REVIEW

doi:10.1038/nature11117

# Engineering the third wave of biocatalysis

U. T. Bornscheuer[1], G. W. Huisman[2], R. J. Kazlauskas[3,4], S. Lutz[5], J. C. Moore[6] & K. Robins[7]

**Over the past ten years, scientific and technological advances have established biocatalysis as a practical and environmentally friendly alternative to traditional metallo- and organocatalysis in chemical synthesis, both in the laboratory and on an industrial scale. Key advances in DNA sequencing and gene synthesis are at the base of tremendous progress in tailoring biocatalysts by protein engineering and design, and the ability to reorganize enzymes into new biosynthetic pathways. To highlight these achievements, here we discuss applications of protein-engineered biocatalysts ranging from commodity chemicals to advanced pharmaceutical intermediates that use enzyme catalysis as a key step.**

Biocatalysis is the application of enzymes and microbes in synthetic chemistry, and uses nature's catalysts for new purposes: applications for which enzymes have not evolved[1–5]. The field of biocatalysis has reached its present industrially proven level through several waves of technological research and innovations.

During the first wave of biocatalysis (Fig. 1), which started more than a century ago, scientists recognized that components of living cells could be applied to useful chemical transformations (in contrast to the fermentation processes, which had been commonplace for millennia already). For example, Rosenthaler synthesized (*R*)-mandelonitrile from benzaldehyde and hydrogen cyanide using a plant extract[6]; hydroxylation of steroids[7] occurring within microbial cells was also known. More recent examples are the use of proteases in laundry detergents[8], glucose isomerase to convert glucose to the sweeter-tasting fructose[9], and penicillin G acylase to make semisynthetic antibiotics[10]. The main challenge for these applications is the limited stability of the biocatalyst, and such shortcomings were primarily overcome by immobilization of the enzyme, which also facilitated the reuse of the enzyme.

During the second wave of biocatalysis, in the 1980s and 1990s, initial protein engineering technologies, typically structure based, extended the substrate range of enzymes to allow the synthesis of unusual synthetic intermediates. This change expanded biocatalysis to the manufacture of pharmaceutical intermediates and fine chemicals. Examples include the lipase-catalysed resolution of chiral precursors for synthesis of diltiazem (a blood pressure drug), hydroxynitrile-lyase-catalysed synthesis of intermediates for herbicides[11], carbonyl-reductase-catalysed synthesis of enantiopure alcohols for cholesterol-lowering statin drugs, lipase-catalysed synthesis of wax esters such as myristyl myristate or cetyl ricinoleate for cosmetics[12], and nitrile-hydratase-catalysed hydration of acrylonitrile to acrylamide for polymers[13] (where nitrile hydratase was obtained from whole cells of *Rhodococcus rhodochrous*). Apart from stabilization, the challenges now included optimizing the biocatalyst for the non-natural substrates.

The third, and present, wave of biocatalysis started with the work of Pim Stemmer and Frances Arnold in the mid and late 1990s. They pioneered molecular biology methods that rapidly and extensively

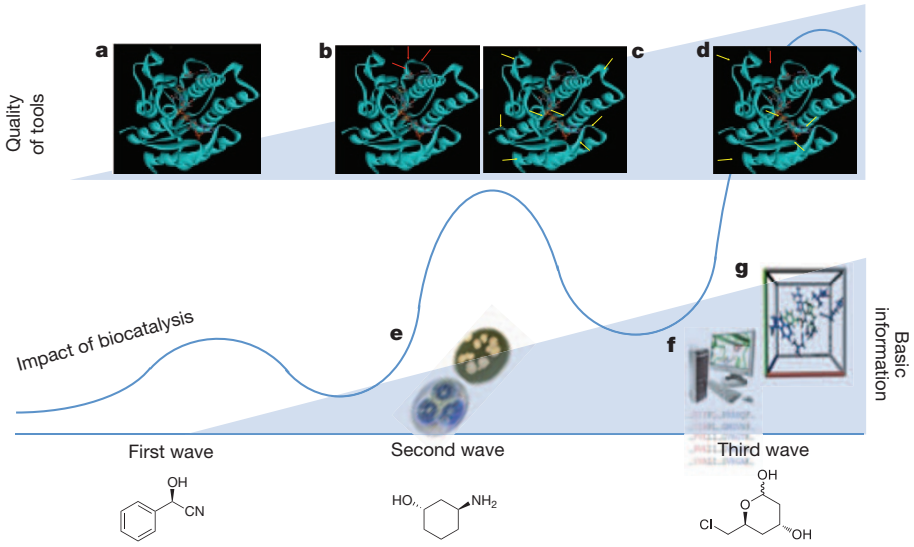



**Figure 1 | The evolution of enzyme discovery and protein engineering strategies used to identify desired catalysts.** Rational design (**b**) identifies distinct point mutations based on protein structures (**a**) or homology models, whereas random mutagenesis (**c**) combined with screening or selection is the basis for directed evolution experiments. Combining these methods makes it possible to create smaller, but smarter, libraries (**d**). The classical screening of enzymes by enrichment cultures (**e**) is now replaced by key motif database searches (**f**) to guide identification of novel enzymes or those with desired properties. Still in its infancy is the computational *ab initio* (or *de novo*) design of enzymes (**g**). The structures at the bottom refer to fine chemicals accessible through the different waves of biocatalysis. (*R*)-mandelonitrile (left) could already be obtained 100 yr ago from a plant extract; (1*S*,3*S*)-3-aminocyclohexanol (centre) is made by Novartis using an immobilized lipase; and 6-chloro-2,4,6-trideoxy-D-erythrohexapyranoside (right) is made by DSM in a process that requires an engineered aldolase to withstand high concentrations of acetaldehyde and to achieve high selectivity.

[1]Institute of Biochemistry, Department of Biotechnology & Enzyme Catalysis, Greifswald University, Felix-Hausdorff-Straße 4, D-17487 Greifswald, Germany. [2]Codexis Inc., 200 Penobscot Drive, Redwood City, California 94063, USA. [3]Department of Biochemistry, Molecular Biology and Biophysics, Biotechnology Institute, University of Minnesota, 1479 Gortner Avenue, Saint Paul, Minnesota 55108, USA. [4]Department of Chemical and Biological Engineering, Seoul National University, Seoul 151-744, Korea. [5]Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, Georgia 30322, USA. [6]Merck Research Laboratories, Merck & Co., Rahway, New Jersey 07065, USA. [7]Lonza AG, Valais Works, CH-3930 Visp, Switzerland.

modify biocatalysts via an *in vitro* version of Darwinian evolution. The methods are now commonly called directed evolution, although this term has been in use since whole-cell experiments in 1972[14]. The initial versions of this technology involve iterative cycles of random amino-acid changes in a protein, followed by selection or screening of the resulting libraries for variants with improved enzyme stability, substrate specificity and enantioselectivity. Subsequent developments, discussed here, have focused on improving the efficiency of directed evolution to create 'smarter' libraries. Industrial-scale biocatalysis focused primarily on hydrolases, a few ketoreductases (KREDs), and cofactor regeneration and protein stability in organic solvents. In some cases, metabolic pathways were optimized; for example, combining genes from various natural strains to produce 1,3-propanediol (a monomer for polyesters) in a new host made it possible to switch from glycerol to the more convenient glucose as the feedstock[15].

As a result of the advances made during the present wave of biocatalysis, remarkable new capabilities can now be engineered into enzymes, such as the ability to accept previously inert substrates (a KRED for montelukast[16] or a transaminase for sitagliptin[17,18]) or to change the nature of the product that is formed (terpene cyclase variants that favour different terpenes[19] or amino-acid metabolism that makes alcohols for biofuels[20]). Novel enzymes are needed today to convert biomass into second- and third-generation biofuels[21,22], materials[23] and chemicals[24]. Key developments that enabled this third wave are advanced protein engineering[25–27] (including directed evolution), gene synthesis, sequence analysis, bioinformatics tools[28] and computer modelling, and the conceptual advance that improvements in enzymes can be more pronounced than previously expected. Engineered enzymes can remain stable at 60 °C in solutions containing organic solvents, can accept new substrates and can catalyse new non-natural reactions. This engineering may now take only a few months, thus greatly expanding the potential applications. In the past, an enzyme-based process was designed around the limitations of the enzyme; today, the enzyme is engineered to fit the process specifications.

About ten years ago, articles in *Nature*[29,30] and *Science*[31] reviewed the first and second waves of biocatalysis and provided a glimpse at what the third wave might bring. Today it is timely to assess the impact of this third wave and to speculate what the next decade might bring (Box 1). Although biocatalysis involves metabolic engineering[22,32,33] and synthetic biology, this Review focuses on enzymatic and whole-cell reactions.

## Engineering enzymes to fit the manufacturing process

To minimize costs, chemical manufacture requires stable, selective and productive catalysts that operate under the desired process conditions. Engineering enzymes for such a process starts by defining the engineering goal, such as increased stability, selectivity, substrate range or, typically, a combination thereof. In 2000, before the third wave, only a few strategies were available to meet these goals. Enzyme immobilization could increase the stability of a protein, but the increases in stability were moderate and often insufficient for most chemical transformations. Directed evolution was also possible, but was still slow because it required construction and screening of large libraries that mostly contained variants with reduced, or even no, activity. Examples of drastic improvements were rarely of industrial relevance. The slow pace meant that the evolved proteins contained only a few changes and, thus, that the enzyme properties changed only slightly. Although several hundred enzymatic processes already had industrial uses[4], most involved enzymes and whole cells that had been marginally altered genetically[29].

