biology

Bedbrook, Claire N

Kato, Mihoko

Ravindra Kumar, Sripriya

Lakshmanan, Anupama

Nath, Ravi D

Sun, Fei

Sternberg, Paul W

Arnold, Frances H

Gradinaru, Viviana

Membrane proteins are the main gatekeepers of cellular state, especially in neurons, serving either to maintain homeostasis or instruct response to synaptic input or other external signals. Visualization of membrane protein localization and trafficking in live cells facilitates understanding the molecular basis of cellular dynamics. We describe here a method for specifically labeling the plasma membrane-localized fraction of heterologous membrane protein expression using channelrhodopsins as a case study. We show that the genetically encoded, covalent binding SpyTag and SpyCatcher pair from the Streptococcus pyogenes fibronectin-binding protein FbaB can selectively label membrane-localized proteins in living cells in culture and in vivo in Caenorhabditis elegans. The SpyTag/SpyCatcher covalent labeling method is highly specific, modular, and stable in living cells. We have used the binding pair to develop a channelrhodopsin membrane localization assay that is amenable to high-throughput screening for opsin discovery and engineering.

# Recent advances in engineering microbial rhodopsins for optogenetics.

2015

Current opinion in structural biology

McIsaac, R Scott

Bedbrook, Claire N

Arnold, Frances H

Protein engineering of microbial rhodopsins has been successful in generating variants with improved properties for applications in optogenetics. Members of this membrane protein family can act as both actuators and sensors of neuronal activity. Chimeragenesis, structure-guided mutagenesis, and directed evolution have proven effective strategies for tuning absorption wavelength, altering ion specificity and increasing fluorescence. These approaches facilitate the development of useful optogenetic tools and, in some cases, have yielded insights into rhodopsin structure-function relationships.

# Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases.

2015

The Biochemical journal

Cahn, Jackson K B

Brinkmann-Chen, Sabine

Spatzal, Thomas

Wiig, Jared A

Buller, Andrew R

Einsle, Oliver

Hu, Yilin

Ribbe, Markus W

Arnold, Frances H

Although most sequenced members of the industrially important ketol-acid reductoisomerase (KARI) family are class I enzymes, structural studies to date have focused primarily on the class II KARIs, which arose through domain duplication. In the present study, we present five new crystal structures of class I KARIs. These include the first structure of a KARI with a six-residue beta2alphaB (cofactor specificity determining) loop and an NADPH phosphate-binding geometry distinct from that of the seven- and 12-residue loops. We also present the first structures of naturally occurring KARIs that utilize NADH as cofactor. These results show insertions in the specificity loops that confounded previous attempts to classify them according to loop length. Lastly, we explore the conformational changes that occur in class I KARIs upon binding of cofactor and metal ions. The class I KARI structures indicate that the active sites close upon binding NAD(P)H, similar to what is observed in the class II KARIs of rice and spinach and different from the opening of the active site observed in the class II KARI of Escherichia coli. This conformational change involves a decrease in the bending of the helix that runs between the domains and a rearrangement of the nicotinamide-binding site.

# Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution.

2015

Angewandte Chemie (International ed. in English)

Renata, Hans

Wang, Z Jane

Arnold, Frances H

High selectivity and exquisite control over the outcome of reactions entice chemists to use biocatalysts in organic synthesis. However, many useful reactions are not accessible because they are not in nature's known repertoire. In this Review, we outline an evolutionary approach to engineering enzymes to catalyze reactions not found in nature. We begin with examples of how nature has discovered new catalytic functions and how such evolutionary progression has been recapitulated in the laboratory starting from extant enzymes. We then examine non-native enzyme activities that have been exploited for chemical synthesis, with an emphasis on reactions that do not have natural counterparts. Non-natural activities can be improved by directed evolution, thus mimicking the process used by nature to create new catalysts. Finally, we describe the discovery of non-native catalytic functions that may provide future opportunities for the expansion of the enzyme universe.

# Enzyme-controlled nitrogen-atom transfer enables regiodivergent C-H amination.

2014

Journal of the American Chemical Society

Hyster, Todd K

Farwell, Christopher C

Buller, Andrew R

McIntosh, John A

Arnold, Frances H

We recently demonstrated that variants of cytochrome P450BM3 (CYP102A1) catalyze the insertion of nitrogen species into benzylic C-H bonds to form new C-N bonds. An outstanding challenge in the field of C-H amination is catalyst-controlled regioselectivity. Here, we report two engineered variants of P450BM3 that provide divergent regioselectivity for C-H amination-one favoring amination of benzylic C-H bonds and the other favoring homo-benzylic C-H bonds. The two variants provide nearly identical kinetic isotope effect values (2.8-3.0), suggesting that C-H abstraction is rate-limiting. The 2.66-A crystal structure of the most active enzyme suggests that the engineered active site can preorganize the substrate for reactivity. We hypothesize that the enzyme controls regioselectivity through localization of a single C-H bond close to the iron nitrenoid.

# Non-natural olefin cyclopropanation catalyzed by diverse cytochrome P450s and other hemoproteins.

2014

Chembiochem : a European journal of chemical biology

Heel, Thomas

McIntosh, John A

Dodani, Sheel C

Meyerowitz, Joseph T

Arnold, Frances H

Recent work has shown that engineered variants of cytochrome P450BM3 (CYP102A1) efficiently catalyze non-natural reactions, including carbene and nitrene transfer reactions. Given the broad substrate range of natural P450 enzymes, we set out to explore if this diversity could be leveraged to generate a broad panel of new catalysts for olefin cyclopropanation (i.e., carbene transfer). Here, we took a step towards this goal by characterizing the carbene transfer activities of four new wild-type P450s that have different native substrates. All four were active and exhibited a range of product selectivities in the model reaction: cyclopropanation of styrene by using ethyl diazoacetate (EDA). Previous work on P450BM3 demonstrated that mutation of the axial coordinating cysteine, universally conserved among P450 enzymes, to a serine residue, increased activity for this non-natural reaction. The equivalent mutation in the selected P450s was found to activate carbene transfer chemistry both in vitro and in vivo. Furthermore, serum albumins complexed with hemin were also found to be efficient in vitro cyclopropanation catalysts.

# Archaerhodopsin variants with enhanced voltage-sensitive fluorescence in mammalian and Caenorhabditis elegans neurons.

2014

Nature communications

Flytzanis, Nicholas C

Bedbrook, Claire N

Chiu, Hui

Engqvist, Martin K M

Xiao, Cheng

Chan, Ken Y

Sternberg, Paul W

Arnold, Frances H

Gradinaru, Viviana

Probing the neural circuit dynamics underlying behaviour would benefit greatly from improved genetically encoded voltage indicators. The proton pump Archaerhodopsin-3 (Arch), an optogenetic tool commonly used for neuronal inhibition, has been shown to emit voltage-sensitive fluorescence. Here we report two Arch variants with enhanced radiance (Archers) that in response to 655 nm light have 3-5 times increased fluorescence and 55-99 times reduced photocurrents compared with Arch WT. The most fluorescent variant, Archer1, has 25-40% fluorescence change in response to action potentials while using 9 times lower light intensity compared with other Arch-based voltage sensors. Archer1 is capable of wavelength-specific functionality as a voltage sensor under red light and as an inhibitory actuator under green light. As a proof-of-concept for the application of Arch-based sensors in vivo, we show fluorescence voltage sensing in behaving Caenorhabditis elegans. Archer1's characteristics contribute to the goal of all-optical detection and modulation of activity in neuronal networks in vivo.

