PROBING PROTEIN DYNAMICS *IN VIVO* USING NON-CANONICAL AMINO ACID LABELING

by

Aya Saleh

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Weldon School of Biomedical Engineering

West Lafayette, Indiana

August 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Tamara L. Kinzer-Ursem, Chair
Weldon School of Biomedical Engineering, Purdue University

Dr. Sarah Calve, Co-Chair
Paul M. Rady Department of Mechanical Engineering, University of Colorado Boulder

Dr. Luis Solorio
Weldon School of Biomedical Engineering, Purdue University

Dr. Chittaranjan Das
Department of Chemistry, Purdue University

Approved by:
Dr. George R. Wodicka

To my family and my beloved husband for their unyielding love and support

ACKNOWLEDGMENTS

"I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale."

– Marie Curie

I have been interested in pursuing a career in scientific discovery for as long as I can remember. This very passion is perhaps what helped me the most to survive the inevitable self-doubts of a Ph.D. student. Indeed, completing my doctoral studies has been a challenging journey yet it has also been very rich and rewarding. This journey would definitely not have been possible without the help and support of many people to whom I would like to express my sincere gratitude.

First, I would like to thank my major advisor, Dr. Tamara Kinzer-Ursem, for accepting me in her research group and for mentoring me during my tenure at Purdue. She has been an academic advisor as well as an elder sister and a dear friend. I would also like to thank my co-advisor, Dr. Sarah Calve, for her guidance throughout my graduate work. I am thankful to my committee members, Drs. Chittaranjan Das and Luis Solorio for their valuable insights and for the fruitful collaborations that culminated into scientific publications. I have been fortunate to have collaborated with each of you on different research projects, which broadened my perspectives and sharpened my skill set.

I am grateful to the past and present members of the Kinzer-Ursem and Clave laboratories for helping me with my different projects. Specifically, I would like to thank Tyler VanDyk, Kathryn Jacobson and Naagarajan Narayanan for their invaluable contributions to the work presented in this thesis. I am also thankful to Uma Aryal, Victoria Hedrick, Amber Jannasch and Bruce Cooper from Bindley Bioscience Center for helping me with the design and analysis of the proteomics and metabolomics experiments. I would also like to thank Tran Nguyen for her valuable friendship and for the getaway lunches.

Throughout my whole life I have been very fortunate to have a wonderful family that loves me unconditionally and supports me in every imaginable way. My parents have always been my

role models who instilled in me the value of scientific integrity and taught me to always do my best. No words of gratitude can ever do them justice and I wish that I can continue to make them proud.

Lastly, I would like to express my deepest gratitude to my husband, Ehab, for his enduring love and support, for his faith in me, and for being my rock, my best friend, and my greatest cheerleader.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACN	acetonitrile
AG	aminoguanidine
Aha	azidohomoalanine
CuAAC	copper(I)-catalyzed azide alkyne cycloaddition
DBA	diazo biotin-alkyne
DTT	dithiothreitol
ECM	extracellular matrix
FA	formic acid
FDR	false discovery rate
GO	gene ontology
HCA	hierarchical clustering analysis
HPLC	high performance liquid chromatography
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LFQ	label-free quantification
Met	methionine
MS	mass spectrometry
$Na_2S_2O_4$	sodium dithionite
NaAsc	sodium ascorbate
ncAA	non-canonical amino acid
NSEPs	newly synthesized extracellular matrix proteins
NSPs	newly synthesized proteins
PCA	principle component analysis
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SILAC	stable isotope labeling by amino acids in cell culture
TBS	tris-buffered saline
TFA	trifluoroacetic acid
THPTA	tris(3-hydroxypropyltriazolylmethyl)amine

ABSTRACT

The cellular protein pool exists in a state of dynamic equilibrium, such that a balance between protein synthesis and degradation is maintained to sustain protein homeostasis. This equilibrium is essential for normal cellular functions and hence alteration in protein dynamics has several pathological implications in developing and adult tissues. Recent progress in mass spectrometry (MS) and metabolic labeling techniques has advanced our understanding of the mechanisms of protein regulation in cultured cells and less complicated multicellular organisms. However, methods for the analysis of the dynamics of intra- and extra-cellular proteins in embryonic and adult tissues remain lacking.

To address this gap, we developed a metabolic labeling technique that enables labeling the nascent murine proteome via injection of non-canonical amino acids (ncAAs), which can be selectively enriched by "clickable" tags for identification and quantification. Using this technique, we developed a MS-based method for the selective identification and quantification of the intra- and extra-cellular newly synthesized proteins in developing murine tissues. We then applied this technique to study the dynamic regulation of extracellular matrix (ECM) proteins during embryonic and adolescent musculoskeletal development. We show that the applied technique enables resolving differences in the nascent proteome of different developmental time points with high temporal resolution. The technique can also reveal protein dynamic information that cannot be captured by the traditional proteomic techniques. Additionally, we identified key ECM components that play roles in musculoskeletal development to provide insights into the mechanisms of musculoskeletal tissue regeneration.

To fully characterize our labeling technique, we developed a mathematical model to describe the biodistribution kinetics of azidohomoalanine (Aha), the most widely used ncAA, in murine tissues. The model enabled measuring the relative rates of protein synthesis and turnover in different tissues and predicting the effect of different dosing regimens of Aha on the degree of protein labeling. Finally, we analyzed the plasma metabolome of Aha-injected mice to investigate the impact of Aha incorporation on normal physiology. The analysis revealed that Aha administration into mice does not significantly perturb metabolic functions. Taken together, the findings presented in this dissertation demonstrate the utility of the ncAA labeling technique in mapping protein dynamics in mammalian tissues. This will ultimately have a significant impact on our understanding of protein regulation in health and disease.

1. INTRODUCTION

This Chapter has been published in part in Saleh et al., "Non-canonical amino acid labeling in proteomics and biotechnology," *Journal of Biological Engineering,* vol. 13, p. 43, 2019. Reprinted with permission.

1.1 Protein Dynamics

Cells maintain normal physiological functions by controlling a balance between the rate of protein synthesis and degradation. Dysregulation in protein homeostasis is accordingly implicated in several developmental disorders as well as various diseases, including cancer, diabetes, neurodegeneration, and cardiovascular disorders [1]. Current progress in quantitative mass spectrometry (MS)-based proteomic techniques has provided substantial insights into the intricate pathways that contribute to the pathogenesis of several diseases. However, accurate understanding of the molecular mechanisms of tissue development and disease progression necessitates employing techniques capable of probing the continuous changes in proteins that are synthesized and degraded as a function of development and disease. This goal is rather challenging because it requires developing methods that can selectively identify and quantify new proteins synthesized during a particular time period with minimal interference from pre-existing proteins. In this regard, the following section discusses two protein labeling techniques devised to probe protein dynamics in response to various biological cues: stable isotope labeling and non-canonical amino acid (ncAA) labeling.

1.1.1 Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC)

Stable isotope labeling with amino acids in cell culture (SILAC) is a MS-based approach that was introduced by Mann and coworkers to quantify variations in protein abundance across

different cell populations [2]. In this technique, one cell population is cultured in a medium containing “heavy” isotopic variants of essential amino acids, whereas another cell population that serves as a control is cultured in a medium containing normal or “light” amino acids. Cells are allowed to grow for several population doublings to ensure that the entire proteome is labeled. Afterwards, the two samples are combined, and the total proteome is extracted, digested, and analyzed using MS. Accordingly, the origin of the resulting peptides can be identified on the basis of the mass difference between the “light” and “heavy” species of each peptide, and the relative abundance of the corresponding proteins is determined based on the peak intensities of the two forms of the peptide.

Selbach and coworkers further modified SILAC to enable comparing the abundance of newly synthesized proteins between two different treatments [3]. The modified method, known as pulsed stable-isotope labeling with amino acids in cell culture (pSILAC), involves initial cultivation of two cell populations in a medium containing light amino acids. One cell population is then switched to a medium containing only “medium” amino acids, while the other population is transferred to a medium containing only “heavy” amino acids. After several cycles of cell divisions, all newly synthesized proteins will be metabolically labeled with either the “medium” or the “heavy” amino acids, whereas old proteins will remain unlabeled. Similar to SILAC, the origin and relative abundance of proteins can be determined based on their mass difference and the ratio of their peak intensities, respectively [3]. pSILAC has been subsequently applied to study protein dynamics in mammalian cell lines [3], nematodes [4], zebrafish [5], as well as mice [6].

Despite the wide utility of isotope labeling, studies in mice have shown that feeding animals with diet containing isotope-labeled amino acids for prolonged periods is needed to achieve complete proteome labeling [7, 8]. Such long labeling time limits the use of isotope-labeled mice

to study proteins with high turnover rates. Moreover, obtaining accurate quantitative data in temporal proteomic studies requires that the method be able to analyze newly synthesized proteins with minimal interference from the more abundant pre-existing proteins. Nevertheless, the co-elution of pre-existing unlabeled and new isotope-labeled peptides precludes the identification of low abundant peptides by MS [9]. This ultimately limits the use of isotope labeling in disease models due to the overlooking of low abundant newly synthesized proteins, which renders interpretations of the molecular determinants of the studied diseases inaccurate. As such, the importance of developing protein labeling techniques that enable selective enrichment of newly synthesized proteins in live mammals cannot be overstated.

1.1.2 Non-Canonical Amino Acid (ncAA) Labeling

To address the insensitivity issue of isotope labeling for newly synthesized proteins, Dieterich et al. introduced the ncAA labeling strategy for selective analysis of *de novo* protein synthesis in mammalian cells [10]. In this method, a ncAA carrying a desired functional group is introduced to the host expression system and is incorporated onto an aminoacyl tRNA synthetase (aaRS) that covalently attaches the ncAA to the corresponding tRNA (Figure 1.1A). The ncAA-tRNA complex is brought into the ribosome where the tRNA recognizes the appropriate mRNA codon sequence and the ncAA is added to the growing polypeptide chain (Figure 1.1B). ncAA labeling can be designed to occur at specific amino acid residues of interest, for example using a methionine (Met) analog that carries an azide (*e.g.* azidohomoalanine, Aha; Figure 1.2) or alkyne functionality (*e.g.* homopropargylglycine, Hpg; Figure 1.2) to replace any Met in a newly synthesized protein [11]. As a result, new proteins synthesized during pulse labeling with the ncAA are tagged with azide or alkyne functional groups [10].

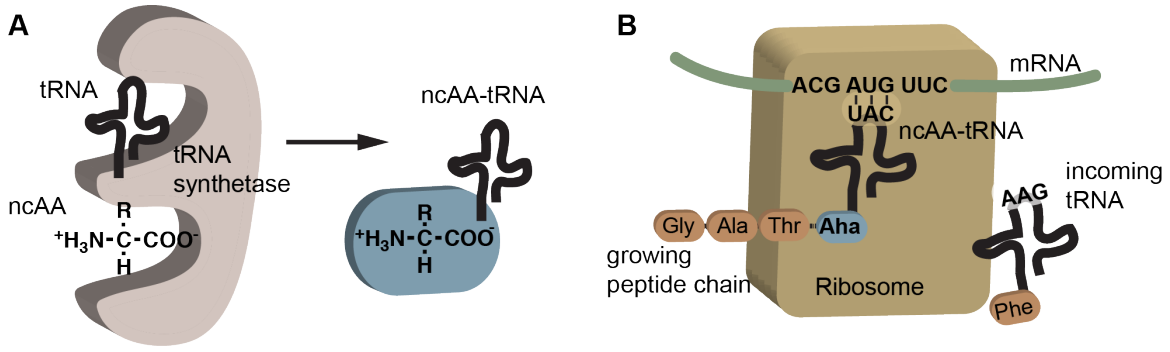


Figure 1.1 Incorporation of non-canonical amino acids (ncAAs) by native cellular machinery. ncAAs are incorporated into the growing polypeptide chain as the protein is synthesized at the ribosome. **(A)** ncAA is attached to a tRNA by amino acyl tRNA synthetase. **(B)** The tRNA, charged with the ncAA (ncAA-tRNA, ncAA in blue), recognizes mRNA codons in the ribosome and the ncAA is added to the growing polypeptide chain. Reprinted with permission from [12].

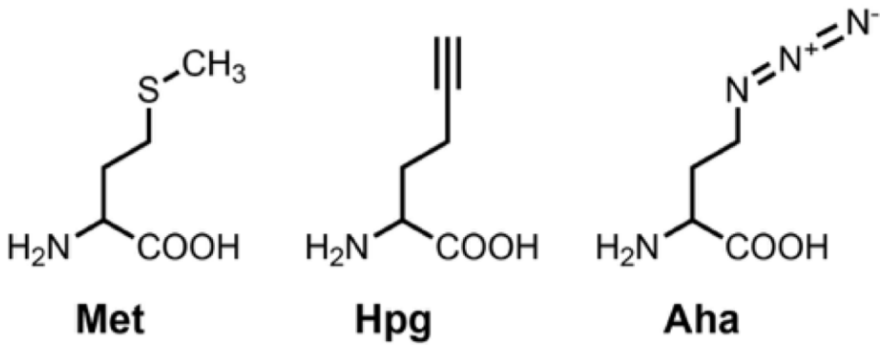


Figure 1.2 Chemical structures of methionine and its analogs. Methionine (Met), homopropargylglycine (Hpg), azidohomoalanine (Aha). Adapted with permission from [12].

The kinetics of Aha and Hpg binding to the methionyl tRNA synthetase (MetRS) are slower than that of Met (k_{cat}/K_M of 1.42×10^{-3} and 1.16×10^{-3} $s^{-1} \cdot \mu M^{-1}$ for Aha and Hpg, respectively vs 5.47×10^{-1} $s^{-1} \cdot \mu M^{-1}$ for Met) [11]. Nonetheless, this is a straightforward labeling method with no need for genetic engineering of the protein or organism under study (Figure 1.3). For applications where 100% Met substitution is not necessary (*e.g.* enrichment for proteomics), adding the ncAA at concentrations where it can outcompete with Met provides sufficient functional incorporation. Alternatives that increase ncAA incorporation include using Met auxotrophic strains of *E. coli* that

cannot make their own Met [13], or Met-free media for mammalian cell culture. Orthogonal aaRSs have also been engineered to bind to ncAAs in cells expressing the mutant aaRS, allowing for protein labeling with ncAAs in specific cell types [14-16].

More importantly, azides and alkynes are bio-orthogonal chemical groups, denoting that they are not available in native biological systems and hence can be selectively modified through specific chemical reactions [17]. The most prominent azide reaction is the "click" chemistry reaction by which azides undergo copper(I)-catalyzed [3+2] cycloaddition (CuAAC) reaction with terminal alkynes to form triazole products (Figure 1.4A) [18]. Accordingly, using click chemistry, azide- and alkyne-labeled proteins can be selectively conjugated to affinity tags even within complex cellular or tissue lysates to enrich the newly synthesized proteins from a pool of pre-existing unlabeled proteins. Additionally, labeled proteins can be ligated to fluorescent dyes for protein visualization [19].

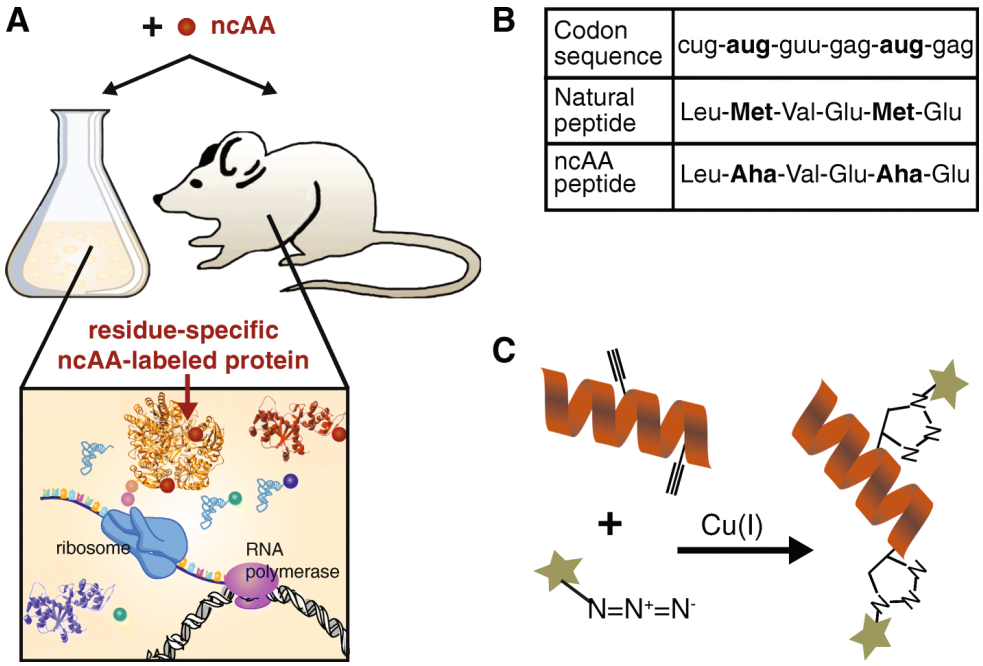


Figure 1.3 Overview of residue-specific ncAA protein labeling. **(A)** A ncAA (red sphere) is added to the system (cell culture or animal model). Newly synthesized proteins will have incorporated the ncAA. **(B)** An example of the codon sequence and corresponding peptides that result from either natural synthesis or synthesis in the presence of the ncAA. **(C)** A peptide labeled at two residue-specific sites with a ncAA carrying an alkyne functional group is conjugated to an azide-containing fluorophore via CuAAC. Reprinted with permission from [12].

Click chemistry functionality

First coined by Sharpless and colleagues in 2001, click chemistries are a set of chemical reactions that are readily catalyzed in aqueous solutions at atmospheric pressure and biologically-compatible temperatures, with few toxic intermediates, and relatively fast reaction kinetics [20]. The suite of specific click chemistry reactions that started with Staudinger ligation of azide and phosphine [11, 21, 22] and copper-catalyzed azide-alkyne cycloaddition [18, 23], has rapidly expanded to include more rapid and biologically friendly chemistries including strain promoted azide-alkyne cycloaddition [24, 25], oxime or hydrazine ligation [26, 27], strain-promoted alkyne nitrone cycloaddition [28, 29], tetrazine ligation [30, 31], and quadricyclane ligation [32, 33].

Azide-alkyne cycloaddition is one of the most widely used click reactions with broad availability of commercial reagents and moderately fast kinetics. CuAAC has been implemented across disciplines, from biomaterials [34] and combinatorial chemistry [35] to polymer synthesis [36], protein activity tagging [37], and proteomics [10]. One disadvantage of CuAAC is that there is significant cytotoxicity with using copper as the catalyst, hampering utilization *in vivo* [38]. To circumvent this limitation, Bertozzi and coworkers introduced a catalyst-free [3+2] cycloaddition reaction between azides and cyclooctyne derivatives, known as strain-promoted azide-alkyne cycloaddition (SPAAC; Figure 1.4B) [24]. Cyclooctynes are alkynes in ring structures that, due to strong ring strain, are highly reactive and react readily with azides in the absence of promoting agents. The biocompatibility of this reaction was first demonstrated in Jurkat cells to label azide-tagged glycoproteins [24]. The strain-promoted azide-alkyne click reaction has since been applied in various *in vivo* settings with no apparent toxicity [19, 39, 40]. Importantly, CuAAC and SPAAC are bio-orthogonal and will not interfere with natural biological chemistries.

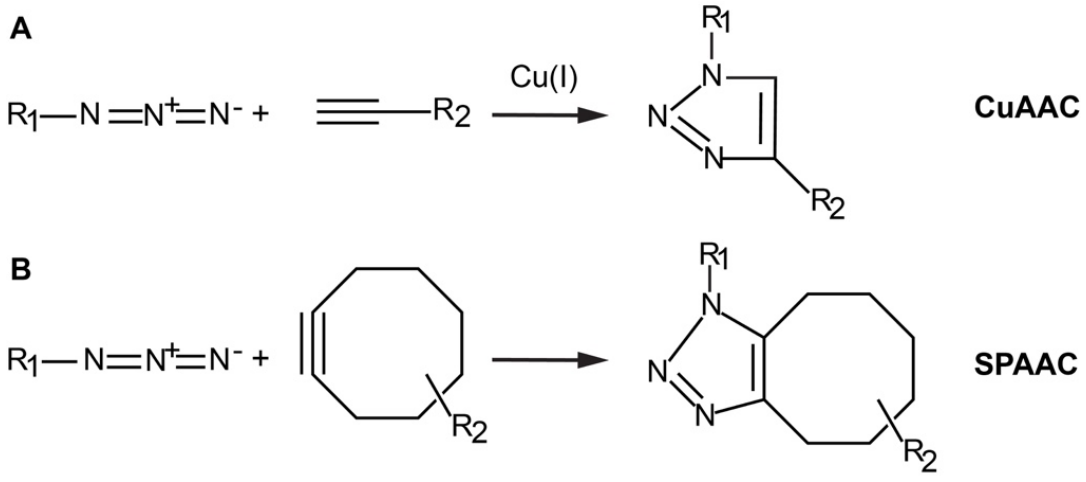


Figure 1.4 Azide-alkyne cycloaddition reactions. **(A)** Copper(I)-catalyzed [3 + 2] azide-alkyne cycloaddition (CuAAC). **(B)** Strain-promoted [3 + 2] azide-alkyne cycloaddition (SPAAC). Reprinted with permission from [12].

Applications of ncAA labeling

Over the last decade, ncAA labeling has gained wide recognition due to its capability of tracking continuous changes in protein expression. It has been applied in various mammalian cell culture systems to study protein acylation [41], lysosomal protein degradation [42], and inflammation [43]. The method has been also used to study protein dynamics in various bacterial systems in order to explore quorum sensing [9], identify bacterial virulence factors [44], and monitor bacterial degradation in phagocytes [45].

In a seminal study, Eichelbaum et al. utilized ncAA labeling in addition to SILAC to study the secretome of different human cell lines, where cells were pulsed with isotope-labeled amino acids and Aha for 24 h [46]. In this study, the performance of dual labeling with non-canonical and isotope-labeled amino acids was compared to labeling with isotope-labeled amino acids only by analyzing the secreted proteins with and without an enrichment step for ncAA-labeled proteins. Without enrichment, the number of quantified proteins decreased from 684 proteins with high correlation between biological replicates ($R = 0.96$) to only 22 proteins with low correlation ($R = 0.02$).

Similarly, Bagert et al. compared the number of quantified proteins from combining ncAA labeling and SILAC to SILAC alone after 4 h and 30 min labeling of HeLa cells [47]. In the ncAA-SILAC experiment, 1931 and 1484 newly synthesized proteins were quantified within the 4 h and 30 min time spans, respectively. Using SILAC alone, the number of quantified proteins was reduced to 589 and 9 proteins for the 4 h and 30 min pulse times, respectively. Together, these studies demonstrate that ncAA labeling can be applied to resolve the small fraction of new proteins synthesized within narrow time intervals and signify the capability of the technique for enriching nascent proteins in a complex background of unlabeled proteins with high temporal resolution.

Deciphering the complexity of the neuronal proteome is another application where ncAA labeling has proven effective. The high complexity of the nervous system as well as the presence of multiple intricate protein networks pose a great challenge for studies aiming at understanding the changes in protein regulation that lead to a perturbed brain activity. In this context, Hodas et al. applied ncAA labeling to identify newly translated proteins following activating dopamine receptors in Aha-labeled hippocampal slices [48]. Further, Kenney et al. combined ncAA and stable isotope labeling to examine the effect of eukaryotic elongation factor 2 kinase (eEF2K) inhibition in primary cortical neurons [49]. In their study, neurons were labeled with Aha for enrichment and with isotopologs of amino acids for absolute quantitation. Their analysis revealed that microtubule-related proteins are specifically sensitive to eFK2K inhibition, signifying the importance of eEF2K in neuronal function.

ncAA labeling for visualization of newly synthesized proteins

At present, methods used for visualizing protein dynamics are commonly based on fusing proteins of interest to fluorescent proteins such as the green fluorescent protein (GFP). While the use of fluorescent proteins has greatly advanced our understanding of the spatial dynamics of proteins in living cells [50], the method has considerable limitations. First, fusing proteins to high molecular weight tags (GFP~27kD) can alter their localization and activity [51]. Additionally, the method relies on ectopic protein expression, thus does not mirror the native cellular milieu [52]. Moreover, this method is restricted to a single or few proteins of interest. As such, ncAA labeling was proposed as a method that can be leveraged to study the spatial dynamics of proteins without the aforementioned limitations. Taking advantage of the small size of “clickable” fluorophore molecules, Dieterich et al. studied the dynamics of neuronal proteins by pulse labeling rat hippocampal neurons with Aha and Hpg for different time intervals [19]. Using alkyne- and azide-

bearing fluorophores, newly synthesized proteins were visualized in the cell bodies of the neurons after only 10 min pulse labeling interval. Increasing the labeling time to 20 min was enough to detect newly synthesized proteins in dendrites, whereas detecting the labeled proteins in dendritic spines required 2 h. The study was further extended to investigate the efficacy of ncAA visualization in intact tissues using hippocampal slices, in which labeled proteins were visualized across the whole tissue following 4 h labeling. These time-dependent experiments in neurons revealed that the method is sensitive in time scales as short as 10 min, and that it can be used to track the subcellular distribution of newly synthesized proteins.

To study the turnover rates of synaptic proteins, Cohen et al. employed ncAA labeling in addition to SILAC to estimate the life-spans of hundreds of synaptic proteins [53]. Using fluorescent ncAA tagging, the fluorescence reduction rates of labeled synaptic proteins were measured to approximately estimate their turnover rates, whereas SILAC was used for accurate MS-based determination of protein half-lives. The estimated half-lives of the identified synaptic proteins were unexpectedly in the range of 2-5 days with no significant difference between the turnover rates of presynaptic and postsynaptic proteins. Collectively, the aforementioned studies suggest that the adaptation of fluorescent visualization of ncAA-labeled proteins in future studies would enable monitoring protein localization and tracking the cellular fate of newly synthesized proteins in normal physiology and in response to various perturbations.

In vivo ncAA labeling

Given the intricacy of human proteome, extending the studies to more complex animals such as rodents is essential to augment our understanding of the dynamics of protein synthesis in response to various stimuli. Until recently, it was presumed that ncAA labeling cannot be applied

for *in vivo* labeling of mammals proteome because mammalian cells would favor incorporating endogenous Met, rather than its analogs, into newly expressed proteins [54]. However, Schiapparelli et al. successfully labeled newly synthesized proteins in the retina of adult rats by intraocular injection of Aha [55]. Further, McClatchy et al. demonstrated that *in vivo* labeling of the murine proteome is feasible by feeding animals for several days an Aha-enriched diet that lacks Met [56]. More recently, our group developed an injection-based technique that enables global labeling of murine proteome with ncAAs [57]. Our results demonstrate that two days of intraperitoneal injection of Aha and Hpg is sufficient for systemic incorporation of the Met analogs into the proteome of both juvenile mice and developing embryos. In this study, neither perturbation of the physiological function of the injected mice nor atypical embryonic development was observed. In addition, both Aha and Hpg were successfully incorporated into different murine tissues in a concentration-dependent manner. Notably, labeling with Hpg was less efficient than Aha, which is in agreement with findings by Kiick et al. that the activation rate of Hpg by methionyl tRNA synthetase is slower than Aha [11]. In light of these results, the successful incorporation of Met analogs into the entire murine proteome through injection will pave the road for using animal models to temporally and spatially probe protein dynamics in both adult and developing tissues.

1.2 Problem Statement and Research Scope

Probing protein dynamics *in vivo* entails developing methods capable of measuring new protein synthesis and degradation in complex organisms. In essence, early alterations in protein expression are usually associated with low-level translational changes. Therefore, these methods should also be able to selectively label and capture the small fraction of nascent proteome synthesized during specified time intervals in a pool of pre-existing proteins. This goal is limited

by the shortcomings of the currently available *in vivo* protein labeling techniques. One established technique relies on feeding animals with isotope-labeled diet for prolonged periods. These long labeling times typically preclude the identification of proteins with high turnover rates. Additionally, these methods do not allow for selective enrichment of the labeled proteins and hence impede detecting those in low abundance. This ultimately hinders the use of isotope labeling in disease models due to the overlooking of low abundant newly synthesized proteins, rendering interpretations of the molecular determinants of the studied disease inaccurate.

Another technique is based on nascent protein labeling with ncAAs that can be selectively enriched via click chemistry. McClatchy et al. showed that labeling of the adult murine proteome can be achieved by feeding mice Met-free, Aha-enriched diet for 4-6 days [56]. Despite the usefulness of this method, it has two main shortcomings. First, similar to isotope labeling, the relatively long labeling period impedes detecting short-lived proteins. Second, several studies have signified the importance of Met for normal embryonic development [58-60], rendering the method limited to adult tissues.

To overcome the limitations of the current techniques, our group developed a technique that enables protein labeling in both adult and embryonic tissues based on systemic injection of ncAAs without the need for Met depletion [57]. This method provides several advantages over introducing the ncAAs in animal diet because injection (1) is relatively easy to perform, (2) achieves global proteome labeling in a shorter duration, (3) enables more accurate dose-effect calculations, thus enhancing experimental reproducibility, and (4) can be applied for adult and developing tissues.

The first report on this technique [57] demonstrated the ability of the method to label adult and embryonic tissues using in-gel fluorescence. However, the application of this technique to

identify the newly synthesized intra- and extra-cellular proteins using MS is yet to be demonstrated. In addition, the kinetics of ncAA biodistribution in murine tissues and the feasibility of using this technique for measuring the relative protein synthesis and turnover rates *in vivo* remain to be investigated. In this vein, the scope of the research presented in this dissertation is to:

1- Develop a MS-based method for the selective enrichment, identification and quantification of the intra- and extra-cellular newly synthesized proteins in the developing mouse.
2- Apply the technique to study the dynamics of forelimb development in embryonic and adolescent murine tissues.
3- Quantitate the biodistribution kinetics of Aha in the adult mouse as it relates to the relative rate of protein synthesis and turnover.
4- Investigate the impact of systemic Aha injection on the normal physiology of the adult mouse.

1.3 Dissertation Overview

The remainder of this dissertation is organized as follows: **Chapter 2** describes the development of a MS-based method for selective enrichment of ncAA-labeled intracellular proteins in murine embryos with minimum background. The developed technique was then applied to resolve differences in the proteome of different embryonic time points. In addition, **Chapter 2** describes the utilization of a tissue fractionation protocol to isolate cytosolic, nuclear, membrane, cytoskeleton and extracellular matrix (ECM) embryonic proteins. The ncAA labeling and tissue fractionation techniques were then combined to examine the dynamics of protein synthesis and turnover in the intra- and extracellular embryonic compartments. In **Chapter 3,** the MS-based

ncAA enrichment method developed in **Chapter 2** was expanded to enrich for the newly synthesized ECM proteins by implementing modifications to overcome the complex biochemical nature of the ECM. **Chapter 3** also describes the application of the modified method to study the dynamic regulation of ECM proteins in embryonic and adolescent forelimbs to provide insights into the mechanisms of musculoskeletal tissue regeneration. **Chapter 4** describes a kinetic analysis of the concentration of the ncAA Aha in plasma and selected tissues following a single injection in adult mice. The concentration profile was then correlated to the degree of tissue labeling over time as measured by fluorescent western blotting. A mathematical model that describes the kinetics of free Aha and Aha-labeled proteins was then developed based on the experimental data to enable estimating the relative protein synthesis and turnover rates as well as predicting the change in protein labeling using different Aha dosing regimens. In addition, **Chapter 4** summarizes the findings of an investigation of the effect of Aha administration on normal physiological functions by comparing the plasma metabolite profile of Aha-injected mice to non-injected controls. Finally, **Chapter 5** summarizes the main research findings and the implications of the work presented in the dissertation, and discusses recommendations for future work in the field of ncAA labeling *in vivo*.

1.4 References

[1] K. Dasuri, L. Zhang, and J. N. Keller, "Oxidative stress, neurodegeneration, and the balance of protein degradation and protein synthesis," in *Free Radical Biology and Medicine* vol. 62, ed, 2013, pp. 170-185.

[2] S. E. Ong *et al.*, "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.," *Molecular & cellular proteomics : MCP,* vol. 1, pp. 376-386, 2002.

[3] B. Schwanhäusser, M. Gossen, G. Dittmar, and M. Selbach, "Global analysis of cellular protein translation by pulsed SILAC," *Proteomics,* vol. 9, pp. 205-209, 2009.

[4] M. Larance *et al.*, "Stable-isotope labeling with amino acids in nematodes," *Nature Methods,* vol. 8, pp. 849-851, 2011.

