SELECTIVE FUNCTIONALIZATION OF THE PROTEIN N-TERMINUS WITH N-MYRISTOYL TRANSFERASE IN BACTERIA

Thesis by

Chethana Kulkarni

In Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2013

(Defended October 5, 2012)

ACKNOWLEDGEMENTS

Every Ph.D. thesis is the product of the efforts of many individuals, and my thesis is no exception. Upon reaching this point in my education and my career, I feel very fortunate to have benefited from the knowledge, inspiration, support, guidance, and good will of many people at different stages of my life. I found that the most straightforward manner in which to organize my thoughts, and this list, was chronologically.

First and foremost, I am greatly indebted to my family for their incredible sacrifices and unwavering support in all aspects of my education. My parents left behind everything and everyone they knew in India to make a better life for themselves and their daughters here in the United States. I could never imagine undertaking such a move myself, and I can never thank them enough for the endless love and generosity they have shown to my sister and me. For my entire life, my sister has been my most vigorous cheerleader. Her fiery spirit, sense of humor, strong will, and big heart inspire me.

I am grateful for the many wonderful teachers and mentors I had before coming to Caltech. In high school, my AP Chemistry teacher, Mr. James Flores, applied an unparalleled love of chemistry, commitment to teaching, and general rigor to every aspect of our coursework. I would not have majored in chemistry in college had it not been for Mr. Flores' influence. At Stanford University, I benefited immensely from the seemingly infinite knowledge of Dan Fandrick and Rajiv Dhawan. Both were very generous with their time as I learned the ropes as an undergrad in the Trost Lab, and they always maintained an upbeat attitude about research. Jason Woertink was also kind and helpful, both as a TA and as a research mentor in the Trost Lab. For my college years, I must also

thank Thierry Glauser, my fantastic internship supervisor at Guidant; his knowledge of polymers and stents and his humorous observations on life helped make for an excellent summer. Finally, during my senior year at Stanford, I had the good fortune of taking a course taught by Tom Wandless as well as completing research in his group. Without Tom's support, I would not have been able to pursue graduate research at Caltech, and I would not have discovered my keen interest in chemical biology.

Of course, the person who has had the largest impact on my thesis is my adviser, David Tirrell, and I am forever grateful that I had the opportunity to be part of his research group. The humility, patience, and humor which Dave exhibits are refreshing, especially at his level of professional success. I appreciate the freedom with which I was allowed to learn, develop, and pursue research in Dave's lab. Furthermore, his approach to mentorship and management have certainly had an impact on me as I begin thinking about my own career. I have no doubt that there are many aspects of Dave's role in my education that I have taken for granted and that will become more apparent as I progress through different stages of my career.

I am also so thankful for the support of my thesis committee members: Professors Peter Dervan, Julie Kornfield, and Sarah Reisman. The three of them have quite different research specialties, experiences, and personalities, and I learned a great deal from each of them. I appreciate their time and their efforts toward helping me grow as a scientist and as a person.

Next, I would like to express my gratitude to the members of the Tirrell Group who helped and supported me over the years. In particular, I am indebted to Rebecca Connor and John Ngo for teaching me nearly everything I know about bacterial protein

expression. Their patience, intelligence, and creativity were an inspiration to me. I greatly appreciate the scientific input and friendship of Stacey Maskarinec and Maren Buck, more than I can express in words. Conversations with Tae Hyeon Yoo, Jim Van Deventer, Julie Champion, and Janek Szychowski were helpful in shaping the direction of some of my thesis projects. Additionally, Frank Truong, Beverly Lu, Alborz Mahdavi, JD Bagert, Kai Yuet, and Brett Babin also provided scientific advice and ideas. In my last two years at Caltech, I had the pleasure of collaborating with Tamara Kinzer-Ursem, from whom I have learned a great deal about science and much more. Tami's infinite passion for research, even (and especially) when things are rough, is admirable. Thanks to Tami, I also had the chance to work alongside the scientists at Maven Biotechnologies and learn more about life and research at a small company. Finally, the summer of 2011 was particularly productive and fun due in part to the hard work and bright personality of Megan Lo, a talented SURF student whom I had the opportunity to mentor.

We have a terrific set of staff members here at Caltech, especially in the CCE Division, and I will miss many of them. The various administrators with whom I interacted—Anne Hormann, Agnes Tong, Karen Baumgartner, Laura Howe, Dian Buchness, and Laura Lutz King—put a superhuman amount of effort into keeping the labs, departments, and division running. I also want to thank Steve Gould, Leah Mentch, and Joe Drew for submitting and delivering my orders for the past six years; I really don't know how they keep track of all of our purchases and remember our names too! Lastly, Cora Carreido and Pat Perrone always did such a wonderful job keeping our stockrooms in shape and our experiments running. All of these folks not only do their jobs well, but

they do them with a smile, and I hope they know how much the students and post-docs appreciate their work.

The scientific facilities staff at Caltech were also instrumental to the research presented in my thesis. The mass spectrometry results critical to some of my projects were obtained with the aid of Mona Shahgoli, Robert Graham, and Jie Zhou. I greatly appreciate David VanderVelde's tireless efforts in maintaining and improving the NMR facilities, which were indispensable for characterizing many of the compounds which I worked with. Finally, Thai Truong and the staff at the Beckman Imaging Center were also helpful over the years.

The process of completing a Ph.D. thesis is made considerably more enjoyable by the support (and commiseration) of friends. I'd like to thank a number of other chemistry graduate students in the incoming class of 2006—Young In Oh, Jean Li, Pam Sontz, Natalie Muren, Hang Song, Nyssa Puskar, Heather Williamson, Charlotte Mason Whited, Narae Park, Chris Daeffler, Ian Tonks, Jacob Bitterman, Justin Chartron, and Patrick Theofanis—for countless lunch breaks, girls' nights, and problem set sessions, especially during our first few years. It was always nice to run into my classmates around campus in later years, too. In addition, I appreciate the camaraderie of so many other students and post-docs in the chemistry department; I'd especially like to acknowledge the Barton Group, who were very welcoming to me over the years, and many members of the Stoltz Group, particularly Jenny Roizen, who was an excellent TA and a dear friend.

I am also extremely grateful for the support, laughs, and perspective offered by my friends outside of Caltech. Conversations with Andrea Burbank, Julie Kim, David Lin, Grace Hsu, and Michelle Bailey Sidwell were integral to my maintaining my sanity

while at Caltech, especially at tough times. As graduate work can become all consuming, it was good and important to be reminded of the world beyond my lab bench, Spalding, and Caltech.

Another important factor in my maintaining some balance during graduate school was the set of student activities in which I participated. The Women Mentoring Women program was a fantastic outlet, offering opportunities to meet other women on campus and learn about navigating the scientific world. A huge thanks goes out to Portia Harris for the time and effort she pours into WMW and other student programming. I enjoyed serving as a mentor, mentee, and advisory board member for WMW over the past six years, and I'm excited to see what the future holds for the program. As I work with future generations of female scientists, I hope to implement what I learned from WMW. I also have no doubt that the closest friends I made through WMW—Kathleen Spencer and Kim Petersen—will continue to be a part of my life for many years to come. Finally, I loved playing flute in the Caltech Concert Band during my first few years of graduate school, and I will always cherish the memories of our magical trip to Carnegie Hall in 2008. It is thanks to William Bing that all of us in band were able to perform at that magnificent venue, and more importantly, were able to unwind and have fun while making music every week here in Pasadena.

I do not exaggerate when I state that I would not be receiving a Ph.D. were it not for my roommate and close friend, Chithra Krishnamurthy. She is like family to me now; no doubt, the most difficult part of transitioning to my post-Caltech life will be her physical absence, though it will be offset slightly by the certainty that we will be lifelong friends. I am so grateful for all the conversations we had about life and research over

meals in our apartment, for all that I learned from her, and for all the fun we had exploring LA when we finally did venture beyond lab and our apartment. She is truly the strongest, wisest, and most graceful person I have ever met. I also want to thank my honorary roommate, Keith Keitz, for his calm demeanor, sense of humor, and willingness to go along on the crazy exploits that Chithra and I planned for the four of us.

And this brings me to my best friend, my love, my husband, Eric Olmon. His infinite patience and steady support served as a solid foundation for me to lay my worries and troubles on every day, and his love for me is apparent in his every word and action. I am so, so thankful for all that he did for us during our time at Caltech, and I am super excited for what great adventures the future holds for us!

ABSTRACT

Proteins are involved in myriad processes in all organisms. They provide structural support in the membrane and scaffolding of each cell; they aid in the transmission of biochemical signals within and between cells; and they play central roles in combating various disease states. The development of techniques enabling selective and site-specific functionalization of proteins is an active area of investigation, as such modifications are critical to many studies and uses of proteins. For instance, with the addition of a unique reactive handle, a protein may be conjugated to a polymer for the production of protein-based therapeutics exhibiting improved bioavailability. Alternatively, proteins may be attached to slides to prepare diagnostic microarrays, reacted with hydrogels to create functional biomaterials, or decorated with fluorophores for *in vivo* imaging. Site-specific protein tagging techniques have already contributed greatly to biomedical research and will continue to advance the state of the field.

The focus of my thesis research has been the development of a novel site-specific protein labeling method centered on the eukaryotic enzyme N-myristoyl transferase (NMT). In a process called myristoylation, NMT appends a fatty acid to the N-terminus of numerous substrate proteins. Previous work demonstrated that NMT tolerates a wide range of chemically functionalized analogs of its natural fatty acid substrate. Here, we describe efforts that exploit various features of NMT: its ability to bind and utilize reactive fatty acid analogs, its exquisite selectivity toward its protein substrates, and its orthogonality toward those proteins naturally present in bacteria.

First, in Chapter II, we discuss the development of a model system for NMT-mediated protein labeling in the bacterium *Escherichia coli*. We synthesized an azide fatty acid analog that can participate in bioorthogonal chemistries, and we prepared two GFP-based substrate proteins, each displaying a recognition sequence derived from a known substrate of NMT. Our experiments indicate that labeling by NMT is site-specific, quantitative, and highly selective for each engineered substrate within the bacterial milieu.

As summarized in Chapter III, the model system was extended to the N-terminal labeling of two neuronal proteins, calcineurin (CaN) and calmodulin (CaM). While CaN is naturally myristoylated, CaM was engineered to achieve labeling by NMT. Experiments with CaN and CaM confirmed that our NMT-based system is quantitative and selective in its labeling of both natural and engineered substrate proteins. Extensive characterization of each protein allowed us to identify constructs that retain wild-type levels of activity even after labeling with the azide fatty acid.

Three of the protein constructs reported in Chapters II and III were utilized for microarray studies, as described in Chapter IV. We achieved rapid surface immobilization of each azide-labeled protein directly from lysate, a significant advantage when considering the time and resources normally required to purify proteins for downstream applications. The experiments and methods summarized in this chapter will be adapted for high-throughput biochemical research with protein microarrays.

Finally, the orthogonality of NMT toward bacterial systems was probed further for the purpose of selective labeling of individual bacterial proteins for live-cell imaging. In addition to identifying an azide fatty acid suitable for such studies, we also selected

two bacterial proteins that exhibit interesting functions and localization patterns, and we developed corresponding protein constructs for NMT-mediated labeling. Progress toward the use of NMT for *in vivo* imaging applications in bacteria is described in Chapter V.

Ultimately, our objective throughout the design and execution of these projects was to create and validate a new technique to achieve site-specific protein labeling. The particular advantages of NMT include its tolerance of reactive fatty acid analogs and engineered substrate proteins, and its lack of interaction with proteins present in the widely used *E. coli* expression host. We believe that the ideas and results presented in this thesis establish NMT-mediated protein labeling as a valuable tool for addition to the existing set of site-specific protein labeling methods. Development of such methods represents an important and exciting area within the field of modern chemical biology.

TABLE OF CONTENTS

LIST OF FIGURES, SCHEMES, CHARTS, AND TABLES

CHAPTER I

Protein Engineering:

Techniques and Applications

ABSTRACT

Proteins are an important class of cellular biomolecules, playing diverse roles in the structure, motility, communication, defense, division, and destruction of cells. They are found in every cell of every organism, from bacteria to humans. The incredible range of protein structures and functions is all the more remarkable considering that all proteins are built from the same simple canon of twenty amino acids. To study proteins, both in their natural cellular context as well as within *in vitro* systems, scientists have long taken advantage of the limited chemical functionalities present in amino acid side chains to covalently attach proteins to other molecules. Furthermore, protein engineering techniques have expanded the range of chemical moieties that may be incorporated into proteins. Such techniques enable the differentiation of one subset of proteins from a larger population, and they allow for the selective conjugation of individual proteins to reactive partners ranging from polymers to fluorophores to microarray chips. Given the central role played by proteins in so many cellular processes, advances in protein engineering have important implications for biomedical research and human health.

This introductory chapter highlights some examples of protein engineering methods developed to date—specifically, the incorporation of non-canonical amino acids into proteins—as well as their applications. Because the focus of this thesis is the development and use of a chemoenzymatic approach to N-terminal protein labeling, other strategies enabling functionalization of the N-terminus are also discussed. Finally, we present an overview of the chapters comprising this thesis and summarize contributions to individual projects.

PROTEIN ENGINEERING

Incorporation of Non-Canonical Amino Acids into Proteins

Perhaps the most direct strategy for introducing non-natural functionalities into proteins is the incorporation of amino acid building blocks displaying such groups. Natural amino acids may be replaced by non-canonical analogs in a residue-specific or a site-specific manner. Hundreds of non-canonical amino acids (ncAAs) have now been utilized for protein engineering; a selection of these is displayed in Figure I-1.[1]

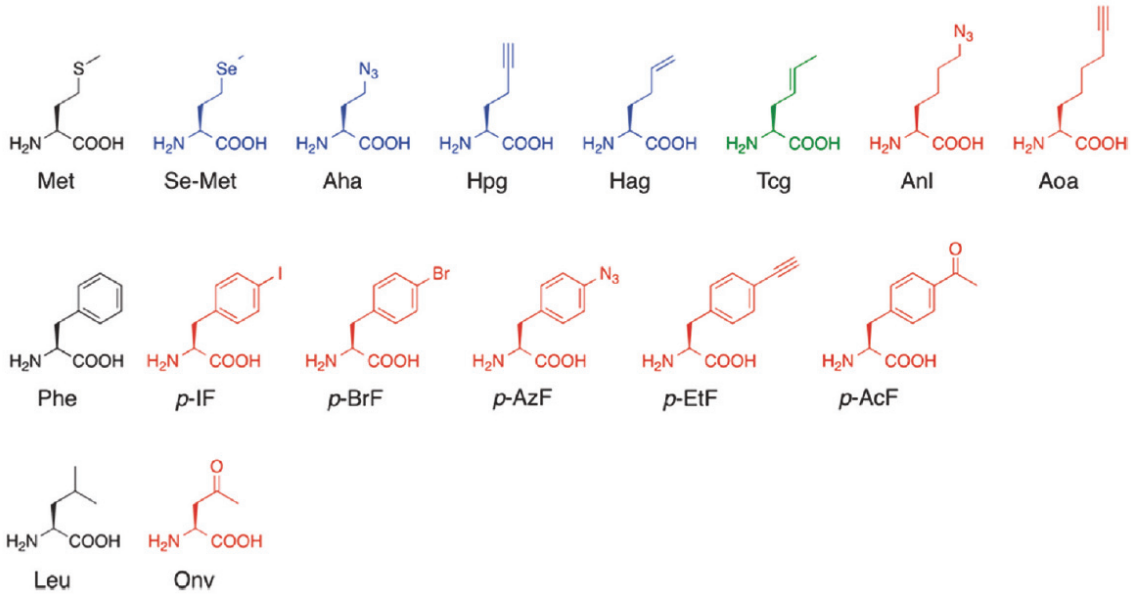


Figure I-1. Numerous non-canonical surrogates for the amino acids methionine (Met), phenylalanine (Phe), and leucine (Leu) have been developed. They may be bound by endogenous tRNA synthetases (those colored blue or green), or they may require the use of mutant tRNA synthetases (those colored red). Adapted from Reference 1.

Global replacement of an amino acid by a non-natural counterpart is useful for proteomics investigations or in other contexts requiring a snapshot in time of the entire protein population. In contrast, single-site incorporation of an artificial amino acid is

valuable for detailed structural studies of a particular protein or controlled conjugation of a protein to a reactive partner. Both methods are described here.

Residue-specific incorporation of ncAAs generally requires no manipulation of the genetic material (DNA or mRNA) corresponding to the proteins that are to be synthesized. Instead, as the ribosomal complex translates an mRNA strand into a protein, the tRNAs corresponding to a given codon deliver a pre-determined ncAA in place of the natural amino acid encoded by that codon. The aminoacylation of tRNAs with ncAAs may be accomplished by endogenous tRNA synthetases or may require the use of engineered tRNA synthetases, as noted in Figure I-1. Experimental parameters may be modulated to tune the timing and extent of ncAA incorporation into proteins.[1–3]

The Tirrell Lab, the Schuman Lab, and others have contributed greatly to the development and application of global amino acid replacement methods, particularly with the establishment of the BONCAT (bio-orthogonal non-canonical amino acid tagging) technique in 2006.[4] BONCAT has since been utilized for numerous proteomics studies, including examination of mixed-cell populations and zebrafish larvae, to name a few.[5–7] Imaging studies have also been described, wherein ncAAs were utilized in conjunction with reactive dyes to visualize proteins in bacteria and mammalian cells.[8,9] The reagents, methods, and outcomes associated with residue-specific incorporation of ncAAs have proven to be of great use and interest to both biologists and chemists.[10–12]

As a complementary technique to residue-specific ncAA incorporation, site-specific incorporation of a ncAA may be desirable when a single protein is under investigation. The ability to change one amino acid systematically has been integral to some structural studies, such as the Dougherty Lab's research on ion channels.[13] Adding

a single reactive ncAA to a protein has also allowed for controlled protein-polymer conjugation,[14] an important step in the preparation of some protein therapeutics.[15] Largely developed by the Schultz Lab, methods for site-specific incorporation of ncAAs generally require that the gene encoding the protein of interest be modified to display a Stop codon at the desired site for ncAA incorporation. In conjunction, suitable tRNAs that display the anticodon for the Stop codon must be chemically or enzymatically amino-acylated with a ncAA. When the ribosomal complex encounters the Stop codon during translation, the ncAA is incorporated at the corresponding site into the protein product.[16–18] Single-site replacement methods have been central to a number of biological studies during the past few decades,[19] with their use recently extended to live fruit flies.[20] However, these methods also have some limitations: in particular, they are not effective for replacement of the N-terminal residue, a transformation that has been achieved with the use of complementary approaches.

Modification of the Protein N-Terminus

The protein N- and C-termini are attractive targets for the conjugation of proteins to other substrates, including polymers, beads, slides, and fluorophores, because even in the folded state, the termini of many proteins are surface-accessible.[21] Early attempts at selective modification of the N-terminus took advantage of the slight difference in pK_a between the N-terminal amine and lysine amines, though success was generally limited to cases in which few lysine side chains presented competition.[22] More recently, the Francis Group has reported their use of chemical approaches to achieve transformation of the protein N-terminus.[23] They utilized their strategy to label antibodies, which retained their

ability to bind targets even after labeling.[24] In contrast, the Tirrell Group described a biological approach to N-terminal labeling: the L,F-transferase enzyme was employed *in vitro* and in *E. coli* to add analogs of Leu, Phe, and Met to the N-terminus of peptides and proteins.[25] Other site-specific chemoenzymatic labeling techniques have been developed; these methods and others are evaluated in a recent review, which also details the suitability of different approaches for N-terminal protein modification.[26–30] For our work, we focused on the enzyme N-myristoyl transferase to achieve N-terminal protein functionalization.

N-MYRISTOYL TRANSFERASE

Structure and Function of N-Myristoyl Transferase

The eukaryotic enzyme N-myristoyl transferase (NMT) catalyzes the transfer of myristic acid, a 14-carbon unbranched saturated fatty acid, to the N-terminus of various substrate proteins. Members of many classes of proteins undergo this co-translational and irreversible transformation, which is called myristoylation (Figure I-2). Substrates include protein kinases A and G, subunits of heterotrimeric G proteins, multiple tyrosine kinases, phosphatases, and even viral capsid proteins.[31]

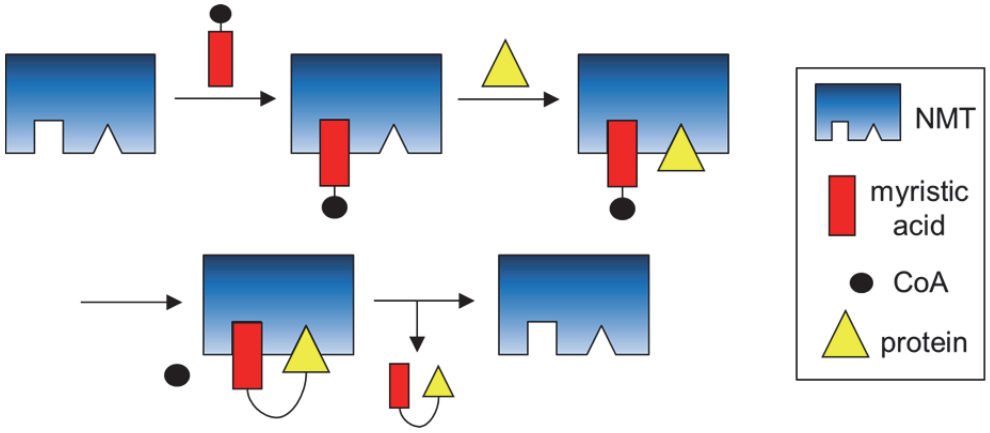


Figure I-2. Schematic overview of the process of myristoylation. NMT is colored blue, the fatty acid is colored red, and the substrate protein is colored yellow. NMT first binds myristoyl-CoA and only then binds a substrate protein. An amide linkage forms between the protein N-terminus and the fatty acid, the labeled protein is expelled, and NMT begins a new enzymatic cycle.

The hydrocarbon tail of myristic acid is thought to aid in the localization of some substrates to lipid membranes, underscoring the importance of myristoylation to the proper functioning of proteins that are involved in signal transduction and viral infectivity.[32] Endogenous expression of NMT has been confirmed in 15 different eukaryotic species, ranging from yeast to humans, and dozens of substrate proteins have been identified.[33,34] Notably, native expression of NMT has not been observed in *Escherichia coli*.[35]

NMT substrate proteins share certain features within their N-terminal sequence domain; these characteristics play an important role in proper recognition by NMT. The only absolute and universal constraint is that all NMT substrates must display a glycine (Gly, G) residue at the N-terminus; the initiating Met residue normally found in proteins must be removed by a separate enzyme, methionine amino-peptidase (Met-AP), prior to

myristoylation.[36] Previous research using site-directed mutagenesis demonstrated that the amino acids at positions 3, 6, and 7 in the substrate protein (where position 1 is the initial Met) play a particularly important role in NMT binding. More specifically, charged residues, aromatic residues, and proline are not permitted at position 3; serine or threonine is usually found at position 6; and positively charged residues are preferred at position 7.[33,37] The combination of amino acids at positions 3, 6, and 7 may influence species-specific recognition of substrate proteins by NMT.[37] The crystal structure of NMT, with both of its binding pockets highlighted, is depicted in Figure I-3.

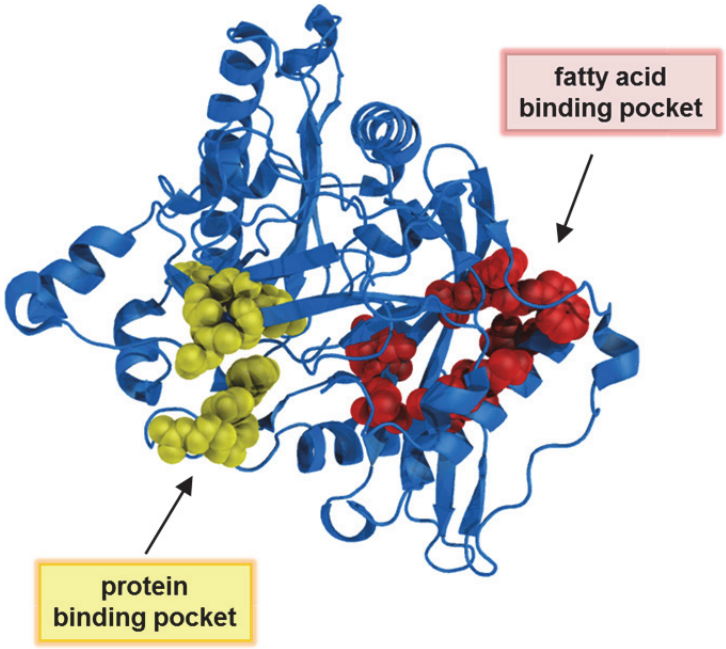


Figure I-3. Crystal structure of NMT, with the binding pocket for myristic acid colored red and the binding pocket for the substrate protein colored yellow. PDB ID: 1IID.

Substitution of Fatty Acid Analogs for Myristic Acid

At the outset of this work, two aspects of NMT were especially pertinent in drawing our attention to its possible use in a novel protein labeling method: its absence

from bacteria, and its tolerance toward reactive variants of myristic acid. The ability of NMT to transfer functionalized fatty acid analogs was tested extensively by the Gordon Lab using octapeptide substrates with *Saccharomyces cerevisiae* NMT in an *in vitro* system.[38–40] Of particular interest, analogs with useful chemical moieties, such as those seen in the ncAAs shown in Figure I-1, e.g., azide, alkyne, and ketone groups, were bound and transferred by NMT. These functional groups can participate in powerful bioorthogonal reactions that allow for selective attachment of proteins bearing these groups to appropriately derivatized partners.[41]

Building on the Gordon Lab's *in vitro* work completed 15 years prior, the Ploegh Lab in 2007 employed azide-functionalized fatty acids for selective labeling and detection of myristoylated proteins in a cancer cell line.[42] Azido and alkynyl fatty acids were also utilized in the past five years for N-terminal labeling by NMT in bacteria.[43,44] To the best of our knowledge, all NMT-based protein labeling studies completed to date with reactive analogs of myristic acid in cells have involved natural substrate proteins only, either for proteomic profiling projects or proof-of-principle purposes. The primary goal of the work described in this thesis is to expand the scope of NMT-mediated labeling to simultaneously utilize non-natural variants of both the fatty acid *and* protein substrates. We envisioned a system in which any protein of interest, engineered to display a short recognition sequence, could be co-expressed with NMT in the presence of a reactive fatty acid to achieve selective and site-specific functionalization.

THESIS ORGANIZATION

Overview of Thesis Chapters

The first task that we set out to accomplish was the development of a robust bacterial co-expression system to serve as a foundation for subsequent projects. Initial work was attempted with natural NMT substrates, but we quickly progressed to the use of engineered non-natural substrates. Chapter II details the development and evaluation of two GFP-based substrate proteins that were selectively and quantitatively labeled by NMT with an azido fatty acid. This is the first example of the use of a non-natural substrate protein and a fatty acid analog together for NMT-mediated protein labeling.

Chapter III describes an application of the system established in Chapter II. Building on our work with GFP, a common model protein, we undertook the task of labeling two proteins implicated in learning and memory: calcineurin (CaN) and calmodulin (CaM). Both proteins are under investigation by neuroscientists for their important roles in the brain. CaN is a natural substrate of NMT, but CaM is not. We engineered CaM to display different NMT recognition sequences and demonstrated selective and quantitative N-terminal labeling of both CaN and CaM-based substrates with an azido fatty acid. We also investigated the activity of the labeled proteins. CaN was shown to be equally active whether it was labeled with the azido fatty acid or with myristic acid. In addition, a CaM construct was identified that retained wild-type activity, even with the addition of an NMT recognition sequence and labeling with the azido fatty acid.

Chapter IV summarizes our work to date in immobilizing N-terminally functionalized proteins on surfaces, with the ultimate goal of studying protein activity in a high-throughput format. Preliminary experiments were completed with agarose beads, though for the majority of our work in this area, we utilized protein microarrays. Surface capture of N-terminally labeled GFP, CaN, and CaM-based constructs was achieved. Furthermore, the orthogonality of NMT in bacteria enabled selective coupling of these proteins directly out of lysate, with no prior purification.

Finally, Chapter V describes our progress toward the use of NMT-mediated protein labeling to study the localization patterns of individual proteins in bacteria. Recent work has demonstrated that bacterial cells exhibit far more control over the spatiotemporal organization of their proteins than was previously thought. Again, the absence of any natural NMT substrates from prokaryotes makes NMT a perfect candidate for the selective functionalization and visualization of appropriately engineered bacterial proteins of interest.

A schematic overview summarizes the projects described in Chapters II, III, IV, and V of this thesis (Figure I-4).

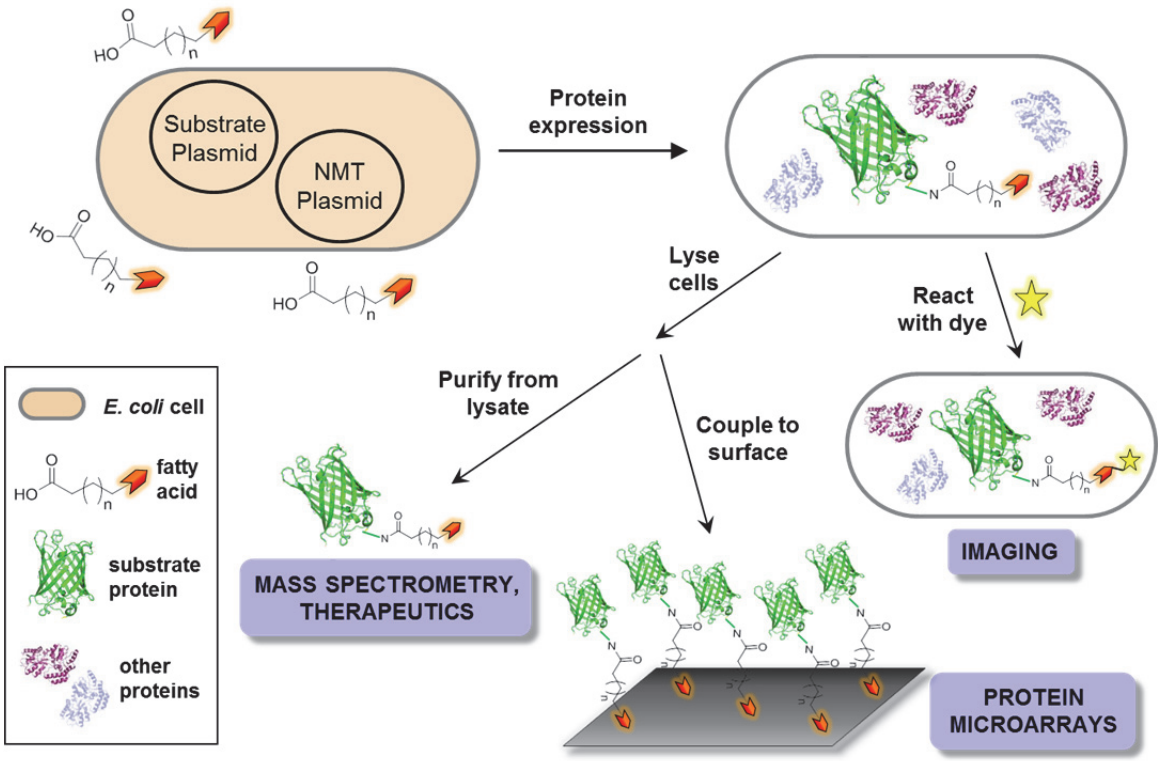


Figure I-4. Overview of thesis projects. NMT-mediated labeling of both natural and engineered substrate proteins, as well as various applications of N-terminal protein labeling, are described herein. Transformation of *E. coli* cells with the appropriate plasmids enables co-expression of NMT and a substrate protein. NMT appends certain fatty acid analogs to the N-terminus of substrate proteins. Cells may be lysed or harvested intact for *in vivo* imaging experiments. Mass spectrometry and therapeutic applications require purification of the substrate protein from lysate, while surface capture of the substrate protein may be performed from lysate due to the orthogonality of NMT in bacterial systems.

Thesis Contributions

For Chapter II, I conceived the design of the model system and completed all cloning, expression, purification, mass spectrometry, and fluorescence gel experiments. I also synthesized and characterized the azide fatty acid used in this chapter as well as in Chapters III and IV. I analyzed all data.

For Chapter III, I designed the CaM constructs and completed all cloning, expression, mass spectrometry, and fluorescence gel experiments; I analyzed the resultant data. My collaborator, Dr. Tamara Kinzer-Ursem, contributed intellectually throughout this project and experimentally to the purification of the CaN and CaM constructs. We conducted the activity assays and analyzed the results together. Megan Lo, an undergraduate student, also aided in the completion of experiments with CaN and CaM.

For Chapter IV, I completed preliminary experiments with agarose beads and the accompanying data analysis, while Dr. Kinzer-Ursem led our efforts in microarray preparation as well as protein microarray experiments and data analysis.

For Chapter V, I conceived the project design and completed all cloning and protein expression experiments. Professor Zemer Gitai of Princeton University provided input regarding which proteins to study. The library of azide fatty acids and the cyclooctyne-coumarin dye were prepared by a former post-doctoral researcher, Dr. Janek Szychowski. I completed all lysate and live-cell dye-labeling experiments with the azide fatty acids, and I analyzed all associated data.

REFERENCES

1. Ngo, J. T. & Tirrell, D. A. Noncanonical amino acids in the interrogation of cellular protein synthesis. *Acc. Chem. Res.* **44**, 677–85 (2011).

