Published online 25 September 2009 Nucleic Acids Research, 2009, Vol. 37, No. 18 **e125**
doi:10.1093/nar/gkp602

TrimerDimer: an oligonucleotide-based saturation mutagenesis approach that removes redundant and stop codons

Paul Gaytán*, Casandra Contreras-Zambrano, Mónica Ortiz-Alvarado, Alfredo Morales-Pablos and Jorge Yáñez

Instituto de Biotecnología-Universidad Nacional Autónoma de México, Ap. Postal 510-3 Cuernavaca, Morelos 62250, México

Received January 6, 2009; Revised June 15, 2009; Accepted June 30, 2009

ABSTRACT

9-fluorenylmethoxycarbonyl (Fmoc) and 4,4′-dimethoxytrityl (DMTr) are orthogonal hydroxyl protecting groups that have been used in conjunction to assemble oligonucleotide libraries whose variants contain wild-type and mutant codons randomly interspersed throughout a focused DNA region. Fmoc is labile to organic bases and stable to weak acids, whereas DMTr behaves oppositely. Based on these chemical characteristics, we have now devised TrimerDimer, a novel codon-based saturation mutagenesis approach that removes redundant and stop codons during the assembly of degenerate oligonucleotides. In this approach, five DMTr-protected trinucleotide phosphoramidites (dTGG, dATG, dTTT, dTAT and dTGC) and five Fmoc-protected dinucleotide phosphoramidites (dAA, dTT, dAT, dGC and dCG) react simultaneously with a starting oligonucleotide growing on a solid support. The Fmoc group is then removed and the incorporated dimers react with a mixture of three DMTr-protected monomer phosphoramidites (dC, dA and dG) to produce 15 trinucleotides: dCAA, dAAA, dGAA, dCTT, dATT, dGTT, dCAT, dAAT, dGAT, dCGC, dAGC, dGGC, dCCG, dACG and dGCG. After one mutagenic cycle, 20 codons are generated encoding the 20 natural amino acids. TrimerDimer was tested by randomizing the four contiguous codons that encode amino acids L64–G67 of an engineered, nonfluorescent GFP protein. Sequencing of 89 nonfluorescent mutant clones and isolation of two fluorescent mutants confirmed the principle.

INTRODUCTION

Several strategies have been developed to accelerate artificial modification of proteins, either by enzymatic mutagenesis of the encoding genes or by amplification from synthetic oligonucleotides (1). Enzymatic methods are mainly based on error-prone PCR (2) producing quasi-random single base changes throughout a gene sequence. A similar effect directed at a focused region is done with libraries of spiked oligos, where each wild-type nucleotide is substoichiometrically replaced with the other three bases during the chemical assembly (3,4). Unfortunately, a large number of amino acid (aa) replacements do not occur in these approaches, such as those requiring 2–3-bp changes per codon (5), and some proteins do require this kind of non-natural aa substitutions to change their properties and improve their function (6–8).

Covering of the whole sequence space in a focused protein region has been achieved by some methods of saturation mutagenesis that produce pools of degenerate oligos. NNG/C and NNG/T approaches are widely used and commercially available options that generate mixtures of 31 sense codons and one stop codon, encoding the 20 natural aa's at varied frequency. In consequence, protein libraries created with ordinary degenerate oligos are highly biased toward those aa sequences encoded by redundant codons (9). To avoid this drawback, Hughes *et al.* (9) described MAX randomization, a mutagenic approach that only requires mixtures of conventional oligos to remove redundant and stop codons via hybridization selection with an oligo template that is completely randomized. The process seems interesting and easy to implement, but is limited to alternate codons or only two contiguous codons in the template and, disappointingly, the frequency of mutated codons was still very biased.

Another approach for the elimination of codon redundancy and stop codons uses chemical incorporation of

*To whom correspondence should be addressed. Tel: 52 777 3291604; Fax: 52 777 3172388; Email: paul@ibt.unam.mx

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2009. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

DMTr-protected trinucleotide phosphoramidites (DMTr stands for the 4,4′-dimethoxytrityl protecting group) into growing oligos in traditional solid-phase synthesis. These reagents may encode all 20 aa's (10,11) or a subset of them (12,13). Recently, Krumpe *et al.* (14) demonstrated that this trinucleotide approach significantly enhanced the functional diversity of a 12-mer peptide library when directly compared with the same library assembled via the conventional NNG/T approach.

After a lengthy development undertaken by Alexei Kayushin's group, a set of 20 DMTr-protected trinucleotide phosphoramidites has recently become commercially available from Glen Research Corporation. However, these reagents are very expensive and often cost-prohibitive for developing countries and poorly funded research groups.

Eliminating redundancy and stop codons was also achieved by Neuner *et al.* (15), who developed an oligonucleotide-based approach in which the degeneracy was created by the laborious resin-splitting approach. In this method, groups of five different codons are synthesized in four parallel columns. In total, 11 DMTr-protected dimer phosphoramidites and four monomers were required to complete the set of 20 codons.

Taking advantage of our experience employing two orthogonal 5′-hydroxyl protecting groups simultaneously in several nonsaturating mutagenesis approaches for the synthesis of degenerate oligonucleotides (16–18), we have now developed an approach that removes redundant and stop codons in a scheme of saturation mutagenesis. Contrary to the DMTr-trinucleotide approach, the method presented here only requires five DMTr-protected trinucleotide phosphoramidites, five Fmoc-protected dinucleotide phosphoramidites and a mixture of three ordinary DMTr-protected monomer phosphoramidites to produce a set of 20 codons, each encoding a different aa. Fmoc (which stands for 9-fluorenylmethoxycarbonyl group) is labile to basic solutions and stable to acidic solutions whereas DMTr behaves oppositely. The method was dubbed TrimerDimer, a composite word that highlights the simultaneous use of trimers and dimers for the assembly of degenerate oligos. Contrary to Neuner's method (15), TrimerDimer is suitable for complete automation.

MATERIAL AND METHODS

5′-*O*-(4,4′-*O*-dimethoxytrityl)-*N*-benzoyl-2′-deoxyadenosine-3′-methyl-*N*,*N*-diisopropylamino-phosphoramidite (DMTr-dAp), 5′-*O*-(4,4′-*O*-dimethoxytrityl)-*N*-benzoyl-2′-deoxycytidine-3′-methyl-*N*,*N*-diisopropylamino-phosphoramidite (DMTr-dCp), 5′-*O*-(4,4′-*O*-dimethoxytrityl)-*N*-isobutyryl-2′-deoxyguanosine-3′-methyl-*N*,*N*-diisopropylamino-phosphoramidite (DMTr-dGp), 5′-*O*-(4,4′-*O*-dimethoxytrityl)-2′-deoxythymidine-3′-methyl-*N*,*N*-diisopropylamino-phosphoramidite (DMTr-dTp), 5-ethylthiotetrazole (ETT), diethoxy-*N*,*N*-diisopropylamino-phosphoramidite (DDP) and the support for oligonucleotide synthesis (DMTr-dC-LCAA-CPG, 1000 Å) were purchased from Chemgenes (Wilmington, MA). 5′-*O*-DMTr-protected nucleosides of *N*-benzoyl-2′-deoxycytidine (DMTr-dC^{bz}), *N*-isobutyryl-2′-deoxyguanosine (DMTr-dG^{ibu}) and 2′-deoxythymidine (DMTr-dT) were a gift from Glen Research Corporation (Sterling, VA). Iodine (I_2), pyridine (py), acetic anhydride (Ac_2O), *N*-methylimidazole (NMI), trichloroacetic acid (TCA), ammonium hydroxide (NH_4OH), methanolic ammonia, 9-fluorenylmethoxycarbonyl chloride (Fmoc-Cl), 1,8-diazabicyclo (5.4.0)undec-7-ene (DBU), thiophenol, triethylamine (TEA), *N*,*N*-diisopropylethylamine (DIPEA) and chloro (diisopropylamino)methoxyphosphine (MeOPClN^{i}Pr$_2$) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Dichloromethane (DCM) and tetrahydrofuran (THF) were purchased from Honeywell Burdick & Jackson (Muskegon, MI, USA), while anhydrous acetonitrile was from American Bioanalytical (Natick, MA, USA).

