Spiked Genes: A Method to Introduce Random Point Nucleotide Mutations Evenly throughout an Entire Gene Using a Complete Set of Spiked Oligonucleotides for the Assembly

Edson Cárcamo,†,§ Abigail Roldán-Salgado,†,§ Joel Osuna,† Iván Bello-Sanmartín,† Jorge A. Yáñez,† Gloria Saab-Rincón,† Héctor Viadiu,‡ and Paul Gayta*†

†Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Cuernavaca, Morelos 62210, México
‡Instituto de Química, Universidad Nacional Autónoma de México, Circuito Exterior, Ciudad Universitaria, Ciudad de México 04510, México

Supporting Information

ABSTRACT: In vitro mutagenesis methods have revolutionized biological research and the biotechnology industry. In this study, we describe a mutagenesis method based on synthesizing a gene using a complete set of forward and reverse spiked oligonucleotides that have been modified to introduce a low ratio of mutant nucleotides at each position. This novel mutagenesis scheme named “Spiked Genes” yields a library of clones with an enhanced mutation distribution due to its unbiased nucleotide incorporation. Using the far-red fluorescent protein emKate as a model, we demonstrated that Spiked Genes yields richer libraries than those obtained via enzymatic methods. We obtained a library without bias toward any nucleotide or base pair and with even mutations, transitions, and transversion frequencies. Compared with enzymatic methods, the proposed synthetic approach for the creation of gene libraries represents an improved strategy for screening protein variants and does not require a starting template.

INTRODUCTION

The ability to produce a large number of protein variants that are screened to optimize the desired protein function continues to expand our technological progress. Directed molecular evolution relying on random and semirandom mutagenesis of genes has become a powerful strategy for modifying protein properties.¹,² There are already many proteins whose function has been modified by selecting variants from a created protein library. For example, variant proteins with stability to high temperatures, extreme pH,⁴,⁵ and organic solvents⁶,⁷ have been obtained. Moreover, variant enzymes with improved activities⁸,⁹ or specificity changes¹⁰,¹¹ are available. Other physical properties, such as oligomerization of a target protein¹²,¹³ or color of fluorescent proteins have also been modified.¹⁴,¹⁵

Although screening and selection are intrinsically related to the particular properties of the target protein, the generation of genetic diversity at the gene level is a more general issue and depends on the mutagenesis method employed.¹,²,¹⁶ Soon after the advent of the polymerase chain reaction (PCR) technique, it was determined that random point mutations were generated during the in vitro amplification process¹⁷,¹⁸ and protocols were developed to take advantage of such mutations, creating the error-prone PCR (epPCR) mutagenesis approach.¹⁹,²⁰ Now, it is well known that manganese ions and unbalanced dNTP ratios,²¹,²² nucleotide analogs,²³ and low-fidelity DNA polymerases²⁴,²⁵ promote the random incorporation of erroneous nucleotides.

On the other hand, early studies on oligonucleotide synthesis performed on a solid phase demonstrated that when added in equimolar concentrations, the four DNA monomer phosphoramidites dA, dC, dG, and dT have similar chemical reactivity.²⁶ In 1986, taking advantage of the equal nucleotide reactivity, the 30 bp glucocorticoid response element of the mouse mammary tumor virus was randomized by spiked oligonucleotides, wherein each position was doped during automated DNA synthesis by the low levels of the other three phosphoramidites.²⁷ This work confirmed the robustness of the method to produce all of the expected mutations by sequencing 546 clones and finding 88 different base substitutions out of the 90 possible single mutants. A similar mutagenesis approach was reported the following year by Hill et al.²⁸ and 4 years later by Dale et al.²⁹ Using genes as templates, researchers incorporated spiked oligonucleotides by PCR-based methods to modify specific regions of the encoded proteins.³⁰ More recently, Hidalgo et al.³¹ and Jin et al.³² reused the spiked oligonucleotides to achieve a focused and directed evolution in certain regions of the regulatory elements and enzymes via in
Oligonucleotides rapidly increase the number of variants at the nucleotide level, synthesized with saturated degeneracies at some positions contrary to the spiked oligonucleotides, ordinary primers precise annealing required to form full-length genes is achieved selected by sequence alignments of homologous genes. The ing saturated degeneracies at the desired nucleotide positions, assembled by appropriately prepared oligonucleotides contain ing saturated degeneracies at the desired nucleotide positions, selected by sequence alignments of homologous genes. The precise annealing required to form full-length genes is achieved by controlled overlapping of the designed oligomers. However, contrary to the spiked oligonucleotides, ordinary primers synthesized with saturated degeneracies at some positions rapidly increase the number of variants at the nucleotide level and enrich the libraries with protein variants harboring multiple amino acid mutations, which usually affect either the folding or function of most encoded proteins. The fundamental difference between spiked and ordinary degenerated oligonucleotides relies on the level of doping of the wild-type nucleotides. In spiked oligonucleotides each wild-type nucleotide is “doped” during the chemical synthesis, with a very low ratio of the other three bases, yielding oligonucleotide mutants harboring very few changes per chain and a high ratio containing the wild-type sequence. However, in ordinary commercial degenerated oligonucleotides, the target nucleotides are completely replaced with an equimolar ratio of the desired mixed bases: N: A/G/C/T, V: A/C/G, D: A/G/T, H: A/T/C, B: C/G/T, M: A/C, R: A/G, W: A/T, S: C/G, Y: C/T, or K: G/T, producing oligonucleotide mutants harboring multiple mutations per chain and practically nothing of the wild-type sequence when more than 12 nucleotides are randomized with N.

