

---

CHAPTER 1

*Chemical Tools for Temporally  
and Spatially Resolved Mass  
Spectrometry-Based Proteomics*

---

This chapter first appeared as an article in *Annals of Biomedical Engineering*: Kai P. Yuet and David A. Tirrell, *Annals of Biomedical Engineering*, February 2014, Volume 42, Issue 2, Pages 299–311 (doi: 10.1007/s10439-013-0878-3).

## 1.1 Abstract

Accurate measurements of the abundances, synthesis rates, and degradation rates of cellular proteins are critical for understanding how cells and organisms respond to changes in their environments. Over the past two decades, there has been increasing interest in the use of mass spectrometry for proteomic analysis. In many systems, however, protein diversity as well as cell and tissue heterogeneity limit the usefulness of mass spectrometry-based proteomics. As a result, researchers have had difficulty in systematically identifying proteins expressed within specified time intervals, or low abundance proteins expressed in specific tissues or in a few cells in complex microbial systems. In this chapter, we present recently-developed tools and strategies that probe these two subsets of the proteome: proteins synthesized during well-defined time intervals – *temporally resolved proteomics* – and proteins expressed in predetermined cell types, cells, or cellular compartments – *spatially resolved proteomics* – with a focus on chemical and biological mass spectrometry-based methodologies.

## 1.2 Introduction

Messenger RNA (mRNA) profiling at the systems level (transcriptomics) with microarray or deep-sequencing technologies offers a high-throughput route to the analysis of gene expression [1, 2]. However, transcriptomic methods are blind to post-transcriptional phenomena such as translational regulation, protein modification, protein-protein interactions, or protein interactions with other molecular components. Furthermore, because mRNA abundance correlates poorly with protein abundance [3, 4, 5], reliable quantitative information about changes in protein abundance cannot be derived from microarray analysis.

Recent advances in genomic sequencing and high-resolution mass spectrometry have enabled rapid progress in the study of proteins and their abundances, modifications, interactions, and functions [6, 7]. Sample analysis workflows that employ high-resolution liquid chromatography-tandem mass spectrometry (LC-MS/MS) use site-specific endopeptidases (e.g., trypsin or Lys-C) to convert mixtures of proteins into peptides [8]. Microscale or nanoscale high-performance liquid chromatography (HPLC) [9] and/or ion exchange (IEX) chromatography [10] separates the peptides and injects them into the mass spectrometer. After an initial ion scan ( $MS^1$ ), peptide ions fragment in the mass spectrometer and yield secondary “MS/MS” ( $MS^2$ ) spectra. Peptide search engines like Mascot or Andromeda match the MS/MS spectra with theoretical MS/MS spectra and recover the corresponding peptide sequences from a database [11, 12]. Finally, proteins are identified using two or more unique peptides. Mass spectrometry-based proteomic methods have been applied to a wide range of problems including deciphering the protein composition of organelles [13], systematically mapping protein-protein interactions [14], and large-scale decoding of post-translational events in response to stimuli [15].

As of this writing, no eukaryotic proteome has been mapped with 100% coverage [14, 16]. The dynamic range of protein expression can span many orders of magnitude (e.g., up to 12 orders of magnitude in serum [17]). The resulting variations in protein abundance challenge the dynamic range and sequencing speed of contemporary mass spectrometers. Sequence coverage becomes increasingly important in the analysis of higher eukaryotes where proteins exhibit high levels of sequence homology due to evolution of protein families, alternative splicing, and differential processing [18, 19, 20]. Furthermore, while genome-wide transcriptome analyses are now routinely performed on small samples of RNA, even from single cells [21], low abundance proteins cannot be amplified to improve identification rates. To expand the dynamic range of identified proteins, researchers have focused on information-rich

subsets of the proteome, such as the glycoproteome [22], phosphoproteome [23], and ubiquitome [24]. In this chapter, we highlight recent work that leverages mass spectrometry and chemical biology to examine two additional subsets of the proteome: proteins synthesized during predetermined time intervals – *temporally resolved proteomics* – and proteins expressed in specific cell types, cells, or cellular compartments – *spatially resolved proteomics*.

## 1.3 Temporally Resolved Proteomic Analysis

### 1.3.1 Stable-Isotope Labeling with Amino Acids in Cell Culture

Interest in quantitative modeling and analysis of biological processes has motivated the development of tools for quantitative mass spectrometry-based proteomics such as “label-free” and isotopically labeled protein profiling [25, 26]. Comparison of protein abundances in different samples is most accurately accomplished by using stable isotopes [25, 26, 27]. Stable-isotope labeling of peptides relies on either chemical methods that attach isotopically labeled linkers to peptides (e.g., ICAT [28] or iTRAQ [29]) or metabolic methods that incorporate isotopically labeled amino acids into peptides (e.g.,  $^{15}\text{N}$ -labeling [30] or SILAC [31]).

Stable-isotope labeling by amino acids in cell culture (SILAC, **Figure 1.1, Top Left**) introduces the label at the amino acid level and overcomes complex isotopic clusters found in alternative metabolic methods like  $^{15}\text{N}$ -labeling. Distinct isotopologs of the labeled amino acids – typically “light” arginine ( $^{12}\text{C}_6^{14}\text{N}_4^1\text{H}_{14}^{16}\text{O}_2$ ) and “light” lysine ( $^{12}\text{C}_6^{14}\text{N}_2^1\text{H}_{14}^{16}\text{O}_2$ ) in one sample, and “heavy” arginine ( $^{13}\text{C}_6^{15}\text{N}_4^1\text{H}_{14}^{16}\text{O}_2$ ) and heavy lysine ( $^{13}\text{C}_6^{15}\text{N}_2^1\text{H}_{14}^{16}\text{O}_2$ ) in the other – are added to the samples of interest. After several generations of growth, all cellular proteins have uniformly incorporated the labeled amino acids. Cells are lysed and subjected to proteolysis, and the resulting peptide pools are mixed. In theory, both the “heavy” and “light” forms

of each peptide should behave identically with respect to processing and LC-MS/MS analysis. However, the origin of each peptide can be determined from the mass of its isotope label, and the relative signal intensities of the “heavy” and “light” peptides indicate the relative abundances of the corresponding proteins in the two samples. Its simplicity and accuracy have made SILAC an increasingly popular method, both for cell culture-based quantitative proteomics [32], and for whole organism quantitative proteomics [3, 33, 34, 35, 36].

