Chemistry & Biology, Vol. 10, 917–926, October, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2003.09.007

Protein Evolution by "Codon Shuffling": A Novel Method for Generating Highly Variant Mutant Libraries by Assembly of Hexamer DNA Duplexes

Sidharth Chopra and Anand Ranganathan*
Recombinant Gene Products Group
International Centre for Genetic Engineering and Biotechnology
Aruna Asaf Ali Marg
New Delhi 110067
India

Summary

The laboratory evolution of functional proteins holds great fascination as an effective tool for searching and discovering novelty in the protein space. Here, we employ a simple molecular approach for generating a multitude of structurally diverse yet functionally similar proteins that have all "evolved" from the parent enzyme, in this case a β-lactamase. By randomly combining a set of 14 DNA-hexamer duplexes, each corresponding to judiciously chosen amino acid pairs, we were able to generate functional proteins that contained large regions of previously unknown sequences. Some of the generated proteins were not only more active than the parent, they were also significantly smaller in size. Our approach could therefore be effectively used in searching for novel structural as well as functional proteins.

Introduction

Charles Darwin first introduced the concept of "Unconscious Evolution" to describe the sudden improvement of an individual member of a species through generations of selective breeding and natural stress selection, by means of processes beyond the control and understanding of man [1]. The idea of laboratory evolution or directed evolution [2–4] of proteins and nucleic acids, however, is only as recent as a decade old, although much progress has since been made [5–9]. The use of degenerate oligonucleotides for cassette mutagenesis, sometimes referred to as "saturation mutagenesis" has been helpful for directed evolution as well as protein structure and function studies, particularly as it allows one to obtain mutants that may differ in every base position in a chosen stretch of DNA sequence [10–12]. Seminal studies by Loeb and coworkers as well as by Struhl and Oliphant have demonstrated the advantage of saturation mutagenesis over the earlier conventional methods [11, 12]. However, the variance is limited to a "portion" of the total gene length, which rarely exceeds 50 bp (correspondingly ~15–20 amino acids). This may be seen as a possible drawback especially in cases where mutations in the whole gene and not just one particular region are desired. Stemmer and coworkers were the first to address this problem by expanding upon the combinatorial potential of earlier techniques such as site-directed mutagenesis, error-prone PCR, and degenerate-oligonucleotide (or cassette) mutagenesis in order to create libraries of protein and nucleic acid mutants, with their gene-shuffling (or sexual PCR) technique [5, 13]. Since its inception, the gene-shuffling method has been used extensively to select for protein mutants, and there are now a host of new techniques that are variants of the gene-shuffling method [6–8]. The mutational rate in gene-shuffling ranges from 0.3% to 2% of the total length of the gene, a factor that may be seen as a drawback of the technique. Nonetheless, gene shuffling and saturation mutagenesis are currently the methods of choice in directed evolution experiments. More recently, Szostak and coworkers have developed a molecular evolution technique (mRNA display) by utilizing a large collection of synthetic degenerate DNA molecules in order to produce a library of proteins of equivalent size [14, 15]. However, these advances in molecular evolution of proteins are all not entirely devoid of limitations, ranging from low on degeneracy (i.e., error rate) to high on complexity of experimental design and expense [5, 14].

We now report an altogether novel approach toward protein evolution, one that introduces the concept of "codon shuffling" in order to obtain a vast number of shuffled degenerate gene sequences. The approach entails in vitro assembly of hexamer DNA duplexes that represent all 20 amino acids as codon pairs. The assembled "dicodons" are ligated with an inactive gene scaffold and the ligation mixture used to transform a bacterial host. Upon stress, the cell selects the appropriate shuffled gene whose length has been increased by the dicodon addition, for eventual translation into a catalytic protein that is fit to combat the provided stress (it may be noted that the term "codon shuffling" has been used here to represent the events that occur during the in vitro assembly process, i.e., the dicodons are shuffled in a random manner; the earlier usage of the term "shuffling," for example in gene shuffling, was to denote shuffling of the wild-type DNA sequences). We have applied this technique to an inactive β-lactamase parent gene scaffold, and upon codon shuffling, have selected progenies that are vastly different from the parent with some differing in their primary sequence by as much as 94%, while still retaining the β-lactamase activity.

Results and Discussion

Protein Evolution through Codon Shuffling

In order to design a concept of laboratory evolution of proteins that would either address or circumvent the limitations of earlier methods, we began by asking a simple question: can the degeneracy provided for by the genetic code at the nucleotide level be exploited to yield a library of proteins? After all, the degeneracy of a protein at its primary sequence level is the direct fallout of the degeneracy of its gene that is being translated (e.g., the assembly of a 300 base pair DNA fragment,

*Correspondence: anand@icgeb.res.in

made up of 100 codon positions that are to be filled up by a set of 20 provided codons having complete freedom to combine with each other, will result in a degeneracy of 20^{100} different fragments, each of which, when translated, will produce a unique protein). In theory, if a set of codons were to be ligated with each other, and the assembled DNA efficiently translated, one could end up with a collection of proteins whose degeneracy is a factor of the number of codons chosen in the particular set, as well as the manner in which the codons themselves have assembled. In practice, however, the ligation of a 3 base pair DNA (i.e., a codon) duplex with another codon duplex using T4 DNA ligase has not yet been reported, although it is common to ligate 6 base pair DNA duplexes with other DNA fragments of varying lengths [16].

Therefore, in an effort to develop over-and-above this theme of so-called "codon-shuffling," we carried out the following experiment: we randomly assembled in vitro, a collection of 6 base pair DNA duplexes (called dicodons) that represented all 20 amino acids, to form degenerate DNA fragments. The collection of dicodons was devoid of the three stop codon sequences so that all degenerate DNA fragments may be read always in a single desired open reading frame. This collection, comprising 14 dicodons, was based on the following "design rules": any given codon was paired up with a codon that is the exact complement of the former and in addition is a "neutral" codon; high-usage codons were used as representative of amino acids, with *E. coli* usage as the guide (the codon usage cable can be accessed at http://www.kazusa.or.jp/codon/); the collection was a reflection of the percentage of the type (e.g., neutral, polar, hydrophobic, etc.) of amino acids present in the parent protein that was to be evolved, as well as a reflection of the distribution of the various types of amino acids used in *E. coli*. The comparison of the percentage of amino acids present in an equimolar dicodon pool with *E. coli* total codon usage is shown in Supplemental Figure S1 at http://www.chembiol.com/cgi/content/full/10/10/917/DC1. As can be seen from the figure, the percentage values are similar to a large extent, thereby indicating that any progeny developed from the use of the equimolar dicodon pool would be as well represented in the various protein attributes (hydrophobic, acidic, basic nature, etc.), as are most proteins synthesized by *E. coli*. The collection of dicodons chosen for the experiment is shown in Table 1. Wherever possible, care was taken to associate any given codon with one that would be representative of a hydrophobic, small, or a nonpolar amino acid (thus termed as a "neutral" codon). The reasoning behind this was two fold. First, there is a preponderance of neutral codons in *E. coli*—the 7 amino acids Leu, Val, Ile, Met, Pro, Ala, and Gly together make up 50% of total codons used. Through such a pairing, therefore, the equimolar pool of 14 dicodons would work as a truer representative of the *E. coli* codon usage. Second, these so-called neutral codons represent amino acids that, among them, enjoy superior secondary-structure forming propensities (see reference 4 in the Supplemental Data available at *Chemistry & Biology's* website). Next, every codon was paired with its exact complement, as this would allow such single-stranded palindromic DNA molecules to associate spontaneously and form a homodimer, or duplex DNA upon cooling. Therefore, only those palindromes were selected wherein the possibility of a mismatched duplex was energetically unfavorable or nonexistent. The complete absence of a mismatched dicodon duplex in the ligation pool would mean that a frameshift in the reading frame was not possible, as only perfectly matched, hence blunt-ended dicodons would be present in the ligation reaction. All 20 natural amino acids were represented by the 14 dicodons (with amino acids Glu, Gly, Ala, Pro, Thr, and Val being represented twice and the amino acid Leu three times); the percentage hydrophobicity, defined as the percentage of amino acids having a positive value on a Kyte-Doolittle hydrophobicity scale in an equimolar dicodon pool, was 39%, compared with a value of 38% for the parent enzyme (see below; [17]).

