Title

A Liquid Chromatography-Mass Spectrometry Platform for Identifying Protein Targets of Small-Molecule Binding Relevant to Disease and Metabolism

Author

O'Brien Johnson, Reid

Publication Date

2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

A Liquid Chromatography-Mass Spectrometry Platform for Identifying Protein Targets of Small-Molecule Binding Relevant to Disease and Metabolism

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Biochemistry and Molecular Biology

by

Reid Lee O'Brien Johnson

2018

ABSTRACT OF THE DISSERTATION

A Liquid Chromatography-Mass Spectrometry Platform for Identifying Protein Targets of Small-Molecule Binding Relevant to Disease and Metabolism

by

Reid Lee O'Brien Johnson

Doctor of Philosophy in Biochemistry and Molecular Biology

University of California, Los Angeles, 2018

Professor Joseph Ambrose Loo, Chair

Identifying small-molecule binders to protein targets remains a daunting task due to the huge diversity in compound structure, activity, and mechanisms of action. Affinity-based target identification techniques are limited by the necessity to modify each drug individually (without losing bioactivity), while non-affinity based approaches are dependent on the drug's ability to induce specific biochemical/cellular readouts. To overcome these limitations, we have developed a high-throughput liquid chromatography-mass spectrometry (LC-MS) platform coupled to a universally applicable target identification approach that analyzes direct small-molecule binding to its protein target(s). DARTS (drug affinity responsive target stability) relies on a well-known phenomenon in which ligand binding causes thermodynamic stabilization of its target protein's structure such that the protein becomes resistant to a variety of insults, including proteolysis.

DARTS-MS is performed by screening for proteins that become resistant to proteolysis in the presence of the ligand. Protection from proteolysis is specific to the target protein(s) while proteolysis of non-target proteins is unchanged, allowing for unbiased, proteome-wide screening for small-molecule binders. Samples are proteolyzed and fractionated according to molecular weight (MW). Each molecular weight range is subsequently analyzed by LC-MS/MS (Q-Exactive q-Orbitrap) to identify and quantify the proteins present. Proteins that are differentially digested between a small-molecule-protected sample and the control are determined using Protomap (Protein Topography and Migration Analysis Platform) based on differences in the total amount of each protein and/or differences in sizes of the protein fragments. To date we have utilized our DARTS-MS strategy both to re-discover known protein targets of existing small-molecules as well as discovering novel targets for established drugs and important metabolites.

This dissertation of Reid Lee O'Brien Johnson is approved.

James Akira Wohlschlegel

Robert Thompson Clubb

Joseph Ambrose Loo, Committee Chair

University of California, Los Angeles

2018

DEDICATION

I would like to dedicate this work to all my friends and family for their love and support over the years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost I would like to thank my advisor, Professor Joseph Loo, for giving me this research opportunity and for his constant support and guidance throughout my graduate career. Thanks to Joe's expertise and connections, I have had the honor to work on numerous interesting projects throughout my studies and thanks to his calm and friendly demeanor I have always felt encouraged to seek advice and engage in open discussions. I will be forever grateful to him for allowing me to work on such interesting research projects with such high-end equipment.

I would also like to express my gratitude to Dr. Rachel Loo for all the knowledge and expertise she has shared with me over the years; especially for her detailed answers to the countless questions I have asked on everything from sample preparation, to data analysis, to equipment. And to all the Loo lab members, past and present, I would like to say thank you for the help, support and laughs; in particular, Dr. Sean Shen, Dr. Hong Nguyen, Dr. Carly Daniels (Ferguson), Dr. Pete Wongkongkathep, Dr. Dyna Shirasaki, Michael Nshanian, Neil Quebbemann, and Kate Liu. I would also like to thank my committee members, Professors James Wohlschlegel, Robert Clubb, Jim Bowie, Feng Guo and Joseph Loo for their guidance and advice over the last several years.

The original idea and early development of DARTS is the work of Drs. Jing Huang and Brett Lomenick. The further development and refinement of DARTS-MS is a collaboration between our labs. The work presented in this dissertation was supported in part by the NIH Cellular and Molecular Biology Training Grant.

APPENDIX 1 of this dissertation is a version of Jiang, J., Chan, H., Cash, D.D., Miracco, E.J., Loo Ogorzalek, R.R., Upton, H.E., Cascio, D., O'Brien Johnson, R.L., Collins, K., Loo, J.A. and Zhou, Z.H., Feigon, J. Structure of *Tetrahymena* telomerase reveals previously unknown subunits, functions, and interactions in *Science*, 2015 Vol. 350, Issue 6260, p. aab4070. Reprinted with permissions. I would like to acknowledge my co-authors Jiansen Jiang, Henry Chan, Darian D. Cash, Edward J. Miracco, Rachel R. Ogorzalek Loo, Heather E. Upton, Duilio Cascio, Kathleen Collins, Joseph A. Loo, Z. Hong Zhou, and Juli Feigon.

I would like to acknowledge Jing Huang and Brett Lomenick for their work on the early development of the DARTS system described in Chapters 1 and 2 of this dissertation and for their work identifying anti-aging metabolites described in Chapter 3 of this dissertation.

VITA

2008	Bachelors of Science, Biochemistry University of Oregon, Eugene, Oregon
2004-2008	University of Oregon Dean's Scholarship
2008	University of Oregon Clark Honors College Graduate
2008	University of Oregon Thesis Defended with Distinction
2013	Teaching Assistant, Department of Chemistry and Biochemistry University of California, Los Angeles, California
2013-2016	Ruth L. Kirschstein National Research Service Award
2015	Master of Science, Biochemistry, Molecular & Structural Biology University of California, Los Angeles, California
2016-2018	Teaching Assistant, Department of Chemistry and Biochemistry University of California, Los Angeles, California
2017	American Society for Mass Spectrometry Graduate Student Travel Award

PUBLICATIONS

Jiang, J., H. Chan, D.D. Cash, E.J. Miracco, R.R. Ogorzalek Loo, H.E. Upton, D. Cascio, R. O'Brien Johnson, K. Collins, J.A. Loo, Z.H. Zhou, J. Feigon, "Structure of Tetrahymena telomerase reveals previously unknown subunits, functions, and interactions", *Science*, 350(6260), aab4070 (2015).

Brow, C., R. O'Brien Johnson, R.L. Johnson & H.M. Simon, "Assessment of Anaerobic Toluene Biodegradation Activity by bssA Transcript/Gene Ratios", *Appl Environ Microbiol*., 79(17), 5338-5344 (2013).

Johnson, R.L., J.T. Nurmi, G. O'Brien Johnson, D. Fan, R. O'Brien Johnson, Z. Shi, A.J. Salter-Blanc, P.G. Tratnyek, and G.V. Lowry, "Field-Scale Transport and Transformation of Carboxymethylcellulose-Stabilized Nano Zero-Valent Iron", *Environ. Sci. Technol*., 47(3), 1573-80, (2013).

Johnson, R.L., C.R. Brow, R. O'Brien Johnson, H.M. Simon, "Cryogenic Core Collection and Preservation of Subsurface Samples for Biomolecular Analyses", *Groundwater Monitoring and Remediation*, 33(2), 38-43, (2012).

Brow, C., R. O'Brien Johnson, M. Xu, R.L. Johnson, and H.M. Simon, "Effects of cryogenic preservation and storage on the molecular characteristics of microorganisms in sediments", *Environ. Sci. Technol*., 44(21), 8243-8247, (2010).

Johnson, R.L., P.G. Tratnyek, R. O'Brien Johnson, "Persulfate Persistence under Thermal Activation Conditions", *Environ. Sci. Technol.*, 42, 9350-9356, (2008).

Johnson, R.L., R.B. Thoms, R. O'Brien Johnson, J.T. Nurmi, P.G. Tratnyek, "Mineral Precipitation Upgradient from a Zero-Valent Iron Permeable Reactive Barrier", *Ground Water Monitoring and Remediation*, *2*8(3), 56-64, (2008).

Johnson, R.L., R.B. Thoms, R. O'Brien Johnson, T. Krug, "Field Evidence for Flow Reduction through a Zero Valent Iron Permeable Reactive Barrier", *Groundwater Monitoring and Remediation*, *2*8(3), 47-55, (2008).

SELECTED PRESENTATIONS

O'Brien Johnson, R., M. Nshanian, R. Ogorzalek Loo, J. Loo, "Improving TFA-Based LC-MS of Proteins by Addition of Supercharging Agents", 66th ASMS Conference on Mass Spectrometry and Allied Topics, San Diego, CA (2018)

O'Brien Johnson, R., B. Lomenick, J. Huang, J. Loo, "LC-MS Platform for Identifying Protein Targets of Small-Molecule Binding Relevant to Disease and Metabolism", 65th ASMS Conference on Mass Spectrometry and Allied Topics, Indianapolis, IN (2017)

CHAPTER 1 : CURRENT LIMITATIONS IN IDENTIFYING PROTEIN TARGETS OF SMALL MOLECULES

1.1 INTRODUCTION

The discovery of agents that treat or prevent human disease is of paramount importance to both the scientific community and the general public. However, understanding how these agents interact within living systems is a formidable task. It is perhaps unsurprising then that, of the approximately 20-30 new drugs that are approved every year by the US FDA, only 3-5 address a new molecular target or novel mechanism [1]. Moreover, even if we consider all drugs for which there is a target that is known, or thought to be known, it is estimated that there are only 300-400 drug targets at present [1, 2]. This is in spite of the fact that, since the completion of the Human Genome Project, we know that humans are encoded by approximately 25,000 genes and their products and, undoubtedly, many of these gene or gene products represent potential new drug targets. While the development of a drug to target every gene or gene product is clearly unnecessary, it is equally clear that we are severely limited in developing new cures due to current drug development paradigms. Given the limited understanding of our own biology we desperately need new tools to develop drugs and understand their interactions within living systems.

Small-molecule drugs constitute the majority of today's medicines, however, identifying the underlying molecular targets of the drug's therapeutic effects remains a serious challenge. And it is an even greater challenge to identify off-target interactions

that could result in adverse side effects. Since most drug molecules produce a biological response by binding to a protein or protein complex, the development of new methods for the identification of proteins to which small-molecules bind is an important area of research and a wide variety of methods have been investigated [3-5]. In spite of significant technological advances, however, identifying the protein targets of small molecules remains a serious challenge due to the diversity of compound structures, functions and mechanisms of action.

Current target identification methods can broadly be grouped into affinity-based and non-affinity based methods. While useful, these methods also have inherent limitations. Affinity-based target identification methods require the modification of the drug or drug target without modifying their bioactivity while non-affinity based methods necessitate a specific biochemical or cellular readout. Due to limitations in the current technology, we are developing a universally applicable target identification platform that can directly analyze the binding of a small-molecule to its protein target(s). One idea is that if a protein becomes less susceptible to proteolysis when bound to drug, that alone could allow for the identification of protein drug targets [6, 7]. Thus the method, called DARTS (Drug Affinity Responsive Target Stability), is not dependent on synthetic chemistry nor on a biological response [8]. It doesn't necessitate the immobilization or modification of the small-molecule; it assumes only that the small-molecule will bind to its target and that this will induce some conformational stability in the complex. As a result, DARTS can theoretically be used with native proteins under physiologically relevant conditions and can potentially be used for the target identification of any small molecule, from drugs to pesticides to natural metabolites to peptides and beyond.

While it will be necessary to develop technological tools and practical knowledge to utilize protease susceptibility as a metric for small-molecule target identification, the potential upside is huge. For example, the increasing popularity of high-throughput screening for the identification of biologically active small molecules in the pharmaceutical industry has resulted in hundreds of thousands of compounds with potential therapeutic value [3, 9]. However, the targets of these small molecules are generally unknown and further development of these compounds into useful drugs is often marred by a lack of biological understanding. Thus, a widely applicable target identification platform like DARTS is desperately needed. Once a universal drug-target identification strategy is in hand and has been demonstrated to function well on a variety of small-molecules with known targets, it can be utilized on promising new bioactive small molecules with unknown targets including the hundreds of thousands identified by high-throughput screening. This could significantly improve drug development pipelines by introducing new protein targets for therapeutic research far beyond the 300-400 currently known.

This work aims to test our hypothesis that the phenomenon of a protein becoming more resistant to proteolysis when it is bound to a drug than in its drug-free state can be exploited for proteome-wide unbiased drug target identification. If our hypothesis holds true, entirely new classes of protein drug targets could be revealed at an unprecedented rate. The ultimate goal of medicine is the development of safe and effective treatments and it is our belief that the development of useful therapeutic treatments is greatly enhanced when a protein target is known. The purpose of this work is thus to facilitate the discovery of new protein targets and map interactions of existing small molecules in order to aid drug development and biological research.

1.2 CURRENT APPROACHES

Existing target identification methods can broadly be grouped into two categories, affinity-based and non-affinity based methods. While both categories have proved useful in numerous situations, they also have some fundamental limitations. Affinity chromatography (also known as affinity purification) is the classic example of an affinity-based method and has traditionally played a major role in identifying the binding targets of biologically active small molecules [5]. With this approach, the small molecule or its target is modified in some way, such as covalently linking it to a solid matrix or labeling the molecule (e.g. with an affinity, fluorescent or radioisotope tag). The modified molecule is then used to pull down its target. However, in order for this method to be successful, the molecule must be able to form a derivative and that derivative must retain its biological activity and binding specificity. It is worth noting that the former is often not the case and the latter cannot be predicted ahead of time. Indeed, most compounds have not been successfully derivatized for the purposes of affinity chromatography. It requires time and expertise to design a derivative and it is often difficult or expensive to verify that the bioactivity of a derivative is unchanged. Even if it is successful, this process must be repeated for every molecule of interest.

There are additional challenges with affinity chromatography such as difficulty synthesizing the molecules (in particular natural products) and non-specific interactions that confound the identification of the real protein target. The latter is particularly problematic as most small molecules or matrices will pull down a large number of non-specific binders. This typically necessitates the use of a structurally similar but inactive

analogue as a negative control. There have been improvements in this area through the use of heavy labeled cell cultures [10], however, these methods are still limited by the required derivatization of each individual compound to be analyzed.

Non-affinity target identification methods generally rely on a biochemical or cellular readout to infer a drug target or pathway interaction. Phenotypic screens are an example of a non-affinity based method used widely in chemical genetics studies. For example, one might measure changes in cell growth due to a genetic mutation and screen for small molecules that overcome the effect of the mutation. While classical genetic and modern genomic methods generally have the advantage of using unmodified small molecules and proteins under native conditions, they are limited to cases where a specific mechanism leads to measurable cellular effects [11-13]. They also only provide information on potential targets, so the direct interaction between target and therapeutic molecule must later be verified by other means. These limitations significantly reduce the number of therapeutics and targets that can be identified. For example, while many naturally occurring products are likely to have therapeutic effects, they are unlikely to directly affect fitness through changes in cell growth or viability and thus their efficacy cannot be determined in this way. Furthermore, for many complex systems, such as aging or cancer prevention, there are no easily rescuable phenotypes to be detected.

Genome-wide expression profiling is another example of a non-affinity target identification method. A common example of this technique uses DNA microarrays to identify specific gene expression signatures that indicate the inhibition of a particular pathway, protein or complex [14-16]. However, this technique is only applicable to drugs that induce major changes in the transcriptome. Additionally, the measured response is

often far downstream of the actual drug target. Furthermore, these large scale genetic and transcriptomic approaches are often limited to relatively simple and well-characterized model organisms that can limit their ability to identify targets in mammalian systems.

In short, while the established target identification methods are useful in many cases, they also have inherent limitation. Affinity-based target identification methods are limited to molecules with derivatives that retain their biological activity and binding specificity. The process of creating these derivatives is often difficult and time consuming, if even possible, and must be repeated for every molecule of interest. Non-affinity target identification methods are indirect and require measureable biological responses. Therefore, the development of an unbiased proteome-wide target identification platform free of these limitations is highly desirable.

Recently, there have been a number of attempts to develop such a platform for target identification [17-21]. Of these methods, thermal proteome profiling (TPP) is by far the most widely used. The thermal profiling method starts off with a collection of proteins split into two fractions. One fraction receives the small molecule while the other fraction receives just the vehicle and both are incubated for a period of time (typically ten minutes to an hour). For samples with membrane proteins, a detergent is also added. Both fractions are split into around ten aliquots, each aliquot is heated at a different temperature for a few minutes (typically for three minutes between 40°C and 80°C). The aliquots are then spun down and the soluble fraction is digested with trypsin. Digested samples are tagged with tandem mass tags (TMT) such that both control and small molecule samples have the same TMT tag for a given temperature but all temperatures

have a different tag. All aliquots from the same initial sample are then pooled back together (i.e. all small molecule aliquots are pooled and all control aliquots are pooled). There is often an additional fractionation step prior to mass spectrometry (MS) analysis to increase analytical depth in complex samples. While some methods vary slightly on the tagging or fractionation steps, e.g. some methods use gels while others do not, conceptually they are all quite similar. For each protein detected, the abundance in both the small molecule and control samples are plotted at each temperature point. As the protein is degraded, e.g. unfolded to cause precipitation, at higher temperatures, less of it will remain in the soluble fraction creating a 'melting curve.' Any significant difference between the melting curves of the small molecule and control sample could indicate an interaction with the small molecule.

The thermal profiling method has been used successfully in a variety of systems including whole cell and lysate analysis of metabolites and membrane proteins [22, 23]. There are a number of limitations, however. The most basic limitation is the reliance on temperature to generate melting curves of proteins. In the laboratory setting, it is impractical to heat samples above 100°C, yet there are many proteins that are stable even at these temperatures. Indeed, for some extreme organisms, the optimal growth conditions are in excess of 110°C [24]. Clearly in these cases, thermal profiling would not work. Furthermore, the choice of temperature steps can dramatically change results which may necessitate repeating the experiment at multiple temperature ranges [22]. TPP also has considerable difficulty with low abundance proteins, which often means there is not enough sensitivity to determine if there is a difference between the small molecule and control sample. Additionally, much of the data interpretation is subjective. For

example, some papers consider only melting curves of ‘good-quality’ which is a term that is never clearly defined [23]. There are also sometimes known targets of drugs that are not detected for unknown reasons [25]. Also, since TPP compares melting curves, it depends on very accurate quantitation more than other methods such as DARTS. That is to say, accuracy for DARTS depends mostly on comparing the presence or absence of a protein in a particular fraction. One does not need to know the exact amount of protein present in each fraction for DARTS, but one does to create a good melting curve. Thus while recent proteome-wide target identification methods such as TPP are promising, there is still the need for new methods and further development.

1.3 DARTS DEVELOPMENT

In order to overcome the limitations of current target identification approaches, we are developing a universally-applicable, proteome-wide strategy to detect the direct interaction of a small-molecule and its target. Our hypothesis is if a protein becomes more resistant to proteolysis when it is bound to a drug than in its drug-free state [6, 7], that this phenomenon can be exploited for unbiased drug target identification and drug screening. Drug affinity responsive target stability (DARTS) [8] is conceptually simple and is summarized in Figure 1-1. A collection of proteins (e.g. a whole cell lysate) is split into two fractions, one of which receives a drug molecule while the other receives only the vehicle for the drug and serves as a control. The mixtures are allowed to equilibrate, during which time the small molecule binds to its target inducing conformational stability in the complex. Both fractions are then subjected to a limited proteolysis and the relatively intact drug-target complex is isolated and identified. One of the major advantages to such a system is that it does not require the modification or immobilization of the small molecule or target. The interactions can occur under physiologically relevant conditions with native proteins. Another major advantage is that the system depends only on the direct interaction of the drug and protein target. There is no need to measure biological changes in the system nor is any prior knowledge of the system, target, or mechanism required.

One of the first proof-of-concept experiments for DARTS involved the immunophilin protein, FKBP12 (Figure 1-2). FKBP12 is a chaperone protein involved in protein folding and trafficking and is a common target of immune suppressing drugs such as rapamycin and FK506 [11, 26]. The protease activity of subtilisin was markedly

decreased in the presence of either rapamycin or FK506 when compared to the drug vehicle on its own (DMSO) or in the presence of an unrelated drug that does not bind FKBP12 (wortmannin). This demonstrates that the protection from proteolysis is specific to drugs that bind FKBP12 and will occur at different levels of protease concentration. Additionally, rapamycin and FK506 have dramatically different binding affinities (K_i = 0.2 nM and 400 nM respectively) and yet, at least qualitatively, have similar levels of protection from proteolysis. Lastly, based on the X-ray crystal structures of FKBP12 both bound and unbound to rapamycin and FK506, it does not appear that there are any major conformational changes in the structure of the protein, which suggests that the binding of the drug alone is sufficient to protect the protein from proteolysis [27].

While this first proof-of-concept experiment demonstrate the utility of DARTS, it examined only a single protein and drug interaction. In order for DARTS to be a generally useful drug target identification tool, it must also be effective in more complex systems, such as whole cell lysates. Thus for the next experiment, DARTS was used to examine a whole cell lysate (human Jurkat T cells) for interaction with didemnin B (DB). DB is a natural product isolated from tunicate with strong antiviral and anti-cancer bioactivity and is known to bind EF-1 alpha (EF1A) [28]. DB on its own was found to have no impact on the whole cell lysate, however in the presence of both DB and a protease (thermolysin) a protected band of ~50 kDa was observed (Figure 1-3). The protected band and matching region in the control sample were analyzed by mass spectrometry and the DB sample was found to be strongly enriched for EF1A isoforms. Similar results were obtained from an immunoblot; increasing concentrations of DB had no impact on the abundance of intact EF1A in control samples. However, in the presence of thermolysin, increasing

concentrations of DB led to increasing protection of intact EF1A. The total protein abundance did not change substantially between samples as indicated by a loading control (GAPDH) that is highly resistant to thermolysin. While this experiment is a good demonstration of DARTS in a more complex system, the possibility remains that other proteins are also protected from proteolysis by DB but are not evident on the gel due to low concentration or resolution.

1.4 ALTERNATIVES AND IMPROVEMENTS

The previous examples of DARTS demonstrate its utility in identifying protein targets in complex mixtures and under various small molecule and protease concentrations. It also suggests that this target identification process is possible even in the absence of conformational changes or with small-molecules with low binding affinity. In addition, all reactions take place under native conditions without modification of the small molecule or target and it is independent of any biological response from the cell, relying solely on the direct interaction of the small molecule and its protein target.

However there are limitations; in the preceding examples, only highly abundant proteins were identified (such as elongation and transcription factors). This is due primarily to the low resolving power and qualitative results of SDS-PAGE. Additionally, the putative targets were identified solely by the presence of an additional protein band. Clearly, it would be desirable to detect lower abundance proteins (such as regulatory proteins) and to have a more objective measure of protease protection. Thus the goal of this project going forward is to develop a more sensitive system with ability to efficiently analyze the results. One alternative approach that avoids the limitations of gel electrophoresis and potentially provides higher sensitivity and throughput, would be to couple DARTS with a shotgun LC-MS/MS strategy (as will be discussed in CHAPTER 2). In addition, it would be beneficial to expand the platform and make it more generally useful by applying it to a wide variety of small molecules and systems. For example, while the previous proof-of-concept experiments used highly specific drugs, DARTS should also be applicable to naturally occurring small molecules (such as metabolites). This could prove

highly informative and aid in the difficult task of mapping the complex interactions of natural products within the cell.

In developing this improved DARTS platform, it is important to address potential limitations of the system. One fundamental limitation of DARTS is that the drug target must be susceptible to proteolysis and there is evidence that select groups of proteins have evolved considerable proteolytic resistance due to their importance under stressed conditions [29]. However, while a protein may be resistant to one or two proteases this is likely less of an issue when many proteases are used together. Thus a commercially available mixture of proteases, pronase (Roche), will be utilized in future experiments. Pronase contains both endo- and exo-proteases and is able to digest both folded and unfolded proteins. Since pronase is able to break down virtually all proteins into individual amino acids when at high enough concentrations, highly protease resistant proteins should not be an issue. This should allow for a more even level of proteolysis across the proteome. It is still conceivable, however, that some proteins will be completely degraded before others have undergone any proteolysis. This could necessitate using multiple concentrations of protease in each analysis.

