A one-pot, simple methodology for cassette randomisation and recombination for focused directed evolution

Aurelio Hidalgo¹², Anna Schließmann¹, Rafael Molina³, Juan Hermoso³ and Uwe T. Bornscheuer¹⁴

¹Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, Ernst-Moritz-Arndt University Greifswald, Felix-Hausdorff-Str. 4, D-17487 Greifswald, Germany, ²Centro de Biología Molecular ‘Severo Ochoa’ (UAM-CSIC), Nicolás Cabrera 1, 28049 Madrid, and ³Grupo de Cristalografía Macromolecular y Biología Estructural, Instituto de Química Física ‘Rocasolano’, Consejo Superior de Investigaciones Científicas (CSIC), Serrano 119, 28006 Madrid, Spain

Protein engineering is currently performed either by rational design, focusing in most cases on only a few positions modified by site-directed mutagenesis, or by directed molecular evolution, in which the entire protein-encoding gene is subjected to random mutagenesis followed by screening or selection of desired phenotypes. A novel alternative is focused directed evolution, in which only fragments of a protein are randomised while the overall scaffold of a protein remains unchanged. For this purpose, we developed a PCR technique using long, spiked oligonucleotides, which allow randomisation of one or several cassettes in any given position of a gene. This method allows over 95% incorporation of mutations independently of their position within the gene, yielding sufficient product to generate large libraries, and the possibility of simultaneously randomising more than one locus at a time, thus originating recombination. The high efficiency of this method was verified by creating focused mutant libraries of Pseudomonas fluorescens esterase I (PFEI), screening for altered substrate selectivity and validating against libraries created by error-prone PCR. This led to the identification of two mutants within the OSCARR library with a 10-fold higher catalytic efficiency towards p-nitrophenyl dodecanolate. These PFEI variants were also modelled in order to explain the observed effects.

Keywords: esterase/focused directed evolution/megaprimer PCR/substrate selectivity

Introduction

Rational design of protein structure and function is highly dependent on the knowledge of the three-dimensional structure of the protein of interest and its corresponding mechanism of action. Moreover, even in the cases where such information is available, predictions of perturbations caused by limited amino acid substitutions are highly uncertain. Directed evolution on the other hand is not subject to structural preconditions. Essentially, the gene encoding the protein of interest is subjected to random mutagenesis to create a library of mutants, which is subsequently screened to identify desired variants. However, random mutagenesis techniques favour the discovery of distant mutations, not because they are better but because they are more likely, as there are more amino acids far from the active site than close to the active site (Morley and Kazlauskas, 2005).

As the number of available protein sequences and structures is tremendously increasing, a more focused approach is feasible by homology studies of related proteins, which can be performed by aligning gene sequences or similar protein structures. This eventually leads to the identification of regions which have to be altered to change the function of the protein, while other areas of the protein must not be changed in order to ensure that the protein folding and function is not negatively affected (Bernhardt et al., 2005). Consequently, this combination of rational protein design and focused directed evolution requires to identify important regions for the function of the protein and suitable methodologies to randomise only selected parts of the protein-encoding gene (Dunn et al., 1998). The concept of CASTing (Reetz et al., 2005, 2006) may effectively help to reduce the size of mutant libraries, but has mostly found application to develop enantioselectivity through the study of residues that hinder the access of the substrate to the active site, considering the type of secondary structure they are located on. However, high-resolution structures of proteins are not always available, and this technique would not be applicable. In this case, random mutagenesis focused around the active site is proposed as an alternative. It has been previously shown that saturation mutagenesis near the binding site was more effective in the alteration of catalytic properties such as enantioselectivity in the Pseudomonas fluorescens esterase I (PFEI) (Park et al., 2005), or in general in the development of an alternate catalytic activity and substrate selectivity (Morley and Kazlauskas, 2005). Comprehensive libraries (~10¹⁰ individuals) have only been screened by in vitro compartmentalisation/FACS, phage display selects 10⁹ individuals in libraries, whereas activity screens are suitable for a maximum of 10⁹ clones using an agar plate assay or 10³ clones using liquid assays (Aharoni et al., 2005). Thus, it is an additional advantage to turn to focused libraries not only because of the increased chance for success, but also for the reduced library size, thus helping overcome limited screening capacities, such as in the case of complex analytical methods (Reetz).

Although it is not the optimum method for generating comprehensive libraries, the most frequently used method on a laboratory scale and as a first approach is still error-prone PCR (epPCR) (Cadwell and Joyce, 1995). However, this method can hardly be used to create ‘focused’ mutants in a relatively short region of a gene, as its mutation frequency (usually ranging from 3 to 10 mutations/kb) is too low, and any attempts to increase it, would involve
experimental conditions (heavy nucleotide imbalance and higher concentrations of manganese ions) resulting in smaller amounts of a heavily biased product (Vartanian et al., 1996; Wong et al., 2006). Considering a (relatively large) region of 33 amino acids to be randomised (corresponding to 100 bp), a 1% mutation frequency in epPCR would statistically lead to only one nucleotide exchange per round of mutation. Due to the technique bias and the degenerated code, the chance to find just one amino acid mutation in the library is thus extremely unlikely. Furthermore, in whole genes or parts of a gene with high GC content, such protocols are less effective or result in bias. For instance, the frequency of replacement of C or G with the most used epPCR protocol (Cadwell and Joyce, 1992) using 0.5 mM MnCl₂ is approximately five times less than that of A or T (Brakmann and Lindemann, 2004).