In this article, for the first time, we report the strategy of using a complete set of spiked oligonucleotides to introduce mutations throughout an entire synthetic gene, an approach that according to a review written by Wong et al. has not been reported yet. This method is not limited to a target region, as previously described, or to a reduced number of target nucleotides. This synthetic strategy, which we have named “Spiked Genes”, was applied to the monomeric enhanced red fluorescent protein emKate, a mutation that makes it more fluorescent than its parental protein. Moreover, some amino acid replacements on its scaffold produce variants of a different color, similar to the protein DsRed that served as a scaffold for producing a palette of fluorescent proteins termed mFruits.

### RESULTS AND DISCUSSION

#### Gene Library Synthesis.

The synthetic emKate gene library was assembled, as described in the Methods section, by a one-step strategy using a diluted equimolar pool of 22 internal spiked oligonucleotides and two outermost spiked oligonucleotides as extension primers (Figure 1). The optimal concentration of the internal oligonucleotides was 4 nM, 100-fold less concentrated than that of the external primers. The assembly efficiency was sensitive at higher internal oligonucleotide concentrations likely due to mispairing between some of them and the external primers. We did not observe any assembly efficiency difference between the wild-type and spiked mutant oligonucleotides.

To evaluate the higher efficiency of Spiked Genes for generating random point nucleotide substitutions, we compared our method with the mutagenesis efficiencies, over the emKate gene, of three enzymatic epPCR-based methods: two commercial mutagenesis kits (GeneMorph II and Diversify) and the traditional epPCR method. The Diversify PCR Random Mutagenesis Kit is based on the methods of Leung et al. and Cadwell and Joyce and relies on the use of Titanium Taq polymerase and different concentrations of manganese and dGTPs to achieve different mutation rates. GeneMorph II is a commercial mutagenesis kit containing Mutazyme II: a blend of two error-prone DNA polymerases. One of these enzymes is Mutazyme I, which favors mutations at G’s and C’s, whereas the other is a novel Taq DNA polymerase mutant that exhibits increased misinsertion and misextension frequencies at A’s and T’s compared with those of the wild-type Taq.

#### Analysis of the Libraries.

To evaluate the mutagenic efficiency of the Spiked Genes method versus the three error-prone enzymatic methods, we used three parameters to define the quality of DNA libraries: (1) the mutation distribution, which evaluates the capacity of the method to spread point mutations along the complete gene sequence, (2) the mutation rate, which evaluates the prediction ability of the method to achieve an expected number of point mutations per kb, and (3) the mutational diversity, also known as the mutational

---

Figure 1. Strategy for the assembly of the synthetic gene library. Each continuous arrow represents a spiked oligonucleotide, wherein each position was doped with 0.25% of each of the other three bases. Primer sequences are listed in Table S1. (A) Scheme of the single-step PCR assembly reaction to synthesize the emKate gene library. (B) PCR reactions, using three different starting concentrations of the internal primers (4, 8, and 16 nM) and outermost primers at 400 nM, analyzed by agarose gel electrophoresis and the GeneRuler 100 bp Plus DNA ladder as the molecular marker.
Table 1. Comparison of Mutational Spectra and Mutation Rate Generated by the Spiked Genes and Three Different Enzymatic Mutagenesis Approaches

