ABSTRACT

Title of Document: THE DEVELOPMENT OF NEW TOOLS FOR THE INVESTIGATION OF PROTEIN FUNCTION USING PHOTO-REACTIVE UNNATURAL AMINO ACIDS.

Bryan Jason Wilkins, Doctor of Philosophy, 2010

Directed By: Dr. T. Ashton Cropp, Department of Chemistry and Biochemistry

Reported here is the direct synthesis and application of unnatural amino acids for the development of exploratory tools for protein studies. This work takes advantage of an expanded genetic code to extract a more precise chemical understanding of protein function with novel additions to the unnatural amino acid catalogue, as well as the expansion of techniques with previously developed compounds. The photochemical crosslinker, [D_{11}]-*p*-benzoylphenylalanine (pBpa), is synthesized for isotopic labeling in proteins. When [D_{11}]-pBpa is co-incorporated into protein with [D_0]-pBpa it is a mass spectral tool for rapid and conclusive identification of crosslinked fragments. Following enzymatic digestion the fingerprint of M, M+ 11 is readily identified allowing for rapid peak identification and the determined site of crosslink formation with single amino acid accuracy. In a means to extract a level of spatiotemporal control over fluorescent labeling of protein,

the photo-protected unnatural amino acid, *o*-nitrobenzyl cysteine (ONBC), is introduced to a small amino acid tag sequence CCPGCC. This tag is required and specifically binds the pro-fluorescent compound 5-bis(1,3,2-dithiasolan-2-yl)fluorescein (FlAsH). This work takes advantage of the inability of FlAsH to bind the cysteine-tag motif in the presence of an ONBC mutation. The photo-protected amino acid is deprotected with light, affording natural cysteine and the successful binding of FlasH to the tetracysteine tag only following ultraviolet irradiation. Finally, fluorinated tyrosine derivatives are synthetically modified to contain photo-protecting groups, which act as a disguise during unnatural amino acid mutagenesis techniques. Fluorinated tyrosines are recognized by endogenous tyrosyl-tRNA synthetases and incorporated globally throughout a protein at tyrosine positions. To circumvent this problem the *o*-nitrobenzyl photo-protecting group is installed on the tyrosine derivatives 2-fluorotyrosine, 3-fluorotyrosine, and 2,6-difluorotyrosine. The directed evolution of an orthogonal amber-tRNA synthetase, specific for these unnatural amino acids, is performed, providing the translational machinery for site-specific incorporation of these compounds. Following expression of protein with the protected tyrosine derivatives, protein exposed to ultraviolet irradiation subsequently loses the protecting group affording the site-specific incorporation of fluorinated tyrosine. Fluorinated tyrosines are introduced to the critical trysoine residue in the chromophore of super-folder green fluorescent protein to determine how the altered pKa affects its fluorescent properties.

THE DEVELOPMENT OF NEW TOOLS FOR THE INVESTIGATION OF PROTEIN FUNCTION USING PHOTO-REACTIVE UNNATURAL AMINO ACIDS.

By

Bryan Jason Wilkins

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Assistant Professor T. Ashton Cropp, Chair
Professor Catherine Fenselau
Assistant Professor Barbara Gerratana
Associate Professor Jason Kahn
Associate Professor Eric Haag, Dean's Representative

Dedication

This work is first dedicated to William Earl Schuler and Leslie William Wilkins.

Your memories serve to remind me of who I aspire to be. You are missed.

Mia-Vasiliki Wilkins, never give up on your dreams and always strive to be exactly the person you want to be.

Acknowledgements

The accomplishments over the past five years would not have been possible without the help of numerous individuals. I entered this chapter of my life naively, without the full realization of the sacrifices that it would entail. Without the support of my family and friends, as well as the everyday contacts with coworkers and mentors, the final goal would have never been achieved.

I would like to first thank Dr. T. Ashton Cropp who has served as a remarkable advisor. Your instruction and support has guided me through this work. I am especially thankful for your continued patience with my schedule as I have undoubtedly held atypical hours for several years. Your understanding has helped me maintain a more comfortable home life. Your relaxed nature has made it possible for me to find my way through a very stressful number of years. I am extremely thankful for your introduction to chemical synthesis and helping me realize that my true interests lie in chemical biology. Without your mentorship I would not have grown as a researcher and a teacher.

I would like to thank my committee members for all their support and advice. I am grateful for all the letters of recommendation for which you have written on my behalf. Your expertise and participation in each of my meetings and exams have helped me to become a better researcher. I would like to thank Dr. Gerratana for her help with independent proposal and the kindness of sharing chemicals and enzymes throughout the years. Dr. Kahn and Dr. Fenselau, thank you for agreeing to take the time out of your busy schedules to help guide me and keep me focused on the tasks at hand. I would also like to thank Dr. Haag for agreeing to sit on my committee as the Dean's Representative. Dr. Lyle Isaacs, although you have not served on my committee, you have been a tremendous teaching mentor. Thank you for allowing me the opportunity to lecture for you and help me grow as a teacher.

Thank you to all those who have worked with me in the Cropp lab. Kelly Daggett, I am especially thankful to you for helping me stay sane throughout this entire process. I am glad I was not the only first-time graduate student in a new lab and that I could share the experience with someone else. Your input on so many

subjects has helped tremendously. I also want to thank you for being so diligent with the lab appearance for so many years. If you weren't there to keep things semi-organized we would never be able to find anything. Good luck at Rockefeller. Thanks to Sam Marionni for all your help on various projects, but especially with the synthesis of the fluorotyrosines. Good luck in your own graduate career. Jia Liu, Zijun Chen and Karina Herrear, thanks for helping with all sorts of different aspects of daily lab life. I wish you all the best of luck with the rest of your graduate studies.

Thanks to all my fellow graduate students throughout the years. A special thanks to those of the Gerratana lab and Julin Lab, from which I have borrowed supplies numerous times. Thanks to Will Shadrick for being the fix-it man, and always having a solution when any apparatus may have been broken.

Paskalini, my thanks to you goes beyond words. Without you, and your patience and love, I would have never made it to this point. You have been my most precious friend and companion and I truly am "the luckiest". Thank you to my daughter for being so much fun and reminding me that it cannot always be work and no play. Barrett, I thank you for supporting me when no one else would. You are the definition of a true friend.

Finally I would like to thank my family for supporting me throughout my life. Mom and Dad, you have been generous and loving and always accepted me for who I was. Thanks for putting up with me and all my crazy adventures when I was young. Gram, I wouldn't be anywhere today if it was not for you. Haley and Joe, thanks for all your hospitality and leaving an open door invitation to your couch for so many years. You guys are the best. Elise thanks for being such a great niece and Adyson, please stay adorable. Brad, Cliff, Audrey, and Nana, thank you each for everything throughout the years. I love all of you and would not be the person I am today without any of you.

To all those not mentioned, and there are so many, thank you!

Table of Contents

List of Tables

List of Figures

List of Schemes

Abbreviations

DNA	deoxyribonucleic acid
RNA	ribonucleic acid
tRNA	transfer RNA
ATP	adenosine triphosphate
ONB	*ortho*-nitrobenzyl
ONBC	*ortho*-nitrobenzyl cysteine
ONBY	*ortho*-nitrobenzyl tyrosine
ONB-2FY	*ortho*-nitrobenzyl-2-fluorotyrosine
ONB-3FY	*ortho*-nitrobenzyl-3-fluorotyrosine
ONB-3,5-diFY	*ortho*-nitrobenzyl-3,5-difluorotyrosine
ONB-2,3,5,6-tetraFY	*ortho*-nitrobenzyl-2,3,5,6-tetrafluorotyrosine
ONB-YRS	*ortho*-nitrobenzyl tyrosine tRNA synthetase
ONB-2FY-RS	*ortho*-nitrobenzyl-2-fluorotyrosine tRNA synthetase
aaRS	aminoacyl-tRNA synthetase
UAA	unnatural amino acid
NAA	near-natural amino acid
EDTA	ethylenediaminetetraacetic acid
PCR	polymerase chain reaction
GFP	green fluorescent protein
PA-GFP	photo-activatable green fluorescent protein
sfGFP	super folder green fluorescent protein
GST	glutathione *S*-transferase

kDa	kilodaltons
LB	Luria-Bertani broth
MALDI-TOF	matrix assisted laser desorption ionization-time of flight
SDS-PAGE	sodium dodecyl sufate-polyacrylamide gel electrophoresis
UV	ultra violet
CAT	chloramphenicol acetyl transferase
FRET	fluorescence resonance energy transfer
pBpa	*para*-benzoylphenylalanine
FlAsH	5-bis(1,3,2-dithiasolan-2-yl)fluorescein
ReAsH	4,5-bis(1,3,2-dithiarsolan-2-yl)resorufin
EDT	1,2-ethanedithiol
TC-Tag	tetracysteine tag
WT	wild-type
NMR	nuclear magnetic resonance
GMML	glycerol minimal media + leucine
Amp	ampicillin
Kan	kanamycin
Chl	chloramphenicol
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
TAPS	N-tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid
ONPG	*ortho*-nitrophenyl-β-galactoside
IPTG	isopropyl β-D-1-thiogalactopyranoside

Chapter 1: Introduction

Understanding the structure and function of proteins is one of the most heavily studied disciplines of biochemistry. Biological systems rely on an infrastructure of vast assemblies of complex networks that operate on a multitude of levels. At the base of this scaffolding lies an atomic network where simple molecules, such as amino acids and nucleotides, act as the monomeric subunits of larger macromolecules, including proteins and DNA, respectively. In the case of proteins, the arrangement of amino acids and their atomic interactions within the configuration dictate folding patterns and overall structure and function. A single amino acid mutation, deletion, or repeat can alter conformation, causing structural defects that modify or destroy the native function of the protein. Where proteins are only one facet of a large biological realm, biochemistry attempts to explain the underlying basis of biomolecular assemblies, exposing their structure and function through chemical mechanisms that ultimately fashion all living organisms.

Chemical biology is a specialized discipline that sits at the interface of chemistry and biology, approaching biological issues from a chemical prospective. While this may sound synonymous with biochemistry, it differs in the fact that chemical synthesis is often employed for the design and implementation of small molecules to manipulate a system to help elucidate its biological importance. This field revolves around the development of techniques that uniquely address questions in biology, where each new small molecule, or probe, is an addition to a molecular "toolbox" for biochemical exploration. This work spans several disciplines including chemistry, biochemistry, and

molecular biology, focusing on the application of synthesized molecules for the perturbation of chemical environments within protein. While the scope of chemical biology spans a vast assortment of chemical explorations, this dissertation will focus on the synthesis and application of unnatural amino acids as a defined chemical approach with which to investigate protein function.

1.1 Unnatural amino acids

The genetic code of all living organisms is a template by which protein assembly is based. This sequenced code of nucleotides in RNA is translated through ribosomal machinery yielding a polymeric peptide chain. The construction of all protein is carried out using the same 20 naturally occurring amino acids, with the addition of selenocysteine[3] and pyrrolysine[4], considered to be the 21st and 22nd natural amino acids, respectively. While evolution has dictated that this small set of building blocks is sufficient to maintain life, the side chain chemistries have a finite number of functionalities. In many instances, the chemistries introduced are not adequate to carry out their natural intended function, dependent upon post-translational modifications, providing irregular amino acids[5-10]. This limitation is more obvious for biochemical assays, where the conventional approach to probing protein function is based on site-directed mutagenesis. Restrictions are placed on the available functionality by the limited number of amino acids provided in nature.

Chemistry allows for an almost infinite number of chemically synthesized amino acids that contain unnatural side chains. The direct perturbation of specific amino acids aids in the elucidation of structure-function relationships, as well as in the generation of proteins with novel properties. Non-native chemistry, introduced at specific residues,

greatly augments the ability to focus on specific biological questions. The implication that any non-native functional group can be generated and in turn engineered into protein, provides the basis for the development of strategies which genetically incorporate these compounds into an expanded genetic code.

1.1.1 Current applications of non-natural amino acids in biological studies

While there have been several approaches developed to incorporate unnatural amino acids (UAA) into protein, this section will focus primarily on the successes and advances toward *in vivo* incorporation of UAAs site-specifically through suppressor tRNA machinery (methodology will be discussed in further sections). Greater than forty UAAs have been site-specifically incorporated at the genetic level of *Escherichia coli* (*E. coli*), *Saccharomyces cerevisiae* (yeast), and mammalian cells, introducing a myriad of new functionality into proteins.

Nature does not encode for amino acids which contain highly reactive side chains under mild or neutral conditions. This, of course, is inherent in the processes of cellular activity, as most reactions are controlled under strict general acid/based catalyzed mechanisms. Functional groups such as ketones and aldehydes are not prevalent as their chemistries would prove over-reactive and fatal to biological function. This lack of reactivity in nature creates a limit on the biosynthetic additions that can be carried out due to the unavailability of chemical handles. UAAs with chemically reactive side chains have been introduced to circumvent this limitation. Examples of ketone moieties having been genetically introduced are *p*-acetylphenylalanine[11] (**1-1**), *m*-acetylphenylalanine[11] (**1-2**), and *p*-(3-oxobutanoyl)-L-phenylalanine[12] (**1-3**). These new chemical environments allow for the specific labeling of proteins with numerous reagents under mild conditions,

and without the need for protecting native functional groups. Specific reactions have shown the selectivity of the keto and diketo groups to react with both hydraazides and alkoxyamines, and the ability to label these groups with a number of fluorophores[11] and tags, including the direct addition of polyethylene glycol (PEG)[13]. These specific examples help illustrate how the introduction of novel chemistry allows for specificity along a peptide chain for a desired chemical reaction and the ability to site-specifically modify residues of interest.

Two UAAs have particular applications in the study of protein-protein interactions, where their side chain chemistries allow for the ability to chemically crosslink protein at interfaces in spatial proximity to the residue. The UAA *p*-azido-phenylalanine (**1-4**) contains an aryl-azide functional group, which is a photo-activated crosslinker[14]. When irradiated with light at ~ 310 nm a dehydroazepine is formed which is readily reactive with adjacent amines.

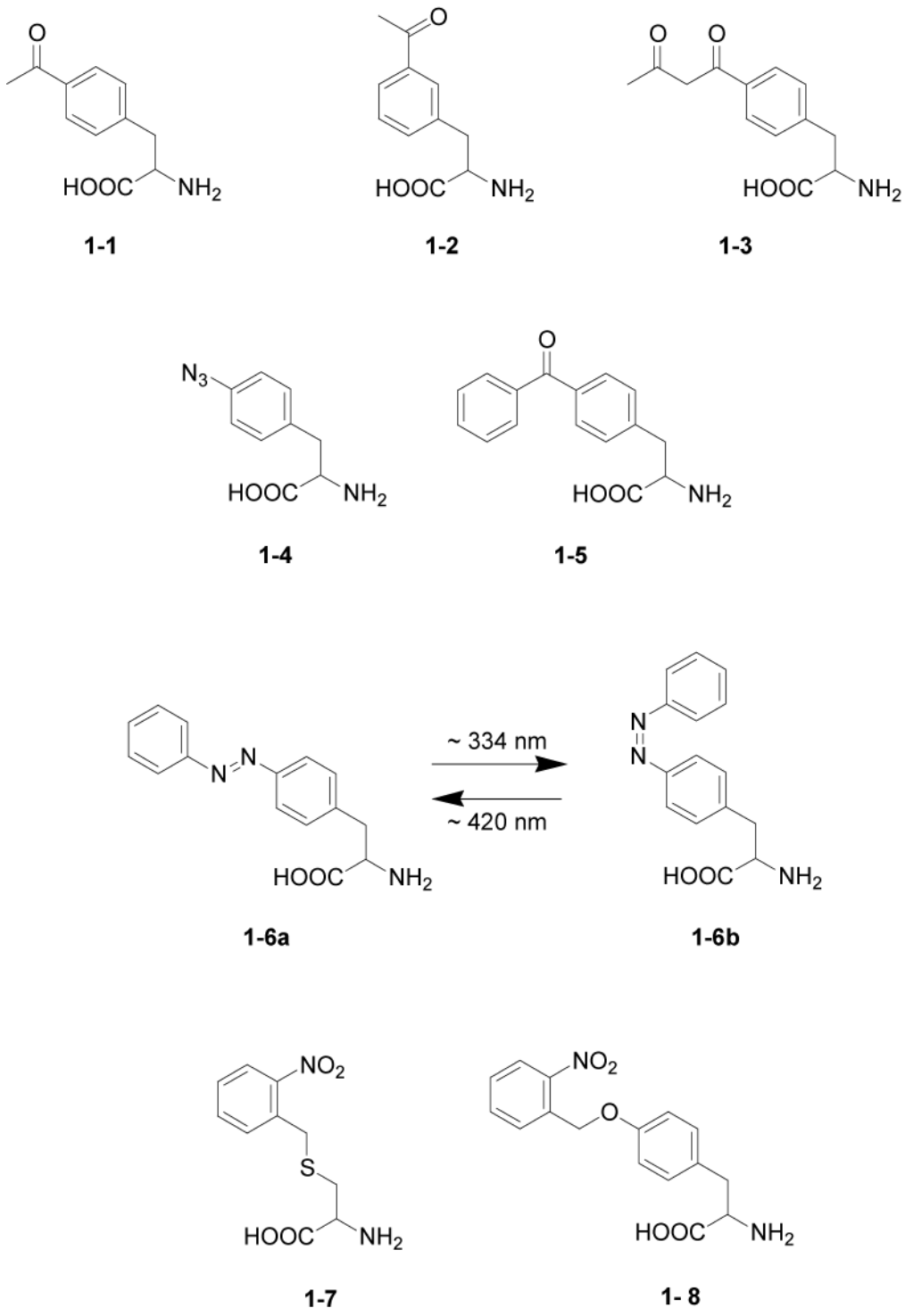


Figure 1-1. Chemical structures of the unnatural amino acids represented in Chapter 1.

Of particular interest to this work is the photo-crosslinking agent *p*-benzoylphenylalanine[15, 16] (**1-5**). The side chain of this residue contains a benzophenone

group which is activated with light at ~360 nm and reacts irreversibly with adjacent carbon-hydrogen bonds (the chemistry and details to be outlined in Chapter 2). Both of these UAAs have the ability to act as powerful *in vitro* and *in vivo* probes for potential protein-protein interactions. *p*-Benzoylphenylalanine has recently been utilized to capture important ligand binding events in G-protein coupled receptors in yeast[17]. Another photo-activated UAA is *p*-azophenyl-phenylalanine (**1-6a/b**) which can be light induced to toggle between geometric isomers[18], with the potential to act as a photo-switch.

Perhaps one of the most intriguing groups of UAAs are those which contain photo-protecting groups. These moieties are generally termed "caging" groups, and are photolysed to afford a native amino acid residue. The rationale of a caging group is to inhibit the natural activity of the amino acid until removed with light. Inhibition is based on the introduction of this group to a critical atom on the amino acid side chain so as to block its natural chemical ability. Photo-caged cysteine[19] (**1-7**) and tyrosine[20] (**1-8**) have both been introduced into proteins where the thiol and hydroxyl groups of the prospective amino acids were blocked by a nitrobenzyl group. These caged amino acids were then photo-chemically decaged upon irradiation with light at 365 nm. The photo-caged cysteine was specifically introduced at an essential cysteine residue in the cysteine dependent protease, caspase 3[19]. Introduction of the caged variant of cysteine inhibited protein activity, which was photo-activated upon irradiation with light. There is an extensive amount of research, and numerous reviews, dedicated to the caged inhibition of molecules, in particular, amino acids[19, 21-27]. Light activation of critical residues in a protein is a very attractive area for the investigation of membrane channels, transcription

events, and signaling pathways as this approach introduces a level of spatiotemporal control over protein function[21, 25, 28-32].

Each of the above UAAs is a specific case of novel chemistry being introduced at the protein level. Each new addition to the growing number of UAAs that have been successfully incorporated into protein allows for a unique probe and extended level of insight into biochemical questions relating to protein chemistry. This work attempts to expand the depth of an UAA molecular toolbox by extending the applications of already developed UAAs while also developing new UAAs for novel probes in protein investigations.

1.1.2 Caged amino acids

While the previous section explored some of the available UAAs used in recent studies, this section will narrow that field to those which contain photolabile protecting groups, in particular, the *o*-nitrobenzyl group (ONB). As stated previously, these groups are referred to as "caging" groups and are photo-released when irradiated with 365 nm light. The caging moiety inhibits the native function of the molecule to which it is covalently attached. The caged amino acid is inactive until decaged by light, thereby restoring the molecules activity. In this approach light provides an external control over biological function when the caged group is placed on a critical residue involved in protein function. Moreover, light irradiation can be controlled in a spatial and temporal fashion providing precise control over processes where location and timing are variables which often have distinct biological implications.

The first instance of this strategy being applied in a biological context was the introduction of an ONB protecting group on the γ-phosphate of adenosine triphosphate

(ATP)[33]. This derivative of ATP was not recognized as a substrate by the ATP dependent enzyme Na/K-ATPase, which is responsible for cellular outflow of Na^+ ions and the influx of K^+ ions. Photolysis released ATP, restoring substrate recognition, and, in turn, restoring activity of the ion pump. This study jump started an appealing movement toward biologically active molecules which can act as photo-switches providing direct control over cellular events.

There is a wide-range of photolabile protecting groups available for the application of light activated biomolecules[23, 34-36]. Some of the more popular variants are derivatives of coumarin (**1-9**) and ONB (**1-10**), where nitrobenzyl groups are by far the most common (Figure 1-2a) due to the ease by which they can be installed on the molecule of interest. This is especially straightforward in the addition of ONB to amino acids with nucleophilic potential, under basic conditions, where simple S_N2 type reactions occur with compounds such as 2-nitrobenzyl bromide (Figure 1-2b, **1-11**), which is commercially available. It is important to note, that the caging group must be installed on the atom of functional interest, thereby inhibiting its native chemistry. For example, figure 1-2b illustrates caging at the thiol of a cysteine molecule thereby inhibiting its nucleophilic potential while protected. The decaging mechanism of ONB groups also occurs at a relatively low energy wavelength which does not have adverse affects on other cellular components. Herein we will focus on the application of ONB groups as their presence in this research is extensively employed.

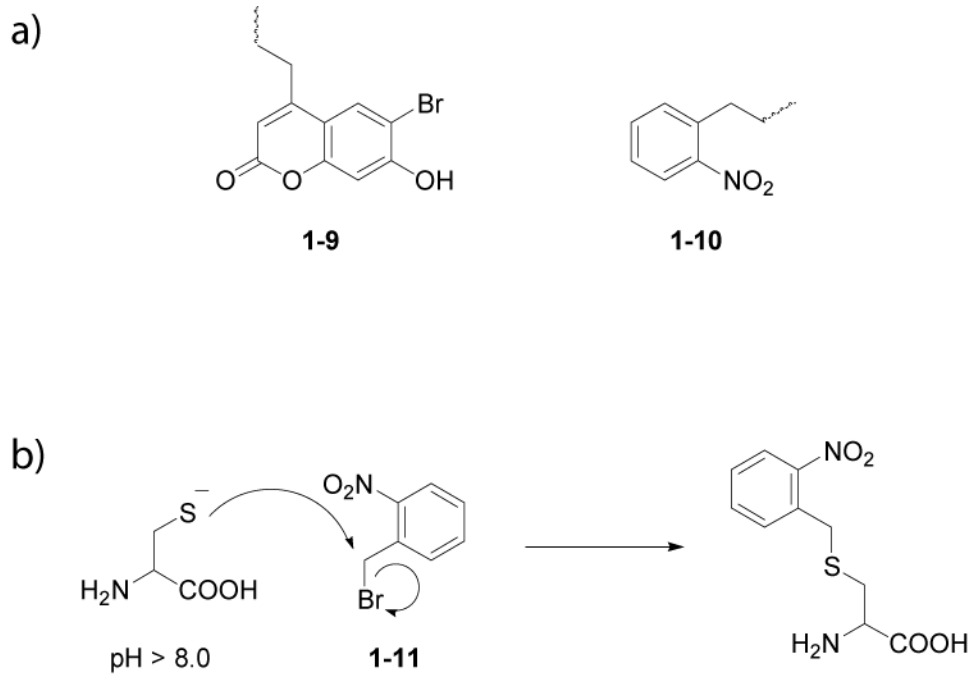
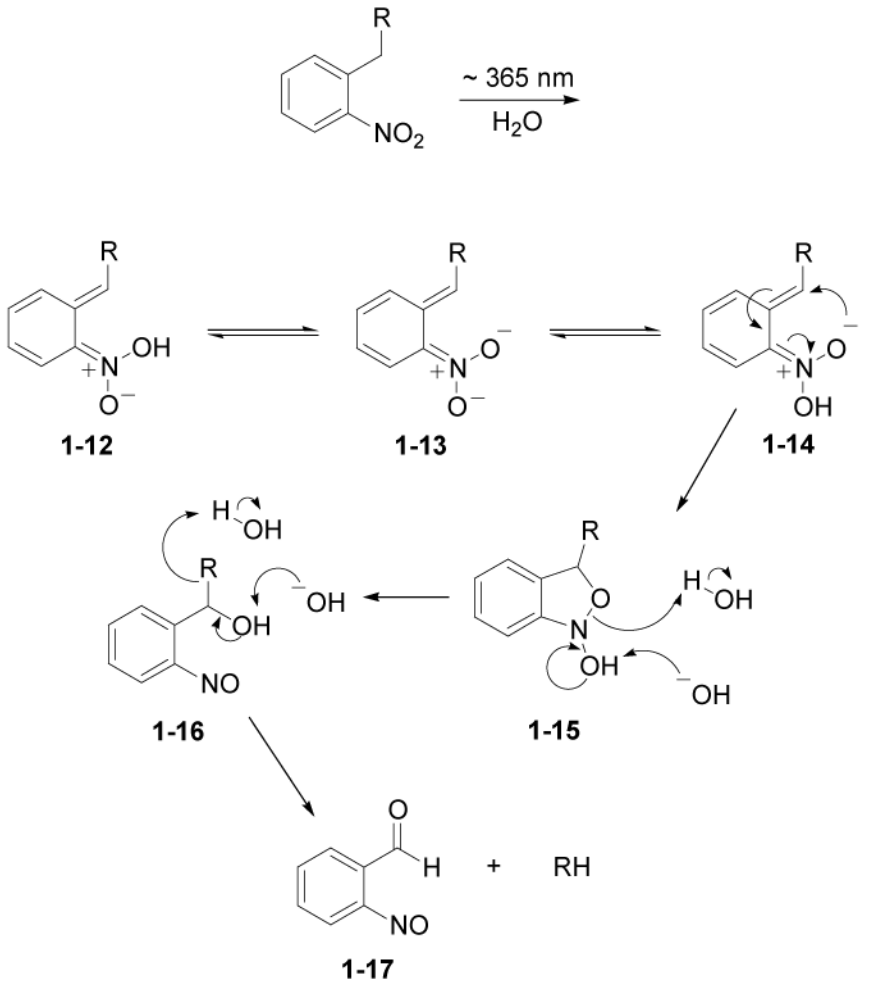


Figure 1-2. a) Commonly used photo-protecting groups: 6-bromo-7-hydroxycoumarin-4-ylmethyl (**1-9**) and ONB (**1-10**). b) Reaction schematic of the general approach to covalent modification of amino acids with 2-nitrobenzylbromide.

The mechanism of photochemical decaging is detailed in Scheme 1-1[37]. The initial excitation of the caged molecule promotes the formation of an *aci*-nitro tautomer (**1-12**), where the generation of this species occurs in the range of <1 ns[23]. The decay of **1-12** proceeds via **1-13** and the subsequent equilibration to yield the isomeric nitronic acid (**1-14**). **1-14** then continues through a cyclization step forming the bicyclic product **1-15** which is broken down to form **1-16**. The final step is a base-catalyzed breakdown of **1-16**. One of the glaring disadvantages to this approach is the generation of a nitrosaldehyde (**1-17**) which has the potential to be harmful in biological context due to its highly reactive functional group. The base catalyzed reaction mechanism is suggested in line with previous studies on the hydrolysis of hemiacetals of acetophenone and its analogues[23, 37].



Scheme 1-1. Decaging mechanism of the *o*-nitrobenzyl group.

From a synthetic standpoint, designing a caged compound is rather clear-cut and seemingly uncomplicated. One must simply identify an essential functional group required for biological activity and covalently modify that group with a photolyzing moiety. This strategy has been successfully utilized to provide an array of caged biomolecules, ranging from small molecule effectors[29, 30, 33, 38, 39] to biological macromolecules, including nucleic acids[28, 31, 32, 40, 41] and amino acids, of which the latter will be addressed in greater detail.

As stated previously, there have been several methods employed for the direct incorporation of UAAs into proteins, and several different caged proteins have been

reported. The differing methodologies will be introduced in further sections. For the application and delivery of caged UAAs, site-specifically at the genetic level, there have been several successful attempts utilizing the caged versions of tyrosine[20, 22, 42-44], cysteine[19, 22], serine[25], aspartic acid[45-47], and lysine[48]. For each of these amino acids the direct caging of the side chain resulted in the inhibition of the protein's biological activity. Subsequent decaging events lead to the release of the native amino acid, restoring the proteins function.

The first genetically encoded caged amino acid was *o*-nitrobenzyl cysteine (ONBC, **1-7**)[19]. In a proof of principle experiment an essential active-site cysteine residue was substituted with ONBC on the human proaprotic protein caspase 3. This protein is a member of the cysteine-aspartic acid protease family, which undergoes proteolitic activation from an inactive proenzyme to its active conformation. When caspase 3 is activated it initiates cysteine mediated protease activity in apoptotic events[49]. The presence of ONBC in the caspase active site inhibited the enzyme's ability to act as a protease, blocking *in vitro* cleavage events. Photolysis resulted in the cleavage of the benzylic carbon-thiol bond and the release of the nitrosaldehyde, reestablishing nearly 40 % of the enzyme activity. This work outlined the possibility of spatiotemporal control for cellular processes that involve essential cysteine residues. In this manner, a portion of the research performed for this dissertation will take advantage of this technology, directly applying the ability to decage a specific cysteine residue involved in protein-small molecule interactions.

More extensive work has been performed with the encoded caged amino acid *o*-nitrobenzyl tyrosine (ONBY, **1-8**). Its first introduction was used to test the inhibition of

β-galactosidase (β-gal), an enzyme responsible for the hydrolysis of β-galactosides to monosaccharides[50-52]. Following irradiation, and subsequent loss of the protecting group; nearly 70 % of the native activity was restored[20]. In more recent studies the exploits of ONBY has been used to control polymerase activity setting the stage for temporal control over DNA replication[44]. DNA polymerase is one of the key enzymes that catalyze the addition of deoxyribonucleotides to the 3' end of a newly forming strand. In an essential spatial arrangement, during the loading of a sequential nucleotide in the extension process, a vital tyrosine residue helps position the template-primer pair into the active site, enabling the DNA template to line up with the incoming nucleotide[53]. Mutational studies of this residue indicate polymerase loss of function[54]. Once again, the introduction of a caged tyrosine at a critically important tyrosine residue inhibited the enzyme's native function. While this residue is not mechanistically important, its conformational position is vital and the caging group is thought to block the access of a free nucleotide into the correct positioning for DNA strand extension. Light induced decaging of polymerase, followed by polymerase chain reaction (PCR) with the caged enzyme, revealed that DNA amplification process could indeed be photo-activated. Caged enzyme that was not subjected to irradiation yielded no PCR product. A wide-range of other enzymatic activities have been reported with several caged derivatives of amino acids. This approach still remains a small subset of biological research but is steadily moving toward the forefront of the chemical biology arena.

The ability to control biological function with light is an extremely seductive approach. There are many active biomolecules that have essential functional groups, all of which have the potential to be restricted in nature due to the addition of a photo-caging

group. The constraint here being that only one specific residue is required for its particular activity. The approach is not well suited for interactions that occur over large molecular surfaces, including many protein-protein and nucleic acid-protein interactions. One of the more clever uses in this work is the attempt to mask amino acids that have only slight modifications, changing their electronic properties, but are still recognized by their endogenous aminoacyl-tRNA synthetases. These UAAs will be termed *near-natural amino acids*.

1.2 Near-natural amino acids

As the term unnatural amino acid could fittingly be applied to the molecules being discussed in this section, it is appropriate to augment their label for distinct reasons. It is true that these amino acids are non-natural; however, they are only slight variants of their native structures. Subtle perturbations allow for the alteration of electronic properties, which can help elucidate biological activity with a means that could not be achieved through other methods. The concern with near natural amino acids (NAA) is that they are recognized by the endogenous aminoacyl-tRNA synthetase machinery that naturally incorporates the non-modified amino acid. This creates toxic effects when a NAA is introduced to cells, as the NAA will be integrated to all cellular protein at its respective native amino acid position.

1.2.1 Fluorinated analogues of histidine and tyrosine

Histidine residues have the only side chain pKa (~6.5) that is near physiological pH. Because of this, the imidazole group has the ability to act as a general acid/base catalyst for enzymatic activity. The substitution of one hydrogen atom of the imidazole ring with a fluorine atom significantly lowers the pKa of the side chain, increasing its

acidity[55] (Figure 1-3). Histidine has long been thought to play a role in acid/base catalysis in the mechanism of RNA hydrolysis by RNAse[56]. The NAA 4-fluorohistidine (**1-19**) was substituted for the active site histidine in RNAse A and used to probe this catalytic reaction[57]. As the mutant exhibited nearly the same catalytic rate (k_{cat}) of the wild-type, its optimal pH was lowered, corresponding to the decrease in pKa of the active site fluorohistidine. This data provides overwhelming support to the proposed mechanism of this enzyme and the implication that histidine does indeed act as a general acid/base catalyst. This supporting information would not be obtainable by any other means as the direct fluorination of a histidine, posttranslationally, is experimentally impossible.

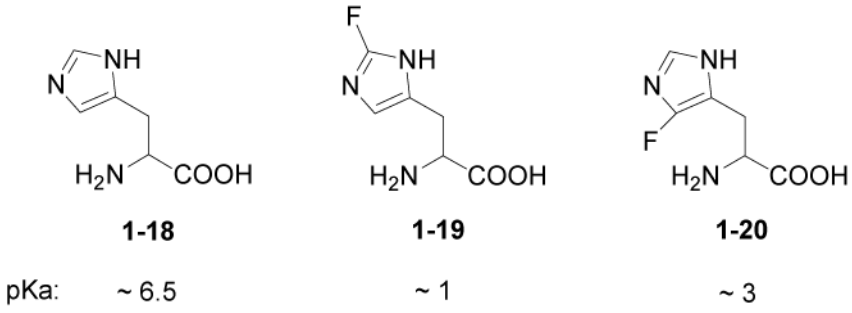


Figure 1-3. Histidine (**1-18**) and its fluorinated analogues. The pKas of the near-natural histidines decrease as shown for 4-fluorohistidine (**1-19**) and 2-fluorohistidine (**1-20**).

Fluorinated tyrosines have also found their way into an extensive amount of research involved in probing protein structure and function. Some of the advantages of developing fluorinated proteins are the ability to directly investigate protein folding by ^{19}F NMR[58-61], as well as introducing altered electronic properties resulting in lower pKa values of the phenolic proton[62-64]. As demonstrated in Figure 1-4, the pKa of a tetrafluorinated derivative (2,3,5,6-tetrafluorotyrosine, **1-25**) decreases the pKa of native tyrosine (~10) to a pKa of nearly 5. This is a substantial change as the new pKa

approaches that of a carboxylic acid. In a manner similar to that of fluorohistidine, the fluorinated tyrosines can act as powerful probes in the investigation of tyrosine active sites. Indeed, this was performed with insulin receptor tyrosine kinase where direct control over the pKa of the substrate tyrosine (via fluorinated derivatives) exposed that a neutral tyrosine was essential for enzyme activity[65].

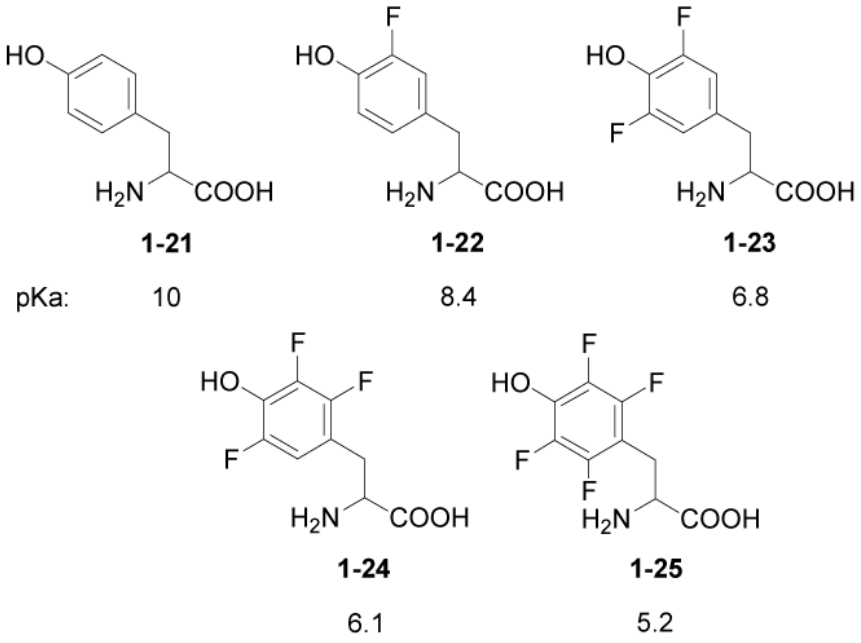


Figure 1-4. Tyrosine (**1-21**) and examples of its fluorinated analogues. The pKa of each derivative decreases as more fluorines are substituted on the ring.