When the desired mutation rate or quality is not reached, an alternative is to introduce the desired mutations by means of synthetic oligonucleotides carrying mismatches in the target sequence. The use of synthetic-spiked oligos was described as early as 1989 (Hermes et al., 1989), and since then several technical variations (Dunn et al., 1998; Gaytan et al., 1998; Gaytan et al., 2001) have been described as well as design algorithms in order to optimise the resulting library (Arkin and Youvan, 1992; Tomandl et al., 1997; Jensen et al., 1998). Furthermore, mutagenesis studies at an industrial scale are often carried out using synthetic oligonucleotides (Saboulard et al., 2005). On the laboratory scale, the incorporation of a synthetic oligo into a target sequence may be carried out by megaprimer PCR (Kammann et al., 1989; Landt et al., 1990; Sarkar and Sommer, 1990; Sarkar et al., 1990; Lai et al., 2003). This technique is based on the generation of a megaprimer by amplification of part of a gene, introducing the desired mutations in one of the primers. After purification of the megaprimer, it is extended to full length by means of an ‘external’ primer. Already, variations have been introduced to avoid purification of the intermediate megaprimer, using either primers of different annealing temperature (Ke and Madison, 2005), in limiting concentration (Tyagi et al., 2004) or exonuclease I (Nabavi and Nazar, 2005). Such one-pot alternatives have only afforded 80% efficiency and were only targeted at creating a single mutation and not at randomising a cassette within a gene.

The OSCARR methodology (One-pot Simple methodology for CASSette Randomization and Recombination) presented here was established for the gene encoding a PFEI serving as model, for convenience reasons. The PFEI is an α/β fold-hydrolase, commonly used in our group for biocatalytical purposes, with which error-prone methods have had limited success (Henke and Bornscheuer, 1999) or none at all. This held especially true on some parts of the gene, where the GC content rises over 75% and that were shown to be refractory to conventional epPCR, even using a specific DNA polymerase for GC-rich templates (data not shown).

The aim of this work was to establish a simple methodology based on the one-pot megaprimer PCR to randomise a gene in a focused way, i.e. to randomise one or several cassettes in any given position of a gene. To illustrate the applicability of this method, the esterase libraries created were screened for altered substrate range of PFEI.

Materials and methods

Materials

All restriction enzymes were from New England Biolabs (Ipswich, MA, USA). Fast Blue and α-naphthylacetate were purchased from Fluka (Buchs, Switzerland). Deoxynucleotide triphosphate stocks (dNTPs) and 1 kb DNA ladder for agarose electrophoresis were purchased from Carl Roth (Karlsruhe, Germany). Generuler Xpress 1 kb was supplied by Fermentas (Vilnius, Lithuania). Non-spiked oligonucleotide synthesis and DNA automated sequencing were carried out by MWG Biotech GmbH ( Martinsried, Germany). Spiked oligonucleotides were purchased from Thermo Electron GmbH (Ulmm, Germany).

Expression strains

Escherichia coli strains DH5α [supE44 ΔlacU169 (φ80lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1] and BL21(DE3) [hsdS gal (ΔcI857 ind1) Sam7nin5 lacUV5-T7 gene 1)] were used as hosts for genetic manipulation of plasmids and for the overexpression of proteins.

Templates

 Constructs containing the PFEI gene (pfeI, accession number U12.537, PDB 1VA4) in plasmids pGast (pGPFEI) and pET22 (Novagen, pET22PFEI) were used as templates for PCR. On both constructs, site-directed mutagenesis was performed in order to delete the Nar site situated at 363 and at 726 bp in the pfe gene by means of a silent mutation, leading to plasmids pGPFEIΔNar and pET22PFEIΔNar, respectively. Target cassettes for mutagenesis were chosen according to their distance to the catalytic serine (Ser 94), on condition that they were within 10 Å from this residue (Fig. 1), except in the case of cassettes 4 and 5, which are at a distance of 15 Å from the catalytic Ser.

Primers

All primers used are shown in Table I. Long, mutagenic primers were designed that would introduce a stop codon in the final product to yield a truncated product after cloning and expression, if the mutagenesis strategy was successful. Amplification of the wild-type and unspecific primers has been described as well as design algorithms in order to optimise the resulting library (Arkin and Youvan, 1992; Jensen et al., 1998). Furthermore, mutagenesis studies at an industrial scale are often carried out using synthetic oligonucleotides (Saboulard et al., 2005). On the laboratory scale, the incorporation of a synthetic oligo into a target sequence may be carried out by megaprimer PCR (Kammann et al., 1989; Landt et al., 1990; Sarkar and Sommer, 1990; Sarkar et al., 1990; Lai et al., 2003). This technique is based on the generation of a megaprimer by amplification of part of a gene, introducing the desired mutations in one of the primers. After purification of the megaprimer, it is extended to full length by means of an ‘external’ primer. Already, variations have been introduced to avoid purification of the intermediate megaprimer, using either primers of different annealing temperature (Ke and Madison, 2005), in limiting concentration (Tyagi et al., 2004) or exonuclease I (Nabavi and Nazar, 2005). Such one-pot alternatives have only afforded 80% efficiency and were only targeted at creating a single mutation and not at randomising a cassette within a gene.

