Engineering Enzymes for Biocatalysis

Paul A. Dalby*

The Advanced Centre for Biochemical Engineering, Department of Biochemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK

Abstract: Protein engineering techniques have been available for over two decades beginning with the development of methods for genetic engineering. Since that time, the engineering of enzymes has advanced rapidly along with a revolution in the range and efficiency of new techniques and strategies for designing and evolving proteins in the laboratory. While recent advances in high-throughput screening techniques are permitting larger libraries of enzymes to be screened more rapidly, a combination of new genetic tools and computational methods are enabling the efficient application of random mutagenesis targeted to areas of enzyme structures that are more likely to elicit the desired enhancement of their biocatalytic properties. Meanwhile the rational design of enzyme properties, in particular, by computational design is showing great potential. This review article summarises recent and important patents relating to the engineering of enzymes for biocatalysis.

INTRODUCTION

Natural enzymes are capable of catalysing reactions with up to 10^{17}-fold rate accelerations [1], and with exquisite control of regio- and stereochemistry. Such control makes the use of enzymes very attractive as an alternative to traditional catalysts in the synthesis of complex and high-value molecules, especially where chemical routes are difficult to implement [2,3]. Enzymes have evolved over billions of years to operate most effectively under physiological conditions, on a narrow range of natural substrates, and usually at concentrations in the low mM range. By contrast, an efficient industrial synthetic process may require a biocatalytic enzyme to operate on non-natural substrates, and under conditions in which the enzyme becomes unstable or inactive, such as extremes of temperature, pH and pressure, after repeated or prolonged use, or in the presence of organic solvents which facilitate substrate solubility or product extraction [4,5]. As a result of these requirements, by far the majority of naturally occurring enzymes are not suitable for industrial-scale biocatalytic processes without further modification of the enzyme itself. Many examples of the engineering of enzymes for their improved use as biocatalysts have been published including: increased activity on novel substrates [6,7]; altered enantioselectivity [8-10]; enhanced stability at extreme temperatures [11,12], pH [13] or in organic solvents [14,15]; and the preference for alternative enzyme cofactors [16]. Many more examples have been reviewed in detail [5,17,18] and cannot all be discussed here.

Two broad approaches have been taken to engineer the amino-acid sequences of enzymes, namely rational design and directed evolution. Rational design applies the growing knowledge of enzyme structure and function to make one or more amino-acid changes that are predicted to elicit the desired improvements to enzyme function. Such knowledge is often based upon the bioinformatical analysis of protein sequences or amino-acid propensities, generalised rules derived by characterising the effect of mutations upon enzyme properties, and by implementation of molecular potential functions that enable the effect of mutations upon structure to be predicted [19-21].

On the other hand, directed evolution techniques mimic natural evolution processes such as random mutagenesis [22] and sexual recombination [23-26]. Just as Nature has produced its enzymes by the process of evolution without using any 'knowledge' of enzyme structure and function, directed evolution similarly permits us to engineer enzymes without understanding them in great detail. Computational methods also provide a valuable tool for the evolutionary design of proteins. Such tools can potentially be used to obtain useful information from protein sequences and structures that can guide the random mutagenesis of proteins making directed evolution more efficient [27]. Alternatively, the use of molecular potential functions can be used to predict the effects of mutations on protein structure and stability for libraries of enzyme variants generated *in silico*. The latest developments and applications of these tools will also be discussed.

Recent advances in both the rational and evolutionary design of enzymes will be outlined below. However, these two general approaches are not mutually exclusive and techniques for improving directed evolution strategies may include the knowledge-based design of enzyme variant libraries, as will also be discussed. Overall, this review is intended to discuss the protein engineering of enzymes for biocatalysis and associated new patents. While many of the methods are of general applicability to protein engineering,

*Address correspondence to this author at The Advanced Centre for Biochemical Engineering, Department of Biochemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK; Tel: +44 020 7679 2962; Fax: +44 020 7383 2348; E-mail: p.dalby@ucl.ac.uk

only those of interest to modifying enzymes for biocatalysis are included. The final section brings together examples of the applications of protein engineering to various biocatalytic enzymes, particularly those appearing as published patents.

RATIONAL DESIGN STRATEGIES

The degree of success of rational design studies depends largely on the enzyme property to be altered. For example, the expression level of a protein is readily improved by ensuring the use of the DNA codons for which the host cell has the greatest preference [28]. Many successful studies have applied rational design to the improved conformational stability of enzymes, through the introduction of disulphide bonds [29,30], proline residues [31,32], or the mutation of protein sequences towards the consensus for a given enzyme family [33]. An alternative approach that uses combinatorial mutation towards the consensus sequence is discussed later as a directed evolution strategy [34]. Since these successful applications, progressively more challenging goals have been achieved using rational design.

An ability to improve the aqueous solubility of membrane-bound enzymes is of potential use in the application of these enzymes as biocatalysts, and also for determining their three-dimensional structures which would critically enable further enzyme design work. An automated computational design strategy has recently been demonstrated to effectively solubilise two membrane-associated proteins, phospholamban and KcsA [35]. In the example of phospholamban (PLB), the authors first predicted the sites for mutagenesis, in the absence of a PLB structure [36]. They used the experimentally observed effects of mutations upon pentameric protein structure formation, to define a perturbation index as a sinusoidal function of residue position in helices. This enabled eleven solvent exposed residues to be predicted as shown in Fig. (**1**). While one residue, F32, was mutated to a more chromogenic tyrosine residue to assist experimental characterisation, the remaining ten were subjected to an energy minimisation designed to optimise the net free energy, -helical propensity, and water solubility of the protein. Sequence diversity was introduced using a sequence entropy term in the potential function. In the PLB example, the new water-soluble variant was found to retain the pentameric form and degree of helical structure, and also to be stabilised upon phosphorylation, in the same manner as the wild-type PLB. More recently, the formation of this water-soluble variant of PLB has subsequently enabled the crystal structure of its tetrameric form to be obtained (Fig. (**1**)), providing information about the PLB structure that was previously difficult to obtain [37]. While the techniques described are limited to the solubilisation of helical membrane-associated proteins, there is a clear potential for the application of this approach to the solubilisation of membrane associated enzymes with highly -helical content.

