PART I

INTRODUCTION CHAPTERS

CHAPTER 1

BIOTECHNOLOGY TOOLS FOR GREEN SYNTHESIS: ENZYMES, METABOLIC PATHWAYS, AND THEIR IMPROVEMENT BY ENGINEERING

ROMAS J. KAZLAUSKAS
Department of Biochemistry, Molecular Biology and Biophysics and
The Biotechnology Institute, University of Minnesota, Saint Paul, Minnesota

BYUNG-GEE KIM
Department of Chemical and Biological Engineering, Seoul National University,
Seoul, Korea

1.1 INTRODUCTION

Green chemistry is the design of products and processes that eliminate or reduce waste, toxic, and hazardous materials. Green chemistry is not a cleanup approach, but a prevention approach. Preventing problems is inevitably easier and less expensive than contending with difficulties after they occur.

The risk associated with a chemical depends both on how dangerous it is (hazard) and on one's contact with it (exposure) (Figure 1.1). In the past, governments and industry focused on reducing risk by minimizing exposure. Rules limit the exposure of workers to hazardous chemicals and the release of these chemicals into the environment. This approach is expensive; it is difficult to establish a safe level of hazardous chemicals, and currently, only a small fraction of the chemicals manufactured are regulated.

The green chemistry approach focuses on reducing risk by reducing or eliminating the hazard. Hazardous materials are eliminated by, for example, replacing them with nonhazardous ones. Hazardous materials are eliminated also by increasing the yield of a reaction, as higher yield eliminates some of the waste from that reaction. In addition, higher yield allows any preceding reactions to

Biocatalysis for Green Chemistry and Chemical Process Development, First Edition.
Edited by Junhua (Alex) Tao and Romas Kazlauskas.

risk = hazard • exposure

Figure 1.1 The risk associated with a chemical depends on both how dangerous it is (hazard) and one's contact with it (exposure). In the past, the focus was on minimizing exposure by rules that limit the amounts of hazardous chemicals in air and water. The green chemistry approach is to eliminate or reduce hazardous materials. This change requires redesigning of synthetic approaches.

be carried out on a smaller scale, thus eliminating some of the waste from these steps as well. This prevention approach saves money, as fewer raw materials are needed and the cost of treatment of waste is reduced.

Reducing costs and being environmentally friendly are goals that everyone agrees on. Why has this not been done before? One reason is that environmental costs were ignored in the early days of the chemical industry. Now that more of the cleanup cost falls on the manufacturer, there is a big financial incentive to be greener. Another reason is that chemists in research and design laboratories did not view environmental hazards as their problems. It was something to be fixed later in the scale-up stage. The green chemistry approach changes this thinking. By thinking about hazards and environmental consequences at the research and design stage, many problems are prevented and do not need a fix later. The principles of green chemistry outlined by Anastas and Warner [1] provide specific guidelines for what to look for at the research and design stage to make a greener process. These principles are discussed below in the context of biocatalysis.

The first use of biochemical reactions for organic synthesis was probably in 1858, when Louis Pasteur resolved tartaric acid by using a microorganism to destroy one enantiomer [2]. In spite of this early demonstration, chemists have used biocatalysis only sporadically. Chemists gradually recognized the potential of biochemical reactions, but there were both practical and conceptual hurdles. Practical problems were how to get the enzymes and how to stabilize them. The conceptual problems were beliefs that enzymes accept only a narrow range of biochemical intermediates as substrates, and that enzymes are too complex to consider engineering them for key properties like stability, stereoselectivity, substrate range, and even reaction type. The recent advances in biotechnology have solved many of the practical problems, and the increased understanding of biochemical structures and mechanisms has made biocatalysts more understandable to chemists. This chapter surveys the state of the art for engineering biocatalysts for chemistry applications. If you find an enzyme that catalyzes your desired reaction, regardless of how poor the enzyme is, it is highly likely that it can be engineered into an enzyme suitable for industrial and large-scale use.

1.2 THE NATURAL FIT OF BIOCATALYSIS WITH GREEN CHEMISTRY

Biotechnology methods fit naturally to the goals and principles of green chemistry. Green chemistry, or sustainable chemistry, seeks to integrate industrial

manufacturing practice with the natural world. This natural world is the biological world, where sustainability and recycling are integral parts. Use of the biological methods for industrial manufacturing is an excellent starting point to create a green process. In some cases, biotechnology tools, unlike chemical tools, are even edible. Baker's yeast, used to make bread, also catalyzes the reduction of various carbonyl compounds. Lipases are the most commonly used enzymes for biocatalysis. These enzymes are also eaten in multigram amounts by patients with pancreatic insufficiency and in smaller amounts when food-grade lipases are used in the manufacture of cheese.

Although the "bio" part of biocatalysis makes it environmentally friendly, it is the "catalysis" part that provides the green chemistry advantage. Catalysis, in place of reagents, converts many substrate molecules to products and eliminates the need for stoichiometric reagents. Catalysis is fast, so the reaction may not need to be heated. This saves energy and may eliminate side reactions that occur at higher temperatures. Catalysis is selective, eliminating the need to add and remove protective groups or use auxiliaries to control reactivity. Catalysis can enable complex and otherwise difficult reactions. This ability can eliminate steps and simplify syntheses.

The 12 principles of green chemistry outline the design goals for synthesis. Making progress toward any one of these goals will make a synthesis greener; progress toward several goals is of course better. Table 1.1 lists some suggestions on how biocatalysis can help a synthesis toward these goals.

Given the wide-ranging ways in which biocatalysis can contribute to green chemistry, it is no surprise that many recent winners of the US Environmental Protection Agency's Presidential Award in Green Chemistry have used biocatalysis as the key improvement technique (Table 1.2). Many other winners have also used biocatalysis; these examples were selected to show the range of techniques used.