### ***1.3.2 Repurposing SILAC for Temporally Resolved Proteomic Analysis***

Because “old” and “new” copies of cellular proteins are chemically indistinguishable, selective analysis of the subset of the proteome that is expressed within a specified time interval is challenging. A classic method for quantifying specific protein turnover rates involves pulse labeling cells with a radiolabeled amino acid (typically  $^{35}\text{S}$ -methionine or cysteine) [37, 38]. Only those proteins synthesized during the pulse incorporate the radioactive element, and their fates can be monitored by standard radioisotope methods. Analogous pulse labeling with stable isotopes can also determine protein turnover or transport (**Figure 1.1, Top Middle**) [39, 40, 41].

Pulse labeling with amino acids and nucleosides can be combined to yield an unbiased and comprehensive picture of protein and mRNA dynamics. To determine if the cellular abundance of proteins is predominantly controlled at the level of transcription or translation, Schwanhäusser *et al.* pulse labeled mammalian cells with the nucleoside analog 4-thiouridine (**Figure 1.2**) and heavy arginine and lysine to quantify absolute mRNA and protein abundances, half-lives, and transcription and translation rates for more than 5000 genes in mouse fibroblast cells [42]. The thio-substituted nucleoside 4-thiouridine is not a natural component of nucleic acids but is incorporated into RNA biosynthetically; only newly synthesized RNAs are thio-labeled and can be tagged and purified using commercially available reagents [43, 44]. Although mRNA

and protein levels correlated better than in previous studies [3, 4, 5, 45, 46, 47], their half-lives showed no overall correlation. Nevertheless, the authors note that genes with similar combinations of mRNA and protein half-lives share common functions, suggesting that half-lives evolved under similar constraints. For example, many genes involved in constitutive processes like translation, respiration, and central metabolism have stable mRNAs and proteins, consistent with the requirement for conservation of resources. On the other hand, genes involved in transcription regulation, signaling, chromatin modification, and cell cycle-specific processes have unstable mRNAs and proteins, consistent with the requirement for rapid regulation. Correlation of protein abundance with mRNA levels and with translation rates showed that protein copy numbers are determined primarily at the level of translation.

### 1.3.3 Pulsed SILAC

Unlike pulse labeling with a single label to determine protein turnover or transport in the same sample, pulsed stable-isotope labeling with amino acids in cell culture (pSILAC, **Figure 1.1, Top Right**) quantifies differences in protein synthesis between different samples integrated over the measurement time after the pulse. In pSILAC, cells in two different samples previously cultured in light media are transferred to two different pulse media: one containing heavy amino acids (typically arginine-10:  $^{13}\text{C}_6^{15}\text{N}_4^1\text{H}_{14}^{16}\text{O}_2$  and lysine-8:  $^{13}\text{C}_6^{15}\text{N}_2^1\text{H}_{14}^{16}\text{O}_2$ ) and the other containing medium-heavy amino acids (e.g., arginine-6:  $^{13}\text{C}_6^{14}\text{N}_4^1\text{H}_{10}^2\text{H}_4^{16}\text{O}_2$  and lysine-4:  $^{12}\text{C}_6^{14}\text{N}_2^1\text{H}_{10}^2\text{H}_4^{16}\text{O}_2$ ) [48, 49]. During labeling, only newly synthesized proteins incorporate either the heavy or the medium-heavy amino acids. Peptides derived from heavy or medium-heavy proteins are distinguishable during the initial ion scan from preexisting, light proteins on the basis of the mass difference introduced by the isotope label. Intensity ratios for heavy and medium-heavy peptides directly indicate the relative abundances of the corresponding newly synthesized proteins in

the samples of interest. Light preexisting peptides are identified but ignored during quantification.

Selbach and coworkers introduced pSILAC in 2008 as a method to measure changes in protein translation involved in cellular iron homeostasis [49]. Since then, researchers have used pSILAC to assess protein dynamics associated with microRNA overexpression [48, 50, 51], monocyte-macrophage differentiation [52], hyperglycemia-induced stress [53] and mammalian target of rapamycin inhibition [54]. Selbach *et al.* utilized pSILAC to measure changes in the synthesis of several thousand proteins in HeLa cells in response to either microRNA transfection or endogenous microRNA knockdown [48]. HeLa cells cultivated in light media were first transfected with miR-1, miR-155, miR-16, miR-30a, or let-7b, or mock transfected. Eight hours later, cells were transferred to media containing either medium-heavy or heavy amino acids for 24 hours. Unexpectedly, pSILAC in this study and subsequent studies revealed that overexpression of a single microRNA can lead to repression of hundreds of proteins; however, the repressive effect was relatively small and rarely exceeded fourfold [50, 51]. These results cast microRNAs as general orchestrators that tune cellular physiology and metabolism in subtle ways in response to specific cues.

#### ***1.3.4 Bio-Orthogonal Non-Canonical Amino Acid Tagging***

Under some circumstances (e.g., for short labeling times), the dynamic range and complexity of the proteome can preclude reliable identification and analysis of low-abundance isotope-labeled proteins in the presence of more abundant preexisting proteins. Moreover, the use of multiple labels increases the number of distinct peptide forms and the computational complexity associated with protein identification. These limitations can be addressed by *selective enrichment* and identification of newly synthesized proteins.

