Review

Synthetic Metabolism: Engineering Biology at the Protein and Pathway Scales

Collin H. Martin,[1,2] David R. Nielsen,[1] Kevin V. Solomon,[1,2] and Kristala L. Jones Prather[1,2,*]
[1]Department of Chemical Engineering, Synthetic Biology Engineering Research Center (SynBERC), Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Synthetic Biology Engineering Research Center (SynBERC), Massachusetts Institute of Technology, Cambridge, MA 02139, USA
*Correspondence: kljp@mit.edu

Biocatalysis has become a powerful tool for the synthesis of high-value compounds, particularly so in the case of highly functionalized and/or stereoactive products. Nature has supplied thousands of enzymes and assembled them into numerous metabolic pathways. Although these native pathways can be use to produce natural bioproducts, there are many valuable and useful compounds that have no known natural biochemical route. Consequently, there is a need for both unnatural metabolic pathways and novel enzymatic activities upon which these pathways can be built. Here, we review the theoretical and experimental strategies for engineering synthetic metabolic pathways at the protein and pathway scales, and highlight the challenges that this subfield of synthetic biology currently faces.

Introduction

Synthetic biology has emerged as a powerful discipline for the creation of novel biological systems (Endy, 2005; Pleiss, 2006), particularly within the subfield of metabolic pathway and product engineering (Keasling, 2008; Savage et al., 2008). Continuing efforts to characterize and understand natural enzymes and pathways have opened the door for the building of synthetic pathways toward exciting and beneficial compounds such as the antimalarial drug precursor artemisinic acid (Ro et al., 2006) and several branched-chain alcohols for use as biofuels (Atsumi et al., 2008b). The need for synthetic metabolic routes is a consequence of the fact that the array of compounds of interest for biosynthesis vastly outnumbers the availability of characterized pathways and enzymes. Several key building blocks can be made biologically (Patel et al., 2006); however, a recent report from the U.S. Department of Energy highlighted 12 biomass-derived chemical targets, only half of which have known biochemical routes (Werpy and Petersen, 2004).

With the lack of characterized natural pathways to synthesize many high-value compounds, we must learn to forge our own metabolic routes toward these molecular targets. Logically, it follows that for unnatural pathways, we will need new, unnatural enzymes from which these pathways can be composed. The parts-devices framework of synthetic biology lends itself well to this dual-sided problem of synthetic pathway creation (Endy, 2005); that is, pathways can be thought of as metabolic devices composed of individual enzyme-catalyzed reaction parts. Implicit within this framework is the idea that the challenges of pathway creation are best approached at both the part and device levels. In this review, we first discuss efforts at the protein-level for broadening the array of enzyme parts that can be recruited for use in synthetic pathways. The discussion is then expanded to pathway-level synthetic biology, where we review the tools available for designing metabolic pathways from enzyme-level parts and the implementation strategies for realizing these pathways experimentally. The overall process of pathway creation (Figure 1) combines experimental and theoretical components of synthetic biology at both scales.

Synthetic Biology at the Protein Scale

Through natural evolution, organisms have acquired the capacity to catalyze a multitude of diverse chemical reactions as a means to proliferate in a wide range of unique microenvironments. Although only a small fraction of the earth's biodiversity (and an even smaller subset of its composite enzymes) has been characterized, the identification and isolation of novel proteins with unique properties or enzymatic function is a laborious procedure. One particularly promising source of new enzymes and enzymatic activities is the emerging field of metagenomics (Handelsman, 2004). Nonetheless, the physical and catalytic properties of natural enzymes often render them as incompatible or, at the very least, unoptimized for use in engineered pathways and strains. In cases where natural evolution has fallen short of industrial needs, the tools and practices of synthetic biology can be applied to aid in the creation of designer enzymes and cellular phenotypes. The challenge of building new enzymes and reengineering natural ones has been approached with the development of predictive theoretical frameworks and a range of experimental techniques (Figure 2).

Theoretical Approaches

Computational tools exist to adapt the natural array of proteins for use in an increasing number of applications. For example, the effects of codon bias on expression levels (Kane, 1995; Gustafsson et al., 2004) can be resolved by design tools such as Gene Designer (Villalobos et al., 2006). Other effects such as Shine-Dalgarno sequences, promoter strength, and mRNA stability can be similarly optimized. Nonetheless, the application of these tools is still limited to the biochemical diversity found in nature. To increase the number and efficiency of biologically catalyzed reactions, more sophisticated in silico techniques are needed. Although full-scale protein folding and ab initio protein design and modeling are neither trivial nor currently practical, the use of solved protein structures, strong physical models, and experimentally derived libraries allow for the design and improvement of enzymes. These theoretically designed proteins in turn have significant potential to impact pathway-level synthetic metabolism (Yoshikuni et al., 2008).
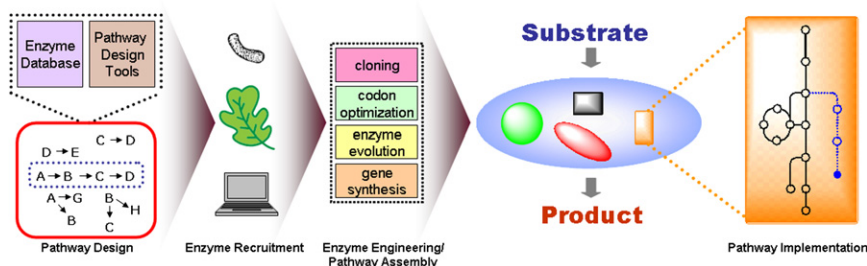
Figure 1. Overall Scheme for Pathway Creation
The creation process includes protein-level recruitment and reengineering of enzymes and pathway-level efforts to design and assemble these enzymes into an unnatural pathway.

An empirical approach to synthetic protein design includes an understanding of the protein sequence/function relationship. One example is the use of a linguistic metaphor to describe a protein sequence (Searls, 1997, 2002). In language, a sentence is composed of a sequence of words whose parsed meaning is a function of not only their individual definitions but also their connotations, which are encoded by their type (part of speech) and their relative location to other words. Similarly, a protein "sentence" is composed of residues that not only have a definitive identity but also possess chemical properties and a relative position that affect the subsequent fold and function of the resultant protein (Przytycka et al., 2002). Building on the successes of a putative protein grammar (Przytycka et al., 2002; Abe and Hiroshi, 1997), Loose et al. (2006) recently demonstrated its use in the design of new antimicrobial peptides. Using the TEIRESIAS algorithm (Rigoutsos and Floratos, 1998), a library with homology to known sequences restricted to below 60% was generated with approximately 50% of designs showing some antimicrobial activity. An alternative approach to modeling protein sequence/function relationships involves the use of folded protein scaffolds and quantum transition state models. Through detailed crystal structures and transition state models, Hederos et al. (2004) noted that the active site of a glutathione transferase was of the appropriate size and structure to stabilize the transition state complex of the hydrolytic degradation of a thioester. By introducing a histidine residue within the active site they were able to impart significant thioesterase activity. Finally, physics-based free energy approaches have been developed to predict protein structure/function relationships in the context of antibody binding strength. Although total free energy models were not a good predictor, Lippow et al. (2007) found that the electrostatic interaction contributions to total energy were well correlated with antibody binding affinity. Using this relationship, they were able to generate an improved lysozyme antibody design that demonstrated a 140-fold increase in binding. Although neither of these examples fully describe protein structure/function relationships, each does offer a unique insight into the problem. Namely, they drastically reduced the sequence space of potential modifications to a manageable subset with a high probability of success. In this manner, such empirical models serve as an important tool in the design and improvement of enzymes.