Another possible concern is that small-molecule binding will alter non-target proteins' susceptibility to proteolysis if they are part of a complex that contains the target. For example, it is conceivable that if drug binding causes a protein complex to disassociate it could render the individual components of that complex more susceptible to proteolysis. This is not necessarily a disadvantage, however, as it could provide details on protein complex formation and disassociation during drug binding. It may complicate

data interpretation, but as long as the difference between experimental and control samples are consistent, the results could provide detail on protein interaction pathways.

There is also considerable evidence that the *in vivo* stability of proteins is highly variable upon drug binding. Indeed, both increases and decreases in susceptibility to proteolysis have been observed upon drug binding for target proteins in cells [30, 31]. Once again, this is not necessarily a disadvantage as a small molecule that destabilizes a protein would just as readily be identified by DARTS as a small molecule that stabilizes a protein. It is conceivable that this would even provide some insight into the molecular mechanism of the drug molecule.

1.5 CONCLUSION

In this chapter, the limitations of current drug target identification methods, the promising early DARTS development, and potential alternatives and improvements to our system have been discussed. The development of useful therapeutics is expensive, time consuming, technically challenging, and often hampered by an incomplete understand of the biology involved. Current methods for drug target identification are limited by the necessity to modify each new molecule individually or the reliance on measurable biologic responses. The initial development of the DARTS protocol suggests that it may overcome these challenges and thus warrants further development.

With future improvements to the DARTS method, specifically by coupling it with mass spectrometry, we hope to expand scientific understanding in at least two major areas. First, we aim to radically change the way we study metabolite interactions within the cell. Since DARTS does not require the modification of small-molecules, it may be possible to systematically map the complex network of protein-metabolite interactions within the cell for the first time. In so doing, we can expand our understanding of the regulatory roles that metabolites and natural products have within the cell and apply this knowledge to better understand the root causes of disease. As with other global mapping projects like genome sequencing or proteome characterization, a cell-wide protein-metabolite interaction network will provide a wealth of information with immediate benefits to biology and medicine.

The other major application of our approach is, of course, drug discovery. With the rapid expansion and enormous output of molecular library screening centers, we now

know of hundreds of thousands of bioactive small molecules that are potentially useful as therapeutics. However, the vast majority of these compounds are not well characterized and developing new assays to elucidate their protein target or pathway is challenging and time consuming. While there exist methods for activity-based protein profiling that can be used to probe enzyme classes [32], these methods are limited to fairly small portions of the proteome. Thus, there is currently an unmet need for a universally-applicable proteome-wide drug target identification system. Our strategy can be used in combination with these vast small-molecule libraries to screen potentially useful compounds for interactions within the cell. Since DARTS functions independent of biology and without any prior knowledge of protein targets, it can be directly applied to the screening of novel molecules without having to design, develop or implement a new assay. Therefore, this technique could allow drug screening to begin as soon as a molecule is implicated as a potential therapeutic.

We also hope to greatly expand the limited number of useful drug targets from the 300-400 available at present [1]. By investigating the protein targets of both naturally occurring molecules and novel therapeutics, it should be possible to identify new pathways that are well-suited to drug development. To achieve these goals, however, further development of our DARTS system by combining it with mass spectrometry will be necessary. Specifically we hope to increase the sensitivity of the platform (and thereby increase our ability to detect both proteolysis and lower abundance proteins) as well as improve quantification and data analysis. Coupling DARTS to mass spectrometry could allow for the unbiased analysis of therapeutic protein targets across multiple proteomes in ways no other tool has been able to achieve.

1.6 FIGURES

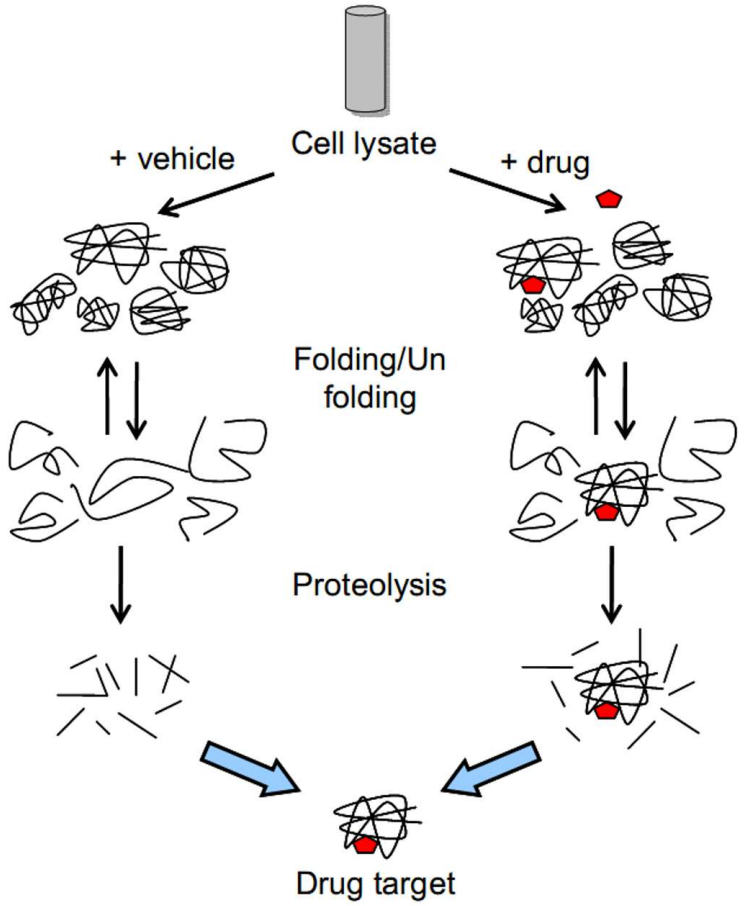


Figure 1-1: Scheme for DARTS A collection of protein (e.g. a cell lysate) is split into two fractions one of which is treated with a small molecule (e.g. drug) and the other receives an equal amount of solvent and serves as a control. The samples are allowed to incubate and then are treated with a protease. The control and treated samples are compared and proteins protected from proteolysis in the treated sample are isolated and identified [8].

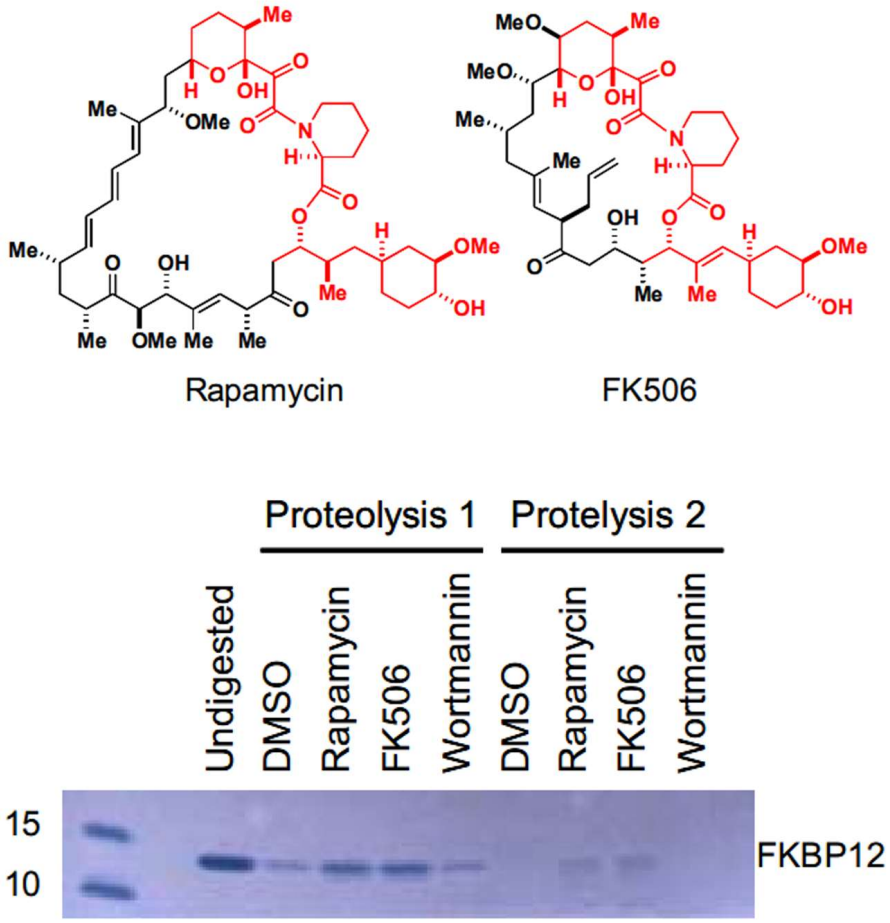


Figure 1-2: DARTS proof-of-principle using FKBP12. (Top) Structures of rapamycin and FK506. (Bottom) Human FKBP12 protein was incubated with the indicated compounds and digested with subtilisin. Similar results were obtained under two arbitrary proteolysis conditions tested [8].

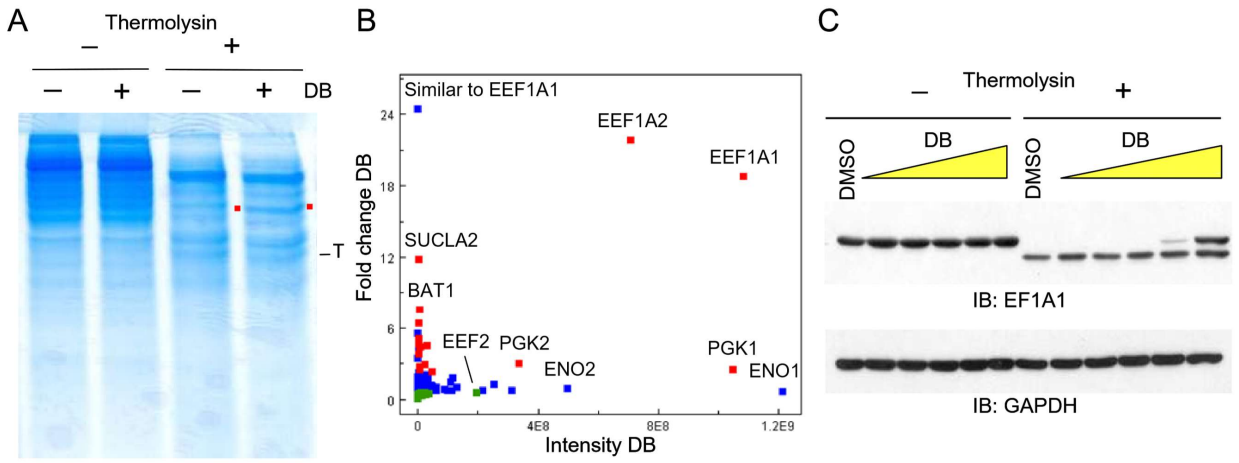


Figure 1-3: DARTS using whole-cell lysate. (A) Intact human Jurkat cells were treated with didemnin B (DB) (1µg/mL) or DMSO and lysates were subjected to thermolysin digestion and Coomassie staining. (B) Enrichment of EF1A isoforms in the protected band (red dots) from (A) revealed by mass spectrometry. Red, protein enriched > 2-fold with p-value < 0.001; green, protein depleted > 2-fold with p-value < 0.001; blue, unchanged protein. (C) Intact human Jurkat cells were treated with DMSO or DB (0.1, 1, 10, 100, or 1000 ng/mL), and lysates were subjected to thermolysin digestion and western blotting. GAPDH is resistant to thermolysin under these conditions and served as a loading indicator [8].

1.7 REFERENCES

1. Overington, J.P., Al-Lazikani, B. and Hopkins, A.L., 2006. How many drug targets are there?. Nature Reviews Drug Discovery, 5(12), p.993.

2 Pawson, A.J., Sharman, J.L., Benson, H.E., Faccenda, E., Alexander, S.P., Buneman, O.P., Davenport, A.P., McGrath, J.C., Peters, J.A., Southan, C. and Spedding, M., 2013. The IUPHAR/BPS Guide to PHARMACOLOGY: an expert-driven knowledgebase of drug targets and their ligands. Nucleic acids research, 42(D1), pp.D1098-D1106.

3. Burdine, L. and Kodadek, T., 2004. Target identification in chemical genetics: the (often) missing link. Chemistry & Biology, 11(5), pp.593-597.

4. Campillos, M., Kuhn, M., Gavin, A.C., Jensen, L.J. and Bork, P., 2008. Drug target identification using side-effect similarity. Science, 321(5886), pp.263-266.

5. Terstappen, G.C., Schlüpen, C., Raggiaschi, R. and Gaviraghi, G., 2007. Target deconvolution strategies in drug discovery. Nature Reviews Drug Discovery, 6(11), p.891.

6. Park, C. and Marqusee, S., 2005. Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. Nature Methods, 2(3), p.207.

7. Tucker, C.L. and Fields, S., 2001. A yeast sensor of ligand binding. Nature biotechnology, 19(11), p.1042.

8. Lomenick, B., Hao, R., Jonai, N., Chin, R.M., Aghajan, M., Warburton, S., Wang, J., Wu, R.P., Gomez, F., Loo, J.A. and Wohlschlegel, J.A., 2009. Target

identification using drug affinity responsive target stability (DARTS). Proceedings of the National Academy of Sciences, pp.pnas-0910040106.

9. Inglese, J., Johnson, R.L., Simeonov, A., Xia, M., Zheng, W., Austin, C.P. and Auld, D.S., 2007. High-throughput screening assays for the identification of chemical probes. Nature chemical biology, 3(8), p.466.

10. Ong, S.E., Schenone, M., Margolin, A.A., Li, X., Do, K., Doud, M.K., Mani, D.R., Kuai, L., Wang, X., Wood, J.L. and Tolliday, N.J., 2009. Identifying the proteins to which small-molecule probes and drugs bind in cells. Proceedings of the National Academy of Sciences, 106(12), pp.4617-4622.

11. Heitman, J., Movva, N.R. and Hall, M.N., 1991. Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. Science, 253(5022), pp.905-909.

12. Rine, J., Hansen, W., Hardeman, E. and Davis, R.W., 1983. Targeted selection of recombinant clones through gene dosage effects. Proceedings of the National Academy of Sciences, 80(22), pp.6750-6754.

13. Giaever, G., Shoemaker, D.D., Jones, T.W., Liang, H., Winzeler, E.A., Astromoff, A. and Davis, R.W., 1999. Genomic profiling of drug sensitivities via induced haploinsufficiency. Nature Genetics, 21(3), p.278.

14. Bradner, J.E., McPherson, O.M., Mazitschek, R., Barnes-Seeman, D., Shen, J.P., Dhaliwal, J., Stevenson, K.E., Duffner, J.L., Park, S.B., Neuberg, D.S. and Nghiem, P., 2006. A robust small-molecule microarray platform for screening cell lysates. Chemistry & Biology, 13(5), pp.493-504.

15. Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N. and Reich, M., 2006. The Connectivity

Map: using gene-expression signatures to connect small molecules, genes, and disease. science, 313(5795), pp.1929-1935.

16. DeRisi, J.L., Iyer, V.R. and Brown, P.O., 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science, 278(5338), pp.680-686.

17. Molina, D.M., Jafari, R., Ignatushchenko, M., Seki, T., Larsson, E.A., Dan, C., Sreekumar, L., Cao, Y. and Nordlund, P., 2013. Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. Science, 341(6141), pp.84-87.

18. Backus, K.M., Correia, B.E., Lum, K.M., Forli, S., Horning, B.D., González-Páez, G.E., Chatterjee, S., Lanning, B.R., Teijaro, J.R., Olson, A.J. and Wolan, D.W., 2016. Proteome-wide covalent ligand discovery in native biological systems. Nature, 534(7608), p.570.

19. Niphakis, M.J., Lum, K.M., Cognetta III, A.B., Correia, B.E., Ichu, T.A., Olucha, J., Brown, S.J., Kundu, S., Piscitelli, F., Rosen, H. and Cravatt, B.F., 2015. A global map of lipid-binding proteins and their ligandability in cells. Cell, 161(7), pp.1668-1680.

20. Parker, C.G., Galmozzi, A., Wang, Y., Correia, B.E., Sasaki, K., Joslyn, C.M., Kim, A.S., Cavallaro, C.L., Lawrence, R.M., Johnson, S.R. and Narvaiza, I., 2017. Ligand and target discovery by fragment-based screening in human cells. Cell, 168(3), pp.527-541.

21. Li, X., Gianoulis, T.A., Yip, K.Y., Gerstein, M. and Snyder, M., 2010. Extensive in vivo metabolite-protein interactions revealed by large-scale systematic analyses. Cell, 143(4), pp.639-650.

22. Huber, K.V., Olek, K.M., Müller, A.C., Tan, C.S.H., Bennett, K.L., Colinge, J. and Superti-Furga, G., 2015. Proteome-wide drug and metabolite interaction mapping by thermal-stability profiling. Nature Methods, 12(11), p.1055.

23. Reinhard, F.B., Eberhard, D., Werner, T., Franken, H., Childs, D., Doce, C., Savitski, M.F., Huber, W., Bantscheff, M., Savitski, M.M. and Drewes, G., 2015. Thermal proteome profiling monitors ligand interactions with cellular membrane proteins. Nature Methods, 12(12), p.1129.

24. Blöchl, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H.W. and Stetter, K.O., 1997. Pyrolobus fumarii, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 C. Extremophiles, 1(1), pp.14-21.

25. Savitski, M.M., Reinhard, F.B., Franken, H., Werner, T., Savitski, M.F., Eberhard, D., Molina, D.M., Jafari, R., Dovega, R.B., Klaeger, S. and Kuster, B., 2014. Tracking cancer drugs in living cells by thermal profiling of the proteome. Science, 346 (6205)

26. Schreiber, S.L. and Crabtree, G.R., 1995. Immunophilins, ligands, and the control of signal transduction. Harvey Lectures, 91, pp.99-114.

27. Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J., 1993. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. Journal of Molecular Biology, 229(1), pp.105-124.

28. Crews, C.M., Collins, J.L., Lane, W.S., Snapper, M.L. and Schreiber, S.L., 1994. GTP-dependent binding of the antiproliferative agent didemnin to elongation factor 1 alpha. Journal of Biological Chemistry, 269(22), pp.15411-15414.

29. Park, C., Zhou, S., Gilmore, J. and Marqusee, S., 2007. Energetics-based protein profiling on a proteomic scale: identification of proteins resistant to proteolysis. Journal of Molecular Biology, 368(5), pp.1426-1437.

30. Cohen, A.A., Geva-Zatorsky, N., Eden, E., Frenkel-Morgenstern, M., Issaeva, I., Sigal, A., Milo, R., Cohen-Saidon, C., Liron, Y., Kam, Z. and Cohen, L., 2008. Dynamic proteomics of individual cancer cells in response to a drug. Science, 322(5907), pp.1511-1516.

31. Nishiya, Y., Shibata, K., Saito, S., Yano, K., Oneyama, C., Nakano, H. and Sharma, S.V., 2009. Drug–target identification from total cellular lysate by drug-induced conformational changes. Analytical Biochemistry, 385(2), pp.314-320.

32. Cravatt, B.F., Wright, A.T. and Kozarich, J.W., 2008. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. Annu. Rev. Biochem., 77, pp.383-414.

CHAPTER 2 : A LC-MS PLATFORM FOR IDENTIFYING PROTEIN TARGETS OF SMALL-MOLECULE BINDING RELEVANT TO DISEASE AND METABOLISM

2.1 INTRODUCTION

As described in CHAPTER 1, the original DARTS protocol used one-dimensional SDS-PAGE to detect changes in protease susceptibility [1]. However, for complex mixtures of proteins, such as whole cell lysates, it can be difficult or impossible to discern changes in protein abundance from a gel. This is especially difficult with lower abundance proteins or for proteins that are only partially protected from proteolysis. To improve the method we pursued a strategy that coupled gel-free protein fractionation with LC-MS/MS (liquid chromatography-tandem mass spectrometry). We expect that this method can provide higher sensitivity and throughput while reducing sample variability and subjective data analysis. In the protocol detailed in this chapter, each protein sample is incubated either in the presence or absence of a small molecule before being partially digested with pronase and fractionated based on molecular weight (MW). Each molecular weight fraction is then digested to completion with trypsin and analyzed by LC-MS/MS to identify and quantify the proteins/peptides present (Figure 2-1). Identified proteins that underwent different levels of pronase digestion between the control and experimental sample are determined based on the total amount of protein in a given molecular weight fraction and/or the difference in size of a particular protein's fragments.

Our new DARTS-MS platform replaces SDS-PAGE with in-solution molecular weight fractionation and shotgun LC-MS/MS. The protein fractionation is accomplished

using a Gel Elution Liquid Fraction Entrapment Electrophoresis system (GelFree 8100, Protein Discovery/Expedeon) [2]. Each molecular weight fraction is completely digested with trypsin using our modified eFASP (enhanced Filter-Aided Sample Preparation) protocol [3]. The digested peptides are purified and desalted using C18 StageTips before being measured by LC-MS/MS using a Thermo EASY-nLC 1000 attached to a Thermo Q-Exactive Orbitrap instrument (top-10 data dependent acquisition; Acclaim PepMap RSLC Nano C18 column, 2-hr gradient). Proteins with different levels of digestion between the small-molecule protected sample and the control are determined and visualized using Protomap (Protein Topography and Migration Analysis Platform) [4]. The raw MS data can be searched with any proteomic search algorithm (e.g. SEQUEST, Mascot, or ProLuCID) and computer scripts are used to convert the searched MS data into a form Protomap can read. Statistics built in Protomap uses two-tailed homoscedastic t-test to determine p-values (Figure 2-2).

With this approach we readily rediscovered that translation elongation factor 1 alpha (EF1A), a known target of depsipeptide didemnin B (DB), is specifically protected by DB (Figure 2-3 A). In addition, we found that five other proteins of lesser abundance are also protected by DB (Figure 2-3 B) that were missed in previous affinity chromatography and SDS-PAGE DARTS studies [1, 5]. As described in the rest of this chapter, we are continuing to refine the experimental platform outlined in Figure 2-1. Continuing development will improve throughput, sensitivity, and robustness, while reducing experimental variability and false positives/negatives.

2.2 DARTS-MS PROTOCOL

The DARTS-MS protocol is applicable to a wide range of small molecules and systems and, while the procedure is fairly standardized, there are some initial decisions that need to be made. One of the first should be deciding the source of the protein. Since DARTS-MS depends only on the interaction of the drug and drug target, any collection of proteins that contains the drug target should suffice. We have successfully used DARTS-MS on lysates from a number of mammalian cell lines including Jurkat, HeLa, and HEK293. Another consideration is the concentration of the small molecule. While, in general, we try to use physiologically relevant concentrations of the small molecule, DARTS-MS is applicable to any small molecule that binds to a protein target and thus, in many cases, little or no dosage information may be available. In these cases, we recommend using fairly high concentrations of the small molecule (as much as an order of magnitude above the estimated target binding affinity) to insure maximum levels of protection of the drug target.

The type of protease used in the DARTS-MS reaction is another important consideration. The early DARTS development described in CHAPTER 1 used primarily subtilisin or thermolysin. However, we have found that any single protease is not well suited to a whole-cell lysate DARTS-MS experiment. For example, although thermolysin is a fairly non-specific protease, it is only efficient at digesting unfolded proteins [6]. Furthermore, when considering the entire proteome, there are frequently a number of proteins that are resistant to any one protease (e.g. GAPDH is highly resistant to thermolysin). Thus, under most circumstances when looking at a whole-cell lysate or

investigating a novel small molecule, we recommend the use of pronase. Pronase is a commercially available mixture of many different proteases that is effective on a wide variety of substrates and on both folded and unfolded proteins.

