ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

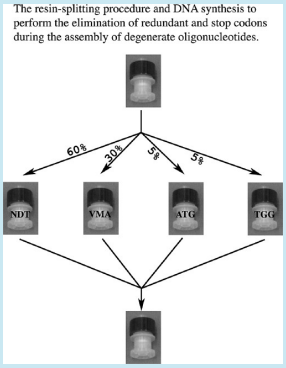
Elimination of Redundant and Stop Codons during the Chemical Synthesis of Degenerate Oligonucleotides. Combinatorial Testing on the Chromophore Region of the Red Fluorescent Protein mKate

Paul Gaytán* and Abigail Roldán-Salgado

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

Supporting Information

ABSTRACT: Although some strategies have been reported for the elimination of stop and redundant codons during the chemical synthesis of degenerate oligonucleotides, incorporating an expensive cocktail of 20 trimer-phosphoramidites is currently a commonly employed and straightforward approach. As an alternative option, we describe here a cheaper strategy based on standard monomer-phosphoramidites and a simplified resin-splitting procedure. The accurate division of the resin, containing the growing oligonucleotide, into four columns represents the key step in this approach. The synthesis of the degenerate codon NDT in column 1, loaded with 60% of the resin, produces 12 codons, while a degenerate codon VMA in column 2, loaded with 30% of the resin, produces 6 codons. Codons ATG and TGG, independently synthesized in columns 3 and 4, respectively, and loaded with 5% each, completes the 20 different codons. The experimental frequency of each mutant codon in the library was assessed by randomizing 12 contiguous codons that encode for amino acids located in the chromophore region of the enhanced red fluorescent protein mKate-S158A. Furthermore, randomization of three contiguous codons that encode for the amino acids Phe62, Met63, and Tyr64, which are equivalent to Phe64, Ser65, and Tyr66 in GFP, gave rise to some red and golden yellow fluorescent mutants displaying interesting phenotypes and spectroscopic properties. The absorption and emission spectra of two of these mutants also suggested that the complete maturation of the red and golden yellow chromophores in mKate proceeds via the formation of a green-type chromophore and a cyan-type chromophore, respectively.



KEYWORDS: *degenerate oligonucleotides, random mutagenesis, directed evolution, fluorescent proteins, mKate, protein engineering*

Several methods have been developed to perform the directed evolution of proteins either by the enzymatic mutagenesis of the encoding genes or by amplification from synthetic oligonucleotides.[1] Enzymatic methods are predominantly based on error-prone PCR,[2] producing random single base changes throughout a gene sequence. A similar effect directed at a focused region is achieved with libraries of spiked oligos, where each wild-type nucleotide is partially replaced with the other three bases during the chemical assembly.[3−5] Unfortunately, a large number of amino acid (aa) replacements do not occur in these approaches, such as those requiring 2−3 base pair changes per codon.[6] For example, methionine is not significantly mutated to Ala, Asp, Asn, Cys, Glu, Gln, Gly, His, Phe, Pro, Ser, Trp, or Tyr.

Covering all of the possible amino acid combinations in a focused protein region is normally achieved using saturation mutagenesis methods that produce pools of degenerate oligonucleotides. NNS and NNK approaches are widely used and commercially available options that generate mixtures of 31 sense codons and one stop codon, encoding the 20 amino acids at different frequencies because of the genetic code degeneracy (N represents any of the four bases, S = G or C, and K = G or T). As a consequence, protein libraries created with ordinary degenerate oligos are highly biased toward those amino acid sequences encoded by redundant codons.[7]

The straightforward approach to generate high quality protein libraries that lack stop and redundant codons[8] resorts to the chemical incorporation of trimer-phosphoramidites (TPs) cocktails during the traditional chemical synthesis of oligonucleotides.[9] TPs may encode for all 20 amino acids or a subset of them. However, these reagents are expensive because they are normally obtained in low overall yields and because their preparation requires many synthetic and purification steps in a conventional liquid phase. Furthermore, the proportion of each TP in the cocktail must be carefully adjusted to produce an even frequency of the 20 codons because of the differential reactivities among them. Thus, the saturation of five codons with a commercial cocktail containing 20 TPs may cost at least $386 USD, which only considers the price of the cocktail.

Krumpe et al. demonstrated that this trinucleotide approach significantly enhances the functional diversity of a 12-mer peptide library when directly compared with the same library assembled via the conventional NNK approach.[10] TPs are so promising that some research groups are interested in improving their synthesis, as was recently reviewed by Müller et al.[8]

Received: December 11, 2012
Published: February 6, 2013

A combination of five DMTr-protected trimer-phosphoramidites and five Fmoc-protected dimer-phosphoramidites appears to be a cheaper option but is not yet commercially available.[11]

Another strategy to eliminate stop and redundant codons at a minor cost, but with more work, requires the utilization of 11 dimer-phosphoramidites (DPs) in combination with the resin-splitting approach.[12] In this method, the resin containing the growing oligonucleotide is separated into four columns, and sets of five degenerate codons are synthesized in each column to complete 20 codons. After the synthesis, the resins are recombined and separated again to receive a new set of degenerate codons. DPs are less expensive than TPs because their preparation requires fewer synthetic and purification steps. Unfortunately, DPs also exhibited different reactivities during solid-phase synthesis, and their proportions in the cocktails must be adjusted to produce a balanced frequency of the resultant codons.

Although the resin-splitting approach was originally created to reduce the complexity of degenerate oligonucleotide-encoding peptide libraries[13] and to spike a DNA stretch with complete mutant codons,[14] it has also been used in combination with standard monomer-phosphoramidites (MPs) to saturate a short DNA region excluding redundant and stop codons.

In contrast to TPs and DPs, the standard MPs of dA, dC, dG, and dT display similar reactivities during solid-phase synthesis[15] and are much more inexpensive. For example, one kit is only $36 USD and is sufficient to prepare four 75-mer degenerate oligonucleotides, randomizing from 1 to 60 internal codons.

In the "multi-line split DNA synthesis", Tabuchi et al.[16] separated the growing oligonucleotide into three columns and employed nine cocktails of MPs, mixed in different molar ratios. Each position of the three degenerate codons was assembled with a different cocktail. Although no stop codons were expected, the theoretical frequency of sense codons was highly biased toward Ala (8.9%) and poorly toward Tyr (1.0%), a worse result compared to the simple NNS or NNK approach.

In the "mix-split-mix" approach, Yin et al.[17] split the growing oligonucleotide into 10 columns, synthesizing sets of two different codons in nine of the columns and TGG (Trp) in the tenth column to exclude redundant, stop, and cysteine codons from the library. Considering that the starting resin for the most common 0.2-μmol scale of synthesis is approximately 5 mg (using ordinary resins with a porosity of 500 or 1000 Å), the accurate separation of 0.5 mg fractions is difficult, and at this level of material, any loss becomes significant in the generation of biased libraries.

To use only custom oligonucleotides, not requiring TPs, DPs, or the resin-splitting approach, Hughes et al. described MAX randomization, a mutagenic approach that requires only mixtures of specific oligos to remove redundant and stop codons via hybridization selection with an oligo template that is completely randomized.[7] This process appears to be interesting and easy to implement but requires many oligonucleotides and is limited to alternate codons or only two contiguous codons in the template.

