Codon-based Mutagenesis

John Randolph[§], Andrey Yagodkin[‡]; Alex Azhayev[‡]; Hugh Mackie[§]
[§]Glen Research Corporation, 22825 Davis Drive, Sterling, Virginia, 20164 USA
[‡]Department of Pharmaceutical Chemistry, University of Kuopio, FIN-70211 Kuopio, Finland

A set of Trimer phosphoramidites was synthesized covering all 20 amino acid codons. These trimer phosphoramidites can be added during synthesis using standard DNA chemistry. A Reaction Factor (RF) was determined empirically for each trimer to compensate for differences in their relative rate of reaction during coupling. It is therefore possible to introduce an equimolar mix of all 20 amino acid codons, or subsets thereof, at any location within a sequence. By mutating a gene at the codon level rather than at individual bases, it is possible to avoid codon bias, frame-shift mutations and the introduction of stop codons, making Trimer phosphoramidites a highly efficient tool for the exploration of sequence space in proteins.

Introduction

The introduction of mutations can be used to fine-tune almost any property in existing proteins. However, generating and finding these improved proteins can be a difficult task. Two popular methodologies are Random Point Mutagenesis and Saturating Mutagenesis:

Random Point Mutagenesis typically employs a low fidelity DNA polymerase, generally in conjunction with Mn^{2+}, to produce point mutations throughout the sequence (Error Prone PCR). The mutation rate is low overall so that the protein will maintain its function, albeit to a greater or lesser extent, and fold properly. This method allows many residues within the protein to be interrogated. However, you cannot adequately explore the protein sequence space. Single base mutations, on average, only afford about six amino acid substitutions, therefore adjacent nucleotides have to be mutated concurrently. This lowers the chance of finding the desired clone. For example:

-The probability of converting Tyr (TAC) to Asn (AAC) in a 100 residue protein is only 1/900
(1/(100 x 3) x 1/3)
-For mutating Tyr (TAC) to Met (ATG), this number drops to 1.37 $x10^{-9}$

Saturating Mutagenesis typically uses a pool of degenerate oligonucleotides which can be incorporated into genes as cassettes or by PCR using the degenerate oligo as a primer. The degenerate oligonucleotides are synthesized as a mixture of A/C/G/T phosphoramidites (N) at the site of the codons to be mutated. There are, however, again difficulties. Out of the 64 possible codon combinations of A, C, G and T, 18 code for leucine, arginine or serine, but only 2 for tryptophan or methionine. As a result, only 3% of the mutagenic oligonucleotides will contain methionine or tryptophan, and over 28% will contain either leucine, arginine or serine. In addition, the three nonsense codons will lead to chain termination in 4.7% of the sequences.

Trimer Phosphoramidites, by covering all 20 amino acids, can offer an elegant solution that circumvents the problems found in both Random Point Mutagenesis as well as Saturating Mutagenesis - because they allow mutation to occur at the *codon level* rather than individual bases. By determining a Reaction Factor (RF) for each Trimer to compensate for differences in the relative rate of reaction during coupling, it is possible to introduce an equimolar mix of all 20 amino acid codons, or subsets thereof, at any location within a sequence. This makes searching a clonal library much more efficient under both saturating and non-saturating mutagenesis conditions. For example, by defining each port of a 20 port DNA synthesizer[1] a particular codon, it is possible to interrogate all amino acids within a protein at a non-saturating level without codon bias, the introduction of stop codons or the generation of frame-shift mutations. This would allow multiple amino acids within a protein to be simultaneously interrogated while maintaining the protein's tertiary structure overall. Under saturating conditions, all 20 trimers or any subset thereof, can be introduced at multiple regions within a sequence using just a single additional port on the synthesizer, weighting the codons as desired - without the introduction of unwanted codon bias or stop codons.

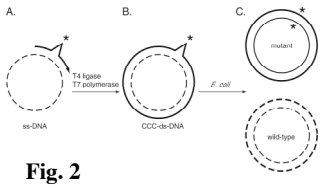


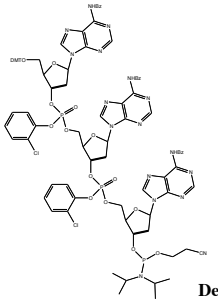
Fig. 2. The mutagenic oligonucleotide (solid line) is annealed to a circularized dU-containing ssDNA template (dashed line) obtained from a *dut/ung*⁻ E. coli strain that lacks Uracil DNA Glycolysase. The mismatched variable region is flanked by complementary regions (A). The plasmid has the variable region filled by the action of T7 polymerase and T4 ligase (B). After introduction into *ung*⁺ E. coli, the template DNA containing dU is selectively destroyed, leading to the enrichment of the mutant plasmid (C).

