Protein evolution by codon-based random deletions

Joel Osuna, Jorge Yáñez, Xavier Soberón and Paul Gaytán*

Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología/UNAM, Ap. Postal 510-3 Cuernavaca, Morelos 62250, México

Received August 13, 2004; Revised September 10, 2004; Accepted September 15, 2004

ABSTRACT

A method to delete in-phase codons throughout a defined target region of a gene has been developed. This approach, named the codon-based random deletion (COBARDE) method, is able to delete complete codons in a random and combinatorial mode. Robustness, automation and fine-tuning of the mutagenesis rate are essential characteristics of the method, which is based on the assembly of oligonucleotides and on the use of two transient orthogonal protecting groups during the chemical synthesis. The performance of the method for protein function evolution was demonstrated by changing the substrate specificity of TEM-1 β-lactamase. Functional ceftazidime-resistant β-lactamase variants containing several deleted residues inside the catalytically important omega-loop region were found. The results show that the COBARDE method is a useful new molecular tool to access previously unexplorable sequence space.

INTRODUCTION

During the last twenty years, we have seen the appearance and disappearance of many approaches for the modification of enzymes. The success of a new mutagenesis method to remain inside the protein engineering toolbox (1–6) depends to a large extent on its simplicity and the quality of DNA diversification.

Protein engineers interested in rational design strategies (structure-based site-specific mutagenesis) have traditionally relied on cassette mutagenesis for introducing single or site-saturation mutagenesis that sometimes can be approached in a combinatorial mode. Recently, the codon-based mutagenesis strategies have been described (7–9) to efficiently explore all amino acid substitutions in a target region. Although it remains true that the structure-based design strategies have had limited capabilities to improve protein properties, as compared with directed evolution approaches, the trend keeps changing. This is shown by the recently reported computational method for the rapid optimization of proteins (10), which eliminates sequences incompatible with the protein fold, reducing the number of functional sequences to a size amenable to experimental screening, and finally helping in the successful redesign of an enzyme.

Unfortunately, this new computational methodology, as well as error-prone PCR (2) and gene shuffling (1), is suitable to evaluate only amino acid replacements. However, nature has widely demonstrated, through the generation of many homologous enzymes and antibodies, that the protein size may also be important to alter properties, such as recognition, activity, specificity and stability. In this regard, interesting methodologies have been developed either to enlarge or to shorten a target protein. These strategies expand the exploration of sequence space through the generation of random segment deletions, insertions or repeats (11–14).

Encouraged by our knowledge on DNA chemistry and driven by a fortuitous active amino acid deletion that arose in a previous research (15), we developed a novel codon-based random deletion (COBARDE) mutagenesis method that is reported here. This approach relies on oligonucleotide synthesis and on the use of two transient, orthogonal protecting groups. The method is able to explore all combinations of amino acid deletions, from singles to multiples, in any protein region comprising several residues and avoids the high rate production of frameshift mutations as occurred in previously described methods (16,17).

Testing of COBARDE was performed in the region 170–178 of the clinically important protein TEM-1 β-lactamase. The aim is to modify the substrate specificity of the enzyme as well as to address the possibility of minimizing part of the enzyme active site with limited or no detriment to the wild-type function.

TEM-1 is a class A β-lactamase composed of 263 residues, which contains among several important residues, an omega-loop structure comprising the residues 164–179. This structure covers the substrate entry to the active site. Class A β-lactamases are commonly known as penicillinases due to their well-known capacity to hydrolyze penicillin substrates, although they still display trace activity toward cephalosporin substrates. Many groups (18–21) have carried out not only systematic amino acid replacements of the omega loop, but also insertion and deletion of segments to study the role of this region to the enzyme–substrate profile. Our results agree to those described in previous works, although we made a better exploration of the sequence space, obtaining never described ceftazidime-resistant mutants and interesting results about ampicillin hydrolysis.