For both of the mentioned NAAs there remains the complication of efficiently generating fluorinated proteins. Synthetic routes have found their utility in the above examples[66-69], but protein size becomes a concern, as larger proteins find synthetic yields are more difficult to obtain. The direct inoculation of cell cultures with these NAAs proves toxic as they are recognized by endogenous translation machinery and they are inserted into all positions which would recognize either a histidine or a tyrosine. This is because the replacement of one (or several) fluorines is an almost perfect isosteric

substitution for hydrogen, of which the aminoacyl-tRNA synthetase cannot distinguish and recognizes as a substrate for the charging of tRNA.

1.2.2 Selenocysteine

While there are potentially limitless possibilities for slight modifications of amino acids, one of the more intriguing analogues is selenocysteine. This amino acid falls into a unique category as it is structurally similar to cysteine; however, it is a natural amino acid. Site-specific encoding of this amino acid is technically difficult as its natural codon is UGA, a universal stop codon, followed by a sequence of mRNA (Selenocysteine Insertion Sequence, SECIS) that specifically recruits cellular machinery required for selenocysteine incorporation. Without the specific SECIS sequence there is no insertion of the amino acid. These represent obvious molecular biological constraints when designing mutational expression sequences. The appeal of this particular residue is the selenol side chain which has a reduced pKa of 5.3[70]. Not only would direct control over insertion of this amino acid provide invaluable data about natural selenocysteine proteins, the reduced pKa and increased nucleophilicity of this side chain group make it an extremely attractive candidate for the evolution of new enzyme function as well as a handle for selective bioconjugation. Synthetic[71, 72] and chemical[73] approaches have been utilized, but in each case there are definitive limitations with their own technical challenges. No direct method exists for site-specific incorporation of NNAs into proteins of larger size, and at any desired position.

1.3 An expanding genetic code

There are several different techniques by which the generation of protein with unnatural amino acids is performed, most of which approach the protein formation from a

synthetic stand point, utilizing solid-phase peptide synthesis. This strategy is extremely effective for small peptides where the diversity of the sequence is unlimited. For larger proteins this method becomes technically restricted as the efficiency and purification yields, following each successive amino acid addition, decreases substantially. There is also the concern that a synthetic protein may not fold properly into its active conformation once it is introduced to biological conditions. Straightforward chemical modification of amino acid side chains can alter functionality. However, direct control over chemical events are often difficult to manage as proteins contain many of the same residues of interest along any one polymer. This can lead to nonselective binding, or multiple modifications, skewing quantitative results. More recently, expressed protein ligation and protein *trans*-splicing techniques combine the use of synthetic peptides, coupled with the expression of a truncated protein[68]. Both of these methods were introduced by Muir, *et al.*[66, 68] and are based on intein splicing mechanisms, which can ultimately fuse synthetic peptides to expressed protein. As clever as this technique may be, the limitation of introducing an UAA is constrained by the size of the peptide being synthesized, resulting in recombinant protein that contains a mutation sequentially located near either termini of the protein. This becomes a dilemma if the mutation of interest is desired near the center of a large protein.

There have been successful attempts at the expression of protein in the presence of chemically misacylated tRNAs, *in vitro*[74] and *in vivo*[75]. This requires the chemical modification, and "false" charging, of a tRNA that recognizes a specific codon. For the *in vitro* studies this was done using a nonsense or frameshift suppressor tRNA and *in vivo* with an amber suppressor tRNA. These systems are appealing but the first is a cell-free

translational mechanism, and both require the chemical additions of mis-acylated tRNAs which limits the production of protein as the tRNA cannot be reacylated once it has been utilized.

A model system would be able to generate unnatural proteins with site-specificity at the genetic level without the need for multiple injections of chemically altered tRNAs. This would allow for the expression of protein containing UAAs at any position, and at sufficient yields, by which isolation and purification methods could be applied. These techniques have been developed and implemented through standard mutagenesis procedures, in order to evolve a unique aminoacyl-tRNA sythentase (aaRS)/suppressor tRNA pair, specific for the amber codon UAG. This system has been termed "an expanding genetic code" (Figure 1-5).

.

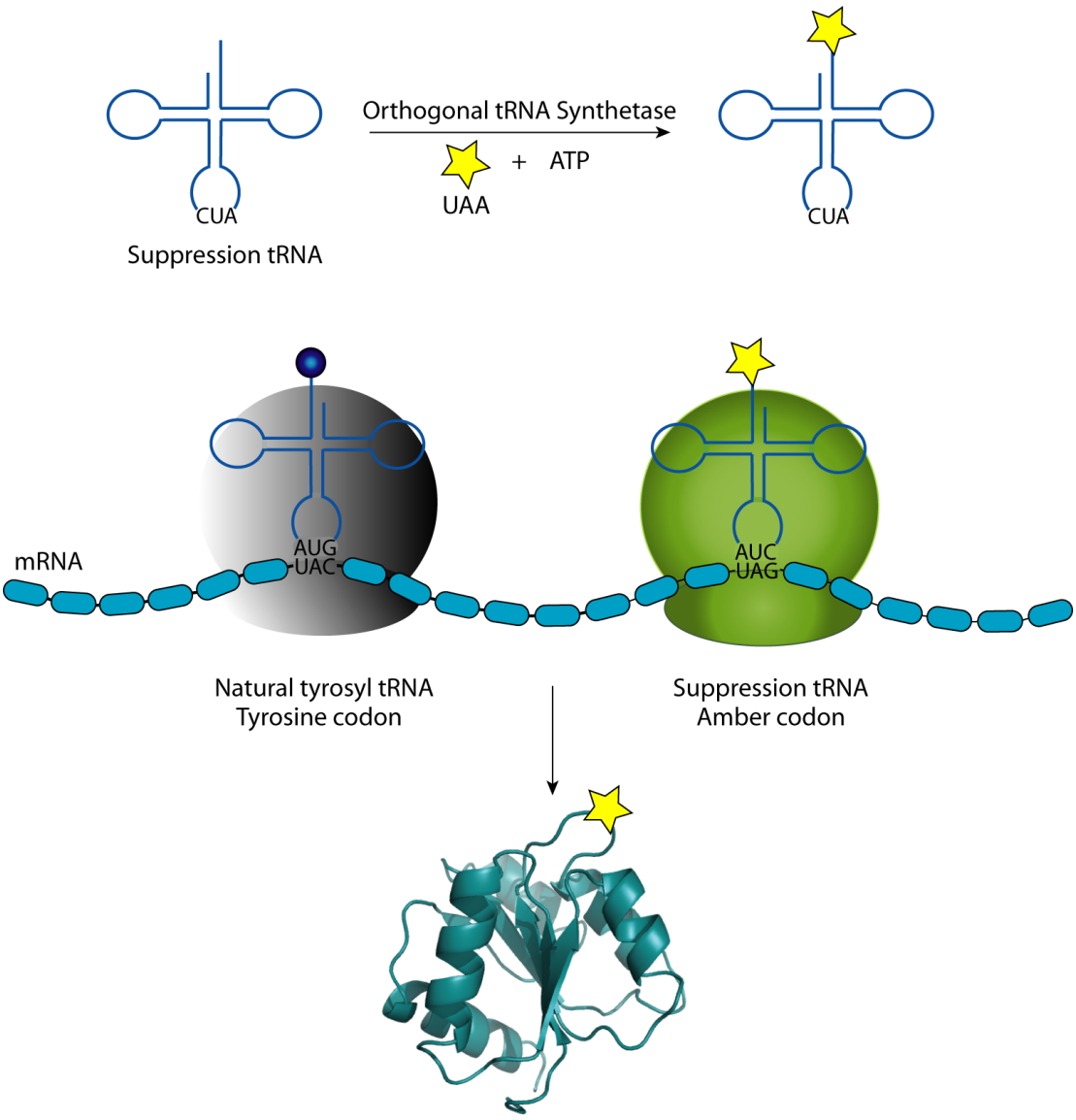


Figure 1-5. A general method for the *in vivo* expression of proteins with an aminoacyl-tRNA sythentase (aaRS)/suppressor tRNA pair, specific for the amber codon UAG. This orthogonal pair works to suppress the amber codon and express full length protein in the presence of unnatural amino acids.

1.3.1 Amber suppression

One of the key components of this new technology is the ability of a tRNA mutant to recognize and suppress the amber stop codon, UAG. This codon is the least utilized terminator in both *E. coli* and yeast. Initial attempts were performed in *E. coli* where it was proposed that an orthogonal tRNA/aaRS pair could be evolved, generating

an aaRS that would recognize an UAA as a substrate and none of the other twenty naturally occurring amino acids. In turn, the charged tRNA specific for a UAG codon would translate and suppress the amber position. Preliminary evolution of these tandem components were performed on existing pairs selected from *E. coli*[76, 77], and yeast[78] with some impediment. Although orthogonal translational partners were generated, these initial pairs found complications with specificity, charging and suppressing the UAG codon with natural amino acids.

It was determined that an existing pair from archael bacteria provided the perfect host by which to evolve orthogonal pairs for the use in *E. coli*. This was due to the fact that no cross-reactions would occur to any significant degree because archael tRNAs will not act as a substrate for an *E. coli* aaRS[79]. The first truly functional orthogonal tRNA/aaRS pair to be used in *E. coli* was derived from the tyrosyl pair from *Methanococcus jannaschii*[80]. One of the defining characteristics of this tyrosine tRNA (*Mj* $tRNA^{Tyr}$) was that its sequence had differing recognition elements in comparison with *E. coli* $tRNA^{Tyr}$. The first base pairing in the acceptor stem is varied, where *E. coli* contains a G-C pairing and *M. jannaschii* contains the reverse, C-G pairing (Figure 1-6). Another point of interest is the fact that the *M. jannaschii* tyrosyl-tRNA synthetase (*Mj* TyrRS) contains a minimal binding domain with the *Mj* $tRNA^{Tyr}$ anticodon loop making it possible to mutate this sequence of the tRNA without a considerable loss of affinity by the synthetase[81]. Indeed, mutation of the anticodon loop was attained altering the codon recognition sequence from UAC to UAG (*Mj* $tRNA^{Tyr}_{GUA}$ → *Mj* $tRNA^{Tyr}_{CUA}$) whereby suppression of the amber codon was observed using the *Mj* $tRNA^{Tyr}_{CUA}$/*Mj* TyrRS pair in *E. coli* protein expression, crafting an efficient amber suppression system[80].

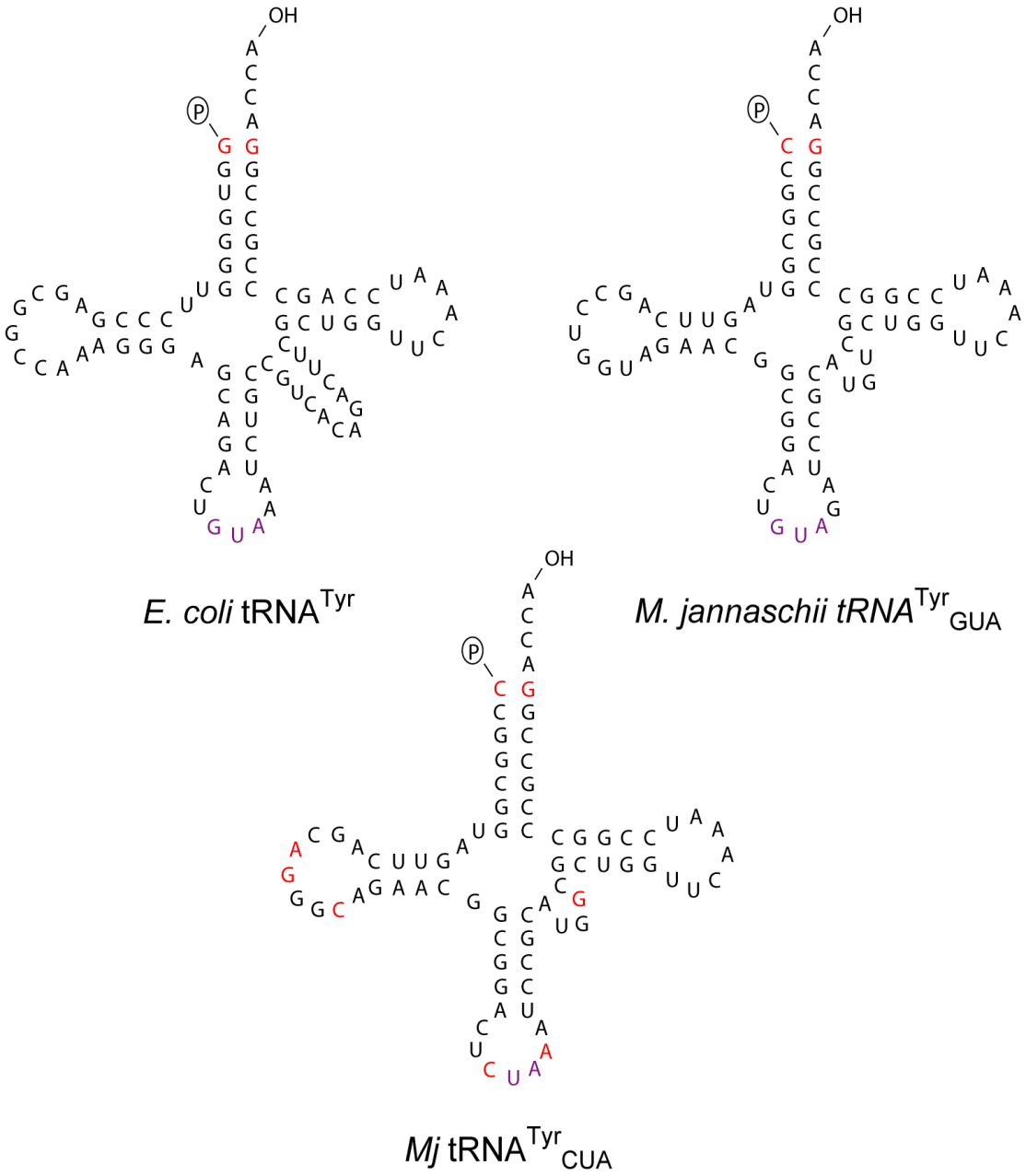


Figure 1-6. Sequence homology of $tRNA^{Tyr}$ from both *E. coli* and *M. jannaschii* as well as the mutant *Mj* $tRNA^{Tyr}_{CUA}$. Mutated nucleotides are shown in red where the anticodon region is in purple.

The *Mj* $tRNA^{Tyr}_{CUA}$ was further evolved to reduce endogenous synthetase recognition of the tRNA. A mutant suppressor tRNA library was designed based around eleven nucleotides which do not have direct interactions with the *Mj* TyrRS[82]. Several rounds of a negative and positive selection methods with definitive markers, afforded a

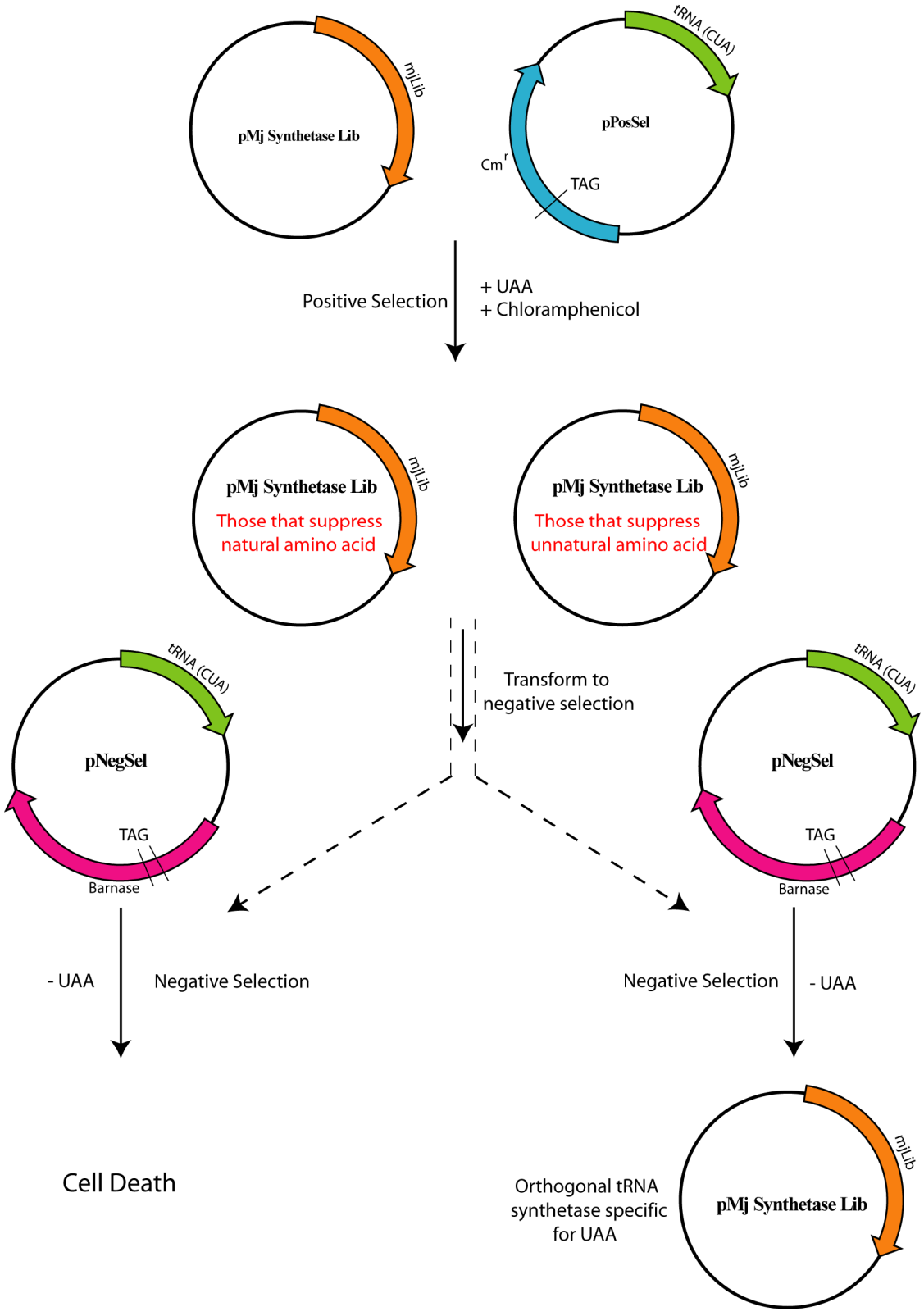
fully optimal, completely orthogonal, *Mj* $tRNA^{Tyr}_{CUA}$ that was recognized efficiently by *Mj* TyrRS to suppress the amber codon (Figure 1-6). Having a system in place lead to the evolution of the *Mj* TyrRS to recognize specific UAAs, and charge them to the cognate *Mj* $tRNA^{Tyr}_{CUA}$.

1.3.2 Synthetase recognition of unnatural amino acids

The substrate recognition of *Mj* TyrRS was evolved to specifically recognize, and charge its tRNA partner, with an UAA and none of the natural amino acids. To accomplish this, mutant libraries were generated with random mutations introduced throughout the tyrosine binding site. These sites were based on crystal structures of the wild-type synthetase. Successive rounds of positive and negative selections were performed with the library of mutant synthetases in the presence of a desired UAA to evolve the most efficient translation pair.

Several different selection techniques have been applied for this system where the first general approach in *E. coli* will be discussed here (Scheme 1-2). This scheme was applied to the evolution of *Mj* TyrRS to recognize, and selectively charge *Mj* $tRNA^{Tyr}_{CUA}$ with *O*-methyl-L-tyrosine (OMe-Tyr)[80]. The positive selection was based on an expression plasmid containing a chloramphenicol acetyl transferase (CAT) gene that included an internal stop codon. When the library was expressed in the presence of this plasmid, only the mutated synthetases that recognize OMe-Tyr suppressed the amber CAT gene and survived in the presence of chloramphenicol. Those library plasmids which survived the positive round were then transformed into a negative selection strain which harbored an amber mutant barnase gene also containing two TAG mutations.

Barnase contains ribonuclease activity, which is toxic to cells when expressed in the absence of an inhibitor[83]. The key to this selection round is that selective pressure is performed in the absence of OMe-Tyr. This step is imperative for removing mutant synthetases which are not orthogonal in *E. coli* and suppress the amber position with an endogenous amino acid. When suppression occurs, the barnase activity is initiated and cell death occurs, ultimately eradicating synthetases which can still recognize native amino acid as a substrate. This round provides cells, which house library plasmids encoding for only orthogonal or nonfunctional aaRS. These two selections are performed multiple times, leading to the evolution of a mutant synthetase, which specifically recognized the UAA as its substrate, charging *Mj* $tRNA^{Tyr}_{CUA}$, and suppressing the amber codon.



Scheme 1-2. General scheme for the positive/negative selections for the evolution of tRNA synthestases specific for UAAs.

A similar approach was later developed which utilized a green fluorescent protein expression (GFPuv) driven by a T7 promoter that controlled expression of a T7 polymerase. The polymerase gene contains an amber codon which is dependent upon suppression for full-length polymerase activity[84]. If the amber position in the polymerase gene is not suppressed there is no subsequent transcription of the GFP gene. This positive selection also maintained the amber-CAT gene, as mentioned above. When this selection was performed in the presence of UAA, suppression of the amber codon in the T7 polymerase gene afforded full length T7-polymerase, and in turn, the expression of GFPuv. Survival was also dependent upon the suppression of the CAT gene, and growth in the presence of chloramphenicol. An advantage to this system was the selective pressure was dependent on two separate events, both of which were contained in the same plasmid. There was also a very clear phenotype, making sorting of fluorescent cells particularly efficient. The plasmids from cells which contained this phenotype were then introduced to the negative selection round, as described above. This alternative approach to the selection method is mention here due to its direct utilization in the work presented within.

1.3.3 Reach of an expanding genetic code

Since the development of this system in *E. coli*, unnatural proteins have seen a surge in popularity throughout biological studies. This is due to the fact that UAAs can now be introduced site-specifically at any position with standard molecular mutagenesis techniques. Success in *E. coli* has lead to the direct development of this system in yeast and mammalian cells.

Although the eukaryotic transcription/translational machinery differs from that of *E. coli*, orthogonal pairs of tRNA/aaRS have been generated. In selection schemes that approach the evolution of orthogonal pairs in similar fashion to those in the above mentioned methodology, tRNA/aaRS pairs from *E. coli* have been imported to yeast which act as the translational components for suppression. Amber suppression in the yeast selections drive GAL4-responsive *his3, ura3,* and *lacZ* reporters[85] where the negative rounds are based on the toxic effects of an amber uracil marker which converts 5-fluoroorotic acid to its toxic product. The details of this selection process will not be outlined in great detail, as this work did not directly apply this selection process; however, an evolved synthetase using this system was utilized for the incorporation of a caged cysteine.

In mammalian cells several different pairs have been generated which have been imported from bacteria[86, 87]. These have seen the specific incorporation of UAAs including 5-hydroxytryptophan. The fact that this selection process can extend into higher species highlights its practical nature and potential for highly specific protein studies using UAAs. Not only does it possess the ability to probe protein function, it may also find its value in the newly emerging field of synthetic biology where its direct application could aid in the development of novel proteins with unprecedented function.

1.4 Specific aims

This work is intended to expand an already impressive molecular toolbox for the exploration of protein structure and function. It spans many fields of study, specifically those of chemistry and biology, utilizing various molecular biological techniques. With the advent of genetic encoding of UAAs the potential for new probes, which are yet

undeveloped, seems limitless. The focus here is the development of new tools, where each specific case, will be applied to a proof-of-principle system to illustrate its usefulness for further applications.

Exploitation of previously evolved tRNA/aaRS pairs will be applied to studies which recognize close analogues of their reported cognate amino acids. This will serve to incorporate isotopic pairs of a crosslinking amino acid for fingerprinting studies via mass spectral analysis. Another system will be used to explore the use of spatiotemporal control with a caged amino acid, where light activated fluorescent labeling of a protein will be addressed. Work in both of these examples extends the applications of existent translational partners.

The ability to increase the number of UAAs available for protein studies is one of the appeals of such a general selection method. It creates an unlimited number of potential candidates for new orthogonal pairings for the suppression of the amber codon. As discussed above, the near-natural analogues of amino acids are not suitable candidates for this system as they are recognized by endogenous synthetases. This work applies directly the selection scheme in *E. coli* for the evolution of a new tRNA/aaRS pair which specifically recognizes the caged version of near-natural fluorinated tyrosine derivatives. This work takes advantage of UAA technology to mask near-natural amino acids for their site-specific incorporation into proteins.

Chapter 2: Mass spectral fingerprinting via the genetic encoding of an isotopically labeled photoaffinity probe

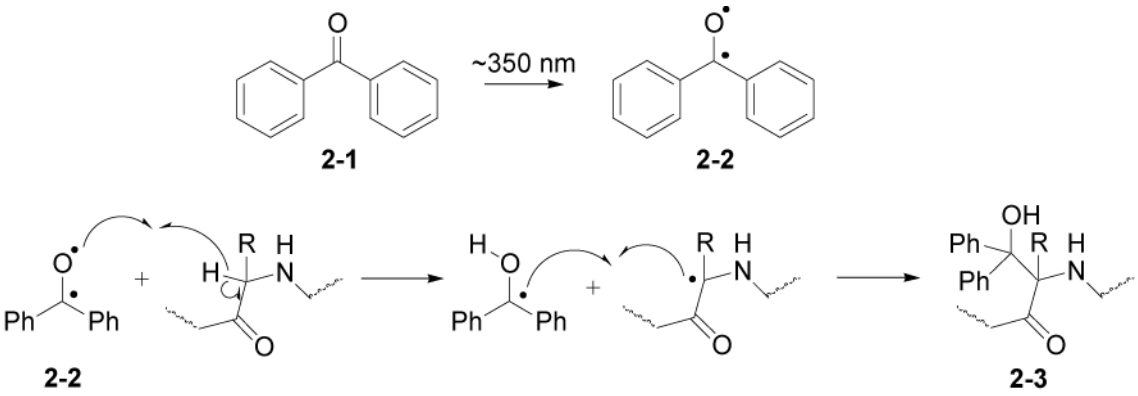
2.1 Introduction

Many methods have been developed to identify, map, and potentially inhibit protein-protein interactions. Classic analysis has been performed through techniques such as co-immunoprecipitation[88], where protein associations are strong enough to survive cell lysis techniques and can be captured by affinity purification assays. This technique is rather powerful but can miss transient protein interactions, which are more difficult to discern using *in vitro* methods. Alternatively, *in vivo* techniques such as the yeast two-hybrid assay[89, 90], protein fragment complementation[91-93], or fluorescence resonance energy transfer (FRET)[94] can be employed, which are not only sensitive, but can survey protein-protein interactions in the physiological environment of the cell. A major limitation of these approaches, however, is that they require candidates to be expressed as fusion partners, and therefore judged as pairs. This can prevent the direct identification of unknown interaction networks that are so often part of large protein complexes and signalling pathways.

While, the previously described techniques help identify potential binding partners, they provide little, if any, information about the molecular detail of the interaction. Chemical crosslinking, paired with mass spectral analysis can identify specific interfacing amino acid residues responsible in the molecular binding events at protein interfaces[95-97]. What would normally be transient interactions can be seized by covalent attachment of reactive compounds. These covalently captured proteins can then be rigorously purified and subjected to enzymatic digestion and analysis by mass

spectrometry to determine peptide fragments that contain crosslinked product. As straightforward as this technique may appear, there are limitations as to the specificity of a desired crosslink. Typically, the protein and crosslinking agent are mixed together *in vitro* and then analyzed for crosslinking activity. Many agents are specific for only one type of functional group, where there are bound to be multiple positions for a group along any one peptide chain. This creates difficulty when interpreting the crosslinked complex due to either the unknown position of the crosslink, or multiple crosslinking within the same complex.

Benzophenone (**2-1**) photochemistry has long been used to probe and identify non-covalent interactions in living systems[98], dating back nearly 35 years to its first application in biological context[99]. This probe has three distinct advantages. The first is the wavelength by which the manipulation of this compound is performed. It is activated with ambient light at a wavelength of 350-360 nm. This relatively low energy ultraviolet radiation is not damaging to protein or DNA. This is an extremely important point for employing this compound in *in vivo* studies. Secondly, benzophenone reacts preferentially with C-H bonds within 4 Å, even when water is used as a solvent. This interaction is favorable for crosslinking biomacromolecules, particularly amino acids at side chain carbons. The last advantage is the stability of the benzophenone. Once activated it readily relaxes to the ground state configuration if an H-donor is not found. This is advantageous for probing *in vivo* because if a connection is not made, the relaxation and excitation can go through cycles until the proper H-donor and geometry of an interaction occurs.



Scheme 2-1. Photochemical pathway of the benzophenone radical mechanism. When activated, one electron in the benzophenone photophore is promoted from a non-bonding sp^2-type orbital on the oxygen to an antibonding π^* orbital of the carbonyl group.

The diradical triplet state (**2-2**) contains an electron deficient oxygen which acts as an electrophile to react with adjacent C-H sigma bonds resulting in the abstraction of hydrogen from the carbon (Scheme 2-1). The resulting radicals readily recombine to form a new C-C bond, affording a benzpinacol-type compound (**2-3**). When the excited oxygen is in spatial orientation with an amine, or other comparable heteroatoms, the reaction may occur through an electron transfer resulting in a hydrogen abstraction from an adjacent alkyl group and a consequent 1,2-radical shift.

Early successes with photoaffinity labeling clearly outlined its potential in biological applications[100, 101]; however, these first attempts at modifying peptides with photolabile probes were dependent upon hetero-bifunctional agents. These probes consisted of both a chemically reactive group as well as a photochemical handle. The photoactive group could only be introduced at a chemically comparative peptide site, which could potentially limit the functionality of the protein, or block ligand binding sites. Another complication, as stated previously, was dependent on how many functionally reactive sites were present along one peptide chain. A bifuntional crosslinking agent's site-specificity is difficult to preserve as chemical control is lost once

the compound is introduced to the system. There are typically many functional sites along a peptide by which the agent can react eliminating specificity control over the molecule.

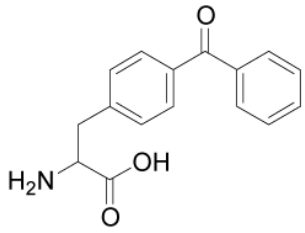


Figure 2-1. *p*-Benzoylphenylalanine

To eliminate site-specific reactivity a photoreactive amino acid, *p*-benzoylphenylalanine (pBpa), was synthesized that could be easily incorporated into proteins through standard solid-phase techniques[102] (Figure 2-1). This arylketone derivative of phenylalanine displayed the stability and the remarkable selectivity of benzophenone when used for crosslinking. Succesful sythentic mutation with pBpa resulted in crosslinking events at a seventy percent yield[102].

As this method proved to be a powerful adaptation of the benzophenone functionality, it too had its limitations. The ability to incorporate this unnatural amino acid through solid-phase technology is hindered by the size of the peptide by which it is to be introduced. Short peptides are typically simple to synthesize with relatively high yields (more than ever with the advent of automated microwave synthetic synthesizers), but solid-phase synthesis of larger protein becomes exhausting. Recently, pBpa has been site-specifically incorporated into proteins in *E. coli*[15], yeast[85], and mammalian cells[103]. Using an orthogonal tRNA/aaRS pair that was constructed to insert pBpa into proteins at the amber codon, TAG. Not only does this allow for probing *in vivo* at physiological levels, it also allows for the site-

specific incorporation of pBpa into any size protein of interest.

As mass spectrometric analysis is increasingly becoming the standard practice to map protein-protein interaction surfaces of chemically cross-linked proteins, a major concern presents itself when trying to assign the cross-linked fragment of an unknown protein to a peak in the spectrum. This is problematic as the complexity of fragments that can result from enzymatically treated complexes is not trivial. Different peptide fragments can have the same nominal mass depending on their sequence. This is further exemplified by missed cleavages during the enzymatic digestion which generates peptides of unpredicted masses.

This chapter will outline a means of simplifying this analysis during the use of unnatural amino acid mutagenesis, an isotopically labelled analogue, [D_{11}]-pBpa, was synthesized and used as an encoded label to identify peptide fragments that have been site-specifically cross-linked and analyzed via mass spectral techniques[104].

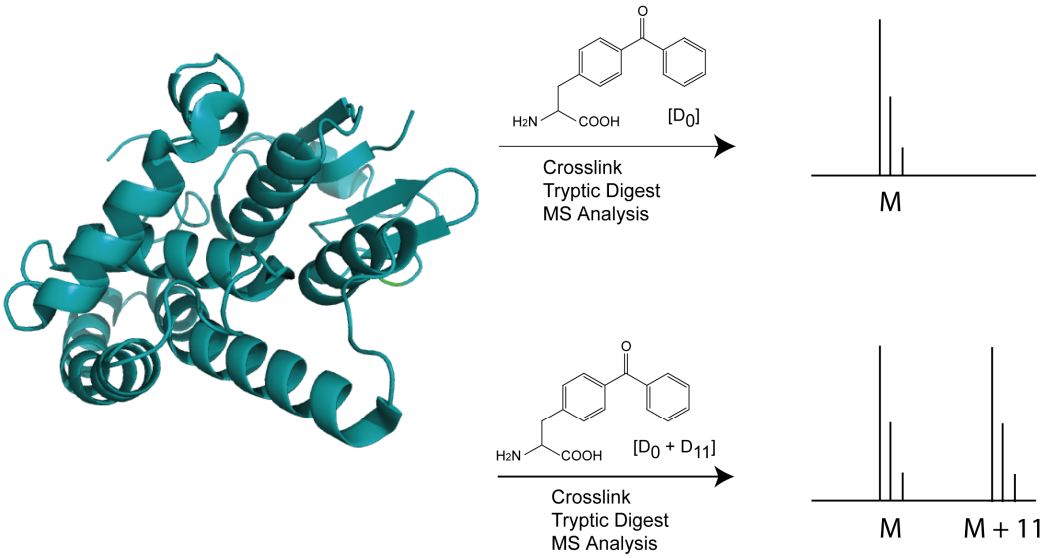


Figure 2-2. General application of labeled pBpa for isotopic fingerprinting. When mixed isotopes of pBpa are introduced to protein, the introduction of an M, M +11 fingerprinting doublet will be observed during mass spectral analysis. This signature doublet will allow for rapid and accurate identification of the mass of interest.

2.2 Experimental

2.2.1 Materials

[D_8]-Toluene and [D_5]-benzoylchloride were purchased from Sigma. All enzymes were purchased from New England Biolabs. Sequencing was performed at the University of Michigan sequencing facility. All other reagents were purchased commercially and used without further purification**.** The plasmid pSUP-pBpa was supplied by Dr. Peter Schultz (The Scripps Institute, La Jolla, CA).

2.2.2 General Methods

The antibiotics employed were chloramphenicol (35 μg mL^{-1}) and ampicillin (100 μg mL^{-1}). PCR reactions were performed with Taq DNA polymerase under the following conditions for 30 cycles: 30 s at 95 °^{o}C, 30 s at a calculated annealing temperature (5° below the lowest oligo Tm), and extension time of 1 min/Kb at 72 °C. DNA isolations were performed using QIAGEN mini-prep kits according to manufacturer's protocols. Ligations were performed using T4 ligase under standard conditions. All transformations were performed into GeneHogs (Invitrogen) electro-competent cells using ~100 ng of DNA transformed into 100 μL of cells. pBpa was dissolved in water with 0.5 M NaOH added dropwise until fully dissolved at room temperature, prior to being added to expression media. UV irradiation (360 nm) was performed with a hand-held 100W Blak Ray lamp used at a distance of 5 cm, at 4 ^{o}C, for 30 min. Purification of GST was performed using Probond Purification resin (Invitrogen) according to the manufacturer's protocol for native isolation. Denaturing discontinuous sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels (12 % acrylamide resolving and 5% acrylamide stacking) and 1X Laemmli buffers were used following standard

procedures, performed at 200 V for 65 min. SDS-PAGE gels were stained with Coomassie Brilliant Blue. Matrix-Assisted Laser Desorption/Ionization-Time of Flight (MALDI) samples were prepared with external standards: insulin (bovine), 25 pmol μL^{-1} in 0.1% TFA; insulin oxidized B, 25 pmol μL^{-1} in 0.1% (v/v) TFA. Sinapinic acid was used for the matrix at a concentration of 10 mg mL^{-1} in 50% (v/v) acetonitrile/0.05% (v/v) TFA. Spotting solutions were 4:1:1:2 μL (tryptic protein: insulin ox B: insulin: matrix) of which 2 μL was spotted on the MALDI plate, in triplicate. MALDI spectra were recorded on an Axima-CFR spectrometer (Shimadzu) operating in reflectron mode at 90% maximum intensity, optimized for m/z 3500. All spectra were the sum of 500 laser shots per spectra.