Using a quantum transition state framework, great strides have been made in the in silico development of enzyme activities (Jiang et al., 2008; Rothlisberger et al., 2008; Kaplan and DeGrado, 2004). At the heart of these efforts is a strong understanding of the desired catalytic mechanism and its associated transition states and reaction intermediates. Once compiled, this information can be used to generate an active site of the appropriate dimensions with critical residues incorporated into appropriate locations for catalysis. At this point, the designer has two options: try to identify a suitable folded scaffold that can accommodate the active site with minimal mutations or generate a protein backbone with correctly folded active site de novo. Each method has its inherent advantages and challenges. Although finding a host scaffold would appear to be the simpler of the two, it requires extensive searches of protein structure libraries with tools such as RosettaMatch (Zanghellini et al., 2006). Nonetheless, this approach has had some success with the catalysis of unnatural reactions such as the retro-aldol catalysis of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (Jiang et al., 2008) and the Kemp elimination (Rothlisberger et al., 2008). Coupled with experimental techniques, in-silico-designed enzymes can have activity levels comparable with
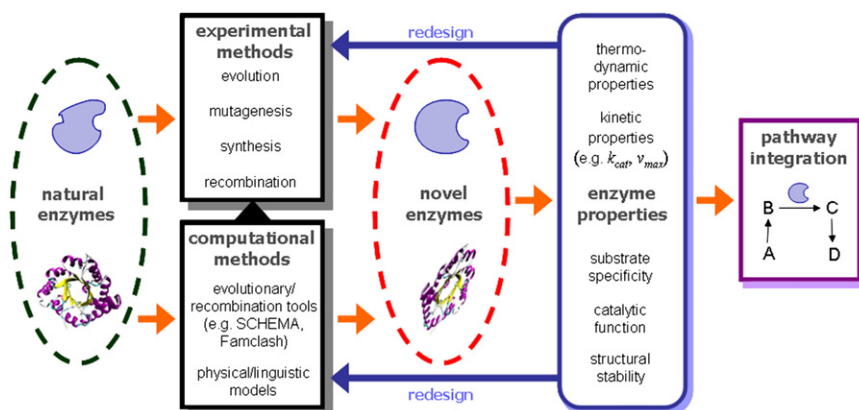
Figure 2. Flowchart for the Creation of New Enzymes with Experimental Techniques and Computational Tools
New enzymes generated with these methods are examined for desired properties and either further reengineered or adapted for use in unnatural pathways.

that of evolved natural enzymes (Rothlisberger et al., 2008). In contrast, de novo protein scaffold development requires significant computational effort to consider not only the stability of the desired conformation of the backbone and active site but also the likelihood of destabilization. Nonetheless, Kaplan and DeGrado (2004) have successfully used such an approach to generate an O_2-dependent phenol oxidase. Despite the computational overhead associated with these methods, their feasibility points to an improving and functional understanding of protein structure/function relationships, leading to increased possibilities for the rational design of enzymes and proteins.

In the absence of rational insight, theoretical tools can assist experimental techniques in generating new and improved proteins. One common technique is protein recombination or in vitro shuffling, which combines the best traits of two or more individual enzymes (Stemmer, 1994a, 1994b). However, successful recombination is contingent on shuffling at domain boundaries to ensure proper folding of each domain. The predictive algorithm SCHEMA, developed by Voigt et al. (2002), was designed to aid in the screening process of such chimeric proteins. By analyzing the nature and number of the disruptions of the intermolecular interactions, Voigt et al. were able to generate a metric correlated with the probability of active β-lactamase hybrids of TEM-1 and PSE-4 (2002). Subsequent studies by Meyer et al. (2003) have confirmed this correlation and used SCHEMA-guided recombination to derive functional and diverse libraries of cytochrome P-450 s (Otey et al., 2004) and β-lactamases (Meyer et al., 2006). Another available predictive algorithm is FamClash (Saraf et al., 2004), which analyzes chimeras for the conservation of charge, volume, and hydrophobicity at a given residue. Generated sequence scores have been demonstrated to be well correlated with the activities of hybrid dihydrofolate reductases. Although experimental techniques are important generators of diverse protein libraries, tools such as FamClash, SCHEMA, and other related sequence analysis programs enrich such chimeric libraries and vastly improve their value in the development of new and improved proteins. Currently, these tools are incapable of predicting hits a priori; however, their importance in successful protein design should not be underestimated.

Experimental Approaches

Rather than focusing on the prediction of protein structure and function, experimental techniques allow the improvement or modification of existing enzymes, in some instances creating entirely new enzymes and enzyme activities. These techniques include mutagenesis, enzyme engineering and evolution, and gene synthesis technology, with each boasting its own distinct advantages and inherent limitations (Bonomo et al., 2006; Alper and Stephanopoulos, 2007). Collectively, they comprise a powerful set of tools for the efficient generation of enzymes with user-specified properties. Protein recombination, for example, provides a means by which secondary structural elements, from natural or evolved proteins, can be rationally assembled in a modular fashion to integrate domains featuring desired attributes (Otey et al., 2004).

The construction of synthetic pathways typically involves the recruitment of genes from an array of sources to provide the required enzymatic function and activity (Figure 1). However, heterologously expressed proteins, particularly those originating

from a source organism belonging to a different kingdom than that of the expression host, often suffer from poor activity as a result of dissimilarities in codon usage. In such cases, the use of synthetic genes with codon optimized sequences has been frequently employed to achieve sufficient levels of functional expression. Synthesis of a codon optimized xylanase gene from *Thermomyces lanuginosus* DSM 5826 led to a 10-fold improvement in expression level in *E. coli* (Yin et al., 2008). Plant genes are often found to be poorly expressed in *E. coli* (Martin et al., 2001). Martin et al. (2003) synthesized a codon optimized variant of amorpha-4,11-diene synthase from the *Artemisia annua* to catalyze the conversion of farnesyl pyrophosphate to amorphadiene, a precursor used for the production of the antimalarial drug artemisinin. As the cost associated with gene synthesis continues to decrease, imaginable applications of synthetic genes and artificial, designer proteins to include increased elements of rational design become increasingly plausible.

The versatility of directed evolution for engineering desired enzyme attributes is highlighted by a multitude of recent works employing this approach for a diverse assortment of applications, including the enhancement of thermal stability (Asako et al., 2008; Shi et al., 2008) and acid tolerance (Liu et al., 2008); promoting higher chemo-, regio-, and enantio-selectivity toward substrates (Asako et al., 2008); elimination of undesired biochemical activities (e.g., side reactions; Kelly et al., 2008); and improving heterologous expression (Mueller-Cajar and Whitney, 2008). In the example of the stereospecific reduction of 2,5-hexanedione to (2S,5S)-hexanediol by alcohol dehydrogenase from the thermophilic bacteria *Pyrcoccus furiosus*, laboratory evolution was used by Machielsen et al. (2008) to alter the enzyme's optimum temperature and improve its activity in recombinant *E. coli* under moderate culture conditions. Meanwhile, Aharoni et al. (2004) have achieved functional expression of mammalian paraoxonases PON1 and PON3 in *E. coli* through a directed evolution scheme that incorporated family DNA shuffling (shuffling of DNA encoding homologous genes from different genetic sources) and random mutagenesis to achieve the first active microbial expression of recombinant PON variants. As a tool, directed evolution continues to benefit from refinements aimed at improving the efficiency at which desired mutations can be obtained from a minimal number of iterations while also reducing screening efforts (Reetz and Carballeira, 2007; Reetz et al., 2008).

