Molecular Evolution Directs Protein Translation Using Unnatural Amino Acids

Vanessa E. Cox[1] and Eric A. Gaucher[2]

[1]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia
[2]School of Biology, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia

Unnatural amino acids have in recent years established their importance in a wide range of fields, from pharmaceuticals to polymer science. Unnatural amino acids can increase the number of chemical groups within proteins and thus expand or enhance biological function. Our ability to utilize these important building blocks, however, has been limited by the inherent difficulty in incorporating these molecules into proteins. To address this challenge, researchers have examined how the canonical twenty amino acids are incorporated, regulated, and modified in nature. This review focuses on achievements and techniques used to engineer the ribosomal protein-translation machinery, including the introduction of orthogonal translation components, how directed evolution enhances the incorporation of unnatural amino acids, and the potential utility of ancient biomolecules for this process.

Keywords: unnatural amino acid • directed evolution • protein translation • ancestral proteins • protein engineering

How to cite this article:
Cox, V.E. and Gaucher, E.A. 2015. Molecular evolution directs protein translation using unnatural amino acids. *Curr. Protoc. Chem. Biol.* 7:223-228.
doi: 10.1002/9780470559277.ch150115

INTRODUCTION

The cellular translation machinery has evolved to translate a canonical set of twenty amino acids (although selenocysteine and pyrrolysine are involved in protein synthesis in certain species). Despite the limited variety of chemical functional groups offered by these amino acids, nearly all organisms utilize the same set of twenty to build proteins for a wide variety of cellular applications. One might question how the functional requirements of biological life are fulfilled using such a restricted set of building blocks. Biological systems use post-translational modifications and cofactors to expand the chemical complexity of protein chemistry, but these cellular techniques can be difficult to harness in the laboratory. A similar question arises when scientists attempt to develop novel proteins in the laboratory—are we limited to 20 canonical amino acids when designing proteins? One method used by researchers to expand or manipulate protein chemistry is to directly incorporate chemical functionality during translation using unnatural amino acids, meaning amino acids other than the canonical twenty. However, unnatural amino acids are frequently rejected by the translation machinery and suffer from low translation rates. Other challenges include incorporation of an unnatural amino acid at multiple positions in the same protein or translating several types of unnatural amino acids into a single protein. Despite the challenges associated with using unnatural amino acids, many scientific fields are actively benefiting from their application (Liu and Schultz, 2010).

Unnatural amino acids offer a wide variety of chemical functionalities that researchers can exploit for a variety of uses. These include chemical tags, enhanced or altered functionality, increased stability, and post-translational modifications (see Fig. 1. For example, biotherapeutics depend on unnatural amino acids

Current Protocols in Chemical Biology 7:223-228, December 2015
Published online December 2015 in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/9780470559277.ch150115

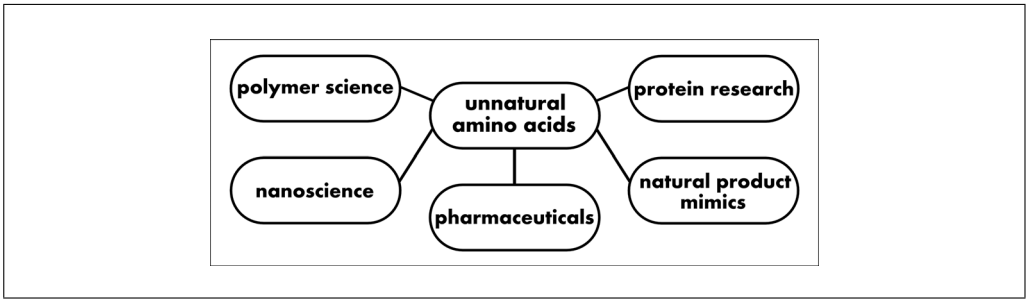

Figure 1 Applications of unnatural amino acids: a diagram illustrating research areas that employ unnatural amino acids.