2.2.3 Synthesis of [D_{11}]-p-benzoylphenylalanine

The precursor bromomethylbenzophenone (**2-4)**, was synthesized as previously reported[105], and used to assemble the amino acid as follows (Scheme 2-2). Diethylacetamidomalonate (614 mg, 2.83 mmol) was added to a stirring solution of sodium ethoxide (385 mg, 5.66 mmol) in ethanol (10 mL). **2-4** (810 mg, 2.83 mmol) was dissolved in ethanol (15 mL) and added dropwise over the course of ten minutes. The reaction mixture was brought to reflux for 24 hours. Upon cooling, water (20 mL) and ether (20 mL) were added and the organic layer was separated. This was washed successively with 1% (w/v) $NaHCO_3$, brine, and then dried over $MgSO_4$ and concentrated under vacuum. Deprotection was performed by the addition of 6 M HCl (40 mL) and then brought to reflux for 24 hours. Upon cooling, the reaction mixture was dried under vacuum. Recrystallization with water afforded the desired amino acid as light yellow crystals (0.501 g, 64%). ^{13}C (MeOD, 100 MHz) δ198.12, 171.03, 140.66, 138.51, 138.05,

133.44, 131.26, 130.59, 130.35, 129.07, 54.72, 36.46. HRMS (ESI) for $C_{16}D_{11}O_3NH_4$ calculated 280.1913, found 280.1928.

2.2.4 Plasmid construction and protein expression

The gene encoding glutathione-S-transferase (GST) was amplified from pGEX-T4-1 (GE Healthcare) using primers (FWD) 5'- CTA GGA TCC CCT ATA CTA GGT TAT TGG -3' and (REV) 5'- CCA GTC GAC GCC TCT AGA AAC CAG ATC CGA TTT -3'. The product was then digested with *Bam*HI and *Sal*I and cloned into the *Bgl*II and *Sal*I sites of pBAD-mycHisA (Invitrogen) resulting in the addition of an N-terminal MDPSSR leader peptide and a C-terminal hexa-histidine tag. Mutation of the F51 codon to TAG (mutation underlined, note: position 51 is in reference to the WT position. In the gene sequence used for these expressions the mutation resides at position F57 due to the introduction of the leader peptide) was performed using standard Quickchange PCR with primers (FWD) 5'-TTG GGT TTG GAG <u>TAG</u> CCC AAT CTT CCT -3' and (REV) 5'-AGG AAG ATT GGG <u>CTA</u> CTC CAA ACC CAA -3'. These mutated plasmids were then transformed along with the pSUP-pBpa plasmid into *E. coli*. Transformants containing pBAD-GST51TAG, and pSUP-pBpa were inoculated in LB medium containing ampicillin and chloramphenicol and grown in the presence of 1 mM pBpa (Bachem), or 1 mM pBpa[D_{11}], or 0.5 mM pBpa/0.5 mM pBpa[D_{11}] and grown to OD_{600} = 0.8. The cultures were induced with 0.2 % (w/v) arabinose and allowed to express for 5 hr. Purified protein was subjected to crosslinking at time intervals of 15 min and 30 min and then subjected to SDS-PAGE analysis.

2.2.5 Tryptic analysis

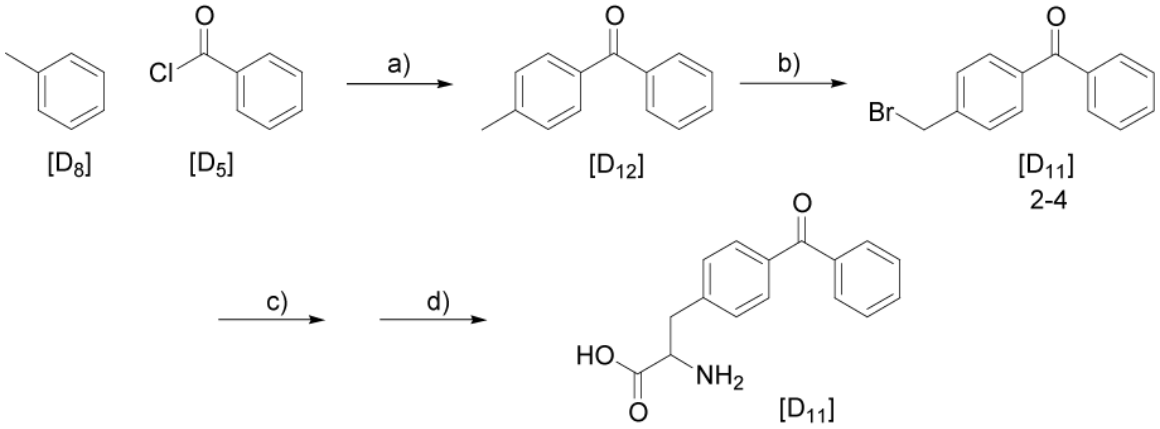
Gel bands to be used for in gel tryptic digests were cut into small cubes and destained twice with 100 mM NH_4HCO_3 in 50% MeCN (200 μL) for 45 min at 37 °C. The gel pieces were dried by Speedvac, and then 10 mM DTT in 100 mM NH_4HCO_3 (100 μL) was added for 15 min at 37 °C. The supernatant was removed and a solution of 55 mM iodoacetamide in 100 mM NH_4HCO_3 (100 μL) was added for 30 min at 37 °C, in the dark. The supernatant was discarded and the gel cubes were washed 3 times with 25 mM NH_4HCO_3 (200 μL) for 15 min at room temp. The gel pieces were dried by speed vac, and then trypsin buffer (12.5 ng/μL trypsin in 50 mM NH_4HCO_3, 60 μL) was added and incubated for 1 hr at 4 °C. The supernatant was removed and the gel pieces were covered with 40% (v/v) MeCN in 40 mM NH_4HCO_3 (50 μL) then incubated for 18 hr at 37 °C. The supernatant was removed and saved and then 0.1% TFA (50 μL) was added to the gel pieces for 1 hr at 37 °C. This supernatant was combined with the previously saved supernatant and dried by speed vac. The tryptic fragments were resuspended in 50% (v/v) MeCN/0.05% (v/v) TFA (15 μL), ready for mass spectrometry analysis.

2.3 Results and Discussion

2.3.1 Synthesis and substrate recognition of isotopically labeled *p*-benzoylphenylalanine

As a means of simplifying mass spectral analysis during the use of unnatural amino acid mutagenesis, an isotopically labelled analogue of pBpa, [D_{11}]-pBpa, was synthesized and use as an encoded label to identify peptide fragments that have been site-specifically cross-linked. The synthesis of the deuterated amino acid began with a Friedel-Crafts acylation of [D_8]-toluene with [D_5]-benzoylchloride (Scheme 2-2). The labeled benzophenone moiety was then brominated and the product coupled with diethyl-

acetamidomalonate to give the protected amino acid. The resulting intermediate was de-protected by acid hydrolysis to afford the amino acid in four steps. This synthesis is non-stereoselective and yields a racemic mixture, of which only L-[D_{11}]-pBpa is incorporated into protein. This approach afforded labeled amino acid on the gram scale, providing sufficient material for many protein expression studies.



Scheme 2-2. Synthetic scheme for the synthesis of [D_{11}]-*p*-benzoylphenylalanine. a) [D_8]-Toluene, [D_5]-benzoyl chloride, $AlCl_3$; b) NBS, AIBN, CCl_4, reflux overnight; c) Diethylacetamidomalonate NaOEt/EtOH, reflux, 24h d) 6N HCl reflux, 24h.

To determine if the cross-linked peptide fragments were easily observable with mass spectrometry, [D_{11}]-pBpa was incorporated into GST. In its native form GST is dimeric, and has been shown to nonspecifically[106] and specifically[15] cross-link. GST was expressed containing a pBpa mutation in place of phenylalanine 57 (F57), a position located in the dimeric interaction surface[107] (Figure 2-3). The residue F57 refers to the mutational position following the introduction of a leader peptide during construction of the plasmid for this study. The position of the mutation is located at F51 on the wild-type protein. It was assumed that with such a small structural change to pBpa, [D_{11}]-pBpa would also act as a substrate, and be accepted by the previously evolved aminoacyl-tRNA synthetase evolved to accept pBpa[15]. Just as NAAs are recognized by their

endogenous aaRS, [D_{11}]-pBpa acts as an isoteric match for substrate recognition of the previously evolved tRNA/aaSR pair. Therefore the amino acid could be inserted into proteins using the previously described machinery.

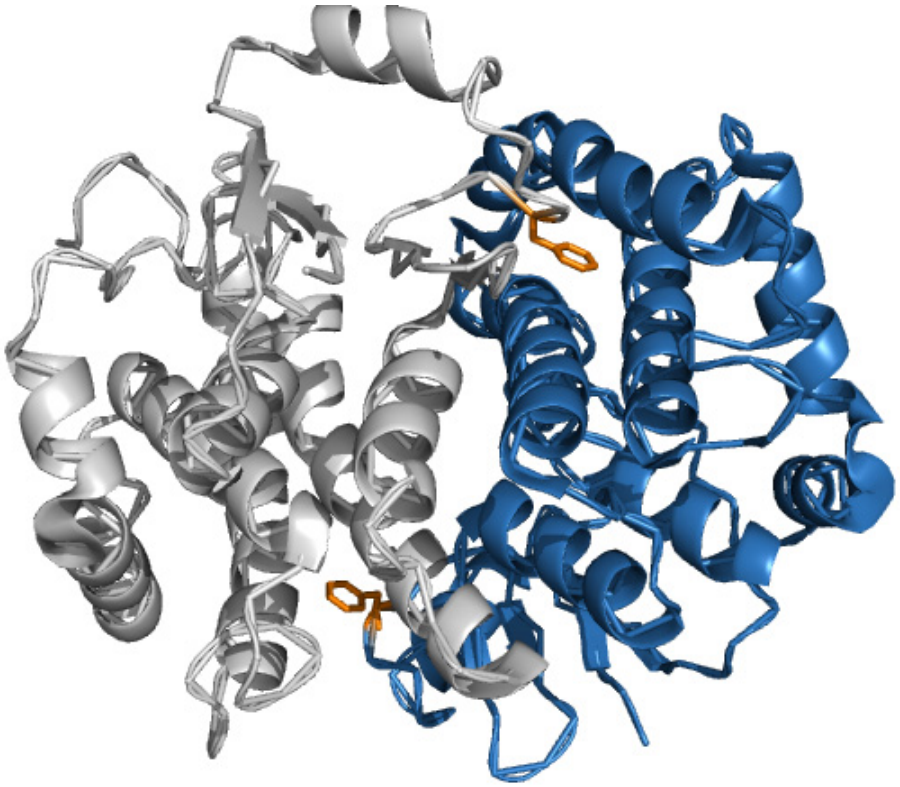


Figure 2-3. Crystal structure of the homodimer of *Schistosoma japonicum* GST. F51, which sits at the dimeric interface on both monomers, is highlighted in orange. PDB entry 1y6e.

Protein expression media was supplemented with either unlabeled [D_0]-pBpa, labelled [D_{11}]-pBpa, or a 1:1 mixture of each of the isotopic variants. Following protein isolation, solutions of protein containing the unnatural amino acid were irradiated with 360 nm light to instigate cross-linking. The wild-type protein was also irradiated as a negative control. As can be seen in Figure 2-4, isolated GST containing unnatural amino acid was irreversibly captured as a covalent dimer. No reaction was observed for the WT protein in the absence of pBpa. These results confirmed that the deuterated form of pBpa

is indeed a substrate for the previously evolved aminoacyl-tRNA synthetase and that it also exhibits identical photoreactivity.

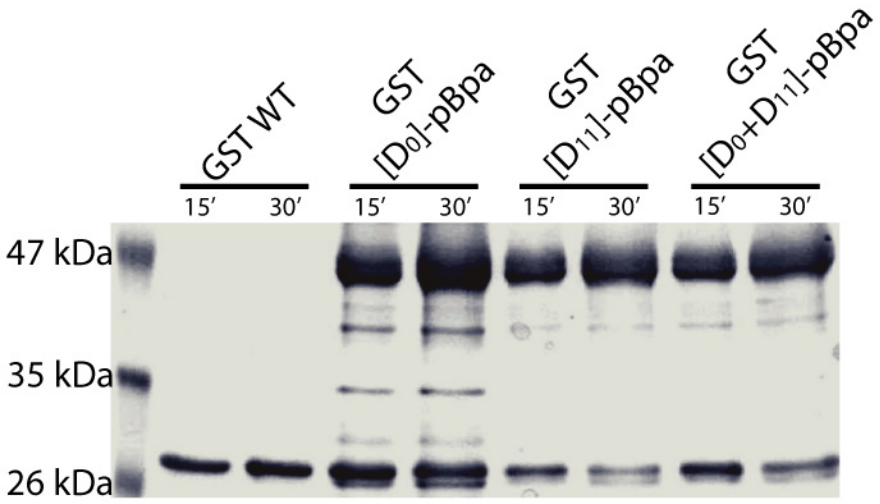


Figure 2-4. Coomassie stained SDS-PAGE gel of GST protein isolation following irradiation at 360 nm for the indicated time.

2.3.2 Mass spectral analysis of crosslinked peptides

The protein bands corresponding to the monomeric and cross-linked proteins were excised from the SDS-PAGE gel and digested with trypsin. The resulting peptides were analyzed by MALDI-TOF mass spectrometry. The residue of interest, F57, was mutated to TAG and resides on the predicted GST fragment 51-FELGLEpBpaPNLPYYIDGDVK-69 with m/z = 2333.2. Upon inspection of the mass spectrum no peak was observed corresponding to this mass. However, there is a larger peak at m/z = 2460.5 that could possibly be assigned to a mis-cleaved fragment containing an extra lysine residue, 50-KFELGLEpBpaPNLPYYIDGDVK-69. Indeed, inspection of the tryptic fragments of protein produced using [D_{11}]-pBpa displayed a M +11 peak at m/z = 2471.4, and a 1:1 mixture of the two variants displayed the characteristic, M and M +11 "doublets", thus confirming the identity of this peak. Using

this strategy, the spectra was searched for the corresponding cross-linked protein at the M and M +11 signature fingerprint quickly identifying peaks at 3028.8 and 3039.5, respectively (Figure 2-5). The observed mass difference of 695.6 between the original peptide (2333.2) and the measured mass of 3028.8 matches the specific capture of GST residues 137-MFDER-141 to the original peptide.

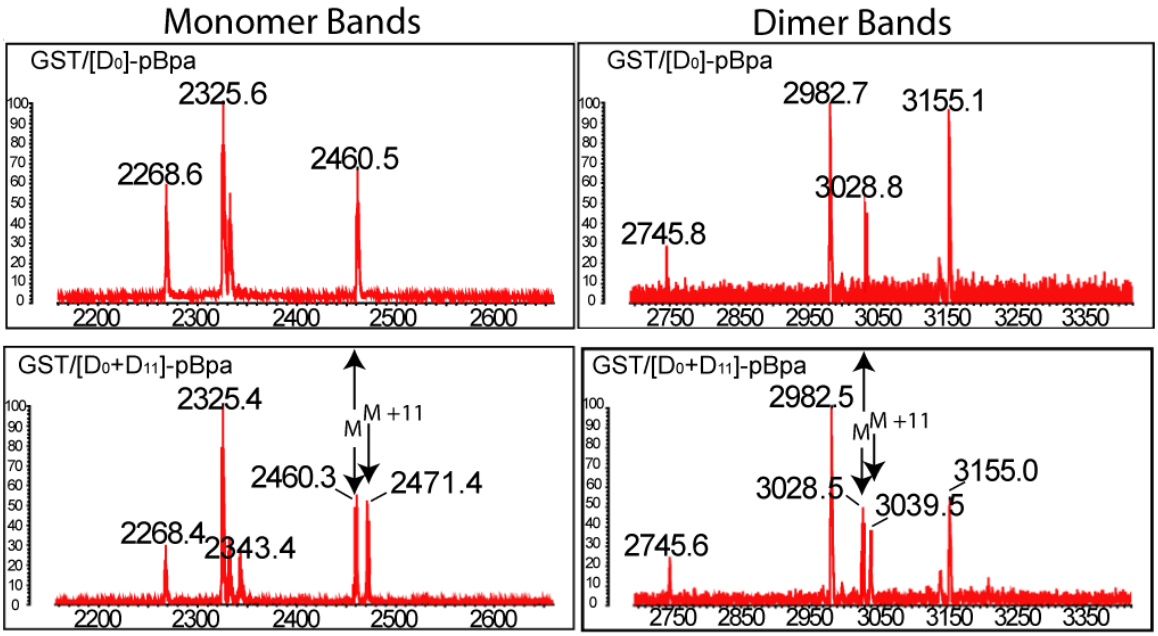


Figure 2-5. The MALDI-TOF spectra of trypsinized monomeric and cross-linked GST, isolated from the SDS-PAGE bands as shown in Figure 2-3. Each spectrum shows a zoom of the area encompassing the assigned peak, where the remaining spectra are identical. The assigned peaks for the peptide shown differ by 11 Da, depending on the unnatural amino acid used.

Inspection of the X-ray crystal structure of the GST dimer interface shows that indeed, this fragment is within 4 Å of the mutated residue, on the opposing monomer (Figure 2-6). No other tryptic fragments, isolated from these gel bands, displayed the M and M +11 signature peaks, therefore these were confident assignments of fragments captured by photo-crosslinking. Intriguingly, non-crosslinked peptide is not seen in these samples, indicating that both peptides of the symmetrical dimer have reacted and doubly

crosslinked. Furthermore, there is no indication of intramolecular crosslinks within the GST monomer. Thus, site-specific crosslinking using these unnatural amino acids can provide accurate MALDI-TOF data that is structurally relevant information, provided the mass peaks are assigned.

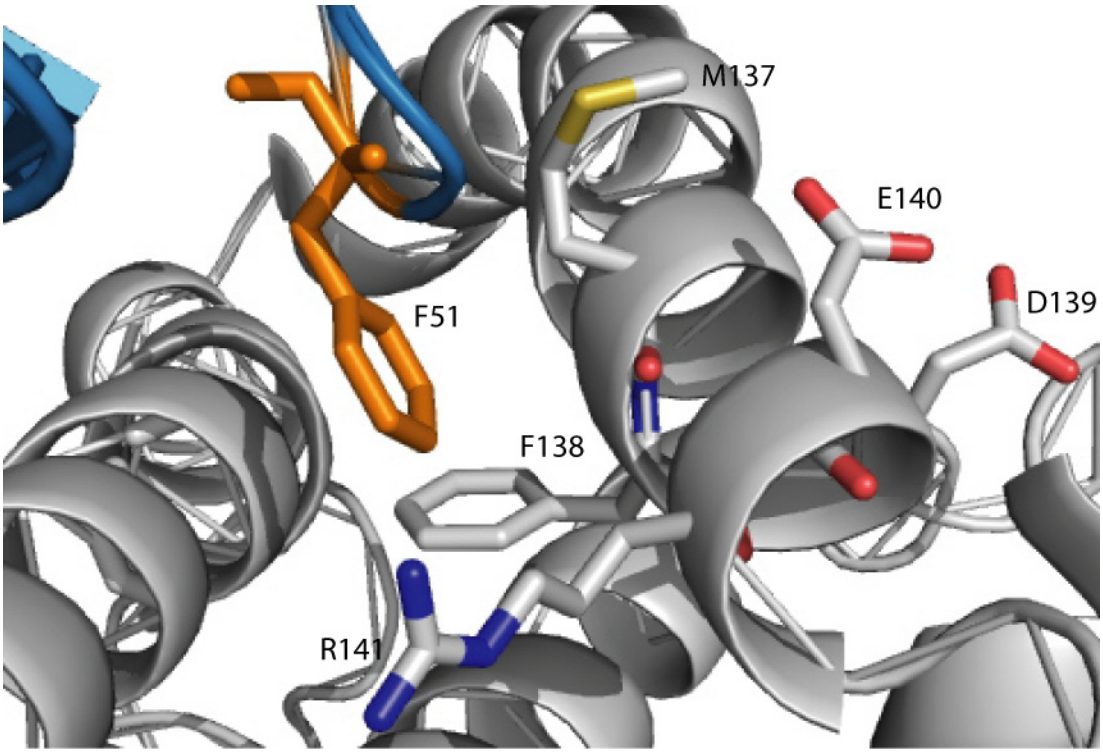


Figure 2-6. Structure of the GST dimer interface where the position of the mutated F51 is depicted in orange on the blue monomer. The residues on the opposing monomer (137-MFDER-141) are shown in grey, where M137 and F138 are within 4 Å of the crosslinking agent. PDB entry ly6e.

2.4 Conclusion

Isotopic encoding is a strategy that can be used in identifying protein interactions via crosslinking with the unnatural amino acid, pBpa. The simplicity of this technique and the ease at which fragments of interest can be identified, significantly reduces the time for spectra analysis. This isotopic fingerprint alleviates the need for multiple calculations that would previously have been needed to deduce the correct fragment. More importantly, the confidence of an assigned fragment is dramatically increased,

where the signature M and M +11 peaks are very clear. Recently several groups have employed photo-crosslinking amino acids to capture unknown protein binding partners *in vivo* [108, 109]. Using a strategy such as the one described here it may now be possible to not only directly identify proteins via tandem mass spectrometry with rapid peak identification, but also determine the site of crosslink formation with single amino acid accuracy. Large-scale proteomic projects that map protein complex connectivities may benefit from such an approach.

The labelled [D_{11}]-pBpa compound synthesized in this work has been requested for use in the lab of Dr. Alan Warren at the MRC laboratories in Cambridge, England. His group has successfully utilized the applications of this work to assign crosslinked peptides in previously unknown protein-protein interactions in the eukaryotic ribosome biogenesis pathway.

Chapter 3: Photochemical control of fluorescent binding to a genetically encoded sequence tag with *o*-nitrobenzyl cysteine

3.1 Introduction

The discovery and evolution of green fluorescent proteins (GFP) has revolutionized the study of biology[110-113]. This protein can be engineered as a fusion tag to fluorescently label a protein of interest *in vivo*. The extensive array of proteins contains a wonderful palette of colors that allows multiple proteins to be tagged with different colors within the same cell. More recently, this collection of proteins has been expanded to include those that change optical properties upon irradiation resulting in photo-activatable GFPs (PA-GFP)[114]. These modified analogs of the wild-type protein exhibit a significant increase in fluorescence when stimulated to their active conformation[114]. Other photo-activatable proteins, similar to GFPs, have also been reported. A coral fluorescent protein has been modified to yield photo-activatable proteins capable of switching from red to green wavelengths (Kaede)[115] as well as exhibiting fluorescence recovery after photobleaching (FRAP) demonstrating the ability to flip-flop between fluorescent signal (Dronpa)[116]. These light induced "switches" provide spatiotemporal control over protein fluorescence. Provided an appropriate microscopy platform, one can track the movements of single protein molecules within a cell at sub-difraction resolution by photo-activation followed by fluorescence tracking[65, 117].

Of course, there are limitations to fluorescent proteins including brightness, rate of chromophore formation, and size (GFP contains 238 amino acids). There is always the concern that a genetic fusion tag this large may alter the function of the protein that is

being studied. It could disrupt a protein-protein interaction, or subcellular localization resulting in data that is not biologically relevant. The rate of GFP formation is also an issue as it must complete several rounds of self-modification to generate fluorophores, taking up to hours to become fluorescent[113]. Further, there are only two available colors of PA-GFPs currently available, limited to assays using only fluorescence spectrometry where the activation levels of PA-GFP are only reported to be a 100-fold difference in fluorescence compared to that of GFP[114].

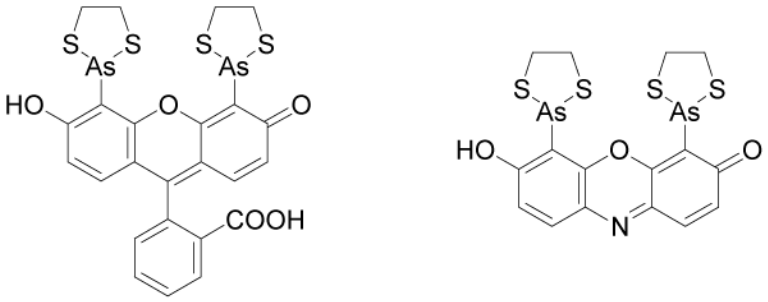


Figure 3-1. Chemical structures of FlAsH and ReAsH.

As an alternative, Tsien and co-workers introduced the biarsenical pro-fluorescent dyes 4,5-bis(1,3,2-dithioarsolan-2-yl)fluorescein and 4,5-bis(1,3,2-dithioarsolan-2-yl)resorufin (FlAsH and ReAsH respectively) that specifically bind to, and are activated by the small amino acid tag sequence, CCPGCC[118, 119] (Figure 3-1). They were inspired by the fact that such binding events were responsible for the toxicity of arsenic compounds, as they covalently bind to cysteine pairs that are in close proximity to one another. These toxic interactions are completely reversible by the introduction of small vicinal dithiols such as 1,2-ethanedithiol (EDT), which associate in a tighter binding-complex than cellular dithiols[120]. It was believed that a protein motif could be designed that had an even higher affinity for an organoarsenical compound than those observed

with EDT, in turn, allowing for a peptide binding domain specific for arsenical compounds, even in the presence of excess cellular dithiols. This domain was designed with four cysteines at a spatial distance as to confer proper orientation of the two thiol-arsenic interactions. Cysteines placed at the *i*, *i*+1, *i*+4, and *i*+5 positions allowed for cooperative and entropically favorable binding[119], allowing for the four thiol groups to react on one side of the helix in correct geometry with the arsenic ligand. This small genetically encoded sequence (TC tag) is orthogonal to cellular proteins and can be labeled extremely selectively and efficiently *in vivo*.

The first successful fluorescent binding partner to the newly designed motif was FlAsH. This nonfluorescent molecule, derived from fluorescein mercuric acetate, exchanges EDT for the tetracysteine motif generating a distinctly fluorescent protein-bound complex[118] (Figure 3-2), becoming more than 50 000 times more fluorescent. It is thought that the fluorescein compound is not fluorescent when bound to EDT due to vibrational quenching or photo-induced electron transfer mechanisms. Presumably the rigid conformation of the peptide-arsenical complex is too constrained to allow for quenching mechanisms to occur.

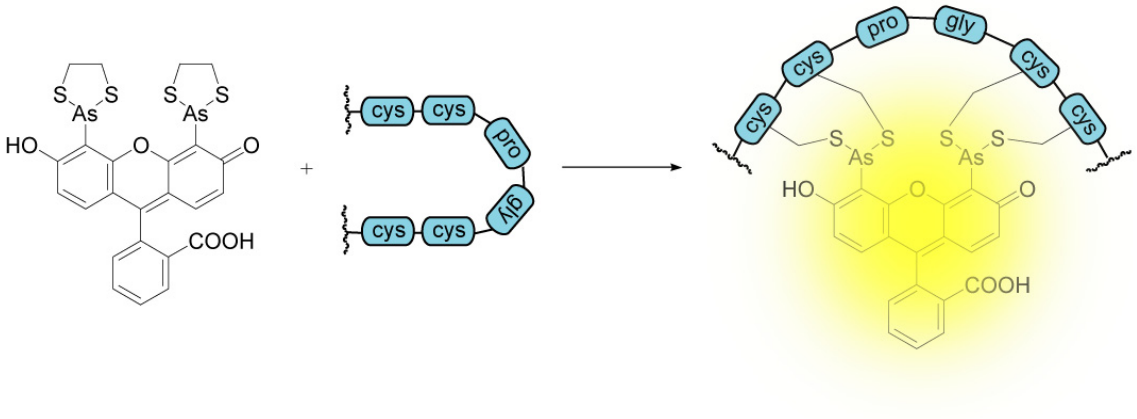


Figure 3-2. FlAsH labeling of tetracysteine motif on any protein of interest.

The nominal size (~700 Da) of the TC-tag has shown less protein perturbation as compared to the relative size of the fluorescent fusion proteins[121-123] and has been applied as a new approach to fluorescent resonance energy transfer (FRET) studies[124, 125]. These short tags have also been shown to serve as useful probes of protein-protein interactions and protein conformational changes using distal Cys-Cys pairs[126]. Whereas these dyes show advantages to GFP fusions, there is no mechanism for the spatiotemporal control of protein labeling, and therefore all proteins within the cell displaying the CCPGCC motif are labeled.

The tetracysteine sequence is critical for the function of these dyes and a single amino acid change causes a major decrease in fluorescence due to destabilization of the biarsenical complexes[118]. This requirement was taken advantage of to extend a level of control to the labeling and fluorescence events by temporarily masking the cysteine residues in the TC-tag. The genetically encoded unnatural amino acid *o*-nitrobenzyl cysteine (ONBC) can be site-specifically introduced into proteins in yeast cells in response to amber stop codons[19]. This "caged" unnatural amino acid *becomes* cysteine

upon irradiation with UV light (365nm) (as detailed in Section 1.1.2) and therefore tag labeling and activation of fluorescence should only occur after irradiation (Figure 3-3). Due to the fast association rate for these complexes (~310000 $M^{-1}s^{-1}$, fluorescence observed in seconds)[118] photo-activation of fluorescence is rapid and therefore might serve as a supplement to PA-GFP. In the case of precise light delivery, this might allow individual protein molecules to be selectively fluorescently labeled.

This chapter will outline the use of photocaged cysteines as a means of blocking small molecule binding to a tetra-cysteine tag. When FlAsH labeling is introduced in the presence of ONBC at a position in the tag, it will inhibit fluorescent labeling of the protein. Once the protein is subjected to light irradiation, the cysteine is freed from the protecting group, thereby allowing for fluorescent labeling of the protein. This light-activated binding adds a level of spatiotemporal control to small molecule binding events to protein (see Figure 3-3).

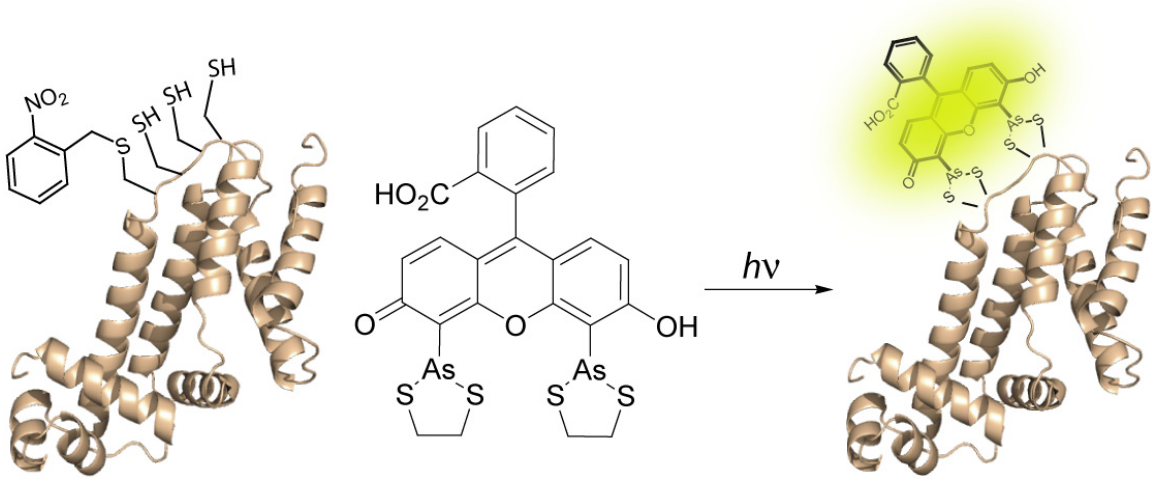


Figure 3-3. Process of UV-induced activation of fluorescence. A caged cysteine mutation in place of one of the cysteines in the TC-tag results in the inhibition of fluorescence. Upon irradiation, the caging group is removed resulting in dithiol exchange, and fluorescence labeling.

3.2 Experimental

3.2.1 Materials

Reagents for the synthesis of FlAsH were purchased from Fisher. All enzymes were purchased from New England Biolabs and Fermentas. Sequencing was performed at the University of Michigan sequencing facility. All other reagents were purchased commercially and used without further purification. The plasmid pESC-PC-Cys was supplied by Dr. Peter Schultz (The Scripps Institute, La Jolla, CA).

3.2.2 General Methods

FlAsH, 4,5-bis(1,3,2-dithiasolan-2-yl)fluorescein, was synthesized and characterized as previously described[127]. The compound was dissolved in 100% DMSO to a final concentration of 10 mM for use. PCR reactions were performed with Deep Vent DNA polymerase (NEB) under the following conditions for 30 cycles: 30 s at 95 ^{o}C, 30 s at a calculated annealing temperature (- 5^{o} of the lowest oligo Tm), and extension time of 1 min/Kb at 72 ^{o}C. DNA isolations were performed using QIAGEN mini-prep kits according to manufacturer's protocols. . Ligations were performed using T4 ligase under standard conditions. All *E. Coli* transformations were performed into GeneHogs (Invitrogen) electro-competent cells using ~100 ng of DNA transformed into 100 μL of cells, and into InvScI chemical-competent cells (Invitrogen) for transformations into *S. cerevisiae*. *o*-Nitrobenzyl cysteine was dissolved in water with 0.5 M NaOH added dropwise until fully dissolved at room temperature, prior to being added to expression media. Expressions for the TAG mutants were carried out in yeast SD media containing 2% glucose in the absence of uracil and tryptophan. All protein concentrations determined by BSA kit (Pierce). UV irradiation (365 nm) was performed with a hand-

held 100W Blak Ray lamp used at a distance of 5 cm, at 4 ^{o}C, for 30 min. Purification of GST was performed using Probond Purification resin (Invitrogen) according to the manufacturer's protocol for native isolation. SDS-PAGE gels (12 % acrylamide resolving and 5% acrylamide stacking) and 1X Laemmli buffers were used following standard procedures, performed at 200 V for 65 min. SDS-PAGE gels were stained with Coomassie Brilliant Blue.

3.2.3 GST-tetracysteine tag plasmid construction

The gene encoding *S. japonicum* GST was amplified using the following primers to introduce the N-terminal tetracysteine tags containing TAG mutations (underlined) and a C-terminal 6 X His tag : WT (no TAG mutations), FWD 5'- ATA TTA AGC TTA CCA TGG GTT GTT GTC CAG GTT GCT GTT CCC CTA TAC TAG GTT ATT GG-3'; C3TAG, FWD 5'-ATA TTA AGC TTA CCA TGG GTT AGT GTC CAG GTT GCT GTT CCC CTA TAC TAG GTT ATT GG-3'; C4TAG, FWD 5'-ATA TTA AGC TTA CCA TGG GTT GTT AGC CAG GTT GCT GTT CCC CTA TAC TAG GTT ATT GG-3'; C7TAG, FWD 5'-ATA TTA AGC TTA CCA TGG GTT GTT GTC CAG GTT AGT GTT CCC CTA TAC TAG GTT ATT GG-3'; C8TAG, FWD 5'-ATA TTA AGC TTA CCA TGG GTT GTT GTC CAG GTT GCT AGT CCC CTA TAC TAG GTT ATT GG-3'; REV (for all) 5'-AAA TCT AGA TCA ATG GTG ATG GTG ATG GTG GTC ACG ATG CGG CCG CTC G. These were each cloned into the yeast expression vector pYES2 (Invitrogen, uracil marker) using the restriction enzymes HindIII and XbaI. The expression vectors were cotransformed along with the plasmid pESC-PC-Cys[19].

3.2.4 Protein expression and isolation

GST-TC tag expressions were carried out in yeast SD media containing 2% glucose in the absence of uracil and tryptophan. The cells were allowed to grow at 30 ^{o}C until OD_{600} = 1.8-2.0 then pelleted and washed with sterile water to remove remaining media containing glucose. The cell pellets were resuspended in media containing 2% galactose (to the same OD_{600} to induce expression) in the absence of uracil and tryptophan, supplemented with 2 mM *o*-nitrobenzyl cysteine (WT carried out in only the absence of uracil with no unnatural amino acid added). Expressions were performed at 30 ^{o}C for 48 hrs in the dark. Cells were lysed for protein purification using acid washed glass beads. Cell pellets were resuspended in 1 mL of native binding buffer (Promega His-tagged resin protocol) where an equal volume of glass beads was added. Shearing was done via vortex at high speed for 15 min. Following cell lysis, cellular debris was pelleted at 15,000 rpm for 30 min. The lysate was then purified using Promega Ni^{2+} resin following the manufacturer's protocol.