# P450-catalyzed asymmetric cyclopropanation of electron-deficient olefins under aerobic conditions.

2014

Catalysis science & technology

Renata, Hans

Wang, Z Jane

Kitto, Rebekah Z

Arnold, Frances H

A variant of P450 from Bacillus megaterium five mutations away from wild type is a highly active catalyst for cyclopropanation of a variety of acrylamide and acrylate olefins with ethyl diazoacetate (EDA). The very high rate of reaction enabled by histidine ligation allowed the reaction to be conducted under aerobic conditions. The promiscuity of this catalyst for a variety of substrates containing amides has enabled synthesis of a small library of precursors to levomilnacipran derivatives.

# Structural, functional, and spectroscopic characterization of the substrate scope of the novel nitrating cytochrome P450 TxtE.

2014

Chembiochem : a European journal of chemical biology

Dodani, Sheel C

Cahn, Jackson K B

Heinisch, Tillmann

Brinkmann-Chen, Sabine

McIntosh, John A

Arnold, Frances H

A novel cytochrome P450 enzyme, TxtE, was recently shown to catalyze the direct aromatic nitration of L-tryptophan. This unique chemistry inspired us to ask whether TxtE could serve as a platform for engineering new nitration biocatalysts to replace current harsh synthetic methods. As a first step toward this goal, and to better understand the wild-type enzyme, we obtained high-resolution structures of TxtE in its substrate-free and substrate-bound forms. We also screened a library of substrate analogues for spectroscopic indicators of binding and for production of nitrated products. From these results, we found that the wild-type enzyme accepts moderate decoration of the indole ring, but the amino acid moiety is crucial for binding and correct positioning of the substrate and therefore less amenable to modification. A nitrogen atom is essential for catalysis, and a carbonyl must be present to recruit the alphaB'1 helix of the protein to seal the binding pocket.

# Directed evolution of a far-red fluorescent rhodopsin.

2014

Proceedings of the National Academy of Sciences of the United States of America

McIsaac, R Scott

Engqvist, Martin K M

Wannier, Timothy

Rosenthal, Adam Z

Herwig, Lukas

Flytzanis, Nicholas C

Imasheva, Eleonora S

Lanyi, Janos K

Balashov, Sergei P

Gradinaru, Viviana

Arnold, Frances H

Microbial rhodopsins are a diverse group of photoactive transmembrane proteins found in all three domains of life. A member of this protein family, Archaerhodopsin-3 (Arch) of halobacterium Halorubrum sodomense, was recently shown to function as a fluorescent indicator of membrane potential when expressed in mammalian neurons. Arch fluorescence, however, is very dim and is not optimal for applications in live-cell imaging. We used directed evolution to identify mutations that dramatically improve the absolute brightness of Arch, as confirmed biochemically and with live-cell imaging (in Escherichia coli and human embryonic kidney 293 cells). In some fluorescent Arch variants, the pK(a) of the protonated Schiff-base linkage to retinal is near neutral pH, a useful feature for voltage-sensing applications. These bright Arch variants enable labeling of biological membranes in the far-red/infrared and exhibit the furthest red-shifted fluorescence emission thus far reported for a fluorescent protein (maximal excitation/emission at approximately 620 nm/730 nm).

# Noncontiguous SCHEMA protein recombination.

2014

Methods in molecular biology (Clifton, N.J.)

Smith, Matthew A

Arnold, Frances H

SCHEMA is a method of designing protein recombination libraries that contain a large fraction of functional proteins with a high degree of mutational diversity. In the previous chapter, we illustrated the method for designing libraries by swapping contiguous sequence elements. Here, we introduce the NCR ("noncontiguous recombination") algorithm to identify optimal designs for swapping elements that are contiguous in the 3-D structure but not necessarily in the primary sequence. To exemplify the method, NCR is used to recombine three fungal cellobiohydrolases (CBH1s) to produce a library containing more than 500,000 novel chimeric sequences.

# Designing libraries of chimeric proteins using SCHEMA recombination and RASPP.

2014

Methods in molecular biology (Clifton, N.J.)

Smith, Matthew A

Arnold, Frances H

SCHEMA is a method for designing libraries of novel proteins by recombination of homologous sequences. The goal is to maximize the number of folded proteins while simultaneously generating significant sequence diversity. Here, we use the RASPP algorithm to identify optimal SCHEMA designs for shuffling contiguous elements of sequence. To exemplify the method, SCHEMA is used to recombine five fungal cellobiohydrolases (CBH1s) to produce a library of more than 390,000 novel CBH1 sequences.

# Synthesis of bioactive protein hydrogels by genetically encoded SpyTag-SpyCatcher chemistry.

2014

Proceedings of the National Academy of Sciences of the United States of America

Sun, Fei

Zhang, Wen-Bin

Mahdavi, Alborz

Arnold, Frances H

Tirrell, David A

Protein-based hydrogels have emerged as promising alternatives to synthetic hydrogels for biomedical applications, owing to the precise control of structure and function enabled by protein engineering. Nevertheless, strategies for assembling 3D molecular networks that carry the biological information encoded in full-length proteins remain underdeveloped. Here we present a robust protein gelation strategy based on a pair of genetically encoded reactive partners, SpyTag and SpyCatcher, that spontaneously form covalent isopeptide linkages under physiological conditions. The resulting "network of Spies" may be designed to include cell-adhesion ligands, matrix metalloproteinase-1 cleavage sites, and full-length globular proteins [mCherry and leukemia inhibitory factor (LIF)]. The LIF network was used to encapsulate mouse embryonic stem cells; the encapsulated cells remained pluripotent in the absence of added LIF. These results illustrate a versatile strategy for the creation of information-rich biomaterials.

# Directed evolution of Gloeobacter violaceus rhodopsin spectral properties.

2014

Journal of molecular biology

Engqvist, Martin K M

McIsaac, R Scott

Dollinger, Peter

Flytzanis, Nicholas C

Abrams, Michael

Schor, Stanford

Arnold, Frances H

Proton-pumping rhodopsins (PPRs) are photoactive retinal-binding proteins that transport ions across biological membranes in response to light. These proteins are interesting for light-harvesting applications in bioenergy production, in optogenetics applications in neuroscience, and as fluorescent sensors of membrane potential. Little is known, however, about how the protein sequence determines the considerable variation in spectral properties of PPRs from different biological niches or how to engineer these properties in a given PPR. Here we report a comprehensive study of amino acid substitutions in the retinal-binding pocket of Gloeobacter violaceus rhodopsin (GR) that tune its spectral properties. Directed evolution generated 70 GR variants with absorption maxima shifted by up to +/-80nm, extending the protein's light absorption significantly beyond the range of known natural PPRs. While proton-pumping activity was disrupted in many of the spectrally shifted variants, we identified single tuning mutations that incurred blue and red shifts of 42nm and 22nm, respectively, that did not disrupt proton pumping. Blue-shifting mutations were distributed evenly along the retinal molecule while red-shifting mutations were clustered near the residue K257, which forms a covalent bond with retinal through a Schiff base linkage. Thirty eight of the identified tuning mutations are not found in known microbial rhodopsins. We discovered a subset of red-shifted GRs that exhibit high levels of fluorescence relative to the WT (wild-type) protein.