One additional consideration for DARTS-MS is the concentration of protease. When using pronase, we have found that nearly all proteins are partially to completely digested in 30 minutes when using a 1:100 to 1:10,000 pronase to protein ratio (weight to weight). However, while some optimization may be required for a particular protein of interest, when performing DARTS-MS it is generally recommended to use a high concentration of pronase (i.e. between 1:100 and 1:1000). This will insure that all proteins are partially digested and easily separated during the gel-free fractionation step. Furthermore, while early DARTS experiments often relied on Western blotting to identify a protein target and thus may require a relatively intact protein for antibody recognition, this is not a concern with mass spectrometry as fully digested proteins can still be readily identified.

Finally, the fractionation step must be considered with DARTS-MS. While it is possible to directly analyze DARTS samples by LC-MS/MS, it is usually advantageous to fractionate first. The reason for this is described in Figure 2-4. Essentially, if DARTS samples are directly analyzed by LC-MS/MS, protein targets will only be identified if a large proportion of the protein is degraded into very small pieces and filtered away prior to LC-MS/MS analysis. In this situation, there will be a substantial difference in protein concentration between the compound-containing sample and the control sample (Figure 2-4 scenario A). However, if there is only a limited amount of proteolysis, a large portion of the protein will remain intact, and thus will not be filtered out prior to LC-MS/MS

analysis. This will result in a similar abundance of the target protein in both the compound-containing sample and the control sample (Figure 2-4 scenario B). While there is still considerable protection from proteolysis in this case, as indicated by the reduction in full length protein in the control sample, this difference is only detectable if the samples are first fractionated by molecular weight. Thus, in order to avoid missing potential target proteins, we highly recommend a system that separates samples by molecular weight as well as determining abundance and identity of the protein. In this protocol we utilize an in-solution molecular weight-based fractionation system called GelFree (Protein Discovery, Inc.). The GelFree fractionation system allows the separation of complex proteomic mixtures of proteins and peptides into discrete molecular weight ranges. When combined with the DARTS-MS protocol, this fractionation allows for the identification of protein targets even if the total abundance of protein is unchanged between the compound-containing sample and the control sample. This has the advantage of increasing the sensitivity of the assay as well as reducing the number of false negatives.

After the initial decisions regarding proteins, small molecules, proteases and fractionation have been made, a standardized DARTS-MS protocol can be followed. This will involve collecting and lysing cells, quantifying the protein concentration in the lysate, incubating the lysates with the compound of interest or the negative control, performing a limited proteolysis, fractionating the lysates with the gel-free fractionation system, performing an addition trypsin digestion and finally analyzing the samples by LC-MS/MS (Figure 2-1). Many general protocols for DARTS have been described previously [1, 7, 8]. However, the specific methods used in the subsequent chapters and described here are primarily from Lomenick et al. 2013 [7]. The following protocol is split into two parts;

the first detailing the DARTS reaction and the second describing the fractionation and LC-MS/MS analysis.

Part 1: DARTS Reaction

Materials

M-PER lysis buffer (Pierce cat. no. 78501)

20x Protease inhibitor solution (one Complete Mini Tablet [Roche, cat. no. 11836153001] in 0.5 mL dH_2O).

Phosphatase inhibitor solutions (100 mM β-pyrophosphate, 50 mM Sodium Pyrophosphate, 200 mM Sodium Vanadate, and 1 M Sodium Fluoride)

Cell culture in logarithmic growth phase (up to 70–80% confluent)

Phosphate Buffered Saline (PBS)

Cell scraper (such as Fisher cat. no. 08-100-241)

Refrigerated centrifuge capable of 18,000 rcf (such as Beckman Microfuge 22R)

10x TNC Buffer (500 mM Tris-HCl (pH 8.0), 500 mM NaCl, 100 mM $CaCl_2$)

BCA Protein Concentration Assay (Pierce cat. no. 23225 or similar assay)

100x stock solution of small molecule(s) in vehicle of choice

Corresponding vehicle

Pronase stock solution (Roche cat. no. 10165921001 dissolved in dH_2O to a final concentration of 10 mg/mL)

1x TNC Buffer (dilute 10x TNC Buffer 10-fold in dH_2O)

Collect and Lyse Cells

1. Prepare 1 mL of M-PER lysis buffer by mixing 883 µL M-PER, 10 µL 100 mM β-pyrophosphate, 50 µL 50 mM Sodium Pyrophosphate, 5 µL 200 mM Sodium Vanadate, 2 µL 1 M Sodium Fluoride, and 50 µL 20x Protease inhibitor solution on ice.

   For DARTS experiments with mammalian cells we typically lyse using the commercial M-PER buffer, which has worked well for a variety of cell lines and small molecule-protein pairs. However, it is not necessary to use M-PER, as DARTS can be successfully performed with a variety of gentle, non-denaturing lysis buffers (e.g. 0.2 - 1% Triton X-100, NP-40). However, harsh, denaturing lysis buffers like RIPA (radio-immuno-precipitation assay) buffer should be avoided for DARTS, as most proteins must retain their native complexes and/or structures in order to bind ligands. The exact protease and phosphatase inhibitors used are also not typically critical, and multiple companies sell inhibitor cocktails that are similar in composition and could be used instead of the individual ones listed here. These inhibitors will not significantly affect pronase (nor thermolysin or subtilisin), which are added in excess during the ensuing DARTS reactions. Whether or not adding phosphatase inhibitors is crucial for DARTS has not been tested.

2. Remove the growth medium from a single 10 cm dish of cells and wash the cells once with 1 mL cold PBS.

   If cells that grow in suspension will be used, simply pellet the cells by centrifuging at 1000 rcf for 10 minutes at 4°C. After centrifuging, remove the growth medium, resuspend the cells in M-PER lysis buffer, and then proceed to step 4 of this protocol. For Jurkat cells, 3 x 10^7 cells lysed in 600 µL M-PER will typically yield soluble lysates of approximately 5 µg protein/µL.

3. Apply 600 µL cold M-PER lysis buffer (with added protease and phosphatase inhibitors) onto the cells and collect the cells with a cell scraper (by holding the plate at an angle and gently scraping from top to bottom).
4. Transfer the M-PER/lysing cells into a 1.5 mL tube pre-chilled on ice, mix it well by pipetting up and down several times, and incubate the tube on ice for 10 minutes.

   Upon completion of the incubation the cells will be completely lysed open and soluble proteins will be extracted into the M-PER solution. You should notice a white, cloudy material floating within the solution, which is primarily DNA. This material must first be separated from the soluble proteins before performing DARTS.

Prepare Cell Lysates for DARTS Experiment

5. Centrifuge the tube at 18,000 x g for 10 minutes at 4°C, best in a refrigerated centrifuge (e.g., Beckman Coulter Microfuge 22R).
6. Transfer 600 µL of the supernatant into a new 1.5 mL tube and discard the pellet.

   The pellet contains the DNA and some other insoluble materials from the cell. The supernatant contains the soluble proteins from the cytoplasm, nucleus, mitochondria, and other cellular compartments.

7. Add 66.7 µL 10x TNC Buffer to the protein lysates and mix well.

   DARTS with pronase works with or without adding the TNC buffer (although we have not done a side-by-side comparison). In our earlier DARTS experiments with thermolysin the TNC buffer was added to ensure that there was enough Ca^{2+} for thermolysin to be fully active, given that the exact composition of M-PER is unknown and that most common lysis buffers do not contain Ca^{2+}.

8. Measure the protein concentration of the lysates using the BCA, Bradford or other assay, according to the manufacturer's instructions.

   We routinely use the BCA protein concentration assay, which is not affected by the M-PER or TNC buffers. The exact protein concentration is not critical, but DARTS

may not work as well if the protein is too dilute. We find that an optimal protein concentration for DARTS may be between 4-6 μg/μL, though we have also used 2 μg/μL with good results (lower concentrations have not been tested extensively). The amount of lysis buffer and cells used in this protocol should result in lysates of approximately 5 μg/μL.

Although the exact protein concentration does not appear to be critical, the use of a fixed protein concentration in this protocol is primarily to standardize the protease amount. We initially optimized the amount of proteolysis for specific proteins (wherein 50–80% of the protein of interest is digested) based on the ratio of protease to protein. However, the amount of proteolysis in a sample depends on a multitude of factors, including the concentration of the protein sample, the concentration of the protease, the buffer components, and the time of proteolysis. Therefore, two samples with different protein concentrations but digested with the same protease to protein ratio will not be digested equally because both the protein and the protease concentrations will be different.

9. Split the lysates into two samples by transferring 297 μL into each of two 1.5 mL tubes.

At this point the lysates should be allowed to warm to room temperature. Some small molecules may have low solubility in aqueous solution and will be more prone to precipitate at lower temperatures. Moreover, hydrophobic interactions are

destabilized at low temperatures, and binding kinetics should be faster at the higher temperature. The rest of the experiment could also be performed at physiological temperature, but this is not necessary.

Incubate protein lysates with the small molecule

10. Add into one tube 3 µL vehicle, and into the other tube 3 µL small molecule at 100x the final concentration for a 1:100 dilution.

    The small molecule stock solution does not have to be exactly 100x, but, if using DMSO as a vehicle, we advise that the final DMSO concentration in the lysates be kept low, around 1-2%, as higher concentrations may interfere with binding and proteolysis. We also prefer to not pipet volumes less than 1 µL, as the chance of significant pipetting error becomes greater. Therefore, do not use a small molecule stock solution that is greater than 300x directly (dilute it first), unless the volume of lysates is also larger.

11. Mix the samples immediately by gently flicking the tubes with your finger several times, spin the tubes briefly in a microfuge to bring all the liquid to the bottom of the tubes, and allow them to incubate at room temperature for 1 hour (or longer if determined necessary from pilot studies).

It is critical that you mix the samples thoroughly but gently, to ensure that the small molecule/protein solution becomes homogeneous and binding equilibrium is reached. Furthermore, it is essential that you do not vortex or use any other harsh mixing method on any sample that contains protein! This will cause many proteins to become denatured and/or aggregated.

Preparation for proteolysis

12. Thaw one aliquot of 10 mg/mL pronase quickly and place on ice.
13. Dilute pronase to 1.25 mg/mL by mixing 12.5 µL pronase with 87.5 µL cold 1x TNC Buffer, which will serve as the 1:100 pronase stock solution.

It is essential to keep all protease solutions on ice at all times to prevent them from starting to digest themselves. The 1.25 mg/mL pronase solution will be the highest concentration used in this experiment. The pronase stock solutions prepared in this and the following step are calculated for lysates of 5 µg protein/µL, and will need to be modified if the protein concentration is significantly different.

Pronase dilutions of 1:300 or 1:1000 can also be used if subsequent analysis indicates a 1:100 stock is too concentrated.

Perform proteolysis

14. Prepare aliquots from both protein samples (the compound-treated sample and control sample), each with 50 µL, and save the remaining 50 µL of each sample as a non-digested control sample.
15. Start a timer and immediately add 2 µL 1:100 pronase solution to one aliquot of compound-treated sample, mix well, and incubate at room temperature.
16. At exactly 1 minute after starting the first digest, add 2 µL 1:100 pronase solution to one aliquot of the control sample, mix well, and incubate at room temperature.
17. After 30 minutes stop the digestion of the first aliquot began by adding 3 µL cold 20x protease inhibitor solution, mixing well, and placing on ice.
18. Stop all the remaining digestions in the order they were started in 1-minute intervals by adding 3 µL cold 20x protease inhibitor solution to each, mixing well, and placing on ice.

Part 2: Fractionation and Mass Spectrometry

Materials

1M dithiothreitol (DTT)

Zeba Spin Desalting Columns, 7K MWCO, 0.5 ml (Thermo Scientific cat. no. 89882)

GelFree 8100 Fractionation Station (Protein Discovery)

10% Tris-Acetate Cartridge Kit (Protein Discovery cat. no. 42105)

1. Perform DARTS for the small molecule of interest using pronase to protein ratios of 1:300 or 1:100.
2. Desalt 200 µg aliquots from each DARTS sample with Zeba Spin Desalting Columns according to the directions provided by Thermo Scientific.

If SILAC or another stable isotope labeling method will be used for quantitative comparison, combine 100 µg labeled protein from the compound-treated sample and 100 µg labeled protein from the control sample into one and process as a single sample throughout this protocol.

3. Prepare the samples for fractionation by bringing their total volumes up to 112 µL with dH_2O, adding 8 µL 1M DTT and 30 µL 5x sample buffer (provided in cartridge kit), mixing, and heating to 50°C for 10 minutes.
4. Allow the samples to cool to room temperature and load each one into an individual lane of a 10% Tris-Acetate cartridge.
5. Separate the samples into a desired number of fractions according to the instructions provided by Protein Discovery.

The GelFree system allows complete user control over how many fractions can be collected and to what molecular weight ranges those fractions correspond. We recommend initially using the general fractionation method provided with the

cartridge kit, which will yield 12 evenly spread fractions of proteins ranging in size from about 5 to 100 kDa. If desired, variations from this method can be tested on a single DARTS sample and the fractions analyzed by traditional SDS-PAGE and silver staining to determine suitability prior to mass spectrometry analysis.

6. Prepare the fractions for mass spectrometry analysis by alkylation and trypsin digestion using the enhanced Filter-Aided Sample Preparation (eFASP) protocol [3] (see below for detailed protocol).

   FASP/eFASP is recommended because it is a robust, efficient method for preparing SDS-containing samples for mass spectrometry. Alternatively, the fractions can be TCA or acetone precipitated and in-solution reduced, alkylated, and digested, but it is critical that the preparation method used removes the SDS from the samples prior to mass spectrometry analysis.

Following is a detailed outline of the eFASP digestion and cleanup procedure used in processing the DARTS-MS samples.

1. Prepare 10 mL 8M Urea in 0.1M Tris-HCl pH 8.5
2. Add 200 µl 8M Urea/0.1M Tris pH 8.5 into a 5 kDa MW-cutoff column
3. Add samples into separate 5 kDa MW-cutoff columns (up to 200 µl)
4. Centrifuge at 14,000 g until about 5 µl remains
5. Add 200 µl 8M Urea/0.1M Tris pH 8.5 into each 5 kDa MW-cutoff column

6. Centrifuge as before down to ~5 µl
7. Prepare 100mM Iodoacetamide in 8M Urea/0.1M Tris pH 8.5
8. Discard flow through and add 100 µl 100mM Iodoacetamide in 8M Urea/0.1M Tris pH 8.5 to each sample
9. Mix samples in thermomixer at 600 rpm for 1 minute
10. Incubate samples for 5-10 minutes in the dark at room temp. without mixing
11. Centrifuge samples at 14,000 g for 20 minutes
12. Prepare 6 mL 8M Urea in 0.1M Tris-HCl pH 8.0
13. Discard flow through and add 100 µl 8M Urea/0.1M Tris pH 8.0 to each sample
14. Centrifuge samples at 14,000 g down to ~5 µl
15. Repeat this wash step twice with 100 µl each time
16. Prepare new collection tubes for each sample for tryptic digest
17. Transfer filters to new collection tubes and bring the total volume up to 40 µl with 8M Urea/0.1M Tris pH 8.0
18. Mix in thermomixer at 600 rpm for 1 minute
19. Add 125 µl 50mM ammonium bicarbonate containing 1:100 Trypsin
20. Mix samples immediately in thermomixer for 1 minute at 600 rpm
21. Incubate digestions 4 hour to overnight at room temp or 37°C
22. Centrifuge samples for 20 minutes at 14,000 g
23. Add 50 µl 0.5M NaCl and repeat centrifugation
24. Discard filters and add 5 µl 5% TFA to each eluate
25. Proceed to Desalting Peptides Protocol or store samples at -20°C

Desalting Peptides Protocol using C18 StageTip

Tip packing material: Empore C18 90 mm Extraction Disc, Model 2315

Tip Adaptors: Protea SpinTip Adaptors (SP-990-8)

1. Dry down the sample in a speedvac, then resuspend in 100 µL loading buffer (0.5% HOAc), check to make sure the sample is acidic (pH 3 or less) and add more HOAc if needed.
2. Prepare the C18/C18 StageTips by packing 200-µL Eppendorf tips with 2 layers of Empore C18 (maximum binding capacity for a piece of membrane is ~20 ug).
3. Condition the C18/C18 tips as follows:
    a. Add 100-µL of 100% MeOH and spin at 2k x g for 25-30 seconds (~10-µL should remain). Do not let the membrane dry out after this point
    b. Add 100-µL of C18 Elution buffer (80% ACN/0.5% HOAc) to remove contaminants on C18 and spin at 2k x g for 25-30 seconds (~10-µL should remain)
    c. Add 100-µL of Loading buffer (0.5% HOAc), spin at 2k x g for 25-30 seconds (~10-µL should remain)
4. Transfer your peptide sample to the conditioned C18/C18 SpinTip. spin and reload the filtrate onto the tip
5. Add 100-µL of Loading buffer and spin at 2K x g (~30 seconds) to wash and transfer the StageTip to a fresh tube.
6. Add 40-µL of C-18 Elution buffer (80% ACN/0.5%HOAc) and collect the flow through

7. Dry down samples in speedvac and resuspend in 25 µL of 1% acetonitrile/0.1% formic acid for MS analysis.

Perform LC-MS/MS analysis of the fractions.

For most of the LC-MS/MS analyses in the subsequent chapters the following equipment and settings were used:

Thermo Easy nano-LC:

Sample volume: 2 µL

Loading volume: 6 µL

Gradient: Start at 5% B and go to 40% B over 2 hours then go to 100% B over 10 minutes and stay at 100% B for 10 minutes. Total run time 140 minutes.

LC flow rate: 300 nL/minute

Solvent A: Water, 0.1% formic acid

Solvent B: Acetonitrile, 0.1% formic acid

Pre-column: Acclaim PepMap RSLC Nano-Trap Column

Particle Size: 3 µm

Pore Size: 100 Å

Stationary Phase: C18

Length: 20 mm

Diameter: 0.075mm

Analytical column: Acclaim PepMap RSLC Nano

Particle Size: 2 µm

Pore Size: 100 Å

Stationary Phase: C18

Length: 250 mm

Column i.d.: 75 µm

Thermo Q-Exactive q-orbitrap:

Top-10 dds

Dynamic exclusion time: 30 s

Polarity: Positive

Full MS resolution: 70K

Full MS AGC target: 1e6

Full MS Scan range: 200-2000 m/z

MS2 resolution: 17.5K

MS2 AGC target: 2e5

Raw mass spectrometry data was extracted with either Proteome Discoverer 1.4 or RawXtract version 1.9. The MS2 files were searched against either a human or mouse Uniprot database (downloaded January 2015 and May 2015 respectively) using SEQUEST, Mascot, or ProLuCID search algorithms. Search results were filtered with DTASelect for a false positive rate of less than 5%. The data from all fractions was compared using Protomap [4].

2.3 CONCLUSION

DARTS-MS combines the advantages that our previous DARTS protocol has over existing affinity and non-affinity protein target identification methods while also providing higher sensitivity and analytical rigor. The method still does not require the modification or immobilization of the small molecule or target and performs all reactions under native, physiologically relevant conditions. Additionally, it provides an unbiased approach to identify all small-molecule binders within the proteome including both on-target and off-target interactions. Whereas previous versions of the DARTS protocol relied on methods with low sensitivity and qualitative results, such as SDS-PAGE, DARTS-MS can potentially delve much deeper into the proteome and identify small-molecule targets with very low abundance while also providing quantitative data to assist analysis. The protocols detailed in this chapter are intended to provide enough detail for replication in any well-equipped laboratory. While adaptation of the protocol may occasionally be necessary, the preceding methods have been well optimized for most systems. The following two chapters detail the use of DARTS-MS on a wide variety of small molecules in both whole cell lysate and subcellular fractions.

2.4 FIGURES

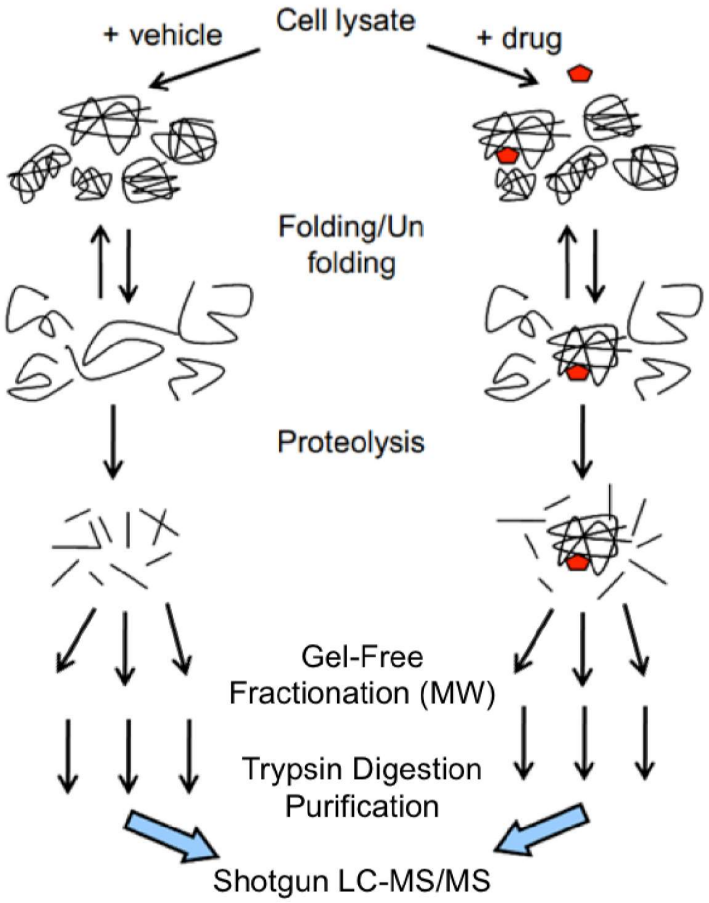


Figure 2-1: Enhanced DARTS-MS platform. A collection of protein (e.g. a cell lysate) is split into two fractions one of which is treated with a small molecule (e.g. drug) and the other receives an equal amount of solvent and serves as a control. The samples are allowed to incubate and then undergo a limited proteolysis. Proteins/peptides are separated according to molecular weight using gel-free fractionation. Each molecular weight fraction is digested completely with trypsin and subjected to LC-MS/MS. The control and treated samples are compared and proteins protected from proteolysis in the treated sample are identified.

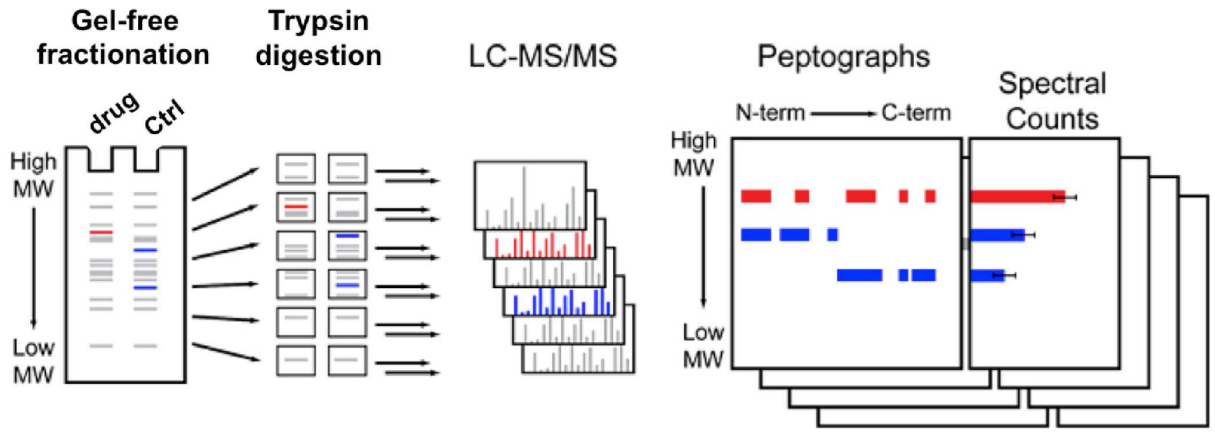


Figure 2-2: Protomap analysis. Samples are subjected to limited proteolysis and then are fractionated according to molecular weight. Each molecular weight fractionation is digested completely by trypsin and measured by LC-MS/MS. Protomap combines information from the fractionation and LC-MS/MS identification to create Peptographs. These compare peptides for a particular protein from the drug-treated sample (red) and control sample (blue). In the example shown, the protein's tryptic peptides are observed in a higher molecular weight fraction in the drug treated sample than the control sample; indicating that the protein is protected from proteolysis in the presence of the drug and thus is likely to bind to the drug [4].