More recently, Tang et al.[18] described a very interesting approach to avoid redundant and stop codons during the saturation mutagenesis of one site, using four custom oligos mixed in the appropriate ratio to represent each of the 20 amino acids evenly with only one codon. The first oligonucleotide contained the degenerate codon NDT (N = A, C, G, or T and D = A, G ,or T), which produces 12 codons: AAT (Asn), GAT (Asp), CAT (His), TAT (Tyr), AGT (Ser), GGT (Gly), CGT (Arg), TGT (Cys), ATT (Ile), GTT (Val), CTT (Leu), and TTT (Phe). The second oligonucleotide contained the degenerate codon VMA (V = A, C or G and M = A or C), yielding six codons: AAA (Lys), CAA (Gln), GAA (Glu), ACA (Thr), CCA (Pro), and GCA (Ala). The third and fourth oligonucleotides contained the specific codons ATG and TGG, which encode for Met and Trp, respectively. This mutagenesis approach is highly attractive for the saturation of one site or distantly separated sites but is not appropriate for the saturation of two or more contiguous sites. For example, the saturation of two contiguous codons required the synthesis of 16 oligonucleotides to generate all possible single and double mutants,[18] whereas the randomization of three, four, or five contiguous sites would have required the synthesis of 64, 256, or 1024 oligos, respectively, increasing the experimental costs to higher levels than using TPs.

Because each oligonucleotide set contained the same 3′ and 5′ flanking regions adjacent to the target codon, we envisaged a combination of "resin splitting" and synthesis of the codons NDT, VMA, ATG, and TGG in each of the four separated fractions to reduce the synthesis of 4^n oligos to the synthesis of only one randomized oligonucleotide when n codons, contiguous or alternated, need to be saturated.

Complete assessment of this mutagenesis approach was performed on the chromophore region of the enhanced far-red fluorescent protein mKate-S158A.[19]

RESULTS AND DISCUSSION

Taking into consideration the work recently reported by Tang et al.[18] and our accessibility to a four-column DNA synthesizer equipped with eight vials for phosphoramidites, four for the clear bases dA, dC, dG, and dT and four for the mixtures N, D, V, and M (Figure 1a), we devised a resin-splitting approach to perform saturation mutagenesis of one or more codons, contiguous or alternated, that at the same time excluded stop and redundant codons. This approach was termed ERASC to highlight the elimination of redundant and stop codons during the synthesis of degenerate oligonucleotides.

As shown in Figure 1b, the key step in ERASC was the separation of the resin containing the growing oligonucleotide into four columns. However, to reduce errors involved in weighing minute amounts of material, the resin was resuspended in the viscous, dense solvent ethylene glycol, and the appropriate volume was pipetted to each of the columns as described in Methods. Better resuspension results were obtained with resins of a 2000 Å porosity, whose density is lower than that of ordinary resins. In addition, 2000 Å resins are more suitable for the synthesis of long oligonucleotides, as are those reported here. Column 1 received 60% of the resin and synthesized the degenerate codon NDT, and column 2 received 30% of the resin and synthesized the degenerate codon VMA. Columns 3 and 4 received 5% of the resin and synthesized ATG and TGG, respectively. When the synthesis in the four columns ended, the resins were recombined, resuspended, and split as before. The mutagenic cycle was repeated as many times as target codons to saturate.

Biological Model. As a protein model to test the degenerate oligonucleotides assembled by ERASC, we chose "emKate",[19] an enhanced version of the red fluorescent protein mKate[20] that contains the amino acid substitution S158A. emKate was selected for the mutagenesis because it is a less

dx.doi.org/10.1021/sb3001326 | ACS Synth. Biol. 2013, 2, 453−462

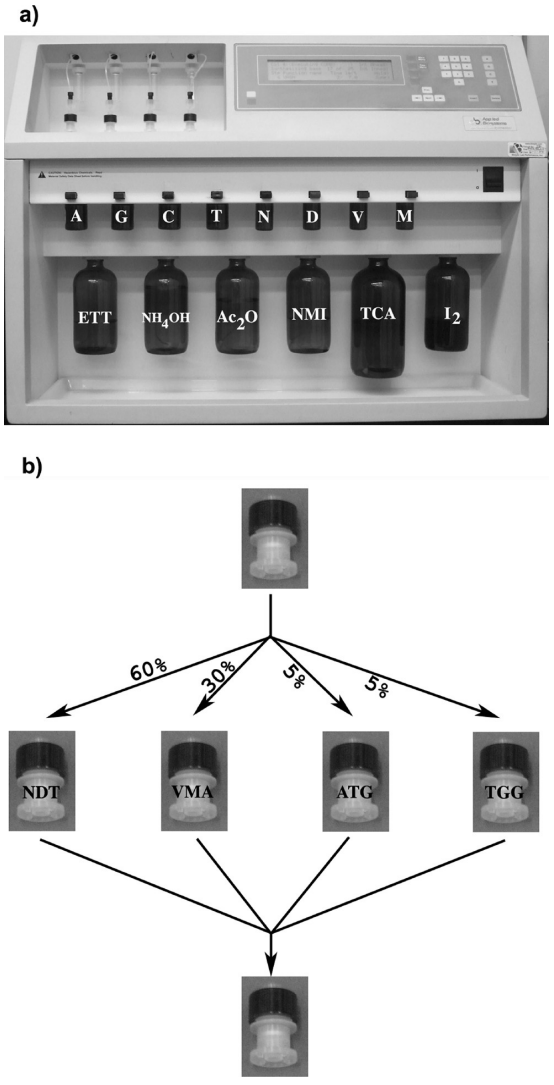


Figure 1. ERASC approach. (a) The DNA synthesizer implemented for the assembly of degenerate oligonucleotides via the ERASC approach, using standard β-cyanoethyl phosphoramidites of dA, dG, dC, dT, and the mixtures N (dA/dG/dC/dT), D (dA/dG/dT), V (dA/dC/dG), and M (dA/dC). The other solutions contained 5-ethylthiotetrazole (ETT) as activating reagent, acetic anhydride (Ac_2O) as capping A, *N*-methylimidazole (NMI) as capping B, trichloroacetic acid (TCA) as detritylating reagent, iodine (I_2) as oxidizing reagent, and ammonium hydroxide (NH_4OH) for final deblockage. Acetonitrile, not visible in the figure, was used as the washing solvent. (b) Cycle of mutagenesis via the resin-splitting approach, which comprises three steps: (1) separation of the resin into four columns, (2) synthesis of the degenerate codons NDT and VMA and the specific codons ATG and TGG into separate columns, and (3) recombination of the resins into one column. The cycle can be repeated if more codons are being randomized or an ordinary synthesis can be resumed in one column to add the 5′ flanking region.

studied fluorescent protein than GFP, is monomeric and more brilliant than the original protein mKate, and emits light in the far-red region of the visible spectrum (Ex 587 nm, Em 621 nm), which is an interesting property for applications in multilabeling techniques and for the potential imaging of deep tissues. emKate's chromophore is generated in a posttranslational cyclization reaction between the amino acids Met63, Tyr64, and Gly65 (equivalent to amino acids Ser65, Tyr66, and Gly67 in GFP) and a second oxidation step that generates an acylimine moiety integrated into the peptide backbone between Phe62 and Met63.

Because the *emKate* gene was used as the template for the subsequent mutagenesis experiments, it was enzymatically assembled as described in Methods[21] using the 24 oligonucleotides shown in Supplementary Figure 1S. Taking advantage of the full synthesis of the gene, we replaced rare *E. coli* codons with high frequency ones according to the codon usage of this bacterium, and a poly histidine tail coding sequence was added at the 3′ end of the gene for purification purposes. After 20 h of growth (37 °C), colonies expressing emKate displayed a purple phenotype at ambient light and a red fluorescent phenotype when illuminated with blue light. The correct assembly of the gene was confirmed by sequencing the plasmid isolated from one of these colonies.

Assessment of the Mutation Frequency Generated by ERASC. For testing the experimental codon diversity generated by ERASC, the degenerate primer 12COD (5′ GGTG-GCCCTTTGCCCTTTGCCTTCGAC $(\mathbf{XXX})_{12}$ TTCA-TTAATCATACCCAGGGCATC 3′) was synthesized to randomize 12 contiguous codons that encode for the amino acids I57-T68. In this primer, XXX represents the codon mixture created by ERASC, being surrounded by the 5′ and 3′ flanking regions included for annealing purposes and subsequent PCR amplification. 12COD and the partial complementary primer PCP-12COD (5′ GTCGAAGGCA-AAGGGCAAAGGGCCACC 3′) were used to create the library I57-T68 via an overlap PCR approach,[22] as shown in Supplementary Figure 2S and described in Methods, using *emKate* as template. After sequencing the plasmid isolated from 31 nonfluorescent clones, the randomized region was analyzed (Table 1); the mutant codons are sorted out into their degenerate codon of origin in Table 2. Supported by these results, the following conclusions were drawn:

1. All expected codons were generated during the diversification process, with only one codon representing each amino acid and no stop codons.