# Engineered thermostable fungal cellulases exhibit efficient synergistic cellulose hydrolysis at elevated temperatures.

2014

Biotechnology and bioengineering

Trudeau, Devin L

Lee, Toni M

Arnold, Frances H

A major obstacle to using widely available and low-cost lignocellulosic feedstocks to produce renewable fuels and chemicals is the high cost and low efficiency of the enzyme mixtures used to hydrolyze cellulose to fermentable sugars. One possible solution entails engineering current cellulases to function efficiently at elevated temperatures in order to boost reaction rates and exploit several other advantages of a higher temperature process. Here, we describe the creation of the most stable reported fungal endoglucanase, a derivative of Hypocrea jecorina (anamorph Trichoderma reesei) Cel5A, by combining stabilizing mutations identified using consensus design, chimera studies, and structure-based computational methods. The engineered endoglucanase has an optimal temperature that is 17 degrees C higher than wild type H. jecorina Cel5A, and hydrolyzes 1.5 times as much cellulose over 60 h at its optimum temperature compared to the wild type enzyme at its optimal temperature. This enzyme complements previously engineered highly active, thermostable variants of the fungal cellobiohydrolases Cel6A and Cel7A in a thermostable cellulase mixture that hydrolyzes cellulose synergistically at an optimum temperature of 70 degrees C over 60 h.The thermostable mixture produces three times as much total sugar as the best mixture of the wild type enzymes operating at its optimum temperature of 60 degrees C, clearly demonstrating the advantage of higher temperature cellulose hydrolysis.

# Enantioselective imidation of sulfides via enzyme-catalyzed intermolecular nitrogen-atom transfer.

2014

Journal of the American Chemical Society

Farwell, Christopher C

McIntosh, John A

Hyster, Todd K

Wang, Z Jane

Arnold, Frances H

Engineering enzymes with novel reaction modes promises to expand the applications of biocatalysis in chemical synthesis and will enhance our understanding of how enzymes acquire new functions. The insertion of nitrogen-containing functional groups into unactivated C-H bonds is not catalyzed by known enzymes but was recently demonstrated using engineered variants of cytochrome P450BM3 (CYP102A1) from Bacillus megaterium. Here, we extend this novel P450-catalyzed reaction to include intermolecular insertion of nitrogen into thioethers to form sulfimides. An examination of the reactivity of different P450BM3 variants toward a range of substrates demonstrates that electronic properties of the substrates are important in this novel enzyme-catalyzed reaction. Moreover, amino acid substitutions have a large effect on the rate and stereoselectivity of sulfimidation, demonstrating that the protein plays a key role in determining reactivity and selectivity. These results provide a stepping stone for engineering more complex nitrogen-atom-transfer reactions in P450 enzymes and developing a more comprehensive biocatalytic repertoire.

# Synthetic biology: Engineering explored.

2014

Nature

Silver, Pamela A

Way, Jeffrey C

Arnold, Frances H

Meyerowitz, Joseph T

# Improved cyclopropanation activity of histidine-ligated cytochrome P450 enables the enantioselective formal synthesis of levomilnacipran.

2014

Angewandte Chemie (International ed. in English)

Wang, Z Jane

Renata, Hans

Peck, Nicole E

Farwell, Christopher C

Coelho, Pedro S

Arnold, Frances H

Engineering enzymes capable of modes of activation unprecedented in nature will increase the range of industrially important molecules that can be synthesized through biocatalysis. However, low activity for a new function is often a limitation in adopting enzymes for preparative-scale synthesis, reaction with demanding substrates, or when a natural substrate is also present. By mutating the proximal ligand and other key active-site residues of the cytochrome P450 enzyme from Bacillus megaterium (P450-BM3), a highly active His-ligated variant of P450-BM3 that can be employed for the enantioselective synthesis of the levomilnacipran core was engineered. This enzyme, BM3-Hstar, catalyzes the cyclopropanation of N,N-diethyl-2-phenylacrylamide with an estimated initial rate of over 1000 turnovers per minute and can be used under aerobic conditions. Cyclopropanation activity is highly dependent on the electronic properties of the P450 proximal ligand, which can be used to tune this non-natural enzyme activity.

# Isobutanol production at elevated temperatures in thermophilic Geobacillus thermoglucosidasius.

2014

Metabolic engineering

Lin, Paul P

Rabe, Kersten S

Takasumi, Jennifer L

Kadisch, Marvin

Arnold, Frances H

Liao, James C

The potential advantages of biological production of chemicals or fuels from biomass at high temperatures include reduced enzyme loading for cellulose degradation, decreased chance of contamination, and lower product separation cost. In general, high temperature production of compounds that are not native to the thermophilic hosts is limited by enzyme stability and the lack of suitable expression systems. Further complications can arise when the pathway includes a volatile intermediate. Here we report the engineering of Geobacillus thermoglucosidasius to produce isobutanol at 50 degrees C. We prospected various enzymes in the isobutanol synthesis pathway and characterized their thermostabilities. We also constructed an expression system based on the lactate dehydrogenase promoter from Geobacillus thermodenitrificans. With the best enzyme combination and the expression system, 3.3g/l of isobutanol was produced from glucose and 0.6g/l of isobutanol from cellobiose in G. thermoglucosidasius within 48h at 50 degrees C. This is the first demonstration of isobutanol production in recombinant bacteria at an elevated temperature.

# Expanding P450 catalytic reaction space through evolution and engineering.

2014

Current opinion in chemical biology

McIntosh, John A

Farwell, Christopher C

Arnold, Frances H

Advances in protein and metabolic engineering have led to wider use of enzymes to synthesize important molecules. However, many desirable transformations are not catalyzed by any known enzyme, driving interest in understanding how new enzymes can be created. The cytochrome P450 enzyme family, whose members participate in xenobiotic metabolism and natural products biosynthesis, catalyzes an impressive range of difficult chemical reactions that continues to grow as new enzymes are characterized. Recent work has revealed that P450-derived enzymes can also catalyze useful reactions previously accessible only to synthetic chemistry. The evolution and engineering of these enzymes provides an excellent case study for how to genetically encode new chemistry and expand biology's reaction space.

# Cytochrome P450-Catalyzed Insertion of Carbenoids into N-H Bonds.

2014

Chemical science

Wang, Z Jane

Peck, Nicole E

Renata, Hans

Arnold, Frances H

Expanding nature's catalytic repertoire to include reactions important in synthetic chemistry will open new opportunities for 'green' chemistry and biosynthesis. We demonstrate enzyme-catalyzed insertion of carbenoids into N-H bonds. This type of bond disconnection, which has no counterpart in nature, can be mediated by variants of the cytochrome P450 from Bacillus megaterium. The N-H insertion reaction takes place in water, provides the desired products in 26-83% yield, forms the single addition product exclusively, and does not require slow addition of the diazo component.

# Innovation by homologous recombination.

2013

Current opinion in chemical biology

Trudeau, Devin L

Smith, Matthew A

Arnold, Frances H

Swapping fragments among protein homologs can produce chimeric proteins with a wide range of properties, including properties not exhibited by the parents. Computational methods that use information from structures and sequence alignments have been used to design highly functional chimeras and chimera libraries. Recombination has generated proteins with diverse thermostability and mechanical stability, enzyme substrate specificity, and optogenetic properties. Linear regression, Gaussian processes, and support vector machine learning have been used to model sequence-function relationships and predict useful chimeras. These approaches enable engineering of protein chimeras with desired functions, as well as elucidation of the structural basis for these functions.

# Controlling macromolecular topology with genetically encoded SpyTag-SpyCatcher chemistry.

2013

Journal of the American Chemical Society

Zhang, Wen-Bin

Sun, Fei

Tirrell, David A

Arnold, Frances H

Control of molecular topology constitutes a fundamental challenge in macromolecular chemistry. Here we describe the synthesis and characterization of artificial elastin-like proteins (ELPs) with unconventional nonlinear topologies including circular, tadpole, star, and H-shaped proteins using genetically encoded SpyTag-SpyCatcher chemistry. SpyTag is a short polypeptide that binds its protein partner SpyCatcher and forms isopeptide bonds under physiological conditions. Sequences encoding SpyTag and SpyCatcher can be strategically placed into ELP genes to direct post-translational topological modification in situ. Placement of SpyTag at the N-terminus and SpyCatcher at the C-terminus directs formation of circular ELPs. Induction of expression at 16 degrees C with 10 muM IPTG yields 80% monomeric cyclic protein. When SpyTag is placed in the middle of the chain, it exhibits an even stronger tendency toward cyclization, yielding up to 94% monomeric tadpole proteins. Telechelic ELPs containing either SpyTag or SpyCatcher can be expressed, purified, and then coupled spontaneously upon mixing in vitro. Block proteins, 3-arm or 4-arm star proteins, and H-shaped proteins have been prepared, with the folded CnaB2 domain that results from the SpyTag-SpyCatcher reaction as the molecular core or branch junction. The modular character of the SpyTag-SpyCatcher strategy should make it useful for preparing nonlinear macromolecules of diverse sequence and structure.

