

# Combinatorial and computational challenges for biocatalyst design

Frances H. Arnold

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

**Nature provides a fantastic array of catalysts extremely well suited to supporting life, but usually not so well suited for technology. Whether biocatalysis will have a significant technological impact depends on our finding robust routes for tailoring nature's catalysts or redesigning them anew. Laboratory evolution methods are now used widely to fine-tune the selectivity and activity of enzymes. The current rapid development of these combinatorial methods promises solutions to more complex problems, including the creation of new biosynthetic pathways. Computational methods are also developing quickly. The marriage of these approaches will allow us to generate the efficient, effective catalysts needed by the pharmaceutical, food and chemicals industries and should open up new opportunities for producing energy and chemicals from renewable resources.**

**B**iological systems are masterful chemists — a fact long appreciated by those who study how living things build complex molecules and systems from simple compounds. Enzymes catalyse the interconversion of a vast number of molecular structures, achieving tasks that range from the fixation of nitrogen to the synthesis of large and intricately structured molecules that ward off predators or attract mates. Such catalysts are models of energy-efficient, environmentally benign chemical agents, as virtually all do their work under mild conditions — in water, at room temperature and atmospheric pressure — and generate few waste products.

In view of increasing environmental and economic pressure to use renewable sources for energy and chemical feedstocks in industry, biocatalysts look like potentially attractive technological tools. But enzymes have evolved to contribute to the survival and reproduction of the organisms that make them; that they might also be useful in laundry detergents or to synthesize a new drug is simply serendipitous. In fact, attempts to use enzymes or whole organisms in applied chemical processes or products reveal some severe disadvantages of biocatalysts. Some of them turn off when a little product accumulates. This feature, so useful in regulating the flow of metabolites inside a cell, quickly derails implementation of a biocatalytic process to make that product. The process engineer is also unlikely to favour a delicate catalyst that must be replaced every few hours or must be coddled to keep it going. And of course nature does not conveniently provide a catalyst for any transformation we wish to conduct.

For many years the identification of new biocatalysts depended on labour-intensive screening of microbial cultures for the desired activities. Almost all the biocatalysts in use today came from the small fraction of organisms that can be grown under controlled conditions, the 'microbial weeds'. (By most counts less than 1% of all microorganisms can be cultured.) Some of these organisms live in harsh environments and their catalysts exhibit remarkable and useful properties, including the ability to function under extreme conditions of temperature, salt or pH<sup>1</sup>. Other organisms have potentially useful new catalysts or enzyme pathways that allow them to produce valuable, biologically active compounds<sup>2</sup>. These catalysts could be recruited to make the natural products in more tractable organisms such as *Escherichia coli*.

Efforts to comb natural biodiversity for useful activities have been greatly facilitated by high-throughput screening technologies and by new methods for collecting genes from the environment and expressing them in recombinant organisms<sup>3-4</sup>. These processes allow faster access to useful catalytic activities from organisms that cannot be cultured. But natural diversity cannot address all practical biocatalytic problems. Screening larger libraries of DNA or microbes may not even be the fastest or most efficient route to obtaining a good catalyst. Some problems can be solved by the right method of implementation — immobilization or crystallization can stabilize weak protein structures, for example. But many problems are best attacked by engineering the catalyst itself, whether it be a single enzyme, multiple enzymes or even a whole cell.

A revolution in biological design possibilities was unleashed by the advent of recombinant DNA technology, with which one can manipulate DNA sequences in a highly specific fashion and express their protein products in a variety of organisms, from animals to bacteria. This provides a means to redesign nature's catalysts at the molecular level according to detailed specifications, and to produce them in large quantities in fast-growing microorganisms. In this review I consider the ways in which biotechnological methods permit the restructuring of enzymes to adapt their functions for applied ends. Broadly speaking, one can identify two philosophies: either existing biocatalysts can be fine-tuned by rational redesign, or combinatorial techniques can be used to search for useful functionality in libraries generated at random and improved by suitable selection methods.

Our ability to manipulate the structures and functions of biological molecules and even whole organisms at will carries the prospect of applications previously not considered in the realm of biocatalysis, such as very large scale chemical production. In the not-too-distant future we can expect custom-made enzymes for gene therapies<sup>5,6</sup>, new reagents for basic science and clinical diagnostics<sup>7,8</sup>, and even new designs of the cellular machinery for making proteins *in vivo*<sup>9</sup>.

## Rational redesign of natural biocatalysts

To take full advantage of recombinant DNA technology for making new enzymes, we need to know the connection between protein sequence and function. In other words, redesigning nature's catalysts rationally — that is, by

specifying the sequence — usually requires detailed understanding of structures and mechanisms. This information is unavailable for the vast majority of enzymes. Even if the target enzyme is well characterized, the molecular basis for the desired function may not be. With hundreds and even thousands of atoms that interact weakly with each other in an ensemble of closely related and interconverting folded conformations, the complex and finely tuned enzyme fades easily in the clumsy hands of the protein engineer.