In 2006, Dieterich *et al.* introduced the bio-orthogonal non-canonical amino acid tagging (BONCAT, **Figure 1.1, Bottom Left**) strategy to enable selective enrichment and identification of newly synthesized proteins in cells and tissues [55]. This strategy relies on bio-orthogonal functional groups – functional groups that react rapidly and selectively with each other but remain inert to the functional groups normally found in biological systems [56]. An exemplary bio-orthogonal group, the azide is small, kinetically stable and absent from living systems, yet can be modified easily and selectively through the Staudinger ligation with triarylphosphine reagents [57], the copper(I)-catalyzed “click” cycloaddition with terminal alkynes [58, 59] or the strain-promoted “click” cycloaddition with strained cyclooctynes [60].

BONCAT employs a two-stage procedure. First, pulse labeling of cells or tissues with the azide-bearing methionine (Met, **Figure 1.2**) analog azidohomoalanine (Aha, **Figure 1.2**) or the alkyne-bearing methionine analog homopropargylglycine (Hpg, **Figure 1.2**) enables metabolic incorporation of either azide or alkyne functional groups into the proteome. In bacterial and mammalian systems, wild-type methionyl-tRNA synthetases (MetRSs) are capable of appending the non-canonical amino acid Aha (or Hpg) to cognate transfer RNAs ( $\text{tRNA}^{\text{Met}}$ ) [55, 61]. Aha (or Hpg) is thereby incorporated into proteins made during the Aha (or Hpg) pulse. Second, a bio-orthogonal ligation with the complementary reactive group conjugated to a probe enables detection of Aha- or Hpg-tagged proteins. Treatment of azide-tagged proteins with alkyne-functionalized fluorescent dyes permits visualization of newly synthesized proteins [62, 63, 64]. Alternatively, treatment of azide-tagged proteins with alkyne-functionalized affinity reagents allows selective enrichment of proteins made during the Aha pulse. Enriched proteins can then be identified by LC-MS/MS [65, 66].

BONCAT has been used in mammalian cells [55], in both unicellular [67, 68] and multicellular organisms [69], and even adapted to measure genome-wide nucleosome turnover dynamics in *Drosophila* S2 cells [70]. Here, we highlight a few recent examples with an emphasis on proteomic discovery. Zhang *et al.* reported a tandem labeling (Aha and alkyne-functionalized palmitic acid) and detection method to monitor the dynamic acylation of Lck, an N-myristoylated and S-palmitoylated non-receptor tyrosine kinase required for T-cell activation [71]. Liu *et al.* expanded on this concept by enriching proteins made in Jurkat cells during a specified time interval, then progressively monitoring their post-translational modifications over time by quantitative in-gel fluorescence scanning with Aha/Hpg and eight additional azide- and alkyne-functionalized post-translational modification probes [72]. The up-regulation of myristoylated protein kinase A was found to be intimately linked to butyric acid-induced apoptosis.

As local protein synthesis is critical for long-term functional synaptic changes, BONCAT is an attractive tool for neurobiological studies. Tcherkezian *et al.* examined the colocalization of DCC (Deleted in Colorectal Cancer) transmembrane receptors with newly synthesized proteins in cultured commissural axon growth cones with Aha-labeling [73]. Hodas *et al.* used BONCAT to identify proteins translated in intact hippocampal neuropil sections upon treatment with the selective D1/D5 dopamine receptor agonist SKF81297 [74]. Yoon *et al.* developed DIGE-NCAT – a combination of 2D difference gel electrophoresis (2D-DIGE) and BONCAT – to examine changes in the proteome of *Xenopus* retinal ganglion cell (RGC) axons in response to stimulation with Engrailed-1 [75]. Engrailed-1 belongs to a family of transcription factors previously shown to cause rapid translation-dependent guidance responses in RGC axons [76]. The authors first severed distal portions of the axon bundles from *Xenopus* eyes and stimulated them with Engrailed-1 for 1 hour along with the addition of Aha. Next, Aha-tagged axonally synthesized proteins were treated with an alkynyl dye and

resolved on 2D-DIGE. Compared to control gels, spots with the greatest difference in fluorescence were analyzed by matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS). Surprisingly, this strategy revealed that the intermediate filament protein lamin B2 (LB2) – normally associated with the nuclear membrane – is axonally synthesized in response to stimulation. Coupled with the finding of LB2s association with mitochondria, these results suggest that LB2-promotion of mitochondrial function is needed for axon maintenance.

### ***1.3.5 Quantitative Non-Canonical Amino Acid Tagging***

How can we quantitatively study proteomic changes during short time intervals (e.g., in response to a stimulus) without the complications that arise from the abundance of preexisting proteins? As labeling times decrease, co-eluting preexisting peptides increasingly obscure low abundance pSILAC-labeled peptides during the first mass spectrometry scan. To overcome this limitation, pSILAC and BONCAT have been combined in an approach designated Quantitative Non-Canonical Amino Acid Tagging (QuaNCAT, **Figure 1.1, Bottom Right**). In a QuaNCAT experiment, two parallel populations of cells in light media are transferred for a limited time to either medium-heavy or heavy media that also contain Aha [77, 78]. During the labeling period, newly synthesized proteins incorporate either the heavy or medium-heavy amino acid as well as Aha. Through BONCAT enrichment of the combined protein pools, preexisting light peptides are greatly reduced in abundance. As in pSILAC, ratios of intensities of heavy and medium-heavy peptides directly indicate the relative abundances of newly synthesized proteins in the two samples.