# Enantioselective intramolecular C-H amination catalyzed by engineered cytochrome P450 enzymes in vitro and in vivo.

2013

Angewandte Chemie (International ed. in English)

McIntosh, John A

Coelho, Pedro S

Farwell, Christopher C

Wang, Z Jane

Lewis, Jared C

Brown, Tristan R

Arnold, Frances H

# A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo.

2013

Nature chemical biology

Coelho, Pedro S

Wang, Z Jane

Ener, Maraia E

Baril, Stefanie A

Kannan, Arvind

Arnold, Frances H

Brustad, Eric M

Whole-cell catalysts for non-natural chemical reactions will open new routes to sustainable production of chemicals. We designed a cytochrome 'P411' with unique serine-heme ligation that catalyzes efficient and selective olefin cyclopropanation in intact Escherichia coli cells. The mutation C400S in cytochrome P450(BM3) gives a signature ferrous CO Soret peak at 411 nm, abolishes monooxygenation activity, raises the resting-state Fe(III)-to-Fe(II) reduction potential and substantially improves NAD(P)H-driven activity.

# General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH.

2013

Proceedings of the National Academy of Sciences of the United States of America

Brinkmann-Chen, Sabine

Flock, Tilman

Cahn, Jackson K B

Snow, Christopher D

Brustad, Eric M

McIntosh, John A

Meinhold, Peter

Zhang, Liang

Arnold, Frances H

To date, efforts to switch the cofactor specificity of oxidoreductases from nicotinamide adenine dinucleotide phosphate (NADPH) to nicotinamide adenine dinucleotide (NADH) have been made on a case-by-case basis with varying degrees of success. Here we present a straightforward recipe for altering the cofactor specificity of a class of NADPH-dependent oxidoreductases, the ketol-acid reductoisomerases (KARIs). Combining previous results for an engineered NADH-dependent variant of Escherichia coli KARI with available KARI crystal structures and a comprehensive KARI-sequence alignment, we identified key cofactor specificity determinants and used this information to construct five KARIs with reversed cofactor preference. Additional directed evolution generated two enzymes having NADH-dependent catalytic efficiencies that are greater than the wild-type enzymes with NADPH. High-resolution structures of a wild-type/variant pair reveal the molecular basis of the cofactor switch.

# Hypocrea jecorina cellobiohydrolase I stabilizing mutations identified using noncontiguous recombination.

2013

ACS synthetic biology

Smith, Matthew A

Bedbrook, Claire N

Wu, Timothy

Arnold, Frances H

Noncontiguous recombination (NCR) is a method to identify pieces of structure that can be swapped among homologous proteins to create new, chimeric proteins. These "blocks" are encoded by elements of sequence that are not necessarily contiguous along the polypeptide chain. We used NCR to design a library in which blocks of structure from Hypocrea jecorina cellobiohydrolase I (Cel7A) and its two thermostable homologues from Talaromyces emersonii and Chaetomium thermophilum are shuffled to create 531,438 possible chimeric enzymes. We constructed a maximally informative subset of 35 chimeras to analyze this library and found that the blocks contribute additively to the stability of a chimera. Within two highly stabilizing blocks, we uncovered six single amino acid substitutions that each improve the stability of H. jecorina cellobiohydrolase I by 1-3 degrees C. The small number of measurements required to find these mutations demonstrates that noncontiguous recombination is an efficient strategy for identifying stabilizing mutations.

# Role of cysteine residues in thermal inactivation of fungal Cel6A cellobiohydrolases.

2013

Biochimica et biophysica acta

Wu, Indira

Heel, Thomas

Arnold, Frances H

Numerous protein engineering studies have focused on increasing the thermostability of fungal cellulases to improve production of fuels and chemicals from lignocellulosic feedstocks. However, the engineered enzymes still undergo thermal inactivation at temperatures well below the inactivation temperatures of hyperthermophilic cellulases. In this report, we investigated the role of free cysteines in the thermal inactivation of wild-type and engineered fungal family 6 cellobiohydrolases (Cel6A). The mechanism of thermal inactivation of Cel6A is consistent with disulfide bond degradation and thiol-disulfide exchange. Circular dichroism spectroscopy revealed that a thermostable variant lacking free cysteines refolds to a native-like structure and retains activity after heat treatment over the pH range 5-9. Whereas conserved disulfide bonds are essential for retaining activity after heat treatment, free cysteines contribute to irreversible thermal inactivation in engineered thermostable Cel6A as well as Cel6A from Hypocrea jecorina and Humicola insolens.

# High-throughput screening for terpene-synthase-cyclization activity and directed evolution of a terpene synthase.

2013

Angewandte Chemie (International ed. in English)

Lauchli, Ryan

Rabe, Kersten S

Kalbarczyk, Karolina Z

Tata, Amulya

Heel, Thomas

Kitto, Rebekah Z

Arnold, Frances H

# Directed evolution of protein-based neurotransmitter sensors for MRI.

2013

Methods in molecular biology (Clifton, N.J.)

Romero, Philip A

Shapiro, Mikhail G

Arnold, Frances H

Jasanoff, Alan

The production of contrast agents sensitive to neuronal signaling events is a rate-limiting step in the development of molecular-level functional magnetic resonance imaging (molecular fMRI) approaches for studying the brain. High-throughput generation and evaluation of potential probes are possible using techniques for macromolecular engineering of protein-based contrast agents. In an initial exploration of this strategy, we used the method of directed evolution to identify mutants of a bacterial heme protein that allowed detection of the neurotransmitter dopamine in vitro and in living animals. The directed evolution method involves successive cycles of mutagenesis and screening that could be generalized to produce contrast agents sensitive to a variety of molecular targets in the nervous system.

# Efficient sampling of SCHEMA chimera families to identify useful sequence elements.

2013

Methods in enzymology

Heinzelman, Pete

Romero, Philip A

Arnold, Frances H

SCHEMA structure-guided recombination is an effective method for producing families of protein chimeras having high sequence diversity, functional diversity, and thermostabilities greater than any of the parent proteins from which the chimeras are made. A key feature of SCHEMA chimera families is their amenability to a "sample, model, and predict" operation that allows one to characterize members of a small chimera sample set and use those data to construct models that accurately predict the properties of every member of the family. In this chapter, we describe applications of this "sample, model, and predict" approach and outline methods for designing chimera sample sets that enable efficient construction of models to identify useful sequence elements. With these models we can also predict the sequences and properties of the most desirable chimeras.

# Engineered thermostable fungal Cel6A and Cel7A cellobiohydrolases hydrolyze cellulose efficiently at elevated temperatures.