[5] H. Nolte *et al.*, "Dynamics of zebrafish fin regeneration using a pulsed SILAC approach," *Proteomics,* vol. 15, pp. 739-751, 2015.
[6] J. M. Basak *et al.*, "Measurement of apolipoprotein e and amyloid β clearance rates in the mouse brain using bolus stable isotope labeling," *Molecular Neurodegeneration,* vol. 7, p. 14, 2012.
[7] D. B. McClatchy, M. Q. Dong, C. C. Wu, J. D. Venable, and J. R. Yates, "15N metabolic labeling of mammalian tissue with slow protein turnover," *Journal of Proteome Research,* vol. 6, pp. 2005-2010, 2007.
[8] M. Krüger *et al.*, "SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function," *Cell,* vol. 134, pp. 353-364, 2008.
[9] J. D. Bagert *et al.*, "Time-resolved proteomic analysis of quorum sensing in Vibrio harveyi," *Chemical Science,* vol. 7, pp. 1797-1806, 2016.
[10] D. C. Dieterich, A. J. Link, J. Graumann, D. A. Tirrell, and E. M. Schuman, "Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT)," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 103, pp. 9482-9487, 2006.
[11] K. L. Kiick, E. Saxon, D. A. Tirrell, and C. R. Bertozzi, "Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, pp. 19-24, 2002.
[12] A. M. Saleh, K. M. Wilding, S. Calve, B. C. Bundy, and T. L. Kinzer-Ursem, "Non-canonical amino acid labeling in proteomics and biotechnology," *Journal of Biological Engineering,* vol. 13, p. 43, 2019.
[13] J. C. M. Van Hest and D. A. Tirrell, "Efficient introduction of alkene functionality into proteins in vivo," *FEBS Letters,* vol. 428, pp. 68-70, 1998.
[14] F. Truong, T. H. Yoo, T. J. Lampo, and D. A. Tirrell, "Two-strain, cell-selective protein labeling in mixed bacterial cultures," *Journal of the American Chemical Society,* vol. 134, pp. 8551-8556, 2012.
[15] K. P. Yuet *et al.*, "Cell-specific proteomic analysis in Caenorhabditis elegans," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 112, pp. 2705-2710, 2015.
[16] A. Mahdavi *et al.*, "Engineered Aminoacyl-tRNA Synthetase for Cell-Selective Analysis of Mammalian Protein Synthesis," *Journal of the American Chemical Society,* vol. 138, pp. 4278-4281, 2016.
[17] A. Borrmann and J. C. M. Van Hest, "Bioorthogonal chemistry in living organisms," *Chemical Science,* vol. 5, pp. 2123-2134, 2014.
[18] V. V. Rostovtsev, L. G. Green, V. V. Fokin, and K. B. Sharpless, "A stepwise huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes," *Angewandte Chemie - International Edition,* vol. 41, pp. 2596-2599, 2002.
[19] D. C. Dieterich *et al.*, "In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons," *Nature Neuroscience,* vol. 13, pp. 897-905, 2010.
[20] H. C. Kolb, M. G. Finn, and K. B. Sharpless, "Click Chemistry: Diverse Chemical Function from a Few Good Reactions," *Angewandte Chemie International Edition,* vol. 40, pp. 2004-2021, 2001.
[21] E. Saxon and C. R. Bertozzi, "Cell surface engineering by a modified Staudinger reaction," *Science,* vol. 287, pp. 2007-2010, 2000.

[22] M. L. Tsao, F. Tian, and P. G. Schultz, "Selective staudinger modification of proteins containing p-azidophenyl-alanine," *ChemBioChem,* vol. 6, pp. 2147-2149, 2005.
[23] Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless, and M. G. Finn, "Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition," *Journal of the American Chemical Society,* vol. 125, pp. 3192-3193, 2003.
[24] N. J. Agard, J. A. Prescher, and C. R. Bertozzi, "A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems," *Journal of the American Chemical Society,* vol. 126, pp. 15046-15047, 2004.
[25] N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo, and C. R. Bertozzi, "A comparative study of bioorthogonal reactions with azides.," *ACS chemical biology,* vol. 1, pp. 644-648, 2006.
[26] M. M. Mahmoodi, M. Rashidian, J. K. Dozier, and M. D. Distefano, "Chemoenzymatic site-specific reversible immobilization and labeling of proteins from crude cellular extract without prior purification using oxime and hydrazine ligation," *Current protocols in chemical biology,* vol. 5, pp. 89-109, 2013.
[27] M. Rashidian, J. K. Dozier, and M. D. Distefano, "Enzymatic labeling of proteins: Techniques and approaches," *Bioconjugate Chemistry,* vol. 24, pp. 1277-1294, 2013.
[28] D. A. MacKenzie, A. R. Sherratt, M. Chigrinova, L. L. W. Cheung, and J. P. Pezacki, "Strain-promoted cycloadditions involving nitrones and alkynes-rapid tunable reactions for bioorthogonal labeling," *Current Opinion in Chemical Biology,* vol. 21, pp. 81-88, 2014.
[29] A. R. Sherratt *et al.*, "Dual Strain-Promoted Alkyne-Nitrone Cycloadditions for Simultaneous Labeling of Bacterial Peptidoglycans," *Bioconjugate Chemistry,* vol. 27, pp. 1222-1226, 2016.
[30] M. L. Blackman, M. Royzen, and J. M. Fox, "Tetrazine ligation: Fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity," *Journal of the American Chemical Society,* vol. 130, pp. 13518-13519, 2008.
[31] B. L. Oliveira, Z. Guo, and G. J. L. Bernardes, "Inverse electron demand Diels-Alder reactions in chemical biology," *Chemical Society Reviews,* vol. 46, pp. 4895-4950, 2017.
[32] E. M. Sletten and C. R. Bertozzi, "A bioorthogonal quadricyclane ligation," *Journal of the American Chemical Society,* vol. 133, pp. 17570-17573, 2011.
[33] F. M. Tomlin, C. G. Gordon, Y. Han, T. S. Wu, E. M. Sletten, and C. R. Bertozzi, "Site-specific incorporation of quadricyclane into a protein and photocleavage of the quadricyclane ligation adduct," *Bioorganic and Medicinal Chemistry,* vol. 26, pp. 5280-5290, 2018.
[34] J. C. M. van Hest and D. A. Tirrell, "Protein-based materials, toward a new level of structural control," *Chemical Communications,* vol. 19, pp. 1897-1904, 2001.
[35] H. C. Kolb and K. B. Sharpless, "The growing impact of click chemistry on drug discovery," *Drug Discovery Today,* vol. 8, pp. 1128-1137, 2003.
[36] J. A. Johnson *et al.*, "Core-clickable PEG-branch-azide bivalent-bottle-brush polymers by ROMP: Grafting-through and clicking-to," *Journal of the American Chemical Society,* vol. 133, pp. 559-566, 2011.
[37] A. E. Speers, G. C. Adam, and B. F. Cravatt, "Activity-based protein profiling in vivo using a copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition," *Journal of the American Chemical Society,* vol. 125, pp. 4686-4687, 2003.
[38] J. C. Jewett, E. M. Sletten, and C. R. Bertozzi, "Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones," *Journal of the American Chemical Society,* vol. 132, pp. 3688-3690, 2010.

[39] K. E. Beatty *et al.*, "Live-cell imaging of cellular proteins by a strain-promoted azide-alkyne cycloaddition," *ChemBioChem,* vol. 11, pp. 2092-2095, 2010.
[40] S. T. Laughlin, J. M. Baskin, S. L. Amacher, and C. R. Bertozzi, "In vivo imaging of membrane-associated glycans in developing zebrafish," *Science,* vol. 320, pp. 664-667, 2008.
[41] M. M. Zhang, L. K. Tsou, G. Charron, A. S. Raghavan, and H. C. Hang, "Tandem fluorescence imaging of dynamic S-acylation and protein turnover," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 107, pp. 8627-8632, 2010.
[42] J. Zhang, J. Wang, S. Ng, Q. Lin, and H. M. Shen, "Development of a novel method for quantification of autophagic protein degradation by AHA labeling," *Autophagy,* vol. 10, pp. 901-912, 2014.
[43] K. Y. Choi, D. N. Lippert, P. Ezzatti, and N. Mookherjee, "Defining TNF-alpha and IL-1beta induced nascent proteins: combining bio-orthogonal non-canonical amino acid tagging and proteomics," *J Immunol Methods,* vol. 382, pp. 189-195, 2012.
[44] A. Mahdavi *et al.*, "Identification of secreted bacterial proteins by noncanonical amino acid tagging," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 111, pp. 433-438, 2014.
[45] D. M. Van Elsland, E. Bos, W. De Boer, H. S. Overkleeft, A. J. Koster, and S. I. Van Kasteren, "Detection of bioorthogonal groups by correlative light and electron microscopy allows imaging of degraded bacteria in phagocytes," *Chemical Science,* vol. 7, pp. 752-758, 2016.
[46] K. Eichelbaum, M. Winter, M. B. Diaz, S. Herzig, and J. Krijgsveld, "Selective enrichment of newly synthesized proteins for quantitative secretome analysis," *Nature Biotechnology,* vol. 30, pp. 984-990, 2012.
[47] J. D. Bagert *et al.*, "Quantitative, time-resolved proteomic analysis by combining bioorthogonal noncanonical amino acid tagging and pulsed stable isotope labeling by amino acids in cell culture," *Molecular and Cellular Proteomics,* vol. 13, pp. 1352-1358, 2014.
[48] J. J. L. Hodas *et al.*, "Dopaminergic modulation of the hippocampal neuropil proteome identified by bioorthogonal noncanonical amino acid tagging (BONCAT)," *Proteomics,* vol. 12, pp. 2464-2476, 2012.
[49] J. W. Kenney, M. Genheden, K. M. Moon, X. Wang, L. J. Foster, and C. G. Proud, "Eukaryotic elongation factor 2 kinase regulates the synthesis of microtubule-related proteins in neurons," *Journal of Neurochemistry,* vol. 136, pp. 276-284, 2016.
[50] D. M. Chudakov, M. V. Matz, S. Lukyanov, and K. A. Lukyanov, "Fluorescent proteins and their applications in imaging living cells and tissues," *Physiological Reviews,* vol. 90, pp. 1103-1163, 2010.
[51] T. Misteli and D. L. Spector, "Applications of The Green Fluorescent Protein In Cell Biology and Biotechnology," *Nature Biotechnology,* vol. 15, pp. 961-964, 1997.
[52] B. N. G. Giepmans, S. R. Adams, M. H. Ellisman, and R. Y. Tsien, "The fluorescent toolbox for assessing protein location and function," in *Science* vol. 312, ed, 2006, pp. 217-224.
[53] L. D. Cohen *et al.*, "Metabolic Turnover of Synaptic Proteins: Kinetics, Interdependencies and Implications for Synaptic Maintenance," *PLoS ONE,* vol. 8, p. e63191, 2013.

[54] J. Liu, Y. Xu, D. Stoleru, and A. Salic, "Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 109, pp. 413-418, 2012.

[55] L. M. Schiapparelli, D. B. McClatchy, H. H. Liu, P. Sharma, J. R. Yates, and H. T. Cline, "Direct detection of biotinylated proteins by mass spectrometry," *Journal of Proteome Research,* vol. 13, pp. 3966-3978, 2014.

[56] D. B. McClatchy *et al.*, "Pulsed azidohomoalanine labeling in mammals (PALM) detects changes in liver-specific LKB1 knockout mice," *Journal of Proteome Research,* vol. 14, pp. 4815-4822, 2015.

[57] S. Calve, A. J. Witten, A. R. Ocken, and T. L. Kinzer-Ursem, "Incorporation of non-canonical amino acids into the developing murine proteome," *Scientific Reports,* vol. 6, pp. 1-7, 2016.

[58] S. Ikeda, M. Sugimoto, and S. Kume, "Importance of methionine metabolism in morula-to-blastocyst transition in bovine preimplantation embryos," *Journal of Reproduction and Development,* vol. 58, pp. 91-97, 2012.

[59] M. Kudo, S. Ikeda, M. Sugimoto, and S. Kume, "Methionine-dependent histone methylation at developmentally important gene loci in mouse preimplantation embryos," *Journal of Nutritional Biochemistry,* vol. 26, pp. 1664-1669, 2015.

[60] S. Tang *et al.*, " Methionine metabolism is essential for SIRT 1-regulated mouse embryonic stem cell maintenance and embryonic development " *The EMBO Journal,* vol. 36, pp. 3175-3193, 2017.

2. DYNAMICS OF NON-CANONICAL AMINO ACID-LABELED INTRA- AND EXTRACELLULAR PROTEINS IN THE DEVELOPING MOUSE

This Chapter has been published in whole in Saleh et al., "Dynamics of non-canonical amino acid-labeled intra- and extracellular proteins in the developing mouse," *Cell Mol Bioeng*, vol. 12, no. 5, pp. 495-509, Oct 2019. Reprinted with permission.

2.1 Introduction

During functional tissue assembly, temporal activation of intracellular protein signaling pathways drive cell division, motility and differentiation, and the surrounding extracellular matrix (ECM) is remodeled to provide cues and support for these cellular processes. However, little is known regarding the synthesis and turnover of proteins in different cellular compartments during the scar-free formation of tissues (*i.e.* during development, repair and regeneration). This information is critical for providing benchmarks for regenerative medicine; only by identifying the dynamics of intra- and extracellular proteins that drive native tissue assembly can therapies that better restore functionality to damaged tissues be designed.

Developmental model systems are ideal for studying the role of proteins in the *de novo* formation of tissues. Liquid chromatography–tandem mass spectrometry (LC-MS/MS), which enables the identification of proteins within complex samples, has been used to investigate development in *D. melanogaster* [1], *X. laevis* [2] and *D. rerio* [3, 4], as well as single time point studies of rat embryonic heart [5] and embryonic murine tissue [6]. More recently, a comprehensive study comparing the proteome and transcriptome of the developing murine stomach was conducted [7]; but these investigations typically focused on unfractionated tissue lysates, which consequently leads to the domination of the MS spectra by proteins that are in higher abundance. This is a critical bottleneck in identifying proteins within developing tissues since

proteins of varying solubility are spread across different cellular compartments (*i.e.* cytosolic, membrane, nuclear, cytoskeletal and matrisome or ECM), and many will not be identified using a single extraction method. To increase the coverage of proteins within complex tissues, buffers of different ionic strength and detergent composition can be used to selectively fractionate proteins. Researchers have used these fractionation methods to characterize the composition and turnover of different cellular compartments of various healthy and pathological adult tissues [8-14]; however, the dynamics of both intra- and extracellular proteins during mammalian embryogenesis remain largely unknown.

LC-MS/MS-based analyses can reveal protein turnover rates through stable isotope labeling by amino acids in cell culture (SILAC) [9, 15, 16]. Unfortunately, it is necessary to feed animals a diet containing isotope-labeled amino acids for prolonged periods to achieve complete proteome labeling [17]. Additional limitations of SILAC labeling are that proteins in low abundance are often undetected and it is not possible to specifically isolate or tag SILAC-labeled molecules [15].

To address the insensitivity of isotope labeling for newly synthesized proteins, non-canonical amino acid (ncAA) labeling was developed for selective analysis of *de novo* protein synthesis [18]. In this method, cells are cultured in media supplied with a methionine (Met) analog such as azidohomoalanine (Aha) which possesses an azide moiety. Due to the structural similarity to Met, cells incorporate Aha into growing protein chains using the endogenous translational machinery [19]. As a result, new proteins synthesized during pulse labeling with Aha are tagged with azides [18]. Importantly, azides are bio-orthogonal, meaning that they are biocompatible but do not cross-react with native biological systems and can be selectively modified through specific chemical reactions [20]. Azides can be conjugated to alkynes, forming a stable triazole product, using copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC), also known as a "click" chemistry

reaction [21]. Accordingly, the newly synthesized Aha-labeled proteins can be either ligated to affinity tags for selective enrichment and identification or fluorescent molecules for protein tracking and visualization [18, 22-24].

Over the last decade, ncAA labeling has been applied to various bacterial and mammalian cell culture systems *in vitro* to study biological processes ranging from quorum sensing to inflammation [25]. Importantly, Bagert et al. demonstrated that ncAA labeling can be applied to resolve protein synthesis within narrow time intervals [15]. Given the limitations of *in vitro* cell culture, extending the studies to the more complex *in vivo* environment is essential to augment our understanding of the dynamics of protein synthesis. McClatchy et al. first showed that *in vivo* labeling of the murine proteome is feasible by feeding animals an Aha-enriched diet for several days [26]. More recently, our group developed an injection-based technique that enables global labeling of murine proteome with ncAAs [27]. Our results demonstrated that two days of intraperitoneal injection of Aha was sufficient for systemic incorporation of Met analogs into the proteome of both juvenile mice and developing embryos with no overt perturbation of physiological functions. Our method provides several advantages over introducing ncAAs or stable isotopes in the animal diet including: ease of intraperitoneal or subcutaneous injection, global proteome labeling in a shorter time period, and more accurate dose-effect calculations.

The successful incorporation of ncAAs into the murine proteome through direct injection provides us with the opportunity to use a developmental model to temporally and spatially probe protein dynamics. However, to resolve dynamics of different intra- and extracellular proteins in the developing embryo, it is first necessary to generate a method that combines ncAA labeling and cellular compartment fractionation. The goals of this study were to build upon our previous results and 1) optimize ncAA enrichment for newly synthesized proteins in murine embryos, 2)

demonstrate this technique could resolve differences in the proteomes of embryonic time points, 3) utilize a tissue fractionation technique to isolate embryonic intra- and extracellular proteins and 4) combine ncAA labeling and tissue fractionation techniques to investigate turnover of different cellular compartments in developing embryos.

Using a cleavable biotin-alkyne linker, we could substantially enrich for Aha-labeled proteins from murine embryos compared with unlabeled controls. Identification of Aha-labeled proteins via LC-MS/MS showed that there was a significant difference in the composition of newly synthesized proteins from Aha-treated E12.5 and E15.5 embryos. The complexity of individual samples was reduced when proteins were fractionated into different cellular compartments. Importantly, we showed the dynamics of Aha labeling in the murine embryo varied as a function of time and cellular compartment. With these tools in hand, we are now poised to conduct a more comprehensive temporal analysis of synthesis and turnover for various intra- and extracellular proteins. Quantification of protein turnover is critical for understanding development, homeostasis and disease progression, and this method has the potential to determine the cues necessary for the formation and maintenance of functional tissues.

2.2 Methods

2.2.1 Animal Model

Animals used in these studies were derived from wild-type C57BL/6 mice (*Mus musculus)* purchased from The Jackson Laboratory. All experimental protocols were performed in compliance with established guidelines and all methods were approved by Purdue Animal Care and Use Committee (PACUC, protocols# 1209000723 and 1801001682). PACUC requires that all animal programs, procedures, and facilities at Purdue University to abide by the policies,

recommendations, guidelines, and regulations of the USDA and the United States Public Health Service in accordance with the Animal Welfare Act and Purdue's Animal Welfare Assurance. To generate embryos of defined ages, female mice were time-mated with males and noon on the date when a copulation plug was found was considered to be embryonic day (E)0.5.

2.2.2 Aha Labeling and Embryonic Tissue Collection

The methionine (Met) analog L-azidohomoalanine (Aha; Click Chemistry Tools, Scottsdale, AZ) was resuspended in phosphate buffered saline (PBS; 10 mg/ml), pH adjusted to 7.4 with NaOH, sterile filtered and stored at -20°C. All injections were administered to pregnant dams subcutaneously at 0.1 mg/g Aha and sterile PBS was used for control injections at 10 µl/g mouse. Embryos were collected at the desired time after injection by euthanizing dams via CO_2 inhalation, which was confirmed using cervical dislocation. The uterine horns were removed and dissected in ice cold PBS, then embryos were snap frozen in liquid nitrogen and stored at -80°C.

2.2.3 Optimization of Aha Enrichment

Pregnant females were injected with Aha or PBS once a day for two days, and embryos were harvested 24 h after the last injection (E12.5 or E15.5). For E12.5, three embryos were pooled for each biological replicate, whereas a single E15.5 embryo was used per biological replicate ($n = 3$). Embryos were homogenized in ice cold lysis buffer [PBS (pH 7.4) with 0.3% sodium dodecyl sulfate (SDS), 1× protease inhibitors (ThermoFisher Scientific, Waltham, MA) and 45 U benzonase (EMD Millipore, Darmstadt, Germany)] using a TissueRuptor (Qiagen, Venlo, Netherlands). Homogenates were rotated end-over-end at 4°C for 1 h and then cleared by centrifugation at 21,100 × *g* for 20 min. Protein concentration of cleared lysates was determined using the Pierce 660 nm Protein Assay (ThermoFisher Scientific). Lysates were alkylated with 30

mM iodoacetamide for 1 h at room temperature (RT) protected from light. The alkylated lysates were then reacted with diazo biotin-alkyne (DBA; Click Chemistry Tools) in a copper-catalyzed azide-alkyne cycloaddition (CuAAC) reaction with [50 µM DBA, 5 mM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA; Click Chemistry Tools), 1 mM $CuSO_4$, 0 – 20 mM aminoguanidine (AG) and 5 or 10 mM sodium ascorbate (NaAsc)] in a reaction volume of 800 µl with a final protein concentration of 2.5 mg/ml. The samples were rotated end-over-end at 4 °C overnight and excess unreacted DBA was removed using Zeba Spin Desalting Columns, 7K MWCO (ThermoFisher Scientific). Desalted samples were supplemented with 1% NP-40 and Aha-labeled proteins were affinity purified by incubation with 100 µl settled NeutrAvidin agarose beads (ThermoFisher Scientific) for 1.25 h with end-over-end mixing at 4°C. The beads were then washed four times with 1 ml (10× bead volume) PBS (pH 7.4) containing 0.05% SDS and 1% NP-40 to remove unlabeled proteins. Aha-labeled proteins were eluted by incubating beads with 400 µl elution buffer [PBS (pH 7.2), 100 mM $Na_2S_2O_4$, 0.05% SDS] for 1 h at RT, protected from light. Eluted proteins were precipitated by adding 1.6 mL 100% acetone (4× elution volume) and incubating overnight at -20 °C. Proteins were pelleted by centrifugation for 20 min at 21,100 × *g* and washed by adding 1.6 ml 80% acetone and incubating for 1.5 h at -20 °C. Samples were pelleted and dried at 26 °C in a centrivap concentrator (Labconco, Kansas City, MO) for 20 min and processed for LC-MS/MS as described below.

2.2.4 Western Blot Analysis of Aha-Labeled Samples

Proteins were resolved on 4 – 20% SDS-PAGE gels (BioRad, Hercules, CA) for 37 min at 170 V, transferred to a PVDF membrane (ThermoFisher Scientific) using a semi-dry transfer system for 40 min and probed overnight at 4°C with IRDye 680 Streptavidin (LICOR, Lincoln,

NE) diluted 1:3000 in Tris-buffered saline (TBS) with 0.05% Tween 20 + Protein-Free TBS Blocking Buffer (ThermoFisher Scientific) at a ratio of 1:1. Membranes were imaged using an Azure Biosystems c600 and then stained with Ponceau S (Sigma-Aldrich) for 20 min to confirm equal protein loading.

2.2.5 LC-MS/MS Analysis of Aha-Enriched Samples

Pellets were suspended in 50 µl 8 M urea with 100 mM ammonium bicarbonate. Proteins were reduced with 10 mM dithiothreitol (DTT) for 2 h at 37°C, shaking at 1,000 rpm (Eppendorf ThermoMixer F1.5, Hauppauge, NY). Samples were cooled to RT and alkylated with 25 mM iodoacetamide for 30 min while protected from light. Each sample was diluted with 100 mM ammonium bicarbonate to a final concentration of 2 M urea. Samples then underwent three subsequent digestion steps, all at 37°C and constant shaking at 1,000 rpm: (1) 2 h with 1 µg Endoproteinase LysC (New England Biolabs, Ipswich, MA); (2) overnight with 3 µg trypsin (MS-grade, ThermoFisher Scientific); and (3) an additional 2 h with 1.5 µg trypsin. Afterwards, enzymes were inactivated with 0.1% trifluoroacetic acid (TFA; VWR, Radnor, PA).

Peptides were processed with Pierce Detergent Removal Spin Columns (ThermoFisher Scientific) per the manufacturer's protocol. Samples were incubated on the column resin for 2 min prior to centrifugation for 2 min at 1,500 × *g*. Following detergent removal, samples were cleaned of excess salts with C-18 MicroSpin Columns (The Nest Group Inc., Southborough, MA). Briefly, columns were conditioned with 100 µl 100% acetonitrile (ACN; ThermoFisher Scientific) and equilibrated with 100 µl 0.1% TFA. Samples were added to the C-18 columns, washed with 300 µl 0.1% TFA, and eluted in 100 µl 80% ACN with 25 mM formic acid (FA; ThermoFisher Scientific). Peptides were dried for 4 h at 45°C in a centrivap concentrator and suspended in 10 µl

of 3% ACN with 0.1% FA. After suspension, peptide concentration was measured with the Pierce Quantitative Colorimetric Peptide Assay (ThermoFisher Scientific) and the most concentrated Aha sample was brought to 0.2 µg/µl by addition of 3% ACN with 0.1% FA. The equivalent volume of 3% ACN with 0.1% FA was added to remaining Aha and PBS samples. Samples were stored at -80°C until analyzed using LC-MS/MS.

Peptides were analyzed using the Dionex UltiMate 3000 RSLC Nano System coupled to the Q Exactive HF Hybrid Quadrupole-Orbitrap Mass Spectrometer (QE HF; ThermoFisher Scientific). Following digestion, 1 µg of peptide was loaded onto a 300 µm i.d. × 5 mm C18 PepMap 100 trap column and washed for 5 min using 2% ACN with 0.01% FA at a flow rate of 5 µl/min. After washing, the trap column was switched in-line with a 75 µm × 50 cm reverse phase Acclaim C18 PepMap 100 analytical column heated to 50°C. Peptides were separated using a 120-min gradient elution method at a flow rate of 300 nl/min. Mobile phase A consisted of 0.01% FA in purified water, while mobile phase B consisted of 0.01 % FA in 80% ACN. The linear gradient started at 2% B and reached 10% B in 5 min, 30% B in 80 min, 45% B in 91 min, and 100% B in 93 min. The column was held at 100% B for the next 5 min before being brought back to 2% B and held for 20 min. Samples were injected into the QE HF through the Nanospray Flex Ion Source fitted with an emission tip from New Objective. Data acquisition was performed monitoring the top 20 precursors at 120,000 resolution with an injection time of 100 ms.

2.2.6 Analysis of LC-MS/MS Spectra

Raw data files were analyzed using MaxQuant (version 1.6.1.0) [28]. Default settings were used unless noted otherwise (see all parameters in Table A.1). Peak lists were searched against the *Mus musculus* Uniprot FASTA database (November 2018), *Gallus gallus* Avidin Uniprot FASTA

protein sequence (May 2018) and a common contaminants database (January 2018). Cysteine carbamidomethylation was included as a fixed modification as were pertinent variable modifications (Table A.1). Peptide and protein false discovery rates (FDR) were set to 0.01 and determined by a reverse decoy database derived from the *Mus musculus* database. Raw protein intensities were analyzed with Microsoft Excel (for filtering and data handling) and GraphPad Prism 8 (for data visualization). Proteins that had less than two razor and unique peptides or proteins marked as a potential contaminant or reverse hit were removed. In addition, proteins identified by match between runs were removed prior to analysis. Proteins were only included in subsequent analyses if identified in at least two biological replicates. Raw intensities were $\log_2$-transformed and were considered enriched if [$\log_2$(Aha raw intensity) – $\log_2$(PBS raw intensity)] > 1 (indicating a >2-fold change). Tissue compartment categories (cytosolic, nuclear, membrane, cytoskeletal, matrisome (ECM)) were assigned to proteins using categorizations derived from the Gene Ontology (GO) Consortium [29] and The Matrisome Project [8]. Percentage of each category was calculated by dividing the summed protein intensities designated to the given compartment by the total protein intensity in the sample.

To compare the distribution of Aha-labeled proteins identified in E12.5 and E15.5 embryos, the raw intensities were normalized and visualized using a volcano plot. All raw intensities in individual samples were summed, then averaged over biological replicates for each time point and a normalization factor was generated by dividing the average raw intensity of E12.5 by E15.5. The raw intensity of each protein in E15.5 samples was multiplied by the normalization factor, $\log_2$-transformed and the fold change (E15.5/E12.5) was calculated. Statistical analysis of transformed intensities was conducted using a two-tailed t-test and corresponding p-values were $\log_{10}$-transformed and visualized using GraphPad Prism 8.

2.2.7 Tissue Fractionation

Control E15.5 embryos, harvested from dams injected with PBS at E13.5 and E14.5, were fractionated using buffers of increasing stringency to selectively enrich for cytosolic (C), nuclear (N), membrane (M), cytoskeletal (CS) or ECM proteins as previously described [14, 30], with some modifications (Table A.2). Embryos were mechanically homogenized with a TissueRuptor in C buffer (400 mg wet weight tissue in 500 µl buffer), rotated end-over-end for 30 min at 4°C, followed by centrifugation at 14,000 × *g* for 20 min. Supernatants were collected, snap frozen and stored at -80°C. The remaining pellet was resuspended in another 500 µl of C buffer and processed as previously described. The supernatant was snap frozen and stored at -80°C, to be combined with the first C fraction before subsequent processing. The pellet was then sequentially processed with N, M and CS buffers following the same protocol as described for the C buffer, with the exception that extractions using CS buffer were performed at RT. The remaining insoluble pellets (ECM) were snap frozen and stored at -80°C.

2.2.8 LC-MS/MS Analysis of Fractionated Tissue

Aliquots of C, N, M, and CS fractions were diluted 2-fold by combining 250 µl lysate with 250 µl 8 M urea containing 100 mM ammonium bicarbonate (final concentration 4 M urea). The ECM fraction was resuspended in 100 µl 8 M urea with 100 mM ammonium bicarbonate. Proteins in C and N fractions were reduced, alkylated, digested, desalted, and dried as described above for the Aha enrichment study. M fractions were processed similarly, but were also cleaned with detergent removal columns prior to desalting. CS and ECM fractions were reduced and alkylated prior to deglycosylation with 0.1 U of chondroitinase ABC (Sigma-Aldrich) for 2 h at 37°C, shaking at 1000 rpm. After deglycosylation, CS and ECM fractions were digested, processed with

detergent removal columns, desalted, and dried. All samples were resuspended in 3% ACN with 0.1% FA and brought to 1 μg/μl.

Samples of each fraction were analyzed on the QE HF as stated above, with each biological replicate representing a separate embryo (*n = 3*). Raw data files were analyzed by MaxQuant as described above. Additionally, samples were grouped by buffer fraction (C, N, M, CS, ECM) for label-free quantification (LFQ) analysis. Data were analyzed using Microsoft Excel and Prism for filtering and data visualization, respectively. Gene ontology (GO) terms and Reactome pathways on the 50 most abundant proteins in each fraction were analyzed using g:Profiler [31, 32].

2.2.9 Temporal Study of Protein Turnover Using Aha

Time-mated females were injected with 0.1 mg/g Aha at E12.5, embryos were harvested 0, 3, 6, 12, 24, and 48 h after injection, snap frozen in liquid nitrogen and stored at -80°C. Embryos were fractionated into different cellular compartments as described above and conjugated with biotin-alkyne in a CuAAC reaction with 25 mM iodoacetamide, 10 mM NaAsc, 50 μM biotin-alkyne, 10 mM THPTA, 2 mM $CuSO_4$ and 20 mM AG for 2 h at RT. Reacted samples were precipitated with methanol-chloroform and air-dried protein pellets of C, N, M and CS fractions were resuspended in 2× Laemmli buffer (BioRad) with 5% β-mercaptoethanol. ECM pellets were resuspended in 100 mM Tris-HCl (pH 7.6), 8% SDS, 0.1 M DTT, 1× native Laemmli with 5% β-mercaptoethanol. All samples were then heated at 95°C for 5 min and analyzed via western blotting as described above. Western blot images were analyzed using ImageJ (NIH, Bethesda, MD) to calculate the mean fluorescence intensities for each time point. A 117 × 816 pixel region of interest (ROI) was generated that was slightly larger than the width of each lane. The sum of the intensity units in the ROI was measured using "RawIntDen." For each blot, the intensity at t = 0 was used

to normalize subsequent time points. Normalized intensity values for each fraction at each time point were plotted as a function of time, and the change in fluorescence intensity between time points were plotted and analyzed using GraphPad Prism 8.0, $n = 3$ biological replicates/blot.