to address a variety of challenges including low permeability across biological barriers and rapid degradation in vivo (Gentilucci et al., 2010; Duro-Castano et al., 2014). Unnatural amino acids are also used to chemically tag proteins at a specific location without labeling non-target amino acids or proteins. Such tagged proteins have application in a range of fields including in vivo cell imaging, proteomics, and drug-delivery systems (Stephanopoulos et al., 2010; Stephanopoulos and Francis, 2011; Elliott et al., 2014; Lang and Chin, 2014). Translation of unnatural amino acids is also used in basic protein research to directly incorporate post-translational modifications during translation, thereby offering reliable protein modification while avoiding the challenges associated with post-translational enzymatic or chemical methods (Park et al., 2011; Yanagisawa et al., 2014). Even nanoscience and polymer science are exploring the expanded possibilities that unnatural amino acids offer (Wallat et al., 2014). Due to these many applications, researchers have sought techniques to insert unnatural amino acids into peptides and proteins.

TRANSLATION OF UNNATURAL AMINO ACIDS

A number of methods exist to facilitate unnatural amino acid incorporation into proteins and peptides, including synthetic techniques such as solid-phase synthesis, in vitro translation systems, and biological methods (Merrifield, 1969; Goto et al., 2011; Smith et al., 2014). Biological approaches to incorporating unnatural amino acids are split into nonribosomal protein synthesis and ribosome-mediated translation (Walsh, 2014). This review focuses on the latter (see Fig. 2. Ribosome-mediated translation requires many cellular components including, but not limited to, the ribosome, transfer RNAs (tRNAs), aminoacyl-tRNA synthetases, and elongation factors. Engineering the ribosome translation machinery to translate unnatural amino acids has required overcoming a number of biological proofreading steps that reject unnatural amino acids during translation. Researchers have addressed these challenges using a variety of techniques including developing orthogonal translation components from other species, directed evolution of translation factors, and, most recently, ancient components of the translation machinery.

Cellular translation utilizes a number of proofreading steps during which unnatural amino acids can be detected and blocked from incorporation. Specific steps in translation allow the ribosome, tRNAs, synthetases, and elongation factor Tu (EF-Tu) to ensure accurate translation of the canonical amino acids and exclusion of any unnatural ones. In the first proofreading step, the synthetase acylates the corresponding tRNA with its cognate amino acid. Next, EF-Tu confirms that the aminoacyl-tRNA is charged with the correct amino acid and delivers the complex to the ribosome for translation. Finally, the ribosome may fail to accommodate an unnatural amino acid and would thus prevent its incorporation into a protein. Each of these three steps is a point at which an unnatural amino acid may be rejected by the translation machinery. Researchers are working to address each of these proofreading steps during translation.

Previous research has demonstrated that the ribosome is fairly promiscuous in accepting canonical and noncanonical (beta, N-methyl, etc.) amino acids. Such indiscriminate behavior has allowed researchers to focus more attention on manipulating other checkpoints in translation. For instance, tRNA synthetases, tRNAs themselves, and EF-Tus often contribute to low rates of unnatural amino acid incorporation (Josephson et al., 2005; Hartman et al., 2007; Park et al., 2011). Researchers have developed methods to engineer these components of the translation

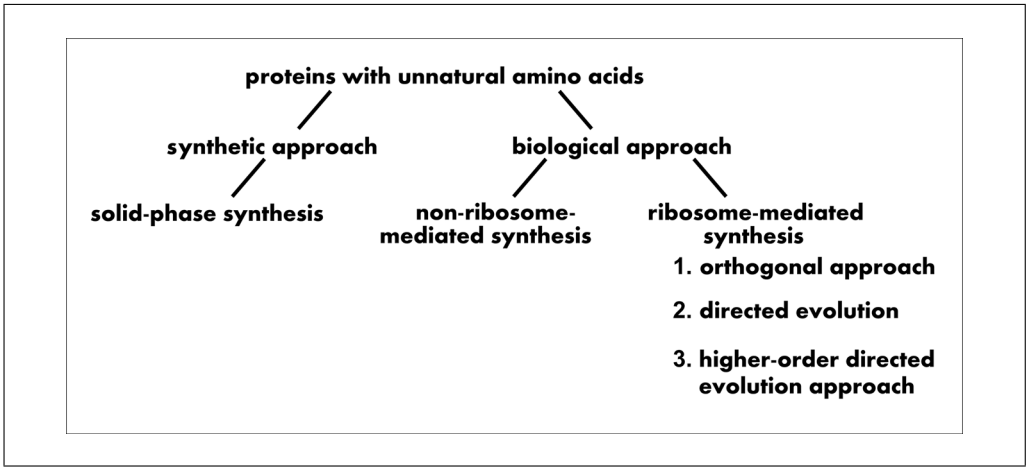

Figure 2 Techniques summary. This flow chart shows methods used to incorporate unnatural amino acids into peptides and proteins. This review focuses on methods highlighted under "ribosome-mediated synthesis" wherein the ribosome translation machinery is engineered to promote efficient translation of unnatural amino acids.

