UNIVERSITY OF CALIFORNIA, SAN DIEGO

Expanding the genetic code of *Caenorhabditis elegans* by incorporating unnatural amino acids

A dissertation submitted in partial satisfaction of the requirements for degree Doctor of Philosophy

in

Biology

by

Angela Rose Parrish

Committee in charge:

Professor Lei Wang, Chair
Professor Andrew Dillin
Professor Yishi Jin
Professor William McGinnis
Professor Yitzhak Tor

2011

The Dissertation of Angela Rose Parrish is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2011

DEDICATION

This dissertation is dedicated to the memory of my grandfather, Severino Santos, and to my father, Mark Parrish, both of whom taught me at an early age never to believe anything that is told to me, but instead to be skeptical of all claims. Memories of the unbelievable stories I fell for now make me search for my own explanations, rather than rely on the explanations of others, and have driven me to be the scientist I am today.

EPIGRAPH

"Results! Why, man, I have gotten a lot of results.
I know several thousand things that won't work."

Thomas Edison

"I have yet to see any problem, however complicated, which, when you looked at it in the right way, did not become still more complicated."

Paul Alderson

"The stereotype of scientists being scruffy nerds with rows of pens in their top pocket is just about as wicked as racist stereotypes."

Richard Dawkins

"Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so that we may fear less."

Marie Curie

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge my committee for their support throughout my graduate career, because without their guidance, advice, optimism, and kind words, this dissertation may never have reached completion.

I would also like to acknowledge my fellow lab mates in the Wang Lab, whose support and filthy jokes allow me to maintain a modicum of sanity. You are truly my second family, especially Jeff Takimoto and David Johnson, the best fake brothers a person could ask for. I would especially like to acknowledge Xingyu She, whose help has proved invaluable and who saved me from being the lone *C. elegans* researcher in our lab, and Vanessa Lacey, the best workout partner I've ever had.

I must acknowledge my husband, Adam Paré, as we both know I would never have made it through graduate school without his stoicism and love. I was able to do this because of you take care of me better than anyone.

The first section of the Introduction (unnatural amino acids) is, in part, a reprint of material as it appears in *Comprehensive Natural Products Chemistry II;* Chemistry and Biology 2010. Parrish, Angela; Wang,Lei., Oxford; volume 5, 2010. The dissertation author was the primary author of this material.

The majority of this dissertation work has been submitted for publication of the material. Parrish, Angela; She, Xingyu; Xiang, Zheng; Coin, Irene; Dillin, Andrew; Wang, Lei. The dissertation author was the primary investigator and author of this material.

VITA

2002	Research Volunteer, Wellcome/Trust Center, Cambridge University
2003	Bachelor of Science, Cornell University
2003-2005	Research Assistant, Georgetown University
2011	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Lacey, V.K., Parrish, A.R., Has, S., Shen, Z., Briggs, S.P., Ma, Y. and Wang, L. (2011). A Fluorescent Reporter of the Phosphorylation Status of the Substrate Protein STAT3. *Angewandte Chemie International Edition* 50(37), 8692-8696.

Parrish, A.R. and Wang, L. Genetic Incorporation of Unnatural Amino Acids into Proteins. (2010) – in *Comprehensive Natural Products Chemistry II*; Chemistry and Biology; Mander, L., Lui, H.-W., Eds.; Oxford; volume 5, pp.587–617

Miermont, A.M., Parrish, A.R., and Furth, P.A. (2010). Role of ERalpha in the differential response of Stat5a loss in susceptibility to mammary preneoplasia and DMBA-induced carcinogenesis. *Carcinogenesis* 31(6), 1124-31.

Wang Q., Parrish, A.R., and Wang, L. (2009). Expanding the Genetic Code for Biological Studies. *Chemistry & Biology* 16(3), 323-36.

Tilli, M.T., Parrish, A.R., Cotarla, I., Jones, L.P., Johnson, M.D., and Furth, P.A. (2008). Comparison of mouse mammary gland imaging techniques and applications: reflectance confocal microscopy, GFP imaging, and ultrasound. *BMC Cancer* 8(21).

Jones, L.P., Tilli, M.T., Assefnia, S., Torre, K., Halama, E.D., Parrish, A., Rosen E.M., and Furth, P.A. (2008) Activation of estrogen signaling pathways collaborates with loss of Brca1 to promote development of ERα negative and ER α positive mammary preneoplasia and cancer. *Oncogene* 27(6), 794-802.

Tilli, M.T., Cabrera, M.C., Parrish, A.R., Torre, K.M., Sidawy, M.K., Gallagher, A.L., Makariou, E., Polin, S.A., Liu, M.C., and Furth, P.A. (2007) Real-time imaging and characterization of human breast tissue by reflectance confocal microscopy.

Journal of Biomedical Optics 12(5).

Parrish, A.R., Wang, W., and Wang, L. (2006) Manipulating Proteins for Neuroscience. *Current Opinion in Neurobiology* 16(5), 585-92.

Parrish, A., Halama, E., Tilli, M.T., Freedman, M., and Furth, P.A. (2005) Reflectance Confocal Microscopy for Characterization of Mammary Ductal Structures and Development of Neoplasia in Genetically Engineered Mouse Models of Breast Cancer. *Journal of Biomedical Optics* 10(5).

ABSTRACT OF THE DISSERTATION

Expanding the genetic code of *Caenorhabditis elegans* by incorporating unnatural amino acids

by

Angela Rose Parrish

Doctor of Philosophy in Biology

University of California, San Diego, 2011

Professor Lei Wang, Chair

The genetic code specifies 20 common amino acids and is largely preserved in both single and multicellular organisms. Unnatural amino acids have been genetically incorporated into proteins by using engineered tRNA/aminoacyl-tRNA synthetase pairs in single cells, enabling new research capabilities and precision inaccessible with common amino acids. Here I show that unnatural amino acids can be genetically incorporated into the coding region of proteins in *Caenorhabditis elegans*, a multicellular model organism extensively used to study basic biology and human diseases. I established that *Escherichia coli* tyrosyl and leucyl amber suppressor tRNA/synthetase pairs can be evolved to incorporate different unnatural amino acids in response to the amber stop codon (UAG) into various proteins in worms. To accurately report unnatural amino acid

incorporation, I found that it is crucial to integrate the UAG-containing reporter gene into the genome rather than to express it on an extrachromosomal array, where expression can be variable. Unnatural amino acid incorporation is greatly affected by bioavailability, which can be improved by delivering in a dipeptide form. Incorporation levels also depend on dosage, exposure time, and tRNA copy number. Unnatural amino acid incorporation can be improved by stabilizing UAG-containing mRNA by RNAi knockdown of *smg-1*. I have generated stable transgenic worms capable of genetically encoding unnatural amino acids, enabling exploitation of new functional groups to address complex biological problems in a metazoan. I anticipate my strategies will be generally applicable to other multicellular organisms.

INTRODUCTION

1. UNNATURAL AMINO ACIDS

With few exceptions, the genetic code is universal. Codons in different organisms have identical meanings, specifying the insertion of one of the canonical 20 amino acids or translational termination. Although amino acids different from the canonical 20 have been discovered in some organisms, only two amino acids, selenocysteine (Bock et al., 1991) and pyrrolysine (Srinivasan et al., 2002), are genetically incorporated into proteins in response to a stop codon in specific mRNA locations. Built by the canonical 20 amino acids, proteins carry out a variety of functions for organismal needs. Nonetheless, proteins are somewhat limited in the reactions they can participate in alone, requiring cofactors and post-translational modifications to fulfill many natural functions.

The introduction of new functional groups apart from those found in the common 20 amino acids into proteins can provide specific and malleable tools to researchers for probing protein structure and function, as well as eventually generating novel protein activities. There are a number of methods to introduce unnatural amino acids (Uaas) into proteins, which can be broadly cataloged as chemical approaches and biosynthetic approaches (Wang and Schultz, 2004). Chemical approaches rely on chemistry to prepare or modify the protein. For instance, reactive side chains of exposed amino acids in proteins can be chemically modified for derivatization (Carne, 1994). Complete chemical

synthesis has been used to introduce Uaas into peptides and small proteins (Kent, 1988). Semisynthetic protein ligation methods, in which two or more protein fragments are chemically ligated to make the full-length protein, further enable Uaas to be introduced into large proteins (Muir, 2003).

On the other hand, biosynthetic approaches harness the endogenous cellular machinery to translate protein and to introduce Uaas. For instance, by using bacterial strains auxotrophic for a particular amino acid, close structural analogues of this amino acid can globally replace the amino acid in proteins (Hortin and Boime, 1983). An *in vitro* biosynthetic method, in which a suppressor tRNA is chemically acylated with an Uaa, allows for the site-specific incorporation of Uaas into proteins in cell extracts supporting translation (Bain et al., 1989; Noren et al., 1989). All of these methods have been proven valuable but are limited by the lack of site selectivity, the *in vitro* nature of the method, or the low incorporation efficiency.

A challenge is to genetically incorporate Uaas at any specified site in the proteome with translational fidelity and efficiency parallel to that of common amino acids - in essence, to expand the genetic code to include Uaas *in vivo*. There are four requirements that must be fulfilled for the genetic incorporation of Uaas *in vivo* to be both specific and effective (Figure 1) (Wang and Schultz, 2002): 1) a tRNA that no endogenous synthetases (RS) will charge with natural amino acids (an orthogonal tRNA) and that decodes a unique codon; 2) a unique codon to signal insertion of the Uaa in the mRNA, such that no endogenous

tRNAs will decode the unique codon with a natural amino acid; 3) an aminoacyl-tRNA synthetase that is specific for the Uaa and charges only the orthogonal tRNA (an orthogonal RS); and 4) Uaa availability in the cytoplasm of the cells of interest, so that the orthogonal synthetase has access to the molecule for acylation. Once these criteria have been met, the orthogonal synthetase will charge the orthogonal tRNA with the desired Uaa only, and the unnatural amino acid-acylated tRNA will incorporate the Uaa in response to the unique codon into proteins by cooperating with the endogenous protein translational machinery. The introduced components function in a similar manner as the counterparts that genetically encode common amino acids. This method is particularly powerful because it enables exploitation of novel properties of Uaas *in vivo*, and has the potential to be applied to almost all genetically tractable organisms.

Efforts to develop a general method for genetically encoding Uaas in live cells first focused on *Escherichia coli* because it is easily manipulated genetically and its translational machinery had been extensively studied. *E. coli* has high transformation efficiency for introducing large DNA libraries, and many genetic selections or screens have been successfully applied in this organism to identify distinct phenotypes. These features of *E. coli* facilitated the generation of orthogonal tRNA, codon, and synthetase, critical components for incorporating Uaas *in vivo*.

A unique codon in mRNA is required to specify the Uaa, and an orthogonal tRNA able to recognize this unique codon is needed for decoding. The

sense codons in the genetic code cannot be used for this purpose because they already all code for an amino acid. Using a “blank” stop codon to encode the Uaa is an attractive strategy, because the orthogonal tRNA introducing the Uaa doesn’t need to compete with an endogenous tRNA. However, the tRNA will still compete with the release factors that terminate protein translation. Initially the amber stop codon (UAG) was selected to specify the Uaa. Among the three stop signals, the amber codon is the least used stop codon in *E. coli*, *Saccharomyces cerevisiae,* and *Caenorhabditis elegans*. Amber suppressor tRNAs, tRNAs capable of reading through the amber stop codon by incorporating a natural amino acid, have been identified in *E. coli*, and these suppressor tRNAs do not significantly affect the growth rates of the host *E. coli* cells (Benzer and Champe, 1962). Moreover, tRNAs have also been engineered to suppress the amber codon for amino acid mutagenesis in proteins (Bossi and Roth, 1980). To make a tRNA recognize the amber stop codon is also straightforward; the anticodon of the tRNA simply needs to be changed to CUA to be complementary to the amber codon UAG.

There are multiple tRNAs and synthetases inside cells, and the synthetase recognizes its cognate tRNAs through specific interactions at various regions of the tRNA (Ibba and Soll, 2000). An orthogonal tRNA for incorporating the Uaa should not interact with any endogenous synthetases, and an orthogonal synthetase should not interact with any endogenous tRNAs. Meanwhile, both the orthogonal tRNA and the synthetase should be compatible with all other

components in the translational machinery. These stringent criteria and the multiple component interactions inside cells make it very challenging to generate an orthogonal tRNA-synthetase pair. Rational engineering of an *E. coli* glutaminyl tRNA-synthetase pair produced an orthogonal tRNA, but no orthogonal synthetase could be generated to selectively recognize this tRNA (Liu et al., 1997).

A successful strategy is to import a tRNA-synthetase pair from species in a different kingdom of life. This strategy is based on the observation that *in vitro* cross-aminoacylation using tRNA-synthetase from different kingdoms is often low. Attention was focused on archaea as a source of orthogonal tRNA-synthetase pairs for use in *E. coli*. Archaeal aminoacyl-tRNA synthetases are more similar to their eukaryotic than prokaryotic counterparts (Ibba and Soll, 2000). Moreover, early work indicated that most tRNAs from the halophile *Halobacterium cutirebrum* cannot be charged by *E. coli* aminoacyl-tRNA synthetases (Kwok and Wong, 1980). The first orthogonal *E. coli* tRNA-synthetase pair generated from archaeal bacteria was derived from the tyrosyl pair taken from *Methanococcus jannaschii* (Wang et al., 2000). *In vitro* experiments showed that the major recognition elements of *M. jannaschii* $tRNA^{Tyr}$ include the discriminator base A73 and the first base pair, C1-G72, in the acceptor stem. The anticodon triplet participates only weakly in identity determination. By contrast, *E. coli* $tRNA^{Tyr}$ uses A73, G1-C72, a long variable arm, and the anticodon as identity elements. The *M. jannaschii* tyrosyl-tRNA synthetase (*Mj* TyrRS) also has a minimalist

anticodon loop binding domain (Steer and Schimmel, 1999), making it possible to change the anticodon loop of its cognate tRNA to CUA with little loss in affinity by the synthetase. In addition, the TyrRS does not have an editing mechanism, which removes amino acids misacylated onto the cognate $tRNA^{Tyr}$. This lack of the editing function can prevent the Uaa from being deacylated from the orthogonal tRNA. Indeed, an amber suppressor *M. jannaschii* $tRNA^{Tyr}_{CUA}$ ($MjtRNA^{Tyr}_{CUA}$) and its cognate *Mj* TyrRS were shown to function efficiently in translation in *E. coli*, but some degree of aminoacylation of this $MjtRNA^{Tyr}_{CUA}$ by endogenous *E. coli* synthetases was observed (Wang and Schultz, 2001).

A general strategy was developed to evolve orthogonal tRNAs in *E. coli* from heterologous precursors (Wang and Schultz, 2001). In this method, a combination of negative and positive selections is applied to a library of tRNA mutants derived from a heterologous suppressor tRNA in the absence and presence of the cognate synthetase, respectively. The tRNA library is first introduced into *E. coli* along with a mutant barnase gene in the negative selection. Amber nonsense codons are introduced in the barnase gene at sites permissive to substitution by other amino acids. When a member of the suppressor tRNA library is aminoacylated by any endogenous *E. coli* synthetase (i.e. it is not orthogonal to the *E. coli* synthetases), the amber codons are suppressed and the ribonuclease barnase is produced, resulting in cell death. Only cells harboring orthogonal or nonfunctional tRNAs can survive. All tRNAs from surviving clones

are then subjected to a positive selection in the presence of the cognate heterologous synthetase and a β-lactamase gene with an amber codon at a permissive site. For a cell to survive the selection pressure from ampicillin, tRNAs must be good substrates for the cognate heterologous synthetase and function in translation to suppress the amber codon and produce active β-lactamase, which hydrolyze ampicillin. Therefore, only tRNAs that 1) are not substrates for endogenous *E. coli* synthetases, 2) can be aminoacylated by the heterologous synthetase of interest, and 3) function in translation will survive both selections.

This approach was applied to the $\mathrm{MjtRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ to further reduce recognition of this tRNA by endogenous *E. coli* synthetases, while preserving activity with both the cognate *Mj* TyrRS and translational machinery. Eleven nucleotides of $\mathrm{MjtRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ that do not interact directly with the *Mj* TyrRS were randomly mutated to generate a suppressor tRNA library. This tRNA library was passed through rounds of negative and positive selections to afford a functional, orthogonal tRNA ($\mathrm{mutRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$) that functions efficiently with *Mj* TyrRS to translate the amber codon (Wang and Schultz, 2001).

The orthogonal $\mathrm{mutRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ has high activity for its cognate *Mj* TyrRS, which is critical for the success of changing the substrate specificity of the synthetase to be specific for a Uaa. Because unnatural amino acid-specific synthetase mutants are generally less active than the wild-type synthetase, these mutants will be difficult to identify if the starting wild-type synthetase-tRNA pair is

not highly active. The $\mathrm{mutRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ -*Mj* TyrRS pair has been evolved to incorporate more than 40 Uaas in *E. coli* to date (Wang et al., 2006b). Additional orthogonal tRNA-synthetase pairs have since been generated and include a $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Asp}}$ -AspRS pair derived from yeast (Pastrnak et al., 2000) and an *E. coli* initiator $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{fMet}}$ -yeast TyrRS pair (Kowal et al., 2001). Orthogonal suppressor tRNAs can also be derived from consensus sequences of multiple archaeal tRNAs and then improved with the above selections (Anderson et al., 2004). This approach has been used to evolve an orthogonal *Methanococcus thermoautotrophicum* $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Leu}}$ -LeuRS pair (Anderson and Schultz, 2003), an orthogonal *Methanosarcina mazei* $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Glu}}$ -GluRS pair (Santoro et al., 2003), and an orthogonal *Pyrococcus horikoshii* $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Lys}}$ -LysRS pair (Anderson et al., 2004). These pairs hold the potential for further expanding the range of Uaas that can be incorporated.

Aminoacyl-tRNA synthetases charge the appropriate tRNA with the correct amino acid, which is important in maintaining the fidelity of protein translation. To genetically encode an Uaa, the substrate specificity of the orthogonal synthetase needs to be altered to charge the orthogonal tRNA with only the desired Uaa and none of the common 20 amino acids. A general scheme was developed for evolving the specificity of aminoacyl-tRNA synthetases that is independent of the structure of the amino acid of interest. In this approach, a library of synthetase active-site mutants is generated, and then subjected to a combination of positive

and negative selections to evolve synthetases that aminoacylate the tRNA with the unnatural and no endogenous amino acids (Wang et al., 2001; Wang and Schultz, 2004). Active synthetases charging unnatural or natural amino acids are identified using positive selection, and variants charging natural amino acids are subsequently eliminated from the sample pool using negative selection.

By analyzing the X-ray crystal structure of the synthetase (or a homologue) complexed with its cognate amino acid or aminoacyl adenylate, residues in the substrate binding pocket of the synthetase are identified. Codons for these residues are then randomized using degenerate oligonucleotides to make a DNA library, which is transformed and expressed in *E. coli* cells to generate the mutant synthetase library. *E. coli* cells harboring the synthetase library are selected using alternating rounds of positive and negative selections (Figure 2). The positive selection is based on resistance to chloramphenicol conferred by suppression of an amber mutation at a permissive site in the chloramphenicol acetyl transferase (CAT) gene; the negative selection uses the barnase gene with amber mutations at permissive sites. When the library of synthetase mutants is passed through the positive selection in the presence of the Uaa, those cells with mutant synthetases that can acylate the orthogonal tRNA with either the Uaa or an endogenous amino acid will be able to make full-length CAT, which enables the cell to survive and propagate in the antibiotic chloramphenicol. Plasmids encoding active mutant synthetases are then transformed into the negative selection strain, and selections are carried out in

the absence of the Uaa. Those cells containing mutant synthetases that recognize endogenous amino acids incorporate the latter in response to the amber codons in the barnase gene, and the produced barnase kills the cell. Therefore, only synthetase variants specific for the Uaa can survive both selections. For further optimization, more mutations can be introduced into the synthetase gene of the initial hits using random mutagenesis or DNA shuffling. Repeated rounds of positive and negative selections can be carried out with increased selection pressure (increasing the concentration of chloramphenicol in the positive selection and decreasing the number of amber codons in the barnase gene in the negative selection), until mutant synthetases that can specifically incorporate the Uaa in response to the amber codon are isolated.

An alternative selection scheme makes use of an amber-T7-GFP (green fluorescent protein) instead of an amber-barnase reporter in the negative selection (Santoro et al., 2002). Suppression of amber codons introduced at permissive sites in T7 RNA polymerase produces full-length T7 RNA polymerase, which drives the expression of GFP under the control of the T7 promoter. In this approach both the amber-CAT reporter and amber-T7-GFP reporter are encoded in a single plasmid. After positive selection in chloramphenicol, surviving cells are grown in the absence of both the Uaa and chloramphenicol. Cells containing mutant synthetases that can acylate the tRNA with any of the 20 common amino acids express GFP, whereas cells containing mutant synthetases that can acylate the tRNA only with the Uaa do not. These nonfluorescent cells are

separated from the fluorescent cells using fluorescence activated cell sorting. One advantage of this latter method is that both reporters are contained within a single genetic construct, eliminating the need for plasmid shuttling between positive and negative selections. However, fluorescence intensity is not as clear-cut as survival/death of bacteria, and effective sorting of tiny bacteria cells is technically challenging. Other selection schemes have also been pursued, including cell surface and phage display systems, but these are less general (i.e., require capture reagents specific for the amino acid of interest) or not as efficient (Pastrnak and Schultz, 2001). The CAT-barnase selection strategy has been most efficient in generating a large number of unnatural amino acid-specific synthetases.

The first Uaa genetically encoded into proteins in *E. coli* is *O*-methyl-l-tyrosine (OMeY, **1**) (Scheme 1) (Wang et al., 2001). A library of synthetase mutants (10^9 in size) was generated by randomizing the five active-site residues (Tyr-32, Glu-107, Asp-158, Ile-159, and Leu-162) in the *Mj* TyrRS based on the crystal structure of the homologous *Bacillus stearothermophilus* TyrRS-tyrosyl adenylate complex. This library was then subjected to two rounds of positive selection and negative screening to afford clones that survived at high chloramphenicol concentrations in the presence of OMeY and at low chloramphenicol concentrations in its absence. To verify the incorporation of OmeY in response to the amber stop codon by the evolved $\mathrm{mutRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$-mutant TyrRS pair, the third codon of the *E. coli* DHFR gene was mutated to TAG, and a

His6 tag was added to the C-terminus to facilitate protein purification and separation from endogenous DHFR. Full-length DHFR was produced only when the mutant TyrRS, the $\mathrm{mutRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$, and OMeY were each present. In the absence of any one component, no DHFR protein could be detected by silver stain or Western blot analysis. Insertion of OMeY in response to the TAG codon was confirmed by mass spectrometric analysis of both the intact protein and tryptic fragments. No incorporation of tyrosine or other amino acids at the TAG position was observed, and OMeY was incorporated only in response to TAG and not any other sites in DHFR. These results demonstrated that one could rationally engineer bacteria that genetically encode an Uaa with high efficiency and translational fidelity.

OMeY is structurally similar to tyrosine and phenylalanine and provided an excellent case to demonstrate the high translational fidelity that can be achieved by this approach. Meanwhile, it was important to test whether the method could be used to genetically encode Uaas that significantly deviate from common amino acids in structure. Efforts were then attempted to genetically encode the second Uaa, l-3-(naphthyl)-alanine (**2**), which represents a significant structural perturbation relative to tyrosine (Wang et al., 2002). A slightly different mutant *Mj* TyrRS library (in which Tyr 32, Asp158, Ile 159, Leu162, and Ala167 were randomized) was constructed, from which four mutant synthetases were identified that have high activity for this Uaa. One round of DNA shuffling of these four synthetase genes followed by selection resulted in a mutant synthetase

(SS12) with enhanced activity for l-3-(naphthyl)alanine and greatly reduced activity for endogenous amino acids. Protein expression and mass spectrometric analysis confirmed that the $\mathrm{mutRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ /SS12 mutant synthetase selectively incorporates l-3-(naphthyl)alanine into proteins with an efficiency and fidelity rivaling that of the common 20 amino acids. This result suggested that the above methodology may be applicable to a large variety of Uaas.

Using the above methodology, more than 40 Uaas have been genetically incorporated into proteins in *E. coli* (Wang et al., 2006b). In general, for most Uaas, suppression efficiencies range from 25% to 75% of wild-type protein and translational fidelity is >99%. Yields of Uaa containing proteins are in the range of several milligrams to tens of milligrams per liter of cell culture. In optimized expression systems, the yield could reach about 1 g/liter, sufficient for large-scale protein production. Other than structural analogs of natural amino acids, which can probe very fine aspects of amino acid structure as it relates to protein function, many Uaas have interesting chemical properties that make them particularly useful for biological research.