|                  | classic epPCR<sup>a</sup> | diversify (Clontech) | GeneMorph II (Stratagene) | Spiked Genes method | expected values<sup>b</sup> |
|------------------|---------------------------|----------------------|---------------------------|---------------------|---------------------------|
| Bias Indicators  |                           |                      |                           |                     |                           |
| Ts/Tv            | 1.81                      | 3.16                 | 1.01                      | 0.61                | 0.5                       |
| AT → GC/GC → AT  | 4.07                      | 8.70                 | 2.14                      | 0.88                | 0.908                     |
| A → N, T → N<sup>c</sup> | 80.2                      | 89.7                 | 68.2                      | 46.9                | 47.58                     |
| G → N, C → N<sup>c</sup> | 19.7                      | 10.3                 | 31.8                      | 53.1                | 52.38                     |
| Types of Mutations |                           |                      |                           |                     |                           |
| transitions<sup>c</sup> |                           |                      |                           |                     |                           |
| A → G, T → C     | 46.0                      | 69.2                 | 29.5                      | 17.7                | 15.86                     |
| G → A, C → T     | 18.4                      | 6.8                  | 20.9                      | 20.4                | 17.46                     |
| transversions<sup>c</sup> |                           |                      |                           |                     |                           |
| A → T, T → A     | 26.3                      | 17.1                 | 34.3                      | 13.6                | 15.86                     |
| A → C, T → G     | 7.9                       | 3.4                  | 4.4                       | 4.4                 | 15.86                     |
| G → C, C → G     | 0                         | 1                    | 4.3                       | 11.6                | 17.46                     |
| G → T, C → A     | 1.4                       | 2.5                  | 6.6                       | 21.1                | 17.46                     |
| insertions and deletions |                           |                      |                           |                     |                           |
| deletions         | 1                         | 8                    | 6                         | 19                  | 0                         |
| insertions        | 0                         | 0                    | 1                         | 1                   | 0                         |
| Data Analyzed     |                           |                      |                           |                     |                           |
| clones            | 18                        | 19                   | 23                        | 22                  |                           |
| sequenced bases   | 13 122                    | 13 395               | 16 038                    | 15 312              |                           |
| point mutations   | 76                        | 204                  | 184                       | 155                 |                           |
| Mutation Frequency (Mutations/kb) | 5.8 (4.5; 7.2) | 15.2 (13.2; 17.4) | 11.4 (9.8; 13.2) | 10.1 (8.6; 11.8) |
| exhibited<sup>d</sup> |                           |                      |                           |                     |                           |
| expected<sup>e</sup> | 6.6                       | 8                    | 16                        | 15                  |                           |

<sup>a</sup>Using the conditions reported by Cadwell and Joyce. References 20, 21<sup>b</sup> Expected values calculated from the proportion of each nucleotide in the emKate gene, assuming an equal mutation probability in every position. In percentage. <sup>c</sup>95% confidence intervals are shown between brackets, assuming that the mutation frequency follows a Poisson distribution. References 22, 23<sup>d</sup> Expected mutation frequency of each approach based on a previous report,<sup>e</sup> instruction manuals of commercial kits and the theoretical rates of Spiked Genes.

The spectrum, which evaluates the capacity of the method to produce all type of transition and transversion mutations. To measure these parameters while avoiding misrepresented data, we randomly selected several clones (independent of the phenotype displayed) from the antibiotic supplemented plate. A total of 57 867 bases were sequenced for this analysis (Table 1).

Mutation Distribution. The first point to highlight from the sequencing results was the homogeneous distribution of mutations generated along the entire gene (Figure 2). Using our Spiked Genes method, we did not observe accumulation or reduction of mutations at any location of the gene, even at the boundaries of the oligonucleotides used. Similar evenly distributed results were also obtained for the libraries created by the three enzymatic mutagenesis methods tested. Although, for the enzymatic methods, some bases were more prone to be replaced than others due to the nucleotide specificity of the DNA polymerases, which is dependent on the surrounding sequence.<sup>24</sup>

Mutation Rate. For our Spiked Genes method, the expected mutation rate was 15 changes/kb due to the 0.75% doping ratio per base per strand and therefore 1.5% per base-pair. The experimental results exhibited an average replacement rate of 10.3 point mutations per kb, which was slightly lower than the expected value, as shown in Table 1. This difference between the theoretical and experimental values reflects a selective pressure during the PCR assembly process to favor pairing of complementary primers with the least ratio of mismatches. The difference between the theoretical and experimental mutation rates would probably decrease using less stringent annealing conditions during the PCR step.

