A Multiplexed Approach for Quantitative Profiling of the Translatome Using Bioorthogonal Non-Canonical Amino Acids

By

Daniel Abram Rothenberg

B.S. Biomedical Engineering
The University of Texas at Austin, 2011

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
OCT 05 2017
LIBRARIES
ARCHIVES

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2017

© Massachusetts Institute of Technology 2017. All Rights Reserved

Author: Signature redacted

Daniel Abram Rothenberg
Department of Biological Engineering
August 24th, 2017

Certified by: Signature redacted

Forest M. White, Ph.D.
Professor of Biological Engineering
Thesis Supervisor

Accepted by: Signature redacted

Mark Bathe, Ph.D.
Professor of Biological Engineering
Chairman, Graduate Program Committee of Biological Engineering

This Doctoral Thesis has been examined by a committee of the Biological Engineering Department

Forest M. White, Ph.D.
Committee Member and Thesis Supervisor
Professor of Biological Engineering
Massachusetts Institute of Technology

Douglas A. Lauffenburger, Ph.D.
Chairperson, Graduate Thesis Committee
Professor and Head of Biological Engineering
Massachusetts Institute of Technology

Roger J. Davis, Ph.D.
Committee Member
Professor of Molecular Medicine
University of Massachusetts Medical School

THIS THESIS IS DEDICATED TO THE MEMORY OF
MAX AUFLICK,
BEATRICE ROTHENBERG,
IRVING ROTHENBERG,
RITA ROTHENBERG-MCGREW,
AND WALTER MCGREW

A Multiplexed Approach for Quantitative Profiling of the Translatome Using Bioorthogonal Non-Canonical Amino Acids

By

Daniel Abram Rothenberg

Submitted to the Department of Biological Engineering
August 2017 in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biological Engineering

Abstract

One of the major goals of systems biology is understanding how a cell changes from a healthy state to a diseased state. Entire fields of cell biology have been built around studying how changes in the type and abundance of specific biomolecules affect disease status. However, one major knowledge gap in systems biology is the quantification of protein synthesis rate at any given time (i.e the *Translatome*). At this time, measurements of protein synthesis rates are limited to methods that use mRNA abundance as a proxy; however, there are regulatory steps on the level of translation that can confound correlation between mRNA abundance and protein synthesis rates. Here, I improve upon a proteomics based method for measuring newly translated proteins, biorthogonal non-canonical amino acid tagging (BONCAT), and adapt it for robust quantitative multiplexing analysis. In the BONCAT method, cells are pulsed with azidohomoalanine (Aha), a methionine analog that contains an azide functional group, such that proteins synthesized for the duration of the pulse incorporate Aha. By coupling pulsed stable isotope labeling of amino acids in cell culture (pSILAC) and Aha metabolic labeling of newly synthesized proteins with strain-promoted azide-alkyne cycloaddition and tandem mass tag (TMT) labeling, I am able to quantitatively interrogate the translatome in a multiplexed manner with high sensitivity and high temporal resolution.

The multiplexed BONCAT protocol was applied to observe changes in temporal protein synthesis during the unfolded protein response (UPR) and epidermal growth factor (EGF) stimulation. Eliciting the UPR by blocking N-glycosylation results in a global downregulation of protein translation, but upregulation of several key protein-folding chaperones. Furthermore, protein translation machinery (ribosomal proteins, initiation factors, and elongation factors) are downregulated to a much greater extent. In contrast to the UPR stress response, pro-growth EGF stimulation resulted in the upregulation of protein translation machinery. EGF stimulation also resulted in waves of temporally distinct protein synthesis, beginning with immediate and delayed early genes and followed by late response genes that determine cell fate. By sampling protein synthesis at both 30 minute and 15 minute intervals, I was able to further elucidate the order of protein synthesis with high temporal resolution. Comparison of protein translation with RNA sequencing and ribosome footprinting revealed tight correlations between RNA,

ribosome occupancy, and protein synthesis. This comparison also allowed the distinction between protein synthesis driven by an increase in transcription versus that driven by an increase in translation. Interestingly, temporal delays between ribosome occupancy and protein synthesis were observed in many genes. These genes also demonstrated a unique codon bias compared to the average codon usage of the genome. An analysis of codon frequency revealed changes in global codon usage over time following EGF stimulation. Changes in chemical modifications of tRNA isoacceptors were also observed which may play a role in regulating protein translation.

Finally, our multiplexed BONCAT method was leveraged to compare the translation response between MEK inhibitor resistant (MelJuso) and sensitive (MM415) melanoma cell lines. Using partial least squares regression (PLSR) and gene set enrichment analysis (GSEA), upregulation of melanoma lineage-dependent transcription factor MITF and MITF targets was observed in MM415s after binimetinib treatment, with no such response in the MelJuso cells. Using a small molecule inhibitor against MITF, we found that MITF inhibitions results in a protective effect against binimetinib in MM415 cells. However, the MM415 cells were resistant to siRNA-mediated knockdown of MITF. Further work needs to be done to characterize the role of MITF in the context of binimetinib sensitivity.

Thesis Supervisor: Forest M. White, Ph.D.
Title: Professor of Biological Engineering

Acknowledgements

My journey into the biological sciences began my sophomore year of high school in Mr. Zumwalt's pre-AP biology class. I was immediately captivated by the intricate and almost fantastical inner workings of a cell that define what makes an entity "alive." Between then and now, I've had so many mentors, teachers, professors, colleagues, and friends who have shaped me and molded me into the scientist that I am today.

My first research experience was through a summer undergraduate research program at The University of Texas Medical School at Houston under the guidance of Dr. Catherine Denicourt and Chris Castle. The following summer, I joined the lab of Dr. Mong-Hong Lee at the MD Anderson Cancer Center and worked closely with two graduate students, Chris Gully and Ed Wang, who took me under their wings and gave me the confidence that I could one day have a research project to call my own. My experiences those two summers ignited a passion for biological research that directed me down the path of graduate school at MIT.

My thesis committee of Roger Davis, Doug Lauffenburger, and Forest White has offered me incredibly valuable guidance, feedback, and encouragement through the many twists and turns of graduate school. As chair of the BE Department, Doug has fostered an incredibly welcoming and collaborative environment that breeds successful research and bright young scientists. I can't imagine a more perfect mentor than Forest. His patience is second to none, and his laughter and enthusiasm made the lab a fun place to come to work every day. The countless hours spent in his office figuring out why certain protocols were failing were instrumental in drawing out the problem-solver in me. My lab mates in the White lab have also been a real pleasure to work with over the years. I've been fortunate to work besides talented minds like Amanda, Jason, Emily, Nate, Scott, Bryan, Nina, Abhinav, Hannah, Dan, Tim, Jose, Aaron, Rebecca, Liz, Jackie, Yang, Raven, Vibhuti, Antje, Kristina, Nader, Lauren, Ishwar, and Jacqueline. An especially big thanks to (Professor) Bryan Bryson, Amanda Del Rosario, and Aaron Gajadhar who were especially important mentors to me during my first year in the White lab.

I've been so fortunate to have an incredible cohort to work with over these last 6 years. The BE entering class of 2011 has provided some amazing support and close friends whose hard work and successes kept me motivated when things got tough. Amongst my friends at MIT, I'm especially grateful to Raven, Gabi, Allison, Alec, Thomas, Jenny, Vyas, Andrew, Nate, Devin, Brian, Raja, Marcus, Kelly, Mark, Tu, the rest of the Funsies crew, Koch Lunch, The Kinnaird Street crew, The 9 Seattle squad, The 23 Tufts guys, the BE Natural Killers, fellow BE Board Members, and really everyone else. There is a lot of redundancy there, and even so I don't think I fully captured to scope of everyone in this community that has shaped me as a scientist and as a person.

I don't know if I would have survived graduate school with my sanity intact if I hadn't had friends outside of MIT to keep me even-keeled. Hiking, skiing, and beer+soup adventures with Ann, Bryan, Allison, Pau, Deidre, Dan, Joe, Ron, Deyo, and Cassie are some of my fondest memories during my time here. When I moved to Boston from Texas, I left a lot

of really close friends back home, but The DLoB a.k.a. The Thread has been awesome for laughs, stimulating conversations, and the occasional argument over who is the real King of the North.

If you've ever spoken to me for more than 10 minutes, then you've probably heard me refer to the November Project at least once. This incredible group of athletes and human beings yanked me out of the grad school bubble and gave me a community to wake up with before the sun rises and to have beers with after the sun sets. This group is also where I met my wonderful girlfriend, Carolyn McCrone, who has been a constant source of love and support throughout the ups and downs over the last 3 years. There are way too many other people to name individually who have given me balance to my life, but know that I'm grateful for all of you. I'm a much better-rounded person with this tribe in my life.

Finally, I absolutely would not be sitting here finishing my thesis with tears in my eyes if it weren't for my loving family. My brothers -- Jared, Ben, and Jacob -- have been my best friends as long as I can remember, and moving half a country away was one of the hardest things I've ever done in my life. Each and every one of them is blazing a path in their careers, whether it's Jared mentoring young student athletes as a high school math teacher and coach, Ben playing an integral role in a young startup digital marketing company, or Jake beginning his senior year of high school as one of the stud pitchers on the varsity baseball squad. I'm so proud of them, and I draw inspiration from them every day. The fact that all four of us have achieved some degree of success is a real testament to how we were raised. My parents, Lisa and Scott, have always had our backs and have been our number one fans. They instilled in us "The Standard of Excellence," and it's a mindset that all four of us take with us to work every day. Thank you, Mom and Dad, for the unconditional love and support. Even though I'm nearly 2000 miles away, this work would not be possible without you. I love you so much.

Table of Contents

Table of Figures

Chapter I

Introduction and Background

Daniel Rothenberg

1.1 The Central Dogma

The Central Dogma of biology describes the flow of information within a cell from a stable storage medium (DNA), to an intermediate transcript (RNA), and finally to functional protein machinery that changes cell behavior[1]. Proteins have a broad spectrum of functionalities including altering the structure and shape of the cell, acting as an enzyme to catalyze biochemical reactions, and serving as intermediates in signaling cascades to further transmit information that affect gene expression and cell phenotype. The three key steps of the central dogma are (1) transcription of a RNA from DNA, (2) processing of a RNA transcript into a mature messenger RNA (mRNA), and (3) translation of mRNA into protein. Each step has its own set of regulatory mechanisms that tightly control protein expression in healthy cells.

1.1.1 Transcription

The concept of an unstable intermediate messenger transferring instructions for protein synthesis from the nucleus to ribosomes was first postulated in 1961 by Jacob and Monod[2]. That same year, the existence of mRNA was proven by Brenner, Jacob, and Meselson[3]. The process of generating an RNA transcript from a DNA template is termed *transcription*. Transcription is the first step in gene expression and is initiated by the binding of transcription factors upstream of the coding sequence in the target gene[4]. Transcription factors recruit an RNA polymerase to the transcription start site (TSS). The RNA polymerase then unwinds the DNA double helix and breaks the hydrogen bonds between base pairs, resulting in two single DNA strands[5]. The RNA polymerase then uses the anti-sense (or noncoding) strand as a template to initiate and elongate the RNA

transcript using ribonucleotides. Transcription is terminated once the RNA polymerase reaches a terminator sequence at the end of the gene. The terminator sequence triggers the release of the RNA polymerase from the DNA, and the newly synthesized nascent transcript continues to further processing prior to translation[6].

1.1.2 RNA Processing

Unlike DNA which is highly stable, RNA transcripts are transient and subject to chemical modifications before being translated into a protein. Being a single-stranded nucleic acid, RNAs are subject to higher rates of hydrolysis and enzymatic degradation than double-stranded DNA. Several chemical modifications are made to pre-mRNA transcripts to enhance its stability before it gets translated. For instance, a 7'-methylguanidine cap is added to the 5' end of the transcript. Unlike traditional 5'-to-3' phosphoester linkages, the 5'-cap consists of a 5'-to-5' linkage between the first base of the transcript and the 7'-methylguandine nucleotide[7]. This cap is added concurrent with transcription within the nucleus[8] and is essential for nuclear export[9], splicing[10], and translation initiation[11]. The cap also protects against 5' exonucleases that would otherwise degrade uncapped RNAs[12,13]. Once the RNA is capped and completely transcribed, it is shuttled through a nuclear pore and into the cytoplasm[9]. In the cytoplasm, the cap binds translation initiation factors to facilitate the recruitment of ribosomes to the RNA transcript[11,14].

A polyadenylate [poly(A)] tail is also added to the 3' terminus of the transcript. A multimeric protein complex recognizes a six nucleotide motif[15] towards the 3' end of the transcript and cleaves it. The newly generated 3' terminus is then modified with a poly(A)

tail[16]. The poly(A) tail is bound by poly(A) binding protein (PABP), facilitating export from the nucleus and protecting the transcript from degradation[17]. PABP also recruits translation initiation factors to facilitate translation[18].

The pre-mRNA transcript also undergoes splicing prior to translation. Nascent pre-mRNA contains a combination of introns (non-coding regions), and exons (coding regions). The process of splicing removes the non-coding introns and ligates the exons together in a single contiguous transcript[19]. Alternative splicing is mechanism by which different combinations of exons can be joined together, allowing many unique isoforms to be generated from a single gene[20].

1.1.3 Translation

The final step in producing a functional protein is the decoding of the mRNA transcript into a sequence of amino acids through the process of *translation*. Translation can occur by two different and distinct mechanisms: cap-dependent translation and cap-independent translation.

Cap-dependent translation is initiated by the binding of eukaryotic translation initiation factors (eIFs) to the 5'-cap, the 40S ribosomal subunit, and the methionine-charged initiator tRNA (Met-$tRNA_i$*). EIF2 is a GTPase that is important for delivery of the Met-$tRNA_i$* to the 40S ribosome subunit. When eIF2 is in the GTP-bound state, it associates with the Met-$tRNA_i$* complex and delivers it to the P-site of the 40S ribosome subunit, forming the 43S preinitiation complex (PIC)[21].

The eIF4F complex, which is bound to the 5'-cap of the mRNA, recruits the 43S PIC and binds the 3'-poly(A) tail to circularize the transcript[17]. EIF4F consists of three

proteins, each with its own functionality. EIF4E recognizes and binds the 5'-cap[22]. EIF4A is a helicase that dissociates secondary structures that may be present in the 5'-UTR[23]. EIF4G serves as bridge for the other eIF4F subunits as well as the 40S ribosome subunit[24] and facilitates circularization of the transcript by binding the 3'-poly(A) tail[25]. Once the 43S PIC is assembled on the transcript, it scans through the 5'-UTR of the transcript until the AUG start codon is found[26]. At that point, eIF2's GTPase activating protein (GAP), eIF5[27], associates with the 43S PIC along with eIF5B, resulting in hydrolysis of GTP to GDP and dissociation of eIF2 and associated factors from the PIC[28]. Following release of eIF2, the 60S subunit binds to the 40S subunit to complete the full 80S initiation complex, and elongation of the nascent polypeptide begins.

In recent years, an alternative mechanism of cap-dependent translation initiation has been uncovered. A subset of mRNA transcripts contain a *cis*-acting RNA structural element that preferentially recruits eIF3D instead of eIF4F as the cap-binding initiation factor. This novel initiation process was first observed in JUN transcripts, and has been proposed to be a mechanism by which specific transcripts can continue to be translated when the eIF4F complex is inhibited during cell stress events[29].

In contrast to cap-dependent translation, cap-independent translation does not require the translation initiation complex associated with the 5'-cap. For a long time, it was thought that eukaryotic translation could not take place in the absence of a 5'-cap. The discovery of a translationally-active uncapped picornaviral mRNA introduced the concept of cap-independent translation[30]. In this scenario, the 40S subunit binds directly to a secondary RNA structure called an internal ribosome entry site (IRES) within the

mRNA transcript. Recruitment of the 40S subunit to an IRES is facilitated by cofactors called IRES transactivating factors (ITAFs)[31].

While prevalent in viral translation systems, cellular cap-independent translation is rare compared to cap-dependent translation. An effort to catalog all the IRESes in the human genome have yielded less than 40 confirmed hits[32]. One of challenges associated with identifying cellular IRESes is that they lack any sort of consensus sequence. Instead, IRESes may be identified based on their secondary and tertiary structures[33]. Genes confirmed to contain an IRES include BIP[34], VEGF[35], MYC[36], FGF-2[37], and eIF4G[38].

In the presence of stress perturbations, the activity of cap-dependent translation is reduced (see Section 1.2 for translational control mechanisms), and cap-independent translation becomes the predominant mode of protein production. Many stress-response proteins are under the control of IRES-like systems, allowing for their production even in the presence of global translation downregulation through inhibition of cap-dependent translation. This phenomenon is investigated further in Chapter 3 of this thesis.

1.2 Mechanisms of Translational Regulation

Regulation of translation initiation can occur through posttranslational modifications of initiation factors. Phosphorylation of initiation factors allows cells to respond rapidly to perturbations by regulating the production of new proteins without having to wait for the transcription and translation of regulatory elements. Structural elements within the mRNA transcripts are also capable of regulating translation.

1.2.1 Phosphorylation of eIF2α

EIF2 is one of the initiation factors that can be phosphorylated in response to cellular stress. It was originally shown that phosphorylation of the eIF2α subunit on serine 51 in rabbit reticulocytes results in its inability to bind Met-tRNA$_i$*[39]. In addition to blocking this interaction, phosphorylation of eIF2 also increases its affinity for its guanine exchange factor (GEF), eIF2B. Phospho-eIF2 serves at its own inhibitor by sequestering eIF2B, preventing it from interacting with the unmodified GTPase[40]. Because eIF2 is expressed at a higher level than eIF2B, only a subset of all cellular eIF2 needs to be phosphorylated to sequester eIF2B and block translation initiation.

Multiple stress pathways result in the phosphorylation of eIF2α S51. In response to hemin depravation, reticulocytes activate the heme-regulated eIF2α kinase (HRI) to halt translation[41]. Viral infections oftentimes result in the production of double stranded RNA (dsRNA). Protein kinase R (PKR) is directly activated by dsRNA and phosphorylates eIF2α as a result[42]. Amino acid depravation causes cells to activate GCN2, which can also phosphorylate eIF2α[43]. Finally, in response to endoplasmic reticulum (ER) stress, cells activate PKR-like ER kinase (PERK), which dissociates from the ER membrane and phosphorylates eIF2α[44]. This mechanism is discussed in more detail in Chapter 3 of this thesis.

1.2.2 4E Binding Proteins

Another mechanism of regulation on the level translational involves disruption of eIF4F complex formation at the 5'-cap of the mRNA. EIF4E directly binds to the cap, and serves as a scaffold for other components of the eIF4F complex[22]. A family of eIF4E

binding proteins (4E-BPs) compete with eIF4G for the binding site on eIF4E[45]. 4E-BPs interact with eIF4E efficiently when they are hypophosphorylated; upon phosphorylation, they lose affinity for eIF4E and become displaced by eIF4G, facilitating the assembly of the eIF4F complex and allowing translation to proceed[17]. In response to pro-growth signals, 4E-BPs are phosphorylated by FKAB/mTOR resulting in an increase in translation[46]. Notably, it has been demonstrated that mTOR regulates the expression of a set of pro-invasion proteins at the level of translation in prostate cancer through phosphorylation of 4E-BP[47].

1.2.3 Alternative Open Reading Frames

Upstream open reading frames (uORFs) are small ORFs that reside in the 5'-UTR of mRNA transcripts. UORFs are characterized by an AUG start codon upstream of the main coding sequence, and regulate translation by diverting ribosomes away from the start codon of the main gene[48]. While less than 10% of genes contain uORFs in their 5'-UTRs, oncogenes appear to be enriched for uORFs, with roughly two-thirds of oncogenes containing a start codon within the 5'-UTR[49].

After translating a uORF, the ribosome can either reinitiate and continue translating subsequent ORFs, or it can fail to reinitiate and dissociate from the transcript[50]. The current prevailing hypothesis is that the limiting step for ribosome re-initiation with the downstream ORF is the reassembly of initiation factors. This hypothesis is supported by observations that longer intercistronic sequences correlate with higher rates of re-initiation by allowing more time for the reassembly of initiation factors with the ribosome[51]. Furthermore, lengthening the uORF has been shown to decrease the efficiency of re-

initiation because the longer uORF allows more time for initiation factors to dissociate from the ribosome[52]. UORFs can also regulate translation by causing ribosomes to stall, preventing translation of the downstream ORF. This model is supported by a study on the cytomegalovirus gpUL4 mRNA that shows an initiation rate of just 10% on a uORF is sufficient to decrease translation of the transcript by 10-fold[53].

In addition to regulating translation of the primary ORF by diverting ribosomes, translation of uORFs also produce a peptide product that in some cases, have been shown to modulate translation in *cis*. For example in *Escherichia coli*, a uORF of the tryptophanase gene codes for a 24-residue peptide that prevents ribosome dissociation following induction by tryptophan, leading to increased translation of the primary ORF[54].

Finally, it has been demonstrated that translation originating from alternative ORFs can trigger nonsense mediated decay of the mRNA transcript. This decay arises when the 5'-UTR is particularly short, causing ribosomes to scan through the canonical AUG codon to a downstream, frame-shifted AUG codon[55].

1.2.4 Codon Bias Switching

Each codon consists of three sequential nucleotides, with four different bases possible at each position, allowing for 64 unique codons. However, there are only 20 different amino acids (plus three stop codons), such that specific amino acids may be encoded by multiple codons, resulting in a degenerate genetic code. For a given amino acid with multiple codons, one might expect that each codon would be used with equal likelihood, but this is not the case. The phenomenon of one particular codon for a given amino acid being used more frequently than others is called *codon bias*. While the exact

cause of codon bias is not fully understood, it is thought to arise as a balance between evolutionary selection for a specific set of codons and random mutations away from the desired codon set[56].

Recent work has shown that codon bias may also play a role in regulating translation of specific proteins in response to perturbations. For example, a set of persistence genes are upregulated in *Mycobacterium bovis* under hypoxic conditions via chemical modification of tRNAs at the wobble position of the anticodon such that the codon bias changes to favor a specific set of mRNA transcripts[57]. Unlike other mechanisms of translational control that primarily regulate translation initiation, this phenomenon of codon bias switching regulates translation at the elongation phase.

1.3 Quantitative Mass Spectrometry

Mass spectrometry (MS) is an incredibly versatile tool used across a broad spectrum of applications. MS analysis is an attractive method for biological applications because of its capability of analyzing biomolecules in a high-throughput, discovery-mode manner. MS has been used to analyze many different types of biomolecules, including RNA, proteins/peptides, lipids, and small molecule metabolites, among many others. In this thesis, MS analysis is used for the identification and quantification of proteins and peptides.

1.3.1 Types of Mass Spectrometers

The core function of a mass spectrometer is to determine the mass to charge ratio (*m/z*) of a molecule or set of molecules. The method of isolation and measurement of specific *m/z* species depends on the type of instrument being used.

Time of Flight (TOF) mass spectrometers leverage the fact that, given equal charges, larger molecules will travel more slowly down an electrical potential gradient. Ions held in a trap are accelerated through an electric field, with conservation of energy dictating that smaller ions will achieve higher velocities than larger ions of the same charge[58]. Furthermore, for ions of a given mass, more highly charged ions will achieve a greater velocity than those with a lower charge state. After accelerating through an electric field, ions are released into a field-free flight tube. The time required for the ions to traverse the flight tube and hit the detector is proportional to the square root of the mass-to-charge ratio. The resolution of m/z species increases with longer flight tubes. The introduction of reflectrons into flight tubes allows ions to travel longer distances in a shorter flight tube, allowing for greater resolution than previous TOF instruments[59]. TOF mass spectrometers have the benefit of high resolution, dynamic range, and short cycle times, but suffer from the fact that small environmental changes can cause fluctuations in the length of the flight tube, requiring frequent calibration or internal standards.

Triple quadrupole (TQ) mass spectrometers measure ions based on their selection through a series of quadrupoles before hitting a detector. The first quadrupole (Q1) uses a ratio of radio frequency (RF) to direct current (DC) voltages to select a single m/z species according to the Mathieu stability diagram[60] to pass through to the second quadrupole (Q2). In Q2, operated in RF-only mode, the selected ion is fragmented by

colliding the ions with an inert gas through a process termed collision induced dissociated (CID). The fragments then pass into the third quadrupole (Q3), where specific m/z species are then selected again prior to measurement on a detector[61]. TQ mass spectrometers benefit from having very high sensitivity and dynamic range, making them especially useful for targeted applications[61]. The main drawback of TQMS is that they can only select a single m/z species in Q1 and Q3, so obtaining a complete mass spectra requires a long cycle time, making TQMS impractical for discovery-mode applications.

Another class of instruments measures the resonant frequency of oscillating ions to calculate the *m/z* ratio by applying the Fourier Transform. Fourier Transform ion cyclotron resonance mass spectrometers (FTICR-MS) collect ions in a Penning trap, consisting of a magnetic field with electrical plates. The ions are excited to circular orbits at their resonant frequency by an oscillating electric field. When the electric field is removed, the ions remain in stable orbits with frequencies dependent upon their *m/z* ratio. The frequency of ion motions are measured as oscillating induced charges in a pair of electrodes, and these frequencies can be transformed into a mass spectrum[62].

Orbitrap mass spectrometers are similar to FTICR-MS instruments, as they measure m/z species based on their harmonic motion in an electric field. An Orbitrap consists of a cylindrical outer electrode surrounding a spindle-like inner electrode at high voltage. Ions are injected tangentially to the inner electrode, resulting in stable orbits with angular frequencies dependent upon the *m/z* ratio. These orbiting ions generate oscillating electric currents that can analyzed in the frequency domain through the Fourier Transform, yielding precise measurements of *m/z* for a wide range of m/z species simultaneously[63]. Furthermore, by pairing an Orbitrap with a quadrupole and a high

energy collision dissociation (HCD) chamber, isolation and fragmentation can be performed for MS/MS analysis. Orbitrap mass spectrometers can make *m/z* measurements with very high resolution, but high resolution measurements require longer cycle times[64].

For the work done in this thesis, a ThermoFisher QExactive Hybrid Quadrupole-Orbitrap MS was used for analysis. This particular MS consists of a front-end quadrupole for selection of precursor ions in tandem with an Orbitrap analyzer and HCD cell for fragmentation. The QExactive is excellent for proteomics analysis because it allows for high resolution *m/z* measurements with quick cycle times along with an option for MS/MS analysis for determination of peptide sequence and isobaric mass tag quantification.

1.3.2 Quantitative Proteomics

Because each peptide has its own unique chemical characteristics, there are variations in ionization potential (leading to changes in ionization efficiency), ion flight, and fragmentation that prevent quantitative comparison between different species within the same run without the addition of internal standards. There are a number of label-free approaches that give an estimate of relative peptide abundance within a single run such as intensity-based absolute protein quantification (iBAQ)[65] and spectral counting[66]. These methods only give a rough estimate of relative abundance, and suffer from variations that arise from each peptide's unique chemical structure as described above. For a more accurate assessment of absolute and/or relative peptide abundance, internal heavy-isotope labeled standards can be spiked into samples in known quantities[67,68]. By

comparing the intensity of the heavy standard to that of the sample, we can directly measure the absolute quantity of a peptide within a sample.

When comparing the relative amounts of the same species, there are more methods available. When run on an autosampler under identical HPLC conditions, peptides of identical sequences should elute at the same time with similar elution profiles across different runs. By taking the area under the curve (AUC) for a particular extracted ion chromatogram (XIC) in the MS1, relative quantification of that species between different experiments can be achieved[69].

For multiplexing samples into a single MS experiment, stable isotope tags can be utilized to differentiate between samples. Stable isotope labeling of amino acids in cell culture (SILAC) uses naturally occurring amino acids that contain heavy isotopes of constituent atoms[70]. In these experiments, cells are cultured in the constant presence of isotopically labeled amino acids until incorporation of the SILAC amino acids reaches a steady state. By using cells grown in the presence of different SILAC labels, samples can be combined early on during processing, reducing technical error between conditions. When analyzed via MS, peptides that incorporate these heavy amino acids will be shifted to a higher mass. Because the peptides are chemically identical except for the heavy isotopes, they co-elute during chromatographic separation allowing for quantitative measurements by taking the AUC of the XIC for each SILAC channel in the MS1. The disadvantage of SILAC is that the number of mass species present in the MS1 is doubled (for two channels) or tripled (for three channels), inherently increasing the complexity of the sample and decreasing the likelihood that lower intensity ions will be selected for MS/MS analysis.

Isobaric mass tag approaches such as tandem mass tags (TMT)[71] or isobaric mass tagging for relative and absolute quantification (iTRAQ)[72] allow for higher order multiplexing. In this scheme, an amine-reactive isotopically-encoded mass tag is used to label the N-terminus and ε-amine of lysines. The intact tags have identical masses, but contain heavy isotopes of carbon and nitrogen differentially distributed between a reporter region and balancer region. Because their intact masses are the same, peptides co-elute and appear as a single peak in MS1 scans. However, upon HCD fragmentation, the tags liberate low-mass reporter ions at different masses for each channel, allowing for quantification of up to 10 channels in a single experiment. While this approach allows for increased multiplexing, the dynamic range of quantitation is oftentimes suppressed by contaminant ions that are co-isolated with the targeted precursor ion.

1.3.3 Reduction of Sample Complexity

The proteome is a highly complex system. The human proteome consists of over 70000 unique proteins[73] with a median protein length of 375 residues[74] with 20 different possible amino acids at each residue. Due to this high degree of complexity, tryptic peptide mixtures from the proteome are usually separated into fractions based on physical characteristics such as size (size exclusion chromatography[75]), charge (ion exchange chromatography[76]), or hydrophobicity (reverse phase chromatography[77]) prior to MS analysis. In the field of proteomics, MS analysis is most effectively used in tandem with one of these chromatographic separation techniques to reduce the complexity of mixtures.

The observation of specific subsets of the proteome is not feasible by shotgun analysis because observations will be limited to only the most abundant proteins. Various sub-specialties of proteomics have been realized by coupling mass spectrometry with enrichment techniques for simplifying sample complexity. Immunoprecipitation-MS (IP-MS) enriches for a particular protein or class of proteins, allowing for examination of binding partners or post-translational modifications. IP-MS relies on antibodies raised against the targeted protein or moiety, so analyses are limited to targets for which such tools exist. In some cases, exogenous expression of the target protein fused to an affinity tag can be used as a bait to search for interacting proteins. Other targeted approaches seek to examine proteins that contain a specific post-translational modification. For example, immobilized metal affinity chromatography (IMAC) specifically enriches phosphorylated proteins and peptides, allowing for analysis of the phosphoproteome regardless of identity of the modified proteins[78]. A recently developed method called SNOTRAP for analyzing global S-nitrosylation of cysteines relies on the redox state of cysteine sulfhydryls to chemically modify nitrosylated cysteines with an affinity tag for enrichment[79]. Subcellular compartments can also be selectively isolated based on physical properties to analyze the resident proteins of that compartment[80]. Without these enrichment and fractionation techniques, analyzing the proteome or a subset thereof with adequate depth would be nearly impossible.

1.4 Computational Modeling

Traditional molecular biology methods allow researchers to gain an intuitive understanding of results by simple manual inspection of the data. Modern 'omics-based

technologies yield datasets that are extremely large. Sometimes, manual inspection of the extreme outliers in these datasets can yield useful insights into the biological system. However, biological insights frequently lie in more nuanced changes that are not readily identified through manual inspection. The massive amounts of data generated by high-throughput 'omics technologies often necessitate further computational modeling in order to draw biological insight from the results.

1.4.1 Clustering

Clustering is one common and easily implementable approach to analyzing complex data sets to see what biological components behave in the same manner. By looking for functional similarities between members of a cluster, one can determine whether they are part of the same pathway, share a common effector, or have any other biologically relevant commonalities.

K-means clustering partitions the dataset into k distinct clusters and refines the clusters through an iterative process. The process begins by initializing cluster centroids randomly in the dataspace. In each step, the data points are assigned to the cluster associated to the nearest centroid, and the centroid is then recalculated to minimize the sum of the squares between the centroid and the members of the cluster. This process is repeated until convergence, at which the within-cluster sum of squares is at a minimum. The location of the centroids at convergence is dependent upon their randomized location at initialization, so in practice, k-means clustering is repeated many times to generate a co-clustering frequency map that can be used to group network components.

Hierarchical clustering is another approach for grouping network components by agglomerating them in a pairwise manner. In this method, the pairwise distance is calculated between all the points in the network, and the two closest are merged into a centroid. The pairwise distance is then recalculated with the centroid replacing the two original points, and this process is repeated until a single centroid remains. The order of agglomeration and associated distances can then be mapped to a dendrogram for easy visualization of clustering. Depending on the distance metric used (e.g. Euclidean, correlation, rank), different insights can be gained through this algorithm.

1.4.2 Dimensionality Reduction

Large biological datasets are oftentimes highly dimensional and difficult to visualize. To simplify datasets and make visualization possible, the data can be projected into a lower dimensional space that represents the largest variation in the dataset through a process called principal component analysis (PCA). In practice, this is done through a linear mapping of the data by calculating the eigenvalues of the covariance matrix. The largest eigenvalues represent the directions with the greatest variation in the data. By selecting the top eigenvalues and making them the basis of a lower dimensional vector space, the data can be projected onto these orthogonal principal components (PCs) for easier visualization. This projection often results in a natural grouping of network components which is amenable to clustering methods described previously. The loadings of the original dataset (projection of the original axes in the new vector space) may also fall within these clusters, giving a better understanding of which network behaviors

associate with which condition. Furthermore, the PCs oftentimes hold biological significance that would have been impossible to realize in the untransformed dataset.

Self-Organizing Maps (SOMs) are artificial neural networks used to project high dimensional input data into lower dimensional space, and then discretize the input by assigning data points to nodes in the network. SOMs are trained through an iterative process where the node closest to a selected data point is moved towards that data point. SOMs are unique from other clustering algorithms in that the neural network is associated with a topology, meaning the process of moving one node towards the data affects adjacent nodes as well. This process is repeated until the neural network is fine-tuned and represents the underlying data. The visualization of a U-matrix representing the distance between adjacent neurons can be useful for showing the relationship between neighboring nodes.

1.4.3 Predictive Models

The models discussed above are *descriptive*, meaning that they don't associate network behavior with a predicted outcome such as a phenotype. They are useful for characterizing behavior of network components, but don't assign biological significance to the observed behavior. Efforts to correlate changes in network behavior with phenotypic outcomes require one or more *predictive* models. These models are especially useful for generating hypotheses on how perturbing the network can result in a desired phenotypic outcome. One of the most commonly used predictive models in systems biology is linear regression. Linear regression relates inputs (network behavior) and outputs (quantitative phenotypes) such that the outputs are represented as a linear

combination of the inputs. In such a model, the weights associated with each input represent the extent to which that particular network component influences the output.

One problem with using linear regression to model biological networks is that such a system is usually underdetermined (more unknowns than equations), and many of the inputs exhibit multicollinearity when they're part of the same pathway. This limitation can be overcome by placing constraints on the model by penalizing larger models and forcing weights towards zero through regularization[81]. Another approach to simplifying models is by projecting the variables to a new space prior to forming the regression model. Partial least squares regression (PLSR) achieves this by mapping both the input and the output onto principal components in order to maximize the covariance between the projected inputs and outputs[82].

1.5 Characterizing Changes in Gene Expression

Before the development of high-throughput technologies and powerful computers, biological research was traditionally performed with a reductionist approach that focused on individual biomolecules or the interactions between several biomolecules. The invention of high-throughput "-omics" technologies along with the concurrent explosion in computational power ushered in a more holistic approach to cell biology known as *systems biology*. Instead of studying individual biological components in isolation, a systems biology approach examines patterns in the whole network to gain insight into drivers of disease. Large datasets collected from high-throughput screens are input into computational models to gain insights that have not have been immediately obvious by looking at individual network components. The real power in systems biology lies in using

this information to make predictions on what interventions may be most effective in treating diseases.

One of the major goals of systems biology is understanding how a cell transitions from a healthy to diseased state. There are a wide variety of technologies that have been developed to interrogate cell biology to understand the factors that can lead to pathogenesis. However, one aspect of cell biology which lacks robust quantitative methods is the translation of mRNA to functional proteins. Studying translation is hugely important because one of the main ways cells alter their phenotype is by the upregulation and downregulation of new proteins. If measurements can be made regarding how a cell is changing its repertoire of protein machinery, one could gain better insight into the mechanisms that drive disease progression.

1.5.1 RNA Abundance

Traditionally, changes in gene expression have been measured by measuring changes in mRNA transcript expression as a proxy for changes in regulation of protein levels. Before the invention of high throughput screens, specific mRNA transcripts were quantified via Northern Blotting, where whole-cell RNA is separated by electrophoresis and then hybridized with a radio-labelled DNA probe against the transcript of interest[83]. In 1983, Kary Mullis developed polymerase chain reaction (PCR), allowing for the amplification of DNA through sequential melting of double-stranded DNA, annealing of DNA primers, and elongation with a thermostable DNA polymerase[84]. The discovery of reverse transcriptase enzymes in the Rous Sarcoma Virus[85] eventually led to the development of reverse transcription-PCR (RT-PCR). In quantitative RT-PCR, RNA is

first reverse-transcribed into intron-free complementary DNA (cDNA), which can then be amplified via traditional PCR with primers specific to the gene of interest[86] and detected via Southern Blotting or with DNA-associated fluorophores. The amount of amplified cDNA is directly proportional to the initial concentration of cDNA, which in turn is proportional to the amount of mRNA. While this approach can yield robust quantitation for a targeted set of transcripts, it cannot be used for high throughput analysis of mRNA quantitation due to the need for gene-specific primers.

The first method that allowed for high throughput detection of mRNA transcripts was cDNA microarrays. In this approach, single stranded DNA oligomers for thousands of genes are spotted on a solid phase support chip. Upon addition of a fluorescently-labeled cDNA library, hybridization with the immobilized DNA oligomers yields a fluorescent signal on the chip that is proportional to the amount of hybridized cDNA[87]. Despite the high-throughput nature of microarrays, they have a limited sensitivity and dynamic range and are restricted to analysis of known genes.

Next Generation Sequencing methods ushered in RNA Sequencing (RNA-Seq), which quantifies mRNA by sequencing the cDNA library, and counting the number of times each sequence was read. RNA-Seq offers a larger dynamic range, doesn't require a predetermined gene set, and has greater sensitivity than microarray technology[88]. Furthermore, RNA-Seq has the capability to distinguish between splice isoforms, a feature that DNA microarrays lack[89]. RNA-Seq is currently the state of the art for quantifying mRNA abundance.

While there are technologies to accurately profile mRNA abundance in a high throughput manner, relative mRNA abundance has been shown to be an incomplete

predictor of protein synthesis and abundance. As discussed in section 1.2, there are a multitude of regulatory steps between the generation of an mRNA transcript and the translation of that transcript into a protein. One study seeking to quantify mammalian gene expression found that the main predictor of protein abundance is the translation rate of that protein rather than the RNA abundance of the corresponding transcript[90]. Another study examining protein and mRNA in LPS-stimulated dendritic cells found that mRNA abundance explains roughly 65% of protein abundance at steady state. After LPS stimulation, correlation at matched time points fell apart, but mRNA abundance at 5 hours post-stimulation accounted for nearly 90% of variation of protein expression at 12 hours, indicating a temporal delay between mRNA transcription and protein translation. Even so, there were still many proteins which demonstrated no change in mRNA abundance, but significant change in protein expression, demonstrating control at a translational level[91]. This phenomenon of increased per mRNA translation has been termed "translation on demand"[92] or "just-in-time translation"[93] and was originally observed in the context of GCN4 translation in yeast due to amino acid starvation, but has now been understood to be a common mechanism of gene regulation.

1.5.2 Polysome Profiling

One way to more closely characterize translation is to identify mRNA transcripts that are bound by ribosomes. Because the assembly of the 80S ribosome initiation complex is a prerequisite for translation, identification of the sequences that are bound by ribosomes should give a more accurate assessment of the translatome. Polysome profiling is a method where entire mRNA transcripts are isolated from bound ribosomes

and analyzed either by microarrays or sequencing[94]. This enrichment is accomplished through sedimentation of monosomes and polysomes (mRNA transcripts with one or more bound ribosomes, respectively) through a sucrose gradient in order to separate them from unbound ribosomal subunits. Polysome profiling gives a more accurate assessment of translation than measuring mRNA abundance, but it is still not a perfect picture of translation. One of the primary challenges associated with isolating polysomes is the process of fractionation. Polysomes can reside across multiple fractions depending on the number of bound ribosomes and length of the mRNA transcript[95]. Furthermore, this method does not account for ribosomes that are bound at uORFs (see section 1.2.3) or stalled ribosomes[48]. A higher resolution method is required to more accurately determine the location of bound ribosomes within an mRNA transcript.

1.5.3 Ribosome Footprinting

Ribosome footprinting is similar to polysome profiling in that monosomes and polysomes are isolated from the total mRNA pool. The key difference in ribosome footprinting is the addition of a ribonuclease digestion step. The ribonuclease degrades all present RNA except for the portions shielded by bound ribosomes[96]. This process yields a library of 30 nucleotide-long fragments that can be sequenced to determine exactly where on the transcript ribosomes were bound[97]. This process allows for the differentiation between ribosomes bound to uORFs which are not translating the main coding sequence, and those that are translating the target gene. Ribosome footprinting has been successfully used to show that mTOR regulates translation of pro-tumorigenic transcripts[47] and that differentiation of mouse embryonic stem cells results in increased

translation of membrane proteins[98]. However, mapping short footprint sequences back to the genome is challenging and requires significant computational horsepower. Furthermore, ribosome footprinting does not give context regarding rate of translation, as the results could be interpreted as many transcripts being translated by few ribosomes, or few transcripts being translated by a large number of ribosomes. In addition, ribosome profiling would interpret ribosomes stalled in the middle of an ORF as actively translating, when in fact there may be some negative regulation resulting in stalling. A recent study comparing ribosome footprints to synthesized proteins via pulsed SILAC (see section 1.5.5) over relatively long time scales (6 hours) demonstrated that ribosome footprints do not fully represent the collection of proteins synthesized over that time[99].

1.5.4 Protein Abundance

All the methods described above depend on the sequencing of mRNA or mRNA fragments as a proxy for the production of new proteins. In contrast to RNA sequencing-based methods, there are a family of proteomic-based methods that observe changes in protein synthesis. The simplest and most straightforward method is the measurement of relative protein abundances between conditions. Prior to the 1980's, proteins were studied primarily with antibody-based immunoassays such as the enzyme-linked immunosorbent assay (ELISA) and Western Blot. However, these methods were limited to studying proteins for which antibodies were available. The invention of mass spectrometry revolutionized the field of proteomics allowing for high throughput, discovery mode analysis of the proteome.

Shotgun proteomics is the most commonly used method for measuring relative protein abundances on a large scale between multiple conditions. In this technique, proteins are digested into peptides with a protease and then analyzed via MS analysis. There are several problems with attempting to measure changes in protein synthesis based on protein abundance measurements. First and foremost, detecting changes in protein synthesis largely depends on the size of the existing pool of a given protein. Changes in protein synthesis rates may be difficult to detect if the protein exists in large amounts. Furthermore, the amount of protein in a cell is the summation of protein synthesis and degradation. Increases in total protein amount could be due to an increase in synthesis rate, a decrease in degradation rate, or a combination of the two. A corollary is that if a change in synthesis is matched by a corresponding change in degradation, this observation would be lost by looking at total protein abundance alone.

1.5.5 Pulsed Stable Isotope Labeling of Amino Acids in Cell Culture (pSILAC)

Current methods used to measure protein synthesis rely on metabolic-labeling schemes. Labelled amino acids added to growth media are incorporated into proteins and used as a readout of proteins synthesized while the amino acid is present. In SILAC-based methods, the labels are naturally occurring amino acids containing heavy isotopes of constituent atoms[70].

Pulsed SILAC (pSILAC) is a variation of SILAC that enables the detection of newly synthesized proteins in the background of the existing proteome. Instead of culturing cells in SILAC amino acids to steady state, the SILAC amino acids are pulsed in for a relatively short duration to label proteins that are synthesized *de novo* during the pulse[100].

This process allows for an estimation of protein turnover by comparing the abundance of the heavy, newly-made peptides to the light, preexisting peptides. By incorporating a third channel, the relative translation rate between two conditions can be compared[101]. Alternatively, the pulse can be followed by a chase with a different isotope label to get an estimate of degradation rates in addition to synthesis rates[102].

These pSILAC methods are useful for gaining estimates of relative protein synthesis, degradation, and turnover rates on longer time scales (hours to days). However, they suffer from the same shortcomings as shotgun proteomics: to analyze the newly synthesized fraction of the proteome, one must also dig through the (usually much larger) pool of existing proteins. This limitation restricts the sensitivity of new proteome analysis for two reasons. First, because some proteins may be synthesized at a slow rate compared to the size of its existing pool, the newly-synthesized fraction may not be detected due to dynamic range limitations of the mass spectrometer. Second, because the MS1 scan contains a combination of newly synthesized proteins and existing proteins, the complexity of the sample is inherently increased, and it's more likely that less abundant proteins will not be selected for MS/MS analysis. In addition to limited sensitivity, measuring relative synthesis rates through pSILAC limits analysis to no more than 3 channels (light, medium and heavy), precluding higher level multiplexing analyses.

1.5.6 Puromycin-Associated Nascent Chain Proteomics (PUNCH-P)

A relatively new technique for mapping protein synthesis, termed PUNCH-P, examines nascent polypeptide chains as they are actively translated by the ribosome[103]. This analysis is accomplished by isolating actively translating ribosomes via

ultracentrifugation, and labeling with a biotinylated variant of the antibiotic puromycin. Puromycin is incorporated into the C-terminus of actively translated proteins, triggering termination of translation and release of the truncated polypeptide from the ribosome. Because this puromycin treatment results in the termination of translation, only one biotin handle is incorporated per peptide. These peptides can be enriched using a streptavidin-conjugated resin and directly analyzed via LC-MS/MS. This method has been used successfully to measure changes in protein synthesis throughout mitosis and cell cycle checkpoints (30-90 minutes), a time scale inaccessible to pSILAC due to the need for high temporal resolution[104].

PUNCH-P is not without its limitations; the primary limitation is the amount of material required for the experiment. The amount of protein associated with actively-translating ribosomes is very small compared to the existing proteome, so in order to get adequate coverage of the nascent proteome, a large amount of starting material (4e7 HeLa cells) is required per sample[104]. Another challenge associated with PUNCH-P is the mapping of truncated peptides to unique proteins.

1.5.7 Bioorthogonal Non-Canonical Amino Acid Tagging (BONCAT)

There are many variations of the BONCAT method since its inception in 2006, but at its core is the labeling of newly synthesized proteins with an amino acid analog that contains a bioorthogonal chemical group for enrichment[105]. The first and most subsequent BONCAT studies used azidohomoalanine (Aha), a methionine analog that contains an azide functional group[106], but other studies have used homopropargylglycine (Hpg)[107] which contains a linear alkyne moiety. Both of these methionine analogs contain

functional groups that can participate in copper-catalyzed azide-alkyne cycloadditon (CuAAC)[108], a relatively bio-orthogonal, high-yield "click" reaction for easy enrichment and detection of proteins that incorporate these analogs. Furthermore, Aha is used by the native methionyl tRNA synthetase (MetRS) without any need for genetic manipulation, albeit nearly 400-fold less efficiently than methionine[109]. This incorporation of an azide into newly synthesized proteins and use of the azide for enrichment and/or detection is the crux of all subsequent derivatives and improvements to the BONCAT method. By adding a chemical handle, the BONCAT approach enables the enrichment of newly translated proteins from the background of the existing proteome, simplifying analysis and increasing sensitivity. One of the major challenges associated with analyzing newly-synthesized proteins is how small the newly-synthesized fraction of the proteome is compared to the existing proteome. Pulsed SILAC requires long labeling times in order to get adequate proteomic coverage, preventing analysis of the translatome with high temporal resolution. PUNCH-P allows for measurements of the translatome with high temporal accuracy, but requires large amounts of input material, suffers from poor proteome coverage, and requires a technically challenging protocol. The BONCAT method allows for measurements of newly-translated proteins with high temporal accuracy and good proteome coverage. Furthermore, the amount of starting material required for BONCAT is generally less than that needed for PUNCH-P because BONCAT is measuring the integration of all synthesized proteins over the duration of the labeling period. Finally, by directly measuring newly translated proteins, the use of BONCAT allows for the integration of all the regulation steps that may be unaccounted for in nucleic-acid based assays such as RNA-Seq and ribosome footprinting.

1.6 Variations of BONCAT

Ever since it has been demonstrated that newly synthesized proteins can be labeling with non-canonical amino acids[109], there have been various modifications of the BONCAT method that serve to improve enrichment of newly translated proteins and allow for quantitative analysis of the translatome.

1.6.1 Enrichment versus Detection

The first study reporting the incorporation of Aha into a newly translated protein measured Aha incorporation by the addition of a FLAG epitope through Staudinger Ligation and detection via Western Blot with an anti-FLAG antibody[109]. Several years later, the first study to coin the term "BONCAT" used the biorthogonal CuAAC reaction to label all newly translated proteins in a cell with a trypsin-cleavable biotin handle[105]. The biotin handle was then used for subsequent detection of newly translated proteins by Western Blot and probing with an anti-biotin antibody. More importantly, the biotin handle was also used for enrichment of newly translated proteins with a streptavidin-conjugated resin and analysis via LC-MS/MS following an on-bead trypsin digest.