2. Glutamate encoded by GAA was the least represented amino acid (2.96%), whereas phenylalanine encoded by TTT was the most represented (8.33%). This result is an acceptable bias from the theoretical 5% mutation frequency per codon when considering the small size of the sample analyzed. However, it was clear that the codons generated by NDT produced 67.5% of the codons and VMA produced 21.5%, whereas the expected frequencies were 60% and 30%, respectively. This difference between the theoretical and experimental frequency indicates a possible manipulation error when the resin for the synthesis of VMA was transferred to column 2. However, this problem is easily resolved by the exhaustive mixing of the stock resin before any transference to the other columns.

3. The independently synthesized codons ATG and TGG were obtained in almost the expected frequency because columns 3 and 4 were the first to receive the separated resin when the stock was freshly resuspended.

4. As originally reported,[15] the four monomer-phosphoramidites exhibited similar reactivities. The addition of N produced 53 dAs, 71 dCs, 63 dGs, and 63 dTs; D produced 73 dAs, 73 dGs, and 105 dTs; V produced 25 dAs, 29 dCs, and 26 dGs; and finally M produced 40 dAs and 40 dCs. Therefore, this result guarantees an almost even generation of the mutant codons.

dx.doi.org/10.1021/sb3001326 | *ACS Synth. Biol.* 2013, 2, 453−462

Table 1. Partial Sequence of 31 Non-fluorescent Mutants from the I57-T68 Library, Which Was Created Using the Degenerate Primer 12COD[a]

emKate	ATC	CTG	GCT	ACC	AGT	TTT	atg	tat	ggc	AGC	AAA	ACG
1	GAT	AAT	GGT	CTT	TTT	CGT	AGT	GTT	CGT	TAT	ATT	GAT
2	AAA	TGG	ATG	CAA	GAT	ATG	TAT	GTT	GCA	TAT	CCA	AAA
3	GCA	GAT	GTT	TAT	CTT	CAA	CCA	TTT	AGT	ATG	TTT	AAA
4	CCA	AAT	ACA	CAT	TGG	CTT	TGT	GCA	AGT	GAT	GAT	AGT
5	TAT	CGT	CAT	TGT	CTT	CTT	TGT	ACA	AAA	CAA	CTT	CGT
6	TTT	CGT	GTT	TTT	TGT	TGT	GGT	AAA	CTT	CTT	CAA	CCA
7	CCA	ACA	ATT	CGT	ATT	CTT	GTT	GTT	AAA	ATT	GAT	GCA
8	GGT	CAT	GGT	GTT	ATT	TTT	AGT	AGT	AGT	GCA	ATT	GAT
9	CGT	CAA	CGT	GAT	TTT	TGG	ATT	CAT	TGT	GTT	CAT	TTT
10	ATG	TGG	ATG	CAT	ATT	CAT	GGT	ACA	ATG	AGT	CAT	GAA
11	TTT	AAT	TAT	TTT	TGT	GAT	CGT	CGT	ACA	AAT	CTT	CAT
12	CAA	AGT	CAA	GA*	GAT	TGT	AGT	CGT	GAT	CAT	AAA	TTT
13	CGT	TTT	TGT	AAT	ACA	ACA	ACA	GAA	CAA	GCA	TGT	CTT
14	AAA	AAT	CGT	GTT	GGT	GGT	GGT	TAT	GAA	ACA	CCA	TGG
15	GGT	AGT	GAT	CAT	ACA	ATG	GGT	CCA	ATG	CAA	GTT	TAT
16	TGT	CAA	TTT	GGT	GTT	GAT	CTT	CAT	GAA	CAT	CTT	AaT
17	CAA	GGT	GTT	TAT	ATG	TTT	TGT	CTT	CAA	CAA	GAT	GGT
18	ATT	GCA	AAA	TTT	CAT	AAA	ATG	GTT	TTT	ATG	CAT	TAT
19	ATT	TGT	GTT	ATT	CAA	CCA	ATG	TTT	GAT	AAT	ATG	CGT
20	GGT	ACA	GTT	CCA	GAT	GGT	CTT	AAA	CGT	GTT	ACA	TGG
21	AAT	CTT	GAA	ATG	TTT	GGT	CAT	ATT	GCA	TGG	TGG	TA*
22	TGT	AAT	TTT	CAT	CGT	CTT	ATT	TTT	CAT	TTT	ATT	ACA
23	ATG	TGG	TGT	AGT	GGT	GAA	TTT	CTT	GCA	CAA	GTT	TTT
24	TAT	ATG	CTT	GTT	AGT	CTT	AGT	CAT	CTT	TGG	AAT	ATG
25	CGT	GAA	ATG	TGG	TTT	ATT	TAT	GTT	GCA	TGG	ATT	TAT
26	TAT	TGG	CGT	CCA	GAA	TGG	CAT	ATT	AGT	TGG	TTT	AGT
27	GCA	GAT	GCA	TTT	TAT	GTT	CTT	AAA	CTT	CTT	TGG	ATT
28	CAA	CAT	AGT	GAT	CTT	ATT	TTT	ATT	CAT	TGT	CCA	TTT
29	AGT	GTT	CAT	TTT	AGT	TGT	GAA	CAA	GTT	GCA	ATT	GCA
30	TTT	ATG	ATT	CAT	GAA	AGT	CTT	CAT	CTT	GGT	AAT	ATG
31	CTT	GAA	TGG	TTT	GTT	GCA	GAT	GAT	CCA	ATG	TGG	GTT

[a]Only the sequence of the randomized region is shown. The wild-type sequence is indicated in bold. The asterisks represent single nucleotide deletions.

5. Only two single nucleotide deletions were observed, representing a deletion rate of 0.18% per base. This is a better result than that previously reported (0.7%) for long oligonucleotides synthesized over more crowded 1000 Å resins.[23] Steric effects during the growth of oligonucleotides likely reduce either the coupling yield of phosphoramidites or the capping efficiency, contributing to the generation of deletions.

6. 12COD was expected to produce a population of 4×10^{15} (20^{12}) variants, a huge number compared to the practical transformation efficiency of 10^7 transformants from ligation reactions. Therefore, finding a functional mutant in library I57-T68 is statistically impossible considering that only 2.44×10^{-7}% of the total population may be transformed and that most of these variants contain multiple changes relative to the parental gene. Many simultaneous changes even in a short DNA stretch normally destroy the function of the encoded protein. In agreement with this analysis, out of the 50,000 colonies plated and analyzed on a blue light transilluminator, none was fluorescent.

Functional Assessment. Knowing that Gly67 is essential for the formation of the chromophore in GFP, whereas Tyr66, Ser65, and Phe64 tolerate some amino acid substitutions,[24] we decided to maintain Gly65 constant and to randomize the equivalent amino acids Tyr64, Met63, and Phe62 in the red fluorescent protein emKate to study the relationship between its chromophore composition and fluorescence.

Randomization of the three contiguous amino acids required the synthesis of the degenerate primer 3COD (5′ TTTGC-CTTCGACATCCTGGCTACCAGT $(\mathbf{XXX})_3$ GGCAG-CAAAACGTTCATTAATC 3′), which in combination with the partial complementary primer PCP-3COD (ACTGG-TAGCCAGGATGTCGAAGGCAAA) gave rise to the gene library F62-Y64 using e*mKate* as template. When *E. coli* was transformed with 1/15 of the ligation reaction and the transformants were grown on several plates for 20 h (37 °C), approximately 10,000 colonies emerged, most being non-fluorescent. Only 19 colonies displayed a red fluorescent phenotype, whereas 36 exhibited a golden yellow fluorescent phenotype when exposed to blue light.

