USER friendly DNA recombination (USERec): a simple and flexible near homology-independent method for gene library construction

B.R.M. Villiers, V. Stein and F. Hollfelder1

Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK

1To whom correspondence should be addressed.
E-mail: fh111@mole.bio.cam.ac.uk

USER friendly DNA recombination (USERec) is introduced as a near homology-independent method that allows the simultaneous recombination of an unprecedented number of 10 DNA fragments (~40–400 bp) within a day. The large number of fragments and their ease of preparation enables the creation of libraries of much larger genetic diversity (potentially \(10^{10}–10^{11}\) sequences) than current alternative methods based on DNA truncation (ITCHY, SCRATCHY and SHIPREC) or type IIb restriction enzymes (SISDC). At the same time, the frequency of frame-shifts in the recombined library is low (90% of the recombined sequences are in frame). Compared to overlap extension PCR, USERec also requires much reduced crossover sequence constraints (only a 5’-AN4-8T-3’ motif) and fewer experimental steps. Based on its simplicity and flexibility, and the accessibility of large and high quality recombinant DNA libraries, USERec is established as a convenient alternative for the combinatorial assembly of gene fragments (e.g. exon or domain shuffling) and for a number of applications in gene library construction, such as loop grafting and multi-site-directed or random mutagenesis.

Keywords: directed evolution/DNA library/exon shuffling/homology-independent DNA recombination/USER enzyme

Introduction

Directed evolution mimics natural evolution through selection in the test tube with the aim of creating genes that code for proteins with new functional traits (Jackel et al., 2008). This involves iterative cycles of gene diversification (Arnold and Georgiou, 2003a; Neylon, 2004) followed by screening or selection of the best variants (Arnold and Georgiou, 2003b). A major challenge remains that only a small fraction of the theoretically and practically accessible sequence space can be sampled during any screening or selection process. Furthermore, the marginal stability of many protein structures results in denaturation of some mutants leading to a high attrition rate in directed evolution experiments. A major research focus has therefore been to improve the functional diversity and the quality of gene libraries to maximize the chances of identifying stable mutants with a desired function. This particularly applies to screening and selection systems in which relatively low throughput assays are used, e.g. in 96-well plates.

A variety of different methods have been developed to generate libraries for directed evolution (Arnold and Georgiou, 2003a; Neylon, 2004; Wong et al., 2007). Error prone PCR (Cadwell and Joyce, 1994) and saturation mutagenesis (e.g. Miyazaki and Arnold, 1999) yield libraries that are relatively similar to the original gene and typically afford minor structural rearrangements along with relatively small improvements in function. In order to achieve more complex structural and functional changes, greater distances need to be covered in sequence space. For this purpose, a range of recombination strategies have been developed that combine the functional properties of two or more genes by swapping individual residues, secondary structure elements, motifs, domains and whole proteins (Ostermeier and Benkovic, 2000). Recombination-based gene diversification can be classified into homologous and non-homologous methods. DNA shuffling (Stemmer, 1994) and StEP PCR (Zhao et al., 1998) are two widely used homology-dependent recombination strategies that rely on hybridization and extension of homologous DNA fragments during PCR. A significant limitation of homology-dependent recombination strategies is that a set or family of genes needs to share at least 70% sequence identity in order to be efficiently recombined. This means that genes sharing high structural homologies but low sequence identities cannot be efficiently recombined. In addition, recombination events are generally biased towards regions of high homology, and the controlled recombination of defined structural motifs is not possible either.

In the evolution of the existing protein repertoire, homology-independent recombination events are thought to have played a central role (Lutz and Benkovic, 2002) and include: (i) gene fusion and domain recruitment as a mechanism to increase functional diversity related to substrate specificity, regulation and chemistry and (ii) exon shuffling which mediates the combinatorial assembly of proteins from independent structural and functional units.

