Title
Engineering Mitochondria Towards A Platform For Unnatural Polymer Synthesis

Author
Rezvani, Ryan

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Engineering Mitochondria Towards A Platform For Unnatural Polymer Synthesis

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Mathematical, Computational, and Systems Biology

by

Ryan Nema Rezvani

Thesis Committee:
Associate Professor Chang Liu, Chair
Assistant Professor Timothy Downing
Assistant Professor Han Li

2021

TABLE OF CONTENTS

	Page
List of Figures	iii
Acknowledgements	v
Abstract	vi
Chapter 1: Introduction to Unnatural Polymer Synthesis Through Expanded Genetic Codes	1
1.1: Overview and Applications of Expanded Genetic Codes	1
1.2: Advancements in Expanded Genetic Codes and Unnatural Polymer Synthesis	4
1.3: Mitochondria as a Center for Orthogonal Translation	7
1.4: References	10
Chapter 2: Design and Implementation of a Mitochondrial Expanded Genetic Code	13
2.1: Aminoacyl-tRNA Synthetase Design and Validation	13
2.2: Application of Mutant aaRSs to a Mitochondrial GFP Reporter System	18
2.3: Transformation of Mitochondria for the Addition of an Affinity-tagged Protein	22
2.4: References	24
Chapter 3: Optimization of the Mitochondrial Expanded Genetic Code Platform	27
3.1: Adopting a Mitochondrial Superfolder GFP Strain	27
3.2: Expression Refinements to the Platform	30
3.3: Click Chemistry Labeling of Unnatural Amino Acids in Mitochondria	35
3.4: References	39
Chapter 4: Biochemical Characterization of the Mitochondrial Expanded Genetic Code Platform	42
4.1: Validating an Affinity Tag on Mitochondrial sfGFP	42
4.2: Developing an Unnatural Peptide Analytical Pipeline with a Bacterial Control	44
4.3: MS Analysis of Purified Mitochondrial Protein	47
4.4: References	52
Chapter 5: Discussion and Future Directions	53
5.1: Further Developing Mitochondrial Unnatural Polymer Synthesis	53
5.2: Pivoting to a Tool for Mitochondrial tRNA Targeting	56
5.3: References	58
Appendix	61

LIST OF FIGURES

		Page
Figure 1.1	Developing Strategies for Genetic Code Expansion	5
Figure 1.2	Flexible *In Vitro* Translation	7
Figure 1.3	Mitochondrial Platform for Genetic Code Expansion	10
Figure 2.1	Alignment of *E. coli* and Mitochondrial Synthetases	14
Figure 2.2	Unnatural Amino Acids Used in this Study	15
Figure 2.3	UAA Sensitivity Plating of MSY1 Variants	16
Figure 2.4	UAA Sensitivity Plating of MSF1 Variants	18
Figure 2.5	UAA-induced Disruption of GFPm-3	20
Figure 2.6	Immunostaining of GFPm-3 and MSF1-containing Mitochondria	21
Figure 2.7	Site of *GFPm-3* Gene Remains Intact After Passaging with UAA	22
Figure 2.8	Confirmation of tandem-Z Integration into BY4741 mtDNA	23
Figure 2.9	*Cox3::tandem-Z* Strains Don't Have Detectable Z-Domain Expression	24
Figure 3.1	UAA Respiration Sensitivity Test in sfGFPmito Strains	28
Figure 3.2	Fluorescence Sensitivity to UAA (2) in sfGFPmito Strain	29
Figure 3.3	Immunostaining of sfGFPmito with nVAR1p and MSF1m1	30
Figure 3.4	Pet111p Acts as a Translational Activator of sfGFPmito	32
Figure 3.5	RNR3 Overexpression Increases sfGFPmito Signal	33
Figure 3.6	CDDO-Me Leads to Protein Accumulation in sfGFPmito + MSF1m1	35
Figure 3.7	Schematic for Azide-labeling of UAAs in Mitochondria	36
Figure 3.8	UAA Click Microscopy in *S. Cerevisiae* Mitochondria	37
Figure 3.9	UAA Click Microscopy in Human Mitochondria	38
Figure 4.1	sfGFPmito-v3 Shows Anti-HIS Signal with UAA	43
Figure 4.2	Two Induction Scheme for Bacterial UAA Incorporation	44

Figure 4.3	4-iodo-phe Incorporation in E. coli sfGFP Peptide GIDFKEDGNILGHK	46
Figure 4.4	IP Eluates of HIS-sfGFP from 4-bromo-phe Mitochondria	48
Figure 4.5	No 4-bromo-phe Incorporation Observed in sfGFPv-3 IP Samples	49
Figure 4.6	No 4-iodo-phe Incorporation Observed in sfGFPmito-v3 Cobalt Samples	51
Figure A.1	Whole sfGFPmito-v3 with UAA on MALDI-TOF	61
Figure A.2	Anti-HIS Immunoblot of Bacterial UAA Induction Cultures	61
Figure A.3	OD-normalized RFU of Bacterial UAA Induction Cultures	62
Figure A.4	OD-normalized RFU of Bacteria Corrected for sfGFP Yield	62
Figure A.5	HIS Purification Eluates from E Coli UAA Induction	63
Figure A.6	4-iodo-phe Incorporation in E. coli sfGFP Peptide FEGDTLVNR	64
Figure A.7	4-iodo-phe Incorporation in E. coli sfGFP Peptide LEYNFNSHNVYITADK	65
Figure A.8	4-azido-phe Incorporation in E. coli sfGFP Peptide FEGDTLVNR	66
Figure A.9	4-azido-phe Incorporation in E. coli sfGFP Peptide GIDFKEDGNILGHK	67
Figure A.10	4-azido-phe Incorporation in E. coli sfGFP Peptide LEYNFNSHNVYITADK	68
Figure A.11	Cobalt Purification Recovers sfGFP from 4-iodo-phe Mitochondrial Samples	69

ACKNOWLEDGEMENTS

I would like to express thanks to my advisor, Professor Chang Liu, for financial support and a place in his lab over the course of my education. His lab was a rigorous learning environment filled with motivated individuals, and I appreciate my time as part of the group. I also thank Rishi Jajoo, for his initial work on the development of the mitochondrial expanded genetic code system, and for his useful comments. Benjamin Katz of the UCI Mass Sectrometry Facility was very helpful in developing protocols and trouble shooting my analytical pipeline.

I would also like to thank CCBS and my program, MCSB, for the financial and logistical support throughout my education. The staff at CCBS was very helpful for organizing meetings and helping me keep track of the administrative aspects of my degree. Thanks to The Beckman Foundation for their funding of this project over a multi-year span.

Lastly, I have to thank my parents for their continued support and unwavering belief in me.

ABSTRACT

Engineering Mitochondria Towards A Platform For Unnatural Polymer Synthesis

by

Ryan Nema Rezvani

Master of Science in Mathematical, Computational, and Systems Biology

University of California, Irvine, 2021

Associate Professor Chang Liu, Chair

The use of an expanded genetic code for unnatural amino acid (UAA) incorporation in living cells has thus far enabled the precise modification of protein function. However, achieving programmed ribosomal synthesis of fully synthetic polymers requires both free space in the genetic code and mutually orthogonal aminoacyl tRNA synthetase (aaRS)/tRNA pairs, two conditions that scale with the number of distinctly encoded novel chemistries in the polypeptide. To address these constraints, we repurposed yeast mitochondria into centers for orthogonal translation, taking advantage of both the nonessentiality of mitochondrial DNA (mtDNA) and the already-distinct ribosome and tRNAs used for mitochondrial translation. By expressing engineered mito-aaRS variants in yeast, we demonstrate the sensitivity of mitochondrially-encoded proteins to multiple UAAs. We also observe tRNA charging via mutant synthetase in both yeast and mammalian mitochondria via bioorthogonal labeling and microscopy. Lastly, we implement a series of expression-boosting genetic optimizations and construct a mitochondrial protein purification-to-MS analysis pipeline. Taken together, these results lay the groundwork for further engineering of mitochondrial translation towards the goal of unnatural polymer synthesis.

1. Introduction to Unnatural Polymer Synthesis Through Expanded Genetic Codes

1.1 Overview and Applications of Expanded Genetic Codes

Though nature has largely settled upon a standard 20 amino acid genetic code, deviations from this format have been known to confer unique functions when encoded in proteins, such as selenocysteine's lowered reduction potential relative to cysteine [1]. The natural occurrence of the nonstandard translational machinery (tRNAs, ribosomes, aaRSs) that enables these and more significant perturbations, like the more diverged genetic code in mitochondria, speaks to the pliability of native protein synthesis and suggests potential for experimental reprogramming [2,3]. Capitalizing on this potential, scientists have made great strides in engineering translational machinery for the direct incorporation of unnatural amino acids (UAAs) into select proteins [4]. These protein modifications confer physiochemical properties not afforded by the natural 20 amino acids, and have been reported in all manner of systems, from bacteria [5] to animals [6].

Before the widespread adoption of expanded genetic codes that allowed for site-specific incorporation solely in target proteins, incorporation of UAAs was achieved proteome-wide via the overexpression of a mutant aaRS [7,8]. In these efforts, led by Dave Tirrell, a substrate-promiscuous mutant of the native *E. coli* aaRS is overexpressed in a bacterial host that is auxotrophic for the native amino acid of the corresponding aaRS, ensuring that no intracellular synthesis of that amino acid occurs during the incorporation period. Additionally, this system involves the overexpression of a target protein shortly after the addition of the UAA, leading to incorporation throughout the protein of interest [8]. This expression window is typically limited, as incorporation of

the UAA through the rest of the *E. coli's* proteome is potentially deleterious to cellular function. This fitness defect during incorporation would scale with the number of mutant synthetases simultaneously employed, making this approach infeasible for multiplexed incorporation of UAAs. Furthermore, control of the incorporation site is only mediated via the chosen synthetase, making specific UAA labeling of targeted regions difficult.

A solution to some of these problems came in the form of a combination strategy involving the use of an unassigned codon (originally the rare amber stop in *E. coli*), and an aaRS/tRNA pair that doesn't exhibit cross-reactivity with the host aaRS/tRNA pairs (thus termed "orthogonal") [9]. By engineering the orthogonal aaRS to be reactive with the UAA of interest, it becomes possible to incorporate a diverse array of unnatural chemistries into only the sites templated by the open codon, creating a flexible and robust tool. Expanding this strategy to include multiple pairs of mutually orthogonal aaRS/tRNA pairs can further expand the simultaneously accessible chemical space for a variety of multiplex incorporation experiments, but the need for additional unassigned codons and mutual orthogonality between all aaRS/tRNA pairs (along with distinct specificities for the UAAs used) makes the problem challenging at scale.

The difficulties of multiplexing aside, just adding 1-2 UAAs at a time still confers a great amount of functionality, ranging from the modulation of protein therapeutics through bioorthogonal conjugation [10,11] to the expansion of a sequence space for directed evolution [12]. For instance, Tian *et al.* utilized site-specific incorporation of para-acetylphenylalanine (pAF) into target antibodies in mammalian cells, enabling a bioorthogonal reaction with a akloxyamine-linked small molecule for the formation of an antibody drug conjugate (ADC). Conjugates synthesized through this method were

found to be more effective *in vivo* than the equivalent ADC made through cysteine-based conjugation chemistries. In another example utilizing pAF conjugation chemistries, the pharmacokinetic properties of human growth hormone (hGH) were modulated through the pAF-mediated addition of PEG moieties [11]. A single added 30 kDa PEG molecule increased the hGH's half life, allowing for maintained effectiveness with a less frequent dosing schedule. In a display of the breadth of expanded genetic code applications, Young *et al.* used a UAG-encoded para-benzoylphenylalanine (pBzF) to expand the sequence space in a library of cyclic-peptide precursors [12]. A selection for cyclic peptide inhibitors against the HIV protease yielded multiple hits, the most potent of which contained a pBzF that was shown to react with the protease's K14 residue.

In addition to the aforementioned, more development-oriented applications, genetic code expansion has proven useful for interrogating function and mechanism in proteins and protein complexes. The incorporation of stable post-translational modifications (PTMs) into proteins gives a quantitative level of control, giving insights on the effect of tyrosine sulfation on anti-GP120 antibody effectiveness [5] and facilitating the solution of acetylated and ubiquitinated proteins' crystal structures [13]. Amino acids with special physical properties can be used to tease out protein dynamics as well. By incorporating a photoactivatable UAA into MAP kinase signaling component MEK1, high temporal resolution translocation of ERK2 was observable [14].

These cases are only a sample of the applications explored thus far with expanded genetic codes. However, the underlying technology is still very much in

development, with several research groups working to expand the potential of the ribosome.

1.2 Advancements in Expanded Genetic Codes and Unnatural Polymer Synthesis

While existing resources in expanded genetic codes are sufficient for incorporating a couple of unique UAAs at a time, there is a great appeal to a designer translation system, where different bespoke biopolymers can by synthesized based on an mRNA template. An extended translational synthesis capacity can make it possible to build molecules reminiscent of the heavily modified, non-ribosomal peptides found in nature, thus sampling new or existing therapeutic properties [15]. Towards this goal of unnatural polymer synthesis, there are multiple ongoing efforts to extensively engineer translational machinery, each with their own challenges.

Continuing from the exploitation of rare codons, like the amber stop in *E. coli*, significant efforts have focused on creating free space in the codon table, either through the systematic removal of certain codons from an organism's genome or through the generation of novel codons. Applying a mutually orthogonal set of aaRS/tRNA pairs to the now unassigned codons expands the coding potential of translation without impairing the host proteome (Figure 1.1A). Efforts on genomic recoding first yielded a bacterium with all of its amber stop codons removed, allowing for the use of UAG for unnatural incorporation without incurring fitness costs [16]. More recently, an *E. coli* genome was recoded via total synthesis to use only 61 of the 64 original codons, with two of the serine codons being systematically removed on top of the amber stop [17]. Pairing this strain with an engineered set of triply orthogonal aaRS/tRNA pairs, Jason Chin's group was able to incorporate three distinct UAAs into the same protein [18].

This work marks an impressive achievement, but the labor associated further removal of codons is considerable, as is the need for an expanding set of mutually orthogonal aaRS/tRNA pairs. In a different approach, there has also been progress made on the creation of novel codons, either through the use non-canonical base pairs that can be read specifically by a corresponding tRNA [19], or by the formation of quadruplet codons and anticodons [20].

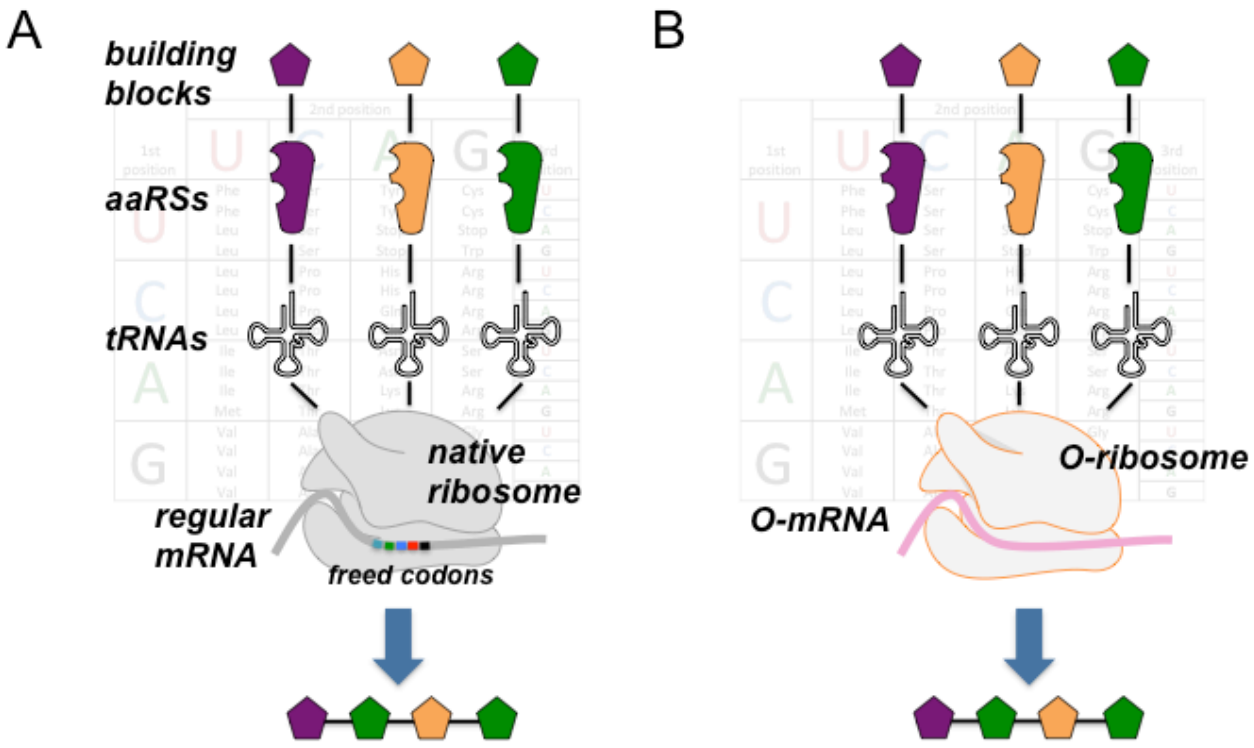


Figure 1.1 | Developing Strategies for Genetic Code Expansion
(A) Open codons are used to selectively encode for specific UAAs while still using the conventional host ribosome. These codons are either "freed" from the host genome, or are created through modified codon/base architecture. **(B)** An engineered orthogonal ribosome (O-ribosome) recognizes a target orthogonal mRNA (O-mRNA), creating a specific site for UAA incorporation. The ribosome can be engineered for relaxed restrictions with regards to quadruplet codons/anticodons.

The potential of quadruplet codons was enhanced with the development of an orthogonal ribosome (Figure 1.1B), which was engineered to more efficiently decode an

orthogonal mRNA containing quadruplet codons [21]. While there have been improvements in the translational efficiency and substrate scope of orthogonal ribosomes [22], extended biopolymer synthesis would require a larger set of aaRS/tRNA specific to the engineered ribosome. Progress has been made on engineering the components of translation [18, 23], but a revamp of translational machinery for the orthogonal ribosome remains an outstanding challenge.

Perhaps the most extensive modifications to ribosomal protein synthesis have come in controlled *in vitro* reactions (Figure 1.2). In this scheme, the necessary components for translation are added in a defined mixture, which includes tRNA that are separately acylated by a promiscuous ribozyme, labeled a flexizyme [24]. This use of precharged tRNAs in the final translation reaction circumvents the need for a panel of orthogonal aaRSs, allowing for scaling to include multiple simultaneous chemistries without engineering new components. Additionally, the wide substrate tolerance of the flexizyme ensures that the scope of chemical diversity is constrained mainly by the ribosome and tRNA. Taking advantage of this, Hiroaki Suga's group has engineered the tRNAs and ribosomes of their *in vitro* translation system for the polymerization of backbone modified peptides, such as polymers with consecutive β-amino acids [25]. This *in vitro* platform has been effective in generating the kinds of bioactive macrocyclic peptides that couldn't typically synthesized by natural ribosomal systems [26], though the process is labor-intensive and challenging at a larger scale.

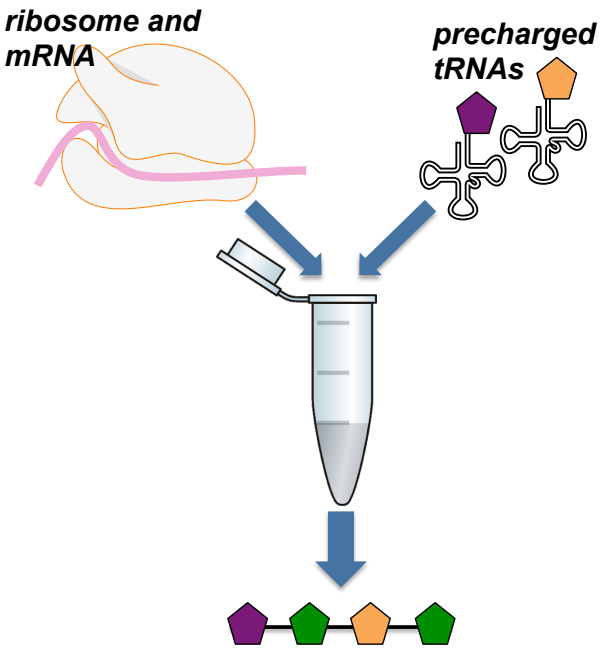


Figure 1.2 | Flexible *In Vitro* Translation
Ribosomal peptide synthesis is performed with select components, including the target mRNA template and various precharged tRNAs, each aminoacylated by a flexizyme in a separate reaction.

1.3 Mitochondria as a Center for Orthogonal Translation

An idealized expanded genetic code platform for unnatural polymer synthesis has convenience and reaction scalability of the aforementioned *in vivo* platforms while also limiting the chance for crosstalk between components, even in multiplexed incorporation of several distinct UAAs. *In vitro* translation platforms maintain that orthogonality via physical separation of tRNA acylation reactions from each other, ensuring that each UAA is linked to the desired anticodon. This concept of spatial separation as reinforcement of orthogonality was explored in a recent work that generated a membrane-less organelle around a target mRNA through the RNA-binding of phase separation-producing proteins [27]. By targeting engineered aaRSs to the designer organelle, researchers were able to get stop suppression specific to target

mRNA. This system, while innovative, will still rely on new aaRS/tRNA pairs to be mutually orthogonal with the existing machinery, and it remains unclear how many unique synthetases can be simultaneously targeted to the mRNA.

With the ideas of spatial separation and an extensive toolkit in mind, we turned to the already-orthogonal platform for protein synthesis in eukaryotes: the mitochondria. The genetic code of mitochondria varies between organisms, with the presence of a non-standard genetic code being a hallmark feature from yeast to humans [2, 28]. For example, in the mitochondrial code of *S. cerevisiae*, the entire CUN block codes for threonine instead of leucine, AUA codes for methionine instead of isoleucine, and UGA codes for tryptophan instead of a stop. Additionally, the CGA and CGC codons that normally encode arginine are absent from the mitochondrial genome, and the corresponding arginine tRNA contains an A in its wobble position, suggesting a loss of efficient recognition of those codons [2]. These systematic coding differences, along with mitochondria's distinctly encoded ribosomes, tRNAs, and aaRSs, are highly suggestive of a natural orthogonality between cytoplasmic and mitochondrial translation, a notion strengthened by the localization of all of these components to a spatially isolated part of the cell.

The system we have developed relies upon the overexpression of mitochondrial aaRS mutants that are designed to charge their native tRNA substrates with UAAs present in the growth medium. Similar to the Tirrell method of incorporation previously described, this would lead to UAA incorporation in all instances of matching codons in mitochondrial transcripts. Thus, it is necessary that the system is able to tolerate these additional incorporation events. In the case of yeast, none of the 8 main proteins

encoded in the mitochondrial genome are necessary for growth in glucose media [29]. In fact, *S. cerevisiae* can survive even when completely devoid of a mitochondrial genome, a condition known as rho^0. 7 of the 8 mitochondrially-translated proteins are involved in respiration, with the last being a protein component of the mito-ribosome's small subunit, Var1p. Fortunately, should UAA incorporation into Var1p pose a problem, it is possible to preserve ribosome functionality through the cytoplasmic expression and import of a nuclearly encoded *VAR1* gene [30].