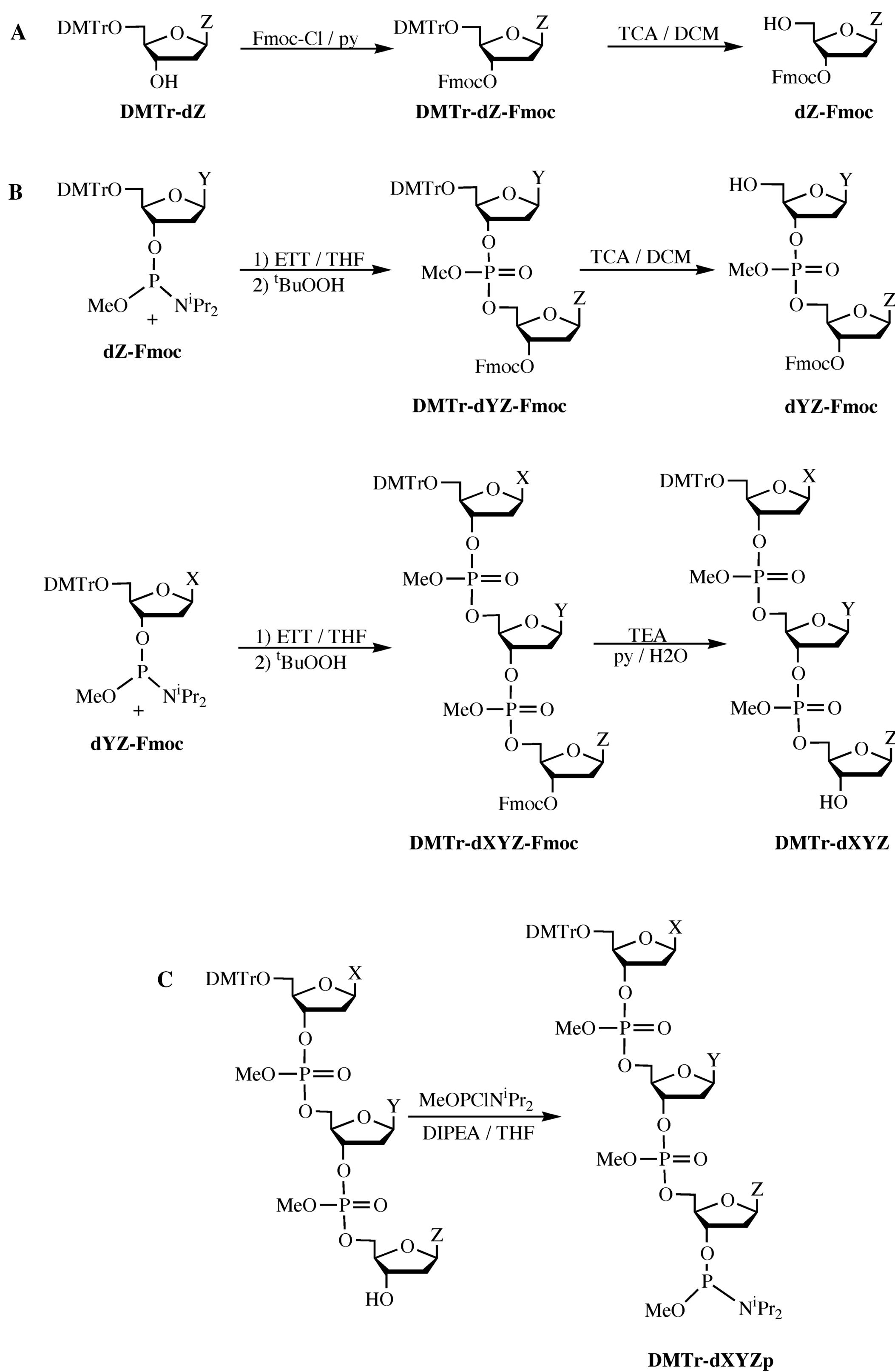
Kanamycin (KAN) was bought from Sigma-Aldrich. Restriction endonucleases, T4 DNA ligase, Vent DNA polymerase, and deoxynucleoside-triphosphates (dNTPs) were purchased from New England Biolabs (Beverli, MA, USA) and used according to standard protocols. *Taq* DNA polymerase was obtained from a batch purified in our laboratory. Purification kits for plasmid and PCR products were purchase from Roche (Mannheim, Germany). Plasmid pT4, which contains the KAN-resistance gene and the ColE1 and f1 origins of DNA replication within 2.0 kb, was produced as described (18). *Escherichia coli* MC1061 strain was used for cloning and plasmid production.

Chemical synthesis

5′-*O*-DMTr-protected trinucleotides (DMTr-dXYZ), precursors of the target DMTr-protected trinucleotide phosphoramidites (DMTr-dXTZp), were synthesized in the 3′→5′ direction using a pathway similar to that reported by Virnekas *et al.* (10) However, the 3′ end of the intermediate compounds were protected with the Fmoc group as shown in Scheme 1, instead of the phenoxyacetyl (PAC) group used by Virnekas *et al.* The procedures for synthesis of DMTr-dTGGp and its precursors are described as a representative example. The other four DMTr-protected trinucleotide phosphoramidites dTTT, dATG, dTAT and dTGC were prepared similarly.

2-N-isobutyryl-2′-deoxyguanosine-3′-YL-9-fluorenylmethoxycarbonate (dG-Fmoc). DMTr-dG^{ibu} (15 mmol, 9.57 g) was co-evaporated twice with 75 ml anhydrous pyridine under high vacuum and re-dissolved in 150 ml of the same solvent. Fmoc-Cl powder (4.66 g, 18 mmol) was added under nitrogen atmosphere and the reaction was magnetically stirred for 2 h at room temperature. The Fmoc-Cl excess was quenched with methanol for 5 min and the reaction was concentrated and co-evaporated twice with 30-ml toluene. The resulting foam was dissolved in 150 ml dichloromethane, cooled to 0°C and treated with 150 ml cold 6% TCA in dichloromethane for 10 min. After this detritylating step, a cold saturated-solution of sodium bicarbonate was added until neutralization. The reaction was washed once more with sodium bicarbonate and twice with brine. The organic phase was

 Nucleic Acids Research, 2009, Vol. 37, No. 18 ***e125***



Scheme 1. Synthetic pathway for the preparation of DMTr-protected trimer phosphoramidites (DMTr-dXYZp). (**A**) Synthesis of 3′ Fmoc-protected deoxynucleosides. (**B**) Synthesis of DMTr-protected trinucleotides (DMTr-dXYZ) by the phophitetriester method in solution-phase. (**C**) Final phosphitylation. X, Y or Z: any of the bases properly protected in the exocyclic amino group. The abbreviations are described in ‘Materials and Methods’ section.

dried over anhydrous sodium sulfate, filtered and concentrated to until dry. The resulting oil was dissolved in 20 ml of dichloromethane and purified by flash column chromatography; the target compound was eluted with 3% methanol in dichloromethane and was precipitated with n-hexane resulting in a white powder (3.13 g, 37% yield).

The other 3′-*O*-Fmoc-protected monomers (dT and dC^{bz}) were synthesized by similar protocol and obtained in yields ranging from 50% to 70% of starting nucleosides.

2-N-isobutyryl-2′-deoxyguanosine-3′-YL-methylphosphate-5′-YL-2-N-isobutyryl-2′-deoxyguanosine-3′-YL-9-fluorenylmethoxycarbonate (dGG-Fmoc). Dry DMTr-dGp (4.8 g, 6 mmol) was dissolved in 20 ml THF and added to a mixture of dG-Fmoc (2.79 g, 5 mmol) and ETT (1.95 g, 15 mmol), which had been previously co-evaporated with anhydrous THF (2 × 20 ml). After 2 h of reaction time, 4 ml 70% tert-butylhydroperoxide were added and stirred for an additional 2 h. The reaction was concentrated under high vacuum, re-dissolved in dichloromethane and washed with sodium bicarbonate and brine. The fully protected dimer was quickly purified by flash column chromatography and subjected to detritylation as described for dG-Fmoc. A second fast purification by flash column chromatography yielded the expected product, dGG-Fmoc, which was eluted with 5% methanol in dichloromethane. After precipitation with n-hexane, the product was obtained as a white powder (3.05 g, 62.2% yield).

5′-O-(4,4′-dimethoxytrityl)-2′-deoxythymidine-3′-YL-methylphosphate-5′-YL-2-N-isobutyryl-2′-deoxyguanosine-3′-YL-methylphosphate-5′-YL-2-N-isobutyryl-2′-deoxyguanosine (DMTr-dTGG). Dry DMTr-dTp (2.03 g, 2.88 mmol) was dissolved in 20 ml THF and added to a mixture of dGG-Fmoc (2.79 g, 2.4 mmol) and ETT (0.93 g, 7.2 mmol), which had been previously co-evaporated with anhydrous THF (2 × 20 ml). After 2 h of reaction, 3 ml 70% tert-butylhydroperoxide were added and the reaction was stirred for an additional 2 h. The reaction was concentrated under high vacuum, re-dissolved in dichloromethane and washed with sodium bicarbonate and brine. The fully protected trimer was quickly purified by flash column chromatography and subjected to removal of the Fmoc group by treatment with 30 ml pyridine:TEA:water 3:1:1 for 15 min.

The reaction mixture was concentrated under high vacuum, re-dissolved in 10 ml dichloromethane and purified by flash column chromatography. The target product was eluted with 6% methanol in dichloromethane and the final product was precipitated with *n*-hexane, and was obtained as a white powder (1.55 g, 47% yield).

Identity analysis of DMTr-protected trinucleotides

Before final phosphitylation, 1–2 mg of each DMTr-protected trinucleotide dTGG, dTTT, dATG, dTAT and dTGC was treated with 1 ml methanolic ammonia for 12 h at 55°C and the supernatant was concentrated until dry in a speedvac. Each residue was re-dissolved in 500 µl water and samples of 25 µl were taken and mixed with the same volume of an identical trinucleotide standard assembled by ordinary solid-phase synthesis in the trityl-on mode. This mixture was analyzed by reverse-phase HPLC on a System Gold chromatograph and an ultrasphere C18 analytical column (4.6 × 250 mm) both from Beckman (Fullerton, CA, USA), with detection at 260 nm. Elution was performed with a linear acetonitrile gradient from 20% to 50% in 0.1 M triethylammonium acetate, pH 7.2, over 20 min at a flow rate of 1 ml/min.