Because of the 1.0% experimental mutation rate, the mutagenesis of the 0.7 kb emKate gene (with our method) resulted in an average of seven nucleotide changes at the DNA level and four or five amino acid changes at the protein level, which is appropriate for performing experiments of directed evolution in most protein targets. However, if a lower mutagenic rate is desired for elucidating protein function–structure relationships, the doping ratio may be simply reduced during the oligonucleotide synthesis to achieve one or two amino acid changes per protein variant.

Such fine control in the mutagenic rate of the Spiked Genes method is hardly achievable by the enzymatic mutagenesis approaches. In our experiment, using the epPCR protocols reported in the literature produced, on an average, 5.8 replacements per kb (0.8 replacements less than the expected 6.6 changes/kb). However, the range of mutagenic rates with the epPCR methods is very broad.<sup>24</sup> Using similar mutagenic conditions, with different target genes and Taq DNA polymerases from different suppliers, we have observed mutation frequencies as high as 25 changes per kb (results not shown). None of the commercial epPCR mutagenesis kits tested could achieve the predicted value even when the mutagenesis conditions were set to reach the maximal mutation rate. Whereas the 4.2 mutations per kb rate of the GeneMorph II kit was lower than the expected mutation rate and the 7.2 mutations per kb rate of the Diversify kit surpassed the expected value (Table 1).
Mutational Diversity. The mutational spectrum is the most important parameter to assess the quality of mutagenesis libraries. The ideal mutagenesis method should be able to replace any base in the gene with any of the other three bases at the same frequency. Mutational diversity can be examined by different parameters.\textsuperscript{41,42} One is by analyzing the ratio of transitions (Ts) to transversions (Tv). Transition mutations are purine (G and A) to purine changes and pyrimidine (T and C) to pyrimidine changes, whereas transversions are changes from purine to pyrimidine and from pyrimidine to purine. In this regard, the theoretical transitions to transversions ratio is 0.5 because there are 4 possible transitions (A → G, G → A, C → T, and T → C) and eight possible transversions (A → C, C → A, T → A, A → T, C → G, G → C, T → G, and G → T). The analysis of the libraries constructed for this work showed that our Spiked Genes method, with a 0.61 Ts/Tv ratio, yielded the most even library. However, the classical and diversify epPCR libraries produced less diverse libraries, with Ts/Tv ratios of 1.81 and 3.16, respectively. In other words, the epPCR libraries produced 64.4 and 76.0% of transitions, although the expected value is 33.3%.

Another way to evaluate the mutational diversity is by calculating the ratio of AT → GC to GC → AT transition mutations (AT → GC/GC → AT ratio), which would be 1 for a perfectly unbiased mutagenic method targeting a gene with the same proportion of AT and GC pairs. As observed with the Ts/Tv ratio, Spiked Genes yielded the most even library, producing an experimental AT → GC/GC → AT ratio of 0.88 versus the expected 0.91 value calculated for the nucleotide composition of the emKate gene (A 26.13\%, T 21.45\%, G 27.72\%, C 24.66\%) (Table 1). Instead, the classical epPCR and diversify methods yielded GC-enriched libraries.

Finally, mutational diversity can also be assessed by measuring the ratio between the frequency of mutating A’s and T’s versus the frequency of mutating G’s and C’s (AT → NN/GC → NN ratio). This ratio should also be 1 for an unbiased mutagenic method that targets a gene with equal concentrations of ATs and GCs. For the Spiked Genes method, we measured the AT → NN value as 46.90\%, which was very close to the 47.58\% expected value for nucleotide composition of the emKate gene, whereas the GC → NN experimental value was 53.10\% as compared with the expected value of 52.38\%.\textsuperscript{21,22,43} For the enzymatic methods, GeneMorph II...
produced less biased results than classical epPCR or the Diversify kit that generated a higher $AT \rightarrow GC$ bias.