Via the expression of substrate-promiscuous mitochondrial aaRSs and accessory proteins for optimal translation, we have created a platform where a mito-localized sequence of interest can be decoded to synthesize unnatural polymers (Figure 1.3). Multiplexing UAA incorporation simply involves the overexpression of a different aaRSs mutant, with the added requirement of checking for UAA-synthetase specificity to minimize crosstalk. We tested mutants of the tyrosyl- and phenylalanyl-tRNA synthetases through their ability to disrupt respiration and reduce fluorescence of mito-encoded GFP reporters only the presence of an UAA [31, 32]. After optimizing mitochondrial expression level, we purified GFP and found suggestions of partial incorporation of phenylalanine analog via mass shift. These indications are ready to be explored with higher resolution modalities. Finally, we used click labeling to observe alkyne incorporation into yeast and human mitochondrial proteins via fluorescence microscopy, suggesting potential utility in labeling of protein synthesis in human mitochondria [33].

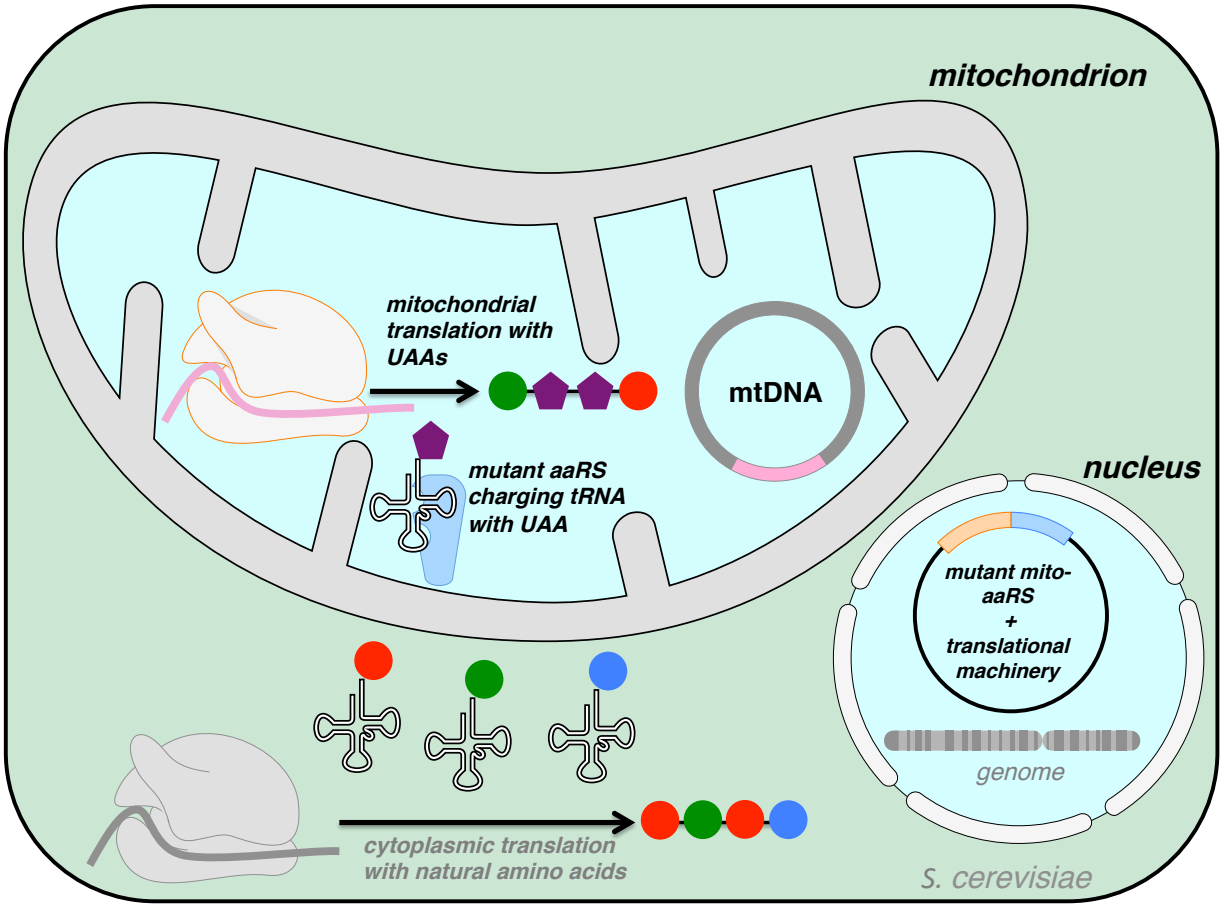


Figure 1.3 | Mitochondrial Platform for Genetic Code Expansion
Substrate-promiscuous aaRS mutants are imported into the mitochondria, where they charge the native tRNAs with a target UAA. This process yields UAA incorporation in mitochondrial proteins, while cytoplasmic translation remains unperturbed.

1.4 References

[1] Johansson L., Gafvelin G., Arnér E.S.J. Selenocysteine in proteins—properties and biotechnological use *Biochimica et Biophysica Acta (BBA) - General Subjects* **1726**, 1-13 (2005)

[2] Bonitz S.G. *et al.* Codon recognition rules in yeast mitochondria *Proc. Natl. Acad. Sci. U.S.A* **77** 3167-3170 (1980)

[3] Osawa, S. *et al.* Evolution of the mitochondrial genetic code III. Reassignment of CUN codons from leucine to threonine during evolution of yeast mitochondria. *J Mol Evol* **30,** 322-328 (1990)

[4] Li, X. & Liu, C.C. Biological Applications of Expanded Genetic Codes. *ChemBioChem* **15**, 2335-2341 (2014)

[5] Li, X., Hitomi, J., Liu C.C. Characterization of a Sulfated Anti-HIV Antibody Using an Expanded Genetic Code *Biochemistry* **57,** 2903-2907 (2018)

[6] Brown W. *et al.* Cell-Lineage Tracing in Zebrafish Embryos with an Expanded Genetic Code *ChemBioChem* **19**, 1244 (2018)

[7] Kiick, K.L., van Hest, J.C.M., Tirrell, D.A. Expanding the Scope of Protein Biosynthesis by Altering the Methionyl-tRNA Synthetase Activity of a Bacterial Expression Host. Angew. *Chem. Int. Ed.* **39**, 2148-2152 (2000)

[8] Kirshenbaum, K., Carrico, I.S., Tirrell, D.A. Biosynthesis of Proteins Incorporating a Versatile Set of Phenylalanine Analogues *ChemBioChem* **3**, 235-237 (2002)

[9] Wang L. *et al.* Expanding the genetic code of Escherichia coli *Science* **292,** 498-500 (2001)

[10] Tian, F. *et al.* A general approach to site-specific antibody drug conjugates. *Proc. Natl. Acad. Sci. U.S.A* **111**,1766-1771 (2014)

[11] Cho, H. *et al.* Optimized clinical performance of growth hormone with an expanded genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9060–5 (2011)

[12] Young, T.S. et al. Evolution of cyclic peptide protease inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **108** 11052-6 (2011)

[13] Davis, L. & Chin, J. Designer proteins: applications of genetic code expansion in cell biology. *Nat Rev Mol Cell Biol* **13,** 168–182 (2012)

[14] Gautier, A., Deiters, A., Chin, J. W. Light-activated kinases enable temporal dissection of signaling networks in living cells. *J. Am.Chem.Soc*. **133**, 2124–2127 (2011)

[15] Sieber, S.A. & Marahiel M.A. Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics *Chemical Reviews* **105,** 715-738 (2005)

[16] Lajoie, M.J. *et al.* Genomically Recoded Organisms Expand Biological Functions *Science* (2013)

[17] Fredens, J. *et al.* Total synthesis of Escherichia coli with a recoded genome. *Nature* **569**, 514–518 (2019)

[18] Dunkelmann, D.L. *et al.* Engineered triply orthogonal pyrrolysyl–tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nat. Chem.* **12**, 535–544 (2020)

[19] Fischer, E.C. *et al.* New codons for efficient production of unnatural proteins in a semisynthetic organism. *Nat Chem Biol* **16**, 570–576 (2020)

[20] Anderson J.C. *et al.* An expanded genetic code with a functional quadruplet codon. *Proc Natl Acad Sci U S A* **101**, 7566-7571 (2004)

[21] Neumann, H. *et al.* Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441–444 (2010)

[22] Schmied, W.H. *et al.* Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature* **564**, 444–448 (2018)

[23] Cervettini, D. *et al.* Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase–tRNA pairs. *Nat Biotechnol* **38**, 989–999 (2020)

[24] Murakami, H. *et al.* A highly flexible tRNA acylation method for non-natural polypeptide synthesis*. Nat Methods* **3**, 357–359 (2006)

[25] Katoh T. & Suga H. Ribosomal Incorporation of Consecutive β-Amino Acids. *J Am Chem Soc.* **140**, 12159-12167 (2018)

[26] Passioura T. *et al.* Selection-Based Discovery of Druglike Macrocyclic Peptides *Annual Review of Biochemistry* 83, 727-752 (2014)

[27] Reinkemeier C.D., Girona G.E., Lemke E.A. Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. *Science*. **363** (2019)

[28] Barrell, B., Bankier, A., Drouin, J. A different genetic code in human mitochondria. *Nature* **282**, 189–194 (1979)

[29] Fox, T.D. Mitochondrial protein synthesis, import, and assembly. *Genetics* **192**, 1203-34 (2012)

[30] Sanchirico M. *et al.* Relocation of the unusual VAR1 gene from the mitochondrion to the nucleus. *Biochem Cell Biol*. **73**, 987-995 (1995)

[31] Cohen, J.S. & Fox, T.D. Expression of green fluorescent protein from a recoded gene inserted into Saccharomyces cerevisiae mitochondrial DNA. *Mitochondrion* **1,** 181-189 (2001)

[32] Suhm, T. *et al.* A novel system to monitor mitochondrial translation in yeast. *Microbial cell* **5**,158-164 (2018)

[33] Deiters, A. *et al.* Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *J Am Chem Soc.* **125**,11782-11783 (2003)

2. Design and Implementation of a Mitochondrial Expanded Genetic Code

2.1 Aminoacyl-tRNA Synthetase Design and Validation

The introduction of mutant aaRSs to the mitochondria is fundamental to our goal of UAA incorporation in mito-translated proteins. Generally, the chosen mutants should have an enhanced substrate scope to facilitate aminoacylation with one or more UAAs while also remaining active enough to outcompete the wild type synthetase, if necessary, when overexpressed. Functional mutants also need to maintain interactions with corresponding native tRNAs, as although there are known examples of $tRNA^{Lys}$(CUU) import, the co-import mechanism necessary for mitochondrial entry is specific to pre-LysRS's interaction with $tRNA_{Lys}$(CUU) [1]. Additionally, there's evidence that the glycolytic enzyme Eno2p acts as an RNA cofactor during pre-import and has specific affinity for the imported tRNA [2]. These targeted molecular interactions made co-import of a foreign aaRS/tRNA pair seem infeasible, lest we were able to reapply key import machinery to adequately recognize the nonnative synthetase and tRNA. There also potentially exists a mechanism for direct import of tRNAs in yeast, as discussed in a report showing cytoplasmic $tRNA^{Gln}$(CUG) suppressing a nonsense mutation in the mitochondrially encoded *cox2* gene [3]. The mechanism for this apparently direct import is not understood, but a followup study was unable to detect the presence of cytoplasmic glutamine tRNA in mitochondria [4], further dissuading exploration of the tRNA import angle.

With these considerations in mind, we chose to focus on the engineering of preexisting mitochondrial aaRSs to relax amino acid specificity. The mutant aaRSs

would then be overexpressed to maximize reactivity with their cognate tRNAs over any wild type (WT) synthetase. We prioritized the mitochondrial phenylalanyl- and tyrosyl-tRNA synthetases, as both of these proteins are significantly homologous to their bacterial counterparts (Figure 2.1), allowing us to apply mutations from engineered *E. coli* synthetases to mitochondrial ones.

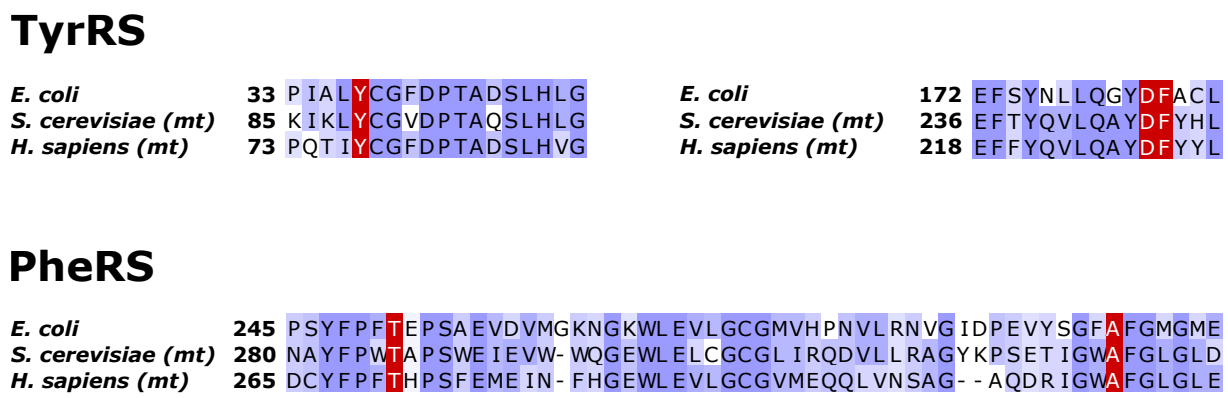


Figure 2.1 | Alignment of *E. coli* and Mitochondrial Synthetases
Protein sequence alignments of the *E. coli* TyrRS and PheRS against mitochondrial homologs in yeast and humans. Residue shading is based on Blosum62 similarity scores between amino acids. Positions highlighted in red are mutated in this study.

Starting with TyrRS engineering, we drew from a previous study that introduced Y37T, D182S, and F183A substitutions into the *E. coli* TyrRS to facilitate incorporation of reactive UAAs **1** and **5** (Figure 2.2) [5], and thus added these mutations to the yeast mitochondrial TyrRS, MSY1. Terming this variant MSY1m, we also created a fusion of the bacterial TyrRS mutant, called EcAz6, and the N terminus of MSY1, based on knowledge that a chimera between the mitochondrial and bacterial TyrRS synthetases is able to complement in yeast [6]. This chimera was labeled MSY1-EcAz6. We then introduced both MSY1m and MSY1-EcAz6 into the standard *S. cerevisiae* chassis BY4741 under control of the strong TDH3 promoter (TDH3pr), such that they are overexpressed relative to the WT MSY1 in the genome [7].

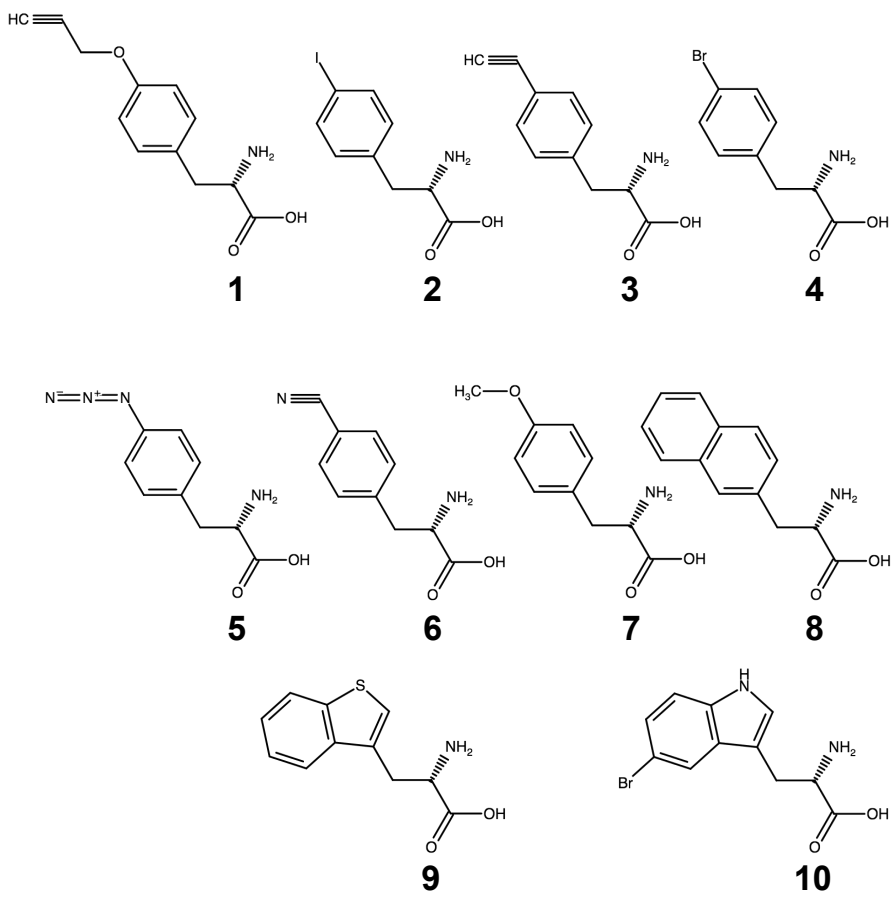


Figure 2.2 | Unnatural Amino Acids Used in this Study
The panel of UAAs used for expanded genetic code testing in the mitochondria. **1**, O-propargyl-L-tyrosine; **2,** 4-iodo-L-phenylalanine; **3**, 4-ethynyl-L-phenylalanine; **4**, 4-bromo-L-phenylalanine; **5**, 4-azido-L-phenylalanine; **6**, 4-cyano-L-phenylalanine; **7**, O-methyl-L-tyrosine; **8**, L-3-(2-naphthyl)alanine; **9**, 3-benzothienyl-L-alanine; **10**, 5-bromo-L-tryptophan

To test the functionality of these mutants, we designed a spot-plating assay to observe sensitivity to a target UAA only in the presence of a mutant aaRS. 7 of the 8 proteins translated in the yeast mitochondria are involved in respiration, with the 8th being a component of the ribosome [8]. We predicted that widespread incorporation of UAAs would interfere with this respiration machinery, thus rendering the cell incapable of growth on respiration-demanding glycerol media. Additionally, a minimal change in fitness on glucose media would indicate that the UAA sensitivity is indeed linked to

mitochondrial translation and not alternative sources of toxicity. In order to minimize the chance of ribosomal breakdown, we opted to express the mitochondrial ribosome small subunit protein VAR1p from the nucleus (nVAR1p), as to ensure that it remains unincorporated and operational [9].

We plated strains expressing nVAR1p and either wild type MSY1, MSY1m, or MSY1-EcAz6 onto glucose and glycerol plates containing either no UAA or 1mM of **1** (Figure 2.3). Interestingly, the chimeric MSY-EcAz6 appears to confer much greater UAA sensitivity than MSY1m, even though there is strong homology in the substituted regions of the synthetase (Figure 2.1). Additionally, the MSY1m variant appears less functional than the chimera, as evidenced by the reduced growth on glycerol in the absence of UAA. It is unlikely that the lowered fitness in this media condition comes as the result of mischarging of tRNAs with other natural amino acids, as that degree of promiscuity would imply a greater sensitivity to the UAA in respiring conditions. Thus, we moved forward with MSY1-EcAz6 as the preferable TyrRS mutant for subsequent experiments.

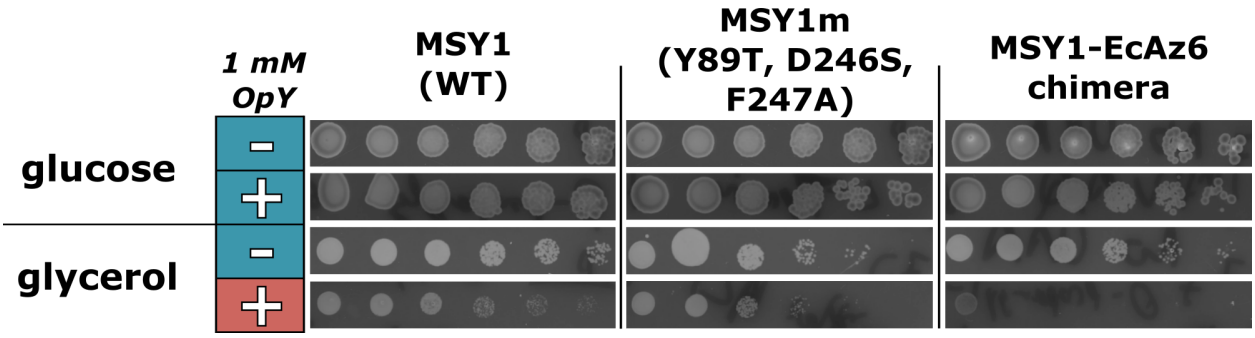


Figure 2.3 | UAA Sensitivity Plating of MSY1 Variants
Wild type (WT) MSY1, MSY1m, and MSY1-EcAz6 are spot plated with serial dilutions on glucose and glycerol media, with and without the UAA OpY, or **1**. All three variants are under TDH3pr expression and are co-expressed with nVAR1p. Conditions labeled in green are expected to accommodate cell growth with successful UAA incorporation, while the condition in red should exhibit reduced growth with incorporation.

Applying a similar logic to engineering of the mitochondrial PheRS, MSF1, we introduced mutations at two sites based on previous work on the *E. coli* PheRS [10,11], producing the MSF1 substitutions T286G and A328G, both thought to create space in the substrate binding pocket of the synthetase. Though the equivalent of the A328G mutation in the bacterial PheRS was the first mutant to show incorporation of phenylalanine analogs **2**-**6** [10], subsequent studies showed preferentially used the T286G equivalent for relaxing substrate specificity in order to incorporate bulkier analogs like **8** [12], partially due to a structural analysis in which that threonine was the top hit [11].

We proceeded to express WT MSF1, the T286G single mutant (MSF1m1), and the T286G, A328G double mutant (MSF1m2) under control of TDH3pr with co-expression of nVAR1p, all in BY4741, as was done with the MSY1 variants. These three strains were then plated on glycerol media supplemented with either no UAA, **2**, or **3** (Figure 2.4). Previous experiments plating confirmed that the mutants and UAAs were not excessively toxic at the concentrations used (data not shown). Though both the single and double mutants confer growth-sensitivity to **2** and **3**, MSF1m2 exhibits poor growth even in the absence of unnatural, possibly indicating that the variant is less active. In any case, we proceeded with MSF1m1 for subsequent experiments given that it both more robustly inhibited growth in the presence of UAA and had less of a defect in the absence of UAA.