machinery. Their techniques include using orthogonal translation components, directed evolution of proteins, and ancient components of the translation system.

ORTHOGONAL TRANSLATION COMPONENTS

Great strides have been made by researchers in addressing the exclusion of unnatural amino acids by tRNAs and synthetases. Synthetases and tRNAs work in cellular translation as an orthogonal pair, specific for an amino acid, be it canonical or unnatural. To incorporate an unnatural amino acid, an orthogonal tRNA and synthetase pair is required (Wang and Schultz, 2001; Santoro et al., 2002). Although the tRNA and synthetase must recognize each other, it is critical that they not interact with other tRNAs or synthetases in the cell, since these pairs are already assigned to a specific amino acid. Rather than develop tRNAs or synthetases de novo, a common approach is to adopt tRNAs and synthetases from other domains of life. Since tRNAs and synthetases are often domain- or even species-specific, a pair from another domain will likely be orthogonal to all other tRNA/synthetase pairs in an organism's translation machinery.

Once an orthogonal pair has been identified, it is engineered to enhance the incorporation of the unnatural amino acid of choice. Briefly, the tRNA's anticodon loop must be altered to a codon that will be exclusively assigned to the unnatural amino acid, often the amber stop codon. Additionally, the synthetase must be engineered to exclusively recognize the engineered tRNA and unnatural amino acid of choice (Liu and Schultz, 2010). While this method has been used to successfully translate many unnatural amino acids, it is not sufficient for all unnatural amino acids. Even with the addition of an effectively engineered tRNA and synthetase, EF-Tu's proofreading capabilities can still prevent unnatural amino acid translation (LaRiviere et al., 2001; Ieong et al., 2012). In such cases, the orthogonal tRNA/synthetase pair method must be supplemented by the addition of an EF-Tu variant capable of accommodating the unnatural amino acid (Doi et al., 2007; Park et al., 2011).

Interestingly, nature has already provided us with one example of an elongation factor exclusive for an unnatural amino acid. SelB, a paralog of EF-Tu, is specific for what is commonly referred to as the twenty-first amino acid, selenocysteine (Bock et al., 1991). EF-Tu does not participate in the translation of selenocysteine, and SelB does not deliver canonical amino acids. Rather, organisms that can directly translate selenocysteine using the ribosomal translation machinery do so using SelB as an elongation factor exclusively for the delivery of selenocysteinyl-tRNA to the ribosome.

DIRECTED EVOLUTION

It has recently become clear that EF-Tu must be engineered to allow certain unnatural amino acids to be efficiently incorporated by the translation machinery. Phosphoserine is an example of an unnatural amino acid that requires a dedicated EF-Tu (Park et al., 2011). An EF-Tu variant specifically engineered to accommodate this unnatural amino acid can

be used in conjunction with an orthogonal tRNA/synthetase pair, creating an orthogonal triplet specific for the unnatural amino acid. To engineer EF-Tu for phosphoserine, random mutagenesis was used to alter the amino acid binding pocket of EF-Tu. An in vivo selection process was then used to identify an EF-Tu variant able to specifically recognize and deliver phosphoserine to the ribosome. Such an approach, however, would require a novel EF-Tu variant for each unnatural amino acid. It is, therefore, valuable to engineer an EF-Tu variant capable of broadly recognizing and delivering many unnatural amino acids to the ribosome. One approach to making such an elongation factor (EF) is to exploit evolutionary information from ancient EFs that existed at a time in evolutionary history when the protein-translation machinery was still developing and expanding the number of amino acids used during translation. This would require higher-order directed evolution techniques, but such an EF may in principle be more promiscuous and thus better able to deliver unnatural amino acids.