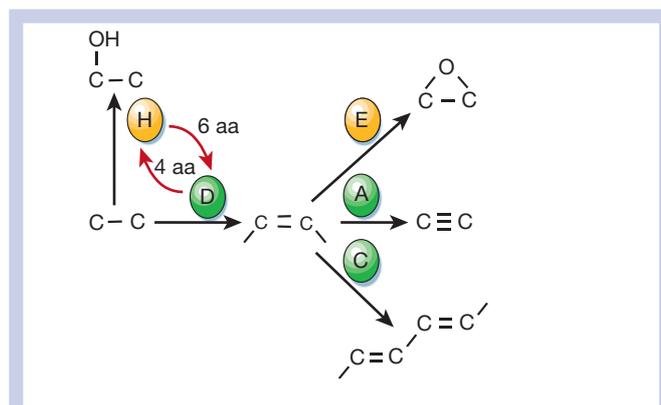
Despite these challenges, biological design is now going through its most exciting period since the introduction of recombinant DNA methods and the invention of site-directed mutagenesis over two decades ago. One factor contributing to this capability is the exponential growth of the databases of protein structures and sequences. By arranging proteins in family trees we have learnt that those proteins existing today evolved from a much smaller set of ancient molecules. Shuffling of segments, fusions with other proteins and accumulation of random mutations all contributed to their diversification. Sometimes the sequences remained sufficiently conserved during this process that we can compare the sequence of a new biocatalyst identified in a screening programme to the thousands that have been deposited in databases and identify related proteins whose functions, and maybe even structures, are already known. With this information we can make inferences about the new catalyst's structure and activity.

There is now ample evidence that new enzymes evolved in nature by relatively minor modification of active-site structures<sup>10,11</sup>. This sequence and structure information can sometimes be used to good effect in transferring activities from one enzyme to another related one. Shanklin and co-workers have exploited this notion to re-engineer membrane-bound di-iron enzymes that exhibit distinct hydroxylase, epoxidase, acetylenase and conjugase activities during fatty-acid biosynthesis in plants<sup>12,13</sup>. It is believed that all these activities arise from a common progenitor enzyme through modification of the active site to allow direction of a common free-radical reaction intermediate into the different end products<sup>14</sup> (Fig. 1). Comparing the sequences of five related oleate desaturases with those of two hydroxylases, Shanklin and co-workers identified seven positions that were strictly conserved within the five desaturases but which differed from the equivalent positions in the two hydroxylases<sup>12</sup>. Reciprocal amino-acid changes between one of the desaturases and one of the hydroxylases at the seven sites yielded pronounced shifts in the ratio of desaturase to hydroxylase activity. Four amino-acid substitutions were sufficient to convert the desaturase to a hydroxylase, and as few as six substitutions turn the hydroxylase into a desaturase.

Another example comes from work on a dehalogenase from the 2-enoyl-coenzyme A (CoA) hydratase/isomerase superfamily<sup>11</sup>. Comparison of the sequences and structures of various family members revealed that their active sites can be viewed as derivatives of a single active-site structure that provides CoA binding, an oxyanion pocket and a chamber containing stations at which substrate binding and catalytic groups have been strategically positioned. Experiments with site-directed mutagenesis show that chemical diversification can be achieved through placement of one or more polar residues along the stations: grafting eight amino-acid substitutions from crotonyl hydratase conferred on its relative, 4-chlorobenzoyl-CoA-dehalogenase, the new ability to catalyse the hydration of crotonyl-CoA.

This comparative approach can be useful for identifying amino acids that control particular enzyme behaviours and demonstrating mechanisms for the diversification of catalytic functions in nature. Catalytic plasticity has clearly contributed to the evolution of chemical diversity. But such comparisons are likely to be of limited use for designing new biocatalysts, because the changes produced by altering the identified amino acids often do not extend beyond the range encoded in the parental genes<sup>13</sup>. One of the main goals of biocatalyst engineering is to endow them with new features that are not found in the natural sequences because they confer no evolutionary advantage.

