Expanding the Genetic Code of *Escherichia coli*

The genetic code of virtually every known organism, from bacterium to human, encodes the same twenty common amino acids. Only in rare instances, selenocysteine or pyrrolysine is used. Nonetheless, it is remarkable that polypeptides synthesized from these simple building blocks can carry out all of the complex processes of life, from photosynthesis and vision to signal transduction and the immune response. Could the genetic code be expanded to include additional amino acids? Could the properties of proteins, or possibly an entire organism, be enhanced by new amino acids with novel chemical, structural or physical properties? My dissertation focused on the development of a general strategy that makes it possible to incorporate unnatural amino acids site-specifically into proteins directly in living cells, and the application of these genetically encoded novel amino acids.

Our method involves the generation of additional components for the biosynthetic machinery, including a novel tRNA-codon pair, an aminoacyl-tRNA synthetase, and an amino acid. This new set of components must function orthogonally to the counterparts of the common 20 amino acids; that is, the orthogonal synthetase (and only this synthetase) aminoacylates the orthogonal tRNA (and only this tRNA) with the unnatural amino acid only, and the resulting acylated tRNA inserts the unnatural amino acid in response to only the unique codon (Figure 1).

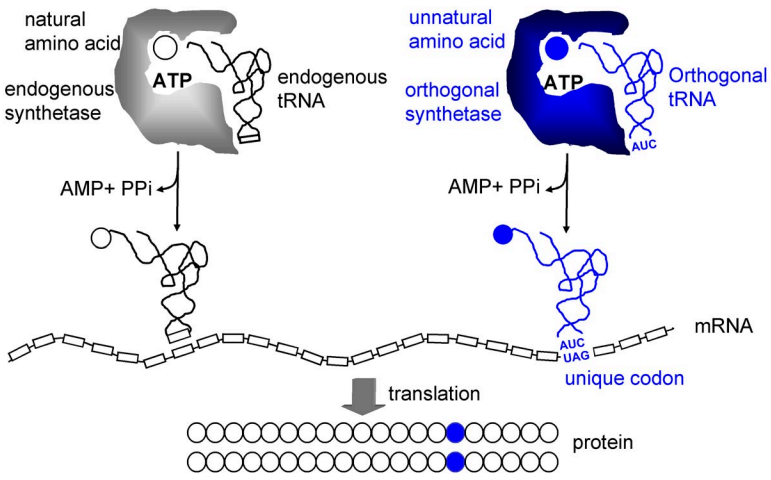

Figure 1. A general approach for the site-specific incorporation of unnatural amino acids into proteins *in vivo*.

We started our experiments using *Escherichia coli* due to its ease of genetic manipulation. The amber nonsense codon was chosen to encode the unnatural amino acid. After the orthogonalities of various tRNA/synthetase pairs from different organisms were tested *in vitro* and in *E. coli*, an orthogonal amber suppressor $tRNA^{Tyr}_{CUA}$ /TyrRS pair was generated by importing a $tRNA^{Tyr}$/TyrRS pair from the archaebacterium *Methanococcus jannaschii* (*Mj*). To reduce further the recognition of the *Mj*- $tRNA^{Tyr}_{CUA}$ by *E. coli* synthetases while maintaining its affinity for *Mj*-TyrRS, a general strategy was then developed, which consists of a combination

of negative and positive selections with a mutant suppressor tRNA library. Eleven nucleotides of the *Mj*-$tRNA^{Tyr}_{CUA}$ were randomly mutated, and from the resulting tRNA library a mutant tRNA ($mutRNA^{Tyr}_{CUA}$) was identified that has almost no affinity for *E. coli* synthetases and is still charged efficiently by the orthogonal *Mj*-TyrRS with tyrosine.

The next step is to alter the amino acid specificity of the orthogonal *Mj*-TyrRS so that it charges the $mutRNA^{Tyr}_{CUA}$ with a desired unnatural amino acid and none of the common twenty. Modification to a synthetase is expected to be difficult due to the high intrinsic fidelity of the natural synthetases and the fact that unnatural amino acids are not required for any cellular function. We have pursued a combinatorial approach to this problem, in which a pool of mutant synthetases is generated from the framework of a wild-type synthetase, and then selected based on their specificities for an unnatural amino acid relative to the common twenty. Mutant synthetases were first selected for their abilities to charge the $mutRNA^{Tyr}_{CUA}$ with either natural or unnatural amino acids, and those with specificities for natural amino acids were selected against subsequently. Genes of synthetases passing both selections were recombined by DNA shuffling, and were subjected to more rounds of selection until a synthetase with desired activity was evolved.

