Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): a versatile tool for generating targeted libraries

Asael Herman1,2 and Dan S. Tawfik1,3

1Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel
2Present address: Department of Pathology, University of Washington, Seattle, WA 98195, USA
3To whom correspondence should be addressed.
E-mail: tawfik@weizmann.ac.il

The directed evolution of proteins has benefited greatly from site-specific methods of diversification such as saturation mutagenesis. These techniques target diversity to a number of chosen positions that are usually non-contiguous in the protein’s primary structure. However, the number of targeted positions can be large, thus leading to impractically large library size, wherein almost all library variants are inactive and the likelihood of selecting desirable properties is extremely small. We describe a versatile combinatorial method for the partial diversification of large sets of residues. Our library oligonucleotides comprise randomized codons that are flanked by wild-type sequences. Adding these oligonucleotides to an assembly PCR of wild-type gene fragments incorporates the randomized cassettes, at their target sites, into the assembled gene. Varying the oligonucleotides concentration resulted in library variants that carry a different average number of mutated positions that comprise a random subset of the entire set of diversified codons. This method, dubbed Incorporating Synthetic Oligons via Gene Reassembly (ISOR), was used to create libraries of a cytosine-C5 methyltransferase wherein 45 individual positions were randomized. One library, containing an average of 5.6 mutated residues per gene, was selected, and mutants with wild-type-like activities isolated. We also created libraries of serum paraoxonase PON1 harboring insertions and deletions (indels) in various areas surrounding the active site. Screening these libraries yielded a range of mutants with altered substrate specificities and indicated that certain regions of this enzyme have a surprisingly high tolerance to indels.

Keywords: directed evolution/rational design/PON1/methyltransferase/insertions/deletions/indels/DNA shuffling

Introduction
Rational design and directed evolution are the two conceptually contrary strategies that underpin protein engineering. Directed evolution requires no prior knowledge of the target protein, yet it relies on selection capabilities that sample only a miniscule fraction of all possible permutations. Rational and computational designs greatly minimize the number of sequence permutations that are explored (often down to one sequence), but are hampered by the complexity of proteins and our limited knowledge regarding sequence–function relationships. An awareness of the relative strengths and weaknesses of these approaches has led workers to combine them, for example, in ‘semi-rational’ protein engineering (Minshull et al., 2004; Chica et al., 2005; Patrick and Firth, 2005) and in targeting library diversity to a given set of residues (e.g. a defined set of active site positions). Recent examples of this ‘targeted library’ approach include the modification of the substrate specificity of a Cre DNA recombinase (Santoro and Schultz, 2002), an epoxide hydrolase (Rui et al., 2004; Park et al., 2005), a lipase (Reetz et al., 2005) and an esterase (Park et al., 2005). Other ‘targeted library’ approaches utilize computational methods and protein design algorithms. Examples include methods that perform a ‘virtual screening’ of otherwise impossibly large libraries (Hayes et al., 2002), computational methods for the design of enzyme active sites (Dwyer et al., 2004) and algorithms that direct recombination by predicting optimal crossover loci (Voigt et al., 2002).

‘Targeted libraries’ are constructed primarily by directing randomization (by saturation mutagenesis) to specific positions within the gene. Saturation mutagenesis uses synthetic oligonucleotides that encode the desired diversity at the specified positions (for example, see Reetz et al., 2001; Santoro and Schultz, 2002; Antikainen et al., 2003; Reetz, 2004; Rui et al., 2004). The diversified oligonucleotides are incorporated by PCR, or directly cloned into the gene of interest as a cassette. The main drawback of this approach is that it often produces library sizes that are too large to explore fully by available selection strategies. It is possible, therefore, that combining computational design with such repertoire selections might, by allowing the former larger degrees of freedom, narrow down the potential library size to a more manageable number.

Rational library design also requires compromises to be made. A typical active site comprises well over 20 non-contiguous residues, and a change of any one of these residues may provide the key to the desired new function. However, simultaneous diversification of many residues creates library sizes that are beyond any available screening capabilities. Typical plate-based screens involve \( \leq 10^4 \) variants and can therefore target only two fully randomized positions. Even high-throughput technologies that allow \( 10^{10} \) variants to be screened can only therefore accommodate six fully randomized positions. Screening only a sample of the library diversity is an option, but one must bear in mind that as the number of diversified positions increases, the number of library variants that are completely inactive (due to the presence of a stop codon, or a mutation that severely undermines stability) increases substantially (Bershtein et al., 2006).