In the past decade, our understanding of proteins and the number of available directed evolution strategies have both increased, making it possible to make large changes in enzyme properties. By and large, enzyme engineering continues to be a collection of case studies resulting from applying one of various possible approaches to the problem at hand, rather than there being a quantitative approach such as those used in disciplines such as civil, electric, software, or chemical engineering. Converting these case studies into engineering principles will require using free energy to connect the design goals to the structural changes needed (Fig. 2). Large changes in properties require large changes in free energy. For example, large changes in stability will require large free-energy changes in the folding–unfolding equilibrium. (Even irreversible protein unfolding starts with a reversible partial unfolding.) A molecular-level understanding of proteins suggests strategies that could be used for the improvements. For example, surface residues contribute to the folding–unfolding equilibrium and adding a proline residue in a loop lowers the entropy of the unfolded form. These strategies replace large libraries of random variants (mostly with poorer properties) with smaller, more focused protein libraries containing a high fraction of active and potentially improved variants (Fig. 1). Finally, by estimating the strength of various interactions (ion pairs on the surface or entropic contributions of adding a proline residue), researchers can estimate the changes needed to reach the goal. Few researchers explicitly use the free-energy-based measures to plan protein-engineering strategies today, but converting case studies into engineering principles requires a quantitative approach.

**BOX 1**

### Requirements and examples of biocatalysis applications

- In traditional biocatalysis, natural products are converted into other natural products using natural reactions and pathways. Technical requirements: maintain microorganism cultures; conceptual requirements: possible to control natural biotransformations; examples: bread and cheese making, leather processing, beer and wine fermentation, and natural antibiotic production.
- In broad-substrate-range biocatalysis, chemical intermediates (non-natural products) are converted into other chemical intermediates using natural reactions and pathways. Technical requirements: use of defined enzymes (no interfering activity present); conceptual requirements: many enzymes have a broad substrate range; examples: manufacture of pharmaceutical intermediates using lipases and carbonyl reductases (alcohol dehydrogenases).
- In multistep biocatalysis, natural products are converted into fuels, materials and chemical feedstocks (non-natural products) using non-natural reactions and pathways. Technical requirements: protein engineering for major changes in stability, substrate range and type of reaction catalysed; conceptual requirements: enzymes can catalyse non-natural reactions and new combinations of enzymes create new pathways; examples: fuel molecules created using the isoprene biosynthesis pathways, amino-acid biosynthesis diverted to fuel alcohols.

## New and improved methodologies

Over the past ten years, major advances in DNA technologies and in bioinformatics have provided critical support to the field of biocatalysis. These tools have promoted the discovery of novel enzymes in natural resources and have substantially accelerated the redesign of existing biocatalysts.

### Advanced DNA technologies

Next-generation DNA sequencing technology has allowed parallel sequence analysis on a massive scale and at dramatically reduced cost. Whereas the cost of a human genome sequence analysis in 2002 was estimated at ~US$70,000,000, the price in 2012 has decreased more than 1,000-fold to less than US$10,000 (ref. 34), and LifeTechnologies, Illumina and Oxford Nanopore Technologies have announced that sequencing machines that are designed to sequence the entire human genome in a matter of hours will be available later in 2012 and will lower the cost per genome to less than US$1,000. Sequences of entire genomes from organisms from different environments, as well as environmental

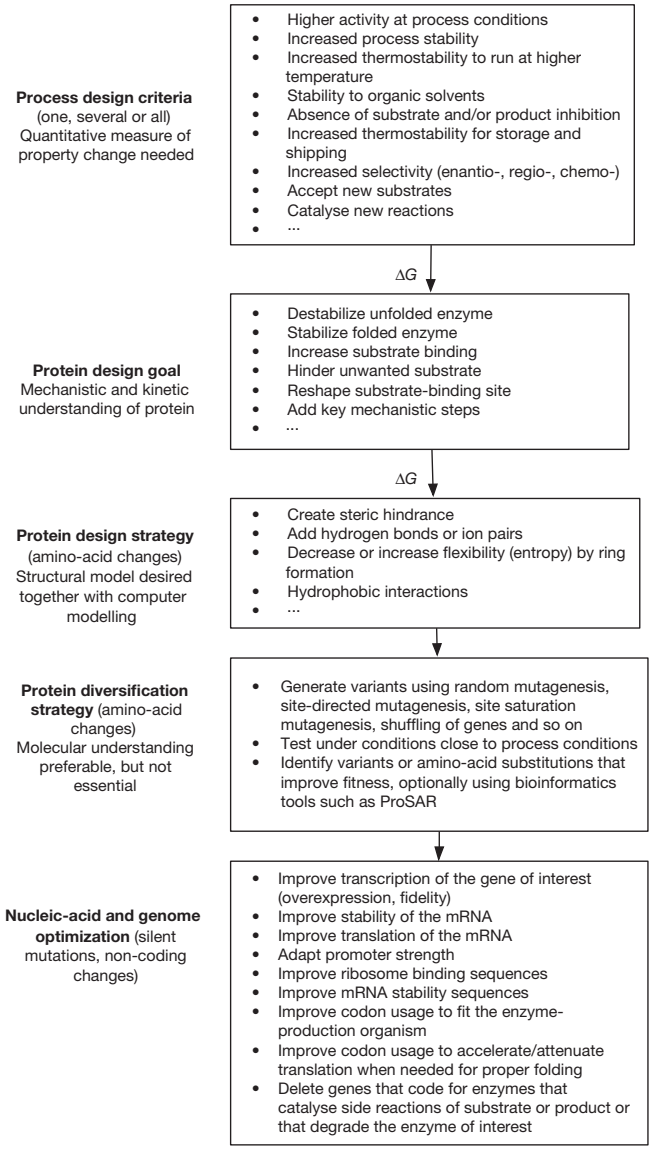



**Figure 2 | Free energy (Δ*G*) connects design goals to the required structural changes via protein engineering strategies.** This reasoning allows the use of more focused libraries. If the design goal lies far from the starting enzyme, then large changes in free energy are needed. Mechanistic and kinetic understanding of the protein unfolding and reaction mechanism identifies an engineering strategy to reach the design goal. Finally, analysis of the structure (which varies from qualitative inspection to extensive computer modelling) can identify the region or amino acids that must be changed. Goals that require large free-energy changes will likewise require more extensive changes in structure. mRNA, messenger RNA.

DNA samples that include unculturable organisms (metagenomes), have created a rich resource in which to search for novel biocatalysts[35], and will continue to do so. Massive high-throughput sequencing (>10,000,000 sequence reads) using the Illumina technology also facilitated the exploration and understanding of protein sequence–function relationships[36].

Low-cost DNA synthesis has replaced isolation of genomic DNA as the starting point for protein engineering. Whole-gene DNA synthesis further allows the codons to be optimized for the host organism and molecular architectural structures such as promoters, terminators, enhancers, restriction sites and so on to be introduced at convenient sites. This DNA synthesis uses traditional phosphoramidite chemistry, but optimized reaction conditions have improved coupling efficiency, increasing the overall quality and quantity of the polymer to make sequences even 200–250 nucleotides long. Parallel DNA synthesis using photolithographic and inkjet printing techniques further cut costs and speed synthesis[37]. DNA synthesis has been used to make entire sections of chromosomal DNA and even complete genomes for metabolic pathway engineering[38]. Whole-gene synthesis can also be used to make high-quality DNA libraries ranging from small, focused, site-saturation libraries to large, comprehensive gene collections. Customized genes and even gene libraries are becoming commodity chemicals similar to reagents and solvents found in today's research laboratories.

### Novel tools in bioinformatics

Complementing the experimental advances, bioinformatics tools have become an integral part of modern protein engineering[39]. Multiple sequence alignments across large enzyme families and homology searches have identified genes with similar catalytic activities, leading to novel, potent biocatalysts[40]. The same sequence information allows the reconstruction of ancestral biocatalysts[40], which may have broader substrate range and catalytic promiscuity (see below). Multiple sequence alignments identify the most common amino acids at each position (the consensus sequence) and amino-acid substitutions that yield stable function enzymes. This data helps in the design of small libraries with a high proportion of catalytically active variants. These libraries have been used to discover biocatalysts with enhanced stabilities, catalytic functions and altered stereoselectivities[41].

Paralleling the advances in sequence-based protein engineering, structure-guided approaches have benefited from a rapid increase in protein structure coordinates deposited in the RCSB Protein Data Bank (http://www.pdb.org). Over the past decade, the repository has grown by over 450% to contain more than 77,000 protein structures. This facilitates both rational protein design and directed evolution, because structural alignment of related proteins helps to identify distinct similarities and differences guiding the more reliable design of mutant libraries.

The utility of smaller libraries was demonstrated in two different approaches to increasing the enantioselectivity of an esterase for resolution of methyl 3-bromo-2-methylpropionic acid, a chiral synthon[42]. Using random mutagenesis and screening 200 out of thousands of variants, the *E*-value for the enzyme (the selectivity of the enzyme for one enantiomer over the other) was increased from 12 to 19 (ref. 43). Recognizing that a relatively small increase (0.5 kcal $mol^{-1}$) in the difference in activation energy ($\Delta\Delta\Delta G^{\ddagger}$) for the two enantiomers was needed to generate a practical enzyme ($E > 30$), mutagenesis was focused at the active site. A library containing all possible single mutations at four positions (76 variants) yielded an enzyme with $E = 61$ ($\Delta\Delta\Delta G^{\ddagger} = 0.96$). Understanding the nature of the problem to be solved focuses the enzyme optimization approach on smaller libraries and gives bigger improvements. In this context, it is worth also mentioning a new method for continuous directed evolution using a combination of a phage infection system and a mutator plasmid in *Escherichia coli*[44].

