Chapter 2

Incorporation of Unnatural Amino Acids into mRNA Display Libraries Using Nonsense Suppression

This chapter is adapted from the following reference:

Li, S., Millward, S.W., Roberts, R.W. (2002) *In vitro* selection of mRNA display libraries containing an unnatural amino acid *J Am Chem Soc 124*, 9972-9973.

Abstract

The incorporation of unnatural amino acids into selectable, amplifiable peptide and protein libraries expands the chemical diversity of such libraries, thus considerably facilitating the process of obtaining ligands with improved properties (affinity, specificity, and function), particularly against therapeutically interesting targets. Here, we report that biocytin, a biotin derivative of lysine, can be inserted into an mRNA-protein fusion through amber stop codon suppression. We also demonstrate that templates containing the codon corresponding to the biocytin tRNA (a UAG stop codon) can be enriched by iterative cycles of selection against a streptavidin agarose matrix.

Introduction

Techniques such as phage (*1*) and mRNA display (*2-4*) now make it possible to generate very large peptide and protein libraries and sieve them for functional molecules. Presently, the chemical diversity that may be programmed into these libraries is limited to the 20 naturally occurring amino acid side chains. Over 10 years ago, Schultz and co-workers demonstrated that unnatural amino acids, residues that do not occur in normal proteins, may be inserted into proteins if they are escorted to the ribosome by an orthogonal suppressor tRNA recognizing a stop codon (*5, 6*). This unnatural strategy allows proteins to be constructed that contain a novel residue at a desired location, enabling insertion of affinity tags (*7*), spectroscopic probes (*8*), and analogues for detailed mechanistic analysis (*9*). This suppression strategy has also been extended to eukaryotic systems both *in vitro* (*10, 11*) and *in vivo* (*12, 13*).

Recently, there has been great interest in extending the unnatural strategy to systems where the new residue may be selected for its function. Here, we demonstrate that *in vitro* selection experiments can be performed on mRNA display libraries containing the unnatural amino acid biocytin (Figure 2.1A). In mRNA display, translation extracts are used to generate combinatorial libraries of peptides and proteins covalently fused to their own mRNA via a 3'-puromycin. These libraries are strictly monovalent and provide for the synthesis of more than 10^{13} independent peptide or protein sequences in a selectable format (*14-18*).