2.3 Results and Discussion

To determine the feasibility of enriching proteins that were newly synthesized within developing embryos, we first used biotin-alkyne to selectively enrich for Aha-labeled proteins. Time-mated C57BL/6 murine dams were injected subcutaneously with either 0.1 mg/g Aha or PBS (control) once a day for two days prior to harvesting at E15.5. Soluble proteins were isolated by homogenizing embryos in 0.3% SDS in PBS and the insoluble portion was removed by centrifugation. Western blot analysis of soluble lysates confirmed the feasibility of conjugating nascent proteins with biotin-alkyne in a CuAAC reaction, enriching with NeutrAvidin beads, and eluting Aha-labeled proteins from embryos (Figure 2.1A). However, LC-MS/MS analysis of eluted proteins revealed avidin peptides dominated the MS/MS spectra, limiting the resolution by which Aha-labeled proteins can be resolved (Figure 2.1B). The excess avidin was attributed to the use of harsh elution conditions, *i.e.* boiling in SDS, which disrupts the biotin-avidin interaction to elute biotinylated proteins. In addition, very few proteins were enriched in Aha-labeled samples as evidenced by Coomassie staining of eluted samples (Figure 2.1A).

To overcome these limitations, diazo biotin-alkyne (DBA), a cleavable biotin linker was used. The diazo group between the biotin and alkyne was cleaved by reducing with sodium dithionite ($Na_2S_2O_4$) under mild conditions preventing the elution of excess avidin (Figure A.1) [33]. Initial optimization experiments using cells cultured *in vitro* confirmed selective enrichment of Aha-labeled proteins. However, when more complex tissue samples were used (E15.5 embryos

from dams injected with Aha or PBS), there was a substantial number of proteins also detected in unlabeled samples, suggesting nonspecific protein binding to NeutrAvidin beads. Several approaches were employed to eliminate nonspecific protein contamination including blocking beads with bovine serum albumin prior to lysate addition, incubating lysates with beads for shorter periods, and increasing the number and stringency of washes (data not shown). Nevertheless, none of these strategies resulted in reducing the level of eluted proteins in unlabeled samples.

We next tested the possibility that side reactions from the CuAAC (click) reaction conditions lead to nonspecific labeling of proteins with the diazo biotin-alkyne linker. Soluble lysates from E15.5 embryos harvested from dams injected with PBS were reacted in the absence or presence of CuAAC and DBA reagents and were enriched with NeutrAvidin beads. A significant amount of unlabeled protein was eluted when using CuAAC and DBA, but not in the control or sample that was reacted with DBA only (Figure 2.1C), suggesting that the alkyne linker conjugates to non-azide targets in CuAAC reaction conditions.

Ascorbate reduction of copper has been reported to produce dehydroascorbate as the oxidation product, which, along with other ascorbate byproducts, interacts with lysine and arginine side chains [34, 35]. These interactions can result in the formation of several covalently cross-linked protein adducts, potentially leading to non-specific binding of unlabeled proteins to the beads. Aminoguanidine (AG), which is structurally similar to the guanidine group of arginine (Figure A.1), was added to the click reaction mixture to investigate if it can reduce the nonspecific labeling by scavenging ascorbate byproducts [36]. Adding AG in the click reaction resulted in substantial reduction in nonspecific protein labeling without a comparable decrease in the amount of eluted proteins in the Aha sample (Figure 2.2B). Furthermore, we found that higher concentrations of NaAsc increase background labeling, leading us to reduce the amount of this

reagent in further reactions. Together, the addition of AG with the use of a cleavable linker provides us with a method that increases the identification of Aha-labeled proteins by reducing nonspecific binding in unlabeled samples as well as minimizing avidin contamination. For subsequent studies, 10 mM AG and 5 mM NaAsc was used in the click reaction.

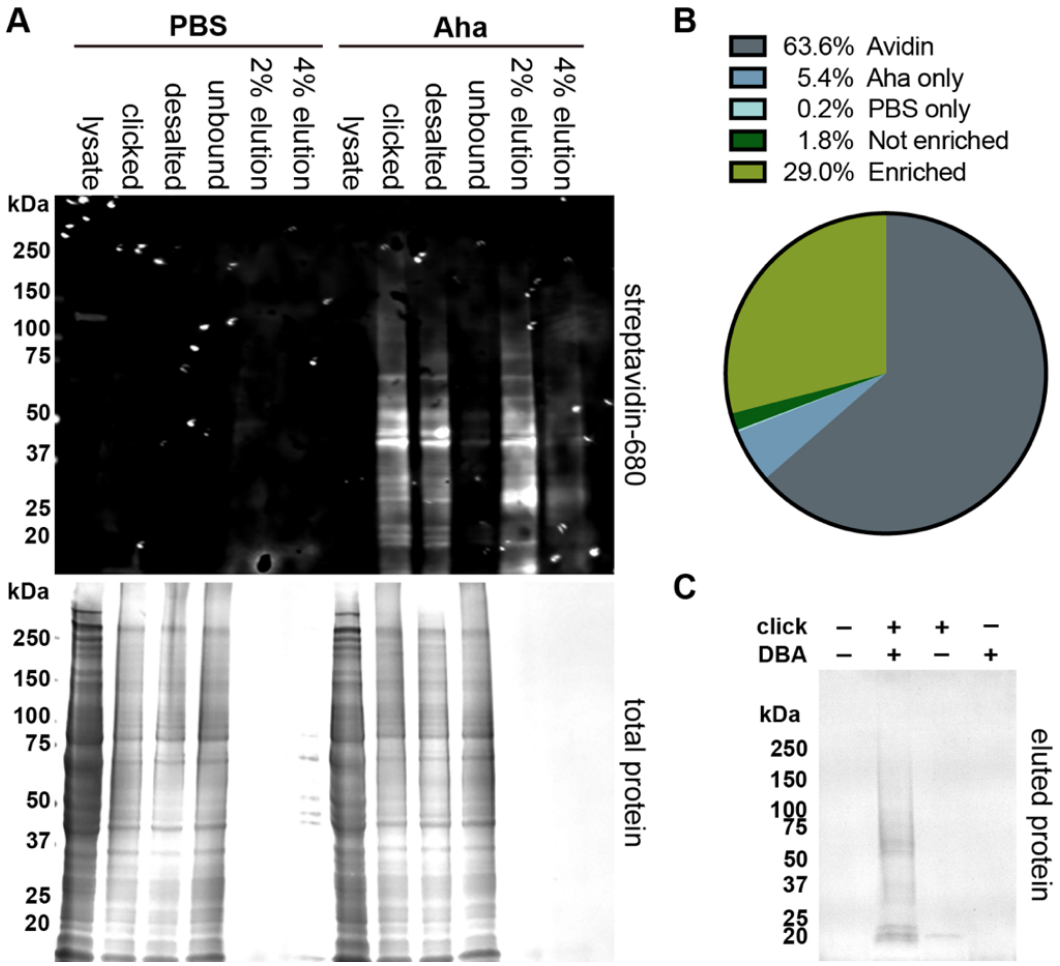


Figure 2.1 Enrichment of Aha-labeled proteins using biotin-alkyne. **(A)** Western blot analysis of E15.5 embryos harvested from dams injected with 0.1 mg/g Aha or PBS. Soluble lysates (lysate) were reacted with biotin-alkyne using CuAAC (clicked) and isolated from unlabeled proteins using NeutrAvidin beads. Removal of unreacted biotin-alkyne via desalting columns (desalted) did not substantially affect total protein concentration. These conditions isolated the majority of Aha-labeled proteins from the lysates as indicated by minimal signal in the unbound lane. Aha-labeled proteins were eluted using 2 or 4% SDS (lanes 5 and 6 for PBS, lanes 11 and 12 for Aha, respectively). **(B)** LC-MS/MS analysis of the eluted proteins revealed a high percentage of avidin contamination. Average raw intensity of $n = 2$ embryos. **(C)** Non-specific binding of unlabeled proteins due to CuAAC. Control E15.5 embryos from PBS-injected dams were reacted using

CuAAC (click) with diazo biotin-alkyne (DBA) and incubated with NeutrAvidin beads. Substantially more proteins were eluted off the beads when the CuAAC and DBA were present in the reaction (lane 2).

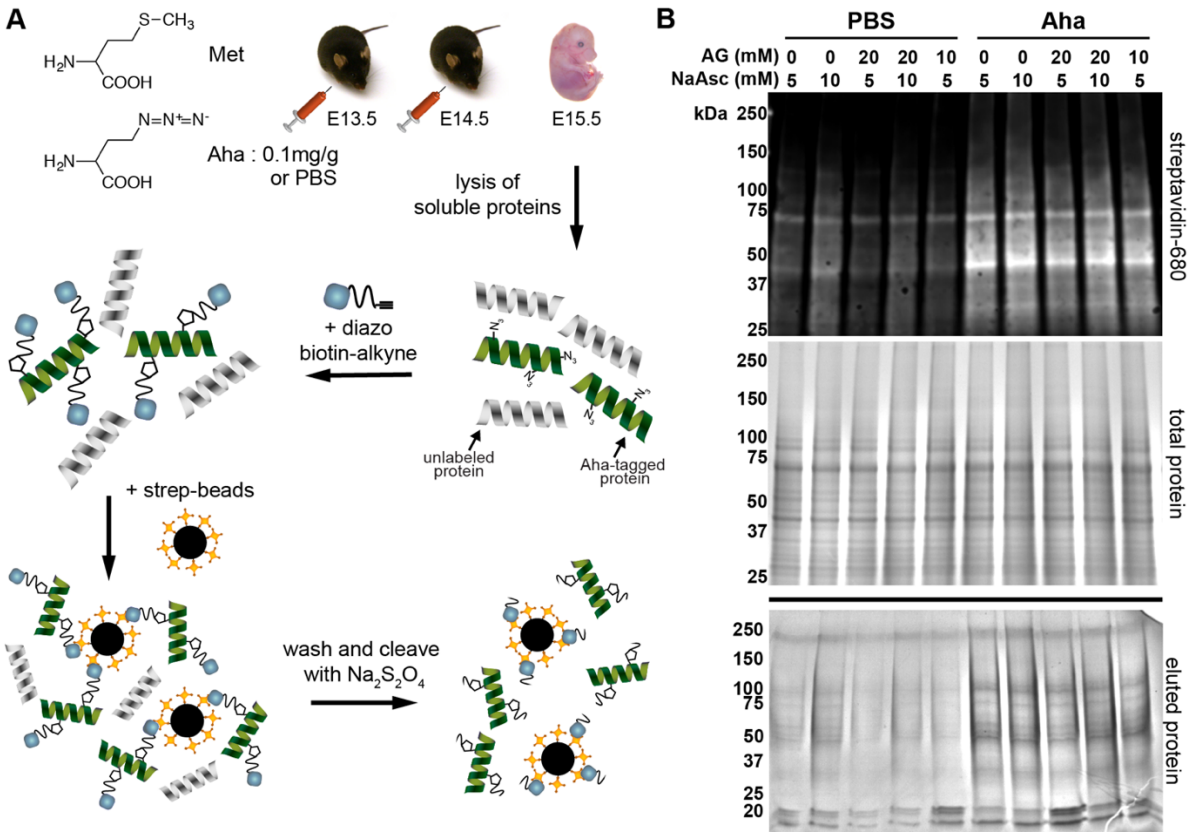


Figure 2.2 Enrichment of Aha-labeled proteins from murine embryos. **(A)** Workflow for enrichment of Aha-labeled proteins using a diazo biotin-alkyne (DBA) linker. The methionine (Met) analog, azidohomoalanine (Aha), was injected into time-mated dams at 0.1 mg/g, once a day at E13.5 and E14.5, whereas 10 µl/g PBS was injected into control animals. E15.5 embryos were harvested and lysates of soluble proteins were reacted with DBA and isolated using NeutrAvidin beads. Unlabeled proteins were washed away and Aha-labeled proteins were released using $Na_2S_2O_4$. **(B)** Soluble lysates from E15.5 embryos generated as described in **(A)** were reacted with DBA using different concentrations of sodium ascorbate (NaAsc) and aminoguanidine (AG). Lysates (18.5 µg/well) were analyzed via western blotting (top) to confirm Aha labeling, a duplicate gel was run at the same time to ensure equal loading (middle, Coomassie-stained gel). The same samples were enriched using the workflow in **(B)** and eluted proteins were run on a gel and stained with Coomassie (bottom). NaAsc increases nonspecific binding of unlabeled proteins to NeutrAvidin beads, whereas addition of AG reduces the nonspecific binding.

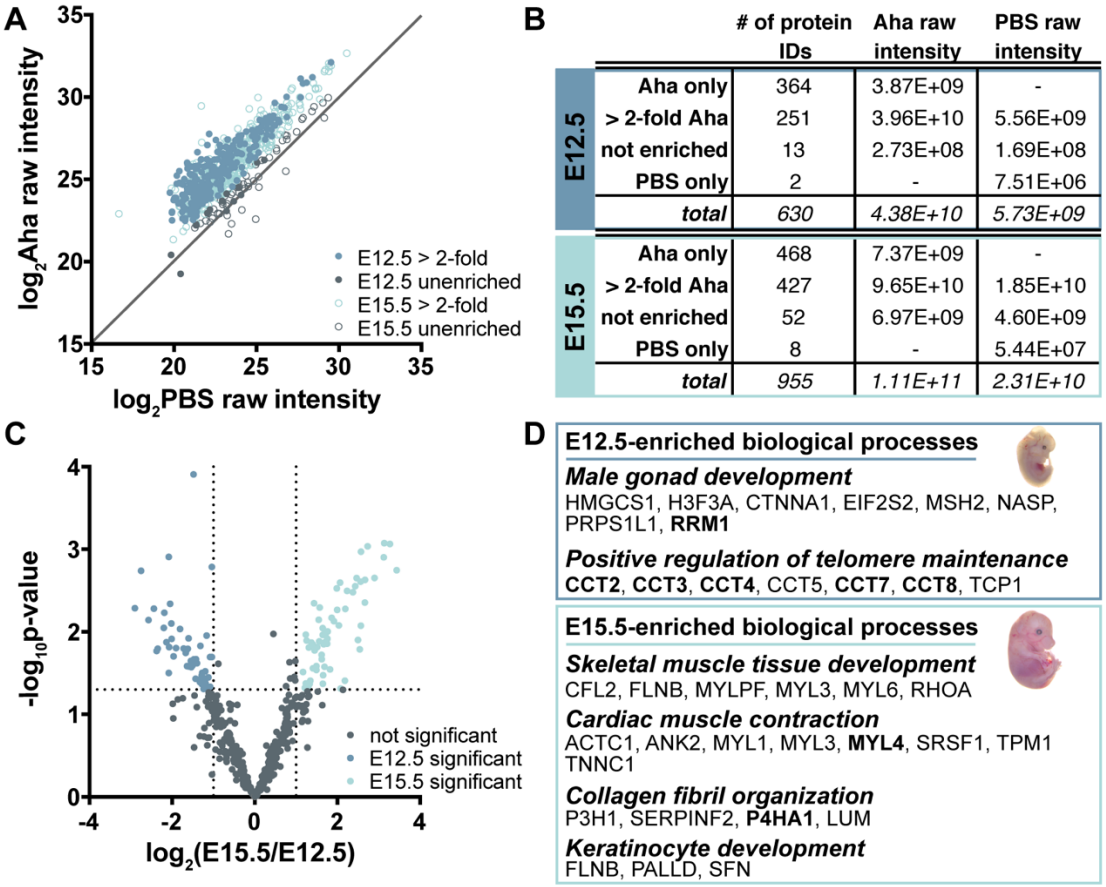
To demonstrate this method could resolve differences in newly synthesized proteins within distinct time windows, dams were time-mated and injected so that Aha-labeling would occur

between E10.5 – E12.5 or E13.5 – E15.5. Dams were injected with Aha or PBS (control) at t = 0 and 24 h and proteins were isolated from embryos at 48 h. Proteins from soluble lysates of E12.5 and E15.5 embryos were enriched using the workflow described in Figure 2.2A, analyzed using LC-MS/MS and the raw intensity of proteins identified in Aha- and PBS-treated samples was compared (Figure 2.3). After filtering out contaminants, reverse hits and proteins that had <2 razor and unique peptides, and imposing a FDR of 0.01, ~ 50% of total proteins identified were found exclusively in Aha-treated embryos and ~ 90% of proteins were >2-fold enriched in Aha compared to PBS samples (Figure 2.3A,B). This degree of enrichment is in line with previous reports that used similar labeling strategies to investigate newly synthesized proteins in adult murine brains [37, 38].

Comparison of newly synthesized proteins isolated from the two embryonic time points revealed that 99 and 379 Aha-labeled proteins were exclusive to E12.5 and E15.5 lysates, respectively. A volcano plot of the 516 Aha-labeled proteins that were common to both time points showed the subsets of proteins were significantly more abundant (> 2-fold, $p < 0.05$) at E12.5 (54 proteins) and E15.5 (60 proteins; Figure 2.3C). A Gene Ontology (GO) analysis was conducted to determine if Aha-enrichment could capture biological processes that pertained to these developmental time frames. Analysis of the newly synthesized proteins in Aha-labeled E12.5 samples identified biological process terms including *male gonad development* and *positive regulation of telomere maintenance* (Figure 2.3D). Notably, genes that regulate sex determination are significantly upregulated around E11.0, and by E12.5 the transcriptomes between testes and ovaries are highly dimorphic [39]. In addition, maintaining telomere length is in line with previous studies that reported an increase in telomerase activity during the early stages of embryonic development [40, 41].

Analysis of the newly synthesized proteins in E15.5 samples generated biological process terms that were more indicative of later development, some of which are listed in Figure 2.3D. Similar to other organ systems, skeletal and cardiac muscle start rapidly differentiating after E11.5 [42, 43], increasing the deposition of the contractile machinery such as the myosin light chain isoform MYL3 that is only expressed in skeletal muscle around E15.5 [44]. To support and direct the growth of tissues, the ECM also needs to increase in density and become more organized [45, 46]. Indeed, the expression of lumican (LUM), one of the small leucine-rich proteoglycans that regulate the assembly of type I collagen fibrils, increases dramatically between E11.5 and E15.5 [47], which is consistent with our identification of LUM only at E15.5 (Figure 2.3D). Notably, there was high correlation of protein intensity distributions between biological replicates, whereas there was low correlation between Aha and PBS samples (Figure 2.4A). Analysis of Aha-labeled proteins revealed large contributions from cytosolic, nuclear and cytoskeletal proteins; however, very few matrisome (ECM) and membrane proteins were identified (Figure 2.4B).

These studies only identified Aha-labeled proteins that were soluble in 0.3% SDS. Even if the proteins in the remaining insoluble portion were analyzed simultaneously with the 0.3% soluble proteins, it is unlikely that the number of IDs would be substantially increased. A limitation of LC-MS/MS is that peptides of highly abundant proteins can mask the signal of those with similar properties but of lower abundance. One way to increase the number of protein IDs is to increase the time of the LC phase, effectively increasing the resolution; however, this also greatly increases cost. Alternatively, proteins can be analyzed in subsets depending on the biological question being asked. Examples include isolating by molecular weight or fractionating based on biochemical characteristics, such as lipid-soluble membrane proteins and the relatively insoluble ECM [30, 48].



		# of protein IDs	Aha raw intensity	PBS raw intensity
E12.5	Aha only	364	3.87E+09	-
	> 2-fold Aha	251	3.96E+10	5.56E+09
	not enriched	13	2.73E+08	1.69E+08
	PBS only	2	-	7.51E+06
	total	*630*	*4.38E+10*	*5.73E+09*
E15.5	Aha only	468	7.37E+09	-
	> 2-fold Aha	427	9.65E+10	1.85E+10
	not enriched	52	6.97E+09	4.60E+09
	PBS only	8	-	5.44E+07
	total	*955*	*1.11E+11*	*2.31E+10*

Figure 2.3 Enrichment of Aha-labeled proteins from different embryonic time points. Newly synthesized proteins isolated from E12.5 and E15.5 embryos were labeled with Aha as shown in Figure 2.2A and the enriched proteins were analyzed using LC-MS/MS. **(A)** Scatter plot comparing the $\log_2$-transformed intensity of proteins identified in both Aha and PBS reveals the substantial enrichment of Aha-labeled proteins at both E12.5 and E15.5. Each point is the average raw intensity of a single protein, $n = 3$ biological replicates. **(B)** Summary of protein IDs in Aha-labeled and PBS control samples. **(C)** Volcano plot of proteins identified in both E12.5 and E15.5 embryos. Proteins were considered significantly more abundant if there was >2-fold difference and $p < 0.05$ between time points. Vertical lines indicate ± 2-fold change, horizontal line indicates $p = 0.05$, calculated using a two-tailed t-test. **(D)** Select biological process terms identified by conducting GO analyses on proteins exclusive to and >2-fold enriched compared with PBS for individual time points. Proteins found at both time points, but >2-fold at either E12.5 or E15.5 are indicated in bold.

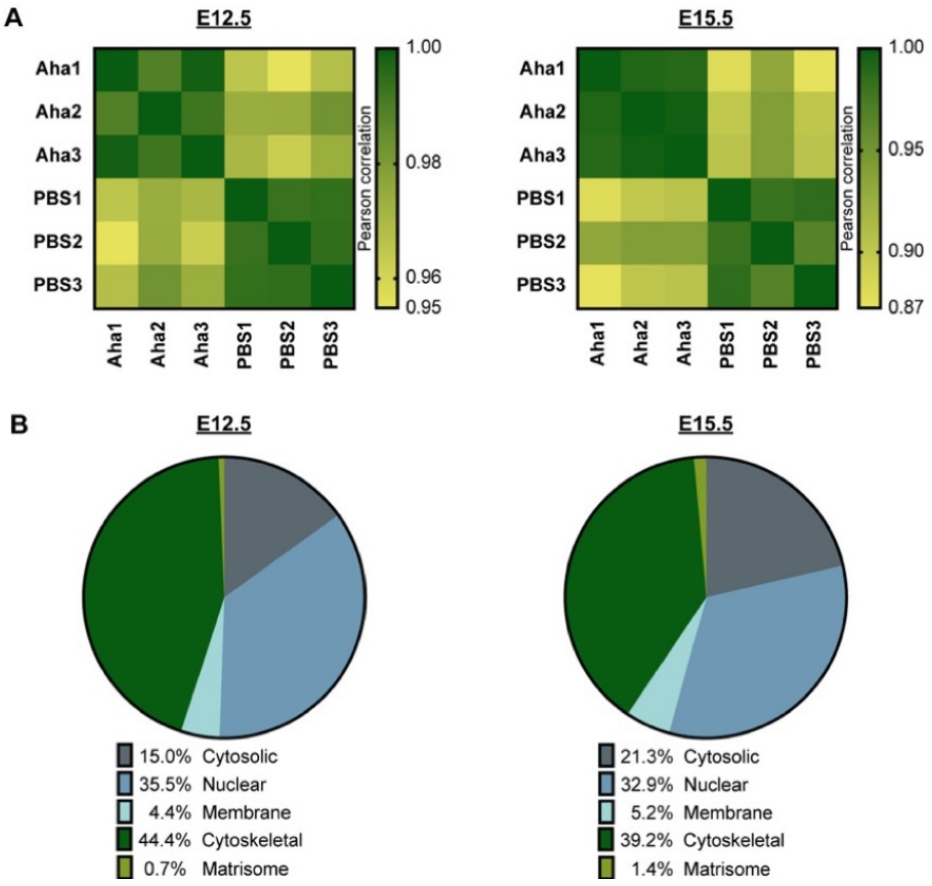


Figure 2.4 Summary of Aha-enriched proteins found at E12.5 (left) and E15.5 (right). **(A)** Heat map comparisons of Pearson Correlation coefficients indicated a higher degree of correlation between samples within the Aha and PBS groups. **(B)** Distribution of raw intensity of proteins found only in Aha as a function of cellular compartment, average of $n = 3$ biological replicates. Note the absence of Avidin.

Expanding the utility of *in vivo* ncAA labeling to identify newly synthesized proteins in different intra- and extracellular compartments holds great potential for increasing the resolution of protein IDs in developmental model systems, but it is unclear if these methods, optimized on adult samples [8-14], can be used on newly assembled embryonic tissues. To demonstrate the applicability of selectively fractionating different cellular compartments, unlabeled E15.5 embryos were homogenized using a protocol designed to isolate the cytosolic (C), nuclear (N), membrane (M), cytoskeleton (CS), and matrisome (ECM) fractions using buffers of increasing stringency [8,

14]. There is a clear distinction in the distribution of proteins of different molecular weight across the fractions (Figure 2.5A). To resolve compartment-specific protein dynamics, we performed LC-MS/MS on the different fractions from E15.5 embryos and annotated which cellular compartment the proteins were predominantly localized to based on the GO database (Figure 2.5B). Cytosolic, membrane and matrisome proteins were enriched in the appropriate fractions (C, M, and ECM respectively; Figure 2.5B); whereas, nuclear and cytoskeletal proteins were more widely distributed. GO analysis for the 50 most abundant proteins in each fraction was performed. The top 10 significant biological process terms were plotted as a function of fraction and there was some enrichment of fraction-specific terms (bold; Figure 2.5C).

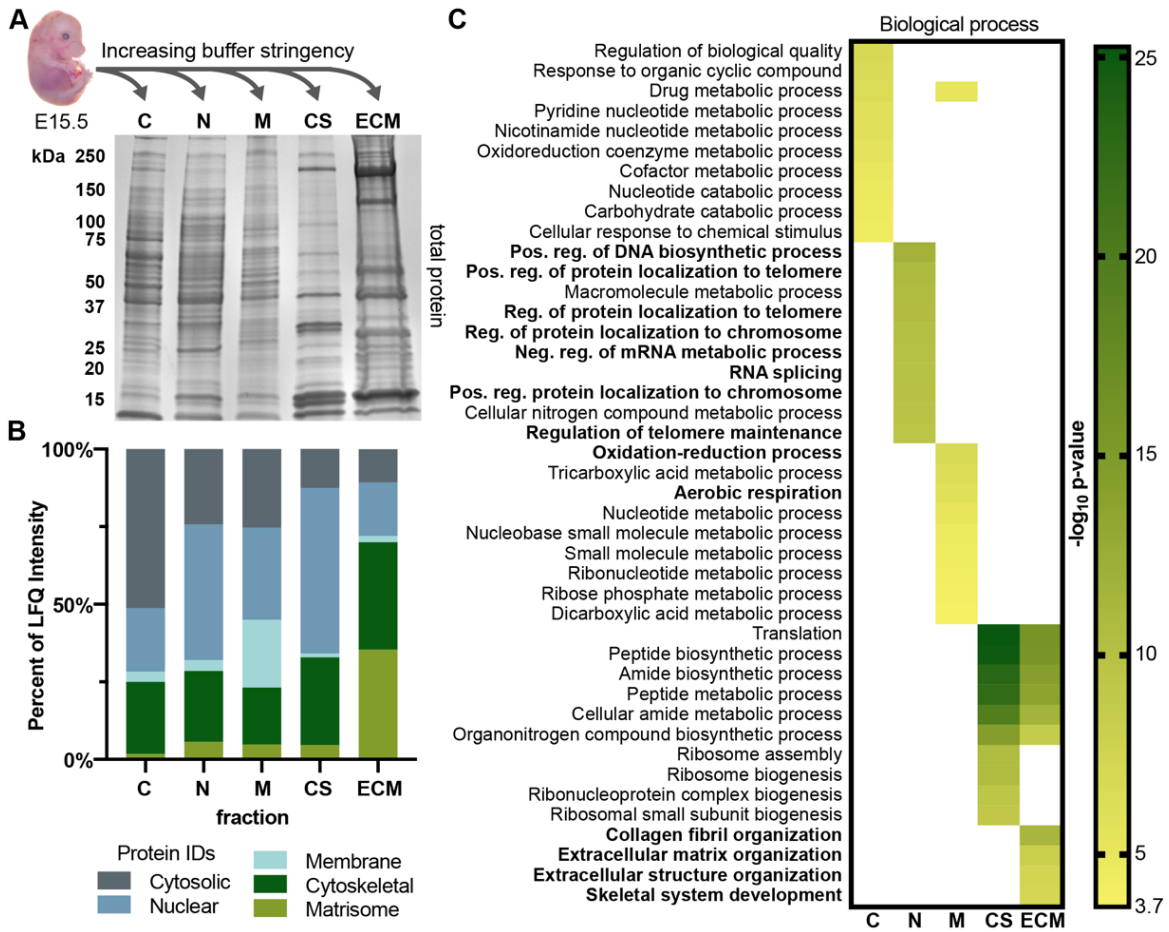


Figure 2.5 Fractionation of embryonic tissue into different cellular compartments. **(A)** Untreated E15.5 embryos were homogenized in buffers of increasing stringency to obtain cytosolic (C), nuclear (N), membrane (M), cytoskeletal (CS) and matrisome (ECM) fractions. **(B)** The distribution of intra- and extracellular proteins across the fractions, plotted as a percent of the average LFQ intensity of $n = 3$ embryos. **(C)** Top 10 significantly enriched non-redundant GO biological process terms of the top 50 proteins from each fraction plotted as a function of their respective fractions. GO terms consistent with fraction isolation are indicated in bold.

In contrast, terms for the same set of proteins derived from the Reactome pathway database [49] revealed trends that were more consistent with the types of proteins identified in each fraction (Figure 2.6). Heat shock proteins and other responses to stress were represented by 5/10 of the most significant pathways in the C fraction. While these terms are typically associated with pathological processes, this large family of proteins that includes HSP90 and HSF1 also regulates normal embryonic development [50]. Pathways associated with the chaperonin family, a group of

multimeric complexes that facilitate protein folding, were found in the N fraction (4/10). These proteins are found in both the nucleus and cytoplasm [51, 52], and, similar to heat shock proteins, have been shown to be critical in response to stress as well as during organ growth [53]. In addition, another 4/10 terms in the N fraction were associated with mRNA metabolism. Interestingly, 6/10 terms in the M fraction are related to mitochondria, which can be attributed to this being a double membrane-bound organelle. The inner membrane forms cristae that invaginate deep within the cell and contain the protein complexes of the respiratory system [54]. In the CS fraction, terms associated with muscle contraction made of 3/9 significant pathways, reflecting the large amount of cytoskeletal elements that striated and smooth muscle need to generate force [55]. Notably, the 9/10 terms in the ECM fraction are related to the matrisome and 5/10 are directly related to collagen metabolism, which is thought to be dynamic during embryogenesis [56].

These results validate that fractionation protocols used on adult tissues are suitable for developing tissues, and the distribution of the Reactome terms (Figure 2.6) is consistent with the cellular components expected to be identified in each fraction [8]. Reducing the complexity of individual samples will increase the overall identification of proteins by LC-MS/MS for more targeted studies. Specifically, we found LC-MS/MS analysis of the M fraction increased the overall intensity of membrane proteins identified when compared with homogenizing whole embryos with 0.3% SDS lysis buffer (Figure 2.4B), as well as an increase in the amount of matrisome components when the ECM fraction was analyzed (Figure 2.5B).