A potential solution to the above obstacle is parsimonious mutagenesis (Balint and Larrick, 1993). As the name suggests, this technique provides a means of partial diversification using
oligonucleotides in which the diversified codons comprise a small proportion of mutating bases in an excess of wild-type bases. However, this technique has not been used extensively, possibly because the high cost of 'doped' oligonucleotides and their limited purity.

Our aim was to develop a cost-effective, facile and general method for the creation of targeted libraries by partial diversification. The approach we took, dubbed ISOR, is a simple adaptation of gene shuffling and allows diversification—by substitution, insertion or deletion—of large sets of residues. Each library variant carries a random, and different, subset of mutated residues, with the entire set represented in the complete library. Here we demonstrate ISOR’s applicability and versatility in two different systems. Following a bioinformatic analysis, we targeted for diversification 45 individual positions in a DNA methyltransferase (M.HaeIII). We created a series of gene-libraries in which the average number of mutations ranged from 1 to 6, and that combinatorially covered the entire set of 45 positions. We also targeted indels (insertions and deletions) to various structural elements in serum paraoxonase (PON1). Both libraries were screened, and functional variants with wild-type-like specificity, or with altered specificities, were isolated.

Materials and methods

Oligonucleotides

The libraries described here were constructed with oligonucleotides obtained from MWG Biotech (Ebersberg, Germany), at the lowest purification grade (HPSF purification). More recently, we have also applied oligonucleotides from IDT (Coralville, IA, USA) of standard desalting grade.

Incorporation of oligonucleotides into the M.HaeIII gene

A schematic outline of ISOR is illustrated in Fig 1. The M.HaeIII gene (992 bp) was initially PCR amplified from the pIVEX2.2-M.HaeIII plasmid using primers LMB2-4 (5’-biotin-labeled) and pIVB-7 (Griffiths and Tawfik, 2003). Approximately 6 μg of this PCR product, in 50 μl of digestion buffer (50 mM Tris–HCl buffer pH 7.5, 10 mM MnCl2), was equilibrated at 20°C in a thermocycler (Eppendorf, Germany). 0.05 U DNaseI was added and DNA digestion allowed to proceed for 5 min at 20°C, then terminated by adding 15 μl 0.5 M EDTA and heating at 90°C for 10 min. The gene fragments were separated in a 2% agarose gel and those of 70–100 bp in size excised and purified using the QIAEX II Gel Extraction Kit (Qiagen). The gene was reassembled by combining 100 ng of purified DNA fragments with varied amounts of oligonucleotides (Supplementary Table II available at PEDS online). The NNS codon (that gives rise to all 20 amino acid residues and minimizes the frequency of stop codons) was incorporated into the reassembled gene.

Construction of M.HaeIII gene-libraries

We synthesized 45 oligonucleotides in order to direct random substitutions to 45 different positions within the M.HaeIII gene (Supplementary Table II available at PEDS online). The NNS codon (that gives rise to all 20 amino acid residues and minimizes the frequency of stop codons) was used for all substitutions. Each of the 45 oligonucleotides was 33 bases long and comprised an NNS codon with 30 flanking bases complementary to the wild-type sequence (15 bases from each side of the NNS). All 45 oligonucleotides were mixed with M.HaeIII gene fragments at equimolar ratios and the whole assembled as described earlier. To help maintain the diversity created in the assembly reaction (>10^10 genes), the full-length assembly product was enriched, and the un-incorporated oligonucleotides removed as follows: 50 μl of assembly reaction was mixed with 2.5 μl M280 streptavidin-coated magnetic beads (Dynal) and 50 μl buffer (10 mM Tris–HCl buffer, pH 7.4 containing 1 M NaCl, 25 mM EDTA and 15 mM EGTA), and incubated at ambient temperature for 1 h. The beads were rinsed three times with the same buffer, and three times with 50 mM Tris–HCl (pH 8), and resuspended in a 50 μl nested PCR.
reaction mixture. The resulting PCR product was then cloned into pIVEX2.2.