We dissected our results to analyze the individual mutations found in the coding strand, normalizing each type of mutation with respect to the wild-type nucleotide content of the gene. Figure 3 shows that the Spiked Genes method yielded the broadest mutagenesis spectrum among all methods, whereas the enzymatic methods had a poor representation of the transversions $C \rightarrow G$, $G \rightarrow T$, $G \rightarrow C$, $T \rightarrow G$, and $C \rightarrow A$, a result that is in concurrence with that reported by Alexander et al.44 for GeneMorph II. At the protein level, the nucleotide bias of enzymatic methods favors neutral amino acid replacements due to the protecting nature found in the degeneracy of the genetic code, leaving unexplored a large fraction of the sequence space. Again, the analysis of individual mutations reflected a clear advantage of the Spiked Genes method over that of the enzymatic-based approaches to produce more complete libraries of protein variants, as reported in the supplementary Table S2, wherein all of the amino acid replacements found in the four libraries were analyzed. Clearly, Spiked Genes created amino acid substitutions unachievable by the other methods.

Although nonbiased mutagenic libraries are mainly used to perform directed evolution studies, controlled-bias mutagenic libraries are occasionally preferred to favor certain amino acid mutations.36,45,46 Thus, it is important to have an adjustable method for setting specific transition/transversion bias to achieve the maximal coverage of desired amino acid substitutions. For this matter, the Spiked Genes method can be finely tuned to produce the desired mutation frequency and mutational bias by simply changing, during oligonucleotide synthesis, the proportion of the spiked bases. Moreover, if an important functional region is discovered after the first screening of the library, our method allows the mutational enrichment of this specific region by reusing some of the spiked oligonucleotides used in the initial assembly, which may be selectively incorporated by additional rounds of mutagenesis via overlapped PCR.47

Figure 3. Experimental/expected frequency ratios for each of the 12 possible nucleotide mutations. The pointed line indicates the ideal normalized value for every substitution.

Figure 4. emKate mutants generated by the Spiked Genes method. (A) E. coli colonies expressing wild-type GFP, a nonfluorescent GFP mutant containing the multiple mutations L64A/C65M/Y66G/G67V, wild-type emKate, and the representative fluorescent mutants found in the library. (B) Absorbance spectra of the mutants and reference proteins shown in panel A.
Together, all of the presented results confirm that the Spiked Genes method is a superior mutagenesis approach to produce libraries with evenly distributed random point nucleotide substitutions and affords an accurate statistical description of the library composition, contrary to the libraries built by epPCR approaches. Unfortunately, our method also showed a higher deletion rate than the enzymatic methods. It is well known that due to the inefficient coupling of the incoming nucleoside phosphoramidites and subsequent inefficient capping of the growing oligonucleotide chains that fail to be blocked synthetic oligonucleotides contain a high rate of single nucleotide deletions. The deletion rate can be reduced during synthesis by increasing the concentration of the nucleoside phosphoramidites to improve coupling yield at every nucleotide addition and using a double capping step to improve blocking. Deletion mutations can also be reduced by postsynthesis purification of the oligonucleotides by size exclusion chromatography, high-performance liquid chromatography, or polyacrylamide gel electrophoresis, as it has already been done to assemble complete virus genomes. These improvements will render higher quality libraries.

**Mutant Analysis.** After transforming *Escherichia coli* competent cells with only one-tenth of the ligation reaction, nonbiased random mutagenesis of the emKate gene using the Spiked Genes method produced 181 colored colonies out of 1995 colonies selected on kanamycin. From the analyzed colonies, 149, 16, 4, and 12 were red, green, yellow, and orange variants, respectively. As expected, one clone containing the wild-type emKate intact sequence emerged after sequencing the plasmid from 12 red colonies, whereas the rest of the red clones contained an average of 2 to 3 amino acid replacements and only 2 clones contained 6 and 8 changes. Most of the substitutions corresponded to neutral replacements.

The color variation observed in some of the mutant clones (Figure 4A) was the result of incomplete maturation of the DsRed-type chromophore present in the emKate protein. In some variants, the chromophore was trapped as a GFP-like immature intermediate (Figure S1). As a consequence, some mutants showed an orange phenotype that corresponded to a mixture of two proteins: one that remained in the GFP-like state, with an absorbance peak of approximately 470 nm, and another that reached the DsRed-like state, with an absorbance peak of approximately 580 nm. The simultaneous emission of green and red lights produced the orange phenotype. Proteins displaying slow and incomplete maturation of their chromophores have been extensively used as fluorescent timers, and perhaps some of our variants could also be used as timers.