2013

Biotechnology and bioengineering

Wu, Indira

Arnold, Frances H

Thermostability is an important feature in industrial enzymes: it increases biocatalyst lifetime and enables reactions at higher temperatures, where faster rates and other advantages ultimately reduce the cost of biocatalysis. Here we report the thermostabilization of a chimeric fungal family 6 cellobiohydrolase (HJPlus) by directed evolution using random mutagenesis and recombination of beneficial mutations. Thermostable variant 3C6P has a half-life of 280 min at 75 degrees C and a T(50) of 80.1 degrees C, a ~15 degrees C increase over the thermostable Cel6A from Humicola insolens (HiCel6A) and a ~20 degrees C increase over that from Hypocrea jecorina (HjCel6A). Most of the mutations also stabilize the less-stable HjCel6A, the wild-type Cel6A closest in sequence to 3C6P. During a 60-h Avicel hydrolysis, 3C6P released 2.4 times more cellobiose equivalents at its optimum temperature (T(opt)) of 75 degrees C than HiCel6A at its T(opt) of 60 degrees C. The total cellobiose equivalents released by HiCel6A at 60 degrees C after 60 h is equivalent to the total released by 3C6P at 75 degrees C after ~6 h, a 10-fold reduction in hydrolysis time. A binary mixture of thermostable Cel6A and Cel7A hydrolyzes Avicel synergistically and released 1.8 times more cellobiose equivalents than the wild-type mixture, both mixtures assessed at their respective T(opt). Crystal structures of HJPlus and 3C6P, determined at 1.5 and 1.2 A resolution, indicate that the stabilization comes from improved hydrophobic interactions and restricted loop conformations by introduced proline residues.

# Navigating the protein fitness landscape with Gaussian processes.

2013

Proceedings of the National Academy of Sciences of the United States of America

Romero, Philip A

Krause, Andreas

Arnold, Frances H

Knowing how protein sequence maps to function (the "fitness landscape") is critical for understanding protein evolution as well as for engineering proteins with new and useful properties. We demonstrate that the protein fitness landscape can be inferred from experimental data, using Gaussian processes, a Bayesian learning technique. Gaussian process landscapes can model various protein sequence properties, including functional status, thermostability, enzyme activity, and ligand binding affinity. Trained on experimental data, these models achieve unrivaled quantitative accuracy. Furthermore, the explicit representation of model uncertainty allows for efficient searches through the vast space of possible sequences. We develop and test two protein sequence design algorithms motivated by Bayesian decision theory. The first one identifies small sets of sequences that are informative about the landscape; the second one identifies optimized sequences by iteratively improving the Gaussian process model in regions of the landscape that are predicted to be optimized. We demonstrate the ability of Gaussian processes to guide the search through protein sequence space by designing, constructing, and testing chimeric cytochrome P450s. These algorithms allowed us to engineer active P450 enzymes that are more thermostable than any previously made by chimeragenesis, rational design, or directed evolution.

# Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes.

2012

Science (New York, N.Y.)

Coelho, Pedro S

Brustad, Eric M

Kannan, Arvind

Arnold, Frances H

Transition metal-catalyzed transfers of carbenes, nitrenes, and oxenes are powerful methods for functionalizing C=C and C-H bonds. Nature has evolved a diverse toolbox for oxene transfers, as exemplified by the myriad monooxygenation reactions catalyzed by cytochrome P450 enzymes. The isoelectronic carbene transfer to olefins, a widely used C-C bond-forming reaction in organic synthesis, has no biological counterpart. Here we report engineered variants of cytochrome P450(BM3) that catalyze highly diastereo- and enantioselective cyclopropanation of styrenes from diazoester reagents via putative carbene transfer. This work highlights the capacity to adapt existing enzymes for the catalysis of synthetically important reactions not previously observed in nature.

# Chimeragenesis of distantly-related proteins by noncontiguous recombination.

2012

Protein science : a publication of the Protein Society

Smith, Matthew A

Romero, Philip A

Wu, Timothy

Brustad, Eric M

Arnold, Frances H

We introduce a method for identifying elements of a protein structure that can be shuffled to make chimeric proteins from two or more homologous parents. Formulating recombination as a graph-partitioning problem allows us to identify noncontiguous segments of the sequence that should be inherited together in the progeny proteins. We demonstrate this noncontiguous recombination approach by constructing a chimera of beta-glucosidases from two different kingdoms of life. Although the protein's alpha-beta barrel fold has no obvious subdomains for recombination, noncontiguous SCHEMA recombination generated a functional chimera that takes approximately half its structure from each parent. The X-ray crystal structure shows that the structural blocks that make up the chimera maintain the backbone conformations found in their respective parental structures. Although the chimera has lower beta-glucosidase activity than the parent enzymes, the activity was easily recovered by directed evolution. This simple method, which does not rely on detailed atomic models, can be used to design chimeras that take structural, and functional, elements from distantly-related proteins.

# A diverse set of family 48 bacterial glycoside hydrolase cellulases created by structure-guided recombination.

2012

The FEBS journal

Smith, Matthew A

Rentmeister, Andrea

Snow, Christopher D

Wu, Timothy

Farrow, Mary F

Mingardon, Florence

Arnold, Frances H

Sequence diversity within a family of functional enzymes provides a platform for elucidating structure-function relationships and for protein engineering to improve properties important for applications. Access to nature's vast sequence diversity is often limited by the fact that only a few enzymes have been characterized in a given family. Here, we recombined the catalytic domains of three glycoside hydrolase family 48 bacterial cellulases (Cel48; EC 3.2.1.176) - Clostridium cellulolyticum CelF, Clostridium stercorarium CelY, and Clostridium thermocellum CelS - to create a diverse library of Cel48 enzymes with an average of 106 mutations from the closest native enzyme. Within this set, we found large variations in properties such as the functional temperature range, stability, and specific activity on crystalline cellulose. We showed that functional status and stability were predictable from simple linear models of the sequence-property data: recombined protein fragments contributed additively to these properties in a given chimera. Using this, we correctly predicted sequences that were as stable as any of the native Cel48 enzymes described to date. The characterization of 60 active Cel48 chimeras expands the number of characterized Cel48 enzymes from 13 to 73. Our work illustrates the role that structure-guided recombination can play in helping to identify sequence-function relationships within a family of enzymes by supplementing natural diversity with synthetic diversity.

# Random field model reveals structure of the protein recombinational landscape.

2012

PLoS computational biology

Romero, Philip A

Arnold, Frances H

We are interested in how intragenic recombination contributes to the evolution of proteins and how this mechanism complements and enhances the diversity generated by random mutation. Experiments have revealed that proteins are highly tolerant to recombination with homologous sequences (mutation by recombination is conservative); more surprisingly, they have also shown that homologous sequence fragments make largely additive contributions to biophysical properties such as stability. Here, we develop a random field model to describe the statistical features of the subset of protein space accessible by recombination, which we refer to as the recombinational landscape. This model shows quantitative agreement with experimental results compiled from eight libraries of proteins that were generated by recombining gene fragments from homologous proteins. The model reveals a recombinational landscape that is highly enriched in functional sequences, with properties dominated by a large-scale additive structure. It also quantifies the relative contributions of parent sequence identity, crossover locations, and protein fold to the tolerance of proteins to recombination. Intragenic recombination explores a unique subset of sequence space that promotes rapid molecular diversification and functional adaptation.

# Structure-guided engineering of Lactococcus lactis alcohol dehydrogenase LlAdhA for improved conversion of isobutyraldehyde to isobutanol.