The OSCARR methodology (One-pot Simple methodology for CASSette Randomization and Recombination) presented here was established for the gene encoding a PFEI serving as model, for convenience reasons. The PFEI is an α/β fold-hydrolase, commonly used in our group for biocatalytical purposes, with which error-prone methods have had limited success (Henke and Bornscheuer, 1999) or none at all. This held especially true on some parts of the gene, where the GC content rises over 75% and that were shown to be refractory to conventional epPCR, even using a specific DNA polymerase for GC-rich templates (data not shown).

The aim of this work was to establish a simple methodology based on the one-pot megaprimer PCR to randomise a gene in a focused way, i.e. to randomise one or several cassettes in any given position of a gene. To illustrate the applicability of this method, the esterase libraries created were screened for altered substrate range of PFEI.

PCR conditions

(i) Single-template, one-pot megaprimer PCR (ST)

Mutation and amplification of the pfeI gene was performed by PCR using 0.5 pmol of the above-mentioned pGASPFEI plasmid in a volume of 100 µl. In the first stage, rrnB_RV was used as a reverse oligo together with one of the casXstop (where X = 1–7) ‘surrogate
mutagenic primers’ (Table I) in concentrations ranging between 5 and 50 pmol, 10 mM of each dNTP and 2.5 units of Pfuplus! DNA polymerase (Roboklon GmbH, Berlin, Germany) and the following temperature program: 3 min at 94°C, 25 cycles comprising each of 45 s at 94°C, 1 min at 53°C and 1 min/kb at 72°C and finally 10 min at 72°C followed by hold at 4°C. Following addition of 1 μl dNTP stock solution (10 mM each dNTP) and 2.5 additional units of Pfuplus!, the reaction tubes were subjected to five asymmetric cycles consisting each of 45 s at 94°C and 1 min 15 s at 72°C and cooled down to 4°C. After addition of 50 pmol of RhaP_long, the mixture was subjected to 25 cycles comprising each of 45 s at 94°C, 1 min at 66°C and 1 kb/min at 72°C and finally 10 min at 72°C followed by hold at 4°C. The overall strategy is illustrated in Fig. 2A.

(ii) Double-template (DT), one-pot megaprimer PCR. In order to reduce amplification of the wild-type sequence, mutation and amplification of the pfeI gene was performed by PCR using 0.5 pmol of the above-mentioned pET22pfEI plasmid in a volume of 100 μl. In the first stage, T7 + 1_RV was used as a reverse ologi together with one of the casXstop ‘surrogate mutagenic primers’ in concentrations ranging between 5 and 50 pmol, 10 mM of each dNTP and 2.5 units of Pfuplus! DNA polymerase and the following temperature program: 3 min at 94°C, 25 cycles comprising each of 45 s at 94°C, 1 min at 53°C and 1 min/kb at 72°C and finally 10 min at 72°C followed by hold at 4°C. Following addition of 0.5 pmol of pGPFEI plasmid, 1 μl dNTP stock solution (10 mM each dNTP) and 2.5 additional units of Pfuplus!, the reaction tubes were subjected to five asymmetric cycles consisting each of 45 s at 94°C and 1 min 15 s at 72°C and cooled down to 4°C. After addition of 50 pmol of RhaP_long, the mixture was subjected to 25 cycles comprising each of 45 s at 94°C, 1 min at 66°C and 1 kb/min at 72°C and finally 10 min at 72°C followed by hold at 4°C. The overall strategy is illustrated in Fig. 2B.

(iii) Competitive single- or double-templated, one-pot megaprimer PCR (compST or compDT). In order to test the possibility to incorporate simultaneously two mutagenic primers in the amplified product, amounts ranging between 0.5 and 50 pmol of each CAS1Xba and CAS3wt primers (Table I) were used as mutagenic

Fig. 1. Distribution of the seven cassettes (shadowed) in the protein sequence of the Pseudomonas fluorescens esterase I on the protein sequence and on the protein structure. The catalytic Ser residue is located at the apex of CAS2.