2. Link, A. J., Mock, M. L. & Tirrell, D. A. Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* **14**, 603–9 (2003).

3. Johnson, J. A., Lu, Y. Y., Van Deventer, J. A. & Tirrell, D. A. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Chem. Biol.* **14**, 774–80 (2010).

4. Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9482–7 (2006).

5. Ngo, J. T. *et al.* Cell-selective metabolic labeling of proteins. *Nat. Chem. Biol.* **5**, 715–7 (2009).

6. Eichelbaum, K., Winter, M., Diaz, M. B., Herzig, S. & Krijgsveld, J. Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat. Biotechnol.* **30**, 984–90 (2012).

7. Hinz, F. I., Dieterich, D. C., Tirrell, D. A. & Schuman, E. M. Non-canonical amino acid labeling in vivo to visualize and affinity purify newly synthesized proteins in larval zebrafish. *ACS Chem. Neurosci.* **3**, 40–9 (2012).

8. Beatty, K. E., Xie, F., Wang, Q. & Tirrell, D. A. Selective dye-labeling of newly synthesized proteins in bacterial cells. *J. Am. Chem. Soc.* **127**, 14150–1 (2005).

9. Beatty, K. E. *et al.* Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew. Chem. Int. Ed. Engl.* **45**, 7364–7 (2006).

10. Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161–4 (2010).

11. Song, W. *et al.* A metabolic alkene reporter for spatiotemporally controlled imaging of newly synthesized proteins in mammalian cells. *ACS Chem. Biol.* **5**, 875–885 (2010).

12. Kar, A. N., MacGibeny, M. A., Gervasi, N. M., Gioio, A. E. & Kaplan, B. B. Intra-axonal synthesis of eukaryotic translation initiation factors regulates local protein

synthesis and axon growth in rat sympathetic neurons. *J. Neurosci.* **33**, 7165–7174 (2013).

13. Dougherty, D. A. Physical organic chemistry on the brain. *J. Org. Chem.* **73**, 3667–73 (2008).

14. Deiters, A., Cropp, T. A., Summerer, D., Mukherji, M. & Schultz, P. G. Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorg. Med. Chem. Lett.* **14**, 5743–5 (2004).

15. Harris, J. M. & Chess, R. B. Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discovery* **2**, 214–21 (2003).

16. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–8 (1989).

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

18. Chin, J. W. *et al.* An expanded eukaryotic genetic code. *Science* **301**, 964–7 (2003).

19. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–44 (2010).

20. Bianco, A., Townsley, F. M., Greiss, S., Lang, K. & Chin, J. W. Expanding the genetic code of Drosophila melanogaster. *Nat. Chem. Biol.* **8**, 748–50 (2012).

21. Jacob, E. & Unger, R. A tale of two tails: why are terminal residues of proteins exposed? *Bioinformatics* **23**, e225–30 (2006).

22. Dixon, H. N-Terminal Modification of Proteins-A Review. *J. Protein Chem.* **3**, 99–108 (1984).

23. Gilmore, J. M., Scheck, R. A., Esser-Kahn, A. P., Joshi, N. S. & Francis, M. B. N-terminal protein modification through a biomimetic transamination reaction. *Angew. Chem. Int. Ed. Engl.* **118**, 5433–7 (2006).

24. Scheck, R. A. & Francis, M. B. Regioselective labeling of antibodies through N-terminal transamination. *ACS Chem. Biol.* **2**, 247–51 (2007).

25. Connor, R. E., Piatkov, K., Varshavsky, A. & Tirrell, D. A. Enzymatic N-terminal addition of noncanonical amino acids to peptides and proteins. *ChemBioChem* **9**, 366–9 (2008).

26. Fernández-Suárez, M. *et al.* Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* **25**, 1483–7 (2007).

27. Duckworth, B. P., Zhang, Z., Hosokawa, A. & Distefano, M. D. Selective labeling of proteins by using protein farnesyltransferase. *ChemBioChem* **8**, 98–105 (2007).

28. Carrico, I. S., Carlson, B. L. & Bertozzi, C. R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **3**, 321–2 (2007).

29. Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E. & Ploegh, H. L. Sortagging: a versatile method for protein labeling. *Nat. Chem. Biol.* **3**, 707–8 (2007).

30. Rabuka, D. Chemoenzymatic methods for site-specific protein modification. *Curr. Opin. Chem. Biol.* **14**, 790–6 (2010).

31. Boutin, J. A. Myristoylation. *Cell Signal* **9**, 15–35 (1997).

32. Knoll, L. J., Russell Johnson, D., Bryant, M. L. & Gordon, J. I. Functional significance of myristoyl moiety in N-myristoyl proteins. *Methods Enzymol.* **250**, 405–35 (1995).

33. Farazi, T. A., Waksman, G. & Gordon, J. I. Structures of Saccharomyces cerevisiae N-myristoyltransferase with bound myristoylCoA and peptide provide insights about substrate recognition and catalysis. *Biochemistry* **40**, 6335–43 (2001).

34. Farazi, T. A., Manchester, J. K., Waksman, G. & Gordon, J. I. Pre-steady-state kinetic studies of Saccharomyces cerevisiae myristoylCoA:protein N-myristoyltransferase mutants identify residues involved in catalysis. *Biochemistry* **40**, 9177–86 (2001).

35. Duronio, R. J. *et al.* Protein N-myristoylation in Escherichia coli: reconstitution of a eukaryotic protein modification in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 1506–10 (1990).

36. Van Valkenburgh, H. A. & Kahn, R. A. Coexpression of proteins with methionine aminopeptidase and/or N-myristoyltransferase in Escherichia coli to increase acylation and homogeneity of protein preparations. *Methods Enzymol.* **344**, 186–93 (2002).

37. Utsumi, T. *et al.* Vertical-scanning mutagenesis of amino acids in a model N-myristoylation motif reveals the major amino-terminal sequence requirements for protein N-myristoylation. *Eur. J. Biochem.* **271**, 863–74 (2004).

38. Kishore, N. S. *et al.* The substrate specificity of Saccharomyces cerevisiae myristoyl-CoA: protein N-myristoyltransferase. *J. Biol. Chem.* **266**, 8835–55 (1991).

39. Devadas, B. *et al.* Substrate specificity of Saccharomyces cerevisiae myristoyl-CoA: protein N-myristoyltransferase. *J. Biol. Chem.* **267**, 7224–39 (1992).

40. Lu, T. *et al.* The substrate specificity of Saccharomyces cerevisiae myristoyl-CoA: protein N-myristoyltransferase. Polar probes of the enzyme's myristoyl-CoA recognition site. *J. Biol. Chem.* **269**, 5346–57 (1994).

41. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed. Engl.* **48**, 6974–98 (2009).

42. Hang, H. C. *et al.* Chemical probes for the rapid detection of fatty-acylated proteins in mammalian cells. *J. Am. Chem. Soc.* **129**, 2744–5 (2007).

43. Heal, W. P. *et al.* Site-specific N-terminal labelling of proteins in vitro and in vivo using N-myristoyl transferase and bioorthogonal ligation chemistry. *Chem. Commun.* **3**, 480–2 (2008).

44. Heal, W. P., Wickramasinghe, S. R., Leatherbarrow, R. J. & Tate, E. W. N-myristoyl transferase-mediated protein labelling in vivo. *Org. Biomol. Chem.* **6**, 2308–15 (2008).

CHAPTER II

Development and Evaluation of a Bacterial Model System for N-Terminal Protein Labeling with N-Myristoyl Transferase

ABSTRACT

This chapter describes the design, development, and evaluation of a model system for NMT-mediated protein labeling. First, we assembled the necessary components; their selection and design are described in the Introduction. Green fluorescent protein (GFP) was chosen to be the test protein, owing to its widespread use and its facile detection by fluorescence measurements as well as conventional protein characterization methods.

We prepared two different engineered GFP substrates, each displaying a different NMT recognition sequence, and achieved selective modification of each protein via co-expression with NMT in the presence of an azido fatty acid. The labeled GFP substrates were purified, digested with a protease, and analyzed by mass spectrometry; identification of the N-terminal peptide fragment confirmed that site-specific labeling had occurred. Whole-protein mass spectrometry was also performed to determine the extent of modification. Finally, lysate samples containing a labeled GFP substrate were treated with a reactive dye and examined by SDS-PAGE. Fluorescence detection indicated that NMT labeled only the engineered GFP substrate *in vivo*. Together, these results demonstrate that our NMT-mediated protein labeling system is site-specific, quantitative, and highly selective. These results also constitute the first example of NMT-mediated labeling of a non-natural substrate protein with a functionalized, non-natural fatty acid, paving the way for future projects with other proteins of interest and other fatty acids analogs.

INTRODUCTION

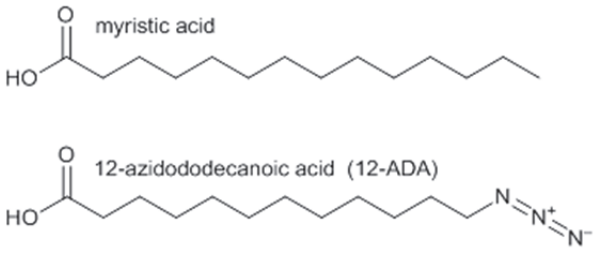
The three components comprising our NMT-mediated protein labeling model system are a reactive fatty acid, NMT, and a substrate protein. The design and development of each component are described below.

Selection of 12-Azidododecanoic Acid (12-ADA)

The natural fatty acid substrate of NMT, myristic acid (Chart II-1), is chemically inert. Fortunately, as described in Chapter I, dozens of reactive fatty acid analogs have been shown to be well-tolerated by NMT in an *in vitro* system.[1–3] Ideally, the analog employed in our protein labeling system would be readily synthesized and purified, permeable to the bacterial cell wall and membrane, bound and transferred by NMT in high yields, and able to participate in bioorthogonal chemistries. The compound 12-azidododecanoic acid (abbreviated 12-ADA henceforth; Chart II-1) was investigated with these criteria in mind.

Previous work indicated that 12-ADA was transferred to octapeptide substrates of NMT in higher yields than 9-, 11-, or 13-carbon analogs.[2] More recent work in mammalian cells yielded similar results, with 12-ADA resulting in more robust protein labeling than closely related azido fatty acids.[4] As described below, we synthesized and purified 12-ADA in high yields from adapted literature protocols. We also attempted the synthesis of an alkynyl compound, 13-tetradecynoic acid, but the reaction yields were low and the product was difficult to purify. Thus, for further work, we focused on the use of 12-ADA.

Chart II-1. Structures of myristic acid, the natural fatty acid substrate of NMT, and 12-azidodecanoic acid (12-ADA), the fatty acid analog prepared for and utilized in our studies.



The azide group can participate in three different bioorthogonal reactions: the Staudinger ligation, the copper-catalyzed azide-alkyne cycloaddition (CuAAC), and the strain-promoted azide-alkyne cycloaddition (SPAAC).[5–10] All three of these reactions may be performed in the challenging environment of cell lysate with exquisite selectivity. The Staudinger ligation and SPAAC can also be conducted in live cells and organisms. By selecting 12-ADA for our protein labeling system, we were well-positioned to explore these chemistries with N-terminally functionalized proteins.

Two-Plasmid Bacterial Co-expression System

Co-expression of NMT and a substrate protein, whether natural or engineered, requires the presence of plasmid(s) harboring genes that encode each one. In order to make our system modular, we decided to utilize two separate plasmids—one encoding NMT and the other encoding a substrate protein—rather than one plasmid encoding both. Changes to the various features of each plasmid (i.e., antibiotic resistance, origin of replication, promoter regions) could be readily made, if desired, to independently control the expression of NMT and the substrate protein. Moreover, we postulated that

replacement of our test protein, GFP, with other substrate proteins for future studies would be more straightforward if NMT were encoded on a separate plasmid.

Two GFP-based substrates, yARF-GFP and Fyn-GFP, were prepared, each displaying an NMT recognition sequence derived from a known substrate protein. The yARF (yeast ADP-ribosylation factor) protein is a GTP-binding protein that helps to control trafficking within the cell.[11] We engineered yARF-GFP to carry the first seven residues of yARF: MGLFASK.[12] The Fyn protein is a member of the Src kinase family,[13] and its first nine residues, MGCVQCKTK, are displayed by Fyn-GFP. A C-terminal 6xHis tag was also added to each protein to aid in affinity purification and detection procedures. The cloning scheme developed to construct the substrate plasmids is described in the Experimental Section of this chapter. The NMT plasmids employed in our system encode human NMT1 or NMT2, as well as methionine aminopeptidase (Met-AP). The genes encoding NMT as well as both substrate proteins were placed under the control of IPTG (isopropyl β-D-1-thiogalactopyranoside)-inducible promoters.

The co-expression protocol that we developed is shown schematically in Figure II-1. Cells were transformed with the NMT and substrate plasmids, and overnight cultures were inoculated from the resultant cell stocks. Expression cultures were then grown until cells reached an appropriate cell density, at which point 12-ADA was added and protein expression was induced. After 3-4 hours of protein expression, cells were harvested. As described in Chapter I, various applications could then be explored, utilizing intact cells, clarified lysates, or purified protein.

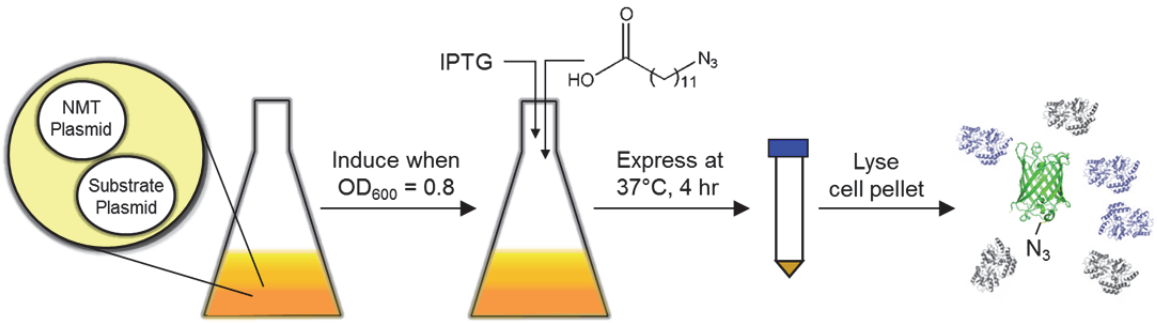


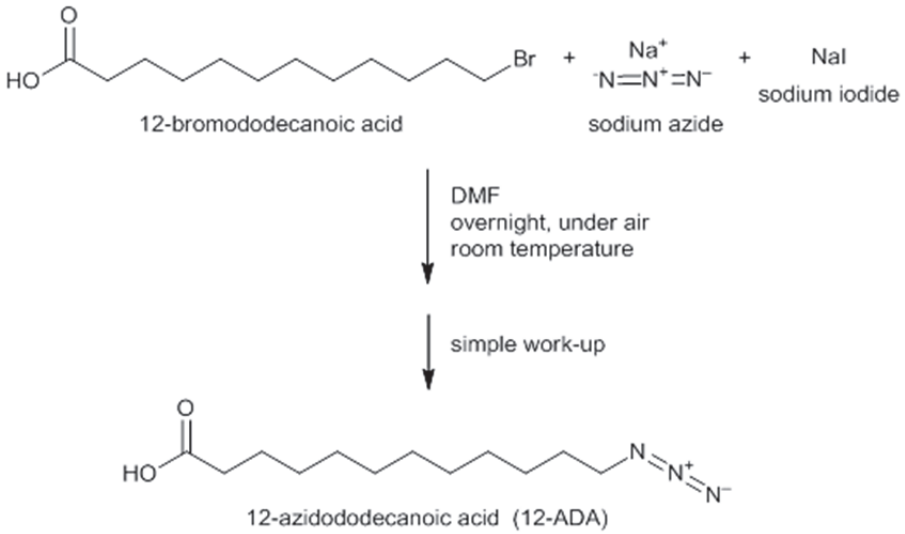
Figure II-1. Experimental overview of our bacterial co-expression system for N-terminal protein labeling with NMT.

RESULTS AND DISCUSSION

Synthesis and Characterization of 12-ADA

The compound 12-ADA was synthesized and purified as shown in Scheme II-1. The protocol we developed was adapted from literature precedent,[2] and the product identity was confirmed by comparison with published IR, ^{1}H NMR, ^{13}C NMR, and ESI-MS data.[2,4] After optimizing the reaction, work-up, and purification steps, near-quantitative yields of 12-ADA were routinely obtained. Though literature protocols described a flash chromatography step for purification, we observed no difference in purity after a work-up only.

Scheme II-1. The azide fatty acid utilized in our studies, 12-azidododecanoic acid (12-ADA), was prepared from simple precursors.



Expression and Purification of GFP-Based Substrate Proteins

The plasmids pQE80_yARF-GFP and pQE80_Fyn-GFP were constructed using standard cloning procedures, as described in the Experimental Section of this chapter. The final plasmids were transformed into *E. coli* BL21(DE3) competent cells that had already been transformed with plasmids encoding human NMT1 or human NMT2, yielding four cell strains: BL21(DE3)/yARF-GFP/hNMT1, BL21(DE3)/yARF-GFP/hNMT2, BL21(DE3)/Fyn-GFP/hNMT1, and BL21(DE3)/Fyn-GFP/hNMT2. We moved forward with two of these cell strains: BL21(DE3)/yARF-GFP/hNMT1 and BL21(DE3)/Fyn-GFP/hNMT2. These two strains were selected due to literature reports regarding the substrate selectivity of each human NMT isoform: yARF has been shown to be myristoylated to a greater extent by hNMT1 than hNMT2,[12] while *in vitro* studies

demonstrated the preference of hNMT2 for a Src peptide substrate.[14] As noted earlier, Fyn is a member of the Src kinase family.

Co-expression of yARF-GFP or Fyn-GFP with NMT was completed as depicted in the schematic overview above (Figure II-1). Briefly, overnight cultures were diluted 1:50 into expression cultures containing the appropriate antibiotics to ensure retention of both plasmids. Protein expression was induced at a cell density of $OD_{600} \approx 0.8$ with the addition of IPTG, and at the same time, 12-ADA was added. After 4 hours of expression at 37°C, cells were harvested from each culture and lysed. Each protein, yARF-GFP or Fyn-GFP, was purified from lysate via denaturing or native Nickel-NTA chromatography. Expression and purification protocols are presented in more detail in the Experimental Section at the end of this chapter. Purification fractions were analyzed by SDS-PAGE, which confirmed the presence of each protein at the expected molecular weight of 32 kDa (Figure II-2).

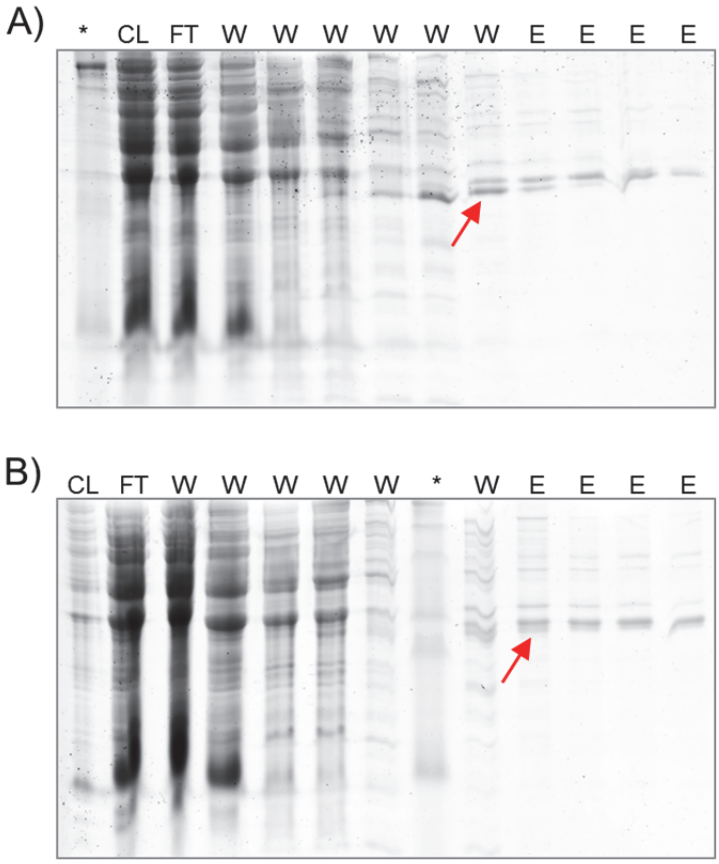


Figure II-2. SDS-PAGE analysis of native Ni-NTA purification fractions. Red arrows indicate the presence of pure yARF-GFP (A) and Fyn-GFP (B) isolated from co-expression cultures grown in the presence of 12-ADA. CL = Clarified Lysate; FT = Flow Through; W = Wash; E = Elution; * = MW marker.

MALDI-MS Analysis of Trypsinized Protein Samples

After expressing and purifying yARF-GFP and Fyn-GFP, we verified that both proteins had been labeled with 12-ADA and determined the extent of their modification. First, to confirm that labeling had occurred only at the protein N-terminus, we employed matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS).

Purified samples of yARF-GFP and Fyn-GFP were digested with the trypsin protease, which cleaves peptides after arginine (Arg, R) and lysine (Lys, K) residues,

yielding a predictable set of fragments. We were particularly interested in detecting the N-terminal fragment of each protein. The expected masses for the N-terminal peptide of each protein—unlabeled, myristoylated, or labeled with 12-ADA—are presented in Table II-1.

Table II-1. Expected masses of the unlabeled (Mass 1), myristoylated (Mass 2), and 12-ADA-labeled (Mass 3) N-terminal peptide fragments of yARF-GFP and Fyn-GFP. Masses account for removal of the initial Met residue, trypsin digestion, reduction, and alkylation. Formation of a covalent bond between Gly and myristic acid or 12-ADA results in the loss of one water molecule. All values are in Daltons (Da).

Protein (N-terminal peptide)	Mass 1	Mass 2 (+Myr)	Mass 3 (+12-ADA)
yARF-GFP (GLFASK)	622.20	832.56	845.52
Fyn-GFP (GCVQCK)	751.22	961.58	974.54

After trypsinization, the resultant peptide fragments for each protein sample were prepared for MALDI-MS analysis using standard clean-up columns and matrix reagents. The MALDI mass spectrum for each protein is shown in Figure II-3.

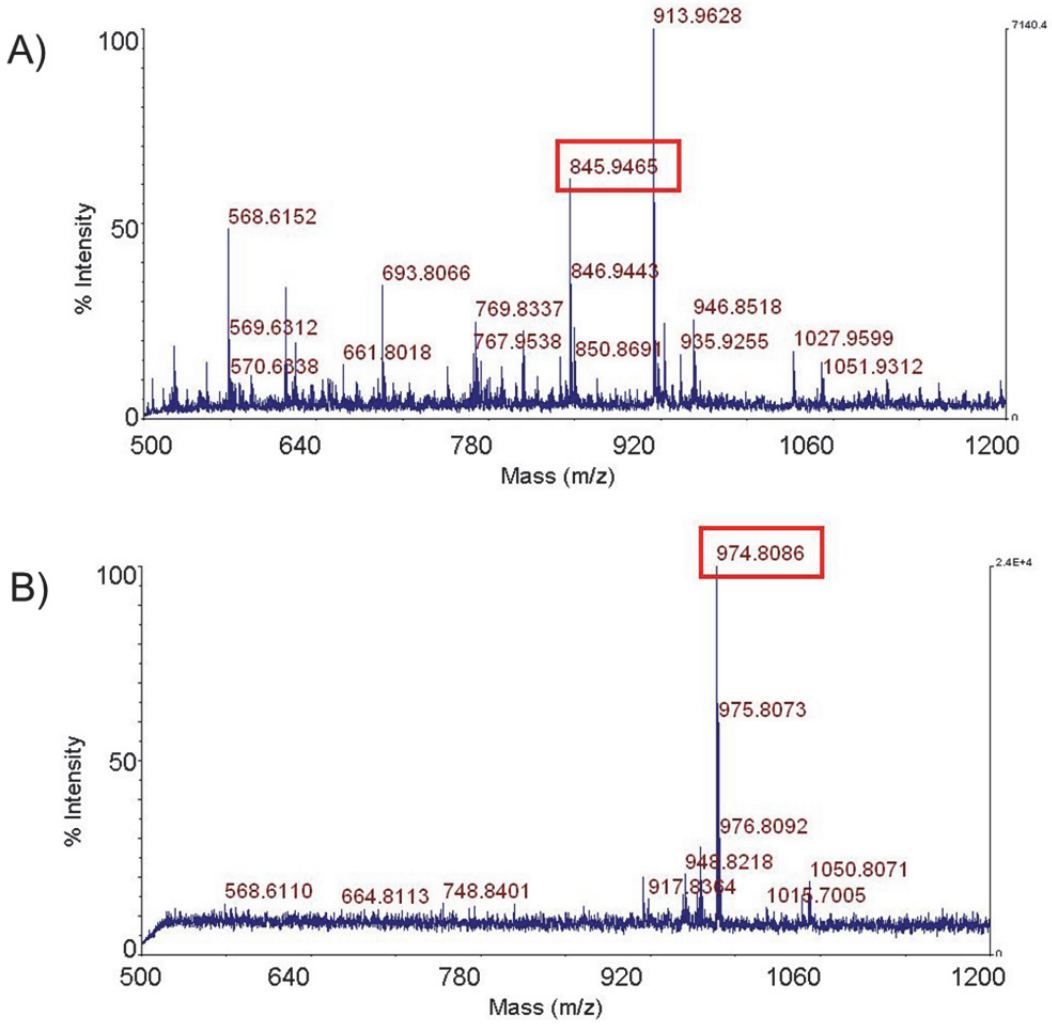


Figure II-3. MALDI mass spectra for yARF-GFP (A) and Fyn-GFP (B) indicate successful N-terminal labeling with 12-ADA. Samples were co-expressed with NMT in the presence of 12-ADA, purified, and trypsinized. The Fyn-GFP sample was also subjected to reduction and alkylation prior to trypsin digestion. No unlabeled or myristoylated N-terminal peptide fragments were observed (see Table II-1).

In analyzing each spectrum, we first looked for peaks corresponding to various expected fragments to ensure that the digestion was effective. We also searched for each N-terminal peptide. As shown in Figure II-3A, the 12-ADA-labeled N-terminal fragment for yARF-GFP was readily identified. For Fyn-GFP, initially, we found that the

12-ADA-labeled N-terminal fragment was consistently 2 Da lower than expected (data not shown), perhaps due to disulfide bonding of the two cysteine residues found in that fragment. Thus, we also examined Fyn-GFP samples that were reduced and alkylated following trypsinization. After such treatment, the 2 Da discrepancy was no longer observed (Figure II-3B). Notably, the myristoylated N-terminal fragment was not detected for either protein, indicating that competition between myristic acid and 12-ADA is not a problem in our system. In addition, negative control samples that were not exposed to 12-ADA during expression yielded spectra that lacked the peaks outlined in red in Figure II-3 (the "Mass 3" values in Table II-1), further confirming that those peaks correspond to the 12-ADA-labeled N-terminal fragments.

Intact LC-MS Analysis of Protein Samples

While MALDI-MS analysis of trypsinized yARF-GFP and Fyn-GFP samples showed that both proteins had been labeled with 12-ADA, they did not indicate what percent of the protein pool was tagged. For both proteins, we were unable to detect the unlabeled N-terminal peptide fragment (Table II-1, "Mass 1"), even for negative control samples that were not exposed to 12-ADA, likely due to the fragments' polarity, charge, and small size. The inability to detect these fragments by MALDI-MS complicated efforts to measure the extent of 12-ADA labeling of yARF-GFP and Fyn-GFP. Fortunately, intact LC-MS of purified samples proved to be a more effective approach. The expected masses of the relevant protein species, as well as those actually observed in whole-protein mass spectra of yARF-GFP and Fyn-GFP, are presented in Table II-2.

Table II-2. Expected masses of yARF-GFP and Fyn-GFP (unlabeled, myristoylated, or labeled with 12-ADA), and masses observed by intact LC-MS. Expected mass values account for removal of initial Met and loss of 20 Da upon formation of GFP chromophore. Analysis revealed quantitative labeling of both proteins with 12-ADA by NMT. All mass values are in Da. N/D = Not Detected.

Protein	Expected Mass	Observed Mass	% 12-ADA-Labeled
yARF-GFP	31,817.85	N/D	> 98 %
+Myr	32,028.20	N/D	
+12-ADA	32,041.16	32,041.31	
Fyn-GFP	32,062.17	N/D	> 98 %
+Myr	32,272.52	N/D	
+12-ADA	32,285.48	32,280.02	

For both proteins, we routinely observed a large peak corresponding to the 12-ADA labeled species, very little evidence of unlabeled protein (<5%), and no evidence of myristoylated protein. Taken together, the two sets of MS data demonstrate that yARF-GFP and Fyn-GFP are both labeled solely and quantitatively at the N-terminus with 12-ADA by NMT, with no competition presented by myristic acid.

As a step toward optimizing our co-expression system, we investigated the use of both minimal medium and rich media for our expression cultures. Certain protocols developed in the Tirrell Lab for residue-specific incorporation of ncAAs recommend the use of minimal medium in order to boost ncAA incorporation, while avoiding the use of rich media that contain a given natural amino acid. We hypothesized that myristic acid might similarly be present in rich media, which could negatively impact the extent of protein labeling with 12-ADA. Thus, we originally used M9 minimal medium for protein expression, but later tested the use of rich media (LB and 2xYT). Identical LC-MS

results were obtained for Fyn-GFP regardless of the expression medium, with no myristoylated protein species observed in any of the samples (Figure II-4).

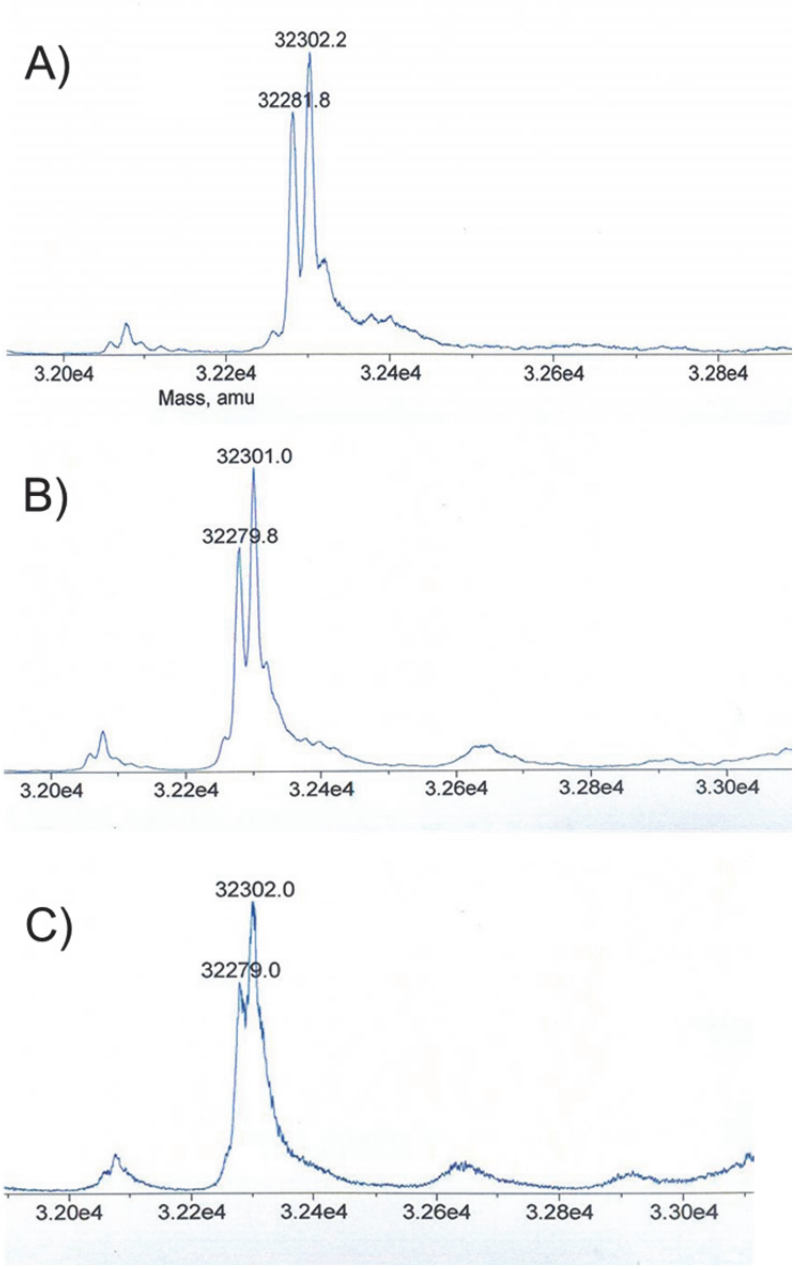


Figure II-4. Intact LC-MS data for Fyn-GFP samples purified from expression cultures grown in M9 (A), LB (B), or 2xYT (C) medium. The compound 12-ADA was added to all three cultures. No myristoylated Fyn-GFP was observed in any sample (see Table II-2).

These results are notable because the use of rich media for protein expression instead of a minimal medium often results in higher protein yields. Moreover, as with the data presented in Table II-2, very little unlabeled protein was detected.

Fluorescence Gel Analysis of Lysate Samples

As mentioned at the beginning of this chapter, one notable feature of NMT is its absence from prokaryotic systems. We were intrigued by the possibility of developing a chemoenzymatic labeling system that is not only site-specific and high-yielding, but also orthogonal to the widely used *E. coli* bacterial expression system. To investigate whether or not the yARF-GFP/hNMT1 and Fyn-GFP/hNMT2 co-expression systems fulfill this additional criterion, we took advantage of existing bioorthogonal chemistries and reagents to detect the presence of 12-ADA-labeled proteins.

