

NONCANONICAL AMINO ACIDS IN  
THE INTERROGATION OF CELLULAR  
PROTEIN SYNTHESIS

Thesis by

John Tuan Ngo

In Partial Fulfillment of the Requirements for  
the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2012

(Defended December 15<sup>th</sup>, 2011)

© 2012

John Tuan Ngo

All Rights Reserved

This thesis is dedicated to David Tirrell.



## ACKNOWLEDGEMENTS

I am eternally grateful to my mother and sisters for their love and encouragement, and for the personal sacrifices they made to provide for me.

To my advisor, Prof. David Tirrell, thank you for your constant support and guidance, I would not have been able to do this without you. Thank you for encouraging my creativity, and for teaching me how to be insightful. Most importantly, thank you for always believing in me.

Colleagues from the Tirrell Lab have become some of my best friends. I have many fond memories of lunchtime brainstorming, late night experiments, and low-budget adventures around Los Angeles. Thank you Julie Champion, Frank Truong, and Rebecca Connor, Shelly Tzlil, Brett Babin, and Kai Yuet for being great labmates, teachers and friends. Thanks also to Caglar Tanrikulu, Kimberly Beatty, Tae Hyeon Yoo, Beverly Lu, Chethana Kulkarni, Maren Buck, Larry Dooling, and Nicholas Ball for their friendship and support.

I thank the Committee for their advice and guidance through my time at Caltech. I am grateful to Prof. Deshaies for his encouragement at times when graduate school seemed overwhelming. Prof. Mazmanian has been a constant source of counsel and inspiration; he is an incredibly creative scientist. I am indebted to Prof. Schuman for her enthusiasm, support, and insight that was constant while she was at Caltech, and continued even after her move to Germany.

I am grateful to Prof. Dan Morse and Dr. Meredith Murr for their continued role in my scientific education since leaving Santa Barbara.

The BMB incoming class of the 2005-06 school year was a great group thanks to Phil Romero, Russ Ernst, Camrbian Liu, Fred Tan, Kelly Matzen, Peera Jaru-Ampornpan. Thanks also to Paul Minor and Mike Kerkmen, two fellow scientists and dear friends.

I have benefited from the expertise of distinguished collaborators and am grateful to Daniela Dieterich, Jennifer Lee-Hodas, Melanie Lee, Paul Sternberg, and Meenakshi Doma.

Thanks also to Daniel M. Smith, Prunie, and Annie.

## ABSTRACT

Proteins in living cells can be made receptive to bioorthogonal chemistries through metabolic labeling with appropriately designed noncanonical amino acids (ncAAs). In the simplest approach to metabolic labeling, an amino acid analog replaces one of the natural amino acids specified by the protein's gene (or genes) of interest. This approach, often termed "residue-specific incorporation," allows the ncAA to be incorporated in controlled proportions into positions normally occupied by the natural amino acid residue. Chapter I of this thesis describes how this strategy has been used to track cellular protein synthesis with reactive ncAAs. In procedures similar to isotopic labeling, translationally active ncAAs are incorporated into proteins during a "pulse" in which newly synthesized proteins are tagged. The set of tagged proteins can be distinguished from those made before the pulse by bioorthogonally ligating the ncAA side chain to probes that permit detection, isolation, and visualization of the labeled proteins.

Chapter II of this thesis describes how the selectivity of the method can be enhanced through the use of mutant aminoacyl tRNA synthetases (aaRSs) that permit incorporation of ncAAs not used by the endogenous biomachinery. Expression of a mutant synthetase in a portion of cells within a complex cellular mixture restricts labeling to that subset of cells. In multicellular environments, this approach permits the identification of the cellular origins of labeled proteins. The work in Chapter III illustrates how the extent of temporal and spatial resolution of protein labeling can be enhanced through controlled expression of mutant synthetases. Use of characterized promoters to direct transcription of mutant synthetase genes can limit labeling to relevant cells and physiological states in settings of increased complexity. Chapter IV presents a novel strategy with which ncAAs can be uniquely incorporated at the N-terminal positions of nascent proteins while excluded from insertion at internal positions. This approach permits "site-selective" tagging of cellular proteins, and its use in tagging and

visualization of cell-cycle dependent protein synthesis is described.

The work described throughout this thesis was designed with the objective of providing powerful and versatile methods for the study of protein synthesis in complex multicellular systems, including live animals. Thus, Chapter V considers how these strategies might be used to dissect protein synthesis in living animals.

## TABLE OF CONTENTS

Acknowledgements .....	iv
Abstract.....	v
Table of Contents .....	vi
List of Tables and Supplementary Tables.....	vii
List of Figures .....	vii
List of Supplementary Figures.....	vii
Chapter I: Noncanonical Amino Acids in the Interrogation of Cellular Protein Synthesis.....	1
Conspectus .....	1
Introduction .....	2
Incorporation of Noncanonical Amino Acids into Proteins.....	4
Profiling Protein Synthesis .....	7
A Chemical and Genetic Toolkit for Profiling Protein Synthesis ..	9
Recent Applications and Future Directions .....	14
Figures .....	17
References .....	22
Chapter II: Cell-Selective Metabolic Labeling of Proteins.....	28
Abstract.....	28
Introduction .....	28
Results and Discussion .....	29
Conclusions.....	33
Figures .....	34
Supplementary Figures.....	37
Materials and Methods .....	41
References .....	49
Chapter III: Promoter-Driven Tagging: profiling cellular protein synthesis under specific physiological states .....	50
Abstract.....	50
Introduction .....	50
Results and Discussion .....	52
Conclusions.....	55
Figures .....	56
Supplementary Figures.....	58
Materials and Methods .....	61

References .....	67
Chapter IV: An Alternative Interpretation of the Start Codon Permits N-terminal Tagging of Nascent Protein Synthesis .....	
Abstract.....	68
Introduction .....	68
Results and Discussion .....	70
Conclusions.....	78
Figures .....	80
Supplementary Figures.....	84
Tables .....	102
Materials and Methods .....	104
References .....	116
Chapter V: Toward Cell-Specific Labeling of Proteins in Multicellular Organisms .....	
Introduction .....	120
Preliminary Work .....	122
Future Work.....	124
Figures .....	125
References .....	127
Concluding Remarks.....	128

## LIST OF TABLES AND SUPPLEMENTARY TABLES

<i>Table</i>	<i>Page</i>
4.1 Detected peptides displaying evidence of Anl-incorporation .....	102
<i>Supplementary Table</i>	<i>Page</i>
4.2 MetAP Processing of Tagged Proteins .....	103

## LIST OF FIGURES

<i>Figure</i>	<i>Page</i>
1.1 Structures of amino acids discussed in this chapter .....	17
1.2 Detection and identification of newly synthesized proteins ...	18
1.3 Visualization of newly synthesized proteins .....	19
1.4 Library screen for MetRS mutants .....	20
1.5 Cell-selective labeling in mixtures of bacterial and mammalian cells .....	21
2.1 Cell-selective metabolic labeling of proteomes with Anl.....	34
2.2 Cell-selective labeling in mixed populations .....	35
2.3 Cell-selective labeling in mixtures of bacterial and mammalian cells .....	36
3.1 Promoter-directed proteomic labeling with Anl .....	56
3.2 Proteomic labeling with Anl under conditions of oxidative stress .....	57
4.1 NLL-MetRS permits incorporation of Anl into proteins synthesized in the mammalian cytoplasm .....	80
4.2 Enrichment and analysis of Anl-tagged proteins.....	81
4.3 NLL-MetRS permits incorporation of Anl into proteins synthesized in the mammalian cytoplasm .....	82
4.4 Protein synthesis in hydroxyurea-treated cells.....	83
5.1 Detection of Aha-tagged <i>C. elegans</i> proteins .....	125
5.2 Visualization of Aha-tagged <i>C. elegans</i> proteins.....	126

## LIST OF SUPPLEMENTARY FIGURES

<i>Supplementary Figure</i>	<i>Page</i>
2.1 Comparison of Anl incorporation by wild-type, L13G, and NLL-MetRS .....	36
2.2 Comparison of Anl-incorporation by cells over-expressing plasmid-borne copies of either WT-MetRS, L13G-MetRS, or NLL-MetRS .....	37
2.3 Detection of Anl-tagged proteins .....	38
2.4 Individual panels from imaging of mixed bacterial cultures from Figure 2.2b.....	39
2.5 Infection of macrophages is performed with control bacteria expressing GFP, but not the NLL-MetRS (Wild-Type/GFP) .....	40
3.1 Proteomic labeling with Anl under conditions of oxidative stress .....	58
3.2 Tagging rate in the SoxRS system is less than 10% .....	59
3.3 Tagging rate in the SoxRS system is dependent on the degree of induction of the NLL-MetRS expression.....	60
4.1 Amino acids and probes used in this study .....	84
4.2 Selective aminoacylation of eukaryotic cytoplasmic initiator tRNA <i>in vitro</i> .....	85
4.3 Heterologous expression of the NLL-MetRS in mammalian cells .....	86
4.4 Anl is incorporated into protein synthesized in HEK293-8D3 cells.....	87
4.5 The specificity of DIBAC-conjugation to Anl-tagged proteins is increased following cysteine alkylation with iodoacetamide .....	88
4.6 Anl-tagging under varying pulse concentrations and durations.....	89
4.7 Tagging and enrichment of constitutively expressed proteins .....	90
4.8 Estimation of the degree of enrichment .....	91
4.9 Structures used to determine the mass modifications applied in the database search for Anl-containing peptides...	92

4.10	Sequence coverage of PTBP .....	93
4.11	Mass spectra of identified Anl-containing peptides .....	94
4.12	Subcellular fractionation of pulsed HEK293-8D3 cells.....	99
4.13	Differences in translational profiles can be accentuated.....	100
4.14	MetAP exhibits attenuated excision activity toward N-terminal Anl .....	101

# CHAPTER I

## *Noncanonical Amino Acids in the Interrogation of Cellular Protein Synthesis*

### CONSPECTUS

Proteins in living cells can be made receptive to bioorthogonal chemistries through metabolic labeling with appropriately designed noncanonical amino acids (ncAAs). In the simplest approach to metabolic labeling, an amino acid analog replaces one of the natural amino acids specified by the protein's gene (or genes) of interest. Through manipulation of experimental conditions, the extent of the replacement can be adjusted. This approach, often termed residue-specific incorporation, allows the ncAA to be incorporated in controlled proportions into positions normally occupied by the natural amino acid residue. For a protein to be labeled in this way with an ncAA, it must fulfill just two requirements: (i) the corresponding natural amino acid must be encoded within the sequence of the protein at the genetic level, and (ii) the protein must be expressed while the ncAA is in the cell. Because this approach permits labeling of proteins throughout the cell, it has enabled us to develop strategies to track cellular protein synthesis by tagging proteins with reactive ncAAs. In procedures similar to isotopic labeling, translationally active ncAAs are incorporated into proteins during a "pulse" in which newly synthesized proteins are tagged. The set of tagged proteins can be distinguished from those made before the pulse by bioorthogonally ligating the ncAA side chain to probes that permit detection, isolation, and visualization of the labeled proteins. Noncanonical amino acids with side chains containing azide, alkyne, or alkene groups have been especially useful in experiments of this kind. They have been incorporated into proteins in the form of methionine analogs that are substrates for the natural translational

machinery. The selectivity of the method can be enhanced through the use of mutant aminoacyl tRNA synthetases (aaRSs) that permit incorporation of ncAAs not used by the endogenous biomachinery. Through expression of mutant aaRSs, proteins can be tagged with other useful ncAAs, including analogs that contain ketones or aryl halides. High-throughput screening strategies can identify aaRS variants that activate a wide range of ncAAs. Controlled expression of mutant synthetases has been combined with ncAA tagging to permit cell-selective metabolic labeling of proteins. Expression of a mutant synthetase in a portion of cells within a complex cellular mixture restricts labeling to that subset of cells. Proteins synthesized in cells not expressing the synthetase are neither labeled nor detected. In multicellular environments, this approach permits the identification of the cellular origins of labeled proteins. In this chapter, we summarize the tools and strategies that have been developed for interrogating cellular protein synthesis through residue-specific tagging with ncAAs. We describe the chemical and genetic components of ncAA-tagging strategies and discuss how these methods are being used in chemical biology.

## INTRODUCTION

Shortly after the discovery of deuterium by Urey and co-workers, Rudolph Schoenheimer suggested that isotopically tagged cellular constituents could be used to trace and identify the products of metabolic reactions.(1) In 1938, Schoenheimer reported the first metabolic labeling of proteins with an isotopically tagged amino acid: in rats fed  $^{15}\text{N}$ -labeled tyrosine, it was found that a fraction of the amino acid was retained within the animal in the form of protein.(2) In the following decades, delineation of the mechanism of protein synthesis would stimulate the prospect that other non-naturally occurring amino acids might be incorporated into proteins. By 1956, this idea was validated by the demonstration that selenomethionine (Se-Met, Figure 1.1) could be used by bacterial cells to make proteins.(3) Today, hundreds of translationally active noncanonical amino acids (ncAAs) have been

identified, and recent progress in the incorporation of reactive ncAAs, combined with developments in bioorthogonal chemistry, have led to new ways to trace the lives of proteins.

Proteins in living cells can be made receptive to bioorthogonal chemistries through metabolic labeling with appropriately designed noncanonical amino acids (ncAAs). In the simplest approach to metabolic labeling, an amino acid analog replaces one of the natural amino acids specified by the protein's gene (or genes) of interest. Through manipulation of experimental conditions, the extent of the replacement can be adjusted. This approach, often termed residue-specific incorporation, allows the ncAA to be incorporated in controlled proportions into positions normally occupied by the natural amino acid residue. For a protein to be labeled in this way with an ncAA, it must fulfill just two requirements: (i) the corresponding natural amino acid must be encoded within the sequence of the protein at the genetic level, and (ii) the protein must be expressed while the ncAA is in the cell.

Because this approach permits labeling of proteins throughout the cell, it has enabled us to develop strategies to track cellular protein synthesis by tagging proteins with reactive ncAAs. In procedures similar to isotopic labeling, translationally active ncAAs are incorporated into proteins during a "pulse" in which newly synthesized proteins are tagged. The set of tagged proteins can be distinguished from those made before the pulse by bioorthogonally ligating the ncAA side chain to probes that permit detection, isolation, and visualization of the labeled proteins.

Noncanonical amino acids with side chains containing azide, alkyne, or alkene groups have been especially useful in experiments of this kind. They have been incorporated into proteins in the form of methionine analogs that are substrates for the natural translational machinery. The selectivity of the method can be enhanced through the use of mutant aminoacyl tRNA synthetases (aaRSs) that permit incorporation of ncAAs not used by the

endogenous biomachinery. Through expression of mutant aaRSs, proteins can be tagged with other useful ncAAs, including analogs that contain ketones or aryl halides. High-throughput screening strategies can identify aaRS variants that activate a wide range of ncAAs.

Controlled expression of mutant synthetases has been combined with ncAA tagging to permit cell-selective metabolic labeling of proteins. Expression of a mutant synthetase in a portion of cells within a complex cellular mixture restricts labeling to that subset of cells. Proteins synthesized in cells not expressing the synthetase are neither labeled nor detected. In multicellular environments, this approach permits the identification of the cellular origins of labeled proteins.

In this chapter, we summarize the tools and strategies that have been developed for interrogating cellular protein synthesis through residue-specific tagging with ncAAs. We describe the chemical and genetic components of ncAA-tagging strategies and discuss how these methods are being used in chemical biology.

## **INCORPORATION OF NONCANONICAL AMINO ACIDS INTO PROTEINS**

Codons are assigned to amino acids through selective aminoacylation of tRNAs followed by accurate base-pairing between charged tRNAs and mRNAs. Amino acids are assigned to individual tRNAs by the aminoacyl-tRNA synthetases (aaRSs). Manipulation of the aminoacylation step to direct the addition of ncAAs to tRNA has facilitated the incorporation of ncAAs into proteins in both site-specific and residue-specific fashion.

### **Site-Specific Incorporation**

Site-specific incorporation methods allow the investigator to insert a single ncAA at a predetermined position in a recombinant protein. In the most

common approach, a TAG stop codon is introduced into the gene of interest. Translation of the full-length protein is enabled by introduction of a suppressor tRNA charged with the ncAA. Introduction of the aminoacyl-tRNA is accomplished either by injection of a chemically misacylated tRNA(4) or by expression of an orthogonal aaRS/tRNA pair.(5)

In site-specific insertion, the ncAA is incorporated into a selected protein at a predetermined site. Since measures can be taken to ensure that incorporation at the selected site does not interfere with protein structure, site-specific approaches are ideal for introducing ncAAs into proteins with minimal perturbation. However, it should be noted that nontarget proteins are labeled by incidental “read-through” on other transcripts that include the blank codon.

Site-specific incorporation of ncAAs is a powerful technology with important applications in protein engineering, biochemistry, and biophysics.(6) Site-specific incorporation of reactive side chains has been used to create useful protein conjugates,(7) and the incorporation of biophysical probes has revealed atomistic details of protein structure and function in bacteria, yeast,(8) and mammalian cells.(9) For a comprehensive review of site-specific technology, see Liu and Schultz.(6)

### **Residue-Specific Incorporation**

Under normal physiological conditions, the aaRSs catalyze high-fidelity aminoacylation of cognate tRNAs with a collective error rate of  $\sim 10^{-4}$ .(10) However, the specificity of the synthetases can be overcome by reducing the concentrations of their natural amino acid substrates and supplementing the cellular medium with substrate analogs. This approach has been used to achieve residue-specific incorporation of a wide variety of useful ncAAs.

In residue-specific incorporation, a ncAA serves as a surrogate for a natural amino acid and is inserted into cellular proteins at multiple sites. The rate at

which the ncAA replaces the natural residue can be varied to suit the application of interest. Under certain conditions, quantitative substitution can be achieved; if partial replacement is preferred (e.g., for protein tagging with minimal perturbation of structure and function), the extent of incorporation can be set by controlling the concentrations of the canonical and noncanonical amino acids in the medium.(11)

For a protein to be labeled in residue-specific fashion, it must fulfill just two requirements: (1) the corresponding natural amino must be encoded within the sequence of the protein at the genetic level, and (2) the protein must be expressed while the ncAA is in the cell. Depending on the properties of the ncAA, the entire proteome may be subject to tagging. Residue-specific methods are therefore ideally suited for interrogating cellular protein synthesis.

Throughout the last decade, ncAAs with reactive side chains have been of particular interest to our laboratory. In combination with rapid progress in the development of bio-orthogonal reactions,(12) the chemistry of ncAAs is providing powerful new tools for tagging and modifying proteins. Although reactive ncAAs have many potential uses in biology and medicine, this chapter will focus on their application in proteomic analysis of cells, tissues, and organisms.

### **Azides and Alkynes in Metabolic Profiling**

The discovery and development of new bioorthogonal reactions has transformed chemical biology. The use of functionalized analogs of biomolecular building blocks, or “chemical reporters”, has permitted the tagging and tracking of cellular components including glycans,(13) lipids,(14) proteins,(15) RNA,(16) and DNA.(17) Azides and alkynes have assumed leading roles in many chemical reporter strategies; these functional groups are small, stable under biologically relevant conditions, and virtually absent

from cells and tissues. Furthermore, many azide and alkyne analogs of cellular building blocks can be incorporated into target biomolecules by the normal biosynthetic machinery of the cell. These analogs mimic the fates of their natural counterparts and can be used to track cellular metabolism. For detection, both azides and alkynes can be revealed by Cu(I)-catalyzed azide-alkyne cycloaddition(18, 19) (CuAAC) to appropriately functionalized probes (haptens, fluorescent dyes, or affinity reagents). The azide is especially versatile; in addition to CuAAC, azides are also susceptible to the Staudinger ligation(20) and to strain-promoted ligation to cyclooctynes and their derivatives.(21-23)

We have developed a set of azide- and alkyne-functionalized ncAAs to tag proteins and track cellular protein synthesis. In 2000, we described incorporation of the methionine analog homopropargylglycine (Hpg) into proteins expressed in *Escherichia coli*.(24) Although kinetic assays with purified *E. coli* methionyl-tRNA synthetase (MetRS) showed Hpg to be activated roughly 2 orders of magnitude more slowly than methionine, quantitative replacement of methionine by Hpg was observed when the marker protein DHFR was expressed in a Met-auxotrophic host strain. Shortly thereafter, in collaboration with the Bertozzi laboratory, we reported incorporation of azidohomoalanine (Aha), a second methionine surrogate.(25) Like Hpg, Aha is activated more slowly than methionine, but can replace methionine essentially quantitatively in recombinant proteins. In our first report on Aha, we demonstrated that azide-labeled proteins could be selectively modified by the Staudinger ligation.

## PROFILING PROTEIN SYNTEHSIS

The cellular proteome dictates cell identity, state, and function; the proteins expressed in the cell drive processes such as cell division and differentiation, and changes in protein expression allow cells and organisms to adapt to signals and stresses. In order to understand how cells execute these and other

functions, the proteins involved in each process, including those that are newly synthesized, modified, or degraded, must be identified. Because the complexity of the proteome frustrates such studies, methods that examine subsets of the proteome (e.g., “phosphoproteomics”,<sup>(26)</sup> “glycoproteomics”,<sup>(13)</sup> and “activity-based protein profiling”<sup>(27)</sup>) have attracted substantial attention.

### **Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT)**

In 2006, Dieterich et al. introduced the BONCAT (bioorthogonal noncanonical amino acid tagging) method, a strategy for selective enrichment and identification of newly synthesized cellular proteins.<sup>(15)</sup> The BONCAT method as originally introduced involved pulse-labeling of cells with the methionine surrogate Aha. Cellular proteins made during the pulse were labeled with Aha and therefore susceptible to selective conjugation to alkyne-functionalized affinity probes via CuAAC (Figure 1.2). Newly synthesized, affinity-tagged proteins can be enriched from the preexisting, unlabeled protein pool by affinity chromatography and identified by mass spectrometry. The BONCAT approach decreases sample complexity and permits direct analysis of the pool of proteins synthesized in response to biological cues.

As the BONCAT method was introduced, complementary techniques for visualization of newly synthesized proteins were also developed. One can visualize newly synthesized proteins in cells by appending fluorescent dyes, rather than affinity reagents, to metabolically labeled proteins. Cells labeled with Aha or Hpg can be selectively conjugated to fluorescent or fluorogenic dyes; imaging of intact cells reveals the spatial distribution of proteins synthesized during the ncAA pulse. We have demonstrated such methods in bacterial cells,<sup>(28)</sup> in mammalian cells,<sup>(29)</sup> and in whole organisms. Figure 1.3 compares mammalian cells pulsed with Aha under conditions that allow (top) or inhibit (bottom) the synthesis of new proteins. Robust labeling of

new proteins is apparent in the absence of protein synthesis inhibitors, indicating that active protein synthesis is underway. The absence of signal in cells in which protein synthesis is inhibited confirms the selectivity of the method with respect to labeling of newly synthesized proteins. Multicolor tracking of two temporally distinct populations of proteins can be achieved by sequential tagging with Aha and Hpg,(30) and live-cell imaging of Aha-tagged proteins has been achieved by ligation to strained-alkyne dyes.(31)

## **A CHEMICAL AND GENETIC TOOLKIT FOR PROFILING PROTEIN SYNTHESIS**

Increasing the sensitivity and selectivity of chemical reporter strategies constitutes a persistent challenge. Increased reaction yields with greater reproducibility,(32) as well as the design and synthesis of improved probes for visualization and enrichment,(33) will extend the limits of what can be detected. In this section, we describe recent work directed toward the development of new chemical and genetic tools for the interrogation of cellular protein synthesis with reactive ncAAs.

### **An Alkene Reporter of Protein Translation**

In 1998, van Hest et al. reported the incorporation of homoallylglycine (Hag), an alkene-functional methionine surrogate, into recombinant proteins.(34) Although it is charged to tRNA<sup>Met</sup> more slowly than Aha or Hpg, Hag can replace methionine quantitatively in proteins expressed in Met-auxotrophic strains of *E. coli*. At the time this work was reported, it was proposed that the Hag side chain might serve as a handle for bioorthogonal protein modification. Recent developments have turned this proposal into reality. In 2009, Davis and co-workers described specific protein conjugation through radical addition of a glycosyl thiol to the Hag side chain.(35) Successful conjugation was demonstrated on a purified Hag-labeled protein, as well as on a virus-like scaffold displaying 180 surface Hag residues; in both cases,

complete modification was observed. In a 2010 report, Song et al. described the use of Hag to label newly synthesized proteins in mammalian cells.(36) Hag-labeled proteins were biotinylated or conjugated to fluorescent dyes via photoinduced tetrazole–alkene cycloaddition. In earlier work, the Lin group demonstrated the bioorthogonality of the tetrazole–alkene cycloaddition by modifying a protein site-specifically labeled with *O*-allyl-tyrosine.(37)

### Engineering the Amino Acyl-tRNA Synthetase

The natural biosynthetic machinery of microbial and mammalian cells permits the incorporation of a surprising number of ncAAs into cellular proteins. Nevertheless, there are substantial advantages to be gained by engineering the aaRS to expand the set of translationally active ncAAs.

In 2000, *in vivo* analysis in *E. coli* suggested that *trans*-crotylglycine (Tcg) could not support protein synthesis in a Met-auxotrophic strain depleted of methionine. However, *in vitro* kinetic measurements revealed that Tcg was activated by the *E. coli* MetRS.(38) Kiick and co-workers proposed that the low rate of activation of Tcg limited the overall rate of protein synthesis and that this limitation might be overcome through overexpression of the synthetase. In cells that expressed a plasmid-encoded wild-type MetRS (in addition to the genome-encoded copy), Tcg was successfully incorporated into protein.

Relaxation of synthetase specificity, either by attenuation of editing activity or by expansion of the amino acid binding site, can also enable incorporation of ncAAs. In nine of the 20 aaRSs, editing activity prevents misacylation of tRNA with noncognate natural amino acids.(9) Mutation in the *E. coli* leucyl-tRNA synthetase (LeuRS) editing domain (T252Y) attenuates the editing activity of the enzyme and permits incorporation of a variety of leucine analogs,(39) including the ketone analog oxonorvaline (Onv).(40) The A294G mutation in the  $\alpha$ -subunit of the *E. coli* phenylalanyl-tRNA synthetase

(PheRS) relaxes substrate specificity by increasing the volume of the amino acid binding pocket.(41) Overexpression of the mutant synthetase enables incorporation of a variety of reactive phenylalanine analogs including *p*-bromophenylalanine (*p*-BrF), *p*-iodophenylalanine (*p*-IF), *p*-ethynylphenylalanine (*p*-EtF), and *p*-azidophenylalanine (*p*-AzF).(42) However, overexpression of this mutant did not allow incorporation of *p*-acetylphenylalanine (*p*-AcF). Because of the versatility of the ketone function in chemoselective ligations, we thought it worthwhile to seek a PheRS mutant that would enable its use in protein labeling. In a computational analysis, the crystal structure of the *Thermus thermophilus* PheRS complexed with phenylalanine was used to construct a backbone-independent rotamer library. Calculations accurately predicted two cavity-forming mutations that would enable *p*-AcF activation.(43) A bacterial strain harboring the PheRS double mutant was prepared and shown to use *p*-AcF as a phenylalanine surrogate. *p*-AcF-labeled proteins were readily modified through hydrazide addition to the side-chain ketone.

The expression of mutant synthetases with altered substrate specificities expands the options available for protein labeling and proteomic analysis. For instance, in cells expressing the A294G mutant *E. coli* PheRS, *p*-AzF and *p*-EtF may be used as substitutes for or complements to Aha and Hpg in protein profiling. Tagging with analogs containing aryl halides, including *p*-BrF and *p*-IF, is also possible.(42) When first reported, these amino acids were proposed as handles for Pd-catalyzed cross-couplings; however, early efforts to adapt cross-coupling procedures to proteins in aqueous solvents were compromised by interference from cysteine thiols, required solvent degassing, and generally resulted in poor yields. In a promising recent development, Davis's adaptation of the Suzuki–Miyaura coupling employing a pyrimidine–Pd catalyst afforded specific and high-yield coupling of aryl bromide and aryl iodide groups on proteins.(44) Since Pd catalysts are poisoned by thiols, it is important that reduction and alkylation of cysteines

(a standard step in sample preparation for proteomic analysis by mass spectrometry) be performed prior to addition of the Pd catalyst.

As described above, the toolkit for ncAA tagging has been expanded by the inclusion of two analogs containing ketones.(40) Chemoselective ligations to *p*-AcF or Onv residues in proteins can be achieved by treatment with hydrazide- or aminoxy-functionalized probes. While ketones and aldehydes are found on some cellular metabolites and cofactors, they are not normally found in proteins; complications in ligation reactions can be avoided by first separating protein from other cellular constituents. However, Bertozzi has suggested that the intracellular presence of ketone-containing metabolites may restrict dye-labeling and imaging applications of ketone- and aldehyde-labeled biomolecules to the cell surface where these functions are usually absent.(45)

### **Combinatorial Approaches to Mutant aaRS**

Efficient screening and selection methods have been developed to enable rapid identification of aaRS mutants of altered substrate specificity.(6, 46, 47) Much of our recent effort in this area has been directed toward isolation of MetRS mutants that activate azidonorleucine (Anl), a long-chain variant of Aha. Using structure-guided mutagenesis, we constructed a library of MetRS mutants by randomization at three positions (L13, Y260, and H301) adjacent to the amino acid binding site. Clones expressing functional mutants were selected on the basis on their ability to incorporate Anl into the outer membrane protein OmpC (Figure 1.4a).(48, 49) Azides displayed on the cell surface were detected by ligation to a biotin–cyclooctyne reagent and treatment with a fluorescent avidin conjugate. Bright clones were isolated by fluorescence-activated cell sorting (FACS). The FACS screen uses the extent of fluorescence labeling as a reporter for Anl incorporation and, in turn, for MetRS activity toward Anl; the highest levels of OmpC expression and fluorescence labeling are expected in cells carrying MetRS mutants most

active toward Anl. Through this approach, more than 40 different MetRS mutants were identified, each allowing incorporation of Anl into proteins with varying efficiency.(47, 50) A combination of mutational and X-ray crystal structure analyses showed that the successful mutations eliminated hydrophilic contacts with the Met side chain, thereby reducing competition with Met and polar amino acids while creating a hydrophobic pocket for Anl (Figure 1.4b,c).(51) The L13N/Y260L/H301L mutant (NLL-MetRS) proved to be especially interesting, in that its specificity constant ( $k_{\text{cat}}/K_M$ ) for activation of Anl was found to be slightly larger than that for methionine.

### Cell-Selective Protein Labeling

Access to the NLL-MetRS mutant allowed us to develop a method for cell-selective proteomic analysis.(52) Cells that express the NLL-MetRS can utilize Anl in protein synthesis; cells that do not express the mutant synthetase are inert to Anl. In a cellular mixture, only proteins from mutant cells are labeled. Through this approach, proteins synthesized in targeted cells can be selectively isolated from complex mixtures for identification by mass spectrometry or conjugated to fluorescent dyes for in situ visualization. In a first demonstration of the method, we confined protein labeling with Anl to bacterial cells cocultured with mammalian macrophages (Figure 1.5). Careful analysis of bacterial proteins in such mixtures requires enrichment because of the disparity in the sizes and protein contents of bacterial and mammalian cells; bacterial proteins typically account for less than 1% of the total protein content in an infection mixture.(53) Anl-tagged proteins were selectively ligated to biotin and separated from mammalian proteins in an infection lysate (Figure 1.3). The Hang laboratory has used the NLL-MetRS to activate the alkyne analog, 2-aminooctynoic acid (Aoa), for cell-selective labeling.(54) The authors used Aoa to label bacterial proteins expressed in macrophage-internalized *Salmonella typhimurium*. While the NLL-MetRS was isolated from a screen for mutants active toward Anl, the Hang experiments show that it also activates Aoa. NLL-MetRS mediated labeling with Aoa required removal

of methionine from the culture medium, whereas Anl-labeling did not. Hang and co-workers found lower background in reactions between azide-functionalized probes and Aoa-tagged proteins than in parallel experiments with alkyne detection reagents and Anl-tagged proteins.

### **Some Considerations**

It is important to keep in mind that profiling technologies are intended to aid and accelerate *discovery*. Proteins identified in profiling experiments of the kind discussed here may be involved in the biological process of interest but may be bystanders or products of secondary effects. The possibility that artifacts can arise in labeling experiments is well-known, and should be a source of caution in interpreting results. In the methods described here, ncAAs are incorporated into many proteins, with unknown effects on cellular behavior. Artifacts are likely to be minimized by short labeling times and modest labeling rates, and since the methods used to track labeled molecules are sensitive (fluorescence microscopy and mass spectrometry), labeling rates of a few percent are sufficient. The labeling rate is easily controlled by adjustment of the relative concentrations of the amino acid to be replaced and the ncAA reporter.

### **RECENT APPLICATIONS AND FUTURE DIRECTIONS**

ncAA-tagging strategies have been employed in recent years to interrogate several complex cellular processes. Examples include kinetic analysis of protein modification and turnover rates and imaging of newly synthesized proteins in neurons.

Deal and co-workers have reported an ncAA-tagging method (designated CATCH-IT, for covalent attachment of tags to capture histones and identify turnover) to examine the dynamics of genome-wide nucleosome disassembly and reassembly, which are important in epigenome maintenance.<sup>(55)</sup> In the CATCH-IT approach, newly translated histones are tagged with Aha and

enter the nucleus where they are incorporated into chromatin. After tagging, total chromatin is isolated, and tagged nucleosomes are ligated to biotin probes. Newly synthesized nucleosomes are isolated by binding to streptavidin beads, and the associated DNA is identified on microarrays to determine the genome positions at which old histones have exchanged with new.

Zhang and colleagues have combined tandem labeling (with Aha and an alkyne analog of palmitic acid) and fluorescence imaging to examine S-palmitoylation dynamics in Jurkat cells.(56) The rate of turnover of H-RasG12V (determined from the Aha signal intensity in a pulse–chase experiment) was correlated with fatty acid cycling (from the palmitate analog). The results were consistent with dynamic S-acylation and minimal protein turnover within the 5 h time interval examined.

Local protein synthesis in neuronal dendrites is thought to contribute to learning and memory, and this provided the initial motivation for development of the BONCAT method. This process has been examined in a recent report by Dieterich and colleagues in which Aha and Hpg were used to visualize protein synthesis in dissociated rat hippocampal neurons and in organotypic slices.(57) Microperfusion of the ncAA permitted direct detection of local protein synthesis in neuronal dendrites. The diffusion dynamics of newly synthesized proteins were examined by conjugation of ncAA-tagged proteins to quantum dots.

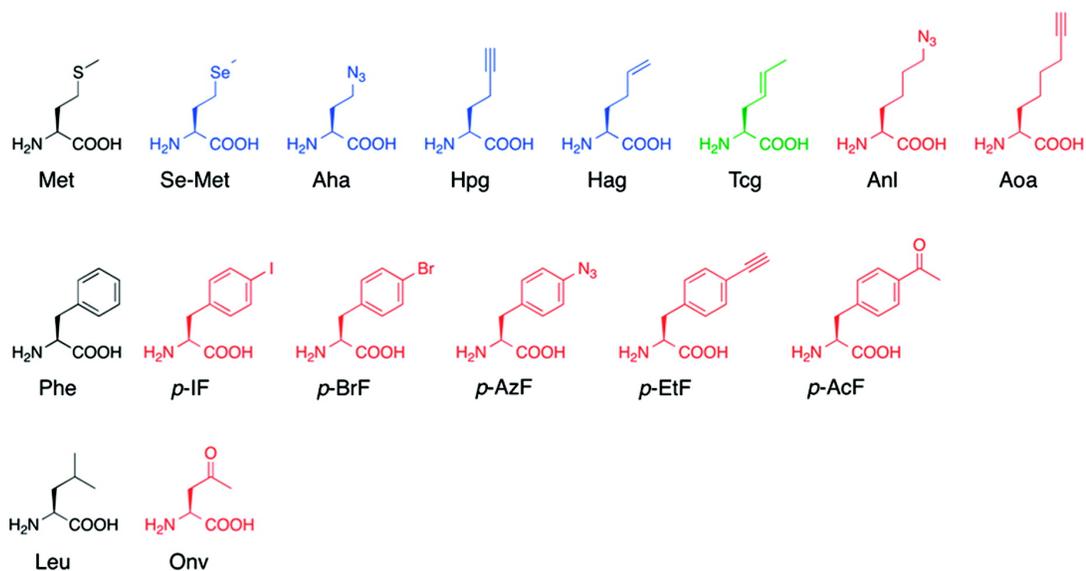
In recent work by Flanagan and co-workers, Aha was used to visualize local protein synthesis in specific subregions of the neuronal cytoplasm in response to external cues.(58) DCC (deleted in colorectal cancer) is a transmembrane receptor that regulates protein translation in axons and dendrites in response to the extracellular ligand netrin. Because DCC colocalizes with components of the translational machinery in neuronal dendrites and axons, it was thought that it might regulate local protein

synthesis in the neuron. To probe the relation between netrin binding and local protein synthesis, Flanagan and co-workers used Aha to tag proteins made in netrin-stimulated neurons. Aha-tagged proteins were found throughout the neuron but exhibited discrete colocalization with DCC along axons, within dendrites, and at the tips of filapodia. Aha tagging was used to measure variations in protein synthesis activity within DCC microdomains by quantifying fluorescence from dye-labeled neurons. Netrin-stimulated cells showed marked increases in protein synthesis in DCC microdomains, further supporting a role for netrin binding in the positive regulation of local translation.

