A General Method for Scanning Unnatural Amino Acid Mutagenesis

Kelly A. Daggett, Mark Layer, and T. Ashton Cropp*
Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

Whether through chemical synthesis or genetic mutagenesis, the ability to rationally change the amino acid sequence of proteins has had a profound impact upon the understanding of structure and function. Diversity-oriented approaches to protein mutagenesis such as error-prone PCR (1), DNA shuffling (2), or cassette mutagenesis (3) have expanded on this and can easily create mixtures of complexity that exceed protein screening capabilities. Scanning mutagenesis utilizes a different approach by making small changes, such as single alanine replacements (4), so that the contribution of individual side chains can be deciphered and later exploited. This generates mixtures of modest, defined complexity and is well-suited for the mapping of protein–protein interactions, solvent accessibility, or active sites using low- to medium-throughput assays. The limitation, however, is the time and effort spent creating the mutations via PCR-based methods. For example there are a clear, defined number of single alanine mutants that could be made of a protein 500 amino acids long, but generating hundreds of amber codon mutations is not practical. Further, it is not possible to randomly introduce these codons via single nucleotide exchanges incorporated with error-prone PCR because of the redundancy and composition of the genetic code. For example the probability of randomly changing a proline codon (CCC) to an amber codon (TAG) by three successive nucleotide mutations is effectively zero. Here we present a rational approach to randomly replace single codons within an open reading frame with TAG such that every site is sampled. This collection of gene mutants then gives rise to the independent mutation of every amino acid residue in the protein with an unnatural amino acid. In contrast to other methods that are used to perform scanning mutagenesis (8, 9), these library members are not redundant or silent and are enriched with members containing only a single mutation. Moreover, we have ensured that the mutation is in the correct reading frame of the protein being studied. This results in a “cleaner” library and allows every possible mutant to be represented in a fairly small collection of clones, particularly when compared to other methods. This convergent process, consisting of relatively simple DNA manipulations, can be applied to any open reading frame encoding a protein.

Previously, transposon methods have been used to map protein domains via photo-crosslinking unnatural amino acids that are genetically encoded by the amber stop codon, TAG (5–7). If one is attempting to map multiple binding sites on a protein, it is important to create mutations that are close but not integral to binding interfaces. Ideally, one would want to sample many different sites, perhaps every site, but making hundreds of amber codon mutations is not practical. Further, it is not possible to randomly introduce these codons via single nucleotide exchanges incorporated with error-prone PCR because of the redundancy and composition of the genetic code. For example the probability of randomly changing a proline codon (CCC) to an amber codon (TAG) by three successive nucleotide mutations is effectively zero. Here we present a rational approach to randomly replace single codons within an open reading frame with TAG such that every site is sampled. This collection of gene mutants then gives rise to the independent mutation of every amino acid residue in the protein with an unnatural amino acid. In contrast to other methods that are used to perform scanning mutagenesis (8, 9), these library members are not redundant or silent and are enriched with members containing only a single mutation. Moreover, we have ensured that the mutation is in the correct reading frame of the protein being studied. This results in a “cleaner” library and allows every possible mutant to be represented in a fairly small collection of clones, particularly when compared to other methods. This convergent process, consisting of relatively simple DNA manipulations, can be applied to any open reading frame encoding a protein.

Previously, transposon methods have been used to map protein domains via

*Corresponding author, acropp@umd.edu.