Table I. Primers used for PCR

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>RhaP_FW</td>
<td>5’-CATCATACTACGTTTCATCTTTCC-3’</td>
</tr>
<tr>
<td>rrnB_RV</td>
<td>5’-CGGCGAGGCAAAATCTGT-3’</td>
</tr>
<tr>
<td>T7term + 1_RV</td>
<td>5’-GCAATTGTGAGATCATACACGTTACATTTCTCCCTGTTGCCAATGGCC-3’</td>
</tr>
<tr>
<td>RhaP_long</td>
<td>5’-GTTGTTCAGCACGTTAGTTACACTGTTGAGATGCGACATGGGAAATCCAG-3’</td>
</tr>
<tr>
<td>cas1stop</td>
<td>5’-CCTGGTGAGTTCTGGGCCGTCGACCCTGTTGCGCAGAAGCCGAC-3’</td>
</tr>
<tr>
<td>cas1Xba</td>
<td>5’-CTTGGTGAGTTCTGGGCCGTCGACCCTGTTGCGCAGAAGCCGAC-3’</td>
</tr>
<tr>
<td>cas2stop</td>
<td>5’-GCTTGGGATCTGGGCCGTCGACCCTGTTGCGCAGAAGCCGAC-3’</td>
</tr>
<tr>
<td>cas3wt</td>
<td>5’-GCTTGGGATCTGGGCCGTCGACCCTGTTGCGCAGAAGCCGAC-3’</td>
</tr>
<tr>
<td>cas3stop</td>
<td>5’-GCTTGGGATCTGGGCCGTCGACCCTGTTGCGCAGAAGCCGAC-3’</td>
</tr>
<tr>
<td>cas4stop</td>
<td>5’-GCTTGGGATCTGGGCCGTCGACCCTGTTGCGCAGAAGCCGAC-3’</td>
</tr>
<tr>
<td>cas5stop</td>
<td>5’-GCTTGGGATCTGGGCCGTCGACCCTGTTGCGCAGAAGCCGAC-3’</td>
</tr>
<tr>
<td>cas6stop</td>
<td>5’-GCTTGGGATCTGGGCCGTCGACCCTGTTGCGCAGAAGCCGAC-3’</td>
</tr>
<tr>
<td>cas7stop</td>
<td>5’-GCTTGGGATCTGGGCCGTCGACCCTGTTGCGCAGAAGCCGAC-3’</td>
</tr>
</tbody>
</table>

Mutations introduced with respect to the wild-type sequence are shown in boldface and stop codons and restriction sites introduced are shown underlined.
Amplified products using the compST or compDT reactions were excised from a 0.8% (w/v) agarose gel, purified and finally digested with NarI and XbaI for 6 h. The whole digestion was run in a 1.5% (w/v) agarose gel, and the restriction pattern was analysed visually.

**Library screening for altered substrate selectivity**

Two hundred microlitres of LB medium containing the necessary antibiotics were dispensed into each well of a 96-well microtiter plate (MTP). OSCARR libraries were generated as described above and transformed into E. coli DH5α competent cells. Each transformant was picked with a sterile toothpick and used to inoculate each well. The plates were grown for 24 h at 37°C, and afterwards 100 μl sterile 60% glycerol (v/v) was added, the plates mixed shortly and stored at –80°C as master plates. With a 96-spoke replicator, new MTP containing 200 μl LB with antibiotics were inoculated and grown for 24 h. From these preinoculum plates, new plates containing 100 μl LB with antibiotics were inoculated with 100 μl, and incubated at 37°C for 3 h. Rhamnose was added to each well up to a final concentration of 0.2% (w/v), and the MTP were incubated for ~6 h until harvested by centrifugation at 213 × g and 4°C for 30 min. After supernatant elimination, the pellets were resuspended with 150 μl of 50 mM sodium phosphate buffer, 300 mM NaCl, pH 8, containing 0.1% DNase. The cells were lysed by incubating the plates for 30 min at 4°C, freezing for 1 h at –80°C, thawing for 30 min at 37°C. The lysates were clarified by centrifugation at 213 × g for 30 min.

Esterase activity was assayed in vitro by monitoring the amount of p-nitrophenol released upon hydrolysis of a 1 mM solution of p-nitrophenyl acetate (pNPA), butyrate (pNPB), octanoate (pNPC) or dodecanoate (pNPL) in 50 mM phosphate buffer, pH 7.5, at room temperature with a FLUOstar Optima spectrofluorimeter (BMG Labtechnologies, Offenburg, Germany). Aliquots (10 μl) of the cell fraction assayed were added to 190 μl of the reaction mixture and the increase in absorbance at 410 nm was measured for min using an apparent extinction coefficient of 1.33 × 10⁻⁴ M⁻¹ cm⁻¹. One unit of hydrolyase activity was defined as the amount of enzyme required to transform 1 μmol of p-nitrophenyl ester to p-nitrophenol per minute at room temperature.

**Library creation by error-prone PCR**

Mutant libraries containing on average three-point mutations per copy of the pef gene were created by epPCR using GoTaq DNA polymerase (Promega), 3 mM MnCl₂ (Cadwell and Joyce, 1992). The PCR product was digested with NdeI and BamHI, ligated into pGASTON vector precut with the same enzymes and transformed into E. coli DH5α.
**Molecular dynamics of esterase variant structures and molecular docking of substrates**

Models of the F125I and F125I/G119S mutants were built on the basis of the crystal structure of PFEI (pdb-code: 1VA4). Amino acid changes were introduced using the O graphic program (Jones et al., 1991) running on a Silicon Graphics workstation. Side-chain rotamers were chosen from a database of more common conformers (Ponder and Richards, 1987). Models for each conformer were energy minimised using the minimiser algorithm implemented in the CNS package (Brunger et al., 1998). The Engh and Huber (Engh and Huber, 1991) force field was used in all energy minimisation calculations. Finally, the conformation exhibiting less variation with respect to the native structure was chosen. The stereochemical quality of the model was checked with the program PROCHECK (Laskowski et al., 1993).