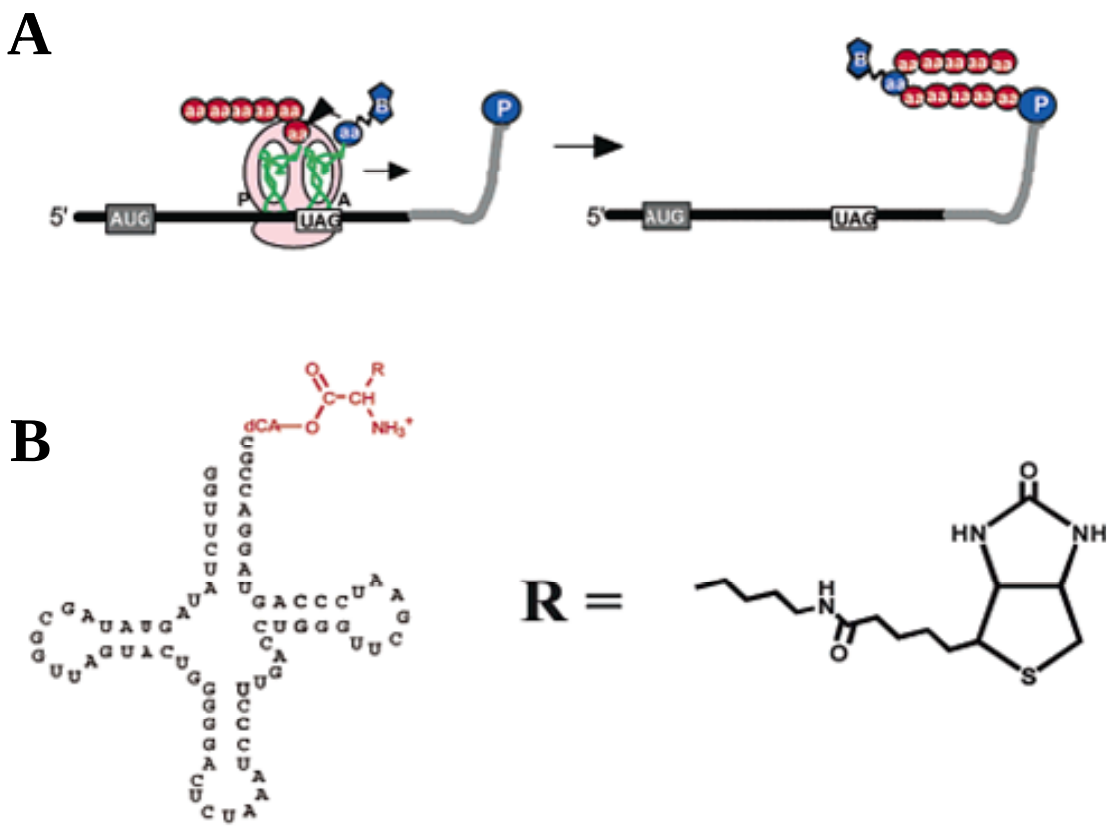

Figure 2.1: (A) Scheme for insertion of unnatural amino acids into mRNA display libraries via amber suppression. (B) Biocytin-charged THG73

Results and Discussion

We began by working to synthesize mRNA-peptide fusions containing an unnatural residue. Biocytin, a biotin derivative of lysine, represented an excellent choice for our target residue as this has been inserted into proteins previously (*7*) and the biotin moiety could be readily used to select sequences that have incorporated this amino acid. We normally construct mRNA display libraries in the rabbit reticulocyte translation extract due to the excellent stability of the template in this media and the efficiency of fusion formation. We chose the amber suppressor tRNA THG73 (a modified *Tetrehymena thermophila* Gln tRNA) to insert our unnatural residues by nonsense UAG suppression (Figure 2.1B) as this construct has high efficiency in eukaryotic translation systems (*12*).

We next constructed two templates to test insertion of the unnatural residue. The first template (Pep1) is a control containing all 20 amino acids, but no stop codon, while the second template (Pep2) contains a similar amino acid composition and a single UAG stop codon at the third position (Figure 2.2A). For Pep2, fusion formation occurs only when the suppressor tRNA is added, consistent with incorporation of biocytin into the Pep2 mRNA-peptide fusion (Figure 2.2B).

We wished to demonstrate that this approach could be used to select peptides from libraries based on the function of the unnatural residue. To do this, we replaced the TTG codon that encoded the Trp residue in position 8 of the template pep1 with an NNS saturation cassette containing 32 possible codons encoding all 20 natural amino acids and the UAG stop (Lib1). We then performed two rounds of *in vitro* selection using streptavidin-agarose as our affinity matrix and biocytin as the 21st amino acid.

A

Pep1 :MGRQEIHWANDLCKPFWVYTSGGGG

Pep2 :MGZLDYKDEDKRQEIHLADLCKPFWVYTSGGGG

B

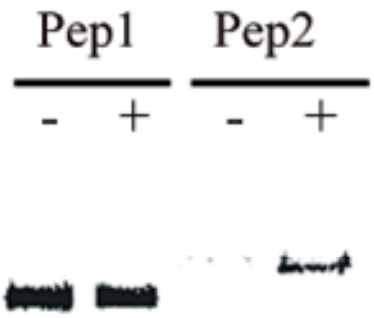

Figure 2.2: Dependence of fusion formation on addition of biocytin-charged THG73. (A) Sequences of fusion templates tested. (B) Gel showing the formation of fusion products, labeled with ^{35}S-methionine, in the presence (+) or absence (-) of biocytin-charged THG73. Fusion formation on the stop codon-containing template (Pep2) occurs only in the presence of the suppressor tRNA.

Sequencing after one round of selection indicated that UAG stop codons were being enriched at both the randomized position and elsewhere in the open reading frame via point mutations. After a second round of selection versus streptavidin agarose, nine clones were sequenced from the library. Eight out of nine (88%) contain a UAG stop codon at the randomized position or elsewhere, including two that contain a GAG-to-UAG transversion at position 5 (Figure 2.3).

