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]

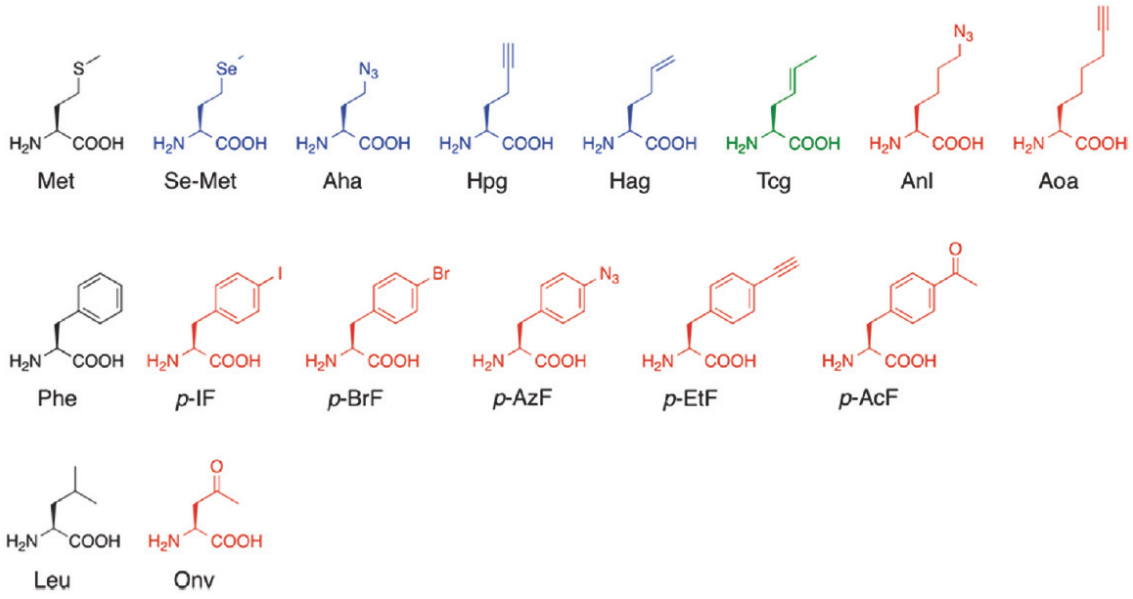

Figure I-1. Numerous non-canonical surrogates for the amino acids methionine (Met), phenylalanine (Phe), and leucine (Leu) have been developed. They may be bound by endogenous tRNA synthetases (those colored blue or green), or they may require the use of mutant tRNA synthetases (those colored red). Adapted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

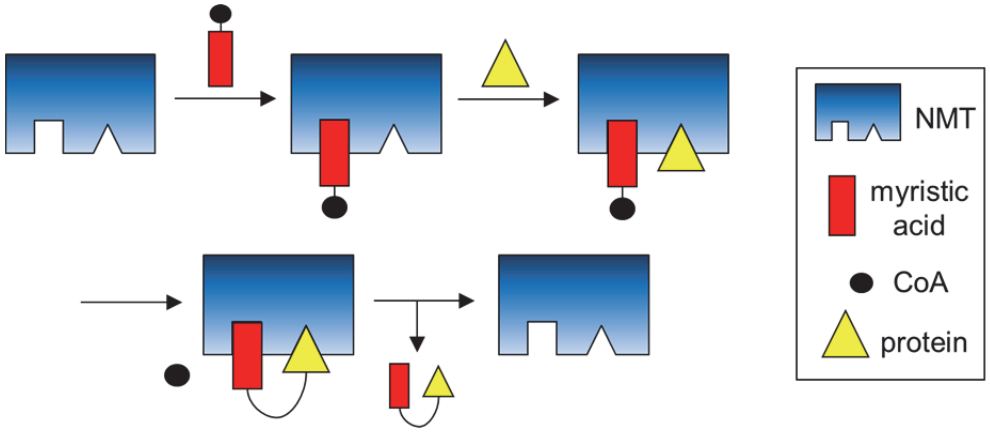

Figure I-2. Schematic overview of the process of myristoylation. NMT is colored blue, the fatty acid is colored red, and the substrate protein is colored yellow. NMT first binds myristoyl-CoA and only then binds a substrate protein. An amide linkage forms between the protein N-terminus and the fatty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

