Directed enzyme evolution

Edgardo T Farinas*, Thomas Bulter* and Frances H Arnold†

Laboratory evolutionists continue to generate better enzymes for industrial and research applications. Exciting developments include new biocatalysts for enantioselective carbon–carbon bond formation and fatty acid production in plants. Creative contributions to the repertoire of evolutionary methods will ensure further growth in applications and expand the scope and complexity of biological design problems that can be addressed. Researchers are also starting to elucidate mechanisms of enzyme adaptation and natural evolution by testing evolutionary scenarios in the laboratory.

Addresses
Division of Chemistry and Chemical Engineering 210-41, California Institute of Technology, Pasadena, CA 91125, USA;
†e-mail: frances@cheme.caltech.edu
*These authors contributed equally to this work.

Current Opinion in Biotechnology 2001, 12:545–551

0958-1669/01/$ – see front matter

Abbreviations

CSR	compartmentalized self-replication
FACS	fluorescence-activated cell sorting
FRET	fluorescence resonance energy transfer
GAO	galactose oxidase
HRP	horseradish peroxidase
KDPG	D-2-keto-3-deoxy-6-phosphogluconate
PCR	polymerase chain reaction
RACHITT	random chimeragenesis on transient templates
SHIPREC	sequence homology independent protein recombination

Introduction

Enzymes can perform intricate regioselective and/or enantioselective chemical transformations and can accelerate reaction rates by enormous factors (up to 10^{12}), all under mild conditions [1]. Although highly attractive for chemical synthesis, enzymes almost always present problems for use on an industrial scale [2,3]. Limitations include sluggish catalysis on nonnatural substrates, low stability or little tolerance for changes in operating parameters, poor activity in nonaqueous media, and requirements for expensive cofactors. The reliable and quick identification of amino acid substitutions that generate desired changes in enzyme performance remains the ultimate goal of protein engineering research.

Evolutionary design approaches have enjoyed considerable success in recent years [4••]. Numerous enzymes have been improved by directing their evolution in the laboratory, which usually involves iterations of random mutagenesis or recombination followed by screening or selection [5]. This review covers recent progress in directed enzyme engineering applications and methods.

Evolved biocatalysts

Carbon–carbon bond formation

D-2-Keto-3-deoxy-6-phosphogluconate (KDPG) aldolase from *Escherichia coli* provides synthetic routes for asymmetric carbon–carbon bond formation. The utility of this enzyme is limited, however, by its high specificity for its natural substrates: KDPG in the cleavage reaction and D-glyceraldehyde-3-phosphate in the addition reaction. Fong *et al.* [6•] evolved the aldolase to accept the nonphosphorylated analog of KDPG and L-glyceraldehyde. Wymer *et al.* [7] also performed directed evolution on KDPG aldolase, broadening its substrate range while retaining enantioselectivity. The crystal structure of the evolved enzyme shows that directed evolution actually relocated the key catalytic lysine residue to another position. Together, these studies open important new opportunities for using aldolases in stereoselective carbon–carbon bond formation.

Oxidation

The need for functional expression in a host suitable for making mutant libraries (usually *Saccharomyces cerevisiae* or bacteria) is a serious bottleneck that has kept several enzymes out of the evolutionist's reach. Until recently, the widely used reporter enzyme horseradish peroxidase (HRP) was one such example. Morawski *et al.* were able to improve functional expression [8] and stability [9] of HRP in *S. cerevisiae* by random mutagenesis, recombination and screening for total HRP activity. The thermostable mutant is also more stable in the presence of H_2O_2 and more active with substrates frequently used in diagnostic applications.

Sun *et al.* [10] expressed the fungal enzyme galactose oxidase (GAO) in *E. coli* and increased the total GAO activity 60-fold by random mutagenesis and screening. Slightly more thermostable than the wild-type enzyme, this mutant is a good starting point for functional improvement by directed evolution. GAO is an interesting catalyst with potential applications in synthesis, biosensors and the labeling of glycoconjugate structures. Delagrave *et al.* [11] report the optimization of GAO for yet another application, selective oxidation of the natural polymer guar, for use in paper manufacturing. Random mutagenesis by error-prone polymerase chain reaction (PCR) and screening on a surrogate substrate, methylgalactose, using solid-phase digital imaging, increased total activity up to 16-fold. Screening the library using a guar-plate assay did not generate mutants that were more active than the ones found using the surrogate substrate. Characterization of the evolved GAO expressed in *Pichia pastoris* showed an approximately threefold decrease in K_m on methylgalactose.

Directed evolution is also being use to expand the substrate ranges of monooxygenases and dioxygenases. Cytochrome P450 BM-3 hydroxylates fatty acids with chain lengths of between C12 and C18. Farinas *et al.* [12] report that P450 BM-3 also slowly oxidizes octane (C8) and that directed evolution can improve this alkane hydroxylation activity. Dioxygenases are useful for the production of

chiral dihydrodiols from a range of aromatic substrates. Sakamoto *et al.* [13] used random and saturation mutagenesis to increase the activity of toluene dioxygenase from *Pseudomonas putida* on a more polar heterocyclic substrate, 4-picoline (see also Update).