Most subsequent studies used the azide functional group as a handle for enrichment, but there are several methods that did not have an affinity-based enrichment step. One study separated an Aha-labeled *E. coli* proteome into 48 fractions via reverse phase HPLC, and then treated each fraction with the reducing agent TCEP. TCEP reduces the azide group of Aha to a primary amine, resulting in a shift in elution time during HPLC fractionation. The fractions were then separated again on an HPLC, with

peptides that eluted outside the original window presumed to be labeled with Aha and collected for MS analysis[110].

Finally, some studies did not aim to determine the identification of newly translated proteins, but rather determined the location of newly translated proteins within a cell or tissue. Fluorescent noncanonical amino acid tagging (FUNCAT) achieved this by labeling proteins that contain azides or alkynes (via Aha or Hpg, respectively) with a fluorophore for direct observation via fluorescent microscopy[111].

1.6.2 Click Chemistry

Of the approaches outlined above, the vast majority of BONCAT-based experiments use the azide group of Aha as a handle for chemical enrichment. Prior to the development of CuAAC, the primary method of ligating moieties to an azide was through Staudinger ligation: the formation of a covalent linkage between a phosphine and an azide. This reaction is bioorthogonal and nontoxic, allowing for the tagging of cell-surface proteins that have been labeled with an azide[112]. However, phosphines are readily reduced under ambient conditions and are difficult to solubilize in aqueous environments, making their applications to biological systems limited. Furthermore, the reaction between the phosphine and azide occurs with slower kinetics than more recently developed techniques[113].

The reaction most commonly used for modification and enrichment of Aha azides is CuAAC, which reacts the azide with a linear alkyne to produce a stable covalent triazole linkage. As its name suggests, this reaction requires a reduced copper(I) catalyst in order to proceed under ambient conditions. The copper catalyst is cytotoxic to cells[114],

precluding it from being used in live-cell applications such as FUNCAT. Furthermore, oxidative intermediates generated from the copper catalyst can react with biomolecules[115], altering their chemical makeup and confounding mass spectrometry analyses. In addition, CuAAC has been shown to have off-target effects, resulting in nonspecific labeling of biomolecules[116].

More recently, strain promoted azide-alkyne cycloaddition (SPAAC) has begun to replace CuAAC as the favored method of azide-alkyne click chemistry. Work out of Carolyn Bertozzi's lab in 2004 demonstrated that a strained cyclical alkyne can form a stable triazole linkage at ambient conditions without the need for additional catalysts[117]. The inherent ring strain of roughly 18 kcal/mol in unsubstituted cyclooctyne[118] provides enough energy to drive the reaction to completion. However, while initial efforts with unsubstituted cyclooctyne resulted in successful labeling of azide-modified glycoproteins, the kinetics of the reaction were still impractical for most biological applications. Various substituents have been added to the cyclooctyne ring in an effort to increase the kinetics of cycloaddition with azides. A cyclooctyne derivative with one of fastest reaction kinetics is dibenzocyclooctyne (DBCO), with reaction kinetics on par with that of CuAAC reactions[119]. Furthermore, an effort to optimize the reaction solvent has revealed that the reaction proceeds most efficiently in a mixture of organic and aqueous solvents[120], although solubilizing proteins in semi-organic solvents may prove challenging because proteins have evolved to have hydrophilic exteriors to interface with the surrounding aqueous environment.

One of the major downsides of SPAAC is that cyclooctyne groups are very large compared to their linear alkyne counterparts. Because of their large size, they cannot be

introduced into biomolecule building blocks in the same manner as linear alkynes. For this reason, cycolooctynes are limited to being used in probes that can react with azides that have been incorporated into biomolecules. Furthermore, DBCO has also been shown to react with sulfhydryls (e.g. reduced cysteines) in addition to azides, but this can be remedied by alkylating reduced sulfhydryls prior to performing the click reaction[121].

1.6.3 Method of Capture and Release

There are two main strategies when it comes to enriching Aha-labeled proteins: (1) ligation of a biotin moiety to the protein and affinity purification with streptavidin-conjugated resins[105,122–125], and (2) direct click reaction with an alkyne-functionalized resin using both CuAAC[126–129] and SPAAC[130,131]. Both have been used extensively with varying degrees of success.

One of the challenges associated with biotin ligation followed by streptavidin affinity purification is the competition of biotin-tagged proteins with free biotin-alkyne remaining in solution. The leftover linker necessitates a further cleanup step, either by protein precipitation (if whole proteins are tagged), or by C18 desalting columns (if peptides are tagged), leading to increased sample losses. Furthermore, if a DBCO-biotin tag is used, the hydrophobic DBCO group will bind to C18 desalting columns and co-elute with peptides, making cleanup challenging.

Another challenge associated with using a biotin tag is the method of elution. The biotin-streptavidin interaction has a dissociation constant on the order of 10^{-15} M, making it one of the strongest non-covalent interactions known in biology[132]. While such a high affinity is excellent for capture of biotinylated targets, disrupting this interaction requires

extremely harsh conditions that are not compatible with MS analysis. On-bead digestion can be used to liberate peptides, but the streptavidin protein will also be digested, leading to contaminating peptides that can confound MS analysis. It's been previously demonstrated that exclusion lists can be used to omit selection of streptavidin peptides for MS/MS fragmentation[133], but exclusion occurs at the expense of dynamic range and isolation of other peptides that may fall within that window.

The alternative to biotin tagging and streptavidin pulldown is a direct pulldown with an alkyne-functionalized resin. This direct method has the advantage of being done in a single step without the need for cleanup of excess alkyne-biotin, but carries its own challenges. Because the alkyne is anchored to a solid-phase support, reaction kinetics may suffer compared to free alkyne in solution. This behavior may be especially evident with DBCO, whose hydrophobic properties may prevent extension of the alkyne into solution, resulting in deceased reaction efficiency. Because this approach results in a covalent linkage of the target protein to a solid-phase support, elution requires on-bead digestion to liberate peptides for MS analysis. One of the downsides of this approach is that the modification identifying the protein as newly-translated (Aha) will remain conjugated to the bead and cannot be analyzed via MS.

1.6.4 Quantitation

The initial BONCAT paper was a proof of principle demonstrating that it was possible to capture newly synthesized proteins, but did not quantitatively compare translation rates between multiple conditions. In this study, the incorporation of SILAC

amino acids was used as a marker for specificity, but not to quantify newly translated proteins[105].

To date, all studies seeking to quantify newly translated proteins use variations of pSILAC to as a quantitative metric. The most common approach, a hybrid of BONCAT and pSILAC referred to as either QuanCAT or BONLAC, involves pulsing medium or heavy isotopes of arginine and/or lysine simultaneously with Aha[128,134]. Translation between two samples can be directly compared by mixing samples pulsed with medium and heavy amino acids and comparing the ratio of their intensities in the MS1. The drawback to this approach is that in order to get quantitation, both a SILAC amino acid and an Aha need to be incorporated into the same protein. The efficiency of this process can be increased by starving the cells of methionine, arginine, and lysine prior to pulsing, but this can cause artifacts that can not be accounted for without proper controls. Since the number of channels in pSILAC is limited to two (medium and heavy; the light channel represents the existing proteome), controls would have to be run separately, increasing the technical complexity of the experiments.

A more recently developed approach uses isotopically-labeled Aha in a process termed heavy isotope labeled azidohomoalanine quantification (HILAQ)[135]. By combining the enrichment handle and isotope-encoding into the same residue, HILAQ removes the need for multiple amino acids to be incorporated into the same protein, thereby increasing sensitivity. The drawback to this approach is that analysis is limited to peptides that contain a methionine (substituted with an Aha), limiting coverage compared to QuanCAT which allows the analysis of any tryptic peptide. Furthermore, this approach requires the

use of a biotin pulldown and elution, which suffers from the challenges described in the previous section.

One method of quantitation conspicuously missing is the use of isobaric mass tag labeling (TMT or iTRAQ, see section 1.3.2). This approach allows for higher level multiplexing than SILAC-based approaches, which is limited to two or three different channels. In this thesis, we introduce multiplexed BONCAT using TMT for robust quantitative analysis of protein synthesis across a variety of contexts and conditions.

1.6.5 Applications

Aha labeling has been performed in a number of different biological systems. The original BONCAT manuscript used mammalian cell culture (HEK293 cells)[105]. Shortly afterwards, Aha was used to label the nascent proteomes of *E. coli*[130] and yeast[136]. BONCAT has also been used to measure local translation in the axons and dendrites of dopaminergic neurons[123]. Moving from cell cultures to entire tissues, BONCAT has recently been used to analyze protein synthesis in brain slices[128]. In live animals, Aha has been used to label new proteins in the retinas of mice through intraocular injection[137]. Most recently, Aha has been used to label the nascent proteome of an entire animal through a process termed pulsed azidohomoalanine labeling in mammals (PALM). Aha was introduced via the mouse's chow, and incorporation was detected as early as two days after introduction into the diet[138]. This technique allows for the analysis of tissue specific of protein synthesis from various organs of a single mouse, albeit with poor temporal resolution compared to *in vitro* approaches.

1.7 Summary

In this thesis, the BONCAT protocol is optimized to increase sensitivity for labeling periods as short as 15 minutes in cell culture. The protocol was also adapted for highly multiplexed quantitation with isobaric mass tags, allowing for the direct comparison of protein synthesis across multiple conditions. Using this approach, changes in protein synthesis were profiled during the unfolded protein response and following EGF stimulation. Protein synthesis was also compared to RNA abundance and ribosome footprinting, the two most commonly used methods to profile changes in gene expression. Finally, BONCAT was used to compare changes in protein synthesis following MEK inhibition between a binimetinib sensitive and resistant NRAS mutant melanoma cell line in an effort to identify potential resistance mechanisms.

1.8 References

1. F. H. Crick. On protein synthesis. *Symp. Soc. Exp. Biol.* **12,** 138–163 (1958).
2. JACOB, F. & MONOD, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3,** 318–56 (1961).
3. Brenner, S., Jacob, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190,** 576–581 (1961).
4. Mitchell, P. J. & Tjian, R. Transcriptional Regulation in Mamnalian Cells by Sequence-Specific DNA Binding Proteins. *Science (80-.).* **245,** 489–385 (1989).
5. Mirkovitch, J. & Darnell, J. E. Mapping of RNA polymerase on mammalian genes in cells and nuclei. *Mol. Biol. Cell* **3,** 1085–94 (1992).
6. Rosonina, E., Kaneko, S. & Manley, J. L. Terminating the transcript: breaking up is hard to do. *Genes Dev.* **20,** 1050–6 (2006).
7. Shatkin, A. J. Capping of eucaryotic mRNAs. *Cell* **9,** 645–653 (1976).
8. Cho, E. J., Takagi, T., Moore, C. R. & Buratowski, S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **11,** 3319–3326 (1997).
9. Visa, N., Izaurralde, E., Ferreira, J., Daneholt, B. & Mattaj, I. W. A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. *J. Cell Biol.*

133, 5–14 (1996).
10. Konarska, M. M., Padgett, R. A. & Sharp, P. A. Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell* **38,** 731–736 (1984).
11. Sonenberg, N. & Gingras, A. C. The mRNA 5' cap-binding protein eIF4E and control of cell growth. *Current Opinion in Cell Biology* **10,** 268–275 (1998).
12. Furuichi, Y., LaFiandra, A. & Shatkin, A. J. 5'-Terminal structure and mRNA stability. *Nature* **266,** 235–239 (1977).
13. Murthy, K. G. K., Park, P. & Manley, J. L. A nuclear micrococcal-sensitive, ATP-dependent exoribonuclease degrades uncapped but not capped RNA substratesx. *Nucleic Acids Res.* **19,** 2685–2692 (1991).
14. Marcotrigiano, J., Gingras, a C., Sonenberg, N. & Burley, S. K. Cocrystal Structure of the Messenger RNA 5` Cap-Binding Protein (eIF4E) Bound to 7-methyl-GDP. *Cell* **89,** 951–961 (1997).
15. Fitzgerald, M. & Shenk, T. The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* **24,** 251–260 (1981).
16. Bienroth, S., Keller, W. & Wahle, E. Assembly of a processive messenger RNA polyadenylation complex. *EMBO J.* **12,** 585–594 (1993).
17. Gingras, A. C., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu Rev Biochem* **68,** 913–963 (1999).
18. Kahvejian, A. Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev.* **19,** 104–113 (2005).
19. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12,** 1–8 (1977).
20. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72,** 291–336 (2003).
21. Pain, V. M. Initiation of protein synthesis in eukaryotic cells. *Eur. J. Biochem.* **236,** 747–771 (1996).
22. Sonenberg, N., Rupprecht, K. M., Hecht, S. M. & Shatkin, A. J. Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 4345–4349 (1979).
23. Rogers, G. W., Komar, A. A. & Merrick, W. C. eIF4A: the godfather of the DEAD box helicases. *Prog. Nucleic Acid Res. Mol. Biol.* **72,** 307–331 (2002).
24. Hentze, M. W. eIF4G: a multipurpose ribosome adapter. *Science (80-.).* **275,** 500–501 (1997).
25. Tarun, S. Z., Wells, S. E., Deardorff, J. A. & Sachs, A. B. Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 9046–51 (1997).
26. Kozak, M. The scanning model for translation: An update. *Journal of Cell Biology* **108,** 229–241 (1989).
27. Das, S., Ghosh, R. & Maitra, U. Eukaryotic Translation Initiation Factor 5 Functions as a GTPase-activating Protein. *J. Biol. Chem.* **276,** 6720–6726 (2001).
28. Pestova, T. V. *et al.* The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* **403,** 332–335 (2000).

29. Lee, A. S., Kranzusch, P. J., Doudna, J. A. & Cate, J. H. D. eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature* **536,** 96–9 (2016).
30. Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334,** 320–325 (1988).
31. Holcik, M. & Sonenberg, N. Translational control in stress and apoptosis. *Nat. Rev. Mol. Cell Biol.* **6,** 318–27 (2005).
32. Mokrejš, M. *et al.* IRESite: the database of experimentally verified IRES structures (www.iresite.org). *Nucleic Acids Res.* **34,** D125–D130 (2006).
33. Le, S. Y. & Maizel, J. V. A common RNA structural motif involved in the internal initiation of translation of cellular mRNAs. *Nucleic Acids Res.* **25,** 362–369 (1997).
34. Macejak, D. G. & Sarnow, P. Internal initiation of translation mediated by the 5′ leader of a cellular mRNA. *Nature* **353,** 90–94 (1991).
35. Akiri, G. *et al.* Regulation of vascular endothelial growth factor (VEGF) expression is mediated by internal initiation of translation and alternative initiation of transcription. *Oncogene* **17,** 227–236 (1998).
36. Nanbru, C. *et al.* Alternative translation of the proto-oncogene c-myc by an internal ribosome entry site. *J. Biol. Chem.* **272,** 32061–32066 (1997).
37. Vagner, S. *et al.* Alternative translation of human fibroblast growth factor 2 mRNA occurs by internal entry of ribosomes. *Mol. Cell. Biol.* **15,** 35–44 (1995).
38. Byrd, M. P., Zamora, M. & Lloyd, R. E. Translation of Eukaryotic Translation Initiation Factor 4GI (eIF4GI) Proceeds from Multiple mRNAs Containing a Novel Cap-dependent Internal Ribosome Entry Site (IRES) That Is Active during Poliovirus Infection. *J. Biol. Chem.* **280,** 18610–18622 (2005).
39. Farrell, P., Balkow, K., Hunt, T., Jackson, R. J. & Trachsel, H. Phosphorylation of initiation factor eIF-2 and the control of reticulocyte protein synthesis. *Cell* **11,** 187–200 (1977).
40. ROWLANDS, A. G., MONTINE, K. S., HENSHAW, E. C. & PANNIERS, R. Physiological stresses inhibit guanine-nucleotide-exchange factor in Ehrlich cells. *Eur. J. Biochem.* **175,** 93–99 (1988).
41. Chen, J. J. & London, I. M. Regulation of protein synthesis by heme-regulated eIF-2 alpha kinase. *Trends Biochem. Sci.* **20,** 105–108 (1995).
42. Feng, G. S., Chong, K., Kumar, A. & Williams, B. R. Identification of double-stranded RNA-binding domains in the interferon-induced double-stranded RNA-activated p68 kinase. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 5447–5451 (1992).
43. Berlanga, J. J., Santoyo, J. & De Haro, C. Characterization of a mammalian homolog of the GCN2 eukaryotic initiation factor 2alpha kinase. *Eur. J. Biochem.* **265,** 754–762 (1999).
44. Harding, H. P., Zhang, Y. & Ron, D. Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature* **397,** 271–274 (1999).
45. Pause, A. *et al.* Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function. *Nature* **371,** 762–767 (1994).
46. Gingras, A.-C., Raught, B. & Sonenberg, N. Regulation of translation initiation by FRAP / mTOR. *Genes Dev.* **15,** 807–826 (2001).
47. Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer

initiation and metastasis. *Nature* **485,** 55–61 (2012).
48. MEIJER, H. A. & THOMAS, A. A. M. Control of eukaryotic protein synthesis by upstream open reading frames in the 5′-untranslated region of an mRNA. *Biochem. J.* **367,** 1–11 (2002).
49. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15,** 8125–8148 (1987).
50. Morris, D. R. & Geballe, A. P. Upstream Open Reading Frames as Regulators of mRNA Translation. *Mol. Cell. Biol.* **20,** 8635–8642 (2000).
51. Kozak, M. Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. *Mol. Cell. Biol.* **7,** 3438–45 (1987).
52. Luukkonen, B. G., Tan, W. & Schwartz, S. Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. *J. Virol.* **69,** 4086–94 (1995).
53. Cao, J. & Geballe, A. P. Translational inhibition by a human cytomegalovirus upstream open reading frame despite inefficient utilization of its AUG codon. *J. Virol.* **69,** 1030–6 (1995).
54. Konan, K. V & Yanofsky, C. Role of ribosome release in regulation of tna operon expression in Escherichia coli. *J. Bacteriol.* **181,** 1530–6 (1999).
55. Arribere, J. A. & Gilbert, W. V. Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res.* **23,** 977–987 (2013).
56. Hershberg, R. & Petrov, D. A. Selection on Codon Bias. *Annu. Rev. Genet.* **42,** 287–299 (2008).
57. Chionh, Y. H. *et al.* tRNA-mediated codon-biased translation in mycobacterial hypoxic persistence. *Nat. Commun.* **7,** 13302 (2016).
58. Dawson, J. H. J. & Guilhaus, M. Orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **3,** 155–159 (1989).
59. Mamyrin, B. A., Karataev, V. I., Shmikk, D. V & Zagulin, V. A. The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Sov. Phys. - JETP* **37,** 45–48 (1973).
60. Richards, J. A. Stability Diagram Approximation for the Lossy Mathieu Equation. *SIAM J. Appl. Math.* **30,** 240–247 (1976).
61. Kero, F. A., Pedder, R. E. & Yost, R. A. in *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics* (John Wiley & Sons, Ltd, 2005). doi:10.1002/047001153X.g301319
62. Marshall, A. G., Hendrickson, C. L. & Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **17,** 1–35 (1998).
63. Makarov, A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* **72,** 1156–1162 (2000).
64. Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **8,** 61–80 (2015).
65. Malmström, J. *et al.* Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **460,** 762–765 (2009).
66. Sadygov, R. G., Liu, H. & Yates, J. R. A model for random sampling and

estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76,** 4193–201 (2004).
67. Curran, T. G., Zhang, Y., Ma, D. J., Sarkaria, J. N. & White, F. M. MARQUIS: A multiplex method for absolute quantification of peptides and posttranslational modifications. *Nat. Commun.* **6,** 5924 (2015).
68. Stemmann, O., Zou, H., Gerber, S. A., Gygi, S. P. & Kirschner, M. W. Dual inhibition of sister chromatid separation at metaphase. *Cell* **107,** 715–726 (2001).
69. Bondarenko, P., Chelius, D. & Shaler, T. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography− tandem mass. *Anal. Chem.* (2002). at <http://pubs.acs.org/doi/abs/10.1021/ac0256991>
70. Zhu, H., Pan, S., Gu, S., Morton Bradbury, E. & Chen, X. Amino acid residue specific stable isotope labeling for quantitative proteomics. *Rapid Commun. Mass Spectrom.* **16,** 2115–2123 (2002).
71. Thompson, A. *et al.* Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75,** 1895–1904 (2003).
72. Ross, P. L. Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **3,** 1154–1169 (2004).
73. Uniprot Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45,** D158–D169 (2017).
74. Brocchieri, L. & Karlin, S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* **33,** 3390–3400 (2005).
75. Lathe, G. H. & Ruthven, C. R. The separation of substances and estimation of their relative molecular sizes by the use of colums of starch in water. *Biochem. J.* **62,** 665–674 (1956).
76. Inamuddin & Luqman, M. Ion-exchange technology II. (2012). at <http://public.eblib.com/choice/publicfullrecord.aspx?p=971755>
77. Molnar, I. & Horvath, C. Reverse phase chromatography of polar biological substances: separation of catechol compounds by high performance liquid chromatography. *Clin. Chem.* **22,** 1497–1502 (1976).
78. Ficarro, S. B. *et al.* Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat. Biotechnol.* **20,** 301–305 (2002).
79. Seneviratne, U. *et al.* S-nitrosation of proteins relevant to Alzheimer's disease during early stages of neurodegeneration. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 4152–7 (2016).
80. Kumar, A., Baycin-Hizal, D., Shiloach, J., Bowen, M. A. & Betenbaugh, M. J. Coupling enrichment methods with proteomics for understanding and treating disease. *PROTEOMICS - Clin. Appl.* **9,** 33–47 (2015).
81. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. **58,** 267–288 (1996).
82. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: A basic tool of chemometrics. in *Chemometrics and Intelligent Laboratory Systems* **58,** 109–130 (2001).
83. Alwine, J. C., Kemp, D. J. & Stark, G. R. Method for detection of specific RNAs in

agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5350–4 (1977).
84. Mullis, K. *et al.* Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51 Pt 1,** 263–73 (1986).
85. Temin, H. M. & Mizutani, S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226,** 1211–1213 (1970).
86. Becker-Andre ', M. & Hahlbrock, K. Absolute raRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided n-anscript ritratkm assay (PATTY). *Nucleic Acids Res.* **17,** (1989).
87. Schena, M., Shalon, D., Davis, R. W. R. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science (80-.).* **270,** 467 (1995).
88. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5,** 621–628 (2008).
89. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
90. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).
91. Jovanovic, M. *et al.* Dynamic profiling of the protein life cycle in response to pathogens. *Science (80-.).* **347,** 1259038–1259038 (2015).
92. Beyer, A., Hollunder, J., Nasheuer, H.-P. & Wilhelm, T. Post-transcriptional expression regulation in the yeast Saccharomyces cerevisiae on a genomic scale. *Mol. Cell. Proteomics* **3,** 1083–92 (2004).
93. Brockmann, R., Beyer, A., Heinisch, J. J., Wilhelm, T. & Sberro, H. Posttranscriptional Expression Regulation: What Determines Translation Rates? *PLoS Comput. Biol.* **3,** e57 (2007).
94. Arava, Y. *et al.* Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 3889–3894 (2003).
95. Ingolia, N. T. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.* **15,** 205–213 (2014).
96. Steitz, J. A. Polypeptide Chain Initiation: Nucleotide Sequences of the Threee Ribosomal Binding Sites in Bacteriophage R17 RNA. *Nature* **224,** 957–964 (1969).
97. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science (80-.).* **324,** 218–223 (2009).
98. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147,** 789–802 (2011).
99. Liu, T. Y. *et al.* Time-Resolved Proteomics Extends Ribosome Profiling-Based Measurements of Protein Synthesis Dynamics. *Cell Syst.* **4,** 636–644.e9 (2017).
100. Doherty, M. K., Whitehead, C., McCormack, H., Gaskell, S. J. & Beynon, R. J. Proteome dynamics in complex organisms: Using stable isotopes to monitor individual protein turnover rates. *Proteomics* **5,** 522–533 (2005).

101. Schwanhäusser, B., Gossen, M., Dittmar, G. & Selbach, M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9,** 205–209 (2009).
102. Fierro-Monti, I. *et al.* A novel pulse-chase SILAC strategy measures changes in protein decay and synthesis rates induced by perturbation of proteostasis with an Hsp90 inhibitor. *PLoS One* **8,** e80423 (2013).
103. Aviner, R., Geiger, T. & Elroy-Stein, O. Genome-wide identification and quantification of protein synthesis in cultured cells and whole tissues by puromycin-associated nascent chain proteomics (PUNCH-P). *Nat. Protoc.* **9,** 751–760 (2014).
104. Aviner, R., Geiger, T. & Elroy-Stein, O. Novel proteomic approach (PUNCH-P) reveals cell cycle-specific fluctuations in mRNA translation. *Genes Dev.* **27,** 1834–1844 (2013).
105. Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9482–7 (2006).
106. Dieterich, D. C. *et al.* Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat. Protoc.* **2,** 532–40 (2007).
107. Wiltschi, B., Wenger, W., Nehring, S. & Budisa, N. Expanding the genetic code of Saccharomyces cerevisiae with methionine analogues. *Yeast* **25,** 775–786 (2008).
108. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67,** 3057–64 (2002).
109. Kiick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci.* **99,** 19–24 (2002).
110. Kramer, G. *et al.* Identification and quantitation of newly synthesized proteins in Escherichia coli by enrichment of azidohomoalanine-labeled peptides with diagonal chromatography. *Mol. Cell. Proteomics* **8,** 1599–1611 (2009).
111. Dieterich, D. C. *et al.* In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat. Neurosci.* **13,** 897–905 (2010).
112. Saxon, E. Cell Surface Engineering by a Modified Staudinger Reaction. *Science (80-.).* **287,** 2007–2010 (2000).
113. Van Berkel, S. S., Van Eldijk, M. B. & Van Hest, J. C. M. Staudinger ligation as a method for bioconjugation. *Angewandte Chemie - International Edition* **50,** 8806–8827 (2011).
114. Seo, T. S., Li, Z., Ruparel, H. & Ju, J. Click chemistry to construct fluorescent oligonucleotides for DNA sequencing. *J. Org. Chem.* **68,** 609–612 (2003).
115. Hong, V., Presolski, S. I., Ma, C. & Finn, M. â. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew. Chemie - Int. Ed.* **48,** 9879–9883 (2009).
116. Li, L. & Zhang, Z. Development and applications of the copper-catalyzed azide-alkyne cycloaddition (CuAAC) as a bioorthogonal reaction. *Molecules* **21,** 1393 (2016).

117. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126,** 15046–15047 (2004).
118. Turner, R. B., Jarrett, A. D., Goebel, P. & Mallon, B. J. Heats of hydrogenation. IX. Cyclic acetylenes and some miscellaneous olefins. *J. Am. Chem. Soc.* **95,** 790–792 (1973).
119. Debets, M. F. *et al.* Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloadditionw. (2009). doi:10.1039/b917797c
120. Davis, D. L. *et al.* Effect of Buffer Conditions and Organic Cosolvents on the Rate of Strain-Promoted Azide–Alkyne Cycloaddition. *J. Org. Chem.* acs.joc.6b01112 (2016). doi:10.1021/acs.joc.6b01112
121. van Geel, R., Pruijn, G. J. M., van Delft, F. L. & Boelens, W. C. Preventing Thiol-Yne Addition Improves the Specificity of Strain-Promoted Azide–Alkyne Cycloaddition. *Bioconjug. Chem.* **23,** 392–398 (2012).
122. Choi, K.-Y. (Grace), Lippert, D. N. D., Ezzatti, P. & Mookherjee, N. Defining TNF-α and IL-1β induced nascent proteins: Combining bio-orthogonal non-canonical amino acid tagging and proteomics. *J. Immunol. Methods* **382,** 189–195 (2012).
123. Hodas, J. J. L. *et al.* Dopaminergic modulation of the hippocampal neuropil proteome identified by bioorthogonal noncanonical amino acid tagging (BONCAT). *Proteomics* **12,** 2464–2476 (2012).
124. Bagert, J. D. *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* **13,** 1352–8 (2014).
125. Wang, J. *et al.* Quantitative chemical proteomics profiling of de novo protein synthesis during starvation-mediated autophagy. *Autophagy* **12,** 1931–1944 (2016).
126. Eichelbaum, K. & Krijgsveld, J. Rapid temporal dynamics of transcription, protein synthesis, and secretion during macrophage activation. *Mol. Cell. Proteomics* **13,** 792–810 (2014).
127. Zhang, G. *et al.* In-depth quantitative proteomic analysis of de novo protein synthesis induced by brain-derived neurotrophic factor. *J. Proteome Res.* **13,** 5707–5714 (2014).
128. Bowling, H. *et al.* BONLAC: A combinatorial proteomic technique to measure stimulus-induced translational profiles in brain slices. *Neuropharmacology* **100,** 76–89 (2016).
129. Kumar, A. *et al.* Dataset generated using hyperplexing and click chemistry to monitor temporal dynamics of newly synthesized macrophage secretome post infection by mycobacterial strains. *Data in Brief* **9,** 349–354 (2016).
130. Nessen, M. A. *et al.* Selective Enrichment of Azide-Containing Peptides from Complex Mixtures. *J. Proteome Res.* **8,** 3702–3711 (2009).
131. Glenn, W. *et al.* BONCAT enables time-resolved analysis of protein synthesis in native plant tissue. *Plant Physiol.* **19,** pp.01762.2016 (2017).
132. Green, N. M. in 85–133 (1975). doi:10.1016/S0065-3233(08)60411-8
133. Fukuyama, H. *et al.* On-bead tryptic proteolysis: An attractive procedure for LC-MS/MS analysis of the Drosophila caspase 8 protein complex during immune response against bacteria. *J. Proteomics* **75,** 4610–4619 (2012).

134. Howden, A. J. M. *et al.* QuaNCAT: quantitating proteome dynamics in primary cells. *Nat. Methods* **10,** 343–346 (2013).
135. Ma, Y., McClatchy, D. B., Barkallah, S., Wood, W. W. & Yates, J. R. HILAQ: A Novel Strategy for Newly Synthesized Protein Quantification. *J. Proteome Res.* **16,** 2213–2220 (2017).
136. Ma, Y., Biava, H., Contestabile, R., Budisa, N. & Di Salvo, M. L. Coupling bioorthogonal chemistries with artificial metabolism: Intracellular biosynthesis of azidohomoalanine and its incorporation into recombinant proteins. *Molecules* **19,** 1004–1022 (2014).
137. Schiapparelli, L. M. *et al.* Direct detection of biotinylated proteins by mass spectrometry. *J. Proteome Res.* **13,** 3966–3978 (2014).
138. McClatchy, D. B. *et al.* Pulsed azidohomoalanine labeling in mammals (PALM) detects changes in liver-specific LKB1 knockout mice. *J. Proteome Res.* **14,** 4815–4822 (2015).

Chapter II

Adapting the BONCAT Protocol for Robust Multiplexing Analysis

Daniel Rothenberg

2.1 Introduction

A method for accurate and robust quantitative measurements of the protein synthesis rate has thus far eluded life science researchers. RNA sequencing-based approaches seek to use mRNA abundance as a proxy for protein expression, but the correlation between mRNA abundance and protein translation varies widely and may be temporally offset due to regulation at the level of translation. Proteomics methods based on measuring protein abundance or incorporation of labeled amino acids without enrichment suffer from poor sensitivity and low temporal resolution. To date, the most promising approach for the direct measurement of protein synthesis rates is BONCAT (introduced in section 1.5.7).

BONCAT methods use pulses of non-canonical amino acids (usually azidohomoalanine, Aha) to insert bioorthogonal chemical handles into newly translated proteins for enrichment and analysis. Because newly translated proteins are labeled with Aha for the duration of the pulse, the actual value being measured is the integration of all proteins synthesized during that time period. In order to measure translation *rates*, the cumulative amounts must be normalized to the length of the Aha pulse. In the absence of quantitative labeling approaches (discussed in section 1.6.4), the enrichment of the nascent proteome is a purely qualitative exercise. The introduction of multiple SILAC labeling channels – such as in QuanCAT[1], BONLAC[2], and HILAQ[3] – allow for the comparison of translation rates across multiple conditions. However, the nature of SILAC-based approaches limits quantitation to 2 or 3 conditions per analysis. Furthermore, these quantitative studies used pulses no shorter than 1 hour, preventing the analysis of rapid (minutes to hours) translational responses to cellular perturbations.

A wide variety of different approaches can be taken for the labeling, enrichment, and analysis of newly synthesized proteins, many of which were discussed in depth in section 1.6. Here I seek to optimize the enrichment of the nascent proteome for highly multiplexed quantitative analyses with unprecedented temporal resolution.

2.2 Methods

2.2.1 Cell Culture

Unless otherwise noted, all protocol development experiments were performed using lysates from MCF10a cells. Cells were cultured in DMEM:F12 media supplemented with 5% horse serum, 20 ng/mL EGF, 500 µg/mL hydrocortisone, 100 ng/mL cholera toxin, 10 µg/mL insulin, 1% penicillin/streptomycin, and 2 mM glutamine. Cells were passaged every third day at a 1:4 ratio, and all experiments were performed on the third day following passaging.

2.2.2 Metabolic Labeling and Lysis

For most experiments during protocol development, Aha and Met were added directly to the growth media. In these experiments, media volume was reduced to 5 mL, and an appropriate amount of Aha or Met was added to the cells (usually 1 mM-5 mM depending on experiment) for various durations. For cycloheximide (CHX) controls, CHX was added to the cells at 300 µg/mL for 30 minutes prior to the addition of Aha or Met. For experiments using Lys/Arg/Met (KRM)-free media, growth media was removed and replaced with 5 mL of KRM-free media 30 minutes prior to the addition of Met or Aha. For experiments using pSILAC labels, $^{15}N_4{}^{13}C_6$ arginine (R10) and $^{15}N_2{}^{13}C_6$ lysine (K8) were

added to 0.5 mM concurrently with Aha or Met. In experiments with CHX negative controls and KRM-free media, CHX was added concurrently with the media change. Following Aha or Met labeling, the media was aspirated, and the cells were washed in ice cold PBS supplemented with 300 µg/mL CHX. Cells were lysed in differing lysis buffers depending on the experiment, but the most common one was 1% SDS in PBS supplemented with 50 mM N-ethylmaleimide (NEM) and 300 µg/mL CHX.

2.2.3 Western Blots

Protein concentration of lysates were measured via bicinchroninic acid (BCA) assays and normalized to an appropriate concentration (usually 1 mg/mL unless otherwise noted). For measurements of Aha incorporation, DBCO-biotin (or derivatives thereof) was added to the lysates for 1 hour at a concentration of 25 µM unless otherwise noted. In experiments where the reaction was quenched, quenching was performed by adding sodium azide to roughly 100 mM. LDS buffer was added to 1x, and beta-mercaptoethanol was added to 1% except in experiments using a disulfide linker. Samples were loaded onto a NuPAGE Novex 4-12% Bis-Tris Midi Protein Gels (Invitrogen) and run at 160 volts for about an hour. Samples were transferred to a nitrocellulose membrane at 100 volts for 1 hour. Membranes were blocked with LICOR PBS blocking buffer for 30 minutes at room temperature. Biotinylated proteins were probed with IRDye 680RD-conjugated streptavidin (LICOR #32230). The following primary antibodies were used: GAPDH 1:10,000 (Cell Signaling #5174), The following secondary antibodies were used: IRDye 800CW-conjugated goat-anti-rabbit 1:10,000

(LICOR # 926-32211) and IRDye 680LT-conjugated goat-anti-mouse 1:10,000 (LICOR #926-68031). Western blots were imaged on a LICOR Odyssey instrument.

2.2.4 Mass Spectrometry and Peptide Identification

Samples were manually loaded onto a C18 trapping column and placed in-line with an Agilent 1260 Infinity HPLC. When using an autosampler, samples were transferred to autosampler vials and loaded into a Thermo Fisher Easy nLC 1000. Specific gradients depended on the particular experiment, but focused on a transition from 0% B to 42% B (solvent A = 0.2 M acetic acid (AcOH) or 0.1% formic acid (FA); solvent B = 70% acetonitrile (MeCN) in 0.2 M AcOH or 80% MeCN in 0.1% FA). Unless otherwise noted, all samples were analyzed on a ThermoFisher QExactive or QExactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer. Mass spectra were searched using the Mascot search engine v2.4 through ThermoFisher Proteome Discoverer software.

2.3 Protocol Development and Results

In section 1.6, I discussed the many variations of the BONCAT method. Given the large number of tools available for click chemistry and enrichment of newly translated proteins, there are many combinations that can be adopted. Here I discuss some of the tools and approaches tested during the process of optimizing the quantitative BONCAT protocol as well as the limitations associated with these tools. The final protocol used for subsequent experiments can be found in appendix 2.7.

2.3.1 Aha and pSILAC Labeling Optimization

In order to visualize Aha-labeled proteins, a DBCO-biotin linker (Click Chemistry Tools #A116) was used to chemically conjugate a biotin group onto proteins that incorporate Aha. To see if time-dependent changes in Aha incorporation could be visualized, Aha was directly added to the media of cells to 5mM for 15, 30, 60, 90, 120, and 240 minutes. As a negative control, Aha was replaced with Met for all 6 time points. To demonstrate that this incorporation was a translationally controlled mechanism, cells were pretreated for 30 minutes with 300 µg/mL CHX prior to the addition of Aha to the media (CHX+Aha). Cells were lysed in 1% SDS, and total protein was normalized by BCA to 1 mg/mL. DBCO-biotin was added to the lysate to 25 µM for 1 hour. The reaction was quenched with sodium azide, and the samples were analyzed via Western Blot. A time-dependent increase of biotin incorporation in Aha-labeled cells was observed, but

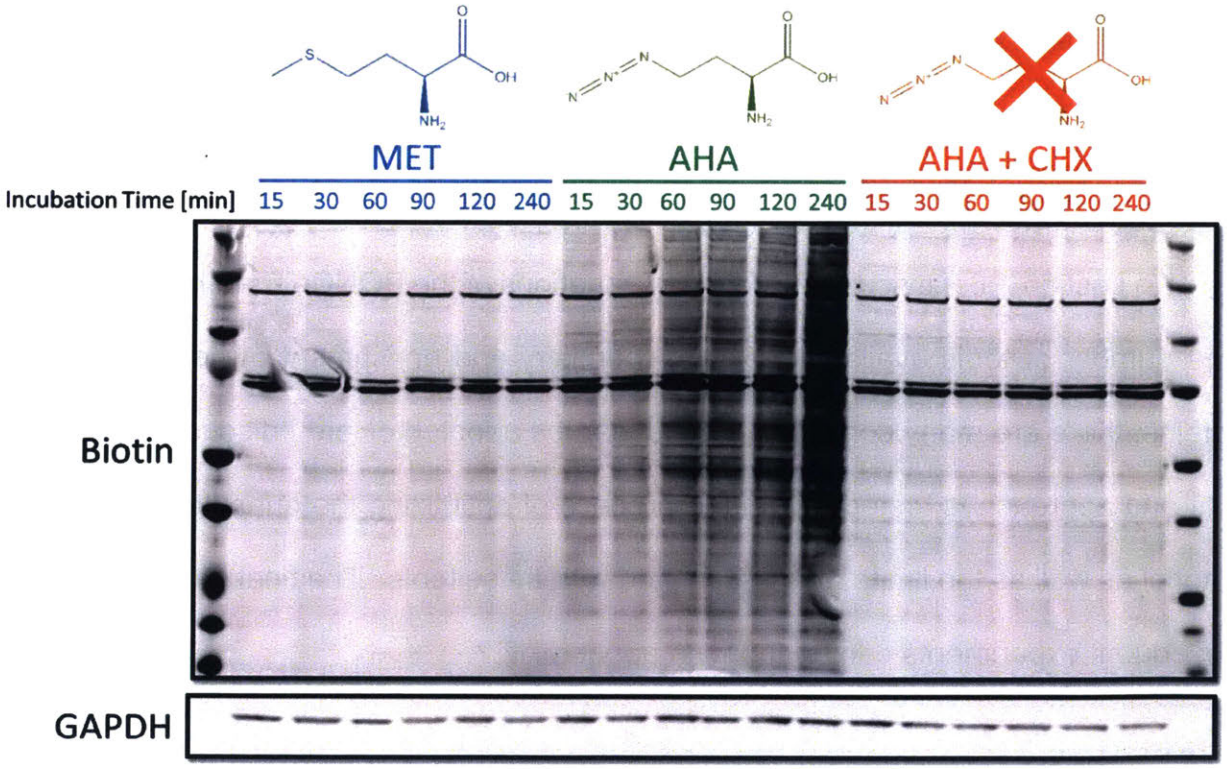


Figure 2-1 – Time dependent incorporation of Aha into newly synthesized proteins. Biotin labeling only occured when cells were labeled with Aha in the absence of CHX.

not in Met-labeled cells or in CHX-pretreated cells (Figure 2-1). The only biotin observed in the negative control samples were naturally biotinylated carboxylases[4] around 75 and 140 kDa that were present in all subsequent western blots. This result demonstrates that Aha was incorporated into proteins in a time-dependent manner, and that Aha incorporation is dependent upon the translation of new proteins.

Previous reports have demonstrated that DBCO is reactive towards sulfhydryl groups in addition to azides[5]. To test this hypothesis, cells were labeled with 5mM Aha, Met and CHX+Aha for 4 hours. Lysates were normalized to 1 mg/mL, and blocked with either 55 mM iodoacetamide (IAA) or 10 mM N-ethylmaleimide (NEM) for 60, 90, or 120 minutes at room temperature. DBCO-biotin was clicked on as before, and the sample was analyzed via Western blot. A normal exposure appears to show the reaction is highly specific, with biotin labeling only observed in the Aha samples. However, a long exposure shows that in the absence of thiol alkylation, there was substantial incorporation of biotin label into proteins. IAA reduced this nonspecific incorporation to some extent, and NEM appeared to prevent all detectable nonspecific incorporation of DBCO-biotin (Figure 2-2). To ensure all proteins were sufficiently alkylated, all subsequent experiments included NEM in the lysis buffer.

Ionic chaotropic detergents such as SDS lyse mammalian cells and solubilize membrane proteins efficiently, making them desirable for most cell biology applications. However, most detergents are incompatible with MS analysis and cannot be removed via traditional C18-based desalting methods. For most MS applications, cells are lysed in non-detergent chaotropic buffers such as 8 M urea. However, comparing click reactions in 0.5% SDS buffer and 8 M urea buffer reveals that the reaction was severely limited in

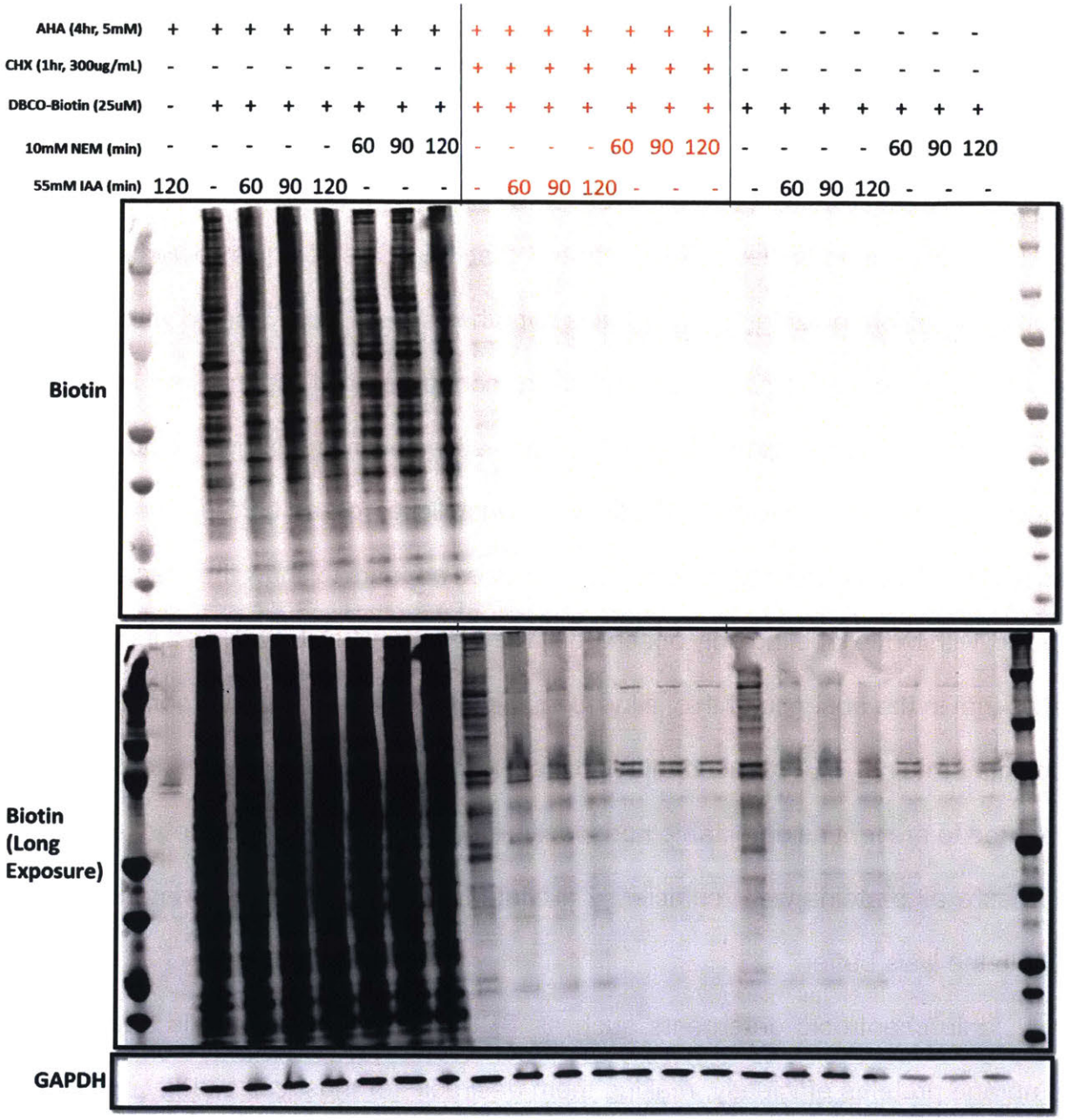


Figure 2-2 – IAA and NEM increased the specificity of the DBCO-azide click reaction by alkylating reduced sulfhydryls that would otherwise react with DBCO.

urea-based buffers (Figure 2-3). For this reason, only detergent-based buffers were considered going forward, and the process of enriching Aha-labeled proteins on a solid-state scaffold was used as a detergent-removal step.

For Non-SILAC MS-based experiments, the only way to positively identify Aha-labeled proteins was to actually observe the peptide containing the Aha residue. By limiting observations to peptides containing the Aha residue, the sensitivity of the experiments was restricted because of the limited coverage of observable peptides. In order to increase observations to all tryptic peptides, heavy arginine (R10) and heavy lysine (K8) SILAC labels were added concurrently with Aha. By labeling lysine and arginine residues with SILAC amino acids, every tryptic peptide from a protein could be identified rather than just the peptide containing the Aha residue. Initial experiments involved spiking K8, R10, and Aha directly into the media on cells. However, because

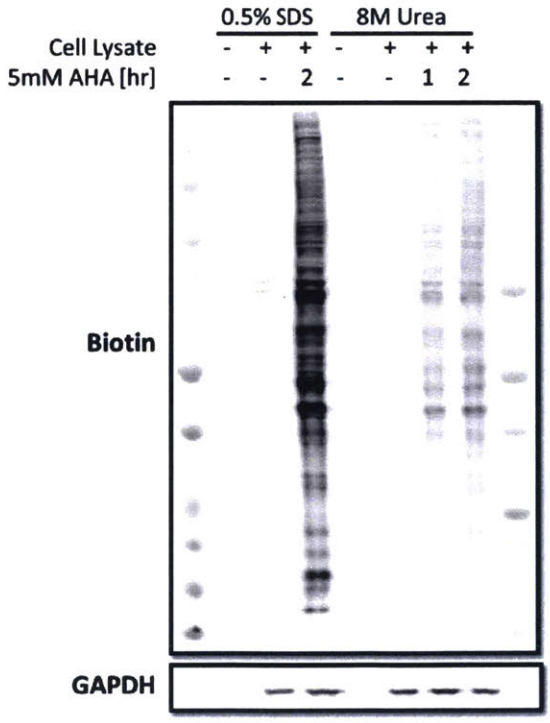


Figure 2-3 – DBCO-azide click reaction was less efficient in 8M urea than SDS-based lysis buffers.

cell culture media contained arginine and lysine on the order of 500-700 μM, K8 and R10 had to be added in excess of this concentration in order to outcompete the light amino acids present in the media. It was noted over several experiments that the addition of K8 and R10 to media appeared to reduce Aha incorporation in a dose-dependent manner (Figure 2-4). Arginine and Lysine are both transported into the cell by the cationic amino acid transporter 1 (CAT-1)[6]. While not cationic, Aha is zwitterionic, with both a positive and negative charge residing on the azide group. These results suggest that Aha may also be transported into the cell by CAT-1, and that by saturating the transporter with high concentrations of K8 and R10, Aha is not efficiently transported into the cell.

