REVIEW

Combinatorial approaches to engineering hybrid enzymes

James D. Stevenson and Stephen J. Benkovic*

Department of Chemistry, The Pennsylvania State University, 414 Wartik Laboratory, University Park, PA, 16802, USA. E-mail: sjb1@psu.edu; Fax: (814) 865 2973

Received (in Cambridge, UK) 27th May 2002
First published as an Advance Article on the web 18th July 2002

Covering: 1994–2002

1 Introduction
2 Methods of fragment recombination
2.1 *In vitro* homologous recombination
2.2 Non-homologous recombination
3 Improving enzyme activity
4 Changing substrate specificity
5 Improving thermostability
6 Enzymes in organic media
7 Modulating activity
8 Conclusions
9 References

1 Introduction

Enzymes are finding increasing application in both the academic and industrial communities as a means of overcoming the demanding challenges facing modern synthetic chemistry.[1,2] The benefits of enzyme catalysed processes are hard to overstate, with high levels of rate acceleration and chemo-, regio- and stereo-control possible under the mildest of reaction conditions. With increasing availability and reduced cost enzymes have become a viable alternative to chemical catalysts.

Choosing the right enzyme for the right task is often the main hurdle. The highly specific nature of protein catalysts, in terms of their substrate preference and tolerance to reaction environment, can significantly limit the broader application of many enzymes. Finding enzymes which can utilize different substrates, or operate in organic media or at different temperatures and pH's, is a significant task. Searching the diversity provided by Nature is one approach, but the advent of *in vitro* molecular evolution has made it possible to direct the evolution of existing enzymes in order to obtain new catalysts with tailored properties.[3,4]

Early attempts to alter protein function involved using site directed mutagenesis to introduce specific changes to the amino acid sequence based on protein structure analysis.[5] This approach had limited success however due to our limited understanding of protein structure–function relationships. The use of random mutagenesis to create large libraries of variants, coupled with efficient selection methods, later became more prevalent. This approach mimics the natural process of evolution by first generating molecular diversity, and then applying a selection criterion, directing the evolutionary path of the enzyme towards a desired target.

Making random mutations, however, soon incurs the problems associated with expanding library size. For example, a library of 3.2×10^6 (20^5) variants is generated if only 5 amino acids are randomly substituted with all possible combinations of amino acids. Analysing libraries of this size for improved function is a significant task, even with powerful genetic selection techniques. Searching a broad region of sequence space is still therefore not feasible using simple random mutagenesis.

A significant development in the capabilities of directed evolution was realised when the natural diversity-generating process of recombination was applied to *in vitro* evolution. Recombination involves the splicing together of multiple DNA sequences to form recombinant molecules containing sequences from different parents. When applied to directed evolution, this process is often referred to as 'molecular breeding',[6,7] as it involves crossing genes in much the same way as animals with desirable qualities are cross-bred in order to combine traits in their offspring. In this case, genes encoding proteins with desired properties are recombined *in vitro* to generate new chimeric DNA molecules, which will encode for hybrid proteins with novel properties.

In comparison with site-directed or random mutagenesis, *in vitro* gene recombination has the key advantage that a larger range of protein sequence space can be sampled in the mutant libraries (see Fig. 1). The chances of identifying hybrid proteins with improved functions is therefore increased. In addition, accumulated deleterious mutations can be eliminated.[8] Using this approach many examples of proteins with altered properties have been generated for use as industrial biocatalysts, protein therapeutics and as tools for biotechnology. In this review, we will focus specifically on the directed evolution of enzyme function, and how the use of combinatorial hybrid enzyme libraries, coupled with powerful selection and screening strategies, can achieve the goal of obtaining new biocatalysts with tailored properties.

2 Methods of fragment recombination

Methods for producing hybrid proteins involve first generating hybrid DNA molecules. These hybrid genes are then expressed *in vivo* to obtain the mutant protein products. A variety of different methodologies for carrying out this *in vitro* recombination process have been reported in the literature, and a selection of the most commonly used methods will be outlined here.

The simplest method of producing a hybrid protein is to use rational design to select one or more positions of crossover by either sequence comparison or analysis of the protein's crystal structure. The hybrid can then be readily generated using overlap extension PCR (polymerase chain reaction).[9] In addition to single crossover hybrids, single amino acid changes, secondary structure elements or whole domains can be introduced into an existing protein sequence using standard molecular biology techniques.[10]

Of more interest to protein engineers is the ability to produce vast libraries of chimeras in which every possible hybrid position is present. The individual hybrids with the desired property can then be isolated from this library. This approach overcomes the possibility that many hybrids may be inactive due to local folding problems at a given crossover position. The methods developed for fragment recombination can be broadly separated into homology-dependent and independent

DOI: 10.1039/b107096g

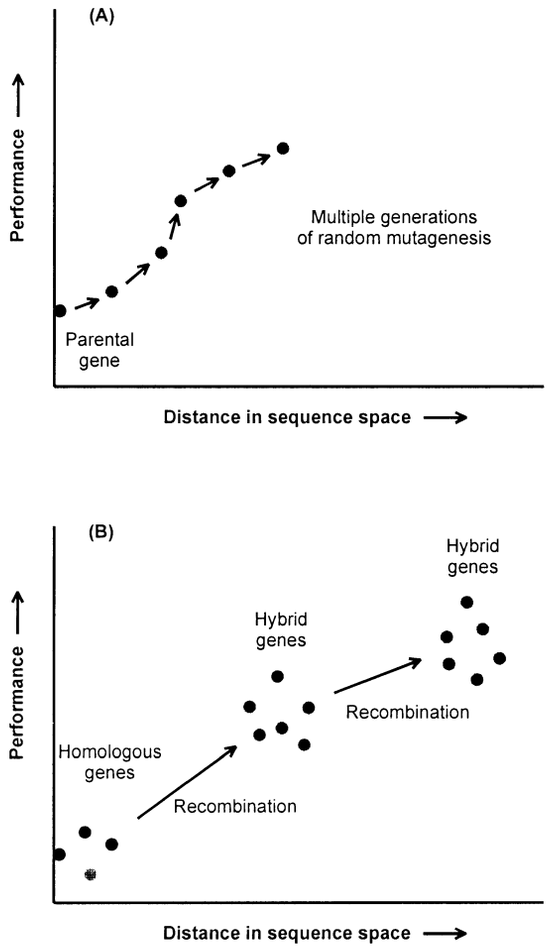

Fig. 1 Strategies for sampling protein sequence space using directed evolution. (A) Random mutagenesis of a single parental gene, followed by selection for improved performance and repeated rounds of mutagenesis. (B) Recombination of a set of homologous genes (family shuffling) resulting in hybrid gene libraries, followed by selection for improved performance and repeated rounds of recombination. For libraries of equivalent size, hybrid gene libraries sample a much broader, but sparser, area of sequence space than that sampled by random mutagenesis.

recombination, depending on the level of sequence homology required between the parental genes. The advantages and shortcomings of both methods will be discussed in the following sections.

2.1 *In vitro* homologous recombination

Similarly to *in vivo* recombination in nature, *in vitro* homologous recombination of gene fragments requires a high level of DNA sequence homology being present in the parental gene sequences. Short DNA fragments from multiple parent genes are used to recreate full length chimeric genes displaying both random positioning and random numbers of crossovers. The assembly step is always driven by local DNA homology, promoting hybridisation between fragments. As a result, high DNA sequence homology is a prerequisite for any candidate genes.

DNA shuffling

The pioneering method, and still the most commonly used, is that of DNA shuffling[11,12] and subsequent modifications[13,14] (also referred to as sexual PCR, molecular breeding or family shuffling, see Fig. 2A). This process takes a repertoire of homologous genes, and fragments them into small pieces by limited digestion with DNase I. Following purification to remove undigested material, the combined fragments are then reassembled in a primer-less PCR reaction in the presence of a thermostable polymerase. Recombination arises as a result of fragments from different parent genes cross-priming to produce a hybrid DNA strand. Parental sized genes are then amplified using terminal primers in a conventional PCR reaction.

In the first published example of DNA shuffling in 1994, Stemmer[12] used repeated rounds of this recombination method, coupled with *in vivo* selection, to enhance the ability of TEM-1 β-lactamase to convey bacterial resistance to the antibiotic cefotaxime. Starting from the single wild-type lactamase gene, shuffling was carried out in order to mix the repertoire of point mutations accumulated during PCR amplification. The most active mutants obtained displayed a 32,000 fold increase in activity, yet contained only 6 point mutations in comparison with the parent protein. This approach, while very successful, was still only able to sample a very limited amount of potential sequence space, as demonstrated by the small number of point mutations observed. Subsequently, Crameri *et al.*[6] expanded the scope of the DNA shuffling method by using an homologous set of related cephalosporinase genes from different species. By shuffling these genes, chimeras were produced with a mosaic-like structure in which blocks of sequences had been exchanged between multiple parents. As a result, a much greater area of sequence space was explored, by exploiting the diversity provided by nature.