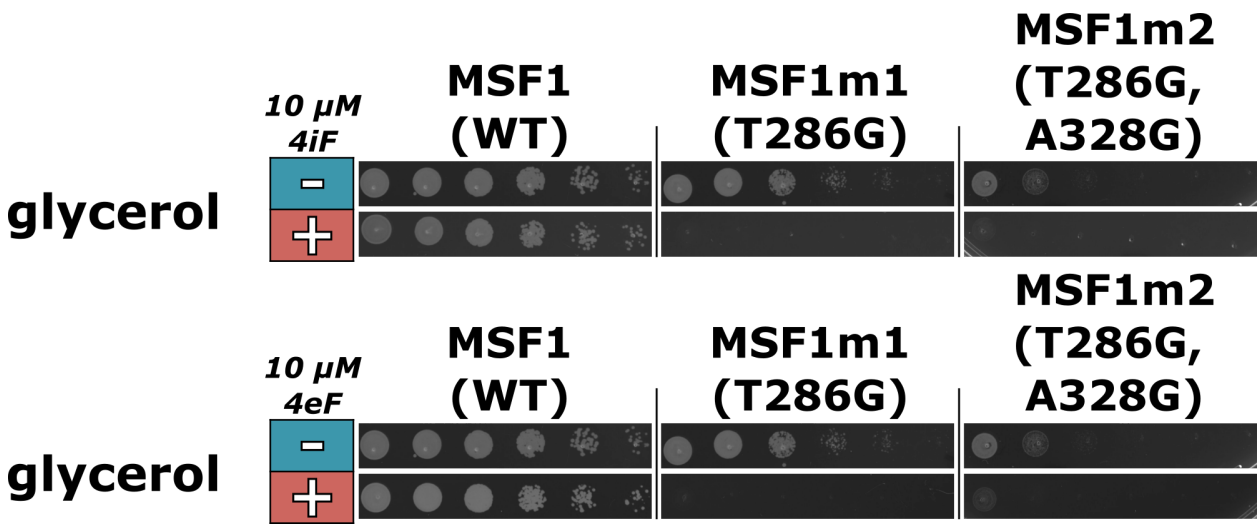


Figure 2.4 | UAA Sensitivity Plating of MSF1 Variants
Wild type (WT) MSF1, MSF1m1, and MSF1m2 are spot plated with serial dilutions on glycerol media, with and without the UAAs 4iF (**2**) or 4eF (**3**). All three variants are under TDH3pr expression and are co-expressed with nVAR1p. Conditions labeled in green are expected to accommodate cell growth with successful UAA incorporation, while the condition in red should exhibit reduced growth with incorporation.

2.2 Application of Mutant aaRSs to a Mitochondrial GFP Reporter System

With the plating experiments suggesting potential UAA incorporation in native mitochondrial genes, we sought to move our platform to a model system that would readily allow for a finer-grained characterization. Furthermore, we wanted confirm incorporation of UAAs via mass shift, and there are fewer tools developed for the purification and downstream analysis of any protein naturally encoded in the mito-genome. To this end, we turned to a yeast strain that has green fluorescent protein (GFP) integrated into the mitochondrial DNA (mtDNA) in place of the cytochrome oxidase gene *Cox3*, creating GFPm-3 [13]. To ensure consistent expression, Cohen & Fox preserved the 5' UTR and first 8 codons of the *Cox3* gene based on previous work showing that the early portions of the *Cox3* mRNA are necessary for translation [14]. With this GFPm-3 strain, we intended to get a more quantitative view on the possible

UAA incorporation. Importantly, GFPm-3 is reported to remain soluble after translation, in contrast with COX3 and the other mitochondrially-translated proteins [13].

We introduced our TDH3pr-driven MSF1 and MSF1m1 into the GFPm-3 reporter strain for subsequent testing of sensitivity UAAs **2** and **4**-**10**. Flow cytometry was performed over a range of concentrations of each UAA, resulting in a panel of differing sensitivities to the multiple UAAs (Figure 2.5). As the concentration of the supplemented amino acids increases, there is a sharp decline in fluorescence for most of the UAAs tested, with the most sensitive cases being for **2**, **4**, and **5**. Additionally, the larger amino acids tested show a reduced response (**8**) or no response at all (**9** and **10**). This seemingly size-correlated effect supports the notion that the observed phenomena are mediated by the enhanced substrate tolerance in MSF1m1's reactivity, but with these results alone it is unclear if the aminoacylated tRNAs are successfully depositing UAAs into GFPm-3 and causing reduced fluorescence/misfolding, or if there is simply a breakdown or stalling of mitochondrial translation as a whole.

To get a better sense of GFP quantity in the presence of UAAs, we isolated mitochondria from the GFPm-3 strain expressing MSF1m1, both with and without the presence of 1mM of **2**. Western blots of these mitochondrial samples show the expression of both GFPm-3 and imported control protein ATP2, but not in all conditions (Figure 2.6). When the mutant synthetase-containing strain is grown with UAA in the media, GFP becomes undetectable while ATP2 signal remains. This can possibly be explained by the absence of nVAR1p in the GFPm-3 tested strains, meaning that incorporation of UAAs into mitochondrially-produced VAR1 could subsequently impair translation. Limitations on open auxotrophic markers in the GFPm-3 strain, as well as

recombination-prone portions of genomic sequence around the *Leu2* ORF [15], created difficulties in readily producing strains co-expressing MSF1m1 and nVAR1p. Eventually a *KanMX*-mediated disruption of *Leu2* was successful, though by that time a newer strain platform was adopted.

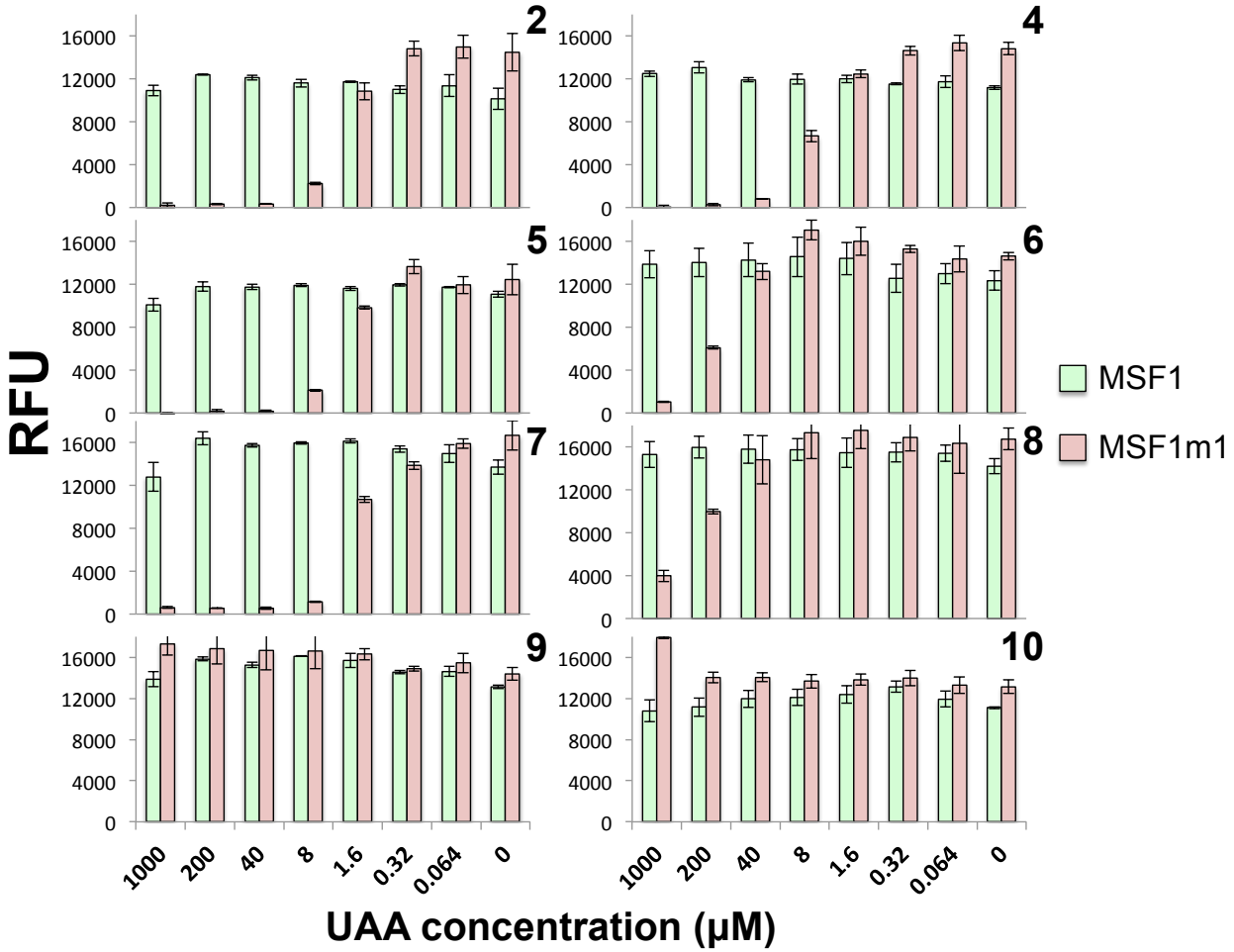


Figure 2.5 | UAA-Induced Disruption of GFPm-3
Flow cytometry of GFPm-3 cells expressing either MSF1 or MSF1m1. The bolded numbers on the top left of each plot denotes the UAA. Measurements are background subtracted using GFP-less BY4741. Bars represent the averaged median of three measurements.

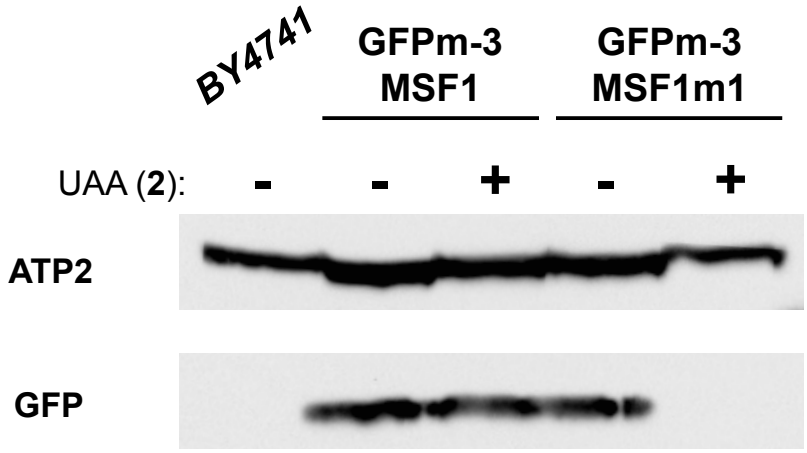


Figure 2.6 | Immunostaining of GFPm-3 and MSF1-containing Mitochondria

Immunoblots of mitochondrial samples from strains containing GFPm-3 and MSF1/MSF1m1. ATP2, an imported mitochondrial protein, is used as a protein quantity control. The BY4741 plasmid contains no GFP and serves as a negative control for GFP signal. Samples labeled as +UAA were grown in 1mM of **2**.

We next wanted to confirm that growth in the presence of a UAA wasn't interfering with the mitochondrial DNA maintenance. There is a well-characterized relationship between the impairment of mitochondrial translation and mito-genome instability, with the potential for genome loss [16,17]. In order to rule out a stark reduction in mtDNA as a possible cause of protein reduction, we passaged GFPm-3 MSF1m1 cultures for 4 rounds in the presence of 1mM of **2** and subsequently performed colony PCRs over the GFPm-3 integration flank in the *Cox3* locus. At the end of passaging, all five of the replicates grown in **2** produced the amplicon specific to the GFPm-3 strain (Figure 2.7A). Additionally, after the 4 rounds of UAA cultures, the replicates were passaged for a final time into UAA-free media, after which they recovered their expected fluorescent signal (Figure 2.7B). Although less extensive instances of mtDNA damage could have occurred over the passaging, this result shows that major loss of the mitochondrial genome is not likely to happen on the timescale of a typical incorporation experiment.

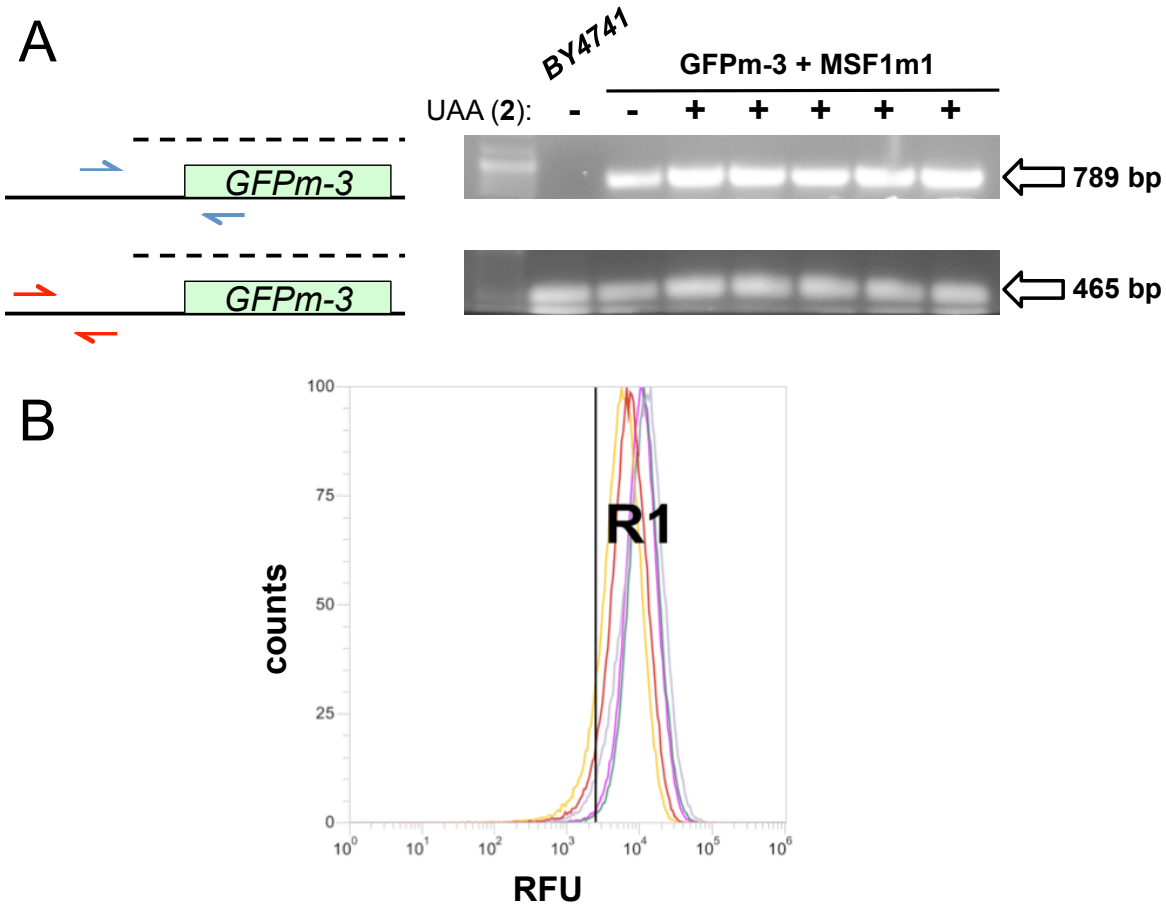


Figure 2.7 | Site of *GFPm-3* Gene Remains Intact After Passaging with UAA
A) Five replicates of the GFPm-3 strain expressing MSF1m1 were passaged 4 times in the presence of 1mM UAA, then two sets colony PCRs were performed based on the left schematic. Dotted lines represent the integration payload. Arrows represent primers used. BY4741 represents a negative control for *GFPm-3* amplification. B) Flow cytometry histogram of the five replicates upon reintroduction to no-UAA media at the end of passaging. R1 refers to the median fluorescence value of the cultures under exposure to UAA.

2.3 Transformation of Mitochondria for the Addition of an Affinity-tagged Protein

In order to both express potentially necessary accessory machinery like nVAR1p and enable affinity purification of our target protein, we set out to introduce a new gene to the mitochondrial DNA of a more tractable strain of our choosing. Z-domain, a synthetic IgG-binding protein based on staphylococcal Protein A that was engineered for enhanced stability [18], is often used as a target for incorporation due to its stability

and easy-of-purification [19-21]. We designed a gene for mitochondrial transformation into the *Cox3* locus comprised of two sequential instances of Z-domain, with a myc tag on the N-terminus and 6xHIS tag on the C-terminus. As is the case with the *GFPm-3* gene, we preserved the 5'UTR of COX3 transcript and included the first 8 amino acids of the COX3 protein. This modified Z-domain was termed the tandem-Z.

In order to introduce the tandem-Z to the mitochondrial genome, we used the microprojectile bombardment method of transformation, where DNA-coated metal microparticles are shot onto a lawn of cells lacking mtDNA [22]. Cells containing the recombination construct are then co-incubated with yeast of the opposite mating type to deliver the transformed construct, which then integrates into the recipient's mitochondrial genome via homologous recombination. We carried out this procedure using BY4741 as the recipient strain and screened transformation clones for successful integration via colony pcr (Figure 2.8). Although all of the screened tandem-Z clones screened had detectable integrant DNA, only those with the correctly sized amplicon spanning beyond the homology arm show signs of successful integration.

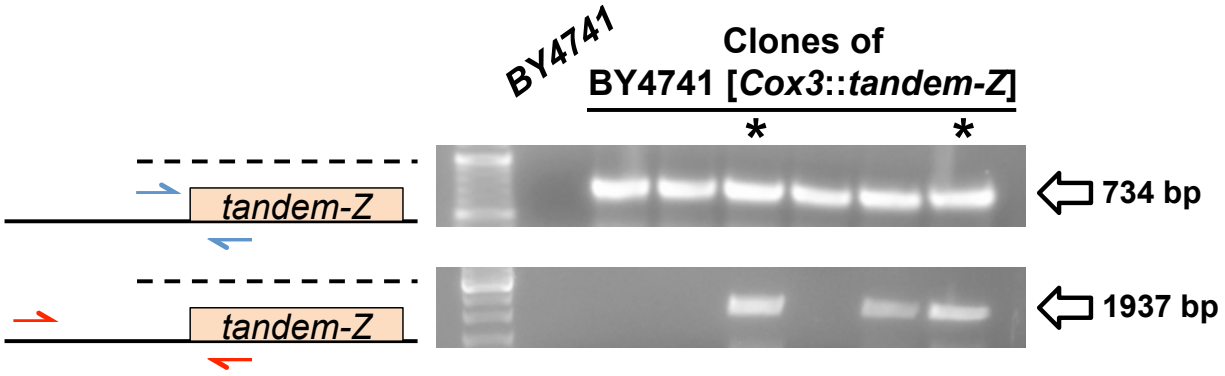


Figure 2.8 | Confirmation of tandem-Z Integration into BY4741 mtDNA
Colony PCRs inside and spanning the integration flank, labeled by the dotted line in the left schematics. Arrows represent primers used. Starred clones are used in subsequent experiments.

Following up on this indication of successful integration, we purified mitochondria of two separate transformants and immunoblotted for HIS- and myc-tags (Figure 2.9). While the control ATP2 signal was detectable in these samples, where was no signal for either of the two affinity tags, a result that we have observed with multiple distinct mito-preparations. We considered the possibility that a small deletion in the *Cox3* 5' UTR could impair recognition of the tandem-Z mRNA and thus reduce protein expression [14], leading us to sequence the PCR represented by the red primers in Figure 2.8. We recovered the correct *Cox3::tandem-Z* sequence for both starred clones, leaving the cause of the absent protein expression uncertain.

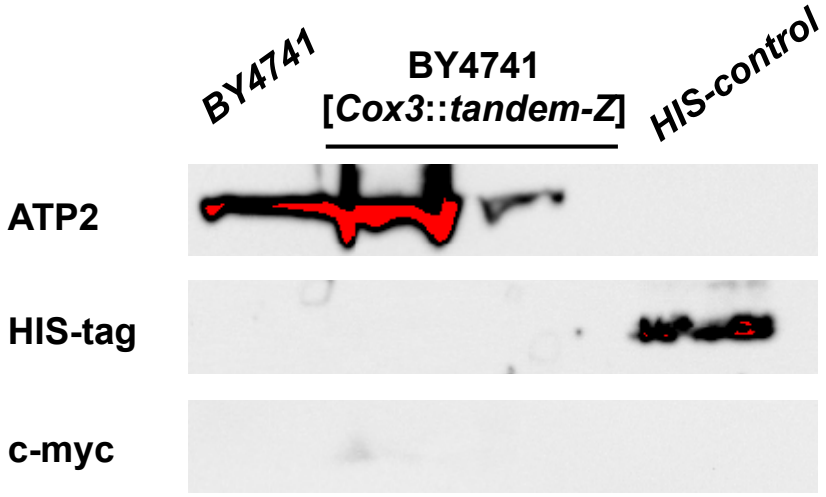


Figure 2.9 | *Cox3::tandem-Z* Strains Don't Have Detectable Z-Domain Expression Immunoblot showing showing the presence of imported mitochondrial protein ATP2, but no tandem-Z via either HIS-tag or myc-tag recognition. HIS-control refers to a purified, HIS-tagged protein of similar size to the tandem-Z (~17kDa).