The 2010 winner of the Greener Reaction Conditions Award was an improved synthesis of sitagliptin, the active ingredient in an oral type 2 diabetes drug. Merck and Codexis collaborated to develop the key step, which is a transaminase-catalyzed formation of a chiral amine (R form) [3]. The starting transaminase required 27 amino acid substitutions to fit the large substrate in the active site, to increase the reaction rate and to stabilize the enzyme to the reaction conditions. The previous synthesis of sitagliptin was already a good synthesis, which won a green chemistry award in 2006 [4]. The new synthesis, which added a biocatalysis step, eliminates four steps, including an asymmetric hydrogenation using a rhodium catalyst that requires a high-pressure reactor.

LS9, Inc., winners of the 2010 prize for a small business, engineered new metabolic pathways into microorganisms to make biofuels [5]. For example, to make *Escherichia coli* bacteria produce biodiesel, they added the genes for plant thioesterases to divert normal fatty acid biosynthesis into synthesis of several fatty acids suitable for biodiesel. Next, they added genes for enzymes to make ethanol and an enzyme to couple the ethanol and fatty acids to make fatty acid ethyl esters, which can be used for biodiesel. Finally, the researchers added genes for

TABLE 1.1 How Biocatalysis Follows the Twelve Principles of Green Chemistry

Prevent waste
The high selectivity of enzymes and their ability to carry out difficult chemical reactions eliminate synthesis steps and the associated waste in multistep reactions. Reagents and solvents for eliminated steps are not needed; higher yield and selectivity in the remaining steps also eliminate waste. Biocatalysis usually uses water as the solvent, which eliminates organic solvent waste
Design safer chemicals and products
Biocatalysts are typically used only for manufacture and are not the product themselves. In some cases, a biocatalyst can be a product, such as an enzyme-based drain cleaner that replaces a drain cleaner based on a strong acid or base. Biocatalysis can enable the manufacture of new products, such as biodegradable polyesters, which is not practical to manufacture chemically
Design less hazardous chemical syntheses
Eliminating hazardous steps or replacing hazardous reagents with biocatalysts makes the manufacturing process safer
Use renewable feedstock
Biocatalysts are typically manufactured by growing microorganisms that secrete the biocatalysts. In other cases, the whole cells can be used as the catalysts. The feedstock to grow microorganisms are sugars and amino acids, which are renewable
Use catalysts, not stoichiometric reagents
Biocatalysts are highly efficient catalysts, that is, each enzyme molecule converts thousands to millions of substrate molecules to product. Thus, reactions are fast and the turnover number is high
Avoid chemical derivatives
The high selectivity of enzymes usually eliminates the need for protective groups and for resolutions involving chemical derivatives
Maximize atom economy
The high selectivity of enzymes and their ability to carry out difficult chemical reactions allow most of the starting material (including all reagents) atoms to be converted to the product. Eliminating synthesis steps, reagents, and derivatives reduces the number of atoms of the starting materials needed
Use safer solvents and reaction conditions
Biocatalysis typically uses water as the solvent, neutral pH, and room temperature. If needed for substrate solubility or faster reaction, biocatalysis can tolerate a wide range of reaction conditions, including organic solvents
Increase energy efficiency
Because biocatalysts are fast catalysts, one does not need to heat a reaction, thus saving energy. One also rarely needs to cool a reaction, which also requires energy, since the reaction rate can be reduced by adding less catalyst. The selectivity of biocatalysts is high, so there is no need to reduce the temperature to increase selectivity as there is with chemical reagents and catalysts

TABLE 1.1 (*Continued*)

Design for degradation
Biocatalysts are biodegradable; some catalysts (e.g., baker's yeast and food-grade proteases) are even edible
Analyze in real time to prevent pollution
No special advantage of biocatalysis. Biocatalytic reactions can be monitored in real time by analyzing pH, oxygen, cofactor concentrations, and other parameters
Minimize the potential for accidents
Mild reaction conditions for biocatalysis typically minimize the potential for explosions, fire, and accidental release. Typical reaction conditions are near room temperature, atmospheric pressure, nonflammable solvent, and dilute solutions

cellulases and xylanases to this *E. coli* bacteria so that instead of using glucose as the starting material for this synthesis, the *E. coli* could use inexpensive biomass. The amount of biodiesel produced is at least 10-fold too low for commercial use, but further engineering will likely increase this amount.

The 2009 prize winner, Eastman Kodak Company, used lipase B from *Candida antarctica*, the most widely used lipase for synthesis. This lipase is highly active in organic solvents and shows high stereoselectivity. The high activity in organic solvents was the key characteristic for this work. It allowed researchers to eliminate the solvent and directly react the starting acid (or an ethyl ester) and alcohol together [6]. The mild reaction conditions allow researchers to use delicate unsaturated fatty acids that would undergo side reactions if an acid or a base is used to catalyze the ester formation. In diols, the enzyme's regioselectivity ensured that only one alcohol group reacted.

The 2006 prize was for a synthesis of a key fragment of atorvastatin (active ingredient of Lipitor), a cholesterol-lowering drug [7]. Codexis engineered 180 variations of the ketoreductase. All variants are stable and highly active, but differ in their substrate specificity. Screening these ketoreductases toward a target substrate identifies which ones have the correct substrate specificity, and further protein engineering to improve the ketoreductases is possible. Codexis demonstrated this approach for the atorvastatin fragment, but a similar approach should work for other problems.

The 2005 prize was for engineering a metabolic pathway. Organisms that naturally produce bacterial polyesters grow slowly and are inconvenient to work with. Researchers at Metabolix transferred the entire pathway into *E. coli* to enable it to produce polyesters [8]. Optimization of the metabolism increased the yield to practical levels for manufacture.