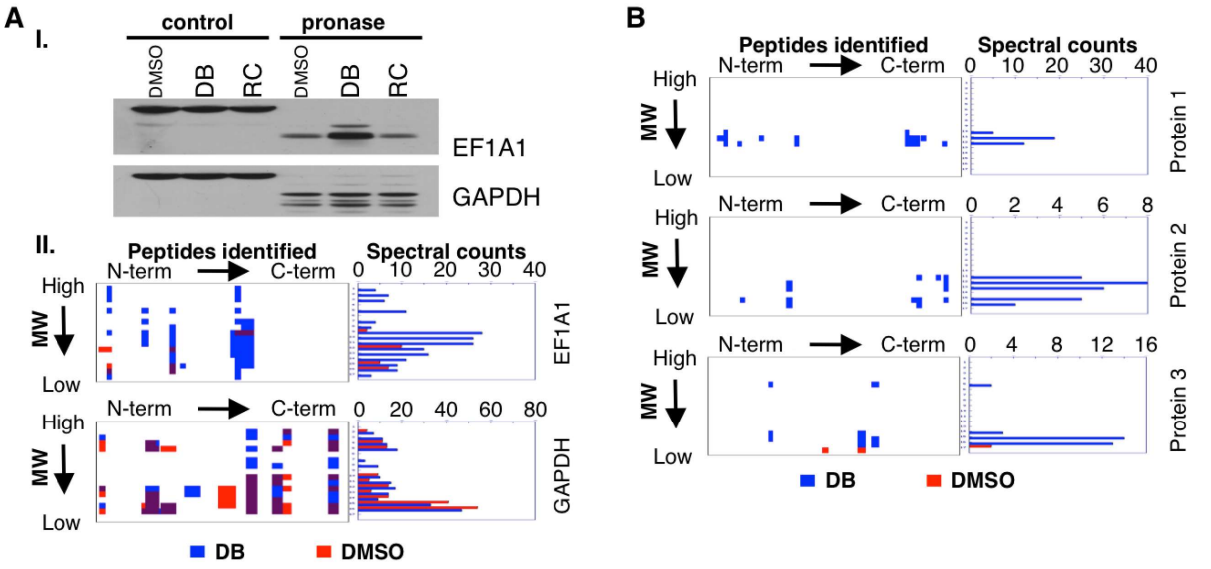


Figure 2-3: Target identification using Gel-Free DARTS-MS. (A) Proof-of principle: rediscovery of EF1A1 as a target of didemnin B (DB). Jurkat cell lysates were digested with pronase in the presence of 1% DMSO, 10 µg/mL DB, or a mixture of random compounds (RC). The lysates were analyzed with Western blotting (I) or LC-MS/MS (II), demonstrating that EF1A1 (but not GAPDH control) is specifically protected by DB (blue). For LC-MS/MS, each sample was separated into 17 liquid fractions according to molecular weight (MW) using the GelFree8100 system and analyzed with LC-MS/MS. (B) Discovery of additional proteins of lower abundance that are specifically protected by DB. Proteins 1-3 correspond to CCT3, CCT5 and CCT7 respectively from the TCP1 complex.

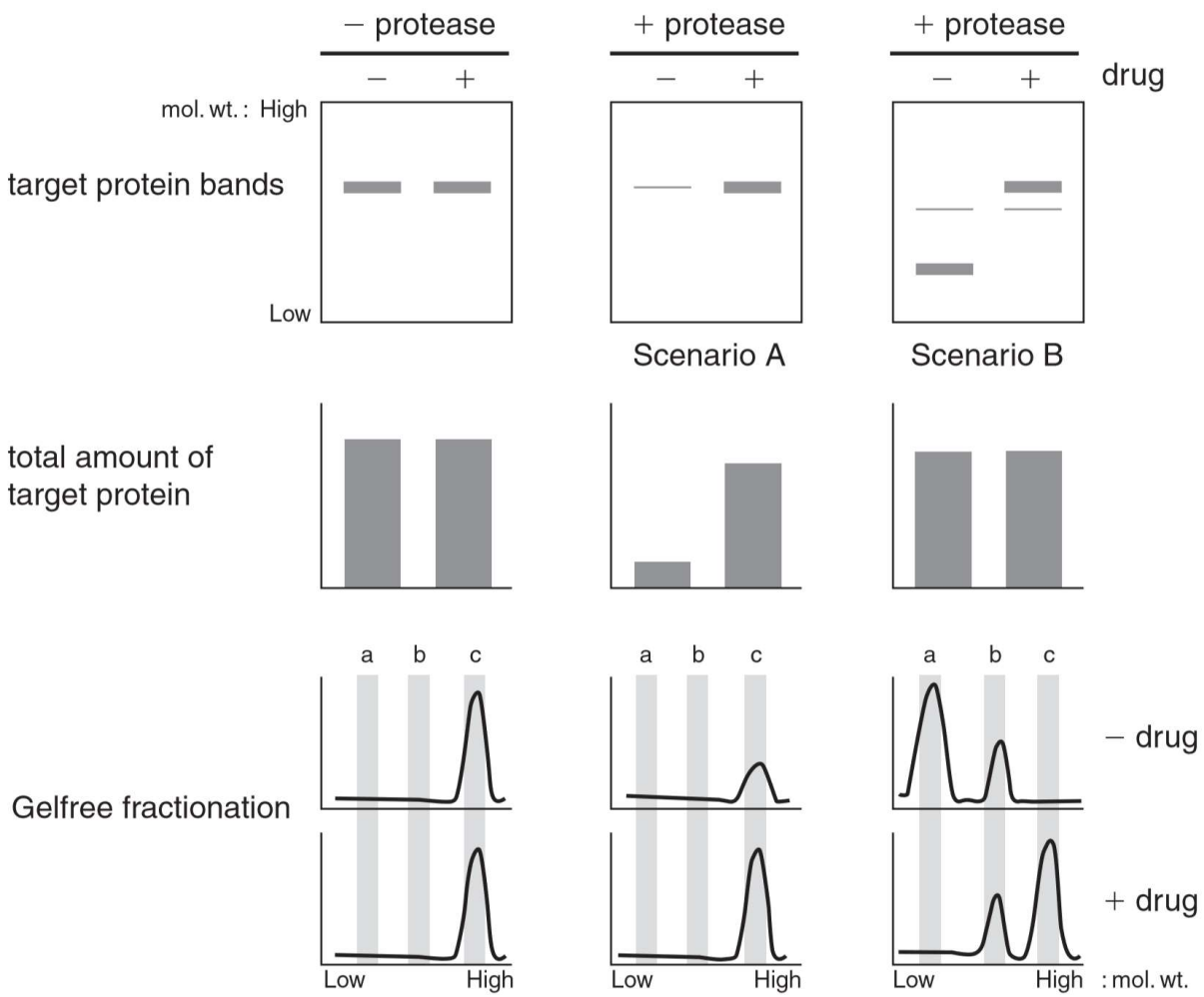


Figure 2-4: DARTS target identification by GelFree fractionation vs. whole-sample shotgun MS. The starting amount of target protein is equivalent in aliquots of lysate containing and lacking the small molecule. Compound binding causes differential proteolysis of the target protein in DARTS. In Scenario A, the target protein is nearly completely digested in the absence of compound, whereas most of the target protein is undigested in the presence of compound. Both GelFree fractionation and whole sample shotgun MS analysis are able to identify the target protein here due to its depletion from the control sample. In Scenario B, however, the target protein is only partially digested in the control sample. Here the total amount of target protein is not significantly different between the compound-treated and control samples despite clear protection from

proteolysis by the compound. While whole sample shotgun MS is unable to identify the target protein in this case, the GelFree fractionation approach that takes molecular weight into consideration can identify the target protein based upon its differently-sized fragments in the two samples [7].

2.5 REFERENCES

1. Lomenick, B., Hao, R., Jonai, N., Chin, R.M., Aghajan, M., Warburton, S., Wang, J., Wu, R.P., Gomez, F., Loo, J.A. and Wohlschlegel, J.A., 2009. Target identification using drug affinity responsive target stability (DARTS). Proceedings of the National Academy of Sciences, pp.pnas-0910040106.
2. Witkowski, C. and Harkins, J., 2009. Using the GELFREE 8100 Fractionation System for molecular weight-based fractionation with liquid phase recovery. Journal of Visualized Experiments: JoVE, (34).
3. Erde, J., Loo, R.R.O. and Loo, J.A., 2014. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. Journal of Proteome Research, 13(4), pp.1885-1895.
4. Dix, M.M., Simon, G.M. and Cravatt, B.F., 2008. Global mapping of the topography and magnitude of proteolytic events in apoptosis. Cell, 134(4), pp.679-691.
5. Crews, C.M., Collins, J.L., Lane, W.S., Snapper, M.L. and Schreiber, S.L., 1994. GTP-dependent binding of the antiproliferative agent didemnin to elongation factor 1 alpha. Journal of Biological Chemistry, 269(22), pp.15411-15414.
6. Arnold, U., Rücknagel, K.P., Schierhorn, A. and Ulbrich-Hofmann, R., 1996. Thermal unfolding and proteolytic susceptibility of ribonuclease A. European Journal of Biochemistry, 237(3), pp.862-869.
7. Lomenick, B., Jung, G., Wohlschlegel, J.A. and Huang, J., 2011. Target identification using drug affinity responsive target stability (DARTS). Current Protocols in Chemical Biology, 3(4), pp.163-180.

8. Lomenick, B., Olsen, R.W. and Huang, J., 2010. Identification of direct protein targets of small molecules. ACS Chemical Biology, 6(1), pp.34-46.

CHAPTER 3 : WHOLE CELL METABOLITE STUDY WITH DARTS-MS

3.1 INTRODUCTION

One of the motivations for developing the DARTS-MS (drug affinity responsive target stability-mass spectrometry) protocol was to have a better tool for identifying and characterizing the interactions of metabolites within the cell. Although it is well accepted that metabolites have a vast array of functions within the cell, including regulation of enzymatic activity, post-translational modifications and gene expression, it is equally apparent that we are far from understanding all of the interactions of even common metabolites. Furthermore, while some areas, such as cancer research, emphasize the importance of metabolomics, it is not generally applied to the area of aging research [1-5]. While aging is an extraordinarily complex process, it has been demonstrated that certain genetic mutations and drugs can markedly extend the lifespan of many model organisms [6, 7]. This result coupled with the importance of aging to human health and disease, offers a tantalizing possibility of improving the treatment of age-related illnesses and our general understanding of the aging process.

While many genes have been found that regulate aging and can affect lifespan, there are relatively few small molecules that are known to be directly involved in the regulation of aging. There are even fewer such molecules that are well characterized and have known molecular targets or mechanisms. However, it was recently discovered that the naturally occurring metabolite α-ketoglutarate (α-KG) has significant pro-longevity effect in *C. elegans* (Figure 3-1). Previous work with DARTS suggests that ATP synthase

subunit beta is directly bound by α-KG [8]. This in turn, may inhibit ATP synthase and reduce cellular respiration, thereby reducing oxidative stress on the cell and increasing lifespan. However, further analysis of the pathway is warranted as it may reveal important mechanisms or therapeutically amenable targets.

ATP synthase is the primary energy-producing enzyme in the cell and is highly conserved across species [9, 10]. While the total suppression of the mitochondrial respiratory chain is detrimental, a partial suppression can have health benefits. In *C. elegans*, for example, deficiencies in the mitochondrial electron transport chain due to mutation or RNAi knockdown result in longer lives than control organisms [11-14]. The numerous parallels between α-KG treatment and ATP synthase knockdown, in addition to the identification of ATP synthase subunit beta as a binding protein of α-KG, have led to the suggestion that inhibition of ATP synthase is the primary cause of lifespan extension in the presence of α-KG [15]. However, there are many other α-KG targets in the cell. HSPD1 (60 kDa heat shock protein) and PKM2 (Pyruvate kinase isozymes M1/M2) have both been found to interact with α-KG by DARTS and are both implicated in aging-related illnesses. To better understand the regulatory role of α-KG within the cell we will use DARTS-MS to identify additional binding proteins that may play a role in lifespan extension.

Increased levels of α-KG have been reported in numerous starved animals (including *C. elegans*) and in exercising humans [16-19]. These increased levels of α-KG are generally explained “by starvation-based anaplerotic gluconeogenesis, which activates glutamate-linked transaminases in the liver to provide carbon derived from amino acid catabolism” [15]. It seems likely that α-KG is one of the key metabolites

responsible for lifespan extension during dietary restriction. However, α-KG may not be unique in this regard. It is possible that other α-keto acids produced during amino acid catabolism are involved in lifespan extension during dietary restriction. In this chapter, we will apply our improved DARTS-MS protocol to further explore the role of α-KG as well as the structurally similar α-ketobutyric acid (α-KB) and α-ketoisocaproic acid (α-KIC), two potentially important endogenous α-keto acids that are not well studied (Figure 3-2).

Dietary restriction, also known as calorie restriction, has been shown to extend lifespan and reduce the impact of age-related illnesses in a diverse group of organisms for nearly a century [20-22]. However, the mechanisms linking dietary restriction and longevity are still incompletely understood. Mapping the interactions of metabolites implicated in dietary restriction could expand this understanding and provide potential avenues for therapeutic development. Since aging is one of the primary risk factors in many common fatal illnesses (including heart disease, cancer, diabetes, and neurodegenerative disorders) it is an important area of research and one in need of new tools [23-27]. Utilizing DARTS-MS to investigate aging-related metabolites is a unique strategy that could uncover fruitful areas for future research into aging or the treatment and prevention of human diseases.

3.2 α-KETOGLUTARIC ACID

As discussed in the previous section, α-ketoglutaric acid (α-KG) is a naturally occurring metabolite involved in many cellular processes including the Krebs cycle and amino acid synthesis. It is also implicated in life-span extension and is thus a small molecule of great interest. However, given the large number of interactions α-KG has within the cell, it is not amenable to traditional target identification strategies. By using DARTS-MS we hope to map α-KG interactions, both novel and previously confirmed, within the cell.

The protocol employed is described in detail in CHAPTER 2. Briefly, Human Embryonic Kidney cells (HEK293) were grown, lysed, incubated with 100 μM α-KG, partially digest with pronase, fractionated by molecular weight, digested with trypsin and analyzed by LC-MS/MS. The raw MS data was analyzed with SEQUEST and Protomap. The data was then organized into “peptographs” that display the sequence coverage for a particular protein on the horizontal axis (from N-terminus on the left to C-terminus on the right) while migration of the partially digested proteins is shown on the vertical axis (with high molecular weight fractions on top and low molecular weight fractions on the bottom). The peptographs also display the average number of spectral counts for the particular protein in each size exclusion fraction. Potential binding targets are identified by changes in proteolytic cleavage as indicated by shifts in migration from higher to lower molecular weight species between the experimental and control samples.

Figure 3-3 is a peptograph for the protein prolyl 3-hydroxylase. In this case, we can see that the majority of peptides from the α-KG containing samples are found in

higher molecular weight fractions than the control samples. Thus it appears that α-KG is protecting the protein prolyl 3-hydroxylase from proteolysis. In fact, prolyl 3-hydroxylase is a known target of α-KG so this is an expected result. However, we also noticed that not all targets of α-KG are protected from proteolysis. In Figure 3-4, for example, we see the peptograph of transferrin receptor protein 1 and, in this case, the majority of peptides from the control samples are found in higher molecular weight fractions than the peptides from the α-KG containing samples. This would suggest that α-KG is rendering transferrin receptor protein 1 more susceptible to proteolysis. One possible explanation for this observation is that α-KG is disrupting a complex of proteins making the individual subunits of that complex more susceptible to proteolysis. Another possibility is that the conformation of the individual protein becomes more exposed and thus more susceptible to proteolysis. It is worth noting however, that the vast majority of proteins in our samples are not made more resistant or sensitive to proteolysis in the presence of α-KG. Figure 3-5 shows the peptograph for glucose-6-phosphate isomerase, and based on the results, it appears this protein is not impacted by the presence or absence of α-KG. Instead we see a fairly even distribution of partially digested glucose-6-phosphate isomerase fragments from both the control and α-KG containing samples.

There were approximately 4000 proteins in total considered in the α-KG experiment. While, there were many more proteins detected across all samples, we only considered those that had high quality spectra and confident protein IDs (false positive rate <5%). Furthermore, in order for the Protomap analysis to have a high degree of statistical confidence, it was necessary to collect many spectra (at least 10 spectra per protein per sample). This is to insure that we can confidently assign the molecular weight

fraction to which a protein localizes. Indeed, as long as we can accurately measure a difference between the average fraction of a protein in the control sample and the experimental sample, this is sufficient to infer protection or sensitization to proteolysis. As a result, the exact quantification of protein is unnecessary for DARTS-MS to identify small-molecule targets (assuming the starting conditions are the same between the control and experimental samples). Quantification will be discussed in more detail in CHAPTER 5.

There were 151 proteins detected that had significant differences in migration between the α-KG and control samples (see Figure 3-6 for an overview of the dataset and Table 3-1 for a complete list of proteins). While this is a large number of putative targets, it is worth noting that α-KG is a highly promiscuous molecule involved in numerous cellular processes. Indeed, based on current literature, there are more than 120 confirmed or suspected targets of α-KG (Table 3-2). Of these known targets, nearly 30 were detected with enough confidence to be analyzed by Protomap (Table 3-3). All but three of the detected protein targets had significant differences between the α-KG and control samples. To put this in context, less than 4% of all proteins detected differed significantly between the experimental and control samples but nearly 90% of expected α-KG protein targets differed significantly between the control and experimental samples. This is an important result as it indicates a fairly low false negative rate for this analysis. Determining the false positive rate is not possible for this analysis as α-KG has many potential targets within the cell most of which have not been confirmed. However, the false positive rate will be investigated in CHAPTER 4.

The validation of the putative α-KG targets identified in this study as well as determining their underlying biological significance is an ongoing project. However, the use of DARTS-MS to characterize additional metabolites with similar anti-aging activity could help identify networks of proteins related to aging as well as conserved pharmacologically amenable targets. Two metabolites structurally similar to α-KG, α-ketobutyric acid (α-KB) and α-ketoisocaproic acid (α-KIC), will be examined in the following sections.

3.3 α-KETOBUTYRIC ACID

Like α-ketoglutaric acid (α-KG), α-ketobutyric acid (α-KB) is a naturally occurring metabolite that forms during amino acid catabolism. Given the similarity in structure and origin of the two molecules, α-KB seems a logical compound to test for a possible role in lifespan extension or dietary restriction. Indeed, there is already some evidence to suggest that α-KB is involved. For instance, α-KB levels are reported to be elevated in long-lived, but not short-lived, *C. elegans* mutants that have a partial disruption in the mitochondrial electron transport chain [28]. Preliminary results from a separate ongoing study indicate that supplementation with α-KB can delay or even reverse certain aging related phenotypes, including hair loss and cataracts, in aged mice (Figure 3-7). This indicates not only possible anti-aging effects of α-KB but also suggests that these effects could be conserved in mammals. If this is the case, α-KB could prove a more effective therapeutic than α-KG, since α-KB is naturally membrane-permeable whereas α-KG is transported slowly across plasma or mitochondrial membranes and is not readily permeable. Given the potential importance of α-KB and its similarity to α-KG, we have used DARTS-MS to help identify targets of α-KB and better understand its mechanism of action. As with α-KG, modification of the α-KB molecule would likely alter its binding affinity making DARTS-MS the ideal platform for target identification.

The protocol employed is described in detail in CHAPTER 2. Briefly, Human Embryonic Kidney cells (HEK293) were grown, lysed, incubated with 100 μM α-KB, partially digest with pronase, fractionated by molecular weight, digested with trypsin and analyzed by LC-MS/MS. The raw MS data was analyzed with SEQUEST and Protomap.

Potential binding targets are identified by changes in proteolytic cleavage as indicated by shifts in migration from higher to lower molecular weight species between the experimental and control samples.

There were around 4000 proteins detected in the α-KB experiment and 37 proteins detected that had significant differences in migration between the α-KB and control samples (see Figure 3-8 for an overview of the dataset and Table 3-4 for a complete list of proteins). There are not many proteins known to interact with α-KB (Table 3-5) and, unfortunately, none of the known targets were detected in this analysis. While α-KB is structurally similar to α-KG, α-KB is a much less polar compound due to the single carboxylic acid group. It is, therefore, not unreasonable to expect that α-KB will have substantially different target proteins than α-KG. Indeed the results suggest that α-KB has many fewer protein targets than α-KG. Given the many cellular processes that α-KG is involved with and the relatively few known targets of α-KB this is not surprising, however, there is still some overlap. Approximately one third of the protein targets of α-KB are also targets of α-KG (Table 3-6). Although assigning biological function to these interactions is beyond the scope of DARTS-MS, identifying and mapping out these networks of metabolite interactions is an important first step towards understanding their role in human health and aging.

3.4 α-KETOISOCAPROIC ACID

Like α-ketoglutaric acid (α-KG) and α-ketobutyric acid (α-KB), α-ketoisocaproic acid (α-KIC) can be produced from the catabolism of amino acids under specific physiological states. All three compounds also have structural similarities (Figure 3-2). α-KIC is known to be an intermediate in metabolism of leucine and has been used as an athletic supplement [29]. However, relatively little is known regarding its interactions within the cell.

The protocol employed is described in detail in CHAPTER 2. Briefly, Human Embryonic Kidney cells (HEK293) were grown, lysed, incubated with 100 μM α-KIC, partially digest with pronase, fractionated by molecular weight, digested with trypsin and analyzed by LC-MS/MS. The raw MS data was analyzed with SEQUEST and Protomap. Potential binding targets are identified by changes in proteolytic cleavage as indicated by shifts in migration from higher to lower molecular weight species between the experimental and control samples.

There were around 4000 proteins detected in the α-KIC experiment and 112 proteins detected that had significant differences in migration between the α-KIC and control samples (see Figure 3-9 for an overview of the dataset and Table 3-7 for a complete list of proteins). Many of the detected proteins are directly involved in branched amino acid synthesis, which is in keeping with previous results. However, given the number of putative interactions, α-KIC may serve direct or related regulatory roles.