2.4 References

[1] Schneider, A. Mitochondrial tRNA Import and Its Consequences for Mitochondrial Translation *Annual Review of Biochemistry* **80**, 1033-1053 (2011)

[2] Entelis, N. *et al.* A glycolytic enzyme, enolase, is recruited as a cofactor of tRNA targeting toward mitochondria in Saccharomyces cerevisiae. *Genes & development* **20**,1609-20 (2006)

[3] Rinehart, J. *et al.* Saccharomyces cerevisiae imports the cytosolic pathway for Gln-tRNA synthesis into the mitochondrion. *Genes & Development* **19**, 583-92 (2005)

[4] Frechin, M. *et al.* Yeast mitochondrial Gln-tRNA(Gln) is generated by a GatFAB-mediated transamidation pathway involving Arc1p-controlled subcellular sorting of cytosolic GluRS. *Genes & Development* **23**, 1119-30 (2009)

[5] Deiters, A. *et al.* Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *J Am Chem Soc.* **125**,11782-11783 (2003)

[6] Edwards, H. & Schimmel, P. An E. coli aminoacyl-tRNA synthetase can substitute for yeast mitochondrial enzyme function in vivo. *Cell* **51**, 643-649 (1987)

[7] Lee, M.E. *et al.* A highly characterized yeast toolkit for modular, multipart assembly. *ACS synthetic biology* **4,** 975-986 (2015)

[8] Fox, T.D. Mitochondrial protein synthesis, import, and assembly. *Genetics* **192**, 1203-34 (2012)

[9] Sanchirico M. *et al.* Relocation of the unusual VAR1 gene from the mitochondrion to the nucleus. *Biochem Cell Biol*. **73**, 987-995 (1995)

[10] Kirshenbaum, K., Carrico, I.S., Tirrell, D.A. Biosynthesis of Proteins Incorporating a Versatile Set of Phenylalanine Analogues *ChemBioChem* **3**, 235-237 (2002)

[11] Datta, D. *et al.* A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins. *J Am Chem Soc.* **124**, 5652-3 (2002)

[12] Kwon, I., Kirshenbaum, K., Tirrell, D. A. Breaking the degeneracy of the genetic code. *J Am Chem Soc.* **125**, 7512-7513 (2003)

[13] Cohen, J.S. & Fox, T.D. Expression of green fluorescent protein from a recoded gene inserted into Saccharomyces cerevisiae mitochondrial DNA. *Mitochondrion* **1,** 181-189 (2001)

[14] Steele, D. F., Butler, C. A., Fox, T. D. Expression of a recoded nuclear gene inserted into yeast mitochondrial DNA is limited by mRNA-specific translational activation. *Proc. Natl. Acad. Sci. U.S.A.* ***93***, 5253-5257 (1996)

[15] Dobson, M. J., Kingsman, S. M., Kingsman, A. J. Sequence variation in the LEU2 region of the Saccharomyces cerevisiae genome. *Gene*, ***16***, 133-139 (1981)

[16] Contamine, V., & Picard, M. Maintenance and integrity of the mitochondrial genome: a plethora of nuclear genes in the budding yeast. *Microbiology and Molecular Biology Reviews* ***64***, 281-315 (2000)

[17] Myers, A. M., Pape, L. K., Tzagoloff, A. Mitochondrial protein synthesis is required for maintenance of intact mitochondrial genomes in Saccharomyces cerevisiae. *The EMBO journal* ***4***, 2087–2092 (1985)

[18] Nilsson, B. *et al.* A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Engineering, Design and Selection* **1**, 107-113 (1987)

[19] Liu, H. *et al*. A method for the generation of glycoprotein mimetics. *Journal of the American Chemical Society* ***125***, 1702-1703 (2003)

[20] Zhang, Z. *et al.* The selective incorporation of alkenes into proteins in Escherichia coli. *Angewandte Chemie (International ed. in English)* ***41*** 2840–2842 (2002)

[21] Wang, L. *et al.* Addition of the keto functional group to the genetic code of Escherichia coli. *Proceedings of the National Academy of Sciences* **100,** 56-61 (2003)

[22] Bonnefoy, N., Remacle, C., Fox, T. D. Genetic transformation of Saccharomyces cerevisiae and Chlamydomonas reinhardtii mitochondria. *Methods in cell biology* **80**, 525–548 (2007)

3. Optimization of the Mitochondrial Expanded Genetic Code Platform

3.1 Adopting a Mitochondrial Superfolder GFP Strain

After difficulties with recovery of GFPm-3 in the presence of UAA, we moved the expanded genetic code platform to a recently developed yeast strain expression superfolder GFP (sfGFP) in the mitochondria [1]. This system sports several advantages over the GFPm-3 chassis. First, superfolder GFP is both more stable and brighter than wild type GFP. Additionally, the mitochondrial sfGFP (sfGFPmito) has a C-terminal HIS tag for affinity capture, and is integrated upstream of the *Cox2* locus in a fashion that avoids the disruption of any genes, thus the strain maintains the ability to respire. This sfGFPmito strain was built on the W303 strain, meaning that four auxotrophic markers are open for use.

We transformed nVAR1p and our mutant synthetase overexpression constructs into the sfGFPmito strain and performed the same growth assay previously done in BY4741, checking for the disruption of respiration solely in the presence of both the UAA and synthetase (Figure 3.1). Compared to the previous respiration fidelity experiment, **2** was added at a higher concentration, likely enhancing its overall toxicity in all backgrounds. We also observe what appears to be cross reactivity between the MSY1-EcAz6 chimera and **2**, though not with the same level of sensitivity as with MSF1m1. This suggests that use of **2** in a multiplexed setting should be done at a lower concentration to reduce the rate of cross-charging of tRNAs.

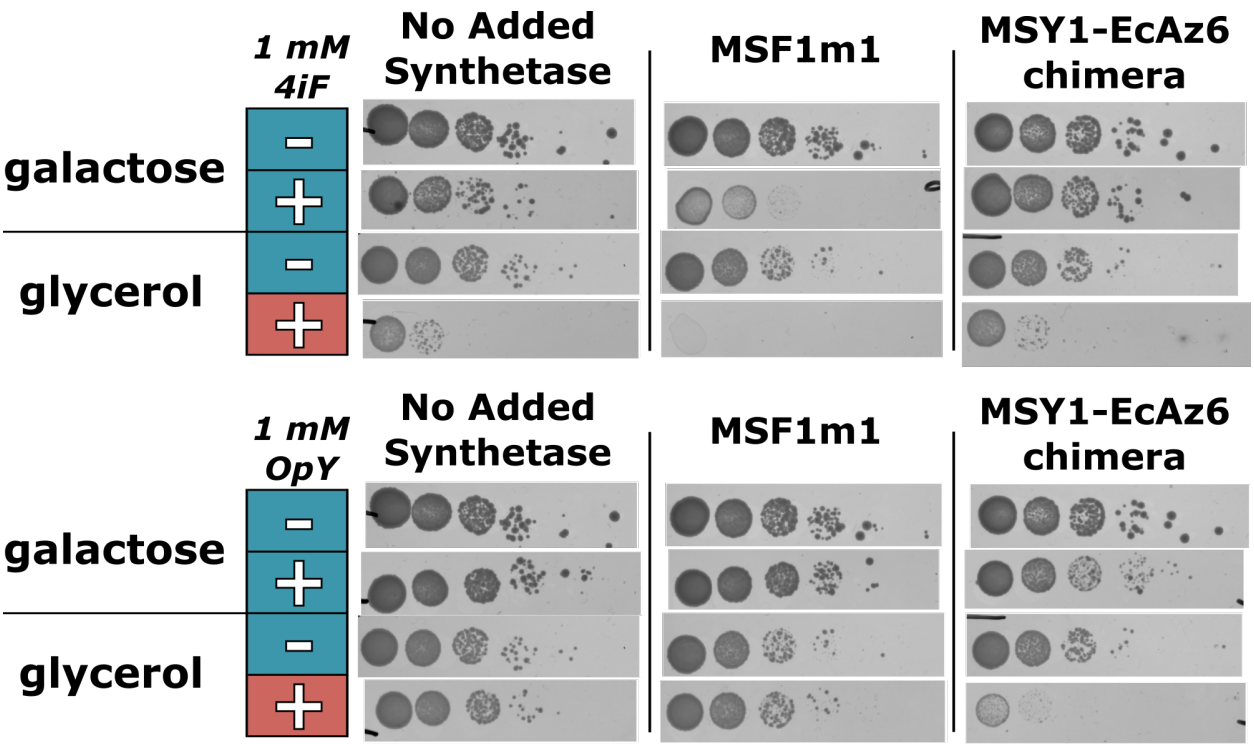


Figure 3.1 | UAA Respiration Sensitivity Test in sfGFPmito Strains
Serial dilution plating of cultures in response to UAA and the presence of either a MSF1 mutant, MSY1 mutant, or no added synthetase. 4iF refers to **2**, OpY refers to **1**. All strains are expressing nVAR1p. Conditions labeled in green are expected to accommodate cell growth with successful UAA incorporation, while the condition in red should exhibit reduced growth with incorporation.

Next, we performed a similar UAA titration cytometry experiment as was conducted with the GFPm-3 strain, but with the inclusion of nVAR1p, which should help prevent direct disruption of the mitochondrial ribosome [2]. Interestingly, a drop in fluorescence upon addition of **2** is still observed, either suggesting that sfGFPmito is being directly disrupted by the unnaturals, or that UAA-charged tRNAs are causing overall translational instability at the peptidyl transferase site (Figure 3.2). Given that the 1 millimolar condition gives the lowest output fluorescence, we proceeded with that concentration for protein characterization.

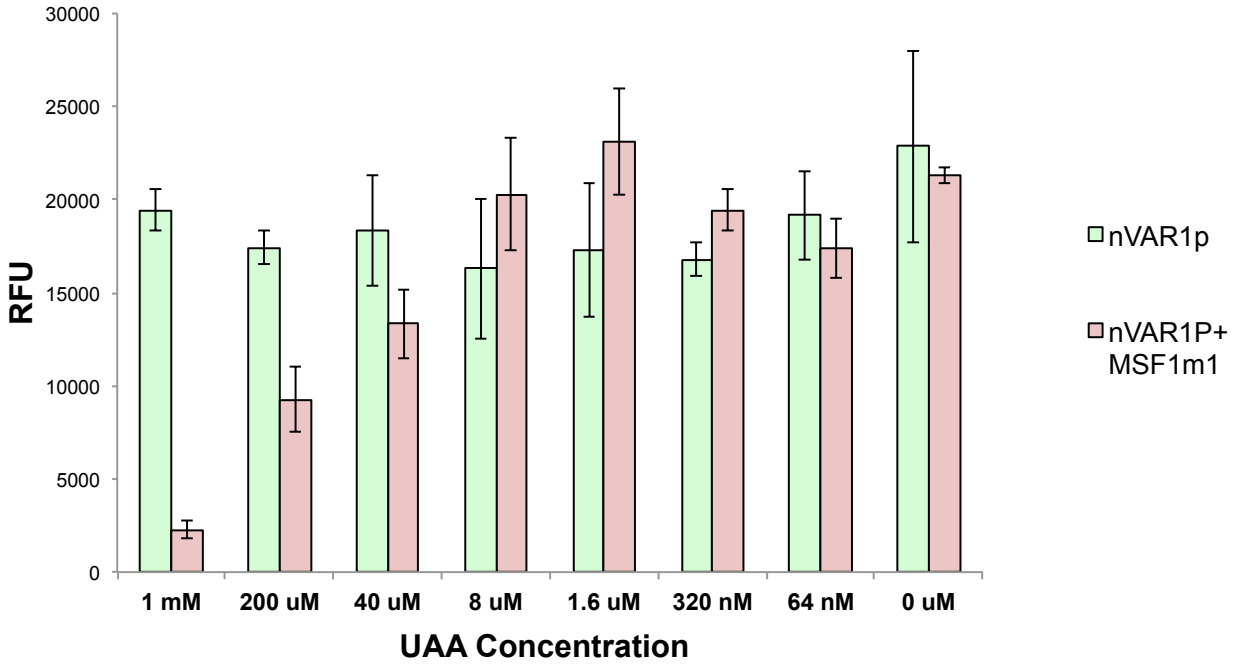


Figure 3.2 | Fluorescence Sensitivity to UAA (2) in sfGFPmito Strain
UAA **2** is titrated into galactose+raffinose growth medium for sfGFPmito strains containing either nVAR1p and MSF1m1, or nVAR1p alone. Bars represent the average of three cytomtery histogram medians.

Proceeding to direct protein characterization, we prepared two batches of purified mitochondria of the dual nVAR1p/MSF1m1 expression strain, with and without **2** (Figure 3.3). In both preparations, GFP was detectable via immunostaining in the +UAA condition, suggesting nVAR1p's absence from the GFPm-3 experiments as a cause for GFP disappearance. While there is a decrease in ATP2-normalized GFP signal in the presence of unnatural, the fold-difference in band intensity has been calculated to be smaller than the difference in fluorescence measured via cytometry. This suggests that while a decrease in overall GFP content can explain some of fluorescence loss, it can't explain the degree of the drop observed. In this reasoning, there is the assumption quantity of a fluorescent protein linear correlates with measured fluorescence intensity, which has been shown previously [3,4].

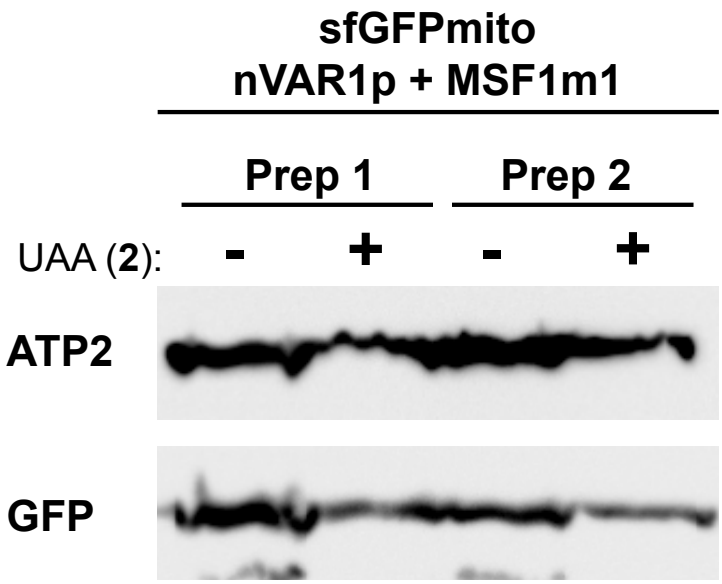


Figure 3.3 | Immunostaining of sfGFPmito with nVAR1p and MSF1m1
Western blot of sfGFPmito preps, with and without 1 mM **2**. ATP2 serves as protein quantity control.

3.2 Expression Refinements to the Platform

We intended to test this GFP for the appropriate mass shift via mass spectrometry, but limitations on protein expression motivated us to search for optimizations, both genetic and otherwise. We were aware that the native genes encoded on the mtDNA have dependencies on translational activators and posttranscriptional processors for proper expression, notably *Cox3* requiring the Pet494p, Pet122p, and Pet54p proteins [5,6]. Given that sfGFPmito is encoded with the 5' UTR of *Cox2*, looking for proteins known to regulate COX2 expression was our next step.

This search produced Pet111p, a *Cox2*-specific translational activator first identified through petite colony phenotype screens [7,8]. This protein is thought to be important for pre-translational processing of the *Cox2* transcript, as COX2 is normally membrane-bound. Previous work has demonstrated that Pet111p is a limiting factor in

COX2 expression via the use of a diploid heterozygous mutant [9]. Promisingly, overexpression of Pet111p was shown to increase expression of a mtDNA-integrated Arg8 enzyme. While additional work suggests that Pet111p overexpression can reduce expression of other cytochrome oxidase genes, thus disrupting the stoichiometry of Complex iV [10], we are not very interested in the fidelity of the other mitochondrially translated genes, since there will be mito-wide UAA incorporation regardless.

We cloned Pet111p into a 2-micron expression cassette under the control of TDH3pr, RPL18Bpr, and REV1pr, a strong, medium, and weak promoter, respectively [4]. Transforming these constructs into both the sfGFPmito strain and its base strain, W303, we used flow cytomtery to check for changes in fluorescence (Figure 3.4). There is a very striking increase in sfGFPmito production upon overexpression of Pet111p, with the background level of W303 not varying at all. Moreover, this increase is dose-dependent, as evidenced by fluorescence intensities matching the promoter strengths.

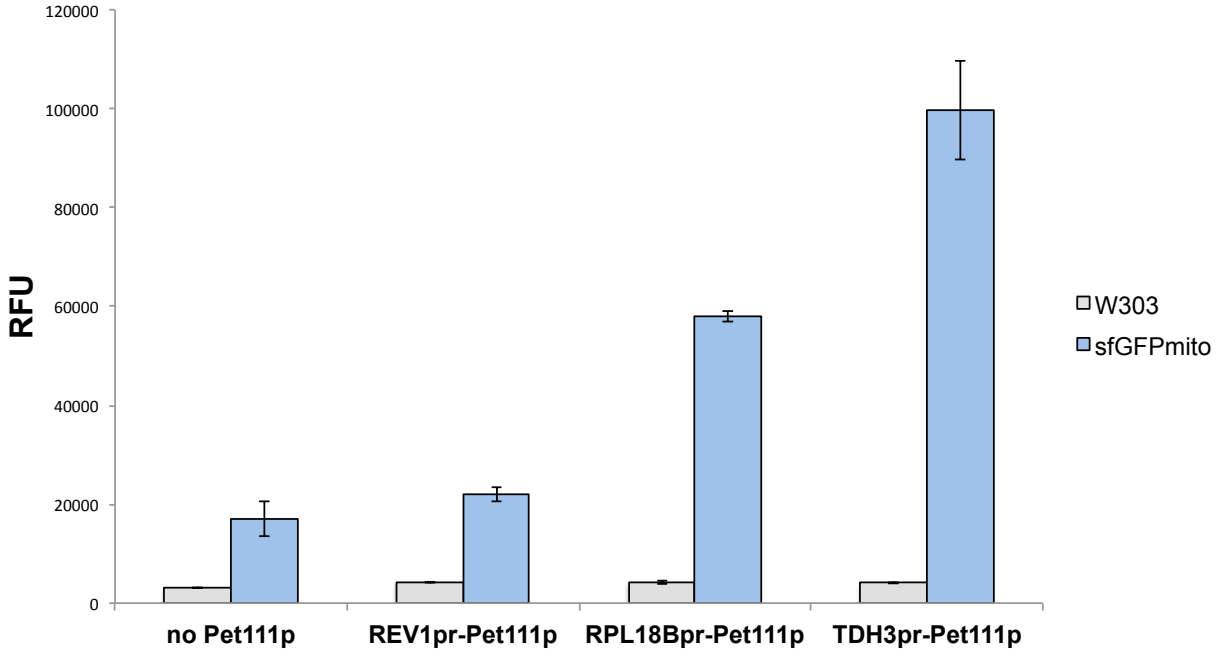


Figure 3.4 | Pet111p Acts as a Translational Activator of sfGFPmito
The effect of different expression levels of Pet111p on sfGFPmito fluorescence. Bars represent the average of three cytometry histogram medians.

In a similar vein of optimizations to expression, we next sought ways of increasing mtDNA copy number with the notion that more abundant copies would lead to more simultaneous transcriptional events, and ultimately more protein [11]. The ribonucleoreductase RNR1 has been implicated in playing a role in controlling copy number, based on reports that overexpression can complement for activity-reduced mutants of mitochondrial DNA polymerase MIP1 [12], and boost mtDNA copy numbers to about 2x WT levels [13]. In addition to potentially increasing protein expression, increased DNA copy numbers has the appeal of reinforcing genome stability, which could become relevant in +UAA growth conditions.

A recent genome-stability screen of the *S. cerevisiae* knockout collection revealed both RNR1 and its less understood paralog, RNR3, as genes whose overexpression can upregulate mtDNA copy number, with the latter being more potent [14]. Based on this, we opted to use RNR3 to tests for boosts in copy number, initially on a 2-micron plasmid under RNR1-level expression, as was done in MIP1 mutant complementation studies [12]. Adding this plasmid to both the sfGFPmito strain and W303, we observed no significant increase in expression (Figure 3.5). Upon swapping out the RNR1 promoter (RNR1pr) for the significantly stronger PGK1 promoter (PGK1pr) [4], we observed a >2x increase in fluorescence. While addition of the 2-micron RNR3 plasmid did seem to boost the W303 background fluorescence, no further increases in background were observed when switching to PGK1pr.

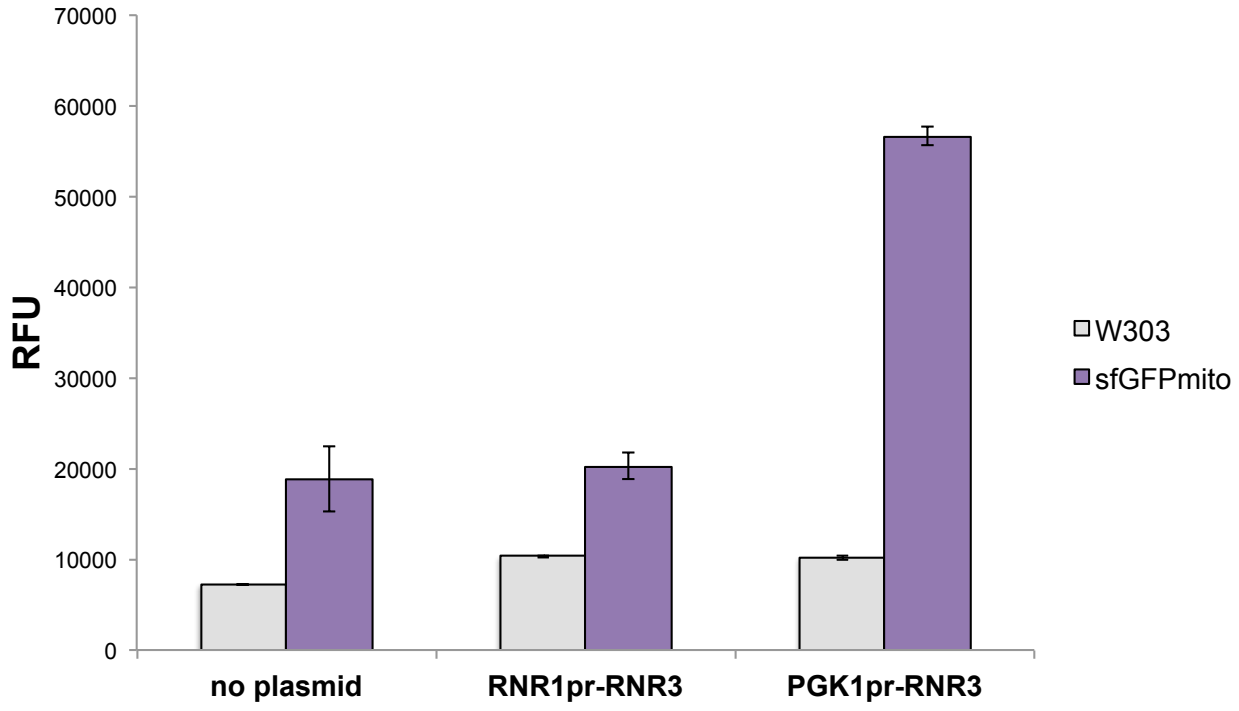


Figure 3.5 | RNR3 Overexpression Increases sfGFPmito Signal
W303 and sfGFPmito strains transformed with either RNR1pr or PGK1pr driven RNR3 on a 2-micron plasmid. Bars represent the average of three cytometry histogram medians

Even if an unnatural-incorporated reporter is highly expressed, there is the chance that introduction of nonnative amino acids accelerates protein degradation, as quality control proteases are known to degrade proteins co-translationally upon the ribosome-mediated detection of errors [15,16]. In the event that this regulation proves prohibitive to accruing significant amounts of incorporated sfGFP, it would be useful to have mechansisms for modulating the quality control machinery. To this end, we focused on the Lon protease, known to be the major quality control protease in mitochondria [17,18]. A search for Lon protease inhibitors revealed the triterpenoid CDDO and its derivative, CDDO-Me, both characterized in mammalian mitochondria [19]. Though never tested in yeast, the high sequence conservation of the Lon protease throughout nature motivated us to try the inhibitor [20].

We added 1 μM of CDDO-me to sfGFPmito cultures expressing MSF1m1, with and without UAA **2**. After purifying mitochondria, we normalized the samples by mitochondrial density, as the intensity of the imported control ATP2 could vary in the presence of a protease inhibitor. Immunostaining for ATP2 and GFP revealed accumulation of both proteins in the presence of CDDO-Me, even when UAA is added (Figure 3.6). While the inhibitor does seem to be effective in accumulating proteins, the cultures grow noticeably slower, which is in accordance with existing knowledge on the effect of Lon interference [20]. Thus, we opted to not use this growth condition unless there was difficulty in detecting incorporated sfGFP with other methods.