The last example is a plant cell fermentation for a multistep synthesis of paclitaxel (Taxol), a complex anticancer drug containing eight stereocenters [9]. Chemical synthesis is impractical (40 steps, $\sim$2% overall yield) as is its isolation from the yew tree, since the amounts are so low. The existing process isolated a precursor of paclitaxel from leaves and twigs followed by 11 chemical

TABLE 1.2 Select Recent Winners of the US Presidential Green Chemistry Awards That Use Biocatalysis

Winner	Discovery	Green Features	Other Advantages	Biotechnology Tools
Merck & Codexis (2010)	Improved synthesis of sitagliptin, an oral diabetes drug.	Eliminates several steps, including an asymmetric hydrogenation using a rhodium catalyst that required a high-pressure reactor 10–13% higher yield and 19% less waste.	Can use existing equipment for synthesis New transaminase may be useful in other syntheses	Combined rational design and directed evolution methods to improve the starting transaminase activity 25,000-fold
LS9 (2010)	Engineered a new metabolic pathway in *Escherichia coli* to make biodiesel	Diverts fatty acid biosynthesis to make fatty acid ethyl esters	Can grow on inexpensive biomass	Cloning enzymes from other organisms into *E. coli*
Eastman Chemical Company (2009)	A solvent-free biocatalytic synthesis of esters as ingredients for cosmetics and personal care products	Eliminates >10 L of organic solvent/kg of ester Purification is often not needed	Allows use of delicate raw materials such as unsaturated fatty acids Regioselective monoesterification of diols such as 4-hydroxy-benzyl alcohol	Immobilized lipase catalyzes esterification and transesterification

Codexis (2006)	A simplified manufacturing process for a key intermediate in Liptor.	Reduces the formation of by-products Reduces generation of waste Eliminates the need for hydrogen gas Reduces solvent use Increases yield	Fewer unit operations, including no fractional distillation of the product Less purification equipment needed	Genetic engineering methods to create up to 4000-fold improved enzymes that catalyze the 3-step synthesis
Metabolix (2005)	Bioplastics (polyhydroxyalkanoates) made by genetically engineered organisms	Made from renewable resources (sugars) instead of petroleum Biodegradable Saves energy	Reduces greenhouse gas emissions	Add new biochemical pathways to microorganisms and optimize metabolism
Bristol Myers Squibb (2004)	Synthesis of paclitaxel using plant cell fermentation technology	Eliminates an estimated 14,000 pounds of hazardous chemicals and materials per year Eliminates 10 solvents and six drying steps Saves energy	Eliminates harvest of yew trees Increases the amount of available drug	Growth of specific paclitaxel-producing cells in fermentation tanks. Optimize metabolism

http://www.epa.gov/greenchemistry/pubs/pgcc/technology.html (accessed 21 September 2010).

transformations to the product. The plant cell fermentation eliminates all of these steps because the specific cell line yields paclitaxel directly.

1.3 WHY BIOCATALYSTS NEED TO BE ENGINEERED

Three of the six examples in Table 1.2 involve unnatural substrates. The ability of the enzymes to accept these synthetic intermediates is due to good luck, not evolutionary pressure in nature. One reason to engineer enzymes is to better accommodate unnatural substrates. Enzymes involved in the uptake of nutrients (lipases, proteases) often have broad substrate specificity because they must act on a broad range of possible food sources. Similarly, enzymes involved in detoxification (*P*450 monooxygenases, glutathione *S*-transferases) also have a broad substrate range to accommodate many possible natural substrates. In contrast, enzymes involved in primary metabolism, such as glycolysis, typically have a narrow substrate range. The broad substrate range of many enzymes is critical to their usefulness because it allows chemists to use enzymes to catalyze reactions on their synthetic intermediates, and not just on biochemical intermediates.

It is also due to luck that these enzymes often show high enantioselectivity toward these synthetic intermediates. The ability of enzymes to both accept a wide range of substrates, which suggests an open active site, and show high selectivity, which suggests a restricted active site, is somewhat surprising. Protein engineering can reshape the active site to increase the selectivity of the enzyme.

Another common reason to engineer enzymes is to increase their stability under the reaction conditions. Reaction conditions can differ dramatically from those present in a cell. Reaction conditions may involve high temperatures, extremes of pH, high substrate and product concentrations, oxidants, and organic cosolvents. Sometimes an enzyme must tolerate these conditions for only a few minutes or hours, but in a continuous manufacturing process, an enzyme may need to tolerate them for months.

A third reason to engineer enzymes and metabolic pathways is to create new reactions or new biochemical pathways. For example, Ran and Frost expanded the substrate range of an aldolase and thus created a new metabolic pathway to make shikimic acid for an influenza drug synthesis (Section 1.5).

Combining enzymes from different organisms and biochemical pathways can create a new biochemical pathway. The artemisinin synthesis discussed in Chapter 7 is an excellent example of this type of protein engineering and several other examples are discussed later in this chapter (Section 1.5).

1.4 STRATEGIES TO ENGINEER ENZYMES

The strategy chosen to engineer an enzyme depends on both the property being engineered and how much information is known about the enzyme, Table 1.3. Some engineering approaches require that the structure of the enzyme known,