For these experiments, lysates were utilized to effectively measure NMT's selectivity toward our engineered substrate proteins. Lysate samples were treated with an alkyne-TAMRA probe in a CuAAC reaction, which is known to conjugate azides and alkynes even in complex biological settings.[15,16] Proteins were subsequently precipitated out of the reaction mixture and analyzed via SDS-PAGE (Figure II-5). For each protein, the same gel was imaged for TAMRA fluorescence, then stained with Coomassie colloidal blue and imaged again. As the gel images indicate, NMT transferred 12-ADA only to yARF-GFP and Fyn-GFP and did not label natural bacterial proteins. These data demonstrated the activity and selectivity of our NMT-mediated protein labeling system, even when using a non-natural fatty acid substrate in conjunction with non-natural substrate proteins.

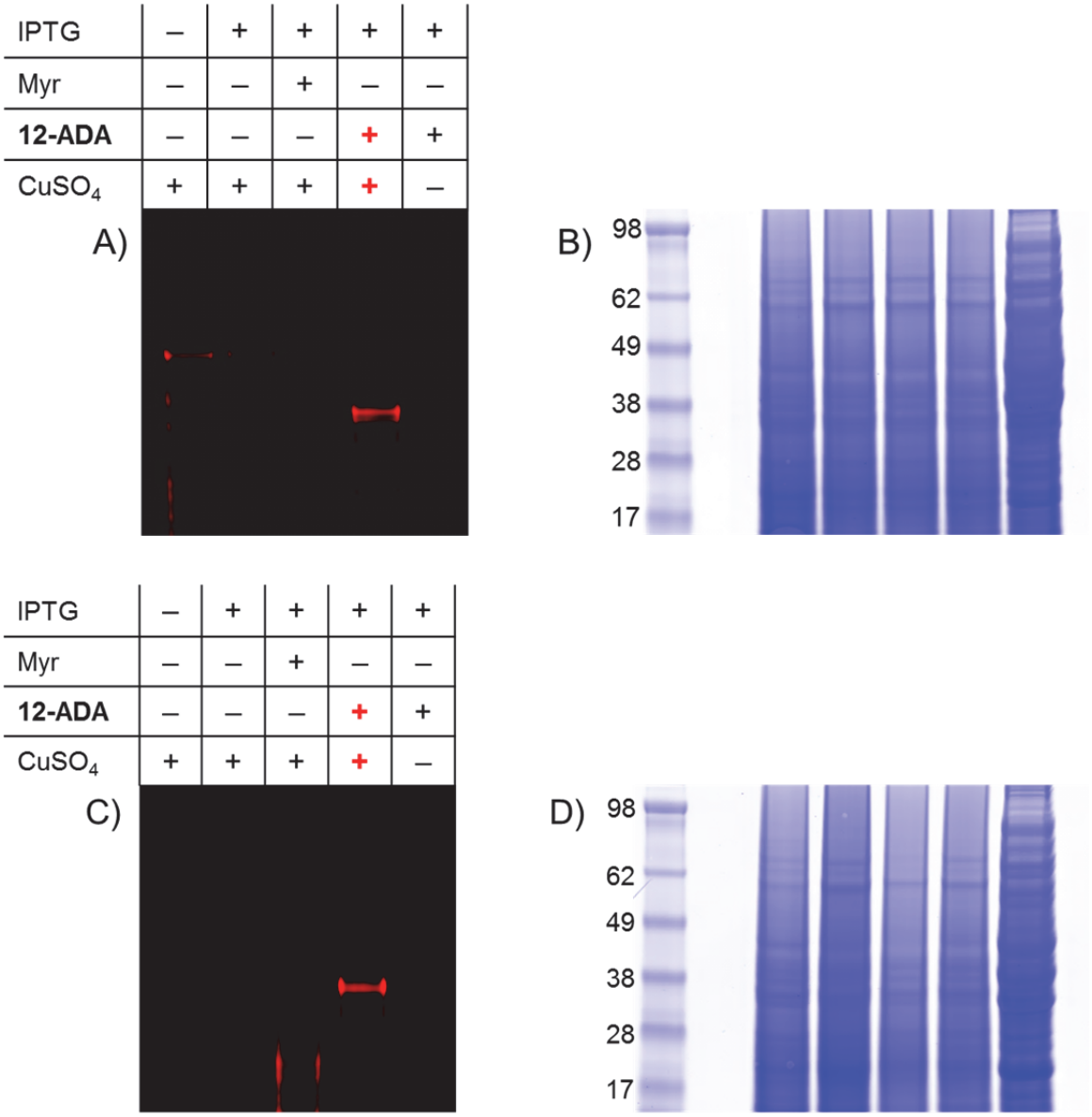


Figure II-5. SDS-PAGE analysis of lysate samples containing yARF-GFP (A, B) or Fyn-GFP (C, D) co-expressed with NMT in the presence of no fatty acid, myristic acid ("Myr"), or 12-ADA. Lysate samples were treated with alkyne-TAMRA for detection of azide-labeled protein (A, C). The same gels were stained with Coomassie colloidal blue (B, D). Comparison of each pair of gel images indicates selective 12-ADA labeling of yARF-GFP and Fyn-GFP.

CONCLUSION

In summary, we designed, constructed, and evaluated a site-specific protein labeling system centered on the eukaryotic enzyme N-myristoyl transferase (NMT). We developed two plasmids encoding GFP-based, non-natural substrates of NMT: yARF-GFP and Fyn-GFP. Both proteins display a recognition sequence derived from a known NMT substrate. We co-expressed each protein with an isoform of human NMT in the presence of the azido fatty acid, 12-ADA, which was readily synthesized. Purified protein samples were subjected to MALDI-MS analysis after trypsinization and were studied via intact LC-MS as well. The MS data sets showed that labeling by NMT was specific for the N-terminus, as expected, and essentially quantitative. Examination of lysate samples treated with an azide-reactive dye confirmed that NMT labels each engineered GFP substrate, but is inactive toward natural bacterial proteins. The site-specific, quantitative, and selective protein labeling system established here is the basis of the further research described in Chapters III, IV, and V.

EXPERIMENTAL SECTION

Materials

Synthesis of 12-ADA. 12-bromododecanoic acid and sodium iodide were purchased from Aldrich. Silica gel 60 was purchased from EMD Chemicals. Sodium azide and all solvents were purchased from VWR.

Cloning. All oligonucleotide primers were ordered from Operon. Polymerase chain reaction (PCR) experiments were carried out in a BioRad DNA Engine Peltier

Thermal Cycler using PfuTurbo DNA Polymerase (Stratagene/Agilent). All restriction enzymes, restriction enzyme buffers, and bovine serum albumin (BSA) were purchased from New England BioLabs (NEB). The pQE60 and pQE80 vectors from Qiagen were used for cloning. NEB DNA Ladders (100 bp and 1 kbp "Quick-Load") were used as markers for all DNA agarose gels, which were visualized with the addition of Plus One ethidium bromide solution from Amersham Biosciences on a UVP UV Transilluminator. Zymo Agarose-Dissolving Buffer (ADB) and Zymo Spin II columns, with their associated buffers, were used to purify DNA out of agarose gels. T4 DNA Ligase from NEB or Roche was used for ligations with equivalent results. All DNA acquisition from cells was completed using the Qiagen Spin Miniprep Kit and columns. All sequencing requests were fulfilled by Laragen.

Protein expression. Plasmids encoding hNMT1 or hNMT2 and methionine-aminopeptidase (Met-AP) were a gift from the laboratory of Professor Richard Kahn at Emory University (Atlanta, GA).[12] *E. coli* BL21(DE3) cells were made chemically competent using the standard Zymo method (Stratagene) and were transformed with either the hNMT1 plasmid or hNMT2 plasmid. M9 minimal medium was composed of M9 salts plus 0.4% dextrose, 100 μM $CaCl_2$, 35 μg/mL thiamine, 1 mM $MgSO_4$, and 4% 20 amino-acid solution (1 g/L each). LB medium was composed of 10 g tryptone (casein hydrolysate), 5 g yeast extract, and 10 g NaCl per liter. 2xYT medium was composed of 16 g tryptone (casein hydrolysate), 10 g yeast extract, and 5 g NaCl per liter. All media were autoclaved before use. Kanamycin (Kan) was used at a working concentration of 35 μg/mL, and ampicillin (Amp) was used at a working concentration of 200 μg/mL. Myristic acid was purchased from Fluka. All optical density (OD) values were measured at 600 nm on a Cary UV-Vis spectrophotometer. All SDS-PAGE gels described in this chapter were 12% acrylamide, Tris-Tricine gels cast in-house or NuPAGE Novex 4-12% Bis-Tris pre-cast gels (Invitrogen). SeeBlue Plus2 Pre-Stained Protein Marker from Invitrogen served as the molecular weight ladder. Gels were stained with Coomassie colloidal blue from Invitrogen.

Protein purification. Nickel-NTA resin manufactured by Qiagen was used for purification of 6xHis-tagged yARF-GFP and Fyn-GFP from lysate. For denaturing Ni-NTA purification, Buffers B, C, D, and E consisted of 8 M urea, 100 mM NaH_2PO_4, and 10 mM Tris-Cl with pH = 8.0, 6.3, 5.9, or 4.5, respectively. For native Ni-NTA purification, buffers contained 50 mM NaH_2PO_4, 300 mM NaCl, and varying concentrations of imidazole (25–500 mM). B-PER lysis buffer from Pierce was usually used for native Ni-NTA purification. Lysozyme was purchased from Aldrich.

Mass spectrometry. The Pierce BCA Assay Kit was used to measure protein concentration in pure protein fractions prior to MS analysis. Promega porcine trypsin was used in digests. Microcon Centrifugal Devices were used to concentrate and buffer-exchange whole-protein samples for intact LC-MS, while Microcon Centrifugal Devices and C_{18} Zip-Tips were both employed to concentrate and de-salt peptide samples for MALDI-MS (both from Millipore). MALDI-MS data were collected on an Applied Biosystems Voyager DE-PRO MALDI TOF-MS. Intact LC-MS data were collected on an Agilent 1100 MSD quadrupole ESI-MS.

Fluorescence detection. Lysate samples were treated with the reagents and according to the protocols of the Click-IT Tetramethylrhodamine (TAMRA) Protein Analysis Detection Kit from Invitrogen. After reaction and precipitation, protein samples were run on Invitrogen NuPAGE Novex 4%–12% Bis-Tris pre-cast gels and imaged on a GE Typhoon laser scanner. Gels were stained with Coomassie colloidal blue from Invitrogen.

Methods

Synthesis of 12-ADA. 12-Azidododecanoic acid (12-ADA) was synthesized according to literature precedent[2] with minor modifications; in particular, the flash chromatography step was found to be unnecessary to obtain pure product. Standard characterization techniques (ESI-MS, ^{1}H NMR, ^{13}C NMR) yielded data that matched published results.[2,4]

Cloning. The pQE80 plasmid possesses an ampicillin resistance gene and a Col E1 Origin of Replication. The GFP gene was PCR-amplified from the vector pQE9_GFP6_lacI_yPheRS_T415G, which was prepared by a former member of the Tirrell Lab, Inchan Kwon. Two rounds of PCR with novel primers yielded a fragment including the GFP gene carrying an EcoRI restriction enzyme site, a ribosome-binding site appropriate for pQE vectors, and base pairs encoding the yeast ARF recognition sequence (MGLFASK, from ATG GGT CTG TTC GCG TCT AAA) or the Fyn recognition sequence (MGCVQCKTK, from ATG GGT TGC GTG CAA TGC AAA ACC AAA) at the 5′ end, and a BglII site at the 3′ end. The PCR product and pQE60 were digested with EcoRI and BglII, and the GFP insert was ligated into pQE60 to allow for the addition of a C-terminal 6xHis tag; each GFP construct was also placed under the control of a T5 promoter, which is inducible by isopropyl β-D-1-thiogalactopyranoside (IPTG). The presence of the insert in the pQE60 construct was confirmed via test digests and sequencing. Then, the entire cassette was digested out of pQE60 using AatII and NheI and ligated into pQE80, which had been digested with the same two enzymes. The final pQE80 plasmids were used for protein expression because pQE80 carries the $lacI^q$ repressor necessary for *cis* regulation of the T5 promoter. Each final construct, pQE80_yARF-GFP or pQE80_Fyn-GFP, was sequenced and transformed into competent cells already harboring an NMT plasmid for co-expression experiments.

Protein expression. Overnight cultures were inoculated in LB supplemented with Kan and Amp and grown in an incubator-shaker (37°C, 250 rpm). The following day, overnight cultures were diluted 1:50 into fresh M9, LB, or 2xYT supplemented with Kan and Amp for expression cultures, which ranged in volume from 5 mL to 250 mL. Cultures were grown in an incubator-shaker (37°C, 250 rpm), and protein expression was induced with IPTG (1 mM, from 1 M stock in water) when the OD_{600} value was between 0.8 and 1.1. Pre-induction samples (1 mL) were collected as needed. The azide fatty acid 12-ADA (500 μM, from 500 mM stock in DMSO) was also added at the time of induction. After 3–4 hr of protein expression, cells were harvested via centrifugation (10 min x 10,000 *g*) and the final OD_{600} value measured. Cell pellets were lysed according to the following formula, regardless of which lysis buffer was used: 50 μL

lysis buffer per mL culture per OD_{600} unit. Crude lysates were centrifuged once more, and the supernatant (clarified lysate, i.e., cytosolic fraction) was saved for further experiments.

Protein purification. Protocols were followed largely as described in *The QIAExpressionist* handbook from Qiagen. For denaturing Ni-NTA purification, cells were lysed in Buffer B. The clarified lysate was incubated with Ni-NTA agarose for 1–2 hr at 4°C with agitation and loaded on an empty column to collect fractions. The protein-agarose mixture was washed with Buffers B, C, and D, and the protein eluted with Buffer E (pH 4.5). Purification fractions were analyzed by SDS-PAGE.

For native Ni-NTA purification, cells were lysed in B-PER buffer or in the standard native Ni-NTA buffer (see "Materials") with 10 mM imidazole. Furthermore, for native Ni-NTA purification, lysozyme was added to the lysis buffer at 1 mg/mL, and the resuspended cell pellet was sonicated to aid in lysis. The clarified lysate was incubated with Ni-NTA agarose for 1–2 hr at 4°C with agitation and loaded on an empty column to collect fractions. The protein-agarose mixture was washed with buffer containing imidazole (25–100 mM) and the protein eluted at 150-200 mM imidazole. Purification fractions were analyzed by SDS-PAGE.

Mass spectrometry. For MALDI-MS experiments, solutions of pure protein were concentrated using Microcon columns (MWCO = 30 kDa). Fyn-GFP samples were reduced and alkylated according to a standard published protocol.[17] Protein samples were digested as follows: 90 μL of 75 mM NH_4CO_3 buffer and 1 μL porcine trypsin were added to 10 μL of a concentrated protein solution, and the mixtures were incubated at 37°C for 2–8 hr, after which they were quenched with 10 μL 10% TFA. C_{18} ZipTips were used to concentrate and de-salt the trypsin digest mixtures in preparation for MALDI-MS (α-cyanohydroxycinnamic acid matrix).

For intact LC-MS experiments, solutions of pure protein were concentrated using Microcon columns (MWCO = 30 kDa) and buffer-exchanged into a 0.1% TFA (trifluoroacetic acid) solution. A final solution of 100 pmol protein in 100 μL was run on the MSD instrument.

Fluorescence detection. Cells were lysed with the buffer recommended in the instructions for the Invitrogen Click-iT kit (1% SDS, 50 mM Tris-HCl, pH 8.0) according to the formula mentioned earlier (50 μL lysis buffer per mL culture per OD_{600} unit). Clarified lysate samples were treated with alkyne-TAMRA and other kit reagents according to the protocols supplied by Invitrogen; the only modification was the use of 15 μL of alkyne-TAMRA dye solution rather than 100 μL. At the conclusion of the 25-min reaction time, samples were precipitated following the methanol-chloroform precipitation protocol described in the same kit instructions; the only modification was the completion of one additional methanol wash of the protein pellet. For SDS-PAGE analysis, protein pellets were resuspended in a denaturing buffer (8 M urea, 100 mM NaH_2PO_4, and 10 mM Tris-Cl) and loaded on NuPAGE Novex 4-12% Bis-Tris pre-cast gels. To detect TAMRA signal on the Typhoon, the 532 nm laser served as the excitation source (filter set: 580 BP 30). Gels were stained with Coomassie colloidal blue, then imaged again, with the 633 nm laser now serving as the excitation source (no filter).

REFERENCES

1. Kishore, N. S. *et al.* The substrate specificity of Saccharomyces cerevisiae myristoyl-CoA: protein N-myristoyltransferase. *J. Biol. Chem.* **266**, 8835–55 (1991).

2. Devadas, B. *et al.* Substrate specificity of Saccharomyces cerevisiae myristoyl-CoA: protein N-myristoyltransferase. *J. Biol. Chem.* **267**, 7224–39 (1992).

3. Lu, T. *et al.* The substrate specificity of Saccharomyces cerevisiae myristoyl-CoA: protein N-myristoyltransferase. Polar probes of the enzyme's myristoyl-CoA recognition site. *J. Biol. Chem.* **269**, 5346–57 (1994).

4. Hang, H. C. *et al.* Chemical probes for the rapid detection of fatty-acylated proteins in mammalian cells. *J. Am. Chem. Soc.* **129**, 2744–5 (2007).

5. Saxon, E. & Bertozzi, C. R. Cell surface engineering by a modified Staudinger reaction. *Science* **287**, 2007–10 (2000).

6. Rostovtsev, V. V, Green, L. G., Fokin, V. V & Sharpless, K. B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.* **41**, 2596–9 (2002).

7. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057–64 (2002).

8. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046–7 (2004).

9. Agard, N. J., Baskin, J. M., Prescher, J. A., Lo, A. & Bertozzi, C. R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **1**, 644–8 (2006).

10. Sletten, E. M. & Bertozzi, C. R. From mechanism to mouse: a tale of two bioorthogonal reactions. *Acc. Chem. Res.* **44**, 666–76 (2011).

11. Pasqualato, S., Renault, L. & Cherfils, J. Arf, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for "front-back" communication. *EMBO Rep.* **3**, 1035–41 (2002).

12. Van Valkenburgh, H. A. & Kahn, R. A. Coexpression of proteins with methionine aminopeptidase and/or N-myristoyltransferase in Escherichia coli to increase acylation and homogeneity of protein preparations. *Methods Enzymol.* **344**, 186–93 (2002).

13. Palacios, E. H. & Weiss, A. Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation. *Oncogene* **23**, 7990–8000 (2004).

14. Giang, D. K. & Cravatt, B. F. A second mammalian N-myristoyltransferase. *J. Biol. Chem.* **273**, 6595–8 (1998).

15. Wang, Q. *et al.* Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 3192–3 (2003).

16. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed. Engl.* **48**, 6974–98 (2009).

17. Kinter, M. & Sherman, N. E. Chapter 6: The Preparation of Protein Digests for Mass Spectrometric Sequencing Experiments. *Protein Sequencing and Identification Using Tandem Mass Spectrometry* 147–165 (2000).

CHAPTER III

Selective N-Terminal Modification of Calcineurin and Calmodulin

ABSTRACT

The NMT-mediated protein labeling system described in Chapter II serves as the foundation of the work described in this chapter. We applied the principles and techniques developed for our GFP-based model system to achieve selective N-terminal labeling of two neuronal proteins, calcineurin (CaN) and calmodulin (CaM). Both proteins are implicated in the complex pathways governing learning and memory, and both have been studied by neuroscience researchers for decades. In the projects described herein, our objective was to utilize NMT and 12-ADA to site-specifically functionalize CaN and CaM for subsequent surface immobilization. (Note that surface immobilization experiments are described in Chapter IV.) We also aimed to create protein constructs that retained wild-type levels of activity even after undergoing engineering and labeling.

Because CaN is a natural substrate of NMT, we proceeded directly to expression and labeling experiments. However, CaM is not naturally myristoylated. Thus, we prepared a family of engineered CaM constructs, each displaying an NMT recognition sequence. CaN and the CaM constructs were co-expressed with NMT in the presence of 12-ADA for N-terminal labeling and purified for intact LC-MS analysis and activity assays. We were pleased to find that CaN and one of the engineered CaM constructs did retain wild-type activity, even after labeling with 12-ADA. We also confirmed, via treatment of lysate samples with an azide-reactive dye, that NMT is selective toward CaN and the engineered CaM constructs in bacteria. In summary, we extended our NMT-based protein labeling system from a model protein, GFP, to proteins that perform important functions in complex organisms.

INTRODUCTION

Calcineurin and Calmodulin

Calcineurin (CaN) is a 79-kDa heterodimeric serine/threonine phosphatase that is enriched in the brain.[1] The 60-kDa A subunit of CaN contains the catalytic domain of the enzyme as well as four other domains: a regulatory domain, an autoinhibitory domain, and domains for binding the CaN B subunit and calmodulin (CaM). Of interest for our work, the 19-kDa B subunit is naturally myristoylated.[2] It is apparent from the crystal structure of CaN (Figure III-1A) that the N-terminus of the CaN B subunit is accessible, a common feature of NMT substrate proteins.

CaN performs important functions in different systems of the body. It is involved in the signal transduction pathways of numerous cell types, including lymphocytes and kidney cells.[3] In the brain, CaN plays a role in synaptic plasticity; CaN knock-out mice experience memory deficits and exhibit characteristics related to schizophrenia, such as impaired social interaction and disrupted nesting behaviors.[4] Because CaN is highly conserved among eukaryotes and is involved in a variety of cellular processes, it has been studied extensively by neuroscientists and behavioral scientists as well as immunologists. Notably, CaN becomes active only upon binding of its A subunit by CaM.

In contrast with CaN, CaM is a small, 16-kDa, monomeric protein.[5] But, like CaN, CaM undergoes conformational changes upon binding Ca^{2+} ions. In fact, CaM is generally inactive toward the more than 100 proteins it regulates until it binds four Ca^{2+} ions (Figure III-1B), at which point it is able to activate other proteins and enzymes, including CaN. CaM is also highly conserved across eukaryotes and is present in all eukaryotic cells; it is enriched in the brain and involved in synaptic plasticity, like CaN.[6]

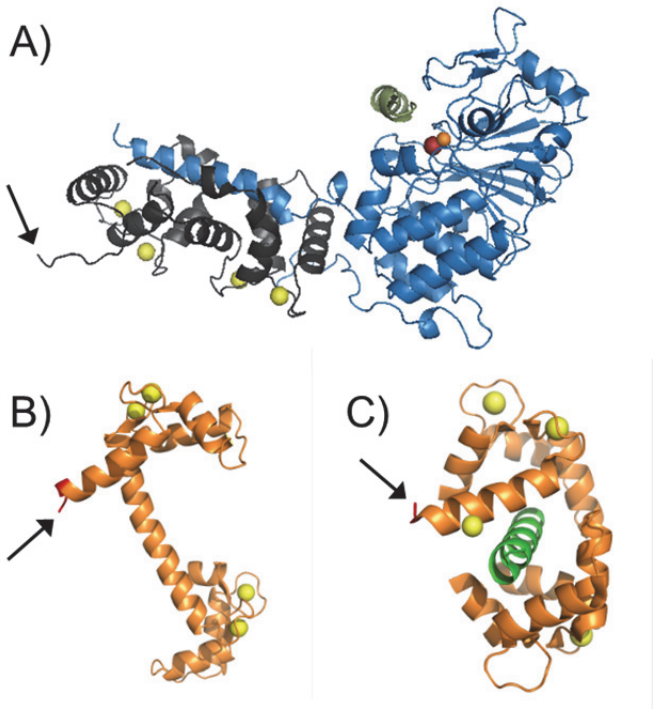


Figure III-1. Crystal structures of CaN (A) and CaM (B, C). Black arrows point to each protein N-terminus. The conformation of CaM shown in (B) is observed upon binding of four Ca^{2+} ions, and that depicted in (C) corresponds to further structural changes upon binding of a partner protein by CaM. For (A): PDB ID = 1AUI, yellow = Ca^{2+}; red = Fe^{3+}; orange = Zn^{2+}; blue = CaN A subunit; grey = CaN B subunit; green = peptide substrate. For (B): PDB ID = 1CLL and (C): PDB ID = 1CDM, yellow = Ca^{2+}; green = peptide derived from CaM kinase II, a CaM binding partner.

We were intrigued by the similarities and differences apparent upon comparison of CaN and CaM. As our interests include expanding the scope of our original NMT-mediated protein labeling system, we were particularly excited by the parallel study of a large natural NMT substrate requiring no engineering, and a small, streamlined protein that could be challenging to engineer without an accompanying loss of function. With the ultimate objective of site-specifically functionalizing both CaN and CaM to prepare

protein microarrays for high-throughput biochemical studies, we began by preparing CaM constructs that would be recognized by NMT.

Engineering CaM for NMT-Mediated Protein Labeling

In designing new CaM-based constructs for labeling with NMT, we drew inspiration from the model system described in Chapter II, in which we prepared two GFP-based constructs: yARF-GFP and Fyn-GFP. For preparation of CaM constructs, we initially worked only with the yARF recognition sequence (MGLFASK) because the Fyn recognition sequence (MGCVQCKTK) contains two cysteine (Cys, C) residues. CaM contains no Cys residues in its native form, and we hypothesized that addition of the nucleophilic thiol side chain of Cys could unfavorably alter the structure of CaM.

The first construct we prepared was yARF-6xHis-CaM (Table III-1), engineered to display the 6xHis affinity purification tag as well as the yARF sequence. Although expression, purification, and mass spectrometry experiments with this protein yielded good results, it was found to be four times less active than wild-type (WT) CaM. It was not apparent whether the loss of activity was caused by the addition of the recognition sequence or the 6xHis tag, or both. Thus, we sought to develop more constructs in order to (a) identify at least one construct that retained WT levels of activity and (b) better understand which sequence modification contributed more significantly to the loss of function observed for yARF-6xHis-CaM.

As summarized below, we selected a second recognition sequence, derived from the N-terminal region of CaN-B (MGNEASYPL), to enable labeling of CaM by NMT. We also explored the use of flexible linkers, postulating that the placement of additional

features upstream of CaM might have a smaller effect on CaM activity if an intervening spacer were present. A number of CaM fusion proteins have been reported, including multiple GFP-CaM fusions.[7–15] In most reports, the identity of the linker was not provided or was simply incidental to the cloning scheme, with minimal quantitation of the impact of the linker on CaM activity. However, one report described an active GFP-CaM fusion in which addition of a linker (SRLIGSA) and GFP to CaM was found "not to significantly affect the functional properties of the CaM molecule."[9] Thus, we also prepared constructs possessing this linker in addition to an NMT recognition sequence. A full summary of constructs is presented in Table III-1.

Table III-1. Summary of engineered CaM constructs developed for N-terminal protein labeling studies. The yARF and hCaNB recognition sequences are derived from known NMT substrates. The 6xHis tag was included for affinity purification and immunodetection purposes. The linker was previously described.[9] Estimated yields are based on actual yields from 100 mL cultures.

Construct Name	Amino Acid Sequence Preceding CaM	Estimated Pure Yield for 1 L Culture (mg)
yARF–CaM	MGLFASK–	2
yARF-Linker–CaM	MGLFASK-SRLIGSA–	4
yARF-6xHis–CaM	MGLFASK-HHHHHH–	35
hCaNB–CaM	MGNEASYPL–	62
hCaNB-Linker–CaM	MGNEASYPL-SRLIGSA–	45

RESULTS AND DISCUSSION

Expression and Purification of CaN

We followed literature protocols for expression and purification of human CaN from an *E. coli* co-expression system.[3] In addition to growing cultures in the presence of myristic acid, we also expressed CaN in the presence of 12-ADA. Identical expression and purification results were obtained for both. Yields of 2–3 mg pure protein per liter of bacterial culture have been reported; we obtained similar yields of pure Myr-CaN and pure 12-ADA-CaN.

Cloning, Expression, and Purification of Wild-Type and Engineered CaM Constructs

To prepare plasmids that encode the engineered CaM proteins listed in Table III-1, we followed different protocols than those described in Chapter II for cloning yARF-GFP and Fyn-GFP. Using a modified site-directed mutagenesis approach, termed "two-step PCR,"[16] we encoded the desired sequences directly into primers, with a plasmid that encodes *Drosophila melanogaster* CaM serving as the template; these steps are described in detail in the Experimental Section of this chapter. This approach was considerably simpler and more efficient than the cloning strategy used to construct yARF-GFP and Fyn-GFP.

After all five constructs had been prepared, the final plasmids were transformed into *E. coli* competent cells already harboring a plasmid encoding one of the two isoforms of human NMT. Following the co-expression protocol outlined in Chapter II, we grew cultures of all five engineered CaM constructs in the presence of 12-ADA and

grew a control culture of WT CaM. We then used phenyl sepharose resin to purify the proteins from lysate.

Phenyl sepharose purification relies on a hydrophobic interaction between CaM and the sepharose resin.[17] When CaM is in its Ca^{2+}-bound form, a hydrophobic patch on the protein is exposed, and it binds the resin. At this point, wash buffers containing high concentrations of salt and Ca^{2+} remove contaminants while CaM remains bound to the column. In contrast, elution buffers contain Ca^{2+} chelators, such as EGTA, which cause CaM to give up its Ca^{2+} ions and undergo a conformational change that conceals its hydrophobic patch. CaM no longer interacts with the resin and elutes off the column. (Purification protocols are presented in more detail in the Experimental Section.)

All proteins were obtained in a very pure form after phenyl sepharose purification, as evidenced by SDS-PAGE analysis (Figure III-2). The amount of pure protein yielded by each 100 mL culture was used to estimate the per-liter yield for each protein, shown in Table III-1. Unexpectedly, the yields for yARF-CaM and yARF-Linker-CaM were consistently and significantly lower than those measured for WT CaM and the other engineered constructs. Fortunately, purification of the 12-ADA-labeled engineered constructs (Figure III-2, B-F) appeared to be unaffected by the presence of the hydrophobic 12-ADA tag when compared to the purification results for WT CaM (Figure III-2A).

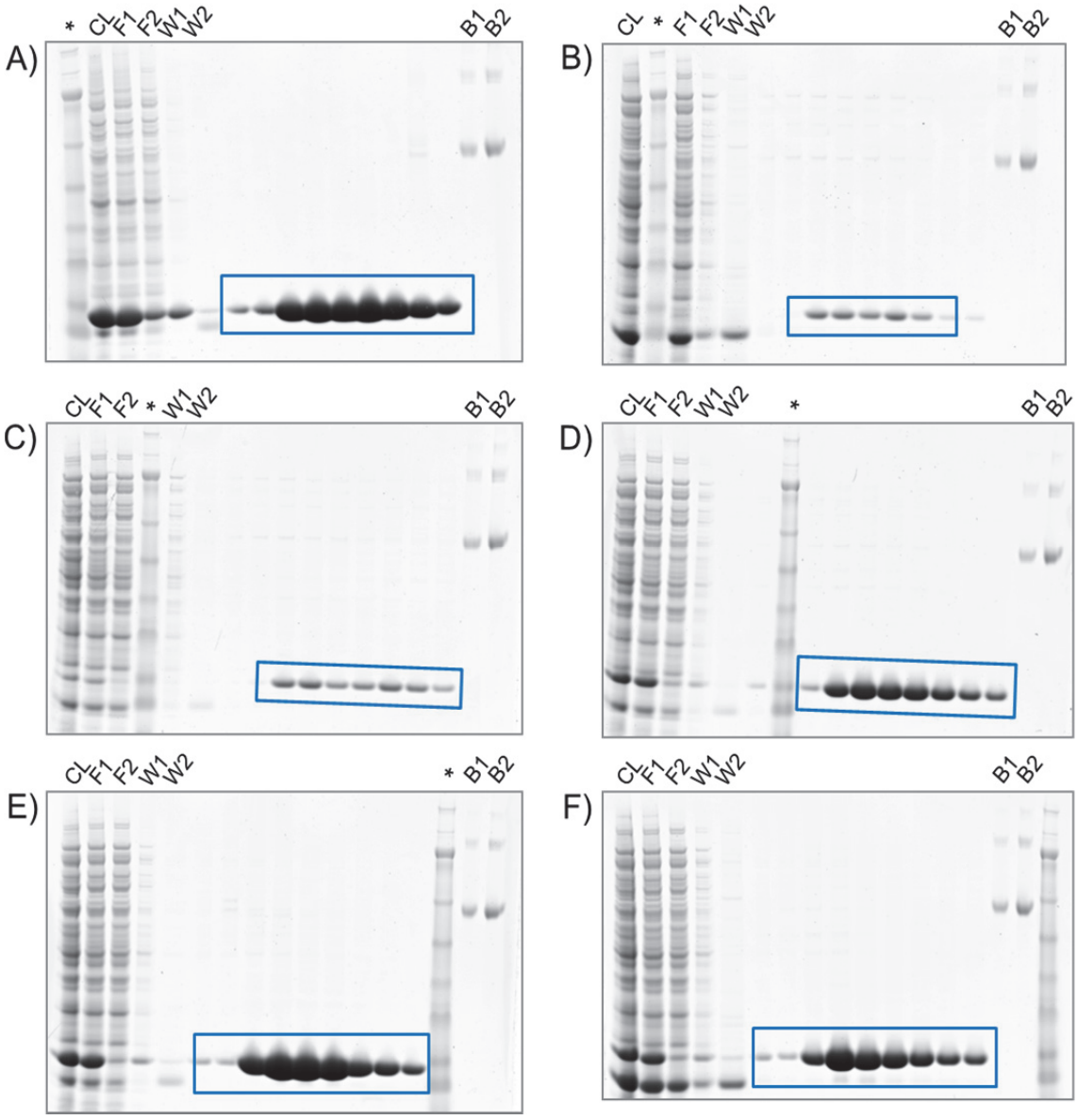


Figure III-2. SDS-PAGE analysis of phenyl sepharose purification fractions for each CaM construct. Pure protein is outlined with a blue box on each gel image: WT CaM (A); yARF-CaM (B); yARF-Linker-CaM (C); yARF-6xHis-CaM (D); hCaNB-CaM (E); hCaNB-Linker-CaM (F). All constructs except WT CaM were expressed in the presence of 12-ADA for N-terminal labeling by NMT. * = MW marker; CL = Clarified Lysate; F = Flow-Through; W = Wash. B1 = 0.05 mg/mL BSA; B2 = 0.1 mg/mL BSA.

