Chapter 1

INTRODUCTION

1.1 The Proteome

Cells, the individual units of life, comprise the tissues and organs that come together cohesively to form entire complex organisms, such as humans. DNA is the genetic material inside cells. According to the central dogma of biology, DNA is transcribed to RNA, which is translated into protein. Proteins are biological macromolecules composed of twenty canonical amino acid building blocks. The simplicity of these building blocks belies the vast diversity of structure and function that results from combining the twenty amino acids into polypeptide chains ranging from tens to thousands of amino acids in length. In cells, proteins have myriad roles, such as providing structure, responding to the environment, transporting molecules, conferring information, directing and enabling motility, regulating cell decisions, and catalyzing chemical reactions. The entire set of proteins contained by a cell or organism is called the proteome. While there are ~30,000 genes in the human genome, there is 10- to 100-fold greater complexity in the proteome, resulting from alternative transcription and splicing of mRNA and from chemically diverse post-translational modifications[1]. Not only is there greater complexity in the proteome than might reasonably be expected from examination of the genome, but further information is contained in the spatial and temporal changes occurring in the proteome through the creation and destruction of

individual proteins. Disease and environment can both cause dynamic changes in the protein content of a cell (**Figure 1.1**).

1.2 Identification of Proteins Using Mass Spectrometry

Analysis of the proteome provides valuable insight into the state of the cell at a given time. Conventional proteomic investigations enable identification of isolated proteins using mass spectrometry[2, 3]. Generally, proteins are isolated from cells and cut into peptide fragments using proteases. Then mass spectrometry is used to identify the peptides based on the masses of the fragments. Unfortunately, existing methods are unable to identify the entire cellular proteome. Instead, researchers rely on subsets of the proteome to reduce the complexity of the sample and to enable the detection of low-abundance proteins. Isolation of individual organelles and separation of proteins by gel electrophoresis or column chromatography are common methods for reducing the complexity of samples before analysis by mass spectrometry.

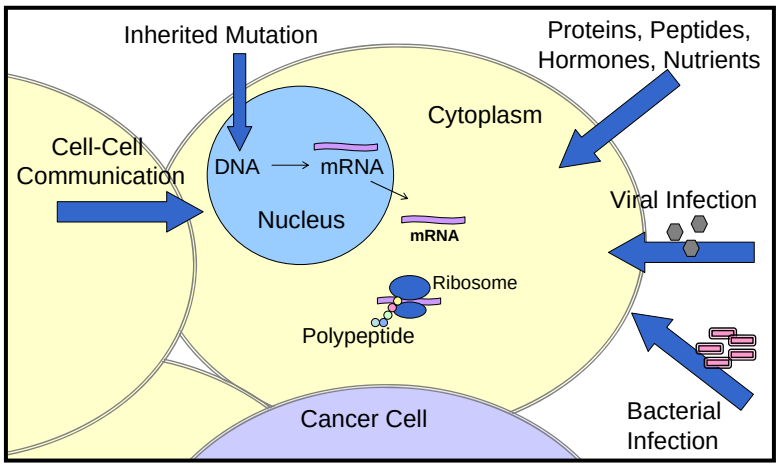

Figure 1.1. Disease and environment can alter the cellular proteome through the translation of new proteins or the degradation of pre-existing proteins.

1.3 Metabolic Tagging of a Subset of the Proteome

Chemical biologists are having a substantial impact on proteomic studies through the synthesis and evaluation of new reactive metabolites as well as the development of new bioconjugation reactions[4]. The knowledge gained by chemical biologists has been applied to tagging subsets of the proteome for isolation and enrichment before identification by mass spectrometry (**Figure 1.2**).

Proteome tagging is facilitated through the introduction of small, reactive metabolic analogues into proteins using the host cell's machinery. Proteins within complex biological mixtures can be specifically tagged with different reactive groups, including ketones[5-8],