Metabolic engineering

The discovery of new natural products continues to be of great interest for the production of pharmaceuticals, new materials, agrochemicals, and consumer products. An exciting approach to generating new natural products is to evolve biosynthetic pathways in microorganisms. For example, torulene was created for the first time in *E. coli* by the recombination of genes encoding carotenoid biosynthetic enzymes followed by screening for novel compounds [14,15]. Evolutionary methods can also be applied to improving production of known compounds. Wang *et al.* [16] previously introduced the biosynthetic genes for enhanced production of isoprenoids into *E. coli*. Geranylgeranyl diphosphate synthase, found to be rate-limiting in the isoprenoid biosynthetic pathway, was targeted for directed evolution [17]. Clones with increased lycopene production of up to 100% were identified by visual screening for orange-colored colonies. One higher-producing variant had mutations only upstream from the coding region, and the enhanced lycopene synthesis was due to increased enzyme expression. Shuffling the best eight variants did not yield further improvements in lycopene production, which was attributed to the limited dynamic range of the visual screen at these high production levels (see also Update).

Metabolically engineered plants are also highly attractive for the production of modified natural products. Plant biosynthetic enzymes can be evolved conveniently in microorganisms and transferred to plants for production. Cahoon and Shanklin [18••] used this strategy to tailor fatty acid production in *Arabidopsis fab1* plants by engineering the acyl-acyl carrier protein desaturase that desaturates C18 fatty acids. Mutagenesis and selection in *E. coli* yielded variants with greater specific activity for shorter-chain C16 fatty acids. One mutant showed a 25-fold preference for C16 chains as compared with the 70-fold preference of the wild type for longer chains. Expression in the plants resulted in an accumulation of unusual monounsaturated fatty acids, which formed more than 25% of the seed oil.

Enzyme adaptation

This review period has seen several interesting applications of laboratory evolution methods to address fundamental questions of enzyme function and evolution. In the laboratory, a particular evolutionary scenario can be simulated by imposing specific selective pressures. In contrast to natural evolution, however, all the intermediates along the evolutionary trajectory are available for study to identify mechanisms of adaptation. The enzymes β-galactosidase and β-glucuronidase from *E. coli* evolved from a common ancestor. Matsumura and Ellington [19•] sought to identify the structural determinants of their substrate specificities by evolving β-glucuronidase in the laboratory to hydrolyze β-galactosides. Error-prone PCR and DNA shuffling generated a variant that hydrolyzed *para*-nitrophenyl-β-D-galactoside 500 times better than wild type and with a strong preference for this substrate over β-glucuronides. (The evolved β-galactosidase activity was still 10 000-fold less active than the natural β-galactosidase.) According to the patchwork hypothesis, modern enzymes evolved by gene duplication and diversification from ancient enzymes with broad substrate specificity [20]. Intermediate variants of β-glucuronidase showed broadened substrate specificity, and the authors propose that adaptation to new substrates was achieved by passing through a functional intermediate similar to their ancestral states [19•].

Jurgens *et al.* [21•] used directed evolution to create an enzyme that might resemble the common ancestor of two isomerases that catalyze similar reactions during the biosynthesis of histidine and tryptophan. Their laboratory-evolved HisA (*N*′-[(5′phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase) mutant has low but equal activity in both pathways, which they claim to support a patchwork hypothesis. But a single amino acid exchange in wild-type HisA isomerase was also shown to change the substrate specificity from one reaction pathway to the other. This switch of specificity without a generalist intermediate demonstrates that a specialized enzyme can also evolve directly from another specialist (see also Update).

Directed evolution can anticipate natural evolution. The development of β-lactam antibiotic resistance in microorganisms is a product of the rapid evolution of β-lactamase to catabolyze the drugs, and laboratory evolution experiments [22,23] have generated a mutant that was subsequently found in clinical isolates [22,23,24•]. In recent work, Orencia *et al.* [24•] combined directed evolution and structural analysis to elucidate the mechanism of β-lactam antibiotic resistance and its development by stepwise mutations. Using an *E. coli* mutator strain and selection on cefotaxime, they generated the same β-lactamase triple mutant (Glu104Lys, Gly238Ser, Met182Thr) whose emergence had been predicted by previous studies using high mutation rates [22] and DNA shuffling [23]. The crystal structure of the evolved β-lactamase revealed that the first two mutations improved access to the active site, whereas the third mutation (Met182Thr) compensates for the deleterious side-effects of the first ones. Understanding the emergence of antibiotic resistance will allow the prediction of mutations in the clinic and should aid in the development of new antibiotics.