### **Future Directions**

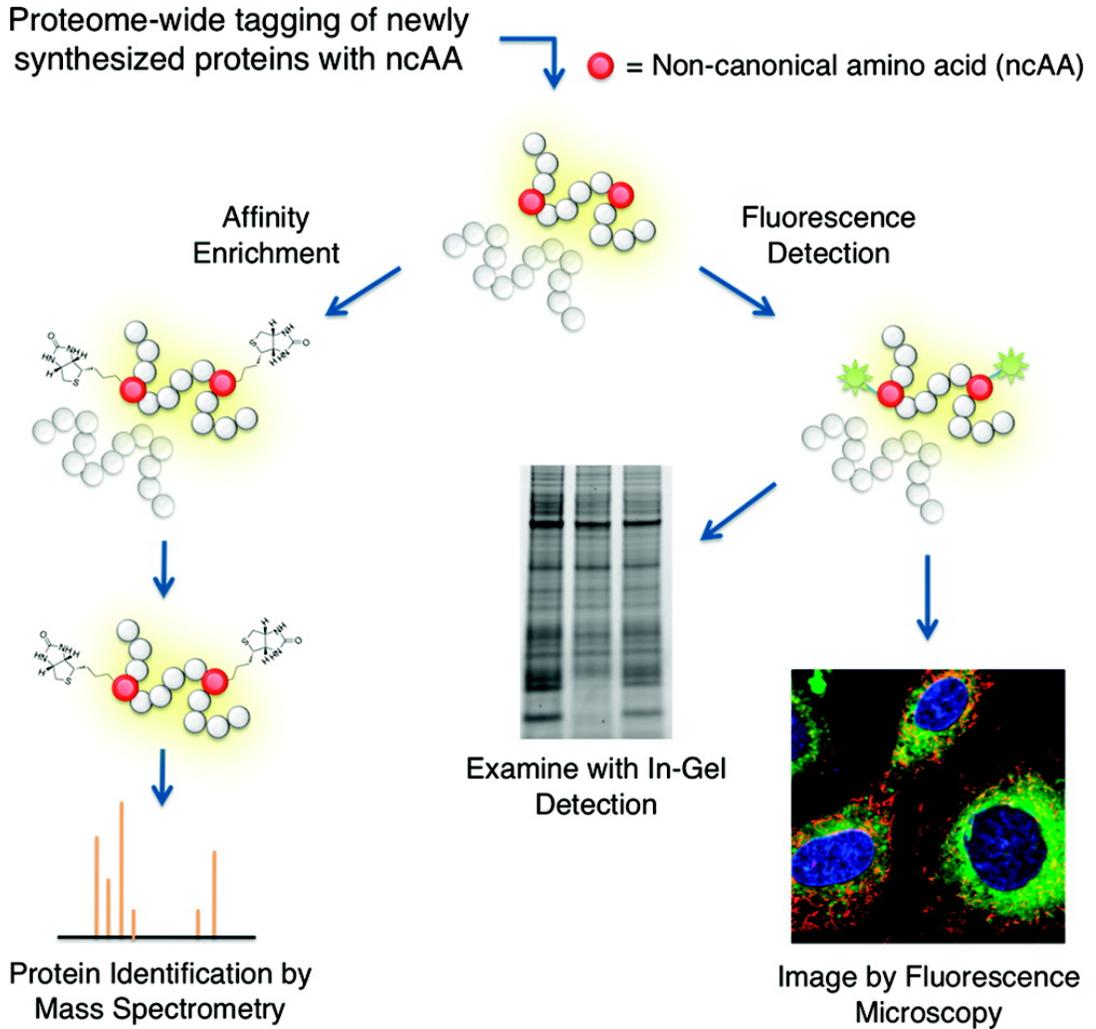
While there are many promising applications of ncAAs in protein engineering and molecular medicine,(7, 59, 60) this Account has focused on the incorporation of reactive ncAAs and their use as analytical tools in chemical biology. We continue to develop ncAA-tagging strategies by creating new ncAAs, engineering new aaRS mutants, and preparing new probes. In collaboration with colleagues, we are pursuing studies of host-pathogen interactions and cell signaling, as well as methods for interrogating protein synthesis in live animals.

Figure 1.1



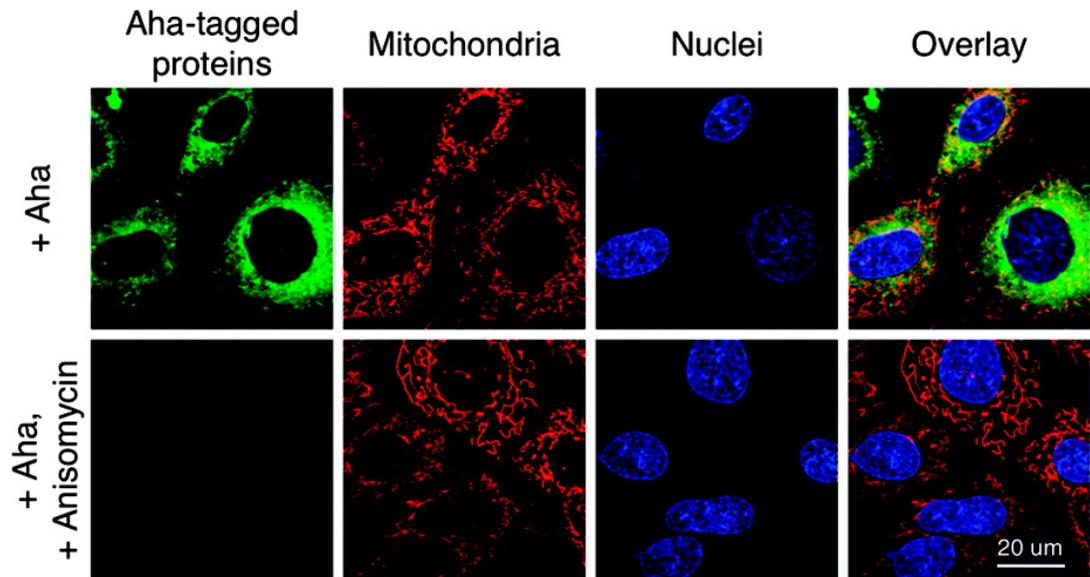
**Structures of the amino acids discussed in this chapter.** ncAAs shown in blue are substrates for the natural translational machinery, the analog shown in green requires overexpression of wild-type MetRS, and those shown in red require expression of mutant aaRSs.

Figure 1.2



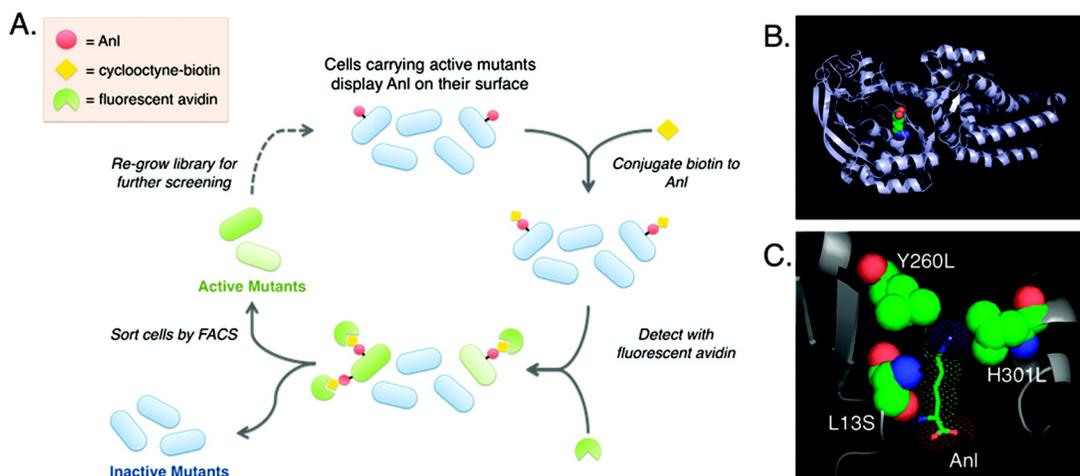
**Detection and identification of newly synthesized proteins.** ncAA-tagged proteins can be ligated to affinity probes for enrichment and identification or to dyes for visualization by in-gel fluorescence scanning or fluorescence microscopy.

Figure 1.3



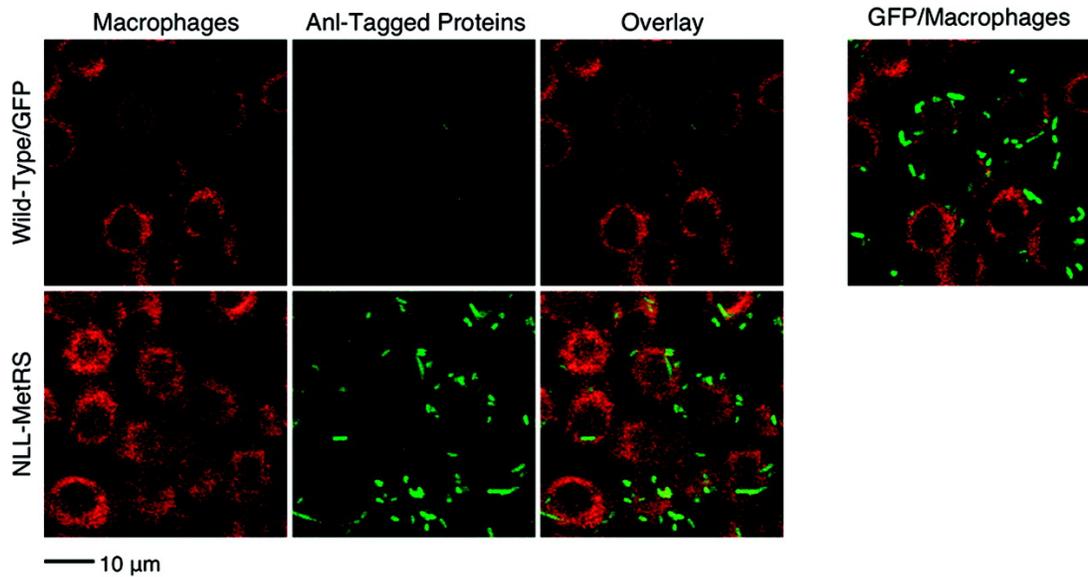
**Visualization of newly synthesized proteins.** Imaging of mammalian cells pulsed with Aha in the absence (top) or presence (bottom) of the protein synthesis inhibitor anisomycin. The left-most panel in each row shows the image formed by dye-labeling of newly synthesized proteins (green). Additional panels show mitochondria (red), nuclei (blue), and panel overlays.

Figure 1.4



**Library screen for MetRS mutants.** (a) Library screen for MetRS mutants that allow Anl incorporation into cellular proteins. (b) X-ray crystal structure of a MetRS mutant (L13S/Y260L/H301L) bound to Anl (spheres). (c) Mutations (spheres) that accommodate Anl (sticks and dots) within the MetRS binding pocket.

Figure 1.5



**Cell-selective labeling in mixtures of bacterial and mammalian cells.** Cell-selective labeling in mixtures of bacterial and mammalian cells. Macrophages (red) infected by *E. coli* cells that express GFP (top) or the NLL-MetRS (bottom) labeled with Anl. Proteins expressed in bacterial cells that express the NLL-MetRS are labeled with Anl. The control bacterial strain was bound and internalized by macrophages (as confirmed by detection of GFP) but exhibited no Anl incorporation. In neither case were macrophage proteins labeled.

## REFERENCES

1. Schoenheimer, R; Rittenberg, D. Deuterium as an indicator of intermediary metabolism. *J. Biol. Chem.* 1935, 111, 163–168.
2. Schoenheimer, R; Ratner, S.; Rittenberg, D. Studies in proteinmetabolism: Themetabolism of tyrosine. *J. Biol. Chem.* 1939, 127, 333–344.
3. Munier, R.; Cohen, G. N. Incorporation of structural analogs of amino acids in bacterial proteins. *Biochim. Biophys. Acta* 1956, 21, 592–593.
4. Saks, M. E.; Sampson, J. R.; Nowak, M. W.; Kearney, P. C.; Du, F; Abelson, J. N.; Lester, H. A.; Dougherty, D. A. An engineered *Tetrahymena* tRNA<sup>Gln</sup> for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression. *J. Biol. Chem.* 1996, 271, 23169–23175.
5. Liu, D. R.; Schultz, P. G. Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. U.S.A.* 1999, 96, 4780–4785.
6. Liu, C. C.; Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* 2010, 79, 413–444.
7. Brustad, E.M.; Lemke, E. A.; Schultz, P. G.; Deniz, A. A. A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J. Am. Chem. Soc.* 2008, 130, 17664–17665.
8. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. An expanded eukaryotic genetic code. *Science* 2003, 301, 964–967.
9. Zhang, Z.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D. S.; Schultz, P. G. Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc Natl. Acad. Sci. U.S.A.* 2004, 101, 8882–8887.
10. Ling, J.; Reynolds, N.; Ibba, M. Aminoacyl-tRNA synthesis and translational quality control. *Annu. Rev. Microbiol.* 2009, 63, 61–78.
11. Nowatzki, P. J.; Franck, C.; Maskarinec, S. A.; Ravichandran, G.; Tirrell, D. A. Mechanically tunable thin films of photosensitive artificial

- proteins: Preparation and characterization by nanoindentation. *Macromolecules* 2008, 41, 1839–1845.
12. Sletten, E. M.; Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chem., Int. Ed.* 2009, 48, 6974–6998.
  13. Khidekel, N.; Ficarro, S. B.; Peters, E. C.; Hsieh-Wilson, L. C. Exploring the O-GlcNAc proteome: Direct identification of O-GlcNAc-modified proteins. *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 13132–13137.
  14. Martin, B. R.; Cravatt, B. F. Large-scale profiling of protein palmitoylation in mammalian cells. *Nat. Methods* 2009, 6, 135–138.
  15. 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.* 2006, 103, 9482–9487.
  16. Jao, C. Y.; Salic, A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci. U.S.A.* 2008, 105, 15779–15784.
  17. Cappella, P.; Gasparri, F.; Pulici, M.; Moll, J. A novel method based on click chemistry, which overcomes limitations of cell cycle analysis by classical determination of BrdU incorporation, allowing multiplex antibody staining. *Cytometry, Part A* 2008, 73, 626–636.
  18. Tornøe, C.W.; Christensen, C.; Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-Triazoles by regioselective copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* 2002, 67, 3057–3064.
  19. 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.* 2002, 41, 2596–2599.
  20. Saxon, E.; Bertozzi, C. R. Cell surface engineering by a modified Staudinger reaction. *Science* 2000, 287, 2007–2010.
  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.* 2004, 126, 15046–15047.

22. Jewett, J. C.; Bertozzi, C. R. Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* 2010, 39, 1272–1279.
23. Debets, M. F.; van der Doelen, C. W.; Rutjes, F. P.; van Delft, F. L. Azide: A unique dipole for metal-free bioorthogonal ligations. *ChemBioChem* 2010, 11, 1168–1184.
24. van Hest, J. C.; Kiick, K. L.; Tirrell, D. A. Efficient incorporation of unsaturated methionine analogues into proteins in vivo. *J. Am. Chem. Soc.* 2000, 122, 1281–1288.
25. Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. U.S.A.* 2002, 99, 19–24.
26. Thingholm, T. E.; Jensen, O. N.; Larsen, M. R. Analytical strategies for phosphoproteomics. *Proteomics* 2009, 9, 1451–1468.
27. Nomura, D. K.; Dix, M. M.; Cravatt, B. F. Activity-based protein profiling for biochemical pathway discovery in cancer. *Nat. Rev. Cancer* 2010, 10, 630–638.
28. Beatty, K. E.; Xie, F.; Wang, Q.; Tirrell, D. A. Selective dye-labeling of newly synthesized proteins in bacterial cells. *J. Am. Chem. Soc.* 2005, 127, 14150–14151.
29. Beatty, K. E.; Liu, J. C.; Xie, F.; Dieterich, D. C.; Schuman, E. M.; Wang, Q.; Tirrell, D. A. Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew. Chem., Int. Ed.* 2006, 45, 7364–7367.
30. Beatty, K. E.; Tirrell, D. A. Two-color labeling of temporally defined protein populations in mammalian cells. *Bioorg. Med. Chem. Lett.* 2008, 18, 5995–5999.
31. Beatty, K. E.; Fisk, J. D.; Smart, B. P.; Lu, Y. Y.; Szychowski, J.; Hangauer, M. J.; Baskin, J. M.; Bertozzi, C. R.; Tirrell, D. A. Live-cell imaging of cellular proteins by a strain-promoted azide-alkyne cycloaddition. *ChemBioChem* 2010, 11, 2092–2095.
32. Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew. Chem., Int. Ed.* 2009, 48, 9879–9883.
33. Szychowski, J.; Mahdavi, A.; Hodas, J. J.; Bagert, J. D.; Ngo, J. T.; Landgraf, P.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. Cleavable

- biotin probes for labeling of biomolecules via azide\_alkyne cycloaddition. *J. Am. Chem. Soc.* 2010, 132, 18351–18360.
34. van Hest, J. C.; Tirrell, D. A. Efficient introduction of alkene functionality into proteins in vivo. *FEBS Lett.* 1998, 428, 68–70.
  35. Floyd, N.; Vijayakrishnan, B.; Koeppe, J. R.; Davis, B. G. Thiyglycosylation of olefinic proteins: S-Linked glycoconjugate synthesis. *Angew. Chem., Int. Ed.* 2009, 48, 7798–7802.
  36. Song, W.; Wang, Y.; Yu, Z.; Vera, C. I.; Qu, J.; Lin, Q. A metabolic alkene reporter for spatiotemporally controlled imaging of newly synthesized proteins in mammalian cells. *ACS Chem. Biol.* 2010, 5, 875–885.
  37. Wang, J.; Zhang, W.; Song, W.; Wang, Y.; Yu, Z.; Li, J.; Wu, M.; Wang, L.; Zang, J.; Lin, Q. A biosynthetic route to photoclick chemistry on proteins. *J. Am. Chem. Soc.* 2010, 132, 14812–14818.
  38. Kiick, K. L.; Van Hest, J. C. M.; Tirrell, D. A. Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host. *Angew. Chem., Int. Ed.* 2000, 39, 2148–2152.
  39. Tang, Y.; Tirrell, D. A. Attenuation of the editing activity of the *Escherichia coli* leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry* 2002, 41, 10625–10645.
  40. Tang, Y.; Wang, P.; Van Deventer, J. A.; Link, A. J.; Tirrell, D. A. Introduction of an aliphatic ketone into recombinant proteins in a bacterial strain that overexpresses an editing-impaired leucyl-tRNA synthetase. *ChemBioChem* 2009, 10, 2188–2190.
  41. Kast, P.; Hennecke, H. Amino acid substrate specificity of *Escherichia coli* phenylalanyl-tRNA synthetase altered by distinct mutations. *J. Mol. Biol.* 1991, 222, 99–124.
  42. Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. Biosynthesis of proteins incorporating a versatile set of phenylalanine analogs. *ChemBioChem* 2002, 3, 2188–2190.
  43. Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins. *J. Am. Chem. Soc.* 2002, 124, 5652–5653.

44. Chalker, J. M.; Wood, C. S. C.; Davis, B. G. A convenient catalyst for aqueous and protein Suzuki\_Miyaura cross-coupling. *J. Am. Chem. Soc.* 2009, 131, 16346–16347.
45. Prescher, J. A.; Bertozzi, C. R. Chemistry in living systems. *Nat. Chem. Biol.* 2005, 1, 13–21.
46. Yoo, T. H.; Tirrell, D. A. High-throughput screening for methionyl-tRNA synthetases that enable residue-specific incorporation of noncanonical amino acids into recombinant proteins in bacterial cells. *Angew. Chem., Int. Ed.* 2007, 46, 5340–5343.
47. Link, A. J.; Vink, M. K.; Agard, N. J.; Prescher, J. A.; Bertozzi, C. R.; Tirrell, D. A. Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 10180–10185.
48. Link, A. J.; Tirrell, D. A. Cell surface labeling of *Escherichia coli* via copper(I)-catalyzed [3 + 2] cycloaddition. *J. Am. Chem. Soc.* 2003, 125, 11164–11165.
49. Link, A. J.; Vink, M. K.; Tirrell, D. A. Presentation and detection of azide functionality in bacterial cell surface proteins. *J. Am. Chem. Soc.* 2004, 126, 10598–10602.
50. Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A., 3rd; Tirrell, D. A. Discovery of *Escherichia coli* methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 2009, 106, 15285–15290.
51. Schmitt, E.; Tanrikulu, I. C.; Yoo, T. H.; Panvert, M.; Tirrell, D. A.; Mechulam, Y. Switching from an induced-fit to a lock-and-key mechanism in an aminoacyl-tRNA synthetase with modified specificity. *J. Mol. Biol.* 2009, 394, 843–851.
52. Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. Cell-selective metabolic labeling of proteins. *Nat. Chem. Biol.* 2009, 10, 715–717.
53. Zhang, C. G.; Chromy, B. A.; McCutchen-Maloney, S. L. Host-pathogen interactions: A proteomic view. *Expert Rev. Proteomics* 2005, 2, 187–202.
54. Grammel, M.; Zhang, M. M.; Hang, H. C. Orthogonal alkynyl amino acid reporter for selective labeling of bacterial proteomes during infection. *Angew. Chem., Int. Ed.* 2010, 49, 5970–5974.

55. Deal, R. B.; Henikoff, J. G.; Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* 2010, 328, 1161–1164.
56. Zhang, M. M.; Tsou, L. K.; Charron, G.; Raghavan, A. S.; Hang, H. C. Tandem fluorescence imaging of dynamic S-acylation and protein turnover. *Proc. Natl. Acad. Sci. U.S.A.* 2010, 107, 8627–8632.
57. Dieterich, D. C.; Hodas, J. J. L.; Gouzer, G.; Shadrin, I. Y.; Ngo, J. T.; Triller, A.; Tirrell, D. A.; Schuman, E. M. In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat. Neurosci.* 2010, 13, 897–905.
58. Tcherkezian, J.; Brittis, P. A.; Thomas, F.; Roux, P. P.; Flanagan, J. G. Transmembrane receptor DCC associates with protein synthesis machinery and regulates translation. *Cell* 2010, 141, 632–644.
59. van Kasteren, S. I.; Kramer, H. B.; Jensen, H. H.; Campbell, S. J.; Kirkpatrick, J.; Oldham, N. J.; Anthony, D. C.; Davis, B. G. Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* 2007, 446, 1105–1109.
60. 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.* 2010, 14, 774–780

## CHAPTER II

### *Cell-Selective Metabolic Labeling of Proteins*

#### ABSTRACT

Metabolic labeling of proteins with the methionine surrogate azidonorleucine can be targeted exclusively to specified cells through expression of a mutant methionyl-tRNA synthetase (MetRS). In complex cellular mixtures, proteins made in cells that express the mutant synthetase can be tagged with affinity reagents (for detection or enrichment) or fluorescent dyes (for imaging). Proteins made in cells that do not express the mutant synthetase are neither labeled nor detected.

#### INTRODUCTION

Time-dependent changes in cellular proteomes can be monitored via a variety of powerful electrophoretic and spectroscopic methods. Traditionally, radiolabeled amino acids have been used to label proteins synthesized during an amino acid 'pulse'; labeled proteins can be distinguished from preexisting (unlabeled) proteins through electrophoretic separation followed by radiographic detection.<sup>1</sup> More recently, mass spectrometry has enabled the use of stable isotopes in amino acid pulse labeling.<sup>2</sup>

In 2006, Dieterich and co-workers introduced the BONCAT (bio-orthogonal non-canonical amino acid tagging) strategy for selective enrichment and identification of newly synthesized proteins in cells.<sup>3,4</sup> The BONCAT approach reduces sample complexity and permits direct analysis of the primary protein synthesis response to stimuli. Bio-orthogonal functional groups<sup>5</sup> are introduced into proteins by pulse-labeling with reactive, non-

canonical amino acids. Labeled proteins are selectively modified with affinity tags for enrichment;<sup>6,7</sup> removal of unlabeled proteins simplifies subsequent analysis and identification by mass spectrometry.

All of these methods suffer from limitations when experiments are performed in systems that contain multiple cell types. Because incorporation of amino acids is non-specific with respect to cell identity, proteins from all cell types are labeled. In studies of interactions between different cell types in a single organism, the origin of the identified proteins can be difficult to ascertain because the cells share a common genome. When interactions between cells of different genomes are studied, detection of low abundance proteins can be problematic. In infection studies, for example, the protein content of the larger host cells can overwhelm that of the pathogen<sup>8</sup> and limit detection and identification of the proteins of primary interest. In complex bacterial communities where hundreds of organisms can occupy a common biological niche,<sup>9</sup> probing the proteome of a single species in its natural context is an even greater challenge.

## RESULTS AND DISCUSSION

To address these difficulties, we describe here a versatile method for cell-selective protein labeling in mixed cellular environments. To achieve selective labeling, we employ non-canonical amino acids that are excluded by the endogenous protein synthesis machinery (Fig. 2.1a). These amino acids face discrimination by the quality control mechanisms found at the level of aminoacyl-tRNA synthetases;<sup>10</sup> they are not charged to tRNA and are not used in protein synthesis. By screening libraries of methionyl-tRNA synthetase (MetRS) mutants from *Escherichia coli*<sup>11,12</sup> we have identified a mutant synthetase (NLL-MetRS) (Supplementary Fig. 2.1) that efficiently appends azidonorleucine (**2**, Fig. 2.1b) to cognate tRNA. Cells bearing the mutant MetRS are able to utilize **2** as a surrogate for methionine (**1**) in protein

synthesis. Wild-type cells are inert to **2**; proteins made in these cells utilize only methionine and are not labeled (Fig. 2.1a). In co-culture, protein labeling is restricted to mutant cells.

To validate this approach, we first confirmed that incorporation of **2** into newly synthesized proteins is dependent on expression of the mutant synthetase. An *E. coli* strain (DH10B/pJTN1) constitutively expressing a plasmid-borne copy of NLL-MetRS was pulse-labeled with **2** and compared to a control strain (DH10B/pQE-80L) that did not express the enzyme. Separate cultures of the two strains were grown in minimal medium containing the twenty canonical amino acids. When the cell density reached OD<sub>600</sub> = 0.5, cells were pulse-labeled with 1 mM **2** for 10 m. Control cells were pulsed in the same fashion with **1**, or incubated with the protein synthesis inhibitor chloramphenicol prior to labeling with **2**. Cell lysates of each culture were probed for incorporation of **2** via Cu (I) catalyzed ligation<sup>13,14</sup> to biotin- FLAG-alkyne (**4**) followed by Western blotting with protein detection by anti-FLAG antibody. The results of these experiments indicated that only proteins synthesized in cells constitutively expressing the NLL-MetRS (DH10B/pJTN1) were labeled with **2** and susceptible to ligation to **4** (Supplementary Fig. 2.2). A second control strain, in which the wild-type synthetase was over-expressed, was also inert to labeling (Supplementary Fig. 2.1).

The behavior observed in separate cultures was maintained when cells were incubated in co- culture to simulate a complex, mixed cellular environment. Two different heterologous proteins were employed as markers for the cells of origin to distinguish between NLL- expressing and wild-type *E. coli*. An *E. coli* strain (DH10B/pJTN2) expressing the NLL- MetRS was programmed to express green fluorescent protein (GFP) upon induction with isopropyl β-D-1-thiogalactopyranoside (IPTG). The control strain (DH10B/pJTN3) carried an IPTG-inducible gene for the marker protein DHFR. Both marker proteins were His- tagged to enable Ni-affinity purification and detection with Penta-

His antibody. Individual cultures of these bacterial strains were grown to  $OD_{600} = 1.0$  and a third culture was created by mixing cells in a volumetric ratio of 1:2 (DH10B/pJTN2:DH10B/pJTN3). To initiate labeling, 1 mM **2** was added to each of the three cultures and expression of marker proteins was induced with 1 mM IPTG for 3 h. Cell lysates from all three samples were subjected to Cu (I) catalyzed azide-alkyne ligation to **4**, and marker proteins were isolated by Ni-affinity purification. Western analysis (Fig. 2.2a) of isolated proteins with Penta-His antibody revealed the expected expression patterns from the three cultures. In striking contrast, analysis of blots with streptavidin-HRP (for detection of conjugation to **4**) revealed exclusive modification of GFP, the marker protein synthesized in cells expressing NLL-MetRS. DHFR isolated from control cells exhibited no such modification. N-terminal protein sequencing indicated 10-20% replacement of **1** by **2** in the GFP marker protein.

To demonstrate further the utility of this approach, we used fluorescence microscopy to distinguish proteins made in cells expressing the NLL-MetRS from those made in cells that do not express the mutant synthetase. The control strain (DH10B/pJTN4) expressed an IPTG-inducible GFP while the strain (DH10B/pJTN5) constitutively expressing the NLL-MRS carried an IPTG-inducible DHFR. Labeling of individual and mixed cultures were performed as described earlier. After pulse-labeling with **2** and induction of protein synthesis, cells were collected by centrifugation and washed prior to Cu (I) catalyzed labeling of cells with dimethylaminocoumarin-alkyne (**5**). After washing with PBS to remove excess dye, cells were imaged by fluorescence microscopy. The results (Fig. 2.2b) were consistent with those of the Western analysis; the coumarin fluorescence was confined to cells that express NLL-MetRS. Control cells expressing GFP were inert with respect to labeling, as indicated by the absence of coumarin emission from these samples (Supplementary Fig. 2.3).

Cell-selective protein labeling can also be accomplished in systems containing mixtures of bacterial and mammalian cells. Murine alveolar macrophages were infected with *E. coli* cells that constitutively express the NLL-MetRS (DH10B/pJTN1) or with control bacterial cells that express a GFP marker protein (DH10B/pJTN4). Prior to infection, 2 mM **2** was added to the macrophage medium; to initiate infection, bacteria were added to the culture medium and co-incubated for 35 m at 37 °C. Cells were fixed, permeabilized, and subjected to Cu (I) catalyzed conjugation to TAMRA-alkyne (**6**, Invitrogen). Bacteria were both bound and internalized by macrophages as confirmed by confocal microscopy and three- dimensional analysis (Supplementary Movie 1 online). Macrophage-associated bacterial cells that express the NLL-MetRS exhibited strong fluorescence emission from **6** (Fig. 2.3a). The control bacterial strain was bound and internalized by macrophages (as seen by detection of GFP, Supplementary Fig. 2.4 online) but exhibited no conjugation to **6** above background (Fig. 2.3a). To confirm protein synthesis by macrophages during the infection period, cells were treated with azidohomoalanine (**3**) in medium lacking **1** (Fig. 2.3b). Both **2** and **3** are susceptible to ligation to alkyne-functionalized probes; however, in contrast to **2**, **3** is activated by wild-type MetRS and does not discriminate between cell types<sup>4</sup>. As shown in Fig 2.3, both bacterial cells and macrophages were labeled with **3**, whereas labeling with **2** was observed only in bacterial cells that express the NLL-MetRS.

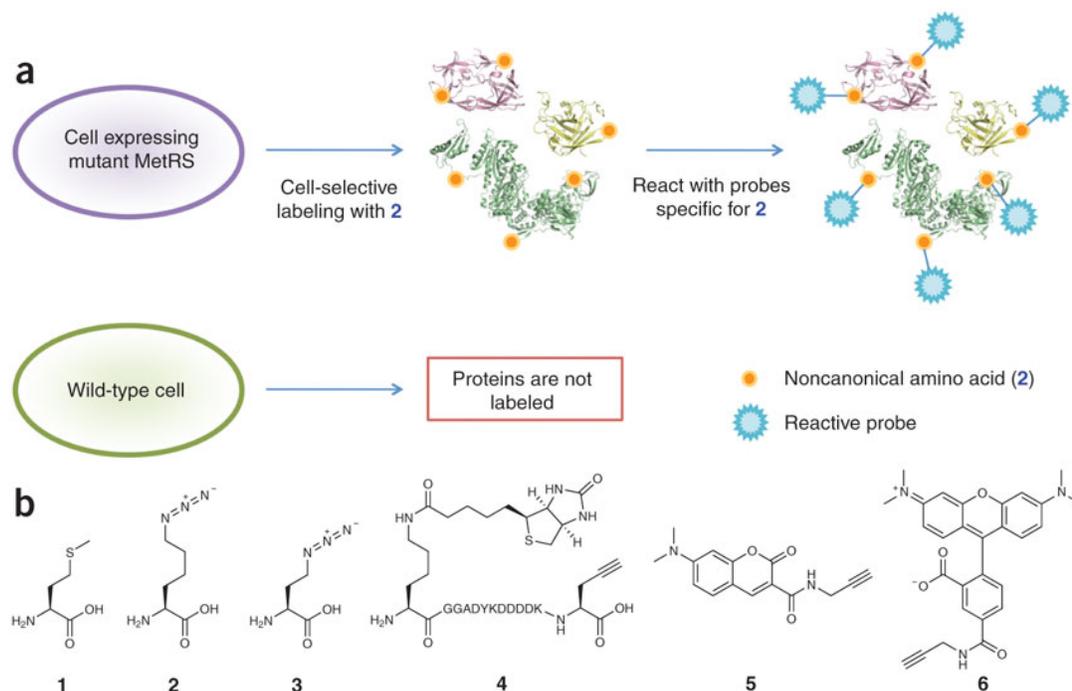
Newly synthesized bacterial proteins can be enriched from such cultures by affinity chromatography. Using an *E. coli* strain that expresses the NLL-MetRS constitutively and GFP under induction with IPTG, we infected macrophages in medium containing 2 mM **2**. Immediately upon infection, IPTG was added to initiate bacterial synthesis of GFP. After 35 m at 37 °C, the total cell mixture was collected by centrifugation and lysed. Proteins were subjected to Cu (I) catalyzed azide-alkyne ligation with alkyne-functionalized biotin. Biotinylated proteins were selectively enriched by collection on

neutravidin-agarose beads. After five washes, proteins were eluted from the resin with 2 mM free biotin and 2% SDS. To examine the extent of enrichment, the lysate, resin flow-through, washes and eluent were analyzed by immunoblot (Fig. 2.3c). The mammalian protein  $\beta$ -actin was detected with anti- $\beta$ -actin and served as a representative macrophage protein. The bacterial marker GFP was detected with anti-Penta-His antibody. No actin was detected in the eluent, indicating at least 50-fold depletion of the mammalian marker. In contrast, comparison of the GFP band intensities in the eluent and lysate confirmed good recovery of the affinity-tagged bacterial protein.

## CONCLUSION

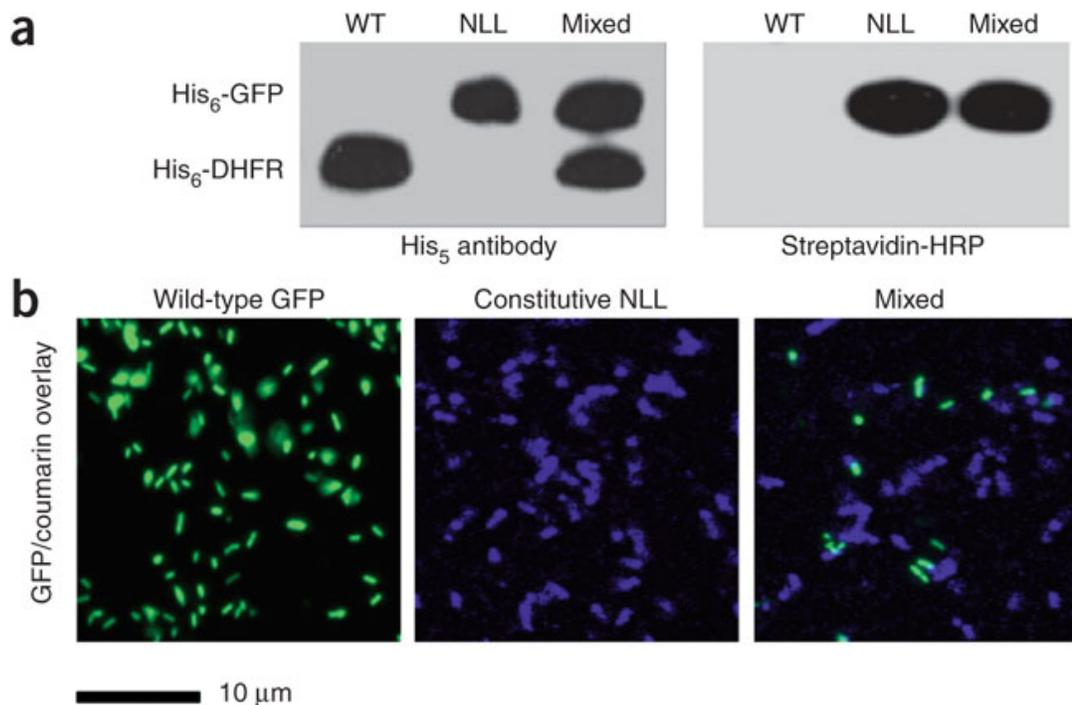
The results described here illustrate the use of mutant aminoacyl-tRNA synthetases to enable cell-specific protein labeling with non-canonical amino acids. In mixed cellular systems, newly synthesized proteins in selected cells can be labeled with affinity reagents or fluorescent dyes for enrichment, identification, and visualization. This approach will enable unambiguous determination of the cellular origins of proteins made in complex multicellular systems and will provide new insight into intercellular communication. We are expanding the studies described here by engineering new amino acid/synthetase pairs and by using the azidonorleucine/NLL-MetRS pair to examine a variety of intercellular interactions.