3.5 FIGURES

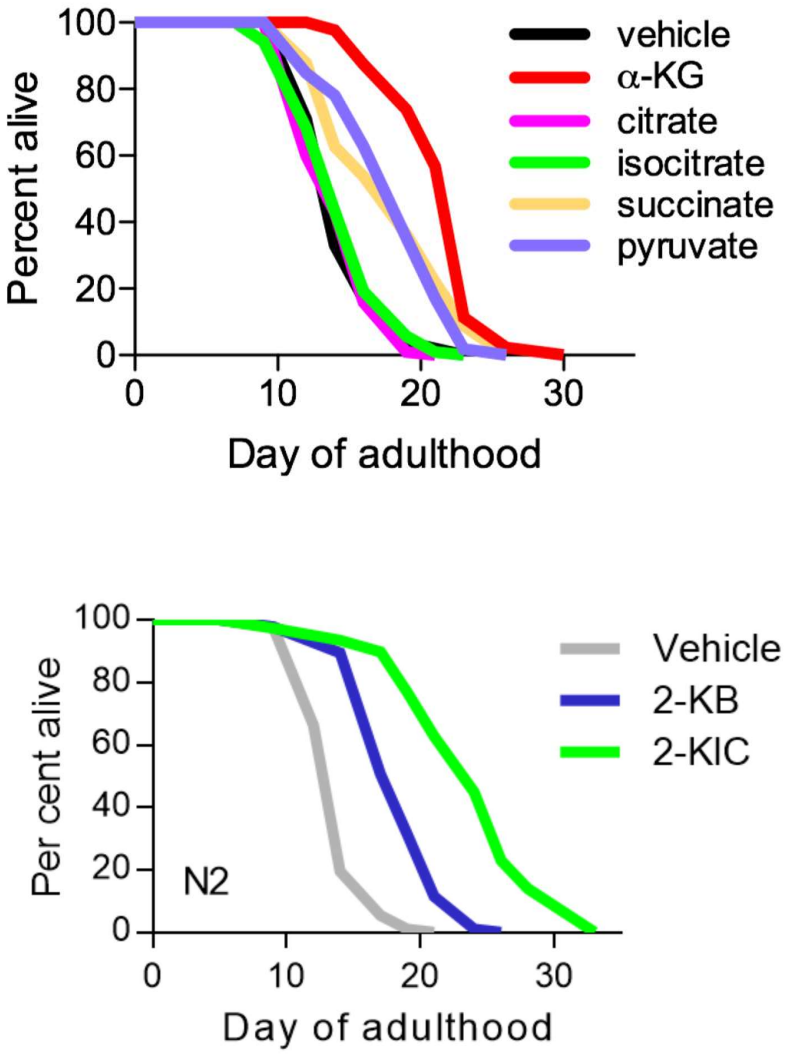


Figure 3-1: Lifespan screen identifies anti-aging metabolites. In order to gain insight into the regulation of aging by endogenous small molecules, the Huang lab screened normal metabolites and aberrant disease associated metabolites for their effects on adult lifespan of *C. elegans*. Molecules that exhibited pro-longevity effects from the screen include pyruvate, which was reported recently [30] and α-KG and succinate, whose effects on lifespan were not known previously. (Top) In particular, α-KG extended the

adult lifespan of *C. elegans* by ~50%, and also delayed the age dependent decline in mobility and youthfulness. (Bottom) α-KB and α-KIC were similarly effective at extending lifespan.

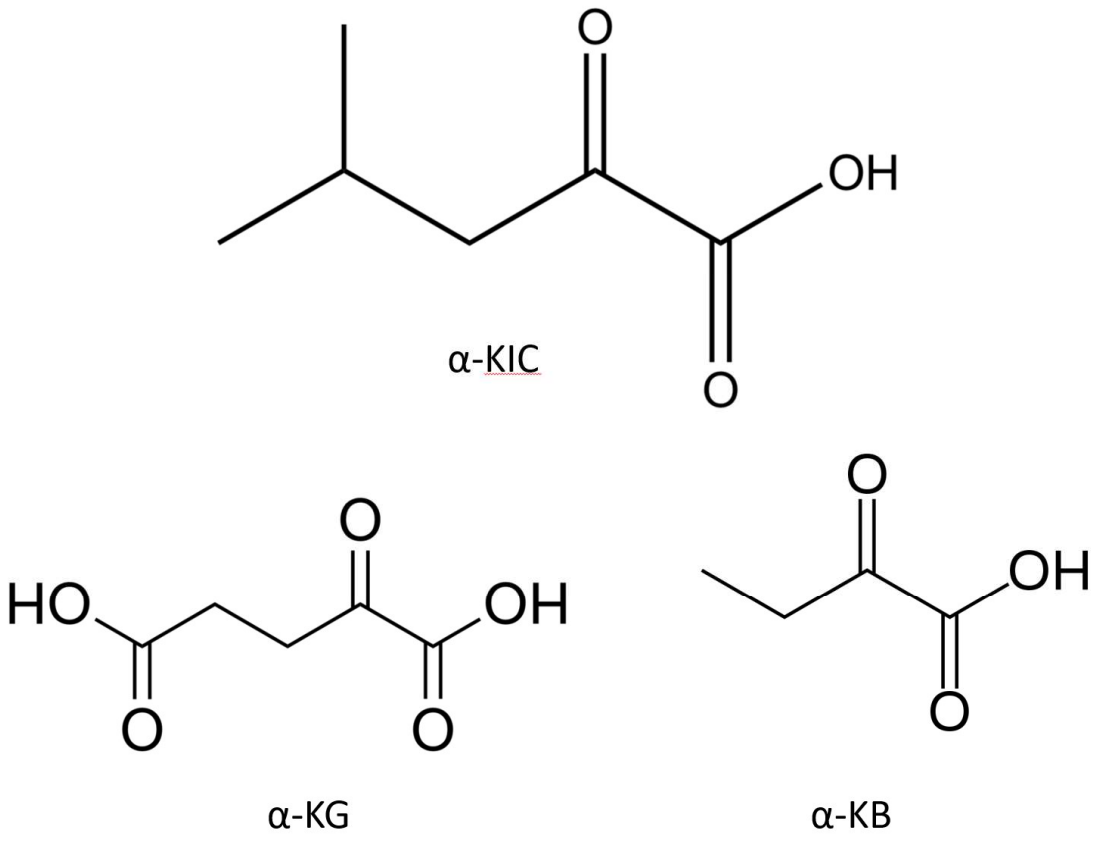


Figure 3-2: The structures of metabolites are depicted (α-KIC) α-ketoisocaproic acid (α-KG) α-ketoglutaric acid, and (α-KB) α-ketobutyric acid

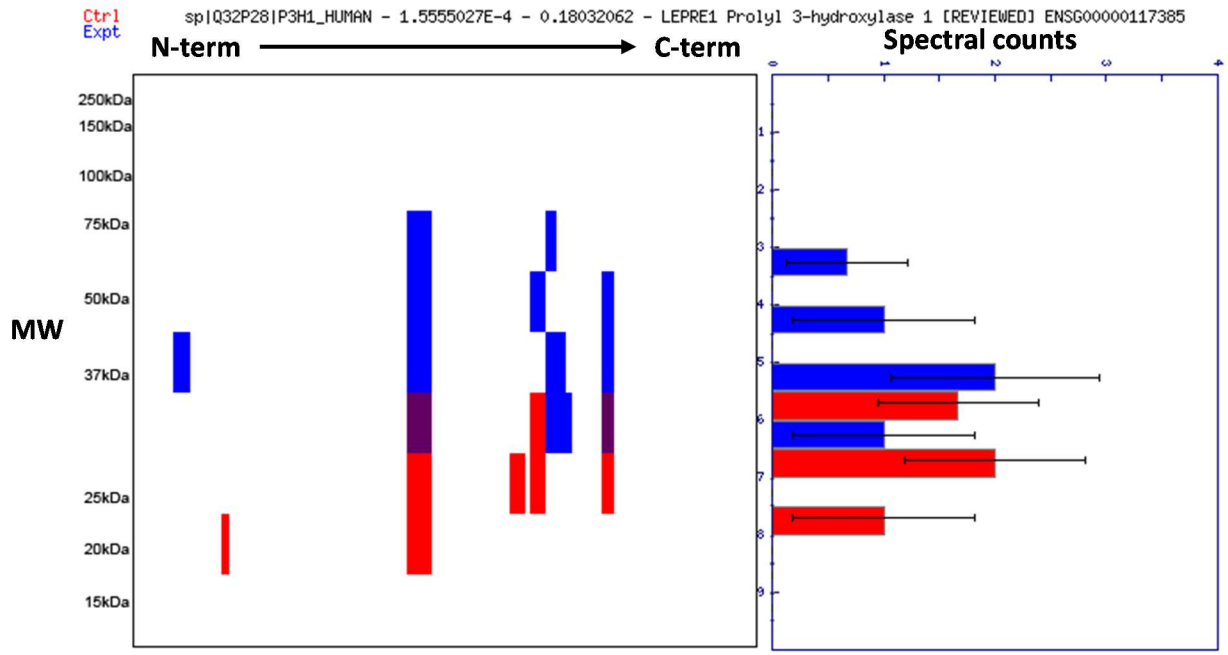


Figure 3-3: Results for prolyl 3-hydroxylase in α-KG DARTS-MS experiments All detected peptides related to prolyl 3-hydroxylase in the α-KG DARTS-MS experiments were integrated into peptographs using Protomap. In the left panel, sequence coverage for the protein is shown on the horizontal axis (from N-terminus on the left to C-terminus on the right) while migration of the partially digested proteins is shown on the vertical axis (with high MW on top and low MW on the bottom). The right panel shows the average number of spectral counts for the particular protein in each size exclusion fraction. Potential binding targets are identified by changes in proteolytic cleavage as indicated by shifts in migration from higher to lower MW species between the experimental (blue) and control (red) samples (purple is the overlap of both samples). In this case, it appears that α-KG is protecting the protein from proteolysis.

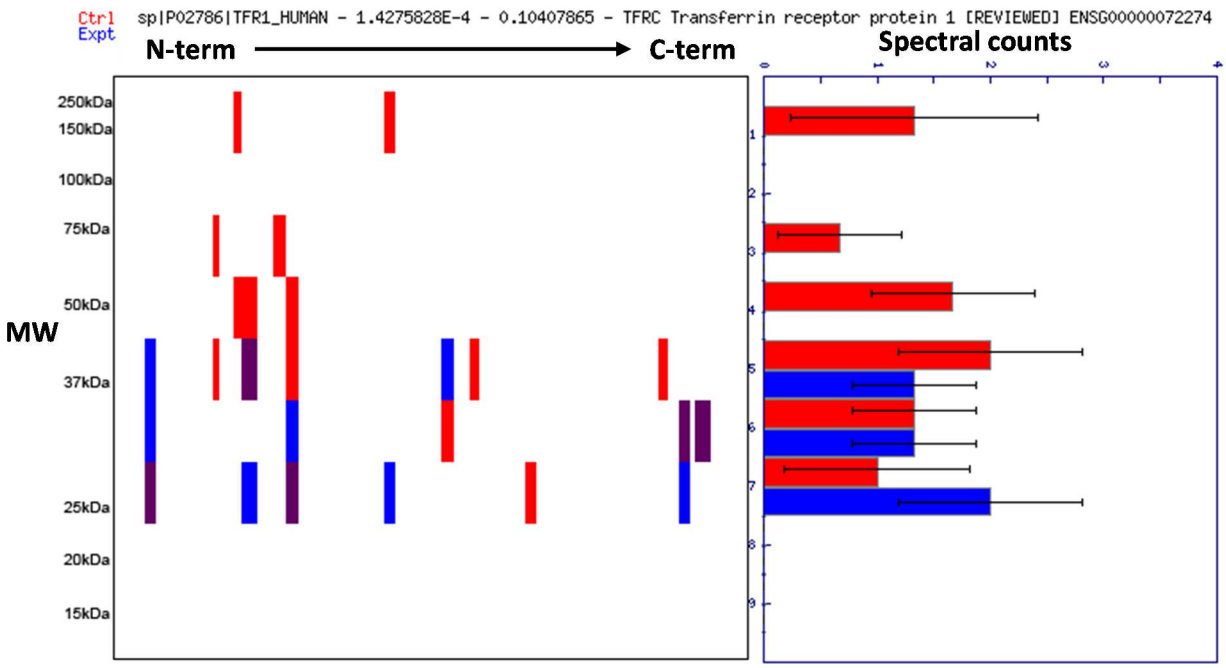


Figure 3-4: Results for transferrin receptor protein 1 in α-KG DARTS-MS experiments All detected peptides related to transferrin receptor protein 1 in the α-KG DARTS-MS experiments were integrated into peptographs using Protomap. In the left panel, sequence coverage for the protein is shown on the horizontal axis (from N-terminus on the left to C-terminus on the right) while migration of the partially digested proteins is shown on the vertical axis (with high MW on top and low MW on the bottom). The right panel shows the average number of spectral counts for the particular protein in each size exclusion fraction. Potential binding targets are identified by changes in proteolytic cleavage as indicated by shifts in migration from higher to lower MW species between the experimental (blue) and control (red) samples (purple is the overlap of both samples). In this case, it appears that α-KG is sensitizing the protein to proteolysis.

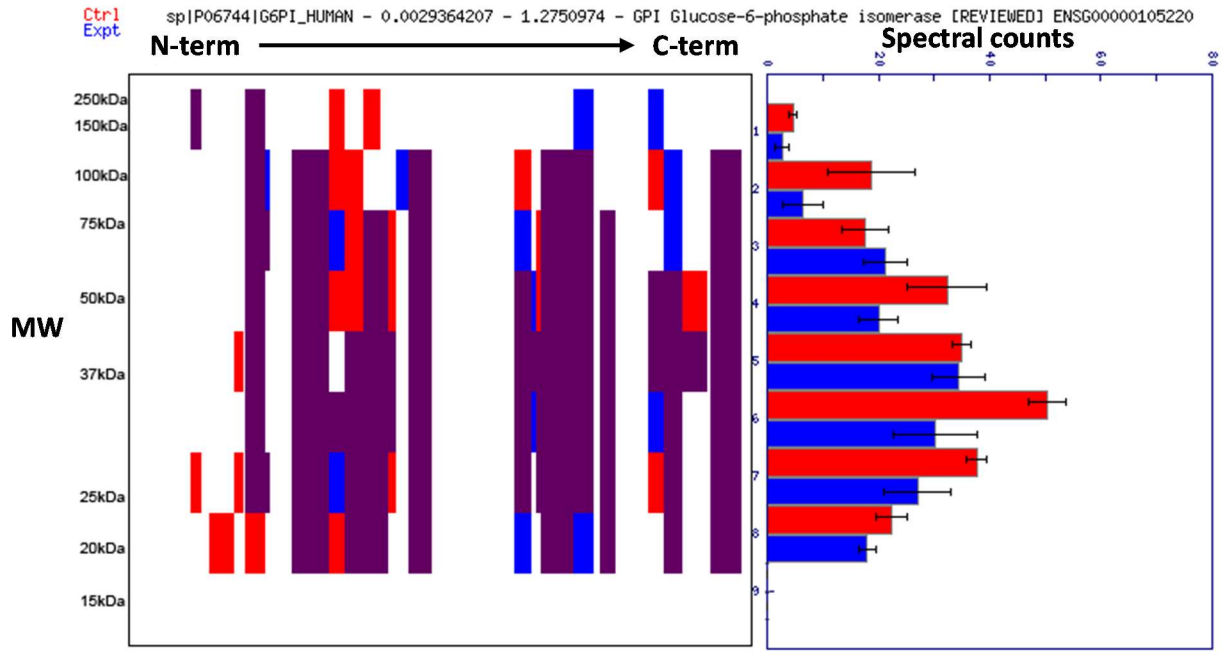


Figure 3-5: Results for glucose-6-phosphate isomerase in α-KG DARTS-MS experiments All detected peptides related to glucose-6-phosphate isomerase in the α-KG DARTS-MS experiments were integrated into peptographs using Protomap. In the left panel, sequence coverage for the protein is shown on the horizontal axis (from N-terminus on the left to C-terminus on the right) while migration of the partially digested proteins is shown on the vertical axis (with high MW on top and low MW on the bottom). The right panel shows the average number of spectral counts for the particular protein in each size exclusion fraction. Potential binding targets are identified by changes in proteolytic cleavage as indicated by shifts in migration from higher to lower MW species between the experimental (blue) and control (red) samples (purple is the overlap of both samples). In this case, it appears that α-KG has no impact on the protein's sensitivity to proteolysis.

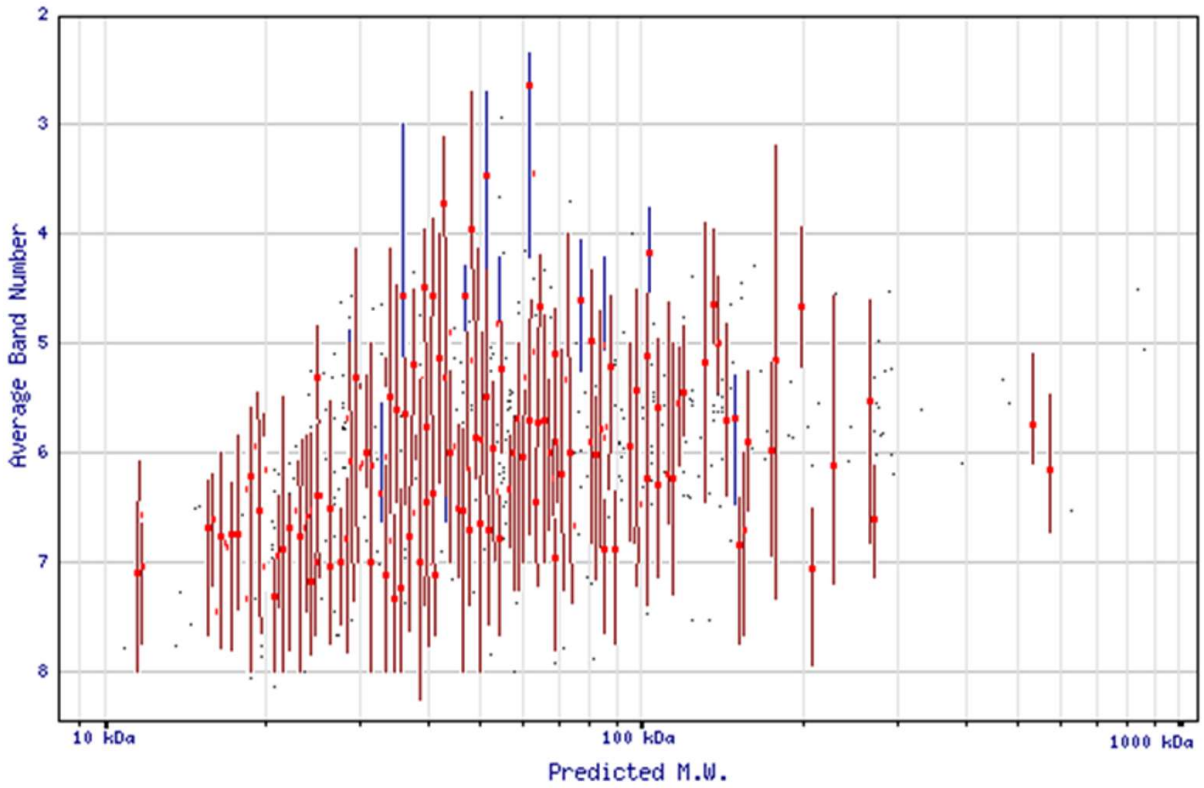


Figure 3-6: Shifts in molecular weight (MW) between α-KG and control samples indicating possible molecular targets. Each spot on the above plot represents a protein detected in both the experimental and control dataset. Red-spots with lines indicate proteins that have been protected from proteolysis by α-KG as evidenced by a shift in migration between control and experimental samples. Blue-spots with lines indicate proteins with an increased sensitivity to proteolysis. The size of the line indicating the magnitude of the shift. Black spots indicate no change in migration

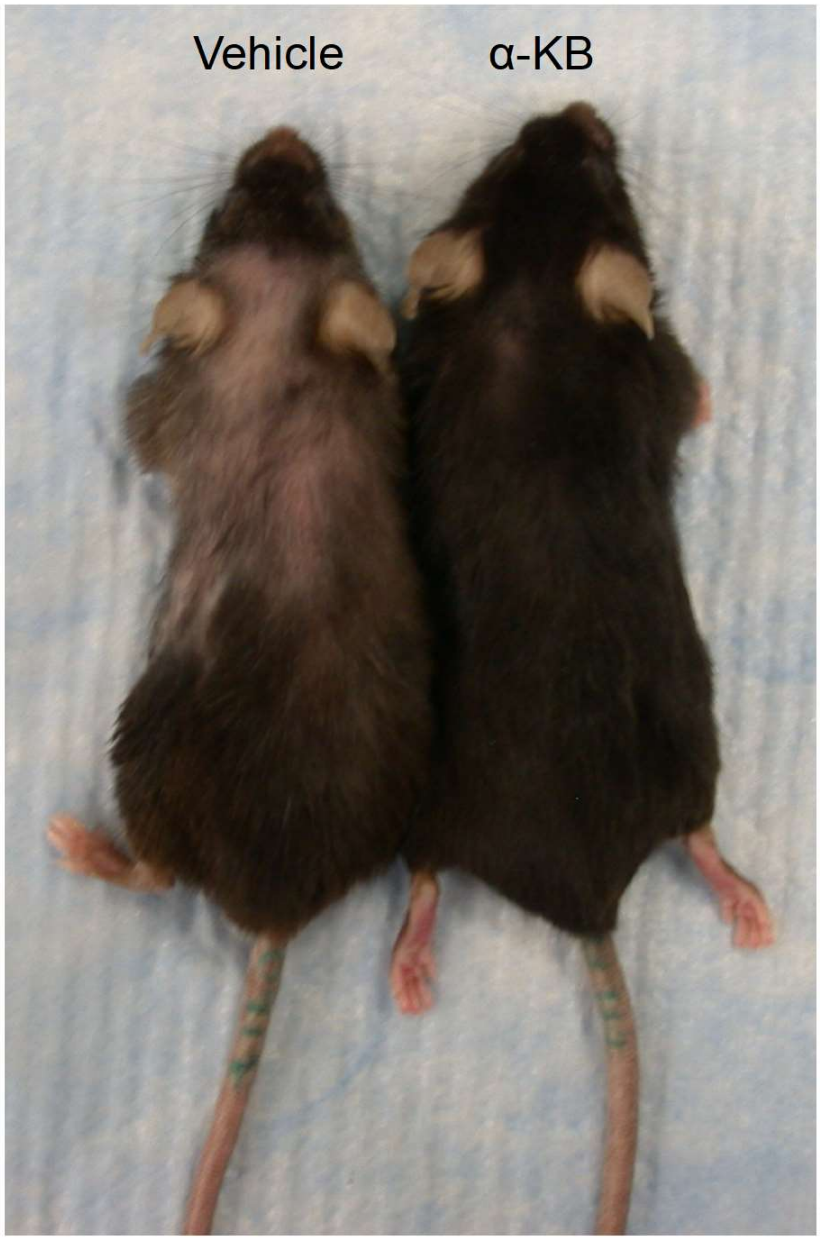


Figure 3-7: α-KB treated mice. Old C57BL/6 mice revert to younger appearance when supplemented with α-KB.