5′-O-(4,4′-dimethoxytrityl)-2′-deoxythymidine-3′-YL-methylphosphate-5′-YL-2-N-isobutyryl-2′-deoxyguanosine-3′-YL-methylphosphate-5′-YL-2-N-isobutyryl-2′-deoxyguanosine-3′-YL-N,N-diisopropylaminomethoxy-phosphoramidite (DMTr-dTGGp). The compound DMTr-dTGG (1.49 g, 1.08 mmol) was co-evaporated twice with 10 ml anhydrous THF and re-dissolved in 20 ml of the same solvent. DIPEA (620 µl, 3.75 mmol) and $MeOPClN^{i}Pr_2$ (520 µl, 2.25 mmol) were added under nitrogen atmosphere and the reaction was magnetically stirred for 2 h. The reaction was then quenched with sodium bicarbonate, diluted with dichloromethane and subsequently washed with sodium bicarbonate and brine. The target compound was purified by flash column chromatography using 20% pyridine in dichloromethane to elute. After precipitation with n-hexane, the compound was obtained as a white powder (800 mg, 47.1% yield) with purity greater than 90% according to HPLC analysis.

Degenerate oligonucleotide synthesis

Two degenerate oligonucleotides with identical sequence were assembled in completely automated mode using the 394 DNA synthesizer from Applied Biosystems (Foster City, CA, USA). The synthesis cycle for column 1, 0.2 µmol scale, is supplied as Supplementary Data.

The synthesizer was loaded with 80 mM acetonitrile solutions of DMTr-dAp, DMTr-dGp, DMTr-dCp and DMTr-dTp in positions 1, 2, 3 and 4, respectively. Position 5 was loaded with 0.15 M DDP in acetonitrile. Position 6 was loaded with a mixture of five DMT-protected trinucleotide phosphoramidites and five Fmoc-protected dinucleotide phosphoramidites in acetonitrile. The individual concentration of each compound for the pair of degenerate oligos is shown (Supplementary Table S1). Position 7 was loaded with an equimolar mixture of the monomer phosphoramidites DMTr-dAp, DMTr-dGp and DMTr-dCp making an 80 mM total solution. Position 10 was loaded with 100 mM DBU in acetonitrile. The other positions of the synthesizer were loaded with ordinary ancillary reagents. 9: 25 mM ETT in acetonitrile; 11: cap A (10% acetic anhydride in THF); 12: cap B (THF:pyridine:N-methylimidazole 8:1:1); 14: detritylating reagent (2% TCA in dichloromethane); 15: oxidizing reagent (20 mM iodine in THF:pyridine:water 8:1:1). Figure 1D shows the DNA synthesizer as described for the assembly.

The DNA sequences were programmed as:

5′ ctgcctgttccgtggccaaccctagtcactact 76 76 76 76 gttcaatgct tttcccgttac 3′.

The number 6 directed the simultaneous addition of trimers and dimers from bottle 6 and inactivation of the detritylation step for this cycle of synthesis. The number 7 directed the alkaline removal of Fmoc and coupling of the three monomers from bottle 7.

Once the degenerate oligonucleotides were assembled, the solid supports were treated with 1 ml dioxane/thiophenol/TEA (2:1:1 v/v) for 2 h at 55°C to remove the internucleotidic methyl and ethyl groups. Supernatants were removed by filtration and the supports were subjected to 55°C treatment overnight with concentrated ammonium hydroxide to remove all remaining nucleobase protecting groups and recover the oligonucleotides in solution. Both oligonucleotides were purified by denaturing polyacrylamide gel (12%) electrophoresis and recovered in deionized water after *n*-butanol desalting.

Construction of plasmids

pT4sgGFPr. Three silent nucleotide substitutions and one missense mutation were achieved on the *sgGFP* gene located in plasmid pQBI25f (QBIOgene) by site-directed mutagenesis employing the overlap extension method (19) and the oligonucleotides shown in Supplementary Figure S3. The substitutions 171a→g and 234t→c destroyed an NcoI and NdeI restriction site, respectively, whereas the substitution 120c→t created a new NdeI site upstream of the target region to be mutated (193c-204t). The contiguous substitutions 297ctt→aag created a HindIII site as well as the aa mutation F99S. The modified gene was amplified by PCR using the external primers NcoGFP and XhoGFP that generated an NcoI site at the 5′ terminus as well as a histidine tag and XhoI site at the 3′ terminus. The product was digested with the restriction enzymes NcoI/XhoI and cloned into the pT4 cloning vector which includes constitutive control of the trc promoter. The protein expressed from this plasmid was named sgGFPr and was considered our reference for subsequent comparisons.

pT4sgGFP^{NF}. This plasmid, containing a gene coding for a nonfluorescent sgGFP (sgGFP^{NF}) protein, was created by replacing the 183-bp segment NdeI/HindIII of plasmid pT4sgGFPr with a synthetic cassette that lacks codons for aa's L64-G67. The cassette was generated by PCR amplifying four partially complementary oligonucleotides as shown in the Supplementary Figure S3. In the first step, primers ΔLCYG (0.5 pmol) and Rev1 (0.5 pmol) were mixed with Vent DNA polymerase (2 U), dNTPs 10 mM (10 µl), ThermoPol Buffer 10X (10 µl) water (75 µl) and were extended via one PCR cycle with the conditions: 95°C, 5 min, 55°C, 5 min and 72°C, 10 min. Next, the mixture was cooled to room temperature, the primers Nde (20 pmol) and Hind (20 pmol) were added and the PCR was continued for 15 cycles with the conditions: 95°C, 1 min, 55°C, 1 min and 72°C, 1 min. The 204-bp fragment was purified by agarose gel (1.7%) electrophoresis, digested with the restriction enzymes NdeI/HindIII, then cloned into the plasmid pT4sgGFPr previously digested with the same enzymes and finally transformed into *E. coli*. After growing one day at 37°C in solid LB supplemented with KAN, three white clones were used to inoculate liquid LB supplemented with KAN and the plasmid was isolated and sequenced.

Libraries of mutant sgGFPs

The first library of mutant sgGFPs was constructed following the protocol for plasmid pT4GFP^{NF}, using the first degenerate oligonucleotide instead of ΔLCYG. However, in this case the recipient plasmid was pT4GFP^{NF} previously digested with the enzymes NdeI and HindIII. After ligation and electroporation, the transformants were recovered as a pool of plasmids by growing in liquid LB supplemented with kanamycin for 16 h at 37°C.

The second library of mutant sgGFPs was constructed by the overlap extension method using two intermediate DNA fragments produced by PCR and the second degenerate oligonucleotide as shown in Supplementary Figure S3. Fragment 1a was produced by extension of the primers NcoGFP (20 pmol, 2 µl) and Rev2 (20 pmol, 2 µl), using plasmid pT4GFP^{NF} (5 ng, 1 µl) as template and Vent DNA polymerase (2 U, 1 µl), dNTPs 10 mM (10 µl), ThermoPol Buffer 10X (10 µl) and water (74 µl) with the PCR conditions: 1X: 95°C, 3 min; 20X: 95°C, 1 min; 60°C, 1 min; 72°C, 1 min. The PCR product was purified by agarose gel (1.7%) electrophoresis and recovered via a PCR Product Purification Kit. Fragment 1b was produced by similar methods, using the second degenerate primer and primer XhoGFP for the PCR reaction. Fragment 2 was assembled by mixing Fragment 1a (5 ng, 1 µl), Fragment 1b (5 ng, 1 µl), *Taq* DNA polymerase (6 U, 3 µl), 10 mM dNTPs (10 µl), 10X Taq DNA polymerase Buffer (10 µl) and water (91 µl). After one PCR cycle (95°C, 5 min; 60°C, 5 min; 72°C, 10 min), the primers NcoGFP and XhoGFP were added (20 pmol each one) and the PCR was completed with 20 cycles of the following conditions: 95°C, 1 min; 60°C, 1 min; 72°C, 1 min. The final product was digested with the restriction enzymes NcoI/XhoI and ligated to plasmid pT4GFP^{NF}. The ligation was transformed into *E. coli*, grown in liquid LB supplemented with KAN and the library was recovered as a library of plasmids.

Screening, DNA sequencing and protein purification

The libraries of mutant plasmids were re-electroporated into *E. coli* cells and the transformants were grown on LB plates supplemented with KAN at 37°C for 24 h. Plasmids of 34 randomly chosen white colonies from the First Library and 55 from the Second Library were isolated and sequenced to analyze the codon diversity generated by the TrimerDimer method. From the Second Library two out of 50 000 colonies displayed a yellowish fluorescent phenotype to the naked eye. Plasmids from these active mutants were isolated, sequenced and the encoded mutant proteins were purified by affinity to Ni-NTA agarose (Qiagen), using a gradient of imidazole in phosphate buffer. Concentration of the proteins and removal of imidazole was achieved simultaneously by filtration with Millipore Amicon Ultra-4 centrifugal filter device (Ultracel –10 k). Two washes with 4 ml of 1X PBS were sufficient to remove the residual imidazole. The two

mutant proteins as well as the reference were re-dissolved in 1.5 ml 1X PBS and their concentration was determined by Bradford assay using the UV/visible spectrophotometer, Ultrospec 210 from Pharmacia.