3.2.5 Fluorescence labeling studies

For each isolated protein containing ONBC half of the protein obtained was subjected to UV irradiation. For the GST-TC-WT, and each of the ONBC proteins (+/- UV irradiation), approximately 2 μM of appropriate protein used for labeling. Protein was added to SDS loading buffer and subjected to 70 ^{o}C for 10 min. FlAsH dye (in DMSO) was then added to a final concentration of 0.5 μM, and allowed to incubate at 70 ^{o}C for 3 min. All samples were then loaded onto an SDS-PAGE gel. The SDS-PAGE gel was scanned in fluorescence mode Blue (450 nM)/ 520 LP, 1000 V, on high sensitivity (Molecular Dynamics Storm 860 Phosphoimager). For fluorimeter studies the

the reaction solution 3.5 μM tris(2-carboxyethyl)phosphine (TCEP), 10 μM EDT, 1 μM FlAsH, 0.5 μM protein, in 100 mM MOPS (pH 7.2) was used (final concentrations listed). Each labeling reaction was allowed to incubate at room temperature for 1 hour prior to scanning. The emission scan was set for an excitation of 508 nm and scanned from 510-600 nm (Ex slit 10.0 nm, Em slit 5.0 nm, PMT voltage 950 v, Hitachi F-4500 FL Spectrophotometer).

3.3 Results and Discussion

3.3.1 Light-activated labeling of TC-tagged protein

Fluorescence labeling with FlAsH occurs through four covalent bonds between the thiols of each cysteine with the two arsenic groups of the dye (see Figure 3-2). This 1,2-dithiol exchange for the four cysteines requires the ability of the side-chain sulfur to be reduced for the binding event to occur. Any change in the *i, i*+ 1, *i* + 4, *i* +5 cysteine arrangement results in the loss of the ability to exchange thiols reducing the quantum yield by up to 500%[118]. The introduction of a caging group on one of the essential cysteines should inhibit the ability of these exchange events.

To test this hypothesis the *Schistosoma japonicum* gene encoding glutathione-*S*-transferse (GST) was PCR amplified such that it contained an N-terminal TC-tag and a C-terminal 6X-histidine tag for purification purposes. In addition, four synthetic genes were created that contain amber stop codons (TAG) in place of either Cys 1, 2, 5, or 6 of the TC-tag (using amino acid numbering in the actual tag). These five genes were then expressed from the yeast expression vector pYES2 in yeast carrying an aminoacyl-tRNA synthetase/tRNA pair specific for ONBC[19]. The orthogonal tRNA/aaRS pair used in this system was originally evolved in yeast. Because FlAsH-protein complexes are stable

under denaturing gel electrophoresis conditions, the fluorescent binding events were first established via SDS-PAGE followed by subsequent fluorescence imaging. FlAsH-protein binding, and thus fluorescence, should be visible at the correct molecular mass of the protein (~28 kDa).

As can be seen in Figure 3-4, fluorescence is observed for GST-TC carrying the wild-type tag with four cysteines. No labeling is seen in protein bearing tags that are masked with ONBC indicating that these mutants are not effectively labeled with FlAsH, despite possessing three of the four cysteine residues. When these same samples are irradiated with 365nm light for 15 min however, clear labeling of the proteins can be seen. Negligible differences of intensity are seen, indicating that the activation ratio of the four mutants is essentially the same. If any difference can be identified, it is that the protein containing the mutation in place of Cys 2 expresses at a higher level than the others, potentially due to codon context effects.

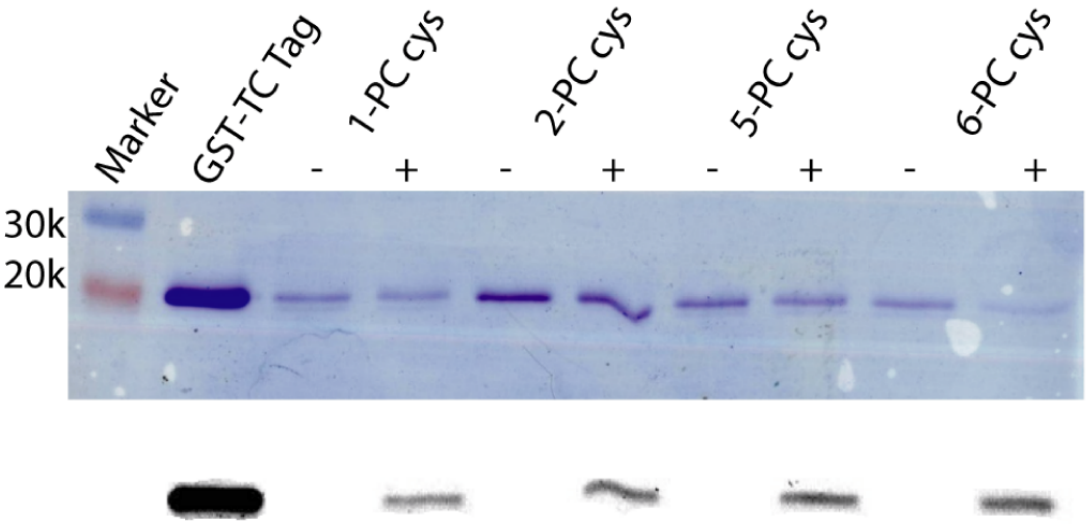


Figure 3-4. SDS-PAGE analysis of FlAsH labeling. Top panel shows a Coomasie stained gel. Bottom panel shows the same imaged for fluorescence (excitation 450 nm). Protein samples contain a TC tag or a caged tag with the location of the unnatural amino acid mutation shown. -/+ indicates the absence/presence of irradiation at 365nm for 15 min.

These results verify that all four cysteine residues are essential for the proper alignment and association of the arsenical compound within the binding domain and that the introduction of a caging group blocks one thiol-arsenic exchange. It is believed that the rigid nature of the complex between FlAsH and the TC-tag actually inhibits a quenching mechanism that occurs when the compound is in complex with EDT. As expected, the introduction of one ONBC molecule at an essential cysteine blocks the exchange events of an arsenic atom on the compound. This was observed at all four potential cysteines indicating that all four residues are required for the fluorescent labeling. Even if the free cysteines (three remaining non-caged residues) are open to undergo chemical interactions, inhibition of fluorescence is observed due to the inability of all four cysteines to form the constraints required to hinder quenching. Photolysis of the caging group is promoted by irradiating with 365 nm affording native cysteine, facilitating the binding of the FlAsH dye. This confirms the ability to control the fluorescent labeling of a TC-tag with FlAsH dye leaning to the notion that this technique may become a powerful application in spatiotemporal control of cellular imaging.

3.3.2 Fluorescence quantification

In order to quantify this activation of protein labeling, purified protein was used to measure the fluorescence intensity derived from the TC-biarsenical complexes following the thiol exchange reaction. Purified protein was again subjected to irradiation (where half was not introduced to decaging events) and was added to an MES buffer solution containing 10 μM EDT and 1 μM FlAsH compound. The specificity of FlAsH binding is improved by the addition of EDT to the solution[119] because it reduces nonspecific labeling with endogenous cellular sites that would recognize other arsenoxide (R – As =

O) compounds for which the motif binding site was based. The labeling was allowed to incubate at room temperature for one hour as fluorescence increases rapidly but requires time to reach maximum levels. An emission scan was performed to determine the fluorescent counts for each labeling event. Concurrent with the data observed for the SDS-PAGE, there was a significant increase in fluorescence (~1000 fold at these concentrations, Figure 3-5) upon decaging for those proteins containing the unnatural amino acid, a value comparable to previous examples using FlAsH. The nonfluorescent FlAsH compound has been reported to increase fluorescence by up to 50000 times upon tetracysteine binding, where we observe a nearly 1000 fold increase, similar to that reported for the optimized sequence, FLNCCPGCCMEP (H2N-Phe-Leu-Asn-Cys-Cys-Pro-Gly-Met-Glu-Pro-CONH2)[127]. Again, there appears to be no apparent difference in the activation ratio between the four positional mutants, indicating that the presence of four thiols is an absolute requirement and that each of the four mutants decages equally well.

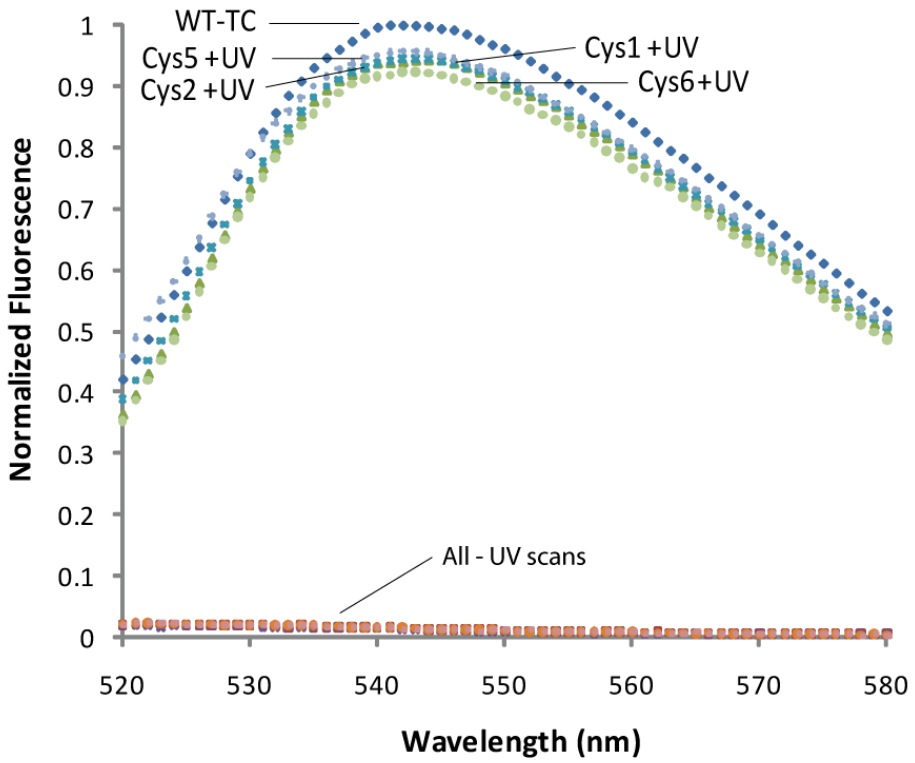


Figure 3-5. Emission scan of TC-tagged proteins (excitation 508nm). The baseline traces are those of no protein (dye only), and all four mutants in the absence of UV irradiation.

3.4 Conclusion

In summary, it has been shown that the ubiquitous FlAsH labeling event can be controlled with photo-caged unnatural amino acids. This represents an attractive alternative to PA-GFP tagging that is smaller, and less likely to disrupt protein function. In addition to FlAsH, several other derivatives of biarsenical dyes have been reported including the red dye ReAsH[118], opening the possibility for multicolor imaging. Indeed this approach might facilitate sequential labeling of multiple TC tags on the same protein or within a cell. Further biarsenical chemistry has recently been used for purposes outside of simple fluorescence labeling such as site-specific bioconjugations and even orthogonal small-molecule control of enzyme function[128]. The use of photo-caged TC-

tags may now bring an unprecedented level of control to these applications as well.

Chapter 4: Site-specific incorporation of 2-fluorotyrosine into *E. coli* via photochemical disguise

4.1 Introduction

Fluorinated amino acids are powerful tools that have been used to investigate protein structure and function. Importantly, the small size of the fluorine atom is expected to not perturb the protein structure, while its high electronegativity and NMR active nuclear spin enable its application as a very precise biological probe. When incorporated into a protein, the fluorine atom provides a unique resonance for structurally investigating selected regions of a protein using ^{19}F NMR spectroscopy. The chemical shielding of this nucleus is highly responsive to the molecular environment and can report on protein conformational changes, and thus folding and unfolding processes[129]. ^{19}F NMR experiments benefit from the high sensitivity and 100% natural abundance of the ^{19}F isotope, thus minimizing complications by background signals from non-labeled biomolecules. The sensitivity is such that it can even be used to track proteins *in vivo* using live cells as the NMR sample[59]. Depending on the amino acid used, the structural disruptions to a protein can be minimized as fluorine is a conservative substitute for hydrogen atoms. Indeed, fluorinated analogues of tyrosine, histidine, phenylalanine and tryptophan have all been incorporated into proteins for this purpose. However, proteins containing extensive fluorination can behave differently,[130] and assignment of ^{19}F resonances in large proteins becomes increasingly complex with increasing numbers of fluorinated residues.

In addition to being an excellent NMR probe, the high electronegativity of fluorine can be used to affect changes in the chemical environment of a protein active

site. As discussed in Chapter 1, the pKa of the tyrosine phenolic proton is approximately 10, but the pKa of fluorotyrosines can range from 5.2-9.0 depending on the extent of fluorination[131, 132]. Therefore, the acidity of individual side chains within a protein can be precisely modulated to investigate the participation of a given residue in acid/base catalysis mechanisms. The same is true of the redox properties of tyrosine residues[64]. In the study of biologically generated tyrosyl radicals, there are no structurally similar amino acids with alternative reduction potentials in the genetic code. Fluorotyrosines have peak reduction potentials that range from 705-968 mV, depending upon the extent and position of fluorination. This allows fluorotyrosines to be used as comparison probes to tyrosine, which has a peak potential of 642 mV[64].

The main limitation preventing the extensive utilization of fluorinated amino acids in the investigation of enzyme function is the process by which they are currently being introduced into proteins, as it is challenging to achieve a site-specific incorporation, while still obtaining significant amounts of protein. Due to the similarities of fluorinated and natural amino acids in size and shape – fluorine has only a 0.15 Å larger van der Waals radius than hydrogen – fluorinated amino acids are often introduced into proteins via global incorporation when introduced to cells. For example, when culturing *E. coli* cells starved of tyrosine in the presence of 2- or 3-fluorotyrosine, the fluorinated amino acids are incorporated into proteins in place of all tyrosine residues[58, 133] (Figure 4-1). This process requires auxotrophic strains of *E. coli* and correct timing of the addition of fluorotyrosine to the media to prevent global incorporation into all cellular proteins. The obvious concern for metabolic labeling is that there is no site-specificity and the sample proteins can contain a multiplicity of fluorinated amino acid substitutions. This can result

in sample heterogeneity and could make it impossible to link functional consequences to a single residue.

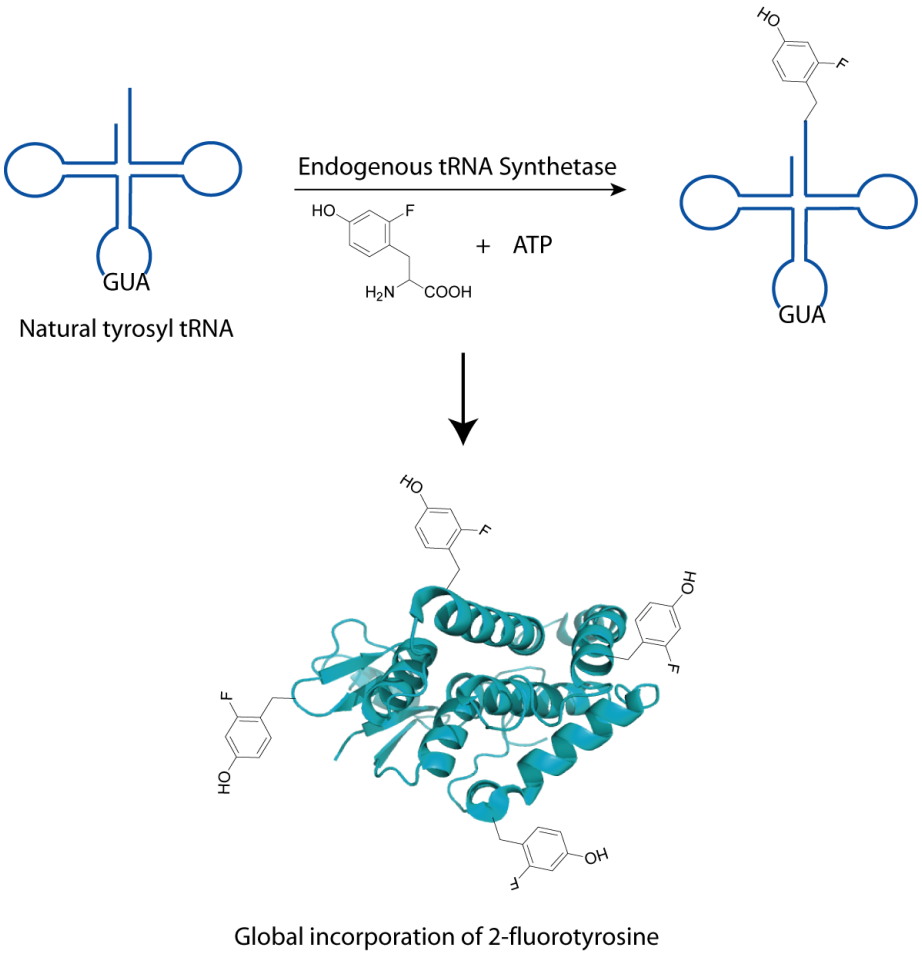


Figure 4-1. General incorporation of 2-fluorotyrosine, which is recognized by the endogenous tyrosyl machinery, encoding fluorotyrosine throughout the protein at all UAC codons.

Several attempts have been made to combat this problem of poor site-selectivity. Homogeneous proteins containing fluorotyrosines have been produced by a combination of chemical peptide synthesis and expressed protein ligation[132]. Unfortunately, this approach is limited to certain locations within a protein, typically at the C-terminus, and can be technically challenging for proteins that are sensitive to denaturation. Previously unnatural amino acid mutagenesis has been performed in an *in vitro* protein expression system using amber suppressor tRNAs that are chemical aminoacylated with

fluorotyrosines[134]. This approach is limited by the requirement for a laborious synthesis of the aminoacylated tRNA and the small protein yields typically obtained.

A more versatile method would be to use *in vivo* unnatural amino acid mutagenesis based on an orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA enabling a protein expression in *E. coli*.[135] This would limit the amount of site-specifically fluorinated protein obtainable only by the size of the bacterial culture and enable the protein production using standard molecular biology techniques. Such an approach, however, is complicated by the fact that NAAs that are similar in structure to endogenous amino acids infiltrate cellular protein biosynthesis. As a solution to this problem, the NAA acid is temporarily converted to a full UAA by the addition of an *o*-nitrobenzyl group. This in turn, masks the fluorinated amino acid from the cellular metabolism using a photo-removable protecting group[34, 35, 136] (Figure 4-2). A chemoenzymatic synthesis of *ortho*-nitrobenzyl-2-fluorotyrosine (**4-1**) was performed for introduction of this amino acid into proteins in response to the amber stop codon, TAG.

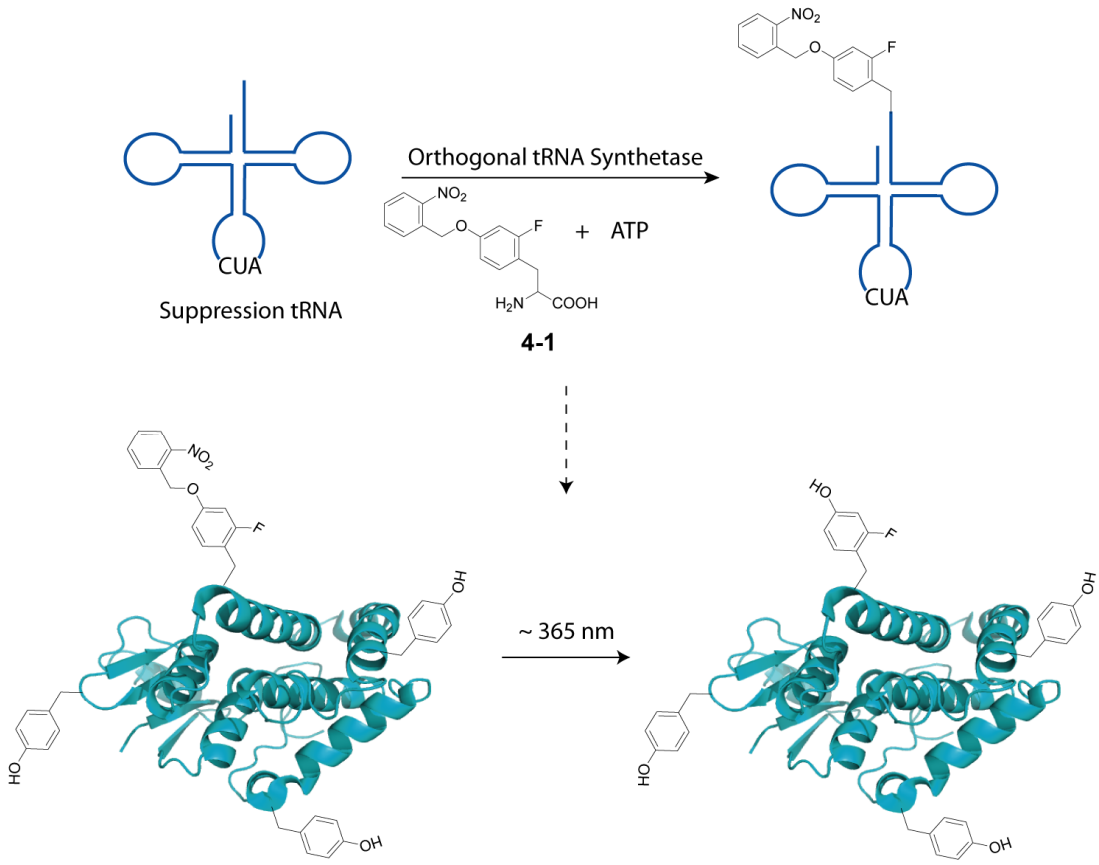


Figure 4-2. Amber suppression for the site-specific incorporation of *o*-nitrobenzyl-2-fluorotyrosine. Once the caged amino acid is encoded it is photo-lysed to afford a singular fluorotyrosine protein.

4.2 Experimental

4.2.1 Materials

Reagents for the synthesis of 2-fluorotyrosine, *o*-nitrobenzyl-2-fluorotyrosine (**4-1**), and *o*-nitrobenzyl tyrosine (**4-2**) were purchased from Fisher. All other reagents were purchased commercially and used without further purification**.** All enzymes were purchased from New England Biolabs and Fermentas. Sequencing was performed at the University of Maryland at the Center for Biosystems research. The positive and negative selection plasmids, as well as pSUP-ONBY, was supplied by Dr. Peter Schultz (The Scripps Institute, La Jolla, CA). The plasmid pTZTPL, which houses the gene for

tyrosine phenol-lyase, was supplied by Dr. Robert S. Phillips (University of Georgia, Athens, GA).

4.2.2 General Methods

The antibiotics employed were kanamycin (25 μg mL^{-1}), tetracycline (25 μg mL^{-1}) and ampicillin (100 μg mL^{-1}). PCR reactions were performed with Taq DNA polymerase under the following conditions for 30 cycles: 95 °C for 30 s, 55 °C for 30 s, 72 °C for 4 min. DNA isolations were performed using Zyppy miniprep kits (Zymo Research). Ligations were performed using T4 ligase under standard conditions. All library transformations were performed into DH-10B electro-competent cells using ~100 ng of DNA transformed into 100 μL of cells. Standard transformations for plasmid amplification were performed into GeneHogs (Invitrogen) electro-competent cells using ~100 ng of DNA transformed into 100 μL of cells. **4-1** and **4-2** were dissolved in 50% DMSO in water (v/v) with 0.5 M NaOH added dropwise until fully dissolved at ~80 °C. The solution of amino acid was then added to hot agar (or liquid) with stirring and allowed to cool to 40 °C before the addition of antibiotics (and cells for expressions). Denaturing discontinuous sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels (12% acrylamide resolving and 5% acrylamide stacking) and 1X Laemmli buffers were used following standard procedures, performed at 200 V for 65 min. Purification of protein was performed using Probond Purification resin (Invitrogen) according to the manufacturer's protocol for native isolation. SDS-PAGE gels were stained with Coomassie Brilliant Blue. GMML media was composed of the following: 1.5% water agar, 1 % glycerol, 1 X M9 salts, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 0.3 mM leucine. For samples subjected to UV irradiation, and subsequent loss of the *o*-

nitrobenzyl group, a hand-held 100W Blak Ray lamp was used at a distance of 5 cm, at 4 ^{o}C, for 30 min. Emission scans were set for an excitation of 482 nm for **4-1** and 490 nm for **4-2.** Scanned was performed from 500-650 nm (Ex slit 10.0 nm, Em slit 5.0 nm, PMT voltage 950 v, Hitachi F-4500 FL Spectrophotometer).

4.2.3 Synthesis of o-nitrobenzyl-2-fluorotyrosine

o-Nitrobenzyl-2-fluorotyrosine (ONB-2FY) was synthesized from 3-fluorophenol and characterized as previously described[62]. Tyrosine phenol-lyase was utilized to enzymatically manufacture ONB-2FY. The enzyme was over-expressed from the plasmid pTZTPL which housed the gene for the protein. The cell pellet was resuspended, and then lysed, in the following buffer: 0.1 M potassium phosphate, pH 7.0, 0.1 mM pyridoxal phosphate, imM EDTA, and 5 mM β-mercaptoethanol. The cellular debris was centrifuged, and the soluble fraction was added to a 1 L aqueous reaction containing the following: 3-fluorophenol (5 g, 44.6 mmol), pyruvic acid (60 mM), pyridoxyl-5'-phosphate (10 mg L^{-1}), ammonium acetate (30 mM), and β-mercaptoethanol (5 mM). The reaction was performed for approximately 5 days at room temperature in the dark. It was monitored by HPLC for the disappearance of phenol (t_R ~15 min) and the appearance of the product (t_R ~ 4 min). The solution was then purified by ion exchange to afford ONB-2FY. ^{1}H NMR (400 MHz, D_2O); δ 6.88 (dd, *J* = 10.17 Hz, J = 8.32 Hz, 1 H), 6.31 (dd, J = 8.32 Hz, J = 2.31 Hz, 1 H), 6.23-6.27 (m, 1 H), 3.35 (dd, *J* = 7.40 Hz, J = 5.55 Hz, 1 H), 2.81 (dd, *J* = 13.87 Hz, *J* = 5.55 Hz, 1 H), 2.64 (dd, *J* = 13.87 Hz, *J* = 7.40 Hz, 1 H). HRMS (ESI) m/z calculated for $C_9H_{10}NO_3F$, 199.0645, found 199.0441.

The *o*-nitrobenzyl protecting group was installed similar to reported methods[137]. Briefly, 2-fluorotyrosine (3.37 g, 16.9 mmol) was dissolved in a 20 mL solution of NaOH

(2.03 g, 50.8 mmol) in water. $CuSO_4$ (1.35 g, 8.47 mmol) was dissolved in 5 mL of water, separately. These two solutions were equilibrated to 60 °C and then the $CuSO_4$ solution was added to the stirring solution of 2-fluorotyrosine. The reaction mixture was maintained at 60 °C for 15 minutes. The solution was then allowed to cool to room temperature and the pH was adjusted to 7.0 using acetic acid. The precipitate was gravity filtered, washed with water, and dried under vacuum. This afforded the copper complex $Cu(2FY)_2$ (3.35 g, 86.2%). The $Cu(2FY)_2$ (3.35g, 7.29 mmol) complex was dissolved in a 200 mL solution of 75% DMF in water (v/v) and K_2CO_3 (2.02 g, 14.6 mmol). 2-Nitrobenzylbromide (3.15 g, 14.6 mmol) was added and the reaction stirred for 48 hours at room temperature in the dark. The precipitate was gravity filtered and then washed with 75% DMF in water (v/v) followed by cold 1 N HCl and water. This product was then stirred in 100 mL of 1 N HCl for two hours at room temperature in the dark. The amino acid was re-crystallized from 8:1 water/ethanol (v/v) with a couple drops of 1 N HCl to afford the desired white crystalline product (2.73 g, 61%, 53% overall). ^{1}H NMR (400 MHz, DMSO); δ 8.13 (d, J = 8.5 Hz, 1 H), 7.76-7.81 (m, 2 H), 7.62-7.66 (m, 1 H), 7.25 (t, J = 9 Hz, 1 H), 6.92 (dd, J = 12 Hz, J = 2 Hz, 1 H), 6.82 (dd, J = 8.5 Hz, J = 2.5 Hz, 1 H), 5.46 (s, 2 H), 3.33 (bs, 3 H), 3.12 (dd, J = 14.5 Hz, J = 6.0 Hz, 1 H), 2.87-2.93 (m, 1 H). HRMS (ESI) m/z calculated for $C_{16}H_{15}N_2O_5F$, 334.0965, found 334.0956.

4.2.4 Directed evolution of aa-tRNA synthetase specific for ONB-2FY

An active site library of *M. janaschii* aminoacyl-tRNA synthetases was constructed on plasmid pBK which encodes the gene for *Mj* YRS under the control of the *Escherichia coli* GlnRS promoter and terminator. For the library the following amino acids were randomized with mixed codons: Leu65, His70, Ile159, and Phe108 to all 20

possible amino acids (NNK codon); Tyr32 to all amino acids except Trp, Phe, Tyr, or Cys (VNK codon); Asp158 was limited to Gly, Ile, Ser, or Val (RKY codon); Leu162 was limited to all polar amino acids (VRK codon). The codes used above for the randomization of specific bases are as follows: N, all possible amino acids; K, only the bases G or T; V, only the bases A, C, or G; R, only the bases A or G; and Y, only the bases C or T. Four rounds of sequential enzymatic inverse PCR were performed to obtain the library with the product of one round serving as the template for the next round. The following primers were used for each round in which a unique *BsaI* site was introduced to each (italics): Round 1 (Tyr32) FWD 5'- GCG CAG GAA A*GG TCT C*AG AAA AAT CTG CTV NKA TAG GTT TTG AAC C and REV 5' – GCG CAG AGT A*GG TCT C*AT TTC ATC TTT TTT TAA AAC CTC TCT TAA CTC TTC C; Round 2 (Phe108) FWD 5' – GCG CAG GAA A*GG TCT C*AG GAA GTG AAV NKC AGC TTG ATA CGG ATT ATA CAC and REV 5' – GCG CAG AGT A*GG TCT C*AT TCC ATA AAC ATA TTT TGC CTT TAA CCC C; Round 3 (Asp158, Ile159, Leu 162) FWD 5' – GCG CAG GAA A*GG TCT C*AC CAA TAA TGC AGG TTA ATR KYN NKC ATT ATV RKG GCG TTG ATC TTG C and REV 5' – GCG CAG AGT A*GG TCT C*AT TGG ATA GAT AAC TTC AGC AAC CTT TGG; Round 4 (Leu65, His70) FWD 5' – GCG CAG GAA A*GG TCT C*AG ATA TAA TTA TAN NKT TGG CTG ATT TAN NKG GCT ATT TAA ACC and REV 5' - GCG CAG AGT A*GG TCT C*AT ATC AAA TCC AGC ATT TTG TAA ATC AAT C. The purified PCR products were treated with *DpnI* to remove the parental DNA. The enzyme *BsaI* was then used to digest the products, which were then ligated, and transformed into *E. coli*. The obtained library size was $>10^8$ colonies. The final library DNA was isolated and transformed into positive

selection cells containing a plasmid (pREP2/YC-JYCUA) encoding a mutant *Mj*tRNA$^{Tyr}_{CUA}$, the gene encoding chloramphenicol acetyltransferase gene with an in-frame TAG stop codon, and the gene T7 RNA polymerase also containing an in-frame TAG stop codon which drives expression of a T7-controlled copy of green fluorescent protein, and a tetracycline resistant marker[84].

The selection cells containing the library were grown in LB supplemented with Kan and Tet to a final $OD_{600} = 0.8$. 1 mL of these cells was plated on four 150 mm plates containing GMML agar supplemented with Tet, Kan and Chl, 0.002% arabinose and a final concentration of 1 mM **4-1** or **4-2** (Scheme 4-1). These plates were incubated at 37 ^{o}C for 48 hrs in the dark and then the cells collected and plasmid DNA isolated. The pBK library was purified via gel electrophoresis. The DNA was then transformed into negative selection cells containing a mutant tRNA$^{Tyr}_{CUA}$, with a TAG stop codon in the barnase gene under control of an arabinose promoter and rrnC terminator, and an ampicillin resistance marker. These transformations were plated on LB agar containing Amp and 0.002% arabinose. Following incubation at 37 °C for 24 hrs, the cells were collected, the library DNA isolated and purified by gel electrophoresis. The DNA from the negative selection round was transformed back into positive selection cells and plated as described above for the first round of positive selections only with 90 μg mL^{-1} chloramphenicol. After two rounds of positive and negative selection, individual colonies were inoculated into a 96-well block and grown in LB media containing Tet and Kan. The cells were screened on GMML Agar supplemented with Tet, Kan and Chl 35 μg mL^{-1}, 0.002% arabinose and a final concentration of 1 mM of **4-1** or **4-2**. Two potential hits (E8 and E10, named ONB-2FYRS-1 and -2, respectively) were observed

and the DNA from these single colonies was isolated and sequenced. The gene encoding ONB-2FYRS-1, which was used for all protein expressions, was excised with *Pst*I and *Nde*I and ligated into the same sites of pSup[138], to generate pSup-ONB-2FYRS-1.

4.2.5 Expression of sfGFP

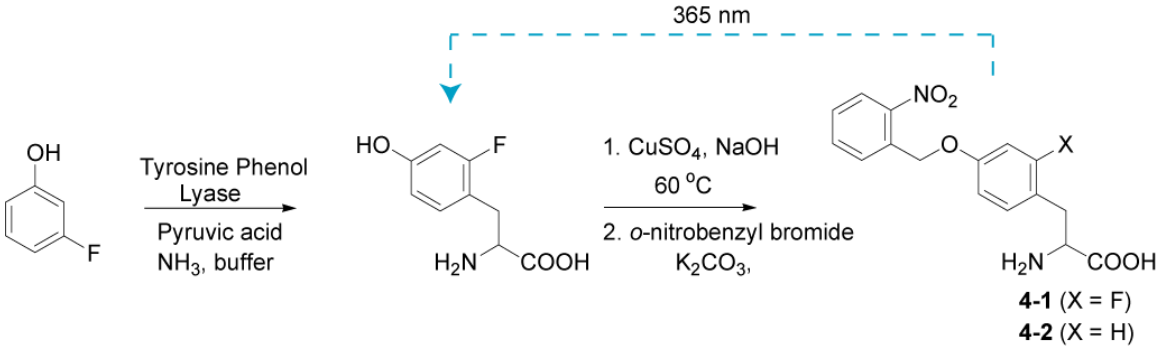
pSup-ONB-2FYRS-1 was transformed into *E. coli* cells along with pTrc-HisA-sfGFP66TAG (or pTrc-HisA-sfGFP150TAG for mass spec studies) (See appendix for plasmid construction). These expressions were performed in 2X-YT media (50 mL) supplemented with Amp, Chl, and a final concentration of 1 mM of **1**. The cells were grown to an OD_{600} = 0.8 and then induced with a final concentration of 1 mM IPTG. Expression was continued at 37 °C, overnight in the dark. Purification of sfGFP was performed using the HisLink purification system (Promega) under native conditions according to the manufacturer's protocol. Isolated protein was then dialyzed against 20 mM TRIS buffer (pH 7.2), analyzed via SDS-PAGE electrophoresis, and used for fluorescence experiments.

4.3 Results and Discussion

4.3.1 Synthesis of ONB-2FY and substrate recognition with pSUP-ONBY

For protein expression multi-gram quantities of ONB-2FY (**4-1**) were needed and therefore it was decided to produce the amino acid using a combination of chemical and enzymatic synthesis. In addition to being scalable, the procedure can be easily adapted to make other fluorotyrosine analogues. Using 3-fluorophenol as the starting material enantiomerically pure **4-1** was synthesized using the enzyme tyrosine phenol lyase[139] and then protected as a copper complex. This complex was then directly subjected to

alkylation with 2-nitrobenzylbromide resulting in the desired amino acid containing a photo-removable protecting group (Scheme 4-1). With sufficient amino acid now synthesized, expressions were performed in the presence of this compound, with genes containing a mutation at a position determined by the amber stop codon mutation, TAG.



Scheme 4-1. Synthesis of ONB-2FY. The addition on *o*-nitrobenzyl bromide was also applied to tyrosine, where X is indicative of the species mentioned in this text. The caging group is removed with exposure to 365 nm light. Step one had a 74% yield, while step two had a 61% yield.

Previously the *Mj* tyrosyl-tRNA synthetase was evolved to accept **4-2** as a substrate, suppressing the amber codon with the UAA[20]. It was assumed that with such a small structural change to the substrate, **4-1** would be accepted by this enzyme and therefore the amino acid could be inserted into proteins using the previously described machinery. This simple change of a single hydrogen to fluorine atom is indeed a conservative substitution since 2- and 3-fluorotyrosine are both substrates for the endogenous *E. coli* TyrRS, hence the requirement for photo-caging in the first place.