## Examples of engineered enzymes in industrial biocatalysis

As predicted by Schmid *et al.* in their forward-looking review in 2001[29], continuous regeneration of cofactors and a wider range of enzymes have been reported in the past ten years. However, the predicted applications of biochips and combinatorial biocatalysis have not yet materialized. The use of non-metabolizing cells for biocatalysis has proven to be more difficult than predicted and preference has instead shifted towards engineered enzymes used in crude and semipurified form. Whereas historically whole cells offered a simple and effective option for cofactor regeneration and enhanced enzyme stability, protein engineering and the use of single enzymes is now considered more economic and practical. The use of isolated enzymes have other advantages: they are easier to remove (less is added because they have more activity per unit mass), they tolerate harsher conditions, they eliminate potential diffusion limitations caused by cell membranes and they are easier to ship around the world. For example, KRED-based processes have now replaced whole-cell reductions and metal–ligand-based chemocatalysis, which were the industry standards during the past decade[45,46]. One exception is a whole-cell process to convert racemic hydantoins into

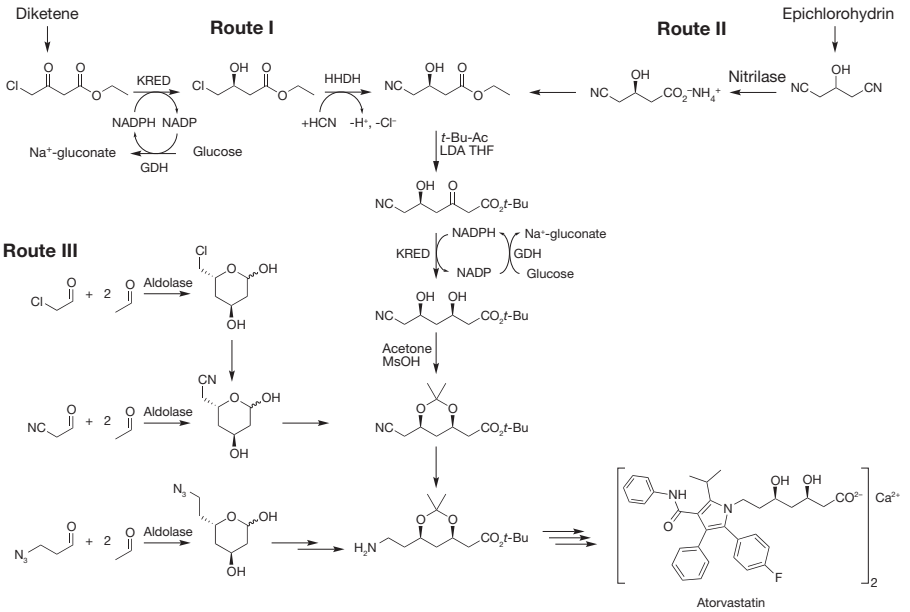



**Figure 3 | Different enzymatic routes to the synthesis of the key side chain of atorvastatin (Lipitor).** These processes use the combination of KRED with a halohydrin dehalogenase (HHDH) (route I), a nitrilase (route II) or an aldolase (route III). They differ not only in class of enzyme, but also in choice of (inexpensive) starting material, activity and selectivity of the biocatalyst, downstream processing, and yield and purity of final product. Routes I and II create one stereocentre, but route III creates both stereocentres required for the advanced pharmaceutical intermediate[52]. Introduction of the second chiral centre following routes I and II is also accomplished using KRED. LDA, lithium diisopropylamide; *t*-Bu, *tert*-butyl; THF, tetrahydrofuran.

optically pure, non-natural α-amino acids. Recombinant *E. coli* concurrently expressing hydantoinase, carbamoylase and racemase was found to be a simple and efficient production system, replacing the original process based on three immobilized enzymes in consecutive fixed-bed reactors[47,48]. Moreover, the whole-cell process required no metabolic flux controls and proceeded without undesired side reactions.

KREDs and other enzymes have been widely investigated for the manufacture of chiral intermediates for pharmaceuticals such as atorvastatin, the active ingredient in Lipitor, which is a cholesterol-lowering drug that had global sales of US$11,900,000,000 in 2010. Seven enzymatic approaches[2,49,50] (Fig. 3), differing not only in the choice of enzyme and starting material but also as to whether the product is a raw material (with a single chiral centre) or an advanced intermediate (with two chiral centres), have been developed. In all cases, success requires protein engineering to improve the reaction rate, the enantioselectivity, the stability to high substrate concentrations (up to 3 M, as in the nitrilase process[51]) or the stability to high solvent concentrations (20% butylacetate in the KRED process[52]). Apart from a highly active biocatalyst, a low-cost process also requires inexpensive raw materials and simple isolation of pure product in high yield. One current process leading to the advanced intermediate uses three biocatalytic steps: first, the combination of KRED and glucose dehydrogenase; second, the combination of this with a halohydrin dehalogenase to make the ethyl (*R*)-4-cyano-3-hydroxybutanoate intermediate (Fig. 3) at a rate of $>100\,t\,yr^{-1}$; and, third, the enzymatic reduction for the advanced diol intermediate[52].

Recent engineering[17] expanded the substrate range of transaminases to ketones with two bulky substituents. The enzyme engineering started with a small ketone substrate, created more space in the active site and used increasingly larger ketones. Several rounds of directed evolution increased the activity ~40,000-fold and yielded an engineered amine transaminase (Fig. 4) that can replace the transition-metal-based hydrogenation catalyst for sitagliptin manufacture. Starting from ATA-117, a close homologue of the wild-type enzyme, which had no detectable activity on the substrate, the first variant provided very low activity (0.2% conversion of $2\,g\,l^{-1}$ substrate using $10\,g\,l^{-1}$ enzyme) towards prositagliptin; the final variant converts $200\,g\,l^{-1}$ ketone to sitagliptin with 99.95% e.e. at 92% yield. The biocatalytic process not only reduced the total waste and eliminated all transition metals, but increased the overall yield and the productivity by 53% by comparison with the metal-catalysed process[18]. The numerous biocatalytic routes scaled up for pharmaceutical manufacturing (Table 1) demonstrate their competitiveness with traditional chemical processes.

Enzyme variants resulting from optimization studies are a unique source of starting points for future programmes, and by using the more stable enzymes engineered for one process the next optimization programme can be even faster. For instance, engineering KREDs to make R3HT (**3**) (see Table 1 for abbreviations and numbering of compounds) created many stable enzyme variants including some that were unsuitable due to low enantioselectivity. However, one of these unsuitable variants was the starting enzyme in engineering a KRED for DCFPE (**4**). One of the DCFPE enzymes was then the starting point for a montelukast (**5**) KRED, which in turn was a starting point for the duloxetine (**6**) KRED. Similarly, the transaminases generated during the evolution of the prositagliptin (**18**) transaminase can make other amines and may serve as starting points for new engineered enzymes for amine synthesis. Starting from a non-natural stabilized enzyme variant that already works in one process thus accelerates catalyst and process development in unprecedented ways.

Enzymatic conversions that simultaneously set two stereocentres are especially efficient ways to make complex molecules. For example,

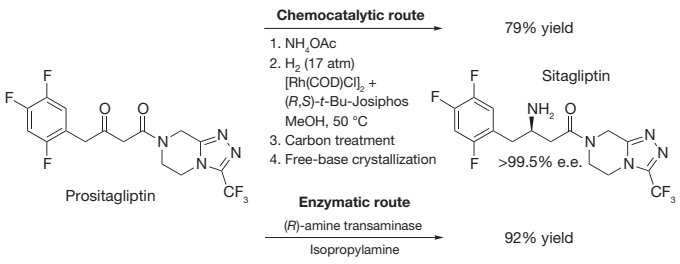



**Figure 4 | Biocatalysis advances synthetic chemistry.** The enzymatic route for the synthesis of sitagliptin using an engineered amine transaminase is superior to the chemical hydrogenation process, resulting in a higher yield of >99.5% e.e. optically pure product, higher productivity, reduced total waste and elimination of a transition-metal catalyst[17,18]. atm, atmospheres; e.e., enantiomeric excess; Me, methyl.

reduction of a ketone catalysed by a KRED sets the alcohol stereocentre. However, if a second stereocentre next to the ketone carbonyl racemizes rapidly in solution and the KRED is highly selective for one configuration, then the reduction reaction can set two stereocentres in one step. Examples include a penem intermediate (**8**), pseudoephedrine (**12**) and phenylephrine (**13**), as well as similar processes for chiral amines (**20**). Also, aldolases are now being used for subsequent reaction steps to generate multiple chiral centres, for example in the synthesis of statin intermediates (**33**). The single-enzyme cascade reactions as catalysed by aldolases have been further expanded to multi-enzyme cascade processes, for instance in the synthesis of 2′-deoxyinosine (**34**) *in vitro* or of complex molecules such as artemisinin (**42**) or Taxol *in vivo*.

## Environmental advantages of biocatalytic processes

In the context of concerns about the environmental aspects of chemical manufacturing, biocatalysis provides an attractive alternative. The US Environmental Protection Agency awards five prizes each year in the Presidential Green Chemistry Award Challenge. The nominations emphasize the 12 Principles of Green Chemistry[53], which consider environmental factors as well as use of renewable feedstocks, energy efficiency and worker safety. Biocatalysis, using either enzyme technology or whole cells, has won 16 awards since 2000 (Table 2). Biocatalysts are made from renewable sources and are biodegradable and non-toxic, and their high selectivities simplify reaction work-ups and provide product in higher yields. Biocatalytic processes are also safe as they typically run at ambient temperature, atmospheric pressure and neutral pH. Hence, it is not surprising that so many of the awards go to biocatalysis.