The sequencing of plasmids isolated from 13 red clones and 8 golden clones revealed the amino acid substitutions shown in Table 3 and the following observations:

1. Only certain amino acid combinations in the chromophore produce fluorescent proteins, as was confirmed by the repetitive finding of some variants. For example, emKate and the mutants emKate_F62M, emKate_F62C/M63Q and emKate_F62C/Y64W were found twice, whereas emKate_F62L and emKate_Y64W were found three times. Figure 2 shows *E. coli* streaks harboring a member of the different mutants

dx.doi.org/10.1021/sb3001326 | *ACS Synth. Biol.* 2013, 2, 453−462

Table 2. Experimental Frequency of Mutant Codons Generated via the ERASC Approach[a]

codon (aa)	found codons	frequency per codon (%)	frequency per subgroup
AAT (N)	12	3.22	NDT 251 codons (67.5%)
CAT (H)	25	6.72	
GAT (D)	21	5.65	
TAT (Y)	15	4.03	
AGT (S)	20	5.38	
CGT (R)	18	4.84	
GGT (G)	18	4.84	
TGT (C)	17	4.57	
ATT (I)	22	5.91	
CTT (L)	28	7.53	
GTT (V)	24	6.45	
TTT (F)	31	8.33	
AAA (K)	12	3.22	VMA 80 codons (21.5%)
CAA (Q)	17	4.57	
GAA (E)	11	2.96	
ACA (T)	13	3.49	
CCA (P)	12	3.23	
GCA (A)	15	4.03	
ATG (M)	21	5.64	5.64%
TGG (W)	18	4.84	4.84%

[a]These codons were found in the I57-T68 library, which was created using the degenerate primer 12COD.

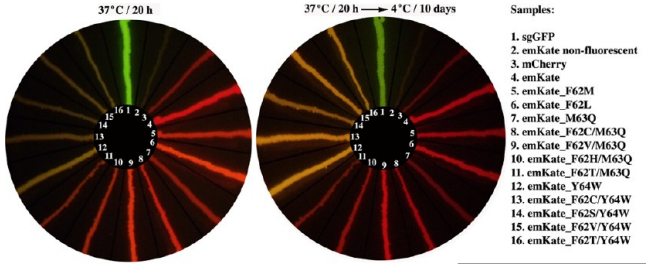


Figure 2. Bacterial streaks of *E. coli* containing the different fluorescent mutants reported in the present work, which were grown on an LB/Km plate. This plate was analyzed after 20 h of incubation at 37 °C and then after 10 days of storage at 4 °C. sgGFP was originally reported as sg25. Nonfluorescent emKate corresponds to one of the mutants isolated from the I57-T68 library. The plates were illuminated over a blue light transilluminator and photographed using the orange filter supplied with the equipment.

isolated in this work. The most fluorescent red and golden mutants were emKate_F62L and emKate_Y64W, respectively.

It is worth mentioning that a library spanning a region of three codons must be composed of 8000 variants if each position is randomized with 20 codons. Therefore, 23,964 clones must be analyzed to complete the screening with 95% coverage. In our case, only 42% of this population was screened because the most fluorescent variants were repeated three times, and we reasoned that no better mutants would be found.

2. All red clones contain tyrosine in position 64, whereas all golden clones contain tryptophan. Because of the substitution Y64W, golden mutants will rarely be obtained by enzymatically directed evolution because any of the tyrosine codons (tat or tac) requires two contiguous nucleotide replacements to change to the tryptophan codon (tgg), an unlikely result even under hypermutagenic conditions.[25] However, because the indole group found in tryptophan is larger than the phenol group of tyrosine, the substitution Y64W crowds the core of the protein and affects its stability, producing a 4-fold reduction in the expressed protein relative to the parental protein, as demonstrated by SDS-PAGE analysis of the total soluble extracts (Figure 3).

3. Red clones tolerate only the wild-type amino acid methionine in position 63 or the mutant amino acid glutamine, which may be essential for the formation of the chromophore.[26] However, the polar substitution M63Q notably reduced the expressed protein (Figure 3, lane 5), and additional substitutions at position 62 were necessary to recover the expression level, as shown in lanes 6−9 of Figure 3. In line with these results, Gln65 (equivalent to 63 in mKate) was the wild-type amino acid found in the first reported red fluorescent protein DsRed,[27] and its widely known mutant mCherry contains the reverse mutation, Q65M.[28]

4. Golden mutants tolerated only the wild-type amino acid methionine at position 63. This result is in agreement with the

Table 3. Fluorescent Functional Mutants Isolated from the F62-Y64 Library, Which Was Created Using the Degenerate Primer 3COD

sample	AA62 (codon)	AA63 (codon)	AA64 (codon)	repetitions[a]	Ex_{max} (nm)	Em_{max}[b] (nm)	phenotype
mKate	F (TTT)	M (ATG)	Y (TAT)		586	616 (0.52)	red
emKate	F (TTT)	M (ATG)	Y (TAT)	2	582	613 (0.36)	
emKate_F62M	**M** (ATG)	M (ATG)	Y (TAT)	2	579	613 (0.12)	
emKate_F62L	**L** (CTT)	M (ATG)	Y (TAT)	3	579	609 (0.09)	
emKate_M63Q	F (TTT)	**Q** (CAA)	Y (TAT)	1	573	614 (0.06)	
emKate_F62C/M63Q	C (TGT)	**Q** (CAA)	Y (TAT)	2	577	609 (0.09)	
emKate_F62 V/M63Q	**V** (GTT)	**Q** (CAA)	Y (TAT)	1	569	605 (0.14)	
emKate_F62H/M63Q	**H** (CAT)	**Q** (CAA)	Y (TAT)	1	576	605 (0.11)	
emKate_F62T/M63Q	T (ACA)	**Q** (CAA)	Y (TAT)	1	573	603 (0.10)	
emKate_Y64W	F (TTT)	M (ATG)	**W** (TGG)	3	479	558	golden yellow
emKate_F62C/Y64W	**C** (TGT)	M (ATG)	**W** (TGG)	1	479	564	
emKate_F62S/Y64W	**S** (AGT)	M (ATG)	**W** (TGG)	2	472	550	
emKate_F62 V/Y64W	**V** (GTT)	M (ATG)	**W** (TGG)	1	473	559	
emKate_F64T/Y64W	**T** (ACA)	M (ATG)	**W** (TGG)	1	480	553	

[a]Some mutants were found two or three times. [b]Red proteins excited at 474 nm generated two emission peaks: a green peak centered at approximately 510 nm and a red peak between 603 and 616 nm. The values reported in parentheses represent the relative ratio of the red peak compared to the green peak.

