

AN ABSTRACT OF THE DISSERTATION OF

Robert J. Blizzard for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on January 8, 2019.

Title: *In Vivo* Reactions of Tetrazines Incorporated through Genetic Code Expansion

Abstract approved: _____

Ryan A. Mehl

The inverse electron demand Diels Alder reaction between tetrazines and strained alkenes is an exceptionally useful tool in functionalizing to biomolecules since it is orthogonal to the chemistry of most living systems and have exceptionally high rate constants. In particular reactions between strained trans-cyclooctenes (sTCO) and tetrazines can achieve second order rate constants of $10^6 \text{ M}^{-1}\text{s}^{-1}$ or greater. This bioorthogonal coupling reaction can be used to quantitatively modify proteins with fluorescent labels, inhibitors, or dimerization agents in live cells with short reaction time. In this work, we explore the use of genetic code expansion to incorporate tetrazine containing amino acids into proteins for the purpose of conjugating sTCO-containing labeling agents to proteins in live cells

In *Escherichia coli*, the tetrazine containing amino acid, 4-(6-methyl-*s*-tetrazin-3-yl)phenylalanine (Tet-v2.0) is incorporated into cells through genetic code expansion. An orthogonal amino acyl-tRNA synthetase is generated that can charge tRNA_{CUA} using Tet-v2.0 as a substrate allowing for the site-specific incorporation of Tet-v2.0 into proteins. Tet-v2.0 on proteins is shown to react with sTCO with a high rate of reaction of $87,000 \text{ M}^{-1}\text{s}^{-1}$ and minimal off-target reactions. To express tetrazine containing proteins in eukaryotic cells, a new set of amino acids based on 3-(6-methyl-*s*-tetrazin-3-yl)phenylalanine (Tet-v3.0), are developed that can be used as substrate by *Methanosarcina. barkeri* Pyrrolysine tRNA synthetases (PylRS). The orthogonal Tet-v3.0-amino acyl-tRNA synthetase/tRNA pair was used to incorporate

Tet-v3.0 amino acids into proteins in HEK293T cells and shown to react with both fluorescent labels and dimerizing sTCO molecules.

We then explore methods to improve the efficiency of the reaction between tetrazines and sTCO, through minimizing side reactions. These side reactions include the reduction of tetrazines to dihydro-tetrazines and irreversible degradation of tetrazines to a yet to be characterized product. To eliminate off-target reactions the reduction of tetrazines is monitored *in cellulo* and proteins are oxidized *in cellulo* through photooxidation. This oxidation reaction has the potential to be an effective method of establishing higher labeling yields. Finally, trapping the tetrazine in the dihydro-state and selectively oxidizing and labeling the tetrazine may be an effective method of minimizing off-target tetrazine reactions in cells by protecting the oxidized amino acid.

©Copyright by Robert J. Blizzard
January 8, 2019
All Rights Reserved

In Vivo Reactions of Tetrazine Incorporated through Genetic Code Expansion

by
Robert J. Blizzard

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented January 8, 2019
Commencement June 2019

Doctor of Philosophy dissertation of Robert J. Blizzard presented on January 8, 2019

APPROVED:

Major Professor, representing Biochemistry and Biophysics

Head of the Department of Biochemistry and Biophysics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Robert J. Blizzard, Author

ACKNOWLEDGEMENTS

Over the past five years, the one person that I absolutely could not have done without is my advisor, Ryan Mehl. I know that I can be a little frustrating to work with, and I thank you for having patience. You have always been a great advocate for all the students that you work with and mentor, and though I can only speak for myself, I'm sure the other students will agree. You are receptive towards my ideas even though they might not always be the greatest. Thank you.

I would like to thank my parents and my brothers. If there is one thing I have missed the most in the years that I have been at graduate school, it is the ability to spend time with you. I want to thank my mom in particular for teaching my brother and I math at night. Also, I appreciate the patience that you had all the times you took us to science museums like the Boonshoft. Dad, I want to thank you for sharing your science magazines and getting me interested in science in general. I love you both and I appreciate what you have done for me. Joey, Jacob, and Sam, I'm sorry that I haven't been able to spend much time with you over the past few years, but I hope to be more available in the future.

The people in the Mehl Lab have been essential to my success. I would like to thank them for teaching me and showing me the basics of genetic code expansion. In particular, I would like to thank Linda Benson and Ryan Mehl for teaching me selections, Joseph Porter for discussing my cloning and molecular biology reactions, and Rick Cooley, Riley Bednar, Nathan Waugh, and Rachel Henson for general lab advice. I also must thank both Subhashis Jana and Hyo Sang Jang for their expertise in chemical synthesis and eukaryotic cell expression which is absolutely essential to my work. I wanted to thank some of the undergraduates that synthesized the tetrazines that I worked with before Jana came along including Dakota, True, and Sam as well as some of the lab members that have made my time here more enjoyable including Wes, Taylor, Hayati, Garth, and Phil.

One aspect of my time here that I often overlook is how cooperative some of the other labs and lab members that I have worked with. Therefore, I wanted to thank the Beckman Lab and Mass Spectrometry Facility for enabling me to get so many essential results. In particular I wanted to thank Joe Beckman, Jared, and Jeff Morre

for helping with mass spectrometry experiments. I want to thank the Johnson Lab for use of their fluorimeter as well as the Barbar, Karplus, Freitag, Loesgen, and Merrill Labs for use of various instruments and reagents.

I want to thank all of the friends I met along the way including my fellow cohort of graduate students, Andrew, Nathan, Rachel, Sara, John, Amanda, and Ryan, the graduate students above me who helped me navigate my way through school including Joey, Kelsey, Chelsea, Steve, Jillian, and Arden, as well as the graduate students after me who I wish luck to in school and other endeavors, Andrew, Michelle, Nathan, Riley, Shauna, Ally, Dan, Heather, Kasie, Amber, Phil, Aiden, and Kayla, and finally the friends I made in spite of graduate school including Kate, Ben, Annie, Alexa, and Paul.

Finally, there is someone who I have left off of several of these lists, not because she is not important to me, but mostly to annoy her. That person is Elise Van Fossen. Elise in particular, I have to thank for taking over for Rachel in putting up with sitting next to me on a day to day basis. Thank you Elise.

CONTRIBUTION OF AUTHORS

In Chapter 2, Wes Brown, Chris Bazewicz, Yi Li, and Dakota Backus planned and performed experiments. Ryan Mehl planned experiments and assisted in writing the paper.

In Chapter 3, Ryan Mehl, Subhashis Jana, and Hyo Sang Jang planned experiments. Subhashis Jana and Hyo Sang Jang performed experiments. Ryan Mehl, Subhashis Jana, and Hyo Sang Jang assisted in writing the text.

In Chapter 4, Ryan Mehl, Subhashis Jana, John Perona, and Camden Driggers planned experiments. Subhashis Jana and Camden Driggers performed experiments. Ryan Mehl and Subhashis Jana assisted in writing the text.

TABLE OF CONTENTS

Chapter 1	1
Thesis Overview	1
Advantages from Bioorthogonal Ligations	2
Common Bioorthogonal Ligations	3
Inverse Electron Demand Diels Alder Reactions	6
The Use of Tetrazines in Bioorthogonal Reactions.....	8
Methods of Introducing Bioorthogonal Functional Groups	10
Site-Specific Genetic Code Expansion.....	12
Orthogonal tRNA/Aminoacyl tRNA Synthetase Pairs.....	14
Generating Aminoacyl tRNA Synthetases for Incorporating Amino Acids	15
Disseration Overview	17
Chapter 2.....	19
Ideal Bioorthogonal Reactions Using A Site-Specifically Encoded Tetrazine Amino Acid.....	19
Abstract.....	20
Introduction	20
Supplemental	22
Results	43
Discussion.....	50
Acknowledgements	51
Chapter 3.....	52
Tetrazine Incorporation and Reaction in Eukaryotic Cells.....	52
Introduction	53
Results/Discussion.....	54
Conclusion.....	73
Materials and Methods	74

Chapter 4.....	94
Side Reactions of Tetrazines.....	94
Introduction	95
Results/Discussion.....	95
Conclusion.....	110
Materials and Methods	111
Chapter 5.....	120
Conclusions and Outlook.....	120
Summary.....	121
Future Experiments	123
Potential Research Directions.....	125
Directions of the Field	127

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Fig 1.1. Common Bioorthogonal Ligation Reactions.....	4
Fig 1.2. Structures of Common IEDDA Functional Group.....	7
Fig 1.3. Labeling Biomolecules with IEDDA Reactions.....	9
Fig 1.4. Methods of Introducing Bioorthogonal Functional Groups.....	11
Fig 1.5. Site Specific Genetic Code Expansion.....	13
Fig 1.6. Selections for aaRSs Incorporating an Amino Acid of Interest.....	17
Fig 2.1. Synthesis of 4-(6-methyl- <i>s</i> -tetrazin-3-yl)phenylalanine (Tet-v2.0).....	23
Fig 2.2. ¹ H NMR of 4-(6-methyl- <i>s</i> -tetrazin-3-yl)phenylalanine (Tet-v2.0).....	24
Fig 2.3. ¹³ C NMR of 4-(6-methyl- <i>s</i> -tetrazin-3-yl)phenylalanine (Tet-v2.0).....	24
Fig 2.4. MS of 4-(6-methyl- <i>s</i> -tetrazin-3-yl)phenylalanine (Tet-v2.0).....	25
Fig 2.5. Fluorescence Incorporation of Tet-v2.0 Synthetases.....	29
Fig 2.6. Reaction of GFP-Tet-v2.0 with sTCO and dTCO.....	31
Fig 2.7. <i>In vitro</i> incubation of GFP-Tet-v2.0 with sTCO.....	32
Fig 2.8. <i>In vitro</i> kinetic analysis of GFP-Tet-v2.0 with sTCO and dTCO.....	33
Fig 2.9. Calibration Curve for Determining sTCO Concentration.....	35
Fig 2.10. Synthesis of TAMRA linked sTCO.....	36
Fig 2.11. ¹ H NMR of TAMRA-sTCO.....	36
Fig 2.12. SDS-PAGE analysis of GFP-Tet-v2.0 reactivity.....	37
Fig 2.13. Fluorescent SDS-PAGE analysis of GFP-Tet-v2.0 reaction.....	38
Fig 2.14. Quenching of the reaction between tetrazine and sTCO.....	40
Fig 2.15. 15 min reaction between sTCO-TAMRA and GFP-tetrazine.....	41
Fig 2.16. 5 min reaction between sTCO-TAMRA and GFP-tetrazine.....	41

LIST OF FIGURES (continued)

<u>Figure</u>	<u>Page</u>
Fig 2.17. ¹ H NMR of Tet-v2.0 in dPBS.....	42
Fig 2.18. Genetic incorporation of Tet-v2.0, into proteins and labeling with sTCO..	44
Fig 2.19. <i>In vitro</i> and <i>in cellulo</i> rate constant determination for reaction of GFP-Tet-v2.0 with sTCO.....	47
Fig 2.20. Sub-stoichiometric characterization of GFP-Tet-v2.0 reaction with sTCO.	49
Fig 3.1: Eukaryotic Incorporation of Tetrazines Enables Control of Proteins	53
Fig 3.2. Amino acid libraries of PylRS active sites.	55
Fig 3.3. Structures of Tetrazine Amino Acids in PylRS.....	56
Fig 3.4: Efficiency of Generated Synthetases for Tet-v3.0 Amino Acids	59
Fig 3.5: Mass Spectra of Tetrazine GFP and Reactions	60
Fig 3.6: Mobility Shift Assay of GFP150-Tet-v3.0.....	62
Fig 3.7: Protein Kinetics of the Reaction between GFP150-Tet-v3.0 and sTCO.....	63
Fig 3.8: Toxicity of Tet-v3.0 Amino Acids	64
Fig 3.9: Concentration Dependence Curves of Tet-v3.0	65
Fig 3.10: HEK293T Cell Production of GFP150-Tet-v3.0.....	66
Fig 3.11: The Effect of a Nuclear Export Sequence on GFP Production	67
Fig 3.12. Flow Cytometry of TAMRA labeled HEK293T cells.	68
Fig 3.13. Two-Dimensional Flow Cytometry of Tet-v3.0 Expressing HEK293T Cells.	69
Fig 3.14. Background Reactions of TAMRA Fluorophores on HEK Cell Lysate.	70
Fig 3.15. In Cell Fluorescence Increase of GFP150-Tet-v3.0-butyl upon Reaction. .	71
Fig 3.16. Mobility Shift of HEK293T cell expressed GFP150-Tet-v3.0-butyl.....	72
Fig 3.17. Concentration Dependent Dimerization of GFP150-Tet-v3.0-butyl.....	73
Fig 3.18. Synthesis of s-tetrazine derivatives of phenylalanine (Tet-v3.0).	74

LIST OF FIGURES (continued)

<u>Figure</u>	<u>Page</u>
Fig 3.19. Synthesis of TAMRA linked strained alkenes (sCCO, sTCO, dTCO and oxoTCO).	79
Fig 3.20. Synthesis of PEGylated sTCO, di-sTCO and TAMRA-sTCO.	82
Fig 4.1. Time Dependent Mobility Shift Assay of GFP150-Tet-v2.0.	96
Fig 4.2. Tet-v3.0-methyl Kinetic Comparison.....	97
Fig 4.3. Fitting of <i>In cellulo</i> Reactions to 1 st Order Equations.	97
Fig 4.4. Reaction of Purified GFP150-Tet-v2.0 with TCEP.....	100
Fig 4.5. Photooxidation of Tet-v2.0 and Tet-v3.0 in <i>E. coli</i>	101
Fig 4.6. Time Dependent Photooxidation.	103
Fig 4.7. <i>In vitro</i> Aminoacylation of tRNA.....	105
Fig 4.8. Effect of Dihydro-Tet-v2.0 on Cellular Expression.	106
Fig 4.9. Expression Dependent Gel Shift of GFP150-Tet-v2.0.....	107
Fig 4.10. Purification of the Unreactive Protein.	108
Fig 4.11. Mass Spectra of the Unreactive Protein.	109
Fig 4.12. Incorporation of Tet-v2.0 Derivatives.	110
Fig 5.1. Dimerization with Bioorthogonal Ligations.....	124
Fig 5.2. Future Routes of Tetrazine Synthesis.....	126

LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table 2.1: Components for autoinducing and non-inducing mediums.....	28
Table 2.2: Sequence of top performing Tet-RSs.....	29
Table 3.1: Mutants of the PylRS Libraries.....	55
Table 3.2: Stopped Flow Rate Characterization and Cyclic Voltammetry Characterization of Tet-v3.0 Derivatives	57
Table 3.3: Sequences of Generated <i>Mb</i> Synthetases.....	58
Table 3.4: Observed Masses of Tet-v3.0 containing GFP.....	61
Table 4.1: Reduction Potentials of Tet-v2.0 and Tet-v3.0.....	99

***In Vivo* Reactions of Tetrazine Amino Acids using Genetic Code Expansion**

Chapter 1

Thesis Overview

Advantages from Bioorthogonal Ligations

Bioorthogonal ligation chemistry is a term first described in 2003 by Hang et al. which encompasses reactions where the functional groups involved in the reaction undergo minimal off-target reactions in biological environments¹. These reactions must form a covalent linkage between two functional groups that are reactive to each other, but unreactive to all other functional groups found in living systems². The main utility of such a reaction is the ability to selectively label biological macromolecules *in vivo*, though this chemistry is often used to label biomolecules *in vitro* as well. Such chemistry has been used to label every major biological macromolecule, including proteins³⁻⁷, DNA⁸⁻¹⁰, RNA^{11,12}, lipids¹³, and carbohydrates¹⁴⁻¹⁶. Rationales for the development of these reactions include live cell fluorescent labeling^{4,17,18}, mimicking native macromolecular modifications^{19,20}, generating of antibody drug conjugates²¹⁻²³, *in vitro* labeling of biomolecules with radioisotopes²⁴⁻²⁷, and selective inhibition of enzymes within a related family²⁸. As such, the generation and characterization of effective bioorthogonal reactions is of interest to biotechnology and pharmaceutical communities.

Many bioorthogonal functional groups can undergo side reactions with components of biological systems. These reactions can result in degrees of cellular toxicity^{29,30} such that they are only suitable for labeling biomolecules *in vitro*³¹. In some cases, even reactions or functional groups with some degree of cellular toxicity can be used to label biomolecules on viruses³² or on cell surfaces³³. In addition to complications due to toxicity, many bioorthogonal functional groups undergo some degree of degradation *in vivo*. This degradation results in lower reaction yields and limits utility of bioorthogonal reactions.^{34,35} The propensity towards degradation is asymmetric in that the functional group attached to the biomolecule of interest requires a greater stability than the cognate bioorthogonal label because the ability to incorporate this functional group into the biomolecule of interest often takes substantially more time than the labeling reaction itself. An unstable functional group can often be used as the labeling reagent³⁶⁻³⁸, limiting its exposure to the *in vivo* environment. Functional groups may be more bioorthogonal towards certain organisms

or environments than others. Factors such as pH, reduction potential, or concentrations of biological species may impact the rate or equilibria of side reactions. Chapter 4 of this dissertation addresses tetrazine bioorthogonal reactions that undergo reduction reactions in an environment dependent manner.

A highly desirable quality of bioorthogonal ligation reactions is a high rate of the reaction. Higher second order rate constants result in faster labeling experiments and the ability to study shorter-lived cellular processes. Given a second order rate constant of $0.1 \text{ M}^{-1}\text{s}^{-1}$ and concentrations of $10 \text{ }\mu\text{M}$ for both bioorthogonal functional groups, a reaction would require 14.6 hours to reach 95% completion. In comparison, a reaction with a rate constant of $1000 \text{ M}^{-1}\text{s}^{-1}$ can reach 95% completion in approximately 5 s. When attempting to perform experiments in live cells, the time an experiment takes dictates the cellular events that such an experiment can effectively report on. A 14.6 hour labeling experiment would be ineffective at measuring events such as cell division. To overcome this, higher concentrations of labeling reagent can be used. This approach, however, can generate its own problems such as excess fluorophore interfering with the ability to observe a labeled macromolecule, or excess inhibitor resulting in off target inhibition toward unlabeled enzyme. The cleanest solution to address these issues is to develop bioorthogonal reactions with high rate constants, though this is often limited by available bioorthogonal chemistries.

Common Bioorthogonal Ligations

While numerous bioorthogonal chemistries and variations on bioorthogonal chemistries have been developed³⁹⁻⁴¹ (Fig 1), only a handful have seen extensive use. Commonly used reactions tend to be rapid reactions between small, synthetically basic functional groups with high stability. Understanding the benefits and limitations of newly developed bioorthogonal chemistries is important in evaluating their utility. Additionally, there is interest in the development of reactions in which two or more bioorthogonal ligation chemistries can be performed on the same biomolecule or in the same environment. These reactions can be used, for instance, to introduce multiple fluorophores for FRET experiments, or to induce the formation of heterodimers. Such

“orthogonal bioorthogonal chemistries” require a set of at least four bioorthogonal functional groups that show no cross reactivity⁴²⁻⁴⁴. An alternate method of achieving orthogonal bioorthogonal ligations is to take a set of bioorthogonal chemistries whose cross reactivity, though present, is slow relative to the rate of the corresponding labeling reactions.⁴³ Sequential addition of labels can then be used to perform dual-labeling reactions. Orthogonal bioorthogonal reactions are difficult to achieve in large part due to the fact that many existing bioorthogonal reactions use the same or similar functional groups. The most common example of this is the azide functional group which is used in Copper Catalyzed Azide-Alkyne Cycloaddition (CUAAC), Strain Promoted Azide Alkyne Cycloaddition (SPAAC), and the Staudinger ligation.⁴⁵

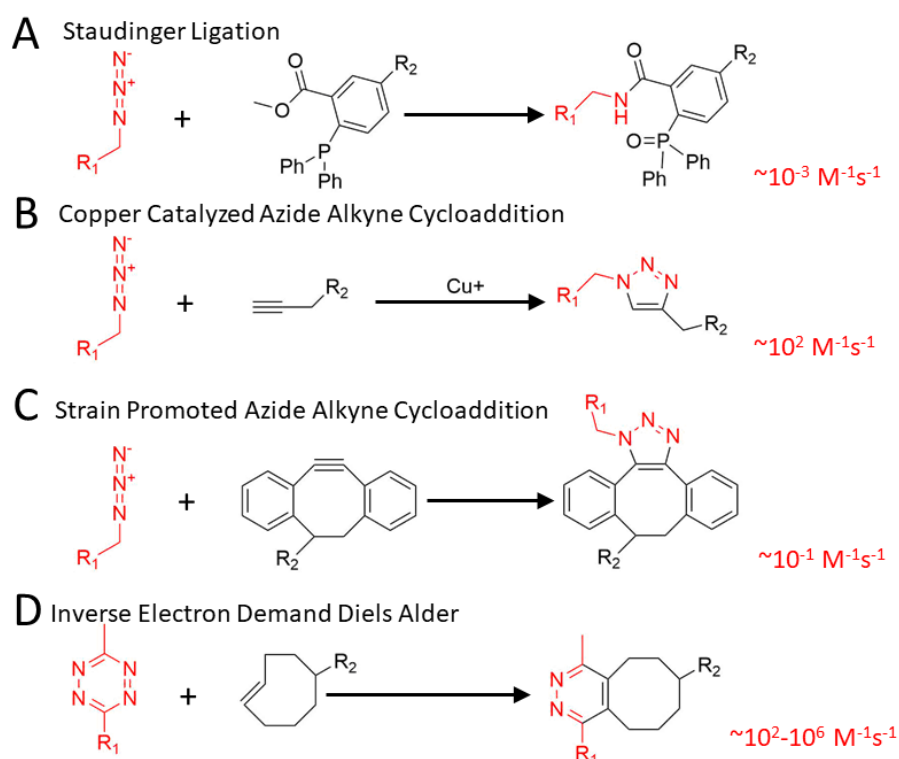


Fig 1.1. Common Bioorthogonal Ligation Reactions.

A) Staudinger ligation between aryl phosphines and azides. B) Copper Catalyzed Azide Alkyne Cycloaddition (CUAAC) commonly referred to as “click” chemistry. C) Strain Promoted Azide Alkyne Cycloaddition (SPAAC) does not refer a Copper(I) catalyst as required by CUAAC reactions. D) Inverse electron demand Diels Alder (IEDDA) reaction between tetrazines and strained alkenes (TCO pictured).

Staudinger ligations are reactions between azides and aryl phosphines substituted by an ester. This chemistry is a variation on the Staudinger reaction which is itself a reduction of the azide by a triaryl phosphine⁴⁶. This reaction can be used as a ligation reaction if the partially reduced intermediate is trapped by nucleophilic attack on the ester (Fig 1.1A). While this reaction sees limited use due to its slow reaction rate⁴⁷ of $10^{-3} \text{ M}^{-1}\text{s}^{-1}$, the small size of the azide functional group is a useful attribute. Small functional groups are useful because they are less likely to perturb the structure or activity of biomolecules³⁹.

The azide functional group is also used in CUAAC reactions, which are among the most commonly used bioorthogonal ligations and are commonly termed “click” chemistry reactions after Kolb et al.⁴⁸, used it to describe synthetic reactions in which heteroatom-carbon bonds are formed in a selective manner with large thermodynamic driving forces. While the term click chemistry used in this manner is an apt description of the CUAAC reactions, bioorthogonal ligations requires more stringent properties including the stability of the functional groups in biological environments as well as the ability to perform the reaction under physiological conditions. The CUAAC is a cycloaddition that requires a toxic copper catalyst^{29,30} (Fig 1.1B). Despite this toxicity, CUAAC chemistry is still termed bioorthogonal. The reaction can proceed quickly with second order rate constants of up to $200 \text{ M}^{-1}\text{s}^{-1}$ depending on Cu^+ concentration⁴⁹. The greatest benefits of this reaction include the small size and synthetic accessibility of both functional groups.

The toxicity of CUAAC reactions has been overcome by Carolyn Bertozzi⁵⁰ through modification of the reaction such that a Cu^+ catalyst is no longer required. This is largely performed by the introduction of strain to the alkyne functional group (Fig 1.1C) resulting in SPAAC chemistry. Doing so accelerates the rate of the reaction. While other solutions to increase the rate or efficacy of the CUAAC chemistry have been developed⁴⁹, these methods still require the presence of the Cu^+ catalyst. This SPAAC chemistry has been widely used but comes with the drawback of having low rate constants in the range of 10^{-2} to $1 \text{ M}^{-1}\text{s}^{-1}$.⁴¹ These slow reaction rates make this impractical for widespread labeling of proteins in living cells.

As far as the ideal reaction for bioorthogonal labeling of macromolecules goes, each chemistry has unique advantages and disadvantages. At the moment, the inverse electron demand Diels Alder (IEDDA) reaction between tetrazines and strained alkenes has become the most popular chemistry for performing ligation reactions *in vivo*^{18,34,51} largely for its high reaction rates and stability.

Inverse Electron Demand Diels Alder Reactions

The IEDDA reaction between tetrazines and strained alkenes was first used as a bioorthogonal ligation in 2008^{36,52} and consists of a [4 + 2] cycloaddition reaction between an electron deficient tetrazine and an electron rich alkene or alkyne^{53,54} (Fig 1.1D). This contrasts with the standard Diels Alder reaction which consists of reactions between electron rich dienes and electron deficient alkenes. In both cases, the reaction can be driven in the forward direction by substitution of the functional groups with electron withdrawing and donating substituents. Whereas, in the normal electron demand Diels Alder reaction, dienes substituted by electron donating groups react fastest, substitution of dienes by electron donating groups slows the IEDDA reaction. The same is true for substitution of the dienophile with electron donating groups.

The IEDDA reaction most commonly consists of the reaction between trans-cyclooctene (TCO) functional groups and tetrazines, but functional groups such as cyclopropenes⁵⁵, norbornenes^{10,56,57}, and spirohexenes⁵⁸ (Fig 1.2A) have been used. Due to substitution of tetrazines with electron withdrawing and donating groups, the IEDDA reactions of the same strained alkene with different tetrazines can result in a difference of several orders of magnitude in the second order reaction rates^{41,55,58}. Additionally, as initially reported by the Joseph Fox, the introduction of additional strain to the alkene-containing rings can result in a further increase in the second order reaction rate. This has been used to generate strained alkenes including dioxolane fused TCO (dTTCO)⁵⁹, strained TCO (sTCO)⁶⁰, and bicyclononynes (BCN)⁶¹ that show higher second order rate constants in their reactions with tetrazines than their unstrained counterparts.

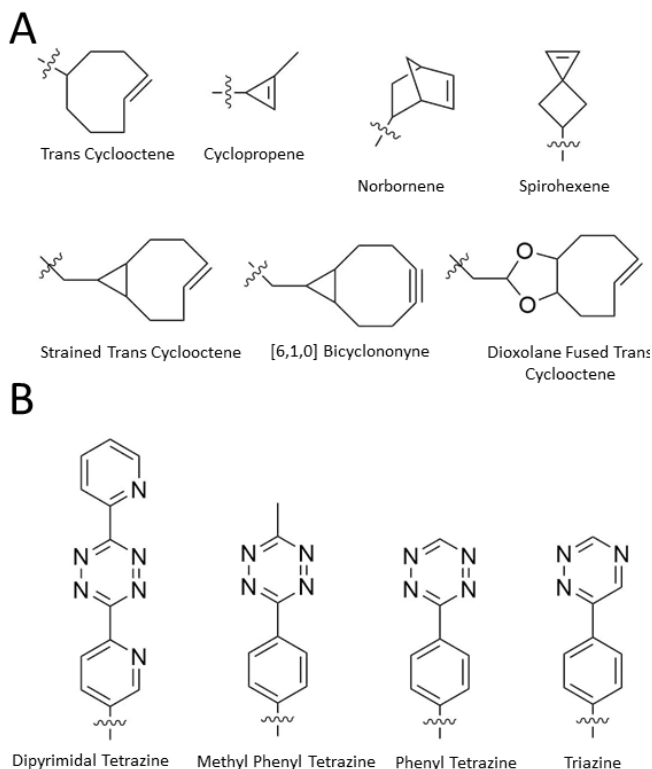


Fig 1.2. Structures of Common IEDDA Functional Group.

A) Structures of commonly used strained alkene functional groups. Strained Trans Cyclooctene and Dioxane Fused Trans Cyclooctene are used in this work. B) Structures of commonly used tetrazines and dienes in IEDDA reactions.

There is some structural variation in the functional groups capable of performing IEDDA chemistry. Structural variation in tetrazine is largely limited to substitutions of 1,2,4,5-tetrazines with substituents with varying degrees of electron withdrawing character⁶². Triazines have been shown to undergo these reactions as well⁶³, though they are less electron withdrawing and have lower rate constants. Substitution of tetrazines has been explored to a limited extent. Tetrazines substituted by 3,6-dipyrimidal tetrazines, 3-phenyl tetrazines, and 3,6-phenylmethyl tetrazines have been used in bioorthogonal reactions(Fig 1.2B)⁶². Most IEDDA functional group variation lies in the strained alkene.

The great structural diversity in the functional groups that undergo IEDDA reactions results in a variety of reaction properties. IEDDA reactions can have second order rate constants spanning from 10^{-6} to $10^{-2} \text{ M}^{-1}\text{s}^{-1}$.^{35,41,63} Functional groups with higher rate constants often undergo degradation *in vivo*^{35,38,64}. Because of this, use of

extremely reactive functional groups may not be appropriate for reactions inside cells. The functional group used in IEDDA reactions should be selected to match the desired use. As an example, the fastest bioorthogonal reactions published are the reactions between silanated trans-cycloheptenes and tetrazines. These reactions can have a second order rate constant of up to $1.14 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$.⁶⁵ The silanated trans-cycloheptenes degrade *in vivo* which limits their use to *in vitro* reactions.

The Use of Tetrazines in Bioorthogonal Reactions

There are many uses of IEDDA reactions in live cells in which the strained alkene functional group is attached to the biomolecule of interest and subsequently reacted with a tetrazine containing label^{4,34,59} (Fig 1.3A). There are a few advantages to using this approach of tetrazine functional groups as labels including the fact that tetrazines undergo nucleophilic attack by thiols in biological environments^{52,66}. By attaching the tetrazine ring to the label the exposure of the tetrazine to biological environments and off-target reactions is minimized. Tetrazines often quench the fluorescence of common fluorescent labels, and then there is a subsequent increase in fluorescence upon labeling the biomolecule of interest¹⁷. This can result in greater signal to noise ratios for cellular fluorescent labeling.

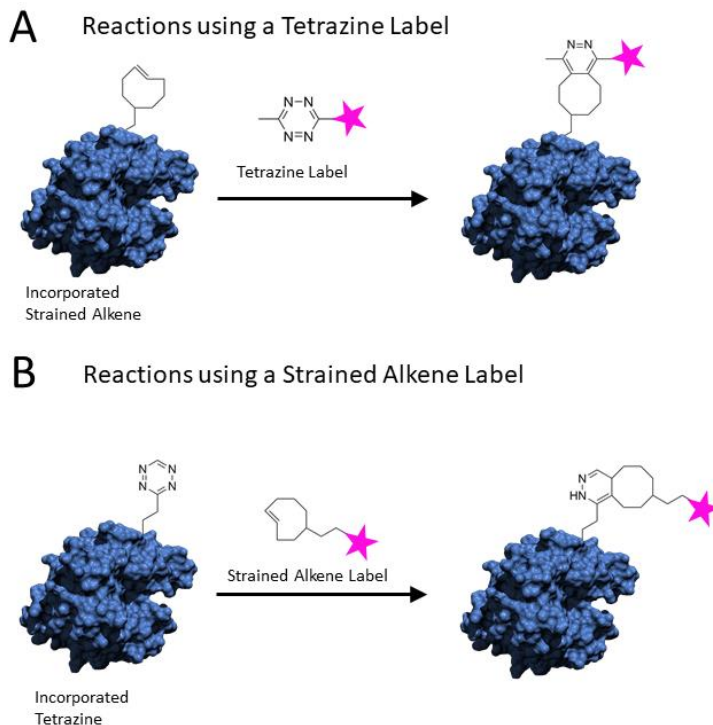


Fig 1.3. Labeling Biomolecules with IEDDA Reactions

A) The incorporation of a strained alkene onto a biomolecule require subsequent reaction a tetrazine label. B) The incorporation of a tetrazine onto a biomolecule require subsequent reaction with a strained alkene label.

The opposite approach in which a tetrazine functional group is incorporated onto biomolecules and subsequently reacted with a strained alkene containing label is relatively rare (Fig 1.3B). The fastest bioorthogonal ligations reported to date are the reactions of highly strained alkenes with tetrazines. Trans-cyclooctenes like sTCO are unstable under physiological conditions and can degrade via isomerization to form the corresponding cis-cyclooctene (strained cis-cyclooctene (sCCO)^{64,65} for the isomerization of sTCO). This isomerization event limits the ability to incorporate sTCO onto biomolecules *in vivo*. To enable these faster bioorthogonal reactions, an IEDDA reaction with a tetrazine functional group incorporated onto the biomolecule should be employed³⁴.

Methods of Introducing Bioorthogonal Functional Groups

One challenge of employing bioorthogonal ligation chemistry in live cells is the introduction of the desired bioorthogonal functional group onto the biomolecule of interest. Many strategies have been employed to achieve this, each of which is specific to the class of biomolecule to be modified. The introduction of bioorthogonal functional groups to DNA is often performed *in vitro* during oligonucleotide synthesis reactions⁶⁷. The modified DNA can then be introduced into cells via transformation or transfection. The introduction of bioorthogonal functional groups to RNA is commonly performed through hijacking the 5' capping enzyme for modifying mRNA⁶⁸. Modification of lipid membranes can be relatively simple as lipids containing bioorthogonal handles will incorporate into cellular membranes¹³. The introduction of bioorthogonal functional groups into carbohydrates is commonly performed using native or repurposed glycosylation enzymes to introduce or extend glycans containing the functional group of interest^{69,70}.

In contrast, proteins can be a difficult class of biomolecules to modify through the introduction of bioorthogonal ligation handles. As such, numerous strategies have been developed to introduce these handles (Fig 1.4), each method having its own strengths and weaknesses. Translationally fusing the protein of interest to enzymatic tags such as the Halo tag is one such method. Halo tag is a 34 kDa protein that reacts with reagents containing a chloroalkane group to form a covalent bond⁷¹. Halo tag is relatively large and has the potential to impact the function of the protein to which it is fused. Because Halo tag is a genetic fusion, it is limited to modification of the N-terminus, C-terminus, or a loop region.

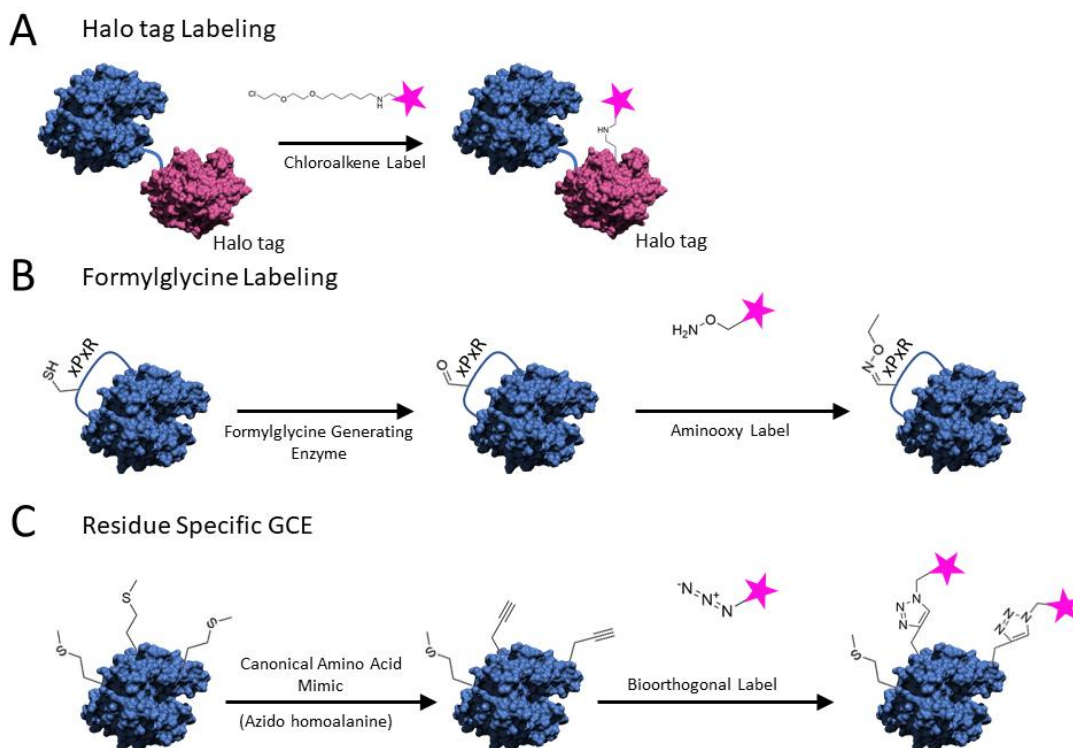


Fig 1.4. Methods of Introducing Bioorthogonal Functional Groups.

A) Halo tag labeling requires the genetic fusion with a Halo tag protein. Addition of chloroalkene reagents results in modification. B) Formylglycine labeling requires the insertion of a CxPxR motif that is recognized by formylglycine generating enzyme. Upon formation of formylglycine, reaction with aminoxy labels are possible. C) Residue specific GCE utilizes native aminoacyl tRNA synthetases to incorporate structurally similar amino acids in a stochastic pattern. Labeling reagent is dependent on the incorporated amino acid.

Site specific introduction of bioorthogonal ligation handles on proteins is a problem that has been addressed by Carolyn Bertozzi wherein the insertion of a CxPxR motif⁷ and subsequent treatment with formylglycine generating enzyme results in the conversion of the motif cysteine to formylglycine^{7,72}. This formylglycine amino acid can then be reacted with aminoxy functional groups. As a method of introducing bioorthogonal handles, this system is advantageous in that the formylglycine generating enzyme can be genetically encoded, and the recognized motif is small and can be inserted in a wider variety of sites than Halo tags. A major disadvantage of this system includes that it is limited to reactions of formylglycine and aminoxy functional groups whose chemistry suffers from low rate constants and large degrees of off target reactions.

Genetic Code Expansion (GCE) techniques have been used to incorporate structurally diverse bioorthogonal functional groups into proteins in the form of noncanonical amino acids (ncAAs). Two main methods of GCE have been developed with unique advantages: residue-specific GCE and site-specific GCE. Residue specific GCE consists of the addition to cells of an amino acid that is structurally similar to canonical amino acids^{5,73}. Aminoacyl tRNA synthetases (aaRSs) responsible for charging tRNAs with canonical amino acid are permissive to the ncAA of interest and incorporate it into proteins at codons specific for the canonical amino acid⁷⁴. Residue-specific GCE is an excellent method for studying newly synthesized proteins and protein turnover^{5,74,75}. While residue-specific GCE has the advantage of specific incorporation into newly synthesized proteins, protein products are not chemically homogenous in that sites may contain the bioorthogonal functional group of interest or the canonical amino acid. Residue-specific GCE is also incapable of modifying a single protein, and instead results in the modification of every cellular protein containing the residue of interest. Finally, since residue specific GCE uses the native aaRSs, incorporation is limited to ncAAs structurally similar to canonical amino acids. In contrast, site-specific GCE consists of the introduction of an ncAA containing the desired bioorthogonal functional group at a single site on a protein⁷⁶.

Site-Specific Genetic Code Expansion

Site-specific genetic code expansion was pioneered by Schultz⁷⁶ and Furter⁷⁷ in the late 1990s. It consists of selectively incorporating an ncAA of interest into proteins usually through the suppression of a stop codon^{78,79} though rare sense codons are also targets⁸⁰. This technique has been used to incorporate hundreds⁸¹⁻⁸³ of structurally unique amino acids into a variety of organisms including *e. coli*⁷⁶, *s. cerevisiae*⁸⁴, *d. melanogaster*⁸⁵, and *m. musculus*⁸⁶ among others. While bioorthogonal functional groups are an important subset of the ncAAs that have been successfully incorporated, ncAAs have also been designed to mimic posttranslational modifications,⁸⁷⁻⁸⁹ perform photocrosslinking reactions^{90,91}, and explore effects of minute perturbations to amino acid structure⁹².

Site-specific GCE takes advantage of natural ribosomal translation to produce proteins containing ncAAs (Fig 1.5). The ribosome accepts tRNAs charged with the ncAA of interest in response to an amber stop codon. If no ncAA is present, translation is terminated at the amber codon by release factor 1 resulting in termination of translation⁹³. To accomplish this incorporation, four components are necessary:

- 1) The amino acid of interest added to cells.
- 2) An expressed orthogonal tRNA molecule with an anticodon complementary to the stop codon interrupting the gene of interest and that is incapable of being aminoacylated by native aaRSs.
- 3) An aaRS that uses the amino acid of interest as a substrate to aminoacylate the orthogonal tRNA but cannot use canonical amino acids as substrates. Additionally, this aaRS must not charge any of the native tRNAs as substrates.
- 4) The gene of interest interrupted by a stop codon (usually amber stop codons).

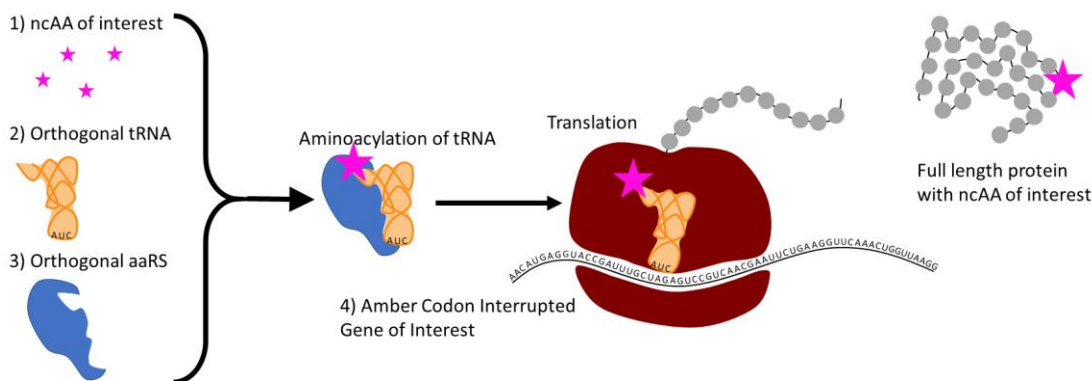


Fig 1.5. Site Specific Genetic Code Expansion.

Site specific GCE relies on the presence of a ncAA of interest which is charged onto an orthogonal tRNA via an orthogonal aaRS. Once charged with the ncAA of interest, the tRNA can be used in the ribosome to decode amber stop codons in the gene of interest as the amino acid of interest. Together this results in the production of full length protein containing the amino acid of interest at the site of amber stop codons.

Orthogonal tRNA/Aminoacyl tRNA Synthetase Pairs

Together the combination of generated aaRS and tRNA are considered orthogonal pairs and are generally developed in tandem with one another. The active site of the aaRS typically interacts only with the 3' end of tRNAs. Modifications to the active site do not negatively impact tRNA binding⁸¹. This allows orthogonal pairs, once generated, to incorporate a variety of ncAA structures. The ability to recognize the cognate tRNA is distributed throughout the aaRS⁸¹ making the generation of orthogonal tRNA/aaRS pairs difficult. As such relatively few orthogonal tRNA/aaRS pairs have been developed⁹⁴, but those tRNA/aaRS pairs have been extensively used to incorporate hundreds of amino acids⁸³. Furthermore, while these tRNA/aaRS pairs are often termed orthogonal pairs, they are rarely if ever orthogonal to all living systems. Often a tRNA/aaRS pair is taken from one organism and cloned into another organism to generate an orthogonal system. These tRNA/aaRS pairs are orthogonal to the organism they are cloned into, but not to organism in which they originated.⁸⁴ As an example of this, the tyrosyl tRNA/aaRS pair from *Methanocaldococcus jannaschii* is orthogonal to *E. coli* and other bacteria⁷⁶, yet demonstrates cross reactivity with native tRNA/aaRS pairs in archaeal and eukaryotic cells⁸⁴.

The *M. jannaschii* tyrosyl synthetase pair was the first such tRNA/aaRS pair developed⁷⁶, and as it was based on a tyrosine incorporating active site, many of the ncAAs that have been successfully incorporated with it have a similar structure⁸³. Very few, if any, amino acids have been incorporated with *M. jannaschii* system that are not based on a tyrosine or phenylalanine base⁸³. For incorporation of amino acids in eukaryotic systems, the *Methanosarcina barkeri* and the *Methanosarcina mazei* pyrrolysyl aaRS are the most commonly used enzymes. These proteins are homologous to one another and have high degrees of sequence similarity in the active site such that when mutations to the *M. barkeri* active site resulting in the ability to incorporate acetyl-lysine were made in the active site of *M. mazei*, the ability to incorporate the acetyl-lysine was also conferred to *M. mazei*⁹⁵. These enzymes are commonly grouped together as the PylRS system. It is of note that only the *M. mazei* active site has been

successfully crystalized, though homology suggests a similar structure for the *M. barkeri* active site^{81,95,96}.

Other synthetase systems have been developed such as the *E. coli* tyrosyl⁹⁷ aaRS as well as an *E. coli* tryptophanyl⁹⁸ aaRS both of which are orthogonal to eukaryotic cells. These systems have potential due to their orthogonality to eukaryotic cells as well as orthogonality to specific cell lines of bacteria. These systems can be used alongside the more rigorously characterized PylRS system as their active sites are more amenable to tyrosine-based and tryptophan-based structures, respectively, making selections for these amino acids more fruitful.

Generating Aminoacyl tRNA Synthetases for Incorporating Amino Acids

The process of selecting aaRS mutants capable of incorporating the amino acid of interest can be a bottleneck to genetically encoding the amino acid. The process typically involves generating large libraries of mutants by mutating residues in the active site of the aaRS. Residues are typically mutated to codons with nucleotide sequence NNK where N is any nucleotide and K is guanine or thymine. In doing so, all possible amino acids as well as amber stop codons (TAG/UAG) are present at each mutated codon. While more effective methods exist^{99,100}, the NNK is the most common. Typically, the number of codons mutated is limited to five per library generated as this results in a library size of approximately 33.5 million.. Given that the largest bottleneck in the selection process is transformation of library members into *E. coli* cells, extensive library sizes can result in incomplete library transformation. Typically, the number of cells transformed should be 10-100 fold greater than the nucleotide size of the library, because a 25 fold coverage of a library this size ensures a >99.9% chance of covering the entire library assuming an equal probability of colony formation for each library member.

When generating a library for the selection process, the decision of which residues in the aaRS to mutate is based on knowledge of the aaRS structure and activity. Residues are generally selected that make-up of the amino acid binding pocket. Residues that are essential for catalysis are avoided as mutations to these residues

cannot be performed without impacting the ability to aminoacylate tRNA. Residues that are distant from the binding pocket should be avoided for mutations as the impact of these residues on ncAA binding is not clear. To make these evaluations a crystal structure or other structural information regarding the native aaRS is required¹⁰¹.

The selection processes usually consist of a double sieve selection with alternating rounds of positive and negative⁷⁶ selection (Fig 1.6). In the positive selection, a life/death selection is performed in which the library is co-transformed with a plasmid encoding for an amber codon interrupted antibiotic resistance gene (usually chloramphenicol acetyl transferase). The transformed cells are then plated on media containing chloramphenicol as well as the amino acid of interest. Functional synthetases capable of incorporating either the amino acid of interest or a canonical amino acid suppresses the amber stop codon resulting in the production of full length antibiotic resistance gene and therefore cell survival. Historically, chloramphenicol concentration has been varied to tune the stringency of such selections^{76,102}. Negative selections consist of transformation of the library members remaining after the first positive selection into cells containing a barnase gene interrupted by multiple amber codons. The cells are then plated on media in the absence of the amino acid of interest. Production of full length barnase results in cleavage of mRNA resulting in cell death. Library members that incorporate canonical amino acids are expected to result in the production of barnase and ultimately cell death. Multiple rounds of positive and negative selection are typically performed, though it has been found that too many rounds of selection can result in lower quality synthetases¹⁰³.

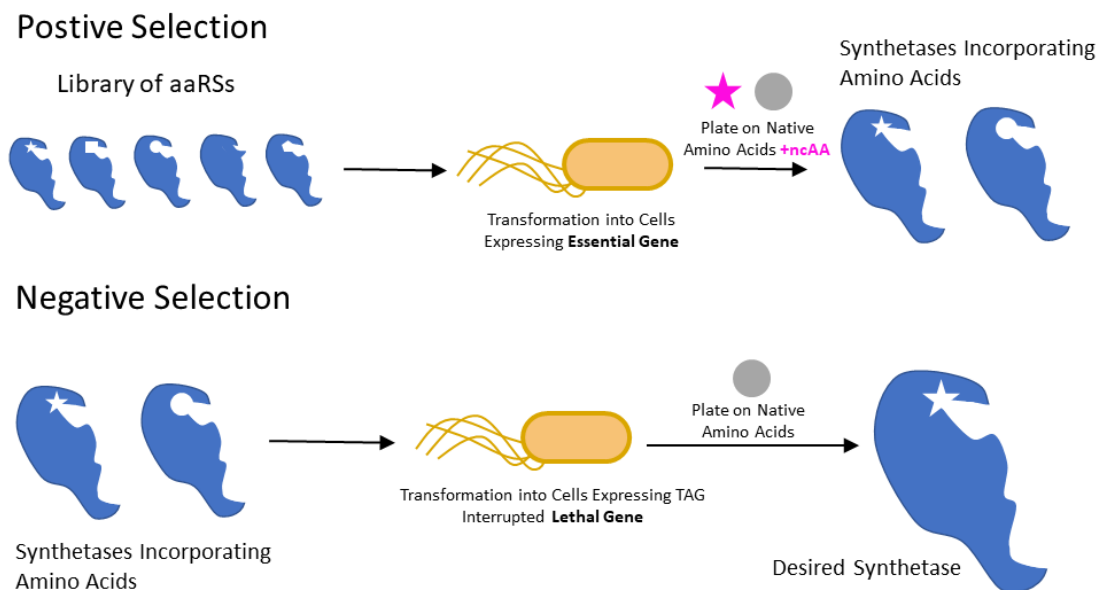


Fig 1.6. Selections for aaRSs Incorporating an Amino Acid of Interest.