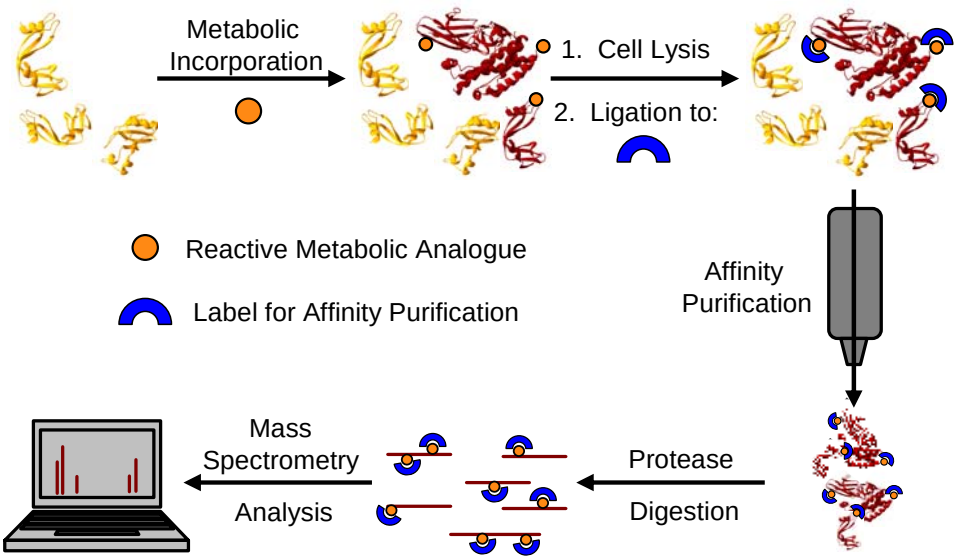

Figure 1.2. Metabolic tagging of proteins for identification by mass spectrometry. First, a reactive metabolic analogue (orange circle) is used to tag a subset of the proteome (red), enabling those proteins to be distinguished from pre-existing proteins (yellow). Second, cells are lysed, and a selective ligation reaction attaches an affinity purification tag (blue symbol) to proteins displaying the metabolic analogue. Next, tagged proteins are enriched by affinity purification and cut into fragments using protease digestion. Finally, peptide fragments are identified and analyzed by mass spectrometry.

azides[9-12], or alkynes[13-15]. After incorporation of the reactive analogue, the protein is ligated to an affinity purification tag. Labeling can be achieved by an azide-alkyne ligation or a Staudinger ligation; both enable selective labeling in complex biological mixtures[10, 16-19]. This metabolic labeling strategy has enabled the identification of many different subsets of the proteome (e.g., glycoproteins[20-22], phosphoproteins[23, 24], farnesylated proteins[25], or fatty acylated proteins[26-29]).

Similarly, reactive amino acid analogues can be incorporated into newly synthesized proteins to track spatial and temporal changes in the proteome[30-35]. Dieterich and coworkers described the use of a reactive methionine (Met) analogue, azidohomoalanine (Aha), to identify newly synthesized proteins in HEK 293 cells[36]. This tagging method is reminiscent of conventional pulse-labeling with radioactive amino acids; the endogenous cellular machinery places a reactive Met analogue at sites normally occupied by Met within polypeptide chains. The newly synthesized proteins, which contained Aha, were then labeled with a biotin affinity purification tag by a copper-catalyzed azide-alkyne ligation[16, 17]. A total of 195 proteins were then identified by mass spectrometry analysis. Cells did not exhibit changes in morphology, viability, or ubiquitination during their pulse with azidohomoalanine. This work demonstrates how a non-canonical amino acid can be used to profile a subset of a mammalian proteome.

1.4 Imaging the Proteome

While there are many methods for identifying proteins, a more complete examination of the proteome would include complementary, time-resolved images of proteins in their natural environment. The distinct characteristics of each proteome

cannot be described simply by protein identification. The spatial and temporal nature of protein synthesis and localization are important aspects to consider.

Spatially-resolved protein profiling and imaging can be achieved using imaging mass spectrometry on cells or tissues[37-39]. This method enables quantitative comparative analysis and visualization of proteins within diseased and healthy tissues, including mammalian brain tumors[38]. While this strategy allows imaging of proteomes, acquisition of high resolution images demands exquisite technical ability, specialized software and equipment, and lengthy analysis.

Fluorescence microscopy provides a far more accessible way to visualize the cellular proteome. Fluorescence imaging of a diverse set of proteins, many of which have unknown sequences and structures, requires a labeling strategy free of genetic manipulation. For this reason, commonly used protein labels, such as green fluorescent protein (GFP) or tetracysteine motifs, cannot be used for proteomic imaging[40, 41]. The proteome can be imaged using autoradiography. Although addition of a radioactive amino acid does not require genetic manipulation or prior knowledge of which proteins are being expressed, imaging by high-resolution autoradiography is technically arduous and time consuming[42].