Molecular docking of pNPB and pNPC was carried out by homology modelling on the spatial distribution of the active site in the *Candida rugosa* lipase isozymes. Based on a previous binding model of *Bacillus subtilis* BS2 esterase (pdb-code: 1QE3) with pNPB (Kourist et al., 2007), catalytic residues of BS2 esterase and PFEI were superimposed and a preliminary ligand-binding model was generated for PFEI. Due to the serious differences in ligand-binding region between the two proteins compared, this previous model was discarded and molecular docking of C4 and C8 in BS2 was carried out on native PFEI and F125I mutant using GOLD program (Verdonk et al., 2003). Discussions about GOLD results were made from fitness parameter values and number of docking solutions provided by this program.

**Results**

As stated in the Introduction section, there is a considerable gap in mutagenesis methodologies between randomising techniques such as epPCR, whose mutation rate is not enough to mutate a 30 bp cassette, and site-directed mutagenesis, in which 1–6 bp can be mutated. The interest in performing such mutagenesis resides along the lines of ‘focused directed evolution’, in which determined regions of a gene (generally closer to the active site) are targeted for mutagenesis, in order to maximise the results and minimise library size. With the aim of bridging this gap, the gene coding for the PFEI was chosen as a model, because of the biocatalytic potential of this protein and its interest in protein evolution, since it belongs to the α/β hydrolase fold family, which is a very widespread structural scaffold among hydrolases. Assuming the ‘focused directed evolution’ hypothesis that ‘closer is often better’ (Morley and Kazlauskas, 2005), target regions for mutagenesis where chosen based exclusively on their distance (<10 or <15 Å) to the active site (Fig. 1). A first approach was made on the region termed CAS3, not only because of its proximity to the active site, but also because it had previously proved refractory to epPCR, likely due to a local high GC content (76.7%). In order to test the incorporation of mutagenic spiked oligos, the CAS3stop primer was designed to introduce a stop codon in the final product. Thus, a successful incorporation yielded a truncated product after cloning and expression, whereas amplification of the wild-type would lead to intact copies of the *pfeI* gene, resulting in active PFEI esterase.

Full-length product was obtained when using the single template (ST) strategy and different amounts of CAS3stop primer in the reaction (Fig. 3A). However, although specificity was gained in the amplification, efficiency of the mutation incorporation decreased with the concentration of CAS3stop present in the reaction (Fig. 3B). Additionally, the template concentration was optimised to 1 ng of plasmid. Higher concentrations of plasmid (up to 100 ng) and length of template (whether whole plasmid or just the fragment of interest) had no effect either on amplification specificity or on incorporation of the mutation (data not shown).

In order to increase the percentage of incorporation by decreasing the competing amplification of the wild-type sequence, two templates (pET22PFEI and pGPFEI) were used in order to avoid simultaneous annealing of both outer primers on the same template, and thus amplification of the wild-type (Eggert et al., 2005). As shown in Fig. 4A, the specificity of amplification increased at higher concentrations of CAS3stop primer, although the yield decreased. When the percentage of incorporation was measured, the double-templated reaction proved to be more effective than the

![Fig. 3. One-pot megaprimer PCR of the *pfeI* gene using a single template. Experimental conditions are detailed in Materials and methods. Arrows indicate the size of the expected product. (A) Lane 1: reaction with 50 pmol CAS3stop and 50 pmol rmb_RV; lane 2: reaction with 25 pmol CAS3stop and 50 pmol rmb_RV; lane 3: reaction with 5 pmol CAS3stop and 50 pmol rmb_RV; lane 4: 1 kb DNA ladder; lane 5: reaction with 50 pmol CAS3stop and 50 pmol rmb_RV; lane 6: reaction with 25 pmol CAS3stop and 25 pmol rmb_RV; lane 7: reaction with 5 pmol CAS3stop and 5 pmol rmb_RV. (B) Percentage of mutation incorporation for the reactions mentioned above. A minimum of 200 clones were counted for each bar.](https://academic.oup.com/peds/article-abstract/21/9/567/1461482/571)
In order to improve the specificity of the amplification and reduce the amount of leftover megaprimer, the amount of reverse primer was simultaneously reduced together with the amount of CAS3stop primer. As shown in Figs 2B and 3B (rightmost bars), although the percentage of mutation incorporation did not change significantly and the amplification became more specific, the amount of product decreased.

The effect of PCR additives on the ST amplification was studied as well. The effect of some additives, such as glycerol, was detrimental on the incorporation of the mutagenic primer, but 2% v/v formamide showed a marginal increase in specificity while allowing the primer to incorporate to the same extent as in a control without additives (Supplementary data, available at PEDS online, Table S2). Finally, five asymmetric cycles (consisting only of denaturation and extension) were added at the beginning of the best reaction, to yield 97% incorporation.