Intact LC-MS Analysis of Protein Samples

After purifying 12-ADA-CaN, WT CaM, and 12-ADA-labeled engineered CaM constructs, we analyzed all seven proteins via intact LC-MS. As with yARF-GFP and Fyn-GFP, we observed near-quantitative labeling of all proteins displaying an NMT recognition sequence. No myristoylated protein species were detected for any samples. Across multiple expressions and purifications, yARF-CaM was the only protein for which greater than 10% unlabeled species was regularly observed; it is possible that this result is somehow correlated to the very low expression yields for this particular construct.

Table III-2. Intact LC-MS results for WT CaM and 12-ADA-labeled CaN and CaM constructs. All expected masses account for removal of the initial Met, excluding the WT CaM mass. "+12-ADA (red.)" entries correspond to labeled proteins on which the azide group was reduced to an amine. "% Labeled" = sum of the relative abundances of both 12-ADA-labeled species observed for each protein.

Protein	Expected Mass (Da)	Observed Mass (Da)	Relative Abundance	% Labeled
Human CaN, B subunit	19,168.72	N/D	N/A	> 98 %
+12-ADA	19,392.03	19,392.12	0.59	
+12-ADA (red.)	*19,366.03*	*19,363.70*	*0.41*	
WT CaM	16,679.80	16,677.38	1.00	N/A
yARF-CaM	17,283.53	17,281.28	0.12	88%
+12-ADA	17,506.84	17,504.78	0.75	
+12-ADA (red.)	*17,480.84*	*17,478.84*	*0.13*	
yARF-Linker-CaM	17,968.34	N/D	N/A	> 98 %
+12-ADA	18,191.65	18,188.88	0.75	
+12-ADA (red.)	*18,165.65*	*18,165.27*	*0.25*	
yARF-6xHis-CaM	18,106.30	N/D	N/A	> 98 %
+12-ADA	18,329.68	18,327.36	0.72	
+12-ADA (red.)	*18,303.68*	*18,302.44*	*0.28*	
hCaNB-CaM	17,511.71	N/D	N/A	> 98 %
+12-ADA	17,735.02	17,731.59	0.45	
+12-ADA (red.)	*17,709.02*	*17,706.47*	*0.55*	
hCaNB-Linker-CaM	18,196.52	18,193.50	0.07	93 %
+12-ADA	18,419.83	18,416.77	0.48	
+12-ADA (red.)	*18,393.83*	*18,391.58*	*0.45*	

An interesting pattern in the intact LC-MS results for the 12-ADA-labeled CaN and CaM constructs is the abundance of the reduced 12-ADA-labeled species across samples. We investigated the cause of the azide reduction, examining the steps in our experimental protocol that could be responsible. We wondered if the reduction was caused by the presence of DTT in the purification buffers, but the use of buffers lacking DTT did not decrease the extent of azide reduction (data not shown). We also hypothesized that 12-ADA might be reduced in the cell or growth media during the

expression itself. Thus, we conducted a time-course study with hCaNB-CaM in which 12-ADA was added to the expression culture 2 hr or 1 hr prior to inducing protein expression, or at the time of induction. We found that there was no time-dependent trend in the extent of 12-ADA reduction (Table III-3), indicating that reduction was unlikely to be occurring in the expression flask. In related work, a sample of pure 12-ADA (in which the azide group was confirmed to be intact) was analyzed on the same instrument used for intact LC-MS analysis of proteins. The major species detected was the reduced form of 12-ADA (data not shown), underscoring the notion that azide reduction most likely occurred during the actual LC-MS run.

Table III-3. Intact LC-MS results for hCaNB-CaM samples purified from expression cultures exposed to 12-ADA for different lengths of time. 12-ADA was added 2 hr or 1 hr prior to inducing protein expression, or at the time of induction (0 hr). No unlabeled hCaNB-CaM was detected in any sample, and no trend was observed for the extent of reduction of the azide moiety with respect to time. N/D = Not Detected. N/A = Not Applicable.

Time Point	Protein Species	Expected Mass (Da)	Observed Mass (Da)	Relative Abundance	% Labeled	% Reduced
2 hr	hCaNB-CaM +12-ADA *+12-ADA (red.)*	17,511.71 17,735.02 *17,709.02*	N/D 17,732.40 *17,706.87*	N/A 0.64 *0.36*	> 98 %	36%
1 hr	hCaNB-CaM +12-ADA *+12-ADA (red.)*	17,511.71 17,735.02 *17,709.02*	N/D 17,732.41 *17,706.90*	N/A 0.59 *0.41*	> 98 %	41%
0 hr	hCaNB-CaM +12-ADA *+12-ADA (red.)*	17,511.71 17,735.02 *17,709.02*	N/D 17,732.68 *17,707.13*	N/A 0.61 *0.39*	> 98 %	39%

Evaluation of CaN and Engineered CaM Constructs via Phosphatase Activity Assays

Next, we investigated the potential impact of protein engineering and labeling on the activity of our various constructs. To address this issue, we utilized a phosphatase activity assay that tests the ability of CaN to dephosphorylate a phosphopeptide substrate derived from one of its natural substrate proteins. The resultant free phosphate reacts colorimetrically with a malachite green reagent, producing a change in absorption that may be measured on a standard plate reader.[18,19] Dephosphorylation depends on proper functioning of CaN, which in turn must be bound and activated by a functional form of CaM. Moreover, both CaN and CaM are fully active only in their Ca^{2+}-bound form, a fact that is important for the assays described here and especially pertinent for the assays described in the next section.

For all of the assay results depicted in Figure III-3, the concentration of CaN was held constant, while the concentration of CaM was varied over roughly four orders of magnitude. Ca^{2+} was also present at a saturating level. After CaN, CaM, and Ca^{2+} were equilibrated, the phosphopeptide substrate was added to the solution to initiate the enzymatic reaction; the reaction was quenched after 10 minutes via addition of malachite green.

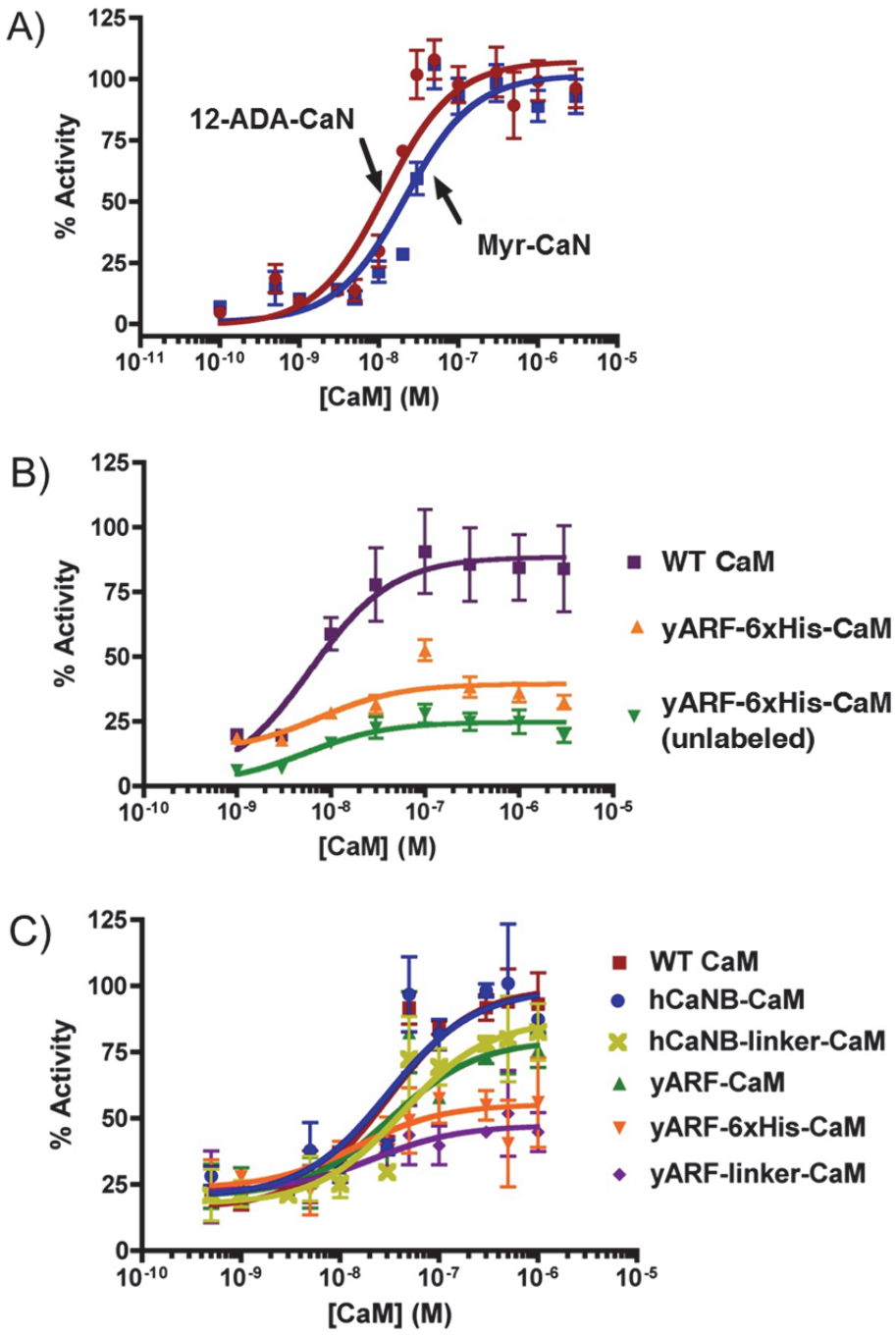


Figure III-3. Free phosphate generated upon incubation of CaN, CaM, and Ca^{2+} with RII phosphopeptide was measured in a standard Biomol Green (malachite green) phosphatase activity assay. (A) Myr-CaN or 12-ADA-CaN was incubated with saturating Ca^{2+} and varying concentrations of WT CaM. (B) CaN was incubated with saturating Ca^{2+} and varying concentrations of WT CaM, 12-ADA-labeled yARF-6xHis-CaM, or unlabeled yARF-6xHis CaM. Results in (A) and (B)

indicate that 12-ADA has little effect on CaN or CaM activity. (C) CaN was incubated with saturating Ca^{2+} and varying concentrations of WT CaM or a 12-ADA-labeled engineered CaM protein. For (A), results are presented as % activity of Myr-CaN. For (B) and (C), results are presented as % activity of WT CaM. For all graphs, $n \geq 4$.

The first graph above indicates that myristoylated CaN (Myr-CaN) and 12-ADA-CaN behave identically (Figure III-3A); we did not expect there to be much difference between these protein species because CaN is naturally myristoylated, and the azide moiety is largely inert. For our initial studies with CaM, we examined only yARF-6xHis-CaM (Figure III-3B). We found that addition of the yARF recognition sequence and the 6xHis tag resulted in a four-fold loss of activity relative to WT CaM, though the 12-ADA label did not appear to impact CaM activity. These data motivated us to prepare more engineered CaM constructs for NMT labeling, as described earlier.

After expressing and purifying the other members of the family of engineered CaM constructs, we tested them in the same phosphatase activity assay and were pleased to find that one construct, hCaNB-CaM, was as active as WT CaM (Figure III-3C). The ability to functionalize a CaM construct in a site-specific manner without an accompanying loss of activity should enable researchers to study CaM in a variety of settings, such as in single-molecule fluorescence experiments or on protein microarrays. Considering the ubiquity of CaM in eukaryotes, and given the large number of proteins that CaM binds and activates, the hCaNB-CaM construct has the potential to be useful for researchers in a number of fields.

Finally, in analyzing the data presented in Figure III-C, we were also able to answer a question posed near the beginning of this chapter: is the yARF sequence or the

6xHis tag responsible for the diminished activity of yARF-6xHis-CaM? The yARF-CaM construct, which lacks a linker or affinity tag, is 25% less active than WT CaM. Addition of the linker sequence or a 6xHis tag further decreases the activity of the corresponding constructs relative to the yARF-CaM and hCaNB-CaM parent constructs. Thus, it seems that both the yARF sequence and the 6xHis tag contribute to the poor activity of yARF-6xHis-CaM. With a different recognition sequence and no linker or 6xHis tag, and with wild-type levels of activity, hCaNB-CaM is clearly the construct of choice for future studies.

Calcium-Binding Behavior of Engineered CaM Constructs

To obtain a better understanding of our engineered CaM constructs, we investigated their Ca^{2+}-binding behavior with two different methods: a conventional electrophoretic mobility assay ("gel shift assay") described in the literature to probe other CaM mutants,[9,20] and the phosphatase activity assay described in the previous section.

The electrophoretic mobility assay exploits the change in the apparent molecular weight of CaM during SDS-PAGE depending on the buffer environment. As noted above, CaM undergoes a significant conformational change upon binding Ca^{2+}, exposing a hydrophobic patch on the protein. This structural change results in CaM running to a lower molecular weight in the presence of saturating levels of Ca^{2+}. In contrast, when CaM is in the presence of Ca^{2+} chelators, it runs to a higher apparent molecular weight. By utilizing SDS-PAGE buffers containing high concentrations of Ca^{2+} or EDTA, we found that our engineered and 12-ADA-labeled CaM constructs underwent a shift similar

to that of WT CaM: all six proteins appear to bind Ca^{2+} similarly at saturating levels of Ca^{2+} (Figure III-4A).

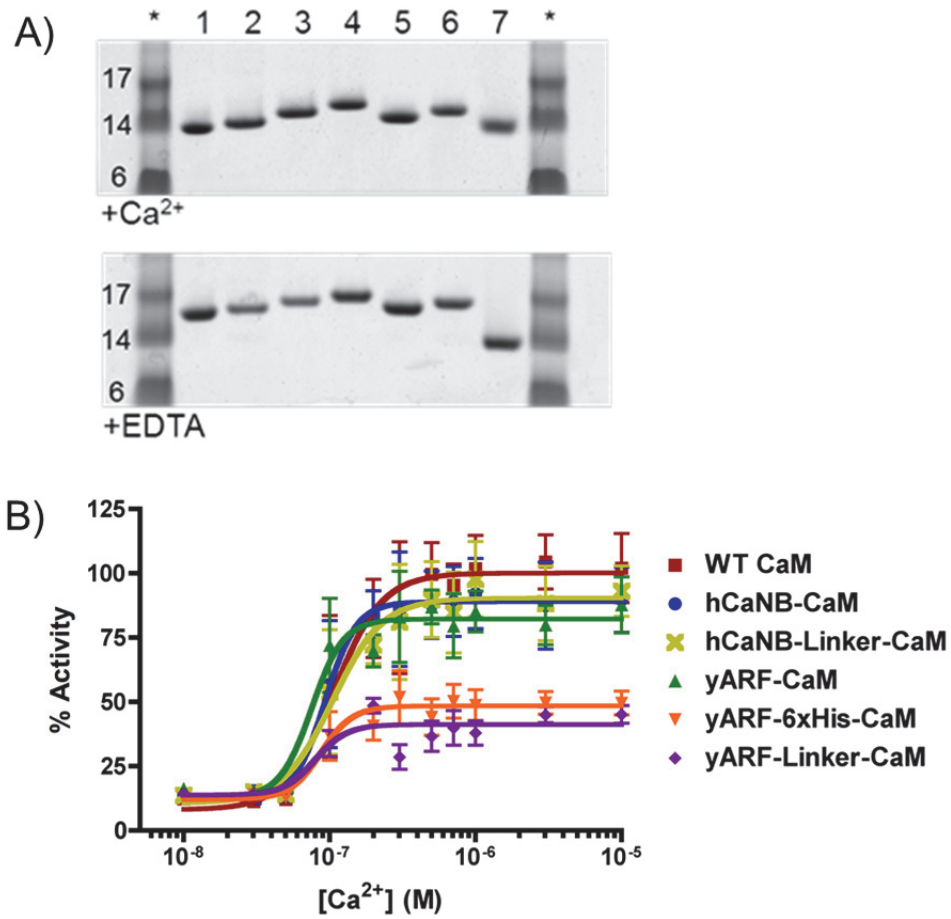


Figure III-4. (A) WT CaM and 12-ADA-labeled engineered CaM proteins were examined in an electrophoretic mobility ("gel shift") assay. SDS-PAGE buffers contained either Ca^{2+} (top) or EDTA (bottom). Coomassie staining revealed a Ca^{2+}-dependent shift in apparent mass for WT CaM (1), as well as for all 12-ADA-labeled engineered CaM proteins: yARF-CaM (2), yARF-Linker-CaM (3), yARF-6xHis-CaM (4), hCaNB-CaM (5), and hCaNB-Linker-CaM (6). Lysozyme (7) served as a negative control for Ca^{2+}-dependent mobility. Protein marker lanes are denoted by *. (B) CaN was incubated with saturating CaM (WT CaM or a 12-ADA-labeled engineered CaM protein) and varying concentrations of Ca^{2+}; then, free phosphate was detected in a standard Biomol Green (malachite green) activity assay. Results are presented as % activity of WT CaM; $n \geq 4$.

The phosphatase activity assay provided a more quantitative measure of the Ca^{2+}-binding behavior of our family of CaM constructs. The assay was carried out exactly as previously described, but instead of varying the concentration of CaM, we used a saturating level of CaM and varied the concentration of Ca^{2+} (Figure III-4B). Again, hCaNB-CaM was found to be the most active construct relative to WT CaM, with the other proteins exhibiting diminished activity in roughly the same order as was observed in the [CaM]-dependent phosphatase assay (Figure III-3C). All of the K_D values measured in our activity assay studies are summarized in Table III-4.

Table III-4. Binding constants for activity assay graphs (Figures III-3A, III-3C, and III-4B). Values are reported as ± standard error. ND = Not Determined.

Protein	[CaM]-Dependent Activity Assay: K_D (nM)	[Ca^{2+}]-Dependent Activity Assay: K_D (nM)
Myr-CaN	21 ± 12	-
12-ADA-CaN	12 ± 12	-
WT CaM	31 ± 14	107 ± 12
hCaNB-CaM	31 ± 14	93 ± 12
hCaNB-Linker-CaM	43 ± 16	105 ± 12
yARF-CaM	33 ± 17	75 ± 12
yARF-6xHis-CaM	ND	87 ± 12
yARF-Linker-CaM	ND	79 ± 13

Fluorescence Gel Analysis of Lysate Samples

Before proceeding to surface capture experiments with 12-ADA-CaN and 12-ADA-hCaNB-CaM, we wanted to confirm that NMT is selective toward our constructs in *E. coli*. In order to couple our proteins to microarrays directly from lysate,

it was important to establish that proteins other than our substrates were not labeled by NMT. Utilizing experimental protocols optimized for the GFP/NMT model system, we examined clarified lysate samples of CaN and the engineered CaM constructs after *in vivo* labeling with 12-ADA. For CaM, we focused on the two most active constructs: hCaNB-CaM and hCaNB-Linker-CaM. (Similar results were obtained for the remainder of the constructs, though yARF-CaM and yARF-Linker-CaM were difficult to detect in this experiment due to their low expression levels.) The gel images presented in Figure III-5 confirm that NMT is indeed selective toward CaN and the engineered CaM constructs in bacteria: fluorescent bands appear only at the molecular weight values for CaN-B and the engineered CaM proteins.

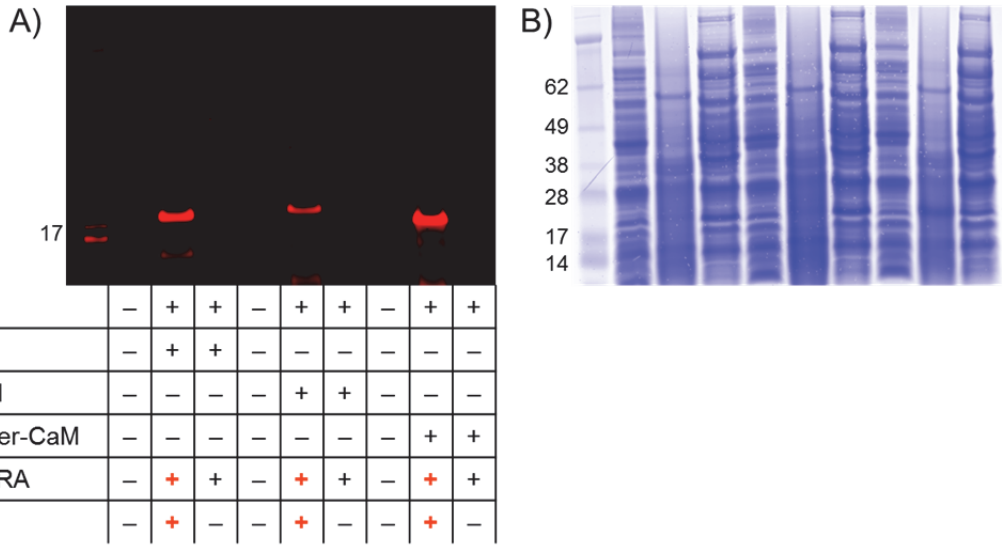


Figure III-5. SDS-PAGE analysis of lysate samples of CaN, hCaNB-CaM, and hCaNB-Linker-CaM co-expressed with NMT in the presence of 12-ADA. Lysate samples were treated with alkyne-TAMRA for detection of azide-labeled protein. The gel was imaged for TAMRA (A) and stained with Coomassie colloidal blue (B). Comparison of both gel images indicates selective 12-ADA labeling of each natural or engineered substrate protein. Similar results were obtained for other engineered CaM constructs.

CONCLUSION

The work presented in this chapter significantly expands the scope of the NMT-mediated protein labeling system developed in Chapter II. In moving beyond the test protein, GFP, we demonstrated the versatility and power of the system with proteins of real biomedical interest, calcineurin (CaN) and calmodulin (CaM). A natural substrate of NMT, CaN was shown to be equally active in its myristoylated and 12-ADA-labeled forms. Engineering of CaM to display different NMT recognition sequences also resulted in robust protein labeling with 12-ADA, as measured by intact LC-MS. Phosphatase activity assays investigating the behavior of the engineered CaM constructs showed that hCaNB-CaM was as active as WT CaM in both CaM-dependent and Ca^{2+}-dependent assays. Finally, treatment of lysate samples with an azide-reactive dye confirmed that only the CaN and CaM-based substrates were labeled with 12-ADA by NMT, enabling the preparation of protein microarrays from lysate, as described in the next chapter.

EXPERIMENTAL SECTION

Materials

Cloning. All oligonucleotide primers were ordered from IDT. The pET-15b plasmid encoding *Drosophila melanogaster* wild-type calmodulin was a gift from Professor Steven Mayo's lab at Caltech. The QuikChange site-directed mutagenesis kit was used as is from Stratagene/Agilent. Polymerase chain reaction (PCR) experiments were carried out in a BioRad DNA Engine Peltier Thermal Cycler using PfuTurbo DNA Polymerase (Stratagene/Agilent). All restriction enzymes, restriction enzyme buffers,

bovine serum albumin (BSA), and ligase were purchased from New England BioLabs (NEB). NEB DNA Ladders (100 bp and 1 kbp "Quick-Load") were used as markers for all DNA agarose gels, which were visualized with the addition of Plus One ethidium bromide solution from Amersham Biosciences on a UVP UV Transilluminator. Zymo Agarose-Dissolving Buffer (ADB) and Zymo Spin II columns, with their associated buffers, were used to purify DNA out of agarose gels. All DNA acquisition from cells was completed using the Qiagen Spin Miniprep Kit and columns. All sequencing requests were fulfilled by Laragen.

Protein expression. Plasmids encoding hNMT1 or hNMT2 and methionine-aminopeptidase (Met-AP) were a gift from the laboratory of Professor Richard Kahn at Emory University (Atlanta, GA).[21] The plasmid encoding human CaN^3 was purchased from Addgene. *E. coli* BL21(DE3) chemically competent cells were prepared using the standard Zymo method (Stratagene) and were transformed with either the hNMT1 plasmid or hNMT2 plasmid. LB medium was composed of 10 g tryptone (casein hydrolysate), 5 g yeast extract, and 10 g NaCl per liter. Media were autoclaved before use. Kanamycin (Kan) was used at a working concentration of 35 μg/mL, and ampicillin (Amp) was used at a working concentration of 200 μg/mL. Myristic acid was purchased from Fluka. All optical density (OD) values were measured at 600 nm on a Cary UV-Vis spectrophotometer. All SDS-PAGE gels described in this chapter were NuPAGE Novex 4%–12% Bis-Tris pre-cast gels (Invitrogen). SeeBlue Plus2 Pre-Stained Protein Marker from Invitrogen served as the molecular weight ladder. Gels were stained with Coomassie colloidal blue from Invitrogen.

Protein purification. CaN was purified with Talon cobalt affinity resin (Clontech) and calmodulin-sepharose 4B resin (GE Healthcare); Nickel-NTA resin from Qiagen could be used in place of Talon resin, if desired. All lysis, wash, and elution buffers were prepared exactly as reported.[3] Lysozyme was purchased from Aldrich.

CaM and engineered CaM constructs were purified using phenyl sepharose resin from GE Healthcare. (The yARF-6xHis-CaM protein was also purified using Qiagen Ni-NTA resin, in a manner identical to that described in Chapter II for yARF-GFP and Fyn-

GFP.) For CaM purification, the Lysis Buffer was 50 mM Tris (pH 7.5), 100 mM KCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mg/mL lysozyme, and 0.5 mM PMSF; Wash Buffer 1 was 50 mM Tris (pH 7.5) and 1 mM $CaCl_2$; Wash Buffer 2 was 50 mM Tris (pH 7.5), 1 mM $CaCl_2$, and 500 mM NaCl; and Elution Buffer was 50 mM Tris (pH 7.5) and 1.5 mM EGTA. Lysozyme was purchased from Aldrich.

Mass spectrometry analysis. The Pierce BCA Assay Kit was used to measure protein concentration in pure protein fractions prior to MS analysis. Millipore Microcon Centrifugal Devices were used to concentrate and buffer-exchange whole-protein samples for intact LC-MS analysis, which was carried out on an Agilent 1100 MSD quadrupole ESI-MS.

Phosphatase activity assays. Human recombinant WT CaM, human recombinant WT myristoylated CaN, and Biomol Green (malachite green) reagent were purchased from Enzo Life Sciences. The phosphorylated CaN-specific substrate peptide (pRII peptide) was purchased from GenScript. All other reagents were reagent grade and purchased from Sigma-Aldrich. Assay Buffer was composed of 50 mM Tris (pH 7.5), 100 mM NaCl, 6 mM $MgCl_2$, and 0.5 mM DTT.

Electrophoretic mobility assay. Buffers for the electrophoretic mobility assay was prepared as described.[9,20] In summary, 5x Loading Buffer was 0.225 M Tris (pH 6.8) containing 50% glycerol, 5% SDS, and 0.05% bromophenol blue. To 1 mL of the Loading Buffer, either 2 μL of 1 M $CaCl_2$ was added for a final concentration of 2 mM $CaCl_2$ (for the $+Ca^{2+}$ gel), or 6 μL of 0.5 M EDTA was added for a final concentration of 3 mM EDTA (for the +EDTA gel). Running Buffer was MES running buffer (Boston BioProducts); to 1 L of the commercially available running buffer, which already contained 1 mM EDTA, either 3 mL of 1 M $CaCl_2$ was added to achieve a final effective concentration of 2 mM $CaCl_2$ (for the $+Ca^{2+}$ gel), or 4 mL of 0.5 M EDTA was added to achieve a final concentration of 3 mM EDTA (for the +EDTA gel). SDS-PAGE gels were NuPAGE Novex 4%–12% Bis-Tris pre-cast gels, and SeeBlue Plus2 Pre-Stained Protein Marker served as the molecular weight ladder (both from Invitrogen). Gels were stained with Coomassie colloidal blue, also from Invitrogen.

Fluorescence detection. Lysate samples were treated with the reagents and according to the protocols of the Click-IT Tetramethylrhodamine (TAMRA) Protein Analysis Detection Kit from Invitrogen. After reaction and precipitation, protein samples were run on Invitrogen NuPAGE Novex 4%–12% Bis-Tris pre-cast gels and imaged on a GE Typhoon laser scanner. Gels were stained with Coomassie colloidal blue from Invitrogen.

Methods

Cloning. The template plasmid for all engineered CaM constructs was pET-15b encoding *Drosophila melanogaster* wild-type CaM. Primers were designed to encode the amino acid sequences corresponding to the constructs outlined in Table III-1: the yARF recognition sequence (MGLFASK, from ATG GGT CTG TTC GCG TCT AAA), the hCaNB recognition sequence (MGNEASYPL, from ATG GGT AAC GAA GCG TCT TAC CCG CTG), a 6xHis tag, and/or the linker sequence (SRLIGSA, from TCT CGT CTG ATC GGT TCT GCT) at the 5′ end of the gene. The QuikChange site-directed mutagenesis kit was used in conjunction with published protocols for the "two-step PCR" method; this approach circumvents "the tendency of the perfectly complementary mutagenic primers to dimerize with each other, rather than anneal to the target sequence [in the parent plasmid]," and thus enables the addition of long insertions.[16] Briefly, for each construct, two single-primer PCR reactions were carried out to produce "hybrid" plasmids, comprised of one original (wild-type CaM) strand and one new (mutant) strand possessing the given sequence(s). Then, in a second PCR step, the two reactions from the first step were combined and more polymerase was added to the reaction mixture. After digestion with DpnI, the mixture was transformed into XL1-Blue competent cells and plated. Colonies were selected for inoculation of cultures from which DNA was isolated and submitted for sequencing. Each final construct was transformed into competent cells already harboring an NMT plasmid for co-expression experiments.

Protein expression. Overnight cultures were inoculated in LB supplemented with Kan and Amp and grown in an incubator-shaker (37°C, 250 rpm). The following

day, overnight cultures were diluted 1:50 into fresh LB supplemented with Kan and Amp for expression cultures, which ranged in volume from 5 mL to 100 mL. Cultures were grown in an incubator-shaker (37°C, 250 rpm), and protein expression was induced with IPTG (1 mM, from 1 M stock in water) when the OD_{600} value was between 0.8 and 1.1. Pre-induction samples (1 mL) were collected as needed. Myristic acid or the azide fatty acid 12-ADA (500 μM, from 500 mM stock in DMSO) was also added at the time of induction. After 3–4 hr of protein expression, cells were harvested via centrifugation (10 min x 10,000 *g*) and the final OD_{600} value was measured. Cell pellets were lysed according to the following formula, regardless of which lysis buffer was used: 50 μL lysis buffer per mL culture per OD_{600} unit. Crude lysates were centrifuged once more, and the supernatant (clarified lysate) was saved for further experiments.

Protein purification. Published protocols were followed with minor modifications for purification of CaN and CaM constructs.

For CaN, a three-step protocol is described to purify the enzyme from bacterial lysate.[3] Briefly, harvested cells were lysed using a probe sonicator and lysozyme (1 mg/mL). The three steps are an ammonium sulfate (high-salt) precipitation, a Talon cobalt affinity resin purification step, and CaM–sepharose chromatography. All buffers were prepared and used exactly as reported in the literature to yield pure human CaN. All purification fractions were analyzed via SDS-PAGE for detection of pure protein.

For CaM, a one-step phenyl sepharose purification is reported to isolate the protein from bacterial lysate.[17,22] When CaM is in its Ca^{2+}-bound form, a hydrophobic patch on the protein is exposed, and it binds the phenyl sepharose resin; CaM elutes from the resin in the presence of buffers containing Ca^{2+} chelators, which cause CaM to undergo a conformational change that hides its hydrophobic patch. Briefly, harvested cells were lysed in Lysis Buffer using a probe sonicator and lysozyme (1 mg/mL). Clarified lysate was incubated with phenyl sepharose at 4°C for 30 min to ensure binding between CaM and the resin. The slurry was poured into a column and washed alternately with Wash Buffers 1 and 2. Then elution fractions were collected upon addition of Elution Buffer to the column. All purification fractions were analyzed via SDS-PAGE for detection of pure protein.

Mass spectrometry analysis. For intact LC-MS experiments, solutions of pure protein were concentrated using Microcon columns (MWCO = 30 kDa) and buffer-exchanged into a 0.1% TFA (trifluoroacetic acid) solution. A final solution of 100 pmol protein in 100 μL was run on the MSD instrument.