The different levels of protonation of the chromophore also influenced the observed phenotypes, as revealed by the absorption properties of the green and yellow variants (Figure 4B), which exhibited different ratios of the GFP-type chromophore. In the case of the emKate gene library assembly, forward and reverse (shown in Table S1), were synthesized with a MerMade 192 DNA synthesizer (BioAutomation), using standard DMT-protected phosphoramidites dABz, dCAc, dGdmf, and dT. However, contrary to conventional synthesis, each nucleotide was doped with 0.25% of each of the other three nucleotides, producing an overall contamination of 0.75% per base. After deprotection and analysis by denaturing polyacrylamide gel electrophoresis, all oligonucleotides were observed as a single band and were used for the subsequent experiments without further purification.

**CONCLUSIONS**

In this study, we describe a new method to perform random mutagenesis that we call the Spiked Genes method, and we compared its efficiency against that of enzymatic epPCR mutagenesis methods. Using oligonucleotides with 0.75% mutant nucleotides at each position, we obtained a mutant library that (with respect to other methods) showed: first, a higher control of mutation rate; second, a wider and homogeneous mutation distribution along the entire gene; and third, a more even mutation diversity, with minimal mutation bias. Moreover, our method can be modulated to achieve the desired mutation and transition/transversion ratios. In this era of synthetic biology, the continuous drop in the price of oligonucleotides makes this random mutagenesis strategy a suitable option for the improvement of protein properties. The power of the strategy was manifested in the high phenotypic variations found in the library. We have created a rich library of emKate variant proteins that once characterized could potentially be used as reporters of cellular processes.

**METHODS**

**Oligonucleotide Synthesis.** All oligonucleotides used for the gene library assembly, forward and reverse (shown in Table S1), were synthesized with a MerMade 192 DNA synthesizer (BioAutomation), using standard DMT-protected phosphoramidites dABz, dCAc, dGdmf, and dT. However, contrary to conventional synthesis, each nucleotide was doped with 0.25% of each of the other three nucleotides, producing an overall contamination of 0.75% per base. After deprotection and analysis by denaturing polyacrylamide gel electrophoresis, all oligonucleotides were observed as a single band and were used for the subsequent experiments without further purification.
Library of Synthetic Genes Encoding emKate. The
library of synthetic genes encoding emKate (accession number
EU383029, DNA Data Bank of Japan) was assembled in a
single PCR step, as described in a previous report.20
Oligonucleotides in one strand were designed to overlap 30
nt with the oligonucleotides in the complementary strand.
To improve gene expression, the least frequent E. coli codons
were substituted with more frequently used codons. An overlapping
extension was carried out using an equimolar pool of the 22
internal oligonucleotides as templates (mFw2 to mFw12 and
mRv1 to mRv11) (Table S1). Three different equimolar
concentrations of the internal oligonucleotides were tested (4,
8, and 16 nM), plus 400 nM of the outermost primers (mFw1
and mRv12), which, for cloning purposes, also included
Ndel and XhoI restriction sites at the S’ and S’ ends of the synthesis
genes, respectively. mRv2 also contained the information to
introduce a polyhistidine tag –SGGSHHHHHH at the carboxy
end of the variant proteins. The final 100 μL of PCR mixture
had Vent DNA polymerase (New England BioLabs), buffer,
MgSO4, and dNTPs, as recommended by the supplier. The
PCR reaction was performed under the following conditions: 1
cycle: 94 °C for 3 min; 25 cycles: 94 °C for 1 min, 58 °C for 1
min, 72 °C for 1 min; 1 cycle: 72 °C for 5 min.
Both the gene library and the pJQO plasmid were double-
digested with the Ndel and XhoI restriction endonucleases for
12 h at 37 °C and purified from agarose gel, using the EZ-10
spin column PCR purification kit (Bio Basic Inc.). Next, 200 ng
of the gene library was ligated into 200 ng of the vector in a 20
μL reaction at 16 °C for 20 h. Using electroporation, 2 μL of
the ligation reaction was used to transform 100 μL of MC1061
E. coli cells. After recovering in 1 mL of LB media at 37 °C for 1
h, the transformed cells were grown at 37 °C for 20 h on LB
Kanamycin plates. After transformation, plasmids from several
independent colonies (fluorescent and nonfluorescent) were
isolated for DNA sequencing.
Enzymatic Mutagenesis. The wild-type gene emKate
was subjected to three different epPCR-based mutagenesis
approaches, two using commercial kits and one established in
our laboratory. We followed the instructions to obtain the
highest mutation rate recommended for two commercial kits:
16/kb for the GeneMorph II mutagenesis kit from Stratagene
and 8/kb for the Diversify PCR Random Mutagenesis Kit from
Clontech. As a third protocol, we evaluated the classical epPCR
approach at the predicted mutation rate of 6.6/kb, as described
by Cadwell and Joyce.20 All libraries generated from the three
ePCR methods were cloned, selected, and evaluated, as
previously described for the synthetic library.