Under these optimised conditions, the incorporation percentage was tested for all of the remaining six cassettes, to check whether there was any dependence on the position of the cassette, i.e. on the length of the megaprimer. Incorporation proceeded at over 95%, regardless of the cassette position (Supplementary data are available at PEDS online, Fig. S1).

Finally, a spiked oligonucleotide was designed to fully randomise the cassette 3 region, and was used in the one-pot reaction, which was carried out under the established optimal conditions, i.e. 25 pmol spiked oligo, 50 pmol rnb_RV primer, 2% formamide. The product was purified, digested and ligated into precut pGaston, transformed into E.coli DH5α and 26 transformants were sequenced. The distribution of mutations introduced in PFEI at the amino acid level is shown in Fig. 5 and the sequences themselves are shown in the Supplementary data available at PEDS online, Fig. S2. In all 26 nucleotide sequences, one wild-type product was found, as well as one insertion, two single-nucleotide deletions and one deletion of practically the whole cassette 3. The population distribution is centred on two amino acid changes and due to the degeneracy of the genetic code 15% of the sequenced products, although mutated, translate into wild-type PFEI protein. The percentage of ‘useful’ products (mutated, but excluding frameshifts or deletions) totalled 64%.

As a practical example, the library generated on the CAS3 region (~1200 clones) was screened for altered substrate range using PNPA, butanoate, PNPC and PNPL. No hits were found in the epPCR library (~1500 clones) but two variants with striking differences in their substrate preference were found in the OSCARR library. Sequencing revealed one mutant to contain a single-point mutation F125I, and the other to be a G119S/F125I cumulative mutant. Detailed kinetic characterisation of the mutants revealed a 2-fold decrease in catalytic efficiency for short-chain esters (C2 and C4) and a 10-fold increase for C8 and C12 substrates (Supplementary data available at PEDS online, Table S3). Thus, the typical preference of a carboxyl esterase for short-chain esters was substantially shifted towards medium-chain fatty acids yielding variants with more lipase-like properties. Furthermore, this broadening of the substrate range did not totally destroy the acceptance of short-chain esters and thus makes these PFEI variants more versatile biocatalysts. epPCR failed to alter this property of PFEI and hence the focused directed evolution using the OSCARR method clearly represents a highly efficient approach.

Last but not least, the potential of this method to provide recombination simultaneously with randomisation was probed. For that, two markers were used one in each target region of the gene, whose incorporation can be evidenced independently. Surrogate mutagenic primers were designed, so that CAS1Xba would introduce an XbaI site in the cassette 1 region (Fig. 1) and CAS3wt, would re-introduce in the cassette 3 region, a NarI site that had been previously deleted in the templates pGPFEIΔNar and pET22PFEIΔNar.

As a preliminary study, the effect of the proportion of CAS1Xba:CAS3wt on the size of megaprimer obtained was studied. For that purpose, the concentration of CAS1Xba was
fixed at 50 pmol and decreasing concentrations of CAS3wt were used (50–6.2 pmol). As shown in Fig. 6A, already at a 1:0.5 ratio of primers, a 881 bp megaprimer was observed and its concentration increased as the amount of CAS3wt used decreased. Incorporation of CAS3wt in the megaprimer was studied by \textit{NarI} digestion of the purified 818 bp megaprimer obtained in (A) when the reaction was carried out with 25 pmol CAS3wt (lane 3), 12.5 pmol CAS3wt (lane 5), 8.3 pmol CAS3wt (lane 7). Lanes 2, 4 and 6 correspond to undigested controls, containing the same amount of DNA as lanes 3, 5 and 7, respectively. Lane 1: 1 kb DNA ladder. (C), (D) and (E) Restriction pattern of the final product obtained when the reaction was carried out with 25 pmol CAS3wt (C), 12.5 pmol CAS3wt (D), 8.3 pmol CAS3wt (E). Lane 1: undigested product; lane 2: partial digestion with \textit{XbaI}; lane 3: partial digestion with \textit{NarI}; lane 4: partial digestion with both \textit{XbaI} and \textit{NarI}; lane 5: Fermentas 1 kb GeneRuler [20, 10, 7, 5 (bright), 4, 3, 2, 1.5 (bright), 1, 0.7, 0.5 (bright), 0.4, 0.3, 0.2 and 0.075 kDa].

<table>
<thead>
<tr>
<th>Mutagenic primers incorporated</th>
<th>Fragments obtained after digestion (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (wt)</td>
<td>1086</td>
</tr>
<tr>
<td>CAS1Xba</td>
<td>859, 227</td>
</tr>
<tr>
<td>CAS3wt</td>
<td>590, 496</td>
</tr>
<tr>
<td>CAS1Xba + CAS3wt</td>
<td>227, 269, 590</td>
</tr>
</tbody>
</table>

The digestion is shown in Fig. 6C–E. Thus, decreasing the amount of CAS3wt surrogate mutagenic primer resulted in the increasing digestion of the 496 bp fragment when the full-length product was simultaneously incubated with \textit{NarI} and \textit{XbaI}, as evidenced by disappearance of the corresponding band (lane 4 in all cases). Therefore, in Fig. 6C, it may be concluded that only products incorporating the mutation in cassette 1 or cassette 3 were present, whereas in Fig. 6D,
the intensity of the band decreases and therefore, the presence of the three possible products can be expected in the PCR product mix, and in Fig. 6E, only products with mutations in the cassette 1, or in both cassettes 1 and 3.