Some striking examples of the *de novo* design of protein structures have been achieved based on existing knowledge [38-41], and using computational design algorithms [42-44]. Computational design has been successfully used to introduce binding functionality into natural protein structures, such as the design of TNT, L-lactate, D-lactate, and serotonin binding-sites into *Escherichia coli* periplasmic
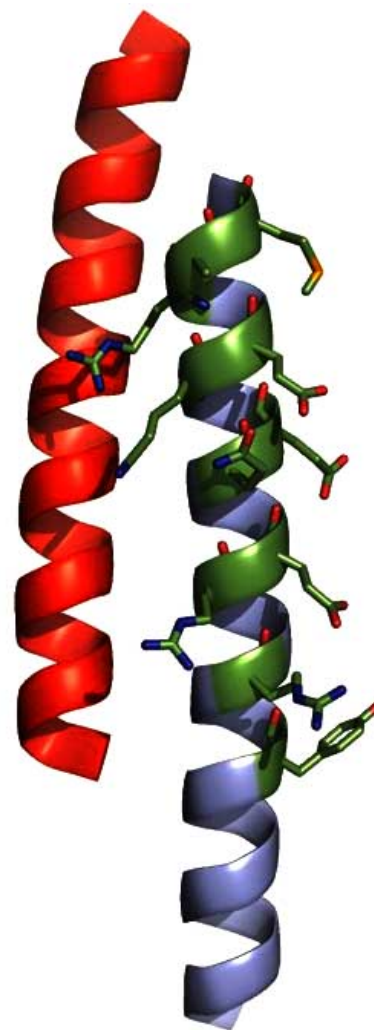
Fig. (1). Water-soluble phospholamban structure.
One half of the tetrameric water-solubilised phospholamban structure (1YOD.pdb), showing antiparallel helices. Residues identified as lipid-exposed by mutagenesis studies and a perturbation index analysis are indicated (green) for one helix only.

binding proteins (PBP) [45,46], and also to redesign the DNA-sequence specificity of an endonuclease [47]. Furthermore, computational design has been used to add binding functionality, such as calcineurin-binding [48], to protein scaffolds that were themselves obtained previously by rational design.

Enzyme functionality has also been introduced into protein scaffolds using computational design methods, demonstrating the potential application of such methods to biocatalysis. For example, the ability to hydrolyse para-nitrophenyl acetate ester has been engineered into the catalytically inert thioredoxin protein scaffold [49]. In a spectacular example, Hellinga and coworkers introduced triose phosphate isomerase (TIM) activity into the bacterial ribose-binding protein (RBP), a periplasmic receptor that has no known catalytic activity, using the computational design

algorithm described above [45,50]. The nascent enzyme functionality was then further improved by experimental directed evolution to achieve a rate enhancement of 3.4×10^5 over the uncatalysed reaction.

The *de novo* design of a protein structure that also has partial enzyme functionality has been demonstrated for a four-helix bundle with stoichiometric ferridoxase activity, obtained by introducing a diiron-binding site [51]. However, the creation of an enzyme functionality that has an efficiency comparable with natural enzymes, remains elusive using rational design techniques alone, most likely due to the greater complexity of the problem, and the reliance on static protein structure models and transition-state theory for predicting enzyme structure-function relationships. These models omit the potentially significant contributions of enzyme dynamics and quantum tunnelling towards catalysis [52,53]. Recent efforts to study the function of enzymes using molecular dynamics simulations and experimental probes of structural dynamics are expected to lead to improved rational design methods for enzyme functionality [54]. While the rational design of a highly efficient catalytic activity into an inert protein scaffold remains challenging, it is worth noting that the improvement of an existing enzyme activity using computational design methods has been recently demonstrated [55]. In this example, the activity of chorismate mutase from *E. coli* was improved by 50% despite modelling an *ab initio* calculated transition-state structure bound into only a static active-site configuration of the enzyme.

DIRECTED EVOLUTION STRATEGIES

While rational design methods seek to design beneficial mutations or protein sequences by applying empirically derived rules or theoretical models, directed evolution uses a combinatorial approach to create libraries of enzymes from which enhanced variants can be identified experimentally. The two most important aspects to consider when devising directed evolution programmes are: a) the availability of suitable screening or selection-based methods for identifying 'hits' with improvements of the desired enzyme property; and b) an appropriate mutagenesis strategy given the extent of change required in the target property, and the existing availability of information relating to the target enzyme structure and function. These two aspects are discussed below along with recent related innovations that impact on the application of directed evolution to enzyme biocatalysis.

SCREENING AND SELECTION METHODS

For directed evolution to be successful and efficient, it is essential that good screening or selection tools are employed for accurately identifying enzymes with improvements in the desired properties. Furthermore, the screening or selection-based method available for the target activity or enzyme property critically dictates the enzyme variant library size that can be practically tested and, therefore, the most appropriate strategy for library design. In general, the ability to isolate variants from larger libraries will improve the chances of success in finding an enzyme with the desired enhanced properties. Screening methods analyse individual colonies cultured on agar plates, in microtitre plates, or in microfluidic chips [56], and are typically used to screen in

the range of 10^3-10^6 individual enzyme variants [57]. However, selection-based methods can use phage display, cell surface display, ribosomal display, or plasmid display, to isolate high-performing enzymes from libraries ranging up to 10^{14} variants. While selection-based methods enable much greater sequence diversity to be explored, they rely upon indirect selection for catalytic activity such as binding to transition-state analogues, whereas screening methods bring the advantage of direct evaluation of catalytic proficiency [58]. Screening methods also permit enzyme variants to be assessed in accurately controlled, and non-physiological environments [4] which may otherwise prohibit the use of selection-based techniques.

The gap between library sizes that can be practicably analysed by screening and by selection-based techniques is gradually being diminished, for example by using fluorescence-activated cell-sorting (FACS) of cell surface displayed enzymes [59], capable of sorting up to 10^7 enzyme variants per hour. A recent high-throughput screening technique has been developed in which water-in-oil-in-water emulsion droplets are sorted by FACS as outlined in Fig. (**2**), to identify the best enzyme variants [58,60]. This technique takes advantage of *in vitro* compartmentalization, in which library encoding DNA and an *in vitro* transcription/ translation reaction mixture create enzyme variants within an emulsion droplet [61]. The droplets also contain a fluorogenic enzyme substrate, such that enzyme variants can then be assessed and sorted by FACS. The DNA that encodes the highest performing variants can then also be isolated from the same droplet and used for sequencing. In an alternative version of this technique, single bacterial cells expressing enzyme variants cytoplasmically or displayed on the cell surface, can be encapsulated in water-in-oil-in-water emulsions and screened by FACS [62].