dx.doi.org/10.1021/sb3001326 | *ACS Synth. Biol.* 2013, 2, 453−462

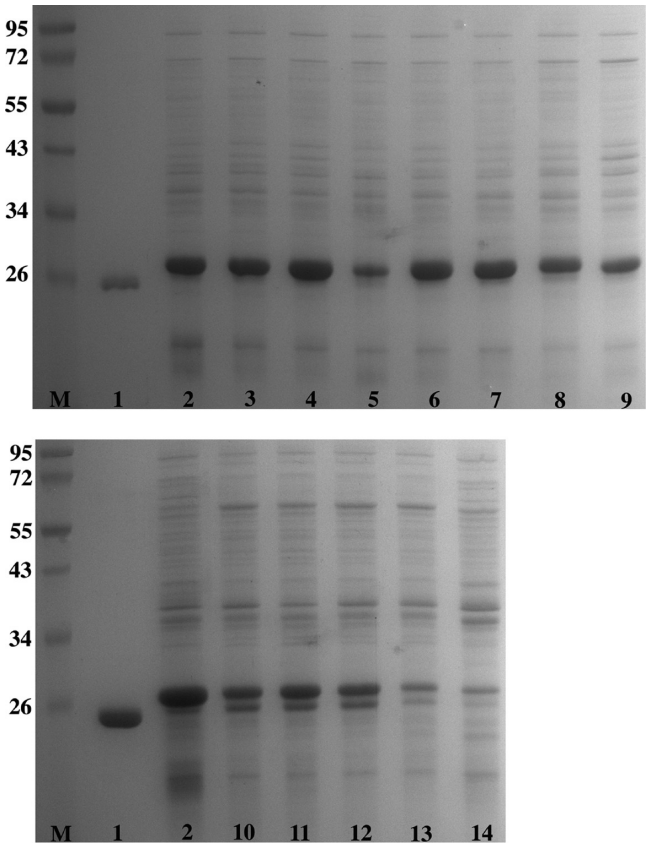


Figure 3. SDS-PAGE analysis of total soluble protein extracts. The top gel was loaded with mutants displaying a red phenotype, and the bottom gel with mutants displaying a golden yellow phenotype. The protein ladder and the reference proteins sgGFP and emKate were loaded in both gels. Molecular weights are indicated at the left of the first lane. M: protein ladder, 1: sgGFP pure, 2: emKate, 3: emKate_F62M, 4: emKate_F62L, 5: emKate_M63Q, 6: emKate_F62C/M63Q, 7: emKate_F62 V/M63Q, 8: emKate_F62H/M63Q, 9: emKate_F62T/M63Q, 10: emKate_Y64W, 11: emKate_F62C/Y64W, 12: emKate_F62 V/Y64W, 13: emKate_F62S/Y64W, 14: emKate_F62T/Y64W. Before being loaded on the gel, the samples were mixed with denaturing loading buffer and boiled for 5 min; 1/100 of the total extract per sample was analyzed.

chromophore's composition of the fluorescent protein mHoneydew,[28] which evolved from DsRed and, therefore, contains the double amino acid replacement Q65M/Y66W. We do not know yet if this sequence restriction is due to steric effects in the core of the protein or because only Met63 can ensure the appropriate maturation of the new chromophore. A crystal structure would likely provide more information about this sequence restriction.

5. Phe62 tolerated many amino acid replacements in both the red and golden clones. This amino acid was more permissive to amino acid substitutions because only its carbonyl group becomes part of the chromophore and not its side chain. However, whereas the substitution F64L notably improved the protein expression of GFP at 37 °C,[29] the equivalent replacement in emKate produced only a slight increase relative to the parental protein, as can be observed in lanes 2 and 4 of Figure 3. In general terms, the SDS-PAGE analysis of total soluble extracts demonstrated that Phe62 tolerates conservative amino acid replacements with similar levels of expressed protein, whereas it is very sensitive to polar amino acid substitutions, which significantly reduce the soluble protein in either the red or golden mutants, as can be observed in lanes 5, 6, 7, 8, and 9 of Figure 3.

6. In a similar manner to mKate, emKate, and the well-known monomeric protein sgGFP, all mutants generated in this work migrated as monomeric proteins when analyzed by SDS-PAGE under semidenaturing conditions (data not shown). In fact, no significant changes were expected in the oligomerization nature of the mutants because the amino acid substitutions took place in the core of the protein and not on the surface.

Fluorescence Characterization of Mutants. Because some red mutants, either *E. coli* streaks or protein extracts, exhibited different hues with visible light and because some appeared more fluorescent than the parental emKate protein when exposed to blue light, we expected a significant change in the maximum emission of these mutants relative to the parental protein. However, when these mutants were excited at 474 nm, the emission spectrum for each of them produced two peaks, at approximately 510 and 610 nm, including the parental protein emKate and the original mKate protein, as shown in Figure 4 (the reference protein mKate was constructed from emKate by reverting the mutation S158A via site-directed mutagenesis). These results were unexpected because (1) the green peak at 510 nm was not reported in the original research,[20] and (2) the maximum emission for the red peak of emKate and mKate was blue-shifted 8 and 19 nm, respectively, to the maxima reported.[19,20] This difference in emission was likely due to our particular experimental conditions; however, the maximum excitation and emission peaks, which were also determined for the reference proteins mCherry[28] and sgGFP,[30] closely matched the previously reported data. Furthermore, because the 510 nm peak accounts for an incomplete mature chromophore remaining in the intermediate GFP-type green state[26] and the 610 nm peak for the complete mature red chromophore, the different red hues observed in the emKate mutants are due to a different proportion in the two emission peaks,[31] resembling a mixing of green and red paint at different proportions.

For golden mutants, the green and red peaks found in the emission spectrum of the parental protein emKate disappeared, and only one emission peak was observed between 550 and 564 nm when the proteins were excited at 474 nm. Similar to mHoneydew[28] and the enhanced cyan fluorescent protein (eCFP),[32] these mutants also displayed broad excitation and emission peaks due to the substitution Y64W, suggesting the existence of more than one excited state species.

Because of the large difference between the excitation and emission wavelengths found in the golden mutants, these proteins may be classified as large Stokes shift-yellow fluorescent proteins (LSS-YFPs) and may be used in combination with a green-emitting protein (such as eGFP[29] or sgGFP[30]) and a red-emitting large Stokes shift protein[33] for multilabeling experiments, exciting only at 474 nm. Multilabeling experiments with excitation at a single wavelength have recently been described using an elaborated LSS-orange fluorescent mutant derived from mOrange.[34]

Finally, the most fluorescent red and golden mutants, emKate_F62L, emKate_Y64W, and emKate_F62C/Y64W, and the reference proteins mKate and emKate were purified to determine their fluorescent properties. However, whereas the reference proteins produced similar absorption and excitation spectra, the mutants produced quite different results, i.e., one additional peak at a shorter wavelength was observed in the absorption spectra (Supplementary Figure 3S). For example,

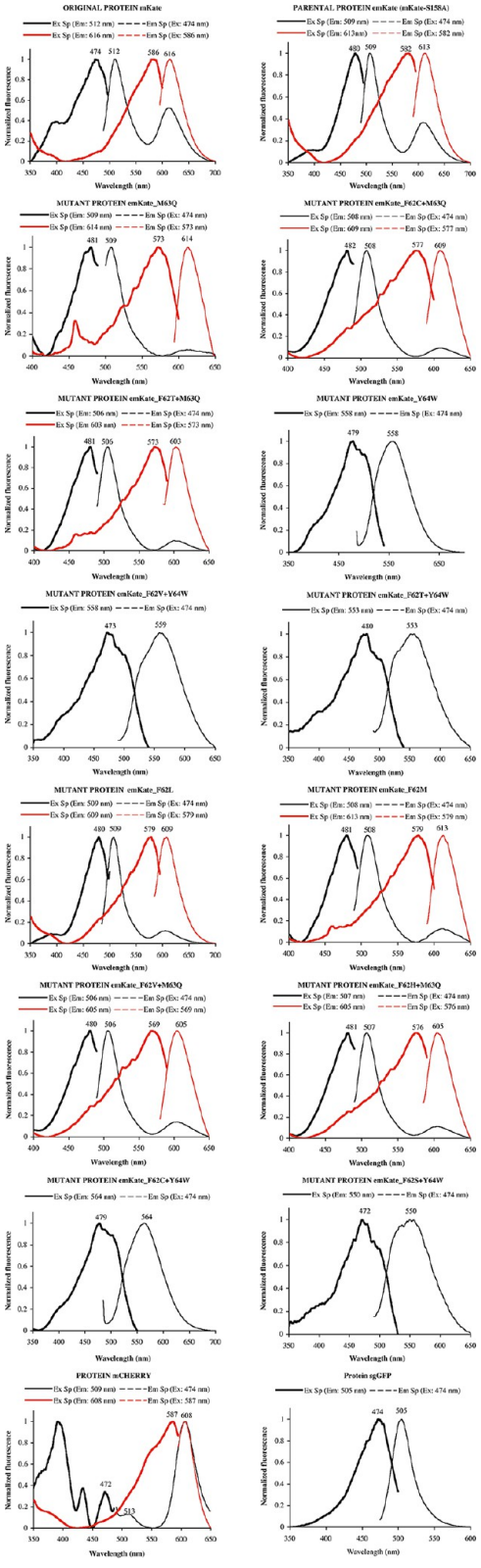


Figure 4. Excitation and emission spectra of the reference and mutant fluorescent proteins generated in this work. Only the proteins mKate, emKate, emKate_F62L, emKate_Y64W, and emKate_F62C/Y64W were analyzed in pure form, whereas all of the other proteins were analyzed as total soluble extracts. Initially, all proteins were excited at 474 nm, and the emission spectra were recorded. After identifying the maximum emission peaks, the excitation spectrum for each peak was recorded.