Our experiments represent the first combination of an *in vitro* selection experiment and nonsense suppression of which we are aware. Our approach now allows selectable peptide and protein libraries to be constructed containing any unnatural amino acid that is compatible with the translation apparatus. These libraries should facilitate the discovery of novel ligands with functionalities beyond those provided by the 20 naturally occurring residues. For example, peptides containing N-substituted amino acids are protease-resistant (*19*) and can show enhanced affinity for natural protein interaction modules such as SH3 domains (*20*). Our approach should also aid physical organic dissection of protein interfaces, particularly where molecular interactions display context dependence. Additionally, this approach provides a convenient way to construct new molecular tools based on known pharmacaphores, in that therapeutically useful small molecules may be presented adjacent to the chemical diversity present in a 10^{13}-member peptide or protein library. Finally, translation systems that allow insertion of two or more unnatural amino acids (*21*) now provide the intriguing possibility for construction of wholly unnatural libraries in a selectable mRNA display format.

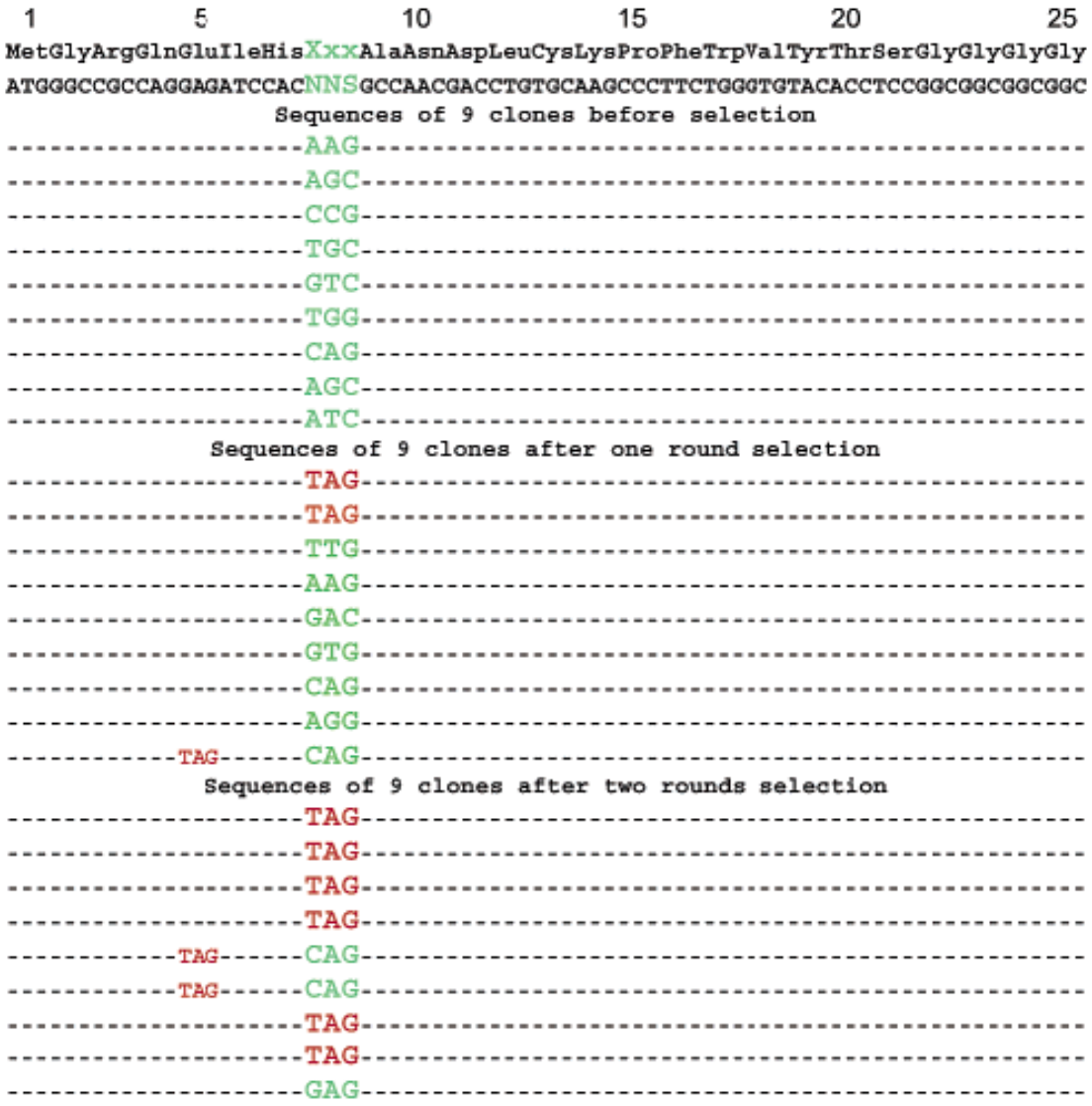

Figure 2.3: Sequences present in Lib1 before and after selection vs. streptavidin-agarose. The Lib1amino acids sequence before selection is shown in the second line and the DNA sequence is shown in the third line. Xxx represents all 20 amino acid residues plus the UAG stop codon. N is an equal amount of all four nucleotides, and S is 50% G plus 50% C in that position. The sequences in the NNS saturation region are highlighted in green, and the emerging UAG stop codons are highlighted in red. Sequences the same as the original template are labeled with a dash (-).

Materials and Methods

Synthesis of Biocytin-tRNA Suppressor

Synthesis of NVOC-Biocytin Cyanomethyl Ester

Biocytin (100 mg, 0.26 mmol, Molecular Probes) and sodium carbonate (56 mg, 0.54 mmol) were dissolved in a mixture of water (15 ml) and THF (10 ml). A solution of 6-nitroveratryloxycarbonyl chloride (NVOC-Cl) (74 mg, 0.26 mmol, Sigma) in 10 ml THF was added slowly. After 3 hr, solvents were removed *in vacuo*. Then, 3 ml of dry DMF, 3 ml of chloroacetonitrile, and 800 μL triethylamine (TEA) were added to the residue. After overnight stirring, the solvents were removed *in vacuo* and the remaining solid was purified by flash chromatography (silica gel, 10% MeOH in CH_2Cl_2).