Ideally, K8, R10, and Aha would simply be spiked into the media without changing the media in order to limit artefactual perturbations to the cell. However, the need to reduce the concentration of total arginine and lysine (to allow for Aha incorporation) along with the desire for efficient incorporation of K8, R10, and Aha means metabolic labeling

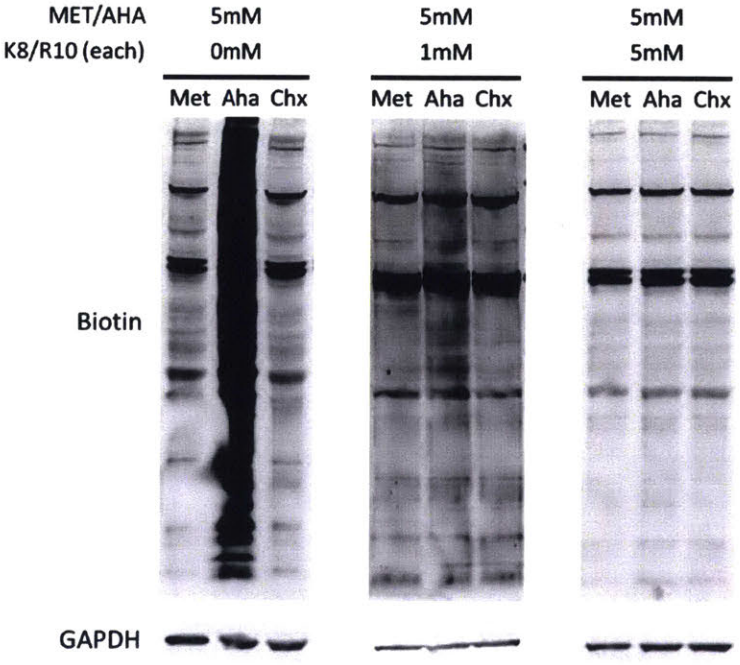


Figure 2-4 – Increasing the amount of pSILAC amino acids decreased the efficiency of Aha incorporation into proteins, even as the concentration of Aha remains constant.

needed to be done in media lacking arginine and lysine. Furthermore, since MetRS loads methionine onto tRNA's at a roughly 400-fold higher rate than Aha[7], then it would also be desirable to perform labeling in methionine-free media as well. Since the media was changed to lysine- and arginine-free media anyway to accommodate K8 and R10 labeling, it could also be changed to methionine-free media as well. An experiment comparing Aha incorporation in the presence of 0.5 mM arginine and 0.5 mM lysine in regular media versus KRM-free media demonstrated vastly superior Aha labeling in the KRM-free media (Figure 2-5). Therefore, all subsequent experiments were done in KRM-free media with a 30 minute starvation step to deplete internal amino acid stores prior to Aha labeling.

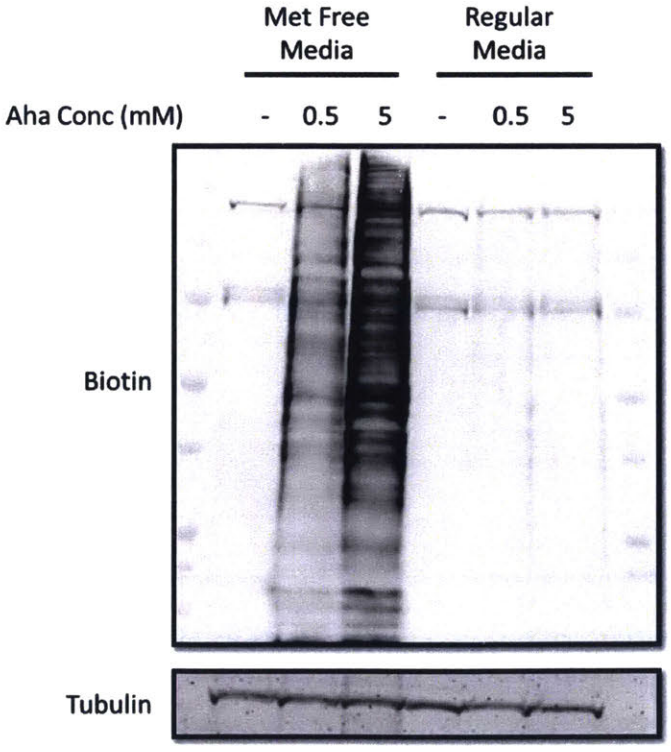


Figure 2-5 – Methionine-free media greatly enhanced Aha incorporation into newly synthesized proteins compared to spiking Aha into normal media.

Having changed to KRM-free media, I then sought to optimize the concentration of Aha for maximal protein labeling. Cells were labeled with 0.5 mM arginine and lysine in addition to either 0.1, 0.5, 1, 3, or 5 mM Aha in KRM-free media for 2 hours. Negative controls were labeled with 5 mM Met or 5 mM Aha with a CHX pretreatment. Western blotting revealed that Aha-labeling increased in a dose-dependent manner until 3 mM, at which point additional Aha did not appear to increase protein labeling. Aha-labeled proteins from these samples were also enriched using DBCO-magnetic beads (described in section 2.3.6) following acetone precipitation. Examining the supernatant revealed good depletion compared to the input at all concentrations (Figure 2-6). An on-bead trypsin digest liberated peptides for MS analysis. For each increase in Aha concentration, a corresponding increase in protein and peptide hits was observed via MS analysis,

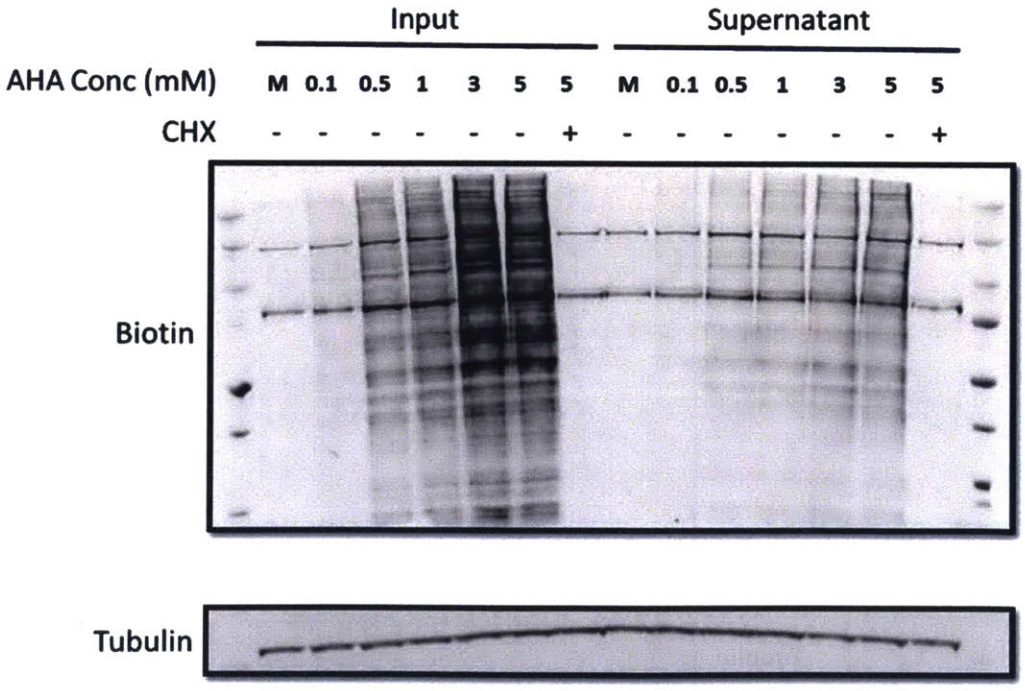


Figure 2-6 – Aha incorporation into proteins appeared to saturate around 3mM. Using the DBCO-magnetic beads from section 2.3.6, Aha-labeled proteins were depleted from the supernatant. M = 5mM methionine.

reaching a maximum at 3mM. This analysis indicated that an Aha concentration of 3 mM maximized the extent of incorporation into newly-translated proteins.

2.3.2 Biotin Capture and Elution

There are many different approaches to enriching Aha-labeled newly translated proteins from a whole cell lysate. The most straightforward approach was to use DBCO-biotin to modify Aha-labeled proteins with a biotin group and use streptavidin beads to enrich the modified proteins. Biotinylated samples were prepared as described previously, and 500 µg total protein was incubated over 30 µL streptavidin magnetic beads (ThermoFisher #88816) for 1 hour. The beads were washed in 1 mL each 1% NP-40 in 150 mM Tris, 2x 150 mM Tris, and water. Previous reports have shown that the biotin-streptavidin interaction can be broken in water at 70°C[8]. The beads were incubated in 25 µL 70°C deionized water for 10 minutes, the supernatant was removed, and the beads were boiled in 1x LDS buffer for 10 minutes at 98°C. Equal amounts of each sample were analyzed by Western blot (Figure 2-7). There appeared to be very little, if

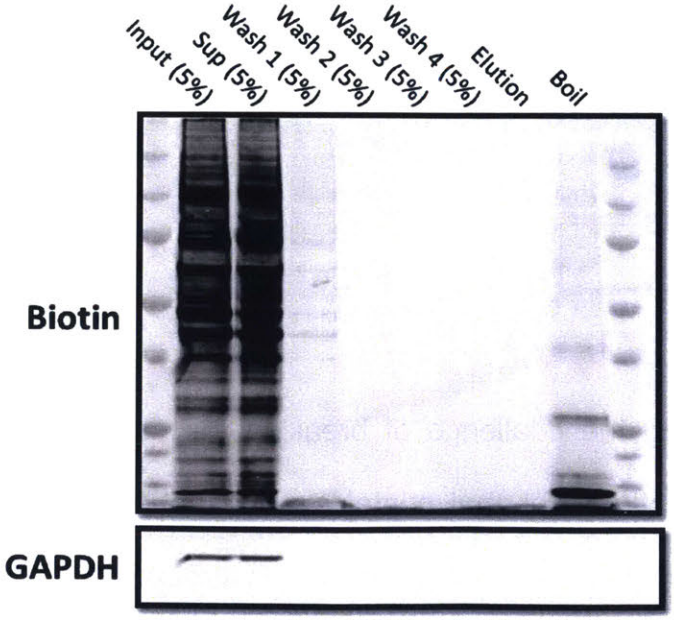


Figure 2-7 – Pulldown and elution of biotin-tagged proteins was inefficient. Proteins were eluted by incubating in 70°C water for 10 minutes.

any, depletion in the supernatant compared to the input lanes, and the elution in water was completely empty. The boiling step appeared to elute some biotinylated proteins, but not much compared to the amount observed in the input.

One potential cause for inefficient enrichment is the presence of unreacted DBCO-biotin that competes for binding on the streptavidin beads. To prevent this competition from occurring, the approach was changed to pre-loading the streptavidin beads by incubating 30 µL beads with 25 µM DBCO-biotin, and subsequently enriching by clicking proteins onto the beads. Aha-labeled lysate was applied directly to the pre-loaded beads for 1 hour at room temperature and washed as before. Once again, very little depletion was observed compared to the input, and the elution remained inefficient compared to the boiling sample (Figure 2-8).

Figure 2-8 – Preloading streptavidin beads with DBCO-Biotin failed to enhance enrichment, and the elution remains inefficient.

2.3.3 Photocleavable Linker

One approach to overcoming the challenge of breaking the streptavidin-biotin interaction is to utilize a DBCO-biotin linker that contains a cleavable moiety. This approach would allow for the elution of Aha-labeled proteins from streptavidin under mild

conditions, increasing specificity and eliminating the need for detergent removal. The first attempt utilized a photocleavable linker designed to dissociate when exposed to ultraviolet (UV) light (Figure 2-9, top). Aha-labeled lysates were incubated with 25 μM either DBCO-biotin or photocleavable DBCO-biotin (PC DBCO-biotin, Click Chemistry Tools #1120) and then exposed to a UV lamp for a variable amount of time. DBCO-biotin appeared to label the lysates more efficiently than PC DBCO-biotin in the absence of UV exposure, but UV exposure greatly increased biotin labeling of proteins (Figure 2-9,

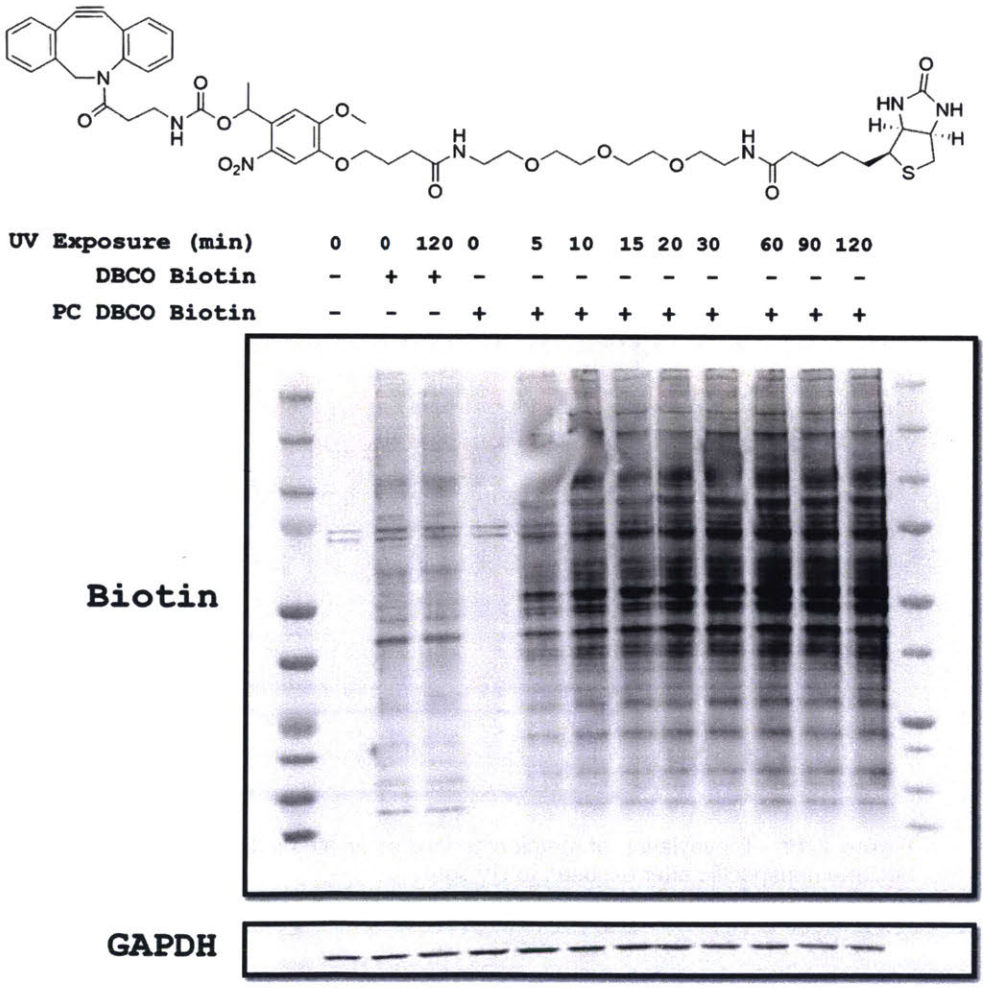


Figure 2-9 – Tested photocleavable linker contains a UV-labile chemical group separating the DBCO group and biotin (above). Exposure to UV light increased rather than decreased biotinylation of proteins (below).

bottom). This result was unexpected, as the exposure to UV light should have cleaved biotin from the proteins, resulting in less signal rather than more signal.

To test whether the increased biotin incorporation was specific to Aha-labeled proteins, the same assay was performed on lysates from Met-labeled cells. While labeling with the regular DBCO-biotin linker was no longer observed, UV light exposure resulted in extensive biotin labeling of proteins when using PC DBCO-biotin (Figure 2-10). These results indicate that the photocleavage process resulted in nonspecific biotinylation of proteins, perhaps through a reactive free-radical intermediate.

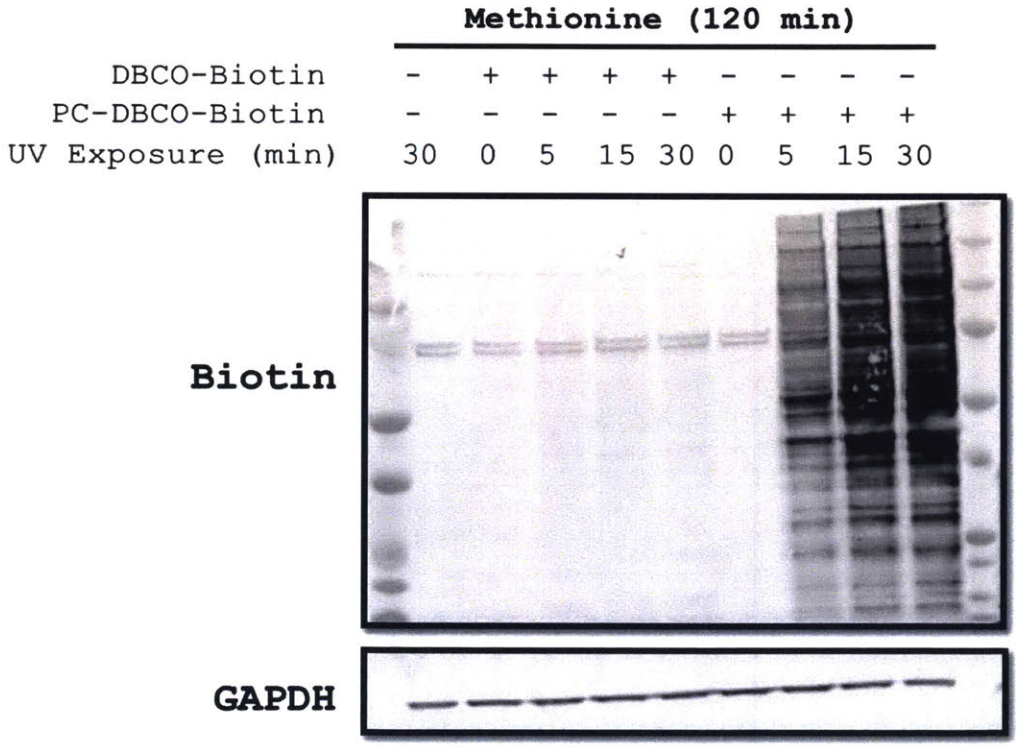


Figure 2-10 – Biotinylation of methionine-labeled lysates indicated that PC DBCO-biotin became nonspecific after exposed to UV light

2.3.4 Disulfide Cleavable Linker

Due to the nonspecificity of the PC DBCO-biotin linker, an alternative approach was explored using a disulfide-based chemical cleavable linker (DBCO-SS-biotin, Click Chemistry Tools #A112) (Figure 2-11, top). This linker is designed to be cleaved under reducing conditions, such as treatment with DTT or TCEP, allowing for the efficient elution from streptavidin beads under mild conditions. To test whether the linker could indeed be cleaved under reducing conditions, lysates with or without Aha-labeling were incubated with 25 µM either DBCO-biotin or DBCO-SS-biotin for 1 hour. Following the click reaction, the lysates were aliquoted into tubes, and treated with 10 mM TCEP, 10 mM DTT, or water. Western blotting revealed that biotin labeling was strongly reduced

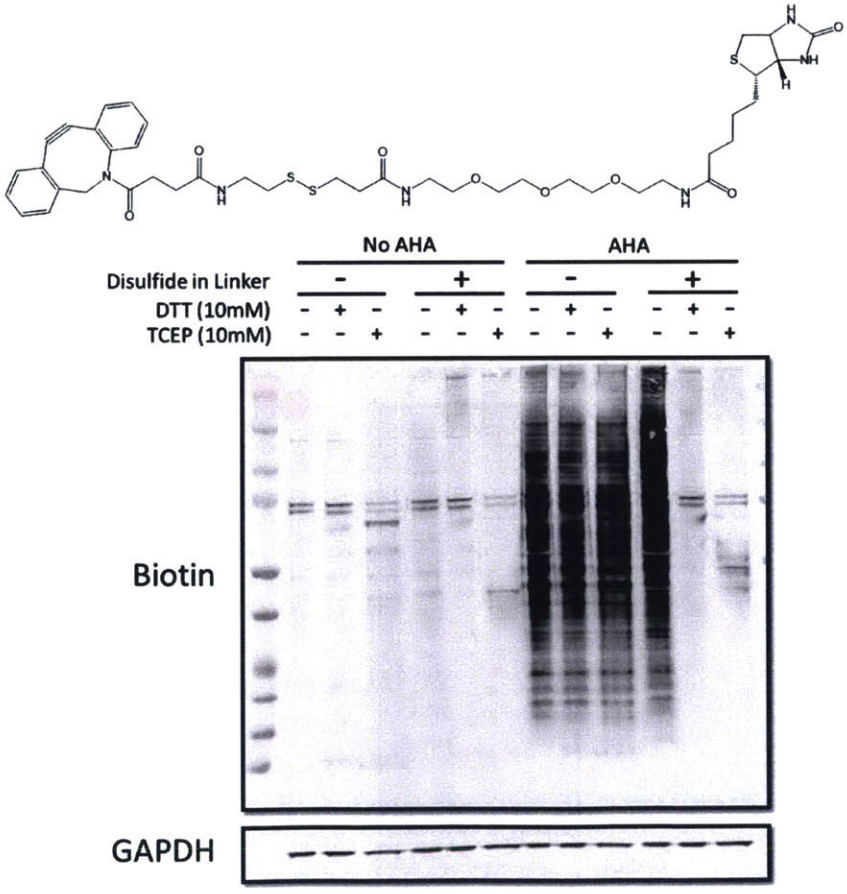


Figure 2-11 – DBCO-SS-biotin contains a disulfide bridge (top) that could be cleaved under reducing conditions, such a treatment with DTT or TCEP (bottom, right two lanes)

following treatment with DTT or TCEP only in DBCO-SS-biotin samples, indicating that the linker was indeed readily cleaved under reducing conditions (Figure 2-11, bottom).

After closer examination, it was noted that DBCO-SS-biotin did not have the robust specificity that was observed with regular DBCO-biotin. A side-by-side comparison of biotin incorporation into both Aha-labeled and control Met-labeled lysates clearly showed that DBCO-SS-biotin resulted in nonspecific incorporation of biotin into proteins that did not occur with DBCO-biotin (Figure 2-12). There has been evidence disulfides in similar linkers may exchange with disulfide bridges present in proteins[9], resulting in nonspecific incorporation of biotins independent of Aha labeling status. A longer exposure of Figure 2-11 also revealed the nonspecific incorporation of biotin in lysates lacking an Aha label (not shown). Normally, disulfide bridges could be reduced and alkylated, but the treatment of lysates with reducing agents would also result in the reduction of azides to primary amines, eliminating the chemical handle required for enrichment of newly-translated proteins.

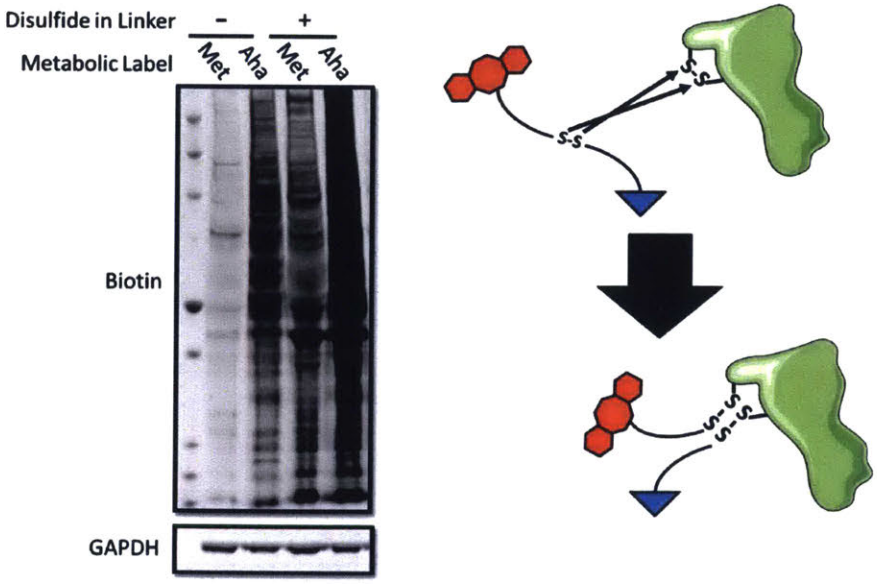


Figure 2-12 – The presence of a disulfide bridge in the DBCO-biotin linker introduced nonspecific biotin incorporation in methionine-labeled lysates. The most likely cause for this nonspecific incorporation was disulfide exchange between the linker and disulfide bridges present in the proteins.

2.3.5 Dde Cleavable Linker

As yet another alternative, a novel chemical linker was used incorporating a bioorthogonal Dde functional group designed to be cleaved in the presence of 2% hydrazine[10]. Dde groups are traditionally used as a protecting group for ε-amines of lysine residues during peptide synthesis, but have been repurposed here to create a Dde-based cleavable DBCO-biotin linker (DBCO-Dde-biotin, Click Chemistry Tools #1138) (Figure 2-13, top). 25 µM DBCO-biotin or DBCO-Dde-biotin was reacted with Aha- and Met-labeled lysates for 1 hour followed by the addition of hydrazine to 2% for 1 hour at room temperature. As a control, DTT was added to 10 mM or water was added at an equal volume. The addition of Dde to the DBCO-biotin linker did not appear to confer non-specificity as seen in the Met-labeled samples. Furthermore, cleavage of biotin from the proteins was observed only in the Aha sample labeled with DBCO-Dde-biotin and treated with 2% hydrazine (Figure 2-13, bottom). The specificity of labeling and efficiency of

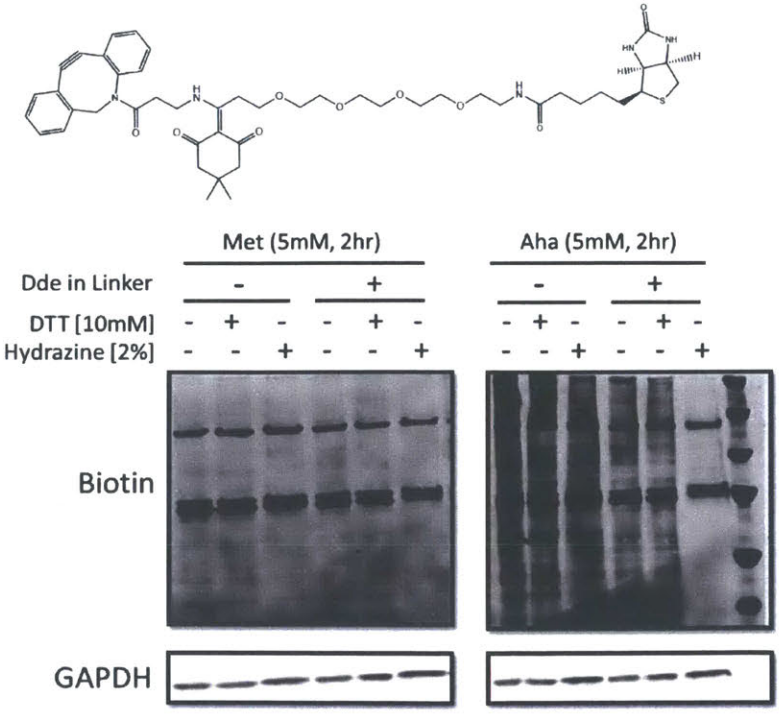


Figure 2-13 – DBCO-Dde-biotin contained a Dde functional group (top) that could be cleaved efficiently in the presence of 2% hydrazine (bottom).

cleavage made DBCO-Dde-biotin an excellent candidate for further protocol development.

Unfortunately, upon MS analysis of the linker itself, it was discovered that the predicted mass ([M+H] = 872.42 Da) was a very small fraction compared to other species present in the reagent (Figure 2-14, top). As a comparison, MS analysis was performed on DBCO-biotin and found that the predicted mass ([M+H] = 654.20) was the dominant mass species (Figure 2-14, bottom). This observation brings into question the purity of the DBCO-Dde-biotin reagent. Because this reagent was relatively novel, it was not widely available from commercial vendors. Therefore, I was forced to discard the DBCO-Dde-biotin reagent and take a different approach to enriching Aha-labeled proteins.

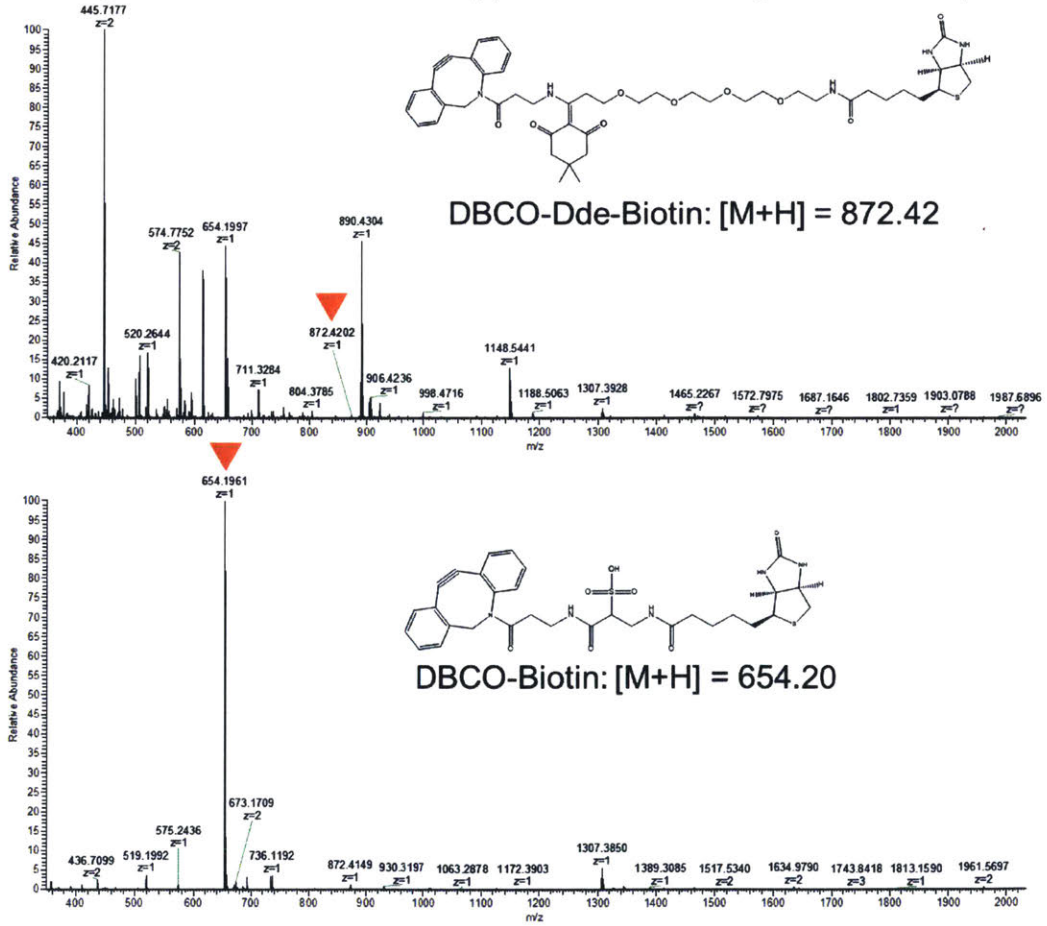


Figure 2-14 – The mass species associated with DBCO-Dde-biotin (872.42, top red arrow) made up a very small fraction of the total species present in the reagent (top). As a comparison, the expected mass of DBCO-biotin (654.20, bottom red arrow) was the dominant species present in the reagent (bottom).

2.3.6 DBCO Functionalized Magnetic Beads

Due to the limitations associated with the chemically cleavable biotin-DBCO linkers, I changed my approach to directly clicking proteins to DBCO-functionalized beads. Magnetic beads are smaller and easier to manipulate with magnetic stands and liquid handling robots than agarose beads, making them preferable for potential highly multiplexed quantitative analyses. Because of the limited availability of quality DBCO-magnetic reagents, I attempted to functionalize magnetic beads in-house.

To generate our own DBCO-functionalized magnetic beads, DBCO-PEG_{13}-NHS (Click Chemistry Tools #1015) (Figure 2-15, top) was coupled directly to amine-functionalized magnetic beads (Dynabeads M270 Amine, Thermo Fisher #14307). The 13xPEG spacer allowed for extension of the DBCO group into solution, preventing steric blockages that may occur with shorter spacers. To functionalize the beads, a variable concentration of DBCO-PEG_{13}-NHS in 10µL was incubated with 20 µL amine beads in 40 µL 500 mM TEAB for one hour. The reaction was quenched with the addition of Tris pH 7.4 to 0.5 M. Following the functionalization of the beads, the extent of DBCO-functionalization was assayed by clicking on an azide-conjugated fluorophore and measure fluorescence via flow cytometry of the magnetic beads. 30 µM Oregon Green azide (ThermoFisher #O10180) was added to the beads in 100 µL 1% SDS in PBS and allowed to react for 1 hour. The beads were washed 2x in 1% SDS, 1x in 0.1% NP-40, and then resuspended in 1 mL 0.1% NP-40 for flow cytometry analysis. It was found that beads were partially functionalized with DBCO-PEG_{13}-NHS starting around 300 µM, and

reached saturation around 2.4 mM (Figure 2-15, bottom). To ensure complete functionalization, a concentration of 4.8 mM was used going forward.

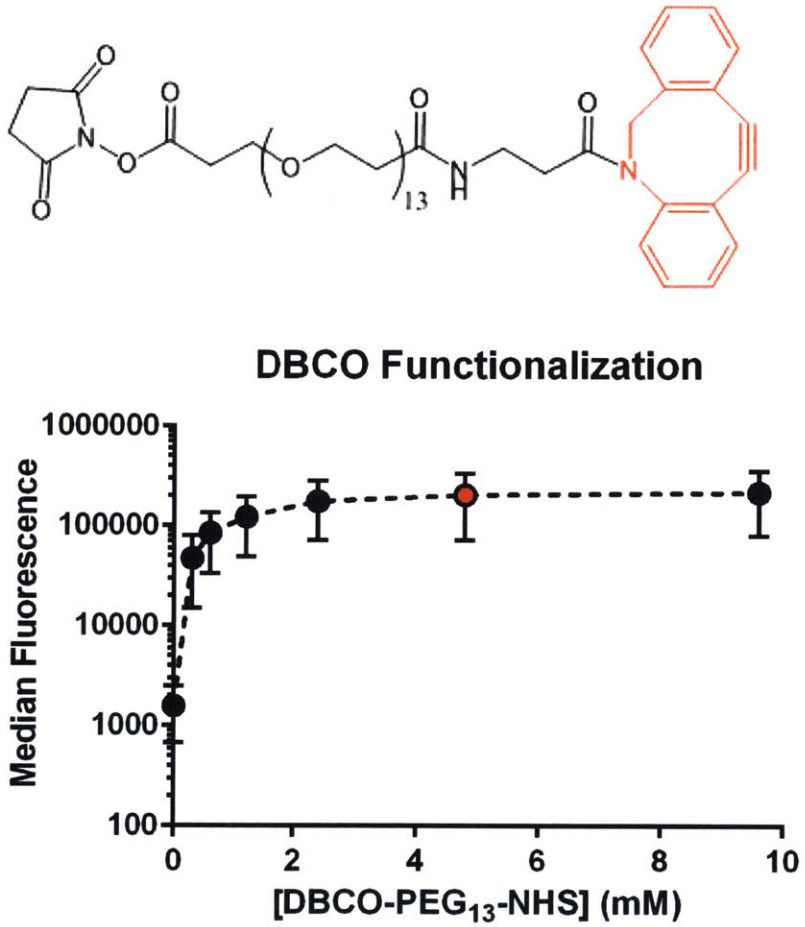


Figure 2-15 – DBCO-PEG$_{13}$-NHS (top) functionalizes amine-coated magnetic beads as measured by flow cytometry following the ligation of a azide-fluorophore to the DBCO groups (bottom). A concentration of 4.8 mM DBCO-PEG$_{13}$-NHS was used for functionalization (red data point).

I next sought to determine if a shift in fluorescence could be detected after enriching Aha-labeled proteins from a lysate. Aha-labeled proteins should react with the DBCO groups, quenching them and preventing them from reacting with the azide-fluorophore. Control lysates treated with Met or CHX+Aha should not react with the beads, and the beads' capacity to react with the azide-fluorophore should not be

diminished. Lysates were prepared by labeling with 5 mM Aha or Met for 30 minutes, lysing in 1% SDS + 50 mM NEM + 300 µg/mL CHX, and normalizing to 200 µg total protein in 100 µL. Beads were functionalized with DBCO-PEG_{13}-NHS as previously described above. Following functionalization, the beads were incubated with the lysate overnight at 40°C. The following day, the supernatant was removed, and the beads were washed with the following regimen: 2x 1% SDS in PBS, 15 min washes each of 100% MeCN, 50% MeCN+0.5% SDS in PBS, 1% SDS in PBS, 2 M NaCl in PBS, and quick 2x 100 mM ammonium acetate pH 8.9. After washing, the beads were again incubated in 3 µM Oregon Green azide for 1 hour at 40°C and analyzed via flow cytometry. Among the DBCO-functionalized beads, it was observed that beads incubated with the Met-labeled cells exhibited fluorescence, and the beads incubated with the Aha-labeled lysates were indeed quenched. However, it was also observed that the beads that had been incubated with the Aha+CHX control, which has translation inhibited, exhibited quenching as well (Figure 2-16). This observation suggests that the intracellular Aha that had not been incorporated into proteins was quenching the beads rather than Aha-labeled proteins.

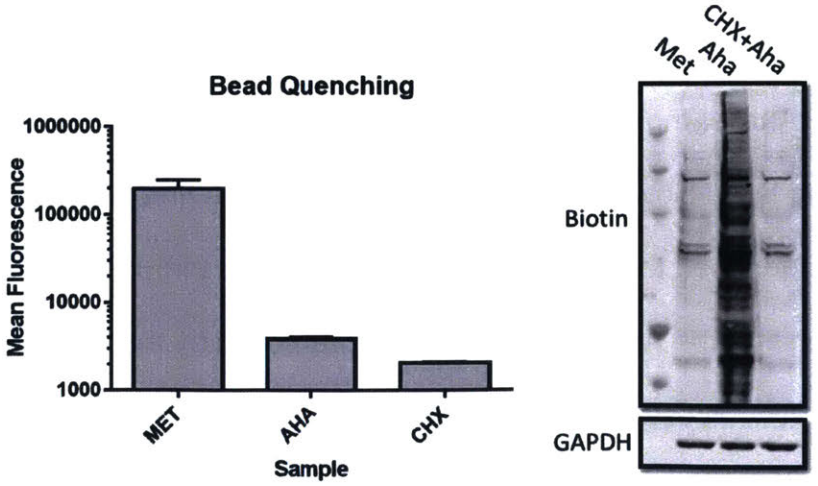


Figure 2-16 – While the Met-labeled sample did not quench the DBCO magnetic beads, both the Aha-labeled sample and CHX-pretreated Aha-labeled sample quenched the beads (left). Because the CHX+Aha sample did not have Aha-incorporated proteins (right), the quenching was most likely coming from excess Aha.

Two approaches were used to remove the excess free Aha from the lysates: acetone precipitation and molecular weight cutoff (MWCO) filters. Samples were labeled with either 5 mM Aha or Met for 2 hours. For acetone precipitation, 280 µL lysate was combined with 1.5 mL -20°C acetone and precipitated for at least an hour at -20°C. Following precipitation, the samples were centrifuged at maximum speed (~21,000 x g) for 15 minutes at room temperature to pellet the precipitated proteins. The supernatant was aspirated off the pellet, and the pellet was allowed to air dry for 15 minutes to remove residual acetone. 250 µL of 1% SDS in PBS supplemented with 50mM NEM was added to the pellet and vortexed vigorously to resuspend the pellet. For MWCO filtration, 280 µL lysate was added to a Vivaspin Hydrosart 2 kDa MWCO filters (Sartorius #VN01H92). The filter was spun for 30 minutes at 14,000 x g. The flow through was saved for analysis, and 300 µL lysis buffer was added to the top reservoir. This process was repeated again, and the eluate (~50 µL) was diluted in 150 µL lysis buffer. The lysates were all normalized to 2 mg/mL (or equivalent dilution for the flow through samples). The samples (except for the CHX flow through samples) were applied to functionalized magnetic beads and prepared for flow cytometry as described previously. For this experiment, neither the Met, Aha, nor CHX+Aha samples quenched the beads as had been previously observed (Figure 2-17, top left). However, it was observed that the first flow through sample from the Aha lysate quenched the beads, and the second flow through quenched it to a lesser extent, indicating that the Aha was indeed being removed by the filtration process (Figure 2-17, bottom left). Western blots comparing Aha-labeled proteins in the input and supernatant for each sample indicated a moderate depletion in the supernatant compared to the inputs for both the MWCO filtered samples as well as the acetone precipitated

sample, but a substantial amount of Aha-labeled protein still remained in the supernatant (Figure 2-17, right). This result suggests that the removal of residual Aha was an important step for the enrichment of Aha-labeled proteins.

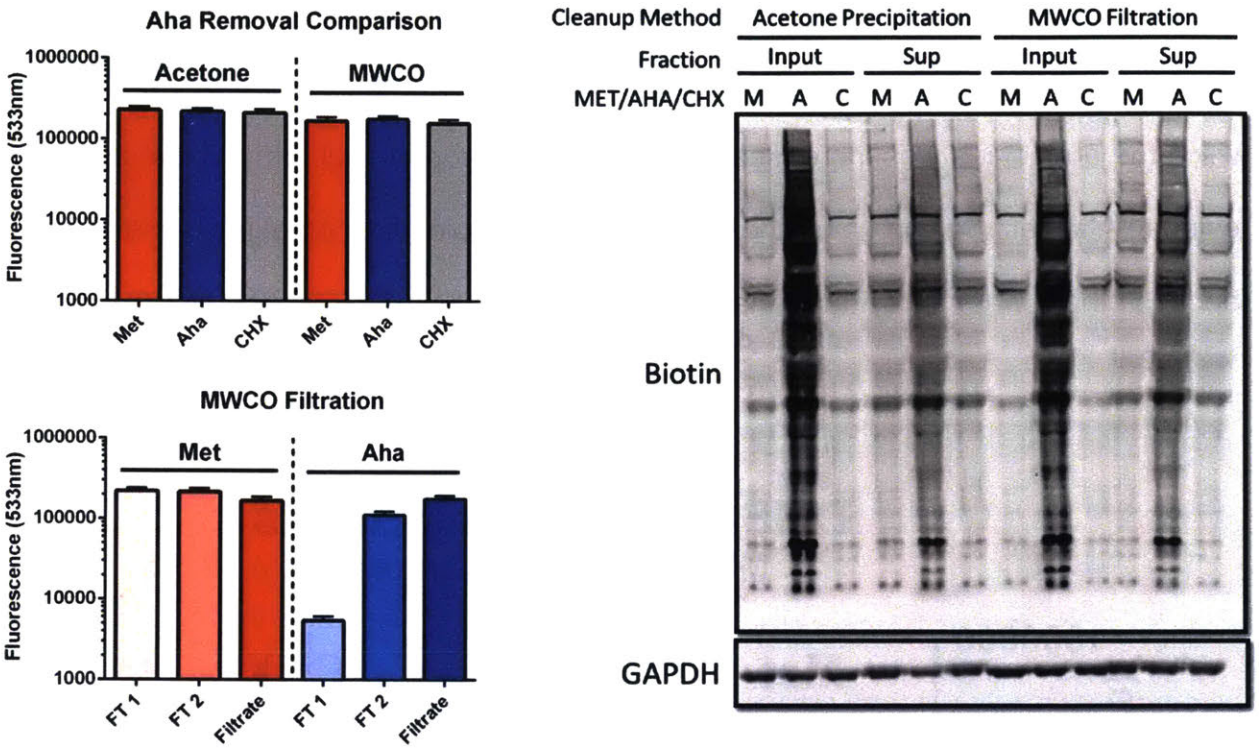


Figure 2-17 – Both acetone precipitation and MWCO filtration appeared to efficiently remove excess Aha because the beads were no longer quenched (top left). The MWCO filtration flow through samples from the Aha sample, but not the Met sample, quenched the beads, showing that excess Aha was indeed removed by filtration (bottom left). However, while depletion of Aha-labeled proteins was observed from the supernatants using both schemes, Aha-labeled proteins were observed that were not enriched (right), despite there being remaining bead binding capacity as measured by flow cytometry.

Another potential factor that could increase efficiency of the click pulldown is optimization of the solvent in which the reaction takes place. Proteins were readily soluble in DMSO up to 50% (data not shown). Lysates were prepared as described previously, using MWCO filters and filtering the samples 2x. After normalizing total protein via BCA to 2 mg/mL, samples were diluted 2-fold in either DMSO or lysis buffer, and the samples were applied to DBCO-functionalized magnetic beads. Sample-loaded beads were

washed and prepared for flow cytometry as described above. In this experiment, the MWCO filters did not completely remove the residual Aha from the sample, as indicated by the slight quenching of the Aha and CHX+Aha samples compared to the Met sample. Interestingly, in the presence of DMSO, the quenching was greatly enhanced despite the same amount of protein being present (Figure 2-18, left). Furthermore, the degree to which the quenching was increased appeared to be proportional to the initial amount of residual Aha left behind, and the Met labeled sample did not exhibit this increase (because there was no residual Aha in this sample).

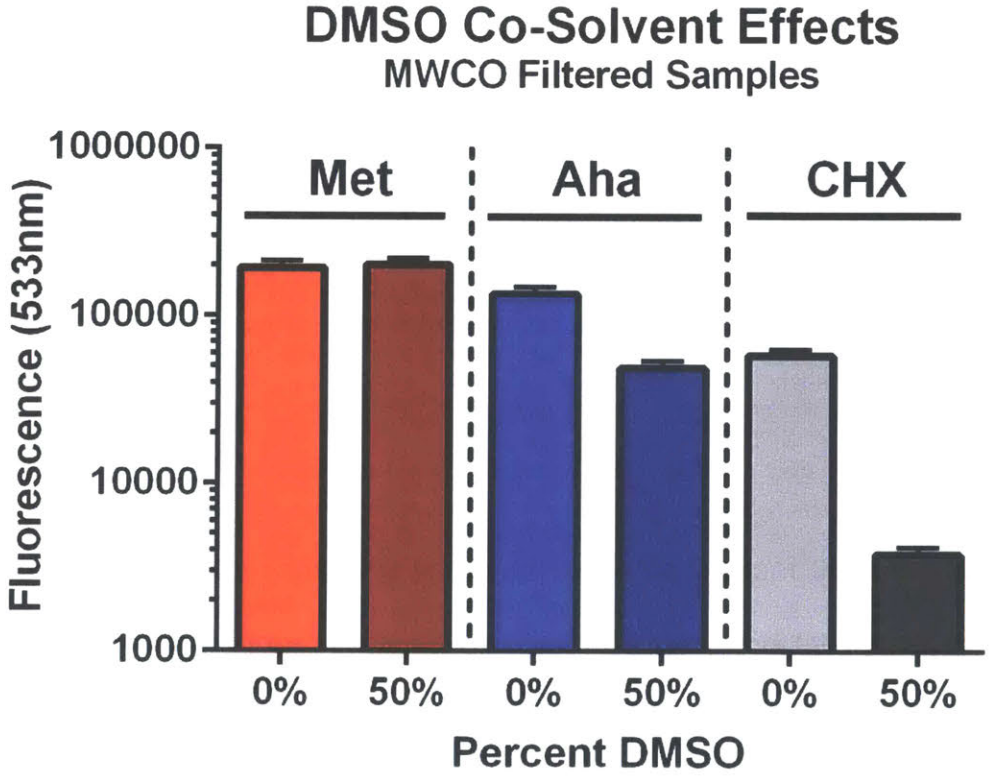


Figure 2-18 – The addition of DMSO as a solvent to the click enrichment buffer increased the degree of DBCO quenching. This result was most likely due to incomplete removal of Aha from the lysate by filtration, as the CHX-treated sample showed quenching as well.

A pilot MS experiment combining pSILAC with Aha labeling used all the optimized conditions previously developed. Cells were labeled for 2 hours with 3 mM Aha, 0.5 mM K8, and 0.5 mM R10 in KRM-free media. Negative controls used 3 mM Met and CHX pretreatment before Aha labeling. Following labeling, cells were lysed and proteins precipitated in acetone. It had previously found that acetone precipitation was more reliable and efficient at removing excess Aha than MWCO filters, so this, and all subsequent experiments used acetone precipitation. Proteins were resuspended in 1% SDS + 50 mM NEM and normalized to 2 mg/mL in 500 µL 50% DMSO + 0.5% SDS in PBS. Samples were applied to DBCO-magnetic beads and washed as previously described. The sample was resuspended in 50 µL 100 mM ammonium acetate pH 8.9 + 625 ng trypsin and digested overnight at room temperature. The digest was transferred to a new tube and loaded onto a C18 trapping column at 400 psi. The column was thoroughly washed with 0.1% AcOH for 20 minutes at 400 psi prior to being placed in line with an Agilent 1260 Infinity HPLC for MS/MS analysis. The sample was run using the following gradient at 0.2 mL/min with a splitter (A = 0.2 M AcOH, B = 70% MeCN in 0.2 M AcOH):

Time	**Percent B**
0 min	0% B
10 min	13% B
105 min	42% B
115 min	60% B
120 min	100% B
128 min	100% B
130 min	0% B

The MS/MS spectra were generated using HCD set at 29 NCE. All three samples (Aha, Met, and Aha+CHX) were run consecutively.