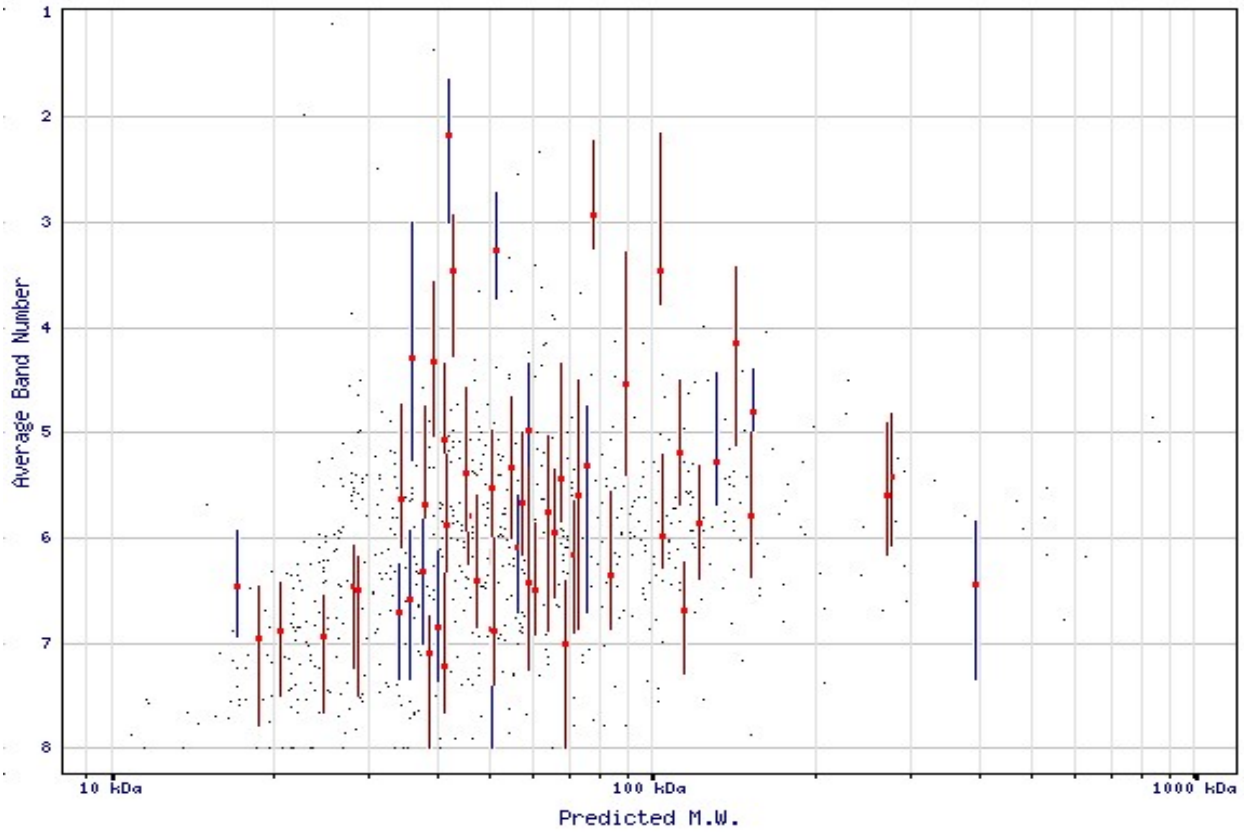


Figure 3-8: Shifts in molecular weight (MW) between α-KB and control samples indicating possible molecular targets. Each spot on the above plot represents a protein detected in both the experimental and control dataset. Red-spots with lines indicate proteins that have been protected from proteolysis by α-KB as evidenced by a shift in migration between control and experimental samples. Blue-spots with lines indicate proteins with an increased sensitivity to proteolysis. The size of the line indicating the magnitude of the shift. Black spots indicate no change in migration

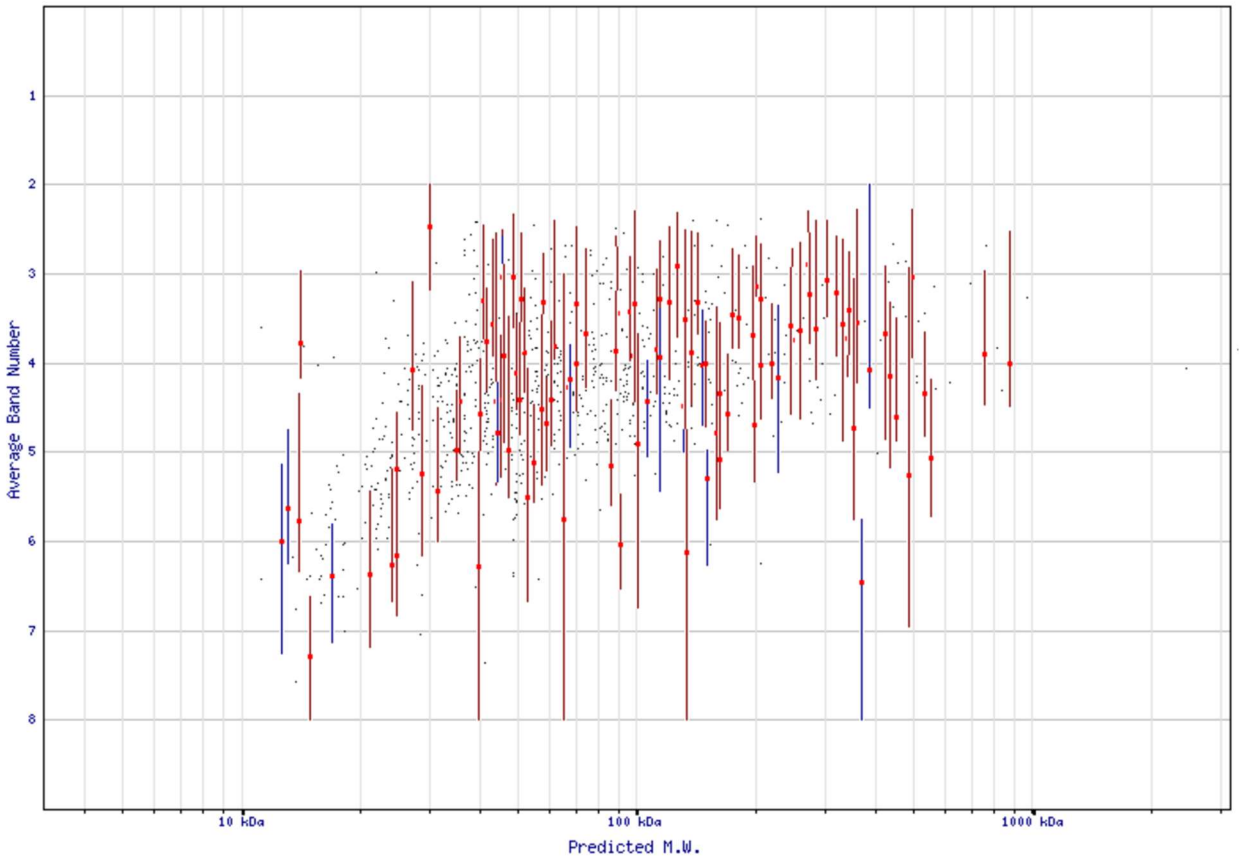


Figure 3-9: Shifts in molecular weight (MW) between α-KIC and control samples indicating possible molecular targets. Each spot on the above plot represents a protein detected in both the experimental and control dataset. Red-spots with lines indicate proteins that have been protected from proteolysis by α-KIC as evidenced by a shift in migration between control and experimental samples. Blue-spots with lines indicate proteins with an increased sensitivity to proteolysis. The size of the line indicating the magnitude of the shift. Black spots indicate no change in migration

3.6 TABLES

Table 3-1: Putative targets of α-KG based on shifts in migration. Each protein in the following table had a statistically significant difference between the α-KG samples and the control samples. The absolute value of the distance corresponds to the difference in average MW fraction between the experimental and control samples. Negative distances indicates a sensitization to proteolysis in the presence of the small molecule while a positive value corresponds to a protection from proteolysis.

Protein	Distance
ACACA Acetyl-CoA carboxylase 1	1.7
ACO2 Aconitate hydratase, mitochondrial	-1.1
ACTL6A Actin-like protein 6A	1.6
ACTN3 Alpha-actinin-3	-1.1
ACTR1B Beta-centractin	1.1
AHSG Alpha-2-HS-glycoprotein	1
AIFM1 Apoptosis-inducing factor 1, mitochondrial	1.1
AK1 Adenylate kinase isoenzyme 1	1.1
ANP32E Acidic leucine-rich nuclear phosphoprotein 32 family member E	1
AP2A1 AP-2 complex subunit alpha-1	1.1
ARF1 ADP-ribosylation factor 1	1
ARHGEF1 Rho guanine nucleotide exchange factor 1	1.3
ATP5O ATP synthase subunit O, mitochondrial	1.3
BCAT1 Branched-chain-amino-acid aminotransferase, cytosolic	1
BCAT2 Branched-chain-amino-acid aminotransferase, mitochondrial	1
CACYBP Calcyclin-binding protein	1.1
CFL2 Cofilin-2	1.4
COPB1 Coatomer subunit beta	1
COPG1 Coatomer subunit gamma-1	1.3
COPS2 COP9 signalosome complex subunit 2	2.2
COPS3 COP9 signalosome complex subunit 3	1.2
CRKL Crk-like protein	1.4
CTSD Cathepsin D	1
CXADR Coxsackievirus and adenovirus receptor	1.3
DDX17 Probable ATP-dependent RNA helicase DDX17	1.1

Protein	Distance
DHFR Dihydrofolate reductase	1
DNAJA2 DnaJ homolog subfamily A member 2	1.3
DNAJB11 DnaJ homolog subfamily B member 11	1.1
EIF3D Eukaryotic translation initiation factor 3 subunit D	1.4
EIF3K Eukaryotic translation initiation factor 3 subunit K	1.2
EIF4B Eukaryotic translation initiation factor 4B	1.1
EIF5B Eukaryotic translation initiation factor 5B	1
EPPK1 Epiplakin	1.1
ERLIN2 Erlin-2	1.2
ERP29 Endoplasmic reticulum resident protein 29	1.1
FKBP10 Peptidyl-prolyl cis-trans isomerase FKBP10	2
FKBP14 Peptidyl-prolyl cis-trans isomerase FKBP14	1.1
FTO Alpha-ketoglutarate-dependent dioxygenase FTO	1
GAMT Guanidinoacetate N-methyltransferase	1.1
GBAS Protein NipSnap homolog 2	1.3
GBE1 1,4-alpha-glucan-branching enzyme	1.4
GLRX3 Glutaredoxin-3	1.3
GPD2 Glycerol-3-phosphate dehydrogenase, mitochondrial	1.6
GNA14 Guanine nucleotide-binding protein subunit alpha-14	1.8
GNS N-acetylglucosamine-6-sulfatase	1.1
GPD2 Glycerol-3-phosphate dehydrogenase, mitochondrial	1.2
HADH Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	1.1
HEBP1 Heme-binding protein 1	1
HIBADH 3-hydroxyisobutyrate dehydrogenase, mitochondrial	1.1
HLA-A HLA class I histocompatibility antigen, A-2 alpha chain	-1
HMGB1 High mobility group protein B1	1
IDH1 Isocitrate dehydrogenase [NADP] cytoplasmic	1
IDH3A Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	1
IDH3B Isocitrate dehydrogenase [NAD] subunit beta, mitochondrial	1
IGF2BP1 Insulin-like growth factor 2 mRNA-binding protein 1	1
IGF2BP3 Insulin-like growth factor 2 mRNA-binding protein 3	1.1
KDM3B Lysine-specific demethylase 3B	1
KDM6B Lysine-specific demethylase 6B	1
KRT14 Keratin, type I cytoskeletal 14	-2
KRT5 Keratin, type II cytoskeletal 5	1.2
LAMB1 Laminin subunit beta-1	1.2
LEPRE1 Prolyl 3-hydroxylase 1	2.1
LRPPRC Leucine-rich PPR motif-containing protein, mitochondrial	1
MAT2A S-adenosylmethionine synthase isoform type-2	-1.1
MCM4 DNA replication licensing factor MCM4	1.2

Protein	Distance
MCM5 DNA replication licensing factor MCM5	1.6
MCM7 DNA replication licensing factor MCM7	1
MLLT4 Afadin	1.4
MYO6 Unconventional myosin-VI	-1.1
NASP Nuclear autoantigenic sperm protein	1.2
NCAPD2 Condensin complex subunit 1	1.2
NRD1 Nardilysin	2.5
NUDT5 ADP-sugar pyrophosphatase	1.2
OGDH 2-oxoglutarate dehydrogenase, mitochondrial	1
OGFOD1 2-oxoglutarate and iron-dependent oxygenase domain-containing protein 1	1
OGT UDP-N-acetylglucosamine--pe	1
OSGEP Probable tRNA threonylcarba	1.4
OTUB1 Ubiquitin thioesterase OTU	1.2
P4HA1 Prolyl 4-hydroxylase subunit alpha-1	1
P4HA2 Prolyl 4-hydroxylase subunit alpha-2	1
PABPC4 Polyadenylate-binding protein 4	-1.2
PEPD Xaa-Pro dipeptidase	-1.1
PGP Phosphoglycolate phosphatase	1.2
PHF2 Lysine-specific demethylase PHF2	1
PHGDH D-3-phosphoglycerate dehydrogenase	1
PLEC Plectin	1
PLOD1 Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	1
PLOD2 Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	1.1
PLOD3 Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	1
PPIB Peptidyl-prolyl cis-trans isomerase B	1.6
PPP2R4 Serine/threonine-protein phosphatase 2A activator	1.4
PPP2R5C Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform	-1.1
PPP4R1 Serine/threonine-protein phosphatase 4 regulatory subunit 1	1.4
PPP5C Serine/threonine-protein phosphatase 5	1
PRDX6 Peroxiredoxin-6	1.1
PRPS1 Ribose-phosphate pyrophosphokinase 1	1.9
PRPS2 Ribose-phosphate pyrophosphokinase 2	-1
PSAT1 Phosphoserine aminotransferase	1
PSMA1 Proteasome subunit alpha type-1	1.5
PSMA5 Proteasome subunit alpha type-5	1.6
PSMD11 26S proteasome non-ATPase regulatory subunit 11	1.1
PTPLAD1 3-hydroxyacyl-CoA dehydratase 3	1.7
PTPN11 Tyrosine-protein phosphatase non-receptor type 11	1
RAC1 Ras-related C3 botulinum toxin substrate 1	1.2

Protein	Distance
RANBP1 Ran-specific GTPase-activating protein	1.1
RBBP4 Histone-binding protein RBBP4	1.3
RCN2 Reticulocalbin-2	1
RPA2 Replication protein A 32 kDa subunit	1.2
RPL12 60S ribosomal protein L12	1.5
RPL21 60S ribosomal protein L21	1.5
RPL27A 60S ribosomal protein L27a	1.3
RPL5 60S ribosomal protein L5	1.1
RPL7A 60S ribosomal protein L7a	1.6
RPLP0 60S acidic ribosomal protein P0	1
RPLP2 60S acidic ribosomal protein P2	1.4
RPS13 40S ribosomal protein S13	1
RPS17L 40S ribosomal protein S17-like	1.4
RPS2 40S ribosomal protein S2	1
RPS5 40S ribosomal protein S5	1
RPS6 40S ribosomal protein S6	-1.1
RRM2B Ribonucleoside-diphosphate reductase subunit M2 B	1.2
RRP12 RRP12-like protein	1.5
SCRIB Protein scribble homolog	1.7
SEC23B Protein transport protein Sec23B	1.3
SEC24A Protein transport protein Sec24A	1
SERPINB6 Serpin B6	1.1
SERPINH1 Serpin H1	1.8
SET Protein SET	1.4
SF3A1 Splicing factor 3A subunit 1	1.4
SKP1 S-phase kinase-associated protein 1	1
SMC2 Structural maintenance of chromosomes protein 2	1
SNRPA1 U2 small nuclear ribonucleoprotein A'	1.5
SOD1 Superoxide dismutase [Cu-Zn]	1
SRM Spermidine synthase	1.2
SRP68 Signal recognition particle 68 kDa protein	1
SRSF3 Serine/arginine-rich splicing factor 3	1
Signal transducer and activator of transcription 1-alpha/beta	1.4
TFRC Transferrin receptor protein 1	-1.7
TMX3 Protein disulfide-isomerase TMX3	1
TNPO1 Transportin-1	1.6
TPM3 Tropomyosin alpha-3 chain	-1
Tubulin--tyrosine ligase-like protein 12	1.1
TUBA8 Tubulin alpha-8 chain	1.7
TUBB6 Tubulin beta-6 chain	1.3

Protein	Distance
U2AF2 Splicing factor U2AF 65 kDa subunit	1
UBA2 SUMO-activating enzyme subunit 2	1.1
UBC Polyubiquitin-C	-1.1
UBE2N Ubiquitin-conjugating enzyme E2 N	1
UBR4 E3 ubiquitin-protein ligase UBR4	1.2
USP15 Ubiquitin carboxyl-terminal hydrolase 15	1.4
XPNPEP1 Xaa-Pro aminopeptidase 1	1.1

Table 3-2: Known targets of α-KG Previously elucidated targets of α-KG

Protein names	Gene names
Kynurenine/alpha-aminoadipate aminotransferase, mitochondrial (KAT/AadAT) (2-aminoadipate aminotransferase) (2-aminoadipate transaminase) (EC 2.6.1.39) (Alpha-aminoadipate aminotransferase) (AadAT) (Kynurenine aminotransferase II) (Kynurenine--oxoglutarate aminotransferase II) (Kynurenine--oxoglutarate transaminase 2) (EC 2.6.1.7) (Kynurenine--oxoglutarate transaminase II)	AADAT KAT2
Alpha-aminoadipic semialdehyde synthase, mitochondrial (LKR/SDH) [Includes: Lysine ketoglutarate reductase (LKR) (LOR) (EC 1.5.1.8); Saccharopine dehydrogenase (SDH) (EC 1.5.1.9)]	AASS
4-aminobutyrate aminotransferase, mitochondrial (EC 2.6.1.19) ((S)-3-amino-2-methylpropionate transaminase) (EC 2.6.1.22) (GABA aminotransferase) (GABA-AT) (Gamma-amino-N-butyrate transaminase) (GABA transaminase) (GABA-T) (L-AIBAT)	ABAT GABAT
Hydroxyacid-oxoacid transhydrogenase, mitochondrial (HOT) (EC 1.1.99.24) (Alcohol dehydrogenase iron-containing protein 1) (ADHFe1) (Fe-containing alcohol dehydrogenase)	ADHFE1 HMFT2263
Alkylated DNA repair protein alkB homolog 1 (EC 1.14.11.33) (Alpha-ketoglutarate-dependent dioxygenase ABH1) (DNA lyase ABH1) (EC 4.2.99.18) (DNA oxidative demethylase ALKBH1)	ALKBH1 ABH ABH1 ALKBH
Alpha-ketoglutarate-dependent dioxygenase alkB homolog 2 (EC 1.14.11.33) (Alkylated DNA repair protein alkB homolog 2) (DNA oxidative demethylase ALKBH2) (Oxy DC1)	ALKBH2 ABH2
Alpha-ketoglutarate-dependent dioxygenase alkB homolog 3 (EC 1.14.11.-) (Alkylated DNA repair protein alkB homolog 3) (DEPC-1) (Prostate cancer antigen 1)	ALKBH3 ABH3 DEPC1
Alpha-ketoglutarate-dependent dioxygenase alkB homolog 4 (EC 1.14.11.-) (Alkylated DNA repair protein alkB homolog 4)	ALKBH4 ABH4
RNA demethylase ALKBH5 (EC 1.14.11.-) (Alkylated DNA repair protein alkB homolog 5) (Alpha-	ALKBH5 ABH5 OFOXD1

Protein names	Gene names
ketoglutarate-dependent dioxygenase alkB homolog 5)	
Alpha-ketoglutarate-dependent dioxygenase alkB homolog 6 (EC 1.14.11.-) (Alkylated DNA repair protein alkB homolog 6)	ALKBH6 ABH6
Alpha-ketoglutarate-dependent dioxygenase alkB homolog 7, mitochondrial (EC 1.14.11.-) (Alkylated DNA repair protein alkB homolog 7) (Spermatogenesis cell proliferation-related protein) (Spermatogenesis-associated protein 11)	ALKBH7 ABH7 SPATA11 UNQ6002/PRO3456 4
Alkylated DNA repair protein alkB homolog 8 (EC 1.14.11.-) (Probable alpha-ketoglutarate-dependent dioxygenase ABH8) (S-adenosyl-L-methionine-dependent tRNA methyltransferase ABH8) (tRNA (carboxymethyluridine(34)-5-O)-methyltransferase ABH8) (EC 2.1.1.229)	ALKBH8 ABH8
Aspartyl/asparaginyl beta-hydroxylase (EC 1.14.11.16) (Aspartate beta-hydroxylase) (ASP beta-hydroxylase) (Peptide-aspartate beta-dioxygenase)	ASPH BAH
Aspartate beta-hydroxylase domain-containing protein 1 (EC 1.14.11.-)	ASPHD1
Aspartate beta-hydroxylase domain-containing protein 2 (EC 1.14.11.-)	ASPHD2
Gamma-butyrobetaine dioxygenase (EC 1.14.11.1) (Gamma-butyrobetaine hydroxylase) (Gamma-BBH) (Gamma-butyrobetaine,2-oxoglutarate dioxygenase)	BBOX1 BBH BBOX
Branched-chain-amino-acid aminotransferase, cytosolic (BCAT(c)) (EC 2.6.1.42) (Protein ECA39)	BCAT1 BCT1 ECA39
Branched-chain-amino-acid aminotransferase, mitochondrial (BCAT(m)) (EC 2.6.1.42) (Placental protein 18) (PP18)	BCAT2 BCATM BCT2 ECA40
Complement C1r subcomponent (EC 3.4.21.41) (Complement component 1 subcomponent r) [Cleaved into: Complement C1r subcomponent heavy chain; Complement C1r subcomponent light chain]	C1R
Complement C1s subcomponent (EC 3.4.21.42) (C1 esterase) (Complement component 1 subcomponent s) [Cleaved into: Complement C1s subcomponent heavy chain; Complement C1s subcomponent light chain]	C1S

Protein names	Gene names
Kynurenine--oxoglutarate transaminase 1 (EC 2.6.1.7) (Cysteine-S-conjugate beta-lyase) (EC 4.4.1.13) (Glutamine transaminase K) (GTK) (Glutamine--phenylpyruvate transaminase) (EC 2.6.1.64) (Kynurenine aminotransferase I) (KATI) (Kynurenine--oxoglutarate transaminase I)	CCBL1
Kynurenine--oxoglutarate transaminase 3 (EC 2.6.1.7) (Cysteine-S-conjugate beta-lyase 2) (EC 4.4.1.13) (Kynurenine aminotransferase III) (KATIII) (Kynurenine--glyoxylate transaminase) (EC 2.6.1.63) (Kynurenine--oxoglutarate transaminase III)	CCBL2 KAT3
Cadherin EGF LAG seven-pass G-type receptor 1 (Cadherin family member 9) (Flamingo homolog 2) (hFmi2)	CELSR1 CDHF9 FMI2
Cadherin EGF LAG seven-pass G-type receptor 2 (Cadherin family member 10) (Epidermal growth factor-like protein 2) (EGF-like protein 2) (Flamingo homolog 3) (Multiple epidermal growth factor-like domains protein 3) (Multiple EGF-like domains protein 3)	CELSR2 CDHF10 EGFL2 KIAA0279 MEGF3
D-2-hydroxyglutarate dehydrogenase, mitochondrial (EC 1.1.99.-)	D2HGDH D2HGD
Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	DBT
Probable 2-oxoglutarate dehydrogenase E1 component DHKTD1, mitochondrial (EC 1.2.4.2) (Dehydrogenase E1 and transketolase domain-containing protein 1)	DHTKD1 KIAA1630
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial (EC 2.3.1.61) (2-oxoglutarate dehydrogenase complex component E2) (OGDC-E2) (Dihydrolipoamide succinyltransferase component of 2-oxoglutarate dehydrogenase complex) (E2K)	DLST DLTS
Egl nine homolog 1 (EC 1.14.11.29) (Hypoxia-inducible factor prolyl hydroxylase 2) (HIF-PH2) (HIF-prolyl hydroxylase 2) (HPH-2) (Prolyl hydroxylase domain-containing protein 2) (PHD2) (SM-20)	EGLN1 C1orf12 PNAS-118 PNAS-137
Egl nine homolog 2 (EC 1.14.11.29) (Estrogen-induced tag 6) (HPH-3) (Hypoxia-inducible factor prolyl hydroxylase 1) (HIF-PH1) (HIF-prolyl	EGLN2 EIT6