Selections require alternate positive and negative rounds of selection where a library of synthetases is reduced to a few desired synthetases by eliminating synthetases that use native amino acids as substrates and synthetases that are inactive with native amino acids or the amino acid of interest. The positive selection steps result in the elimination of synthetases that are inactive with native amino acids or the amino acid of interest, while the negative selection eliminates synthetases that use native amino acids as substrates.

While this selection scheme is the most common method of generating synthetases, it is by no means the only method. Bioorthogonal functional group containing amino acids have been selected for via reaction with biotin containing labels after cell surface display¹⁰⁴. Others have turned to deep sequencing to analyze increases in frequency of desirable hits¹⁰⁵. Finally, fluorescence activated cell sorting is a technique that has been used to evaluate suppression of fluorescence reporter genes in selections¹⁰⁶. Additionally, many have turned to directed evolution approaches as opposed to more traditional library selection approaches^{107,108}.

Disseration Overview

The objective of my thesis work was to develop tetrazine-based bioorthogonal reactions for use on proteins in living systems. A genetic code expansion approach was undertaken in large part due to the site-selective nature of this system, the ability to

incorporate a number of structural variants of tetrazine, and to minimally perturb the incorporated protein's function. In addition to industrial and pharmaceutical applications, a robust and well characterized ability to rapidly label proteins *in cellulo* would enable precise control of protein function, the ability to perturb cellular pathways in a controlled manner and would provide a useful tool for fluorescent imaging of proteins *in cellulo*. Chapters 2, 3, and 4 of this dissertation address key steps in encoding tetrazine-containing ncAAs in prokaryotes and eukaryotes. Chapter 2 introduces the tetrazine chemistry used and demonstrates that this chemistry demonstrates ideal reaction rates between tetrazines and strained alkenes when labeling proteins in live *E. coli* cells. Chapter 3 overcomes difficulties associated with genetically encoding and reacting tetrazine-containing amino acids in eukaryotic cells. Chapter 4 explores side reactions of tetrazine-containing amino acids and methods to control reactivity to ensure the labeling reactions behave ideally. Chapter 5 describes future experiments that should be performed as direct extensions of this work as well as future directions that this work may go. Of the chapters presented here, chapter 2 has been published, while the work presented in chapters 3 and 4 are in the process of being written for future publication.

Chapter 2

Ideal Bioorthogonal Reactions Using A Site-Specifically Encoded Tetrazine Amino Acid

Robert J. Blizzard, Dakota R. Backus, Wes Brown, Christopher G. Bazewicz, Yi Li,
Ryan A. Mehl

Published in *J. Am. Chem. Soc.*, 2015, 137, pp 10044-10047
© 2015 American Chemical Society. All rights reserved.

Abstract

Bioorthogonal reactions for labeling biomolecules in live cells have been limited by slow reaction rates or low component selectivity and stability. Ideal bioorthogonal reactions with high reaction rates, high selectivity, and high stability would allow for stoichiometric labeling of biomolecules in minutes and eliminate the need to wash out excess labeling reagent. Currently, no general method exists for controlled stoichiometric or sub-stoichiometric labeling of proteins in live cells. To overcome this limitation, we developed a significantly improved tetrazine-containing amino acid (Tet-v2.0) and genetically encoded Tet-v2.0 with an evolved aminoacyl-tRNA synthetase/tRNA_(CUA) pair. We demonstrated *in cellulo* that protein containing Tet-v2.0 reacts selectively with cyclopropane-fused trans cyclooctene (sTCO) with a bimolecular rate constant of $72,500 \pm 1660 \text{ M}^{-1}\text{s}^{-1}$ without reacting with other cellular components. This bioorthogonal ligation of Tet-v2.0-protein reacts *in cellulo* with sub-stoichiometric amounts of sTCO-label fast enough to remove the labeling reagent from media in minutes, thereby eliminating the need to wash out label. This ideal bioorthogonal reaction will enable the monitoring of a larger window of cellular processes in real time.

Introduction

The development of bioorthogonal reactions and strategies to apply them in the study of biopolymers has transformed our ability to study and engineer biomolecules. The early successes of this technology inspired nearly two decades of research toward building faster and more selective reactions^{39,41,109}. The broadly defined bioorthogonal reaction is a selective reaction between functional groups in the presence of biological entities. Great progress has been made at increasing the rate and selectivity of bioorthogonal reactions, but the vast majority of reactions still cannot be used inside living cells because: i) high molecular concentrations in cellular environments increase off target side-reactions ii) the reactive functional groups introduced compromise the cellular reducing environment and/or catalytic processes iii) the cell interior is challenging to access efficiently with the necessary functionalized molecules.^{39,41,109} A

few chemoselective reactions have cleared the more stringent *in cellulo* hurdle, but their sluggish reaction rates prevent utility.^{110,111} The ideal bioorthogonal reaction which functions *in cellulo* with quantitative yields at low concentrations and with exquisite chemoselectivity is said to represent the holy grail of chemical synthesis.¹¹²

The goal of the ideal bioorthogonal reaction should be to label molecules *in cellulo* faster than the rate constants of cellular processes but without side reactions or degradation of reagents. To compete effectively with cellular processes, ideal bioorthogonal reactions need i) fast kinetics ($>10^4 \text{ M}^{-1}\text{s}^{-1}$) to react completely on biological time scales of seconds to minutes and to function at biological concentrations (μM to nM) of both biomolecule and label, ii) high selectivity to ensure only the target biomolecules are modified, iii) functional groups stable enough to enable the labeling of quantitative portions of biomolecules *in vivo* and iv) small structural components as to not adversely affect the structure and function of the biomolecule under investigation.

As defined, ideal bioorthogonal reactions would enable access to new scientific inquiry because they could turn on or trap typical biological events *in vivo* at rates comparable to enzymatic reactions (typically 10^3 - $10^6 \text{ M}^{-1}\text{s}^{-1}$). In addition, many applications, such as delivery of visual probes in organisms for nuclear medicine, single molecule spectroscopy, and fluorescent imaging, demand extremely fast reaction rates because low concentrations of labeling reagents are required.^{41,109,113} The ideal bioorthogonal reaction presented here will allow short reaction times even at sub-stoichiometric concentrations of labeling reagents. The use of stoichiometric concentrations of labeling reagent reduces background signal and side reactions from excessive unreacted label.

An exciting class of bioorthogonal ligations, inverse-electron demand Diels-Alder (IED-DA), posts rate constants up to $10^6 \text{ M}^{-1}\text{s}^{-1}$ between tetrazines and strained trans cyclooctenes.^{52,59} Current functional groups that provide these exceptional rates lack the *in vivo* stability and selectivity to meet the requirements of the ideal bioorthogonal reaction. More stable transcyclooctene (TCO)-containing amino acids have been site-specifically incorporated into proteins by using genetic code expansion and react *in vivo* with dipyrimidyl-tetrazines, showing labeling rates of $5200 \text{ M}^{-1}\text{s}^{-1}$.^{17,111}

Unfortunately, when the reaction rate is increased by adding electron-withdrawing groups to the tetrazine or strain to TCO, these components lose significant *in vivo* selectivity. The commonly used 3-phenyl-s-tetrazine and 3,6-(dipyridin-2-yl)-s-tetrazine are extremely reactive with strained alkenes but act as electrophiles for cellular thiols.^{37,64} A strained version of transcyclooctene, sTCO, (cyclopropane-fused transcyclooctene) is also not compatible with genetic code expansion as an amino acid because its isomerization *in vivo* results in a half-life of 0.67 days.^{52,59,64} If instead, a modestly active tetrazine amino acid is encoded into the protein, the short half-life of sTCO is acceptable because the sTCO-attached labelling reagent will be consumed prior to significant decomposition.

Supplemental

Synthesis of 4-(6-methyl-s-tetrazin-3-yl)phenylalanine (Tet-v2.0)

(Fig 2.1) *N-(tert-Butoxycarbonyl)-4-(6-methyl-1,2,4,5-tetrazin-3-yl)-L-phenylalanine* (**2**)¹ A dry, 75 mL heavy walled reaction tube was equipped with a stir bar and was charged with **1** (500 mg, 1.72 mmol), Ni(OTf)₂ (307.3 mg, 0.86 mmol), and acetonitrile (1.8 mL, 34.4 mmol). The flask was purged with argon for 20 min. Anhydrous hydrazine was added to the mixture (2.7 mL, 86 mmol), the tube was sealed, and the reaction mixture was heated to 60°C for 24 hr. The reaction was allowed to cool to room temperature and slowly opened to air. Sodium nitrite (2 M, 8 mL) was added to the flask and the contents were cooled to 0°C. 1 N HCl was added slowly until gas evolution ceased and the pH of the mixture was acidic (~3). The mixture was then diluted with EtOAc and the layers separated. The aqueous layer was extracted with EtOAc (2x). The combined organic layers were washed with brine, dried with Na₂SO₄, and concentrated under reduced pressure. Silica gel flash column chromatography (50% ethyl acetate in hexanes with 1% acetic acid) yielded 370 mg of **2** (1.03 mmol, 60%) in the form of a red oil. R_f=0.33 in 50% ethyl acetate in hexanes with 1% acetic acid; ¹H NMR (400 MHz, CDCl₃) δ 8.53 (d, 2H), 7.39 (d, 2H), 5.0 (d, 1H), 4.66 (m, 1H), 3.37-3.15 (dd, 2H), 3.09 (s, 3H), 1.39 (s, 9H).

4-(6-methyl-1,2,4,5-tetrazin-3-yl)-L-phenylalanine hydrochloride salt (3) A dry 100 mL flask containing **2** (480 mg, 1.34 mmol) was charged with and then purged with argon for 10 min. This oil was dissolved in 4 mL of 1,4 dioxane then 4 M HCl in 1,4 dioxane (6 mL) was added over 20 sec. The reaction was allowed to stir for 12 hours at room temperature. The reaction was concentrated under reduced pressure, then dissolved in EtOAc, and concentrated down to a red powder. The red solid was washed with pentanes and collected via filtration to afford 377 mg of **Tet-v2.0** (1.18 mmol, 95%). ¹H NMR (400MHz, CD₃OD) δ 8.48 (d, 2H), 7.48 (d, 2H), 4.29 (t, 1H), 3.41-3.20 (m, 2H), 2.97 (s, 3H) (Fig 2.2). The product was further characterized via ¹³C NMR and Mass Spectrometry (Fig 2.3 and 2.4).

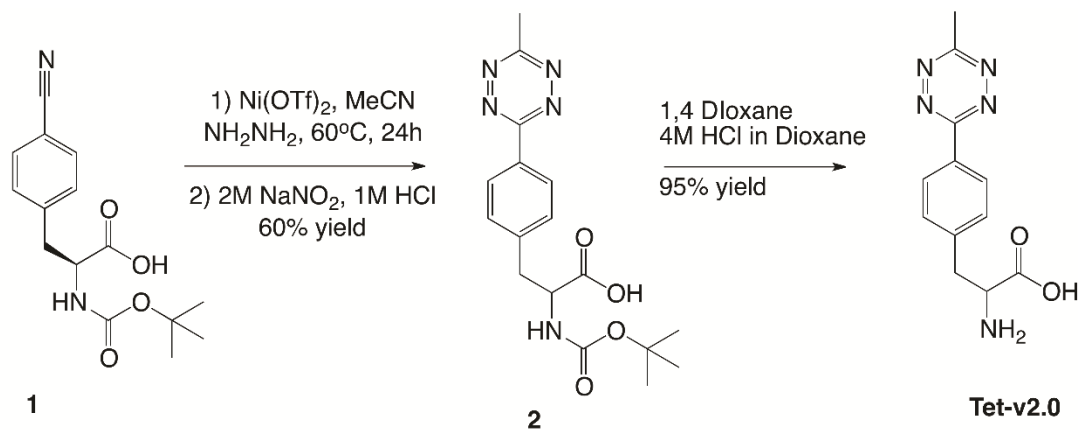


Fig 2.1. Synthesis of 4-(6-methyl-*s*-tetrazin-3-yl)phenylalanine (Tet-v2.0).

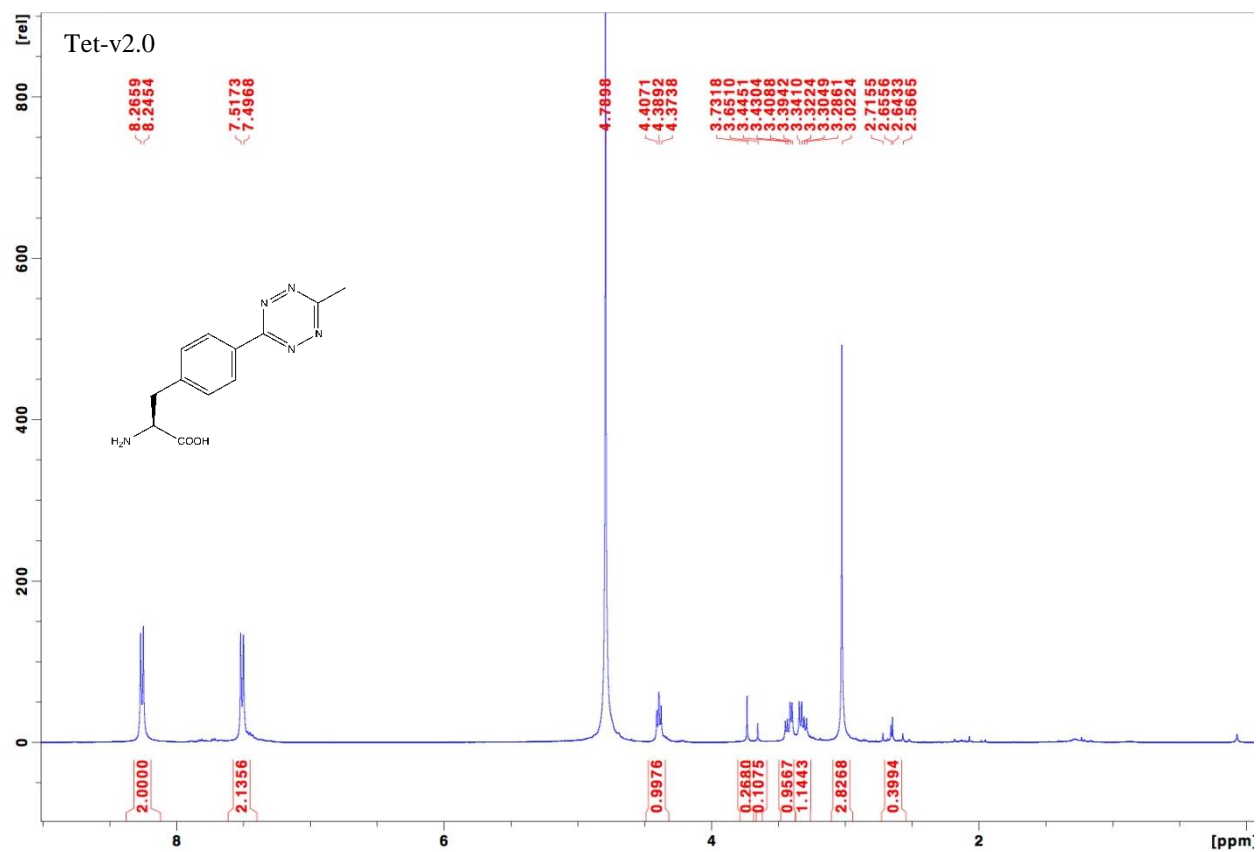


Fig 2.2. ^1H NMR of 4-(6-methyl-*s*-tetrazin-3-yl)phenylalanine (Tet-v2.0).

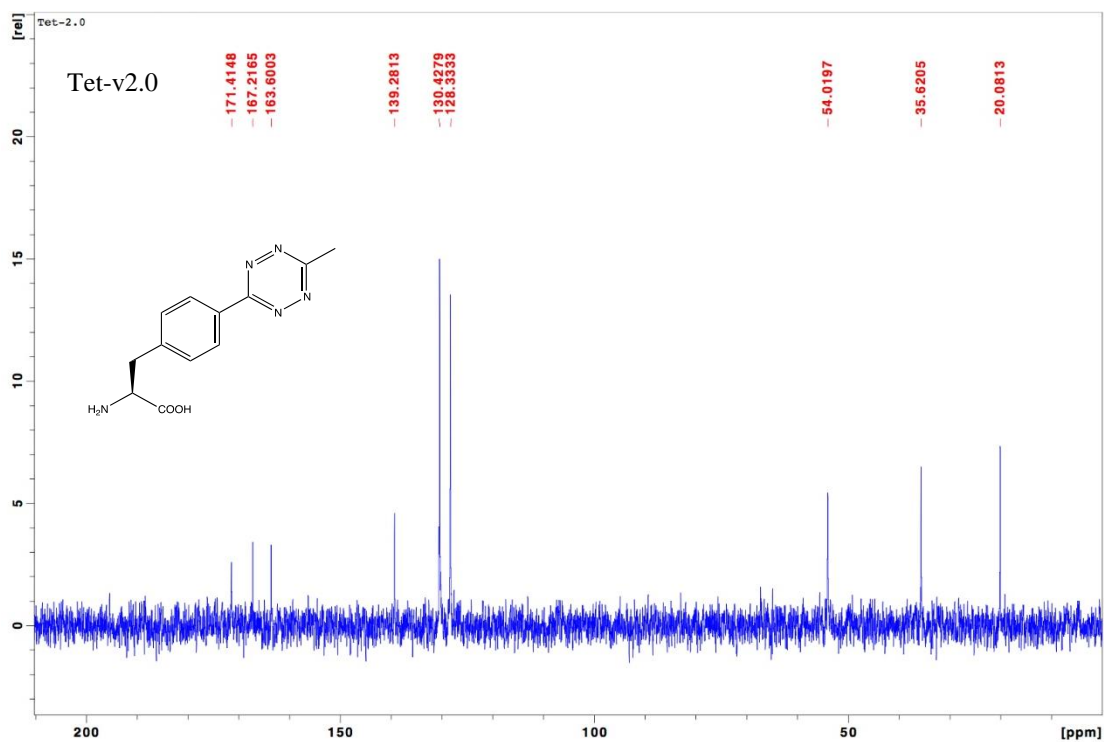


Fig 2.3. ^{13}C NMR of 4-(6-methyl-*s*-tetrazin-3-yl)phenylalanine (Tet-v2.0).

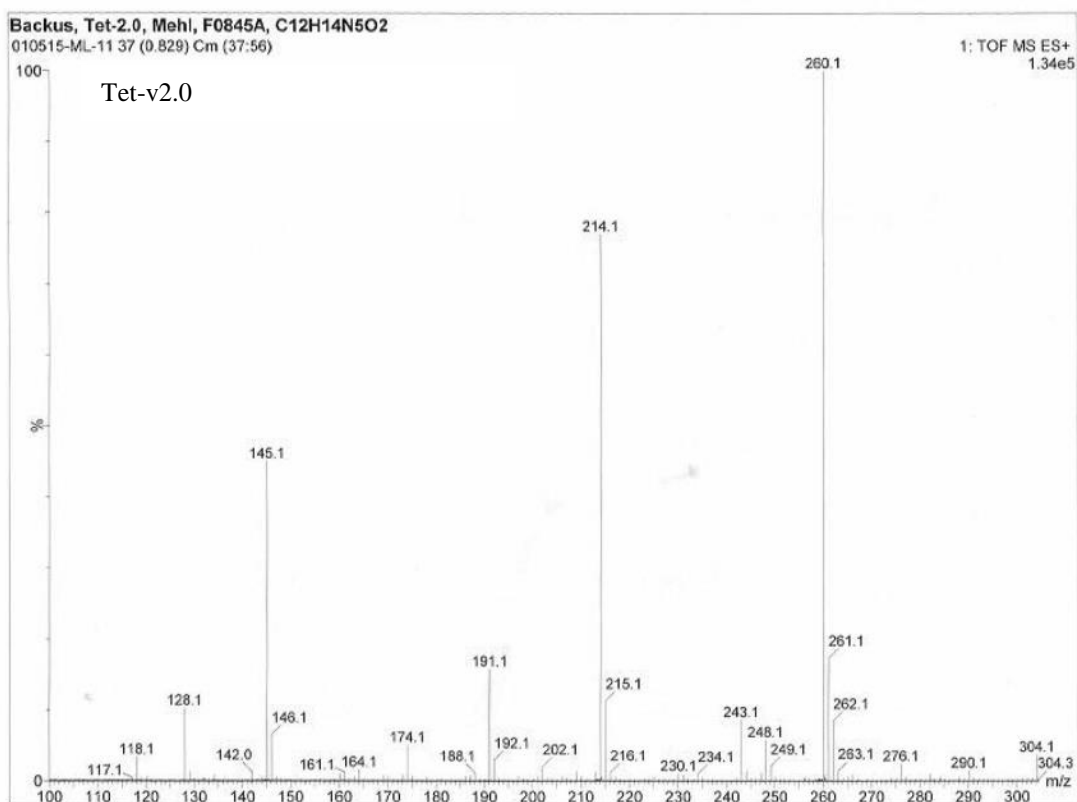


Fig 2.4. MS of 4-(6-methyl-s-tetrazin-3-yl)phenylalanine (Tet-v2.0).

Selection of aminoacyl-tRNA synthetases specific for Tet-v2.0

The library of aminoacyl-tRNA synthetases was encoded on a kanamycin (Kn) resistant plasmid (pBK, 3000 bp) under control of the constitutive *Escherichia coli* GlnRS promoter and terminator. The aminoacyl synthetase library (3D-Lib) was randomized as follows: Leu65, His70, Gln155, and Ile159 were randomized to all 20 natural amino acids; Tyr32 was randomized to 15 natural amino acids (less Trp, Phe, Tyr, Cys, and Ile); Asp158 was restricted to Gly, Ser, or Val; Leu162 was restricted to Lys, Ser, Leu, His, and Glu; and Phe108 and Gln109 were restricted to the pairs Trp-Met, Ala-Asp, Ser-Lys, Arg-Glu, Arg-Pro, Ser-His, or Phe-Gln. The library plasmid, pBK-3D-Lib, was moved between cells containing a positive selection plasmid (pCG) and cells containing a negative selection plasmid (pNEG).

The positive selection plasmid, pCG (10000 bp), encodes a mutant *Methanococcus jannaschii* (Mj) tyrosyl-tRNA_{CUA}, an amber codon-disrupted chloramphenicol acetyltransferase, an amber codon-disrupted T7 RNA polymerase, a T7 promoter controlled green fluorescent protein gene, and the tetracycline (Tet) resistance marker. The negative selection plasmid, pNEG (7000 bp), encodes the mutant tyrosyl-tRNA_{CUA}, an amber codon-disrupted barnase gene under control of an arabinose promoter and *rrnC* terminator, and the ampicillin (Amp) resistance marker. pCG electrocompetent cells and pNEG electrocompetent cells were made from DH10B cells carrying the respective plasmids and stored in 100 μ L aliquots at -80 °C for future rounds of selection.

The synthetase library in pBK-3D-Lib was transformed by electroporation into DH10B cells containing the positive selection plasmid, pCG. The resulting pCG/pBK-3D-Lib-containing cells were amplified in 1 L of 2 \times YT with 50 μ g/mL Kn and 25 μ g/mL Tet with shaking at 37 °C. The cells were grown to saturation, then pelleted at 5525 rcf, resuspended in 30 mL of 2 \times YT and 7.5 mL of 80% glycerol, and stored at -80 °C in 1 mL aliquots for use in the first round of selections.

For the positive selection, 2 mL of pCG/pBK-3D-Lib cells were thawed on ice before addition to 1.2 L of room temperature 2 \times YT media containing 50 μ g/mL Kn and 25 μ g/mL Tet. After incubation (11 h, 250 rpm, 37 °C), a 200 μ L aliquot of these cells was plated on eleven 15 cm GMMML-agar plates containing 50 μ g/mL Kn, 25 μ g/mL Tet, and 60 μ g/mL chloramphenicol (Cm). The positive selection agar medium also contained 1 mM **3**. After spreading, the surface of the plates was allowed to dry completely before incubation (37 °C, 15 h). To harvest the surviving library members from the plates, 10 mL of 2 \times YT (50 μ g/mL Kn, 25 μ g/mL Tet) was added to each plate. Colonies were scraped from the plate using a glass spreader. The resulting solution was incubated with shaking (60 min, 37 °C) to wash cells free of agar. The cells were then pelleted, and plasmid DNA was extracted. For the first positive selection a Qiagen midiprep kit was used to purify the plasmid DNA. For all other plasmid purification steps a Thermo Scientific miniprep kit was used to purify the plasmid DNA. The smaller pBK-3D-Lib plasmid was separated from the larger pCG plasmid by agarose

gel electrophoresis and extracted from the gel using the Thermo Scientific gel extraction kit.

The purified pBK-3D-Lib was then transformed into pNEG-containing DH10B cells. A 100 μ L sample of pNEG electrocompetent cells was transformed with 50 ng of purified pBK-3D-Lib DNA. Cells were rescued in 1 mL of SOC for 1 h (37 °C, 250 rpm) and the entire 1 mL of rescue solution was plated on three 15 cm LB plates containing 100 μ g/mL Amp, 50 μ g/mL Kn, and 0.2% L-arabinose. Cells were collected from plates and pBK-3D-Lib plasmid DNA was isolated in the same manner as described above for positive selections.

In order to evaluate the success of the positive and negative selection based on variation in synthetase efficacy (as opposed to traditional survival/death results), the synthetases resulting from the selection rounds were tested with the pALS plasmid. This plasmid contains the sfGFP reporter with a TAG codon at residue 150 as well as tyrosyl-tRNA_{CUA}. When a pBK plasmid with a functional synthetase is transformed with the pALS plasmid and the cells are grown in the presence of the appropriate amino acid on autoinduction agar, sfGFP is expressed and the colonies are visibly green.

One microliter of each library resulting from the second positive and the second negative rounds of selection was transformed with 60 μ L of pALS-containing DH10B cells. The cells were rescued for 1 hr in 1 mL of SOC (37 °C, 250 rpm). A 250 μ L and 50 μ L volume of cells from each library were plated on autoinducing minimal media with 25 μ g/mL Kn, 50 μ g/mL Tet, and 1 mM **Tet-v2.0**. Plates were grown at 37 °C for 24 hours and then grown on the bench top, at room temperature, for an additional 24 hours. Autoinducing agar plates were prepared by combining the reagents in Table 2.1 with an autoclaved solution of 40 g of agarose in 400 mL water. Sterile water was added to a final volume of 500 mL. Antibiotics were added to a final concentration of 25 μ g/mL Tet and 50 μ g/mL Kan.

	A) Autoinduction medium	B) Non-inducing medium	C) Autoinducing plates
5% aspartate, pH 7.5	25 mL	25 mL	25 mL
10% glycerol	25 mL	-	25 mL
25× 18 amino acid mix	20 mL	20 mL	20 mL
50× M	10 mL	10 mL	10 mL
leucine (4 mg/mL), pH 7.5	5 mL	5 mL	5 mL
20% arabinose	1.25 mL	-	1.25 mL
1 M MgSO ₄	1 mL	1 mL	1 mL
40% glucose	625 µL	6.25 mL	125 µL
Trace metals	100 µL	100 µL	100 µL

Table 2.1. Components for autoinducing and non-inducing mediums.

For final volume of 500 mL.

A total of 96 visually green colonies were selected from the two 1 mM **Tet-v2.0** plates and used to inoculate a 96-well plate containing 0.5 mL per well non-inducing minimal media (Sup. Table 1B, with sterile water added to a final volume of 500 mL) with 25 µg/mL Kn, 25 µg/mL Tet. After 24 hours of growth (37 °C, 250 rpm), 5 µL of these non-inducing samples were used to inoculate 96-well plates with 0.5 mL autoinduction media (Table 2.1, with sterile water added to a final volume of 500 mL) containing 25 µg/mL Kn, 25 µg/mL Tet with and without 1 mM **Tet-v2.0**.

Fluorescence measurements of the cultures were collected 36 hours after inoculation using a BIOTEK® Synergy 2 Microplate Reader. The emission from 528 nm (20 nm bandwidth) was summed with excitation at 485 nm (20 nm bandwidth). Samples were prepared by diluting suspended cells directly from culture 2-fold with phosphate buffer saline (PBS). Seven expressions showed high efficiency with **Tet-v2.0** and good fidelity without ncAA.

Fluorescence analysis of highest-fluorescing clones

Non-inducing cultures (3 mL) with 25 µg/mL Kn and 25 µg/mL Tet were grown to saturation (37 °C with shaking at 250 rpm) from the top seven expressions in the 96 well plate analysis. Autoinduction cultures (5 mL) with 25 µg/mL Kn and 25 µg/mL Tet were inoculated with 30 µL of non-inducing cultures and grown with and without 1 mM **Tet-v2.0** at 37 °C with shaking at 250 rpm. After approximately 40 hours,

fluorescence was assessed as described above (Fig 2.5). The top seven performing clones were sequenced revealing seven unique clones (Table 2.2).

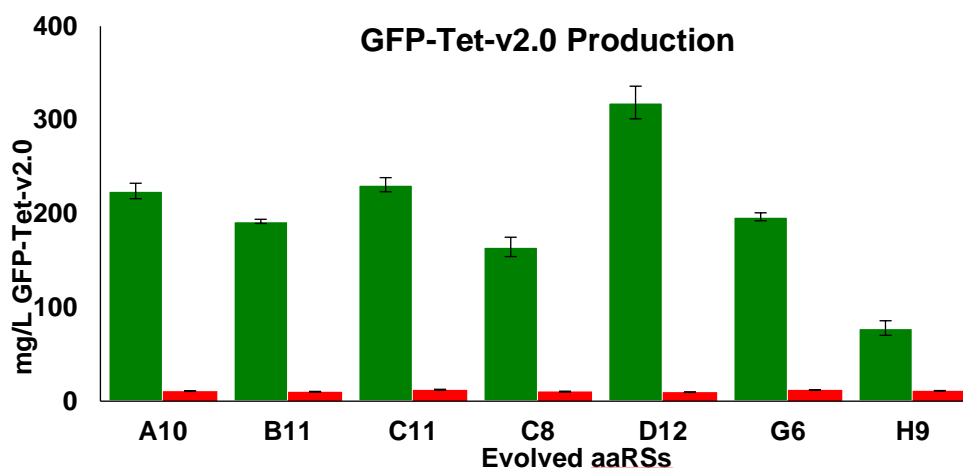


Fig 2.5. Fluorescence Incorporation of Tet-v2.0 Synthetases.

Fluorescence measurements of seven synthetases with GFP ncAA-reporter. Green represents colonies induced in media containing 1 mM **Tet-v2.0**, while red represents colonies induced in the absence of UAA. Expressions of 500 μ L were grown for 36 hours before dilution of suspended cells directly from culture 2-fold with phosphate buffer saline (PBS). Fluorescence measurements were collected using a BIOTEK® Synergy 2 Microplate Reader.

Mj parent	Tyr	Tyr32	Leu65	Phe108	Gln109	Asp158	Leu162
A10	Ala	Ala	Ala	Glu	His	Gly	Ser
B11	Ala	Ala	Ser	Gln	Glu	Gly	Ala
C8	Ala	Ala	Ala	His	Ser	Ser	Gly
C11	Ala	Ala	Val	Asp	His	Gly	Ser
D12	Gly	Gln	Ser	Asp	Ser	Ser	Asn
E12	Ala	Ala	Ala	Leu	Pro	Gly	Gly
G6	Ala	Ala	Ser	Ala	Glu	Asn	Ala
H9	Ala	Ala	Ser	Gln	Asp	Ser	Ala

Table 2.2. Sequence of top performing Tet-RSs.

The D12 synthetase (bold) was moved into the pDule plasmid for protein expression.

Generation of pDule-tet-v2.0

The top performing Tet-RS was moved from the *pBK-D12* plasmid to the pDule plasmid (*pDule-tet2.0*). The *pDule* plasmid was generated by amplifying the *MjYRS* gene from the *pBK* plasmid isolated from the library using primers RSmovf (5'-CGCGCGCCATGGACGAATTTGAAATG-3') and RSmover (5'-GACTCAGTCTAGGTACCCGTTTGAAACTGCAGTTATA-3'). The amplified

DNA fragments were cloned into the respective sites on the *pDule* plasmids using the incorporated NcoI and KpnI sites.

Expression and Purification of GFP-Tet-v2.0

DH10B *E. coli* cells co-transformed with *pDule-tet2.0* and *pBad GFP-TAG150* vector were used to inoculate 5 mL of non-inducing medium containing 100 µg/mL Amp and 25 µg/mL Tet. The non-inducing medium culture was grown to saturation with shaking at 37 °C, and 1 mL was used to inoculate 0.1 L autoinduction medium with 100 µg/mL Amp and 25 µg/mL Tet, and 1 mM **Tet-v2.0** (0.1 L of media grown in 500 mL plastic baffled flasks). After 40 hours of shaking at 37 °C, cells were collected by centrifugation. The protein was purified using BD-TALON cobalt ion-exchange chromatography. The cell pellet was resuspended in wash buffer (50 mM sodium phosphate, 300 mM sodium chloride, pH 7) and lysed using a microfluidizer (final volume 35 mL). The lysate was clarified by centrifugation, applied to 0.5 mL bed-volume resin, and bound for 20 min. Bound resin was washed with >50 volumes wash buffer. Protein was eluted from the bound resin with 2.0 mL of elution buffer (50 mM sodium phosphate, 300 mM sodium chloride, 150 mM imidazole pH 7) until the resin turned pink and the color of the eluent the column was no longer green. The elution concentrations were checked with a Bradford protein assay. The protein was desalted into PBS pH 7.4 using PD10 columns.

Rapid in vitro labeling of Tet-v2.0 containing protein with dTCO

Pure GFP-Tet-v2.0 at 3 µM in 20 mM ammonium acetate pH 7 was incubated with dioxolane fused trans cyclooctene (dTCO) at a final concentration of 39 µM (Fig 2.6). The reaction was run for 5 min at room temperature before desalting with PD10 columns into 20 mM ammonium acetate pH 7. These samples were frozen and vacuum dried. WT GFP and GFP-Tet-v2.0 with no addition of **dTCO** were run in parallel as controls.

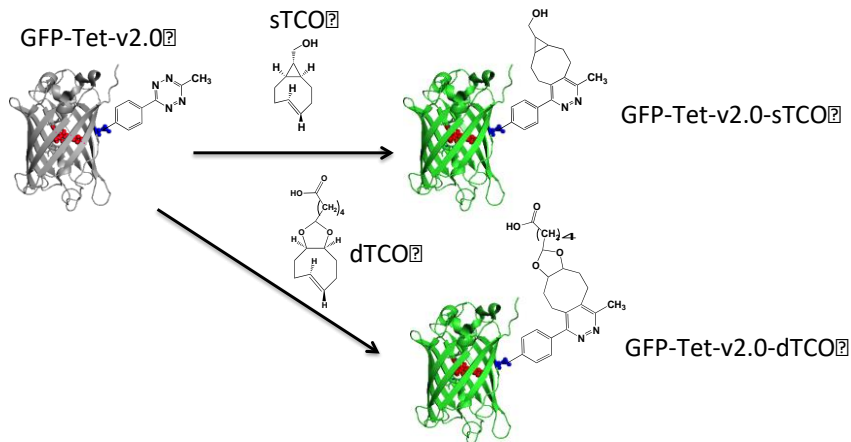


Fig 2.6. Reaction of GFP-Tet-v2.0 with sTCO and dTCO.

Rapid *in vivo* labeling of Tet-v2.0 containing protein with sTCO.

DH10B cells in 50 mL autoinduction media expressing cytosolic GFP-Tet-v2.0 were pelleted at 2000 rcf for 5 min. The cells were washed 3 times with 5 mL PBS. Aliquots of 1 mL cells were centrifuged (2000 rcf, 5 min) and stored at -80 °C. Cells were thawed on ice and resuspended in 5 mL PBS. The cells were incubated at 37 °C for 3 hours with heavy aeration. A stock solution of sTCO was prepared in MeOH with a concentration of 1.0 mM. Cells were diluted in a cuvette (100 μ L cells to 2.9 mL PBS). The fluorescence of the reaction was monitored (488 nm excitation, 509 nm emission, 5 mm slit width). Excess addition of sTCO was performed as a positive control (50 μ L 1.0 mM). The first three sub-stoichiometric additions (3 μ L 1.0 mM) were allowed 3 min to equilibrate before 250 μ L of buffer was removed to assess the sTCO concentration. After 5 min the subsequent addition was performed. The fourth addition of sTCO consisted of an excess of sTCO (41 μ L 1.0 mM). Control additions of sTCO and subsequent removal were performed for GFP TAG150 cells grown in the absence of Tet-v2.0 and for PBS buffer alone.

MS Analysis of WT GFP, GFP-Tet-v2.0, and GFP-Tet-v2.0-sTCO.

Purified WT GFP and GFP-Tet-v2.0 were diluted to a concentration of 1 mg/mL and reacted with ~5 equivalents of sTCO overnight. Protein was desalted on Millipore C₄ zip tips and analyzed using an FT LTQ mass spectrometer at the Oregon State University mass spectrometry facility (Fig 2.7).

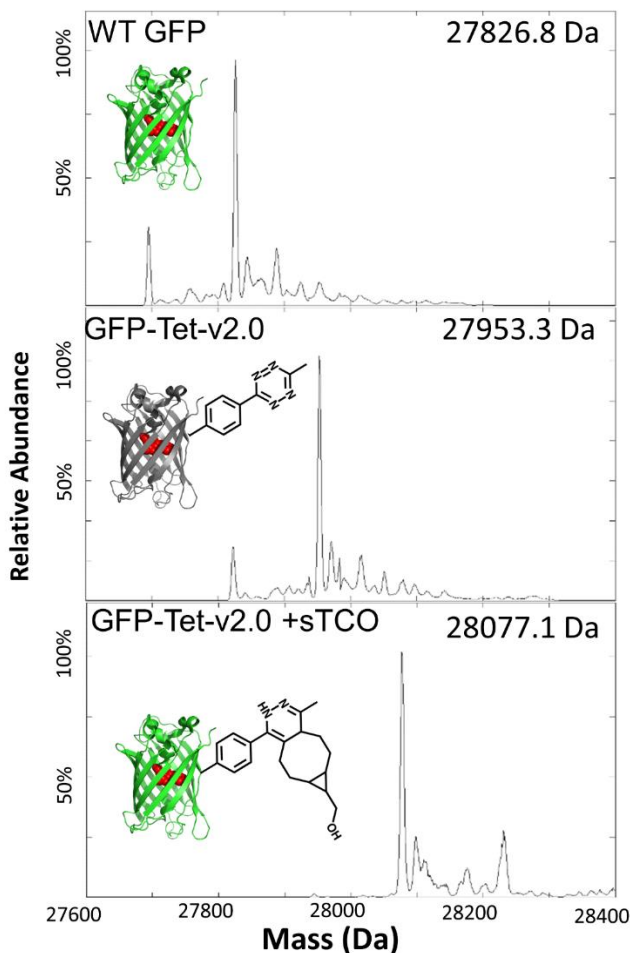


Fig 2.7. *In vitro* incubation of GFP-Tet-v2.0 with sTCO.

ESI-MS of proteins; GFP-Tet-v2.0, GFP-Tet-v2.0-sTCO, and GFP-Tet-v2.0-dTCO, demonstrates specific and quantitative labeling of GFP-Tet-v2.0 and no background labeling of WT GFP. A: ESI-MS TOF analysis of WT GFP shows a single major peak at 27826.8 Da \pm 1 Da (expected 27827.3) B: ESI-MS-TOF analysis of GFP-Tet-v2.0 shows a single major peak at 27953.3 Da \pm 1 Da which is in agreement with the expected mass (27954.5 Da). C: ESI-MS-TOF analysis of GFP-Tet-v2.0-sTCO shows a single major peak at 28077.1 Da \pm 1 Da (expected 28078.7 Da). This shows the expected molecular weight difference of 124.2 Da from GFP-Tet-v2.0 demonstrating specific and quantitative conversion to GFP-Tet-v2.0-sTCO. C: ESI-MS-TOF analysis of GFP-Tet-v2.0 incubated with dTCO shows a single major peak at 28181.9 Da \pm 1 Da (expected 28180.7 Da). This shows the expected molecular weight difference of 226.2 Da from GFP-Tet-v2.0 demonstrating specific and quantitative conversion to GFP-Tet-v2.0-dTCO. Each sample did show a small peak at -131 \pm 1 Da indicating minor amounts of peptidase-based removal of N-terminal methionines and +22 sodium adducts.

***In vitro* kinetic analysis of GFP-Tet-v2.0 with sTCO**

Seven stock solutions of sTCO in methanol were prepared (0.585 mM, 0.293 mM, 0.146 mM, 0.0731 mM, 0.0366 mM, 0.0183 mM, and 0.00914 mM). Solid dTCO was dissolved in methanol to generate seven stock dilutions (1.73 mM, 0.864 mM, 0.432 mM, 0.216 mM, 0.108 mM, 0.054 mM, 0.027 mM). Kinetic trials containing 30 μ L of GFP-Tet-v2.0 (1 μ M in PBS) in 2.5 mL of 1x PBS buffer at 21°C were initiated by adding 50 μ L of an sTCO or dTCO stock solution. Reactions were monitored by observing the fluorescence increase from product formation (Excitation 488 nm with a 1.25 nm slit, Emission 510 nm with a 4 nm slit). Fluorescence measurements for each trial were run until a constant emission intensity was reached indicating a completed reaction. A unimolecular rate constant was obtained for each concentration and all seven unimolecular rate constants were used to obtain a bimolecular rate constant (Fig 2.8).

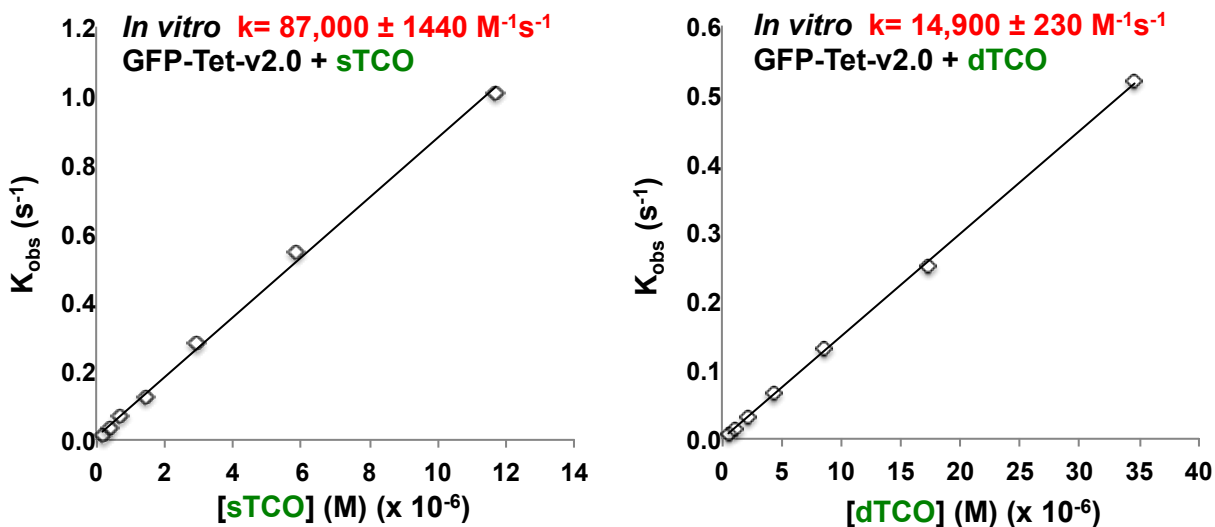


Fig 2.8. *In vitro* kinetic analysis of GFP-Tet-v2.0 with sTCO and dTCO.

Determined first order rate constants were plotted against sTCO or dTCO concentrations. Standard curves were fit to the data and second order rate constants were calculated.

***In vivo* kinetic analysis of GFP-Tet-v2.0 with sTCO.**

A 50 mL GFP-150-Tet-v2.0 cell pellet was resuspended and washed in 50 mL PBS buffer three times. Six kinetic trials containing 100 μ L of the cell solution added to 2.85 mL of 1x PBS buffer were initiated by adding 50 μ L of one of six stock solutions

of sTCO (613, 306, 153, 76.6, 38.3, 19.1 μM). Kinetics trials were stirred continuously and monitored by observing the fluorescence increase from product formation (Excitation 488 nm with 2 nm slit, Emission 506 nm with 5 nm slit, 0.1 sec integration time, and 2 sec increments). Fluorescence measurements for each trial were run until a constant emission intensity indicative of a complete reaction was obtained. A unimolecular rate constant was obtained for each concentration and all six unimolecular rate constants were used to obtain a bimolecular rate constant.

Assay of sTCO concentration in supernatant of sub-stoichiometric additions

The buffer samples from the sub-stoichiometric reactions described above were centrifuged (2000 rcf, 5 min) and 200 μL of supernatant was removed for analysis. To assay the concentration the supernatant was reacted with excess purified GFP-Tet-v2.0 and the fluorescence increase of the GFP-Tet-v2.0 was compared to a standard curve. To generate the standard curve 10 μL GFP-Tet-v2.0 was added to 2990 μL PBS. Known amounts of pure sTCO (10-1000 pmol) were added in volumes of 10-50 μL . The increase in fluorescence of the GFP-Tet-v2.0 was measured and the dilution factor was accounted for to result in an adjusted fluorescence increase. The quantity of sTCO was plotted against the adjusted increase and the data were fit with a least squares regression line ($R^2 = 0.9969$) (Fig 2.9). The adjusted fluorescence increases were measured for 20-50 μL additions of supernatant from the PBS, GFP-Tet-v2.0 cells, and the minus amino acid cells. GFP-Tet-v2.0 cells had significant background due to trace amounts of GFP in the supernatant. The background was accounted for by subtracting the adjusted fluorescence increase from addition of the supernatant to 3000 μL PBS buffer alone. The standard curve was used to determine quantities and concentrations of sTCO present in the supernatants and error was determined as the error in the standard curve.

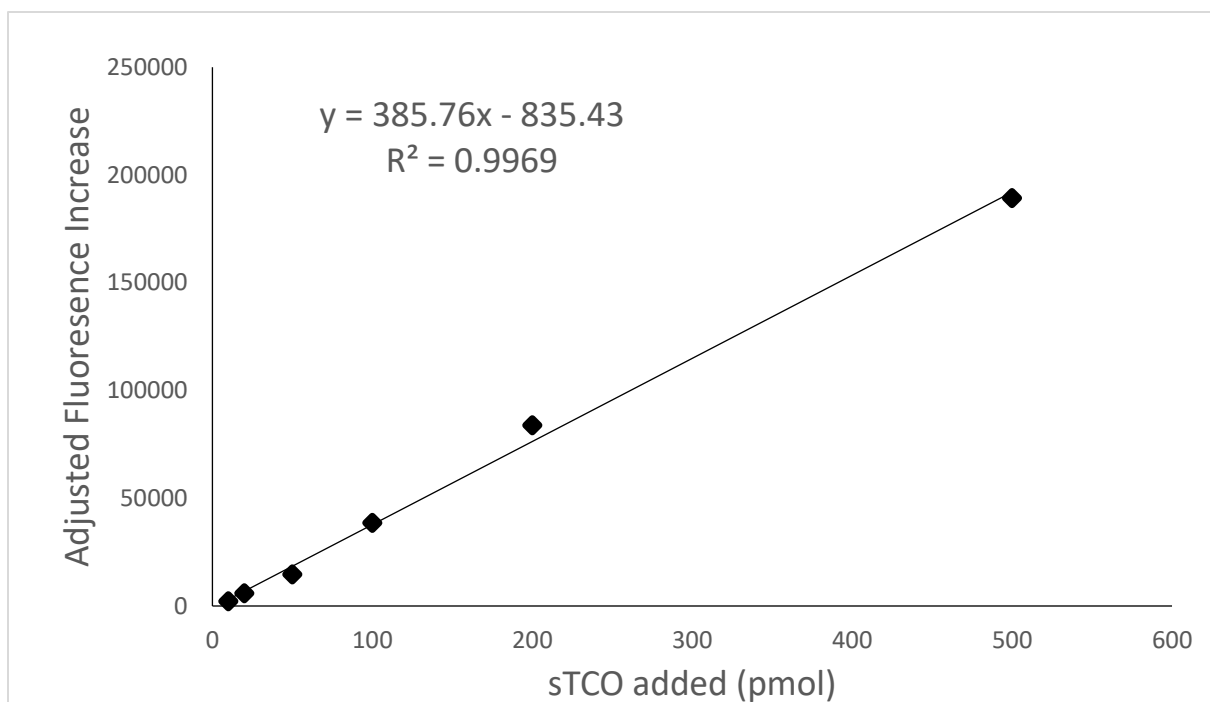


Fig 2.9. Calibration Curve for Determining sTCO Concentration.