LC-MS/MS raw data files were searched on MASCOT with fixed modifications for NEM alkylation on cysteines (+125.047 Da) and variable modifications for SILAC R10 on arginine residues (+10.008 Da), SILAC K8 on lysine residues (+8.014 Da), Aha substitution for methionine residues (-4.986 Da), diaminobutyrate (reduced Aha) substitution for methionine residues (-30.976 Da), and phosphorylation on tyrosine, threonine, and serine residues (+79.966 Da). Precursor tolerance was 10 ppm, fragment tolerance was 15 mmu, two miss cleavages were allowed, and the enzyme was set to trypsin. Peptides were considered to be positively identified if they had a score of at least 25.

The Aha labeled sample yielded 26 unique proteins with SILAC label at 60% specificity. The Met sample yielded 8 proteins, and the CHX+Aha sample yielded 7 unique proteins, all lacking SILAC labels. This pilot demonstrated that I could indeed enrich for newly-synthesized proteins with an Aha label, although the sensitivity needed to be greatly improved. The western blot comparing the input to the supernatant

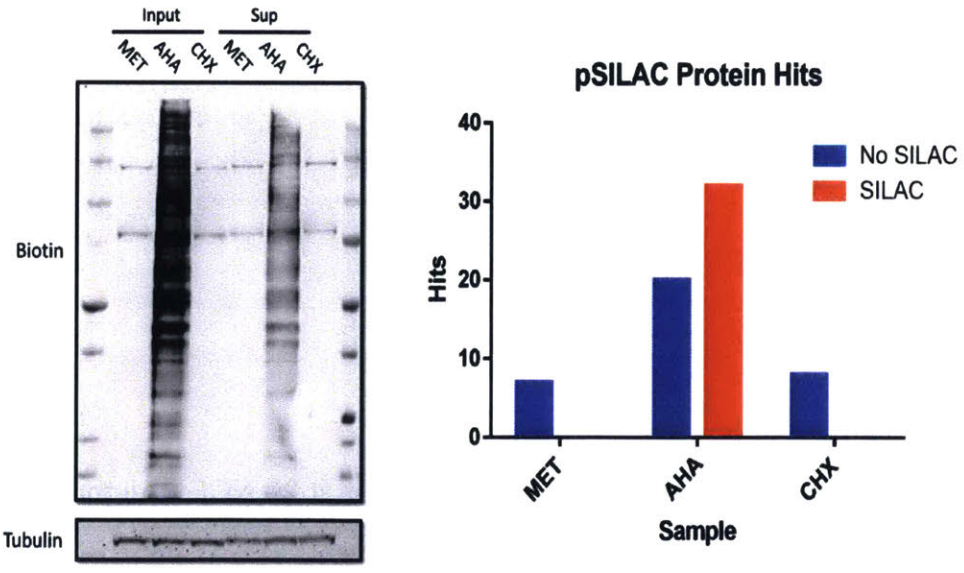


Figure 2-19 – A pilot pSILAC experiment using the optimized conditions (removal of Aha, DMSO in buffer) demonstrated a moderate depletion of Aha-labeled proteins from the supernatant following click enrichment (left). SILAC labeled peptides were detected exclusively in the Aha-labeled sample without a CHX pretreatment (right).

demonstrated that there was still a lot of Aha-labeling protein in the supernatant (Figure 2-19). It is unclear why so much Aha-labeled protein remained in the supernatant even though flow cytometry of the beads showed that there was additional binding capacity.

2.3.7 DBCO Functionalized Agarose Beads

While attempting to optimize the conditions for click enrichment on house-made DBCO-magnetic beads, a study was published that successfully used commercially available DBCO-agarose beads (Click Chemistry Tools #1034) to enrich newly-synthesized proteins from Aha-labeled plant seedlings[11]. Inspired by the success in this study, I opted to change the approach and used the DBCO-agarose resin instead of making the magnetic DBCO resin. Because the agarose beads average 200 µm in diameter, they are not compatible with flow cytometry analysis, so bead quenching was not able to be measured as had been previously done with the magnetic resin.

Aha-labeled lysates were prepared as described in section 2.2.2. For these samples, Aha was added at 3 mM for 120 minutes to the cells. As negative controls Met was added for 120 minutes, and Aha was added for 120 minutes to cells that had been pretreated with CHX. After the appropriate labeling period, cells were lysed in 250 µL 1% SDS supplemented with 50 mM NEM and 300 µg/mL CHX. Immediately following lysis, 1.5 mL -20°C acetone was added to the tube, and the lysate was stored at -20°C for at least an hour to precipitate proteins.

Following precipitation, the samples were centrifuged at maximum speed (~21,000 x g) for 15 minutes at room temperature to pellet the precipitated proteins. The supernatant was aspirated off the pellet, and the pellet was allowed to air dry for 15

minutes to remove residual acetone. 250 µL of 1% SDS in PBS supplemented with 50mM NEM was added to the pellet, along with ~1/10th of a vial of 0.7 mm garnet homogenizing beads to aid in pellet disruption. Samples were alternately vortexed and centrifuged at maximum speed until the pellet was completely solubilized.

Following resuspension, protein concentration was measured via BCA assay, and total protein was normalized to 300 µg in 500 µL of 1% SDS in PBS. Each sample was diluted 2-fold in 8 M urea + 850 mM NaCl to bring the final concentration to 300 µg in 1 mL. At this point, 12.5 µL of the lysate was aliquoted for biotin labeling of Aha-labeled proteins and western blot analysis (see section 2.2.3 and Figure 2-20, left). The sample was applied to 50 µL of DBCO-agarose beads that had previously been washed 3x in 1mL 0.8% SDS in PBS. Click enrichment took place overnight at room temperature on a rotor.

After the overnight click incubation, the bead/supernatant mixture was transferred to an empty spin column and allowed to drain into Eppendorf tubes to collect the supernatant. 12.5 µL of supernatant was aliquoted for biotin labeling of Aha-labeled proteins and western blot analysis (see section 2.2.3 and Figure 2-20, left). The western blot demonstrated revealed excellent depletion of Aha-labeled proteins following an overnight click enrichment. The tubes were rinsed out with 1mL MilliQ water and added to the spin column, allowing to drain into a waste container. To reduce disulfide bridges, 10 mM DTT in 0.8% SDS in PBS was added to the columns and the columns were capped and placed on a rotor at 50°C. Following reduction, the column was drained, and 50 mM NEM in 0.8% SDS in PBS was added to the column to alkylate the newly reduced cysteines. The columns were placed on a rotor at room temperature for 30 minutes. The

column was once again drained, and the beads were then washed 8x with 1 mL 0.8% SDS in PBS, 8x with 1 mL 8 M urea, and 8x with 1 mL 20% MeCN. After the second wash in each step, the column was capped and allowed to stand for 10 minutes.

Following the final wash step, the beads were completely dried by spinning the excess wash buffer into an empty Eppendorf tube. The beads were resuspended in 300 µL of 100 mM ammonium acetate + 10% MeCN (digest buffer) and transferred to a fresh tube. The column was twice rinsed with 300 µL digest buffer, with the rinses being combined with the sample. The beads were pelleted by centrifuging at 500 x g for 5 minutes at room temperature. The supernatant was carefully removed and replaced with 100 µL 1 ng/µL trypsin in digest buffer. The on-bead digest proceeded overnight at room temperature on a rotor.

After the overnight digest, the digest was acidified with the addition of AcOH to 10%. The beads were pelleted by briefly centrifuging the sample, and the digest was directly loaded onto a C18 trapping column at 400 PSI by inserting the end of a fritted column directly into the bead pack. The column was washed, placed in line with the

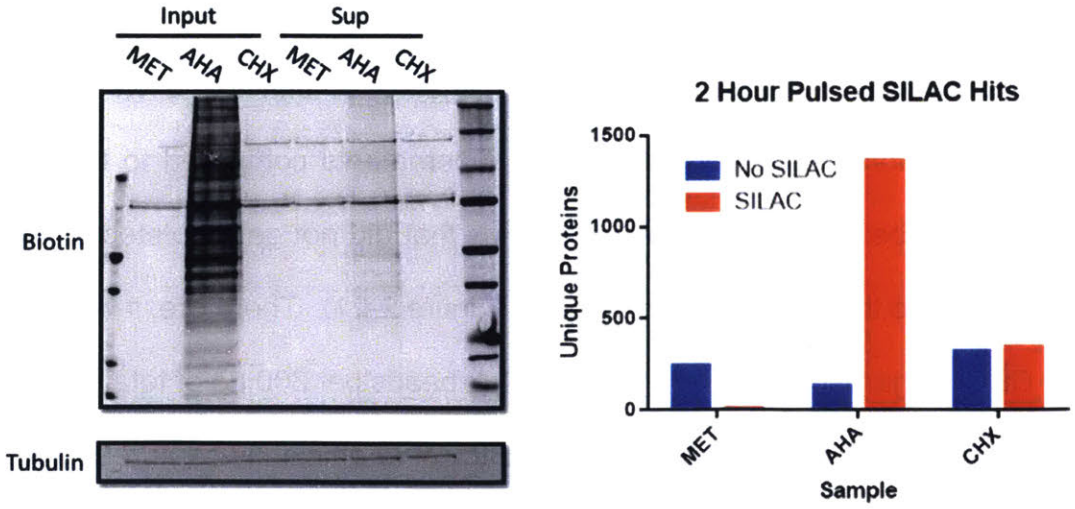


Figure 2-20 – Using commercially available DBCO-agarose beads, I observed excellent depletion of Aha-labeled proteins from the supernatant following click enrichment (left), and nearly 1400 unique SILAC-labeled proteins in the Aha-labeled sample (right).19

HPLC-MS setup, analyzed and searched as described in the previous section. A massive increase was observed in the number of SILAC-labeled proteins found in our Aha sample compared to the Met and CHX controls (Figure 2-20, right), with approximately 1363 unique SILAC labeled proteins identified in the Aha sample, compared to 244 in the CHX sample and 10 in the Met sample. As a measurement of nonspecificity, 135 non-SILAC labeled proteins were identified in Aha sample, 317 in the CHX sample, and 244 in the Met sample. This experiment demonstrated that newly-translated proteins could be enriched with a high sensitivity and specificity.

One final parameter that required optimization was the amount of DBCO-agarose beads to use for enrichment. Using too few beads could lead to bead saturation and loss of dynamic range, while using too many wastes reagent and lead to an increase of nonspecific binding. Aha, but not SILAC amino acids, were used to label newly synthesized proteins for 30 minutes, and lysates were prepared and processed as previously described. The lysate was split into 3x300 µg aliquots in 1 mL final volume. The samples were applied to 10, 30 and 50 µL of DBCO-agarose resin. Following overnight click enrichment, the supernatant was recovered and a 12.5 µL aliquot of each supernatant and input was used for click conjugation of biotin and western blotting analysis. A depletion was observed from all supernatants compared to the input, but there appeared to be some Aha labeled proteins that did not get depleted in the 10 µL sample compared to the 30 and 50 µL samples (Figure 2-21). Therefore, it was estimated that using DBCO-agarose beads at a ratio of 30µL beads per 300 µg of total protein when labeling for 30 minutes with 3 mM Aha was sufficient to efficiently captures Aha-labeled proteins.

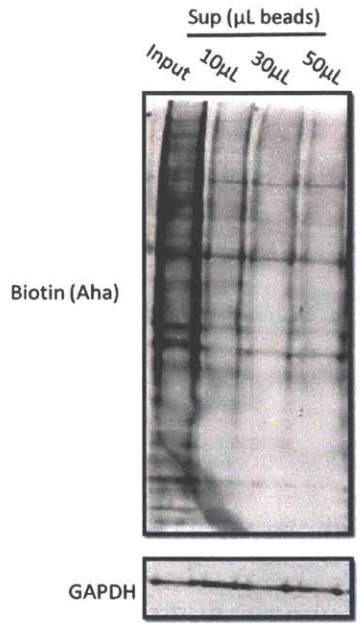


Figure 2-21 – Titrating the amount of beads used for a given sample revealed that 30 µL DBCO-agarose beads maximizes depletion of Aha-labeled proteins from 300 µg sample labeled with 3 mM Aha for 30 minutes.20

2.3.8 TMT Quantitation of Aha Time Course

To assess the robustness of this quantitative approach, An Aha labeling time course was performed (Figure 2-22, top). Using the approach described in section 2.2.2, 3 mM Aha was added to the cells for 15, 30, 60, or 120 minutes. As negative controls Met was added for 120 minutes, and Aha was added for 120 minutes to cells that had been pretreated with CHX. After the appropriate labeling period, cells were lysed, precipitated, resuspended, and newly synthesized proteins were enriched with 30 µL DBCO-agarose beads per condition as previously described. Aliquots were removed for Western blotting (Figure 2-22, middle). Following an overnight incubation, the digest was carefully transferred to a new tube. The beads were washed 2x with 150 µL MeCN and pooled with the digest. The sample was placed in a vacuum centrifuge to reduce the volume to below 200 µL, and the samples were subsequently snap frozen and lyophilized.

Peptides were resuspended in 70 µL ethanol + 30 µL 0.5 M TEAB. Each aliquot of a TMT 6-plex set was resuspended in 30 µL MeCN and added to the samples. The

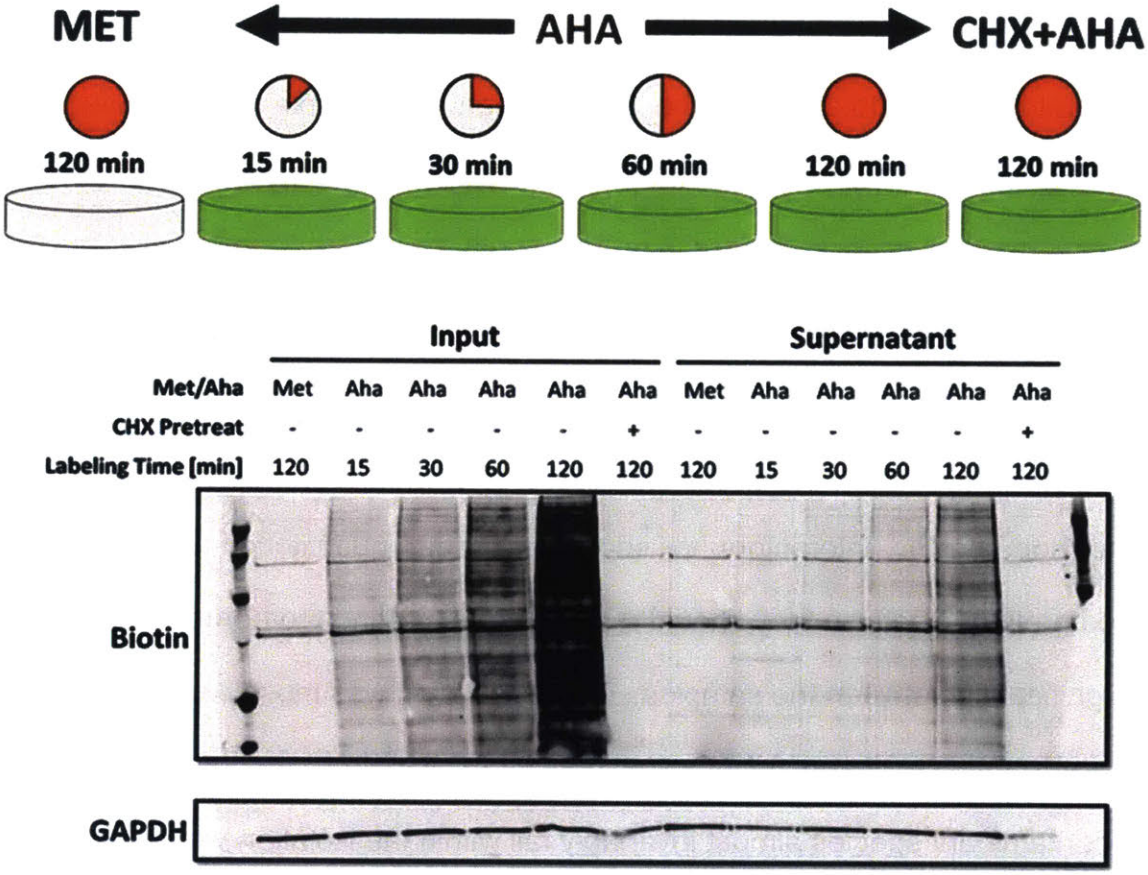
reaction was incubated for 1 hour at room temperature to allow complete TMT labeling of peptides. Following labeling, the reaction was quenched with the addition of 15 µL of 1 M Tris pH 7.4 and the samples were placed in a vacuum centrifuge for 30 minutes to reduce the volume to roughly 50 µL. The samples were then pooled into a single tube, with each tube being rinsed 3x with 40% MeCN + 0.1% AcOH and pooled with the samples. The labeled peptides were then vacuum centrifuged to dryness. The pellet was fully resuspended in 0.1% AcOH + 10% MeCN and loaded onto a C18 trapping column at 400 psi. The column was washed for 20 min with 0.1% AcOH at 400 psi prior to being placed in line with an Agilent 1260 Infinity HPLC for MS/MS analysis. The analysis was performed using the same settings as described in section 2.3.9, except the NCE was increased to 33 to ensure fragmentation of the TMT reporter ions.

The LC-MS/MS raw data files were searched using MASCOT version 2.4 with the same parameters and modifications described in section 2.3.9, except the following changes: fixed addition of TMT 6-plex to N-termini (+229.163 Da), fixed addition of TMT to lysine residues (+229.163 Da), and variable addition of TMT 6-plex to SILAC K8 lysine residues (+237.177 Da) replacing unmodified SILAC K8.

While the enrichment of newly translated proteins was successful (618 SILAC labeled proteins, 186 without SILAC), there were still issues with the quantitation that needed to be fixed. For each doubling in Aha labeling time, I would expect roughly a doubling in enrichment of newly synthesized proteins. However, large deviations were observed from that expected value, with abnormally high enrichment in the 30 minute sample and depletion in the 60 minute channel resulting in a poor linear fit (Figure 2-22, bottom left). These deviations most likely occurred due to technical error in sample

handling, specifically in the removal of the digest from the beads and transfer to a new tube for drying prior to TMT labeling. In some samples, beads were inadvertently aspirated and transferred to the new tube, and in others, the sample may not have been completely transferred. The variations that arose from this step may have been the key contributors to the deviations observed in the quantitation.

The second major issue with the quantitation was a strong correlation between the quantitation in SILAC-labeled proteins (newly translated proteins) and the quantitation in the proteins that lack SILAC (nonspecific interactors) (Figure 2-22, bottom right). I did not expect the non-specific interactors to show the same quantitation as true newly synthesized proteins because they should represent the existing proteome, which should not vary significantly between the samples. This correlation was most likely coming from isolation interference between SILAC and non-SILAC labeled samples. In complex mixtures, contaminant species almost invariably fall within the isolation window around a targeted precursor. TMT quantitation represents the sum of all the ions that fall within the isolation window, and thus isolation interference can lead to suppression of quantitative dynamic range. While nonspecific, non-SILAC labeled proteins may not actually exhibit the predicted quantitation based on our time course, if SILAC-labeled contaminant peptides were isolated along with the targeted non-SILAC peptide, the contaminating ions would have skewed the quantitation towards that of the SILAC-labeled peptide. The converse is true as well: if non-SILAC labeled contaminants fell within the isolation window of a SILAC labeled peptide, then the dynamic range would be suppressed by these contaminants.



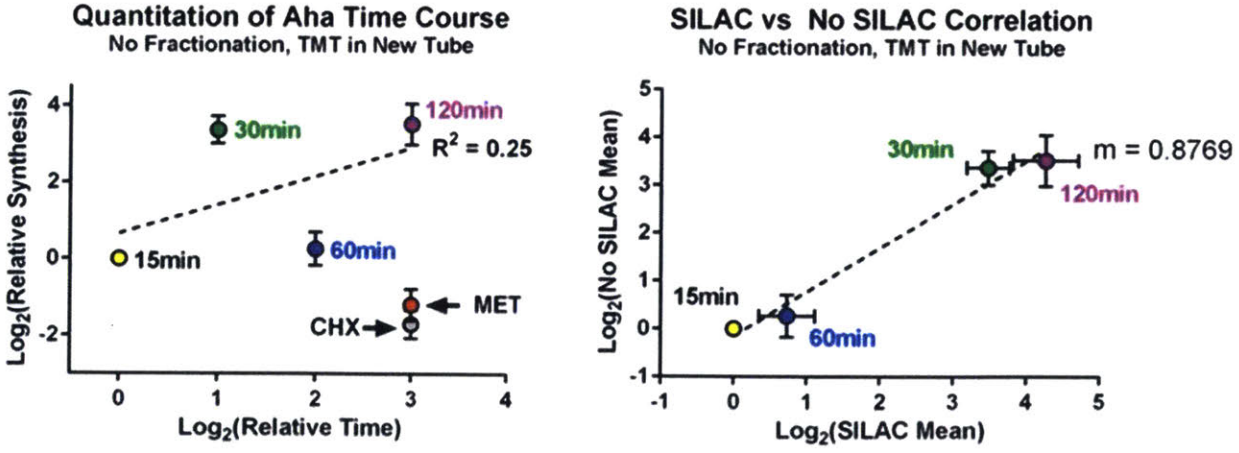


Figure 2-22 – Quantitative analysis was done by TMT 6-plex labeling of an Aha-labeling time course (top). We observed a time dependent increase of biotin incorporation of Aha into newly-translated proteins, and excellent depletion of Aha-labeled proteins from the supernatant following enrichment (middle). However, the quantitation was inconsistent, with a poor linear fit (bottom left) and a strong correlation between SILAC and non-SILAC labeled peptides (bottom right).

To overcome these challenges, two modifications were made to the protocol. First, the digest buffer was changed from 100 mM ammonium acetate pH 8.9 to 200 mM TEAB + 10% MeCN. The key difference between ammonium acetate and TEAB is that ammonium acetate has primary amines and TEAB does not. This difference is important because the TMT reagent is amine reactive and will be completely quenched in amine-based buffers. By changing the digest to an amine-free buffer, TMT labeling could be performed in the same tube as the digest, removing the transfer step that may be leading to technical variation.

The second key modification was the introduction of a high pH reverse phase fractionation step following TMT labeling. The typical HPLC setup for MS analysis used either acetic acid or formic acid as the aqueous mobile phase, resulting in low pH solvents. Under acidic conditions, most acidic and basic groups are protonated as long as their pKa is greater than the pH. Under basic conditions where the pH of the solvent is greater than the pKa of acidic and basic groups, these groups are deprotonated. The difference in hydrophobicity between the protonated and deprotonated state of peptides is sufficient to make high pH and low pH reverse phase HPLC orthogonal methods of separation[12].

To test if these changes improved the robustness and dynamic range of quantitation, the time course experiment was repeated with these changes included. The samples were prepared, enriched, reduced, alkylated, washed, and digested as previously described, replacing the ammonium acetate buffer with TEAB. Following the overnight digest, an aliquot of TMT dissolved in 30 µL anhydrous MeCN was added directly to the tube (beads included). TMT labeling proceeded for 1 hour at room

temperature, after which the reaction was quenched with the addition of 15 µL 1 M Tris pH 7.4. The volumes were reduced to about 50 µL in a vacuum centrifuge. All samples (including beads) were combined into a single Eppendorf tube. The individual tubes were rinsed 3x with 40 µL 50% MeCN + 0.1% AcOH, with the rinses being combined with the pooled sample. The sample was completely dried in a vacuum centrifuge.

Following drying, the sample (beads included) was resuspended in 500 µL 10 mM TEAB pH 8. The fritted end of a 200 µm i.d. capillary was placed in the bead pack, and the sample was loaded onto a ZORBAX Extend 300 C18 column (Agilent #770995-902) at 750 psi. The C18 column was placed in line with an HPLC, and the following gradient was run at a flow rate of 1mL/min. (A = 10mM TEAB pH 8, B = 99% MeCN, 10mM TEAB).

Time	Percent B
0min	0% B
1min	1% B
5min	5% B
50min	40% B
59min	70% B
64min	70% B
65min	1% B

Fractions were collected every minute between 5 min and 65 min. Every 12th fraction was concatenated together to give 12 total samples (5 fractions per sample). Samples were placed into a vacuum centrifuge overnight or until the sample reached dryness.

Each sample was resuspended in 50 µL 0.1% FA, transferred to an autosampler vial, and loaded into a Thermo Fisher Easy nLC 1000 autosampler. The MS settings were the same as previously described, but the gradient was changed to the following (A = 0.1% FA, B = 80% MeCN in 0.1% FA):

Time	Percent B
0 min	0% B
4 min	13% B
50 min	42% B
57 min	60% B
60 min	100% B
68 min	100% B
69 min	0% B
75 min	0% B

Each fraction was run consecutively, and 12 runs were combined into a single search using the same parameters listed previously.

Using this protocol, 2802 unique SILAC labeled proteins were detected with nearly 70% specificity. Furthermore, the dynamic range of quantitation was excellent, with a near doubling of protein abundance for each doubling in Aha labeling time, and far less protein in the Met and CHX negative control (Figure 2-23, left). Comparing the SILAC to non-SILAC labeling quantitation, the two groups were still correlated, but the slope of the line decreased to 0.584 from 0.877 (Figure 2-23, center). Furthermore, a line fitted to the Aha-labelled time points had a slope 0.87, which was slightly less than an expected value of 1.00 for perfect quantitation (Figure 2-23, right). This result indicates there still was dynamic range suppression to some degree, but the quantitation was sufficient to be used for practical purposes going forward. Further fractionation may increase the fitted slope closer to 1.00, but at the expense of more instrument time and longer experiments.

The degree of correlation between the SILAC and non-SILAC labeled proteins may also be explained by the fact that every Aha labeled protein may not contain a SILAC label at every arginine and lysine. Incorporation of Aha, arginine, and lysine are independent stochastic events, and although intracellular stores are depleted with the initial KRM-free media starvation, there remains a residual pool of arginine and lysine.

Even if a protein incorporates an Aha and is enriched, if the observed tryptic peptides do not have a SILAC amino acid, they will appear as non-specific. This hypothesis is supported by observations that the ratio of spectral counts of peptides containing a non-SILAC arginine to peptides containing a non-SILAC lysine was about 60:40 repeated across four replicates, which matches the ratio of arginine to lysine in DMEM:F12 media. On the other hand, the ratio of spectral counts of peptides containing R10 to peptides containing K8 was 50:50 across these replicates, matching the ratio of K8 and R10 in the labeling media. Therefore, peptides that do not contain a SILAC amino acid may have been a combination of enriched nascent proteins and nonspecific binders, resulting in a degree of correlation with specific SILAC-labeled peptides. This labeling could potentially be improved upon by further optimizing the concentration of SILAC amino acids used during metabolic labeling, but for my purposes, the current protocol was satisfactory. A complete protocol can be found in appendix 2.6.

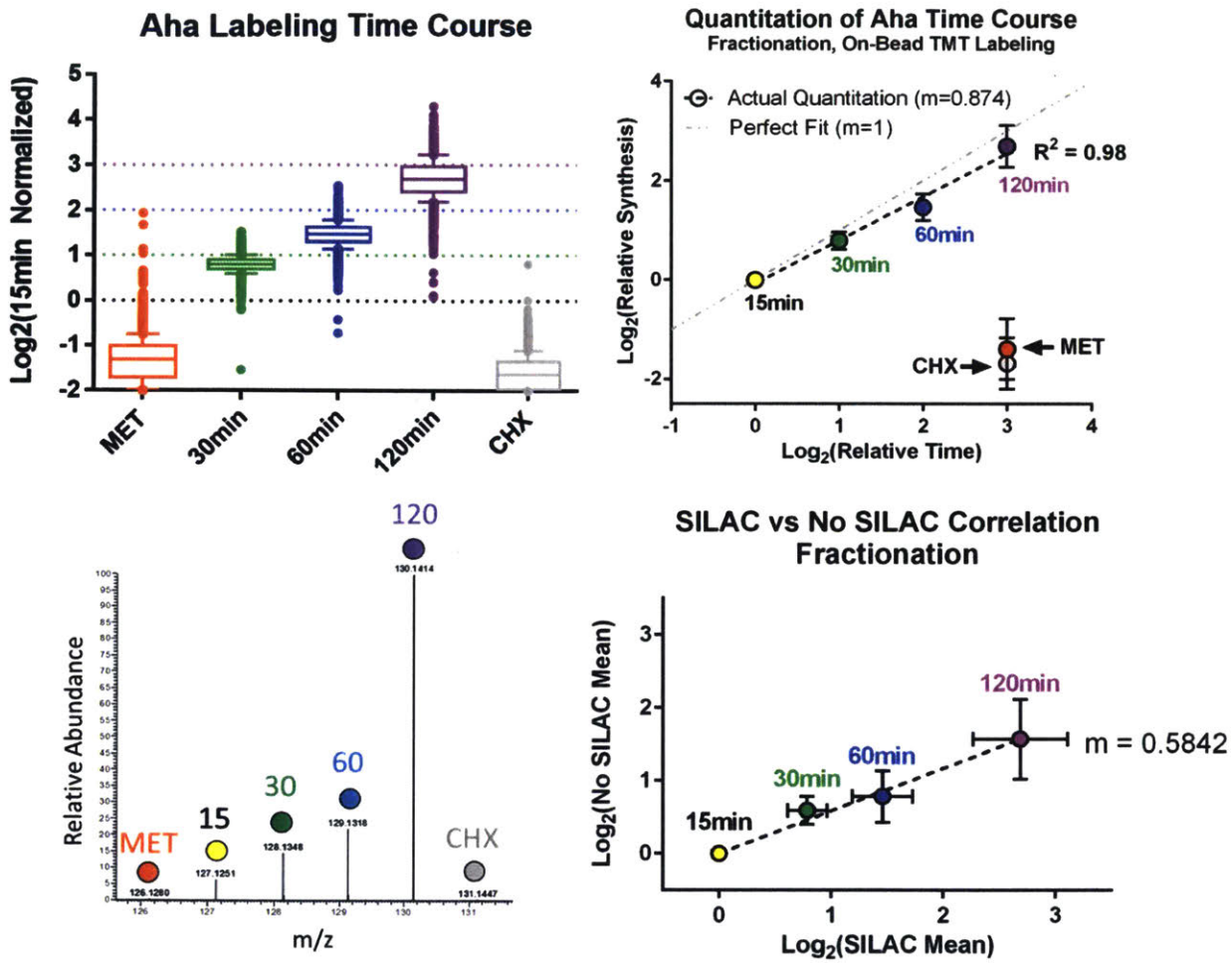


Figure 2-23 – With the modified protocol, Aha labeling increased quantitatively with a near doubling of protein enrichment for each doubling in labeling time (top left). A linear fit to the Aha-labeled time points yielded a slope of 0.87, which was slightly suppressed from the ideal slope of 1, but with an excellent fit with $R^2 = 0.98$ (top right). An exemplar spectra shows the quantitative increase in reporter ions (bottom left). The slope of the line fitting the quantitation from SILAC labeled proteins versus non-SILAC labeled proteins decreased from 0.87 to 0.58, demonstrating less interference between nonspecific and specific peptides (bottom right)

2.4 Conclusions

In this section, I have described the various tools available for the enrichment and detection of newly synthesized proteins, both for MS analysis and visualization via Western blotting. Ultimately, for the most quantitative approach, I combined SILAC labeling (for specificity) along with TMT labeling (for quantitation), and used commercially available DBCO-agarose beads for enrichment. After several modifications to the protocol, I was able to generate a robust, quantitative analysis of an Aha-labeling time course.

While the protocol as described here is satisfactory to our needs, several further optimization steps can be done for a more complete analysis. The concentration of K8 and R10 SILAC amino acids can be further adjusted to optimize the overlap between Aha-incorporation and SILAC labeling. In addition, to increase sensitivity and quantitative dynamic range, more individual fractions can be analyzed rather than pooling every 5^{th} fraction following high pH reverse phase fractionation.

In the following chapters, this quantitative BONCAT analysis will be applied to various applications to examine the nascent proteome in a time-dependent manner, as well as across multiple drug treatment conditions.

2.5 References

1. Howden, A. J. M. *et al.* QuaNCAT: quantitating proteome dynamics in primary cells. *Nat. Methods* **10,** 343–346 (2013).
2. Bowling, H. *et al.* BONLAC: A combinatorial proteomic technique to measure stimulus-induced translational profiles in brain slices. *Neuropharmacology* **100,** 76–89 (2016).
3. Ma, Y., McClatchy, D. B., Barkallah, S., Wood, W. W. & Yates, J. R. HILAQ: A Novel Strategy for Newly Synthesized Protein Quantification. *J. Proteome Res.* **16,** 2213–2220 (2017).
4. Chapman-Smith, A. & Cronan, J. E. The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. *Trends Biochem. Sci.* **24,** 359–363 (1999).
5. van Geel, R., Pruijn, G. J. M., van Delft, F. L. & Boelens, W. C. Preventing Thiol-Yne Addition Improves the Specificity of Strain-Promoted Azide–Alkyne Cycloaddition. *Bioconjug. Chem.* **23,** 392–398 (2012).
6. Closs, E. I., Simon, A., Vé, N. & Rotmann, A. Arginine Metabolism: Enzymology, Nutrition, and Clinical Significance Plasma Membrane Transporters for Arginine 1,2. *J. Nutr* **134,** 2752–2759 (2004).
7. Kiick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci.* **99,** 19–24 (2002).
8. Holmberg, A. *et al.* The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* **26,** 501–510 (2005).
9. Szychowski, J. *et al.* Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *J. Am. Chem. Soc.* **132,** 18351–18360 (2010).
10. Yang, Y. & Verhelst, S. H. Cleavable trifunctional biotin reagents for protein labelling, capture and release. *Chem. Commun.* **49,** 5366–5368 (2013).
11. Glenn, W. *et al.* BONCAT enables time-resolved analysis of protein synthesis in native plant tissue. *Plant Physiol.* **19,** pp.01762.2016 (2017).
12. Gilar, M., Olivova, P., Daly, A. E. & Gebler, J. C. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J. Sep. Sci.* **28,** 1694–1703 (2005).

2.6 Complete Quantitative BONCAT Protocol

1. Remove media from cells and replace with 5mL methionine/arginine/lysine free media for 30 minutes to deplete intracellular stores

2. Spike in treatment of choice along with Aha to 3mM (30uL of 300mM in PBS), K8 and R10 to 0.5mM (50uL each of 5mM stock in PBS) and incubate for desired amount of time for incorporation into newly translated proteins. Aha/K8/R10 labeling can also be offset from treatment to measure delayed effects of treatment.

3. After incubation period, aspirate off media and replace with 5mL ice cold PBS supplemented with 300ug/mL cycloheximide to halt translation

4. Aspirate off PBS (remove as much as possible) and add 250uL 1% SDS in PBS supplemented with 300ug/mL cycloheximide, 50mM NEM, and 1:1000 benzonase (added immediately before lysis).

5. Transfer lysate to 2mL Eppendorf tube. Add 1500uL of -20C acetone to precipitate proteins. Vortex well and incubate at -20C for at least one hour, and overnight if desired.

6. Pellet proteins by centrifuging at maximum speed (>13,000 rpm) for 15 minutes.

7. Aspirate or decant off acetone, careful not to disrupt protein pellet. Allow pellet to air dry for 15 minutes, but not longer to prevent over-drying.

8. Resuspend pellet in 250uL 1% SDS in PBS supplemented with 50mM NEM. Use a clean metal spatula or homogenizing beads to mechanically break apart pellet if necessary. Vortex and spin down at maximum speed for 1 minute. Repeat as necessary until pellet is fully resuspended.

9. Perform BCA on lysate to determine protein concentration. Normalize concentration to 100ug-500ug protein in 500uL 1% SDS in PBS. Add 500uL 8M urea + 850mM NaCl to bring final volume to 1mL.

10. (Optional): Remove 12.5uL aliquot for western blot analysis.

11. Aliquot 30uL DBCO-Agarose beads per sample. Wash 3x1mL 0.8% SDS in PBS. Split into equal aliquots during third wash.

12. Apply sample to DBCO-Agarose resin and incubate overnight at room temperature on rotor (at least 16 hours).

13. Transfer pulldown to 1mL spin column (Bio-Rad catalog #7326204). Allow supernatant to drain into fresh Eppendorf tube if desired or into waste container.

14. Add 1mL MilliQ water to the sample and allow to drain.

15. Add 500uL 10mM DTT in 0.8% SDS in PBS. Cap the column and incubate for 30 minutes on rotor at 50C. Drain the column after incubation.

16. Add 0.5mL 50mM NEM in 0.8% SDS in PBS. Cap the column and incubate for 30 minutes on rotor at RT. Drain the column after incubation.

17. Wash each column 8x1mL 0.8% SDS in PBS. After 2nd wash, cap column and incubate for 10 minutes.

18. Wash each column 8x1mL 8M urea. After 2nd wash, cap column and incubate for 10 minutes

19. Wash each column 8x1mL 20% MeCN. After 2nd wash, cap column and incubate for 10 minutes

For unlabeled MS analysis:

20. Transfer resin to a fresh Eppendorf tube in 300uL 100mM NH4CH3CO2 + 10% MeCN. Rinse column with 2x300uL 100mM NH4CH3CO2 + 10% MeCN, pooling each rinse with sample. Centrifuge for 5 minutes at 5000x g to pellet.

21. Remove all liquid using gel loading pipette tip. Resuspend in 100uL 100mM NH4CH3CO2 + 10% MeCN + 100ng trypsin. Digest overnight at RT.

22. Acidify digest with 10% AcOH.

23. Either load directly onto C18 precolumn, or filter out beads by flowing sample through fritted 200um column directly into an autosampler vial.

For TMT labeled MS analysis:

20. Transfer resin to a fresh Eppendorf tube in 300uL 200mM TEAB + 10% MeCN. Rinse column with 2x300uL 200mM TEAB+10% MeCN, pooling each rinse with sample. Centrifuge for 5 minutes at 5000x g to pellet.

21. Remove supernatant carefully using gel loading pipette tip. Resuspend in 100uL 200mM TEAB + 10% MeCN + 100ng trypsin. Digest overnight at RT.

22. Resuspend TMT aliquots in 30uL anhydrous MeCN, vortex and spin down. Add aliquot directly to digest. Label for 1hr at RT.

23. Reduce volume to ~40uL in vacuum centrifuge (30min). Combine samples in a fresh tube, washing each individual tube 3x with 50% MeCN.

24. Vacuum centrifuge to dryness

25. Resuspend bead/peptide pellet in 500uL 10mM TEAB pH 8. Briefly centrifuge.

26. Insert fritted end of 200um capillary into bead pack, and attach other end to input of a ZORBAX Extend 300 C18 column (Agilent catalog #770995-902). Load supernatant onto column at 500-1000 PSI.

27. Transfer C18 column to HPLC for high pH fractionation. Solvent A = 10mM TEAB pH 8; Solvent B = 99% MeCN, 10mM TEAB. Flow Rate = 1mL/min. Gradient:

0 min	0% B
1 min	1% B
5 min	5% B
50 min	40% B
59 min	70% B
64 min	70% B
65 min	1% B

Collect fractions every minute (1mL/fraction) from 5min to 65min.

28. Concatenate every 12th fraction down to 12 samples (5 fractions per sample). Vacuum centrifuge samples to dryness overnight.

29. Resuspend pellets in 40uL 0.1% formic acid. Transfer to autosampler vial and load into autosampler.

30. Analyze samples via LC-MS/MS. We used a Thermo EASY-nLC 1000 coupled to a Thermo QExactive mass spectrometer. For HPLC separation, we used low pH solvents. Solvent A = 0.1% formic acid; Solvent B = 80% MeCN, 0.1% formic acid. Flow Rate = 200nL/min. Gradient:

0 min	0% B
4 min	13% B
50 min	42% B
57 min	60% B
60 min	100% B
68 min	100% B
69 min	0% B
75 min	0% B

Chapter III

Validation of Multiplexed BONCAT with the Unfolded Protein Response

Daniel Rothenberg

3.1 Introduction

Following the development of a quantitative method for measuring newly translated proteins in Chapter 2, I sought to validate the robustness of this method in a biological context. The selected application should have a well-characterized response and demonstrate regulation on the level of translation. One application that fits both of these requirements is the unfolded protein response (UPR).

When a cell detects an accumulation of unfolded or misfolded proteins in the endoplasmic reticulum (ER), it triggers the UPR stress response in an attempt to return the cell to homeostasis. The UPR seeks to achieve three goals: 1) the inhibition of protein translation to prevent further misfolding, 2) the degradation of misfolded or unfolded proteins, and 3) the upregulation of protein folding chaperones to aid in the correction of misfolded proteins. These responses result from the integration of three distinct axes of the UPR[1].

The first axis is the inositol-requiring transmembrane kinase/endonuclease-1 (IRE1α) arm of the UPR. Under normal conditions, IRE1α is bound by BIP (also known as HSPA5 or GRP78) in an inhibited state. In the presence of ER stress, BIP dissociates from IRE1α and binds to the misfolded proteins. The dissociation of BIP from IRE1α also activates IRE1α, allowing it to dimerize and form an active endonuclease[2]. The endonuclease mediates a non-canonical splicing event, removing an intron from cytoplasmic unspliced XBP1 mRNAs, resulting in a transcriptionally active spliced XBP1 transcript[3]. The product of unspliced XBP1 (XBP1u) is highly unstable and rapidly degraded by the proteasome. The translated spliced XBP1 (XBP1s) codes for an active

transcription factor that is more stable than XBP1u and upregulates other protein-folding chaperones, disulfide isomerases, and genes involved in glycosylation.

The second axis of the UPR is the activating transcription factor 6 (ATF6) pathway. Under normal conditions, ATF6 exists as a transmembrane protein spanning the ER. Upon ER stress, ATF6 is translocated to the Golgi where it gets cleaved, releasing a fragment (ATF6f) that acts as a transcription factor. ATF6f upregulates proteins associated with the ER-associated degradation (ERAD) pathway that mediates degradation of misfolded proteins in the ER[4].

The third and final axis of the UPR is via PKR-like ER kinase (PERK). PERK is a transmembrane kinase, with the kinase domain in the cytoplasm, and the luminal domain bound to BIP. In response to ER stress, BIP dissociates from PERK, and PERK dimerizes and undergoes phosphorylation, becoming an active kinase. The kinase directly phosphorylates eIF2α on serine 51[5], resulting in inhibition of translation through a mechanism described in section 1.2.1. It has been shown that this process is very rapid, allowing for a decrease in global translation within 30 minutes following ER stress[6]. Phosphorylation of eIF2α also results in the upregulation of ATF4, another transcription factor that targets genes involved in amino acid transport and metabolism as well as resistance to oxidative stress[7]. ATF4 is paradoxically upregulated in response to eIF2α because it contains multiple uORFs (described in section 1.2.3), one of which overlaps with the canonical ORF. By reducing overall translation, ribosomes bypass these uORFs, and instead translate the ORF coding for ATF4, increasing overall expression of ATF4[8].

Ultimately, the hallmark of the UPR is the upregulation of key proteins associated with protein folding and degradation in the background of global translation repression.

By looking for this characteristic translational response, I can validate the robustness of our multiplexed quantitative BONCAT method.

3.2 Methods

3.2.1 Cell Culture

MCF10a cells were cultured in DMEM:F12 media supplemented with 5% horse serum, 20 ng/mL EGF, 500 µg/mL hydrocortisone, 100 ng/mL cholera toxin, 10 µg/mL insulin, 1% penicillin/streptomycin, and 2 mM glutamine. Cells were passaged every third day at a 1:4 ratio, and all experiments were performed on the third day following passaging.

3.2.2 Drug Treatment and Metabolic Labeling

30 minutes prior to tunicamycin treatment, growth media was removed and replaced with lysine/arginine/methionine (KRM)-free DMEM:F12 media. Tunicamycin was added to 10 µg/mL or an equivalent volume of DMSO was added as a control. At the appropriate time points following treatment, Aha was added to 3 mM, and $^{15}N_4{}^{13}C_6$ arginine (R10) and $^{15}N_2{}^{13}C_6$ lysine (K8) were added to 0.5 mM. After 30 minutes of Aha/K8/R10 labeling, the media was aspirated, and the cells were washed in ice cold PBS supplemented with 300 µg/mL cycloheximide (CHX). Cells were lysed in 1% SDS in PBS supplemented with 50 mM N-ethylmaleimide (NEM) and 300 µg/mL CHX. 1.5 mL of -20°C acetone was added to each tube immediately following lysis to precipitate proteins. Proteins were precipitated at -20°C for at least an hour.

3.2.3 Western Blots

Western blotting was performed as described in section 2.2.3. In addition to the reagents outlined previously, the following primary antibodies were used: RPL13a 1:1000 (Cell Signaling #2765) and GRP78 1:1000 (Cell Signaling #3177).

3.2.4 RNA Sequencing

Cells were subjected to identical KRM-free media pretreatment, Aha/K/R labeling, and PBS+CHX washing conditions described previously to account for any effects that may be caused by these treatments. Cells were lysed in 250 µL TRIzol reagent (ThermoFisher #15596018) supplemented with 300 µg/mL CHX. 50 µL chloroform was added to each lysate, vortexed, and centrifuged for 20 minutes at maximum speed (~13000 x g) for 20 minutes at 4°C. The top aqueous layer was transferred to a new tube. Total RNA was purified using the QIAgen RNEasy Mini kit (QIAgen #74104) according to the manufacturer's protocol and including a DNase digest. Samples were submitted to the MIT BioMicroCenter facility for library preparation and sequencing. The library was prepared using using Illumina NeoPrep System and sequenced on an Illumina MiSeq2000 instrument with 40 nt single-end sequencing and 6 nt barcodes. A single replicate of RNA-Seq was performed for this application.

3.2.5 Sample Preparation and MS Analysis

Samples were processed, enriched, washed, digested, TMT labeled, and fractionated as described in the protocol in appendix 2.6. The pooled fractions were placed in a ThermoFisher Easy nLC 1000 autosampler and analyzed on a ThermoFisher

QExactive Plus mass spectrometer. The instrument was operated in data dependent acquisition mode, with the top 15 most abundant precursors with charge of +2 or greater selected for fragmentation and dynamic exclusion set to 15 s. Precursors were isolated with a window of 0.4 *m/z* and fragmented via HCD at 33 NCE.

The Thermo .raw files were searched on MASCOT version 2.4 with fixed modifications for NEM alkylation on cysteines (+125.047 Da), addition of TMT 6-plex to N-termini and lysine residues (+229.163 Da). Variable modifications were SILAC R10 on arginine residues (+10.008 Da), SILAC K8 on lysine residues (+8.014 Da), addition of TMT 6-plex to SILAC K8 lysine residues (+237.177 Da), Aha substitution for methionine residues (-4.986 Da), diaminobutyrate (reduced Aha) substitution for methionine residues (-30.976 Da), oxidation of methionine residues (+15.995 Da) and phosphorylation on tyrosine, threonine, and serine residues (+79.966 Da). Precursor tolerance was 10 ppm, fragment tolerance was 15 mmu, two missed cleavages were allowed, and the enzyme was set to trypsin. Peptides were considered to be positively identified if they had a score of at least 25 and newly translated if they contained a SILAC or Aha residue. For analysis, all counts were normalized to the DMSO 30 minute channel and log2 transformed.

3.2.6 Clustering with Self-Organizing Maps

A self-organizing map (SOM) was used to cluster proteins that exhibited similar protein synthesis dynamics following tunicamycin treatment. Analysis was performed using the SOM Toolbox Matlab package (http://www.cis.hut.fi/projects/somtoolbox). A 5-by-5 neural network was initiated with hexagonal lattice structure. The input was the log-2 fold-change in protein synthesis following tunicamycin treatment relative to DMSO

controls for each time point. The network was randomly initiated and used Euclidean distance as the metric for classifying proteins to specific neurons. The SOM algorithm was repeated 1000 times, and a co-clustering map was generated indicating the frequency with which any two proteins clustered in the same neuron. This co-clustering map was then subjected to hierarchical clustering using Euclidean distance as the metric for clustering proteins.

3.2.7 Gene Ontology (GO) Term Enrichment Analysis

GO term enrichment was performed using the PANTHER classification system version 12.0[9]. Cluster members were queried against the background of all proteins included in the clustering analysis. P-values were generated from PANTHER.