The broad range of awards in Table 2 shows the application of biocatalysts beyond the pharmaceutical industry. Several applications involve polymers, especially polyesters. The optimized fermentation of lactic acid is the basis for a polylactic acid plant in Nebraska with a capacity of 141,000 tonnes per year, and a new non-natural metabolic route allows synthesis of 1,3-propanediol for the manufacture of SORONA polymer. The 1,4-butanediol fermentation yields a component of another polymer, spandex. In the synthesis of polyhydroxyalkanoates, the biocatalyst catalyses not just the monomer synthesis but also its polymerization. Yang and co-workers recently engineered these polyhydroxyalkanoate synthases to polymerize lactic acid into polylactic acid[54]. Several other awards involve new metabolic pathways to manufacture biofuels. For example, LS9, Inc. engineered *E. coli* bacteria to produce biodiesel. Adding the genes for plant thioesterases to *E. coli* diverted normal fatty-acid biosynthesis into synthesis of several fatty acids suitable for biodiesel. Then genes were added 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[22]. The amount of biodiesel produced is at least tenfold too low for the process to be commercially viable, but further engineering will probably increase the yield.

## New concepts in protein engineering

Large changes in enzyme properties usually require multiple amino-acid substitutions because they make larger changes to the protein structure. However, simultaneous multiple amino-acid substitutions create exponentially more variants for testing. There are 7,183,900 possibilities for two substitutions anywhere in a 200 amino-acid protein and 9,008,610,600 possibilities for three substitutions. Many of these variants are inactive, and either all are created and tested to find the improved variants, or the library is screened only partially and incremental improvements in subsequent rounds of evolution are required.

The simplest solution to this problem is more efficient screening. Changes in substrate specificity may be monitored by high-throughput methods, such as fluorescence-activated cell sorting[55–57], which can screen tens of millions of variants in a short amount of time. Whittle and Shanklin made six simultaneous substitutions in the active site of a desaturase and then screened for growth on a different substrate. Only those variants with altered substrate specificities could grow[58]. Seelig and Szostak used very large random libraries (up to $10^{13}$ variants), from which they could select variants that catalysed an RNA ligation[59] based on binding of the product, but not the starting materials, to a column.

At present, the best approach to creating multiple mutations is to add them simultaneously but to limit the choices using statistical or bioinformatic methods. One statistical correlation approach is based on the ProSAR (protein structure activity relationship) algorithm used by Codexis researchers to improve the reaction rate of a halohydrin dehalogenase >4,000-fold[60]. Researchers made random amino-acid substitutions (an average of ten) in the dehalogenase and measured the rate of catalysis by the variants. Then statistical methods identified whether a particular substitution was beneficial. For example, variants that contained a Phe 186 Tyr substitution were, on average, better than those that did not. Some variants that contained such a substitution were not beneficial, owing to the detrimental effects of other mutations, but the statistical analysis identified that, on average, Phe 186 Tyr is a beneficial mutation. The final improved enzyme contained 35 amino-acid substitutions among its 254 amino acids.

γ-Humulene synthase catalyses the cyclization of farnesyl diphosphate via cationic intermediates to γ-humulene in 45% yield, but forms 51 other sesquiterpenes in smaller amounts. Keasling and co-workers substituted amino-acid residues in the active site stepwise and identified the contribution of each one to the product distribution[61]. Substitutions were combined to favour formation of one of the other sesquiterpenes. For example, one triple substitution created an enzyme that formed 78% sibirene; the natural enzyme forms 23% sibirene.

Another approach is to limit the location of changes to the active site and the type of changes to those known from sequence comparisons to occur often at these sites. Jochens and Bornscheuer used this approach to increase the enantioselectivity of a *Pseudomonas fluorescens* esterase. There were 160,000 ($20^4$) ways simultaneously to vary the four amino acids adjacent to the substrate in the active site. The researchers aligned the amino-acid sequences of >2,800 related enzymes to identify which amino acids are most common at these locations. This analysis limited the possibilities to several hundred variants, which were tested to find a double and a triple mutant with the desired selectivities[41]. Another important advance that allows multiple mutations is the recognition that mutations often destabilize proteins[62–64] and that starting with a very stable protein therefore allows it to tolerate a greater number and range of changes[65,66].

Because the workload for screening larger libraries containing multiple mutations increases exponentially with library size, most researchers work on the assumption that beneficial mutations are mostly additive[67] and that synergistic effects are rare, except for nearby changes. Indeed, combining beneficial single mutations often yields additive improvements (for example increasing the stability of an esterase from *Bacillus subtilis* to an organic solvent[68] or increasing the enantioselectivity of a lipase from *Pseudomonas aeruginosa*[69]). However, often the contributions do not add up exactly or have unexpected behaviour. For example, mutations A and B by themselves may be deleterious, but together they may be beneficial. Weinreich and co-workers[70] investigated such cooperative interactions in the evolution of a β-lactamase with higher activity. Mutation A increased the reaction rate but destabilized the β-lactamase. The overall effect was slightly beneficial. Mutation B did not affect rate but stabilized the β-lactamase; by itself it had no effect. Together mutations A and B were highly beneficial because the β-lactamase was faster and maintained its stability, but adding mutations stepwise will most likely miss these types of synergy. Reetz and Sanchis came to similar conclusions in testing the stepwise addition of mutations to increase enantioselectivity[71]. Synergistic effects are important when one of the mutations is a stabilizing mutation and when the mutations are nearby one another. The complication of non-additivity due to stabilizing mutations can be minimized by stabilizing the protein before starting mutagenesis, but the complication of non-additivity due to nearby mutations is the most common one and not easily avoided.

Consequently, the extent of useful changes made during the improvement of a protein has increased drastically in the past decade. In the early 2000s, 1–5 mutations were typical, whereas by 2010, 30–40 amino-acid