Codon Shuffling of a Parent β-Lactamase

As a proof-of-concept experiment, we chose the TEM-1 enzyme, a 286 amino acid long class A β-lactamase as the parent protein [18]. TEM-1 is the best studied of all β-lactam hydrolyzing enzymes, with its active site comprising of a catalytic serine residue (S^{45}, Figure 1) to which the open form of the β-lactam is covalently linked [19]. Numerous structural and functional studies have indicated active involvement of a further 3–4 amino acid residues in catalysis, along with a string of 15 residues lying almost 100 residues downstream of the active site serine—the so-called ω loop—thought to play a crucial role during the binding and hydrolysis of the lactam moiety (Figure 2A) [20–22]. In fact, all of the residues that are involved in catalysis lie downstream of S^{45}, including a lysine (involved in the acylation step) that is just 2 residues away. In our belief, TEM-1 represents an attractive target for laboratory evolution using codon shuffling for many reasons. First, it is a reasonably large, monomeric, and soluble protein that can be assayed for its activity both in vivo and in vitro with ease using a variety of substrates. Second, it possesses an active site that is composed of residues that are well spread out in the primary sequence of the protein. Finally, as the results of codon shuffling would be such so as to overwhelmingly change the primary sequence of the parent protein, any new evolved protein would yield important structural and functional insights into the functioning of β-lactamases in general.

As a first step, we severely truncated the parent TEM-1 enzyme, so that it was now only 86 amino acids long (including the 23 amino acid signal peptide at the N terminus and a His×6 tag at the C terminus, respectively). The plasmid containing the gene for this protein, *pSC2*, was unable to confer resistance against any of the β-lactams tested (Table 2). This was not at all surprising considering that the truncated protein was all of only the first 53 amino acids of the wild-type TEM-1, with S^{45} and K^{48} being the only remnants left from the active site of the parent (Figure 2B). In order to build on this inactive protein scaffold, we engineered a dicodon insertion point in its gene, at a position 15 bases downstream of the K^{48} codon. The insertion point was in the form of the sequence for the restriction enzyme SnaBI (5′TAC$^{\downarrow}$GTA3′)

Table 1. The Dicodon Set Used for Codon-Shuffling Experiments

Oligonucleotide	Duplex	DC	Oligonucleotide	Duplex	DC
5′-GAGCTC-3′	5′-GAGCTC-3′ 3′-CTCGAG-5′	EL	5′-CCCGGG-3′	5′-CCCGGG-3′ 3′-GGGCCC-5′	PG
5′-GATATC-3′	5′-GATATC-3′ 3′-CTATAG-5′	DI	5′-ATGCAT-3′	5′-ATGCAT-3′ 3′-TACGTA-5′	MH
5′-AAGCTT-3′	5′-AAGCTT-3′ 3′-TTCGAA-5′	KL	5′-CAGCTG-3′	5′-CAGCTG-3′ 3′-GTCGAC-5′	QL
5′-AACGTT-3′	5′-AACGTT-3′ 3′-TTGCAA-5′	NV	5′-TGGCCA-3′	5′-TGGCCA-3′ 3′-ACCGGT-5′	WP
5′-GGCGCC-3′	5′-GGCGCC-3′ 3′-CCGCGG-5′	GA	5′-TACGTA-3′	5′-TACGTA-3′ 3′-ATGCAT-5′	YV
5′-AGTACT-3′	5′-AGTACT-3′ 3′-TCATGA-5′	ST	5′-TTCGAA-3′	5′-TTCGAA-3′ 3′-AAGCTT-5′	FE
5′-TGCGCA-3′	5′-TGCGCA-3′ 3′-ACGCGT-5′	CA	5′-CGTACG-3′	5′-CGTACG-3′ 3′-GCATGC-5′	RT

Amino acids are listed as single letter codes. DC, dicodon.

such that cleavage of the gene with this enzyme would result in blunt-ended products on either side of the excision—a TAC^{tyr} and a GTA^{val} codon, respectively. Plasmid *pSC2* was then excised with SnaBI and introduced to a soup of randomly assembled dicodons, followed by transformation of *E. coli* DH5α and plating on media

Figure 1. Primary Sequence Analyses of Obtained Mutant β-Lactamases

(A) Primary sequence of wild-type and mutant β-lactamases. For the proteins other than TEM-1, the regions in red depict the evolved sequences, while those in black represent the wild-type sequence. The catalytic serine (S^{45}) and lysine (K^{48}) residues are shown in bold. The conserved Tyr residue (boxed) in the mutant sequences is a result of the SnaBI dicodon insertion site design. %MR, percentage mutagenic rate, defined as the percentage of primary sequence that is different from the wild-type TEM-1; GRAVY, grand average of hydropathicity [17].

(B) A register of the number and type of dicodons found in the mutant β-lactamases. The numbers for the second generation mutants BlaSC7-9 do not include the first generation dicodon numbers.

(C) Profiling of mutant lactamases according to the amino acid attribute patterning developed by Hecht and coworkers. Polar amino acids, namely S, T, Y, W, Q, N, C, H, R, K, D, and E, are depicted as open circles while nonpolar amino acids, namely, G, A, I, V, L, P, F, and M, are shown as filled circles. The dicodon sequences are in red. The sequences are numbered according to (B). As a reference, sequence no. 7 shows the patterning of an insoluble protein that contains β sheet repeats and forms amyloid structures. This sequence was designed by Hecht and coworkers to test their hypothesis of obtaining amyloid type structures when alternating polar-nonpolar repeats (shown boxed) are used [24].