Spectral measurements

The molar absorptivity (ε_{max}) of each protein, at their maximum absorbance peak, was determined using the spectrophotometer by measuring the absorbance of five duplicated dilutions. Plotting concentration (μM) versus absorbance rendered a linear graph whose slope is ε_{max} in units $\mu M^{-1} cm^{-1}$. Each of the dilutions was further diluted 25-fold and their fluorescence was measured at their maximum emission wavelength with excitation at their maximum absorbance wavelength. The plot of absorbance versus fluorescence produced a linear graph with a particular slope.

As a standard to determine the quantum yield (QY) of the proteins, we determined the relationship between absorbance (490 nm) and fluorescence (511 nm) for fluorescein, whose reported QY in 0.1 M NaOH is 0.79 (20). Dividing the sample's slope by fluorescein's slope and multiplication by 0.79 gives the sample's QY.

RESULTS AND DISCUSSION

Rationale of TrimerDimer mutagenesis

In a previous report (18), we described a mutagenesis approach to generate spiked codon-based amino acid substitutions in a protein region spanning six residues, in the active site of the enzyme TEM-1 ß-lactamase. Some ceftazidime-resistance mutants were found, with 2 or 3 aa replacements per variant. At the DNA level, codon variation was generated by reacting a growing oligo with a mixture of five Fmoc-protected dinucleotide phosphoramidites (dAA, dTT, dAT, dGC and dCG) and then with a mixture of four DMTr-protected monomer phosphoramidites (dA, dC, dG and dT) to produce a pool of 20 codons encoding 18 out of the 20 aa's (Supplementary Figure 1S; top). Only the codons for tryptophan and methionine were not generated and serine was the only aa represented twice. One stop codon was also produced, although in a nonsaturation mutagenesis approach the percentage of truncated proteins is not significant. For instance, if the library comprising a region of six amino acids is biased toward variants containing two amino acid replacements on average, only around 10% [$100(1–0.95^2)$] of this subpopulation will correspond to truncated proteins, whereas a saturation mutagenesis approach will produce a high ratio of truncated proteins as shown with 32 sequenced variants within the library polylinker100 (Figure 1S; bottom). This means 37.5% of the sequenced mutants contained the stop codon TAA. However, the situation gets worse for sequence space. Due to the lack of codons representing 2 aa, randomizing six codons at saturation will produce only 53% (100×0.90^6) of all possible aa sequences, and this percentage decreases linearly for longer regions.

For these reasons of stop codons, decreased variability of mutations produced, and because DMTr-protected trinucleotide phosphoramidites are expensive reagents for use in saturation mutagenesis approaches, we devised an alternative, novel method of saturation mutagenesis dubbed TrimerDimer. In this approach, five DMTr-protected trinucleotide phosphoramidites (DMTr-trimers) and five Fmoc-protected dinucleotide phosphoramidites (Fmoc-dimers) are simultaneously incorporated into an oligonucleotide growing in the ordinary direction $3' \rightarrow 5'$ (Figure 1A). Simultaneous addition of trimers and dimers will increase a fraction of the oligonucleotide by three nucleotides, whereas the remaining fraction will be increased by two nucleotides. After this step, the longer oligos are protected with DMTr whereas the shorter ones are protected with Fmoc. Because the DMTr group is stable to organic bases and Fmoc is labile, short treatment of the growing oligo with a basic solution and subsequent coupling with a mixture of three DMTr-protected monomer phosphoramidites (DMTr-monomers) of dC, dA and dG will produce 15 different trinucleotides. At this point, one mutagenesis cycle is completed, the 20 codons shown in Figure 1C have been incorporated into the sequence and the oligonucleotide is ready for either another mutagenesis cycle or for continuing the wild-type sequence assembly. Furthermore, most of the codons generated with TrimerDimer are the preferred codons used by *E. coli*, CTT being the exception, although its preference is similar to the second and third most used codons for leucine.

Chemical synthesis of DMTr-trimers

Because DMTr-monomers are commercially inexpensive reagents and Fmoc-dimers were already prepared, for this project we focused on production of DMTr-trimers (Scheme 1). These compounds were essentially synthesized as previously described (10), changing the 3′ hydroxyl protecting group of the starting nucleosides. Contrary to Virnekas *et al.* (10) we were able to prepare pure 3′-*O* Fmoc-protected monomers not only of dT and dC^{bz}, but also of dG^{ibu}. In the previous report the dG^{ibu} monomer was completely depurinated during the acylation reaction with Fmoc-chloride as demonstrated by mass spectrometry. We agree that depurination of dG^{ibu} as well as dA^{bz} is possible if a large excess of Fmoc-chloride is used and a long reaction time is allowed to produce hyperacylation of the purines. The electron attracting effect of the Fmoc group and the effect of the existing amino protecting group can weaken the glycosidic bond. In these conditions, the basicity of pyridine, the solvent used for the reaction, can complete the depurination reaction. However, the equimolar conditions used in our experiments as well as the short reaction times likely diminished this secondary reaction allowing acylation of the free 3′ hydroxyl to be the primary reaction. 3′-*O* Fmoc-monomers (dZ-Fmoc) were obtained in acceptable academic yields (>37%) via acylation of 5′-*O* DMTr-protected monomers (DMTr-dZ) and its subsequent detritylation using standard liquid conditions with trichloroacetic acid in dichloromethane rather than the strongly acidic ion exchange resin used in the previous report. The PAC group used to protect the 3′ end of the starting nucleosides was the main disadvantage of the

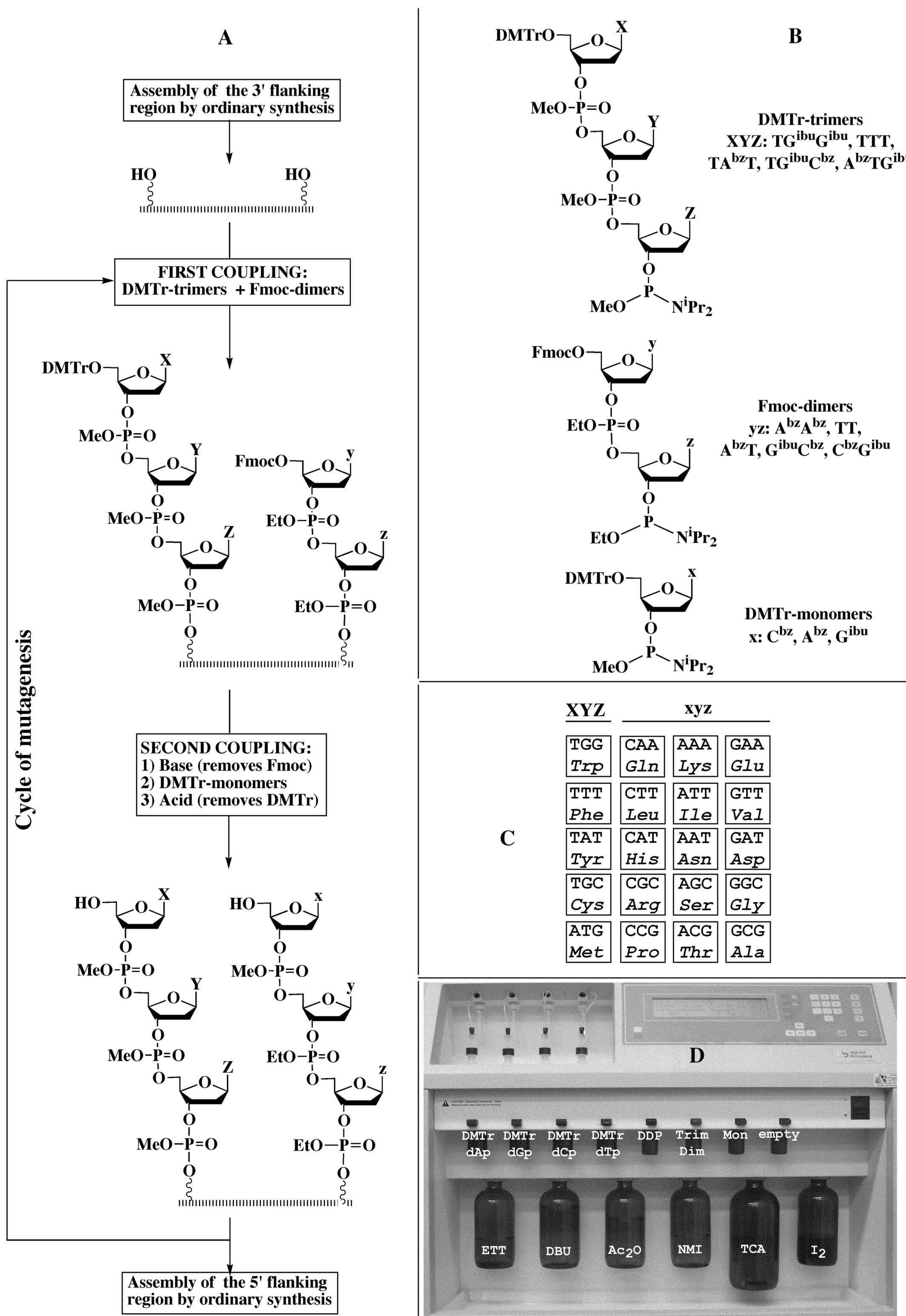


Figure 1. TrimerDimer mutagenesis approach. (**A**) The mutagenic cycle begins with a 5′ de-protected oligonucleotide growing on a solid support and is composed of two couplings. In the first coupling, a mix of DMTr-trimers and Fmoc-dimers is reacted with the oligonucleotide in the presence of ETT as an activating reagent. A phosphoramidite-based capping step, an ordinary capping step and ordinary oxidation step complete this cycle. In the second coupling, the Fmoc group is removed by alkaline hydrolysis. The recently incorporated dimers are converted in trimers by the addition of three DMTr-protected monomers and ETT, the cycle is completed as the first coupling, the DMTr group is removed by acid hydrolysis and the mutagenic cycle can be repeated or ordinary synthesis can be resumed to assemble the 5′ wild-type flanking region. (**B**) Chemical structures of monomers, dimers and trimers employed in the TrimerDimer mutagenesis approach. In these compounds, cytosine and adenine are protected with the benzoyl group and guanine with the isobutyryl group. Thymine is not protected. (**C**) Codons and encoded amino acids generated by TrimerDimer mutagenesis. (**D**) DNA synthesizer implemented to automate TrimerDimer mutagenesis. All abbreviations are described in 'Materials and Methods' section.