The general strategy developed to genetically encode Uaas in bacteria is applicable to higher organisms. Orthogonal tRNA-codon-synthetase sets need to be generated for each organism of interest by following the same principle. A number of orthogonal tRNA-synthetase pairs have been generated for use in eukaryotic organisms, and most of them are derived from bacterial pairs, again due to the low cross-kingdom aminoacylation. For instance, it has been shown

that the *E. coli* tyrosyl amber suppressor ($tRNA^{Tyr}_{CUA}$) is not aminoacylated by eukaryotic synthetases and acts as an amber suppressor in *S. cerevisiae* in the presence of *E. coli* TyrRS (Edwards and Schimmel, 1990; Edwards et al., 1991). In addition, *E. coli* TyrRS does not aminoacylate yeast tRNAs (Kwok and Wong, 1980). Thus, *E. coli* $tRNA^{Tyr}_{CUA}$-TyrRS functions as an orthogonal pair in yeast. A human initiator tRNA-derived amber suppressor together with *E. coli* GlnRS form another orthogonal pair for use in yeast (Kowal et al., 2001). In addition, the *E. coli* $tRNA^{Gln}_{CUA}$-GlnRS pair has also been used in mammalian cells for efficient suppression of the amber codon (Drabkin et al., 1996). To selectively introduce an Uaa into proteins in eukaryotes, Yokoyama and co-workers screened a small collection of designed active-site variants of *E. coli* TyrRS in a wheat-germ translation system and discovered a mutant synthetase that utilizes 3-iodotyrosine (**3**) more effectively than tyrosine (Kiga et al., 2002). This mutant synthetase was used with the *B. stearothermophilus* $tRNA^{Tyr}_{CUA}$ to incorporate 3-iodotyrosine into proteins in mammalian cells (Sakamoto et al., 2002).

To develop a general selection scheme in yeast analogous to that used in *E. coli* for evolving synthetases specific for Uaas, two amber stop codons were introduced into the permissive sites of the transcriptional activator protein GAL4 (Chin et al., 2003). The GAL4(2TAG) mutant was expressed in yeast MaV203, a commercially available yeast strain which has the endogenous GAL4 deleted and contains three GAL4 inducible reporter genes, HIS3, URA3, and LACZ.

Suppression of these amber codons leads to the production of full-length GAL4, which drives transcription of the reporters. Expression of HIS3 and URA3 complements the histidine and uracil auxotrophy in this strain and provides a positive selection for clones expressing active synthetase mutants. On the other hand, addition of 5-fluoroorotic acid (5-FOA), which is converted into a toxic product by URA3, results in the death of cells expressing active synthetases. In the absence of the Uaa, this serves as a negative selection to remove synthetases specific for endogenous amino acids. Like GFP, the lacZ reporter can serve as another marker for colormetrically distinguishing active synthetases from inactive ones.

This selection scheme was used to evolve the orthogonal *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$-TyrRS pair in yeast (Chin et al., 2003). A synthetase library (10^8 in size) was similarly constructed by randomizing five active-site residues in the *E. coli* TyrRS corresponding to the five residues randomized in the *Mj* TyrRS. Mutant synthetases were identified after several rounds of positive and negative selection that incorporate a number of Uaas into proteins, albeit with rather low protein yields (about 0.05 mg/L). A similar approach has been used to evolve orthogonal *E. coli* leucyl $\mathrm{tRNA}_{\mathrm{CUA}}$-LeuRS pairs that selectively incorporate photochromic and fluorescent amino acids into proteins in yeast (Summerer et al., 2006; Wu et al., 2004).

Low incorporation efficiency creates difficulties in making use of the genetically encoded Uaas in yeast. To address this problem, new methods have

been developed to efficiently express orthogonal prokaryotic tRNA and to improve the target mRNA stability in yeast, respectively. Prokaryotes and eukaryotes differ significantly in tRNA transcription and processing (Figure 3a). *E. coli* tRNAs are transcribed by the sole RNA polymerase (Pol) through promoters upstream of the tRNA gene. However, the transcription of eukaryotic tRNAs by Pol III depends principally on promoter elements within the tRNA known as the A- and B-box (Galli et al., 1981). The A- and B-box identity elements are conserved among eukaryotic tRNAs but are lacking in many *E. coli* tRNAs. Creating the consensus A- and B-box sequences in *E. coli* tRNAs through mutation could cripple the tRNA (Sakamoto et al., 2002). In addition, all *E. coli* tRNA genes encode full tRNA sequences, whereas eukaryotic tRNAs have the 3′-CCA trinucleotide enzymatically added after transcription (Li et al., 2002). The *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ has the B-box but no fully matched A-box. In the previously described system (Chin et al., 2003), only the structural gene of *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ was inserted in the plasmid. The resultant basal level of tRNA expression is likely driven by a cryptic promoter elsewhere on the plasmid, and should contribute to the low incorporation efficiency of Uaas.

One method to increase the tRNA expression uses the 5'- and 3'-flanking sequences of an endogenous yeast suppressor $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ (SUP4) (Chen et al., 2007). The *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ without the -CCA trinucleotide was flanked by these sequences, and this arrangement was repeated 3 to 6 times in tandem (Figure

3b). A strong PGK1 RNA Polymerase II promoter was additionally placed upstream of the tandem arrangements. This combination gave an overall >50 fold increase in tRNA levels compared to the previous tRNA alone scheme. Together with the optimized expression of the synthetase, this approach improved the yield of Uaa containing proteins to 6-8 mg/L of culture.

A new method to efficiently express prokaryotic tRNAs in yeast involves an external promoter containing the consensus A- and B-box sequences (Figure 3c) (Wang and Wang, 2008). When placed upstream of the *E. coli* tRNA (without the 3'-CCA trinucleotide), the promoter drives transcription of a primary RNA transcript consisting of the promoter and the tRNA. The promoter is then cleaved post-transcriptionally to yield the mature tRNA. Two yeast Pol III promoters, the SNR52 promoter and the RPR1 promoter, have been shown to drive the expression of the *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ efficiently in yeast (Wang and Wang, 2008). The expressed *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ are 6 to 9 fold more active in protein translation than the tRNA transcribed using the SUP4 5'-flanking sequence. The increased activity is not through an increase in tRNA transcription level based on Northern blot results, underscoring the importance of proper tRNA processing. The cleavage of the SNR52 or RPR1 promoter from the primary RNA transcript could directly generate the correct 5' end of the tRNA. When a 5' flanking sequence is used, RNaseP is required to generate the 5' tRNA end. It may be difficult for yeast RNaseP to process prokaryotic tRNA efficiently. This external promoter

method has also been used to express *E. coli* $tRNA^{Leu}_{CUA}$ in yeast, and the resultant $tRNA^{Leu}_{CUA}$ functions efficiently in protein translation as well, suggesting it can be a general method for functionally expressing different prokaryotic tRNAs in yeast and other eukaryotes.

The stability of mRNA is an important issue for Uaa incorporation in eukaryotes. Eukaryotic cells have an mRNA surveillance mechanism, nonsense mediated mRNA decay (NMD), to identify mRNA with premature stop codons and target the mRNA for rapid degradation. This mechanism is intended to protect cells from generating undesired truncated proteins. When stop codons are used to encode Uaas, NMD could result in a shorter lifetime for the target mRNA and thus a lower protein yield. Inactivation of NMD would preserve the stability of the UAG-containing target mRNA and thus enhance the incorporation efficiency of Uaas. An NMD-deficient yeast strain was generated by knocking out the UPF1 gene, an essential component for NMD in yeast. The Uaa incorporation efficiency was indeed increased more than 2 fold in the upf1Δ strain compared to the wild-type yeast (Wang and Wang, 2008). NMD in yeast shows a polar effect of nonsense codon position (Cao and Parker, 2003; Peltz et al., 1993). The reduction in steady-state mRNA levels is larger when the nonsense codon is closer to the 5' end than to the 3' end. The increase of Uaa incorporation efficiency in the upf1Δ strain similarly correlates with the position of the UAG codon, with > 2 fold increase when the amber codon is within the N-terminal two

thirds of the gene and no significant increase when within the C-terminal fourth of the coding region (Wang, Q., Wang, L., unpublished results). By using the external SNR52 promoter and the NMD-deficient upf1Δ strain, the overall purified yields of Uaa containing proteins have reached ~ 15 mg/L of yeast cells, ~300-fold higher than the previous system and comparable to the yield in *E. coli*.

Genetic incorporation of Uaas into proteins in mammalian cells is a big leap, as mammalian cells are biologically complex, and therefore all of the hurdles to incorporation in yeast will be amplified. Specifically, one challenge is the efficient expression of prokaryotic orthogonal tRNAs that are functional in translation in mammalian cells, again due to the aforementioned difference in tRNA transcription and processing between prokaryotes and eukaryotes. In addition, the transfection efficiency of mammalian cells is much lower than the transformation efficiency of *E. coli* and yeast, making it impractical to generate large synthetase mutant libraries inside mammalian cells. Survival-death selection in mammalian cells is also not as efficient as in *E. coli* and yeast. These factors lead to a second challenge: how to evolve mutant orthogonal synthetases to be specific for Uaas for use in mammalian cells.

Initial attempts at introducing Uaas into proteins in mammalian cells involved the transfection of an amber suppressor tRNA that is chemically acylated with an Uaa *in vitro* (Kohrer et al., 2001; Monahan et al., 2003). This approach limits the amount of overall protein that can be produced since the acylated suppressor tRNA is consumed stoichiometrically and cannot be

regenerated inside cells. It is also technically demanding to prepare the chemically acylated tRNA.

Since the *E. coli* $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$-TyrRS pair is orthogonal in mammalian cells, it should be possible to evolve this pair to incorporate Uaas in mammalian cells. However, the *E. coli* $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ does not express well in mammalian cells, presumably due to the lack of a matched A-box inside the tRNA structural gene. Another bacterial $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ derived from *B. stearothermophilus* happens to contain the consensus A- and B-box sequences in eukaryotic tRNAs. This tRNA is not aminoacylated by any endogenous synthetase in mammalian cells, and functions with the *E. coli* TyrRS for suppressing the TAG codon with tyrosine. The *B. stearothermophilus* $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ (lacking the 3'-CCA) was linked to the 5'-flanking sequence of the human $\mathrm{tRNA}^{\mathrm{Tyr}}$, and nine tandem repeats of this gene arrangement was used for tRNA expression (Sakamoto et al., 2002). A small collection of designed active-site variants of the *E. coli* TyrRS was screened using *in vitro* biochemical assays, and a mutant synthetase that uses 3-iodotyrosine more effectively than Tyr was identified. This mutant synthetase was used with nine tandem *B. stearothermophilus* $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ to incorporate 3-iodotyrosine (**3**) into proteins in Chinese hamster ovary (CHO) cells and HEK293 cells with approximately 95% fidelity. In another attempt, mutations were introduced into the A-box region of the *B. subtilis* tryptophan opal suppressor tRNA ($\mathrm{tRNA}_{\mathrm{UCA}}^{\mathrm{Trp}}$) to create a consensus A-box (Zhang et al., 2004). The mutant

tRNA (lacking the 3'-CCA) was expressed using the 5'- and 3'-flanking sequences from the *Arabidopsis* $tRNA^{Trp}$. Together with a rationally designed *B. subtilis* TrpRS mutant, this tRNA incorporated 5-hydroxytryptophan (**4**) into the foldon protein in HEK293 cells in response to the UGA opal codon with ~97% fidelity.

The above methods may not be generally applied to other tRNA-synthetase pairs and various Uaas. Most bacterial tRNAs do not have the consensus A- and B-box sequences, and heterologous pairing of the tRNA and synthetase often results in low activity. Mutations to create such consensus sequences in tRNA, although they did not dramatically decrease the activity of the *B. subtilis* $tRNA^{Trp}_{UCA}$ (Zhang et al., 2004), greatly impaired the suppression ability of *E. coli* $tRNA^{Tyr}_{CUA}$ (Sakamoto et al., 2002) since the A- and B-box lie at nucleotides involved in tertiary interactions that support the L-shaped structure of the tRNA. For generating synthetase variants that are specific for Uaas, it is difficult to predict *a priori* which active site residues need to be mutated. As mutations are often required at multiple sites to achieve high substrate specificity, small collections of synthetase mutants will likely fall short. Mutants thus generated may still recognize common amino acids, as is the case with the synthetase used to incorporate 3-iodotyrosine (Sakamoto et al., 2002). Therefore, it remained a challenge to incorporate different Uaas in mammalian cells with high efficiency and fidelity.

For efficient expression of functional prokaryotic tRNAs in mammalian cells, a novel method was developed that involves the use of an external type-3 Pol III promoter (Wang et al., 2007). A type-3 Pol III promoter does not require intragenic elements (such as the A- and B-box) for transcription; its promoter sequence resides exclusively upstream of the coding sequence. Thus tRNAs with or without the consensus A- and B-box sequences can be transcribed by this type of promoter in mammalian cells. In addition, the transcription initiation site of some type-3 Pol III promoters, such as the H1 promoter (Myslinski et al., 2001) and the U6 small-nuclear RNA promoter (Paule and White, 2000), is well defined, which could be used to generate the correct 5' end of the tRNA without further post-transcriptional processing.

A fluorescence-based translation assay was set up to systematically identify expression elements that are required to efficiently drive the transcription of *E. coli* tRNAs and to generate tRNAs functional in protein translation in mammalian cells. The H1 promoter, which drives the expression of human H1 RNA (the RNA component of the human nuclear RNase P), the 3'-CCA trinucleotide, and the 5'- and 3'-flanking sequence from the human $tRNA^{fMet}$ are linked to the *E. coli* $tRNA^{Tyr}_{CUA}$ in different combinations for tRNA expression. The gene arrangement that yields tRNAs with the highest translational activity is the H1 promoter, followed by the *E. coli* $tRNA^{Tyr}_{CUA}$ (without 3'-CCA) and by the 3'-flanking sequence of $tRNA^{fMet}$. The *E. coli* $tRNA^{Tyr}_{CUA}$ expressed using this method

is ~70 fold more active in translation than the one expressed with the 5' and 3'-flanking sequences of human $tRNA^{fMet}$. This method has also been used to express different *E. coli* tRNAs successfully, and it is effective in various mammalian cells such as HEK293, HeLa, and primary neurons.

A transfer strategy was developed to address the challenge of evolving Uaa specific synthetases directly in mammalian cells. Since the *E. coli* $tRNA^{Tyr}_{CUA}$ - TyrRS pair is orthogonal in mammalian cells (in addition to yeast), and the translational machinery of yeast is homologous to that of higher eukaryotes, it should be possible to evolve the specificity of this synthetase in yeast (which is well-suited for genetic selections with large libraries) and transfer the optimized tRNA-synthetase pairs directly to mammalian cells. In one report, mutant synthetases that were evolved in yeast from the *E. coli* TyrRS specific for different Uaas, together with the *B. stearothermophilus* $tRNA^{Tyr}_{CUA}$, have been successfully used in CHO cells and HEK293 cells for incorporating the corresponding Uaa (Liu et al., 2007). Mass spectrometry confirmed a high translational fidelity for the Uaa, and proteins could be produced with efficiencies up to 1 μg per 2 x 10^7 cells.

In another report, mutant synthetases evolved in yeast from the *E. coli* TyrRS were used together with the cognate *E. coli* $tRNA^{Tyr}_{CUA}$ expressed by the type-3 Pol III H1 promoter in mammalian cells (Wang et al., 2007). Because the H1 promoter enabled the functional expression of the *E. coli* $tRNA^{Leu}_{CUA}$, mutant

synthetases evolved in yeast from the *E. coli* LeuRS were also successfully transferred in mammalian cells to incorporate the fluorescent Uaa dansylalanine (**5**). The incorporation efficiencies of the Uaas were in the range of 13% to 41%, depending on the activity of the evolved synthetase, and the incorporation fidelity was also high judging by a sensitive fluorescence assay. Uaas have been incorporated in HEK293 cells, HeLa cells, and even in primary neurons. Finally, Uaas were then incorporated into a voltage sensitive K^+ channel, Kv1.4, in HEK293 cells, which allowed for a mechanistic analysis of the fast inactivation of the ion channel.

The total number of Uaas that can be encoded in any organism is limited by the number of noncoding codons. It should be possible to use quadruplet codons and cognate suppressor tRNAs with expanded anticodon loops to encode additional amino acids. There are many examples of naturally occurring +1 frameshift suppressors including UAGN (N = A, G, C, or T) suppressors derived from Su7 which encodes glutamine (Curran and Yarus, 1987), *sufJ*-derived suppressors of ACCN codons encoding threonine (Bossi and Roth, 1981), and CAAA suppressors derived from $tRNA^{Lys}$ and $tRNA^{Gln}$ (O'Connor, 2002). Moreover, genetic selections have been used to identify efficient four and five-base codon suppressor tRNAs from libraries of mutant tRNAs (Anderson et al., 2002; Magliery et al., 2001). Frameshift suppressor tRNAs can efficiently incorporate a number of the common 20 amino acids into proteins *in vivo* (Atkins et al., 1991), and chemically aminoacylated frameshift suppressors have also

been used to incorporate Uaas into proteins using *in vitro* translation systems with four- and five-base codons (Hohsaka et al., 2001a; Hohsaka et al., 2001b). An orthogonal four-base suppressor tRNA-synthetase pair was generated from the $tRNA^{Lys}$-LysRS of archaebacteria *Pyrococcus horikoshii*, which efficiently incorporates the Uaa homoglutamine (**6**) into proteins in *E. coli* in response to the quadruplet codon AGGA (Anderson et al., 2004). Frameshift suppression with homoglutamine does not significantly affect protein yields or cell growth rates, and is mutually orthogonal with amber suppression. This approach allows the simultaneous incorporation of two Uaas at distinct sites within proteins. The orthogonal AGGA-specific $\mathrm{tRNA}^{\mathrm{Lys}}_{\mathrm{UCCU}}$ and the UAG-specific $\mathrm{mutRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ were expressed with a mutant myoglobin gene containing the Gly24(AGGA) and Ala75(TAG) mutations. In the presence of both of the corresponding orthogonal homoglutamine-specific and *o*-methyl-L-tyrosine-specific synthetases and the two Uaas, 1.7 mg/L of mutant myoglobin was produced with an overall suppression efficiency ~25%. Electrospray mass spectrometric analysis of the full-length protein confirmed that myoglobin contained both Uaas.

As our ability to synthesize large DNA molecules keeps improving, it may be possible to generate a synthetic *E. coli* variant in which rare or degenerate codon-tRNA pairs are eliminated from the wild-type genome by replacing them with sense codons specifying the same amino acid. The liberated codons can instead be used to encode Uaas. All these possibilities suggest that neither the

number of available triplet codons nor the translational machinery itself represents a significant barrier to further expansion of the code.

Uaas are added to the growth medium in most experiments. There are a large number of amino acid and amine transporters that are relatively nonspecific, which may help to transport the Uaas into cells. From measurements of cytoplasmic levels of amino acids, it is found that a large number of Uaas are efficiently transported to the *E. coli* cytoplasm in millimolar concentrations. Highly charged or hydrophilic amino acids may require derivatization (e.g., esterification, acylation) with groups that are hydrolyzed in the cytoplasm. Metabolically labile amino acids or analogues (e.g., α-hydroxy acids, *N*-methyl amino acids) may require strains in which specific metabolic enzymes are deleted.

An alternative to adding exogenous amino acids to the growth media involves engineering a pathway for the biosynthesis of the Uaa directly in the host organism. For example, genes for the biosynthesis of *p*-amino-L-phenylalanine (*p*AF) from simple carbon sources can be heterologously expressed in *E. coli* (Mehl et al., 2003). *p*AF was biosynthesized from the metabolic intermediate, chorismic acid, using the *papA, papB,* and *papC* genes from *Streptomyces venezuelae* in combination with a nonspecific *E. coli* transaminase. *E. coli* containing these genes produced *p*AF at levels comparable to those of other aromatic amino acids and had normal growth rates. In the presence of a *p*AF specific, orthogonal $\mathrm{mutRNA}_{\mathrm{CUA}}^{\mathrm{Tyr}}$ -synthetase pair, *E. coli* transformed with *papA-C* produced mutant proteins containing *p*AF at sites

encoded by the amber codon with excellent yield and fidelity. In addition to *p*AF, it should be possible to biosynthesize and genetically encode other amino acids *in vivo* as well.

Uaas have been used to study various biological problems involving proteins. As virtually every cellular process involves proteins, this technology has very broad applicability. A subset of applications of Uaas that are genetically encoded using the above methodology is highlighted below. As the selected examples illustrate, the novel physical, chemical, and biological properties embodied by Uaas could benefit research in a diverse set of disciplines.

Site-specific incorporation of biophysical probes into proteins allows the structure and function of proteins to be probed with greater precision and accuracy using various biophysical means. Genetically encoded biophysical probes in the form of Uaas further extend the potential of these studies into live cells, the native environment of proteins. A handful of Uaas that can serve as biophysical probes have been genetically incorporated into proteins in *E. coli*, yeast, and mammalian cells, and representative applications of some probes are discussed.

Standard techniques to introduce heavy atoms into protein crystals for phase determination use telluromethionine or selenomethionine to replace methionine during protein translation, or soak the protein crystal in solutions containing ions for heavy atoms. The soaking method can produce a number of low-occupancy sites whose positions must be determined before phase can be

derived, while quantitative replacement of methionine is difficult. A novel method to introduce heavy atoms into proteins is to genetically incorporate *p*-iodo-L-phenylalanine (**3**) into proteins in *E. coli* or yeast cells. This approach ensures that the heavy atom iodine is quantitatively introduced at a specific site of the target protein, and a large quantity of proteins can be conveniently prepared for single-wavelength anomalous dispersion (SAD) phasing. *p*-iodo-L-phenylalanine was incorporated into T4 lysozyme, and the mutant protein was crystallized (Xie et al., 2004). Diffraction data were collected with a laboratory CuKα X-ray source, and the structure was solved using SAD. A single iodinated amino acid among 164 residues resulted in a strong anomalous signal, about 3% of the total intensities, which compares favorably with the level achieved with selenomethionine using synchrotron beams. In addition, this amino acid caused little structural perturbation when substituted for Phe in the core of T4 lysozyme. The strong anomalous signal, the possibility of incorporation at multiple sites and in different cell types, and the use of an in-house X-ray source should facilitate solving novel protein structures in a high-throughput manner, providing more structural data to the scientific community.

Nuclear magnetic resonance (NMR) can be used to determine the structure and dynamics of a protein or complex of proteins in solution. Transient binding or conformational changes associated with binding to small molecules or other proteins could be measured, in principle, if NMR spectra from large proteins or complexes were less complicated to decode. The ability to introduce one or

several specific NMR labels at defined locations in a protein would greatly reduce the complexity of its NMR spectra and simplify the signal assignment. Genetic incorporation of Uaas containing isotopic labels or fluorine can allow such NMR labels to be introduced site-selectively into proteins with high fidelity and efficiency.

Tri-fluoromethyl-L-phenylalanine (tfm-Phe, **8**) was placed into the binding interface of two well studied obligate dimers, nitroreductase and histidinol dehydrogenase, both of which contain active sites at the interface of these dimers (Jackson et al., 2007). Possible decreases in enzyme activity from Uaa incorporation, as well as substrate or inhibitor induced conformational changes, could be measured by NMR, and there were clear conformation-based changes in signal output. In another report, the binding of a small molecule ligand to the thioesterase domain of fatty acid synthase was studied by NMR using three different NMR-active Uaas, tfm-Phe, ^{13}C and ^{15}N labeled *o*-methyl-phenylalanine (**9**), and ^{15}N labeled *o*-nitrobenzyl tyrosine (**10**), at 11 different sites (most solvent exposed), none of which seriously perturbed the structure of the protein (Cellitti et al., 2008). By comparing the spectra of a number of the mutants and the changes in conformation upon addition of a small molecule, the binding site of this molecule was mapped on the protein. In addition, photo-decaging of the ^{15}N labeled *o*-nitrobenzyl tyrosine to regenerate ^{15}N labeled tyrosine provides a new method of isotopic labeling of individual residues without altering the protein sequence. The utility of these methods for the identification of active small

molecules and the exclusion of those binding to non-desired regions of the protein, as well as their effects on conformation, are likely to be used in the future for NMR structural studies.

The side chain size of Uaas can be conveniently altered at atomic precision to probe effects of amino acid bulk to protein structure and function. For instance, the fast inactivation mechanism of the voltage-gated potassium channel Kv1.4 was examined using Uaas in HEK293 cells (Wang et al., 2007). The classic ball-and-chain model for channel inactivation suggests that the N-terminal inactivation peptide forms a ball-like domain to occlude the channel exit for ions. In contrast, a new model hypothesizes that the inactivation peptide threads through a side portal and extends into the inner pore of the channel to block ion flow. To experimentally test the new model, tyrosine 19, a residue in the inactivation peptide, was initially mutated to phenylalanine or tryptophan, which showed no difference in channel inactivation. However, mutation of this residue to *o*-methyl-L-tyrosine (**1**) resulted in a markedly slower inactivation, as did the incorporation of dansylalanine (**5**) at this site. Modeling suggested that the diameter of the inactivation peptide, which is unchanged for each of the canonical amino acids incorporated at this site but is larger for *o*-methyl-L-tyrosine and dansylalanine, is important for channel inactivation. This is likely due to the narrow width of the side portal in the channel, supporting the new model for channel inactivation. Elucidation of structure-function relationships involving specific residues in other proteins can also benefit from Uaa incorporation, as in

this case, because conventional mutagenesis is unable to provide the desired increase or decrease residue bulk to clearly delineate the effect.

The ability to tag a protein with fluorescence has been a boon to molecular and cell biology, allowing *in vivo* imaging studies to detect protein expression, localization or trafficking. Fluorescent amino acids would provide a similar level of genetic encodability, but without the bulkiness and potential perturbations of normal protein function and interaction that fluorescent proteins may introduce. Additionally, these Uaas can be engineered to be sensitive to microenvironments, lending themselves to potential use as sensors of post-translational modification, dimerization, protein folding, pH, and a number of other physical and biochemical properties. Two fluorescent amino acids, 2-amino-3-(5-(dimethyl-amino)naphthalene-1-sulfonamide) propanoic acid (dansylalanine, **5**) and L-(7-hydroxycoumarin-4-yl) ethylglycine (coumarin, **11**), have been genetically encoded into proteins. Each of these Uaas possesses a fluorescent moiety that is excited in the UV spectrum and emits blue or green light. The emission intensity or wavelength of these Uaas is dependent on the microenvironment where the Uaa is located. Dansylalanine (DanAla) is brighter in hydrophobic environments than hydrophilic ones, making it a feasible probe of dimerization or protein folding, while coumarin is sensitive to the pH in addition to the polarity of the surroundings.