**Table 1 | Recently developed biocatalytic processes in the pharmaceutical industry**

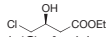
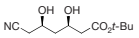
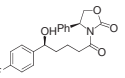
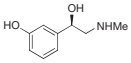
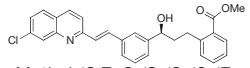
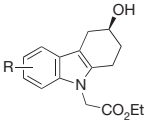
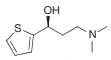
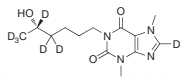
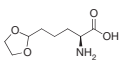
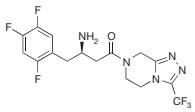
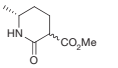
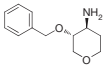
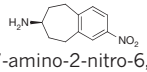
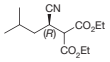
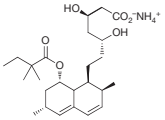
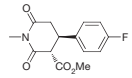
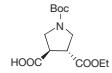
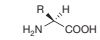
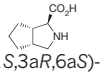
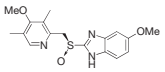
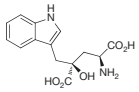
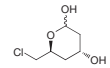
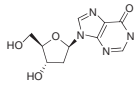
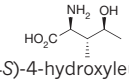
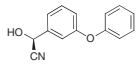
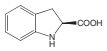
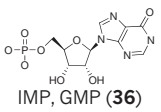
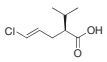
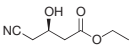
| Product | Technology (company, reference) | Product | Technology (company, reference) |
|---|---|---|---|
| KREDs | | | |
|  Ethyl (*S*)-4-chloro-3-hydroxybutanoate (**1**) | Various protein engineering approaches, including directed evolution, to increase activity, stability and coenzyme specificity. Initial intermediate for atorvastatin (Lipitor) manufacture on the industrial scale. (Kaneka[81], Codexis[52]) |  *Tert*-butyl (2*S*,3*R*)-4-chloro-3-hydroxy-1-phenylbutan-2-yl-carbamate (**10**) | Recombinant expression and also protein engineering for activity. Intermediate for atazanavir (Reyetaz). (Bristol Myers Squibb[82]; Codexis, WO/2011/005527) |
|  *Tert*-butyl 6-cyano-3,5-dihydroxyhexanoate (**2**) | Protein engineering for activity and stability to 30% organic substrate (the hydroxyketone substrate and a *tert*-butyl acetoacetate impurity, both liquids). Intermediate for atorvastatin. (Codexis, US/2011/7879585; Pfizer, US/2008/0118962) |  (4*S*)-3[(5*S*)-5-(4-fluorophenyl)-5-hydroxy-pentanoyl]-4-phenyl-1,3-oxazolidin-2-one (**11**) | Protein engineering for activity and solvent stability. Intermediate for ezetimibe (Zetia, Vytorin). (Codexis, WO/2010/025085) |
|  (*R*)-3-hydroxythiolane (R3HT) (**3**) | Directed evolution primarily for enantioselectivity. Intermediate for sulopenem-type antibiotics (since discontinued). (Codexis[83], WO/2009/029554) |  D-pseudoephedrine (**12**) | Protein engineering to overcome product inhibition. The ketone substrate is racemic at $C_2$ and racemizes during the reaction to give high-e.e. product. The product of the reaction is a generic API. (Daiichi[84]) |
|  (*S*)-1-(2,6-dichloro-3-fluorophenyl)-ethanol (DCFPE) (**4**) | Directed evolution of a variant identified from the R3HT (**3**) project for increased activity. Note that the ketone is hindered owing to the ortho chloro-substituents. Raw material for crizotinib (Xalkori). (Codexis, WO/2009/036404) |  (*R*)-phenylephrine (**13**) | Protein engineering for activity and tolerance to aminoketone substrate. The product of the reaction is a generic API. (Codexis, WO/2011/022548) |
|  Methyl (*S*,*E*)-2-(3-(3-(2-(7-chloroquinolin-2-yl)vinyl)phenyl)-3-hydroxypropyl) benzoate (**5**) | Directed evolution of a variant identified from the DCFPE (**4**) project. Protein engineering for activity on the poorly soluble substrate (~10 mg $l^{-1}$ in 40% toluene, 10% IPA). Intermediate for montelukast (Singulair). (Codexis[16], WO/2009/042984) |  Ethyl-((7*S*)-7-hydroxy-6,7,8,9-tetrahydro-pyrido[1,2-a]indol-10-yl) acetate (**14**) | A recombinantly expressed, engineered variant was used to produce 1.6 kg of this intermediate for a CRTH2 receptor antagonist. (Merck, US/2010/0234415) |
|  (*S*)-*N*,*N*-dimethyl-3-hydroxy-3-(2-thienyl)-1-propanamine (**6**) | Protein engineering for activity at pH 9 and IPA, conditions where the ketoamine is stable. Intermediate for duloxetine (Cymbalta). (Codexis, WO/2010/025238) |  (3*R*)-4,4-dimethoxy-tetrahydro-2H-pyran-3-ol (**15**) | Recombinantly produced KRED was used for the stereoselective production of 38 kg of this intermediate. (Merck, US/2008/0138866) |
|  (*R*)-2-methylpentanol (**7**) | Protein engineering for activity on the aldehyde substrate as well as enantiospecificity towards the desired enantiomer. Intermediate for imagabalin (since discontinued). (Codexis[85], WO/2010/027710) |  *N*-Cbz-4*R*-azepanol (**16**) | A recombinantly expressed enzyme was used to produce this intermediate for a novel β-lactamase inhibitor. (Merck, WO/2008/039420) |
|  Methyl (2*S*,3*R*)-2-benzamidomethyl-3-hydroxybutyrate (for penem antibiotics) (**8**) | Protein engineering for activity on the racemic ketone as well as for enantiospecificity. The process involves a dynamic kinetic resolution at $C_2$ to give the desired product in high e.e. and d.e. (Codexis, US/2011/7883879) |  (*R*)-8-$d_1$-1-(4,4,5,5,6,6,6-$d_6$-5-hydroxyhexyl)-3,7-dimethyl-1H-purine-2,6-(3H,7H)dione (**17**) | A recombinantly expressed enzyme was used to produce this deuterium containing pentoxyfylline analogue on the gram scale. (Concert Pharmaceuticals, US/2011/0053961) |
|  (*S*)-Allysine ethylene acetal (for omapatrilat) (**9**) | L-acylase (a hydrolase), large-scale commercial manufacturing. Key intermediate for angiotensin-1-converting enzymes and neutral endopeptidase inhibitors. (Chirotech/DRL[86]) | — | — |
| Transaminases | | | |
|  (2*R*)-4-oxo-[3-(trifluoro-methyl)-5,6-dihydro[1,2,3]triazolo[4,3-a]pyrazine-7(8H)-yl]-1-(2,4,5-trifluorophenyl) butan-2-amine (**18**) | Protein engineering to develop enzyme with initial activity, followed by activity improvement in increasing DMSO, substrate and isopropylamine, as well as increased thermostability. The product is sitagliptin (Januvia) free base. (Codexis; Merck[17], WO/2010/099501) |  Methyl (6*R*)-6-methyl-2-oxopiperidine-3-carboxylate (**20**) | A codon-optimized mutant enzyme was used to manufacture 29 kg of this orexin receptor inhibitor. (Merck, WO/2009/143033) |
|  (3*R*,4*S*)-3-benzyloxy)tetrahydro-2H-pyran-4-amine (**19**) | An enzyme obtained by directed evolution was used for this intermediate for an M1-receptor-positive allosteric regulator. (Merck, WO/2011/062853) |  (7*S*)-7-amino-2-nitro-6,7.8.9-tetrahydro-5H-benzo[7] annulene (**21**) | A recombinantly expressed (codon-optimized) enzyme was used to manufacture this intermediate for an Axl inhibitor on the gram scale. (Rigel, US/2010/0196511) |

**Table 1 | Continued**

| Product | Technology (company, reference) | Product | Technology (company, reference) |
|---|---|---|---|
| **Hydrolases** | | | |
|  Intermediate for pregabalin (**22**) | A highly productive process was enabled for this pregabalin intermediate using a natural enzyme after significant reaction engineering. The process produces quantities of the order of 100 tonnes. (Pfizer, WO/2006/000904) |  Simvastatin ammonium salt (**25**) | Reaction engineering and directed evolution enabled a simple process for manufacture of this high-volume API. The enzyme was improved for reaction conditions and a non-natural acyl donor. (UCLA[87]; Codexis, WO/2011/041231) |
|  Intermediate for paroxetine (**23**) | Intermediates for paroxetine. (BioVerdant, WO/2009/005647) |  L-fluoro-leucine (**26**) | A highly productive process for this odanacatib intermediate was enabled using a natural enzyme after enzyme immobilization and reaction engineering. (Merck[88]) |
|  (1*S*,3*S*)-3-aminocyclohexanol (**24**) | Enzymatic kinetic resolution; lipase from *Thermomyces lanuginosus*; produced on the kilogram scale. (Novartis[89]) |  (3*R*,4*R*)-Boc-pyrrolidine-3,4-dicarboxylic acid 3-ethyl ester (**27**) | Cascade reaction using two commercial lipases; >100 kg produced. (Roche[90]) |
| **Oxidative enzymes** | | | |
|  (1*R*,2*S*,5*S*)-6,6-dimethyl-3-azabicyclo[3,1,0] hexane-2-carboxylic acid (**28**) | A fungal amine oxidase was improved for expression in *E. coli* and activity under process conditions, including substrate and product tolerance. The intermediate is used for boceprevir synthesis. (Codexis, WO/2010/008828) |  (*S*)-amino acids (**30**) | (*R*)-amine oxidase/(*S*)-amino acid transferase; key intermediate for antidiabetic drug. (Bristol Myers Squibb[91]) |
|  (1*S*,3a*R*,6a*S*)-octahydrocyclopenta[c]pyrrole-1-carboxylic acid (**29**) | An amine oxidase was improved for activity under process conditions, including substrate and product tolerance. The intermediate is used for telaprevir synthesis. (Codexis, WO/2010/008828) |  Esomeprazole (**31**) | The enantioselectivity of the Baeyer–Villiger monooxygenase was inverted and the enzyme was then improved for activity, stability and chemoselectivity. The product is an API (Nexium). (Codexis, WO/2011/071982) |
| **Aldolases** | | | |
|  (2*S*,4*S*)-monatin (**32**) (and (2*R*,4*R*)-monatin) | CLEC-subtilisin (chemo-biocatalytic route); kilogram quantities; natural sweetener. (CSIR Biosciences; Altus[92]) |  6-chloro-2,4,6-trideoxy-D-erythrohexapyranoside (**33**) | Processes based on 2-deoxy-D-ribose 5-phosphate aldolase were developed by various companies (DSM, Diversa, Pfizer) for this and other statin intermediates. The enzyme was engineered to withstand high concentrations of acetaldehyde and for product selectivity[93]; industrial scale. (DSM, US/2009/0209001) |
| **Other** | | | |
|  2′-deoxyinosine (**34**) | Deoxyribose aldolase, phosphopentomutase and purine nucleoside phosphorylase were used in a combinatorial biosynthesis. (Yuki Gosei[94]) |  (2*S*,3*R*,4*S*)-4-hydroxyleucine (**38**) | L-isoleucine dioxygenase. (Ajinomoto[95]) |
|  2-hydroxy-(3′-phenoxy-2-phenylacetonitrile) (**35**) | Protein engineering of a hydroxynitrile lyase for activity and expression. Used for manufacture of cypermethrin. (DSM, EP/2000/0969095) |  (*S*)-2,3-dihydro-1H-indole-2-carboxylic acid (**39**) | Phenylalanine ammonia lyase; recombinant enzyme from metagenome; industrial scale. (DSM[96]) |
|  IMP, GMP (**36**) | AP/PTase random mutagenesis increased transferase activity and decreased phosphatase activity; industrial scale. (Ajinomoto[97]) |  (*S*)-5-chloro-2-isopropylpent-4-enoate (**40**) | Recombinant pig liver esterase; aliskiren intermediate. (DSM[98], WO/2010/10122175) |
|  D-amino acids (**37**) | D-carbamoylase, thermo- and pH-stability; industrial scale. (Kaneka[99]) |  Ethyl (*R*)-4-cyano-3-hydroxybutyrate (**41**) | Directed evolution of a halohydrin dehalogenase for improved activity, stability and tolerance to substrate and product provided a catalyst that is now used for commercial manufacture of this atorvastatin intermediate. (Codexis[52], US/2010/7807423) |
| **Whole cells** | | | |
|  Artemisinin (**42**) | Pathway engineering; overexpression of additional genes; knockouts; PE. (Amyris; Sanofi-Aventis[100]) |  Vanillin (**43**) | Increased productivity; decreased by-product formation. (Givaudan, US/2001/6235507) |

API, active pharma ingredient; Boc, butyloxycarbonyl; Cbz, carbobenzyloxy; CLEC, crosslinked enzyme crystal; d.e., diastereomeric excess; DMSO, dimethylsulphoxide; Et, ethyl; GMP, guanosine-5′-monophosphate; IMP, inosine-5′-monophosphate; IPA, isopropyl alcohol; PE, protein engineering.