The adjusted fluorescence increase of additions of sTCO to GFP-Tet-v2.0 were plotted against the amount of sTCO. Unknown concentrations of sTCO were then added and the resulting adjusted fluorescence increase was used to determine the concentration of the unknown values.

Synthesis of TAMRA linked sTCO

(Fig 2.10) A dry 25 mL pear shaped flask was charged with (19.7 mg, 31.3 μmol) tetramethylrhodamine 5-(and 6)-carboxamide cadaverine TFA salt that was partially dissolved in anhydrous dichloromethane (1 mL, 15 mmol). Diisopropylethylamine (20.6 μL , 0.118 mmol) was added and resulted in the complete dissolution of tetramethylrhodamine 5-(and 6)-caboxamide cadaverine. A solution was prepared of (activated sTCO) (18.8 mg, 59.2 μmol) in dichloromethane (1 mL, 15 mmol) and was added to the mixture. The reaction was allowed to proceed at room temperature under argon for 16 hours. The mixture was dried under reduced pressure and purified using silica gel flash column chromatography (9:1 dichloromethane:methanol) to yield TAMRA-sTCO (13.2 mg, 61%), a red solid. ^1H NMR (400MHz, $\text{C}_2\text{D}_6\text{OS}$) ^1H NMR(400MHz, DMSO D_6) δ - 8.82 (t, 1H), 8.69 (t, 1H), 8.44 (s, 1H), 8.22 (d, 1H), 8.17 (d, 1H), 8.06 (d, 1H), 7.63 (s, 1H), 7.32 (d, 1H), 7.08 (m, 2H), 5.776 (m, 2H), 5.067

(m, 2H). MS (ESI) $[M+H]^+$ calcd. for $C_{41}H_{49}N_4O_6^+$, $[M+H]^+$: 693.36 ; found: 693.40 (Fig 2.11).

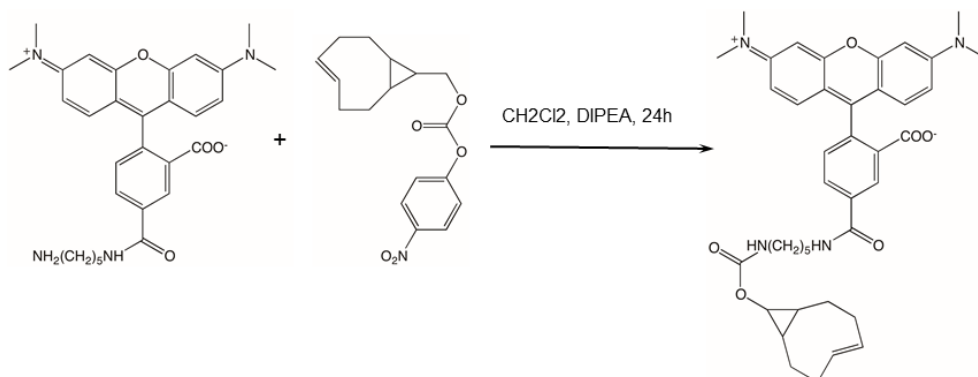


Fig 2.10. Synthesis of TAMRA linked sTCO.

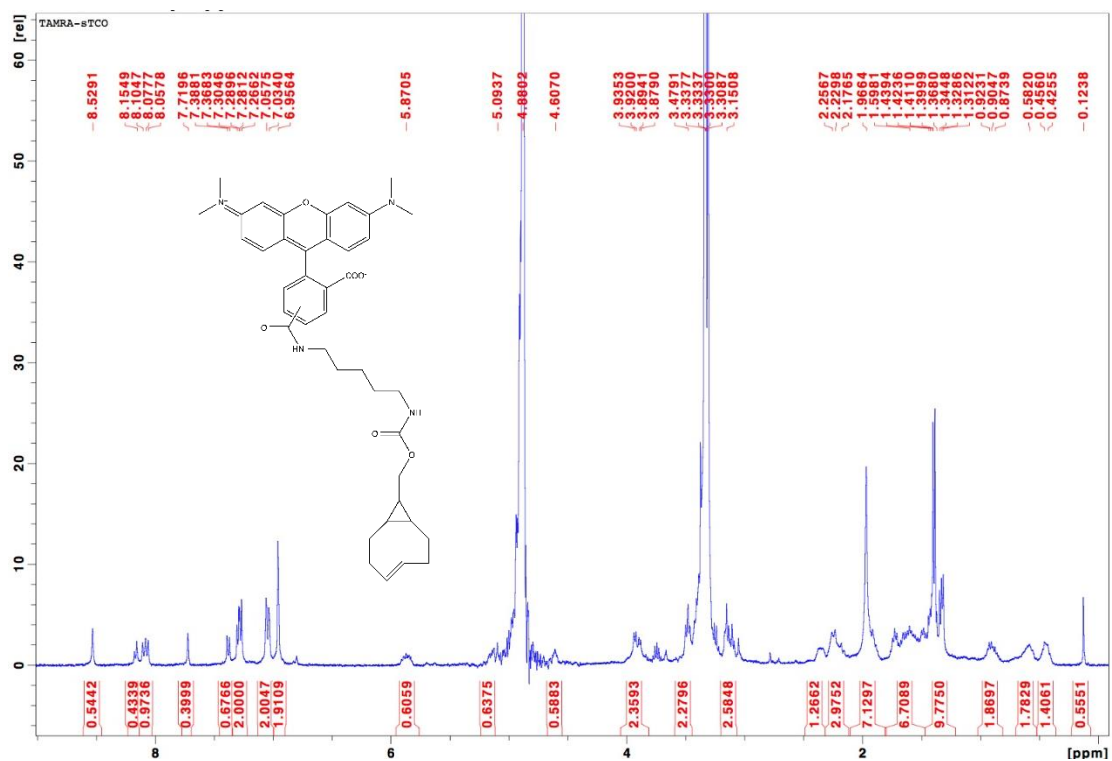


Fig 2.11. 1H NMR of TAMRA-sTCO.

SDS-PAGE analysis of GFP-Tet-v2.0 incorporation and reactivity

DH10B cells expressing GFP-Tet-v2.0 from 50 mL autoinducing media with 1 mM Tet-v2.0 were washed 3 times with 5 mL PBS and resuspended in 5 mL PBS. The cells were incubated at 37°C for 3 hours. Cells were aliquotted into 1 mL samples and centrifuged (2000 rcf, 5 min) and stored at -80°C. Similarly, WT-GFP expressed cells

and cells expressed in the absence of Tet-v2.0 were also aliquotted and centrifuged. TAMRA-sTCO (173 μ L, 1 mM) was added to a WT-GFP and GFP-Tet-v2.0 aliquot and allowed to react overnight. Additionally, sTCO was added to an aliquot of GFP-Tet-v2.0 and allowed to react overnight. The aliquots were subsequently lysed and purified using BD-TALON cobalt ion exchange chromatography eluting into 0.5 mL. Equal volumes of purified protein were mixed with 2x Laemmli Buffer and heated at 95°C for 20 min. Samples were analyzed using SDS-PAGE (7 μ L sample, 15% Acrylamide gel, 180 V, 70 min). The gel was fluorescently imaged prior to staining with Coomassie. After staining and destaining, the gel was imaged again and the fluorescent image was aligned with the Coomassie-stained image using the ladders visible in the unstained image. Fluorescent signal from the TAMRA label was only present when the dye was reacted with GFP-Tet-v2.0 (Fig 2.12).

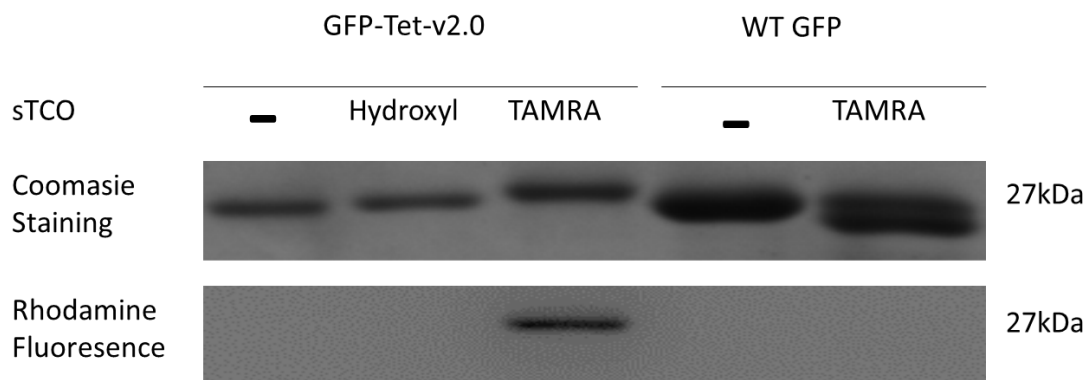


Fig 2.12. SDS-PAGE analysis of GFP-Tet-v2.0 reactivity.

Reaction of TAMRA-sTCO with GFP-Tet-v2.0 resulted in a slight upward gel shift in comparison to the unreacted and sTCO reacted forms. In addition a fluorescent band was visible for the TAMRA-sTCO reacted GFP-Tet-v2.0. No shift was present or fluorescence was observed for WT GFP incubated with TAMRA-sTCO.

TAMRA labeling of GFP-Tet-v2.0 in cell lysate

GFP-Tet-v2.0 expressing *E. coli* cells were resuspended in PBS and lysed. The lysate was clarified via centrifugation (21036 rcf, 45 min) and the supernatant was decanted and stored at 4°C. A mixture of protease inhibitors (400 μ L) were added to 20 mL of supernatant. The clarified lysate was divided into 8 x 1 mL aliquots. TAMRA-sTCO (1 mg/mL in methanol, 0-100 μ L) was added to each aliquot. Each sample was allowed to react for 1 hr at room temperature. The samples were mixed with 2x Laemmli Buffer

and heated at 95°C for 20 min. The samples were analyzed via SDS-PAGE (10 μ L sample, 15% Acrylamide gel, 200V, 50 min). The dye front was not allowed to run off the gel. Samples were fluorescently imaged prior to staining with Coomassie (Fig 2.13).

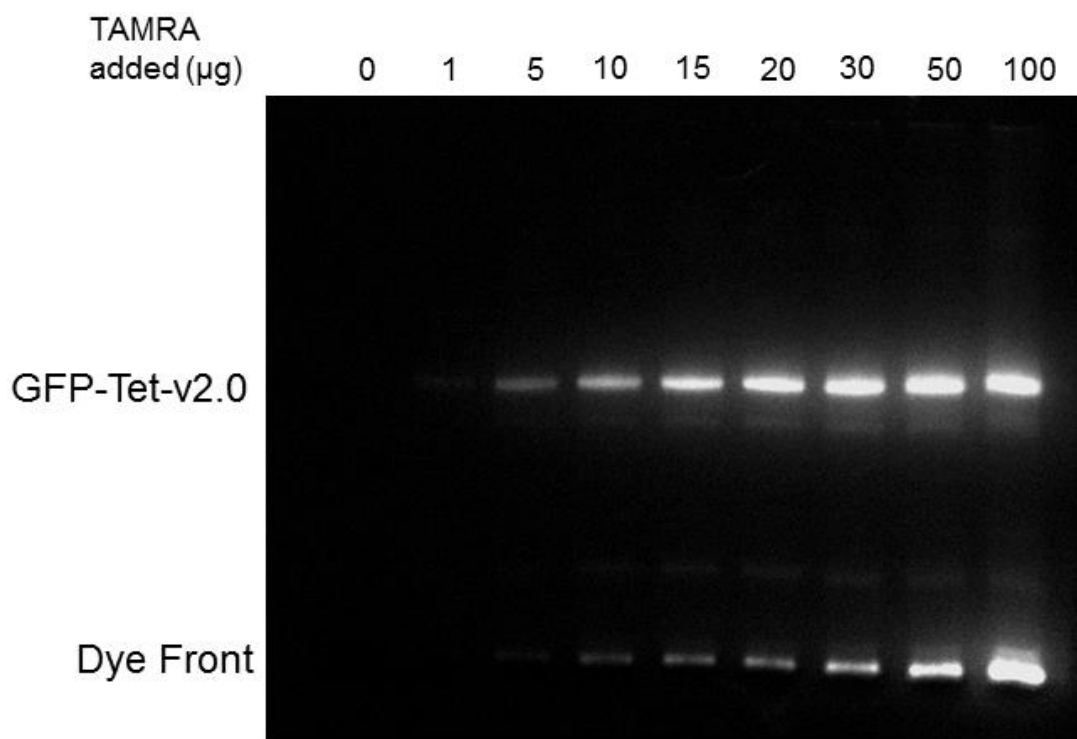


Fig 2.13. Fluorescent SDS-PAGE analysis of GFP-Tet-v2.0 reaction.

Reaction of TAMRA-sTCO with GFP-Tet-v2.0 resulted in two prominent fluorescent bands. The bands correspond to migration of TAMRA-sTCO reacted with GFP-Tet2.0 and unreacted TAMRA-sTCO.

We compared the ability of the reaction between sTCO and tetrazines to label GFP-Tet-v2.0 and GFP-Tet-v1.0 with a fluorescent dye in cell lysate. In order to analyze the reactions quantitatively, we evaluated two different methods of quenching the reaction. The first method involved addition of excess 3,6-Di-2-pyrimidal-1,2,4,5-tetrazine to react with and remove unreacted sTCO from solution. The second method of quenching involves the addition of an excess amount of sTCO to remove unreacted tetrazine functional groups from solution. The effectiveness of these quenching methods was tested by reacting GFP-Tet-v1.0 and GFP-Tet-v2.0 in cell lysate with TAMRA-sTCO and quantifying the relative fluorescent intensities of their corresponding bands using SDS-PAGE. Cell lysate containing GFP-Tet-v1.0 and GFP-Tet-v2.0 (20 μ L, \sim 1.25 μ M GFP), as well as WT GFP (20 μ L, \sim 1.25 μ M GFP) was

quenched via the addition of either sTCO (1 mM in 5 μ L methanol) or 3,6-Di-2-pyrimidal-1,2,4,5-tetrazine (1 mM in 5 μ L dimethylformamide) and after 1 min reacted with \sim .33 equivalents of TAMRA-sTCO in 5 μ L mehtanol. As a negative control, samples were run without the addition of TAMRA-sTCO, and as a positive control, samples were reacted with TAMRA-sTCO 1 min prior to addition of the 3,6-Di-2-pyrimidal-1,2,4,5-tetrazine quenching agent (1mM in 5 μ L dimethylformamide). Samples were allowed to react 1 min after the addition of the last reagent before mixing with 30 μ L 2x Laemmli Buffer and heating for 10 min at 99.5 $^{\circ}$ C. The samples were analyzed via SDS-PAGE (10 μ L sample, 15% Acrylamide gel, 200V, 40 min). Samples were fluorescently imaged prior to staining with Coomassie (Fig 2.14). Both the absence of fluorescent bands at the expected molecular weight outside of the positive control indicates that the addition of excess sTCO and excess tetrazine are sufficient to quench this reaction. In addition, the absence of fluorescently labeled bands outside of GFP and the dye front indicate that this reaction is selective towards Tet-v2.0. To further demonstrate this we have increased the brightness on the fluorescently imaged gel to the point where there is a clear halo effect around the two observed bands and from the dye front at the bottom of the image. Even with this bright image, no bands corresponding to side reactions other cellular components are observed.

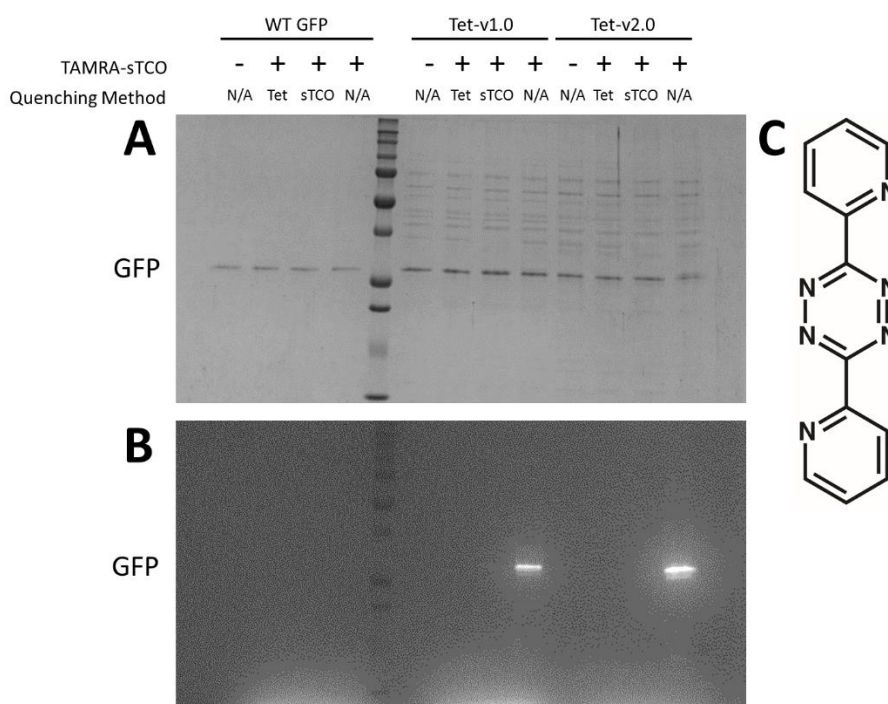


Fig 2.14. Quenching of the reaction between tetrazine and sTCO.

(A) SDS-PAGE of WT GFP, GFP-Tet-v1.0, and GFP-Tet-v2.0 reacted with TAMRA-sTCO in cell lysate. (B) Rhodamine fluorescence of the above gel. Bands are present at the expected molecular weight of GFP only when TAMRA-sTCO is added prior to the addition of the quenching agent. (C) Structure of 3,6-Di-2-pyrimidal-1,2,4,5-tetrazine.

GFP-Tet-v1.0 and GFP-Tet-v2.0 (20 μ L, \sim 1.25 μ M GFP in cell lysate) were then reacted with various concentrations of TAMRA-sTCO spanning from 0.07 to 1.33 equivalents. The reactions were quenched through the addition of 3,6-Di-2-pyrimidal-1,2,4,5-tetrazine (1 mM, 5 μ L in dimethylformamide) after 15 sec and 15 min, and analyzed using SDS-PAGE. The fluorescent intensity of bands corresponding to reaction of tetrazine-GFP and sTCO-TAMRA was quantitated. After 15 min of reaction, GFP-Tet-v1.0 and GFP-Tet-v2.0 had similar intensities when labeled with the same amount of TAMRA-sTCO (Fig 2.15). Both Tet-v1.0 and Tet-v2.0 underwent sub-stoichiometric labeling. However, after shorter 15 s reactions, Tet-v2.0 reacted to completion at sub-stoichiometric quantities of label whereas Tet-v1.0 did not completely react (Fig 2.16).

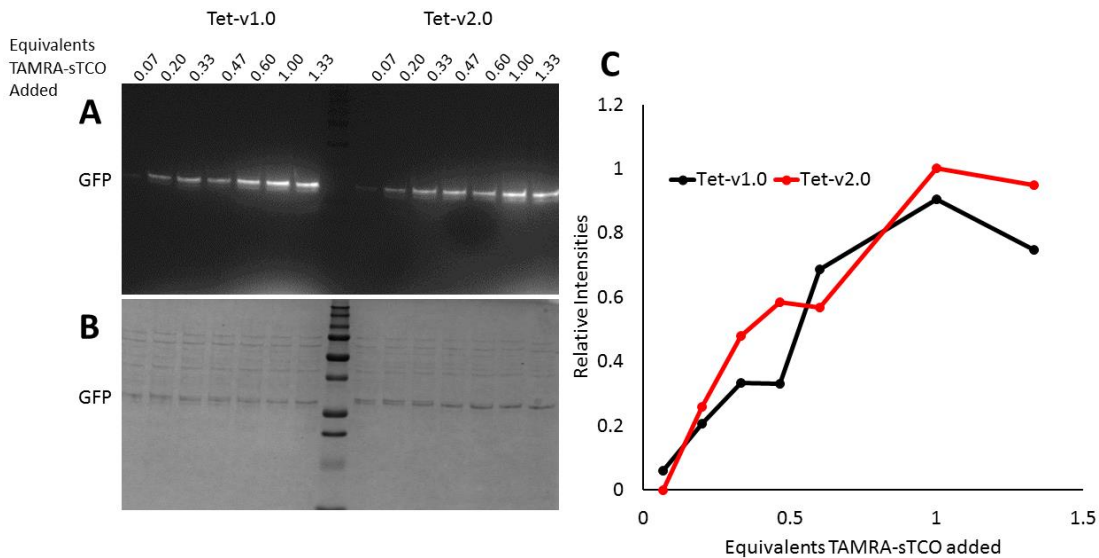


Fig 2.15. 15 min reaction between sTCO-TAMRA and GFP-tetrazine.

(A) Fluorescent SDS-PAGE analysis of GFP-Tet-v1.0, and GFP-Tet-v2.0 reacted with TAMRA-sTCO in cell lysate. Bands are only observed at the expected molecular weight of GFP and at the dye front. (B) Coomassie staining of the above gel. (C) Quantitation of the fluorescent intensity of the gel in A at the position of GFP. As the concentration of TAMRA-sTCO increases, both Tet-v1.0 and Tet-v2.0 react completely. Fluorescence intensity is normalized to the highest observed intensity.

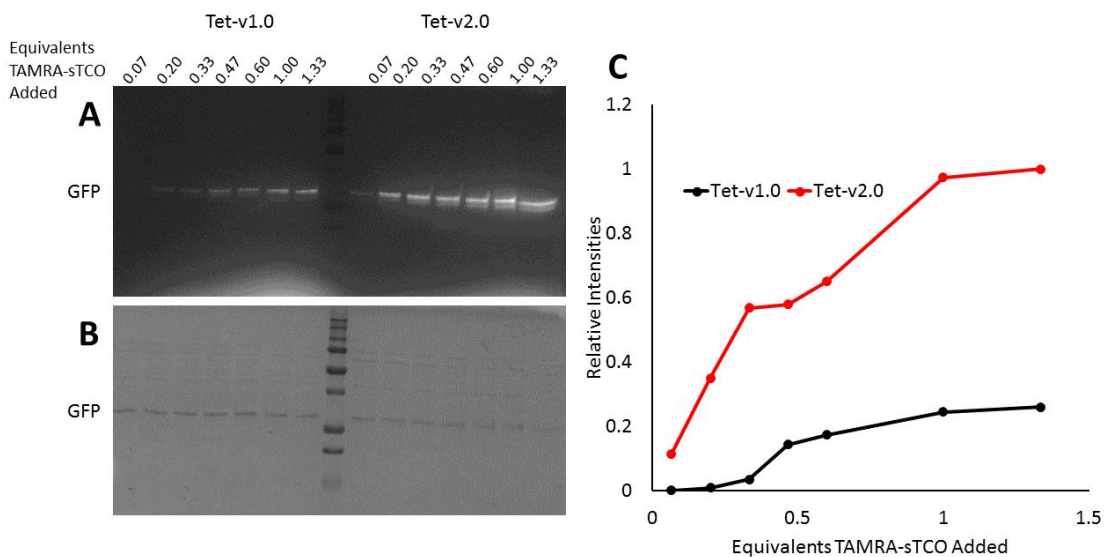


Fig 2.16. 5 min reaction between sTCO-TAMRA and GFP-tetrazine.

(A) Fluorescent SDS-PAGE analysis of GFP-Tet-v1.0, and GFP-Tet-v2.0 reacted with TAMRA-sTCO in cell lysate. Bands are only observed at the expected molecular weight of GFP and at the dye front. (B) Coomassie staining of the above gel. (C) Quantitation of the fluorescent intensity of the gel in A at the position of GFP. As the concentration of TAMRA-sTCO increases, only Tet-v2.0 is observed to react completely. Partial reaction is seen for Tet-v1.0. Fluorescence intensity is normalized to the highest observed intensity.

Stability assessment of Tet-v2.0 in aqueous solutions.

A solution of Tet-v2.0 (5 mM) in deuterated PBS was created. ^1H NMRs were obtained after 0 and 10 days (Fig 2.17). The sample showed no significant degradation over time.

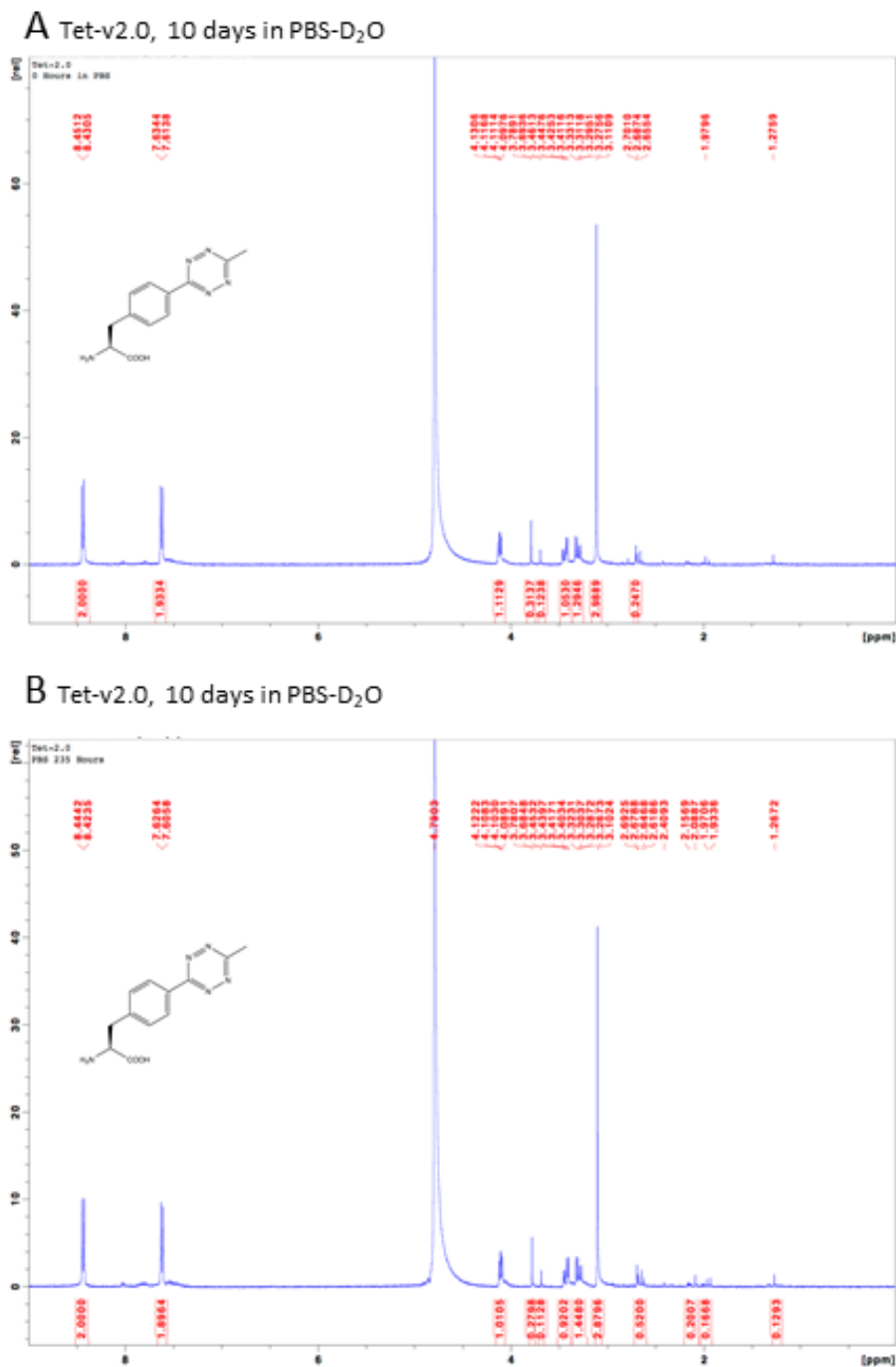


Fig 2.17. ^1H NMR of Tet-v2.0 in dPBS..

(A) ^1H NMR of Tet-v2.0 upon dissolution in deuterated PBS. (B) ^1H NMR of Tet-v2.0 in deuterated PBS after 10 days. Tet-v2.0 shows no significant degradation after 10 days.

Results

We site-specifically encoded the first tetrazine amino acid (Tet-v1.0) into proteins showing this functionality is compatible with genetic code expansion (Fig 2.18A).^{17,111} The *in cellulo* reaction rate of Tet-v1.0 with sTCO was faster than most bioorthogonal ligations at $880 \text{ M}^{-1}\text{s}^{-1}$, but was not fast enough to probe biological processes as an ideal bioorthogonal reaction. A maximum synthetic yield of 3% and low levels of hydrolysis at the amine linkage are additional weaknesses of Tet-v1.0 that ultimately limit its utility. To overcome these shortcomings and push the limits of *in vivo* bioorthogonal reaction rates, we generated a second tetrazine amino acid (Tet-v2.0) using a robust synthetic route. We genetically incorporated Tet-v2.0 into proteins and characterized the reactivity of Tet-v2.0-GFP *in cellulo* to show that it qualifies as an ideal bioorthogonal ligation.

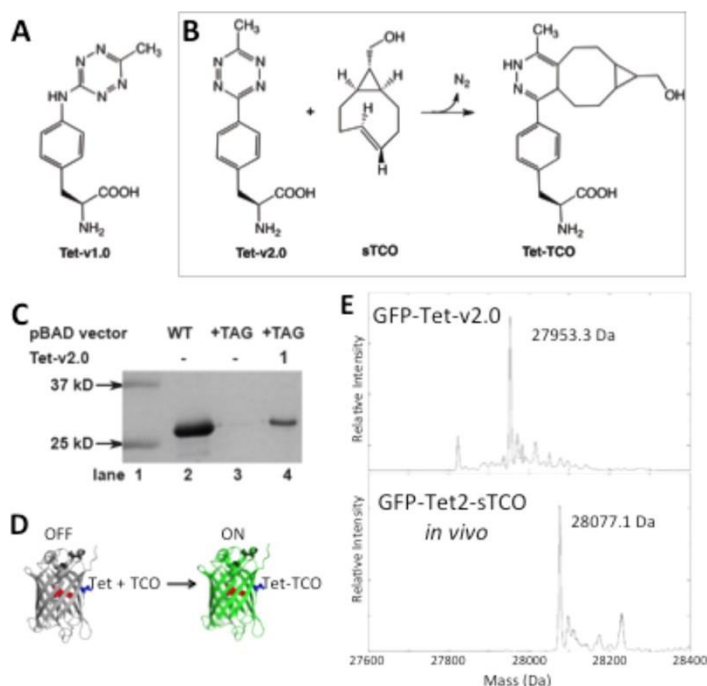


Fig 2.18. Genetic incorporation of Tet-v2.0, into proteins and labeling with sTCO.

(A) Structure of Tet-v1.0 (B) Reaction of Tet-v2.0 with sTCO to form the stable conjugate Tet-sTCO. (C) SDS-PAGE analysis of site-specific incorporation of Tet-v2.0 in response to the amber codon. Lane 2 shows expression levels of GFP-wt from pBad-GFP-His₆. Lanes 3 and 4 show the Tet-v2.0 dependent production of GFP-Tet-v2.0 (D) Excitation at 488 nm produces low fluorescence for GFP-Tet, while the reaction forming GFP-Tet-sTCO produces full fluorescence for GFP. (E) ESI-Q MS analysis of GFP-Tet-v2.0 shows a single major peak at 27953.3 ± 1 Da. *In cellulo* reaction of GFP-Tet-v2.0 with sTCO shows a single major peak at 28077.1 ± 1 Da consistent with the expected mass increase from specific and quantitative reaction with sTCO. Each sample did show $+22 \pm 1$ Da and -131 ± 1 Da peaks consistent with the mass of a sodium adduct and the removal of N-terminal methionine. No other peaks were observed that would correlate with background incorporation of natural amino acids

We predicted that removing the amine linkage of Tet-v1.0 would increase the tetrazine reaction rate and prevent hydrolysis at that junction. Replacing the strongly electron donating secondary amine linkage with the weakly donating phenyl substituent is expected to significantly accelerate the IED-DA reaction.⁶² Additional rate enhancement is achieved by removing the polar amine linkage because “enforced hydrophobic interactions” are improved.¹¹⁴ Using a nickel triflate catalyst for generating tetrazines from nitriles,¹¹⁵ we were able to produce 4-(6-methyl-*s*-tetrazin-3-yl)phenylalanine (Tet-v2.0) in two steps in a 57% yield from commercially available starting materials (Fig 2.1). Tet-v2.0 proved to be highly stable in PBS exhibiting no

degradation over 10 days in contrast to 3-phenyl-s-tetrazine and 3,6-(dipyridin-2-yl)-s-tetrazine which show 50% loss after 1 day (Fig 2.14).^{37,38}

In order to genetically incorporate Tet-v2.0 into protein and test its *in cellulo* activity with sTCO (Fig 2.18B), we evolved an orthogonal *Methanococcus jannaschii* (*Mj*) tyrosyl tRNA synthetase (RS)/tRNA_{CUA} pair capable of incorporating Tet-v2.0 in *E. coli* (see Supporting information for details).¹¹⁶ RS plasmids from surviving clones were transformed into cells with a plasmid containing a GFP gene interrupted with an amber codon.^{117,118} Ninety-six colonies assessed for Tet-v2.0-dependent expression of GFP, contained seven clones that had significant GFP-Tet-v2.0 expression in the presence of Tet-v2.0 and no detectable GFP fluorescence over background in the absence of Tet-v2.0 (Fig 2.5). Sequencing revealed that all seven RS sequences were unique (Table 2.2).

To facilitate robust expression of site-specifically encoded Tet-v2.0 containing proteins, the top performing Tet-RS was cloned into a *pDule* vector that contains one copy of *Mj* tRNA_{CUA} to create *pDule-Tet2.0*.¹¹⁷⁻¹¹⁹ Expression of a GFP gene interrupted by an amber codon at site 150 in the presence of *pDule-Tet2.0* was efficient and dependent on the presence of Tet-v2.0 (Fig 2.18C). Using 1 mM Tet-v2.0, 13.0 mg of GFP-Tet-v2.0 was purified per liter of medium, while GFP-wt yielded 161 mg/L under similar conditions (no GFP was produced in the absence of Tet-v2.0). To demonstrate that Tet-v2.0 can be stably incorporated into recombinant proteins using *pDule-Tet2.0*, we compared the masses of GFP-Tet-v2.0 to GFP-wt using ESI-Q mass analysis. The native GFP-wt has the expected mass of 27827 ±1 Da and GFP-Tet-v2.0 exhibits the expected mass increase to 27955 ±1 Da, verifying that Tet-v2.0 is incorporated at a single site (Fig 2.18E and Fig 2.7A). Overall, the results of protein expression, MS analysis and SDS PAGE demonstrate the cellular stability and efficient, high fidelity incorporation of Tet-v2.0 into proteins using a *pDule* system.

Previously, we showed that tetrazine amino acids quench GFP fluorescence when encoded close to its chromophore, and fluorescence returns when reacted with TCO-labels (Fig 2.18D). This increase in fluorescence exhibited by GFP-Tet-v2.0 upon reaction enables quantification of labeling reactions and reaction rates *in vitro* and *in cellulo*. Incubation of GFP-Tet-v2.0 (1.25 μM) with 13 μM sTCO in PBS buffer

showed a complete return of fluorescence in less than 10 seconds indicating that GFP-Tet-v2.0-sTCO was formed. ESI-Q of the desalted reaction mixture confirmed the quantitative conversion of GFP-Tet-v2.0 (expected 27954.5 Da; observed 27955.7±1 Da) to GFP-Tet-v2.0-sTCO (expected 28078.7 Da; observed 28078.3 ±1 Da) (Fig 2.7A). This demonstrates, the reaction between GFP-Tet-v2.0 and sTCO is quantitative *in vitro*.

To determine if this bioorthogonal ligation is also quantitative *in cellulo*, *E. coli* cells containing expressed GFP-Tet-v2.0 were incubated with 3.3 μM sTCO in PBS buffer at room temperature. Complete fluorescence returned in less than 10 seconds, indicating that GFP-Tet-v2.0-sTCO had been formed. After incubation at room temperature for 24 hours the cells were lysed, GFP-Tet-v2.0-sTCO-His₆ was affinity purified and analyzed by ESI-Q MS. The resulting molecular mass matched the expected molecular mass of GFP-Tet-sTCO (Fig 2.18E). This verifies that the *in cellulo* reaction is facile, quantitative, and produces a stable conjugated product.

An ideal bioorthogonal reaction requires an *in cellulo* rate of $>10^4 \text{ M}^{-1}\text{s}^{-1}$ to reach completion in seconds to minutes at biological concentrations (μM to nM) of both biomolecule and label. To determine if reactions of Tet-v2.0 on a protein are fast enough to meet these rates, the reaction of GFP-Tet-v2.0 with sTCO was measured. The kinetics of the reaction were performed under pseudo-first-order conditions as verified by a single exponential fit for return of product fluorescence. The *in vitro* second order rate constant for GFP-Tet-v2.0 with sTCO was calculated to be $87,000 \pm 1440 \text{ M}^{-1}\text{s}^{-1}$ (Fig 2.19A). Surprisingly the site-specific Tet-v2.0-protein reaction with sTCO is two orders of magnitude faster than Tet-v1.0.

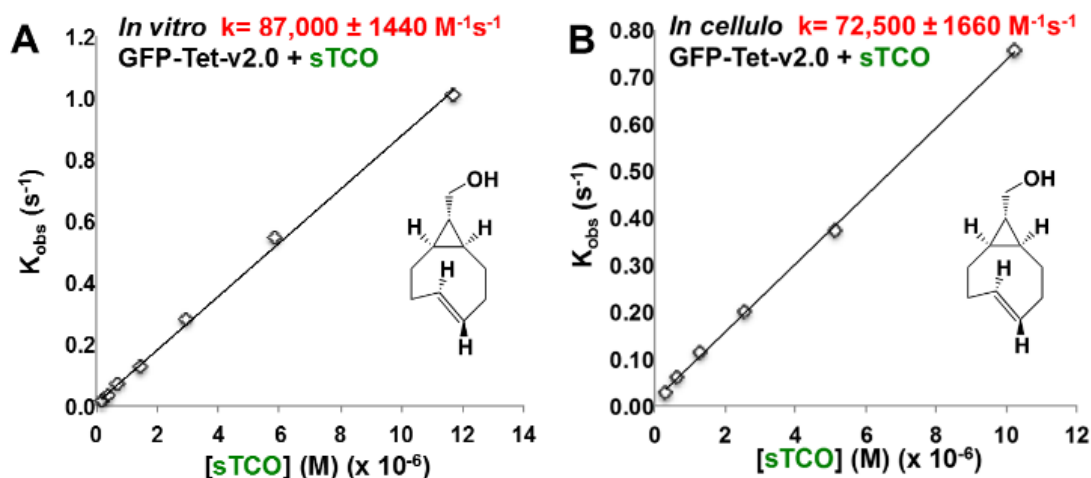


Fig 2.19. *In vitro* and *in cellulo* rate constant determination for reaction of GFP-Tet-v2.0 with sTCO.

(A) Kinetics of GFP-Tet-v2.0 with sTCO *in vitro* resulted in a rate constant of $k = 87,000 \pm 1440 \text{ M}^{-1}\text{s}^{-1}$ in a PBS buffer at pH 7 at 21°C. (B) Kinetics of GFP-Tet-v2.0 with sTCO *in cellulo* resulted in a rate constant of $k = 72,500 \pm 1660 \text{ M}^{-1}\text{s}^{-1}$. For both experiments, unimolecular rate constants were calculated by fitting the rate of product formation to a single exponential at different concentrations of sTCO, and the bimolecular rate constant was determined using the observed unimolecular rate constants ($k_{\text{obs}} = k[\text{TCO}]$).

To date, no bioorthogonal rate constants greater than $10^3 \text{ M}^{-1}\text{s}^{-1}$ have been measured *in cellulo*.^{41,55} To determine the rate constant for this reaction inside live cells, *E. coli* expressing GFP-Tet-v2.0 was washed, resuspended in PBS buffer, and reacted with sTCO. The *in cellulo* bimolecular rate constant for this reaction is $72,500 \pm 1660 \text{ M}^{-1}\text{s}^{-1}$ and is fast enough to meet the needs of the ideal bioorthogonal ligation (Fig 2.19B). This *in cellulo* reaction rate will allow 95% labeling in less than a minute at 1 μM Tet-v2.0-protein and sTCO label. The short reaction time is enabled by a $t_{1/2}$ of 12-14 seconds. Ideal bioorthogonal reaction rates eliminate the need for time consuming washing steps prior to cell analysis and allow for immediate monitoring of cellular events since the labeling reaction is rapidly completed at stoichiometric concentrations of label.

To verify that the Tet-v2.0-protein/sTCO reaction rate is sufficient to effectively use sub-stoichiometric concentrations of label in live cells, we reduced the amount of sTCO added to *E. coli* cells containing GFP-Tet-v2.0 (Fig 2.20A). For comparison, traditional labeling conditions using an excess of sTCO show complete labeling in ~ 1 minute (red trace). The green trace shows four additions of sTCO to cells containing GFP-Tet-v2.0.

The first three sTCO additions are $1/5^{\text{th}}$ the molar amount of GFP-Tet-v2.0 and the fourth addition is an excess of sTCO. The sub-stoichiometric labeling reproducibly showed complete labeling within 1 minute. When reacting sTCO with Tet-v2.0-protein sub-stoichiometrically *in cellulo*, all sTCO-label should bind to Tet-v2.0-protein *in cellulo* leaving none in extracellular solution. To verify that this was the case in our experiment, we assayed samples of the solution for sTCO after fluorescence plateau from each sTCO addition (points 1-4 Fig 2.20A). Following sub-stoichiometric additions of sTCO, (points 1-3) negligible concentrations of sTCO were detected in solution (Fig 2.20B). This contrasts with the stepwise increase in concentration of sTCO detected in solution when identical amounts of sTCO were added to PBS buffer in the absence of Tet-v2.0-protein. This feature of Tet-v2.0 thus eliminates the need for a wash out step when labeling protein *in vivo* if sTCO is conjugated to a fluorescent dye.

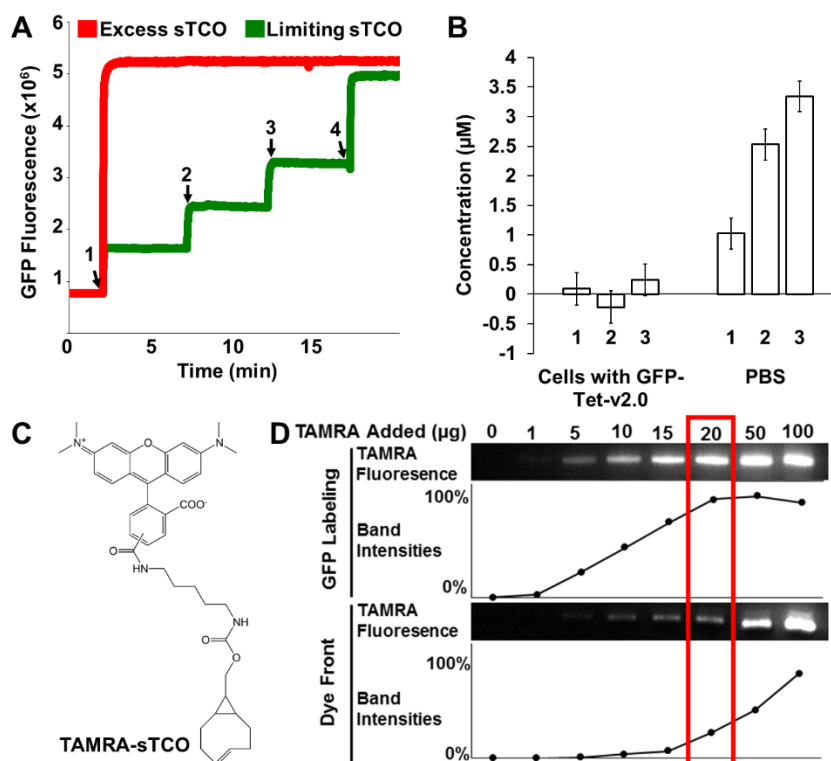


Fig 2.20. Sub-stoichiometric characterization of GFP-Tet-v2.0 reaction with sTCO.

(A) Red trace shows fluorescent change from sTCO added in excess. Green trace shows fluorescent change from the first three additions of 1/5 eq. of sTCO and the fourth addition of excess sTCO. (B) Concentrations of sTCO in medium were determined for samples removed after sTCO additions 1-3. Concentrations of sTCO were determined for identical additions of sTCO to buffer alone. (C) Structure of TAMRA-sTCO (D) Sub-stoichiometric labeling of *E. coli* lysate containing expressed GFP-Tet-v2.0 with TAMRA-sTCO. Lysate incubated with TAMRA-sTCO was separated on SDS-PAGE and imaged fluorometrically. Displayed regions correspond to GFP and dye front migration with their relative band intensities. The red box highlights the point of 100% protein labeling.

To demonstrate that the wash out step of a conjugated dye is nonessential when reaction rates of this magnitude are employed, a tetramethyl-rhodamine (TAMRA)-linked sTCO label was synthesized (Fig 2.20C). TAMRA-sTCO was incubated with purified GFP-Tet-v2.0 *in vitro* and analysis by SDS-PAGE demonstrated a reaction between GFP-Tet-v2.0 and TAMRA-sTCO (Fig 2.12). Fluorescence imaging of the gel showed a band present only when TAMRA-sTCO and GFP-Tet-v2.0 were present. Labeling of protein in living cells with low concentrations of dyes is often slow and incomplete because dye diffusion into cells at these concentrations and timescales is limiting.¹²⁰ Conjugated TAMRA dyes have previously been shown to enter mammalian

cells, but slower bioorthogonal reaction rates required higher concentrations of TAMRA-labels and longer reaction times.^{17,121} As suggested by others, improved fluorescent dyes are needed to overcome the rate limiting steps of cellular uptake with fast bioorthogonal ligations.¹²⁰ To circumvent this problem for this sub-stoichiometric demonstration, we reacted TAMRA-sTCO with *E. coli* lysate containing GFP-Tet-v2.0 at quantities of TAMRA-sTCO ranging from 5-500% of the total GFP-Tet-v2.0 concentration. The lysate was analyzed by SDS-PAGE and showed two rhodamine fluorescence bands; a ~27 kDa band corresponding to GFP-Tet-v2.0 conjugated to TAMRA-sTCO and a dye front migrating band corresponding to unreacted TAMRA-sTCO (Fig 2.20D). As expected, the fluorescent TAMRA-GFP band increased incrementally in intensity with additions of TAMRA-sTCO until the intensity plateaued at ~100% labeled GFP-Tet-v2.0 (20 μ g TAMRA-sTCO Fig 2.20D). While TAMRA-sTCO was added to the full lysate, only Tet-v2.0-GFP was labeled and TAMARA dye did not accumulate at the dye front of the gel until GFP-Tet-v2.0 was completely labeled (Fig 2.13). After this point, TAMRA fluorescence at the dye front increased rapidly with the amount of TAMRA-sTCO added as would be expected from a reaction with excess label. The low background signal detected at the dye front in the sub-stoichiometric reactions likely resulted from incomplete purification of TAMRA-sTCO from isomerized cyclopropane-fused cis-cyclooctene-linked-TAMRA and unreacted TAMRA starting materials. Together these data indicate that efficient sub-stoichiometric reactions of protein-Tet-v2.0 with TAMRA-sTCO are possible in the presence of cellular components.

Discussion

In summary, we have developed an *in cellulo* bioorthogonal reaction based on a genetically encodable tetrazine amino acid that meets the demands of an ideal bioorthogonal ligation. GFP-Tet-v2.0 does not cross-react with any cellular components or degrade in the cellular environment as demonstrated with mass spectrometry. Tet-v2.0 is small enough that it does not perturb the structure of GFP when positioned at site 150. The on-protein bimolecular rate constant of $87,000 \pm 1440 \text{ M}^{-1}\text{s}^{-1}$ gives this robust reaction the speed it needs to compete with cellular processes.

The same attributes that make this reaction ideal open the door to a variety of applications. The bimolecular rate constant is a significant improvement over previous *in vivo* bioorthogonal ligations. This speed affords complete labeling of Tet-v2.0-protein in minutes even with low concentration of the sTCO label or concentrations below that of the protein being labeled. A sub-stoichiometric *in vivo* bioorthogonal ligation has applications towards drug-antibody conjugates where it could minimize the clearance time of drugs or radioactive labels targeted to specific cells. Additionally, the high rate combined with *in cellulo* reactivity enable one to probe various pathways on a biologically relevant time scale. To our knowledge, this is the first demonstration of a bioorthogonal ligation with sufficient selectivity and a high enough reaction rate to sub-stoichiometrically label proteins in live cells, thereby eliminating the need to wash out excess label prior to imaging. At this point, the ability of the fluorescent probe to enter the cytosol is the limiting factor to *in cellulo* sub-stoichiometric labeling. Combining the flexibility of genetic code expansion with the diversity of labels in live cells allows for numerous creative applications that modulate cellular function.