A further problem is that the amino acids care about their context — their neighbours influence their contributions to an enzyme's



**Figure 1** Catalytic plasticity in a family of fatty-acid synthesis enzymes. A family of closely related (>50% amino-acid identity) Fad2-desaturase-like lipid-modification enzymes can mediate a range of functional outcomes. Enzymes: D, oleate desaturase; H, oleate hydroxylase; E, linoleate epoxygenase, A, linoleate acetylenase; C, linoleate conjugase. Yellow enzymes denote oxygen transfer; green enzymes denote hydrogen abstraction. Red arrows indicate the number of amino-acid (aa) substitutions shown to substantially affect reaction outcome<sup>13,14</sup>.

activity. Also, many important biocatalyst properties are not localized in a small number of catalytic residues, but reflect contributions from many residues distributed over large parts of the protein. Even when large functional changes can be obtained with a few amino-acid substitutions, it will often be difficult or impossible to discern the specific mutations responsible. Most sequence changes accumulated during evolution have little or no effect on the property of interest, and their presence makes it difficult to pick out the key positions<sup>15</sup>. A good example is stability. Hundreds of amino acids using many types of interactions can contribute to the stability of a protein, and useful design rules for stabilization have not yet been extracted from sequence comparisons<sup>16</sup>. The factors determining stability have, however, been good targets for powerful computational methods of protein modelling that can handle the large numbers of competing interactions<sup>17,18</sup>.

Nearly all engineered enzymes that are used today came out of structure-based protein-engineering efforts of the 1980s. The successes have been notable, but the results were costly and came far too slowly. Although some properties, notably enzyme specificity, respond relatively well to structure-based design and site-directed mutagenesis<sup>19</sup>, this approach is often cumbersome and unsatisfactory for engineering industrial biocatalysts which must meet a long list of performance specifications and for which the windows of opportunity are all too brief. In the pharmaceuticals industry, new catalysts must be selected and implemented often within a few months. Predictive capabilities are still rudimentary for catalysis, and even when successful, the desired changes in activity and specificity often come at the cost of other, equally important properties, such as stability or expression level.

### Breeding a better catalyst

Another key factor contributing to expanding biological design capabilities is the development of 'evolutionary' protein design methods<sup>20</sup> using random mutagenesis, gene recombination and high-throughput screening.

Unlike natural evolution, laboratory evolution is directed — more like breeding<sup>21,22</sup>. A 'generation' of molecules can be bred in a few days, with large numbers of progeny subject to selective pressures not encountered in nature. Because the molecules are produced in recombinant cells and are decoupled from their biological functions, they can be bred for non-natural but useful properties, including the ability to carry out reactions on substrates not encountered in nature, or to function under highly unusual conditions. Because molecules can be bred for multiple traits simultaneously by changing the

conditions of the screen or selection, this approach is particularly attractive for engineering industrial biocatalysts.

Although there are many ways to evolve a biocatalyst in the laboratory, they all involve two main steps: making a set of mutant biocatalysts and searching that set for mutants with the desired properties. The process can be iterative, so that large changes in function are obtained by accumulating small changes over many generations.

Sequential rounds of random mutagenesis carried out on ever-improved mutants is a simple and highly effective strategy that has been applied successfully to a number of catalyst design problems. Particularly relevant to biocatalysis — and particularly difficult to manipulate using structure-based design — is enzyme enantioselectivity<sup>23</sup>. Subtle changes in enzyme structure and even changes in reaction conditions can influence enantioselectivity, but these effects are almost impossible to predict. Enantioselectivity can, however, be tuned by laboratory evolution. Starting from a naturally occurring lipase with almost no selectivity for the hydrolysis of racemic 2-methyldecanoic ester, Liebeton *et al.*<sup>24</sup> evolved an enzyme that catalysed the reaction at more than 90% enantiomeric excess using several rounds of mutagenesis and screening. In another recent study, three generations of mutagenesis and screening actually inverted the enantioselectivity of a hydantoinase to prefer L- over D-5-(2-methylthioethyl) hydantoin and increased its activity fivefold<sup>25</sup>. Degussa AG is currently evaluating a whole-cell catalyst incorporating the evolved enzyme for commercial production of enantiopure L-methionine.

Laboratory evolution has also been effective in altering other key biocatalyst properties, including stability, function in non-natural environments (such as organic solvents; see accompanying review by Klivanov, pages 241–246), product inhibition, expression in a recombinant host and substrate specificity<sup>21,23,26</sup>. A particularly impressive example is the evolution of an aspartate transaminase to have  $2.6 \times 10^6$ -fold higher activity towards the non-native substrate valine<sup>27,28</sup>. The crystal structure of the evolved enzyme shows how the active site was remodelled through the cumulative effects of mutations distributed over much of the enzyme structure<sup>27</sup>. Yet only one of the 17 mutated residues contacts the substrate, and none contact the pyridoxal 5'-phosphate cofactor. This study illustrates well how complex the solutions to enzyme design problems can be, a point echoed in structural analyses of other laboratory-evolved enzymes<sup>29</sup>. In the right places, amino acids serve as 'molecular shims'<sup>13</sup> to tune