To determine the benefit of selectively enriching and quantifying secreted proteins, Eichelbaum *et al.* labeled two human cell lines (PC3 and WPMY-1) with Aha and isotopologs of arginine and lysine for 24 hours [77]. With enrichment via on-bead azide-alkyne cycloaddition and trypsin digestion, 684 secreted proteins were quantified with high correlation ( $R = 0.96$ ) between biological duplicates. Without enrichment, only 22 proteins were quantified with low correlation ( $R = 0.02$ ). The authors identified as many as 500 proteins, even at short labeling times (2 hours), including several known lipopolysaccharide (LPS) effector proteins as well as many other previously unassociated proteins during LPS stimulation of mouse macrophages. Howden *et al.* tested QuaNCAT by examining changes in expression of more than 600 proteins enriched from freshly isolated human CD4-positive T cells stimulated by activation with phorbol 12-myristate 13-acetate and ionomycin during 2 and 4 hour pulses [78]. Many transcription factors and transcriptional regulators, cytokines, activation surface markers, protein chaperones, and proteins involved in cytoskeleton dynamics and vesicle transport, were among the proteins found to be substantially increased in expression following stimulation. While less than 1% of the proteome was labeled in a 2 hour pulse, the post-enrichment protein pool consisted of 10-20% heavy and medium-heavy proteins. The ability to resolve protein abundance changes over very short time spans should prove useful when studying systems previously thought to be “unquantifiable” by SILAC such as short-lived progenitor or primary cells [79].

### 1.3.6 *O*-Propargyl-Puromycin Labeling

Salic and coworkers have developed an alternative approach to the selective labeling of newly synthesized proteins, in which a puromycin analog (*O*-propargyl-puromycin, **Figure 1.3, Top**) bearing a terminal alkyne is used to label nascent polypeptide chains [80]. Puromycin mimics aminoacyl-tRNAs and terminates translational elongation after entering the acceptor site of the ribosome. Salic and coworkers showed

that O-propargyl-puromycin not only terminates nascent polypeptide chains but also appends an alkyne label at the C-terminus. Liu and coworkers injected mice intraperitoneally with O-propargyl-puromycin and harvested tissues 1 hour later. Tissues from these mice displayed specific patterns of O-propargyl-puromycin incorporation into nascent proteins. For example, in the small intestine, the most intense labeling occurred in cells in the crypts and at the base of intestinal villi, consistent with the high proliferative and secretory activity of these cells (**Figure 1.3, Bottom**).

## 1.4 Spatially Resolved Proteomic Analysis

### 1.4.1 *Coupling Flow Cytometry and Mass Spectrometry*

Proteomic analysis of rare cells in heterogeneous environments presents important, difficult challenges. Combining flow cytometry with quantitative proteomics provides a solution in some systems of this kind. For example, Rechavi *et al.* combined quantitative proteomics and high-purity cell sorting to discover proteins transferred from human B cells to natural killer cells [81]. These authors developed a strategy called “trans-SILAC”, in which one cell type – capable of transferring proteins – is labeled with heavy arginine and lysine and the other cell type – the recipient of transferred proteins – remains unlabeled. To initiate contact- and actin cytoskeleton-dependent protein transfer, the heavy “donor” B cells and freshly isolated light “recipient” natural killer cells were co-incubated for 1.5 hours. Fluorescence-activated cell sorting separated recipient cells from donor cells before the proteomic workflow, and transferred proteins in the recipient cells were identified by their mass shifts. Analysis of the transferred proteins revealed significant enrichment for the annotation term “MHC class II protein complex”, as expected.

### 1.4.2 *Cell-Selective BONCAT*

Although some cell types can be isolated for proteomic analysis through cell sorting or laser-capture techniques, others cannot. Subsets of neurons or glia in the central nervous system, for example, are difficult to isolate by dissection or dissociation methods. Or perhaps one would like to analyze proteomic responses of pathogenic bacteria hiding inside host cells, or of specific cells in multicellular animals, without interference from proteins derived from the host or from all of the other cells in the animal. Can we perform these kinds of tasks without prior separation of the cells of interest?

Promising solutions to analogous problems in genomics and transcriptomics rely on spatially restricted enzymatic labeling. In 2010, Henikoff and coworkers developed INTACT (isolation of nuclei tagged in specific cell types) – a method that allows affinity-based isolation of nuclei from individual cell types in a tissue [82]. In INTACT, the spatial restriction of the *Escherichia coli* biotin ligase BirA and a nuclear targeting fusion protein results in biotin-labeled nuclei in the cell type of choice. First described in *Arabidopsis*, INTACT has been extended to profile gene expression and histone modifications in adult *Caenorhabditis* muscle [83] and Kenyon cells and octopaminergic neurons in the adult *Drosophila* brain [84].

For RNA, Miller *et al.* developed TU-tagging, a method that allows affinity-based isolation of RNA from individual cell types of a tissue [85]. In TU-tagging, the spatial restriction of the *Toxoplasma gondii* nucleotide salvage enzyme uracil phosphoribosyltransferase enables RNA in the cell type of choice to be labeled with 4-thiouracil (**Figure 1.2**). Unlike 4-thiouridine, 4-thiouracil is not recognized by the endogenous biosynthetic machinery. Recently, Gay *et al.* used TU-tagging to purify transcripts from rare (<5%) cells, such as *Tie2:Cre+* brain endothelia/microglia in mice [86].

Analogous strategies have been developed for cell-selective proteomic analysis. Because endogenous methionyl-tRNA synthetases charge Aha to cognate tRNAs in all cell types, Aha-based BONCAT is not cell-selective – newly synthesized proteins in all cell types are labeled. To enable cell-selective proteomic analysis, Tirrell and coworkers engineered a family of mutant *E. coli* MetRSs capable of appending the azide-bearing methionine analog azidonorleucine (Anl, **Figure 1.2**) to tRNA<sup>Met</sup> [87, 88]. Anl is not a good substrate for any of the wild-type aminoacyl-tRNA synthetases in bacteria to mammals; it is excluded from proteins made in wild-type cells but incorporated readily into proteins made in cells that express an appropriately engineered MetRS. Anl-labeling does not require depletion of methionine if the mutant MetRS activates Anl faster than methionine.