To mimic these natural events, several homology-independent recombination methods have been developed (Lutz and Benkovic, 2000) to create chimeric proteins containing crossovers. Truncation-based methods that generate hybrid proteins of distantly related genes include ITCHY (Östermeier et al., 1999), SCRATCHY (Lutz et al., 2001a) and SHIPREC (Sieber et al., 2001). Disadvantages of these methods are that only two parent genes can be recombined at a time, and that only few crossovers are generated, typically one or two. Despite hybrid-purifying selection systems or size selection by agarose gel excision (Lutz and Östermeier, 2003), a significant portion of mutants is rendered non-functional as a result of base insertions and deletions since recombination occurs at random locations, irrespective of the reading frame and any underlying structural homology. Other methods have been developed for the simultaneous recombination of several fragments. Overlap extension PCR (Horton et al., 1989; Kolkman and Stemmer, 2001) requires ~15 nucleotide overlaps between fragments, and successive assembly PCRs. Random multi-recombinant PCR (RM-PCR) (Tsuji et al.,
does not require homologous sequences at the recombination sites but the accessible number and size of the recombined fragments is very limited due to the reliance on synthetic DNA sequences. In sequence-independent site-directed chimeragenesis (SISDC) (Hiraga and Arnold, 2003; Meyer et al., 2006a), type IIb restriction enzymes that cut outside their recognition site have been used, but their introduction is laborious and time consuming: the DNA fragments flanked by suitable restriction sites are first prepared by PCR, and then reassembled into their parent genes by overlap extension PCR. Moreover, only four fragments can be efficiently recombined at a time and intermediate cloning steps are required (Meyer et al., 2006a).

We report a simpler and more flexible near homology-independent DNA recombination strategy which allows the creation of large gene libraries (\(>10^6\)) through the combinatorial assembly of elementary gene fragments within a day while maintaining low error rates leading to unwanted mutations. Assembly of elementary DNA fragments is achieved with USER enzyme (Nour-Eldin et al., 2006; Bitinaite et al., 2007; Geu-Flores et al., 2007) and T4 DNA ligase (Fig. 1). USER enzyme is a commercially available enzyme mixture composed of uracil DNA glycosylase (UDG) and endonuclease VIII. In combination, the two enzymes catalyze the excision of uracil residues from DNA thereby generating a single base-pair gap. If a uracil residue is introduced close to the 5’-end by PCR using uracil-containing oligonucleotides, the ends of a linear DNA fragment can be endowed with unique single-stranded 3’ extensions (Fig. 1, step 2). T4 DNA ligase is used to ligate the different fragment variants simultaneously, leading to the recombination of the genetic diversity (step 3). The DpnI digestion allows removal of the PCR templates that were used to produce the fragments (step 4). Finally, the correct recombination product is amplified by PCR with external oligonucleotides and PfTurbo DNA polymerase, generating a large quantity of the recombined DNA library (step 5).

![Fig. 1. Principle of the USER friendly DNA recombination method (USERec) shown with two fragments (of 10 in total, see Fig. 2). Each fragment is produced by standard PCR using uracil-containing primers and PfTurbo Cx Hotstart DNA Polymerase (a mutant of Pf DNA polymerase able to amplify uracil-containing templates) (step 1). The genetic diversity is introduced for each fragment (F1, F2) by using a mixture of DNA templates (shown with different colors) leading to different fragment variants. For recombination between F1 and F2, the reverse uracil-containing primer of F1 and the forward uracil-containing primer of F2 are designed to be complementary for \(~8\) bases (Bitinaite et al., 2007) at the 5’-end, starting with A/U and finishing with U/A (5’-AN4-gU-3’). The PCR-amplified fragments are mixed and incubated with USER enzyme that generates a single nucleotide gap at the location of the uracil (U), creating complementary 3’ single-stranded extensions (step 2). T4 DNA ligase is used to ligate the different fragment variants simultaneously, leading to the recombination of the genetic diversity (step 3). The DpnI digestion allows removal of the PCR templates that were used to produce the fragments (step 4). Finally, the correct recombination product is amplified by PCR with external oligonucleotides and PfTurbo DNA polymerase, generating a large quantity of the recombined DNA library (step 5).](https://academic.oup.com/peds/article-abstract/23/1/1/1491461)
TycA were prepared, each of them containing a single alanine mutation in 1 of the 10 DNA fragment regions. Figure 2 shows an outline of this test case for the USERec procedure. In this way, $2^{10} (= 1024)$ possible combinations between wild-type and mutant gene fragments could be generated. This efficient procedure requires a minimal sequence pattern for recombination ($5'\text{-AN}_{4-8}\text{-T}-3'$), and only relies on simple, PCR-generated fragments for assembly.

Materials and methods

Oligonucleotides

All oligonucleotides (Supplementary data available at PEDS online, Table SI) were supplied salt-free by Operon Biotechnologies GmbH (Cologne, Germany).