the red mutant emKate_F62L produced two absorption peaks at 382 and 579 nm, reinforcing the idea of incomplete maturing. In this context, the 579 nm peak accounts for the absorption of the red chromophore, whereas the 382 nm peak accounts for the absorption of the GFP-type chromophore in a protonated state.[24] Excitation of the protein at 382 nm rendered a principal emission peak at 509 nm, confirming the presence of a GFP-type intermediate according to the route of chromophore maturation shown in Supplementary Figure 4S. The fluorescent properties of emKate_F62L are reported in Supplementary Table 1S, revealing a larger quantum yield than the parental protein emKate but a minor extinction coefficient because of the incomplete mature chromophore.

With respect to the golden mutants emKate_Y64W and emKate_F62C/Y64W, both exhibited a major absorption peak at 421 nm and a smaller shoulder at approximately 499 nm, as shown in Supplementary Figure 3S. Because none of these peaks matched the excitation peaks shown in Figure 4, we decided to postpone the determination of the fluorescence properties of these mutants until we could devise an explanation for these results. Meanwhile, excitation of both proteins at 421 nm rendered one emission peak at 474−476 nm and a smaller shoulder at approximately 550 nm. These results suggest that most of the protein remains in the intermediate CFP-type state[24] and that only a small amount reaches the complete maturation state, according to the mechanism shown in Supplementary Figure 4S for the mutant emKate_Y64W. The sluggish maturation of the golden mutants was also qualitatively confirmed when the *E. coli* streaks harboring the different mutants were grown at 37 °C for 20 h and then stored at 4 °C for 10 days. Whereas the fluorescence intensity of the reference proteins remained constant during this period, the brightness of the golden variants was notably increased, as can be observed in Figure 2. More brilliant and useful LSS-YFPs proteins will likely be obtained via directed evolution experiments focused on enhancing the maturation process.

Conclusion. We have demonstrated the robustness of the ERASC approach for the elimination of redundant and stop codons during the assembly of degenerate oligonucleotides. The reagents used for this approach are inexpensive and commercially available from several companies, and the ideal DNA synthesizer to perform ERASC is one of the most common pieces of equipment in many laboratories and companies. Because oligonucleotide synthesis is currently available even for nonchemists, anyone can synthesize his/her own primers, including degenerate oligonucleotides, such as those assembled by ERASC. The accurate separation of the resin, in different proportions into four columns, was made possible by pipetting the appropriate volume of material resuspended in the viscous solvent ethylene glycol. Contrary to two recently published reports,[18,35] randomizing *n* codons by ERASC requires the synthesis of only one degenerate oligonucleotide, whereas the other approaches require the synthesis of 4^n or 3^n oligonucleotides, respectively, increasing the experimental costs.

In addition, using the separation strategy described herein, accurate experiments for codon-based doping to mutate large amino acid regions randomly, in "spiked mode", can easily be performed, separating the resin into four columns and synthesizing the wild-type codon in column 1, NDT in column 2, VHG in column 3, and TGG in column 4.[35] The average number of replaced amino acids per variant will be determined

dx.doi.org/10.1021/sb3001326 | *ACS Synth. Biol.* 2013, 2, 453−462

by the amount of resin separated for the synthesis of the wild-type codon.

By using the absorption and fluorescence analyses of the emKate_F62L and emKate_Y64W mutants, we have demonstrated that the complete maturation of the red or golden yellow chromophore proceeds via the formation of the green-type or cyan-type chromophore, respectively, and not via a blue intermediate, as previously suggested.[36]

Finally, if ERASC was useful for the creation of fluorescent protein reporters with interesting properties, such as the LSS-YFPs, it will also be useful for improving the properties of enzymes and antibodies with potential industrial and clinical applications.

METHODS

Standard β-cyanoethyl phosphoramidites of dA^{bz}, dC^{Ac}, dG^{dmf}, and dT and empty TWIST columns for 0.2-μmol scale synthesis were purchased from the Glen Research Corporation (Sterling, VA). TWIST columns are appropriate for the resin-splitting process because they are easily opened and closed by hand. 5-Ethylthiotetrazole (ETT) and the support for the oligonucleotide synthesis (DMTr-dC-LCAA-CPG, 2000 Å, loading 13 μmol/g) were purchased from Chemgenes (Wilmington, MA). The other ancillary reagents for oligonucleotide synthesis and the kanamycin (Km) were purchased from Aldrich.

The restriction endonucleases, Vent DNA polymerase, and deoxynucleoside-triphosphates (dNTPs) were purchased from New England Biolabs (Ipswich, MA) and used according to standard protocols. T4 DNA ligase was purchased from Fermentas. The purification kits for the plasmid and PCR products were purchased from either Roche (Mannheim, Germany) or BioBasic Inc. (Ontario, Canada). The constitutive plasmid pJOQ, kindly donated by Dr. Joel Osuna-Quintero, was used for all cloning and expression experiments. pJOQ is a 2.0-kb pUC-like plasmid that contains the kanamycin-resistance gene instead of the ampicillin gene. Cloned genes were inserted as *Nde*I/*Xho*I-digested fragments and were transcribed from the trc promoter. The *E. coli* MC1061 strain was used for cloning and plasmid and protein production.

The screening of fluorescent colonies was performed over a dark reader transilluminator (Clare Chemical Research; Dolores, CO), whereas the excitation and emission spectra were recorded using a luminescence spectrometer (LS50B, Perkin-Elmer).

Synthesis of Degenerate Oligonucleotides. Two degenerate oligonucleotides were synthesized on a 394 DNA synthesizer (Applied Biosystems; Foster City, CA) using the ERASC approach described herein. This synthesizer was originally designed for the parallel assembly of four oligos. Because it is equipped with eight vials for nucleotide phosphoramidites, it is ideal for the ERASC approach.

The synthesizer was loaded with 80 mM acetonitrile solutions of β-cyanoethyl phosphoramidites of dA, dG, dC, and dT in positions 1, 2, 3, and 4, respectively. These solutions were also used as stock solutions to prepare the equimolar mixtures N (dA + dC + dT + dG, loaded in position 5), D (dA + dG + dT, position 6), V (dA + dC + dG, position 7), and M (dA + dC, position 8). The other positions of the synthesizer were loaded with standard reagents, as shown in Figure 1A.

Both degenerate oligonucleotides were designed on the basis of the sense strand of the gene and contained a randomized codon (XXX) instead of each wild-type target codon, a 3′ flanking region for a first PCR amplification, and a 5′ flanking region for a second PCR amplification. These primers were labeled as 3COD and 12COD, which is in agreement with the numbers of the codons to be replaced.

3COD: 5′TTTGCCTTCGACATCCTGGCTACCAGT $(\mathbf{XXX})_3$ GGCAGCAAAACGTTCATTAATC 3′

12COD: 5′GGTGGCCCTTTGCCCTTTGCCTTCGAC $(\mathbf{XXX})_{12}$ TTCATTAATCATACCCAGGGCATC 3′

Only the synthesis of the degenerate oligonucleotide 3COD will be described because the 12COD was synthesized in the same mode.