synthetic pathway described by Virnekas *et al.* (10) PAC was removed in such harsh alkaline conditions (methanolic ammonia) that a significant fraction of internucleotidic protecting groups and some of the nucleobase protecting groups were also removed, reducing the yield of the valuable DMTr-protected trinucleotides (DMTr-dXYZ) by up to 47%. Rather than using PAC, we cleanly removed the Fmoc group by ß-elimination using TEA, a bulky, weak and non-nucleophilic organic base that attacks neither the internucleotidic methyl groups nor the nucleobase protecting groups.

Although several pathways for the synthesis of trinucleotides have been published (11–13,21,22), we choose the strategy described by Virnekas *et al.* because the reaction of commercial DMTr-protected methyl phosphoramidites with 3′ *O*-protected compounds renders internucleotidic phosphates protected with the methyl group. This group is more resistant than o-chlorophenyl or β-cyanoethyl to DBU, a second organic base used in solid-phase synthesis for the removal of Fmoc. Since TEA and DBU are used in excess with respect to Fmoc-carrying compounds, the choice of the basic reagent is determined by its reactivity and the way the excess is eliminated after the reaction has been completed. TEA is more volatile but less reactive than DBU; therefore, after a prolonged treatment (15 min) of the fully protected trimers (DMTr-dXYZ-Fmoc) with TEA in solution-phase, the excess is easily removed by evaporation under high vacuum. DBU is not volatile, but in solid-phase removes the Fmoc group in only 15 s and the excess is discarded by simple filtration.

Therefore, the pathway shown in Scheme 1 represents an improved process for the synthesis of DMTr-trimers. The identity of the compounds DMTr-dXYZ was confirmed by comparative reverse-phase HPLC analysis, co-injecting each compound with its respective trinucleotide standard assembled by ordinary, solid-phase synthesis in the trityl-on mode. Such analysis was performed with compounds and standards independently de-protected with methanolic ammonia, which in the case of identity will produce the same de-protected trinucleotide. An example of this is shown for the trimer DMTr-dTGG (Supplementary Figure S2; top). Since the five analyses gave rise to only one peak, it was concluded that these precursors were synthesized correctly, so they were subsequently phosphitylated to the final products. The five DMTr-trimers were obtained with purities higher than 95% according to HPLC analysis. The reaction of an equimolar mixture of these compounds with each of the four nucleosides (dA, dG, dC and dT) linked to a solid support yielded a pool of five different tetramers as shown (Supplementary Figure S2; bottom). This experiment demonstrated the successful reactivity of the five DMTr-trimers in solid-phase synthesis and, therefore, potential application in the TrimerDimer mutagenesis approach.

Assembly of degenerate oligonucleotides

To assess the reproducibility and feasibility of the TrimerDimer mutagenesis approach, two degenerate primers were assembled using a mix of DMTr-trimers/Fmoc-dimers in approximately molar ratio 3:1 and the same sequence:

5′ *ctgcctgttccgtggccaaccctagtcactact* **xyz/XYZ xyz/XYZ xyz/XYZ xyz/XYZ** *gttcaatgcttttcccgttac* 3′

Here, xyz/XYZ represents the mix of 20 codons generated by the mixture of DMTr-trimers/Fmoc-dimers followed by the DMTr-monomers as shown in Figure 1A. Specifically, **x** represents dC, dA or dG, **yz** each of the dimers dAA, dTT, dAT, dGC or dCG and **XYZ** each of the trimers dTGG, dATG, dTTT, dTAT or dTGC shown in Figure 1B. The italicized sequences flanking the mutated region corresponded to wild-type segments included for priming purposes. Fmoc-dimers were used in 3-fold higher concentration than DMTr-trimers because each dimer had to generate three codons. Therefore, assuming similar reactivity between DMTr-trimers and Fmoc-dimers, at the end of each mutagenic cycle an even frequency of the 20 codons was expected. The assembly process was accomplished in completely automated mode in a DNA synthesizer equipped with eight vials to accommodate the four ordinary monomer phosphoramidites, the mix of trimers/dimers, the mix of monomers, and diethoxy-*N,N*,-diisopropylamino-phosphoramidite as a second 'capping reagent'. This phosphoramidite-based capping reagent was used in addition to the conventional capping reagent (acetic anhydride and 1-methylimidazole) to reduce the ratio of frameshift mutations generated by n-1 subproducts that usually reduce the fraction of useful variants in-phase. The DBU solution that had been used to remove the Fmoc group, was loaded in the position originally assigned for ammonium hydroxide and all the other positions were loaded with ordinary ancillary reagents as shown in Figure 1D. The synthesis program was similar to that recommended by the manufacturer, except in three points: (i) between the coupling step and ordinary capping step, the phosphoramidite-based capping step was incorporated and activated for every nucleotide addition; (ii) the detritylating step was deactivated only when the mix DMTr-trimers/Fmoc-dimers was incorporated to the growing oligo; and (iii) a brief DBU treatment was included and activated prior to the addition of the DMTr-monomers mix in order to remove the Fmoc group that was protecting the incorporated dimers. Once assembled, the degenerate primers were de-ethylated and de-methylated with thiophenol and finally deprotected with concentrated ammonium hydroxide. Analysis and purification of both oligonucleotides by polyacrylamide gel electrophoresis revealed one main spot of the expected size and indicated successful synthesis.

It is worth mentioning that prior to the assembly, HPLC analysis of the five Fmoc-dimers showed deterioration of only 5–10% of the original material in 7 years of storage at –20°C. This amount of deterioration is not significantly detrimental for oligonucleotide assembly, because of the large excess of reagents used in solid-phase synthesis and therefore the stability of these compounds is an additional value of the TrimerDimer mutagenesis approach.

Biological assessment

As a protein model to evaluate the feasibility of TrimerDimer as a mutagenesis tool, superglo green fluorescent protein (sgGFP) was used. sgGFP, originally reported as sg25 (23), is an engineered variant of *Aequorea victoria* GFP (avGFP) that carries the amino acid substitutions F64L, S65C, I167T and an alanine insertion numbered 1b, located between M1 and S2. Due to these changes, sgGFP is efficiently translated and expressed in mammalian cells, matures faster than avGFP and displays a unique excitation peak at 474 nm and a unique emission peak at 508 nm. With the naked eye, concentrated sgGFP, or cells that express it, display a green-brilliant phenotype. Like most fluorescent proteins, sgGFP contains a stretch of four contiguous amino acids (L64, C65, Y66 and G67) located in the center of the β-barrel structure that are important for the fluorescence activity. Cyclization of residues C65, Y66, G67 and subsequent oxygen oxidation gives rise to the protein's chromophore. Residue L64 is responsible for efficient protein expression at 37°C (24). Of the four residues, only G67 is essential for formation of the chromophore, whereas the other three residues can accept some amino acid replacements that usually modify the excitation, emission or expression of the original protein. With this background, the region L64-G67 was determined to be an excellent site to test random amino acid saturation by TrimerDimer. Finding new mutants was not expected because this region was extensively mutated at the early stage of recombinant GFP development (25).

In order to avoid using mammalian cells and amplify adequate sgGFP gene for our particular purposes, the original gene located in plasmid pQBI25f was modified via four PCR-based site-directed mutagenesis steps and cloned into the bacterial plasmid pT4 as a NcoI/XhoI insert to produce the plasmid pT4sgGFPr. This plasmid expresses the reference protein sgGFPr with the addition of a C-terminal histidine tail and modified with the folding mutation F99S. The site-directed mutations were intended to destroy two restriction sites (NcoI: nt 169–174, NdeI: nt 232–237) that would interfere with the subsequent work and to create two more sites (NdeI: nt 116–121, HindIII: nt 297–302) that flank the nucleotide region 193–204 which were to be substituted. The histidine tail was incorporated for use in purification and the mutation F99S was a result of the generation of the HindIII site. Although quantitative brightness comparison is not presented, cells expressing sgGFPr appeared more green and larger than cells expressing wt sgGFP, after growing 24 h at 37°C.