Received for review November 3, 2008 and accepted January 7, 2009.
Published online January 28, 2009
10.1021/cb800271f CCC: $40.75
© 2009 American Chemical Society
randomly distributed amino acid linker insertions. Recently Jones and co-workers described a transposon-based system for removal of triplet nucleotides from plasmid DNA using typeIIS asymmetric restriction enzymes and have gone on to replace those deletions with new sequence in an approach referred to as Tri-NEX. While this method replaces nucleotides, our goal was to perform single unnatural amino acid mutations in a protein and not affect the sequence that flanks the mutation. We adopted a similar strategy in which we perform nucleotide substitutions consisting of an amber stop codon and then ensure that this replacement is in-frame with the original coding sequence by genetic selection. The process consists of (i) a transposon reaction to randomly insert two unique MlyI restriction sites, (ii) digestion with MlyI, which randomly breaks a coding DNA sequence and removes three nucleotides, (iii) replacement and reassembly with three new nucleotides consisting of the desired mutation, and (iv) simultaneous verification that this replacement occurred in the correct reading frame of the gene (Figure 1). As a proof of principle, we performed this mutagenesis reaction on a plasmid containing a target gene encoding glutathione S-transferase (GST) from Schistosoma japonicum. The plasmid (pIT-GST) we used is a derivative of pInSALect, which expresses proteins as N-terminal fusions to a β-lactamase-VMA intein fusion protein. This is a system that has been used in the past for isolating intact open reading frames. Using a modified Mu transposon, a library of this plasmid was generated and recovered in E. coli. This insertion event results in the duplication of five nucleotides, N1N2N3N4N5, at the site of insertion (shown in panel b). The library can be purified to only contain members where the transposon was inserted into the gene of interest by restriction digest. b) The transposon fragment with recognition sequences, R1 and R2, is modified to have MlyI restriction sites on either end. Three base pairs, N1N2N3, of the original plasmid sequence are then removed by digestion with MlyI, and c) a new frame selectable TAG linker segment of DNA (orange) is ligated into this digestion site, resulting in a library of plasmids that are selected for correct reading frame in E. coli. The basis for the reading frame selection is assembly of a functional β-lactamase-intein chimera that splices to yield colonies resistant to ampicillin. d) This library is then subjected to a second digestion and religated, thus resulting in selective removal of the linker and a net replacement of a new codon.

Figure 1. Schematic diagram of codon scanning mutagensis. a) A plasmid containing the gene (blue) to be mutated is first targeted with a modified Mu transposon (green) that is inserted in vitro by the action of MuA transposase, resulting in a library of plasmids that are recovered in E. coli. This insertion event results in the duplication of five nucleotides, N1N2N3N4N5, at the site of insertion (shown in panel b). The library can be purified to only contain members where the transposon was inserted into the gene of interest by restriction digest. b) The transposon fragment with recognition sequences, R1 and R2, is modified to have MlyI restriction sites on either end. Three base pairs, N1N2N3, of the original plasmid sequence are then removed by digestion with MlyI, and c) a new frame selectable TAG linker segment of DNA (orange) is ligated into this digestion site, resulting in a library of plasmids that are selected for correct reading frame in E. coli. The basis for the reading frame selection is assembly of a functional β-lactamase-intein chimera that splices to yield colonies resistant to ampicillin. d) This library is then subjected to a second digestion and religated, thus resulting in selective removal of the linker and a net replacement of a new codon.
gene that must be ligated in-frame with the partner on the target plasmid to confer resistance to ampicillin (Figure 1). Upon successful in-frame ligation with the fusion gene on the original plasmid, constructs are generated that express a fusion protein where two halves of the β-lactamase resistance marker are separated by a self-splicing intein and the remnant piece of the coded target protein. This precursor protein is rapidly spliced in vivo by the intein (15), generating functional β-lactamase. Out-of-frame ligations do not allow for this preprotein to be expressed correctly and are therefore removed. As expected, when we replated this library containing the linker replacement on media containing ampicillin, approximately 90% of all library members were removed (as judged by colony count comparison to nonselective plates), consistent with removal of the five-sixths of the library derived from out-of-frame and reversed ligations. Moreover, DNA sequence analysis of the plasmids derived from the surviving colonies verified that all clones enriched in this step consisted of only in-frame fusion genes that were randomly distributed.