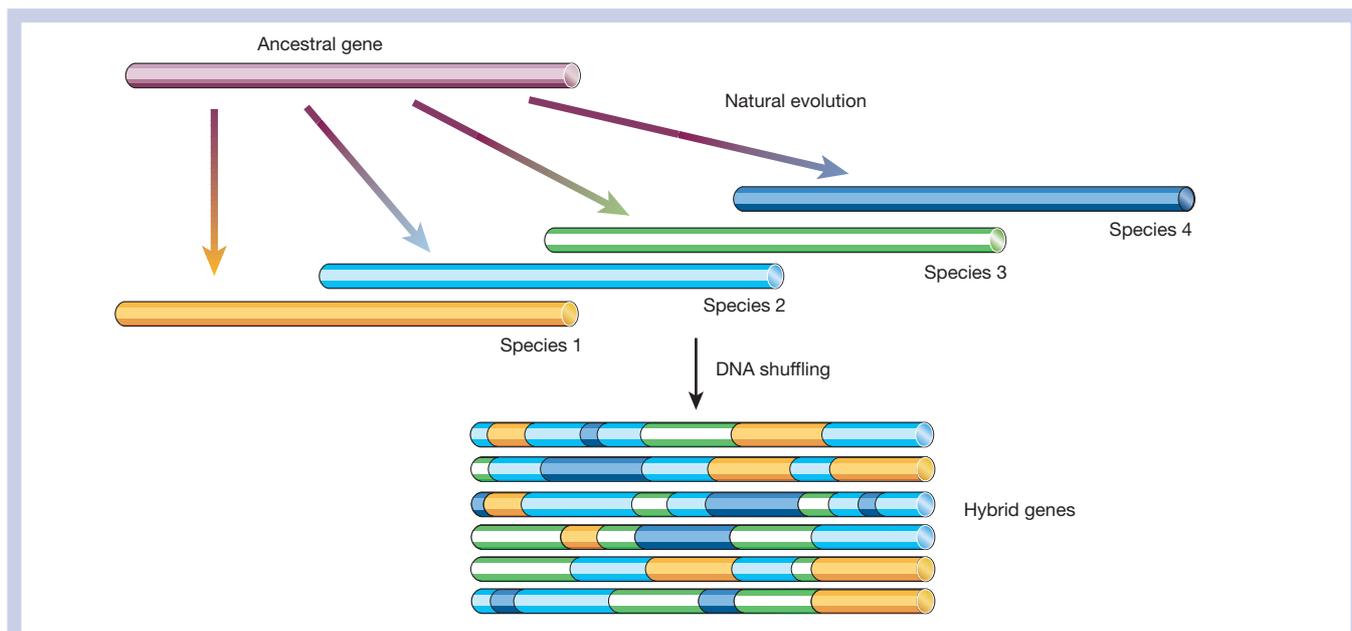
substrate and reaction specificity; beneficial amino-acid substitutions easily identified by random mutagenesis and screening may have minute structural consequences, beyond the resolution of structural analysis and certainly beyond our ability to predict.

Accumulating point mutations is an effective fine-tuning mechanism, but nature also uses other means to create new molecular diversity on which evolution can act. One of those is recombination. Recent studies show that recombination is an extremely useful operation for laboratory evolution. So-called DNA shuffling methods<sup>30</sup> pioneered by Stemmer create hybrid gene libraries by homologous recombination of related parent genes (ref. 31 and Fig. 2). This 'molecular sex' creates new genes that code for proteins with sequence information from any or all parents. Genes from multiple parents and even from different species can be shuffled in a single step, operations that are forbidden in nature but may be very useful for rapid adaptation. DNA shuffling is used widely to generate highly improved biocatalysts, as well as ones with features not present in the parent enzymes and not known to occur in nature<sup>21,22</sup>.

This molecular breeding concept extends nicely to more complex problems involving many interacting genes. A good example is creating new, multienzyme pathways for making chemicals. Microorganisms, plants and animals produce a wide range of compounds that could function as new drugs, dyes, fragrances, flavourings and cosmetics. But many are found only in trace quantities in their natural sources and are difficult or impossible to synthesize chemically. An important goal for biocatalysis is to produce these compounds in fast-growing organisms suitable for large-scale production.

Genes encoding the enzymes that catalyse the series of chemical reactions necessary to make a particular compound can be transferred to more amenable host organisms, conferring on them the new ability to synthesize the desired product<sup>32,33</sup>. Molecular breeding can optimize the engineered pathways, and it can also create new pathways, capable of synthesizing novel compounds.