For these reasons, proteomic imaging may best be accessed through metabolic incorporation of a small, bioorthogonal tag, such as a non-canonical amino acid. In 2005, we first described the use of homopropargylglycine (Hpg) for metabolic tagging of newly synthesized proteins (**Figure 1.3**)[43]. Protein tagging with Hpg is operationally similar

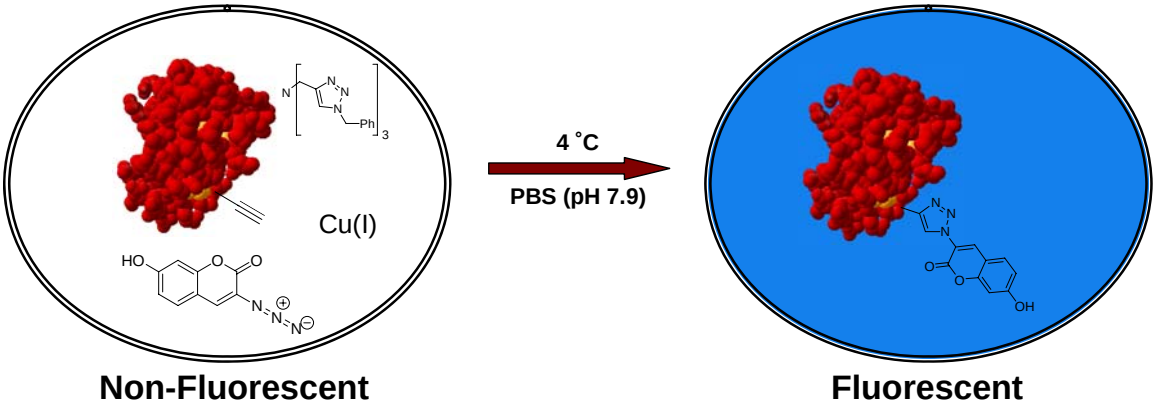

Figure 1.3. Selective dye-labeling of newly synthesized proteins in bacterial cells.

to conventional pulse-labeling with ^{35}S-methionine; the absence of Met synthesis in Met auxotrophic *E. coli* cells and the promiscuity of the methionyl-tRNA synthetase make it straightforward to incorporate Hpg into proteins in competition with Met[13, 14]. However, of the two amino acids, only Hpg undergoes a selective ligation reaction with a membrane-permeant fluorophore for *in situ* imaging. This method is described in Chapter 2. Next, we expanded the method to enable the selective fluorescence labeling and imaging of newly synthesized proteins in a diverse set of mammalian cells (Chapter 3)[44]. More recently, our work has enabled imaging of two distinct subsets of the proteome through two-dye labeling (Chapter 4). Lastly, we developed methods to enable dynamic, live-cell imaging of the proteome in mammalian cells (Chapters 5 and 6).

All of the methods developed in my thesis research should enable the scientific community to gain a deeper understanding of how temporal and spatial changes in protein synthesis affect cellular proteomes. In the future, proteomic imaging may elucidate the role of protein translation in disease[45, 46], memory[47], or protein trafficking[48-50]. The proteomic imaging methods described in these pages could be

used to clarify many processes in biology that rely on spatially and temporally controlled protein synthesis.