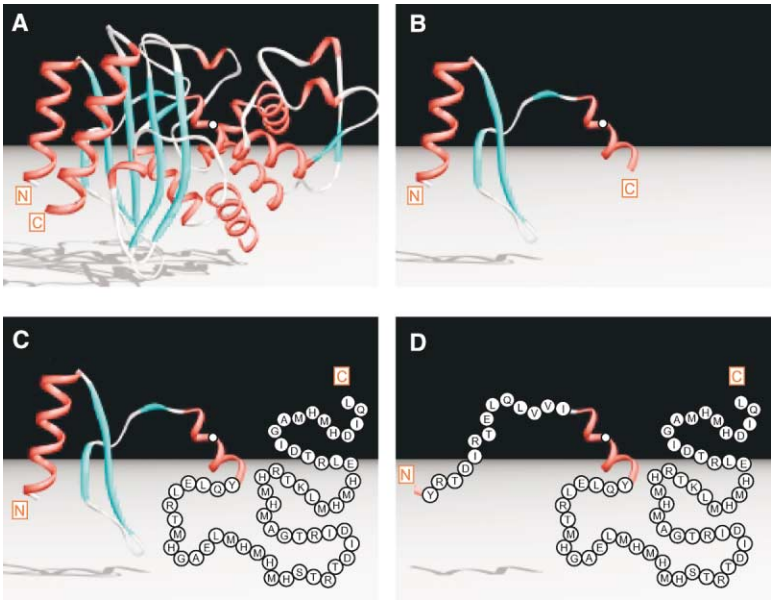

Figure 2. Hypothetical Representation of the Primary Sequence of Evolved Mutants

(A) Ribbon drawing of the 3D X-ray crystal structure of the 263 amino acid long TEM-1 β-lactamase at 1.8 Å resolution [32]. The N- and C termini are pointed in squares. The position of the active site serine residue is depicted with a white circle.

(B) Representation of the 57 amino acid long inactive scaffold BlaSC2, the corresponding gene of which was used for dicodon insertion experiment.

(C) The 115 residues long active mutant lactamase BlaSC3. The dicodon incorporate amino acid sequence is shown linked to the inactive BlaSC2 scaffold. The primary sequence of this protein is 79% different from that of its TEM-1 parent.

(D) The 91 residues long active second generation mutant lactamase BlaSC7. The ribbon that is flanked on both sides by the dicodon incorporate amino acid sequence is the TEM-1 active site "left-over" region and contains the active site serine residue. The primary sequence of this protein is 94% different from that of its TEM-1 parent. The figures were drawn using the Viewerlite and POV-Ray computer programs.

containing antibiotics (see Experimental Procedures). Out of the initial library size of approximately 10^3 transformants, we isolated three cfu that were highly resistant to ampicillin and amoxycillin (Table 2). It may be noted that the expression vector *pET28a*(+) was chosen as a starting vector for all our experiments as it possess a kanamycin-resistance gene and *not* a β-lactamase gene as the selection marker. However, in order to discount for the theoretical possibility that our observations were a result of an external contaminating β-lactamase gene, we isolated the plasmid DNA from the resistant strains and PCR amplified the causative agent for the resistance by using universal primers (the so-called universal primers—T7 forward and reverse primers—do not bind to the regions flanking the TEM-1 gene in any of the commonly used cloning vectors, like for example *pGMT*, *pBluescript*, *pUC*, or *pET*-based vectors, all of them being possible sources of a contaminating lactamase gene. On the other hand, the universal primers bind perfectly to the T7 promoter and terminator regions of a *pET28*-based vector. These regions thus flank any "evolving" gene and are therefore ideal primers for amplification of such a gene). The amplified genes were cloned first in a different vector (*pGEMT-Easy*) and then inserted back into the *pET28a* expression vector, followed by introduction of these new plasmids into *E. coli* DH5α. The resulting strains too were resistant to β-lactams, thus confirming our initial observations that the resis-

Table 2. Minimum Inhibitory Concentration Values for Various β-Lactams in Units of μg/ml

Strain/Plasmid (Protein)	AMP	AMX	PEN	CTX
DH5α/pSC1 (TEM-1)	20,000	>20,000	10,000	0.03
DH5α/pSC3 (BlaSC3)	40,000	>20,000	5,000	0.03
DH5α/pSC4 (BlaSC4)	10,000	10,000	500	0.03
DH5α/pSC5 (BlaSC5)	10,000	10,000	1,000	0.03
DH5α/pSC7 (BlaSC7)	20,000	>20,000	6,000	0.40
DH5α/pSC8 (BlaSC8)	10,000	>20,000	2,000	0.03
DH5α/pSC9 (BlaSC9)	10,000	>20,000	2,000	0.03
DH5α/pSC2	—[a]	—[a]	—[a]	—[b]
DH5α/pSC6	—	—	—	—
DH5α/pSC10 (BlaSC10)	—	—	—	—
DH5α/pSC1$^{S\rightarrow A}$ (TEM-1$^{S\rightarrow A}$)	—	—	—	—
DH5α/pSC3$^{S\rightarrow A}$ (BlaSC3$^{S\rightarrow A}$)	—	—	—	—
DH5α/pSC3$^{K\rightarrow A}$ (BlaSC3$^{K\rightarrow A}$)	—	—	—	—
DH5α/pSC7$^{S\rightarrow A}$ (BlaSC7$^{S\rightarrow A}$)	—	—	—	—
DH5α/pSC7$^{K\rightarrow A}$ (BlaSC7$^{K\rightarrow A}$)	—	—	—	—
DH5α/pET28a(+)	—	—	—	—

MIC, minimum inhibitory concentration.
[a] MIC values of less than 10 μg/ml.
[b] MIC values of less than 0.02 μg/ml.

tance in each case was only due to *pSC2* plasmid alterations brought about by codon shuffling. Additionally, we have also found that MBP gene fusion with the mutant genes leads to excellent overexpression of periplasmic fusion proteins of the predicted molecular weight. Although the interpretation of this result is discussed at a later point, the recloning of the mutant genes in different vectors, and as fusions with other genes, does indeed rule out the possibility of a lactamase contaminant carrying over repeatedly and through a series of PCR amplifications and reclonings. Furthermore, one of the first generation mutant genes, *BlaSC3*, has itself been used as a scaffold, to generate active, second generation mutants (see later), thus demonstrating that the active *BlaSC3* can be converted into an inactive scaffold and then "reactivated" upon further evolution—a switch that would not have been possible were a contaminating lactamase to be present. This is because the presence of the contaminant would always show colony growth on ampicillin, even at the point where an inactive *BlaSC3* was being tested for its resistance to the antibiotic.

Library Size

Our estimate of the library size was based on the observation that greater than 10^3 cfu were obtained when the vector-locked dicodon fragments (i.e., the ligation mixture) were plated on media containing *only* kanamycin, while plating of just the self-ligated SnaBI-cut and dephosphorylated vector on media containing kanamycin resulted in not more than 20 cfu. This number, when subtracted from the first value, gives the library size. It can now be configured that 3 cfu out of this library were able to overcome the ampicillin as well as kanamycin stress in the dicodon experiment. This is at best a crude estimate of the library size as extraneous factors like the efficiency of the ligation process as well as of the competent cells will affect the size estimate.