MATERIALS AND METHODS

The antibiotics ampicillin (AMP) and kanamycin (KAN) were purchased from Sigma Chemical Co., whereas ceftazidime (CAZ) was from Glaxo. All restriction endonucleases and
modification enzymes were from New England Biolabs. Plasmid pBR322 and deoxynucleoside triphosphates (dNTPs) were bought from Boehringer Mannheim. Plasmid pT4BlaSac was the parent plasmid for the library of TEM-1 genes carrying codon deletions. Plasmid pT4BlaSac contains the TEM-1 β-lactamase gene, including its original promoter, a KAN resistance gene and the CoIE1 and f1 origins of DNA replication within 3.2 kb. This plasmid contains two unique restriction sites of SacI (496-gaacctg-501) and PstI (541-ctgcag-546) within the blaTEM-1 gene. *Escherichia coli* XL1-Blue cells (recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lacI[F proAB, lacY1 lacZΔM15, Tn 10(ter101)]) was obtained from Stratagene Inc., and used for the propagation of the plasmids.

The chemical reagents used for oligonucleotide assembly were purchased from different suppliers. 5-(ethylthio)-1H-tetrazole (ETT) and 5'-O-(4,4'-dimethoxytrityl)-methyl amidites of dAbz, dCbz, dGiBu and dT were from Chemgenes. Iodine, pyridine, acetic anhydride (Ac2O), 1-methylimidazolide (NMI), trichloroacetic acid, ammonium hydroxide (NH4OH), 9-fluorenylmethoxycarbonyl chloride (Fmoc-Cl), 4-(dimethylamino)pyridine (DMAP), 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU), thiophenol and triethylamine were from Aldrich. Dichloromethane and tetrahydrofuran were from Burdick & Jackson, whereas anhydrous acetonitrile was from American Bioanalytical.

**Oligonucleotides carrying COBARDEs**

The oligonucleotide library carrying codon deletions was assembled in complete automated mode using the ASM-800 DNA synthesizer from Biosett Ltd (Novosibirsk, Russia), in a single synthesis column at 0.2 µmol scale. The synthesis protocol is supplied as Supplementary Material.

While some DNA synthesizers are supplied with five vial positions for loading phosphorimidates and six bottle positions for ancillary reagents, ASM-800 is supplied with eight vial positions and 10 bottle positions to perform special DNA chemistries (www.biossett.com). The synthesizer, shown in a Supplementary Figure, was loaded with conventional reagents for oligonucleotide synthesis, excepting monomers. DMTr-methylamidites of dAbz, dCbz, dGiBu and dT were from Chemgenes. Iodine, pyridine, acetic anhydride (Ac2O), 1-methylimidazolide (NMI), trichloroacetic acid, ammonium hydroxide (NH4OH), 9-fluorenylmethoxycarbonyl chloride (Fmoc-Cl), 4-(dimethylamino)pyridine (DMAP), 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU), thiophenol and triethylamine were from Aldrich. Dichloromethane and tetrahydrofuran were from Burdick & Jackson, whereas anhydrous acetonitrile was from American Bioanalytical.