1.5 References

1. Walsh, C.T., *Posttranslational Modification of Proteins: Expanding Nature's Inventory*. 2006, Englewood, Colorado: Roberts and Company Publishers.
2. Hamdan, M. and P. Righetti, *Proteomics Today: Protein Assessment and Biomarkers Using Mass Spectrometry, 2D Electrophoresis, and Microarray Technology*. Wiley-Interscience Series in Mass Spectrometry. 2005, Hoboken: John Wiley and Sons, Inc.
3. Warnock, D.E., E. Fahy, and S.W. Taylor, *Identification of protein associations in organelles, using mass spectrometry-based proteomics.* Mass Spectrom. Rev., 2004. **23**(4): p. 259-280.
4. Prescher, J.A. and C.R. Bertozzi, *Chemistry in living systems.* Nature Chemical Biology, 2005. **1**(1): p. 13-21.
5. Lemieux, G.A. and C.R. Bertozzi, *Chemoselective ligation reactions with proteins, oligosaccharides and cells.* Trends Biotechnol., 1998. **16**(12): p. 506-513.
6. Rodriguez, E.C., L.A. Marcaurelle, and C.R. Bertozzi, *Aminooxy-, hydrazide-, and thiosemicarbazide-functionalized saccharides: Versatile reagents for glycoconjugate synthesis.* J. Org. Chem., 1998. **63**(21): p. 7134-7135.
7. Datta, D., et al., *A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins.* J. Am. Chem. Soc., 2002. **124**(20): p. 5652-5653.
8. Wang, L., et al., *Addition of the keto functional group to the genetic code of Escherichia coli.* Proc. Natl. Acad. Sci. USA, 2003. **100**(1): p. 56-61.
9. Kiick, K.L., et al., *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation.* Proc. Natl. Acad. Sci. USA, 2002. **99**(1): p. 19-24.
10. Saxon, E. and C.R. Bertozzi, *Cell surface engineering by a modified Staudinger reaction.* Science, 2000. **287**(5460): p. 2007-2010.
11. Chin, J.W., et al., *Addition of p-azido-L-phenylaianine to the genetic code of Escherichia coli.* J. Am. Chem. Soc., 2002. **124**(31): p. 9026-9027.
12. Speers, A.E. and B.F. Cravatt, *Profiling enzyme activities in vivo using click chemistry methods.* Chemistry & Biology, 2004. **11**(4): p. 535-546.

13. Kiick, K.L., R. Weberskirch, and D.A. Tirrell, *Identification of an expanded set of translationally active methionine analogues in Escherichia coli.* Febs Letters, 2001. **502**(1-2): p. 25-30.

14. van Hest, J.C.M., K.L. Kiick, and D.A. Tirrell, *Efficient incorporation of unsaturated methionine analogues into proteins in vivo.* J. Am. Chem. Soc., 2000. **122**(7): p. 1282-1288.

15. Deiters, A. and P.G. Schultz, *In vivo incorporation of an alkyne into proteins in Escherichia coli.* Bioorg. Med. Chem. Lett., 2005. **15**(5): p. 1521-1524.

16. Rostovtsev, V.V., et al., *A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes.* Angew. Chem. Int. Ed., 2002. **41**(14): p. 2596-2599.

17. Tornoe, C.W., C. Christensen, and M. Meldal, *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., 2002. **67**(9): p. 3057-3064.

18. Agard, N.J., J.A. Prescher, and C.R. Bertozzi, *A strain-promoted [3+2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems.* J. Am. Chem. Soc., 2004. **126**(46): p. 15046-15047.

19. van Swieten, P.F., et al., *Bioorthogonal organic chemistry in living cells: Novel strategies for labeling biomolecules.* Organic & Biomolecular Chemistry, 2005. **3**(1): p. 20-27.

20. Khidekel, N., et al., *Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics.* Nat Chem Biol, 2007. **3**(6): p. 339-348.

21. Khidekel, N., et al., *Exploring the O-GlcNAc proteome: Direct identification of O-GlcNAc-modified proteins from the brain.* Proc. Natl. Acad. Sci. USA, 2004. **101**(36): p. 13132-13137.

22. Dube, D.H., et al., *Probing mucin-type O-linked glycosylation in living animals.* Proc. Natl. Acad. Sci. USA, 2006. **103**(13): p. 4819-4824.

23. Green, K.D. and M.K.H. Pflum, *Kinase-catalyzed biotinylation for phosphoprotein detection.* J. Am. Chem. Soc., 2007. **129**(1): p. 10-11.

24. Warthaka, M., P. Karwowska-Desaulniers, and M.K.H. Pflum, *Phosphopeptide Modification and Enrichment by Oxidation-Reduction Condensation.* ACS Chem. Biol., 2006. **1**(11): p. 697-701.

25. Kho, Y., et al., *A tagging-via-substrate technology for detection and proteomics of farnesylated proteins.* PNAS, 2004. **101**(34): p. 12479-12484.

26. Hang, H.C., et al., *Chemical probes for the rapid detection of fatty-acylated proteins in mammalian cells.* J. Am. Chem. Soc., 2007. **129**(10): p. 2744-2755.