2012

Journal of biotechnology

Liu, Xiang

Bastian, Sabine

Snow, Christopher D

Brustad, Eric M

Saleski, Tatyana E

Xu, Jian-He

Meinhold, Peter

Arnold, Frances H

We have determined the X-ray crystal structures of the NADH-dependent alcohol dehydrogenase LlAdhA from Lactococcus lactis and its laboratory-evolved variant LlAdhA(RE1) at 1.9A and 2.5A resolution, respectively. LlAdhA(RE1), which contains three amino acid mutations (Y50F, I212T, and L264V), was engineered to increase the microbial production of isobutanol (2-methylpropan-1-ol) from isobutyraldehyde (2-methylpropanal). Structural comparison of LlAdhA and LlAdhA(RE1) indicates that the enhanced activity on isobutyraldehyde stems from increases in the protein's active site size, hydrophobicity, and substrate access. Further structure-guided mutagenesis generated a quadruple mutant (Y50F/N110S/I212T/L264V), whose KM for isobutyraldehyde is approximately 17-fold lower and catalytic efficiency (kcat/KM) is approximately 160-fold higher than wild-type LlAdhA. Combining detailed structural information and directed evolution, we have achieved significant improvements in non-native alcohol dehydrogenase activity that will facilitate the production of next-generation fuels such as isobutanol from renewable resources.

# Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods.

2012

Protein engineering, design & selection : PEDS

Komor, Russell S

Romero, Philip A

Xie, Catherine B

Arnold, Frances H

Building on our previous efforts to generate thermostable chimeric fungal cellobiohydrolase I (CBH I, also known as Cel7A) cellulases by structure-guided recombination, we used FoldX and a 'consensus' sequence approach to identify individual mutations present in the five homologous parent CBH I enzymes which further stabilize the chimeras. Using the FoldX force field, we calculated the effect on DeltaG(Folding) of each candidate mutation in a number of CBH I structures and chose those predicted to be stabilizing in multiple structures. With an alignment of 41 CBH I sequences, we also used amino acid frequencies at each candidate position to calculate predicted effects on DeltaG(Folding). A combination of mutations chosen using these methods increased the T(50) of the most thermostable chimera by an additional 4.7 degrees C, to yield a CBH I with T(50) of 72.1 degrees C, which is 9.2 degrees C higher than that of the most stable native CBH I, from Talaromyces emersonii. This increased stability resulted in a 10 degrees C increase in the optimal temperature for activity, to 65 degrees C, and a 50% increase in total sugar production from crystalline cellulose at the optimal temperature, compared with native T.emersonii CBH I.

# SCHEMA-designed variants of human Arginase I and II reveal sequence elements important to stability and catalysis.

2012

ACS synthetic biology

Romero, Philip A

Stone, Everett

Lamb, Candice

Chantranupong, Lynne

Krause, Andreas

Miklos, Aleksandr E

Hughes, Randall A

Fechtel, Blake

Ellington, Andrew D

Arnold, Frances H

Georgiou, George

Arginases catalyze the divalent cation-dependent hydrolysis of L-arginine to urea and L-ornithine. There is significant interest in using arginase as a therapeutic antineogenic agent against L-arginine auxotrophic tumors and in enzyme replacement therapy for treating hyperargininemia. Both therapeutic applications require enzymes with sufficient stability under physiological conditions. To explore sequence elements that contribute to arginase stability we used SCHEMA-guided recombination to design a library of chimeric enzymes composed of sequence fragments from the two human isozymes Arginase I and II. We then developed a novel active learning algorithm that selects sequences from this library that are both highly informative and functional. Using high-throughput gene synthesis and our two-step active learning algorithm, we were able to rapidly create a small but highly informative set of seven enzymatically active chimeras that had an average variant distance of 40 mutations from the closest parent arginase. Within this set of sequences, linear regression was used to identify the sequence elements that contribute to the long-term stability of human arginase under physiological conditions. This approach revealed a striking correlation between the isoelectric point and the long-term stability of the enzyme to deactivation under physiological conditions.

# Structure-guided directed evolution of highly selective p450-based magnetic resonance imaging sensors for dopamine and serotonin.

2012

Journal of molecular biology

Brustad, Eric M

Lelyveld, Victor S

Snow, Christopher D

Crook, Nathan

Jung, Sang Taek

Martinez, Francisco M

Scholl, Timothy J

Jasanoff, Alan

Arnold, Frances H

New tools that allow dynamic visualization of molecular neural events are important for studying the basis of brain activity and disease. Sensors that permit ligand-sensitive magnetic resonance imaging (MRI) are useful reagents due to the noninvasive nature and good temporal and spatial resolution of MR methods. Paramagnetic metalloproteins can be effective MRI sensors due to the selectivity imparted by the protein active site and the ability to tune protein properties using techniques such as directed evolution. Here, we show that structure-guided directed evolution of the active site of the cytochrome P450-BM3 heme domain produces highly selective MRI probes with submicromolar affinities for small molecules. We report a new, high-affinity dopamine sensor as well as the first MRI reporter for serotonin, with which we demonstrate quantification of neurotransmitter release in vitro. We also present a detailed structural analysis of evolved cytochrome P450-BM3 heme domain lineages to systematically dissect the molecular basis of neurotransmitter binding affinity, selectivity, and enhanced MRI contrast activity in these engineered proteins.

# Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes.

2012

Protein engineering, design & selection : PEDS

Chen, Mike M Y

Snow, Christopher D

Vizcarra, Christina L

Mayo, Stephen L

Arnold, Frances H

Three semi-rational approaches, combinatorial site-saturation mutagenesis (CSSM) using a reduced amino acid set and two libraries based on C(orbit) and CRAM computational design algorithms targeting up to 10 active site residues, were used to engineer cytochrome P450 BM3 to demethylate dimethyl ether and hydroxylate propane and ethane. These small libraries (343-1028 variants) were all enriched with respect to the fraction functional and maximal activities compared with a random mutagenesis library and individual site-saturation libraries targeting the same residues. Despite high average amino acid substitution levels of 2.6, 5 and 7.5, the CSSM, C(orbit) and CRAM libraries had at least 75% of library members properly folded. Propane- and ethane-hydroxylating P450 BM3 variants were identified using all three mutagenesis approaches, with as few as two amino acid substitutions. The library designed using the CRAM algorithm, which sought to reduce the size of the binding pocket, produced both a higher number of active variants and variants supporting the greatest number of catalytic turnovers. The most active variant E32 supports 16 800 propane turnovers at 36% coupling, which rivals the activity of variants obtained after 10-12 rounds of directed evolution using random and site-saturation mutagenesis. None of the variants in this study achieved the complete re-specialization for propane hydroxylation (including 93% coupling) previously obtained via multiple rounds of mutagenesis and screening. However, these semi-rational approaches allowed for large jumps in sequence space to variants with the desired functions.

# Reversal of NAD(P)H cofactor dependence by protein engineering.

2011

Methods in molecular biology (Clifton, N.J.)

Bastian, Sabine

Arnold, Frances H

There is increasing interest in utilization of engineered microorganisms for the production of renewable chemicals and next-generation biofuels. However, imbalances between the cofactor consumption of the engineered production pathway and the reducing equivalents provided by the cell have been shown to limit yields. This imbalance can be overcome by adjusting the cofactor dependencies of the pathway enzymes to match the available cofactors in the cell. We show how cofactor preference can be reversed by structure-guided directed evolution of the target enzyme.

# A structural study of Hypocrea jecorina Cel5A.

2011

Protein science : a publication of the Protein Society

Lee, Toni M

Farrow, Mary F

Arnold, Frances H

Mayo, Stephen L

Interest in generating lignocellulosic biofuels through enzymatic hydrolysis continues to rise as nonrenewable fossil fuels are depleted. The high cost of producing cellulases, hydrolytic enzymes that cleave cellulose into fermentable sugars, currently hinders economically viable biofuel production. Here, we report the crystal structure of a prevalent endoglucanase in the biofuels industry, Cel5A from the filamentous fungus Hypocrea jecorina. The structure reveals a general fold resembling that of the closest homolog with a high-resolution structure, Cel5A from Thermoascus aurantiacus. Consistent with previously described endoglucanase structures, the H. jecorina Cel5A active site contains a primarily hydrophobic substrate binding groove and a series of hydrogen bond networks surrounding two catalytic glutamates. The reported structure, however, demonstrates stark differences between side-chain identity, loop regions, and the number of disulfides. Such structural information may aid efforts to improve the stability of this protein for industrial use while maintaining enzymatic activity through revealing nonessential and immutable regions.