Protein names	Gene names
hydroxylase 1) (HPH-1) (Prolyl hydroxylase domain-containing protein 1) (PHD1)	
Egl nine homolog 3 (EC 1.14.11.29) (HPH-1) (Hypoxia-inducible factor prolyl hydroxylase 3) (HIF-PH3) (HIF-prolyl hydroxylase 3) (HPH-3) (Prolyl hydroxylase domain-containing protein 3) (PHD3)	EGLN3
Endothelial PAS domain-containing protein 1 (EPAS-1) (Basic-helix-loop-helix-PAS protein MOP2) (Class E basic helix-loop-helix protein 73) (bHLHe73) (HIF-1-alpha-like factor) (HLF) (Hypoxia-inducible factor 2-alpha) (HIF-2-alpha) (HIF2-alpha) (Member of PAS protein 2) (PAS domain-containing protein 2)	EPAS1 BHLHE73 HIF2A MOP2 PASD2
Coagulation factor X (EC 3.4.21.6) (Stuart factor) (Stuart-Prower factor) [Cleaved into: Factor X light chain; Factor X heavy chain; Activated factor Xa heavy chain]	F10
Coagulation factor VII (EC 3.4.21.21) (Proconvertin) (Serum prothrombin conversion accelerator) (SPCA) (Eptacog alfa) [Cleaved into: Factor VII light chain; Factor VII heavy chain]	F7
Coagulation factor IX (EC 3.4.21.22) (Christmas factor) (Plasma thromboplastin component) (PTC) [Cleaved into: Coagulation factor IXa light chain; Coagulation factor IXa heavy chain]	F9
Alpha-ketoglutarate-dependent dioxygenase FTO (EC 1.14.11.-) (Fat mass and obesity-associated protein)	FTO KIAA1752
Growth hormone receptor (GH receptor) (Somatotropin receptor) [Cleaved into: Growth hormone-binding protein (GH-binding protein) (GHBP) (Serum-binding protein)]	GHR
Glutamate dehydrogenase 1, mitochondrial (GDH 1) (EC 1.4.1.3)	GLUD1 GLUD
Glutamate dehydrogenase 2, mitochondrial (GDH 2) (EC 1.4.1.3)	GLUD2 GLUDP1
Aspartate aminotransferase, cytoplasmic (cAspAT) (EC 2.6.1.1) (EC 2.6.1.3) (Cysteine aminotransferase, cytoplasmic) (Cysteine transaminase, cytoplasmic) (cCAT) (Glutamate oxaloacetate transaminase 1) (Transaminase A)	GOT1
Putative aspartate aminotransferase, cytoplasmic 2 (EC 2.6.1.-) (Glutamate oxaloacetate	GOT1L1

Protein names	Gene names
transaminase 1-like protein 1) (Transaminase A-like protein 1)	
Aspartate aminotransferase, mitochondrial (mAspAT) (EC 2.6.1.1) (EC 2.6.1.7) (Fatty acid-binding protein) (FABP-1) (Glutamate oxaloacetate transaminase 2) (Kynurenine aminotransferase 4) (Kynurenine aminotransferase IV) (Kynurenine--oxoglutarate transaminase 4) (Kynurenine--oxoglutarate transaminase IV) (Plasma membrane-associated fatty acid-binding protein) (FABPpm) (Transaminase A)	GOT2
Alanine aminotransferase 1 (ALT1) (EC 2.6.1.2) (Glutamate pyruvate transaminase 1) (GPT 1) (Glutamic--alanine transaminase 1) (Glutamic--pyruvic transaminase 1)	GPT AAT1 GPT1
Alanine aminotransferase 2 (ALT2) (EC 2.6.1.2) (Glutamate pyruvate transaminase 2) (GPT 2) (Glutamic--alanine transaminase 2) (Glutamic--pyruvic transaminase 2)	GPT2 AAT2 ALT2
Hypoxia-inducible factor 1-alpha (HIF-1-alpha) (HIF1-alpha) (ARNT-interacting protein) (Basic-helix-loop-helix-PAS protein MOP1) (Class E basic helix-loop-helix protein 78) (bHLHe78) (Member of PAS protein 1) (PAS domain-containing protein 8)	HIF1A BHLHE78 MOP1 PASD8
Hypoxia-inducible factor 1-alpha inhibitor (EC 1.14.11.30) (EC 1.14.11.n4) (Factor inhibiting HIF-1) (FIH-1) (Hypoxia-inducible factor asparagine hydroxylase)	HIF1AN FIH1
4-hydroxy-2-oxoglutarate aldolase, mitochondrial (EC 4.1.3.16) (Dihydrodipicolinate synthase-like) (DHDPS-like protein) (Probable 2-keto-4-hydroxyglutarate aldolase) (Probable KHG-aldolase) (Protein 569272)	HOGA1 C10orf65 DHDPSL
Lysine-specific demethylase hairless (EC 1.14.11.-)	HR
Isocitrate dehydrogenase [NADP] cytoplasmic (IDH) (EC 1.1.1.42) (Cytosolic NADP-isocitrate dehydrogenase) (IDP) (NADP(+)-specific ICDH) (Oxalosuccinate decarboxylase)	IDH1 PICD
Isocitrate dehydrogenase [NADP], mitochondrial (IDH) (EC 1.1.1.42) (ICD-M) (IDP) (NADP(+)-specific ICDH) (Oxalosuccinate decarboxylase)	IDH2
Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial (EC 1.1.1.41) (Isocitric	IDH3A

Protein names	Gene names
dehydrogenase subunit alpha) (NAD(+)-specific ICDH subunit alpha)	
Isocitrate dehydrogenase [NAD] subunit beta, mitochondrial (EC 1.1.1.41) (Isocitric dehydrogenase subunit beta) (NAD(+)-specific ICDH subunit beta)	IDH3B
Isocitrate dehydrogenase [NAD] subunit gamma, mitochondrial (EC 1.1.1.41) (Isocitric dehydrogenase subunit gamma) (NAD(+)-specific ICDH subunit gamma)	IDH3G
Probable JmjC domain-containing histone demethylation protein 2C (EC 1.14.11.-) (Jumonji domain-containing protein 1C) (Thyroid receptor-interacting protein 8) (TR-interacting protein 8) (TRIP-8)	JMJD1C JHDM2C KIAA1380 TRIP8
Bifunctional arginine demethylase and lysyl-hydroxylase JMJD6 (EC 1.14.11.-) (Histone arginine demethylase JMJD6) (JmjC domain-containing protein 6) (Jumonji domain-containing protein 6) (Lysyl-hydroxylase JMJD6) (Peptide-lysine 5-dioxygenase JMJD6) (Phosphatidylserine receptor) (Protein PTDSR)	JMJD6 KIAA0585 PTDSR
Lysine-specific demethylase 2A (EC 1.14.11.27) (CXXC-type zinc finger protein 8) (F-box and leucine-rich repeat protein 11) (F-box protein FBL7) (F-box protein Lilina) (F-box/LRR-repeat protein 11) (JmjC domain-containing histone demethylation protein 1A) ([Histone-H3]-lysine-36 demethylase 1A)	KDM2A CXXC8 FBL7 FBXL11 JHDM1A KIAA1004
Lysine-specific demethylase 2B (EC 1.14.11.27) (CXXC-type zinc finger protein 2) (F-box and leucine-rich repeat protein 10) (F-box protein FBL10) (F-box/LRR-repeat protein 10) (JmjC domain-containing histone demethylation protein 1B) (Jumonji domain-containing EMSY-interactor methyltransferase motif protein) (Protein JEMMA) (Protein-containing CXXC domain 2) ([Histone-H3]-lysine-36 demethylase 1B)	KDM2B CXXC2 FBL10 FBXL10 JHDM1B PCCX2
Lysine-specific demethylase 3A (EC 1.14.11.-) (JmjC domain-containing histone demethylation protein 2A) (Jumonji domain-containing protein 1A)	KDM3A JHDM2A JMJD1 JMJD1A KIAA0742 TSGA
Lysine-specific demethylase 3B (EC 1.14.11.-) (JmjC domain-containing histone demethylation protein	KDM3B C5orf7 JHDM2B JMJD1B KIAA1082

Protein names	Gene names
2B) (Jumonji domain-containing protein 1B) (Nuclear protein 5qNCA)	
Lysine-specific demethylase 4A (EC 1.14.11.-) (JmjC domain-containing histone demethylation protein 3A) (Jumonji domain-containing protein 2A)	KDM4A JHDM3A JMJD2 JMJD2A KIAA0677
Lysine-specific demethylase 4B (EC 1.14.11.-) (JmjC domain-containing histone demethylation protein 3B) (Jumonji domain-containing protein 2B)	KDM4B JHDM3B JMJD2B KIAA0876
Lysine-specific demethylase 4C (EC 1.14.11.-) (Gene amplified in squamous cell carcinoma 1 protein) (GASC-1 protein) (JmjC domain-containing histone demethylation protein 3C) (Jumonji domain-containing protein 2C)	KDM4C GASC1 JHDM3C JMJD2C KIAA0780
Lysine-specific demethylase 4D (EC 1.14.11.-) (JmjC domain-containing histone demethylation protein 3D) (Jumonji domain-containing protein 2D)	KDM4D JHDM3D JMJD2D
Lysine-specific demethylase 4E (EC 1.14.11.-) (KDM4D-like protein) (Lysine-specific demethylase 4D-like)	KDM4E KDM4DL
Lysine-specific demethylase 5A (EC 1.14.11.-) (Histone demethylase JARID1A) (Jumonji/ARID domain-containing protein 1A) (Retinoblastoma-binding protein 2) (RBBP-2)	KDM5A JARID1A RBBP2 RBP2
Lysine-specific demethylase 5B (EC 1.14.11.-) (Cancer/testis antigen 31) (CT31) (Histone demethylase JARID1B) (Jumonji/ARID domain-containing protein 1B) (PLU-1) (Retinoblastoma-binding protein 2 homolog 1) (RBP2-H1)	KDM5B JARID1B PLU1 RBBP2H1
Lysine-specific demethylase 5C (EC 1.14.11.-) (Histone demethylase JARID1C) (Jumonji/ARID domain-containing protein 1C) (Protein SmcX) (Protein Xe169)	KDM5C DXS1272E JARID1C SMCX XE169
Lysine-specific demethylase 5D (EC 1.14.11.-) (Histocompatibility Y antigen) (H-Y) (Histone demethylase JARID1D) (Jumonji/ARID domain-containing protein 1D) (Protein SmcY)	KDM5D HY HYA JARID1D KIAA0234 SMCY
Lysine-specific demethylase 6A (EC 1.14.11.-) (Histone demethylase UTX) (Ubiquitously-transcribed TPR protein on the X chromosome) (Ubiquitously-transcribed X chromosome tetratricopeptide repeat protein)	KDM6A UTX
Lysine-specific demethylase 6B (EC 1.14.11.-) (JmjC domain-containing protein 3) (Jumonji domain-containing protein 3) (Lysine demethylase 6B)	KDM6B JMJD3 KIAA0346

Protein names	Gene names
Lysine-specific demethylase 7A (EC 1.14.11.-) (JmjC domain-containing histone demethylation protein 1D) (Lysine-specific demethylase 7)	KDM7A JHDM1D KDM7 KIAA1718
Lysine-specific demethylase 8 (EC 1.14.11.27) (JmjC domain-containing protein 5) (Jumonji domain-containing protein 5)	KDM8 JMJD5
L-2-hydroxyglutarate dehydrogenase, mitochondrial (EC 1.1.99.2) (Duranin)	L2HGDH C14orf160
Prolyl 3-hydroxylase 1 (EC 1.14.11.7) (Growth suppressor 1) (Leucine- and proline-enriched proteoglycan 1) (Leprecan-1)	LEPRE1 GROS1 P3H1 PSEC0109
Prolyl 3-hydroxylase 2 (EC 1.14.11.7) (Leprecan-like protein 1) (Myxoid liposarcoma-associated protein 4)	LEPREL1 MLAT4 P3H2
Prolyl 3-hydroxylase 3 (EC 1.14.11.7) (Leprecan-like protein 2) (Protein B)	LEPREL2 P3H3
Latent-transforming growth factor beta-binding protein 1 (LTBP-1) (Transforming growth factor beta-1-binding protein 1) (TGF-beta1-BP-1)	LTBP1
Mannan-binding lectin serine protease 1 (EC 3.4.21.-) (Complement factor MASP-3) (Complement-activating component of Ra-reactive factor) (Mannose-binding lectin-associated serine protease 1) (MASP-1) (Mannose-binding protein-associated serine protease) (Ra-reactive factor serine protease p100) (RaRF) (Serine protease 5) [Cleaved into: Mannan-binding lectin serine protease 1 heavy chain; Mannan-binding lectin serine protease 1 light chain]	MASP1 CRARF CRARF1 PRSS5
Mannan-binding lectin serine protease 2 (EC 3.4.21.104) (MBL-associated serine protease 2) (Mannose-binding protein-associated serine protease 2) (MASP-2) [Cleaved into: Mannan-binding lectin serine protease 2 A chain; Mannan-binding lectin serine protease 2 B chain]	MASP2
Bifunctional lysine-specific demethylase and histidyl-hydroxylase MINA (EC 1.14.11.-) (60S ribosomal protein L27a histidine hydroxylase) (Histone lysine demethylase MINA) (MYC-induced nuclear antigen) (Mineral dust-induced gene protein) (Nucleolar protein 52) (Ribosomal oxygenase MINA) (ROX)	MINA MDIG MINA53 NO52
Nuclear factor NF-kappa-B p105 subunit (DNA-binding factor KBF1) (EBP-1) (Nuclear factor of kappa	NFKB1

Protein names	Gene names
light polypeptide gene enhancer in B-cells 1) [Cleaved into: Nuclear factor NF-kappa-B p50 subunit]	
Bifunctional lysine-specific demethylase and histidyl-hydroxylase NO66 (EC 1.14.11.-) (EC 1.14.11.27) (60S ribosomal protein L8 histidine hydroxylase) (Histone lysine demethylase NO66) (Myc-associated protein with JmjC domain) (Nucleolar protein 66) (hsNO66) (Ribosomal oxygenase NO66) (ROX)	NO66 C14orf169 MAPJD
2-oxoglutarate dehydrogenase, mitochondrial (EC 1.2.4.2) (2-oxoglutarate dehydrogenase complex component E1) (OGDC-E1) (Alpha-ketoglutarate dehydrogenase)	OGDH
2-oxoglutarate dehydrogenase-like, mitochondrial (EC 1.2.4.-) (2-oxoglutarate dehydrogenase complex component E1-like) (OGDC-E1-like) (Alpha-ketoglutarate dehydrogenase-like)	OGDHL KIAA1290
Prolyl 3-hydroxylase OGFOD1 (EC 1.14.11.-) (2-oxoglutarate and iron-dependent oxygenase domain-containing protein 1) (Termination and polyadenylation 1 homolog)	OGFOD1 KIAA1612 TPA1
2-oxoglutarate and iron-dependent oxygenase domain-containing protein 2 (EC 1.14.11.-)	OGFOD2
2-oxoglutarate and iron-dependent oxygenase domain-containing protein 3 (EC 1.14.11.-)	OGFOD3 C17orf101
2-oxoglutarate receptor 1 (Alpha-ketoglutarate receptor 1) (G-protein coupled receptor 80) (G-protein coupled receptor 99) (P2Y purinoceptor 15) (P2Y15) (P2Y-like GPCR) (P2Y-like nucleotide receptor)	OXGR1 GPR80 GPR99 P2RY15 P2Y15
Prolyl 4-hydroxylase subunit alpha-1 (4-PH alpha-1) (EC 1.14.11.2) (Procollagen-proline,2-oxoglutarate-4-dioxygenase subunit alpha-1)	P4HA1 P4HA
Prolyl 4-hydroxylase subunit alpha-2 (4-PH alpha-2) (EC 1.14.11.2) (Procollagen-proline,2-oxoglutarate-4-dioxygenase subunit alpha-2)	P4HA2 UNQ290/PRO330
Prolyl 4-hydroxylase subunit alpha-3 (4-PH alpha-3) (EC 1.14.11.2) (Procollagen-proline,2-oxoglutarate-4-dioxygenase subunit alpha-3)	P4HA3 UNQ711/PRO1374
Protein disulfide-isomerase	P4HB
Transmembrane prolyl 4-hydroxylase (P4H-TM) (EC 1.14.11.-) (Hypoxia-inducible factor prolyl hydroxylase 4) (HIF-PH4) (HIF-prolyl hydroxylase 4) (HPH-4)	P4HTM PH4

Protein names	Gene names
Lysine-specific demethylase PHF2 (EC 1.14.11.-) (GRC5) (PHD finger protein 2)	PHF2 KIAA0662
Histone lysine demethylase PHF8 (EC 1.14.11.27) (PHD finger protein 8)	PHF8 KIAA1111 ZNF422
D-3-phosphoglycerate dehydrogenase (3-PGDH) (EC 1.1.1.95)	PHGDH PGDH3
Phytanoyl-CoA dioxygenase, peroxisomal (EC 1.14.11.18) (Phytanic acid oxidase) (Phytanoyl-CoA alpha-hydroxylase) (PhyH)	PHYH PAHX
Phytanoyl-CoA dioxygenase domain-containing protein 1	PHYHD1
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 (EC 1.14.11.4) (Lysyl hydroxylase 1) (LH1)	PLOD1 LLH PLOD
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 (EC 1.14.11.4) (Lysyl hydroxylase 2) (LH2)	PLOD2
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3 (EC 1.14.11.4) (Lysyl hydroxylase 3) (LH3)	PLOD3
Vitamin K-dependent protein C (EC 3.4.21.69) (Anticoagulant protein C) (Autoprothrombin IIA) (Blood coagulation factor XIV) [Cleaved into: Vitamin K-dependent protein C light chain; Vitamin K-dependent protein C heavy chain; Activation peptide]	PROC
Vitamin K-dependent protein S	PROS1 PROS
Vitamin K-dependent protein Z	PROZ
Phosphoserine aminotransferase (EC 2.6.1.52) (Phosphohydroxythreonine aminotransferase) (PSAT)	PSAT1 PSA
Mitochondrial 2-oxoglutarate/malate carrier protein (OGCP) (Solute carrier family 25 member 11)	SLC25A11 SLC20A4
Mitochondrial 2-oxodicarboxylate carrier (Solute carrier family 25 member 21)	SLC25A21 ODC
Solute carrier family 25 member 39	SLC25A39 CGI-69 PRO2163
Signal transducer and activator of transcription 5A	STAT5A STAT5
Signal transducer and activator of transcription 5B	STAT5B
Tyrosine aminotransferase (TAT) (EC 2.6.1.5) (L-tyrosine:2-oxoglutarate aminotransferase)	TAT
Methylcytosine dioxygenase TET1 (EC 1.14.11.n2) (CXXC-type zinc finger protein 6) (Leukemia-associated protein with a CXXC domain) (Ten-eleven translocation 1 gene protein)	TET1 CXXC6 KIAA1676 LCX
Methylcytosine dioxygenase TET2 (EC 1.14.11.n2)	TET2 KIAA1546 Nbla00191
Methylcytosine dioxygenase TET3 (EC 1.14.11.n2)	TET3 KIAA0401

Protein names	Gene names
Thrombomodulin (TM) (Fetomodulin) (CD antigen CD141)	THBD THRM
Trimethyllysine dioxygenase, mitochondrial (EC 1.14.11.8) (Epsilon-trimethyllysine 2-oxoglutarate dioxygenase) (Epsilon-trimethyllysine hydroxylase) (TML hydroxylase) (TML-alpha-ketoglutarate dioxygenase) (TML dioxygenase) (TMLD)	TMLHE TMLH
Tankyrase-2 (TANK2) (EC 2.4.2.30) (ADP-ribosyltransferase diphtheria toxin-like 6) (ARTD6) (Poly [ADP-ribose] polymerase 5B) (TNKS-2) (TRF1-interacting ankyrin-related ADP-ribose polymerase 2) (Tankyrase II) (Tankyrase-like protein) (Tankyrase-related protein)	TNKS2 PARP5B TANK2 TNKL
tRNA wybutosine-synthesizing protein 5 (hTYW5) (EC 1.14.11.42) (tRNA(Phe) (7-(3-amino-3-carboxypropyl)wyosine(37)-C(2))-hydroxylase)	TYW5 C2orf60
Histone demethylase UTY (EC 1.14.11.-) (Ubiquitously-transcribed TPR protein on the Y chromosome) (Ubiquitously-transcribed Y chromosome tetratricopeptide repeat protein)	UTY

Table 3-3: Known targets of α-KG detected All detected proteins from our DARTS-MS analysis that are thought to directly interact with α-KG. All but three have a noticeable difference between the control sample and α-KG containing sample.

Protein	Protection
Branched-chain-amino-acid aminotransferase, cytosolic	Slightly protected
Branched-chain-amino-acid aminotransferase, mitochondrial	Slightly protected
Alpha-ketoglutarate-dependent dioxygenase FTO	Protected
Isocitrate dehydrogenase [NADP] cytoplasmic	Protected
Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	Protected
Isocitrate dehydrogenase [NAD] subunit beta, mitochondrial	Protected
Lysine-specific demethylase 3B (EC 1.14.11.-)	Migration change
Lysine-specific demethylase 6B (EC 1.14.11.-)	Migration change
Prolyl 3-hydroxylase 1 (EC 1.14.11.7)	Protected
2-oxoglutarate dehydrogenase, mitochondrial (EC 1.2.4.2)	Protected
Prolyl 3-hydroxylase OGFOD1 (EC 1.14.11.-)	Protected
Prolyl 4-hydroxylase subunit alpha-1 (4-PH alpha-1)	Slightly protected
Prolyl 4-hydroxylase subunit alpha-2 (4-PH alpha-2)	Migration change
Protein disulfide-isomerase	Slightly changed
Lysine-specific demethylase PHF2	Slightly protected
D-3-phosphoglycerate dehydrogenase (3-PGDH)	Protected
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	Migration change
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	Sensitized
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	Slightly protected
Phosphoserine aminotransferase	Protected
Glutamate dehydrogenase 1, mitochondrial	Unchanged
Aspartate aminotransferase, cytoplasmic	Unchanged
Aspartate aminotransferase, mitochondrial	Unchanged

Table 3-4: Putative targets of α-KB based on shifts in migration. Each protein in the following table had a statistically significant difference between the α-KB samples and the control samples. The absolute value of the distance corresponds to the difference in average MW fraction between the experimental and control samples. Negative distances indicates a sensitization to proteolysis in the presence of the small molecule while a positive value corresponds to a protection from proteolysis.

Protein	Distance
ACTN3 Alpha-actinin-3	1.6
AHSG Alpha-2-HS-glycoprotein	1.4
ALKBH7 Probable alpha-ketoglutarate-dependent dioxygenase ABH7	1
ARHGAP1 Rho GTPase-activating protein 1	1.1
BCAT1 Branched-chain-amino-acid aminotransferase, cytosolic	1
CKAP4 Cytoskeleton-associated protein 4	1.2
CUL4B Cullin-4B	1
DNAJC7 DnaJ homolog subfamily C member 7	-1.1
FKBP10 Peptidyl-prolyl cis-trans isomerase FKBP10	1.8
GNA14 Guanine nucleotide-binding protein subunit alpha-14	1.6
GNAS Guanine nucleotide-binding protein G(s) subunit alpha isoforms short	1
HEPHL1 Hephaestin-like protein 1	-1.2
IDH3B Isocitrate dehydrogenase [NAD] subunit beta, mitochondrial	1
IGF2R Cation-independent mannose-6-phosphate receptor	1.2
ILVBL Acetolactate synthase-like protein	1.5
KDM3B Lysine-specific demethylase 3B	1
Keratin, type I cytoskeletal 14	-1
LEPRE1 Prolyl 3-hydroxylase 1	1
NAA25 N-alpha-acetyltransferase 25, NatB auxiliary subunit	1.1
NAP1L1 Nucleosome assembly protein 1-like 1	1.3
NME1 Nucleoside diphosphate kinase A	-1
P4HA1 Prolyl 4-hydroxylase subunit alpha-1	1
PHGDH D-3-phosphoglycerate dehydrogenase	1
PLOD1 Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	1
PLOD2 Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	1
POTEE POTE ankyrin domain family member E	1
PPM1G Protein phosphatase 1G	-1.4

Protein	Distance
PPP2CA Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	-1.4
PSAT1 Phosphoserine aminotransferase	1
RARS Arginine--tRNA ligase, cytoplasmic	-1
RPLP0 60S acidic ribosomal protein P0	1.3
SERPINB6 Serpin B6	1.3
SPR Sepiapterin reductase	1.1
ST13 Hsc70-interacting protein	1.2
TMX3 Protein disulfide-isomerase TMX3	1
TNPO3 Transportin-3	-1
UTRN Utrophin	-1.5

Table 3-5: Known targets of α-KB Previously elucidated targets of α-KB

Protein names	Gene names
2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial (EC 1.2.4.4) (Branched-chain alpha-keto acid dehydrogenase E1 component alpha chain) (BCKDE1A) (BCKDH E1-alpha)	BCKDHA
2-oxoisovalerate dehydrogenase subunit beta, mitochondrial (EC 1.2.4.4) (Branched-chain alpha-keto acid dehydrogenase E1 component beta chain) (BCKDE1B) (BCKDH E1-beta)	BCKDHB
[3-methyl-2-oxobutanoate dehydrogenase [lipoamide]] kinase, mitochondrial (EC 2.7.11.4) (Branched-chain alpha-ketoacid dehydrogenase kinase) (BCKD-kinase) (BCKDHKIN)	BCKDK
Cystathionine gamma-lyase (EC 4.4.1.1) (Cysteine-protein sulfhydrase) (Gamma-cystathionase)	CTH
L-serine dehydratase/L-threonine deaminase (SDH) (EC 4.3.1.17) (L-serine deaminase) (L-threonine dehydratase) (TDH) (EC 4.3.1.19)	SDS SDH
Serine dehydratase-like (L-serine deaminase) (L-serine dehydratase/L-threonine deaminase) (L-threonine dehydratase) (TDH) (EC 4.3.1.19) (Serine dehydratase 2) (SDH 2) (EC 4.3.1.17)	SDSL
Threonine synthase-like 2	THNSL2
Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	DBT

Table 3-6: Protein targets of α-KG and α-KB Putative targets identified in both α-KG and α-KG DARTS-MS experiments.