**Cloning of mutant cassettes**

Duplexes of the target region were generated by the extension of the complementary primer over the mutagenic oligonucleotide, using the Klenow fragment (3′→5′ exo-) of DNA polymerase I. Briefly, 500 pmol each of the complementary and mutagenic oligonucleotides were mixed in Klenow buffer containing all four dNTPs. The mixture was heated to 70°C for 15 min and allowed to cool to room temperature. Klenow fragment was added and the mixture was incubated at 37°C for 1.5 h. The enzyme was inactivated by heating to 70°C for 15 min, and the restriction enzymes SacI and PstI were added using the appropriate enzyme buffer. The restriction enzyme reactions were continued for several hours at 37°C and then the enzymes were inactivated by heating at 70°C for 15 min. Finally, ~40 pmol of the putative double-stranded and double-digested mutagenic fragment DNAs were ligated, using the T4 DNA ligase, to 2 pmol of pT4BlaSac cloning vector previously digested with restriction enzymes SacI and PstI. The recombinant plasmids were electroporated into XL1-Blue cells. A library of a few thousand clones should be enough to represent each member of the 512 (29) possible variants. Several clones coming from this library were sequenced to evaluate the frequency and distribution of mutations. The plasmid library was re-electroporated to XL1-Blue cells, and different amounts of transformants were allowed to grow on selective plates containing either ampicillin or ceftazidime.
Antibiotic susceptibility
An aliquot of 5 μl of an overnight culture of XL1-blue harbor-
ing different variants were spotted after a 10⁻³ or 10⁻⁵ fold
dilution (10–50 colonies appeared without antibiotic in the last
dilution) with fresh Luria–Bertani (LB) medium onto LB KAN
plates containing increasing concentrations of ampicillin or
ceftazidime. The minimum concentration of antibiotic com-
pletely inhibiting the growth of cells was taken as the MIC.

Preparation of cell extracts and western blotting
The cells from an overnight culture were used to inoculate 5 ml
of LB medium containing the appropriate antibiotics. The
cells were allowed to reach stationary phase. The final optical
density (OD) of the culture was measured and the cells were
collected by centrifugation. The cell pellet was resuspended
with 300–500 μl of phosphate-buffered solution (PBS) buffer,
depending on the OD reached by any individual culture. The
cell suspension was sonicated (three 30 s pulses at 0°C) with
a model 450 Branson Sonifier. After one centrifugation step,
soluble (S) protein samples were taken for the western
blotting.

Western blotting was performed according to the estab-
lished procedures (22,23). Three percent BSA in PBS was
used to block any non-specific sites in the nitrocellulose
paper. For β-lactamase protein detection, a rabbit polyclonal
antibody specific for TEM-1 β-lactamase was used as the first
antibody. The protein band was revealed using an anti-rabbit
secondary alkaline phosphatase-conjugated antibody and
ready-to-use alkaline phosphatase substrates from ICN. The
expression levels were determined by densitometric analysis.

RESULTS
COBARDE method
Taking advantage of our experience in DNA chemistry, we
devised a powerful and easily automated mutagenesis method
to generate COBARDEs through the assembly of synthetic
oligonucleotides. Subsequent incorporation of the mutant
oligonucleotides into structural genes and their respective
expression in an appropriate host would permit the study
of the effect of random amino acid deletions in important
functional regions of proteins. Considering that ordinary oli-
gonucleotide synthesis is performed assembling monomer-
amidites protected with the DMTr group (24) and that
Fmoc and DMTr are orthogonal protecting groups (8)
(Fmoc is labile to base and stable to acid, whereas DMTr
displays opposite properties), we thought it would be possible
to arrest a fraction of the growing oligo, precisely in the
nucleotide previous to the codon to be deleted, by reacting
the solid support with a diluted solution of Fmoc-Cl. The
remaining 5'-OH unprotected oligonucleotide chains would
then be subjected to three cycles of ordinary synthesis to
incorporate 3 nt according to the wild-type sequence. At
this point, both groups, Fmoc and DMTr, would be sub-
sequently removed and the process of mutagenesis is
repeated as many times as necessary. The process is outlined
in Figure 1.