TABLE 1.3 Examples of Protein Engineering of Biocatalysts

Strategy	Enzyme	Details	Improvement
Rational design	Phosphodiesterase	Decreases the size of one pocket to increase enantioselectivity; reverses the sizes of two pockets to reverse enantioselectivity	Increase or reverse enantioselectivity
Rational design	T4-lysozyme	Introduces proline residues and/or removed glycine residues in loop regions, introduces disulfide bonds	Higher thermostability
Rational design		Designs an enzyme that catalyzes a retro-aldol cleavage; not all designs gave catalytic activity	New catalytic activity
Directed evolution	Subtilisin	Random mutagenesis using error-prone PCR and screening to identify more stable variants	>250-fold higher activity in 60% DMF
Directed evolution	Phosphite dehydrogenase for NAD(P)H regeneration	Random mutagenesis using error-prone PCR and screening to identify more stable variants; combines the 12 best substitutions using site-directed mutagenesis	>7000-fold higher thermal stability
Directed evolution	Nitrilase for synthesis of Lipitor side chain	Saturation mutagenesis at each position and screening to find the best amino acid substitution	Increases enantioselectivity to >100
Directed evolution		Saturation mutagenesis at selected locations in the active site	Increases E from 12 to >50
Bioinformatics	Phytase	Amino acid substitutions to bring the amino acid sequence closer to the consensus sequence	Higher thermostability
Bioinformatics	Thermolysin	Replaces amino acids with those found at equivalent positions in naturally occurring, more thermostable variants	Stable to boiling water >2 h
Statistical correlation	Halohydrin dehalogenase for synthesis of Lipitor side chain	Identifies beneficial or detrimental effects of various amino acid substitutions (ProSAR) and combines 35 beneficial mutations	>4000-fold higher volumetric productivity
Statistical correlation	Humulene synthase	Correlates amino acid substitution in the active site to which product is formed	Changes major product from humulene to sibirene
Statistical correlation	Cellulase	Recombines eight blocks from three parents; identifies those blocks that increase thermostability	Higher thermostability

while others do not require any structural information, but require larger libraries and more screening. In some cases, typically for thermostability, screening is easy, while in other cases, for example, less product inhibition under process conditions, the screening can be slow and limited to several hundred variants.

1.4.1 Rational Design

Rational design requires structural information and a molecular level understanding of how the protein structure influences the property to be engineered. This requires answers to questions such as "What is the molecular basis of catalysis?" and "What determines efficient protein expression in different hosts?" These questions cannot be completely answered. This inability limits the success of rational protein engineering.

In spite of these difficulties, there are a number of examples where rational design has worked. For example, changing the pocket sizes in an organophosphorus hydrolase predictably altered the enantioselectivity (E) [10]. The wild-type enzyme favors the S_P enantiomer of ethyl phenyl 4-nitrophenyl phosphate ($E = 21$). Decreasing the site of the small site using a Gly60Ala mutation increased enantioselectivity to $E > 100$. Four other amino acid substitutions to reverse the relative sizes of the subsites reversed the enantiopreference.

Protein stability can also be engineered rationally in many cases. Protein stability depends on the equilibrium between the folded and unfolded states. The hydrophobic effect and inter-residue interactions stabilize the folded state, while the entropy associated with main chain flexibility favors the unfolded state. Rational strategies that reduce the flexibility of the unfolded state increase protein stability. In particular, introducing disulfide cross-links or proline residues (a less flexible residue) or removing glycine residues (the most flexible residue) all stabilized enzymes [11].

The most advanced rational design uses quantum mechanical models of the transition state and search strategies that test many possible conformations of the enzyme. Baker and collaborators designed enzymes that catalyzed an unnatural model reaction, the Kemp elimination [12], a retro-aldol cleavage [13], and a Diels–Alder cycloaddition [14] using this approach. These calculations modeled hundreds of thousands of possibilities and narrowed them to several dozen candidates. Synthesis of these predicted variants revealed a number with detectable catalytic activity.

1.4.2 Directed Evolution

Directed evolution needs no structure. One makes random changes in an enzyme to create many enzymes, and then screens these enzymes for the ones that show improvements. This has to be repeated until the enzyme is sufficiently improved. Chen and Arnold were the first to demonstrate that directed evolution can improve a biocatalyst [15]. They improved the stability of subtilisin in organic solvents.

This problem was well suited to directed evolution because the molecular basis for enzyme stability in organic solvents is not well understood and therefore cannot be engineered rationally. Chen and Arnold created amino acid substitutions throughout the protein and screened to find the more solvent-stable variants. After three rounds of mutagenesis and screening, they had a subtilisin variant with 10 amino acid substitutions that was 250-fold more active in 60% dimethylformamide (DMF) than the wild type. Using a similar approach, Zhao and coworkers increased the stability of a dehydrogenase >7000-fold [16] and DeSantis and coworkers increased the enantioselectivity of a nitrilase [17]. In the last example, researchers did not have structural information for the nitrilase, so they could not use rational design.

If some structure information is available, it can speed up directed evolution by eliminating regions in the protein that are less likely to contain a solution. For example, residues close to substrate (either direct contact or the next sphere of residues) are more likely to yield big improvements in enantioselectivity than distant residues [18]. For example, Horsman *et al.* [19] increased the enantioselectivity of a *Pseudomonas fluorescens* esterase by targeting only four amino acid residues in the active site ($4 \times 19 = 76$ possibilities). They found two variants with an enantioselectivity of ~60 toward 3-bromo-2-methyl propanoate esters, which was a dramatic increase over the wild-type enantioselectivity of 12. The focus on the active site is a good approach to changing enantioselectivity or substrate specificity, but mutations that improve thermal stability, catalytic activity, and probably stability toward organic solvents are scattered throughout the protein.

Amino acid substitutions are the most common way to create enzyme variants (review [20]). The most common methods to make these substitutions are error-prone polymerase chain reaction (epPCR) and saturation mutagenesis. The epPCR approach is simple and makes random single amino acid substitutions throughout the protein, but likely misses some substitutions because they require the statistically unlikely two nucleotide substitutions in a single codon [21]. Site saturation mutagenesis gives all possible substitutions at selected locations. Saturation mutagenesis at every position (one position at a time) is also possible, but is expensive and requires a large screening effort. Another approach to making protein variants is DNA shuffling, which exchanges longer sections of proteins between two or more parents. The sections will differ in several amino acid substitutions, but may also include insertions and deletions.

1.4.3 Bioinformatics Approaches

Bioinformatics compares the amino acid or DNA sequence of the biocatalyst with other known sequences. Genome sequencing projects, enabled by the rapid advances in DNA sequencing methods, have created vast amounts of sequence information. The amount of this information is much larger and broader than the amount of functional or structural information. Extracting information for protein engineering from these sequences relies on the hypothesis that biocatalyst with

similar sequences are related in evolutionary history. This relatedness means that they will have similar function and similar structure.