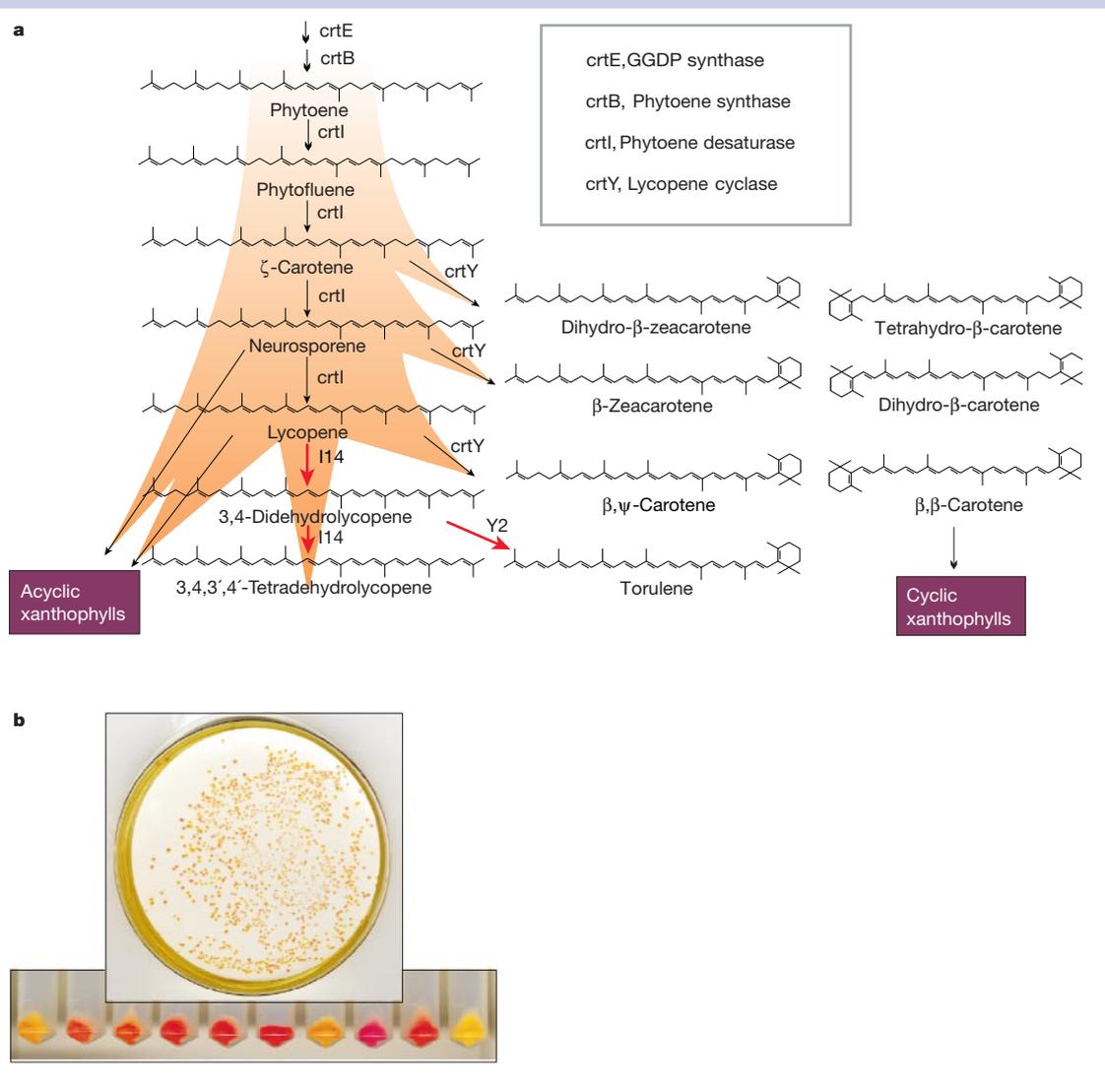
Schmidt-Dannert *et al.*<sup>34</sup> have evolved the pathways that synthesize carotenoid pigments. Using a small set of bacterial genes that produce  $\beta$ -carotenes, they were able to exploit the remarkable plasticity of carotenoid biosynthetic pathways to generate pathways for a number of related carotenoids and precursors (Fig. 3). The two genes from *Erwinia* sp. that produce phytoene were engineered into



**Figure 2** Molecular breeding by DNA shuffling. Diverse gene libraries for laboratory evolution can be created by recombination of related genes<sup>31</sup>. This approach generates highly diverse sequences, but conserves function. Improved or altered enzymes have been identified by screening such hybrid protein libraries.