These 'family shuffling' experiments yielded hybrid cephalosporinase enzymes with between 270–540 fold improved activity against the moxalactam antibiotic. The most active hybrids contained 8 fragments from 3 of the parental enzymes and an additional collection of 33 amino acid point mutations. As a comparison, DNA shuffling was also carried out separately on the individual parental genes. This yielded mutant enzymes with only eightfold improved activity, arising from a handful of point mutations. This clearly demonstrates that the family shuffling approach, by undertaking a 'sparse sampling' of sequence space, has a much greater chance of generating hybrid enzymes with improved function.

Random priming recombination (RPR)

Since the original DNA shuffling method was demonstrated, other related techniques of *in vitro* homologous recombination have been developed. Shao *et al.* reported the random priming recombination (RPR) method,[15] which varies from DNA shuffling in that the gene fragments are obtained by random priming synthesis instead of DNase I digestion. Following template removal, the fragments are then used to reassemble parent sized hybrid genes as in DNA shuffling. This has the benefit that only small quantities of template are required, and the use of DNase I is avoided.

Staggered extension process (StEP)

Zhao *et al.*[16] eliminated the use of gene fragments, and instead used template-switching as a way of building chimeric sequences from multiple parents. In this method, termed staggered extension process (see Fig. 2B), repeated terminal-primed synthesis is carried out with extremely short extension times, such that each newly synthesized DNA strand switches templates multiple times. This results in new chimeric strands containing sequences from all the templates to which each individual strand annealed during its synthesis. In comparison with RPR,[15] StEP also has the benefits that only a small amount of template is needed, and DNase I is avoided. The need to remove the template strands is also avoided, and the increased level of experimental automation significantly simplifies the overall procedure.

Random chimeragenesis on transient templates (RACHITT)

This most recent addition[17] to the gene shuffling toolkit combines elements of both DNA shuffling and StEP

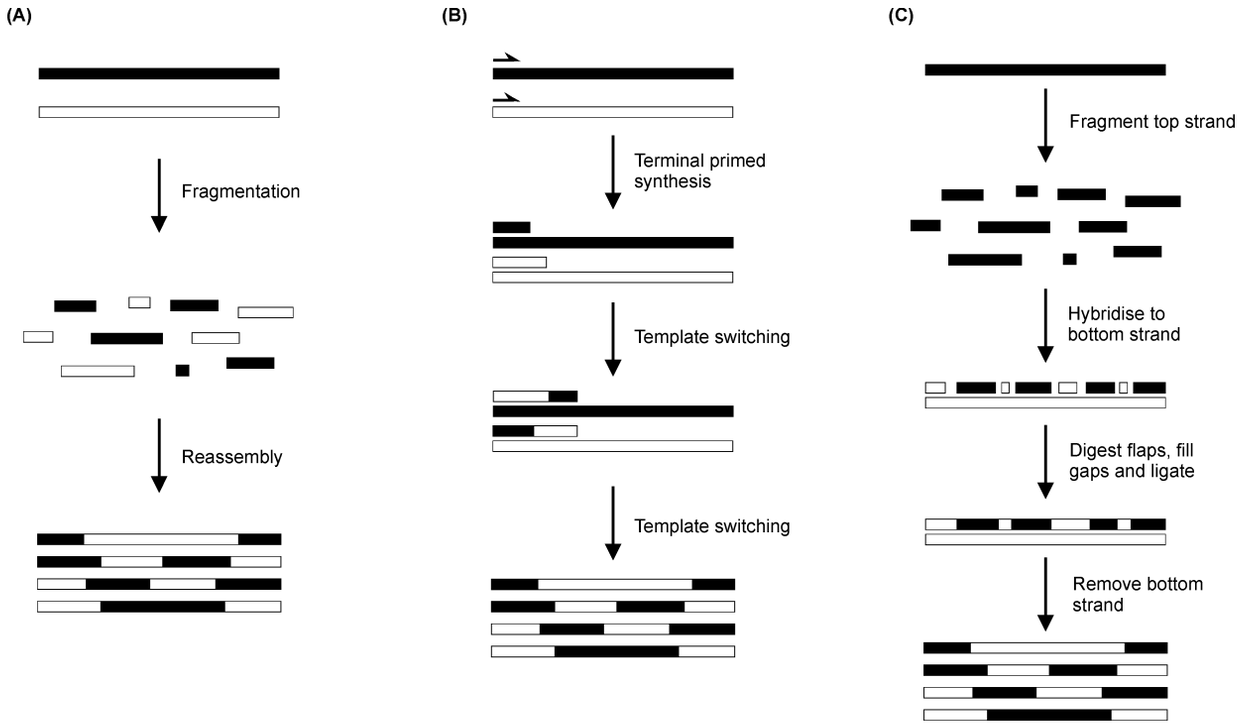

Fig. 2 Homologous *in vitro* recombination methods used to create hybrid gene libraries from multiple parental sequences (only 2 shown). (A). DNA shuffling:[11] parental genes are fragmented with DNaseI, and then reassembled in a primer-less PCR reaction. (B). Staggered extension protocol (StEP):[16] limited terminal primed extension is followed by multiple rounds of template switching between parental templates. (C). Random chimera-genesis on transient templates (RACHITT):[17] a transient parental bottom strand is used as a template to reassemble fragments from multiple parents, followed by flap removal, gap filling, nick ligation and destruction of the template strand. All methods require high DNA sequence homology to generate hybrid crossovers.

recombination. First, parental genes are fragmented using DNase I, then a uracil-containing template strand is used to direct assembly of chimeras from the fragment pool (see Fig. 2C). Fragments annealed to the transient template are trimmed and joined by treatment with nuclease competent polymerases and DNA ligase. The template strand is then glycosylated by uracil-DNA-glycosylase, rendering it non-amplifiable in a subsequent PCR amplification reaction. As a result the parent templates are effectively eliminated from the gene pool.

Experimentally, this method appears more demanding than those discussed previously. However, the reported quality of the resulting library, in terms of crossover numbers and diversity, is much improved over other shuffling methods. An average of 14 crossovers per gene were reported, in comparison with a figure of between 1–5 for DNA shuffling and StEP recombination. Clearly the more crossovers per hybrid, the greater the sequence diversity of the library, and the greater the chances of obtaining mutants with improved activity. In addition, no parental gene sequences were identified from the library, a common problem encountered with other methods of homologous recombination.

Restriction fragment shuffling

The problem of regenerating parental genes from a DNA shuffling reaction was highlighted recently in work by Kikuchi *et al.*,[18] in which a recombination frequency of less then 1% was observed. Two catechol dioxygenase genes sharing 84% identity were shuffled by conventional methods, yielding almost exclusively wild-type genes. In an attempt to overcome this, the authors modified the shuffling procedure by generating the fragment pools using sets of different restriction endonucleases, instead of DNase I. The digested DNA fragments were then used in a reassembly reaction to generate almost 100% recombinant genes. While successful in this case, the main limitation of this approach is that the formation of sequence crossovers becomes biased to existing restriction sites. Sampling crossovers along the whole length of the target genes is therefore not possible, limiting library size and diversity.

2.2 Non-homologous recombination

The main limitation of all the aforementioned methods is that while related proteins may share common 3D structures and amino acid sequences, the essential homology at the DNA level may have been lost as neutral mutations accumulated. As a result, the homology-dependent assembly step, common to all the techniques, may not be possible. In such cases where homology is limited between the parental sequences, library bias toward parental sequences, or crossovers being limited to pockets of local DNA homology becomes prevalent.[19,20] The chances of generating comprehensive mosaic sequence libraries are thus reduced, impacting significantly on the success of evolving enzyme function using this strategy.

Recently a number of methods have been reported which allow the creation of hybrid protein libraries independent of the DNA sequence homology of the parental genes. This is achieved by eliminating the need for homologous fragment hybridisation, and instead uses a blunt-ended ligation step, to bring together gene fragments. No bias towards the composition of those fragments is therefore encountered. This makes it possible for more distantly related, or even totally unrelated, protein sequences to be recombined in a combinatorial fashion.