Preliminary results using pT4sgGFPr as cloning vector for creation of the libraries demonstrated that incomplete digestion of the plasmid with the pair of restriction enzymes NdeI/HindIII or NcoI/XhoI generated some false-positive active mutants that, in fact, corresponded to sgGFPr. To avoid this kind of fluorescent background, a new cloning vector expressing a nonfluorescent sgGFP mutant ($sgGFP^{NF}$) was constructed by replacing the stretch NdeI/HindIII in pT4sgGFPr with a synthetic cassette exclusively assembled by PCR of four partially complementary oligonucleotides, as shown in Supplementary Figure S3. One of these primers, ΔLCYG, lacked the nucleotide region 193–204 that encodes amino acids L64-G67. *Escherichia coli* cells transformed with the plasmid $pT4sgGFP^{NF}$ displayed a white, non-fluorescent phenotype quite different from green colonies expressing the reference protein.

Repeating the process for construction of $pT4sgGFP^{NF}$, but replacing the primer ΔLCYG with one of the degenerate oligonucleotides described above and using plasmid $pT4sgGFP^{NF}$ as the cloning vector, gave rises to the 'first library' of mutant sgGFPs. Although ~20 000 colonies from this library were grown on plates and visually screened under UV excitation at 360 nm, none of them was fluorescent. However, DNA sequencing of plasmids isolated from 34 white clones demonstrated that TrimerDimer mutagenesis was successful as the 20 expected codons were found in the mutants (Table 1).

Unfortunately, these sequences also revealed a high ratio of single nucleotide deletions outside of the mutagenesis region, generating frameshift mutations that reduced the ratio of useful mutant proteins. This result was consistant with confirmation of the original $sgGFP^{NF}$ gene sequence that required the analysis of three samples to find one without frameshift mutations. Most deleterious mutations were found in the cassette region which had been assembled with ordinary oligonucleotides and whose synthesis did not include the double capping step. In other projects, we have used the same approach to generate libraries, observing a low ratio of single nucleotide deletions because the cassette size was shorter than that used here (26). These results are also consistent with the general observation that single nucleotide deletions are the main bottleneck in the production of synthetic genes assembled with only oligonucleotides (27). The ratio of these deleterious mutations is directly proportional to the size of the gene.

In order to improve the quality of the library, the cassette mutagenesis was abandoned and a second library was constructed by enzymatic amplification (19) of the second degenerate oligonucleotide, using $sgGFP^{NF}$ gene as template. Nucleotide substitutions errors performed by polymerases are definitely more tolerated in the encoded proteins due to the protective effect of the genetic code than nucleotide deletions. Thus, whereas in the first library a segment of $sgGFP^{NF}$ was replaced with a synthetic cassette, the second library was created by replacing the complete gene in $pT4sgGFP^{NF}$ with a library of PCR-amplified mutant genes. Through the generation of two intermediate PCR fragments as shown in Figure S3 and a second PCR using two external primers, the gene library was generated and the template virtually eliminated. The DNA sequence of 55 plasmids isolated from white colonies (Table 1) not only confirmed the reproducibility of TrimerDimer but also confirmed the good quality of the library because only four mutants contained frameshift mutations.

When the mutations of both libraries were grouped by codons, as shown in Table 2, and the concentration of the Fmoc-dimers and DMTr-trimers used for the

Table 1. DNA sequence of mutants created with TrimerDimer mutagenesis and isolated from the first and second library

Library	Sample	DNA sequence (amino acid)			
	sgGFPr	CTG(L)	TGC(C)	TAC(Y)	GGT(G)
First	1	TAT(Y)	TAT(Y)	ATG(M)	*TCC*(S)
	2	AGC(S)	CTT(L)	AAA(K)	CAA(Q)
	3	CCG(P)	CTT(L)	TGC(C)	GTT(V)
	4	AAA(K)	CAT(H)	CAT(H)	TGC(C)
	5	ATT(I)	CTT(L)	CAT(H)	TGC(C)
	6	CTT(L)	GTT(V)	GAT(D)	TGC(C)
	7	AAA(K)	CAT(H)	CCG(P)	TAT(Y)
	8	GCG(A)	TAT(Y)	GTT(V)	ACG(T)
	9	GCG(A)	CTT(L)	GCG(A)	ACG(T)
	10	AGC(S)	CAT(H)	CTT(L)	CAA(Q)
	11	ATG(M)	CCG(P)	AAT(N)	CCG(P)
	12	GAA(E)	CGC(R)	CAA(Q)	TGC(C)
	13	TAT(Y)	CAA(Q)	GAA(E)	TTT(F)
	14	AAT(N)	AAA(K)	CTT(L)	AAA(K)
	15	AAT(N)	CAT(H)	CAA(Q)	GTT(V)
	16	TGG(W)	AAA(K)	GGC(G)	GAT(D)
	17	CAA(Q)	GAA(E)	CAT(H)	ATT(I)
	18	GAA(E)	GTT(V)	AAA(K)	GCG(A)
	19	AGC(S)	ACG(T)	GAA(E)	AAA(K)
	20	AAA(K)	ATG(M)	ATG(M)	TTT(F)
	21	ATG(M)	TGG(W)	AAT(N)	ACG(T)
	22	GAT(D)	TGG(W)	ATG(M)	ATG(M)
	23	TTT(F)	GAT(D)	TTT(F)	TGC(C)
	24	GGC(G)	ACG(T)	GAT(D)	GCG(A)
	25	TGC(C)	AAT(N)	CTT(L)	GAA(E)
	26	AAA(K)	TTT(F)	CAA(Q)	CGC(R)
	27	GCG(A)	TTT(F)	TGC(C)	GAA(E)
	28	ATG(M)	CCG(P)	CAA(Q)	TGC(C)
	29	CGC(R)	AAT(N)	GCG(A)	GAT(D)
	30	CAA(Q)	CGC(R)	CGC(R)	CAT(H)
	31	ACG(T)	ATT(I)	CAA(Q)	CAA(Q)
	32	GCG(A)	GCG(A)	ACG(T)	AAA(K)
	33	*CCA*(P)	CGC(R)	GCG(A)	ACG(T)
	34	CCG(P)	CTT(L)	GCG(A)	GTT(V)
Second	1	GAT(D)	ATG(M)	GAA(E)	GAA(E)
	2	ACG(T)	TGC(C)	AAA(K)	TGC(C)
	3	AAT(N)	AAA(K)	GTT(V)	TGG(W)
	4	*CTG*(L)	AAT(N)	GTT(V)	GCG(A)
	5	TAT(Y)	CCG(P)	GTT(V)	ATT(I)
	6	AAA(K)	GAA(E)	ACG(T)	AGC(S)
	7	TGC(C)	CCG(P)	CCG(P)	GAT(D)
	8	CAT(H)	AAA(K)	TAT(Y)	GAA(E)
	9	AGC(S)	GAA(E)	GAA(E)	TGC(C)
	10	*TTA*(L)	CAA(Q)	GCG(A)	AGC(S)
	11	AAT(N)	CAA(Q)	CGC(R)	GAA(E)
	12	AGC(S)	CAT(H)	AGC(S)	CGC(R)
	13	ATT(I)	GGC(G)	CAA(Q)	CAA(Q)
	14	CTT(L)	TGC(C)	TGG(W)	GTT(V)
	15	TGC(C)	ATG(M)	TAT(Y)	TGG(W)
	16	GCG(A)	ATG(M)	GGC(G)	GTT(V)
	17	GAT(D)	GTT(V)	ACG(T)	CAA(Q)
	18	AAT(N)	CAA(Q)	AAA(K)	GTT(V)
	19	AGC(S)	GTT(V)	TGG(W)	GAT(D)
	20	AGC(S)	CAT(H)	GCG(A)	CTT(L)
	21	CGC(R)	CCG(P)	GAA(E)	TGG(W)
	22	GTT(V)	AAT(N)	TGC(C)	TGG(W)
	23	ATT(I)	AGC(S)	AAT(N)	AGC(S)
	24	TGG(W)	TGG(W)	CGC(R)	GAA(E)
	25	ACG(T)	TGG(W)	GGC(G)	ATT(I)
	26	CCG(P)	CGC(R)	AAA(K)	AGC(S)
	27	GGC(G)	CTT(L)	ATG(M)	ggcgc
	28	TAT(Y)	AAT(N)	AAT(N)	GAA(E)
	29	AAT(N)	GTT(V)	CGC(R)	GCG(A)
	30	GAA(E)	GAA(E)	GGC(G)	GGC(G)
	31	A--	GCG(A)	GCG(A)	AGC(S)
	32	AAT(N)	TTT(F)	TAT(Y)	GTT(V)
	33	ATG(M)	ACG(T)	CAA(Q)	CTT(L)
	34	CGC(R)	GTT(V)	AAA(K)	CAA(Q)
	35	GAA(E)	ATG(M)	CGC(R)	GAT(D)
	36	AGC(S)	GTT(V)	CAA(Q)	ATG(M)
	37	AGC(S)	GTT(V)	CAA(Q)	ATG(M)
	38	AGC(S)	GAT(D)	AAA(K)	ACG(T)
	39	CTT(L)	GGC(G)	GGC(G)	GGC(G)
	40	gttat	GAA(E)	TGC(C)	AGC(S)
	41	GCG(A)	CGC(R)	CAA(Q)	ATG(M)
	42	AAT(N)	TTT(F)	AAT(N)	CAA(Q)
	43	CAT(H)	TTT(F)	AAA(K)	AAT(N)
	44	AAA(K)	CGC(R)	CGC(R)	GTT(V)
	45	GTT(V)	CTT(L)	GTT(V)	ACG(T)
	46	TGC(C)	GTT(V)	CCG(P)	GTT(V)
	47	GAA(E)	TGG(W)	TGG(W)	CCG(P)
	48	GAA(E)	CGC(R)	*TAG*(X)	GTT(V)
	49	AAA(K)	ACG(T)	CAA(Q)	ACG(T)
	50	AAA(K)	GTT(V)	TGG(W)	CAA(Q)
	51	GTT(V)	GAA(E)	AGC(S)	CGC(R)
	52	ATG(M)	TGG(W)	GGC(G)	AAT(N)
	53	AGC(S)	GCG(A)	CCG(P)	GAT(D)
	54	AGC(S)	ATG(M)	TGC(C)	AAT(N)
	55	TAT(Y)	GAA(E)	TGG(W)	TGG(W)
	1576	CTT(L)	GCG(A)	TAT(Y)	GGC(G)
	1723	GCG(A)	GCG(A)	TAT(Y)	GGC(G)