Mutant synthetase libraries were generated by mutating five active-site residues of *Mj*-TyrRS. After two rounds of iteration, a synthetase was evolved that when coexpressed with the $mutRNA^{Tyr}_{CUA}$, leads to the incorporation of *O*-methyl-L-tyrosine into proteins in *E. coli* in response to an amber codon with translational fidelity and yield rivaling that of natural amino acids. The result was confirmed by gel electrophoresis, Western blotting, and mass spectral analyses of the expressed mutant protein, and enzymatic assay of the evolved synthetase. The genetic code of *E. coli* was expanded for the first time. A second evolved synthetase is capable of inserting L-3-(2-naphthyl)-alanine, an amino acid structurally distinct from tyrosine, into proteins selectively, suggesting that our methodology should be generalizable to a variety of unnatural amino acids.

We subsequently explored unnatural amino acids with novel functionalities or properties, and have been successful to add more than thirteen new amino acids to the genetic code of *E. coli* to date (Figure 2). For instance, the nonproteinogenic alkene group was genetically encoded in the form of *O*-allyl-L-tyrosine, which can afford new chemistry such as olefin metathesis for protein derivatization. The keto group is ubiquitous in organic chemistry, yet is absent from common amino acids. Two keto-containing amino acids (*p*- and *m*-acetyl-L-phenylalanine) were added to the genetic code. Using keto's unique reactivity, we demonstrated that proteins could be selectively labeled with fluorophores, biotin, and carbohydrates for imaging protein conformation change, protein detection, and generation of homogeneous glycoprotein mimetics, respectively. Two heavy atom containing amino acids (*p*-bromo and *p*-iodo-L-phenylalanine) were site-specifically incorporated into proteins, providing a general and reliable method for preparing suitable isomorphous heavy atom derivatives of proteins, which is of fundamental importance in protein crystallography. Photocrosslinker *p*-benzoyl-L-phenylalanine was also included to map protein-protein interactions.

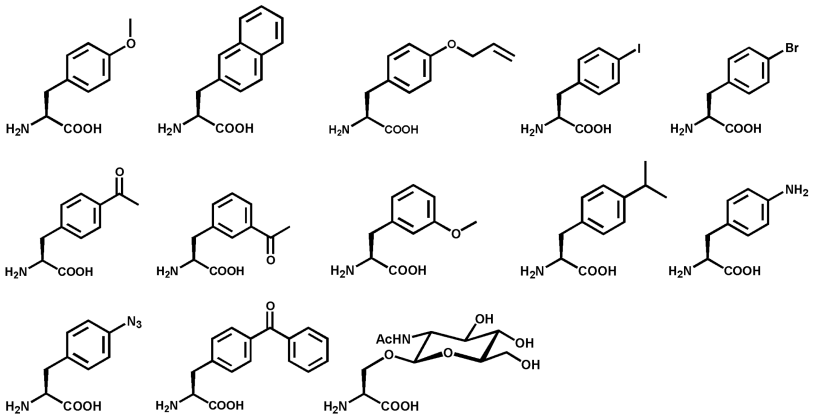

Figure 2. Structure of unnatural amino acids that have been added to the genetic code of *E. coli* to date.

To begin to study the effects of such novel amino acids on protein structure and function, we selectively substituted Tyr66 of the green fluorescent protein (GFP) with five unnatural amino acids. The spectral properties of the mutant GFPs, including the absorbance and fluorescence maxima and quantum yields, correlate with the structural and electronic properties of the substituents on the amino acids. These mutant GFPs may further expand the applications of this class of bioluminescent proteins.

For evolutionary studies, it would be ideal for the host cell to be able to synthesize the unnatural amino acid from simple carbon sources itself. Biosynthetic pathways were generated for *O*-methyl-L-tyrosine and *p*-amino-L-phenylalanine in *E. coli* to create a completely autonomous 21-amino-acid bacterium. Such organisms may provide an opportunity to examine the evolutionary consequences of adding novel amino acids to the genetic repertoire.

In summary, we have shown that it is possible to augment the protein biosynthetic machinery of *E. coli* to accommodate additional genetically encoded amino acids. This methodology should be generalizable to other amino acids as well as other cell types. It is providing powerful new tools for analyzing protein and cellular function both *in vitro* and *in vivo*, and an opportunity for directed protein and organism evolution with unnatural amino acids in search for properties not existed before.