Incremental truncation for the creation of hybrid enzymes (ITCHY)

Developed by Ostermeier and coworkers,[21–23] ITCHY was the first homology-independent combinatorial method of recombining gene fragments *in vitro*. The methodology was originally designed to map functional dissection points along protein backbones, as a way of converting existing monomeric

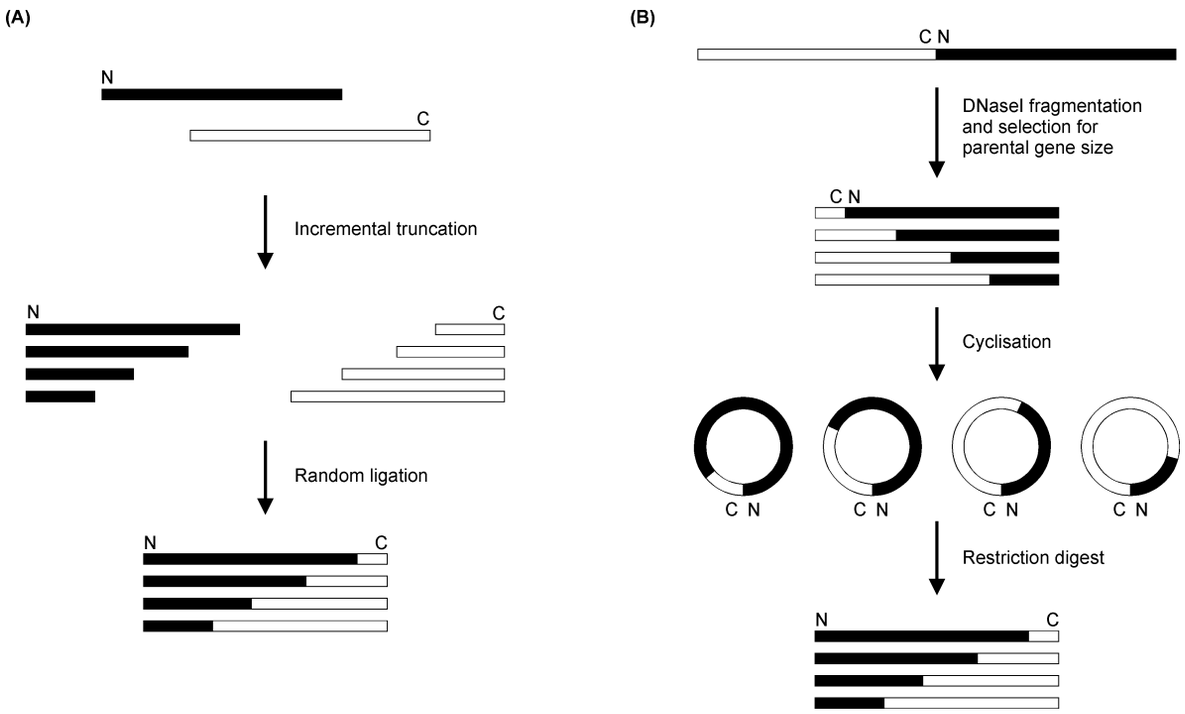

Fig. 3 Non-homologous *in vitro* recombination methods used to create single crossover gene libraries between two parental genes. (A). Incremental truncation for the creation of hybrid enzymes (ITCHY):[21] two opposing incremental truncation libraries are generated from two parental genes using Exonuclease III, and are then randomly ligated together to generate a comprehensive hybrid library. (B). Sequence homology independent protein recombination (SHIPREC):[26] a gene dimer is fragmented with DNaseI and parental sized fragments are recovered and inverted by circular ligation and restriction digest. For both ITCHY and SHIPREC, crossovers are formed independently of local DNA sequence homology by blunt ligation.

proteins into new heterodimeric forms.[24] Subsequently, the technique was extended to the creation of hybrid protein libraries.[21]

The method involves the production of two opposing incremental gene truncation libraries, in which every single DNA base pair deletion over a desired range is generated using controlled digestion with Exonuclease III (see Fig. 3A). Following blunting with single-stranded nuclease, the two fragment libraries are then randomly fused head-to-tail to generate a single crossover hybrid gene library. As the ligation step is totally random, crossovers are not limited to either positions of homology, or positions at which the two sequences align, *i.e.* hybrids of all possible lengths and reading frames are generated. ITCHY therefore has the benefit of not only forming crossovers within regions of low homology, but also of generating hybrids with sequence insertions and deletions.

The ITCHY technique was directly compared with DNA shuffling by creating interspecies hybrids of glycinamide ribonucleotide transformylase (GART) from *E. coli* (PurN) and human (hGART). The genes for these two enzymes share only 50% DNA sequence identity, making them unsuitable for family DNA shuffling, which would ordinarily require at least 70% DNA homology. Analysis of active hybrid GARTs generated by both methods revealed that crossovers were much more widely distributed in the ITCHY hybrids, while the DNA shuffled hybrids contained crossovers within only one short stretch of high DNA sequence homology and total amino acid sequence identity. Furthermore, the highest level of enzyme activity was demonstrated by ITCHY hybrids, with fusion points at regions of low sequence homology at both the DNA and protein level.

Interestingly, of all the possible hybrids generated by ITCHY, only those displaying perfect sequence alignment were active, *i.e.* parent sized proteins. While this observation may be specific to this enzyme case, a modified ITCHY protocol[25] was developed in order to produce libraries containing primarily parent sized hybrid genes. While more experimentally demanding than the original method, this approach allows much of the redundancy to be eliminated from a library.

Sequence homology independent protein recombination (SHIPREC)

Sieber *et al.*[26] developed an additional method of producing single crossover hybrid protein libraries between two parental genes, again with no requirement for DNA sequence homology. The recombination method (see Fig. 3B) uses an end-to-end gene fusion of the two genes with a linker sequence positioned in between which contains unique restriction sites. The gene dimer is digested with DNase I and fragments corresponding to parental sized genes are isolated. Intramolecular blunt ligation, followed by restriction digest, then flips the orientation of the two fragments to generate the hybrid gene library.

As with the ITCHY method, fragment recombination is achieved by blunt ligation, resulting in a loss of reading frame in a large proportion of the library members. To eliminate these members prior to activity screening, the SHIPREC library was subjected to a pre-selection step. The antibiotic resistance gene chloramphenicol acetyltransferase (CAT) was fused at the C-terminus of the hybrid gene library, allowing selection *in vivo* for library members maintaining a correct reading frame using the chloramphenicol resistance phenotype. In addition, this step favours the selection of soluble hybrids, as the fusion of insoluble proteins to CAT is known to result in decreased resistance *in vivo*.[27] Similar methods of selecting for soluble in-frame hybrid proteins have been reported in the literature.[28,29]

Using the SHIPREC method, the authors created a library of interspecies hybrids between a membrane-associated human cytochrome P450 (1A2) and a soluble *B. megaterium* P450 (BM3), the two enzymes sharing only 16% amino acid sequence identity. Hybrids were identified which displayed the substrate specificity of the human cytochrome, but which also showed increased solubility in the bacterial cytoplasm in comparison with the native human enzyme.

Homology independent DNA shuffling

The main limitation of the homology-independent recombination methods outlined so far, is that only single crossover hybrids from two parents can be generated. This greatly limits the potential sequence diversity which can be explored for improving enzyme function. In all reported cases of successful DNA family shuffling, multiple crossovers between two or more parental sequences have been required in order to achieve the desired properties.

To overcome this limitation, Lutz and coworkers[30] used ITCHY and DNA shuffling in a sequential manner in order to create multiple crossover libraries between two genes of low sequence homology (see Fig. 4). Termed SCRATCHY, this approach first uses ITCHY to install a repertoire of single crossovers at purely random locations between the two genes. The ITCHY hybrids are then used as a set of pseudo-homologous genes for DNA family shuffling. This step recombines the pre-existing ITCHY crossovers, forming multiple crossover hybrids. The homology-dependent DNA shuffling process is therefore used as a method of combining crossovers originally generated independently of DNA sequence homology. The combination of these two techniques results in a hybrid library with much greater sequence diversity than that which would be generated using ITCHY or DNA shuffling alone.