One straightforward application of bioinformatics is to search for additional variants of an enzyme in the sequence databases. Starting from one or a few enzymes that catalyze the desired reaction, one searches for similar amino acid sequences. These sequences likely correspond to enzymes that also catalyze the desired reaction, but may have altered substrate specificity, stability to reaction conditions, or thermostability. For example, Fraaije *et al*. identified a thermostable Baeyer–Villiger monooxygenase enzymes by searching for a sequence characteristic of Baeyer–Villiger monooxygenases in the genome of a thermophile [22].

When the current variant of an enzyme already has many desirable characteristics, it is preferable to engineer the remaining characteristics instead of discovering new enzymes that may be lacking in other ways. Bioinformatics can also guide the engineering of more stable enzymes in two ways. First, sequence comparison can identify similar enzymes in thermophilic organisms. Some of the sequence differences account for the different stability. If one can identify which differences are most likely to lead to increase stability, then making these changes in the less stable enzyme can increase its stability. For example, Bae and Phillips stabilized adenylate kinase (T_m increased by 5°C.) by adding three salt bridges that they identified in the adenylate kinase from a thermophile [23].

The second approach to using bioinformatics to engineer more stable enzymes is the consensus sequence concept. The underlying hypothesis is that conserved amino acids contribute most to stability. If an amino acid substitution yields an unstable protein, the organism will not survive, so the conserved amino acids are more likely to contribute to stability than the nonconserved amino acids. The stabilization strategy is to compare the related sequences and to engineer the target protein to resemble the consensus sequence, that is, the most abundant amino acid at each position. Lehman and coworkers dramatically increased the thermostability of phytase using this approach; T_m increased up to 35°C from 55 to 90°C. The new enzyme contained >10 amino acid substitutions, where each substitution contributed slightly to the stability [24].

Bioinformatics is also the basis for homology modeling—the extrapolation of three-dimensional structures to proteins that have similar amino acid sequences (>35% identical amino acids). These three-dimensional structures are less accurate than experimentally derived structures, but can be a good starting point for rational design. A homology model can identify the active site of an enzyme, so saturation mutagenesis to change the shape of the active site can be attempted. For example, Keasling and coworkers made a homology model of γ-humulene synthase, and then targeted the residues predicted to be in the active site for mutagenesis [25]. More details about this experiment are in the next section.

1.4.4 Statistical Correlation Approaches

Unlike rational design, which uses molecular design principles to improve a desired property, statistical correlation approaches take a more empirical

approach. A set of enzyme variants are tested, and then statistics are used to correlate the changes with the improvements. Replacement of changes that degrade the property with neutral or beneficial changes improves the enzyme.

The best example is the ProSAR (protein structure activity relationship) approach used by Codexis researchers to improve the reaction rate of a halohydrin dehalogenase >4000-fold [4]. Researchers made random amino acid substitutions (an average of 10) in the dehalogenase and measured catalysis by the variants. Then, they made a statistical correlation whether a particular substitution was beneficial. For example, variants that contained a F186Y substitution were, on average, better than those that did not. Some variants that contained F186Y were poor because of the detrimental effects of other mutations, but the statistical analysis identified that, on average, it was a beneficial mutation. The final improved enzyme contained 35 amino acid substitutions among its 254 amino acids. Combining beneficial mutations and removing deleterious ones yields an improved enzyme.

γ-Humulene synthase catalyzes the cyclization of farnesyl diphosphate via cationic intermediates to γ-humulene in 45% yield, but forms 51 other sesquiterpenes in smaller amounts. Keasling and coworkers substituted amino acid residues in the active site and identified the contribution of each one to the product distribution [25]. Substitutions were combined to favor formation of one of the other sesquiterpenes. For example, a triple substitution created an enzyme that formed 78% sibirene. The contribution of each substitution was additive.

In another example, Arnold and coworkers improved the thermal stability of cellulases [26]. Instead of single amino acid substitutions, they recombined peptide fragments from three parents to make the variants that contained several substitutions as a group. They divided three parent cellulases into eight blocks and then recombined these blocks to make new cellulases. They made a set of 24 variants, measured their thermal stability, and identified the contribution of each block. For example, block 6 from parent 3 tended to stabilize the cellulases. Combining the stabilizing blocks and removing the destabilizing blocks yielded more stable cellulases.

Although this is an empirical approach that focuses on what works rather than why it works, it does not preclude the use of molecular design principles in choosing which changes to make. For example, the Codexis researchers hypothesized that changes nearest the active site would likely change the substrate fit for the halohydrin dehalogenase and therefore directed many of the amino acid substitutions to this region. Finally, by carefully examining the changes in protein structure that work, researchers may learn the molecular basis for the improvements.

1.4.5 Multiple Criteria

Practical applications require that the biocatalysts should meet multiple criteria: they must be stable at the temperature for the reaction, they must tolerate the pH and solvent of the reaction, they must tolerate high concentration of the starting material and product, they must show high selectivity, and they must react quickly with the substrate even when it is an unnatural compound. Often, one starts with

a catalyst that meets some of these criteria and tries to engineer in the missing properties. These multiple requirements make engineering more difficult than engineering a single property because engineering to improve one property must not destroy other properties in the process. For rational engineering, this means that you must understand the molecular basis of all of them, which is currently not possible. For methods that rely on screening, this means that you must screen for all the important properties.

For example, Schmidt and coworkers identified an esterase variant with three amino acid substitutions that showed higher enantioselectivity (E) toward the synthetic building block (S)-but-3-yn-2-ol ($E = 89$ as against $E = 3$ in wild type). Unfortunately, these substitutions degraded protein expression. Bacteria produced high amounts of the soluble starting esterase, but only small amounts of the variant, and most of it was in an unfolded insoluble form [27]. Additional experimentation revealed a variant with only two amino acid substitutions from wild type that was both highly enantioselective and efficiently produced in soluble form.