Surprisingly, after repeated attempts to produce protein using ONB-YRS, expressions were unsuccessful. This was indeed an unexpected result as there was no intent to have to evolve a new synthetase for suppression with **4-1**. Since this UAA would not act as a substrate for the ONB-YRS, the only alternative was the generation of a new tRNA/aaRS pair.

4.3.2 Evolution of a tRNA/aaRS pair for the recognition of ONB-2FY

Taking advantage of the selection methods detailed in Chapter 1.3.2, a library was constructed in which the residues within the active site of the *Mj* YRS (Y32, L65, H70, F108, D158, I159 and Ll62) were randomized or selectively diversified yielding $>10^8$ variants. This library was built through successive rounds of inverse PCR, where each consecutive round acted as the template for the next. These mutations were based on those, which were identified for the evolution of *Mj* YRS to recognize **4-2** (ONB-YRS), all of which are considered important residues in the active site binding domain (Figure 4-3A). Y32 and D158 are especially critical as they form hydrogen bonds with the hydroxyl group of the cognate tyrosine. Random mutations at these sites hoped to provide an expanded space for access of a larger amino acid, as well as displacing the hydrogen bond associations at the phenol. The observed mutations in ONB-YRS were as follows: Y32G, L65G, F108E, D158S, and L162E. These mutations indeed disfavor the recognition of native tyrosine as the D158S and Y32G changes likely disrupt the hydrogen bonding needed for the substrate recognition. As a caged group was introduced to the compound a larger active site is warranted for binding to the synthetase. The Y32G and L65G mutations seem to fit this role and increase the size of the binding pocket to better house the bulky addition. The compound to be selectively paired with a new aaRS has nearly equivalent structural identity with the mentioned system, so it was assumed that only slight mutations at the same positions would be necessary to facilitate this evolved recognition event. It should be noted that the original ONB-YRS was derived from a library which did not include randomized positions H70 and I159, resulting in a slightly different starting pool of synthetase mutants. These new introductions to the randomized library were also chosen based on the crystal structure of

the *Mj* YRS evolved to recognize pBpa (Figure 4-3b). H70 sits directly adjacent to the phenylalanine benzene group when bound in the active site. The electronic interactions between the imidazole group and the fluorine on the tyrosine could be too disruptive for binding events to occur. Altering this site was thought to allow for less sterics. The I159 was mutated to help potentially open the binding pocket a little further. In addition, these mutations are part of a large diverse library that selectively partitions for the best fit, so there was no risk in diversifying the library even greater than its previous efforts.

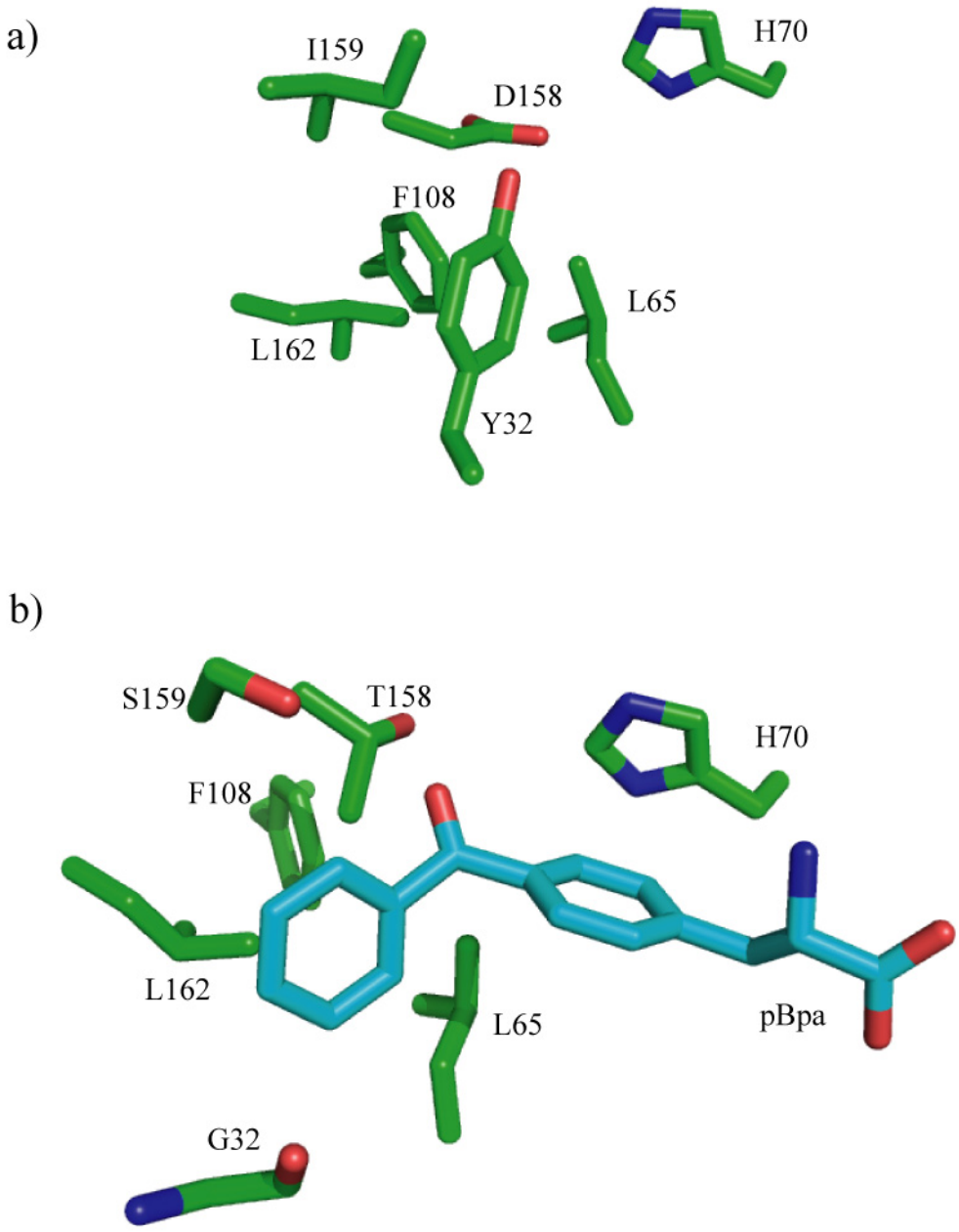


Figure 4-3. Crystal structure of the binding pocket of (a) wild-type *Mj* TyrRS (PDB 1J1U), and (b) pBpa *Mj* TyrRS bound to pBpa, shown in blue (PDB 2HGZ).

The new library was applied in a double-sieve selection designed to isolate a new enzyme that was capable of accepting **4-1** as a substrate, but no endogenous amino acids. The pool of variants was repeatedly passed through a positive selection (based on

suppression of a TAG codon in the gene encoding chloramphenicol acetyl transferase, and a T7-polymerase gene in direct control of a partnered GFP gene expression) in the presence of **4-1** and a negative selection (using a TAG codon in the toxic gene barnase) in the absence of **4-1**, as described in Chapter 1[20]. After two rounds of positive selection and two rounds of negative selection, individual clones were screened for chloramphenicol resistance in a third positive selection, dependent on the presence of **4-1**.

Initial trials with this selection process appeared to have afforded plasmid synthetases which specifically recognize **4-1**. As can be seen in figure 4-4, there are several colonies which have survived the final positive selection round, growing only in the presence of **4-1**, and at elevated levels of chloramphenicol. These results indicated the suppression of the CAT gene as well as the gene encoding the T7-polymerase, as indicated by the green fluorescence of the cells. The library plasmid DNA was isolated from these final survivors and sent for sequencing. Again, data was obtained which was contrary to the expected assumptions. Sequencing results for these library members proved unreasonable, where nearly all contained no sequence homology, or similarity to the previously evolved synthetase for **4-2**. This was surprising indeed, with the only explanation being that these synthetases survived the negative round selections, suppressing the amber codon with endogenous amino acids.

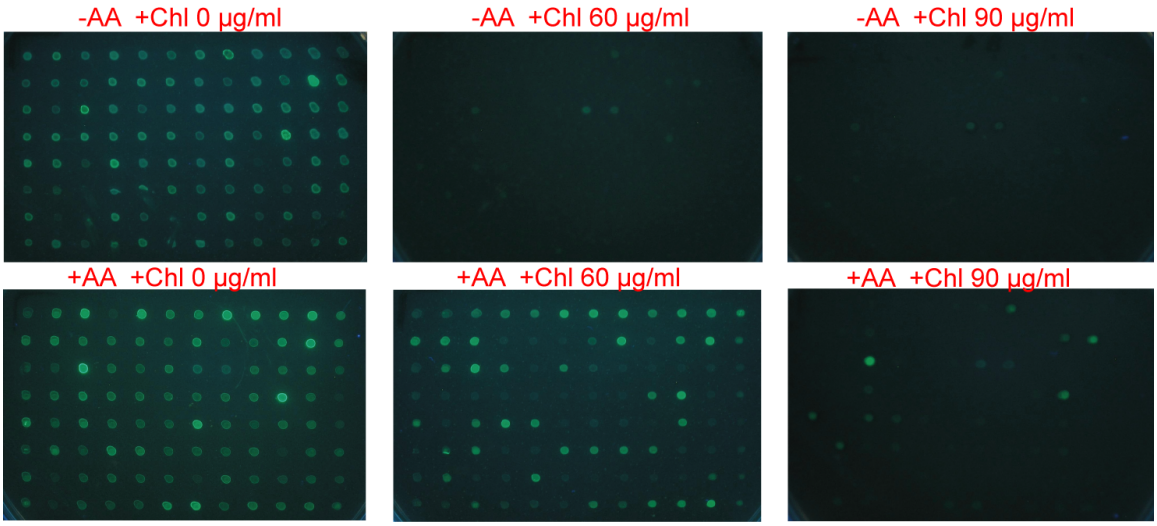


Figure 4-4. Final positive selection round for the suppression of a T7-promoter and CAT gene in the presence of the 4-1. The top row is grown in the absence of amino acid, while the bottom is in the presence of 1 mM, final concentration.

At this point, there was apprehension over the fidelity of the library. Even though sequencing results of the library DNA were correct, these unsuccessful attempts at selecting a synthetase for **4-1** was concerning. To test the library it was selected against **4-2**, as previously described, to act as a marker for its integrity. Following successive rounds of selections, nearly twenty colonies survived, where four of the plasmids were sent for sequencing. Two of the sequencing results provided the desired synthetase sequence, with mutations as reported above. Interestingly, two other sequences were returned which had only slight variations in the mutational residues which held precedence that perhaps the library of plasmids selected for against **4-2** would have the mutation of interest for recognizing **4-1**. With this, an alternative selection process was initiated where all the surviving colonies of the final negative round from a selection against **4-2** were then screened in the presence of **4-1**. This idea was rationalized by the fact that a similar two-step approach has proven useful in the past to modify substrate specificity in aaRSs[140].

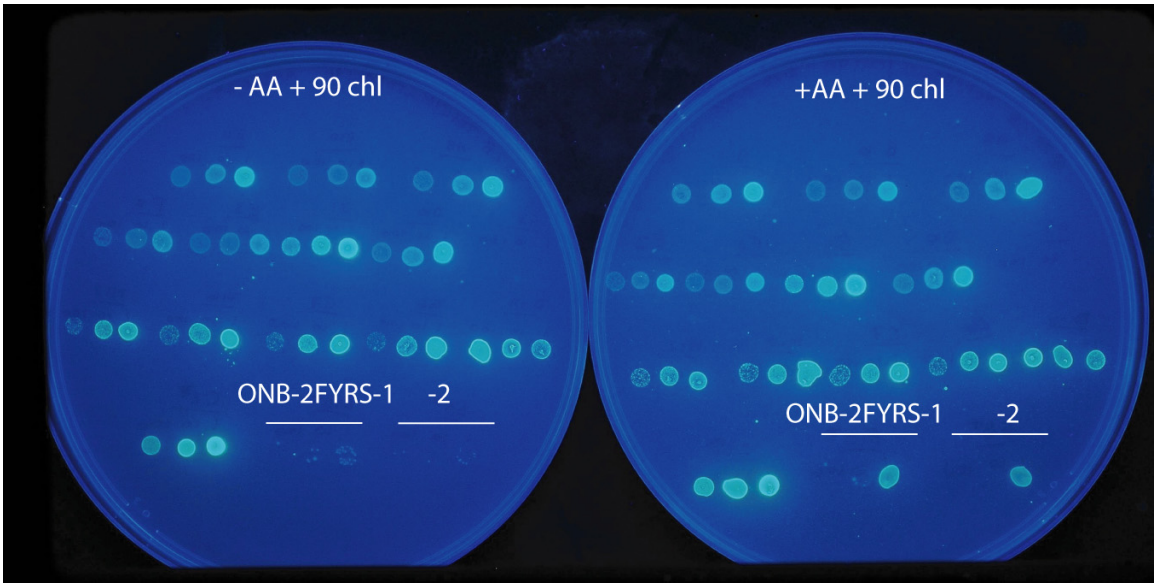


Figure 4-5. Final positive selection round for the suppression of a T7-promoter and CAT gene in the presence of the 4-1, from previous rounds selected against 4-2. Two surviving colonies were selected in the presence of chl 90 μg mL^{-1}.

Indeed, two clones that grew on agar containing 90 μg mL^{-1} chl, in the presence of **4-1** but only 20 μg mL^{-1} with no exogenous amino acid (Figure 4-5). Upon examining the sequence of these two aaRSs (ONB-2FYRS-1 and -2) it was determined that they were very close homologues of the previously evolved enzyme[20]. A noticeable difference, however, was the presence of a F108G mutation in both clones. This is contrasted with the original ONB-YRS, which contained a F108E mutation (Table 4-1). It is quite possible that this mutation to glycine results in a larger, more flexible active site capable of accepting the slightly larger fluorinated substrate. In addition, the mutations of H70N or H70M were identified, and I159M, which could not have been identified in the previous study as those residues were not included in the original library for **4-2**. The mutations at H70 help justify a potentially disruptive interaction between the fluorine and rigid, electron rich imidazole ring when **4-1** is introduced to the original ONB-YS. The mutation to N7) or M70 moves the electronic environment further away

from the spatial arrangement near the fluorine atom. What is most intriguing is that the natural *E. coli* TyrRS does not discriminate against fluorinated analogues despite being selected by evolution, while the engineered ONB-YRS does.

MjYRS	Tyr 32	Leu 65	His 70	Phe 108	Asp 158	Ile 159	Leu 162
ONBYRS[20]	Gly	Gly	His	Glu	Ser	Ile	Glu
ONB-2FYRS-1	Gly	Gly	Asn	Gly	Ser	Met	Asn
ONB-2FYRS-2	Gly	Gly	Met	Gly	Ser	Met	Glu

Table 4-1. Sequence comparison of evolved TyrRS synthetases. Differences from the original ONB-YRS sequence are indicated in red.

To verify the site-specific insertion of **4-1** into a protein an over-expression of the gene encoding 6×-histidine tagged sfGFP with a TAG codon in place of residue V150 was performed. This was carried out using the variant ONB-2FYRS-1, sub-cloned into a pSUP plasmid[138] for compatibility with our expression system. The pSUP plasmid simultaneously harbors the orthogonal amber tRNA gene, enabling the ability to co-express the pair of essential suppression elements in parallel. Expression of full-length protein was observed when **4-1** was included in the growth medium at a 1 mM concentration (Figure 4-6). The protein yield under these conditions is approximately 2 mg L^{-1}, which is comparable to protein expressions using **4-2** [20]. In the absence of the unnatural amino acid, no production of full-length protein is observed, indicating a high fidelity of the evolved synthetase since it does not accept any endogenous amino acid as a substrate. To further verify the correct product, an in-gel tryptic digest on the sfGFP protein was performed, where the isolated fragments were subjected to mass spectrometric sequencing confirming the site-specific incorporation of the fluorotyrosine

(Figure 4-6). A comparison of the mass spectrometry analysis of sfGFP expressed with ONB-YRS in the presence of **4-2** and expressed with ONB-2FYRS in the presence of **4-1** revealed identical peptide sequences for both proteins apart from those fragments containing the mutation which differ by a m/z of 18, equal to the difference of a fluorine and hydrogen atom. The arrows in figure 4-6 indicate two ion fragments which differ by a mass of 18 Da due to the introduction of fluorine on the tyrosine of the peptide.

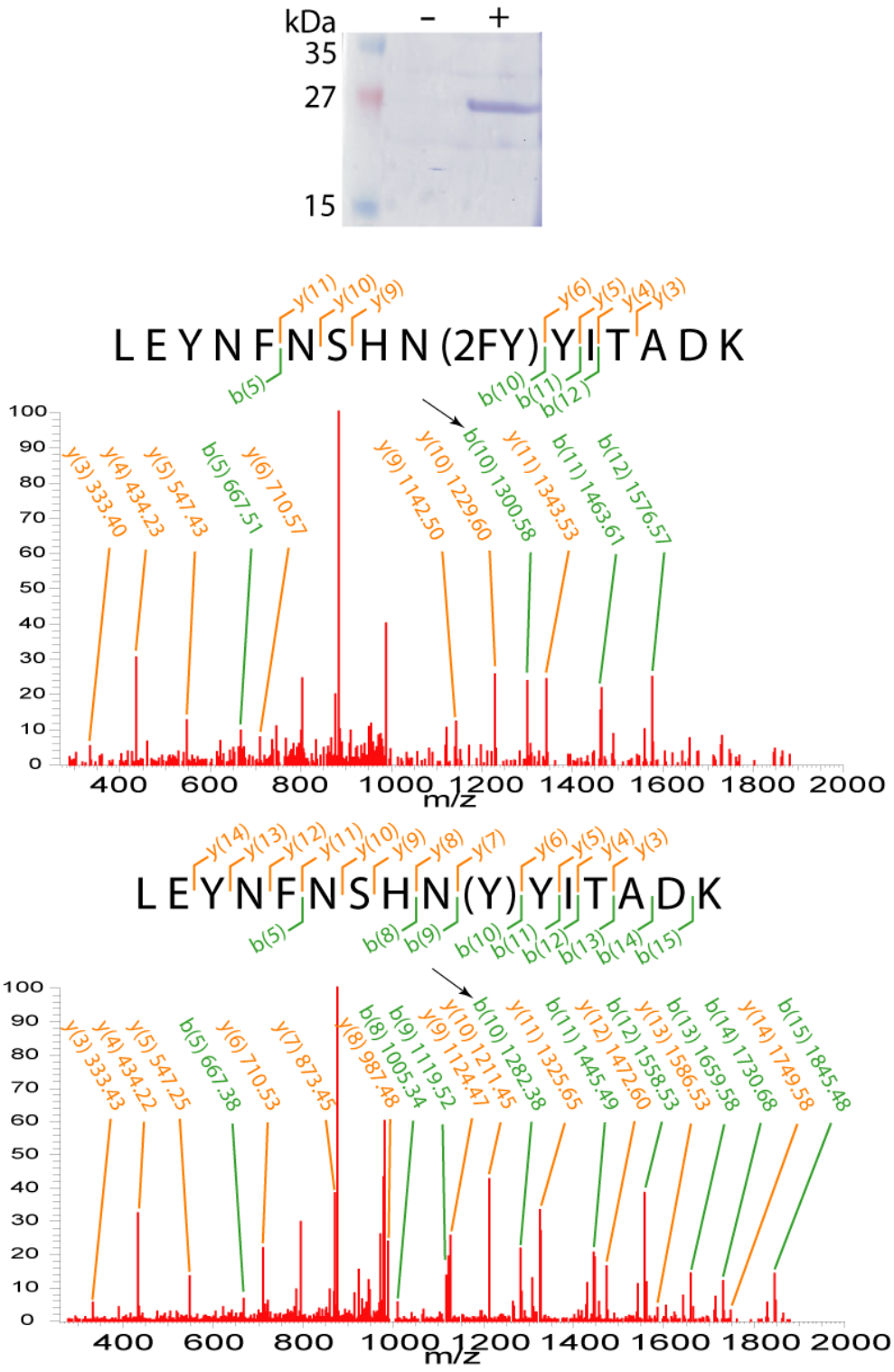


Figure 4-6. Coomassie stained SDS-PAGE gel of purified sfGFPV150TAG following an expression in both the absence and presence of 1 mM **4-1**. Mass spectral analysis of tryptic fragments (top) bearing **4-1** and (bottom) bearing **4-2**. The b and y ions change for fragments containing the fluorinated tyrosine by a mass of 18 Da. This is illustrated by the arrows at one ion fragment b(10), where the b fragmentation for protein containing **4-1** is 18 Da heavier.

4.3.3 Investigation of altered protein function through fluorescence studies

To test whether site-specific mutations of fluorotyrosine could be used to alter protein function the site-specific introduction of **4-1** was performed at an essential tyrosine in the chromophore of GPF. Previously, global mutagenesis studies have given rise to altered activity, such as pH rate profiles, but these experiments used proteins containing multiple mutations. One such study used global incorporation of 3-fluorotyrosine into EGFP[60, 133], which contains eleven total tyrosine residues. One of these, Y66, is central to the chromophore. In order to conduct a more precise investigation, sfGFP, an improved stability version of GFP[141], was expressed and isolated in the presence of **4-1** (Figure 4-7 c-d). This protein, unlike any previous studies, contains only a single residue mutation to a fluorotyrosine. In addition, a control sfGFP, which instead contained **4-2**, was also generated for comparison purposes as a surrogate for the wild-type tyrosine residue. Both of these proteins were purified and then "decaged" by irradiation at 365 nm for 30 minutes, after which were assayed for fluorescence properties. In the case of the fluorinated chromophore mutation the λ_{em} is blue shifted (503 nm) when compared to the tyrosine control (513 nm) (Figure 4-7 a-b). This demonstrates the ability of a single fluorotyrosine residue to change the electronic properties of the chromophore and ultimately affect the protein function.

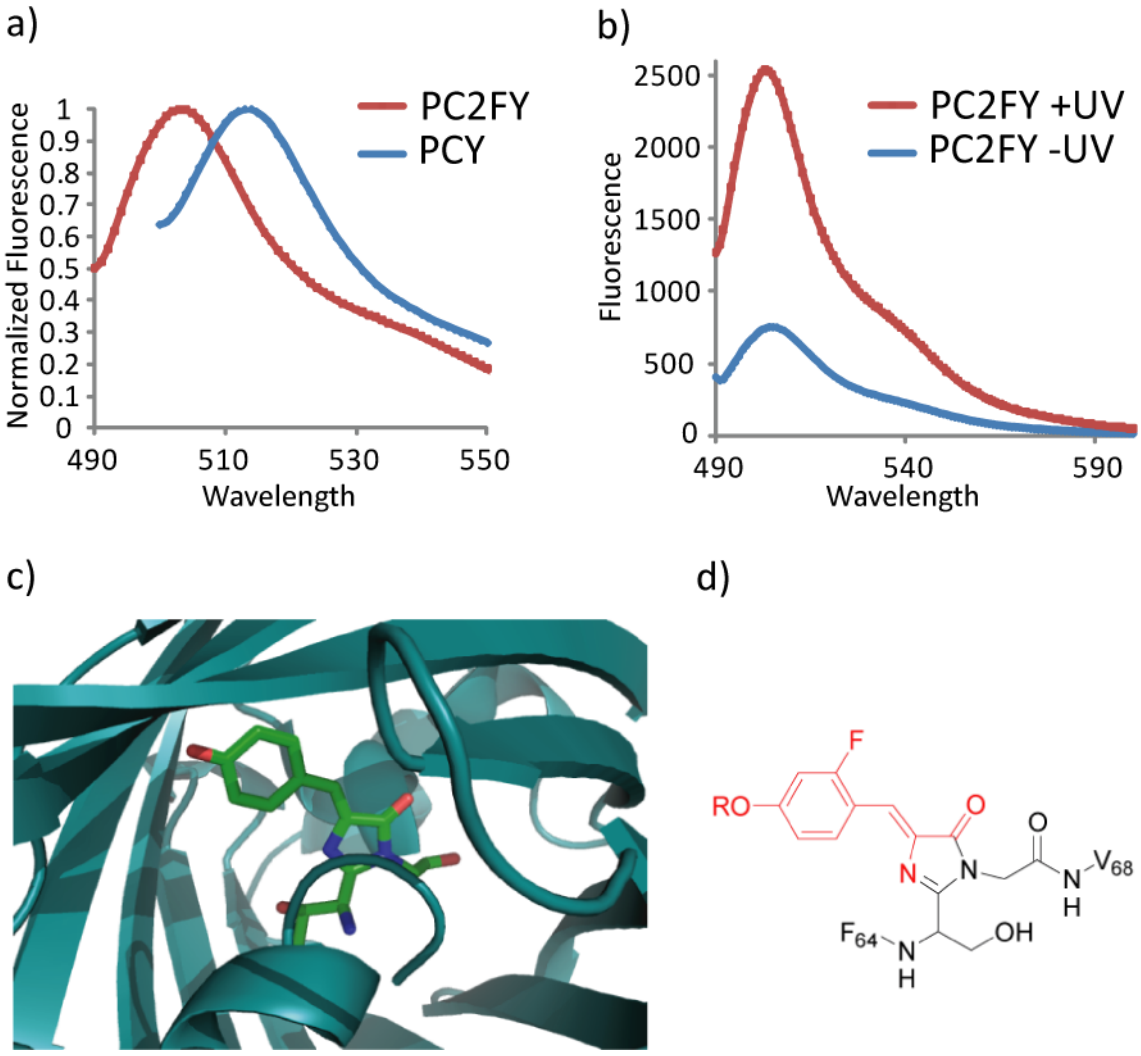


Figure 4-7. a) Normalized fluorescence of decaged protein containing **4-1** or **4-2**. b) The actual fluorescence of protein containing **4-1** before and after irradiation at 365 nm for 30 minutes, showing activation levels. c) The structure of sfGFP mature chromophore, shown in green. PDB entry 2B3P. d) The structure of the mature chromophore with a 2-flurotyrosine mutation at Y66.

One interesting observation is that the sfGFP bearing **4-1** is fluorescent at 503 nm, and that this fluorescence increases by 343% after removal of the ONB caging group through a brief irradiation at 365 nm (Figure 4-7). This increase in fluorescence is most likely due to the generation of the phenol hydroxyl group which can delocalize charge throughout the chromophore ring system, and is not possible when the amino acid is caged. Surprisingly though, one could imagine that mutation of the chromophore

tyrosine of sfGFP, which is in the core of the protein, would be destabilizing to the protein. Indeed, sfGFP is known to lose all fluorescence upon denaturation[142]. Thus, the observation of fluorescence in the case of GFP containing **4-1** indicates that the large ONB caging group does not prevent correct protein folding and can be accommodated in the active site. The photo-activation by the removal of a protecting group from the chromophore of the protein could be used as a general method to "switch" fluorescent proteins, and the fact that **4-1** and **4-2** are genetically encoded amino acids enables the possibility of using directed evolution to improve the photo-switching properties.

4.3.4 Investigation of altered protein function through enzymatic studies

As an attempt to further explore the use of fluorinated tyrosine, it will be used as a probe to investigate its electronic nature in the active site of *E.* coli β-galactosidase. The aim is to control tyrosine acidity within the context of an enzyme active site by incorporating **4-1** in place of Y503. This tyrosine residue is critical for catalysis and previous studies have, shown that Y503 is important in orienting active site residues, including E537, and serving as a general acid[143, 144] (Figure 4-8). Global introduction of fluorotyrosines would surely alter statistical data, as there are 32 tyrosines in this rather large protein. However, using the general approach of disguise, with photo-protecting groups, **4-1** is to selectively replace one tyrosine residue.

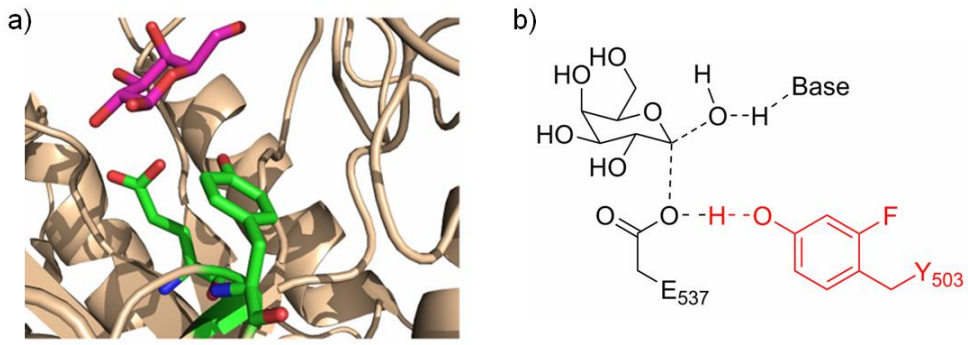


Figure 4-8. a) Structure of β-gal with the position of tyrosine mutation and catalytic glutamate shown in green, substrate in magenta. b) Schematic of the transition state for the regeneration of the free enzyme from the glycosylated enzyme intermediate. PDB entry 1JZ7.

4.4 Conclusion

Previously, the investigation of tyrosines in protein, using fluorotyrosine substitutions, was limited by the global incorporation of this amino acid. The general methodology described above enables individual tyrosines in proteins to be selectively mutated to 2-fluorotyrosine. The concept of using a photo-removable protecting group to temporarily "hide" amino acids from the endogenous cellular biosynthetic machinery may be generally applicable to other fluorotyrosine analogues, or any other close analogue of one of the twenty genetically encoded amino acids. Finally, it was intriguing to discover that the fluorescence of sfGFP containing the caged 2-fluorotyrosine in the chromophore active site was greatly reduced compared to wild-type sfGFP, but could be activated upon a brief irradiation at 365 nm. This observation holds promise as a novel rational method for the creation of photo-activatable green fluorescent proteins for cellular biology studies with high spatial and temporal resolution.

Chapter 5: Site-specific incorporation of other tyrosine derivatives into *E. coli* via photochemical disguise

5.1 Introduction

The previous chapter outlined the general application of photo-disguising 2-fluorotyrosine for efficient incorporation into protein at the amber stop codon. It was originally assumed that the small perturbation on the molecule would not disrupt substrate recognition by the previously evolved synthetase for ONBY. Due to the inability to express protein with this synthetase, a new synthetase was evolved to specifically recognize *o*-nitrobenzyl-2-fluorotyrosine (ONB-2FY). There are several other fluorotyrosine derivates of interest, all of which warrant the ability to be site-specifically introduced at the genetic level. Each analogue possesses differing pKa values which can be used as probes for protein function, just as outlined above.

For the rationale addressed in the previous chapter, 3-fluorotyrosine (**5-1**) and 2,6-difluorotyrosine (**5-2**) cannot be directly introduced to proteins as they are recognized by the endogenous translational machinery responsible for charging a tyrosyl-RS. These two NAAs are also not suitable candidates for UAA mutagenesis for this same reason. As the general method for masking NAAs proved successful with ONB-2FY, the same application was directly applied to both **5-1** and **5-2**. Each of these compounds were synthesized using a combination of chemical and enzymatic synthesis, and then caged in the same fashion as that for the 2-fluorotyrosine. With the evolution of two new synthetases, which possess altered binding-pockets, it was assumed that these new spatial arrangements may be promiscuous, and in turn also accept the caged versions of both **5-1** and **5-2** (Figure 5-1). As the binding interference from histidine in the original synthetase

has been modified it was worth performing expressions with these new synthetases to examine their direct recognition of **5-3** and **5-4**. Suppression of the amber codon with these UAAs would expand the molecular toolbox with available NAAs for protein exploration.

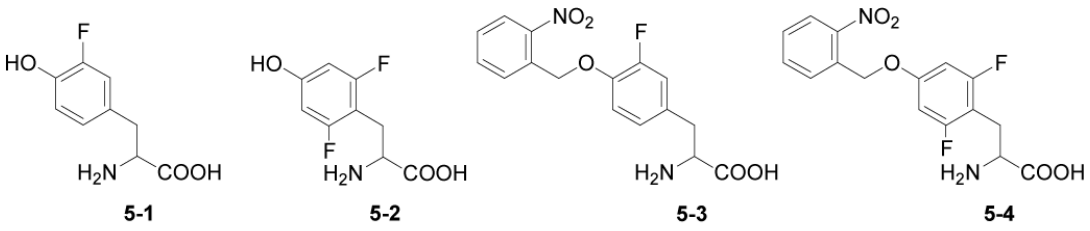


Figure 5-1. The fluorotyrosine derivatives 3-fluorotyrosine (**5-1**) and 2,6-difluorotyrosine (**5-2**).

5.2 Experimental

5.2.1 Materials

Reagents for the synthesis of 3-fluorotyrosine, 2,6-difluorotyrosine, *o*-nitrobenzyl-3-fluorotyrosine (**5-3**), and *o*-nitrobenzyl-2,6-difluorotyrosine (**5-4**) were purchased from Fisher. All other reagents were purchased commercially and used without further purification**.** pSUP-ONB-2YRS-1 and -2 were employed from previous exploits.

5.2.2 General Methods

The antibiotics employed were chloramphenicol (35 μg mL^{-1}) and ampicillin (100 μg mL^{-1}). Transformations for were performed into GeneHogs (Invitrogen) electro-competent cells using ~100 ng of DNA transformed into 100 μL of cells. **5-1** and **5-2** were dissolved in 50% DMSO in water (v/v) with 0.5 M NaOH added dropwise until fully dissolved at ~80 °C. The solution of amino acid was then added to hot agar (or liquid) with stirring and allowed to cool to 40° C before the addition of antibiotics (and

cells for expressions). Denaturing discontinuous sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels (12 % acrylamide resolving and 5% acrylamide stacking) and 1X Laemmli buffers were used following standard procedures, performed at 200 V for 65 min. Purification of protein was performed using Probond Purification resin (Invitrogen) according to the manufacturer's protocol for native isolation. SDS-PAGE gels were stained with Coomassie Brilliant Blue. For samples subjected to UV irradiation, and subsequent loss of the *o*-nitrobenzyl group, a hand-held 100W Blak Ray lamp was used at a distance of 5 cm, at 4 °C, for 30 min. Emission scans were set for an excitation of 490 nm**.** Scanning was performed from 500-650 nm (Ex slit 10.0 nm, Em slit 5.0 nm, PMT voltage 950 v, Hitachi F-4500 FL Spectrophotometer).

5.2.3 Synthesis of o-nitrobenzyl-3-fluorotyrosine and o-nitrobenzyl-2,6-difluorotyrosine

5-1 was synthesized from the starting material 2-fluorophenol and **5-2** was synthesized from the starting material 3,5-fluorophenol, where both were characterized as previously described in the previous chapter (4.2.3).

3-Fluorotyrosine. ^{1}H NMR (400 MHz, D_2O); δ 6.79 (dd, J = 12.95 Hz, J = 2.31 Hz, 1 H), 6.67-6.70 (m, 1 H), 6.59 (dd, J = 10.17 Hz, J = 8.32 Hz, 1 H), 3.33 (dd, J = 7.40 Hz, J = 5.55 Hz, 1 H), 2.74-2.78 (m, 1 H), 2.85-2.63 (m, 1 H). HRMS (ESI) m/z calculated for $C_9H_{10}NO_3F$, 199.0645, found 199.0237.

2,6-Difluorotyrosine. ^{1}H NMR (400 MHz, D_2O); δ 6.04-6.09 (m, 2 H), 3.32 (dd, J = 7.63 Hz, J = 6.01 Hz, 1 H), 2.77-2.82 (m, 1 H), 2.62-2.67 (m, 1 H). HRMS (ESI) m/z calculated for $C_9H_9NO_3F_2$, 217.0550, found 217.0297.

The *o*-nitrobenzyl protecting group was installed similar to reported methods[137], and described in detail in section 4.2.3.

***o*-Nitrobenzyl-3-fluorotyrosine.** ^{1}H NMR (400 MHz, DMSO); δ 8.14 (d, *J* = 7.86 Hz, 1 H), 7.78-7.83 (m, 2 H), 7.62-7.66 (m, 1 H), 7.12-7.18 (m, 2 H), 6.99 (d, *J* = 8.79 Hz, 1 H), 5.50 (s, 2 H), 3.36 (bs, 3 H), 3.06 (dd, *J* = 14.33 Hz, *J* = 4.62 Hz, 1 H), 2.08 (dd, *J* = 14.33 Hz, *J* = 7.86 Hz, 1 H). HRMS (ESI) m/z calculated for $C_{16}H_{15}N_2O_5F$, 334.0965, found 334.0258.

***o*-Nitrobenzyl-2,6-difluorotyrosine.** ^{1}H NMR (400 MHz, DMSO); δ 8.14 (d, *J* = 8.03 Hz, 1 H), 7.74-7.81 (m, 2 H), 7.64 (t, *J* = 7.53 1 H), 6.84 (d, *J* = 9.54, 2 H), 5.48 (s, 2 H), 3.35 (bs, 3 H), 3.06-3.19 (m, 2 H). HRMS (ESI) m/z calculated for $C_{16}H_{14}N_2O_5F_5$, 352.0871, found 352.0758.