**Figure 3** Evolution of pathways that synthesize carotenoid pigments.

**a**, Diversity of carotenoid structures produced by molecular breeding of carotenoid biosynthetic genes from *Erwinia* sp.<sup>34</sup>. C<sub>40</sub> carotenoid biosynthesis branches into a variety of pathways to acyclic and cyclic carotenoids for which biosynthetic genes from bacteria have been cloned. Red arrows indicate how molecular breeding of the desaturase extended the central desaturation pathway to generate fully conjugated 3,4,3',4'-tetrahydrolycopene. Subsequent branching of this pathway by a member of a library of shuffled cyclase genes allowed synthesis of torulene, a carotenoid not made by the parent genes and not known in any bacteria. **b**, Coloured bacteria containing shuffled genes synthesize carotenoids not made by the *Erwinia* parents.



*E. coli*, together with a large library of gene hybrids created by shuffling two versions of a third *Erwinia* gene encoding a desaturase, which normally introduces double bonds into phytoene to make lycopene. Among the thousands of coloured bacterial progeny, they found some that were more yellow and pink than the orange *E. coli* containing the three naturally occurring carotenoid biosynthetic genes. Different members of the bacterial library made one or more of the carotenoids that contained double bonds at the various positions, and all the possible desaturation products were represented.

Combination of the genes that made the pink carotenoid (tetrahydrolycopene) with a new library of mated gene hybrids of a fourth (cyclase) gene generated an even greater variety of coloured bacteria: yellow, orange, pink and bright red (Fig. 3). The bright red cells produce torulene, a carotenoid not made by *Erwinia*, and not known in any bacteria but found in some red yeasts. Yet the pathway created by molecular breeding is not the same as that used by yeast to make torulene.

The combination of gene assembly (pathway engineering) and molecular evolution can solve very complex problems of biological design. By generating efficient pathways to make natural and non-natural products, it can greatly extend the applications of biocatalysis into the discovery and production of new biologically active compounds.

### De novo catalyst design

There are likely to be many problems for which natural molecules cannot even offer a suitable starting point for evolution. In some

cases, the whole enzyme frameworks are not suitable because, with many hundreds of amino acids, they are too unwieldy to produce or use in large quantities or, in the case of protein-based drugs, cannot be delivered efficiently to their targets. In other cases, the biological pathways are too cumbersome for practical use. For example, biological oxidation reactions are usually catalysed by large multiprotein complexes and use expensive cofactors that few would consider for an industrial process. In general, the many, sometimes conflicting demands and the contingent nature of evolution means that enzyme structures are not necessarily optimized as chemical reagents for a specific transformation, and there may be much better functional solutions that use completely different sequences and structures. How can one find them?

Some possible routes are evolutionary. Catalytic function can be coaxed out of protein frameworks evolved for different, non-catalytic roles. In one of the first evolutionary approaches to making new biocatalysts, catalytic antibodies or 'abzymes' were generated in response to molecules that mimic the transition state of a reaction<sup>35</sup>. But the development of commercially useful antibody catalysts has been hampered by their low expression, limited stability and generally low turnover rates, although there are a few notable exceptions<sup>36</sup>. The basic idea of targeting a transition-state analogue can be extended to generate catalytic activity from other, perhaps more tractable frameworks. But the activities of these new enzymes may still be low, reflecting the fact that transition-state binding is only one aspect of the catalytic process.

Other approaches to designing new protein catalysts use different breeding practices. There are many ways to create molecular diversity beyond point mutation and homologous recombination. Several groups are, for example, investigating nonhomologous recombination of distantly related<sup>37,38</sup> and even unrelated sequences<sup>39</sup> as a means to generate new functional proteins. Others are developing techniques to generate<sup>40,41</sup> and screen<sup>42</sup> larger libraries so as to be able to identify rarer but possibly more useful solutions.

But the number of possible protein sequences inevitably dwarfs any existing or even conceivable technology for searching it experimentally. So one must make intelligent choices about what and how to search. This may be where 'rational' design will be crucial: identifying the most likely places to search combinatorially for desired functions.

That rudimentary structure-based designs can be improved through evolutionary tuning is well accepted, if not yet widely practised. This blend of approaches was demonstrated by Altamirano and co-workers, who converted an  $\alpha/\beta$ -barrel enzyme with one activity (indole-3-glycerol phosphate synthase) into another with equally efficient activity (phosphoribosylanthranilate isomerase)<sup>43</sup>.

Conversely, structure-based computational methods can be used to identify likely sites for evolutionary improvement, thereby supporting the generation of specific 'targeted' libraries and greatly reducing the experimental search. Voigt *et al.* (ref. 44 and unpublished data) have used powerful computational methods<sup>17</sup> to search vast regions of sequence space to identify the most probable solutions to protein-design problems. They use the computational methods where they work best — solving the generic problems of identifying protein sites that are tolerant to mutation or that will tolerate crossover without significant disruption — and the evolutionary methods to find specific solutions within the generic ones.