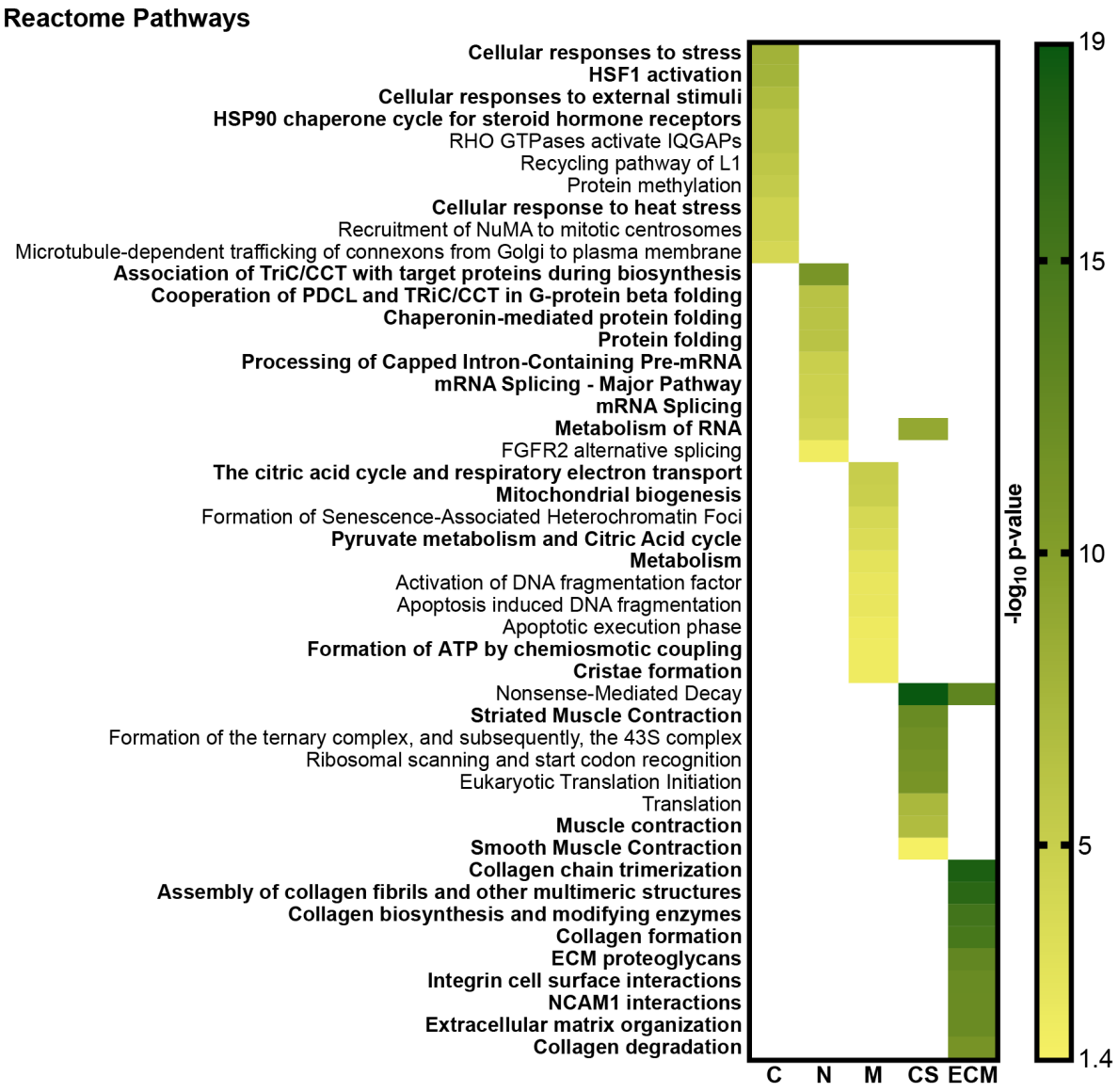


Figure 2.6 Reactome pathways reflect differential distribution of proteins across fractions. The top 10 significantly enriched, non-redundant Reactome terms based on the top 50 proteins from each fraction were determined for untreated E15.5 samples. Fraction-specific biological pathways discussed in text are indicated in bold.

To investigate that protein turnover can be resolved within each cellular fraction, time-mated dams were injected once with 0.1 mg/g Aha at E12.5, and embryos were harvested at 0, 3, 6, 12, 24, and 48 h following injection. Lysates were fractionated, reacted with biotin-alkyne, and analyzed via western blotting using streptavidin-680. Overall, the highest degree of Aha labeling

was found 6 h after injection (Figure 2.7A). Between 3 – 24 h, the relative amount of labeling in different fractions varied considerably, as shown when the change in fluorescence intensity was plotted as a function of time (Figure 2.7B,C). While measurements based on western blotting are semi-quantitative, these results indicate that proteins of different solubility vary in turnover rate. Two-way ANOVA revealed that the effect of time and fraction were significant for both fluorescence intensity ($p<0.0001$ for time and fraction; Figure 2.7B) and change in fluorescence with respect to time ($p<0.0001$ for time, $p<0.01$ for fraction; Figure 2.7C). Between E12.5 and E14.5, murine embryos increased in weight from 95.0 ± 9.2 mg to 275.3 ± 20.8 mg (average ± S.D.; $N \geq 4$). Even with this large increase, Aha-labeled proteins were still present in all fractions 48 h after injection, which is consistent with a recent study demonstrating the lifetime of proteins in adult murine tissues for days [9].

Overall, the results from this study set the stage for future investigations to combine the methods of Aha labeling and enrichment with tissue fractionation to identify and quantify the key proteins involved with various developmental processes.

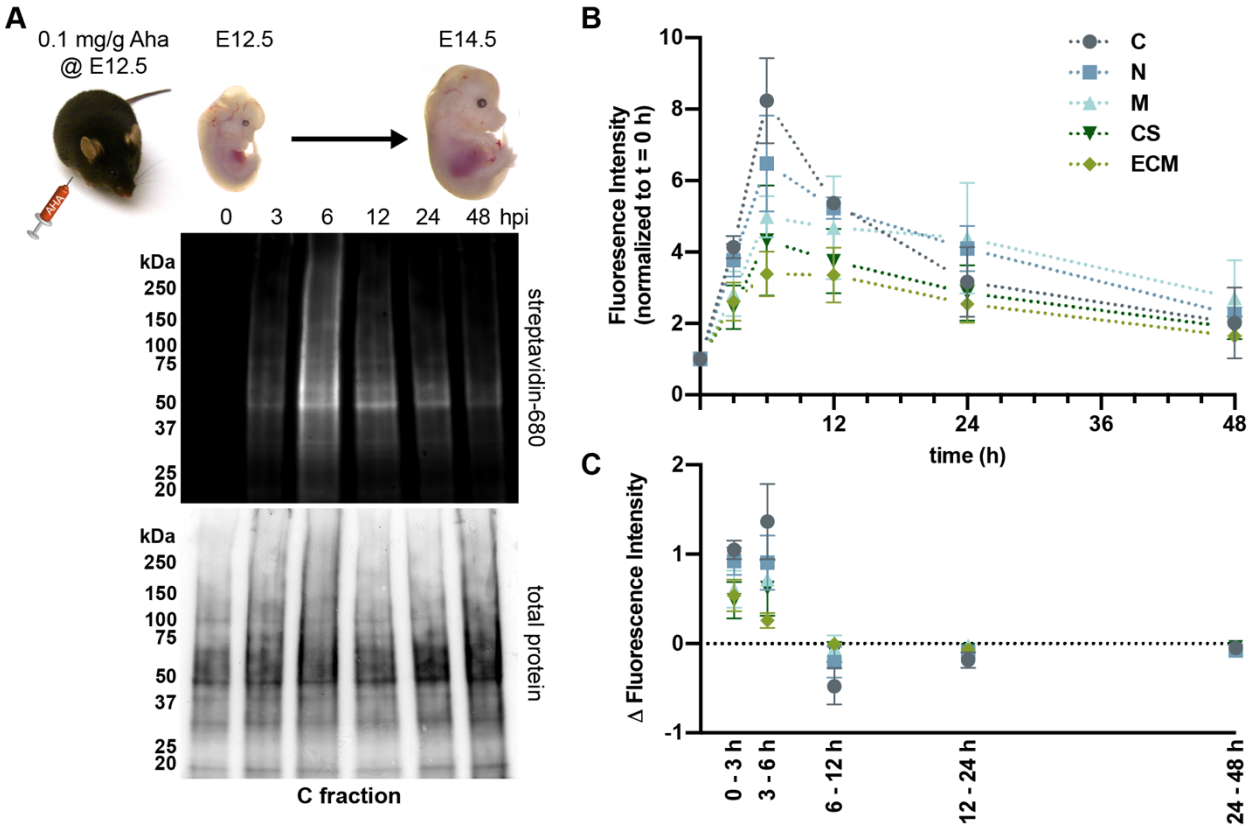


Figure 2.7 Persistence of Aha-labeled proteins in developing embryos. **(A)** Pregnant dams were injected one time with 0.1 mg/g Aha and embryos were harvested 0 – 48 h post injection (hpi). Proteins were isolated following Figure 2.5A, fractions were clicked with biotin-alkyne and analyzed via western blotting (30 µg/well). Shown is a representative image of the blot for the cytosolic (C) fraction (top). Ponceau S staining of same membrane used to confirm equal loading (bottom). **(B, C)** Fluorescence intensity of western blot lanes were plotted as function of time and fraction, normalized to t = 0 (*n = 3* biological replicates).

2.4 Conclusions

We previously demonstrated that ncAA labeling of proteins can be readily performed *in vivo* in a murine model with facile intraperitoneal injection, allowing for labeling of newly synthesized proteins in a variety of tissues and at varying stages of development from embryonic to juvenile mice [27]. Here, we extend these results to show that subcutaneous injection of the Met analog Aha, in combination with an optimized enrichment protocol, can be used to isolate newly synthesized proteins for LC-MS/MS identification with very little background relative to PBS

injected controls. Approximately 50% of total protein IDs were found exclusively in Aha-labeled tissue, and 90% of proteins found in both Aha and PBS samples were enriched >2-fold in the Aha-labeled samples. The proteins that were selectively enriched from E12.5 and E15.5 embryos were related to developmental processes occurring around each time point. In addition, we established the feasibility of using cellular fractionation, which enabled a broader investigation of intra- and extracellular proteins in developing embryos than standard isolation protocols. Reactome pathway analysis of LC-MS/MS data demonstrated that proteins identified in each fraction corresponded to the expected cellular compartment, *i.e.* cytosolic (C), nuclear (N), membrane (M), cytoskeleton (CS), and extracellular matrix (ECM) fractions. Finally, we combined *in vivo* ncAA labeling and cellular fractionation of embryos to investigate the turnover of proteins within each compartment. Significant labeling of newly synthesized proteins was observed by 3 h post-injection and persisted for varying durations depending on cellular compartment.

Future efforts will combine our *in vivo* labeling approach with that of embryo fractionation and LC-MS/MS analysis to precisely determine the turnover rates of individual proteins. To do so, it will first be necessary to characterize the metabolism of Aha *in vivo*, which remains unknown. In addition, we will need to normalize for changes in mass/protein content that occur during development. Together, these approaches will provide insight into functional tissue assembly by enabling the mapping of protein synthesis and turnover at specific developmental time points in different cellular compartments.

2.5 References

[1] N. Casas-Vila *et al.*, "The developmental proteome of Drosophila melanogaster," *Genome Res,* vol. 27, no. 7, pp. 1273-1285, Jul 2017.

[2] L. Sun, M. M. Bertke, M. M. Champion, G. Zhu, P. W. Huber, and N. J. Dovichi, "Quantitative proteomics of Xenopus laevis embryos: expression kinetics of nearly 4000 proteins during early development," *Sci Rep,* vol. 4, p. 4365, Mar 14 2014.

[3] M. B. Lucitt *et al.*, "Analysis of the zebrafish proteome during embryonic development," *Mol Cell Proteomics,* vol. 7, no. 5, pp. 981-94, May 2008.

[4] M. Y. Kessels *et al.*, "Proteomics analysis of the zebrafish skeletal extracellular matrix," *PLoS One,* vol. 9, no. 3, p. e90568, 2014.

[5] C. Williams, K. P. Quinn, I. Georgakoudi, and L. D. Black, 3rd, "Young developmental age cardiac extracellular matrix promotes the expansion of neonatal cardiomyocytes in vitro," *Acta Biomater,* vol. 10, no. 1, pp. 194-204, Jan 2014.

[6] T. Geiger *et al.*, "Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse," *Mol Cell Proteomics,* vol. 12, no. 6, pp. 1709-22, Jun 2013.

[7] X. Li *et al.*, "A time-resolved multi-omic atlas of the developing mouse stomach," *Nat Commun,* vol. 9, no. 1, p. 4910, Nov 21 2018.

[8] A. Naba, K. R. Clauser, S. Hoersch, H. Liu, S. A. Carr, and R. O. Hynes, "The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices," *Mol Cell Proteomics,* vol. 11, no. 4, p. M111 014647, Apr 2012.

[9] E. F. Fornasiero *et al.*, "Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions," *Nat Commun,* vol. 9, no. 1, p. 4230, Oct 12 2018.

[10] C. M. Warren, D. L. Geenen, D. L. Helseth, Jr., H. Xu, and R. J. Solaro, "Sub-proteomic fractionation, iTRAQ, and OFFGEL-LC-MS/MS approaches to cardiac proteomics," *J Proteomics,* vol. 73, no. 8, pp. 1551-61, Jun 16 2010.

[11] R. C. Hill, E. A. Calle, M. Dzieciatkowska, L. E. Niklason, and K. C. Hansen, "Quantification of extracellular matrix proteins from a rat lung scaffold to provide a molecular readout for tissue engineering," *Mol Cell Proteomics,* vol. 14, no. 4, pp. 961-73, Apr 2015.

[12] M. L. Decaris *et al.*, "Proteomic analysis of altered extracellular matrix turnover in bleomycin-induced pulmonary fibrosis," *Mol Cell Proteomics,* vol. 13, no. 7, pp. 1741-52, Jul 2014.

[13] A. Didangelos, X. Yin, K. Mandal, M. Baumert, M. Jahangiri, and M. Mayr, "Proteomics characterization of extracellular space components in the human aorta," *Mol Cell Proteomics,* vol. 9, no. 9, pp. 2048-62, Sep 2010.

[14] H. B. Schiller *et al.*, "Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair," *Mol Syst Biol,* vol. 11, no. 7, p. 819, Jul 14 2015.

[15] J. D. Bagert *et al.*, "Quantitative, time-resolved proteomic analysis by combining bioorthogonal noncanonical amino acid tagging and pulsed stable isotope labeling by amino acids in cell culture," *Mol Cell Proteomics,* vol. 13, no. 5, pp. 1352-8, May 2014.

[16] X. Chen, S. Wei, Y. Ji, X. Guo, and F. Yang, "Quantitative proteomics using SILAC: Principles, applications, and developments," *PROTEOMICS,* vol. 15, no. 18, pp. 3175-3192, 2015.

[17] D. B. McClatchy, M. Q. Dong, C. C. Wu, J. D. Venable, and J. R. Yates, 3rd, "15N metabolic labeling of mammalian tissue with slow protein turnover," *J Proteome Res,* vol. 6, no. 5, pp. 2005-10, May 2007.

[18] D. C. Dieterich, A. J. Link, J. Graumann, D. A. Tirrell, and E. M. Schuman, "Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT)," *Proc Natl Acad Sci U S A,* vol. 103, no. 25, pp. 9482-7, Jun 20 2006.

[19] K. L. Kiick, R. Weberskirch, and D. A. Tirrell, "Identification of an expanded set of translationally active methionine analogues in Escherichia coli," *FEBS Lett,* vol. 502, no. 1-2, pp. 25-30, Jul 27 2001.

[20] A. Borrmann and J. C. M. van Hest, "Bioorthogonal chemistry in living organisms," *Chem Sci,* 10.1039/C3SC52768A vol. 5, no. 6, pp. 2123-2134, 2014.

[21] V. V. Rostovtsev, L. G. Green, V. V. Fokin, and K. B. Sharpless, "A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes," *Angew Chem Int Ed Engl,* vol. 41, no. 14, pp. 2596-9, Jul 15 2002.

[22] D. C. Dieterich, J. J. Lee, A. J. Link, J. Graumann, D. A. Tirrell, and E. M. Schuman, "Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging," *Nat Protoc,* vol. 2, no. 3, pp. 532-40, 2007.

[23] K. L. Kiick, E. Saxon, D. A. Tirrell, and C. R. Bertozzi, "Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation," *Proc Natl Acad Sci U S A,* vol. 99, no. 1, pp. 19-24, Jan 8 2002.

[24] D. C. Dieterich *et al.*, "In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons," *Nat Neurosci,* vol. 13, no. 7, pp. 897-905, Jul 2010.

[25] A. M. Saleh, K. M. Wilding, S. Calve, B. C. Bundy, and T. L. Kinzer-Ursem, "Non-canonical amino acid labeling in proteomics and biotechnology," *J Biol Eng,* vol. 13, p. 43, 2019.

[26] D. B. McClatchy *et al.*, "Pulsed Azidohomoalanine Labeling in Mammals (PALM) Detects Changes in Liver-Specific LKB1 Knockout Mice," *J Proteome Res,* vol. 14, no. 11, pp. 4815-22, Nov 6 2015.

[27] S. Calve, A. J. Witten, A. R. Ocken, and T. L. Kinzer-Ursem, "Incorporation of non-canonical amino acids into the developing murine proteome," *Sci Rep,* vol. 6, p. 32377, Aug 30 2016.

[28] J. Cox and M. Mann, "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nat Biotechnol,* vol. 26, no. 12, pp. 1367-72, Dec 2008.

[29] M. Ashburner *et al.*, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat Genet,* vol. 25, no. 1, pp. 25-9, May 2000.

[30] A. Naba, K. R. Clauser, and R. O. Hynes, "Enrichment of Extracellular Matrix Proteins from Tissues and Digestion into Peptides for Mass Spectrometry Analysis," *Journal of visualized experiments : JoVE,* no. 101, pp. e53057-e53057, 2015.

[31] J. Reimand *et al.*, "g:Profiler-a web server for functional interpretation of gene lists (2016 update)," *Nucleic Acids Res,* vol. 44, no. W1, pp. W83-9, Jul 8 2016.

[32] A. Fabregat *et al.*, "The Reactome Pathway Knowledgebase," *Nucleic Acids Res,* vol. 46, no. D1, pp. D649-D655, Jan 4 2018.

[33] Y. Y. Yang, M. Grammel, A. S. Raghavan, G. Charron, and H. C. Hang, "Comparative analysis of cleavable azobenzene-based affinity tags for bioorthogonal chemical proteomics," *Chem Biol,* vol. 17, no. 11, pp. 1212-22, Nov 24 2010.

[34] J. A. Dunn *et al.*, "Reaction of ascorbate with lysine and protein under autoxidizing conditions: formation of N epsilon-(carboxymethyl)lysine by reaction between lysine and products of autoxidation of ascorbate," *Biochemistry,* vol. 29, no. 49, pp. 10964-70, Dec 11 1990.
[35] O. Reihl, M. O. Lederer, and W. Schwack, "Characterization and detection of lysine-arginine cross-links derived from dehydroascorbic acid," *Carbohydr Res,* vol. 339, no. 3, pp. 483-91, Feb 25 2004.
[36] V. Hong, S. I. Presolski, C. Ma, and M. G. Finn, "Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation," *Angew Chem Int Ed Engl,* vol. 48, no. 52, pp. 9879-83, 2009.
[37] B. Alvarez-Castelao *et al.*, "Cell-type-specific metabolic labeling of nascent proteomes in vivo," *Nat Biotechnol,* vol. 35, no. 12, pp. 1196-1201, Dec 2017.
[38] T. P. Krogager *et al.*, "Labeling and identifying cell-specific proteomes in the mouse brain," *Nat Biotechnol,* vol. 36, no. 2, pp. 156-159, Feb 2018.
[39] S. C. Munger, A. Natarajan, L. L. Looger, U. Ohler, and B. Capel, "Fine time course expression analysis identifies cascades of activation and repression and maps a putative regulator of mammalian sex determination," *PLoS Genet,* vol. 9, no. 7, p. e1003630, 2013.
[40] L. Martin-Rivera, E. Herrera, J. P. Albar, and M. A. Blasco, "Expression of mouse telomerase catalytic subunit in embryos and adult tissues," *Proc Natl Acad Sci U S A,* vol. 95, no. 18, pp. 10471-6, Sep 1 1998.
[41] E. Varela, R. P. Schneider, S. Ortega, and M. A. Blasco, "Different telomere-length dynamics at the inner cell mass versus established embryonic stem (ES) cells," *Proc Natl Acad Sci U S A,* vol. 108, no. 37, pp. 15207-12, Sep 13 2011.
[42] M. Buckingham *et al.*, "The formation of skeletal muscle: from somite to limb," *J Anat,* vol. 202, no. 1, pp. 59-68, Jan 2003.
[43] S. Siedner *et al.*, "Developmental changes in contractility and sarcomeric proteins from the early embryonic to the adult stage in the mouse heart," *J Physiol,* vol. 548, no. Pt 2, pp. 493-505, Apr 15 2003.
[44] G. E. Lyons, M. Ontell, R. Cox, D. Sassoon, and M. Buckingham, "The expression of myosin genes in developing skeletal muscle in the mouse embryo," *J Cell Biol,* vol. 111, no. 4, pp. 1465-76, Oct 1990.
[45] D. E. Birk and P. Brückner, "Collagens, Suprastructures, and Collagen Fibril Assembly," in *The Extracellular Matrix: an Overview*, R. P. Mecham, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 77-115.
[46] N. H. Brown, "Extracellular matrix in development: insights from mechanisms conserved between invertebrates and vertebrates," *Cold Spring Harb Perspect Biol,* vol. 3, no. 12, Dec 1 2011.
[47] S. Ying *et al.*, "Characterization and expression of the mouse lumican gene," *J Biol Chem,* vol. 272, no. 48, pp. 30306-13, Nov 28 1997.
[48] J. A. Paulo, "Sample preparation for proteomic analysis using a GeLC-MS/MS strategy," *Journal of biological methods,* vol. 3, no. 3, p. e45, 2016.
[49] A. Fabregat *et al.*, "The Reactome Pathway Knowledgebase," *Nucleic Acids Res,* vol. 46, pp. D649-D655, 2018.
[50] W. Rupik, K. Jasik, J. Bembenek, and W. Widłak, "The expression patterns of heat shock genes and proteins and their role during vertebrate's development," *Comparative Biochemistry and Physiology, Part A,* vol. 159, pp. 349-366, 2011.

[51] C. Dekker *et al.*, "The interaction network of the chaperonin CCT," *EMBO J,* vol. 27, pp. 1827-1839, 2008.
[52] J. Frydman, "Folding of newly translated proteins in vivo: the role of molecular chaperones," *Annu Rev Biochem,* vol. 70, pp. 603-647, 2001.
[53] A.-R. Kim and K.-W. Choi, "TRiC/CCT chaperonins are essential for organ growth by interacting with insulin/TOR signaling in Drosophila," *Oncogene,* 2019.
[54] W. Kuhlbrandt, "Structure and function of mitochondrial membrane protein complexes," *BMC Biol,* vol. 13, p. 89, 2015.
[55] H. L. Sweeney and D. W. Hammers, "Muscle Contraction," *Cold Spring Harb Perspect Biol,* vol. 10, 2018.
[56] T. Rozario and D. W. DeSimone, "The extracellular matrix in development and morphogenesis: a dynamic view," *Dev Biol,* vol. 341, pp. 126-140, 2010.

3. IDENTIFICATION AND VISUALIZATION OF NASCENT EXCTRACULLAR MATRIX PROTEINS IN THE DEVELOPING MOUSE

The content of this Chapter has been submitted in part for publication.

3.1 Introduction

The extracellular matrix (ECM) is the non-cellular component of multicellular organisms that defines tissue architecture and directs cell function [1], and is comprised of a complex network of collagens and non-collagenous glycoproteins, proteoglycans and polysaccharides. It is also a source of crucial biochemical and mechanical cues that instruct cell morphogenesis, proliferation, and adhesion [1, 2]. The ECM is highly dynamic and undergoes continuous controlled remodeling to maintain normal physiology. Dysregulation of ECM remodeling has been linked to several developmental abnormalities as well as various diseases such as fibrosis, cancer and musculoskeletal disorders [1, 3, 4]. Therefore, understanding the regulation of ECM synthesis and turnover is key to elucidating the mechanisms driving ECM-associated disorders. However, the complex nature of ECM proteins (large size, insolubility, heavy glycosylation, high cross-linking) has limited progress towards mapping ECM dynamics [5, 6].

ECM proteins, collectively termed the matrisome [7], make up a small percentage of the total protein content in most tissues. Accordingly, detection of ECM proteins using liquid chromatography-tandem mass spectrometry (LC-MS/MS) is limited without prior fractionation of tissues to remove intracellular and membrane-bound proteins. Recently, we demonstrated that tissue fractionation combined with LC-MS/MS is suitable to analyze the matrisome of embryonic tissues [8]; however, this only provides a static snapshot of the total matrisome at a given developmental time point. Knowledge of the proteins synthesized at a specific time point will

provide more information about protein dynamics during tissue development that may not be revealed in the static matrisome.

The non-canonical amino acid (ncAA) labeling technique has been used by our group and others to identify the newly synthesized proteome in embryonic, adult and pathological tissues [8-10]. The technique utilizes ncAAs that are incorporated into proteins using the endogenous cellular translational machinery [11]. These ncAAs possess biorthogonal moieties (*i.e.* azides) that enable the enrichment of newly synthesized proteins (NSPs) through click chemistry reactions with complementary chemical groups (*i.e.* alkynes) [12]. ncAA labeling has been previously used to enrich for soluble proteins, but the application of this technique to the insoluble matrisome has not yet been achieved.

In this Chapter, we describe for the first time a MS-based method that enables selective enrichment and identification of newly synthesized ECM proteins (NSEPs) in murine tissues. First, we demonstrate the successful application of the method to characterize the NSEPs of whole embryos. Next, we applied the method to forelimb tissues to study the dynamics of musculoskeletal development in embryonic and adolescent developmental time points. We show that this method enables high temporal resolution, which reveals ECM dynamics that could not be identified by traditional static proteomic analysis. Finally, we further extended the method to study the spatial distribution of NSEPs in decellularized forelimbs. The method described here can be broadly applied to study the temporal and spatial dynamics of ECM proteins in normal and pathological tissues.

3.2 Methods

3.2.1 Animal Models

Animals used in this study were derived from wild-type C57BL/6 mice (*Mus musculus)* purchased from The Jackson Laboratory. All experimental protocols were performed in compliance with established guidelines and all methods were approved by Purdue Animal Care and Use Committee (PACUC, protocols# 1209000723 and 1801001682). PACUC requires that all animal programs, procedures, and facilities at Purdue University to abide by the policies, recommendations, guidelines, and regulations of the USDA and the United States Public Health Service in accordance with the Animal Welfare Act and Purdue's Animal Welfare Assurance. To generate embryos of defined ages, female mice were time-mated with males and noon on the date when a copulation plug was found was considered to be embryonic day (E)0.5.

3.2.2 Aha Labeling and Tissue Collection

The methionine (Met) analog L-azidohomoalanine (Aha; Click Chemistry Tools) was resuspended in phosphate buffered saline (PBS; 10 mg/ml), pH adjusted to 7.4 with NaOH, sterile filtered and stored at -20 °C. All injections were administered to pregnant dams subcutaneously at 0.1 mg/g Aha and sterile PBS was used for control injections at 10 µl/g mouse. For whole embryos labeling and collection, pregnant dams were injected once a day with Aha or PBS at embryonic days (E) 13.5 and E14.5, and embryos were harvested 24 h after the last injection (E15.5). Embryos ($n = 2$ biological replicates) were collected by euthanizing dams via CO2 inhalation, which was confirmed using cervical dislocation. The uterine horns were removed and dissected in ice cold PBS, then embryos were snap frozen in liquid nitrogen and stored at -80 °C.

For embryonic forelimbs, pregnant dams were injected once subcutaneously at E13.5 with Aha or PBS. Dams were euthanized and forelimbs (*n = 3* biological replicates) were collected by dissecting tissues from the scapula to the digit tips following 6 h (E13.75) or 24 h (E14.5) of injection. For adolescent forelimbs, non-pregnant dams were injected once with Aha or PBS at postnatal day (P) 35 and forelimbs (*n = 3* biological replicates) were collected 6 h (P35.25) or 24 h (P36) post injection. Forelimb tissues were snap frozen in liquid nitrogen and stored at -80 °C.

3.2.3 Tissue Fractionation

E15.5 embryos, harvested from dams injected with Aha or PBS at E13.5 and E14.5, were fractionated using buffers of increasing stringency to selectively enrich for cytosolic (C), nuclear (N), membrane (M), cytoskeletal (CS) or ECM proteins as previously described [13] with some modifications (Table A.2). Embryos were mechanically homogenized with a TissueRuptor (Qiagen) in C buffer, rotated end-over-end for 1 h at 4 °C, followed by centrifugation at 21,100 × *g* for 20 min. The remaining pellet was resuspended in C buffer and processed as previously described. The pellet was then sequentially processed with N, M and CS buffers following the same protocol as described for the C buffer, with the exception that extractions using CS buffer were performed at RT. The remaining insoluble pellets (ECM) were snap frozen and stored at -80 °C.

For forelimb tissue fractionation, the Compartment Protein Extraction Kit (Millipore EMD) was used according to manufacturer instructions. In brief, forelimbs were homogenized with a TissueRuptor (Qiagen) in the cytosolic (C) buffer. Homogenized tissues were rotated end-over-end for 1 h at 4°C then centrifuged at 21,100 × *g* for 20 min at 4°C. Pellets were resuspended in the wash (W) buffer, rotated end-over-end for 5 minutes at 4°C, and centrifuged. Pellets were then

sequentially resuspended twice in the nuclear (N) buffer and once in the membrane (M) buffer, rotated end-over-end with each buffer for 30 minutes at 4°C, and centrifuged. Pellets were finally resuspended in cytoskeletal (CS) buffer, rotated end-over-end for 20 at RT, and centrifuged. The remaining pellet represented the ECM fraction was snap frozen and stored at -80°C. All buffers were supplemented with 1× Halt protease inhibitor cocktail (ThermoFisher Scientific) and 0.1% benzonase (Millipore EMD) except the CS buffer was supplemented with 1× Halt protease inhibitor cocktail only.

3.2.4 Enrichment of Aha-Labeled ECM Proteins

E15.5 embryos

ECM pellets obtained from the tissue fractionation protocol were resuspended in 8 M urea/100 mM ammonium bicarbonate buffer and sonicated 3× on ice for 10 sec using 50% duty cycle and output 3. Samples were centrifuged at 16,000 × *g* for 20 min at RT, the supernatant was transferred to a new tube and protein concentration was measured using the Pierce 660 nm Quantitative Colorimetric Assay (ThermoFisher Scientific). For click reaction, proteins were first alkylated with 45 mM iodoacetamide (IAA) for 30 min at RT in dark then reacted with (50 μm diazo biotin-alkyne, 10 mM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA; Click Chemistry Tools), 2 mM copper sulfate, 20 mM aminoguanidine, 5 mM sodium ascorbate) for 2-3 h at RT while rotating end-over-end. Following click reaction, ice-cold acetone was added to a final concentration of 80% to precipitate proteins and remove excess click reagents. Samples were incubated overnight at -20°C followed by centrifugation at 21,000 × *g* for 20 min at 4°C. Protein pellets were dried, resuspended in 4 M urea/100 mM ammonium bicarbonate buffer and centrifuged at 16,000 × *g* for 20 min at RT to remove insoluble particles. Supernatants were added

to 100 µl settled NeutrAvidin beads (ThermoFisher Scientific) previously washed 3× with 100 mM ammonium bicarbonate and rotated end-over-end for 1.5 hr at RT. Following incubation, beads were washed extensively first in (4 M urea, 0.1% SDS, 100 mM ammonium bicarbonate pH 8) then in (0.1% SDS, 1× PBS pH 7.15). Aha-labeled proteins were eluted by adding 400 µL (50 mM sodium dithionite, 0.1% SDS, 1× PBS pH 7.2) and rotating end-over-end for 1 hr at RT in dark. The first elution fraction was collected and the elution step was repeated twice. Elution fractions were combined and proteins were precipitated with acetone overnight at -20°C.

Forelimb tissues

Aha-labeled ECM proteins from forelimb tissues were enriched as described for E15.5 embryos with some modifications. ECM forelimb pellets were resuspended in Chondroitinase ABC digestion buffer (0.1 M Tris-HCl pH 8, 0.03 M sodium acetate, 1× Halt protease inhibitor cocktail). 0.2 U Chondroitinase ABC were added to each sample and incubated overnight at 37°C with constant agitation. Following incubation, proteins were acetone precipitated then pelleted by centrifugation at 21,100 × *g* for 20 min at 4°C. Dried pellets were resuspended in 8 M urea/100mM ammonium bicarbonate and sonicated 4× on ice for 10 sec using 50% duty cycle and output 3. Samples were centrifuged at 16,000 × *g* for 20 min, the supernatant was transferred to a new tube and protein concentration was measured using the Pierce 660 nm Quantitative Colorimetric Assay. Aliquots of each supernatant were saved, snap frozen and stored at -80°C. These aliquots represent the "unenriched proteome" of each sample. Click reaction, enrichment and elution of Aha-labeled proteins were then performed as indicated for E15.5 embryos.

3.2.5 Sample Preparation for LC-MS/MS Analysis

Aha-enriched and unenriched samples were denatured in (8 M urea/100 mM ammonium bicarbonate, pH 8) buffer, reduced with 10 mM DTT for 2 h at 37°C and then alkylated with 25 mM IAA for 30 minutes at RT in dark. Samples were diluted to 2 M urea with 100 mM ammonium bicarbonate pH 8 prior to the addition of 0.1 U Chondroitinase ABC and incubated for 2 h at 37°C. Samples were then digested at 37°C with: (1) 1 µg Endoproteinase LysC (New England Biolabs) for 2 h; (2) 3 µg MS-grade trypsin (ThermoFisher Scientific) for overnight; and (3) an additional 1.5 µg trypsin for 2 h. Afterwards, enzymes were inactivated with 0.1% trifluoroacetic acid (TFA; VWR). Peptides were cleaned from detergents using Pierce Detergent Removal Spin Columns (ThermoFisher Scientific) and desalted with C-18 MicroSpin Columns (The Nest Group Inc.). Peptides were dried for 4 h at 45°C in a centrivap concentrator and suspended in 10 µl of 3% acetonitrile (ACN) with 0.1% formic acid (FA). After suspension, peptide concentration was measured with the Pierce Quantitative Colorimetric Peptide Assay (ThermoFisher Scientific).