Figure 2.1



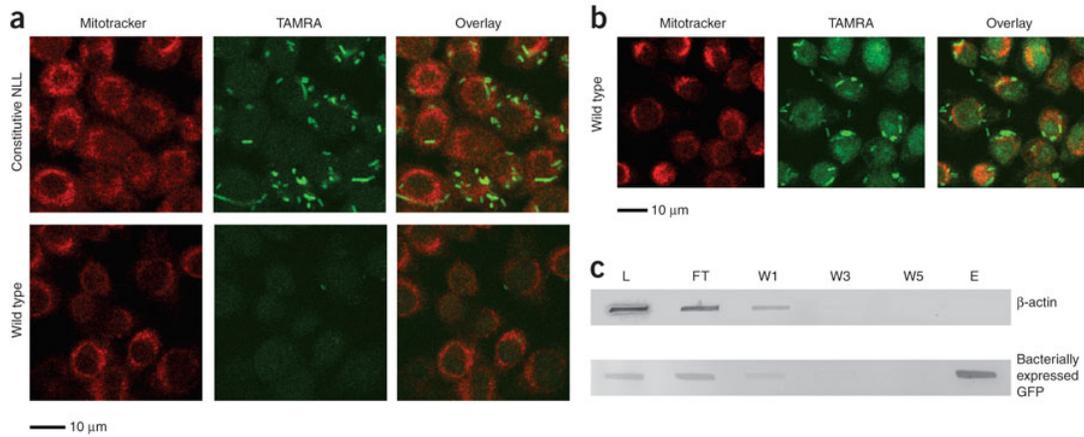
**Cell-selective labeling of proteomes with azidonorleucine.** (a) Schematic representation of incorporation of azidonorleucine exclusively in cells expressing NLL-MRS. (b) Structures of amino acids and probes used in this study: (1) methionine, (2) azidonorleucine, (3) azidohomoalanine, (4) biotin-FLAG-alkyne, (5) dimethylaminocoumarin-alkyne, (6) TAMRA-alkyne.

Figure 2.2



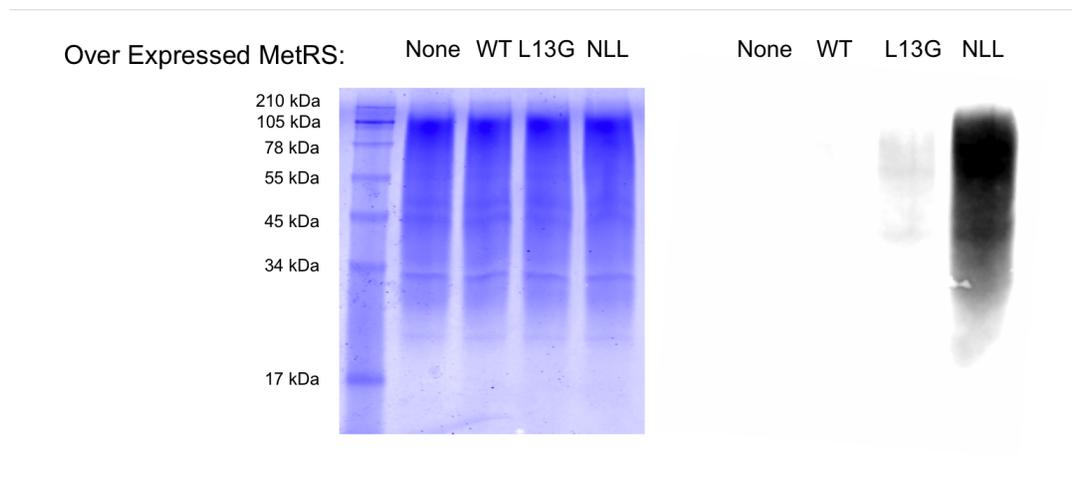
**Cell-selective protein labeling in mixed populations.** (a) Western blot detection of marker protein expression with Penta-His antibody (left) and with streptavidin-HRP (right). DHFR was made in cells lacking the NLL-MetRS (WT) and GFP in cells expressing the NLL-MetRS (NLL); both proteins contain multiple ATG (Met) codons (7 in DHFR and 8 in GFP). Azidonorleucine was added to all three cultures upon induction of protein synthesis. Samples from the mixed culture contain both proteins (as shown by Penta-His antibody detection), but only GFP is labeled with azidonorleucine and susceptible to labeling with **4**. (b) Fluorescence images of cells expressing GFP but not NLL-MetRS (Wild-Type), cells expressing DHFR and NLL-MetRS (NLL), and a mixed culture of the two (Mixed). Azidonorleucine was added to all three cultures upon induction of protein synthesis. Cells from all three cultures were treated with **5**, but only cells expressing the NLL-MetRS were labeled. Note that the marker proteins co-expressed with NLL-MetRS are different in the Western blotting and fluorescence imaging experiments.

**Figure 2.3**



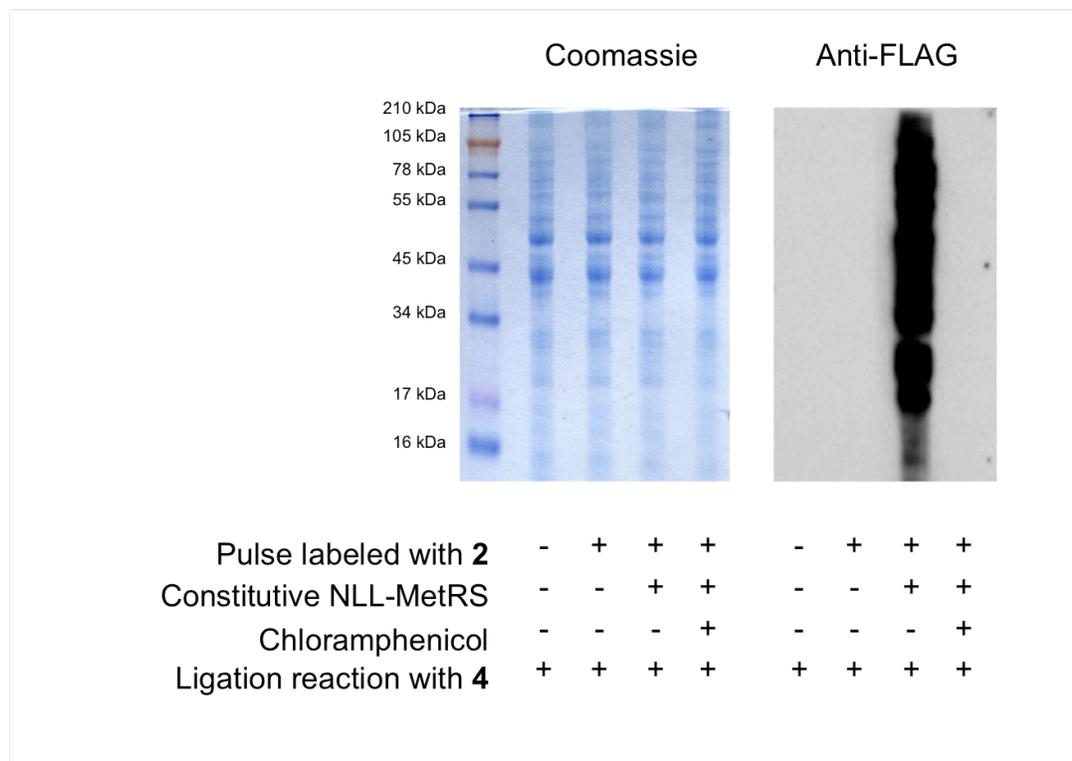
**Cell-selective labeling in mixtures of bacterial and mammalian cells. (a)** Fluorescence images of mixed cultures containing bacteria attached to or internalized by murine alveolar macrophages. Infection was performed in medium containing azidonorleucine. Bacterial cells constitutively expressing the NLL-MetRS were labeled by TAMRA-alkyne (Constitutive NLL) while cells lacking the NLL-MetRS (Wild type GFP) are visible only in the GFP channel (Supplementary Fig. 2.4). Macrophages were labeled with Mitotracker Deep Red (Invitrogen) and exhibited very low TAMRA background emission. In all cases, conjugation of TAMRA-alkyne was confined to bacterial cells expressing the NLL-MetRS. **(b)** Fluorescence images of macrophage infection with wild-type bacteria performed in the presence of azidohomoalanine. Protein synthesis by macrophages is indicated by strong TAMRA-alkyne emission from both bacterial cells and macrophages. **(c)** Macrophages were infected with bacterial cells that express GFP under induction with IPTG and constitutively express the NLL-MetRS. Infection was performed in medium containing IPTG and azidonorleucine to facilitate bacterial synthesis and labeling of GFP. Total cell lysate from the infection was subjected to conjugation with alkyne-functionalized biotin; labeled proteins were enriched with streptavidin avidity. Bacterially expressed GFP and mammalian  $\beta$ -actin were followed by immunoblot. Analysis of the lysate (L), unbound flow-through (FT), washes (W1, W3, W5), and eluent (E) reveal a separation of bacterial and mammalian representative proteins. Bacterially expressed GFP was labeled with azidonorleucine and thus subject to conjugation to biotin and enrichment with streptavidin. Protein originating from macrophages, including  $\beta$ -actin, were not labeled with 2 and therefore not conjugated to alkyne-functionalized biotin.

## Supplementary Figure 2.1



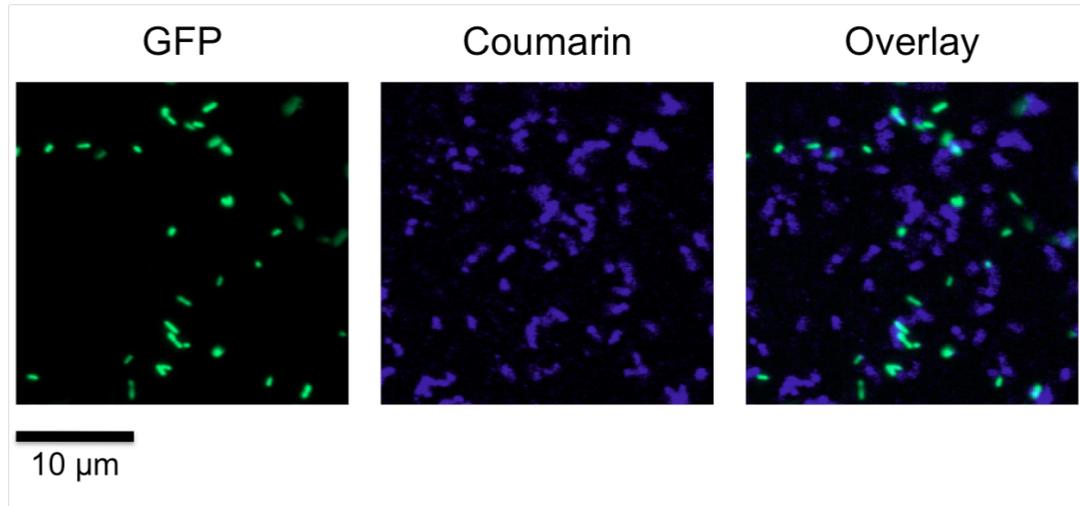
**Comparison of cells Anl-incorporation by cells over-expressing plasmid-borne copies of either WT-MetRS, L13G-MetRS, or NLL-MetRS.** The NLL mutant MetRS (NLL-MetRS) carries the mutations Leu13Asn, Tyr260Leu, and His301Leu and was isolated by methods described in Reference 11. Multiple rounds of high-throughput screening were performed in 1 mM azidonorleucine with increasing concentrations of methionine (from 0.01 to 0.1 mM) for isolation of highly active clones capable of charging azidonorleucine in the presence of methionine. In the work described in Reference 11 (Link *et al*, 2006), screening was done in methionine-depleted medium. The L13G mutant identified in that work affords only low levels of protein labeling with azidonorleucine in media supplemented with methionine. Over-expression of WT MetRS is insufficient for incorporation of Anl as shown by the absence of signal in western blots visualized with Anti-FLAG-HRP. Previously-isolated MetRS mutant L13G activates Anl, but yields low levels of Anl incorporation in media supplemented with Met.

## Supplementary Figure 2.2



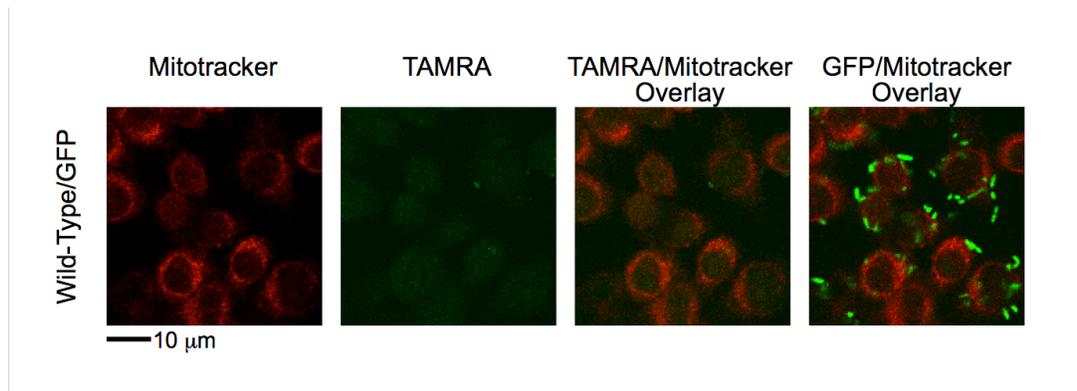
**Detection of Anl-tagged proteins.** Coomassie detection of proteins from cells pulse-labeled with azidonorleucine (left) and Western blot detection of proteins conjugated to biotin-FLAG-alkyne (right). Azidonorleucine labels newly synthesized proteins in cells expressing the NLL-MetRS. Wild-type cells, cells not exposed to azidonorleucine, and mutant cells exposed to the protein synthesis inhibitor chloramphenicol are not labeled with azidonorleucine; proteins from these cells do not conjugate with biotin-FLAG-alkyne.

Supplementary Figure 2.3



**Individual panels from imaging of mixed bacterial cultures from Figure 2.2b.** Cu-catalyzed ligation to coumarin dye is restricted to cells expressing the NLL-MetRS. GFP-expressing wild-type cells not expressing the mutant enzyme do not exhibit fluorescence from the coumarin dye.

## Supplementary Figure 2.4



**Infection of macrophages is performed with control bacteria expressing GFP, but not the NLL-MetRS (Wild-Type/GFP).** The absence of emission from the TAMRA indicates that both macrophages and bacteria not expressing the NLL-MetRS are unable to utilize **2** in protein synthesis. The presence of Wild-Type/GFP bacteria is confirmed through detection of GFP fluorescence.

## MATERIALS AND METHODS

### Reagents

Azidonorleucine (2) and azidohomoalanine (3) were prepared as previously described<sup>1</sup>, both were confirmed by <sup>1</sup>H NMR and mass spectrometry, and stored at 100 mM in water at 4 °C. The triazole ligand (Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine) was prepared as described previously<sup>2</sup>, confirmed by <sup>1</sup>H NMR and mass spectrometry, and stored at 20 mM in DMSO at -20 °C. The biotin-FLAG-alkyne reagent (4) was synthesized by GenScript Corporation, was determined pure by mass spectrometry, and stored at 5 mM in water at -20 °C. The dimethylaminocoumarin-alkyne (5) was synthesized as previously described<sup>3</sup>, confirmed by <sup>1</sup>H NMR and mass spectrometry, and stored at 40 mM in DMSO at -20 °C. The Click-iT Tetramethylrhodamine (TAMRA) Protein Analysis Detection Kit dye was purchased from Invitrogen and the TAMRA-alkyne (6) dye was used as obtained. CuBr (99.999% purity) was purchased from Sigma and suspended in water at 20 mM immediately before use.

### Strains and Plasmids

pQE-80L: was used as obtained from Qiagen.

pJTN1: The gene encoding wild-type *E. coli* MetRS was isolated by digestion of pAJL611 with NheI. The purified fragment containing the gene cassette for constitutive expression of *E. coli* MetRS under control of its natural promoter was then ligated into pQE-80L. The mutations L13N, Y360L, H301L were generated by three sequential site-directed mutageneses according to the manufacturer's protocol from the commercially available QuikChange kit (Stratagene).

pJTN2: pJTN1 was digested with NheI and the gene cassette for constitutive expression of NLL-MetRS was isolated by gel purification. This fragment was inserted into the NheI site of pJTN4 to generate pJTN2.

pJTN3: Renamed from pAJL60 previously reported<sup>4</sup>.

pJTN4: Renamed from pQE-80L/GFP<sub>rm</sub>\_AM previously reported<sup>5</sup>.

pJTN5: pJTN1 was digested with NheI and the gene cassette for constitutive expression of NLL-MetRS was isolated by gel purification. This fragment was ligated into the NheI site of pJTN3 to generate pJTN5.

pJTN6: pAJL61 was digested with NheI and the gene cassette for constitutive expression of WT-MetRS was isolated by gel purification. This fragment was inserted into the NheI site of pJTN4 to generate pJTN6.

pJTN7: pAJL61 was subjected to site-directed mutagenesis to generate the L13G mutation on the region that encodes the MetRS.

Sequencing of plasmids was performed by Laragen. All plasmids were transformed into the XL-1 Blue strain of *E. coli* (Stratagene) for cloning. Transformants were grown on plates containing 200 ug ml<sup>-1</sup> ampicillin. For pulse-labeling experiments, plasmids were transformed into the DH10B strain of *E. coli* (Invitrogen). Transformants were grown on ampicillin agar plates or maintained in media with 200 ug ml<sup>-1</sup> ampicillin.

### **Proteomic Pulse Labeling and Protein Conjugation Reactions**

DH10B/pQE-80L and DH10B/pJTN1 were used for this experiment. Cells were diluted 1:50 from an overnight culture in LB into M9 minimal medium and grown at 37°C with agitation at 250 rpm. When OD<sub>600</sub> = 0.5 was reached, cells were pulse-labeled by addition of 1 mM azidonorleucine to the medium for 10 m. A culture of DH10B/pQE-80L in which no azidonorleucine was added was used as a control. As an additional control, a

culture of DH10B/pJTN1 was treated with 100 ug ml<sup>-1</sup> of chloramphenicol for 10 m prior to addition of azidonorleucine, to inhibit protein synthesis. Cells were collected by centrifugation at 5,000 g at 4°C for 5 m, resuspended in PBS (pH 7.8) with 10% SDS, and heated at 95°C for 5 m for lysis. PBS (pH 7.8) was added to each sample to bring the concentration of SDS to 1%. Azidonorleucine-labeled proteins were then conjugated to probe by addition of 200 uM triazole ligand, 25 uM biotin-FLAG-alkyne, and 400 uM CuBr. The reaction was allowed to proceed for 16 h at room temperature with gentle agitation. Proteins were precipitated by addition of trichloroacetic acid (TCA) to 10%, and the resulting pellets were washed twice with cold acetone. Proteins were resuspended in SDS-PAGE loading buffer and separated on two parallel 12% Tris-Tricine gels. One gel was stained with Coomassie-Brilliant Blue R-250 for non-specific detection of proteins. Proteins from the second gel were transferred to nitrocellulose and blocked in PBS-Tween containing 5% milk for Western blot analysis with anti-FLAG-HRP conjugate (Sigma) at a dilution of 1:20,000. Detection was achieved by chemiluminescence with Super Signal West Pico Chemiluminescent Substrate (Pierce) and x-ray BioMax film (Kodak).

### **Comparison of AnI Incorporation in Strains Over-Expressing the WT-MetRS, L13G Met-RS, or NLL-MetRS**

DH10B/pJTN6, DH10B/pJTN7 and DH10B/pJTN2 were used for this experiment. Procedures for growth and pulse labeling of cells, as well as conjugation chemistry, processing and Western analysis of proteins were performed as described above.

### **Model Protein Expression, Labeling, Conjugation, and Purification**

DH10B/pJTN2 and DH10B/pJTN3 were used in this experiment. Cells were diluted 1:50 from an overnight LB culture into M9 minimal medium and grown at 37°C with agitation at 250 rpm. When OD<sub>600</sub> = 1.0 was reached a

third culture was generated by mixing DH10B/pJTN2 and DH10B/pJTN3 in a 1:2 volumetric ratio. To initiate protein labeling and synthesis, 1 mM azidonorleucine and 1 mM IPTG were added to the individual cultures of DH10B/pJTN2 and DH10B/pJTN3, as well as to the mixed culture. After 3 h, cells were collected by centrifugation at 5,000 g for 5 m, resuspended in PBS (pH 7.8) with 10% SDS, and heated at 95°C for 5 m to lyse cells. PBS (pH 7.8) was added to each sample to bring the concentration of SDS to 1%. Azidonorleucine-labeled proteins were conjugated to probe by addition of 200 uM triazole ligand, 25 uM biotin-FLAG-alkyne, and 400 uM CuBr suspension in water. The reaction was allowed to proceed for 16 h at room temperature with gentle agitation. Proteins were precipitated by addition of TCA to 10%, and the resulting pellets were washed twice with cold acetone. Proteins were resuspended in Buffer B6 with 10 uM  $\beta$ -mercaptoethanol and His-tagged proteins were isolated on Ni-NTA Spin Kit columns (Qiagen) as directed by the manufacturer's protocol. Proteins were separated on 12% Tris-Tricine gels, transferred to nitrocellulose, blocked with PBS-Tween containing 5% milk, and analyzed by Western blotting with Penta-His Antibody (Qiagen) at a dilution of 1:10,000 and with Streptavidin-HRP (Pierce) at a dilution of 1:5,000. Antibodies were detected as described above.

### **Bacterial Dye Labeling and Imaging**

DH10B/pJTN4 and DH10B/pJTN5 were used in this experiment. Cells were diluted 1:50 from overnight cultures in LB into M9 minimal medium and grown at 37°C with agitation at 250 rpm. When OD600 of 1.0 was reached, a third culture was generated by mixing DH10B/pJTN5 and DH10B/pJTN4 in a 1:2 volumetric ratio. Each culture was then aliquoted into a clear-bottom, black 96-well plate and cells were allowed to settle for 2 h at 37°C. To initiate protein labeling and synthesis, 1 mM azidonorleucine and 1 mM IPTG were added to each well. After 4 h, cells were collected by centrifugation at 3,000 g for 5 m. Cells were washed twice and resuspended in PBS (pH 7.8). Dye labeling was performed by addition of 200 uM triazole ligand, 10 uM

dimethylaminocoumarin-alkyne, and 400  $\mu$ M CuBr suspension with rocking for 16 h at 4°C. Prior to imaging, cells were washed twice with cold PBS (pH 7.8). Images were obtained using a Zeiss LSM510 confocal microscope (LSCM) with a 100x oil immersion lens. Images represent overlays of detected GFP and coumarin emission as previously described.<sup>7,8</sup>

### **Macrophage Infection Pulse with Azidonorleucine, Dye Labeling and Imaging**

The mouse alveolar macrophage cell line MH-S was obtained from ATCC and maintained in RPMI medium supplemented with 10% FBS, 0.05 mM  $\beta$ -mercaptoethanol, and 1% penicillin/streptomycin (Sigma) under standard culture conditions (37°C, 5% CO<sub>2</sub>, humidified). DH10B/pJTN1 and DH10B/pJTN4 were inoculated 1:40 from an overnight culture into 2XYT medium. GFP synthesis was induced by the addition of 1 mM IPTG to the culture of DH10B/pJTN4 at an OD<sub>600</sub> = 0.3. Bacterial cells were allowed to grow for 3.5 h after initial induction, diluted into 2XYT to obtain OD<sub>600</sub> = 1.0, and placed on ice. At 20 h prior to infection, macrophages were passaged into wells on a LabTek II Chamber Slide System (Nalge-Nunc) at 4x10<sup>5</sup> cells ml<sup>-1</sup> in 0.5 ml maintenance media and kept under standard culture conditions. Medium was removed and replaced with antibiotic-free medium containing 100 nM Mitotracker Deep Red (Invitrogen). Mitochondrial staining took place for 30 m under standard culture conditions. Macrophages were rinsed twice with fresh antibiotic-free medium and chilled to 4°C for 15 m. Prior to infection, 2 mM of azidonorleucine was added to the cold macrophage cultures. To initiate infection, E. coli cells were immediately added to the medium at a multiplicity of infection of 1:100. Cells were spun at 4°C at 100 g for 10 m to promote attachment of bacteria to macrophages and synchronize infection. To allow internalization of bacteria, cells were warmed to 37°C for 35 m. Following infection, cells were washed twice with PBS to remove unbound bacteria and fixed with 3.7% paraformaldehyde for 15 m at 37°C. Fixative was removed by washing twice with PBS. Cells were

permeabilized with 0.1% Triton-X100 at room temperature for 3 m and washed three times with PBS (pH 7.8). The dye labeling solution was prepared 5 m before addition to each well and contained 200 uM ligand, 400 uM tris(2-carboxyethyl)phosphine hydrochloride, 200 uM CuSO<sub>4</sub>, and TAMRA-alkyne diluted 1:750 as obtained from the manufacturer. Dye solution was added to each well and labeling was performed for 16 h at room temperature with gentle agitation. The following day, cells were rinsed three times with PBS containing 0.5 mM EDTA and 1% Tween-20. Well dividers were removed and samples were covered with mounting media (composed of 50% glycerol and 50% PBS) and coverslips. Samples were imaged on a Zeiss LSM 510 NLO three-channel confocal microscope with a 63x oil immersion lens. Mitotracker, TAMRA, and GFP were excited at 633, 543, and 488 nm, respectively. Images of Mitotracker and TAMRA or Mitotracker and GFP from the middle of a confocal stack were overlaid to create the figures. Bacteria appearing within the Mitotracker stain are internalized while those appearing on the edge are attached to the outside of macrophages.

### **Macrophage Infection Pulse with Azidohomoalanine, Dye Labeling and Imaging**

An additional control was performed for the E.coli/macrophage experiment to ensure that macrophages were indeed synthesizing protein and that the lack of macrophage fluorescence emission in the TAMRA channel was not due to an inability of the macrophages to synthesize protein in co-culture. The control experiment was performed as above except pulse labeling was performed with 2 mM azidohomoalanine instead of azidonorleucine. Azidohomoalanine can be incorporated by the endogenous protein synthesis machinery and can label proteins synthesized in macrophages. The experiment was performed in methionine-free medium (Dulbecco's Modified Eagle's Medium supplemented with 1% Glutamax (Sigma)). Conditions for infection, dye labeling, and analysis were identical to those described above.

### **Isolation of Labeled, Bacterially Expressed Proteins from Infection**

The mouse alveolar macrophage cell line MH-S and the E. coli strain DH10B/pJTN2 were used for this experiment. DH10B/pJTN2 from an overnight culture in LB was inoculated 1:50 in fresh LB and grown for three hours. Macrophages ( $2.12 \times 10^8$  cells) were suspended in RPMI medium supplemented with 10% FBS, and 1% penicillin/streptomycin in a 125 mL shake flask and agitated gently with rotation at 60 rpm at 37°C. DH10B/pJTN2 cells were then pelleted by centrifugation at 6,000 g for 1 m, washed once in 0.9% NaCl, and resuspended in 0.9% NaCl. Azidonorleucine was added to the macrophage medium to 2 mM, DH10B/pJTN2 was then added to the macrophage culture ( $2.12 \times 10^9$  cells) and 1 mM IPTG was added immediately to induce bacterial expression of GFP. After a 35 m with gentle agitation at 60 rpm at 37°C, all cells were collected by two sequential centrifugation steps at 100 g and 6000 g respectively. Cells were combined and lysed in PBS pH 7.4 with 2% SDS with heating to 85°C for 5 m. The SDS concentration was diluted to 0.5%, and insoluble lysate debris was removed by centrifugation at 15,000 g for 30 m at room temperature. The resulting solution was subjected to conjugation to the commercially available biotin-alkyne (7) (Invitrogen). Conjugation was performed with 200  $\mu$ M triazole ligand, 100  $\mu$ M biotin-alkyne, and 400  $\mu$ M CuBr. The reaction was allowed to proceed over night with rotation. The following day, excess unreacted biotin was removed from the mixture using PD-10 (GE) desalting columns pre-equilibrated with PBS pH 7.4 with 0.5% SDS. The proteins were then subjected to affinity purification with Neutravidin Resin (Pierce). Prior to binding, 2 mL of resin was washed with PBS pH 7.4 with 0.5% SDS then added to the protein solution and allowed to bind at room temperature for 1 h with rotation. The unbound proteins were then removed and the resin was subjected to five 10 mL washes with PBS pH 7.4 with 0.2% SDS. The proteins were then eluted in 2 mL of PBS pH 7.4 with 2% SDS and 2 mM free biotin with boiling for 10 m. The unreacted sample, unbound proteins (flow

through), wash 1, wash 3, wash 5, and the eluent were then subjected to immuno-blot using the Bio-Dot Slot-Format apparatus (Bio-Rad) to immobilize proteins on two nitrocellulose membranes according to the manufacturer's protocol. Both blots were blocked in PBS-Tween containing 5% milk. One blot was probed with Anti-Penta-His Alexa 647 Conjugate (Qiagen) according to supplier's protocols. The second blot was probed first with Mouse Monoclonal Anti- $\beta$ -Actin Antibody (Sigma) then with a secondary Rabbit Anti-Mouse Cy5 (Millipore) both according to suppliers' recommendations. Blots were then analyzed by fluorescence scanning using the Typhoon Trio (GE). Densitometric analysis was performed using the companion software ImageQuant (GE).

## REFERENCES

1. Beynon, R.J. & Pratt, J.M. *Mol. Cell. Proteomics* 4, 857–872 (2005).
2. Mann, M. *Nat. Rev. Mol. Cell Biol.* 7, 952–958 (2006).
3. Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A. & Schuman, E.M. *Proc. Natl. Acad. Sci. USA* 103, 9482–9487 (2006).
4. Dieterich, D.C. et al. *Nat. Protoc.* 2, 532–540 (2007).
5. Prescher, J.A. & Bertozzi, C.R. *Nat. Chem. Biol.* 1, 13–21 (2005).
6. Beatty, K.E., Xie, F., Wang, Q. & Tirrell, D.A. *J. Am. Chem. Soc.* 127, 14150–14151 (2005).
7. Beatty, K.E. et al. *Angew. Chem. Int. Ed.* 45, 7364–7367 (2006).
8. Zhang, C.G., Chromy, B.A. & McCutchen-Maloney, S.L. *Expert Rev. Proteomics* 2, 187–202 (2005).
9. Macpherson, A.J. & Harris, N.L. *Nat. Rev. Immunol.* 4, 478–485 (2004).
10. Ibba, M. & Söll, D. *Science* 286, 1893–1897 (1999).
11. Link, A.J. et al. *Proc. Natl. Acad. Sci. USA* 103, 10180–10185 (2006).
12. Yoo, T.H. & Tirrell, D.A. *Angew. Chem. Int. Ed.* 46, 5340–5346 (2007).
13. Tornøe, C.W., Christensen, C. & Meldal, M. *J. Org. Chem.* 67, 3057–3064 (2002).
14. Lewis, W.G. et al. *Angew. Chem. Int. Ed.* 41, 1053–1057 (2002).

## CHAPTER III

### *Promoter-Driven Tagging: Profiling Cellular Protein Synthesis Under Specific Physiological States*

#### ABSTRACT

Transcriptional activity from a specified promoter can provide a useful marker for the physiological state of a cell. Here we introduce a method for tagging proteins made in cells in which specified promoters are active. Tagged proteins can be modified with affinity reagents for enrichment or with fluorescent dyes for visualization. The method allows “genetic dissection” of the proteome, whereby proteins synthesized in pre-determined physiological states can be identified.

#### INTRODUCTION

Cell function and survival rely on the proper regulation of gene expression to ensure that proteins are synthesized in response to internal and external demands. Proteins that execute basic cellular functions (“housekeeping proteins”) are constitutively expressed, while expression of other genes may be restricted to specific physiological states. For instance, in bacterial pathogens, ribosomal proteins may be synthesized continually whereas expression of virulence factors may be limited to a particular stage of infection of a host. These differences are controlled in large part by transcriptional regulation and *cis*-regulatory elements, such as up-stream promoters. The level of transcription from a promoter can reflect cellular demand for the associated protein, and may signify a cell’s physiological state. For example, transcription from the *soxS* promoter in *Escherichia coli* drives expression of the SoxS protein, a transcription factor responsible for

directing expression of dozens of genes involved in protection against damage by free radicals<sup>1</sup>; correspondingly, activation of the *soxS* promoter is used as an indicator of cell oxidative stress.<sup>1</sup> In other cases, transcriptional activity of a particular promoter may be indicative of cell physiology while having no direct role in the regulation of other genes. For example, expression of SspH1 (a secreted virulence factor) by *Salmonella typhimurium* is limited to bacteria residing within mammalian cells<sup>2</sup>; while SspH1 has no known role in the regulation of other *Salmonella* genes, its expression is concurrent with that of other proteins that promote intracellular survival.

## RESULTS AND DISCUSSION

Here we describe a method for “genetic dissection” of the proteome. We elaborate on an existing strategy in which expression of a specific transgene permits incorporation of a “metabolic tag” into newly translated proteins. In this approach, we synchronize tagging with specific genetic activities to direct the tag into proteins that are expressed concomitantly with the activation of a particular promoter (and thus a certain physiological state of the host cell).

We previously described bio-orthogonal non-canonical amino acid tagging (BONCAT), a strategy for selective enrichment and identification of newly synthesized cellular proteins.<sup>3</sup> In procedures similar to those used in isotopic labeling, azide- or alkyne-containing non-canonical amino acids (ncAAs) are introduced to cells during a “pulse” in which actively synthesized proteins are tagged. Tagged proteins are distinguished from those made prior to the pulse through bio-orthogonal ligation of the ncAA side chain to probes that permit their detection, isolation,<sup>4</sup> and visualization.<sup>5, 6</sup> In 2009, we reported a genetically-targeted strategy for confining protein labeling to specified cells within heterogeneous mixtures,<sup>7</sup> by using the methionine (Met, **Fig. 3.1a**) surrogate azidonorleucine (Anl, **Fig. 3.1a**) as the metabolic label. In this

approach, we relied on expression of the L13N/Y260L/H301L mutant form of the *E. coli* methionyl-tRNA synthetase (NLL-MetRS) to enable cells to use Anl in competition with Met during translation.<sup>8</sup> Cells that do not express the mutant enzyme are inert to Anl. In cellular mixtures, only those proteins made in cells that express the mutant synthetase are labeled. Through this approach, proteins synthesized in targeted cells can be selectively isolated from complex mixtures for identification by mass spectrometry or conjugated to fluorescent dyes for *in situ* visualization.

Here we describe methods for “state-selective” labeling of cellular proteins. We placed the gene encoding NLL-MetRS under control of two promoters of interest and compared the patterns of protein synthesis observed in active and inactive transcriptional states. We anticipated that when the promoter is inactive (or repressed), NLL-MetRS would not be expressed and proteins would not be subject to Anl-labeling (**Fig. 3.1b**). However, under inducing conditions where transcription from the promoter is active, NLL-MetRS is expressed and newly synthesized proteins can be tagged with Anl (**Fig. 3.1c**).

To demonstrate this approach, we first used arabinose induction of the  $P_{BAD}$  promoter to drive expression of the mutant synthetase. The combined negative and positive control of transcription from the  $P_{BAD}$  promoter yields an “off/on” genetic switch that permits tightly controlled target gene expression in *E. coli*.<sup>9</sup> We inserted a DNA sequence encoding NLL-MetRS into a plasmid downstream of the  $P_{BAD}$  and *ara* operon regulatory elements. We transformed the resulting construct into *E. coli* cells and compared protein labeling under non-inducing and inducing conditions (**Fig. 3.1d**). Following a 10-minute pulse with Anl, cell lysates were subjected to Cu (I)-catalyzed ligation<sup>10</sup> to an alkyne probe (Alkyne-TAMRA, **Fig. 3.1a**) for selective modification of Anl-tagged proteins. After separation of proteins by SDS-PAGE, in-gel fluorescence scanning revealed proteome-wide incorporation of Anl in arabinose-induced cells. Proteins from cells grown under non-inducing conditions did not exhibit significant labeling. NLL-

MetRS expression, as determined by western blot analysis, was consistent with both  $P_{BAD}$  transcriptional behavior and fluorescence detection of Anl-tagged proteins (**Fig. 3.1e**).