Ngo *et al.* first achieved cell type-selectivity by outfitting an *E. coli* strain with the L13N/Y260L/H301L mutant form of the *E. coli* MetRS (NLL-*Ec*MetRS). Proteins made in this strain could be labeled with Anl in co-culture with murine alveolar macrophages, which were not labeled (**Figure 1.4**) [89]. Bacterial proteins were effectively separated from murine proteins by treatment of mixed lysates with alkyne-functionalized biotin reagents and subsequent affinity chromatography on NeutrAvidin resin.

System-wide identification of bacterial proteins expressed or secreted during infection can provide new insight into mechanisms of bacterial pathogenesis. Grammel *et al.* studied cultures of Raw264.7 murine macrophages infected with *Salmonella typhimurium* (a Gram-negative intracellular pathogen) outfitted with NLL-*Ec*MetRS [90]. Cultures were pulse-labeled with 2-aminooctynoic acid (Aoa, **Figure 1.2**), an alkyne analog of Anl, for 1 hour. Of the 218 proteins identified, 185 (85%) were *Salmonella* proteins and 33 were mouse proteins. Five of the *Salmonella* proteins (SodM, SsrB, SseA, PipB2, and PhoP) had been previously described as virulence factors.

In more recent work, Ngo and coworkers demonstrated that heterologous expression of the NLL-*EcMetRS* enables incorporation of Anl into proteins expressed in human (HEK293) cells, permitting enrichment and visualization of proteins made during various stages of the cell cycle [91]. Interestingly, Anl replaces methionine selectively at N-terminal positions – not at internal sites. Site-selectivity occurs because NLL-*EcMetRS* catalyzes aminoacylation only of the mammalian initiator tRNA<sup>Met</sup>, not the mammalian elongator tRNA<sup>Met</sup>. Through judicious selection of regulatory elements, systems of this kind will enable cell-selective or “cell-state-selective” [92] interrogation of protein synthesis in co-cultures of mammalian cells, in virally transfected tissues or even in living animals.

To expand the set of tools available for cell-selective BONCAT, Truong *et al.* engineered a MetRS variant capable of activating propargylglycine (Pra, **Figure 1.2**) but not Anl [93]. Pra is an alkynyl amino acid smaller than methionine and is not activated by wild-type aminoacyl-tRNA synthetases or by NLL-*EcMetRS*. Using directed evolution, Truong *et al.* isolated a MetRS variant (designated propargylglycyl-tRNA synthetase; PraRS, L13P/A256G/P257T/Y260Q/H301F/A331V/ $\Delta$ 548E) capable of near-quantitative replacement of Met by Pra in proteins. In protein mixtures that have been labeled both with Anl and with Pra, treatment with cyclooctyne-functionalized probes selectively tags Anl side chains, and subsequent treatment with azide-functionalized probes selectively tags Pra side chains. By using one promoter to drive expression of NLL-*EcMetRS* in one cell and a different promoter to drive expression of PraRS in another cell, researchers can perform differential, cell-selective BONCAT in complex multicellular systems without prior separation.

Restricting expression of mutant methionyl-tRNA synthetases by using promoters active only in specific cells or tissues should restrict Anl/Aoa labeling to those cells or tissues. Additional specificity should be possible by exploiting the combinatorial action of multiple regulators. To this end, Mahdavi *et al.* bisected the NLL-*EcMetRS*

and found several split variants capable of charging Anl to tRNA<sup>Met</sup> [94]. Because labeling requires expression of both the N- and C-terminal fragments of the synthetase, this system allows the investigator to restrict labeling to cells in which two promoters of interest are active.

### ***1.4.3 Ascorbate Peroxidase Labeling***

Ting and coworkers have recently used an engineered variant of ascorbate peroxidase (APEX) to selectively tag proteins that localize to the mitochondrial matrix in living cells [95]. The Ting laboratory had previously used APEX to catalyze the H<sub>2</sub>O<sub>2</sub>-dependent polymerization of diaminobenzidine to provide contrast in electron microscopy [96]. They genetically targeted APEX to the mitochondrial matrix of human embryonic kidney (HEK) cells, initiated labeling by adding biotin-phenol and H<sub>2</sub>O<sub>2</sub>, and stopped labeling after 1 minute by cell fixation or lysis (**Figure 1.5**). APEX processes biotin-phenol in a hydrogen peroxide-dependent manner to yield highly reactive products that covalently link biotin to electron-rich amino acids such as tyrosine, tryptophan, histidine, and cysteine. Unlike cell-selective BONCAT, APEX labeling is not time-selective. Any proteins accessible to the phenoxy radical – including pre-existing proteins made before the biotin-phenol pulse – are biotinylated in the 1 minute “snapshot”. Enrichment and mass spectrometry of the biotinylated proteins led to identification of 464 known mitochondrial proteins as well as 31 that were not previously known to localize to mitochondria. Demonstrating APEX-based proteomics as a viable discovery tool, a random subset of these 31 “mitochondrial orphans” were verified by fluorescence imaging to have complete or partial mitochondrial localization.

## 1.5 Conclusions

This chapter highlights chemical biological strategies that allow researchers to identify, isolate, and quantitatively analyze proteins within defined windows of time and space. Many of these strategies rely on amino acids with isotopic signatures or functional groups that are not found in biological systems. Incorporation of these “tags” permits the investigator to distinguish proteins made in different time intervals or in different cells or organelles. In combination with these tagging methods, mass spectrometry becomes an even more powerful tool for addressing complex biological questions.