1.5 ENGINEERING OF METABOLIC PATHWAYS

Multistep metabolic pathways offer the opportunity for complex syntheses, but require use of whole cells. Examples of multistep processes on commercial scales are fermentation of glucose to ethanol by yeast, fermentation of glucose to citric acid by *Aspergillus* fungi, and production of penicillin G by *Penicillium* fungi. Although isolated enzymes can be used much like chemical catalysts in a wide range of temperatures, solvents, and reactors, whole cells are more limited in the reaction conditions they tolerate. In addition, they require the ability to work with microorganisms and typically use dilute aqueous solutions. The key advantages of whole cells are that they can stabilize enzymes that are difficult to isolate and that they can contain multiple enzymes. An isolated enzyme approach with multiple enzymes would require multiple enzyme isolations and optimization, which would eliminate the advantages of isolated enzymes.

Metabolic pathways are not limited to those existing in nature; the sections below highlight current strategies to create new metabolic pathways [28]. These pathways may be more efficient routes to existing products, or they may be new routes to create new products, Table 1.4.

1.5.1 New Pathways To Increase Yields

Combining existing pathways from different organisms created a pathway that yielded high concentrations of 1,3-propanediol used in polymers on a multi-ton scale. Several microorganisms convert glycerol to 1,3-propandiol, but Dupont and Genencor engineered a new pathway that allows use of glucose, a less expensive carbon source [29], Figure 1.2a. The pathway combines three different pathways from three different organisms in an *E. coli* strain. The first pathway is the naturally occurring glycolysis pathway in *E. coli*, which converts glucose to dihydroxyacetone phosphate. The second pathway consists of two enzymes added from the

TABLE 1.4 Examples of Engineering Metabolic Pathways

Product/Substrate	Microorganism	Details
1,3-propanediol from glucose	Recombinant *E. coli*	The natural glycerol to 1,3-propandiol pathway is combined with a glucose to glycerol pathway. Glucose as a carbon source is less expensive than fermentation grade glycerol. 1,3-Propanediol is used for synthesis of a polyester (poly(trimethylene terephthalate)). This biocatalytic process uses 40% less energy than conventional processes, and reduces greenhouse gas emissions by 20%
Shikimate from glucose	Recombinant *E. coli*	Shikimate, used for the synthesis of oseltamivir phosphate (Tamiflu®), is made by the aromatic amino acid biosynthetic pathway. The natural pathway uses phosphoenolpyruvate as a starting compound, but the engineered pathway uses the more abundant pyruvate and gives larger amounts of product
Copolymer of β-hydroxybutyrate and β-hydroxyvalerate from glucose and propionic acid	*Rastonia eutrophus*	Excess glucose is converted to poly(β-hydroxybutyrate) via acetyl-CoA. Co-feeding with propionic acid makes propionyl-CoA, which results in a copolymer that is more flexible and useful than the natural homopolymer
Higher alcohols from glucose	Recombinant *E. coli*	An added decarboxylase diverts 2-keto acids from the normal amino acid biosynthesis to form aldehydes. Reduction of the aldehydes yields alcohols, which may be useful as biofuels

yeast *Saccharomyces cerevisiae*, which convert the dihydroxyacetone phosphate to glycerol. The final pathway consists of several enzymes from *Klebsiella pneumonia*, which convert the glycerol to 1,3-propandiol. Surprisingly, researchers found that an uncharacterized oxidoreductase in *E. coli* worked better than 1,3-propandiol dehydrogenase from *Klebsiella*, so the production strain uses this *E. coli* oxidoreductase in this pathway. The production strain also incorporates gene deletions that eliminate nonproductive reactions. Another reason to move a pathway from one organism to another is to circumvent native regulation and thus make higher concentrations of product.

Ran and Frost designed a new pathway to shikimic acid by inventing a new step [30], Figure 1.2b. Shikimic acid is a precursor of oseltamivir phosphate (Tamiflu®), an influenza drug. The normal pathway to shikimic acid uses phosphoenol pyruvate as a starting compound. Unfortunately, low levels of

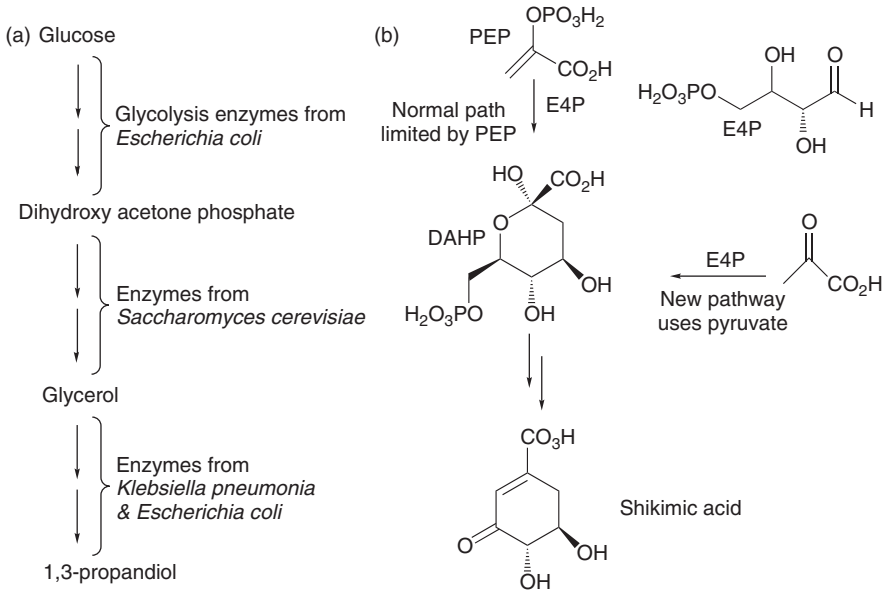

Figure 1.2 Two examples of strategies to engineer more efficient metabolic pathways to existing products. (a) Combining existing pathways from different organisms creates new pathways. Three pathways—conversion of glucose to dihydroxyacetone phosphate, conversion of dihydroxyacetone phosphate to glycerol, and glycerol to 1,3-propandiol—are combined in a strain of *E. coli* for the manufacture of 1,3-propandiol. (b) The amount of phosphoenolpyruvate limits the yield of shikimic acid from the natural pathway. A new pathway, created by expanding the substrate range of an aldolase, uses the more abundant pyruvate and generates higher amounts of shikimic acid.