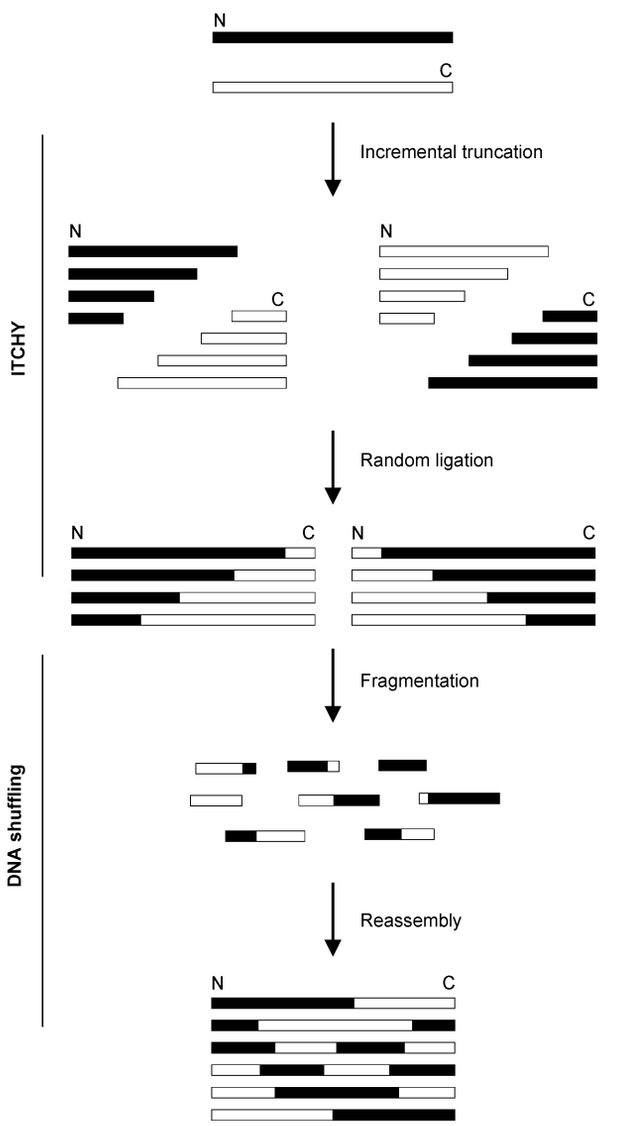

Fig. 4 Homology-independent fragment shuffling (SCRATCHY)[30] combines ITCHY and DNA shuffling in order to create multiple crossover libraries between genes of low DNA sequence homology. First, ITCHY libraries introduce single crossovers between two genes in both orientations, *i.e.* each gene at the N- and C-terminus. Then the ITCHY libraries are combined and fragmented with DNaseI and reassembled in a DNA shuffling reaction to introduce multiple crossovers into the hybrid libraries.

Using SCRATCHY on the GAR transformylase model system, functional enzymes were identified which contained multiple crossovers at positions of both high and low sequence homology. Although the highest number of crossovers identified was three, only a limited library was sampled, and computer modelling of SCRATCHY recombination indicated that formation of higher numbers of crossovers was possible. In addition, reassembly of parental sequences was not predominant following the DNA shuffling step, despite the low level of sequence homology between the two parental genes.

3 Improving enzyme activity

As discussed previously, the original work on DNA shuffling carried out by Stemmer[11] was demonstrated by improving the activity of a lactamase to convey bacterial resistance to the poorly hydrolysed antibiotic cefotaxime. Starting from a minimum inhibitory concentration (MIC) of 0.02 μg ml^{-1} for *E. coli* cells expressing the TEM-1 β-lactamase, mutants obtained using DNA shuffling displayed new MIC's of as high as 640 μg ml^{-1}, representing a 32,000-fold improvement compared with the parental enzyme.

Since this original demonstration, numerous reports by Stemmer on improving the activity of enzymes, signalling proteins,[31] antibodies,[32] other proteins[33] and even whole viruses,[34,35] have appeared in the literature and have been reviewed extensively elsewhere.[7,36,37] The remit of this review will focus specifically on evolving enzyme properties.

Thymidine kinase

In an attempt to obtain improved thymidine kinase (TK) enzymes for potential HIV therapies using 3′-azidothymidine (AZT), Christians *et al.*[38] used DNA family shuffling to create a library of hybrids between TK's from herpes simplex virus types 1 and 2 (HSV-1,2). Both of these viral TK enzymes display poor substrate selectivity, and readily phosphorylate nucleoside analogues, although not with high efficiency. Chimeric enzymes selected from the library displayed enhanced activity with AZT, reducing the concentration required to kill host cells by 32-fold compared with HSV-1 TK, and 16,000 fold less than HSV-2 TK.

Kinetic analysis of the best hybrid TK's revealed reduced K_M^{AZT} values compared with parental enzymes, indicating an increased preference for AZT as a substrate. This was supported using molecular modelling, which suggested that the hybrid enzymes' active sites were more able to accommodate the azido group of AZT.

Arsenate detoxification

In a more elaborate study, evolution of a multi-enzyme process was attempted by simultaneously shuffling all the genes within a whole operon. Crameri *et al.*[39] subjected the *Staphylococcus aureus* arsenic detoxification operon to multiple rounds of DNA shuffling and selection for increased resistance to arsenic in *E. coli*. The operon is composed of three genes: *arsR* (encoding the repressor regulatory protein), *arsB* (encoding the arsenite membrane efflux protein) and *arsC* (encoding the arsenate reductase enzyme which converts intracellular arsenate [As(v)] to arsenite [As(III)]). Clones were obtained which allowed growth in media containing up to 0.5 M arsenate, a 40-fold increase in resistance compared to the parental operon.

Analysis of the evolved operon revealed most of the mutations were located within the arsenite membrane pump (ArsB). In addition, the operon itself had become integrated into the bacterial chromosome, allowing for more efficient expression of the detoxification proteins. The arsenate reductase (ArsC) contained no mutations, but its expression level had become elevated, and arsenate reduction had increased 12-fold. This combination of factors, cooperatively enhancing arsenate resistance, would be difficult to engineer in a rational fashion, clearly demonstrating the strength of the directed evolution approach in tackling such complex problems. The bacterial strains evolved in this study could find potential application in the bioremediation of arsenic contaminated effluents generated by the chemical industry.

Biphenyl dioxygenases

A similar study, attempting to enhance the activity of detoxification enzymes for the purposes of bioremediation, was reported by Kumamaru *et al.*[40] In this case, hybrid biphenyl dioxygenases (BP Dox) were generated in order to improve the *in vivo* oxidative degradation of polychlorinated biphenyls (PCB's), common environmental pollutants. Two genes, encoding the large substrate binding subunit of BP Dox, from *Pseudomonas pseudoalcaligenes* and *Burkholderia cepacia* were recombined by DNA shuffling. The resulting hybrids were assayed for their ability to degrade a variety of biphenyl compounds *in vivo* (see Fig. 5).

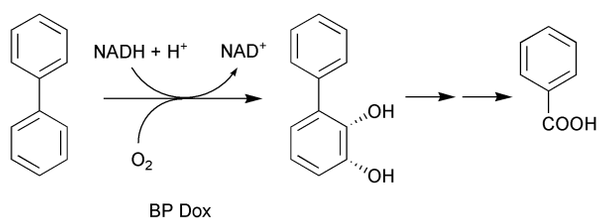

Fig. 5 Oxidative catabolic pathway of biphenyl compounds initiated by biphenyl dioxygenases (BP Dox). DNA shuffling of two homologous BP Dox genes from different species resulted in variants with improved activity towards polychlorinated biphenyl compounds and single ring aromatics.[40]

Hybrid enzymes were obtained with improved activity towards PCB and related compounds, although no detailed kinetic characterization was reported. In addition, some of the evolved hybrids displayed activity with other single ring aromatics such as benzene and toluene, which are poor substrates for the parental enzymes. Therefore, not only an increase in activity, but an expansion of specificity, was obtained by generating hybrid BP Dox enzymes. Similar results were obtained in subsequent studies by the same group using random priming recombination[41] and site directed mutagenesis.[42] For a recent review of engineering oxygenase enzyme see Cirino and Arnold.[43]

Subtilisin

An extensive report on the simultaneous evolution of multiple enzyme properties was attempted with the protease subtilisin by Ness *et al.*[44] Using the family shuffling approach, 25 subtilisin genes (ranging from 63.7–99.5% identity) from different *Bacillus* species, were shuffled to generate a small library of 654 active hybrid proteases. The hybrids were screened using a high throughput fluorescence assay for four different properties: activity at 23 °C, thermostability, solvent stability and pH dependence.

Under all four assay conditions, hybrids were obtained which outperformed all of the parental proteases. In addition, the most active hybrids at 23 °C and pH 10 also displayed improved thermostability and activity in the presence of 35% dimethylformamide (DMF). This demonstrates that family shuffling can potentially improve more than one enzyme property at once. In addition, this study allowed the authors to correlate each property in terms of parental sequence composition in order to determine which parent sequences would be best used in a shuffling reaction to optimise each individual property.

4 Changing substrate specificity

Modifying an existing enzyme to utilize unnatural substrates, while retaining the same chemistry displayed by the parental protein, is a powerful method of producing new biocatalysts. Rational[45–47] and semi-rational[48] strategies have both been used extensively to obtain enzymes with new substrate specificities, generally by altering the amino acids residues which either interact directly, or are in close proximity to the substrate, *i.e.* first sphere residues.

This approach relies heavily on tertiary structure information however, and even in cases where protein structures are available, the outcome of such mutagenesis cannot always be accurately predicted. In addition, amino acid substitutions at positions distant from the active site, which would not be expected to elicit any alteration in substrate specificity, can very often have a significant impact. Achieving desired changes in substrate specificity using rational design is therefore not always a reliable strategy. In the absence of such a predictive method, combinatorial hybrid gene libraries represent a method of attaining desired changes in substrate specificity. A selection of the most significant works in this area will be highlighted in the following paragraphs.