ERASC Approach. The cycle of synthesis that was used for the assembly was similar to the one recommended by the manufacturer for the 0.2-μmol scale. However, the first acetonitrile wash was increased from 10 to 20 s to remove the residual ethylene glycol after the splitting process.

Because oligonucleotide synthesis is performed in the 3′→5′ direction, the 3′ flanking fragment 5′ GGCAGCAAAA-CGTTCATTAATC-3′ was first synthesized on column 1, in the Trityl-on mode, starting with 15 mg of dC resin (equivalent to 0.2 μmol of the starting nucleoside). At the end of the synthesis, the resin was dried by argon flushing for 1 min, and the column was opened and sealed at the bottom end with an empty disposable syringe. The resin was resuspended in 200 μL of ethylene glycol and agitated using a pipetter. Sixty microliters was transferred into column 2, while columns 3 and 4 received 10 μL each. The remainder of the volume was left in column 1. Each column was closed again and fitted on the synthesizer. The coupling of degenerate codon NDT was programmed for column 1, VMA for column 2, ATG for column 3, and TGG for column 4. When the synthesis ended, the resins were dried as before and combined with 200 μL of ethylene glycol, starting from column 4 and ending in column 1. The process of splitting, synthesis, and recombining was repeated twice to complete the three degenerate codons. In the case of 12COD, the steps were repeated 12 times. At the end of the mutagenic process, the resins were combined in column 1, and the 5′ flanking fragment was added (5′TTTGCCTTCGACATC-CTGGCTACCAGT-3′).

Both degenerate oligonucleotides were deprotected as normal and purified by denaturing polyacrylamide gel electrophoresis (PAGE).

mKate-S158A Gene (*emKate*). The synthetic gene encoding for the red fluorescent protein mKate-S158A (GenBank accession number EU383029) was assembled in a similar mode to that reported by Stemmer et al.;[21] however, the assembly and amplification reaction was performed in only one step. The 100-μL PCR reaction contained 22 internal oligonucleotides (60 mers) at individual 4 nM concentrations, two outermost primers (400 nM), 2 U of vent DNA polymerase and dNTPs, $MgSO_4$, and buffer as recommended by New England Biolabs. The amplification process was achieved under the following conditions: 94 °C/3 min; 25 cycles of 94 °C/1 min, 58 °C/1 min, 72 °C/1 min; and finally 72 °C/3 min. The internal oligonucleotides, shown in Supplementary Figure S1, were designed to partially overlap each other by 30 nt and to eliminate rare *E. coli* codons, whereas the outermost primers (Pr1-Fw and Pr12-Rv) were designed to code part of the gene and to introduce *Nde*I and *Xho*I restriction sites at the 5′ and 3′ end of the gene, respectively. Pr12-Rv also contained the information to introduce a polyhistidine tag, -SGGSHHHHHH, at the carboxy end of the protein.

dx.doi.org/10.1021/sb3001326 | *ACS Synth. Biol.* 2013, 2, 453−462

Because the PCR produced only one fragment of the expected size (756 bp), the DNA was cleaned and recovered with 60 μL of water using an EZ-10 spin column for the purification of PCR products. After being digested with the enzymes *Nde*I/*Xho*I (37 °C/12 h) under standard conditions, ligation of the agarose-purified fragment to the pJOQ vector, the electroporation of the direct ligation (2 μL) with 100 μL of MC1061 electrocompetent cells and growth on LB plates supplemented with kanamycin, several red fluorescent colonies emerged after 20 h of incubation (37 °C).

Gene Libraries. The first library of mutants was named I57-T68 and was constructed via the overlap PCR approach shown in Supplementary Figure 2S using *emKate* as template, the degenerate oligonucleotide 12COD, the partial complementary primer PCP-12COD (5′GTCGAAGGCAAAGGGCAAAG-GGCCACC 3′), the outermost primers Pr1-Fw and Pr12-Rv, and the Vent DNA polymerase for the amplification process. A first PCR reaction, employing the primers Pr1-Fw and PCP-12COD, produced a 179-bp fragment, whereas a second PCR reaction, employing the primers 12COD and Pr12-Rv, generated a 604-bp fragment. Both products were purified from an agarose gel, and 10 ng of the first fragment and 30 ng of the second were mixed and used as template for a third PCR reaction using only the outermost primers, yielding a 756-bp fragment. Because the third PCR reaction produced only one visible product, the DNA was subjected to the same steps described above to clone the expected library of mutants, I57-T58.

Because this library was constructed to assess the experimental frequency of each codon generated by ERASC, the plasmids from several nonfluorescent colonies were isolated and sequenced.

The second library of mutants, named F62-Y64, was created as the library I57-T68 using the primer PCP-3COD (ACTGG-TAGCCAGGATGTCGAAGGCAAA) instead of PCP-12COD for the first PCR reaction and the degenerate oligonucleotide 3COD instead of 12COD for the second PCR reaction. Plasmids from some fluorescent colonies displaying different phenotypes were isolated and sequenced.

***mKate* Gene.** The gene encoding for mKate, the original protein reported by Shcherbo et al.,[20] was constructed via the overlap PCR approach shown in Figure 2S using *emKate* as template, the pair of complementary primers A158S-Fw (5′ ggcctggaaggTcgttcggacatggccctg-3′) and A158S-Rv (5′-caggg-ccatgtccgaacgAccttccaggcc-3′), the outermost primers Pr1-Fw and Pr12-Rv, and the Vent DNA polymerase for the amplification process. A first PCR reaction, employing the primers Pr1-Fw and A158SRv, produced a 497-bp fragment, whereas a second PCR reaction, employing the primers A158S-Fw and Pr12-Rv, generated a 289-bp fragment. Both products were purified from an agarose gel, and 20 ng of each fragment was used as template for a third PCR reaction using only the outermost primers, yielding a 756-bp fragment. Because the third PCR reaction generated only one visible product, the DNA was subjected to the same steps described above to clone the original *mKate* gene. After sequencing the plasmid isolated from one of the fluorescent red colonies, the DNA sequence of the gene was confirmed.

Protein Purification. The fluorescent protein mutants emKate_F62L, emKate_Y64W, and emKate_F62C/Y64W, the parental emKate (mKate-S158A), and the original mKate protein expressed in the *E. coli* MC1061 strain were produced in 50-mL LB cultures supplemented with kanamycin (37 °C) during a 20-h incubation period under agitation (200 rpm). The cells were recovered by centrifugation and lysed with 5 mL of B-PER reagent (Thermo Scientific) for 1 h at room temperature. After centrifugation at 13,000 rpm, the clear supernatant was loaded on a HisTrap HP/1 mL column (GE Healthcare), and the proteins were purified using a gradient from 10% to 100% of buffer B in 30 min (flow rate of 0.75 mL/min) with detection at 280 nm. Buffer A: 100 mM phosphate buffer (pH 7.2) containing 0.5 M NaCl; buffer B: 300 mM imidazole in buffer A. Fractions containing the pure protein were combined, concentrated on centrifugal filter devices (amicon Ultra-4 10k, Millipore), and washed with PBS 1X (2 × 4 mL) to remove the residual imidazole. The pure proteins were redissolved in 1.5 mL of PBS 1X, and the concentration was determined using the Bradford assay with BSA standards and an UV−vis spectrophotometer (Ultrospec 210, Pharmacia).