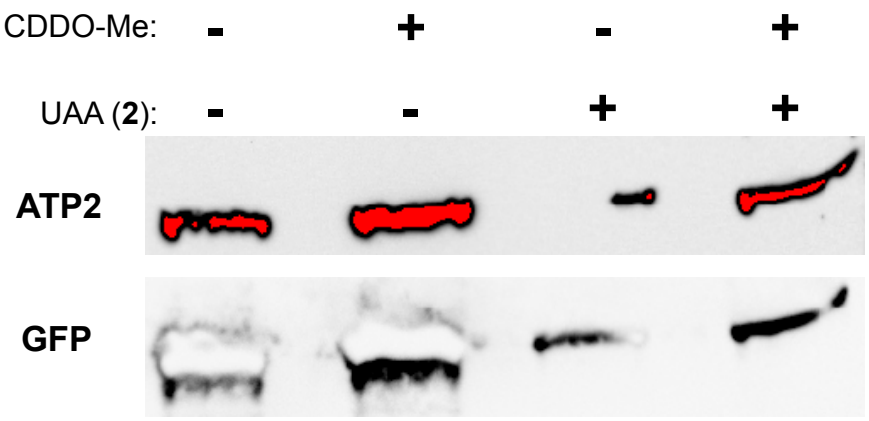


Figure 3.6 | CDDO-Me Leads to Protein Accumulation in sfGFPmito + MSF1m1 CDDO-Me added to cultures at a concentration of 1 μM, UAA **2** added at 1 mM. Red pixels indicate signal saturation of the detector.

3.3 Click Chemistry Labeling of Unnatural Amino Acids In Mitochondria

To showcase the potential of our mitochondrial translation system as an investigative tool, we sought to label amino acids in the mitochondrial via a bioorthogonal labeling chemistry. The “click” chemistry reaction has been demonstrated previously in various expanded genetic code contexts via the use of UAAs bearing either azide or alkyne functional groups [21,22]. A reaction between these amino acids and a fluorophore bearing the matching labeling chemistry allows for targeted fluorescent labeling of a newly synthesized protein of interest [23].

By adapting the methods in Yuet *et al.* for yeast, we created a fluorescent microscopy workflow to label UAAs **1** and **3** in the mitochondria (Figure 3.7). In this scheme, the mutants of the MSY1 and MSF1 synthetase are used according to their experimentally determined substrate compatibilities, as discussed in Chapter 2. For both amino acids, the click reaction appends an azide-conjugated Alexa 488 dye to the alkyne-containing R group, which then fluoresces green upon excitation with blue light.

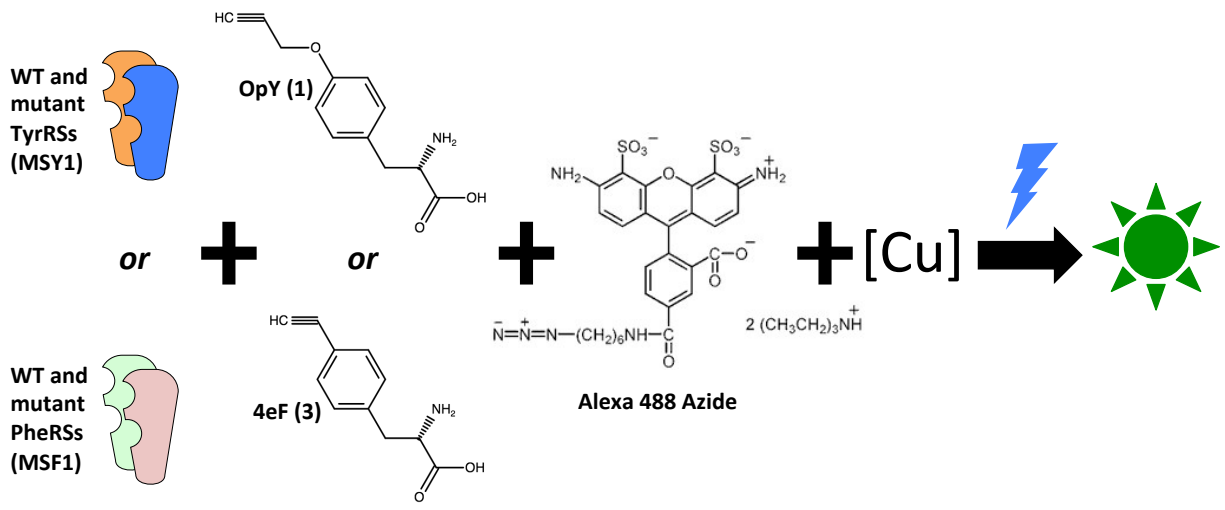


Figure 3.7 | Schematic for Azide-labeling of UAAs in Mitochondria
Amino acids 1 and 3 are paired with the WT and mutants variants of the MSY1 and MSF1 synthetases, respectively. In the presence of copper, the alkyne groups react with an azide-conjugated Alexa 488.

We compared the fluorescent output of wild type vs. mutant aaRS-expressing yeast exposed to equivalent concentrations of either **1** or **3**. After cells were grown overnight with UAA, they were fixed and click-labeled with the Alexa 488 fluorophore. In the resulting images, we observed clear mitochondrial puncta only when cells were expressing the mutant synthetases (Figure 3.8). This morphology is commonly observed in mitochondria, along with rod and network structures [24]. The diffuse signal observed in both the WT and mutant conditions can be attributed to loose fluorophore. As expected, cells grown in UAA-free media did not display fluorescent puncta, even when expressing the mutant aaRSs (data not shown).

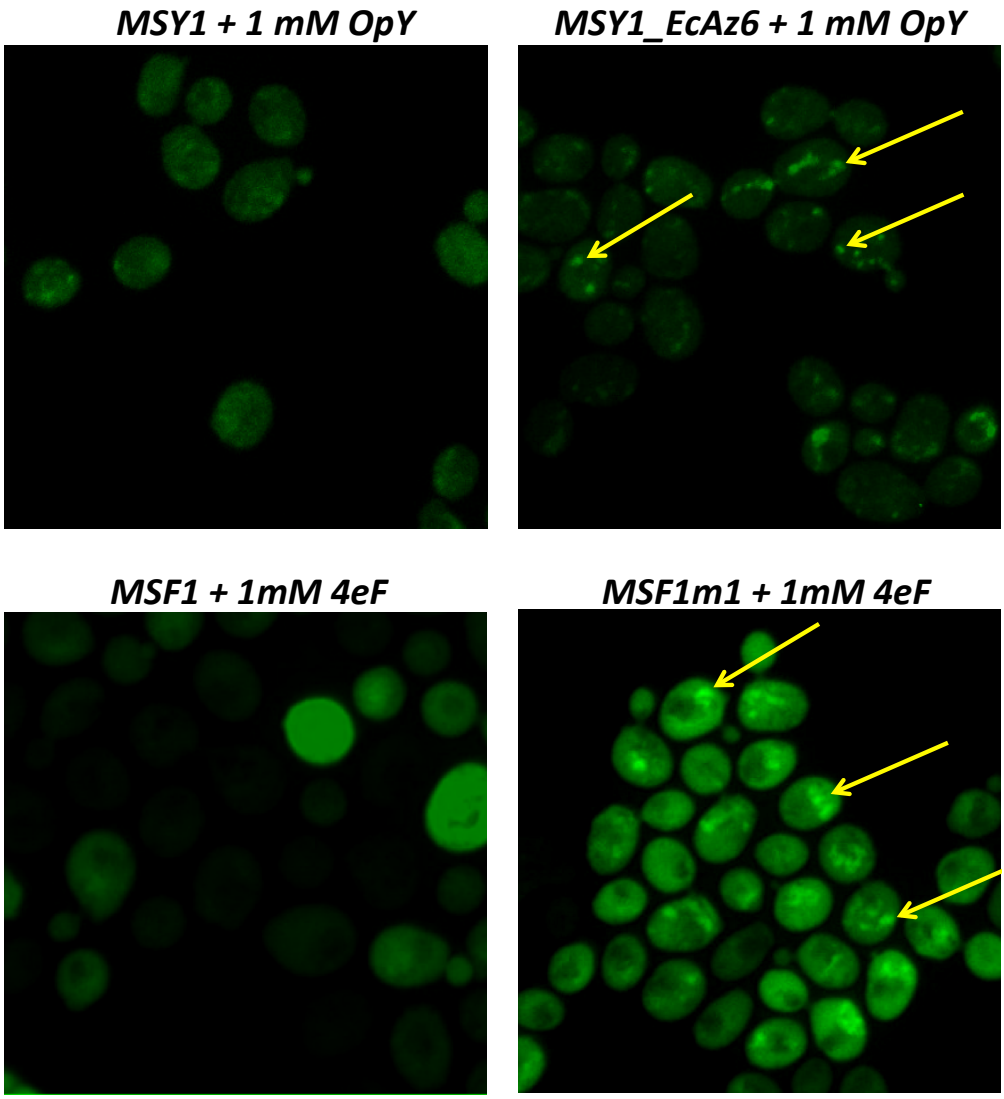


Figure 3.8 | UAA Click Microscopy in *S. Cerevisiae* Mitochondria
OpY and 4eF refer to UAAs **1** and **3**, respectively. Cells were grown overnight in the presence of UAA, washed three times with PBS, then stained with Alexa 488 and washed again three times. Arrows designate examples of mitochondrial puncta. All images captured with the same exposure and background subtracted with the same background level.

We next investigated whether the putative mitochondrial labeling observed in yeast was extensible to mammalian cells. Implementing the same synthetase engineering strategy described in Chapter 2, we added the T271G mutation in the human phenylalanyl tRNA-synthetase, FARS2 (designated FARS2m1). We then implemented the same click microscopy approach to HEK293T cells with either a CMV

promoter-driven FARS2m1 or an empty vector. Based off of some previously-observed cross-activity of the PheRS mutant with **1**, we opted to use both UAAs **1** and **3** with the same synthetase.

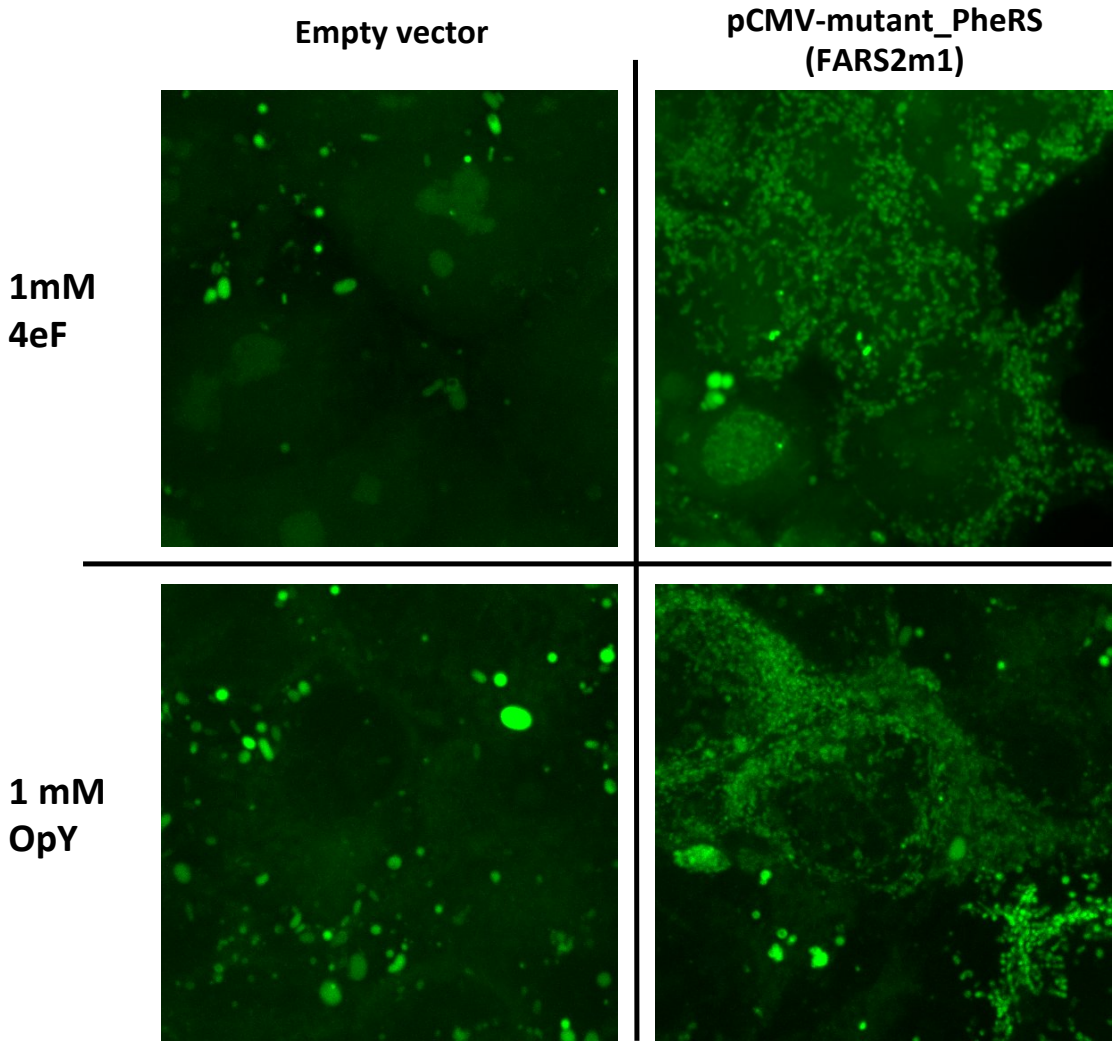


Figure 3.9 | UAA Click Microscopy in Human Mitochondria
OpY and 4eF refer to UAAs 1 and 3, respectively. Cells were grown overnight in the presence of UAA, washed three times with PBS, then stained with Alexa 488 and washed again three times. All images captured with the same exposure.

As observed in yeast, the human cell microscopy showed mitochondrial signal accumulation only when FARS2m1 was overexpressed (Figure 3.9). The click-labeling results were shown to be mutation-dependent in both cellular models, which together

with previous respiration and fluorescence interference data strongly suggest that mitochondrial tRNA charging with UAAs is indeed mediated by mutant synthetases. While it is possible that the signal observed in the click labeling experiments can be partially attributed to UAAs inside of protein polymers, the fluorescent microscopy readout makes it impossible to differentiate between UAAs in a peptide and UAAs on charged mitochondrial tRNAs. The accumulation of these tRNAs, which are transcribed in the mitochondria and are not known to be exported, is likely to be the major contributor to the fluorescent signal, based on the preexisting difficulties with UAA-containing protein expression prior to any optimization.

3.4 References

[1] Suhm, T. *et al.* A novel system to monitor mitochondrial translation in yeast. *Microbial cell* **5**,158-164 (2018)

[2] Sanchirico M. *et al.* Relocation of the unusual VAR1 gene from the mitochondrion to the nucleus. *Biochem Cell Biol*. **73**, 987-995 (1995)

[3] Soboleski, M. R., Oaks, J., Halford, W. P. Green fluorescent protein is a quantitative reporter of gene expression in individual eukaryotic cells. *The FASEB journal*, ***19***, 1-20 (2005)

[4] Lee, M.E. *et al.* A highly characterized yeast toolkit for modular, multipart assembly. *ACS synthetic biology* **4,** 975-986 (2015)

[5] Fox, T.D. Mitochondrial protein synthesis, import, and assembly. *Genetics* **192**, 1203-34 (2012)

[6] Steele, D. F., Butler, C. A., Fox, T. D. Expression of a recoded nuclear gene inserted into yeast mitochondrial DNA is limited by mRNA-specific translational activation. *Proc. Natl. Acad. Sci. U.S.A.* ***93***, 5253-5257 (1996)

[7] Jones, Julia L *et al.* Yeast mitochondrial protein Pet111p binds directly to two distinct targets in *COX2* mRNA, suggesting a mechanism of translational activation. *The Journal of biological chemistry* **294**, 7528-7536 (2019)

[8] Dekker, P. J. T., Papadopoulou, B., Grivell, L. A. Properties of an abundant RNA-binding protein in yeast mitochondria. *Biochimie* **73**, 1487-1492 (1991)

[9] Green-Willms, N. S. *et al.* Pet111p, an Inner Membrane-bound Translational Activator That Limits Expression of the Saccharomyces cerevisiae Mitochondrial Gene COX2. *Journal of Biological Chemistry* **276**, 6392-6397 (2001)

[10] Fiori, A., Perez-Martinez, X., Fox, T. D. Overexpression of the COX2 translational activator, Pet111p, prevents translation of COX1 mRNA and cytochrome c oxidase assembly in mitochondria of Saccharomyces cerevisiae. *Molecular microbiology **56***, 1689-1704 (2005).

[11] Ravikumar, Arjun, et al. "Scalable, continuous evolution of genes at mutation rates above genomic error thresholds." *Cell* 175.7 (2018): 1946-1957.

[12] Lecrenier, N. & Françoise F. Overexpression of the RNR1 gene rescues Saccharomyces cerevisiae mutants in the mitochondrial DNA polymerase-encoding MIP1 gene. *Molecular and General Genetics MGG* **249**, 1-7 (1995)

[13] Taylor, S. D. *et al.* The conserved Mec1/Rad53 nuclear checkpoint pathway regulates mitochondrial DNA copy number in Saccharomyces cerevisiae. *Molecular biology of the cell* **16**, 3010-3018 (2005)

[14] Puddu, F. *et al.* Genome architecture and stability in the *Saccharomyces cerevisiae* knockout collection. *Nature* **573,** 416–420 (2019)

[15] Gandin V & Topisirovic I. Co-translational mechanisms of quality control of newly synthesized polypeptides. *Translation* **2** (2014)

[16] Vazquez-Calvo, C. *et al.* The basic machineries for mitochondrial protein quality control. *Mitochondrion* **50**, 121-131 (2020)

[17] Van Melderen, L. & Aertsen, A. Regulation and quality control by Lon-dependent proteolysis. *Research in microbiology* **160**, 645-651 (2009)

[18] Bender, T. *et al.* Mitochondrial enzymes are protected from stress-induced aggregation by mitochondrial chaperones and the Pim1/LON protease. *Molecular biology of the cell* **22**, 541-554 (2011)

[19] Bernstein, S. H. *et al*. The mitochondrial ATP-dependent Lon protease: a novel target in lymphoma death mediated by the synthetic triterpenoid CDDO and its derivatives. *Blood, The Journal of the American Society of Hematology* **119**, 3321-3329 (2012)

[20] Pinti, M. *et al.* Emerging role of Lon protease as a master regulator of mitochondrial functions. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1857**, 1300-1306 (2016)

[21] Deiters, A. *et al.* Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *J Am Chem Soc.* **125**,11782-11783 (2003)

[22] Li, X. & Liu, C.C. Biological Applications of Expanded Genetic Codes. *ChemBioChem* **15**, 2335-2341 (2014)

[23] Yuet, K. P. *et al.* Cell-specific proteomic analysis in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.* **12**, 2705-2710 (2015)

[24] Leonard A.P. *et al.* Quantitative analysis of mitochondrial morphology and membrane potential in living cells using high-content imaging, machine learning, and morphological binning. *Biochim Biophys Acta*. **1853**, 348-360 (2015)

4. Biochemical Characterization of the Mitochondrial Expanded Genetic Code Platform

4.1 Validating an Affinity Tag on Mitochondrial sfGFP

After combining the previously described expression-boosting modifications, sans CDDO-Me, into one strain labeled sfGFPmito-v3, it became clear that an affinity purification pipeline was needed for unbiased recovery of unnatural-containing peptides. Our prior attempts at GFP purification included immunoprecipitation (IP) via an anti-GFP antibody, which could possibly be biased against destabilized UAA-containing copies, and two step liquid chromatography, which could discriminate against the differing physical properties of unnatural-containing peptides. When engineering sfGFPmito-v3, it was revealed that the base strain's sfGFP was actually integrated into the mitochondrial genome with a C-terminal HIS-tag, a detail not disclosed in the manuscript describing the strain's construction [1]. With an affinity tag lacking phenylalanine, we were presented with a purification method with reduced bias, though the C-terminus location ensured that all recovered proteins would be fully translated sequences.

We first verified that our sfGFPmito-v3 strain had detectable protein via anti-GFP and anti-HIS staining (Figure 4.1). With the UAA-containing cultures yielding protein quantities seemingly sufficient for downstream analysis, we proceeded to an IP approach utilizing an anti-HIS antibody for targeted capture. This method, while limited in its total yield compared to HIS-affinity resins, offers the superior specificity of an antibody, thus reducing nonspecific background signal.

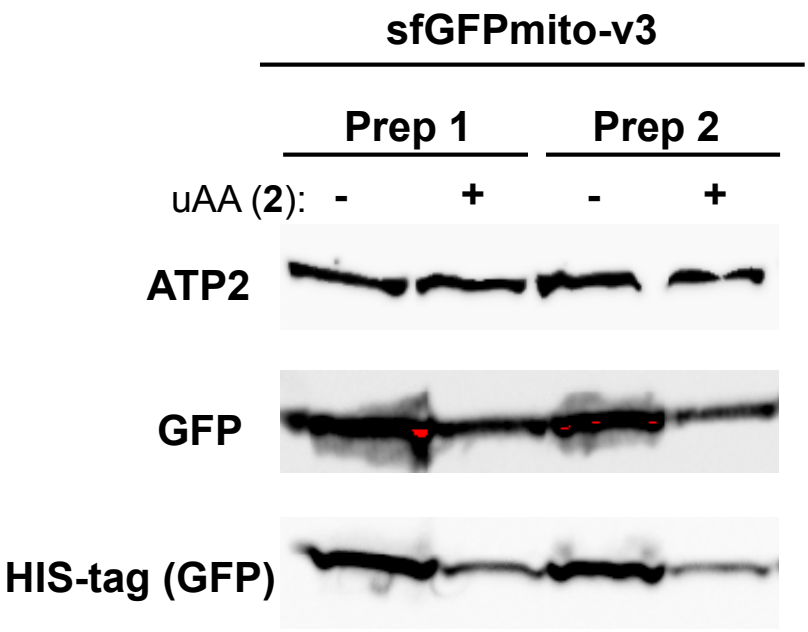


Figure 4.1 | sfGFPmito-v3 Shows Anti-HIS Signal with UAA
sfGFPmito-v3 contains MSF1m1, nVAR1p, Pet111p, and RNR3 overexpression. UAA used at 1 mM. Prep 1 and 2 refer to two replicates.

We first analyzed HIS-IP eluates of mitochondrial preps +/- UAA with MALDI-TOF MS, where despite poor signal intensity, a peak at the correct sfGFP size was observed, with some leeway for the limited mass resolution of MALDI in this mass-range (Appendix 1). A peak of about the same size was also seen in the sample derived from +UAA mitochondria, although there is a noticeable broadening towards higher masses. While the wider peak could be an indication of a mixture between incorporated and unincorporated copies of sfGFP, the MALDI-based approach lacks the granularity required to identify exact mass shifts. With the possibility of partial UAA incorporation patterns, it became clear that a fragmented protein analysis pipeline would need to be developed and validated.