phosphoenol pyruvate limit the amount of shikimic acid that can be produced. The solution was introducing a new enzyme, an aldolase that can use the more abundant pyruvate as the starting compound. The new enzyme was created by changing the substrate specificity of an existing aldolase by directed evolution. The new pathway gave higher yields of the desired compound on glucose.

1.5.2 New Pathways To Make Different Products

One way to get a new product from a metabolic pathway is to add a new intermediate that can enter the pathway. One example is the production of a mixed bacterial polyester, Figure 1.3a. Many bacteria, such as *Rastonia eutropha*, store excess carbon as granules of poly(β-hydroxybutyrate). This polyester is a potential biodegradable, renewable replacement for polypropylene. However, this natural polyester is crystalline and too brittle for most applications. A solution is to feed propionic acid along with the normal carbon source [31]. The acetyl-CoA intermediate forms from the normal carbon source, while propionyl-CoA forms from the propionic acid. Both enter the polyester synthesis pathway resulting in

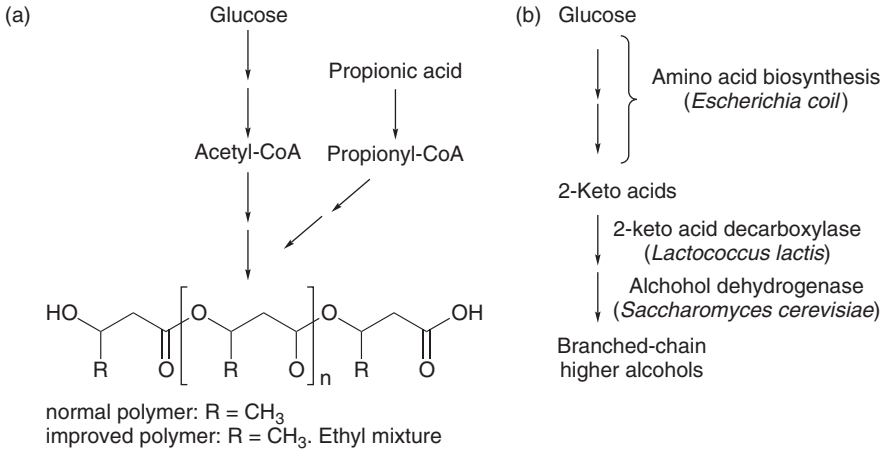

Figure 1.3 Two examples of strategies to engineer new metabolic pathways that create new products. (a) Adding propionic acid to the bacteria (*Rastonia eutropha*) making poly(β-hydroxybutyrate) leads to a mixed copolymer of β-hydroxyvalerate and β-hydroxybutyrate. This polymer is more flexible and useful than the natural poly(β-hydroxybutyrate). (b) Adding two new enzymes to *E. coli* diverts the 2-keto acids from amino acid biosynthesis to the synthesis of alcohols. The 2-keto acids are decarboxylated to the aldehyde and then reduced to the alcohol. The alcohols with five to eight carbons are potential replacements for ethanol as a biofuel

a copolymer of 4-carbon (β-hydroxybutyrate) and 5-carbon (β-hydroxyvalerate) subunits. This mixed polymer is less crystalline, more flexible, and more useful for application. This unnatural polymer does not form naturally because acyl-CoA intermediates with an odd number of carbon atoms are rare in natural biosynthetic pathways.

Another way to get new products from a metabolic pathway is to add new steps that divert an intermediate to the newly added pathway. One example is the diversion of 2-ketoacids from amino acid biosynthesis to the synthesis of alcohols for biofuels [32], Figure 1.3b. Adding a 2-keto acid decarboxylase converts these acids to aldehydes, which are reduced to alcohols. The higher alcohols (five to eight carbon atoms) are potential second generation biofuels to replace ethanol.

1.6 OUTLOOK

Biocatalysis tools have improved dramatically in the last 10 years and continuing advances in biology indicate that there will be additional improvements. Biocatalysis will be an increasingly important strategy and will become one of the key core technologies for chemical manufacturing in the next decade. It is a green chemistry technology that works perfectly with the emerging trend of bio-based sustainable feedstocks (biorefinery). The subsequent chapters detail additional examples where biocatalysis enables greener organic syntheses.

ACKNOWLEDGMENT

The authors thank the Korean Ministry of Education, Science and Technology, for funding this work (WCU program grant R32-10213).