LIBRARY CONSTRUCTION

The first methods of directed evolution employed mutagenesis that was targeted across entire genes by error-prone PCR [63] or recombination of homologous genes by DNA shuffling [23,64-66]. While these techniques provided the groundbreaking benefits of being able to enhance the properties of enzymes without the need for any knowledge of the enzyme structure or its relationship to enzyme function, they still suffer from a number of deficiencies that present a barrier to their wider application. For example, the gains in enzyme activity or other desired properties obtained by these techniques, may not always sufficiently improve an industrial biocatalytic process such that an investment in them is seen as worthwhile. On the other hand, performing directed evolution experiments require a degree of experience sufficient to overcome the difficulties in creating large libraries with techniques that depended on inefficient DNA ligation reactions [67].

Since the early directed evolution experiments a number of advances have been made that help to overcome these issues, in particular, the development of mutagenesis techniques that avoid the need for DNA ligations. One example was the introduction of mutator strains which enable mutations to accumulate in plasmids harboured within a bacterial strain deficient in DNA repair mechanisms [68]. However, such strains were of limited use as the chromo-
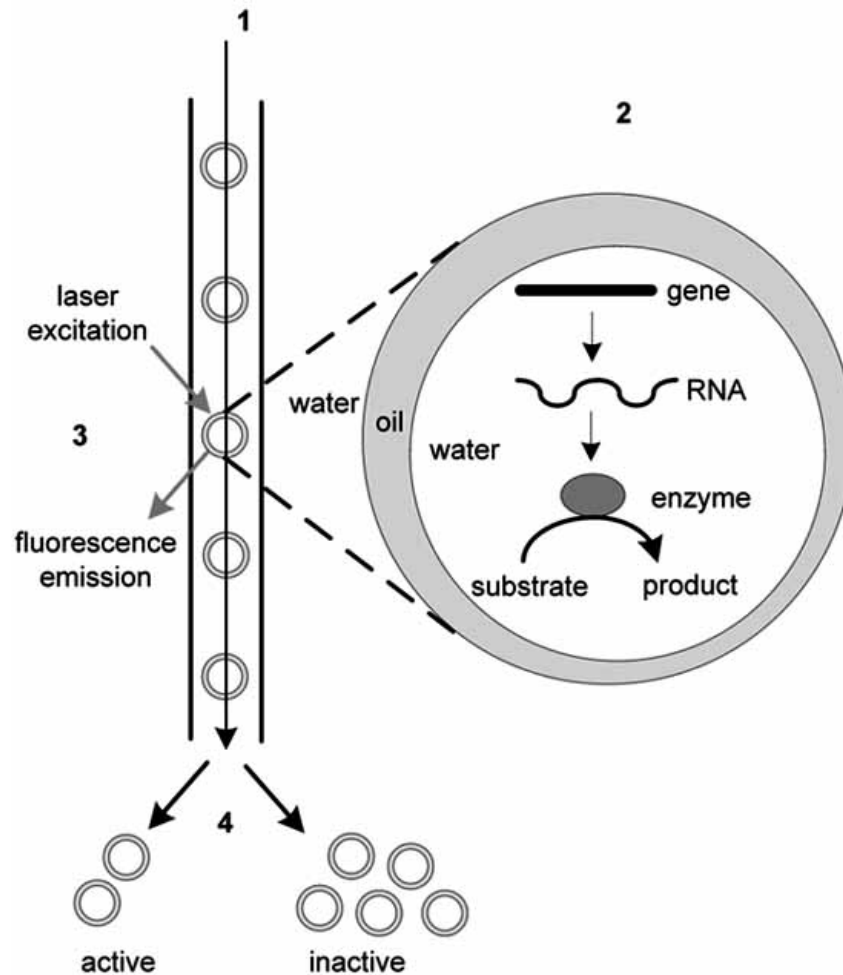
Fig. (2). Selection of active enzymes in double emulsion microdroplets by fluorescence-activated cell sorting (FACS).
(1) Water-in-oil-in-water emulsion droplets are passed through a fluorescence-activated cell sorter. (2) Each of these double emulsion droplets contains a variant gene which is transcribed and translated to a mutant enzyme. Active enzymes convert fluorogenic substrate into fluorescent product. (3) Laser excitation produces fluorescence for double emulsion droplets containing active enzyme. (4) Droplets are sorted by FACS into those containing active and inactive enzyme variants.

somal DNA of the host strain was also susceptible to mutagenesis, resulting in long-term instability in cell viability. A recent technique significantly reduces this problem by using a highly error-prone DNA polymerase that preferentially mutates ColE1 plasmids [69,70]. The error-prone DNA polymerase was created by introducing three point mutations based upon homology with Taq polymerase I. It has been shown to introduce an 80,000-fold greater level of mutagenesis on average, than wild-type DNA polymerase I, and targets plasmid genes with a 400-fold preference over chromosomal gene mutagenesis.

Many developments have also been made that improve the potential gains obtained by directed evolution strategies. While the ability to isolate variants from larger libraries will lead to a greater chance of success in finding an enzyme with the desired enhanced properties, there may be limitations imposed by the library size that can be practicably assessed by the available screening or selection-based methods. It is, therefore, recognised that an ability to improve the quality of enzyme libraries would be beneficial, for example, by promotion of the diversity or uniqueness of clones [71], or

by reducing the number of enzyme residues subjected to random mutagenesis, without incurring any loss in evolutionary potential.

Most DNA shuffling methods utilise the homology between parent sequences to reassemble fragments that have been generated using either restriction-enzyme digests, or short bursts of PCR. Consequently it is difficult to recombine sequences with less than approximately 70% DNA sequence similarity, thus potentially limiting the range of genetic diversity, and therefore, enzyme functionality that can be explored. Recent methods have enabled the non-homologous random recombination (NRR) of DNA sequences [26,72,73]. While most NRR methods typically achieve only a single crossover [26,72], a few enable multiple crossovers to occur, including a technique in which DNA hairpins are added to a mixture of DNA fragments before reassembly by ligation with T4 DNA ligase [74,75]. The use of T4 DNA ligase enables the recombination of DNA fragments independently of their homology, whereas the addition of the DNA hairpins, which can only ligate at one end, permits the range of DNA sequence lengths obtained to be carefully controlled.