Analysis of the Progeny β-Lactamase

The protein sequence from the three evolved genes is shown in Figure 1A. Indeed, all three genes, *BlaSC3-5*, were dicodon incorporates (in red), sharing little sequence similarity between them or for that matter with any of the proteins in the database. The longest gene among the three, contained in *pSC3*, yields a protein that is less than half the size of the parent TEM-1 but with a similar GRAVY value (Figure 1A). It is worth noting that the evolved proteins from previous directed-evolution techniques are always of the same size as their parent and differ only in their primary sequences [5]. Growing cultures of *E. coli* strains harboring *pSC3* were found to be consistently more resistant than the other two TEM-1 derivative genes, and displayed an MIC value double that of the TEM-1 gene-containing strains (Table 2). The BlaSC3 protein was also found to be more soluble than the other BlaSC proteins and could routinely be purified for assaying (see below). As is clear from the sequence of BlaSC3, there is an overwhelming deviation from the wild-type TEM-1 sequence with no hint as to the nature of the new active site residues, apart from S^{45} and K^{48}. We confirmed the active role of these 2 residues during catalysis by BlaSC3 using site-directed mutagenesis (see Experimental Procedures). The S45A and K48A BlaSC3 mutants showed no catalytic turnover (Table 3). While the entire nature of the BlaSC3 active site can only be made clear through a more detailed structural analysis, it is evident that altogether new active site structural motifs have taken the place of their counterparts that were present in the TEM-1 enzyme (like the ω loop). Furthermore, primary sequence analysis of BlaSC3 and its comparison with wild-type TEM-1 show that although BlaSC3 is 79% different from TEM-1, it is strikingly similar in the overall attributes that pertain to hydrophobicity, charge, and secondary structure propensities (Figure 3). This finding does indeed point tantalizingly to the suggestion that the natural course for selecting a progeny from the parent during protein evolution is in fact an emphasis on the overall "functional" similarity between the two proteins even though the progeny is different from the parent by as much as 79% (or for that matter 94% as in the case of the second generation mutant described later). In other words, the dicodon-led protein evolution mimics natural evolution that generates any given protein family. A relevant example of this is the protein family of β-lactamases itself, where we find tens of proteins among the four different classes of lactamases, some differing dramatically in primary protein sequence as well as size, yet all acting as β-lactam hydrolyzing enzymes. The profiling of obtained progeny sequences in accordance with the elegant hypothesis of Hecht and coworkers [23, 24], whereby particular secondary structures can be predicted to occur on the basis of patterning of polar and nonpolar amino acids in the primary sequence, also does not conclusively point toward any particular structural repeat. Encouragingly, comparison of progeny profiles with that of an insoluble amyloid-forming β sheet repeat structure (no. 7, Figure 1C) [24] indicates that such structural elements may not be present in the obtained sequences. One reason why such repeats are not obtained may be because of the patterning of the 14 dicodons itself (Figure 1C)—not all dicodons are patterned as polar-nonpolar pairs. Indeed, it may be that this facet is in part responsible for the progenies being water soluble.

As a last check, we truncated the wild-type TEM-1 enzyme, to bring it to a length similar to that of BlaSC3 (see Supplemental Data at http://www.chembiol.com/cgi/content/full/10/10/917/DC1). The truncated protein BlaSC10 was found to be catalytically inactive, and *E. coli* cultures bearing the plasmid *pSC10* were not able to grow on ampicillin plates. This shows that the activity of the evolved β-lactamase is because of the nature of the primary sequence of the protein and *not* its length.

Expression and Purification of the Evolved β-Lactamase

As a result of its construction design, the *BlaSC3* mutant is directly under the control of a T7 promoter. Under induction with IPTG in growing cultures of the T7 RNA polymerase-carrying *E. coli* BL21(DE3) strain, we found most of the formed protein to be in the insoluble fraction. Indeed, pervious reports concerning the purification of β-lactamase point to its strong promoter-led overexpression resulting preferentially in the formation of inclu-

Table 3. Kinetic Parameters of TEM-1 and Other β-Lactamases

Enzyme	AMP			NIT			CTX		
	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\mu M^{-1}s^{-1}$)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\mu M^{-1}s^{-1}$)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\mu M^{-1}s^{-1}$)
TEM-1	48	892	18.4	52	575	10.9	1316	2.9	0.002
BlaSC3	23	1011	43.5	29	390	13.1	ND	ND	ND
BlaSC7	76	400	5.24	109	439	4.0	467	6.2	0.013
$TEM\text{-}1^{S\rightarrow A}$	—[a]	—		—	—				
$BlaSC3^{S\rightarrow A}$	—	—		—	—				
$BlaSC3^{K\rightarrow A}$	—	—		—	—				

AMP, ampicillin; NIT, nitrocefin; CTX, cefotaxime; ND, not determinable ($K_m > 1600$ μM).
[a] No measurable activity.

sion bodies. The β-lactamases are therefore commonly purified through uninduced expression of the gene, from the resultant periplasmic fractions. We therefore expressed the *BlaSC3* gene in uninduced growing cultures of *E. coli* DH5α and isolated soluble BlaSC3 as C-terminally His-tagged protein (MW 14.2 kDa) using Ni-NTA chromatography. We found, for reasons as yet unexplained, the expression levels of BlaSC3 to be around 30–40 times less than TEM-1 levels under similar conditions. The isolated protein was adjudged to be >90% pure by densitometric analysis using SDS-PAGE. The identity of this protein was also confirmed by Western blot using anti-His×6 antibodies (Figure 4; see Supplemental Data at http://www.chembiol.com/cgi/content/full/10/10/917/DC1). Purified periplasmic BlaSC3 was assayed for its activity using ampicillin, nitrocefin, and cefotaxime as substrates. The k_{cat}/K_m values were in the micromolar range, similar to those obtained for the parent TEM-1 enzyme in our hands, as well as those previously reported for class A lactamases (Table 3) [25].