DanAla has been genetically encoded as a protein folding reporter in human superoxide dismutase (SOD), which was expressed in *S. cerevisiae*,

purified, and treated with guanidinium chloride (Summerer et al., 2006). Placing the Uaa in a site on the surface of a β-barrel in the protein showed little change in fluorescence after protein denaturation. When the Uaa was placed in an internal site of the β-barrel there was a red-shift in the emission wavelength of the Uaa and a large decrease in fluorescence intensity upon denaturation. The change from a buried to exposed environment was reported by changes in wavelength and intensity of the fluorescence from the incorporated DanAla. The potential to use DanAla or similar Uaas as sensors for protein-protein dimerization by incorporation at a binding interface, as a fluorescent reporter of receptor-ligand binding or to identify membrane associated proteins, and a number of other possible applications, make this Uaa attractive for further study.

The coumarin based fluorescent Uaa **11** has properties similar to DanAla, and has also been used to monitor protein folding due to its solvent polarity-dependent fluorescence emission (Wang et al., 2006a). The coumarin Uaa was incorporated into sperm whale myoglobin at two different helices, A or C. An increase in fluorescence correlates to local unfolding of the region containing the coumarin due to increased exposure of the Uaa to solvent. At 5M urea, both sites show an equal fluorescence increase, indicating unfolding of the protein. At 2M urea fluorescence increases only when the Uaa is incorporated in helix A, while coumarin which is incorporated at helix C shows little change in fluorescence intensity. However, at 3M urea, coumarin placed in helix C shows a similar increase in fluorescence. NMR data indicates that at urea concentrations above

2.2M, helices A and B are largely unfolded, while the remaining helices (including C) are not destabilized until urea concentrations above 3M, suggesting that the fluorescent emissions of coumarin incorporated in different locations are an accurate reporter of local protein destabilization. Coumarin can be used as a site-specific probe of local protein folding and conformational changes, and may be useful for other fluorescence based protein assays sensitive to polarity.

Coumarin has also been used as a substrate phosphorylation sensor, based on its pH and polarity-dependent fluorescence attributes. Signal transducer and activator of transcription-3 (STAT3) is a transcription factor that dimerizes upon phosphorylation of Tyr 705 and translocates to the nucleus to bind DNA and activate transcription. When incorporated into the SH2 domain of STAT3 and purified from *E. coli*, *in vitro* phosphorylation of Tyr 705 by Src can be detected by monitoring coumarin fluorescence change by fluorometer (Lacey et al., 2011). Additionally, this purified protein can be used as a fluorescent sensor to detect phosphorylated STAT3 in mammalian cell extracts after stimulation with IL-6, and is one of the first reports demonstrating a substrate level sensor, as most kinase activity sensors detect a specific kinase, not phosphorylation of a specific substrate.

Though still in its infancy, the use of genetically encoded fluorescent amino acids as biosensors is promising, and it is likely these and similar Uaas will be used as optical reporters directly in live cells in the future for studying protein trafficking and function in live cells and in real time.

It would be advantageous to control protein interaction or activity with an easy to administer, non-invasive and selective method. Application of light of a specific wavelength would be an ideal way to control protein activities directly *in vivo* without disrupting the cellular environment. Many functional groups can respond to light by crosslinking with nearby molecules, removing protecting groups, or changing molecular conformation. A number of Uaas bearing such functional groups as side chains have been genetically incorporated into proteins. They have the potential to be very useful in a wide variety of applications, including the identification of protein-protein or protein-nucleic acid interactions and the optical control of protein structure or function *in vivo*.

Crosslinking has been an important technique for uncovering protein interactions in cells. Using specific antibodies and formaldehyde, interactions between proteins or proteins and DNA can be examined. However, formaldehyde crosslinking lacks both spatial and temporal resolution. In general, it is difficult to pinpoint transient or weak interactions between proteins, to identify which domains of the protein are involved in the interaction, and to determine whether the interaction is direct or involves an unknown protein acting as a scaffold in protein complexes. Site-specific crosslinking enabled by genetically encoded Uaas would allow subregions of a protein to be examined for binding of proteins, and would distinguish direct interactions from indirect ones. In addition, transient or weak interactions can be enriched by a prolonged crosslinking time, as only

the proteins containing the Uaa will be covalently linked, rather than capturing a snapshot of the whole cell at once.

Three Uaas, *p*-azidophenylalanine (Azi, **7**), *p*-benzoylphenylalanine (pBpA, **12**), and *p*-(3-trifluoromethyl-3*H*-diazirin-3-yl)-phenylalanine (TfmdPhe, **13**) with photocrosslinking side chains have been genetically incorporated into proteins. Upon excitation by UV light, Azi crosslinks to C-H and N-H bonds, pBpA to C-H bonds, and TfmdPhe to C-H or O-H bonds. A mutant homodimeric glutathione S-transferase with *p*-azidophenylalanine or pBpA substituted site-specifically at the dimer interface could be crosslinked *in vitro* (Chin et al., 2002a; Chin et al., 2002b), as well as in the cytoplasm of *E. coli* in case of pBpa (Chin and Schultz, 2002).

Photocrosslinkers have been incorporated in *E. coli* to confirm close contact between specific residues of a protein and its substrate. ClpB, a heat shock protein that aids in the disaggregation and refolding of proteins during the heat shock response, has a conserved aromatic residue (Tyr 251) in the central pore of its AAA+ domain, considered to be the main substrate recognition residue (Schlieker et al., 2004). Using *in vivo* Uaa incorporation of pBpA to replace Tyr 251, biotinylated substrate peptides were shown to be crosslinked upon UV light exposure, but not if pBpA was incorporated elsewhere in the AAA+ domain of this protein.

Photocrosslinkers have also been used in *S. cerevisiae* to specifically crosslink a ligand to its receptor. Ste2p, a G-protein coupled receptor, is known to

bind a short peptide pheromone, α-factor, which affects downstream signaling and leads to growth arrest. Several extracellular loops of Ste2p, based on available topology, were mutagenized to the amber stop codon for pBpA incorporation and α-factor photo-capture (Huang et al., 2008). Two sites were able to specifically crosslink biotinylated α-factor and were identified by Western blot. It is possible in the future that GPCRs or other cell membrane receptors can be mutagenized to incorporate a photo-crosslinker and identify novel ligands through mass spectrometry.

In mammalian cells, pBpA has been incorporated into Grb2 (growth-factor-receptor-bound protein 2), in the ligand binding pocket of the SH2 domain, which binds to specific, phosphorylated tyrosine residues. It was shown that pBpA incorporation into a site on the SH2 domain did not interfere with Grb2 binding to the epidermal growth factor receptor (EGFR) by co-immunoprecipitation, and that a larger molecular weight band can be seen when activated EGFR and Grb2 containing pBpA were co-expressed and treated with UV light (Hino et al., 2005). Several positions on the SH2 domain were chosen to place pBpA, with differing crosslinking capacities based on their distance from the ligand. Therefore, closely interacting versus further interacting residues in ligand binding may be distinguished using photocrosslinkers placed in various locations on the binding domain.

A separate report in mammalian cells uses Azi to map the ligand binding region of the corticotropin releasing factor receptor-type 1 (CRF-R1), a G-coupled

protein receptor (GPCR) which exhibits promiscuity in ligand binding. By substituting the photocrosslinker for a number of positions in extracellular regions of the receptor, radioactive ligands were covalently captured and visualized on a gel (Coin et al., 2011). Intriguingly, certain sites appear to bind all ligands examined, while others are specific for only certain ligands, challenging the current model of how this receptor binds ligands and transduces signals. This method of receptor mapping to identify ligand binding sites cannot be performed with conventional formaldehyde crosslinking and has great potential to elucidate how promiscuous receptors transduce multiple responses.

These studies suggest that fine structural and mechanistic information can be obtained from site-specific photo-crosslinking using genetically incorporated Uaas. The potential of photocrosslinkers to identify biomolecular interactions more specifically than general formaldehyde crosslinking, combined with its compatibility with a wide range of biological processes, suggests that this *in vivo* photocrosslinking strategy will be used more extensively in the future to very specifically probe protein complexes and the interactions within.

To regulate protein activity using light, one strategy is to attach a photoremovable protecting group to the suitable amino acid in the target protein, which masks the amino acid and renders the protein inactive. Photolysis releases the caging group and converts the amino acid to an active form, which generates abrupt or localized changes to the target protein. The nitrobenzyl derivative is the most prevalent form for caged compounds. The side chain hydroxy or thiol

groups of Cys, Ser and Tyr are blocked by substituted nitrobenzyl groups that can be cleaved on irradiation with 365-nm light. These photocaged Cys, Ser and Tyr have been genetically incorporated into different proteins.

In one example, mutation of the active-site cysteine residue in the proapoptotic protease caspase 3 to *o*-nitrobenzylcysteine (**14**) in yeast led to a catalytically inactive enzyme. UV illumination of the cell lysate converted ~40% of the caged caspase to the active enzyme (Wu et al., 2004). In another report, *o*-nitrobenzyltyrosine (**15**) was incorporated into β-galactosidase to activate its enzymatic activity by using light both *in vitro* and in *E. coli* (Deiters et al., 2006). β-galactosidase activity is dependent on a critical tyrosine residue (Tyr503), and placing *o*-nitrobenzyltyrosine in this site effectively reduced the activity of this enzyme to 5% of the wild-type form. After a 30 minute exposure of bacterial cells to long wavelength UV light, the enzyme regained activity to around 70% of wild-type levels. In a third example, protein phosphorylation in yeast is controlled by a photocaged serine, 4,5-dimethoxy-2-nitrobenzylserine (DMNB-Ser, **16**) (Lemke et al., 2007). DMNB can be removed by light with a wavelength of 405 nm, reducing the side effect of UV light to cells. This DMNB-Ser blocks phosphorylation and is incorporated at different phosphoserine sites in the transcription factor Pho4. Upon photo-decaging, serine is regenerated and subsequently phosphorylated, triggering the nuclear export of Pho4 (Figure 4).

This strategy will likely be used to light-activate other enzymes of interest in living cells provided that enzymatic function is dependent on just one or a few

residues of the protein. Another useful application could be the decaging of phosphorylated, ubiquitinated, or otherwise post-translationally modified residues with light at specific time points for examination of signal transduction pathways.

Photolysis of a caged amino acid residue is an irreversible process. Reversible modulation can be achieved with the photochromic azobenzene compounds. Azobenzene undergoes a reversible *cis*-*trans* isomerization: The more stable *trans* isomer can be converted to the *cis* isomer upon illumination at 320–340 nm, and the *cis*-form can revert to *trans*-form either thermally or by irradiation at >420 nm. The resultant change in geometry and/or dipole of the compound can be used for regulating protein activity in a reversible manner. The azobenzene group has been genetically encoded in the form of *p*-azophenyl-phenylalanine (AzoPhe, **17**). AzoPhe was incorporated at the Ile71 site of the *E. coli* catabolite activator protein, a transcriptional activator (Bose et al., 2006). Its binding affinity for the promoter sequence decreased fourfold after irradiation at 334 nm, which converts the predominant *trans* AzoPhe to the *cis*-form. The isomerized *cis* AzoPhe then was switched back to the *trans*-state by irradiation at >420 nm, after which the affinity of the protein for the promoter was completely recovered. Similarly, it should be possible to photomodulate the activity of other enzymes, receptors and transcription factors.

One vast challenge with studying protein function by expression and purification is the extremely large number of post-translational modifications found on proteins, especially those from eukaryotes. These modifications often

play important roles in protein stability, localization, and function, but cannot be easily mimicked or produced in bacterial cells. It is also difficult to prepare homogenously modified proteins from eukaryotic cells. Uaas with the properties of post-translational modifications, or chemical handles for the addition of these modifications, may allow the generation of homogenous proteins with the desired modifications to uncover the functional roles of post-translational modifications, and may enable us to control protein function by controlling the modifications appended to the protein of interest. Many post-translational tags, like glycosylation, lipidation, acetylation, methylation, or phosphorylation can be incorporated into proteins using genetically encoded Uaas.

Phosphorylation is a nearly ubiquitous post-translational modification that can control protein function, conformation, and stability. Phosphorylation of serine, threonine, and tyrosine residues are the key switches in many signal transduction cascades. The ability to precisely control which residues are phosphorylated would enable dissection of which signaling events are tied to modifications on specific residues. Using photocaged serines, threonines, or tyrosines at critical sites, phosphorylation of specific residues could be prevented until a desired moment, a more attractive and versatile method than replacing these residues with natural amino acids that cannot be phosphorylated. On the other hand, it can be difficult to generate a signaling molecule that is constitutively active due to the reversible nature of phosphorylation and the large number of phosphatases in the cell. Creating a mimetic for a phosphorylated residue that cannot be cleaved by

endogenous phosphatases can be used to identify the role of phosphorylation of a single residue, and can create more stable activated proteins for a variety of functional assays.

The Uaa *p*-carboxymethyl-L-phenylalanine (*p*CMF, **18**) is a nonhydrolyzable analog of phosphotyrosine and was found capable of mimicking the phosphorylated state of Tyr (Xie et al., 2007). This capability was demonstrated in a model phosphoprotein, the human signal transducer and activator of transcription-1 (STAT1). STAT1 has only a weak affinity for DNA, but during phosphorylation of Tyr701, STAT1 forms a homodimer and strongly binds a DNA duplex that contains M67 sites. The mutant STAT1 with Tyr701 substituted with *p*CMF also bound the M67-containing DNA duplex tightly, which suggests that *p*CMF could replace phosphotyrosine in the generation of constitutively active phosphoproteins. Non-hydrolyzable analogs of phosphorylated residues can be similarly used to study the functional consequences of constitutive phosphorylation of a number of proteins, and to study phosphorylation induced changes in binding partners or binding affinity more easily.

Acetylation is a reversible modification on proteins that can also contribute to protein localization and function. Acetylation of lysine residues in histone proteins can control the secondary structure of chromatin as well as gene expression levels from certain loci, and chromatin remodeling and its consequences in a variety of molecular and cell biological questions are intensely

researched. Many other proteins undergo reversible acetylation, and the functional consequences of these modifications are poorly understood in many cases. The Uaa *N*-acetyllysine (**19**) was genetically incorporated into proteins in *E. coli* (Neumann et al., 2008). In the presence of a de-acetylase inhibitor, manganese superoxide dismutase (MnSOD) containing *N*-acetyllysine at residue Lys44 was prepared from *E. coli*. Acetylation of Lys44 in MnSOD, based on experiments comparing recombinant MnSOD with or without *N*-acetyllysine at lysine 44, did not appear to affect the activity of the enzyme, however. *N*-acetyllysine and other lysine analogues were also incorporated into GRB2 in HEK293 cells (Mukai et al., 2008).

Uaas can possess such diverse structures that there are a number of them that may be used for purposes previously unidentified. For instance, *p*-nitrophenylalanine (**20**), originally used as a distance probe to quench the fluorescence of tryptophan, has recently been found to stimulate potent immune responses for novel immunogenic applications. In general, an animal will not mount a substantial immune response against a protein that is normally present in the body. It was shown that the incorporation of *p*-nitrophenylalanine into antigens is capable of breaking immune self-tolerance (Grunewald et al., 2008). Incorporation of this Uaa into tumor necrosis factor α (TNF-α) led to an immune response directed at both the Uaa mutant and the natural version of the protein, even in the absence of adjuvant. Incorporation of the nitro group creates a T-cell epitope that was not previously recognized, and mobilizes a robust immune

response against TNF-α. The immune response was retained in mice after exposure, because after lipopolysaccharide challenge, which leads to a TNF-α mediated septic shock, animals immunized with the mutant TNF-α had improved survival rates compared with mock-immunized animals, or those immunized with wild-type TNF-α. These striking observations have the potential for generating new therapies for a number of human diseases. This work further shows the immense diversity that Uaas can play in a number of disciplines and suggests that there are many more unknown applications to be identified.

Since 2001, when the first Uaa was genetically encoded in live cells (Wang et al., 2001), more than 40 new Uaas have been incorporated into proteins in prokaryotic and eukaryotic cells. The number of novel experiments that can be performed increases each day another functional group is made available. Proof-of-principle experiments have demonstrated the utility of many of these Uaas, and their application in addressing challenging biological questions and uncovering unknowns should emerge in the coming years.

The use of this technology in mammalian cells is a burgeoning area, in which a variety of cell biology questions can be answered in the native cellular environment. Proteins containing fluorescent Uaas that behave as sensors of small molecules and protein-protein interactions, or proteins whose activity can be photoregulated *in vivo* are particularly attractive. Mimics of post-translational modifications or photocaged amino acids may be used to control signal transduction components. Genetically encoded Uaas can aid in the identification

of proteins that are critical in the development of different tissue and cell lineages, can tease apart at the molecular level the process of cellular differentiation, and may generate new therapies for a number of human diseases. The ability to control signal transduction events directly in live cells would enable a better understanding of diseases in which signal transduction has gone awry, including cancer.

Although single cells are valuable models, research into development, intercellular communication, differentiation, cancerous transformation, and a variety of other processes necessitates multicellular organisms. The ability to genetically incorporate unnatural amino acids in a multicellular organism would open the field to biological problems concerning how cells interact with one another in a variety of settings, allowing aspects of development, neural connectivity, and cellular signaling to be studied, to name only a few. While a recent paper has incorporated Uaas into *C. elegans* using a pyrrolysyl tRNA/synthetase pair derived from *Methanosarcina mazei* (Greiss and Chin, 2011), the ability to use this technology in multicellular organisms is still in its infancy.

The genetically encoded Uaas, as new building blocks for proteins, should allow the design or evolution of proteins with novel properties. For instance, glycosylated or PEGylated therapeutic proteins may be rationally designed to improve pharmacological properties; various proteins can be engineered with light responsiveness to work as noninvasive molecular tools inside cells; protein

properties never required and that have never existed in nature may be evolved from random Uaa mutagenesis. It may even be possible to experimentally test whether there is an evolutionary advantage for organisms with more than the 20 genetically encoded amino acids.

2. *CAENORHABDITIS ELEGANS* AS A MODEL ORGANISM

In order to expand the genetic code of a multicellular organism, it is important that the organism in question have a number of attributes to enable the successful incorporation of Uaas into proteins. *Caenorhabditis elegans* is a model organism widely used for researching development, neurobiology, aging, meiosis, and epigenetics (Jin, 2005; Riddle, 1997). The nematode has an invariant cell lineage during development (Sulston et al., 1983), making it an attractive model system for many aspects of cellular signaling and cell fate specification. A wealth of knowledge is available on the *C. elegans* genome (Antoshechkin and Sternberg, 2007;), and strong forward and reverse genetics makes the animal particularly amenable to a fine dissection of many biological problems. A number of the attractive aspects of *C. elegans* biology identifying the animal as a likely candidate for genetic code expansion are discussed below.

Certain characteristics of *C. elegans* make the animal attractive for the incorporation of Uaas. It is the only metazoan in which endogenous amber suppressor tRNAs have been identified, demonstrating a tolerance for amber suppression. The first loci identified in genetic screens were *sup-5* and *sup-7*

(Waterston, 1981; Waterston and Brenner, 1978) due to their ability to suppress a subset of mutations in a number of different genes that seemed unrelated. These suppressor mutations cause some toxicity and reproductive problems, especially at lower temperatures, presumably due to the suppression of legitimate stop codons, leading to elongated proteins. The mutated loci associated with this suppression were later identified as mutations in the anticodon loop of tryptophanyl tRNA, changing the sequence to CUA (Bolten et al., 1984; Wills et al., 1983). This mutation enabled the tRNA to pair with the amber stop codon UAG, and by extension each allele that these mutations suppressed were amber nonsense mutations.

Several other $tRNA^{Trp}$ loci have been uncovered in genetic screens searching for additional amber suppressors (Kondo et al., 1988; Kondo et al., 1990). These suppressors exhibit a range of suppression efficiency depending on the locus and mutation being suppressed, suggesting that the flanking sequences around the tDNA gene may play a role in both expression level and tissue specificity. The level of suppression *in vivo*, detected by a lacZ amber reporter, correlates with transcription in cell-free extracts, indicating that differences in suppressor strength is tied to transcriptional activity in the animal (Li et al., 1998). The strength of the suppressor is also inversely correlated to the level of toxicity seen in colder temperatures, indicating that amber suppression in *C. elegans* is likely to be more efficient at colder temperatures. Transgenic tRNA

suppressors derived from *Drosophila melanogaster* $tRNA^{Ser}$ are also functional as amber suppressors in *C. elegans* (Pilgrim and Bell, 1993).

In 2006, the non-coding RNA transcriptome of *C. elegans* was cloned and classified based on a number of features, including biogenesis and expression (Deng et al., 2005). They identified nearly 2700 non-protein coding RNAs, the majority of which are developmentally controlled. The upstream motifs associated with these RNAs were classified based on common elements. Many of these genes are transcribed by RNA Pol I or III, or are processed out of the introns of protein-coding transcripts. It was shown that the cap structure of non-coding RNAs correlates to their upstream motif structure, and a large number of RNA Pol III promoters were identified and catalogued.

Nematodes, while enclosed in a hard cuticle to protect their hydrostatic bodies from environmental assaults, nonetheless take up a diverse set of macromolecules. *C. elegans* is known to take up fluorescein dyes through exposed amphid neurons (Hedgecock et al., 1985), and has been used as a model for drug discovery screens to identify protein targets of small molecules (Kwok et al., 2006). Additionally, it is known that RNAi by feeding *E. coli* expressing double-stranded RNA is quite effective at specifically knocking down gene expression (Timmons and Fire, 1998). The ease with which many types of molecules are taken into the animal, through a number of routes, suggests a strong candidate for Uaa uptake to at least a subset of cells.

The intestine of *C. elegans* expresses a number of transporters to bring digested nutrients from the lumen into the intestinal cells themselves. Both verified and predicted amino acid transporters (*aat* genes) have been identified in *C. elegans* (Veljkovic et al., 2004b). Many are expressed on intestinal cells, though one, *aat-9*, appears to be expressed exclusively on muscle and neuronal cells and is specific for aromatic amino acid transport (Veljkovic et al., 2004a). Furthermore, genes encoding peptide transporters (*opt-1* and *opt-2*) are expressed on a number of tissues in the animal (Fei et al., 1998). These proteins belong to a family of oligopeptide-protein symporters, and the binding determinants for a number of peptide substrates have been well characterized in mammalian homologues (Biegel et al., 2006; Daniel et al., 2006). Oligopeptide transporters are known to recognize and transport a wide range of peptide substrates across the cell membrane, providing a potential target for uptake of exogenously applied and structurally divergent Uaas.

A number of methods are in use currently to introduce transgenes into *C. elegans*, each with their own advantages and disadvantages. Historically, injections of plasmid DNA into the gonad leads to the formation of an extrachromosomal array (Mello et al., 1991) through homologous recombination and concatamerization. This array functions as a miniature chromosome, with multiple (80-300) copies of the injected plasmids present. However, unlike a legitimate chromosome, the array does not follow Mendelian laws of segregation, leading to mosaic inheritance at each meiotic or mitotic event, and mosaic

expression both within and between animals. The array can be integrated by irradiation, leading to random insertion of the entire concatamer into a chromosome. Integration eliminates the mosaicism, but the copy number is still extremely high. Additionally, both extrachromosomal and integrated arrays can be subject to DNA silencing by chromatin modification due to the repetitive sequence elements.

Recently, two new methods for *C. elegans* transgenesis have been reported. Microparticle bombardment involves coating gold beads with the plasmid DNA of interest and shooting the animals using a gene gun (Praitis et al., 2001). This method is ideal to screen large numbers of plasmids, because the gene gun allows a high throughput of animals provided the screening method is robust. Many transgenesis events identified will be extrachromosomal arrays; however, in some cases the bombarded DNA will be integrated in a few copies (1-4) into a chromosome in a random location in the genome, generally when the gold bead is projected directly into a nucleus in the gonad.

Another new method for transgenesis is known as the Mos Single Copy Insertion (MosSCI) method (Frokjaer-Jensen et al., 2008). MosSCI relies on DNA damage created by mobilization of the *MosI* transposon (Bessereau et al., 2001) and subsequent DNA repair from a template (Robert et al., 2008). A library of transposon insertions is available from the Bessereau lab. While the majority of these insertions interfere with gene expression, some insertion sites are between two transcribed genes. These sites are amenable to single copy insertion, which

occurs when a plasmid encoding the transposase is injected with a targeting vector. The mobilized transposon creates a double strand break, which is repaired at some frequency using the targeting vector as a template. The targeting vector contains sequences homologous to the regions flanking the transposon location, with the genes of interest inserted at this locus. This method is advantageous because a single copy can be inserted into a known location in the chromosome, minimizing the positional effects of random gene insertion.

The transparent body of *C. elegans* makes the use of light for experimentation particularly attractive. Since first being expressed in *C. elegans* (Chalfie et al., 1994), fluorescent proteins have been used in the nematode to study numerous biological problems, including synaptogenesis, calcium signaling, and protein localization and trafficking, all in live animals (Hutter, 2006). These same strategies can be applied to Uaas with functional groups that respond to light. Photo-responsive Uaas can be encoded for fluorescent imaging, photocrosslinking and modulation via photolysis, and should greatly expand research in *C. elegans* with broader biophotonic technologies. The model organism *C. elegans* possesses a number of qualities, in addition to the vast amount of genetic resources and an excellently curated library of mutations (*Caenorhabditis* Genetics Center), which make it particularly suited to Uaa incorporation, including the existence of genetic amber suppressor tRNA mutations and a number of excellent transgenesis methods.