In vitro methylation assays

Libraries, gene pools from selected libraries and individual cloned genes were all assayed for methylation activity by the same digoxigenin–biotin ELISA-based method (Tawfik and Griffiths, 1998). PCR-amplified DNA (2 nM) was transcribed and translated in vitro with Eco Pro T7 extract (Novagen) for 1 h at 30°C. The temperature was adjusted to 24°C and the reaction mixed with an equal volume of ‘methylation mixture’ (100 mM Tris–HCl buffer, pH 8.5 containing 100 mM NaCl, 20 mM diithiothreitol, 20 mM EDTA, 0.3 mM S-adenosyl-l-Methionine and 30 nM of a 1 kb DIG-folA-3-biotin DNA substrate). Aliquots were collected at various time points, then quenched and incubated in streptavidin-coated 96-well plates (Nunc). The bound DNA was digested with the restriction endonuclease HaeIII (New England Biolabs). Methylation progress was followed by ELISA, using anti-DIG-HRP-conjugated antibodies (Roche). The ELISA signal was plotted against time and the time required to methylate 50% of the restriction/methylation sites (t50) determined.

In vitro compartmentalization and selection for M.HaeIII activity

Selection for M.HaeIII activity by in vitro compartmentalization took a modified form of that described by Tawfik and Griffiths (1998). Briefly, 100 μl of EcoPro T7 in vitro transcription/translation system (Novagen), plus 0.1 nM biotinylated library DNA and 0.3 mM S-adenosyl-l-Methionine, were added to 1 ml of ice-cold oil mix [4.5% (w/w) Span80, 0.5% (w/w) Tween80 in light mineral oil (Sigma)]. The mixture was homogenized on ice for 5 min at 8000 rpm in an Ultra Turrax T25 (IKA) homogeniser equipped with a disposable shaft (OmmiTip). The resulting emulsions were incubated at 25°C for 4 h, and then broken and the biotinylated genes therein captured on streptavidin-coated beads. Non-methylated genes were ‘neutralized’ by digesting with the restriction endonuclease HaeIII (NEB). The methylated and, therefore, undigested genes were subsequently amplified by PCR and re-cloned into pIVEX2.2.

Construction of PON1 libraries

The PON1 variant gene G3C9 (Aharoni et al., 2004) was PCR amplified from a pET32b plasmid using primers S'-biotin pET-fw (GGCAGCCAATCGCTTCC) and pET-bac (CGAACGCCAGCACATGG). The 1065 bp PCR product was used as a template for the creation of ISOR gene libraries harboring indels in different structural elements (Supplementary Table IV available at PEDS online). Once assembled (as earlier), the full-length products were re-cloned into pET32b plasmid.

Screening PON1 libraries

A screen for a range of PON1 activities was applied, essentially as described elsewhere (Aharoni et al., 2004; Harel et al., 2004). Briefly, libraries were transformed into E. coli cells; then grown on nutrient agar plates and replicated with velvet cloth for the esterase screen. A layer of soft agar (0.5%) in activity buffer (50 mM Tris buffer, pH 8.0 containing 1 mM CaCl2) was supplemented with 0.3 mM 2-naphthylacetate (2NA) and 1.3 mg/ml Fast Red (Sigma Aldrich) then poured onto the original agar plates. Colonies that turned red first were picked from the replica plate and used to inoculate 500 μl of LB medium in a 96-deep-well plate. Following growth overnight on a shaker plate (200 rpm) at 30°C, plates were duplicated and lysed with BugBuster (Novagen). The hydrolysis of three substrates, 2NA, paraoxon and γ-thiobutyrolactone, was monitored in cleared crude cell lysates as described elsewhere (Aharoni et al., 2004; Harel et al., 2004). Briefly, aliquots (10–100 μl) of cleared lysate were transferred into transparent polystyrene 96-well plates, and mixed with substrate in activity buffer. Product release was monitored in a plate reader at 405 nm for p-nitrophenol, and 320 nm for 2-naphthol. Hydrolysis of γ-thiobutyrolactone was detected using 5,5'-dithio-bis-2-nitro benzoic acid at 412 nm as described (Aharoni et al., 2005a, 2005b).