In an alternative approach to NRR, a method that mimics exon shuffling [76] has been described in an example which recombined two distantly related polymerase genes, to identify novel DNA polymerases [77,78]. The approach, called structure-based combinatorial protein engineering (SCOPE), requires structural knowledge or sequence analysis of the parent enzymes to design PCR primers that enable recombination to occur at defined structural boundaries.

Amino-acid mutations introduced by traditional error-prone PCR tend to be biased towards those that can be obtained by single-base substitutions, typically accessing only 5.7 alternative residues on average [79]. Greater sequence diversity can be obtained by applying saturation mutagenesis to one or more specific sites using degenerate oligonucleotides that encode for potentially all twenty amino-acids. In one approach, every single site in a protein can be systematically mutagenised independently to identify the optimal residues at each position [80]. The best mutations can then be recombined with the aim of combining the enhancements endowed by each individual mutation.

Another recent method enables sequence saturation mutagenesis (SeSaM) to occur with less bias than error-prone PCR, at random positions within the targeted sequence, yet with only a single 1-3bp mutation appearing in each variant [81]. The method, outlined schematically in Fig. (**3**), creates random-length DNA gene fragments by PCR doped with -phosphothioate nucleotides (dNTP S), and selective cleavage of the sites of dNTP S incorporation using iodine. The 3' ends of each DNA fragment are then 'tailed' with typically 1-3 units of the universal base deoxy-inosine (dITP), before elongation to the full length genes by PCR with a single-stranded template. A subsequent PCR to amplify the product allows promiscuous incorporation of standard bases at only the sites of dITP in the template DNA.

The incorporation of random saturation mutations at specific multiple sites provides the potential of identifying pairs of mutations that result in synergistic effects on enzyme properties. This type of library can be achieved using oligonucleotide PCR primers that encode for potentially all twenty amino-acids at each position, in a modification [82] of the original Quikchange® protocol as supplied by Stratagene (La Jolla, CA). The original protocol employs Pfu DNA polymerase to extend mutagenic primers annealed to a template plasmid [83]. This then creates a non-methylated nicked plasmid progeny containing mutations defined by the oligonucleotide primers. Subsequent digestion with the methylated-DNA specific DpnI restriction enzyme then removes the methylated parental plasmid DNA that is obtained from cell lysates, leaving only the mutated plasmid DNA progeny. The improved method significantly enhances the ability to introduce mutations at multiple sites by including a blend of Pfu DNA polymerase with Taq polymerase, Pfu flap endonuclease (PfuFEN-1), and Taq DNA ligase in the Quikchange® PCR reaction. Adaptation of the Quikchange methods are also available for performing site-directed mutagenesis of genes in plasmids, with improved transformation directly into *Bacillus* strains, and without the need for an intermediate transformation of *E. coli* [84].
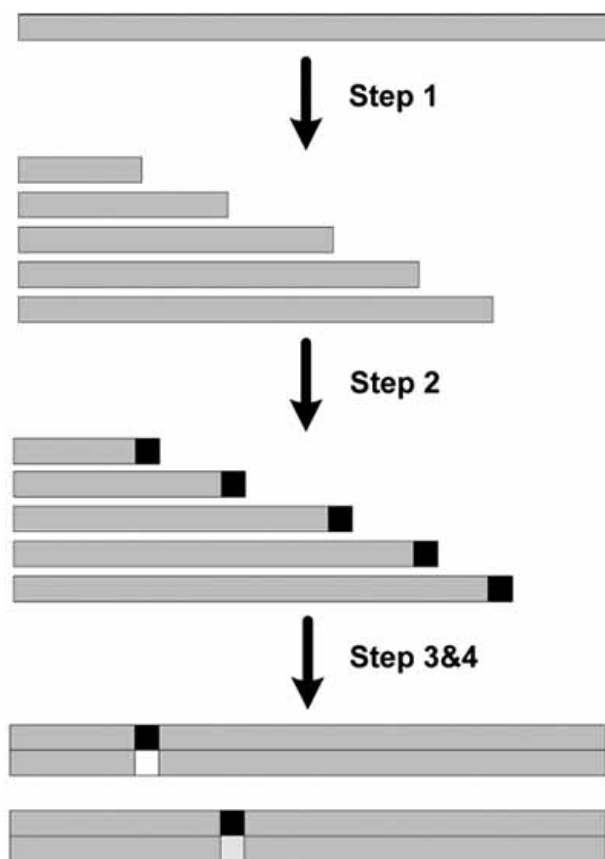
Fig. (3). Sequence saturation mutagenesis (SeSaM) method.
The SeSaM method is schematically outlined. Step 1: creation of random length DNA gene fragments by PCR doped with -phosphothioate nucleotides (dNTP S), and selective cleavage of the sites of dNTP S incorporation using iodine. Step 2: Tailing of 3' ends with the universal base deoxyinosine (dITP) (black squares). Step 3: Elongation to the full length genes by PCR with a single-stranded template. Step 4: PCR to amplify the product, allowing promiscuous incorporation of standard bases (white & gray squares) at sites that base-pair only with the dITP in the template DNA.

Mutagenesis at specific sites of a plasmid DNA generally makes use of PCR for amplification of mutagenised DNA, as discussed in the examples above. However, the use of mismatch endonucleases, mixed with a proofreading poly-merase, dNTPs and a ligase, potentially obviates the PCR step by simply excising and repairing mismatched DNA in the heteroduplex formed between parental template plasmid

and a mutagenic oligonucleotide [85]. This new method presents an efficient alternative for site-directed mutagenesis, and also for the recombination of point mutations from several gene variants.

The functionality represented by the standard twenty amino-acids may be limiting in terms of engineering novel enzyme functions. For example, the incorporation of non-natural amino-acids into proteins may potentially permit enzyme chemistries that are not observed in Nature. It is now possible, through groundbreaking work, to incorporate non-natural amino-acids into proteins at specific sites during their expression from bacterial cells [86,87]. This was achieved by the directed evolution of a tyrosyl t-RNA synthetase to enable it to charge the t-RNA which is complementary to the little used amber codon TAG, with a non-natural amino-acid.

IDENTIFYING TARGET SITES FOR MUTAGENESIS

Alongside the development of techniques for creating different types of library as discussed above, the methods available for identifying key residues within an enzyme to target with mutagenesis have also advanced considerably. Decreasing the redundancy of enzyme libraries by targeting mutagenesis has the potential to allow more useful sequence-space to be searched, and makes more efficient use of the practical limits often imposed by the throughput of analytical tools.