3.2.6 LC-MS/MS Analysis

Peptides were analyzed using the Dionex UltiMate 3000 RSLC Nano System coupled to the Q exactive HF hybrid quadrupole-orbitrap mass spectrometer (QE HF; ThermoFisher Scientific). Following digestion, peptides were loaded onto a 300 µm i.d. × 5 mm C18 PepMap 100 trap column and washed for 5 min using 2% ACN with 0.01% FA at a flow rate of 5 µL/min. After washing, the trap column was switched in-line with a 75 µm × 50 cm reverse phase Acclaim C18 PepMap 100 analytical column heated to 50 °C. Peptides were separated using a 120-min gradient elution method at a flow rate of 300 nl/min. Mobile phase A consisted of 0.01% FA in purified water, while mobile phase B consisted of 0.01% FA in 80% ACN. The linear gradient started at 2% B and reached 10% B in 5 min, 30% B in 80 min, 45% B in 91 min, and 100% B in 93 min.

The column was held at 100% B for the next 5 min before being brought back to 2% B and held for 20 min. Samples were injected into the QE HF through the Nanospray Flex Ion Source fitted with an emission tip from New Objective. Data acquisition was performed monitoring the top 20 precursors at 120,000 resolution with an injection time of 100 ms.

3.2.7 LC-MS/MS Data Search

Raw data files were analyzed using MaxQuant (version 1.6.1.0) [14]. Peak lists were searched against the *Mus musculus* Uniprot database (November 2018), *Gallus gallus* Avidin Uniprot protein sequence (May 2018) and a common contaminants database (January 2018). Cysteine carbamidomethylation was included as a fixed modification. Methionine oxidation, hydroxylysine, hydroxyproline, asparagine deamidation, glutamine to pyro-glutamic acid conversion, Aha substitution for methionine and cleaved tagged Aha substitution for methionine were included as variable modifications. Enzyme specificity was set to LysC and trypsin with up to 2 missed cleavages allowed. Peptide and protein false discovery rates (FDR) were set to 0.01 as determined by a reverse decoy database derived from the *Mus musculus* database. Proteins that had less than two razor and unique peptides or marked as a potential contaminant or reverse hit were removed. Proteins were only included in subsequent analyses if identified in at least two biological replicates. Raw intensities were $\log_2$-transformed and proteins were considered enriched if $[\log_2(\text{Aha raw intensity}) - \log_2(\text{PBS raw intensity})] > 1$ (indicating a > 2-fold change). Tissue compartment categories (cytosolic, nuclear, membrane, cytoskeletal, matrisome) were assigned to proteins using categorizations derived from the Gene Ontology (GO) Consortium [15] and the Matrisome Project [7]. GO analysis was performed using the database for annotation, visualization and integrated discovery (DAVID) [16]. GraphPad Prism 8 was used for data visualization.

3.2.8 Forelimb Decellularization, Staining and Imaging

Aha-labeled and PBS (control) E14.5 embryos were collected and forelimbs were dissected as described above. Forelimbs were rinsed in 1× PBS pH 7.4 for 1 h at RT, transferred to Tissue-Tek Cryomold (Sakura Finetek) and embedded in 1% low melting agarose (Sigma Aldarich) to maintain tissue architecture. Agarose-embedded forelimbs were rocked in (0.05% SDS, 0.02% sodium azide, 1× PBS) decellularization solution. The solution was replaced every day until a complete decellularization was achieved after 5 days. Following decellularization, samples were fixed overnight with 4% paraformaldehyde, rinsed in 1× PBS and stored at 4°C until use.

For visualization of Aha-labeled ECM proteins, fixed forelimbs were alkylated overnight with 30 mM IAA to block cysteine residues and decrease background staining [17]. Alkylated samples were rinsed 3× in 1× PBS and incubated with 30 μm Alexa Fluor (AF)647 DBCO (Click Chemistry Tools) overnight at 4°C. For immunohistochemistry staining, samples were rinsed 4× in 1× PBS and incubated in blocking buffer (10% donkey serum, 0.2% bovine serum albumin, 1× PBS) for 1 h at RT. Samples were then incubated with a primary antibody against elastin microfibril interface-1 (Emilin1) (1:50; Novus Biologicals) for two days at 4°C. Samples were rinsed 3× in 1× PBS and incubated for two days at 4°C with donkey anti-rabbit AF546 (1:500; Invitrogen) to probe for Emilin1 and AF488 conjugate of wheat germ agglutinin (WGA; 1:100; Invitrogen) to visualize proteoglycans. Samples were rinsed 3× in 1× PBS and stored at 4°C protected from light until imaging. For image acquisition, samples were transferred to 8 well chamber slide (Ibidi) and imaged using an inverted Zeiss LSM 880 confocal microscope (Carl Zeiss). Images were analyzed using ImageJ (National Institute of Health).

3.3 Results and Discussion

3.3.1 Enrichment of Newly Synthesized ECM Proteins from Whole Embryos

The complex biochemical nature of the ECM renders attempts to identify and quantify newly synthesized ECM proteins (NSEPs) very challenging. In particular, ECM proteins are large in size, cross-linked, heavily modified with polysaccharides chains, and highly insoluble compared to intracellular proteins [13]. Recently, several studies have applied MS-based techniques to provide a more comprehensive analysis of the ECM composition in various tissues [7, 18, 19]. These studies have shown that breaking down large insoluble ECM proteins into small peptide fragments by enzymatic digestion results in a clear solution without the need of prior full solubilization of the complex ECM protein mixture. Adopting this technique to enrich for NSEPs, however, is not optimal because it entails digesting Aha-labeled proteins into peptides then clicking and enriching only peptides containing Aha residues. This and the fact that Met is a relatively uncommon amino acid [20] limit the attainment of comprehensive profiling of NSEPs due to overlooking the more abundant non Aha-containing peptides.

Accordingly, a more effective strategy to study NSEPs would require enriching labeled samples at the protein level rather than the peptide level; an approach that has been hindered by the insolubility of the ECM proteins. In this regard, 2-day labeled whole embryos at embryonic day (E) 15.5 were used to assess the feasibility of enriching Aha-labeled ECM at the protein level. The embryos were fractionated using buffers of increasing stringency to extract the ECM-rich protein pellets as discussed in Section 2.3. Prior to click reaction, ECM pellets were resuspended in a strong denaturing buffer (8 M urea) to partially solubilize the highly insoluble ECM proteins. To further enhance the solubility and extraction of urea-insoluble proteins, the resuspended pellets were then subjected to brief cycles of ultrasonic energy (Figure 3.1A). Notably, the combination

of urea solubilization and ultrasonication resulted in very small remaining pellets, indicating that most of the ECM proteins went into solution.

Solubilized ECM proteins were clicked with a cleavable diazo biotin-alkyne linker and Aha-labeled proteins were enriched via NeutrAvidin beads, eluted by selective cleavage of the diazo group and analyzed by LC-MS/MS. A total of 216 intra- and extracellular proteins were identified. Proteins were considered newly synthesized if exclusively detected in Aha-labeled samples or > 2-fold enriched in Aha-labeled compared to control samples (Figure 3.1B). This analysis assigned 207 proteins as newly synthesized (NSPs), of which 52 proteins were ECM. Notably, the raw intensity of the NSEPs comprised about 57% of the total protein intensity (Figure 3.1C). In addition, Reactome Pathway analysis of the 207 NSPs revealed that several ECM-specific terms such as ECM organization, collagen chain trimerization and collagen degradation were significantly enriched (Figure 3.1D).

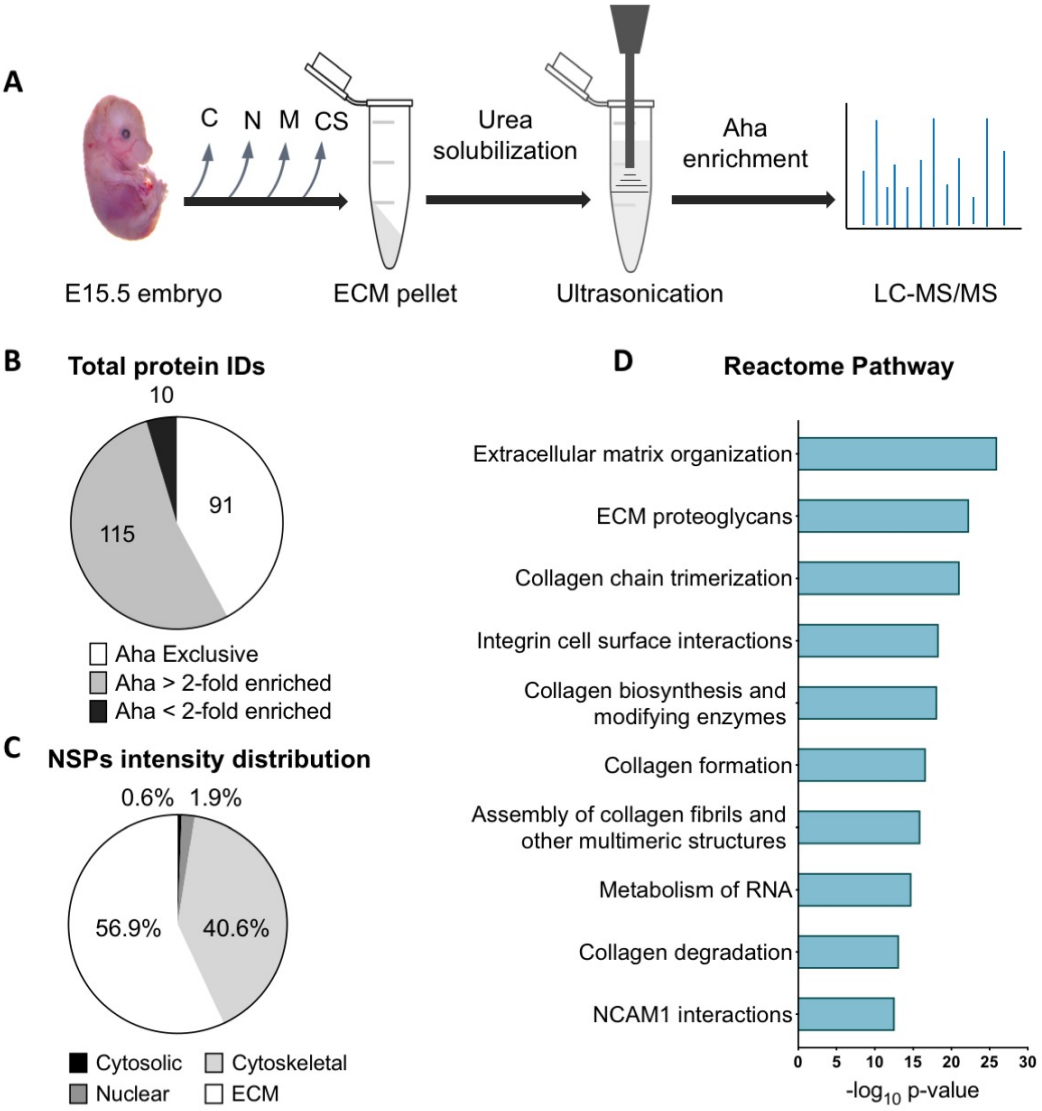
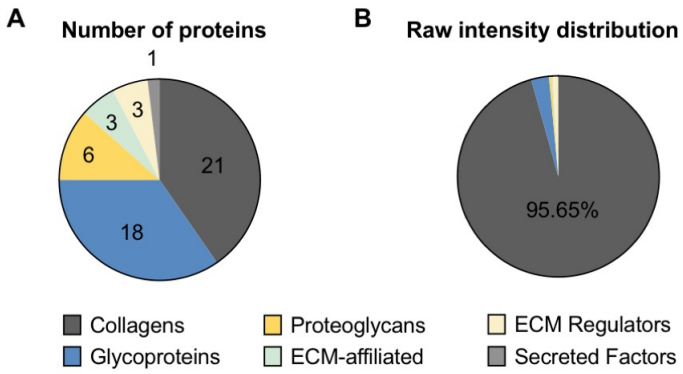


Figure 3.1 LC-MS/MS analysis of Aha-labeled ECM proteins isolated from E15.5 embryos. **(A)** Experimental work flow combining tissue fractionation of E15.5 embryos to isolate ECM, urea solubilization and ultrasonication to enhance the solubility of ECM proteins, enrichment of Aha-labeled proteins and LC-MS/MS analysis. **(B)** Number of proteins identified by LC-MS/MS analysis either exclusively, > or < 2-fold enriched in Aha-labeled embryos compared to control. **(C)** Percentage of the total raw intensity distribution of different cellular fractions for the identified newly synthesized proteins (NSPs). The membrane fraction is not shown on the pie chart as it comprised only 0.03% of the total intensity. **(D)** Top 10 enriched terms of the Reactome Pathway analysis of the NSPs. Note the significant enrichment of multiple ECM-specific processes.

The ensemble of ECM proteins, termed the matrisome, is further classified into two categories: (1) the core matrisome proteins (comprising collagens, glycoproteins and

proteoglycans) and (2) the matrisome-associated proteins (comprising ECM-affiliated proteins, ECM regulators and ECM secreted factors) [7]. Each of these matrisome components has a distinctive biochemical composition. For example, the majority of core matrisome proteins are extensively post-translationally modified by hydroxylation, sulfation and glycosylation [6]. In addition, collagens and some glycoproteins undergo post-translational cross-linking by disulfide bonding and tissue transglutaminases, which further contributes to their insolubility [5]. On the contrary, several matrisome-associated proteins such as growth factors and ECM-remodeling enzymes are more soluble and easier to extract [13].

To evaluate whether the method is able to enrich for the different ECM components or limited to the more soluble, less modified proteins, the 52 identified NSEPs were assigned to their matrisome categories. This analysis revealed that each matrisome category was detected in our dataset (Figure 3.2A). Notably, 46 proteins belonged to the core matrisome, whereas 6 proteins were matrisome-associated. Collagens, the most abundant ECM proteins, comprised about 96% of the total raw intensity (Figure 3.2B). Additionally, several NSEPs, including some of the most complex collagens and glycoproteins, were identified with a sequence coverage greater than 25% (Figure 3.2C).



C

Core Matrisome

Collagens

Gene name	MW [kDa]	Peptides number	Sequence coverage
Col1a1	138.03	58.5	56.95
Col1a2	129.56	48.5	56.15
Col2a1	141.97	40	41.1
Col3a1	138.94	38	35.6
Col4a1	149.16	3	3.5
Col4a2	167.32	6	5.3
Col5a1	183.67	11	11.8
Col5a2	145.02	23.5	32.35
Col6a1	108.49	17	19
Col6a2	110.33	12	14.3
Col6a3	353.93	54.5	21.95
Col6a6	246.42	9.5	6.9
Col9a1	92.094	8	16.35
Col9a2	65.313	4	10
Col9a3	63.474	5.5	15.2
Col11a1	181.03	14	14.3
Col11a2	171.53	10	8.5
Col12a1	333.73	42.5	21.1
Col14a1	192.75	5	4.1
Col16a1	155.8	3	3.75
Col26a1	45.809	2.5	9

Glycoproteins

Gene name	MW [kDa]	Peptides number	Sequence coverage
Comp	80.674	3	4.6
Emilin1	107.58	18	25.2
Emilin3	77.395	3	6.9
Fbn1	312.3	3	0.85
Fbn2	313.82	2	0.6
Fga	87.428	6	8.4
Fgb	54.752	3	10
Fgg	50.349	4	12.2
Fn1	262.84	32.5	20.45
Lama5	404.05	3	1.1
Lamb1	202.44	5	3.5
Lamc1	177.19	9	6.8
Matn1	54.421	19.5	45
Matn3	51.845	2.5	8.25
Matn4	64.68	12.5	21.15
Nid2	128.41	6	8.7
Postn	93.142	19.5	32.15
Tnc	231.8	15.5	10.8

Proteoglycans

Acan	221.94	5	3.2
Bgn	41.639	2	8.7
Dcn	39.809	3	9.9
Epyc	36.762	2	5
Hapln1	40.477	5	16
Hspg2	469.02	21	8.35

Matrisome-associated

ECM-affiliated

Gene name	MW [kDa]	Peptides number	Sequence coverage
Anxa1	38.734	2.5	9.7
Anxa2	21.856	2	14.8
Frem2	350.64	4	2.2

ECM regulators

Serpinh1	46.533	12	36.5
Tgm2	77.06	10	18.2
Try10	26.221	2	14.6

Secreted factors

Tgfbi	74.596	5.5	12.4

Value low to high

Figure 3.2 Matrisome categorization of the identified newly synthesizes ECM proteins (NSEPs). **(A)** NSEPs were classified into collagens, glycoproteins, proteoglycans, ECM-affiliated proteins, ECM-regulators and secreted factors, and the number of proteins identified in each matrisome category was estimated. **(B)** Percentage of the total raw intensity distribution of the different matrisome categories. Collagens, glycoproteins, proteoglycans, ECM regulators, ECM-affiliated and secreted factors comprised 95.65%, 2.86%, 0.43%, 0.84%, 0.19% and 0.04% of the total raw intensity, respectively. **(C)** Table of the number of identified peptides and the percentage of the protein sequence covered by the identified peptides for each NSEP. Proteins are denoted by their gene symbols. Color key indicates value from low (yellow) to high (blue).

3.3.2 Mapping ECM Dynamics During Forelimb Development Using ncAA Labeling

Having established that the method can enrich and identify the newly synthesized ECM proteins in developing embryos, we next investigated whether it can resolve the temporal dynamics of ECM synthesis and degradation during forelimb musculoskeletal development. In the musculoskeletal system, the ECM plays a pivotal role by providing crucial biochemical cues that direct cell morphogenesis and proliferation. It is also essential for maintaining the structural integrity of tissues in response to mechanical stimuli to ensure effective force transmissions during muscle contraction [21]. The ECM mediates its function by undergoing continuous remodeling in its composition in adaptation to the various instructive and structural signals [1]. This dynamic fine-tuning of the synthesis and turnover of ECM proteins ultimately orchestrates the formation of functional musculoskeletal tissues.

In essence, the regeneration mechanisms through which injured adult muscles restore their functionality emulate the native biological processes occurring during embryonic skeletal muscle development. [22, 23]. Therefore, mapping the change in ECM composition as a function of development is essential for designing regenerative therapies that aim at restoring the functionality of damaged tissues in the musculoskeletal system. In such studies, the employed method should have high temporal resolution to capture the early changes in protein composition occurring during different developmental stages. To this end, the forelimb study was designed to (1) enrich the small fraction of the new ECM proteins synthesized following a short labeling window (6 h post Aha injection (hpi)) in the complex musculoskeletal system that is composed of bone, muscle, tendons, ligaments, and cartilage, (2) resolve differences in the new proteins synthesized within adjacent developmental time points (6 and 24 hpi), (3) establish a baseline of the ECM dynamics during musculoskeletal development in embryonic compared to adolescent time points, and (4)

demonstrate that the ncAA labeling technique reveals ECM dynamics that cannot be resolved by traditional static proteomic analysis.

To expand the Aha enrichment method to forelimb tissues, we first considered the ECM composition in the musculoskeletal system. It has been demonstrated that proteoglycans, macromolecules comprised of a core protein attached to glycosaminoglycan chains, play a critical role in the modulation of limb skeletal muscle differentiation [24, 25]. Proteoglycans are classified according to the type of glycosaminoglycan chains they possess into chondroitin sulfate, dermatan sulfate, keratan sulfate, heparin sulfate and heparan sulfate proteoglycans [26]. Among proteoglycans, chondroitin sulfate proteoglycans are present within the different tissues of the musculoskeletal system and are essential for chondrogenesis and joint morphogenesis [24, 27-29]. Typically, the proteoglycan backbone is heavily modified by glycosaminoglycan chains [30]. For example, aggrecan, the major cartilage proteoglycan, has about 100 chondroitin sulfate chains, each is around 20 kDa [31]. Such degree of glycosylation could affect the accessibility of Aha residues to the clickable tags and hence decreases the efficiency of enrichment of Aha-labeled proteins. To optimize the technique for musculoskeletal tissues, we therefore added a deglycosylation step before the click reaction to digest chondroitin sulfate chains.

Using the optimized method, we sought to determine which proteins are being actively synthesized at different stages of musculoskeletal development. Specifically, pregnant mice were injected at embryonic day (E) 13.5 with either Aha or PBS (control) and forelimbs were harvested from embryos 6 and 24 hpi (E13.75 and E14.5, respectively) to identify proteins synthesized during morphogenesis. In addition, non-pregnant mice were injected with either Aha or PBS at postnatal day (P) 35 and harvested 6 and 24 hpi (P35.25 and P36, respectively) to identify proteins synthesized during late growth. Forelimbs were fractionated to isolate the ECM and the ECM

fraction was deglycosylated with chondroitinase ABC enzyme. Deglycosylated proteins were solubilized as described for whole embryos and split into two samples: (1) an "unenriched" sample that represented the static proteome containing both old and newly synthesized proteins and (2) an "enriched" sample from which the NSPs were isolated. The proteome of unenriched and enriched samples was analyzed by LC-MS/MS (Figure 3.3A).

LC-MS/MS analysis of the embryonic time points

Analysis of the static proteome of the E13.75 samples identified 57 ECM proteins, all of which were among the 70 ECM proteins identified in the E14.5 samples (Figure 3.3B). In contrast, analysis of the Aha-enriched samples identified NSEPs that were exclusive to each time point (Figure 3.3C), indicating that Aha enrichment enabled detecting the small fraction of the nascent proteins synthesized following a narrow labeling window as well as proteins that were distinctively synthesized within adjacent developmental time points. In addition, laminin subunit alpha-5 (Lama5) was identified in the enriched proteins of the E13.5 – E14.5 labeling window (Figure 3.3C) but not identified in the static proteome of the E14.5, thus signifying the capability of this technique to capture and identify NSPs that are in lower abundance. Notably, Lama5 is critical for digit septation and embryos lacking it showed syndactyly at E14.5 [32].

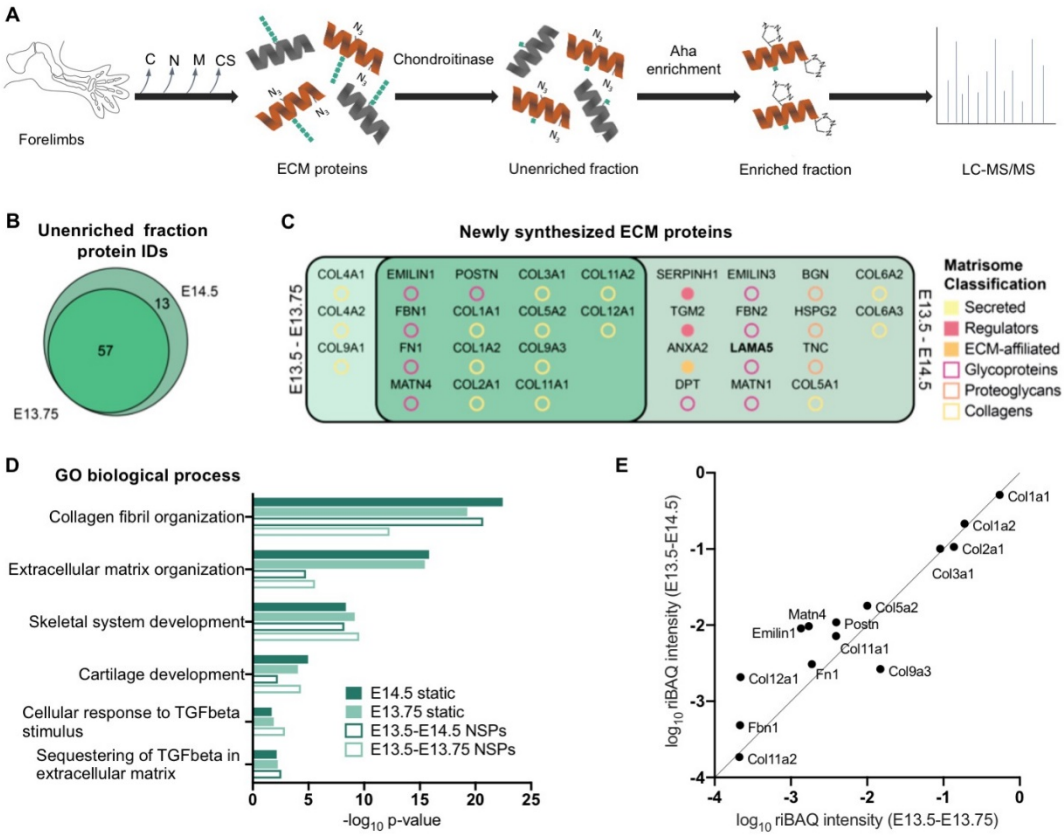


Figure 3.3 Newly synthesized ECM proteins in embryonic (E13.5 – E14.5) forelimbs. **(A)** Experimental workflow combining *in vivo* Aha-labeling, tissue fractionation, deglycosylation with chondroitinase ABC, protein solubilization and enrichment of Aha-labeled ECM proteins for LC-MS/MS analysis. **(B)** Venn diagrams comparing the number of ECM proteins identified in the unenriched fraction at each time point. **(C)** List of the newly synthesized ECM proteins, both unique and shared, between E13.5 – E13.75 and E13.5 – E14.5 labeling windows in the embryonic musculoskeletal system. **(D)** Selected biological process Gene Ontology (GO) terms of the static and newly synthesized E13.5 – E14.5 proteomes **(E)** Correlation plot comparing the relative iBAQ (riBAQ) intensity of the newly synthesized ECM proteins identified in the two embryonic time points.

Gene ontology (GO) analysis was performed to identify the biological processes associated with the Aha-enriched proteomes (Figure 3.3D). Several significant biological processes terms were shared between the two embryonic time points such as collagen fibril organization, skeletal system development and cartilage development. Yet, some of these GO terms showed distinctions in the group of NSEPs linked to the annotated process. For example, collagens (Col) Col2a1,

Col11a1 and Col11a2 were associated with the cartilage development term for the E13.5 – E14.5 labeling time point. The same set of proteins were identified for the E13.5 – E 13.75 time point in addition to Col9a1, which was exclusively detected at this time point (Figure 3.3C). Noticeably, Col9a1 has been shown to be essential for proper cartilage development [33, 34]. Furthermore, the GO analysis revealed terms unique to every embryonic time point due to the distinctive proteins identified in each. Particularly, cellular response to transforming growth factor beta (TGFbeta) stimulus was a significantly enriched term for the E13.5 – E 13.75 time point. Interestingly, sequestering of TGFbeta in extracellular matrix was a significant term for the E13.5 – E 14.5 time point. Markedly, both cellular response to TGFbeta and sequestering of TGFbeta in extracellular matrix were significantly enriched terms for the static proteome of the two embryonic time points (Figure 3.3D). Together, these results substantiate that Aha-labeling and enrichment of NSEPs enabled identification of ECM dynamics that were not revealed by the static proteome.

Identifying key ECM components at different stages of development as well as the relative composition of these components is critical for understanding tissue regeneration mechanisms during repair. In this regard, the intensity-based absolute quantification (iBAQ) algorithm was used to examine the relative abundance of the NSEPs identified at both embryonic time points. The iBAQ value is estimated by dividing protein intensity by the number of protein's theoretical tryptic peptides, and has on average a high accuracy for estimating protein abundance [35]. For relative abundance measurements, a relative iBAQ value (riBAQ) was calculated by dividing the iBAQ of each shared NSEP by the sum of iBAQ values for all NSEPs identified in an embryonic time point [36, 37]. Remarkedly, over 24 h of embryonic development, the relative abundance of several NSEPs such as Col12a1, matrilin-4 (Matn4) and elastin microfibril interface-1 (Emilin1) varied significantly ($p < 0.05$) between the two embryonic time points, whereas other NSEPs

remained unchanged (Figure 3.3E). This analysis further validates the efficiency of the method in understanding the regulation of the composition and abundance of the nascent proteins during development with high temporal resolution.

LC-MS/MS analysis of the adolescent time points

Analysis of Aha-labeled adolescent forelimb proteins revealed late development-related ECM dynamics. Particularly, of the 58 ECM proteins identified in both P35.25 and P36 unenriched samples, 41 proteins (~70%) were shared between the two time points (Figure 3.4A). In contrast, out of the total 33 NSEPs detected in the P35.25 and P36 Aha-enriched samples, only 6 NSPs (~18%) were found in both time points (Figure 3.4B). Notably, the newly synthesized prolargin (Prelp) protein was exclusively detected in the P35 – P36 labeling window although found in both P35.25 and P36 unenriched samples. Prelp is a proteoglycan that regulates bone development and remodeling by inhibiting late-stage osteoclast differentiation [38]. In addition, the relative abundance of most of the shared NSEPs (Col1a1, Col1a2, Col2a1, dermatopontin (Dpt)) was not significantly different ($p > 0.05$) between the adolescent time points (Figure 3.4C), suggesting that the change in some ECM components is less dynamic during postnatal forelimb development, which is a characteristic of late growth [39].

Overall, 16 NSEPs were unique to the embryonic forelimbs, 18 were only found in the adolescent and 15 were shared between embryonic and adolescent time points (Figure 3.4D). About 67% of the shared proteins were collagens (Col1a1, Col1a2, Col2a1, Col3a1, Col4a2, Col5a1, Col5a2, Col6a2, Col6a3 and Col12a1), signifying that they are critical for musculoskeletal tissue formation. Notably, collagen is the major ECM protein in musculoskeletal tissues [40]. Types I, III, IV, V and VI collagen have been shown to be expressed in skeletal muscles throughout

embryonic and postnatal development [41, 42]. In embryonic forelimbs, ECM proteins associated with morphogenesis such as fibronectin-1 (Fn1), fibrillin-2 (Fbn2) and Emilin1 were among the exclusively detected. In the early stages of morphogenesis, these proteins set the stage for various ECM networks to polymerize and facilitate the formation of distinct tissues [43-45]. In contrast, some of the NSEPs unique to adolescent tissues such as cartilage intermediate layer protein-2 (Cilp2) and Prelp are associated with bone maturation and homeostasis [38, 46], indicating that skeletal elements are approaching maturity.

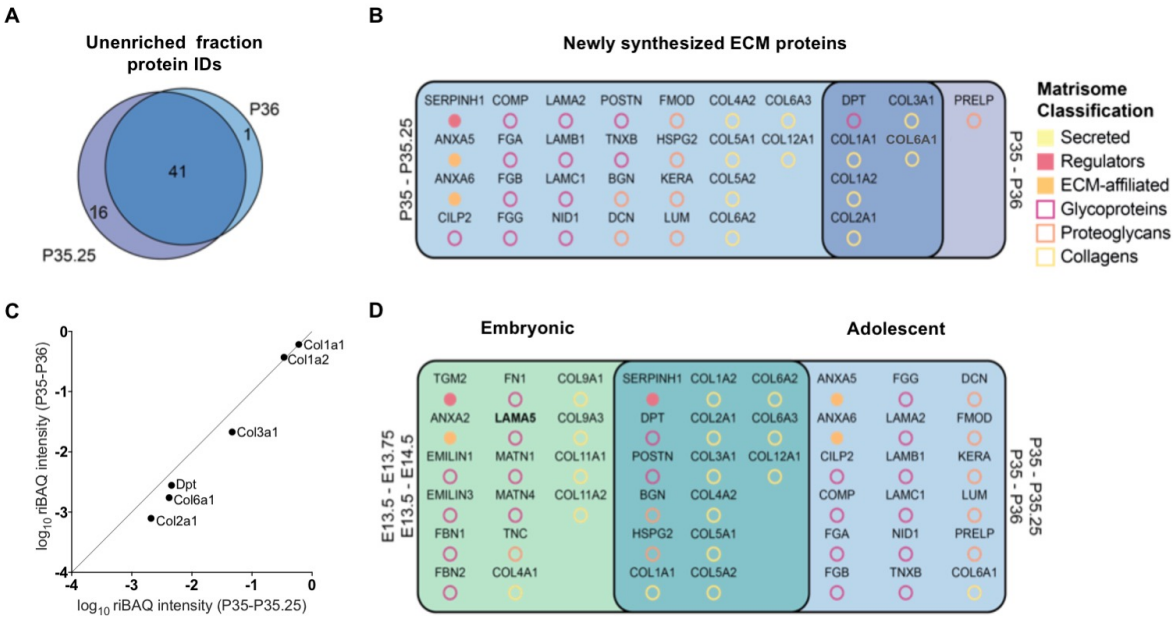


Figure 3.4 Newly synthesized ECM proteins in adolescent (P35 – P36) forelimbs **(A)** Venn diagrams comparing the number of ECM proteins identified in the unenriched fraction at each time point. **(B)** List of the newly synthesized ECM proteins, both unique and shared, between P53 – P35.25 and P35 – P36 labeling windows in the adolescent musculoskeletal system. **(C)** Correlation plot comparing the relative iBAQ (riBAQ) intensity of the newly synthesized ECM proteins identified in the two adolescent time points. **(D)** Identified newly synthesized ECM proteins, both unique and shared, in embryonic and adolescent forelimbs.