Results

Incorporation of synthetic oligonucleotides in the process of gene reassembly

Figure 1 presents a schematic of ISOR where a biotinylated PCR product of the target gene is subjected to fragmentation by digestion with DNaseI. The DNaseI fragments are then mixed with a set of synthetic oligonucleotides, and assembled in a process of self-primed extension by Taq polymerase. The assembled genes can be enriched by capture on streptavidin-coated magnetic beads, thereby maintaining the diversity created in the assembly reaction by minimizing mispriming and the amplification of short products. It is worth noting, however, that in most cases magnetic bead separation need not be applied, particularly if the required library diversity is ≤10⁶ genes. In any case, the product (enriched or not) is then amplified in a nested PCR using internal, non-biotinylated primers. It should be noted that the assembly reaction and subsequent PCR amplification will introduce additional point mutations at random. The frequency of such can, however, be controlled by the choice of polymerase. Here, we primarily used a high-fidelity polymerase that gave an average of 0.5-point mutations per gene. The application of ordinary polymerases may result in a much higher frequency of mutations (>2 per 1 kb genes).

Our first goal was to optimize reaction conditions and tune the frequency of oligonucleotide incorporation. To this end, we tested the incorporation of two oligonucleotides, oligo I (40 bases) and oligo II (60 bases), which bore a unique EcoRI restriction site plus two (oligo I) and three (oligo II) randomized (NNS) codons (Supplementary Table II available at PESDS online). Varying amounts of these oligonucleotides were mixed with 100 ng (120 nM) of M.HaeIII gene fragments (generated by DNaseI digestion), and the gene reassembled by self-primed extension. The expected EcoRI restriction pattern was then compared with that obtained with the assembly products (Fig. 2). At an initial oligonucleotide concentration of 320 nM, DNA products with one restriction site appeared, indicating a single oligonucleotide had been incorporated. Seventy percent DNA products at this concentration, however, had no oligonucleotide incorporated at all. On the other hand, at the highest oligonucleotide concentration tested (800 nM), intact DNA that had no
oligonucleotides incorporated was scarce (4%), and the products containing oligo I, oligo II or both were found in equal proportions. It is also notable that the strand used as template for the oligonucleotides’ synthesis is generally of no importance, and the libraries described in the subsequent parts of this work were made with oligonucleotides that were all complementary to the same strand. In cases where neighboring residues should be targeted independently, the usage of oligonucleotides complementing the opposite strands is recommended. We were able to successfully modify two consecutive codons, using two oligonucleotides that were complementary to opposite strands (data not shown). Using an approach similar to that described earlier, we were able to incorporate oligonucleotides encoding insertions or deletions of 3 or 12 bases (Supplementary Fig. I available at Peds online). Thus, both substitutions and indels could be incorporated into genes, and the level of incorporation could be tuned using different oligonucleotide concentrations.

**Generation of targeted libraries of M.HaeIII**

The M.HaeIII gene was diversified by ISOR to demonstrate that our technology could be used to construct highly diverse, yet targeted gene libraries. Our goal was to direct diversity to a subset of residues that are only moderately conserved; working by the assumption that a library with mutations in highly conserved residues would contain a high ratio of inactive variants, whereas mutations in non-conserved residues would be largely neutral. A library based on moderately conserved residues therefore might be an optimal starting point for the evolution of new enzyme variants. We used ConSurf (Landau *et al.*, 2005) to search for M.HaeIII-homologous cytosine-5 DNA methyltransferase sequences. The results were aligned against M.HaeIII, a phylogenetic tree was constructed and the degree of conservation for each M.HaeIII residue calculated. This investigation led to the identification of a group of 45 residues which, judging by ConSurf scores, were moderately conserved (Supplementary Table I available at Peds online). A set of 45 oligonucleotides was synthesized to target NNS codons to each one of these 45 positions. A range of libraries was made by including different concentrations of an equimolar mixture of all 45 oligonucleotides in the assembly of the M.HaeIII gene. A range of oligonucleotide concentrations (9–144 nM) gave a concentration-dependent average incorporation rate of 1–6 randomizing oligonucleotides (Table I). Oligonucleotide concentrations under 9 nM (0.2 nM each oligo) resulted in no detectable incorporation, whereas concentrations above 360 nM inhibited the assembly reaction, reducing yield significantly. Sequence analysis of library variants did not reveal any bias either in the location of incorporation or in the nature of the amino acids incorporated.