Construction of the 10 alanine mutant templates

Ten alanine mutant templates were produced by site-directed mutagenesis based on the QuikChange method (Stratagene). Each of the 10 point mutations was introduced in 1 of the 10 DNA fragment regions (see Supplementary data available at PEDS online)—the following numbering corresponds to the base-pair numbers in the plasmid pSU18-tycA-PheATE-His (shown in the Supplementary data available at PEDS online) (Grünewald et al., 2004), F1: bp no. 1533–1723, length: 191 bp; F2: no. 1716–1828, 113 bp; F3: no. 1821–1934, 114 bp; F4: no. 1927–2291, 365 bp; F5: no. 2284–2389, 106 bp; F6: no. 2382–2466, 85 bp; F7: no. 2459–2676, 218 bp; F8: no. 2669–2772, 104 bp; F9: no. 2765–2806, 42 bp; F10: no. 2799–3114, 316 bp. Each 50 μl PCR contained 30 ng of the DNA template pSU18-tycA-PheATE-His (Grünewald et al., 2004), 0.2 mM dNTPs each, 0.24 μM of the forward and reverse mutagenic oligonucleotides (Supplementary data available at PEDS online, Table SI) and 2.5 U of PfuTurbo DNA polymerase in cloned Pfu DNA polymerase reaction buffer (Stratagene). The following cycling protocol was used: initial denaturation for 30 s at 95 °C; 18 cycles of denaturation for 30 s at 95°C, annealing for 1 min at 55°C and polymerization for 7 min at 68°C. The 50 μl PCR products were incubated with 1 μl (20 U) of DpnI restriction enzyme (New England Biolabs) for 1 h at 37°C. XL1-Blue cells (100 μl) were transformed by heat shock at 42°C with 8 μl of DpnI digestion product and plated on 2YT agar plates containing 20 μg/ml chloramphenicol. Plasmid DNA was prepared from a 5 ml overnight cell culture with the QIAprep Spin Miniprep Kit (Qiagen), and the concentration was measured by absorbance at 260 nm. The expected mutations were confirmed by sequencing with the oligonucleotides S_forward or S_reverse (Supplementary data available at PEDS online, Table SI).

PCR amplification of the 10 fragments (F1–F10) as substrates for recombination

The 10 recombination substrates (F1–F10) were prepared by PCR using an equimolar mixture of the WT TyCA (Mootz and Marahiel, 1997) and alanine mutant plasmid templates (i.e. 10 plasmids each containing one alanine point mutation), with the oligonucleotides indicated in Supplementary data available at PEDS online, Table SI. Each 50 μl PCR contained 5 ng of DNA template (i.e. 2.5 ng of plasmid pSU18-tycA-PheATE-His and 2.5 ng of the alanine mutant plasmid), 0.2 mM dNTPs each, 0.3 μM of the forward and reverse oligonucleotides and 2.5 U of Pfu Turbo Cx Hotstart DNA polymerase in Pfu Turbo Cx reaction buffer (Stratagene). The following cycling protocol was used: initial denaturation for 2 min at 95°C; 30 cycles of
denaturation for 30 s at 95°C, annealing for 30 s at 55°C and polymerization for 1 min at 72°C; final polymerization for 10 min at 72°C. An aliquot (10 μl) of each PCR product was analyzed by 3.5% TAE-agarose gel electrophoresis with 5 μl of HyperLadder V DNA marker (Bioline) (Fig. 3A). To make the procedure as practical as possible and minimize the number of purification steps following PCR amplification of the recombination substrates, we sampled all reactions into a single volume of 500 μl without quantifying the yield of individual PCRs. Three different aliquots (namely 50 μl, 100 μl or 250 μl) of this fragment mixture were then purified with the QIAquick PCR Purification Kit (Qiagen) and eluted with 40 μl of Milli-Q H₂O resulting in three purified fragment mixtures of varying concentrations; these eventually corresponded to an average 1.2 × 2.4 x and 6 x DNA concentration relative to the original PCR.

Recombination of the 10 fragments

The assembly efficiency was tested for the three purified fragment mixture concentrations (obtained from 50, 100 or 250 μl of fragment mixture), covering a range above and below the recommended amounts of uracil-containing DNA (cf. USER enzyme protocol; New England Biolabs). The purified fragment mixture (40 μl) was incubated in a PCR thermocycler with 5 μl of 10 × T4 DNA ligase buffer (New England Biolabs) and 5 μl (5 U) of USER enzyme (New England Biolabs) giving a final volume of 50 μl. After 20 min at 37°C and 20 min at 25°C, T4 DNA ligase (5 μl, 2000 U) was added and the resulting mixture (55 μl) was further incubated for 15 min at 25°C. The recombination product was purified with the QIAquick PCR Purification Kit (Qiagen) and eluted with 45 μl of Milli-Q H₂O.