5.2.4 Expression of sfGFP

pSup-ONB-2FYRS-1 (or -2) was transformed into *E. coli* cells along with pTrc-HisA-sfGFP66TAG (or pTrc-HisA-sfGFP150TAG for mass spec studies) (See appendix for plasmid construction). These expressions were performed in 2X-YT media (50 mL) supplemented with Amp, Chl, and a final concentration of 1 mM of appropriate amino acid. The cells were grown to an OD_{600} = 0.8 and then induced with a final concentration of 1 mM IPTG. Expression was continued at 37 ^oC, overnight in the dark. Purification of sfGFP was performed using the HisLink purification system (Promega) under native conditions according to the manufacturer's protocol. Isolated protein was then dialyzed against 20 mM TRIS buffer (pH 7.2), analyzed via SDS-PAGE electrophoresis, and used for fluorescence experiments.

5.3 Results and Discussion

To test the substrate recognition of both **5-3** and **5-4**, protein expressions were carried out in the presence of each amino acid. These compounds were tested against

both of the newly evolved synthetases, evaluating the efficiency of each. The expressions were performed with sfGFP, as previously described.

Expression of full-length protein was observed when **5-3** was included in the growth medium at a 1 mM concentration when paired with either the ONB-2FRS-1, or -2, synthetase (Figure 5-2). In the absence of the unnatural amino acid, no production of protein was observed, indicating a high fidelity of binding interactions with **5-3**. It appears however, that **5-4** is only recognized as a substrate for the ONB-2FRS-2 synthetase (Figure 5-2). When **5-4** was included in the growth medium for expression with RS-1, there is no observed isolation of protein. Full-length protein is indeed observed for the expression with RS-2. Again, in the absence of the unnatural amino acid with RS-2, no production of protein was determined. Although RS-2 appears to recognize both **5-3** and **5-4**, the efficiency of protein expression seems impaired.

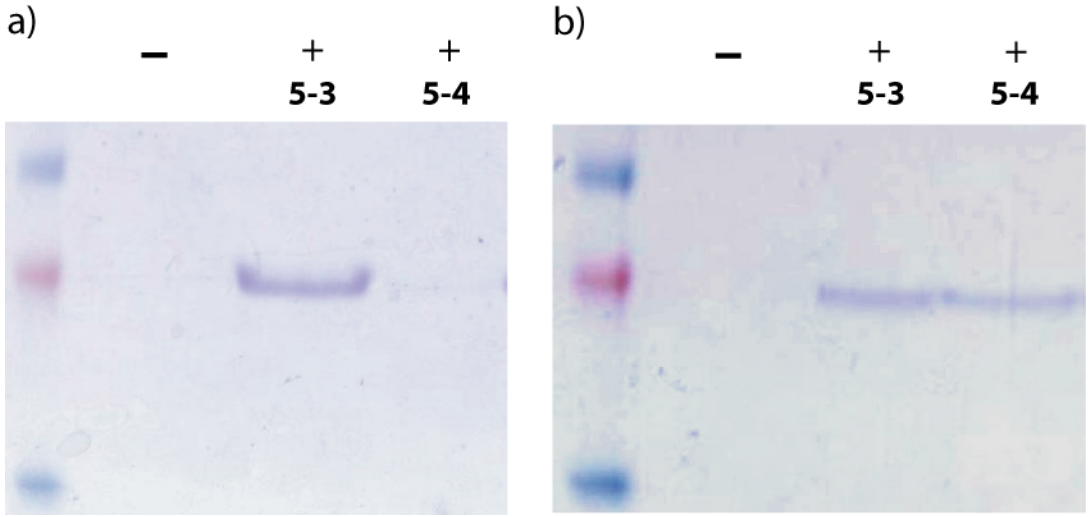


Figure 5-2. Coomassie stained SDS-PAGE gels of isolated protein expression for sfGFP. Both gels represent protein expressed in the presence of the indicated amino acid, where the negative sign represents no addition of amino acid. a) performed with ONB-2FYRS-1 and b) with ONB-2FYRS-2.

To further verify the correct product, an in-gel tryptic digest on the sfGFP protein was performed as before, where the isolated fragments were subjected to mass

spectrometric sequencing confirming the site-specific incorporation of the two fluorotyrosines (Figure 5-3). A comparison of the mass spectrometry analysis of sfGFP expressed in the presence of **5-3** and **5-4** revealed identical peptide sequences for both proteins apart from those fragments containing the mutation which differ by a m/z of 18, equal to the difference of a fluorine and hydrogen atom, as the di-fluoro and 3-fluoro derivatives have a single fluorine variation. When these results were compared to the mass spectral data for ONB-2FY in the previous chapter, the masses were identical to those with **5-3**, where the only variation between these two molecules is the position of the fluorine atom. As expected, a mass differences of m/z = 36 was observed when the di-fluoro protein was compared to that of the non-fluorinated protein (see Figure 4-6).

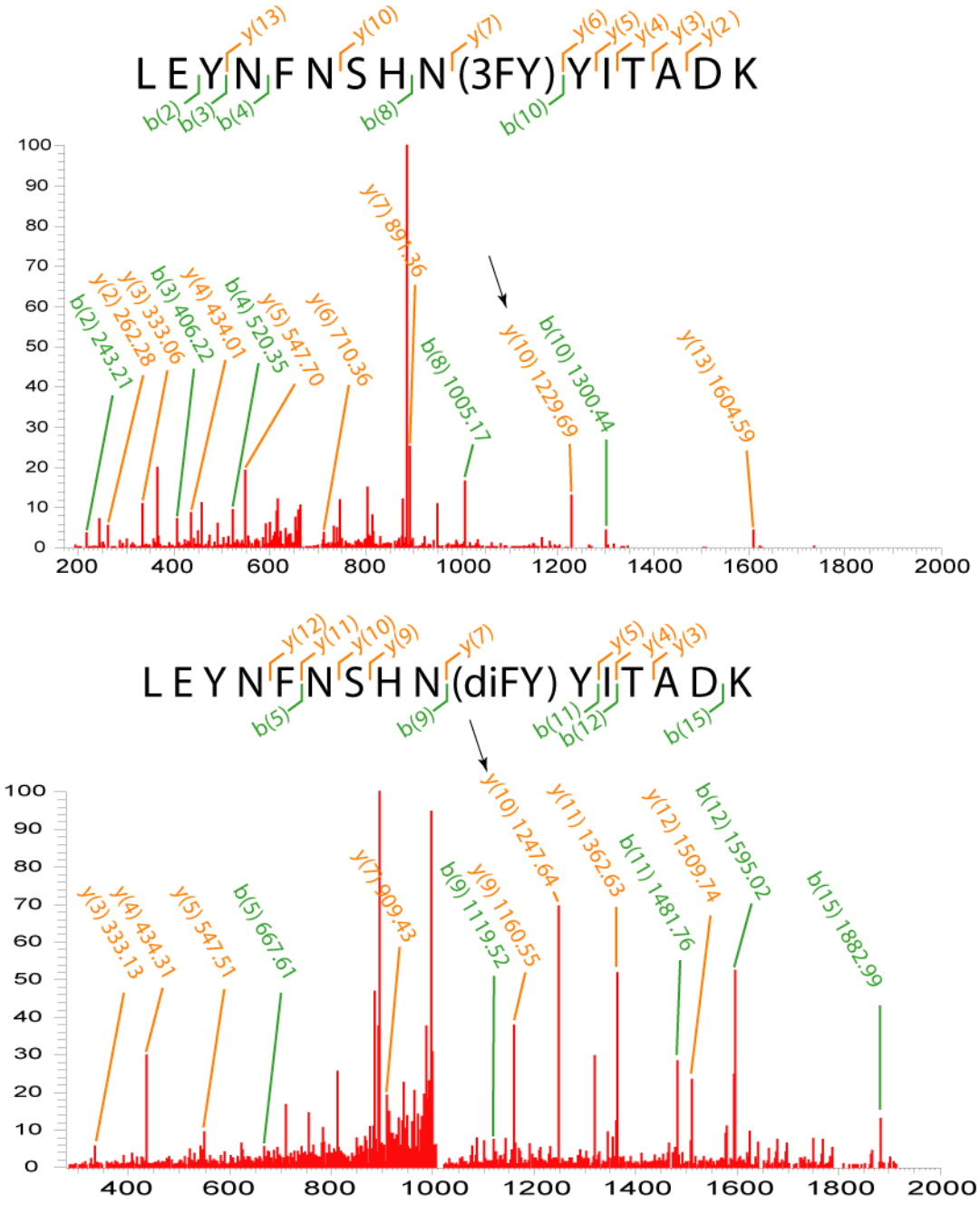


Figure 5-3. Mass spectral analysis of tryptic fragments of sfGFP (top) bearing **5-1** and (bottom) bearing **5-2**. The b and y ions change for fragments containing the **5-1** by a mass of 18 Da and 36 Da for fragments containing **5-2**, as compared to native tyrosine. The arrows at one ion fragment y(10), are illustrated here, where the y fragmentation for protein containing **5-1** is 18 Da heavier than that for protein with **5-2**, due to difluorination of the tyrosine.

To further evaluate the potential of **5-3** and **5-4**, these amino acids were introduced to the Y66 position in the chromophore of sfGFP, as before. Both of the

expressions were purified and then "decaged" by irradiation at 365 nm for 30 minutes, after which they were assayed for fluorescence properties. For the 2-fluoro chromophore mutation the λ_{em} was shifted from the wild-type 513 nm to 503 nm. For the 3-fluoro chromophore mutation the λ_{em} was red shifted from 513 nm to 515 nm (Figure 5-4). This is a rather conservative change indeed, but is consistent with previous data for the red shifting of 3-fluorotyrosine by 2 units in EGFP[62]. The ability to induce fluorescence intensity upon UV irradiation was also observed for the 3-fluoro protein, as previously determined for the 2-fluoro derivative. sfGFP bearing **5-3** is fluorescent at 515 nm and the fluorescence increases by nearly 500% after removal of the ONB caging group (Figure 5-4). This data is consistent with that of the previous decaging events with ONB-2FY further justifying the possibility that the photo-activation of the protein could be used as a general method to control fluorescent proteins in cells providing a level of spatiotemporal control.

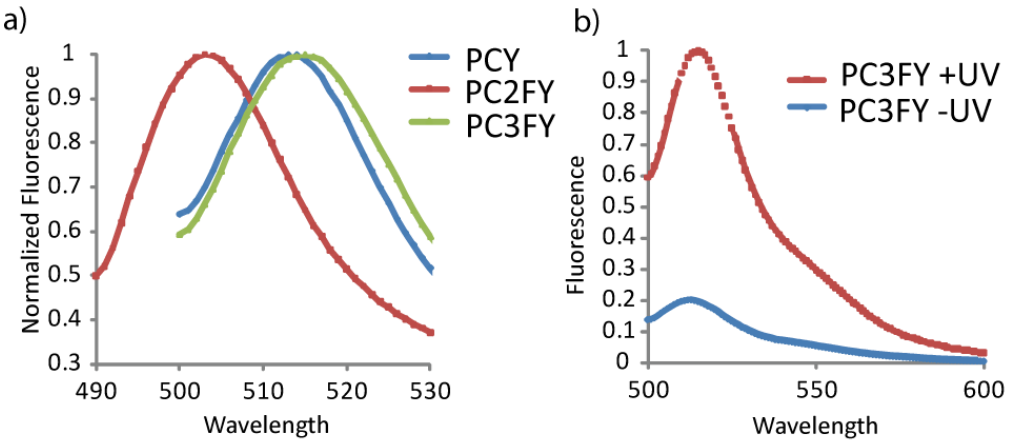


Figure 5-4. a) Normalized fluorescence of decaged protein containing **4-1**, **4-2**, and **5-1**. Also displayed is the normalized fluorescence of protein containing **5-1**, before and after decaging (the caged fluorescence was adjusted against the peak fluorescence observed for the decaged sample).

Although protein was isolated for the expression in the presence of **5-2**, fluorescence was observed, at levels below those seen for the caged samples of both 2-

and 3-fluoro proteins, prior to decaging events. No increase in fluorescence occurred following irradiation with 365 nm light. It is believed that the di-fluoro addition to the critical tyrosine position may cause destabilizing effects in the chromophore of the protein. If the new polarization interferes with the covalent formation of the chromophore, no fluorescence would be expected.

As this is a continuing project, the activity of β-gal is currently being probed by the site-specific insertion of **5-3** and **5-4** at the active site Y66. These mutants will be applied to enzymatic assays as described in the previous chapter to determine the pH rate profiles for each fluorinated derivative.

5.4 Conclusion

The concept of using a photo-removable protecting group to temporarily mask amino acids from the endogenous cellular biosynthetic machinery may be generally applicable to any near-natural analogues of amino acids. It has been demonstrated here, with three fluorinated derivatives of tyrosine that the covalent modification of the phenolic hydroxyl group can act as a bypass mechanism for which UAA mutagenesis can be applied. Once these compounds have been utilized in protein expression they can be photo-activated to become powerful probes in protein explorations. While a fluorinated caged tyrosine was not recognized by the synthetase for its cognate non-fluoro partner, several fluorinated caged compounds seem to be recognized by newly evolved synthetases. Where there are two distinctly different synthetases which recognize both ONB-2FY and **5-3**, only one of them recognizes **5-4**. Since there is some promiscuity in these binding sites, it is worth exploring other fluorinated tyrosines with these newly evolved components. It is easy to imagine that the synthetase which recognizes **5-4** may

also bind derivatives such as ONB-3,5-diFY, and ONB-2,3,5,6-tetraFY, the latter of which has a significant change in pKa at the phenol. Not only may these compounds act as new tools for altering active site electronics, but the caging group generates the ability to control function with an external stimulus.

Chapter 6: Significance and Conclusion

With the advent of an expanding genetic code it is now possible to directly select for an amino acid of interest for protein translation at the genetic level. The mutational selection system for an orthogonal amber-tRNA/aminoacyl-tRNA synthetase pair has revolutionized protein discovery through unnatural amino acid mutagenesis. The growing number of these new additions to the genetic code each adds their own insight into biochemical function. This work has taken advantage of these processes to exact a more precise chemical understanding of protein function with novel additions to the unnatural amino acid catalogue, as well as the expansion of techniques with previously developed compounds.

The overlapping theme of each compound utilized in this work is the ability to augment its native state with light. This is a powerful weapon in the biological world as biochemical activity occurs as a cascade of cellular events which happen at specific intervals depending on the stimulus. This spatial and temporal association is extremely difficult to monitor with typical biochemical assays. Once an external agent is introduced to a system the control of a defined instance for its use is lost. With the onset of photochemistry being applied to biological molecules it has become possible to manipulate amino acid side chains which contain photo-affinity labels, or photo-labile protecting groups, with the power of light. In one instance, it becomes an invaluable tool for site-specific photo-crosslinking, and in the latter a potential "light switch" for biological activity. In a more cunning manner, the photo-protecting group may simply

act as a mask by which to introduce near-natural amino acids to protein, when disguised as an unnatural amino acid.

As mass spectrometric analysis is increasingly becoming paired with chemically cross-linked protein analysis, a major concern presents itself when trying to assign the cross-linked fragment of an unknown protein to a peak in the spectrum. Complex fragmentation patterns can often result from enzymatically treated complexes. This work developed a new tool for the simplification of these analytic techniques during the use of unnatural amino acid mutagenesis, with an isotopically labeled analogue, [D_{11}]-pBpa. While isotopic labeling is widely used for mass spectral studies, this is the first example of this method with unnatural amino acids. The labeled derivative of pBpa creates a fingerprint during mass spectral analysis, clearly identifying a crosslinked peptide fragment of interest by the appearance of M, M +11 signature peaks in an array of masses.

The ability to control biological function with light is an intriguing application in biochemical studies. There are many active biomolecules that have essential functional groups, all of which have the potential to be restricted in nature due to the addition of a photo-protecting group. In the work accomplished here, a caged cysteine residue was introduced to a genetically encoded fluorescent affinity tag which is dependent upon four cysteine residues. The compound 4,5-bis(1,3,2-dithioarsolan-2-yl)fluorescein, FlAsH, specifically binds to, and is activated by, the small amino acid tag sequence, CCPGCC. This occurs through a 1,2-dithiol exchange with the four cysteines in the tag. It was shown that the ability of cysteine/dye reaction events was inhibited, due to the presence of an *o*-nitrobenzyl group on a cysteine residue. This hindered the required binding for

the FlAsH dye, and prohibited fluorescence. When full-length protein was isolated, which contained ONBC in the tetracysteine tag, the direct control over small molecule binding to the protein was observed. FlAsH labeling events could be manipulated with a photo-caged unnatural amino acid, as the fluorescence was restored following cysteine decaging with light irradiation. This new application for the genetic encoding of ONBC could be an attractive alternative to other photo-activatable fluorescent proteins such as PA-GFP.

Near-natural amino acids are non-natural; however, they are only slight variants of their native structures. The concern with these amino acids is that they are recognized by the endogenous aminoacyl-tRNA synthetase machinery that naturally incorporates the non-modified amino acid. To circumvent this dilemma, these near-natural amino acids were temporarily disguised as unnatural amino acids, so as to be directly applied to the general method of unnatural amino acid mutagenesis. As part of this work a new aminoacyl-tRNA synthetase was evolved to specifically recognize the fluorinated derivative of a caged 2-fluorotyrosine. It was demonstrated that this fluorinated tyrosine could indeed be site-specifically introduced by first translating the protein with *o*-nitrobenzyl-2-fluorotyrosine. Subsequent illumination with light proved successful in the decaging of the molecule, affording protein with the site-specific placement of 2-fluorotyrosine.

The newly evolved synthetase has also proven promiscuous to other fluorinated derivatives with the expression of protein in the presence of both a caged 3-fluorotyrosine and 2,6-difluorotyrosine. This simple "deception" of introducing a photo-removable group to near-natural amino acids affords an efficient method for the site-specific

introduction of these amino acids to protein. This opens the door for an array of new amino acids with powerful probing potential.

The direct expansion of any toolbox provides a wider range of options when attempting to elucidate problems of interest. Chemically, there is an unlimited ceiling to the number of unnatural amino acids which can be synthesized for chemical studies. Each newly acquired chemical function can potentially be introduced to a protein through an expanded genetic code.

This work has spanned several disciplines including chemistry, biochemistry, and molecular biology. The applications here should find a broad range of potential guests for further studies. It is believed that this work has successfully focused on the synthesis and application of unnatural amino acids as a defined chemical approach in which to investigate protein function. One of the major aspects of this work is the first attempt to use unnatural amino acid mutagenesis as a general approach to mask near-natural amino acids for their site-specific introduction to proteins. This system provides the backdrop for an overwhelmingly large range of new small molecules (through synthetic generation of amino acids) for future chemical biology research.

Appendices

Other work addressed during this study

Selenocysteine:

As discussed earlier, the site-specific encoding of selenocysteine is technically difficult as its natural codon is UGA, a universal stop codon, followed by a sequence Selenocysteine Insertion Sequence, SECIS, mRNA that specifically recruits cellular machinery required for selenocysteine incorporation. Without the specific SECIS sequence there is no insertion of the amino acid. These represent obvious molecular biological constraints when designing mutational expression sequences. The site-specific incorporation of this amino acid during mutational studies is particularly appealing due to its reduced side chain pKa of 5.3. This would not only help studies involving natural selenocysteine proteins, but also has the potential to be an attractive site for bioconjugations.

This amino acid is a near-natural analogue of cysteine; however, it is naturally occurring. In a similar fashion to the work applied in the fluorotyrosine studies, a caged version of the amino acid was synthesized so as to be altered to an unnatural amino acid. Considering this amino acid is the same general size as *o*-nitrobenzyl cysteine (ONBC),

protein expression was performed in the presence of caged selenocysteine, with the previously evolved synthetase which recognizes ONBC. It was determined that caged selenocysteine was indeed a substrate for the ONBC synthetase when introduced for the over expression of human superoxide dismutase containing a TAG codon (Figure A-1). The protein is only fully expressed when in the presence of caged selenocysteine. This is yet another example of how disguising amino acids with photo-removable protecting groups may act as a powerful method for the direct insertion of near-natural amino acids. Further exploration of this particular application needs to be pursued, as this indeed sets the basis for particularly interesting protein exploration.

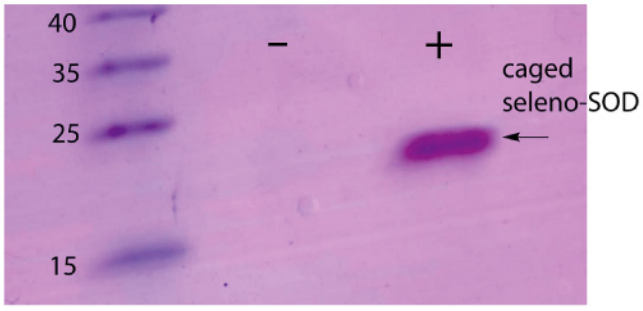


Figure A-1. Coomassie stained SDS-PAGE gel of isolated hSOD protein which contains an amber codon in its gene. Expressions performed with ONBC synthetase in the presence (+) and absence (-) of selenocysteine.

Photo-induced protein function and site-specific cysteine labeling:

Biological systems are dependent upon "switch" controlled protein function to transduce appropriate cellular responses. Theses switches can range from small peptide ligands, to carbohydrates, to external stimuli such as light. Artificial control of biological function allows for chosen activity by the researcher and not by nature. Currently a system has been developed by which light can be used as a switch to activate protein function, posttranslationally. This was done by taking advantage of an intein function,

which is a protein splicing element. Inteins catalyze their own excision from a polypeptide chain, through cysteine protease type activity, subsequently joining the flanking exteins[145-147]. A proteins function is generally disrupted prior to excision of the intein; however, excision is spontaneous and rapid upon complete translation of the protein. Controlling the excision events of the intein presumably will allow for control over the function of the protein.

It was decided to apply the photo-caged system to an expression linked to an intein construct which contains the amber codon at the essential cysteine. If a photo-caged cysteine may be introduced, the caging group would act to inhibit intein excision, as it would block the nucleophilic ability of the thiol in the reaction mechanism. Once the caging group was removed, via light, native cysteine would be afforded, activating the intein excision, "turning on" protein function (Figure A-2). The expression vectors for this system have been created, and will be addressed in future studies. Theoretically, it may be possible to extend a level of spatiotemporal control to protein function which could play a powerful role in the elucidation of biological pathways.

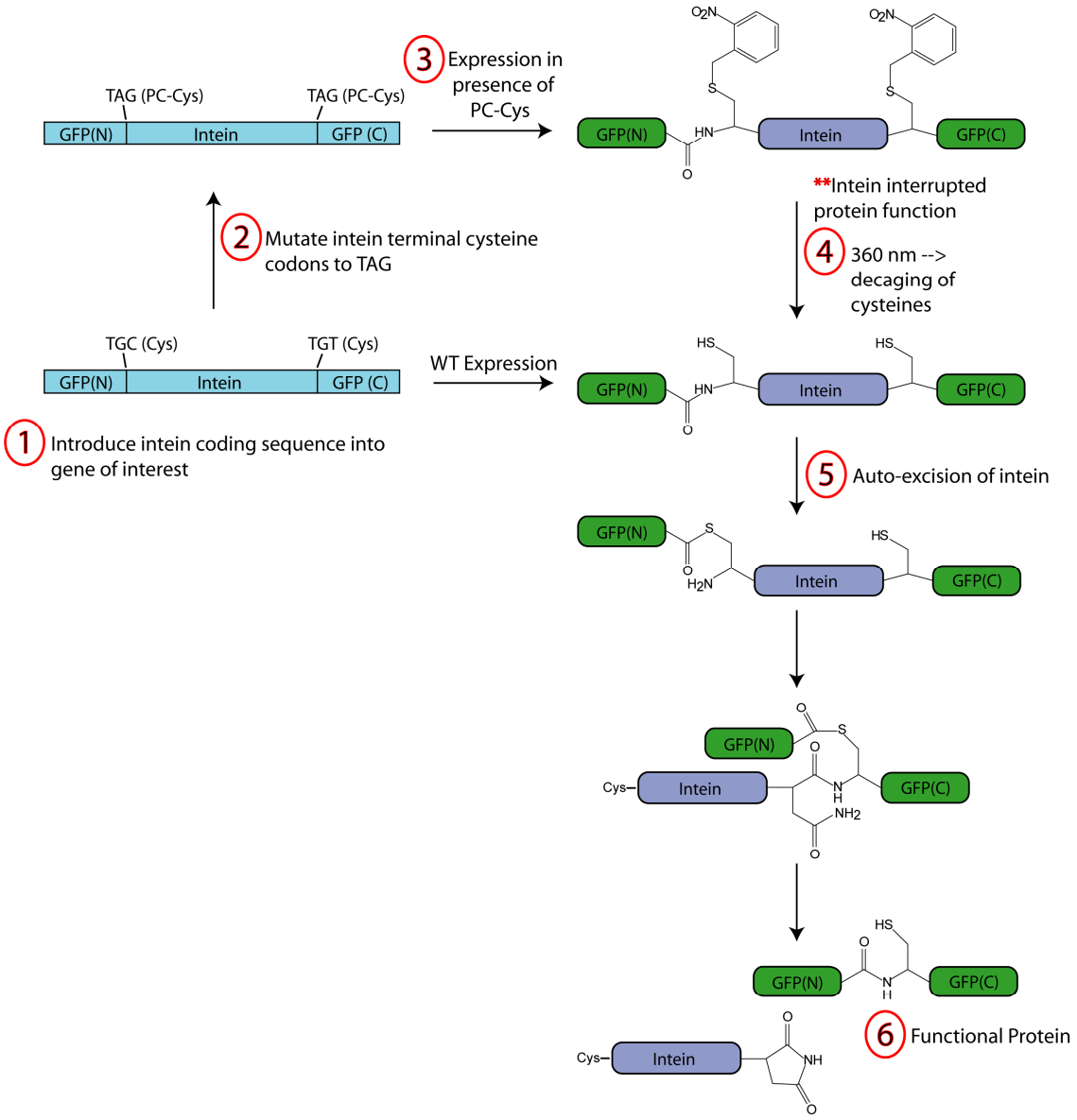


Figure A-2. General scheme of an intein based expression system where photo-caged cysteine may be used as a "switch" for posttranslational protein functionality.

In a similar fashion, photo-caged cysteine may prove to be an efficient tool for site-specific labeling with small molecules, or specific sites for bioconjugation reactions. If a protein contains multiple cysteine residues, the site of interested may be mutated to photo-caged cysteine, where the remaining cysteines (not already associated in a disulfide bond) can be chemically protected with reagents such as 5,5'-dithio-*bis*-2-nitrobenzoic

acid (Ellman's reagent). The protein can then be photochemically modified so that a single cysteine residue remains available for additional bioconjugation/modification reactions (Figure A-3). In this general scheme it should be possible to site-specifically modify cysteine residues with tags such as polyethylene glycol (PEG).

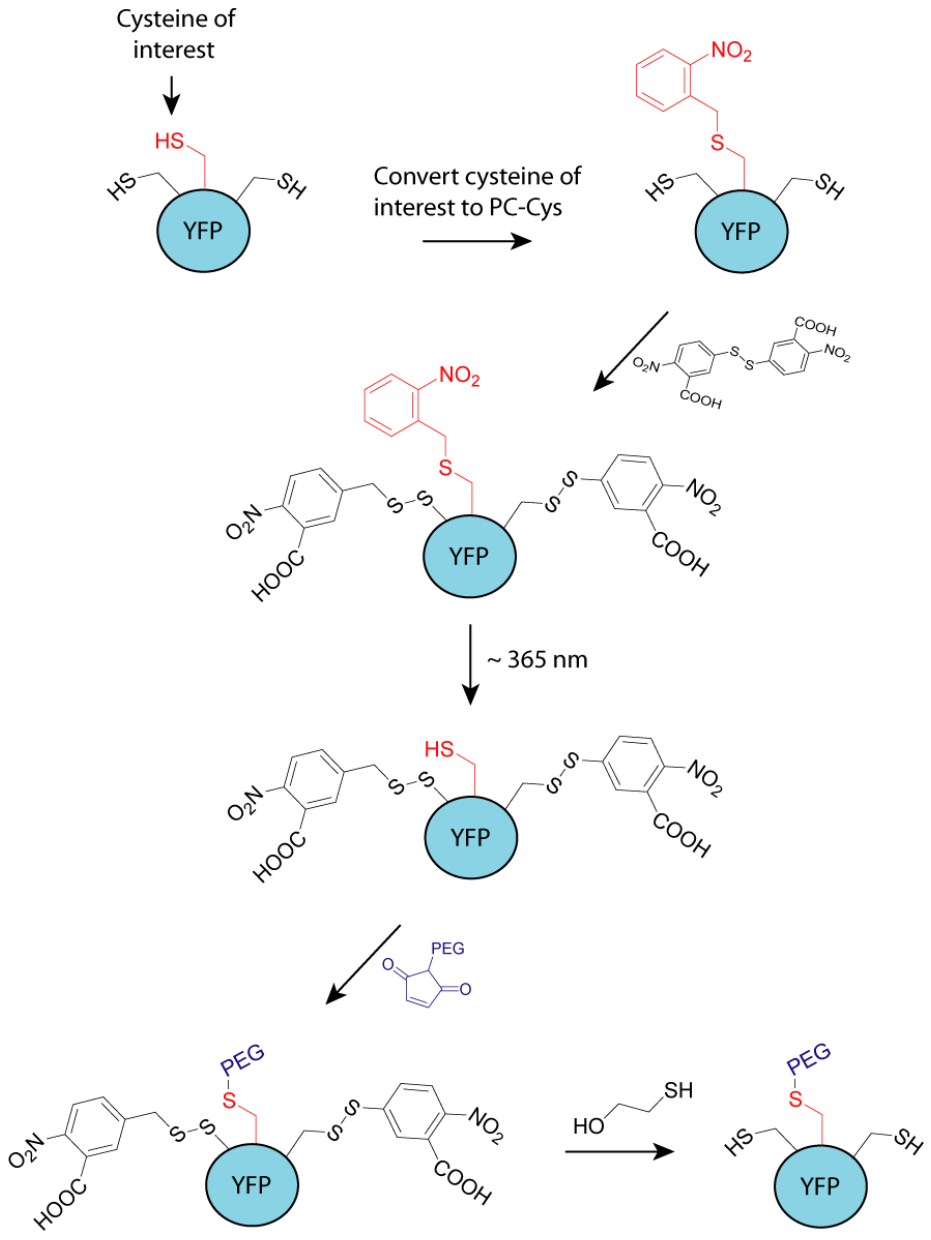


Figure A-3. General scheme of potential site-specific labeling through the use of photo-caged cysteine. Shown here is Ellman's reagent for the cysteine protecting group, and PEG as the specific addition.

Experimental for ONBC control of intein function:

PCR Reactions and Homologous Recombination: Three PCR reactions were performed to obtain the overlapping sections of a GFP-VMA-GFP construct. VMA refers to the VMA intein from *S. Cerevisiae,* which was introduced as an insert, flanked by the N-term and C-term of the GFPuv gene. The intein coding sequence replaced that of the amino acids 107-109 of GFP(uv). The following primers were used to PCR amplify the N-terminal portion of GFP: FWD – CCC GGA TCG GAC TAC TAG CAG CTG TAA TAC GAC TCA CTA TAG GG AAT ATT AAG CTT ACC ATG AGT AAA GGA GAA GAA CTT TTC; REV – TTC CAG CAC ACT GGC GGC CGT TAC TAG TGG ATC CGA GCT CGG TAC CGT AGT TCC CGT CAT C T TTG AAA GA. The red sequence represents the homologous overlapping sequence which matches that in the multiple cloning site of the expression vector used, pYES2 (Invitrogen). The following primers were used to PCR amplify the C-terminal portion of GFP: FWD – GCT GAA GTC AAG TTT GAA GGT GAT ACC CTT GTT AA; REV – ATT ACA TGA TGC GGC CCT CTA GAT GCA TGC TCG AGC GTT AAT GGT GAA TGG TGA TGG TGT TTG TAG AGC TCA TCC ATG CC (CL648). The red sequence on the REV primer represents the homologous overlapping sequence which matches that in the multiple cloning site of pYES2, and the poly his-tag is underlined. The plasmid pGFP(uv) was used as the template for both GFP amplifications with Phusion high-fidelity polymerase (NEB). The following primers were used to PCR amplify the VMA intein portion of the construct: FWD – CAG GAA CGC ACT ATA TCT TTC AAA GAT GAC GGG AAC TAC GCG TGC TTT GCC AAG GGT ACC AAT GTT TTA ATG G; REV – TTA ACA AGG GTA TCA CCT TCA AAC TTG ACT TCA GTT AAC

AAG GGT ATC ACC TTC AAA CTT GAC TTC AGC TCG GCA ATT ATG GAC GAC AAC CTG GTT GGC AAG C. The red sequence represents the homologous overlapping sequence which matches that of GFP, with the cysteine codons underlined. Yeast genomic DNA was used as the template for these PCR reactions with Phusion polymerase.

Three different PCR reactions were performed to obtain the VMA mutants. The following primers were used to PCR amplify VMA with the 5'- cys replaced by TAG (underlined): FWD – CAG GAA CGC ACT ATA TCT TTC AAA GAT GAC GGG AAC TAC GCG TAG TTT GCC AAG GGT ACC AAT GTT TTA ATG G; REV – TTA ACA AGG GTA TCA CCT TCA AAC TTG ACT TCA GTT AAC AAG GGT ATC ACC TTC AAA CTT GAC TTC AGC TCG GCA ATT ATG GAC GAC AAC CTG GTT GGC AAG C. The following primers were used to PCR amplify VMA with the 3'-Cys replaced by TAG: FWD – CAG GAA CGC ACT ATA TCT TTC AAA GAT GAC GGG AAC TAC GCG TGC TTT GCC AAG GGT ACC AAT GTT TTA ATG G; REV – TTA ACA AGG GTA TCA CCT TCA AAC TTG ACT TCA GCT CGC TAA TTA TGG ACG ACA ACC TGG TTG GCA AGC. The following primers were used to PCR amplify VMA with both the 5' and the 3' terminal cys residues mutated to TAG: FWD – CAG GAA CGC ACT ATA TCT TTC AAA GAT GAC GGG AAC TAC GCG TAG TTT GCC AAG GGT ACC AAT GTT TTA ATG G; REV – TTA ACA AGG GTA TCA CCT TCA AAC TTG ACT TCA GCT CGC TAA TTA TGG ACG ACA ACC TGG TTG GCA AGC. The red sequence represents the homologous overlapping sequences which match that of GFP. pYES2-GFP-VMA was used as the template for these reactions with Phusion polymerase.

Overlap PCR reactions were performed with all of the above VMA PCR products and the GFP(C) PCR product. These reactions were carried out using 1 μL of each half reaction and the GFP(C) REV primer. The FWD primer used was that which corresponded to the VMA introduced to PCR reaction. The products obtained were VMA(WT)-GFP(C), VMA5'-GFP(C), VMA3'-GFP(C), and VMA3/5'-GFP(C), all of which were used in homologous recombination events.