## 1.6 Figures

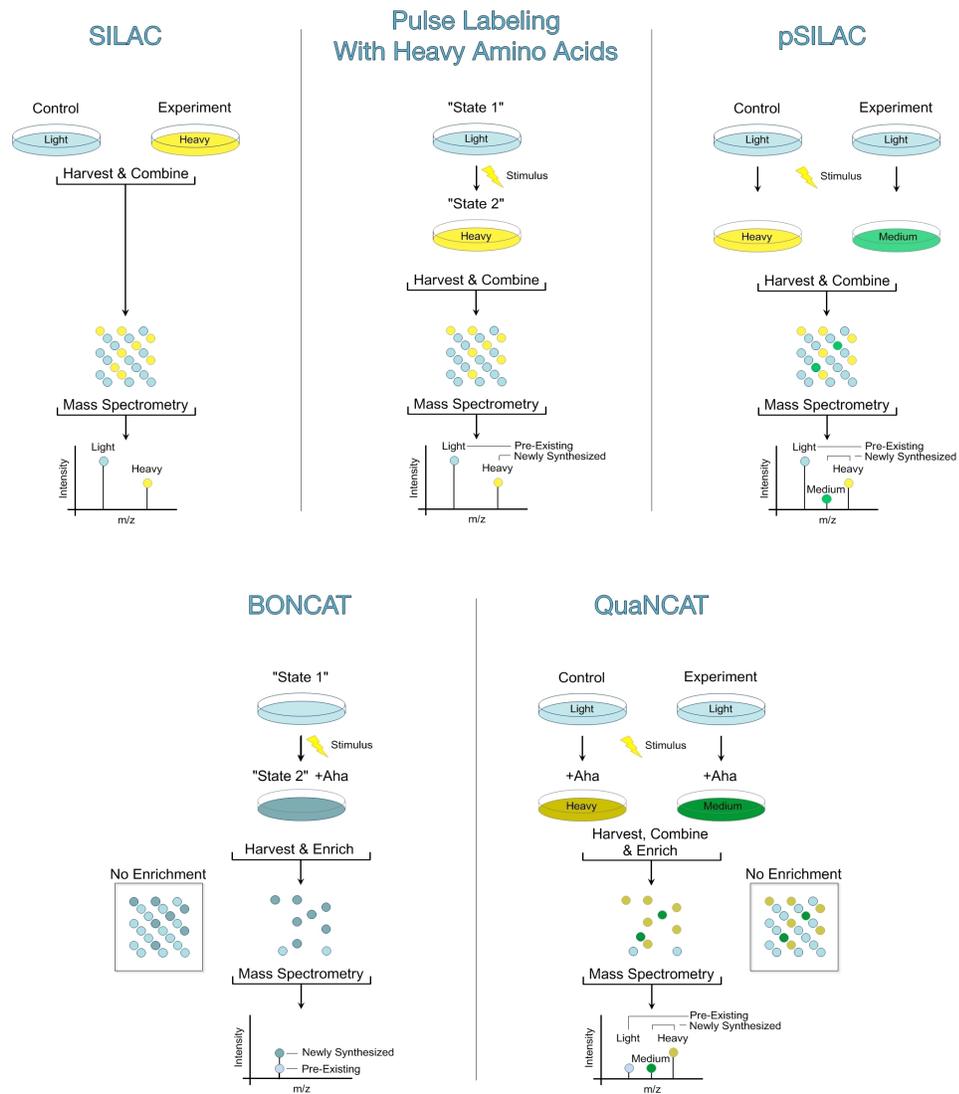


Figure 1.1: Cells can be metabolically labeled with a combination of Aha and/or stable isotopic variants of arginine and lysine in 5 workflows: standard SILAC (top left), pulse labeling with heavy amino acids (top middle), pSILAC (top right), BONCAT (bottom left), and QuaNCAT (bottom right).