Discussion

Recent trends in directed evolution are showing a preference for small, focused libraries, rather than oversize randomized libraries, that can only be partly screened (Qian et al., 2007). In the light of such a trend, and motivated by the impossibility of mutating certain parts of the gene coding for the PFEI by conventional epPCR, the OSCARR methodology was developed to efficiently randomise cassettes of any given gene.

Focusing the mutagenesis towards a discrete area may be more effective in order to develop enantioselectivity (Park et al., 2005), an alternate catalytic activity or a different substrate selectivity (Morley and Kazlauskas, 2005).

Therefore, there is a need for the improvement of the existing methods or development of viable alternatives in order to implement an efficient method, all with the aim of generating domain libraries instead of position-specific, saturation libraries, while maintaining a high yield.

For that purpose, a concept was developed, in which mutations would be delivered by a spiked oligonucleotide, that would be incorporated in the final product efficiently and quantitatively (i.e. with minimum side-products and high yield) by means of a modified one-pot megaprimer PCR (Tyagi et al., 2004; Ke and Madison, 2005; Nabavi and Nazar, 2005). However, in the examples found in the literature, the aim was to introduce single mutations in a gene, with a yield of approximately 80%. Unfortunately, having a content of only 80% variants in a library implies that one-fifth of the screened clones will still be wild-type and this is undesired, especially when dealing with expensive or technically complex screening methods (Reetz; Reetz et al., 2000).

Consequently, the reported megaprimer PCR techniques had to be customised to the quality requirements of a library for directed evolution.

First of all, an algorithm was used to design spiked oligos customised to our specific needs (Jensen et al., 1998). In the case of the CAS3 region of the gene coding for PFEI, a complete randomisation was chosen, but in other cases, special conserved motifs can be kept unchanged while randomising other positions. The use of spiked oligos in which defined percentages of each base are used becomes of special interest when overriding problems caused by mutation-refractory GC-rich sequences (such as in our case) or to overcome the mutational bias that causes adjacent amino acid replacements to be underrepresented in epPCR libraries.

Afterwards, a model was devised that allowed accurate estimation of the incorporation of the mutagenic primer into the final product, since the determination of mutation rates for the methods described in the literature was limited by a very low number of sequenced clones (between 1 and 5). To that purpose, the pfeI gene was chosen because its expression was easy, and the presence of active product in the transformants could be tested on plate through an activity-staining assay. The introduction of a stop codon as a model mutation in the pfeI gene (Table I) would yield a truncated and therefore inactive product (leading to a colourless colony on the assay agar-plate). Similar tests have been carried out introducing mutations into genes conferring different resistances (Selifonova et al., 2001; Denamur et al., 2002).

Starting from the mutation rate (80%) previously reported for the modification of the one-pot megaprimer PCR (Laskowski et al., 1993), conditions were improved reducing template to 1 ng plasmid DNA, and reducing primer concentrations (Fig. 3) until mutation rate values were situated above 90%. Then, additives were chosen that could exert positive effects especially on PCR specificity (Dieffenbach and Dveksler, 1995). Some authors describe the use of formamide concentrations of 5 up to 30% (Sarkar and Sommer, 1990; Sarkar et al., 1990; Zhang et al., 1991) using Taq polymerase, but with Pfu+ polymerase only 2% formamide could be used. Other strategies that are not based on the addition of reagents were implemented in order to improve product purity, such as addition of asymmetric cycles in the PCR (Ke and Madison, 2005). Seemingly, by asymmetrically generating only single-stranded mutated megaprimer first, the reverse primer has a higher likelihood of annealing to a mutated strand rather than amplifying the wild-type sequence. Secondly, the use of two templates to avoid the simultaneous annealing of both external primers (Zhang et al., 1991; Qian et al., 2007) in the proposed PCR protocol, remarkably improved the percentage of incorporation of the surrogate mutagenic primer, especially at low primer concentrations (5 pmol) (Fig. 4). Under the optimal conditions, an almost quantitative incorporation (97%) of the mutation-carrying oligonucleotide to the final product was achieved, and this was determined with a sufficient number of transformants (>200).

When introducing a new technique, a key issue is the comparison with other pre-existing methods and if it effectively fills gaps not covered by other techniques. The OSCARR concept is not affected by several limitations currently present in other techniques because it profits from custom-made spiked oligo design. For instance, in the Stratagene multi-site mutagenesis kit, for a 3-site mutation, a standard protocol yields 50% triple mutants, and the other 50% is made up of wt, single and double mutants. The multiple mutation events are not finely controlled, but rather obtained and optimised on a trial and error basis, either by increasing the percentage of DMSO or by increasing the amount of template, which would reduce the probability of simultaneous hybridisation, maximising the product with the highest number of mutations. The similar if not identical approach of Sawano and Miyawaki (2000) reached 70% multiple (double) mutants and the protocol described by Hames et al. (2005) introduced nine mutations with a frequency of four in five sequenced clones. Overlap extension (OE) has also often been used to create mutant libraries (Santoro and Schultz, 2002). Although OE–PCR was considered at first when devising OSCARR, the one-pot variation on the megaprimer PCR proved to be more advantageous in terms of even better yield and did not require purification of the intermediate products, which is even more convenient in the DT approach we proposed. In our case, the number of mutations within the cassette is controlled by oligo design in the non-recombinative OSCARR approach, and together with the high percentage of incorporation, results in the percentage of wt clones being significantly lower than in the above-mentioned techniques. In terms of crossover of mutations in
different cassettes, a recombinative approach was developed. This approach is based on the generation of different mega-primers that would allow the possibility of hybridisation and recombination among them during the same PCR (Urban et al., 1997; Eggert et al., 2005).