Homologous Recombination: Homologous recombination was first performed using linearized pYES2 plasmid (digested with HindIII) cotransformed with the purified GFP(N) PCR product into InvSC cells. The GFP(N) PCR product contained flanking overhangs homologous to the plasmid. This recombination event gave pYES2-GFP(N), which contained the N-terminal portion of the GFP gene inserted into the expression vector. The plasmid pYES2-GFP(N) was digested overnight with EcoRV to obtain linearized plasmid. This was then transformed chemically into InvScI cells along with the overlapping PCR products mentioned above. InvScI cells were inoculated in 100 mL of YPD media and grown to an OD_{600} = 0.80. These cells were then spun down and washed three times with sterile water. The cell pellet was resuspended in 1.5 mL of a lithium acetate solution (10 mM, Tris-HCl, 1 mM EDTA, 100 mM lithium acetate, pH 7.5 in water). To 100 μL of the cells was added 5 μL of linearized pYES2-GFP(N), 5 μL of the desired PCR product, and 600 μL of a PEG solution (10 mM, Tris-HCl, 1 mM EDTA, 100 mM lithium acetate, pH 7.5, in 50% PEG – 3500). This solution was incubated at 30 ^{o}C for 30 min. Following incubation, 70 μL of DMSO was added and mixed thoroughly via pipetting up and down. This was then heat shocked at 42 ^{o}C for 15 min and then placed on ice for 2 min. The resulting transformation was then pelleted and

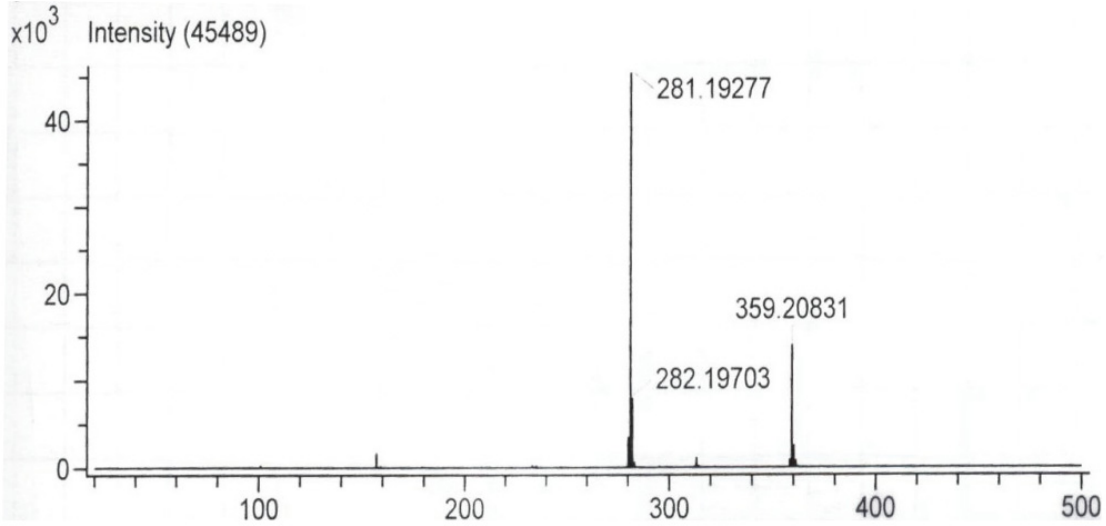
washed three times with 0.9% NaCl. The pellet was then resuspended in 100 µL of 0.9% NaCl and the entire volume was plated on –uracil dropout agar media. These plates were incubated at 30 ^{o}C for 48 hrs.

Colonies from these transformations were inoculated into –uracil dropout media and colony PCR was performed on each to determine if the correct recombination events had occurred. The primers used for these PCR reactions were the GFP(N) FWD primer and the GFP(C) REV primer. For those inoculations in which the correct products were observed, the DNA was then analyzed by the restriction enzymes HindIII and XbaI. Sequencing was then performed on all DNA consistent with the expected restriction sizes. Sequencing results confirmed that the correct mutations were inserted.

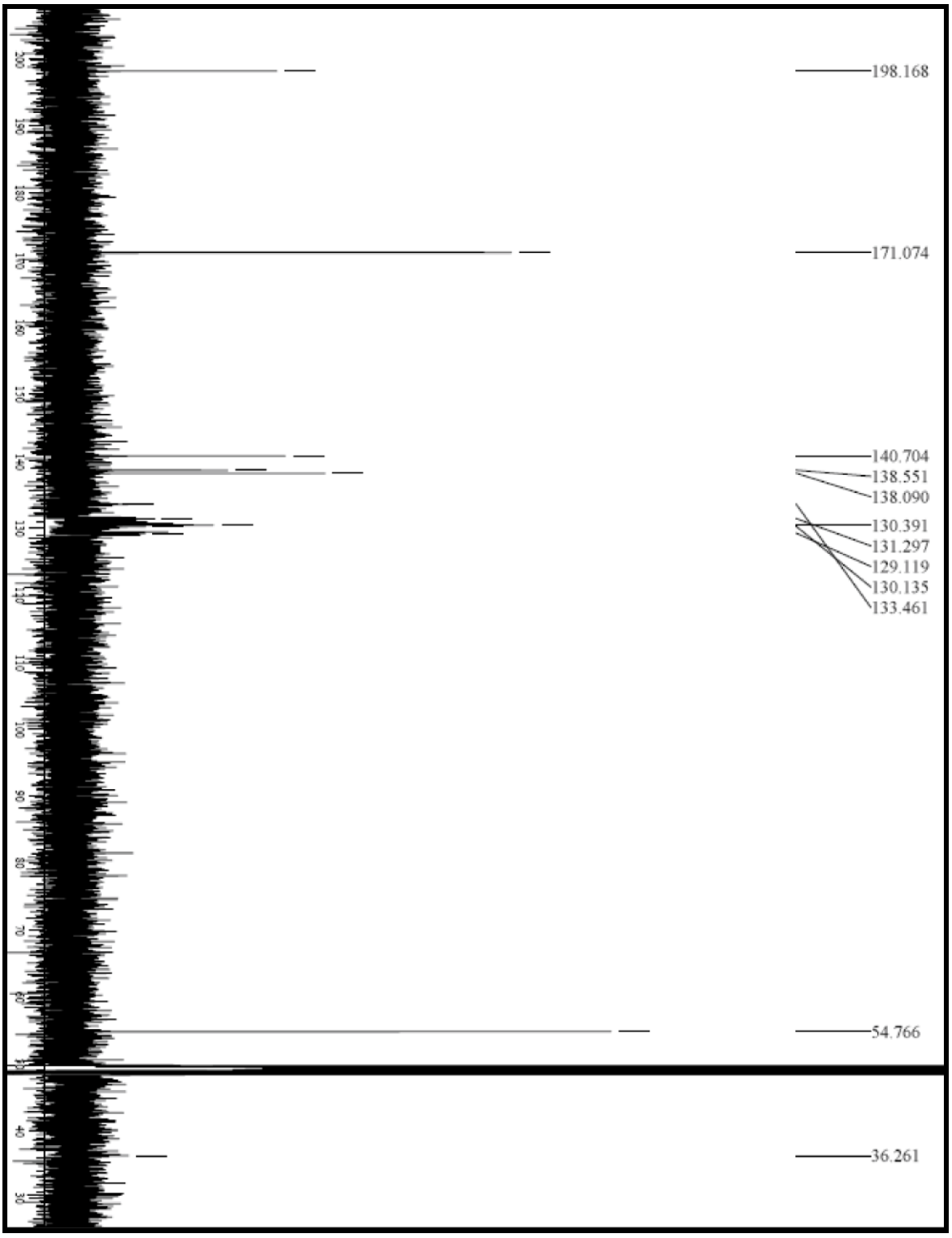
GST sequence used for isotopic study (Phenylalanine of interest is highlighted)

MDPSSRSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELG
LEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRY
GVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYD
ALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFG
GGDHPPKSDLVSRGVDHHHHHH

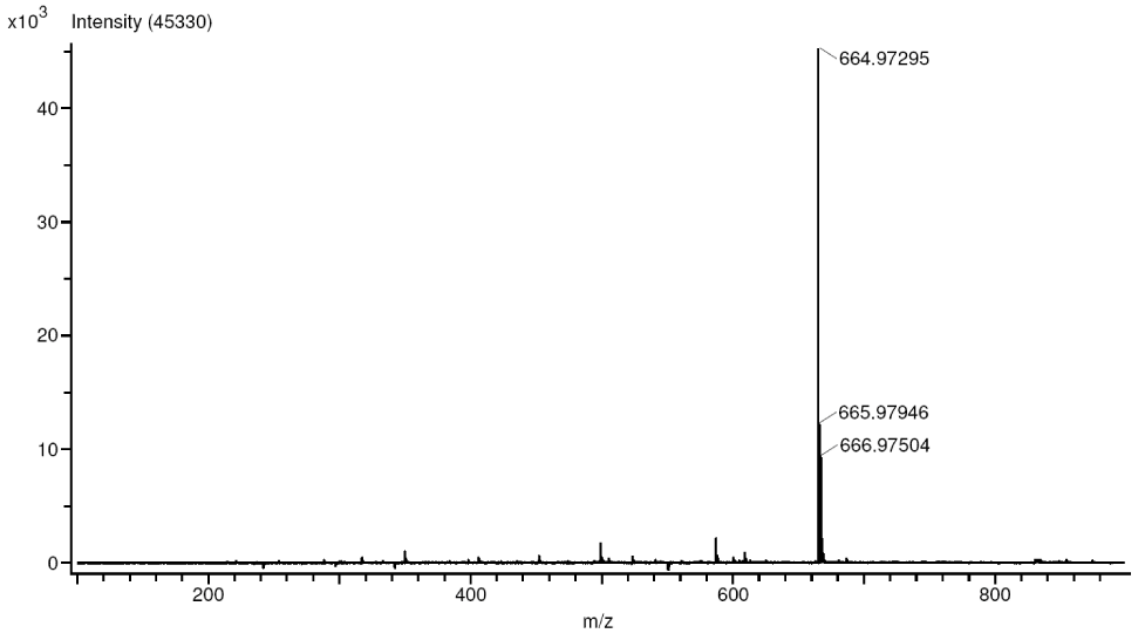
ESI Mass Spectrum for [D_{11}]-pBpa



^{13}C NMR for [D_{11}]-pBpa



ESI Mass Spectrum for 4,5-bis(1,3,2-dithiasolan-2-yl)fluorescein (FlAsH)



GST sequence used for FlAsH labeling study (cysteines of interest are highlighted)

MGCCPGCCSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE
LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDI
RYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLY
DALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATF
GGGDHPPKSDLERPHRDHHHHHH

Construction of the sfGFP expression vectors

The gene encoding sfGFP was synthetically prepared by overlap PCR using homologous oligonucleotides. Oligos 2 through 31 (see below) were combined in a reaction mixture under the following conditions; 0.2 mM of each dNTPs, 0.02 U/μL of DeepVent polymerase (NEB), 1 μM of each #2~#31 oligos, 20 mM Tris-HCl (pH 8.8 at 25 ºC), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$ and 0.1% Triton X-100 in 50 μL reaction. This mixture was then thermal cycled; initial denaturation at 95 ºC for 5 min,

25 cycles of 95 ºC for 30 s, 57 ºC for 30 s and 72 ºC for 1 min and final extension at 72 ºC for 10 min. The synthetic gene was then amplified by PCR using oligos 1 and 32 in the following reaction; 1 μL of assembly PCR product, 0.2 mM of each dNTPs, 0.02 U/μL of DeepVent polymerase (NEB), 1 μM of each #1 and #32 oligos, 20 mM Tris-HCl (pH 8.8 at 25 ºC), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$ and 0.1% Triton X-100 in 50 mL reaction, under the same cycle conditions. The second PCR yielded a single DNA fragment which was digested with *Nhe*I and *Hind*III and was cloned into the same sites on the pTrcHisA vector (Invitrogen).

The mutations to the sfGFP gene, 66TAG or 150TAG were performed via Quikchange mutagenesis with the corresponding oligos, Tyr66TAG-Fwd and -Rev or Val150TAG-Fwd and -Rev, respectively. The reaction conditions were; 2.4 ng pTrcHisA-sfGFP, 0.2 mM of each dNTPs, 0.01 U/μL of Phusion polymerase (NEB) and 1X Phusion HF buffer in 20 mL reaction. The following cycling conditions were used; initial denaturation at 98 ºC for 30 s, 18 cycles of 98 ºC for 10 s, 58 ºC for 30 s and 72 ºC for 3 min and final extension at 72 ºC for 10 min. This PCR product was digested overnight with DpnI and transformed into E. coli (DH-10B) to provide the correct mutations. The identity of mutation was confirmed by DNA sequencing.

Sequences of Oligonucleotides:

#1: CTA*gctagc*ATGATGAGCAAAGGCGAAGAACTGTTTACCGGCGTGGTTCCG (*Nhe* I)

#2: ATCGCCATCCAGTTCCACCAGAATCGGAACCACGCCGGTAAA

#3: GGTGGAACTGGATGGCGATGTAAATGGCCACAAGTTTAGCGT

#4: TCGCCCTCCCCTTCCCCACGCACGCTAAACTTGTGGCCATTT

#5: GGGAAGGGGAGGGCGATGCGACCAATGGCAAACTGACCCTGA

#6: TTCCCGGTCGTGCAAATAAACTTCAGGGTCAGTTTGCCATTG

#7: GTTTATTTGCACGACCGGGAAACTGCCGGTTCCTTGGCCCAC

#8: GCCATACGTCAGGGTGGTGACAAGGGTGGGCCAAGGAACCGG

#9: CACCACCCTGACGTATGGCGTGCAATGCTTTAGCCGTTACCC

#10: TCATGCCGCTTCATGTGGTCCGGGTAACGGCTAAAGCATTGC

#11: ACCACATGAAGCGGCATGACTTCTTCAAAAGCGCCATGCCTG

#12: CGTCCGTTCCTGAACATAGCCTTCAGGCATGGCGCTTTTGAA

#13: GCTATGTTCAGGAACGGACGATCTCGTTTAAGGATGACGGCA

#14: CCTCCGCACGGGTCTTATAGGTGCCGTCATCCTTAAACGAGA

#15: CTATAAGACCCGTGCGGAGGTCAAATTCGAAGGCGATACCCT

#16: TCAGCTCAATGCGGTTCACCAGGGTATCGCCTTCGAATTTGA

#17: GGTGAACCGCATTGAGCTGAAGGGCATCGACTTCAAAGAGGA

#18: TTGTGCCCCAGTATGTTGCCATCCTCTTTGAAGTCGATGCCC

#19: GGCAACATACTGGGGCACAAGCTGGAGTACAACTTCAACAGC

#20: GCGGTGATGTAGACGTTGTGGCTGTTGAAGTTGTACTCCAGC

#21: CACAACGTCTACATCACCGCCGACAAGCAGAAGAACGGCATT

#22: CCGAATCTTGAAGTTGGCCTTAATGCCGTTCTTCTGCTTGTC

#23: AAGGCCAACTTCAAGATTCGGCACAATGTGGAGGACGGAAGC

#24: GATAATGATCCGCCAGCTGAACGCTTCCGTCCTCCACATTGT

#25: GTTCAGCTGGCGGATCATTATCAACAGAATACCCCCATTGGC

#26: CCGGCAGAAGCACGGGACCGTCGCCAATGGGGGTATTCTGTT

#27: CCCGTGCTTCTGCCGGATAATCATTACTTGAGCACCCAGAGC

#28: ATTCGGGTCCTTGCTCAGCACGCTCTGGGTGCTCAAGTAATG

#29: GCTGAGCAAGGACCCGAATGAGAAACGGGATCACATGGTGCT

#30: CGCAGCGGTCACAAATTCCAGCAGCACCATGTGATCCCGTTT

#31: GGAATTTGTGACCGCTGCGGGCATTACACATGGCATGGATGA

#32: CCC*aagctt*TTATTTATACAGTTCATCCATGCCATGTGTAATGCCCG

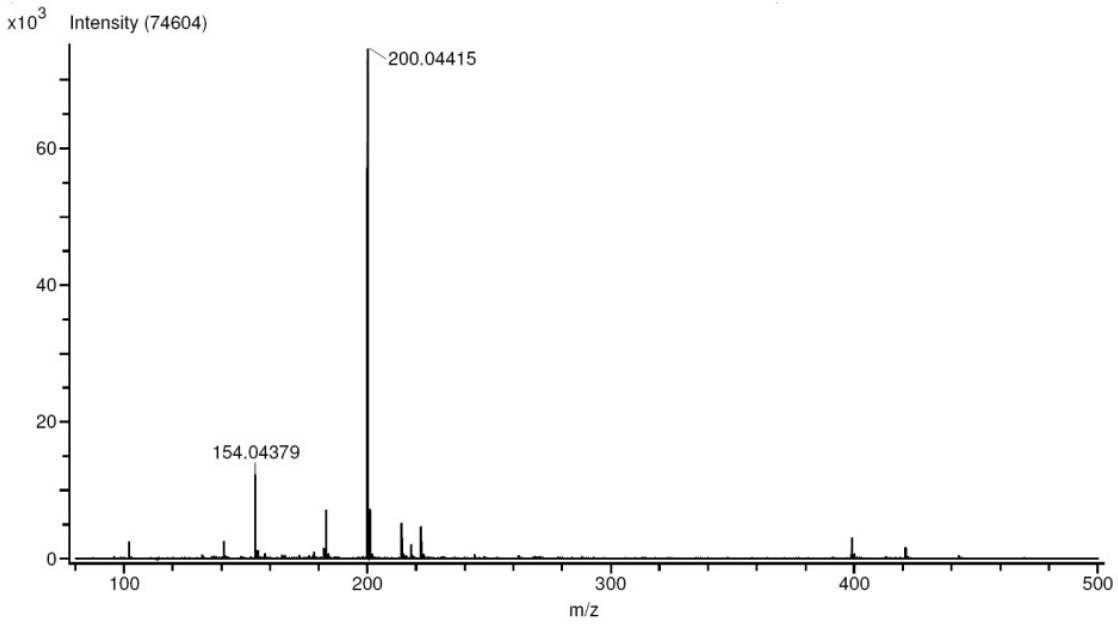
Tyr66TAGFwd: CCCTTGTCACCACCCTGACG*tag*GGCGTGCAATGC

Tyr66TAGRev: CGTCAGGGTGGTGACAAGGGTGGGCCAAGGAA

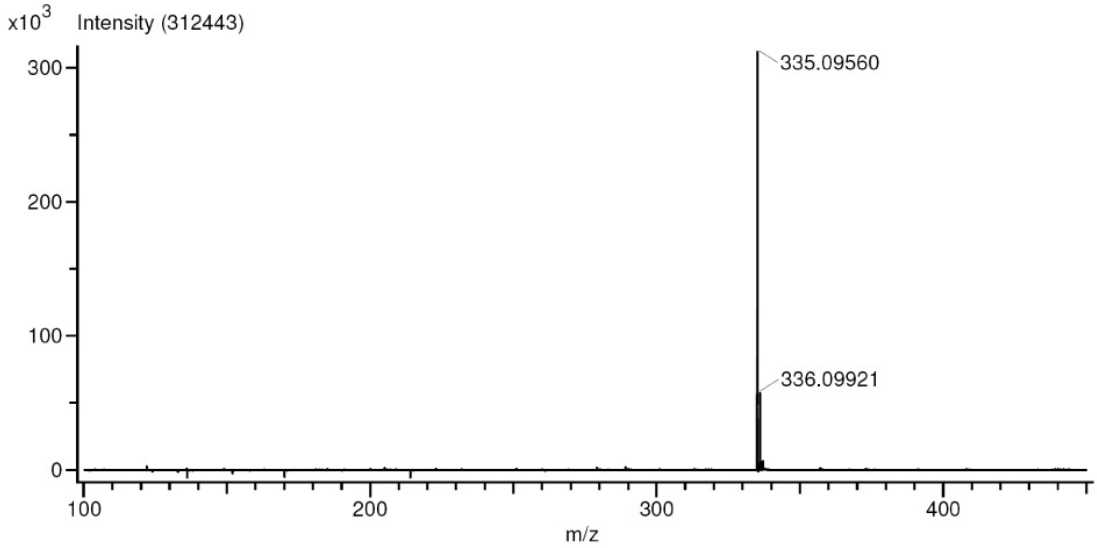
Val150TAGFwd: TTCAACAGCCACAAC*tag*TACATCACCGCC

Val150TAGRev: GTTGTGGCTGTTGAAGTTGTACTCCAG

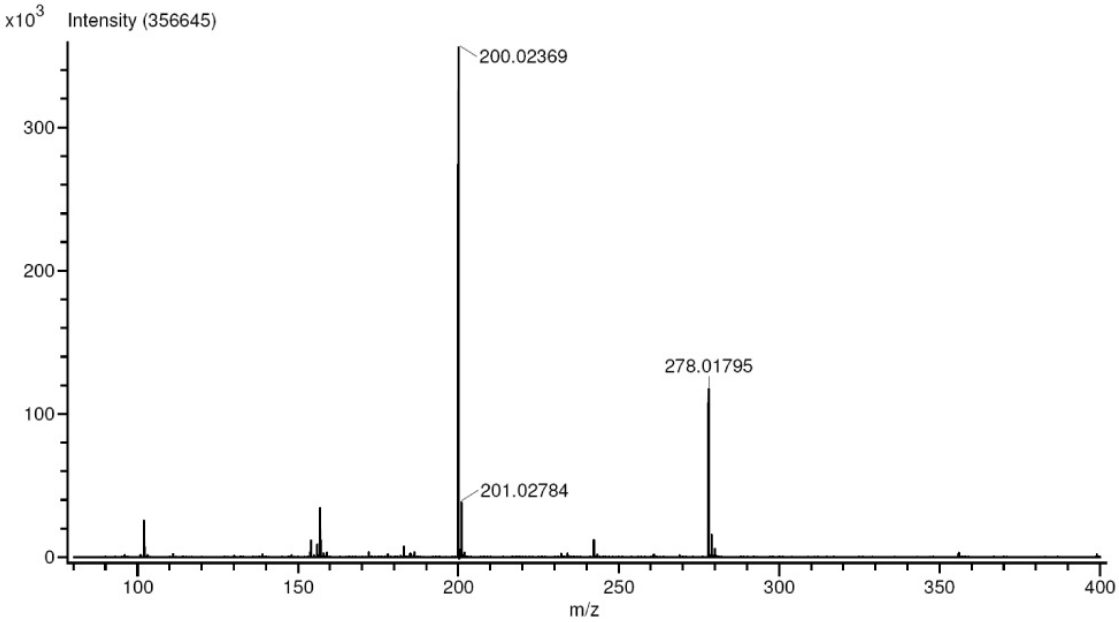
ESI Mass Spectrum for 2-fluorotyrosine



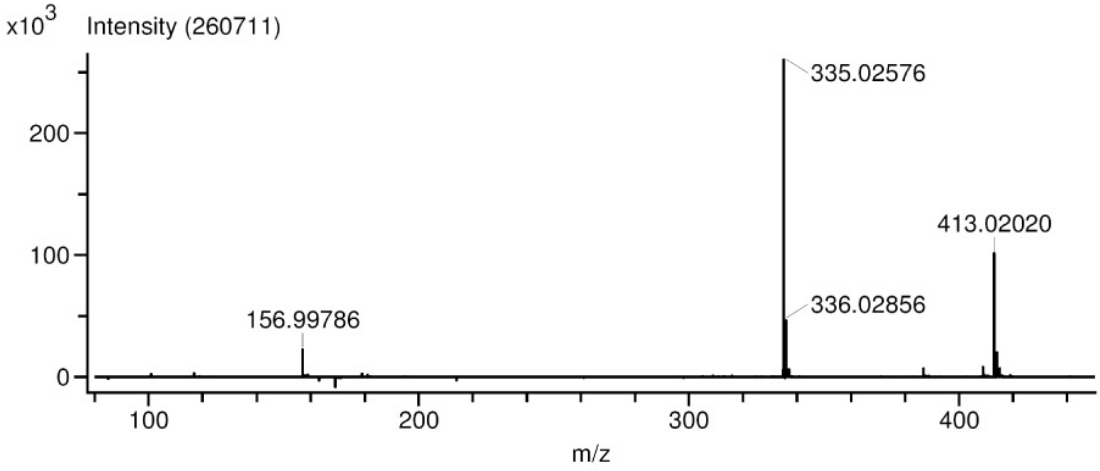
ESI Mass Spectrum for *o*-nitrobenzyl-2-fluorotyrosine



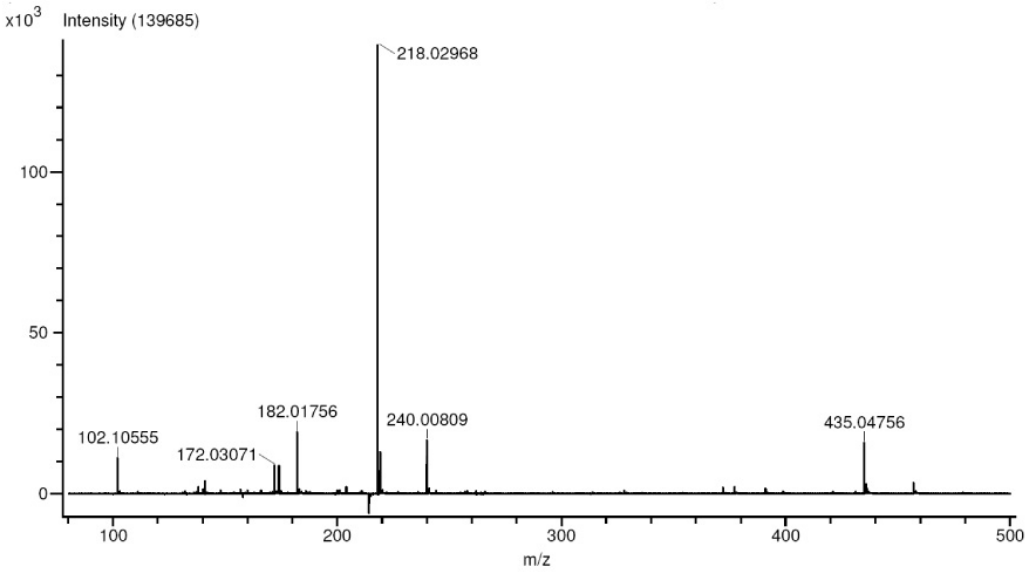
ESI Mass Spectrum for 3-fluorotyrosine



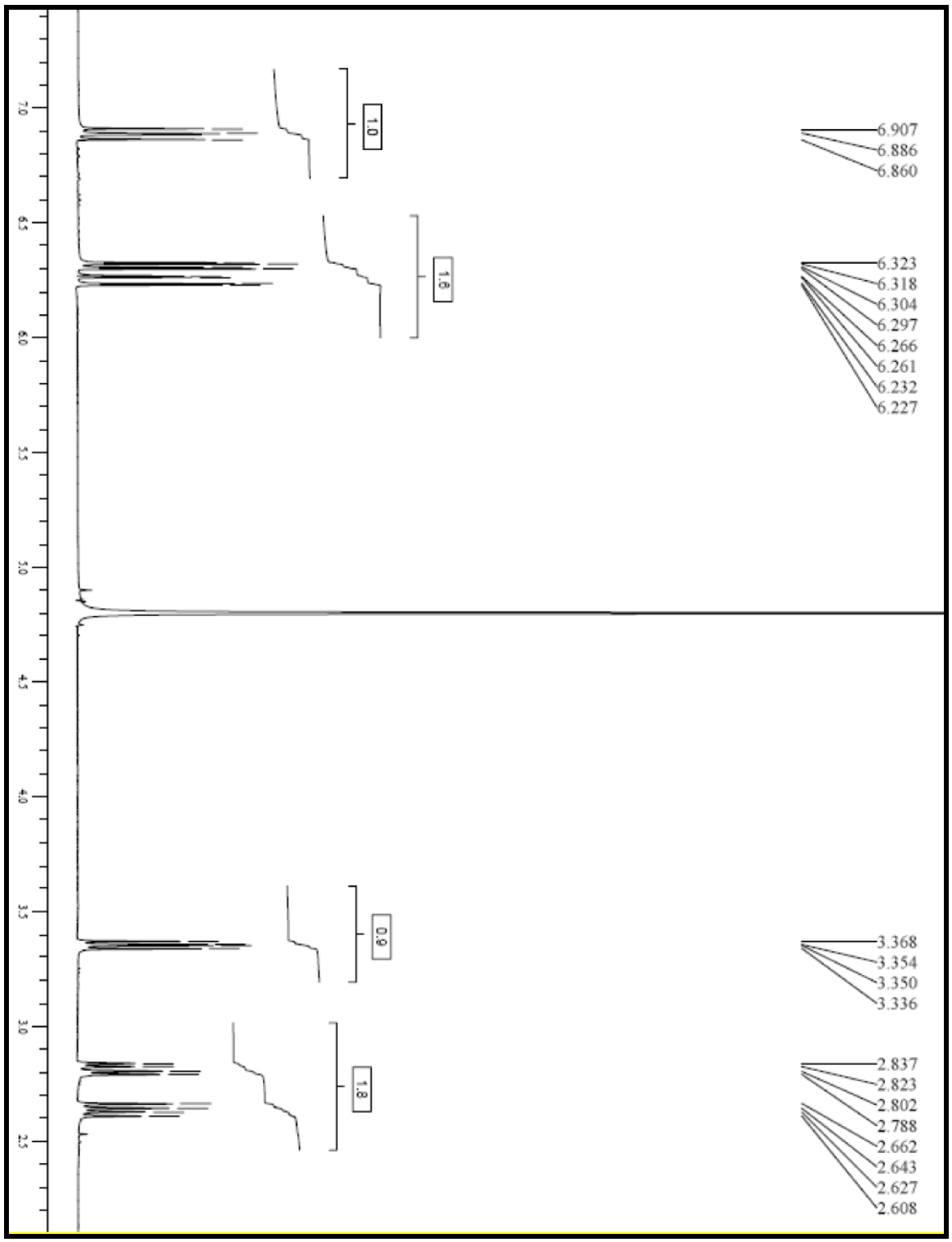
ESI Mass Spectrum for *o*-nitrobenzyl-3-fluorotyrosine



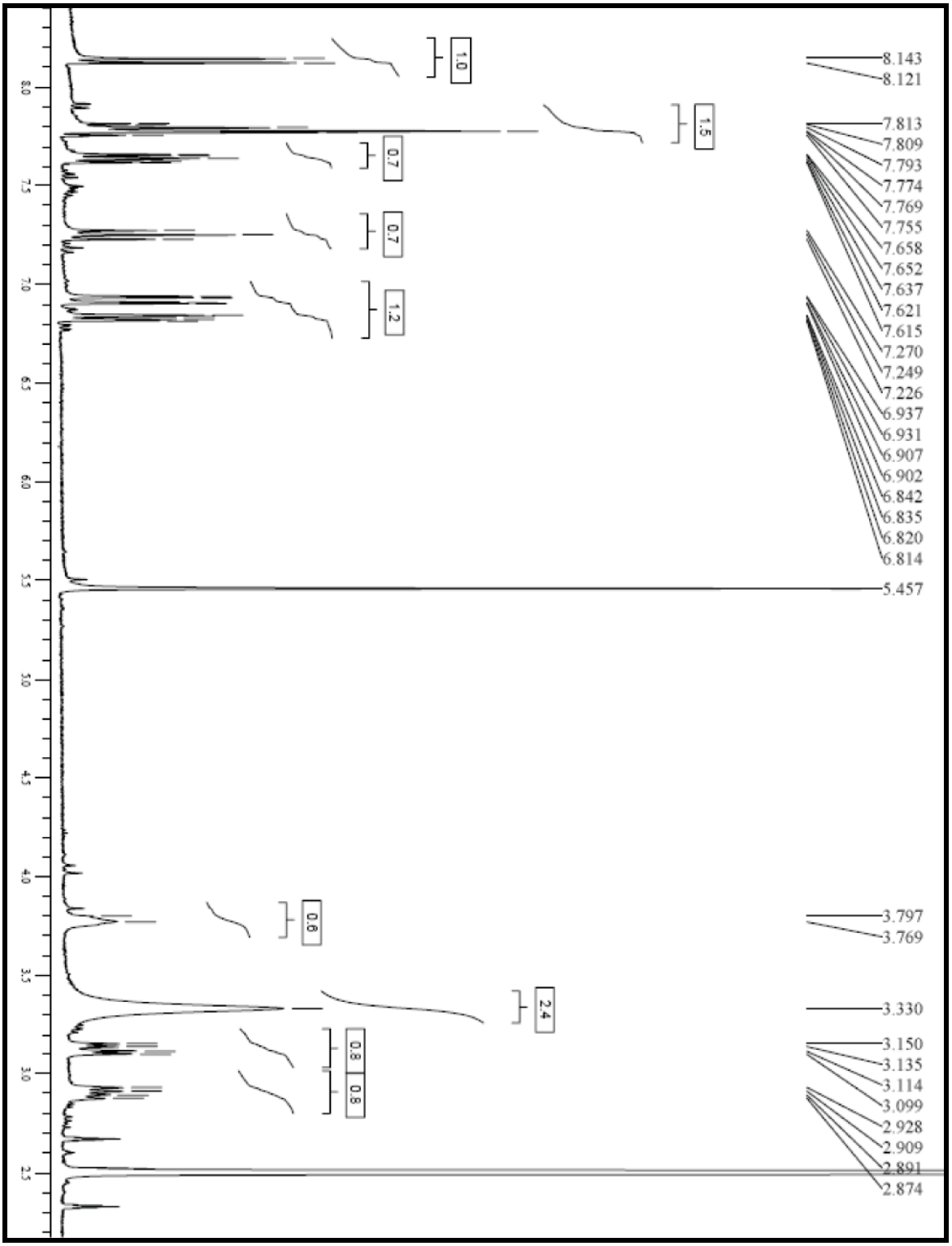
ESI Mass Spectrum for 2,6-difluorotyrosine



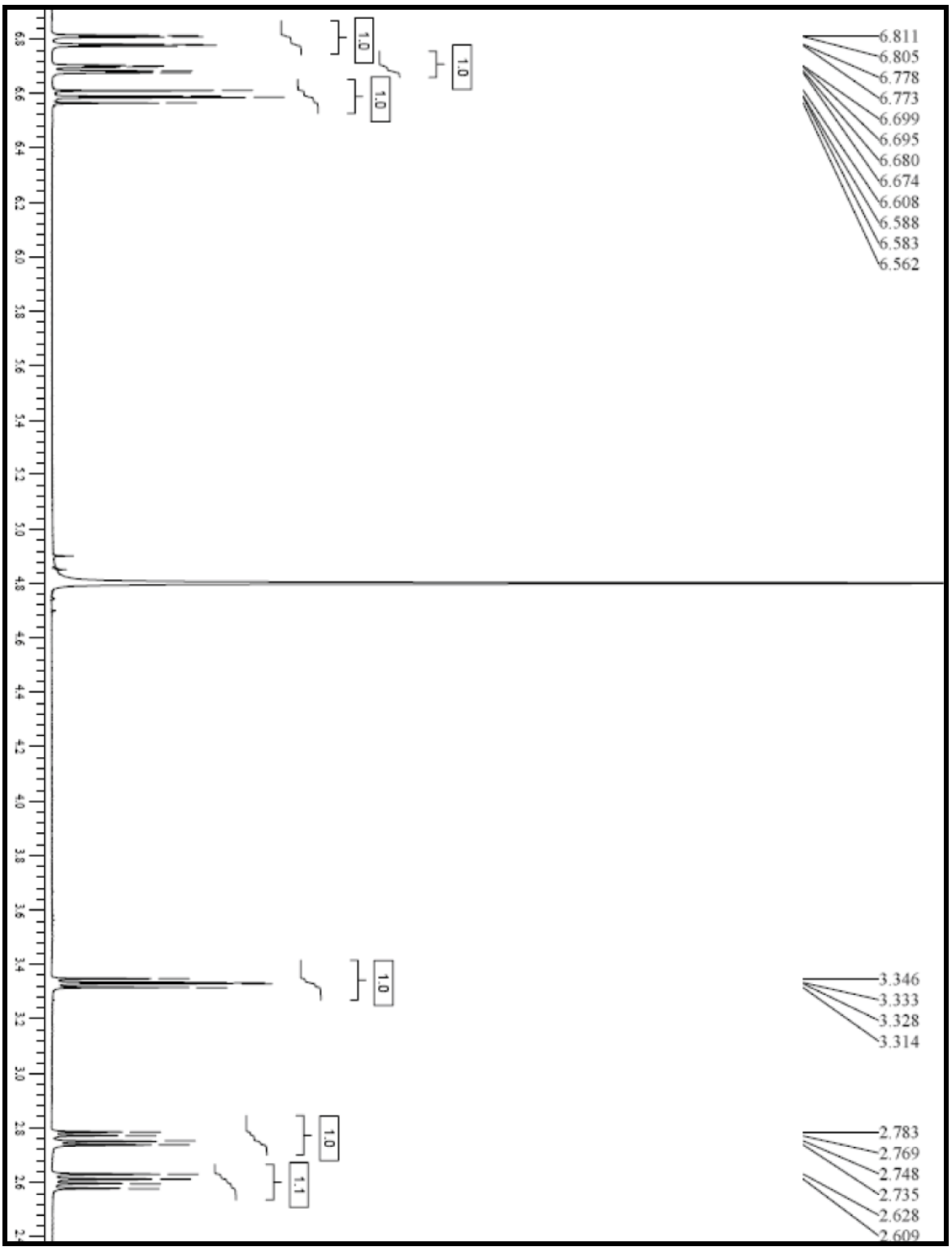
^{1}H NMR for 2-fluorotyrosine



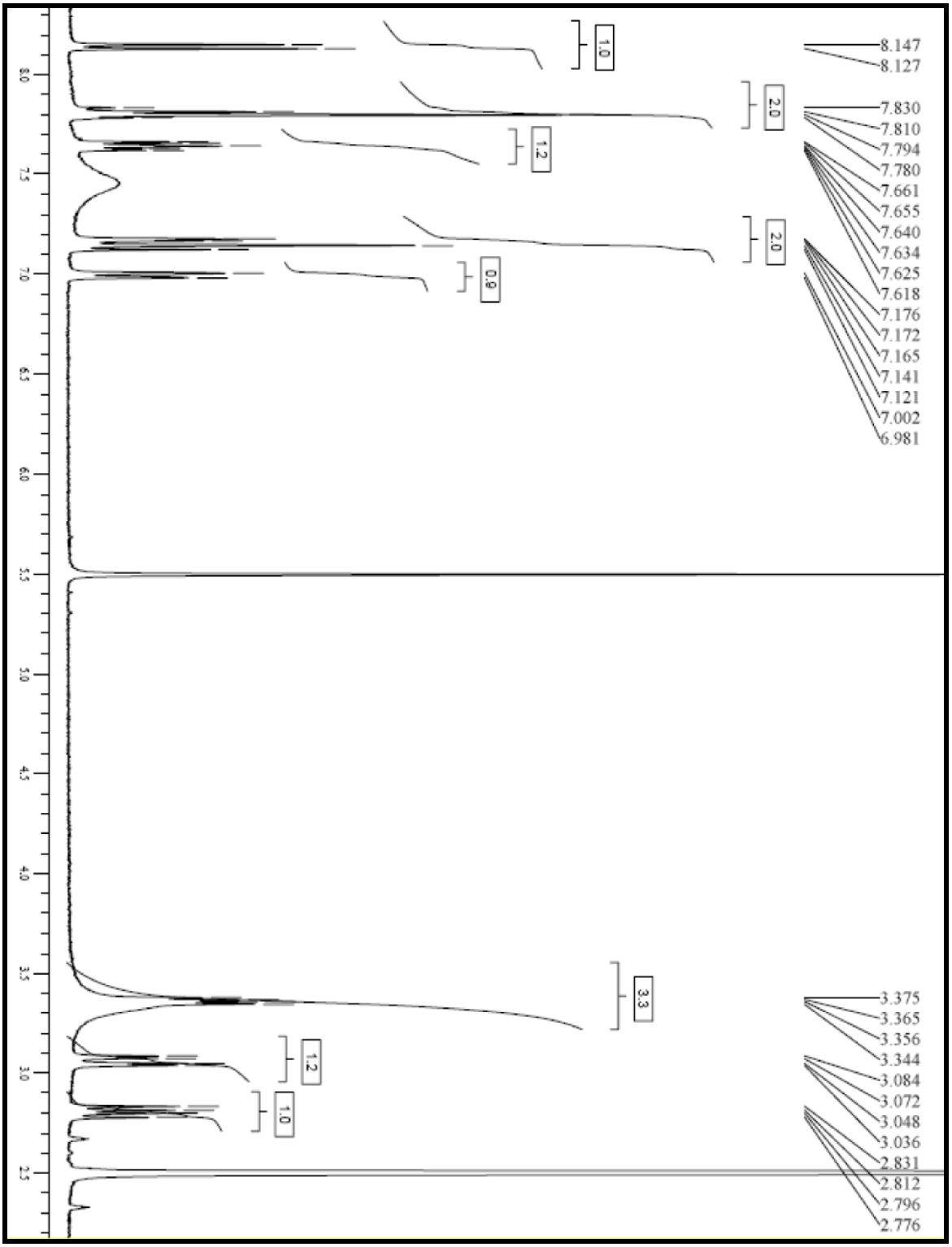
[1]H NMR for *o*-nitrobenzyl-2-fluorotyrosine