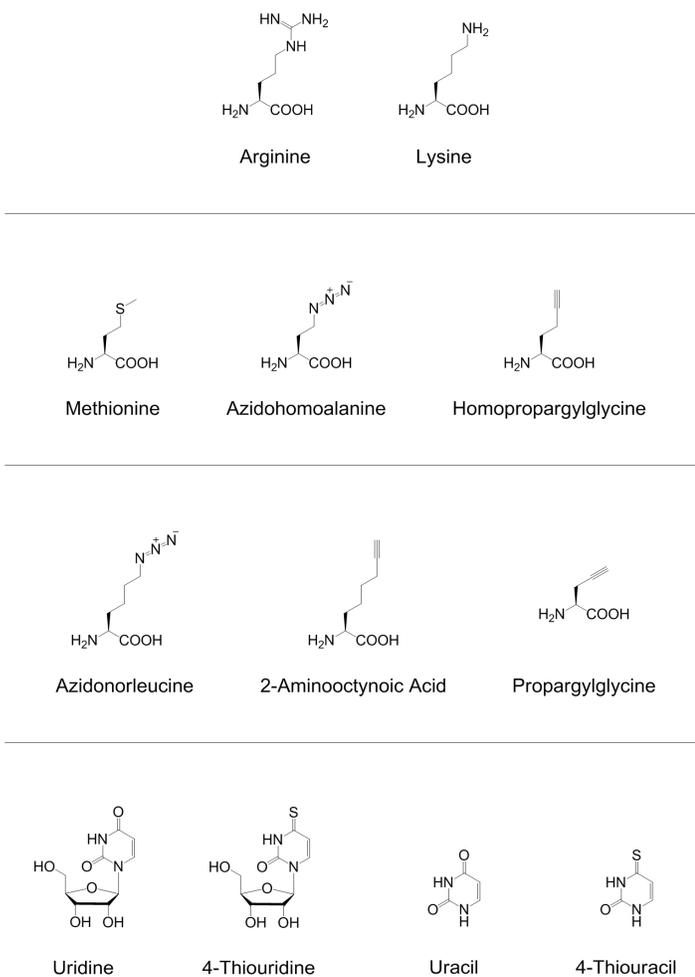
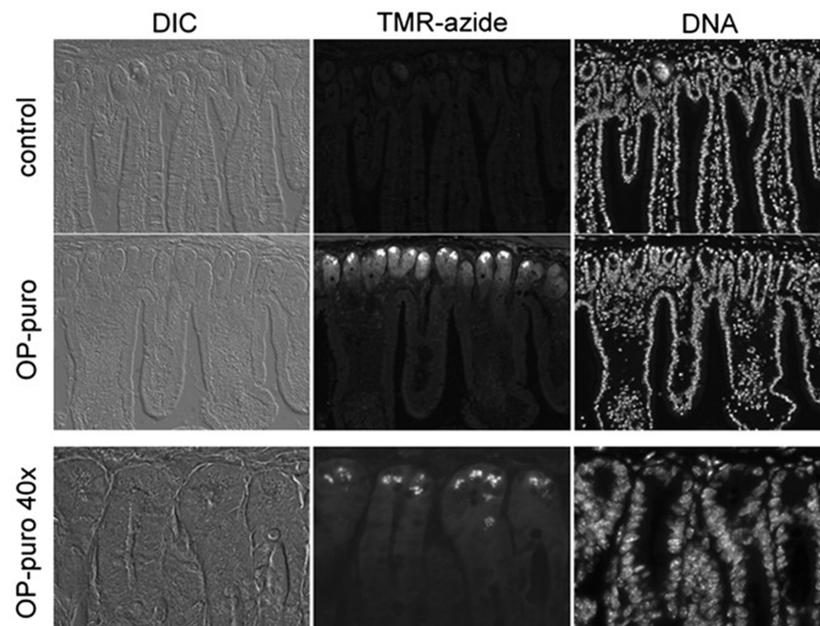
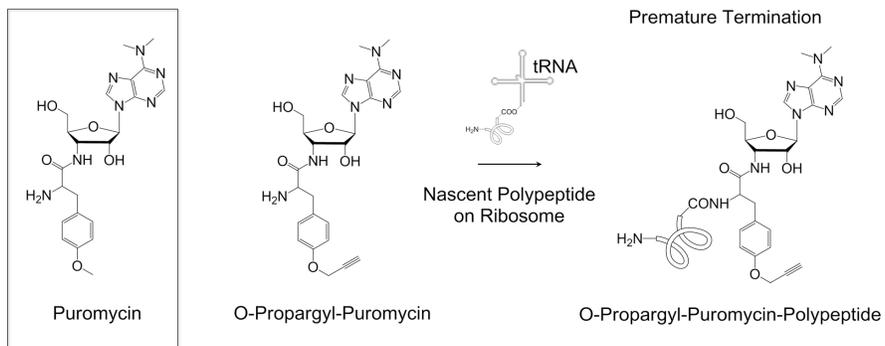


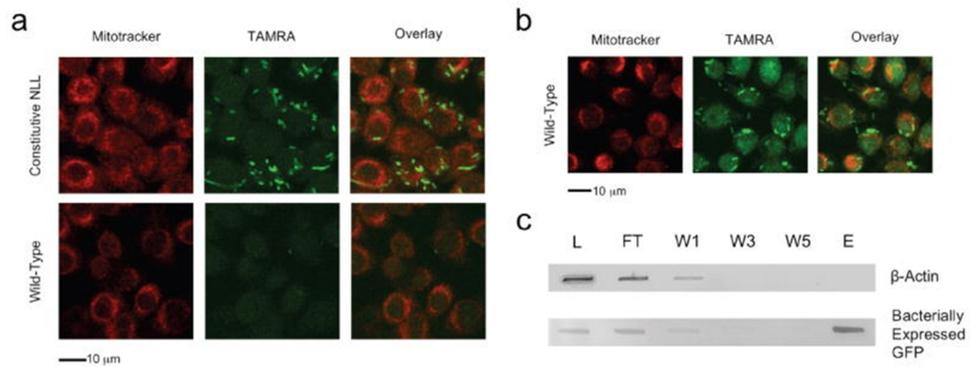
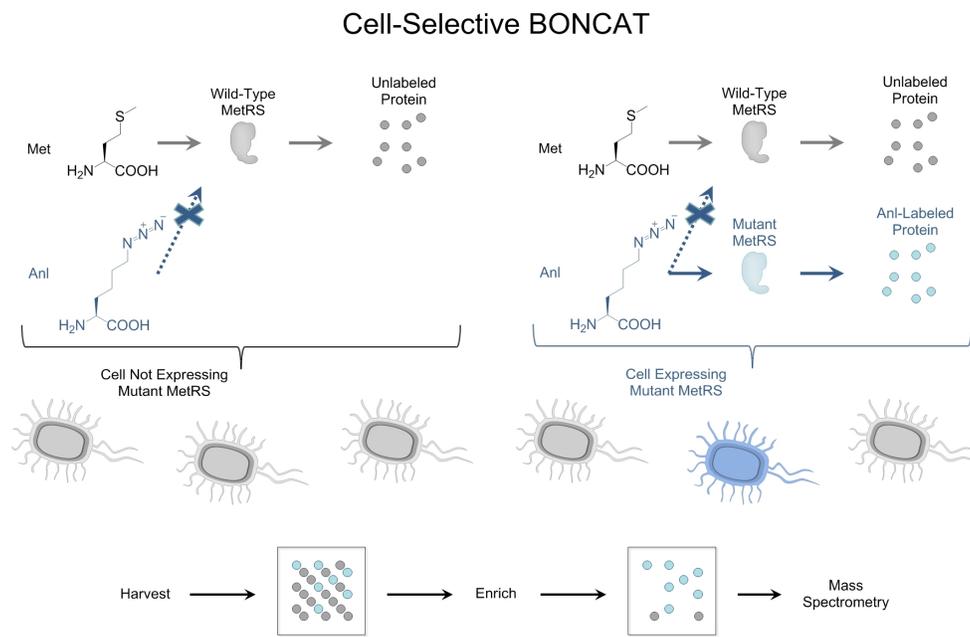
Figure 1.2: *Structures discussed in this chapter: amino acids for stable isotopic labeling (top row), methionine and analogs that are substrates for wild-type methionyl-tRNA synthetases (second row), methionine analogs that require the expression of mutant methionyl-tRNA synthetases for proteomic incorporation (third row), and uridine and uracil as well as their thio-substituted analogs 4-thiouridine and 4-thiouracil (last row).*