Acknowledgements

NSF CHE-1112409, NSF MCB-1518265, OHSU-MRF and the OSU Cell Imaging and Analysis Facilities of the Environmental Health Sciences Center, P30 ES00210. We want to thank Dr. Joseph Fox for providing sTCO and Nathan Jespersen, John Gamble and Dr. Scott Brewer for their assistance with kinetics measurements and data analysis.

Chapter 3

Tetrazine Incorporation and Reaction in Eukaryotic Cells

Introduction

Ideal bioorthogonal ligations, while useful in *E. coli* have the ability to truly shine when used in eukaryotic cells (Fig 3.1 A). In eukaryotic cells there are numerous cellular pathways relevant to human health and disease that require a more complete understanding. Selective inhibition of a particular protein or activation of a protein through dimerization are useful tools for exploring protein function, that while possible with ideal bioorthogonal reactions in *E. coli*, would be at their greatest potential in live human cells.

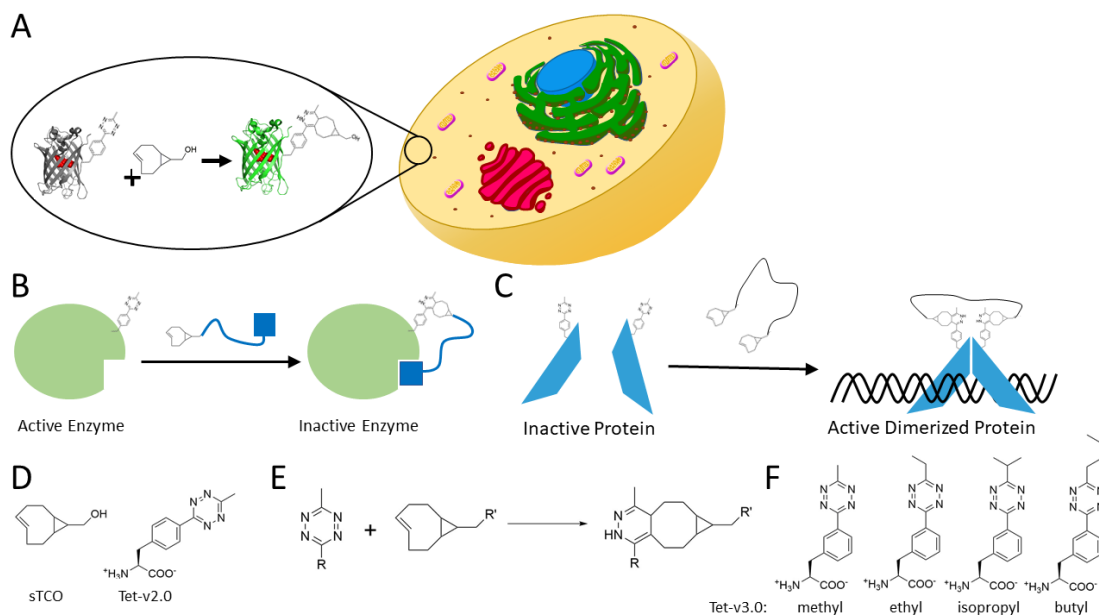


Fig 3.1: Eukaryotic Incorporation of Tetrazines Enables Control of Proteins

A) Reactions of tetrazines in live eukaryotic cells enables modification of proteins. B) Addition of an inhibitor to an enzyme via bioorthogonal ligations enables the selective inhibition of the protein of interest over other cellular proteins. C) Homodimerization of proteins results in activation and DNA binding. D) Structure of sTCO and Tet-v2.0. E) Reaction of tetrazines with sTCO. F) Structures of Tet-v3.0 amino acids.

Bioorthogonal ligations have been genetically encoded in eukaryotic cells for the purposes of fluorescent microscopy⁴ and protein inhibition²⁸. Common bioorthogonal functional groups that have been incorporated include TCOs¹²², BCN¹⁷, and alkynes³. Such labeling reactions tend to use an excess of label followed by thorough washing steps^{18,34,51}, though fluorogenic reactions have been used such that no washing step is required¹²³. Any secondary step increases the overall time the experiment takes and as such decreases the temporal resolution of events that can be observed. Additionally, requiring an excess of labeling reagent locks the user into

labeling the biomolecule of interest completely as opposed to a defined sub-stoichiometric reaction such that a fraction of the cellular protein could be modified. There is room for improvement.

Difficulties in eukaryotic cell work are not limited to bioorthogonal chemistry. The most common method for introducing tRNAs, aaRSs, or amber codon interrupted genes is through transient transfection. Doing so results in modification of a fraction of available cells. Even when successfully transfected, these genes can suffer a variety of problems not found in bacterial systems including nonsense mediated decay of mRNA¹²⁴, nuclear importation of the PylRS¹²⁵, or ratios of tRNA to aaRS that are not ideal¹²⁶. While steps can be taken to minimize these effects, ideal conditions for eukaryotic expression of GCE components are still under investigation.

Many of the difficulties of *E. coli* labeling of protein are also applicable in eukaryotic cells including cell permeability to dyes. In *E. coli* it was found that TAMRA-sTCO is not permeable to the membrane. In general, many of the fluorescent reporters that are suitable *in vitro* are not suitable in cells¹²⁷. This is an important hurdle to overcome with respect to fluorescent labeling of tetrazine containing cells.

Results/Discussion

The Tet-v2.0 amino acid was an obvious starting point since this tetrazine amino acid exhibited ideal bioorthogonal reactivity on protein in *E. coli* cells. To identify an orthogonal *Methanosarcina barkeri* pyrrolysyl tRNA synthetase/tRNA_{CUA} pair (Mb-PylRS/tRNA) able to incorporate Tet-v2.0 in response to an amber codon, we screened three libraries of MbPylRS variants in which five active-site residues were randomly mutated to all 20 amino acids (Table 3.1). Each codon was mutated to all 32 sequences following an NNK motif where N is any nucleotide and K is guanine or cytosine resulting in 3.35×10^7 Mb-PylRS members per library (Fig 3.2). Selections were performed by performing alternating rounds of positive and negative selection using chloramphenicol acetyl transferase and barnase, respectively^{76,103}. While control selections performed in parallel with a smaller aromatic ncAA, 3-nitroTyrosine, resulted in functional tRNA/aaRS pairs, no mutant tRNA/aaRS pairs

capable of suppressing amber codons with Tet-v2.0 were identified from multiple attempts with these three aaRS libraries.

Library	Library Residues (<i>M. mazei</i> / <i>M. barkeri</i>)				
A	L305/L270	Y306/Y271	L309/L274	N346/N311	C348/C313
B	M276/M241	A302/A267	Y306/Y271	L309/L274	C348/C313
C	N346/N311	C348/C313	V401/V366	W417/W382	G421/G386

Table 3.1: Mutants of the PylRS Libraries

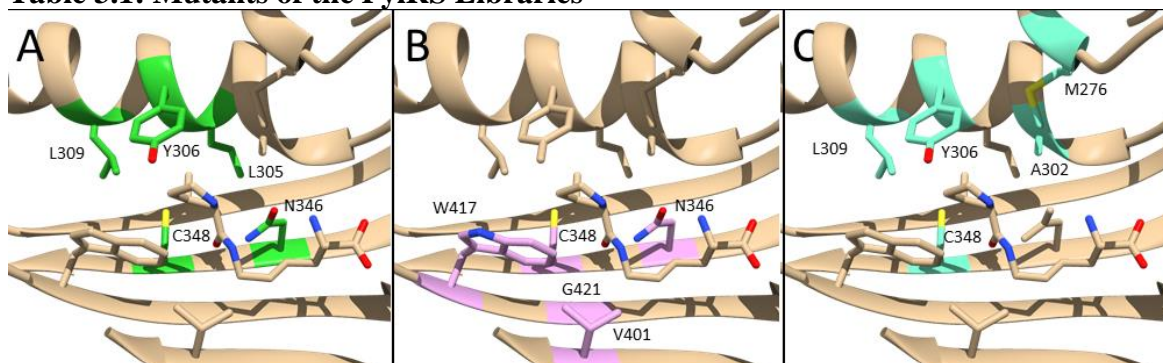


Fig 3.2. Amino acid libraries of PylRS active sites.

A) Structure of *M. mazei* PylRS with mutants of the library A. B) Structure of *M. mazei* PylRS with mutants of the library B. Structure of *M. mazei* PylRS with mutants of the library C.

The crystal structure of *M. mazei* synthetase was evaluated using Tet-v2.0 docked in the active site to identify key interactions that might improve the selection process (Fig 3.3) ¹²⁸. Tet-v2.0 was superimposed in place of the backbone of the native pyrrolysine substrate (Fig 3.3A). Potential clashes were identified with the side chains of W417, N346, and C348 as well as with backbones of G419 and I418 (Fig 3.3B). Unfortunately, potential clashes with the peptide backbone of the beta-strand in the RS active side with the methyl substituent and tetrazine ring were observed. Since the aaRS libraries are not designed to restructure the peptide backbone in the aaRS active-site

this conflict was believed to be the reason no successful selections were possible with Tet-v2.0 and these RS libraries.

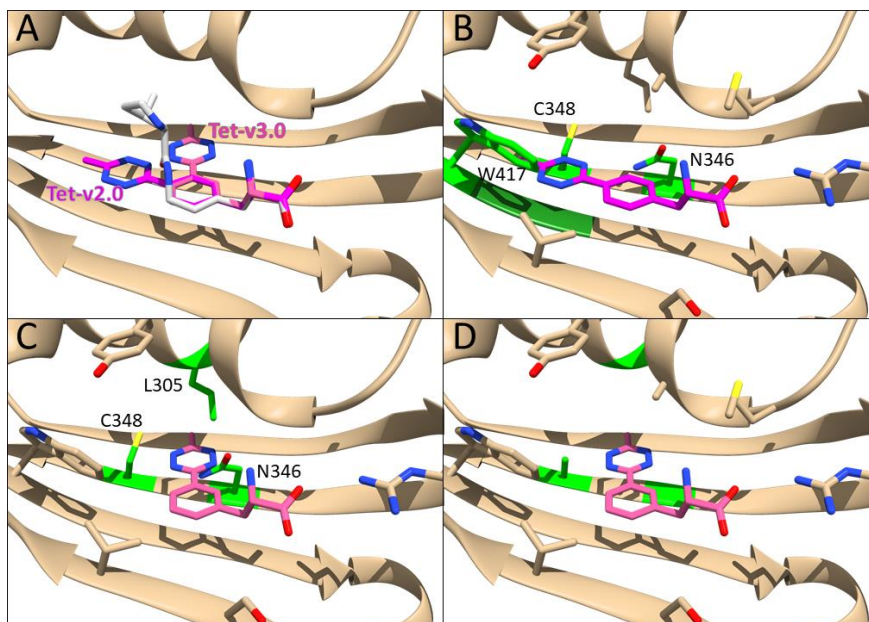


Fig 3.3. Structures of Tetrazine Amino Acids in PylRS

A) Tet-v2.0 and Tet-v3.0 superimposed upon Pyrrolysine (white) in the PylRS active site (2ZCE). B) Potential clashes between Tet-v2.0 and the side chains of active site residues (light green) as well as potential clashes with the backbone of active site residues (dark green). C) Potential clashes between Tet-v3.0 and the side chains of active site residues (light green). D) Mutations found in R2-84 synthetase (light green) widen the active site to accommodate Tet-v3.0.

Instead of generating new active site libraries to accommodate Tet-v2.0, we decided to alter the structure of the tetrazine amino acid to limit overlap with the protein backbone. One option would be a shorter tetrazine amino acid without a first aromatic ring. While this structural change has a good chance of allowing the resulting amino acid to fit into the active site it would also result in a tetrazine amino acid more difficult to synthesize. Synthesis of tetrazines not substituted by an aromatic ring have lower yields¹¹⁵. A second option that should maintain the tetrazine stability and rate constant of the IEDDA reaction is shifting the para substituted tetrazine ring on the phenylalanine to the meta position. Tet-v3.0 amino acids (Table 3.3) were designed such that the meta substituted tetrazine ring would be positioned deeper into the binding pocket of the PylRS active site. Additionally, this amino acid has greater flexibility in that rotation about the CB-CG bond results in a wider variety of available positions of the tetrazine ring. We synthesized the 3-(6-alkyl-*s*-tetrazin-3-yl)phenylalanine, Tet-

v3.0 and a series of molecules derived from Tet-v3.0 in two steps using standard coupling conditions.

Stopped flow measurements were used to determine the rate constants of the reaction between the free Tet-v3.0 amino acids and sTCO (Figure 2B). These determined second order rate constants have similar values to Tet-v2.0-methyl with the exception of Tet-v3.0-isopropyl that has a rate constant of $5830 \text{ M}^{-1}\text{s}^{-1}$. This 80% drop in rate constant we attributed to the increased steric bulk of the isopropyl group hindering reaction with sTCO since steric hindrance reduced the IEDDA reaction rate⁵⁵. Since Tet-v3.0-methyl, Tet-v3.0-ethyl, Tet-v3.0-butyl have similar reaction rate with sTCO as compared to Tet-v2.0 it is expected that when integrated into a protein these will also have similar rates. The reported rates for Tet-v2.0 were obtained for amino acid in a protein environment, and that IEDDA reactions are known to be solvent dependent⁶⁰.

Name of Compounds	Structure	Reaction rates with sTCO in PBS (pH 7.4)
Tet-v2.0		$29450 \text{ M}^{-1}\text{S}^{-1}$
Tet-v3.0-methyl		$24780 \text{ M}^{-1}\text{S}^{-1}$
Tet-v3.0-ethyl		$22202 \text{ M}^{-1}\text{S}^{-1}$
Tet-v3.0-isopropyl		$5830 \text{ M}^{-1}\text{S}^{-1}$
Tet-v3.0-butyl		$20569 \text{ M}^{-1}\text{S}^{-1}$

Table 3.2: Stopped Flow Rate Characterization of Tet-v3.0 Amino Acids

Standard selection methods using Tet-v3.0-methyl as the amino acid of interest against the aaRS libraries A and B generated many useful synthetases¹⁰³. Two rounds of positive and negative selections were performed before 96 synthetases were screened for the ability to suppress an amber codon in a GFP gene. Of the evaluated synthetases, 15 were found to have degrees of incorporation and were sequenced. Six unique sequences were identified (Table 3.2). Of those, the R2-74 and the R2-84 synthetases demonstrated the best efficiency and fidelity. Not surprisingly the best functioning Tet-v3.0-RSs had mutations in N346 and C348 residues, consistent with opening the active site in the meta position required for Tet-v3.0 (Fig 3.3C).

Site (<i>Mb/Mm</i>)	R1-8	R1-35	R2-3	R2-40	R2-74	R2-84
L270/L305	-	-	G	-	G	G
Y271/Y306	-	-	-	-		-
L274/L309	-	-	-	-	-	-
N311/N346	A	A	G	G	G	G
C313/C348	S	V	A	A	S	A
Other (<i>Mb</i>)	-	-	-	K296R	-	R263C

Table 3.3: Sequences of Generated *Mb* Synthetases

Residues are listed in both as *M. mazei* and *M. barkeri* for reference to the above structural information.

Many synthetases generated for genetic code expansion demonstrate permissivity, the ability to incorporate ncAAs that have similar structures to the ncAA used in selections. We tested the top aaRS variants selected to incorporate Tet-v3.0-methyl for their ability to incorporate the Tet-v3.0 amino acids (Figure 3.4). All tested synthetases demonstrated permissivity towards the Tet-v3.0 amino acids, with similar levels of incorporation of the Tet-v3.0 amino acids suggests that a wider variety of structural variants of Tet-v3.0 can be incorporated than those tested here. The enables

the ability to tune the rate constant by substituting Tet-v3.0 with electron donating or withdrawing functional groups.

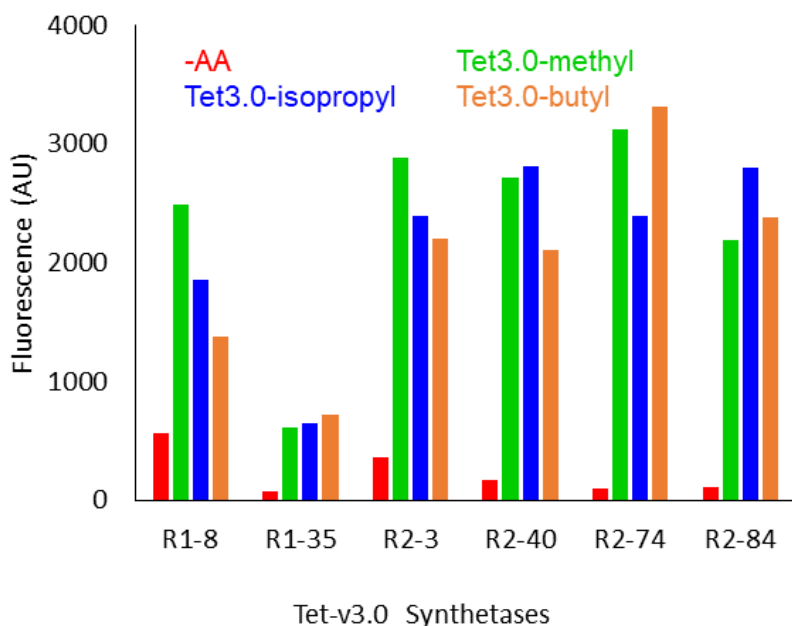


Fig 3.4: Efficiency of Generated Synthetases for Tet-v3.0 Amino Acids

Cells expressing synthetases selected for Tet-v3.0-methyl were grown in the presence of 1mM of the Tet-v3.0 amino acids. The corresponding fluorescence was measured and reported.

To verify the efficiency and fidelity of the R2-84 aaRS, GFP150-Tet-v3.0 was expressed and purified by affinity chromatography. Site specific incorporation of Tet-v3.0 as verified by MS analysis (Figure 3.5). The GFP150-Tet-v3.0-methyl protein shows a final mass of 27950.1 Da, a 127.1 Da increase relative to WT GFP (Figure 3.5). This corresponds to the expected mutation from the native asparagine residue to Tet-v3.0-methyl. To verify complete reactivity of GFP150-Tet-v3.0-methyl with sTCO, an excess of sTCO was incubated with the protein for 30 minutes. As expected the MS analysis of the resulting product GFP150-Tet3.0-methyl-sTCO shows an upward shift in mass to 124.2 Da. This shift of 124.2 Da corresponds to reaction of the tetrazine ring of Tet-v3.0-methyl with the added sTCO molecule. Similar shifts in the molecular weight were observed for Tet-v3.0-ethyl, Tet-v3.0-isopropyl, and Tet-v3.0-butyl (Table 3.4) No unreacted GFP150-Tet-v3.0 is detected, indicating a quantitative reaction. Reactions of Tet-v3.0 amino acids with sTCO on proteins also show complete labeling. The reactivity of the incorporated Tet-v3.0 amino acids was also verified by

adding sTCO-PEG₅₀₀₀ to purified GFP150-Tet-v3.0 (Fig 3.6). Upon reaction with sTCO-PEG₅₀₀₀ an increase in the apparent mobility of GFP150-Tet-v3.0 was observed. This shift in mobility corresponds to the formation of GFP150-Tet-v3.0- sTCO-PEG₅₀₀₀.

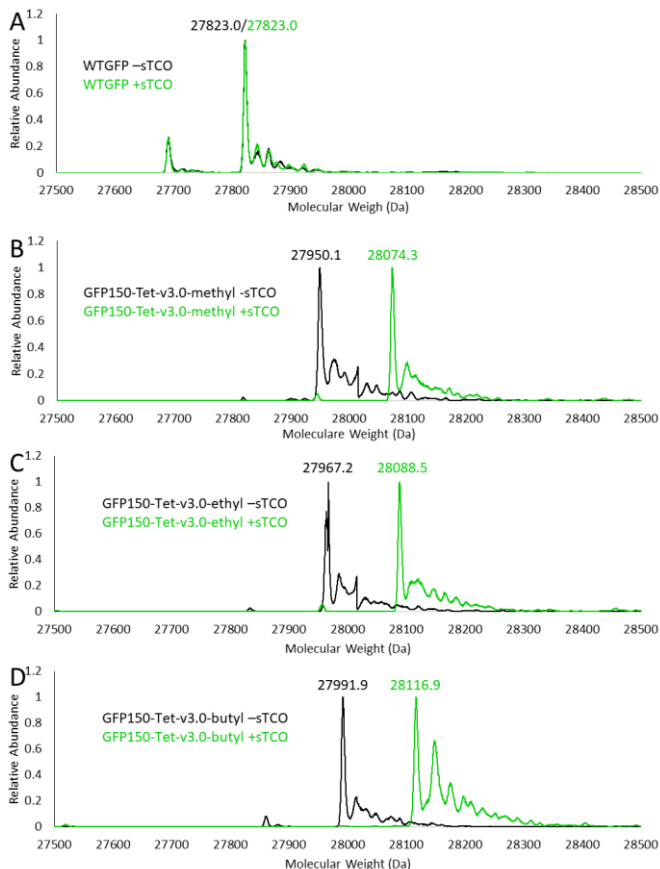


Fig 3.5: Mass Spectra of Tetrazine GFP and Reactions

A) WT GFP was reacted with sTCO without change to the molecular weight. B) GFP150-Tet-v3.0-methyl upon reaction with sTCO resulted in an increase in the protein molecular weight of 124.2 Da compared to an expected molecular weight increase of 124.2 Da. C) GFP150-Tet-v3.0-ethyl upon reaction with sTCO resulted in an increase in the protein molecular weight of 124.2 Da compared to an expected molecular weight increase of 124.2 Da. D) GFP150-Tet-v3.0-butyl upon reaction with sTCO resulted in an increase in the protein molecular weight of 124.2 Da compared to an expected molecular weight increase of 124.2 Da. Prominent peaks are detectable that correspond to cleavage of the N-terminal methionine. Incorporation of Tet-v3.0 variants at site 150 results in the expected masses, but no peaks consistent with mis-incorporation of canonical amino acids.

Protein	Observed Mass	Expected Mass	Mass Above WT	Expected Mass Above WT
WTGFP	27823.0	27827.3	0	0
WTGFP +sTCO	27823.0	27827.3	0	0
GFP150-Tet- v3.0-methyl	27950.1	27954.4	127.1	127.1
GFP150-Tet- v3.0-ethyl	27967.2	27968.4	144.2	141.1
GFP150-Tet- v3.0-butyl	27991.9	27996.5	168.9	169.2
GFP150-Tet- v3.0-methyl +sTCO	28074.3	28078.6	251.3	251.3
GFP150-Tet- v3.0-ethyl +sTCO	28088.5	28092.6	265.5	265.3
GFP150-Tet- v3.0-butyl +sTCO	28116.9	28120.7	293.9	293.4

Table 3.4: Observed Masses of Tet-v3.0 containing GFP

While GFP masses were approximately 4 Da less than expected, WTGFP also showed this mass difference, and the difference in mass between WTGFP and the GFP150-Tet-v3.0 molecules were approximately what was expected.

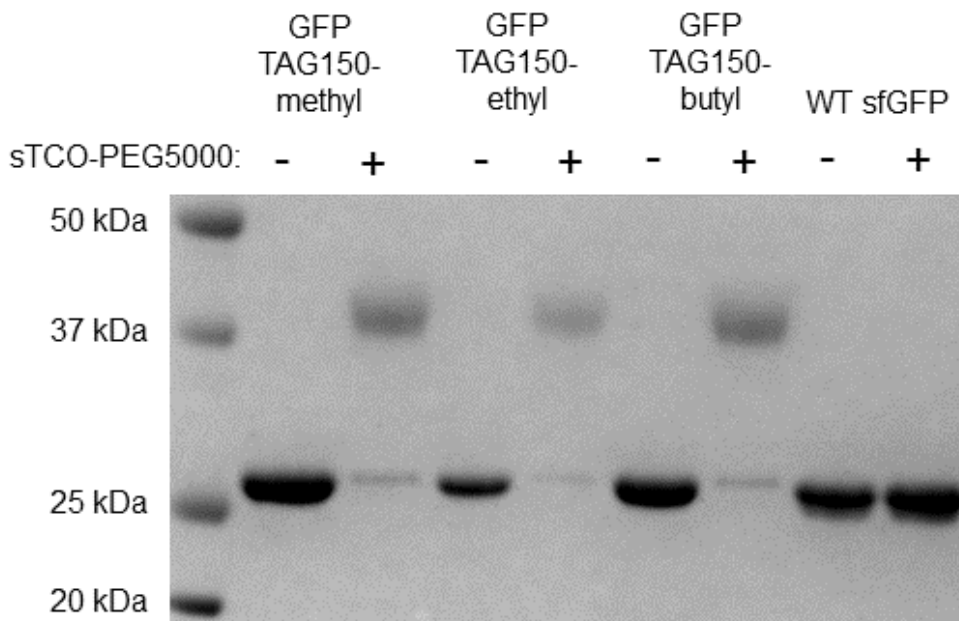


Fig 3.6: Mobility Shift Assay of GFP150-Tet-v3.0

GFP150-Tet-v3.0 amino acids were reacted with sTCO-PEG₅₀₀₀ and analyzed via SDS-PAGE. An upwards shift in mobility is observed consistent with reaction. A small unshifted band remains.

The reaction of tetrazine amino acids in proteins can be monitored when a tetrazine amino acid is incorporated at site 150 on GFP. The presence of a tetrazine amino acid quenches GFP fluorescence. Full fluorescence is returned by reaction of the tetrazine.¹¹¹ Upon reaction of GFP150-Tet-v3.0 and sTCO a 4-6 fold increase in fluorescence is observed. Using this fluorescence increase upon reaction the reaction rates of GFP-Tet-v3.0 amino acids with sTCO were measured. (Fig 3.7). The observed rates are significantly faster than the rates observed between the reactions of the free amino acids and sTCO. The differences in rate may be attributable to differences in polarity between the protein environment and buffer. Once again, the rate constant of the reaction between Tet-v3.0-isopropyl and sTCO is approximately 25% that of the other Tet-v3.0 variants, potentially due to steric hinderance.

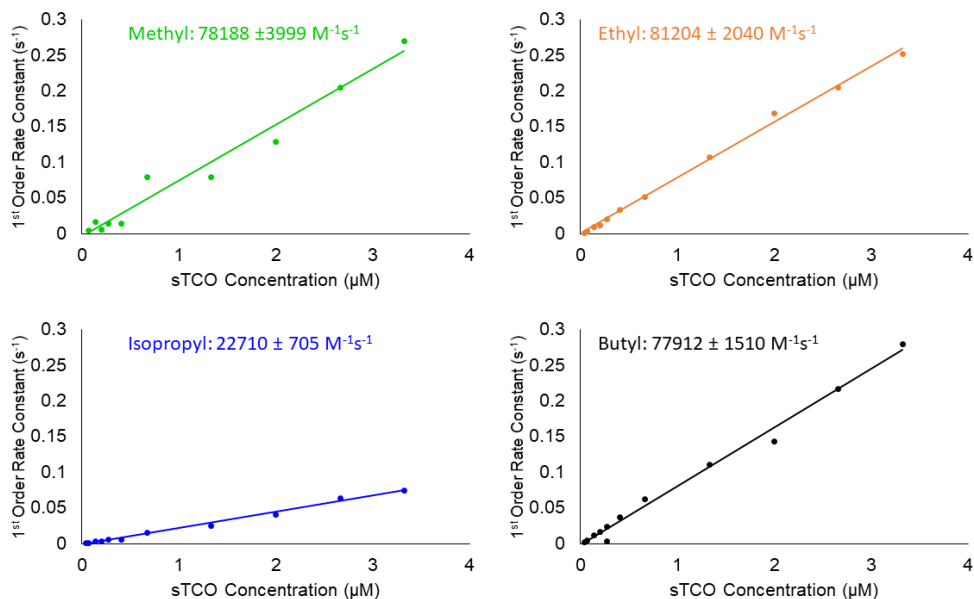


Fig 3.7: Protein Kinetics of the Reaction between GFP150-Tet-v3.0 and sTCO

Kinetic Characterization of the reaction rates between GFP150-Tet-v3.0 amino acids and sTCO. Individual sTCO concentrations were fit to 1st order exponential equations and the observed rate constants were plotted against the sTCO concentration. The reaction rates are similar to the reaction between Tet-v2.0 and sTCO on proteins with the exception of the reaction between sTCO and Tet-v3.0-isopropyl.

The Tet-v3.0 amino acids when incorporated into proteins maintain the reactivity and stability components to match that of the ideal reactivity of Tet-v2.0.

The synthetases R2-74 and R2-84 were codon optimized for mammalian cells and cloned into the pAcBac1 plasmid¹²⁹. This plasmid contains four genes corresponding to the orthogonal tRNA as well as the synthetase gene. A corresponding pACBAC plasmid containing four tRNA genes as well as a TAG interrupted GFP gene was used in conjunction with synthetase pACBAC.

The resulting pAcBac1-R2-84 plasmid and pAcBac1-R2-74 plasmid were independently cotransfected into HEK293T cells with the reporter pAcBac1-sfGFP-TAG150 plasmid. A positive control for full expression efficiency was evaluated using pAcBac1-sfGFP uninterrupted by amber stop codons. The Tet-v3.0 amino acids showed HEK293 cell toxicity when added to the media at the 1 mM ncAA concentration used for the *E. coli* expression. To identify the maximum allowable Tet-v3.0 concentration, cell toxicity was evaluated for decreasing amounts of each Tet-v3.0 amino acid (Fig 3.8).

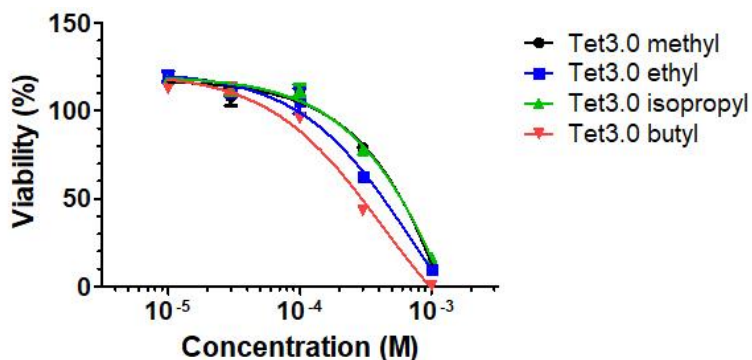


Fig 3.8: Toxicity of Tet-v3.0 Amino Acids

Cells incubated with varying concentrations of Tet-v3.0 amino acids for 72 hours were assayed with a CellTiter Glo Imaging Kit to test viability. Decreases in viability are observed at concentrations at or above 100 μ M.

Since HEK293 cells could only tolerate up to 0.1 mM Tet-v3.0 amino acid we predicted that the reduced concentration of the ncAA in the media could result in lower suppression efficiency. To assess the concentration dependence of R2-74 and R2-84 for the Tet-v3.0 series we measured GFP-150-Tet-v3.0 incorporation efficiency in *E. coli* at reduced concentrations of Tet-v3.0. Doing so generated an ncAA dependency relationship for each aaRS (Fig 3.9). Consistent high cellular GFP-Tet-v3.0 fluorescence at above 0.2 mM ncAA indicated that all Tet-v3.0 amino acids were good substrates for both Tet-v3.0RSs at these concentrations. At concentrations below 100 μ M Tet-v3.0 the efficiency of the two Tet-3.0 RSs was different for producing GFP-Tet-v3.0. Notably, the R2-84-RS was more effective at producing protein at low

substrate concentrations and the Tet-v3.0-butyl ncAA was the most effective substrate at low concentrations.

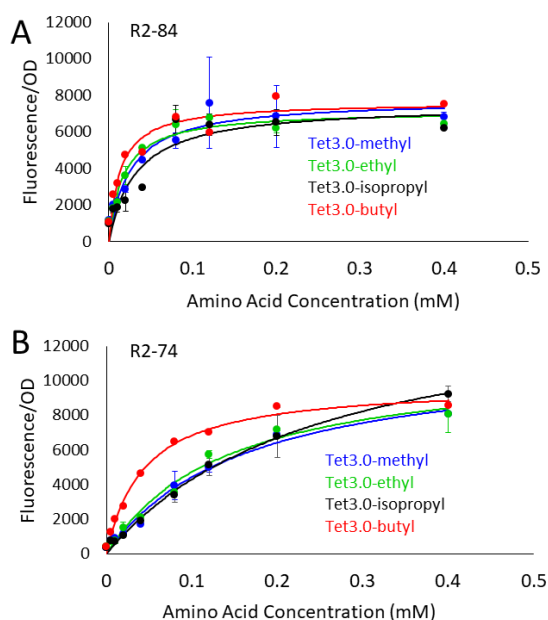


Fig 3.9: Concentration Dependence Curves of Tet-v3.0

A) Concentration dependence of Tet-v3.0 amino acids on expression of GFP150 with the R2-84 synthetase in *E. coli*. B) Concentration dependence of Tet-v3.0 amino acids on expression of GFP150 with the R2-74 synthetase in *E. coli*. The R2-84 synthetase was found to be able to produce more GFP at lower concentrations of Tet-v3.0-butyl than the other Tet-v3.0 derivatives.

Amino Acid	R2-84	R2-74
Tet-v3.0-methyl	25.7 μ M	163.5 μ M
Tet-v3.0-ethyl	17.4 μ M	140.0
Tet-v3.0-isopropyl	32.4 μ M	251.0 μ M
Tet-v3.0-butyl	13.5 μ M	43.5 μ M

Table 3.5: UP50 values for the R2-84 and R2-74 Synthetases

UP₅₀ values consist of the concentration at which the half-maximal amount of GFP was produced.

Using the maximum concentration of 0.1 mM Tet-v3.0 amino acid tolerated by HEK293 cells, we assessed the incorporation efficiency and fidelity of the top performing Tet-v3.0-RSs (R2-74 and R2-84) using the GFP-TAG interrupted reporter. The amber codon suppression in HEK293 cells was evaluated for all four different Tet-v3.0 amino acid structures using both Tet-v3.0 tRNA/aaRS pairs (Fig 3.10). The fidelity for both Tet-3.0-aaRSs was the same, showing no background incorporation of natural amino acids. HEK293 cells suppression efficiency varied considerably for the different Tet-v3.0 amino acids with the longest ncAA Tet-v3.0 butyl showing clear

preference over the shortest Tet-v3.0 methyl in agreement with the concentration dependence seen in *E. coli*.

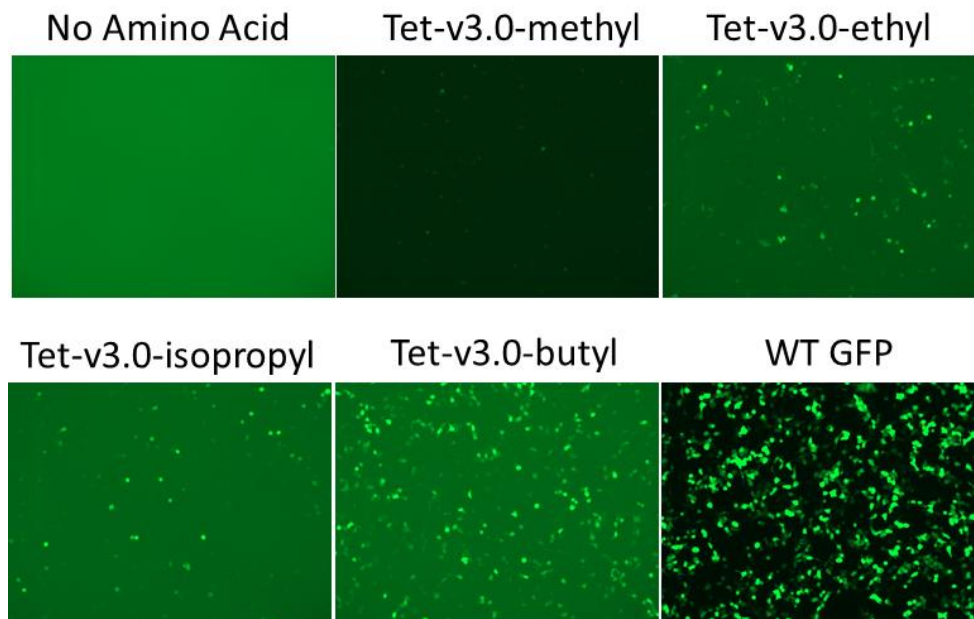


Fig 3.10: HEK293T Cell Production of GFP150-Tet-v3.0

HEK293T cells were grown in the presence of 30 μM of Tet-v3.0 amino acids. Expression was observed for each of the Tet-v3.0 amino acids, though low levels of expression were observed for Tet-v3.0-methyl. Tet-v3.0-butyl was observed to have the greatest level of fluorescence.

In an effort to improve yields, nuclear export sequences were cloned N-terminal to the counteract nuclear importation present natively in the PylRS. Transfection of HEK293T cells in with a nuclear exportation fused PylRS resulted in greater fluorescence than in the absence of the nuclear exportation sequence (Fig 3.11).

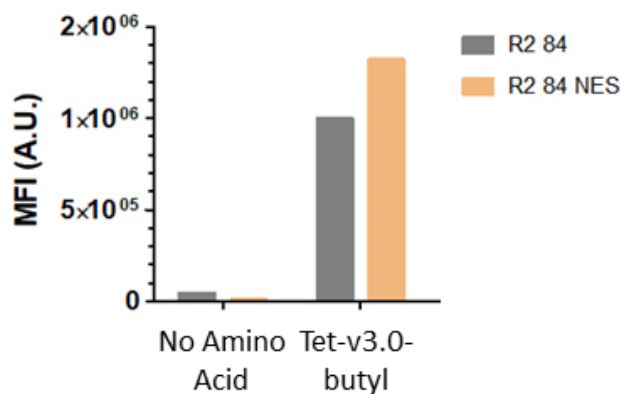


Fig 3.11: The Effect of a Nuclear Export Sequence on GFP Production

Nuclear export sequences fused to R2-84 results in improved expression of GFP-Tet-v3.0-butyl in cells.

To assess *in cellulo* reactivity, GFP expressed in HEK cells was labeled with TAMRA-sTCO. The previously synthesized TAMRA-sTCO molecule (Chapter 2) was used to as the label of choice due to the ability of TAMRA-fluorophores to diffuse through cell membranes¹²⁷. TAMRA-sTCO was incubated with HEK293T cells expressing GFP150-Tet-v3.0-butyl that had been washed to remove the free tetrazine amino acid. While the expected result of TAMRA fluorescence for cells expressing GFP150-Tet-v3.0-butyl was observed, fluorescence was also observed for transfected cells in a non-specific labeling reaction. This non-specific labeling may be a result of either reactions between the sTCO and cellular components or non-covalent interactions between the TAMRA-sTCO molecule and the cell.

To further characterize this interaction, untransfected HEK293T cells were labeled with TAMRA-sTCO as well as the free TAMRA acid (Fig 3.12). In addition, an unreactive isomerized version of the TAMRA-sTCO molecule (TAMRA-sCCO) (Fig 3.12) was synthesized and used to label untransfected cells. Cells were analyzed via flow cytometry. Whereas the free TAMRA acid showed no to little nonspecific labeling of cells, both the TAMRA-sTCO and TAMRA-sCCO showed significant labeling of the cells in a concentration dependent manner. As TAMRA-sCCO is unreactive, this result indicates that non-covalent interactions are responsible for observed non-specific interactions. In addition, TAMRA-dTCO and TAMRA-oxoTCO were incubated with untransfected HEK293T cells with some degree of non-specific

adhesion. These compounds are less reactive than the TAMRA-sTCO molecule. Though non-covalent interactions may differ from the TAMRA-sTCO molecule as well.

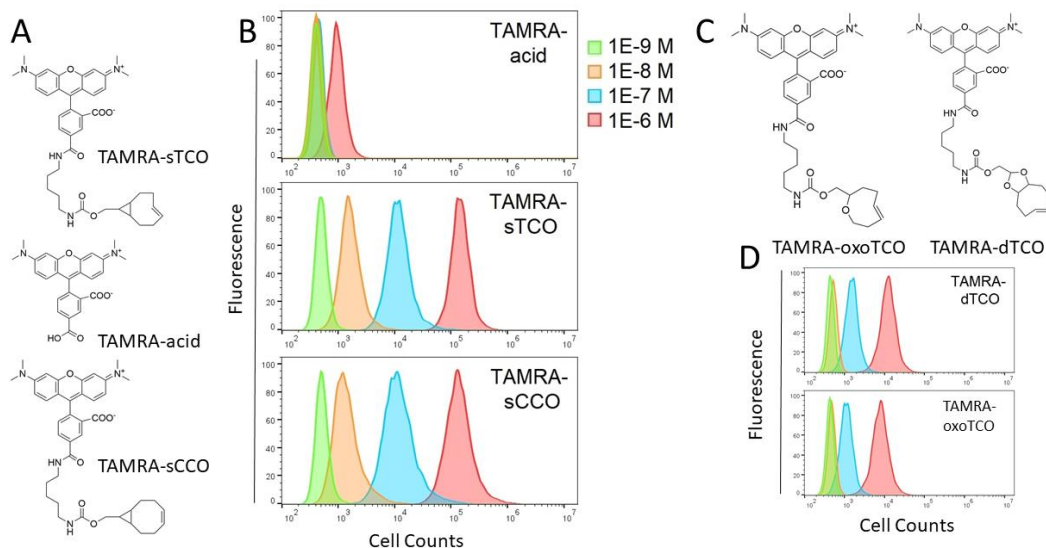


Fig 3.12. Flow Cytometry of TAMRA labeled HEK293T cells.

A) Structures of TAMRA molecules used to label cells. B) Flow cytometry of TAMRA treated cells shows adhesion of TAMRA-sTCO and TAMRA-sCCO to HEK293T cells. C) Structures of TAMRA-oxoTCO and TAMRA dTCO. D) TAMRA-dTCO and TAMRA-oxoTCO show moderate levels of adhesion to HEK293T cells.

HEK293T cells expressing GFP150-Tet-v3.0-butyl were washed and labeled with the TAMRA dyes. The labeled cells were then subject to two-dimensional flow cytometry measuring both the TAMRA fluorescence and the GFP fluorescence. For each set of cells GFP fluorescence can be broken down into two distinct subsets: Cells that have not been successfully transfected resulting in low fluorescence, and cells expressing GFP. Upon labeling with TAMRA acid, no fluorescence increase is observed in either subset of cells (Fig 3.13). Labeling with the unreactive TAMRA-sCCO results in fluorescence labeling of all cells regardless of GFP expression. When the cells were labeled with TAMRA-sTCO, all cells were labeled with TAMRA fluorescence, though cells expressing GFP150-Tet-v3.0-butyl were labeled to a higher degree. This indicates that labeling of cells with TAMRA-sTCO results in reaction with GFP-Tet-v3.0-butyl, though there is a high degree of non-specific labeling.

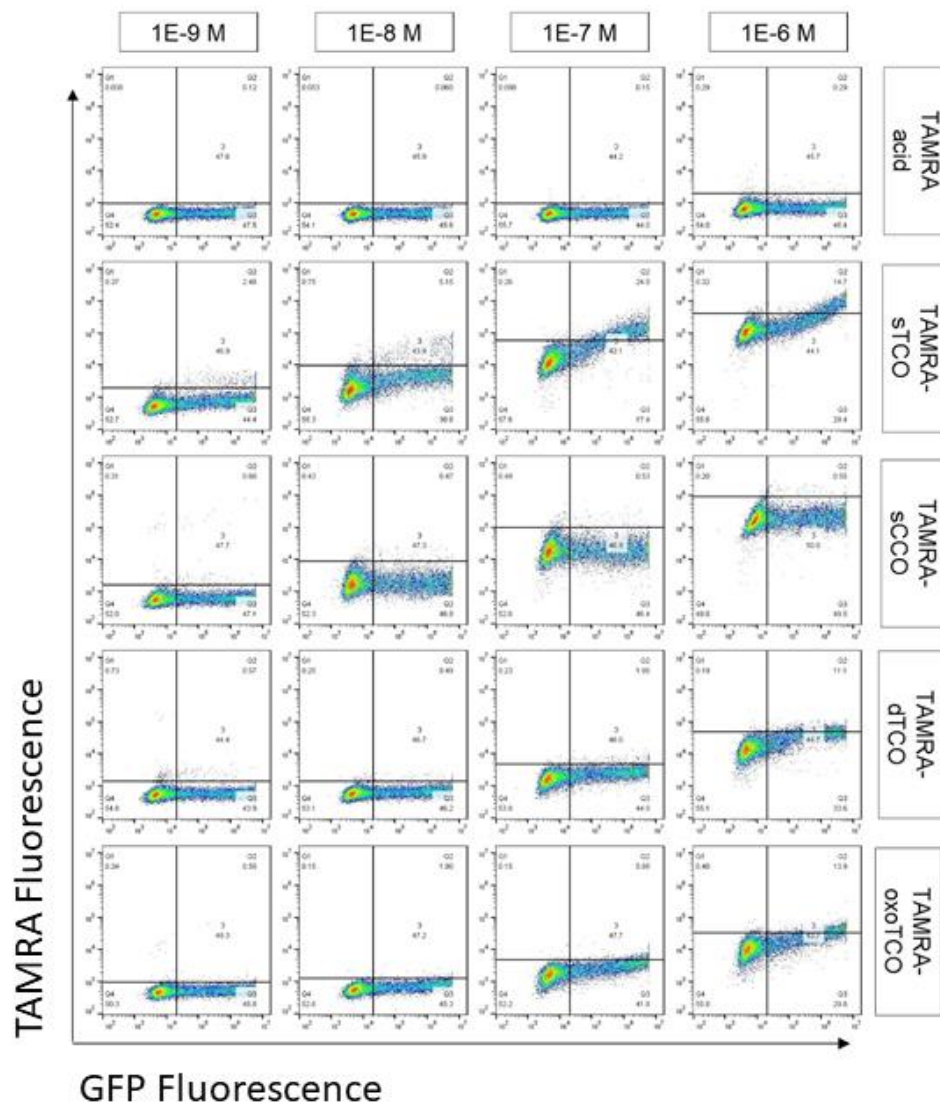


Fig 3.13. Two-Dimensional Flow Cytometry of Tet-v3.0 Expressing HEK293T Cells.

HEK293T cells expressing GFP150-Tet-v3.0-butyl were incubated for 30 min with varying concentrations of TAMRA fluorophores. Cells were washed and GFP and TAMRA fluorescence measured via flow cytometry. GFP fluorescence exhibits two distinct groups of cells: A head of cells that are either untransfected or not expressing GFP and a tail of cells expressing GFP-Tet-v3.0-butyl to varying degrees. TAMRA fluorescence as measured on the y-axis largely increases with the concentration of TAMRA fluorophores with the exception of the addition of TAMRA acid. With TAMRA-sTCO incubation, the TAMRA fluorescence of cells expressing GFP150-Tet-v3.0-butyl increases more than non-expressing cells. This behavior is seen to a slight degree in TAMRA-dTCO and TAMRA-oxoTCO.

To verify that cellular TAMRA labeling is not labeling cellular proteins non-specifically, cells expressing GFP150-Tet-v3.0-butyl were labeled with TAMRA-sTCO and washed to remove unreactive label. The cells were then treated with an excess of the Tet-v2.0 amino acid to react with any unreacted sTCO functional groups. A subset of cells was pre-incubated with the Tet-v2.0 amino acid and subsequently labeled with TAMRA-sTCO. The cells were lysed, and the lysate was analyzed via SDS-PAGE. Fluorescent imaging of the gels shows that solely the GFP150-Tet-v3.0-butyl was labeled with TAMRA-sTCO and that no non-specific protein labeling occurred (Fig 3.14). Further, TAMRA-sCCO labeling of GFP150-Tet-v3.0-butyl was non-existent.

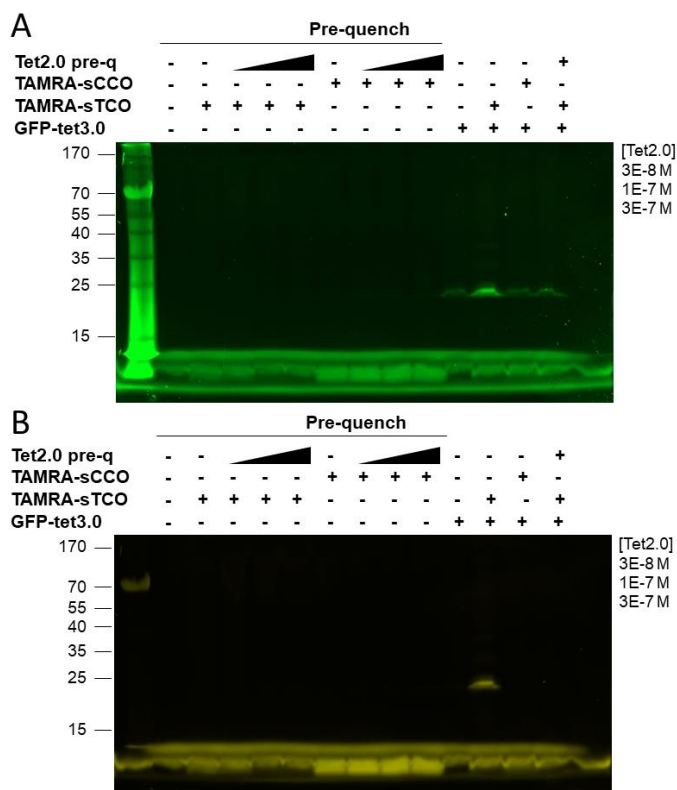


Fig 3.14. Background Reactions of TAMRA Fluorophores on HEK Cell Lysate.