This frame-selected collection of clones was pooled, and the plasmid DNA was extracted. The pooled library DNA was then again digested with MlyI to remove the linker segment (Figure 1), religated, and retransformed into E. coli. Simultaneously, the plasmid is recircularized, resulting in no net addition or subtraction of nucleotides and a seamless replacement of the original codon with a new amber stop codon, TAG. To verify that each of these steps proceeded as expected, we isolated 48 individual clones as representatives of the library and resequenced the ligation sites, one of which is shown in Figure 2. Out of the 48 clones, we observed 15 different clones (>30%) that were entirely correct sequences containing in-frame mutations. In the other clones we did observe unexpected deformations in the coding sequence such as the addition or removal of single nucleotides at the ligation sites (see Supporting Information). We believe that this is due to the fidelity of the MlyI restriction digests in the final step of library assembly. Nevertheless, the transposon insertions and the resulting mutations are distributed across the protein coding sequence, consistent with previous reports of the very limited specificity of Mu transposomes (17). Further, the stepwise nature of this process allows one to recover and amplify the plasmid DNA in cells and therefore count colonies to ensure that the modest yet statistically sampled diversity (10^3 clones) is more than maintained throughout the process.

After library construction, 10 plasmids containing GST with in-frame TAG mutations were chosen on the basis of sequencing results to be used for unnatural amino acid incorporation. The 10 mutant clones obtained from the library were subcloned into an overexpression vector and then used to transform E. coli cells that harbor the translational machinery capable of inserting the unnatural photo-crosslinking amino acid p-benzoylphenylalanine (pBpa) (6). The resulting double transformants were then grown as individual cultures and assayed for protein expression in the presence and absence of pBpa. As can be seen in Figure 3 these clones only express protein in the presence of the unnatural amino acid. The protein expression yields of the mutants...
effects of amber suppression (L64TAG) failed to produce at all. We believe this variability is based on the context of amber suppression (18) and, in the case of no expression, the stability of the mutant protein. Just as with standard site-directed mutagenesis, one would not expect all mutants to be equally stable. In order to probe the photoreactivity of those mutants that showed good expression while displaying pBpa (W7, W40, L49, P85, and P173), we irradiated them at 365 nm and analyzed for the formation of covalent cross-links across the dimeric GST interface (Figure 3). Of the five mutants we assayed, only P85 showed an indication of successful cross-link formation, albeit minor when compared to a mutant that is known to cross-link very well (F51) (6). This is consistent with the expectation that only those mutations in or near the dimeric interface (such as P85 and F51) would be capable of capturing the interface. More importantly, however, these individual protein expression results can be viewed as members of the larger population of clones that exist in the library mixture, each expressing a unique protein mutant containing a single unnatural amino acid mutation at a random position. That is, overexpression of the pool of clones created using this approach would generate a mixture of proteins (in this case GST) all of which are the same apart from a single, random unnatural amino acid mutation.

In summary, we have created an alternative, non-PCR-based approach for creating mixtures of (all possible) single residue, unnatural protein mutants to facilitate the initiation of large-scale scanning mutagenesis projects. The requirement for the application of this method is the lack of the unique and rare restriction sites used in the cloning process. If present in a candidate open reading frame, these sites can be removed easily by traditional site-directed silent mutagenesis or total gene synthesis (19, 20). In this case, we chose to express mutant proteins bearing photoreactive groups at the site of mutation. One could envision other uses for scanning unnatural amino acid mutagenesis, including strategies for bioconjugation, isotope or fluorophore labeling, and post-translational modifications. There is also evidence that enzymes bearing unnatural amino acids can exhibit improved catalytic activity, above and beyond that available using the 20 genetically encoded amino acids (21). In addition to amber codon mutations, an analogous approach (using a different codon linker segment) could be taken to perform alanine scanning to map protein epitopes (4), cysteine scanning to map protein surfaces (22), or lysine or aspartic acid scanning to “super-charge” proteins (23). Indeed, we have now created the full complement of the required 20 reading-frame-selectable linkers, such that a combined approach in which multiple linkers coding for custom mixtures of amino acids (or all 20) could be used as a method for creating defined molecular diversity in protein libraries. This represents an alternative to error-prone mutagenesis that is not limited by the redundancy of the genetic code. This allows a much more efficient sampling of protein sequence space. Finally, because the approach is convergent, it could be iterated to generate libraries of proteins that contain two, three, or four mutations, etc. or collections of single mutants recombined in vitro (24, 25) to create a distribution of these possibilities. Rational protein diversification methods such as these could generate protein libraries that are much more likely to contain new or improved protein function.