HIGHER-ORDER DIRECTED EVOLUTION

In addition to orthogonal translation components and directed evolution of proteins, fresh approaches that exploit ancient evolutionary information are proving their utility in protein engineering. Instead of using an exhaustive or random search for mutations that impart novel functionality, researchers can use ancestral sequence information to mine sequence space that has already been tested for viability by nature. These higher-order directed evolution strategies, such as ancestral sequence reconstruction (ASR), can be used to direct protein engineering to design variants with desirable phenotypes such as increased stability or promiscuity (Liberles, 2007).

ASR is a computational technique that can rewind the tape of evolution and resurrect ancient proteins that theoretically once existed. This method uses extant protein sequences to computationally determine protein sequences that existed at various speciation events (Cole and Gaucher, 2011). The advantage of this technique is that these ancestral proteins have already been tested by nature and deemed fit to survive. In several labs, ASR has been successfully used to identify protein sequences with increased thermostability, promiscuous enzymatic activity, or greater stability during directed evolution, see Figure 3 (Gaucher et al., 2003; Goldsmith and Tawfik, 2013; Risso et al., 2013).

One way ASR has facilitated protein engineering is through directing the design of more stable protein variants. For example, ancestral proteins can have increased thermostability and kinetic stability, which is thought to be derived from existing on a hotter planet during the Precambrian period (Gaucher et al., 2003; Risso et al., 2013). Over billions of years, proteins evolved under conditions that were slowly becoming more moderate in terms of temperature, leading to modern proteins that are often stable only within a very narrow temperature range.

Additionally, ASR has produced ancestral proteins that demonstrate greater functional promiscuity. Resurrected β-lactamases from the Precambrian period were not only found to confer resistance to penicillin antibiotics, but also demonstrated activity against third-generation human-made antibiotics. In contrast, modern β-lactamases have more limited enzymatic capabilities and are not effective against third-generation antibiotics (Risso et al., 2013). Since promiscuous function in enzymes is thought to be advantageous when seeking to evolve novel protein function, ancestral proteins may be better scaffolds to use in directed evolution protein engineering (Khersonsky and Tawfik, 2010).

Another technique combines directed evolution and ASR by introducing stabilizing, ancestral mutations into an evolving protein, thereby compensating for destabilizing mutations when seeking novel function. Mutations that generate novel functionality can be destabilizing, which makes their identification difficult. By bringing a protein sequence closer to the ancestral sequence, the protein can better mitigate the effects of destabilizing mutations that arise during directed evolution (Goldsmith and Tawfik, 2013). This ancestral-modern hybrid can serve as a better scaffold on which to introduce mutations during directed evolution. A more stable protein can accumulate a wider range of mutations, thereby allowing a broader search of sequence space, which in turn enhances the search for novel enzymatic function.

The increased promiscuity of ancestral proteins is a feature that may be harnessed in the design of protein variants able to translate unnatural amino acids. Although the modern translation machinery displays specificity for only twenty amino acids, ancient translation components may be more promiscuous and may accept a wider range of substrates,

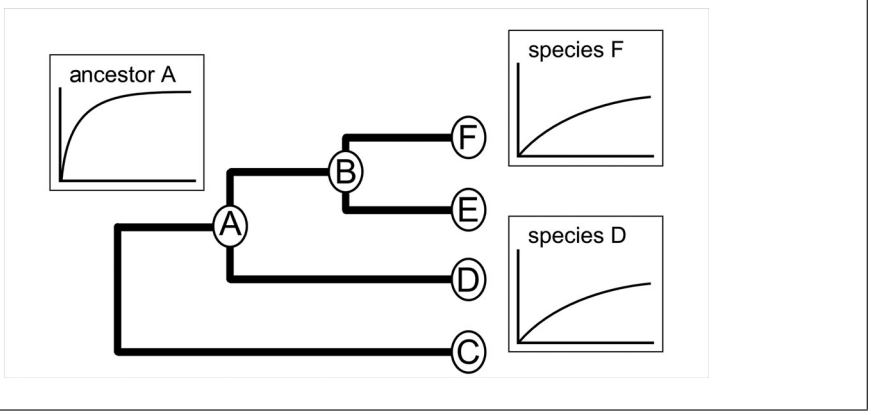

Figure 3 Phylogenetic tree showing increased K_M and V_{max} for ancestor A. This example of a protein phylogeny illustrates an improved K_M and V_{max} for ancestral proteins over modern proteins. All graphs show the rate of reaction versus substrate concentration. Modern proteins from species F and D show low K_M and V_{max}. The ancestral protein inferred at node A shows increased K_M and V_{max}.