4.2 Developing an Unnatural Peptide Analytical Pipeline with a Bacterial Control

We opted to model our analytical pipeline on the Tirrell Lab's methods, namely the tryptic digest to LC-MS approach [2,3]. In order to confirm effective protein purification and validate the mass shift detection capability, a bacterial control strain from Yuet *et al.* with an inducible copy of the PheRS T251G single mutant (PheRSm1) was used. By introducing a separately inducible HIS-tagged sfGFP into the strain, the production of our target protein could be restricted to only when the system is incorporation-capable and exposed to UAA. We can ensure this by adhering to a two-induction protocol where a pool of PheRSm1 is first produced, the cells are transferred to UAA-containing media, and then sfGFP is translated (Figure 4.2). This approach minimizes the quantity of unincorporated sfGFP molecules while also limiting toxicity from proteome-wide UAA incorporation [3].

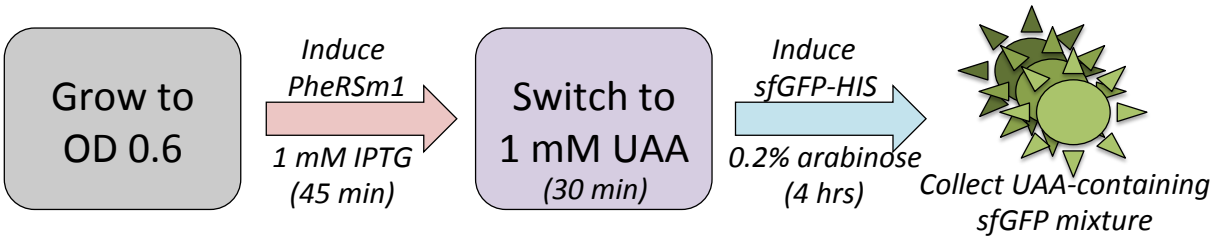


Figure 4.2 | Two Induction Scheme for Bacterial UAA Incorporation
Cells are first grown to log phase, and are washed 3 times in 0.9% NaCl before transferring to UAA-containing media. Resultant sfGFP mixture is potentially comprised of varying degrees of incorporation (up to 12 sites).

We applied this induction method to cells exposed to phenylalanine (WT), 4-iodo-phe (**2**), 4-bromo-phe (**4**), or 4-azido-phe (**5**). These UAAs were chosen based on ease-of-access and experimental demonstration of their compatibility with the T251G mutant, as described in Chapter 2 and in previous literature [3]. After harvesting the cells, immunoblots confirmed consistent expression of the PheRSm1 mutant between the four

cultures (Appendix 2). Consistent with the yeast mitochondrial experiments, sfGFP expression was decreased in samples containing UAAs, though the overall protein yield from arabinose induction still far exceeded anything observed in the mitochondrial platform. Additionally, a small amount of sfGFP leak was observed in the un-induced sample, though the faint band compared to the induced +UAA conditions suggests that this residual sfGFP is a minority product by the end of the induction protocol.

To further compare the bacterial control to the mitochondrial system, we measured the OD-normalized fluorescence of the induced cultures (Appendix 3). As this bacterial platform had been previously validated as a means of producing UAA-incorporated protein, we sought to determine possible effects of UAA insertion into sfGFP [2,3]. Indeed, we observed a markedly reduced cellular fluorescence in the UAA-containing samples, with levels comparable to the un-induced culture. In order to control for the effects of reduced sfGFP expression in the presence of the UAAs, we normalized the fluorescence with the immunoblot band intensities from Appendix 2. Even with this adjusted metric, we observe a reduction in cellular fluorescence, suggesting that the drop in RFU can't be explained solely by the decreased expression of sfGFP (Appendix 4).

Next, we designed a protein purification-to-analysis pipeline capable of parsing sfGFP molecules with variable degrees of incorporation. To this end, we subjected the cell lysates to cobalt affinity resin purification. Cobalt's reduced affinity for the HIS tag cuts down on the nonspecific binding observed, making it more amenable to eventual application for mitochondria, where off-target protein is exceedingly abundant relative to the tagged-sfGFP [4]. Per manufacturer's recommendation, we carried out the

purification in denaturing concentrations of urea to further reduce nonspecific binding. These adjustments yielded purified samples where the two HIS-tagged proteins, PheRSm1 and sfGFP, were the predominant proteins observed via PAGE (Appendix 5).

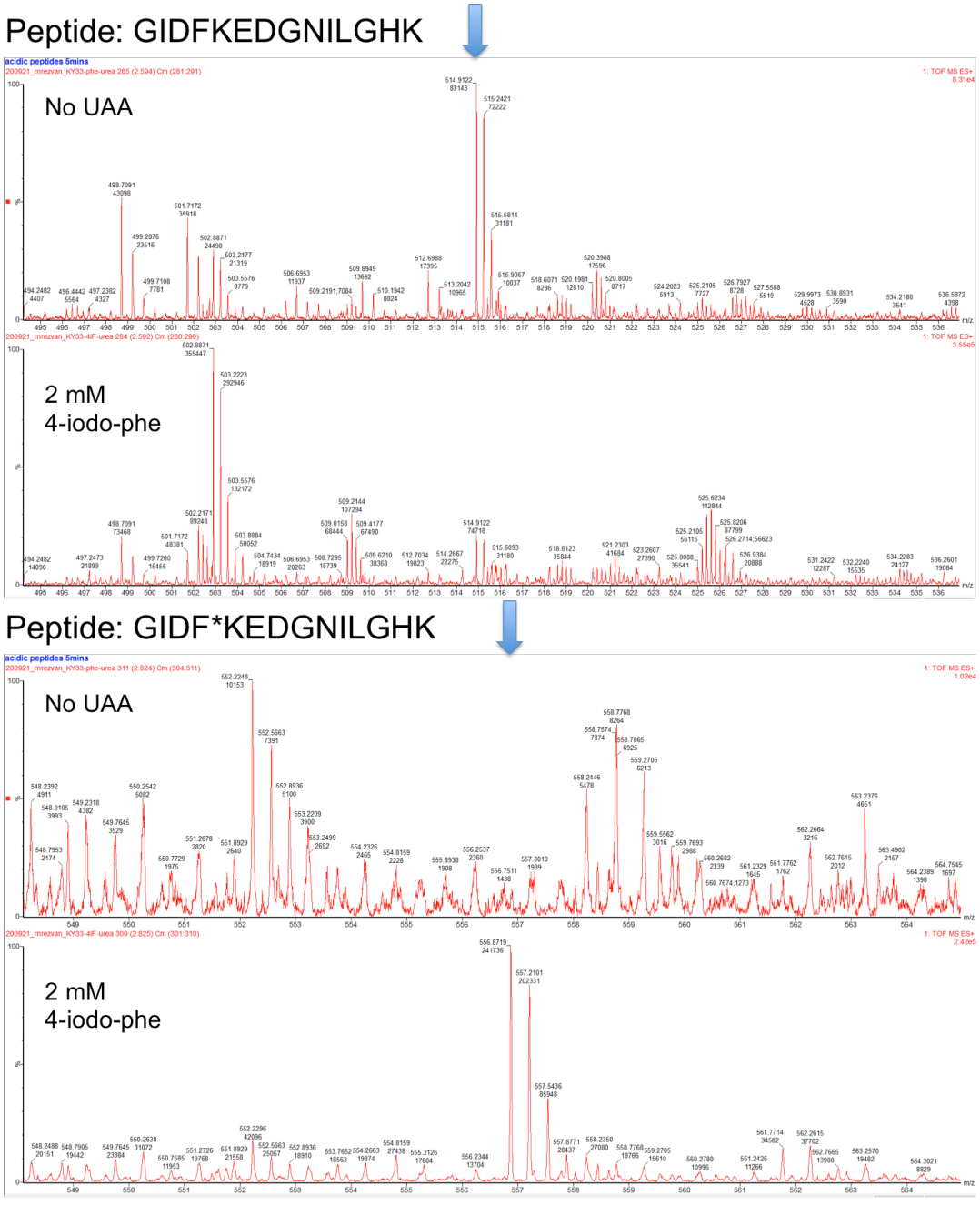


Figure 4.3 | 4-iodo-phe Incorporation in E. coli sfGFP Peptide GIDFKEDGNILGHK Mass spectra captured with LC-ESI-TOF after tryptic digest. Arrows designate target peak locations. F* represents 4-iodo-phenylalanine.

For proteomic analysis of the purified sfGFP, we opted for LC-MS of tryptic digests, based upon existing methods involving multi-incorporation products [2,3,5]. After LC-ESI-TOF of digested peptides, the resultant peaks were analyzed via BioPharmaLynx, and then manually validated in the raw mass spectra. For all three UAA conditions, at least 3 distinct peptides were identified with the expected mass shifts. One example peptide comparison for UAA **2** shows that the mass-shifted peak is unique to the sampled prepared with UAA, as expected (Figure 4.3). The other peptide comparisons show the same specificity, though the WT peptide mass peak is sometimes observable in the +UAA sample (Appendix 6-7). This same set of comparisons was applied to UAAs **4** and **5**, yielding identifiable unique mass shifts as expected for those phenylalanine analogs (Appendix 8-10). Expected mass shifts for each unnatural peptide were searched against the data from the other two analogs, yielding no matches as expected. Additionally, unnatural-containing peptides consistently eluted later off of the LC step, as expected with these modified phenylalanines [2,3].

4.3 MS Analysis of Purified Mitochondrial Protein

With a purification-to-MS pipeline capable of detecting multiple incorporations with distinct UAAs, we returned to our sfGFPmito-v3 yeast mitochondrial expression platform. Preparing cultures with the aforementioned three UAAs, we first performed an anti-HIS IP on cultures grown with 1 mM **4** or **5**, as well as a no-UAA control (Figure 4.4). While the capture antibody's heavy and light chains were co-eluted, a distinct

sfGFP band was identifiable in the mitochondria prepared with **4**. The **5** sample didn't yield recoverable sfGFP, and thus wasn't subjected to downstream digest and analysis.

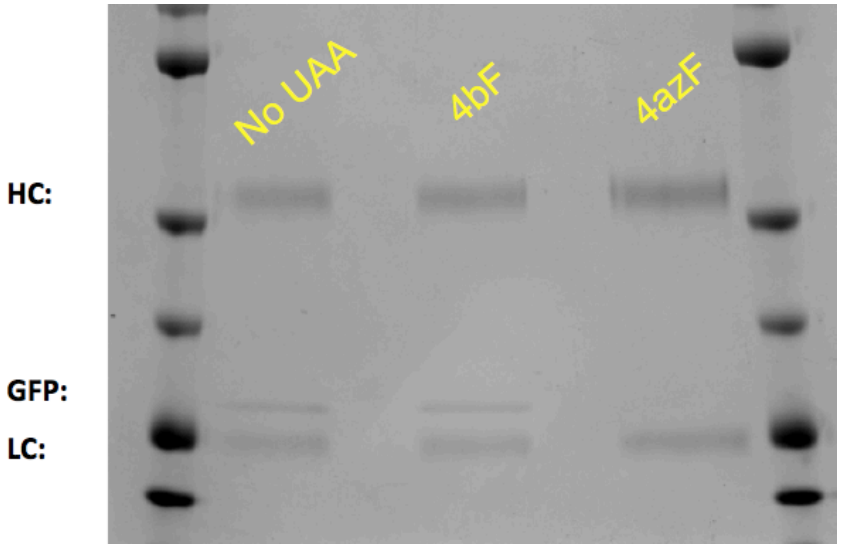


Figure 4.4 | IP Eluates of HIS-sfGFP from 4-bromo-phe Mitochondria
Polyacrylamide gel stained with Coomassie R250. IP loading normalized by anti-GFP immunoblot signal, except in the case of 4azF. HC and LC refer to heavy chain and light chain of the capture antibody, respectively. 1 mM of UAA used in the mitochondrial preparations. 4bF: **4**, 4azF: **5**.

Upon tryptic digest and LC-ESI-TOF, we observed peaks corresponding to unincorporated peptide masses in both the +/- **4** conditions, at comparable intensities (Figure 4.5). Furthermore, we did not detect any mass shifted peptides, even for equivalent target sequences that were of high abundance in the bacterial controls. This finding, while alarming, does not unequivocally point to the mitochondrial incorporation system not working. If the IP antibody used captures only a subset of the total sfGFP in solution, copies of the protein with UAA incorporation could be selected against due to reduced stability and potential indirect interference with the antibody-binding site. Thus, we sought to apply the cobalt resin purification strategy to our subsequent mitochondrial preparations in order to capture all available sfGFP in solution.

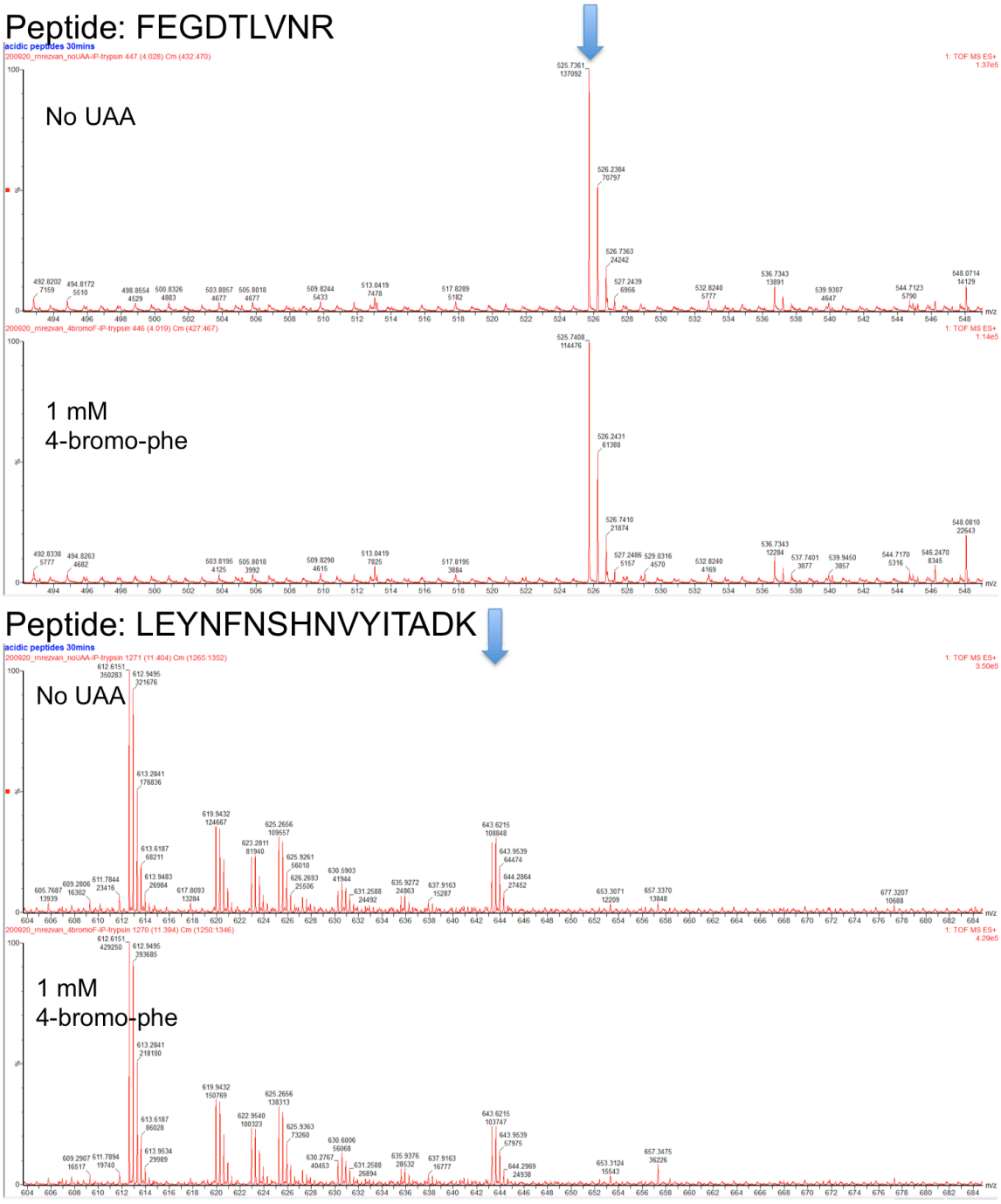


Figure 4.5 | No 4-bromo-phe Incorporation Observed in sfGFPv-3 IP Samples Mass spectra captured with LC-ESI-TOF after tryptic digest. Arrows designate target peak locations.

To minimize off target resin binding in mitochondrial samples, we applied the cobalt HIS purification protocol used for the bacterial incorporation controls, with

denaturing 8 M urea. This optimized approach was sufficient for recovering the mitochondrial sfGFP as the dominant species in solution, even for the sample grown in 1 mM of **2** (Appendix 11). For this purification it was confirmed via the absence of signal on a follow-up immunoblot that all of the sfGFP in the original sample was either washed out or captured in the eluate, the majority being the latter (data not shown). This confirmation, plus the presence of Coomassie-visible sfGFP bands, gave us confidence in the comprehensiveness of the potential sfGFP peptide readout.

Upon MS analysis of the sfGFPmito-v3 no UAA and **2** samples, it became clear that there were no high quality peaks representing tryptic peptides with incorporation. Instead, all the major matching peptides recovered corresponded to sequences with natural phenylalanine, in both the control and UAA samples (Figure 4.6). The limitations in the system might be explained by some of the key differences between the mitochondrial platform and the bacterial controls, as they both seem to facilitate mutation-mediated tRNA charging with UAAs. In particular, the recovered sfGFP in the mitochondrial experiments can potentially be attributed to proteins synthesized prior to UAA exposure, or mid-exposure using the on-hand natural phenylalanine. The bacterial control is a phenylalanine auxotroph, so controlling the media composition can greatly reduce the presence of intracellular phenylalanine [2,3]. Additionally, having both the synthetase mutant and the gene of interest under inducible control allows for the maximization of incorporated to unincorporated product. It could be the case that the mitochondrial system is indeed incorporating, but underneath the detection limit of our pipeline. Further analysis and optimization is needed to surmise the exact bottle necks of the existing system.

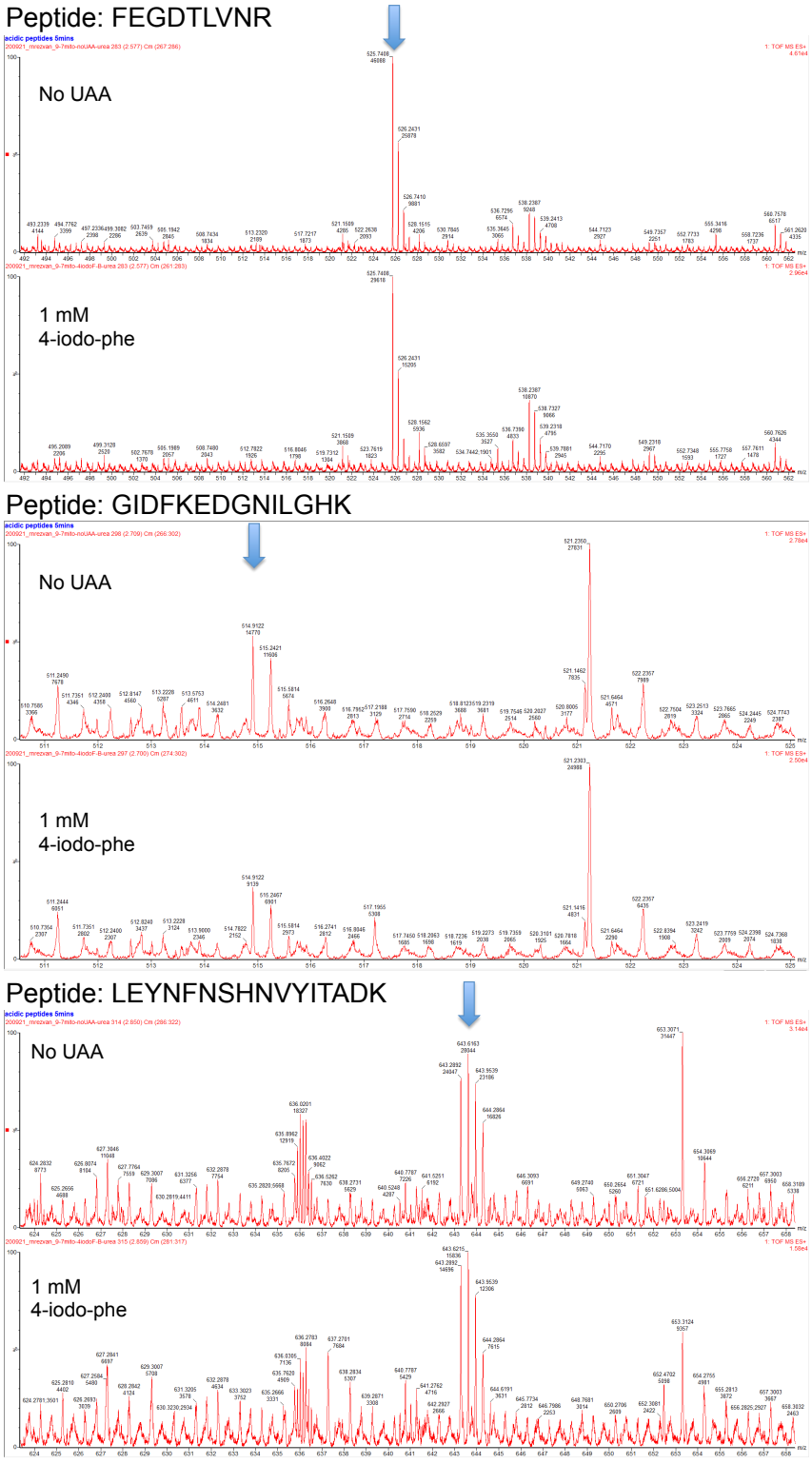


Figure 4.6 | No 4-iodo-phe Incorporation Observed in sfGFPmito-v3 Cobalt Samples
Mass spectra captured with LC-ESI-TOF after tryptic digest. Arrows designate target peak locations.

4.4 References

[1] Suhm, T. *et al.* A novel system to monitor mitochondrial translation in yeast. *Microbial cell* **5**,158-164 (2018)

[2] Kirshenbaum, K., Carrico, I.S., Tirrell, D.A. Biosynthesis of Proteins Incorporating a Versatile Set of Phenylalanine Analogues *ChemBioChem* **3**, 235-237 (2002)

[3] Yuet, K. P. *et al.* Cell-specific proteomic analysis in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.* **12**, 2705-2710 (2015)

[4] Andersen, K.R., Leksa, N.C. and Schwartz, T.U. Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins 81, 1857-1861 (2013)

[5] Datta, D. *et al.* A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins. *J Am Chem Soc.* **124**, 5652-3 (2002)

5. Discussion and Future Directions

5.1 Further Developing Mitochondrial Unnatural Polymer Synthesis

In its current state, the mitochondrial expanded genetic code platform does not produce enough incorporated target protein to be detectable via an LC-ESI-MS pipeline. It potentially is the case that, as is, the system is incorporating trace amounts of UAA that would need to be measured another way. To that end, submitting the tryptic digests for amino acid analysis can give a more exact information on the percent of phenylalanine residues throughout a target protein's sequence that have been replaced by an analog [1,2]. In the event that incorporation is indeed occurring in trace amounts, a series of potential optimizations can be explored to either improve protein expression or system stability.