The residual methylation activity of the proteins encoded by the library was determined. The PCR amplified libraries transcribed and translated in vitro, and the resulting proteins assayed in pool for methylation activity. Table I shows the relationship between oligonucleotide concentration and methylation activity. Thus, a library containing, on average, one mutation per gene exhibited 11% of wild-type activity, whereas the incorporation of ≥5 mutations per gene reduced the residual activity below the detection threshold.

**Selection of the targeted M.HaeIII gene-libraries**

Having established the ability of ISOR to controllably incorporate site-specific mutations, we wanted to uncover the potential of ISOR libraries, particularly those containing a high number of mutated residues, to yield active protein variants. We created an initial library by combining the two libraries with the average mutation frequencies of 5 and 6 mutations per gene. Prior to selection, the combined library exhibited no detectable methylation activity. A pool of \( \sim 10^{10} \) of these genes was taken through two rounds of selection for M.HaeIII activity (methylation of GGCC) by compartmentalization in emulsions (Tawfik and Griffiths, 1998). The gene pools from the first, and second, round of selection exhibited 3%, and 6%, of the M.HaeIII wild-type activity.

Ten clones from the second rounds were randomly picked and their sequence and methylation activity determined (Supplementary Table III available at Peds online). Of these ten clones, three were not active (two of these encoded truncated proteins due to the incorporation of stop codons in one of the diversified positions). The remaining seven clones had activities ranging from 3% to 42% of wild-type activity, and contained 2–5 mutations per gene. None of the 10 clones had a wild-type sequence, and no mutation repeated among the clones. As expected, most of the substitutions observed in the targeted positions (68%) were to amino acids that appear in homologous C5 methyltransferases, as judged by the library.

**Table I. Gene-libraries derived from the M.HaeIII gene**

<table>
<thead>
<tr>
<th>Oligonucleotides in assembly (nM)</th>
<th>Average no. of mutations per gene</th>
<th>Methytransferase activity (wild-type = 1) ( ^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>1</td>
<td>0.11</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>0.058</td>
</tr>
<tr>
<td>36</td>
<td>3.6</td>
<td>0.033</td>
</tr>
<tr>
<td>72</td>
<td>5</td>
<td>0.004</td>
</tr>
<tr>
<td>144</td>
<td>6</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

\( ^* \) Methytransferase activity is expressed as the time required to methylate 50% of the R/M sites by wild type M.HaeIII divided by the time required by the sample proteins.
by our multiple sequence alignment (data not shown). These results demonstrate the capability of ISOR to create large, highly complex and diverse, yet functional libraries.

Generation and selection of PON1 indel libraries

We also employed ISOR to generate serum paraoxonase (PON1) gene libraries containing indels in various structural elements related to PON1’s active site and which are not an integral part of the β-propeller scaffold. Thirty different active site positions were chosen along PON1’s primary structure (Supplementary Table IV available at PEDS online) targeting two major locations of PON1’s active site (Supplementary Fig. II available at PEDS online).

The first targeted group is the active site canopy, apparently unique to PONs, which is defined by helices H2 and H3, and the loops connecting them to the β-propeller scaffold (Harel et al., 2004). We included in this group a long surface loop (residues 68–83) that is mostly disorder in the crystal structure, but its location suggests that it may comprise part of the active site. The canopy, including the 68–83 loop, is thought to be of principle importance to the function of PONs as it contains most of the residues that seem to determine the substrate specificity of the various PON family members (Harel et al., 2004).

The second group locates to relatively short loops (2–9 residues) at the top of the tertiary structure that are typical to β-propellers (the upper loops). These loops connect either the outer β-strand of each blade (strand D) with the inner strand of the next blade (A), or the two strands in the middle of the blades (strands B and C). Many of the residues within these loops face the entrance to the central tunnel of the propeller, and some may influence enzymatic activity (Harel et al., 2004).

Ultimately, 16 canopy positions and 14 upper loops positions were chosen for the introduction of indels. Three different oligonucleotides were synthesized for each of the 30-targeted positions (Supplementary Table IV available at PEDS online) as follows: one oligonucleotide introduced an insertion of a single randomized codon (NNS), one inserted two NNS codons and the third was designed to delete the targeted position.