Glycosidases

A common target enzyme in protein engineering is the *Escherichia coli* β-galactosidase LacZ, due to the well characterised nature of its function, mechanism and structure. This enzyme cleaves glycosidic linkages with high specificity for β-D-galactosyl moieties. In an attempt to switch this specificity from galactose to another naturally occurring sugar, D-fucose, Zhang *et al.*[49] employed DNA shuffling in combination with colony screening using a chromogenic fucose substrate. D-fucose differs in structure from D-galactose in only one position, it lacks a 6′ hydroxy group. All other functional groups and stereochemistry are maintained (see Fig. 6).

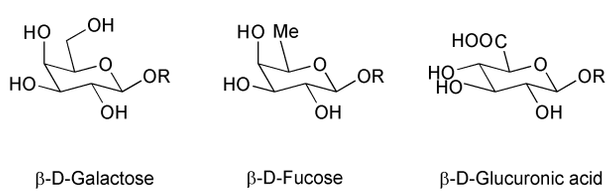

Fig. 6 Sugar substrate structures for engineered glycosidases. The β-galactosidase from *Escherichia coli* (LacZ) was evolved to accept D-fucose as a substrate,[49] and the *Escherichia coli* β-glucuronidase (GusA) was similarly adapted to utilise D-galactose.[50] R = *p*-nitrophenyl.

Converting a galactosidase enzyme into a fucosidase should therefore not be a great task, in fact similar k_{cat} values for both substrates are displayed by native β-galactosidase (~200 s^{-1} for both *p*-nitrophenyl-β-D-galactopyranoside and *p*-nitrophenyl-β-D-fucopyranoside).

After seven rounds of DNA shuffling and screening, an evolved fucosidase enzyme was obtained which displayed a 1000 and 300 fold increase in substrate specificity (k_{cat}/K_M) for the two fucose substrates relative to the corresponding galactose substrates in comparison with the native β-galactosidase. Kinetic characterisation showed that this was mostly due to K_M changes, seeing an enhanced affinity of the enzyme for fucose at the expense of galactose. The amino acid sequence of the evolved enzyme contained 6 point mutations, one of which, Asn604Ser, was located within the substrate

binding site, presumably reshaping the pocket to accommodate fucose.

In a related study Matsumura and coworkers attempted to create a new β-galactosidase enzyme starting from a β-glucuronidase.[50] In this case the structural differences between the two substrates are more pronounced (see Fig. 6). Glucuronic acid has an additional 6′ carboxylate group and the stereochemistry at the 4′ position is inverted relative to galactose. Adapting the binding pocket of the enzyme to accommodate this new substrate would therefore be expected to be more difficult.

After three rounds of DNA shuffling the catalytic efficiency of the *Escherichia coli* β-glucuronidase (GusA) with *p*-nitrophenyl-β-D-galactoside was improved 500-fold. This represents a 52 million fold inversion of the substrate specificity from galactose to glucuronic acid. In addition, the evolved β-galactosidase was found to display a substrate specificity profile very similar to that of the native *Escherichia coli* β-galactosidase, recognising fucose, but not galactoside-6-phosphate, as a substrate.

The authors then investigated the effects of the point mutations responsible for the change in specificity by reintroducing the individual mutations into the parental GusA gene and monitoring the change in substrate preference. They found that the 6 reoccurring mutations operated synergistically to elicit the functional properties. Clones containing only a few of the point mutations tended to display much broader substrate specificity than either the parental enzyme or the fully evolved β-galactosidase, utilising multiple substrates with intermediate efficiencies. Only when all the selected mutations were in place did the differentiation between substrates become apparent. This supports the 'patchwork' hypothesis of enzyme family evolution by apparently returning the enzyme to a less specialised state in which it can utilise many substrates with similar efficiencies, rather than the high level of specialisation displayed by modern enzymes.

Triazine hydrolases

The ability to recreate the properties of such ancestor-like enzymes, displaying broadened substrate specificities, was the goal of a recent study. Raillard and coworkers carried out DNA family shuffling using two highly homologous triazine hydrolases,[51] AtzA (from a *Pseudomonas* species) and TriA (from *Acidovorax avenea sitrulli*), which catalyse the hydrolytic dechlorination and deamination respectively of derivatives of the herbicide atrazine (see Fig. 7A). Interestingly, despite only differing at 9 amino acid positions (98.1% identity) the two enzymes have very little overlap in their substrate specificity profiles.

An extensive study of the substrate specificity of the chimeric library members generated by DNA shuffling was carried out against a panel of 15 *s*-triazine derivatives (see Fig. 7B). The small substrate library contained every combination of 5 different leaving groups and 3 different side groups, representing a broad range of chemical and structural diversity. For all of the substrates investigated, hybrid enzymes were identified with improved activity, up to 150-fold greater than either parental enzyme. Hybrids were also identified which utilised all of the substrates used by both parents, in effect combining their substrate profiles. In addition, activity with 5 out of 8 new substrates not used by either parent, was obtained. This study clearly demonstrates that activity with novel substrates can be elicited by recombining sequences from existing enzymes, even when none of the parental enzymes display activity with that substrate.

Aspartate aminotransferase

Yano and coworkers modified an aspartate aminotransferase to accept other amino acids as substrates.[52] The *E. coli*

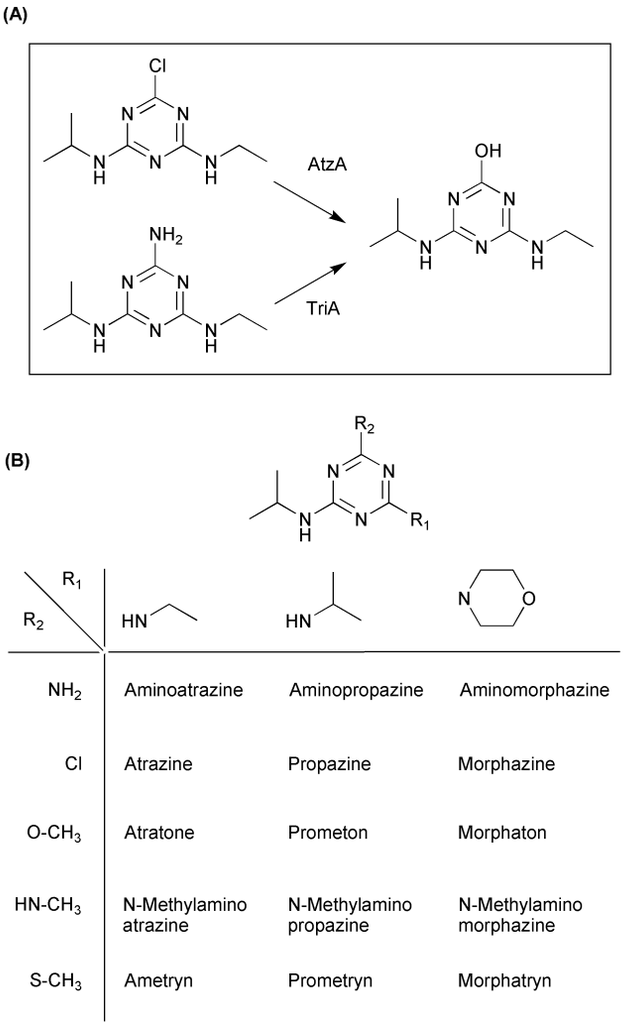

Fig. 7 Hydrolysis of triazine derivatives catalysed by triazine hydrolases. (A). Hydrolysis reaction catalysed by the two parental enzymes: dechlorination of atrazine by AtzA and deamination of aminoatrazine by TriA. (B). Triazine derivative library used to explore the substrate specificities of the hybrid triazine hydrolase library members generated by DNA shuffling. The library contains a range of side groups (R_1) of increasing steric bulk and a range of leaving groups (R_2). Adapted from Raillard *et al.*[51]

aspartate aminotransferase (AspAT) is highly specific for the acidic amino acids L-aspartate and L-glutamate. It has almost no activity with other structural classes of amino acids, in particular the β-branched amino acids valine, leucine and isoleucine, for which a different aminotransferase exists. The authors carried out DNA shuffling on the AspAT gene and used an *in vivo* selection system to identify new variants which use β-branched substrates.

In order to achieve this the authors generated a bacterial strain which had the native β-branched amino acid aminotransferase gene knocked out. Plasmid-borne AspAT variants capable of using β-branched substrates were then identified by selection on amino acid deficient media supplemented with 2-oxovaline. The most active clone (AV5A-7) displayed an impressive 1.3×10^5 fold increase in activity (k_{cat}/K_M) with 2-oxovaline and a 7.3×10^4 fold increase for 2-oxoisoleucine in comparison with the native AspAT enzyme. In addition a 20-fold loss of activity with its natural substrate oxaloacetate was observed. Of the six point mutations shown to be responsible for this functional conversion, only one was in direct proximity to the substrate.