# High throughput screening of fungal endoglucanase activity in Escherichia coli.

2011

Journal of visualized experiments : JoVE

Farrow, Mary F

Arnold, Frances H

Cellulase enzymes (endoglucanases, cellobiohydrolases, and beta-glucosidases) hydrolyze cellulose into component sugars, which in turn can be converted into fuel alcohols. The potential for enzymatic hydrolysis of cellulosic biomass to provide renewable energy has intensified efforts to engineer cellulases for economical fuel production. Of particular interest are fungal cellulases, which are already being used industrially for foods and textiles processing. Identifying active variants among a library of mutant cellulases is critical to the engineering process; active mutants can be further tested for improved properties and/or subjected to additional mutagenesis. Efficient engineering of fungal cellulases has been hampered by a lack of genetic tools for native organisms and by difficulties in expressing the enzymes in heterologous hosts. Recently, Morikawa and coworkers developed a method for expressing in E. coli the catalytic domains of endoglucanases from H. jecorina, an important industrial fungus with the capacity to secrete cellulases in large quantities. Functional E. coli expression has also been reported for cellulases from other fungi, including Macrophomina phaseolina and Phanerochaete chrysosporium. We present a method for high throughput screening of fungal endoglucanase activity in E. coli. This method uses the common microbial dye Congo Red (CR) to visualize enzymatic degradation of carboxymethyl cellulose (CMC) by cells growing on solid medium. The activity assay requires inexpensive reagents, minimal manipulation, and gives unambiguous results as zones of degradation ("halos") at the colony site. Although a quantitative measure of enzymatic activity cannot be determined by this method, we have found that halo size correlates with total enzymatic activity in the cell. Further characterization of individual positive clones will determine , relative protein fitness. Traditional bacterial whole cell CMC/CR activity assays involve pouring agar containing CMC onto colonies, which is subject to cross-contamination, or incubating cultures in CMC agar wells, which is less amenable to large-scale experimentation. Here we report an improved protocol that modifies existing wash methods for cellulase activity: cells grown on CMC agar plates are removed prior to CR staining. Our protocol significantly reduces cross-contamination and is highly scalable, allowing the rapid screening of thousands of clones. In addition to H. jecorina enzymes, we have expressed and screened endoglucanase variants from the Thermoascus aurantiacus and Penicillium decumbens, suggesting that this protocol is applicable to enzymes from a range of organisms.

# Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli.

2011

Metabolic engineering

Bastian, Sabine

Liu, Xiang

Meyerowitz, Joseph T

Snow, Christopher D

Chen, Mike M Y

Arnold, Frances H

2-methylpropan-1-ol (isobutanol) is a leading candidate biofuel for the replacement or supplementation of current fossil fuels. Recent work has demonstrated glucose to isobutanol conversion through a modified amino acid pathway in a recombinant organism. Although anaerobic conditions are required for an economically competitive process, only aerobic isobutanol production has been feasible due to an imbalance in cofactor utilization. Two of the pathway enzymes, ketol-acid reductoisomerase and alcohol dehydrogenase, require nicotinamide dinucleotide phosphate (NADPH); glycolysis, however, produces only nicotinamide dinucleotide (NADH). Here, we compare two solutions to this imbalance problem: (1) over-expression of pyridine nucleotide transhydrogenase PntAB and (2) construction of an NADH-dependent pathway, using engineered enzymes. We demonstrate that an NADH-dependent pathway enables anaerobic isobutanol production at 100% theoretical yield and at higher titer and productivity than both the NADPH-dependent pathway and transhydrogenase over-expressing strain. Our results show how engineering cofactor dependence can overcome a critical obstacle to next-generation biofuel commercialization.

# Cytochrome P450: taming a wild type enzyme.

2011

Current opinion in biotechnology

Jung, Sang Taek

Lauchli, Ryan

Arnold, Frances H

Protein engineering of cytochrome P450 monooxygenases (P450s) has been very successful in generating valuable non-natural activities and properties, allowing these powerful catalysts to be used for the synthesis of drug metabolites and in biosynthetic pathways for the production of precursors of artemisinin and paclitaxel. Collected experience indicates that the P450s are highly 'evolvable' - they are particularly robust to mutation in their active sites and readily accept new substrates and exhibit new selectivities. Their ability to adapt to new challenges upon mutation may reflect the nonpolar nature of their active sites as well as their high degree of conformational variability.

# Self-organization, layered structure, and aggregation enhance persistence of a synthetic biofilm consortium.

2011

PloS one

Brenner, Katie

Arnold, Frances H

Microbial consortia constitute a majority of the earth's biomass, but little is known about how these cooperating communities persist despite competition among community members. Theory suggests that non-random spatial structures contribute to the persistence of mixed communities; when particular structures form, they may provide associated community members with a growth advantage over unassociated members. If true, this has implications for the rise and persistence of multi-cellular organisms. However, this theory is difficult to study because we rarely observe initial instances of non-random physical structure in natural populations. Using two engineered strains of Escherichia coli that constitute a synthetic symbiotic microbial consortium, we fortuitously observed such spatial self-organization. This consortium forms a biofilm and, after several days, adopts a defined layered structure that is associated with two unexpected, measurable growth advantages. First, the consortium cannot successfully colonize a new, downstream environment until it self-organizes in the initial environment; in other words, the structure enhances the ability of the consortium to survive environmental disruptions. Second, when the layered structure forms in downstream environments the consortium accumulates significantly more biomass than it did in the initial environment; in other words, the structure enhances the global productivity of the consortium. We also observed that the layered structure only assembles in downstream environments that are colonized by aggregates from a previous, structured community. These results demonstrate roles for self-organization and aggregation in persistence of multi-cellular communities, and also illustrate a role for the techniques of synthetic biology in elucidating fundamental biological principles.

# Improved product-per-glucose yields in P450-dependent propane biotransformations using engineered Escherichia coli.

2011

Biotechnology and bioengineering

Fasan, Rudi

Crook, Nathan C

Peters, Matthew W

Meinhold, Peter

Buelter, Thomas

Landwehr, Marco

Cirino, Patrick C

Arnold, Frances H

P450-dependent biotransformations in Escherichia coli are attractive for the selective oxidation of organic molecules using mild and sustainable procedures. The overall efficiency of these processes, however, relies on how effectively the NAD(P)H cofactors derived from oxidation of the carbon source are utilized inside the cell to support the heterologous P450-catalyzed reaction. In this work, we investigate the use of metabolic and protein engineering to enhance the product-per-glucose yield (Y(PPG)) in whole-cell reactions involving a proficient NADPH-dependent P450 propane monooxygenase prepared by directed evolution [P450(PMO)R2; Fasan et al. (2007); Angew Chem Int Ed 46:8414-8418]. Our studies revealed that the metabolism of E. coli (W3110) is able to support only a modest propanol: glucose molar ratio (YPPG ~ 0.5) under aerobic, nongrowing conditions. By altering key processes involved in NAD(P)H metabolism of the host, considerable improvements of this ratio could be achieved. A metabolically engineered E. coli strain featuring partial inactivation of the endogenous respiratory chain (Deltandh) combined with removal of two fermentation pathways (DeltaadhE, Deltaldh) provided the highest Y(PPG) (1.71) among the strains investigated, enabling a 230% more efficient utilization of the energy source (glucose) in the propane biotransformation compared to the native E. coli strain. Using an engineered P450(PMO)R2 variant which can utilize NADPH and NADH with equal efficiency, we also established that dual cofactor specificity of the P450 enzyme can provide an appreciable improvement in Y(PPG). Kinetic analyses suggest, however, that much more favorable parameters (K(M), k(cat)) for the NADH-driven reaction are required to effectively compete with the host's endogenous NADH-utilizing enzymes. Overall, the metabolic/protein engineering strategies described here can be of general value for improving the performance of NAD(P)H-dependent whole-cell biotransformations in E. coli.