Several gene-libraries of PON1 were prepared by incorporating different concentrations of the 90 oligonucleotides (at equal ratios). The relationship between oligonucleotide concentration and mutation frequency was nearly linear within the range of 9–72 nM (or, 0.1–0.8 nM of each oligonucleotide). In addition to the ‘designed’ indels, the libraries carried an average of 0.5 additional random point mutations due to PCR errors. The residual activity of these libraries also correlated with the average number of mutations (Fig. 3).

Mapping PON1’s tolerance to indels

The versatility of ISOR allowed us to create various gene-libraries with indels in different structural elements of the protein, and then to compare the residual activity of libraries carrying a similar (~0.4%) mutation frequency (Table II).

The residual activity of the canopy library (19%) was similar to the residual activity arising from diversification of its various components [H2 (16%), H3 and connecting loop (13%)]. Interestingly, this level of residual activity (13–19%) is very similar to residual activities observed in libraries with the same mutation frequency, but created by error-prone PCR along the entire gene (L.Gaidukov, unpublished data). Somewhat surprisingly, therefore, the canopy region appears to be as tolerant to indels as PON1 is to point mutations across its entire length. Introducing indels to other regions in the PON1 fold produced rather different residual activities. The upper loops were found to be less tolerant to indels (7% residual activity), possibly a function of their short length. The long, mobile surface loop, on the other hand, was exceptionally tolerant to indels (37.5% residual activity).

Screening the PON1 indel libraries

To demonstrate the capability of our ISOR indel libraries to generate functional variants, we screened a library with indels at all the structural elements described and further characterized variants markedly differ from wild-type PON1 in their substrate specificity. A library was prepared from all 90 oligonucleotides at a total concentration of 36 nM, this produced an average of 2.75 directed indels and 0.62-point mutations per gene. We applied a screen on agar plates for esterase activity using 2NA. This screen is highly sensitive, and even clones with very weak esterase activity (k_{cat}/K_{M} > 25 M^{-1}s^{-1} (Gould and Tawfik, 2005) appear positive. This method was used to screen ~2000 clones, and we

### Table II. Residual activity of PON1 libraries with indels incorporated into various structural elementsa

<table>
<thead>
<tr>
<th>Structural element</th>
<th>Residual relative activityb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper loops</td>
<td>0.07</td>
</tr>
<tr>
<td>Helix 3</td>
<td>0.13</td>
</tr>
<tr>
<td>Helix 2</td>
<td>0.16</td>
</tr>
<tr>
<td>Canopy</td>
<td>0.19</td>
</tr>
<tr>
<td>Surface loop</td>
<td>0.37</td>
</tr>
</tbody>
</table>

a All libraries have the same mutation frequency of 0.4%, or 1.5 mutations per gene.

b Paraoxonase hydrolysis activity, the wild-type-like G3C9 variant’s activity equal 1.
picked 300 active clones that exhibited the highest rate of color formation. The crude cell lysates of each of the 300 active clones was assayed spectrophotometrically for hydrolysis of three different substrates (2NA, paraoxon and γ-thiobutyrolactone). Eighteen out of the 300 active clones had markedly different substrate specificities from wild-type PON1. These were assayed in triplicate and sequenced. None had the wild-type sequence. Five clones harbored indels and substitutions (not shown) and the rest comprised PON1 sequences with indels only; nine of which were unique sequences. Of the nine unique sequences, six indels occurred in Helix 3, highlighting its importance in substrate recognition (Fig. 4). In addition, four of the six indels identified in Helix 3 occurred at residue 291, including insertions of one and two amino acids (Trp in NA087 and Thr-Gly in NA235) and a deletion of 291 (NA311 and NA332). All three of these clones showed higher esterase activity than wild-type, whereas the two other assayed activities (paraoxon and γ-thiobutyrolactone) were reduced by more than 50-fold. Conversely, two other Helix 3 mutants showed a bias toward thiolactonase and phosphotriesterase activities (NA070 and NA079, respectively).