**ASSOCIATED CONTENT**

 Supporting Information

The Supporting Information is available free of charge on the
ACS Publications website at DOI: 10.1021/acsomega.7b00508.

Spiked oligonucleotides synthesized to assemble the
eKate gene via the Spiked Genes method. Each wild-
type nucleotide was doped with 0.25% of each of the other
nucleotides (Table S1); experimental amino acid
replacements obtained with the four mutagenesis
methods tested over the eKate gene (Table S2); the
mechanism proposed for the post-translational matura-
tion of the red chromophore found in eKate protein
(Figure S1); alignment of sequences corresponding to
the mutants streaked in Figure 4 (Figure S2) (PDF)
540. Protein Eng., Des. Sel. Mutagenic DNA Polymerases to Generate Enzyme Variant Libraries. Fluorescent Protein. Nat. Biotechnol. N. G.; Palmer, A. E.; Tsien, R. Y. Improved Monomeric Red, Orange and Yellow Fluorescent Proteins Derived from Discosoma sp Red Fluorescent Protein. Nat. Biotechnol. 2004, 22, 1567–1572.

Subach, O. M.; Gundorov, I. S.; Yoshimura, M.; Subach, F. V.; Zhang, J.; Gruenwald, D.; Soutslova, E. A.; Chudakov, D. M.; Verkhusha, V. Y. Conversion of Red Fluorescent Protein into a Bright Blue Probe. Chem. Biol. 2008, 15, 1116–1124.

Lutz, S.; Patrick, W. M. Novel Methods for Directed Evolution of Enzymes: Quality, not Quantity. Curr. Opin. Biotechnol. 2004, 15, 291–297.

Koehavong, P.; Thilly, W. G. Fidelity of DNA Polymerases in DNA Amplification. Proc. Natl. Acad. Sci. U.S.A. 1989, 86, 9253–9257.

Eckert, K. A.; Kunkel, T. A. High Fidelity DNA Synthesis by the GeneMorph II Random Mutagenesis Kit. Methods Enzymol. 1995, 255, 589–603.

Zaccolo, M.; Williams, D. M.; Brown, D. M.; Gherardi, E. An Approach to Random Mutagenesis of DNA Using Mixtures of Triphosphate Derivatives of Nucleoside Analogues. J. Mol. Biol. 1996, 255, 818–8196.

Houtmann, P.; Roldan-Salgado, A. Elimination of Redundant and Stop Codons during the Chemical Synthesis of Degenerate Oligonucleotides. Combinatorial Testing on the Chromophore Region of the Red Fluorescent Protein mKate. ACS Synth. Biol. 2013, 2, 453–462.

Pletnev, S.; Shcherbo, D.; Chudakov, D. M.; Pletneva, N.; Merzlyak, E. M.; Wlodawer, A.; Dauter, Z.; Pletnev, V. A Crystallographic Study of Bright Far-Red Fluorescent Protein mKate Reveals pH-Induced cis-trans Isomerization of the Chromophore. J. Biol. Chem. 2008, 283, 28980–28987.

Shcherbo, D.; Merzlyak, E. M.; Chepurnykh, T. V.; Fradkov, A. F.; Ermakova, G. V.; Solovieva, E. A.; Lukyanov, K. A.; Bogdanova, E. A.; Zaraisky, A. G.; Lukyanov, S.; Chudakov, D. M. Bright Far-Red Fluorescent Protein for Whole-Body Imaging. Nat. Methods 2007, 4, 741–746.

Nov, Y. Probabilistic Methods in Directed Evolution: Library Size, Mutation Rate, and Diversity. Methods Mol. Biol. 2014, 1179, 261–278.

Verma, R.; Wong, T. S.; Schwaneberg, U.; Roccatano, D. The Mutagenesis Assistant Program. Methods Mol. Biol. 2014, 1179, 279–290.

Copp, J. N.; Hansson-Manful, P.; Ackerley, D. F.; Patrick, W. M. Error-Prone PCR and Effective Generation of Gene Variant Libraries for Directed Evolution. Methods Mol. Biol. 2014, 1179, 3–22.

Casson, L. P.; Manser, T. Evaluation of Loss and Change of Specificity Resulting from Random Mutagenesis of an Antibody VH Region. J. Immunol. 1995, 155, 5647–5654.