# Optimizing non-natural protein function with directed evolution.

2010

Current opinion in chemical biology

Brustad, Eric M

Arnold, Frances H

Developing technologies such as unnatural amino acid mutagenesis, non-natural cofactor engineering, and computational design are generating proteins with novel functions; these proteins, however, often do not reach performance targets and would benefit from further optimization. Evolutionary methods can complement these approaches: recent work combining unnatural amino acid mutagenesis and phage selection has created useful proteins of novel composition. Weak initial activity in a computationally designed enzyme has been improved by iterative rounds of mutagenesis and screening. A marriage of ingenuity and evolution will expand the scope of protein function well beyond Mother Nature's designs.

# Metal-substituted protein MRI contrast agents engineered for enhanced relaxivity and ligand sensitivity.

2010

Journal of the American Chemical Society

Lelyveld, Victor S

Brustad, Eric

Arnold, Frances H

Jasanoff, Alan

Engineered metalloproteins constitute a flexible new class of analyte-sensitive molecular imaging agents detectable by magnetic resonance imaging (MRI), but their contrast effects are generally weaker than synthetic agents. To augment the proton relaxivity of agents derived from the heme domain of cytochrome P450 BM3 (BM3h), we formed manganese(III)-containing proteins that have higher electron spin than their native ferric iron counterparts. Metal substitution was achieved by coexpressing BM3h variants with the bacterial heme transporter ChuA in Escherichia coli and supplementing the growth medium with Mn3+-protoporphyrin IX. Manganic BM3h variants exhibited up to 2.6-fold higher T1 relaxivities relative to native BM3h at 4.7 T. Application of ChuA-mediated porphyrin substitution to a collection of thermostable chimeric P450 domains resulted in a stable, high-relaxivity BM3h derivative displaying a 63% relaxivity change upon binding of arachidonic acid, a natural ligand for the P450 enzyme and an important component of biological signaling pathways. This work demonstrates that protein-based MRI sensors with robust ligand sensitivity may be created with ease by including metal substitution among the toolkit of methods available to the protein engineer.

# Comparison of family 9 cellulases from mesophilic and thermophilic bacteria.

2010

Applied and environmental microbiology

Mingardon, Florence

Bagert, John D

Maisonnier, Cyprien

Trudeau, Devin L

Arnold, Frances H

Cellulases containing a family 9 catalytic domain and a family 3c cellulose binding module (CBM3c) are important components of bacterial cellulolytic systems. We measured the temperature dependence of the activities of three homologs: Clostridium cellulolyticum Cel9G, Thermobifida fusca Cel9A, and C. thermocellum Cel9I. To directly compare their catalytic activities, we constructed six new versions of the enzymes in which the three GH9-CBM3c domains were fused to a dockerin both with and without a T. fusca fibronectin type 3 homology module (Fn3). We studied the activities of these enzymes on crystalline cellulose alone and in complex with a miniscaffoldin containing a cohesin and a CBM3a. The presence of Fn3 had no measurable effect on thermostability or cellulase activity. The GH9-CBM3c domains of Cel9A and Cel9I, however, were more active than the wild type when fused to a dockerin complexed to scaffoldin. The three cellulases in complex have similar activities on crystalline cellulose up to 60 degrees C, but C. thermocellum Cel9I, the most thermostable of the three, remains highly active up to 80 degrees C, where its activity is 1.9 times higher than at 60 degrees C. We also compared the temperature-dependent activities of different versions of Cel9I (wild type or in complex with a miniscaffoldin) and found that the thermostable CBM is necessary for activity on crystalline cellulose at high temperatures. These results illustrate the significant benefits of working with thermostable enzymes at high temperatures, as well as the importance of retaining the stability of all modules involved in cellulose degradation.

# Combinatorial alanine substitution enables rapid optimization of cytochrome P450BM3 for selective hydroxylation of large substrates.

2010

Chembiochem : a European journal of chemical biology

Lewis, Jared C

Mantovani, Simone M

Fu, Yu

Snow, Christopher D

Komor, Russell S

Wong, Chi-Huey

Arnold, Frances H

# Enzymatic functionalization of carbon-hydrogen bonds.

2010

Chemical Society reviews

Lewis, Jared C

Coelho, Pedro S

Arnold, Frances H

The development of new catalytic methods to functionalize carbon-hydrogen (C-H) bonds continues to progress at a rapid pace due to the significant economic and environmental benefits of these transformations over traditional synthetic methods. In nature, enzymes catalyze regio- and stereoselective C-H bond functionalization using transformations ranging from hydroxylation to hydroalkylation under ambient reaction conditions. The efficiency of these enzymes relative to analogous chemical processes has led to their increased use as biocatalysts in preparative and industrial applications. Furthermore, unlike small molecule catalysts, enzymes can be systematically optimized via directed evolution for a particular application and can be expressed in vivo to augment the biosynthetic capability of living organisms. While a variety of technical challenges must still be overcome for practical application of many enzymes for C-H bond functionalization, continued research on natural enzymes and on novel artificial metalloenzymes will lead to improved synthetic processes for efficient synthesis of complex molecules. In this critical review, we discuss the most prevalent mechanistic strategies used by enzymes to functionalize non-acidic C-H bonds, the application and evolution of these enzymes for chemical synthesis, and a number of potential biosynthetic capabilities uniquely enabled by these powerful catalysts (110 references).

# A general mechanism for network-dosage compensation in gene circuits.

2010

Science (New York, N.Y.)

Acar, Murat

Pando, Bernardo F

Arnold, Frances H

Elowitz, Michael B

van Oudenaarden, Alexander

Coping with variations in network dosage is crucial for maintaining optimal function in gene networks. We explored how network structure facilitates network-level dosage compensation. By using the yeast galactose network as a model, we combinatorially deleted one of the two copies of its four regulatory genes and found that network activity was robust to the change in network dosage. A mathematical analysis revealed that a two-component genetic circuit with elements of opposite regulatory activity (activator and inhibitor) constitutes a minimal requirement for network-dosage invariance. Specific interaction topologies and a one-to-one interaction stoichiometry between the activating and inhibiting agents were additional essential elements facilitating dosage invariance. This mechanism of network-dosage invariance could represent a general design for gene network structure in cells.

# Combinatorial recombination of gene fragments to construct a library of chimeras.

2010

Current protocols in protein science

Farrow, Mary F

Arnold, Frances H

Recombination of distantly related and nonrelated genes is difficult using traditional PCR-based techniques, and truncation-based methods result in a large proportion of nonviable sequences due to frame shifts, deletions, and insertions. This unit describes a method for creating libraries of chimeras through combinatorial assembly of gene fragments. It allows the experimenter to recombine genes of any identity and to select the sites where recombination takes place. Combinatorial recombination is achieved by generating gene fragments with specific overhangs, or sticky ends. The overhangs permit the fragments to be ligated in the correct order while allowing independent assortment of blocks with identical overhangs. Genes of any identity can be recombined so long as they share 3 to 5 base pairs of identity at the desired recombination sites. Simple adaptations of the method allow incorporation of specific gene fragments.