Another area where evolutionary experiments have been able to shed light on ancient processes and adaptive mechanisms is adaptation to different temperatures [25–27]. Enzymes are readily evolved in the laboratory for higher

thermostability and high-temperature activity [28•,29]; activity at low temperatures can also be improved. Because relatively few amino acid substitutions (1–5% of the sequence) can convert a mesophilic enzyme into its thermophilic or psychrophilic counterpart, laboratory-evolved enzymes provide a superb, almost noise-free dataset for comparative studies to uncover adaptive mechanisms. Also of great interest is the fact that thermostability can be improved at no cost to activity at low temperature and vice versa. Naturally occurring homologous enzymes adapted to different temperature niches often show trade-offs in activity and stability, which are thought to reflect conflicting demands of enzyme flexibility. The directed evolution studies clearly show that any trade-offs are much more likely to reflect a lack of selective pressure on both properties and genetic drift [26].

The room temperature activity of a hyperthermophilic β-glucosidase from *Pyrococcus furiosis* (CelB) was improved by directed evolution [30]. After error-prone PCR, DNA shuffling, and screening on *p*-nitrophenol-β-D-glucopyranoside, mutants with up to three times more activity at 20°C than wild type were identified. 3-Isopropylmalate dehydrogenase from *Thermus thermophilus* (LeuB) was also adapted to lower temperatures [31•]. To select for enzyme function, Suzuki *et al.* [31•] replaced the *leuB* locus of *E. coli* with the genes coding for *T. thermophilus* LeuB and kanamycin nucleotidyl transferase. The resulting *E. coli* strain only grew at 42°C and above in the absence of leucine, because the dehydrogenase was inactive at lower temperatures. The strain was treated with a chemical mutagen and screened on leucine-free plates containing kanamycin at 40°C. After three days, approximately 100 colonies appeared on the plates. Sequencing *leuB* from five randomly picked colonies revealed 0–3 base substitutions per gene. One gene had no mutations at all, illustrating the potential hazards of using a growth-based selection and mutagenesis that is not limited to the gene of interest, but which instead targets the entire genome. The remaining four mutant enzymes showed increases of up to 4.1-fold in specific activity at 40°C. Two had melting temperatures similar to wild type, demonstrating that activity at lower temperatures can be acquired without sacrificing the high thermostability of the thermophilic enzyme.

Several evolution experiments have increased enzyme thermostability. A psychrophilic subtilisin from Antarctic *Bacillus* TA41 (S41) was converted into its thermophilic counterpart over eight generations [28•,32]. The final variant was as stable as a thermophilic subtilisin while retaining, and even improving, low-temperature activity towards a peptide substrate. Sequencing revealed 13 amino acid substitutions, some of which are believed to stabilize a surface loop and increase calcium affinity. Notably, six of the 13 were substitutions of serine by another amino acid. A reduced occurrence of serine in thermophilic enzymes was observed when sequences and structures were compared with their naturally occurring mesophilic homologs [33,34].

Prolyl endopeptidase from *Flavobacterium meningosepticum* cleaves specifically at proline residues; no other endopeptidase has shown this selectivity. The low stability of prolyl endopeptidase compared with other peptidases such as subtilisin and trypsin made it a target for laboratory evolution [29]. The best mutant after three cycles of random mutagenesis and screening in *E. coli* increased its half-life 60-fold at 60°C. Furthermore, the thermostable mutant retained 50% of its activity after 3 h in 50% glycerol (necessary for *in vitro* amidation), whereas the wild-type enzyme rapidly lost activity.

New methods for directed evolution

Although widely used for improving biocatalysts, directed evolution is still an emerging technology. Methods for mutagenesis and recombination are constantly developing. Identifying improved clones from a vast pool of mutants is only feasible with a sensitive screen or selection. Screening methods are of great importance, but have been reviewed recently [35–37] and will not be covered here.