3.3.3 Spatial Dynamics of Newly Synthesized ECM Forelimb Proteins

Monitoring the localization and tracking the fate of newly synthesized ECM within forelimb tissues would provide insights into the spatial organization of ECM proteins in the musculoskeletal system in both development and disease. A key advantage of Aha labeling is that labeled proteins can be covalently attached to small-size fluorophores bearing alkyne functionalities and hence allowing selective, irreversible conjugation without perturbing protein activity or spatial distribution. To establish the feasibility of visualizing NSEPs, Aha-labeled E14.5 forelimbs were used. Tissues were decellularized by treatment with 0.05% SDS for several days to remove the intracellular components and enhance the visibility of ECM. For nascent protein visualization, fixed decellularized tissues were reacted with Alexa Fluor (AF)647 conjugated to the strain-promoted alkyne dibenzocyclooctyne (DBCO) (DBCO; blue). To confirm that decellularization retained the native ECM architecture, tissues were counterstained with the ECM glycoprotein Emilin1 (red) and AF488-conjugated wheat germ agglutinin (WGA; green), which binds to sialic acid and N-acetylglucosamine residues of proteoglycans. Fluorescent imaging of labeled and control forelimbs indicated that NSPs were selectively tagged with AF647-DBCO with minimal background signal detected in control samples (Figure 3.5). In addition, successful staining with Emilin1 and WGA following click reaction denotes that clicked samples can be immunolabeled using established techniques to examine the spatial distribution of ECM components of interest.

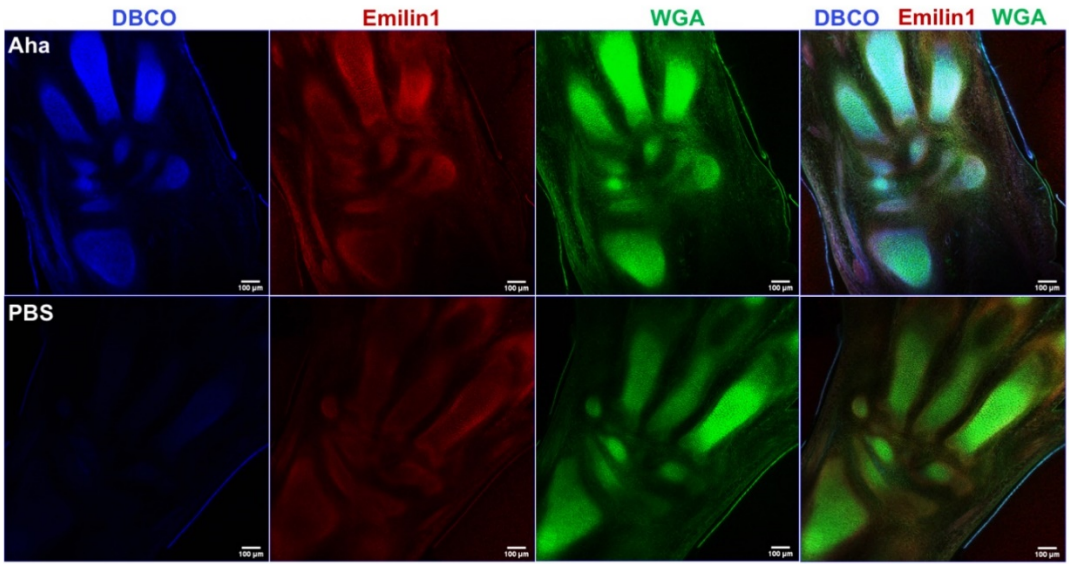


Figure 3.5 Fluorescent imaging of the musculoskeletal system in decellularized E14.5 forelimbs. Aha-treated and PBS (control) E14.5 forelimbs were decellularized and reacted with AF647-DBCO (DBCO; blue) followed by staining with an antibody against Emilin1 (red) and WGA (green). Bar = 100 μm.

3.4 Conclusions

Despite the fundamental role the ECM plays in maintaining proper cellular functions, our understanding of the dynamics of ECM regulation in development and disease remains lacking. This is primarily attributed to the complex biochemical nature of the ECM (insolubility, high molecular weight, glycosylation and cross-linking) that has hindered comprehensive temporal profiling of the nascent ECM proteome. In this study, we report for the first time the development of MS-based method that enables the analysis of ncAA-labeled newly synthesized ECM proteins *in vivo*. We show that the method enables enriching and identifying NSEPs in developing embryos with low background from unlabeled proteins while covering the different matrisome categories. Applying this method to forelimbs enabled detecting the key ECM components that play roles in musculoskeletal tissues at different stages of embryonic and adolescent development. The results demonstrate that the method has high temporal resolution as indicated by resolving the small

fraction of new proteins synthesized following only 6 h of Aha injection as well as those synthesized within adjacent developmental time points. Comparing our method to traditional static proteomic analysis, we show that Aha labeling revealed ECM dynamics that could not be captured by the static proteome. In addition, we further extended the labeling technique to visualize the NSEPs in decellularized forelimb tissues. Future studies will focus on combining identifying Aha-labeled ECM proteins using MS and tracking their spatial distribution as a function of musculoskeletal development. Employing these methods will provide substantial insights into the mechanisms of musculoskeletal repair and guide the design of regenerative scaffolds.

3.5 References

[1] T. Rozario and D. W. DeSimone, "The extracellular matrix in development and morphogenesis: a dynamic view," *Dev Biol,* vol. 341, no. 1, pp. 126-40, May 1 2010.

[2] N. Felsenthal and E. Zelzer, "Mechanical regulation of musculoskeletal system development," *Development,* vol. 144, no. 23, pp. 4271-4283, Dec 1 2017.

[3] C. Bonnans, J. Chou, and Z. Werb, "Remodelling the extracellular matrix in development and disease," *Nat Rev Mol Cell Biol,* vol. 15, no. 12, pp. 786-801, Dec 2014.

[4] R. V. Iozzo and M. A. Gubbiotti, "Extracellular matrix: The driving force of mammalian diseases," *Matrix Biol,* vol. 71-72, pp. 1-9, Oct 2018.

[5] R. O. Hynes and A. Naba, "Overview of the matrisome--an inventory of extracellular matrix constituents and functions," *Cold Spring Harb Perspect Biol,* vol. 4, no. 1, p. a004903, Jan 1 2012.

[6] A. Naba, K. R. Clauser, H. Ding, C. A. Whittaker, S. A. Carr, and R. O. Hynes, "The extracellular matrix: Tools and insights for the "omics" era," *Matrix Biol,* vol. 49, pp. 10-24, Jan 2016.

[7] A. Naba, K. R. Clauser, S. Hoersch, H. Liu, S. A. Carr, and R. O. Hynes, "The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices," *Mol Cell Proteomics,* vol. 11, p. M111 014647, 2012.

[8] A. M. Saleh, K. R. Jacobson, T. L. Kinzer-Ursem, and S. Calve, "Dynamics of Non-Canonical Amino Acid-Labeled Intra- and Extracellular Proteins in the Developing Mouse," *Cell Mol Bioeng,* vol. 12, no. 5, pp. 495-509, Oct 2019.

[9] D. B. McClatchy *et al.*, "Pulsed Azidohomoalanine Labeling in Mammals (PALM) Detects Changes in Liver-Specific LKB1 Knockout Mice," *J Proteome Res,* vol. 14, no. 11, pp. 4815-22, Nov 6 2015.

[10] D. B. McClatchy, Y. Ma, D. A. Liem, D. C. M. Ng, P. Ping, and J. R. Yates, "Quantitative temporal analysis of protein dynamics in cardiac remodeling," *Journal of Molecular and Cellular Cardiology,* vol. 121, pp. 163-172, 2018.

[11] D. C. Dieterich, A. J. Link, J. Graumann, D. A. Tirrell, and E. M. Schuman, "Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT)," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 103, pp. 9482-9487, 2006.
[12] A. M. Saleh, K. M. Wilding, S. Calve, B. C. Bundy, and T. L. Kinzer-Ursem, "Non-canonical amino acid labeling in proteomics and biotechnology," *Journal of Biological Engineering,* vol. 13, p. 43, 2019.
[13] A. Naba, K. R. Clauser, and R. O. Hynes, "Enrichment of extracellular matrix proteins from tissues and digestion into peptides for mass spectrometry analysis," *Journal of Visualized Experiments,* p. e53057, 2015.
[14] J. Cox and M. Mann, "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nature Biotechnology,* vol. 26, pp. 1367-1372, 2008.
[15] M. Ashburner *et al.*, "Gene ontology: Tool for the unification of biology," in *Nature Genetics* vol. 25, ed, 2000, pp. 25-29.
[16] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols,* vol. 4, pp. 44-57, 2009.
[17] R. Van Geel, G. J. M. Pruijn, F. L. Van Delft, and W. C. Boelens, "Preventing thiol-yne addition improves the specificity of strain-promoted azide-alkyne cycloaddition," *Bioconjugate Chemistry,* vol. 23, pp. 392-398, 2012.
[18] J. D. Hebert *et al.*, "Proteomic Profiling of the ECM of Xenograft Breast Cancer Metastases in Different Organs Reveals Distinct Metastatic Niches," *Cancer Res,* Feb 4 2020.
[19] A. Naba, K. R. Clauser, D. R. Mani, S. A. Carr, and R. O. Hynes, "Quantitative proteomic profiling of the extracellular matrix of pancreatic islets during the angiogenic switch and insulinoma progression," *Sci Rep,* vol. 7, p. 40495, Jan 10 2017.
[20] F. J. Veredas, F. R. Canton, and J. C. Aledo, "Methionine residues around phosphorylation sites are preferentially oxidized in vivo under stress conditions," *Sci Rep,* vol. 7, p. 40403, Jan 12 2017.
[21] M. Kjaer, "Role of extracellular matrix in adaptation of tendon and skeletal muscle to mechanical loading," *Physiol Rev,* vol. 84, no. 2, pp. 649-98, Apr 2004.
[22] A. J. Wagers and I. M. Conboy, "Cellular and molecular signatures of muscle regeneration: current concepts and controversies in adult myogenesis," *Cell,* vol. 122, no. 5, pp. 659-67, Sep 9 2005.
[23] M. Karalaki, S. Fili, A. Philippou, and M. Koutsilieris, "Muscle regeneration: cellular and molecular events," *In Vivo,* vol. 23, no. 5, pp. 779-96, Sep-Oct 2009.
[24] H. Olguin and E. Brandan, "Expression and localization of proteoglycans during limb myogenic activation," *Dev Dyn,* vol. 221, no. 1, pp. 106-15, May 2001.
[25] E. Brandan and J. Gutierrez, "Role of skeletal muscle proteoglycans during myogenesis," *Matrix Biol,* vol. 32, no. 6, pp. 289-97, Aug 8 2013.
[26] J. D. Esko, K. Kimata, and U. Lindahl, "Proteoglycans and Sulfated Glycosaminoglycans," in *Essentials of Glycobiology*, nd *et al.*, Eds. Cold Spring Harbor (NY), 2009.
[27] J. Melrose *et al.*, "Chondroitin sulphate and heparan sulphate sulphation motifs and their proteoglycans are involved in articular cartilage formation during human foetal knee joint development," *Histochem Cell Biol,* vol. 138, no. 3, pp. 461-75, Sep 2012.

[28] K. Choocheep *et al.*, "Versican facilitates chondrocyte differentiation and regulates joint morphogenesis," *J Biol Chem,* vol. 285, no. 27, pp. 21114-25, Jul 2 2010.
[29] H. E. Snow, L. M. Riccio, C. H. Mjaatvedt, S. Hoffman, and A. A. Capehart, "Versican expression during skeletal/joint morphogenesis and patterning of muscle and nerve in the embryonic mouse limb," *Anat Rec A Discov Mol Cell Evol Biol,* vol. 282, no. 2, pp. 95-105, Feb 2005.
[30] J. A. Klein, L. Meng, and J. Zaia, "Deep Sequencing of Complex Proteoglycans: A Novel Strategy for High Coverage and Site-specific Identification of Glycosaminoglycan-linked Peptides," *Mol Cell Proteomics,* vol. 17, no. 8, pp. 1578-1590, Aug 2018.
[31] C. Kiani, L. Chen, Y. J. Wu, A. J. Yee, and B. B. Yang, "Structure and function of aggrecan," *Cell Res,* vol. 12, no. 1, pp. 19-32, Mar 2002.
[32] J. H. Miner, J. Cunningham, and J. R. Sanes, "Roles for laminin in embryogenesis: exencephaly, syndactyly, and placentopathy in mice lacking the laminin alpha5 chain," *J Cell Biol,* vol. 143, no. 6, pp. 1713-23, Dec 14 1998.
[33] R. Dreier, A. Opolka, J. Grifka, P. Bruckner, and S. Grassel, "Collagen IX-deficiency seriously compromises growth cartilage development in mice," *Matrix Biol,* vol. 27, no. 4, pp. 319-29, May 2008.
[34] A. Opolka *et al.*, "Collagen IX is indispensable for timely maturation of cartilage during fracture repair in mice," *Matrix Biol,* vol. 26, no. 2, pp. 85-95, Mar 2007.
[35] J. F. Krey *et al.*, "Accurate label-free protein quantitation with high- and low-resolution mass spectrometers," *J Proteome Res,* vol. 13, no. 2, pp. 1034-1044, Feb 7 2014.
[36] J. F. Krey *et al.*, "Mass spectrometry quantitation of proteins from small pools of developing auditory and vestibular cells," *Sci Data,* vol. 5, p. 180128, Jul 17 2018.
[37] J. B. Shin *et al.*, "Molecular architecture of the chick vestibular hair bundle," *Nat Neurosci,* vol. 16, no. 3, pp. 365-74, Mar 2013.
[38] N. Rucci *et al.*, "The glycosaminoglycan-binding domain of PRELP acts as a cell type-specific NF-kappaB inhibitor that impairs osteoclastogenesis," *J Cell Biol,* vol. 187, no. 5, pp. 669-83, Nov 30 2009.
[39] R. B. White, A. S. Bierinx, V. F. Gnocchi, and P. S. Zammit, "Dynamics of muscle fibre growth during postnatal mouse development," *BMC Dev Biol,* vol. 10, p. 21, Feb 22 2010.
[40] J. A. Babraj *et al.*, "Collagen synthesis in human musculoskeletal tissues and skin," *Am J Physiol Endocrinol Metab,* vol. 289, no. 5, pp. E864-9, Nov 2005.
[41] T. Nishimura, K. Ojima, A. Hattori, and K. Takahashi, "Developmental expression of extracellular matrix components in intramuscular connective tissue of bovine semitendinosus muscle," *Histochem Cell Biol,* vol. 107, no. 3, pp. 215-21, Mar 1997.
[42] A. Listrat, B. Picard, and Y. Geay, "Age-related changes and location of type I, III, IV, V and VI collagens during development of four foetal skeletal muscles of double-muscled and normal bovine animals," *Tissue Cell,* vol. 31, no. 1, pp. 17-27, Feb 1999.
[43] S. S. Chaudhry *et al.*, "Mutation of the gene encoding fibrillin-2 results in syndactyly in mice," *Hum Mol Genet,* vol. 10, no. 8, pp. 835-43, Apr 1 2001.
[44] J. Sottile and D. C. Hocking, "Fibronectin polymerization regulates the composition and stability of extracellular matrix fibrils and cell-matrix adhesions," *Mol Biol Cell,* vol. 13, no. 10, pp. 3546-59, Oct 2002.
[45] A. Schiavinato, D. R. Keene, T. Imhof, R. Doliana, T. Sasaki, and G. Sengle, "Fibulin-4 deposition requires EMILIN-1 in the extracellular matrix of osteoblasts," *Sci Rep,* vol. 7, no. 1, p. 5526, Jul 17 2017.

[46] B. C. Bernardo, D. Belluoccio, L. Rowley, C. B. Little, U. Hansen, and J. F. Bateman, "Cartilage intermediate layer protein 2 (CILP-2) is expressed in articular and meniscal cartilage and down-regulated in experimental osteoarthritis," *J Biol Chem,* vol. 286, no. 43, pp. 37758-67, Oct 28 2011.

4. KINETICS OF AZIDOHOMOALANINE BIODISTRIBUTION AND ITS METABOLIC IMPLICATIONS *IN VIVO*

The content of this Chapter will be submitted for publication. Tyler G. VanDyk has contributed to the development of the mathematical model presented in this Chapter.

4.1 Introduction

The use of non-canonical amino acid (ncAA) labeling for selective identification of newly synthesized proteins (NSPs) in mammalian cells was first introduced by Dieterich et al. in 2006 [1] and has been employed since to study several biological systems (reviewed in Saleh et al. 2019 [2]). In this technique, an amino acid analog, typically methionine (Met), is introduced to the biological system of interest to be incorporated into the nascent polypeptide chains using the endogenous translational machinery of the cell. ncAAs carry reactive chemical groups such as azides and alkynes, which can be covalently modified via the azide-alkyne cycloaddition reaction (a click chemistry reaction) [2]. As such, ncAA-labeled NSPs can be selectively ligated to "clickable" affinity or fluorescent tags for identification or visualization, respectively [1, 3]. ncAA labeling has been successfully employed to probe protein dynamics in a variety of bacterial [4-6] and mammalian cells [7-9] as well as several model organisms such as zebrafish [10] and *Xenopus* [11]. More recently, ncAA labeling has been shown to be effective in probing NSPs in rodents [12-14]. Notably, extending the studies to more complex organisms such as rodents will pave the road to answer several biological questions wherein understanding the temporal dynamics of protein synthesis and turnover is critical.

To label the rodent proteome, the ncAA can be administered to animals by adding it to the diet [12, 15]. In this method, a Met-free diet is usually used to enhance the ncAA incorporation. While this approach could result in high labeling efficiency, it may alter normal physiology

especially when longer labeling periods are required [16, 17]. In addition, the presence of Met in animal diet is essential for normal embryonic development [18-20], which renders the method limited to adult tissues.

To address this gap, our group previously demonstrated that labeling the adult and embryonic murine proteome can be achieved via systemic injection of ncAAs without the need for Met depletion [13, 21]. Compared to feeding with ncAA-enriched diet, the injection method achieves global proteome labeling in a shorter period of time, which enables the detection of proteins synthesized shortly after injection and proteins with high turnover rates [21]. In addition, the injection method allows accurate dosing calculations, which eliminates the inherent variability of the feeding method due to fluctuations in feeding patterns and intestinal absorption.

Despite the application of ncAA labeling in a number of studies to decipher complex cellular processes in animal models [12, 14, 15, 22], understanding of the kinetics of ncAA distribution in tissues, especially as it pertains to rates of protein incorporation and loss by degradation, is lacking. Determination of the timescale of ncAA uptake by tissues following administration and the lag time before maximum protein labeling are critical information for the design of robust temporal experiments to study the nascent proteome. Such understanding will also enable optimizing the dosing regimen to attain the ideal ncAA concentration that achieves the required degree of protein labeling over the course of the study.

In addition to the limited knowledge of the distribution kinetics of ncAAs *in vivo*, evaluation of the physiological impact of ncAA administration to animals has been limited to examining changes in gross behavior, physical appearance and body weight [12-14]. To address these

limitations, systemic analysis of the effect of ncAA incorporation on cellular functions is required to demonstrate the suitability of the method for *in vivo* studies.

In the work presented in this chapter, we aimed to characterize the distribution kinetics of azidohomoalanine (Aha), the most widely used Met analog, in mice following subcutaneous injection, and to investigate the impact on normal physiology. To study the biodistribution of Aha, we measured the concentration in the plasma, liver, kidney, brain and skeletal muscles using liquid chromatography-tandem mass spectrometry (LC-MS/MS) over a period of 24 h. In addition, we used fluorescent western blotting to measure protein labeling in these tissues during the same period of time. We then utilized these data to develop a mathematical model to describe the biodistribution kinetics of Aha, and to calculate the relative rates of protein synthesis and turnover.

To investigate whether Aha incorporation into proteins perturbs normal physiological functions, we compared the plasma metabolome 24 h after Aha injection to that of non-injected mice to identify the metabolic pathways that could potentially be dysregulated due to protein labeling with Aha. This analysis showed that few metabolites were dysregulated in the injected mice, indicating that Aha administration does not have a significant impact on the normal physiology. Taken together, the results presented in this chapter provide fundamental understanding of the ncAA injection method with regards to the interrelation between the distribution kinetics of the ncAA into murine tissues and the associated degree of protein labeling, as well as the impact of the injection of the ncAA on physiological functions.

4.2 Methods

4.2.1 Animal Model

Animals used in these studies were derived from female age-matched wild-type C57BL/6 mice (*Mus musculus*) purchased from The Jackson Laboratory. All experimental protocols were performed in compliance with established guidelines and all methods were approved by Purdue Animal Care and Use Committee (PACUC, protocols# 1209000723 and 1801001682). PACUC requires that all animal programs, procedures, and facilities at Purdue University to abide by the policies, recommendations, guidelines, and regulations of the United States Department of Agriculture (USDA) and the United States Public Health Service (USPHS) in accordance with the Animal Welfare Act and Purdue's Animal Welfare Assurance.

4.2.2 Aha Injection, and Plasma and Tissue Collection

The methionine (Met) analog L-azidohomoalanine (Aha; Click Chemistry Tools) was resuspended in 1 × phosphate buffered saline (PBS) to a 10 mg/ml concentration, pH adjusted to 7.4 with NaOH, sterile filtered and stored at -20°C. All injections were administered to mice subcutaneously at 0.1 mg Aha /g mouse. Mice ($n = 3$ biological replicates) were euthanized 0.5, 1, 2, 4, 6, 12 and 24 h post injection (hpi). Blood was collected by cardiac puncture into EDTA-treated tubes and centrifuged at 1,500 × g for 10 min at 4°C. The supernatant (plasma) was transferred into a new tube using a Pasteur pipette, snap frozen in liquid nitrogen and stored at -80°C. Liver, brain, kidney and hindlimb skeletal muscle tissues were dissected at each time point, snap frozen in liquid nitrogen and stored at -80°C. Plasma and tissues were collected as described above from non-injected mice ($n = 3$ biological replicates) to be used as a control.

4.2.3 Sample Preparation for Aha Analysis

For plasma sample preparation, 50 µl of plasma were mixed with 10 µl of 1 × PBS, pH 7.4 and 5 µl of 100 ng/µl L-α-aminobutyric acid (α-ABA; Sigma Aldarich) that was used as an internal standard. 12.5 µl of trichloroacetic acid (TCA; Sigma Aldarich) were added to the mixture to precipitate proteins. The mixture was incubated for 10 min at 4°C and centrifuged at 16,000 × *g* for 10 min at RT. The supernatant was then mixed with 100% acetonitrile (ACN; Fisher Scientific) at a 1:1 ratio (v/v). The mixture was transferred to an HPLC autosampler vial for LC-MS/MS analysis. For calibration curve generation, Aha standards were prepared by mixing 50 µl of non-injected plasma with 10 µl of a known concentration of Aha and 5 µl of α-ABA. Proteins were then precipitated with TCA and prepared for LC-MS/MS analysis as described above.

For tissue sample preparation, tissues were rinsed with ice-cold 1 × PBS, pH 7.4 to remove residual blood and homogenized in ice-cold 1 × PBS, pH 7.4 using a TissueRuptor (Qiagen). The final homogenate weight was measured and converted to volume by using a homogenate density of 1 g/ml. Samples were then prepared for LC-MS/MS analysis as described for plasma by using 50 µl of the tissue homogenate. The remaining plasma samples and tissue homogenates were snap frozen and stored at -80°C until use for western blot and untargeted metabolomic analyses as described below.

4.2.4 LC-MS/MS Targeted Analysis of Aha

An Agilent 1260 Rapid Resolution liquid chromatography (LC) system coupled to an Agilent 6470 series QQQ mass spectrometer was used for Aha analysis (Agilent Technologies). An Intrada Amino Acid 2.0 mm x 150 mm, 3.0 µm column (Imtakt Corporatio) was used for LC separation. The buffers were (A) ACN, 0.3 % formic acid (FA; Sigma Aldarich) and (B) ACN/100

mM ammonium formate (20/80 v/v). The linear LC gradient was as follows: time 0 min, 20 % B; time 5 min, 20 % B; time 11 min, 35 % B; time 20 min, 100 % B; time 22 min, 100 % B; time 22.5 min, 20 % B; time 30 min, 20% B. The flow rate was 0.3 ml/min. Multiple reaction monitoring (MRM) was used for MS analysis. Data were acquired in a positive electrospray ionization (ESI) mode according to Table 4.1. The jet stream ESI interface had a gas temperature of 325°C, gas flow rate of 9 L/minute, nebulizer pressure of 35 psi, sheath gas temperature of 250°C, sheath gas flow rate of 7 L/minute, capillary voltage of 3500 V in a positive mode, and nozzle voltage of 1000 V. The delta electron multiplier voltage was 300 V. Agilent MassHunter Quantitative Analysis software was used for data analysis (version 8.0).

Table 4.1 Multiple reaction monitoring (MRM) table for amino acid LC-MS/MS data acquisition

Compound name	Precursor ion (*m/z*)	Product ion (*m/z*)	Collision energy (eV)
Aha	145.1	101.3	5
Aha	145.1	71.3	10
Aha	145.1	58.3	40
Ala	90	44	15
Arg	175	116	18
Asn	133	87	12
Asp	134	88	14
Cys	122	76	15
Cys-Cys	241.1	152	15
Gln	147	84	22
Glu	148	130	12
Gly	76	30	15
His	156	110	19
Ile	132	86	15
Leu	132	86	15
Lys	147	84	20
Met	150	104	15
Phe	166	120	15
Pro	116	70	15

4.2.5 Western Blot Analysis of Aha-Labeled Tissues

Tissue homogenates were thawed and protein concentration was measured using the Pierce 660 nm Protein Assay (ThermoFisher Scientific). 200 µg of tissue homogenate were alkylated with 40 mM iodoacetamide for 30 min at RT in the dark with end-over-end rotation. Samples were then reacted for 2 h at RT with the following click reagents: 50 µM biotin-alkyne (ThermoFisher Scientific), 5 mM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA; Click Chemistry Tools), 2 mM copper sulfate, 20 mM aminoguanidine and 10 mM sodium ascorbate. Following click reaction, proteins were precipitated by adding ice-cold 100% acetone to the samples at a 4:1 ratio (v/v). Samples were incubated overnight at -20ºC, centrifuged at 21,100 × *g* for 20 min at 4ºC, supernatants were discarded and protein pellets were vacuum-dried for 15 min at RT using CentriVap (Labconco). Dried pellets were resuspended in (8 M urea in 1× PBS) and centrifuged at 16,000 × *g* for 15 min at RT to remove insoluble particles. The supernatants were transferred into new tubes and protein concentration was measured using the Pierce 660 nm Protein Assay (ThermoFisher Scientific). Proteins were resolved on 4 – 20% SDS-PAGE gels (BioRad), transferred to a PVDF membrane (ThermoFisher Scientific) using the Trans-Blot Turbo Transfer System (BioRad) and probed overnight at 4°C with IRDye 680 Streptavidin (LICOR) (1:3000 dilution). Membranes were imaged using an Azure Biosystems c600. Western blot images were analyzed using ImageJ (National Institutes of Health) to calculate the mean fluorescence intensities of each time point. The intensity of the control sample was used to normalize the intensity of each time point ($n = 3$ biological replicates per blot).

4.2.6 Kinetic Analysis of Aha Distribution

MATLAB r2018b was used for the analysis of Aha concertation profiles in the plasma and tissues and for the calculations of relative protein synthesis and turnover rates. A system of

ordinary differential equations (ODEs, see Section 4.3 for a list of the equations) was defined and solved using MATLAB predefined solver ode15s, and was fit using least square estimation (LSE) with the MATLAB package fmincon.

4.2.7 Plasma Sample Preparation for Untargeted Metabolomic Analysis

The plasma metabolome of non-injected control samples ($n = 3$ biological replicates) and samples collected 24 h post Aha injection ($n = 3$ biological replicates) was extracted by adding methanol: chloroform: water (1:1:1 v/v) to 80 µl of each plasma sample. Samples were vortexed briefly and centrifuged at 8,000 × *g* for 5 min at RT. The upper layer was transferred into a new tube and vacuum-dried overnight at RT. The dried fraction was reconstituted in 75 ul of a diluent composed of 5% ACN and 0.1% FA. Reconstituted samples were sonicated for 5 minutes, centrifuged at 16,000 × *g* for 8 min at RT, and the supernatants were transferred to HPLC autosampler vials.

4.2.8 LC-MS Untargeted Metabolomic Analysis

Separations were performed on an Agilent 1290 UPLC system (Agilent Technologies). The metabolites were analyzed using a Waters Acquity HSS T3 column (1.8 µm, 2.1 × 100 mm), with a mobile phase flow rate of 0.45 ml/min, where the mobile phase A and B were 0.1% FA in double distilled water (ddH_2O) and ACN, respectively. Initial conditions were 100:0 A:B, held for 1 minute, followed by a linear gradient to 20:80 at 16 min, then 5:95 at 22.5 min. Column re-equilibration was performed by returning to 100:0 A:B at 23.5 minutes and holding until 28.5 minutes.

The mass analysis was obtained using an Agilent 6545 Quadrupole Time of Flight (Q-TOF) MS with ESI capillary voltage +3.2 kV, nitrogen gas temperature 325 °C, drying gas flow rate 8.0 L/min, nebulizer gas pressure 30 psig, fragmentor voltage 130 V, skimmer 45 V, and OCT RF 750 V. MS data scans (*m/z* 70-1000) were collected using Agilent MassHunter Acquisition software (v.B.06). Mass accuracy was improved by infusing Agilent Reference Mass Correction Solution (G1969-85001). MS/MS was performed in a data-dependent acquisition mode on composite samples.

4.2.9 Metabolomics Data Analysis

Peak deconvolution and integration was performed using Agilent ProFinder (v.10.0). Bioinformatic analyses were performed using Agilent Mass Profiler Professional (v.13.1). Chromatographic peaks were aligned across all samples. Peak areas were normalized by $\log_2$-transformation and applying a 75% percentile shift. Metabolites were filtered out if present in only one sample. Only metabolites present in all 3 replicates of at least one group were included in the analysis. Statistical analysis was performed using unpaired student's *t*-test. Metabolites with $P < 0.05$ and fold change ≥ 2 were considered significant. Peak annotations were performed using the METLIN metabolite database [23], with a mass error of less than 5 ppm. Identifications were aided by MS/MS spectra comparisons. Principal component analysis (PCA), hierarchal clustering analysis (HCA) and metabolic pathway analysis were performed using MetaboAnalyst v.4.0 [24].

4.3 Results and Discussion

4.3.1 The Biodistribution of Free Aha Differs Between Different Murine Tissues

The concentration profile of free Aha (Aha not incorporated into proteins) in the plasma and tissues was determined by injecting Aha subcutaneously into mice at 0.1 mg/g body weight

followed by sacrificing mice 0.5, 1, 2, 4, 6, 12 and 24 h post injection (hpi). LC-MS/MS multiple reaction monitoring (MRM) analysis was used for accurate identification of free Aha by monitoring the product ion spectra at *m/z* 101.3, 71.3 and 58.3 generated from the precursor ion at *m/z* 145.1 eluting at ~4 min (Figure 4.1).

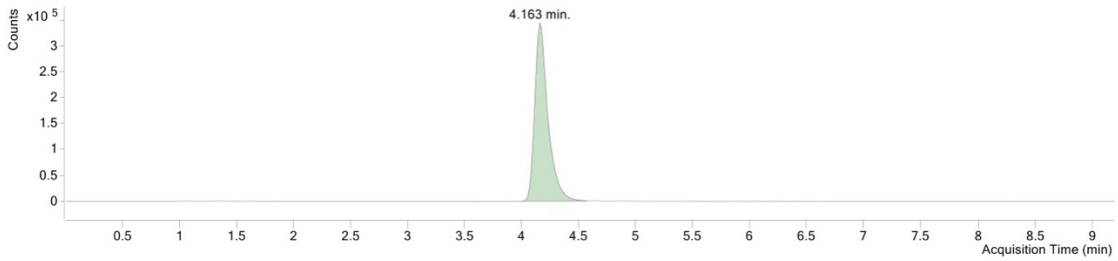


Figure 4.1 Representative multiple reaction monitoring (MRM) chromatogram of free Aha in plasma showing an elution time of 4.163 min. The concentration of free Aha was determined by calculating the area under the peak in the sample and using a calibration curve of known Aha concentrations.