Preparation of NVOC-Biocytin-pdCpA

The tetrabutylammonium salt of dinucleotide (pdCpA) was a gift from Prof. Dennis Dougherty's group (Division of Chemistry and Chemical Engineering, California Institute of Technology). At room temperature, 10 mg pdCpA (8.3 μmol) and 16 mg NVOC-biocytin cyanomethyl ester (25 μmol) were mixed in dry DMF under argon. 20 μL ammonium acetate (25 mM, pH 4.5) was added to quench the reaction after 1 hr. The crude product was purified by semi-preparative RP-HPLC using a gradient from 25 mM NH4OAc (pH=4.5) to CH_3CN. The appropriate fractions were combined and lyophilized. The resulting solid was re-dissolved in 10 mM acetic acid/CH3CN and lyophilized to give 3 mg biocytin-pdCpA as a pale yellow solid. The product was confirmed by ESI-MS: $[M+H]^+ = 1230.4$.

In vitro Transcription of tRNA

THG73 tRNA was synthesized *in vitro* from FokI linearized plasmid harboring THG73 tRNA gene (gift from Prof. Dennis Dougherty's group) using T7 MEGAshortscript kit (Ambion). The product was purified by polyacrylamide gel electrophoresis and dissolved in water.

Ligation of NVOC-Biocytin-pdCpA to THG73 tRNA

The mixture of THG73 tRNA (25 μg in 10 μL water) and HEPES-KOH (20 μL, 10 mM, pH 7.5) was heated at 94 ºC for 3 min and cooled down to 37 ºC slowly. 8 μL NVOC-biocytin-pdCpA (3 mM in DMSO), 32 μL 2.5X reaction buffer (25 μL 400 mM HEPES-KOH (pH 7.5); 10 μL 100 mM DTT; 25 μL 200 mM MgCl2; 3.75 μL 10 mM ATP; 10 μL 5 mg/ml BSA; 26.25 μL water; 1 μL RNasin (Promega)), 5 μL water, as well as 5 μL T4 RNA ligase (New England Biolabs) was added. After 1 hr at 37 ºC, the reaction mixture was extracted once with an equal volume of phenol (saturated with 300 mM sodium acetate, pH 5.0):CHCl3:isoamyl alcohol (25:24:1), then precipitated with 3 volume of cold ethanol at -20 ºC. The precipitate was washed with cold 70% (v/v) ethanol, dried under vacuum, and resuspended in 5 μL 1mM sodium acetate (pH=5.0). The amount of NVOC-biocytin-tRNA was quantified by measuring A260 and the concentration was adjusted to 1 μg/μL with 1 mM sodium acetate (pH=5.0). Prior to the suppression reaction, the NVOC-biocytin-tRNA was deprotected by xenon lamp equipped with a 315 nm cut-off filter for 5 minutes.