The first section of the Introduction (unnatural amino acids) is, in part, a reprint of material as it appears in *Comprehensive Natural Products Chemistry II;* Chemistry and Biology 2010. Parrish, Angela; Wang,Lei., Oxford; volume 5, 2010. The dissertation author was the primary author of this material.

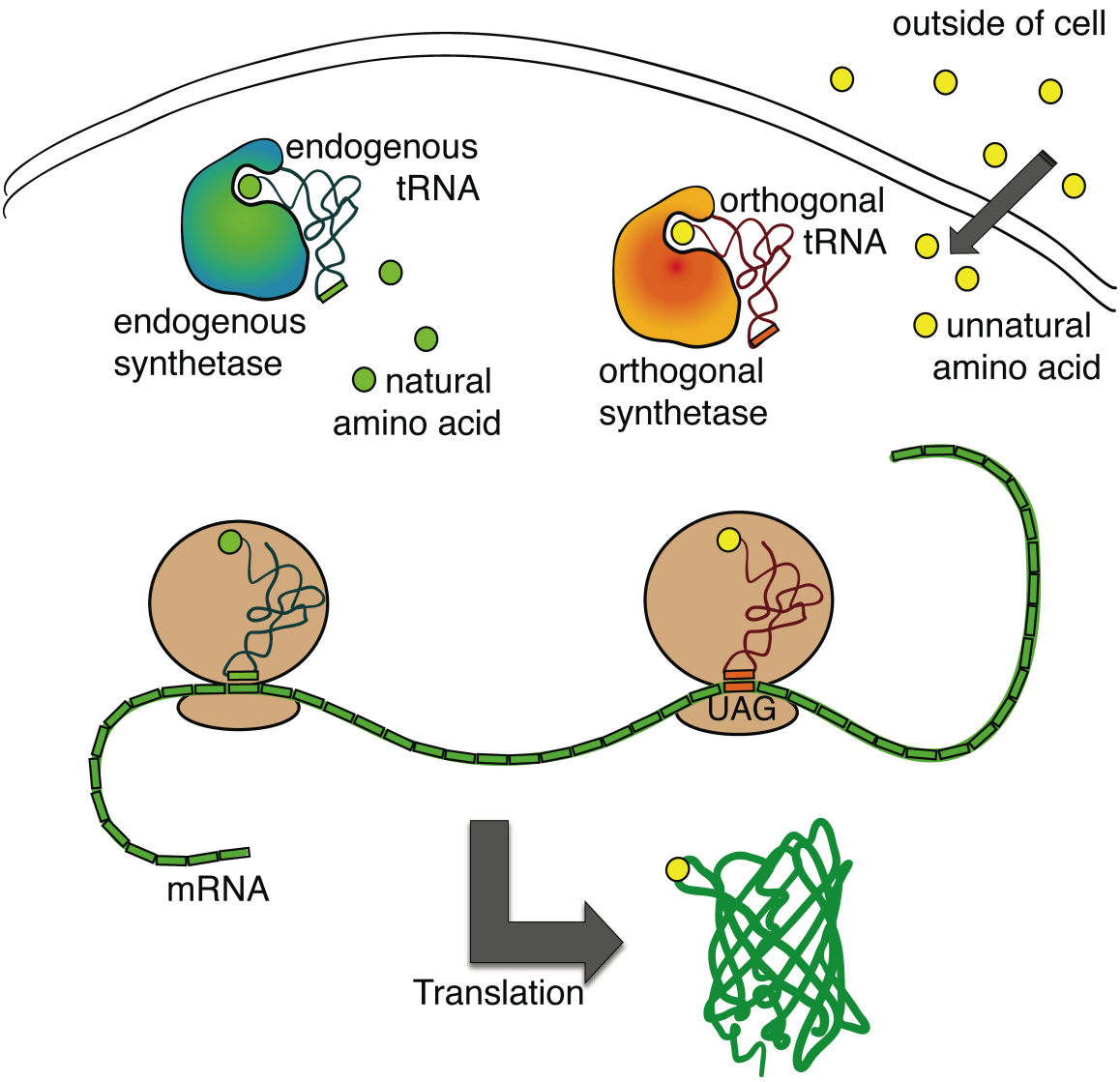


Figure 1: A general method for genetically encoding unnatural amino acids in live cells.

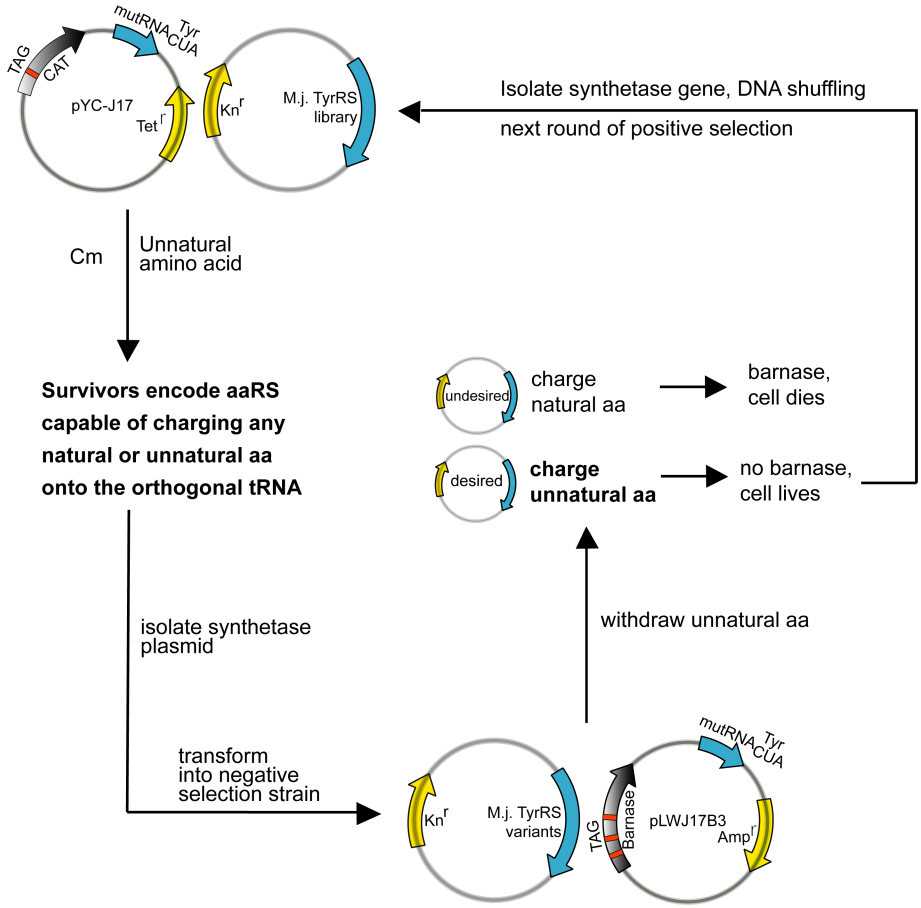


Figure 2: A general positive and negative selection strategy for evolving aminoacyl-tRNA synthetase variants specific for an unnatural amino acid.

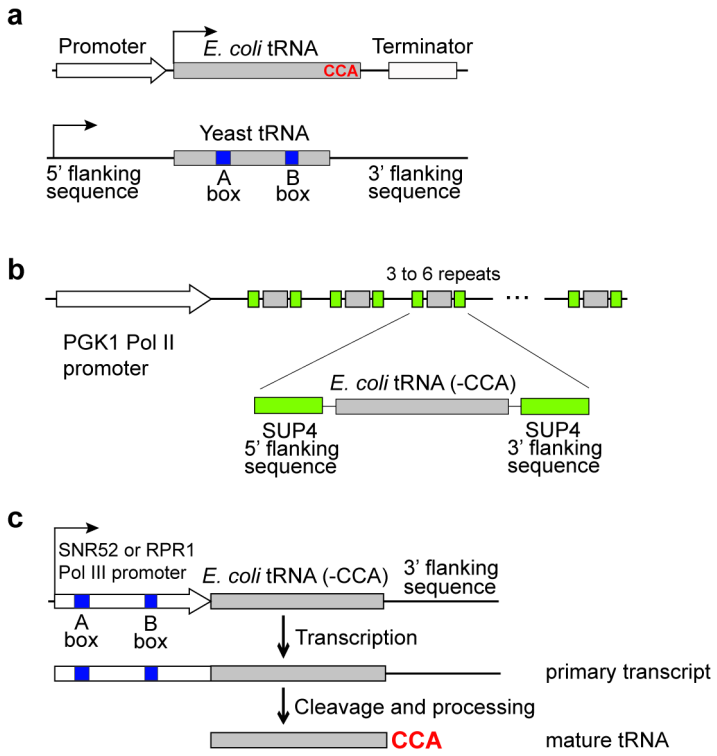


Figure 3: a) Gene elements for tRNA transcription in *Escherichia coli* and yeast. b) Increasing tRNA expression in yeast. A tandemly repeated (3-6 times) tRNA cassette containing 5' and 3' yeast tRNA flanking sequences around an *E. coli* tRNA and driven by an upstream RNA Polymerase II promoter is shown. c) A general method for expressing prokaryotic tRNAs in yeast using an external RNA Pol III promoter that contains the consensus A- and B-box sequences and that is cleaved from the primary transcript.

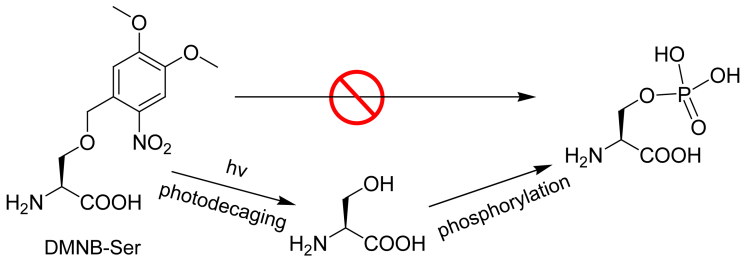
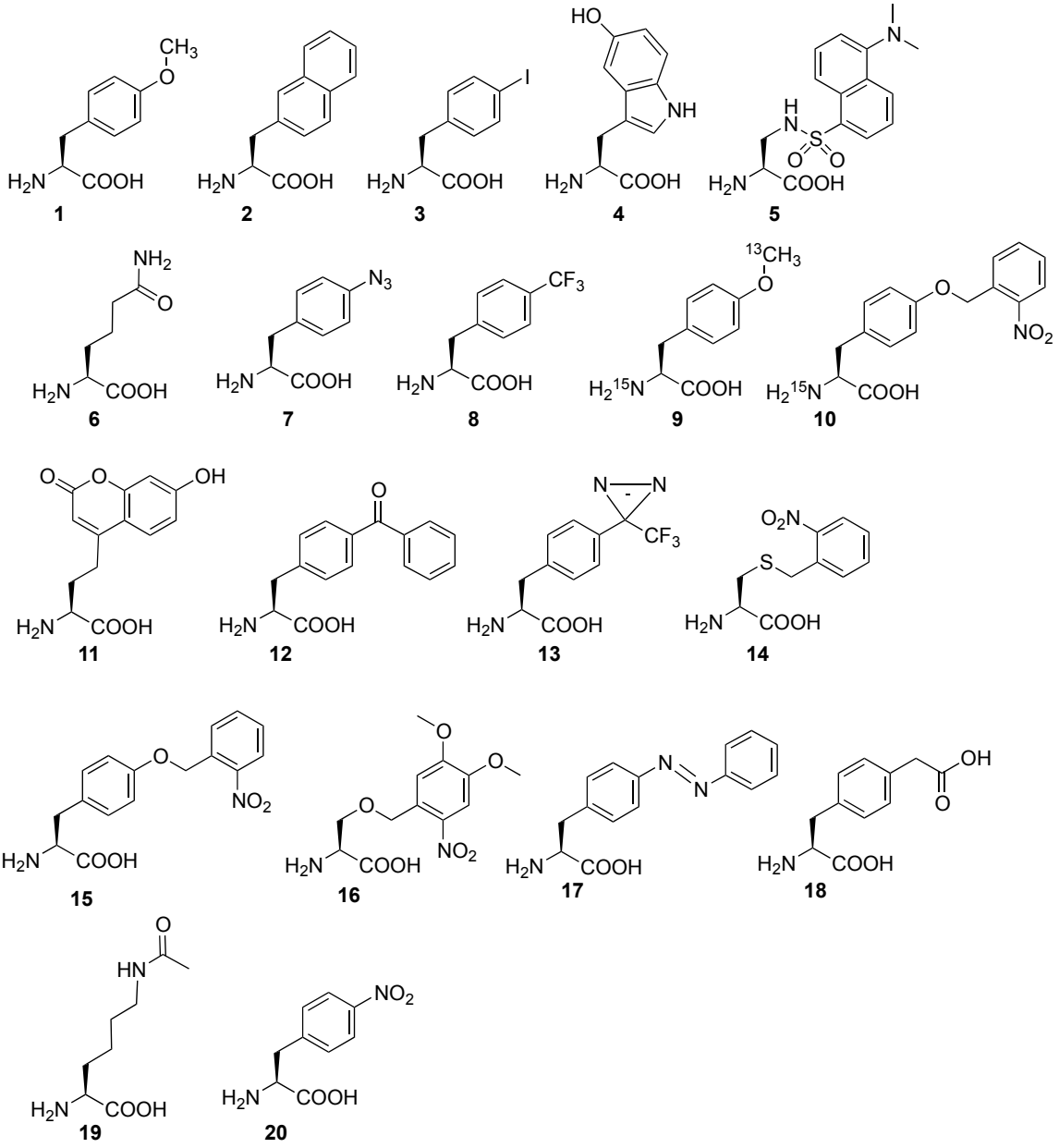


Figure 4: Photocaged serine is used to control when the target serine is regenerated and phosphorylated.



Scheme 1: Structure of unnatural amino acids

RESULTS

1. SCREENING AND DEVELOPMENT OF TRNA/RS PAIRS

In order to identify tRNA/RS pairs that are both functional and orthogonal in *C. elegans*, the first step is the generation of a reliable reporter. It is important that this reporter be sensitive to amber stop codon suppression, yet not have a high false positive rate, obscuring positive hits with a high background level of suppression. For this reason, I chose to use the easily detectable fluorescent protein mCherry. I used the amber stop codon UAG to encode Uaas in *C. elegans* because amber suppression has been successfully employed to incorporate Uaa in single cells, and natural amber suppressors have been isolated in *C. elegans*. Additionally, the amber stop codon is the least frequently used in the *C. elegans* genome, which will ideally mitigate any serious issues from suppressing legitimate stop codons.

I constructed the fluorescent reporter for amber suppression by introducing the UAG codon at a permissive site (codon 156, Tyr) in the fluorescent protein mCherry for body wall muscle expression using the well-characterized promoter for the *unc-54* gene. Due to the toxicity of genuine stop codon suppression, which has been seen in the endogenous tRNA mutants uncovered in forward genetic screens (Waterston, 1981), experiments were done only in body wall muscle to minimize developmental perturbations.

Initially, the reporter (*unc-54p::mCherryTAG156::unc-54 3'UTR*) (see Figure 5 for a schematic of all constructs used) was injected alone to generate extrachromosomal arrays. These injected animals, however, showed diffuse red fluorescence in body wall muscles, with some muscles showing fairly bright fluorescence (Figure 6A). Of worms expressing the co-injection marker, approximately half exhibited detectable fluorescence in the absence of any amber suppressor tRNA/RS. The UAG codon may have mutated or been deleted during array formation (Mello et al., 1991). Moreover, nonspecific stop codon readthrough, to some degree, has been reported in worms (Hodgkin, 1985). To avoid these problems, I integrated a single copy of the fluorescent reporter gene within Chromosome II using the MosSCI method (Frokjaer-Jensen et al., 2008) and confirmed integrity of the UAG codon with DNA sequencing. The resultant reporter strain, LWA1560 (*wISi151[unc54::mCherryTAG156 cb-unc-119(+)]II*) (see Table 1 for the full genotype of each strain used), showed zero detectable fluorescence (other than intestinal autofluorescence) in the absence of a suppressor tRNA/RS, but consistent red body wall muscle fluorescence when crossed to the *C. elegans* endogenous amber suppressor *sup-7(st5)* (Figure 6B,C) (Bolten et al., 1984), indicating that this reporter reliably reflected amber suppression. I therefore used a single integrated copy of reporter genes for all experiments.

Effective Uaa incorporation requires that a tRNA/RS does not interact with any endogenous tRNA/RS pairs. A tRNA/RS pair taken from a different kingdom

of life is likely to be orthogonal (Wang and Schultz, 2004), directing my focus to bacterial and archaeal pairs. Due to the difficulty of expressing prokaryotic tRNAs in eukaryotic systems, I needed to identify a method for tRNA expression in *C. elegans*, as simply transplanting bacterial tRNA genes into eukaryotic tRNA expression cassettes rarely generates functional tRNAs. Previously, this problem has been circumvented in yeast (Wang and Wang, 2008), mammalian (Wang et al., 2007), and stem cells (Shen et al., 2011) using external Pol III promoters, which contain all necessary promoter elements.

I tested Pol III *C. elegans* promoters to drive the expression of *E. coli* tRNA in worms. The promoters for the non-coding RNA genes *cen-4, cen-38,* and *rpr-1* were identified from the *C. elegan*s noncoding transcriptome (Deng et al., 2005) as candidates. These promoters were placed at the 5' end of the *E. coli* amber suppressor tRNA (without the last trinucleotide CCA) and followed by the 3' flanking sequence of the *C. elegans* $tRNA^{Lys}$. The endogenous 3' flanking sequences of tRNA encourage proper processing of prokaryotic tRNAs in eukaryotic systems, as does deleting the last trinucleotide from the introduced tRNA (Wang and Wang, 2008; Wang et al., 2007). Plasmids containing the candidate promoter driving the *E. coli* leucyl amber suppressor tRNA ($tRNA^{Leu}_{CUA}$) and the *E. coli* leucyl-tRNA synthetase (LeuRS) driven by the *unc-54* promoter (*unc-54p::LeuRS::unc-54 3'UTR_promoter::* $tRNA^{Leu}_{CUA}$) were microinjected into the mCherry reporter line and maintained as extrachromosomal arrays to create

LWA1561(*wISi151 II, wIEx1[unc-54::LeuRS_rpr-1::* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ *+ pRF4(rol-6)]*). Only if the tRNA is correctly expressed and processed will it be charged with Leu to suppress the amber codon within mCherry, generating red fluorescence. After screening I found that the *rpr-1* promoter showed strong reporter activation (Figure 6D), indicating that it can drive the functional expression of *E. coli* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ in *C. elegans*. Moreover, when a plasmid containing *rpr-1* driven $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ was injected in the absence of LeuRS, no red fluorescence was detected (Figure 6E), indicating that the *E. coli* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ is orthogonal to *C. elegans.*

To test if the *rpr-1* promoter can be generally used to express other *E. coli* tRNAs, I replaced the *E. coli* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ with *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$. When the *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ was expressed with its cognate *E. coli* TyrRS, the strain LWA1562 (*wISi151 II, wIEx22[unc-54::TyrRS_rpr-1::* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ *+ pRF4(rol-6)]*) showed red fluorescence in body wall muscle (Figure 6F). Additionally, in the absence of the *E. coli* TyrRS, no red fluorescence was detected, indicating that the *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ is orthogonal to *C. elegans*. The tyrosyl tRNA/RS pair from the archael organism *Methanococcus jannaschii* was also tested using the same experimental approach, and while fluorescence was detected, it was the weakest suppressor of the three pairs tested and therefore was not pursued.

Amber suppression in *C. elegans* using my orthogonal tRNA/RS system showed robust temperature dependence: at 15 ^{o}C there was strong fluorescence

in body wall muscles, but at 20 ^{o}C fluorescence was weak. This result is consistent with the previously reported temperature dependence of endogenous *C. elegans* amber suppressors (Waterston, 1981), further suggesting our system is functional for amber suppression.

Delivery of Uaas into the cells of interest is another potential hurdle to Uaa incorporation in multicellular animals. For tissue culture cells and unicellular organisms, addition of the Uaa to the growth medium is sufficient for the uptake of many Uaas (Takimoto et al., 2010). However, *C. elegans* has a protective cuticle that excludes many compounds from internalization, and an active digestion system that could potentially degrade Uaas before assimilation by the intestine. We reasoned that Uaas might be taken in as food via the intestinal route and trafficked to multiple tissues. However, intestinal transporters may not recognize Uaas with drastically different structures. We chose two Uaas to study their uptake in *C. elegans*: *o*-methyl-L-tyrosine (OmeY, **1**), structurally similar to tyrosine, and 2-amino-3-(5-(dimethylamino) napththalene-1-sulfonamido) propanoic acid (DanAla, **5**), a fluorescent amino acid that shares no structural similarity to any natural amino acids (see Scheme 1, also Figure 7A).

Orthogonal synthetases specific for Uaas have been evolved in *E. coli* and later in yeast from large mutant synthetase libraries consisting of >10^9 members (Wang et al., 2001). Similar evolution strategies cannot be used in *C. elegans* due to inefficient transgene delivery and low throughput of phenotypic selection. It has been shown that transferring a synthetase evolved in bacteria or yeast to

tissue culture cells is a useful strategy to generate Uaa-specific synthetases for systems where directed evolution is not possible (Liu et al., 2007; Takimoto et al., 2011; Wang et al., 2007). Mutant synthetases specific for OmeY (OmeRS, derived from *E. coli* TyrRS) and DanAla (DanRS, derived from *E. coli* LeuRS) have been generated in *S. cerevisiae* (Summerer et al., 2006; Wang and Wang, 2008), suggesting that they could be transferred into *C. elegans* for incorporation of the respective Uaa.

To incorporate OmeY into mCherry, I exposed gravid mCherry reporter animals harboring an array encoding *unc-54p::OmeRS::unc-54 3'UTR_rpr-1p::* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ (LWA1563) to OmeY in liquid culture. One millimolar OmeY showed no reporter activation. Five millimolar OmeY led to reporter activation in more than half of L3-L4 progeny containing the array (Figure 7B,C). Only animals reared on OmeY from the embryo stage exhibited reporter activation, whereas animals placed in OmeY culture after hatching lacked fluorescence, indicating that there is a time in development after which Uaas cannot be transferred to body wall muscle. In addition, most negative controls grown in the absence of OmeY lacked fluorescence; only 10% showed very faint fluorescence in 1 or 2 muscle cells. These results indicate that the *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ and the evolved OmeRS specifically insert OmeY into mCherry in *C. elegans*. To test the delivery method of OmeY, *E. coli* expressing proteins containing OmeY (ydii and Z-domain) were plated on OmeY media, though the reporter activation seen at 5 mM OmeY was not greater than free OmeY.

For the incorporation of DanAla, feeding free Uaa at 1 or 2.5 mM did not activate the reporter in animals with *unc-54p::DanRS::unc-54 3'UTR_rpr-1p::* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ (LWA1564). OmeY might enter *C. elegans* and reach muscle cells through pathways used by Tyr or Phe due to structural similarity, but DanAla is too divergent to follow natural uptake routes. Initial attempts to localize DanAla were hindered by the overlap of gut granule fluorescence (Figure 8A). Using a *glo-4 (ok623)* (Hermann et al., 2005) mutant that lacks fluorescent gut granules (Figure 8B), I found that DanAla was sequestered in the intestinal cells, which became fluorescent only after feeding DanAla (Figure 8C).

OPT-1 and OPT-2 are dipeptide transporters present on the surface of many *C. elegans* cells (Fei et al., 1998). I reasoned that an Ala-Uaa dipeptide should enter cells through dipeptide transporters due to their wide range of substrates. The dipeptide would then be hydrolyzed by cellular peptidases to generate the free Uaa inside cells. I devised a strategy for intestinal uptake of DanAla by using an Ala-DanAla dipeptide. Unlike free DanAla, the Ala-DanAla dipeptide was not trapped within the intestine unless high concentrations were used (Figure 8D-E). When gravid worms were fed 1 mM Ala-DanAla dipeptide, mCherry reporter activation was observed in body wall and vulval muscle cells of L3-L4 stage progeny (Figure 7D,E). In the absence of the dipeptide, no mCherry expression was detected. As the DanRS is specific for DanAla, these results suggest that DanAla entered muscle cells and was incorporated into mCherry. A number of additional methods were attempted to deliver dansylalanine to muscle

cells, most of them chemical or electrical methods to disrupt the worm cuticle, but these methods were not effective.

Several experiments to incorporate Uaas into endogenous *C. elegans* proteins were attempted, which are briefly mentioned here. I tried incorporating DanAla into the Wnt protein MOM-2 by injecting plasmids encoding the gene in order to visualize protein trafficking using fluorescence; however, the expression level of the full-length protein was extremely low, and as described above, DanAla overlaps significantly with autofluorescent intestinal granules, which are just forming at the developmental time point in which I was interested. Furthermore, as a maternally expressed gene, it is likely the extrachromosomal arrays I injected were silenced at the developmental timepoint needed for protein labeling.