The number of deleted codons per oligonucleotide in the
library would be directly related to the rate of Fmoc block-
ing. For example, if the Fmoc-Cl solution is very dilute then
very few oligonucleotide chains would be blocked and the
library would mainly contain oligonucleotides of wild-type

Figure 1. COBARDE method exemplified for a region of three codons. (A) Sequence to be assembled oriented in a 3'→5' direction. Bold, italic and underlined bases
represent codons to be deleted. Plain codons represent wild-type flanking regions. All mutant oligos will bear these flanking regions for cloning purposes, either to
introduce them as cassettes or as PCR mutagenic primers. (B) Assembly process. Synthesis starts by assembling the 3' flanking region through ordinary chemistry,
using DMTr-monomer amidites. Last DMTr group is removed and a fraction of the sequence is transiently blocked by a reaction with diluted Fmoc-Cl. The unblocked
sequence is extended 3 nt by ordinary chemistry. Both Fmoc and DMTr protecting groups are removed by sequential treatment with alkali and acid, respectively. The
mutagenic cycle is repeated twice and all the variants are ended with the 5' flanking sequence. (C) Expected mutants after three cycles of mutagenesis.
sequence or lacking only one codon in different places of the evaluated region. Conversely, if the Fmoc-Cl solution is very concentrated, most of the chains would be blocked at every mutagenic cycle and most oligonucleotide variants would contain several codon deletions in different positions of the sequence. Therefore COBARDE is able, through fine-tuning of the Fmoc-Cl concentration, to give rise to a controlled frequency of variants following a binomial distribution. In this regard, one could generate oligonucleotide libraries enriched for one, two, three or more codons deleted. Furthermore, a simple gel purification step could easily help to concentrate the characterization effort to any particular subset of mutants.

To demonstrate the viability of COBARDE method, we developed a synthesis program to fully automate the process in a DNA synthesizer equipped with enough vials and bottles to perform special DNA chemistries. The protocol is detailed in Materials and Methods, whereas the program is available as Supplementary Material. This mutagenic approach may be easily implemented in any commercial DNA synthesizer equipped with three additional vials or bottles to the 10 vials/bottles ordinarily used. In addition to the Fmoc-Cl solution, it was necessary to use a solution of 4-DMAP in acetonitrile as activating agent to complete the transient blocking. In fact, in a preliminary effort to simplify the process, we tested conventional capping B (made up of 1-NMI) as putative activator, but there was no blocking reaction. From this result, it is obvious that 1-NMI is able to activate anhydrides such as acetic anhydride, capping A, but not acyl chlorides. In the third additional bottle, a DBU solution in acetonitrile was loaded to remove the Fmoc group before a chlorides. In the third additional bottle, a DBU solution in acetonitrile was loaded to remove the Fmoc group before a chlorides. In the third additional bottle, a DBU solution in acetonitrile was loaded to remove the Fmoc group before a chlorides.

For testing of the method, we assembled the nucleotide sequence coding for the amino acids 170–178 of the β-lactamase protein: S′-gaacccagtctaatgacgctataccaagagactgtgatccagctgc-3′. Bold bases generated partial blockage of the previous 3′ base using the Fmoc group. For this experiment, the Fmoc-Cl solution was 20 mM, giving a mutagenesis rate of 41% for the first deleted codon (cgt). Because measurement of mutagenesis rate for the following codons is less reliable due to the presence of DMTr and Fmoc groups in the growing chain, we considered 41% as the average mutagenesis rate for all the codons. At this point, it is important to recall that oligonucleotide synthesis proceeds in 3′→5′ direction and that a blocked chain will give rise to deletion of the 5′ nearest codon. The real combinatorial effect takes place 3 nt ahead when the Fmoc group and the DMTr group that protects the last coupled nucleotide are removed and the mutagenesis cycle is repeated.

After complete assembly, the oligonucleotide was submitted to a two-step deprotection procedure to eliminate the internucleotidic methyl-protecting groups and the base-protecting groups. The crude purification was purified by electrophoresis in a polyacrylamide gel, in order to eliminate truncated sequences (Figure 2). This library must contain a population of 512 variants: 1 wild-type, 9 singles, 36 doubles, 84 triples, 126 quadruples, 126 quintuples, 84 hextuples, 36 heptuples, 9 octuples and 1 nonuple.