3.3 Results

Following tunicamycin treatment, Aha/K8/R10 was applied to the cells at 0, 60, 120, 180, and 240 minutes for a duration of 30 minutes prior to lysis (Figure 3-1, top). In order to account for the effects of the KRM-free media starvation and Aha labeling (Figure

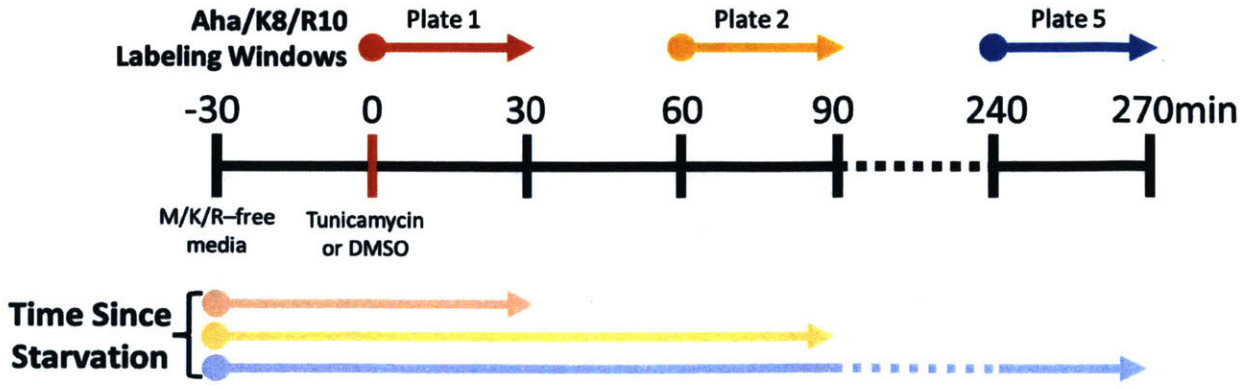


Figure 3-121 – Following a 30 minute M/K/R starvation, cells were treated with 10 µg/mL tunicamycin and incubated with Aha and SILAC amino acids in 30 minute intervals to label newly synthesized proteins.

3-1, bottom), matched DMSO controls were also collected at the same time points. By labeling in sequential 30 minute windows rather than one long 4.5 hour window, the temporal resolution of protein synthesis measurements is greatly increased, enabling differentiation between proteins synthesized in the first 30 minutes from those synthesized at later time points.

Following processing and MS analysis, 3178 unique SILAC labeled proteins were observed across all three replicates, with 2007 unique proteins appearing in at least two of the three replicates. Only proteins appearing in at least two replicates were retained for subsequent analyses. When comparing the log2 transformed medians, a global decrease in protein synthesis was observed following tunicamycin compared to DMSO controls starting between 30 and 90 minutes (Figure 3-2). This observation was consistent with reports that global translation decreases due to phosphorylation of eIF2α around 30 minutes following ER stress[6].

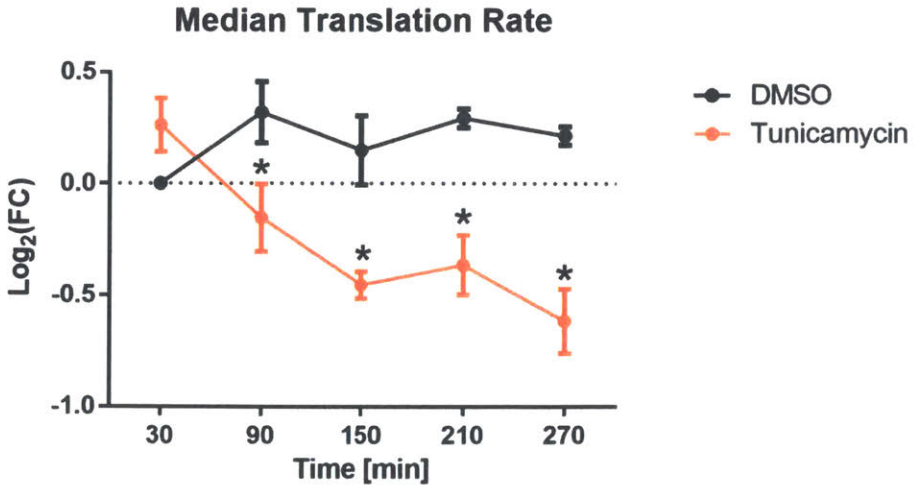


Figure 3-222 – The median translation rate decreased around 90 minutes following tunicamycin treatment (* = $p < 0.05$).

To determine whether particular sets of proteins exhibited temporally distinct behaviors, all proteins were subjected to clustering via SOMs as described in section 3.2.6. Because SOMs use a random initial seed, SOM analysis was repeated 1000 times and used to generate a co-clustering frequency map. The co-clustering map was subjected to hierarchical clustering in order to group proteins that most often cluster together. Following this analysis, two distinct clusters were observed: one characterized by protein synthesis downregulated below the median (cluster A), and one characterized by protein synthesis upregulated above the median (cluster B) (Figure 3-3).

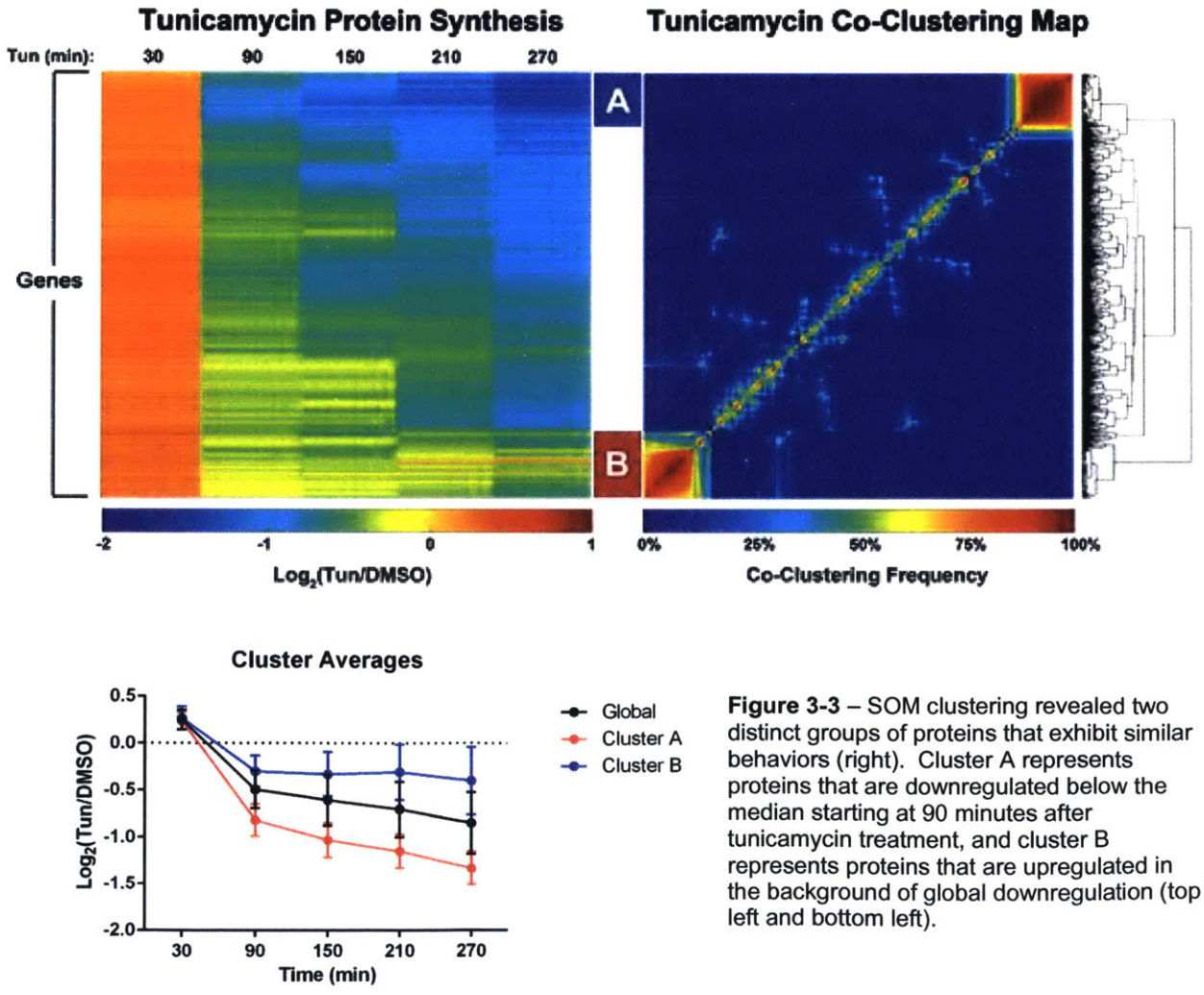


Figure 3-3 – SOM clustering revealed two distinct groups of proteins that exhibit similar behaviors (right). Cluster A represents proteins that are downregulated below the median starting at 90 minutes after tunicamycin treatment, and cluster B represents proteins that are upregulated in the background of global downregulation (top left and bottom left).

Gene ontology (GO) term analysis revealed that downregulated cluster A is highly enriched for the terms *protein targeting to the ER* (p = 2.87e-20), *translational initiation* (p = 2.80e-19), and *ribosome biogenesis* (p = 7.15e-10). The downregulation of proteins involved in translation such as ribosomal proteins, translation initiation factors, and translation elongation factors (Figure 3-4, top) may be another mechanism by which cells undergoing the UPR suppress protein synthesis.

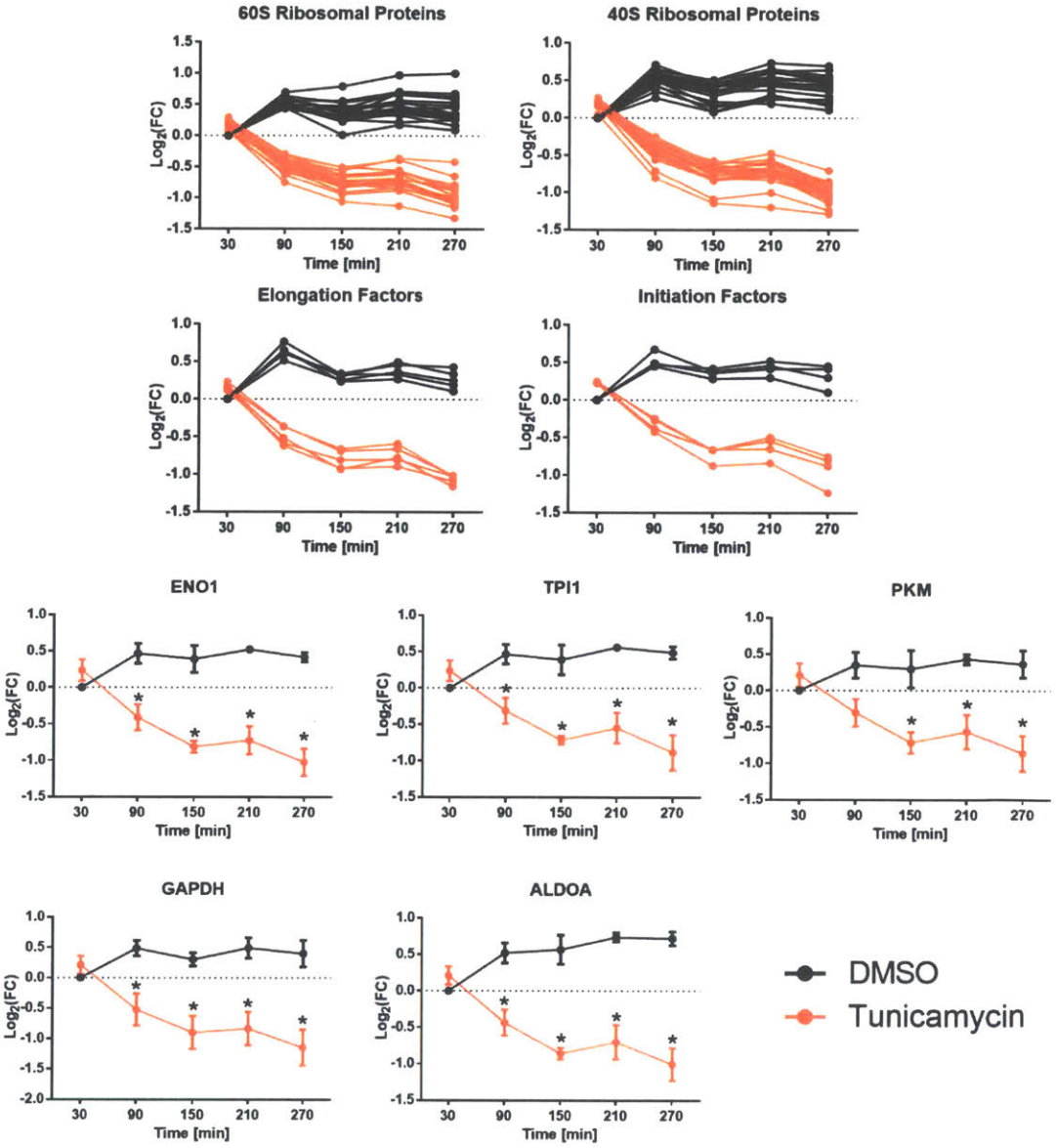


Figure 3-423 – Cluster A is enriched for proteins involved in translation, such as ribosomal proteins, initiation factors, and elongation factors (top). Cluster A is also enriched for proteins involved in glycolysis (bottom), suggesting downregulation of metabolic flux in response to ER stress.

GO enrichment analysis also showed a statistically significant enrichment for the terms *canonical glycolysis* (p = 8.54e-3), *cellular metabolic process* (p = 3.29e-3), and *ATP biosynthesis process* (p = 6.16e-3). The downregulation of proteins involved in glycolysis – including GAPDH, ENO1, TPI1, PKM, and ALDOA (Figure 3-4, bottom) – suggests that metabolic flux through the glycolytic pathway may be reduced during the UPR. It has been previously reported tunicamycin treatment reduces glucose uptake, lactate production, and ATP levels[10], consistent with the observed downregulation of proteins associated with these pathways. The mechanism behind the specific downregulation of these proteins is not known, but many of these proteins are classified as housekeeping genes that are expressed at high levels across tissue types[11]. Perhaps because these proteins were expressed at such high levels, global inhibition of translation resulted in a greater degree of downregulation of these proteins compared to other more highly regulated proteins.

Cluster B contained proteins that are downregulated to a lesser extent, and also included proteins that are even upregulated in the background of global translation repression. GO term analysis shows that this cluster was enriched for the terms *endoplasmic reticulum unfolded protein response* (p = 5.82e-4), *response to endoplasmic reticulum stress* (p = 7.09e-3), and *cellular response to unfolded protein* (p = 5.82e-4). One of the most highly upregulated proteins in this cluster was BIP, which is the canonical ER stress response chaperone. Another protein that exhibited a large increase is HERPUD1, which is involved in targeting proteins for degradation via the ERAD pathway. This cluster also included proteins involved in protein folding (HSP90B/Endoplasmin), disulfide bridge formation (PDIA4), and survival (HYOU1) (Figure 3-5). Interestingly, I

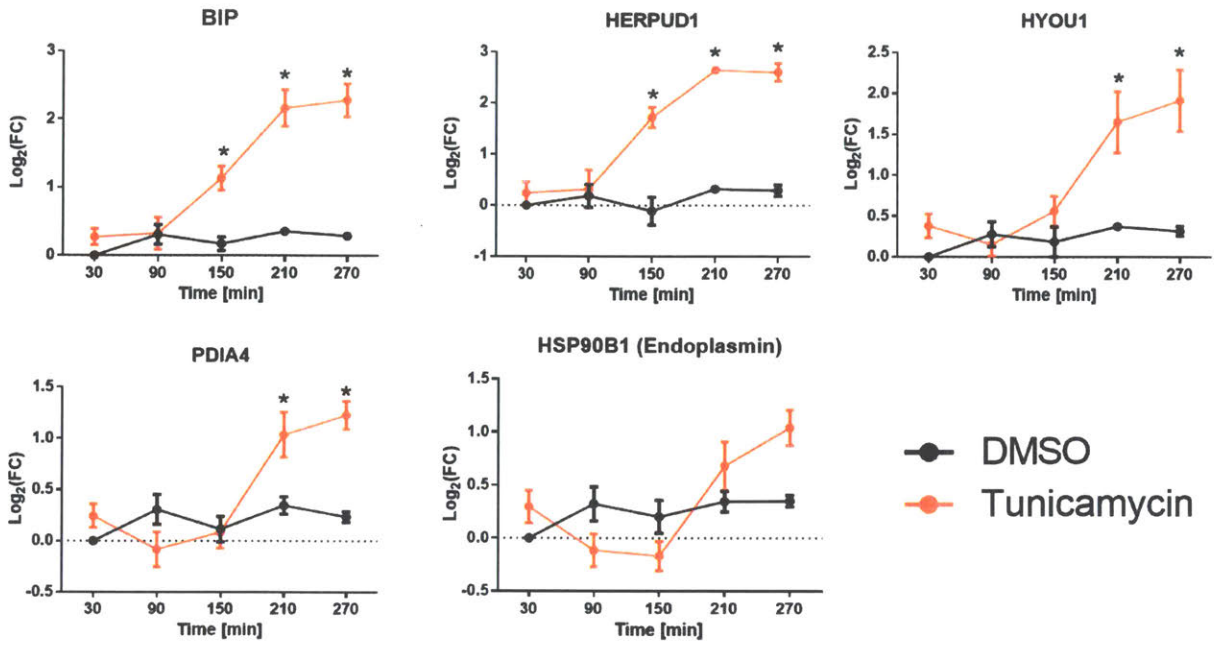


Figure 3-5 – Cluster B contained proteins that are upregulated in response to ER stress, including protein folding chaperones (HSPA5 and HSP90B1), disulfide isomerases (PDIA4), components of the ERAD (HERPUD1), and stress survival (HYOU1). * = $p < 0.05$

also observed translation initiation factors eIF5 and DAP5 (also known as eIF4G2) in this upregulated cluster as well. EIF5 is the GTPase that hydrolyzes eIF2-bound GTP to GDP, resulting in the release of eIF2 from the preinitiation complex. EIF5 has also been shown to be induced in response to genotoxic stress[12]. DAP5 shares homology to eIF4G and binds the 5'-cap of mRNA transcripts, but unlike eIF4G, DAP5 acts as a translational repressor by forming complexes that are not translationally active[13]. Furthermore, DAP5 has been shown to be upregulated in response to Tunicamycin treatment via an IRES in its 5'-UTR, allowing for selective translation during stress-induced repression of cap-dependent translation[14].

Proteins that did not fall into either of these clusters still showed a decrease in protein synthesis consistent with the global median, but varied in times of downregulation (Figure 3-3). Interestingly, members of the tricarboxylic acid (TCA) cycle fell within this

group, indicating that large decreases in metabolic protein synthesis was limited to those involved in glycolysis. However, studies examining metabolic flux 48 hours after tunicamycin treatment reveal a decrease in both glycolysis and the TCA cycle[10]. These conflicting reports may be reconciled if TCA cycle downregulation occured at later time points not assayed in this experiment.

Unique peptides from the same protein should show similar quantitation because they are typically present in equimolar ratios. However, sometimes peptides from a single protein are observed to have divergent behaviors. This divergence can occur via multiple mechanisms, including when one peptide contains a posttranslational modification. If a protein is phosphorylated in addition to being upregulated, then the quantification of most unmodified peptides would reflect this upregulation. However, peptides that contain the residue where phosphorylation occurs may not have a quantitation consistent with the others because in addition to upregulation, a proportion the peptide would be chemically modified and therefore a different chemical species. Other potential reasons for discrepancies between two peptides from the same protein may be alternative splicing and multiple isoforms of the same protein.

In this dataset, two unique peptides of XBP1 were observed that show different dynamics. Across two different replicates, one peptide (VVVAAAPNPADGTPK) showed robust upregulation in response to tunicamycin treatment, consistent with reported upregulation through IRE1α-mediated splicing of XBP1 mRNA[3]. However, a second peptide (GA**S**PEAASGGLPQAR) did not show the same upregulation (Figure 3-6). Both peptides can reside on either XBP1s or XBP1u; they are not in the intron of XBP1u that

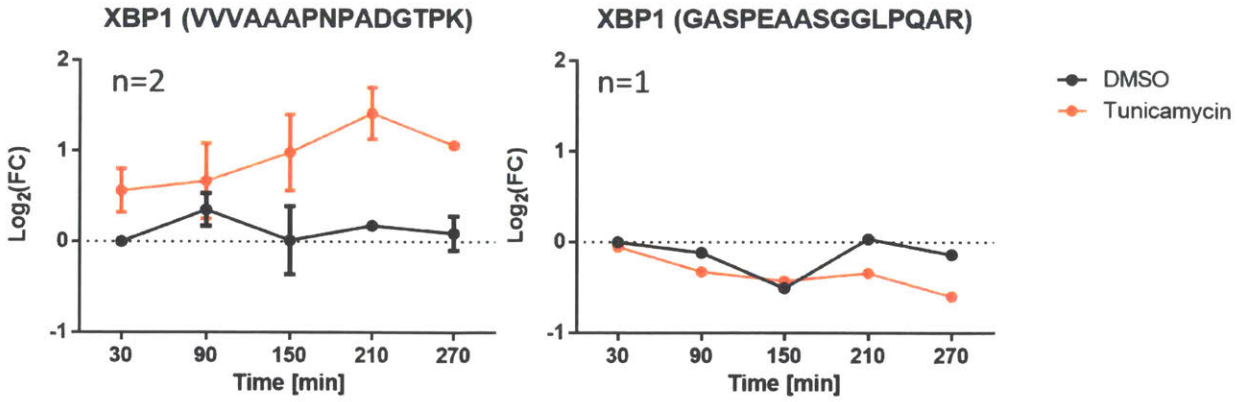


Figure 3-6 – XBP1 is known to be upregulated in response to ER stress through a non-canonical splicing of its mRNA by IRE1α. This upregulation was observed in one peptide across multiple replicates (left), but not in a second peptide (right). This discrepancy may be explained by a phosphorylation site present in the second peptide.

is spliced out by IRE1α. PhosphositePlus (www.phosphosite.org) indicates that serine 47 of XBP1 residing on this peptide (bolded above) can be phosphorylated, but the function of this modification is unknown. This phosphorylation may explain the divergent behavior between these two peptides from XBP1. Divergences between multiple peptides from the same peptide may be used to identify novel post-translational modifications, splice variants, and isoforms of a protein.

One of the advantages of BONCAT is the ability to observe changes in protein synthesis regardless of the amount of the given protein present in the cell (e.g., the size of the existing protein pool). Traditional housekeeping proteins such as GAPDH and ribosomal proteins are usually present in high quantities, buffering their levels to changes in synthesis and degradation. Indeed, Western blots against GAPDH and RPL13a did not demonstrate a change in total protein levels despite the large decrease in synthesis observed following Tunicamycin treatment. This lack of detectable change suggests that the degradation rate of these proteins (which was not measured) was small compared to the existing pool. This observation was in contrast to a blot of BIP which shows a marked

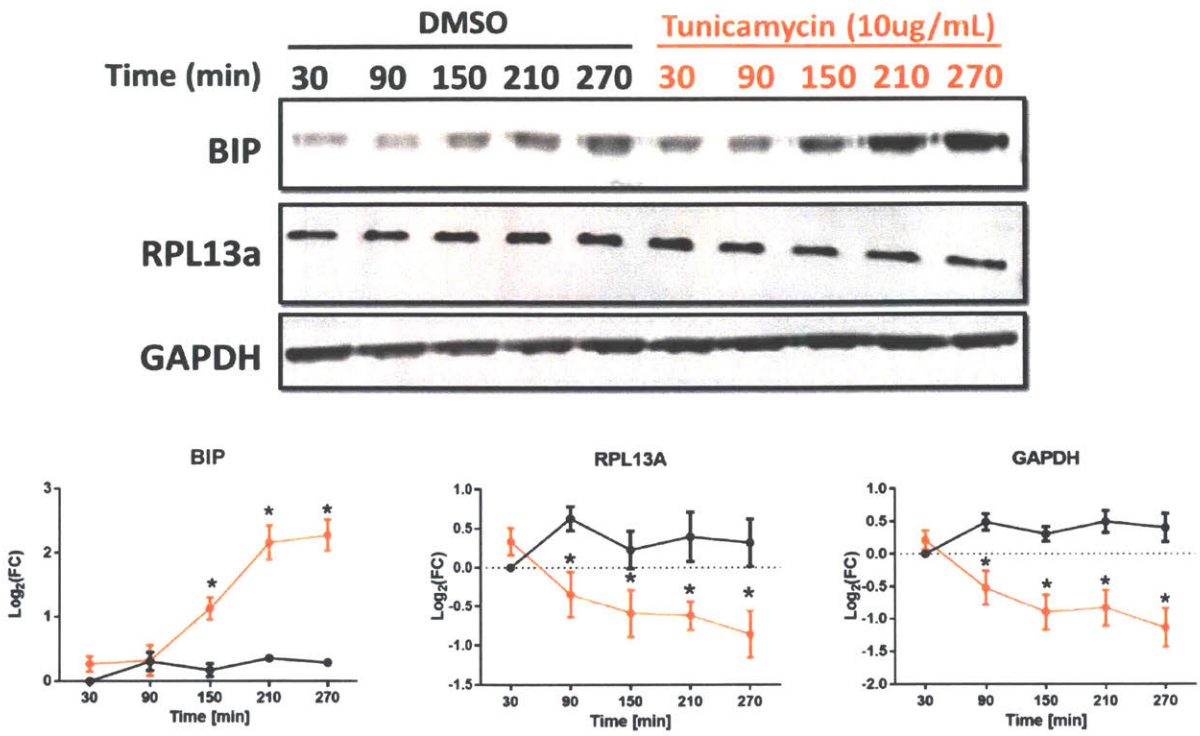


Figure 3-7 - Changes measured by BONCAT may not be observable by measuring total protein amount. For proteins with large pools of existing protein, changes in synthesis may not be reflected in the total protein abundance (GAPDH, RPL13a). Proteins with smaller existing pools showed changes in response to changes in synthesis (GRP78).

increase in protein levels following tunicamycin treatment (Figure 3-7). Because the increase in synthesis rate of BIP was large compared to the existing pool of BIP, the change in total BIP levels was observable by measuring the total protein pool. Increases in protein levels could also be bolstered by decreases in protein degradation, a process not measured by this technique.

The downregulation of protein synthesis during the UPR is mediated by phosphorylation of eIF2α by PERK, resulting in regulation at the level of translation. To investigate whether mRNA levels correlated with observed changes in protein synthesis, RNA-Seq was performed to quantify mRNA expression in response to tunicamycin treatment. At the global scale, mRNA abundance decreased modestly between 0 and 90 minutes following tunicamycin treatment, and then increased back to initial levels by 150

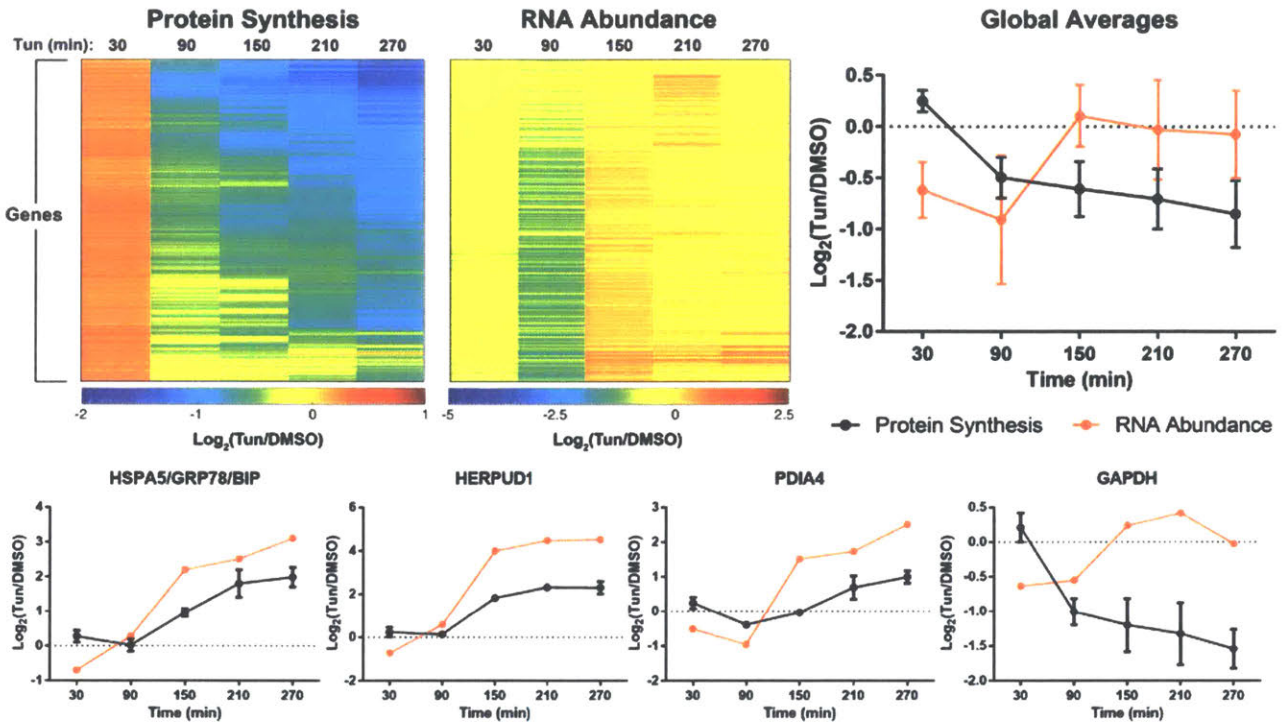


Figure 3-8 – Despite a robust downregulation of protein synthesis following Tunicamycin treatment, there was not a downregulation of mRNAs on the level of transcription, demonstrating repression on the level of translation (top). Proteins that were upregulated from cluster B show a matched increase in RNA abundance, but the same behavior was not observed by proteins outside the cluster, exemplified by GAPDH (bottom).

minutes. By comparison, total protein synthesis was downregulated by 90 minutes, and continued to be repressed for the duration of the time course (Figure 3-8, top). Notably, mRNA abundance for the upregulated proteins in cluster B was also increased, suggesting that these proteins were regulated on the level of transcription (Figure 3-8, bottom). On a global level, the difference between transcription and protein synthesis supports a model of regulation at the level of translation in response to ER stress.

3.4 Conclusion

The unfolded protein response is a well-studied phenomenon characterized by the global downregulation of protein translation and the upregulation of key stress response proteins, making it an excellent system to test my multiplexed quantitative BONCAT

protocol. Newly synthesized proteins were sampled for 30 minutes every hour following UPR induction via tunicamycin treatment. Indeed, global downregulation of protein synthesis was observed that was not measurable via RNA-Seq, which is the current standard for the high throughput measurement of changes in gene expression. Clustering of the temporal response of protein synthesis to tunicamycin treatment reveals two major clusters: one downregulated far below the median, and one upregulated compared to the median. The downregulated cluster contained proteins involved in the translational machinery and glycolysis, consistent with the downregulation of translation and metabolic flux associated with a stress response. These proteins largely consisted of traditional housekeeping proteins (*e.g.* ribosomal and metabolic proteins) that are expressed constitutively at high levels. Therefore, global inhibition of translation may have caused synthesis of these proteins to decrease further than others. Interestingly, members of the TCA cycle did not fall within this cluster, indicating that metabolic flux is rapidly regulated in response to ER stress by a decrease in glycolysis. Furthermore, I was able to measure an increase in expression of key protein folding chaperones and stress response proteins present in the upregulated cluster. Altogether, these data demonstrate the reliability of this new protocol for quantifying changes in protein synthesis rates in a complex biological system.

3.5 References

1. Hetz, C. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nat. Rev. Mol. Cell Biol.* **13,** 89–102 (2012).
2. Hetz, C., Martinon, F., Rodriguez, D. & Glimcher, L. H. The Unfolded Protein Response: Integrating Stress Signals Through the Stress Sensor IRE1. *Physiol. Rev.* **91,** 1219–1243 (2011).
3. Yoshida, H., Matsui, T., Yamamoto, A., Okada, T. & Mori, K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* **107,** 881–891 (2001).
4. Yoshida, H. *et al.* ATF6 activated by proteolysis binds in the presence of NF-Y (CBF) directly to the cis-acting element responsible for the mammalian unfolded protein response. *Mol. Cell. Biol.* **20,** 6755–67 (2000).
5. Harding, H. P., Zhang, Y. & Ron, D. Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature* **397,** 271–274 (1999).
6. Novoa, I. *et al.* Stress-induced gene expression requires programmed recovery from translational repression. *EMBO J.* **22,** 1180–1187 (2003).
7. Harding, H. P. *et al.* An Integrated Stress Response Regulates Amino Acid Metabolism and Resistance to Oxidative Stress. *Mol. Cell* **11,** 619–633 (2003).
8. Harding, H. P. *et al.* Regulated Translation Initiation Controls Stress-Induced Gene Expression in Mammalian Cells. *Mol. Cell* **6,** 1099–1108 (2000).
9. Mi, H. *et al.* PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45,** D183–D189 (2017).
10. Wang, X. *et al.* ER stress modulates cellular metabolism. *Biochem. J* **435,** 285–296 (2011).
11. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends in Genetics* **29,** 569–574 (2013).
12. Sheikh, M. S. *et al.* Identification of several human homologs of hamster DNA damage-inducible transcripts. Cloning and characterization of a novel UV-inducible cDNA that codes for a putative RNA-binding protein. *J. Biol. Chem.* **272,** 26720–6 (1997).
13. Imataka, H., Olsen, H. S. & Sonenberg, N. A new translational regulator with homology to eukaryotic translation initiation factor 4G. *EMBO J.* **16,** 817–825 (1997).
14. Lewis, S. M. *et al.* The eIF4G homolog DAP5/p97 supports the translation of select mRNAs during endoplasmic reticulum stress. *Nucleic Acids Res.* **36,** 168–178 (2008).

Chapter IV

Investigating Translation Dynamics Following EGF Stimulation

Daniel Rothenberg
Matthew Taliaferro
Sabrina Huber

4.1 Introduction

Cells respond to their environments by detecting extracellular cues, transmitting information through signaling pathways, and integrating these signals into a response that induces phenotypic changes to ensure their survival. Receptor tyrosine kinases (RTKs) play an important role in initiating such responses through binding external ligands and triggering phosphorylation cascades[1]. Epidermal growth factor (EGF) is an example of an RTK-binding extracellular ligand that is important in cell growth and proliferation[2]. The binding of EGF to its RTK (EGFR) causes the dimerization of the receptor and transactivation of the intracellular kinase domains[3]. The kinase domains of EGFR undergo autophosphorylation on tyrosine residues within the C-terminal tails, creating docking sites for adaptor proteins and SOS, the GEF for the GTPase RAS. Upon docking with the EGFR-bound complex, SOS facilitates the release of GDP from RAS so that it can be replaced with a GTP, resulting in activated RAS and signaling through the RAF/MEK/ERK or PI3K/AKT pathways[4]. Both of these pathways promote cell growth, proliferation, and survival, and are frequently hyper-activated in different types of cancer[5,6].

Because EGFR is one of the RTKs that sits atop key signaling pathways involved in cancer development, many cancers arise through increased expression or mutations of EGFR[7]. There have been many studies analyzing the effects of EGFR activation, examining both the relatively short-term signaling effects as well as the longer-term changes in gene expression. Studies analyzing signaling downstream of EGFR activation have examined phosphotyrosine modifications of adaptor proteins, kinases, and other downstream effectors within hours, minutes, and even seconds following EGF treatment[8–

[10]. Other studies have sought to determine the result of these signaling cascades and characterize which genes are expressed following EGFR activation. Unlike signaling activation, in which entire pathways can be activated on the order of seconds to minutes, gene expression is thought to be a slower process with many regulated steps. Signals must be transduced and integrated by transcription factors. Transcription initiation complexes must be assembled, and the mRNA transcript must be transcribed, processed, and spliced. The translation preinitiation complex must be assembled, and the ribosome must translate the mRNA into a protein. Altogether, gene expression is usually measured on the order of minutes to hours following a perturbation.

Prior studies profiled gene expression in response to EGFR activation using RNA-Seq and protein abundance as readouts[11,12]. RNA-Seq analysis following EGFR activation revealed temporally distinct waves of transcription. Immediate early genes (IEGs) showed upregulated transcripts as early as 45 minutes following EGF stimulation. IEGs consist of pro-growth transcription factors including EGR1, MYC, FOS, and JUN[13]. Following IEG activation, delayed early genes (DEGs) are upregulated between 90 and 120 minutes and serve as negative feedback regulators for IEGs, resulting in a sharp attenuation of IEG transcription. Examples of DEGs include DUSP1, a phosphatase targeting ERK[14], and ZFP36, which recruits mRNA-degrading enzymes to IEG transcripts[15]. Finally, DEGs also activate late response genes (LRGs), which are upregulated 2 to 4 hours after stimulation. LRGs are responsible for modulating cell phenotypes, and include proteins that interact with the cytoskeleton and proteins involved in the epithelial to mesenchymal transition (EMT)[16,17]. IEGs and DEGs are characterized by not requiring the synthesis of new proteins to induce their expression, whereas LRGs

(also known as secondary response genes), require an intermediate protein to be synthesized in order to be upregulated[18].

As discussed in chapter 1, RNA-Seq analysis does not account for many of the regulatory steps between transcription and translation of a new protein. Furthermore, ribosome footprinting also may not completely capture the spectrum and timing of new protein synthesis[19]. In this chapter, multiplexed quantitative BONCAT will be applied to EGFR activation to characterize protein synthesis following activation pro-growth pathways. Furthermore, I will directly compare protein synthesis as measured by BONCAT to RNA-Seq and ribosome footprinting data to look for novel insights gained through the BONCAT approach.

4.2 Methods

4.2.1 Cell Culture

HeLa cells were cultured in DMEM supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin, and 2 mM glutamine. Cells were passaged every 3 days and split between 1:4 and 1:6 ratio.

4.2.2 EGF Stimulation and Metabolic Labeling

All experiments were performed following a 24 hour serum starvation (same media as above sans FBS) initiated 2 days following passaging. 30 minutes prior to EGF stimulation, growth media was removed and replaced with lysine/arginine/methionine (KRM)-free DMEM:F12 media. EGF was added to 20 nM or an equivalent volume of PBS was added as a control. At the appropriate time points following treatment, Aha was

added to 3 mM, and $^{15}N_4{}^{13}C_6$ arginine (R10) and $^{15}N_2{}^{13}C_6$ lysine (K8) were added to 0.5 mM. After 30 minutes of Aha/K8/R10 labeling, the media was aspirated, and the cells were washed in ice cold PBS supplemented with 300 µg/mL cycloheximide (CHX). Cells were lysed in 1% SDS in PBS supplemented with 50 mM N-ethylmaleimide (NEM) and 300 µg/mL CHX. 1.5 mL of -20°C acetone was added to each tube immediately following lysis to precipitate proteins. Proteins were precipitated at -20°C for at least an hour.

4.2.3 Ribosome Footprinting and RNA Sequencing

Cells were subjected to identical KRM-free media pretreatment, Aha/K/R labeling, and PBS+CHX washing conditions as described previously to account for any effects that may be caused by these treatments. Cells were lysed and processed using the Illumina TruSeq Ribo Profile kit (Illumina #RPHMR12126) according to the manufacturer's protocol. RFP samples were sequenced on an Illumina NextSeq instrument with 50 nt single-end reads and 6 nt barcodes with 6 samples per lane. Adapter sequences were removed from the 3' end using Cutadapt. Reads were then mapped to hg38 rRNA sequences using STAR. Reads that aligned to rRNA sequences were removed. The remaining reads were then mapped to an hg38 annotation (Gencode release 26). Reads that uniquely map to this annotation were then quantified using Salmon and a FASTA file containing all hg38 cDNA sequences to generate transcripts per million (TPM) values and counts. A library for total RNA was prepared using the Illumina NeoPrep System and sequenced on an Illumina NextSeq instrument with 40 nt paired-end reads and 6 nt barcodes with 12 samples per lane. For analyses requiring alignment, these reads were mapped to the same hg38 (Gencode release 26) annotation. Quantification was done

using Salmon and a FASTA file containing all hg38 cDNA sequences to generate TPM and count values. Ribosome profiling TE changes and associated p-values were calculated using the Salmon-derived count data and the Xtail package[20].

4.2.4 Analysis of tRNA Modifications

Purification of tRNA from total RNA was carried out on a 1200 Agilent HPLC system fitted with an Agilent size-exclusion SEC3 column (7.8 x 300 mm, 3.0 µm particle size, 300 Å pore size). The tRNA was eluted under isocratic conditions at 60 °C with 100 mM ammonium acetate and a flow rate of 0.5 mL/min. The collected tRNA fractions were concentrated to about 200 µL by vacuum centrifugation. After addition of 0.1 volumes of 3M sodium acetate and 2.5 volumes ethanol, tRNA was precipitated overnight at -20 °C. The tRNA was pelleted at 12,000 x g for 30 min at 4 °C, washed with 75% ethanol, air-dried and resuspended in water. A total of 6 µg of tRNA per sample was digested using the following master mix: 2.5 µL 100 mM $MgCl_2$, 5 µL 1M Tris pH 8, 0.5 µL 0.1 mg/mL Coformycin, 0.75 µL 100 mM Deferoxamine, 0.15 µL 100 mM butylated hydroxytoluene, 5 µL 1 µM ^{15}NdA, 0.85 µL water, 2.25 µL 5 U/µL Benzonase, 1.5 µL 17 U/µL calf intestinal alkaline phosphatase, 1.5 µL 0.1 U/µL phosphodiesterase I. Aqueous solutions of tRNA (6 µg in 30 µL water) were mixed with digestion enzyme master mix (20 µL) and incubated at 37 °C for 3 h. The nucleoside mixture was subsequently cleaned-up by filtration over 10 kDa MWCO spin columns. Ribonucleosides were resolved with a Phenomenex Synergi Fusion-reverse phase column (100 x 2 mm, 2.5 µm particle size, 100 Å pore size) eluted with the following gradient of acetonitrile in 5mM ammonium acetate (pH 5.3) at a flow rate of 0.35 ml/min and 35 °C: 0-1 min, 0%; 1-10 min, 0-10%, 10-14 min, 10-40%,

14-15min, 40-80%. The HPLC column was coupled to an Agilent 6430 Triple Quadrupole LC/MS spectrometer with an electrospray ionization source operated in positive ion mode with the following parameters for voltages and source gas: gas temperature, 350 °C; gas flow, 10 L/min; nebulizer, 45 psi; and capillary voltage, 3500 V. The first and third quadrupoles (Q1 and Q3) were fixed to unit resolution and the modifications were quantified by pre-determined molecular transitions. The dwell time for each ribonucleoside was 500 ms. The retention time, *m/z* of the transmitted parent ion, *m/z* of the monitored product ion, fragmentor voltage, and collision energy of each modified nucleoside are as follows: ac4C, 6.5 min, *m/z* 286 → 154, 80 V, 6 V; acp3U, 1.7 min, *m/z* 346 → 214, 80 V, 10 V; Am, 8.5 min, *m/z* 282 → 136, 100 V, 15 V; Cm, 4.2 min, *m/z* 258 → 112, 80 V, 8 V; D, 1.5 min, *m/z* 247 → 115, 80 V, 5 V; Gm, 6.2 min, *m/z* 298 → 152, 80 V, 7 V; I, 4.4 min, *m/z* 269 → 137, 80 V, 10 V; i6A, 14 min, *m/z* 336 → 204, 100 V, 17 V; m1A, 3.6 min, *m/z* 282 → 150, 100 V, 16 V; m1G, 6.1 min, *m/z* 298 → 166, 90 V, 10 V; m1I, 5.9 min, *m/z* 283 → 151, 80 V, 10 V; m227G, 7.1 min, *m/z* 326 → 194, 80 V, 10 V; m22G, 7.8 min, *m/z* 312 → 180, 100 V, 8 V; m2A, 8.7 min, *m/z* 282 → 150, 100 V, 16 V; m2G, 6.5 min, *m/z* 298 → 166, 90 V, 10 V; m2s2i6A, 15.4 min, *m/z* 382 → 250, 100 V, 17 V; m3C, 2.6 min, *m/z* 258 → 126, 80 V, 8 V; m3U, 5.5 min, *m/z* 259 → 127, 80 V, 7 V; m5C, 3.7 min, *m/z* 258 → 126, 80 V, 7 V; m5s2U, 7.6 min, *m/z* 275 → 143, 80 V, 5 V; m5U, 4.8 min, *m/z* 259 → 127, 80 V, 7 V; m5Um, 7.3 min, *m/z* 273 → 127, 100 V, 10 V; m6A, 9.3 min, *m/z* 282 → 150, 100 V, 16 V; m6t6A, 9.8 min, *m/z* 427 → 295, 80 V, 10 V; m7G, 3.9 min, *m/z* 298 → 166, 90 V, 10 V; mcm5s2U, 8.7 min, *m/z* 333 → 201, 100 V, 10 V; mcm5U, 6.2 min, *m/z* 317 → 185, 90 V, 8 V; mcm5Um, 9 min, *m/z* 331 → 185, 90 V, 8 V; ms2t6A, 10.8 min, *m/z* 459 → 327, 100 V, 8 V; OHyW, 12.5 min, *m/z* 525 → 393,

100 V, 10 V; Q, 5.3 min, *m/z* 410 → 295, 60 V, 8 V; t6A, 8.1 min, *m/z* 413 → 281, 100 V, 8 V; Um, 5.4 min, *m/z* 259 → 113, 80 V, 7 V; Y, 1.7 min, *m/z* 245 → 191, 80 V, 10 V; yW, 13.1 min, *m/z* 509 → 377, 90 V, 10 V. Modified ribonucleosides were identified using commercially available nucleoside standards or by HPLC retention time and collision-induced dissociation (CID) fragmentation patterns. Quantitative comparisons between biological replicates from various time points were made possible by correcting for biological variation in tRNA quantities by dividing raw peak area for each ribonucleoside by the ultraviolet absorbance peak areas for the four canonical ribonucleosides. Fold changes in the levels of ribonucleosides were calculated for each time point by dividing the normalized peak areas for EGF-exposed samples by their corresponding PBS control. Significant changes were determined Student's t-test ($p \leq 0.05$).

4.2.5 Sample Preparation and MS Analysis

Samples were processed, enriched, washed, digested, TMT labeled, and fractionated as described in the protocol in appendix 2.6. Due to limitations on amount of material, 100-300 µg of total protein were used in each channel depending on the replicate. The pooled fractions were placed in a ThermoFisher Easy nLC 1000 autosampler and analyzed on a ThermoFisher QExactive Plus mass spectrometer. The instrument was operated in data dependent acquisition mode, with the top 15 most abundant precursors with charge of +2 or greater selected for fragmentation and dynamic exclusion set to 15 s. Precursors were isolated with a window of 0.4 *m/z* and fragmented via HCD at 33 NCE.

The Thermo .raw files were searched on MASCOT version 2.4 with fixed modifications for NEM alkylation on cysteines (+125.047 Da), addition of TMT 6-plex to N-termini and lysine residues (+229.163 Da). Variable modifications were SILAC R10 on arginine residues (+10.008 Da), SILAC K8 on lysine residues (+8.014 Da), addition of TMT 6-plex to SILAC K8 lysine residues (+237.177 Da), Aha substitution for methionine residues (-4.986 Da), diaminobutyrate (reduced Aha) substitution for methionine residues (-30.976 Da), oxidation of methionine residues (+15.995 Da) and phosphorylation on tyrosine, threonine, and serine residues (+79.966 Da). Precursor tolerance was 10 ppm, fragment tolerance was 15 mmu, two missed cleavages were allowed, and the enzyme was set to trypsin. Peptides were considered to be positively identified if they had a score of at least 25 and newly translated if they contained a SILAC or Aha residue.

4.2.6 K-Means Clustering

K-means clustering was used to group proteins into clusters with distinct temporal responses. Data was filtered by removing proteins that appeared in less than two out of four replicates, as well as removing proteins that did not show a statistically significant change (according to Student's T test) in synthesis between EGF and PBS controls in at least one time point. After optimizing to reduce within-cluster distance, 6 clusters were used for analysis. Cluster centroids were initialized randomly, and Pearson correlation was used as the distance metric. K-means clustering was repeated 10000 times, and a co-clustering map was generated indicating the frequency with which any two proteins shared the same cluster. This co-clustering map was then subjected to hierarchical clustering using Euclidean distance as the metric for clustering proteins.

4.3 Results

4.3.1 Analysis of EGF-Induced Protein Synthesis

Aha/K8/R10 was applied to the cells at 0, 30, 60, 90, and 120 minutes after EGF stimulation for a duration of 30 minutes (Figure 4-1, top). In order to account for the effects of the KRM-free media starvation and Aha labeling (Figure 4-1, bottom), matched PBS controls were also collected at the same time points. In all subsequent analyses, the samples were referred to by the time at which they were lysed.

Following processing and analysis, 1749 unique proteins were identified across four replicates, with 1007 observed in at least two replicates. Only proteins identified in two or more replicates were retained for subsequent analysis. Unlike the response to tunicamycin treatment (see chapter 3), there was no statistically significant change in global protein synthesis following EGF stimulation. Therefore, to control for technical variations between channels, all values were normalized to the median of each channel within each replicate.