Alexander, D. L.; Lilly, J.; Hernandez, J.; Romsdahl, J.; Troll, C. J.; Camps, M. Random Mutagenesis by Error-Prone Pol Plasmid Replication in Escherichia coli. Methods Mol. Biol. 2014, 1179, 31–44.

Wong, T. S.; Roccatano, D.; Zacharias, M.; Schwaneberg, U. A Statistical Analysis of Random Mutagenesis Methods Used for Directed Protein Evolution. J. Mol. Biol. 2006, 355, 885–871.

Wong, T. S.; Roccatano, D.; Schwaneberg, U. Are transversion mutations better? A Mutagenesis Assistant Program analysis on P450 BM-3 heme domain. Biotechnol. J. 2007, 2, 133–142.

Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R. Site-Directed Mutagenesis by Overlap Extension Using the Polymerase Chain Reaction. Gene 1989, 77, 51–59.

Ellington, A.; Pollard, J. D. Introduction to the Synthesis and Purification of Oligonucleotides. In Current Protocols in Nucleic Acid Chemistry; John Wiley & Sons, Inc., 2001.

Young, L.; Dong, Q. Two-Step Total Gene Synthesis Method. Nucleic Acids Res. 2004, 32, No. e99.

Caruthers, M. H.; Barone, A. D.; Beauchage, S. L.; Dodds, D. R.; Fisher, E. F.; McBride, L. J.; Matteucci, M.; Stabinsky, Z.; Tang, J. Y. Chemical Synthesis of Deoxyoligonucleotides by the Phosphoramidite Method. Methods Enzymol. 1987, 154, 287–313.

Gaytan, P.; Contreras-Zambrano, C.; Ortiz-Alvarado, M.; Morales-Pablos, A.; Yanez, J. TrimerDimer: an Oligonucleotide-Based Saturation Mutagenesis Approach that Removes Redundant and Stop Codons. Nucleic Acids Res. 2009, 37, No. e125.

Yu, D.; Tang, J. Y.; Iyer, R. P.; Agrawal, S. Diethoxy N-Diisopropyl Phosphoramidite as an Improved Capping Reagent in the
Synthesis of Oligonucleotides Using Phosphoramidite Chemistry. *Tetrahedron Lett.* **1994**, *35*, 8565−8568.

(53) Smith, H. O.; Hutchison, C. A., 3rd; Pfannkoch, C.; Venter, J. C. Generating a Synthetic Genome by Whole Genome Assembly: phiX174 Bacteriophage from Synthetic Oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15440−15445.

(54) Terskikh, A.; Fradkov, A.; Ermakova, G.; Zaraisky, A.; Tan, P.; Kajava, A. V.; Zhao, X. N.; Lukyanov, S.; Matz, M.; Kim, S.; Weissman, I.; Siebert, P. "Fluorescent Timer": Protein that Changes Color with Time. *Science* **2000**, *290*, 1585−1588.

(55) Miyawaki, A.; Shcherbakova, D. M.; Verkhusha, V. V. Red Fluorescent Proteins: Chromophore Formation and Cellular Applications. *Curr. Opin. Struct. Biol.* **2012**, *22*, 679−688.

(56) Wiedenmann, J.; Ivanchenko, S.; Oswald, F.; Schmitt, F.; Röcker, C.; Salih, A.; Spindler, K. D.; Nienhaus, G. U. EosFP, a Fluorescent Marker Protein with UV-inducible Green-to-Red Fluorescence Conversion. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15905−15910.

(57) Subach, F. V.; Malashkevich, V. N.; Zencheck, W. D.; Xiao, H.; Filonov, G. S.; Almo, S. C.; Verkhusha, V. V. Photoactivation Mechanism of PAmCherry Based on Crystal Structures of the Protein in the Dark and Fluorescent States. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 21097−21102.

(58) Wang, Q.; Byrnes, L. J.; Shui, B.; Röhrig, U. F.; Singh, A.; Chudakov, D. M.; Lukyanov, S.; Zipfel, W. R.; Kotlikoff, M. I.; Sondermann, H. Molecular Mechanism of a Green-Shifted, pH-Dependent Red Fluorescent Protein mKate Variant. *PLoS One* **2011**, *6*, No. e23513.

(59) Brown, L. D.; Cai, T. T.; DasGupta, A. Interval Estimation for a Binomial Proportion. *Statistical Science* **2001**, *16*, 101−133.