To remove the plasmid templates used for the production of the 10 DNA fragments, the recombination product (45 μl) was incubated with 5 μl 10 × NEBuffer 4 and 1 μl (20 U) DpnI restriction enzyme (New England Biolabs) for 1 h at 37°C. The digestion product was purified with the QIAquick PCR Purification Kit (Qiagen) and eluted with 30 μl of Milli-Q H₂O. An aliquot (10 μl) of this purified DpnI digested recombination product was analyzed by 1% TAE-agarose gel electrophoresis with 5 μl of HyperLadder I DNA marker (Bioline) (Fig. 3B). The best assembly was obtained for the purified fragment mixture concentration corresponding to 100 μl (Fig. 3B, white arrow), so this purified DpnI digested recombination product was used for the later steps.

PCR amplification of the recombination product

An aliquot (10 μl) of the purified DpnI digested recombination product was amplified in a 50 μl PCR with 0.3 μM of the external oligonucleotides F1_forward and F10_reverse (C, 1% agarose gel). The major band obtained from this amplification (the recombined DNA library) had the correct size (1582 bp), was purified by gel extraction using the QIAquick Gel Extraction Kit (Qiagen) and eluted with 50 μl of Milli-Q H₂O. The DNA concentration was measured by absorbance at 260 nm.

Cloning of the recombed DNA library

The recombed DNA library was cloned by a method analogous to the QuickChange method, using the recombed DNA library as megaprimers. The 50 μl PCR contained 250 ng of recombed DNA library, 50 ng of DNA template [plasmid pSU18-rcA-PheATE-His (Gruenwald et al., 2004)], 0.2 mM dNTPs each and 2.5 U of PfuTurbo DNA polymerase in cloned Pfu DNA polymerase reaction buffer (Stratagene). The following cycling protocol was used: initial denaturation for 2 min at 95°C; 30 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 55°C and polymerization for 2 min at 72°C; final polymerization for 10 min at 72°C. All of the PCR product (50 μl) was loaded on a 1% TAE-agarose gel with 10 μl of HyperLadder I DNA marker (Bioline) (Fig. 3C). The major band, corresponding to the recombed DNA library (1582 bp), was purified by gel extraction using the QIAquick Gel Extraction Kit (Qiagen) and eluted with 50 μl of Milli-Q H₂O. The DNA concentration was measured by absorbance at 260 nm.
chloramphenicol. Plasmid DNA was prepared from a 5 ml overnight cell culture with the QIAprep Spin Miniprep Kit (Qiagen). The quality of the recombined DNA library was evaluated by sequencing with the oligonucleotides 
5′_forward and S_reverse (Supplementary data available at PEDS online, Table S1).

Results

Recombination of 10 DNA fragments by the USER enzyme

Ten DNA recombination substrates were generated by PCR (Fig. 1, step 1) using uracil-containing oligonucleotides and the proofreading PfuTurbo Cx Hotstart DNA polymerase. The identity of the fragments was ascertained by gel electrophoresis (Fig. 3A). PfuTurbo Cx is a mutant of Pfu DNA polymerase which is able to read through uracils in the DNA template. Each amplification reaction was supplemented with an equal amount of WT TyC and an alanine mutant as a marker for a possible recombination event. Assuming equal amplification efficiencies and unbiased recombination, each variant then has an anticipated recombination probability of 50% (see what follows). The amount of DNA template in the PCR had a significant impact on the efficiency of the subsequent assembly step: e.g. use of 50 ng DNA template instead of 5 ng decreased the assembly efficiency (data not shown).