In addition to improving expression and altering the thermal properties of heterologous enzymes, novel biochemical activities can be similarly engineered by the aforementioned strategies. For example, cytochrome P450 BM3 from *Bacillus megaterium* has been engineered via directed evolution using several sequential rounds of mutagenesis to alter its regioselectivity for the hydroxylation of n-alkanes from subterminal positions to that of the terminus (Meinhold et al., 2006). The approach has been employed to convert several different n-alkanes to their corresponding n-alcohols, including the hydroxylation of ethane to ethanol as a means for producing more tractable transportation fuels from petrochemical feedstocks (Meinhold et al., 2005). To promote high end-product specificity while maximizing metabolite flux, the preferential activity of an enzyme between multiple competing substrates can also be tailored. For instance, the substrate specificity of pyruvate oxidase (PoxB) from *E. coli* was altered via localized random mutagenesis to decrease its

A

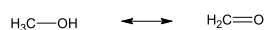**Enzyme** | **Reaction Catalyzed**
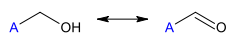
Homoserine Dehydrogenase
(E.C. 1.1.1.3)
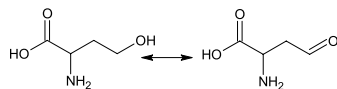
1,3-Propanediol Dehydrogenase
(E.C. 1.1.1.202)

Methanol Dehydrogenase
(E.C. 1.1.1.244)

Core Generalized
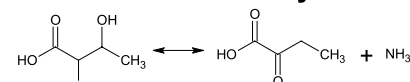Reaction

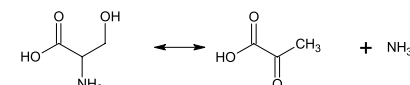Conserved Structure
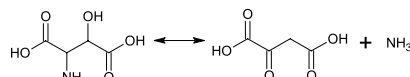Generalized Reaction

B

Enzyme | **Reaction Catalyzed**

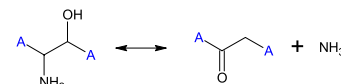L-Serine Ammonia-Lyase
(E.C. 4.3.1.17)
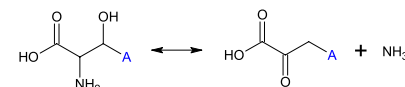
Threonine Ammonia-Lyase
(E.C. 4.3.1.19)

Erythro-3-Hydroxyaspartate
Ammonia-Lyase
(E.C. 4.3.1.20)

Core Generalized
Reaction

Conserved Structure
Generalized Reaction

C

Enzyme | **Reaction Catalyzed**

Cysteine Synthase
(E.C. 2.5.1.47)
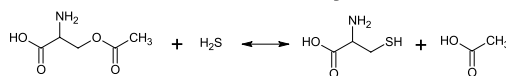
Cystathionine gamma-Synthase
(E.C. 2.5.1.48)
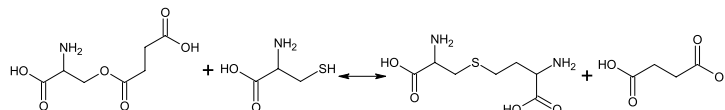
O-Acetylhomoserine
Aminocarboxypropyltransferase
(E.C. 2.5.1.49)
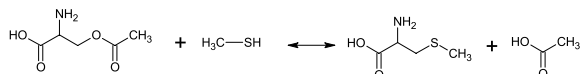
beta-Pyrazolylalanine Synthase
(E.C. 2.5.1.51)
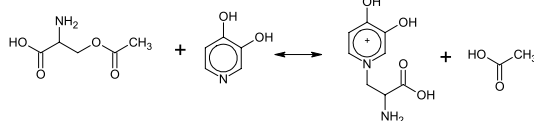
L-Mimosine Synthase
(E.C. 2.5.1.52)
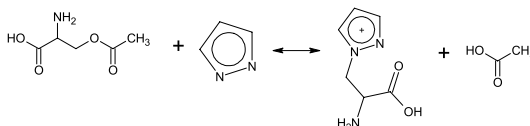
First Core Generalized
Reaction (E.C. 2.5.1.47-
2.5.1.49)
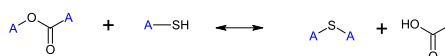
First Conserved
Structure Generalized
Reaction (E.C. 2.5.1.47-
2.5.1.49)
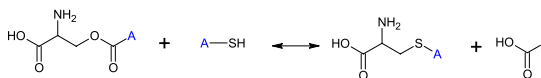
Second Core
Generalized Reaction
(E.C. 2.5.1.51 and
2.5.1.52)
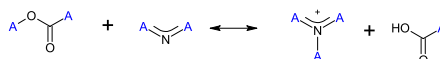
Second Conserved
Structure Generalized
Reaction (E.C. 2.5.1.51
and 2.5.1.52)

activity on pyruvate in favor of an alternative endogenous metabolite, 2-oxo-butanoate (Chang and Cronan, 2000). Synthetic pathways incorporating this PoxB mutant will accordingly display preferential synthesis of products from the four-carbon precursor. Meanwhile, Tsuge et al. (2003) utilized site-directed mutagenesis to shift the substrate specificity of PhaJ, an *R*-specific enoyl-coenzyme A (CoA) hydratase from *Aeromonas caviae* from short-chain 3-hydroxyacyl-CoA precursors toward those with longer carbon chain lengths (8 to 12). When incorporated into an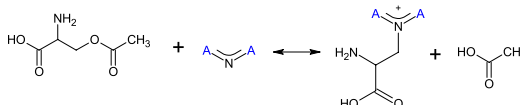 engineered polyhydroxyalkanoate (PHA) synthesis pathway in *E. coli*, increased molar fractions of C_8 and C_{10} 3-hydroxyacid monomer units were found to be incorporated into PHA. In this case, the capacity to distinctly manipulate the composition of PHAs makes possible the synthesis of novel bioplastics with customizable physical properties to meet commercial requirements. The ability to finely tune the substrate specificity of an engineered enzyme is of particular importance for promoting high selectivity and product yield, as well as for reducing the ill-effects of molecular cross-talk between engineered and endogenous pathways.

At the protein level, synthetic biology aims to expand the catalog of well-characterized enzymes while also engineering novel biochemistries. Subsequent incorporation of engineered enzymes into synthetic pathways leads to the construction of devices that can be implemented to achieve a user-specified function, such as the production of biofuels or high-value pharmaceutical compounds. The design and construction of new metabolic routes from individual enzymes represents synthetic biology at the next scale, the pathway scale, and has unique challenges of its own.