Next, we examined promoter-driven tagging following activation of the oxidative stress response in *E. coli*. We placed NLL-MetRS expression under control of SoxR, an oxidative stress sensor and transcription factor that is constitutively expressed in inactive form.<sup>1,11</sup> SoxR contains two iron-sulfur clusters that are subject to one-electron oxidation or nitrosylation by superoxide and nitric oxide, respectively<sup>12</sup>. SoxR binds the promoter that controls expression of the *soxS* gene; transcription from this promoter is low when SoxR is in its reduced, inactive state. Upon activation of SoxR by superoxide or nitric oxide, strong expression of SoxS is driven from the *soxS* promoter. The SoxS protein, also a transcription factor, is responsible for activating the cellular response to superoxide and nitric oxide, and coordinates expression of more than forty genes involved in detoxification and oxidative-damage repair.

To synchronize labeling with the oxidative stress response, we placed the NLL-MetRS gene under control of the SoxRS regulon by positioning it directly downstream of the *soxS* promoter (see Methods). The resulting construct was introduced to *E. coli* cells via a low-copy number plasmid such that, in addition to control of chromosomally encoded SoxS, the SoxRS regulon was also responsible for directing expression of NLL-MetRS (**Fig. 3.2a**). As described earlier for the  $P_{BAD}$  system, we examined promoter-directed protein tagging under both non-inducing and inducing conditions. We activated the SoxRS regulon in *E. coli* cells grown in liquid culture by adding the superoxide generating agent paraquat (PQ).<sup>12</sup> After a 15-minute pulse with Anl, tagged proteins were modified in lysates by strain-promoted conjugation to a dibenzoazacyclooctyne<sup>13</sup> (DIBAC)-functionalized fluorescent dye (DIBAC-TAMRA, **Fig. 3.1a**) and separated by SDS-PAGE. In-gel fluorescence scanning yielded results analogous to those observed with the

$P_{\text{BAD}}$  system; activation of the *soxS* promoter with PQ (**Fig. 3.2b**) induced proteome-wide tagging of proteins synthesized by *E. coli* in the oxidative stress state. Labeling was also detected in proteins from cells treated with nitric oxide (**Supplementary 3.1**). Proteins from cells cultured under non-inducing conditions exhibited minimal evidence of tagging. NLL-MetRS expression, as revealed by western blot analysis (**Fig. 3.2c**), was consistent with the behavior of the SoxRS regulon and fluorescence analysis of protein tagging. Together with observations from fluorescence microscopy (**Fig. 3.2d**), these data demonstrate that a promoter-driven strategy can be used to direct protein tagging to cells in specific physiological states.

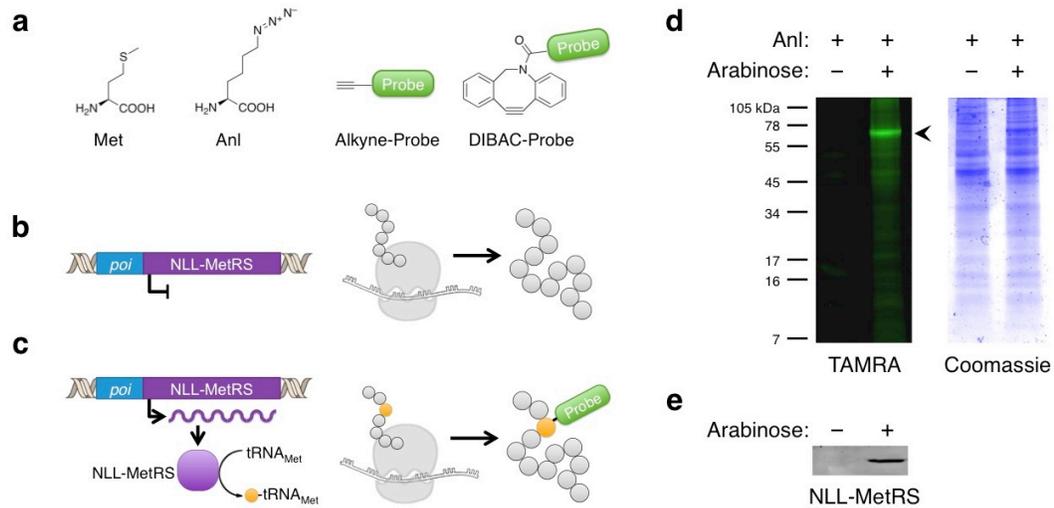
We achieved useful levels of protein-tagging specificity in spite of low levels of “leaky” expression in both the  $P_{\text{BAD}}$  and SoxRS systems. Reduction of background labeling required optimization of the relative amounts of Anl and Met in the culture medium; we screened conditions by adjusting the concentration of Anl until negligible tagging was observed in non-induced cultures. Background signal was also reduced by using glucose (for  $P_{\text{BAD}}$  catabolite repression) or thiamine (an antioxidant that reduces basal levels of SoxR oxidation). Use of low copy-number plasmids was essential for minimizing background labeling; we anticipate that a single copy of the NLL-MetRS gene by chromosomal recombination would also yield satisfactory results.

We estimate that, under the conditions used in this study, the extent of replacement of Met by Anl is less than 10%. This estimate was made by comparing the intensities of TAMRA fluorescence emitted by cells labeled with the SoxRS system to those of cells labeled under conditions previously shown to yield 10% substitution (**Supplementary Fig 3.2**). Similar results were obtained for the arabinose system (data not shown). The rate of substitution was sensitive to the level of NLL-MetRS induction (**Supplementary Fig. 3.3**).

## CONCLUSION

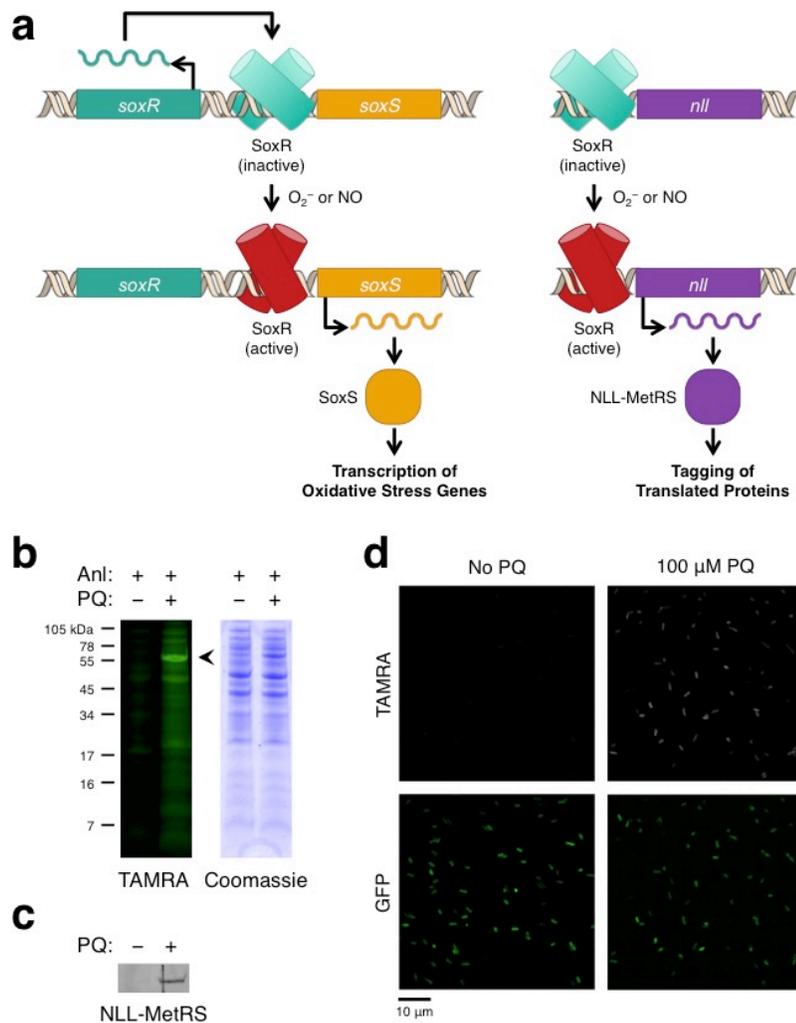
In this work, we demonstrated that controlled expression of NLL-MetRS can be used to restrict protein labeling to cells that are in a particular physiological state. The linking of protein tagging to promoter activity permits “genetic dissection” of the proteome such that proteins synthesized in specific cell states can be tagged and identified. In multi-cellular organisms, this system may be useful for identifying proteins expressed by specific cell types or translated during specified stages of development.

**Figure 3.1**



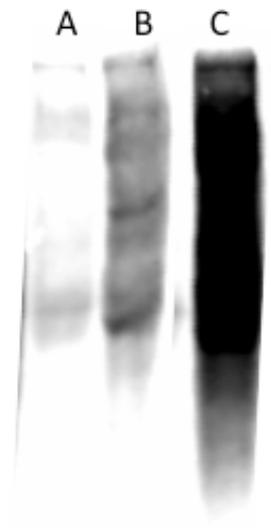
**Promoter-directed proteomic labeling with Anl.** (a) Structures of amino acids and simplified representations of probes used in this study. (b) NLL-MetRS expression is placed under control of a promoter of interest (*poi*). When transcriptional activity of the *poi* is “off,” the NLL-MetRS is not expressed and proteins are not tagged. (c) When transcription of the *poi* is “on,” the NLL-MetRS is expressed and newly synthesized proteins are tagged. (d) When under control of  $P_{BAD}$  promoter and *ara* operon regulatory elements, the NLL-MetRS can be induced by arabinose. Induced (+ arabinose to 0.2% (w/v)) and uninduced (- arabinose) cells were pulsed with 100  $\mu$ M Anl. Incorporation of Anl was assessed by conjugation to alkyne-TAMRA and subsequent detection by in-gel fluorescence scanning (left). Arrowhead indicates labeling of the induced NLL-MetRS. Coomassie staining was used to confirm even loading of protein between lanes (right). (e) Western analysis of NLL-MetRS expression in uninduced (-) and induced (+) cells.

Figure 3.2



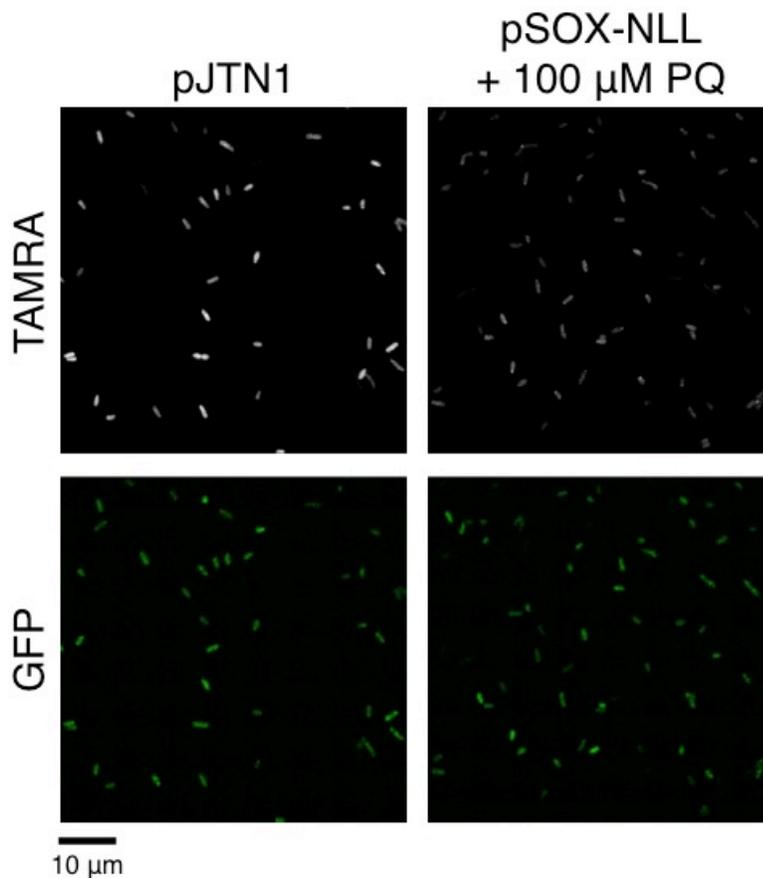
**Proteomic labeling with Anl under conditions of oxidative stress.** (a) The NLL-MetRS is under control of the SoxRS regulon and is activated by superoxide or nitric oxide. (b) Anl-tagging is activated in cells treated with paraquat (PQ), a superoxide generating agent. Cells were treated with PQ or were uninduced and subsequently pulsed with 125  $\mu$ M Anl. Incorporation of Anl was assessed by conjugation to DIBAC-TAMRA and subsequent detection by in-gel fluorescence scanning (left). Arrowhead indicates labeling of the induced NLL-MetRS. Coomassie staining was used to confirm even loading of protein between lanes (right). (c) Western analysis of NLL-MetRS expression in induced (+PQ) and uninduced (-PQ) cells. (d) Fluorescence visualization of tagged proteins in induced (100  $\mu$ M PQ) and uninduced (No PQ) cells. Detection of constitutively expressed GFP confirms presence of cells in both induced and uninduced cultures.

### Supplementary Figure 3.1



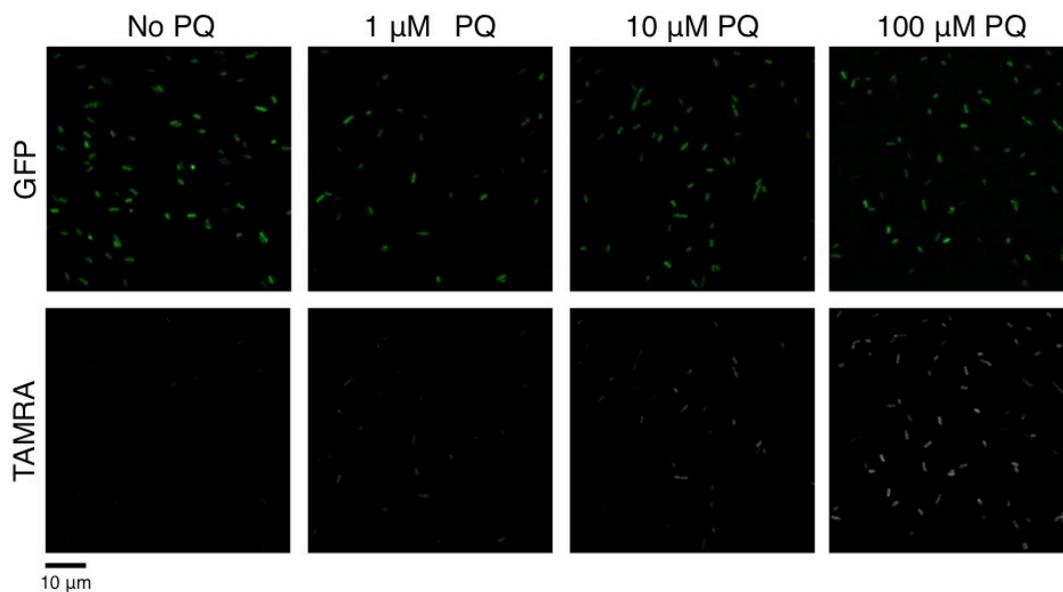
**Proteomic labeling with Anl under conditions of oxidative stress.** The NLL-MetRS is under control of the *soxS* promoter and can be activated by superoxide or nitric oxide. Cells in medium containing Anl were treated with nitric oxide (b), paraquat, a superoxide-generating agent (c), or were uninduced (a). Incorporation of Anl is detected by conjugation to biotin-PEG-alkyne and subsequent detection by western blotting with streptavidin-HRP

Supplementary Figure 3.2



**Tagging rate in the SoxRS system is less than 10%.** *E. coli* cells harboring the pJTN1 plasmid pulsed with AnI under conditions that yield a 10% substitution rate are compared with *E. coli* harboring pSOX-NLL induced with paraquat and pulsed with AnI. Incorporation of AnI is assessed by conjugation with alkyne-TAMRA and subsequent detection by fluorescence microscopy. SoxRS-directed labeling with AnI yields a substitution rate of less than 10%, as fluorescence emission from pJTN1 cells is more intense than that observed with SoxRS-directed labeling. Cells constitutively expressed GFP, which was separately detected to confirm the presence of cells.

### Supplementary Figure 3.3



**Tagging rate in the SoxRS system is dependent on the degree of induction of NLL-MetRS expression.** The NLL-MetRS is under control of the *soxS* promoter and is activated by addition of paraquat (PQ) to the culture medium. As the degree of transcription from the *soxS* promoter is dependent on the concentration of PQ used, so is the level of NLL-MetRS induction. Cells induced with PQ and pulsed with 125  $\mu$ M AnI exhibit increasing levels of AnI incorporation as more PQ is added. Incorporation of AnI is assessed by conjugation to alkyne-TAMRA and subsequent detection by fluorescence microscopy. Cells constitutively expressed GFP, which was separately detected to confirm the presence of cells.

## MATERIALS AND METHODS

### Materials

AnI was synthesized as previously described (Link, A.J., *et al*, *Nat. Protoc.* **2**, 1879—1883) but using N $\alpha$ -boc-Lysine as the starting material instead of N $\alpha$ -diaminobutyric acid. The THPTA ligand was synthesized as previously described<sup>9</sup>. Structures were verified by mass spectrometry and <sup>1</sup>H-NMR spectroscopy.

TAMRA-alkyne and DIBAC-TAMRA were purchased from Click Chemistry Tools. Anti-PentaHis monoclonal antibody AlexaFluor647-conjugate was purchased from Qiagen. Coomassie Colloidal Blue staining kit was purchased from Invitrogen. BCA Protein Assay Kit was used to measure protein concentrations and was from Pierce. pREP4 encoding the bacterial *lacIq* gene was obtained from Qiagen (not to be confused with the commercially available mammalian pREP4). Biotin-PEG-alkyne was purchased from Invitrogen. pACYC was obtained from New England Biolabs free of cost. All other reagents were purchased from Sigma and used without further purification.

### Cloning

All cloned constructs were confirmed by DNA sequencing by Laragen before use.

Cloning of pJTN1 was previously described.<sup>7</sup>

To clone pBAD33-NLL-MetRS, the gene encoding an N-terminally His-tagged NLL-MetRS was amplified by PCR from pQE80-NLL<sup>8</sup> using the following primers: 5'-atatggtaccctatcatttagaggcttccacc-3' and 5'-atatgagctcggccacgaaggccaggagtgaaacgatgagaggatcgcatc-3'. The resulting

fragment was inserted between the *SacI* and *KpnI* sites of pBAD18 using standard cloning procedures.

To clone pREP4-*soxS*-NLL, the *soxS* promoter and ribosomal binding site sequence (as encoded on the DH10B *E. coli* genome) was added to the 5' end of the N-terminally His-tagged NLL-MetRS by PCR using primer overhangs. The fragment was inserted into the *NheI* site of pREP4. While this construct exhibited leaky expression, NLL-MetRS could still be activated using paraquat and nitric oxide (this contrast was used in preparing Supplemental Figure 3.1). Because of leaky expression, this construct was not used in subsequent experiments.

To clone pSOX-NLL-MetRS, the IDT\_Sox\_Regulon sequence (below) was custom synthesized by Integrated DNA Technologies and received from the manufacturer on pSOX\_IDT. The gene encoding the NLL-MetRS was amplified from pQE80-NLL by PCR using primers: 5'-atatggatccatgactcaagtcgcaaga-3' and 5'-ggtggaagcctctaatgaagatct-3'. The resulting fragment was inserted between the *BamHI* and *BglII* sites of pSOX\_IDT to form pSOX\_IDT\_NLL. The cassette encoding the NLL-MetRS under control of the SoxRS regulon was amplified from pSOX\_IDT\_NLL by PCR using the primers: 5'-atatgctagccccgtgtaaacgacggccagt-3' and 5'-atatgctagcagtattgagcctcaggaacagctatgac-3'. The resulting fragment was inserted at the *NheI* site of the low-copy number pACYC177 plasmid to form pSOX-NLL-MetRS. DH10B *E. coli* cells harboring the pSOX-NLL-MetRS plasmid were used to generate the data displayed in Figure 3.2, Supplemental Figure 3.2, and Supplemental Figure 3.3. A schematic of the construct is shown in Figure 3.2a.

IDT\_Sox\_Regulon sequence:

```
5' ATATgctagcCCCGTGTA AACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAG
CTGATCGTGGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAG
TCTGAGGCTCGTCCTGAATGATATGCGACCGCCGGAGGGTTGCGTTTGAGACGGGCG
```

ACAGATCGACACTGCTCGATCCGCTCGCACCCAAAAACCCCTCAAGACCCGTTTAG  
 AGGCCCAAGGGGTTATGCTATCATTATCATTAGTTTTGTTTCATCTTCCAGCAAGCG  
 TGCGCCGTACCTTCTTCTCCTAAGCGGTCGCCGGGTTACGCAACGGGCAATCACT  
 GCGCGAAAGGCAGCCACAACCAATACATCCGTCCAGTTCGTACGCAGCGCCACTAA  
 GGTATGAATGCGCCGATCCAACCTTCTCGCCATTGGGACGAAAGCTGTTTCCACTC  
 TTTTCGCACTTAACGTATGCCCTTCGGGCAACACGCCAAACGCTTCACCAATGGTCGC  
 CAGCGGAATGCCAATACGCTGAGCAATTTTGATAATTGCAACATATCGCAACACATC  
 ACGTTTATATCGCCGCTGATTGCCGCTGTTACGGATACTGGTAATCAACCCTTTACT  
 TTCATAGAAATGCAGCGCCGATACCGCCACACCGCTGCGTTTTCGCCACTTCGCCGGG  
 GGTTAGCAGCGCTTTAATGCGGGGTAATTTCTTTTCCATAAATCGCTTTACCTCAAG  
 TTAACCTTGAGGAATTATACTCCCCAACAGATGAATTAACGAACTGAACACTGAAAAG  
 AGGCAGATTTCCATGGGAGGATCCAGATCTCATCACCATCACCATCACTAAGCTTAA  
 TTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTG  
 GATTTGTTTCAGAACGCTCGGTTGCCGCGGGCGTTTTTTATTGGTGAGAATCGGATC  
 GACGAGAGCAGCGGACTGGATCACTACTGGACC GCGAGCTGTGCTGCGACCCGTGA  
 TCTTACGGCATTATACGTATGATCGGTCCACGATCAGCTAGATTATCTAGTCAGCTT  
 GATGTCATAGCTGTTTCCCTGAGGCTCAATACgctagcATAT-3'

5'-Blue: transcriptional termination sequence (reverse)  
 Green: SoxR transcription factor gene from (reverse)  
 Yellow: Bi-directional regulatory sequence containing the *soxR* and *soxS*  
 promoters  
 3'-Blue: transcriptional termination sequence (forward)

### Protein Tagging with AnI

P<sub>BAD</sub> system:

DH10B *E. coli* cells transformed with pBAD33-NLL-MetRS were grown in LB medium supplemented with 0.2% glucose and 100 mg L<sup>-1</sup> of ampicillin overnight at 37 °C with orbital shaking at 250 rpm. The following day, the culture was diluted 1:50 into M9 minimal medium supplemented with 40 mg L<sup>-1</sup> of each of the twenty canonical amino acids, 1 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 30 mg L<sup>-1</sup> thiamine hydrochloride, and 0.4% (v/v) glycerol. This culture was grown at 37°C with orbital shaking at 250 rpm until OD<sub>600</sub> was ~ 0.5, at which point the culture was divided into two. To one culture, arabinose was added to a final concentration of 0.2% for induction of the NLL-MetRS. Ten minutes later, both cultures were pulsed by addition of AnI to a final concentration of 100 mM. After a 10 m pulse, cells were pelleted by

centrifugation at 4 °C, rinsed once with cold PBS, and pelleted again. Cell pellets were frozen at - 20 °C until further processing.

SoxRS system:

DH10B *E. coli* cells transformed with pSOX-NLL were grown overnight in SOB medium supplemented with 30 mg L<sup>-1</sup> thiamine hydrochloride and 100 mg L<sup>-1</sup> ampicillin at 37 °C with orbital shaking at 250 rpm. The following day, the culture was diluted 1:50 into fresh SOB medium supplemented with 10 mg L<sup>-1</sup> thiamine hydrochloride and 100 mg L<sup>-1</sup> ampicillin and grown at 37°C with orbital shaking at 250 rpm. When the culture reached an OD<sub>600</sub> of ~ 1.0, the culture was diluted 1:10 into RPMI medium supplemented with 10% (v/v) fetal bovine serum and 10 mg L<sup>-1</sup> thiamine hydrochloride. This culture was allowed to grow for 1 h. The culture was divided into two and paraquat was added to one culture at a final concentration of 100 mM (from a freshly prepared 10 mM stock in water). Ten minutes later, both cultures were pulsed by addition of AnI to a final concentration of 125 μM. After a 15 m pulse, cells were pelleted by centrifugation at 4 °C, rinsed once with ice cold PBS and pelleted again. Cell pellets were frozen at - 20 °C until further processing.

For Supplementary Figure 3.1, DH10B cells were transformed with pREP4-*soxS*-NLL and maintained in 35 mg L<sup>-1</sup> kanamycin. Paraquat treatment was done as described for pSOX-NLL, and exposure of cells to nitric oxide was done as described by Ding *et al* (ref. 11).

### **Protein Conjugation and Detection**

For protein conjugation to alkyne-TAMRA, cells were lysed by suspension in PBS containing 1% (w/v) SDS and heated to 75 °C. Lysates were briefly sonicated to shear DNA and reduce the viscosity of the solution. Protein concentrations were normalized by BCA assay. The conjugation reaction was performed according to the protocol suggested by Vu *et al* (ref. 9) using 10

$\mu\text{M}$  alkyne-TAMRA and a 20 m reaction time. Reactions were stopped by addition of Anl to a concentration of 100  $\mu\text{M}$  followed by brief vortexing.

For protein conjugation to DIBAC-TAMRA, cells were lysed by suspension in 100 mM Tris buffer (pH 8.0) containing 1% (w/v) SDS, and heated to 75 °C. Lysates were briefly sonicated to shear DNA and reduce the viscosity of the solution. Prior to addition of DIBAC-TAMRA, cysteine thiols were blocked by addition of iodoacetamide to a final concentration of 25 mM. The blocking reaction was allowed to proceed for 1 h in the dark with mild agitation. Protein conjugation reactions were initiated by addition of DIBAC-TAMRA to a final concentration of 20  $\mu\text{M}$  (from a 5 mM stock in DMSO). Reactions were allowed to proceed for 20 m with mild agitation at room temperature, protected from light. Reactions were stopped by addition of Anl to 100  $\mu\text{M}$  followed by brief vortexing.

For detection, ~ 10  $\mu\text{g}$  of protein was separated on 12% Tris-Tricine SDS-PAGE gels. Following electrophoresis, gels were imaged by fluorescence scanning on a GE Typhoon Trio+ fluorescence scanner with detection for TAMRA. Subsequently, proteins were stained using the Coomassie Colloidal Blue staining kit according to the manufacturer's protocol. Images of Coomassie-stained gels were also obtained on the GE Typhoon Trio+ using red-laser excitation of the gel and omitting the emission filter.

## **Imaging**

Cultures prepared as described earlier were pulse labeled with 100  $\mu\text{M}$  Anl for 20 m. To halt protein synthesis, cell pellets were washed with cold PBS containing 30  $\mu\text{g mL}^{-1}$  chloramphenicol. Cells were fixed in 3.7% (w/v) formaldehyde at room temperature for 30 m, permeabilized in 0.1% (v/v) Triton-X 100 in PBS at room temperature for 5 m, sedimented by centrifugation, and resuspended in PBS. Reagents were added to the cell suspension to the following concentrations: 100  $\mu\text{M}$   $\text{CuSO}_4$ , 500  $\mu\text{M}$  THPTA

ligand, and 40  $\mu$ M alkyne-TAMRA. The reaction was initiated by addition of aminoguanidine and sodium ascorbate (from 20X stocks in water) to final concentrations of 5 mM each. The reaction was allowed to proceed for 30 m at room temperature, protected from light. Cells were washed twice with PBS to remove excess dye and resuspended in PBS. Each cell suspension was applied to a 5% agarose slab and covered with a coverslip. Images were obtained with a Zeiss LSM510 confocal microscope using a 63X oil immersion lens.

### **Western Blot Analysis**

Protein contents in cell lysates were normalized by BCA assay and 5  $\mu$ g of protein from each lysate was separated on 12% Tris-Tricine SDS-PAGE gels. Following electrophoresis, proteins were transferred to nitrocellulose membranes and subsequently blocked with 5% milk in PBST. Membranes were probed for NLL-MetRS by addition of Anti-PentaHis AlexaFluor647-conjugate to the blocking buffer at a dilution of 1:5,000. After 1 h incubation with rocking at room temperature, membranes were washed for 5 m with PBST. The wash was repeated three times. Washed membranes were scanned on a GE Typhoon Trio+ fluorescence scanner with detection for AlexaFluor647.

## REFERENCES

1. Pomposiello, P.J. & Demple, B. *Trends. Biotechnol.* **19**, 109—114 (2001).
2. McGhie, E.J., Brawn, L.C., Hume, P.J., Humphreys, D., Koronakis, V. *Curr. Opin. Microbiol.* **12**, 117—124 (2009).
3. Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A. & Schuman, E.M. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9487—9487 (2006).
4. Szychowski, J., et al. *J. Am. Chem. Soc.* **132**, 18351—18360 (2010).
5. Beatty, K.E., Xie, F., Wang, Q. & Tirrell, D.A. *J. Am. Chem. Soc.* **127**, 14150—14151 (2005).
6. Dieterich, D. C., et al. *Nat. Neurosci.* **13**, 897-905 (2010).
7. Ngo, J. T., et al. *Nat. Chem. Biol.* **5**, 715—717 (2009).
8. Tanrikulu, I.C., Schmitt, E., Mechulam, Y., Goddard, W.A., 3<sup>rd</sup> & Tirrell, D.A. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15285—15290 (2009).
9. Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. *J. Bacteriol.* **177**, 4121—4130 (1995).
10. Hong, V., Presolski, S.I., Ma, C. & Finn, M.G. *Angew. Chem. Int. Ed. Engl.* **48**, 9879—9883 (2009).
11. Ding, H., Hidalgo, E. & Demple, B. *J. Biol. Chem.* **271**, 33173—33175 (1996).
12. Ding, H. & Demple, B. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5146—5150 (2000).
13. Debets, M.F., van Berkel, S.S., Schoffelen, S., Rutjes, F.P.J.T, van Hest, J.C.M. & van Delft, F.L. *Chem. Commun.*, **46**, 97-99 (2010).

## CHAPTER IV

### *An Alternative Interpretation of the Start Codon Permits N-terminal Tagging of Proteins in Mammalian Cells*

#### ABSTRACT

The AUG codon serves two functions: it encodes methionine (Met), and signals the start of protein synthesis. Correspondingly, AUG is decoded by two forms of methionyl-tRNA: the initiator and the elongator. Selective intracellular charging of the initiator tRNA with the non-canonical amino acid azidonorleucine (Anl) permits incorporation of Anl at the N-terminal position of proteins. Because Anl is not charged to the elongator tRNA, it is not introduced at internal positions. Here we show that Anl can be globally introduced at the N-termini of nascent proteins and used as a metabolic “tag” to track the up- and down-regulation of protein translation in mammalian cells.

#### INTRODUCTION

While codons are degenerate, they are rarely ambiguous. With few exceptions, each codon designates a single amino acid and is decoded as such with high fidelity. However, one codon serves two functions: AUG encodes methionine (Met) and also signals the start of translation of the RNA message. Correspondingly, two forms of tRNA<sup>Met</sup> are used to decode AUG codons: initiator tRNA and elongator tRNA<sup>Met</sup>. The initiator decodes the start codon and incorporates Met at the N-terminal positions of proteins.<sup>1,2,3</sup> The elongator form is used in decoding internal AUG codons and delivers Met to internal positions within polypeptide chains. As start codons and internal

AUG codons are interpreted through distinct mechanisms, it is possible that they could encode separate amino acids. For instance, if initiator tRNA were aminoacylated with a non-canonical amino acid and elongator tRNA<sup>Met</sup> with Met, then start codons and internal AUG codons would be interpreted accordingly. This kind of differential decoding can be achieved in cell-free translation systems by adding chemically misacylated initiator tRNA to direct N-terminal incorporation of a non-canonical amino acid.<sup>4,5</sup> Unfortunately, cell-free systems produce limited amounts of protein and do not allow direct interrogation of cellular processes. Extension of the N-terminal incorporation strategy to cellular systems would be an important advance. Here we demonstrate that heterologous expression of a mutant methionyl-tRNA synthetase (MetRS) from *Escherichia coli* in the mammalian cytoplasm facilitates selective intracellular aminoacylation of initiator tRNA with the non-canonical amino acid azidonorleucine (Anl, **Supplementary Fig. 4.1**). This strategy permits site-selective incorporation of Anl at the N-terminal position of nascent cellular proteins.

In early studies of tRNA recognition, researchers performed inter-species aminoacylations to uncover the molecular determinants governing tRNA specificity.<sup>6</sup> In the methionine system, studies showed that the *E. coli* MetRS efficiently aminoacylates initiator tRNAs from different organisms (bacteria, yeast, plants, and animals) and various eukaryotic organelles (mitochondria and chloroplasts).<sup>7,8,9,10</sup> In contrast, eukaryotic cytoplasmic elongator tRNA<sup>Met</sup> from yeasts, plants, and mammals were poor substrates for the *E. coli* MetRS.<sup>7,8,10,11</sup> A particularly poor substrate was the mammalian cytoplasmic elongator tRNA<sup>Met</sup>, which has been reported not to bind *E. coli* MetRS.<sup>9</sup> A mutational analysis of tRNA<sup>Met</sup> identified the enlarged anticodon loop of the mammalian cytoplasmic elongator as the structural element prohibiting recognition by the *E. coli* MetRS.<sup>12</sup> Given the selectivity of the *E. coli* MetRS for the mammalian initiator, we imagined that it might enable selective aminoacylation of initiator-tRNA in the mammalian cytoplasm. To explore

this idea, we used the L13N/Y260L/H301L triple mutant of the *E. coli* MetRS (NLL-MetRS) to activate Anl in human embryonic kidney (HEK 293) cells. We found Anl to be incorporated at the N-terminal positions of proteins and not be excluded from internal Met positions (**Fig. 4.1A**).

## RESULTS AND DISCUSSION

### Function of NLL-MetRS in mammalian cells

We had previously identified the NLL-MetRS through structure-guided mutagenesis of the *E. coli* MetRS amino acid binding pocket.<sup>13,14,15</sup> When expressed in bacterial cells, the mutant synthetase permits incorporation of Anl into protein at both N-terminal and internal positions encoding Met.<sup>13,16,17</sup> To confirm that the mutant bacterial synthetase was also selective for mammalian initiator tRNA (like the wild-type enzyme), we aminoacylated synthetically prepared tRNA<sup>Met</sup> following previously described procedures.<sup>18</sup> Our analysis revealed that, despite mutation, the NLL-MetRS exhibited tRNA selectivity similar to that of the wild-type enzyme (**Supplementary Fig. 4.2**).

Next, we asked whether or not the NLL-MetRS was functional when expressed in the cytoplasm of mammalian cells. RajBhandary and coworkers observed that wild-type *E. coli* MetRS was active when expressed in monkey cells and that its expression increased translational initiation of a reporter protein.<sup>19</sup> In order to test intracellular aminoacylation with Anl, we generated by transfection a human embryonic kidney (HEK 293)-derived cell line that constitutively expressed NLL-MetRS as a fusion with mCherry (**Supplementary Fig. 4.3A-B**). Following selection for stably transfected cells, we isolated a single clone to generate the HEK293-8D3 cell line (**Supplementary Fig. 4.3C-D**). Expression of the NLL-MetRS-mCherry fusion was confirmed by fluorescence microscopy, western analysis, and affinity

purification of the protein from HEK293-8D3 cells. The enzyme was soluble, fully intact, and broadly distributed within the HEK293-8D3 cytoplasm.

We assessed the activity of the NLL-MetRS in HEK293-8D3 by seeking evidence of Anl incorporation into cellular proteins. Although previous studies have shown that misacylated initiator tRNAs misacylated can be inactive in initiation because of poor binding to eIF2,<sup>20</sup> we were optimistic regarding Anl because efficient initiation with Anl has been observed in bacterial cells.<sup>16,17</sup> In a preliminary analysis, we incubated HEK293-8D3 and control cell cultures with Anl in a 4 h “pulse.” Subsequently, cells were washed, fixed, permeabilized, and examined for Anl incorporation by Cu (I)-catalyzed azide-alkyne ligation<sup>21</sup> of AlexaFluor488-alkyne (**Supplementary Fig. 4.1**). Imaging of cells by fluorescence microscopy revealed that AlexaFluor488-alkyne was attached throughout HEK293-8D3 cells and absent from non-transfected cells (**Fig. 4.1B**). To confirm that Anl was retained within cells in the form of protein, HEK293-8D3 cells were co-incubated with Anl and the protein synthesis inhibitor anisomycin (**Supplementary Fig. 4.4**). Following conjugation of AlexaFluor488-alkyne, we observed markedly decreased fluorescence emission from anisomycin-treated cells, indicating that active protein synthesis is required for cellular retention of Anl.