To make the procedure as practical as possible and minimize the number of purification steps following PCR amplification of the recombination substrates, we sampled all reactions into a single volume of 500 µl without quantifying the yield of individual PCRs. Three different amounts of this DNA fragment mixture (namely 50 µl, 100 µl or 250 µl) were then purified and used for the assembly carried out with USER enzyme (Figure 1, step 2) and T4 DNA ligase (Figure 1, step 3). The strongest intensity of the band observed on the agarose gel was obtained for 100 µl (Fig. 3B, white arrow), while no correct recombination product could be detected for 250 µl, indicating that too much DNA inhibits the assembly reaction. According to the intensity of the band observed on the agarose gel (Fig. 3B, white arrow), about 30–40 ng of correct recombination product can be obtained in a 10 µl aliquot. This significant amount of DNA template limits the number of times the same sequence is amplified during the amplification reaction, thus making the polymerization-induced errors negligible (~1% of the sequences; see Supplementary data available at PEDS online). For a DNA fragment of 1582 bp (or about 10^3 kDa), 30–40 ng represent 10^10 double-stranded DNA copies, which matches or exceeds the throughput that is achieved with the majority of screening and selection systems.

Purifying the fragment mixture prior to the assembly process was necessary to remove the polymerase. It was observed that in the absence of a purification step, the number of single nucleotide deletions (dA or dT) at the ends of the fragment regions was significantly increased (data not shown). Before the amplification step, a DpnI restriction digestion was carried out (Fig. 1, step 4) to remove the parent genes that served as templates for the amplification of the 10 fragments.

Amplification of the recombination product

The recombination product obtained from 100 µl of fragment mixture (Fig. 3B, white arrow) was amplified with the external oligonucleotides F1_forward and F10_reverse (Supplementary data available at PEDS online, Table S1) (Fig. 1, step 5). The major PCR product obtained had the correct size (1582 bp, Fig. 3C) and was purified by gel extraction. The DNA concentration was measured by absorbance at 260 nm and showed that 1.25 µg of recombined DNA library could be produced from one single 50 µl amplification reaction. This result means that ~19 µg of recombined DNA library can be obtained from the initial 500 µl mixture of fragments. This value is based on using 10 µl of recombination product (of a total volume of 30 µl) for the amplification reaction, and 100 µl of fragment mixture (of a total volume of 500 µl) to produce the 30 µl of recombination product (i.e. 1.25 µg × 3 × 5 = 19 µg). This large quantity of DNA makes USERec suitable for constructing large libraries.

Cloning of the recombined DNA library

USERec allows a wide range of cloning strategies since specially designed external oligonucleotides can be used during the amplification step. For example, the most common cloning strategy based on type II restriction enzymes and a ligase could be applied by using external oligonucleotides containing a unique restriction site. Considering the high degree of DNA sequence identity between our recombined DNA library and the WT sequence, we decided to optimize a cloning strategy suitable for libraries with point mutations, since an important potential application of USERec is the construction of complex multiple site-directed mutagenesis libraries (Bhat, 1996). The strategy is inspired by the QuikChange method (Stratagene) in which the complementary mutagenic primers have been replaced by the double-stranded recombined DNA library (acting as megaprimers). About 6 × 10^4 clones could be obtained from 1.25 µg of the recombined DNA library and lab-made XL1-Blue competent cells with an efficiency of ~10^6 cfu/µg pUC19.

Analysis of the quality of the recombined DNA library

Table I shows the recombination patterns obtained in 28 randomly picked clones of the recombined DNA library. The average numbers of alanine mutations (A) per clone (4.61) or per fragment (12.90) are very close to their theoretical values (5 and 14, respectively), suggesting that there is no recombination bias. All the sequences (28) contained the 10 fragments in the right order, indicating perfect directionality of the method, and 21 of them (75%) showed no error. The seven remaining sequences (25%) contained one error (0.16 error per kb) and all errors were different: three deletions of one single nucleotide (dA) at the 5′-end of fragment regions (0.068 deletion per kb), and four substitutions (C to T, T to C, A to C, G to T) in oligonucleotide regions (0.09 substitution per kb), so ~90% of the sequences are in frame.

We suspect the deletions to be due to traces of PfuTurbo Cx Hotstart DNA polymerase, the 3′ overhangs being particularly sensitive to 3′→5′ exonuclease activity. All the substitutions occurred in oligonucleotide regions, which represent only ~24% of the sequence, suggesting that
Table I. Recombination patterns in 28 clones of the DNA library generated by USERec (see Fig. 2)

<table>
<thead>
<tr>
<th>Clone no.</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
<th>No. of A(^a) per fragment(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td></td>
<td>A</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>A</td>
<td>A</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>A</td>
<td>A</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>A</td>
<td>A</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

No. of A\(^a\) per clone\(^b\)

\(^{a}\)Alanine mutation.