The myosin heavy chain gene *unc-54* has several amber alleles with a range of phenotypes that were reported to be suppressible by *sup-7(st5)* (Bolten et al., 1984). These alleles were crossed to animals harboring the extrachromosomal arrays encoding the Uaa machinery. A number of assays to detect phenotypic reversion of the *unc-54* mutations were performed, including examining the muscle structure by immunofluorescence for actin and UNC-54, animal mobility assays, and mass spectrometry, but none of these methods indicated Uaa incorporation. After integrating the Uaa machinery (next section), these assays were repeated, with identical results. It is likely that reversion in *unc-54* mutants was not observed because, as a structural gene, a large amount

of protein product is necessary to detect a phenotypic change and Uaa incorporation is simply not robust enough. Attempts to visualize the protein by Western blot following immunoprecipitation had similarly inconclusive results.

2. OPTIMIZATION OF UAA INCORPORATION IN *C. ELEGANS*

To improve incorporation efficiency and generate animals with 100% transmission of the Uaa incorporation machinery, the extrachromosomal arrays harboring the orthogonal tRNA/RS genes were integrated using γ-irradiation. In contrast to the mosaic expression of the extrachromosomal array caused by random mitotic segregation, the integrated strain LWA1565 (*wlSi151 II; wlls1565[unc-54::DanRS_rpr-1::* $\mathrm{tRNA}_{\mathrm{CUA}}^{\mathrm{Leu}}$ *+ myo-2::GFP]X*) displayed uniform non-mosaic expression of the fluorescent reporter in all body wall muscles (Figure 9). After integration, expression of the reporter gene was less variable, ensuring all animals within a population had similar levels of Uaa incorporation and target gene expression.

Further optimization indicated that the stably integrated animals exhibited more robust reporter activation on solid medium compared to liquid (Figure 10). The effect of *smg-1* RNAi was also examined at this point, to identify if stabilizing mRNA containing a premature termination codon would dramatically affect mCherry expression. No difference between control and treated groups were seen. I also noted that LWA1565 animals grown in the presence of Ala-DanAla dipeptide developed slower with fewer progeny than those grown without Uaa,

but these effects were more pronounced in liquid than solid medium. Animals grown on plates appeared healthier for most conditions with dipeptide concentrations up to 3 mM.

A quantifiable reporter would enable me to optimize the components for amber suppression and protein expression. A different target protein for Uaa incorporation would also test whether my strategy can be generally applied to other proteins. I decided to use a luciferase because the enzymatic oxidation of luciferin offers sensitive detection of Uaa incorporation by measuring the resultant luminescence.

A permissive site for mutagenesis (codon 185, Ser) in the *Photinus pyralis* (North American) firefly luciferase was identified based on its structure and verified using amber suppression in mammalian cells. A single copy *unc-54p::luciferaseTAG185::unc-54 3'UTR* was integrated within Chromosome I using MosSCI to generate the stable luciferase reporter strain LWA1850 (*wlSi1852[unc-54::luciferaseTAG185 cb-unc-119 (+)]I*). Wild type luciferase was similarly integrated as a positive control, from which robust luciferase activity was measured after worm lysis. When the (Ser185TAG) luciferase reporter line was crossed to a line harboring an integrated $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$/TyrRS array to generate LWA1851 (*wlSi1852 I; wlls1851[unc-54::TyrRS_rpr-1::* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ *+ myo-2::GFP]X*), luciferase activity was readily detected in LWA1851 (Figure 11). In the absence of the suppression machinery, no luciferase activity was detected from the reporter line, confirming that single integration of the amber-containing

luciferase gene had no nonspecific UAG readthrough. These results demonstrate the suitability of the luciferase reporter strain for quantification.

To test for Uaa incorporation efficiency and the optimal Uaa dosage, I crossed the luciferase reporter line to integrated tRNA/RS array lines, creating strain LWA1852 (*wlSi1852; wlls1853[unc-54::OmeRS_rpr-1::* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ *+ myo-2::GFP]X*) and LWA1031 (*wlSi1852 I; wlls1565[unc-54::DanRS_rpr-1::* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ *+ myo-2::GFP]X*). Animals were grown on solid media with varying concentrations of Uaas. A dose dependent increase in luciferase activity was observed for both OmeY and Ala-DanAla dipeptide, up to concentrations where growth and fertility became limiting (Figure 12). For instance, at concentrations above 3 mM the Ala-DanAla dipeptide made the medium too acidic, inhibiting bacterial lawn growth and leading to few surviving animals; however it is also possible that the animals simply do not thrive at high concentrations of any Uaa. Similarly increasing the dosage of Phe or Tyr had little to no effect on luciferase activity (Figure 13). Uaa dosage dependence together with the established specificity of the evolved RS for the Uaa suggests that the Uaa was not metabolized but incorporated unaltered into luciferase in *C. elegans*.

I also analyzed incorporation efficiency in relation to Uaa placement within the dipeptide. Surprisingly, I found a dramatic increase of DanAla incorporation using DanAla-Ala instead of Ala-DanAla (Figure 14), which may be caused by enhanced accessibility of DanAla-Ala to cellular peptidases or dipeptide transporters.

Previously, it was found that a higher ratio of tRNA to RS improves Uaa incorporation efficiency in mammalian cells (Coin et al.). To test if this extends to *C. elegans,* the ratio of tRNA to RS expression cassettes in the animal was altered. A single $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ expression cassette (*rpr-1p::* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$), as well as the wild type LeuRS and the luciferase amber reporter, both driven by the *unc-54* promoter, were integrated using MosSCI into different chromosomes to create LWA1008 (*wlSi1852 I; wlSi17[(rpr-1::* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ *)]II; wlSi550[unc-54::LeuRS]IV*). Another strain, LWA1009 (*wlSi1852 I; wlSi13[rpr-1::* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ *3x]II; wlSi550*), was similarly generated except that the single $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ expression cassette was replaced with 3 tandem copies. A comparison of these 2 lines showed that an increase in tRNA copy number leads to 19 times higher luciferase activity (Figure 15A).

Next, I tested if this effect could be recapitulated in animals incorporating Uaas. Three tandem copies of the *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ expression cassette were integrated into Chromosome IV for use with OmeRS. I crossed the integrated tRNA/RS array lines to those with the triple tRNA expression cassettes, generating LWA1855 (*wlSi1852 I; wlSi1082[(rpr-1::* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ *)3x]IV; wlls1853 X*) to incorporate OmeY and LWA1856 (*wlSi1852I; wlSi17 II; wlls1565 X*) to incorporate DanAla. When incorporating either OmeY or DanAla (using either dipeptide orientation), there was indeed an increase in luciferase activity with a higher tRNA copy number (Figure 15B,C).

Unlike single cell systems, a multicellular organism will alter its gene expression and small molecule uptake throughout the course of development and aging. To perform precisely timed studies, it is necessary to discern the timing of Uaa uptake and clearance. Understanding Uaa uptake can improve incorporation efficiency and permit pulse-chase experiments using Uaas.

I previously observed that only animals raised in the presence of Uaa from embryonic stages were responsive to Uaa incorporation, whereas those exposed to Uaa during later stages showed no reporter activation. Aliquoting fewer animals initially allowed worms to grow for 3 generations on Uaa before assaying for luciferase. Whereas negative control worms showed no change, worms exposed to the Uaa over 3 generations showed an increase in luciferase activity compared to those exposed for only 1 generation (Figure 16A,B). This increase suggests enrichment or transmission of Uaas from hermaphrodite to progeny occurred. The effect was observed for both Uaas examined and may be general for other Uaas. However, there was no further increase in luciferase activity with growth for 6 generations on Uaa containing media (Figure 16C).

To determine if the Uaa incorporation persists after worms have been removed from Uaa plates, animals were placed on unsupplemented NG for 1 generation after exposure to 2 mM OmeY. For animals grown on Uaa plates for 1 or 3 generations, the luciferase activity dropped to background levels after a single generation on normal plates, with no lingering protein expression (Figure 16C). This effect was consistent for all concentrations of Uaa examined (Figure

17A,B). The ability to turn reporter activity on and off with Uaa exposure further confirms the necessity of the Uaa and its incorporation into the reporter.

Experiments to use small molecule chaperones to aid in target protein folding were performed. It is known that when protein translation is slowed due to a large number of rare codons, protein folding, stability, and function can be compromised (Coleman et al., 2008; Kimchi-Sarfaty et al., 2007). Glycerol has been shown to function as a molecular chaperone in both mammalian tissue culture and *C. elegans* (Deocaris et al., 2008). In order to aid in luciferase folding, animals were grown on plates containing 400 mM glycerol. A modest effect was seen (Figure 18A-C), but overall luciferase production was not increased dramatically. Additionally, animals grew more slowly on glycerol but appeared slightly healthier.

To biochemically demonstrate that robust expression of target proteins is dependent on Uaa supplementation, I integrated the *Luciola cruciata* (Japanese) firefly luciferase protein as a third reporter. This protein (JFFluc) is tolerant to a C-terminal 3xFLAG tag and is well expressed in *C. elegans* body wall muscle. A permissive site (codon 158, Val) was identified in mammalian cells using amber suppression, and cloned using the Gateway system to make the plasmid *unc-54::JFFluciferaseTAG158_ICRH2B:GFP*. This plasmid was used for MosSCI integration into the same site in Ch. I to generate line LWA1582 (*wlSi1582[unc-54::JFFluciferaseTAG158 cb-unc-119 (+)]I*), which was crossed to LWA1852 to generate LWA717 (*wlSi1582 I, wlSi1082 IV, wlls1853 X*). Luciferase assays

comparing the suppression activity of JFFluc versus the original luciferase shows a dramatic increase in luciferase activity, though a concomitant increase in background activity is also detected (Figure 19).

These animals were incubated at 15 ^{0}C on 4 mM or 0 mM OmeY plates (35 for each sample), and after two generations immunoprecipitation with FLAG beads was performed on lysates. Incubating the animals in 4 mM OmeY leads to large amounts of JFF luciferase protein puled down from lysates, while in the absence of OmeY, small amounts of JFF can be detected in the first two elutions, as expected based on the luciferase assays (Figure 20A), further indicating that the tRNA/OmeRS pair is not completely orthogonal in *C. elegans*. There is an obvious response to OmeY, however, as substantially more protein is pulled down in samples that have been exposed to 4 mM OmeY. Band densitometry determines that at 0mM OmeY, elutions 1 and 2 show 3-5% of the intensity of the corresponding elutions at 4 mM OmeY, elution 3 has 0.08% of the intensity, and a band could not be detected in the 0 mM sample for elution 4.

The effect of *smg-1* RNAi was tested in this optimized system. Unlike previous experiments, RNAi knockdown of *smg-1*, a component of the NMD machinery in *C. elegans*, leads to an increase in immunoprecipitated protein when suppressed with both tyrosine and OmeY, while a GFP control shows no enrichment (Figure 20B). In two independent experiments, the fold change in band density on the Western blot ranges from 2.6 to 8.6 for OmeY incorporation, likely dependent on the effectiveness of the RNAi knockdown, when compared to

the GFP control. The tyrosyl synthetase shows a fold change of 10, which is unsurprising as tyrosine is not a limiting factor in the cell. In my optimized system, RNAi knockdown of nonsense mediated decay components can improve Uaa incorporation.

3. DISCUSSION

A common challenge for incorporating Uaas in eukaryotes is the expression of a functional orthogonal tRNA, which often lacks the intragenic promoter elements conserved in eukaryotic tRNAs necessary for transcription by Pol III. A previously developed general method to express orthogonal tRNAs uses a post-transcriptionally cleavable Pol III promoter in yeast (Wang and Wang, 2008) and type-3 Pol III promoters in mammalian and stem cells (Shen et al., 2011; Wang et al., 2007). Here I demonstrate that this strategy can be extended to *C. elegans*. The *rpr-1* promoter efficiently drives the expression of *E. coli* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ and $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ for functional translation in worms. The *rpr-1* promoter is a homolog of the H1 promoter in human cells that is used to express orthogonal tRNAs in mammalian cells. Both promoters are type-3 Pol III promoters for expressing the RNA component of RNaseP, which is involved in tRNA processing and required for all cells at various developmental stages. Therefore, the use of *rpr-1* promoter for orthogonal tRNA expression should be generally compatible for Uaa incorporation in different tissues and stages in *C. elegans*. Other members of the type-3 Pol III family may also work in a similar manner.

The method of target gene expression containing a UAG codon in *C. elegans* is crucial for accurate reporting of Uaa incorporation. Microinjection and microparticle bombardment can both result in the formation of extrachromosomal arrays, which contain multiple copies of the transgene for high-level expression. When a UAG-containing gene is expressed in an extrachromosomal array, background UAG readthrough is detectable even in the absence of an amber suppressor tRNA. Low levels of stop codon readthrough have been reported in *C. elegans* (Hodgkin, 1985), and my observation of red fluorescence in worms injected only with the mCherry amber reporter confirms this phenomenon. The nonspecific UAG readthrough is likely due to insertions or deletions near UAG and/or context codons, which can occur during the recombination that forms extrachromosomal arrays (Mello et al., 1991; Stinchcomb et al., 1985). In contrast, single integration of the same mCherry reporter gene showed absolutely no background red fluorescence. Consistently, single integration of the luciferase amber reporter in worms showed virtually no luciferase activity. Therefore, a low copy number of the UAG-containing reporter gene is necessary to avoid false positives in reporting amber suppression and Uaa incorporation.

Greiss *et al.* recently reported the incorporation of Lys analogs into a linker region between two fluorescent proteins in *C. elegans* (Greiss and Chin, 2011). However, the authors expressed the amber reporter gene in extrachromosomal arrays, which can result in false positives, as I found here. In addition, when the amber stop codon is placed upstream of the open reading frame of a fluorescent

protein, it has been observed that the protein is expressed through internal initiation of translation in the absence of amber suppression in mammalian cells (Coin et al., 2011) and stem cells (Shen et al., 2011). Greiss *et al.* placed the amber codon upstream of the mCherry; internal initiation of translation, rather than Uaa incorporation at UAG, could lead to their observed red fluorescence in *C. elegans*. Moreover, the positive rate (1-5 per hundreds of worms) in Greiss's report is very low and comparable to the rate of apparent UAG readthrough due to mutational effects caused by array formation. This low positive rate will make it extremely difficult, if not impossible, to identify a strain for expressing an exogenous gene that has no sensitive and straightforward readout like fluorescence. In contrast, my integrated single copy reporter had zero false positive signals.

In addition, I placed the amber codon within the coding region of three different reporters, thereby eliminating the possibility of generating functional proteins through internal initiation of translation. When an orthogonal tRNA/RS was injected into the reporter line, I obtained strains that incorporate Uaas with a success rate of 25-50%. This high rate should facilitate the generation of transgenic worms to incorporate Uaa into different target genes. The mCherry fluorescence and luciferase activity I observed in *C. elegans* were Uaa-dependent with dosage and on-and-off switch effects, indicating that Uaas were incorporated into the reporter proteins. Both a fluorescent protein and two enzymes retained activity upon incorporation of an Uaa in the coding sequence

of the protein. The different tRNA/RS pairs, Uaas and target proteins I employed suggest that my approach can be generally applied to various Uaas and proteins.

Low-copy integration of amber-containing target genes is desirable for the generation of Uaa-containing target proteins to investigate biological processes in *C. elegans*. Nonspecific amber readthrough in extrachromosomal arrays will result in natural amino acids at the amber site in the target protein, and this lack of fidelity for Uaa incorporation will impair rigorous studies of the target protein. In addition, array expression results in inconsistency both between animals of the same genotype and within a single animal, making it difficult to interpret results. By integrating the extrachromosomal array expressing the orthogonal tRNA/RS pair, I generated a population of genetically identical animals, abrogating the need to constantly select for transgenic animals. A stable transgenic worm capable of Uaa incorporation should facilitate analysis wherein mosaic transgenic worms can impose complications. Moreover, the ability to detectably suppress the amber codon in a single copy gene suggests a fairly high efficiency for my Uaa incorporation machinery. Therefore, I may be able to recapitulate natural levels of gene expression, rather than overexpression, when generating Uaa-containing proteins, especially if this mRNA is stabilized by a knockdown of NMD machinery. This close-to-native condition will help obtain physiologically relevant results.

A multicellular organism such as *C. elegans* imposes more challenges on the bioavailability of Uaa than single cells. DanAla, a Uaa deviating from natural

amino acids in structure, was sequestered in the intestine. Supplying DanAla in dipeptide form enables DanAla to traverse the intestine and reach muscle cells, suggesting that dipeptide transporters tolerate structural deviations of substrates more readily than amino acid transporters. Since DanAla is so drastically different from natural amino acids, its successful uptake in dipeptide form suggests that other Uaas can be delivered into *C. elegans* using this strategy. As dipeptide transporters are present on most of *C. elegans* cells, this method has the potential to deliver Uaas to different tissues.

The development of a multicellular animal involves many complex interactions between proteins. Genetically encoded Uaas will afford new chemistries to enable finer investigation and manipulation of protein functions directly in living *C. elegans*. It is likely that photo-responsive Uaas will be used for to study a number of biological problems in *C. elegans* in a non-invasive manner. My strategies should be generally applicable for the incorporation of various Uaas in worms, and should be valuable in guiding the genetic code expansion of other multicellular organisms.

The majority of this section has been submitted for publication of the material. Parrish, Angela; She, Xingyu; Xiang, Zheng; Coin, Irene; Dillin, Andrew; Wang, Lei. The dissertation author was the primary investigator and author of this material.

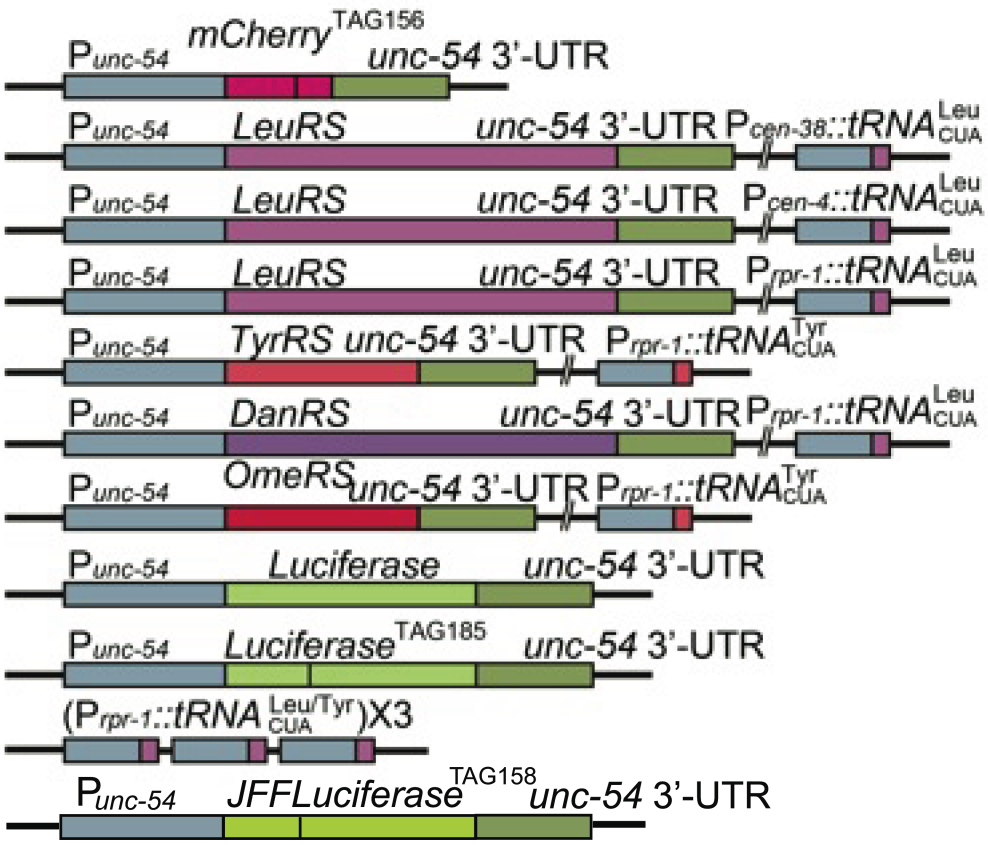


Figure 5: Representation of the plasmids used to make the animals in this work. All tRNA 3' flanking sequences are taken from *C. elegans* $tRNA^{Lys}$.

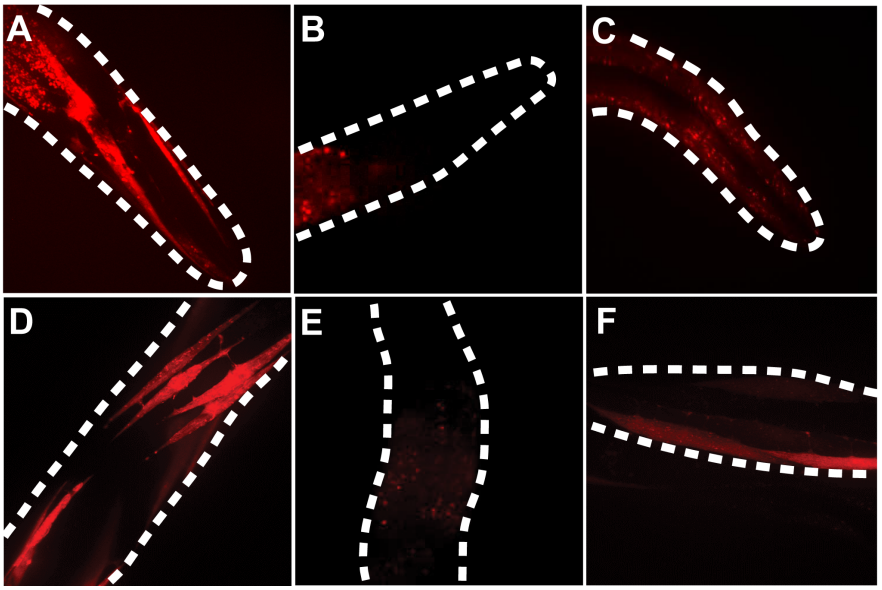


Figure 6: A) The extrachromosomal array *unc-54p::mCherryTAG156::unc-54 3'UTR* showed fluorescence in muscle cells in the absence of amber suppressor tRNA/RS. B) A single integrated copy of the same mCherry amber reporter in strain LWA1560 had no detectable fluorescence in muscle cells. Weak autofluorescence is from intestinal granules. C) LWA1560 crossed with an endogenous amber suppressor (*sup-7(st5); wISi151*) showed uniform red fluorescence in muscle cells. D) LWA1561 showed strong fluorescence in some body wall muscles. Only a subset of muscles were fluorescent due to the mosaic nature of the extrachromosomal array expressing the *E. coli* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ /LeuRS. E) LWA1560 with *rpr-1p:* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ + *pRF4(rol-6)* had no muscle fluorescence. F) LWA1562 showed red fluorescence in some body wall muscles. Dashed lines indicate worm outline. All images are confocal Z stacks and representative of each genotype at young adulthood.

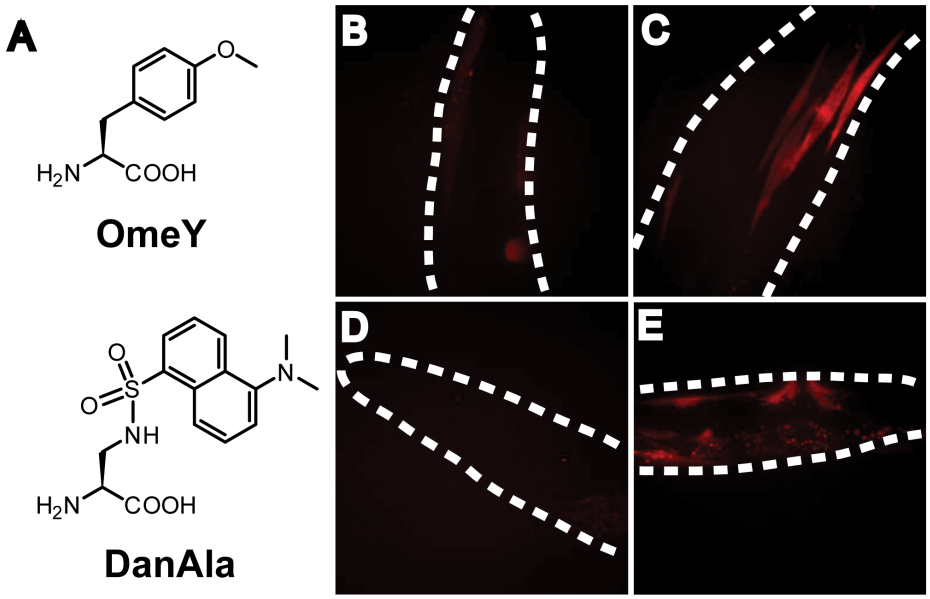


Figure 7: A) Structure of Uaas. B) LWA1563 showed very little fluorescence in the absence of OmeY. C) LWA1563 grown with 5 mM OmeY showed strong mosaic activation of the mCherry reporter in muscle cells. D) LWA1564 showed no fluorescence in the absence of Uaa. E) LWA1564 grown with 1 mM Ala-DanAla dipeptide showed strong mosaic activation of the mCherry reporter in body wall and vulval muscles. All images are confocal Z stacks and representative of each genotype. Dashed lines indicate worm outline. Animals were grown in liquid culture for 1 generation and imaged at young adulthood.