However, our method could not override the bias caused by the low probability of two adjacent nucleotide substitutions occurring when mutation rates are kept low. A careful design of target consensus sequence using the design algorithm might try to mitigate, but never fully overcome this bias.

On a practical level, the technique was tested using a spiked oligo in the CAS3 region to alter the substrate selectivity of PFEI. While in an epPCR-generated library, no hits were found, in the OSCARR library, two hits were found with a 10-fold higher catalytic efficiency towards pNPL. In order to explain the change in selectivity, the mutants were modelled and substrate docking was performed with pNPB and pNPC in both the wild-type enzyme and the F125I mutant. Although the lack of mutants in the epPCR library can be attributed to lack of coverage, attempts to randomise only a 76 bp stretch centred on CAS3 failed to deliver mutations (data not shown), evidencing that perhaps the high GC content of the CAS3 sequence, together with the deficiency of T and A in the epPCR protocol might be the cause for this bias.

The substitution F125I in the first variant seems to create a larger entrance to the active site (Fig. 7), and likely altering the aromatic substrate selectivity of this arylesterase. A similar phenylalanine-rich mouth of a substrate tunnel has been observed in the structure of the C. rugosa lipase, with a varying number of Phe residues in the different isoenzymes effectively modulating the substrate specificity of this enzyme (Mancheno et al., 2003). Furthermore, higher score parameter values and, in the docking of pNPC, number of docking solutions for F125I mutant than natives ones, confirmed the easier capability of the F125I mutant to fix pNPB or pNPC (Supplementary data are available at PEDS online, Table S4).

As for the second variant found (G119S/F125I), the G119S mutation would create a catalytic tetrad together with the other three active site residues S94, D222 and H251, in a similar way to other hydrolases such as B. subtilis esterase BS2, elastase, kallikrein, proteinase K and subtilisin Carlsberg. In our case, the introduction of G119S slightly decreased $K_m$. This may be in accordance with reports of reversion of the tetrad to triad having negative effects on $K_m$, depending on the extent on which the electrostatic potential of the catalytic Asp is affected (Krem et al., 2002). However, active site catalytic function seemed to be compromised by the introduction of the tetrad, as evidenced from the reduction of $k_{cat}$ values. These results are well in agreement with the areas closer to the active site targeted by Park et al. (2005) in order to alter the enantioselectivity of PFEI by acting on residues residues Trp28, Val121, Phe198 and Val225 by means of site-directed mutagenesis. Trp28 would

Fig. 7. (A) $p$-Nitrophenyl octanate docked into the active site of the wild-type enzyme and (B) $p$-nitrophenyl octanate docked into the active site of the F125I mutant. Note the enlarged cavity in the F125I mutant. (C) $p$-Nitrophenyl octanate is shown docked into the PFEI active site with the residues that define the narrow entrance (F125, F140, F143, F158, I224) shown in cyan.
be within the CAS1 region and Val 121 comprised in the CAS3 region. In our and Park’s case, mutations closer to the active site proved to create a greater effect than other mutations which were more distant to the active site, with rather moderate effects.

The combination of megaprimers sharing a common primers involved (Fig. 6A and B). Further experiments will be needed in order to confirm our first impression that obtaining recombined libraries (Fig. 6C–E) in a very straightforward, tunable and simple way is possible and to generate recombined mutant libraries of a protein of interest.

In summary, a new technique of broad applicability is presented here that yields high-quality libraries for focused directed evolution with the desired mutation rate, over the desired region of a gene, with the possibility of carrying out such mutations simultaneously in more than one region and therefore with the possibility of obtaining recombined libraries. Furthermore, the quality of the obtained libraries is ensured by the almost quantitative incorporation of the mutation-carrying oligonucleotide into the final product.

Acknowledgements

The authors would like to thank Dr Lars J. Jenssen (EMBL, Heidelberg, Germany) for his help with the spiked oligo design algorithm and Prof. José Berenguer (Centro de Biología Molecular ‘Severo Ochoa’, Madrid, Spain) for helpful discussion of PCR strategies.

Funding

European Commission’s Human Potential Programme (HPRN-CT2002-00239); Deutsche Bundestiftung Umwelt (AZ13159); Spanish Ministry of Education and Science (‘Ramón y Cajal’ Programme); ‘Ramón Areces’ Foundation.

References


Received December 28, 2007; revised April 28, 2008; accepted May 19, 2008

Edited by Andrew Griffiths