Synthetic Biology at the Pathway Scale

Pathway-scale synthetic biology aims to create novel metabolic routes toward both existing metabolites and unnatural compounds. Traditionally, pathway engineering has been synonymous with metabolic engineering and its toolbox has been composed of the same tools: gene knockouts, flux optimization, gene overexpression, and the like. The ability to manipulate natural metabolism has seen many useful applications, such as improving ethanol production in *Saccharomyces cerevistiae* (Bro et al., 2006), solventogenesis in *Clostridium acetobutylicum* (Mermelstein et al., 1993; Woods, 1995), and penicillin production in *Penicillium chrysogenum* (Casqueiro et al., 2001). A key limitation in all of these examples is the confinement of pathway engineering to the manipulation of natural metabolism. Continuing advances in characterizing, modifying, and even creating enzymes (several of them discussed in the previous section of this review) now allow us to build unnatural pathways for the biological production of compounds. Understanding synthetic biology at the protein scale affords us the opportunity to apply it at the pathway scale.

As at the protein scale, pathway-level synthetic biology has been approached from both theoretical and experimental fronts. The theoretical work centers on the concept of pathway design—assembling a logical series of enzyme-catalyzed reactions to convert an accessible substrate into a valued final compound. Theoretical pathway design probes what conversions are possible and what enzyme parts need to be assembled to create a functional metabolic device. In contrast, experimental efforts focus on the construction and application of unnatural pathways and serve as powerful real-world examples of what these pathways can accomplish. Experimental approaches enable the exploration of enzyme behaviors such as substrate promiscuity and activity, both useful properties for creating unnatural pathways that cannot readily be predicted with theoretical approaches.

Theoretical Approaches

Before an unnatural metabolic pathway can be built in the laboratory, it must first be designed. The goal of pathway design is to use a series of biochemically catalyzed reactions to connect a target product molecule to either a cellular metabolite (such as acetyl-CoA, α-ketoglutarate, or L-alanine) or to a feasible feedstock (such as glucose or glycerol). This can be accomplished using either natural enzymes or engineered ones. The sheer number of known enzymes (both natural and engineered) and enzyme-catalyzed reactions available means that there will almost certainly exist many possible theoretical pathways toward a given target compound (Li et al., 2004; Hatzimanikatis et al., 2005). Identifying and ranking these different possibilities are the central challenges in pathway design.

One of the first steps in pathway design is obtaining knowledge of the enzymes and enzyme-catalyzed reactions available for use in a pathway. Comprehensive protein and metabolism databases, such as BRENDA (Schomburg et al., 2004), KEGG (Kanehisa et al., 2006), Metacyc (Caspi et al., 2006), and Swiss-Prot (Wu et al., 2006), provide a wealth of information on the pool of natural, characterized enzymes that can be recruited. More importantly, these databases reveal chemical conversions that are achievable with enzymes. As of the preparation of this manuscript, there are approximately 398,000 protein entries in Swiss-Prot (build 56.2), from which the enzymes are organized into 4757 four-digit enzyme classification (E.C.) groups in the most recent version of BRENDA (build 2007.2). Because of the large number of characterized enzymes, those performing similar reaction chemistries are typically organized into generalized enzyme-catalyzed reactions for the purposes of pathway construction (Li et al., 2004). A generalized enzyme-catalyzed reaction is defined as the conversion of one functional group or structural pattern in a substrate into a different group or structure in its product (Figure 3). Structural information about the nonreacting portions of the substrate is ignored, making the identification of enzymes to carry out a desired chemical conversion a much

Figure 3. Generalized Enzyme-Catalyzed Reactions for a Subset of E.C. 1.1.1 Alcohol Dehydrogenases, E.C. 4.3.1 Ammonia-Lyases, and E.C. 2.5.1 Synthases
The "A" atoms present in the molecular structures are wildcards.
(A) Two different methods of assigning generalized reactions, one considering only the reacting parts of the molecule (core generalized reaction) and one identifying patterns of conserved molecular structure in addition to the reacting structural elements (conserved structure generalized reaction), arrive at the same generalized reaction.
(B) The two methods arrive at different generalized reactions, illustrating the need for a generalization standard.
(C) A set of five enzymes within a three-digit E.C. class result in two different sets of generalized reactions, illustrating that the E.C. system does not necessarily correlate with reaction generalization.

more tractable problem. However, the logical rules for assigning enzymes to a generalized reaction can be subjective (Figure 3). One could for instance differentiate between reactions solely on the reacting functional groups (i.e., aldehyde to alcohol) as Li et al. (2004) did, or one could also include information about conserved patterns of molecular structure between similar enzyme-catalyzed reactions. Furthermore, generalized enzymatic reactions do not all fall cleanly into the existing E.C. system (Figure 3C).

Despite the need for a universal standard in reaction generalization, several publically available tools utilize this approach to address the problem of pathway design. The BNICE (*Biochemical Network Integrated Computational Explorer*) framework allows for the discovery of numerous possible metabolic routes between two compounds (Li et al., 2004; Hatzimanikatis et al., 2005). This framework was applied to aromatic amino acid biosynthesis to find over 400,000 theoretical biochemical pathways between chorismate and phenylalanine, tyrosine, or tryptophan (Hatzimanikatis et al., 2005) and it was used to explore thousands of novel linear polyketide structures (González-Lergier et al., 2005). Our group has developed a database of over 600 conserved structure generalized enzyme-catalyzed reactions called ReBiT (Retro-Biosynthesis Tool, http://www.retro-biosynthesis.com), which accepts as input a molecular or functional group structure and returns as output all three-digit E.C. groups capable of reacting with or producing that structure. The University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD) uses a series of generalized reaction rules to propose pathways step by step, with particular emphasis on analyzing the degradation trajectories of xenobiotics (Ellis et al., 2006; Fenner et al., 2008).

Typically multiple, and indeed in some cases, several thousand, metabolic routes can be proposed for a given compound. How does one distinguish logical, feasible pathways from frivolous, improbable ones? What metrics can be applied to judge one computationally generated pathway as superior (i.e., more likely to be functionally constructed) to another? One way of narrowing the choice of pathways is to apply natural precedent to filter out unlikely pathway steps. In this strategy, a large set of experimentally validated enzyme-catalyzed reactions are examined for patterns of structural change and a series of rules are developed to give preference to reaction steps containing structural changes that follow these rules. This methodology is implemented in the UM-BBD to avoid the "combinatorial explosion" that results when considering all the possible pathways that any given compound can take (Fenner et al., 2008). Another ranking strategy is to calculate the thermodynamic favorability of the steps and to penalize pathways involving steps that are energetically unfavorable. This approach is taken by the BNICE framework (Hatzimanikatis et al., 2005) using a functional group contribution method (Jankowski et al., 2008) to compute the overall change in Gibbs energy for each individual pathway step. A new pathway modeling tool, DESHARKY, quantifies and employs metabolic burden as a metric for judging unnatural pathways and, in particular, how they are connected to cellular metabolism (Rodrigo et al., 2008). DESHARKY is a Monte-Carlo-based algorithm that estimates the transcriptomic and metabolic loads on cells expressing unnatural pathways and calculates the decrease in specific growth rate as a result of

these additional burdens. There are still other possibilities for pathway ranking, such as the number of pathway steps taken, the known substrate specificities (or lack thereof) of the enzymes involved in each pathway, or the availability and diversity of homologous enzymes to test at each pathway step. One of the key challenges in pathway design is scoring pathways in a robust and balanced manner, and only as more nonnatural pathways are designed and built will there be a better understanding as to which of these metrics are relevant and useful.