Two general approaches are taken. In the first, aligned sequence information from homologous enzymes can be used to identify residues that vary in Nature, which may be linked to various enzyme properties. It has recently been observed that the consensus protein sequence for a particular enzyme has greater stability in many cases than the wild-type enzymes [88]. A recent method for improving enzyme stability takes advantage of this observation by creating libraries of enzymes that contain variations only at sites that deviate from the consensus sequence [89]. Interestingly, a similar method has also recently been used to invert the cofactor specificity of a lactate dehydrogenase enzyme from NAD^+ to $NADP^+$ [34].

The second general approach for identifying key residues uses computational techniques, often though not necessarily, the same as those applied to rational design. In these examples, the conformational stabilities of enzyme variants generated *in silico*, are estimated. The degree of loss in conformational stability estimated by such a method has been shown to correlate well with mutagenic hot-spots that were determined experimentally, and could therefore potentially be used for identifying amino acids in a protein that are tolerant to mutagenesis [90-92].

A similar concept has been used for pre-screening libraries *in silico* in a technique dubbed Protein Design Automation (PDA), which eliminates sequences that are incompatible with the protein fold of the target enzyme [93,94]. In this example, the $7x10^{23}$ potential variants created by *in silico* mutagenesis of nineteen active-site residues of -lactamase, were reduced down to just 172,800. The remaining variants were designed and screened experimentally in a single round of directed evolution to produce entirely novel enzyme variants with a 1280-fold increase in cefotaxime resistance.

EXAMPLES OF ENZYME ENGINEERING FOR BIOCATALYSIS

Many enzymes have been modified by protein engineering methods and some recently patented examples are summarised below.

The targeted random mutagenesis of seven amino-acid residues of human butyrylcholinesterase expressed in mammalian cells, was recently used to evolve an increased activity towards cocaine hydrolysis [95]. Such an enzyme has therapeutic potential for treating severe cases of cocaine toxicity. Seven residues that line the active-site gorge when aligned to the available structure of the homologous acetylcholinesterase, were chosen for mutagenesis. Previous biochemical data that identified sites important for cocaine hydrolysis were also used to guide the choice of the seven target sites. Several enzyme variants were found that had up to 100-fold greater activity towards cocaine hydrolysis than the wild-type butyrylcholinesterase.

Using a range of protein engineering techniques, including error-prone PCR, saturation mutagenesis and DNA shuffling, mutants of toluene 4-monooxygenase, toluene-o-xylene monooxygenase and toluene-4-monooxygenase from various *Pseudomonads* have been obtained with improved activities towards benzene, toluene, nitrobenzene, nitrophenols, catechols, o-methoxyphenol, and o-Cresol [96]. Such biocatalytic reactions can be used in the synthesis of phenol, catechol, nitrophenols, nitrocatechols, 3-methoxy-catechol, 1,2,3-trihydroxybenzene, methoxyhydroquinone, methyl-hydroquinone, 4-methylresorcinol, methylhydroquinone, and pyrogallol.

B12-dependent dehydratases with improved reaction kinetics have been obtained using error-prone PCR and oligonucleotide-directed mutagenesis to target the DhaB1 gene, encoding the -subunit of glycerol dehydratase [97]. Such enzymes are useful in the production of glycerol and 1,3-propanediol.

Nitrilases can be used in biocatalytic processes to convert nitriles to carboxylic acids. Recently, error-prone PCR has been used to generate mutants of an *Acidovorax facilis* nitrilase having improved activity towards 3-hydroxynitriles [98]. Isolated enzyme variants were capable of up to 1.9-fold improved activities when used to convert 3-hydroxynitrile to 3-hydroxycarboxylic acid, 3-hydroxyvaleronitrile to 3-hydroxyvaleric acid, or 3-hydroxybutyronitrile to 3-hydroxybutyric acid.

In another example highlighting the general applicability of error-prone PCR, two carotenoid biosynthesis genes were simultaneously mutated randomly to improve the production of astaxanthin from beta-carotene [99]. Targeting of the two enzymes, CrtO ketolase and CrtZ hydroxylase, simultaneously resulted in mutations within both genes and a net increase in the astaxanthin biosynthesis yield of up to 20% above that obtained with the wild-type enzyme pathway.

Error-prone PCR has also been used to improve the activity of 5'-xanthylic acid (XMP) aminase, which is potentially useful for the biocatalytic synthesis of 5'-guanylic acid (GMP) from XMP in microbial fermentations [100]. Six improved mutants were found, with between two and six

accumulated mutations, and the best mutant having 3.8-fold greater activity resulting from three mutations. Xylanases are useful enzymes added to animal feed to aid the digestion of feed grains containing xylan. The growing need to sterilise the feed grain requires a xylanase enzyme that can retain activity after heat treatment, as well as being active at physiological digestive conditions (pH 3.5-5.0, 40°C). A recent patent [101] has combined an engineered disulphide bond, previously introduced into a homologous xylanase enzyme, and the Q162H mutation previously reported for xylanases [102]. Previously, the disulphide bond was shown to only marginally increase the xylanase thermostability, whereas the Q162H mutation had no effect on thermostability. However, the two mutations combined together result in a xylanase enzyme with a thermal denaturation midpoint increase of 14°C, and which also retains 40% of the original activity after a 30 minute treatment at 70°C, and at least 30% activity at pH3.5-6, and 40-60°C.

A similar product, 5'-inosinic acid (IMP), is synthesised in an industrial process that uses mutant acid phosphatases from *E. blattae*, for which the phosphomonoesterase activity has been reduced to less than 40% of the wild-type enzyme due to the mutations G74D or I153T [103]. Decreased esterase activity allows the mutant enzymes to synthesise IMP from inosine.

Another enzyme that can be used as a supplement to animal feeds is phytase. This enzyme is useful during digestion for releasing inorganic phosphate from phytate, which contains over 70% of the phosphate in plant material. A variant of the *E. coli* pH 2.5 acid phosphatase has been shown, rather unexpectedly, to exhibit increased phytase activity upon digestion of the enzyme with a mixture of trypsin and pepsin [104].