Spectral Measurements. The extinction coefficient (ε_{max}) of the pure proteins at their maximum absorbance peak was determined using a spectrophotometer by measuring the absorbance of six duplicate dilutions. Plotting concentration (M) versus absorbance produced a linear graph whose slope was ε_{max} in units M^{-1} cm^{-1}. Each of the dilutions was further diluted 25-fold, and fluorescence was measured at the maximum emission wavelength with excitation at the 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 (ϕ) of the mutants, we determined the relationship between absorbance (586 nm) and fluorescence (616 nm) for the pure protein mKate, whose reported value is 0.33.[20] Dividing the sample's slope by mKate's slope and multiplying by 0.33 gives the sample's quantum yield.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Oligonucleotides used for the assembly of the synthetic gene emKate; overlap extension PCR strategy for the assembly of libraries, exemplified by the I57-T68 library; absorption spectra for the pure proteins mKate, emKate, emKate_F62L, emKate_Y64W and emKate_F62C/Y64W; proposed intermediates and final chromophores generated during the maturation of the mutant proteins emKate_F62L and emKate_Y64W; and quantum yields and molar extinction coefficients for the parental protein emKate and the mutant emKate_F64L isolated and purified in this work. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: paul@ibt.unam.mx.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Technical assistance from Jorge Yáñez, Santiago Becerra, Ana Yanci Alarcón, Oskar Rivera, Michelle Reyes-Ruiz, and Iván Bello San Martín is highly appreciated. Leopoldo Güereca provided advice for the purification and analyses of proteins, and Dr. Humberto Flores kindly donated the MC1061 electrocompetent cells. Dr. Hugh Mackie and Dr. Joel Osuna kindly reviewed and edited this manuscript.

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., 3rd, 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. U.S.A. 83*, 710−714.

(5) Dale, S.-J., and Belfield, M. (1996) Oligonucleotide-directed random mutagenesis using the phosphorothioate method. *Methods Mol. Biol. 57*, 369−374.

(6) Sirotkin, K. (1986) Advantages to mutagenesis techniques generating populations containing the complete spectrum of single codon changes. *J. Theor. Biol. 123*, 261−279.

(7) 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.

(8) Arunachalam, T.-S., Wichert, C., Appel, B., and Muller, S. (2012) Mixed oligonucleotides for random mutagenesis: best way of making them. *Org. Biomol. Chem. 10*, 4641−4650.

(9) 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.

(10) 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.*, DOI: 10.1186/1472-6750-7-65.

(11) Gaytan, P., Contreras-Zambrano, C., Ortiz-Alvarado, M., Morales-Pablos, A., and Yanez, J. (2009) TrimerDimer: an oligonucleotide-based saturation mutagenesis approach that removes redundant and stop codons. *Nucleic Acids Res.*, DOI: 10.1093/nar/gkp602.

(12) Neuner, P., Cortese, R., and Monaci, P. (1998) Codon-based mutagenesis using dimer-phosphoramidites. *Nucleic Acids Res. 26*, 1223−1227.

(13) Hooft van Huijsduijnen, R.-A., Ayala, G., and DeLamarter, J.-F. (1992) A means to reduce the complexity of oligonucleotides encoding degenerate peptides. *Nucleic Acids Res. 20*, 919.

(14) Glaser, S.-M., Yelton, D.-E., and Huse, W.-D. (1992) Antibody engineering by codon-based mutagenesis in a filamentous phage vector system. *J. Immunol. 149*, 3903−3913.

(15) Zon, G., Gallo, K.-A., Samson, C.-J., Shao, K.-L., Summers, M.-F., and Byrd, R.-A. (1985) Analytical studies of 'mixed sequence' oligodeoxyribonucleotides synthesized by competitive coupling of either methyl- or beta-cyanoethyl-N,N-diisopropylamino phosphoramidite reagents, including 2′-deoxyinosine. *Nucleic Acids Res. 13*, 8181−8196.

(16) Tabuchi, I., Soramoto, S., Ueno, S., and Husimi, Y. (2004) Multi-line split DNA synthesis: a novel combinatorial method to make high quality peptide libraries. *BMC Biotechnol.*, DOI: 10.1186/1472-6750-4-19.

(17) Yin, C.-C., Ren, L.-L., Zhu, L.-L., Wang, X.-B., Zhang, Z., Huang, H.-L., and Yan, X.-Y. (2008) Construction of a fully synthetic human scFv antibody library with CDR3 regions randomized by a split-mix-split method and its application. *J. Biochem. 144*, 591−598.

(18) Tang, L., Gao, H., Zhu, X., Wang, X., Zhou, M., and Jiang, R. (2012) Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. *BioTechniques 52*, 149−158.

(19) Chu, J., Zhang, Z., Zheng, Y., Yang, J., Qin, L., Lu, J., Huang, Z.-L., Zeng, S., and Luo, Q. (2009) A novel far-red bimolecular fluorescence complementation system that allows for efficient visualization of protein interactions under physiological conditions. *Biosens. Bioelectron. 25*, 234−239.

(20) 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., and Chudakov, D.-M. (2007) Bright far-red fluorescent protein for whole-body imaging. *Nat. Methods 4*, 741−746.

(21) Stemmer, W.-P., Crameri, A., Ha, K.-D., Brennan, T.-M., and Heyneker, H.-L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene 164*, 49−53.

(22) 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.

(23) Hecker, K.-H., and Rill, R.-L. (1997) Synthetic polynucleotide templates for characterizing sequence-selective small molecule/nucleic acid interactions. *Anal. Biochem. 244*, 67−73.

(24) Zacharias, D.-A., and Tsien, R.-Y. (2006) Molecular biology and mutation of green fluorescent protein. *Methods Biochem. Anal. 47*, 83−120.

(25) Zaccolo, M., and Gherardi, E. (1999) The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase. *J. Mol. Biol. 285*, 775−783.

(26) Stepanenko, O.-V., Shcherbakova, D.-M., Kuznetsova, I.-M., Turoverov, K.-K., and Verkhusha, V.-V. (2011) Modern fluorescent proteins: from chromophore formation to novel intracellular applications. *BioTechniques 51*, 313−327.

(27) Matz, M.-V., Fradkov, A.-F., Labas, Y.-A., Savitsky, A.-P., Zaraisky, A.-G., Markelov, M.-L., and Lukyanov, S.-A. (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol. 17*, 969−773.

(28) Shaner, N.-C., Campbell, R.-E., Steinbach, P.-A., Giepmans, B.-N., Palmer, A.-E., and Tsien, R.-Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol. 22*, 1567−1572.

(29) Cormack, B.-P., Valdivia, R.-H., and Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene 173*, 33−38.

(30) 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.

(31) Merola, F., Levy, B., Demachy, I., and Pasquier, H. (2010) Photophysics and spectroscopy of fluorophores in the green fluorescent protein family. In *Advanced Fluorescence Reporters in Chemistry and Biology I: Fundamentals and Molecular Design* (Demchenko, A. P., Ed.). Springer Ser. Fluoresc. 8: pp 347−384, Springer-Verlag, Berlin, Heidelberg.

(32) Tramier, M., Gautier, I., Piolot, T., Ravalet, S., Kemnitz, K., Coppey, J., Durieux, C., Mignotte, V., and Coppey-Moisan, M. (2002) Picosecond-hetero-FRET microscopy to probe protein-protein interactions in live cells. *Biophys. J. 83*, 3570−3577.

(33) Piatkevich, K.-D., Hulit, J., Subach, O.-M., Wu, B., Abdulla, A., Segall, J.-E., and Verkhusha, V.-V. (2010) Monomeric red fluorescent proteins with a large Stokes shift. *Proc. Natl. Acad. Sci. U.S.A. 107*, 5369−5374.

(34) Shcherbakova, D.-M., Hink, M.-A., Joosen, L., Gadella, T.-W., and Verkhusha, V.-V. (2012) An orange fluorescent protein with a large Stokes shift for single-excitation multicolor FCCS and FRET imaging. *J. Am. Chem. Soc. 134*, 7913−7923.

(35) Kille, S., Acevedo-Rocha, C.-G., Parra, L.-P., Zhang, Z.-G., Opperman, D.-J., Reetz, M.-T., and Acevedo, J.-P. (2012) Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.*, DOI: 10.1021/sb300037w.

(36) Subach, F.-V., and Verkhusha, V.-V. (2012) Chromophore transformations in red fluorescent proteins. *Chem. Rev. 112*, 4308−4327.