Experimental Approaches

With a target compound and a proposed metabolic route to reach that compound in hand, one is now ready to begin experimental implementation of that pathway. Synthetic pathway construction occurs over several shades of novelty—from recreating natural pathways in heterologous hosts to creating synthetic pathways that parallel natural ones to building completely novel metabolic routes toward unnatural compounds from multiple, ordinarily unrelated enzymes (Figure 4). Here we discuss the situations in which nonnatural pathways prove useful and several general strategies for creating these pathways.

Through the course of evolution, nature has assembled many pathways toward several useful compounds, such as the biofuel and solvent 1-butanol in *Clostridium acetobutylicum* (Jones and Woods, 1986; Dürre et al., 2002; Lee et al., 2008), the C_5 terpenoid building block isopentenyl pyrophosphate in *Saccharomyces cerevisiae* (Seker et al., 2005), and the biopolymer polyhydroxybutyrate (PHB) in *Ralstonia eutropha* (Wang and Yu, 2007). These pathways have physiological roles within their native hosts; for example, the butanol pathway from acetyl-CoA in *C. acetobutylicum* serves as an electron sink to regenerate NAD^+ for glycolysis while deacidifying its environment (Jones and Woods, 1986). Pathways in nature are optimized through evolution to accomplish their physiological objectives, yet in most cases of pathway engineering, it is desired to maximize the production of a target molecule in a pathway rather than to accomplish a physiological goal. Butanol production in *C. acetobutylicum*, for instance, is constricted by cellular regulation tying it to pH, redox conditions, and sporulation (Dürre et al., 2002; Lee et al., 2008). The transference of natural pathways into heterologous hosts isolates these pathways from their regulatory elements and represents a first small step toward the creation of nonnatural metabolism. Although heterologous pathway expression is limited to pathways found in nature, it nonetheless has proven effective in enhancing product titers and/or deregulating compound production for a wide array of products, including the compounds in the examples above (Atsumi et al., 2008b; Kang et al., 2008; Martin et al., 2003; Pitera et al., 2007).

The next level of novelty in synthetic pathway construction is creating metabolic routes that parallel natural pathways, typically by capitalizing on enzymatic promiscuity or enzyme engineering to operate natural or near-natural pathways on nonnatural substrates. This pathway construction strategy allows for the biosynthesis of truly unnatural compounds. Returning to the PHB example, recombinant *R. eutropha* have been shown to incorporate sulfur-containing short- and medium-chain length thioacids into polythioester copolymers (Ewering et al., 2002). The synthesis of these completely unnatural polymers was made possible by taking advantage of the relatively broad substrate specificity of PHA synthases (Hazer and Steinbüchel,
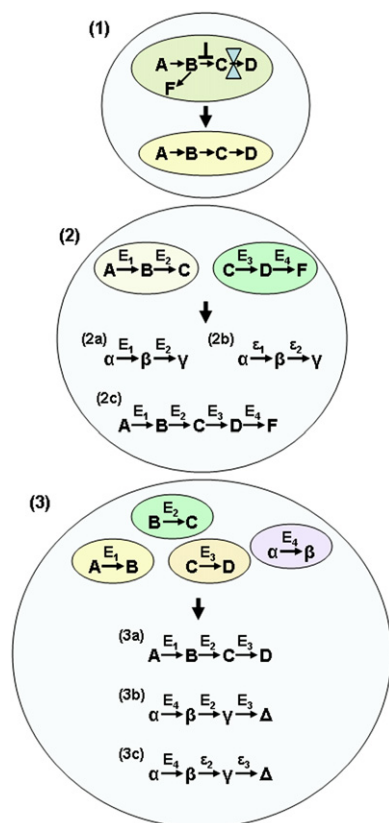
Figure 4. Strategies for Synthetic Pathway Creation Arranged in Increasing Degrees of Departure from Nature

A, B, C, D, F, α, β, γ, and Δ represent metabolites, E represents an enzyme catalyzing a reaction, and ε represents an engineered enzyme catalyzing a reaction. In (1), a natural pathway in its native host is transferred to a heterologous host, decoupling it from native regulation. This strategy is limited to the production of natural products using natural pathways. In (2), new pathways are made in parallel to natural ones through the use of promiscuous enzymes (2a), enzyme engineering (2b), or combinations of natural pathways (2c). Strategies 2a and 2b allow for the synthesis of new, nonnatural products, whereas 2c allows for the creation of new metabolic routes between natural metabolites. Strategy 3 represents de novo pathway construction, where individual unrelated enzymes are recruited to form entirely unnatural pathways. This can be done using native enzyme activities (3a), promiscuous enzymes (3b), engineered enzymes (3c), or combinations thereof.

2007), and because of that broad substrate specificity, hundreds of different monomer units of various sizes (C_3–C_{16}) and substituents have been incorporated into PHA copolymers (Steinbüchel and Valentin, 1995). Another example of parallel pathway construction is the synthesis of triacetic acid lactone from acetyl-CoA by expressing an engineered fatty acid synthase B from *Brevibacterium ammoniagenes* (Zha et al., 2004). This multifunctional enzyme has many domains designed to catalyze the various reductions and condensations necessary for fatty acid synthesis (Meurer et al., 1992). By specifically inactivating the ketoacyl-reductase domain of this fatty acid synthase, the enzyme could no longer use NADPH to reduce its acetyl-CoA condensation products, causing them to circularize into triacetic acid lactone rather than forming linear fatty acids. Finally, natural products can be synthesized by arranging whole or partial pathways to form a mixed, synthetic metabolic route. For example,

the theoretical yield of L-glutamate was improved from 1 mol glutamate per mol glucose to 1.2 mol per mol by augmenting the native *Corynebacterium glutamicum* pentose phosphate pathway with a phosphoketolase from *Bifidobacterium lactis* (Chinen et al., 2007). This strategy allowed for the production of acetyl-CoA without the loss of carbon caused by pyruvate decarboxylation to acetyl-CoA and resulted in increased glutamate titers and productivity.

One of the most promising (and challenging) strategies for building synthetic pathways is de novo pathway construction: the creation of pathways using disparate enzymes to form entirely unnatural metabolic routes toward valuable compounds. This method of pathway building does not rely upon natural precedent, but rather allows one to build entirely new metabolite conduits from individual enzymatic pieces. As a result, this approach allows for the biosynthesis of the widest array of compounds. However, this strategy is the most difficult to realize given that for a completely unnatural pathway, there might not be a complete set of appropriate known enzymes in nature to build it. De novo pathway construction illustrates the need for a more complete set of enzymatic tools for use in building synthetic pathways, and frequently this strategy is coupled with enzyme engineering or the exploitation of enzymatic promiscuity to compensate for the absence of a natural enzyme to execute a desired conversion step.