**METHODS**

**Scanned Library Construction.** To create a codon-scanned library, the target plasmid (in this case pIT-GST, which is described in the Supporting Information) was used in an in vitro transposon mutagenesis reaction. In a 20 μL reaction, 400 ng of target plasmid was mixed with a 1.3 molar excess of the transposon, which is a BglII-digested PCR product. To the DNA mixture were added 2 μL of 10X HyperMu reaction buffer (1.5 M potassium acetate, 0.5 M Tris-acetate (pH 7.5), 0.1 M magnesium acetate and 40 mM spermidine) and 1 μL of Mu transposase (Epibenter), and the reaction was incubated at 30 °C for 4 h. The reaction was then stopped by adding SDS to a final concentration of

---

**Figure 3.** SDS/PAGE analysis of proteins. a) Ten amber codon mutants in the presence (+) and absence (−) of pBpa. Amino acid residues mutated to TAG codons are shown. One mutant, L65TAG, did not express for unknown reasons. b) Photoactivity of high expression mutants derived from scanning compared to a mutant known to cross-link (F51).

**Methods:**

**Scanned Library Construction.** To create a codon-scanned library, the target plasmid (in this case pIT-GST, which is described in the Supporting Information) was used in an in vitro transposon mutagenesis reaction. In a 20 μL reaction, 400 ng of target plasmid was mixed with a 1.3 molar excess of the transposon, which is a BglII-digested PCR product. To the DNA mixture were added 2 μL of 10X HyperMu reaction buffer (1.5 M potassium acetate, 0.5 M Tris-acetate (pH 7.5), 0.1 M magnesium acetate and 40 mM spermidine) and 1 μL of Mu transposase (Epibenter), and the reaction was incubated at 30 °C for 4 h. The reaction was then stopped by adding SDS to a final concentration of.

**Scanned Library Construction.** To create a codon-scanned library, the target plasmid (in this case pIT-GST, which is described in the Supporting Information) was used in an in vitro transposon mutagenesis reaction. In a 20 μL reaction, 400 ng of target plasmid was mixed with a 1.3 molar excess of the transposon, which is a BglII-digested PCR product. To the DNA mixture were added 2 μL of 10X HyperMu reaction buffer (1.5 M potassium acetate, 0.5 M Tris-acetate (pH 7.5), 0.1 M magnesium acetate and 40 mM spermidine) and 1 μL of Mu transposase (Epibenter), and the reaction was incubated at 30 °C for 4 h. The reaction was then stopped by adding SDS to a final concentration of.

---

**Methods:**

**Scanned Library Construction.** To create a codon-scanned library, the target plasmid (in this case pIT-GST, which is described in the Supporting Information) was used in an in vitro transposon mutagenesis reaction. In a 20 μL reaction, 400 ng of target plasmid was mixed with a 1.3 molar excess of the transposon, which is a BglII-digested PCR product. To the DNA mixture were added 2 μL of 10X HyperMu reaction buffer (1.5 M potassium acetate, 0.5 M Tris-acetate (pH 7.5), 0.1 M magnesium acetate and 40 mM spermidine) and 1 μL of Mu transposase (Epibenter), and the reaction was incubated at 30 °C for 4 h. The reaction was then stopped by adding SDS to a final concentration of.