The ideal would be to specify a catalyst *de novo*: purely from its primary sequence. In principle this should be possible, and indeed the first *de novo* proto-enzymes are now being reported<sup>45</sup> — although they are not particularly impressive catalysts. Primitive iron- and oxygen-binding sites introduced into the small protein thioredoxin by computational design show varying selectivities in oxidation processes<sup>45</sup>. Such designed sites might be adequate starting points for evolutionary methods.

## Conclusions

Biocatalysts need to become predictable and routine tools. At present they are neither, and biocatalyst design is still more of an art than a science. But things are changing. Laboratory evolution methods are now sufficiently robust that improved biocatalysts can be obtained with confidence on a reasonable timescale. Further developments, especially miniaturization and automation of high-throughput screening, will accelerate the acceptance and widespread application of biocatalysis.

Today, evolutionary methods seem the most fertile approach for developing new commercial biocatalysts. But the capabilities of rational design, particularly computational techniques and *de novo* design are expanding too. And emerging design methods that marry the best of the computational and the combinatorial approaches promise to make biocatalysis a key tool for synthetic chemistry in the century ahead. □

- Adam, M. W. W. & Kelly, R. M. Finding and using hyperthermophilic enzymes. *Trends Biotechnol.* **16**, 329–332 (1998).
- Rondon, M. R., Goodman, R. M. & Handelsman, J. The Earth's bounty: assessing and accessing soil microbial diversity. *Trends Biotechnol.* **17**, 403–409 (1999).
- Short, J. Recombinant approaches for accessing biodiversity. *Nature Biotechnol.* **15**, 1322–1323 (1997).
- Bull, A. T., Ward, A. C. & Goodfellow, M. Search and discovery strategies for biotechnology: the paradigm shift. *Microbiol. Mol. Biol. Rev.* **64**, 573–606 (2000).
- Black, M. E., Newcomb, T. G., Wilson, H.-M. P. & Loeb, L. A. Creation of drug-specific Herpes Simplex Virus type 1 thymidine kinase mutants for gene therapy. *Proc. Natl Acad. Sci. USA* **93**, 3525–3529 (1996).
- Encell, L. P., Coates, M. M. & Loeb, L. A. Engineering human DNA alkyltransferases for gene therapy using random sequence mutagenesis. *Cancer Res.* **58**, 1013–1020 (1998).