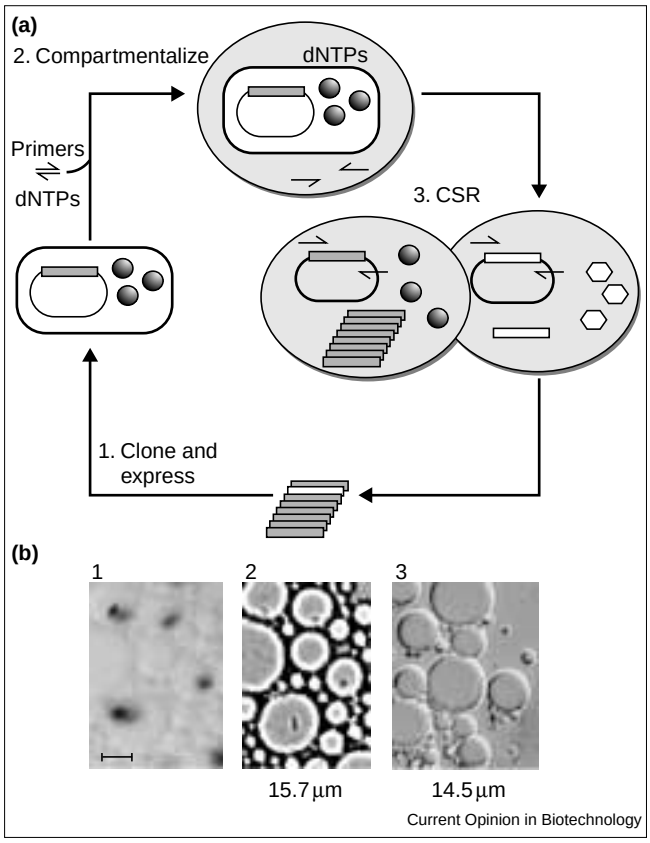
An elegant strategy called CSR (compartmentalized self-replication) has been demonstrated for directed evolution of enzymes involved in DNA replication (e.g. polymerases) [38••] (Figure 1). *E. coli* cells were used to deliver a library of *Taq* DNA polymerases into separate aqueous compartments, formed from a water-in-oil emulsion [39], which also contained flanking PCR primers and deoxynucleotide triphosphates. PCR of these samples amplified mutant *Taq* DNA polymerase genes, depending on the activity of their gene product. CSR increased the thermostability of *Taq* DNA polymerase 11-fold and enhanced resistance to heparin 130-fold, without reducing catalytic activity. CSR could be extended to the evolution of other enzymes that can couple to the gene replication reaction, for example, by providing substrates required for replication of their own genes [38••].

Cell-surface and phage-display methods have been limited mostly to evolution of binding rather than catalysis. Olsen *et al.* [40•] devised a strategy to screen large enzyme libraries displayed on bacterial cell surfaces by fluorescence-activated cell sorting (FACS). The *E. coli* cell-surface protease OmpT was evolved for novel substrate specificity using a specially designed fluorescence resonance energy transfer (FRET) substrate that allows the genotype to remain linked to the phenotype by association with the cell surface. The peptide substrate contained a fluorophore, the scissile bond, a quenching fluorophore, and an overall +3 charge that retains the fluorescent product on the cell surface. This FRET/FACS system nicely exploits the high-throughput of FACS screening for catalysis; however, limitations on surface expression of enzymes and constraints on substrate design make this approach somewhat specialized.

New methods to create enzyme libraries

The method used to create enzyme diversity determines the quality of the mutant library and strongly influences the

Figure 1



Evolution of polymerases by compartmentalized self-replication (CSR) [38••]. (a) *E. coli* cells expressing a library of polymerases are separated into heat stable, aqueous compartments of an oil-in-water emulsion together with flanking primers and deoxynucleotide triphosphates (dNTPs) (steps 1 and 2). Thermocycling lyses the cells and empties them into the individual compartments where the polymerase variants amplify their coding genes (step 3). Genes amplified by the most active polymerase variants are selectively enriched, isolated from the emulsion and subjected to the next CSR cycle. (b) Panels 1, 2 and 3 illustrate the aqueous compartments of an oil-in-water emulsion. Scale bar 10 µm; the size of the compartments is indicated below. (The figure was reproduced from [38••] with permission 2001 National Academy of Sciences USA.)

efficiency of the evolutionary search. Methods for gene diversification are far from perfect. *In vitro* recombination by DNA shuffling [41] creates libraries of chimeric genes from any number of parent sequences, including naturally occurring homologous genes. A recently published method, RACHITT (random chimeragenesis on transient templates) [42•], claims to overcome some limitations of previous shuffling protocols. In RACHITT one of the parent genes is used as a transient scaffold to which complementary fragments of the other parents anneal. Unhybridized fragment termini are trimmed and gaps are filled enzymatically. These trimming reactions have the potential to reduce the size of the shuffled fragments and increase the number of recombinations per gene. The transient single-stranded scaffold is expected to avoid amplification of parent sequences. However, the choice of the scaffold will probably influence the outcome of the experiment, as sequences that are similar to the scaffold will be incorporated more frequently into the chimeras. When used to shuffle two bacterial monooxygenases, RACHITT generated an average of 14 crossovers per gene, which is higher than that reported for other *in vitro* recombination methods. However, the benefit of such a high crossover frequency remains to be demonstrated. Coco *et al.* [42•] found 29% of their crossovers in regions of 10 or fewer bases of sequence identity; however, DNA shuffling of dioxygenases using Stemmer's approach [41] also generated a similar frequency (32%) of crossovers in regions of 10 or fewer bases (J Joern, FH Arnold, personal communication).

Another recombination method, CLERY (combinatorial libraries enhanced by recombination in yeast), combines DNA shuffling *in vitro* and *in vivo* shuffling in yeast [43•]. The procedure was used to shuffle two human cytochrome P450 enzymes. The average number of crossovers was around 4.4 and the library contained only 14% nonchimeric (parental) sequences. Requiring two hosts, this procedure is most suitable for shuffling eukaryotic genes that cannot be expressed in bacteria. A high mutation rate associated with the *in vitro* shuffling reduced the fraction of functional clones (12%). This paper also describes a useful probe hybridization method for characterizing large numbers of sequences from the chimeric library.