\(^{b}\)The average number of A per clone in these 28 clones is 4.61. The theoretical no. is 5.

substitutions are not induced by polymerization but more likely by initial errors in oligonucleotide sequences. So it appears that the percentage of errors with USERec depends directly on the number of crossovers (or recombination sites, nine in our case). In other words, more crossovers would lead to more fragments, so more oligonucleotides and more potential deletion sites. The unprecedented number of 10 DNA fragments that can be simultaneously recombined should not be considered an upper limit. However, we anticipate that the number of errors (substitutions and deletions) in the recombined DNA library will increase with the number of fragments. Finally, no wild-type sequence was found in the 28 clones, indicating an efficient removal of the PCR templates by the DpnI digestion.

Discussion

The simple near homology-independent DNA recombination method based on USER enzyme and T4 DNA ligase described here provides a new route to DNA libraries for directed evolution. USER enzyme has previously been applied for cloning purposes (Nour-Eldin et al., 2006; Bitinaite et al., 2007; Geu-Flores et al., 2007) and required USER-compatible vectors. Bitinaite et al. (2007) assembled and cloned seven PCR fragments simultaneously. Our work expands a method originally optimized for gene synthesis to a DNA recombination tool. It is shown that the number of recombinant clones per assembly reaction—originally limited to a few hundred—is now sufficient to access large libraries due to the large amount of recombined DNA library produced. The previous limited cloning efficiency could be ascribed to the large number of DNA fragments that had to hybridize simultaneously together with the vector, and to the presence of nicks and gaps in the resulting plasmid. In addition, the error rate was too high to be useful for directed evolution experiments based on small libraries using low-throughput screening (~60%; 8 out of 10 recombiant clones contained the right PCR fragment, and 3 out of 6 clones with the right PCR fragment contained no error). Here, an unprecedented number of 10 DNA fragments were recombined with a lower error rate (25%).

In comparison to alternative recombination strategies, the USERec method carries a number of unique advantages: First, it is simple and fast as the PCRs leading to a recombinant DNA library can be produced within a single day. Given the large number of DNA fragments that can be recombined, the simplicity of their preparation by PCR and the yield of the recombination product, highly diverse recombination libraries become easily accessible. This is crucial for high-throughput screening experiments that search for very rare functional variants such as newly folded proteins created from the recombination of secondary structure elements (Riechmann and Winter, 2000; de Bon et al., 2005). Secondly, only minimal sequence requirements are imposed on the crossover sites allowing a range of different DNA fragments to be recombined. Considering the consensus sequence of 5'-AN\(_{4-8}\)T-3', crossover sites can easily be identified in the sequence of any gene as the distance between A and T is flexible (Bitinaite et al., 2007). The consensus sequence requirements of USERec are significantly less demanding than those needed in a conventional overlap extension PCR (Horton et al., 1989; Kolkman and Stemmer, 2001) where at least five identical amino acid residues are needed for an overlap of 15 nucleotides. Thirdly, USERec exhibits low error rates in terms of base substitutions, insertions and deletions, whereas frameshifts during the construction (Tsuji et al., 2001) or the recombination (Lutz et al., 2001a, b; Sieber et al., 2001; Hiraga and Arnold, 2003) of the fragments have been major issues for previous recombination techniques. For instance, the methods ITCHY (Ostermeier et al., 1999; Ostermeier and Lutz, 2003), SCRATCHY (Lutz et al., 2001) or the recombination (Lutz et al., 2001a, b; Ostermeier et al., 2003) do not control the reading frame of the recombination sites, generating a frameshift at the crossover in two-thirds of the library members. In-frame selection systems based on genetic fusions to antibiotic resistance genes have been developed to reduce this problem in SCRATCHY (Lutz and Ostermeier, 2003) and SHIPREC (Sieber et al., 2001; Udit et al., 2003) do not control the reading frame of the recombination sites, generating a frameshift at the crossover in two-thirds of the library members. In-frame selection systems based on genetic fusions to antibiotic resistance genes have been developed to reduce this problem in SCRATCHY (Lutz and Ostermeier, 2003) and SHIPREC (Sieber et al., 2001), but they can generate false positives (Lutz et al., 2001a) and represent an additional step in a method that is already labor intensive. In both techniques, an agarose gel excision step was also carried out prior to in-frame selection to select the hybrid genes of parental size. In these hybrid genes, the crossovers are more likely to occur at structurally related sites and thus less likely to generate unfolded chimeras, but the size selection criterion is approximate and problematic when parental genes have
different sizes. The error rate, especially frameshifts, and more generally the proportion of full-length and stably folded protein variants are the most important criteria for low-throughput screening experiments. In our case, the number of frameshifts as a result of insertions or deletions is low (~10%). This is particularly remarkable considering the large number of 10 DNA fragments that are recombined simultaneously, indicating that our method is suitable for constructing highly functional recombined DNA libraries. A single nucleotide substitution is a much less problematic mutation than a frameshift since protein functionality is less likely to be affected. In our method, these errors are quite limited as well (~15%), and are an expected consequence of the large number (20) of oligonucleotide sequences used. They do not seem to arise from the proofreading polymerase, since none of the errors was found in polymerization regions. The amplification of the recombination product is crucial to produce a large amount of recombined DNA library of the right size. In overlap extension PCR, a higher amplification factor (arising from low recombination efficiency) results in a larger number of polymerase-induced errors (Geu-Flores et al., 2007).