Cytochrome P450

A common problem faced when attempting to use enzymes on an industrial scale is to make the process economically viable.

This is especially true when the enzyme in question requires expensive cofactors. Cytochrome P450's for example, which have great synthetic potential in the hydroxylation of unactivated organic compounds, require the nicotinamide adenine dinucleotide (NAD) cofactor. *In vivo* this cofactor is regenerated by a complex redox system involving multiple proteins. Reproducing this *in vitro* for synthetic applications would be impractical, so Joo and co-workers[53] attempted to overcome this problem by replacing NAD and O_2 with hydrogen peroxide as the oxidising source. This alternate reaction pathway, termed the 'peroxide shunt' pathway, is inefficient with the native cytochrome P450 when naphthalene is the substrate. In an attempt to improve this reaction the authors used random mutagenesis, coupled with StEP recombination, to evolve the enzyme *in vitro* to hydroxylate naphthalene (see Fig. 8).

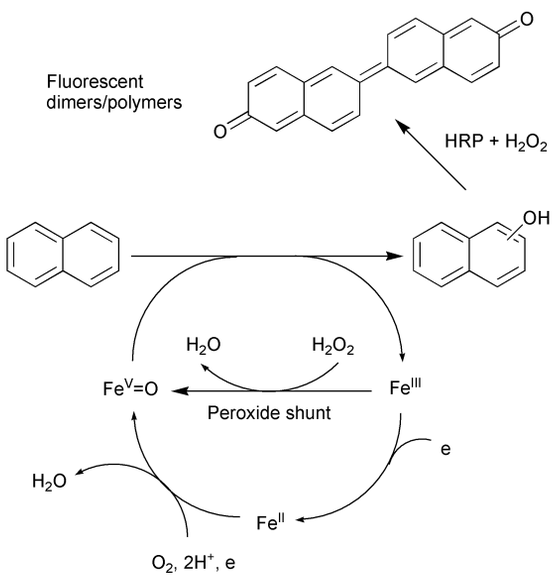

Fig. 8 Cytochrome P450 catalysed hydroxylation of naphthalene, and subsequent horseradish peroxidase (HRP) catalysed coupling of naphthol to form fluorescent dimer/polymer products, which can be detected *in vivo* by fluorescence assay. Mutant cytochrome P450's were obtained with improved efficiency for the peroxide shunt pathway,[53] bypassing the use of NADH and O_2 to regenerate the reduced iron centre of the enzyme.

Using a digital imaging screening system,[54] bacterial colonies expressing a mutant library of cytochrome P450 from *Pseudomonas putida*, and an active variant of horseradish peroxidase (HRP1A6),[55] were screened by monitoring fluorescence at 460 nm. Fluorescence arises as a result of the HRP catalysed coupling of hydroxylated naphthalene products, generating highly coloured dimeric and polymeric species. Variants of cytochrome P450 were identified with over 20-fold greater activities than the native enzyme in a whole cell assay. In addition, the versatility of the fluorescence screening system allowed mutants to be identified with novel hydroxylation selectivities, by monitoring the unique spectral properties of the different naphthalene derivatives produced. The enzymes generated in this study could be useful catalysts for the synthesis of various molecules in the absence of a cofactor.

5 Improving thermostability

Thermophilic micro-organisms have been a rich source of thermostable enzymes for carrying out reactions at elevated reaction temperatures. However, as in the case of enzyme substrate specificity, the molecular determinants of protein thermostability[56] are not well understood, and attempts to rationally design thermostable variants from existing thermolabile enzymes are not always successful. Screening combinatorial libraries has therefore become a common alternate approach[57] of obtaining enzymes displaying improved resistance to heat-denaturation (increased half-life of denaturation) and increased reaction temperature optima, both of which significantly impact enzyme utility in high-temperature processes.

Subtilisin

A common feature of enzymes isolated from thermophilic organisms is that, while the enzymes are as active as thermolabile enzymes at their optimal temperature, they often display minimal activity at lower temperatures. This is thought to be due to the reduced flexibility of the thermostable proteins at low temperatures. Obtaining enzymes which are both thermostable and active at lower temperatures is therefore a challenging goal for protein engineers. This was recently achieved by Miyazaki *et al.*[58] who improved the thermostability of the psychrophilic protease subtilisin S41 while simultaneously increasing the enzyme's activity at low temperatures.

Random mutagenesis and StEP recombination were used to generate a library of hybrid subtilisin S41 variants. The most active variant (3-2G7) displayed a 10 °C increase in its optimum temperature, and a 500 times longer half-life of denaturation at 60 °C than the native enzyme. In addition to this improved thermostability, the catalytic efficiency also showed improvement over a broad temperature range (10–60 °C). This demonstrates that engineering increased temperature stability doesn't always necessitate a loss in enzyme activity at lower temperatures, just that this property is not essential in a biological context.

A similar result was obtained with subtilisin E from the mesophile *Bacillus subtilis*.[59] Here the optimum temperature of the best evolved hybrid enzyme was increased by 17 °C to 76 °C. A half-life of denaturation of 3.5 minutes at 83 °C was also reported, equalling the thermostability of the protease thermitase from the thermophile *Thermoactinomyces vulgaris* (see Fig. 9). In addition, the activity profile of this mutant across the whole temperature range investigated (20–90 °C) showed improved activity compared with the parental protease. Improving activity at lower temperatures, while simultaneously increasing thermal stability, is therefore clearly a possibility using protein engineering strategies.

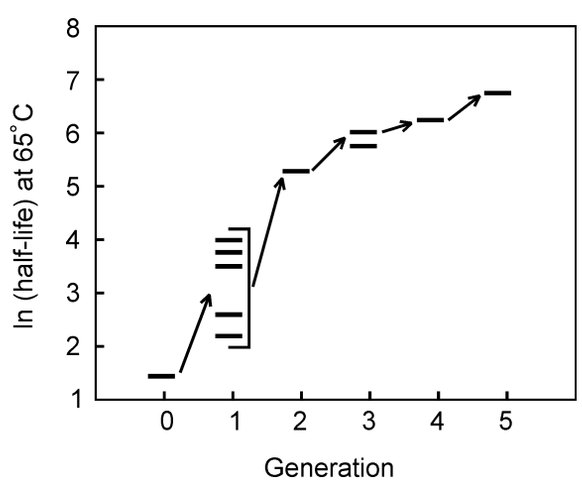

Fig. 9 Improving the thermostability of subtilisin E from *Bacillus subtilis* by directed evolution. The parental enzyme (generation 0) was subjected to random mutagenesis, resulting in five thermostable variants (generation 1). These beneficial mutations were then recombined using StEP to obtain the most thermostable combination (generation 2), which was then further mutated to achieve additional improvements in thermostability (generations 3–5). Each cycle of mutation and screening resulted in incremental increases in the enzyme's half-life of denaturation at 65 °C. Adapted from Zhao and Arnold.[59]

In a related study, Wintrode and coworkers[60] sought to adapt a mesophilic subtilisin-like protease from *Bacillus sphaericus* to operate at low temperatures as a functional equivalent of a psychrophilic protease. One round of random mutagenesis and StEP recombination resulted in a variant with a 10-fold increase in catalytic efficiency (k_{cat}/K_M) at 10 °C, which exceeds the activity of the psychrophilic subtilisin S41. A simultaneous loss in thermal stability was also displayed by this mutant which exhibited a denaturation half-life at 70 °C, 3.3-fold less then the parent protease. Similarly, indoleglycerol phosphate synthase from the hyperthermophile *Sulfolobus solfataricus* was evolved using DNA shuffling to increase its activity at low temperatures.[61] Again the thermostability of the parent enzyme was lost as a result of the mutations introduced to elicit increased activity.

Esterases

Giver and coworkers[62] reported that they observed an initial loss in activity at ambient temperatures as they enhanced the thermostability of a mesophilic *p*-nitrobenzyl esterase by directed evolution. But that in subsequent rounds of mutagenesis this activity was recovered, and exceeded that of the parent enzyme, while still retaining the improved structural stability. Clearly, thermostability and low temperature activity are compatible objectives so long as selection pressure for both criteria is applied to each generation of mutants.

In this case, the *p*-nitrobenzyl esterase from *Bacillus subtilis*, which is involved in cephalosporin antibiotic synthesis, was evolved in six rounds of random mutagenesis and DNA shuffling, coupled with selection for activity at 30 °C following heat treatment. The most active variant displayed a 14 °C increase in melting temperature (T_m), with an optimum temperature of ~55 °C. In addition, kinetic analysis revealed a 2-fold increase in catalytic efficiency at 30 °C compared to the parent enzyme.