In the context of the bacterial proteome-wide labeling, the sequential induction of protein expression plays a huge role in the quality and quantity of the final incorporated-product [2]. In earlier studies, the Tirrell Lab's methods involved the timing of simultaneous expression for both the mutant aaRS and the gene of interest (GOI) [1]. By waiting until a culture's log phase to express both of these key components in fresh UAA-containing media, errant unnatural free-GOI expression from the initial outgrowth is avoided. However, this protocol gives the cells a limited window to generate functional copies of the synthetase mutant before the natural amino acid is completely spent. Additionally, that same intracellular pool of the native amino acid could be allocated towards copies of the GOI, which would reduce the total level of incorporation observed. This was

rectified with the two-step system, where the synthetase is first generated in the UAA-free phase of the culture growth, followed by a UAA acclimation period [2].

Currently the mitochondrial orthogonal translation platform is free of induction steps, mainly because of transcriptional activation limitations on the mtDNA. The closest approximation, however, could potentially be taking advantage of the glucose-repressible expression of mitochondrially-localized genes [3,4]. By first growing large cultures of cells in glucose media, expression of mito-encoded genes would be repressed. Then cells could be washed and transferred to a non-repressible sugar media with UAA, potentially maximizing the mitochondrial translation with unnatural incorporation. The sfGFPmito-v3 strain has the sfGFP integrated into the COX2 locus with the appropriate 5' UTR, ensuring that it would be subject to the same regulation as the native mitochondrial genes [5].

Improvements to the overall protein yield of the mitochondrial orthogonal translation platform could help in recovering UAA-incorporated protein. Beyond the improvements discussed in Chapter 3, we have also tested supplementation with various iron salts and refinement of the carbon source combination, both leading to no significant improvements. In the real of further genetic modulation, however, there are some immediate next targets for potential overexpression. Changes in the expression of mitochondrial RNA polymerase RPO41 have been positively correlated to fluctuations in mito-transcription [6]. In an *S. pombe* model, overexpression of RPO41 increased direct DNA binding of the polymerase, leading to increased expression [7]. This evidence makes the

mitochondrial RNA polymerase a prime target for dosage increase in our orthogonal translation platform, though it's active structure as a holoenzyme with the mitochondrial transcription factor MTF1 in *S. cerevisiae* suggests the need for dual overexpression [7].

Much like how boosting the mtDNA copy number led to indirect increases is mitochondrial gene expression, modulation of organelle biogenesis could serve as a means to raise the average mitochondria count per cell. The transcriptional activators HAP1, HAP2, and HAP3 have been shown to bind the upstream regions of the cytochrome c (CYC1) locus in response to a switch to a non-fermentable carbon source [8]. We can also take advantage of the extensive study of mitochondrial biogenesis in mammalian cells, as many of the key regulators have yeast homologs of similar function. At the forefront of these transcription factors are NRF-1 and NRF-2, which recognize different DNA binding sites but both induce expression of a host of respiration-associated genes [9]. The PGC-1 family of nuclear co-activators binds DNA to facilitate transcription factor (TF) binding, making them potentially suitable targets for co-expression with their associated TFs [9]. Boosting the mitochondrial count in the absence of non-fermentable media involves mimicking the natural transcriptional response to a change in carbon source, which could potentially be accomplished by the constitutive overexpression of the aforementioned regulators.

5.2 Pivoting to a Tool for Mitochondrial tRNA Targeting

Even if multiple improvements to the mitochondrial expanded genetic code platform results in detectable levels of UAA-incorporation, the total quantity of protein recovered will likely be too insignificant for regular use as a polymer synthesis tool. As it currently stands, multi-liter yeast cultures just for analytical quantities of a target protein, for which no incorporation was observed. When compared to an equivalent bacterial platform that yielded measureable incorporation with <50 mL cultures, the goal of producing functional quantities of UAA-incorporated protein seems out of reach [2]. However, when comparing the limited protein yield to the striking click-microscopy (Chapter 3.3), it is clear that the tRNA aminoacylation component of the system is behaving as intended. Thus, the introduction of mutant synthetases can provide a way to probe the biology of mitochondrial tRNA transcription, modification, and aminoacylation.

Unlike in yeast, direct transformation into the mitochondrial genome has not been demonstrated in mammalian cell models [10]. Traditional protocols for ballistic DNA transformation into yeast mitochondria require a rho 0 recipient strain that lacks a mitochondrial genome, and such a genotype is not viably tolerated in mammalian cells in the absence of pyruvate [11,12]. While alternative approaches to generating mtDNA mutations in mammalian cells, like bacterial conjugation strategies, have shown some promise, indirect labeling strategies for mitochondrially-derived biomolecules remain appealing [13].

tRNA biology is known to play a significant role in human health: over 200 mutations to mitochondrial tRNAs are linked to known pathologies [14]. The

disease outcomes for these mutations are far-reaching, ranging from cardiomyopathy to hearing loss [15,16]. In addition to direct mutations to mt-tRNAs that modulate structure and function, genomic mutations that regulate transcription in the mitochondria can also interfere with standard mitochondrial processes [17]. A bioorthogonal labeling strategy utilizing reactive UAAs can give readouts on both specific mt-tRNA quantities and aminoacylation, and when applied to a tissue-specific mammalian disease model, can inform on how the pathology is altered/alleviated by exposure to different drugs or research compounds [14].

For example, the T12201C mutation in mt-tRNA^{HIS} disrupts the acceptor stem, leading to destabilization and a ~75% reduction in the steady-state levels of functional tRNA [18]. A substrate-promiscuous mutant copy of the histidyl-tRNA synthetase can be introduced to a cell model of this tRNA pathology for labeling with an alkyne-bearing histidine analog. Labeling with an azide can thus provide a look into how this disease model responds to screens of different compounds or genetic modulators that can have a compensatory effect on the expression of functional mt-tRNA^{HIS}. By tracking other parameters like total mitochondrial content and membrane potential with existing probes, a multidimensional readout can be constructed to give unique insight on how the physiology of the disease model is responding to user-defined conditions [19,20].

Going forward, a pivot towards a tool for mt-tRNA biology appears to be the most feasible direction for a mitochondrial expanded genetic code. The wide array of potential disease models, coupled with continuous improvements to *in*

vivo bioorthogonal labeling, offer many avenues for scientific exploration and technology-development [21]. Although the existence of promiscuous aaRS mutants hasn't been demonstrated for all the clinically relevant tRNAs, there is potential in the active site-targeted mutational strategies previously used to relax substrate specificity for bacterial synthetases [1,22]. Of course, there are outstanding questions regarding the overexpression of mutant mito-aaRSs in mammalian models. Namely, will normal mitochondrial processes be so perturbed that the resulting data isn't biologically relevant? Further investigation in healthy vs. disease mitochondrial models is needed to address these concerns.

5.3 References

[1] Kirshenbaum, K., Carrico, I.S., Tirrell, D.A. Biosynthesis of Proteins Incorporating a Versatile Set of Phenylalanine Analogues *ChemBioChem* **3**, 235-237 (2002)

[2] Yuet, K. P. *et al.* Cell-specific proteomic analysis in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.* **12**, 2705-2710 (2015)

[3] Ulery T.L., Jang S.H., Jaehning J.A. Glucose repression of yeast mitochondrial transcription: kinetics of derepression and role of nuclear genes. *Mol Cell Biol.* **2**, 1160-70 (1994)

[4] Fox, T.D. Mitochondrial protein synthesis, import, and assembly. *Genetics* **192**, 1203-34 (2012)

[5] Suhm, T. *et al.* A novel system to monitor mitochondrial translation in yeast. *Microbial cell* **5**,158-164 (2018)

[6] Wilcoxen S.E., *et al*. Two forms of RPO41-dependent RNA polymerase. Regulation of the RNA polymerase by glucose repression may control yeast mitochondrial gene expression. *J Biol Chem*. **25**, 12346-51 (1988)

[7] Jiang H., *et al*. Identification and characterization of the mitochondrial RNA polymerase and transcription factor in the fission yeast Schizosaccharomyces pombe. *Nucleic Acids Res*. **12**, 5119-30 (2011)

[8] Grivell, L.A. Nucleo‐mitochondrial interactions in yeast mitochondrial biogenesis. *European Journal of Biochemistry* **182**, 477-493 (1989)

[9] Scarpulla R.C. Nuclear activators and coactivators in mammalian mitochondrial biogenesis. *Biochim Biophys Acta.* **1576**, 1-14 (2002)

[10] Yoon Y.G., Koob M.D. Toward genetic transformation of mitochondria in mammalian cells using a recoded drug-resistant selection marker. *J Genet Genomics.* **38**,173-9 (2011)

[11] Johnston S.A., *et al*. Mitochondrial transformation in yeast by bombardment with microprojectiles. *Science*. **240**, 1538-41 (1988)

[12] Pearce S.F., *et al*. Regulation of Mammalian Mitochondrial Gene Expression: Recent Advances. *Trends Biochem Sci.* **42**, 625-639 (2017)

[13] Yoon Y.G., Koob M.D., Yoo Y.H. Re-engineering the mitochondrial genomes in mammalian cells. *Anat Cell Biol*. **43**, 97-109 (2010)

[14] Abbott J.A., Francklyn C.S., Robey-Bond S.M. Transfer RNA and human disease. *Front Genet*. **5** (2014)
[15] Taylor, R. W., *et al.* A homoplasmic mitochondrial transfer ribonucleic acid mutationas a cause of maternally inherited hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol.* **41**, 1786–1796 (2003)

[16] Lax, N. Z., *et al*. Early-onset cataracts, spastic paraparesis, and ataxiacaused by a novel mitochondrial tRNAGlu (MT-TE) gene mutation caus-ing severe complex I deficiency: a clinical, molecular, and neuropathologicstudy. *J. Neuropathol. Exp. Neurol.* **72**, 164–175 (2013)

[17] Bestwick M.L, Shadel G.S. Accessorizing the human mitochondrial transcription machinery. *Trends Biochem Sci.* **38**, 283-91 (2013)

[18] Yan X., *et al.* Maternally transmitted late-onset non-syndromic deafness is associated with the novel heteroplasmic T12201C mutation in the mitochondrial tRNAHis gene. *J Med Genet.* **48**, 682-90 (2011)

[19] Chazotte B. Labeling mitochondria with MitoTracker dyes. *Cold Spring Harb Protoc*. **8**, 990-2 (2011)

[20] Sivandzade, F., Bhalerao, A., Cucullo, L. Analysis of the Mitochondrial Membrane Potential Using the Cationic JC-1 Dye as a Sensitive Fluorescent Probe. *Bio-protocol* **9**, 3128 (2019)

[21] Mushtaq S., Yun S.J., Jeon J. Recent Advances in Bioorthogonal Click Chemistry for Efficient Synthesis of Radiotracers and Radiopharmaceuticals. *Molecules* **24**, 3567 (2019)

[22] Deiters, A. *et al.* Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *J Am Chem Soc.* **125**,11782-11783 (2003)

Appendix

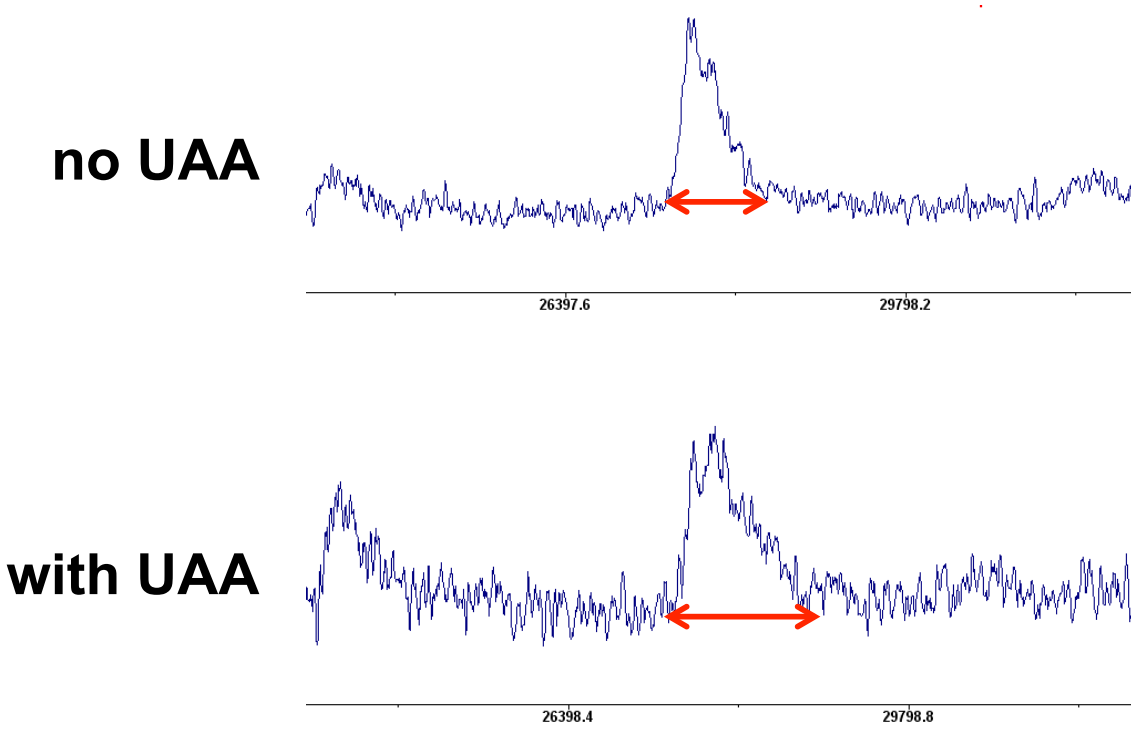


Figure A.1 | Whole sfGFPmito-v3 with UAA on MALDI-TOF
MALDI-TOF mass spectra of sfGFPmito-v3. Red arrows indicate peak broadening of GFP mass. UAA is 1 mM of **2**.

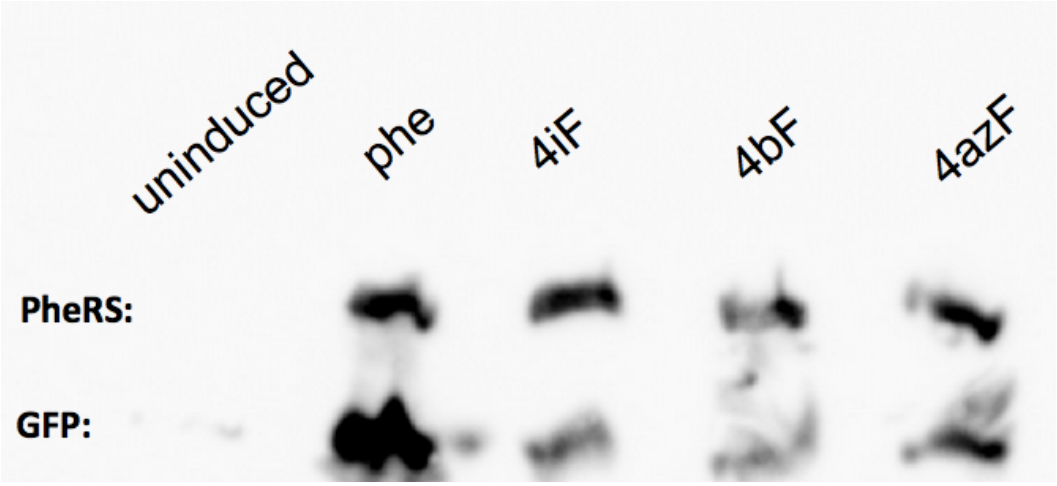


Figure A.2 | Anti-HIS Immunoblot of Bacterial UAA Induction Cultures
Loaded lysate normalized by cell count. PheRS and GFP both expressed with HIS tag. Phe: phenylalanine, 4iF: **2**, 4bF: **4**, 4azF: **5**.

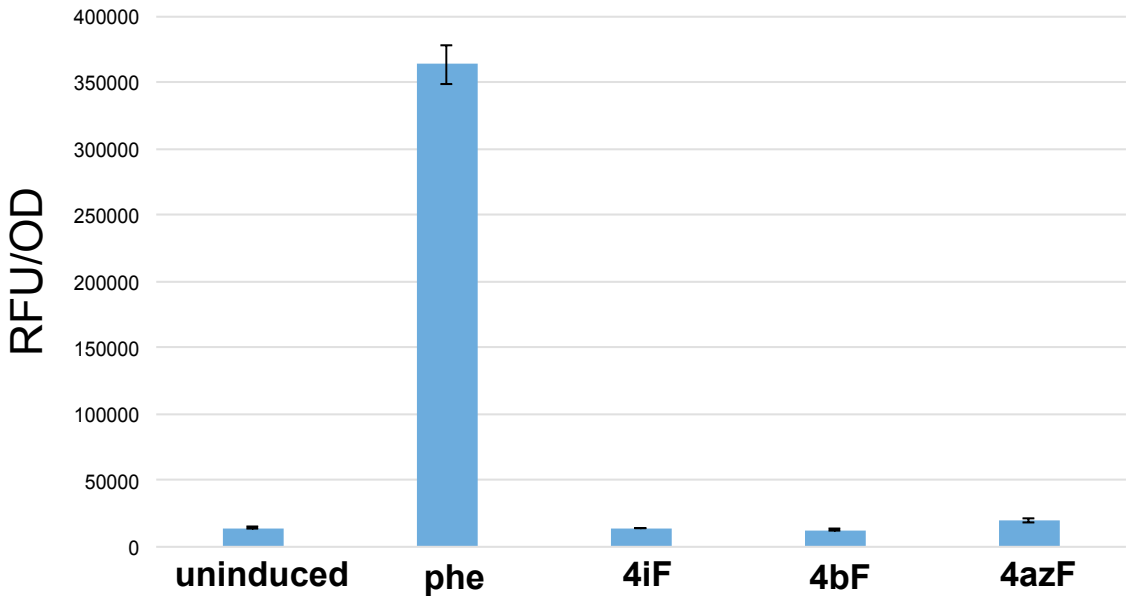


Figure A.3 | OD-normalized RFU of Bacterial UAA Induction Cultures
Data collected via TECAN spectrophotometer. Background subtracted with bacteria lacking the sfGFP expression plasmid. Supplemented amino acids were added to 2 mM. Phe: phenylalanine, 4iF: **2**, 4bF: **4**, 4azF: **5**.

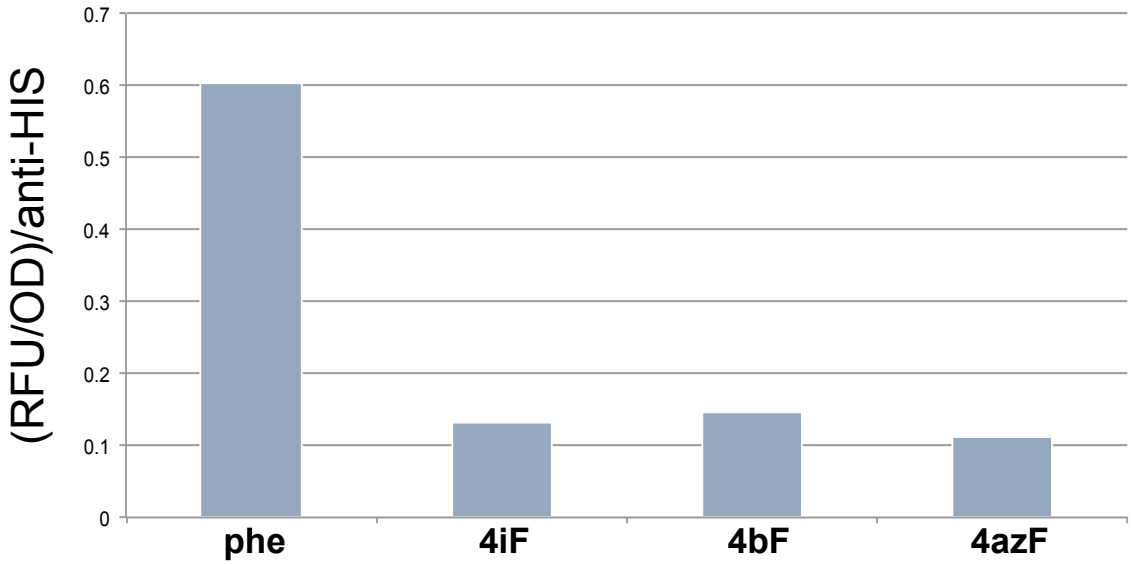


Figure A.4 | OD-normalized RFU of Bacteria Corrected for sfGFP Yield
Fluorescence data from Appendix 3 adjusted with sfGFP band intensity from Appendix 2. Everything shown is relative to values from un-induced sample. Supplemented amino acids were added to 2 mM. Phe: phenylalanine, 4iF: **2**, 4bF: **4**, 4azF: **5.**

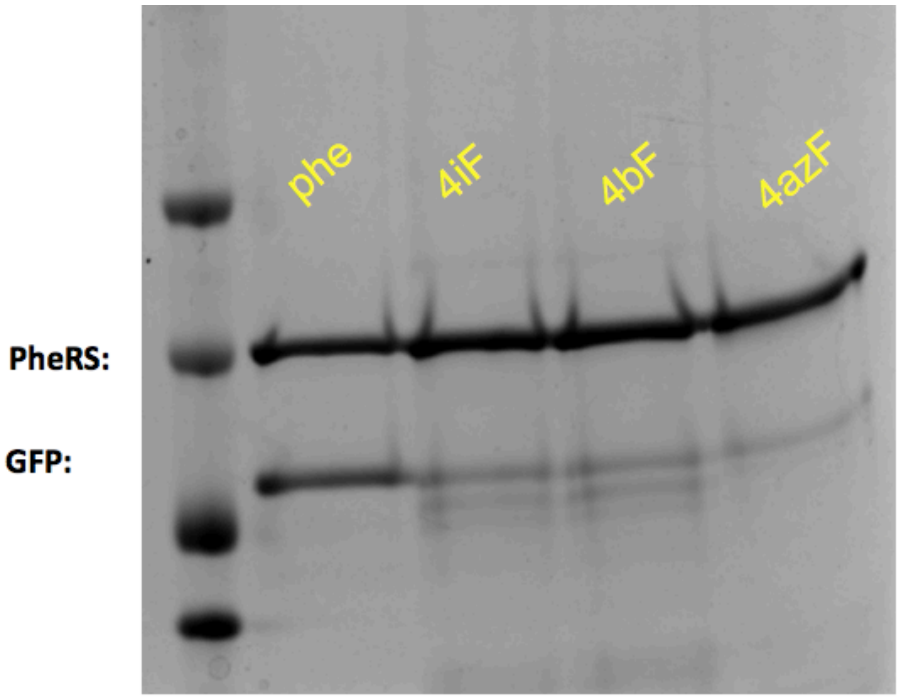


Figure A.5 | HIS Purification Eluates from E Coli UAA Induction
Protein loaded into purification normalized by sfGFP immunoblot intensities. Polyacrylamide gel stained with Coomassie R250. 4iF: **2**, 4bF: **4**, 4azF: **5**.