Phosphatase activity assays: [CaM]-dependent assays. The Ca^{2+}/CaM-activated phosphatase activity of CaN was determined using the malachite green assay, a colorimetric technique employed for quantitatively measuring the amount of inorganic phosphate released by dephosphorylation of a CaN-specific peptide substrate, pRII. This assay takes advantage of the green color produced by the complex formed between malachite green, molybdate, and free phosphate (PO_4).[18,19]

Briefly, varying concentrations of WT CaM were incubated with saturating Ca^{2+} (10 μM $CaCl_2$) and 10 nM Myr-CaN or 12-ADA-CaN in Assay Buffer for 10 min at 37°C. To initiate the reaction, phosphorylated RII peptide substrate was added at a final concentration of 0.5 mM in 50 μL and allowed to react for 10 min. A standard curve of inorganic phosphate (PO_4) in Assay Buffer was made on each day of experiments. A quantity of 50 μL of Biomol Green reagent (containing malachite green and molybdate) was added to the standard curve and experimental samples to terminate the reaction, and color was allowed to develop for 30 min. The absorbance of all samples was measured at 620 nm on a Tecan 96-well plate reader. The absorbance values of the standard curve were plotted against the (known) concentrations of the PO_4 standards and fit to a second-order polynomial, from which the amount of CaN-mediated release of PO_4 in the samples was calculated. The CaN phosphatase activity was plotted as a function of CaM concentration using Prism (GraphPad Software). The dose-response of CaN activity at varying CaM concentrations was calculated according to the sigmoidal dose response (Equation (1)):

$$Y = \min + \frac{max - min}{1 + 10^{Log(X50 - X)}} \qquad (1)$$

where X is the concentration, Y is the response, min is the lower asymptote of the curve, max is the upper asymptote of the curve, and X50 is the x-coordinate of the inflection point (x, y). X50 represents the concentration at which CaN is half-maximally activated and is directly related to the ability of CaM to bind and activate CaN.

Phosphatase activity assays: [Ca^{2+}]-dependent assays. The ability of the 12-ADA-labeled engineered CaM proteins to bind and activate CaN was measured using the malachite green assay described above in order to determine the CaM constructs' Ca^{2+}-dependent activation. Briefly, 10 nM Myr-CaN, 1 μM wild-type CaM or engineered labeled CaM, and varying concentrations of free Ca^{2+} were incubated at 37°C for 10 min. The concentration of free Ca^{2+} was tightly controlled by titration of the calcium chelator EGTA.[23] The reaction was initiated by addition of phosphorylated RII peptide substrate at a final concentration of 0.5 mM in 50 μL. The reaction was allowed to proceed for 10 min at 37°C. A standard curve of inorganic phosphate (PO_4) in Assay Buffer was made on each day of experiments. Biomol Green reagent (50 μL) was added to the standard curve and experimental samples to terminate the reaction, and color was allowed to develop for 30 min. The absorbance of all samples was measured at 620 nm on a Tecan 96-well plate reader. The absorbance values of the standard curve were plotted against known concentrations of PO_4 and fit to a second-order polynomial, from which the amount of CaN-mediated release of PO_4 in the samples was interpolated. The CaM-mediated CaN phosphatase activity as a function of Ca^{2+} concentration was plotted using Prism (GraphPad Software). The dose-response of activity of the same was calculated according to Equation (1).

Electrophoretic mobility assay. These experiments were conducted according to literature protocols.[9,20] Two different gels were run: a +Ca^{2+} gel, in which samples were exposed to an environment containing saturating levels of Ca^{2+}, and a +EDTA gel, in which samples were exposed to an environment containing excess EDTA. Samples of pure wild-type CaM, engineered CaM proteins, and lysozyme were prepared at equal concentrations, and 10 μg of each protein was loaded on a protein gel using the Loading Buffers described in the Materials section. SDS-PAGE was performed using the Running Buffers described in the Materials section. Gels were stained with Coomassie colloidal blue and imaged on the Typhoon with the 633 nm laser serving as the excitation source (no filter).

Fluorescence detection. Cells were lysed with the buffer recommended in the instructions for the Invitrogen Click-iT kit (1% SDS, 50 mM Tris-HCl, pH 8.0) according to the following formula: 50 μL lysis buffer per mL culture per OD_{600} unit. Lysate samples were reacted with alkyne-TAMRA and other kit reagents according to the protocols supplied by Invitrogen; the only modification was the use of 15 μL of alkyne-TAMRA dye solution rather than 100 μL. At the conclusion of the 25-min reaction time, samples were precipitated following the methanol-chloroform precipitation protocol described in the same kit instructions; the only modification was the completion of one extra methanol wash of the protein pellet. For SDS-PAGE analysis, protein pellets were resuspended in a denaturing buffer (8 M urea, 100 mM NaH_2PO_4, and 10 mM Tris-Cl) and loaded on a NuPAGE Novex 4%–12% Bis-Tris pre-cast gel. To detect TAMRA signal on the Typhoon, the 532 nm laser served as the excitation source (filter set: 580 BP 30). Gels were stained with Coomassie colloidal blue, then imaged again, with the 633 nm laser now serving as the excitation source (no filter).

REFERENCES

1. Squire, L. R. *et al. Fundamental Neuroscience, 2nd Ed.* (Academic Press/Elsevier Science: San Diego, CA, 2003).

2. Towler, D., Gordon, J. I., Adams, S. P. & Glaser, L. The biology and enzymology of eukaryotic protein acylation. *Annu. Rev. Biochem.* **57**, 69–99 (1988).

3. Mondragon, A. *et al.* Overexpression and purification of human calcineurin alpha from Escherichia coli and assessment of catalytic functions of residues surrounding the binuclear metal center. *Biochemistry* **36**, 4934–42 (1997).

4. Miyakawa, T. *et al.* Conditional calcineurin knockout mice exhibit multiple abnormal behaviors related to schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8987–92 (2003).

5. Means, A. R., Tash, J. S. & Chafouleas, J. G. Physiological implications of the presence, distribution, and regulation of calmodulin in eukaryotic cells. *Physiol. Rev.* **62**, 1–39 (1982).

6. Xia, Z. & Storm, D. R. The role of calmodulin as a signal integrator for synaptic plasticity. *Nat. Rev. Neurosci.* **6**, 267–76 (2005).

7. Mori, M. X., Imai, Y., Itsuki, K. & Inoue, R. Quantitative measurement of Ca(2+)-dependent calmodulin-target binding by Fura-2 and CFP and YFP FRET imaging in living cells. *Biochemistry* **50**, 4685–96 (2011).

8. Bal, M., Zaika, O., Martin, P. & Shapiro, M. S. Calmodulin binding to M-type K+ channels assayed by TIRF/FRET in living cells. *J. Physiol.* **586**, 2307–20 (2008).

9. Li, C. J. *et al.* Dynamic redistribution of calmodulin in HeLa cells during cell division as revealed by a GFP-calmodulin fusion protein technique. *J. Cell Sci.* **112**, 1567–77 (1999).

10. Miyawaki, A. *et al.* Fluorescent indicators for Ca(2+) based on green fluorescent proteins and calmodulin. *Nature* **388**, 882–7 (1997).

11. Griesbeck, O., Baird, G. S., Campbell, R. E., Zacharias, D. A. & Tsien, R. Y. Reducing the environmental sensitivity of yellow fluorescent protein: mechanism and applications. *J. Biol. Chem.* **276**, 29188–94 (2001).

12. Erickson, M. G., Alseikhan, B. a, Peterson, B. Z. & Yue, D. T. Preassociation of calmodulin with voltage-gated Ca(2+) channels revealed by FRET in single living cells. *Neuron* **31**, 973–85 (2001).

13. Jenikova, G., Lao, U. L., Gao, D., Mulchandani, A. & Chen, W. Elastin-calmodulin scaffold for protein microarray fabrication. *Langmuir* **23**, 2277–9 (2007).

14. Moser, M. J., Flory, M. R. & Davis, T. N. Calmodulin localizes to the spindle pole body of Schizosaccharomyces pombe and performs an essential function in chromosome segregation. *J. Cell Sci.* **110**, 1805–12 (1997).

15. Nagai, T., Sawano, A., Park, E. S. & Miyawaki, A. Circularly permuted green fluorescent proteins engineered to sense Ca(2+). *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3197–202 (2001).

16. Wang, W. & Malcolm, B. A. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *Biotechniques* **26**, 680–2 (1999).

17. Gopalakrishna, R. & Anderson, W. B. Ca(2+)-induced hydrophobic site on calmodulin: application for purification of calmodulin by phenyl-sepharose affinity chromatography. *Biochem. Biophys. Res. Commun.* **104**, 830–6 (1982).

18. Martin, B., Pallen, C. J., Wang, J. H. & Graves, D. J. Use of fluorinated tyrosine phosphates to probe the substrate specificity of the low molecular weight phosphatase activity of calcineurin. *J. Biol. Chem.* **260**, 14932–7 (1985).

19. Harder, K. W. *et al.* Characterization and kinetic analysis of the intracellular domain of human protein tyrosine phosphatase beta (HPTP beta) using synthetic phosphopeptides. *Biochem. J.* **298**, 395–401 (1994).

20. Rhyner, J. A., Koller, M., Durussel-Gerber, I., Cox, J. A. & Strehler, E. E. Characterization of the human calmodulin-like protein expressed in Escherichia coli. *Biochemistry* **31**, 12826–32 (1992).

21. Van Valkenburgh, H. A. & Kahn, R. A. Coexpression of proteins with methionine aminopeptidase and/or N-myristoyltransferase in Escherichia coli to increase acylation and homogeneity of protein preparations. *Methods Enzymol.* **344**, 186–93 (2002).

22. Putkey, J. A., Slaughter, G. R. & Means, A. R. Bacterial expression and characterization of proteins derived from the chicken calmodulin cDNA and a calmodulin processed gene. *J. Biol. Chem.* **260**, 4704–12 (1985).

23. Bers, D. M. A simple method for the accurate determination of free [Ca] in Ca-EGTA solutions. *Am. J. Physiol. Cell Physiol.* **242**, C404–8 (1982).

CHAPTER IV

Surface Capture of N-Terminally Functionalized Proteins out of Lysate

ABSTRACT

In this chapter, we focus on the covalent conjugation of N-terminally labeled proteins to functionalized slides for the creation of protein microarrays. Specifically, 12-ADA-labeled yARF-GFP, CaN, and hCaNB-CaM, prepared via bacterial co-expression with NMT, were selectively coupled to cyclooctyne-spotted glass slides. Moreover, rapid surface capture was achieved directly out of lysate, without prior purification of the recombinantly expressed proteins. Our ability to prepare protein microarrays directly from cellular extracts exploits the orthogonality of NMT toward bacterial systems and the exquisite selectivity of NMT toward both natural and engineered substrate proteins. We also describe experiments completed with an azide dye and acetylene- or cyclooctyne-derivatized agarose beads, which informed our subsequent work with microarrays; these studies provided a comparison of the relative efficiency of reacting azides with terminal alkynes versus cyclic strained alkynes, and they shed light on how reaction efficiency is affected by the presence of lysate proteins.

The microarray studies described in this chapter were performed at Maven Biotechnologies in collaboration with Dr. Tamara Kinzer-Ursem. Maven has developed an instrument for protein measurements based on a technology called LFIRE, or Label-Free Internal Reflection Ellipsometry. Successful coupling of our 12-ADA-labeled proteins to slides was confirmed using the LFIRE instrument, which enables sensitive and high-throughput detection of changes in height on the surface of a slide. Our results in this area provide a strong foundation for future work evaluating the biochemical activity of CaM, CaN, and other labeled proteins in a microarray format.

INTRODUCTION

Immobilization of Proteins on Surfaces

Protein chips and microarrays have been utilized in a variety of contexts. They may serve as components in medical diagnostic systems,[1–3] and they may be useful research tools for *in vitro* studies of the interactions of proteins with small molecules and other proteins.[4–7] Elucidating such interactions is a key step toward discovering new small-molecule drugs that bind specific protein targets and dissecting intricate protein–protein networks that govern complex biological processes, such as memory formation, and disease states, such as oncogenesis.

As such, the development of techniques to couple proteins to surfaces has been an area of active research for the past twenty years, continuing to the present day.[8–10] In 2000, the Schreiber Lab published a landmark paper describing the immobilization of proteins on aldehyde slides via lysine (Lys) side-chain amines.[11] While this work represented a major advance at the time for high-throughput studies of protein-protein interactions, it also utilized a chemical reaction that is neither site-specific nor selective for a single protein: Lys residues are prevalent across the proteome and are often found at multiple sites within a single protein. The use of the Lys-aldehyde reaction and other non-specific protein chemistries yields microarrays that display the protein of interest in various orientations, depending on the site of surface attachment on the protein; only a fraction of those protein molecules exhibit a useful or active orientation, rendering most of the microarray useless in some cases. Thus, there has been a growing need for new chemistries offering greater selectivity within biological molecules to improve the utility of protein microarrays.

A major advance for protein chemistry occurred in 2001, when the Sharpless and Meldal Groups independently reported the selective copper-catalyzed reaction of azide and alkyne groups,[12,13] which was soon shown to be bioorthogonal (i.e., neither reactive partner interacts with chemical groups normally present in biological settings).[14] Numerous other bioorthogonal reactions have been established that enable selective and controlled chemical reactions of biomolecules.[15] Importantly, methods have been developed in concert to incorporate the appropriate chemical groups into proteins and other biomolecules, enabling their participation in bioorthogonal reactions.[16–20] One notable outcome of these interdisciplinary advances is that bioorthogonal reactions have now been utilized to achieve site-specific and selective immobilization of proteins on surfaces, as described in a recent review.[10] Clearly, much progress has been made since the Schreiber Lab's report of aldehyde–protein microarrays thirteen years ago.

Our objective for the work described here was to utilize NMT-mediated protein labeling as a step towards site-specific immobilization of functionalized proteins on surfaces for downstream applications. We developed methods using yARF-GFP, a protein from our original model system, and applied them to the CaN and hCaNB-CaM proteins described in Chapter III. Given that some surface immobilization techniques currently in use require purification of the protein of interest,[10] we were also intrigued by the possibility of coupling our recombinant proteins to surfaces directly from lysate. In general, the process of isolating a protein from lysate has three primary drawbacks: purification often requires considerable time and resources; it may result in a significant loss of net protein product; and it may cause adverse changes in protein structure and function. Thus, advancing methods for protein–surface coupling that do not require prior

protein purification is highly desirable. To that end, we describe our preliminary conjugation experiments with functionalized beads and a small-molecule dye, and then shift our focus to our work with protein microarrays.

RESULTS AND DISCUSSION

Experiments with Alkyne-Functionalized Agarose Beads

For all of our protein conjugation experiments, we planned to utilize the azide–alkyne cycloaddition reaction.[12,13,21] In order to gain a better understanding of how this reaction behaves when one partner is immobilized, we completed preliminary experiments with alkyne-functionalized agarose beads and a small-molecule azide dye. As shown in Figure IV-1, agarose beads displaying N-hydroxy succinimidyl (NHS) ester groups were reacted with amine-bearing alkyne reagents. Both acetylene (terminal alkyne) and azadibenzocyclooctyne (ADIBO) beads were prepared. Either ethanolamine or methyl-PEG4-amine was used to quench unreacted NHS groups; this quenching step is important for protein experiments, because NHS groups also react readily with lysine amine groups. Finally, beads were blocked with BSA or left unblocked. In total, eight types of alkyne-derivatized beads were prepared.

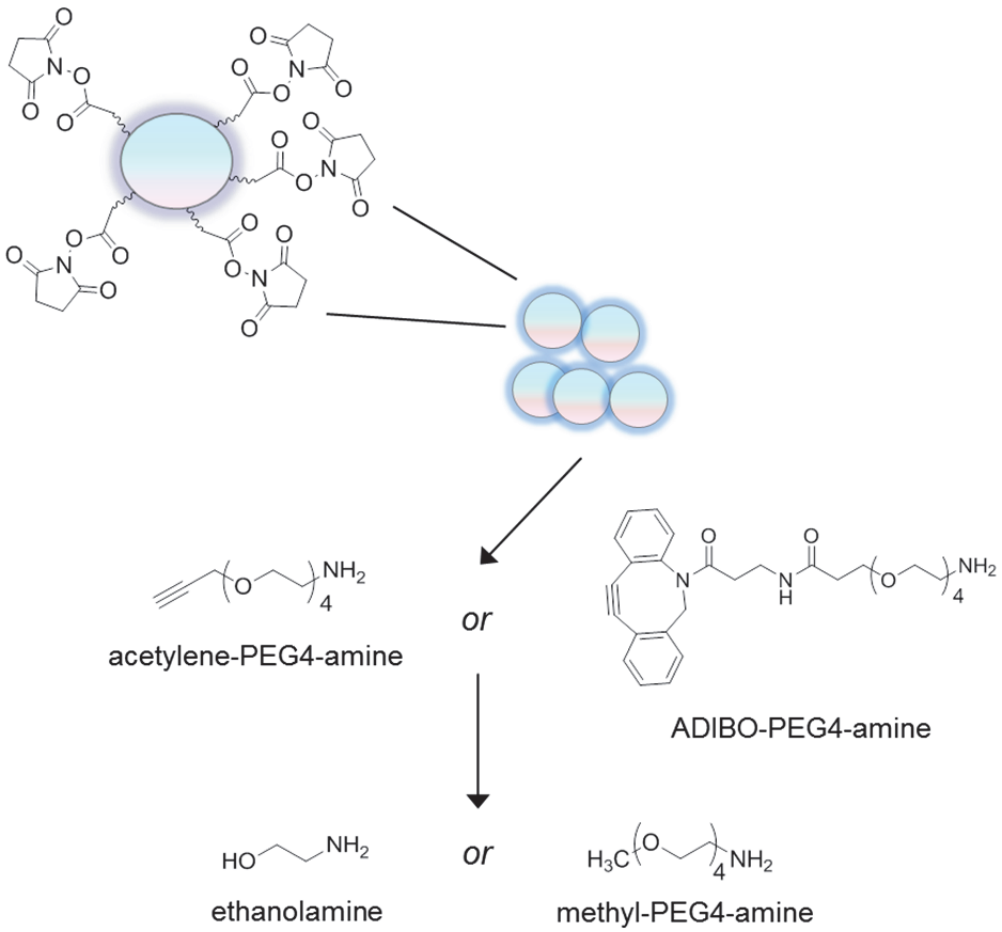


Figure IV-1. Schematic overview of functionalization of agarose beads. NHS-ester agarose beads were reacted with an acetylene or a cyclooctyne-bearing amine, followed by treatment with a quenching agent to ensure that no unreacted NHS-ester groups remained.

After completing the bead preparation protocol, we confirmed successful functionalization of the beads by reacting them with a small-molecule dye, Azide-Fluor488. We also wanted to determine whether or not the different quenching conditions and the BSA blocking step had any effect on azide-alkyne reaction efficiency. Either the copper-catalyzed or strain-promoted azide-alkyne reaction was carried out, depending on the bead type, with no-copper negative controls included for the acetylene

beads. Equivalent levels of fluorescence were observed across all four bead types for each type of alkyne (acetylene or ADIBO), indicating that functionalization was successful and that the type of quencher and the BSA blocking step did not affect reaction efficiency (data not shown). For subsequent experiments, we moved forward with both acetylene and ADIBO beads that had been quenched with methyl-PEG4-amine.

Next, we investigated how different bead preparation conditions affected the nonspecific binding of lysate protein to the beads. We hypothesized that the BSA blocking step would decrease nonspecific interaction of lysate proteins with the beads. We also expected to observe a higher background signal for ADIBO beads than for acetylene beads because the thiol groups in cysteine side chains are known to react with activated cyclooctynes.[22,23] To test these hypotheses, we reacted NHS-AlexaFluor633 with lysate protein collected from *E. coli* BL21(DE3) cells in which no protein expression had been induced, and we incubated beads with the dyed lysate. The fluorescence of the beads was measured before and after washing. The results, presented in Figure IV-2, indicate that the BSA blocking step did not have an impact on the interaction of our beads with lysate protein. However, ADIBO beads did indeed show a four-fold higher lysate background signal, even after washing, as compared to acetylene beads: it is likely that some lysate proteins react covalently with the ADIBO beads.

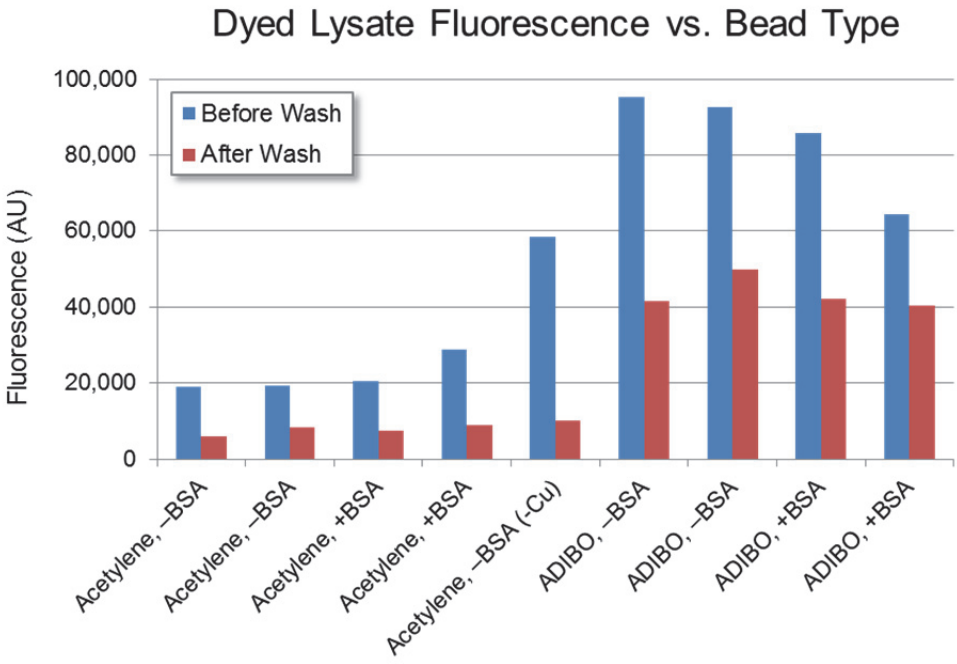


Figure IV-2. Fluorescence of beads after incubation with AlexaFluor633-dyed lysate, before/after washing. Duplicate pairs are shown for each bead type, except the no-Cu negative control. Beads were functionalized with acetylene-PEG4-amine or ADIBO-PEG4-amine, quenched with methoxy-PEG-amine, and blocked with BSA ("+BSA") or not blocked ("–BSA"). All fluorescence measurements were normalized by number of beads. After washing, roughly four times as much lysate remained on ADIBO beads as compared to acetylene beads.

We were also interested in how the presence of lysate protein affected the reaction efficiency between the azido dye and the alkynyl beads, an important consideration for achieving successful capture of 12-ADA-labeled proteins from lysate on alkyne-derivatized surfaces. Acetylene and ADIBO beads were reacted with Azide-Fluor488 in the absence (Figure IV-3A) or presence (Figure IV-3B) of lysate protein. We measured a two-fold lower fluorescence signal for beads that had been reacted with the azido dye in

the presence of lysate, for both acetylene and ADIBO beads. Again, very consistent results were obtained within each bead type, regardless of whether or not the beads had been blocked with BSA. These results were considered in the design of subsequent microarray experiments.

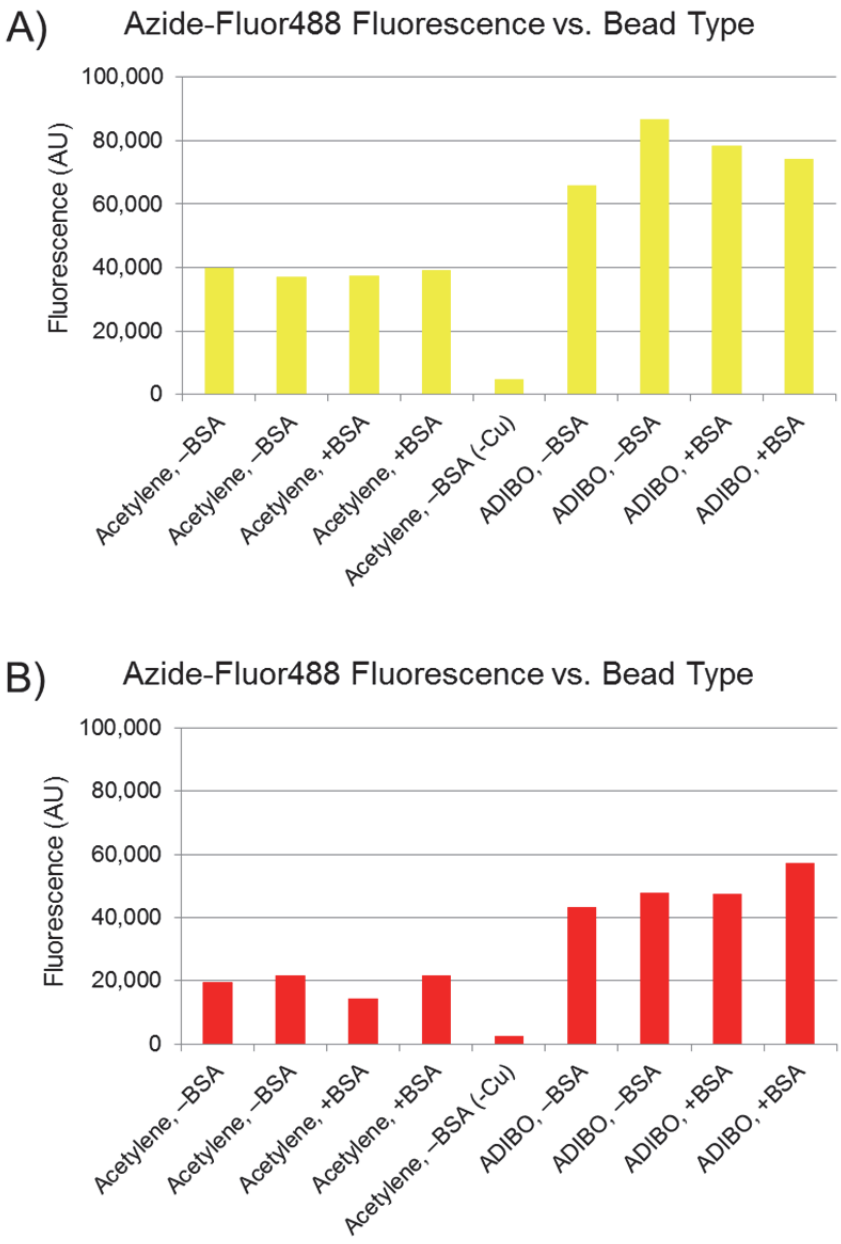


Figure IV-3. Fluorescence of beads after reaction with Azide-Fluor488 in the absence (A) or presence (B) of lysate. Duplicate pairs are shown for each bead type, except the no-Cu negative controls. Beads were functionalized with acetylene-PEG4-amine or ADIBO-PEG4-amine, quenched with methoxy-PEG-amine, blocked with BSA ("+BSA") or not blocked ("–BSA"), and washed, prior to reaction with Azide-Fluor488. All fluorescence measurements were normalized by number of beads. The azide-alkyne reaction yield in lysate is approximately half the reaction yield in buffer only.

Examination of the graphs in Figure IV-3 reveals a clear difference in average signal for ADIBO beads versus acetylene beads after reaction with Azide-Fluor488. Strained alkynes such as ADIBO are known to react with azides more rapidly and to a greater extent than terminal alkynes do.[24] However, as noted earlier and as evidenced by Figure IV-2, strained alkynes can also react nonspecifically with nucleophiles present in biological systems, particularly the thiol groups of Cys residues.[23] Thus, there is a trade-off between signal and background that is worth considering when selecting or designing an alkynyl molecule for reaction with azides, depending on the context of a given cycloaddition reaction. For our purposes, we found that the copper catalyst required for reaction of terminal alkynes with azides strongly interfered with activation of CaN by CaM (data not shown); both proteins possess multiple metal-binding centers, as discussed in Chapter III, so it is likely that either or both proteins bound the Cu^{2+} and Cu^{1+} ions yielded by the copper catalyst. Thus, we utilized cyclooctyne compounds exclusively when working with CaM and CaN, as described in the next section.

Overview of LFIRE Instrumentation and Experimental Set-up

In progressing from the use of an azido dye to the use of azide-labeled proteins, we also moved to a higher-throughput format than beads: protein microarrays. More specifically, a robotic spotter was used to print cyclooctyne molecules on glass slides at a density of approximately 16 spots per square mm, with a distance of 300 μm between spots. Glass slides were treated with a proprietary optical coating enabling their use in the Label-Free Internal Reflection Ellipsometry instrument, or LFIRE, developed by Maven Biotechnologies (Figure IV-4A).[25] The protocols for preparing amine-coated

glass slides and printing NHS-cyclooctyne molecules on the slides are described in more detail in the Experimental Section of this chapter.

Our LFIRE microarray experiments began with affixing a well structure to the printed slide, mounting the slide on a glass prism, and inserting the prism into the instrument (Figure IV-4B). The slide was kept hydrated and remained mounted on the prism throughout the experiment. Buffer was added to the wells, followed by a BSA blocking solution. LFIRE measurements were collected during this time to establish a consistent baseline. After the slide was thoroughly washed, cell lysate was added to each well, with simultaneous initiation of further LFIRE measurements. An image of the entire slide was collected with a predetermined frequency, depending on the size of the total printed area; for our experiments, an image was collected approximately once every 90 seconds. Data analysis entailed stacking all of the images in a software program such as ImageJ, subtracting a background image, and detecting localized rises in signal corresponding to protein deposition on the slide surface. The LFIRE experimental protocol is depicted schematically in Figure IV-4B, and the following sections describe our microarray work with 12-ADA-labeled yARF-GFP, CaN, and hCaNB-CaM.

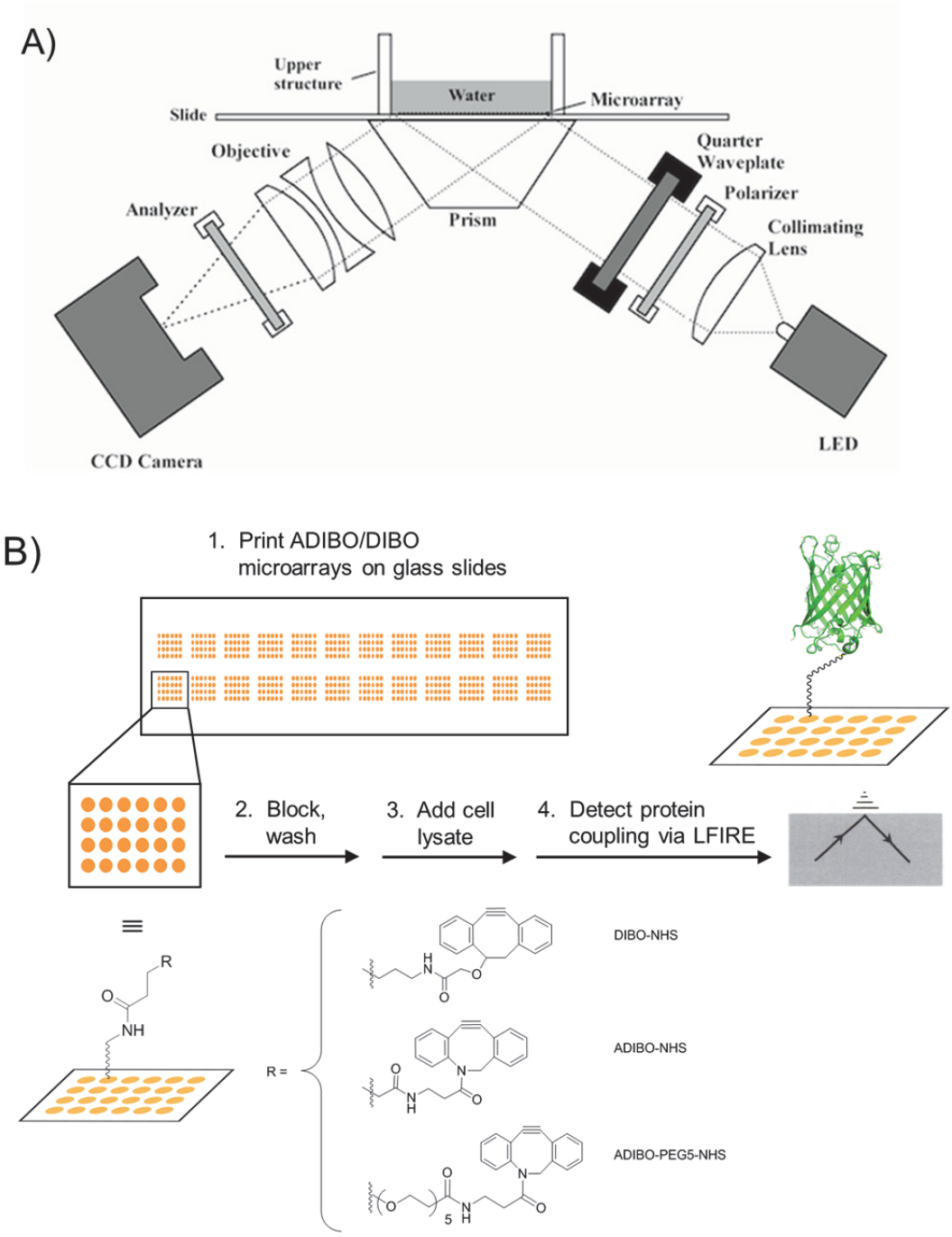


Figure IV-4. (A) Schematic overview of LFIRE instrument (adapted from Reference 25). Polarized light aids in the detection of changes in height, i.e., protein deposition, on the microarray surface. (B) Schematic overview of surface coupling experiments. Slides are printed with cyclooctyne spots, blocked, incubated with cell lysate, and imaged on the LFIRE.

Surface Capture and LFIRE Analysis of yARF-GFP Lysates

Lysates were prepared from *E. coli* cells in which yARF-GFP was expressed in the presence of no fatty acid, myristic acid, or 12-ADA; additionally, a negative control lysate was prepared from cells in which no test protein had been overexpressed. The protein concentration of each of the four lysate samples was measured, so that equal concentrations of lysate protein could be added to different wells on the microarray slide. For these experiments, we used slides that had been printed with DIBO-NHS or ADIBO-NHS (structures shown in Figure IV-4B). As described above, LFIRE data were collected throughout the duration of the blocking and lysate-coupling steps. Background-subtracted data are presented in Figures IV-5 and IV-6.