Because of the challenge in creating functional de novo pathways, few examples exist. However, those that are available describe the biosynthesis of a wide range of useful compounds and illustrate the utility of the approach. For instance, a pathway for the biosynthesis of 1,2,4-butanetriol from D-xylose and L-arabinose was assembled using pentose dehydrogenases and dehydratases from *Pseudomonas fragi* and *E. coli* and benzoylformate decarboxylase from *Pseudomonas putida* (Niu et al., 2003). In this case, multiple decarboxylases were screened to find a promiscuous decarboxylase from *P. putida* capable of acting on a 3-deoxy-glyceropentulosonic acid intermediate in the pathway. Another example of exploiting substrate promiscuity in de novo pathway design is in the synthesis of several higher biofuels such as 2-methyl-1-butanol, isobutanol, and 2-phenylethanol from glucose in *E. coli* (Atsumi et al., 2008b). Here, several 2-keto-acid decarboxylases were screened to identify one from *Lactococcus lactis* for use in creating alcohols from 2-ketoacids (when combined with native *E. coli* alcohol dehydrogenase activity). In a third example, a synthetic pathway for the unnatural aminoacid phenylglycine from phenylpyruvate was made by combining hydroxymandelate synthase, hydroxymandelate oxidase, and D-(4-hydroxy)phneylglycine aminotransferase activities from *Amycolatopsis orientalis*, *Streptomyces coelicolor*, and *P. putida* (Müller et al., 2006). Finally, engineered enzymes can be employed to create de novo pathways, as in the recent case of the synthesis of 3-hydroxypropionic acid from alanine in *E. coli* (Liao et al., 2007). Here, a lysine 2,3-aminomutase from *Porphyromonas gingivalis* (Brazeau et al., 2006) was evolved to have alanine 2,3-aminomutase activity, allowing for the biosynthesis of β-alanine. Combining this evolved enzyme with β-alanine aminotransferase and endogenous alcohol dehydrogenase activities afforded the final 3-hydroxypropionic acid product. Another very recent work utilizes engineered pyruvate decarboxylase and 2-isopropylmalate synthase

for the synthesis of non-natural alcohols from 2-ketoacids in *E. coli* (Zhang et al., 2008). By engineering the enzymes responsible for elongating 2-ketoacids and carrying out their decarboxylation and reduction, the production of a broader array of longer-chain alcohols was enabled.

Conclusions

The design and assembly of unnatural metabolic pathways represents a young and exciting field with the potential to supplement, expand upon, or even replace current industrial processes for the production of fine and commodity chemicals. Synthetic pathway engineering integrates many components and consequently is highly interdisciplinary (Figure 1). Key issues that need to be overcome in pathway design are (1) establishing a standard for generalized enzyme-catalyzed reactions, (2) capturing enzyme substrate preferences in these generalized reactions, and (3) determining the pathway metrics that correlate with successful pathway construction. Overcoming the first two challenges will allow for the creation of the next generation of pathway design tools that better account for enzyme behavior, whereas conquering the last challenge will afford us the ability to rank and choose metabolic pathways and refine the results from design tools. For experimentally implementing unnatural pathways, the central challenge is the limited number of characterized enzymes for the construction of new pathways. In particular, there is great demand for both promiscuous natural enzymes and engineered enzymes to perform specific desired reactions.

The need for new enzymes has given rise to several theoretical frameworks for relating protein sequence, structure, and function. These frameworks each address a piece of the problem—energetics, active site catalysis, and protein backbone structure, etc.—but the ability to routinely build whole enzymes is still in the distant future. In the meantime, mimicking active sites, backbones, and protein linguistics from nature has proven fruitful in creating novel proteins. Experimental evolution and chimeragenesis of enzymes are standard ways of imparting unnatural properties, particularly in the absence of detailed information about the protein. The power of these experimental techniques is primarily limited by the size of the resulting enzyme libraries and the throughput of the screen to analyze them. Computational tools such as SCHEMA (Voigt et al., 2002) and Famclash (Saraf et al., 2004) can assist in focusing and enriching these libraries.

Because biotechnology is increasingly relied upon as a means for chemical production, progress on the creation of new enzymes and unnatural pathway design and construction will flourish. These new pathways must still be expressed within a cellular context, thus improving and understanding unnatural pathway efficacy at a systems level will be important for shattering barriers in pathway expression and product titer. For example, application of flux balance analysis (Edwards et al., 2002) can guide systems-level integration of nonnatural pathways with host metabolism. Furthermore, redox balancing and cofactor regeneration with respect to new pathways are critical to minimize their burden on the host cell (Endo and Koizumi, 2001). Systems-level functionality can also be coupled with unnatural pathways, for instance in the delivery of recombinant microbes to a cancerous tumor (Anderson et al., 2006). Such

microbes could be engineered to simultaneously produce and deliver a drug. Established and recent advances in metabolic engineering, such as global transcription machinery engineering (Alper and Stephanopoulos, 2007), can complement synthetic biology in this regard, leading to improved performance of novel pathways.

REFERENCES

Abe, N., and Hiroshi, M. (1997). Predicting protein secondary structure using stochastic tree grammars. Mach. Learn. *29*, 275–301.

Aharoni, A., Gaidukov, L., Tagur, S., Toker, L., Silman, I., and Tawfik, D.S. (2004). Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization. Proc. Natl. Acad. Sci. USA *101*, 482–487.

Alper, H., and Stephanopoulos, G. (2007). Global transcription machinery engineering: A new approach for improving cellular phenotype. Metab. Eng. *9*, 258–267.

Anderson, J.C., Clarke, E.J., Arkin, A.P., and Voigt, C.A. (2006). Environmentally controlled invasion of cancer cells by engineered bacteria. J. Mol. Biol. *355*, 619–627.

Asako, H., Shimizu, M., and Itoh, N. (2008). Engineering of NADPH-dependent aldo-keto reductase from *Penicillium citrinum* by directed evolution to improve thermostability and enantioselectivity. Appl. Microbiol. Biotechnol. *80*, 805–812.

Atsumi, S., Hanai, T., and Liao, J.C. (2008b). Non-fermentive pathways for synthesis of branched-chain higher alcohols as biofuels. Nature *451*, 86–90.

Bonomo, J., Warnecke, T., Hume, P., Marizcurrena, A., and Gill, R.T. (2006). A comparative study of metabolic engineering anti-metabolite tolerance in *Escherichia coli*. Metab. Eng. *8*, 227–239.

Brazeau, B.J., Gort, S.J., Jessen, H.J., Andrew, A.J., and Liao, H.H. (2006). Enzymatic activation of lysine 2,3-aminomutase from *Porphyromonas gingivalis*. Appl. Environ. Microbiol. *72*, 6402–6404.

Bro, C., Regenberg, B., Förster, J., and Nielsen, J. (2006). In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. Metab. Eng. *8*, 102–111.

Caspi, R., Foerster, H., Fulcher, C.A., Hopkinson, R., Ingraham, J., Kaipa, P., Krummenacker, M., Paley, S., Pick, J., Rhee, S.Y., et al. (2006). MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. *34*, D511–D516.

Casqueiro, J., Bañuelos, O., Gutiérrez, S., and Martín, J.F. (2001). Metabolic engineering of the lysine pathway for β-lactam overproduction in *Penicillium chrysogenum*. In Focus on Biotechnology, A. Van Broekhoven, F. Shapiro, and J. Anne, eds. (Dordrecht, Netherlands: Kluwer Academic Publishers), pp. 147–159.

Chang, Y.Y., and Cronan, J.E., Jr. (2000). Conversion of *Escherichia coli* pyruvate oxidase to an 'alpha-ketbutyrate oxidase'. Biochem. J. *352*, 717–724.

Chinen, A., Kozlov, Y.I., Hara, Y., Izui, H., and Yasueda, H. (2007). Innovative metabolic pathway design for efficient l-glutamate production by suppressing CO_2 emission. J. Biosci. Bioeng. *103*, 262–269.