All the available methods for DNA shuffling require high sequence similarity for recombination (> ~60%). There is some interest, however, in shuffling parent sequences of lower identity or even with no discernible sequence homology [44,45]. During this period, Sieber *et al.* [46•] introduced SHIPREC (sequence homology independent protein recombination). To maximize the abundance of functional hybrids in a library made by recombining two homologous genes of low sequence identity, the chimeras are selected by size to retain the length of the parent genes. SHIPREC therefore combines the N-terminal fragment of the parent protein with the appropriate C-terminal fragment and, in principle, makes all such possible combinations. With a chloramphenicol acetyl transferase preselection for properly folded and soluble variants, SHIPREC was able to generate hybrids that combined the activity of a membrane-bound human cytochrome P450 enzyme with the increased solubility of a bacterial cytochrome P450. The two P450s share only 16% amino acid identity. The method as described is limited to creating libraries containing only one crossover between two parent genes.

It is proposed that proteins evolved naturally by the assembly of nonhomologous genes. Protein domains might also have developed by the exchange and assembly of gene segments [47]. One natural process that leads to nonhomologous recombination is exon shuffling. The *in vitro* recombination of domain-encoding exons might prove useful for creating enzymes with novel functions [48], although this approach would be limited to modular proteins which contribute only a small subset to the diversity of enzymes.

One shortcoming of *in vitro* mutagenesis is the limited size of the variant libraries created ($<10^9$ and usually considerably smaller). Fabret *et al.* [49•] report an *in vivo* mutagenesis method that is capable of generating large collections of mutants (10^{12}–10^{13} per litre of cell culture). In contrast to mutator strains commonly used, the *E. coli* strain employed in this approach is genetically stable. The target gene is propagated in a ColE1-type plasmid in which DNA polymerase I is essential for the initiation of replication. After replication of the initial 1 kb of the plasmid insertion, polymerase I is replaced by polymerase III, which completes the replication. The *E. coli* strain expresses a mutant polymerase I with negligible proof-reading activity, and the mismatch repair system is disabled. As polymerase I replicates only 1% of the host chromosome, the target gene was mutagenized without changing the cell phenotype. This makes it possible to cultivate the cells for a longer time, so that ~10% of the cells in the culture are mutated after 150 generations. The claim that the method can generate libraries of 10^{12} mutants is overstated, because the library diversity is limited by the low mutation rate and by the restriction that only 1 kb can be mutated. Achieving this with the reported system does not seem practical owing to the low error rate. Also because of the low mutation rate, this approach is suitable for selections but not screening (see also Update).

Computational approaches to evolution

Screening is usually the most labor-intensive part of the evolution experiment, and there is considerable value in finding ways to use structural or mechanistic information to increase the efficiency of the search for improved mutants. This will reduce the cost of obtaining a given functional improvement or, even better, lead to much greater improvements for the same amount of effort. Structural information is therefore often used to choose regions or even individual amino acids for mutagenesis. The risk, of course, is that targeting specific sites for mutation is based on limited understanding of the molecular basis for the desired function and often fails. Directed evolution experiments have shown time and time again that beneficial mutations often occur where least expected.

A new, computational approach to the problem uses powerful protein design software to identify sites on proteins that are most likely to accept mutations. Voigt *et al.* [50•] observed that sites calculated to be highly tolerant to amino acid substitution are very often the sites where beneficial mutations are found during directed evolution. If one can identify a limited number of mutation sites computationally, simultaneous saturation mutagenesis becomes feasible. This has the benefits that all possible amino acid substitutions can be accessed (rather than the 5.7 on average that are available by point-mutation [51]) and that multiple synergistic mutations can be discovered. This method requires a crystal structure or a good homology model, but does not make *a priori* decisions as to where beneficial mutations will appear (e.g. in a substrate recognition site).

Conclusions

Over the past decade, different implementations of directed evolution have improved the specific properties of a wide variety of proteins [4••], offering a robust means to transfer enzymes out of the laboratory and into industrial processes. More recently, these same approaches have provided insights into past as well as future evolutionary events. Directed evolution is merging molecular biology, chemistry and engineering to forge many opportunities for further growth.

Update

In recent work, Suenaga *et al.* [52] shuffled the biphenyl dioxygenase genes from *Pseudomonas pseudoalcaligenes* and *Burkholderia cepacia*. *E. coli* expressing biphenyl dioxygenase chimeric genes showed increased activity for monocyclic aromatic compounds (benzene, toluene and alkylbenzenes) that are poor substrates for the wild-type enzymes. Wang and Liao [53] used directed evolution to alter the product specificity of phytoene desaturase from *Rhodobacter sphaeroides* from neurosporin to lycopene. The gene was expressed in *E. coli* together with another plasmid containing the necessary isoprenoid pathway genes. Two rounds of directed evolution and visual selection of the colored colonies followed by recombination of the mutations by site-directed mutagenesis resulted in a mutant that produced 90% lycopene.