The concentration profiles in the plasma and tissues show that Aha peaks between 0.5 and 1 hpi and that Aha is nearly cleared from the system by 12 hpi with the liver having the earliest peak compared to the other tissues (Figure 4.2). The early peak observed in the liver can be attributed to its high blood perfusion, which likely results in faster distribution equilibrium of Aha into the liver compared to other tissues.

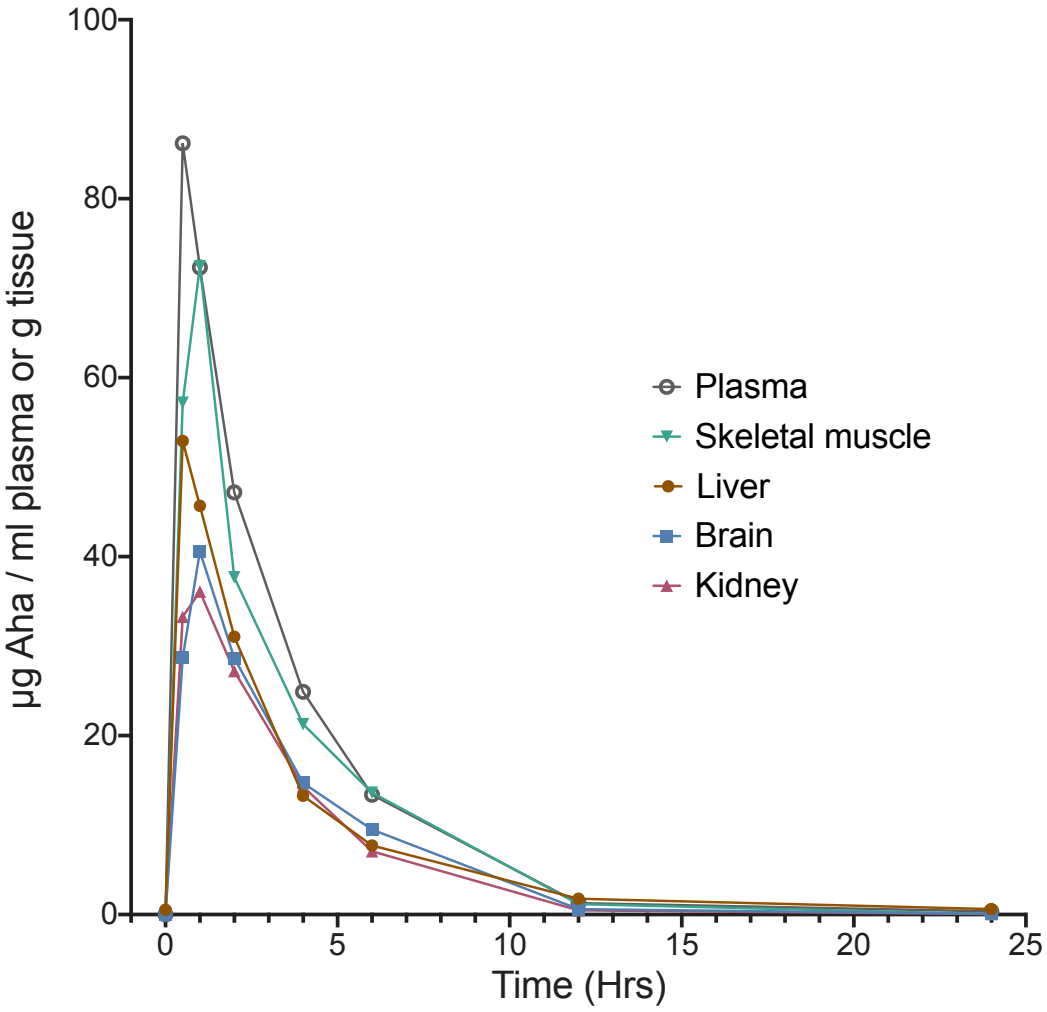


Figure 4.2 The concentration profile of free Aha in the plasma and different tissues. The amount of Aha (µg) measured by LC-MS/MS analysis was normalized by the total plasma volume or tissue mass and plotted over time.

4.3.2 Aha Labeling Reveals Different Relative Protein Synthesis and Turnover Dynamics in Murine Tissues

To investigate whether introducing Aha into mice via injection without Met depletion is suitable for probing the dynamics of protein synthesis and turnover in different tissues, the pattern of protein labeling within each tissue was monitored over the duration of the study. Tissue homogenates were reacted with biotin-alkyne via copper-catalyzed click reaction, analyzed by western blotting using streptavidin fluorophore and the change in fluorescence intensity relative

to the non-injected controls was measured. In all tissues, maximum protein labeling was observed around 6 hpi (Figure 4.3A). However, the degree of labeling, represented by the maximum fold increase in fluorescence intensity compared to control, and the kinetics of protein synthesis and turnover varied considerably between tissues. In particular, the liver showed the highest degree of labeling as well as the highest relative rates of protein synthesis and turnover, whereas skeletal muscle had the lowest degree of labeling and the slowest relative rates of protein synthesis and turnover (Figure 4.3B). These results are in agreement with previous isotope labeling studies that showed faster protein turnover rates in liver and kidney compared to brain and skeletal muscle [25-27]. It is important to note here that measurements based on western blotting are typically semi-quantitative and relatively variable. As such, accurate quantitation of protein turnover rates using MS is required for the precise estimation of protein half-lives in different murine tissues.

Interestingly, while skeletal muscle showed the lowest degree of labeling, the amount of free Aha (μg) per unit mass tissue (g) was higher in skeletal muscle than in liver (Figure 4.1A), indicating that the low degree of labeling observed in skeletal muscle is predominantly due to the slow rate of muscle protein synthesis rather than the lack of availability of free Aha in muscle tissues. This result also underlines the importance of tailoring the dosing regimen of Aha (*i.e.* amount per dose and dosing frequency) to the tissue of interest and the biological question under investigation.

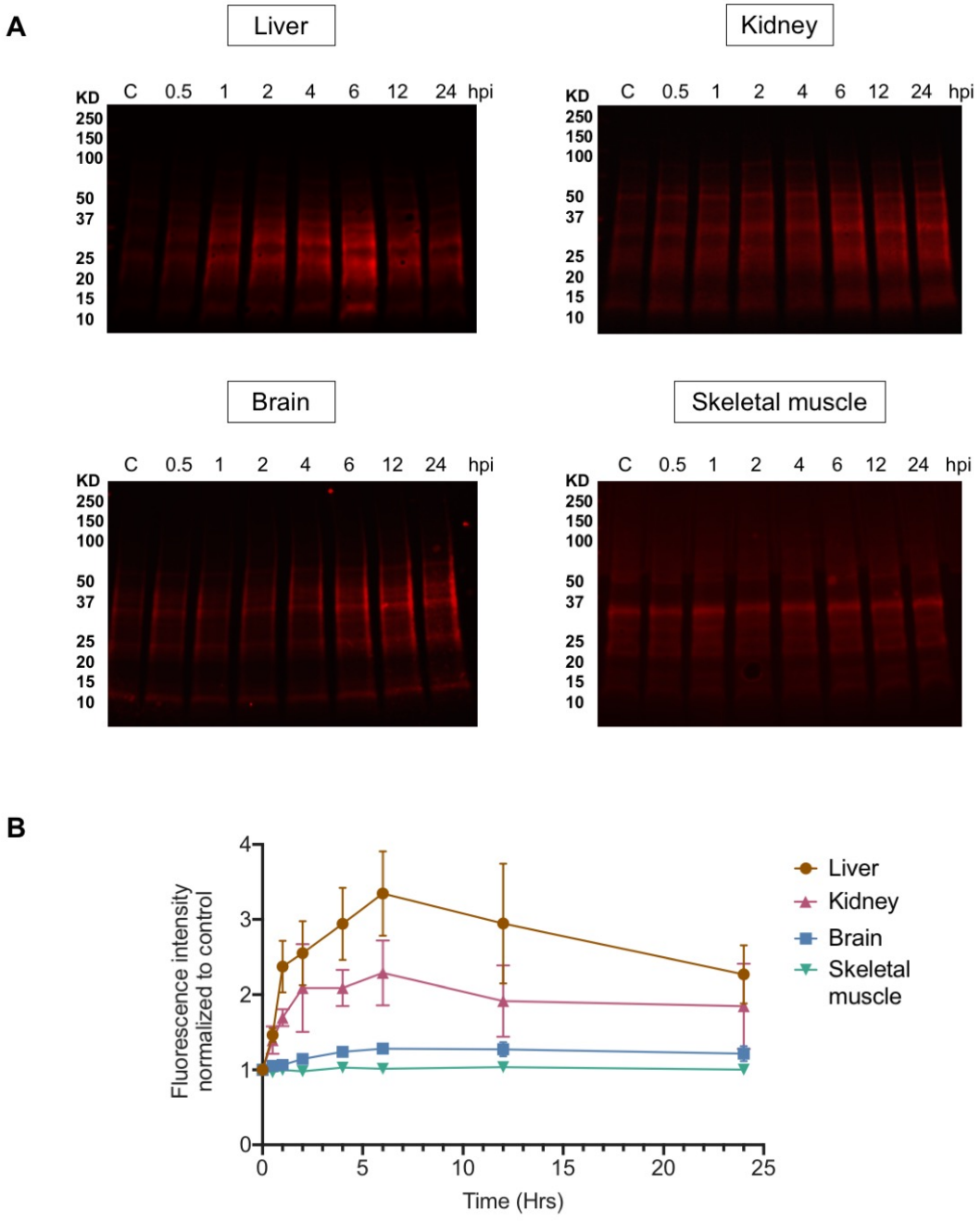


Figure 4.3 Protein synthesis and turnover kinetics in murine tissues. **(A)** Fluorescent western blots of the tissue homogenates of control non-injected samples (C) and samples collected 0.5 – 24 h post Aha injection (hpi). The samples were reacted with biotin-alkyne using click chemistry and analyzed via western blotting using streptavidin-680 fluorophore. **(B)** Fluorescence intensity of western blot lanes were normalized to that of the respective control samples and plotted as function of time. Error bars represent standard error of the mean ($n = 3$ biological replicates).

4.3.3 Mathematical Modeling of Aha Biodistribution, and Relative Protein Synthesis and Turnover Rates

The results in Section 4.3.2 show that Aha is successfully incorporated into the NSPs in different tissues shortly after a single injection. These results demonstrate the potential of the method to enable the identification of NSPs and quantitation of the turnover rates of individual proteins *in vivo* using MS. Depending on the tissue type and biological processes to be studied, multiple injections of Aha may be required to attain a high enough degree of labeling that results in a suitable MS signal. Therefore, optimizing the dose and frequency of Aha injections is critical for the appropriate design of labeling studies. To this end, we sought to develop a quantitative framework to calculate the relative rates of protein synthesis and turnover, and to quantify the relation between the biodistribution profile of Aha and the degree of protein labeling. As such, this framework can be divided into two underlying modules: (1) Aha biodistribution and (2) relative protein synthesis and degradation. Both modules can then be utilized to define a general model that relates the Aha biodistribution profile to the degree of protein labeling. In this section, the development of this model is described.

Kinetic modeling of Aha biodistribution in murine tissues

Modeling of the biodistribution kinetics of Aha in murine tissues was performed using a system of ordinary differential equations (ODEs) describing small molecule transport in mice. After subcutaneous injection, free, protein-unbound Aha (fAha) enters and circulates the blood stream, wherein it transports to and exchanges with different tissue (compartments). Therefore, it is reasonable to break down the distribution model of fAha into two stages: transport and exchange.

Within each tissue, the time rate of change of the fAha plasma concentration ([$fAha_p$]) available for exchange with the tissue can be described as a mass balance. The simple transport case in the absence of tissue exchange can modeled for each tissue (x) as follows:

$$\left(\frac{d[fAha_p]_x}{dt}\right)_{transport} = \frac{q_x}{V_x} \cdot \left(\left[fAha_p\right]_{blood\ reservoir} - \left[fAha_p\right]_x\right) \quad \text{Eq. 4.1}$$

where q_x is the blood flow rate ($\mu l \cdot min^{-1} \cdot mg^{-1}$) perfusing the region, and V_x is the corresponding volume of plasma ($\mu l \cdot mg^{-1}$). The liver, gut and kidney plasma compartments all have additional elimination terms ($min^{-1} \cdot mg^{-1}$) accounting for excretion and metabolism of fAha. All kinetic parameters for circulatory transport were normalized by tissue mass to compare relative perfusion rates between tissue compartments of differing size. Once localized to a tissue, fAha in the plasma may then be exchanged across the cell membrane with the intracellular Aha ($fAha_t$) and can be described as follows:

$$\left(\frac{d[fAha_p]_x}{dt}\right)_{exchange} = k_{out} \cdot [fAha_t]_x - k_{in} \cdot \left[fAha_p\right]_x \quad \text{Eq. 4.2}$$

$$\left(\frac{d[fAha_t]_x}{dt}\right)_{exchange} = k_{in} \cdot \left[fAha_p\right]_x - k_{out} \cdot [fAha_t]_x \quad \text{Eq. 4.3}$$

where k_{in} and k_{out} are the rates of Aha transport ($min^{-1} \cdot mg^{-1}$) into and outside of the tissue, respectively. The two stages of fAha distribution were then combined into a single system of ODEs, which was parameterized and bound within reasonable ranges for a model of small molecule pharmacokinetics [28-31]. Parameters were then fit to best match the fAha concentration profile measured using LC-MS/MS.

Kinetic modeling of protein labeling

Following its distribution to tissues, fAha is incorporated into proteins via protein synthesis, resulting in a protein-bound form (pAha). As a Met analog, Aha is able to bind to methionyl tRNA synthase, albeit at a much slower rate than Met (k_{cat}/K_m is 5.47×10^{-1} and 1.42×10^{-3} $s^{-1} \cdot \mu M^{-1}$ for Met and Aha, respectively) [32]. The concentration of free Met (fMet) in mice tissues reported in literature (~ 8 and 12 $\mu g \cdot g^{-1}$ in liver and kidney, respectively) [33] is lower than but within the same order of magnitude of the peak concentrations of Aha that were measured in these tissues (Figure 4.2). To our knowledge, however, there is no reliable experimental measurements reported in literature of the amount of Met incorporated into proteins (pMet) in murine tissues. Met is a relatively rare amino acid [34], with a tissue concentration comparable to that of the peak Aha in our study but binds to methionyl tRNA synthase approximately 400 times faster than Aha [32]. Therefore, it is reasonable to assume that the concentration of pAha is small compared to that of fAha in tissues. It follows then that the concentration of pAha is dependent upon the concentration of fAha available for protein synthesis, but the depletion of fAha due to its incorporation into proteins is likely negligible compared to the systemic transport and elimination of fAha. In addition, despite vast differences in the degree of protein labeling between tissues (for instance liver and skeletal muscle) (Figure 4.3), there is no discernable difference in their fAha concentration profiles (Figure 4.2), supporting the assumption of negligible depletion of fAha by protein synthesis. As such, using these assumptions, the rates of change of the concentrations of fAha and pAha in the tissues can be described as follows:

$$[fAha_t]_x \gg [pAha_t]_x \qquad \text{Eq 4.4}$$

$$\left(\frac{d[fAha_t]_x}{dt}\right)_{synthesis} = -k_{syn} \cdot [fAha_t]_x \approx 0 \qquad \text{Eq 4.5}$$

$$\left(\frac{d[pAha_t]_x}{dt}\right)_{synthesis} = k_{syn} \cdot [fAha_t]_x - k_{deg} \cdot [pAha_t]_x \quad \text{Eq 4.6}$$

where k_{syn} and k_{deg} are the tissue-specific relative rate constants of protein synthesis and degradation, respectively. Furthermore, if the degree of fluorescent signal relative to the background (rF) from pAha labeled proteins is linearly proportional to the concentration of pAha by a factor, k_f, then the generation of rF signal can also be described as follows:

$$rF_x = \frac{flourescent\ signal}{flourescent\ background} - 1 = k_f[pAha_t]_x \quad \text{Eq 4.7}$$

$$\frac{d(rF_x)}{dt} = k_f\left(\frac{d[pAha_t]_x}{dt}\right) = k_f \cdot k_{syn} \cdot [fAha_t]_x - k_{deg} \cdot rF_x \quad \text{Eq 4.8}$$

This model was fit to fluorescence data acquired for each tissue via western blotting to (1) estimate relative protein synthesis and degradation rates in each tissue and (2) establish a time-resolved predictive model of protein labeling given a variety of input dosing paradigms.

Estimation of relative protein synthesis and turnover rates

Using the model described above, relative protein synthesis and turnover rates were calculated for the different tissues (Table 4.2). Relative protein synthesis and turnover rates differed among tissues. Notably, the estimated protein half-lives are consistent with previous studies that utilized isotope labeling to measure tissue protein half-lives using MS [25-27]. In these studies, measurements based on absolute quantitation using isotope labeling revealed that liver and kidney have higher turnover rates compared to brain and skeletal muscle. For example, Price et al. showed that the average half-life of brain proteins is 3 times higher than liver (9 and 3 days for

brain and liver, respectively) [26]. Markedly, our model estimated an average brain protein half-life that is 2.7 higher than the liver (38.3 and 14.1 hours for brain and liver, respectively) (Table 4.2). The discrepancies between previously reported values and the half-lives values estimated here can be attributed to the shorter timescale of our experimental setup and the lower accuracy of western blotting measurements compared to MS.

Table 4.2 Tissue relative protein synthesis rates and protein half-lives

Parameter	Liver	Kidney	Brain	Skeletal muscle
Protein turnover rate (h^{-1})	0.0491	0.0488	0.0181	0.0229
Protein half-life (h)	14.1	14.2	38.3	30.2
Relative protein synthesis rate* $\left(\frac{\mathrm{RU}}{\mu g_{\mathrm{pAha}}} \cdot \frac{1}{\mu g_{\mathrm{fAha}} \cdot \mathrm{h}}\right)$	4.19×10^{-3}	5.41×10^{-3}	6.62×10^{-4}	4.53×10^{-5}

*RU: relative fluorescence unit.

Predictive simulations of dosing paradigms

Attaining good protein labeling is critical to adequately enrich Aha-labeled proteins with high signal-to-noise ratio for accurate quantitative MS measurements. Therefore, it is imperative to optimize the labeling strategy for the tissue of interest. Using the developed model, fAha biodistribution and tissue protein labeling can be predicted following different dosing regimens. Figure 4.4 shows predictions of protein labeling for three examples of multiple dosing regimens. As expected, increasing the dosing frequency increases the degree of protein labeling, which is critical for tissues with lower protein synthesis rates such as brain and skeletal muscle. For the different tissues and biological processes to be studied, researchers should be able to use this model to guide the optimization of tissue labeling with only minimal preliminary experiments.

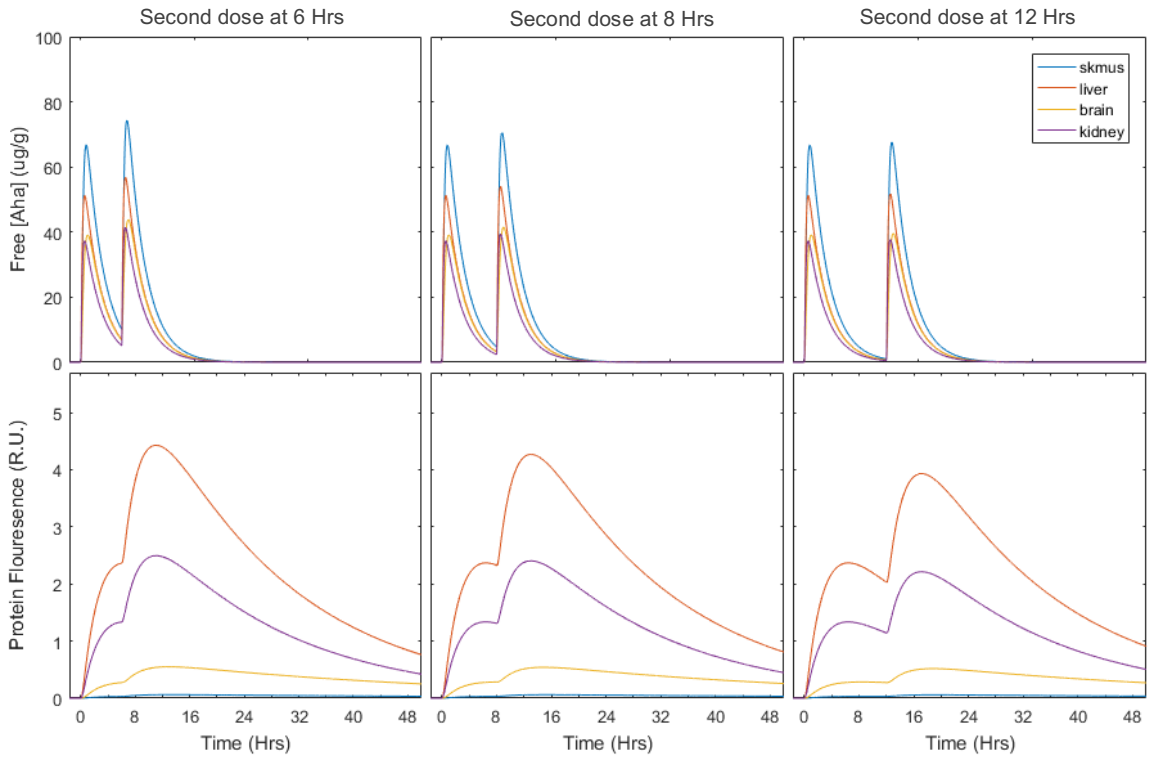


Figure 4.4 The effect of different dosing regimens on the biodistribution of fAha and protein labeling. Top panels display the best fit model of [fAha] in each tissue over time and bottom panels show the respective predicted rF signal in each tissue. Each column displays the results of a separate dosing paradigm. Left to right, an initial dose of 0.1 mg/g Aha injected at 0, and then a second dose at 6, 8 and 12 hours.

4.3.4 Aha Administration Does Not Perturb Normal Physiology in Mice

Based on the timescale of Aha biodistribution (Figure 4.2) and incorporation into proteins (Figure 4.3), the physiological impact of Aha incorporation into newly synthesized proteins was evaluated using untargeted plasma metabolomic analysis. Since metabolites are the end products of the cellular biological processes, we reasoned that a lag time is expected between potential changes in protein functions due to the incorporation of Aha and any associated effects on metabolism. Therefore, given that maximum protein labeling occurs ~ 6 hpi (Figure 4.3), we analyzed the plasma metabolome 24 hpi to enable identifying changes in metabolic pathways in response to altered protein functions (if any).

LC-MS metabolomic analysis of the plasma identified a total of 1268 mass features (*i.e.* metabolites). The peak area of each mass feature is proportional to the amount of the corresponding ion in the sample and is used as a measurement for the relative abundance of each identified metabolite across samples. Principal component analysis (PCA) was performed to examine whether control and Aha-treated samples can be grouped according to variation in metabolite abundance. The analysis revealed no distinct segregation between the control and Aha mice, indicating that there are no global differences in the plasma metabolome between the two groups (Figure 4.5A).

It should be noted here that PCA cannot be performed in the presence of missing values. The occurrence of missing values is common in untargeted metabolomic data. It often results from the presence of metabolites with concentrations that are lower than the MS detection limit or due to technical reasons such as incomplete ionization or inaccurate peak detection [35]. In our dataset, a total of 194 (2.5%) missing values were detected across all samples. Since the percentage of missing values is low, it is deemed appropriate to assume that the potential impact of missing values is insignificant [36]. Therefore, these values were filtered out, resulting in a remaining 1112 mass features that were used for PCA. To confirm the validity of this approach, PCA was also conducted on the dataset after missing value imputation using the K-nearest neighbor (KNN) method and showed similar indistinct grouping of the injected and control mice (Figure B.1).

In addition to PCA, unsupervised hierarchical clustering analysis (HCA) was conducted and a heatmap was generated to examine variations in metabolic patterns between the Aha and control groups (Figure 4.5B). HCA and heatmap visualization showed no clustering between the biological replicates of each group and no distinct differential abundance patterns between the two groups. Accordingly, this result further establishes that there are no substantial metabolic

differences between control and injected mice. Similar to PCA, HCA performed using the dataset imputed via the KNN method resulted in indistinct clustering of the mice (Figure B.2).

Following global analysis using PCA and HCA, Student's *t*-test was employed to identify metabolites that may be dysregulated between the two groups. In this analysis, a total of 15 metabolites were differentially abundant using a *p*-value of 0.05 and a fold change of ≥ 2 as cut-offs (Figure 4.5C). Of the 15 metabolites, 3 were upregulated and 12 were downregulated in Aha-treated mice compared to the control. Searching the 15 metabolites in the METLIN metabolite database using a mass tolerance of 5 ppm did not result in any known metabolite. The presence of a large number of unknown mass features is an intrinsic characteristic of untargeted metabolomic studies due to the complexity of the mammalian metabolome and the lack of structure characterization of a large number of metabolites [37, 38]. Yet, the fact that only ~1.3% of metabolites were dysregulated and that all dysregulated metabolites did not belong to any of the known major metabolic pathways signify that minimal metabolic alterations occur due to Aha administration.

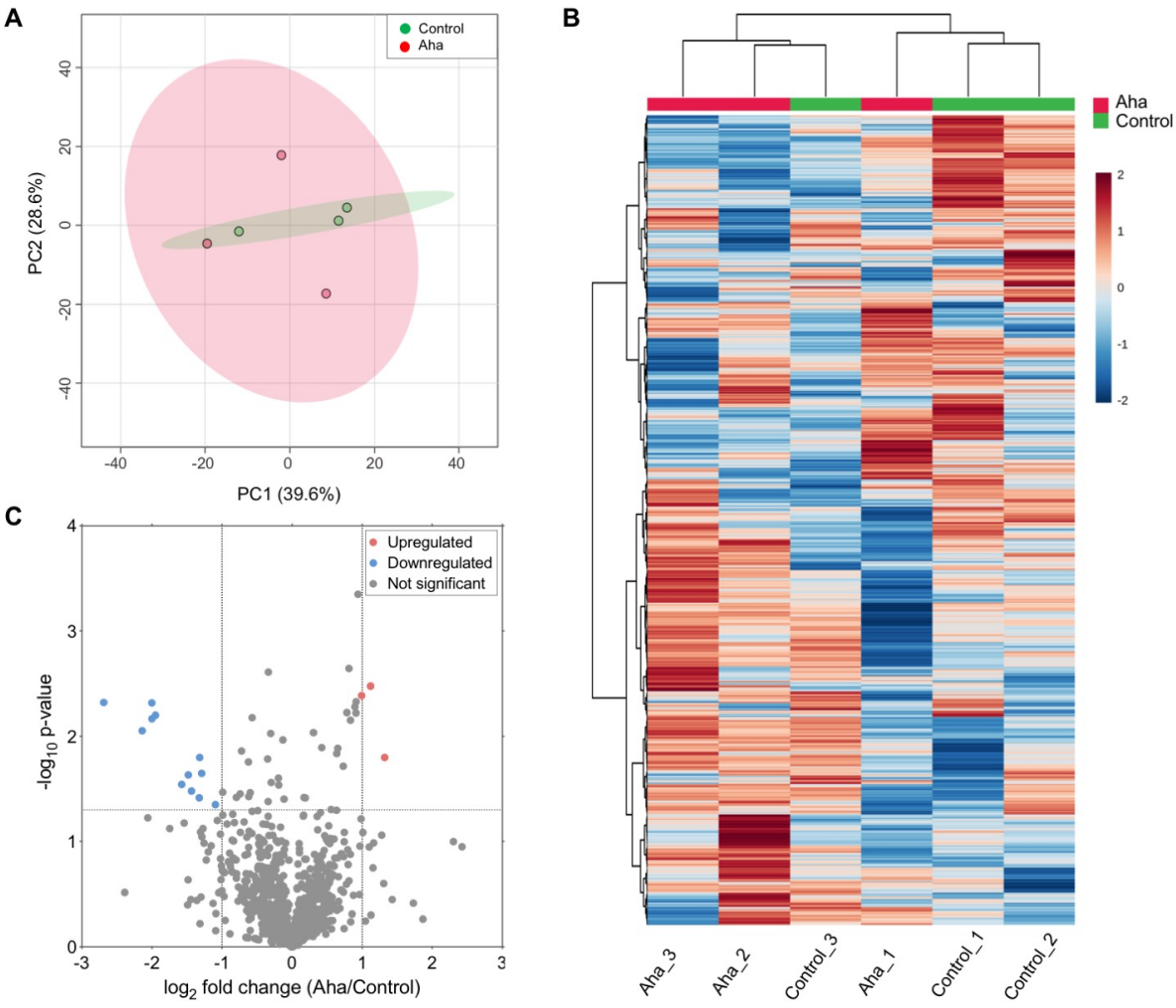


Figure 4.5 Aha administration does not significantly change mouse plasma metabolome. **(A)** Principle component analysis (PCA) shows no clear separation between control and Aha-treated groups in the first two components. Components 1 and 2 account for 39.6% and 28.6% of the total data variability, respectively. Green and red dots denote control and Aha samples, respectively. Green and red shaded areas represent the 95% confidence bands of the control and Aha samples, respectively. **(B)** Unsupervised hierarchal clustering and heatmap of the identified metabolites show lack of clustering between replicates of each group. Color scale indicates metabolite abundance; blue: lowest, red: highest. **(C)** Volcano plot comparing the relative abundance of the identified metabolites between control and Aha groups according to statistical significance and fold change. Horizontal line indicates *p*-value = 0.05 and vertical lines indicate $\pm$ 2-fold change. Grey, red and blue circles denote equally-abundant, upregulated and downregulated metabolites, respectively.

Finally, to identify the metabolic pathways covered by the LC-MS analysis, pathway analysis of the equally-abundant metabolites was conducted using the Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathway database [39]. Several metabolic pathways were detected in this analysis, including cysteine and Met metabolism, valine, leucine and isoleucine biosynthesis, aminoacyl-tRNA biosynthesis, vitamin B6 metabolism, and glycerophospholipid metabolism. This result indicates that the LC-MS analysis identified metabolites that belong to various metabolic pathways and that these pathways are not significantly changed in mice injected with Aha.

Collectively, the metabolomic analyses demonstrate that Aha administration into mice does not significantly alter the plasma metabolome. This is in agreement with a recent study that investigated the metabolic effect of growing *E.coli* in media supplemented with ncAAs [40]. A major advantage of our labeling technique is that it does not involve Met restriction or depletion as the case with other labeling strategies that feed animals on Aha-enriched Met-free diet. Met dietary restriction has been shown to alter the metabolism in mouse models and in humans [16, 17, 41]. Being a principle sulfur-containing amino acid, Met restriction specifically alters Met and sulfur metabolism [16, 17]. Notably, the results of the pathway analysis identified several unchanged metabolites such as 5'-methylthioadenosine, dehydroalanine and 2-oxobutanoate that belong to cysteine (another sulfur-containing amino acid) and Met metabolism, indicating the advantage of the injection method with regards to its potential impact on metabolic functions.

4.4 Conclusions

In this Chapter, we report for the first time the biodistribution kinetics of Aha, the most widely used Met analog, in murine tissues as well as the associated relative rates of protein

synthesis and turnover. These results show that liver and kidney have faster synthesis and turnover rates compared to brain and skeletal muscle, which is consistent with the results of previous studies that utilized isotope labeling for analyzing protein turnover rates. We also show that the injection technique allows observing maximum protein labeling in a relatively short time (~ 6 h), which enables studying proteins with shorter half-lives, in contrast to the traditional method of introducing the ncAA in diet or using isotope labeled amino acids. Additionally, we report the development of a mathematical framework that describes the distribution kinetics of Aha in murine tissues and its relation to the degree of protein labeling, and computes the relative rates of protein synthesis and turnover.

Finally, we investigated the impact of Aha administration on the plasma metabolome and show that Aha incorporation into cellular proteins does not have adverse effects on the normal physiology of mice. This observation further confirms previous results from our group that demonstrated that ncAAs do not affect the gross behavior nor the physical appearance of treated mice.

4.5 References

[1] D. C. Dieterich, A. J. Link, J. Graumann, D. A. Tirrell, and E. M. Schuman, "Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT)," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 103, pp. 9482-9487, 2006.