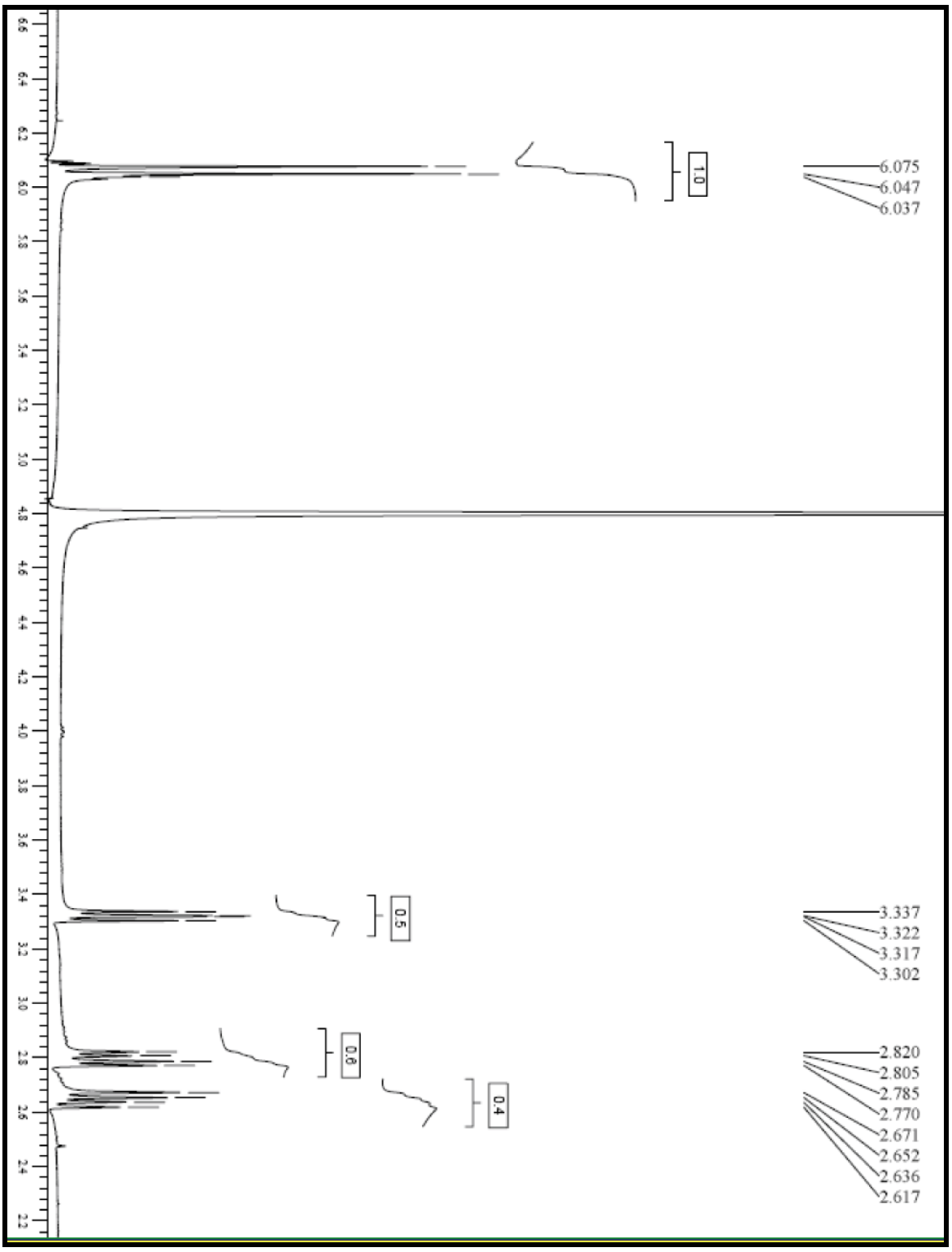
^{1}H NMR for 3-fluorotyrosine



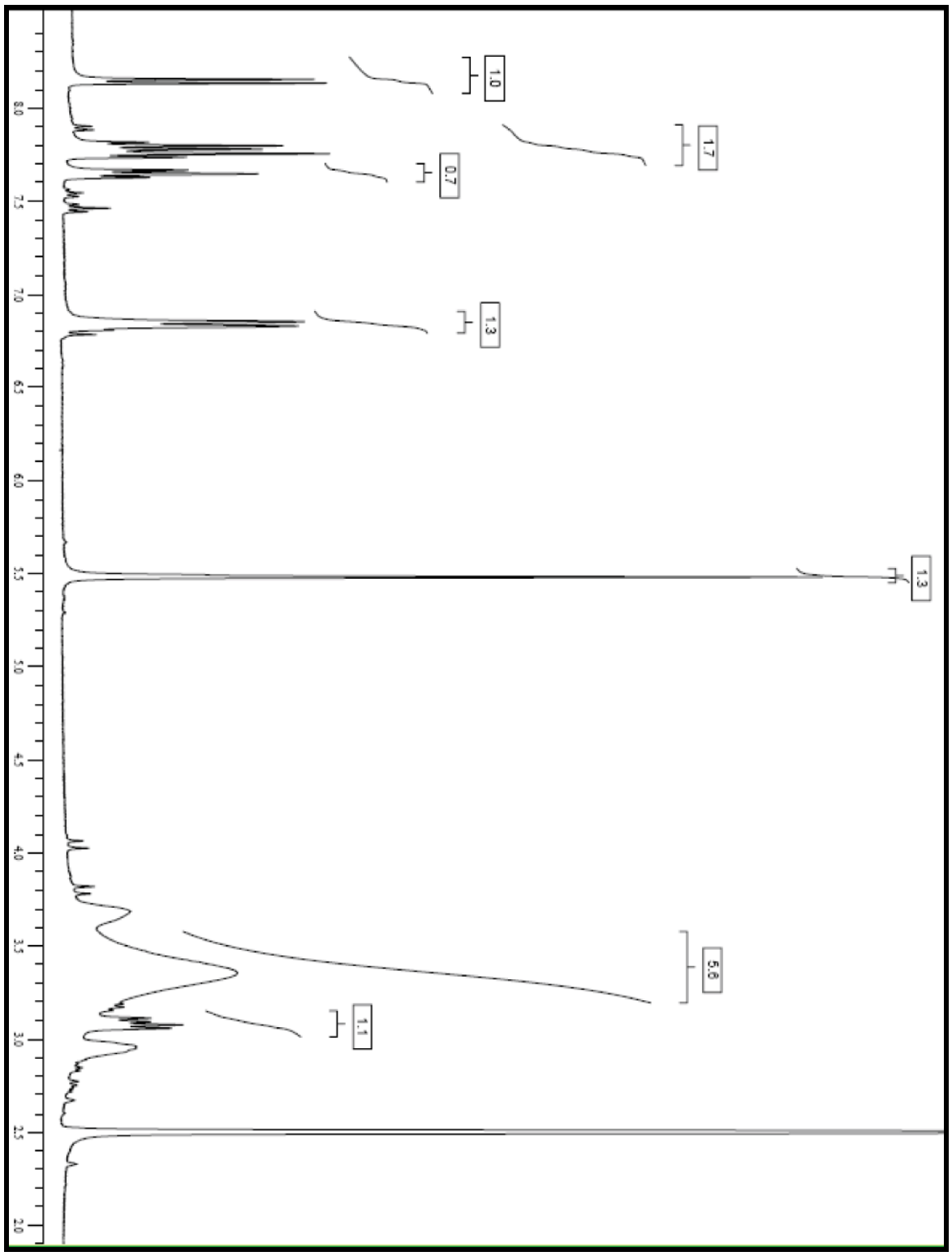
^{1}H NMR for *o*-nitrobenzyl-3-fluorotyrosine



^{1}H NMR for 2,6-difluorotyrosine



^{1}H NMR for *o*-nitrobenzyl-2,6-difluorotyrosine



Bibliography

1. Turro, N. J., and Schuster, G. (1975) Photochemical Reactions as a Tool in Organic Syntheses, *Science 187*, 303-312.

2. Turro, N. J. (1969) Mechanisms of sensitized photochemical cis-trans-isomerization in solution, *Photochem Photobiol 9*, 555-563.

3. Bock, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B., and Zinoni, F. (1991) Selenocysteine: the 21st amino acid, *Mol Microbiol 5*, 515-520.

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

5. Vermeer, C. (1990) Gamma-carboxyglutamate-containing proteins and the vitamin K-dependent carboxylase, *Biochem J 266*, 625-636.

6. Bhattacharjee, A., and Bansal, M. (2005) Collagen structure: the Madras triple helix and the current scenario, *IUBMB Life 57*, 161-172.

7. Blenis, J., and Resh, M. D. (1993) Subcellular localization specified by protein acylation and phosphorylation, *Curr Opin Cell Biol 5*, 984-989.

8. Park, M. H. (2006) The post-translational synthesis of a polyamine-derived amino acid, hypusine, in the eukaryotic translation initiation factor 5A (eIF5A), *J Biochem 139*, 161-169.

9. Brosnan, J. T., and Brosnan, M. E. (2006) The sulfur-containing amino acids: an overview, *J Nutr 136*, 1636S-1640S.

10. Kivirikko, K. I., and Pihlajaniemi, T. (1998) Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases, *Adv Enzymol Relat Areas Mol Biol 72*, 325-398.

11. Wang, L., Zhang, Z., Brock, A., and Schultz, P. G. (2003) Addition of the keto functional group to the genetic code of Escherichia coli, *Proc Natl Acad Sci U S A 100*, 56-61.

12. Zhang, Z., Smith, B. A., Wang, L., Brock, A., Cho, C., and Schultz, P. G. (2003) A new strategy for the site-specific modification of proteins in vivo, *Biochemistry 42*, 6735-6746.

13. Deiters, A., Cropp, T. A., Summerer, D., Mukherji, M., and Schultz, P. G. (2004) Site-specific PEGylation of proteins containing unnatural amino acids, *Bioorg Med Chem Lett 14*, 5743-5745.

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

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

16. Chin, J. W., and Schultz, P. G. (2002) In vivo photocrosslinking with unnatural amino acid mutagenesis, *Chembiochem 3*, 1135-1137.

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

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

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

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

21. Miller, J. C., Silverman, S. K., England, P. M., Dougherty, D. A., and Lester, H. A. (1998) Flash decaging of tyrosine sidechains in an ion channel, *Neuron 20*, 619-624.

22. Philipson, K. D., Gallivan, J. P., Brandt, G. S., Dougherty, D. A., and Lester, H. A. (2001) Incorporation of caged cysteine and caged tyrosine into a transmembrane segment of the nicotinic ACh receptor, *American Journal of Physiology-Cell Physiology 281*, C195-C206.

23. Pelliccioli, A. P., and Wirz, J. (2002) Photoremovable protecting groups: reaction mechanisms and applications, *Photochem Photobiol Sci 1*, 441-458.

24. Goeldner, M., and Givens, R. (2005) *Dynamic studies in biology : phototriggers, photoswitches and caged biomolecules*, Wiley-VCH, Weinheim.

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

26. Wilkins, B. J., Yang, X., and Cropp, T. A. (2009) Photochemical control of FlAsH labeling of proteins, *Bioorg Med Chem Lett 19*, 4296-4298.

27. Buskirk, A. R., and Liu, D. R. (2005) Creating small-molecule-dependent switches to modulate biological functions, *Chem Biol 12*, 151-161.

28. Chaulk, S. G., and MacMillan, A. M. (1998) Caged RNA: photo-control of a ribozyme reaction, *Nucleic Acids Res 26*, 3173-3178.

29. Lin, W., Albanese, C., Pestell, R. G., and Lawrence, D. S. (2002) Spatially discrete, light-driven protein expression, *Chem Biol 9*, 1347-1353.

30. Goedhart, J., and Gadella, T. W., Jr. (2004) Photolysis of caged phosphatidic acid induces flagellar excision in Chlamydomonas, *Biochemistry 43*, 4263-4271.

31. Hobartner, C., and Silverman, S. K. (2005) Modulation of RNA tertiary folding by incorporation of caged nucleotides, *Angew Chem Int Ed Engl 44*, 7305-7309.

32. Wenter, P., Furtig, B., Hainard, A., Schwalbe, H., and Pitsch, S. (2006) A caged uridine for the selective preparation of an RNA fold and determination of its refolding kinetics by real-time NMR, *Chembiochem 7*, 417-420.

33. Kaplan, J. H., Forbush, B., 3rd, and Hoffman, J. F. (1978) Rapid photolytic release of adenosine 5'-triphosphate from a protected analogue: utilization by the Na:K pump of human red blood cell ghosts, *Biochemistry 17*, 1929-1935.

34. Mayer, G., and Heckel, A. (2006) Biologically active molecules with a "light switch", *Angew Chem Int Ed Engl 45*, 4900-4921.

35. Young, D. D., and Deiters, A. (2007) Photochemical control of biological processes, *Org Biomol Chem 5*, 999-1005.

36. Lusic, H., and Deiters, A. (2006) A new photocaging group for aromatic N-heterocycles, *Synthesis-Stuttgart*, 2147-2150.

37. Corrie, J. E., Barth, A., Munasinghe, V. R., Trentham, D. R., and Hutter, M. C. (2003) Photolytic cleavage of 1-(2-nitrophenyl)ethyl ethers involves two parallel pathways and product release is rate-limited by decomposition of a common hemiacetal intermediate, *J Am Chem Soc 125*, 8546-8554.

38. Engels, J., and Schlaeger, E. J. (1977) Synthesis, structure, and reactivity of adenosine cyclic 3',5'-phosphate benzyl triesters, *J Med Chem 20*, 907-911.

39. Goard, M., Aakalu, G., Fedoryak, O. D., Quinonez, C., St Julien, J., Poteet, S. J., Schuman, E. M., and Dore, T. M. (2005) Light-mediated inhibition of protein synthesis, *Chem Biol 12*, 685-693.

40. Ghosn, B., Haselton, F. R., Gee, K. R., and Monroe, W. T. (2005) Control of DNA hybridization with photocleavable adducts, *Photochem Photobiol 81*, 953-959.

41. Ando, H., Furuta, T., Tsien, R. Y., and Okamoto, H. (2001) Photo-mediated gene activation using caged RNA/DNA in zebrafish embryos, *Nat Genet 28*, 317-325.

42. Chou, C. J., Young, D. D., and Deiters, A. (2009) A light-activated DNA polymerase, *Angewandte Chemie-International Edition 48*, 5950-5953.

43. Miller, J. C., Silverman, S. K., England, P. M., Dougherty, D. A., and Lester, H. A. (1998) Flash decaging of tyrosine sidechains in an ion channel, *Neuron 20*, 619-624.

44. Chou, C., Young, D. D., and Deiters, A. (2009) A light-activated DNA polymerase, *Angew Chem Int Ed Engl 48*, 5950-5953.

45. Mendel, D., Ellman, J. A., and Schultz, P. G. (1991) Construction of a light-activated protein by unnatural amino acid mutagenesis, *J Am Chem Soc 113*, 2758-2760.

46. Short, G. F., Lodder, M., Laikhter, A. L., Arslan, T., and Hecht, S. M. (1999) Caged HIV-1 protease: Dimerization is independent of the ionization state of the active site aspartates, *J Am Chem Soc 121*, 478-479.

47. Lodder, M., Golovine, S., Laikhter, A. L., Karginov, V. A., and Hecht, S. M. (1998) Misacylated transfer RNAs having a chemically removable protecting group, *Journal of Organic Chemistry 63*, 794-803.

48. Endo, M., Nakayama, K., and Majima, T. (2004) Design and synthesis of photochemically controllable restriction endonuclease BamHI by manipulating the salt-bridge network in the dimer interface, *Journal of Organic Chemistry 69*, 4292-4298.

49. Yuan, J., Shaham, S., Ledoux, S., Ellis, H. M., and Horvitz, H. R. (1993) The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 beta-converting enzyme, *Cell 75*, 641-652.

50. Fowler, A. V., and Zabin, I. (1970) The amino acid sequence of beta galactosidase. I. Isolation and composition of tryptic peptides, *J Biol Chem 245*, 5032-5041.

51. Matthews, B. W. (2005) The structure of E. coli beta-galactosidase, *C R Biol 328*, 549-556.

52. Yuan, J., Martinez-Bilbao, M., and Huber, R. E. (1994) Substitutions for Glu-537 of beta-galactosidase from Escherichia coli cause large decreases in catalytic activity, *Biochem J 299 (Pt 2)*, 527-531.

53. Li, Y., Korolev, S., and Waksman, G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation, *EMBO J 17*, 7514-7525.

54. Suzuki, M., Baskin, D., Hood, L., and Loeb, L. A. (1996) Random mutagenesis of Thermus aquaticus DNA polymerase I: concordance of immutable sites in vivo with the crystal structure, *Proc Natl Acad Sci U S A 93*, 9670-9675.

55. Yeh, H. J. C., Kirk, K. L., and Cohen, L. A. (1975) *J Chem Soc Perkin Trans 2*, 928-934.

56. Irie, M. (1997) Structures and functions of ribonucleases, *Yakugaku Zasshi 117*, 561-582.

57. Jackson, D. Y., Burnier, J., Quan, C., Stanley, M., Tom, J., and Wells, J. A. (1994) A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues, *Science 266*, 243-247.

58. Song, L., Teng, Q., Phillips, R. S., Brewer, J. M., and Summers, A. O. (2007) (19)F-NMR Reveals Metal and Operator-induced Allostery in MerR, *J Mol Biol*.

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

60. Khan, F., Kuprov, I., Craggs, T. D., Hore, P. J., and Jackson, S. E. (2006) (19)F NMR studies of the native and denatured states of green fluorescent protein, *J Am Chem Soc 128*, 10729-10737.

61. Frieden, C., Hoeltzli, S. D., and Bann, J. G. (2004) The preparation of 19F-labeled proteins for NMR studies, *Methods Enzymol 380*, 400-415.

62. Kim, K., and Cole, P. A. (1998) Kinetic analysis of a protein tyrosine kinase reaction transition state in the forward and reverse directions, *J Am Chem Soc 120*, 6851-6858.

63. Seyedsayamdost, M. R., and Stubbe, J. (2006) Site-specific replacement of Y356 with 3,4-dihydroxyphenylalanine in the beta2 subunit of E. coli ribonucleotide reductase, *J Am Chem Soc 128*, 2522-2523.

64. Seyedsayamdost, M. R., Reece, S. Y., Nocera, D. G., and Stubbe, J. (2006) Mono-, di-, tri-, and tetra-substituted fluorotyrosines: new probes for enzymes that use tyrosyl radicals in catalysis, *J Am Chem Soc 128*, 1569-1579.

65. Ablooglu, A. J., Till, J. H., Kim, K., Parang, K., Cole, P. A., Hubbard, S. R., and Kohanski, R. A. (2000) Probing the catalytic mechanism of the insulin receptor kinase with a tetrafluorotyrosine-containing peptide substrate, *J Biol Chem 275*, 30394-30398.

66. Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) Synthesis of proteins by native chemical ligation, *Science 266*, 776-779.

67. Dawson, P. E., and Kent, S. B. (2000) Synthesis of native proteins by chemical ligation, *Annu Rev Biochem 69*, 923-960.

68. Muralidharan, V., and Muir, T. W. (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins, *Nat Methods 3*, 429-438.

69. Boerema, D. J., Tereshko, V. A., and Kent, S. B. (2007) Total synthesis by modern chemical ligation methods & high resolution (1.1Angstrom) X-ray structure of ribonuclease A, *Biopolymers*.

70. Huber, R. E., and Criddle, R. S. (1967) Comparison of the chemical properties of selenocysteine and selenocystine with their sulfur analogs, *Arch Biochem Biophys 122*, 164-173.

71. Hondal, R. J., and Raines, R. T. (2002) Semisynthesis of proteins containing selenocysteine, *Methods Enzymol 347*, 70-83.

72. Eckenroth, B., Harris, K., Turanov, A. A., Gladyshev, V. N., Raines, R. T., and Hondal, R. J. (2006) Semisynthesis and characterization of mammalian thioredoxin reductase, *Biochemistry 45*, 5158-5170.

73. Ping, Z., and Hilvert, D. (1989) Conversion of a protease into an acyl transferase: selenosubtilisin, *J Am Chem Soc 111*, 4513-4514.

74. Cornish, V. W., Mendel, D., and Schultz, P. G. (1995) Probing protein-structure and function with an expanded genetic-code, *Angewandte Chemie-International Edition in English 34*, 621-633.

75. Beene, D. L., Dougherty, D. A., and Lester, H. A. (2003) Unnatural amino acid mutagenesis in mapping ion channel function, *Current Opinion in Neurobiology 13*, 264-270.

76. Liu, D. R., Magliery, T. J., and Schultz, P. G. (1997) Characterization of an 'orthogonal' suppressor tRNA derived from E. coli tRNA2(Gln), *Chem Biol 4*, 685-691.

77. Liu, D. R., Magliery, T. J., Pastrnak, M., and Schultz, P. G. (1997) Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo, *Proc Natl Acad Sci U S A 94*, 10092-10097.

78. Liu, D. R., and Schultz, P. G. (1999) Progress toward the evolution of an organism with an expanded genetic code, *Proc Natl Acad Sci U S A 96*, 4780-4785.

79. Kwok, Y., and Wong, J. T. F. (1980) Evolutionary relationship between halobacterium-cutirubrum and eukaryotes determined by use of aminoacyl transfer RNA-synthetases as phylogenetic probes, *Canadian Journal of Biochemistry 58*, 213-218.

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

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

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

83. Schreiber, G., Frisch, C., and Fersht, A. R. (1997) The role of Glu73 of barnase in catalysis and the binding of barstar, *J Mol Biol 270*, 111-122.

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

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

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

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

88. Phizicky, E. M., and Fields, S. (1995) Protein-Protein Interactions - Methods For Detection And Analysis, *Microbiological Reviews 59*, 94-123.

89. Fields, S., and Song, O. (1989) A novel genetic system to detect protein-protein interactions, *Nature 340*, 245-246.

90. Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T., Goldberg, D. S., Li, N., Martinez, M., Rual, J. F., Lamesch, P., Xu, L., Tewari, M., Wong, S. L., Zhang, L. V., Berriz, G. F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H. W., Elewa, A., Baumgartner, B., Rose, D. J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S. E., Saxton, W. M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K. C., Harper, J. W., Cusick, M. E., Roth, F. P., Hill, D. E., and Vidal, M. (2004) A map of the interactome network of the metazoan C. elegans, *Science 303*, 540-543.

91. Ghosh, I., Hamilton, A. D., and Regan, L. (2000) Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein, *J Am Chem Soc 122*, 5658-5659.

92. Paulmurugan, R., Umezawa, Y., and Gambhir, S. S. (2002) Noninvasive imaging of protein-protein interactions in living subjects by using reporter protein complementation and reconstitution strategies, *Proc Natl Acad Sci U S A 99*, 15608-15613.

93. Rossi, F., Charlton, C. A., and Blau, H. M. (1997) Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation, *Proc Natl Acad Sci U S A 94*, 8405-8410.

94. Yan, Y., and Marriott, G. (2003) Analysis of protein interactions using fluorescence technologies, *Curr Opin Chem Biol 7*, 635-640.

95. Friedhoff, P. (2005) Mapping protein-protein interactions by bioinformatics and cross-linking, *Anal Bioanal Chem 381*, 78-80.

96. Sinz, A. (2003) Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes, *J Mass Spectrom 38*, 1225-1237.

97. Burdine, L., Gillette, T. G., Lin, H. J., and Kodadek, T. (2004) Periodate-triggered cross-linking of DOPA-containing peptide-protein complexes, *J Am Chem Soc 126*, 11442-11443.

98. Dorman, G., and Prestwich, G. D. (1994) Benzophenone photophores in biochemistry, *Biochemistry 33*, 5661-5673.

99. Galardy, R. E., and Jamieson, J. D. (1977) Photoaffinity labeling of a peptide secretagogue receptor in the exocrine pancreas, *Mol Pharmacol 13*, 852-863.

100. Haynes, F. J., and Yip, C. C. (1985) Photoaffinity labelling of hepatic plasma membranes suggests two classes of hepatic insulin receptor, *Diabetologia 28*, 786-792.

101. Yip, C. C., Laing, L. P., and Flynn, T. G. (1985) Photoaffinity labeling of atrial natriuretic factor receptors of rat kidney cortex plasma membranes, *J Biol Chem 260*, 8229-8232.

102. Kauer, J. C., Erickson-Viitanen, S., Wolfe, H. R., Jr., and DeGrado, W. F. (1986) p-Benzoyl-L-phenylalanine, a new photoreactive amino acid. Photolabeling of calmodulin with a synthetic calmodulin-binding peptide, *J Biol Chem 261*, 10695-10700.

103. Lawrence, M. S., Phillips, K. J., and Liu, D. R. (2007) Supercharging proteins can impart unusual resilience, *J Am Chem Soc 129*, 10110-10112.

104. Wilkins, B. J., Daggett, K. A., and Cropp, T. A. (2008) Peptide mass fingerprinting using isotopically encoded photo-crosslinking amino acids, *Mol Biosyst 4*, 934-936.

105. Lamos, S. M., Krusemark, C. J., McGee, C. J., Scalf, M., Smith, L. M., and Belshaw, P. J. (2006) Mixed isotope photoaffinity reagents for identification of small-molecule targets by mass spectrometry, *Angew Chem Int Ed Engl 45*, 4329-4333.

106. Maru, Y., Afar, D. E., Witte, O. N., and Shibuya, M. (1996) The dimerization property of glutathione S-transferase partially reactivates Bcr-Abl lacking the oligomerization domain, *J Biol Chem 271*, 15353-15357.

107. McTigue, M. A., Williams, D. R., and Tainer, J. A. (1995) Crystal structures of a schistosomal drug and vaccine target: glutathione S-transferase from Schistosoma japonica and its complex with the leading antischistosomal drug praziquantel, *J Mol Biol 246*, 21-27.

108. Schlieker, C., Weibezahn, J., Patzelt, H., Tessarz, P., Strub, C., Zeth, K., Erbse, A., Schneider-Mergener, J., Chin, J. W., Schultz, P. G., Bukau, B., and Mogk, A.

(2004) Substrate recognition by the AAA+ chaperone ClpB, *Nat Struct Mol Biol 11*, 607-615.

109. Chen, H. T., Warfield, L., and Hahn, S. (2007) The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex, *Nat Struct Mol Biol 14*, 696-703.

110. Heim, R., Cubitt, A. B., and Tsien, R. Y. (1995) Improved green fluorescence, *Nature 373*, 663-664.

111. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) Crystal structure of the Aequorea victoria green fluorescent protein, *Science 273*, 1392-1395.

112. Yang, F., Moss, L. G., and Phillips, G. N. (1996) The molecular structure of green fluorescent protein, *Nat Biotechnol 14*, 1246-1251.

113. Tsien, R. Y. (1998) The green fluorescent protein, *Annu Rev Biochem 67*, 509-544.

114. Patterson, G. H., and Lippincott-Schwartz, J. (2002) A photoactivatable GFP for selective photolabeling of proteins and cells, *Science 297*, 1873-1877.

115. Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H., and Miyawaki, A. (2002) An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein, *Proc Natl Acad Sci U S A 99*, 12651-12656.

116. Ando, R., Mizuno, H., and Miyawaki, A. (2004) Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting, *Science 306*, 1370-1373.

117. Manley, S., Gillette, J. M., Patterson, G. H., Shroff, H., Hess, H. F., Betzig, E., and Lippincott-Schwartz, J. (2008) High-density mapping of single-molecule trajectories with photoactivated localization microscopy, *Nat Methods 5*, 155-157.

118. Adams, S. R., Campbell, R. E., Gross, L. A., Martin, B. R., Walkup, G. K., Yao, Y., Llopis, J., and Tsien, R. Y. (2002) New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications, *J Am Chem Soc 124*, 6063-6076.

119. Griffin, B. A., Adams, S. R., and Tsien, R. Y. (1998) Specific covalent labeling of recombinant protein molecules inside live cells, *Science 281*, 269-272.

120. Thompson, R. H., and Whittaker, V. P. (1947) Antidotal activity of British anti-lewisite against compounds of antimony, gold and mercury, *Biochem J 41*, 342-346.

121. Andresen, M., Schmitz-Salue, R., and Jakobs, S. (2004) Short tetracysteine tags to beta-tubulin demonstrate the significance of small labels for live cell imaging, *mol biol cell 15*, 5616-5622.

122. Enninga, J., Mounier, J., Sansonetti, P., and Van Nhieu, G. T. (2005) Secretion of type III effectors into host cells in real time, *Nat Methods 2*, 959-965.

123. Tour, O., Adams, S. R., Kerr, R. A., Meijer, R. M., Sejnowski, T. J., Tsien, R. W., and Tsien, R. Y. (2007) Calcium Green FlAsH as a genetically targeted small-molecule calcium indicator, *Nat Chem Biol 3*, 423-431.

124. Nakanishi, J., Takarada, T., Yunoki, S., Kikuchi, Y., and Maeda, M. (2006) FRET-based monitoring of conformational change of the beta(2) adrenergic receptor in living cells, *Biochem Biophys Res Commun 343*, 1191-1196.

125. Hoffmann, C., Gaietta, G., Bunemann, M., Adams, S. R., Oberdorff-Maass, S., Behr, B., Vilardaga, J. P., Tsien, R. Y., Eisman, M. H., and Lohse, M. J. (2005) A FlAsH-based FRET approach to determine G protein - coupled receptor activation in living cells, *Nat Methods 2*, 171-176.

126. Luedtke, N. W., Dexter, R. J., Fried, D. B., and Schepartz, A. (2007) Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH, *Nat Chem Biol 3*, 779-784.

127. Adams, S. R., and Tsien, R. Y. (2008) Preparation of the membrane-permeant biarsenicals FlAsH-EDT2 and ReAsH-EDT2 for fluorescent labeling of tetracysteine-tagged proteins, *Nat Protoc 3*, 1527-1534.

128. Zhang, X. Y., and Bishop, A. C. (2007) Site-specific incorporation of allosteric-inhibition sites in a protein tyrosine phosphatase, *J Am Chem Soc 129*, 3812-3813.

129. Danielson, M. A., and Falke, J. J. (1996) Use of 19F NMR to probe protein structure and conformational changes, *Annu Rev Biophys Biomol Struct 25*, 163-195.

130. Yoo, T. H., Link, A. J., and Tirrell, D. A. (2007) Evolution of a fluorinated green fluorescent protein, *Proc Natl Acad Sci U S A 104*, 13887-13890.

131. Brooks, B., Phillips, R. S., and Benisek, W. F. (1998) High-efficiency incorporation in vivo of tyrosine analogues with altered hydroxyl acidity in place of the catalytic tyrosine-14 of Delta 5-3-ketosteroid isomerase of Comamonas (Pseudomonas) testosteroni: effects of the modifications on isomerase kinetics, *Biochemistry 37*, 9738-9742.

132. Seyedsayamdost, M. R., Yee, C. S., and Stubbe, J. (2007) Site-specific incorporation of fluorotyrosines into the R2 subunit of E. coli ribonucleotide reductase by expressed protein ligation, *Nat Protoc 2*, 1225-1235.

133. Pal, P. P., Bae, J. H., Azim, M. K., Hess, P., Friedrich, R., Huber, R., Moroder, L., and Budisa, N. (2005) Structural and spectral response of Aequorea victoria green fluorescent proteins to chromophore fluorination, *Biochemistry 44*, 3663-3672.

134. Thorson, J. S., Shin, I., Chapman, E., Stenberg, G., Mannervik, B., and Schultz, P. G. (1998) Analysis of the role of the active site tyrosine in human glutathione transferase A1-1 by unnatural amino acid mutagenesis., *J Am Chem Soc 120*, 451-452.

135. Cropp, T. A., and Schultz, P. G. (2004) An expanding genetic code, *Trends Genet 20*, 625-630.

136. Lee, H. M., Larson, D. R., and Lawrence, D. S. (2009) Illuminating the Chemistry of Life: Design, Synthesis, and Applications of "Caged" and Related Photoresponsive Compounds, *ACS Chem Biol*.

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

138. Ryu, Y., and Schultz, P. G. (2006) Efficient incorporation of unnatural amino acids into proteins in Escherichia coli, *Nat Methods 3*, 263-265.

139. Nagasawa, T., Utagawa, T., Goto, J., Kim, C. J., Tani, Y., Kumagai, H., and Yamada, H. (1981) Syntheses of L-tyrosine-related amino acids by tyrosine phenol-lyase of Citrobacter intermedius, *Eur J Biochem 117*, 33-40.

140. Xie, J., Liu, W., and Schultz, P. G. (2007) A genetically encoded bidentate, metal-binding amino acid, *Angew Chem Int Ed Engl 46*, 9239-9242.

141. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein, *Nat Biotechnol 24*, 79-88.

142. Andrews, B. T., Schoenfish, A. R., Roy, M., Waldo, G., and Jennings, P. A. (2007) The rough energy landscape of superfolder GFP is linked to the chromophore, *J Mol Biol 373*, 476-490.

143. Penner, R. M., Roth, N. J., Rob, B., Lay, H., and Huber, R. E. (1999) Tyr-503 of beta-galactosidase (Escherichia coli) plays an important role in degalactosylation, *Biochem Cell Biol 77*, 229-236.

144. Juers, D. H., Heightman, T. D., Vasella, A., McCarter, J. D., Mackenzie, L., Withers, S. G., and Matthews, B. W. (2001) A structural view of the action of Escherichia coli (lacZ) beta-galactosidase, *Biochemistry 40*, 14781-14794.

145. Paulus, H. (2000) Protein splicing and related forms of protein autoprocessing, *Annu Rev Biochem 69*, 447-496.

146. Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F. B., and Xu, M. Q. (1996) Protein splicing involving the Saccharomyces cerevisiae VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system, *J Biol Chem 271*, 22159-22168.

147. Xu, M. Q., and Perler, F. B. (1996) The mechanism of protein splicing and its modulation by mutation, *Embo J 15*, 5146-5153.