We anticipate USERec will be of high practical value for a number of applications in gene engineering where several genetically diverse DNA fragments need to be assembled. These include exon shuffling and recombination of DNA fragment libraries that have been randomized by error prone PCR and/or site-directed mutagenesis. The size of the DNA fragments that were recombined in our method (~40–400 bp) matches the domain size (30–130 amino acids) that is observed in natural modularly assembled proteins (Kolkman and Stemmer, 2001), suggesting that USERec is suitable for exon shuffling. Moreover, given the high flexibility in the choice of recombination sites, USERec might also be a valuable tool for identifying the most functional domain fusion sites e.g. in case of loop grafting for the humanization of antibodies (i.e. CDR grafting) (Kettleborough et al., 1991). Other potential targets of this method include folds like TIM-barrels that are structurally homologous, but have diverged during evolution at the DNA level, often precluding DNA shuffling. Likewise swapping of domains in modular biosynthetic systems such as polyketide synthases or non-ribosomal peptide synthetases can be considered. More generally, the flexibility of the USERec signature sequence allows the definition of recombination sites according to the logic of protein structure, maintaining e.g. highly conserved elements and placing recombination sites in locations that are less likely to disrupt protein structure.

In the laboratory, this is increasingly supported by computational methods developed to identify the most favorable recombination sites in protein sequences i.e. the ones that minimize folding pathway disruptions and steric and/or physicochemical clashes at the module surfaces. SCHEMA (Voigt et al., 2002; Meyer et al., 2006b) calculates from the three-dimensional structure, the number of interactions that are disrupted upon recombination, when specified elements of the protein are replaced by counterparts from another parent. FamClash (Saraf et al., 2004) minimizes residue–residue clashes in engineered protein hybrids by using protein family sequence data. Domain-finding algorithms (Swindells, 1995; Emmert-Streib and Mushegian, 2007) have also been used to identify mobile protein units that fold independently. The recombination sites identified in silico can be tested experimentally by protein fragmentation (Tasayco and Carey, 1992), and the peptide backbone cleavage at these sites should not cause the unfolding of the resulting fragments. The potential recombination sites should also be able to tolerate peptide insertions or deletions. Protein sequence alignments showed that these sites are most likely located in loops (Pascarella and Argos, 1992). Strategies are available to increase the tolerance of recombined libraries to recombination events: e.g. by low error-rate random mutagenesis or by the introduction of known stabilizing mutations [e.g. back-to-consensus/ancestor mutations (Bershtein et al., 2008)] non-functional or unstable chimeras have been shown to be rescued (Meyer et al., 2006b).

USERec is set up to allow facile access to such experiments that are designed to take advantage of the ability to combine protein ‘building blocks’ at will. The generation of near homology-independent recombination libraries composed of multiple DNA fragments in one step, within a day and with minimal sequence constraints renders this method potentially valuable for a number of applications in gene library construction.

Acknowledgements
We thank members of the Hollfelder group for comments on the manuscript.

Funding
This work was supported by the MRC, an EPSRC platform grant, the EU network ProSA and the EU NEST project MiFem. B.R.M.V. is supported by the EU Early-stage Training Site ChemBioCam. F.H. is an ERC Starting Investigator.

References

Received April 14, 2009; revised September 29, 2009; accepted September 30, 2009

Edited by Alan Berry