## O-Propargyl-Puromycin Labeling



---

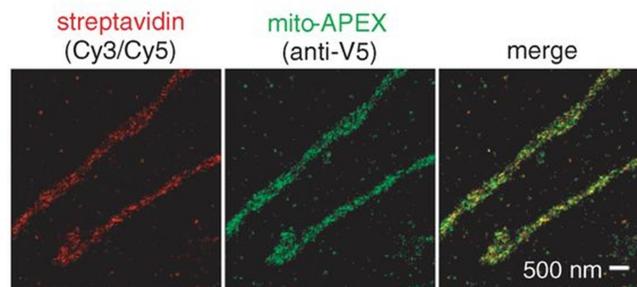
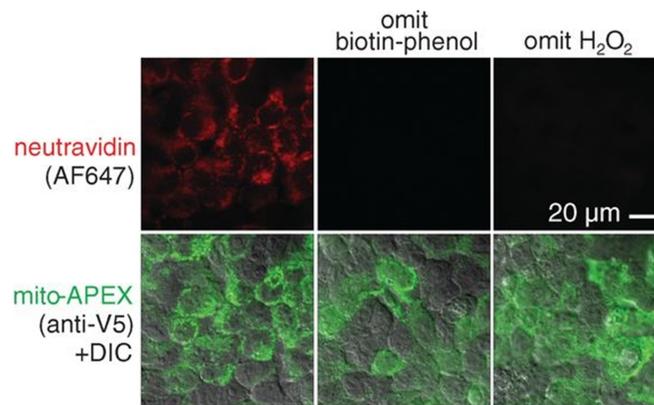
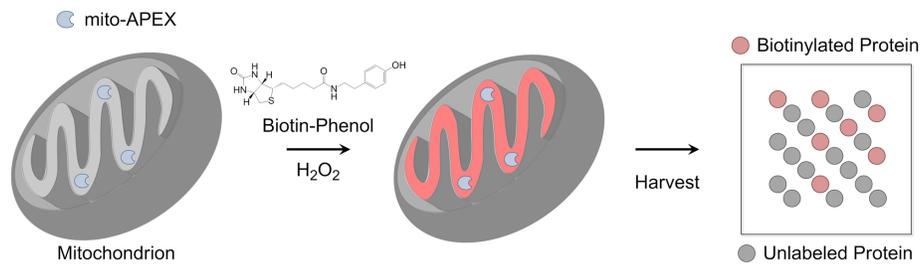
Figure 1.3 (preceding page): *(Top) Puromycin and its alkyne analog O-propargyl-puromycin incorporate into nascent polypeptide chains on translating ribosomes, resulting in premature termination of nascent polypeptide chains. (Bottom) Sectioning of mouse small intestine showed that OP-Puro labeling occurred primarily in cells in the crypts and the cells at the base of the villi. (Adapted with permission from Liu et al., Proc. Natl. Acad. Sci. U.S.A., 109, 413-418, 2012. Copyright 2012 Proceedings of the National Academy of Sciences USA.)*



---

Figure 1.4 (preceding page): (Top) Cell-selective BONCAT performed in a mixture of cells. Restricting expression of a mutant synthetase to a certain cell restricts Anl labeling to that cell (highlighted in blue). Proteins synthesized in cells (highlighted in gray) that do not express the mutant synthetase are neither labeled nor detected following enrichment. (Bottom) Cell-selective labeling in mixtures of bacterial and mammalian cells. (a) In Anl-containing mixed cultures of *E. coli* and mouse alveolar macrophages, only *E. coli* cells constitutively expressing the mutant NLL-EcMetRS were labeled by TAMRA-alkyne. Macrophages were labeled with Mitotracker Deep Red and displayed low TAMRA-alkyne background emission. (b) In Aha-containing mixed cultures of *E. coli* and mouse alveolar macrophages, both wild-type *E. coli* cells and macrophages exhibited strong TAMRA-alkyne emission; incorporation of Aha occurs in both cell types. (c) Mixed cell lysate was subjected to conjugation with alkyne-functionalized biotin, and labeled proteins were enriched by NeutrAvidin affinity chromatography. Immunoblotting of unbound flow-through (FT), washes (W1, W3, W5) and eluent (E) reveals enrichment of the bacterial marker protein GFP. (Adapted with permission from Ngo et al., *Nat. Chem. Biol.*, **5**, 715-717, 2009. Copyright 2009 Nature Publishing Group.)

## Engineered Ascorbate Peroxidase (APEX) Labeling



---

Figure 1.5 (preceding page): *(Top) Selective labeling of the mitochondrial matrix proteome in living cells requires 1) genetically targeting APEX to the mitochondrial matrix (mito-APEX), 2) initiating biotinylation by adding biotin-phenol and H<sub>2</sub>O<sub>2</sub> to the medium, and 3) stopping biotinylation by cell fixation or lysis. (Middle) In human embryonic kidney cells, only mitochondria that expressed mito-APEX and were exposed to both biotin-phenol and H<sub>2</sub>O<sub>2</sub> contained biotinylated proteins (stained with NeutrAvidin-Alexa Fluor 647). Both confocal fluorescence imaging (Middle) and stochastic optical reconstruction microscopy (STORM) (Bottom) showed that biotinylated proteins (stained with Streptavidin-Cy3/Cy5 for STORM images) overlapped with mito-APEX only in the mitochondrial matrix. (Adapted with permission from Rhee et al., Science, **339**, 1328-1331, 2013. Copyright 2013 The American Association for the Advancement of Science.)*