Raillard *et al.* [54•] recently published their *tour de force* study of an enzyme library made by shuffling two very closely related enzymes involved in s-triazine hydrolysis. The resulting library exhibited considerable functional diversity, including enzymes that hydrolyze substrates not accepted by either parent. 'Reverse engineering' the functional enzymes provided useful and comprehensive information on sequence–function relationships.

Using saturation mutagenesis at every site in the protein, Gray *et al.* [55•] dramatically increased the thermostability of an haloalkane dehalogenase from *Rhodococcus rhodochrous*, without compromising its activity. The enzyme also became fully reversible in its denaturation. Eight single-site mutations were combined by site-directed mutagenesis. The variant was immobilized on alumina, and the enzymatic conversion of 1,2,3-trichloropropane to 2,3-dichloropropanol in a bioreactor was investigated at 55°C. The operational productivity of the mutant (lb product/ lb of enzyme) was 25-fold greater and the half-life was 18 times longer compared with wild type.

Acknowledgements

This work was supported by fellowships from the US National Science Foundation DBI-9807460 (ETF) and the Deutsche Forschungsgemeinschaft (TB).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Faber K: *Biotransformations in Organic Chemistry*. New York: Springer-Verlag; 2000.

2. Schmid A, Dordick JS, Hauer B, Kiener A, Wubbolts M, Witholt B: Industrial biocatalysis today and tomorrow. *Nature* 2001, 409:258-268.

3. Arnold FH: Combinatorial and computational challenges for biocatalyst design. *Nature* 2001, 409:253-257.

4. •• Arnold FH (Ed): *Evolutionary Protein Design*. New York: Academic Press; 2001. [Advances in Protein Chemistry vol. 55.]
An excellent compendium of evolutionary protein design approaches, with contributions from top evolutionary engineering laboratories world-wide.

5. Sutherland JD: Evolutionary optimisation of enzymes. *Curr Opin Chem Biol* 2000, 4:263-269.

6. • Fong S, Machajewski TD, Mak CC, Wong C: Directed evolution of D-2-keto-3-deoxy-6-phosphogluconate aldolase to new variants for the efficient synthesis of D- and L-sugars. *Chem Biol* 2000, 7:873-883.
An aldolase was evolved for enantioselective carbon–carbon bond formation. The enzyme provides new routes for the enzymatic synthesis of D- and L-sugars.

7. Wymer N, Buchanan LV, Henderson D, Mehta N, Botting CH, Pocivavsek L, Fierke CA, Toone EJ, Naismith JH: Directed evolution of a new catalytic site in 2-keto-3-deoxy-6-phosphogluconate aldolase in ***Escherichia coli***. *Structure* 2001, 9:1-9.

8. Morawski B, Lin Z, Cirino P, Joo H, Bandara G, Arnold FH: Functional expression of horseradish peroxidase in ***Saccharomyces cerevisiae*** and ***Pichia pastoris***. *Protein Eng* 2000, 13:377-384.

9. Morawski B, Quan S, Arnold FH: Functional expression and stabilization of horseradish peroxidase by directed evolution in ***Saccharomyces cerevisiae***. *Biotechnol Bioeng* 2001, 76:99-107.

10. Sun L, Petrounia IP, Yagasaki M, Arnold FH: Expression and stabilization of galactose oxidase in ***Escherichia coli*** by directed evolution. *Protein Eng* 2001, in press.

11. Delagrave S, Murphy DJ, Pruss JL, Maffia AM, Marrs BL, Bylina EJ, Coleman WJ, Grek CL, Dilworth MR, Yang MM *et al.*: Application of a very high-throughput digital imaging screen to evolve the enzyme galactose oxidase. *Protein Eng* 2001, 14:261-267.

12. Farinas ET, Schwaneberg U, Glieder A, Arnold FH: Directed evolution of a cytochrome P450 monooxygenase for alkane oxidation. *Adv Synth Catal* 2001, 343:601-606.

13. Sakamoto T, Joern J, Arisawa A, Arnold FH: Laboratory evolution of toluene dioxygenase to accept 4-picoline. *Appl Environ Microbiol* 2001, in press.

14. Schmidt-Dannert C, Umeno D, Arnold FH: Molecular breeding of carotenoid biosynthetic pathways. *Nat Biotechnol* 2000, 18:750-753.

15. Schmidt-Dannert C: Engineering novel carotenoids in microorganisms. *Curr Opin Biotechnol* 2000, 11:255-261.

16. Wang CW, Oh M-K, Liao JC: Engineered isoprenoid pathway enhances astaxanthin production in ***Escherichia coli.*** *Biotechnol Bioeng* 1999, 62:235-241.