Analysis of the LFIRE data clearly indicated that significant coupling occurred only in the areas where DIBO or ADIBO was spotted, and only in wells to which lysate containing 12-ADA-yARF-GFP was added; the images shown in Figure IV-5 are representative of results obtained across multiple wells. Furthermore, the similarity in appearance of the control wells (no test protein, yARF-GFP, and Myr-yARF-GFP) indicates that the majority of background protein deposition is unlikely to be yARF-GFP, but rather, is likely the result of nonspecific reaction between cysteine-containing lysate proteins and DIBO or ADIBO. Finally, the large difference in appearance between Myr-yARF-GFP wells and 12-ADA-yARF-GFP wells demonstrates that the strong signal measured for the latter is not simply due to hydrophobic attraction between the fatty acid tag and the slide. Post-wash LFIRE data from multiple wells exposed to the same four lysate samples were analyzed quantitatively; those results are presented in Figure IV-6 and confirm that selective coupling is achieved with 12-ADA-yARF-GFP out of lysate.

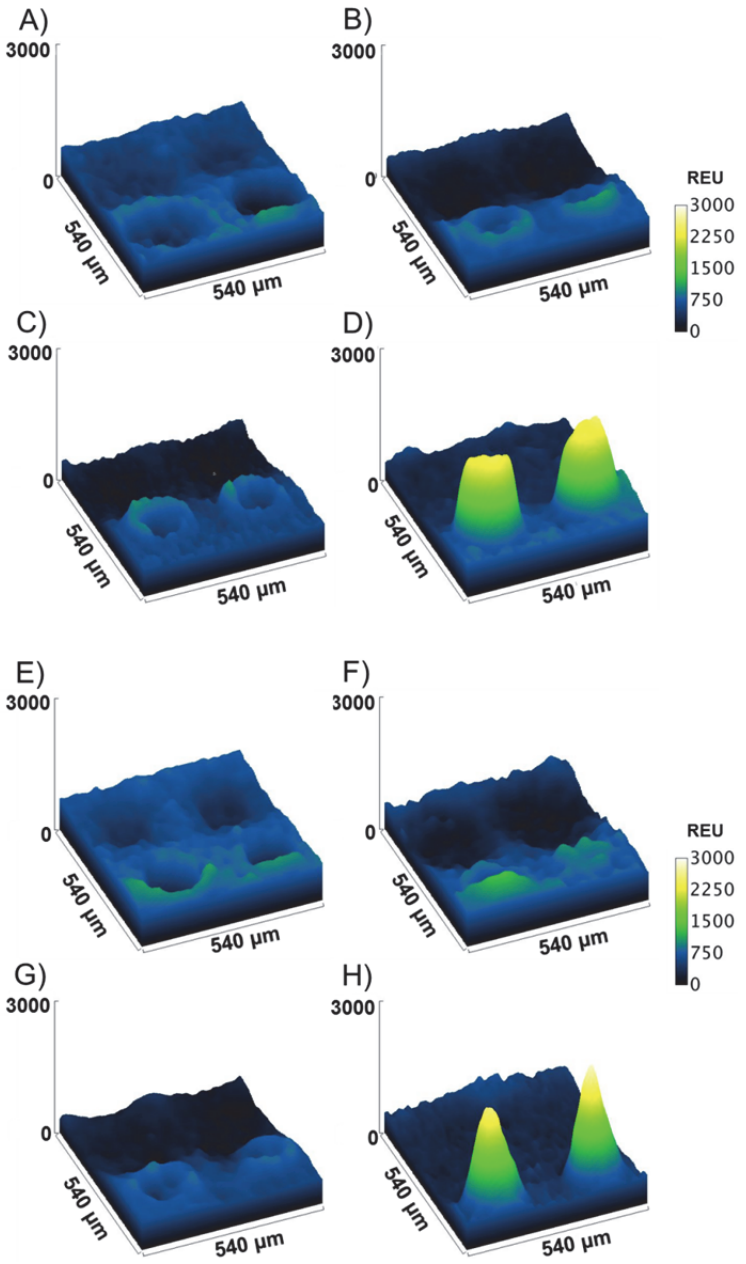


Figure IV-5. LFIRE 3-D surface plots of microarrays after incubation with cell lysate containing no overexpressed protein (A, E); yARF-GFP (B, F); Myr-yARF-GFP (C, G); or 12-ADA-yARF-GFP (D, H). The two microarray spots in back are BSA for all panels, and the two spots in front are DIBO (A–D) or ADIBO (E–H). Significant protein coupling was observed only in the presence of 12-ADA-labeled yARF-GFP lysate, and only within cyclooctyne-derivatized areas.

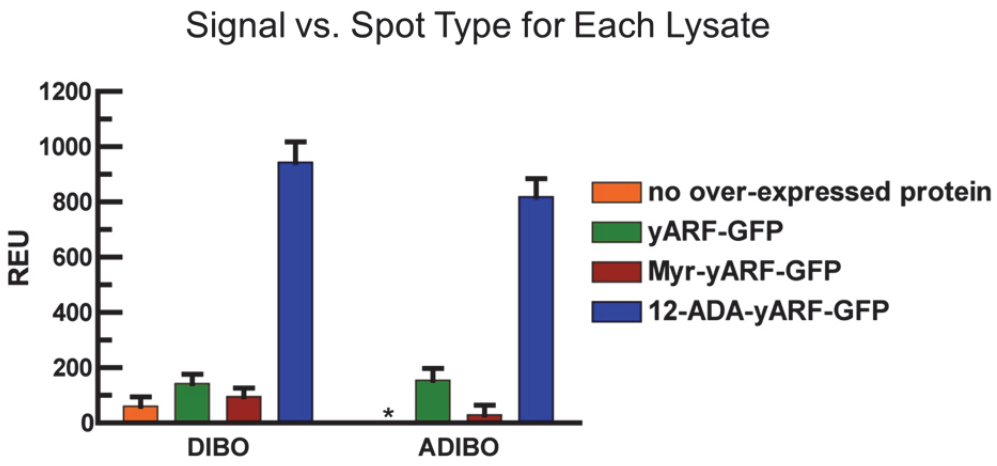


Figure IV-6. Average post-wash LFIRE signal for DIBO and ADIBO microarray spots after exposure to different lysate samples. For both DIBO and ADIBO slides, significant coupling was observed only in the presence of 12-ADA-labeled yARF-GFP. Each bar represents the average of $\geq$ 16 spots, $\pm$ standard error of the mean. * = no detectable signal.

Surface Capture and LFIRE Analysis of CaN Lysates

LFIRE experiments were performed with lysates containing Myr-CaN or 12-ADA-CaN in a manner similar to that described above for the yARF-GFP lysates. (Controls with no protein or unlabeled protein were not performed; attempts to express CaN in a non-myristoylated/non-labeled form have been shown to result in significant protein aggregation.[26]) Again, results indicated strong coupling only between cyclooctyne spots and 12-ADA-labeled protein (Figure IV-7). For these experiments, data analysis was performed before and after a final wash step. The strong post-wash signal observed in Figure IV-7B for 12-ADA-CaN, and its similarity to the pre-wash signal in Figure IV-7A, suggests that 12-ADA-CaN is covalently conjugated to the slide.

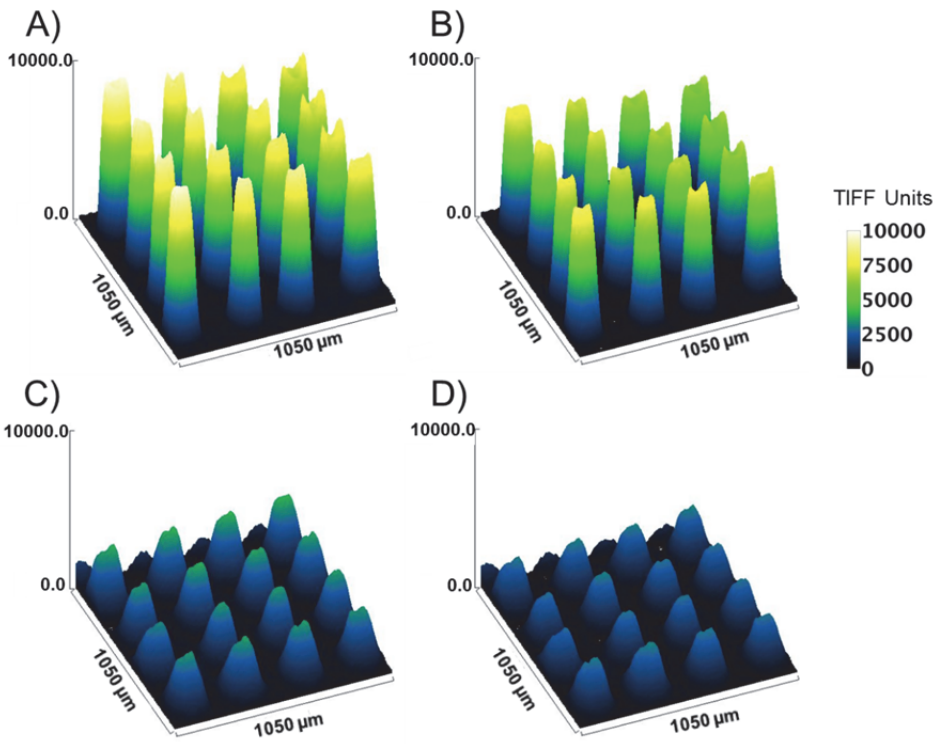


Figure IV-7. LFIRE 3-D surface plots of ADIBO-PEG5-NHS microarrays after incubation with cell lysate containing 12-ADA-CaN (A, B) or Myr-CaN (C, D). For all panels, the 16 spots shown are a representative subset from a larger array of 289 spots. (A) and (C) display signal immediately after incubation, prior to washing, while (B) and (D) display post-wash signal intensity. Significant protein coupling was observed only in the presence of 12-ADA-CaN, only within cyclooctyne-derivatized areas.

Further analysis of the CaN LFIRE data set was performed, as presented in Figure IV-8. Plotting the signal as a function of time provided insight into the relative efficiency of the azide-cyclooctyne reaction versus background reactions between cyclooctynes and other molecules present in lysate (Figure IV-8A). In particular, the data for the first 15 minutes of the incubation period indicate that the azide-cyclooctyne reaction proceeds considerably faster than background reactions in our system. Considering the significant

interest in and widespread use of the biocompatible azide-cyclooctyne reaction,[15,27] we believe these results could be of interest to the larger chemistry/chemical biology community.

We were also interested in gaining a better understanding of the enrichment factor achieved via protein coupling—that is, the amount of 12-ADA-labeled protein present on the slide versus present in lysate. As shown in Figure IV-8, the final signal of 12-ADA-CaN is roughly three times that of Myr-CaN: this result indicates that two-thirds of the 12-ADA-CaN signal corresponds to the labeled species, while one-third is background. Quantitative Western blotting indicated that CaN constitutes roughly 7% of the total lysate protein for both Myr-CaN and 12-ADA CaN (data not shown). Comparison of the ~7% figure in lysate (i.e., background:labeled = 93:7 = 13.3:1) with the ~66% figure after coupling (i.e., background:labeled = 34:66 = 0.51:1) yields a 26-fold reduction in contaminating proteins relative to 12-ADA-CaN upon coupling to cyclooctyne microarrays from lysate.

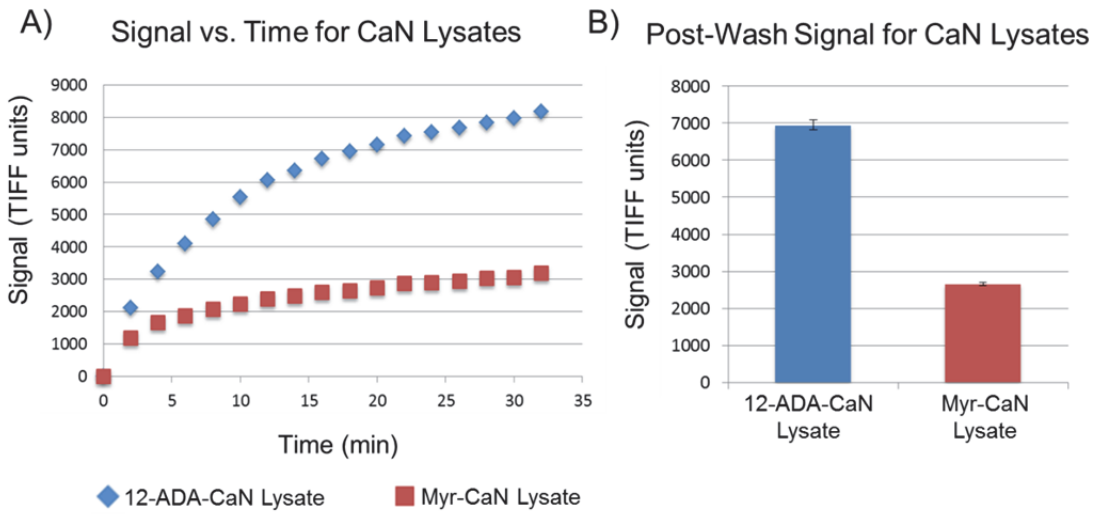


Figure IV-8. Quantitative analysis of CaN LFIRE surface plots. (A) Average signal for ADIBO microarray spots during incubation with lysate containing 12-ADA-CaN or Myr-CaN, prior to washing. Lysate was added at 0 min. (B) Average signal for ADIBO spots after lysate incubation and wash. For both panels, data points represent the average signal of the 16 spots shown in Figure IV-7. The results show that 12-ADA-CaN couples more quickly and more specifically to ADIBO spots as compared to the control protein, Myr-CaN. Error bars represent the standard deviation.

Surface Capture and LFIRE Analysis of hCaNB-CaM Lysates

Finally, LFIRE experiments similar to those outlined above were performed with lysates containing hCaNB-CaM expressed in the presence of myristic acid or 12-ADA; similar results were obtained as well. Lysate containing 12-ADA-labeled hCaNB-CaM (henceforth referred to as "12-ADA-CaM") coupled specifically to cyclooctyne spots, and the signal remained strong after washing (Figure IV-9, A and B). Lower signal intensities were detected for the negative control lysate containing myristoylated hCaNB-CaM (Myr-CaM) (Figure IV-9, C and D).

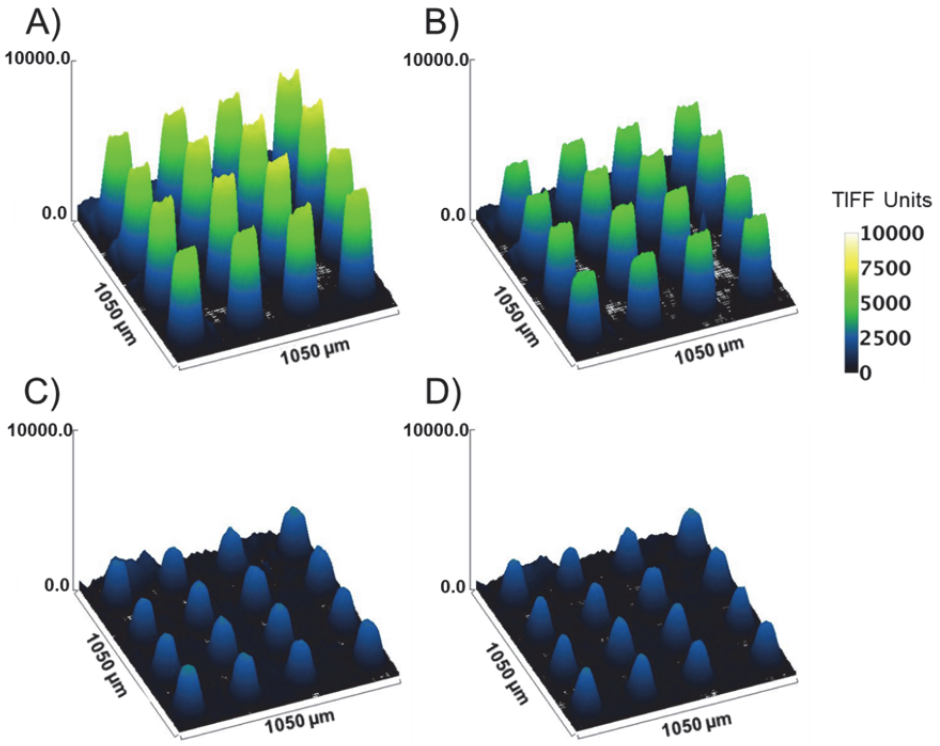


Figure IV-9. LFIRE 3-D surface plots of ADIBO-PEG5-NHS microarrays after incubation with cell lysate containing 12-ADA-CaM (A, B) or Myr-CaM (C, D). For all panels, the 16 spots shown are a representative subset from a larger array of 289 spots. (A) and (C) display signal immediately after incubation, prior to washing, while (B) and (D) display post-wash signal intensity. Significant protein coupling was observed only in the presence of 12-ADA-CaM, only within cyclooctyne-derivatized areas.

The representative images in Figure IV-9 indicate that the "signal-to-noise" ratio (final signal of 12-ADA-CaM versus final signal of Myr-CaM) is lower than the corresponding ratio for CaN, although the mass spectrometry results presented in Chapter III confirmed that both CaN and hCaNB-CaM are labeled quantitatively with 12-ADA. It is possible that the functionalized N-terminus of 12-ADA-CaM is less accessible for subsequent conjugation than that of 12-ADA-CaN; because CaN is a natural substrate of

NMT, it is likely that its native folded structure renders its N-terminus available for reaction. Further surface coupling studies with natural NMT substrate proteins and engineered substrates will elucidate the extent to which this theory is correct.

LFIRE data for the CaM lysates were also analyzed in a more quantitative manner (Figure IV-10). Again, we observed a more rapid reaction rate for the azide-cyclooctyne reaction than for background reactions, with the difference in reaction rates most apparent during the first 15 minutes (Figure IV-10A). As would be expected from the representative images shown in Figure IV-9, the average final post-wash signal of 12-ADA-CaM lysate is lower than that of 12-ADA-CaN lysate, though it is considerably higher than that of Myr-CaM lysate (Figure IV-10B). About 60% of the final signal for 12-ADA-CaM is the recombinant protein (i.e., background:labeled = 40:60 = 0.67:1). Quantitative Western blots of 12-ADA-CaM and Myr-CaM indicated that CaM constitutes ~13% of the total lysate protein (background:labeled = 87:13 = 6.7:1). Thus, we achieved a 10-fold reduction in contaminating proteins relative to 12-ADA-CaM upon coupling to cyclooctyne microarrays from lysate.

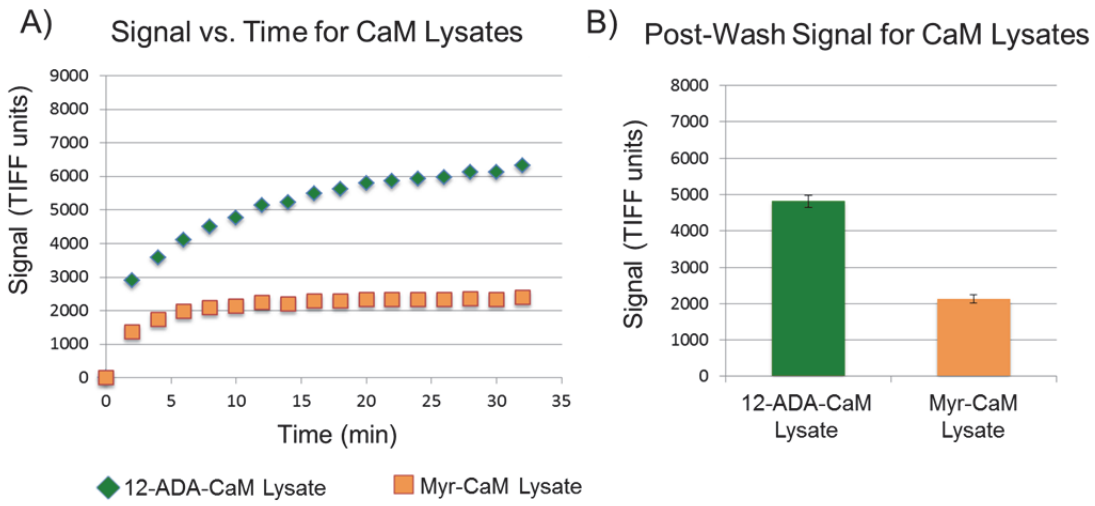


Figure IV-10. Quantitative analysis of CaM LFIRE surface plots. (A) Average signal for ADIBO microarray spots during incubation with lysate containing 12-ADA-CaM or Myr-CaM, prior to washing. Lysate was added at 0 min. (B) Average signal for ADIBO spots after lysate incubation and wash. For both panels, data points represent the average signal of the 16 spots shown in Figure IV-9. The results show that 12-ADA-CaM couples more quickly and more specifically to ADIBO spots as compared to the control protein, Myr-CaM. Error bars represent the standard deviation.

CONCLUSION

This chapter describes experiments completed with functionalized beads and microarrays, two different systems for surface capture of proteins. Using Maven Biotechnologies' LFIRE instrument, we detected selective coupling of 12-ADA-yARF-GFP, 12-ADA-CaN, and 12-ADA-hCaNB-CaM to cyclooctyne-spotted glass slides. The orthogonality exhibited by NMT toward bacterial proteins allowed us to couple 12-ADA-labeled proteins to derivatized surfaces directly out of lysate, a useful feature for studies

of proteins that are difficult to purify or that are negatively affected by isolation from the lysate environment. Experiments conducted with lysates containing no recombinant protein, unlabeled yARF-GFP, Myr-yARF-GFP, and 12-ADA-yARF-GFP resulted in significant surface coupling of 12-ADA-yARF-GFP only. The similar appearance of all three controls indicated that the coupling observed for 12-ADA-yARF-GFP was indeed specific to the azide moiety and was related to neither the protein sequence nor the presence of a fatty acid tag. Experiments conducted with 12-ADA-CaN and 12-ADA-CaM demonstrated that enrichment factors of 26 and 10, respectively, were achieved upon selective coupling of each protein from lysate to cyclooctyne arrays. These results provide a strong foundation for the completion of biochemical studies with CaN and CaM microarrays, as well as a set of methods for the surface capture of other proteins.

EXPERIMENTAL SECTION

Materials

Experiments with alkyne-derivatized agarose beads. NHS-ester agarose resin was purchased from GE Healthcare. Azide-Fluor488, Acetylene-PEG4-amine, and ADIBO-PEG4-amine were purchased from Click Chemistry Tools. Methyl-PEG4-amine was purchased from Pierce. Ethanolamine and fluorescein salt were purchased from Sigma-Aldrich. "Wash Buffer 1" was 1 mM HCl; "Wash Buffer 2" was 100 mM Tris-HCl, pH 8.5; "Wash Buffer 3" was 100 mM NaOAc, pH 4.5. "Coupling Buffer" was 200 mM $NaHCO_3$, 500 mM NaCl, pH 8.2. For CuAAC reactions, $CuSO_4$ was purchased from VWR, sodium ascorbate and aminoguanidine-HCl were purchased from Aldrich, and the ligand was bathophenanthroline sulfonated sodium salt from GFS Chemicals. For dye-labeling of lysate, NHS-AF633 (AlexaFluor-633 carboxylic acid, succinimidyl

ester) was purchased from Invitrogen. The BCA Assay kit was purchased from Pierce. Methanol, chloroform, and other solvents were purchased from VWR.

Cell lysate preparation. Bacterial cultures were grown as described in Chapters II and III. Lysis Buffer was composed of 50 mM Tris (pH 7.5), 100 mM NaCl, 0.1 mM PMSF, and Roche Complete Protease Inhibitor. OmniLyse cell lysis kits were purchased from Claremont BioSolutions. The BCA Assay kit was purchased from Pierce.

Amine functionalization of glass slides. Ammonium hydroxide was purchased from VWR. All other reagents were reagent grade and purchased from Sigma-Aldrich.

Microarray printing. "PBS-T" was composed of 50 mM phosphate, 150 mM NaCl, and 0.01% Tween-20. "TBS-T" was composed of 50 mM Tris (pH 7.5), 150 mM NaCl, and 0.01% Tween-20. The DIBO-NHS (Click-iT Succinimidyl Ester DIBO Alkyne) reagent was purchased from Life Technologies/Invitrogen. ADIBO-NHS and ADIBO-PEG5-NHS were purchased from Click Chemistry Tools. Methyl-PEG4-NHS (MS(PEG)4 succinimidyl ester) was purchased from Pierce. Anhydrous DMSO was purchased from Cambridge Isotope Laboratories. The SpotBot Personal Microarray System is manufactured by ArrayIt.

Methods

Experiments with alkyne-derivatized agarose beads. Beads were resuspended in the bottle, and 5 mL of slurry was transferred to a new conical vial. Packing buffer was removed, and beads were then washed 3 x 10 mL with Wash Buffer 1 and divided into two batches. Each batch was reacted with 25 mM acetylene-PEG4-amine or ADIBO-PEG4-amine in Coupling Buffer for 5 hr at room temperature. Quenching was accomplished with 100 mM ethanolamine or methyl-PEG4-amine in Coupling Buffer overnight at 4°C. Beads were then washed 3 x 10 mL alternating between Wash Buffers 2 and 3. Some beads were blocked with BSA (0.1% in 500 mM NaCl, pH 8.2). After a final set of washes with 500 mM NaCl, pH 8.2, beads were ready for reaction with the Azide-Fluor488 fluorescent probe.

To all beads, Azide-Fluor488 was added at a final concentration of 50 μM. For ADIBO-functionalized beads, no additional reagents were required. For acetylene-functionalized beads, CuAAC click reactions were performed as follows: 200 μM $CuSO_4$, 500 μM ligand, 5 mM sodium ascorbate, and 5 mM aminoguanidine-HCl (the Azide-Fluor488, $CuSO_4$, and ligand were combined separately and pre-mixed for 3 min). All reactions were performed in standard Phosphate-Buffered Saline (PBS). Reactions were allowed to proceed for 1.5 hr at room temperature, then beads were washed 8 times with PBS.

For experiments with dye-labeled lysate, *E. coli* BL21(DE3) cells were grown but no protein expression was induced. Cells were lysed in lysis buffer (1% SDS, 50 mM Tris-HCl, pH 8.0) according to the following formula: 50 μL lysis buffer per mL culture per OD_{600} unit. Lysate was dye-labeled with NHS-AF633 in PBS, then precipitated according to the methanol-chloroform precipitation protocol described in the Invitrogen Click-iT kit handbook. The protein pellets were resuspended in 8 M urea buffer, and protein concentration was measured using the BCA Assay kit. Dyed lysate was used at a concentration of 0.25 mg/mL in experiments with beads.

To measure the beads per unit volume, OD_{400} was measured on the plate reader. To measure the signal of Fluor488, samples were excited at 488 nm (bandwidth: 20 nm) and signal was read at 530 nm (bandwidth: 20 nm). To measure the signal of AF633, samples were excited at 633 nm (bandwidth: 5 nm) and signal was read at 660 nm (bandwidth: 5 nm). Fluorescence values were divided by OD_{400} values for normalization.

Cell lysate preparation. A cell pellet corresponding to approximately 5 mL of a given *E. coli* culture was resuspended in 500 μL Lysis Buffer at 4°C, lysed with mechanical disruption using OmniLyse cell lysis kits, and clarified of cellular debris via centrifugation (10 min x 10,000 *g* at 4°C). The concentration of soluble protein was determined with the BCA Assay kit. If lysates were not used the same day they were prepared, they were stored at -80°C and used within two weeks.

Amine functionalization of Maven glass slides. Microscope slides from Maven Biotechnologies were prepared for vapor phase deposition of APTES

((3-aminopropyl)triethoxysilane) by cleaning with RCA-1 (30% hydrogen peroxide, ammonium hydroxide, and RO water at a 1:1:10 ratio) and washing with DI water 3 times. Excess water was removed with N_2 gas, and slides were dried in an oven at 70-80°C for 1 hr. Slides were allowed to equilibrate to room temperature and placed in a vacuum desiccator. Then 2 microcentrifuge tubes, each containing 100 µL APTES, were placed inside the vacuum desiccator. The chamber was evacuated to 22 mm Hg and slides were incubated for at least 2 hr. Slides were stored at room temperature away from light prior to use.

Microarray printing. Microarrays were contact printed with a SpotBot Personal Microarray System using a 946 MP4 pin with approximately a 135 µm spot size on a 300 µm pitch. Microscope slides were washed 3 times with PBS-T and DI water and dried completely with N_2 gas. Control spots of 0.1 mg/mL BSA in TBS-T and 0.5 mg/mL polyethylene glycol (PEG, MW 35000) were printed, as were DIBO-NHS, ADIBO-NHS, ADIBO-PEG5-NHS, and methyl-PEG4-NHS in anhydrous DMSO and 0.5 mg/mL PEG. All arrays were printed in 55–65% humidity in under 1.5 hr. Spots were allowed to dry slowly in a humidified chamber, and slides were stored under N_2 gas with desiccant at -20°C.

LFIRE experiments with lysate samples. Maven amine-coated glass slides with printed microarrays were washed 3 times with PBS-T, mounted on a glass prism with index matching oil, and loaded into the LFIRE instrument. The optimal angle for the total internal reflection measurement, one that maximized the difference in signal between background and microarray spots, was determined. Each microarray was blocked with 0.25 mg/mL BSA in TBS-T for 1 hr and washed 5 times with TBS-T. Cell lysates containing recombinant proteins of interest were added at a final concentration of 0.5 mg/mL. The microarrays were incubated with lysates for approximately 1 hr, then washed 8 times with TBS-T. LFIRE data from each microarray were captured at approximately 80–120 second intervals throughout the experiment. All experiments were performed at room temperature.

LFIRE data analysis. The LFIRE instrument captures data in 16-bit grayscale TIFF units. These data were analyzed using the ImageJ software package. First, images from one experiment were placed in sequential order, stacked, and image stabilized. The image immediately preceding the addition of cell lysate was subtracted as a baseline from the sequence. Plots of the change in microarray surface and surface profile pixel intensities were generated using the Surface Plot and Plot Profile functions in ImageJ. To extract data from each microarray spot, regions of interest (ROIs) of uniform size were drawn around each spot, and the average pixel intensity from each ROI at each time point was exported to Microsoft Excel. For GFP-containing samples, data from replicate microarray spots were averaged and normalized to the average pixel intensity of the BSA spots, yielding data in relative ellipsometry units (REU). For CaN- and CaM-containing samples, data were not normalized to BSA spot intensities and are reported as TIFF units. All graphs were prepared with Prism software (GraphPad Software).

REFERENCES

1. Zhu, H. & Snyder, M. Protein arrays and microarrays. *Curr. Opin. Chem. Biol.* **5**, 40–5 (2001).

2. Espina, V. *et al.* Protein microarrays: molecular profiling technologies for clinical specimens. *Proteomics* **3**, 2091–100 (2003).

3. Zhu, H. & Qian, J. Chapter Four – Applications of Functional Protein Microarrays in Basic and Clinical Research. *Advances in Genetics* 123–55 (2012).

4. MacBeath, G. Protein microarrays and proteomics. *Nat. Genet. Supp.* **32**, 526–32 (2002).

5. Jason-Moller, L., Murphy, M. & Bruno, J. Overview of Biacore systems and their applications. *Curr. Protoc. Prot. Sci.* **19**, 1–14 (2006).

6. Ray, S., Mehta, G. & Srivastava, S. Label-free detection techniques for protein microarrays: prospects, merits and challenges. *Proteomics* **10**, 731–48 (2010).

7. Mehan, M. R. *et al.* Highly Multiplexed Proteomic Platform for Biomarker Discovery, Diagnostics, and Therapeutics. *Adv. Exp. Med. Biol.* **734**, 283–300 (2013).

8. Bertone, P. & Snyder, M. Advances in functional protein microarray technology. *FEBS J.* **272**, 5400–11 (2005).

9. Jenikova, G., Lao, U. L., Gao, D., Mulchandani, A. & Chen, W. Elastin-calmodulin scaffold for protein microarray fabrication. *Langmuir* **23**, 2277–9 (2007).

10. Chen, Y., Triola, G. & Waldmann, H. Bioorthogonal chemistry for site-specific labeling and surface immobilization of proteins. *Acc. Chem. Res.* **44**, 762–73 (2011).

11. MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760 (2000).

12. Rostovtsev, V. V, Green, L. G., Fokin, V. V & Sharpless, K. B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.* **41**, 2596–9 (2002).

13. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057–64 (2002).

14. Wang, Q. *et al.* Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 3192–3 (2003).

15. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed. Engl.* **48**, 6974–98 (2009).

16. Duckworth, B. P., Zhang, Z., Hosokawa, A. & Distefano, M. D. Selective labeling of proteins by using protein farnesyltransferase. *ChemBioChem* **8**, 98–105 (2007).

17. Ngo, J. T. & Tirrell, D. A. Noncanonical amino acids in the interrogation of cellular protein synthesis. *Acc. Chem. Res.* **44**, 677–85 (2011).

18. Jing, C. & Cornish, V. Chemical tags for labeling proteins inside living cells. *Acc. Chem. Res.* **44**, 784–792 (2011).

19. Yao, J. Z. *et al.* Fluorophore targeting to cellular proteins via enzyme-mediated azide ligation and strain-promoted cycloaddition. *J. Am. Chem. Soc.* **134**, 3720–8 (2012).