Structural analysis was carried out on one of the evolved esterase mutants (8g8) in order to study the mechanism of thermostability.[63] Many of the 13 amino acid substitutions present in 8g8 were shown to be involved in stabilising flexible loops on the surface of the protein. This allowed unstructured regions, not visible in the electron density of the parent enzyme, to be viewed in the structure of 8g8. In addition, helix stabilisation and improved core residue packing appeared to account for the improved structural stability. The reduced flexibility of the thermostable esterase variant was also confirmed by a tryptophan fluorescence study,[64] although no correlation between flexibility and catalytic efficiency at ambient temperatures was observed.

β-Glucuronidase

In most cases, enzymes selected for directed evolution experiments have been monomeric species. However, one of the key mechanisms by which proteins from thermophiles achieve increased protein stability is oligomerisation. A potential route towards improving the thermostability of multimeric enzymes would therefore be to increase quaternary structural stability by enhancing subunit association.

This property was observed in thermostable mutants of the tetrameric enzyme β-glucuronidase from *Escherichia coli* obtained by Flores *et al.*[65] Following multiple rounds of random mutagenesis, DNA shuffling and screening, a range of mutant enzymes were obtained which retained activity at ambient temperatures following heat treatment. The best variants retained activity up to temperatures as high as 80 °C, whereas the parent enzyme was completely inactivated at 65 °C. The majority of the accumulated point mutations responsible for the increase in stability were located at the subunit interface, and many of the variants displayed increased resistance to thermal disruption of their oligomeric structure. Interestingly, in contrast to the thermostable subtilisin variants discussed previously,[58] the mutant β-glucuronidases all displayed reduced catalytic proficiency at ambient temperatures. Simultaneous improvement of activity and thermostability therefore does not appear feasible in this case.

6 Enzymes in organic media

Carrying out enzyme catalysed transformations in organic media has many potential benefits to the synthetic chemist.[66] Firstly, organic reagent solubility and stability will be significantly improved in comparison with aqueous media. In addition, effects on reaction thermodynamics and enzyme specificity can result in new products not available to aqueous reactions. Although many enzymes have proved to be viable catalysts in pure organic solvents, often aqueous–organic mixtures are most problematic in terms of enzyme denaturation. Adapting enzymes for use under these conditions has recently been a goal for directed enzyme evolution.

Moore and coworkers used random mutagenesis and DNA shuffling to modify the activity of an esterase for use in aqueous DMF solution.[67,68] The esterase from *Bacillus subtilis* catalyses the deprotection of the *p*-nitrobenzyl ester of the cephalosporin antibiotic Loracarbef (see Fig. 10). The best variant esterase

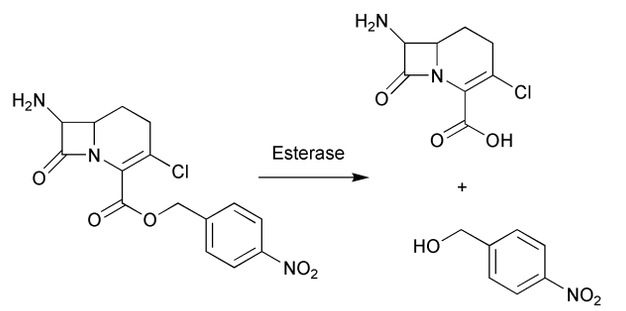

Fig. 10 Deprotection of the *p*-nitrobenzyl ester of the cephalosporin antibiotic Loracarbef, catalysed by the *p*-nitrobenzyl esterase from *Bacillus subtilis*. Random mutagenesis and DNA shuffling were used to increase the activity of this enzyme 60-fold in 30% aqueous dimethylformamide solution[67] in comparison with the wild-type enzyme.

displayed up to a 60-fold increase in activity in 30% aqueous DMF solution, equalling the performance of the parental enzyme in purely aqueous media. The detrimental effects of the organic co-solvent appeared to be overcome by increasing the overall stability of the protein. A structural study[63] revealed that many of the mutations were acting to stabilise flexible surface loops, a similar outcome to that observed for thermostable variants of the same enzyme.

In addition to denaturation, enzymes are also susceptible to chemical modification in the presence of highly reactive reagents. Matsumura *et al.*[69] sought to evolve resistance to chemical modification by aldehydes in the common reporter enzyme β-glucuronidase (GUS) from *Escherichia coli*. Tissue fixation (preservation) with either formaldehyde or glutaraldehyde results in a significant reduction in GUS activity, limiting the sensitivity of tissue staining. To reduce this effect on GUS, random mutagenesis and DNA shuffling were employed to alter the surface chemistry of the enzyme by selecting for variants which retained activity following treatment with increasing concentrations of glutaraldehyde.

The most active mutant (GUS^{AR}) retained the majority of its activity (78%) following treatment with 0.2% glutaraldehyde, whereas the parental enzyme lost 99.6% of its activity after treatment with only 0.04% glutaraldehyde. In addition, the catalytic efficiency of GUS^{AR} was not significantly different to that of the parent. Interestingly, only one of the eight mutations responsible for increasing aldehyde resistance involved substi-

tution of a lysine residue. Aldehydes are known to form Schiff bases with amine groups in proteins, so it would be expected that surface lysines would be eliminated in aldehyde resistant mutants. However, other mutations were in close proximity to conserved surface lysines, which could potentially modify the reactivity of the amino groups sufficiently to prevent reaction with glutaraldehyde.

7 Modulating activity

In order to use antibody-binding as a method of selectively regulating the activity of an enzyme, a number of reports have demonstrated that by inserting a peptide epitope sequence into the amino acid sequence of an enzyme, addition of an antibody which binds to that peptide sequence can efficiently inhibit catalytic activity. Brennan and coworkers inserted a 13 amino acid sequence from the HIV-1 gp120 protein into various surface loops around the active site of the *Escherichia coli* alkaline phosphatase.[70] In the presence of an antibody which binds to the gp120 sequence, one of the hybrid phosphatase enzymes exhibited a significant reduction in activity, up to 50% of that observed in the absence of the antibody. In a subsequent report[71] the same authors showed that by introducing two point mutations (D101S and D153G) into the parental AP, the effect of antibody binding could be transformed from an inhibitory effect into a stimulatory one. Up to a 300% increase in enzyme activity was reported. This was suggested to be due to a conformation change induced by antibody binding.

In a similar study, Legendre *et al.*[72] inserted short random peptide sequences into the TEM-1 β-lactamase at positions around the active site. Biopanning[73] of functional lactamases against an anti-prostate specific antigen (PSA) antibody identified variants of the enzyme which could be inhibited by antibody binding. The authors used these hybrid enzymes to develop an immunoassay which allowed them to detect extremely low concentrations (10^{-9} M) of PSA, without prior knowledge of the epitope specificity of the antibody.

8 Conclusions

The recent expansion in methodologies available for the rapid generation of vast enzyme libraries has made it feasible to access an ever expanding region of protein sequence space in search of improved biocatalysts. Coupled with powerful genetic selection and high-throughput screening systems, a diverse range of effects can be elicited with existing enzymes using these *in vitro* directed evolutionary techniques. Changes in catalytic efficiency, substrate specificity, stability and the ability to modulate activity are now possible. Expanding these approaches to engineer whole metabolic pathways[74] and even whole organisms,[75] offers a powerful route to obtaining new catalytic systems in the future.