To verify protein labeling more directly, Anl-tagged proteins in cell lysates were modified via strain-promoted cycloaddition<sup>22</sup> with a dibenzo-azacyclooctyne-functionalized<sup>23</sup> fluorescent dye (DIBAC-TAMRA, **Supplementary Fig. 4.1**). Prior alkylation of protein thiols with iodoacetamide eliminated the background reaction of the DIBAC-group with cysteine residues (**Supplementary Fig. 4.5**). Modified proteins were separated by electrophoresis and examined by in-gel fluorescence scanning. Proteins derived from HEK293-8D3 cells showed strong fluorescence emission (**Fig. 4.1C**), indicating incorporation of Anl throughout the

proteome. In contrast, proteins from non-transfected cells showed little evidence of labeling.

### **N-terminal Tagging**

In 2006, we described bio-orthogonal non-canonical amino acid tagging (BONCAT), a strategy for enrichment and identification of newly synthesized proteins.<sup>24</sup> In the BONCAT method, cells are pulsed-labeled with a non-canonical amino acid tag, and tagged proteins are separated from those made prior to the pulse by selective modification of the non-canonical side chain.<sup>24,25,26,27</sup> Use of Anl/NLL-MetRS system to achieve BONCAT labeling in mammalian cells appeared particularly attractive as we anticipated that labeling would be restructured to N-terminal positions. Since the N-termini of proteins tend to be solvent exposed and flexible,<sup>28</sup> introduction of Anl at these positions would be expected to cause minimal perturbation of protein function.

Pulse labeling of cells with Anl is achieved by simple addition of the amino acid to the growth medium (fresh medium contains at least 115  $\mu$ M Met). We tested Anl concentrations over a 128-fold range and observed incorporation at concentrations as low as 20  $\mu$ M (**Supplementary Fig. 4.6A**). The extent of labeling could be easily adjusted by varying the amount of Anl used in the pulse. The band profiles of tagged proteins were independent of Anl, suggesting that Anl does not cause marked changes in protein expression. Anl was rapidly taken up by cells and incorporated into proteins; Anl=tagging was detected following a 15 m pulse (**Supplementary Fig. 4.6B**). Since we observed increasing accumulation of tagged proteins with increasing pulse length, we inferred that Anl-tagging did not cause rapid protein degradation.

### **Enrichment of Tagged Proteins**

We used a DIBAC-functionalized biotin reagent (DIBAC-biotin, **Supplementary Fig. 4.1**) to modify and enrich tagged proteins from the lysates of Anl-treated HEK293-8D3 cells. To assess the selectivity of the enrichment procedure, a “mock enrichment” of proteins from unlabeled cells was processed in parallel. Each step of the enrichment process was followed by coomassie staining of protein gels and streptavidin detection of biotinylated proteins (**Fig. 4.2A-B**). We did not detect coomassie-stained proteins in the elution of the mock-enrichment (**Fig. 4.2A**, lane “E”), suggesting that proteins enriched from HEK293-8D3 cells (**Fig. 4.2B**, lane “E”) contained very few (if any) tagged proteins. In order to confirm that the enriched fraction did indeed contain proteins synthesized during the pulse, we employed specific immunodetection of two constitutively expressed “housekeeping” proteins, GAPDH and  $\beta$ -actin. While GAPDH and  $\beta$ -actin were absent from the elution of the mock enrichment, both were present in the enriched fraction of HEK293-8D3 proteins (**Supplementary Fig. 4.7**). We estimated the extent of enrichment by adding purified (singly-tagged) GFP to untagged lysate. Following enrichment, the amount of untagged protein in the elution was diluted at least 120-fold and enriched GFP was concentrated by a more than 3-fold (**Supplementary Fig. 4.8**).

### **Analysis of Tagged Proteins by Mass Spectrometry**

In order to confirm that Anl was introduced at the initiator position, we pursued direct evidence of its incorporation at protein N-termini. SDS-PAGE gels containing enriched proteins were divided into individual gel slices and treated with protease. Digested peptides were extracted before analysis by tandem mass spectrometry (LC-MS/MS). The resulting fragmentation data were used to perform a peptide-identification search in which the masses corresponding to DIBAC-biotin modified and unmodified Anl were included as acceptable substitutions for Met (**Supplementary Fig. 4.9**). Although we did not expect to find proteins containing unmodified Anl residues (as only streptavidin-enriched proteins were digested), it was included in the search

nonetheless. Also, since many proteins are subject to enzymatic N $\alpha$ -acetylation,<sup>29</sup> the N $\alpha$ -acetyl derivative of each modification was also included in the search (**Supplementary Fig. 4.9**). In an initial analysis, we identified more than 3,000 unique peptides of which only one exhibited evidence of Anl-incorporation. The N-terminal tryptic fragment of the polypyrimidine tract binding protein (PTBP) was detected containing N $\alpha$ -acetylated Anl modified with DIBAC-biotin (**Fig. 4.2C, Table 4.1**). Western analysis confirmed that PTBP was indeed present in the enriched fraction (**Fig. 4.2D**), and we inferred that detection of this fragment was favorable in part because of PTBP's high relative abundance in the analyzed gel slice (**Fig. 4.2D**, and red box in **Fig 4.2B**). We were confident in the authenticity of the observed spectrum (**Fig. 4.2D**) for the following reasons: 1) the mass of parent ion matched exactly with the expected mass of the substituted peptide, 2) the substitution was directly detected following fragmentation and appeared as both b- and y-ions (**Fig. 4.2D**, ions b1 and y13), and 3) the ion of highest intensity (y9) indicated peptide fragmentation N-terminal to proline, a cleavage that occurs with high propensity when preceded by valine.<sup>30</sup> In total, we detected 15 unique PTBP peptides amounting to 43% sequence coverage of intact PTBP (**Supplementary Fig. 4.10**). Given the nature of selective N-terminal incorporation, expected that internal PTBP peptides would not contain substitutions. Thus, we were pleased to find that of the 14 peptides corresponding to internal PTBP fragments, none displayed evidence of Anl-incorporation (four contained one or more Met residues).

### Detection of additional modifications

In an attempt to encourage detection of modified peptides, we enriched proteins using a cleavable probe in which a disulfide is positioned between the DIBAC group and the biotin handle (DIBAC-S-S-biotin, **Supplemental Fig. 4.1**). Following enrichment, reduction of the disulfide bond releases a substantial portion of the probe from proteins and the resulting structure is considerably smaller than that obtained with the original probe (232 amu

less). We performed a similar analysis as described in the preceding section using DIBAC-S-S-biotin instead of DIBAC-biotin. A database search for modifications corresponding to reduced and alkylated DIBAC-S-S-biotin (**Supplemental Fig. 4.9**) identified six additional peptides exhibiting evidence of Anl incorporation (**Supplemental Fig. 4.11A-G, Table 4.1**). Each peptide was the N-terminal tryptic fragment of a protein, and equivalent peptides containing N-terminal Met were not detected. For each protein, we identified 9 or more unique internal peptides, with sequence coverage ranging between 30-67%. While N-terminal peptides containing Anl were detected, no N-terminal peptides containing Met were observed.

### Tagging of Histones

In our early imaging experiments, we noticed that a fraction of cells exhibited increased fluorescence localized in the center of the cell. In dye-labeled HEK293-8D3 cells, these intensities co-localized with fluorescence emission from Hoechst 33342-stained nuclei (**Fig. 4.3A**), suggesting that certain nuclei contained higher amounts of tagged protein. Analysis of subcellular fractions revealed a set of highly fluorescent bands within the 10-25 kDa mass range in the fraction containing nuclear proteins (**Supplementary Fig. 4.12**). We verified that these proteins were histones by acid-extraction of histones from nuclei and DIBAC-TAMRA labeling of the isolated proteins (**Fig. 4.3B**). TAMRA fluorescence was detected in bands corresponding to the core-histone proteins, as well as the linker histone (data not shown). Since histone synthesis is tightly synchronized with DNA synthesis, histones are translated only by cells in S phase of the cell cycle.<sup>31,32</sup> Correspondingly, we observed that only a fraction of cells within imaged populations exhibited increased nuclear fluorescence. Also, our findings demonstrate that Anl-tagged histones are properly transported into the nucleus following synthesis (histone mRNAs are translated in the cytoplasm).

Since excess amounts of histones can lead to genomic instability, histone synthesis must be tightly controlled to match the amount required for chromatin assembly. Through transcriptional, post-transcriptional, and translational mechanisms, histone is synchronized with DNA replication. To further explore histone tagging, we treated cells with the DNA synthesis inhibitor hydroxyurea (HU). Since histone translation is tightly coupled to DNA synthesis, HU is also an inhibitor of histone translation. We pulse labeled cells with Anl in the presence and absence of HU and analyzed proteins by in-gel fluorescence. A side-by-side comparison of the band intensities showed a significant decrease in fluorescence at bands corresponding to histones, indicating that histone synthesis was decreased by HU treatment (**Fig. 4.4A-B**).

Upon initial inspection of protein gels, HU appeared to selectively down-regulate histone translation while having little or no effect on the synthesis of other proteins (**Fig. 4.4A-B**). However, after close inspection of gels, additional variation in band intensity between lanes could be identified. In order to accentuate differences within individual mass ranges, we separated proteins using gradient and non-gradient gels of various acrylamide concentrations. Using a 15% polyacrylamide gel, we detected an unambiguous difference at ~50 kDa (**Fig. 4.4C**). Because the band was absent from the lane containing proteins from untreated cells, and present in the lane containing proteins from HU-treated cells, we wondered whether it represented stabilized p53 tumor suppressor. The p53 tumor suppressor is a transcription factor that is continually synthesized in cells and is targeted for rapid degradation in its non-phosphorylated state.<sup>33</sup> Under conditions of genotoxic stress (such as during HU treatment), p53 is stabilized by phosphorylation, permitting cellular accumulation of the protein.<sup>33</sup> Detection of p53 expression by western blotting confirmed its stabilization in HU-treated cells (**Fig. 4C**). Additional band differences between lanes were made

apparent after tagged proteins were enriched from cell lysates and detected by streptavidin (**Supplementary Fig. 4.13**).

Lastly, we pulsed HEK293-8D3 in the presence or absence of HU and imaged the cellular distribution of Anl-tagged protein by fluorescence microscopy (**Fig. 4.4D**). In a population of HEK293-8D3 cells not treated with HU, we observed cells exhibiting increased nuclear fluorescence that is associated with histone synthesis. However, increased nuclear fluorescence was not seen when cells were treated with HU. Instead, fluorescence was broadly distributed throughout the cell with nuclear boundaries indiscernible. Individual punctate intensities were visible in HU-treated cells, as well as untreated cells that were not in S-phase (**Fig. 4.4D**). These intensities may represent localized synthesis of proteins at the rough endoplasmic reticulum, or membrane-bound N-terminal signal sequences of proteins that have entered the secretory pathway.

### **Post-translational Removal of N-terminal Anl**

While all proteins are initiated with Met, N-terminal Met is post-translationally removed from many proteins by methionyl-aminopeptidase (MetAP). Recognition by MetAP is largely dependent on the identity of the second amino acid in the polypeptide chain; proteins containing glycine, alanine, or serine at the second position are most efficiently processed.<sup>34</sup> In cases where the N-terminal Met of a substrate protein is replaced by a structurally analogous non-canonical amino acid, processing can still occur (albeit at lower rates).<sup>35,36</sup> In an enzymatic assay using recombinant human MetAP-2 (the cytoplasmic form), we measured excision activity toward N-terminal Anl using a model substrate. While Anl was removed by MetAP, processing was notably less efficient than it was for Met (**Supplemental Fig. 4.14**). Of the proteins for which we had direct evidence of Anl incorporation (by mass spectrometry, or by confirming the identity of a TAMRA-labeled protein band), many were known substrates of MetAP (**Supplementary**

**Table 4.2).**<sup>37</sup> It is likely that N-terminal Anl is also removed intracellularly, although not entirely. An illustrative example of Anl retention can be seen in **Supplemental Figure 4.7** with the enrichment of GAPDH and  $\beta$ -actin. Despite experiencing complete removal of their N-terminal Met residues under natural conditions, N-terminal Anl is retained on GAPDH and  $\beta$ -actin at some fraction. Additionally, the high abundance of PTBP in enriched fractions (red box in **Fig. 4.2B**, lane “E” in **Fig. 4.2C**) may be in part because PTBP is not a MetAP substrate (and thus completely retains Anl). Furthermore, we found that use MetAP inhibitors could reduce the level of intracellular N-terminal excision of Anl (data not shown).

## CONCLUSIONS

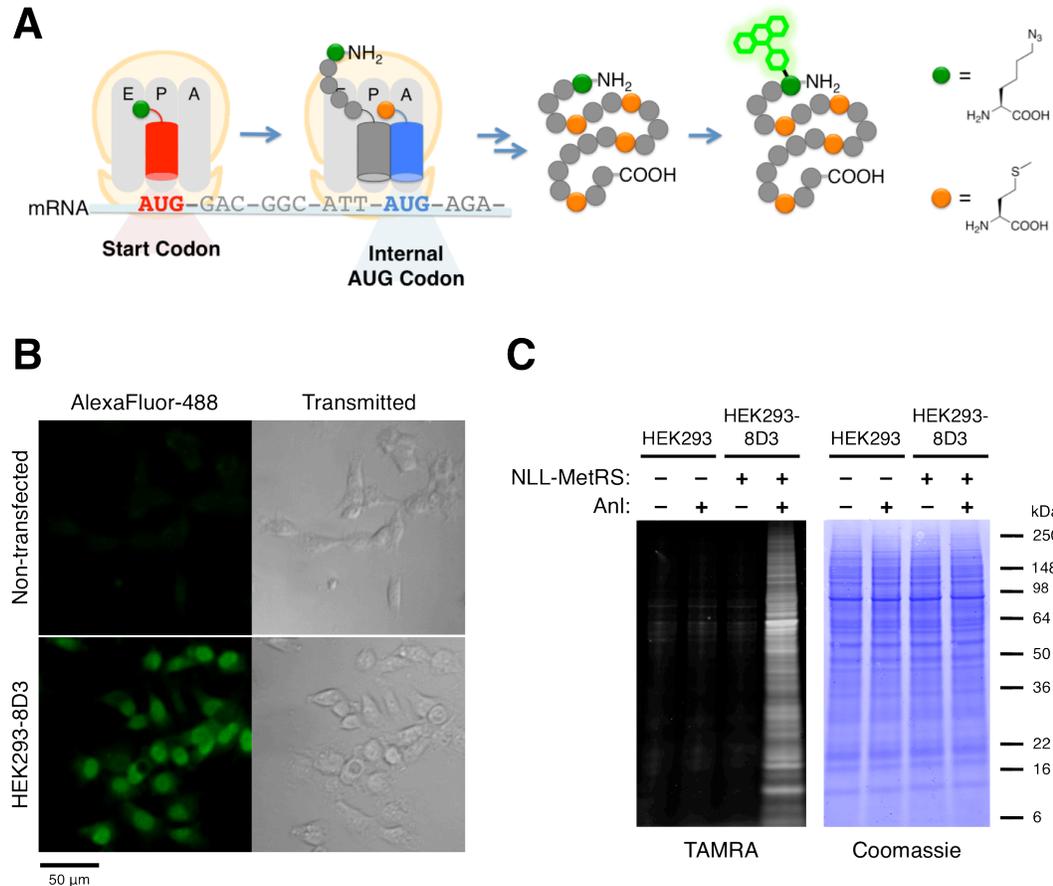
We have demonstrated that intracellular aminoacylation of mammalian cytoplasmic initiator tRNA by the *E. coli* NLL-MetRS can direct site-selective incorporation of Anl at the N-terminal positions of nascent proteins. Discrimination against elongator-tRNA<sup>Met</sup> by the bacterial enzyme prohibits incorporation of Anl at internal sites encoded by AUG, and only Met is inserted at these positions. Proteins synthesized by the cell can be tagged in a “pulse” and subsequently revealed through bio-orthogonal ligations to the Anl side chain. Conjugation of affinity reagents to tagged proteins permits their selective enrichment, and tagged protein present at less than 0.1% of total protein can be isolated from cell lysates. Enrichment of tagged proteins is advantageous for two primary reasons: 1) it permits direct analysis of proteins synthesized in response to internal and external stimuli, and 2) it decreases sample complexity and facilitates identification of proteins by mass spectrometry. Alternatively, conjugation of tagged proteins to detection reagents permits examination of translational profiles on SDS-PAGE gels. Differences in protein synthesis between conditions can be identified through side-by-side comparison of labeled protein bands. In this

work, we detected proteins that were subject to translational up-regulation, down-regulation, or stabilization following treatment with HU.

In order to visualize newly synthesized proteins, fluorescent dyes can be attached to Anl-tagged proteins in fixed cells. We examined global protein synthesis with temporal and spatial resolution and found that translation during S phase is primarily devoted to histone production. Histone synthesis and nuclear translocation were detected as increased fluorescence intensity within cell nuclei, and these intensities co-localized with nuclear staining, resembled the distribution of GFP-tagged histones, and were abolished by inhibition of histone translation. These observations were consistent with the amount of histones required for new chromatin assembly, as well as the cell-cycle dependence of histone translation. In cells that were not in S phase, the distribution of newly synthesized proteins was broadly distributed with individual localized puncta that may represent proteins within the endoplasmic reticulum.

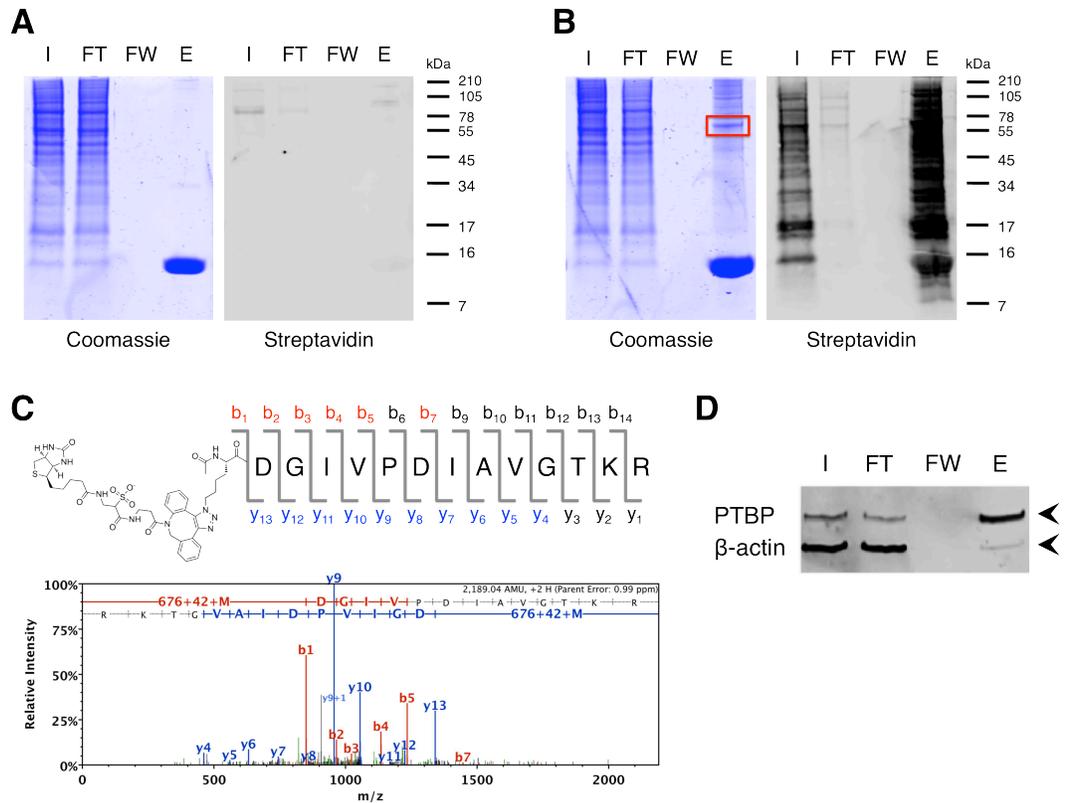
N-terminal incorporation may be extended for interrogating protein synthesis in a cell-selective manner within cellular mixtures, heterogeneous tissue, or even multicellular organisms. This technique complements existing strategies for translational analysis of individual proteins,<sup>38</sup> and combination of these methods may be especially powerful. Since non-canonical amino acids have many additional applications in protein engineering and molecular medicine,<sup>39,40</sup> site-selective N-terminal incorporation may serve as a useful technology for producing modified protein therapeutics. N-terminal incorporation via selective intracellular aminoacylation of initiator tRNA is a novel strategy for expression of proteins containing non-canonical amino acids, and we anticipate future developments may extend this technology to introduce non-natural amino acids beyond Anl.

Figure 4.1



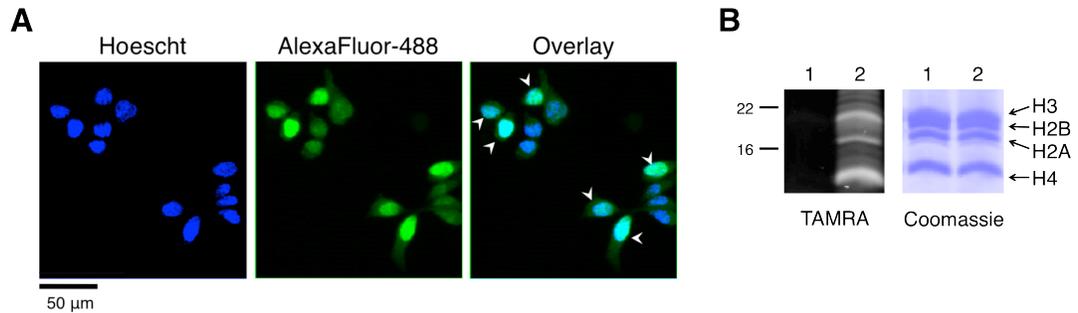
**NLL-MetRS permits incorporation of Anl into proteins synthesized in the mammalian cytoplasm.** (A) Initiator tRNA (red cylinder) aminoacylated with Anl (green circle) is used to initiate protein synthesis from the start codon. Elongator-tRNA<sup>Met</sup> (blue cylinder) aminoacylated with Met (orange circle) is used for decoding internal AUG codons. N-terminally tagged proteins can be ligated to specific probes that permit their detection. (B) HEK293-8D3 and non-transfected cells were pulsed with Anl and incorporation of Anl is assessed by conjugation to AlexaFluor488-alkyne followed by fluorescence microscopy. HEK293-8D3 cells expressing the NLL-MetRS are able to incorporate Anl into proteins whereas non-transfected cells are not. (C) Proteins from pulsed HEK293-8D3 and control cells are examined for Anl incorporation by conjugation to DIBAC-TAMRA followed by in-gel fluorescence scanning. Numerous proteins from pulsed HEK293-8D3 cells are tagged with Anl. Proteins from non-transfected cells or cells not exposed to Anl did not exhibit TAMRA fluorescence.

**Figure 4.2**



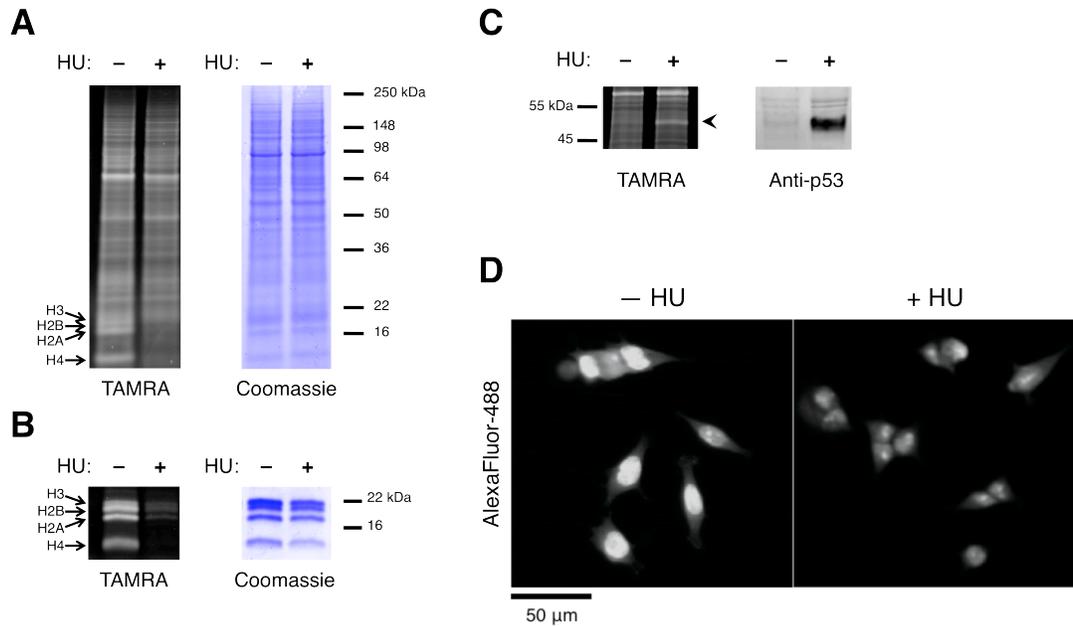
**Enrichment and analysis of Anl-tagged proteins.** Lysates were reacted with DIBAC-biotin (input, “I”), and biotinylated proteins were collected on immobilized streptavidin. The unbound fraction (flow-through, “FT”), and final wash “FW” were collected and compared to the elution “E.” The enrichment was followed by coomassie staining and by detection of biotinylated proteins using fluorescent streptavidin. (A) Proteins from non-transfected cells pulsed with Anl were analyzed in a “mock enrichment” to assess the selectivity of the procedure. Untagged proteins were neither conjugation to DIBAC-biotin nor enriched. Consequently, no proteins were present in the elution except naturally biotinylated proteins and streptavidin. (B) Tagged proteins present in a lysate of HEK293-8D3 cells were conjugated to DIBAC-biotin and could be enriched. The red box indicates the band corresponding to PTBP. (C) Ion fragmentation spectrum of the N-terminal tryptic peptide of PTBP displaying mass shifts corresponding to Anl substitution and conjugation to DIBAC-biotin (+676), as well as N $\alpha$ -acetylation (+42). (D) The enrichment is followed by western detection of PTBP and  $\beta$ -actin. Both PTBP and  $\beta$ -actin are present in the enriched fraction.

Figure 4.3



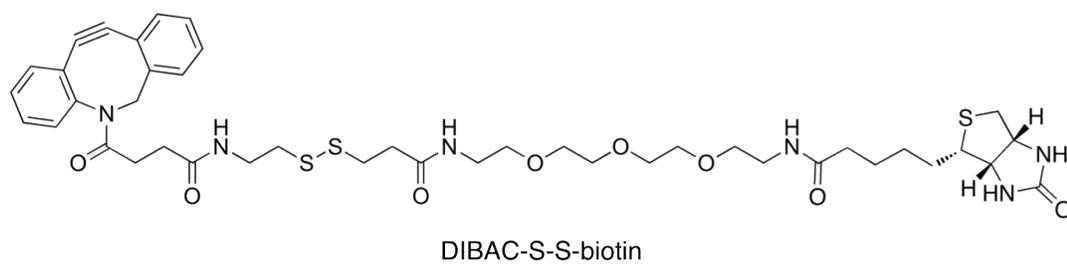
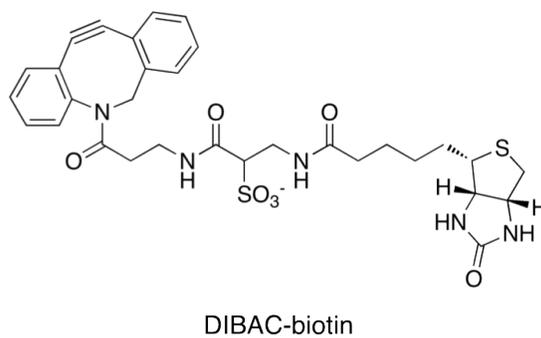
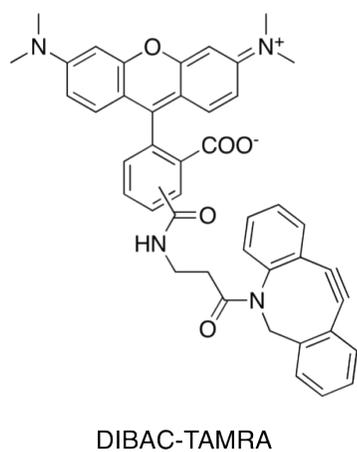
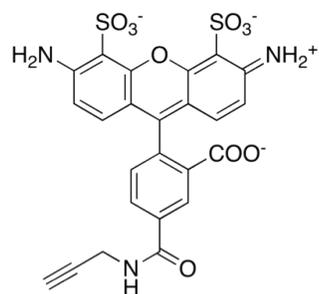
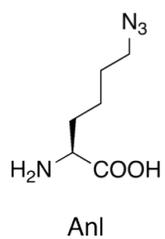
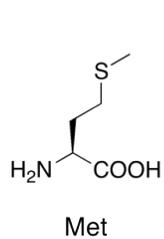
**Histones synthesized by cells in S-phase are tagged with Anl.** (A) HEK293-8D3 cells were dye-labeled with AlexaFluor488-alkyne to reveal the distribution of tagged proteins synthesized during the pulse. Localized increased fluorescence intensity from AlexaFluor488 was observed in a subset of cells. These intensities co-localized with the emission of the nuclear stain Hoechst 33342, indicating that certain nuclei contained increased amounts of tagged protein. (B) Extracted histones were assessed for Anl-incorporation by conjugation to DIBAC-TAMRA followed by detection by in-gel fluorescence scanning. Lane 1 contains histones from HEK293-8D3 cells that were not treated with Anl. Lane 2 contains histones from Anl-treated HEK293-8D3 cells. Bands corresponding to the core histone proteins are labeled, indicating that histones synthesized by cells in S-phase are tagged during the pulse.

Figure 4.4



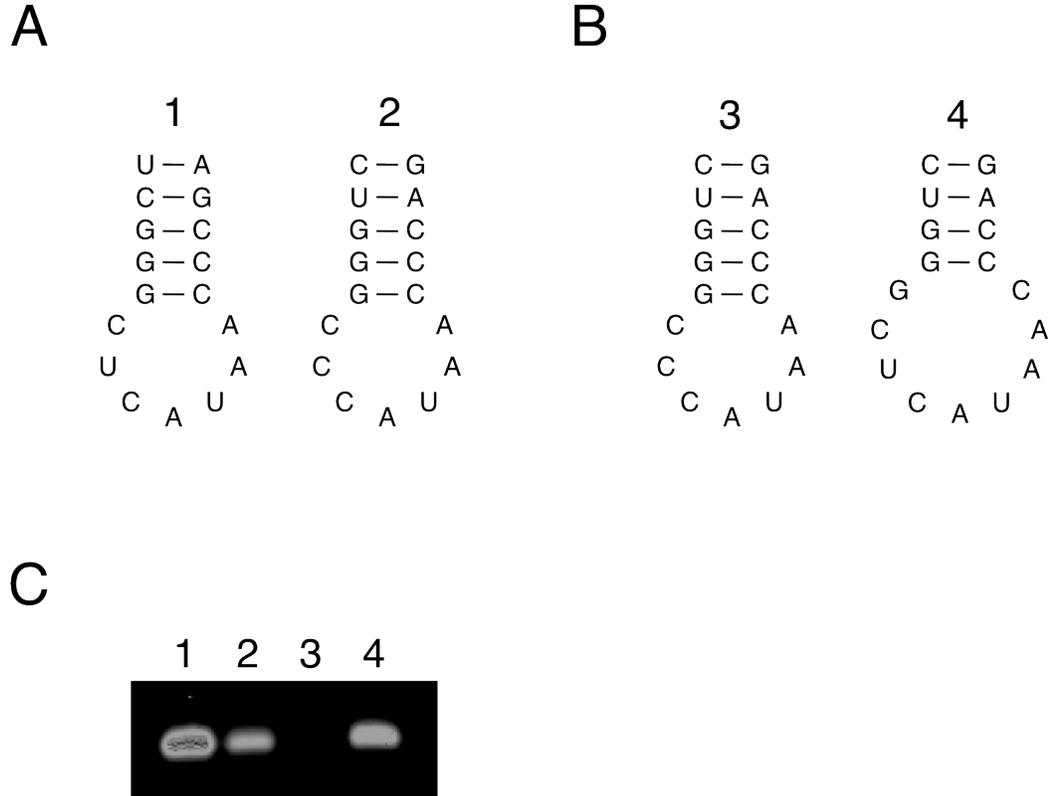
**Protein Synthesis in hydroxyurea-treated cells.** Proteins from untreated and hydroxyurea (HU)-treated cells are assessed for AnI-tagging by conjugation to DIBAC-TAMRA followed by fluorescence scanning. HU is an inhibitor of DNA synthesis and histone translation. Total protein (A), and isolated histones (B) of HU-treated cells exhibit decreased fluorescence intensity at bands corresponding to core histone proteins. (C) Close inspection of TAMRA-labeled proteins reveals stabilization of p53, which is confirmed by western detection with anti-p53 antibody. (D) Cells in S-phase exhibit increased nuclear AlexaFluor488-alkyne fluorescence from tagged histones translated in synchrony with DNA replication. When DNA synthesis is inhibited (+HU), histone translation is rapidly inhibited and nuclear intensities are not detected.

## Supplementary Figure 4.1



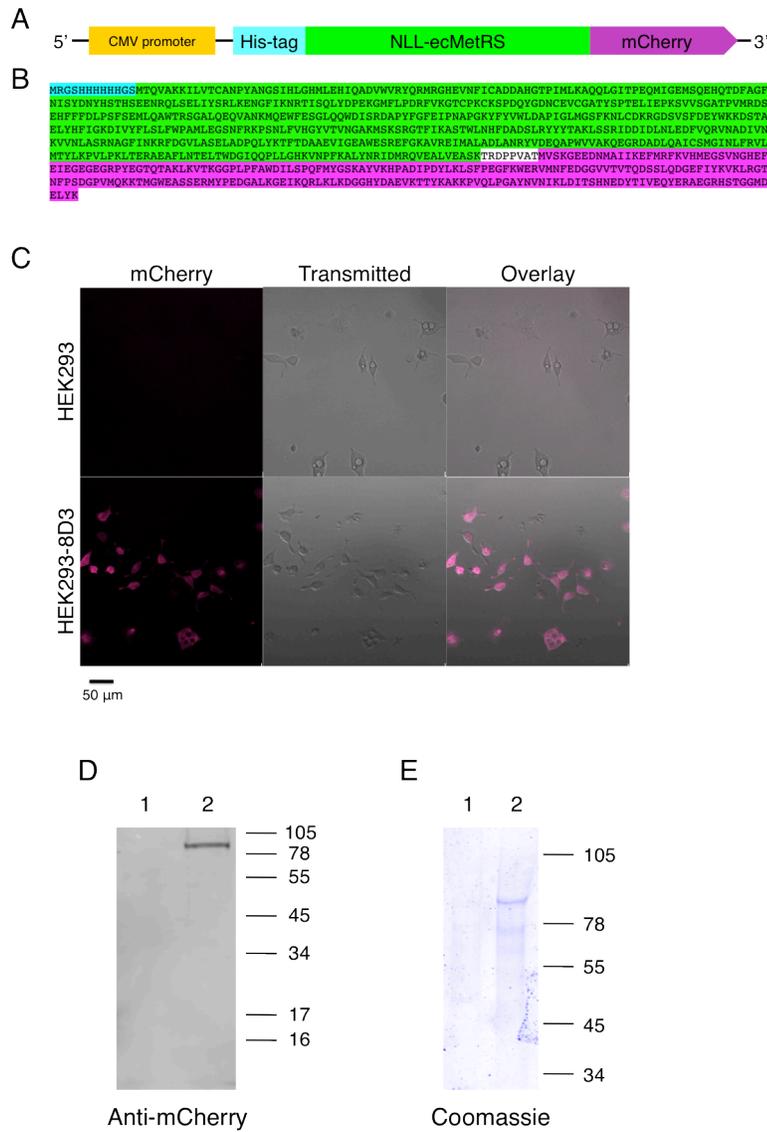
**Amino acids and probes used in this study.**

Supplementary Figure 4.2



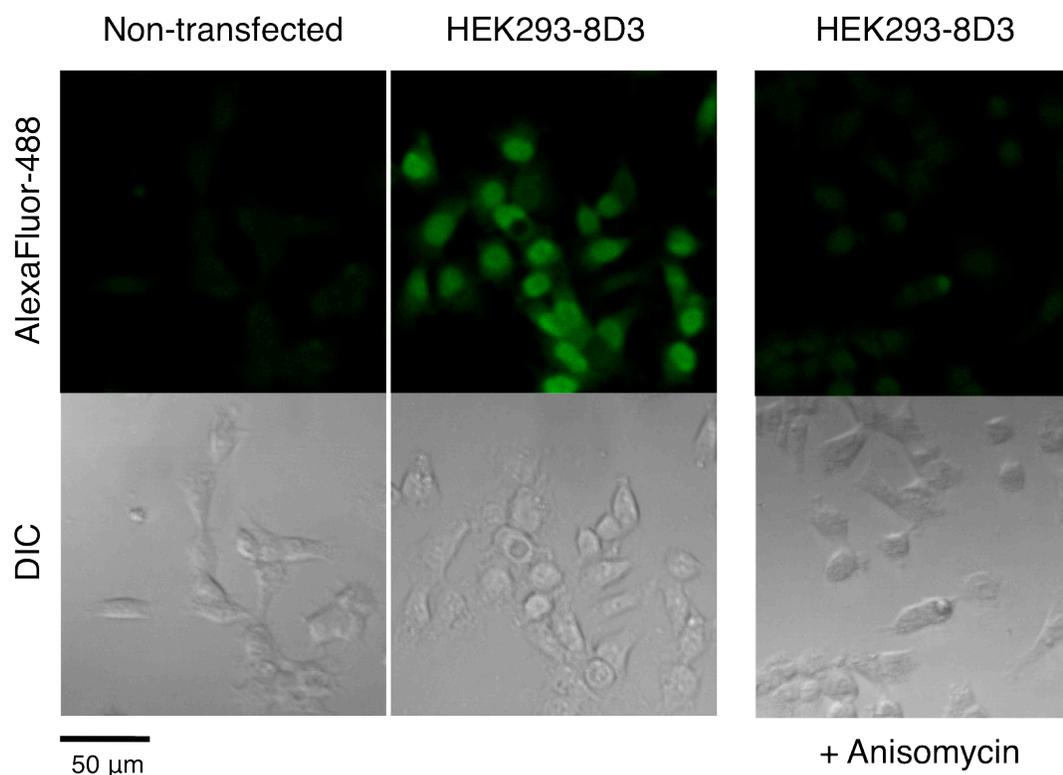
**Selective aminoacylation of eukaryotic cytoplasmic initiator tRNA *in vitro*.** Anticodon stem-loop structures of (A) *E. coli* initiator ("1"), and elongator-tRNA<sup>Met</sup> ("2"), and (B) eukaryotic cytoplasmic initiator ("3"), and elongator-tRNA<sup>Met</sup> ("4"). (C) Synthetically prepared tRNAs are aminoacylated using purified NLL-MetRS and Anl attachment is assessed by conjugation to a cyclooctyne-functionalized AlexaFluor488 and subsequently detected by in-gel fluorescence scanning. The NLL-MetRS, like the wild-type *E. coli* MetRS, does not charge eukaryotic cytoplasmic elongator-tRNA<sup>Met</sup>.