including unnatural amino acids. Presumably, early life contended with a variety of amino acids far more diverse than what is currently found in vivo. The ribosome has been argued to be an entropy trap, suggesting it is inherently promiscuous (Sievers et al., 2004). Based upon the assumption of the ribosome's promiscuity, other components of translation may have originally been more functionally restrictive in terms of which amino acids they would incorporate into proteins, such that only a small number of amino acids (presumably less than twenty) were incorporated into proteins. Over time, the components of protein translation functionally expanded and the number of amino acids used in translation increased. This promiscuity of the translation components led to the efficient translation of increased numbers of amino acids until life settled on the canonical twenty. To maintain this restrictive list, the modern translation machinery has evolved again to be exclusive for the modern twenty amino acids. As such, ancient translation components, designed through methods like ASR, may be more promiscuous and may be able to translate a wider range of amino acids, including unnatural amino acids.

CONCLUSION

Methods to incorporate unnatural amino acids continue to evolve as our understanding of protein synthesis continues to develop. Initially, the tRNAs and synthetases were targeted; researchers knew an orthogonal pair was required for each amino acid regardless of whether that amino acid was natural or unnatural. The notion of manipulating translation components was recently expanded to include engineered EF-Tu variants, developed using directed evolution. Currently, higher-order directed evolution approaches offer ancient proteins as a new direction for the incorporation of unnatural amino acids. Ancestral sequences can offer both increased stability and promiscuity, thereby acting upon a wider range of substrates than their modern counterparts. Stabilizing, ancestral mutations can also compensate for destabilizing mutations, a vital component when seeking altered or novel function via directed evolution. Ultimately, the process by which directed evolution impacts unnatural amino acid translation will progress as our understanding of both translation and evolution continues to develop.

Acknowledgements

We thank Dieter Söll, Jesse Rinehart, and Michael Jewett for helpful discussions and materials. We also thank NASA Exobiology, NASA Astrobiology, and DuPont for funding.

Literature Cited

Bock, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B., and Zinoni, F. 1991. Selenocysteine—the 21st amino-acid. *Mol. Microbiol.* 5:515-520. doi: 10.1111/j.1365-2958.1991.tb00722.x.

Cole, M.F. and Gaucher, E.A. 2011. Utilizing natural diversity to evolve protein function: Applications towards thermostability. *Curr. Opin. Chem. Biol.* 15:399-406. doi: 10.1016/j.cbpa.2011.03.005.

Doi, Y., Ohtsuki, T., Shimizu, Y., Ueda, T., and Sisido, M. 2007. Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system. *J. Am. Chem. Soc.* 129:14458-14462. doi: 10.1021/ja075557u.

Duro-Castano, A., Conejos-Sanchez, I., and Vicent, M.J. 2014. Peptide-based polymer therapeutics. *Polymers* 6:515-551. doi: 10.3390/polym6020515.

Elliott, T.S., Bianco, A., and Chin, J.W. 2014. Genetic code expansion and bioorthogonal labelling enables cell specific proteomics in an animal. *Curr. Opin. Chem. Biol.* 21:154-160. doi: 10.1016/j.cbpa.2014.07.001.

Gaucher, E.A., Thomson, J.M., Burgan, M.F., and Benner, S.A. 2003. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* 425:285-288. doi: 10.1038/nature01977.

Gentilucci, L., De Marco, R., and Cerisoli, L. 2010. Chemical modifications designed to improve peptide stability: Incorporation of non-natural amino acids, pseudo-peptide bonds, and cyclization. *Curr. Pharm. Des.* 16:3185-3203. doi: 10.2174/138161210793292555.

Goldsmith, M. and Tawfik, D.S. 2013. Enzyme engineering by targeted libraries. *Methods Enzymol.* 523:257-283.