Further Codon Shuffling of a Progeny β-Lactamase

We then decided to further minimize the presence of TEM-1 structural remnants in BlaSC3. We shifted the dicodon insertion point to the N terminus of the S^{45} region while additionally reducing the size of the region to just 13 residues. This time the codon shuffling resulted in the generation of three new genes with their predicted proteins, BlaSC7-9, showing little sequence similarity between them (Figure 1A). However, the primary sequence analysis of one of the second generation mutants, BlaSC7, suggests good similarity at the "sequence attribute" level with the wild-type TEM-1 as well as the first generation mutant BlaSC3 (Figure 3). BlaSC7 is the smallest (91 aa) among the three proteins and it is 94% different from TEM-1, yet it shares a similar hydrophobicity and charge profile. In retrospect, these similarities are quite possibly responsible for the export of these proteins into the periplasmic space as well as their presence in the periplasm as soluble proteins. The strains harboring the shuffled genes displayed similar albeit lower MIC values compared with the TEM-1 or the earlier BlaSC series, with one notable exception. We found that the strain containing the gene for the BlaSC7 protein showed an MIC value for cefotaxime that was more than a fold higher than the TEM-1 enzyme itself (Table 2). Cefotaxime is a third generation β-lactam that displays high potency against the class A β-lactamases (MIC of around 0.02 μg/ml), but is more readily hy-

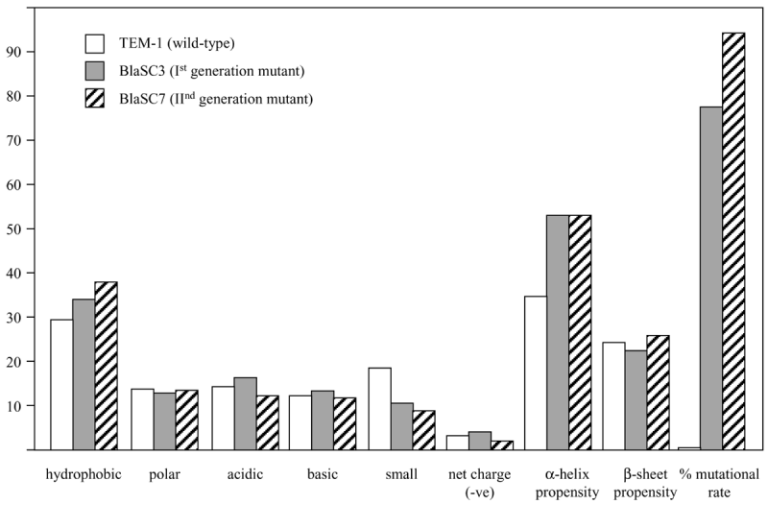

Figure 3. Primary Sequence Analysis of Evolved β-Lactamases Using the SAPS Program

Comparison of the two representative first and second generation evolved mutants (BlaSC3 and BlaSC7, respectively) with the wild-type TEM-1 enzyme is shown. The *y* axis represents the percentage amino acids of the total number of amino acids in the proteins. Hydrophobic: defined in SAPS (Statistical Analysis of Protein Sequences) [33] as the L, V, I, F, M set; polar: S, T, C; acidic: D, E; basic: K, R; small: A, G; net charge: DE − KR; α helix and β sheet propensity: % total of amino acids values defined in the modified version of the Chou-Fasman scale (see reference 4 in the Supplemental Data available at http://www.chembiol.com/cgi/content/full/10/10/917/DC1) that have a tendency to partake in the formation of α helices or β sheets; % mutational rate: defined as the percentage of primary sequence that is different from the wild-type TEM-1.

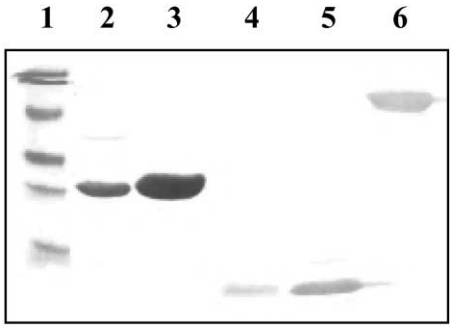

Figure 4. Western Blot Analysis of Purified β-Lactamase Proteins with Anti-His Mab

Lane 1: BioRad prestained MW marker, comprised of 113, 92, 52.9, 35.4, 30 and 21.5 kDa; lane 2: Signal-less TEM-1 (cytoplasmic, predicted MW of 30.52 kDa); lane 3: mature TEM-1 (periplasmic, predicted MW of 30.24 kDa); lane 4: Signal-less BlaSC3 (cytoplasmic, predicted MW of 14.51 kDa); lane 5: mature BlaSC3 (periplasmic, predicted MW of 14.24 kDa); lane 6: MBP-BlaSC3 fusion (cytoplasmic, predicted MW of 58.21 kDa).

drolyzed by the class C enzymes [25]. In fact, in one of the earliest laboratory evolution experiments, the resistance to cefotaxime by TEM-1 was shown to be increased by almost 32,000-fold through gene shuffling [5]. The evolution of proteins such as BlaSC7 clearly shows that "codon-shuffled" protein variants are indeed capable of counteracting a historically overwhelming stress, like the TEM-1 sensitivity to cefotaxime, to a much greater extent than the parent. BlaSC7 was purified as a mature protein from the periplasm and assayed using different β-lactams as substrates. The results corroborated well with the MIC values, especially as we found a lower K_m value for cefotaxime as compared with the wild-type TEM-1 enzyme (Table 3).

Expression of MBP-Fusion Proteins of Evolved Lactamases

In order to obtain greater amounts of soluble BlaSC3 and BlaSC7 proteins, we fused these mutant genes with *malE* gene that codes for the maltose binding protein. Previous work has indicated that N-terminal MBP fusion results in increased amounts of soluble proteins that were previously difficult to obtain as single entities [26]. In the first instance, the fusion was made with the *malE* gene devoid of its 25 amino acid long periplasmic secretion signal so that the fusion protein would exclusively be present in the cytoplasmic fraction. We were able to obtain sufficiently large amounts of the fusion protein in soluble form (see Supplemental Figure S2 at http://www.chembiol.com/cgi/content/full/10/10/917/DC1). Unfortunately, upon cleavage from MBP, we found BlaSC3 to be preferentially in the insoluble fractions. However, treatment with urea, followed by refolding of the protein resulted in BlaSC3 regaining its catalytic activity (see Supplemental Data at *Chemistry & Biology's* website). This was also found to be the case for signal-less TEM-1 and BlaSC3 proteins that were found to be insoluble when isolated from the cytoplasm but regained hydrolytic activity upon urea treatment and refolding (see Supplemental Data at *Chemistry & Biology's* website). Indeed, the regaining of activity upon refolding of TEM-1 has previously been reported [27]. We then fused the mutant as well as the wild-type TEM-1 genes with the periplasmic signal-containing *malE* gene. All three fusion proteins are overexpressed in the cell and are generously visible at their predicted lengths on SDS-PAGE (see Supplemental Figure S3 at *Chemistry & Biology's* website). Preliminary results indicate that the MBP fusions are found predominantly in the soluble fraction. In addition to removing any doubts regarding the absence of a contaminating lactamase, this finding also means that progenies can now be constructed on inactive "MBP-parent protein" scaffolds, thus yielding greater amounts of soluble protein, a prerequisite for conducting 3D structural studies. Such studies are currently underway in our laboratory.

Comparison with Other Notable Directed Evolution Methods

Cassette Mutagenesis

One of the earliest examples of the utility of cassette mutagenesis involved the "reformatting" of the 17 amino acid long active site region of TEM-1 using a 51 bp degenerate oligonucleotide [12]. From an initial library of 10^5, 2000 cfu were able to confer moderate to high ampicillin resistance. Out of the 50 clones that were sequenced, it was found that any given clone contained not more than three amino acid changes (mutation rate around 10%–15% over the 17 residue stretch). The comparison can now be made on many counts. First, while the library size is two orders of magnitude larger than what is obtained from our method, the "whole-gene" mutation rate is 1% compared to values of up to 94% obtained from our method. Second, the lengths of the progenies are all identical to the parent, as this parameter is dictated by the length of the degenerate oligonucleotide itself. Here, there is a striking difference between the two methods—the lengths of the active lactamases obtained through our method range from being 40% to just 25% of the total TEM-1 length.