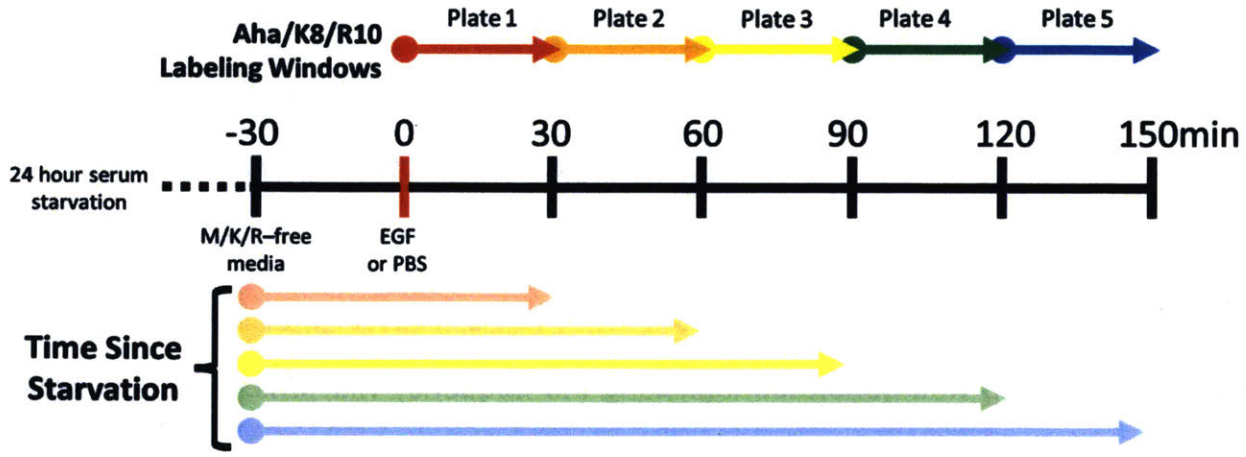


Figure 4-1 – Following a 24 hour serum starvation, cells were transitioned to KRM-free media for 30 minutes prior to stimulation with 20 nM EGF (or PBS control). Cells were labeled with Aha and SILAC amino acids in consecutive 30 minute intervals prior to lysis.

In order to group proteins into temporally distinct clusters, k-means clustering analysis was performed using 6 clusters and Pearson correlation as the distance metric. Pearson correlation was selected because proteins should be clustered based on their temporal response independent of the magnitude of their change. This analysis revealed four distinct clusters, each with a unique temporal profile (Figure 4-2, top left and bottom left).

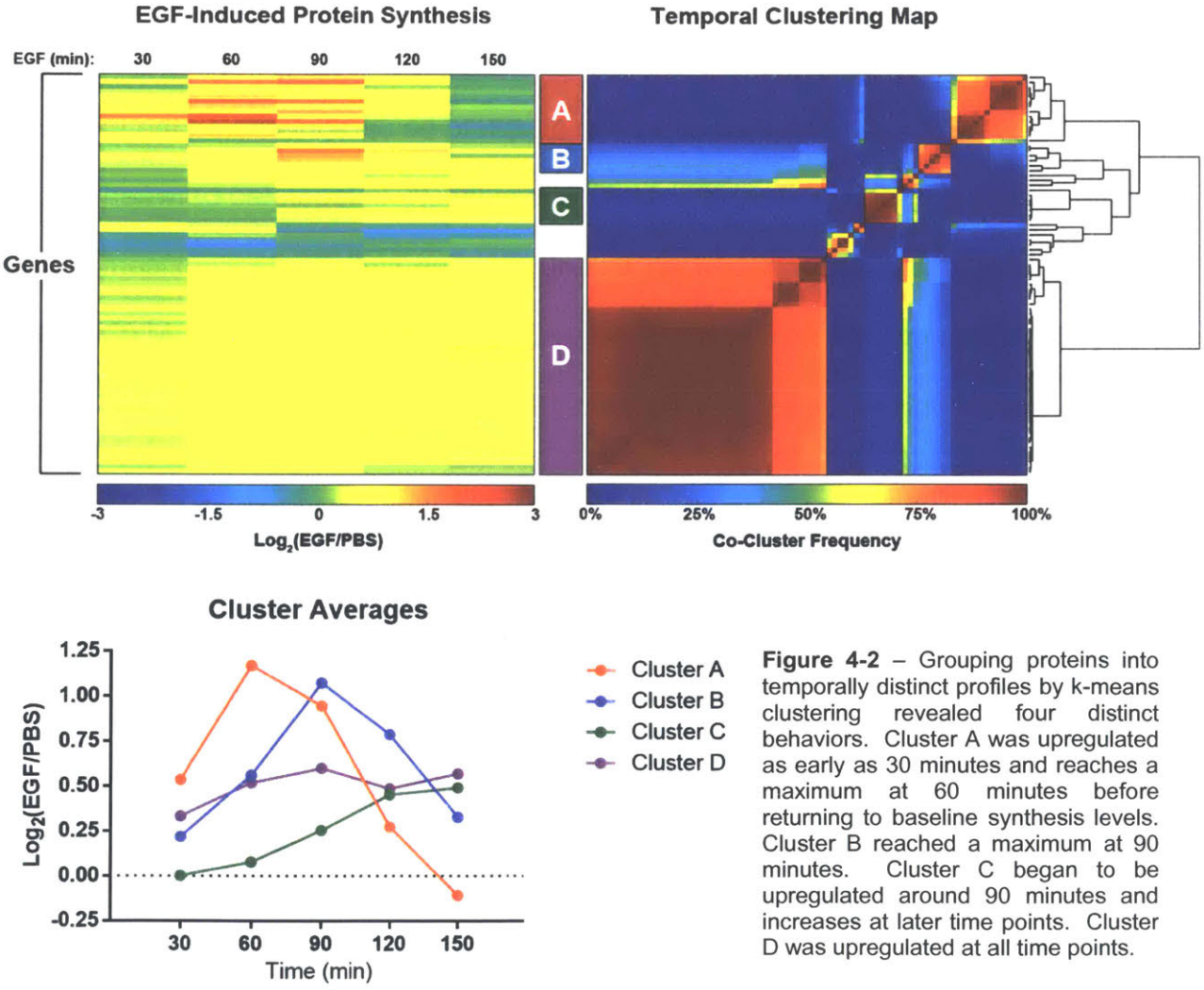


Figure 4-2 – Grouping proteins into temporally distinct profiles by k-means clustering revealed four distinct behaviors. Cluster A was upregulated as early as 30 minutes and reaches a maximum at 60 minutes before returning to baseline synthesis levels. Cluster B reached a maximum at 90 minutes. Cluster C began to be upregulated around 90 minutes and increases at later time points. Cluster D was upregulated at all time points.

Cluster A was characterized by increased protein synthesis as early as 30 minutes following EGF stimulation, and peaking at 60 minutes before returning to baseline levels by 150 minutes. This cluster contained many of the canonical IEGs, including EGR1, JUN, CYR61, and IER2, consistent with a model of rapid upregulation of these genes within the first hour following stimulus. However, this cluster also contained many of the canonical DEGs, such as DUSP1, ATF3, and ZFP36L1. This rapid upregulation of DEGs was in contrast to established literature that shows DEG transcripts being upregulated between 1 and 2 hours following EGF stimulation[16,17]. One notable IEG absent from the dataset was MYC, which has previously been shown to be upregulated on the transcript level in response to EGF stimulation[12]. MYC was modestly upregulated between 30 and 90 minutes (not shown), but not to an extent that was statistically significant in this analysis. This observation suggests that previously observed increases in MYC protein levels may be through increased stability rather than synthesis.

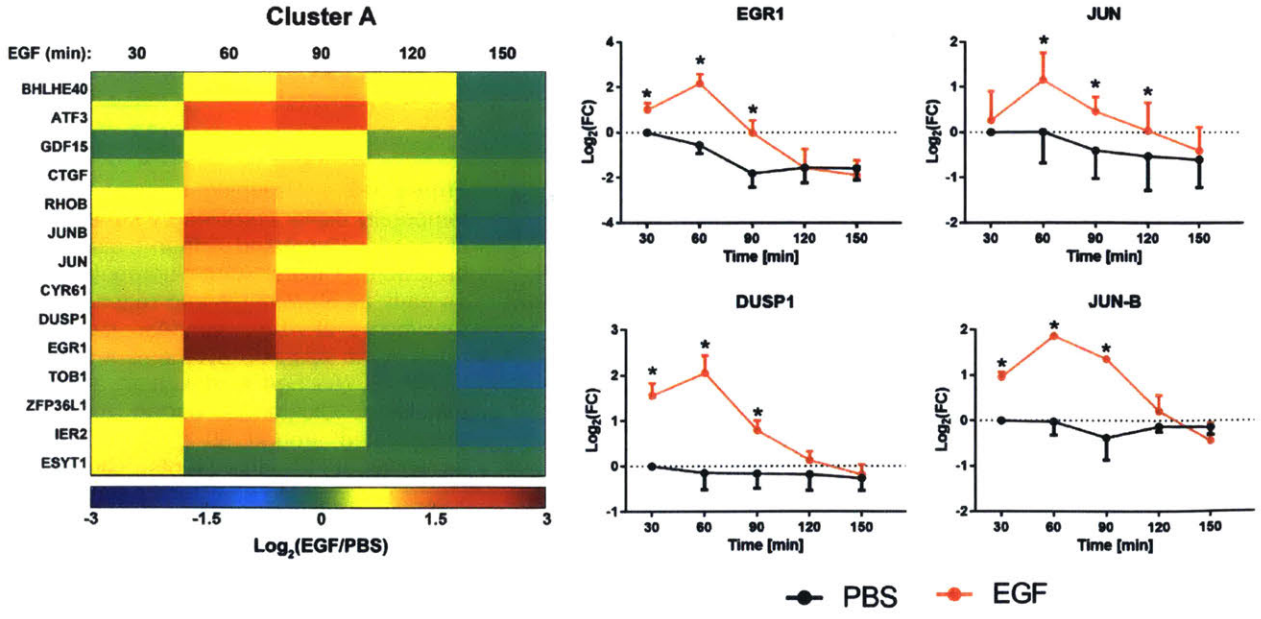


Figure 4-3 – Cluster A was characterized by increased synthesis as early as 30 minutes after EGF stimulation, with synthesis peaking around 60 minutes, and returning to basal levels by 150 minutes. This cluster contained canonical IEGs (e.g. EGR21 and JUN, top right) and DEGs (e.g. DUSP1, JUN-B, bottom right). * = $p < 0.05$

Cluster B was smaller, containing only 6 proteins, but had similar synthesis dynamics as cluster A. Proteins in cluster B began to be expressed by 60 minutes, and peak their levels at 90 minutes. Like cluster A, cluster B contained both IEGs (NR4A1 and NR4A3) and DEGs (KLF10) (Figure 4-4). The NR4A proteins are nuclear orphan receptors that have been shown to be upregulated in response to a wide variety of stimuli, including growth factor stimulation[21,22]. KLF10 is a tumor suppressor canonically upregulated in response to TGF-β treatment, but EGF treatment has also been shown to modestly increase KLF10 expression[23].

Within clusters A and B, synthesis of both IEGs and DEGs were observed between 30 and 90 minutes following EGF stimulation. Rather than a wave of IEGs followed by a wave of DEGs, IEGs and DEGs appeared to be collectively expressed within the same time frame with maximal synthesis occurring between 60 and 90 minutes following EGFR activation.

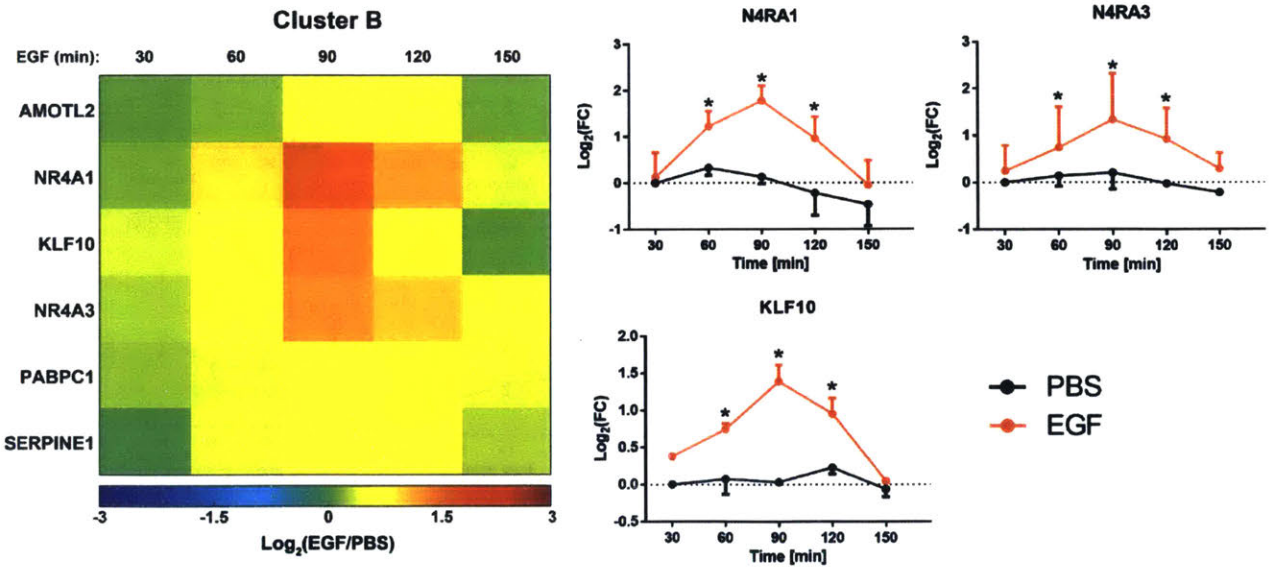


Figure 4-4 – Cluster B was characterized by increased synthesis by 60 minutes, with a peak at 90 minutes. This cluster included the IEGs NR4A1 and NR4A3 as well as the DEG KLF10. * = $p < 0.05$

Cluster C was characterized by a delayed response to EGFR activation, with synthesis beginning at around 90 minutes and increasing through 150 minutes after stimulation. This cluster contained several LRGs involved in cytoskeletal dynamics and cell motility, such as CTEN, VASP, EZR, and EPPK1 (Figure 4-5).

EZR is a membrane-associated protein that also interacts with the actin cytoskeleton, allowing for dynamic changes in cell morphology. It has been demonstrated that the AP-1 transcription factor (a heterodimer consisting of IEGs including JUN, ATF, and FOS) can activate EZR, resulting in an invasive phenotype[24]. EPPK1 has been shown to crosslink intermediate filaments of the cytoskeleton, suggesting a role in cell structure[25]. VASP is a protein that associates with the barbed ends of the actin cytoskeleton and prevents binding of capping proteins. Increased expression of VASP is consistent with processes associated with cytoskeleton dynamics, including changes in cell morphology, migration, and invasion[26].

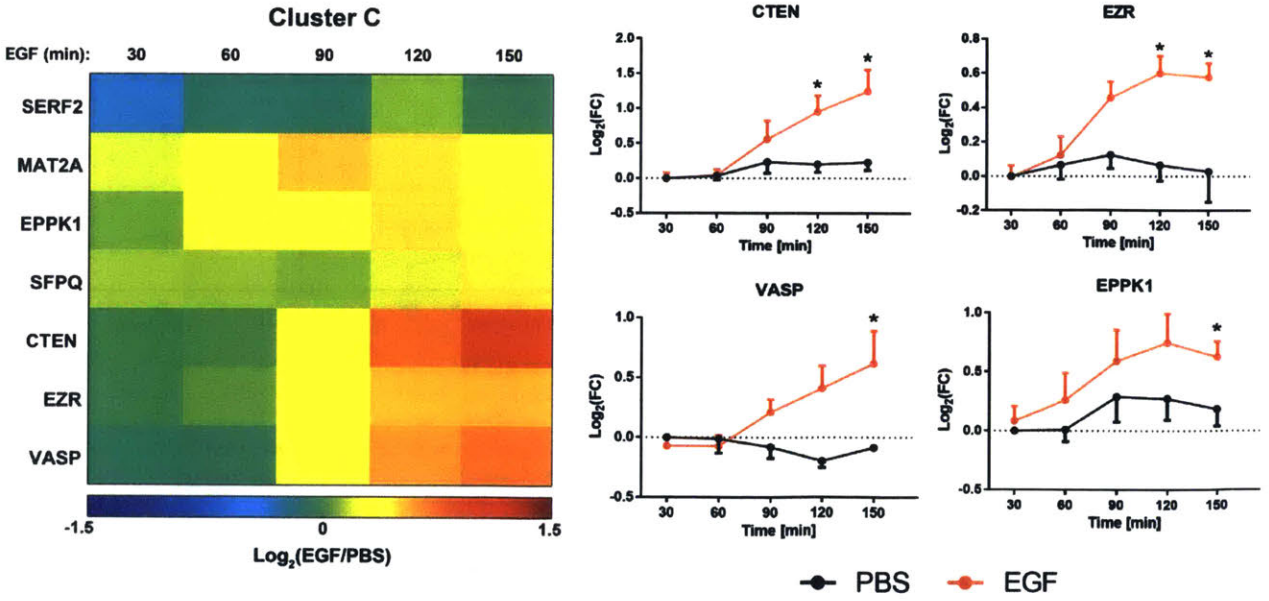


Figure 4-5 - Cluster C was characterized by increased synthesis by 90 minutes that continued through 150 minutes. This cluster included the LRGs that control cytoskeleton dynamics such as CTEN, EZR, VASP, and EPPK1. * = $p < 0.05$

CTEN is a member of the Tensin family of proteins that couple focal adhesions to the actin cytoskeleton. However, unlike other Tensins, CTEN lacks an actin binding domain. As a result, CTEN migration to focal adhesions uncouples the plasma membrane from the cytoskeleton, resulting in increased cell motility. In response to EGFR activation, CTEN has been shown to migrate to focal adhesions and bind to the cytoplasmic tails of integrin β_1, displacing TNS3 which crosslinks focal adhesions to the cytoskeleton. Furthermore, it has been observed that CTEN transcripts are upregulated following EGFR activation concurrent with TNS3 transcript downregulation[27]. Indeed, CTEN synthesis was upregulated starting around 90 minutes after EGF treatment. Furthermore, a modest decrease in TNS3 synthesis was observed starting around 120 minutes and continuing through 150 minutes (not shown). Although this TNS3 downregulation was not statistically significant during the assayed time points, continued observation of protein synthesis rates may demonstrate a further downregulation of TNS3 at later time points.

Another notable member of this cluster was methionine adenosyltransferase 2A (MAT2A), an enzyme important for S-adenosylmethionine (SAM) synthesis. SAM is the primary methyl group donor for a wide variety of cellular processes, and plays an important role in cell homeostasis. An upregulation of MAT2A was observed following EGFR activation and continued through 150 minutes (Figure 4-6). It has been demonstrated that EGF treatment can induce MAT2A expression[28], but an induction of MAT2A synthesis was observed in the PBS control as well. The procedure for labeling newly synthesized proteins with Aha included a transfer to methionine-free media (Figure

4-1), and previous work has shown that methionine depletion can induce expression of MAT2A[29]. Similar changes were also observed in the PBS controls of EGR1, and CTGF (Figure 4-6). These observations highlight the importance of including matched negative controls for every time point in this analysis.

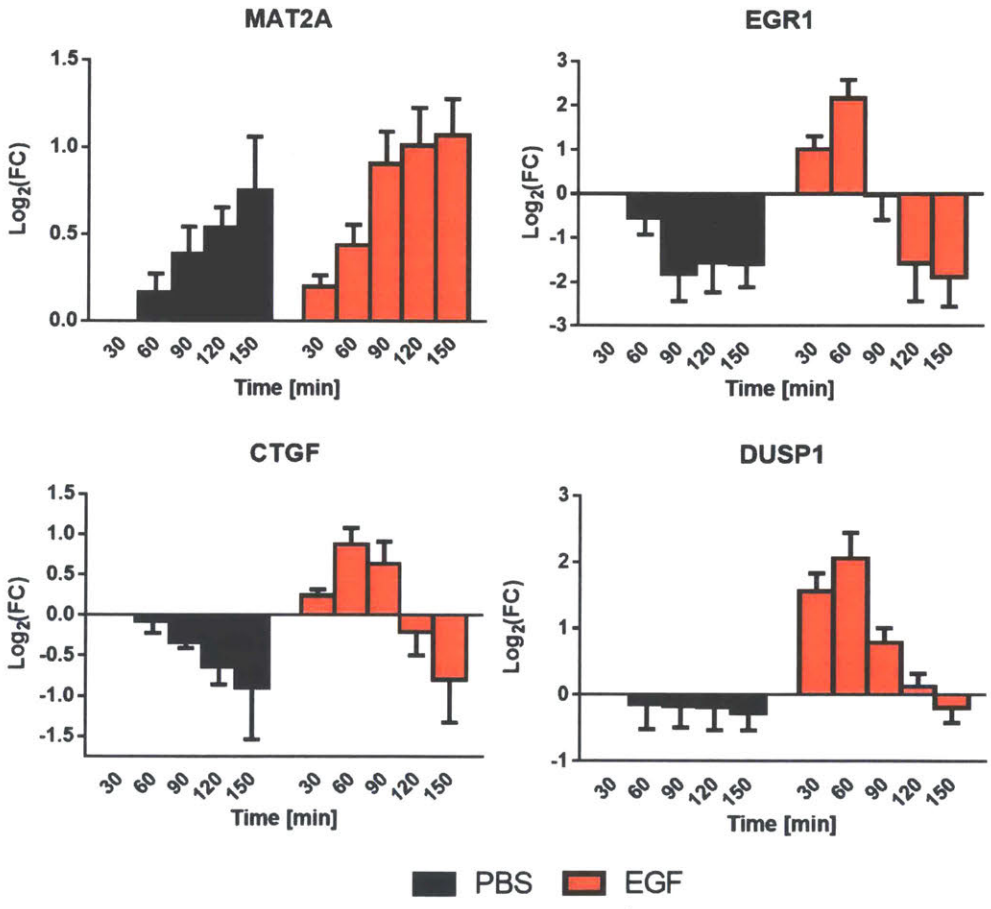


Figure 4-6 – Some proteins showed time-dependent changes in protein synthesis even in PBS controls (MAT2A, EGR1, CTGF). These changes may be induced by the transfer of cells to KRM- free media prior to labeling with Aha. DUSP1 is an example of a protein that did not show changes in the PBS control

Finally, cluster D was characterized by an increase in synthesis beginning around 30 minutes and remaining elevated across all time points. This cluster almost exclusively contained proteins associated with translational machinery including ribosomal proteins, translation initiation factors, and translation elongation factors (Figure 4-7).

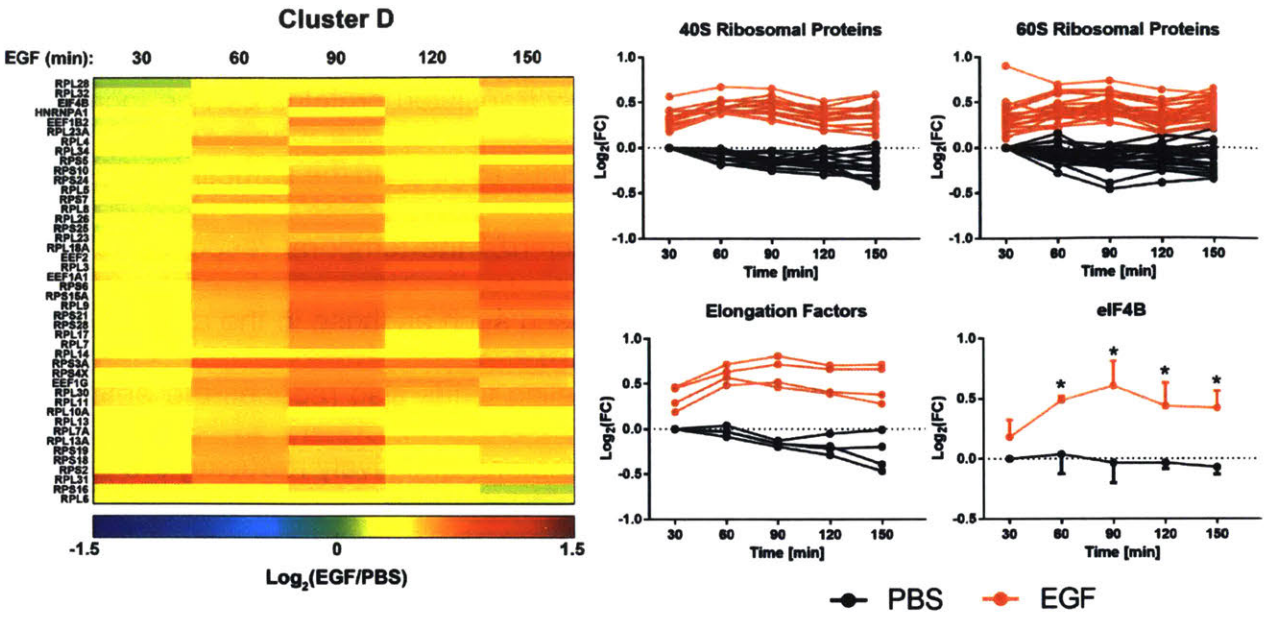


Figure 4-7 – Cluster D was characterized by increased synthesis beginning at 30 minutes and elevated synthesis at all time points of the experiment. This cluster consisted of almost exclusively ribosomal proteins and associated translation factors. * = $p < 0.05$ for eIF4B graph

The phenotypic changes in response to pro-growth cues including EGF stimulation require the synthesis of new proteins. To accommodate this increase in protein synthesis, cells also appeared increase their translational capacity by synthesizing more ribosomal proteins and associated translation factors. This observed increase in ribosomal protein synthesis is consistent with the observation that ribosomal RNAs (rRNAs) also show an increase in transcription within 30 minutes following EGFR activation[30].

This pro-growth response contrasts with ribosomal protein synthesis in the context of cellular stress when cells are conserving energy by globally repressing translation. Under stress conditions, ribosomal proteins and associated translation factors are sharply downregulated (see chapter 3).

4.3.2 Increasing Temporal Resolution

The temporal resolution of analyzing newly translated proteins can be increased by reducing the Aha/K8/R10 labeling periods and thus increasing the sampling frequency. Increased sampling is desirable to better characterize the temporal response to EGF, especially for proteins that are transiently expressed such as those in the above defined clusters A and B. However, decreasing the labeling times also reduces the amount of proteins that incorporate an Aha handle for enrichment, effectively reducing the amount of protein available for pulldown and subsequent analysis.

To better characterize the temporal responses of proteins expressed within the first hour after EGFR activation, Aha/K8/R10 labeling times were reduced from 30 minutes to 15 minutes. Samples were collected at 15, 30, 45, 60, and 75 minutes following EGF treatment with matched negative controls as described previously. After processing three replicates, 1860 unique proteins were identified, with 1136 present in more than one replicate. It was surprising that marginally more total proteins are observed in this 15 minute analysis than in the 30 minute labeling experiments given that the total amount of Aha-labeled protein available for enrichment should be decreased by a factor of two. One possible explanation for this observation is that the samples may be so complex that the number of positive identifications was limited not by instrument sensitivity, but rather by instrument cycle time. Under this assumption, more newly synthesized proteins may be identified by increasing the number of fractions analyzed separately.

To determine if increasing temporal resolution to 15 minutes further stratifies newly translated proteins into unique clusters not accessible at 30 minute resolution, k-means

clustering using Pearson correlation was performed on each dataset independently. Proteins were filtered to include only those in both datasets, those present in more than one replicate, and those with a statistically significant change between PBS and EGF samples in at least one time point of the 30 minute dataset. 74 proteins passed these requirements.

The clustering patterns between the two datasets were very similar. For instance, the clustering of IEGs and DEGs from cluster A that were rapidly upregulated was retained in the 15 minute dataset, as was the clustering of IEGs and DEGs from cluster B that are upregulated slightly later. Because the time course ended at 75 minutes -- before both the cluster B proteins and LRGs in cluster C reach their maximum -- both LRGs and late IEG/DEGs were clustered together. Finally, the ribosomal proteins and translation factors from cluster D remained clustered. However, there appeared to be a bifurcation between those that began upregulation at 15 minutes and those that began at 30 minutes (Figure 4-8, right and center).

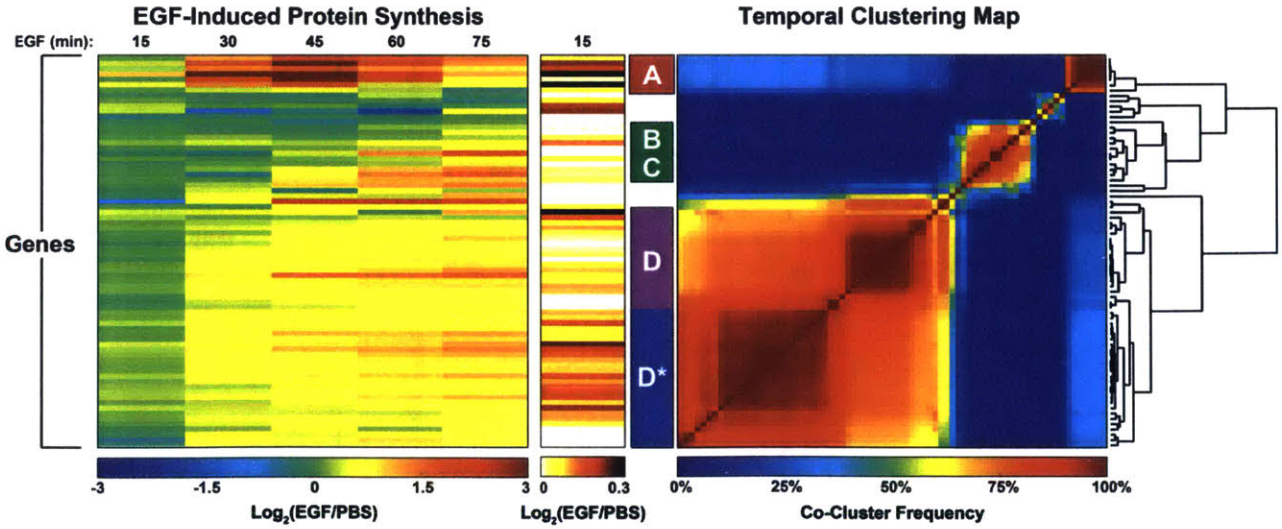


Figure 4-8 – Clustering the 15 minute resolution dataset yielded similar results as clustering with the 30 minute resolution dataset. The main differences were the grouping of IEGs/DEGs that peak at 90 minutes and LRGs into the same cluster (cluster B/C). The large ribosomal cluster D also showed further subgrouping depending on the extent of upregulation at the 15 minute time point (cluster D/D* and center heatmap).

Despite the relatively small changes in clustering, increasing the temporal resolution did yield interesting insights at specific time points. One notable change in clustering was the removal of ATF3 from cluster A into its own cluster. This change appeared to be driven by the 15 minute time point, where ATF3 synthesis was decreased below basal levels, before increasing and peaking by 45 minutes. This decrease of expression within 15 minutes, followed by a rapid increase over the next 30 minutes and subsequent return towards basal levels demonstrated the rapid dynamics of protein synthesis over short time periods. Other proteins were observed to be rapidly upregulated within the first 15 minutes including DUSP1, JUN, and several ribosomal proteins (Figure 4-9, black arrowheads). DUSP1 was one of the fastest responders, with expression increasing nearly 60% within the first 15 minutes. This rapid upregulation of a DEG even before most IEGs had not been previously observed and stands in contrast to the classical model of DEGs being upregulation following the activation of IEGs.

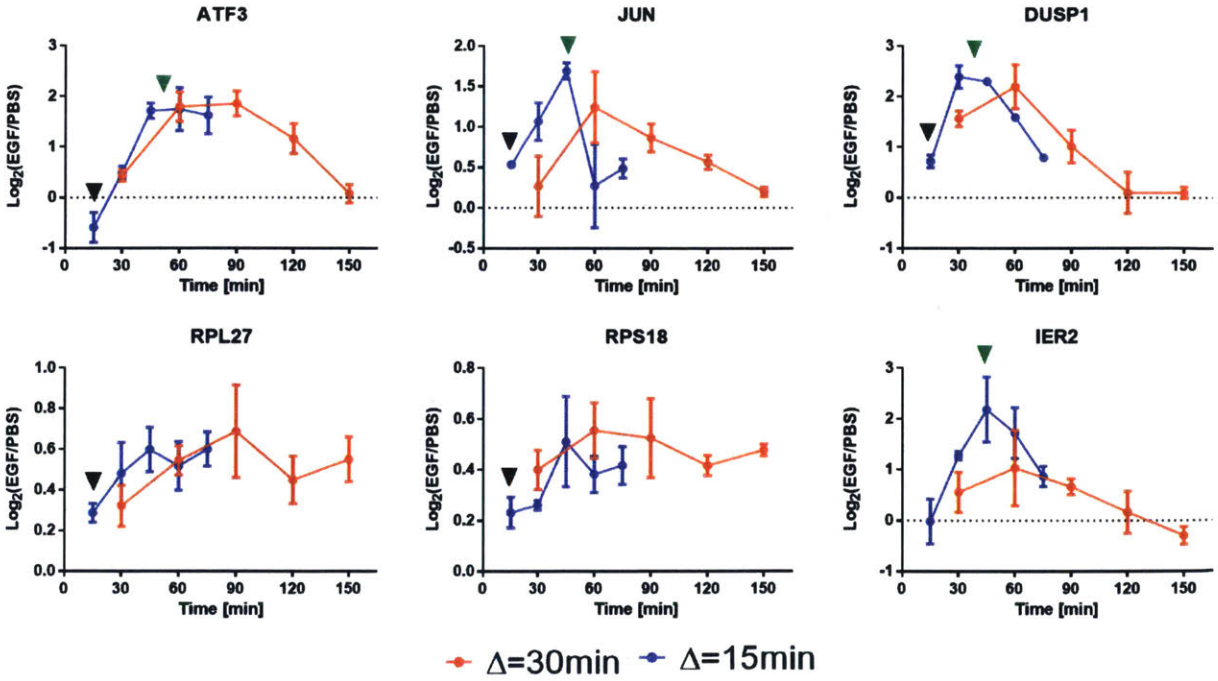


Figure 4-9 – Increasing temporal resolution revealed changes in protein synthesis within the first 15 minutes after EGF stimulation (black arrowheads), as well as improved estimation of peak protein synthesis at times inaccessible by sampling every 30 minutes (green arrowheads).

Increasing temporal resolution also allowed for improved characterization of when transient IEGs and DEGs reach their peak synthesis rates. For example, proteins such as EGR1, JUN, and IER2 appeared to reach peak synthesis around 45 minutes, a time point not assayed when sampling every 30 minutes. Based on the equal intensities of adjacent time points, it can be estimated that DUSP1 peaked between 30 and 45 minutes, and that ATF3 peaked between 45 and 60 minutes (Figure 4-9, green arrowheads).

Finally, increased temporal resolution allowed for better determination of when LRGs become initially upregulated. Measuring protein synthesis every 30 minutes suggests that LRGs become upregulated starting 90 minutes following EGFR activation. However, increasing the sampling resolution to every 15 minutes shows that LRGs such as EZR and VASP may be upregulated as early as 60 or 75 minutes after EGF treatment (Figure 4-10, black arrowheads).

Given that Aha/K8/R10 labeling duration was decreased from 30 minutes to 15 minutes without a significant loss of sensitivity, I sought to determine whether increasing the sampling frequency to 5 minute intervals would be feasible. To capture the dynamic changes in protein synthesis that occur between 30-60 minutes, cells were treated with 20 nM EGF and Aha/K8/R10 was applied to the cells for 5 minutes at 25, 30, 35, 40, and

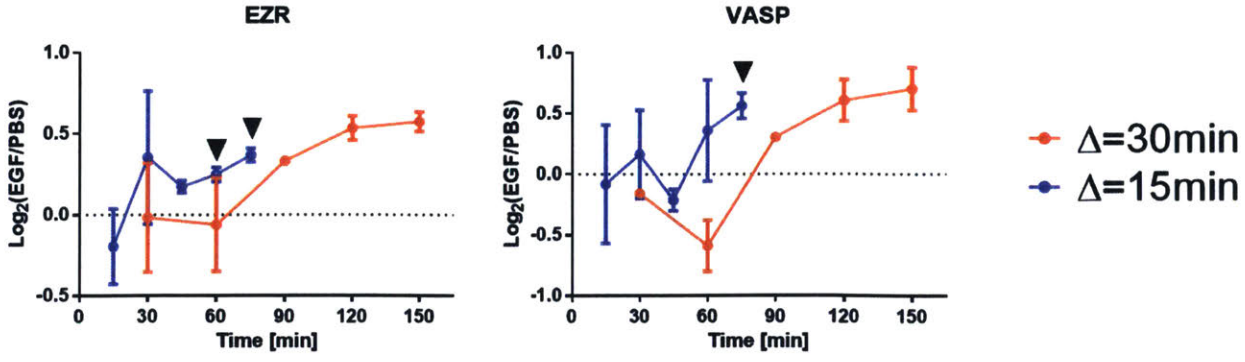


Figure 4-10 – Data from sampling every 30 minutes suggests that LRGs were upregulated around 90 minutes. However, increasing the labeling frequency to every 15 minutes revealed that LRGs may be upregulated as early as 75 minutes after EGF treatment (black arrowheads).

45 minutes after stimulation. Samples were lysed and processed as described previously. Over 3 replicates, only 319 unique proteins were identified, with only 11 appearing in more than one replicate. Furthermore, non-specific interactions (as measured by non-SILAC labeled proteins) increased significantly in these samples, with only 4-12% of identified protein containing a SILAC label. As labeling times decreased, the amount of protein containing an Aha residue decreased as well, and the ratio of unlabeled proteins to labeled proteins increased, increasing the opportunity for nonspecific binding to the beads. Increasing the amount of input material may result in better sensitivity and specificity.

Despite the poor sensitivity of this analysis, the IEGs CTGF and CYR61 were observed in multiple replicates. While upregulation was observed at all time points between 30-50 minutes, the amount variation between replicates compared to the magnitude of the changes between time points confounded drawing conclusions regarding peak synthesis time. Because so little material was enriched and analyzed in these analyses, small variations that were introduced during processing were magnified. Further improvements in sample handling, perhaps through automation, may improve the quality and consistency of the data.

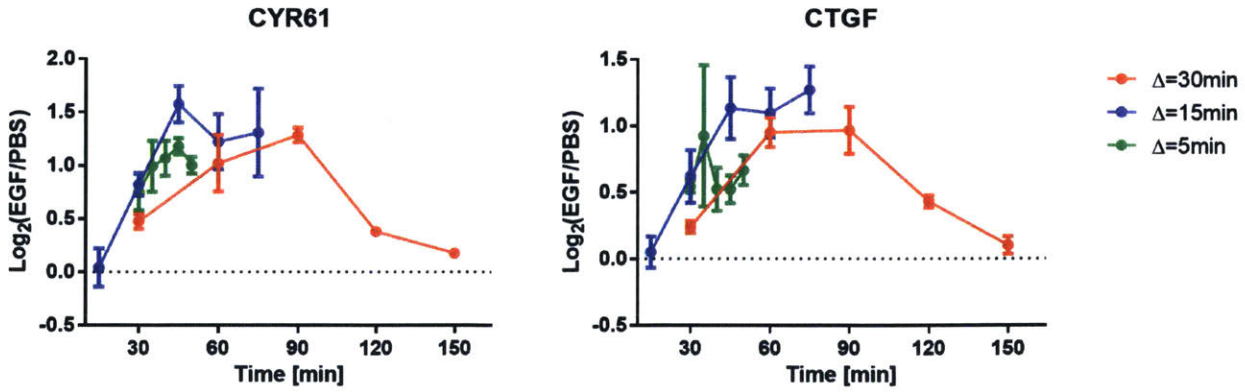


Figure 4-11 – Data from sampling every 5 minutes was largely consistent with previously acquired data, but the variations between replicates made drawing conclusions over short time scales challenging.

4.3.3 Comparison to RNA Sequencing and Ribosome Footprints

As discussed in section 1.5, a wide variety of technologies are available to measure changes in protein expression. Two of the most commonly used ones are RNA sequencing (RNA-Seq) and ribosome footprint (RFP) analysis. These methods are discussed in detail in sections 1.5.1 and 1.5.3, respectively. To briefly summarize, RNA-Seq uses Next Generation Sequencing technologies to sequence and count the number of times an mRNA transcript was read in transcripts per million reads (TPMs). This approach to analyze changes in protein expression assumes that regulation occurs on the level of transcription and does not account for variations in translational regulation. In contrast, RFPs map where ribosomes are bound on an mRNA transcript by including a ribonuclease digest that degrades all RNA except for portions shielded by bound ribosomes. By sequencing these short protected fragments, they can be counted (also in units of TPM) and mapped back to the genome to determine which transcripts are bound by ribosomes.

RNA-Seq and RFP samples were prepared, sequenced, and analyzed as described in section 4.2.3. Due to the limited capacity of the Illumina kits used, only the first three time points were assayed for RNA-Seq and RFP analysis. However, these time points included the most dynamic changes in protein translation. The data was filtered to only include proteins observed in two or more replicates of BONCAT experiments, proteins that showed a 25% increase or decrease in synthesis in at least one time point, and transcripts that were present in both the RFP and RNA-Seq analysis. Translational efficiency (TE) was calculated using the Xtail algorithm[20], and is roughly equivalent to ribosomes per mRNA transcript. The data were clustered by concatenating all four

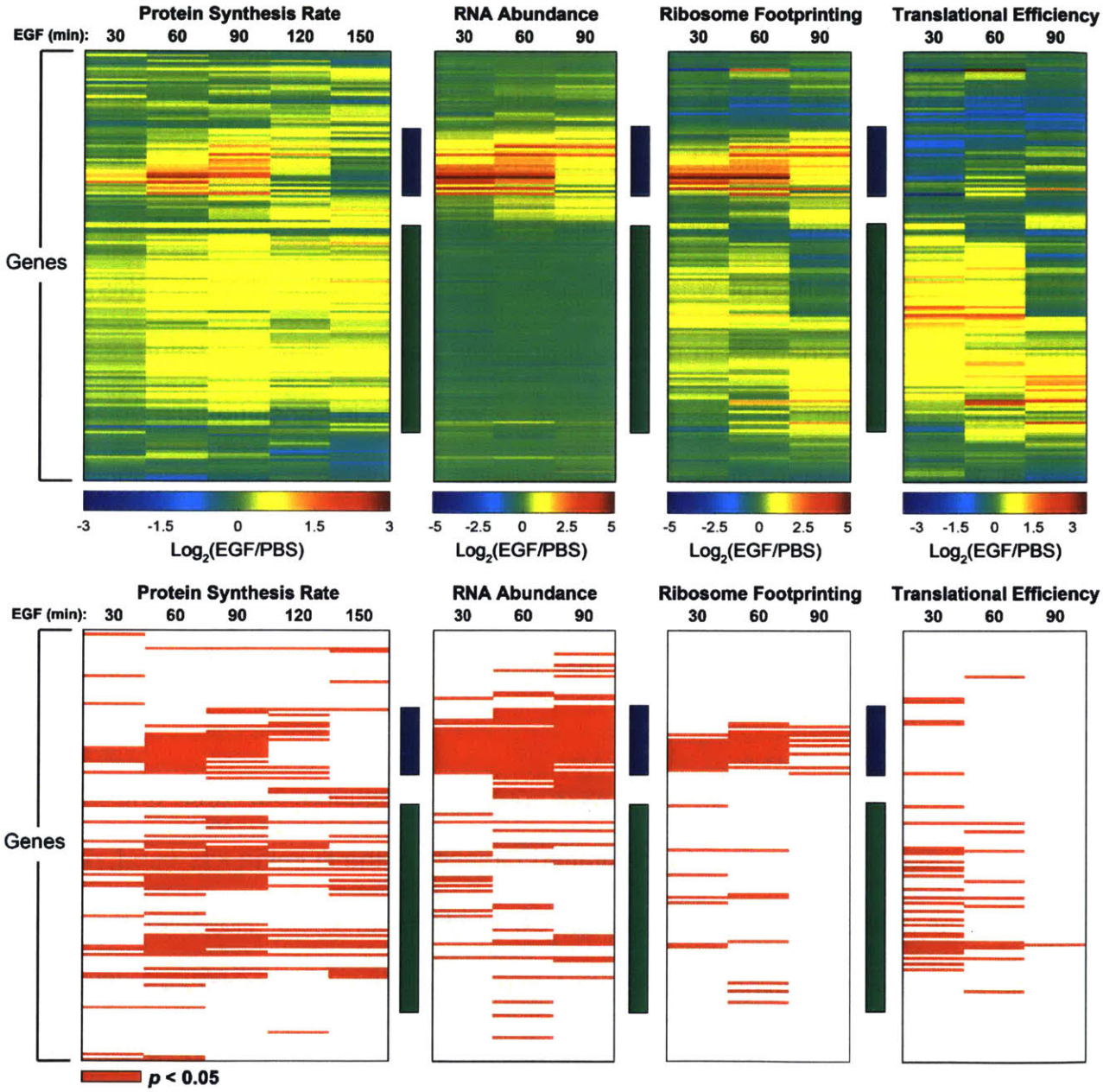


Figure 4-12 - Direct comparison between protein synthesis, RNA abundance (from RNA-Seq), and translational efficiency (from RFPs) showed distinct control mechanisms over different sets of genes. One set of proteins were controlled at the level of transcription, with RNA abundance correlating with protein synthesis (blue bar). Another set of proteins was controlled at the level of translation, with variations in translation efficiency correlating with protein synthesis with little change in RNA abundance (green bar). Statistically significant changes compared to PBS controls are marked by red boxes in the bottom diagram.

datasets (synthesis, RNA-Seq, RFP,and TE) and performing hierarchical clustering with Pearson correlation distance metric.

Comparing protein synthesis to RNA abundance and TE measurements, two distinct groups of genes emerged. In the first group, changes in protein synthesis correlated strongly with changes in RNA abundance (Figure 4-12, blue bar and Figure 4-13), suggesting control predominantly at the level of transcription. In the second group, changes in protein synthesis correlated with variations in TE, with little change in RNA abundance (Figure 4-12, green bar and Figure 4-14), suggesting regulation on the level of translation. Furthermore, statistically significant changes in RNA abundance were equally likely to occur across all three time points assayed, whereas changes in TE occurred primarily within the first 30 minutes. This observation may suggest that translational regulation may be a mechanism to control rapid changes in protein synthesis.

Proteins in the transcriptionally controlled cluster were the IEGs and DEGs that increased in the first 30-90 minutes after EGFR activation (Figure 4-13). Previous reports have characterized IEGs and DEGs based on their transcription profiles, so the correlation of their RNA abundance and protein synthesis was not unexpected. One notable observation was the decrease in TE of ATF3 within the first 30 minutes. This rapid translational downregulation of ATF3 was also observed in the 15 minute resolution BONCAT dataset (Figure 4-9, top left). The mechanism and consequence of this rapid fluctuation in ATF3 protein synthesis is unknown, but appears to be driven by a rapid decrease in translation followed by an upregulation of transcription.

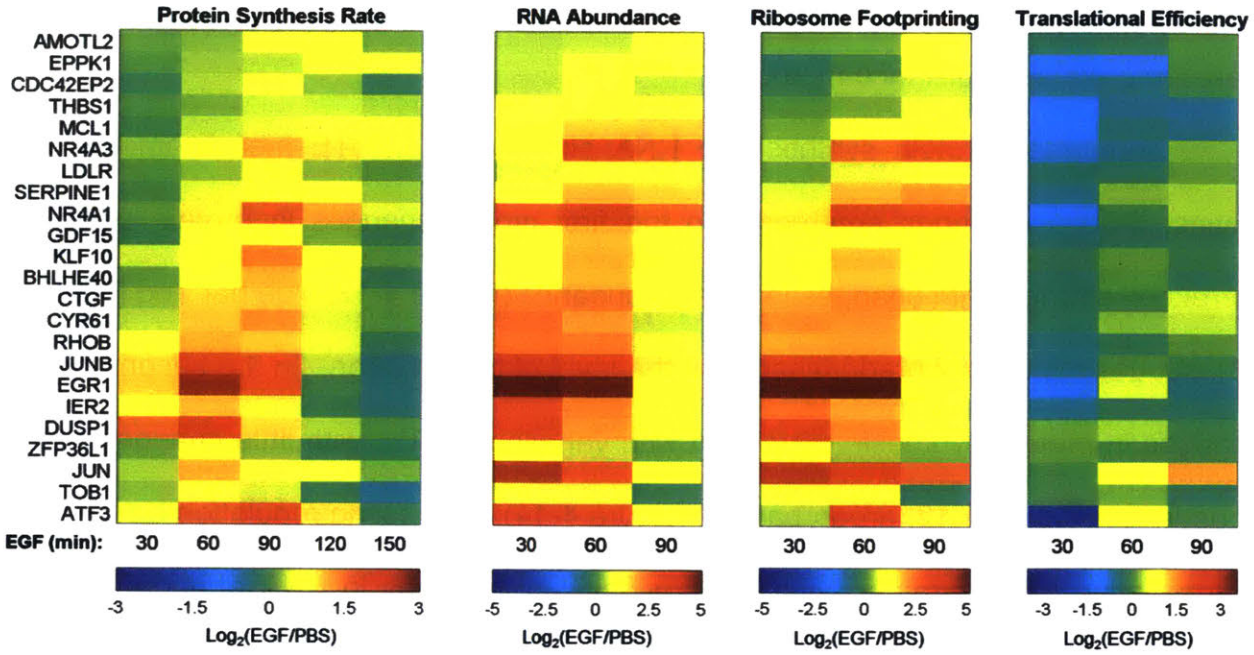


Figure 4-13 – This cluster shows a strong correlation between protein synthesis and RNA abundance, suggesting that regulation of protein expression was primarily on the level of transcription.