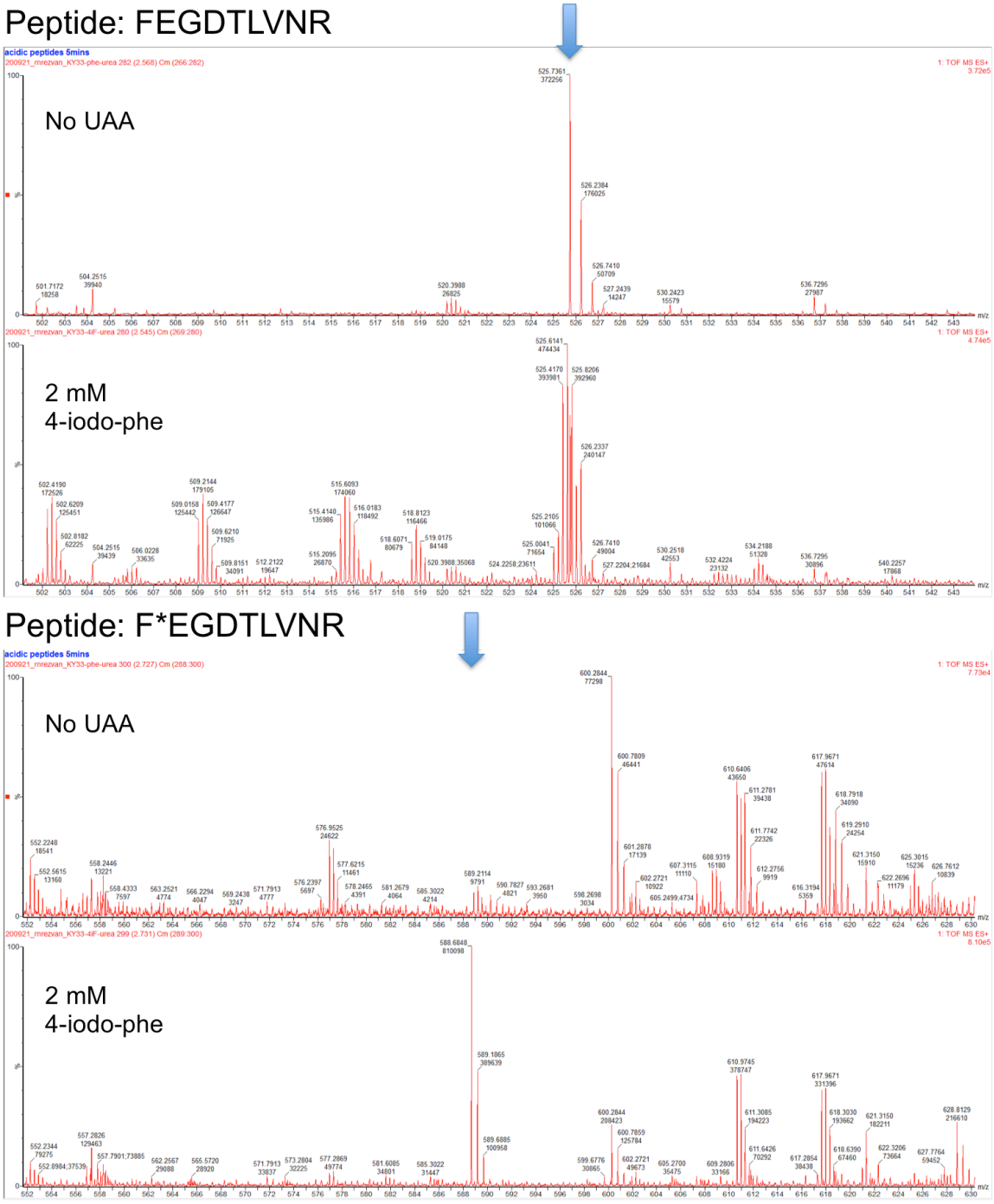
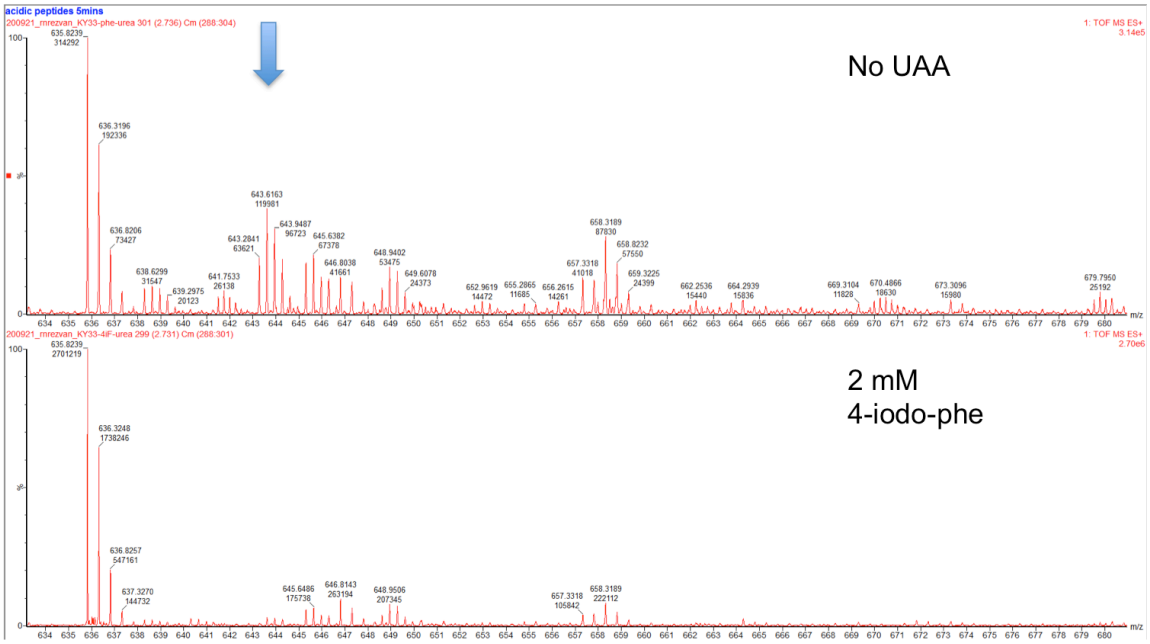


Figure A.6 | 4-iodo-phe Incorporation in E. coli sfGFP Peptide FEGDTLVNR
Mass spectra captured with LC-ESI-TOF after tryptic digest. Arrows designate target peak locations. F* represents 4-iodo-phenylalanine.

Peptide: LEYNFNSHNVYITADK



Peptide: LEYNF*NSHNVYITADK

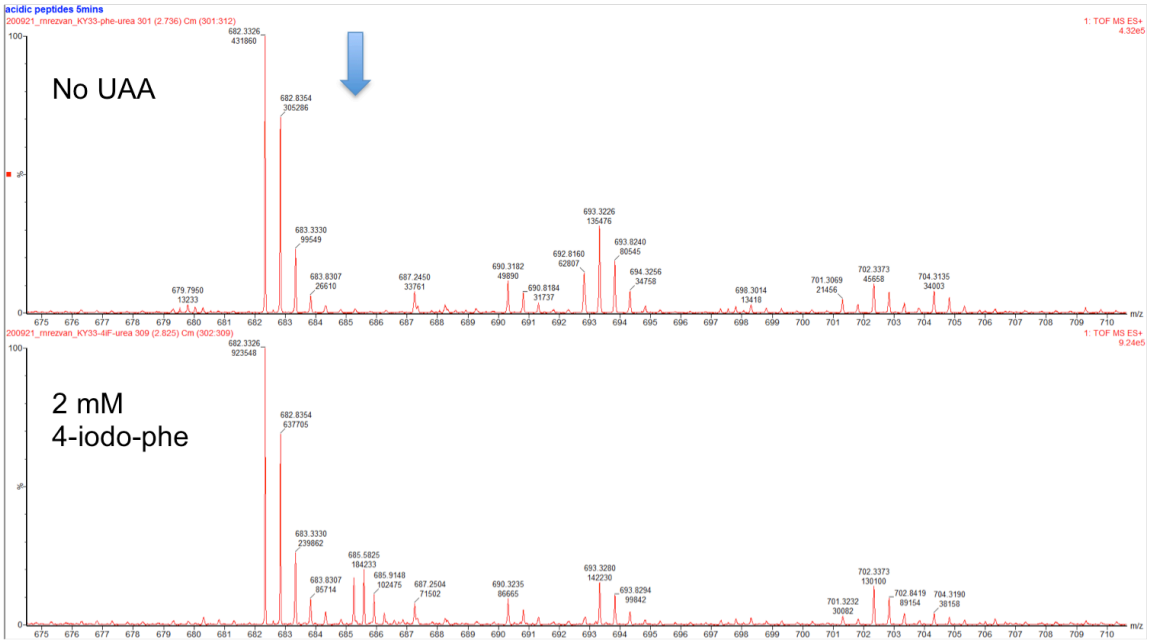


Figure A.7 | 4-iodo-phe Incorporation in E. coli sfGFP Peptide LEYNFNSHNVYITADK

Mass spectra captured with LC-ESI-TOF after tryptic digest. Arrows designate target peak locations. F* represents 4-iodo-phenylalanine.

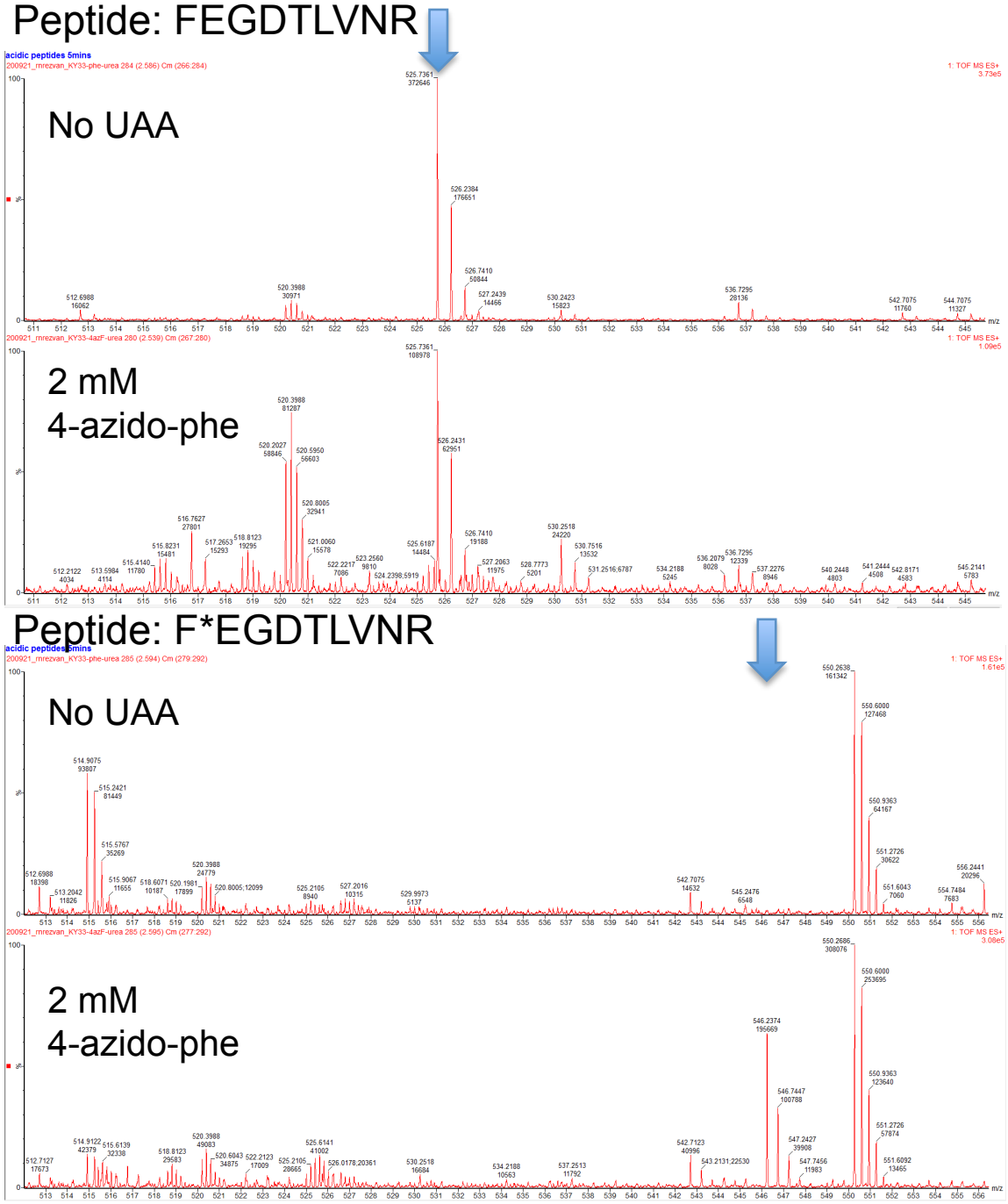


Figure A.8 | 4-azido-phe Incorporation in E. coli sfGFP Peptide FEGDTLVNR
Mass spectra captured with LC-ESI-TOF after tryptic digest. Arrows designate target peak locations. F* represents 4-azido-phenylalanine.

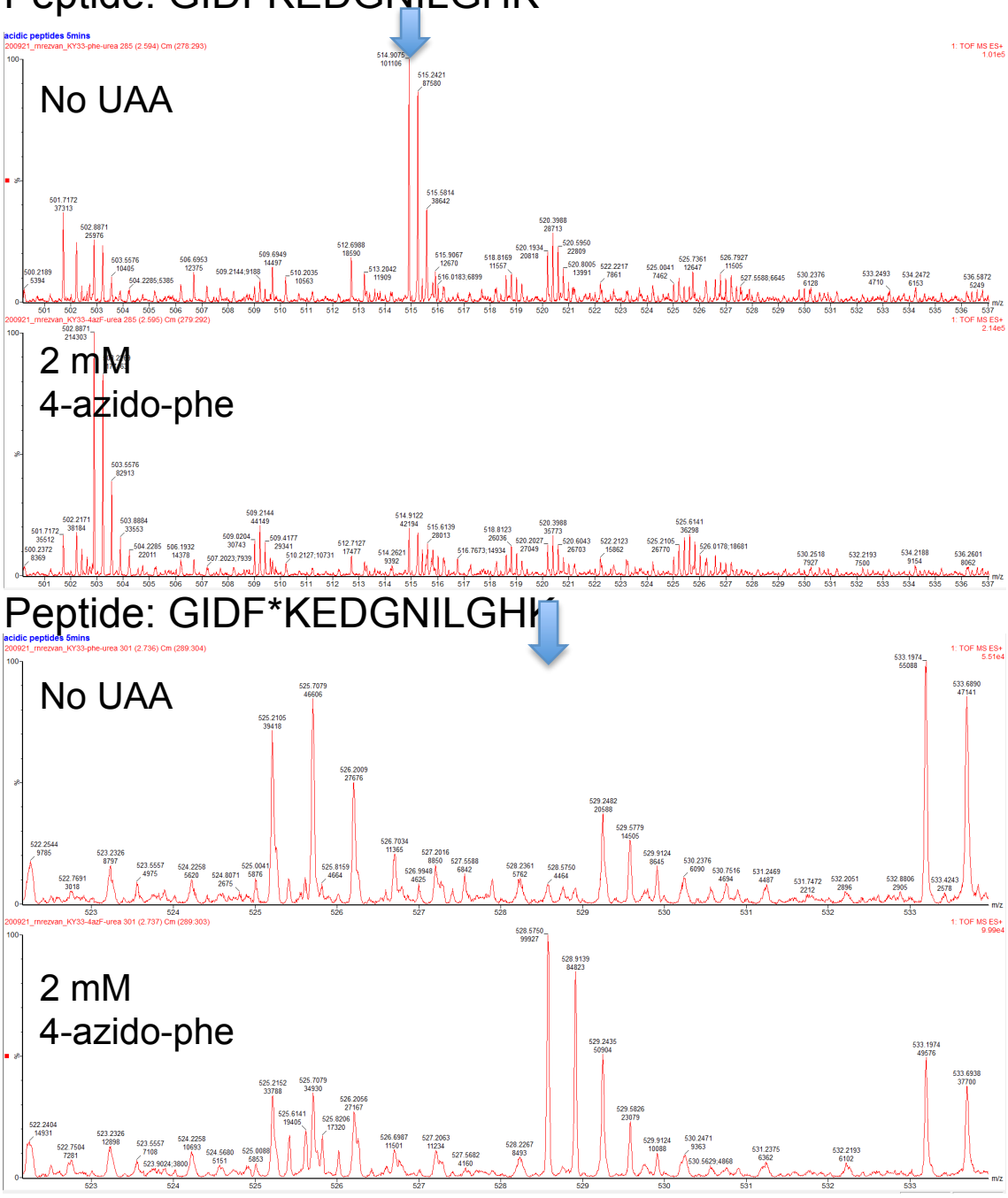


Figure A.9 | 4-azido-phe Incorporation in E. coli sfGFP Peptide GIDFKEDGNILGHK

Mass spectra captured with LC-ESI-TOF after tryptic digest. Arrows designate target peak locations. F* represents 4-azido-phenylalanine.

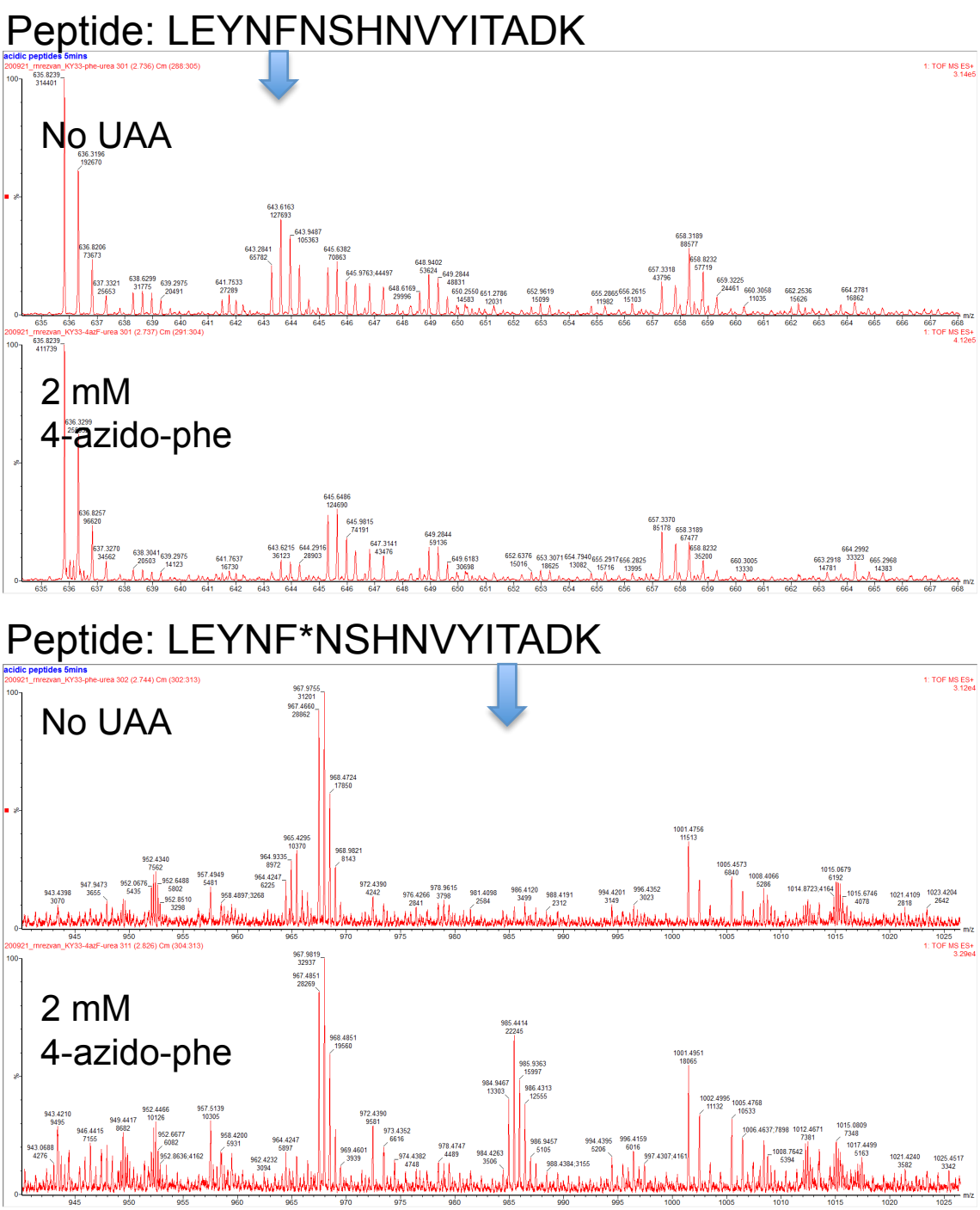


Figure A.10 | 4-azido-phe Incorporation in E. coli sfGFP Peptide LEYNFNSHNVYITADK

Mass spectra captured with LC-ESI-TOF after tryptic digest. Arrows designate target peak locations. F* represents 4-azido-phenylalanine.

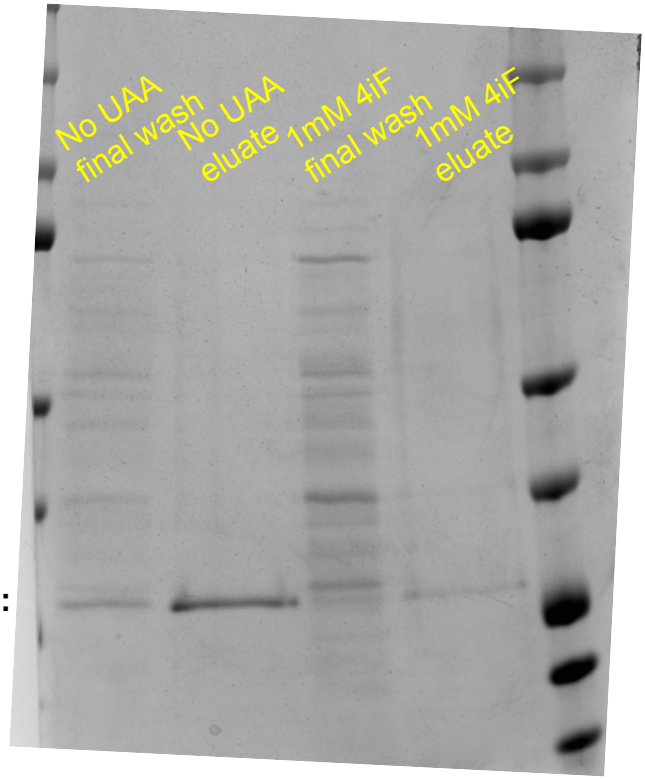


Figure A.11 | Cobalt Purification Recovers sfGFP from 4-iodo-phe Mitochondrial Samples
Polyacrylamide gel stained with Coomassie R250. “Final wash” refers to the last of 4 washes with urea-containing buffer. Input sample normalized by sfGFP immunoblot intensity. 4iF: **2**.