Gene Shuffling

The very first application of gene shuffling was to the TEM-1 lactamase [5]. Although the initial library size was not determined, it is likely that the number would have been as high as the transformation efficiency of the host, i.e., around 10^8. This number is much higher than what is obtained using our method. There is also a clear advantage when using gene shuffling, in that all areas of the parent gene are prone to mutagenesis, unlike in our method or in cassette mutagenesis. However, the mutation rate per cycle was in the range of 1%, albeit variant methods like family shuffling, ITCHY, etc. yield much higher mutagenesis rates. Finally, the progeny length is always identical to that of the parent, unlike in our method. As of the present, we believe that gene shuffling, in its simplicity and control, remains as the method of choice for a truly combinatorial approach toward improvement of any given protein.

Methods for De Novo Protein Synthesis

Perhaps a more pertinent comparison would be with regard to the prospect of generating totally de novo proteins or chimeras wherein large regions of the se-

quence are created de novo. Here, attention is drawn toward the seminal works of Szostak and coworkers [14], Riechmann and Winter [28], and finally of Hecht and coworkers [23]. In a pioneering effort, Szostak and coworkers created totally de novo proteins that were able to counter stress, which in their case amounted to the binding of ATP moiety by the proteins [14]. The successful proteins were obtained from a maximum possible set of 10^{12} sequences, which in their estimate is tantamount to obtaining a fully folded functional protein from a collection of 10^{11} totally random sequences. In another significant study, Reichmann and Winter demonstrated the possibility of obtaining protein chimeras, starting from an inactive protein scaffold, in their case a cold shock protein, cspA [28]. The scaffold was made active by "filling up" the missing regions of the protein sequence through the use of a library of small segments comprising the total *E. coli* genome. From a library of 10^8 sequences, as many as 600 chimeras were able to counter the provided stress, given that nearly half of each progeny sequence was always that of the wild-type cspA. We on the other hand have obtained functional proteins from a library size that is several degrees of magnitude lower than the starting libraries in the above-mentioned experiments. Why then are functional proteins still able to emerge? We believe this may be in large because of the following two reasons. First, although the chimeras BlaSC3-7 differ greatly from their TEM-1 parent in the primary sequence, all of them without exception possess the two most critical residues required for β-lactamase activity, namely the active site serine and lysine. In the context of a directed evolution experiment, this may be construed as the inactive scaffold being "sufficiently close" to becoming an active protein. Thus, all that was required was a "little push" in the right direction that came from the dicodon-assembled amino acid addenda. It therefore remains to be seen whether our approach can be extended to generate totally de novo functional proteins. Second, for such amino acid attachments to result in functional proteins, given that the library size is not very large, it perhaps is helpful if the selected dicodons themselves have an inherent preference for forming favorable structures. As has been mentioned earlier, the 14 dicodons do not all conform to one single identical pattern in terms of their polar-nonpolar nature. Therefore, it may very well be that the obtained amino acid attachments are from the beginning itself able to form foldable structures, thereby reducing the need for a large primary library. This is presently at best a conjecture that nonetheless can be tested by employing a range of different dicodon sets. For such future experiments, the hypothesis of Hecht and coworkers mentioned earlier would be particularly helpful for dicodon design.

Significance

We have reported here a novel and simple method for protein evolution capable of generating variants that differ vastly from their parent protein. We have also shown that the variance can be applied at multiple regions on the parent sequence depending upon the position of the dicodon insertion point. The method can easily be programmed to skew the type of amino acids one wishes to be present in the evolved proteins, depending upon the nature, amount, and number of dicodons used. Ours is a method that yields proteins of vastly varying lengths and primary sequence. In many cases, therefore, the "mutation rate" is as high as 94%. Indeed, this variance in progeny length makes for an attractive prospect for generating "minimized" enzymes that contain a bare minimum catalytic center, yet are able to perform as well or even better than their large parents. Our method can also be utilized in the context of elegant, wholly in vitro systems of Szostak and coworkers, thereby further extending the amount of degeneracy accessible at the DNA level. There is also the attractive possibility of applying our method for the construction of combinatorial protein libraries, based on the binary patterning developed by Hecht and coworkers [23]. For example, alternating patterns of polar (○)-nonpolar (●) amino acid sequences (e.g., ○●○●○●○●) that give rise to β sheet repeats can be obtained by ligating only those dicodons that are themselves polar-nonpolar pairs (e.g., DI, KL, EL, NV, etc). On the other hand, patterns leading to α helices would have to obtained by assembly of modules like ○○●●, so that the assembled sequence corresponds to a nonpolar residue every three or four residues (e.g., ○○●●○○●●○○●●). Success in such a strategy promises exciting implications for de novo protein design.

Experimental Procedures

All antibiotics, with the exception of Nitrocefin (Merck) were purchased from Sigma. Ni-NTA resin and Anti-His monoclonal antibody were purchased from Qiagen. PGEMT-Easy cloning kit was purchased from Promega. T4 DNA Ligase was purchased from New England Biolabs. *E. coli* BL21(DE3) strain, *pET21c(+)* and *pET28a(+)* expression vectors were from Novagen. QuickChange-XL site-directed mutagenesis kit was purchased from Stratagene.

DNA Cloning

Routine cloning and transformation procedures for *E. coli* were as described earlier [29]. All new plasmids generated during this work were sequenced using the di-deoxy method in order to validate their authenticity.

Construction of Plasmid pSC1

The TEM-1 β-lactamase gene was PCR amplified (25 cycles) from plasmid pET21c using *taq* DNA polymerase and the oligonucleotides 5′-AACCATGGGTATTCAACATTTCCGTGTCGCCCTT-3′ and 5′-AACTCGAGTACGTACCAATGCTTAATCAGTGAGGC-3′ as forward and reverse primers respectively. The resulting 875 bp long PCR product was cloned in *pGEMT-Easy* vector. The *TEM-1* gene was excised using NcoI and XhoI restriction enzymes and cloned in *pET28a* vector previously cut with the same two enzymes. The resulting expression plasmid was designated *pSC1*.

Construction of Plasmid pSC2

Using plasmid *pSC1* as template, a fragment from *TEM-1* gene was amplified using the oligonucleotides 5′-AAGCATGCAAGGAGATGGCGCCCAACAGTCCC-3′ and 5′-AATACGTAGCCACATAGCAGAACTTTAAAAGT-3′. The resulting 537 bp long PCR product was cloned in *pGEMT-Easy* vector. The desired fragment of the *TEM-1* gene was excised using SphI and SnaBI and cloned in *pSC1* previously cut with the same two enzymes. The resulting expression plasmid was designated *pSC2*. This vector served as the inactive *TEM-1* scaffold for the first-generation dicodon experiments.