Fig. 2

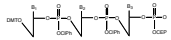
Results

A set of *E. coli* optimized trimer phosphosphoramidites coding for all 20 amino acids was synthesized using a variant of the synthesis of Kayushin[2]. Shown in Fig. 1 is an example of the trimer AAA.

Fig. 1 Structure of Trimer Phosphoramidites



Detailed structure of trimer AAA



General structure of Trimer Phosphoramidites where B =A^{bz}, C^{bz}, G^{ibu}, or T

Trimer phosphoramidites, with a chiral center at each phosphorus, contain 2^3 diastereomers, making validation by RP HPLC difficult. To ensure the identity of the trimer and its correct orientation - i.e. CAT (His) could not be mistaken for TAC (Tyr), HPLC MS and MS/MS was used[3]. The principal dissociation pathway using Collision-induced Dissociation is in the 3' to 5' direction, allowing the orientation to be determined unequivocally. The codon set used was optimized for E. coli. Starting from the RFs initially determined by Kayushin[2], the RFs were adjusted after obtaining sequencing results in which custom trimer mixes were used to transform *E. coli*. Shown in Table 1 are the Trimers, their corresponding amino acids, RF and an example of sequencing data obtained for a custom mix provided by Glen Research Corporation[4]. Shown in Fig. 2 is a recommended mutagenesis strategy[5]. Trimer Phosphoramidites have been particularly useful in phage-display libraries. A recent paper by Krumpe found that using trimer phosphoramidites for generating a phage-display library increased diversity by an order of magnitude over traditional methods while exhibiting excellent amino acid uniformity [6].

Table 1

Trimer		mw	RF	mw x RF	number sequenced	Observed	Expected
AAA (Lys)	K	1911.5	1.10	2102.65	51	3.2%	5.3%
AAC (Asn)	N	1887.5	1.00	1887.50	52	3.3%	5.3%
ACT (Thr)	T	1774.5	1.30	2306.85	74	4.7%	5.3%
ATC (Ile)	I	1774.5	1.20	2129.40	88	5.6%	5.3%
ATG (Met)	M	1780.5	1.30	2314.65	78	4.9%	5.3%
CAG (Gln)	Q	1869.5	2.00	3739.00	99	6.3%	5.3%
CAT (His)	H	1774.5	1.90	3371.55	113	7.1%	5.3%
CCG (Pro)	P	1845.5	1.80	3321.90	79	5.0%	5.3%
CGT (Arg)	R	1756.5	1.10	1932.15	64	4.0%	5.3%
CTG (Leu)	L	1756.5	1.20	2107.80	73	4.6%	5.3%
GAA (Glu)	E	1893.5	1.90	3597.65	98	6.2%	5.3%
GAC (Asp)	D	1869.5	1.30	2430.35	81	5.1%	5.3%
GCT (Ala)	A	1756.5	1.50	2634.75	76	4.8%	5.3%
GGT (Gly)	G	1762.5	1.10	1938.75	51	3.2%	5.3%
GTT (Val)	V	1667.5	1.90	3168.25	107	6.8%	5.3%
TAC (Tyr)	Y	1774.5	1.60	2839.20	226	14.3%	10.5%
TCT (Ser)	S	1661.4	1.30	2159.82	161	10.2%	10.5%
TGC (Cys)	C	1756.5	1.50	2634.75	0	0.0%	0.0%
TGG (Trp)	W	1762.5	2.40	4230.00	0	0.0%	0.0%
TTC (Phe)	F	1661.4	1.30	2159.82	3	0.2%	0.0%
Stop Codon					0	0.0%	0.0%

[1]For example, the MerMade 12 (BioAutomation, Plano, TX)
[2]A.L. Kayushin, M.D. Korosteleva, A.I. Miroshnikov, W. Kosch, D. Zubov and N. Piel, *Nucleic Acids Research*, 1996, **24**, 3748-3755.
[3]T. Mauriala, S. Auriola, A. Azhayev , A. Kayushin, M. Korosteleva, and A. Miroshnikov , *J. Pharmaceutical and Biomedical Analysis*, 2004, **34**, 199-206.
[4]For a quote on a Custom Trimer Phosphoramidite Mix, please contact: support@glenres.com or go to http://www.glenres.com.
[5]S.S. Sidu, H.B. Lowman, B.C. Cunningham and J.A. Wells, *Methods Enzymol.*, 2000, **328**, 333-363.
[6]L.R. Krumpe, K.M. Schumacher, J.B. McMahon, L. Makowski, T. Mori, *BMC Biotechnology*, 2007 Oct 5, **7**:65.