**Table 2 | Presidential Green Chemistry Challenge Awards in biocatalysis over the past ten years**

| Product | Technology | Company | Year |
|---|---|---|---|
| Succinic acid as chemical feedstock | Fermentation | BioAmber | 2011 |
| 1,4-butandiol for polymers and chemical feedstock | Fermentation | Genomatica | 2011 |
| Higher alcohols as fuels and chemical feedstocks | Fermentation | UCLA (Prof. Dr J. Liao) | 2010 |
| Renewable petroleum from fatty-acid metabolism intermediates | Fermentation | LS9 | 2010 |
| Sitagliptin: a pharmaceutical ingredient for treatment of type 2 diabetes | Enzyme | Merck and Codexis | 2010 |
| Esters for cosmetics and personal care products | Enzyme | Eastman Chemical Co. | 2009 |
| Atorvastatin intermediate for treatment of high cholesterol | Enzyme | Codexis | 2006 |
| Polyhydroxyalkanoates as biodegradable plastics and chemical feedstock | Fermentation | Metabolix | 2005 |
| Low trans fats and oils for human nutrition | Enzyme | ADM and Novozymes | 2005 |
| Rhamnolipids: biobased, biodegradable industrial surfactants | Fermentation | Jeneil Biosurfactant Company | 2004 |
| Taxol for treatment of breast cancer | Fermentation | Bristol Myers Squibb | 2004 |
| Improved paper recycling using enzymes to remove sticky contaminants | Enzyme | Buckman Laboratories International | 2004 |
| Polyester synthesis using lipases | Enzyme | Polytechnic University (Prof. Dr R. Gross) | 2003 |
| 1,3-propanediol for polymers | Fermentation | Dupont | 2003 |
| Lactic acid for poly(lactic acid) polymers | Fermentation | NatureWorks | 2002 |
| Removal of natural waxes and oils from cotton before it is made into fabric | Enzyme | Novozymes | 2001 |

See the US Environmental Protection Agency's website (http://www.epa.gov/greenchemistry/pubs/pgcc/past.html).

substitutions were not unusual. For example, directed evolution of the halohydrin dehalogenase for manufacture of the atorvastatin (Lipitor) side chain (Fig. 3) changed at least 35 of the 254 amino acids[60] (>14%) and directed evolution of the transaminase for sitagliptin manufacture (Fig. 4) changed 27 of the 330 amino acids[17] (8.2%). Similarly, computational design of a retro aldolase required 8 or 12 amino acid substitutions (4–6%) in the starting enzyme, which was a xylanase composed of 197 amino acids[72].

A second approach investigated in the past ten years is the creation of new, often non-natural, catalytic activities. The starting point for this new activity is usually a catalytically promiscuous reaction. Catalytic promiscuity is the ability of one active site to catalyse more than one reaction type. Typically, the enzyme catalyses one normal reaction and additional side reactions, which may involve common catalytic steps. The new reaction type is not just a substituent added to the substrate, but involves a different transition state and/or forms different types of chemical bond. For example, pyruvate decarboxylase normally converts pyruvate to acetaldehyde and carbon dioxide. However, a promiscuous catalytic activity of pyruvate decarboxylase is the coupling of this acetaldehyde to another aldehyde in an acyloin condensation. Such a non-natural pyruvate-decarboxylase-catalysed condensation of acetaldehyde with benzaldehyde is the basis for an industrial process developed at BASF in the 1920s to make a precursor of the drug Ephedrine. Recent protein engineering enhanced the promiscuous ability of pyruvate decarboxylase to catalyse the acyloin condensation[73]. The normal reaction requires a proton transfer, but the promiscuous reaction does not. A single amino-acid substitution to remove the proton donor disabled the natural activity and increased the promiscuous activity about fivefold.

The method of disabling unwanted pathways to increase flux to the desired product is further developed by the third advance, metabolic pathway engineering. This allows more complex pathways from secondary metabolism to be transferred into new organisms and entirely new biochemical pathways to be created to make pharmaceutical intermediates and biofuels. The normal metabolisms of terpenes, amino acids and fatty acids have been re-engineered to make hydrocarbons, alcohols and polyesters for use as fuels, bulk chemicals and plastics (see above).

## Challenges remaining in biocatalyst engineering

Despite the advances, there remain major challenges to harnessing the advantages of biocatalysis fully. Enzyme engineering is much faster than it was ten years ago, but changing 30–40 amino acids and screening tens of thousands of candidates still requires a large research team. Many, if not all, engineering strategies will yield improved variants, but some will yield better variants and find them faster. Which ones are the better strategies is still unclear. Directly comparing strategies for the same problem and testing the assumptions behind different strategies will identify the most efficient ones.

The first assumption is that the goal can be achieved using enzyme engineering. The thermodynamics of reactions involving non-natural substrates may be less favourable than that of reactions involving natural substrates, and attaining certain enzyme activities may be thermodynamically impossible. Diffusion sets an upper limit to reaction rates. A closer integration of thermodynamics and biocatalytic process development is highly desirable in designing new processes.

Protein engineering often relies on knowing the quaternary structure of the enzyme because residues at the protein–protein interface can contribute to stability. Researchers assume that the structure of the enzyme under reaction conditions (low enzyme concentrations, high substrate concentration, organic solvents and so on) is very similar to that of the crystallized enzyme (high enzyme concentration, no substrate and/or organic solvent). Because proteins crystallize only under narrow conditions found by extensive experimentation, in solution they probably adopt many conformations besides the ones seen in the crystal structures. Furthermore, our understanding of protein dynamics is still very limited and this makes predictions difficult.

Third, enzyme engineering assumes that individual mutations are additive[67]. Although mutations are mostly non-interactive, many interactive mutations are highly useful but difficult to study. One way of identifying cooperative effects involves statistical analysis using the ProSAR algorithm[60], but better techniques are needed to predict at an early stage of protein engineering which additional mutations are possibly additive and which lead to a dead end.

Fourth, computer design of new enzyme activities is not accurate. Design still requires testing 10–20 predictions and usually results in an enzyme with low activity, which then requires substantial further engineering. For example, the initial computer-based design of an enzyme for the manufacture of sitagliptin[17] yielded an enzyme that converted only 0.1 substrate molecule per day, yet that substrate fits well in the active site within the computer model derived from the crystal structure. New enzymes can be designed to catalyse reactions not found in nature (Kemp-elimination[28], new Diels–Alder reactions[74]), but such activities are so far too low for practical use. Better understanding of the mechanistic, dynamic and structural aspects of enzymatic catalysis is needed.

Technical challenges also limit biocatalysis. The current DNA synthesis methods are close to their efficiency limits, but still cost approximately US$0.35 per base (~US$300 per 1,000-nucleotide gene), which is too high for large-scale applications requiring thousands of genes. Longer and cheaper DNA fragments would simplify and speed up experiments. Next-generation DNA synthesis methods may involve synthesis of oligodeoxynucleotides by codons (trinucleotides) rather than individual nucleotides. This approach was first suggested two decades ago but never reached the mainstream, presumably owing to instrumental limitations (synthesis starts with 64 phosphoramidite trinucleotides). Nevertheless, the concept recently was used in the rapid assembly of entire genes in a single synthesis and in the preparation of high-quality mutagenesis libraries[75], and thus seems feasible today.

New ideas for the integration of biocatalysts with nanodevices and in complex multi-enzyme assemblies hold promise for the future. Enzyme immobilization has been a strategy since the early days of biocatalysis, but it

may be more effective when the biocatalyst's surface orientation is controlled[76]. Similarly, using proteins and nucleic-acid scaffolds to control the number and orientation of enzymes within multi-enzyme pathways also improves efficiency[77]. Separately, functional matrices such as carbon nanotubes and quantum dots can substitute for complex biological electron transfer systems, offering new methods for regenerating redox catalysts and interfacing enzymes with semiconductors[78]. Nevertheless, the integration of enzymes with non-biological matrices and nanomaterials, and as part of metabolic engineering, is still inefficient. Future protein engineering has to address challenges emerging through the interfacing of individual biocatalysts with other proteins in a metabolic pathway or support matrices.

Protein engineering solves the previous weaknesses of biocatalysts: low stability and low activity towards unusual substrates. Large amounts of protein were used to compensate for low activity and this caused emulsions that hampered work-up and reduced yield. Highly active enzymes solve this problem because emulsions do not form using smaller amounts of protein. Training chemists in both biocatalysis and chemocatalysis will help them choose the best solution in each case. Improved enzymes with a long shelf life and good activity and stability in organic solvents should help biocatalysis to spread further into industrial laboratories.

Recent advances in protein engineering have achieved the equivalent of converting mouse proteins into human proteins. The amino-acid sequences of similar proteins in mice and human typically differ by ~13% (ref. 79). Today's advanced protein engineering makes similar changes in converting a wild-type enzyme into an enzyme suitable for chemical process applications. This protein engineering is equivalent to compressing the 75,000,000-yr evolution of an early mammal into modern-day mice and humans into several months of laboratory work. Consistent with the more extensive changes made in these proteins, the properties have also changed more dramatically. The catalytic properties of the enzymes have improved quantitatively by factors of thousands to millions[80], and the engineered enzymes now can act in unusually harsh conditions. The understanding of protein engineering built over the third wave of biocatalysis allows dramatic improvements in enzymatic performance to be realized in parallel with the development of chemical syntheses requiring these catalysts, allowing biocatalysis to develop as an increasingly important tool in chemical synthesis.