Construction of Libraries

100 ng of each of the 14 5′-phosphorylated DNA hexamers (referred to as dicodons or DC in text) in a 20 μl reaction mixture containing 2 μl of 10× ligase buffer and 14 μl of double-distilled water was gently heated to 55°C. The temperature was then slowly brought down to 4°C. 14 μl of each DC was then mixed and distributed equally into seven tubes. To each of these tubes were added 2 μl of 10× ligase buffer and T4 DNA ligase respectively and the tubes incubated at 4°C. This was considered as time t = 0 hr. 2 μl (~250 ng) of SnaBI-cut and dephosphorylated plasmid *pSC2* was added to each tube after defined time intervals of 0.15, 0.5, 1, 4, 8, 16, and 24 hr. All tubes were then incubated at 4°C for a further 8 hr. The contents of each tube were then independently used to transform *E. coli* DH5α-competent cells. The transformation efficiency of these competent cells had earlier been standardized to 2×10^8 cfu per 100 μl of cells per μg of DNA. The transformation mixture was plated on LB agar plates containing 50 μg/ml kanamycin, 0.1 mM IPTG (see reference 1 in the Supplemental Data available at http://www.chembiol.com/cgi/content/full/10/10/917/DC1), and increasing amounts of ampicillin (50–20,000 μg/ml). Initially 3 cfu, exhibiting growth at ampicillin concentrations of 20,000; 5,000; and 2,000 μg/ml respectively were isolated, plated again on ampicillin/kanamycin plate, and plasmid DNA recovered from them. The plasmids were designated as *pSC3′*, *pSC4′*, and *pSC5′*. Using these plasmids as DNA template, the three genes of unknown sequence were independently PCR amplified by employing the universal forward and reverse primers 5′-AATTAATACGACTCACTATAGGGGAATTGT-3′ and 5′-AGCAGCCAACTCAGCTTCCTTTCGGGCTTT-3′ and the PCR products cloned in *pGEMT-Easy* vector. The complete genes were then excised using the restriction enzymes NcoI and XhoI and cloned in *pET28a* expression vector previously cut with the same two enzymes. The newly derived plasmids were sequenced using the T7 forward and reverse primers and were designated *pSC3*, *pSC4*, and *pSC5*. The sequence of the genes harbored in these plasmids—designated *BlaSC3*, *BlaSC4*, and *BlaSC5*—was found to be identical to the gene sequence contained in plasmids *pSC3′*, *pSC4′*, and *pSC5′*. As a final check, we digested plasmids *pSC3-5* with restriction enzymes in order to destroy the newly formed β-lactamase genes. The digested plasmids were self-ligated and used to transform DH5α. The transformants were able to grow on kanamycin but not on ampicillin.

Construction of Plasmid pSC6

Using plasmid *pSC1* as template, a fragment from TEM-1 gene was amplified using the oligonucleotides 5′-AAGCATGCAAGGAGATGG CGCCCAACAGTCCC-3′and 5′-TTGATATCACTACGTATGGGTGAG CAAAAACAGGAAG-3′. The resulting 392 bp long PCR product was cloned in *pGEMT-Easy* vector and the new vector designated *pSC6a*. A 937 bp XmnI fragment from plasmid *pSC3* was ligated with *pSC6a* previously cut with EcoRV, and the new plasmid was designated as *pSC6b*. The desired fragment of the altered *BlaSC3* gene was excised using NcoI and XhoI and cloned in *pET28a* previously cut with the same two enzymes. The resulting expression plasmid was designated *pSC6*. This vector served as the inactive *BlaSC3* scaffold for the second generation dicodon experiments.

Construction of Plasmids pSC7, pSC8, and pSC9

The generation of these plasmids was carried using the DC-ligation method described previously except that 2 μl (~250 ng) of SnaBI-cut and dephosphorylated plasmid *pSC6* was added to each tube containing the DC-ligated products. The *E. coli* DH5α transformants were plated on increasing amounts of cefotaxime, in addition to their selection on ampicillin and kanamycin. Following on, the three plasmids, *pSC7*, *pSC8*, and *pSC9* were generated and sequenced. The genes harbored in these plasmids were designated as *BlaSC7*, *BlaSC8*, and *BlaSC9*.

Site-Directed Mutagenesis

The active site serine and lysine residues of TEM-1, BlaSC3, and BlaSC7 were changed to alanine using the primer pairs 5′-GTTTTCC AATGATGGCGACTTTTAAAGTTCTGC-3′ and 5′-GCAGAACTTTAA AAGTCGCCATCATTGGAAAAC-3′, and 5′-ATGATGAGCACTTTTGC GGTTCTGCTATGTGGC-3′ and 5′-GCCACATAGCAGAACCGCAAA AGTGCTCATCAT-3′, respectively. The altered genes were sequenced and the changes validated.

Determination of MIC Values

MICs were determined by microdilution method as well as by plating cells on LB agar plates containing varying amounts of antibiotic. In the former method, a bacterial inoculum of ~10^5 cfu per tube was grown in the presence of 2-fold increases of each antibiotic used. The results were interpreted according to the guidelines provided [30]. In the latter method, procedures were followed as published previously, especially for determining the MIC for cefotaxime [5]. The plates or growing cultures were examined after 18 hr incubation at 37°C. The lowest concentration of antibiotic that inhibited growth was adjudged as the MIC.

Purification of TEM-1, BlaSC3, and BlaSC7 Proteins

6 liters of LB media containing 200 μg/ml of ampicillin and 50 μg/ml of kanamycin were inoculated with overnight cultures of DH5α strains harboring the plasmids *pSC1*, *pSC3*, and *pSC7*. All cultures were grown for 16 hr at 37°C under shaking. Cells were harvested by centrifugation, washed twice, and resuspended in buffer A (30 mM Tris-HCl, pH 8.0, 1 mM EDTA, 20% w/v sucrose) and purified using the osmotic-shock procedure [31]. The supernatant was collected and applied onto a 5 ml Ni^{+2}-NTA column that had been equilibrated with buffer B (30 mM Tris-HCl, pH 8.0, 1 mM EDTA, 150 mM NaCl). The column was extensively washed with buffer B and eluted with two bed volumes of 250 mM imidazole in buffer B. Fractions displaying β-lactamase activity against nitrocefin were collected and pooled. The presence of β-lactamases was also detected by 15% SDS-PAGE. The Ni^{+2}-NTA-purified material was concentrated using YM-5 concentrators (Vivascience, UK). The purified proteins were pooled and dialyzed against buffer C (50 mM phosphate, pH 7.0). All proteins were stored at 4°C until further use.