Alpha-amylase is an enzyme that is commonly used for the degradation of starch during textile and paper desizing, and also as a component of household detergents. Several examples of engineered alpha-amylases have been recently patented, including variants with improved activity and/or thermostability [105,106]; solvent stability [107], and resistance to multimerisation [108].

Polysaccharide lyases have potential uses in polysaccharide sequencing and degradation. A recent patent describes rationally designed variants of the polysaccharide lyase, chondroitinase B, and also potential uses in the inhibition of anticoagulation, angiogenesis and maternal malarial infection [109]. Active-site residues involved in catalysis and substrate binding were targeted by site-directed mutagenesis. While mutagenesis of the catalytic residues were damaging to enzyme activity, the mutation of a substrate binding arginine (R364) to alanine resulted in an altered product profile to that of wild-type enzyme after the digestion of dermatan sulfate [110].

CURRENT & FUTURE DEVELOPMENTS

In recent years, significant advances have been made in the understanding of protein structure and function that have also enabled rational design techniques to become more successful and further reaching than previously. In particular, the use of computational design algorithms with advanced features has made it possible to introduce nascent enzyme functionality into proteins, and in at least one example, to improve the efficiency of an existing enzyme activity. Advances in rapid screening techniques, new methods for creating genetic diversity, and computational tools for predicting hot-spots in protein sequences, have also pushed forward the boundaries that can be reached using directed evolution. The challenge now resides in refining computational methods for more accurate prediction of the effect of mutations on enzyme activity, potentially including the effects of protein dynamics. Meanwhile, improved understanding of natural enzyme evolution and better prediction of important residues within enzymes will enable directed evolution techniques to become much more powerful.

ABBREVIATIONS

CrtO	=	Carotene ketolase gene
CrtZ	=	Carotene hydroxylase gene
dITP	=	Deoxyinosine triphosphate
dNTP S	=	-Phosphothioate nucleotides
FACS	=	Fluorescence-activated cell-sorting
GMP	=	5'-Guanylic acid
IMP	=	5'-Inosinic acid
KcsA	=	Potassium ion-channel protein gene
NAD^+	=	Nicotinamide adenine dinucleotide
NADP	=	Nicotinamide adenine dinucleotide phosphate
NRR	=	Non-homologous random recombination
PCR	=	Polymerase chain reaction
PDA	=	Protein design automation
Pfu	=	*Pyrococcus furiosus*
PLB	=	Phospholamban
RBP	=	Ribose binding protein
SCOPE	=	Structure-based combinatorial protein engineering
SeSaM	=	Sequence saturation mutagenesis
TIM	=	Triose phosphate isomerase
TNT	=	Trinitrotoluene
XMP	=	5'-Xanthylic acid

ACKNOWLEDGEMENTS

REFERENCES

[1]   Radzicka A, Wolfenden R. A proficient enzyme. Science 1995; 267: 90-93.
[2]   Schoemaker HE, Mink D, Wubbolts MG. Dispelling the myths-biocatalysis in industrial synthesis. Science 2003; 299: 1694-1697.
[3]   Straathof AJ, Panke S, Schmid A. The production of fine chemicals by biotransformations. Curr Opin Biotech 2002; 13: 548-556.
[4]   Hibbert EG, Baganz F, Hailes HC, *et al.* Directed evolution of biocatalytic processes. Biomol Eng 2005; 22: 11-19.

[5]   Sylvestre J, Chautard H, Cedrone F, Delcourt M. Directed evolution of biocatalysts. Org Process Res Dev 2006; 10: 562-571.

[6]   Glieder A, Farinas ET, Arnold FH. Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. Nat Biotechnol 2002; 20: 1135-1139.

[7]   Yoshikuni Y, Ferrin TE, Keasling JD. Designed divergent evolution of enzyme function. Nature 2006; 440: 1078-1082.

[8]   Reetz MT, Zonta A, Schimossek K, Liebeton K, Jaeger KE. Creation of enantioselective biocatalysts for organic chemistry by *in vitro* evolution. Angew Chem Int Edit 1997; 36: 2830-2832.

[9]   May O, Nguyen PT, Arnold FH. Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. Nat Biotechnol 2000; 18: 317-320.

[10]   Carr R, Alexeeva M, Enright A, Eve TSC, Dawson MJ, Turner NJ. Directed evolution of an amine oxidase possessing both broad substrate specificity and high enantioselectivity. Angew Chem Int Edit 2003; 42: 4807-4810.

[11]   Merz A, Yee MC, Szadkowski H, *et al.* Improving the catalytic activity of a thermophilic enzyme at low temperatures. Biochemistry 2000; 39: 880-889.

[12]   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.

[13]   Cherry JR. Directed evolution of microbial oxidative enzymes. Current Opin Biotech 2000; 11: 250-254.

[14]   Moore JC, Arnold FH. Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. Nat Biotechnol 1996; 14: 458-467.

[15]   Hao JJ, Berry A. A thermostable variant of fructose bisphosphate aldolase constructed by directed evolution also shows increased stability in organic solvents. Protein Eng Des Sel 2004; 17: 689-697.

[16]   Joo H, Lin Z, Arnold FH. Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. Nature 1999; 399: 670-673.

[17]   Dalby PA. Optimising enzyme function by directed evolution. Curr Opin Struct Biol 2003; 13: 500-505.

[18]   Turner NJ. Directed evolution of enzymes for applied biocatalysis. Trends Biotech 2003; 21: 474-478.

[19]   Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Res 2005; 33: W382-W388.

[20]   Monsellier E, Bedouelle H. Improving the stability of an antibody variable fragment by a combination of knowledge-based approaches: validation and mechanisms. J Mol Biol 2006; 362: 580-593.

[21]   Poole AM, Ranganathan R. Knowledge-based potentials in protein design. Curr Opin Struct Biol 2006; 16: 508-513.

[22]   Chen K, Arnold FH. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc Natl Acad Sci USA 1993; 90: 5618-5622.

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

[24]   Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. Molecular evolution by staggered extension process (StEP) *in vitro* recombination. Nat Biotechnol 1998; 16: 258-261.

[25]   Coco WM, Levinson WE, Crist MJ, *et al.* DNA shuffling method for generating highly recombined genes and evolved enzymes. Nat Biotechnol 2001; 19: 354-359.

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

[27]   Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH. Evolving strategies for enzyme engineering. Curr Opin Struct Biol 2005; 15: 447-452.