---

**Methods:**

**Scanned Library Construction.** To create a codon-scanned library, the target plasmid (in this case pIT-GST, which is described in the Supporting Information) was used in an in vitro transposon mutagenesis reaction. In a 20 μL reaction, 400 ng of target plasmid was mixed with a 1.3 molar excess of the transposon, which is a BglII-digested PCR product. To the DNA mixture were added 2 μL of 10X HyperMu reaction buffer (1.5 M potassium acetate, 0.5 M Tris-acetate (pH 7.5), 0.1 M magnesium acetate and 40 mM spermidine) and 1 μL of Mu transposase (Epibenter), and the reaction was incubated at 30 °C for 4 h. The reaction was then stopped by adding SDS to a final concentration of.

---

**Methods:**

**Scanned Library Construction.** To create a codon-scanned library, the target plasmid (in this case pIT-GST, which is described in the Supporting Information) was used in an in vitro transposon mutagenesis reaction. In a 20 μL reaction, 400 ng of target plasmid was mixed with a 1.3 molar excess of the transposon, which is a BglII-digested PCR product. To the DNA mixture were added 2 μL of 10X HyperMu reaction buffer (1.5 M potassium acetate, 0.5 M Tris-acetate (pH 7.5), 0.1 M magnesium acetate and 40 mM spermidine) and 1 μL of Mu transposase (Epibenter), and the reaction was incubated at 30 °C for 4 h. The reaction was then stopped by adding SDS to a final concentration of.

---

**Methods:**

**Scanned Library Construction.** To create a codon-scanned library, the target plasmid (in this case pIT-GST, which is described in the Supporting Information) was used in an in vitro transposon mutagenesis reaction. In a 20 μL reaction, 400 ng of target plasmid was mixed with a 1.3 molar excess of the transposon, which is a BglII-digested PCR product. To the DNA mixture were added 2 μL of 10X HyperMu reaction buffer (1.5 M potassium acetate, 0.5 M Tris-acetate (pH 7.5), 0.1 M magnesium acetate and 40 mM spermidine) and 1 μL of Mu transposase (Epibenter), and the reaction was incubated at 30 °C for 4 h. The reaction was then stopped by adding SDS to a final concentration of.

---

**Methods:**

**Scanned Library Construction.** To create a codon-scanned library, the target plasmid (in this case pIT-GST, which is described in the Supporting Information) was used in an in vitro transposon mutagenesis reaction. In a 20 μL reaction, 400 ng of target plasmid was mixed with a 1.3 molar excess of the transposon, which is a BglII-digested PCR product. To the DNA mixture were added 2 μL of 10X HyperMu reaction buffer (1.5 M potassium acetate, 0.5 M Tris-acetate (pH 7.5), 0.1 M magnesium acetate and 40 mM spermidine) and 1 μL of Mu transposase (Epibenter), and the reaction was incubated at 30 °C for 4 h. The reaction was then stopped by adding SDS to a final concentration of.

---

**Methods:**

**Scanned Library Construction.** To create a codon-scanned library, the target plasmid (in this case pIT-GST, which is described in the Supporting Information) was used in an in vitro transposon mutagenesis reaction. In a 20 μL reaction, 400 ng of target plasmid was mixed with a 1.3 molar excess of the transposon, which is a BglII-digested PCR product. To the DNA mixture were added 2 μL of 10X HyperMu reaction buffer (1.5 M potassium acetate, 0.5 M Tris-acetate (pH 7.5), 0.1 M magnesium acetate and 40 mM spermidine) and 1 μL of Mu transposase (Epibenter), and the reaction was incubated at 30 °C for 4 h. The reaction was then stopped by adding SDS to a final concentration of.
Unnatural Amino Acids and Photo-Crosslinking.

For the pIT backbone and GST used to transform chemically competent isolated by gel electrophoresis, religated, and then covered in 4 mL SOC for 1 h and then plated on library was digested with the transposon inserts randomly into the target nies were picked and analyzed by restriction to en-...