Codon-based random amino acid deletions in the omega-loop region of β-lactamase

As described in Materials and Methods, a SacI restriction site was introduced by site-directed mutagenesis at the β-lactamase gene region coding for amino acids 168 and 169. This new restriction site and the PsiI site, naturally present at the gene region coding for amino acids 183–185, facilitated to use the C-end (amino acids 170–178) of the catalytically important omega-loop of β-lactamase as a model target to test our new mutagenesis strategy.

Using standard molecular biology methods, a pool of mutant cassettes was constructed containing deletions in the β-lactamase gene region coding for positions 170–178 of the omega loop. This library was then characterized under selective and non-selective growth conditions.

As shown in Table 1, every amino acid present in the mutagenesis target region was deleted in about one-third of the randomly picked non-selected clones. Of the nine target positions, six were deleted in ~50% of the analyzed clones, correlating well with the expected deletion rate obeying a binomial distribution, planned for the mutagenesis experiment. Only 2 from 34 sequenced clones carried undesired not-designed alterations. Four clones (~11% of the analyzed clones) presented no deletions at all, but the absence in these clones of a designed silent nucleotide change introduced with the mutagenic oligos (471C→T) indicated that a poorly prepared cloning vector was the source of this wild-type omega-loop contamination. Table 1 shows an apparently random distribution of the different-sized gaps around the mutagenesis region.

We understand that this library is a source of never before explored sequence space in the enzyme active site, and we challenged it in selection experiments aimed at looking for
functional variants able to catalyze the hydrolysis of a normal penicillin substrate (ampicillin) or to find variants showing a substrate specificity change (ceftazidime hydrolysis).

Fifteen clones were obtained that grow in ceftazidime concentrations (1 mg/ml) six times above the minimum inhibitory concentration of clones containing the wild-type enzyme (0.15 mg/ml). Sequence analysis resulted in only eight different clones (one single clone appeared five times, and two clones appeared twice during the selection step). Due to the wild-type omega-loop contamination problem above described, we were able to pick up only three clones (from 150 analyzed clones) growing at ampicillin concentrations (100 mg/ml) well above the reported ampicillin minimum inhibitory concentration (1 mg/ml) of the host cell.

Inspection of the nucleotide sequences shown in Table 2 indicates that amino acid positions 170 and 178 were rarely deleted in the functional variant group. On the other hand, all the variants selected with ceftazidime lost the amino acid at position 173. In general, the ceftazidime-selected group shows a tendency to have large gaps (4–6 deleted residues) around positions 172–177. Opposite to the results obtained with the ceftazidime-selected group, the three variants obtained from the ampicillin-selected experiments indicate that only small gaps still allow the enzyme to keep the ability to hydrolyze the substrate. Despite showing smaller gaps, MIC analysis (Table 2) indicates that the ampicillin-selected group is functionally poorer than the ceftazidime-selected group. Two ceftazidime-selected clones (CAZ2 and CAZ7) presented high MIC values against the two tested antibiotics. Low expression levels relative to the wild-type enzyme were observed for most of the mutants (Table 2). This outcome might indicate that most of the amino acid deletions resulted in impairment of the protein stability. Inspection of the solvent accessibility surface area calculated for all the amino acid residues present in the mutagenesis target region shows that all the side chains are partially or totally exposed to solvent. Furthermore, the main-chain atoms of positions 171–176 are also partially or totally exposed to solvent. Amino acids 170, 177 and 178 are the ones with larger main-chain surface area excluded from the solvent as compared with the rest of residues in the mutagenesis target window. The above results could explain the observed gap position preference around residues 171–177 in the ceftazidime-selected group. Despite the very small number of ampicillin-selected clones, there is also an observed preference to tolerate residue deletions around positions 171–175.