20. Popp, M. W.-L., Karssemeijer, R. A. & Ploegh, H. L. Chemoenzymatic site-specific labeling of influenza glycoproteins as a tool to observe virus budding in real time. *PLoS Pathogens* **8**, e1002604 (2012).

21. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046–7 (2004).

22. Beatty, K. E. *et al.* Live-cell imaging of cellular proteins by a strain-promoted azide-alkyne cycloaddition. *ChemBioChem* **11**, 2092–5 (2010).

23. Van Geel, R., Pruijn, G. J. M., Van Delft, F. L. & Boelens, W. C. Preventing thiol-yne addition improves the specificity of strain-promoted azide-alkyne cycloaddition. *Bioconjugate Chem.* **23**, 392–8 (2012).

24. Jewett, J. C. & Bertozzi, C. R. Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* **39**, 1272–79 (2010).

25. LFIRE Technology. *Maven Biotechnologies* at <http://www.mavenbiotech.com/technology.htm>

26. Mondragon, A. *et al.* Overexpression and purification of human calcineurin alpha from Escherichia coli and assessment of catalytic functions of residues surrounding the binuclear metal center. *Biochemistry* **36**, 4934–42 (1997).

27. Sletten, E. M. & Bertozzi, C. R. From mechanism to mouse: a tale of two bioorthogonal reactions. *Acc. Chem. Res.* **44**, 666–76 (2011).

CHAPTER V

Progress toward the In Vivo *Visualization of Individual Bacterial Proteins after N-Terminal Labeling*

ABSTRACT

In previous chapters, we have taken advantage of the orthogonality exhibited by NMT toward the bacterial proteome. This feature of NMT enabled in-lysate surface coupling of three different proteins that had been labeled by NMT, as summarized in Chapter IV. Here, we describe efforts toward applying NMT-mediated protein labeling to the selective functionalization of bacterial proteins of interest in order to visualize them in live cells and study their localization patterns.

Our initial attempts at such imaging experiments were complicated by the fact that thorough washing was not sufficient to remove unbound 12-ADA from cells; in fact, 12-ADA may be incorporated into the cell membrane by endogenous enzymes. Thus, we sought a different azide fatty acid that would be transferred by NMT to substrate proteins, but that could also be washed out of cells. We found that 7-azidoheptanoic acid fulfilled both of these criteria, as described in this chapter. We also selected two proteins for initial studies, PyrG and MreB, both of which are known to undergo spatiotemporal localization in bacterial cells. Cloning and expression were completed for both proteins, and an engineered MreB construct was found to be robustly labeled by NMT in live cells with both 7-azidoheptanoic acid and 12-ADA. These results will be useful in future applications of the NMT-mediated protein labeling system to detailed imaging studies of bacterial protein organization.

INTRODUCTION

Organization and Localization of Bacterial Proteins

It has long been believed that bacteria do not exhibit sophisticated spatiotemporal orchestration of their proteins and other biomolecules. Biology textbooks have generally depicted mammalian cells in great detail, with complex organelles that migrate precisely in space and time, while describing prokaryotes in more simple terms. In a marked shift, recent research has shown that bacteria actually do organize and localize their proteins to a much greater degree than previously thought.[1–3] Powerful imaging techniques, such as stochastic optical reconstruction microscopy (STORM) and electron cryotomography (ECT), have enabled scientists to study bacteria at resolutions that are orders of magnitude higher than those offered by traditional light microscopy.[2,4] These developments and changes in the state of the field are summarized in the following selection from an essay written by Professor Bonnie Bassler, a leading microbiologist at Princeton University:

> "Eukaryotes have long been known to possess sophisticated subcellular architecture in which DNA, RNA, and proteins are localized to the right place at the right time. Bacteria, in contrast, have until recently been thought to be unorganized bags of goop. Consequently, cell biology was generally restricted to eukaryotes. However, remarkable recent advances in imaging technologies… have made it so that we can now peer into bacterial cells as we traditionally peered into bigger eukaryotic cells. These technologies have revealed that bacteria are decidedly organized." *(Adapted from Reference 5.)*

NMT-Mediated Protein Labeling for Imaging Studies in Bacteria

After developing the GFP/NMT model system described in Chapter II and confirming that NMT is selective toward engineered substrate proteins in bacteria, we recognized that NMT-mediated protein labeling could be a very useful tool for visualizing and studying individual bacterial proteins. By adding an NMT recognition sequence to a bacterial protein of interest, we postulated that we could selectively label the protein with 12-ADA and react it with a cyclooctyne dye for imaging. To test this hypothesis, we first attempted to utilize the original model system: Fyn-GFP and hNMT2 were co-expressed in bacteria in the presence of 12-ADA, and cells were treated with a cyclooctyne-lissamine-rhodamine dye. Initial results seemed promising, though the control experiment shown in Figure V-1 yielded an unexpected result: some cells appeared fluorescent even when no substrate protein was expressed.

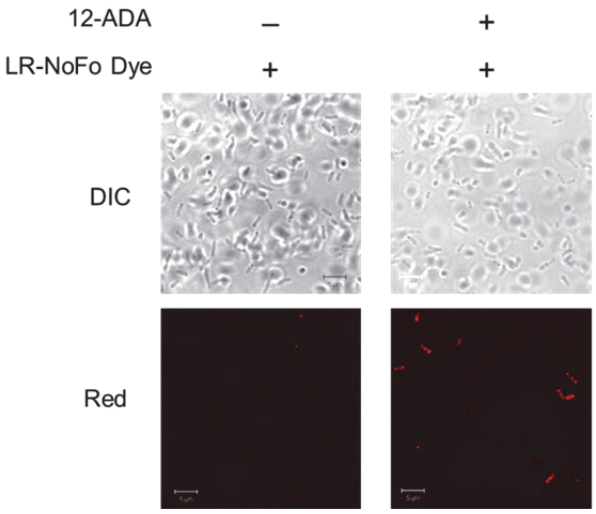


Figure V-1. Confocal microscope images of live cells expressing no NMT substrate protein, after reaction with lissamine-rhodamine non-fluorinated cyclooctyne dye (LR-NoFo). Cells that were not exposed to 12-ADA were dark (left panels), while prior exposure to 12-ADA yielded a fluorescence signal (right panels), despite the lack of a substrate protein for NMT to label with 12-ADA.

Based on the work described in Chapter II, we were confident that NMT was not labeling endogenous bacterial proteins. Thus, we concluded that 12-ADA, unlike the LR-NoFo dye, was not removed from cells by standard wash conditions that are compatible with live cells. In fact, 12-ADA may be transported into cells by bacterial enzymes responsible for the uptake of exogenous fatty acids; these fatty acids may then serve different functions within the cell, including structural roles in the cell membrane.[6] In any case, 12-ADA appeared to be interacting with bacterial cells in a manner that rendered it unsuitable for our proposed live-cell imaging studies.

In subsequent sections of this chapter, we describe our investigation of other azide fatty acids (FAs) to use in place of 12-ADA for imaging experiments. Specifically, we tested azide FAs with shorter chain lengths, searching for at least one that would be (a) bound and transferred by NMT to a substrate protein, and (b) washed out of cells in its free form without the use of harsh wash conditions. Using a combination of lysate and live-cell dye-labeling experiments, we identified a promising candidate, 7-azidoheptanoic acid. The azide FAs that we tested and the experimental outcomes are described below.

PyrG and MreB

In addition to finding a suitable azide FA for imaging experiments, we also undertook the task of engineering two bacterial proteins for labeling by NMT. The proteins we selected are PyrG and MreB, shown in Figure V-2.

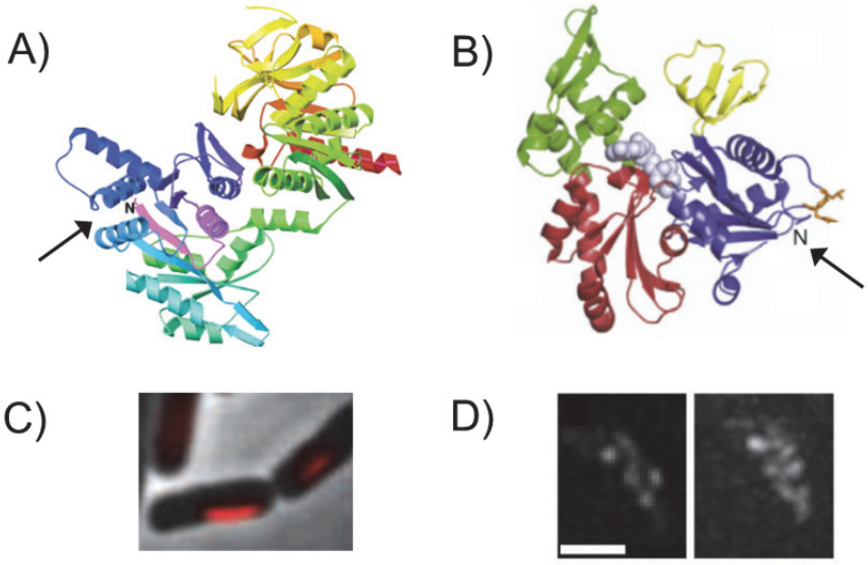


Figure V-2. Overview of the structures and localization patterns of PyrG and MreB. The black arrows point to the N-terminus on the crystal structures of PyrG (A) and MreB (B). In both proteins, the N-terminus appears to be surface-accessible. Confocal images of mCherry fluorescent protein fusions to PyrG (C) and MreB (D) in *E. coli* show that PyrG forms straight filamentous structures, while MreB adopts a helical pattern along the length of the cell. Adapted from References 7-10.

PyrG is a 60-kDa enzyme involved in converting UTP to CTP, and it also plays important structural roles in the cell; the interplay between its enzymatic and physical functions is currently under investigation.[9] PyrG is the *E. coli* homolog of a protein known as Ctp synthase (CtpS) in other bacterial species, such as *C. crescentus*.[9] As shown in Figure V-2C, an mCherry-PyrG fusion protein was shown to form long filamentous structures along the cell membrane; PyrG also self-assembles into filaments in its purified form.[9] We selected PyrG for further studies involving NMT labeling because it assembles into clear structures and also appears to perform interesting functions in the cell.

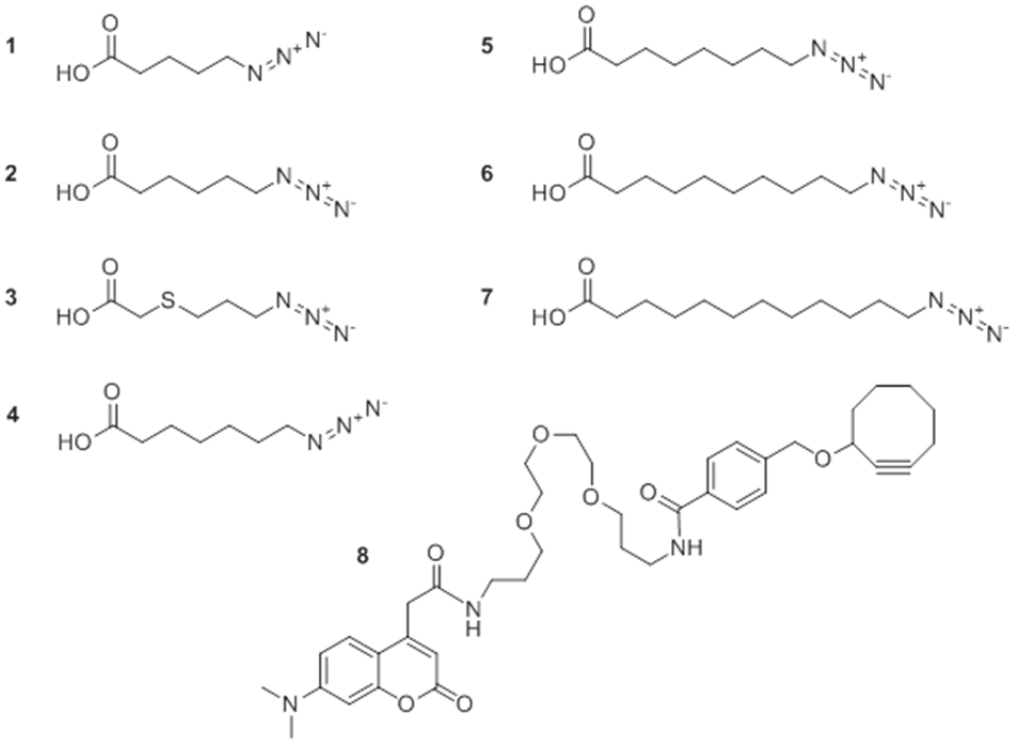
MreB is the primary bacterial homolog of actin.[11] Like PyrG, this 37-kDa protein has been expressed, purified, and structurally characterized. The crystal structure depicted in Figure V-2B is of the *Thermotoga maritima* form of MreB, which is 54% identical (70% related) to the *E. coli* form. In cells, MreB appears to assemble into filaments that in turn organize into a helical structure along the length of the cell (Figure V-2D). The length of the individual filaments is under continued investigation,[12] but they appear to play a role in maintaining the overall structure of the cell as well as contributing to cell motility.[10] We selected MreB for NMT-mediated protein labeling studies because, like PyrG, it forms interesting structures in the cell, and it is an early example of a bacterial protein whose localization patterns are actively regulated *in vivo*.

RESULTS AND DISCUSSION

Studies of Azide Fatty Acid Analogs: Dye-Labeling Lysate

The structures of the azide FAs described in this chapter are shown in Chart V-1.

Chart V-1. Azide fatty acids (azide FAs) and cyclooctyne-coumarin dye utilized for *in vivo* labeling studies in bacteria. FA **7** is 12-ADA.



The azide FAs (**1-7**) were prepared by a former post-doctoral scholar in the Tirrell Group, Dr. Janek Szychowski, in a manner similar to that described in Chapter II for the synthesis of 12-ADA. The cyclooctyne-coumarin dye (**8**) utilized for the *in vivo* experiments described later in this chapter was prepared as previously described, also by Dr. Szychowski.[13] Cyclooctyne **8** is known to cross the cell membrane of mammalian cells and react selectively with azide-tagged biomolecules without harming cells. Its use in bacteria has not yet been reported, though we surmised that it would behave similarly in *E. coli* as in mammalian cells.

First, we investigated the extent to which each of the azide FAs was transferred by NMT to a substrate protein. For these experiments, we utilized the Fyn-GFP/hNMT2

co-expression system described in Chapter II. Protein expression, cell harvesting, and cell lysis were carried out as previously described; the only difference was that no fatty acid or one of the azide FAs in Chart V-1 was added to the expression flasks instead of 12-ADA when inducing protein expression. Lysates were treated with alkyne-TAMRA to enable detection of azide-labeled protein, as described in Chapters II and III, and analyzed by SDS-PAGE (Figure V-3).

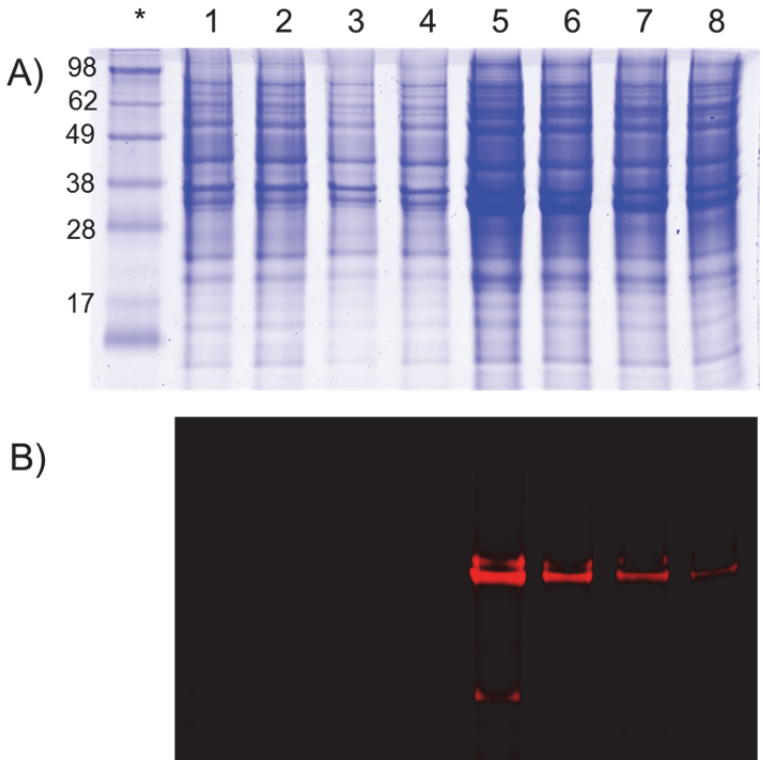


Figure V-3. SDS-PAGE analysis of lysate samples from Fyn-GFP/ hNMT2 co-expression cultures exposed to different azide FAs. Samples were treated with alkyne-TAMRA for detection of azide-labeled Fyn-GFP. The gel was stained with Coomassie colloidal blue (A) and imaged for TAMRA signal (B). Fluorescent bands indicate successful and selective transfer of an azide FA onto Fyn-GFP. Lanes correspond to cultures exposed to the following azide FAs: 1: no fatty acid; 2: FA **1**, 3: FA **2**, 4: FA **3**, 5: FA **4**, 6: FA **5**, 7: FA **6**, 8: FA **7**.

We were pleased to find that azide FA **4**, i.e., 7-azidoheptanoic acid, and FAs longer than **4** were appended by NMT to Fyn-GFP, as evidenced by the fluorescent bands in Figure V-3B (Lanes 5-8). The brighter appearance of the fluorescent band in the FA **4** lane (Lane 5) compared to the fluorescent bands for FAs **6** and **7** (Lanes 7-8) is simply an artifact of differences in protein loading across lanes, as confirmed by similar experiments culminating in SDS-PAGE analysis; when lanes were equally loaded with protein samples, we observed brighter TAMRA fluorescence for FAs **6** and **7**, as might be expected (data not shown) . The sharp drop-off in signal for azide FAs shorter than **4** is in agreement with *in vitro* work carried out by the Gordon Lab with FAs possessing alkyl chains of varying lengths.[14]

One other notable feature of the gel image in Figure V-3B is the weak second band in the FA **4** lane (Lane 5). It is possible that FAs **4** and **5**, 7-azidoheptanoic acid and 8-azidooctanoic acid, respectively, may be utilized by lipoic acid ligase (also known as lipoate protein ligase) to label its substrate proteins; use of octanoic acid by this enzyme instead of lipoic acid has been documented when the concentration of lipoic acid is low or the concentration of octanoic acid is particularly high.[15] We found in separate experiments that titrating just 100 μM free lipoic acid into the expression system eliminated the second band altogether (data not shown). Thus, if background labeling of lipoic acid ligase substrates with FA **4** does appear to interfere with visualizing a protein of interest during live-cell imaging, then addition of a small amount of lipoic acid should address the problem. However, given that the lower band in Lane 5 is significantly weaker than the Fyn-GFP band, this issue may not even arise in the context of microscopy experiments.

Studies of Azide Fatty Acid Analogs: Dye-Labeling Live Cells

Next, we examined how effectively each of the azide FAs **1-7** could be washed out of cells. In the first set of experiments described here, no substrate protein was expressed, yielding an experimental set-up similar to that in Figure V-1. Four hours after the addition of no fatty acid or one of the azide FAs **1-7** to the expression flask, cells were harvested, but were not lysed. Instead, samples were thoroughly washed, treated with cyclooctyne-coumarin **8**, washed again, and analyzed on a 96-well plate reader. Coumarin fluorescence as well as cell density were measured so that fluorescence measurements could be normalized. Both FA **1** and FA **4** produced a very low fluorescence signal (Figure V-4), indicating that those FAs were effectively washed out of cells prior to dye-labeling. Interesting, FAs **1** and **4** are both relatively short and possess an odd number of carbons.

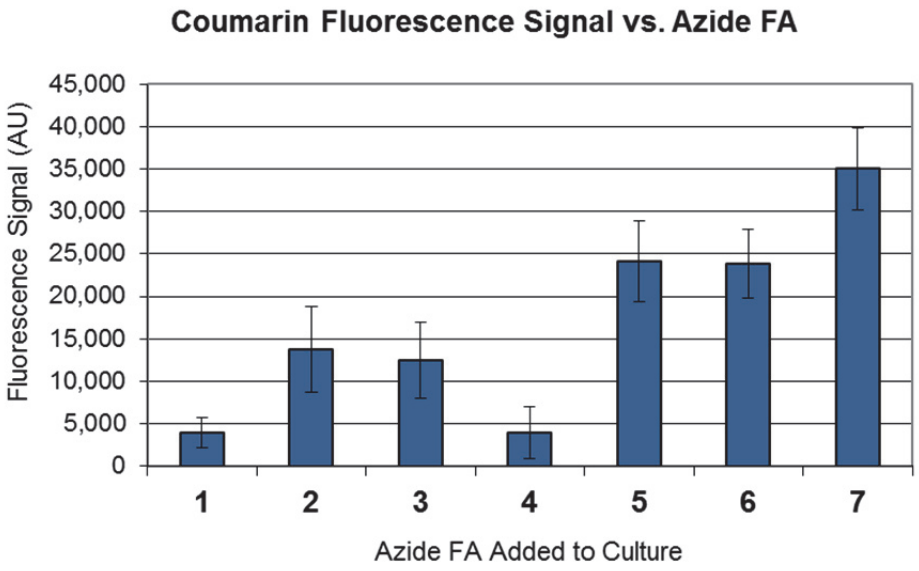


Figure V-4. Fluorescence signal of live *E. coli* cells after addition of an azide FA to the growth culture and subsequent dye-labeling with cyclooctyne-coumarin **8**. The azide FA added to each culture is

denoted as in Chart V-1. Each fluorescence value represents the average of a triplicate set of identical cultures, ± standard deviation. Fluorescence values were normalized by cell density (OD_{600}) and were background-corrected via subtraction of the fluorescence value of a control culture that was treated with **8** but not exposed to a FA. These experiments identified azide FA **4** as a promising candidate for *in vivo* labeling studies.

In a follow-up set of experiments, we again tested FAs **4** and **7** (7-azidoheptanoic acid and 12-ADA, respectively), now alongside additional controls. We grew cultures and dye-labeled cells as above, but we also prepared a sample that was exposed to neither a fatty acid nor cyclooctyne-coumarin **8,** and a sample that was exposed to myristic acid and **8**. These controls provided some insight into the potential interference of cellular autofluorescence (i.e., the sample not treated with **8**), the degree of nonspecific interaction between **8** and cellular components (i.e., the sample exposed to **8** but not to a fatty acid), and the degree of nonspecific interaction between **8** and an unreactive fatty acid (i.e., sample exposed to **8** and myristic acid). Furthermore, we grew parallel cultures in which no substrate protein was expressed or Fyn-GFP and hNMT2 were co-expressed.

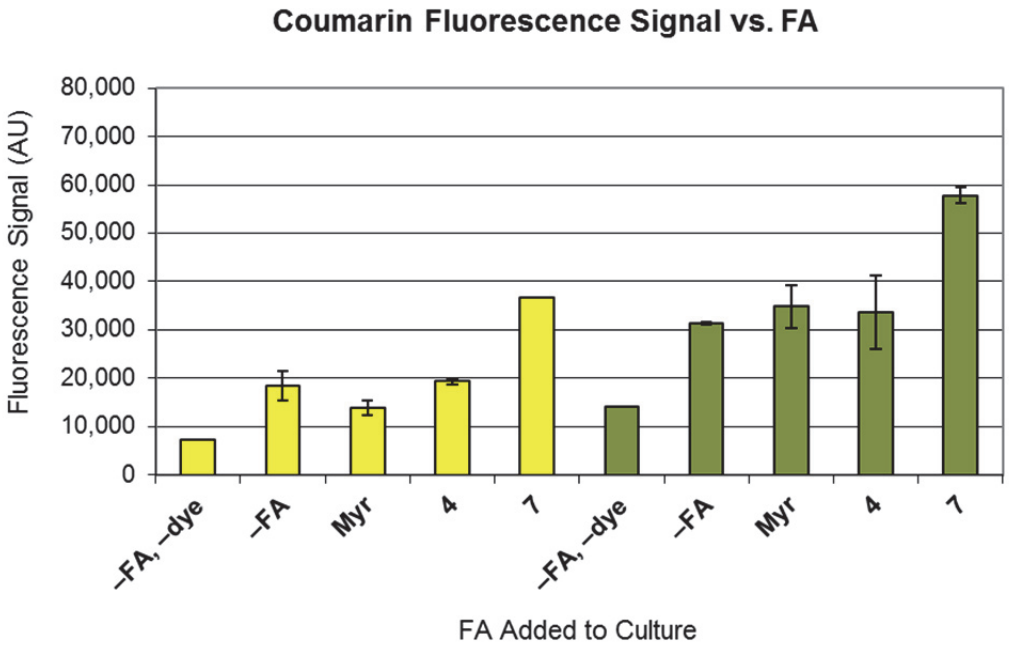


Figure V-5. Fluorescence signal of live *E. coli* cells after addition of no FA ("–FA"), myristic acid ("Myr"), or an azide FA (**4** or **7**), followed by dye-labeling with cyclooctyne-coumarin **8**. Cells exposed to neither a FA nor **8** ("–FA, –dye") served as a control. Cells expressed no substrate protein (yellow bars) or co-expressed Fyn-GFP and hNMT2 (green bars). Each fluorescence value represents the average of a duplicate set of identical cultures, $\pm$ standard deviation. All fluorescence signals were normalized by cell density (OD_{600}).

The data, summarized in Figure V-5, provide some interesting results. We were pleased to find that the fluorescence of cultures to which **4** was added was similar to that of cultures exposed to no fatty acid or to myristic acid prior to dye-labeling with **8**. This similarity indicates that the fluorescence observed for cultures exposed to **4** is essentially background fluorescence due to some coumarin dye remaining in cells after washing, unrelated to the fatty acid alkyl chain or the azide moiety of **4**. The results also indicate that almost half the background signal in each set of samples is simply cellular

autofluorescence, as illustrated by the "–FA, –dye" samples. As expected, cultures exposed to **7** were considerably brighter than cultures exposed to **4** and the three controls. This result is in line with the observations associated with Figure V-1: azide FA **7** is not removed from cells by standard wash procedures and consequently reacts with cyclooctyne-coumarin **8**.

Surprisingly, the fluorescence signal for the Fyn-GFP/hNMT2 culture exposed to **7** was almost double that of the culture exposed to **4**, despite dye-labeling experiments in lysate (Figure V-3B) indicating that NMT transfers **4** to Fyn-GFP as well as it transfers **7**. It is possible that cyclooctyne-coumarin **8**, though membrane-permeable in mammalian cells, does not cross the bacterial cell membrane/cell wall to the same extent; in this scenario, most of the fluorescence in our live-cell experiments would arise from alkyne-reactive cell membrane components (i.e. azide FA **7**) rather than proteins inside the cell (i.e. Fyn-GFP labeled with **4** or **7**) reacting with **8**. Future work probing the behavior and membrane permeability of coumarin **8** in live bacteria should elucidate the cause of this discrepancy.

Finally, it may be worthwhile to confirm that FA **4** is not elongated or otherwise processed within the cell prior to binding by NMT. The methods described in Chapters II and III for whole-protein LC-MS studies would be readily applicable to this problem: Fyn-GFP co-expressed with NMT in the presence of FA **4** would be purified and subjected to LC-MS, in order to verify the appearance of a mass shift corresponding to the addition of one molecule of FA **4**. It seemed unlikely that enzymatic elongation is occurring, given the difference in data sets obtained for FAs **4** and **7** in live-cell dye-

labeling experiments. Thus, we moved forward as planned with experiments involving engineered substrate proteins.

For experiments with engineered PyrG and MreB constructs, described in the next section, we proceeded with FA **4**, 7-azidoheptanoic acid, which fulfills both criteria established earlier for an azide FA to be suitable for imaging studies: FA **4** is transferred to substrate proteins by NMT, and free FA **4** is removed from live cells upon completion of a standard wash protocol.

Preparation and Evaluation of yARF-PyrG and yARF-MreB Constructs

As noted earlier, the bacterial proteins PyrG and MreB were selected for NMT-mediated protein labeling studies. Both proteins were engineered to display the yARF NMT recognition sequence (MGLFASK) described in Chapters II and III. Rather than using a recombinant plasmid as the PCR template, as we had done for all previous cloning projects, we used genomic DNA isolated from *E. coli* cells as the template for PCR amplification. The final yARF-PyrG and yARF-MreB constructs were transformed into *E. coli* BL21(DE3) competent cells already harboring the hNMT1 plasmid. Co-expression cultures were grown in LB medium as described in Chapters II and III, with the addition of no fatty acid, **4**, or **7** when protein expression was induced. Samples were collected at regular time points (30, 60, 90, and 120 min) to monitor cell density, protein expression, and extent of protein labeling for each culture. Growth curves are presented first, in Figure V-6.

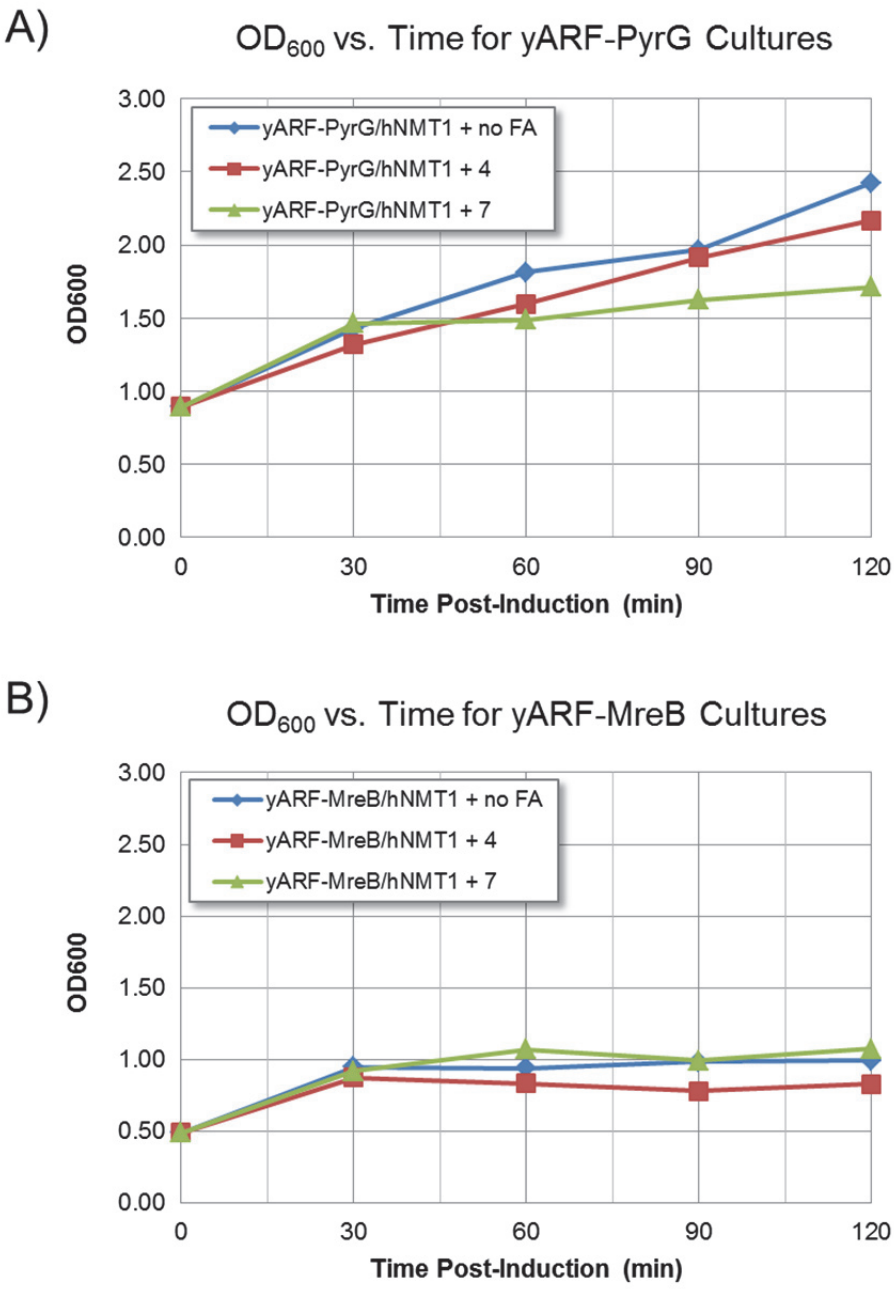


Figure V-6. Growth rates for co-expression cultures of yARF-PyrG/hNMT1 (A) and yARF-MreB/hNMT1 (B). Cell density (OD_{600}) was measured at 30-min intervals and plotted as a function of time. Addition of an azide FA (**4** or **7**) does not appear to impact cell growth for either protein. Expression of yARF-MreB has a larger effect on cell growth than expression of yARF-PyrG.

In contrast with our earlier work involving GFP, CaN, and CaM, the experiments in this chapter involve engineered versions of proteins that are endogenous to bacteria and that perform important functions for proper cell functioning. Thus, we were curious how the expression of yARF-PyrG or yARF-MreB and addition of **4** or **7** would affect cell growth, as measured by cell density (OD_{600}). For both the yARF-PyrG/hNMT1 and yARF-MreB/hNMT1 cultures, addition of an azide FA appeared to have little or no impact on growth rates, an encouraging sign with regard to utilizing NMT-mediated protein labeling to study bacterial proteins. Expression of yARF-PyrG did not appear to impact cellular health as measured by cell density, which increased normally with time. However, expression of yARF-MreB did have a negative effect on growth, evidenced by a plateau in cell density after induction. For future work, it will be worthwhile to place each construct under the control of native promoters in order to achieve endogenous levels of protein expression, as well as to limit expression to the appropriate times and places in the cell. It will also be important to examine any changes to protein structure and function resulting from addition of an NMT recognition sequence and 12-ADA.