- Matsumura, I., Wallingford, J. B., Surana, N. K., Vize, P. D. & Ellington, A. D. Directed evolution of the surface chemistry of the reporter enzyme  $\beta$ -glucuronidase. *Nature Biotechnol.* **17**, 696–701 (1999).
- Lanio, T., Jeltsch, A. & Pingoud, A. Towards the design of rare cutting restriction endonucleases: using directed evolution to generate variants of *EcoRV* differing in their substrate specificity by two orders of magnitude. *J. Mol. Biol.* **283**, 59–69 (1998).
- Liu, D. R., Magliery, T. J., Pastrnak, M. & Schultz, P. G. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*. *Proc. Natl Acad. Sci. USA* **94**, 10092–10097 (1997).
- Babbitt, P. New functions from old scaffolds: how nature re-engineers enzymes for new functions. *Adv. Protein Chem.* **55**, 1–28 (2000).
- Xiang, H., Luo, L., Taylor, K. L. & Dunaway-Mariano, D. Interchange of catalytic activity within the 2-enoyl-coenzyme A hydratase/isomerase superfamily based on a common active site template. *Biochemistry* **38**, 7638–7652 (1999).
- Broun, P., Shanklin, J., Whittle, E. & Somerville, C. Catalytic plasticity of fatty acid modification enzymes underlying chemical diversity of plant lipids. *Science* **282**, 1315–1317 (1998).
- Shanklin, J. Exploring the possibilities presented by protein engineering. *Curr. Opin. Plant Biol.* **3**, 243–248 (2000).
- Shanklin, J. & Cahoon, E. B. Desaturation and related modifications of fatty acids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 611–641 (1998).
- Wintropde, P. & Arnold, F. H. Temperature adaptation of enzymes: lessons from laboratory evolution. *Adv. Protein Chem.* **55**, 161–226 (2000).
- Jaenicke, R. Stability and stabilization of globular proteins in solution. *J. Biotechnol.* **79**, 193–203 (2000).
- Dahiyat, B. I. & Mayo, S. L. De novo protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
- Malakaukas, S. M. & Mayo, S. L. Design, structure and stability of a hyperthermophilic protein variant. *Nature Struct. Biol.* **5**, 470–475 (1998).
- Cedrone, E., Ménez, A. & Quéméneur, E. Tailoring new enzyme functions by rational redesign. *Curr. Opin. Struct. Biol.* **10**, 405–410 (2000).
- Arnold, F. H. (ed.) *Evolutionary Protein Design* (*Adv. Protein Chem.* **55**) (Academic, San Diego, 2000).
- Affholter, J. & Arnold, F. H. Engineering a revolution. *Chem. Britain* **35**, 48–51 (1999).
- Ness, J. E., del Cardayre, S. B., Minshull, J. & Stemmer, W. P. C. Molecular breeding—the natural approach to enzyme design. *Adv. Protein Chem.* **55**, 261–292 (2000).
- Reetz, M. T. & Jaeger, K.-E. Superior biocatalysts by directed evolution. *Top. Curr. Chem.* **200**, 31–57 (1999).
- Liebeton, K. *et al.* Directed evolution of an enantioselective lipase. *Chem. Biol.* **7**, 709–718 (2000).
- May, O., Nguyen, P. & Arnold, F. H. Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. *Nature Biotechnol.* **18**, 317–320 (2000).
- Arnold, F. H. Directed enzyme evolution <<http://www.che.caltech.edu:80/groups/fha/Enzyme/directed.html>>
- Oue, S., Okamoto, A., Yano, T. & Kagamiyama, H. Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues. *Biol. Chem.* **274**, 2344–2349 (1999).
- Yano, T., Oue, S. & Kagamiyama, H. Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl Acad. Sci. USA* **95**, 5511–5515 (1998).
- Orencia, M. C. & Stevens, R. C. Structural changes in directed and immune evolved enzymes. *Adv. Protein Chem.* **55**, 227–260 (2000).
- Stemmer, W. P. C. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389–391 (1994).
- Cramer, A., Raillard, S.-A., Bermudez, E. & Stemmer, W. P. C. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998).
- Sandmann, G., Albrecht, M., Schnurr, G., Knörzer, O. & Böger, P. The biotechnological potential and design of novel carotenoids by gene combination in *Escherichia coli*. *Trends Biotechnol.* **17**, 233–237 (2000).
- Cane, D. E., Walsh, C. T. & Khosla, C. Harnessing the biosynthetic code: combinations, permutations and mutations. *Science* **282**, 63–68 (1998).
- Schmidt-Dannert, C., Umeno, D. & Arnold, F. H. Molecular breeding of carotenoid biosynthetic pathways. *Nature Biotechnol.* **18**, 750–753 (2000).
- Patten, P. A. *et al.* The immunological evolution of catalysis. *Science* **271**, 1086–1091 (1996).
- Barbas, C. F. III, List, B., Rader, C., Segal, D. J. & Turner, J. M. From catalytic asymmetric synthesis to the transcriptional regulation of genes: *in vivo* and *in vitro* evolution of proteins. *Adv. Protein Chem.* **55**, 317–366 (2000).
- Ostermeier, M. & Benkovic, S. J. Evolution of protein function by domain swapping. *Adv. Protein Chem.* **55**, 29–78 (2000).
- Sieber, V., Martinez, C. A. & Arnold, F. H. Libraries of hybrid proteins from distantly-related sequences. *Nature Biotechnol.* (in the press).
- Riechmann, L. & Winter, G. Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc. Natl Acad. Sci. USA* **97**, 10068–10073 (2000).
- Plückthun, A., Schaffitzel, C., Hanes, J. & Jermutus, L. In vitro selection and evolution of proteins. *Adv. Protein Chem.* **55**, 367–404 (2000).
- Roberts, R. W. Totally *in vitro* protein selection using mRNA-protein fusions and ribosome display. *Curr. Opin. Chem. Biol.* **3**, 268–273 (1999).
- Olsen, M., Iverson, B. & Georgiou, G. High-throughput screening of enzyme libraries. *Curr. Opin. Biotechnol.* **11**, 331–337 (2000).
- Altamirano, M. M., Blackburn, J. M., Aguayo, C. & Fersht, A. R. Directed evolution of new catalytic activity using the  $\alpha/\beta$ -barrel scaffold. *Nature* **403**, 617–622 (2000).
- Voigt, C. A., Mayo, S. L., Arnold, F. H. & Wang, Z.-G. Computational method to reduce the search for directed protein evolution. *Proc. Natl Acad. Sci. USA* (in the press).
- Benson, D. E., Wisz, M. S. & Hellinga, H. W. Rational design of nascent metalloenzymes. *Proc. Natl Acad. Sci. USA* **97**, 6292–6297 (2000).

## Acknowledgements

I thank the many talented students and postdocs who have contributed to the development of new biocatalyst engineering tools in my laboratory, and the following organizations for their financial support: the US Office of Naval Research, the US National Science Foundation, the Army Research Office, Maxygen, Inc., The Biotechnology Research & Development Corporation, British Petroleum, Degussa AG and Procter & Gamble Co. I also thank C. Voigt and J. Shanklin for thoughtful comments, and J. Shanklin for Fig. 1.