HEK293T cells were labeled with TAMRA-sTCO and TAMRA-sCCO after being pre-quenched with Tet-v2.0. GFP150-Tet-v3.0-butyl expressing cells were labeled with TAMRA-sTCO and TAMRA-sCCO. Cells not pre-quenched with Tet-v2.0 were quenched after labeling with TAMRA dyes. The cells were then lysed and analyzed via SDS-PAGE. A) GFP fluorescence of cell lysates. B) TAMRA fluorescence of labeled cell lysates. It is of note that no non-specific labeling of proteins occurred except for a large band at the dye front and a set of small bands corresponding to GFP. No TAMRA labeling was observed in pre-quenched cells. GFP fluorescence is greatest in cells that were successfully reacted with TAMRA-sTCO. This may be due to the fluorescence increase upon reaction observed in GFP150-Tet-v3.0-butyl or bleed-through of the TAMRA fluorescence.

An alternative method to assess the *in cellulo* reaction of the labeled protein, utilizes the tetrazine quenching of GFP used to evaluate the rate of the reaction. HEK293T cells expressing GFP150-Tet-v3.0-butyl were reacted with TAMRA-sTCO and sTCO-OH. SDS-PAGE was performed on the lysate of the reacted cells (Fig 3.15). An increase in the fluorescence of the GFP in a sTCO and TAMRA-sTCO dependent manner was observed. Additionally, the increase in GFP fluorescence corresponds to the increase in the TAMRA fluorescence for the TAMRA-sTCO labeled cells. This demonstrates that the reaction between Tet-v3.0-butyl and sTCO can be used for labeling reactions in live eukaryotic cells.

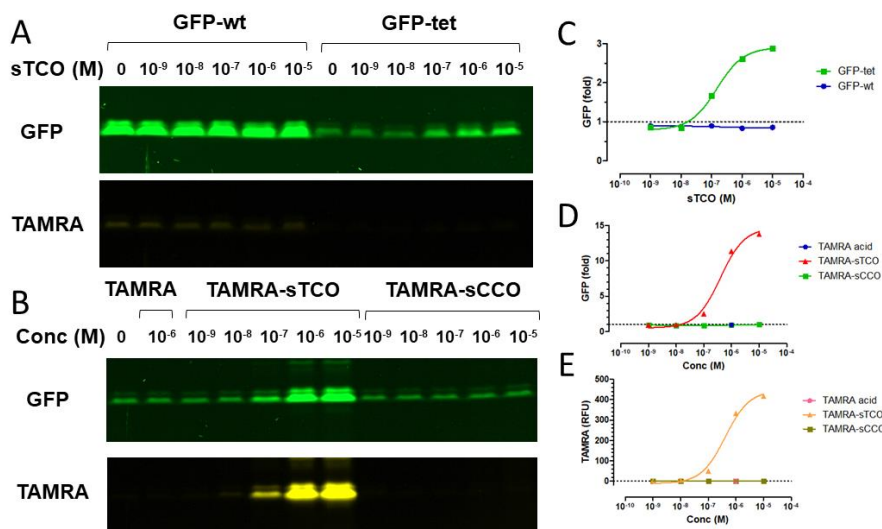


Fig 3.15. In Cell Fluorescence Increase of GFP150-Tet-v3.0-butyl upon Reaction.

A) WT GFP expressing cells and cells expressing GFP-Tet-v3.0 were labeled with sTCO-OH. Fluorescence increase was observed upon reaction. B) GFP150-Tet-v3.0-butyl expressing cells were labeled with TAMRA acid, TAMRA-sTCO, and TAMRA-sCCO dyes. Only TAMRA-sTCO showed notable labeling at any concentration. GFP Fluorescence was also observed. C) Quantitation of GFP fluorescence upon reaction with sTCO-OH in A). D) Quantitation of GFP fluorescence upon reaction with TAMRA-sTCO in B). E) Quantitation of TAMRA fluorescence upon reaction with TAMRA-sTCO in B).

Many proteins form homodimers *in vivo* as a method of regulating activity¹³⁰. While methods exist to mimic dimer formation and study this regulation¹³¹, site specific GCE for the introduction of bioorthogonal ligations is an excellent tool that would allow the controlled formation of artificial homodimers with minimal perturbation to native protein structure and function. To perform such dimerization reactions double-headed sTCO molecules with 100 Da and 500 Da PEG linkers (dh-sTCOPEG₁₀₀ and dh-sTCOPEG₅₀₀) were synthesized. Cells expressing GFP150-Tet-v3.0-butyl were

reacted with these linkers as well as sTCO-PEG₁₀₀₀ and sTCO-PEG₅₀₀₀ labels. The cells were quenched, lysed, and analyzed via SDS-PAGE. Fluorescent analysis shows that while the double-headed PEG linkers were capable of dimerizing GFP, the sTCO-PEG₁₀₀₀ and sTCO-PEG₅₀₀₀ label failed to shift the apparent molecular weight of intracellular GFP (Fig 3.16). This suggests that the double-headed linkers were able to pass through the cell membrane while the sTCO-PEG₁₀₀₀ and sTCO-PEG₅₀₀₀ were not.

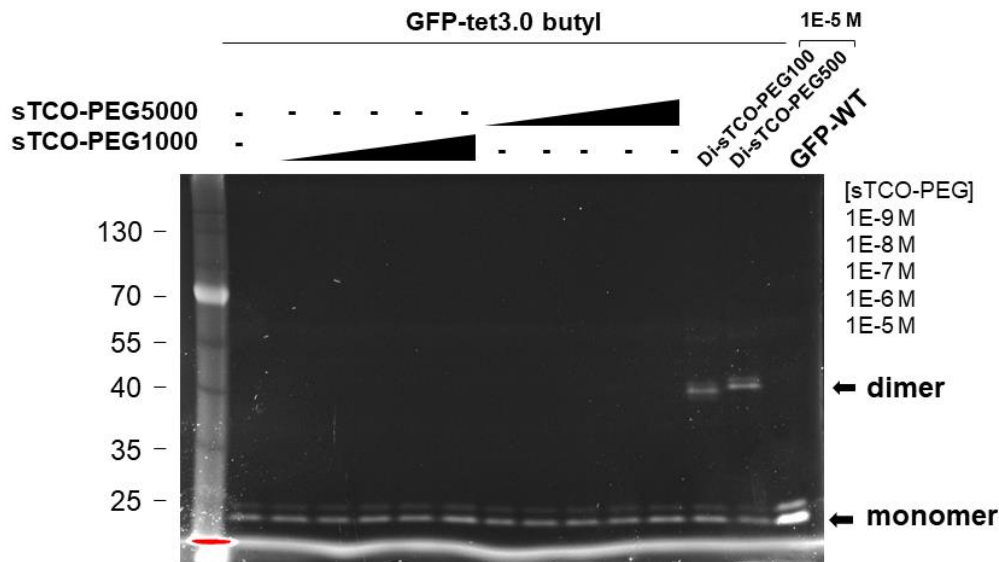


Fig 3.16. Mobility Shift of HEK293T cell expressed GFP150-Tet-v3.0-butyl.

Cells expressing GFP150-Tet-v3.0-butyl were reacted with sTCO-PEG compounds and the cell lysate was analyzed by SDS-PAGE. No mobility shift was observed for sTCO-PEG₁₀₀₀ or sTCO-PEG₅₀₀₀. Dimerization was observed for both dh-sTCOPEG₁₀₀ and dh-sTCOPEG₅₀₀.

Reactions between excess dh-sTCO linkers and GFP150-Tet-v3.0-butyl would only be expected to form dimers if the dh-sTCO linker is limiting. Excess of the dh-linker would result in the majority of GFP150-Tet-v3.0-butyl reacted with a dh-sTCO molecule that is not linked to a second protein. If diffusion through the cell membrane is limiting, reaction of dh-sTCO-PEG linkers would remain limiting *in cellulo* for the duration of the reaction. This enables investigation of the diffusion through HEK cell membranes by reacting expressed GFP-150-Tet-v3.0-butyl with excess linker. When dh-sTCOPEG₁₀₀ and dh-sTCOPEG₅₀₀ were titrated into cells expressing GFP150-Tet-v3.0-butyl, low concentrations of both linkers showed low levels of dimer formation (Fig 3.17). As the linker concentrations rose, the amount of dimer formed increased. Large excess of linker resulted in a drop in dimer formation for the dh-sTCOPEG₁₀₀

whereas the dh-sTCOPEG₅₀₀ showed no such concentration dependent decrease. This shows differential rates of diffusion between the two linkers such that the rate of dh-sTCOPEG₅₀₀ diffusion across the cell membrane is limiting. This suggests a size dependent diffusion rate. As additional evidence neither the sTCO-PEG₁₀₀₀ and sTCO-PEG₅₀₀₀ showed any cellular diffusion.

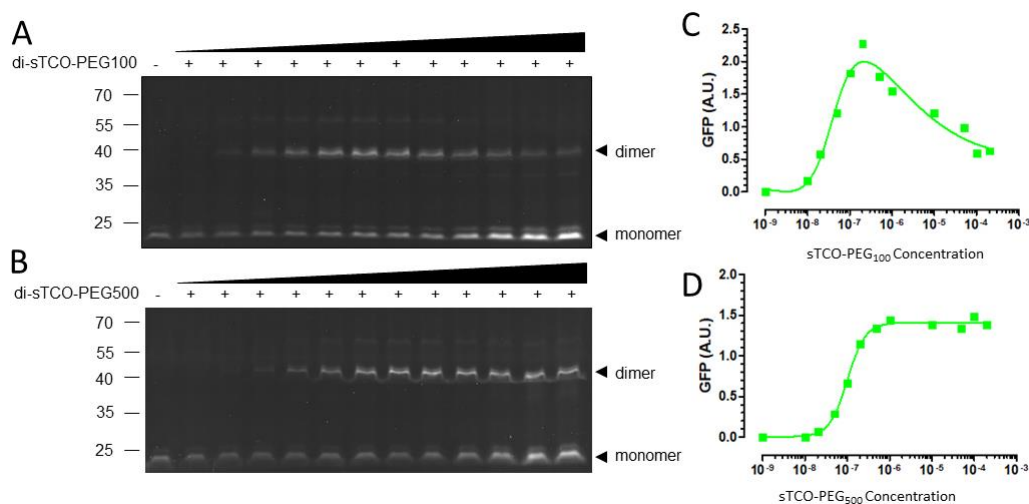


Fig 3.17. Concentration Dependent Dimerization of GFP150-Tet-v3.0-butyl.

A) HEK293T cells expressing GFP150-Tet-v3.0-butyl were expressed and reacted with varying concentrations of sTCO-PEG₁₀₀. B) HEK293T cells expressing GFP150-Tet-v3.0-butyl were expressed and reacted with varying concentrations of sTCO-PEG₅₀₀. C) Quantitation of dimer fluorescence of A). D) Quantitation of dimer fluorescence of B).

Conclusion

In total, a tetrazine containing amino acid has been generated that both has the ideal properties demonstrated with Tet-v2.0 and is capable of being expressed in proteins in eukaryotic cells. This can enable modification of proteins in live eukaryotic cells within a matter of minutes and can potentially be used for sub-stoichiometric labeling reactions. While the benefit of an ideal reaction is enormous, the amino acid used here is not without drawbacks. The Tet-v3.0 derivatives show toxicity to eukaryotic cells that may alter native metabolism. Additionally, concentrations of Tet-v3.0 used to overcome this toxicity result in limited production of protein incorporating Tet-v3.0 compared to WT controls. Finally, many of the experiments may be limited by poor cellular diffusion or by fluorescence adhesion to cells. Better fluorophores that

show limited cellular adhesion should be developed to enable sub-stoichiometric protein labeling.

Materials and Methods

General Synthetic Methods:

All purchased chemicals were used without further purification. Anhydrous dichloromethane was used after overnight stirring with calcium hydride and distillation under argon atmosphere. Thin-layer chromatography (TLC) was performed on silica 60F-254 plates. The TLC spots of alkene were charred by potassium permanganate staining. Flash chromatographic purification was done on silica gel 60 (230-400 mesh size). ^1H NMR spectra were recorded at Bruker 400MHz and 700 MHz and ^{13}C NMR spectra were recorded at 175 MHz. Coupling constants (J value) were reported in hertz. The chemical shifts were shown in ppm and are referenced to the residual non-deuterated solvent peak CDCl_3 ($\delta = 7.26$ in ^1H NMR, $\delta = 77.23$ in ^{13}C NMR), CD_3OD ($\delta = 3.31$ in ^1H NMR, $\delta = 49.2$ in ^{13}C NMR), d_6 -DMSO ($\delta = 2.5$ in ^1H NMR, $\delta = 39.5$ in ^{13}C NMR) as an internal standard. Splitting patterns of protons are designated as follows: s-singlet, d-doublet, t-triplet, q-quartet, quin-quintet, sext-sextet, sept-septet, m-multiplet.

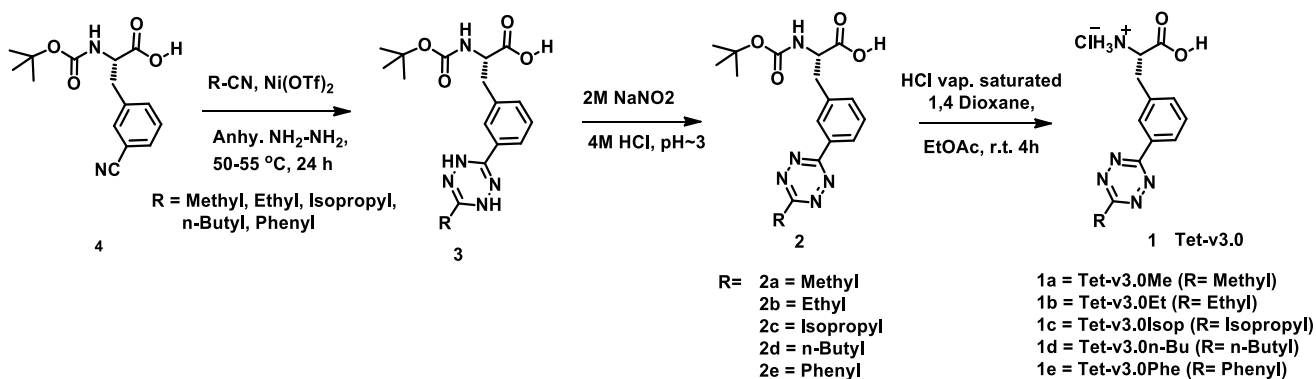


Fig 3.18. Synthesis of s-tetrazine derivatives of phenylalanine (Tet-v3.0).

Synthetic Procedure

1. General Synthetic Procedure of Boc-protected Tet-v3.0 Derivatives

The starting material Boc-protected 3-cyano phenylalanine (1.0 eqv.) was taken in a dried heavy walled reaction tube under argon atmosphere. Maintaining inert atmosphere inside the reaction vessel, catalyst Ni(OTf)₂ (0.5 eqv.) and corresponding nitrile derivatives (7 - 10 eqv.) were added. Then anhydrous hydrazine (50 eqv.) was added slowly to reaction mixture using a glass syringe and under stirring condition. Purged argon for another 10 minutes and immediately sealed the tube. The reaction vessel was immersed into the preheated oil bath at 50- 55 °C for 24 hours. After that, the reaction vessel was lifted from the oil bath and allowed to come at room temperature. The reaction mixture was poured into a beaker and added 10 eqv. 2M NaNO₂ aqueous solution and 5 mL water. The aqueous phase was washed with ethyl acetate (20 ml) to remove the homo coupled byproduct. Then the aqueous phase was acidified with 4M HCl (pH~2) under ice-cold condition with homogeneous mixing and extracted with ethyl acetate (3x 20 mL). The combined organic layer was washed with brine, dried with anhydrous Na₂SO₄ and concentrated under reduced pressure. Silica gel flash column chromatography (20-25% ethyl acetate in hexanes with 1% acetic acid) yielded desired tetrazine derivatives in the form of a pinkish red gummy material.

2. General Procedure of Boc-deprotection of Tet 3.0 Derivatives

In a dry RB, Boc-protected Tet-v3.0 derivatives were dissolved in 3 mL ethyl acetate and 1 mL HCl gas saturated 1,4 Dioxane was added to the solution under argon atmosphere. The reaction mixture was stirred at room temperature until the starting materials were consumed monitoring by TLC (normally 3 to 4 h). After completion the Boc-deprotection, remove the solvent under reduced pressure and re-dissolved in ethyl acetate (2x 10 ml) and similarly concentrated to remove excess HCl gas. Finally, added 5 mL pentane and dried which made the pink color solid material of chloride salt of **Tet-v3.0** derivatives in quantitative yield (~ 97-98%).

(S)-2-((*tert*-butoxycarbonyl)amino)-3-(3-(6-methyl-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid (**2a**): Following the general synthetic procedure **1**, starting from 0.30 g (1.03 mmol) of Boc-protected 3-cyano phenyl alanine **4** and 0.53 mL (10.3 mmol) of acetonitrile afforded 0.274 g (0.76 mmol) of the title compound **2a** (methyl derivative of **Tet -v3.0**) as a pink gummy material. Yield 74%. ¹H NMR (400MHz, CDCl₃) δ 8.41 (2H, t, *J* = 7.2 Hz), 7.51-7.44 (2H, m),

5.17 (1H, d, $J = 7.4$), 4.71 (1H, d, $J = 5.2$), 3.33 (1H, dd, $J = 13.6, 5.2$ Hz), 3.21 (1H, dd, $J = 13.2, 6.4$ Hz), 3.07 (3H, s), 1.39 (9H, s). ^{13}C NMR (175MHz, CDCl_3) δ 175.6, 167.3, 164.1, 155.5, 137.7, 133.8, 132.1, 129.5, 129.1, 126.7, 80.4, 54.4, 38.1, 28.4, 21.1.

Chloride salt of (S)-2-amino-3-(3-(6-methyl-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid

(1a): Following the general procedure **2**, 0.274 gm (0.76 mmol) of Boc-protected methyl derivative of Tet -v3.0 **2a** yielded 0.218 gm (0.74 mmol) of the title compound **1a**. Yield 97%. ^1H NMR (400MHz, CD_3OD) δ 8.52-8.49 (2H, m), 7.66-7.60 (2H, m), 4.38 (1H, dd, $J = 7.2, 6$ Hz), 3.46 (1H, dd, $J = 14.4, 5.6$ Hz), 3.35 (1H, dd, $J = 14.4, 7.2$ Hz), 3.05 (3H, s). ^{13}C NMR (175 MHz, CD_3OD) δ 171.1, 169.1, 165.2, 137.1, 134.7, 134.4, 131.2, 129.8, 128.4, 55.1, 37.3, 21.1. ESI-MS calculated for $\text{C}_{12}\text{H}_{14}\text{N}_5\text{O}_2$ ($[\text{M} + \text{H}]^+$) 260.1142, found 260.1133.

(S)-2-((tert-butoxycarbonyl)amino)-3-(3-(6-ethyl-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid

(2b): Following the general synthetic procedure **1**, starting from 0.30 g (1.03 mmol) of Boc-protected 3-cyano phenyl alanine **4** and 0.72 mL (10.3 mmol) of propionitrile afforded 0.294 g (0.79 mmol) of the title compound **2b** (ethyl derivative of Tet -v3.0) as a pink gummy material. Yield 72%. ^1H NMR (400MHz, CDCl_3) δ 8.45 (1H, d, $J = 7.6$ Hz), 8.41 (1H, s), 7.50 (1H, t, $J = 7.6$ Hz), 7.45 (1H, d, $J = 7.6$ Hz), 5.14 (1H, d, $J = 7.6$ Hz), 4.70 (1H, d, $J = 5.6$ Hz), 3.38 (2H, q, $J = 7.6$ Hz), 3.32 (1H, d, $J = 4.4$ Hz), 3.19 (1H, dd, $J = 12.8, 5.6$ Hz), 1.53 (3H, t, $J = 7.6$ Hz), 1.39 (9H, s). ^{13}C NMR (175MHz, CDCl_3) δ 176.1, 171.1, 164.3, 155.5, 137.6, 133.8, 132.3, 129.6, 129.1, 126.8, 80.5, 54.5, 38.1, 28.5, 28.4, 12.4.

Chloride salt of (S)-2-amino-3-(3-(6-ethyl-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid (1b):

Following the general procedure **2**, 0.270 gm (0.72 mmol) of Boc-protected ethyl derivative of Tet -v3.0 **2b** yielded 0.220 gm (0.71 mmol) of the title compound **1b**. Yield 98%. ^1H NMR (400MHz, CD_3OD) δ 8.52 (1H, s), 8.51 (1H, d, $J = 1.6$ Hz), 7.66-7.61 (2H, m), 4.38 (1H, d, $J = 6$ Hz), 3.47 (1H, dd, $J = 14.8, 5.6$ Hz), 3.37 (3H, q, $J = 7.6$ Hz), 1.53 (3H, t, $J = 7.6$ Hz). ^{13}C NMR (175 MHz, CD_3OD) δ 172.3, 165.4, 137.2, 134.7, 134.3, 131.2, 129.9, 128.3, 55.3, 37.3, 29.2, 12.4. ESI-MS calculated for $\text{C}_{13}\text{H}_{16}\text{N}_5\text{O}_2$ ($[\text{M} + \text{H}]^+$) 274.1299, found 274.1290.

(S)-2-((tert-butoxycarbonyl)amino)-3-(3-(6-isopropyl-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid (2c): Following the general synthetic procedure **1**, starting from 0.20 g (0.69 mmol) of Boc-protected 3-cyano phenyl alanine **4** and 0.62 mL (6.9 mmol) of isobutyronitrile afforded 0.152 g (0.39 mmol) of the title compound **2c** (isopropyl derivative of **Tet -v3.0**) as a pink gummy material. Yield 57%. ¹H NMR (400MHz, CDCl₃) δ 8.47 (1H, d, *J* = 7.6 Hz), 8.42 (1H, s), 7.52 (1H, t, *J* = 7.6 Hz), 7.45 (1H, d, *J* = 7.6 Hz), 5.08 (1H, d, *J* = 6.4 Hz), 4.69 (1H, d, *J* = 4 Hz), 3.68 (1H, sept, *J* = 6.8 Hz), 3.35 (1H, dd, *J* = 13.2, 4.4 Hz), 3.2 (1H, dd, *J* = 12.8, 5.6 Hz), 1.55 (6H, d, *J* = 7.2 Hz), 1.39 (9H, s). ¹³C NMR (175MHz, CDCl₃) δ 175.8, 173.8, 164.3, 155.5, 137.5, 133.7, 132.4, 129.7, 129.1, 126.7, 54.5, 38.1, 34.4, 28.4, 21.4.

Chloride salt of (S)-2-amino-3-(3-(6-isopropyl-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid (1c): Following the general procedure **2**, 0.150 gm (0.39 mmol) of Boc-protected isopropyl derivative of **Tet -v3.0 2c** yielded 0.122 gm (0.38 mmol) of the title compound **1c**. Yield 97.5%. ¹H NMR (400MHz, CD₃OD) δ 8.52 (1H, s), 8.51 (1H, d, *J* = 6.4 Hz), 7.66-7.62 (2H, m), 4.36 (1H, bs), 3.65 (1H, sept, *J* = 6.8 Hz), 3.46 (1H, dd, *J* = 14, 4.4 Hz), 3.4 (1H, dd, *J* = 12.8, 4 Hz), 1.54 (6H, d, *J* = 7.2 Hz). ¹³C NMR (175MHz, CD₃OD) δ 175.1, 165.4, 137.3, 134.7, 134.4, 131.2, 129.9, 128.3, 55.6, 37.4, 35.5, 21.5. ESI-MS calculated for C₁₄H₁₈N₅O₂ ([M + H]⁺) 288.1455, found 288.1451.

(S)-2-((tert-butoxycarbonyl)amino)-3-(3-(6-butyl-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid (2d): Following the general synthetic procedure **1**, starting from 0.40 g (1.37 mmol) of Boc-protected 3-cyano phenyl alanine **4** and 1.4 mL (13.3 mmol) of valeronitrile afforded 0.43 g (1.07 mmol) of the title compound **2d** (butyl derivative of **Tet -v3.0**) as a pink gummy material. Yield 78%. ¹H NMR (400MHz, CDCl₃) δ 8.46 (1H, d, *J* = 7.6 Hz), 8.42 (1H, s), 7.51 (1H, t, *J* = 7.6 Hz), 7.46 (1H, d, *J* = 7.6 Hz), 5.12 (1H, d, *J* = 6.8 Hz), 4.7 (1H, d, *J* = 4 Hz), 3.34 (3H, t, *J* = 7.6 Hz), 3.20 (1H, t, *J* = 6.4 Hz), 1.95 (2H, quin, *J* = 7.6 Hz), 1.47 (2H, sext, *J* = 7.6 Hz), 1.39 (9H, s), 0.98 (3H, t, *J* = 7.2 Hz). ¹³C NMR (175MHz, CDCl₃) δ 175.8, 170.4, 164.2, 155.5, 137.7, 133.7, 132.3, 129.7, 129.1, 126.8, 80.5, 54.5, 38.2, 34.6, 30.4, 28.4, 22.4, 13.8.

(S)-2-((tert-butoxycarbonyl)amino)-3-(3-(6-butyl-1,4-dihydro-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid (3d): Synthetic procedure is quite similar to the general procedure **1** except

oxidation step. So, after completion the reaction it was poured into a beaker. Then 10 mL water and 10 mL ethyl acetate were added to the beaker to disconnect the contact of air with reaction mixture. Then acidification was done with (4N) HCl (pH~2) under ice-cold and stirring condition. The aqueous phase was extracted with ethyl acetate (3 times). The combined organic layer was washed with brine, dried with anhydrous Na₂SO₄ and concentrated under reduced pressure. Silica gel flash column chromatography (40-45% ethyl acetate in hexanes with 1% acetic acid) yielded dihydro-tetrazine derivative **3d** in the form of a yellowish gummy material. During column purification 15-20% product was oxidized. We got 55-60% dihydro-tetrazine compound. ¹H NMR (400MHz, CDCl₃) δ 7.52 (2H, d, *J* = 7.2 Hz), 7.36 (1H, t, *J* = 7.6 Hz), 7.31 (1H, s), 5.34 (1H, d, *J* = 8 Hz), 4.68 (1H, q, *J* = 5.2 Hz), 3.27-3.13 (2H, dd, *J* = 13.6, 5.2 Hz), 2.23 (2H, t, *J* = 8 Hz), 1.55 (2H, quin, *J* = 7.6 Hz), 1.41 (9H, s), 0.98 (2H, sext, *J* = 7.2 Hz), 0.90 (3H, t, *J* = 7.2 Hz).

Chloride salt of (S)-2-amino-3-(3-(6-butyl-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid (1d):

Following the general procedure **2**, 0.400 gm (0.99 mmol) of Boc-protected butyl derivative of Tet -v3.0 **2d** yielded 0.326 gm (0.97 mmol) of the title compound **1d**. Yield 98%. ¹H NMR (400MHz, CD₃OD) δ 8.52 (1H, s), 8.51 (1H, d, *J* = 6 Hz), 7.66-7.61 (2H, m), 4.38 (1H, t, *J* = 6.4 Hz), 3.46 (1H, dd, *J* = 8.8, 5.6 Hz), 3.35 (3H, dd, *J* = 15.6, 8 Hz), 1.96 (2H, quin, *J* = 7.6 Hz), 1.47 (2H, sext, *J* = 7.2 Hz), 1.02 (3H, t, *J* = 7.8 Hz). ¹³C NMR (175MHz, CD₃OD) δ 171.7, 165.3, 137.2, 134.7, 134.3, 131.2, 129.9, 128.4, 55.2, 37.3, 35.4, 31.3, 23.4, 14.1. ESI-MS calculated for C₁₅H₂₀N₅O₂ ([M + H]⁺) 302.1612, found 302.1604.

(S)-2-((tert-butoxycarbonyl)amino)-3-(3-(6-phenyl-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid (2e):

Following the general synthetic procedure **1**, starting from 0.25g (0.86 mmol) of Boc-protected 3-cyano phenyl alanine **4** and 0.62 mL (6.1 mmol) of benzonitrile afforded 0.156 g (0.37 mmol) of the title compound **2e** (phenyl derivative of Tet -v3.0) as a pink gummy material. Yield 43%. ¹H NMR (700MHz, CDCl₃) δ 8.63 (2H, bs), 8.51-8.46 (2H, m), 7.6 (3H, bs), 7.52-7.47 (2H, m), 5.14 (1H, bs), 4.72 (1H, bs), 3.47-3.11 (2H, m), 1.42 (9H, s). ¹³C NMR (175MHz, CDCl₃) δ 174.9, 164.1, 163.9, 155.6, 137.6, 134.1, 132.9, 132.2, 131.9, 129.8, 129.5, 129.1, 128.2, 126.9, 80.7, 54.5, 38.1, 28.4.

Chloride salt of (S)-2-amino-3-(3-(6-phenyl-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid (1e):

Following the general procedure **2**, 0.150 gm (0.36 mmol) of Boc-protected phenyl derivative of Tet -v3.0 **2e** yielded 0.125 gm (0.35 mmol) of the title compound **1e**. Yield 97%.

^1H NMR (700MHz, $\text{d}_6\text{-DMSO}$) δ 8.55 (2H, d, $J = 8.4$ Hz), 8.49 (1H, s), 8.46 (1H, d, $J = 6.3$ Hz), 7.73-7.68 (3H, m), 7.67 (1H, t, $J = 7.7$ Hz), 7.64 (1H, d, $J = 7.7$ Hz), 4.32 (1H, bs), 3.32 (2H, dd, $J = 6.3, 3.5$ Hz). ^{13}C NMR (175MHz, $\text{d}_6\text{-DMSO}$) δ 170.3, 163.5, 163.4, 136.5, 134.1, 132.8, 132.2, 131.9, 129.9, 129.6, 128.8, 127.7, 126.7, 53.1, 35.7. ESI-MS calculated for $\text{C}_{17}\text{H}_{16}\text{N}_5\text{O}_2$ ($[\text{M} + \text{H}]^+$) 322.1299, found 322.1303.

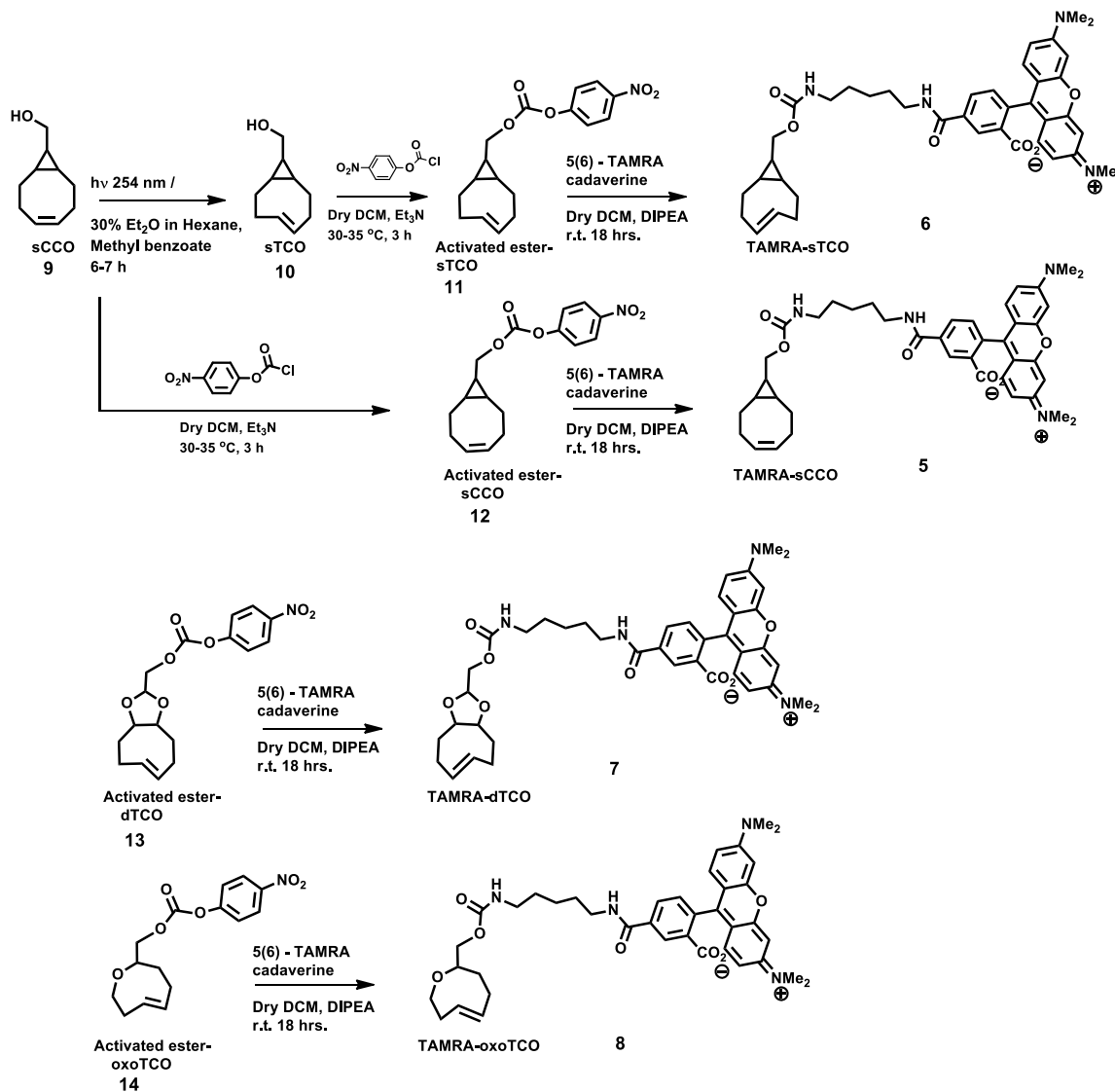


Fig 3.19. Synthesis of TAMRA linked strained alkenes (sCCO, sTCO, dTCO and oxoTCO).

(Z)-bicyclo[6.1.0]non-4-en-9-ylmethanol (sCCO, 9): Synthetic procedure,¹ ¹H NMR (700MHz, CD₃OD) δ 5.64-5.60 (2H, m), 3.38 (2H, d, $J = 7$ Hz), 2.30-2.26 (2H, m), 2.19-2.15 (2H, m), 2.09-2.04 (2H, m), 1.73-1.43 (2H, m), 0.79-0.74 (2H, m), 0.58-0.55 (1H, m).

(E)-bicyclo[6.1.0]non-4-en-9-ylmethanol (sTCO, 10): Synthetic procedure,² ¹H NMR (700MHz, CD₃OD) δ 5.89-5.84 (1H, m), 5.15-5.10 (1H, m), 3.44-3.39 (2H, m), 2.37 (1H, d, $J = 13.3$ Hz), 2.27 (1H, dt, $J = 12.6, 4.2$ Hz), 2.25-2.23 (1H, m), 2.19-2.15 (1H, m), 1.94-1.87 (2H, m), 0.92-0.87 (1H, m), 0.64-0.58 (1H, m), 0.50-0.46 (1H, m), 0.37-0.31 (2H, m).

(E)-bicyclo[6.1.0]non-4-en-9-ylmethyl (4-nitrophenyl) carbonate (11): In a dry round-bottom flask, sTCO **10** (0.3 gm, 1.97 mmol) was dissolved in anhydrous dichloromethane (DCM) under inert atmosphere. Subsequently, trimethylamine (Et₃N) (650 μ L, 4.9 mmol) and 4-Nitrophenyl chloroformate (0.43 gm, 2.16 mmol) were added to the solution and stirred at 30-35 °C for 2-3 hrs. After consumption of all starting material (monitored by TLC), added 15 mL DCM to the reaction mixture and washed with water. The aqueous layer was re-extracted twice with DCM. The organic layers were combined, dried with anhydrous Na₂SO₄, concentrated using rotary evaporator. Purification was done using silica gel flash column chromatography (5% ethyl acetate in hexane) yielded yellowish white solid material **11** (0.51 gm, 1.6 mmol). Yield 81%. ¹H NMR (400MHz, CDCl₃) δ 8.27 (2H, d, $J = 9.6$ Hz), 7.37 (2H, d, $J = 9.6$ Hz), 5.88-5.82 (1H, m), 5.18-5.14 (1H, m), 4.18 (2H, d, $J = 7.2$ Hz), 2.43-2.39 (1H, m), 2.35-2.22 (3H, m), 1.96-1.90 (2H, m), 0.94-0.83 (1H, m), 0.69-0.64 (1H, m), 0.62-0.49 (3H, m).

3. General synthetic procedure of TAMRA linked strained alkenes:

A dry 10 mL round-bottom flask was charged with Tetramethylrhodamine 5 - (and - 6) - carboxamide cadaverine (TAMRA) and activated ester of strained alkene (**11**, **12**, **13**, **14**) (2 eqv.) under N₂ atmosphere. Anhydrous dichloromethane (2 mL) and N,N-diisopropylethylamine (DIPEA) (3 eqv.) were added to the reaction mixture and allowed to stirrer at room temperature for 18 hours. After that, solvent was concentrated onto silica gel under reduced pressure and directly loaded on the silica gel column chromatography for purification. Using the solvent gradient 30-35% methanol in dichloromethane isolated desired molecules as a red solid material. Yield 46-57%.

TAMRA linked sCCO (5): Using general procedure **3**, 5 mg (9.7 μ mol) of 5(6) - TAMRA cadaverine yielded 3.4 mg (4.9 μ mol) of the title compound (**5**). Yield 51%. ¹H NMR (400MHz,

CD₃OD) δ 8.14 (1H, d, $J = 7.2$ Hz), 8.07 (1H, dd, $J = 8.0, 1.6$ Hz), 7.69 (1H, d, $J = 2.0$ Hz), 7.27 (1H, s), 7.25 (1H, s), 7.03 (1H, d, $J = 2.4$ Hz), 7.01 (1H, d, $J = 2.4$ Hz), 6.92 (2H, d, $J = 2.4$ Hz), 5.62-5.58 (2H, m), 3.82 (2H, d, $J = 7.6$ Hz), 3.38 (2H, t, $J = 7.2$ Hz), 3.28 (12H, s), 3.08 (2H, t, $J = 7.2$ Hz), 2.27-2.22 (2H, m), 2.16-2.02 (3H, m), 1.67-1.59 (2H, m), 1.56-1.49 (2H, m), 1.42-1.34 (5H, m), 0.83-0.78 (2H, m), 0.64-0.59 (1H, m).

TAMRA linked sTCO (6): Using general procedure **3**, 10 mg (0.019 mmol) of 5(6) - TAMRA cadaverine yielded 7.5 mg (0.010 mmol) of the title compound **(6)**. Yield 57%. ¹H NMR (400MHz, CD₃OD) δ 8.50 (1H, s), 8.05 (1H, t, $J = 8.4$ Hz), 7.35 (1H, d, $J = 7.2$ Hz), 7.25 (2H, dd, $J = 8.8, 4.0$ Hz), 7.01 (2H, d, $J = 9.2$ Hz), 6.91 (2H, s), 5.89-5.78 (1H, m), 5.16-5.04 (1H, m), 3.90 (1H, bs), 3.61 (1H, s), 3.49-3.42 (2H, m), 3.27 (12H, s), 3.16-3.05 (2H, m), 2.37-2.29 (1H, m), 2.26-2.16 (2H, m), 1.95 (2H, s), 1.92-1.87 (1H, m), 1.74-1.43 (5H, m), 1.35 (3H, d, $J = 6$ Hz), 0.93-0.82 (1H, m), 0.62-0.50 (1H, m), 0.46-0.36 (1H, m).

TAMRA linked dTCO (7):³ Using general procedure **3**, 5 mg (9.7 μ mol) of 5(6) - TAMRA cadaverine yielded 3.2 mg (4.4 μ mol) of the title compound **(7)**. Yield 46%. ¹H NMR (400MHz, CD₃OD) δ 8.15 (1H, d, $J = 8$ Hz), 8.07 (1H, dd, $J = 8.0, 1.6$ Hz), 7.69 (1H, d, $J = 1.6$ Hz), 7.27 (1H, s), 7.25 (1H, s), 7.01 (2H, dd, $J = 8.4, 2.0$ Hz), 6.92 (2H, d, $J = 2.4$ Hz), 5.62-5.53 (2H, m), 4.01-3.96 (2H, m), 3.93-3.88 (1H, m), 3.71 (1H, quin, $J = 6.8$ Hz), 3.81 (2H, t, $J = 6.8$ Hz), 3.28 (12H, s), 3.22 (1H, q, $J = 7.6$ Hz), 3.11-3.07 (2H, m), 2.38-2.33 (1H, m), 2.24-2.19 (1H, m), 2.15-2.10 (2H, m), 1.96 (2H, bs), 1.76-1.71 (1H, m), 1.66-1.59 (2H, m), 1.57-1.51 (2H, m), 1.42-1.38 (2H, m), 1.31-1.28 (1H, m).

TAMRA linked oxoTCO (8):⁴ Using general procedure **3**, 5 mg (9.7 μ mol) of 5(6) - TAMRA cadaverine yielded 3.1 mg (4.5 μ mol) of the title compound **(8)**. Yield 47%. ¹H NMR (400MHz, CD₃OD) δ 8.14 (1H, d, $J = 8$ Hz), 8.06 (1H, dd, $J = 8.0, 1.6$ Hz), 7.69 (1H, d, $J = 1.6$ Hz), 7.27 (1H, s), 7.25 (1H, s), 7.01 (2H, dd, $J = 9.2, 2.4$ Hz), 6.92 (2H, d, $J = 2.4$ Hz), 5.69-5.61 (1H, m), 5.41-5.34 (1H, m), 4.01-3.96 (1H, m), 3.89-3.81 (3H, m), 3.37 (2H, t, $J = 7.2$ Hz), 3.23 (12H, s), 3.23-3.19 (1H, m), 3.08 (2H, t, $J = 6.8$ Hz), 2.45-2.40 (1H, m), 2.30-2.24 (1H, m), 2.22-2.12 (1H, m), 1.82 (1H, dd, $J = 14, 4.4$ Hz), 1.66-1.59 (2H, m), 1.56-1.49 (2H, m), 1.42-1.38 (3H, m), 1.30-1.27 (1H, m).

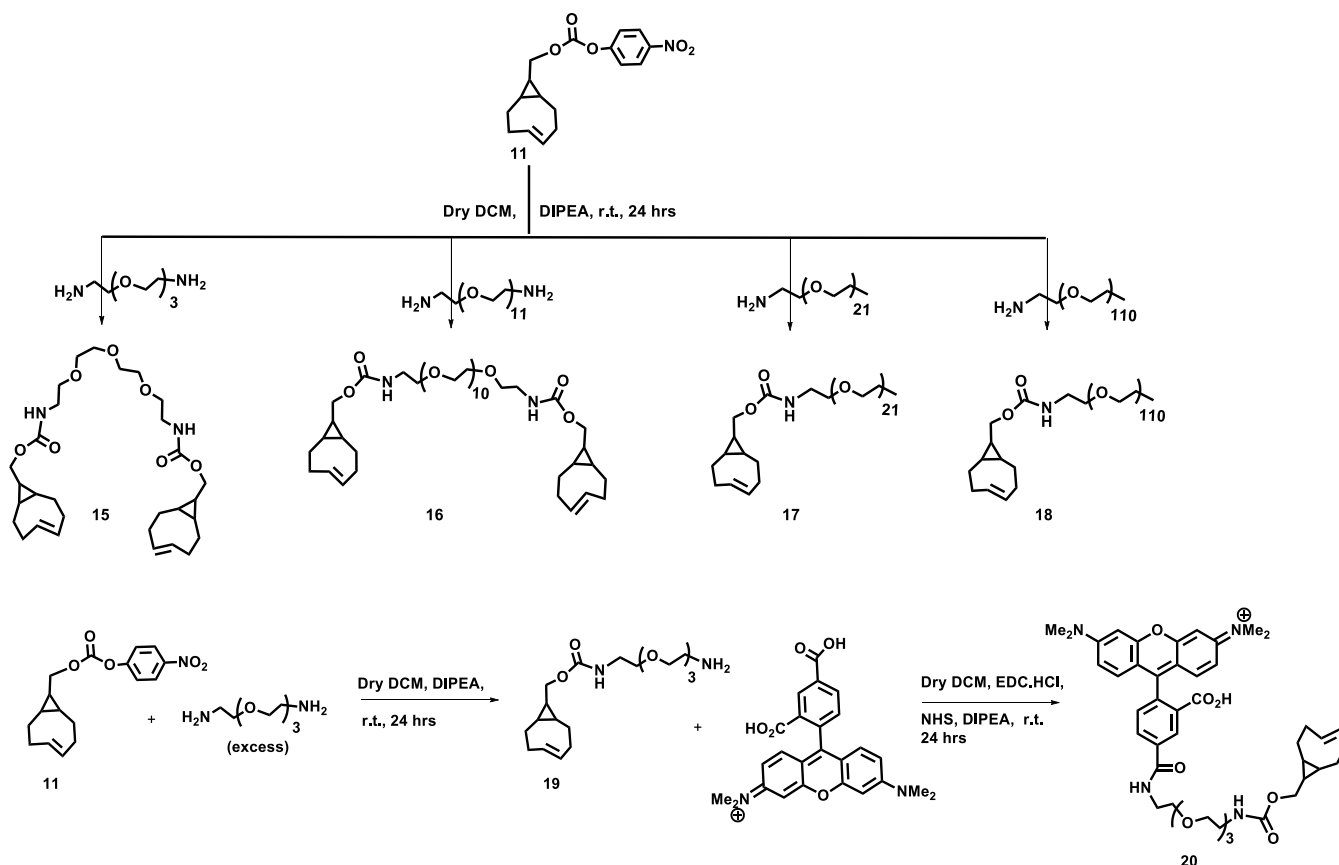


Fig 3.20. Synthesis of PEGylated sTCO, di-sTCO and TAMRA-sTCO.

PEG192 linked di-sTCO (15): In 3 mL anhydrous dichloromethane, Amino-PEG192-Amine (15 mg, 0.07 mmol), the activated ester of sTCO **11** (86 mg, 0.27 mmol) and followed by triethylamine (30 μ L, 0.212 mmol) were added under argon atmosphere. The reaction mixture was stirred at room temperature for 24 hours. After that, the solvent was concentrated onto silica gel under reduced pressure and purified our desired molecule **15** (25 mg, 0.045 mmol) by silica gel column chromatography (10-15% methanol in dichloromethane). Yield 64%. ^1H NMR (400 MHz, CD_3OD) δ 5.91-5.83 (2H, m), 5.18-5.10 (2H, m), 3.93 (4H, d, $J = 6.4$ Hz), 3.66-3.62 (8H, m), 3.54 (4H, t, $J = 5.6$ Hz), 3.29 (4H, q, $J = 5.6$ Hz), 2.36 (2H, dd, $J = 14, 2$ Hz), 2.30-2.17 (6H, m), 1.98-1.87 (4H, m), 0.94-0.85 (2H, m), 0.67-0.55 (4H, m), 0.49-0.44 (4H, m).

PEG550 linked di-sTCO (16): Similarly, using 40 mg of Amino-PEG550-Amine (0.072 mmol), 80 mg of the activated ester of sTCO **11** (0.254 mmol) and 30 μ L of triethylamine (0.218 mmol) produced 40 mg of the title molecule **16** (0.042 mmol). Yield 58%. ^1H NMR

(400MHz, CD₃OD) δ 5.90-5.82 (2H, m), 5.16-5.09 (2H, m), 3.91(4H, d, J = 6.0 Hz), 3.65-3.61 (40H, m), 3.52 (4H, t, J = 5.6 Hz), 3.26 (4H, t, J = 5.6 Hz), 2.36-2.33 (2H, m), 2.28-2.14 (6H, m), 1.94-1.87 (4H, m), 0.94-0.83 (2H, m), 0.65-0.56 (4H, m), 0.47-0.42 (4H, m).

mPEG1000 linked sTCO (17): Following the above procedure, using 115 mg (0.115 mmol) of mPEG1000-Amine, 43 mg (0.138 mmol) of the activated ester of sTCO **11** and 35 μ L triethylamine (0.23 mmol) made 98 mg (0.083 mmol) of the title molecule **17**. Yield 72%. Compound was purified by silica gel column chromatography (5% methanol in dichloromethane). ¹H NMR (400MHz, CD₃OD) δ 5.90-5.82 (1H, m), 5.17-5.09 (1H, m), 3.91(2H, d, J = 6.8 Hz), 3.63 (82H, bs), 3.55-3.50 (4H, m), 3.35 (3H, s), 3.25 (2H, q, J = 5.6 Hz), 2.36-2.36 (1H, d, J = 13.6 Hz), 2.28-2.15 (3H, m), 1.98-1.86 (2H, m), 0.94-0.84 (1H, m), 0.65-0.53 (2H, m), 0.48-0.42 (2H, m).

mPEG5000 linked sTCO (18): Following the above procedure, using 130 mg (0.026 mmol) of mPEG5000-Amine, 10 mg of the activated ester of sTCO (0.031 mmol) and 10 μ L triethylamine (0.052 mmol) produced 104 mg of the title molecule **18** (0.019 mmol). Yield 74%. Compound was purified by silica gel column chromatography (5% methanol in dichloromethane). ¹H NMR (400MHz, CD₃OD) δ 5.90-5.82 (1H, m), 5.17-5.09 (1H, m), 3.91(2H, d, J = 6.8 Hz), 3.82-3.79 (3H, m), 3.63 (411H, bs), 3.54-3.50 (4H, m), 3.45 (2H, t, J = 4.8 Hz), 3.35 (3H, s), 3.26 (2H, t, J = 5.6 Hz), 3.16 (2H, q, J = 7.6 Hz), 2.35 (1H, d, J = 15.2 Hz), 2.27-2.15 (3H, m), 1.97-1.88 (2H, m), 0.95-0.84 (1H, m), 0.66-0.54 (2H, m), 0.48-0.41 (2H, m).

sTCO linked PEGylated-Amine (19): Following the above procedure, 40 mg of the activated ester of sTCO **11** (0.126 mmol), 49 mg (0.252 mmol) of Amino-PEG192-Amine and 26 μ L of triethylamine (0.189 mmol) made 29 mg (0.078 mmol) of the title molecule **19**. Yield 62%. Compound was purified by silica gel column chromatography (30-35% methanol in dichloromethane). ¹H NMR (400MHz, CD₃OD) δ 5.91-5.83 (1H, m), 5.18-5.10 (1H, m), 3.94(2H, d, J = 6.4 Hz), 3.72 (2H, t, J = 5.2 Hz), 3.69 (4H, bs), 3.67-3.64 (4H, m), 3.54 (2H, t, J = 5.6 Hz), 3.30 (2H, t, J = 5.6 Hz), 3.13 (2H, t, J = 5.2 Hz), 2.35 (1H, dd, J = 14, 2.4 Hz), 2.29-2.19 (3H, m), 1.96-1.89 (2H, m), 0.96-0.85 (1H, m), 0.64-0.56 (2H, m), 0.50-0.43 (2H, m).