1. Buchholz, K., Kasche, V. & Bornscheuer, U. T. *Biocatalysts and Enzyme Technology* 2nd edn (Wiley-VCH, 2012).
2. Drauz, K., Gröger, H. & May, O. (eds) *Enzyme Catalysis in Organic Synthesis* Vols 1–3, 3rd edn (Wiley-VCH, 2012).
3. Bornscheuer, U. T. & Kazlauskas, R. J. *Hydrolases in Organic Synthesis - Regio- and Stereoselective Biotransformations* 2nd edn (Wiley-VCH, 2006).
4. Liese, A., Seelbach, K. & Wandrey, C. (eds) *Industrial Biotransformations* 2nd edn (Wiley-VCH, 2006).
5. Wenda, S., Illner, S., Mell, A. & Kragl, U. Industrial biotechnology—the future of green chemistry? *Green Chem.* **13,** 3007–3047 (2011).
6. Rosenthaler, L. Durch Enzyme bewirkte asymmetrische Synthese. *Biochem. Z.* **14,** 238–253 (1908).
7. Sedlaczek, L. Biotransformations of steroids. *Crit. Rev. Biotechnol.* **7,** 187–236 (1988).
8. Estell, D. A., Graycar, T. P. & Wells, J. A. Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *J. Biol. Chem.* **260,** 6518–6521 (1985).
   **This paper puts forward the basis for the first application of protein engineering in industrial biotechnology.**
9. Jensen, V. J. & Rugh, S. Industrial scale production and application of immobilized glucose isomerase. *Methods Enzymol.* **136,** 356–370 (1987).
10. Bruggink, A., Roos, E. C. & de Vroom, E. Penicillin acylase in the industrial production of β-lactam antibiotics. *Org. Process Res. Dev.* **2,** 128–133 (1998).
11. Griengl, H., Schwab, H. & Fechter, M. The synthesis of chiral cyanohydrins by oxynitrilases. *Trends Biotechnol.* **18,** 252–256 (2000).
12. Hills, G. Industrial use of lipases to produce fatty acid esters. *Eur. J. Lipid Sci. Technol.* **105,** 601–607 (2003).
13. Nagasawa, T., Nakamura, T. & Yamada, H. Production of acrylic acid and methacrylic acid using *Rhodococcus rhodochrous* J1 nitrilase. *Appl. Microbiol. Biotechnol.* **34,** 322–324 (1990).
14. Francis, J. C. & Hansche, P. E. Directed evolution of metabolic pathways in microbial populations. I. Modification of the acid phosphatase pH optimum in *S. cerevisiae*. *Genetics* **70,** 59–73 (1972).
15. Nakamura, C. E. & Whited, G. M. Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol.* **14,** 454–459 (2003).
16. Liang, J. *et al.* Development of a biocatalytic process as an alternative to the (-)-DIP-Cl-mediated asymmetric reduction of a key intermediate of montelukast. *Org. Process Res. Dev.* **14,** 193–198 (2010).
17. Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329,** 305–309 (2010).
18. Desai, A. A. Sitagliptin manufacture: a compelling tale of green chemistry, process intensification, and industrial asymmetric catalysis. *Angew. Chem. Int. Ed.* **50,** 1974–1976 (2011).
    **This highlight article compares in detail two processes for the production of sitagliptin, one catalysed by rhodium and one catalysed by a transaminase.**
19. O'Maille, P. E. *et al.* Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nature Chem. Biol.* **4,** 617–623 (2008).
20. Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451,** 86–89 (2008).
    **This paper describes the efficient diversion of amino-acid metabolism to the production of alcohols.**
21. Lee, S. K., Chou, H., Ham, T. S., Lee, T. S. & Keasling, J. D. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr. Opin. Biotechnol.* **19,** 556–563 (2008).
22. Steen, E. J. *et al.* Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463,** 559–562 (2010).
23. Bohmert-Tatarev, K., McAvoy, S., Daughtry, S., Peoples, O. P. & Snell, K. D. High levels of bioplastic are produced in fertile transplastomic tobacco plants engineered with a synthetic operon for the production of polyhydroxybutyrate. *Plant Physiol.* **155,** 1690–1708 (2011).
24. McKenna, R. & Nielsen, D. R. Styrene biosynthesis from glucose by engineered *E. coli*. *Metab. Eng.* **13,** 544–554 (2011).
25. Kazlauskas, R. J. & Bornscheuer, U. T. Finding better protein engineering strategies. *Nature Chem. Biol.* **5,** 526–529 (2009).
26. Lutz, S. & Bornscheuer, U. T. (eds) *Protein Engineering Handbook* (Wiley-VCH, 2009).
27. Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nature Chem. Biol.* **5,** 567–573 (2009).
28. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453,** 190–195 (2008).
29. Schmid, A. *et al.* Industrial biocatalysis today and tomorrow. *Nature* **409,** 258–268 (2001).
30. Arnold, F. H. Combinatorial and computational challenges for biocatalyst design. *Nature* **409,** 253–257 (2001).
31. Schoemaker, H. E., Mink, D. & Wubbolts, M. G. Dispelling the myths—biocatalysis in industrial synthesis. *Science* **299,** 1694–1697 (2003).
32. Keasling, J. D. Manufacturing molecules through metabolic engineering. *Science* **330,** 1355–1358 (2010).
33. Madison, L. L. & Huisman, G. W. Metabolic engineering of poly(3-hydroxyalkanoates): from DNA to plastic. *Microbiol. Mol. Biol. Rev.* **63,** 21–53 (1999).
34. Wetterstrand, K. A. DNA sequencing costs: data from the NHGRI large-scale genome sequencing program. *National Human Genome Research Project* ⟨http://www.genome.gov/sequencingcosts⟩ (2011).
35. Lorenz, P. & Eck, J. Metagenomics and industrial applications. *Nature Rev. Microbiol.* **3,** 510–516 (2005).
36. Fowler, D. M. *et al.* High-resolution mapping of protein sequence- function relationships. *Nature Methods* **7,** 741–746 (2010).
37. Richmond, K. E. *et al.* Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Res.* **32,** 5011–5018 (2004).
38. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329,** 52–56 (2010).
39. Lutz, S. Beyond directed evolution—semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **21,** 734–743 (2010).
40. Höhne, M., Schätzle, S., Jochens, H., Robins, K. & Bornscheuer, U. T. Rational assignment of key motifs for function guides in silico enzyme identification. *Nature Chem. Biol.* **6,** 807–813 (2010).
    **In this paper, careful analysis of key motifs of 5,000 pyridoxal-5-phosphate-dependent transaminase sequences in public databases identified 20 novel enzymes for which substrate preference (ketone, not α-keto acid) and enantiopreference ((*R*), not (*S*)) could be predicted and experimentally confirmed.**
41. Jochens, H. & Bornscheuer, U. T. Natural diversity to guide focused directed evolution. *ChemBioChem* **11,** 1861–1866 (2010).
42. Park, S. *et al.* Focusing mutations into the *P. fluorescens* esterase binding site increases enantioselectivity more effectively than distant mutations. *Chem. Biol.* **12,** 45–54 (2005).
43. Horsman, G. P., Liu, A. M. F., Henke, E., Bornscheuer, U. T. & Kazlauskas, R. J. Mutations in distant residues moderately increase the enantioselectivity of *Pseudomonas fluorescens* esterase towards methyl 3-bromo-2-methylpropanoate and ethyl 3-phenylbutyrate. *Chemistry* **9,** 1933–1939 (2003).
44. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472,** 499–503 (2011).
45. Moore, J. C., Pollard, D. J., Kosjek, B. & Devine, P. N. Advances in the enzymatic reduction of ketones. *Acc. Chem. Res.* **40,** 1412–1419 (2007).
46. Strohmeier, G. A., Pichler, H., May, O. & Gruber-Khadjawi, M. Application of designed enzymes in organic synthesis. *Chem. Rev.* **111,** 4141–4164 (2011).
    **This is a review about protein engineering to create biocatalysts for industrial applications.**
47. Wiese, A., Wilms, B., Syldatk, C., Mattes, R. & Altenbuchner, J. Cloning, nucleotide sequence and expression of a hydantoinase and carbamoylase gene from