General procedure to make mRNA-peptide fusions

Construction of Fusion Template

Synthetic DNA templates–including Pep1 (5'- ACTATTTACAACCACCATGGGCCGCCAGGA GATCCACTGGGCCAACGACCTGTGCAAGCCCTTCTGGGTGTACACCTCC-3'), Pep2 (5'-ACTAT TTACAACCACCATGGGCTAGCTTGACTACAAGGACGAGGACAAGCGCCAGGAGATCCACTTG GCCAACGACCTGTGCAAGCCCTTCTGGGTGTACACCTCC-3'), and Lib1 (5'-ACTATTTACAACCACCATGGGCCGCCAGGAGATCCACNNSGCCAACGACCTGTGCAAGCCCTT CTGGGTGTACACCTCC-3')–were purified by preparative polyacrylamide gel electrophoresis. Polymerase chain reaction (PCR) of these templates with two synthetic primers, sd2 (5'-GGATTCTAATACG ACTCACTAT AGGGACAATTAC TATTTACAACCACCATG-3') and sd26 (5'-GCCGCCGCCGCC GGAGGTGTACACCCAGAAG-3'), generates double-stranded DNA. mRNA was produced by T7 runoff transcription of these templates in the presence of RNAsecure (Ambion) followed by size exclusion column purification (NAP25 column, Amersham Pharmacia Biotech). The flexible DNA linker containing puromycin, F30P (5'-dA21[C9]3dACdCP; C9=triethylene glycol phosphate, Glen Research; P = CPG-puromycin, Glen Research), was synthesized using standard chemistry. The oligonucleotide was chemically phosphorylated using phosphorylation reagent II (Glen Research) and purified by OPC cartridge. Ligation of pF30P to transcribed mRNA was done by mixing mRNA, pF30P, and a splint (5'-TTTTTTTTTTTTGCCGCCGCCGCC-3') in a 1:0.5:1.2 ratio with 2 units of T4 DNA ligase (New England Biolabs) per picomole of template mRNA. After ligation, the fusion template was gel-purified, electroeluted, and desalted by ethanol precipitation.

Translation and Fusion Formation

The fusion template was translated in reticulocyte lysate (Novagen) using standard conditions (800 nM template) with the addition of ^{35}S-methionine as the labeling reagent. In the case of templates containing UAG stop codon, 2 μg of biocytin-tRNA suppressor was also added. On completion of translation, fusion formation was stimulated by addition of $MgCl_2$ and KCl to 50 mM and 0.6 M, respectively, followed by incubation at -20 ºC overnight. The resulting ^{35}S-labeled mRNA peptide fusions were directly loaded to 15% tricine SDS-PAGE for separation. After running, the gel was dried and exposed to phosphor screen (Molecular Dynamics) for several hours. The phosphor screen was then scanned to give the image shown on Figure 2.2B.

Enrichment of UAG Stop Codon by Selection against Streptavidin-Agarose Matrix

Template-Based (dT) Purification

To isolate fusion, the lysate was diluted in binding buffer (1M NaCl, 20 mM Tris (pH= 8.0), 1 mM DTT, 10 mM EDTA, 0.2% Triton X-100) and incubated with dT-cellulose at 4 ºC for 1 hr. Bound fusions were washed with washing buffer (0.3M NaCl, 20 mM Tris (pH=8.0)) and eluted by ddH2O.

Reverse Transcription and Selective Step

After dT purification, the purified fusions were concentrated and used for reverse transcription with Superscript II RNase H$^-$ reverse transcriptase (BRL, life Technologies) following the standard conditions recommended by the manufacturer. The reaction mixture (50 μL) was directly added into 1 ml phosphate buffer (50 mM, (pH=7.0)) and streptavidin-agarose matrix (Pierce). After 1 hr at 4 ºC, the matrix was washed 6 times with 500 μL washing buffer (50 mM phosphate (pH=8.0), 100 mM NaCl, 0.1% SDS).