TAMRA linked PEGylated –sTCO (20): In dry DCM (2 mL), (5(6)-Carboxytetramethyl-rhodamine) (TAMRA Acid) (6mg, 0.014 mmol), 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC.HCl) (4mg, 0.02 mmol) and N-Hydroxysuccinimide (NHS) (2.5 mg, 0.020 mmol) were added under argon atmosphere and allowed to stirrer for 30 minutes under ice cold condition. After that, sTCO linked PEGylated-Amine **19** (5.6 mg, 0.016 mmol) and followed by N,N-Diisopropylethylamine (DIPEA) (10 μ l, 0.05 mmol) were added to the reaction mixture. After 15 minutes ice bath was removed and stirring was continued for another 24 hours at room temperature. After that, the solvent was concentrated onto silica gel under reduced pressure and purified the title compound **20** (4.8 mg, 6.13 μ mol) by silica gel column chromatography (20-25% methanol in dichloromethane). Yield 44%. ^1H NMR (700MHz, CD_3OD) δ 8.16 (1H, d, $J = 8.4$ Hz), 8.12 (1H, d, $J = 8.4$ Hz), 7.74 (1H, s), 7.27 (2H, dd, $J = 13.3, 9.1$ Hz), 7.08-7.04 (2H, m), 6.97 (2H, bs), 5.91-5.79 (1H, m), 5.17-5.09 (1H, m), 3.97-3.87(2H, m), 3.74 (4H, quin, $J = 7.0$ Hz), 3.68-3.63 (6H, m), 3.56-3.52 (2H, m), 3.31 (12H, s), 3.23 (4H, q, $J = 7.7$ Hz), 2.49-2.23 (2H, m), 2.28-2.20 (1H, m), 1.44 (2H, q, $J = 7.0$ Hz), 1.16-1.06 (1H, m), 0.94-0.84 (2H, m), 0.70-0.53 (2H, m), 0.47-0.36 (1H, m).

Selection of Aminoacyl-tRNA Synthetases Specific for a Representative Tetrazine Amino Acid and Bioorthogonal Labeling

The D3 library (Leu270, Tyr306, Leu309, Asn346, and Cys348) from the *Methanosarcina barkeri* (*Mb*) system was chosen for its previously demonstrated ability to incorporate large aromatic amino acids into proteins. The D3 library was encoded on a kanamycin (Kn) resistant plasmid (pBK, 3000 bp) under control of the constitutive *Escherichia coli* GlnRS promoter and terminator. The aminoacyl synthetase library consists of the codons at the aforementioned sites mutated to NNK codons corresponding to all 20 natural amino acids where N is A, C, G, or T and K is G or T. The library plasmid, pBK D3 lib, was moved between cells containing a positive selection plasmid (pRep pylT) and cells containing a negative selection plasmid (pYOBB2 pylT).

The positive selection plasmid, pRep pylT (10000 bp), encodes a mutant (*Mb*) pyrrolysyl-tRNA_{CUA}, an amber codon-disrupted chloramphenicol acetyltransferase, an amber codon-disrupted T7 RNA polymerase, a T7 promoter controlled GFP gene, and the tetracycline (Tcn) resistance marker. The negative selection plasmid, pYOBB2 pylT (7000 bp), encodes the mutant pyrrolysyl-tRNA_{CUA}, an amber codon-disrupted barnase under control of an arabinose promoter and rrnC terminator, and the ampicillin (Amp) resistance marker. pRep

pylT electrocompetent cells and pYOBB2 pylT electrocompetent cells were made from DH10B cells carrying the respective plasmids and stored in 100 μ L aliquots at -80 $^{\circ}$ C.

pRep pylT/pBK D3 library cells were prepared by transforming 1.6 μ g of the pBK D3 lib plasmid into freshly prepared pRep pylT competence cells. The transformation yielded greater than 1000-fold coverage of the genotypic pRep pylT library (3.36×10^7 library members). Transformed cells (10 mL saturated solution) were used to inoculate 300 mL 2xYT containing 50 μ g/mL Kn and 25 μ g/mL Tcn. The OD of the 2xYT was monitored until it reached 2.8 at which point 250 μ L aliquots of media were plated on eleven 15 cm LB-agar plates containing 50 μ g/mL Kn, 25 μ g/mL Tcn, and 40 μ g/mL chloramphenicol (Cm). The positive selection agar medium also contained 1 mM Tet-v3.0. After spreading, the surface of the plates was allowed to dry completely before incubation (37 $^{\circ}$ C, 18 h). To harvest the surviving library members from the plates, 5 mL of 2xYT was added to each plate. Colonies were scraped from the plate using a glass spreader. The resulting solution was incubated with shaking (30 min, 37 $^{\circ}$ C) to wash cells free of agar. The cells were then pelleted, and plasmid DNA was extracted using a Thermo Scientific miniprep kit. The smaller pBK D3 lib plasmid was separated from the larger pRep pylT plasmid by agarose gel electrophoresis (1% (w/v) agarose gel, 120 V, 40 min) and extracted from the gel using a Thermo Scientific gel extraction kit.

The purified pBK D3 library plasmid was then transformed into pYOBB2 pylT-containing DH10B electrocompetent cells. A 100 μ L sample of pYOBB2 pylT electrocompetent cells was transformed with 5.0 ng of purified pBK D3 lib DNA. Cells were rescued in 1 mL of SOC for 1 h (37 $^{\circ}$ C, 250 rpm) and the rescue solution was plated on three 15 cm LB plates containing 100 μ g/mL Amp, 25 μ g/mL Cm, and 0.2% (w/v) L-arabinose (250 μ L rescue solution per plate). Cells were scraped and the PBK D3 library DNA was isolated as described above for positive selections. At this point, the remaining library was retransformed into pREP pylT containing cells to start a second round of positive selections and subsequently, a second round of negative selections. After the second round of selections DNA from the first and second negative rounds was further evaluated.

In order to evaluate the success of the positive and negative selections based on variation in synthetase efficiency (as opposed to traditional survival/death results) the resulting pBK D3 library members were tested with the pALS plasmid. This plasmid contains the GFP reporter with a TAG codon at residue 150 as well as pyrrolysyl-tRNA_{CUA}. When a pBK plasmid with a functional synthetase is transformed with the pALS plasmid and the cells are

grown in the presence of the appropriate amino acid on autoinduction agar, sfGFP is expressed and the colonies are visibly green.

The pBK D3 library plasmid (70 ng) from the negative selection was used to transform 100 μ L of pALS-containing DH10B cells. The cells were rescued for 1 hour in 1 mL of SOC (37 °C, 250 rpm). Both 250 μ L and 50 μ L aliquots of cells from each library were plated on autoinducing agar plates with 25 μ g/mL Kn and 50 μ g/mL Tcn. The plates were further divided by the presence or absence of amino acid (1mM Tet-v3.0). Plates were grown at 37 °C for 24 hours and then grown on the bench top, at room temperature, for an additional 24 hrs. Autoinducing agar plates were prepared by combining the reagents in Table 2.1 with an autoclaved solution of 4.5 g of agar in 400 mL water. Sterile water was added to a final volume of 500 mL.

A total of 72 visually green colonies from the two 1 mM Tet-v3.0 plates and 24 visually white colonies from the two plates without Tet-v3.0 were used to inoculate a 96-well plate containing 0.5 mL per well non-inducing media (NIM) (Table 2.1) containing 50 μ g/mL Kn and 25 μ g/mL Tcn. After 24 hours of growth (37 °C, 250 rpm), 50 μ L of these non-inducing samples were used to inoculate two 96-well plates with 0.5 mL autoinducing media(AIM) (Table 2.1) containing 50 μ g/mL Kn, 25 μ g/mL Tcn. One 96-well plate was created with and one 96-well plate was created without 1 mM Tet-v3.0.

Fluorescence measurements of the cultures were collected 24, 48, and 72 hours after inoculation using a BIOTEK® Synergy 2 Microplate Reader. The emission from 528 nm (20 nm bandwidth) was summed with excitation at 485 nm (20 nm bandwidth). Samples were prepared by diluting suspended cells directly from culture 4-fold with sterile water. Fifteen colonies were selected for their high fluorescence with Tet-v3.0 present and low fluorescence in the absence of ncAAs. Selected hits were grown overnight (37 °C for 24 hours) in LB media containing 50 μ g/mL Kn and 25 μ g/mL Tcn. Grown media (800 μ L) was mixed with 80% (v/v) glycerol (200 μ L) and stored at -80 °C.

Selections were also performed using the Susan2 library: Asn311, Cys313, Val366, Trp382, and Gly 386. However, hits from the Susan2 library demonstrated poorer fidelity and were not characterized further.

Characterization of Efficiency and Fidelity

After the completion of the selections, efficiency and fidelity were measured in larger, more aerated cultures. Cell stocks were used to inoculate 5 mL of NIM containing 50 μ g/mL Kn and 25 μ g/mL Tcn and allowed to grow overnight (37 °C, 18 hours). Saturated NIM (50 μ L) was then used to inoculate AIM containing 50 μ g/mL Kn, 25 μ g/mL Tcn, and Tet-v3.0 (1

mM). Fluorescence was measured every 12 hours for 48 hours. The six unique synthetases are compared in (Fig 3.4). Fluorescence measurements were collected using a Turner Biosystems Picofluor fluorimeter diluting 100 μ L cell culture in 1.9 mL water.

Sequencing of Tet-v3.0 hits

The pBK plasmids were sequenced stocks were sequenced. Of the fifteen colonies sequenced, six unique sequences were identified, and the sequences can be found in Table 3.2. It is of note that the hits termed R2-3 and R2-84 differ only by the mutation Arg263Cys which is not predicted to directly contact the amino acid or play an essential role in catalysis.

Permissivity Screen of Select Synthetases

Cell stocks stored at -80 °C containing Tet-v3.0 selection hits in the pBK/pALS system, were used to inoculate 5 mL of NIM containing kanamycin (50 μ g/mL) and tetracycline (25 μ g/mL). Cells were grown for 16 hours at 37 °C shaking at 250 rpm. The grown NIM cultures were used to inoculate 5 mL cultures of AIM (50 μ L inoculate volume) containing kanamycin (50 μ g/mL) and tetracycline (25 μ g/mL). A separate culture was inoculated for every amino acid tested and 1 mM amino acid was introduced to those cultures. The amino acids tested include Tet-v3.0-methyl, Tet-v3.0-isopropyl, Tet-v3.0-butyl, and Tet-v3.0-phenyl. All amino acids were dissolved in DMF prior to the addition to media. Upon addition to media Tet-v3.0-phenyl had significant precipitation. This alters the effective concentration of amino acid in the media but were not able to solubilize Tet-v3.0-phenyl to a significant degree via other methods. Cultures were grown for 30-36 hours at 37 °C and 250 rpm. Fluorescence was assessed every 12 hours by removing 100 μ L of media and diluting it to 2 mL total volume. Fluorescence was measured using a Turner Biosystems Picofluor fluorimeter.

Expression and purification of GFP-TAG150-Tet-v3.0

A cell stock of DH10B cells cotransformed with the R2-84 pBK plasmid, and pALS plasmid was used to inoculate a 5 mL culture of NIM containing kanamycin (50 μ g/mL) and tetracycline (25 μ g/mL), which was then grown 16 hours at 37 °C shaking at 250 rpm. A 50 mL AIM culture containing kanamycin (50 μ g/mL) and tetracycline (25 μ g/mL) was inoculated with 0.5 mL of the grown NIM. The AIM was supplemented with 1 mM Tet-v3.0 that was pre-dissolved in DMF to a concentration of 100 mM. The AIM culture was allowed to grow shaking at 250 rpm for 48 hours at 37 °C. Additionally, cultures without amino acid and containing only the WT pALS plasmid were grown simultaneously. All cells were harvested by centrifugation 5000 rcf for 5 min. Supernatant was decanted and cell pellets were stored at -80 °C. To purify, cells were resuspended in TALON wash buffer (NaCl 300 mM, NaH₂PO₄ 50

mM, pH 7.0). Cells were lysed using a Microfluidics M-110P microfluidizer (18,000 psi) and the lysate was collected in TALON wash buffer. The lysate was clarified by centrifugation (21036 rcf, 1 hour) and the supernatant was decanted and stored. The pellet was discarded. To the cell lysate, 100 μ L bed volume TALON resin was added. Lysate was incubated with the resin for 1-2 hours gently rocking at 4° C. Resin and lysate were applied to a column and flow through was discarded. Resin was washed with 3 x 10 mL TALON wash buffer. Protein was eluted with 4 x 250 μ L TALON elution buffer (TALON wash buffer containing imidazole 250 mM). Protein concentration was measured using a Bradford assay. Protein purity was assessed using SDS-PAGE.

Mass spectra of GFP-Tet-v3.0 from different synthetases

GFP-TAG150-Tet-v3.0 was expressed and purified as described above from each of the R2-40, R2-74, and R2-84 synthetases. Protein was diluted to 10 μ M and desalted on C₄ zip tips and analyzed using an FT LTQ mass spectrometer at the Oregon State University mass spectrometry facility. Of the six synthetases identified, three (R2-40, R2-74, and R2-84) produced GFP-TAG150-Tet-v3.0 that appears as a single peak of the expected molecular weight.

Measuring the impact of amino acid concentration on GFP production

Auto-induction media (3 mL, tetracycline 25 ng/ μ L, kanamycin 50 ng/ μ L) was inoculated using 30 μ L of non-induction media cultures of DH10B cells cotransformed with the pALS plasmid and a pBK plasmid corresponding to the synthetases of R2-84 and R2-74. Autoinduction media contained various amounts of the amino acid of interest ranging from 0-1 mM. Cultures were grown at 37 °C and shaking at 250 rpm. OD₆₀₀ and fluorescence measurements were taken every 12 hours and the fluorescence was normalized to the OD.

Measuring Tet-v3.0 kinetics

Fluorescence of 1.9 nmol of purified GFP-Tet-v3.0 diluted in 3 mL of PBS was measured (488 nm excitation, 509 nm emission, 5 points/second) for 60 seconds prior to the addition of various quantities (2 nmol-200 nmol) of sTCO. Fluorescence was measured until no fluorescence increase was observed. Curves were fit using the curve-fitting program Igor to determine kinetic constants.

Cloning into pDule1/2

The primers 5'-GAGTTTACGCTTTGAGGAATCCCCATGGATAAAAAACCGCTGGATG-3' (Forward primer) and 5'- CCTCTTCTGAGATGAGTTTTTGTCTTACAGGTTTCGTGCTAATGC-3' (Reverse primer) were used to amplify the synthetase gene from the pBK plasmid. The

amplified fragment was gel purified and cloned into the pDule1/pDule2 plasmid using a SLICE reaction.

Cloning into pACBac

A G-block was ordered of the mammalian codon optimized pylRS synthetases R2-74 and R2-84. The primers 5'-CTTCCTGGAAATCAAGAGCCCCATCCT-3' (Forward primer) and 5'-GTTCCAGGTGCGCCGTGCATGATGT-3' (Reverse primer) were used to amplify the synthetase gene from the G-block. The PCR products as well as a pUC plasmid backbone were digested with the restriction enzymes EcoR1 and Xho1. The desired fragments were gel purified and a ligation reaction was performed between the pUC backbone and the synthetase genes. The ligated product was sequenced to confirm identity.

The pUC plasmids containing both the R2-74 and R2-84 synthetases, as well as pACBac plasmid with no insert were digested using the restriction enzymes Nhe1 and EcoR1. The fragments containing the synthetase gene from the pUC plasmids and the fragments containing the pACBac backbone were gel purified. A ligation reaction was performed between individual synthetase genes and the pACBac backbone. Identity of ligation products was confirmed by sequencing.

Mobility Shift Assay

Purified GFP was diluted to 50 μ M in PBS. The protein was reacted with excess sTCO-PEG₅₀₀₀ (250 μ M) for 30 minutes in PBS. Reactions were not quenched. Protein was denatured through the addition of Laemmli buffer and heating at 95 °C for 15 minutes. Samples were then analyzed using SDS-PAGE.

Sequences of Genes Used

WT GFP:

Protein:

MVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTCLKFICTTGKLPVP
WPTLVTTLTLYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEV
KFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNHNVYITADKQKNGIKANFKIRHN
VEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMLLEFVTA
AGITHGMDELYKGSHHHHHH

DNA:

ATGGTTAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGAAC
TGGATGGTGTATGTGAATGGCCATAAATTTAGCGTTCGTGGCGAAGGCCGAAGGTG
ATGCGACCAACGGTAAACTGACCCTGAAATTTATTTGCACCACCGGTAAACTGCC
GGTCCGTGGCCGACCCTGGTGACCACCCTGACCTATGGCGTTCAGTGCTTTAGC

CGCTATCCGGATCATATGAAACGCCATGATTTCTTTAAAAGCGCGATGCCGGAAG
 GCTATGTGCAGGAACGTACCATTAGCTTCAAAGATGATGGCACCTATAAAAACCC
 GTGCGGAAGTTAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAAG
 GTATTGATTTTAAAGAAGATGGCAACATTCTGGGTCATAAACTGGAATATAATTT
 CAACAGCCATAATGTGTATATTACCGCCGATAAACAGAAAAATGGCATCAAAGC
 GAACTTTAAAATCCGTCACAACGTGGAAGATGGTAGCGTGCAGCTGGCGGATCA
 TTATCAGCAGAATACCCCGATTGGTGTATGGCCCGGTGCTGCTGCCGGATAATCAT
 TATCTGAGCACCCAGAGCGTTCTGAGCAAAGATCCGAATGAAAAACGTGATCAT
 ATGGTGCTGCTGGAATTTGTTACCGCCGCGGGCATTACCCACGGTATGGATGAAC
 TGTATAAAGGCAGCCACCATCATCATCACCAT

R2-84 PylRS:

Protein:

MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTA
 RAFRHHKYRKTCKRCRVSDDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSR
 APKPLENSVSAKASTNTRSVPSPAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDK
 ISLNMAKPFRELEPELVTRRKNDFQRLYTNDREDYLGKLERDITKFFVDRGFLEIKSPI
 LIPAEYVERMGINNDTELSKQIFRVDKNLCLCPMLAPTGYNYLRKLDRLPGPIKIFEV
 GPCYRKESDQKEHLEEFMVGFAQMGSGCTRENLEALIKEFLDYLEIDFEIVGDSCM
 VYGDITLDIMHGDLELSSAVVGPVSLDREWIDKWPWIGAGFGLERLLKVMHGFKNIK
 RASRSESYNGISTNL

DNA:

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGC
 CGTACCGGCACCCTGCATAAAATCAAACATCATGAAGTGAGCCGCAGCAAATC
 TATATTGAAATGGCGTGCGGCGATCATCTGGTGGTGAACAACAGCCGTAGCTGCC
 GTACCGCGCGTGCCTTTCGTCATCATAAATACCGCAAACCTGCAAACGTTGCCG
 TGTGAGCGATGAAGATATCAACAACCTTTCTGACCCGTAGCACCGAAAGCAAAAA
 CAGCGTGAAAGTGCGTGTGGTGTGAGCGCGCCGAAAGTGAAAAAAGCGATGCCGA
 AAAGCGTGAGCCGTGCGCCGAAACCGCTGGAAAATAGCGTGAGCGCGAAAGCG
 AGCACCAACACCAGCCGTAGCGTTCCGAGCCCGGCGAAAAGCACCCCGAACAGC
 AGCGTTCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCAGCTGGATCGT
 GTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCG
 TTTCGTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCC
 TGTATACCAACGATCGTGAAGATTATCTGGGCAAACCTGGAACGTGATATCACCA
 AATTTTTTGTGGATCGCGGCTTTCTGGAAATTAAGCCCGATTCTGATTCCGGC

GGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGAGCAAACAAAT
 TTTCCGTGTGGATAAAAACCTGTGCCTGTGTCCGATGCTGGCCCCGACCGGTTAT
 AACTATTTGCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAG
 TGGGCCCGTGCTATCGCAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCA
 CCATGGTTGGTTTTGCTCAAATGGGCAGCGGCTGCACCCGTGAAAACCTGGAAGC
 GCTGATCAAAGAATTCCTGGATTATCTGGAAATCGACTTCGAAATTGTGGGCGAT
 AGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGA
 GCAGCGCGGTGGTGGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAAC
 CGTGGATTGGCGCGGGTTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTT
 CAAAACATTAAACGTGCGAGCCGTAGCGAAAGCTACTATAACGGCATTAGCAC
 GAACCTGTAA

Toxicity Screen of Tetrazines

HEK293T cells plated in a 96-well plate at approximately 40% confluency. Cells were incubated for 48 h with tet3.0 butyl or 1% DMSO, and the cell viability was measured using CellTiter Glo assay kit (Promega) according to the manufacturer's instruction. Briefly, 25 μ l of CellTiter Glo reagent was added to each well and incubated for 10 min at RT. The signal was measured for 1 sec using TR717 microplate luminometer (Berthold, Germany) and WinGlow software version 1.25 (Berthold Technologies). The data was normalized to vehicle control and fitted to a curve using non-linear regression method using GraphPad Prism 5. $n = 3 \pm$ SEM

Transfection of HEK cells

HEK293T cells were plated in a 24-well plate at about 40% confluency so that they reach 70% ~ 90% confluency at the time of transfection. Cells were transfected using Lipofectamine 2000 (Thermo Fisher) using the manufacturer's protocol with minor modification. Briefly, 600 ng of plasmid DNA was diluted in 25 μ l of serum free DMEM and 1.8 μ l of Lipofectamine 2000 reagent was diluted in 25 μ l of serum free DMEM. They were combined and incubated for 10 min at RT before adding to cells. Tet3.0 amino acid was added to the cells immediately and incubated for 24 h ~ 48 h.

Reaction with TAMRA-strained alkenes

To assess the background staining untransfected HEK293T cells were incubated for 30 min with TAMRA acid, TAMRA-sCCO, TAMRA-dTCO, TAMRA-oxoTCO, or TAMRA-sTCO. Cells were washed once with PBS and dissociated into single cells by 0.05% trypsin/0.53 mM EDTA. Cells were washed twice with PBS and

subjected to flow cytometry as described below to analyze the level of TAMRA staining.

Flow Cytometry Assessment (Include 2D)

HEK293T cells were transfected as described above with R2 84 RS pAcBac1 and sfGFP-TAG150 pAcBac1 for 24 h. GFP expression and fluorescence was confirmed by fluorescence microscope. Adherent cells were washed three times with ncAA-free DMEM supplemented with 10% FBS and penicillin/streptomycin. Cells were incubated for 30 min with TAMRA acid, TAMRA-sCCO, TAMRA-dTCO, TAMRA-oxoTCO, or TAMRA-sTCO. Cells were washed once with PBS and dissociated into single cells by trypsin/EDTA. To analyze the expression of GFP protein and its labeling with TAMRA-strained alkenes, cells were analyzed by flow cytometry using CytoFLEX flow cytometer and CytExpert software version 2.2 (Beckman Coulter). GFP and TAMRA signal was collected on FL1-H FITC and FL10-H PE channel, respectively. Doublets were identified by plotting FSC-H versus FSC-A, and excluded from the subsequent analysis. In the scatter plot of GFP versus TAMRA, the basal level of GFP or TAMRA signal was determined from untransfected cells or no ncAA control cells. Any cells that have higher level of GFP than negative control were defined to be a GFP expressing cells. Background TAMRA staining level was determined from GFP-negative cell population. Any GFP expressing cells that have higher TAMRA signal than the background TAMRA signal were regarded as TAMRA-labeled cell population. On the other hand, any GFP expressing cells that have same TAMRA signal as the background TAMRA signal were regarded as non-labeled cell population. The relative ratio of the two-cell population and mean fluorescence intensity were calculated by using FlowJo or CytExpert software.

***In cellulo* Reactions (Dimerization/sTCO-PEG1000)**

HEK293T cells were transfected as described in the flow cytometry assessment section. To evaluate the feasibility of protein dimerization, transfected cells were washed three times with DMEM supplemented with 10% FBS and penicillin/streptomycin followed by 30 min incubation with di-sTCO-PEG100 or di-sTCO-PEG500. The reactive dimerization agents were quenched by reacting excessive

amount of tet2.0 methyl for 10 min at 37C. Cells were washed with PBS and lysed in beta-mercaptoethanol-free Laemmli buffer. Without heat denaturation, the lysates were analyzed by SDS-PAGE on 12% gel. The GFP fluorescence image was captured using ChemiDoc imaging system (Bio-Rad) using UV transillumination and the standard emission filter of 580/120 nm. Equal amount of protein loading was confirmed by staining the gel with Coomassie Brilliant Blue R.

To further evaluate the *in cellulo* labeling of GFP-Tet-v3.0 protein in HEK293T cells, transfected HEK293T cells were washed three times with DMEM, 10% FBS, pen/strep, and incubated for 30 min with sTCO-PEG1000 or sTCO-PEG5000. In gel fluorescence was performed to visualize the mobility shift of GFP-Tet-v3.0 protein that reacted with the agent as described above. (Neither agent worked)

Cell Lysis and Fluorescence Gels

HEK293T cells were washed with PBS and lysed in beta-mercaptoethanol-free Laemmli buffer. Without thermal denaturation, the lysates were separated by 12% SDS-PAGE. The fluorescence images were captured using ChemiDoc imaging system (Bio-Rad). UV transillumination and the emission filter of 580/120 nm were used to visualize GFP, and green epi illumination and the emission filter of 605/50 nm were used to detect TAMRA labeled protein band.

Densitometry Analysis

The images were analyzed using ImageJ software 1.52a (NIH) to detect and quantify the band. Briefly, the lane was manually defined using rectangle tool and the densitogram was generated using the same tool. The specific protein peak above the base line was manually set using a straight-line tool. The band intensity was calculated by the software. The signal was normalized to the no labeling control.

Chapter 4

Side Reactions of Tetrazines

Introduction

The reactions of tetrazines with sTCO molecules have been demonstrated to have near ideal reactions in that: 1) The rate constant of the reactions is extremely fast to the degree that the reaction can reach 99% completion in under a minute with sub-stoichiometric concentrations. 2) The reaction reaches completion such that no unreacted protein is observed via mass spectrometry. 3) The functional groups are non-toxic and stable *in vivo* during the timescale of reaction. However, instances of incomplete reaction have been observed. Mobility shift assays show a portion of protein that remains unreacted (Fig 3.7). Stability of the tetrazine and sTCO functional groups is not absolute. Isomerization of sTCO¹¹² as well as reduction of tetrazines¹³² have previously been observed. These side reactions must be characterized and controlled.

While isomerization of the functional group used as a label can be minimized through short exposure times *in vivo*, the incorporated functional group must be stable *in vivo*. The incorporation of tetrazines in proteins required further characterization of tetrazine side reactions. The most prominent side reaction of tetrazines is the reduction of tetrazines to form dihydro-tetrazines. This reaction takes place when a tetrazine is exposed to reducing agents (commonly thiols). The tetrazines may be reoxidized with reagents such as sodium nitrite¹¹⁵ or by exposure to atmospheric oxygen. Given that cellular environments are reducing in nature, this reaction may be relevant for *in vivo* labeling reactions.

Results/Discussion

Previously change in GFP fluorescence has been used to characterize the rate of the labeling reaction of an encoded Tet-v2.0 amino acid at site 150 with strained alkenes. The presence of an oxidized tetrazine on the surface of GFP results in fluorescence quenching of the GFP molecule. The GFP fluorescence can be restored through reaction of the GFP with sTCO or other strained alkenes (Fig 4.1A). Using this method, the reactivity of Tet-v2.0 has been assessed *in cellulo*. However, separate *in cellulo* preparations of proteins show inconsistencies in the degree of fluorescence return. This can result in difficulty interpreting results due to the incomplete reaction.

Yet, upon purification, expressed GFP-Tet-v2.0 always shows complete labeling via mass spectrometry when reacted with strained alkenes.

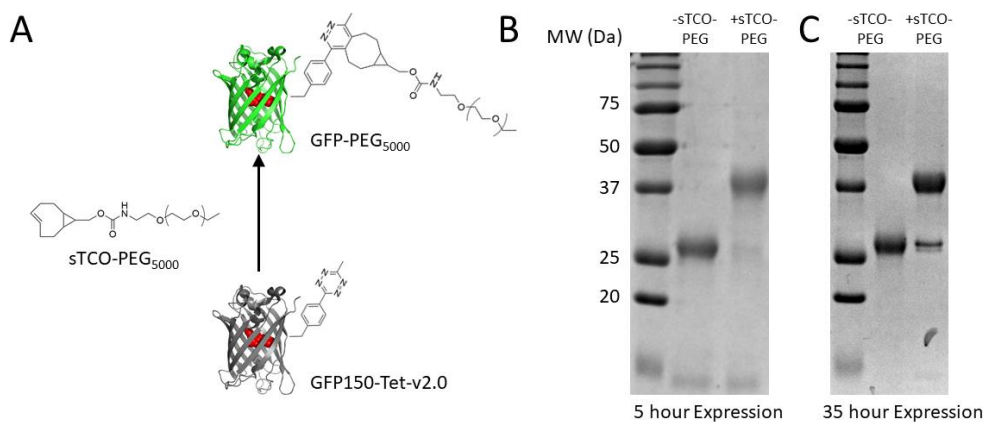


Fig 4.1. Time Dependent Mobility Shift Assay of GFP150-Tet-v2.0.

A) Reaction of GFP150-Tet-v2.0 with sTCO-PEG₅₀₀₀ results in a fluorescence increase and a protein with greater molecular weight. B) Mobility shift assay of reactions of GFP150-Tet-v2.0 expressions varying by the expression time. The 35-hour expression results in a larger unshifted band.

Similar to Tet-v2.0, Tet-v3.0-methyl has been shown to demonstrate many properties of ideal reactions (Chapter 3). Upon reaction with sTCO in *E. coli*, Tet-v3.0-methyl shows an increase in cellular fluorescence (Fig 4.2). Traces of this fluorescence increase vary from the fluorescence increase observed for the reaction of GFP150-Tet-v3.0-methyl *in vitro*. When the reaction curves corresponding to the reaction between sTCO and GFP150-Tet-v3.0-methyl *in cellulo* are fit to exponential equations from the expected model, patterned residuals occur (Fig 4.3) indicating that the reaction does not follow the expected pseudo 1st order reaction kinetics. In contrast, no pattern is observed in the residuals of reactions of Tet-v2.0 (Fig 4.3). These data suggest a more complex phenomenon is occurring in the reaction between tetrazines and sTCO in *E. coli*.

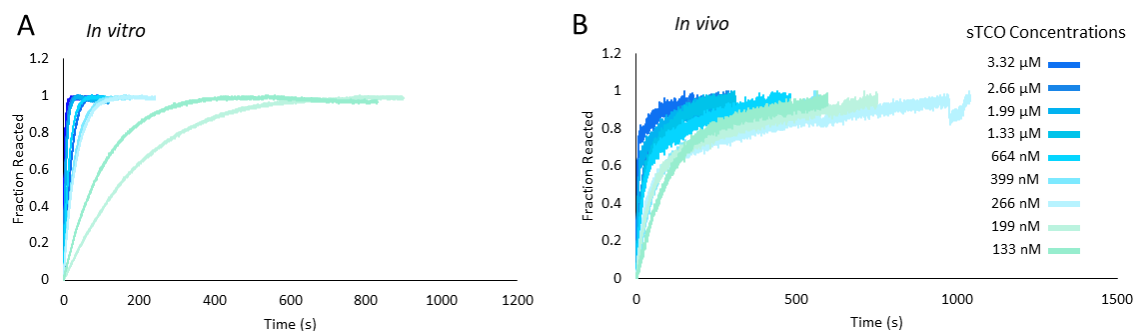


Fig 4.2. Tet-v3.0-methyl Kinetic Comparison.

A) Fluorescent traces of the pseudo 1st order reaction between GFP150-Tet-v3.0-methyl and sTCO *in vitro* at varying sTCO concentrations. The reactions occur as expected and fit well to exponential equations. B) Fluorescent traces of the reaction between GFP150-Tet-v3.0-methyl and sTCO *in cellulo* at varying sTCO concentrations. In addition to only partially reacting, the concentration dependence of the reaction is relatively weak and does not fit well to exponential equations.

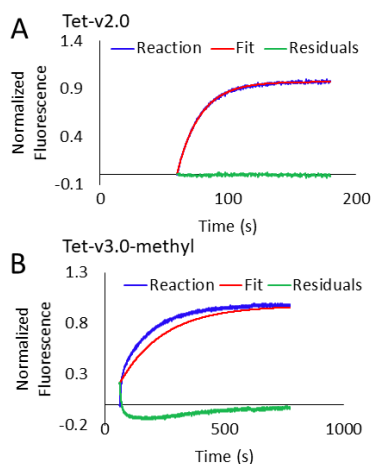


Fig 4.3. Fitting of *In cellulo* Reactions to 1st Order Equations.

A) Curve fit of the reaction between GFP150-Tet-v2.0 and excess sTCO to an exponential equation. Residuals are small and show no patterns demonstrating that this reaction fits well to a 1st order equation. B) Curve fit of the reaction between GFP150-Tet-v3.0-methyl and excess sTCO to an exponential equation. Residuals show a clear pattern indicating that model of fit is not appropriate. This suggests that this reaction is not a pseudo 1st order reaction.

While using mass spectrometry, GFP150-Tet-v2.0 protein was shown to undergo complete labeling when reacted with sTCO. When monitored using mobility shift assays, an incomplete reaction is observed (Fig 4.1). This incomplete reaction varied from preparation to preparation. Some protein preparations resulted in near 100% labeling as measured by complete upward shift of the GFP-Tet-v2.0 band to a higher molecular weight consistent with the addition of sTCO-PEG₅₀₀₀ (Fig 4.1B).

Others showed 80-90% completion when reacted with a large excess of sTCO-PEG₅₀₀₀ (Fig 4.1C). The unreacted protein could be a result of a low fidelity aaRS resulting in misincorporation of natural amino acids at the amber codon site. However, mass spectrometry analysis of the purified GFP150-Tet-v2.0 has revealed no natural amino acid incorporation as would be indicated by proteins with lower molecular weights. Further, expression of amber codon interrupted protein in the absence of ncAA results in the production of little to no full-length protein, as determined by purification using a C-terminal His Tag.

Together these results suggest that the tetrazine functionality on proteins was becoming unreactive either through reduction to the corresponding dihydro-tetrazine or through reaction with other species, making it unreactive to strained alkenes such as sTCO. Given the prevalence of oxidation and reduction reactions of tetrazines in literature¹³², these reactions are an obvious starting point for investigating potential causes of limited reactivity. Using cyclic voltammetry, the ability of the tetrazine amino acids to undergo oxidation and reduction reactions was assessed. It was found that Tet-v2.0 undergoes irreversible oxidation and reduction events at potentials of -380 mV and 45 mV (Table 4.1). In comparison, Tet-v3.0-methyl has similar potentials, though it is slightly more susceptible to reduction and less susceptible to oxidation than Tet-v2.0. These values appear to be relevant *in vivo* as common cellular redox pairs can have potentials on the range of -80 to -400 mV¹³³. Determination of the cellular redox state of tetrazines is not possible with these values as the cellular environment does not exist at a fixed redox potential because individual cellular redox pairs are not in equilibria with each other¹³³. However, this suggests that some amount of GFP150-Tet-v3.0 and even GFP150-Tet-v2.0 may exist as dihydro-tetrazines *in vivo*.

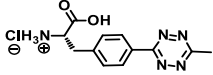
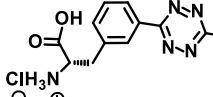
Name of Compounds	Structure	Ag/AgCl, 0.13(M) NaCl	
		10 % dist. DMF in PBS(pH~7.4)	
		$E_{c\ red}^n$ (mV)	$E_{a\ oxd}^n$ (mV)
Tet-v2.0 Me		-380/-380	+45/+40
Tet-v3.0 Me		-370	+15

Table 4.1. Reduction Potentials of Tet-v2.0 and Tet-v3.0.

Efforts to address the reduction reactions of tetrazines by the Joseph Fox¹³² have shown that photooxidation of dihydro-tetrazines is possible. These reactions use photoreactive dyes such as methylene blue, Rose Bengal, or fluorescein in combination with visible light of varying wavelengths to form oxidized tetrazines. To evaluate the ability to oxidize tetrazines on proteins, the reduced state of the proteins must be characterized. GFP150-Tet-v2.0 was treated with tris(2-carboxyethyl)phosphine (TCEP). The reaction with excess TCEP generated GFP-dihydro-Tet-v2.0. An increase in the fluorescence of GFP was observed upon treatment with TCEP consistent to the fluorescence increase observed upon reaction of GFP150-Tet-v2.0 with sTCO (Fig 4.4). This corresponds to the loss of GFP quenching due to tetrazine proximity. Because similar levels of fluorescence were observed for reactions of GFP-Tet-v2.0 with TCEP and sTCO, it can be inferred that the TCEP is not detrimental to the GFP fluorophore. Attempts to oxidize this protein using photooxidation were unsuccessful. This may be due to the continued presence of TCEP reducing any oxidized GFP150-dihydro-Tet-v2.0 prior to the ability to measure fluorescence. This shows that while GFP-Tet-v2.0 is largely oxidized when purified, it can be readily reduced. In cells, the oxidation state of Tet-v2.0 is less apparent and more difficult to manipulate as cellular treatment with

TCEP will not necessarily have the same impact on the intracellular redox environment as it does *in vitro*.

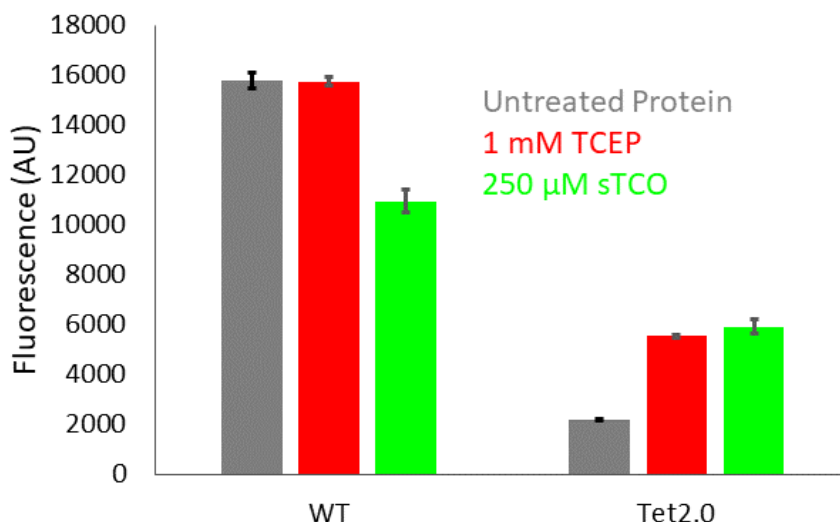


Fig 4.4. Reaction of Purified GFP150-Tet-v2.0 with TCEP.

GFP150-Tet-v2.0 reacted with TCEP results in an increase in fluorescence similar to the reaction of sTCO as measured via increase in fluorescence.

To determine whether tetrazines becomes reduced *in cellulo*, *E. coli* cells expressing GFP150-Tet-v2.0 and GFP150-Tet-v3.0-methyl were treated with methylene blue and red light (Fig 4.5). A decrease in fluorescence was observed consistent with the oxidation of dihydro-tetrazines. When light is removed, and the same cells are treated with sTCO an increase in fluorescence was observed consistent with the reaction of oxidized tetrazines with sTCO. When cells were treated with methylene blue or red light alone, no fluorescence decrease corresponding to oxidation was observed. Upon addition of sTCO, photooxidized cells show a fluorescence equivalent to unoxidized cells treated with excess sTCO. Together this data indicates that in cells expressing Tet-v2.0, a mixture of oxidized and reduced Tet-v2.0 is present. In contrast, Tet-v3.0-methyl is predominantly in the reduced state. This may be due to the observed differences in reduction potential of the free amino acids.

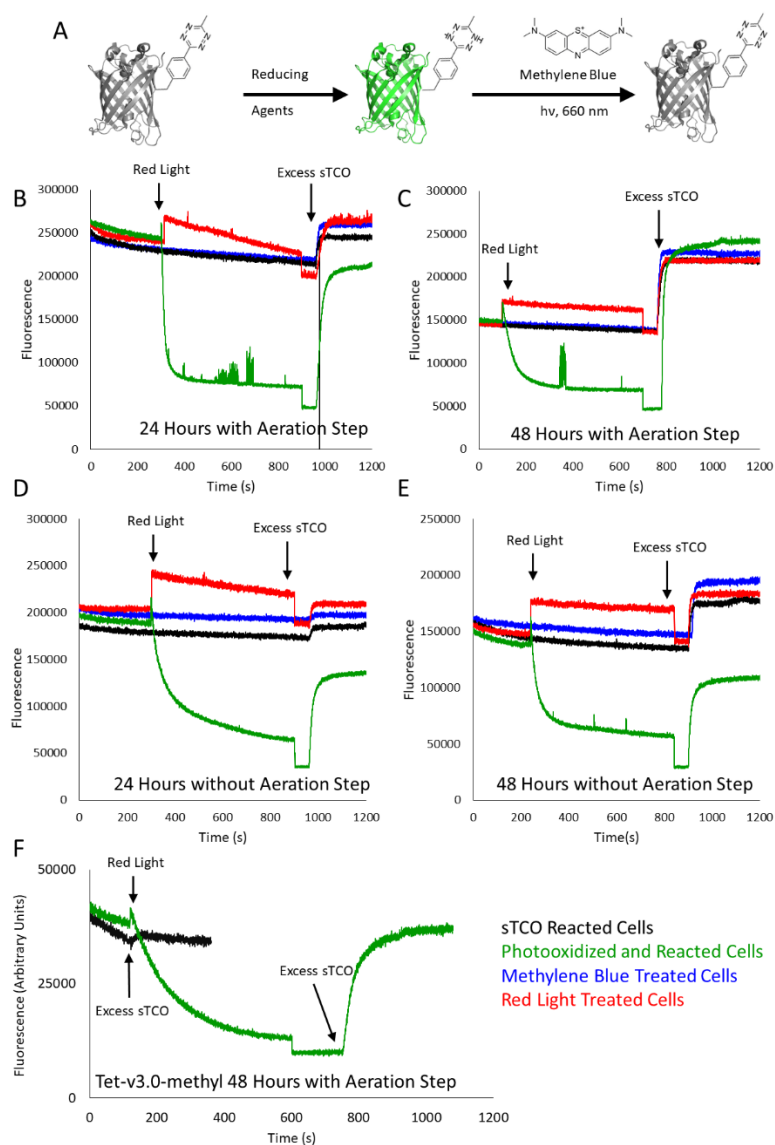


Fig 4.5. Photooxidation of Tet-v2.0 and Tet-v3.0 in *E. coli*.

A) Photooxidation of GFP with methylene blue and red light. Oxidized tetrazine quenches GFP until reduced. Photooxidation results in further quenching. B) Photooxidation of cells expressing GFP150-Tet-v2.0 for 24 hours. The cells were incubated for 2 hours with high aeration prior to use. Upon addition of sTCO, sTCO treated cells (black) and Methylene blue treated cells (green) increase in fluorescence corresponding to the IEDDA reaction. Cells treated with red light (blue) show an increase in fluorescence signal upon application of light, and a corresponding decrease upon removal. Cells treated with methylene blue and red light (red) decrease in fluorescence upon exposure to red light. Subsequent reaction with sTCO results in a return in fluorescence to the level of untreated cells. C) Photooxidation of cells expressing GFP150-Tet-v2.0 for 48 hours. The cells were incubated for 2 hours with high aeration prior to use. D) Photooxidation of cells expressing GFP150-Tet-v2.0 for 24 hours without an additional incubation step. E) Photooxidation of cells expressing GFP150-Tet-v2.0 for 48 hours without an additional incubation step. F) Photooxidation of cells expressing GFP150-Tet-v3.0-methyl for 48 hours. The cells were incubated for 2 hours with high aeration prior to use. Reaction with sTCO in the absence of methylene blue and red light shows little increase in fluorescence corresponding to reaction.

The dihydro-Tet-v2.0 can be oxidized through photooxidation such that subsequent reaction with sTCO returns it to a level of fluorescence that is equivalent to untreated cells reacted with sTCO. As these cells must undergo a greater fluorescence increase to return to the same level upon reaction with sTCO, more of the protein must be oxidized to allow for such a reaction. Cells were compared based on expression time. It was found that shorter expression lengths resulted in a greater proportion of the GFP-Tet-v2.0 that was in the dihydro state (Fig 4.5).

An important step in the preparation of cells for reaction with sTCO is an aeration step in which the expressed and washed cells are incubated in buffer with high levels of aeration. When the effect of this treatment on *in cellulo* oxidation state was evaluated it was found that in absence of the extended wash step photooxidation resulted in a decrease in fluorescence, but full fluorescent return to an equivalent level to untreated cells is dependent on this extended wash step (Fig 4.5 D,E). Additionally, a time dependence was observed for reactivity of these proteins towards sTCO in the absence of photooxidation. Cells that were expressed for 24 hours showed limited reaction with sTCO in comparison to cells expressed for 48 hours. This suggests that the oxidation state of the cellular environment varies over the course of an expression.

While differences were observed in the cellular abundance of oxidized and reduced tetrazines at 24 and 48 hours after the start of an expression, limited data exists describing how the reduction of Tet-v2.0 changes over course of an expression. Such measurements would detail the oxidation state of Tet-v2.0 during translation. When the fluorescence of GFP150-Tet-v2.0 producing cells is measured an increase in the fluorescence occurs corresponding to production of protein (Fig 4.6). Overtime this fluorescence decreases, seeming to indicate a loss of GFP. However, it is possible that this decrease in fluorescence corresponds to an oxidation event resulting in quenching of the GFP fluorescence. When the protein in the cells are reacted with excess sTCO, an increase in the fluorescence is observed corresponding to reaction of oxidized tetrazine to form a high fluorescent product (Fig 4.6). As this fluorescence does not decrease as a factor of expression time, this data suggests that the drop in GFP fluorescence of the unmodified cells corresponds to oxidation. As a control, the samples were also treated with methylene blue and red light to ensure that the amino

acid was in the oxidized state. As this treatment resulted in a steady lower limit of fluorescence (Fig 4.6), the fluorescence of the unmodified culture approximates the amount of protein in the oxidized (methylene blue treated) or reduced/reacted (sTCO treated) states. When GFP150-Tet-v3.0-methyl expressing cells were subject to the same treatment, similar results were observed. It is notable that the Tet-v3.0-methyl cells do not result in a continuing drop in GFP fluorescence, but rather level off after 28 hours.

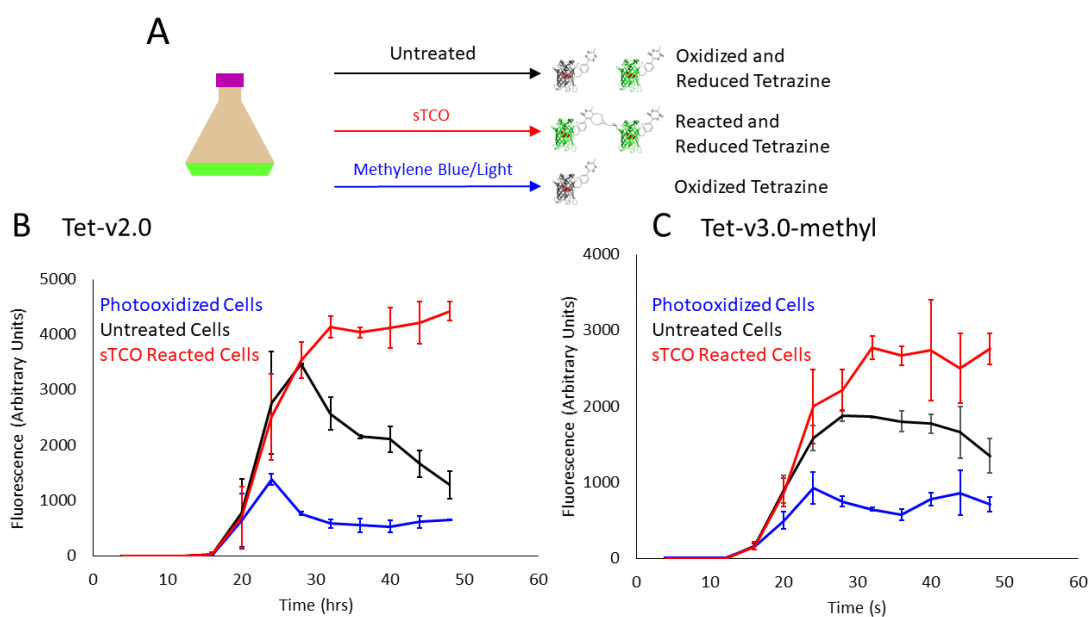


Fig 4.6. Time Dependent Photooxidation.