[28]   Hale RS, Thompson G. Codon optimization of the gene encoding a domain from human type 1 neurofibromin protein results in a threefold improvement in expression level in *Escherichia coli*. Protein Expres Purif 1998; 12: 185-188.

[29]   Shimaoka M, Lu CF, Salas A, Xiao T, Takagi J, Springer TA. Stabilizing the integrin alpha M inserted domain in alternative conformations with a range of engineered disulfide bonds. Proc Natl Acad Sci USA 2002; 99: 16737-16741.

[30]   Clarke J, Fersht AR. Engineered Disulfide Bonds As Probes of the Folding Pathway of Barnase - Increasing the Stability of Proteins Against the Rate of Denaturation. Biochemistry 1993; 32: 4322-4329.

[31]   Suzuki Y. The proline rule - A strategy for protein thermal stabilization. P Jpn Acad B-Phys 1999; 75: 133-137.

[32]   Choi EJ, Mayo SL. Generation and analysis of proline mutants in protein G. Protein Eng Des Sel 2006; 19: 285-289.

[33]   Lehmann M, Wyss M. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Current Opin Biotech 2001; 12: 371-375.

[34]   Flores H, Ellington AD. A modified consensus approach to mutagenesis inverts the cofactor specificity of *Bacillus stearothermophilus* lactate dehydrogenase. Protein Eng Des Sel 2005; 18: 369-377.

[35]   Slovic, A.M., Summa, C.M., Saven, J.G., DeGrado, W.F.: US2004215400 **(2004)**.

[36]   Slovic AM, Summa CM, Lear JD, DeGrado WF. Computational design of a water-soluble analog of phospholamban. Protein Sci 2003; 12: 337-348.

[37]   Slovic AM, Stayrook SE, North B, DeGrado WF. X-ray structure of a water-soluble analog of the membrane protein phospholamban: Sequence determinants defining the topology of tetrameric and pentameric coiled coils. J Mol Biol 2005; 348: 777-787.

[38]   DeGrado WF, Wasserman ZR, Lear JD. Protein design, a minimalist approach. Science 1989; 243: 622-628.

[39]   Hecht MH, Richardson JS, Richardson DC, Ogden RC. Denovo Design, Expression, and Characterization of Felix - A 4-Helix Bundle Protein of Native-Like Sequence. Science 1990; 249: 884-891.

[40]   Kamtekar S, Schiffer JM, Xiong HY, Babik JM, Hecht MH. Protein Design by Binary Patterning of Polar and Nonpolar Amino-Acids. Science 1993; 262: 1680-1685.

[41]   Quinn TP, Tweedy NB, Williams RW, Richardson JS, Richardson DC. Betadoublet - De-Novo Design, Synthesis, and Characterization of A Beta-Sandwich Protein. Proc Natl Acad Sci USA 1994; 91: 8747-8751.

[42]   Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. Science 2003; 302: 1364-1368.

[43]   Kortemme T, Ramirez-Alvarado M, Serrano L. Design of a 20-amino acid, three-stranded beta-sheet protein. Science 1998; 281: 253-256.

[44]   Summa CM, Rosenblatt MM, Hong JK, Lear JD, DeGrado WF. Computational de novo design, and characterization of an A(2)B(2) diiron protein. J Mol Biol 2002; 321: 923-938.

*[45]   Hellinga, H.W., Looger, L.L., Dwyer, M.A.: US2004229290 **(2004)**.

[46]   Looger LL, Dwyer MA, Smith JJ, Hellinga HW. Computational design of receptor and sensor proteins with novel functions. Nature 2003; 423: 185-190.

[47]   Ashworth J, Havranek JJ, Duarte CM, *et al.* Computational redesign of endonuclease DNA binding and cleavage specificity. Nature 2006; 441: 656-659.

[48]   Ghirlanda G, Lear JD, Lombardi A, DeGrado WF. From synthetic coiled coils to functional proteins: Automated design of a receptor for the calmodulin-binding of calcineurin. J Mol Biol 1998; 281: 379-391.

[49]   Bolon DN, Mayo SL. Enzyme-like proteins by computational design. P Natl Acad Sci USA 2001; 98: 14274-14279.

[50]   Dwyer MA, Looger LL, Hellinga HW. Computational design of a biologically active enzyme. Science 2004; 304: 1967-1971.

[51]   Pasternak A, Kaplan S, Lear JD, DeGrado WF. Proton and metal ion-dependent assembly of a model diiron protein. Protein Sci 2001; 10: 958-969.

[52]   Sutcliffe MJ, Scrutton NS. Enzyme catalysis: over-the-barrier or through-the-barrier? Trends Biochem Sci 2000; 25: 405-408.

[53]   Sutcliffe MJ, Scrutton NS. A new conceptual framework for enzyme catalysis. Hydrogen tunnelling coupled to enzyme dynamics in flavoprotein and quinoprotein enzymes. Eur J Biochem 2002; 269: 3096-3102.

[54]   Agarwal PK. Enzymes: An integrated view of structure, dynamics and function. Microb Cell Fact 2006; 5: 2.

[55]   Lassila JK, Keeffe JR, Oelschlaeger P, Mayo SL. Computationally designed variants of *Escherichia coli* chorismate mutase

show altered catalytic activity. Protein Eng Des Sel 2005; 18: 161-163.

[56] Fu AY, Spence C, Scherer A, Arnold FH, Quake SR. A microfabricated fluorescence-activated cell sorter. Nat Biotechnol 1999; 17: 1109-1111.

[57] Aharoni A, Griffiths AD, Tawfik DS. High-throughput screens and selections of enzyme-encoding genes. Curr Opin Chem Biol 2005; 9: 210-216.

[58] Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT, Griffiths AD. High-throughput screening of enzyme libraries: *In vitro* evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. Chem Biol 2005; 12: 1291-1300.

[59] 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.

*[60] Griffiths, A.D., Weitz, D., Link, D., Ahn, K., Bibette, J.: US2006078888 **(2006)**.

[61] Tawfik, D.S., Bernath, K., Magdassi, S., Peisajovich, S.G.: WO06051552 **(2006)**.

[62] Aharoni A, Amitai G, Bernath K, Magdassi S, Tawfik DS. High-throughput screening of enzyme libraries: Thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. Chem Biol 2005; 12: 1281-1289.

[63] Zhou YH, Zhang XP, Ebright RH. Random mutagenesis of gene-sized DNA-molecules by use of PCR with Taq DNA-polymerase. Nucleic Acids Res 1991; 19: 6052-6052.