17. Wang C, Oh M, Liao JC: Directed evolution of metabolically engineered ***Escherichia coli*** for carotenoid production. *Biotechnol Prog* 2000, 16:922-926.

18. •• Cahoon EB, Shanklin J: Substrate-dependent mutant complementation to select fatty acid desaturase variants for metabolic engineering of plant seed oils. *Proc Natl Acad Sci USA* 2000, 97:12350-12355.
A plant enzyme involved in the biosynthesis of unsaturated fatty acids was evolved in *E. coli* with the aid of a clever selection method. Expression in plants led to the production of unusual unsaturated fatty acids.

19. • Matsumura I, Ellington AD: ***In vitro*** evolution of β-glucuronidase into a β-galactosidase proceeds through non-specific intermediates. *J Mol Biol* 2001, 305:331-339.
Directed evolution is used to probe evolutionary pathways for changing enzyme specificity.

20. Jensen RA: Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* 1976, 30:409-425.

21. • Jurgens C, Strom A, Wegener D, Hettwer S, Wilmanns M, Sterner R: Directed evolution of a $(\beta\alpha)_8$-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci USA* 2000, 97:9925-9930.
This study explores natural enzyme evolution by attempting to recreate a common ancestral phenotype in the laboratory. A single point-mutation is sufficient to switch substrate specificity.

22. Zaccolo M, Gherardi E: The effect of high-frequency random mutagenesis on ***in vitro*** protein evolution: a study on TEM-1 β-lactamase. *J Mol Biol* 1999, 285:775-783.

23. Stemmer WPC: Rapid evolution of a protein ***in vitro*** by DNA shuffling. *Nature* 1994, 370:389-391.

24. • Orencia MC, Yoon JS, Ness JE, Stemmer WP, Stevens RC: Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat Struct Biol* 2001, 8:238-242.
The emergence of β-lactamase mutants that result in antibiotic resistance can be anticipated by directed evolution and structural analysis.

25. Wintrode PL, Arnold FH: Temperature adaptation of enzymes: lessons from laboratory evolution. *Adv Protein Chem* 2000, 55:161-225.

26. Arnold FH, Wintrode PL, Miyazaki K, Gershenson A: How enzymes adapt: lessons from directed evolution. *Trends Biochem Sci* 2001, 26:100-106.

27. Sterner R, Liebl W: Thermophilic adaptation of proteins. *Crit Rev Biochem Mol Biol* 2001, 36:39-106.

28. • Miyazaki K, Wintrode PL, Grayling RA, Rubingh DN, Arnold FH: Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J Mol Biol* 2000, 297:1015-1026.
Directed evolution converted a poorly stable, psychrophilic protease into a highly thermostable one without compromising low-temperature activity.

29. Uchiyama H, Inaoka T, Ohkuma-Soyejima T, Togame H, Shibanaka Y, Yoshimoto T, Kokubo T: Directed evolution to improve the thermostability of prolyl endopeptidase. *J Biochem (Tokyo)* 2000, 128:441-447.

30. Lebbink JHG, Kaper T, Kengen SWM, van der Oost J, de Vos WM: β-Glucosidase CelB from ***Pyrococcus furiosus***: production by ***Escherichia coli***, purification, and ***in vitro*** evolution. *Hypertherm Enz Pt A* 2001, 330:364-379.

31. • Suzuki T, Yasugi M, Arisaka F, Yamagishi A, Oshima T: Adaptation of a thermophilic enzyme 3-isopropylmalate dehydrogenase to low temperatures. *Protein Eng* 2001, 14:85-91.
Enzyme libraries in directed evolution experiments are usually contained in a transformed plasmid. In this study, the gene of interest was recombined into the genome. This eliminates the possibility of isolating false-positives due to increased copy number. Furthermore, the integrated gene should be more stable in the host genome.

32. Wintrode PL, Miyazaki K, Arnold FH: Patterns of adaptation in a laboratory evolved thermophilic enzyme. *Biochim Biophys Acta* 2001, 1549:1-8.

33. Szilagyi A, Zavodszky P: Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* 2000, 8:493-504.

34. McDonald JH, Grasso AM, Rejto LK: Patterns of temperature adaptation in proteins from Methanococcus and Bacillus. *Mol Biol Evol* 1999, 16:1785-1790.

35. Reetz MT: Combinatorial and evolution-based methods in the creation of enantioselective catalysts. *Angew Chem Int Ed Engl* 2001, 40:284-310.

36. Olsen M, Iverson B, Georgiou G: High-throughput screening of enzyme libraries. *Curr Opin Biotechnol* 2000, 11:331-337.

37. Wahler D, Reymond JL: Novel methods for biocatalyst screening. *Curr Opin Chem Biol* 2001, 5:152-158.

38. •• Ghadessy FJ, Ong JL, Holliger P: Directed evolution of polymerase function by compartmentalized self-replication. *Proc Natl Acad Sci USA* 2001, 98:4552-4557.
This strategy links polymerase adaptive evolutionary improvement to amplification of its own gene. Genotype–phenotype linkage is effected by compartmentalization in an emulsion.