Goto, Y., Katoh, T., and Suga, H. 2011. Flexizymes for genetic code reprogramming. *Nat. Protoc.* 6:779-790. doi: 10.1038/nprot.2011.331.

Hartman, M.C.T., Josephson, K., Lin, C.W., and Szostak, J.W. 2007. An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PloS One* 2:e972.

Ieong, K.W., Pavlov, M.Y., Kwiatkowski, M., Forster, A.C., and Ehrenberg, M. 2012. Inefficient delivery but fast peptide bond formation of unnatural L-Aminoacyl-tRNAs in translation. *J. Am. Chem. Soc.* 134:17955-17962. doi: 10.1021/ja3063524.

Josephson, K., Hartman, M.C.T., and Szostak, J.W. 2005. Ribosomal synthesis of unnatural peptides. *J. Am. Chem. Soc.* 127:11727-11735. doi: 10.1021/ja0515809.

Khersonsky, O. and Tawfik, D.S. 2010. Enzyme promiscuity: A mechanistic and evolutionary perspective. *Ann. Rev. Biochem.* 79:471-505.

Lang, K. and Chin, J.W. 2014. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* 114:4764-4806. doi: 10.1021/cr400355w.

LaRiviere, F.J., Wolfson, A.D., and Uhlenbeck, O.C. 2001. Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science* 294:165-168. doi: 10.1126/science.1064242.

Liberles, D.A. 2007. Ancestral Sequence Reconstruction. Oxford University Press, New York.

Liu, C.C. and Schultz, P.G. 2010. Adding new chemistries to the genetic code. *Ann. Rev. Biochem.* 79:413-444.

Merrifield, R.B. 1969. Solid-phase peptide synthesis. *Adv. Enzymol. Relat. Areas Mol. Biol.* 32:221-296.

Park, H.S., Hohn, M.J., Umehara, T., Guo, L.T., Osborne, E.M., Benner, J., Noren, C.J., Rinehart, J., and Soll, D. 2011. Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* 333:1151-1154. doi: 10.1126/science.1207203.

Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A., and Sanchez-Ruiz, J.M. 2013. Hyperstability and substrate promiscuity in laboratory resurrections of precambrian beta-lactamases. *J. Am. Chem. Soc.* 135:2899-2902. doi: 10.1021/ja311630a.

Santoro, S.W., Wang, L., Herberich, B., King, D.S., and Schultz, P.G. 2002. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* 20:1044-1048. doi: 10.1038/nbt742.

Sievers, A., Beringer, M., Rodnina, M.V., and Wolfenden, R. 2004. The ribosome as an entropy trap. *Proc. Natl. Acad. Sci. U.S.A.* 101:7897-7901. doi: 10.1073/pnas.0402488101.

Smith, M.T., Wilding, K.M., Hunt, J.M., Bennett, A.M., and Bundy, B.C. 2014. The emerging age of cell-free synthetic biology. *Febs Lett.* 588:2755-2761. doi: 10.1016/j.febslet.2014.05.062.

Stephanopoulos, N. and Francis, M.B. 2011. Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* 7:876-884. doi: 10.1038/nchembio.720.

Stephanopoulos, N., Tong, G.J., Hsiao, S.C., and Francis, M.B. 2010. Dual-surface modified virus capsids for targeted delivery of photodynamic agents to cancer cells. *ACS Nano* 4:6014-6020. doi: 10.1021/nn1014769.

Wallat, J.D., Rose, K.A., and Pokorski, J.K. 2014. Proteins as substrates for controlled radical polymerization. *Polym. Chem.* 5:1545-1558. doi: 10.1039/C3PY01193C.

Walsh, C.T. 2014. Blurring the lines between ribosomal and nonribosomal peptide scaffolds. *ACS Chem. Biol.* 9:1653-1661. doi: 10.1021/cb5003587.

Wang, L. and Schultz, P.G. 2001. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* 8:883-890. doi: 10.1016/S1074-5521(01)00063-1.

Yanagisawa, T., Umehara, T., Sakamoto, K., and Yokoyama, S. 2014. Expanded genetic code technologies for incorporating modified lysine at multiple sites. *Chembiochem* 15:2181-2187. doi: 10.1002/cbic.201402266.