[64] Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. Molecular evolution by staggered extension process (StEP) *in vitro* recombination [see comments]. Nat Biotechnol 1998; 16: 258-261.

[65] Stemmer, W.P.: US6995017 **(2006)**.

[66] Short, J.M.: US2004248143 **(2004)**.

[67] Hibbert EG, Dalby PA. Directed evolution strategies for improved enzymatic performance. Microb Cell Fact 2005; 4: 29.

[68] Greener A, Callahan M, Jerpseth B. An efficient random mutagenesis technique using an *Escherichia coli* mutator strain. Mol Biotechnol 1997; 7: 189-195.

*[69] Camps M, Naukkarinen J, Johnson BP, Loeb LA. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. P Natl Acad Sci USA 2003; 100: 9727-9732.

[70] Camps, M., Loeb, L.A.: WO03102213 **(2003)**.

[71] Drummond DA, Iverson BL, Georgiou G, Arnold FH. Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. J Mol Biol 2005; 350: 806-816.

[72] Sieber V, Martinez CA, Arnold FH. Libraries of hybrid proteins from distantly related sequences. Nat Biotechnol 2001; 19: 456-460.

[73] Lutz S, Ostermeier M, Moore GL, Maranas CD, Benkovic SJ. Creating multiple-crossover DNA libraries independent of sequence identity. P Natl Acad Sci USA 2001; 98: 11248-11253.

[74] Bittker JA, Le BV, Liu DR. Nucleic acid evolution and minimization by nonhomologous random recombination. Nat Biotechnol 2002; 20: 1024-1029.

[75] Liu, D.R., Bittker, J.A., Liu, J.M.: US2005260655 **(2005)**.

[76] Kolkman JA, Stemmer WP. Directed evolution of proteins by exon shuffling. Nat Biotechnol 2001; 19: 423-428.

[77] O'Maille PE, Bakhtina M, Tsai MD. Structure-based combinatorial protein engineering (SCOPE). J Mol Biol 2002; 321: 677-691.

[78] O'Maille, P.E., Noel, J.P.: WO05118861 **(2005)**.

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

[80] Delcourt, M.: WO06027471 **(2006)**.

[81] Schwaneberg, U.: WO05035757 **(2005)**.

[82] Cline, J.M., Hogrefe, H.H.: WO03025118 **(2003)**.

[83] Bauer, J.C., Wright, D.A., Braman, J.C., Geha, R.S.: US5932419 **(1999)**.

[84] Leeflang, C., Van der Kleij, W.A.H.: WO04064744 **(2004)**.

[85] Padgett, H.S., Vaewhongs, A.A., Vojdani, F.S., Smith, M.L., Lindbo, J.A., Fitzmaurice, W.P.: WO03066809 **(2003)**.

[86] Wang L, Brock A, Herberich B, Schultz PG. Expanding the Genetic Code of *Escherichia coli*. Science 2001; 292: 498-500.

*[87] Schultz, P.G., Wang, L., Andersen, J.C., Chin, J.W., Liu, D.R., Magliery, T.J., Meggers, E.L., Mehl, R.A., Pastrnak, M., Santoro, S.W., Zhang, Z.: US2005250183 **(2005)**.

[88] Lehmann M, Pasamontes L, Lassen SF, Wyss M. The consensus concept for thermostability engineering of proteins. Biochim Biophys Acta Protein Struct Mol Enzymol 2000; 1543: 408-415.

[89] Aehle, W., Ramer, S.W., Schellenberger, V.: WO05040344 **(2006)**.

[90] 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.

[91] Voigt CA, Mayo SL, Arnold FH, Wang ZG. Computationally focusing the directed evolution of proteins. J Cell Biochem Suppl 2001; Suppl 37: 58-63.

[92] Voigt, C.A., Wang, Z.G., Arnold, F.H., Mayo, S.L.: WO03091835 **(2003)**.

[93] Hayes RJ, Bentzien J, Ary ML, Hwang MY, Jacinto JM, Vielmetter J, Kundu A, Dahiyat BI. Combining computational and experimental screening for rapid optimization of protein properties. P Natl Acad Sci USA 2002; 99: 15926-15931.

*[94] Mayo, S.L., Dahiyat, B.I., Gordon, D.B., Street, A., Su, Y.: US2006019316 **(2006)**.

[95] Watkins, J.D., Pancook, J.D.: US2003153062 **(2003)**.

[96] Wood, T.K., Vardar, G.: US2006051782 **(2006)**.

[97] Der-ing, L.: WO04056963 **(2004)**.

[98] Payne, M.S., Dicosimo, R., O'Keefe, D.P.: WO06023520 **(2006)**.

[99] Tang, X.-S., Cheng, Q., Shyr, J.Y., Tao, L.: WO06072078 **(2006)**.

[100] Pan, J.G., Jung, H.C., Kim, E.J., Lee, H.S., Park, Y.H., Kim, H.S., Han, J.K., Lee, J.N., Oh, K.H., Kim, J.H., Oh, Y.S., Sim, J.I., Hong, K.K., Choi, K.O., Kim, H.S., Baek, M.J., Kang, T.S.: WO06065076 **(2006)**.

[101] Sung, W.L., Tolan, J.S.: US20067060482 **(2006)**.

[102] Sung, W.L., Yaguchi, M., Ishikawa, K.: US5759840 **(1998)**.

[103] Ajinomoto, K.K., Mihara, Y., Utagawa, T., Hideaki, Y., Asano, Y.: WO9637603 **(1996)**.

[104] Lei, X.G.: US2003072844 **(2003)**.

[105] van der Laan, J.M., Aehle, W.: WO9535382 **(1995)**.

[106] Svendsen, A., Kjaerulff, S., Bisgaard-Frantzen, H., Andersen, C.: EP1676913 **(2006)**.

[107] Bessler, C., Wieland, S., Maurer, K.-H.: WO06037484 **(2006)**.

[108] Bessler, C., Wieland, S., Maurer, K.-H.: WO06037483 **(2006)**.

[109] Pojasek, K., Raman, R., Sasisekharan, R.: WO03102160 **(2003)**.

[110] Pojasek K, Raman R, Kiley P, Venkataraman G, Sasisekharan R. Biochemical characterization of the chondroitinase B active site. J Biol Chem 2002; 277: 31179-31186.