---

**Table 1. Nucleotide sequence of non-selected clones**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence</th>
<th>Del*</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>GAGCTC GAC CGT GATACCACGATGCCTGCAG</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>GAGCTC GAA CGT GATACCACGATGCCTGCAG</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>GAGCTC GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>GAGCTC GCC CGT GATACCACGATGCCTGCAG</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>GAGCTC GCC CGT GATACCACGATGCCTGCAG</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>5</td>
</tr>
<tr>
<td>16</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>6</td>
</tr>
<tr>
<td>17</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>7</td>
</tr>
<tr>
<td>18</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>5</td>
</tr>
<tr>
<td>21</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>6</td>
</tr>
<tr>
<td>22</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>7</td>
</tr>
<tr>
<td>23</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>8</td>
</tr>
<tr>
<td>24</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>9</td>
</tr>
<tr>
<td>25</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>10</td>
</tr>
<tr>
<td>26</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>11</td>
</tr>
<tr>
<td>27</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>12</td>
</tr>
<tr>
<td>28</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>13</td>
</tr>
<tr>
<td>29</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>14</td>
</tr>
<tr>
<td>30</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>15</td>
</tr>
<tr>
<td>31</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>16</td>
</tr>
<tr>
<td>32</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>17</td>
</tr>
<tr>
<td>33</td>
<td>GAGCTC AAT GAA GCC ATA CCA AAC GAC GAG CGT GACACCCAGATGCTGCAG</td>
<td>18</td>
</tr>
</tbody>
</table>

*Del indicates the number of deleted codons per variant.

**Clones 2 displayed an insertion of 12 bases (TTGGTGCCATTG) in the 3' flanking region. This mutation may have arisen during the enzymatic assembly of the cassettes, due to mis-pairing between the primer and the mutant oligo.**

**Clones 8, 11, 18 and 32 resulted from religation of the vector. Clone 15 shows six deleted codons and 1 nt insertion as a by-product of the synthesis.**
DISCUSSION

Shortle and Sondek (25) envisioned the importance of insertions and deletions as new tools for experimentally altering protein structure and function. Insertions and deletions (indels) are important events in protein evolution as are substitutions, although the latter emerge more frequently. Recently, it has been described that indels can be partially responsible for dramatic structural changes observed within evolutionary related proteins (26). Pascarella and Argos (27) found that, in general, proteins tend to tolerate short indels (1–5 residues) and that interruption in helices and strands were very rare events. Exhaustive work from Matthews (28), Shortle (29) and Sauer (30) groups evaluating the importance of protein compactness for the maintenance of functional properties indicated that amino acid substitutions are more easily adjusted in the proteins than amino acid deletions. Although this conclusion may be right, what is clear is that altering the residue spacing in proteins can also contribute to diversify their functions (25). For instance, it was recently shown that single-codon deletions in CDRH2 of human antibodies are structurally well tolerated and the protein variants still retain good antigen recognition (31).

To date, the technology and reagents to generate codon-based random amino acid substitutions and insertions (32) is well established. This is not so with the technology to generate amino acid deletions, either in a single or combinatorial mode where there is still a large gap. Perhaps the major problem associated with amino acid deletions is the known fact that oligonucleotide synthesis is performed adding nucleotide by nucleotide. If nucleotide-based deletions are deliberately induced along the assembly, e.g. through incomplete monomer coupling (16) or incomplete detritylation (17), then a high ratio of frameshift deletions will be generated resulting in very low yields of genes encoding fully complete polypeptides.

The only solution to generate in-phase amino acid deletions is employing a real codon-based approach. To our knowledge, there are only two reported methods that could be used to generate random combinatorial amino acid deletions. One of them is the resin-splitting method reported by Glaser et al. (7), in which a small amount of the growing oligo can be removed from the synthesis column and reincorporated after three cycles of synthesis. If the process is repeated several times, a library of oligonucleotides bearing random codon deletions is created. However, this method is so cumbersome, laborious and difficult to automate that has been only occasionally used even to perform amino acid substitutions. The second method, patented by Shortle and Sondek (33), although theoretically viable has never been experimentally demonstrated. He proposes to block part of the growing oligo with a monomer-amidite protected in its 5'-OH with a group resistant to all chemical steps used during conventional synthesis, and after four addition cycles of wild-type monomers the group must be removed and the synthesis resumed as normal. However, a careful analysis of this strategy of mutagenesis immediately reveals an important shortcoming of the proposed mutagenesis protocol: the method is only able to delete alternated but not contiguous codons in a region of a gene, and in this sense many interesting combinations of amino acid deletions in the protein could be lost.