Only the mutated region is shown and the encoded amino acids are indicated in parenthesis in the one-letter code.
sgGFPr is the reference protein.
Samples 1–34 from the first library and samples 1–55 from the second library were nonfluorescent mutants isolated from the first and second library, respectively.
1576 and 1723 are the fluorescent mutants, sgGFP-C65A and sgGFP-L64A/C65A respectively, isolated from the second library.
Unexpected TrimerDimer codons are indicated in italics, nucleotide deletions are indicated with hyphens and nucleotide insertions are in lowercase letters. Mutant 22 from the second library contained the single nucleotide deletion 182delT, located 10 nt upstream from the mutated region.

oligonucleotide assemblies were averaged, the following conclusions were drawn:

(i) The five Fmoc-dimers and five DMTr-trimers were successfully incorporated into the degenerate oligonucleotides. Subsequent reaction of dimers with the DMTr-monomers gave rise to the complete set of 20 codons. The occurrence of the 20 codons in both libraries, using different degenerate oligonucleotides, confirmed the reproducibility of TrimerDimer.

(ii) An ideal randomization approach must yield 5% of each mutant codon. TrimerDimer was close to produce this ideal frequency because most of the mutant codons occurred around this rate. Only ATT was underrepresented (1.97%) and GTT over represented (7.58%), whereas CTT, the third codon generated with dimer TT occurred at 4.21%, close to the expected value. Therefore, oligonucleotides ending with T seem to preferentially react with G rather than with A. However, more experiments will be needed to address this unexpected bias.

Table 2. Mutant codons found in 89 nonfluorescent variants randomly isolated from the first and second libraries

Codon (aa)	Found codons	Frequency[a] (%)	Frequency[b] per dimer or group of trimers (%)
AAA (K)	23	6.61	20.69
CAA (Q)	24	6.90	(12.6)
GAA (E)	25	7.18	
ATT (I)	7	2.01	14.08
CTT (L)	15	4.31	(13.6)
GTT (V)	27	7.76	
AAT (N)	21	6.03	13.79
CAT (H)	14	4.02	(13.8)
GAT (D)	13	3.74	
AGC (S)	20	5.75	14.95
CGC (R)	19	5.46	(17.6)
GGC (G)	13	3.74	
ACG (T)	18	5.17	14.94
CCG (P)	14	4.02	(15.5)
GCG (A)	20	5.75	
TGG (W)	18	5.17 (7.9)	21.55
ATG (M)	18	5.17 (5.0)	(26.8)
TTT (F)	9	2.59 (4.7)	
TAT (Y)	11	3.16 (4.4)	
TGC (C)	19	5.46 (4.8)	
Not expected: ctg, tta, tag, tcc, cca	5	1.40	2.24
INDELs	3	0.84	
Total	356		

[a]Values in parenthesis are the average molar ratio per trimer in the mixtures of DMTr-trimers/Fmoc-dimers that were used to synthesize the pair of degenerate oligonlucleotides.
[b]Values in parenthesis are the average molar ratio per dimer or group of trimers in the mixtures of DMTr-trimers/Fmoc-dimers that were used to synthesize the pair of degenerate oligonucleotides.

(iii) DMTr-trimers were 5.7% less reactive than Fmoc-dimers. In the mutagenic mixture, trimers represented 26.8% of the total molar concentration, whereas in the libraries the codons generated with these compounds occurred at a rate of 21.1%. Thus, the molar ratio of trimers in the mutagenic mixture must be increased to 32% to yield 25% of codons. However, not all trimers displayed the same reactivity. For example, TTT and TAT were less abundant than ATG and TGC. TGG was close to the ideal frequency (5.06%) because this was the most concentrated trimer (7.9%) in the mutagenic mixture. Based on these experimental results, we have calculated the ideal concentration of each dimer and trimer to yield an equal frequency of codons (Supplementary Table S1).

(iv) AA was the most reactive Fmoc-dimer because the codons AAA, CAA and GAA were the most hugely represented, even when the AA dimer was the least concentrated in the mutagenic mixture.

(v) GC was the least reactive dimer because a higher concentration of this compound in the mutagenic mixture was required to yield 14.61% of the three codons AGC, CGC and GGC. The ideal frequency per group of codons generated with each Fmoc-dimer is 15%.

(vi) The frequency of the other codons generated with Fmoc-dimers TT, AT and CG corresponded to the molar ratio of each dimer in the mutagenic mixture.

(vii) Since the mixture of DMTr-monomers employed in the conversion of dimers to trimers was equimolar, a rate of 33.3% was expected for each base. However, dC occurred at a rate of 27.1%, dG at 38.8% and dA at 34.1%, indicating differential reactivity. Therefore, the mixture of monomers should also be modified to obtain an equal frequency of each monomer and, therefore, an equal frequency of codons.

(viii) Five codons different from those expected by TrimerDimer were found in the mutated region. These unexpected codons represented only 1.4% of all mutant codons. It is likely these codons arose by the natural substitution error rate of the DNA polymerases employed during the assembly of the mutant genes or the DNA cassette. For instance, codon CTG may have arisen from CTT, CCG or ATG, with CCG being the most probable origin because enzymatic transitions (e.g. c→t) are favored over enzymatic transversions. TTA had to be generated from TTT, TCC from TGC, TAG from TGG or TAT and CCA from CCG or CAA. These results correlate with the polymerase experimental error rate (10^{-3}) found in the second library outside of the mutated segment.

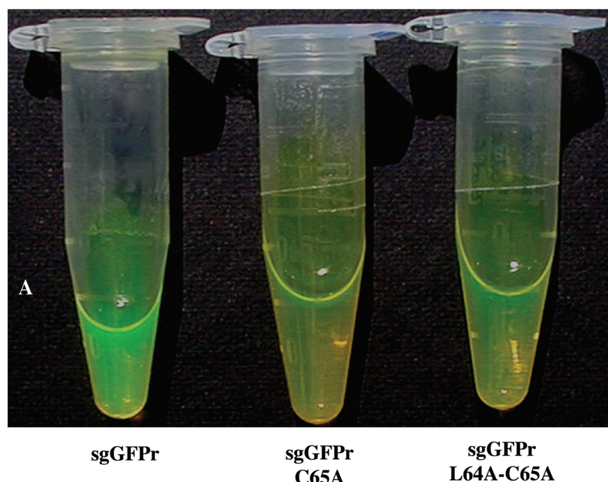
(ix) Contrary to the first library, the second library assembled by the overlap PCR approach displayed good quality. Only 7.3% of the clones contained single or double nucleotide insertions or deletions versus 56% of the first library. Therefore, the overlap PCR approach is better than the cassette approach for the generation of libraries.

Once the generation of diversity by TrimerDimer mutagenesis was confirmed, the second library was screened to search for fluorescent mutants. Only two out of ~50 000 colonies displayed a fluorescent phenotype when illuminated with long wavelength UV light. When their plasmids were isolated and sequenced, it was found that one of these mutants corresponded to the single mutant sgGFPr-C65A and the other one to the double mutant sgGFPr-L64A/C65A. As a consequence of mutation L64A, mutant sgGFPr-L64A/C65A was expressed 60% less in culture than mutant sgGFPr-C65A at 37°C. This finding is congruent with previous experiments which found that the mutation F64L markedly improved the solubility of avGFP at 37°C (24).

Both mutant proteins, as well as the reference protein sgGFPr, were purified by affinity to Ni and their excitation maxima (λ_{exc}), emission maxima (λ_{em}), molar absortivity (ε_{max}) and QY were determined and reported in Table 3. Whereas the molar absorptivity was similar for the three proteins, the QY of the mutants was notably lower than that of the reference protein. This demonstrates that the presence of a hydrophobic amino acid at position 65 reduces protein brightness.