The matrix was then used for PCR amplification with sd2 and sd26. The PCR product was cloned with TOPO Clone kit (Invitrogen) for sequencing.

References

(1) Smith, G. P., and Petrenko, V. A. (1997) Phage Display. *Chem Rev 97*, 391-410.

(2) Nemoto, N., Miyamoto-Sato, E., Husimi, Y., and Yanagawa, H. (1997) In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett 414*, 405-8.

(3) Roberts, R. W. (1999) Totally in vitro protein selection using mRNA-protein fusions and ribosome display. *Curr Opin Chem Biol 3*, 268-73.

(4) Roberts, R. W., and Szostak, J. W. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci U S A 94*, 12297-302.

(5) Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C. J., and Schultz, P. G. (1991) Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. *Methods Enzymol 202*, 301-36.

(6) Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989) A general method for site-specific incorporation of unnatural amino acids into proteins. *Science 244*, 182-8.

(7) Gallivan, J. P., Lester, H. A., and Dougherty, D. A. (1997) Site-specific incorporation of biotinylated amino acids to identify surface-exposed residues in integral membrane proteins. *Chem Biol 4*, 739-49.

(8) Cornish, V. W., Benson, D. R., Altenbach, C. A., Hideg, K., Hubbell, W. L., and Schultz, P. G. (1994) Site-specific incorporation of biophysical probes into proteins. *Proc Natl Acad Sci U S A 91*, 2910-4.

(9) Arslan, T., Mamaev, S. V., Mamaeva, N. V., and Hecht, S. M. (1997) Structurally Modified Firefly Luciferase. Effects of Amino Acid Substitution at Position 286. *J Am Chem Soc 119*, 10877-10887.

(10) Bain, J. D., Diala, E. S., Glabe, C. G., Dix, T. A., and Chamberlin, A. R. (1989) Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J Am Chem Soc 111*, 8013-8014.

(11) Karginov, V. A., Mamaev, S. V., and Hecht, S. M. (1997) In vitro suppression as a tool for the investigation of translation initiation. *Nucleic Acids Res 25*, 3912-6.

(12) Saks, M. E., Sampson, J. R., Nowak, M. W., Kearney, P. C., Du, F., Abelson, J. N., Lester, H. A., and Dougherty, D. A. (1996) An engineered Tetrahymena tRNAGln for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression. *J Biol Chem 271*, 23169-75.

(13) Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of Escherichia coli. *Science 292*, 498-500.

(14) Barrick, J. E., Takahashi, T. T., Ren, J., Xia, T., and Roberts, R. W. (2001) Large libraries reveal diverse solutions to an RNA recognition problem. *Proc Natl Acad Sci U S A 98*, 12374-8.

(15) Cho, G., Keefe, A. D., Liu, R., Wilson, D. S., and Szostak, J. W. (2000) Constructing high complexity synthetic libraries of long ORFs using in vitro selection. *J Mol Biol 297*, 309-19.

(16) Keefe, A. D., and Szostak, J. W. (2001) Functional proteins from a random-sequence library. *Nature 410*, 715-8.

(17) Liu, R., Barrick, J. E., Szostak, J. W., and Roberts, R. W. (2000) [19] Optimized synthesis of RNA-protein fusions for in vitro protein selection, in *Methods in Enzymology* (Academic Press) pp 268.

(18) Tabuchi, I., Soramoto, S., Nemoto, N., and Husimi, Y. (2001) An in vitro DNA virus for in vitro protein evolution. *FEBS Lett 508*, 309-12.

(19) Miller S. M., Simon, R.J., Ng, S., Zuckermann, R.N., Kerr, J.M., Moos, W.H. (1995) Comparison of the proteolytic susceptibilities of homologous L-amino acid, D-amino acid, and N-substituted glycine peptide and peptoid oligomers. *Drug Development Research 35*, 20-32.

(20) Nguyen, J. T., Turck, C. W., Cohen, F. E., Zuckermann, R. N., and Lim, W. A. (1998) Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science 282*, 2088-92.

(21) Hohsaka, T., Ashizuka, Y., Sasaki, H., Murakami, H., and Sisido, M. (1999) Incorporation of Two Different Nonnatural Amino Acids Independently into a Single Protein through Extension of the Genetic Code. *J Am Chem Soc 121*, 12194-12195.