In this work, we have described a new fully automatable mutagenesis method capable of producing combinatorial random deletions of amino acids in any chosen protein target region. The method was able to experimentally explore a very rich portion of sequence space never before produced in the β-lactamase active site (Table 1). Furthermore, since our mutagenic approach produces oligonucleotides of different length with regard to the wild-type sequence, the method could be easily adjusted to explore a specific subset of variants, selecting the appropriate band by gel electrophoresis (Figure 2). This selection step at the oligonucleotide level, before cloning, could be of great value for strategies that depend on screening to find the improved variants.

Our results show that β-lactamase is able to tolerate several amino acid deletions in the omega loop without entirely losing its catalytic properties (as shown by bacterial antibiotic resistance levels). Analysis of the protein functional variants illustrates the power of this new technology: the ampicillin-selected variant AMP2, containing a double deletion (amino acids 172 and 173), confers to bacteria ampicillin resistance levels to ~300 μg/ml. Introducing a new deletion at position 177 (ceftazidime-selected variant CAZ8) roughly maintains the ampicillin MIC and at the same time alters the substrate
specificity (ceftazidime MIC ~0.8 μg/ml). Moving the new deletion to position 175 (ceftazidime-selected variant CAZ2 containing deletions at positions 172, 173 and 175) results in a protein able to confer the wild-type enzyme ampicillin resistance levels to bacteria and keeping the substrate specificity change. A weakness of the present technology is the trend to obtain protein variants showing putative stability problems (as shown by the very low expression levels for most of the variants). Fortunately, by using present protein evolution technologies it could be easy to recover the stability properties of any interesting variant. Another drawback of this new technology is that previous knowledge of protein structure and active site of an enzyme are required to identify the potential target regions. Recently, a random insertion/deletion mutagenesis (RID) method (13) was described, which is capable of deleting three randomly selected consecutive bases in a target gene. If knowledge of the protein structure or mechanism of the enzyme is lacking, random single amino acid deletions using the RID method will help to identify some interesting protein target regions that can be optimized with the new technology we are describing.

Recently, Kitamura et al. (34), using the Y-ligation-based block shuffling technology, were unable to find functional green fluorescent proteins (GFPs) containing random deletions at three consecutive amino acids in three different protein regions. The proposed method described here is ideally poised to search whether the same GFP regions are able to tolerate three amino acid deletions but distributed in a combinatorial way in the target region.

Finally, we note that a combination of synthetic oligonucleotides made using this mutagenesis strategy with gene shuffling (1) or synthetic shuffling (35) should lead to a protein evolution process that allows the exploration of a much larger protein sequence space.

In conclusion, a combinatorial COBARDE mutagenesis method was developed for exploring an entirely new sequence space of an enzyme active site. The method was able to produce functional β-lactamases with amino acid deletions at the catalytically important omega loop. The robustness of the method and its ease of implementation should make COBARDE a technique of choice in any enzyme engineering effort.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank Eugenio López and Santiago Becerra for their technical assistance and René Hernández for DNA sequencing. Viviana Escobar kindly supplied us the rabbit polyclonal antibody specific for TEM-1 β-lactamase. We are also indebted to Dr Hugh Mackie and Dr Lorenzo Segovia for critical reviewing of the manuscript. CONACYT grant NC230 and PAPIIT grant IN223199 are fully appreciated.

REFERENCES