A) The fluorescence of cultures expressing GFP150-Tet-v2.0 or GFP150-Tet-v3.0-methyl were measured every four hours. A subset of cells was reacted with sTCO or methylene blue prior to measurement. B) Fluorescence of GFP150-Tet-v2.0 shows greatest fluorescent in cells treated with sTCO and least fluorescence in cells that were photooxidized. Untreated cells decrease in fluorescence overtime after fully expressed. C) Fluorescence of GFP150-Tet-v3.0-methyl exhibits many of the same behaviors as Tet-v2.0. Untreated cells do not show as significant of a decrease in fluorescence.

When the greatest increase in protein fluorescence occurs during an expression, the fluorescence of the untreated cells approximates that of those cells treated with sTCO. This suggests that newly expressed protein contains tetrazine in the high fluorescent reduced state. There are two options consistent with this result; the amino acid undergoes translation in the oxidized state and the resulting tetrazine protein is oxidized but becomes reduced soon after expression, or counter to our previous

understanding, the Tet-v2.0 and Tet-v3.0-methyl amino acids undergo translation in the reduced state. This would indicate that the evolved aaRSs load the dihydro-tetrazine amino acids and that the newly translated protein contains tetrazine in the dihydro form. Subsequently the dihydro-tetrazine on protein can equilibrate with the oxidized form *in cellulo*. Over the course of an expression, that equilibrium changes resulting in a higher abundance of oxidized tetrazine later in the expression. This shift in oxidation is potentially a result in metabolic changes of the *E. coli* over the course of the expression.

To verify that the reduced tetrazine amino acid is the substrate for the orthogonal tRNA/RS pairs that were evolved for the incorporation of Tet2.0 and Tet-v3.0-methyl, the D12 Tet-v2.0 aaRS was kinetically characterized *in vitro*. The R2-84 Tet-v3.0-methyl aaRS was not similarly characterized due to difficulty in overexpression of the *M. barkeri* PylRS. The D12 aaRS was cloned into the pBAD expression plasmid with a His tag for purification. The D12 aaRS was expressed and purified. The corresponding orthogonal tRNA was produced via *in vitro* transcription using T7 RNA polymerase. Additionally, the dihydro form of Tet-v2.0 was synthesized. In an anaerobic environment, the D12 aaRS was assayed for its ability to aminoacylate the corresponding tRNA with both the dihydro-Tet-v2.0 and the oxidized Tet-v2.0. The dihydro-Tet-v2.0 was found to be a substrate for the D12 enzyme at an amino acid concentration of 5 μM (Fig 4.7). In contrast at concentrations of Tet-v2.0 as high as 240 μM , no activity was found for aminoacylation of the corresponding

tRNA with the oxidized form of Tet-v2.0. With the dihyrdo-Tet-v2.0 it was found that up to 70% of tRNA was aminoacylated by the D12 aaRS.

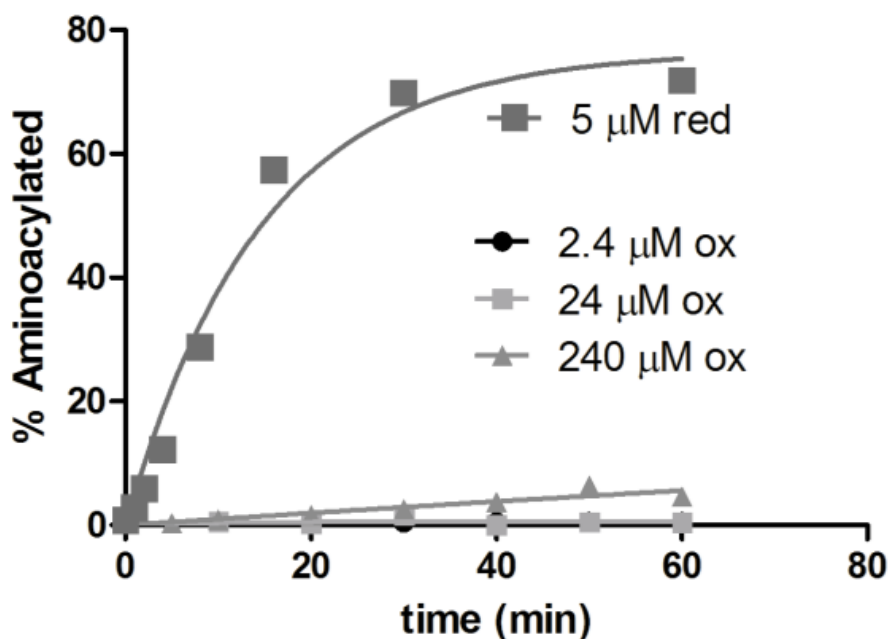


Fig 4.7. *In vitro* Aminoacylation of tRNA.

At varying concentrations of reduced and oxidized Tet-v2.0, the percent of tRNA that was aminoacylated was measured. The dihydro-Tet-v2.0 acts as a substrate where the oxidized form does not.

The selectivity of the D12 aaRS for the dihydro-Tet-v2.0 suggests that the dihydro-Tet-v2.0 rather than the oxidized form is the primary substrate for the D12 aaRS *in cellulo*. This is corroborated by the fact that during protein expression, the majority of newly produced protein shows limited to no reactivity with sTCO as would be expected of dihydro-tetrazines (Fig 4.6). Given that the D12 aaRS was selected for the oxidized form of the amino acid rather than the reduced, it would be expected that the amino acid becomes reduced *in cellulo* during the selection process. Additional selections would be expected to result in aaRSs specific for the reduced form. To generate an aaRS that can use the oxidized form of tetrazines as a substrate an alternative structure is required. Tetrazines with a more negative reducing potential would more strongly favor the oxidized state, enabling selections for a perpetually oxidized tetrazine amino acid.

Given that the dihydro-Tet-v2.0 is the substrate for the D12 aaRS, it is unclear whether the addition of the dihydro-Tet-v2.0 to media is more effective at producing protein as opposed to the addition of the oxidized amino acid. To test this, cultures

expressing the D12 aaRS and an amber codon interrupted GFP gene were supplemented with Tet-v2.0 and dihydro-Tet-v2.0. No major differences in GFP expression or cell growth were observed (Fig 4.8) indicating that the amino acids undergo reduction and oxidation readily such that the reduction state of the supplemented amino acid is not limiting.

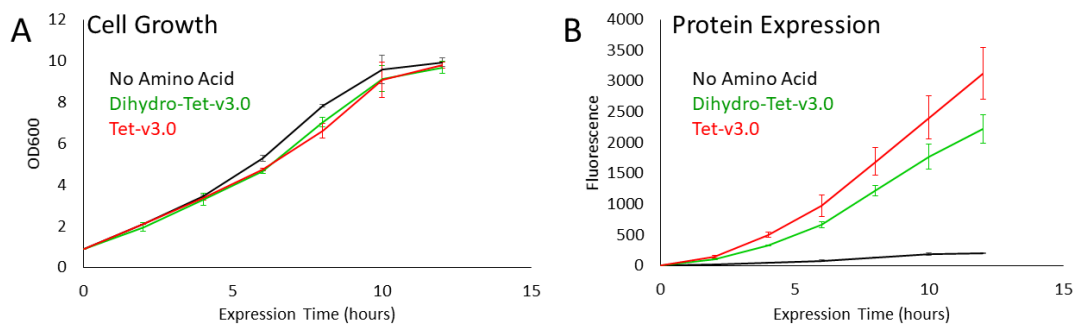


Fig 4.8. Effect of Dihydro-Tet-v2.0 on Cellular Expression.

A) Optical density of cells expressing GFP150-Tet-v2.0 in the presence of 1 mM Tet-v2.0 and dihydro-Tet-v2.0. No change in growth is observed. B) Fluorescence of cells expressing GFP150-Tet-v2.0 in the presence of 1 mM Tet-v2.0 and dihydro-Tet-v2.0.

Indeed, in cultures, it is observed that dihydro-Tet-v2.0 oxidizes in the presence of atmospheric oxygen as observed by a color change characteristic to both molecules. This suggests that atmospheric oxygen is sufficient to oxidize Tet-v2.0. However, even purified protein shows incomplete reaction with sTCO when measured using mobility shift assays (Fig 4.1). By performing mobility shift assays on protein purified throughout an expression it was shown that the unreacted protein accumulates as a product of expression time (Fig 4.9). This suggests that the cause of the unreacted protein is not GFP150-dihydro-Tet-v2.0, but rather a separate side reaction of Tet-v2.0 that results in a product that is unreactive to sTCO. Three possibilities exist as to what this modification may consist of: The modification could be the misincorporation of natural amino acids in place of the amber stop codon. The lack of reactivity could be due to degradation of the tetrazine ring to a smaller byproduct. Finally, the modification could be an addition reaction of the tetrazine functional groups with cellular

components. While the exact reaction causing this effect is unclear, insight in to the nature of this modification will enable steps to minimize its occurrence.

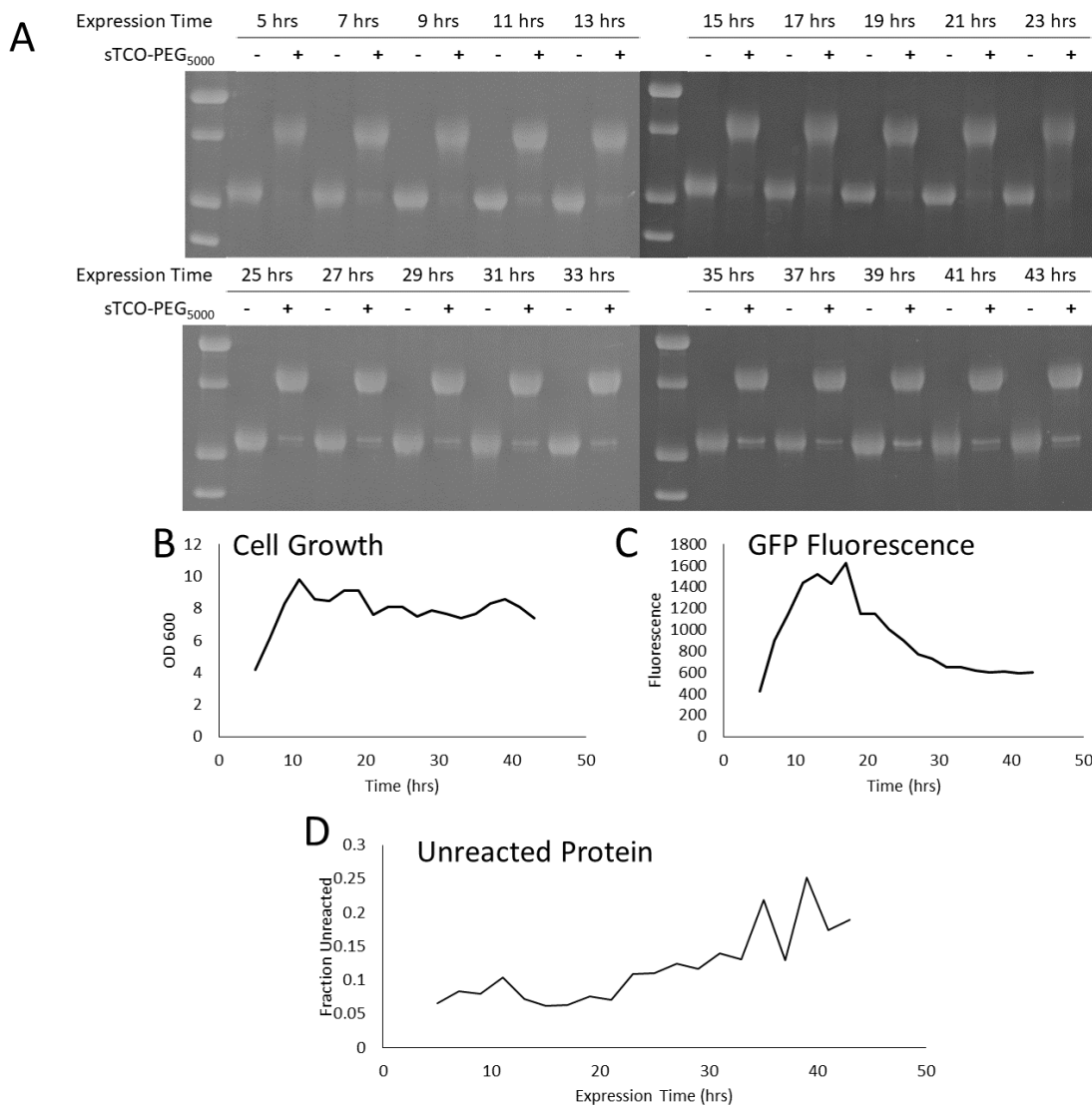
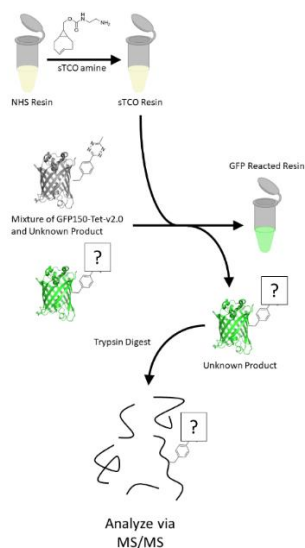


Fig 4.9. Expression Dependent Gel Shift of GFP150-Tet-v2.0.

A) An expression of GFP150-Tet-v2.0 was sampled every 2 hours. Purified protein was reacted with the sTCO-PEG₅₀₀₀. Accumulation of unreacted protein increased with expression time. B) Optical density of harvested cells as a function of expression time. C) Fluorescence of harvested cells as a function of expression time. Fluorescence follows the same pattern as found in Fig 4.6. D) Densitometry of the unshifted band shows an increase in the amount of GFP that remains unreactive overtime.

To determine the nature of the reaction with GFP-Tet-v2.0, a peptide containing the unreactive tetrazine was purified for mass spectrometry analysis. To isolate specifically the protein that is unreactive to sTCO, sepharose resin displaying an sTCO functional group was generated through a coupling reaction of amine-functionalized sTCO with commercially available NHS resin (Fig 4.10). The resin was washed to remove unreacted amine-functionalized sTCO. GFP-Tet-v2.0 was expressed for 72 hours to increase the proportion of the protein that was unreactive to sTCO. This protein was purified via His-tag, buffer exchanged into ammonium bicarbonate buffer, and applied to the generated sTCO resin. In doing so, the portion of protein reactive to sTCO was expected to bind to the resin. The remaining protein was collected and analyzed via mass spectrometry (Fig 4.11). The protein was found to consist of a set of higher molecular weight adducts with masses between 100 to 400 Da (Fig 4.11). While the resolution is of too low of quality to identify these peaks, a tryptic digestion was performed on the purified protein. The resulting peptide was analyzed using a Thermo Scientific Orbitrap Fusion Lumos Mass Spectrometer. The collected data suggests that no adducts are forming and that the majority of the unreacted bands consist of misincorporation of canonical amino acids at the amber codon.

Fig 4.10. Purification of the Unreactive Protein.



GFP150-Tet-v2.0 that shows limited reactivity via mobility shift assays was reacted with an sTCO resin generated by reacting sTCO-amine with NHS-Sepharose. The GFP that was not reacted was collected and proteolytically digested with trypsin for mass spectrometry analysis.

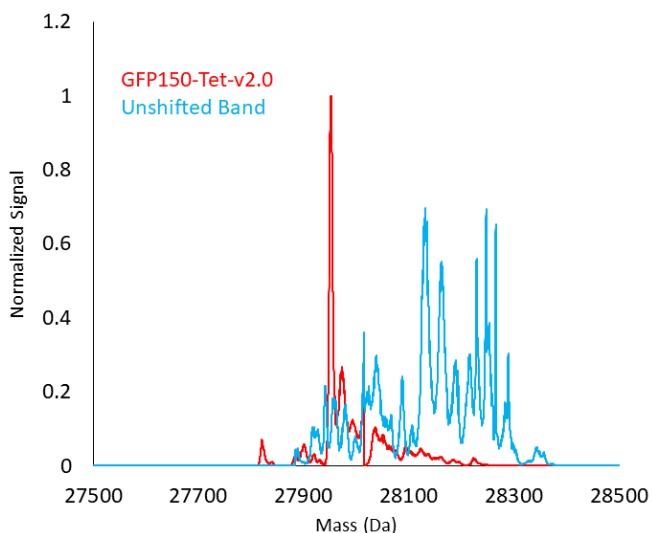


Fig 4.11. Mass Spectra of the Unreactive Protein.

In contrast to GFP150-Tet-v2.0, the unreactive protein appears to consist of several higher molecular weight adducts.

Substitution of the tetrazine amino acid can result in tetrazines with vastly different chemical properties. In general, tetrazines with more electron withdrawing substituents are expected to result in tetrazines with faster reaction rates as well as reaction equilibria that disfavors the oxidized state⁶². Thus, by modifying the tetrazine, there exists the potential to change the oxidation state of the amino acid in living systems. There are two routes to control the reduction reaction. A tetrazine that is less susceptible to reduction can be generated such that the amino acid is stable *in cellulo* in the oxidized form, or a tetrazine can be generated that is more susceptible to reduction and more stable in the reduced state such that it can be photooxidized in a controlled manner to trigger reaction with strained alkenes. Given that we see an accumulation of unreactive tetrazines over the course of protein expression that corresponds to the proportion of the protein in the oxidized state, a reasonable hypothesis would be that the oxidized state of tetrazines is responsible for the irreversible reactions that result in the formation of unreactive protein. Therefore, by generating a tetrazine that is stable in and predominately found in the reduced state, off target reactions can be reduced or eliminated.

Two amino acids were synthesized with a greater predicted stability in the reduced state (Fig 4.12A). Both the fluoro-Tet-v2.0 and the Tet-v2.0-acetyl are predicted to favor the dihydro state due to the electron withdrawing character of their substituents.

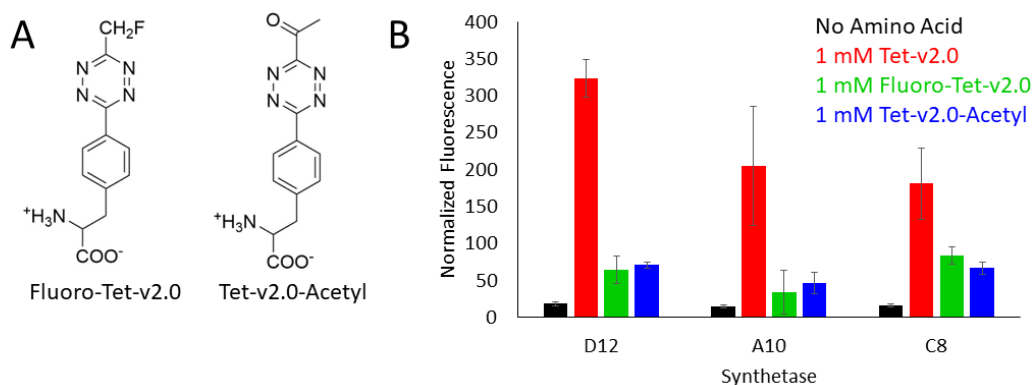


Fig 4.12. Incorporation of Tet-v2.0 Derivatives.

A) Structures of Fluoro-Tet-v2.0 and Tet-v2.0-Acetyl. B) Permissivity of select Tet-v2.0 aaRSs for Fluoro-Tet-v2.0 and Tet-v2.0-Acetyl. The C8 aaRS was most effective at incorporating the new amino acids.

To determine if protein can be produced with these amino acids, an aaRS capable of incorporating these amino acids must be generated. Rather than performing selections, we evaluated the Permissivity of existing Tet-v2.0 aaRSs. Previously Tet-v2.0 aaRSs have been shown to incorporate other amino acids⁶³ which suggests that they may be able to incorporate our amino acids of interest. Cells cotransformed with a pBK plasmid containing Tet-v2.0 aaRSs as well as a pALS plasmid containing an amber codon suppressed GFP gene were grown in the presence of Tet-v2.0, fluoro-Tet-v2.0, and Tet-v2.0-acetyl (Fig 4.12B). The C8 aaRS was shown to be permissive to both the fluoro-Tet-v2.0 and Tet-v2.0-acetyl, though with significantly lower efficiency compared to the Tet-v2.0 amino acid.

Conclusion

While tetrazines are useful as bioorthogonal ligations, they may be limited by some of the off-target reactions that they undergo. Reduction to form the dihydro-tetrazine is the most well characterized of these reactions. This reaction is especially relevant *in vivo* where the reducing cellular environment often favors the dihydro-tetrazine as opposed to the oxidized form. Tools exist to address this problem in the form of photooxidation as well as the development of tetrazines with altered redox

activity. Photooxidation, can be a powerful tool, not just for enabling control of when a reaction occurs. Some of the most reactive tetrazines, that can have limited *in vivo* half-lives, may be stable in the dihydro state. Photooxidation enables the use of these tetrazines *in cellulo* such that higher rates of reaction are possible. Photooxidation can also result in the spatial control of the reaction, which has utility in performing reactions in live organisms as well as spatial control for microscopy work.

While evidence exists for other irreversible side reactions that tetrazines undergo, this evidence is not absolute. The lack of complete reaction observed in mobility shift assays could be due to reaction with small molecules in cells, or it could be due to misincorporation of canonical amino acids. Investigation into the nature of this unreacted protein is necessary to not only make the reaction between tetrazines and strained alkenes more accessible but may have implications to genetic code expansion as a whole.

Materials and Methods

Mobility Shift Assay

Purified protein was diluted in PBS to a concentration of 50 μ M. Protein was reacted with an excess of sTCO-PEG₅₀₀₀ (250 μ M) for 30 minutes. No quenching reaction was performed. Samples were mixed with Laemmli buffer and heated to 95 °C for 15 minutes to denature. Samples were run on SDS-PAGE.

GFP-Tet-v2.0/3.0 Cell Expression

Cultures of non-inducing media (5 mL, Table 2.1) containing tetracycline (25 μ g/mL) and ampicillin (100 μ g/mL) was inoculated with cells co-transformed with the pBAD-GFP-TAG150 plasmid containing a GFP gene interrupted at site 150 by an amber stop codon and an ampicillin resistance gene as well as a pDule1 plasmid containing the D12 aaRS gene and a tetracycline resistance gene. For expression of GFP150-Tet-v3.0, the pDule1 plasmid contained the R2-84 aaRS gene. Cells were grown for 16-24 hours at 37 °C.

The grown non-inducing media (0.5 mL inoculation volume) was used to inoculate 50 mL cultures of auto-inducing media (Table 2.1) containing tetracycline (25 μ g/mL) and ampicillin (100 μ g/mL). Amino acid was added to the media to a

concentration of 0.5 mM from a stock solution of 100 mM amino acid in DMF. For expression of GFP150-Tet-v2.0, the amino acid Tet-v2.0 was added. For expression of GFP150-Tet-v3.0-methyl, the Tet-v3.0-methyl amino acid was added. Cultures were grown for 24-72 hours at 37 °C. Cells were harvested through centrifugation at 5000 rcf for 10 minutes. Cell pellets were stored at -80 °C until use.

For the generation of purified protein, stored cells were thawed and resuspended in 5 mL of TALON wash buffer (300 mM NaCl, 50 mM Na₂HPO₄, pH 7.0). The resuspended cells were lysed using a Microfluidics M-110P microfluidizer. Lysed cells were clarified via centrifugation at 20000 rcf for 1 hour. The pellet was discarded. TALON resin (100 µL bed volume per 50 mL culture) was incubated with the supernatant for 1 hour at 4 °C with gentle rocking. The mixture of supernatant and resin was applied to a column and the flow through was discarded. The resin was washed with 3 x 10 mL of TALON wash buffer. Protein was eluted with 4 x 250 µL TALON elution buffer. The elutions were subsequently pooled. Purified proteins were buffer exchanged into PBS using GE Healthcare PD-10 columns. Buffer exchanged protein was concentrated using a VWR 3 kDA cutoff centrifugal filter.

Preparation of Cells for Reaction

Expressed cells were washed prior use in *in cellulo* labeling experiments. Cells from a 50 mL culture were resuspended in 5 mL of PBS and aliquoted into 5 x 1 mL aliquots. Aliquots were pelleted through centrifugation for 5 minutes at 2000 rcf. The supernatant was discarded. The pellet was washed an additional two times via resuspension in 1 mL PBS and centrifugation. After the third wash the pellet was resuspended in 1 mL PBS. Some pellets were subjected to an additional aeration step in which the cells resuspended in PBS were placed into an aerated culture tube. The cells were incubated for 2 hours at 37 °C. The aerated cells were stored on ice until ready for use.

***In cellulo* Reactions (Kinetics)**

Cells containing GFP150-Tet-v3.0 derivatives were added to a quartz cuvette and diluted to a volume of 3 mL in PBS. Fluorescence was monitored over time in a PTI Quntamaster fluorimeter (Excitation 488 nm, Emission 509 nm). To the cuvette, sTCO was added at final concentrations ranging from 133 nM to 3.32 µM. An increase

with fluorescence was observed and individual runs were stopped when no additional fluorescence increase was observed. Curves corresponding to the reaction were fit to exponential equations using the program Igor. From the fit, pseudo 1st order rate constants were obtained for each run. Pseudo 1st order rate constants were plotted against the concentration of sTCO to obtain a linear curve. The slope of the linear curve was taken as the second order rate constant.

TCEP Reduction *in vitro*

GFP150-Tet-v2.0 was diluted to a concentration of 50 μM in PBS. TCEP was added to a concentration of 1 mM or sTCO was added to a concentration of 250 μM . Reactions were allowed to proceed at room temperature for 30 minutes prior to detection. A Biotek Synergy2 microplate reader was used to make measurements. Approximately 10 minutes elapsed in between the end of the reaction and when measurements were taken.

Photooxidation and Reaction in cells

Cells containing GFP150-Tet-v2.0 and GFP150-Tet-v3.0-methyl were added to a quartz cuvette and diluted to a volume of 3 mL in PBS containing 1 μM methylene blue. Fluorescence was monitored over time in a PTI Quantamaster fluorimeter (Excitation 488 nm, Emission 509 nm). Samples were irradiated using a Luxeon Rebel Color LED (660 nm, 300 mW) for 10 minutes. After irradiation was complete, the light source was removed and 1 minute was allowed to establish a post-irradiation baseline. A solution of sTCO (2 mM, 10 μL) in methanol was then added to the cuvette for a final sTCO concentration of 6.64 μM . Samples were allowed to react until no increase in fluorescence was observed.

Cultured Cells Photooxidation

Non-inducing media cultures of *E. coli* cells containing a pBAD plasmid encoding a GFP-TAG150 gene and a pDule1 plasmid encoding either the D12 synthetase (Tet-v2.0) or R2-84 synthetase (Tet-v3.0-methyl) genes were used to inoculate 2 x 50 mL auto-inducing media containing ampicillin (100 $\mu\text{g}/\text{mL}$) and tetracycline (25 $\mu\text{g}/\text{mL}$). The amino acid of interest (1 mM) was added to the media and the OD₆₀₀ and the fluorescence of the cultures were measured every 4 hours for 48 hours. Fluorescent measurements were made by taking 50 μL of media and mixing

with either 50 μL H_2O , 50 μL of 20 μM methylene blue, or 50 μL of 2 mM sTCO in methanol. Samples that were mixed with methylene blue were also irradiated with a Luxeon Rebel Color LED (660 nm, 300 mW) for 10 minutes prior to measurement of fluorescence. The final samples consisting of 100 μL total volume were mixed with H_2O to a final volume of 2 mL and the fluorescence was measured using a Turner Biosystems Picofluor fluorimeter.

Expression and Purification of D12 synthetase

BL21AI cells were transformed with a kanamycin resistant pET28 plasmid encoding a His-tagged D12 aaRS gene. The transformed cells were plated on LB agar containing kanamycin (50 $\mu\text{g}/\text{mL}$) and grown overnight at 37 $^\circ\text{C}$. Colonies from the transformation were used to inoculate 5 mL of non-inducing media containing kanamycin (50 $\mu\text{g}/\text{mL}$). Non-inducing media cultures were grown for 24 hours shaking at 37 $^\circ\text{C}$. Auto-inducing media cultures 2 x 500 mL were inoculated with 500 μL of grown non-inducing culture. The auto-inducing media was supplemented with 50 $\mu\text{g}/\text{mL}$ kanamycin and 0.2 % (w/v) lactose. These cultures were expressed for 24 hours before pelleting cells through centrifugation (5000 rcf, 10 minutes). The supernatant was discarded, and cell pellets were stored at -80 $^\circ\text{C}$ until needed.

Cell pellets were thawed and resuspended in 5 mL of TALON wash buffer. Cells were lysed using a Microfluidics M-110P microfluidizer. Lysate was clarified via centrifugation (1 hour, 20000 rcf). Clarified lysate was filtered through 0.2 μm sterile filters and His-tag purified on an Äkta Explorer using a 5 mL HisTrap column. Purified protein was buffer exchanged into 20 mM Tris, 50 mM NaCl, 10 mM β -mercaptoethanol at pH 8.5. The protein was then concentrated in a VWR 3 kDA cutoff centrifugal filter to a concentration of 680 μM . Protein was flash frozen in liquid nitrogen until use.

Dihydro-Tet-v2.0 Toxicity and Oxidation Measurements

Cells containing the pBAD and pDule1 plasmids used above was used to inoculate a non-inducing media culture containing ampicillin (100 $\mu\text{g}/\text{mL}$) and tetracycline (25 $\mu\text{g}/\text{mL}$). The culture was grown at 37 $^\circ\text{C}$ for 24 hours shaking at 250 rpm. Autoinduction media cultures (300 mL) containing ampicillin (100 $\mu\text{g}/\text{mL}$) and tetracycline (25 $\mu\text{g}/\text{mL}$) were inoculated with 0.5 mL of grown non-inducing media.

Cells were grown at 37 °C with shaking at 250 rpm until an OD₆₀₀ of 1.0 was observed. At this point the culture was split into 6 x 50 mL cultures and each culture was supplemented with either DMF (0.5 mL), 1 mM Tet-v2.0 (in 0.5 mL DMF), or 1 mM dihydro-Tet-v2.0 (in 0.5 mL DMF). Fluorescence and OD₆₀₀ of these cultures was measured every 2 hours for 12 hours.

Expression Dependent Mobility Shift

Cells cotransformed with an ampicillin resistant pBAD plasmid containing an amber codon interrupted GFP gene as well as a tetracycline resistant pDule1 plasmid containing the gene for the D12 synthetase were used to inoculate a 5 mL culture of non-inducing media. The culture was grown for 24 hours at 37 °C while shaking at 250 rpm. An auto-inducing media culture (1 L, 100 µg/mL ampicillin, 25 µg/mL tetracycline) was inoculated with 1 mL of grown non-inducing media culture. The auto-inducing media contained 100 µg/mL ampicillin, 25 µg/mL tetracycline, and 1 mM Tet-v2.0. The OD₆₀₀ of the culture was monitored. When an OD₆₀₀ of 1.0 was observed, the 1 L culture was divided into 20 x 50 mL cultures. Five hours after an OD₆₀₀ of 1.0 was observed and subsequently every two hours for 40 hours, the OD₆₀₀ and fluorescence of one of the cultures was measured and the culture was harvested via centrifugation (5000 rcf, 10 minutes). Cell pellets were stored at -80 °C until purified as described above.

Once purified the GFP150-Tet-v2.0 was allowed to oxidize at 4 °C to ensure that no dihydro-Tet-v2.0 was present. The concentrations of GFP150-Tet-v2.0 from all samples were measured and the samples were diluted to a final concentration of 30 µM protein. Each sample was reacted with 100 µM sTCO-PEG₅₀₀₀ for 30 minutes at room temperature. Samples were denatured with Laemmli buffer and heated at 95 °C for 15 minutes prior to analysis via SDS-PAGE.

NHS resin purification/Unshifted Purification

NHS-sepharose resin (1.5 mL bed volume) was reacted with reacted with 480 µL of sTCO-amine. The reaction was allowed to proceed for 4 hours gently rocking at 4 °C. The resin was pelleted via centrifugation (900 rcf, 2 minutes). The unreacted supernatant was removed. The resin was washed three times using addition of 3 mL 100 mM ammonium bicarbonate buffer (100 mM, pH 8.5). Resin was pelleted via

centrifugation (900 rcf, 2 minutes) in between wash steps to remove ammonium bicarbonate buffer. The resulting product is an sTCO-resin.

GFP150-Tet-v2.0 (100 μ L, 800 μ M) was reacted with 250 μ L bed volume of generated sTCO-resin. Reaction was allowed to proceed for 15 minutes at 4 °C. Resin was pelleted via centrifugation (900 rcf, 2 minutes) and unreacted GFP supernatant was saved. The supernatant was then applied sequentially to 3 x 250 μ L bed volume of resin to ensure that no unreacted GFP150-Tet-v2.0 remained. The final supernatant was fluorescent indicating the presence of GFP, though most GFP appeared to adhere to the first set of resin.

Trypsin Digestion/Peptide Purification

GFP that was purified using sTCO-resin above was diluted to a concentration of 100 μ M in ammonium bicarbonate buffer (100 mM, pH8.5). The protein was heated to 90 °C for 30 minutes to denature the GFP. The sample was cooled on ice for approximately 15 seconds before the addition of 5 μ M trypsin. The mixture of GFP and trypsin was incubated at 37 °C for 16 hours. Upon completion of the digestion reaction, the mixture was analyzed by SDS-PAGE to ensure complete digestion (Data not shown).

The trypsin digested protein was diluted to 1 mL final volume in equilibration buffer (99.9% MS-grade H₂O, 1% MS-grade trifluoroacetic acid). The pH of the corresponding solution was confirmed to be 0 by pH test strips. An Oasis 10 mg Extraction Cartridge was wetted through the addition of 2 x 1 mL wetting buffer (99.9% MS-grade acetonitrile, 0.1% trifluoroacetic acid). The column was then washed with 3 x 1 mL equilibration buffer. Acidified sample was applied to the column. The column was washed with 3 x 1 mL equilibration buffer prior to elution with 4 x 250 μ L elution buffer (70% MS-grade acetonitrile, 29.9% MS-grade H₂O, 0.1% MS-grade formic acid). The eluted sample was freeze dried prior to MS analysis.

(S)-2-((tert-butoxycarbonyl)amino)-3-(4-(6-methyl-1,2-dihydro-1,2,4,5-tetrazin-3-yl)phenyl)propanoic acid: In a dry, 15 mL heavy walled reaction tube containing Boc-protected 4-CN phenylalanine (200 mg, 0.688 mmol) was charged with Ni(OTf)₂ (122 mg, 0.344 mmol) and acetonitrile (0.36 ml, 6.8 mmol) under argon atmosphere. Then anhydrous hydrazine (1.1 mL, 34.4 mmol) was added to the reaction mixture and

purged with argon for 5 minutes and immediately sealed the tube. The reaction mixture was heated to 50 °C for 24 hr. Then the reaction mixture was cooled in ice bath, opened slowly and added 5 ml of water and ethyl acetate. Under ice cooled condition reaction mixture was acidified by 4 N HCl (pH~3). The aqueous part was extracted by ethyl acetate (3 times) and combined organic layer was dried with MgSO₄ and concentrated under reduced pressure. Silica gel flash column chromatography (45 % ethyl acetate in hexanes with 1% acetic acid) yielded 150 mg of **title compound** (0.417 mmol) in the form of a yellowish red gummy material. Yield 60%. ¹H NMR (400MHz, CDCl₃) δ 7.50 (d, 2H), 7.22 (d, 2H), 5.31 (d, 1H), 4.59 (m, 1H), 3.20-3.06 (m, 2H), 1.93 (s, 3H), 1.40 (s, 9H).

Hydrochloride salt(S)-2-amino-3-(4-(6-methyl-1,2-dihydro-1,2,4,5-tetrazin-3-yl)phenyl) propanoic acid: In a dry round bottom flask (RB), Boc protected dihydro tetrazene amino acid (120 mg, 0.33 mmol) was charged with 1 mL dioxane-HCl (dioxane saturated with HCl gas) and 3 mL ethyl acetate under argon atmosphere. The reaction mixture was allowed to stirrer at room temperature for 2 hr. After completion of deprotection, the reaction mixture was concentrated under reduced pressure and re-dissolved in ethyl acetate (2 times) and subsequently concentrated to remove excess HCl gas which afforded solid yellowish red material in quantitative yield. ¹H NMR (700MHz, MeOD) δ 7.40 (d, 2H), 7.10 (d, 2H), 3.97 (t, 1H), 3.03- 2.89 (ddd, 2H), 1.76 (s, 3H). ¹³C NMR (175MHz, MeOD) δ 171.1, 159.8, 154.0, 141.1, 131.2, 129.1, 128.6, 68.3, 54.9, 37.1, 14.7.

Hydrochloride salt of (S)-2-amino-3-(4-(6-(fluoromethyl)-1,2,4,5-tetrazin-3-yl)phenyl) propanoic acid: The starting material Boc-protected 4-cyano phenylalanine (200 mg, 0.68 mmol) was taken in a dried heavy walled reaction tube under argon atmosphere. Maintaining inert atmosphere of the reaction vessel, catalyst Ni(OTf)₂ (0.5 eqv.) and fluoroacetonitrile (6.8 mmol) were added. Then anhydrous hydrazine (50 eqv.) was added very slowly to reaction mixture under stirring condition at 0 °C. Purged argon for another 10 minutes and immediately sealed the tube. The reaction vessel was immersed into the preheated oil bath at 50 °C for 24 hours. After

that, the reaction vessel was lifted from the oil bath and allowed to come at room temperature. The reaction mixture was poured into a beaker and added 10 eqv. 2M NaNO₂ aqueous solution and 5 mL water. Then the aqueous phase was acidified with 4M HCl (pH~2) under ice-cold condition with homogeneous mixing and extracted with ethyl acetate (3x 20 mL). The combined organic layer was washed with brine, dried with anhydrous Na₂SO₄ and concentrated under reduced pressure. Silica gel flash column chromatography (30 % ethyl acetate in hexanes with 1% acetic acid) yielded desired Boc-protected tetrazine amino acid derivative (155 mg, 0.303 mmol) in the form of a pinkish red gummy material. Yield 44%. Similar way, Boc- group was deprotected using the dioxane-HCl (mentioned above) and formed the title amino acid as a pinkish solid material in quantitative yield (~ 97-98 %).

Hydrochloride salt of (S)-3-(4-(6-acetyl-1,2,4,5-tetrazin-3-yl)phenyl)-2-aminopropanoic acid: Following the synthetic procedure of fluoromethyl derivative, starting from 0.25 g (0.86 mmol) of Boc-protected 4-cyano phenyl alanine and 0.61 mL (8.6 mmol) of pyruvitrile afforded Boc-protected version of the title amino acid (0.140 g, 0.36 mmol). Yield 42%. Then Boc group was deprotected using same method (dioxane -HCl) which made chloride salt of the title amino acid as a pinkish solid in quantitative yield (~ 97-98 %).

Permissivity Acetyl Tet/FluoroTet

Cells containing both pBK plasmids encoding a *M. jannaschii* derived TyrRS gene with a kanamycin resistance gene as well as a pALS plasmid encoding an amber codon interrupted GFP gene and a tetracycline resistance gene were used to inoculate 5 mL of non-inducing media containing kanamycin (50 µg/mL) and tetracycline (25 µg/mL). Three different sets of cells (D12, A10, C8) each encoding a separate *M. jannaschii* TyrRS sequence that was selected for incorporation of Tet-v2.0 were used to start these cultures. The non-inducing cultures were grown for 24 hours at 37 °C. Auto-inducing media cultures 36 x 5 mL containing kanamycin (50 µg/mL) and tetracycline (25 µg/mL) were each inoculated with 50 µL of one of the non-inducing media cultures. Each of the A10, C8, and D12 cultures were used to inoculate 12 cultures. These 12 cultures were then supplemented with 50 µL of DMF, 1 mM Tet-

v2.0 in DMF, 1 mM fluoro-Tet-v2.0 in DMF, or 1 mM Tet-v2.0-acetyl in DMF (3 cultures per condition). The fluorescence and OD₆₀₀ were measured every 24 hours for 72 hours. Data reported consists of fluorescence and OD₆₀₀ after 48 hours.

Chapter 5

Conclusions and Outlook

Summary

This dissertation explored reactions between tetrazines and strained alkenes *in vivo* for generating tools for the imaging of proteins and manipulation of their activities in live cells and organisms. In chapter 2, ideal bioorthogonal reactions are described as having four qualities:

1) Fast reaction rates

2) Functional groups selective for their reaction with each other

3) Functional groups stable to degradation *in vivo*

4) Small functional groups

These qualities can, to an extent, describe the reactions between Tet-v2.0 or Tet-v3.0 and sTCO. These reactions occur with second order rate constants in the range of 70,000-90,000 M⁻¹s⁻¹ if Tet-v3.0-isopropyl is excluded. While these reactions are fast relative to the bioorthogonal reactions of SPAAC or CUAAC (10⁻¹ to 200 M⁻¹s⁻¹) the rate constants are orders of magnitude lower than the ideal rate constant of diffusion limited reactions (~10⁹ M⁻¹s⁻¹). The second and third criteria relate to the selectivity and stability of the sTCO and tetrazine functional groups. These groups are selective enough that minimal off-target reactions are observed via protein mass spectrometry and stable enough that after 48 hours of expression in *E. coli* approximately 85% of tetrazines are still reactive with sTCO molecules. Nonetheless there is room for improvement. Reduction of tetrazines to dihydro-tetrazines can compete with *in cellulo* labeling of tetrazines and approximately 15% of tetrazines lose reactivity after 48 hours. The ideal bioorthogonal ligation would not suffer these drawbacks. Finally, the tetrazine functional group consists of just six atoms. While azides and alkynes present smaller functional groups of 2-3 atoms, the reactions of these functional groups are not as fast, selective, and stable as the IEDDA reaction.

In *E. coli*, proteins incorporating Tet-v2.0 were labeled with sTCO in a reaction with high rate constants and excellent selectivity. Attempts to label proteins with a fluorophore in the form of TAMRA-sTCO were limited by adhesion of TAMRA-sTCO to *E. coli* cells. This labeling reaction was performed successfully in cell lysate. We

predict that this adhesion is due to hydrophobic interactions between the TAMRA-sTCO molecule and cellular components. To overcome potential interactions sTCO can be coupled to other fluorophores known to diffuse across membranes such as BODIPY fluorophores¹²⁷ or more hydrophilic linkers could be generated. While *E. coli* is an excellent system for testing this IEDDA reaction, performing the IEDDA reaction in eukaryotic cells would be more relevant for studying human diseases.

The first challenge to overcome towards performing IEDDA reactions in eukaryotic cells is the generation of an aaRS that is both specific for incorporation of a tetrazine amino acid and orthogonal to translational machinery in eukaryotes. The Tet-v2.0 synthetases used in Chapter 2 lack orthogonality to eukaryotes. While the *M. barkeri* tRNA/aaRS pair is orthogonal to eukaryotic cells, no synthetases specific for Tet-v2.0 were generated. To overcome this, the Tet-v3.0 amino acids were synthesized and effective synthetases were selected. Once tetrazines were incorporated onto proteins in HEK293T cells, labeling of protein with TAMRA-sTCO was attempted. As with *E. coli* cells, TAMRA-sTCO was found to adhere to cells and the ability to assess fluorescent labeling reactions was limited. This adhesion was found to be independent of sTCO reactivity suggesting that hydrophobic interactions are responsible for labeling. Though fluorescent labeling was limited, dimerization of proteins with double-headed sTCO molecules was found to be effective in HEK293T cells. Even so, larger dimerizing agents were shown to be limited by diffusion into cells. In both *E. coli* and HEK293T cells non-specific adhesion of sTCO to cells as well as cellular diffusion of sTCO labels are problems that must be addressed before this technology can be used to study meaningful biological systems.

In *E. coli*, tetrazines undergo side reactions with reducing agents to form dihydro-tetrazines, as well as side reactions that result in a subset of protein that is unreactive to sTCO as measured by mobility shift assays. The formation of dihydro-tetrazines is dominant in cells to the degree that the aaRS developed to incorporate Tet-v2.0 was found to be specific for dihydro-Tet-v2.0. Tools to address the reduction of tetrazines have been developed in the form of photooxidation reactions. These reactions can be used *in cellulo* to oxidize dihydro-tetrazines. With respect to other side reactions of tetrazines, the cause of these reactions remains unclear. Data suggests that this

reaction accumulates as oxidized tetrazine builds up in cells and that the reaction takes the form of an addition reaction. A potential solution to this problem is the development of tetrazine-containing amino acids that are stable in cells as dihydro-tetrazines. These amino acids can then be photooxidized and reacted at will.

Future Experiments

In both *E. coli* and HEK293T cells diffusion and adhesion of sTCO labels limit the use of this technology. Larger and more hydrophobic labels show non-specific adhesion to cells. To alleviate this, the structure of the linker will be varied and the adhesion of synthesized linkers to cells will be measured with flow cytometry. Preliminary data suggests that the addition of more polar or charged functional groups to TAMRA-sTCO labels results in less adhesion (Data not shown). An additional way to address this problem is to create an sTCO-fluorophore that consists of a non-TAMRA fluorophore such as BODIPY or the Alexa series of fluorophores. While TAMRA has been shown to easily diffuse across cell membranes¹²⁷, the TAMRA-sTCO molecule does not share this property.

Unlike TAMRA-sTCO, doubleheaded-sTCO-PEG₁₀₀ shows no signs of adhesion to cells. The addition of this molecule to cells can be used to create artificial protein homodimers. This is an ability shared with a technique called chemical induced dimerization¹³¹. In chemical induced dimerization, two proteins are genetically fused to dimerization domains (Fig 5.1A). A small molecule linker is added that binds to both dimerization domains. In this way, the two proteins of interest associate by proximity. The use of bioorthogonal ligation reactions to achieve this result, would have the advantage of not requiring the use of dimerization domains which may affect the activity of the protein of interest (Fig 5.1B). STAT1 is an excellent candidate molecule for this experiment. STAT1 natively forms both an inactive and active homodimer. Upon phosphorylation *in vivo*, STAT1 switches from the inactive to active homodimer, migrates to the nucleus, and regulates transcription¹³⁴. A key question in the field is whether phosphorylation or dimerization is responsible for nuclear importation and

changes in activity. Selectively inducing the formation of one of these dimerization states, could enable not just controlled activation, but controlled regulation as well.

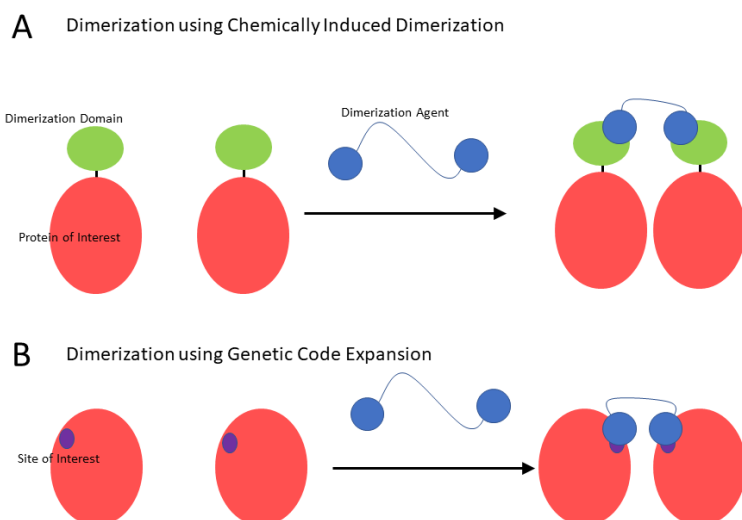


Fig 5.1. Dimerization with Bioorthogonal Ligations.

A) Chemically Induced Dimerization uses genetically attached binding domains to bind to a small molecule dimerizing agent. B) Dimerization with GCE uses bioorthogonal ligations that consist of the addition of a single amino acid to proteins. Dimerization is induced via a double-headed bioorthogonal linker.

The question of the unknown side reaction of Tet-v2.0 that accumulates over time remains. As this amount of side product increases over the course of an expression, modification to an existing tetrazine-containing protein appears to be occurring. Similarly, whole-protein mass spectrometry of the side product indicates that several higher molecular weight adducts are being formed over the course of an expression. Finally, peptide mass spectrometry suggests that no adducts are being formed but rather misincorporation of canonical amino acids at site 150 is responsible for the unreacted protein. Given that the peptide mass spectrometry and the whole protein mass spectrometry differ, future experiments will be directed towards resolving this difference. These experiments will consist of performing high resolution whole-protein mass spectrometry on the same sample that is then trypsin digested and subjected to high resolution peptide mass spectrometry. Additionally, more controls are necessary including high resolution mass spectrometry of untreated GFP150-Tet-v2.0, to ensure that the purification of the unreacted protein is not removing any relevant proteins. Finally, to confirm that this band is not a byproduct of genetic code expansion techniques an experiment in which another genetically encoded amino acid will be

incorporated at site 150 and subject to whole protein and peptide mass spectrometry. As infrequent peptides and proteins may be difficult to detect among large quantities of peptides and proteins without misincorporation, a bioorthogonal amino acid should be used to enrich the sample with potentially misincorporated peptides and proteins. This can be performed through reaction with resin as was performed for Tet-v2.0. In addition to providing insight to side reactions that tetrazines undergo *in cellulo*, these experiments may provide insight into the degree and nature of translational errors caused by GCE.