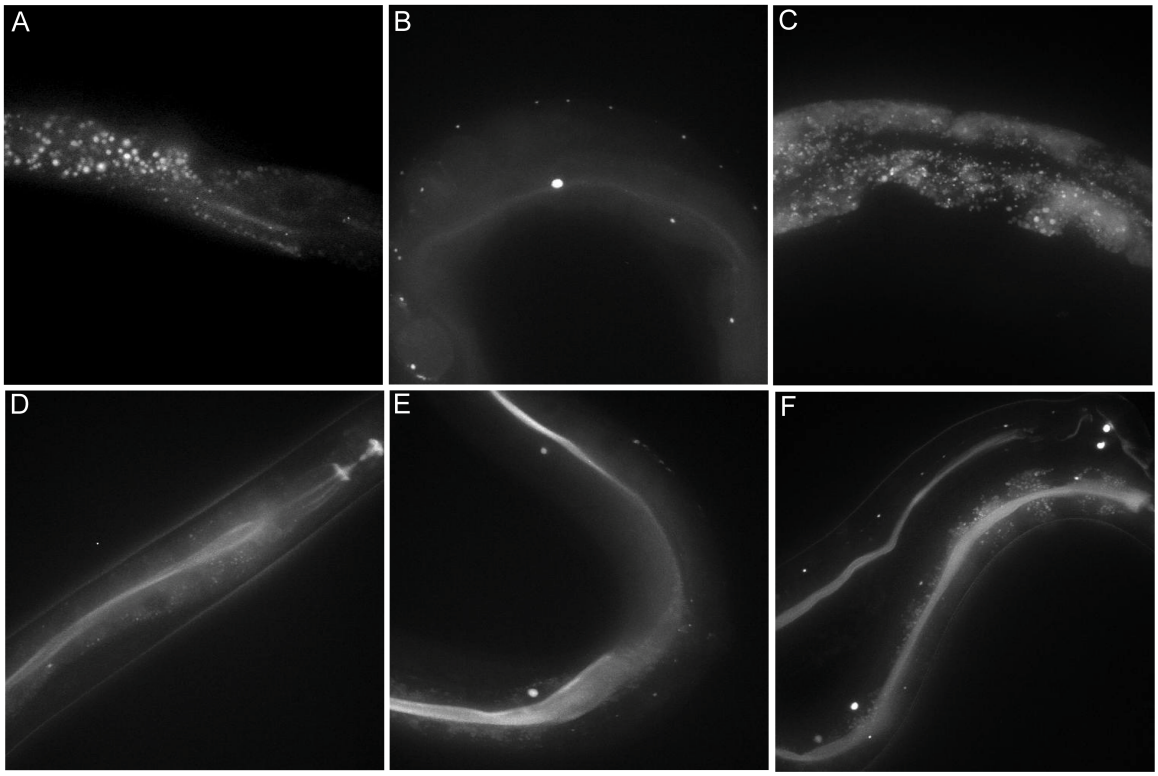


Figure 8: A) N2 worms have autofluorescent intestinal granules. B) *glo-4 (ok623)* does not have autofluorescent intestinal granules. C) After exposure to 1 mM DanAla, intestinal granules of sequestered DanAla appeared in *glo-4 (ok623)*. D) *glo-4 (ok623)* after exposure to 1 mM Ala-DanAla dipeptide. There was some diffuse fluorescence in the lumen and intestinal cells, but few punctate granules. E) *glo-4 (ok623)* after exposure to 2.5 mM Ala-DanAla dipeptide. F) *glo-4 (ok623)* after exposure to 5 mM Ala-DanAla dipeptide. At this high concentration of dipeptide punctate fluorescent granules began to appear. All animals were exposed for 1 generation on solid plates and assayed as young adults. Images are confocal Z stacks.

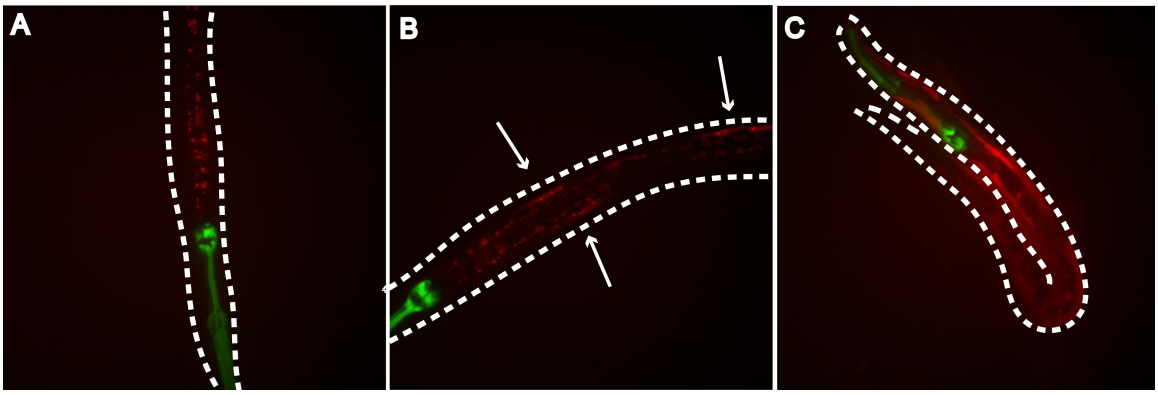


Figure 9: A) LWA1565 with the $tRNA^{Leu}_{CUA}$/DanRS array integrated. Red dots are intestinal autofluorescence. B) mCherry reporter line LWA1560 with the extrachromosomal array *wlEx1565* exposed to 2 mM Ala-DanAla dipeptide. Arrows point to individual red muscle cells; mCherry expression was mosaic. C) LWA1565 exposed to 2 mM Ala-DanAla dipeptide. Note red fluorescence throughout head and body, rather than isolated muscles. All animals were raised in liquid culture for 1 generation at 15 ^{o}C. Dashed lines indicate the outline of the worm. All images are confocal Z stacks and representative of each genotype at larval stage 3-4.

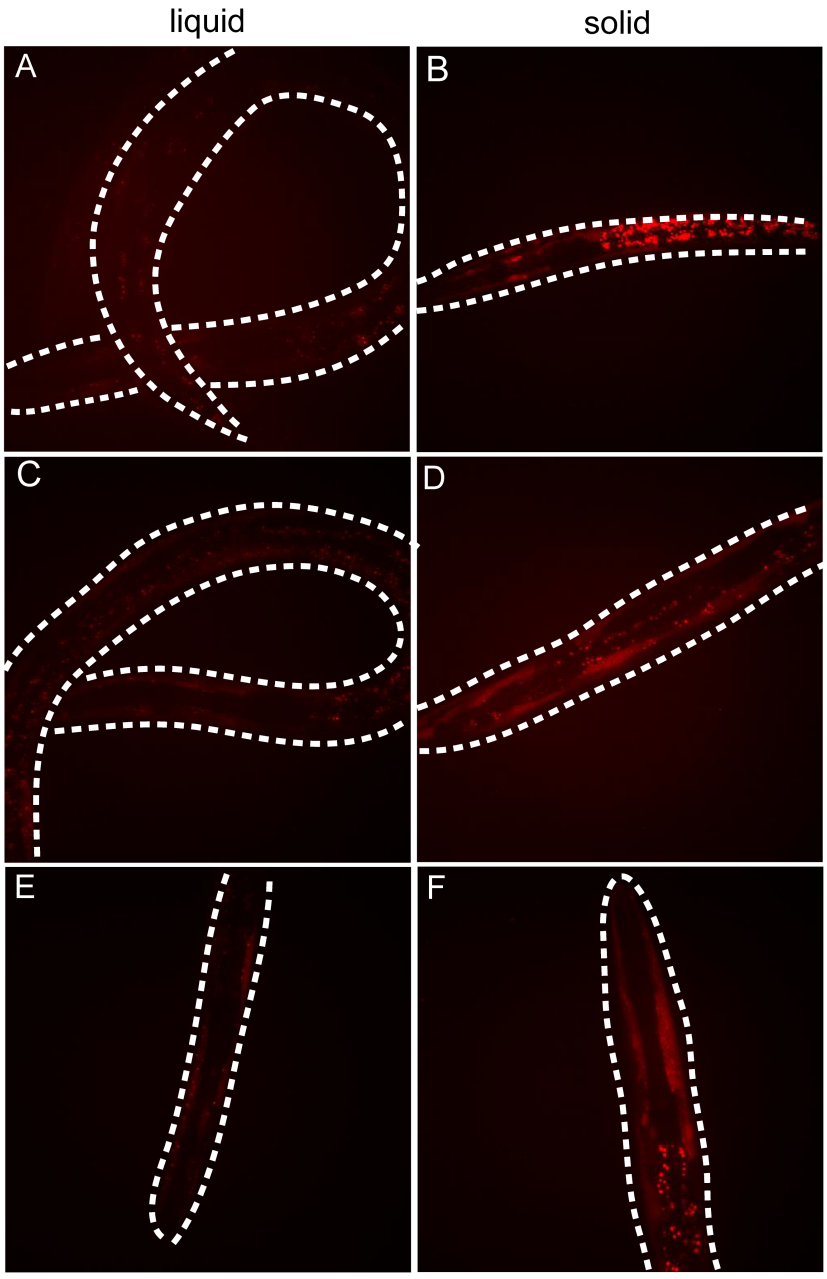


Figure 10: All animals are LWA1565. A-B) 0.5 mM Ala-DanAla dipeptide. C-D) 1 mM Ala-DanAla dipeptide. E-F) 2.5 mM Ala-DanAla dipeptide. Punctate dots are intestinal autofluorescence. In each case, there was more uniform and bright mCherry expression when Uaa is delivered with solid food, and animals appeared healthier on solid food. All animals were raised for 1 generation at 15 ^{o}C. Dashed lines indicate the outline of the worm. All images are confocal Z stacks and the healthiest animals from each condition were imaged, ranging from larval stage 2-4.

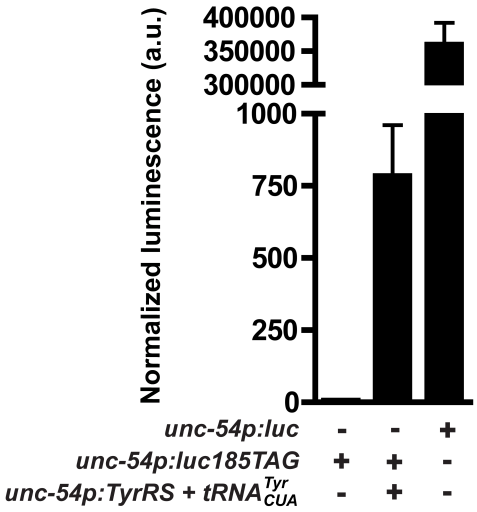


Figure 11: Luciferase activity was readily detectable when the wild type luciferase gene was expressed from a single integrated copy in body wall muscle by the *unc-54* promoter. LWA1850 has a singly integrated luciferase gene with a TAG mutation at permissive site 185 driven by the *unc-54* promoter. No luciferase activity was detected from this reporter strain. Only when crossed to a strain with an integrated *E. coli* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ /TyrRS array (to generate LWA1851) was luciferase detectable. Animals were grown in duplicate. Luminescence was normalized to total protein concentration. Error bars represent s.e.m.; $n \geq 2$.

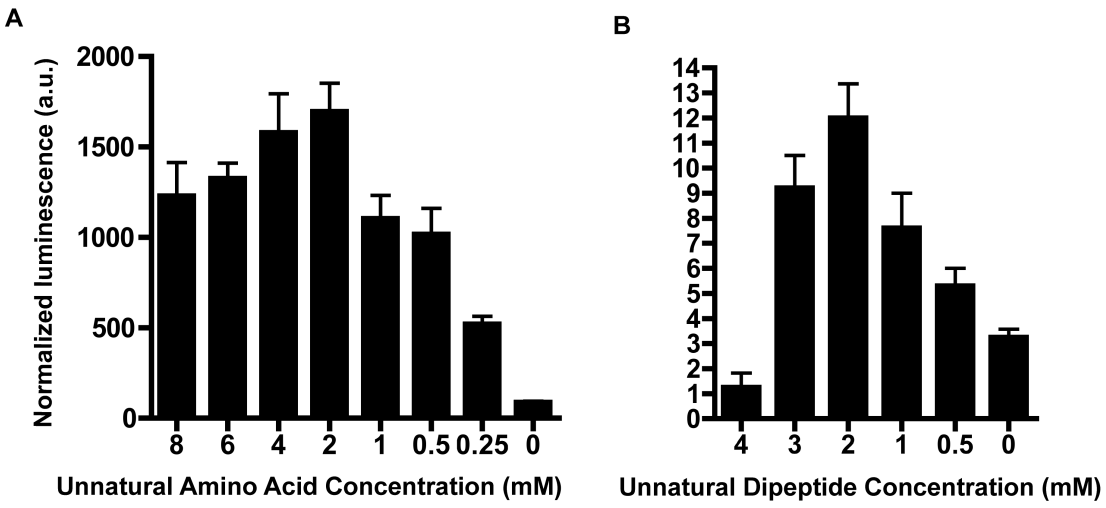


Figure 12: A) Luciferase activities measured from strain LWA1852 grown at different OmeY concentrations. B) Luciferase activities measured from strain LWA1031 grown at different Ala-DanAla concentrations. Both strains were grown on solid plates at 15 ^{o}C for 1 generation in duplicate. Luminescence was normalized to total protein concentration. Error bars represent s.e.m.; n ≥ 2.

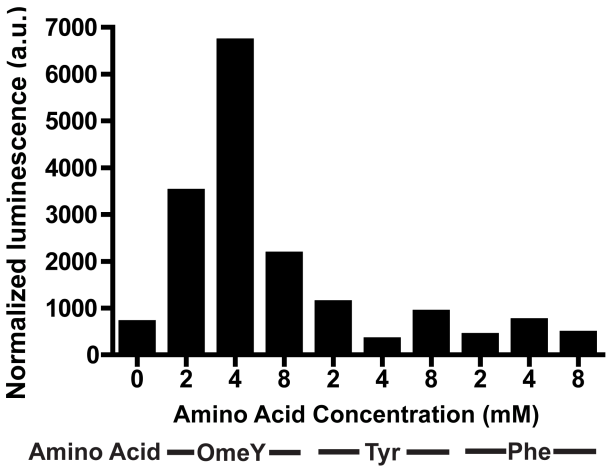


Figure 13: Luciferase activities measured from strain LWA1852 grown at different concentrations of OmeY, Tyr, or Phe for 3 generations. Exposure to natural amino acids does not show dose dependent luciferase activity as exposure to OmeY does, indicating that the OmeRS is specific for OmeY, and is not inserting Tyr or Phe in response to the stop codon in the luciferase reporter. Luminescence was normalized to total protein concentration.

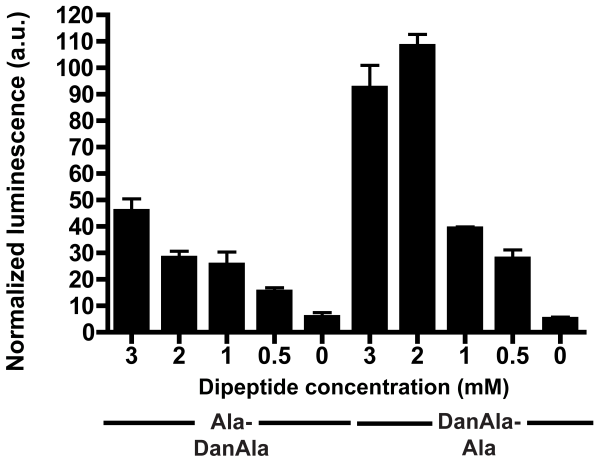


Figure 14: LWA1031 showed differential luciferase activity depending on the Uaa position in the dipeptide. Animals were exposed to the indicated concentrations of dipeptides for 3 generations at 15 °C on solid media. Animals were grown in duplicate. Luminescence was normalized to total protein concentration. Error bars represent s.e.m.; n ≥ 2.

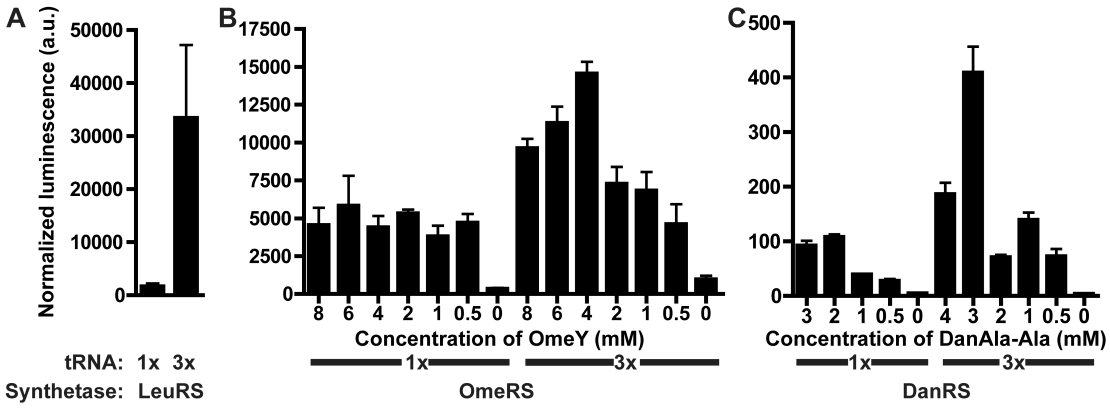


Figure 15: A) Comparison of LWA1008 with LWA1009 indicates that 3x $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ increased Leu incorporation into luciferase. B) Comparison of LWA1852 (1x $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$) with LWA1855 (with *wISi1082[(rpr-1::* $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ *)3x]IV*) indicates that additional $\mathrm{tRNA}^{\mathrm{Tyr}}_{\mathrm{CUA}}$ increased OmeY incorporation into luciferase. C) Comparison of LWA1031 with LWA1856 (with *wISi13 II*) indicates that additional $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ increased DanAla incorporation into luciferase. All animals were grown for 3 generations on solid plates with indicated concentration of Uaa when applicable at 15 °C in duplicate. Luminescence was normalized to total protein concentration. All error bars represent s.e.m., n ≥ 2.

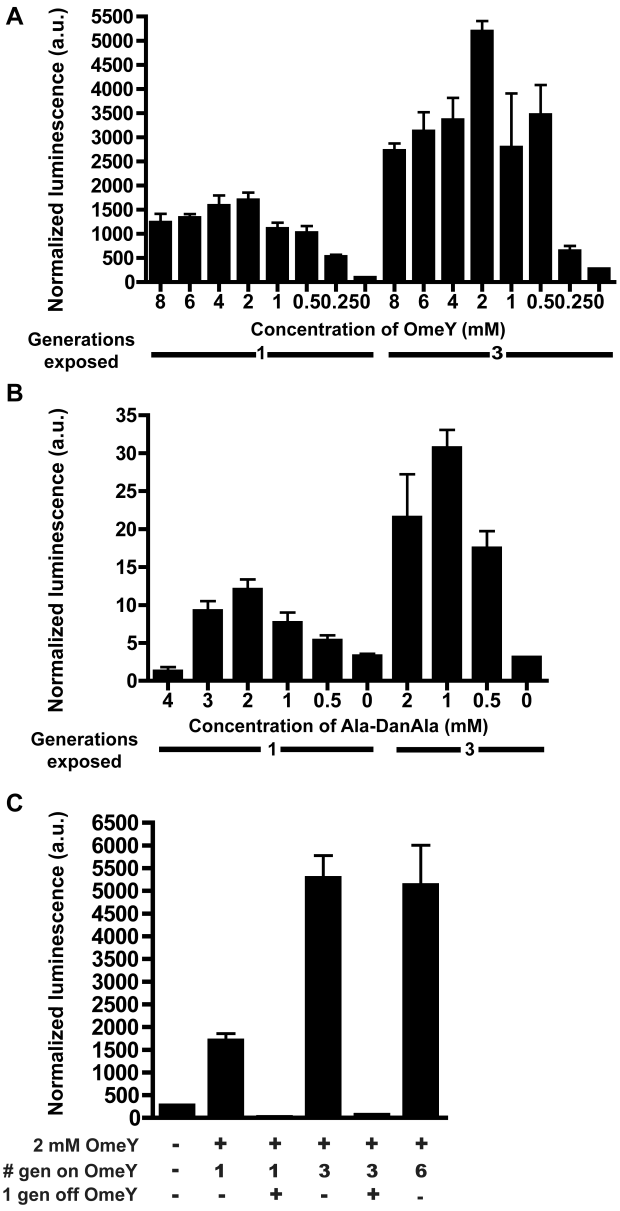


Figure 16: A) LWA1852 animals show an increase in OmeY-dependent luciferase activity over 3 generations compared to 1. B) LWA1031 animals show an increase in Ala-DanAla-dependent luciferase activity over 3 generations compared to 1. For A) and B), animals were grown in duplicate on solid plates of indicated Uaa concentrations at 15 ^{o}C, and assayed after 1 or 3 generations of exposure to Uaa. C) Comparison of multi-generation data and reversibility of Uaa incorporation. LWA1852 animals were grown in duplicate at 15 ^{o}C with 2 mM OmeY on solid plates for indicated generations. Removal of worms from the Uaa for 1 generation dropped the luciferase activity to baseline levels. Luminescence was normalized to total protein concentration. All error bars represent s.e.m.; n ≥ 2.

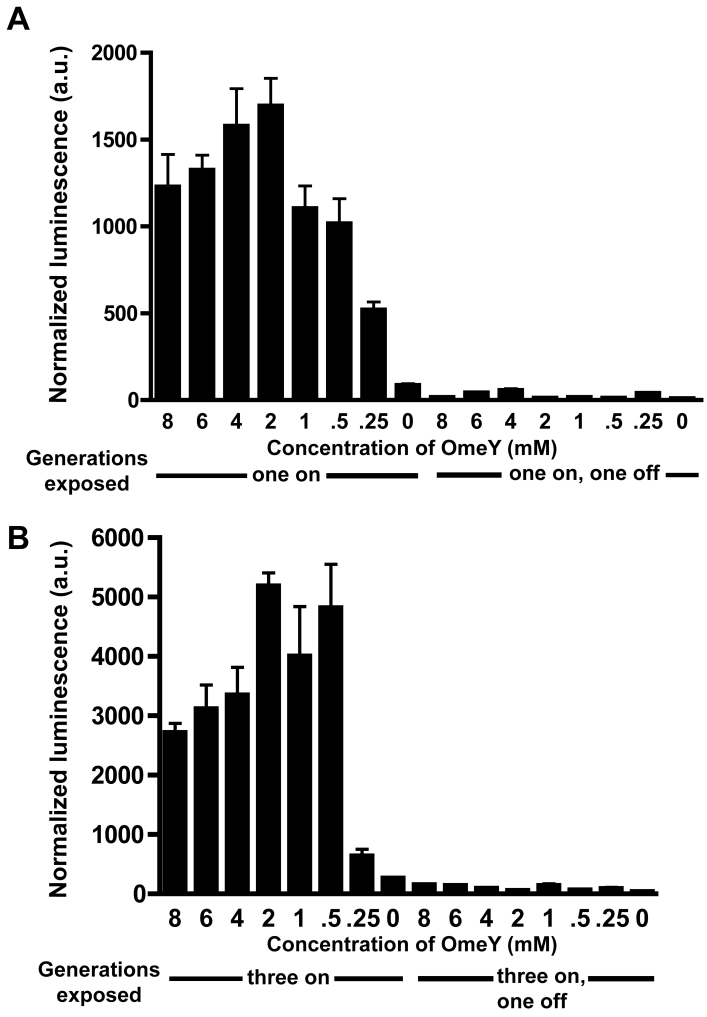


Figure 17: A) LWA1852 was grown at the indicated concentrations of OmeY for 1 generation in duplicate. Animals were removed from each plate to 0 mM plates for 1 generation and assayed. B) LWA1852 was grown at the indicated concentrations of OmeY for 3 generations in duplicate. Animals were removed from each plate to 0 mM plates for 1 generation and assayed. All animals were grown in duplicate at 15 ^{o}C on solid media. After 1 generation in the absence of Uaa, the luciferase activity dropped back to background levels in all cases. Luminescence was normalized to total protein concentration. Error bars represent s.e.m.; $n \geq 2$.

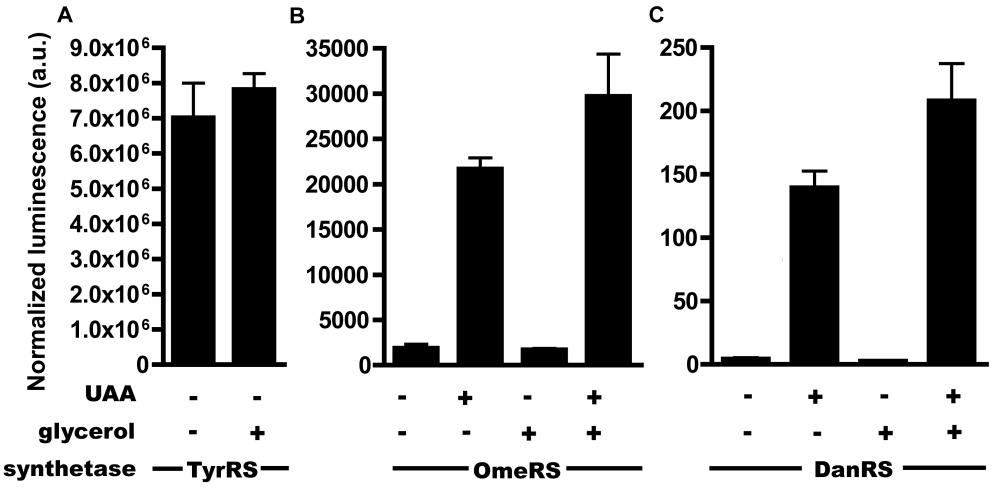


Figure 18: A) LWA1857 animals were raised in the presence and absence of 400 mM glycerol. B) LWA1855 animals were raised in the presence and absence of 2 mM OmeY and 400 mM glycerol. C) LWA1856 animals were raised in the presence and absence of 1 mM DanAla-ala dipeptide and 400 mM glycerol. In each case, there is a modest increase in luciferase activity in the presence of glycerol. All animals were grown for 3 generations on solid plates with indicated concentration of Uaa when applicable at 15 °C in duplicate. Luminescence was normalized to total protein concentration. All error bars represent s.e.m., n ≥ 2.