At each time point, cells were also collected and lysed in order to determine the level of protein expression and the extent of labeling by NMT. Lysate samples were treated with alkyne-TAMRA, precipitated, resuspended, and analyzed by SDS-PAGE, as described in Chapters II and III. Examination of the results for yARF-PyrG (Figure V-7) indicate that the protein is expressed rapidly and robustly, though almost no NMT labeling is observed. It is possible that the yARF-PyrG protein N-terminus is inaccessible to NMT in the protein's folded state; it could be that the assembly of PyrG monomers into filaments blocks access to the N-terminus of individual protein

monomers; or in either of these scenarios, it is possible that the N-terminus is indeed labeled by NMT but is inaccessible for reaction with alkyne-TAMRA. One or more of these factors, or others, could contribute to the lack of yARF-PyrG labeling by NMT, despite high expression levels.

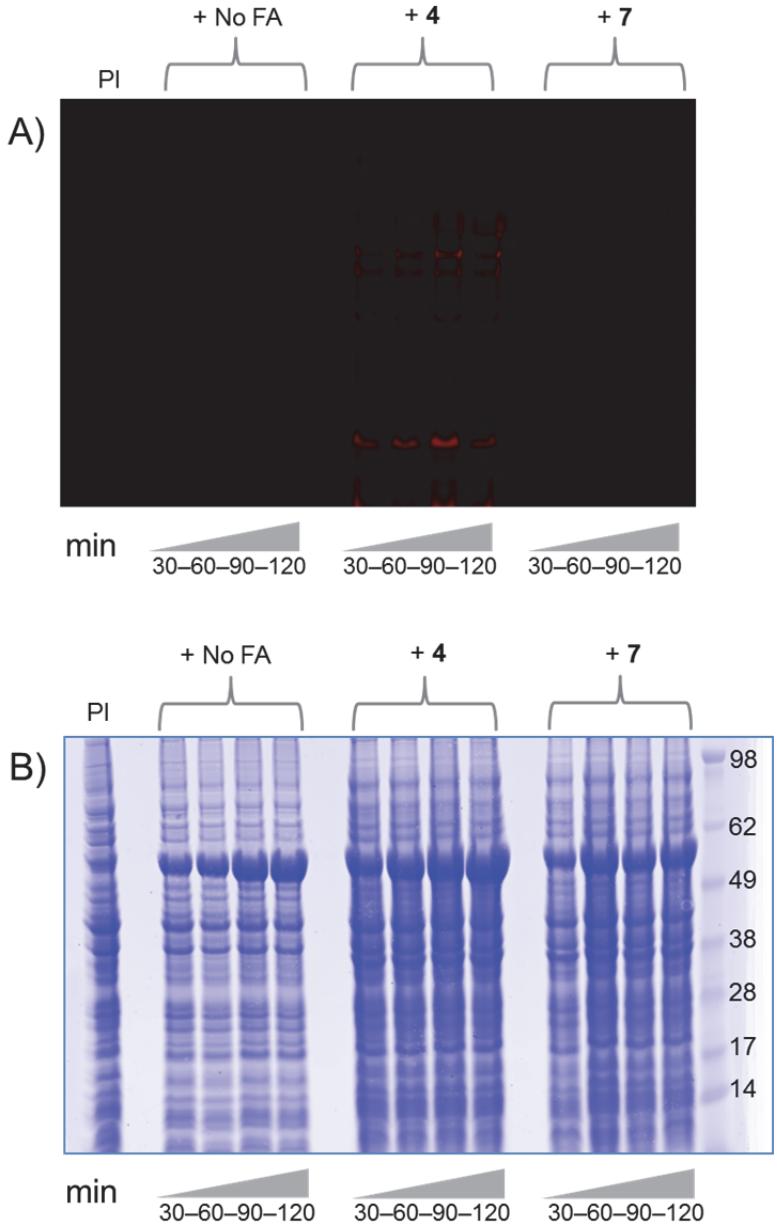


Figure V-7. SDS-PAGE analysis of lysate samples from yARF-PyrG/ hNMT1 co-expression cultures to which no FA, azide FA **4**, or azide FA **7** was added. Bacterial culture samples were collected at 30, 60, 90, and 120 min, and resultant lysates were treated with alkyne-TAMRA for detection of azide-labeled yARF-PyrG. The gel was imaged for TAMRA signal (A) and stained with Coomassie colloidal blue (B). The lack of fluorescent bands indicates poor labeling of yARF-PyrG by NMT, though the strong Coomassie-stained bands near 62 kDa indicate successful expression.

Identical experiments were carried out with samples from yARF-MreB/hNMT1 co-expressions. Again, lysate samples were treated with alkyne-TAMRA, precipitated, resuspended, and analyzed by SDS-PAGE. The results shown in Figure V-8 indicate that yARF-MreB expresses well and is labeled by NMT with either FA **4** or **7**, a promising finding with regard to using **4** for future imaging studies. Furthermore, high levels of protein expression and labeling were observed for yARF-MreB within just 30 min. We anticipate that even shorter expression times could be used, thus providing a more accurate snapshot of yARF-MreB localization at a given point in time or in response to a particular stimulus.

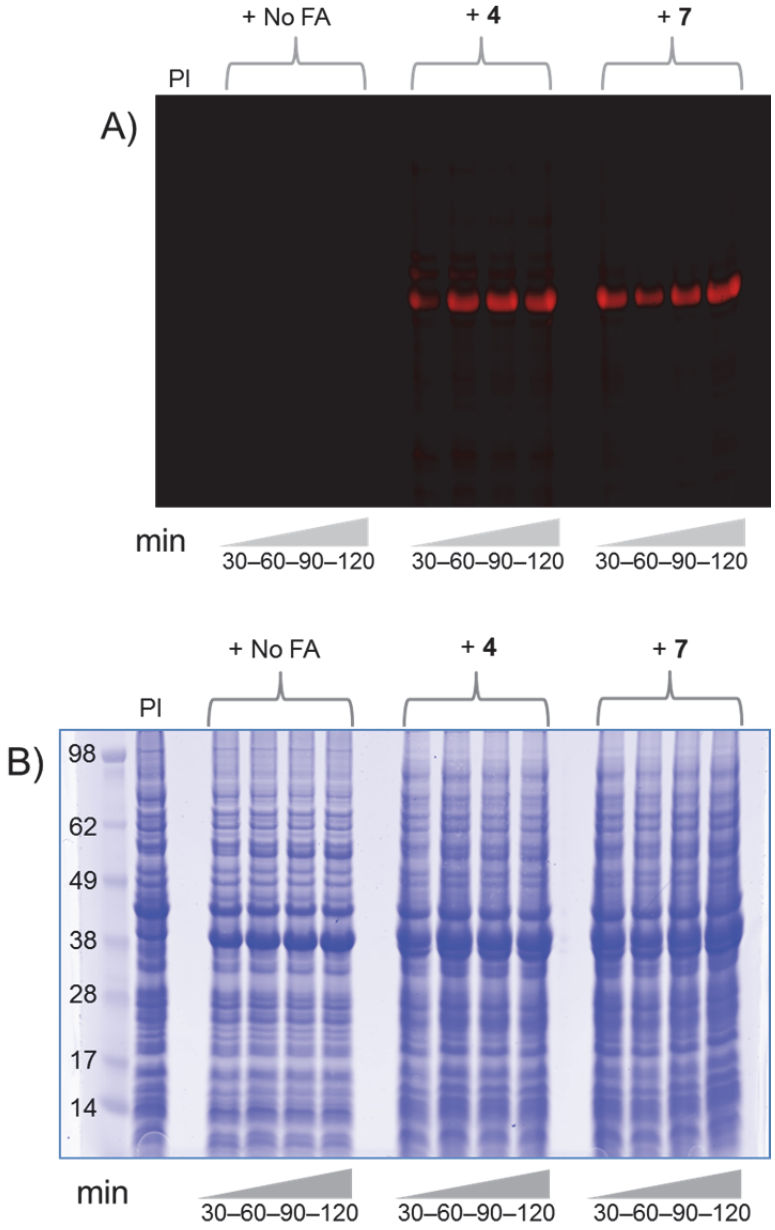


Figure V-8. SDS-PAGE analysis of lysate samples from yARF-MreB/ hNMT1 co-expression cultures to which no FA, azide FA **4**, or azide FA **7** was added. Bacterial culture samples were collected at 30, 60, 90, and 120 min, and resultant lysates were treated with alkyne-TAMRA for detection of azide-labeled yARF-MreB. The gel was imaged for TAMRA signal (A) and stained with Coomassie colloidal blue (B). The strong Coomassie-stained bands near 38 kDa and the corresponding TAMRA bands indicate rapid expression of yARF-MreB and robust labeling by NMT.

CONCLUSION

We have developed and evaluated nearly all of the necessary components to apply NMT-mediated protein labeling to the selective *in vivo* visualization of specific bacterial proteins. As with the GFP-based model system described in Chapter II, a reactive fatty acid, NMT, and a substrate protein are required. We evaluated a family of azide fatty acids with varying chain lengths and identified 7-azidoheptanoic acid as a promising candidate for imaging studies: our experiments demonstrate that this azide fatty acid is transferred to substrate proteins by NMT, like 12-ADA, and can be washed out of cells, unlike 12-ADA. We also prepared two novel non-natural substrates for NMT labeling, yARF-PyrG and yARF-MreB, derived from proteins that exhibit interesting localization patterns in cells. Cloning and expression of both constructs was successful, and yARF-MreB was shown to be labeled efficiently by NMT with 7-azidoheptanoic acid.

The final steps in this project include placement of yARF-MreB under the control of endogenous promoters, investigation of the impact of protein engineering and labeling on yARF-MreB, and ultimately, completion of live-cell imaging experiments. Options for protein visualization include the treatment of cells with a cyclooctyne fluorophore followed by confocal microscopy, or possibly the use of ECT with an appropriate probe. In addition, the cloning and protein labeling methods summarized herein could be applied to other proteins of interest in bacteria. The progress described here is a promising start towards employing NMT in detailed studies of the complex and fascinating internal environment of bacterial cells.

EXPERIMENTAL SECTION

Materials

Synthesis of azide fatty acids. The starting materials (bromo-, iodo-, or chloro-acids) were purchased from Aldrich. Silica gel 60 was purchased from EMD Chemicals. Sodium azide and all solvents were purchased from VWR.

Lysate studies with azide fatty acids. The BL21(DE3)/Fyn-GFP/hNMT2 bacterial co-expression system was described in Chapter II. A control cell strain lacking pQE80_Fyn-GFP was also used. LB medium was composed of 10 g tryptone (casein hydrolysate), 5 g yeast extract, and 10 g NaCl per liter. All media was autoclaved before use. Kanamycin (Kan) was used at a working concentration of 35 μg/mL, and ampicillin (Amp) was used at a working concentration of 200 μg/mL. Myristic acid was purchased from Fluka. All optical density (OD) values were measured at 600 nm on a Cary UV-Vis spectrophotometer. Lysate samples were treated with the reagents and according to the protocols of the Click-IT Tetramethylrhodamine (TAMRA) Protein Analysis Detection Kit from Invitrogen. Lysis Buffer was 1% SDS, 50 mM Tris-HCl, pH 8.0, as recommended by Invitrogen. After reaction with alkyne-TAMRA and precipitation, protein samples were run on Invitrogen NuPAGE Novex 4–12% Bis-Tris pre-cast gels and imaged on a GE Typhoon laser scanner. Gels were stained with Coomassie colloidal blue from Invitrogen.

Live-cell dye-labeling studies with azide fatty acids. Materials for expression cultures are as indicated above for "Lysate studies with azide fatty acids," though cells were not lysed after harvesting. Instead, cells were dye-labeled with cyclooctyne-coumarin, prepared by Dr. Janek Szychowski as previously reported,[13] and washed with phosphate-buffered saline (PBS). Fluorescence measurements were collected on a Tecan 96-well plate reader.

Cloning. All oligonucleotide primers were ordered from IDT. Polymerase chain reaction (PCR) experiments were carried out in a BioRad DNA Engine Peltier Thermal Cycler using PfuTurbo DNA Polymerase (Stratagene/Agilent). Genomic DNA, isolated

from *E. coli* BL21 (DE3) cells with the Qiagen DNEasy kit, served as the template DNA for PCR reactions. All restriction enzymes, restriction enzyme buffers, bovine serum albumin (BSA), and ligase were purchased from New England BioLabs (NEB). NEB DNA Ladders (100 bp and 1 kbp "Quick-Load") were used as markers for all DNA agarose gels, which were visualized with the addition of Plus One ethidium bromide solution from Amersham Biosciences on a UVP UV Transilluminator. Zymo Agarose-Dissolving Buffer (ADB) and Zymo Spin II columns, with their associated buffers, were used to purify DNA out of agarose gels. All plasmid DNA acquisition from cells was completed using the Qiagen Spin Miniprep Kit and columns. All sequencing requests were fulfilled by Laragen.

Time-course growth and expression studies. BL21(DE3)/yARF-PyrG/hNMT1 and BL21(DE3)/yARF-MreB/hNMT1 cell strains were prepared by transforming each final construct in pQE80 into BL21(DE3) competent cells already harboring the hNMT1 plasmid. The plasmid encoding hNMT1 and methionine-aminopeptidase (Met-AP) was a gift from the laboratory of Professor Richard Kahn at Emory University (Atlanta, GA).[16] Otherwise, all materials are identical to those listed above for "Lysate studies with azide fatty acids."

Methods

Synthesis of azide fatty acids. Azide fatty acids were prepared by Dr. Janek Szychowski with methods similar to those described in Chapter II for the synthesis of 12-azidododecanoic acid (12-ADA). The starting material for each compound was the bromo-, iodo-, or chloro-acid. After reaction with sodium azide and a work-up procedure, each compound was isolated by rotary evaporation and characterized by ESI-MS and ^{1}H NMR.

Lysate studies with azide fatty acids. The BL21(DE3)/Fyn-GFP/hNMT2 bacterial co-expression system, described in Chapter II, was utilized for studies with the different azide FAs. (Control cultures lacking pQE80_ Fyn-GFP were also grown.) In summary, overnight cultures were diluted into fresh LB, and expression cultures were

grown in an incubator-shaker (37°C, 250 rpm). Protein expression was induced with IPTG (1 mM, from 1 M stock in water) when the OD_{600} value was between 0.8 and 1.1. (In control cultures lacking pQE80_Fyn-GFP, no protein expression was induced.) Either no fatty acid or one of the azide fatty acids in Chart V-1 (500 μM, from 500 mM stock in DMSO) was also added at the time of induction. After 4 hr of protein expression, cells were harvested via centrifugation and the final OD_{600} value was measured. Cell pellets were lysed according to the following formula: 50 μL Lysis Buffer per mL culture per OD_{600} unit. Crude lysates were centrifuged once more, and the supernatant (clarified lysate) was reacted with alkyne-TAMRA according to the protocols supplied by Invitrogen, as described in Chapter II. Samples were precipitated following the methanol-chloroform precipitation protocol described in the same kit instructions, and they were analyzed by SDS-PAGE. To detect TAMRA signal on the Typhoon, the 532 nm laser served as the excitation source (filter set: 580 BP 30). Gels were stained with Coomassie colloidal blue, then imaged again, with the 633 nm laser now serving as the excitation source (no filter).

Live-cell dye-labeling studies with azide fatty acids. Cultures were grown identically as indicated above for "Lysate studies with azide fatty acids," except that cells were not lysed after harvesting. Instead, cells were washed with PBS 3 times to remove excess azide FA, dye-labeled with 50 μM cyclooctyne-coumarin for 30 min at 37°C, and washed again to remove excess dye. Absorption and fluorescence measurements were collected on a plate reader. To measure the cells per unit volume, OD_{600} was measured on the plate reader. To measure cyclooctyne-coumarin fluorescence, samples were excited at 380 nm (bandwidth: 5 nm) and signal was read at 475 nm (bandwidth: 5 nm). Fluorescence values were divided by OD_{600} values for normalization.

Cloning. Genomic DNA isolated from *E. coli* BL21(DE3) cells served as the template DNA for preparation of both yARF-PyrG and yARF-MreB. Primers were designed to encode the yARF recognition sequence (MGLFASK, from ATG GGT CTG TTC GCG TCT AAA) at the 5′ end of the gene and a 6xHis tag at the 3′ end of the gene, as well as appropriate restriction sites. The yARF-PyrG and yARF-MreB PCR products

were digested with EcoRI and HindIII and ligated into pQE80 digested with the same enzymes. (Constructs were also prepared in pET-15b, though they were not utilized for further protein expression studies; similar PCR products were prepared, digested with NcoI and XhoI, and ligated into pET-15b that had been digested with the same enzymes.) All four constructs (pQE80_yARF-PyrG, pQE80_yARF-MreB, pET-15b_yARF-PyrG, and pET-15b_yARF-MreB) were transformed into DH-10b competent cells and plated. Colonies were selected for inoculation of cultures from which DNA was isolated and submitted for sequencing. Each final construct was transformed into competent cells already harboring an NMT plasmid for co-expression experiments.

Time-course growth and expression studies. Co-expression of yARF-PyrG or yARF-MreB with hNMT1 was performed in a manner identical to that described in Chapter II and summarized above in the "Lysate studies with azide fatty acids" section. When inducing protein expression with IPTG, no fatty acid, FA **4** (7-azidoheptanoic acid), or FA **7** (12-ADA) was also added to the expression flask (500 μM, from 500 mM stock in DMSO). Samples were collected after 30 min, 60 min, 90 min, and 120 min of protein expression. At each time point, cells were harvested via centrifugation (10 min x 10,000 *g* at 4°C). To prepare growth curves, the OD_{600} of each time-point sample was also measured on a UV-Vis spectrophotometer. Cell pellets were lysed according to the following formula: 50 μL Lysis Buffer per mL culture per OD_{600} unit. Crude lysates were centrifuged once more, and the supernatant (clarified lysate) was reacted with alkyne-TAMRA according to the protocols supplied by Invitrogen, as described in Chapter II. Samples were precipitated following the methanol-chloroform precipitation protocol described in the same kit instructions, and they were analyzed by SDS-PAGE. To detect TAMRA signal on the Typhoon, the 532 nm laser served as the excitation source (filter set: 580 BP 30). Gels were stained with Coomassie colloidal blue, then imaged again, with the 633 nm laser now serving as the excitation source (no filter).

REFERENCES

1. Gitai, Z. The new bacterial cell biology: moving parts and subcellular architecture. *Cell* **120**, 577–86 (2005).

2. Jensen, G. J. & Briegel, A. How electron cryotomography is opening a new window onto prokaryotic ultrastructure. *Curr. Opin. Struct. Biol.* **17**, 260–7 (2007).

3. Shapiro, L., McAdams, H. H. & Losick, R. Why and how bacteria localize proteins. *Science* **326**, 1225–8 (2009).

4. Gitai, Z. New fluorescence microscopy methods for microbiology: sharper, faster, and quantitative. *Curr. Opin. Microbiol.* **12**, 341–6 (2009).

5. Bassler, B. L. Small cells—big future. *Mol. Biol. Cell* **21**, 3786–7 (2010).

6. Dirusso, C. C. & Black, P. N. Bacterial long chain fatty acid transport: gateway to a fatty acid-responsive signaling system. *J. Biol. Chem.* **279**, 49563–6 (2004).

7. Endrizzi, J. A., Kim, H., Anderson, P. M. & Baldwin, E. P. Crystal structure of Escherichia coli cytidine triphosphate synthetase, a nucleotide-regulated glutamine amidotransferase/ATP-dependent amidoligase fusion protein and homologue of anticancer and antiparasitic drug targets. *Biochemistry* **43**, 6447–63 (2004).

8. Van den Ent, F., Amos, L. A. & Löwe, J. Prokaryotic origin of the actin cytoskeleton. *Nature* **413**, 39–44 (2001).

9. Ingerson-Mahar, M., Briegel, A., Werner, J. N., Jensen, G. J. & Gitai, Z. The metabolic enzyme CTP synthase forms cytoskeletal filaments. *Nat. Cell Biol.* **12**, 739–46 (2010).

10. Gitai, Z., Dye, N. & Shapiro, L. An actin-like gene can determine cell polarity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8643–8 (2004).

11. Shaevitz, J. W. & Gitai, Z. The structure and function of bacterial actin homologs. *Cold Spring Harb. Perspect. Biol.* a000364 (2010). doi:10.1101/cshperspect.a000364

12. Swulius, M. T. *et al.* Long helical filaments are not seen encircling cells in electron cryotomograms of rod-shaped bacteria. *Biochem. Biophys. Res. Commun.* **407**, 650–5 (2011).

13. Beatty, K. E. *et al.* Live-cell imaging of cellular proteins by a strain-promoted azide-alkyne cycloaddition. *ChemBioChem* **11**, 2092–5 (2010).

14. Kishore, N. S. *et al.* Comparison of the acyl chain specificities of human myristoyl-CoA synthetase and human myristoyl-CoA: protein N-myristoyltransferase. *J. Biol. Chem.* **268**, 4889–902 (1993).

15. Perham, R. N. Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu. Rev. Biochem.* **69**, 961–1004 (2000).

16. Van Valkenburgh, H. A. & Kahn, R. A. Coexpression of proteins with methionine aminopeptidase and/or N-myristoyltransferase in Escherichia coli to increase acylation and homogeneity of protein preparations. *Methods Enzymol.* **344**, 186–93 (2002).

CHAPTER VI

Concluding Remarks and Future Work

SUMMARY OF THESIS WORK

All of the projects presented in this thesis involve the use of N-myristoyl transferase (NMT) to achieve selective and site-specific protein labeling in bacteria. Both *in vitro* and in live cells, NMT catalyzes the transfer of myristic acid, a fatty acid, to the N-terminus of its substrate proteins. While NMT plays important roles in eukaryotes, such as regulating signal transduction via myristoylation of its substrate proteins and enabling the infectivity of HIV and other viruses,[1,2] we examined NMT from the perspective of protein engineering. Specifically, we sought to exploit the tolerance of NMT toward reactive analogs of myristic acid and the ability of NMT to label engineered proteins displaying a recognition sequence derived from a natural substrate. Additional strengths that NMT naturally possesses are its specificity for the protein N-terminus, an attractive site for subsequent protein conjugation,[3] and its orthogonality toward endogenous bacterial proteins.[4] We took advantage of all of these features to develop a novel site-specific protein labeling system, and we evaluated our NMT-based system for different applications.

First, we constructed a bacterial model system for NMT-mediated protein functionalization, which required NMT itself, a test protein, and a reactive fatty acid tolerated by NMT. For these initial studies, summarized in Chapter II, we prepared two GFP-based substrate proteins, yARF-GFP and Fyn-GFP, named for the corresponding known NMT substrates, yARF and Fyn. In addition, we elected to incorporate 12-azidododecanoic acid (12-ADA), which we synthesized and purified, in place of myristic acid; we were drawn to the azide moiety because it can participate in three different bioorthogonal reactions.[5] Each GFP-based substrate protein was co-expressed

with NMT in *E. coli* in the presence of 12-ADA. Purification of each substrate protein and subsequent mass spectrometry experiments confirmed that NMT transfers 12-ADA to the protein N-terminus in quantitative yields. Treatment of lysate samples containing either engineered substrate protein with an azide-reactive fluorescent dye showed that NMT labeled only the substrate protein within the complex cellular environment of bacteria. The work presented in Chapter II constitutes the first example of NMT-mediated labeling of an engineered substrate protein with an analog of myristic acid.

As described in Chapter III, the methods developed for the GFP-based model system were applied to two proteins, calcineurin (CaN) and calmodulin (CaM), that are involved in learning and memory formation in mammals. CaN is naturally myristoylated; successful expression of CaN in bacteria actually requires co-expression with NMT in the presence of myristic acid.[6] We found that substitution of 12-ADA for myristic acid had no impact on the activity of CaN, as measured in a phosphatase activity assay. CaM is not naturally myristoylated, so we created a family of engineered CaM constructs and examined the impact that engineering and labeling had on the activity of each new protein. One construct, hCaNB-CaM, retained wild-type levels of activity and was carried forward for surface coupling work. Experiments similar to those described in Chapter II confirmed that NMT-mediated labeling of the CaN and CaM constructs was quantitative and selective.

Chapter IV details our work with protein microarrays. The yARF-GFP, CaN, and hCaNB-CaM proteins were each labeled with 12-ADA by NMT. Lysate samples containing each labeled substrate protein were incubated with cyclooctyne-spotted glass slides, and protein deposition was detected with the LFIRE (Label-Free Internal

Reflection Ellipsometry) instrument as a change in height at the slide surface. In this context, the orthogonality of NMT toward bacterial proteins is especially useful, as it enables surface immobilization of 12-ADA-labeled proteins without prior purification. To gain a better understanding of the kinetics of the azide-cyclooctyne reaction versus background reactions, we also measured the LFIRE signal over time. We found that significant surface capture of our 12-ADA-labeled proteins occurred within only 15 minutes, while the background signal remained relatively low. Furthermore, by comparing the percent of lysate protein constituted by CaN or hCaNB-CaM with the percent of coupled protein determined to be CaN or hCaNB-CaM, we calculated enrichment factors of 26 and ten for CaN and hCaNB-CaM, respectively. The methods we developed for our surface-coupling experiments comprise a strong foundation for completing high-throughput biochemical measurements with 12-ADA-labeled proteins captured directly out of lysate in a microarray format.

In the final section, Chapter V, we summarize our progress toward using NMT-mediated protein labeling to probe bacterial protein organization. Recent research has demonstrated that prokaryotes orchestrate the expression and movement of their proteins with greater sophistication than previously thought.[7,8] We believe that NMT would be an immensely useful tool for studying these phenomena, as NMT can functionalize a single predetermined substrate protein in bacteria for subsequent reaction with a dye or a probe. To that end, we identified an azide fatty acid, 7-azidoheptanoic acid, that is reactive, transferred by NMT to substrate proteins, and readily washed out of cells in its free form; in contrast, initial imaging experiments with 12-ADA showed that it is not removed from cells by standard wash protocols. We also prepared constructs encoding the bacterial

proteins PyrG and MreB modified with the yARF recognition sequence; both PyrG and MreB assemble into complex structures in cells and exhibit interesting localization patterns. Both yARF-PyrG and yARF-MreB expressed well, and the latter was robustly labeled with 7-azidoheptanoic acid or 12-ADA within 30 minutes. The tools and materials prepared for these experiments constitute a strong start toward the use of NMT-mediated protein labeling for *in vivo* imaging of bacterial proteins of interest.

In summary, the chapters comprising this thesis provide versatile methods and a strong case for the use of NMT in research projects involving site-specific functionalization of proteins. Potential applications include the conjugation of proteins to polymers for therapeutic ends, to surfaces for diagnostic microarrays, or to fluorophores for imaging studies. The N-terminus is well-suited for the covalent attachment of proteins to reactive partners because it is often surface-accessible, even in the folded state of a protein. A number of chemoenzymatic protein labeling techniques have been developed in the past decade, each with its own benefits and drawbacks.[9] The advantages of using NMT include the wide range of reactive myristic acid analogs that may serve as the fatty acid substrate, the small size of the recognition sequence required to achieve labeling of non-natural substrate proteins, and the orthogonality of NMT toward bacterial proteins. We envision many exciting possibilities for future research involving NMT-mediated protein labeling.

FURTHER STUDIES

Activity Measurements of Surface-Immobilized CaN and hCaNB-CaM

The larger goal of the work described in Chapter IV is the preparation of protein microarrays for high-throughput protein characterization studies, with an initial focus on measuring the activity of surface-bound CaN and hCaNB-CaM. Various methods are available for measuring CaN and CaM activity levels; as described in Chapter III, we employed an assay involving colorimetric detection of free phosphate resulting from Ca^{2+}/CaM-dependent dephosphorylation of a phosphopeptide substrate by CaN. This approach could be adapted to measure free phosphate generated by surface-bound proteins, rather than proteins in solution, perhaps by making use of suitably derivatized 96-well plates. Other colorimetric assays have also been developed, such as dephosphorylation by CaN of *para*-nitrophenyl phosphate (pNPP).[10] Alternatively, CaN or hCaNB-CaM could be immobilized on Biacor chips, enabling precise measurements of binding events.[11] These applications and others will be explored in the future by my collaborator, Dr. Tamara Kinzer-Ursem, in her lab at Purdue University.

Protein-Specific Imaging Studies in Bacteria

As described in Chapter V, we have developed nearly all of the necessary components for selective *in vivo* labeling and visualization of yARF-MreB. Dye-labeling experiments with lysates confirmed that bacterial co-expression of NMT and yARF-MreB in the presence of 7-azidoheptanoic acid results in labeling of only yARF-MreB; dye-labeling experiments with intact live cells indicated that the use of 7-azidoheptanoic acid yields no fluorescence signal above background. Thus, the next step is to dye-label

intact cells with a membrane-permeable and cytocompatible dye that reacts specifically with azides, such as cyclooctyne-coumarin[12] or cyclooctyne-BODIPY,[13] and conduct confocal imaging studies to elucidate where and how yARF-MreB is localized. Alternatively, other imaging methods with enhanced spatial resolution, such as electron cryotomography (ECT), could be utilized with suitable probes.[14,15] Finally, to gain a more accurate understanding of how yARF-MreB behaves in the cell, it would be advisable to replace the gene encoding MreB in the bacterial genome with the gene encoding yARF-MreB. This step would ensure that yARF-MreB is under the control of native promoters, so that its expression is turned "on" and "off" appropriately. Following the cloning methods described in Chapters II, III, and V, more constructs could be readily prepared that encode engineered bacterial proteins for labeling by NMT and subsequent visualization.

Controlled Conjugation and Release of N-Terminally Labeled Proteins from Hydrogels

During the course of this thesis, the development of functional biomaterials benefitting from NMT-mediated protein labeling had not been explored. However, work is currently underway to build upon the model system described in Chapter II to decorate hydrogels with site-specifically labeled proteins. In this context, the orthogonality of NMT toward bacterial systems is again beneficial: hydrogels displaying cyclooctyne or alkyne sites can simply be incubated with lysate containing a 12-ADA-labeled protein of interest for attachment of the labeled protein to the material. This project, led by Cole DeForest, a post-doctoral fellow in the Tirrell Lab, holds a great deal of potential.

Beyond the further studies described here, we are also excited by the idea of NMT-mediated protein labeling playing a role in entirely new applications. Moreover, while all of the projects described in this thesis took advantage of azide–alkyne reactions, other bioorthogonal chemistries could be explored, given the tolerance shown by NMT toward a wide variety of reactive fatty acid analogs. Similarly, dozens of recognition sequences have been identified, each possessing different steric and electrostatic characteristics depending on the residues comprising the sequence. Finally, nearly any protein can be engineered to display an NMT recognition sequence, with minimal perturbation of protein structure and function. The inherent characteristics of NMT, and its fascinating balance of specificity, selectivity, and promiscuity, render it a very powerful tool for protein-labeling studies. We hope that the reagents, methods, and results presented in this thesis for NMT-mediated protein labeling will be of great use in future biomedical research.

REFERENCES

1. Towler, D., Gordon, J. I., Adams, S. P. & Glaser, L. The biology and enzymology of eukaryotic protein acylation. *Annu. Rev. Biochem.* **57**, 69–99 (1988).

2. Boutin, J. A. Myristoylation. *Cell Signal* **9**, 15–35 (1997).

3. Jacob, E. & Unger, R. A tale of two tails: why are terminal residues of proteins exposed? *Bioinformatics* **23**, e225–30 (2006).

4. Duronio, R. J. *et al.* Protein N-myristoylation in Escherichia coli: reconstitution of a eukaryotic protein modification in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 1506–10 (1990).

5. Agard, N. J., Baskin, J. M., Prescher, J. A., Lo, A. & Bertozzi, C. R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **1**, 644–8 (2006).

6. Mondragon, A. *et al.* Overexpression and purification of human calcineurin alpha from Escherichia coli and assessment of catalytic functions of residues surrounding the binuclear metal center. *Biochemistry* **36**, 4934–42 (1997).

7. Shapiro, L., McAdams, H. H. & Losick, R. Why and how bacteria localize proteins. *Science* **326**, 1225–8 (2009).

8. Vendeville, A., Larivière, D. & Fourmentin, E. An inventory of the bacterial macromolecular components and their spatial organization. *FEMS Microbiol. Rev.* **35**, 395–414 (2011).

9. Rabuka, D. Chemoenzymatic methods for site-specific protein modification. *Curr. Opin. Chem. Biol.* **14**, 790–6 (2010).

10. Mercan, F. & Bennett, A. M. Analysis of protein tyrosine phosphatases and substrates. *Curr. Protoc. Mol. Biol.* **18**, 1–17 (2010).

11. Jason-Moller, L., Murphy, M. & Bruno, J. Overview of Biacore systems and their applications. *Curr. Protoc. Prot. Sci.* **19**, 1–14 (2006).

12. Beatty, K. E. *et al.* Live-cell imaging of cellular proteins by a strain-promoted azide-alkyne cycloaddition. *ChemBioChem* **11**, 2092–5 (2010).

13. Beatty, K. E., Szychowski, J., Fisk, J. D. & Tirrell, D. A. A BODIPY-cyclooctyne for protein imaging in live cells. *ChemBioChem* **12**, 2137–9 (2011).

14. Jensen, G. J. & Briegel, A. How electron cryotomography is opening a new window onto prokaryotic ultrastructure. *Curr. Opin. Struct. Biol.* **17**, 260–7 (2007).

15. Gitai, Z. New fluorescence microscopy methods for microbiology: sharper, faster, and quantitative. *Curr. Opin. Microbiol.* **12**, 341–6 (2009).