### Supplementary Figure 4.3



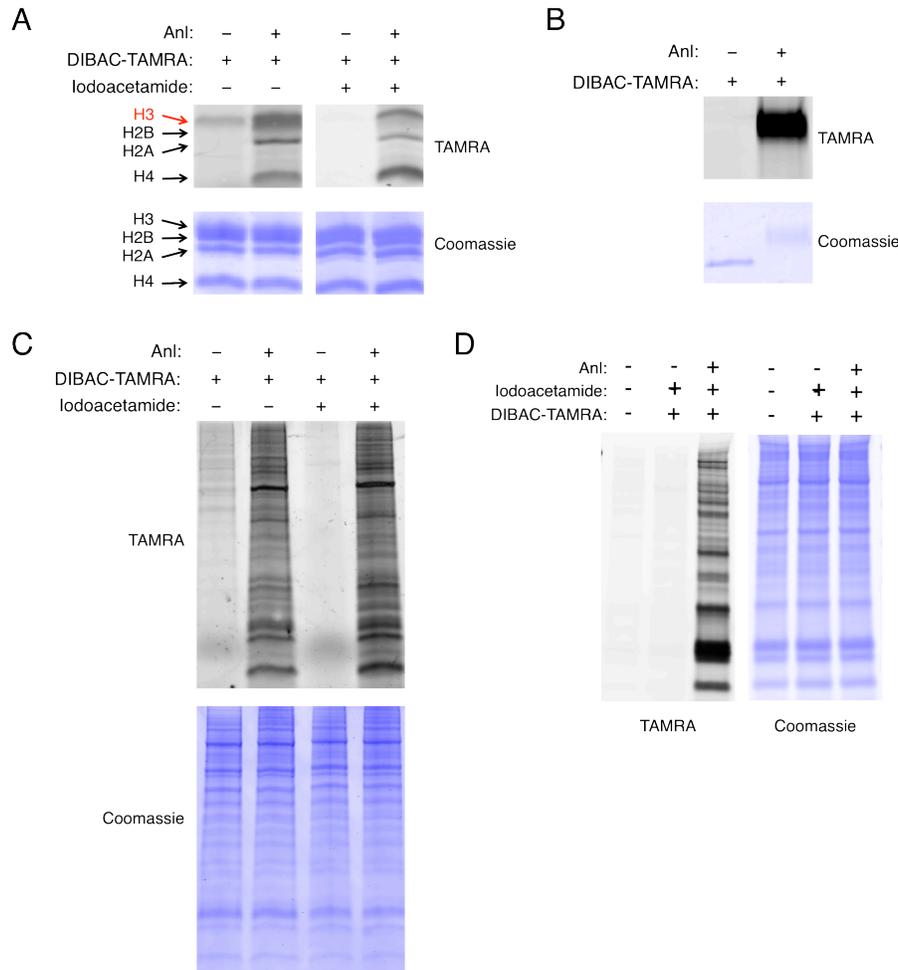
**Heterologous expression of the NLL-MetRS in mammalian cells.** (A) Schematic of the gene construct encoding the His-tagged NLL-MetRS-mCherry fusion as driven by the CMV promoter. (B) Color coded protein sequence of the encoded gene. (C) HEK293-8D3 cells exhibit mCherry fluorescence from expression of the synthetase-mCherry fusion and the protein is distributed throughout the cell. (D) Western analysis of non-transfected (1) and HEK293-8D3 cells (2) with anti-mCherry revealed that the His-tagged NLL-MetRS-mCherry is translated to full length (91.4 kDa). (E) The NLL-MetRS-mCherry is fully intact and soluble, as this protein was affinity purified from the soluble cytoplasmic fraction of proteins via an N-terminal His-tag.

#### Supplementary Figure 4.4



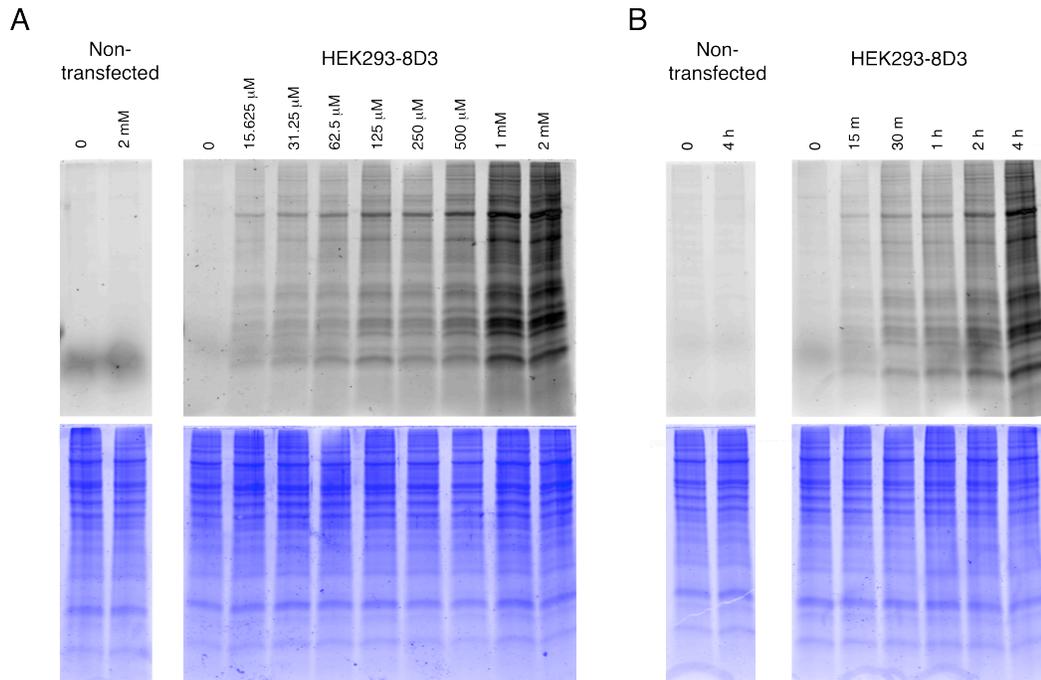
**Anl is incorporated into proteins synthesized in HEK293-8D3 cells.** Cells were pulsed with Anl and assessed for Anl-incorporation in fixed cells by conjugation to AlexaFluor488-alkyne followed by detection by fluorescence microscopy. Non-transfected cells that do not express the NLL-MetRS do not incorporate Anl into proteins. HEK293-8D3 cells are able to introduce Anl into proteins as indicated by cellular fluorescence from attached AlexaFluor488-alkyne. To confirm that Anl was incorporated into proteins, HEK293-8D3 was co-incubated with Anl and the protein synthesis inhibitor anisomycin. Protein synthesis inhibited-cells exhibited significantly diminished levels fluorescence from AlexaFluor488-alkyne.

## Supplementary Figure 4.5



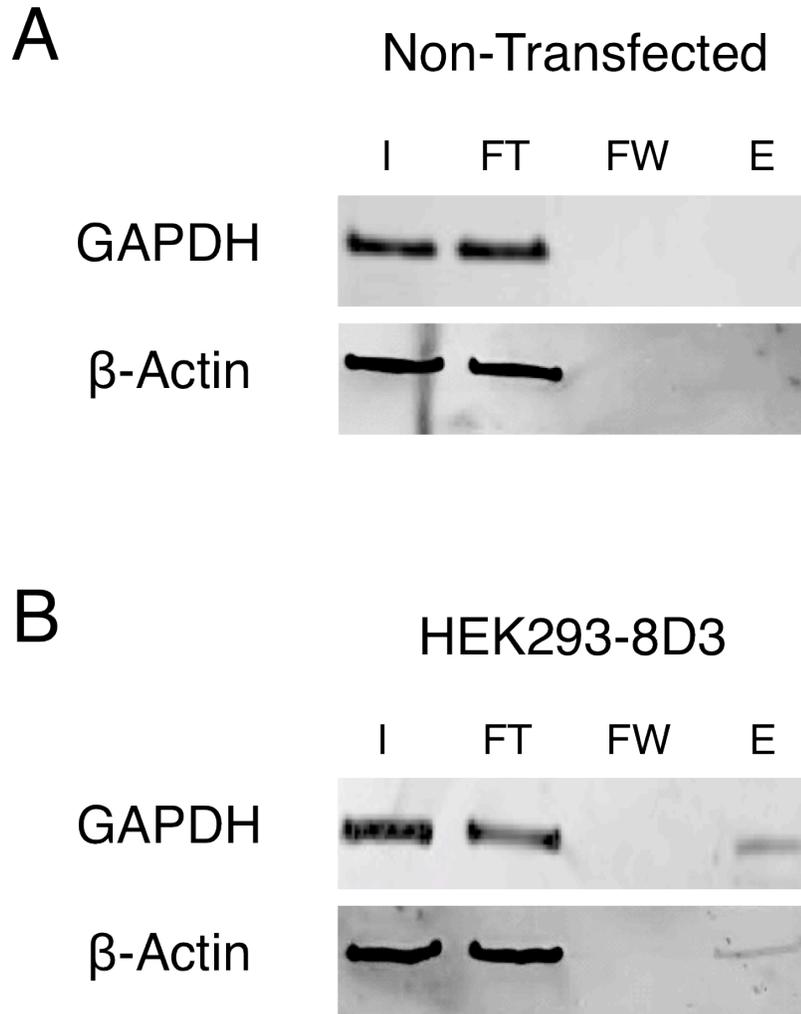
**The specificity of DIBAC-conjugation to Anl-tagged proteins is increased following cysteine alkylation with iodoacetamide.** (A) Untagged histones exhibit background labeling with DIBAC-TAMRA at the band corresponding to Histone H3, the only cysteine-containing histone. This background labeling is diminished following alkylation of cysteine with iodoacetamide. (B) Recombinant DHFR does not contain cysteine and untagged DHFR does not experience background labeling with DIBAC-TAMRA. A DHFR containing 7 azidohomoalanine residues is conjugated to DIBAC-TAMRA. A complete band shift of the azide-containing DHFR indicates that the reaction is highly efficient. (C) Untagged proteins in lysates exhibit decreased background labeling at cysteine residues following alkylation with iodoacetamide. (D) DIBAC-TAMRA labeling is highly specific following alkylation of proteins. Fluorescence intensity of alkylated proteins following reaction with DIBAC-TAMRA is compared to tagged proteins DIBAC-TAMRA, and proteins not exposed to DIBAC-TAMRA.

## Supplementary Figure 4.6



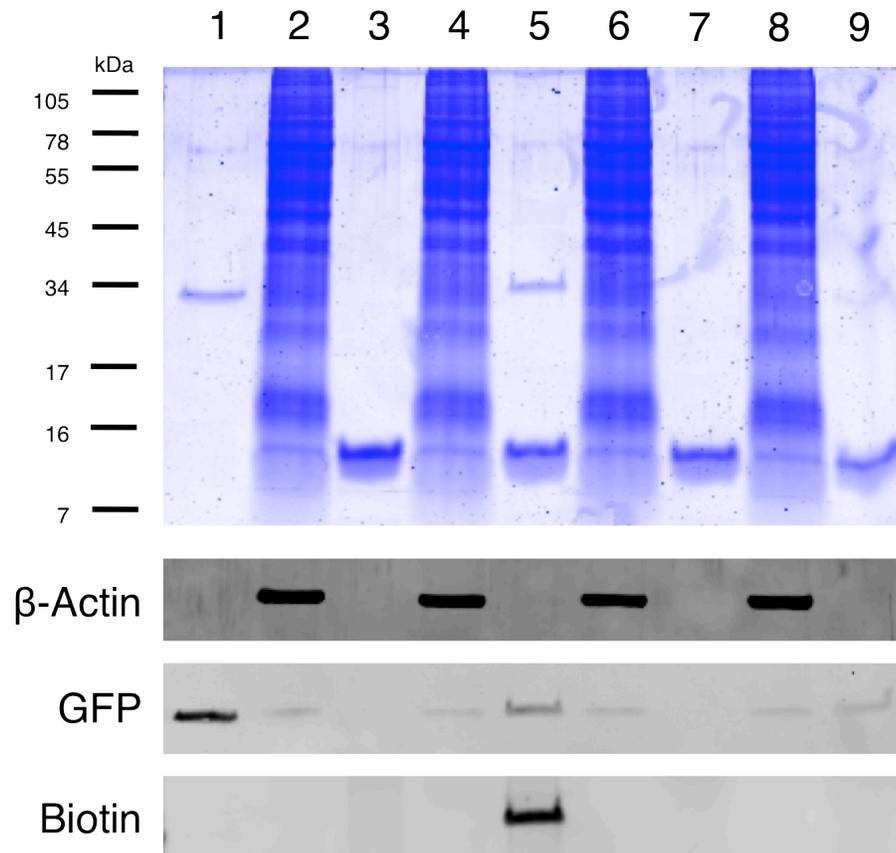
**Anl-tagging under varying pulse concentrations and durations.** Non-transfected HEK293 and HEK293-8D3 cells were pulsed with the indicated concentrations of Anl for 4 h (A), or pulsed with 1 mM Anl for the indicated times (B). To assess Anl-tagging, lysates alkylated with iodoacetamide, reacted with DIBAC-TAMRA, and protein were examined by in-gel fluorescence scanning. Proteins gels were subsequently stained with coomassie to ensure even loading (10  $\mu$ g of protein was loaded per lane).

Supplementary Figure 4.7



**Tagging and enrichment of constitutively expressed proteins.** Proteins from pulsed (A) non-transfected cells (“mock enrichment”), and (B) pulsed HEK293-8D3 cells were processed in an enrichment protocol using DIBAC-biotin and immobilized streptavidin. The enrichment was analyzed by western detection of the constitutively expressed “housekeeping” proteins GAPDH and  $\beta$ -actin. The lysate (input, “I”), flow-through (“FT”), final-wash (“FW”), and eluted (“E”) fractions were compared. As GAPDH and  $\beta$ -actin are continually synthesized, these proteins were tagged in pulsed HEK293-8D3 cells and enriched from the corresponding lysate. The absence of GAPDH and  $\beta$ -actin from the eluted fraction of the “mock enrichment” demonstrates the selectivity of the method.

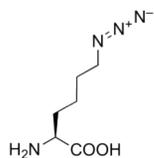
Supplementary Figure 4.8



**Estimation of the degree of enrichment.** Purified GFPs (28.2 kDa) containing a single Anl or Met were mixed with untagged HEK293 lysate at a ratio of 1:1,000 (w/w) and subjected to the enrichment procedure using DIBAC-biotin or DIBAC-S-S-biotin. The extent of enrichment was examined by comparing the “spiked” lysate with the eluted fraction following enrichment. Proteins were stained with coomassie, or transferred to membranes analyzed by western detection with  $\beta$ -actin antibody, anti-PentaHis antibody (GFP), or streptavidin (biotin). (1) Unmodified GFP. (2) lysate spiked with Met-GFP reacted with DIBAC-biotin, and (3) the corresponding elution. (4) lysate spiked with Anl-GFP reacted with DIBAC-biotin, and (5) the corresponding elution. (6) lysate spiked with Met-GFP reacted with DIBAC-S-S-biotin, and (7) the corresponding elution. (8) lysate spiked with Anl-GFP reacted with DIBAC-S-S-biotin, and (9) the corresponding elution. Anl-GFP was enriched from lysates with both DIBAC-probes, although less GFP was enriched with DIBAC-S-S-biotin. Using DIBAC-biotin, GFP was concentrated in the elution by over 3-fold. In all cases, untagged  $\beta$ -actin was not detected in the elution indicating that it was diluted by at least 120-fold.

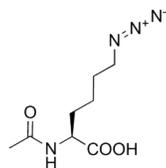
## Supplementary Figure 4.9

Anl



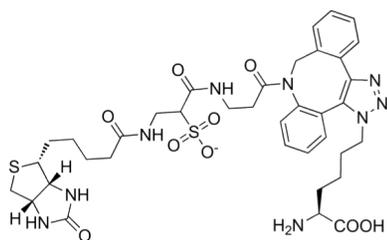
Chemical Formula:  $C_6H_{12}N_4O_2$   
 Exact Mass: 172.10  
 Molecular Weight: 172.19

Na-acetylated Anl



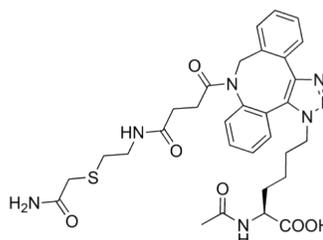
Chemical Formula:  $C_8H_{14}N_4O_3$   
 Exact Mass: 214.11  
 Molecular Weight: 214.22

DIBAC-biotin modified Anl



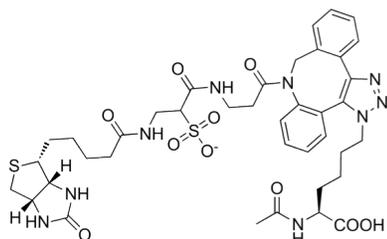
Chemical Formula:  $C_{37}H_{46}N_9O_9S_2^-$   
 Exact Mass: 824.29  
 Molecular Weight: 824.95

Na-acetylated  
 DIBAC-biotin modified Anl



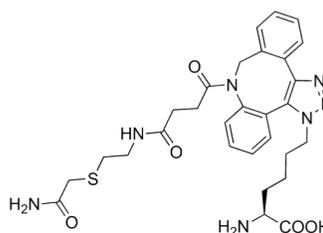
Chemical Formula:  $C_{31}H_{37}N_7O_6S$   
 Exact Mass: 635.25  
 Molecular Weight: 635.73

Cleaved and alkylated  
 DIBAC-S-S-biotin modified Anl



Chemical Formula:  $C_{39}H_{48}N_9O_{10}S_2^-$   
 Exact Mass: 866.30  
 Molecular Weight: 866.98

Na-acetylated, cleaved and alkylated  
 DIBAC-S-S-biotin modified Anl



Chemical Formula:  $C_{29}H_{35}N_7O_5S$   
 Exact Mass: 593.24  
 Molecular Weight: 593.70

**Structures used to determine the mass modifications applied in the database search for Anl-containing peptides.**

## Supplementary Figure 4.10

```

      10      20      30      40      50      60
MDGIVPDIAV GTKRGSDELF STCVTNGPFI MSSNSASAAN GNSKPKFGD SRSAGVPSRV
      70      80      90     100     110     120
IHIRKLPIDV TEGEVISLGL PFQKVTNLLM LKGKNQAFIE MNTEEAANTM VNYTTSVTPV
      130     140     150     160     170     180
LRGQPIYIQF SNHKELKTDS SPNQARAQAA LQAVNSVQSG NLALAASAAA VDAGMAMAGQ
      190     200     210     220     230     240
SPVLRRIIVEN LFYPVTLQVL HQIFSKFGTV LKIITFTKNN QFQALLQYAD PVSAQHAKLS
      250     260     270     280     290     300
LDGQNIYNAC CTLRIDFSKL TSLNVKYNND KSRDYTRPDL PSGDSQPSLD QTMAAAFGAP
      310     320     330     340     350     360
GIISASPYAG AGFPPTFAIP QAAGLSVPNV HGALAPLAIP SAAAAAAAAAG RIAIPGLAGA
      370     380     390     400     410     420
GNSVLLVSNL NPERVTQPSL FILFGVYGDV QRVKILFNKK ENALVQMADG NQAQLAMSHL
      430     440     450     460     470     480
NGHKLHGKPI RITLSKHQNV QLPREGQEDQ GLTKDYGNSP LHRFKKPGSK NFQNIFFPSA
      490     500     510     520     530     540
TLHLSNIPPS VSEEDLKVLF SSNGGVVKGK KFFQKDRKMA LIQMGSVVEA VQALIDLHNN
      550
DLGENHHLRV SFSKSTI

```

**Sequence coverage of PTBP.** Tryptic peptides are underlined in blue, and chymotryptic peptides are underlined in yellow. Met positions are highlighted in red.

### Supplementary Figure 4.11

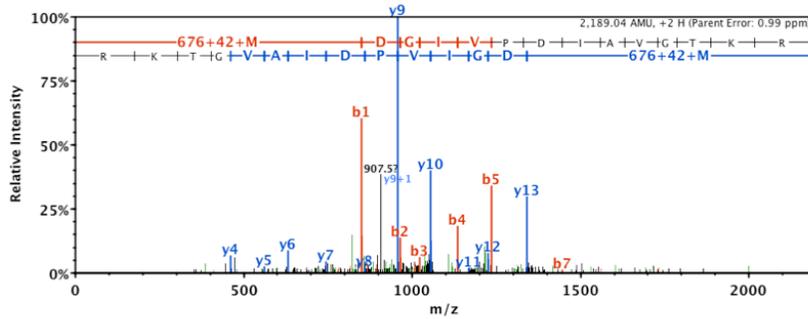
**Mass spectra of identified Anl-containing peptides.** (A-G) (top) Anl-containing peptides are highlighted yellow and the position of the Anl containing peptide is displayed in green. (middle) Spectrum corresponding to highlighted Anl-containing peptide. (bottom) Total sequence coverage of the associated protein is displayed with individual peptides highlighted in yellow. Modifications of any kind are shown in green (including Met-oxidation, cysteine alkylation, Anl incorporation). Modifications corresponding to Anl-incorporation were detected only at N-terminal positions.

# Supplementary Figure 4.11 cont.

**A**

IPI00183626 (100%), 59,634.4 Da  
 Tax\_Id=9606 Gene\_Symbol=PTBP1 polypyrimidine tract-binding protein 1 isoform a  
 1 unique peptides, 1 unique spectra, 1 total spectra, 14/557 amino acids (3% coverage)

<b>M</b> DGILV <b>P</b> DI <b>A</b> V	<b>G</b> T <b>K</b> R	GSD <b>E</b> L <b>F</b>	S <b>T</b> C <b>V</b> T <b>N</b> G <b>P</b> F <b>I</b>	M <b>S</b> S <b>N</b> S <b>A</b> S <b>A</b> A <b>N</b>	G <b>N</b> D <b>S</b> K <b>K</b> F <b>K</b> G <b>D</b>
S <b>R</b> S <b>A</b> G <b>V</b> P <b>S</b> R <b>V</b>	I <b>H</b> I <b>R</b> K <b>L</b> P <b>I</b> D <b>V</b>	T <b>E</b> G <b>E</b> V <b>I</b> S <b>L</b> G <b>L</b>	P <b>F</b> G <b>K</b> V <b>T</b> N <b>L</b> L <b>M</b>	L <b>K</b> G <b>K</b> N <b>Q</b> A <b>F</b> I <b>E</b>	
M <b>N</b> T <b>E</b> E <b>A</b> A <b>N</b> T <b>M</b>	V <b>N</b> Y <b>T</b> S <b>V</b> T <b>P</b> V	L <b>R</b> G <b>Q</b> P <b>I</b> Y <b>I</b> Q <b>F</b>	S <b>N</b> H <b>K</b> E <b>L</b> K <b>T</b> D <b>S</b>	S <b>P</b> N <b>Q</b> A <b>R</b> A <b>Q</b> A <b>A</b>	
L <b>Q</b> A <b>V</b> N <b>S</b> V <b>Q</b> S <b>G</b>	N <b>L</b> L <b>A</b> A <b>S</b> A <b>A</b> A	V <b>D</b> A <b>G</b> M <b>A</b> M <b>A</b> G <b>Q</b>	S <b>P</b> V <b>L</b> R <b>I</b> I <b>V</b> E <b>N</b>	L <b>F</b> Y <b>P</b> V <b>T</b> L <b>D</b> V <b>L</b>	
H <b>Q</b> I <b>F</b> S <b>K</b> F <b>G</b> T <b>V</b>	L <b>K</b> I <b>I</b> T <b>F</b> T <b>K</b> N <b>N</b>	Q <b>F</b> Q <b>A</b> L <b>L</b> Q <b>Y</b> A <b>D</b>	P <b>V</b> S <b>A</b> Q <b>H</b> A <b>K</b> L <b>S</b>	L <b>D</b> G <b>Q</b> N <b>I</b> Y <b>N</b> A <b>C</b>	
C <b>T</b> L <b>R</b> I <b>D</b> F <b>S</b> K <b>L</b>	T <b>S</b> L <b>N</b> V <b>K</b> Y <b>N</b> N <b>D</b>	K <b>S</b> R <b>D</b> Y <b>T</b> R <b>P</b> D <b>L</b>	P <b>S</b> G <b>D</b> S <b>Q</b> P <b>S</b> L <b>D</b>	Q <b>T</b> M <b>A</b> A <b>A</b> F <b>G</b> A <b>P</b>	
G <b>I</b> L <b>S</b> A <b>S</b> P <b>Y</b> A <b>G</b>	T <b>S</b> L <b>N</b> V <b>K</b> Y <b>N</b> N <b>D</b>	Q <b>A</b> A <b>G</b> L <b>S</b> V <b>P</b> N <b>V</b>	H <b>G</b> A <b>L</b> A <b>P</b> L <b>A</b> I <b>P</b>	S <b>A</b> A <b>A</b> A <b>A</b> A <b>A</b> G	
R <b>I</b> A <b>I</b> P <b>G</b> L <b>A</b> G <b>A</b>	G <b>N</b> S <b>V</b> L <b>L</b> V <b>S</b> N <b>L</b>	N <b>P</b> E <b>R</b> V <b>T</b> P <b>Q</b> S <b>L</b>	F <b>I</b> L <b>F</b> G <b>V</b> Y <b>G</b> D <b>V</b>	Q <b>R</b> V <b>K</b> I <b>L</b> F <b>N</b> K <b>K</b>	
E <b>N</b> A <b>L</b> V <b>Q</b> M <b>A</b> D <b>G</b>	N <b>Q</b> A <b>Q</b> L <b>A</b> M <b>S</b> H <b>L</b>	N <b>G</b> H <b>K</b> L <b>H</b> G <b>K</b> P <b>I</b>	R <b>I</b> T <b>L</b> S <b>K</b> H <b>Q</b> N <b>V</b>	Q <b>L</b> P <b>R</b> E <b>G</b> Q <b>E</b> D <b>Q</b>	
G <b>L</b> T <b>K</b> D <b>Y</b> G <b>N</b> S <b>P</b>	L <b>H</b> R <b>F</b> K <b>K</b> P <b>G</b> S <b>K</b>	N <b>F</b> Q <b>N</b> I <b>F</b> P <b>P</b> S <b>A</b>	T <b>L</b> H <b>L</b> S <b>N</b> I <b>P</b> P <b>S</b>	V <b>S</b> E <b>D</b> L <b>K</b> V <b>L</b> F	
S <b>S</b> N <b>G</b> G <b>V</b> V <b>K</b> G <b>F</b>	K <b>F</b> F <b>Q</b> K <b>D</b> R <b>K</b> M <b>A</b>	L <b>I</b> Q <b>M</b> G <b>S</b> V <b>E</b> E <b>A</b>	V <b>Q</b> A <b>L</b> I <b>D</b> L <b>H</b> N <b>H</b>	D <b>L</b> G <b>E</b> N <b>H</b> H <b>L</b> R <b>V</b>	
S <b>F</b> S <b>K</b> S <b>T</b> I					



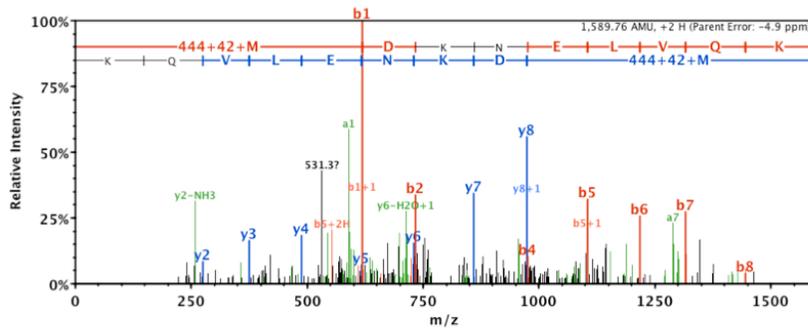
IPI00183626 (100%), 59,634.4 Da  
 Tax\_Id=9606 Gene\_Symbol=PTBP1 polypyrimidine tract-binding protein 1 isoform a  
 15 unique peptides, 17 unique spectra, 21 total spectra, 239/557 amino acids (43% coverage)

<b>M</b> DGILV <b>P</b> DI <b>A</b> V	<b>G</b> T <b>K</b> R	GSD <b>E</b> L <b>F</b>	S <b>T</b> C <b>V</b> T <b>N</b> G <b>P</b> F <b>I</b>	M <b>S</b> S <b>N</b> S <b>A</b> S <b>A</b> A <b>N</b>	G <b>N</b> D <b>S</b> K <b>K</b> F <b>K</b> G <b>D</b>
S <b>R</b> S <b>A</b> G <b>V</b> P <b>S</b> R <b>V</b>	I <b>H</b> I <b>R</b> K <b>L</b> P <b>I</b> D <b>V</b>	T <b>E</b> G <b>E</b> V <b>I</b> S <b>L</b> G <b>L</b>	P <b>F</b> G <b>K</b> V <b>T</b> N <b>L</b> L <b>M</b>	L <b>K</b> G <b>K</b> N <b>Q</b> A <b>F</b> I <b>E</b>	
M <b>N</b> T <b>E</b> E <b>A</b> A <b>N</b> T <b>M</b>	V <b>N</b> Y <b>T</b> S <b>V</b> T <b>P</b> V	L <b>R</b> G <b>Q</b> P <b>I</b> Y <b>I</b> Q <b>F</b>	S <b>N</b> H <b>K</b> E <b>L</b> K <b>T</b> D <b>S</b>	S <b>P</b> N <b>Q</b> A <b>R</b> A <b>Q</b> A <b>A</b>	
L <b>Q</b> A <b>V</b> N <b>S</b> V <b>Q</b> S <b>G</b>	N <b>L</b> L <b>A</b> A <b>S</b> A <b>A</b> A	V <b>D</b> A <b>G</b> M <b>A</b> M <b>A</b> G <b>Q</b>	S <b>P</b> V <b>L</b> R <b>I</b> I <b>V</b> E <b>N</b>	L <b>F</b> Y <b>P</b> V <b>T</b> L <b>D</b> V <b>L</b>	
H <b>Q</b> I <b>F</b> S <b>K</b> F <b>G</b> T <b>V</b>	L <b>K</b> I <b>I</b> T <b>F</b> T <b>K</b> N <b>N</b>	Q <b>F</b> Q <b>A</b> L <b>L</b> Q <b>Y</b> A <b>D</b>	P <b>V</b> S <b>A</b> Q <b>H</b> A <b>K</b> L <b>S</b>	L <b>D</b> G <b>Q</b> N <b>I</b> Y <b>N</b> A <b>C</b>	
C <b>T</b> L <b>R</b> I <b>D</b> F <b>S</b> K <b>L</b>	T <b>S</b> L <b>N</b> V <b>K</b> Y <b>N</b> N <b>D</b>	K <b>S</b> R <b>D</b> Y <b>T</b> R <b>P</b> D <b>L</b>	P <b>S</b> G <b>D</b> S <b>Q</b> P <b>S</b> L <b>D</b>	Q <b>T</b> M <b>A</b> A <b>A</b> F <b>G</b> A <b>P</b>	
G <b>I</b> L <b>S</b> A <b>S</b> P <b>Y</b> A <b>G</b>	A <b>G</b> F <b>P</b> P <b>T</b> F <b>A</b> I <b>P</b>	Q <b>A</b> A <b>G</b> L <b>S</b> V <b>P</b> N <b>V</b>	H <b>G</b> A <b>L</b> A <b>P</b> L <b>A</b> I <b>P</b>	S <b>A</b> A <b>A</b> A <b>A</b> A <b>A</b> G	
R <b>I</b> A <b>I</b> P <b>G</b> L <b>A</b> G <b>A</b>	G <b>N</b> S <b>V</b> L <b>L</b> V <b>S</b> N <b>L</b>	N <b>P</b> E <b>R</b> V <b>T</b> P <b>Q</b> S <b>L</b>	F <b>I</b> L <b>F</b> G <b>V</b> Y <b>G</b> D <b>V</b>	Q <b>R</b> V <b>K</b> I <b>L</b> F <b>N</b> K <b>K</b>	
E <b>N</b> A <b>L</b> V <b>Q</b> M <b>A</b> D <b>G</b>	N <b>Q</b> A <b>Q</b> L <b>A</b> M <b>S</b> H <b>L</b>	N <b>G</b> H <b>K</b> L <b>H</b> G <b>K</b> P <b>I</b>	R <b>I</b> T <b>L</b> S <b>K</b> H <b>Q</b> N <b>V</b>	Q <b>L</b> P <b>R</b> E <b>G</b> Q <b>E</b> D <b>Q</b>	
G <b>L</b> T <b>K</b> D <b>Y</b> G <b>N</b> S <b>P</b>	L <b>H</b> R <b>F</b> K <b>K</b> P <b>G</b> S <b>K</b>	N <b>F</b> Q <b>N</b> I <b>F</b> P <b>P</b> S <b>A</b>	T <b>L</b> H <b>L</b> S <b>N</b> I <b>P</b> P <b>S</b>	V <b>S</b> E <b>D</b> L <b>K</b> V <b>L</b> F	
S <b>S</b> N <b>G</b> G <b>V</b> V <b>K</b> G <b>F</b>	K <b>F</b> F <b>Q</b> K <b>D</b> R <b>K</b> M <b>A</b>	L <b>I</b> Q <b>M</b> G <b>S</b> V <b>E</b> E <b>A</b>	V <b>Q</b> A <b>L</b> I <b>D</b> L <b>H</b> N <b>H</b>	D <b>L</b> G <b>E</b> N <b>H</b> H <b>L</b> R <b>V</b>	
S <b>F</b> S <b>K</b> S <b>T</b> I					

**B**

IPI00021263 (100%), 27,745.9 Da  
 Tax\_Id=9606 Gene\_Symbol=YWHAZ 14-3-3 protein zeta/delta  
 1 unique peptides, 1 unique spectra, 1 total spectra, 9/245 amino acids (4% coverage)