[2] A. M. Saleh, K. M. Wilding, S. Calve, B. C. Bundy, and T. L. Kinzer-Ursem, "Non-canonical amino acid labeling in proteomics and biotechnology," *Journal of Biological Engineering,* vol. 13, p. 43, 2019.

[3] D. C. Dieterich *et al.*, "In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons," *Nat Neurosci,* vol. 13, no. 7, pp. 897-905, Jul 2010.

[4] J. D. Bagert *et al.*, "Time-resolved proteomic analysis of quorum sensing in Vibrio harveyi," *Chemical Science,* vol. 7, pp. 1797-1806, 2016.

[5] D. M. Van Elsland, E. Bos, W. De Boer, H. S. Overkleeft, A. J. Koster, and S. I. Van Kasteren, "Detection of bioorthogonal groups by correlative light and electron microscopy allows imaging of degraded bacteria in phagocytes," *Chemical Science,* vol. 7, pp. 752-758, 2016.

[6] A. Mahdavi *et al.*, "Identification of secreted bacterial proteins by noncanonical amino acid tagging," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 111, pp. 433-438, 2014.

[7] J. D. Bagert *et al.*, "Quantitative, time-resolved proteomic analysis by combining bioorthogonal noncanonical amino acid tagging and pulsed stable isotope labeling by amino acids in cell culture," *Molecular and Cellular Proteomics,* vol. 13, pp. 1352-1358, 2014.

[8] L. D. Cohen *et al.*, "Metabolic Turnover of Synaptic Proteins: Kinetics, Interdependencies and Implications for Synaptic Maintenance," *PLoS ONE,* vol. 8, p. e63191, 2013.

[9] J. W. Kenney, M. Genheden, K. M. Moon, X. Wang, L. J. Foster, and C. G. Proud, "Eukaryotic elongation factor 2 kinase regulates the synthesis of microtubule-related proteins in neurons," *Journal of Neurochemistry,* vol. 136, pp. 276-284, 2016.

[10] F. I. Hinz, D. C. Dieterich, D. A. Tirrell, and E. M. Schuman, "Noncanonical amino acid labeling in vivo to visualize and affinity purify newly synthesized proteins in larval zebrafish," *ACS Chemical Neuroscience,* vol. 3, pp. 40-49, 2012.

[11] W. Shen *et al.*, "Acute Synthesis of CPEB Is Required for Plasticity of Visual Avoidance Behavior in Xenopus," *Cell Reports,* vol. 6, pp. 737-747, 2014.

[12] D. B. McClatchy *et al.*, "Pulsed Azidohomoalanine Labeling in Mammals (PALM) Detects Changes in Liver-Specific LKB1 Knockout Mice," *J Proteome Res,* vol. 14, no. 11, pp. 4815-22, Nov 6 2015.

[13] S. Calve, A. J. Witten, A. R. Ocken, and T. L. Kinzer-Ursem, "Incorporation of non-canonical amino acids into the developing murine proteome," *Scientific Reports,* vol. 6, pp. 1-7, 2016.

[14] B. Alvarez-Castelao *et al.*, "Cell-type-specific metabolic labeling of nascent proteomes in vivo," *Nature Biotechnology,* vol. 35, pp. 1196-1201, 2017.

[15] D. B. McClatchy, Y. Ma, D. A. Liem, D. C. M. Ng, P. Ping, and J. R. Yates, "Quantitative temporal analysis of protein dynamics in cardiac remodeling," *Journal of Molecular and Cellular Cardiology,* vol. 121, pp. 163-172, 2018.

[16] X. Gao *et al.*, "Dietary methionine influences therapy in mouse cancer models and alters human metabolism," *Nature,* vol. 572, no. 7769, pp. 397-401, Aug 2019.

[17] T. Olsen *et al.*, "Effects of dietary methionine and cysteine restriction on plasma biomarkers, serum fibroblast growth factor 21, and adipose tissue gene expression in women with overweight or obesity: a double-blind randomized controlled pilot study," *J Transl Med,* vol. 18, no. 1, p. 122, Mar 11 2020.

[18] S. Ikeda, M. Sugimoto, and S. Kume, "Importance of methionine metabolism in morula-to-blastocyst transition in bovine preimplantation embryos," *Journal of Reproduction and Development,* vol. 58, pp. 91-97, 2012.

[19] M. Kudo, S. Ikeda, M. Sugimoto, and S. Kume, "Methionine-dependent histone methylation at developmentally important gene loci in mouse preimplantation embryos," *Journal of Nutritional Biochemistry,* vol. 26, pp. 1664-1669, 2015.

[20] S. Tang *et al.*, " Methionine metabolism is essential for SIRT 1-regulated mouse embryonic stem cell maintenance and embryonic development " *The EMBO Journal,* vol. 36, pp. 3175-3193, 2017.
[21] A. M. Saleh, K. R. Jacobson, T. L. Kinzer-Ursem, and S. Calve, "Dynamics of Non-Canonical Amino Acid-Labeled Intra- and Extracellular Proteins in the Developing Mouse," *Cell Mol Bioeng,* vol. 12, no. 5, pp. 495-509, Oct 2019.
[22] H. T. Evans, L. G. Bodea, and J. Gotz, "Cell-specific non-canonical amino acid labelling identifies changes in the de novo proteome during memory formation," *Elife,* vol. 9, Jan 6 2020.
[23] C. A. Smith *et al.*, "METLIN: a metabolite mass spectral database," *Ther Drug Monit,* vol. 27, no. 6, pp. 747-51, Dec 2005.
[24] J. Chong, D. S. Wishart, and J. Xia, "Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis," *Curr Protoc Bioinformatics,* vol. 68, no. 1, p. e86, Dec 2019.
[25] D. B. McClatchy, M. Q. Dong, C. C. Wu, J. D. Venable, and J. R. Yates, "15N metabolic labeling of mammalian tissue with slow protein turnover," *Journal of Proteome Research,* vol. 6, pp. 2005-2010, 2007.
[26] J. C. Price, S. Guan, A. Burlingame, S. B. Prusiner, and S. Ghaemmaghami, "Analysis of proteome dynamics in the mouse brain," *Proceedings of the National Academy of Sciences,* vol. 107, pp. 14508-14513, 2010.
[27] D. E. Hammond *et al.*, "Proteome Dynamics: Tissue Variation in the Kinetics of Proteostasis in Intact Animals," *Mol Cell Proteomics,* vol. 15, no. 4, pp. 1204-19, Apr 2016.
[28] R. P. Brown, M. D. Delp, S. L. Lindstedt, L. R. Rhomberg, and R. P. Beliles, "Physiological parameter values for physiologically based pharmacokinetic models," *Toxicol Ind Health,* vol. 13, no. 4, pp. 407-84, Jul-Aug 1997.
[29] J. U. Streif *et al.*, "In vivo assessment of absolute perfusion in the murine skeletal muscle with spin labeling MRI. Magnetic resonance imaging," *J Magn Reson Imaging,* vol. 17, no. 1, pp. 147-52, Jan 2003.
[30] C. R. Kirman *et al.*, "Physiologically based pharmacokinetic model for rats and mice orally exposed to chromium," *Chem Biol Interact,* vol. 200, no. 1, pp. 45-64, Oct 25 2012.
[31] C. A. Boswell *et al.*, "Comparative physiology of mice and rats: radiometric measurement of vascular parameters in rodent tissues," *Mol Pharm,* vol. 11, no. 5, pp. 1591-8, May 5 2014.
[32] K. L. Kiick, E. Saxon, D. A. Tirrell, and C. R. Bertozzi, "Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, pp. 19-24, 2002.
[33] E. Takach, T. O'Shea, and H. Liu, "High-throughput quantitation of amino acids in rat and mouse biological matrices using stable isotope labeling and UPLC-MS/MS analysis," *J Chromatogr B Analyt Technol Biomed Life Sci,* vol. 964, pp. 180-90, Aug 1 2014.
[34] F. J. Veredas, F. R. Canton, and J. C. Aledo, "Methionine residues around phosphorylation sites are preferentially oxidized in vivo under stress conditions," *Sci Rep,* vol. 7, p. 40403, Jan 12 2017.
[35] R. Wei *et al.*, "Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data," *Sci Rep,* vol. 8, no. 1, p. 663, Jan 12 2018.

[36] E. G. Armitage, J. Godzien, V. Alonso-Herranz, A. Lopez-Gonzalvez, and C. Barbas, "Missing value imputation strategies for metabolomics data," *Electrophoresis,* vol. 36, no. 24, pp. 3050-60, Dec 2015.
[37] Z. Yang *et al.*, "Toward better annotation in plant metabolomics: isolation and structure elucidation of 36 specialized metabolites from Oryza sativa (rice) by using MS/MS and NMR analyses," *Metabolomics,* vol. 10, no. 4, pp. 543-555, 2014.
[38] R. Chaleckis, I. Meister, P. Zhang, and C. E. Wheelock, "Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics," *Curr Opin Biotechnol,* vol. 55, pp. 44-50, Feb 2019.
[39] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: New perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Research,* vol. 45, pp. D353-D361, 2017.
[40] K. F. Steward *et al.*, "Metabolic Implications of Using BioOrthogonal Non-Canonical Amino Acid Tagging (BONCAT) for Tracking Protein Synthesis," *Front Microbiol,* vol. 11, p. 197, 2020.
[41] S. J. Mentch *et al.*, "Histone Methylation Dynamics and Gene Regulation Occur through the Sensing of One-Carbon Metabolism," *Cell Metab,* vol. 22, no. 5, pp. 861-73, Nov 3 2015.

5. CONCLUSIONS AND FUTURE RECOMMENDATIONS

Understanding the regulation of protein synthesis and turnover in developing and adult tissues has been limited due to a lack of tools capable of probing protein dynamics *in vivo*. To address this gap, in this dissertation I sought to develop and characterize an injection-based protein labeling technique that utilizes non-canonical amino acids (ncAAs) to label the nascent proteome of developing and adult murine tissues. In this section, the main findings of the dissertation are summarized and recommendations for future work in the field of *in vivo* protein dynamics are discussed.

5.1 Conclusions

Using the methionine analog azidohomoalanine (Aha), we developed a mass spectrometry (MS)-based method that enables selective identification and quantification of newly synthesized proteins (NSPs) in developing murine tissues. The results presented in **Chapter 2** demonstrate that the method enables isolation of intracellular NSPs with minimum background signal from the unlabeled proteome. The method was then used to isolate NSPs from the embryonic (E) time points E12.5 and E15.5, and successfully identified NSPs that are associated with the developmental processes occurring around each time point, indicating that the method can resolve differences in the nascent proteome of different embryonic time points. Furthermore, we employed a cellular fractionation protocol to fractionate embryonic tissues to different cellular fractions (cytosolic, nuclear, membrane, cytoskeleton, and extracellular matrix (ECM)), and combined it with the Aha labeling technique to probe the dynamics of protein synthesis and turnover in each fraction. The results of this investigation showed that proteins in the different cellular fractions vary in their

synthesis and turnover rates, signifying the capability of the ncAA labeling method to probe protein dynamics within various cellular fractions *in vivo*.

In **Chapter 3,** the Aha enrichment method developed in **Chapter 2** was expanded to enrich for the newly synthesized ECM proteins. To this end, several approaches were methodically tested to overcome the limitation the complex biochemical nature of the ECM poses on the use of MS-based methods to study NSPs. Using the optimized method, we report for the first time the successful use of ncAA labeling to identify NSPs in the ECM *in vivo* and investigate the dynamic regulation of ECM proteins in embryonic and adolescent musculoskeletal tissues. The results of this investigation showed that the method detects differences in proteins synthesized within adjacent developmental time points with high temporal resolution. Additionally, the MS results of the Aha-enriched and unenriched samples demonstrate that Aha enrichment reveals information about ECM dynamics that are missed in traditional static proteomic analyses. Finally, to gain insights into the spatial dynamics of the ECM NSPs during musculoskeletal development, we extended the labeling technique to visualize the labeled ECM proteins in decellularized forelimb tissues.

In **Chapter 4**, we investigated the applicability of the method to probe protein dynamics in adult murine tissues. To this end, we quantified the biodistribution kinetics of Aha in adult mice following a single Aha injection and measured the relative degree of protein labeling in different tissues using fluorescent western blotting. Using these measurements, we developed a quantitative model of Aha biodistribution and incorporation into cellular proteins. The model enabled estimating the relative rates of protein synthesis and turnover in the different tissues and predicting the degree of tissue protein labeling using various dosing regimens.

In our experimental setup, the calculated relative rates of protein synthesis and turnover in brain and skeletal muscle are lower than those in liver and kidney, which are in agreement with previous reports that utilized isotope labeling. Finally, we investigated the impact of Aha incorporation into cellular proteins on normal physiology by analyzing the plasma metabolome following Aha injection and showed that Aha administration does not significantly perturb cellular functions as reflected by an unchanged plasma metabolome compared to non-injected controls.

5.2 Recommendations for Future Work

Collectively, the results presented in this dissertation demonstrate the effectiveness of the injection-based ncAA labeling technique in mapping the spatiotemporal dynamics of intra- and extra-cellular proteins in the developing mouse and demonstrates the applicability of the method to study NSPs and protein turnover in adult mouse models. Indeed, further work is warranted to realize the full potential of ncAA labeling in tracking changes in protein regulation in both development and disease. Herein, two recommendations for future work are highlighted:

The first recommendation focuses on the application of ncAA labeling to advance the field of regenerative medicine by understanding the role of ECM in musculoskeletal tissue regeneration. To achieve this goal, the method should be utilized to map the changes in protein expression across different stages of development. In the work presented in this dissertation, we have established the feasibility of analyzing the temporal and spatial dynamics of ECM proteins in forelimb tissues of selected developmental time points. As such, future studies in this field should focus on the comprehensive profiling of the nascent ECM proteome synthesized during the course of musculoskeletal development and the resolution of its spatial distribution within the different

tissues of the musculoskeletal system. Such analyses will provide valuable insights into the design of scaffolds for tissue regeneration, which importance cannot be overstated.

The second recommendation is directed towards advancing the ncAA labeling technique in mouse models. Measurements of protein turnover in pathological states will enable better understanding of the mechanisms of protein dysregulation underlying the disease and identifying therapeutic targets to restore homeostasis. In this dissertation, the kinetics of protein synthesis and turnover were analyzed using western blotting. Therefore, the development of MS-based methods is of utmost importance to enable accurate quantitation of the turnover rates of individual proteins in embryonic and adult mouse. Calculation of turnover rates in animals is challenging as it requires designing the labeling experiment in a way that allows mapping both high and low turnover proteins. High turnover proteins may become fully unlabeled even before the cellular protein pool is sufficiently labeled. On the other hand, low turnover proteins have a slow rate of label incorporation such that longer labeling periods are typically required to enable adequate labeling. As such, according to the tissue under investigation, tissue sampling should be optimized to allow covering both high and low turnover proteins.

Turnover studies performed in adult tissues assume a state of constant cellular protein concentrations. Such steady-state assumption is, however, not valid in embryonic tissues where cells undergo rapid change in protein abundance due to growth. Therefore, methods that distinguish between changes in protein abundance as a result of growth and that due to actual protein turnover need to be employed. One potential strategy for measuring abundance-corrected turnover rates is using a secondary labeling step with isobaric tags such as isobaric tags for relative and absolute quantitation (iTRAQ) and tandem mass tag (TMT). In this method, quantitative reporter ions derived from the isobaric tags in the MS/MS spectra can be used to determine the

proportion of protein that is degraded and that is newly synthesized at each time point. This method has been successfully applied to measure turnover rates in in the fast growing *Streptomyces coelicolor* bacterium by combining SILAC and iTRAQ [1]. To the best of our knowledge, dual labeling with ncAA and isobaric tags has not yet been performed in embryonic tissues. Accordingly, research in this area should focus on evaluating the feasibility of this dual labeling approach for measuring turnover rates during embryonic development.

5.3 References

[1] K. P. Jayapal *et al.*, "Multitagging proteomic strategy to estimate protein turnover rates in dynamic systems," *J Proteome Res,* vol. 9, no. 5, pp. 2087-97, May 7 2010.

APPENDIX A

This appendix contains supplementary information to Chapter 2.

Supplementary Figure

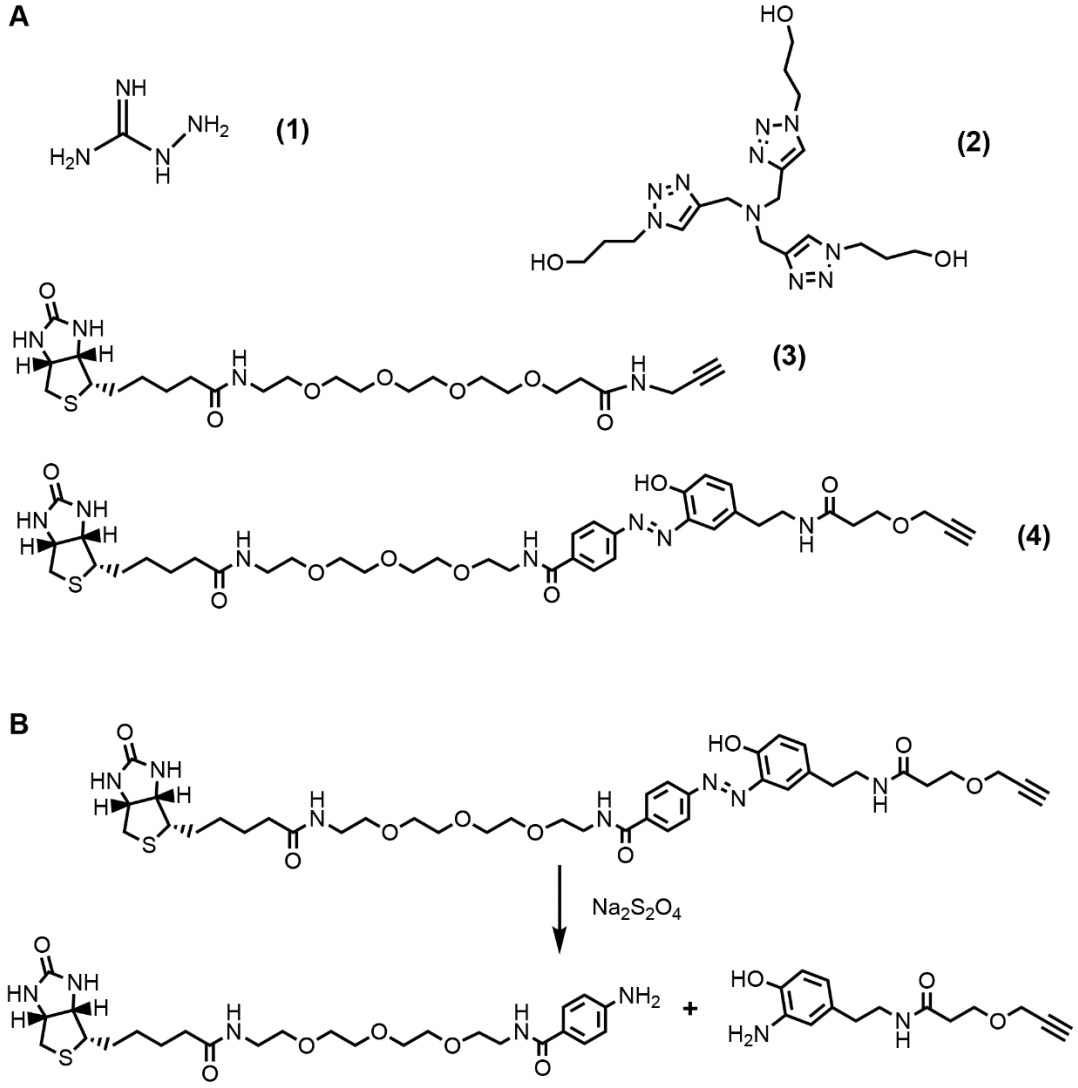


Figure A.1 **(A)** Chemical species used for CuAAC. (1) Aminoguanidine (AG); (2) tris(3-hydroxypropyltriazolylmethyl)amine (THPTA); (3) Non-cleavable biotin-alkyne linker (4) Diazo biotin-alkyne (DAB) linker. **(B)** The diazobenzene group of DAB is cleaved by $Na_2S_2O_4$ under neutral pH conditions.

Supplementary Tables

Table A.1 MaxQaunt parameters

Aha/PBS Samples		
Parameter	**Aha**	**PBS**
General		
MaxQuant Version	1.6.1.0	
Group	0	
Fraction	Yes	
LC-MS run type	Standard	
Multiplicity	1	
Labels	N/A	
Digestion		
Enzyme mode	Specific	
Enzymes	LysC, Trypsin	
Maximum missed cleavages	2	
Separate enzyme for first search	FALSE	
Modifications		
Fixed modifications	Carbamidomethyl on Cys (+57.012)	
Variable modifications	Oxidation of Met (+15.995 Da)	
	Met replacement by Aha (-4.986Da)	
	Met replacement by cleaved diazobiotin triazole [alkyne-azide] (+257.145Da)	
Maximum number of modifications per peptide	5	
Separate variable modifications for first search	FALSE	
Sequences		
Fasta files	Uniprot *Mus musculus* (November 2018)	
	Uniprot Avidin *Gallus gallus* (May 2018)	
	Uniprot contaminants (January 2018)	
Include contaminants	TRUE	
Decoy mode	Revert	
Special AAs	FALSE	
Minimum peptide length	7	
Maximum peptide mass	4600 Da	
Minimum peptide length for unspecific search	8	
Maximum peptide length for unspecific search	25	
Identification		
PSM FDR	0.01	

Protein FDR	0.01
Site FDR	0.01
Minumum unique peptides	0
Minumum razor + unique peptides	1
Minimum peptides	1
Minimum score for unmodified peptides	0
Minimum score for modified peptides	40
Minimum delta score for unmodified peptides	0
Minimum delta score for modified peptides	6
Main search maximum combinations	200
Base FDR calculations on delta score	FALSE
Razor protein FDR	TRUE
Second peptides	TRUE
Find dependent peptides	FALSE
Match between runs	TRUE
Labeled amino acid filtering	FALSE
uantification	
Modifications included in protein quantification	Oxidation of Met (+15.995 Da) Met replacement by Aha (-4.986Da) Met replacement by cleaved diazobiotin triazole [alkyne-azide] (+257.145Da)
Peptides used for protein quantification	Unique + razor
Discard unmodified counterpart peptides	TRUE
Label minimum ratio count	2
Label-free quantification	FALSE
LFQ min. ratio count	N/A
Fast LFQ	N/A
Separate LFQ in parameter groups	N/A
Stabilize large LFQ ratios	N/A
Require MS/MS for LFQ comparisons	N/A
LFQ norm for sites and peptides	N/A
iBAQ	FALSE
iBAQ log fit	N/A
Re-quantify	FALSE
Advanced ratio estimation	FALSE
Instrument	
Instrument type	Orbitrap
First search peptide tolerance	20
Main search peptide tolerance	4.5 ppm
Individual peptide mass tolerance	TRUE
Isotope match tolerance	2 ppm
Centroid match tolerence	8 ppm
Centroid half width	35 ppm

Time valley factor	1.4
Isotope time correlation	0.6
Theoretical isotope correlation	0.6
Recalibration unit	ppm
Use MS1 centroids	FALSE
Use MS2 centroids	FALSE
Intensity dependent calibration	FALSE
Min. peak length	2
Max. charge	7
Min score for recalibration	70
Cut peaks	TRUE
Gap scans	1
Advanced peak splitting	FALSE
Intensity threshold	0
Intensity determination	Value at maximum
MS/MS match tolerance (FTMS)	20 ppm
MS/MS de novo tolderance	10 ppm
MS/MS deisotoping tolerance (FTMS)	7 ppm
Top peaks per Da interval (FTMS)	12
Top x mass window [Da] (FTMS)	100
MS/MS deisotoping (FTMS)	TRUE
Higher charges (FTMS)	TRUE
Water loss (FTMS)	TRUE
Ammonia loss (FTMS)	TRUE
Dependent losses (FTMS)	TRUE
FTMS recalibration (FTMS)	FALSE
MS/MS match tolerance (ITMS)	0.5 Da
MS/MS de novo tolerance (ITMS)	0.25 Da
MS/MS deisotoping tolerance (ITMS)	0.15 Da
Top peaks per Da interval (ITMS)	8
Top x mass window [Da] (ITMS)	100
MS/MS deisotoping (ITMS)	FALSE
Higher charges (ITMS)	TRUE
Water loss (ITMS)	TRUE
Ammonia loss (ITMS)	TRUE
Dependent losses (ITMS)	TRUE
ITMS recalibration	FALSE

Table A.2 Cellular fractionation buffers

Buffer	Components
Cytosolic (**C**)	50 mM HEPES (pH 7.9), 1 mM $MgCl_2$, 50 mM KCl, 2 mM EDTA, 25 mM sucrose, 5% glycerol, 5 mM sodium orthovanadate, 1× protease inhibitors, 0.1% benzonase
Nuclear (**N**)	50 mM HEPES (pH 7.9), 1mM $MgCl_2$, 100 mM KCl, 2 mM EDTA, 5% glycerol, 5 mM sodium orthovanadate, 1× protease inhibitors, 0.1% benzonase
Membrane (**M**)	50 mM HEPES (pH 7.9), 1 mM $MgCl_2$, 150 mM KCl, 2 mM EDTA, 5% glycerol, 0.5% sodium deoxycholate, 0.5% NP-40, 5mM sodium orthovanadate, 1× protease inhibitors, 0.1% benzonase
Cytoskeletal (**CS**)	50 mM Tris-HCl (pH 6.8), 500 mM NaCl, 2 mM EDTA, 5% glycerol, 2% sodium deoxycholate, 1% NP-40, 1% SDS, 5mM sodium orthovanadate, 1× protease inhibitors
Extracellular matrix (**ECM**)	8 M urea in 100 mM ammonium bicarbonate

APPENDIX B

This appendix contains supplementary information to Chapter 4.

Supplementary Figures

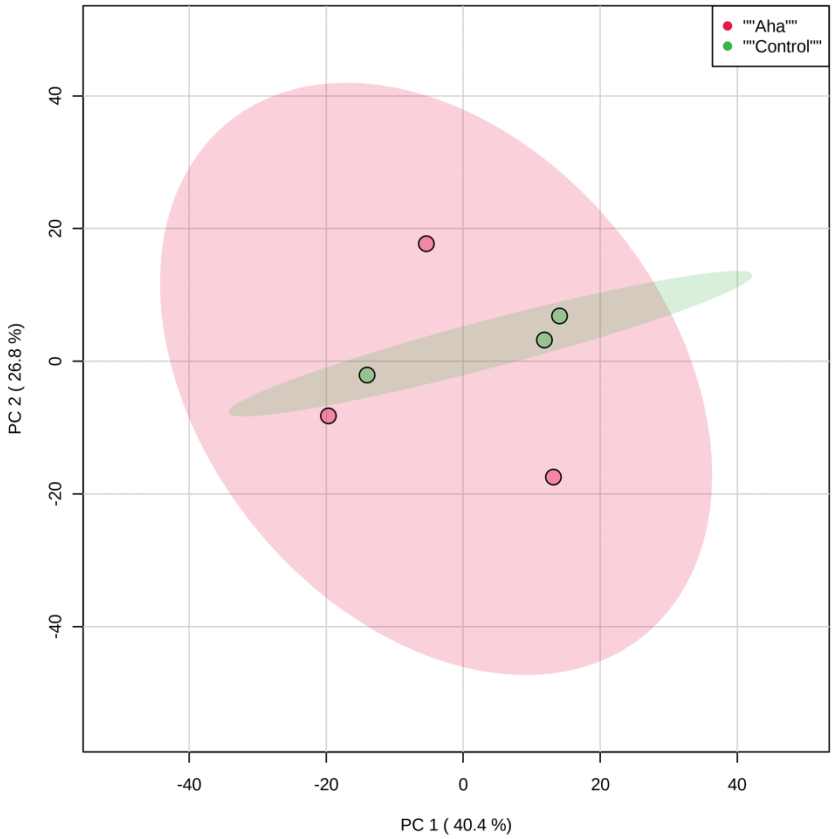


Figure B.1 Principle component analysis (PCA) shows no clear separation of control and Aha-treated groups in the first two components. Components 1 and 2 account for 40.4% and 26.8% of the total data variability, respectively. Green and red dots denote control and Aha samples, respectively. Green and red shaded areas represent the 95% confidence bands of the control and Aha samples, respectively. For the dataset used in the analysis, missing values were imputed using the K-nearest neighbor (KNN) method.

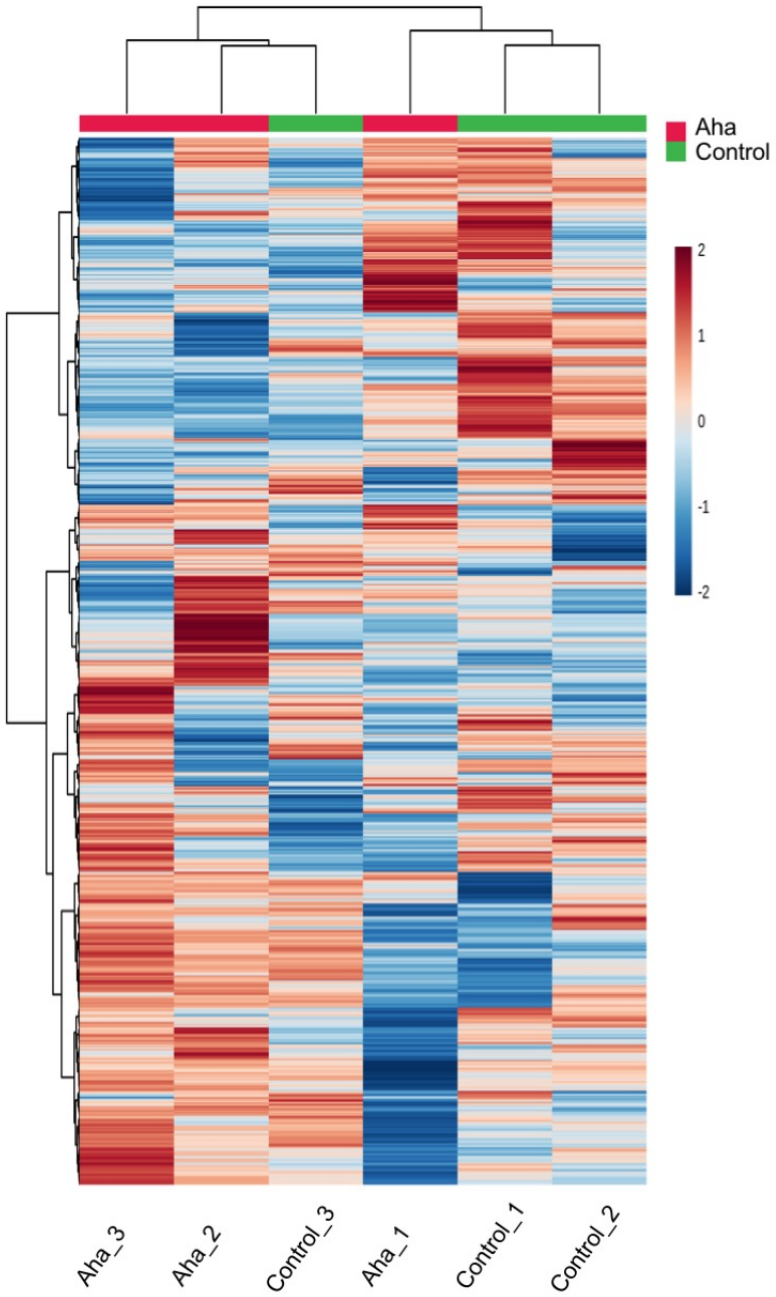


Figure B.2 Unsupervised hierarchal clustering and heatmap of the identified metabolites show lack of clustering between replicates of each group. Color scale indicates metabolite abundance; blue: lowest, red: highest. For the dataset used in the analysis, missing values were imputed using the K-nearest neighbor (KNN) method.

PUBLICATIONS

1. Saleh et al., "Non-canonical amino acid labeling in proteomics and biotechnology," *Journal of Biological Engineering*, vol. 13, p. 43, 2019.

2. Saleh et al., "Dynamics of non-canonical amino acid-labeled intra- and extracellular proteins in the developing mouse," *Cellular and Molecular Bioengeneering*, vol. 12, no. 5, pp. 495-509, 2019.

3. *Puvar and *Saleh et al., "Fluorescent probes for monitoring serine ubiquitination," *Biochemistry*, vol. 59, pp. 1309-1313, 2020. *Equal contribution first author.

4. Libring et al. "The dynamic relationship of breast cancer cells and fibroblasts in fibronectin accumulation at primary and metastatic tumor sites," *Cancers*, vol. 12, p. 1270, 2020.