Two clones had insertions at the surface loop: NA024 had an insertion of two amino acids between residues 82 and 83 (Val-Tyr) and ‘specialized’ as esterase. Clone NA311 carried an insertion also at the surface loop (Gly, between residues 74 and 75) as well as a deletion of H3 residue 291. The substrate specificity of this mutant is very similar to that of NA332 and is therefore most likely dictated by the deletion in H3, rather than by the surface loop insertion (although this change, not having been examined here, may exert a subtle influence). As expected from the analysis of whole libraries (Table II), only a minority of clones (2/9) harbored insertions in the indel-sensitive upper loops, although interesting improvements were recorded. NA073 had an insertion of Leu in loop 2D3A (i.e. the loop connecting the D strand of the second blade with the A strand of the third blade) and exhibited increased thiolactonase activity. NA198 harbored a Gly insertion at loop 2B2C and exhibited increased phosphotriesterase activity.

Discussion

Optimization and application of the ISOR protocol

Directed evolution methodologies need not be seen necessarily as a way of circumventing design, but rather as a way of complementing it by allowing for a much larger margin of error. However, despite the availability of numerous methods for directed and random mutagenesis (Arnold and Georgiou, 2003; Neylon, 2004), there is still a need for techniques that allow a systematic design of gene libraries informed by inputs from rational, or computational design. Prior to this study, we had made attempts to perform gene assembly from long synthetic oligonucleotides (60–80 bp) that were designed to introduce the targeted diversified residues in a manner similar to the synthetic shuffling method (Ness et al., 2002; Zha et al., 2003). We found, however, that the libraries constructed in this way were very sensitive to oligonucleotide quality, and that long oligonucleotides contain a significant fraction of \((n-1)\) and \((n+1)\) products. What is more, purification of these oligonucleotides by PAGE resulted in an even higher frequency of frame-shifts. Chastened by these experiences, we directed our efforts at developing a general

![Fig. 4. Activity and sequence changes in a representative group of active variants isolated from the PON1 indels library. Activities of each variant are indicated, relative to wild-type PON1 (=1). Three substrates—an aryl ester, 2NA (grey bar), a phosphotriester (paraoxon; white bar) and γ-thiobutyrolactone (black bar)—were tested. Activities were measured in cleared cell lysates in triplicate. Lower panel: the indels found in each clone variant and its location.](https://academic.oup.com/peds/article-abstract/20/5/219/2330127)
and versatile technique that targets diversity to pre-define, and specific positions, thereby creating the desired gene libraries with high precision. ISOR is the result of this effort, and is based on the incorporation of synthetic oligonucleotides via gene-reassembly (Fig. 1). The addition of synthetic oligonucleotides to a mixture of gene fragments prior to DNA shuffling was suggested in Stemmer’s original report (Stemmer, 1994). Perhaps, due to the lack of a systematic, well-established protocol, this approach has been only very rarely applied (van den Beucken et al., 2001; Stutzman-Engwall et al., 2005). In this work, we describe the optimization of this method, and its application toward the generation of a range of different targeted libraries, while incorporating base substitutions, insertions and deletions.

As shown here, relatively short synthetic oligonucleotides (~30 bp) can encode substitutions, insertions or deletions, at any given position with a high degree of precision. The frequency of errors in such short oligonucleotides, and their cost, is much lower than of long ones, and they require no chromatographic purification that increases costs and biases library content. ISOR, therefore, begins from a reliable starting point, yet is extremely versatile and adaptable. Once a set of oligonucleotides has been synthesized, it can be used for the assembly of various libraries with different rates of diversification, or indeed libraries created with different subsets of the same oligonucleotides. Since most of the gene sequence is reassembled from DNaseI fragments, and because the oligonucleotides used are short, the method is not very sensitive to oligonucleotide quality.

The major advantage of ISOR is its ‘tuneability’ in that it allows a ‘parsimonious’ representation of diversity in many positions (30–45, as demonstrated here), while affording the opportunity to control the mutation frequency at each targeted residue. In saturation mutagenesis, randomization of more than a few positions results in impossibly high library sizes and high mutation frequencies that render almost all library variants inactive and the identification of positives therefore depends on screening of an extremely large number of variants. Attempts have been made to overcome this drawback. For example, using oligonucleotides in which a small proportion of randomizing bases were ‘doped’ into the wild-type sequence. However, the applicability of this method, often called parsimonious mutagenesis (Balint and Larrick, 1993), has been limited, possibly due to the high cost and limited quality of doped oligonucleotides. The power of ISOR is in the fact that the concentration of each oligonucleotide in the assembly reaction determines the frequency of modification at each position (Fig. 2, Tables I and II). This is much more difficult to achieve with synthetic shuffling, especially when mutations in adjacent codons are encoded by the same oligonucleotide.