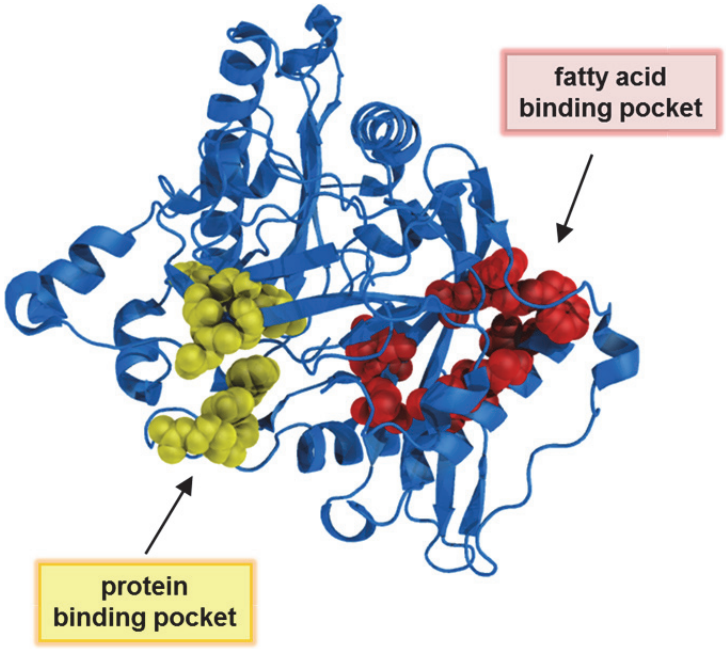

Figure I-3. Crystal structure of NMT, with the binding pocket for myristic acid colored red and the binding pocket for the substrate protein colored yellow. PDB ID: 1IID.

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).

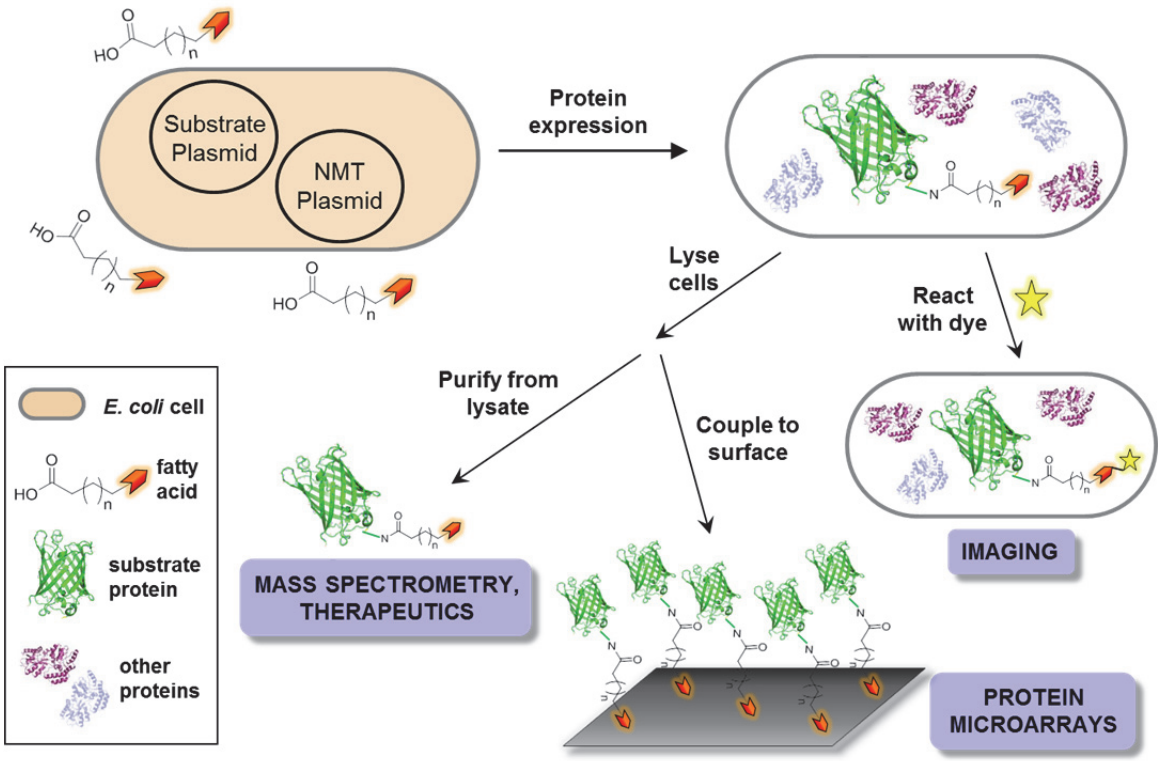

Figure I-4. Overview of thesis projects. NMT-mediated labeling of both natural and engineered substrate proteins, as well as various applications of N-terminal protein labeling, are described herein. Transformation of *E. coli* cells with the appropriate plasmids enables co-expression of NMT and a substrate 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).