Table 3. Fluorescent properties of the reference protein sgGFPr and functional mutants found by TrimerDimer mutagenesis

Protein	λ_{exc} (nm)	λ_{em} (nm)	ε_{max}	QY
sgGFP wt	474	504	45 104	0.737
sgGFP-C65A	471	502	46 170	0.388
sgGFP-L64A/C65A	470	500	40 798	0.457



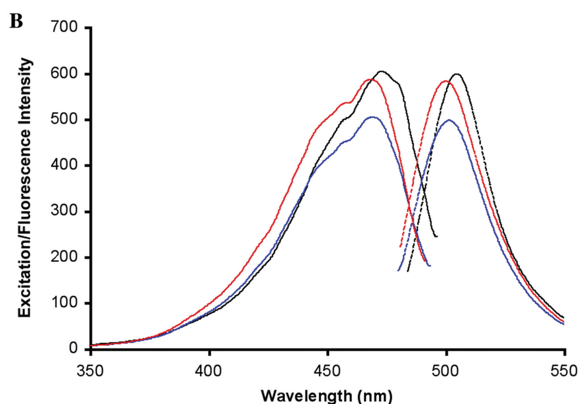


Figure 2. (**A**) Pure proteins sgGFPr, sgGFPr-C65A and sgGFPr-L64A/C65A photographed under sunlight. (**B**) Fluorescence excitation and emission spectra (solid and dashed lines, respectively) of the pure proteins sgGFPr (black), sgGFPr-C65A (blue) and sgGFPr-L64A/C65A (red).

Under visual inspection, the purified mutant proteins sgGFPr-C65A and sgGFPr-L64A/C65A (Figure 2A) as well as *E. coli* cells that expressed them displayed a brilliant yellowish color obviously different from the green color of sgGFPr. The mutation C65A, common to both mutants, is responsible for this apparent color change. However, this result is intriguing as all previously reported yellow mutants have replaced T203 with an aromatic aa and they show an emission batochromic shift from around 500 nm to 520 nm, a consequence of a stacking effect over the chromophore (28). Our mutants exhibit emission maximum around 500 nm and they look yellow. Perhaps the intersection between the excitation and emission spectrum is causing this unusual effect as seen in Figure 2B.

Finally, it is worth mentioning that a library spanning a region of four codons must be composed of 160 000 variants if each position is randomized with 20 codons. For screening with 99% coverage, at least 736 825 clones must be analyzed. In our experiments only 7% of this population was analyzed and for that reason neither the original sgGFPr protein nor other fluorescent mutants were found. A high-throughput screening approach such as FACS would likely find all possible active mutants.

CONCLUSION

We have demonstrated the robustness of TrimerDimer mutagenesis for the elimination of redundant and stop codons when gene regions are subjected to random saturation. The reagents used for this approach are easily synthesized and highly stable when stored properly. Contrary to the resin-splitting approach developed by Neuner *et al.* (15), the assembly process of degenerate primers by TrimerDimer can be completely automated in any of several DNA synthesizers and once the mixture DMTr-trimers/Fmoc-dimers is prepared, the process is accesable even for nonexpert chemists. Only minor codon bias is generated and can be corrected by adjusting the concentration of the appropriate trimer or dimer.

TrimerDimer is highly versatile. As shown in Figure 1C, one can infer that selection of certain trimers as well as certain dimers and monomers can yield interesting subsets of amino acids. Additionally, TrimerDimer is not only applicable to the coding strand but also to the noncoding strand if specific combinations of trimers, dimers and monomers are used to produce a set of 20 anticodons that, after generation of the complementary strand are converted into codons for the 20 different aa's. These possible combinations are outlined in the Supplementary Figure S4. Currently, we have the five Fmoc-protected trimers (dAAA, dCCA, dATA, dGCA and dCAT), the five Fmoc-protected dimers (dAA, dTT, dAT, dGC and dCG) and the three DMTr-protected monomers (dC, dT and dG) to perform such a strategy.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Technical assistance by Eugenio López, Santiago Becerra and Soledad Juárez-Ramírez is highly appreciated. Filiberto Sánchez helped in some cloning steps, and Leopoldo Guereca provided advice for purification and analyses of proteins. Dr. Hugh Mackie from Glen Research Corporation, Thomas Godby from Bioautomation and Dr. Mari L. Salmi from our Institute kindly reviewed this manuscript.

FUNDING

Funding for open access charge: The Open Access publication charges for this paper has been waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

1. Neylon,C. (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res*, **32**, 1448–1459.
2. Cadwell,R.C. and Joyce,G.F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl.*, **2**, 28–33.
3. Derbyshire,K.M., Salvo,J.J. and Grindley,N.D. (1986) A simple and efficient procedure for saturation mutagenesis using mixed oligodeoxynucleotides. *Gene*, **46**, 145–152.
4. Hutchison,C.A. III, Nordeen,S.K., Vogt,K. and Edgell,M.H. (1986) A complete library of point substitution mutations in the glucocorticoid response element of mouse mammary tumor virus. *Proc. Natl Acad. Sci. USA*, **83**, 710–714.
5. Sirotkin,K. (1986) Advantages to mutagenesis techniques generating populations containing the complete spectrum of single codon changes. *J. Theor. Biol.*, **123**, 261–279.
6. Reetz,M.T. and Carballeira,J.D. (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.*, **2**, 891–903.
7. Miyazaki,K. and Arnold,F.H. (1999) Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J. Mol. Evol.*, **49**, 716–720.
8. Wachter,R.M., Elsliger,M.A., Kallio,K., Hanson,G.T. and Remington,S.J. (1998) Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure*, **6**, 1267–1277.
9. Hughes,M.D., Nagel,D.A., Santos,A.F., Sutherland,A.J. and Hine,A.V. (2003) Removing the redundancy from randomised gene libraries. *J. Mol. Biol.*, **331**, 973–979.
10. Virnekas,B., Ge,L., Pluckthun,A., Schneider,K.C., Wellnhofer,G. and Moroney,S.E. (1994) Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res.*, **22**, 5600–5607.
11. Kayushin,A.L., Korosteleva,M.D., Miroshnikov,A.I., Kosch,W., Zubov,D. and Piel,N. (1996) A convenient approach to the synthesis of trinucleotide phosphoramidites—synthons for the generation of oligonucleotide/peptide libraries. *Nucleic Acids Res.*, **24**, 3748–3755.
12. Ono,A., Matsuda,A., Zhao,J. and Santi,D.V. (1995) The synthesis of blocked triplet-phosphoramidites and their use in mutagenesis. *Nucleic Acids Res.*, **23**, 4677–4682.
13. Lyttle,M.H., Napolitano,E.W., Calio,B.L. and Kauvar,L.M. (1995) Mutagenesis using trinucleotide beta-cyanoethyl phosphoramidites. *Biotechniques*, **19**, 274–281.
14. Krumpe,L.R., Schumacher,K.M., McMahon,J.B., Makowski,L. and Mori,T. (2007) Trinucleotide cassettes increase diversity of T7 phage-displayed peptide library. *BMC Biotechnol.*, **7**, 65.
15. Neuner,P., Cortese,R. and Monaci,P. (1998) Codon-based mutagenesis using dimer-phosphoramidites. *Nucleic Acids Res.*, **26**, 1223–1227.
16. Yanez,J., Arguello,M., Osuna,J., Soberon,X. and Gaytan,P. (2004) Combinatorial codon-based amino acid substitutions. *Nucleic Acids Res.*, **32**, e158.
17. Gaytan,P., Yanez,J., Sanchez,F. and Soberon,X. (2001) Orthogonal combinatorial mutagenesis: a codon-level combinatorial mutagenesis method useful for low multiplicity and amino acid-scanning protocols. *Nucleic Acids Res.*, **29**, E9.
18. Gaytan,P., Osuna,J. and Soberon,X. (2002) Novel ceftazidime-resistance beta-lactamases generated by a codon-based mutagenesis method and selection. *Nucleic Acids Res.*, **30**, e84.
19. Ho,S.N., Hunt,H.D., Horton,R.M., Pullen,J.K. and Pease,L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51–59.
20. Umberger,J.Q. and LaMer,V.K. (1945) The kinetics of diffusion controlled molecular and ionic reactions in solution as determined by measurements of the quenching of fluorescence. *J. Am. Chem. Soc.*, **67**, 1099–1109.
21. Zehl,A., Starke,A., Cech,D., Hartsch,T., Merkl,R. and Fritz,H. (1996) Efficient and flexible access to fully protected trinucleotides suitable for DNA synthesis by automated phosphoramidite chemistry. *Chem. Commun.*, 2677–2678.
22. Yagodkin,A., Azhayev,A., Roivainen,J., Antopolsky,M., Kayushin,A., Korosteleva,M., Miroshnikov,A., Randolph,J. and Mackie,H. (2007) Improved synthesis of trinucleotide phosphoramidites and generation of randomized oligonucleotide libraries. *Nucleosides Nucleotides Nucleic Acids*, **26**, 473–497.
23. Palm,G.J., Zdanov,A., Gaitanaris,G.A., Stauber,R., Pavlakis,G.N. and Wlodawer,A. (1997) The structural basis for spectral variations in green fluorescent protein. *Nat. Struct. Biol.*, **4**, 361–365.
24. Cormack,B.P., Valdivia,R.H. and Falkow,S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, **173**, 33–38.
25. Palm,G.J. and Wlodawer,A. (1999) Spectral variants of green fluorescent protein. *Methods Enzymol.*, **302**, 378–394.
26. Flores-Ramirez,G., Rivera,M., Morales-Pablos,A., Osuna,J., Soberon,X. and Gaytan,P. (2007) The effect of amino acid deletions and substitutions in the longest loop of GFP. *BMC Chem. Biol.*, **7**, 1.
27. Smith,H.O., Hutchison,C.A. III, Pfannkoch,C. and Venter,J.C. (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl Acad. Sci. USA*, **100**, 15440–15445.
28. Tsien,R.Y. (1998) The green fluorescent protein. *Annu. Rev. Biochem.*, **67**, 509–544.