27. Martin, D.D.O., et al., *Rapid detection, discovery, and identification of post-translationally myristoylated proteins during apoptosis using a bio-orthogonal azidomyristate analog.* FASEB J., 2007. **22**: p. 1-10.

28. Martin, D.D.O., et al., *Rapid detection, discovery, and identification of post-translationally myristoylated proteins during apoptosis using a bio-orthogonal azidomyristate analog.* FASEB J., 2007: p. fj.07-9198com.

29. Kostiuk, M.A., et al., *Identification of palmitoylated mitochondrial proteins using a bio-orthogonal azido-palmitate analogue.* FASEB J., 2007. **22**: p. 1-10.

30. Link, A.J., M.L. Mock, and D.A. Tirrell, *Non-canonical amino acids in protein engineering.* Current Opinion in Biotechnology, 2003. **14**(6): p. 603-609.

31. Hohsaka, T. and M. Sisido, *Incorporation of non-natural amino acids into proteins.* Current Opinion in Chemical Biology, 2002. **6**(6): p. 809-815.

32. Hendrickson, T.L., V. de Crecy-Lagard, and P. Schimmel, *Incorporation of nonnatural amino acids into proteins.* Annual Review of Biochemistry, 2004. **73**: p. 147-176.

33. Budisa, N., *Engineering the Genetic Code: Expanding the Amino Acid Repertoire for the Design of Novel Proteins*. 2006: Wiley-VCH.

34. Budisa, N., *Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire.* Angew. Chem. Int. Ed., 2004. **43**(47): p. 6426-6463.

35. Mathews, M.B., N. Sonenberg, and J.W.B. Hershey, *Translational Control in Biology and Medicine*. 2007: Cold Springs Harbor Laboratory Press.

36. Dieterich, D.C., et al., *Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal non-canonical amino acid tagging (BONCAT).* Proc. Natl. Acad. Sci. USA, 2006. **103**(25): p. 9482-9487.

37. Chaurand, P., S.A. Schwartz, and R.M. Caprioli, *Assessing protein patterns in disease using imaging mass spectrometry.* Journal of Proteome Research, 2004. **3**(2): p. 245-252.

38. Stoeckli, M., et al., *Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues.* Nat. Med., 2001. **7**(4): p. 493-496.

39. Pacholski, M.L. and N. Winograd, *Imaging with mass spectrometry.* Chem. Rev., 1999. **99**(10): p. 2977-3006.

40. Griffin, B.A., S.R. Adams, and R.Y. Tsien, *Specific covalent labeling of recombinant protein molecules inside live cells.* Science, 1998. **281**(5374): p. 269-272.

41. Tsien, R.Y., *The green fluorescent protein.* Annual Review of Biochemistry, 1998. **67**: p. 509-544.

42. Caro, L.G., J.A. Kolb, and R.P. van Tubergen, *High resolution autoradiography.* J. Cell Biol., 1962. **15**: p. 173-188.

43. Beatty, K.E., et al., *Selective dye-labeling of newly synthesized proteins in bacterial cells.* J. Am. Chem. Soc., 2005. **127**(41): p. 14150-14151.

44. Beatty, K.E., et al., *Fluorescence visualization of newly synthesized proteins in mammalian cells.* Angew. Chem. Int. Ed., 2006. **45**(44): p. 7364-7367.

45. Angelichio, M.J. and A. Camilli, *In Vivo Expression Technology.* Infect. Immun., 2002. **70**(12): p. 6518-6523.

46. Hay, N. and N. Sonenberg, *Upstream and downstream of mTOR.* Genes Dev., 2004. **18**(16): p. 1926-1945.

47. Steward, O. and E.M. Schuman, *Compartmentalized synthesis and degradation of proteins in neurons.* Neuron, 2003. **40**: p. 347-359.

48. Palade, G., *Intracellular Aspects of the Process of Protein Synthesis* Science, 1975. **189**(4200): p. 347-358.

49. Lippincott-Schwartz, J., T.H. Roberts, and K. Hirschberg, *Secretory Protein Trafficking and Organelle Dynamics in Living Cells.* Annu. Rev. Cell Dev. Biol., 2000. **16**(1): p. 557-589.

50. Baskin, J.M., et al., *Copper-free click chemistry for dynamic in vivo imaging.* Proc. Natl. Acad. Sci. USA, 2007. **104**(43): p. 16793-16797.