*Arthrobacter aurescens* DSM 3745 in *Escherichia coli* and comparison with the corresponding genes from *Arthrobacter aurescens* DSM 3747. *Appl. Microbiol. Biotechnol.* **55,** 750–757 (2001).
48. Martinez-Gomez, A. I. *et al.* Recombinant polycistronic structure of hydantoinase process genes in *Escherichia coli* for the production of optically pure D-amino acids. *Appl. Environ. Microbiol.* **73,** 1525–1531 (2007).
49. Panke, S. & Wubbolts, M. Advances in biocatalytic synthesis of pharmaceutical intermediates. *Curr. Opin. Chem. Biol.* **9,** 188–194 (2005).
50. Breuer, M. *et al.* Industrial methods for the production of optically active intermediates. *Angew. Chem. Int. Ed.* **43,** 788–824 (2004).
51. DeSantis, G. *et al.* Creation of a productive, highly enantioselective nitrilase through gene site saturation mutagenesis (GSSM). *J. Am. Chem. Soc.* **125,** 11476–11477 (2003).
**This paper describes how a single amino-acid substitution can create a nitrilase with high enantioselectivity at 3 M substrate concentration, for synthesis of an intermediate for atorvastatin.**
52. Ma, S. K. *et al.* A green-by-design biocatalytic process for atorvastatin intermediate. *Green Chem.* **12,** 81–86 (2010).
53. Anastas, P. & Warner, J. (eds) *Green Chemistry: Theory and Practice* (Oxford Univ. Press, 1998).
54. Yang, T. H. *et al.* Biosynthesis of polylactic acid and its copolymers using evolved propionate CoA transferase and PHA synthase. *Biotechnol. Bioeng.* **105,** 150–160 (2010).
55. Bernath, K. *et al. In vitro* compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. *Anal. Biochem.* **325,** 151–157 (2004).
56. Becker, S. *et al.* Single-cell high-throughput screening to identify enantioselective hydrolytic enzymes. *Angew. Chem. Int. Ed.* **47,** 5085–5088 (2008).
57. Fernández-Alvaro, E. *et al.* A combination of *in vivo* selection and cell sorting for the identification of enantioselective biocatalysts. *Angew. Chem. Int. Ed.* **50,** 8584–8587 (2011).
58. Whittle, E. & Shanklin, J. Engineering delta 9–16:0-acyl carrier protein (ACP) desaturase specificity based on combinatorial saturation mutagenesis and logical redesign of the castor delta 9–18:0-ACP desaturase. *J. Biol. Chem.* **276,** 21500–21505 (2001).
59. Seelig, B. & Szostak, J. W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **448,** 828–831 (2007).
60. Fox, R. J. *et al.* Improving catalytic function by ProSAR-driven enzyme evolution. *Nature Biotechnol.* **25,** 338–344 (2007).
61. Yoshikuni, Y., Ferrin, T. E. & Keasling, J. D. Designed divergent evolution of enzyme function. *Nature* **440,** 1078–1082 (2006).
62. Bava, K. A., Gromiha, M. M., Uedaira, H., Kitajima, K., &. Sarai, A. ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.* **32,** D120–D121 (2004).
63. Guo, H. H., Choe, J. & Loeb, L. A. Protein tolerance to random amino acid change. *Proc. Natl Acad. Sci. USA* **101,** 9205–9210 (2004).
**This paper shows that about 34% of random amino-acid replacements inactivate a protein's functions, indicating the importance of starting with a stabilized protein for protein engineering experiments.**
64. Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O. & Arnold, F. H. Why highly expressed proteins evolve slowly. *Proc. Natl Acad. Sci. USA* **102,** 14338–14343 (2005).
65. Gupta, R. D. & Tawfik, D. S. Directed enzyme evolution via small and effective neutral drift libraries. *Nature Methods* **5,** 939–942 (2008).
66. Bloom, J. D., Romero, P. A., Lu, Z. & Arnold, F. H. Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol. Direct* **2,** 17 (2007).
67. Wells, J. A. Additivity of mutational effects in proteins. *Biochemistry* **29,** 8509–8517 (1990).
68. Moore, J. C. & Arnold, F. H. Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotechnol.* **14,** 458–467 (1996).
69. Reetz, M. T., Zonta, A., Schimossek, K., Liebeton, K. & Jaeger, K.-E. Creation of enantioselective biocatalysts for organic chemistry by *in vitro* evolution. *Angew. Chem. Int. Edn Engl.* **36,** 2830–2832 (1997).
70. Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312,** 111–114 (2006).
71. Reetz, M. T. & Sanchis, J. Constructing and analyzing the fitness landscape of an experimental evolutionary process. *ChemBioChem* **9,** 2260–2267 (2008).
72. Jiang, L. *et al. De novo* computational design of retro-aldol enzymes. *Science* **319,** 1387–1391 (2008).
73. Meyer, D. *et al.* Conversion of pyruvate decarboxylase into an enantioselective carboligase with biosynthetic potential. *J. Am. Chem. Soc.* **133,** 3609–3616 (2011).
74. Siegel, J. B. *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329,** 309–313 (2010).
75. Randolph, J., Yagodkin, A., Lamaitre, M., Azhayev, A. & Mackie, H. Codon based mutagenesis using trimer phosphoramidites. *Nucleic Acids Symp. Ser.* **52,** 479 (2008).
76. Rana, S., Yeh, Y. C. & Rotello, V. M. Engineering the nanoparticle-protein interface: applications and possibilities. *Curr. Opin. Chem. Biol.* **14,** 828–834 (2010).
77. Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnol.* **27,** 753–759 (2009).
78. McDonald, T. J. *et al.* Wiring-up hydrogenase with single-walled carbon nanotubes. *Nano Lett.* **7,** 3528–3534 (2007).
79. Mouse Genome Sequencing Consortium.. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
80. Savile, C. K. & Lalonde, J. J. Biotechnology for the acceleration of carbon dioxide capture and sequestration. *Curr. Opin. Biotechnol.* **22,** 818–823 (2011).
81. Morikawa, S. *et al.* Highly active mutants of carbonyl reductase S1 with inverted coenzyme specificity and production of optically active alcohols. *Biosci. Biotechnol. Biochem.* **69,** 544–552 (2005).
82. Patel, R. N., Chu, L. & Mueller, R. Diastereoselective microbial reduction of (S)-[3-chloro-2-oxo-1-(phenylmethyl)propyl]carbamic acid, 1,1-dimethylethyl ester. *Tetrahedr. Asymm.* **14,** 3105–3109 (2003).
83. Liang, J. *et al.* Highly enantioselective reduction of a small heterocyclic ketone: biocatalytic reduction of tetrahydrothiophene-3-one to the corresponding (*R*)-alcohol. *Org. Process Res. Dev.* **14,** 188–192 (2010).
84. Urano, N. *et al.* Directed evolution of an aminoalcohol dehydrogenase for efficient production of double chiral aminoalcohols. *J. Biosci. Bioeng.* **111,** 266–271 (2011).
85. Gooding, O. W. *et al.* Development of a practical biocatalytic process for (*R*)-2-methylpentanol. *Org. Process Res. Dev.* **14,** 119–126 (2010).
86. Cobley, C. J., Hanson, C. H., Lloyd, M. C. & Simmonds, S. The combination of hydroformylation and biocatalysis for the large-scale synthesis of (*S*)-allysine ethylene acetal. *Org. Process Res. Dev.* **15,** 284–290 (2011).
87. Xie, X., Watanabe, K., Wojcicki, W. A., Wang, C. C. & Tang, Y. Biosynthesis of lovastatin analogs with a broadly specific acyltransferase. *Chem. Biol.* **13,** 1161–1169 (2006).
88. Truppo, M. D. & Hughes, G. Development of an improved immobilized CAL-B for the enzymatic resolution of a key intermediate to odanacatib. *Org. Process Res. Dev.* **15,** 1033–1035 (2011).
89. Brocklehurst, C. E., Laumen, K., Vecchia, L. L., Shaw, D. & Vögtle, M. Diastereoisomeric salt formation and enzyme-catalyzed kinetic resolution as complementary methods for the chiral separation of cis-/trans-enantiomers of 3-aminocyclohexanol. *Org. Process Res. Dev.* **15,** 294–300 (2011).
90. Iding, H. *et al.* in *Asymmetric Catalysis on Industrial Scale* (eds Blaser, H. U. & Federsel, H. J.) 377–396 (Wiley-VCH, 2010).
91. Chen, Y. J. *et al.* Enzymatic preparation of an (*S*)-amino acid from a racemic amino acid. *Org. Process Res. Dev.* **15,** 241–248 (2011).
92. Rousseau, A. L. *et al.* Scale-up of a chemo-biocatalytic route to (*2R,4R*)- and (*2S,4S*)-monatin. *Org. Process Res. Dev.* **15,** 249–257 (2011).
93. Greenberg, W. A. *et al.* Development of an efficient, scalable, aldolase-catalyzed process for enantioselective synthesis of statin intermediates. *Proc. Natl Acad. Sci. USA* **101,** 5788–5793 (2004).
94. Horinouchi, N. *et al.* Biochemical retrosynthesis of 2′-deoxyribonucleosides from glucose, acetaldehyde, and a nucleobase. *Appl. Microbiol. Biotechnol.* **71,** 615–621 (2006).
95. Hibi, M. *et al.* Characterization of *Bacillus thuringiensis L*-isoleucine dioxygenase for production of useful amino acids. *Appl. Environ. Microbiol.* **77,** 6926–6930 (2011).
96. de Lange, B. *et al.* Asymmetric synthesis of (*S*)-2-indoline carboxylic acid by combining biocatalysis and homogeneous catalysis. *ChemCatChem* **3,** 289–292 (2011).
97. Asano, Y., Mihara, Y. & Yamada, H. A new enzymatic method of selective phosphorylation of nucleosides. *J. Mol. Catal. B* **6,** 271–277 (1999).
98. Hermann, M. *et al.* Alternative pig liver esterase (APLE) - cloning, identification and functional expression in *Pichia pastoris* of a versatile new biocatalyst. *J. Biotechnol.* **133,** 301–310 (2008).
99. Ikenaka, Y. *et al.* Thermostability reinforcement through a combination of thermostability-related mutations of N-carbamyl-D-amino acid amidohydrolase. *Biosci. Biotechnol. Biochem.* **63,** 91–95 (1999).
100. Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440,** 940–943 (2006).

**Acknowledgements** We thank H.-P. Meyer and R. Fox for discussions. R.J.K. thanks the US National Science Foundation (CBET-0932762) and the Korea Science and Engineering Foundation funded by the Ministry of Education, Science and Technology (WCU programme R32-2008-000-10213-0). U.T.B. and S.L. thank the German Research Foundation (SPP 1170, Bo1864/4-1) and, respectively, the US National Science Foundation (CBET-0730312) for financial support.

**Author Contributions** U.T.B., R.J.K. and S.L. drafted the text; K.R., G.W.H. and J.C.M. collected the examples for industrial applications; and the authors together wrote and edited the review. All authors contributed equally.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to U.T.B. (uwe.bornscheuer@uni-greifswald.de).