Dürre, P., Böhringer, M., Nakotte, S., Schaffer, S., Thormann, K., and Zickner, B. (2002). Transcriptional regulation of solventogenesis in *Clostridium acetobutylicum*. J. Mol. Microbiol. Biotechnol. *4*, 295–300.

Edwards, J.S., Covert, M., and Palsson, B. (2002). Metabolic modeling of microbes: the flux-balance approach. Environ. Microbiol. *4*, 133–140.

Ellis, L.B.M., Roe, D., and Wackett, L.P. (2006). The University of Minnesota Biocatalysis/Biodegradation Database: the first decade. Nucleic Acids Res. *34*, D517–D521.

Endo, T., and Koizumi, S. (2001). Microbial conversion with cofactor regeneration using genetically engineered bacteria. Adv. Synth. Catal. *343*, 521–528.

Endy, D. (2005). Foundations for engineering biology. Nature *438*, 449–453.

Ewering, C., Lütke-Eversloh, T., Luftmann, H., and Steinbüchel, A. (2002). Identification of novel sulfur-containing bacterial polyesters: biosynthesis of poly(3-hydroxy-S-propyl-ω-thioalkanoates containing thioether linkages in the side chains. Microbiology *148*, 1397–1406.

Fenner, K., Gao, J., Kramer, S., Ellis, L., and Wackett, L. (2008). Data-driven extraction of relative reasoning rules to limit combinatorial explosion in biodegradation pathway prediction. Bioinformatics *24*, 2079–2085.

González-Lergier, J., Broadbelt, L.J., and Hatzimanikatis, V. (2005). Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways. J. Am. Chem. Soc. *127*, 9930–9938.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous protein expression. Trends Biotechnol. *22*, 346–353.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. *68*, 669–685.

Hatzimanikatis, V., Li, C., Ionita, J.A., Henry, C.S., Jankowski, M.D., and Broadbelt, L.J. (2005). Exploring the diversity of complex metabolic networks. Bioinformatics *21*, 1603–1609.

Hazer, B., and Steinbüchel, A. (2007). Increase diversification of polyhydroxyalkanoates by modification reactions for industrial and medical applications. Appl. Microbiol. Biotechnol. *74*, 1–12.

Hederos, S., Broo, K.S., Jakobsson, E., Kleywegt, G.J., Mannervik, B., and Baltzer, L. (2004). Incorporation of a single His residue by rational design enables thiol-ester hydrolysis by human glutathione transferase A1-1. Proc. Natl. Acad. Sci. USA *101*, 13163–13167.

Jankowski, M.D., Henry, C.S., Broadbelt, L.J., and Hatzimanikatis, V. (2008). Group contribution method for thermodynamic analysis of complex metabolic networks. Biophys. J. *95*, 1487–1499.

Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J.L., Betker, J.L., Tanaka, F., Barbas, C.F., III, et al. (2008). *De novo* computational design of retro-aldol enzymes. Science *319*, 1387–1391.

Jones, D.T., and Woods, D.R. (1986). Acetone-butanol fermentation revisited. Microbiol. Rev. *50*, 484–524.

Kane, J.F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Curr. Opin. Biotechnol. *6*, 494–500.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. *34*, D354–D357.

Kang, Z., Wang, Q., Zhang, H., and Qi, Q. (2008). Construction of a stress-induced system in *Escherichia coli* for efficient polyhydroxyalkanoates production. Appl. Microbiol. Biotechnol. *79*, 203–208.

Kaplan, J., and DeGrado, W.F. (2004). *De novo* design of catalytic proteins. Proc. Natl. Acad. Sci. USA *101*, 11566–11570.

Keasling, J.D. (2008). Synthetic Biology for Synthetic Chemistry. ACS Chem. Biol. *3*, 64–76.

Kelly, R.M., Leemhuis, H., Rozeboom, H.J., van Oosterwijk, N., Dijkstra, B.W., and Dijkhuizen, L. (2008). Elimination of competing hydrolysis and coupling side reactions of a cyclodextrin glucanotransferase by directed evolution. Biochem. J. *413*, 517–525.

Lee, S.Y., Park, J.H., Jang, S.H., Nielsen, L.K., Kim, J., and Jung, K.S. (2008). Fermentive Butanol Production by Clostridia. Biotechnol. Bioeng. *101*, 209–228.

Li, C., Henry, C.S., Jankowski, M.D., Ionita, J.A., Hatzimanikatis, V., and Broadbelt, L.J. (2004). Computational discovery of biochemical routes to specialty chemicals. Chem. Eng. Sci. *59*, 5051–5060.

Liao, H.H., Gokarn, R.R., Gort, S.J., Jensen, H.J., and Selifonova, O. (2007). Alanine 2,3-aminomutase. U.S. patent 7,309,597.

Lippow, S.M., Wittrup, K.D., and Tidor, B. (2007). Computational design of antibody-affinity improvement beyond *in vivo* maturation. Nat. Biotechnol. *25*, 1171–1176.

Liu, Y.H., Lu, F.P., Li, Y., Wang, J.L., and Gao, C. (2008). Acid stabilization of *Bacillus licheniformis* alpha amylase through introduction of mutations. Appl. Microbiol. Biotechnol. *80*, 795–803.

Loose, C., Jensen, K., Rigoutsos, I., and Stephanopoulos, G. (2006). A linguistic model for the rational design of antimicrobial peptides. Nature *443*, 867–869.

Machielsen, R., Leferink, N.G., Hendriks, A., Brouns, S.J., Hennemann, H.G., Daubetamann, T., and van der Oost, J. (2008). Laboratory evolution of *Pyrococcus furiosus* alcohol dehydrogenase to improve the production of (2S,5S)-hexanediol at moderate temperatures. Extremophiles *12*, 587–594.

Martin, V.J., Yoshikuni, Y., and Keasling, J.D. (2001). The *in vivo* synthesis of plant sesquiterpenes by *Escherichia coli*. Biotechnol. Bioeng. *75*, 497–503.

Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D., and Keasling, J.D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for the production of terpenoids. Nat. Biotechnol. *21*, 796–802.

Meinhold, P., Peters, M.W., Chen, M.M., Takahashi, K., and Arnold, F.H. (2005). Direct conversion of ethane to ethanol by engineered cytochrome P450 BM3. ChemBioChem *6*, 1765–1768.

Meinhold, P., Peters, M.W., Hartwick, A., Hernandez, A.R., and Arnold, F.H. (2006). Engineering cytochrome P450 BM3 for terminal alkane hydroxylation. Adv. Synth. Catal. *348*, 763–772.

Mermelstein, L.D., Papoutsakis, E.T., Petersen, D.J., and Bennett, G.N. (1993). Metabolic engineering of *Clostridium acetobutylicum* ATCC 824 for increased solvent production by enhancement of acetone fromation enzyme activities using a synthetic acetone operon. Biotechnol. Bioeng. *42*, 1053–1060.

Meurer, G., Biermann, G., Schütz, A., Harth, S., and Schweizer, E. (1992). Molecular structure of the mutlifunctional fatty acid synthetase gene of *Brevibacterium ammoniagenes*: its sequence of catalytic domains is formally consistent with a head-to-tail fusion of the two yeast genes *FAS1* and *FAS2*. Mol. Gen. Genet. *232*, 106–116.