Determination of Kinetic Parameters

Kinetic measurements were performed on a Shimatzu UV-1601 double-beam spectrophotometer using a 1.0 cm path-length cuvette at room temperature in 50 mM phosphate buffer (pH 7.0) using enzyme concentrations of either 1 or 5 nM. The rate constants were obtained for the following antibiotics: ampicillin (232 nm, $\Delta\epsilon = 820\ M^{-1}\ cm^{-1}$), nitrocefin (486 nm, $\Delta\epsilon = -20{,}500\ M^{-1}\ cm^{-1}$), and cefotaxime (264 nm, $\Delta\epsilon = 6{,}900\ M^{-1}\ cm^{-1}$). All experiments were recorded in triplicate using six to seven substrate concentrations bracketing the K_m. Kinetic parameters were determined by fitting obtained values to the differential form of the Michaelis-Menten equation using unweighted nonlinear least squares. The data was analyzed by Graphpad-PRISM and by EZ-FIT software programs.

Acknowledgments

We thank Dr. Marcel Luber and his team at Microsynth, Switzerland for DNA sequencing and oligonucleotide synthesis and Dr. S. Pascal, University of Rochester, Rochester, NY for the kind gift of MBP plasmids HMBP-3C and HMBP-3C-Pro. The help and support of RGP Group members during the course of this work is gratefully acknowledged. S.C. would like to thank UGC for financial support.

Received: June 26, 2003
Revised: July 29, 2003
Accepted: July 30, 2003
Published: October 17, 2003

References

1. Darwin, C. (1986). The Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life, 1986 Edition (New York: New American Library).
2. Farinas, E.T., Bulter, T., and Arnold, F.H. (2001). Directed enzyme evolution. Curr. Opin. Biotechnol. *12*, 545–551.
3. Tao, H., and Cornish, V.W. (2002). Milestones in directed enzyme evolution. Curr. Opin. Chem. Biol. *6*, 858–864.
4. Powell, K.A., Ramer, S.W., Del Cardayre, S.B., Stemmer, W.P., Tobin, M.B., Longchamp, P.F., and Huisman, G.W. (2001). Di-

rected evolution and biocatalysis. Angew. Chem. Int. Ed. Engl. *40*, 3948–3959.
5. Stemmer, W.P. (1994). Rapid evolution of a protein in vitro by DNA shuffling. Nature *370*, 389–391.
6. Crameri, A., Raillard, S.A., Bermudez, E., and Stemmer, W.P. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature *391*, 288–291.
7. Ostermeier, M., Shim, J.H., and Benkovic, S.J. (1999). A combinatorial approach to hybrid enzymes independent of DNA homology. Nat. Biotechnol. *17*, 1205–1209.
8. Zhao, H., Giver, L., Shao, Z., Affholter, J.A., and Arnold, F.H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat. Biotechnol. *16*, 258–261.
9. Hanes, J., and Pluckthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. Proc. Natl. Acad. Sci. USA *94*, 4937–4942.
10. Reidhaar-Olson, J.F., and Sauer, R.T. (1988). Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. Science *241*, 53–57.
11. Munir, K.M., French, D.C., and Loeb, L.A. (1993). Thymidine kinase mutants obtained by random sequence selection. Proc. Natl. Acad. Sci. USA *90*, 4012–4016.
12. Oliphant, A.R., and Struhl, K. (1989). An efficient method for generating proteins with altered enzymatic properties: application to beta-lactamase. Proc. Natl. Acad. Sci. USA *9860*, 9094–9098.
13. Stemmer, W.P. (1994). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. USA *91*, 10747–10751.
14. Keefe, A.D., and Szostak, J.W. (2001). Functional proteins from a random-sequence library. Nature *410*, 715–718.
15. Cho, G., Keefe, A.D., Liu, R., Wilson, D.S., and Szostak, J.W. (2000). Constructing high complexity synthetic libraries of long ORFs using in vitro selection. J. Mol. Biol. *297*, 309–319.
16. Gupta, N.K., Ohtsuka, E., Weber, H., Chang, S.H., and Khorana, H.G. (1968). Studies on polynucleotides. LXXXVII. The joining of short deoxyribopolynucleotides by DNA-joining enzymes. Proc. Natl. Acad. Sci. USA *60*, 285–292.
17. Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105–132.
18. Massova, I., and Mobashery, S. (1998). Kinship and diversification of bacterial penicillin-binding proteins and beta-lactamases. Antimicrob. Agents Chemother. *42*, 1–17.
19. Strynadka, N.C., Adachi, H., Jensen, S.E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K., and James, M.N. (1992). Molecular structure of the acyl-enzyme intermediate in beta-lactam hydrolysis at 1.7 Å resolution. Nature *359*, 700–705.
20. Huang, W., Petrosino, J., Hirsch, M., Shenkin, P.S., and Palzkill, T. (1996). Amino acid sequence determinants of beta-lactamase structure and activity. J. Mol. Biol. *258*, 688–703.
21. Petrosino, J.F., and Palzkill, T. (1996). Systematic mutagenesis of the active site omega loop of TEM-1 beta-lactamase. J. Bacteriol. *178*, 1821–1828.
22. Lietz, E.J., Truher, H., Kahn, D., Hokenson, M.J., and Fink, A.L. (2000). Lysine-73 is involved in the acylation and deacylation of beta-lactamase. Biochemistry *39*, 4971–4981.
23. Kamtekar, S., Schiffer, J.M., Xiong, H., Babik, J.M., and Hecht, M.H. (1993). Protein design by binary patterning of polar and nonpolar amino acids. Science *262*, 1680–1685.
24. West, M.W., Wang, W., Patterson, J., Mancias, J.D., Beasley, J.R., and Hecht, M.H. (1999). De novo amyloid proteins from designed combinatorial libraries. Proc. Natl. Acad. Sci. USA *96*, 11211–11216.
25. Bush, K., Jacoby, G.A., and Medeiros, A.A. (1995). A functional classification scheme for beta-lactamases and its correlation with molecular structure. Antimicrob. Agents Chemother. *39*, 1211–1233.
26. Kapust, R.B., and Waugh, D.S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. *8*, 1668–1674.
27. Bowden, G.A., and Georgiou, G. (1990). Folding and aggregation of beta-lactamase in the periplasmic space of Escherichia coli. J. Biol. Chem. *265*, 16760–16766.
28. Riechmann, L., and Winter, G. (2000). Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. Proc. Natl. Acad. Sci. USA *97*, 10068–10073.
29. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).
30. National Committee for Clinical Laboratory Standards. (1997). Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria that Grow Aerobically, Fourth Edition: Approved Standard M7–A4. (Wayne, PA: National Committee for Clinical Laboratory Standards).
31. Neu, H.C., and Heppel, L.A. (1965). The release of enzymes from Escherichia coli by osmotic shock and during the formation of spheroplasts. J. Biol. Chem. *240*, 3685–3692.
32. Jelsch, C., Mourey, L., Masson, J.M., and Samama, J.P. (1993). Crystal structure of Escherichia coli TEM1 beta-lactamase at 1.8 Å resolution. Proteins *16*, 364–383.
33. Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B.E., and Karlin, S. (1992). Methods and algorithms for statistical analysis of protein sequences. Proc. Natl. Acad. Sci. USA *89*, 2002–2006.