<b>M</b> DKNEL <b>V</b> Q <b>K</b> A	K <b>L</b> A <b>E</b> Q <b>A</b> E <b>R</b> Y <b>D</b>	D <b>M</b> A <b>A</b> C <b>M</b> K <b>S</b> V <b>T</b>	E <b>Q</b> G <b>A</b> E <b>L</b> S <b>N</b> E <b>E</b>	R <b>N</b> L <b>L</b> S <b>V</b> A <b>Y</b> K <b>N</b>
V <b>V</b> G <b>A</b> R <b>R</b> S <b>S</b> W <b>R</b>	V <b>V</b> S <b>S</b> I <b>E</b> Q <b>K</b> T <b>E</b>	G <b>A</b> E <b>K</b> K <b>Q</b> O <b>M</b> A <b>R</b>	E <b>Y</b> R <b>E</b> K <b>I</b> E <b>T</b> E <b>L</b>	R <b>D</b> I <b>C</b> N <b>D</b> V <b>L</b> S <b>L</b>
L <b>E</b> K <b>F</b> L <b>I</b> P <b>N</b> A <b>S</b>	Q <b>A</b> E <b>S</b> K <b>V</b> F <b>Y</b> L <b>K</b>	M <b>K</b> G <b>D</b> Y <b>R</b> Y <b>L</b> A	E <b>V</b> A <b>A</b> G <b>D</b> D <b>K</b> K <b>G</b>	I <b>V</b> D <b>S</b> Q <b>Q</b> A <b>Y</b> Q
E <b>A</b> F <b>E</b> L <b>S</b> K <b>K</b> E <b>M</b>	Q <b>P</b> T <b>H</b> P <b>I</b> R <b>L</b> L	A <b>L</b> N <b>F</b> S <b>V</b> F <b>Y</b> E	I <b>L</b> N <b>S</b> P <b>E</b> K <b>A</b> C <b>S</b>	L <b>A</b> K <b>T</b> A <b>F</b> D <b>E</b> A <b>I</b>
A <b>E</b> L <b>D</b> T <b>L</b> S <b>E</b> E <b>S</b>	Y <b>K</b> D <b>S</b> T <b>L</b> I <b>M</b> Q <b>L</b>	L <b>R</b> D <b>N</b> L <b>T</b> L <b>W</b> T <b>S</b>	D <b>T</b> Q <b>G</b> D <b>E</b> A <b>E</b> A <b>G</b>	E <b>G</b> G <b>E</b> N



IPI00021263 (100%), 27,745.9 Da  
 Tax\_Id=9606 Gene\_Symbol=YWHAZ 14-3-3 protein zeta/delta  
 13 unique peptides, 16 unique spectra, 36 total spectra, 148/245 amino acids (60% coverage)

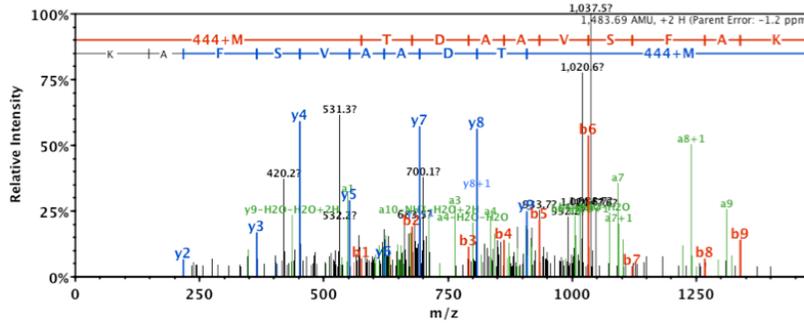
<b>M</b> DKNEL <b>V</b> Q <b>K</b> A	K <b>L</b> A <b>E</b> Q <b>A</b> E <b>R</b> Y <b>D</b>	D <b>M</b> A <b>A</b> C <b>M</b> K <b>S</b> V <b>T</b>	E <b>Q</b> G <b>A</b> E <b>L</b> S <b>N</b> E <b>E</b>	R <b>N</b> L <b>L</b> S <b>V</b> A <b>Y</b> K <b>N</b>
V <b>V</b> G <b>A</b> R <b>R</b> S <b>S</b> W <b>R</b>	V <b>V</b> S <b>S</b> I <b>E</b> Q <b>K</b> T <b>E</b>	G <b>A</b> E <b>K</b> K <b>Q</b> O <b>M</b> A <b>R</b>	E <b>Y</b> R <b>E</b> K <b>I</b> E <b>T</b> E <b>L</b>	R <b>D</b> I <b>C</b> N <b>D</b> V <b>L</b> S <b>L</b>
L <b>E</b> K <b>F</b> L <b>I</b> P <b>N</b> A <b>S</b>	Q <b>A</b> E <b>S</b> K <b>V</b> F <b>Y</b> L <b>K</b>	M <b>K</b> G <b>D</b> Y <b>R</b> Y <b>L</b> A	E <b>V</b> A <b>A</b> G <b>D</b> D <b>K</b> K <b>G</b>	I <b>V</b> D <b>S</b> Q <b>Q</b> A <b>Y</b> Q
E <b>A</b> F <b>E</b> L <b>S</b> K <b>K</b> E <b>M</b>	Q <b>P</b> T <b>H</b> P <b>I</b> R <b>L</b> L	A <b>L</b> N <b>F</b> S <b>V</b> F <b>Y</b> E	I <b>L</b> N <b>S</b> P <b>E</b> K <b>A</b> C <b>S</b>	L <b>A</b> K <b>T</b> A <b>F</b> D <b>E</b> A <b>I</b>
A <b>E</b> L <b>D</b> T <b>L</b> S <b>E</b> E <b>S</b>	Y <b>K</b> D <b>S</b> T <b>L</b> I <b>M</b> Q <b>L</b>	L <b>R</b> D <b>N</b> L <b>T</b> L <b>W</b> T <b>S</b>	D <b>T</b> Q <b>G</b> D <b>E</b> A <b>E</b> A <b>G</b>	E <b>G</b> G <b>E</b> N

# Supplementary Figure 4.11 cont.

**C**

IPI00007188 (100%), 32,896.5 Da  
 Tax\_Id=9606 Gene\_Symbol=SLC25A5 ADP/ATP translocase 2  
 1 unique peptides, 1 unique spectra, 1 total spectra, 10/298 amino acids (3% coverage)

<b>M</b> TDAAVSFAK	DFLAGGVAAA	ISKTAVAPIE	RVKLLQLVQH	ASKQITADKQ
YKGIIDCVVR	IPKEQGLVSF	WRGNLANVIR	YFPTQALNFA	FKDKYKQIFL
GVDKRTQFW	RYFAGNLSG	GAAGATSLCF	VYPLDFARTR	LAADVKGAGA
EREFRGLGDC	LVKIYKSDGI	KGLYQGFNV	VOQIIRAA	YFGIYDTAKG
MLDPDPKNTHI	VISWMIQTV	TAVAGLTSYP	FDTVRRRMMM	QSGRKGTDIM
YTGTLDCWRK	IARDEGGKAF	FKGAWSNVL	GMGGAFVLVL	YDEIKKTY



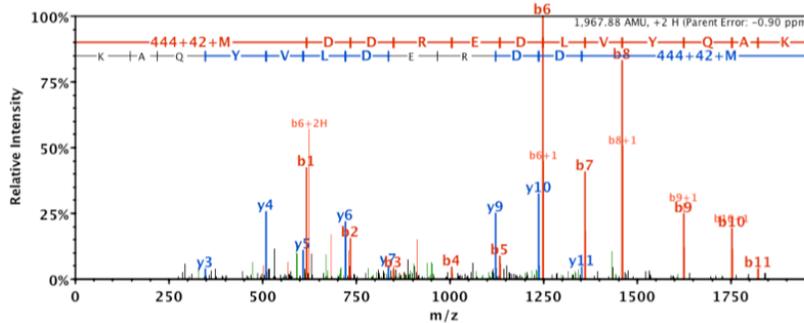
IPI00007188 (100%), 32,896.5 Da  
 Tax\_Id=9606 Gene\_Symbol=SLC25A5 ADP/ATP translocase 2  
 14 unique peptides, 19 unique spectra, 22 total spectra, 108/298 amino acids (36% coverage)

<b>M</b> TDAAVSFAK	DFLAGGVAAA	ISKTAVAPIE	RVKLLQLVQH	ASKQITADKQ
YKGIIDCVVR	IPKEQGLVSF	WRGNLANVIR	YFPTQALNFA	FKDKYKQIFL
GVDKRTQFW	RYFAGNLSG	GAAGATSLCF	VYPLDFARTR	LAADVKGAGA
EREFRGLGDC	LVKIYKSDGI	KGLYQGFNV	VOQIIRAA	YFGIYDTAKG
MLDPDPKNTHI	VISWMIQTV	TAVAGLTSYP	FDTVRRRMMM	QSGRKGTDIM
YTGTLDCWRK	IARDEGGKAF	FKGAWSNVL	GMGGAFVLVL	YDEIKKTY

**D**

IPI0000816 (100%), 29,175.0 Da  
 Tax\_Id=9606 Gene\_Symbol=YWHAE 14-3-3 protein epsilon  
 1 unique peptides, 1 unique spectra, 1 total spectra, 12/255 amino acids (5% coverage)

<b>M</b> DDRDLVYQ	AKLAEQAERY	DEMVESMKKV	AGMDVELTVE	ERNLLSVAYK
NVIGARRASW	RIISIEQKE	ENKGGEDKLL	MTREYRQMV	TEKLKICDI
LDVLDKHLIP	AANTGESKVF	YKMKGDYHR	YLAEFATGND	RKEAAENSLV
AYKAASDIAM	TELPPTHPIR	LGLALNFSVF	YYEILNSPDR	ACLAKAADF
DAIAELDTLS	EESYKDSTLI	MQLLRDNLT	WTSDMQGDGE	EQNKEALQDV
EDENQ				



IPI0000816 (100%), 29,175.0 Da  
 Tax\_Id=9606 Gene\_Symbol=YWHAE 14-3-3 protein epsilon  
 16 unique peptides, 19 unique spectra, 58 total spectra, 172/255 amino acids (67% coverage)

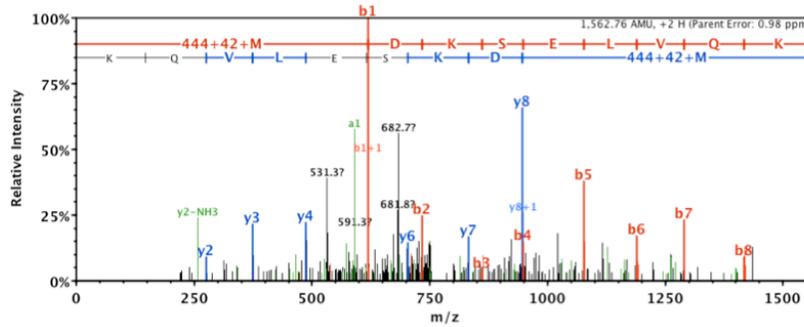
<b>M</b> DDRDLVYQ	AKLAEQAERY	DEMVESMKKV	AGMDVELTVE	ERNLLSVAYK
NVIGARRASW	RIISIEQKE	ENKGGEDKLL	MTREYRQMV	TEKLKICDI
LDVLDKHLIP	AANTGESKVF	YKMKGDYHR	YLAEFATGND	RKEAAENSLV
AYKAASDIAM	TELPPTHPIR	LGLALNFSVF	YYEILNSPDR	ACLAKAADF
DAIAELDTLS	EESYKDSTLI	MQLLRDNLT	WTSDMQGDGE	EQNKEALQDV
EDENQ				

## Supplementary Figure 4.11 cont.

### E

IPI00759832 (100%), 27,850.8 Da  
 Tax\_Id=9606 Gene\_Symbol=YWHAB Isoform Short of 14-3-3 protein beta/alpha  
 1 unique peptides, 1 unique spectra, 1 total spectra, 9/244 amino acids (4% coverage)

<b>M</b> D <b>K</b> S <b>E</b> L <b>V</b> Q <b>K</b> A	K L A E Q A E R Y D	D M A A A M K A V T	E Q G H E L S N E E	R N L L S V A Y K N
V V G A R R S S W R	V I S S I E Q K T E	R N E K K Q Q M G K	E Y R E K I E A E L	Q D I C N D V L E L
L D K Y L I P N A T	Q P E S K V F Y L K	M K G D Y F R Y L S	E V A S G D N K Q T	T V S N S Q Q A Y Q
E A F E I S K K E M	Q P T H P I R L G L	A L N F S V F Y Y E	I L N S P E K A C S	L A K T A F D E A I
A E L D T L N E E S	Y K D S T L I M Q L	L R D N L T L W T S	E N Q G D E G D A G	E G E N



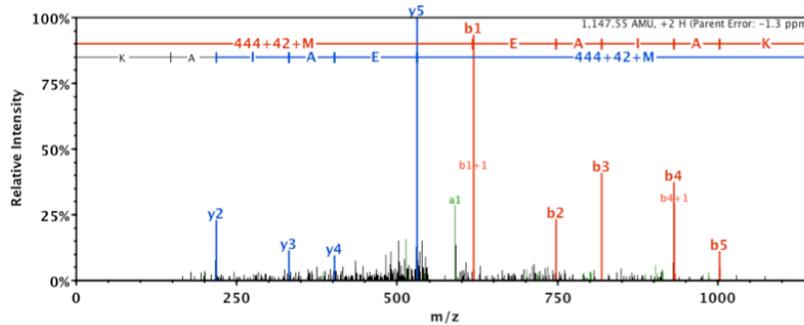
IPI00759832 (100%), 27,850.8 Da  
 Tax\_Id=9606 Gene\_Symbol=YWHAB Isoform Short of 14-3-3 protein beta/alpha  
 9 unique peptides, 10 unique spectra, 23 total spectra, 124/244 amino acids (51% coverage)

<b>M</b> D <b>K</b> S <b>E</b> L <b>V</b> Q <b>K</b> A	K L A E Q A E R Y D	<b>D</b> M <b>A</b> A <b>A</b> M <b>K</b> A <b>V</b> T	E Q G H E L S N E E	R N L L S V A Y K N
V V G A R R S S W R	V I S S I E Q K T E	R N E K K Q Q M G K	E Y R E K I E A E L	Q D I C N D V L E L
L D K Y L I P N A T	Q P E S K V F Y L K	M K G D Y F R Y L S	E V A S G D N K Q T	T V S N S Q Q A Y Q
<b>E</b> A <b>F</b> E <b>I</b> S <b>K</b> K <b>E</b> M	Q P T H P I R L G L	A L N F S V F Y Y E	I L N S P E K A C S	L A K T A F D E A I
A E L D T L N E E S	Y K <b>D</b> S <b>T</b> L <b>I</b> M <b>Q</b> L	<b>L</b> R <b>D</b> N <b>L</b> T <b>L</b> W <b>T</b> S	<b>E</b> N <b>Q</b> G <b>D</b> E <b>G</b> D <b>A</b> G	<b>E</b> G <b>E</b> N

### F

IPI00021327 (100%), 25,206.8 Da  
 Tax\_Id=9606 Gene\_Symbol=GRB2 Isoform 1 of Growth factor receptor-bound protein 2  
 1 unique peptides, 1 unique spectra, 1 total spectra, 6/217 amino acids (3% coverage)

<b>M</b> E <b>A</b> L <b>A</b> K	Y D F K	A T A D D E L S F K	R G D I L K V L N E	E C D Q N W Y K A E	L N G K D G F I P K
N Y I E M K P H P W	F F G K I P R A K A	E E M L S K Q R H D	G A F L I R E S E S	A P G D F S L S V K	
F G N D V Q H F K V	L R D G A G K Y F L	W V V K F N S L N E	L V D Y H R S T S V	S R N Q Q I F L R D	
I E Q V P Q Q T Y	V Q A L F D F D P Q	E D G E L G F R R G	D F I H V M D N S D	P N W W K G A C H G	
Q T C M F P R N Y V	T P V N R N V				



IPI00021327 (100%), 25,206.8 Da  
 Tax\_Id=9606 Gene\_Symbol=GRB2 Isoform 1 of Growth factor receptor-bound protein 2  
 11 unique peptides, 12 unique spectra, 20 total spectra, 101/217 amino acids (47% coverage)

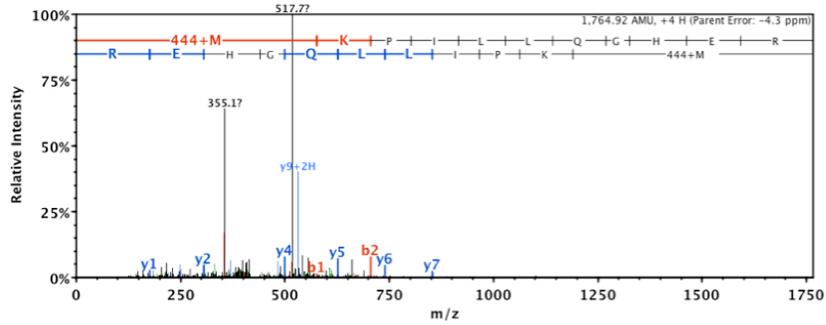
<b>M</b> E <b>A</b> L <b>A</b> K	Y D F K	<b>A</b> T <b>A</b> D <b>D</b> E <b>L</b> S <b>F</b> K	R G D I L K V L N E	<b>E</b> C <b>D</b> Q <b>N</b> W <b>Y</b> K <b>A</b> E	L N G K D G F I P K
N Y I E M K P H P W	F F G K I P R A K A	E E M L S K Q R H D	G A F L I R E S E S	A P G D F S L S V K	
<b>F</b> G <b>N</b> D <b>V</b> Q <b>H</b> F <b>K</b> V	L R <b>D</b> G <b>A</b> G <b>K</b> Y <b>F</b> L	<b>W</b> V <b>V</b> K <b>F</b> N <b>S</b> L <b>N</b> E	<b>L</b> V <b>D</b> Y <b>H</b> R <b>S</b> T <b>S</b> V	S R N Q Q I F L R D	
I E Q V P Q Q T Y	V Q A L F D F D P Q	E D G E L G F R R G	D F I H V M D N S D	P N W W K G A C H G	
<b>Q</b> T <b>C</b> M <b>F</b> P <b>R</b> N <b>Y</b> V	T P V N R N V				

## Supplementary Figure 4.11 cont.

**G**

IPI00012795 (100%), 36,502.0 Da  
 Tax\_Id=9606 Gene\_Symbol=EIF3I Eukaryotic translation initiation factor 3 subunit I  
 11 unique peptides, 1 unique spectra, 1 total spectra, 11/325 amino acids (3% coverage)

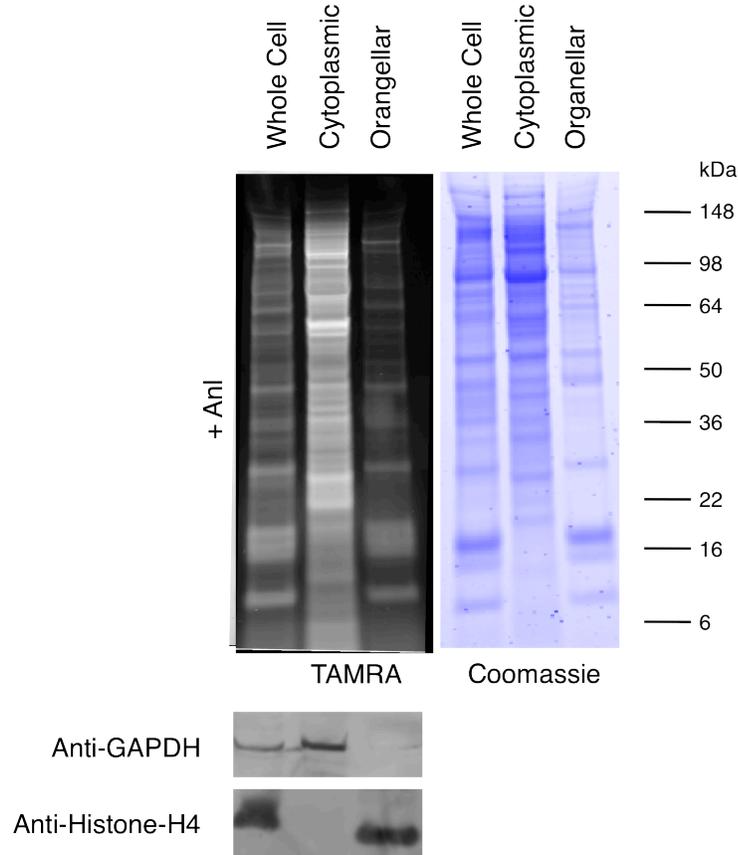
<b>M</b> K <b>P</b> I <b>L</b> L <b>Q</b> G <b>H</b> E	<b>R</b> S I T Q I K Y N R	E G D L L F T V A K	D P I V N V W Y S V	N G E R L G T Y M G
H T G A V W C V D A	D W D T K H V L T G	S A D N S C R L W D	C E T G K Q L A L L	K T N S A V R T C G
F D F G G N I I M F	S T D K Q M G Y Q C	F V S F F D L R D P	S Q I D N N E P Y M	K I P C N D S K I T
S A V W G P L G E C	I I A G H E S G E L	N Q Y S A K S G E V	L V N V K E H S R Q	I N D I Q L S R D M
T M F V T A S K D N	T A K L F D S T T L	E H Q K T F R T E R	P V N S A A L S P N	Y D H V V L G G G Q
E A M D V T T T S T	R I G K F E A R F F	H L A F E E E F G R	V K G H F G P I N S	V A F H P D G K S Y
S S G G E D G Y V R	I H Y F D P Q Y F E	F E F E A		



IPI00012795 (100%), 36,502.0 Da  
 Tax\_Id=9606 Gene\_Symbol=EIF3I Eukaryotic translation initiation factor 3 subunit I  
 11 unique peptides, 13 unique spectra, 16 total spectra, 151/325 amino acids (46% coverage)

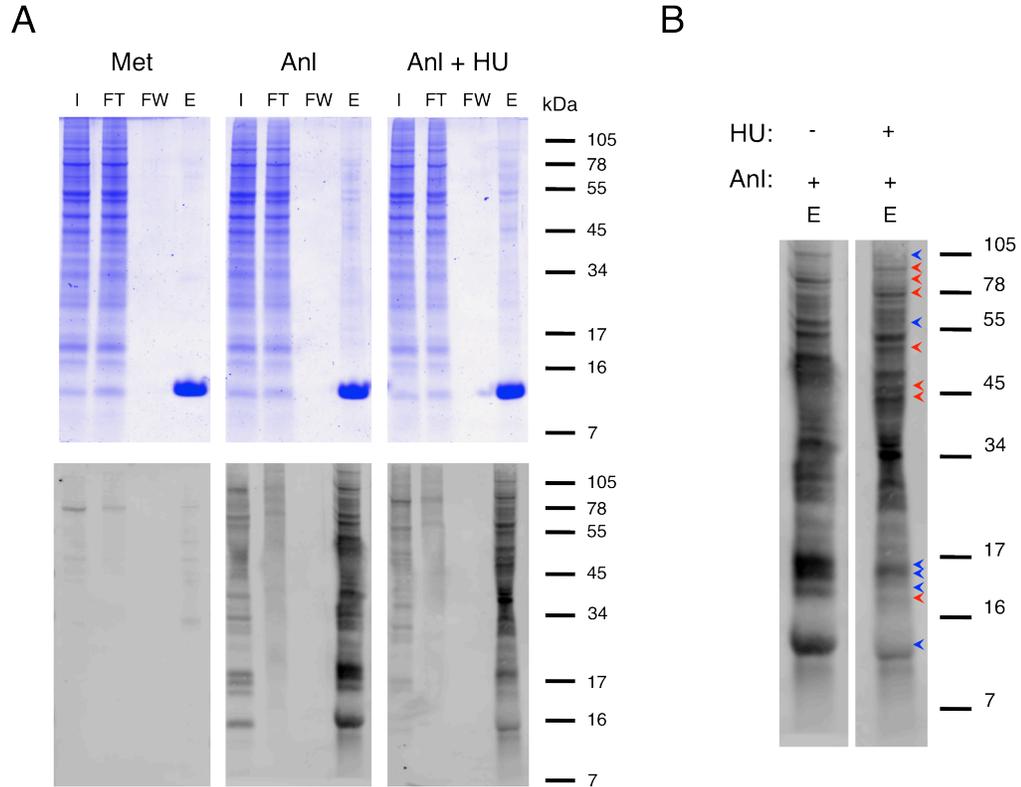
<b>M</b> K <b>P</b> I <b>L</b> L <b>Q</b> G <b>H</b> E	<b>R</b> S I T Q I K Y N R	E G D L L F T V A K	D P I V N V W Y S V	N G E R L G T Y M G
H T G A V W C V D A	D W D T K <b>H</b> V L T G	<b>S</b> A D N S <b>C</b> R L W D	C E T G K Q L A L L	K T N S A V R T <b>C</b> G
<b>F</b> D F G G N I I M F	<b>S</b> T D K Q M G Y Q C	F V S F F D L R D P	S Q I D N N E P Y M	K I P C N D S K I T
S A V W G P L G E C	I I A G H E S G E L	N Q Y S A K <b>S</b> G E V	<b>L</b> V N V K E H S R Q	I N D I Q L S R D M
<b>T</b> M F V T A S K D N	<b>T</b> A K L F D S T T L	E H Q K T F R T E R	<b>P</b> V N S A A L S P N	<b>Y</b> D H V V L G G G Q
<b>E</b> A M D V T T T S T	<b>R</b> I G K F E A R F F	<b>H</b> L A F E E E F G R	<b>V</b> K G H F G P I N S	<b>V</b> A F H P D G K S Y
<b>S</b> S G G E D G Y V R	I H Y F D P Q Y F E	F E F E A		

### Supplementary Figure 4.12



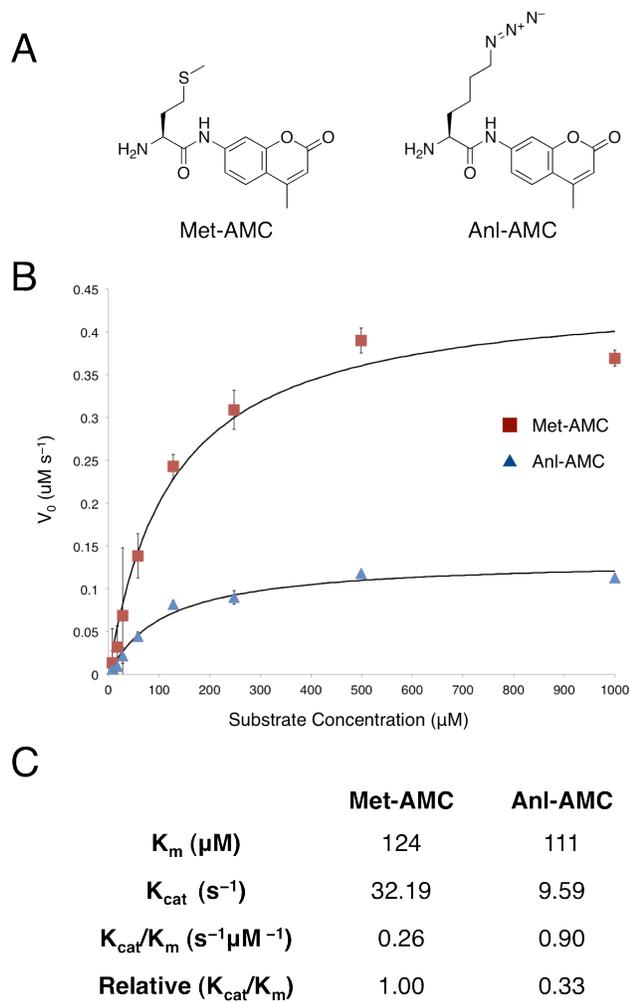
**Subcellular fractionation of pulsed HEK293-8D3 cells.** Anl-containing proteins were conjugated to DIBAC-TAMRA and total protein (whole cell) was compared to cytoplasmic and organellar proteins. Cytoplasmic proteins are present in the whole cell and cytoplasmic fractions, as indicated by western detection of GAPDH. The whole cell and organellar fractions contain nuclear proteins, as indicated by western detection of Histone H4. The highly labeled bands in the 10-20 kDa range present in the whole cell and organellar fractions correspond to the core histones.

Supplementary Figure 4.13



**Differences in translational profiles can be accentuated.** Unpulsed (Met) and pulsed (Anl) HEK293-8D3 cells were untreated or treated with HU (+HU). Tagged proteins enriched from Anl-treated HEK293-8D3 cells using DIBAC-biotin. (A) Enriched proteins were examined by coomassie staining (top, blue) and by streptavidin detection (bottom). (B) Proteins from untreated and HU-treated HEK293-8D3 cells were enriched and analyzed by streptavidin detection. Differences in band intensities are highlighted. Bands exhibiting decreased streptavidin intensity in the +HU lane are indicated with blue arrowheads. Bands exhibiting increased streptavidin intensity in the +HU lane are indicated with red arrowheads.

## Supplementary Figure 4.14



**MetAP exhibits attenuated excision activity toward N-terminal Anl.** (A) Structures of fluorogenic model substrates. (B) Saturation curves of Met-AMC and Anl-AMC in reaction with recombinant human MetAP-2. (C) Kinetic constants determined by reaction of Met-AMC and Anl-AMC with MetAP-2.

**Table 1. Detected peptides displaying evidence of An1-incorporation**

Protein	Total Unique Peptides Detected	Sequence Coverage	Detected Modification of An1	Position of Modification	N-terminal Peptide Containing Met	Internal Positions Encoding Met	Number of Internal Met Detected
PTBP, Polypyrimidine-Tract Binding Protein	15	43%	N $\alpha$ -acetylation, DIBAC-biotin	N-terminal	Not Detected	11	4
GRB2, Growth Factor Receptor Bound Protein 2	11	47%	N $\alpha$ -acetylation, reduced and alkylated DIBAC-S-S-Biotin	N-terminal	Not Detected	4	1
YWHAB Isoform Short Of 14-3-3 Protein Beta/Alpha	9	51%	N $\alpha$ -acetylation, reduced and alkylated DIBAC-S-S-Biotin	N-terminal	Not Detected	6	4
YWHAZ 14-3-3 Protein Zeta/Delta	13	60%	N $\alpha$ -acetylation, reduced and alkylated DIBAC-S-S-Biotin	N-terminal	Not Detected	6	5
SLC25A5 ADP/ATP Translocase 2	14	36%	Reduced and alkylated DIBAC-S-S-Biotin	N-terminal	Not Detected	6	4
YWHAE 14-3-3 Protein Epsilon	16	67%	N $\alpha$ -acetylation, reduced and alkylated DIBAC-S-S-Biotin	N-terminal	Not Detected*	9	6
EIF3I Eukaryotic Translation Initiation Factor 3 Subunit 1	11	46%	Reduced and alkylated DIBAC-S-S-Biotin	N-terminal	Not Detected*	6	4

**Supplementary Table 1. MetAP processing of tagged proteins**

<b>Protein</b>	<b>Direct Evidence of Ani Incorporation by:</b>	<b>Known Substrate of MetAP?</b>
PTBP, Polypyrimidine-Tract Binding Protein	Mass Spectrometry	NO
GRB2, Growth Factor Receptor Bound Protein 2	Mass Spectrometry	NO
YWHAB Isoform Short Of 14-3-3 Protein Beta/Alpha	Mass Spectrometry	YES
YWHAZ 14-3-3 Protein Zeta/Delta	Mass Spectrometry	NO
SLC25A5 ADP/ATP Translocase 2	Mass Spectrometry	YES
YWHAE 14-3-3 Protein Epsilon	Mass Spectrometry	NO
EIF3I Eukaryotic Translation Initiation Factor 3 Subunit 1	DIBAC-TAMRA Fluoresence	YES
Histone H3	DIBAC-TAMRA Fluoresence	YES
Histone H2B	DIBAC-TAMRA Fluoresence	YES
Histone H2A	DIBAC-TAMRA Fluoresence	YES
Histone H4	DIBAC-TAMRA Fluoresence	YES
Histone H1	DIBAC-TAMRA Fluoresence	NO
p53	DIBAC-TAMRA Fluoresence	YES
$\beta$ -Actin	DIBAC-biotin Affinity Purification	YES
GAPDH, Glyceraldehyde 3-Phosphate Dehydrogenase	DIBAC-biotin Affinity Purification	YES

## MATERIALS AND METHODS

### Tagging and Detection Reagents

Azidonorleucine was synthesized as previously described but with *N* $\alpha$ -Boc-L-Lysine (Bachem) as the starting material in place of diaminobutyric acid (Link, A.J., *et al*, *Nat. Protoc.* **2**, 1879—1883). TLC, <sup>1</sup>H NMR, and mass spectrometry were used to confirm the structure and purity of the amino acid. A stock solution of the amino acid was prepared at a concentration of 100 mM in water and sterilized by passage through a 0.2  $\mu$ m filter. The solution were stored at 4 °C until use. Stocks of dibenzo-aza-cyclooctyne-carboxytetramethylrhodamine (DIBAC-TAMRA, Click Chemistry Tools), sulfo-dibenzo-aza-cyclooctyne-biotin (DIBAC-biotin, Click Chemistry Tools), and dibenzo-aza-cyclooctyne-S-S-PEG<sub>3</sub>-biotin (DIBAC-S-S-biotin, Click Chemistry Tools) were prepared at 10 mM in DMSO to 10 mM and stored in aliquots at -20 °C until use. All other materials from commercial suppliers were used without further purification as described in subsequent sections. The Click-iT Alexa Fluor 488 Protein Synthesis HCS Assay (product #C10289, Invitrogen) was used to prepare cells for fluorescence microscopy.

### Western Blotting and Antibodies

Western blots were performed using standard procedures. Briefly, following separation by SDS-PAGE, proteins were transferred to nitrocellulose and blocked in 5% (w/v) dry non-fat milk in TBST for at least 15 m. Membranes were incubated with antibodies at the concentration described below using either blocking buffer or the manufacturer's suggested buffer. Membranes were washed three times with TBST for 5 m each before addition of secondary antibodies or detection on fluorescence scanner (Typhoon Trio+, GE Healthcare). Mass approximations on blots (as well as for SDS-PAGE gels) were made using the pre-stained protein molecular weight marker SeeBlue Plus2 (Invitrogen).

Monoclonal Anti-GAPDH from mouse (Sigma) was used at 1:2,000. Monoclonal Anti-p53 from mouse (Sigma) was used at 1:1,000. Polyclonal Anti-DsRed from rabbit (Clontech) was used at 1:1,000. Monoclonal Anti-Histone H4 (pan) from rabbit (Millipore) was used at 1:1,000. Monoclonal anti- $\beta$ -actin from mouse (Sigma) was used at 1:2,000. Monoclonal Anti-PentaHis-AlexaFlour647 primary conjugate (Qiagen) was used at 1:2,000. Secondary antibodies included Anti-Mouse-AlexaFlour647 conjugate from rabbit (Invitrogen) and Anti-rabbit-DyLight680 conjugate from goat (Cell Signalling). Streptavidin-AlexaFlour488 and Streptavidin-AlexaFlour647 conjugates (Invitrogen) were used at 1:2,000.

### **Cloning**

All primers were custom synthesized from Integrated DNA Technologies. Constructs were verified by DNA sequencing, which was performed by Laragen with the listed primers.

The pmCherry-N1-NLL\_MetRS encodes the N-terminally His-tagged NLL-MetRS-mCherry under control of the CMV promoter and also contains a resistance marker for selection in bacterial and mammalian cells. In order to clone pmCherry-N1-NLL\_MetRS, DNA encoding an N-terminally His-tagged mutant *E. coli* methionyl-tRNA synthetase ( $\Delta$ 547, L13N, Y260L, H301L) was amplified by PCR from pQE80-NLL-MetRS with the primers XhoI\_NLL\_MetRS\_F and XhoI\_NLL\_MetRS\_R. The resulting fragment was inserted into pmCherry-N1 (Clontech) between XhoI and XmaI sites using standard digestion and ligation procedures. The ligation mixture was transformed into competent DH10B *E. coli* (Invitrogen) and plasmid-carrying clones were selected on LB-agar plates containing 35 mg L<sup>-1</sup> of kanamycin sulfate.

pJS2-NLL is a bacterial expression plasmid that encodes an IPTG-inducible His-tagged mutant GFP containing a single Met site and also encodes the

NLL-MetRS under a constitutive promoter. In order to clone pJS2-NLL, DNA encoding a gene cassette for constitutive untagged NLL-MetRS was isolated from pJTN1 by digestion with *NheI*. The isolated fragment was inserted at the *NheI* site of pJS2 using standard digestion and ligation procedures. The ligation mixture was transformed into competent DH10B *E. coli* and plasmid-carrying clones were selected on LB-agar plates containing 200 mg L<sup>-1</sup> of ampicillin.