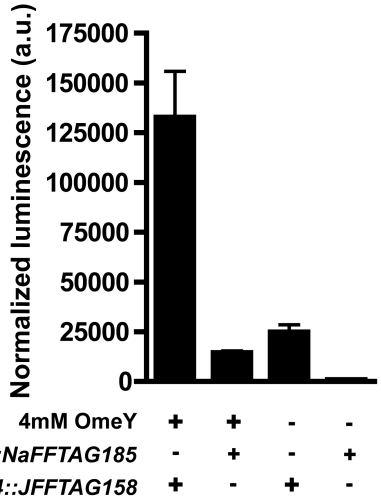


Figure 19: Comparison of North American firefly luciferase reporter (NaFF) to Japanese firefly luciferase reporter (JFF). LWA1855 and LWA717 animals were grown in the presence or absence of 4 mM OmeY as indicated. All animals were grown for 3 generations on solid plates with indicated concentration of Uaa when applicable at 15 °C in duplicate. Luminescence was normalized to total protein concentration. All error bars represent s.e.m., n = 2.

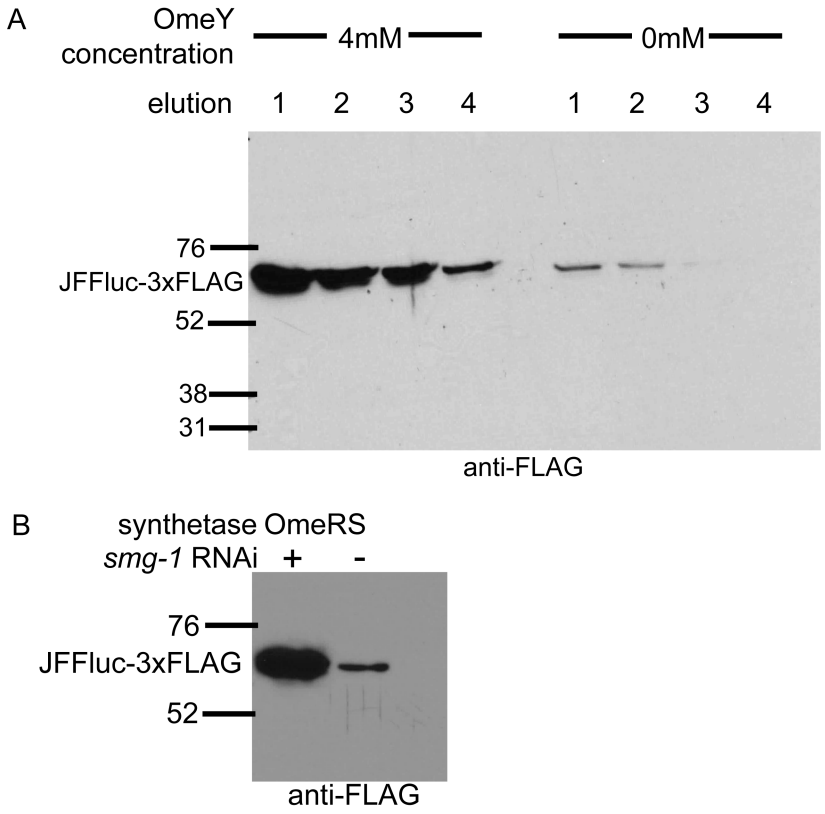


Figure 20: A) LWA717 animals grown on medium containing 4 mM or 0 mM OmeY were lysed and FLAG immunoprecipitation was performed. Four elutions for each sample were collected and equal volumes were loaded into each well. After transferring, the blot was probed with an anti-FLAG antibody to detect purified JFFluc-3xFLAG. At the molecular weight corresponding to the JFFluc-3xFLAG protein (64 kDa) a strong band is visible for all elutions of purified protein from animals grown on 4 mM OmeY, but only a very weak band, corresponding to the first 2-3 elutions, is seen for animals grown in the absence of Uaa. Band densitometry analysis reveals that the band intensity for elutions 1 and 2 for the 0 mM OmeY sample corresponds to 3-5% of the intensity of the elutions for 4 mM OmeY, and elution 3 corresponds to 0.08% of 4 mM OmeY. B) LWA717 animals grown on medium containing 4 mM OmeY show increased JFFluc when treated with RNAi against *smg-1*, compared to RNAi against GFP, to stabilize target mRNA. The mean fold change in band density when comparing *smg-1* treated animals to the GFP control is 5.6, n=2.

Table 1: Genotypes and strain names created for this work.

Strain	Genotype
LWA1560	*wISi151[unc54::mCherryTAG156 cb-unc-119(+)] II.*
LWA1561	*wISi151 II, wIEx1[unc-54::LeuRS_rpr-1::tRNA$_{CUA}^{Leu}$ + pRF4(rol-6)].*
LWA1562	*wISi151 II, wIEx22[unc-54::TyrRS_rpr-1::tRNA$_{CUA}^{Tyr}$ + pRF4(rol-6)].*
LWA1563	*wISi151 II, wIEx13[unc-54::OmeRS_rpr-1::tRNA$_{CUA}^{Tyr}$ + pRF4(rol-6)].*
LWA1564	*wISi151 II, wIEx35[unc-54::DanRS_rpr-1::tRNA$_{CUA}^{Leu}$ + pRF4(rol-6)].*
LWA1565	*wISi151 II, wIls1565[unc-54::DanRS_rpr-1::tRNA$_{CUA}^{Leu}$ + myo-2::GFP]X.*
LWA1850	*wISi1852[unc-54::luciferaseTAG185 cb-unc-119 (+)]I.*
LWA1851	*wISi1852 I, wIls1851[unc-54::TyrRS_rpr-1::tRNA$_{CUA}^{Tyr}$ + myo-2::GFP]X.*
LWA1852	*wISi1852 I, wIls1853[unc-54::OmeRS_rpr-1::tRNA$_{CUA}^{Tyr}$ + myo-2::GFP]X.*
LWA1031	*wISi1852 I, wIls1565[unc-54::DanRS_rpr-1::tRNA$_{CUA}^{Leu}$ + myo-2::GFP]X.*
LWA1008	*wISi1852 I, wISi13[rpr-1::tRNA$_{CUA}^{Leu}$]II, wISi550[unc-54::LeuRS]IV.*
LWA1009	*wISi1852 I, wISi17[(rpr-1::tRNA$_{CUA}^{Leu}$)3x]II, wISi550 IV.*
LWA1855	*wISi1852 I, wISi1082[(rpr-1::tRNA$_{CUA}^{Tyr}$)3x]IV, wIls1853 X.*
LWA1856	*wISi1852 I, wISi17 II, wIls1565 X.*
LWA1857	*wISi1852 I, wISi1082 IV, wIls1851 X.*
LWA1582	*wISi1582[unc-54::JFFluciferaseTAG158 cb-unc-119 (+)]I.*
LWA717	*wISi1582 I, wISi1082 IV, wIls1853 X.*

MATERIALS AND METHODS

C. elegans Strains

EG4322, EG5003, and EG6247 for MosSCI were provided by E. Jorgensen. RB811(*glo-4 (ok623)*) was provided by Y. Jin. N2 (Bristol) was obtained from the *Caenorhabditis* Genetics Center.

Unnatural Amino Acids and Dipeptides

O-methyl-L-tyrosine was purchased from Chem-Impex International. DanAla was synthesized as described previously (Summerer et al., 2006).

Dipeptide Ala-DanAla was synthesized with two methods. In method 1 as shown in Scheme 2A, compound **4** was synthesized from Boc-Dap-OH and dansylchloride via 2 steps according to the known procedure. After deprotection of the Boc group, compound **5** was coupled with Boc-Ala-OH to give compound **6**. Debenzylation and deprotection of the Boc group furnished the Ala-DanAla dipeptide. In method 2 (Scheme 2B), Ala-DanAla was prepared by condensation of Boc-Ala-OSu and DanAla in aqueous bicarbonate (dioxane:water 2:1, pH 8.5). The dipeptide was purified via flash chromatography and deprotected with TFA:CH_2Cl_2 1:1. Solvents were evaporated and the final product re-precipitated from MeOH/Et_2O. The overall yield was about 40%. Ala-DanAla synthesized by both methods behaved the same in *C. elegans* experiments.

Dipeptide DanAla-Ala was synthesized using standard peptide synthesis methods in solution (Scheme 2C). The precursor Boc-Dap(Fmoc)-Ala-OtBu was

prepared by condensation of Boc-Dap(Fmoc)-OH and H-Ala-OH using EDC/HOBt/NMM in DCM, and purified via flash chromatography. The removal of the Fmoc group and the Dansylation were performed in one pot by treatment with 1 equiv. of DBU in DCM followed by the addition of an excess (2.5 equiv.) of dansyl chloride (Coin et al., 2008). After flash chromatographic purification, the protecting groups were removed in a TFA solution (50% in CH_2Cl_2), the volatiles evaporated and the final product re-precipitated from MeOH/Et_2O. The overall yield was about 70%.

Microinjection, Extrachromosomal Array Formation and Integration

Microinjection of plasmid DNA was performed as previously described (Kimble et al., 1982). Construction of strains LWA1560, LWA1850, LWA1008, LWA1009 and LWA1855, LWA1582 was performed as described (Frokjaer-Jensen et al., 2008). Candidate lines were bred to homozygosity, tested by PCR, and sequenced to confirm insertions.

The *unc-54p::RS* contructs were injected as follows: *unc-54p::LeuRS* 15 ng/μL, *unc-54p::DanRS* 20 ng/μL, *unc-54p:*:TyrRS 20 ng/μL, *unc-54p::OmeRS* 15 ng/μL. The injection mix also contained 5 ng/μL *myo-2:GFP* (pPD118.33 from Fire Lab Vector Kit) and 75 ng/μL pBSK2+ to facilitate extrachromosomal array formation. To integrate extrachromosomal arrays containing orthogonal tRNA/RS genes into chromosome, animals harboring each array from injection above were collected. Using γ-irradiation from a Co-60 source, 50 young adult

hermaphrodites were exposed to 4100 rads on an unseeded plate. After a recovery period, the P_o generation animals were picked to single plates. Isolation of F_1 and F_2 animals with high *myo-2::GFP* expression levels led to the identification of animals propagating progeny with 100% co-injection marker expression. The integrated arrays were outcrossed four times to remove any unwanted mutations and compared to the original extrachromosomal arrays in terms of expression pattern and suppression efficiency.

Culture and Uaa Feeding

General culture and manipulation were performed as previously described (Brenner, 1974). Unless indicated, strains were grown at 15 ^{o}C. Animals were grown in liquid culture in the absence of antibiotics (Panowski et al., 2007) to facilitate Uaa uptake. Gravid worms harboring the array were placed in sterile-filtered S basal with cholesterol and incubated in the presence of Uaa or dipeptide with agitation for five days; progeny containing the array were then scored. Small samples were performed in a 24-well tissue culture dish in 0.85 mL total volume, which can support 10-15 gravid adults for one generation. Negative controls used the same procedure but equivalent amounts of NaOH or HCl as in the Uaa were added to the medium to ensure identical pH. For solid culture, NG medium was supplemented with Uaa (also IPTG and carbenicilin for RNAi experiments) immediately before pouring. Plates were seeded with OP50 (or

HT115 for RNAi experiments) according to standard techniques (Stiernagle, 2006).

Luciferase Assays

Animals were grown on Uaa-containing plates in duplicate for 1 or 3 generations. Plates were incubated at 15 ^{o}C until food was exhausted. Animals were washed off of plates with M9 buffer, spun down, and washed twice with M9 and twice with chilled water. After pelleting the worms, 450 μL 1x CCLR lysis buffer (Promega) was added. Animals were lysed in a Precellys24 with ceramic and glass beads for 3 cycles, 10 sec at 6,500 rpm with a 45 sec pause. Samples were spun at 10,000 rpm for 5 min at 4^{o}C. Lysate (80 μL) was placed in a 96-well Nunclon Surface dish. After 100 μL of luciferase assay reagent (Promega) was added to each well, samples were read on a Tecan plate reader using 20 sec integration time. All samples were assayed in duplicate. Luminescence reading was normalized to protein concentration as determined by absorption at 280 nm.

Immunoprecipitation and Western Blotting

LWA717 animals were grown on solid medium until food was exhausted and collected with M9 buffer. Equal numbers of animals were used for each experiment. Animals were flash frozen, then ground into a fine powder with a mortar and pestle on dry ice and resuspended in 400 μl cell lysis buffer (Cell Signaling) supplemented with protease inhibitors (Roche, Sigma). The lysate was

clarified by centrifugation at 16,000*g* for 30 minutes at 4 ^{o}C, and pre-incubated with protein A/G PLUS agarose beads (Santa Cruz). FLAG tagged protein was immunoprecipitated with EZview Red anti-FLAG M2 affinity gel (Sigma) for 2h at 4 ^{o}C, then beads were washed extensively. Samples were eluted with 1 mg/ml 3xFLAG peptide (Sigma) and run on an 8% SDS-PAGE gel. After transfer to a PVDF membrane, the sample was blocked overnight in 5% milk in TBS-T, then incubated with anti-FLAG M2-peroxidase (Sigma), washed for 30 minutes and visualized with SuperSignal West Chemiluminescent Substrate (Thermo Scientific).

RNAi feeding

RNAi feeding was performed as described (Ahringer, 2006) on plates supplemented with IPTG, carbenicilin, and 4mM OmeY, allowing animals to grow as for Uaa feeding experiments until food is exhausted at 15 ^{o}C, equal numbers of animals were used for each experiment. *smg-1* RNAi plasmid is from the Ahringer library (Fraser et al., 2000; Kamath et al., 2003). GFP control plasmid was made by placing the entire coding sequence in vector L4440 (Fire Lab Vector Kit). RNAi plasmids were transformed into *E. coli* strain HT115 for RNAi feeding.

Band densitometry

Quantification of Western blot bands was performed using the gel analyzer tool in the Image J software (NIH).

Microscopy

All fluorescent images are maximum Z-stack projections. Imaging of live worms was performed on 2% agarose pads and 10 mM NaN_3 as an anesthetic using an Olympus IX81 equipped with a spinning disk and a Hamamatsu EM-CCD (Model C9100-02) camera. The light source was a xenon lamp from Sutter Instruments. For DanAla, excitation filter was 360/60 nm and emission filter was 535/40 nm; For mCherry, excitation filter was 580/20 nm and emission filter was 675/130 nm.

Plasmid Construction

Suppression plasmids (*unc-54p::RS*) were constructed using vector pPD30.38 from the Fire lab, which was modified by inserting *Nco* I and *BamH* I sites between the *BamH* I and *Spe* I sites, retaining the *Spe* I site and abolishing the original *BamH* I as a *Bgl* II/*BamH* I hybrid. The *unc-54p::RS* plasmids were produced by inserting each RS gene (Wang and Wang, 2008) in the *Nhe* I/*Sac* I sites between *unc-54* promoter and *unc-54* 3'UTR. The tRNA promoter candidates (500 bp each) were amplified from *C. elegans* genomic DNA and inserted between the *Nco* I and *BamH* I sites, taking care to retain the spacing of the +1 nucleotide and the TATA box, if present. The amber suppressor $tRNA_{CUA}$

without –CCA and immediately followed by the *C. elegans* $tRNA^{Lys}$ 3' flanking sequence (Tranquilla et al., 1982) were inserted between the *BamH* I and *Spe* I sites.

The plasmids for MosSCI integration of *unc-54p*::*mCherryTAG156* were generated by placing the coding sequence of mCherry or mCherryTAG156 in place of the GFP coding sequence in vector pPD95.77, using the *EcoR* I/*Kpn* I sites. The TAG mutation at codon 156 was created using overlapping PCR. The *unc-54* promoter was inserted using *Hind* III/*Kpn* I sites and PCR cloned from vector pPD30.38. This entire fragment was PCR amplified and cloned into the MosSCI vector pCFJ151 (from E. Jorgensen) using *Xho* I and *Spe* I sites.

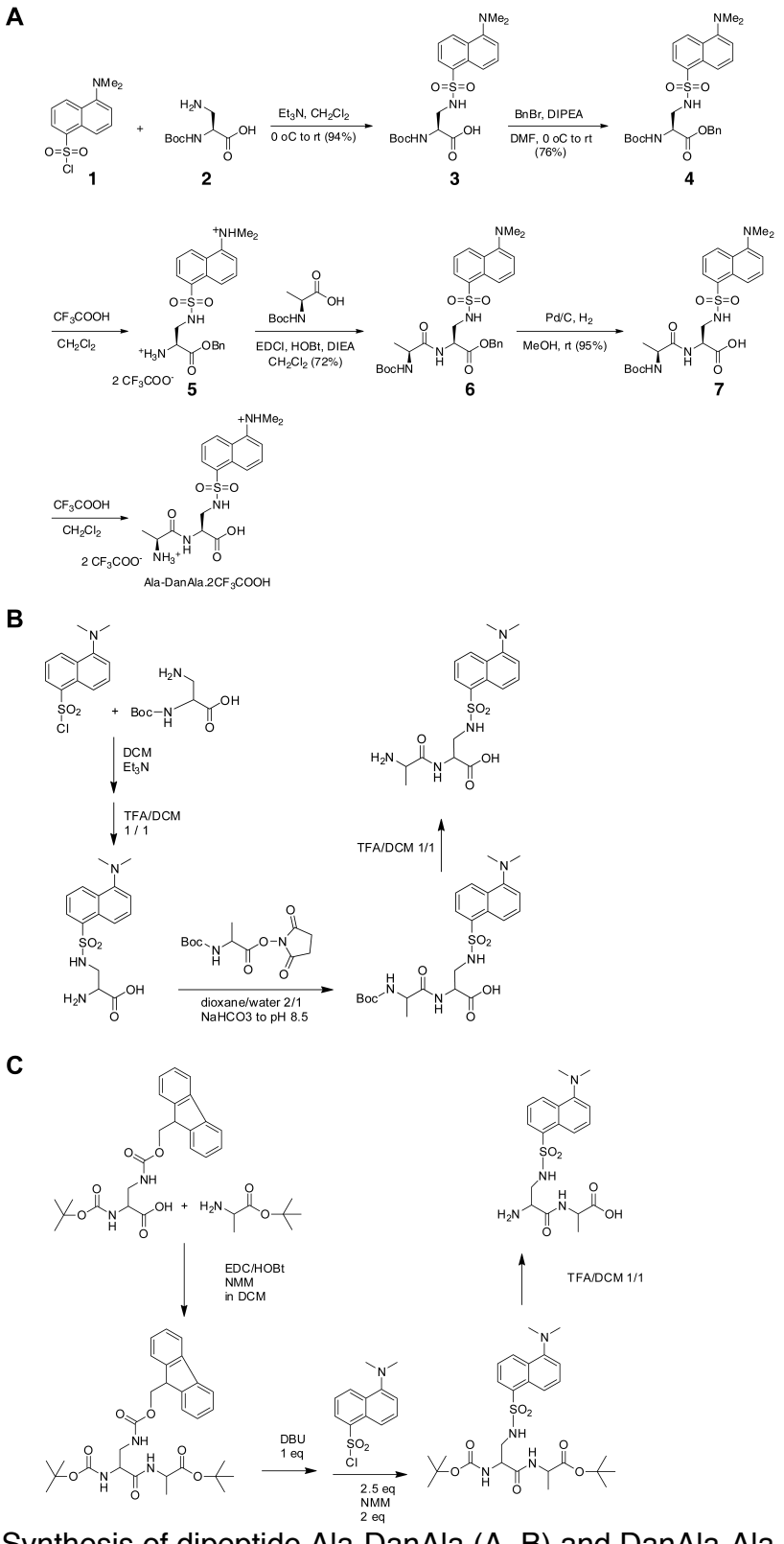
The plasmids for MosSCI integration of *unc-54p::luciferaseTAG185* were generated by PCR amplifying the *unc-54* promoter from *unc-54p*::*mCherryTAG156* using primers with *att*B4 and *att*B1r flanking sites and PCR amplifying the *Photinus pyralis* luciferase gene from pECL1 (Lagido et al., 2001) using primers with *att*B1 and *att*B2 sites. These PCR products were recombined using BP Clonase II into the Gateway vectors pDONR P4-P1r and pDONR221, respectively (Invitrogen). These two plasmids, as well as a plasmid containing the *unc-54* 3' UTR flanked by *att*R2 and *att*L3 sequences (from E. Jorgensen) were recombined with LR Clonase II Plus (Invitrogen) with a MosSCI vector containing homologous regions surrounding the *ttTi*4348 locus on chromosome I in *C. elegans* (from E. Jorgensen).

The plasmids for integrating 3 copies of *rpr-1p::* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ were made using the Gateway system. Identical copies of the gene cassette were PCR amplified for recombination into each of the 3 Gateway vectors and recombined into MosSCI vector pCFJ150 (from E. Jorgensen). The single copy was cut from *unc-54p::*RS using *Nco* I, blunted using Klenow extension (NEB), and then cut with *Spe* I. The *rpr-1p::* $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{CUA}}$ fragment was cloned into pCFJ151, which was cut with *Xho* I, blunted, and then cut with *Spe* I. The *unc-54p::*LeuRS gene cassette was PCR amplified with primers annealing to the *unc-54* promoter and containing an *Xho* I site (5') and the *unc-54* 3'UTR containing an *Spe* I site (3'). This fragment was digested and ligated into pCFJ178 (from E. Jorgensen) for MosSCI integration into chromosome IV.

The *Luciola cruciate* (Japanese) firefly luciferase plasmid (*unc-54p::JFFluciferaseTAG158)* for MosSCI integration was amplified from pJFF12 (from J. Noel) and cloned into pIRES-hrGFP II (Agilent Technologies) using *BamH* I and *Not* I, in frame with the 3xFLAG tag present on the vector. The TAG mutation was made with overlapping PCR and cloned into the same vector. The full gene with the 3xFLAG was amplified and recombined into pDONR221 with BP Clonase II (Invitrogen). This entry plasmid was recombined with the Gateway vector above containing the *unc-54* promoter and an entry vector containing the *tbb-2* 3'UTR followed the *gpd-2* operon linker and a histone H2B:GFP fusion (pCFJ236, from E. Jorgensen) to generate a MosSCI vector targeting Chromosome I, to generate *p4348-unc-54p::JFFluciferase-TAG158::H2B::GFP*.

All fragments were PCR amplified with phusion polymerase (NEB) and cloned with primers containing the restriction sites, and all constructs were sequenced before injection. All Gateway reactions are done according to the manufacturer's instructions (Invitrogen).

A list of primer sequences can be found in Table 2.



Scheme 2: Synthesis of dipeptide Ala-DanAla (A, B) and DanAla-Ala (C).

Table 2: Primers used in this work.

Primer	Sequence 5' → 3'
LeuRSfo	AGCTAGCGGTAGAAAAAATGCAAGAGCAATACCGCC
LeuRSrev	AGAGCTCTTAGCCAACGACCAGATTGAGG
TyrRSfo	AGCTAGCGGTAGAAAAAATGGCAAGCAGTAACTTGATTAAACAATT GCAAG
TyrRSrev	AGAGCTCTTATTTCCAGCAAATCAGACAGTAATTCTTTTTACCGC
tRNAYfo	CAGGATCCCCGGTGGGGTTCCCGAGCGGC
tRNAYrev	TGACTAGTATGAAAAATGCTGTAAATTCAAAAAGATATGGTGGGGG AAGGATTCGAACC
tRNALfo	CCACCATCCGGGCGGGGATCCGTGGTCTCATACAG
tRNALrev	TAACTAGTATGAAAAATGCTGTAAATTCAAAAAGATATACCCGGAG CGGGACTTGAACC
cen4fo	TACCATGGCGACACGATCTTATTTTGATAAAGAACTTGG
cen4rev	AGGATCCGTGCAATAGGAGGCGTTTATATAGGG
cen38fo	TACCATGGCCAATTCTCCATTATGAGCAATTGCTACAGTACC
cen38rev	ATGGATCCATTGAACTGGCGTCTCGTGGGTCAC
rpr1fo	TACCATGGGGCTCAGCCTCAACCAATTTTTAGTG
rpr1rev	AGGATCCACGCGCGCCGCGTCGTT
TAG156fo	GAGCGGATGTAGCCCGAGGACG
TAG156rev	CGTCCTCGGGCTACATCCGCTC
mCherryfo	TAGGTACCGGTAGAAAAAATGGTGAGCAAGGGCGAGGAGG
mCherryrev	AGAATTCTTACTTGTACAGCTCGTCCATGCCG
unc54pfo	CCAAGCTTGTCTTCTTCTAAATTCCC
unc54prev	ATGGTACCCAAGGGTCCTCCTGAAAATGTTC
unc54Xhofo	AACTCGAGCCCATAAAATCCCGAAACTCCTTCCCTC
uncChrev	CGTCATCACCGAAACGCGCG
GWunc54fo	GGGGACAACTTTGTATAGAAAAGTTGCCCATAAAATCCCGAAACT CCTTCC
GWunc54rev	GGGGACTGCTTTTTTGTACAAACTTGCAAGGGTCCTCCTGAAAAT GTTC

Table 2: Primers used in this work, Continued.