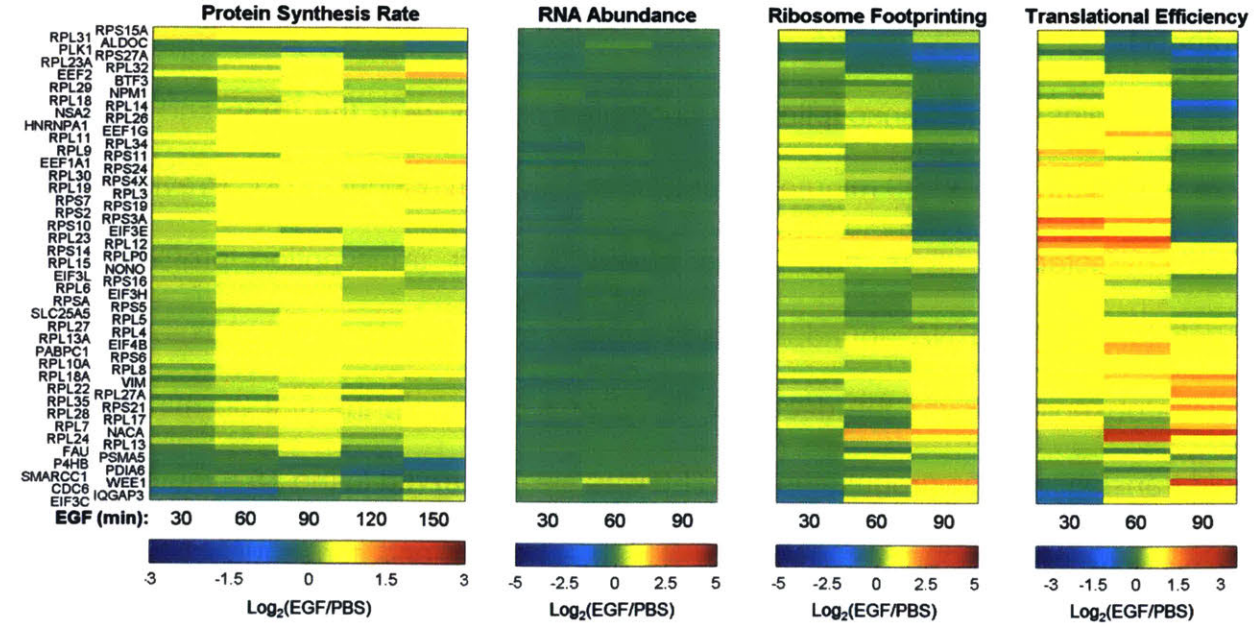


Figure 4-14 – This cluster shows a strong correlation between protein synthesis and translational efficiency, suggesting that regulation of protein expression was primarily on the level of translation.

The second cluster was defined by changes in protein synthesis correlating with changes in TE, suggesting regulation on the level of translation. This cluster was highly enriched for ribosomal proteins and several translation initiation and elongation factors. These proteins were rapidly upregulated and sustained throughout the duration of the analysis, albeit to a lesser magnitude than the IEGs, DEGs, and LRGs. In the 15 minute resolution BONCAT experiments, several of these ribosomal proteins showed increased synthesis within the first 15 minutes after EGF stimulation (Figures 4-8 and 4-9). Increases in translation without the need for increased transcription may be a mechanism to control the quick upregulation of these ribosomal proteins.

Another notable difference between the two clusters was the magnitude of the change in protein synthesis. In the transcriptionally-driven cluster, protein synthesis increased as high as 7-fold, whereas the most highly upregulated protein in the translationally-driven cluster experienced a 2-fold change in synthesis. Upregulation solely through increased translation may be limited by the maximum rate at which ribosomes can bind and translate a transcript. Larger changes in protein expression may require increases the number of mRNA transcripts available to be translated. This observation highlights a potential tradeoff between fast, but limited upregulation through translational control, and slower, but more potent upregulation through transcriptional control.

Generating time course data for ribosome binding (via RFPs) and protein synthesis (via BONCAT) offered the unique opportunity to characterize the temporal relationships between these two processes. Previous studies have modeled the delay between mRNA and protein abundances in the context of EGFR activation[11]. However, the latency

between ribosome binding and protein synthesis has not been previously studied. Because RFP analysis and protein synthesis occur on scales of different magnitudes, all data were log2-mean centered and normalized to the standard deviation for each individual protein. Because the protein synthesis data contain 5 time points and the RNA-Seq and RFP data contain 3 time points, the synthesis data were standardized using the mean and standard deviation of the first three time points.

To compare the dynamics of ribosome binding and protein synthesis, proteins were filtered to only include those that appeared in at least 3 replicates and had at least one statistically significant time point in each of the RFP and protein synthesis datasets. Manual inspection of the 27 genes that fit the criteria resulted in three groups: 10 genes demonstrate overlapping ribosome binding and protein synthesis, 17 genes demonstrate a delay between ribosome binding and protein synthesis, and 3 genes appear to have no relationship between the two datasets (Figure 4-15 and Table 4-1).

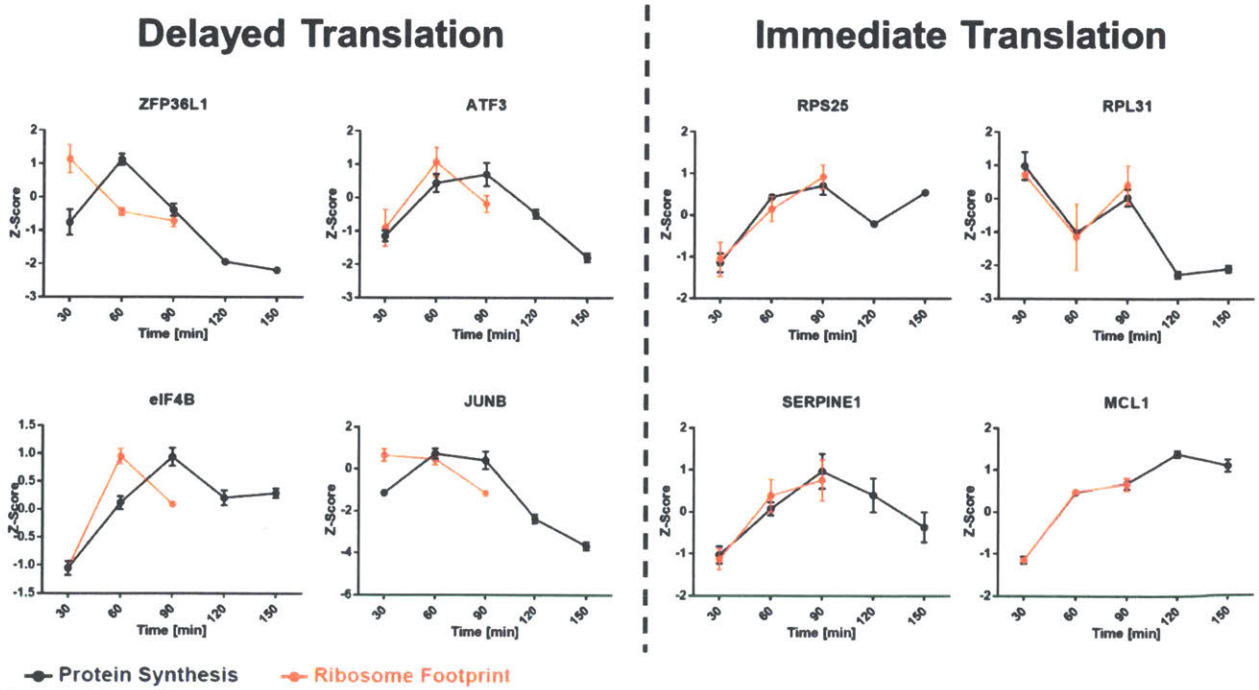


Figure 4-15 – Comparing ribosome binding to protein translation reveals a class of proteins that exhibits synthesis delayed from ribosome binding (left), while others demonstrate overlapping ribosome binding and protein synthesis (right)

Delayed				Concurrent	
ATF3	EGR1	JUNB	RPL11	GDF15	RPL31
DUSP1	EIF4B	KLF10	RPL7A	MCL1	RPS25
CTGF	IER2	NR4A1	TOB1	NR4A3	PABPC1
CYR61	JUN	RHOB	ZFP36L1	SERPINE1	
BHLHE40					

Table 4-1 – Classification of proteins based on temporal relationship between ribosome binding and protein synthesis

4.3.4 Analysis of Codon Bias

Codon bias refers to the preference of one particular codon for a given amino acid to be used more frequently in the genomic sequence of a protein or group of proteins than others. This phenomenon and its potential for regulating gene expression is discussed in detail in section 1.2.4. In Figure 4-15, two different behaviors were observed when comparing ribosome binding, and translation. One factor that may account for these differences is codon bias.

To investigate whether codon bias may be playing a role in the observed differences between the "delayed" and "immediate" translation groups, the average frequency of each codon was compared within the two groups to the average codon frequency in the human genome. To determine whether bias within a particular group was statistically significant, randomized groups of the same size were generated, and the average frequency was compared to the genome average. This process was repeated 1 million times, and a p-value was generated by counting how many randomly generated groups showed a frequency deviation from the genome greater than that of the queried group. Reported p-values were statistically significant if they were less than the Bonferroni corrected α = 7.81e-4 (accounting for 64 different codons).

Proteins that demonstrate an overlap between ribosome binding and protein synthesis do not appear to have any statistically significant codon bias compared to the

genome average. However, proteins that exhibit a delay between ribosome binding and protein synthesis have a codon bias associated with 9 different amino acids (Figure 4-16, left). To determine if these codon bias profiles were unique to these classes, 500 different sets of 17 proteins were selected at random, and their codon usage frequency was compared to that of the human genome. After this analysis, 469 sets demonstrated no significant codon bias and none had a bias in more than 4 amino acids (Figure 4-16, center and right), demonstrating that this codon bias is unique to the "delayed" translation set of proteins. While the effect of this codon bias and temporal delay are not known, it could potentially play a role in fine-tuning the timing of protein expression following EGF stimulation. The relative availability of tRNA isoacceptors is not static, but rather subject to changes in tRNA transcription, localization[32] and post-transcriptional modifications[33]. Therefore, the relationship between tRNA isoacceptor availability and codon usage rates in mRNA transcripts proteins may be another mechanism of control over protein expression.

To asses global changes in codon usage following EGFR activation, the deviation of codon frequency from the genome average was weighted by the log2-fold change compared to the PBS control and summed over all the proteins in our dataset that

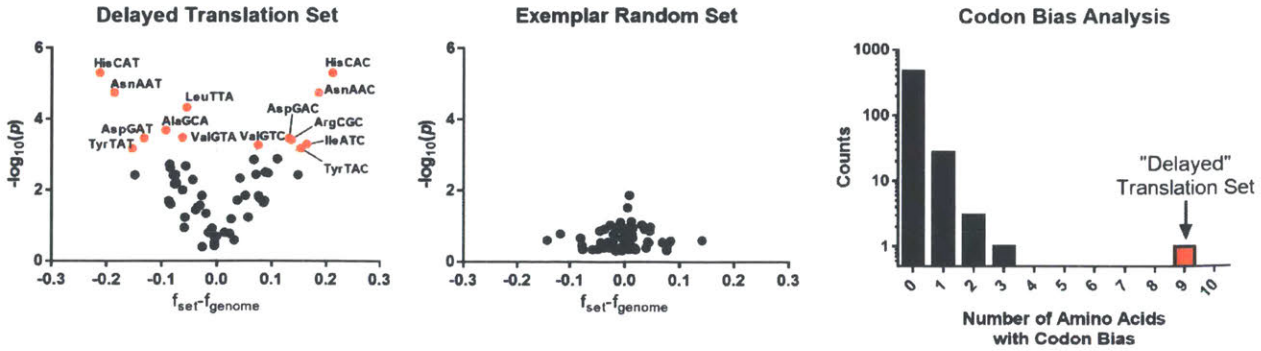


Figure 4-16 – Codon usage analysis reveals that the "delayed" translation set has a codon bias in 9 different amino acids (left). This is unique to this gene set, as randomly selected gene sets tend to not have a large codon bias compared to the genome average (center, left)

appeared in 3 or more replicates. This value was calculated for each codon at each time point and was termed the weighted relative codon bias (WRCB):

$$WRCB_{codon,t} = \sum_{i=proteins} (f_{codon,i,t} - f_{codon,i,genome}) * log_2(\frac{EGF_{i,t}}{PBS_{i,t}})$$

WRCB represents the global change in codon usage compared to the genome average. It was found that codon bias across all proteins in our dataset was dynamic, with changes in preferred codon usage changing with respect to time (Figure 4-17, left). Asparagine was an example of an amino acid with a particularly dynamic codon bias. At 30 minutes, both codons were used with roughly equal frequency, but AsnAAC was preferred between 60 and 90 minutes, before switching to AsnAAT by 150 minutes (Figure 4-17, right). This dynamic changing of codon usage frequency in synthesized proteins may be another point of regulation in protein expression.

In recent years, a new stress-induced translational control mechanism was discovered. In this mechanism, toxicant-exposure leads to the reprogramming of dozens

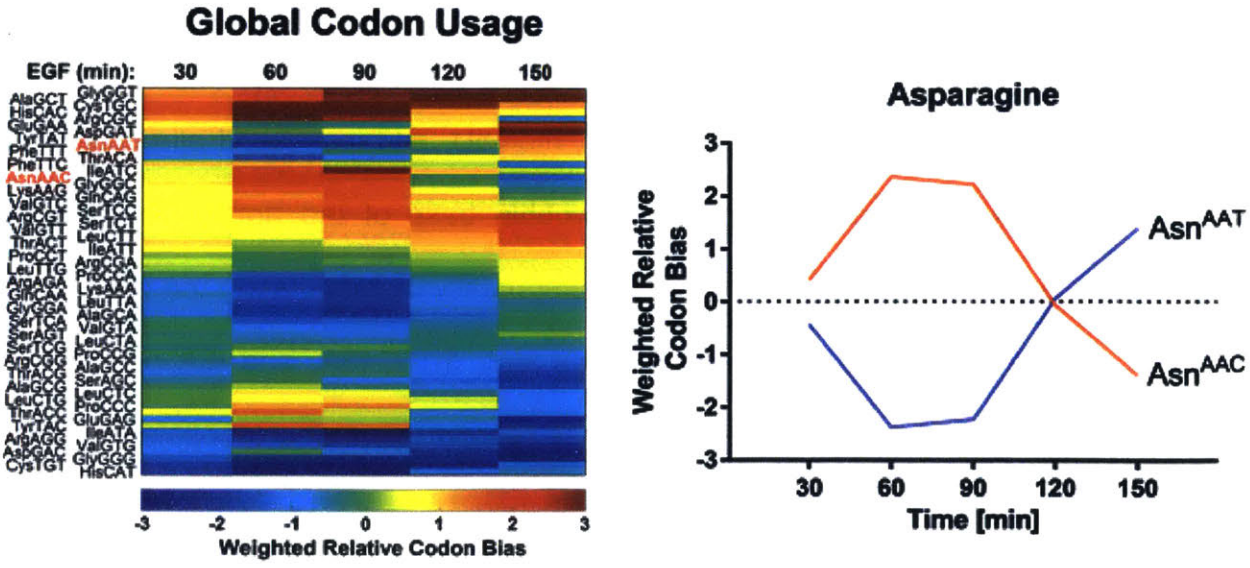


Figure 4-17 – Changes in relative codon frequency weighted by protein synthesis was analyzed over all time points. Changes in codon usage over time suggest that codon bias may be another mechanism of regulation during the process of protein expression.

of modified ribonucleosides in tRNA that regulate the selective translation of codon-biased mRNAs for critical proteins to respond to that specific stress[33,34]. Modified nucleotides in tRNAs can extend the range of recognized synonymous codons by enabling these tRNAs to form wobble base pairs. To determine whether this mechanism also played a role in the response to EGF stimulation, tRNA modification changes were quantified in EGF-stimulated cells using an established bio-analytical platform[35]. In brief, after total RNA isolation, tRNA species were purified by size-exclusion HPLC and subjected to enzymatic hydrolysis. The resulting ribonucleosides were then resolved by HPLC and identified and quantified on a triple quadrupole mass spectrometer. Details on the structure and functions of particular modifications have been discussed in other studies[36,37].

EGF treatment did indeed appear to result in differential enzymatic modification of tRNAs in a time-dependent manner (Figure 4-18). Interestingly, many of the significantly altered modifications are predicted to be wobble modifications or situated close to the anticodon loop and thus have the potential to reprogram codon reading of tRNAs. For example, the modifications I, Gm, Cm, m5C, Y, mcm5U, and mcm5s2U all occured at position 34 of tRNAs, corresponding to the wobble base of the anticodon. These modifications can affect which codon is read preferentially by modulating the reading of the codon wobble position[38]. The modifications m1I, m1G, t6A, and i6A occured at position 37 adjacent to the anticodon, and have been shown to enhance ribosome-binding affinity, reduce misreading, and modulate frame shifting[39].

Future work aims at unveiling a molecular link between the observed tRNA modification changes (Figure 4-18) in response to EGF treatment and the identified codon

biases in different protein classes (Figure 4-16). For this study, significantly altered RNA modifications will be mapped to specific tRNAs by LC-MS/MS analysis of RNase-generated oligonucleotides. RNase-specific digestion products, containing a modified ribonucleoside, will be defined based on the genome sequences of the complete set of expressed tRNAs and published tRNA modification maps[36,37]. Sequence-unique 4- to 5-mer oligoribonucleotides will then be targeted during MS analysis, and their unique fragmentation patterns will be used for sequence confirmation and localization of the modification. Anticodons associated with the modification-mapped tRNAs will then be used to compare to codons identified as over- and underrepresented in the BONCAT proteomics analysis.

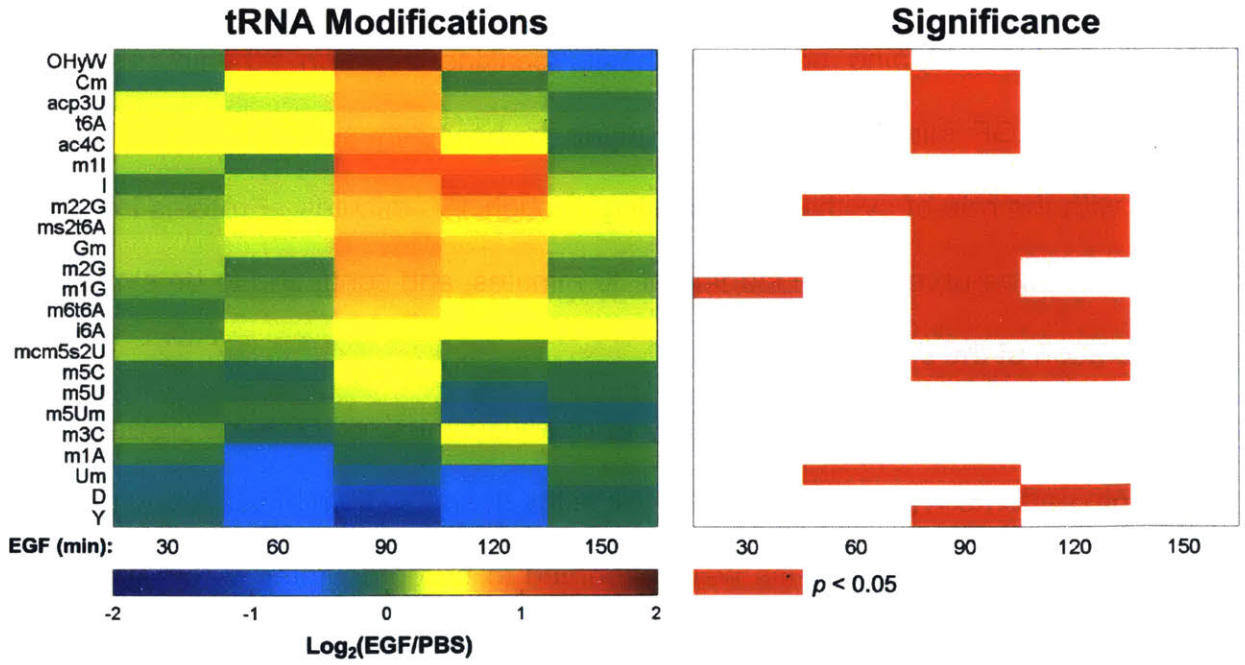


Figure 4-18 – Quantitative LC-MS/MS analysis of tRNA nucleosides revealed many dynamic changes in modifications in or adjacent to the tRNA anticodon loop that have the potential to modulate translation. Further analysis is needed to map specific modifications to specific tRNA isoacceptors. Right heatmap shows statistically significant changes in enzymatic modification of tRNAs as calculated by Student's T-test.

4.4 Conclusions

The transcriptional response to EGF stimulation has been well characterized, with a rapid upregulation of IEGs within the first hour, followed by the rise of DEGs which exert negative feedback on the IEGs to attenuate their expression. DEGs also control the expression of LRGs which result in changes in cell phenotype and play a role in determining cell fate[12,16]. Models have been built that characterize the relationship between transcription and protein abundance in the context of EGFR activation[11], but these models fall short of characterizing how translation affects protein synthesis.

In this chapter, quantitative multiplexed BONCAT was applied to directly analyze protein synthesis following EGFR activation. The proteins were clustered based on their temporal synthesis behaviors. Contrary to traditional transcription models, IEGs and DEGs showed overlapping protein synthesis, peaking between 60 minutes and 90 minutes after EGF stimulation. LRGs began to be synthesized starting around 90 minutes, with the rate of synthesis increasing through 150 minutes. Proteins involved in translation showed upregulation starting at 30 minutes, and continued to be synthesized for the duration of the time course.

Increasing the sampling frequency of newly synthesized proteins yielded further insights into patterns in protein synthesis. While the overall pattern of clustering remained largely unchanged, some proteins were observed to be upregulated within the first 15 minutes after stimulation. Furthermore, the times of peak synthesis of IEGs and DEGs as well as times of initial LRG upregulation were better characterized by decreasing labeling times to 15 minutes.

Protein synthesis as measured by BONCAT was also compared to transcription (via RNA-Seq) and ribosome binding (via RFP). This comparison revealed two distinct groups of genes: those regulated by changes in transcription and those regulated by changes in translation. Genes regulated by transcription showed increases in protein synthesis that correlate with changes in RNA abundance. Proteins in this group were the IEGs and DEGs that show the largest changes in protein synthesis. Genes regulated by translation showed increases in protein synthesis correlating with changes in ribosome binding, but in the absence of changes in RNA abundance. This group consisted primarily of ribosomal proteins, translation initiation factors, and translation elongation factors. Many of these proteins showed changes in synthesis as early as 15 minutes, suggesting that translational control can result in rapid changes in protein synthesis. However, the magnitude of this upregulation was limited compared to proteins in the transcriptionally-driven group, suggesting that large changes in expression may require increases in transcription.

Comparing the temporal relationship of ribosome binding and protein synthesis also resulted in two distinct behaviors. "Immediately" translated proteins showed an overlap between ribosome binding and protein synthesis, whereas "delayed" translation showed a delay between ribosome binding and protein synthesis. An analysis of codon usage frequencies showed a unique bias in the "delayed" translation groups that was not present in other randomly selected groups of proteins, opening the possibility for regulation based on tRNA modifications and tRNA copy numbers. There was no significant codon bias in the "immediately" translated group compared to the genome average which may explain the efficient translation of these proteins. Furthermore, EGFR

activation resulted in dynamic changes in global codon usage as well as chemical modifications of tRNAs within the first 150 minutes. While this analysis did not assign modifications to specific tRNA isoacceptors, future work can examine isoacceptor-specific oligoribonucleotides to map modifications to tRNA isoacceptors. Anticodons associated with the modification-mapped tRNAs can then be used to compare to codons identified as over- and underrepresented in the BONCAT proteomics analysis. By measuring many of the regulatory steps involved in protein expression (transcription, ribosome binding, tRNA modification, codon usage frequency) and comparing them to the magnitude of protein synthesis, a comprehensive model of protein expression can be developed.

4.5 References

1. Cadena, D. L. & Gill, G. N. Receptor tyrosine kinases. *FASEB J.* **6,** 2332–7 (1992).
2. Carpenter, G. & Cohen, S. Epidermal growth factor. *Cell* **48,** 193–216 (1979).
3. Schlessinger, J. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* **110,** 669–672 (2002).
4. Sundaram, M. RTK/Ras/MAPK signaling. *WormBook* 1–19 (2006). doi:10.1895/wormbook.1.80.1
5. Dhillon, A., Hagan, S., Rath, O. & Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **26,** 3279–3290 (2007).
6. Osaki, M., Oshimura, M. & Ito, H. PI3K-Akt pathway: Its functions and alterations in human cancer. *Apoptosis* **9,** 667–676 (2004).
7. Zhang, H. *et al.* ErbB receptors: From oncogenes to targeted cancer therapies. *Journal of Clinical Investigation* **117,** 2051–2058 (2007).
8. Reddy, R. J. *et al.* Early signaling dynamics of the epidermal growth factor receptor. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 3114–9 (2016).
9. Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A. & White, F. M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci.* **104,** 5860–5865 (2007).
10. Zhang, Y. *et al.* Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell. Proteomics* **4,** 1240–50 (2005).
11. Golan-Lavi, R. *et al.* Coordinated Pulses of mRNA and of Protein Translation or Degradation Produce EGF-Induced Protein Bursts. *Cell Rep.* **18,** 3129–3142 (2017).
12. Amit, I. *et al.* A module of negative feedback regulators defines growth factor signaling. *Nat. Genet.* **39,** 503–512 (2007).
13. O'Donnell, A., Odrowaz, Z. & Sharrocks, A. D. Immediate-early gene activation by the MAPK pathways: what do and don't we know? *Biochem. Soc. Trans.* **40,** 58–66 (2012).
14. Liu, C. *et al.* Dual-specificity phosphatase DUSP1 protects overactivation of hypoxia-inducible factor 1 through inactivating ERK MAPK. *Exp. Cell Res.* **309,** 410–418 (2005).
15. Sanduja, S., Blanco, F. F. & Dixon, D. A. The roles of TTP and BRF proteins in regulated mRNA decay. *Wiley Interdisciplinary Reviews: RNA* **2,** 42–57 (2011).
16. Feldman, M. E. & Yarden, Y. Steering tumor progression through the transcriptional response to growth factors and stroma. *FEBS Letters* **588,** 2407–2414 (2014).
17. Avraham, R. & Yarden, Y. Feedback regulation of EGFR signalling: decision making by early and delayed loops. *Nat. Rev. Mol. Cell Biol.* **12,** 104–17 (2011).
18. Lanahan, A., Williams, J. B., Sanders, L. K. & Nathans, D. Growth factor-induced delayed early response genes. *Mol. Cell. Biol.* **12,** 3919–29 (1992).
19. Liu, T. Y. *et al.* Time-Resolved Proteomics Extends Ribosome Profiling-Based Measurements of Protein Synthesis Dynamics. *Cell Syst.* **4,** 636–644.e9 (2017).
20. Xiao, Z., Zou, Q., Liu, Y. & Yang, X. Genome-wide assessment of differential

translations with ribosome profiling data. *Nat. Commun.* **7,** 11194 (2016).
21. Li, Q.-X., Ke, N., Sundaram, R. & Wong-Staal, F. NR4A1, 2, 3--an orphan nuclear hormone receptor family involved in cell apoptosis and carcinogenesis. *Histol. Histopathol.* **21,** 533–540 (2006).
22. Maxwell, M. A. & Muscat, G. E. O. The NR4A subgroup: immediate early response genes with pleiotropic physiological roles. *Nucl. Recept. Signal.* **4,** e002 (2006).
23. Spittau, B. & Krieglstein, K. Klf10 and Klf11 as mediators of TGF-beta superfamily signaling. *Cell and Tissue Research* **347,** 65–72 (2012).
24. Ozanne, B. W., Spence, H. J., McGarry, L. C. & Hennigan, R. F. Transcription factors control invasion: AP-1 the first among equals. *Oncogene* **26,** 1–10 (2007).
25. Wang, W., Sumiyoshi, H., Yoshioka, H. & Fujiwara, S. Interactions between epiplakin and intermediate filaments. *J. Dermatol.* **33,** 518–527 (2006).
26. Krause, M., Dent, E. W., Bear, J. E., Loureiro, J. J. & Gertler, F. B. ENA/VASP PROTEINS: Regulators of the Actin Cytoskeleton and Cell Migration. *Annu. Rev. Cell Dev. Biol.* **19,** 541–564 (2003).
27. Katz, M. *et al.* A reciprocal tensin-3-cten switch mediates EGF-driven mammary cell migration. *Nat. Cell Biol.* **9,** 961–969 (2007).
28. Chen, H. *et al.* Role of Methionine Adenosyltransferase 2A and S-adenosylmethionine in Mitogen-Induced Growth of Human Colon Cancer Cells. *Gastroenterology* **133,** 207–218 (2007).
29. Martínez-Chantar, M. L. *et al.* L-methionine availability regulates expression of the methionine adenosyltransferase 2A gene in human hepatocarcinoma cells: role of S-adenosylmethionine. *J. Biol. Chem.* **278,** 19885–90 (2003).
30. Stefanovsky, V. Y. *et al.* An immediate response of ribosomal transcription to growth factor stimulation in mammals is mediated by ERK phosphorylation of UBF. *Mol. Cell* **8,** 1063–1073 (2001).
31. Karlin, S. & Mrázek, J. Predicted Highly Expressed Genes of Diverse Prokaryotic Genomes Predicted Highly Expressed Genes of Diverse Prokaryotic Genomes. *J. Bacteriol.* **182,** 5238–5250 (2000).
32. Hopper, A. K., Pai, D. A. & Engelke, D. R. Cellular dynamics of tRNAs and their genes. *FEBS Letters* **584,** 310–317 (2010).
33. Chan, C. T. Y. *et al.* A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet.* **6,** 1–9 (2010).
34. Chionh, Y. H. *et al.* tRNA-mediated codon-biased translation in mycobacterial hypoxic persistence. *Nat. Commun.* **7,** 13302 (2016).
35. Cai, W. M. *et al.* A Platform for Discovery and Quantification of Modified Ribonucleosides in RNA: Application to Stress-Induced Reprogramming of tRNA Modifications. *Methods Enzymol.* **560,** 29–71 (2015).
36. Machnicka, M. A. *et al.* MODOMICS: A database of RNA modification pathways - 2013 update. *Nucleic Acids Res.* **41,** (2013).
37. Cantara, W. A. *et al.* The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39,** D195–D201 (2011).
38. Chan, C. T. Y. *et al.* Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat. Commun.* **3,** 937 (2012).

39. Rould, M. A., Perona, J. J. & Steitz, T. A. Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature* **352,** 213–218 (1991).

Chapter V

Examining Potential Resistance Mechanisms in NRAS Mutant Melanoma

Daniel Rothenberg

5.1 Introduction

Each year, nearly 132,000 new patients are diagnosed with melanoma, and 48,000 patients die of the disease worldwide[1]. When detected at an early stage, surgical excision of melanoma lesions results in a 90% 5 year survival rate[2]. However, if not detected at an early stage, malignant melanoma can invade deeper into tissues and metastasize to other organs. When malignant melanoma develops, the median survival is less than one year, and the 5 year survival rate falls to less than 10%.

Melanoma is typically characterized by the mutation status of proteins in the MAP kinase pathway. 50% of melanomas harbor a BRAF mutation, the most common of which is V600E[3]. NRAS mutations occur in 30% of malignant melanomas[4], with the other 20% driven by mutations outside the MAPK pathway. With the introduction of BRAF inhibitors (BRAFi), prognosis for BRAF mutant melanomas have significantly improved. A phase 3 clinical trial comparing the BRAFi vemurafenib to dacarbazine, a DNA methylating chemotherapeutic, showed that vemurafenib treatment resulted in a 6.9 month progression-free survival (PFS) compared to 1.6 months with dacarbazine[5]. However, acquired resistance is nearly certain, with most patients developing BRAFi-resistant melanoma within a year after starting treatment[6,7]. Another approach to treating BRAF mutant melanoma is to target downstream components of the MAPK pathway, particularly MEK. Clinical trials using the MEK inhibitor (MEKi) binimetinib resulted in a partial response, but only increased PFS to 3.7 months[8]. However, the combination of a BRAFi and MEKi improved the rate of PFS compared to treating with either alone[9]. Co-treatment with a MEKi and BRAFi is the current standard of care for BRAF mutant melanomas. Immunotherapies using checkpoint blockade inhibitors, such as nivolumab or ipilimumab

have also proven successful, with a combination of the two inhibitors increasing PFS to 11.5 months[10].

NRAS mutant melanomas require a different treatment strategy. BRAF inhibitors result in paradoxical activation of CRAF, due to the stabilization of BRAF-CRAF dimers in the presence of activated NRAS[11], and therefore are not a suitable option. Immunotherapy is the current first line treatment for NRAS mutant melanomas, but only 30% of patients show progression-free disease 12 months after treatment[12]. For patients that don't respond to immunotherapy, a MEKi is usually employed as the second line treatment. However, a clinical trial showed binimetinib marginally increased PFS to 2.8 months compared to 1.5 months in the dacarbazine treated group, and only 41% of patients responded to binimetinib treatment[13]. Due to the poor response, binimetinib was withdrawn for use in NRAS mutant melanomas. There are currently no approved targeted therapies for NRAS mutant melanomas. The increase in PFS compared to dacarbazine treatment suggests that the treatment initially works but is quickly followed by the emergence of an adaptive resistance mechanism. Understanding the mechanism of how this adaptive resistance arises may lead to the identification of drug targets that can be co-targeted to suppress resistance.

One of the most commonly used unbiased screens for resistance mechanisms is analysis of the transcriptome using RNA-Seq. RNA-Seq allows for the detection of sub-populations harboring mutations (selected resistance) as well as the upregulation or downregulation of genes that may lead to resistance (adaptive resistance)[14]. However, there are examples of resistance mechanisms that arise through increased translation[15]

or that are only observable through analysis of translation[16], and these targets would be missed in a screen looking for changes in the transcriptome.

In this chapter, BONCAT was used to screen for potential drivers of resistance to binimetinib. Two NRAS mutant melanoma cell lines showing different degrees of sensitivity to binimetinib were treated with two concentrations of the drug, and changes in protein synthesis were assayed at 24 hours after initial treatment. Because binimetinib-induced cell death does not occur until around 48 hours after treatment, profiling the translatome at an early time point may yield markers that indicate the rise of adaptive resistance.

5.2 Methods

5.2.1 Cell Culture

MelJuso cells were cultured in RPMI 1640 Glutamax supplemented with 10% FBS and 1% penicillin/streptomycin. MM415 cells were cultured in RPMI 1640 Glutamate supplemented with 10% FBS, 1% pencillin/streptomycin, and 250 mM HEPES. Cells were passaged roughly every 3 days, with MelJuso cells split 1:6-8, and MM415 cells split 1:4.

5.2.2 Binimetinib Treatment and Metabolic Labeling

All experiments were performed starting on the second day following passage. Cells were treated with the appropriate concentration of binimetinib or an equal volume of DMSO added directly to the growth media. 24 hours after treatment, growth media was removed and replaced with lysine/arginine/methionine (KRM)-free DMEM:F12

media. After 30 minutes, Aha was added to 3 mM, and $^{15}N_4{}^{13}C_6$ arginine (R10) and $^{15}N_2{}^{13}C_6$ lysine (K8) were added to 0.5 mM. After 30 minutes of Aha/K8/R10 labeling, the media was aspirated, and the cells were washed in ice cold PBS supplemented with 300 µg/mL CHX. Cells were lysed in 1% SDS in PBS supplemented with 50 mM N-ethylmaleimide (NEM) and 300 µg/mL CHX. 1.5 mL of -20°C acetone was added to each tube immediately following lysis to precipitate proteins. Proteins were precipitated at -20°C for at least an hour.

5.2.3 Western Blots

Western blotting was performed as described in section 2.2.3. In addition to the reagents outlined previously, the following antibodies were used: Erk1/2 1:5000 (Cell Signaling #9107), Erk1/2 pT202/pY204 1:1000 (Cell Signaling #4370), E-Cadherin 1:1000 (Cell Signaling #3195), N-Cadherin 1:1000 (BD Biosciences #610920), Vimentin 1:1000 (Cell Signaling #3932), MITF 1:1000 (Cell Signaling #12590), Zeb1 1:1000 (Cell Signaling #3396), NFκB 1:1000 (Cell Signaling #8242).

5.2.4 Cell Viability Assays

Cells were plated at a density of 10,000 cells in 90 µL per well of a 96 well plate. To control for edge effects, only the interior 60 wells were used, with the 36 outer wells filled with PBS. 24 hours after plating, the drugs to be applied were prepared at 10x concentration in the appropriate growth media. 10 µL of the 10x mixture was added to each well of the 96 well plate. After 72 hours of drug treatment, the plates were removed from the incubator and allowed to equilibrate to room temperature for 10 minutes prior to

the addition of 100 µL of CellTiter Glo reagent (Promega #G7570). Immediately following the addition of CellTiter Glo reagent, plates were placed in a Tecan plate reader and agitated for 120 seconds to lyse the cells, followed by a 10 minute incubation to allow luminescence to develop. Luminescence was measured in each well with 1000 ms integration time. Each drug concentration was measured in duplicate for each experiment. Normalized viability was measured by dividing all luminescence values by the average of DMSO controls.

5.2.5 Sample Preparation and MS Analysis

Samples were processed, enriched, washed, digested, TMT labeled, and fractionated as described in the protocol in appendix 2.6. Due to limitations on amount of material, 100-300 µg of total protein were used in each channel depending on the replicate. The pooled fractions were placed in a ThermoFisher Easy nLC 1000 autosampler and analyzed on a ThermoFisher QExactive Plus mass spectrometer. The instrument was operated in data dependent acquisition mode, with the top 15 most abundant precursors with charge of +2 or greater selected for fragmentation and dynamic exclusion set to 15 s. Precursors were isolated with a window of 0.4 *m/z* and fragmented via HCD at 33 NCE.

The Thermo .raw files were searched on MASCOT version 2.4 with fixed modifications for NEM alkylation on cysteines (+125.047 Da), addition of TMT 6-plex to N-termini and lysine residues (+229.163 Da). Variable modifications were SILAC R10 on arginine residues (+10.008 Da), SILAC K8 on lysine residues (+8.014 Da), addition of TMT 6-plex to SILAC K8 lysine residues (+237.177 Da), Aha substitution for methionine

residues (-4.986 Da), diaminobutyrate (reduced Aha) substitution for methionine residues (-30.976 Da), oxidation of methionine residues (+15.995 Da) and phosphorylation on tyrosine, threonine, and serine residues (+79.966 Da). Precursor tolerance was 10 ppm, fragment tolerance was 15 mmu, two missed cleavages were allowed, and the enzyme was set to trypsin. Peptides were considered to be positively identified if they had a score of at least 25 and newly translated if they contained a SILAC or Aha residue. For analysis, all counts were normalized to DMSO controls and log2 transformed.

5.2.6 Gene Set Enrichment Analysis (GSEA)

GSEA was performed using the GSEA v2.0 desktop from the Broad Institute[17] using the Hallmarks Molecular Signature Database (MSigDB)[18] and a procured list of MITF targets[19]. The input into the algorithm was the Log_2([MelJuso 1 μM]/[MM415 1 μM]) sorted from largest to smallest.

5.2.7 Partial Least Squares Regression (PLSR)

Log2 DMSO-normalized synthesis values were z-scored and used as the X-block of predictors. Normalized viability values from corresponding binimetinib concentrations (10 nM and 1 μM) were also z-scored and used as the Y-block of responses. PLSR modeling was performed using the MATLAB *plsregress* function with 2 principal components.

5.2.8 siRNA Knockdown

Small interfering RNA (siRNA) knockdown of MITF was performed using Silencer Select Validated siRNA CCGUGGACUAUAUCCGAAATT (ThermoFisher #s8790). Silencer Select Negative Control No. 1 siRNA (ThermoFisher #4390843) was used as a negative control. Media was changed on the cells 1 hour prior to transfection (2 mL media for 6 well plate). For each well, 150 µL OptiMEM (ThermoFisher #11058021) was combined with 18 µL Lipofectamine RNAiMax (ThermoFisher #13778030). In a separate tube, 150 µL OptiMEM was combined with 50 pmol total siRNA (desired combination of scramble and siMITF). The two tubes were combined, vortexed, and incubated for 5 minutes prior to application to cells. After 72 hours of growth in the presence of siRNA, cells were lysed in RIPA buffer supplemented with 1x HALT protease cocktail (ThermoFisher #87786) and prepared for analysis via Western blot.

5.3 Results

5.3.1 Characterizing Resistance to Binimetinib

Two NRAS mutant melanoma cell lines were acquired that demonstrate different responses to binimetinib. The MelJuso cell line is characterized to be resistant to binimetinib-induced cell death, whereas the MM415 cell line has been shown to be sensitive to the drug. To quantify their sensitivities, a viability assay was performed to measure cell death at 1 nM, 10 nM, 100 nM, 1 µM, and 10 µM at 72 hours after treatment. MelJuso cells were indeed more resistant to binimetinib, with an estimated IC_{50} of 120 nM compared to the MM415 cells which had an estimated IC_{50} of about 2.6 nM (Figure 5-1). The relative viabilities for 10 nM and 1 µM concentrations were 86% and 56%,

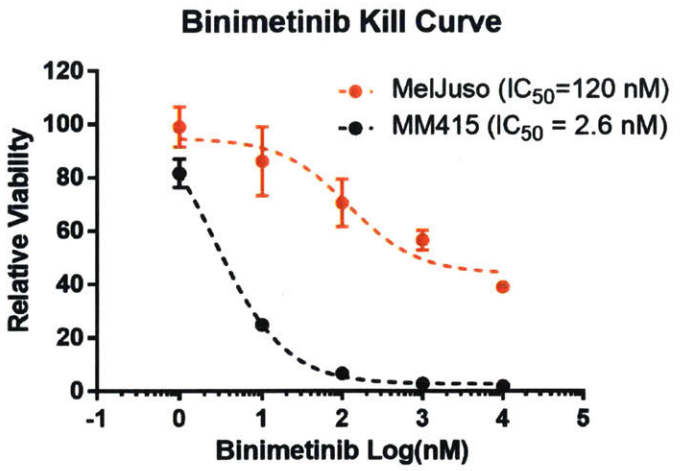


Figure 5-1 – Comparing the responses of MM415 and MelJuso cells across a range of binimetinib concentrations revealed that the MM415 cells show greater sensitivity to the drug than the MelJuso cells.

respectively, for the MelJuso cells, and 25% and 3%, respectively, for the MM415 cells. These values were used in the response matrix in the PLSR model.

5.3.2 Comparing Binimetinib-Induced Changes in Protein Synthesis

We next sought to determine potential mechanisms of resistance by comparing binimetinib induced protein synthesis between the resistant and sensitive cell lines. Two different concentrations, 10 nM and 1 μM, were applied to each cell line for 24 hours. After 24 hours, cells were labeled with Aha/K8/R10 for 30 minutes and lysed as described above. Across 3 replicates, a total of 2606 unique proteins were identified, with 1633 identified in 2 or more replicates. To account for differences in baseline protein synthesis in each cell line, all values were normalized to the respective DMSO controls for each cell line and log2 transformed.

After hierarchical clustering, the samples grouped by binimetinib concentration rather than by cell line, showing that in most cases, the response in protein synthesis was largely in the same direction in the two cell lines (Figure 5-2, left). When comparing the difference in magnitude of change between the two cell lines, there was a clear difference

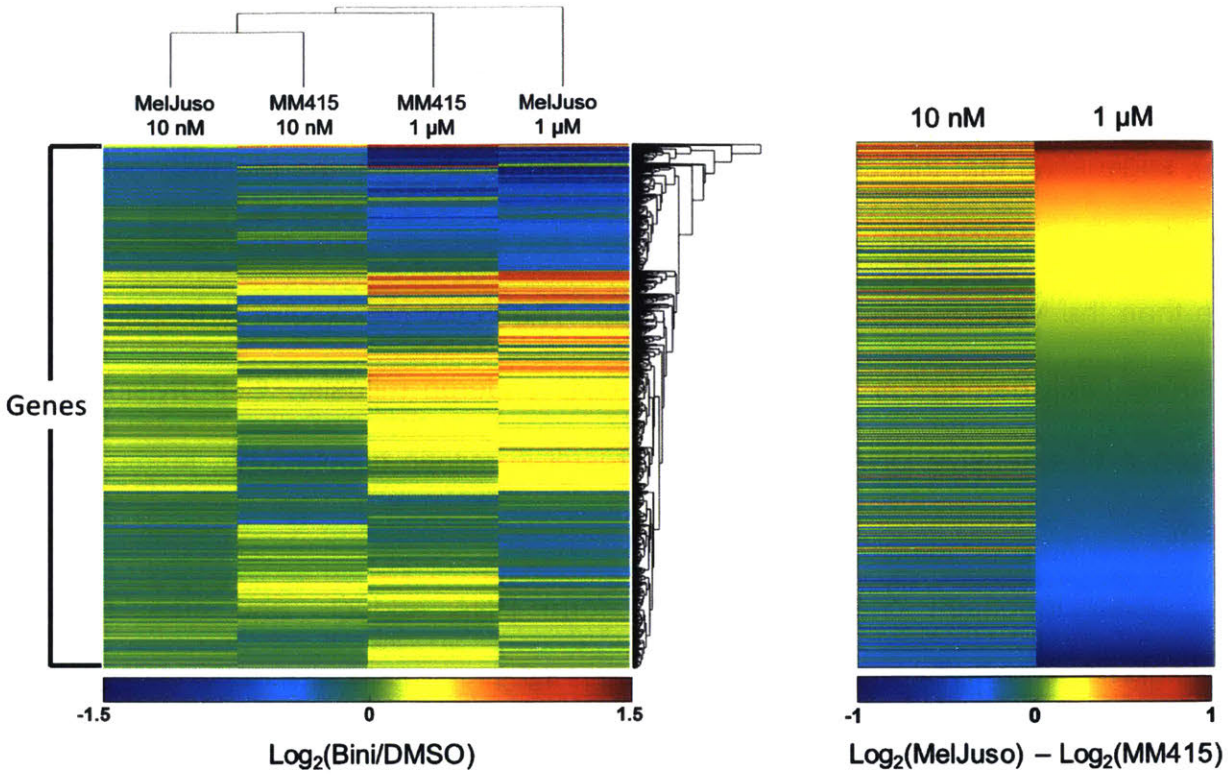


Figure 5-2 – 24 hours after binimetinib treatment, changes in protein synthesis between the MM415 and MelJuso cells were similar in direction (left), but not in degree (right). Furthermore, the magnitude of change in protein synthesis was less when treated with 10 nM compared to treatment with 1 μM.

in degree of upregulation or downregulation of protein synthesis between the two cell lines (Figure 5-2, right). This ranked list of the difference in protein synthesis between the two cell lines were the input into subsequent GSEA analyses.

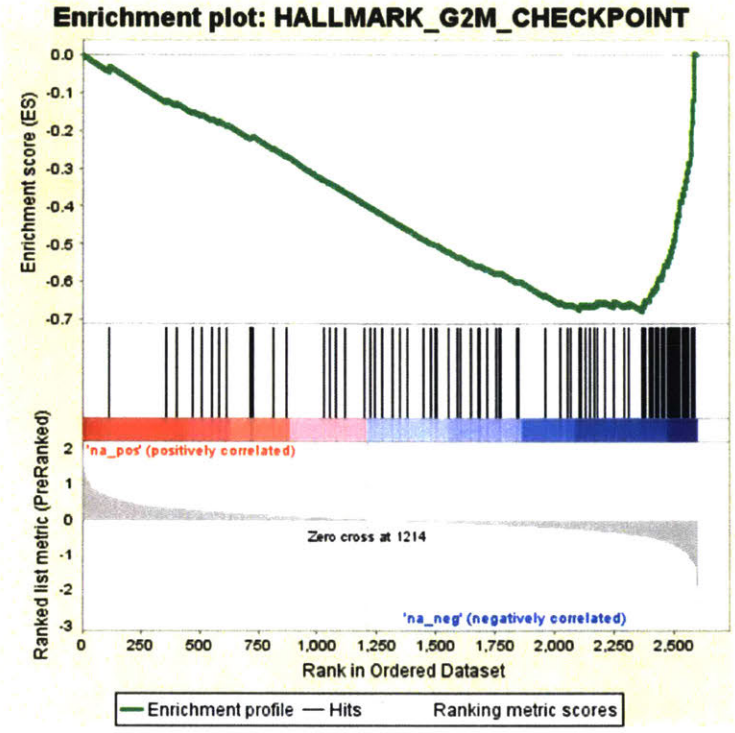
5.3.3 Identification of Drug Target Candidates via GSEA

To identify pathways that might be differentially regulated between the two cell lines, GSEA was used to search for enrichment of gene set members in the ranked list of differential responses between the two cell lines shown in Figure 5-2. The Hallmarks MSigDB[18] was used because of its inclusion of gene sets from diverse biological pathways and processes. The 1 μM concentration was used for GSEA due to the larger changes in protein synthesis compared to the 10 nM concentrations.