#### **PCR Primer Sequences:**

XhoI\_NLL\_MetRS\_F: 5'-AGATCTCGAGATGAGAGGATCGCATCACCA-3'

XmaI\_NLL\_MetRS\_R: 5'-GGATCCCGGGTTTTAGAGGCTTCCACCAGTGC-3'

#### **Sequencing Primers:**

pmCherryN1-NLL\_seq\_1F: 5'-GGTGGCGAAACAGGAAGG-3'

pmCherryN1-NLL\_seq\_2F: 5'-AAGACCATGGGCTGGGAG-3'

pmCherryN1-NLL\_seq\_1R: 5'-CATCAGCCTGGATGTGCTC-3'

pmCherryN1-NLL\_seq\_2R: 5'-GCAGTACCGGCTTCAGGTAA-3'

### **Transfection and Selection**

For transfection, pmCherry-N1-NLL\_MetRS was isolated from DH10B *E. coli* using the EndoFree Plasmid Maxi Kit (Qiagen) according to the supplier's protocol. A 30 mL suspension culture of the HEK293-F cell line (Invitrogen) was grown to a density of 1 x 10<sup>6</sup> cells/mL in 293FreeStyle media (Invitrogen) in a humidified incubator at 37 °C with 8% CO<sub>2</sub> and 130 rpm shaking. Cell were transfected with 150 µg of pmCherry-N1-NLL\_MetRS using polyethylenimine as the transfection reagent. Following transfection, the culture was allowed to grow for 48 h at which point an aliquot of cells were inspected by for mCherry fluorescence by microscopy. Roughly half of

surviving cells exhibited mCherry fluorescence. For selection of stable clones, a 10 mL aliquot of the cell suspension was transferred to a T-75 culture flask and cells were allowed to adhere to the flask surface for 24 hours. Subsequently, the media was exchanged for fresh 293FreeStyle medium containing 50  $\mu\text{g mL}^{-1}$  of geneticin sulfate. Individual clones were isolated into 96-wel plates and maintained in pre-conditioned 293FreeStyle medium before exchange to increasing larger culture vessels. In total, ten lines were generated (including the characterized HEK293-8D3) with varying degrees of mCherry fluorescence. Cell stocks were prepared in 293FreeStyle media containing 5% dimethylsulfoxide and frozen overnight at  $-80\text{ }^{\circ}\text{C}$  in a slow-freezing container before submersion in liquid  $\text{N}_2$  for long-term storage.

### **Cell Culture Maintenance**

Cell stocks of HEK293-8D3 and non-transfected HEK293-F were thawed plated in T-75 flasks containing 15 mL of DMEM +X with 10% FBS. After 48 h, geneticin sulfate was added to the HEK293-8D3 culture to a final concentration of 50  $\mu\text{g mL}^{-1}$ . The culture was maintained over the course of experiments in cell-culture treated T-75 flasks (Corning) or on 100 mm cell-culture treated culture dishes (BD Bioscience). The culture media was replenished every 1 – 3 days, depending on density. Cells passaged at ~80% confluence using trypsin-EDTA solution (0.05% Trypsin-EDTA, Invitrogen) and diluted into fresh media at between 1:10 and 1:20 dilution ratios. Cultures were discarded after ~20 passages and replaced with new cultures from frozen stocks.

### **General Pulse Labeling Procedure**

Cells were prepared by exchanging the culture medium with fresh maintenance medium the day prior to pulse labeling. Cells at between 65 – 85% confluence were pulsed by addition of Anl the culture medium at the indicated final concentration and incubated for the indicated time at  $37\text{ }^{\circ}\text{C}$ ,

with 5% CO<sub>2</sub>. The pulse was stopped by aspiration of the culture medium at which point cells were rinsed once with room temperature PBS. Cells were collected by trypsinization and centrifugation. Pelleted cells were rinsed once with PBS and the rinse solution was removed and the cell pellets were stored at -20 °C.

### **Preparation of Cell Lysates**

Frozen cells were thawed on ice and suspended in 100 mM Tris HCl (pH 8.0) containing protease inhibitors (eComplete ETDA-free, Roche, 1 tablet/50 mL), and 0.2% SDS (diluted from a 10% stock solution in water). Cells were lysed in this solution by pipetting and agitation. Chromosomal DNA was sheared by sonication on ice with 3 x 10 s bursts at 9 W with a 10 s pause between bursts. Protein concentrations were measured by BCA Protein Assay (Pierce) and normalized to 1 mg mL<sup>-1</sup> with lysis buffer. Free thiols were blocked by iodoacetamide, which was added to a final concentration of 25 mM (from a stock solution in water). The alkylation reaction was protected from light and placed on a rotator at room temperature for 1 h. The protein solution was then immediately frozen or further processed as described in subsequent sections.

### **Protein Modification and Detection with DIBAC-functionalized Probes**

Alkylated lysates were reacted with DIBAC-probes by their addition to protein solutions from 100 mM stocks in DMSO. Reactions with DIBAC-TAMRA were performed at a final concentration of 20 μM of the probe and incubated for 20 m protected from light and with rotation at room temperature. Reactions with DIBAC-biotin and DIBAC-S-S-biotin were performed at a final concentration of 20 μM for each probe and incubated for 1 h with rotation at room temperature. Reactions were stopped by the addition of a five-fold excess of free AnI (relative to DIBAC-probe) followed by vortexing.

### **Sub-cellular Fractionation**

Cells were pulsed with Anl for 4 h and collected as described above before differential detergent fractionation using the Qproteome Cell Compartment Kit (Qiagen). Proteins were precipitated from the cytoplasmic and nuclear fractions and resuspended in 100 mM Tris HCl pH 8.0 containing 0.2% (w/v) SDS and alkylated with iodoacetamide as described above. Fractions were compared with unfractionated total protein by reaction with DIBAC-TAMRA (as described above).

### **Dye-Labeling of Fixed Cells with AlexaFluor488-alkyne**

HEK293-8D3 and non-transfected HEK293 cells were grown to ~80% confluence on 100 mm cell culture treated petri dishes, trypsinized and then seeded onto 8-well Lab-Tek II Chamber Slides (ThermoScientific) at a density of  $2.5 \times 10^4$  cells per well. Cells were allowed to adhere by incubation in 293FreeStyle or maintenance medium overnight at 37 °C, with 5% CO<sub>2</sub>. Cells in 293FreeStyle media did not adhere as well as those grown in the serum-containing maintenance medium. The following day, the media was removed by gentle pipetting and replaced with pre-warmed 293FreeStyle or maintenance medium with or without 1 mM Anl. For protein synthesis inhibited controls, anisomycin was added to 40 μM by addition to the culture 30 m prior to exposure to Anl and also after exchange with Anl-containing medium. Cells were incubated 4 h at 37 °C, with 5% CO<sub>2</sub>. Following the “pulse” the culture medium of each well was removed by gentle pipetting. Cells were gently rinsed once with room temperature PBS. Cells were labeled then fixed, permeablized, dye-labeled, and stained with Hoechst 33342 according to the manufacturer’s instructions accompanying the Click-iT Alexa Fluor 488 Protein Synthesis HCS Assay.

### **Fluorescence Microscopy**

All images were obtained with a Zeiss LSM510 confocal microscope using a 20X or 40X lens, or a 63X oil immersion lens.

For detection of mCherry, HEK293-8D3 and non-transfected HEK293 from liquid suspension culture in the 293FreeStyle medium were plated onto black-well clear-bottom 96-well poly-D-Lysine plates (Fisher) at  $\sim 5 \times 10^3$  wells per well. Cells were allowed to adhere overnight before imaging the next day.

### **Histone Extractions**

Frozen cell pellets were resuspended in 3 mL of cold PBS with 0.5% (v/v) Triton X-100 containing eComplete EDTA-free protease inhibitor cocktail at 1 pellet per 50 mL. Cells were lysed gentle pipetting followed by a 10 m incubation on ice. Nuclei were sedimented by centrifugation at 2,000 x g for 10 m at 4° C. The supernatant was collected pelleted nuclei were resuspended in an additional 1.5 mL of cold PBS with 0.5% Triton X-100 to ensure efficient lysis. Nuclei were sedimented again. Histone extraction was performed by resuspension of the nuclear pellet in 0.5 mL 0.2 N HCl and incubated overnight with rotation at 4 °C. The following day, samples were centrifuged at 2,000 x g for 10 m at 4 °C to sediment nuclear debris. The supernatant containing acid-soluble nuclear proteins (including histones) was collected. Prior to protein modification, the solution pH was raised by the addition of an equal volume of 1 M Tris pH 8.0 containing 1% (w/v) SDS.

### **Model Protein Expression in *E. coli* and Purification**

DH10B *E. coli* transformed with pJS2-NLL and pREP4 and maintained using 100 mg L<sup>-1</sup> ampicillin and 35 mg L<sup>-1</sup> kanamycin. Cells grown overnight in LB were diluted 1:50 in M9 minimal medium containing supplemented with 40 mg L<sup>-1</sup> of each of the 20 amino acids 1 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 30 mg L<sup>-1</sup> thiamine hydrochloride, and 0.4% (w/v) glucose. When cells reached an

OD<sub>600</sub> of 1.0, the culture was sedimented by centrifugation at 8,000 × g for 5 min at 4° C and washed with 0.9% NaCl (w/v) solution in water. This was repeated once before cells were resuspended in the same medium without Met and instead containing 1 mM Anl and 1 mM IPTG. After a 4 h expression, cells were collected. Affinity purification of the His-tagged GFP containing 1 Anl was performed on Ni-NTA resin (Qiagen) according to the manufacturer's protocol for purification under native conditions. The purified protein was buffer exchanged into PBS using PD-10 columns (GE Healthcare) and quantified by BCA protein assay.

### **Enrichment of Anl-labeled Proteins**

T-75 flasks of cells were pulsed for 6 h as described above using 2 mM Anl. Lysates were prepared as described above such that the final volume of the solution was ~2 mL following alkylation. Proteins were biotinylated as described above. Excess biotin reagent was removed from the quenched reaction by two successive gel filtration steps using PD-10 Desalting Columns (GE Healthcare). New columns were equilibrated with 100 mM Tris pH 8.0 with 0.2% (w/v) SDS prior to sample loading. In the first step, proteins were desalted according to the "spin protocol" described by the supplier. In the second step, proteins were desalted again on new and equilibrated columns according to the "gravity flow" protocol provided by the manufacturer. A 50 µL aliquot of eComplete protease inhibitor cocktail was added to the elution from a concentrated stock of 1 tablet/mL. Biotinylated proteins were collected on Streptavidin Plus UltraLink Resin (Pierce) by adding 250 µL of a 50% slurry each solution as and incubating the suspension overnight at room temperature with rotation. The following day, the suspension was loaded onto a disposable chromatography column (1 mL capacity) and the flow-through was collected by drip-flow. The resin was washed once with 100 mM Tris pH 8.0 containing 0.2% (w/v) SDS, once with 100 mM Tris pH 8.0 containing 1.0% (w/v) SDS, once with a 6 M urea solution in 100 mM Tris pH 8.0, once with a 1 M NaCl solution in 100 mM Tris pH 8.0, and finally once

again with 100 mM Tris pH 8.0 containing 0.2% (w/v) SDS. Each wash was performed using 20-fold of the resin bed volume (2.5 mL). The entire wash procedure was repeated once before proteins were eluted from the resin by boiling in reducing SDS-PAGE loading buffer.

### **Model Protein Enrichment**

The GFP containing 1 Anl (expressed and purified as described above) was added to HEK293 lysates at a ratio of 1:1,000 (3 µg of GFP added to 3 mg of lysate). Proteins were enriched using DIBAC-biotin or DIBAC-S-S-biotin as described. The enrichment was analyzed by coomassie staining of protein gels, and by western detection of beta-actin, His-tagged GFP, and streptavidin. Coomassie stained gels were analyzed by densitometric analysis using ImageQuant (GE Healthcare).

### **Analysis by Mass Spectrometry**

Proteomic analysis was performed at the Protein Exploration Laboratory (California Institute of Technology, Pasadena, CA). Proteins separated by SDS-PAGE and peptides were generated by in-gel digest according to a previously described detailed procedure (Shevchenko, A., et al. *Nature Protocols* **1**, 2856 – 286). Proteins were digested in separate reactions with either sequencing grade trypsin or chymotrypsin. Extracted peptides were dried by rotary vacuum and resuspended in 0.1% formic acid in water (v/v) before were analysis on a Thermo Scientific LTQFT mass spectrometer.

### **Database Searching**

All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version Mascot). Mascot was set up to search the IPI\_human\_wDEC.v3.54j database (unknown version, 151385 entries) assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.50 Da and a parent ion tolerance of 10.0

PPM. Iodoacetamide derivative of cysteine was specified in Mascot as a fixed modification. Oxidation of methionine, met->ANL of methionine, acetylation of the n-terminus, DSSB cleaved of methionine and DSSBc+Acetyl of methionine were specified in Mascot as variable modifications.

Criteria for Modification: Scaffold (version Scaffold\_3\_00\_07, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 90.0% probability as specified by the Peptide Prophet algorithm (Keller, A et al Anal. Chem. 2002;74(20):5383-92). Protein identifications were accepted if they could be established at greater than 90.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii, AI Anal Chem. 2003 Sep 1;75(17):4646-58). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

### **Hydroxyurea Treatment**

Hydroxyurea (Sigma) was added to cultures of HEK293-8D3 to a final concentration of 2.5 mM from a 1.25 M stock in water and pulsed with Anl five minutes later. Cells were pulsed for 4 hours before proteins were collected and analyzed as described in preceding sections.

### **Aminoacylation of synthetically prepared tRNA**

NLL-MetRS was expressed and purified as previously described.<sup>13</sup> Synthetic tRNAs were prepared by the run-off transcript method and aminoacylated with Met or Anl as previously described.<sup>18</sup> We use the cytoplasmic initiator and elongator-tRNA<sup>Met</sup> sequences from *C. elegans* (which was similar to the mammalian sequences), and initiator and elongator-tRNA<sup>Met</sup> sequences *E. coli*. Aminoacylated tRNAs were stabilized by acetic anhydride as previously described (Walker, S.E., Fredrick, K., *Methods*. **44**, 81–86). Na-

acyl-aminoacyl-tRNAs were precipitated with ice-cold 70% ethanol in water by standard methods. Precipitated tRNAs were normalized and resuspended in 100 mM TRIS pH 7.0 with RNase inhibitor and labeled with DIFO-AlexaFluor488 (a gift from Carolyn Bertozzi, UC Berkeley). Excess dye was removed using BioSpin gel filtration columns (BioRad). tRNAs were separated on 2% agarose and imaged on a Typhoon Trio+ (GE Healthcare).

### **MetAP Enzymatic Assay**

The fluorogenic substrates methionine-7-amino-4-methylcoumarin (Met-AMC, Enzo Life Sciences) and azidonorleucine-7-amino-4-methylcoumarin (Anl-AMC, custom synthesis from Anaspec) were used in kinetic assays of purified recombinant human methionine aminopeptidase 2 (MetAP2, Enzo Life Sciences). Reactions were performed in 30 mM HEPES buffer (pH 7.4), 150 mM NaCl, 100  $\mu$ M CoCl<sub>2</sub>, and carried out in black 96-well plates. Substrate was added to each reaction well at concentrations ranging from 0 - 1000  $\mu$ M and with a final concentration of 4% DMSO to ensure substrate solubility. Reactions were initiated by addition of 1  $\mu$ L of MetAP2 from a 0.5 mg/mL stock, giving a final enzyme concentration of 7.14 nM. The change in fluorescence of each reaction was measured in a plate reader with excitation 380 nm and emission detection at 460 nm. Triplicate measurements were averaged and fit to the Michaelis-Menten equation using Origin.

### **Purification of the MetRS from HEK 293F-8D3**

To collect soluble cytoplasmic proteins, HEK293-8D3 cells were lysed in 1% (v/v) Triton-X100 solution in PBS with protease inhibitors. The N-terminally His-tagged NLL-MetRS-mCherry was purified from the lysate of HEK293-8D3 cells using magnetic Ni-NTA agarose beads (Qiagen) according to the manufacturer's protocol. As a control, the purification was also performed using a lysate of non-transfected HEK293 cells. Proteins were analyzed by separation on 9% polyacrylamide SDS-PAGE gels and stained with

coomassie. We confirmed that the protein purified from HEK293-8D3 cells was indeed the NLL-MetRS-mCherry by in-gel tryptic digestion of the gel fragment containing the band and LC/MS/MS analysis of extracted peptides as described above.

## REFERENCES

1. Housman, D., Jacobs-Lorena, N., RahBhandary, U. L., Lodish, H. F. *Nature* **227**, 913 – 918 (1970).
2. Smith, A.E., Marcker, K.A. *Nature* **226**, 607 – 610 (1970).
3. Kozak, M. *Microbiol. Rev.* **47**, 1 – 45 (1983).
4. Gite, S., Mamaev, S., Olejnik, J., Rothschild, K. *Anal. Biochem.* **279**, 218—225 (2000).
5. Olejnik, J., Gite, Mamaev, S., Rothschild, K.J. *Methods* **36**, 252—260 (2005).
6. Schulman, L.H. *Prog. Nucleic Acid Res.* **41**, 23—87 (1991).
7. RajBhandary, U.L., Ghosh, H.P. *J. Biol. Chem.* **224**, 1104—1113 (1969).
8. Gupta, N.K., Chatterjee, N.K., Bose, K.K. *J. Mol. Biol.* **54**, 145 – 154 (1970).
9. Blanquest, D., Petrissant, G., Waller, J.P. *Eur. J. Biochem.* **36**, 227—233 (1973).
10. Guillemaut, P., Weill, J.H. *Biochim. Biophys. Acta.* **407**, 223—225 (1975).
11. Petrissant, G., Boisnard, M., Puissant, C. *Biochim. Biophys. Acta.* **213**, 223—225 (1970).
12. Meinnel, T., Mechulam, Y., Fayat, G., Blanquet, S. *Nucleic Acids Res.* **20**, 4741—4746 (1992).
13. Tanrikulu, I C., Schmitt, E., Mechulam, Y., Goddard, W.A. and Tirrell, D.A. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15285—15290 (2009).

14. Link, A.J., Vink, M.K.S., Agard, N.J., Prescher, J.A., Bertozzi, C.R. & Tirrell, D.A. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10180—10185 (2006).
15. Schmitt, E., Tanrikulu, I.C., Yoo, T.H., Panvert, M., Tirrell, D.A., Mechulam, Y. *J. Mol. Biol.* **394**, 843—851 (2009).
16. Ngo, J.T., Champion, J. A., Mahdavi, A., Tanrikulu, I. C., Beatty, K. E., Conner, R. E., Yoo, T. H., Dieterich, D. C., Schuman, E. M. & Tirrell, D. A. *Nat. Chem. Biol.* **5**, 715—717 (2009).
17. Abdeljabbar, D.M., Klein, T.J., Link, A.J. *ChemBioChem*, **12**, 1699 – 1702 (2011).
18. Pestova, T.V., Hellen, C.U. *RNA* **7**, 1496—1505 (2001).
19. Drabkin, H.J. and RajBhandary, U.L. *Mol. Cell. Biol.* **18**, 5140 (1998).
20. Wagner, T., Gross, M., Sigler, P.B. *J. Biol. Chem.* **259**, 4706—4709 (1984).
21. Hong, V., Presolski, S.I., Ma, C. & Finn, M.G. *Angew. Chem. Int. Ed. Engl.* **48**, 9879—9883 (2009).
22. Jewett, J.C., Bertozzi, C.R. *Chem. Soc. Rev.* **39**, 1272—1299 (2010).
23. Debets, M.F., van Berkel, S.S., Schoffelen, S., Rutjes, F.P.J.T, van Hest, J.C.M. & van Delft, F.L. *Chem. Commun.*, **46**, 97-99 (2010).
24. Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A. & Schuman, E.M. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9487—9487 (2006).
25. Beatty, K.E., Xie, F., Qian, W., Tirrell, D.A. *J. Am. Chem. Soc.* **127**, 14150—14151 (2005).
26. Ngo, J.T., Tirrell, D.A. *Acc. Chem. Res.* **44**, 677—685 (2011).

27. Szychowski, J., et al. *J. Am. Chem. Soc.* **132**, 18351—18360 (2010).
28. Jacob, E., Unger, R. *Bioinformatics* **23**, e225-e230 (2007).
29. Van Damme, P., Arnesen, T., Gevaert, K. *FEBS J.* **278**, 3822—3834 (2011).
30. Brechi, L.A., Tabb, D.L., Yates, J.R., Wysocki, V.H. *Anal. Chem.* **75**, 1963—1971 (2003).
31. Ewen, M.E., *Genes Dev.* **14**, 2265-2270 (2000)
32. Marxluff, W.A., Duronio, R.J. *Curr. Opin. Cell Biol.* **14**, 692—699 (2002)
33. Yoshia, K., Yoshio, M., *Cancer Sci.* **101**, 831-835 (2010).
34. Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R.C., Giglione, C., Meinel, T. *Mol Cell Proteomics.* **5**, 2336—2349 (2006).
35. Wang, A., Winblade Nairn, N., Johnson, R.S., Tirrell, D.A., Grabstein, K. *Chembiochem.* **9**, 324—330 (2008).
36. Drag, M., Bogyo, M., Ellman, J.A., Salvesen, G.S. *J. Biol. Chem.* **285**, 3310—3318 (2009)
37. To determine if an individual proteins was a MetAP substrate, we searched the Protein Knowledgebase (at <http://www.uniprot.org>) and looked for prior evidence of N-terminal Met retention or cleavage.
38. Lin, M.Z., Glenn, J.S., Tsien, R.Y. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7744—7749 (2008)
39. Liu, C.C., Schultz, P.G. *Annu. Rev. Biochem.* **79**, 413—44 (2010).

40. Johnson, J.A., Lu, Y.Y., Van Deventer, J.A., Tirrell, D.A. *Curr. Opin. Chem. Biol.* **14**, 774—780 (2010).

## CHAPTER V

### *Toward Cell-Specific Labeling of Proteins in Multicellular Organisms*

#### INTRODUCTION

Cell type-specific expression of genes is fundamental to multicellular life. Expression of specific sets of proteins plays an essential role in defining both the identity and the functional state of the cell (Tomancak, 2007). The human body contains more than 200 individual cell types, all sharing a single genome; each cell type is distinguished by the set of genes it expresses. The genes expressed define the cellular proteome, which in turn dictates cellular behavior and gives each cell type its characteristic features (localization, morphology, metabolic state, etc). Because of the heterogeneity of cell types in whole organisms, proteomic analysis on the level of a single cell type is challenging. Data from *in vivo* proteomics (or protein expression data derived from complex cellular mixtures) can be difficult to interpret; small changes in a subset of the cells can be hard to detect and the cellular origins of detected proteins can be difficult to assign. We propose here an approach that we believe will provide a powerful and general solution to the challenge of acquiring time-resolved, cell type-specific proteomic data from living animals.

In 2006, we introduced the BONCAT (bio-orthogonal non-canonical amino acid tagging) method, a strategy for selective enrichment and identification of newly synthesized proteins in cells (Dieterich, et al, 2006). The BONCAT method involves pulse-labeling of cells with the non-canonical amino acid azidohomoalanine (Aha) a surrogate for methionine. Proteins made during

the pulse are labeled with Aha and are therefore subject to selective conjugation to alkyne-functionalized affinity probes via the Cu (I) catalyzed azide-alkyne cycloaddition reaction. Newly synthesized, affinity-tagged proteins can be enriched from the preexisting, unlabeled protein pool and identified by mass spectrometry. For visualization of newly synthesized proteins, proteins pulse-labeled in the same fashion can be conjugated to dyes for tracking by fluorescence microscopy (Beatty, et al, 2005; Beatty, et al, 2006).

Recent work from our laboratory has built on the BONCAT approach to achieve cell-selective labeling of proteins in mixtures of cultured cells (Ngo, et al, 2009). *Escherichia coli* cells expressing the L13N/Y260L/H301L mutant form of the methionyl-tRNA synthetase (NLL-MetRS) can use the non-canonical amino acid azidonorleucine in translation (Tanrikulu, et al, 2009). Cells that do not express the mutant enzyme are inert to AnI; in a cellular mixture, only proteins from mutant cells are labeled. Through this approach, proteins synthesized in targeted cells can be selectively isolated from complex mixtures for identification by mass spectrometry or conjugated to fluorescent dyes for *in situ* visualization.

Cell-selective protein labeling in living animals is perhaps the ultimate test of a selective labeling. We envision that cell-selectivity *in vivo* can be achieved in much the same fashion as in cell culture: confining expression of the NLL-MetRS to selected cells should restrict labeling to these cells even in living animals. The nematode *C. elegans* serves as an ideal model system for the development of cell-selective protein labeling methods, for a variety of reasons. *C. elegans* is a multicellular organism for which full genome sequence data is available, making it suitable for proteomic interrogation. Reliable methods have been developed for rapid generation of transgenic animals, allowing for fast screening of a variety of platforms for NLL-MetRS expression (Mello et al, 1991). The small size and transparency of the worm allow for observation of cellular phenomena in the intact organism with light

microscopy with single cell resolution. Furthermore, *C. elegans* has a defined cell number with a completely characterized cell lineage (Sulston, 1977; Sulston, 1983), and a catalog of cell-specific promoters that drive gene expression in nearly every cell type can be compiled from the literature (Hunt-Newbury et al, 2007). We proposed the use of cell-type specific transcriptional enhancers to regulate expression of NLL-MetRS in *C. elegans*, such that protein labeling with Anl is restricted to selected cell types in Anl-fed worms.

### PRELIMINARY WORK

In a preliminary analysis, we asked whether or not a non-canonical amino acid fed to *C. elegans* could be uptaken and incorporated into proteins. To examine this, we provided worms with Aha and interrogated *C. elegans* proteins for evidence of the azide-containing amino acid. Our intention was to employ Aha as a proxy for Anl in order to develop a feeding strategy that would eventually be used to deliver Anl to worms. Since Aha is a substrate for wild-type MetRS and can be incorporated into proteins using endogenous machinery, standard laboratory strains of *C. elegans* could be tested without needing to create a transgenic organism.

The most common food source for laboratory *C. elegans* is *Escherichia coli*. Thus, we employed *E. coli* as a vehicle for delivery of Aha to worms. In order to maximize the amount of Aha provided to worms, we prepared heavily labeled *E. coli*. Briefly, a culture of methionine auxotrophic *E. coli* at an OD<sub>600</sub> of 1.0 was pelleted and cells were resuspended in synthetic medium that did not contain Met. Aha was added to the culture and expression of a mutant GFP using IPTG. We induced GFP to ensure that bacterial cells contained an abundant level of Aha-containing protein. Following a 6 h labeling, bacterial cells were collected. Amino acid analysis of total protein extracted from the Aha-labeled *E. coli* revealed that Met and Aha were in equal abundance, and that Aha accounted for 0.8% of total amino acid content.

Aha-labeled *E. coli* were fed to a suspension culture of worm in liquid medium (S basal). As a control, a separate culture was fed unlabeled (or Met-labeled) *E. coli*. After an overnight feeding, we noticed much of the bacteria added each culture had been consumed. To assess Aha incorporation in worms, total protein was extracted and conjugated to an alkyne-functionalized fluorescent dye (alkyne-TAMRA) via Cu (I) catalyzed azide-alkyne ligation. Subsequently, proteins were separated by SDS-PAGE and labeled proteins were detected by fluorescence scanning. Proteins from *C. elegans* fed Aha-labeled *E. coli* exhibited incorporation of Aha into numerous proteins as indicated by strong fluorescence intensity from alkyne-TAMRA (Fig. 1A). Proteins from *C. elegans* fed unlabeled *E. coli* did not exhibit TAMRA fluorescence.

In order to be certain that the detected proteins were *C. elegans* proteins, and not proteins from the Aha-labeled *E. coli*, we performed western analysis of protein extracts with streptavidin (Fig. 1B). In lanes containing *C. elegans* proteins, we did not detect the naturally biotinylated biotin-carboxyl carrier protein (17 kDa) characteristic to *E. coli* lysates. This result suggested that the *C. elegans* lysate was predominantly composed of *C. elegans* proteins, and that TAMRA intensity was from dye-conjugation to *C. elegans* proteins.

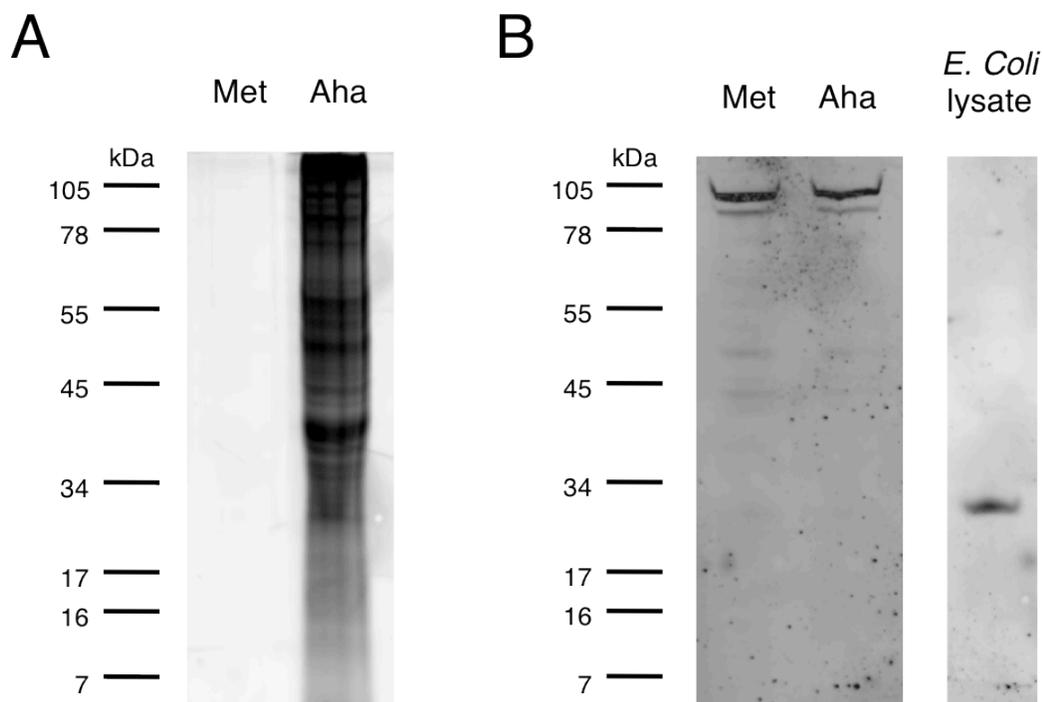
To confirm that *C. elegans* proteins were indeed labeled, we conjugated alkyne-coumarin to proteins in fixed and permeabilized worms. Detection of coumarin by fluorescence microscopy revealed labeling throughout worms fed Aha-labeled *E. coli* (Fig. 2). Control worms fed unlabeled *E. coli* exhibited only background levels of fluorescence.

## FUTURE WORK

Extension of cell-selective labeling to multicellular organisms would constitute an important step forward for proteomic analysis. The preliminary work described here provides a strategy for delivery of non-canonical amino

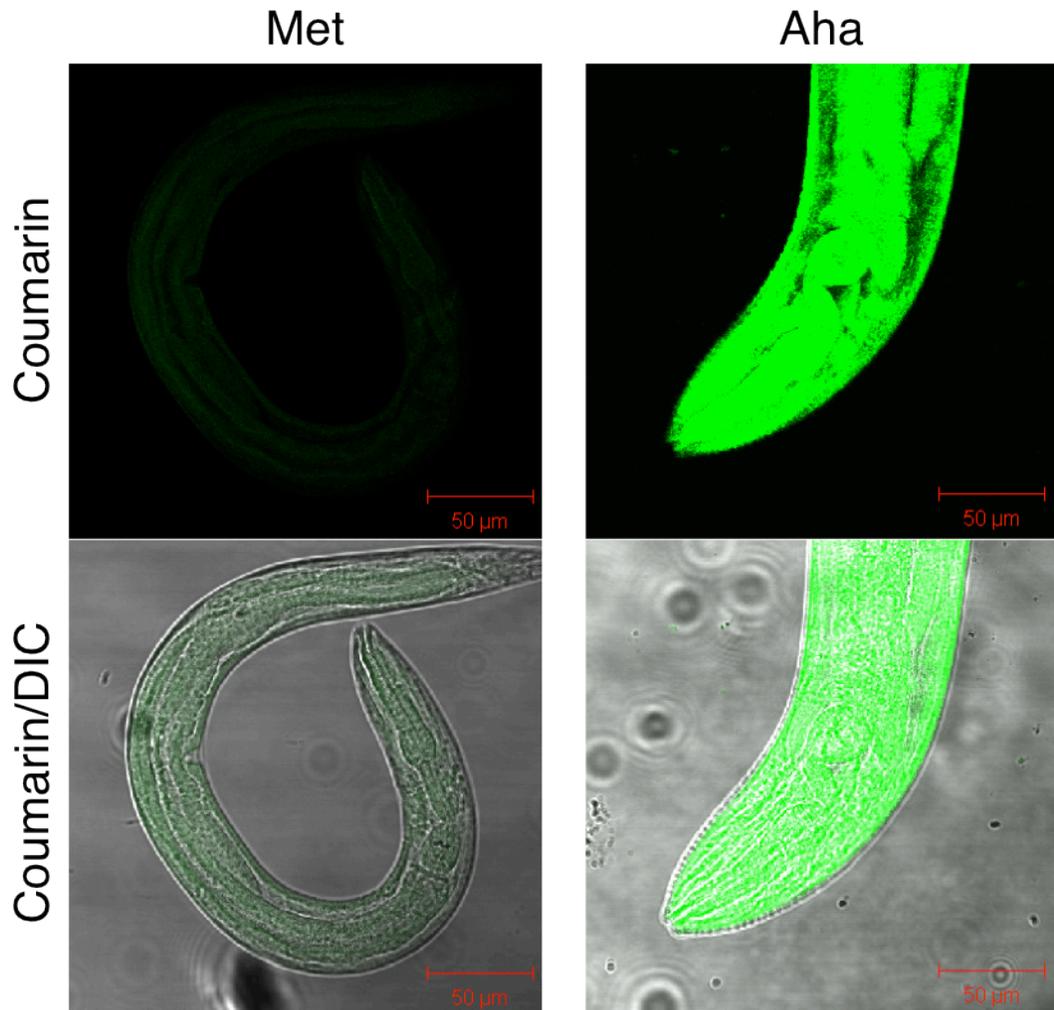
acids to *C. elegans*, as well as validation that non-canonical amino acids can be incorporated into proteins within *C. elegans*. Cell-selective labeling will require generation of transgenic worms expressing NLL-MetRS (or another mutant synthetase) under control by cell-specific drivers. Validation of the proposed model system will establish a powerful platform for cell-selective protein labeling and identification not only in *C. elegans*, but in any number of cell types in various organisms. This approach will allow simplification of proteomic data derived from live animals and permit single-cell dissection of proteome dynamics.

Figure 5.1



**Detection of Aha-tagged *C. elegans* proteins.** (A) Proteins *C. elegans* fed Aha-labeled or Met-labeled *E. coli* cells were extracted and assessed for Aha-incorporation by reaction with alkyne-TAMRA. In-gel fluorescence scanning was used to detect labeled proteins. Numerous proteins from *C. elegans* fed Aha-labeled *E. coli* exhibited TAMRA fluorescence (B) The characteristic 17 kDa biotinylated *E. coli* protein (biotin carboxyl-carrier protein) was not detected the lanes containing to proteins extracted from *C. elegans* fed Aha-labeled or unlabeled *E. coli*.

Figure 5.2



**Visualization of Aha-tagged proteins in *C. elegans*.** *C. elegans* fed Aha-labeled or unlabeled *E. coli* were fixed, permeabilized, and conjugated to alkyne-coumarin. Detection of coumarin by fluorescence microscopy revealed the presence of Aha-containing proteins through worms fed Aha-labeled *E. coli*.

## REFERENCES

- Beatty KE, *et al*, *J Am Chem Soc*, **127**, 14150 – 14151 (2005)
- Beatty KE, *et al*, *Angew Chem Int Ed*, **45**, 7364 – 14151 (2006)
- Chen PR, *et al*, *Angew Chem Int Ed*, **48**, 4052 – 4055 (2009)
- Chin JW, *et al*, *Science*, **301**, 964 – 967 (2003)
- Dieterich DC, *et al*, *Proc Natl Acad Sci USA*, **103**, 9482 – 9487 (2006)
- Epstein HF, *et al*, *J Mol Biol*, **90**, 291 – 300 (1974)
- Hunt-Newbury R, *et al*, *PLoS Biology*, **5**, e237 (2007)
- Jantsch-Plunger V and Fire A, *J Biol Chem*, **269**, 27021 – 27028 (1994)
- Ngo JT, *et al*, *Nat Chem Biol*, **5**, 715 – 717 (2009)
- Sulston JE and Horvitz HR, *Dev Biol*, **56**, 110 – 119 (1977)
- Sulston JE, *et al*, *Dev Biol*, **100**, 64 – 119 (1983)
- Tanrikulu IC, *et al*, *Proc Natl Acad Sci USA*, **106**, 15285 – 15290 (2009)
- Tomancak P, *et al*, *Genome Biol*, **8**, R145 (2007)
- Waterston R, in *The Nematode Caenorhabditis elegans* (Wood WR, ed) pp 281 – 336, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

## CONCLUDING REMARKS

The methods described in this thesis were designed with the objective of providing powerful and adaptable methods for the study of protein synthesis in complex multicellular systems. These strategies combine the exquisite selectivity of bio-orthogonal chemistry with the versatility of genetic control to permit a level of temporal and spatial resolution that is not achieved using traditional techniques. The cell-selective and promoter-driven strategies are currently in use by multiple laboratories for dissection of host-pathogen interactions, bacterial biofilm communities, as well as multicellular organisms. The method for selective N-terminal incorporation is a novel strategy for incorporation of non-canonical amino acids and we anticipate that it will be a valuable tool for both proteomics and protein engineering. A future challenge may include adaptation of these techniques to examine spatial differences in protein translation within individual cells (such as localized protein synthesis within neurons).