REFERENCES

1. Anastas PT, Warner JC. Green chemistry: theory and practice. New York: Oxford University Press; 1998.
2. Pasteur L. Mé'moire sur la fermentation de l'acide tartrique. C R Séances Acad Sci 1858;46:615–618.
3. Savile CK, Janey JM, Mundorff EC, Moore JC, Tam S, Jarvis WR, Colbeck JC, Krebber A, Fleitz FJ, Brands J, Devine PN, Huisman GW, Hughes GJ. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. Science 2010;329:305–309.
4. Hansen KB, Hsiao Y, Xu F, Rivera N, Clausen A, Kubryk M, Krska S, Rosner T, Simmons B, Balsells J, Ikemoto N, Sun Y, Spindler F, Malan C, Grabowski EJJ, Armstrong JD III. Highly efficient asymmetric synthesis of sitagliptin. J Am Chem Soc 2009;131:8798–8804.
5. Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, del Cardayre SB, Keasling JD. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 2010;463:559–562.
6. Clendennen SK, Boaz NW. Cosmetic emulsifiers, United States Patent Eastman Chemical Company. (Kingsport, TN, US 7667067). 2010.
7. Fox RJ, Davis SC, Mundorff EC, Newman LM, Gavrilovic V, Ma SK, Chung LM, Ching C, Tam S, Muley S, Grate J, Gruber J, Whitman JC, Sheldon RA, Huisman GW. Improving catalytic function by ProSAR-driven enzyme evolution. Nat Biotechnol 2007;25:338–344.
8. Madison LL, Huisman GW. Metabolic engineering of poly(3-hydroxyalkanoates): from DNA to plastic. Microbiol Mol Biol Rev 1999;63:21–53.
9. Patel RN. Tour de paclixatel: biocatalysis for semisynthesis. Annu Rev Microbiol 1998;52:361–395.
10. Chen-Goodspeed M, Sogorb MA, Wu F, Raushel FM. Enhancement, relaxation, and reversal of the stereoselectivity for phosphotriesterase by rational evolution of active site residues. Biochemistry 2001;40:1332–1339.
11. (a) Matthews BW, Nicholson H, Becktel WJ. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. Proc Natl Acad Sci USA 1987;84:6663–6667; (b) Wetzel R. Harnessing disulfide bonds using protein engineering. Trends Biochem Sci 1987;12:478–482; (c) Matsumura M, Signor G, Matthews BW. Substantial increase of protein stability by multiple disulphide bonds. Nature 1989;342:291–293; (d) Suzuki Y. The proline rule. A strategy for protein thermal stablization. Proc Jpn Acad Ser B Phys Biol Sci 1999;75:133–137.
12. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D. Kemp elimination catalysts by computational enzyme design. Nature 2008;453:190–195.

13. Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF III, Hilvert D, Houk KN, Stoddard BL, Baker D. De novo computational design of retro-aldol enzymes. Science 2008;319:1387–1391.
14. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St. Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, Houk KN, Michael FE, Baker D. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science 2010;329:309–313.
15. Chen K, Arnold FH. Tuning the activity of an enzyme for unusual environments: Sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc Natl Acad Sci USA 1993;90:5618–5622.
16. Johannes TW, Woodyer RD, Zhao H. Directed evolution of a thermostable phosphite dehydrogenase for NAD(P)H regeneration. Appl Environ Microbiol 2005; 71:5728–5734.
17. DeSantis G, Wong K, Farwell B, Chatman K, Zhu Z, Tomlinson G, Huang H, Tan X, Bibbs L, Chen P, Kretz K, Burk MJ. Creation of a productive, highly enantioselective nitrilase through Gene Site Saturation Mutagenesis (GSSM). J Am Chem Soc 2003;125:11476–11477.
18. Morley KL, Kazlauskas RJ. Improving enzyme properties: when are closer mutations better? Trends Biotechnol 2005;23:231–237.
19. Park S, Morley KL, Horsman GP, Holmquist M, Hult K, Kazlauskas RJ. Focusing mutations into the *P. fluorescens* esterase binding site increases enantioselectivity more effectively than distant mutations. Chem Biol 2005;12:45–52.
20. Neylon C. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. Nucleic Acids Res 2004;32:1448–1459.
21. Wong TS, Roccatano D, Zacharias M, Schwaneberg U. A statistical analysis of random mutagenesis methods used for directed protein evolution. J Mol Biol 2006;355:858–871.
22. Fraaije MW, Wu J, Heuts DPHM, van Hellemond EW, Lutje Spelberg JH, Janssen DB. Discovery of a thermostable Baeyer–Villiger monooxygenase by genome mining. Appl Microbiol Biotechnol 2005;66:393–400.
23. Bae E, Phillips GN Jr. Identifying and engineering ion pairs in adenylate kinases: insights from molecular dynamics simulations of thermophilic and mesophilic homologues. J Biol Chem 2005;280:30943–30948.
24. (a) Lehmann M, Loch C, Middendorf A, Studer D, Lassen SF, Pasamontes L, van Loon APGM, Wyss M. The consensus concept for thermostability engineering of proteins: further proof of concept. Protein Eng 2002;15:403–411; (b) Lehmann M, Wyss M. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Curr Opin Biotechnol 2001;12:371–375.
25. Yoshikuni Y, Ferrin TE, Keasling JD. Designed divergent evolution of enzyme function. Nature 2006;440:1078–1082.
26. Heinzelman P, Snow CD, Wu I, Nguyen C, Villalobos A, Govindarajan S, Minshull J, Arnold FH. A family of thermostable fungal cellulases created by structure-guided recombination. Proc Natl Acad Sci USA 2009;106:5610–5615.
27. Schmidt M, Hasenpusch D, Kähler M, Kirchner U, Wiggenhorn K, Langel W, Bornscheuer UT. Directed evolution of an esterase from Pseudomonas fluorescens yields a mutant with excellent enantioselectivity and activity for the kinetic resolution of a chiral building block. ChemBioChem 2006;7:805–809.

28. Prather KLJ, Martin CH. De novo biosynthetic pathways: rational design of microbial chemical factories. Curr Opin Biotechnol 2008;19:468–474.
29. Nakamura CE, Whited GM. Metabolic engineering for the microbial production of 1,3-propanediol. Curr Opin Biotechnol 2003;14:454–459.
30. Ran N, Frost JW. Directed evolution of 2-keto-3-deoxy-6-phosphogalactonate aldolase to replace 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthase. J Am Chem Soc 2007;129:6130–6139.
31. Rehm BHA. Bacterial polymers: biosynthesis, modifications and applications. Nat Rev Microbiol 2010;8:578–592.
32. Atsumi S, Hanai T, Liao JC. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 2008;451:86–89.