Components of the cell cycle G2/M checkpoint were strongly enriched in the MM415 cells compared to the MelJuso cells following binimetinib treatment (Figure 5-3, top), suggesting that this checkpoint may be downregulated in the MelJuso cells. This observation is consistent with the observation that MelJuso cells proliferate at a faster rate than the MM415 cells (data not shown). The G2/M checkpoint is responsible for ensuring DNA damage is repaired before cells undergo mitosis. Cell division in the presence of extensive DNA damage may result in mitotic catastrophe, leading to cell death[20].

Platinum-based chemotherapeutics form DNA adducts and crosslink DNA molecules[21]. In rapidly proliferating cells, such as cancer cells, the G2/M checkpoint is often dysfunctional, resulting in progression through mitosis in the presence of DNA adducts leading mitotic catastrophe and cell death. To test if the binimetinib-resistant MelJuso cells are susceptible to DNA crosslinking following binimetinib treatment, cells were challenged with cisplatin during binimetinib treatment. Cisplatin and binimetinib were co-administered over a large range of concentrations (1 nM to 10 μM for cisplatin, 1 nM to 1 μM for binimetinib). Cisplatin was freshly dissolved to 10 mM in DMF, diluted to 1 mM in PBS, and subsequent serial dilutions were performed in DMF. Binimetinib was prepared from a 1 mM stock solution and serially diluted in DMSO. After 72 hours of exposure to a wide range of combinations of cisplatin and binimetinib, cisplatin did not appear to have a synergistic effect with binimetinib specific to the MelJuso cells (Figure 5-3, bottom).

GSEA analysis also revealed that proteins associated with the epithelial-to-mesenchymal transition (EMT) may be upregulated in MelJuso cells following binimetinib



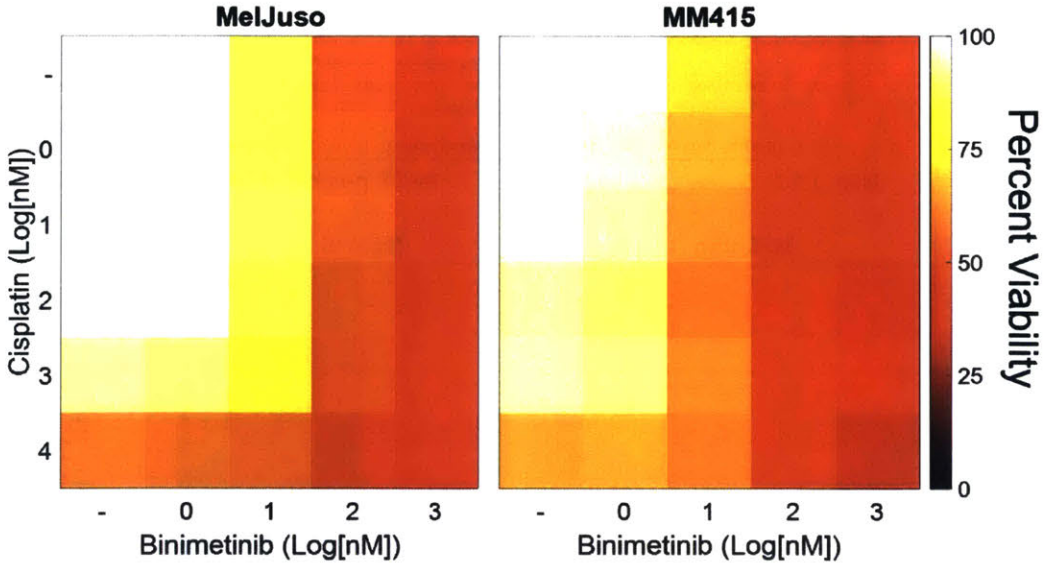


Figure 5-3 - GSEA analysis suggests that the cell cycle G2/M checkpoint may be suppressed in MelJuso cells following binimetinib treatment, resulting in a vulnerability to DNA crosslinking chemotherapeutics. However, co-treatment with cisplatin did not appear to result in synergistic cell death in MelJuso cells.

treatment (Figure 5-4, top). Cells undergoing EMT often acquire resistance to targeted therapeutics[22], so it may be feasible that binimetinib treatment pushed MelJuso cells to undergo an EMT-like transition to acquire resistance. Oftentimes, EMT is mediated by signaling through TGFβ[23], and inhibiting TGFβ signaling pathways can block EMT and metastasis[24].

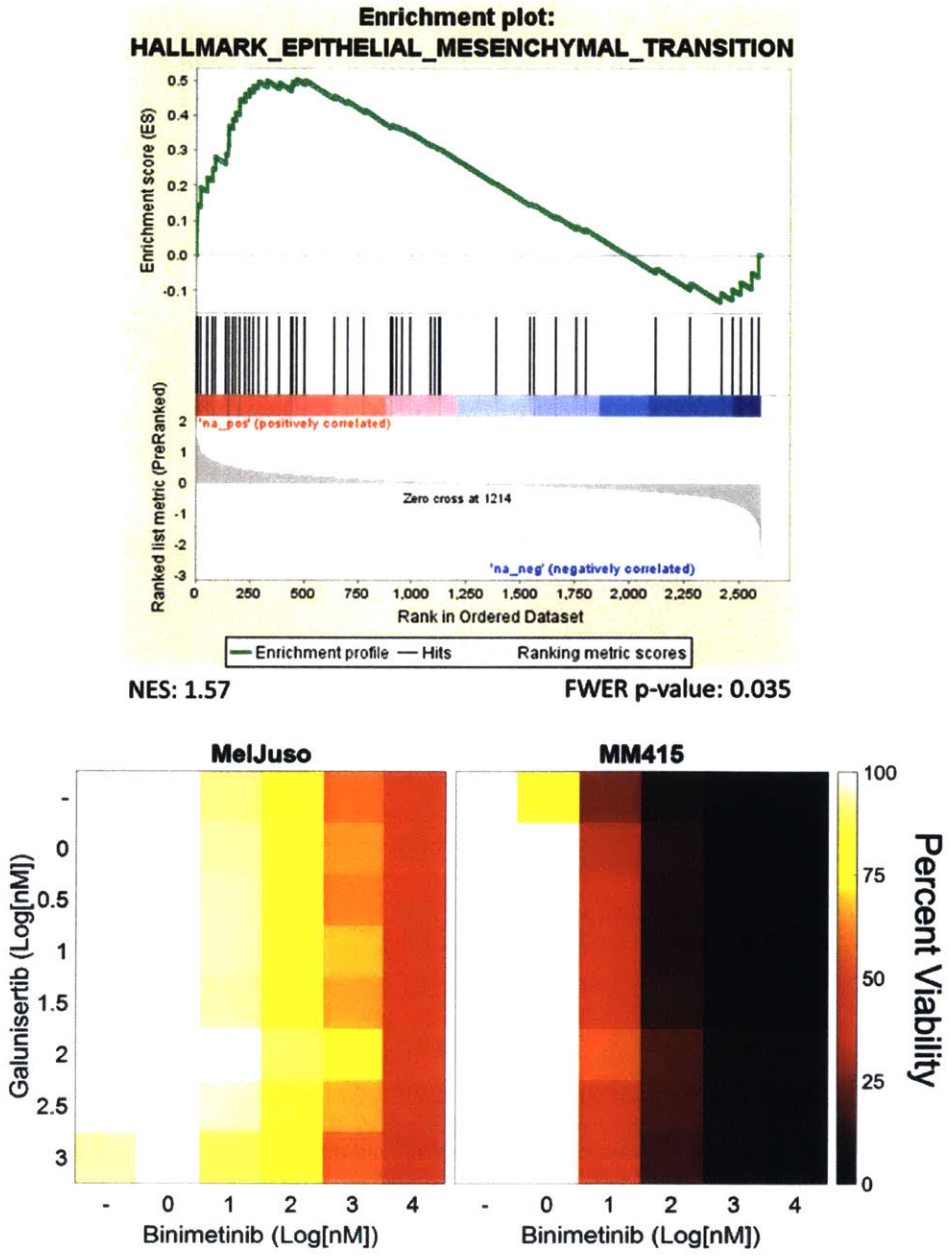


Figure 5-4 - GSEA analysis suggests that MelJuso cells may have been undergoing an EMT-like transition in response to binimetinib treatment. However, attempts to block EMT by inhibiting TGFβ signaling had no effect on either cell line.

To test if MelJuso cells could be sensitized by blocking EMT, cells were treated with a combination of binimetinib and galunisertib, an inhibitor of TGFβRI. Binimetinib was tested from 1 nM to 10 μM, and galunisertib was tested from 1 nM to 1 μM. After 72 hours, galunisertib did not appear to have any effect on cell viability regardless of cell type or binimetinib concentration (Figure 5-4, bottom).

To further investigate whether MelJuso cells were indeed undergoing an EMT in response to binimetinib treatment, Western blotting was performed against markers of EMT including E-Cadherin, N-Cadherin, Vimentin, and Zeb1 24 hours after binimetinib treatment at 10 nM and 1 μM. Surprisingly, MelJuso cells appeared to already be in a mesenchymal-like state with elevated N-Cadherin (and undetectable E-cadherin), Zeb1, and Vimentin. This state was in contrast to the epithelial-like state of MM415 cells, which had elevated E-cadherin, and undetectable N-Cadherin and Zeb1. However, MM415 cells did express Vimentin (Figure 5-5). The total amount of these proteins did not appear to change in response to binimetinib treatment, with the exception of Zeb1, which decreased in a dose-dependent manner. This decrease was consistent with previous observations that MEK inhibition can result in loss of Zeb1 expression (unpublished).

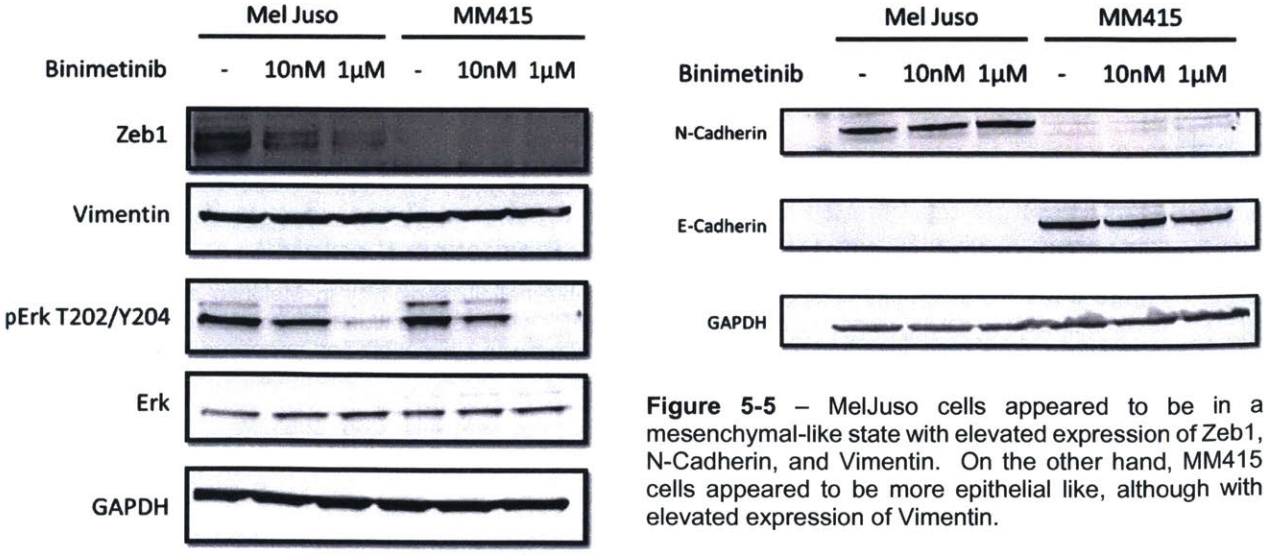


Figure 5-5 – MelJuso cells appeared to be in a mesenchymal-like state with elevated expression of Zeb1, N-Cadherin, and Vimentin. On the other hand, MM415 cells appeared to be more epithelial like, although with elevated expression of Vimentin.

Given that mesenchymal-like cells often acquire drug resistance, and given that the binimetinib-resistant MelJuso cells exhibit a mesenchymal-like phenotype, perhaps attempts to convert MelJuso cells into a more epithelial-like state may result in sensitivity to binimetinib. Reports have indicated that metformin is capable of reversing EMT and multidrug resistance associated with a mesenchymal state[25,26]. To test if metformin can reverse EMT and sensitize MelJuso cells to binimetinib, the combinations of metformin (ranging from 1 μM to 10 mM) and binimetinib (1 nM to 1 μM) were applied to MelJuso and MM415 cells for 72 hours before measuring viability. However, metformin appeared to have no synergistic effect with binimetinib in the in the range of concentrations tested, and the toxicity profiles on both cell lines appear to be identical.

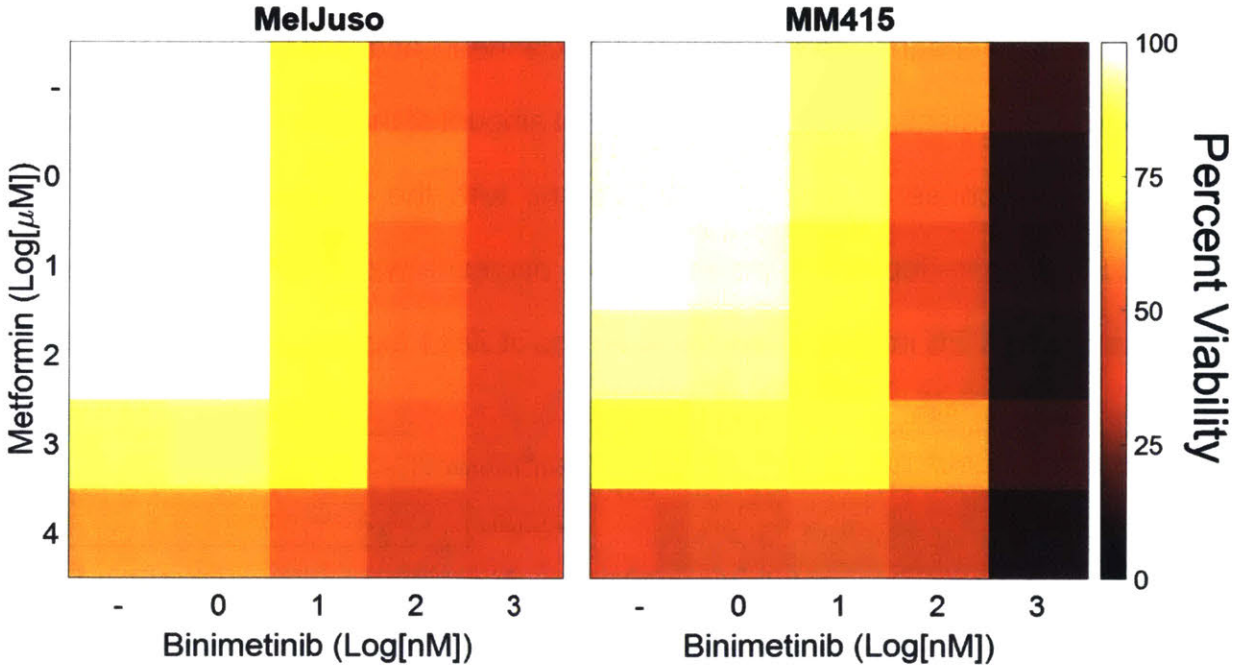


Figure 5-6 – To determine if metformin can induce binimetinib sensitivity in MelJuso cells by reversing EMT, viability was measured after treating with combinations of metformin and binimetinib. Metformin appeared have no synergistic effect with binimetinib and appeared to have the same toxicity on both cell lines.

5.3.4 Identification of Drug Target Candidates via PLSR

Efforts to identify and overcome potential resistance mechanisms via GSEA thus far had not yielded any treatment combinations that effectively overcome binimetinib resistance in the MelJuso cell line. GSEA is useful for identifying pathways that are enriched in the top or bottom of ranked lists, but makes no relation to the degree of resistance shown in these cells. In GSEA, associating a particular gene set with resistance or sensitivity depends on whether that gene set is enriched towards the top or bottom of the list. However, resistance is not a binary response, but rather graded behavior that can vary depending on cell line and drug concentration. PLSR models are better suited for such applications by correlating changes in a set of predictors (protein synthesis) to changes in a response (viability). Therefore, a PLSR model was employed in an effort to better relate changes in protein synthesis to binimetinib resistance (Figure 5-7, left). Details on the parameters of the model are described in section 5.2.7.

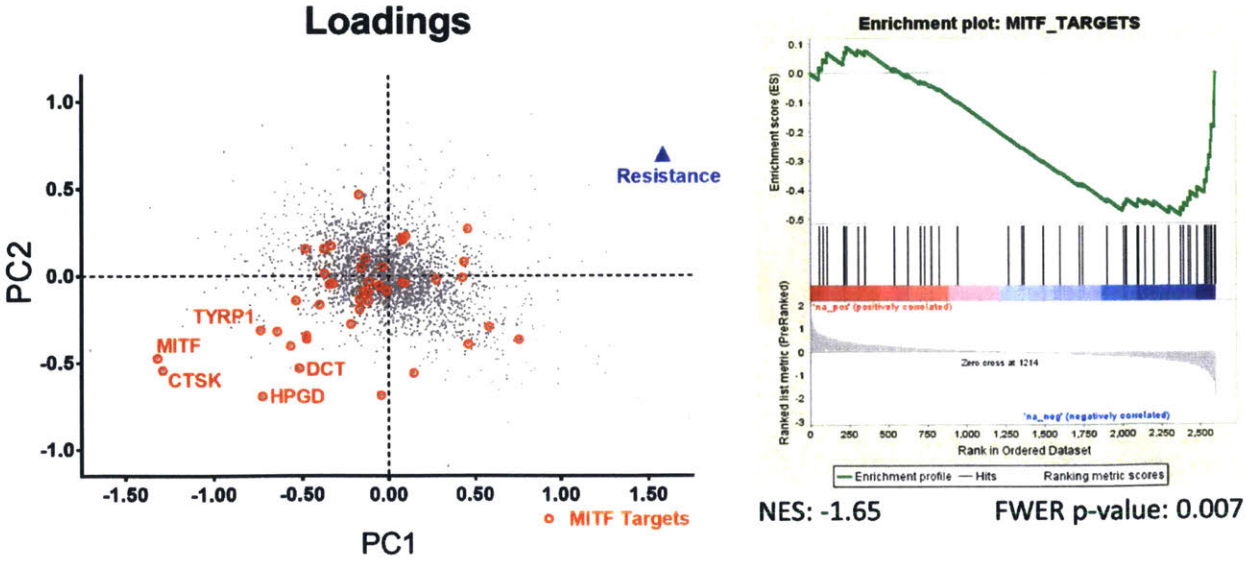


Figure 5-7 – PLSR modeling revealed an anti-correlation between resistance to binimetinib and synthesis of MITF targets, as well as synthesis of MITF itself. These findings were corroborated by GSEA against a list of MITF targets, showing an enrichment of these targets in the MM415 cells (as a depletion of the MelJuso:MM415 ratio).

One of the most notable features of the PLSR model was the anti-correlation between the microphthalmia-associated transcription factor (MITF) and associated targets (*e.g.* HPGD, DCT, TYRP1) with cell viability (Figure 5-7, left). This finding was further supported by GSEA using a gene set of known MITF targets[19] (Figure 5-7, right). MITF is a master regulator of pigment production and differentiation in melanocytes[27], and has also been shown to play a role in the development of malignant melanoma[28]. The role of MITF as an oncogene appears to be contrary to observations that increased MITF expression is a predictor of binimetinib sensitivity. However, another study in BRAF mutant melanoma demonstrated that a MITF low/NFκB high transcription state predicted resistance to MAPK pathway inhibitors, and that a MITF high/NFκB low state predicted sensitivity[29]. Furthermore, this study demonstrated that induction of NFκB with TNFα in the resistant MITF high/NFκB low cells was sufficient to suppress MITF and upregulate resistance markers such as AXL, demonstrating that NFκB induction is driver of this state

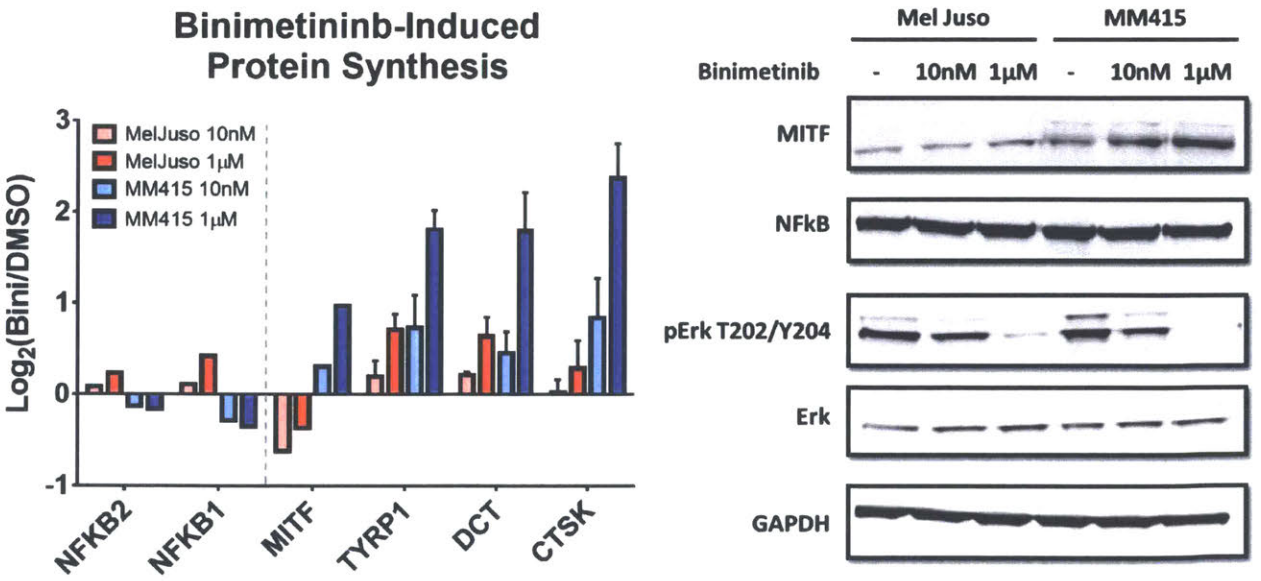


Figure 5-8 – Changes in protein synthesis of NFKB, MITF, and MITF targets show contrasting behaviors between the two cell lines in response to binimetinib. While MITF levels increased in the MM415 cells following binimetinib treatment, there did not appear to be any change in NFKB protein levels despite modest changes in synthesis.

change[29]. Consistent with this finding, NFκB1 and NFκB2 appeared to become upregulated in the resistant MelJuso cells and downregulated in the sensitive MM415 cells, although only in a single replicate (Figure 5-8, left). However, examining the total protein pool by Western blotting showed that NFκB levels appear equal between cell lines and remain unchanged 24 hours after binimetinib treatment. Interestingly, MITF levels appeared elevated in MM415 cells compared to MelJuso cells at basal state, and appeared to increase following binimetinib treatment in MM415 cells (Figure 5-8, right).

To test if the MelJuso cells exhibited a MITF low/NFκB high state, and if this state was driving resistance to binimetinib, cells were co-treated with the NFκB inhibitor SC75741 (1 nM to 10 μM) and binimetinib (1 nM to 1 μM). Interestingly, SC75741 appeared to cause greater cell death in MelJuso cells regardless of binimetinib concentration. Furthermore, SC75741 appeared to induce resistance to binimetinib in the MM415 cells at concentration of 1 μM and above (Figure 5-9). The mechanism behind

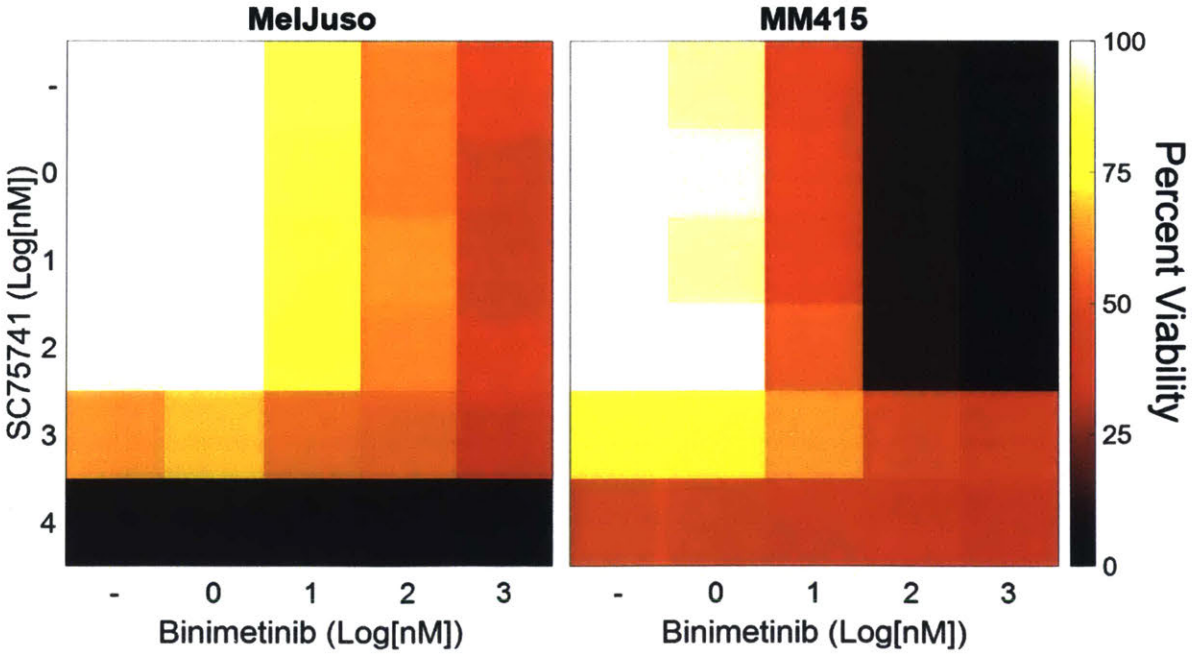


Figure 5-9 – In an attempt to shift cells from a potential MITF low/NFκB high state associated with resistance against MAPK pathway inhibitors, cells were co-treated with NFκB inhibitor SC75741 and binimetinib. Interestingly, NFκB inhibition appeared to promote resistance to binimetinib in MM415 cells.

this behavior is not understood. However, SC75741 did not sensitize MelJuso cells to binimetinib.

If MITF is in fact the driver behind MM415 sensitivity to binimetinib, then blocking the activity of MITF should result in resistance in MM415 cells. One study performed a small molecule screen to find an MITF inhibitor. This screen yielded a compound, ML329, capable of reducing the transcription of MITF target genes and exhibited toxicity only in cell lines dependent on MITF[30]. Due to the ease of using a small molecule inhibitor compared to an siRNA knockdown, ML329 was used as an initial screen to see if MITF inhibition can induce resistance in MM415 cells.

Interestingly, ML329 resulted in increased cell proliferation in the absence of binimetinib in the MM415, but not MelJuso cells. This effect was observed at concentrations up to 100nM, but toxicity was observed at higher concentrations in both cell lines. Furthermore, at higher concentrations of binimetinib (100 nM and higher), 100

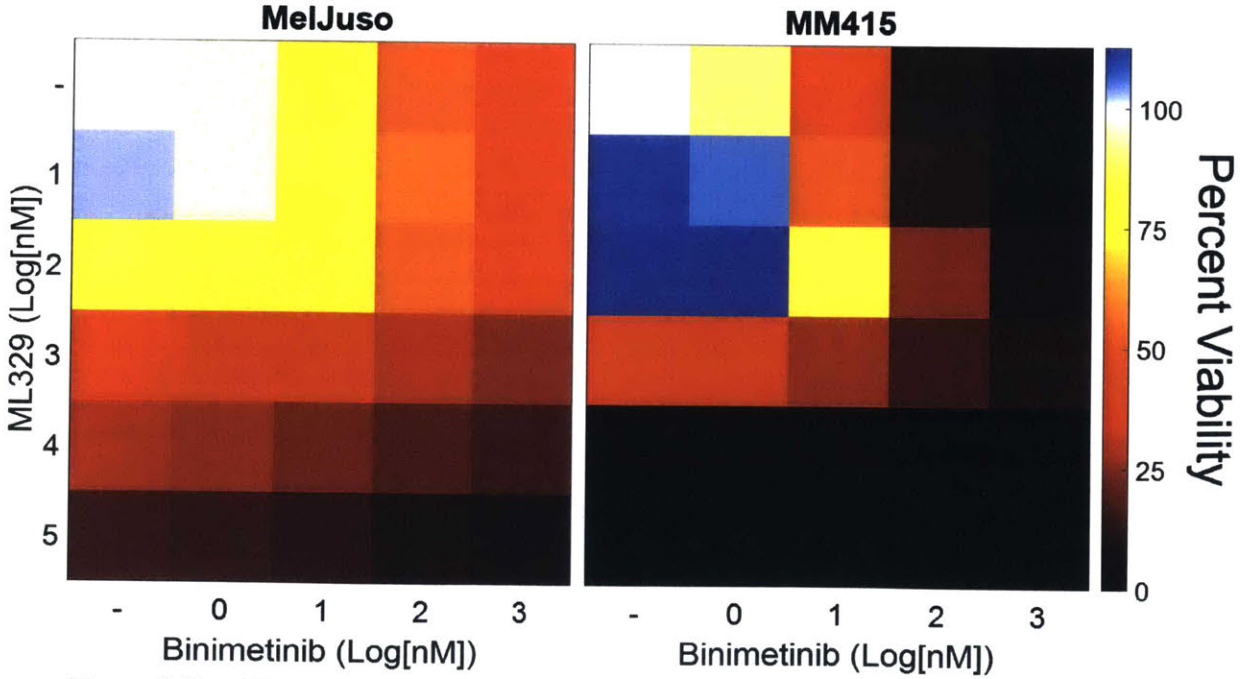


Figure 5-10 – MITF inhibitor ML329 appeared to increase cell proliferation in MM415, but not MelJuso cells. Furthermore, ML329 at concentrations around 100 nM may have resulted in a protective effect against binimetinib treatment

nM ML329 appeared to have a slightly protective effect in MM415 cells (Figure 5-10). These preliminary results suggest that MITF may be a driver of binimetinib sensitivity in MM415 cells.

The mechanism of action of ML329 to inhibit MITF activity is not known, and the off-target effects of ML329 have not been characterized. The screening process for identifying ML329 involved selecting for its action against MITF, but the specificity of the compound was not tested[31]. Therefore, to truly validate that MITF plays a role in sensitizing MM415 cells to binimetinib, siRNA knockdown of MITF was performed. MITF siRNA was tested at 0, 5, 10, 25, and 50 pmol per well of a 6 well plate with a complementary amount of scramble siRNA used as a control (at 50, 45, 10, 25, and 0 pmol, respectively). After 72 hours of incubation in the presence of siRNA, cells were lysed, and a Western blot was performed to examine MITF protein levels. While robust knockdown was observed in the MelJuso cells with as little as 5 pmol siRNA, knockdown was not observed in the MM415 cells (Figure 5-11). The lack of efficient knockdown in MM415 cells may be due to the rapid turnover of MITF proteins, or perhaps MM415 cell

Figure 5-11 – While robust MITF knockdown was observed in MelJuso cells, MM415 cells were resistant to MITF knockdown by siRNA at levels as high as 50 pmol per well.

are dependent upon MITF and therefore counteract siRNA knockdown by some other mechanism.

Because siRNA-mediated knockdown of MITF failed specifically in MM415 cells, confirmation of MITF as the mediator of sensitivity in MM415 cells was not possible. ML329-induced proliferation and protection against binimetinib suggests that MITF plays a key role in mediating binimetinib sensitivity, but further tools to inhibit MITF must be developed before this hypothesis can be validated. Future work investigating MITF could use CRISPR to generate MITF knockout MM415 and MelJuso cell lines. However, if MM415 cells are dependent upon MITF, then MITF knockout may result in cell death. Another avenue for future studies could involve overexpression of MITF in MelJuso cell lines in an attempt to induce sensitivity to binimetinib. While this approach may not be feasible in a therapeutic context, it may yield further insight into the mechanism of drug resistance in these cell lines.

5.5 Conclusions

In this chapter, I investigated the utility of BONCAT to characterize potential mechanisms of binimetinib resistance by comparing protein synthesis between a sensitive and resistant NRAS mutant melanoma cell line. Protein synthesis was assayed at 24 hours after binimetinib treatment, which is 48 hours before most previous phenotypic measurements have been examined. By profiling protein synthesis at this early time point, I sought to observe changes that may indicate the rise of an adaptive resistance mechanism.

Indeed, binimetinib-induced changes in protein expression were observed at 24 hours after treatment. Furthermore, some proteins showed divergent responses between the sensitive MM415 cell line and MelJuso cell line, indicating that they may play a role in resistance to binimetinib. Using a ranked list of the differential response in protein expression between the two cell lines, GSEA was employed to identify several pathways to may be targeted in binimetinib resistant cell lines. However, targeting these pathways did not appear to result in cytotoxicity in the resistant MelJuso cells. Using PLSR to correlate quantitative measurements of binimetinib resistance to changes in protein synthesis, the expression of MITF and MITF targets were observed to be anti-correlated with binimetinib resistance. Inhibition of MITF using a novel but poorly-characterized inhibitor appeared to reduce the toxicity of binimetinib to some degree in MM415 cells, but further studies with more established tools are required to validate this result. Knockdown of MITF was attempted using siRNA, but MM415 cells appeared to be resistant to siRNA-mediated MITF knockdown. Future studies investigating the role of MITF in mediating binimetinib resistance could use CRISPR knockout cell lines or inducible MITF overexpression constructs.

While MITF was identified as a potential mediator of binimetinib sensitivity, the search for a *bona fide* mechanism of resistance was largely unsuccessful. Analysis of changes in protein expression between the two cell lines yielded useful insights into how these cells differ in their response to binimetinib. However, protein expression is merely one dimension of a complex cellular response. Binimetinib targets MEK, a kinase involved in the MAPK signaling pathway that transmits information via the phosphorylation and dephosphorylation of proteins. Solely examining changes in protein

synthesis ignores any potential differences in signaling between these two cell lines. If resistance arises due to the rewiring of signaling networks, then the analysis of protein expression would not be expected to identify that mechanism. Future studies profiling protein synthesis by BONCAT can be paired with phosphoproteomics and other systems biology approaches to more comprehensively characterize changes in cell biology that can lead to drug resistance and other pathologies.

5.5 References

1. Rigel, D. S. & Carucci, J. A. Malignant melanoma: prevention, early detection, and treatment in the 21st century. *CA. Cancer J. Clin.* **50,** 215-36-40 (2000).
2. Mashima, E. *et al.* Nivolumab in the treatment of malignant melanoma: Review of the literature. *OncoTargets and Therapy* **8,** 2045–2051 (2015).
3. H. Davies, G.R. Bignell, C. Cox, P. Stephens, S. Edkins, S. Clegg, et al. Mutations of the BRAF gene in human cancer. *Nature* **417,** 949–954 (2002).
4. Bos, J. L. Ras Oncogenes in Human Cancer : A Review. *CANCER Res.* **49,** 4682–4689 (1989).
5. Chapman, P. B. *et al.* Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *N. Engl. J. Med.* **364,** 2507–16 (2011).
6. Flaherty, K. T. *et al.* Inhibition of Mutated, Activated BRAF in Metastatic Melanoma. doi:10.1056/NEJMoa1002011
7. Lim, S. Y., Menzies, A. M. & Rizos, H. Mechanisms and strategies to overcome resistance to molecularly targeted therapy for melanoma. *Cancer* **123,** 2118–2129 (2017).
8. Ascierto, P. A. *et al.* MEK162 for patients with advanced melanoma harbouring NRAS or Val600 BRAF mutations: A non-randomised, open-label phase 2 study. *Lancet Oncol.* **14,** 249–256 (2013).
9. Long, G. V *et al.* Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. *N. Engl. J. Med.* **371,** 1877–88 (2014).
10. Larkin, J. *et al.* Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N. Engl. J. Med.* **373,** 23–34 (2015).
11. Poulikakos, P. I., Zhang, C., Bollag, G., Shokat, K. M. & Rosen, N. RAF inhibitors transactivate RAF dimers and ERK signaling in cells with wild-type BRAF. doi:10.1038/nature08902
12. DeLeon, T. *et al.* Assessment of treatment response to immunotherapy in melanoma patients with pathogenic mutations of NRAS, BRAF, CDKN2A and P53. *J. Clin. Oncol.* **35,** (2017).
13. Dummer, R. *et al.* Binimetinib versus dacarbazine in patients with advanced NRAS-mutant melanoma (NEMO): a multicentre, open-label, randomised, phase

3 trial. *Lancet Oncol.* **18,** 435–445 (2017).
14. Wacker, S. A., Houghtaling, B. R., Elemento, O. & Kapoor, T. M. Using transcriptome sequencing to identify mechanisms of drug action and resistance. *Nat. Chem. Biol.* **8,** 235–7 (2012).
15. Holcik, M., Yeh, C., Korneluk, R. G. & Chow, T. Translational upregulation of X-linked inhibitor of apoptosis (XIAP) increases resistance to radiation induced cell death. *Oncogene* **19,** 4174–4177 (2000).
16. Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485,** 55–61 (2012).
17. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–50 (2005).
18. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1,** 417–425 (2015).
19. Hoek, K. S. *et al.* Novel MITF targets identified using a two-step DNA microarray strategy. *Pigment Cell Melanoma Res.* **21,** 665–676 (2008).
20. Cuddihy, A. R. & O'Connell, M. J. Cell-cycle responses to DNA damage in G2. *Int. Rev. Cytol.* **222,** 99–140 (2003).
21. Roberts, J. J. & Pascoe, J. M. Cross-linking of Complementary Strands of DNA in Mammalian Cells by Antitumour Platinum Compounds. *Nature* **235,** 282–284 (1972).
22. Singh, A. & Settleman, J. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene* **29,** 4741–4751 (2010).
23. Zavadil, J. & Böttinger, E. P. TGF-beta and epithelial-to-mesenchymal transitions. *Oncogene* **24,** 5764–74 (2005).
24. Bandyopadhyay, A. *et al.* Inhibition of pulmonary and skeletal metastasis by a transforming growth factor-beta type I receptor kinase inhibitor. *Cancer Res.* **66,** 6714–21 (2006).
25. Qu, C. *et al.* Metformin reverses multidrug resistance and epithelial–mesenchymal transition (EMT) via activating AMP-activated protein kinase (AMPK) in human breast cancer cells. *Mol. Cell. Biochem.* **386,** 63–71 (2014).
26. Cufí, S. *et al.* Metformin against TGFβ-induced epithelial-to-mesenchymal transition (EMT): From cancer stem cells to aging-associated fibrosis. *Cell Cycle* **9,** 4461–4468 (2010).
27. Goding, C. R. Mitf from neural crest to melanoma: signal transduction and transcription in the melanocyte lineage. *GENES Dev.* **14,** 1712–1728 (2000).
28. Levy, C., Khaled, M. & Fisher, D. E. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends in Molecular Medicine* **12,** 406–414 (2006).
29. Konieczkowski, D. J. *et al.* A melanoma cell state distinction influences sensitivity to MAPK pathway inhibitors. *Cancer Discov.* **4,** 816–827 (2014).
30. Faloon, P. W. *et al.* in *Probe Reports from the NIH Molecular Libraries Program* (National Center for Biotechnology Information (US), 2010). at <http://www.ncbi.nlm.nih.gov/pubmed/24027801>
31. Frank, W. *et al.* A Small Molecule Inhibitor of the MITF Molecular Pathway. *Mol. Libr.* 1–115 (2010). at <http://www.ncbi.nlm.nih.gov/pubmed/24027801>

Chapter VI

Conclusions and Perspectives

Daniel Rothenberg

6.1 Summary and Conclusions

Cell biology is not the observation of a static state, but rather the study of a dynamic system that is constantly reading a myriad of environmental cues, integrating them through a complex signaling cascade, and responding through the modulation of gene expression to adapt to those inputs. Recent advances in biotechnology have developed platforms for robust, comprehensive analysis of many dimensions of cell biology, ranging from genomics and transcriptomics to proteomics and metabolomics. In the context of measuring changes in protein production, transcriptomics measures changes in mRNA profiles as a proxy for the proteins that are being produced at that moment. However, the correlation between RNA abundance and protein synthesis varies depending on cellular context[1]. A more specialized method of ribosome footprinting reads and measures the mRNA sequences that are bound as ribosomes[2], but even this approach isn't a perfect reflection of protein synthesis[3]. On the proteomics side, shotgun proteomics can measure changes in protein abundance between two conditions, but requires exquisite sensitivity to detect changes in the background of the existing proteome. Metabolic labeling schemes such as pSILAC can measure protein turnover and compare protein synthesis between two conditions, but again requires sifting through the existing proteome to make these measurements[4].

When it comes to directly and accurately measuring newly synthesized proteins, BONCAT is the current state of the art. By incorporating a chemical handle and enriching newly synthesized proteins from the background of the entire proteome, the nascent proteome can be profiled with a high degree of sensitivity. However, the nature of metabolic labeling requires a trade-off between temporal resolution and coverage. The

longer cells are allowed to incorporate Aha into proteins, the more material there is available for enrichment, and the greater the coverage of the nascent proteome. However, gene expression is dynamic, and long labeling times may result in missing rapid changes in synthesis that get averaged out over the course of labeling. Furthermore, current quantification methods for comparing multiple conditions rely on label-free quantitation[5,6], which is not particularly robust, or unique SILAC channels (as in QuanCAT[7], BONLAC[8], or HILAQ[9]), limiting multiplexing to two or three conditions.

In this thesis, I adapted the BONCAT protocol for a high degree of multiplexing using isobaric mass tagging. Furthermore, I improved the sensitivity to allow for analysis of samples labeled for durations as short as 15 minutes. By pairing this improved sensitivity and temporal resolution with highly multiplexed quantitation, I was able to apply BONCAT to time course analyses in response to a number of different perturbations. Furthermore, I was able to directly compare protein synthesis across multiple conditions and cell lines in an effort to gain insight into drug resistance mechanisms.

The robustness of the method was validated in the context of the unfolded protein response (UPR). The UPR was an attractive application for validation because of its well characterized response and regulation of protein translation. In response to ER stress, cells downregulate protein synthesis at the level of translation, but upregulate several key stress response proteins. Indeed, I observed global downregulation of protein synthesis, even in the absence of downregulation of transcription, validating translational control. Furthermore, stress response proteins involved in protein folding, protein degradation, and survival were observed to be upregulated. Together, these observations

demonstrate the reliability of the multiplexed BONCAT protocol for quantifying changes in complex biological systems.

Previous studies analyzing the transcriptional response to EGFR activation have found genes cluster in temporally distinct waves. To compare the transcriptional and translational responses, BONCAT was used to measure protein synthesis in 30 minute intervals following EGF stimulation. We found that there are indeed temporally distinct waves of translation, but the members of each group differ between transcriptional and protein synthesis analyses. Increasing the resolution to 15 minute intervals did not appreciably change the clustering analysis. However, some proteins were observed to be upregulated as soon as 15 minutes after EGF stimulation. Furthermore, the increased temporal resolution allowed for better characterization of times of peak synthesis and upregulation of late response genes.

Direct comparisons to RNA abundance and ribosome footprints revealed two distinct modes of gene regulation. Proteins showing a transient response with large changes in expression correlated strongly with RNA abundance, suggesting control at the level of transcription. Translational machinery components more strongly correlated with increase ribosome binding, indicating control at the level of translation. Furthermore, analysis of the temporal relationship between ribosome binding and protein synthesis stratified proteins into two classes. One of the classes exhibited some delay between either RNA transcription and ribosome binding, or ribosome binding and protein translation. This class also showed significant bias in codon usage frequency, suggesting that codon bias may play a role in regulating expression of these genes. On the other hand, the other class was characterized by overlapping ribosome binding and translation,

and this class of transcripts did not have a statistically significant deviation in codon bias from the genome average. Furthermore, global codon usage was found to by dynamic over the times assayed, and this observation was matched with changes in chemical modifications of tRNA isoacceptors that could affect preferential codon usage.

Finally, BONCAT was applied to compare the translational responses to binimetinib between a resistant and sensitive NRAS mutant melanoma cell line. A comparison of protein synthesis between the two cell lines revealed changes in expression patterns unique to each cell line. Using a PLSR model to correlate resistance and protein synthesis, MITF was identified as a potential marker of sensitivity. While initial validation experiments with a novel small molecule inhibit of MITF yielded promising results, attempted siRNA knockdown of MITF failed in the sensitive cell line. Further tools need to be developed to robustly manipulate MITF expression and activity.

6.2 Perspectives and Future Work

Nucleic acid-based technologies are frequently used to measure changes in gene expression, but these assays oftentimes do not account for regulatory steps that occur throughout the process of transcription and translation. Proteomics based methods measure the output of gene expression, but lack the sensitivity to comprehensively measure new protein synthesis in the background of the genome. BONCAT enriches for newly translated proteins resulting in excellent sensitivity and coverage while integrating all the regulatory steps that may not be accounted for with sequencing-based methods. By robustly measuring the output of gene expression, future studies may be able to build a quantitative model of gene expression that incorporates variables such as regulation

through transcription, ribosome assembly and binding, codon bias, and tRNA isoacceptor availability.

While the method developed in chapter 2 was suitable for the purposes of this thesis, improvements can still be made to the protocol. Future work in protocol development can optimize the amount of SILAC amino acids to add to the media to maximize overlap between SILAC incorporation and Aha incorporation. Furthermore, even after fractioning the sample into less complex mixtures, coverage still seems to be limited by sample complexity rather than sensitivity, as evidenced from the lack of decrease in hits when changing from a 30 minute to 15 minute labeling window in chapter 4. Further optimizing fractionation -- whether it be through increasing the number of fractions, or selecting a more orthogonal fractionation scheme – should increase the coverage of this method. Finally, there are steps in the protocol that can be further optimized to reduce technical variability between samples. Using an automated platform with liquid-handling robots would increase the throughput of this method while reducing the amount of technical error introduced to the analysis.

The unique ability to analyze the output of protein synthesis makes BONCAT especially useful for studying translationally regulated disease models. Pathologies involving signaling through mTOR regulate exert control of translation by regulating assembly of translation initiation factors at the 5'-cap of mRNA transcripts[10]. For example, oncogenic signaling through mTOR has been demonstrated to specifically upregulate the translation of mRNA transcripts that result in increased invasion, metastasis, and metabolism in the context of prostate cancer[11]. In another disease model, Fragile X Syndrome results in mental retardation due to the silencing of FMRP, an RNA-binding

protein that regulates the translation of mRNAs involved in neuronal development. The result of FMRP silencing is an upregulation of the translation of these transcripts, resulting in a mental retardation phenotype[12]. BONCAT could be used to look for the upregulation of these transcripts in the absence of FMRP to better understand the mechanism that drives this mental retardation phenotype. This method could also be applied to studying changes in protein synthesis in infectious diseases. By pre-incubating either host cells or pathogen cells in Aha immediately prior to infection, changes in protein synthesis can be discerned between the host and the pathogen.

While BONCAT is an excellent method for measuring changes in gene expression and protein synthesis, but many diseases are driven by mechanisms other than changes in gene expression. By pairing BONCAT with complementary analytical methods such as phosphoproteomics or metabolomics, a more comprehensive portrait of cell state can be painted.

6.3 References

1. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165,** 535–550 (2016).
2. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science (80-.).* **324,** 218–223 (2009).
3. Liu, T. Y. *et al.* Time-Resolved Proteomics Extends Ribosome Profiling-Based Measurements of Protein Synthesis Dynamics. *Cell Syst.* **4,** 636–644.e9 (2017).
4. Schwanhäusser, B., Gossen, M., Dittmar, G. & Selbach, M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9,** 205–209 (2009).
5. Malmström, J. *et al.* Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **460,** 762–765 (2009).
6. Sadygov, R. G., Liu, H. & Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76,** 4193–201 (2004).
7. Howden, A. J. M. *et al.* QuaNCAT: quantitating proteome dynamics in primary cells. *Nat. Methods* **10,** 343–346 (2013).

8. Bowling, H. *et al.* BONLAC: A combinatorial proteomic technique to measure stimulus-induced translational profiles in brain slices. *Neuropharmacology* **100,** 76–89 (2016).
9. Ma, Y., McClatchy, D. B., Barkallah, S., Wood, W. W. & Yates, J. R. HILAQ: A Novel Strategy for Newly Synthesized Protein Quantification. *J. Proteome Res.* **16,** 2213–2220 (2017).
10. Gingras, A.-C., Raught, B. & Sonenberg, N. Regulation of translation initiation by FRAP / mTOR. *Genes Dev.* **15,** 807–826 (2001).
11. Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485,** 55–61 (2012).
12. Tabolacci, E., Palumbo, F., Nobile, V. & Neri, G. Transcriptional reactivation of the FMR1 Gene. A possible approach to the treatment of the fragile X syndrome. *Genes (Basel).* **7,** 1–16 (2016).