39. Tawfik DS, Griffiths AD: Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* 1998, 16:652-656.

40. • Olsen MJ, Stephens D, Griffiths D, Daugherty P, Georgiou G, Iverson BL: Function-based isolation of novel enzymes from a large library. *Nat Biotechnol* 2000, 18:1071-1074.
A cell-surface display method using a FRET/FACS system for evolution of catalytic function is described. This system can screen 10^9 cells/hour.

41. Stemmer WPC: DNA shuffling by random fragmentation and reassembly – ***in vitro*** recombination for molecular evolution. *Proc Natl Acad Sci USA* 1994, 91:10747-10751.

42. • Coco WM, Levinson WE, Crist MJ, Hektor HJ, Darzins A, Pienkos PT, Squires CH, Monticello DJ: DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotechnol* 2001, 19:354-359.
Libraries of monooxygenase genes shuffled using this method displayed high crossover frequency with no detectable parental genes.

43. • Abecassis V, Pompon D, Truan G: High efficiency family shuffling based on multi-step PCR and ***in vivo*** DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *Nucleic Acids Res* 2000, 28:E88.
In vivo and *in vitro* recombination are combined for directed evolution of eukaryotic genes that are difficult to express in bacteria. Gene libraries are characterized by probe hybridization.

44. Ostermeier M, Benkovic SJ: Evolution of protein function by domain swapping. *Adv Protein Chem* 2001, 55:29-77.

45. Ostermeier M, Shim JH, Benkovic SJ: A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat Biotechnol* 1999, 17:1205-1209.

46. • Sieber V, Martinez CA, Arnold FH: Libraries of hybrid proteins from distantly related sequences. *Nat Biotechnol* 2001, 19:456-460.
This paper demonstrates a method for recombining two parental genes independent of homology by creating chimeras of a soluble microbial enzyme and a membrane-associated mammalian enzyme.

47. Riechmann L, Winter G: Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc Natl Acad Sci USA* 2000, 97:10068-10073.

48. Kolkman JA, Stemmer WP: Directed evolution of proteins by exon shuffling. *Nat Biotechnol* 2001, 19:423-428.

49. • Fabret C, Poncet S, Danielsen S, Borchert TV, Ehrlich SD, Janniere L: Efficient gene targeted random mutagenesis in genetically stable ***Escherichia coli*** strains. *Nucleic Acids Res* 2000, 28:E95.
This approach to *in vivo* mutagenesis restricts replication errors to the targeted gene and thus minimizes effects on cell physiology, allowing a higher percentage (10%) of mutants to appear in a bacterial population.

50. • Voigt CA, Mayo SL, Arnold FH, Wang ZG: Computational method to reduce the search space for directed protein evolution. *Proc Natl Acad Sci USA* 2001, 98:3778-3783.
This method proposes that an evolutionary search can profitably focus on amino acid positions that are highly tolerant to mutations. A structure-based computational method is used to calculate the tolerance of each residue, which is shown to correlate well with beneficial mutations identified during directed evolution.

51. Miyazaki K, Arnold FH: Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J Mol Evol* 1999, 49:716-720.

52. Suenaga H, Mitsuoka M, Ura Y, Watanabe T, Furukawa K: Directed evolution of biphenyl dioxygenase: emergence of enhanced degradation capacity for benzene, toluene, and alkylbenzenes. *J Bacteriol* 2001, 183: 5441-5444.

53. Wang CC, Liao JC: Alteration of product specificity of ***Rhodobacter sphaeroides*** phytoene desaturase by directed evolution. *J Biol Chem* 2001, in press.

54. • Raillard S, Krebber A, Chen Y, Ness JE, Bermudez E, Trinidad R, Fullem R, Davis C, Welch M, Seffernick J *et al*.: Novel enzyme activities and functional plasticity revealed by recombining highly homologous enzymes. *Chem Biol* 2001, 8:891-898.
Shuffling of two triazine hydrolases generated mutants that were much more active and converted substrates that were not accepted by either parent. Permutation of only nine amino acid differences resulted in a functionally diverse mutant library.

55. • Gray KA, Richardson TH, Kretz K, Short JM, Bartnek F, Knowles R, Kan L, Swanson PE, Robertson DE: Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Adv Synth Catal* 2001, 343:607-617.
Saturation mutagenesis at every site was used to identify stabilizing mutations, which were combined by site-directed mutagenesis. The resulting biocatalyst is significantly improved for 1,2,3-trichloropropane hydrolysis in a large-scale process. It would be interesting to know how this approach, which relies on the cumulative effects of multiple mutations identified in the wild-type background, compares in cost and benefits to alternative strategies such as sequential random mutagenesis and selective site saturation.