9 References

1 K. M. Koeller and C. H. Wong, *Nature*, 2001, **409**, 232.
2 A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts and B. Witholt, *Nature*, 2001, **409**, 258.
3 I. P. Petrounia and F. H. Arnold, *Curr. Opin. Biotechnol.*, 2000, **11**, 325.
4 U. T. Bornscheuer and M. Pohl, *Curr. Opin. Chem. Biol.*, 2001, **5**, 137.
5 R. J. Kazlauskas, *Curr. Opin. Chem. Biol.*, 2000, **4**, 81.
6 A. Crameri, S. A. Raillard, E. Bermudez and W. P. Stemmer, *Nature*, 1998, **391**, 288.
7 J. Minshull and W. P. Stemmer, *Curr. Opin. Chem. Biol.*, 1999, **3**, 284.
8 H. Zhao and F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 7997.
9 R. M. Horton, H. D. Hunt, S. N. Ho, J. K. Pullen and L. R. Pease, *Gene*, 1989, **77**, 61.
10 J. Sambrook and D. W. Russell, *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001.
11 W. P. Stemmer, *Nature*, 1994, **370**, 389.
12 W. P. Stemmer, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 10747.
13 H. Zhao and F. H. Arnold, *Nucleic Acids Res.*, 1997, **25**, 1307.
14 I. A. Lorimer and I. Pastan, *Nucleic Acids Res.*, 1995, **23**, 3067.
15 Z. Shao, H. Zhao, L. Giver and F. H. Arnold, *Nucleic Acids Res.*, 1998, **26**, 681.
16 H. Zhao, L. Giver, Z. Shao, J. A. Affholter and F. H. Arnold, *Nat. Biotechnol.*, 1998, **16**, 258.
17 W. M. Coco, W. E. Levinson, M. J. Crist, H. J. Hektor, A. Darzins, P. T. Pienkos, C. H. Squires and D. J. Monticello, *Nat. Biotechnol.*, 2001, **19**, 354.
18 M. Kikuchi, K. Ohnishi and S. Harayama, *Gene*, 1999, **236**, 159.
19 G. L. Moore, C. D. Maranas, S. Lutz and S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 3226.
20 J. M. Joern, P. Meinhold and F. H. Arnold, *J. Mol. Biol.*, 2002, **316**, 643.
21 M. Ostermeier, J. H. Shim and S. J. Benkovic, *Nat. Biotechnol.*, 1999, **17**, 1205.
22 S. Lutz, M. Ostermeier and S. J. Benkovic, *Nucleic Acids Res.*, 2001, **29**, E16.
23 M. Ostermeier, A. E. Nixon and S. J. Benkovic, *Bioorg. Med. Chem.*, 1999, **7**, 2139.
24 M. Ostermeier, A. E. Nixon, J. H. Shim and S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 3562.
25 M. Ostermeier and S. J. Benkovic, *Biotechnol. Lett.*, 2001, **23**, 303.
26 V. Sieber, C. A. Martinez and F. H. Arnold, *Nat. Biotechnol.*, 2001, **19**, 456.
27 K. L. Maxwell, A. K. Mittermaier, J. D. Forman-Kay and A. R. Davidson, *Protein Sci.*, 1999, **8**, 1908.
28 G. S. Waldo, B. M. Standish, J. Berendzen and T. C. Terwilliger, *Nat. Biotechnol.*, 1999, **17**, 691.
29 W. C. Wigley, R. D. Stidham, N. M. Smith, J. F. Hunt and P. J. Thomas, *Nat. Biotechnol.*, 2001, **19**, 131.
30 S. Lutz, M. Ostermeier, G. L. Moore, C. D. Maranas and S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 11248.
31 C. C. Chang, T. T. Chen, B. W. Cox, G. N. Dawes, W. P. Stemmer, J. Punnonen and P. A. Patten, *Nat. Biotechnol.*, 1999, **17**, 793.
32 A. Crameri, S. Cwirla and W. P. Stemmer, *Nat. Med.*, 1996, **2**, 100.
33 A. Crameri, E. A. Whitehorn, E. Tate and W. P. Stemmer, *Nat. Biotechnol.*, 1996, **14**, 315.
34 S. K. Powell, M. A. Kaloss, A. Pinkstaff, R. McKee, I. Burimski, M. Pensiero, E. Otto, W. P. Stemmer and N. W. Soong, *Nat. Biotechnol.*, 2000, **18**, 1279.
35 N. W. Soong, L. Nomura, K. Pekrun, M. Reed, L. Sheppard, G. Dawes and W. P. Stemmer, *Nat. Genet.*, 2000, **25**, 436.
36 S. Harayama, *Trends Biotechnol.*, 1998, **16**, 76.
37 J. E. Ness, S. B. Del Cardayre, J. Minshull and W. P. Stemmer, *Adv. Protein Chem.*, 2000, **55**, 261.
38 F. C. Christians, L. Scapozza, A. Crameri, G. Folkers and W. P. Stemmer, *Nat. Biotechnol.*, 1999, **17**, 259.
39 A. Crameri, G. Dawes, E. Rodriguez, Jr., S. Silver and W. P. Stemmer, *Nat. Biotechnol.*, 1997, **15**, 436.
40 T. Kumamaru, H. Suenaga, M. Mitsuoka, T. Watanabe and K. Furukawa, *Nat. Biotechnol.*, 1998, **16**, 663.
41 H. Suenaga, M. Goto and K. Furukawa, *J. Biol. Chem.*, 2001, **276**, 22500.
42 H. Suenaga, M. Mitsuoka, Y. Ura, T. Watanabe and K. Furukawa, *J. Bacteriol.*, 2001, **183**, 5441.
43 P. C. Cirino and F. H. Arnold, *Curr. Opin. Chem. Biol.*, 2002, **6**, 130.
44 J. E. Ness, M. Welch, L. Giver, M. Bueno, J. R. Cherry, T. V. Borchert, W. P. Stemmer and J. Minshull, *Nat. Biotechnol.*, 1999, **17**, 893.
45 H. M. Wilks, K. W. Hart, R. Feeney, C. R. Dunn, H. Muirhead, W. N. Chia, D. A. Barstow, T. Atkinson, A. R. Clarke and J. J. Holbrook, *Science*, 1988, **242**, 1541.
46 N. S. Scrutton, A. Berry and R. N. Perham, *Nature*, 1990, **343**, 38.
47 J. L. Harris and C. S. Craik, *Curr. Opin. Chem. Biol.*, 1998, **2**, 127.
48 A. S. el Hawrani, R. B. Sessions, K. M. Moreton and J. J. Holbrook, *J. Mol. Biol.*, 1996, **264**, 97.
49 J. H. Zhang, G. Dawes and W. P. Stemmer, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 4504.
50 I. Matsumura and A. D. Ellington, *J. Mol. Biol.*, 2001, **305**, 331.
51 S. Raillard, A. Krebber, Y. Chen, J. E. Ness, E. Bermudez, R. Trinidad, R. Fullem, C. Davis, M. Welch, J. Seffernick, L. P. Wackett, W. P. Stemmer and J. Minshull, *Chem. Biol.*, 2001, **8**, 891.

52 T. Yano, S. Oue and H. Kagamiyama, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 5511.
53 H. Joo, Z. Lin and F. H. Arnold, *Nature*, 1999, **399**, 670.
54 H. Joo, A. Arisawa, Z. Lin and F. H. Arnold, *Chem. Biol.*, 1999, **6**, 699.
55 Z. Lin, T. Thorsen and F. H. Arnold, *Biotechnol. Prog.*, 1999, **15**, 467.
56 R. Jaenicke and G. Bohm, *Curr. Opin. Struct. Biol.*, 1998, **8**, 738.
57 F. H. Arnold, P. L. Wintrode, K. Miyazaki and A. Gershenson, *Trends Biochem. Sci.*, 2001, **26**, 100.
58 K. Miyazaki, P. L. Wintrode, R. A. Grayling, D. N. Rubingh and F. H. Arnold, *J. Mol. Biol.*, 2000, **297**, 1015.
59 H. Zhao and F. H. Arnold, *Protein Eng.*, 1999, **12**, 47.
60 P. L. Wintrode and F. H. Arnold, *Adv. Protein Chem.*, 2000, **55**, 161.
61 A. Merz, M. C. Yee, H. Szadkowski, G. Pappenberger, A. Crameri, W. P. Stemmer, C. Yanofsky and K. Kirschner, *Biochemistry*, 2000, **39**, 880.
62 L. Giver, A. Gershenson, P. O. Freskgard and F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 12809.
63 B. Spiller, A. Gershenson, F. H. Arnold and R. C. Stevens, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 12305.
64 A. Gershenson, J. A. Schauerte, L. Giver and F. H. Arnold, *Biochemistry*, 2000, **39**, 4658.
65 H. Flores and A. D. Ellington, *J. Mol. Biol.*, 2002, **315**, 325.
66 A. M. Klibanov, *Nature*, 2001, **409**, 241.
67 J. C. Moore and F. H. Arnold, *Nat. Biotechnol.*, 1996, **14**, 458.
68 F. H. Arnold and J. C. Moore, *Adv. Biochem. Eng. Biotechnol.*, 1997, **58**, 1.
69 I. Matsumura, J. B. Wallingford, N. K. Surana, P. D. Vize and A. D. Ellington, *Nat. Biotechnol.*, 1999, **17**, 696.
70 C. Brennan, K. Christianson, T. Surowy and W. Mandecki, *Protein Eng.*, 1994, **7**, 509.
71 C. A. Brennan, K. Christianson, M. A. La Fleur and W. Mandecki, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 5783.
72 D. Legendre, P. Soumillion and J. Fastrez, *Nat. Biotechnol.*, 1999, **17**, 67.
73 G. P. Smith and J. K. Scott, *Methods Enzymol.*, 1993, **217**, 228.
74 C. Schmidt-Dannert, D. Umeno and F. H. Arnold, *Nat. Biotechnol.*, 2000, **18**, 750.
75 Y. X. Zhang, K. Perry, V. A. Vinci, K. Powell, W. P. Stemmer and S. B. Del Cardayre, *Nature*, 2002, **415**, 644.