We demonstrated the benefits of ISOR in the preparation of targeted gene libraries of two enzymes. A bioinformatics analysis allowed us to define a set of 45 non-contiguous amino acids as moderately conserved in M.HaeIII. We then used ISOR to target its diversification power to these residues. The result was a series of libraries with a range of mutation frequencies (Table I). Each gene in the library carried different mutations, at a different subset of the targeted positions, and the entire set was therefore explored in a systematic, predictable and combinatorial manner. A library with an average mutation frequency of 5.5 mutations per gene was subjected to iterative rounds of selection, and gave rise to a range of active variants (Supplementary Table III).

**Directed evolution with indels**

Almost all gene libraries described to date have employed point mutations. The application of indels libraries to protein engineering, therefore, has thus far been limited—also due to the lack of an appropriate methodology (for a newly developed technique for random indels incorporation see Fuji et al., 2006). In order to study the undiscovered role of indels in the evolution of new enzyme functions, we generated indels libraries by ISOR. The enzyme under investigation was PON1, a calcium-dependent lactonase that exhibits a range of promiscuous activities (Aharoni et al., 2005a, 2005b; Khersonsky and Tawfik, 2005). ISOR was used to target indels to surface loops, and various structural elements within them, that comprise the wall and perimeter of PON1’s active site. These regions contain the vast majority of residues that are believed to dictate the substrate selectivity of this enzyme family, and seem to have changed over the course of its natural divergence (Harel et al., 2004). Consequently, they seem the most promising for altering the enzyme’s specificity.

We designed a set of 90 oligonucleotides encoding indels throughout the identified structural elements. The versatility of ISOR allowed us to create an array of libraries, including libraries with indels in individual structural elements, and others with indels distributed along several structural elements. Characterization of the constructed libraries before, and after, an activity screen, indicated the potential of ISOR to generate highly complex combinatorial libraries with insertions, deletions or both. Oligonucleotide incorporation proceeded in a highly efficient manner, with no apparent biases and with a minimal number of random point mutations due only to PCR errors.

We also demonstrated the potential of ISOR to elucidate the tolerance of different structural elements to indels. We found that the ‘canopy’ elements, and a highly mobile ‘surface loop’, are much more tolerant to indels than the shorter ‘upper loops’ (Table II). Although these results are obviously preliminary, they seem to indicate that longer, and certainly more mobile loops (the ‘surface loop’ is disordered in the crystal structure), are more tolerant of indels than short, highly ordered loops (for other recent examples see, Scalley-Kim et al., 2003; Mathonet et al., 2006, and references therein).

Underlining the dichotomy between short, ordered and longer disordered loops is the finding that indels in several short loops (e.g. 6D1A and 5D6A) were not found among any of the active library variants (Fig. 4; data not shown). Another interesting observation is that most of the modified variants (66.7%) harbored indels in Helix 3 (H3) which is known to play an important role in substrate recognition (Aharoni et al., 2004; Harel et al., 2004). However, incorporation of indels at this surface loop did not give rise to many variants with altered specificities despite its high tolerance.

In summary, although the above results regarding the role of indels in altering PON1’s catalytic activities are preliminary, they do indicate the potential of indels libraries for directed evolution. Foremost, these libraries, and the one of M.HaeIII, demonstrate the versatility and applicability of ISOR.
Supplementary Data

Supplementary data mentioned in the text is available to online subscribers at http://www.peds.oxfordjournals.org.

Acknowledgments

A.H. thanks Professor Michael Fry and Dr Manel Camps for critical reading of the manuscript, and Professor Lawrence A. Loeb for his support. Financial funding was provided by the Israel Science Foundation, and is gratefully acknowledged. D.S.T. is the incumbent of the Elaine Blonde Career development Chair.

References


Received February 17, 2007; revised February 17, 2007; accepted February 26, 2007

Edited by Phillip Holliger