Primer	Sequence 5' → 3'
GWlucfo	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGTAGAAAAAATGGAAGACGCC
GWlucrev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTACAATTTGGACTTTCCGCCC
lucTAG185fo	CGATTTTGTACCAGAGTAGTTTGATCGTGACAAAAC
lucTAG185rev	GTTTTGTCACGATCAAACTACTCTGGTACAAAATCG
B4RNAsePfo	GGGGACAACTTTGTATAGAAAAGTTGGGCTCAGCCTCAACCAATTTTTAGTG
B1rLtRNArev	GGGGACTGCTTTTTTGTACAAACTTGATGAAAAATGCTGTAAATTCAAAAAGATATACCCGG
B2RNAsePfo	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGCTCAGCCTCAACCAATTTTTAGTG
B1LtRNArev	GGGGACCACTTTGTACAAGAAAGCTGGGTATGAAAAATGCTGTAAATTCAAAAAGATATACCCGG
B2rRNAsePfo	GGGGACAGCTTTCTTGTACAAAGTGGGGCTCAGCCTCAACCAATTTTTAGTG
B3LtRNArev	GGGGACAACTTTGTATAATAAAGTTGATGAAAAATGCTGTAAATTCAAAAAGATATACCCGG
uncRSrev	GCCTAATTTCTTTGCTTATTTTTTACTAGTTTTCCTTCC
B1rYtRNArev	GGGGACTGCTTTTTTGTACAAACTTGATGAAAAATGCTGTAAATTCAAAAAGATATGGTGGG
B2YtRNArev	GGGGACCACTTTGTACAAGAAAGCTGGGTATGAAAAATGCTGTAAATTCAAAAAGATATGGTGGG
B3YtRNArev	GGGGACAACTTTGTATAATAAAGTTGATGAAAAATGCTGTAAATTCAAAAAGATATGGTGGG
JFFlucfo	AAGGATCCATGGAAAACATGGAAAATGATGAAAACATTGTGG

Table 2: Primers used in this work, Continued.

Primer	Sequence 5' → 3'
JFFlucrev	AAGCGGCCGCCAGAGGAGCCCATTTTGGCCACCGGTTTTTTCAGAA
JFFTAG158fo	CCTGGATAGCAAATAGGATTATCGTGGTTATCAG
JFFTAG158rev	CTGATAACCACGATAATCCTATTTGCTATCCAGG
GWJFFlucfo	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGTAGAAAAAAACCATGGAAAACATGGAAAATGATGAAAACATTGTGG
GWFLAGrev	GGGGACCACTTTGTACAAGAAAGCTGGGTCGTCGAGGAATTGCTATTATTTGTCGTCAT

REFERENCES

WORKS CITED

Ahringer, J. (2006). Reverse genetics. In WormBook, T.*C.e.*R. Community, ed. (WormBook).

Anderson, J.C., Magliery, T.J., and Schultz, P.G. (2002). Exploring the limits of codon and anticodon size. Chem Biol *9*, 237-244.

Anderson, J.C., and Schultz, P.G. (2003). Adaptation of an Orthogonal Archaeal Leucyl-tRNA and Synthetase Pair for Four-base, Amber, and Opal Suppression. Biochemistry *42*, 9598-9608.

Anderson, J.C., Wu, N., Santoro, S.W., Lakshman, V., King, D.S., and Schultz, P.G. (2004). An expanded genetic code with a functional quadruplet codon. Proc Natl Acad Sci U S A *101*, 7566-7571.

Antoshechkin, I., and Sternberg, P.W. (2007). The versatile worm: genetic and genomic resources for Caenorhabditis elegans research. Nat Rev Genet *8*, 518-532.

Atkins, J.F., Weiss, R.B., Thompson, S., and Gesteland, R.F. (1991). Towards a genetic dissection of the basis of triplet decoding, and its natural subversion: programmed reading frame shifts and hops. Annu Rev Genet *25*, 201-228.

Bain, J.D., Glabe, C.G., Dix, T.A., Chamberlin, A.R., and Diala, E.S. (1989). Biosynthetic Site-Specific Incorporation of a Non-Natural Amino Acid Into a Polypeptide. J Am Chem Soc *111*, 8013-8014.

Benzer, S., and Champe, S.P. (1962). A change from nonsense to sense in the genetic code. Proc Natl Acad Sci U S A *48*, 1114-1121.

Bessereau, J.L., Wright, A., Williams, D.C., Schuske, K., Davis, M.W., and Jorgensen, E.M. (2001). Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line. Nature *413*, 70-74.

Biegel, A., Gebauer, S., Brandsch, M., Neubert, K., and Thondorf, I. (2006). Structural requirements for the substrates of the H+/peptide cotransporter PEPT2 determined by three-dimensional quantitative structure-activity relationship analysis. J Med Chem *49*, 4286-4296.

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.

Bolten, S., Powell-Abel, P., Fischhoff, D., and Waterston, R. (1984). The sup-7(st5) X gene of Caenorhabditis elegans encodes a tRNATrpUAG amber suppressor. Proc Natl Acad Sci USA *81*, 6784-6788.

Bose, M., Groff, D., Xie, J., Brustad, E., and Schultz, P.G. (2006). The incorporation of a photoisomerizable amino acid into proteins in E. coli. J Am Chem Soc *128*, 388-389.

Bossi, L., and Roth, J.R. (1980). The influence of codon context on genetic code translation. Nature *286*, 123-127.

Bossi, L., and Roth, J.R. (1981). Four-base codons ACCA, ACCU and ACCC are recognized by frameshift suppressor sufJ. Cell *25*, 489-496.

Brenner, S. (1974). The genetics of Caenorhabditis elegans. Genetics *77*, 71-94.

Cao, D., and Parker, R. (2003). Computational modeling and experimental analysis of nonsense-mediated decay in yeast. Cell *113*, 533-545.

Carne, A.F. (1994). Chemical modification of proteins. Methods Mol Biol *32*, 311-320.

Cellitti, S.E., Jones, D.H., Lagpacan, L., Hao, X., Zhang, Q., Hu, H., Brittain, S.M., Brinker, A., Caldwell, J., Bursulaya, B., Spraggon, G., Brock, A., Ryu, Y., Uno, T., Schultz, P. G., and Geierstanger, B. H. (2008). In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. J Am Chem Soc *130*, 9268-9281.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-805.

Chen, S., Schultz, P.G., and Brock, A. (2007). An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in Saccharomyces cerevisiae. J Mol Biol *371*, 112-122.

Chin, J., and Schultz, P. (2002). In vivo photocrosslinking with unnatural amino Acid mutagenesis. ChemBioChem *3*, 1135-1137.

Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z., and Schultz, P.G. (2003). An expanded eukaryotic genetic code. Science *301*, 964-967.

Chin, J.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002a). Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. Proc Natl Acad Sci U S A *99*, 11020-11024.

Chin, J.W., Santoro, S.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002b). Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. J Am Chem Soc *124*, 9026-9027.

Coin, I., Beerbaum, M., Schmieder, P., Bienert, M., and Beyermann, M. (2008). Solid-phase synthesis of a cyclodepsipeptide: cotransin. Org Lett *10*, 3857-3860.

Coin, I., Perrin, M.H., Vale, W.W., and Wang, L. (2011). Photo-Cross-Linkers Incorporated into G-Protein-Coupled Receptors in Mammalian Cells: A Ligand Comparison. Angew Chem Int Ed Engl.

Coleman, J.R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., and Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. Science *320*, 1784-1787.

Curran, J.F., and Yarus, M. (1987). Reading frame selection and transfer RNA anticodon loop stacking. Science *238*, 1545-1550.

Daniel, H., Spanier, B., Kottra, G., and Weitz, D. (2006). From bacteria to man: archaic proton-dependent peptide transporters at work. Physiology *21*, 93-102.

Deiters, A., Groff, D., Ryu, Y., Xie, J., and Schultz, P.G. (2006). A genetically encoded photocaged tyrosine. Angew Chem Int Ed Engl *45*, 2728-2731.

Deng, W., Zhu, X., Skogerbø, G., Zhao, Y., Fu, Z., Wang, Y., He, H., Cai, L., Sun, H., Liu, C., Li, B., Bai, B., Wang, J., Jia, D., Sun, S., He, H., Cui, Y., Wang, Y., Bu, D., and Chen, R. (2005). Organization of the Caenorhabditis elegans small non-coding transcriptome: genomic features, biogenesis, and expression. Genome Res *16*, 20-29.

Deocaris, C.C., Takano, S., Priyandoko, D., Kaul, Z., Yaguchi, T., Kraft, D.C., Yamasaki, K., Kaul, S.C., and Wadhwa, R. (2008). Glycerol stimulates innate chaperoning, proteasomal and stress-resistance functions: implications for geronto-manipulation. Biogerontology *9*, 269-282.

Drabkin, H.J., Park, H.J., and Rajbhandary, U.L. (1996). Amber Suppression in Mammalian Cells Dependent Upon Expression of an Escherichia Coli Aminoacyl-Trna Synthetase Gene. Mol Cell Biol *16*, 907-913.

Edwards, H., and Schimmel, P. (1990). A bacterial amber suppressor in Saccharomyces cerevisiae is selectively recognized by a bacterial aminoacyl-tRNA synthetase. Mol Cell Biol *10*, 1633-1641.

Edwards, H., Trezeguet, V., and Schimmel, P. (1991). An Escherichia coli tyrosine transfer RNA is a leucine-specific transfer RNA in the yeast Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *88*, 1153-1156.

Fei, Y., Fujita, T., Lapp, D., Ganapathy, V., and Leibach, F. (1998). Two oligopeptide transporters from Caenorhabditis elegans: molecular cloning and functional expression. Biochem J *332 (Pt 2)*, 565-572.

Fraser, A.G., Kamath, R.S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. (2000). Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. Nature *408*, 325-330.

Frokjaer-Jensen, C., Davis, M.W., Hopkins, C.E., Newman, B.J., Thummel, J.M., Olesen, S.P., Grunnet, M., and Jorgensen, E.M. (2008). Single-copy insertion of transgenes in Caenorhabditis elegans. Nat Genet *40*, 1375-1383.

Galli, G., Hofstetter, H., and Birnstiel, M.L. (1981). Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature *294*, 626-631.

Greiss, S., and Chin, J.W. (2011). Expanding the genetic code of an animal. J Am Chem Soc *133*, 14196-14199.

Grunewald, J., Tsao, M.L., Perera, R., Dong, L., Niessen, F., Wen, B.G., Kubitz, D.M., Smider, V.V., Ruf, W., Nasoff, M., Lerner, R. A., and Schultz, P. G. (2008). Immunochemical termination of self-tolerance. Proc Natl Acad Sci U S A *105*, 11276-11280.

Hedgecock, E.M., Culotti, J.G., Thomson, J.N., and Perkins, L.A. (1985). Axonal guidance mutants of Caenorhabditis elegans identified by filling sensory neurons with fluorescein dyes. Dev Biol *111*, 158-170.

Hermann, G.J., Schroeder, L.K., Hieb, C.A., Kershner, A.M., Rabbitts, B.M., Fonarev, P., Grant, B.D., and Priess, J.R. (2005). Genetic analysis of lysosomal trafficking in Caenorhabditis elegans. Mol Biol Cell *16*, 3273-3288.

Hino, N., Okazaki, Y., Kobayashi, T., Hayashi, A., Sakamoto, K., and Yokoyama, S. (2005). Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. Nat Methods *2*, 201-206.

Hodgkin, J. (1985). Novel Nematode Amber Suppressors. Genetics *111*, 287-310.

Hohsaka, T., Ashizuka, Y., Murakami, H., and Sisido, M. (2001a). Five-base codons for incorporation of nonnatural amino acids into proteins. Nucleic Acids Res *29*, 3646-3651.

Hohsaka, T., Ashizuka, Y., Taira, H., Murakami, H., and Sisido, M. (2001b). Incorporation of nonnatural amino acids into proteins by using various four-base codons in an Escherichia coli in vitro translation system. Biochemistry *40*, 11060-11064.

Hortin, G., and Boime, I. (1983). Applications of amino acid analogs for studying co- and posttranslational modifications of proteins. Methods Enzymol *96*, 777-784.

Huang, L.Y., Umanah, G., Hauser, M., Son, C., Arshava, B., Naider, F., and Becker, J.M. (2008). Unnatural amino acid replacement in a yeast G protein-coupled receptor in its native environment. Biochemistry *47*, 5638-5648.

Hutter, H. (2006). Fluorescent reporter methods. Methods Mol Biol *351*, 155-173.

I**bba, M., and Soll, D.** (2000). Aminoacyl-tRNA synthesis. Annu Rev Biochem *69*, 617-650.

Jackson, J.C., Hammill, J.T., and Mehl, R.A. (2007). Site-specific incorporation of a (19)F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. J Am Chem Soc *129*, 1160-1166.

Jin, Y. (2005). Synaptogenesis. WormBook, 1-11.

Kamath, R., Fraser, A., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Weichmann, D. P., Zipperlen, P., and Ahringer, J. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature *421*, 231-237.

Kent, S.B. (1988). Chemical synthesis of peptides and proteins. Annu Rev Biochem *57*, 957-989.

Kiga, D., Sakamoto, K., Kodama, K., Kigawa, T., Matsuda, T., Yabuki, T., Shirouzu, M., Harada, Y., Nakayama, H., Takio, K., Hasegawa, Y., Endo, Y., Hirao, I., and Yokowama, S. (2002). An engineered Escherichia coli tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. Proc Natl Acad Sci U S A *99*, 9715-9720.

Kimble, J., Hodgkin, J., Smith, T., and Smith, J. (1982). Suppression of an amber mutation by microinjection of suppressor tRNA in C. elegans. Nature *299*, 456-458.

Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V., and Gottesman, M.M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science *315*, 525-528.

Kohrer, C., Xie, L., Kellerer, S., Varshney, U., and RajBhandary, U.L. (2001). Import of amber and ochre suppressor tRNAs into mammalian cells: a general approach to site-specific insertion of amino acid analogues into proteins. Proc Natl Acad Sci U S A *98*, 14310-14315.

Kondo, K., Hodgkin, J., and Waterston, R. (1988). Differential expression of five tRNA(UAGTrp) amber suppressors in Caenorhabditis elegans. Mol Cell Biol *8*, 3627-3635.

Kondo, K., Makovec, B., Waterston, R.H., and Hodgkin, J. (1990). Genetic and molecular analysis of eight tRNA(Trp) amber suppressors in Caenorhabditis elegans. J Mol Biol *215*, 7-19.

Kowal, A.K., Kohrer, C., and RajBhandary, U.L. (2001). Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. Proc Natl Acad Sci U S A *98*, 2268-2273.

Kwok, T., Ricker, N., Fraser, R., Chan, A., Burns, A., Stanley, E., McCourt, P., Cutler, S., and Roy, P. (2006). A small-molecule screen in C. elegans yields a new calcium channel antagonist. Nature *441*, 91-95.

Kwok, Y., and Wong, J.T. (1980). Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes. Can J Biochem *58*, 213-218.

Lacey, V.K., Parrish, A.R., Han, S., Shen, Z., Briggs, S.P., Ma, Y., and Wang, L. (2011). A fluorescent reporter of the phosphorylation status of the substrate protein STAT3. Angew Chem Int Ed Engl *50*, 8692-8696.

Lagido, C., Pettitt, J., Porter, A.J., Paton, G.I., and Glover, L.A. (2001). Development and application of bioluminescent Caenorhabditis elegans as multicellular eukaryotic biosensors. FEBS Lett *493*, 36-39.

Lemke, E.A., Summerer, D., Geierstanger, B.H., Brittain, S.M., and Schultz, P.G. (2007). Control of protein phosphorylation with a genetically encoded photocaged amino acid. Nat Chem Biol *3*, 769-772.

Li, F., Xiong, Y., Wang, J., Cho, H.D., Tomita, K., Weiner, A.M., and Steitz, T.A. (2002). Crystal structures of the Bacillus stearothermophilus CCA-adding enzyme and its complexes with ATP or CTP. Cell *111*, 815-824.

Li, L., Linning, R., Kondo, K., and Honda, B. (1998). Differential expression of individual suppressor tRNA(Trp) gene gene family members in vitro and in vivo in the nematode Caenorhabditis elegans. Mol Cell Biol *18*, 703-709.

Liu, D.R., Magliery, T.J., and Schultz, P.G. (1997). Characterization of an 'orthogonal' suppressor tRNA derived from E-coli tRNA(2)(Gln). Chem Biol *4*, 685-691.

Liu, W., Brock, A., Chen, S., and Schultz, P.G. (2007). Genetic incorporation of unnatural amino acids into proteins in mammalian cells. Nat Methods *4*, 239-244.

Magliery, T.J., Anderson, J.C., and Schultz, P.G. (2001). Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in Escherichia coli. J Mol Biol *307*, 755-769.

Mehl, R.A., Anderson, J.C., Santoro, S.W., Wang, L., Martin, A.B., King, D.S., Horn, D.M., and Schultz, P.G. (2003). Generation of a bacterium with a 21 amino acid genetic code. J Am Chem Soc *125*, 935-939.

Mello, C., Kramer, J., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. EMBO J *10*, 3959-3970.

Monahan, S.L., Lester, H.A., and Dougherty, D.A. (2003). Site-specific incorporation of unnatural amino acids into receptors expressed in Mammalian cells. Chem Biol *10*, 573-580.

Muir, T.W. (2003). Semisynthesis of proteins by expressed protein ligation. Annu Rev Biochem *72*, 249-289.

Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K., and Yokoyama, S. (2008). Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. Biochem Biophys Res Commun *371*, 818-822.

Myslinski, E., Ame, J.C., Krol, A., and Carbon, P. (2001). An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene. Nucleic Acids Res *29*, 2502-2509.

Neumann, H., Peak-Chew, S.Y., and Chin, J.W. (2008). Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. Nat Chem Biol *4*, 232-234.

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*, 182-188.

O'Connor, M. (2002). Insertions in the anticodon loop of tRNA1Gln(sufG) and tRNA(Lys) promote quadruplet decoding of CAAA. Nucleic Acids Res *30*, 1985-1990.

Panowski, S.H., Wolff, S., Aguilaniu, H., Durieux, J., and Dillin, A. (2007). PHA-4/Foxa mediates diet-restriction-induced longevity of C. elegans. Nature *447*, 550-555.

Pastrnak, M., Magliery, T.J., and Schultz, P.G. (2000). A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code. Helv Chim Acta *83*, 2277-2286.

Pastrnak, M., and Schultz, P.G. (2001). Phage selection for site-specific incorporation of unnatural amino acids into proteins in vivo. Bioorg Med Chem *9*, 2373-2379.

Paule, M.R., and White, R.J. (2000). Survey and summary: transcription by RNA polymerases I and III. Nucleic Acids Res *28*, 1283-1298.

Peltz, S.W., Brown, A.H., and Jacobson, A. (1993). mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. Genes Dev *7*, 1737-1754.

Pilgrim, D., and Bell, J. (1993). Expression of a Drosophila melanogaster amber suppressor tRNA(Ser) in Caenorhabditis elegans. Mol Gen Genet *241*, 26-32.

Praitis, V., Casey, E., Collar, D., and Austin, J. (2001). Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. Genetics *157*, 1217-1226.

Riddle, D.L. (1997). C. elegans II (Plainview, N.Y., Cold Spring Harbor Laboratory Press).

Robert, V.J., Davis, M.W., Jorgensen, E.M., and Bessereau, J.L. (2008). Gene conversion and end-joining-repair double-strand breaks in the Caenorhabditis elegans germline. Genetics *180*, 673-679.

Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, K., Saito, K., Shirouzu, M., Hirao, I., and Yokoyama, S. (2002). Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. Nucleic Acids Res *30*, 4692-4699.

Santoro, S., Wang, L., Herberich, B., King, D., and Schultz, P. (2002). An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. Nat Biotechnol *20*, 1044-1048.

Santoro, S.W., Anderson, J.C., Lakshman, V., and Schultz, P.G. (2003). An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in Escherichia coli. Nucleic Acids Res *31*, 6700-6709.

Schlieker, C., Weibezahn, J., Patzelt, H., Tessarz, P., Strub, C., Zeth, K., Erbse, A., Schneider-Mergener, J., Chin, J.W., Schultz, P.G., Bukau, B., and Mogk, A. (2004). Substrate recognition by the AAA+ chaperone ClpB. Nat Struct Mol Biol *11*, 607-615.

Shen, B., Xiang, Z., Miller, B., Louie, G., Wang, W., Noel, J.P., Gage, F.H., and Wang, L. (2011). Genetically encoding unnatural amino acids in neural stem cells and optically reporting voltage-sensitive domain changes in differentiated neurons. Stem Cells *29*, 1231-1240.

Srinivasan, G., James, C.M., and Krzycki, J.A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science *296*, 1459-1462.

Steer, B., and Schimmel, P. (1999). Major anticodon-binding region missing from an archaebacterial tRNA synthetase. J Biol Chem *274*, 35601-35606.

Stiernagle, T. (2006). Maintenance of C. elegans. WormBook, 1-11.

Stinchcomb, D., Shaw, J., Carr, S., and Hirsh, D. (1985). Extrachromosomal DNA transformation of Caenorhabditis elegans. Mol Cell Biol *5*, 3484-3496.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev Biol *100*, 64-119.

Summerer, D., Chen, S., Wu, N., Deiters, A., Chin, J., and Schultz, P. (2006). A genetically encoded fluorescent amino acid. Proc Natl Acad Sci USA *103*, 9785-9789.

Takimoto, J.K., Dellas, N., Noel, J.P., and Wang, L. (2011). Stereochemical Basis for Engineered Pyrrolysyl-tRNA Synthetase and the Efficient in Vivo Incorporation of Structurally Divergent Non-native Amino Acids. ACS Chem Biol *6*, 733-743.

Takimoto, J.K., Xiang, Z., Kang, J.Y., and Wang, L. (2010). Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation into proteins in mammalian cells. ChemBioChem *11*, 2268-2272.

Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. Nature *395*, 854.

Ting, A.Y., Shin, I., Lucero, C., and Schultz, P.G. (1998). Energetic analysis of an engineered cation-pi interaction in staphylococcal nuclease. J Am Chem Soc *120*, 7135-7136.

Tranquilla, T.A., Cortese, R., Melton, D., and Smith, J.D. (1982). Sequences of four tRNA genes from Caenorhabditis elegans and the expression of C. elegans tRNALeu (anticodon IAG) in Xenopus oocytes. Nucleic Acids Res *10*, 7919-7934.

Veljkovic, E., Bacconi, A., Stetak, A., Hajnal, A., Stasiuk, S., Skelly, P.J., Forster, I., Shoemaker, C.B., and Verrey, F. (2004a). Aromatic amino acid transporter AAT-9 of Caenorhabditis elegans localizes to neurons and muscle cells. J Biol Chem *279*, 49268-49273.

Veljkovic, E., Stasiuk, S., Skelly, P.J., Shoemaker, C.B., and Verrey, F. (2004b). Functional characterization of Caenorhabditis elegans heteromeric amino acid transporters. J Biol Chem *279*, 7655-7662.

Wang, J., Xie, J., and Schultz, P.G. (2006a). A genetically encoded fluorescent amino acid. J Am Chem Soc *128*, 8738-8739.

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

Wang, L., Brock, A., and Schultz, P.G. (2002). Adding L-3-(2-Naphthyl)alanine to the genetic code of E. coli. J Am Chem Soc *124*, 1836-1837.

Wang, L., Magliery, T.J., Liu, D.R., and Schultz, P.G. (2000). A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. J Am Chem Soc *122*, 5010-5011.

Wang, L., and Schultz, P. (2004). Expanding the genetic code. Angew Chem Int Ed Engl *44*, 34-66.

Wang, L., and Schultz, P.G. (2001). A general approach for the generation of orthogonal tRNAs. Chem Biol *8*, 883-890.

Wang, L., and Schultz, P.G. (2002). Expanding the genetic code. Chem Commun, 1-11.

Wang, L., Xie, J., and Schultz, P.G. (2006b). Expanding the Genetic Code. Annu Rev Biophys Biomol Struct *35*, 225-249.

Wang, Q., and Wang, L. (2008). New methods enabling efficient incorporation of unnatural amino acids in yeast. J Am Chem Soc *130*, 6066-6067.

Wang, W., Takimoto, J.K., Louie, G.V., Baiga, T.J., Noel, J.P., Lee, K.F., Slesinger, P.A., and Wang, L. (2007). Genetically encoding unnatural amino acids for cellular and neuronal studies. Nat Neurosci *10*, 1063-1072.

Waterston, R. (1981). A second informational suppressor, SUP-7 X, in Caenorhabditis elegans. Genetics *97*, 307-325.

Waterston, R.H., and Brenner, S. (1978). A suppressor mutation in the nematode acting on specific alleles of many genes. Nature *275*, 715-719.

Wills, N., Gesteland, R.F., Karn, J., Barnett, L., Bolten, S., and Waterston, R.H. (1983). The genes sup-7 X and sup-5 III of C. elegans suppress amber nonsense mutations via altered transfer RNA. Cell *33*, 575-583.

Wu, N., Deiters, A., Cropp, T.A., King, D., and Schultz, P.G. (2004). A genetically encoded photocaged amino acid. J Am Chem Soc *126*, 14306-14307.

Xie, J., Supekova, L., and Schultz, P.G. (2007). A genetically encoded metabolically stable analogue of phosphotyrosine in Escherichia coli. ACS Chem Biol *2*, 474-478.

Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G., and Schultz, P.G. (2004). The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. Nat Biotechnol *22*, 1297-1301.

Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D., and Schultz, P. (2004). Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. Proc Natl Acad Sci USA *101*, 8882-8887.