Meyer, M.M., Silberg, J.J., Voigt, C.A., Endelman, J.B., Mayo, S.L., Wang, Z.-G., and Arnold, F.H. (2003). Library analysis of SCHEMA-guided protein recombination. Protein Sci. *12*, 1686–1693.

Meyer, M.M., Hochrein, L., and Arnold, F.H. (2006). Structure-guided SCHEMA recombination of distantly related β-lactamases. Protein Eng. Des. Sel. *19*, 563–570.

Mueller-Cajar, O., and Whitney, S.M. (2008). Evolving improved *Synechococcus rubisco* functional expression in *Escherichia coli*. Biochem. J. *414*, 205–214.

Müller, U., van Assema, F., Gunsior, M., Orf, S., Kremer, S., Schipper, D., Wagemans, A., Townsend, C.A., Sonke, T., Bovenberg, R., and Wubbolts, M. (2006). Metabolic engineering of the *E. coli* phenylalanine pathways for the production of D-phenylglycine (D-Phg). Metab. Eng. *8*, 196–208.

Niu, W., Molefe, M.N., and Frost, J.W. (2003). Microbial synthesis of the energetic material precursor 1,2,4-butanetriol. J. Am. Chem. Soc. *125*, 12998–12999.

Otey, C.R., Silberg, J.J., Voigt, C.A., Endelman, J.B., Bandara, G., and Arnold, F.H. (2004). Functional evolution and structural conservation in chimeric cytochromes p450: calibrating a structure-guided approach. Chem. Biol. *11*, 309–318.

Patel, M., Dornburg, V., Hermann, B., Roes, L., Hüsing, B., Overbeek, L., Terragni, F., and Recchia, E. (2006). Medium and Long-term Opportunities and Risks of the Biotechnological Production of Bulk Chemicals from Renewable Resources: The Potential of White Biotechnology (Utrecht, Netherlands: Utrecht University), September 2006.

Pitera, D.J., Paddon, C.J., Newman, J.D., and Keasling, J.D. (2007). Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. Metab. Eng. *9*, 193–207.

Pleiss, J. (2006). The promise of synthetic biology. Appl. Microbiol. Biotechnol. *73*, 735–739.

Przytycka, T., Srinivasan, R., and Rose, G.D. (2002). Recursive domains in proteins. Protein Sci. *11*, 409–417.

Reetz, M.T., and Carballeira, J.D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nat. Protoc. *2*, 891–903.

Reetz, M.T., Kahakeaw, D., and Lohmer, R. (2008). Addressing the numbers problem in directed evolution. ChemBioChem *9*, 1797–1804.

Rigoutsos, I., and Floratos, A. (1998). Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm. Bioinformatics *14*, 55–67.

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature *440*, 940–943.

Rodrigo, G., Carrera, J., Prather, K.J., and Jaramillo, A. (2008). DESHARKY: Automatic design of metabolic pathways for optimal cell growth. Bioinformatics *24*, 2554–2556.

Rothlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O., et al. (2008). Kemp elimination catalysts by computational enzyme design. Nature *453*, 190–195.

Saraf, M.C., Horswill, A.R., Benkovic, S.J., and Maranas, C.D. (2004). FamClash: A method for ranking the activity of engineered enzymes. Proc. Natl. Acad. Sci. USA *101*, 4142–4147.

Savage, D.F., Way, J., and Silver, P.A. (2008). Defossiling fuel: how synthetic biology can transform biofuel production. ACS Chem. Biol. *3*, 13–16.

Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., and Schomburg, D. (2004). BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res. *32*, D431–D433.

Searls, D.B. (1997). Linguistic approaches to biological sequences. Comput. Appl. Biosci. *13*, 333–344.

Searls, D.B. (2002). The language of genes. Nature *420*, 211–217.

Seker, T., Møller, K., and Nielsen, J. (2005). Analysis of acyl CoA ester intermediates of the mevalonate pathway in *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. *67*, 119–124.

Shi, C., Lu, X., Ma, C., Ma, Y., Fu, X., and Yu, W. (2008). Enhancing the thermostability of a novel beta-agarase AgaB through directed evolution. Appl. Biochem. Biotechnol. *151*, 51–59.

Steinbüchel, A., and Valentin, H.E. (1995). Diversity of microbial polyhydroxyalkanoic acids. FEMS Microbiol. Lett. *128*, 219–228.

Stemmer, W.P. (1994a). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. USA *91*, 10747–10751.

Stemmer, W.P.C. (1994b). Rapid evolution of a protein in vitro by DNA shuffling. Nature *370*, 389–391.

Tsuge, T., Hisano, T., Taguchi, S., and Doi, Y. (2003). Alteration of chain length substrate specificity of *Aeromonas caviae* R-enantiomer-specific enoyl-coenzyme A hydratase through site-directed mutagenesis. Appl. Environ. Microbiol. *69*, 4830–4836.

Villalobos, A., Ness, J.E., Gustafsson, C., Minshill, J., and Govindarajan, S. (2006). Gene Designer: a synthetic biology tool for construction artificial DNA segments. BMC Bioinformatics *7*, 285.

Voigt, C.A., Martinez, C., Wang, Z.-G., Mayo, S.L., and Arnold, F.H. (2002). Protein building blocks preserved by recombination. Nat. Struct. Biol. *9*, 553–558.

Wang, J., and Yu, H.-Q. (2007). Biosynthesis of polyhydroxybutyrate (PHB) and extracellular polymeric substances (EPS) by *Ralstonia eutropha* ATCC 17699 in batch cultures. Appl. Microbiol. Biotechnol. *75*, 871–878.

Werpy, T., and Petersen, G. (2004). Top Value Added Chemicals from Biomass, Vol 1: Results of Screening for Potential Candidates from Sugars and Synthesis Gas (Oak Ridge, TN: U.S. Department of Energy).

Woods, D.R. (1995). The genetic engineering of microbial solvent production. Trends Biotechnol. *13*, 259–264.

Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., et al. (2006). The Universal Protein Resource (UniProt): an exapanding universe of protein information. Nucleic Acids Res. *34*, D187–D191.

Yin, E., Le, Y.L., Pei, J.J., Shao, W.L., and Yang, Q.Y. (2008). High-level expression of the xylanase from *Thermomyces lanuginosus* in *Escherichia coli*. World J. Microbiol. Biotechnol. *24*, 275–280.

Yoshikuni, Y., Dietrich, J.A., Nowroozi, F.F., Babbitt, P.C., and Keasling, J.D. (2008). Redesigning enzymes based on adaptive evolution for optimal function in synthetic metabolic pathways. Chem. Biol. *15*, 607–618.

Zanghellini, A., Jiang, L., Wollacott, A.M., Cheng, G., Meiler, J., Althoff, E.A., Rothlisberger, D., and Baker, D. (2006). New algorithms and an in silico benchmark for computational enzyme design. Protein Sci. *15*, 2785–2794.

Zha, W., Shao, Z., Frost, J.W., and Zhao, H. (2004). Rational pathway engineering of type I fatty acid synthase allows the biosynthesis of triacetic acid lactone from d-glucose in vivo. J. Am. Chem. Soc. *126*, 4534–4535.

Zhang, K., Sawaya, M.R., Eisenberg, D.S., and Liao, J.C. (2008). Expanding metabolism for biosynthesis of nonnatural alcohols. Proc. Natl. Acad. Sci. USA *105*, 20653–20658.