Potential Research Directions

Substitution of tetrazines with more electron withdrawing and electron donating groups changes the rate of reactions with strained alkenes. More electron withdrawing groups result in higher rate constants. More electron donating group results in lower reaction rates. Electron withdrawing groups should also result in a less negative reduction potential and electron donating groups should result in a more negative reduction potential. Substitution with functional groups such as pyrimidine and pyridine, may result in tetrazines with greater reaction rates, that are stable in cells in the dihydro state (Fig 5.2). Upon photooxidation, these tetrazines could then react with rate constants up to $10^6 \text{ M}^{-1}\text{s}^{-1}$. Currently such amino acids are unusable to due toxicity and stability issues that are associated with the oxidized state. Additionally, more electron-rich tetrazines such as ether substituted tetrazines may be stable in the oxidized

state even in cells (Fig 5.2). These tetrazines would not reduce and could be reacted without the need for an extensive aeration step.

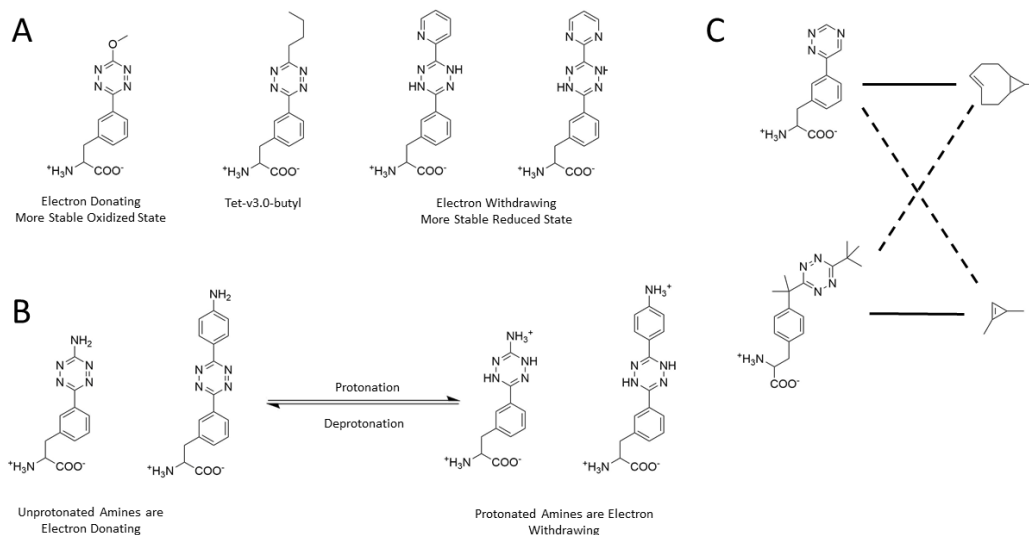


Fig 5.2. Future Routes of Tetrazine Synthesis.

A) Synthesis of tetrazines with electron withdrawing and donating functional groups enables stable tetrazines in the reduced and oxidized states. B) Tetrazines substituted by aromatic amines should undergo a pH dependent transition in reactivity. C) Triazines are reactive with sTCO, but unreactive to cyclopropenes. Sterically hindered tetrazines undergo reactions with cyclopropenes but may have limited reactivity with sTCO.

Another possibility is a tetrazine amino acid that can switch between degrees of reactivity. An amine substituted tetrazine such as amino-Tet-v2.0 or 4-amino-phenyl-Tet-v2.0 would be expected to be electron rich and thus be only mildly reactive (Fig 5.2). As aniline has a pK_a of 4.6, these tetrazines would be susceptible to protonation in biological environments. Upon protonation, the amine is expected to be highly electron withdrawing. This would result in a pH dependent reaction with strained alkenes that would result in a bioorthogonal reaction that functions only in select organelles.

The concept of a set of bioorthogonal reactions that are orthogonal to each other remains desirable. These orthogonal bioorthogonal reactions typically involve two separate reaction chemistries. While it would be ideal to use two IEDDA reactions due to the high rate of reaction, this is limited by cross reactivity. Sterically hindered IEDDA functional groups that lack the ability to react with each other is a possibility (Fig 5.2). A sterically hindered tetrazine could be unreactive to sTCO, but reactive to smaller cyclopropenes or spirohexenes. These functional groups are known to be

unreactive to triazine amino acids⁶³. The triazines would then be reactive with sTCO or TCO functional groups. In this way orthogonal functional groups that all react rapidly could be achieved using IEDDA reactions.

Directions of the Field

Bioorthogonal reactions are seeing increased use in labeling reactions in tissues of live animals. This live animal work is largely directed towards the generation of pre-targeted antibody drug conjugates¹⁰⁹ and radioactive labeling³⁸. Concentrations of label are often low and rapid bioorthogonal reactions with minimal off-target reactions are required. IEDDA reactions between tetrazines and strained alkenes meet these requirements. The continued interest of pharmaceutical communities in antibody drug conjugates and radiolabeling suggests that IEDDA reactions will become increasingly relevant.

The development of bioorthogonal reactions has been a popular topic in chemical biology since the introduction of the concept by Carolyn Bertozzi¹, and an important class of amino acids since the first bioorthogonal amino acid was incorporated into proteins by the Schultz lab¹³⁵. In recent years, much of the work focused on developing bioorthogonal chemistry has been related to the development of faster IEDDA reactions⁶⁵ or the development of orthogonal bioorthogonal reactions⁴². To date, no bioorthogonal reactions have been developed with higher second order rate constants than IEDDA reactions. Therefore, understanding the side reactions of bioorthogonal reactions is relevant not only to understanding limitations of existing IEDDA reactions, but also for the development of future bioorthogonal reactions.

- (1) Hang, H. C.; Yu, C.; Kato, D. L.; Bertozzi, C. R. A Metabolic Labeling Approach toward Proteomic Analysis of Mucin-Type O-Linked Glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (25), 14846–14851. <https://doi.org/10.1073/pnas.2335201100>.
- (2) Sletten, E. M.; Bertozzi, C. R. From Mechanism to Mouse: A Tale of Two Bioorthogonal Reactions. *Acc. Chem. Res.* **2011**, *44* (9), 666–676. <https://doi.org/10.1021/ar200148z>.
- (3) Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G. Adding Amino Acids with Novel Reactivity to the Genetic Code of *Saccharomyces Cerevisiae*. *J. Am. Chem. Soc.* **2003**, *125* (39), 11782–11783. <https://doi.org/10.1021/ja0370037>.
- (4) Uttamapinant, C.; Howe, J. D.; Lang, K.; Beránek, V.; Davis, L.; Mahesh, M.; Barry, N. P.; Chin, J. W. Genetic Code Expansion Enables Live-Cell and Super-Resolution Imaging of Site-Specifically Labeled Cellular Proteins. *J. Am. Chem. Soc.* **2015**, *137* (14), 4602–4605. <https://doi.org/10.1021/ja512838z>.
- (5) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. Selective Identification of Newly Synthesized Proteins in Mammalian Cells Using Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT). *Proc. Natl. Acad. Sci.* **2006**, *103* (25), 9482–9487. <https://doi.org/10.1073/pnas.0601637103>.
- (6) Agarwal, P.; Weijden, J. van der; Sletten, E. M.; Rabuka, D.; Bertozzi, C. R. A Pictet-Spengler Ligation for Protein Chemical Modification. *Proc. Natl. Acad. Sci.* **2013**, *110* (1), 46–51. <https://doi.org/10.1073/pnas.1213186110>.
- (7) Appel, M. J.; Bertozzi, C. R. Formylglycine, a Post-Translationally Generated Residue with Unique Catalytic Capabilities and Biotechnology Applications. *ACS Chem. Biol.* **2015**, *10* (1), 72–84. <https://doi.org/10.1021/cb500897w>.
- (8) Lercher, L.; McGouran, J. F.; Kessler, B. M.; Schofield, C. J.; Davis, B. G. DNA Modification under Mild Conditions by Suzuki–Miyaura Cross-Coupling for the Generation of Functional Probes. *Angew. Chem. Int. Ed.* **2013**, *52* (40), 10553–10558. <https://doi.org/10.1002/anie.201304038>.
- (9) Rieder, U.; Luedtke, N. W. Alkene–Tetrazine Ligation for Imaging Cellular DNA. *Angew. Chem. Int. Ed.* **2014**, *53* (35), 9168–9172. <https://doi.org/10.1002/anie.201403580>.
- (10) Gutmiedl, K.; Wirges, C. T.; Ehmke, V.; Carell, T. Copper-Free “Click” Modification of DNA via Nitrile Oxide–Norborene 1,3-Dipolar Cycloaddition. *Org. Lett.* **2009**, *11* (11), 2405–2408. <https://doi.org/10.1021/ol9005322>.
- (11) Schulz, D.; Rentmeister, A. Current Approaches for RNA Labeling in Vitro and in Cells Based on Click Reactions. *ChemBioChem* **2014**, *15* (16), 2342–2347. <https://doi.org/10.1002/cbic.201402240>.
- (12) Wu, H.; Alexander, S. C.; Jin, S.; Devaraj, N. K. A Bioorthogonal Near-Infrared Fluorogenic Probe for mRNA Detection. *J. Am. Chem. Soc.* **2016**, *138* (36), 11429–11432. <https://doi.org/10.1021/jacs.6b01625>.
- (13) Neef, A. B.; Schultz, C. Selective Fluorescence Labeling of Lipids in Living Cells. *Angew. Chem. Int. Ed.* **2009**, *48* (8), 1498–1500. <https://doi.org/10.1002/anie.200805507>.
- (14) Niederwieser, A.; Späte, A.-K.; Nguyen, L. D.; Jüngst, C.; Reutter, W.; Wittmann, V. Two-Color Glycan Labeling of Live Cells by a Combination of Diels–Alder and Click Chemistry. *Angew. Chem. Int. Ed.* **2013**, *52* (15), 4265–4268. <https://doi.org/10.1002/anie.201208991>.
- (15) Zhou, Q.; Stefano, J. E.; Manning, C.; Kyazike, J.; Chen, B.; Gianolio, D. A.; Park, A.; Busch, M.; Bird, J.; Zheng, X.; et al. Site-Specific Antibody–Drug Conjugation through Glycoengineering. *Bioconjug. Chem.* **2014**, *25* (3), 510–520. <https://doi.org/10.1021/bc400505q>.
- (16) Robinson, P. V.; de Almeida-Escobedo, G.; de Groot, A. E.; McKechnie, J. L.; Bertozzi, C. R. Live-Cell Labeling of Specific Protein Glycoforms by Proximity-Enhanced Bioorthogonal Ligation. *J. Am. Chem. Soc.* **2015**, *137* (33), 10452–10455. <https://doi.org/10.1021/jacs.5b04279>.
- (17) Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W. Genetic Encoding of Bicyclononynes and Trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels–Alder Reactions. *J. Am. Chem. Soc.* **2012**, *134* (25), 10317–10320. <https://doi.org/10.1021/ja302832g>.
- (18) Peng, T.; Hang, H. C. Site-Specific Bioorthogonal Labeling for Fluorescence Imaging of Intracellular Proteins in Living Cells. *J. Am. Chem. Soc.* **2016**, *138* (43), 14423–14433. <https://doi.org/10.1021/jacs.6b08733>.

- (19) Machida, T.; Lang, K.; Xue, L.; Chin, J. W.; Winssinger, N. Site-Specific Glycoconjugation of Protein via Bioorthogonal Tetrazine Cycloaddition with a Genetically Encoded Trans-Cyclooctene or Bicyclononyne. *Bioconjug. Chem.* **2015**, *26* (5), 802–806. <https://doi.org/10.1021/acs.bioconjchem.5b00101>.
- (20) Smith, E. L.; Giddens, J. P.; Iavarone, A. T.; Godula, K.; Wang, L.-X.; Bertozzi, C. R. Chemoenzymatic Fc Glycosylation via Engineered Aldehyde Tags. *Bioconjug. Chem.* **2014**, *25* (4), 788–795. <https://doi.org/10.1021/bc500061s>.
- (21) Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L. Site-Selective Protein-Modification Chemistry for Basic Biology and Drug Development. *Nat. Chem.* **2016**, *8* (2), 103–113. <https://doi.org/10.1038/nchem.2393>.
- (22) Agarwal, P.; Bertozzi, C. R. Site-Specific Antibody–Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development. *Bioconjug. Chem.* **2015**, *26* (2), 176–192. <https://doi.org/10.1021/bc5004982>.
- (23) Zimmerman, E. S.; Heibeck, T. H.; Gill, A.; Li, X.; Murray, C. J.; Madlansacay, M. R.; Tran, C.; Uter, N. T.; Yin, G.; Rivers, P. J.; et al. Production of Site-Specific Antibody–Drug Conjugates Using Optimized Non-Natural Amino Acids in a Cell-Free Expression System. *Bioconjug. Chem.* **2014**, *25* (2), 351–361. <https://doi.org/10.1021/bc400490z>.
- (24) Zeglis, B. M.; Sevak, K. K.; Reiner, T.; Mohindra, P.; Carlin, S. D.; Zanzonico, P.; Weissleder, R.; Lewis, J. S. A Pretargeted PET Imaging Strategy Based on Bioorthogonal Diels–Alder Click Chemistry. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* **2013**, *54* (8), 1389–1396. <https://doi.org/10.2967/jnumed.112.115840>.
- (25) Pretze, M.; Wuest, F.; Peppel, T.; Köckerling, M.; Mamat, C. The Traceless Staudinger Ligation with Fluorine-18: A Novel and Versatile Labeling Technique for the Synthesis of PET-Radiotracers. *Tetrahedron Lett.* **2010**, *51* (49), 6410–6414. <https://doi.org/10.1016/j.tetlet.2010.09.134>.
- (26) Zeng, D.; Zeglis, B. M.; Lewis, J. S.; Anderson, C. J. The Growing Impact of Bioorthogonal Click Chemistry on the Development of Radiopharmaceuticals. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* **2013**, *54* (6), 829–832. <https://doi.org/10.2967/jnumed.112.115550>.
- (27) Denk, C.; Svatunek, D.; Filip, T.; Wanek, T.; Lumpi, D.; Fröhlich, J.; Kuntner, C.; Mikula, H. Development of a ¹⁸F-Labeled Tetrazine with Favorable Pharmacokinetics for Bioorthogonal PET Imaging. *Angew. Chem. Int. Ed.* **2014**, *53* (36), 9655–9659. <https://doi.org/10.1002/anie.201404277>.
- (28) Tsai, Y.-H.; Essig, S.; James, J. R.; Lang, K.; Chin, J. W. Selective, Rapid and Optically Switchable Regulation of Protein Function in Live Mammalian Cells. *Nat. Chem.* **2015**, *7* (7), 554–561. <https://doi.org/10.1038/nchem.2253>.
- (29) Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. Labeling Live Cells by Copper-Catalyzed Alkyne–Azide Click Chemistry. *Bioconjug. Chem.* **2010**, *21* (10), 1912–1916. <https://doi.org/10.1021/bc100272z>.
- (30) Uttamapinant, C.; Tangpeerachaikul, A.; Grecian, S.; Clarke, S.; Singh, U.; Slade, P.; Gee, K. R.; Ting, A. Y. Fast, Cell-Compatible Click Chemistry with Copper-Chelating Azides for Biomolecular Labeling. *Angew. Chem. Int. Ed.* **2012**, *51* (24), 5852–5856. <https://doi.org/10.1002/anie.201108181>.
- (31) Hest, J. C. M. van; Delft, F. L. van. Protein Modification by Strain-Promoted Alkyne–Azide Cycloaddition. *ChemBioChem* **2011**, *12* (9), 1309–1312. <https://doi.org/10.1002/cbic.201100206>.
- (32) Bruckman, M. A.; Kaur, G.; Lee, L. A.; Xie, F.; Sepulveda, J.; Breitenkamp, R.; Zhang, X.; Joralemon, M.; Russell, T. P.; Emrick, T.; et al. Surface Modification of Tobacco Mosaic Virus with “Click” Chemistry. *ChemBioChem* **2008**, *9* (4), 519–523. <https://doi.org/10.1002/cbic.200700559>.
- (33) Li, S.; Wang, L.; Yu, F.; Zhu, Z.; Shobaki, D.; Chen, H.; Wang, M.; Wang, J.; Qin, G.; J. Erasquin, U.; et al. Copper-Catalyzed Click Reaction on/in Live Cells. *Chem. Sci.* **2017**, *8* (3), 2107–2114. <https://doi.org/10.1039/C6SC02297A>.
- (34) Murrey, H. E.; Judkins, J. C.; am Ende, C. W.; Ballard, T. E.; Fang, Y.; Riccardi, K.; Di, L.; Guilmette, E. R.; Schwartz, J. W.; Fox, J. M.; et al. Systematic Evaluation of Bioorthogonal Reactions in Live Cells with Clickable HaloTag Ligands: Implications for Intracellular Imaging. *J. Am. Chem. Soc.* **2015**, *137* (35), 11461–11475. <https://doi.org/10.1021/jacs.5b06847>.

- (35) Selvaraj, R.; Fox, J. M. Trans-Cyclooctene—a Stable, Voracious Dienophile for Bioorthogonal Labeling. *Curr. Opin. Chem. Biol.* **2013**, *17* (5), 753–760. <https://doi.org/10.1016/j.cbpa.2013.07.031>.
- (36) Devaraj, N. K.; Weissleder, R.; Hilderbrand, S. A. Tetrazine-Based Cycloadditions: Application to Pretargeted Live Cell Imaging. *Bioconjug. Chem.* **2008**, *19* (12), 2297–2299. <https://doi.org/10.1021/bc8004446>.
- (37) Karver, M. R.; Weissleder, R.; Hilderbrand, S. A. Synthesis and Evaluation of a Series of 1,2,4,5-Tetrazines for Bioorthogonal Conjugation. *Bioconjug. Chem.* **2011**, *22* (11), 2263–2270. <https://doi.org/10.1021/bc200295y>.
- (38) Rossin, R.; Renart Verkerk, P.; van den Bosch, S. M.; Vulderson, R. C. M.; Verel, I.; Lub, J.; Robillard, M. S. In Vivo Chemistry for Pretargeted Tumor Imaging in Live Mice. *Angew. Chem. Int. Ed.* **2010**, *49* (19), 3375–3378. <https://doi.org/10.1002/anie.200906294>.
- (39) Patterson, D. M.; Nazarova, L. A.; Prescher, J. A. Finding the Right (Bioorthogonal) Chemistry. *ACS Chem. Biol.* **2014**, *9* (3), 592–605. <https://doi.org/10.1021/cb400828a>.
- (40) Sletten, E. M.; Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem. Int. Ed.* **2009**, *48* (38), 6974–6998. <https://doi.org/10.1002/anie.200900942>.
- (41) Lang, K.; Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* **2014**, *114* (9), 4764–4806. <https://doi.org/10.1021/cr400355w>.
- (42) Patterson, D. M.; Prescher, J. A. Orthogonal Bioorthogonal Chemistries. *Curr. Opin. Chem. Biol.* **2015**, *28*, 141–149. <https://doi.org/10.1016/j.cbpa.2015.07.006>.
- (43) Wang, K.; Sachdeva, A.; Cox, D. J.; Wilf, N. M.; Lang, K.; Wallace, S.; Mehl, R. A.; Chin, J. W. Optimized Orthogonal Translation of Unnatural Amino Acids Enables Spontaneous Protein Double-Labeling and FRET. *Nat. Chem.* **2014**, *6* (5), 393–403. <https://doi.org/10.1038/nchem.1919>.
- (44) Kumar Narayanam, M.; Liang, Y.; N. Houk, K.; M. Murphy, J. Discovery of New Mutually Orthogonal Bioorthogonal Cycloaddition Pairs through Computational Screening. *Chem. Sci.* **2016**, *7* (2), 1257–1261. <https://doi.org/10.1039/C5SC03259H>.
- (45) F. Debets, M.; Hest, J. C. M. van; T. Rutjes, F. P. J. Bioorthogonal Labelling of Biomolecules: New Functional Handles and Ligation Methods. *Org. Biomol. Chem.* **2013**, *11* (38), 6439–6455. <https://doi.org/10.1039/C3OB41329B>.
- (46) Saxon, E.; Bertozzi, C. R. Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **2000**, *287* (5460), 2007–2010. <https://doi.org/10.1126/science.287.5460.2007>.
- (47) Lin, F. L.; Hoyt, H. M.; van Halbeek, H.; Bergman, R. G.; Bertozzi, C. R. Mechanistic Investigation of the Staudinger Ligation. *J. Am. Chem. Soc.* **2005**, *127* (8), 2686–2695. <https://doi.org/10.1021/ja044461m>.
- (48) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001**, *40* (11), 2004–2021. [https://doi.org/10.1002/1521-3773\(20010601\)40:11<2004::AID-ANIE2004>3.0.CO;2-5](https://doi.org/10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5).
- (49) McKay, C. S.; Finn, M. G. Click Chemistry in Complex Mixtures: Bioorthogonal Bioconjugation. *Chem. Biol.* **2014**, *21* (9), 1075–1101. <https://doi.org/10.1016/j.chembiol.2014.09.002>.
- (50) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A Strain-Promoted [3 + 2] Azide–Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **2004**, *126* (46), 15046–15047. <https://doi.org/10.1021/ja044996f>.
- (51) Nikić, I.; Plass, T.; Schraidt, O.; Szymański, J.; Briggs, J. A. G.; Schultz, C.; Lemke, E. A. Minimal Tags for Rapid Dual-Color Live-Cell Labeling and Super-Resolution Microscopy. *Angew. Chem. Int. Ed.* **2014**, *53* (8), 2245–2249. <https://doi.org/10.1002/anie.201309847>.
- (52) Blackman, M. L.; Royzen, M.; Fox, J. M. Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels–Alder Reactivity. *J. Am. Chem. Soc.* **2008**, *130* (41), 13518–13519. <https://doi.org/10.1021/ja8053805>.
- (53) Knall, A.-C.; Slugovc, C. Inverse Electron Demand Diels–Alder (IEDDA)-Initiated Conjugation: A (High) Potential Click Chemistry Scheme. *Chem. Soc. Rev.* **2013**, *42* (12), 5131–5142. <https://doi.org/10.1039/C3CS60049A>.
- (54) Sauer, J.; Sustmann, R. Mechanistic Aspects of Diels–Alder Reactions: A Critical Survey. *Angew. Chem. Int. Ed. Engl.* **1980**, *19* (10), 779–807. <https://doi.org/10.1002/anie.198007791>.

- (55) Patterson, D. M.; Nazarova, L. A.; Xie, B.; Kamber, D. N.; Prescher, J. A. Functionalized Cyclopropenes As Bioorthogonal Chemical Reporters. *J. Am. Chem. Soc.* **2012**, *134* (45), 18638–18643. <https://doi.org/10.1021/ja3060436>.
- (56) Han, H.-S.; Devaraj, N. K.; Lee, J.; Hilderbrand, S. A.; Weissleder, R.; Bawendi, M. G. Development of a Bioorthogonal and Highly Efficient Conjugation Method for Quantum Dots Using Tetrazine–Norbornene Cycloaddition. *J. Am. Chem. Soc.* **2010**, *132* (23), 7838–7839. <https://doi.org/10.1021/ja101677r>.
- (57) Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. Genetically Encoded Norbornene Directs Site-Specific Cellular Protein Labelling via a Rapid Bioorthogonal Reaction. *Nat. Chem.* **2012**, *4* (4), 298–304. <https://doi.org/10.1038/nchem.1250>.
- (58) Ramil, C. P.; Dong, M.; An, P.; Lewandowski, T. M.; Yu, Z.; Miller, L. J.; Lin, Q. Spirohexene-Tetrazine Ligation Enables Bioorthogonal Labeling of Class B G Protein-Coupled Receptors in Live Cells. *J. Am. Chem. Soc.* **2017**, *139* (38), 13376–13386. <https://doi.org/10.1021/jacs.7b05674>.
- (59) Darko, A.; Wallace, S.; Dmitrenko, O.; M. Machovina, M.; A. Mehl, R.; W. Chin, J.; M. Fox, J. Conformationally Strained Trans -Cyclooctene with Improved Stability and Excellent Reactivity in Tetrazine Ligation. *Chem. Sci.* **2014**, *5* (10), 3770–3776. <https://doi.org/10.1039/C4SC01348D>.
- (60) Taylor, M. T.; Blackman, M. L.; Dmitrenko, O.; Fox, J. M. Design and Synthesis of Highly Reactive Dienophiles for the Tetrazine–Trans-Cyclooctene Ligation. *J. Am. Chem. Soc.* **2011**, *133* (25), 9646–9649. <https://doi.org/10.1021/ja201844c>.
- (61) Chen, W.; Wang, D.; Dai, C.; Hamelberg, D.; Wang, B. Clicking 1,2,4,5-Tetrazine and Cyclooctynes with Tunable Reaction Rates. *Chem. Commun.* **2012**, *48* (12), 1736–1738. <https://doi.org/10.1039/C2CC16716F>.
- (62) Liu, F.; Liang, Y.; Houk, K. N. Theoretical Elucidation of the Origins of Substituent and Strain Effects on the Rates of Diels–Alder Reactions of 1,2,4,5-Tetrazines. *J. Am. Chem. Soc.* **2014**, *136* (32), 11483–11493. <https://doi.org/10.1021/ja505569a>.
- (63) Kamber, D. N.; Liang, Y.; Blizzard, R. J.; Liu, F.; Mehl, R. A.; Houk, K. N.; Prescher, J. A. 1,2,4-Triazines Are Versatile Bioorthogonal Reagents. *J. Am. Chem. Soc.* **2015**, *137* (26), 8388–8391. <https://doi.org/10.1021/jacs.5b05100>.
- (64) Rossin, R.; van den Bosch, S. M.; ten Hoeve, W.; Carvelli, M.; Versteegen, R. M.; Lub, J.; Robillard, M. S. Highly Reactive Trans-Cyclooctene Tags with Improved Stability for Diels–Alder Chemistry in Living Systems. *Bioconjug. Chem.* **2013**, *24* (7), 1210–1217. <https://doi.org/10.1021/bc400153y>.
- (65) Fang, Y.; Zhang, H.; Huang, Z.; L. Scinto, S.; C. Yang, J.; Ende, C. W. am; Dmitrenko, O.; S. Johnson, D.; M. Fox, J. Photochemical Syntheses, Transformations, and Bioorthogonal Chemistry of Trans -Cycloheptene and Sila Trans -Cycloheptene Ag(i) Complexes. *Chem. Sci.* **2018**, *9* (7), 1953–1963. <https://doi.org/10.1039/C7SC04773H>.
- (66) Yang, J.; Liang, Y.; Šečková, J.; Houk, K. N.; Devaraj, N. K. Synthesis and Reactivity Comparisons of 1-Methyl-3-Substituted Cyclopropene Mini-Tags for Tetrazine Bioorthogonal Reactions. *Chem. – Eur. J.* **20** (12), 3365–3375. <https://doi.org/10.1002/chem.201304225>.
- (67) Schoch, J.; Wiessler, M.; Jäschke, A. Post-Synthetic Modification of DNA by Inverse-Electron-Demand Diels–Alder Reaction. *J. Am. Chem. Soc.* **2010**, *132* (26), 8846–8847. <https://doi.org/10.1021/ja102871p>.
- (68) Marie Holstein, J.; Schulz, D.; Rentmeister, A. Bioorthogonal Site-Specific Labeling of the 5'-Cap Structure in Eukaryotic MRNAs. *Chem. Commun.* **2014**, *50* (34), 4478–4481. <https://doi.org/10.1039/C4CC01549E>.
- (69) Laughlin, S. T.; Agard, N. J.; Baskin, J. M.; Carrico, I. S.; Chang, P. V.; Ganguli, A. S.; Hangauer, M. J.; Lo, A.; Prescher, J. A.; Bertozzi, C. R. Metabolic Labeling of Glycans with Azido Sugars for Visualization and Glycoproteomics. In *Methods in Enzymology; Glycobiology*; Academic Press, 2006; Vol. 415, pp 230–250. [https://doi.org/10.1016/S0076-6879\(06\)15015-6](https://doi.org/10.1016/S0076-6879(06)15015-6).
- (70) Laughlin, S. T.; Baskin, J. M.; Amacher, S. L.; Bertozzi, C. R. In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science* **2008**, *320* (5876), 664–667. <https://doi.org/10.1126/science.1155106>.
- (71) Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; et al. HaloTag: A Novel Protein Labeling

- Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **2008**, *3* (6), 373–382. <https://doi.org/10.1021/cb800025k>.
- (72) Rush, J. S.; Bertozzi, C. R. New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo. *J. Am. Chem. Soc.* **2008**, *130* (37), 12240–12241. <https://doi.org/10.1021/ja804530w>.
- (73) Cohen, G. N.; Munier, R. [Incorporation of structural analogues of amino acids in bacterial proteins]. *Biochim. Biophys. Acta* **1956**, *21* (3), 592–593.
- (74) Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. Residue-Specific Incorporation of Non-Canonical Amino Acids into Proteins: Recent Developments and Applications. *Curr. Opin. Chem. Biol.* **2010**, *14* (6), 774–780. <https://doi.org/10.1016/j.cbpa.2010.09.013>.
- (75) Beatty, K. E.; Liu, J. C.; Xie, F.; Dieterich, D. C.; Schuman, E. M.; Wang, Q.; Tirrell, D. A. Fluorescence Visualization of Newly Synthesized Proteins in Mammalian Cells. *Angew. Chem.* **2006**, *118* (44), 7524–7527. <https://doi.org/10.1002/ange.200602114>.
- (76) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of Escherichia Coli. *Science* **2001**, *292* (5516), 498–500. <https://doi.org/10.1126/science.1060077>.
- (77) Furter, R. Expansion of the Genetic Code: Site-Directed p-Fluoro-Phenylalanine Incorporation in Escherichia Coli. *Protein Sci.* **1998**, *7* (2), 419–426. <https://doi.org/10.1002/pro.5560070223>.
- (78) Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia Coli. *Biochemistry* **2013**, *52* (10), 1828–1837. <https://doi.org/10.1021/bi4000244>.
- (79) Xiao, H.; Chatterjee, A.; Choi, S.; Bajjuri, K. M.; Sinha, S. C.; Schultz, P. G. Genetic Incorporation of Multiple Unnatural Amino Acids into Proteins in Mammalian Cells. *Angew. Chem. Int. Ed.* **2013**, *52* (52), 14080–14083. <https://doi.org/10.1002/anie.201308137>.
- (80) Bohlke, N.; Budisa, N. Sense Codon Emancipation for Proteome-Wide Incorporation of Noncanonical Amino Acids: Rare Isoleucine Codon AUA as a Target for Genetic Code Expansion. *FEMS Microbiol. Lett.* **2014**, *351* (2), 133–144. <https://doi.org/10.1111/1574-6968.12371>.
- (81) Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-TRNA Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2014**, *1844* (6), 1059–1070. <https://doi.org/10.1016/j.bbapap.2014.03.002>.
- (82) Chin, J. W. Expanding and Reprogramming the Genetic Code. *Nature* **2017**, *550* (7674), 53–60. <https://doi.org/10.1038/nature24031>.
- (83) Dumas, A.; Lercher, L.; Spicer, C.; Davis, B. Designing Logical Codon Reassignment – Expanding the Chemistry in Biology. *Chem. Sci.* **2015**, *6* (1), 50–69. <https://doi.org/10.1039/C4SC01534G>.
- (84) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. An Expanded Eukaryotic Genetic Code. *Science* **2003**, *301* (5635), 964–967. <https://doi.org/10.1126/science.1084772>.
- (85) Bianco, A.; Townsley, F. M.; Greiss, S.; Lang, K.; Chin, J. W. Expanding the Genetic Code of *Drosophila Melanogaster*. *Nat. Chem. Biol.* **2012**, *8* (9), 748–750. <https://doi.org/10.1038/nchembio.1043>.
- (86) Ernst, R. J.; Krogager, T. P.; Maywood, E. S.; Zanchi, R.; Beránek, V.; Elliott, T. S.; Barry, N. P.; Hastings, M. H.; Chin, J. W. Genetic Code Expansion in the Mouse Brain. *Nat. Chem. Biol.* **2016**, *12* (10), 776–778. <https://doi.org/10.1038/nchembio.2160>.
- (87) Rauch, B. J.; Porter, J. J.; Mehl, R. A.; Perona, J. J. Improved Incorporation of Noncanonical Amino Acids by an Engineered TRNATyr Suppressor. *Biochemistry* **2016**, *55* (3), 618–628. <https://doi.org/10.1021/acs.biochem.5b01185>.
- (88) Oza, J. P.; Aerni, H. R.; Pirman, N. L.; Barber, K. W.; Haar, C. M. ter; Rogulina, S.; Amrofell, M. B.; Isaacs, F. J.; Rinehart, J.; Jewett, M. C. Robust Production of Recombinant Phosphoproteins Using Cell-Free Protein Synthesis. *Nat. Commun.* **2015**, *6*, 8168. <https://doi.org/10.1038/ncomms9168>.
- (89) Elsässer, S. J.; Ernst, R. J.; Walker, O. S.; Chin, J. W. Genetic Code Expansion in Stable Cell Lines Enables Encoded Chromatin Modification. *Nat. Methods* **2016**, *13* (2), 158–164. <https://doi.org/10.1038/nmeth.3701>.

- (90) Chou, C.; Uprety, R.; Davis, L.; Chin, J. W.; Deiters, A. Genetically Encoding an Aliphatic Diazirine for Protein Photocrosslinking. *Chem. Sci.* **2011**, *2* (3), 480–483. <https://doi.org/10.1039/C0SC00373E>.
- (91) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a Photocrosslinking Amino Acid to the Genetic Code of Escherichia Coli. *Proc. Natl. Acad. Sci.* **2002**, *99* (17), 11020–11024. <https://doi.org/10.1073/pnas.172226299>.
- (92) Lueck, J. D.; Mackey, A. L.; Infield, D. T.; Galpin, J. D.; Li, J.; Roux, B.; Ahern, C. A. Atomic Mutagenesis in Ion Channels with Engineered Stoichiometry. *eLife* **5**. <https://doi.org/10.7554/eLife.18976>.
- (93) Mukai, T.; Hayashi, A.; Iraha, F.; Sato, A.; Ohtake, K.; Yokoyama, S.; Sakamoto, K. Codon Reassignment in the Escherichia Coli Genetic Code. *Nucleic Acids Res.* **2010**, *38* (22), 8188–8195. <https://doi.org/10.1093/nar/gkq707>.
- (94) Willis, J. C. W.; Chin, J. W. Mutually Orthogonal Pyrrolysyl-TRNA Synthetase/TRNA Pairs. *Nat. Chem.* **2018**, *1*. <https://doi.org/10.1038/s41557-018-0052-5>.
- (95) Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. Adding L-Lysine Derivatives to the Genetic Code of Mammalian Cells with Engineered Pyrrolysyl-TRNA Synthetases. *Biochem. Biophys. Res. Commun.* **2008**, *371* (4), 818–822. <https://doi.org/10.1016/j.bbrc.2008.04.164>.
- (96) Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Söll, D.; Steitz, T. A. Structure of Pyrrolysyl-TRNA Synthetase, an Archaeal Enzyme for Genetic Code Innovation. *Proc. Natl. Acad. Sci.* **2007**, *104* (27), 11268–11273. <https://doi.org/10.1073/pnas.0704769104>.
- (97) Italia, J. S.; Latour, C.; Wrobel, C. J. J.; Chatterjee, A. Resurrecting the Bacterial Tyrosyl-TRNA Synthetase/TRNA Pair for Expanding the Genetic Code of Both E. Coli and Eukaryotes. *Cell Chem. Biol.* **2018**, *0* (0). <https://doi.org/10.1016/j.chembiol.2018.07.002>.
- (98) Italia, J. S.; Addy, P. S.; Wrobel, C. J. J.; Crawford, L. A.; Lajoie, M. J.; Zheng, Y.; Chatterjee, A. An Orthogonalized Platform for Genetic Code Expansion in Both Bacteria and Eukaryotes. *Nat. Chem. Biol.* **2017**, *13* (4), 446–450. <https://doi.org/10.1038/nchembio.2312>.
- (99) Kazlauskas, R. J.; Bornscheuer, U. T. Finding Better Protein Engineering Strategies. *Nat. Chem. Biol.* **2009**, *5*, 526–529. <https://doi.org/10.1038/nchembio0809-526>.
- (100) Acevedo-Rocha, C. G.; Reetz, M. T.; Nov, Y. Economical Analysis of Saturation Mutagenesis Experiments. *Sci. Rep.* **2015**, *5*, 10654. <https://doi.org/10.1038/srep10654>.
- (101) Wang, Y.-S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R. A Rationally Designed Pyrrolysyl-TRNA Synthetase Mutant with a Broad Substrate Spectrum. *J. Am. Chem. Soc.* **2012**, *134* (6), 2950–2953. <https://doi.org/10.1021/ja211972x>.
- (102) Wang, L.; Xie, J.; Schultz, P. G. Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35* (1), 225–249. <https://doi.org/10.1146/annurev.biophys.35.101105.121507>.
- (103) Cooley, R. B.; Feldman, J. L.; Driggers, C. M.; Bundy, T. A.; Stokes, A. L.; Karplus, P. A.; Mehl, R. A. Structural Basis of Improved Second-Generation 3-Nitro-Tyrosine TRNA Synthetases. *Biochemistry* **2014**, *53* (12), 1916–1924. <https://doi.org/10.1021/bi5001239>.
- (104) Link, A. J.; Vink, M. K. S.; Agard, N. J.; Prescher, J. A.; Bertozzi, C. R.; Tirrell, D. A. Discovery of Aminoacyl-TRNA Synthetase Activity through Cell-Surface Display of Noncanonical Amino Acids. *Proc. Natl. Acad. Sci.* **2006**, *103* (27), 10180–10185. <https://doi.org/10.1073/pnas.0601167103>.
- (105) Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing : Nature Methods : Nature Research <https://www.nature.com/nmeth/journal/vaop/ncurrent/full/nmeth.4302.html> (accessed Jun 6, 2017).
- (106) Santoro, S. W.; Schultz, P. G. Directed Evolution of the Substrate Specificities of a Site-Specific Recombinase and an Aminoacyl-TRNA Synthetase Using Fluorescence-Activated Cell Sorting (FACS). In *Directed Enzyme Evolution: Screening and Selection Methods*; Arnold, F. H., Georgiou, G., Eds.; Methods in Molecular Biology™; Humana Press: Totowa, NJ, 2003; pp 291–312. <https://doi.org/10.1385/1-59259-396-8:291>.
- (107) Ellefson, J. W.; Meyer, A. J.; Hughes, R. A.; Cannon, J. R.; Brodbelt, J. S.; Ellington, A. D. Directed Evolution of Genetic Parts and Circuits by Compartmentalized Partnered Replication. *Nat. Biotechnol.* **2014**, *32* (1), 97–101. <https://doi.org/10.1038/nbt.2714>.

- (108) Bryson, D. I.; Fan, C.; Guo, L.-T.; Miller, C.; Söll, D.; Liu, D. R. Continuous Directed Evolution of Aminoacyl-TRNA Synthetases. *Nat. Chem. Biol.* **2017**, *13* (12), 1253–1260. <https://doi.org/10.1038/nchembio.2474>.
- (109) Rossin, R.; Robillard, M. S. Pretargeted Imaging Using Bioorthogonal Chemistry in Mice. *Curr. Opin. Chem. Biol.* **2014**, *21*, 161–169. <https://doi.org/10.1016/j.cbpa.2014.07.023>.
- (110) Chang, P. V.; Prescher, J. A.; Sletten, E. M.; Baskin, J. M.; Miller, I. A.; Agard, N. J.; Lo, A.; Bertozzi, C. R. Copper-Free Click Chemistry in Living Animals. *Proc. Natl. Acad. Sci.* **2010**. <https://doi.org/10.1073/pnas.0911116107>.
- (111) Seitchik, J. L.; Peeler, J. C.; Taylor, M. T.; Blackman, M. L.; Rhoads, T. W.; Cooley, R. B.; Refakis, C.; Fox, J. M.; Mehl, R. A. Genetically Encoded Tetrazine Amino Acid Directs Rapid Site-Specific in Vivo Bioorthogonal Ligation with Trans-Cyclooctenes. *J. Am. Chem. Soc.* **2012**, *134* (6), 2898–2901. <https://doi.org/10.1021/ja2109745>.
- (112) Fox, J. M.; Robillard, M. S. Editorial Overview: In Vivo Chemistry: Pushing the Envelope. *Curr. Opin. Chem. Biol.* **2014**, *21*, v–vii. <https://doi.org/10.1016/j.cbpa.2014.07.027>.
- (113) Kobayashi, H.; Choyke, P. L. Target-Cancer-Cell-Specific Activatable Fluorescence Imaging Probes: Rational Design and in Vivo Applications. *Acc. Chem. Res.* **2011**, *44* (2), 83–90. <https://doi.org/10.1021/ar1000633>.
- (114) van der Wel, G. K.; Wijnen, J. W.; Engberts, J. B. F. N. Solvent Effects on a Diels–Alder Reaction Involving a Cationic Diene: Consequences of the Absence of Hydrogen-Bond Interactions for Accelerations in Aqueous Media. *J. Org. Chem.* **1996**, *61* (25), 9001–9005. <https://doi.org/10.1021/jo9614248>.
- (115) Yang, J.; Karver, M. R.; Li, W.; Sahu, S.; Devaraj, N. K. Metal-Catalyzed One-Pot Synthesis of Tetrazines Directly from Aliphatic Nitriles and Hydrazine. *Angew. Chem. Int. Ed.* **2012**, *51* (21), 5222–5225. <https://doi.org/10.1002/anie.201201117>.
- (116) Xie, J.; Schultz, P. G. An Expanding Genetic Code. *Methods* **2005**, *36* (3), 227–238. <https://doi.org/10.1016/j.ymeth.2005.04.010>.
- (117) Miyake-Stoner, S. J.; Refakis, C. A.; Hammill, J. T.; Lusic, H.; Hazen, J. L.; Deiters, A.; Mehl, R. A. Generating Permissive Site-Specific Unnatural Aminoacyl-TRNA Synthetases. *Biochemistry* **2010**, *49* (8), 1667–1677. <https://doi.org/10.1021/bi901947r>.
- (118) Stokes, A. L.; Miyake-Stoner, S. J.; Peeler, J. C.; Nguyen, D. P.; Hammer, R. P.; Mehl, R. A. Enhancing the Utility of Unnatural Amino Acidsynthetases by Manipulating Broad Substrate Specificity. *Mol. Biosyst.* **2009**, *5* (9), 1032–1038. <https://doi.org/10.1039/B904032C>.
- (119) Miyake-Stoner, S. J.; Miller, A. M.; Hammill, J. T.; Peeler, J. C.; Hess, K. R.; Mehl, R. A.; Brewer, S. H. Probing Protein Folding Using Site-Specifically Encoded Unnatural Amino Acids as FRET Donors with Tryptophan. *Biochemistry* **2009**, *48* (25), 5953–5962. <https://doi.org/10.1021/bi900426d>.
- (120) Grimm, J. B.; English, B. P.; Chen, J.; Slaughter, J. P.; Zhang, Z.; Revyakin, A.; Patel, R.; Macklin, J. J.; Normanno, D.; Singer, R. H.; et al. A General Method to Improve Fluorophores for Live-Cell and Single-Molecule Microscopy. *Nat. Methods* **2015**, *12* (3), 244–250. <https://doi.org/10.1038/nmeth.3256>.
- (121) Schmied, W. H.; Elsässer, S. J.; Uttamapinant, C.; Chin, J. W. Efficient Multisite Unnatural Amino Acid Incorporation in Mammalian Cells via Optimized Pyrrolysyl TRNA Synthetase/TRNA Expression and Engineered ERF1. *J. Am. Chem. Soc.* **2014**, *136* (44), 15577–15583. <https://doi.org/10.1021/ja5069728>.
- (122) Zhang, G.; Li, J.; Xie, R.; Fan, X.; Liu, Y.; Zheng, S.; Ge, Y.; Chen, P. R. Bioorthogonal Chemical Activation of Kinases in Living Systems. *ACS Cent. Sci.* **2016**, *2* (5), 325–331. <https://doi.org/10.1021/acscentsci.6b00024>.
- (123) Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W. Genetic Encoding of Bicyclononynes and Trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels–Alder Reactions. *J. Am. Chem. Soc.* **2012**, *134* (25), 10317–10320. <https://doi.org/10.1021/ja302832g>.
- (124) Chang, Y.-F.; Imam, J. S.; Wilkinson, M. F. The Nonsense-Mediated Decay RNA Surveillance Pathway. *Annu. Rev. Biochem.* **2007**, *76* (1), 51–74. <https://doi.org/10.1146/annurev.biochem.76.050106.093909>.
- (125) Nikić, I.; Estrada Girona, G.; Kang, J. H.; Paci, G.; Mikhaleva, S.; Koehler, C.; Shymanska, N. V.; Ventura Santos, C.; Spitz, D.; Lemke, E. A. Debugging Eukaryotic Genetic Code Expansion

- for Site-Specific Click-PAINT Super-Resolution Microscopy. *Angew. Chem. Int. Ed.* **55** (52), 16172–16176. <https://doi.org/10.1002/anie.201608284>.
- (126) Aloush, N.; Schwartz, T.; König, A.; Cohen, S.; Tam, B.; Akabayov, B.; Nachmias, D.; Ben-David, O.; Elia, N.; Arbely, E. Reducing Pyrrolysine tRNA Copy Number Improves Live Cell Imaging of Bioorthogonally Labeled Proteins. *bioRxiv* **2018**, 161984. <https://doi.org/10.1101/161984>.
- (127) Cunningham, C. W.; Mukhopadhyay, A.; Lushington, G. H.; Blagg, B. S. J.; Prisinzano, T. E.; Krise, J. P. Uptake, Distribution and Diffusivity of Reactive Fluorophores in Cells: Implications toward Target Identification. *Mol. Pharm.* **2010**, *7* (4), 1301–1310. <https://doi.org/10.1021/mp100089k>.
- (128) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Crystallographic Studies on Multiple Conformational States of Active-Site Loops in Pyrrolysyl-tRNA Synthetase. *J. Mol. Biol.* **2008**, *378* (3), 634–652. <https://doi.org/10.1016/j.jmb.2008.02.045>.
- (129) Chatterjee, A.; Xiao, H.; Bollong, M.; Ai, H.-W.; Schultz, P. G. Efficient Viral Delivery System for Unnatural Amino Acid Mutagenesis in Mammalian Cells. *Proc. Natl. Acad. Sci.* **2013**, *110* (29), 11803–11808. <https://doi.org/10.1073/pnas.1309584110>.
- (130) Aaronson, D. S.; Horvath, C. M. A Road Map for Those Who Don't Know JAK-STAT. *Science* **2002**, *296* (5573), 1653–1655. <https://doi.org/10.1126/science.1071545>.
- (131) Spencer, D. M.; Wandless, T. J.; Schreiber, S. L.; Crabtree, G. R. Controlling Signal Transduction with Synthetic Ligands. *Science* **1993**, *262* (5136), 1019–1024. <https://doi.org/10.1126/science.7694365>.
- (132) Zhang, H.; Trout, W. S.; Liu, S.; Andrade, G. A.; Hudson, D. A.; Scinto, S. L.; Dicker, K. T.; Li, Y.; Lazouski, N.; Rosenthal, J.; et al. Rapid Bioorthogonal Chemistry Turn-on through Enzymatic or Long Wavelength Photocatalytic Activation of Tetrazine Ligation. *J. Am. Chem. Soc.* **2016**, *138* (18), 5978–5983. <https://doi.org/10.1021/jacs.6b02168>.
- (133) Kemp, M.; Go, Y.-M.; Jones, D. P. Nonequilibrium Thermodynamics of Thiol/Disulfide Redox Systems: A Perspective on Redox Systems Biology. *Free Radic. Biol. Med.* **2008**, *44* (6), 921–937. <https://doi.org/10.1016/j.freeradbiomed.2007.11.008>.
- (134) Mao, X.; Ren, Z.; Parker, G. N.; Sondermann, H.; Pastorello, M. A.; Wang, W.; McMurray, J. S.; Demeler, B.; Darnell, J. E.; Chen, X. Structural Bases of Unphosphorylated STAT1 Association and Receptor Binding. *Mol. Cell* **2005**, *17* (6), 761–771. <https://doi.org/10.1016/j.molcel.2005.02.021>.
- (135) Zhang, Z.; Wang, L.; Brock, A.; Schultz, P. G. The Selective Incorporation of Alkenes into Proteins in Escherichia Coli. *Angew. Chem. Int. Ed.* **2002**, *41* (15), 2840–2842. [https://doi.org/10.1002/1521-3773\(20020802\)41:15<2840::AID-ANIE2840>3.0.CO;2-#](https://doi.org/10.1002/1521-3773(20020802)41:15<2840::AID-ANIE2840>3.0.CO;2-#).