BCAT1 Branched-chain-amino-acid aminotransferase, cytosolic
IDH3B Isocitrate dehydrogenase [NAD] subunit beta, mitochondrial
KDM3B Lysine-specific demethylase 3B
LEPRE1 Prolyl 3-hydroxylase 1
P4HA1 Prolyl 4-hydroxylase subunit alpha-1
PHGDH D-3-phosphoglycerate
PLOD1 Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1
PLOD2 Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2
PSAT1 Phosphoserine aminotransferase
TMX3 Protein disulfide-isomerase TMX3

Table 3-7: Putative targets of α-KIC based on shifts in migration. Each protein in the following table had a statistically significant difference between the α-KIC samples and the control samples. The absolute value of the distance corresponds to the difference in average MW fraction between the experimental and control samples. Negative distances indicates a sensitization to proteolysis in the presence of the small molecule while a positive value corresponds to a protection from proteolysis.

Protein	Distance
40S ribosomal protein S17-like	1.3
40S ribosomal protein S20	-1.4
60S acidic ribosomal protein P2	-2.1
Activated RNA polymerase II transcriptional coactivator p15	2
Acyl-CoA dehydrogenase family member 11	1.1
Alpha-ketoglutarate-dependent dioxygenase alkB homolog 2	1.9
Antigen KI-67	1.9
Antithrombin-III	1.3
Apoptotic protease-activating factor 1	1.7
Arginine and glutamate-rich protein 1	1.1
Bromodomain and WD repeat-containing protein 3	2
cAMP-dependent protein kinase type II-alpha regulatory subunit	-1.1
Caspase-8	2.6
Cell adhesion molecule 1	1.2
Centromere protein F	2.7
Centrosomal protein of 290 kDa	1.2
Centrosome-associated protein 350	1.1
Chloride intracellular channel protein 2	1.6
Chromodomain-helicase-DNA-binding protein 7	2.2
Chromodomain-helicase-DNA-binding protein 9	1.3
CLIP-associating protein 1	1
Coiled-coil domain-containing protein 141	2.3
Coiled-coil domain-containing protein 150	1.7
Condensin complex subunit 3	1.3
Creatine kinase S-type, mitochondrial	1.9
DBF4-type zinc finger-containing protein 2	1.9
DNA-(apurinic or apyrimidinic site) lyase	1
Dynein heavy chain 10, axonemal	1.2
Dynein heavy chain 12, axonemal	1.1
Dynein heavy chain 14, axonemal	-2.4

Protein	Distance
Dynein heavy chain 17, axonemal	1.6
Dynein heavy chain 2, axonemal	4
Dynein heavy chain 3, axonemal	1.3
Dynein heavy chain 9, axonemal	1.6
E3 SUMO-protein ligase RanBP2	1.8
E3 ubiquitin-protein ligase DZIP3	1.2
E3 ubiquitin-protein ligase HERC2	1.1
E3 ubiquitin-protein ligase TRIP12	1
Exportin-T	-1
FERM and PDZ domain-containing protein 3	1.3
Fructose-bisphosphate aldolase B	1.4
Glial fibrillary acidic protein	1.9
Glutamate--cysteine ligase catalytic subunit	1
Golgin subfamily A member 4	1
GTPase-activating protein and VPS9 domain-containing protein 1	1.1
Guanine nucleotide-binding protein subunit alpha-13	1.1
Helicase ARIP4	1.2
Heterogeneous nuclear ribonucleoprotein U	1
Histone-lysine N-methyltransferase 2A	1.8
Inositol 1,4,5-trisphosphate receptor type 1	1
Kinesin-like protein KIF21B	1
Liprin-alpha-1	2.2
Metal-response element-binding transcription factor 2	5
Multiple PDZ domain protein	-1.8
Myb-binding protein 1A	-1.2
Myotubularin-related protein 5	1.1
NAD-dependent protein deacetylase sirtuin-3, mitochondrial	2.8
Nesprin-2	1.5
Nuclease-sensitive element-binding protein 1	1.3
Obscurin	1.9
Peptidyl-prolyl cis-trans isomerase FKBP3	1.5
Phosphoserine phosphatase	1
Plasma membrane calcium-transporting ATPase 2	1.1
Platelet-activating factor acetylhydrolase IB subunit alpha	1.5
Polyadenylate-binding protein 4-like	1.2
Polycystic kidney disease and receptor for egg jelly-related protein	1.8
Potassium-transporting ATPase alpha chain 1	1.2
Prefoldin subunit 2	-1.3
Probable G-protein coupled receptor 158	3.4

Protein	Distance
Probable tRNA N6-adenosine threonylcarbamoyltransferase, mitochondrial	2
Protein dopey-1	1
Protein strawberry notch homolog 2	-1.3
Putative serine/threonine-protein phosphatase 4 regulatory subunit 1-like	1.1
Putative RNA-binding protein Luc7-like 2	1.2
Putative serine/threonine-protein phosphatase 4 regulatory subunit 1-like	-1.3
Ral GTPase-activating protein subunit alpha-2	1
Ras and Rab interactor 2	2.1
RasGAP-activating-like protein 1	1.1
Ras-related C3 botulinum toxin substrate 3	1.7
Ras-related protein Rab-11B	1.4
Replication protein A 70 kDa DNA-binding subunit	-1.1
Rotatin	1.4
Ryanodine receptor 1	1.5
Sarcoplasmic/endoplasmic reticulum calcium ATPase 3	-2
Scaffold attachment factor B1	3
Serine/threonine-protein phosphatase 5	1.1
Serrate RNA effector molecule homolog	1.1
Sodium channel protein type 11 subunit alpha	1.2
Spermatid perinuclear RNA-binding protein	1.5
Stabilin-1	1.7
Sterile alpha motif domain-containing protein 9	1.1
Sterol regulatory element-binding protein 1	1.3
Succinate-semialdehyde dehydrogenase, mitochondrial	1
TAR DNA-binding protein 43	-1
Threonine aspartase 1	1.5
Tight junction protein ZO-2	1
TRAF2 and NCK-interacting protein kinase	1.1
Transcription initiation factor TFIID subunit 1-like	1.1
Transformation/transcription domain-associated protein	1.9
Tropomyosin alpha-3 chain	1.2
Tropomyosin beta chain	1.2
Tubulin polyglutamylase TTLL7	1.6
Ubiquitin-60S ribosomal protein L40	1.1
Uncharacterized protein KIAA1614	-1.1
Usher syndrome type-1G protein	1.2
UV excision repair protein RAD23 homolog A	3
Volume-regulated anion channel subunit LRRC8C	1.1
von Willebrand factor A domain-containing protein 3B	1.1

Protein	Distance
WASP homolog-associated protein with actin, membranes and microtubules	1.4
Xin actin-binding repeat-containing protein 2	-2.2
Zinc finger protein 605	1.5

3.7 REFERENCES

1. Butler, J.A., Mishur, R.J., Bhaskaran, S. and Rea, S.L., 2013. A metabolic signature for long life in the C aenorhabditis elegans M it mutants. Aging Cell, 12(1), pp.130-138.
2. Lucanic, M., Held, J.M., Vantipalli, M.C., Klang, I.M., Graham, J.B., Gibson, B.W., Lithgow, G.J. and Gill, M.S., 2011. N-acylethanolamine signalling mediates the effect of diet on lifespan in Caenorhabditis elegans. Nature, 473(7346), p.226.
3. O'Rourke, E.J., Kuballa, P., Xavier, R. and Ruvkun, G., 2013. ω-6 Polyunsaturated fatty acids extend life span through the activation of autophagy. Genes & Development.
4. Imai, S.I. and Guarente, L., 2014. NAD+ and sirtuins in aging and disease. Trends in Cell Biology, 24(8), pp.464-471.
5. Patti, G.J., Tautenhahn, R., Johannsen, D., Kalisiak, E., Ravussin, E., Brüning, J.C., Dillin, A. and Siuzdak, G., 2014. Meta-analysis of global metabolomic data identifies metabolites associated with life-span extension. Metabolomics, 10(4), pp.737-743.
6. Kenyon, C.J., 2010. The genetics of ageing. Nature, 464(7288), p.504.
7. Harrison, D.E., Strong, R., Sharp, Z.D., Nelson, J.F., Astle, C.M., Flurkey, K., Nadon, N.L., Wilkinson, J.E., Frenkel, K., Carter, C.S. and Pahor, M., 2009. Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. Nature, 460(7253), p.392.

8. Lomenick, B., Hao, R., Jonai, N., Chin, R.M., Aghajan, M., Warburton, S., Wang, J., Wu, R.P., Gomez, F., Loo, J.A. and Wohlschlegel, J.A., 2009. Target identification using drug affinity responsive target stability (DARTS). Proceedings of the National Academy of Sciences, pp.pnas-0910040106.

9. Abrahams, J.P., Leslie, A.G., Lutter, R. and Walker, J.E., 1994. Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. Nature, 370(6491), p.621.

10. Boyer, P.D., 1997. The ATP synthase—a splendid molecular machine. Annual Review of Biochemistry, 66(1), pp.717-749.

11. Tsang, W.Y., Sayles, L.C., Grad, L.I., Pilgrim, D.B. and Lemire, B.D., 2001. Mitochondrial respiratory chain deficiency in Caenorhabditis elegans results in developmental arrest and increased lifespan. Journal of Biological Chemistry.

12. Dillin, A., Hsu, A.L., Arantes-Oliveira, N., Lehrer-Graiwer, J., Hsin, H., Fraser, A.G., Kamath, R.S., Ahringer, J. and Kenyon, C., 2002. Rates of behavior and aging specified by mitochondrial function during development. Science, 298(5602), pp.2398-2401.

13. Lee, S.S., Lee, R.Y., Fraser, A.G., Kamath, R.S., Ahringer, J. and Ruvkun, G., 2003. A systematic RNAi screen identifies a critical role for mitochondria in C. elegans longevity. Nature Genetics, 33(1), p.40.

14. Curran, S.P. and Ruvkun, G., 2007. Lifespan regulation by evolutionarily conserved genes essential for viability. PLoS Genetics, 3(4), p.e56.

15. Chin, R.M., Fu, X., Pai, M.Y., Vergnes, L., Hwang, H., Deng, G., Diep, S., Lomenick, B., Meli, V.S., Monsalve, G.C., Hu, E., Whelan, S.A., Wang, J.X., Jung, G., Solis,

G.M., Fazlollahi, F., Kaweeteerawat, C., Quach, A., Nili, M., Krall, A.S., Godwin, H.A., Chang, H.R., Faull, K.F., Guo F., Jiang, M.S., Trauger, S.A., Saghatelian, A., Braas, D., Christofk, H.R., Clarke, C.F., Teitell, M.A., Petrascheck, M., Reue, K., Jung, M.E., Frand, A.R., Huang, J., 2014. The metabolite α-ketoglutarate extends lifespan by inhibiting ATP synthase and TOR. Nature, 510(7505), p.397.

16. Brauer, M.J., Yuan, J., Bennett, B.D., Lu, W., Kimball, E., Botstein, D. and Rabinowitz, J.D., 2006. Conservation of the metabolomic response to starvation across two divergent microbes. Proceedings of the National Academy of Sciences, 103(51), pp.19302-19307.
17. Kaminsky, Y.G., Kosenko, E.A. and Kondrashova, M.N., 1982. Metabolites of citric acid cycle, carbohydrate and phosphorus metabolism, and related reactions, redox and phosphorylating states of hepatic tissue, liver mitochondria and cytosol of the pigeon, under normal feeding and natural nocturnal fasting conditions. Comparative biochemistry and physiology. B, Comparative Biochemistry, 73(4), pp.957-963.
18. Pechlivanis, A., Kostidis, S., Saraslanidis, P., Petridou, A., Tsalis, G., Mougios, V., Gika, H.G., Mikros, E. and Theodoridis, G.A., 2010. 1H NMR-based metabonomic investigation of the effect of two different exercise sessions on the metabolic fingerprint of human urine. Journal of Proteome Research, 9(12), pp.6405-6416.
19. Brugnara, L., Vinaixa, M., Murillo, S., Samino, S., Rodriguez, M.A., Beltran, A., Lerin, C., Davison, G., Correig, X. and Novials, A., 2012. Metabolomics approach

for analyzing the effects of exercise in subjects with type 1 diabetes mellitus. PloS One, 7(7), p.e40600.

20. McCay, C.M., Crowell, M.F. and Maynard, L.A., 1935. The effect of retarded growth upon the length of life span and upon the ultimate body size: one figure. The journal of Nutrition, 10(1), pp.63-79.

21. Mair, W. and Dillin, A., 2008. Aging and survival: the genetics of life span extension by dietary restriction. Annu. Rev. Biochem., 77, pp.727-754.

22. Colman, R.J., Beasley, T.M., Kemnitz, J.W., Johnson, S.C., Weindruch, R. and Anderson, R.M., 2014. Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys. Nature Communications, 5, p.3557.

23. Guevara-Aguirre, J., Balasubramanian, P., Guevara-Aguirre, M., Wei, M., Madia, F., Cheng, C.W., Hwang, D., Martin-Montalvo, A., Saavedra, J., Ingles, S. and de Cabo, R., 2011. Growth hormone receptor deficiency is associated with a major reduction in pro-aging signaling, cancer, and diabetes in humans. Science Translational Medicine, 3(70), pp.70ra13-70ra13.

24. De Cabo, R., Carmona-Gutierrez, D., Bernier, M., Hall, M.N. and Madeo, F., 2014. The search for antiaging interventions: from elixirs to fasting regimens. Cell, 157(7), pp.1515-1526.

25. Mattson, M.P. and Wan, R., 2005. Beneficial effects of intermittent fasting and caloric restriction on the cardiovascular and cerebrovascular systems. The Journal of Nutritional Biochemistry, 16(3), pp.129-137.

26. Rando, T.A. and Finkel, T., 2013. Cardiac aging and rejuvenation—a sense of humors?. New England Journal of Medicine, 369(6), pp.575-576.

27. Shaw, A.C., Goldstein, D.R. and Montgomery, R.R., 2013. Age-dependent dysregulation of innate immunity. Nature Reviews Immunology, 13(12), p.875.

28. Butler, J.A., Mishur, R.J., Bhaskaran, S. and Rea, S.L., 2013. A metabolic signature for long life in the C aenorhabditis elegans M it mutants. Aging Cell, 12(1), pp.130-138.

29 Wilson, J.M., Fitschen, P.J., Campbell, B., Wilson, G.J., Zanchi, N., Taylor, L., Wilborn, C., Kalman, D.S., Stout, J.R., Hoffman, J.R. and Ziegenfuss, T.N., 2013. International Society of Sports Nutrition position stand: beta-hydroxy-beta-methylbutyrate (HMB). Journal of the International Society of Sports Nutrition, 10(1), p.6.

30 Mouchiroud, L., Molin, L., Kasturi, P., Triba, M.N., Dumas, M.E., Wilson, M.C., Halestrap, A.P., Roussel, D., Masse, I., Dallière, N. and Ségalat, L., 2011. Pyruvate imbalance mediates metabolic reprogramming and mimics lifespan extension by dietary restriction in Caenorhabditis elegans. Aging cell, 10(1), pp.39-54.

CHAPTER 4 : MITOCHONDRIA STUDY WITH DARTS-MS

4.1 INTRODUCTION

Our preliminary whole cell DARTS-MS analysis of α-ketoglutaric acid (α-KG) identified over 100 putative, previously unknown targets of α-KG, which is a full order of magnitude improvement over our previous gel-based DARTS analysis. However, the study was still limited to relatively high abundance cytosolic proteins. Therefore, to better understand the role of metabolites within the cell and improve the depth of our assay, we utilized subcellular fractionation in our next round of experiments. Given the widely divergent abundances of proteins within the cell, and especially within organelles, it may frequently be necessary to perform multiple subcellular DARTS-MS analyses (e.g. on mitochondrial, nuclear, and membrane extracts) in addition to whole-cell analyses. Although as LC-MS/MS techniques become faster and more sensitive this may be less of an issue. For now, subcellular fractionation should allow us to analyze binding to many proteins that were not identified in our preliminary study of cytosolic proteins and enable more detailed mapping of the metabolite-protein interaction network.

In this chapter, we will examine the interactions of several small molecules with proteins in mice liver-cell mitochondria. This will include the naturally occurring metabolite α-ketobutyric acid (α-KB) that was discussed in CHAPTER 3 and that has been implicated in lifespan extension (Figure 3-1, Figure 3-7). If α-KB is involved in the dietary restriction pathway, then, given the relatively few interactions observed in the whole-cell study and the fact that α-KB is naturally membrane permeable, it seems reasonable that we might

identify some of the key metabolite-protein interactions within the mitochondria. We will also investigate the antidiabetic medication metformin, whose interactions are not fully understood but are thought to occur primarily within the mitochondria [1]. It is well established that metformin decreases liver glucose production while increasing insulin-sensitivity but, despite being discovered nearly a century ago, the specific molecular target of metformin remains elusive. Lastly, we will examine the pesticide rotenone which is known to interact with complex I of the electron transport chain (ETC) in the mitochondria [2]. Rotenone is believed to be highly specific and will serve as a sort of control small molecule with a known target for comparison with our other analyses.

The three small molecules investigated in this chapter are structurally distinct (Figure 4-1) and are presumed to have widely varying specificities and mechanisms of action. However, they are all thought to directly target proteins associated with the mitochondria. By focusing on this smaller portion of the cell's proteins we hope to optimize our experimental procedures and demonstrate that the DARTS-MS platform is a powerful method for discovering new protein targets and regulatory functions across a diverse array of small molecules and systems

4.2 α-KETOBUTYRIC ACID

As discussed in CHAPTER 3, α-ketobutyrate (α-KB) is an endogenous metabolite that has been found to extend lifespan in adult *C. elegans* by nearly 50%, which is on the same order as the protection afforded by α-KG (Figure 3-1). It also appears that supplementation with α-KB can delay or even reverse certain aging related phenotypes, including hair loss and cataracts, in aged mice suggesting that the effects are conserved in mammals (Figure 3-7). Since α-KB is transported to the mitochondria following its production in the cell, it is reasonable to expect that many of its targets are mitochondrial proteins. Given these results and the apparent similarity between α-KB and α-KG, we will investigate the targets of α-KB within the mitochondria and try to identify any underlying mechanisms or relation to dietary restriction.

The protocol employed is described in detail in CHAPTER 2. Briefly, mitochondria were extracted from mouse liver, lysed, incubated with 100 μM α-KB, partially digest with pronase, fractionated by molecular weight, digested with trypsin and analyzed by LC-MS/MS. The raw MS data was analyzed with SEQUEST, Mascot, and Protomap. Potential binding targets are identified by changes in proteolytic cleavage as indicated by shifts in migration from higher to lower molecular weight species between the experimental and control samples.

There were around 1500 proteins detected in the α-KB experiment and 24 proteins detected that had significant differences in migration between the α-KB and control samples (see Figure 4-2 for an overview of the dataset and Table 4-1 for a complete list of proteins). It is worth noting that many more proteins were detected in this dataset than

anticipated. While it is possible that all proteins detected are localized to the mitochondria, it is likely that we are also detecting some cytosolic proteins. This should not seriously impact the results as long as the control and experimental samples are consistent. DARTS-MS is designed to work with any collection of protein, sub-cellular fractionation was employed in this study in order to help detect lower abundance proteins localized to the region of interest. As a result, cytosolic proteins may reduce our ability to detect other proteins of interest but should not invalidate any results obtained.

Since many intermediate metabolites serve a regulatory role in the cell, such as modifying the rate of metabolic enzymes in the same or related pathways, one might expect that α-KB will bind or modify the function of metabolic enzymes involved in amino acid synthesis or catabolism. However, many of the proteins detected do not seem to be directly related to this function. Another possibility, given its implication in longevity, is that α-KB directly interacts with proteins known to be involved in dietary restriction. Again, this does not immediately seem to be the case. It is certainly possible that the list of α-KB targets includes proteins that are not yet known to play a direct role in aging. It is also possible that the interactions are several steps removed from direct targets of aging and that multiple α-KB targets are required for its effect. Given the uncertainty of function, it may be necessary to test all possible modulations of α-KB interactions (e.g. by RNAi knockdown experiments with *C. elegans*). This is not an unreasonable result as many important regulator metabolites affect many different targets within the cell. Acetyl coenzyme A, for example, is used as a co-substrate in myriad biological processes and also serves as an allosteric regulator of numerous proteins. Thus, while our DARTS-MS analysis cannot directly uncover the biologically relevant effectors of α-KB, it is an

important first step in the proteome-wide characterization of α-KB targets and is crucial for expanding our understanding of the metabolite interaction network.

4.3 METFORMIN

Metformin is an antidiabetic medication that suppresses hepatic glucose production while increasing insulin sensitivity. Although it has been used in humans for over 50 years the mechanism of action is poorly understood [3]. Many possible protein targets have been suggested including complex I of the mitochondrial electron transport chain, AMP-activated protein kinase (AMPK), cAMP-dependent protein kinase, and glycerophosphate dehydrogenase [1, 4, 5]. In all cases metformin is thought to directly interact with mitochondrial proteins. In this study we used DARTS-MS to identify metformin protein targets and help uncover the underlying mechanism.

The protocol employed is described in detail in CHAPTER 2. Briefly, mitochondria were extracted from mouse liver, lysed, incubated with 10 μM metformin, partially digest with pronase, fractionated by molecular weight, digested with trypsin and analyzed by LC-MS/MS. The raw MS data was analyzed with SEQUEST, Mascot, and Protomap. The data was then organized into “peptographs” that display the sequence coverage for a particular protein on the horizontal axis (from N-terminus on the left to C-terminus on the right) while migration of the partially digested proteins is shown on the vertical axis (with high molecular weight on top and low molecular weight on the bottom). The peptographs also display the average number of spectral counts for the particular protein in each size exclusion fraction. Potential binding targets are identified by changes in proteolytic cleavage as indicated by shifts in migration from higher to lower molecular weight species between the experimental and control samples.

There were around 1500 proteins detected in the metformin experiment and 21 proteins detected that had significant differences in migration between the metformin and control samples (see Figure 4-3 for an overview of the dataset and Table 4-2 for a complete list of proteins). Based on the results, it does not appear that metformin directly interacts with either complex 1 or AMP-activated protein kinase. Although not all subunits for both proteins were detected, there was no significant difference between the control and metformin containing sample for the approximately 30 different subunits detected. Although detected, there was insufficient data to determine if cAMP-dependent protein kinase was effected by metformin. However, glycerophosphate dehydrogenase (glycerol-3-phosphate dehydrogenase) appears to be rendered more susceptible to proteolysis in the presence of metformin (Figure 4-4), which is in agreement with recent results in the literature [4]. Many of the other metformin targets seem unrelated to its therapeutic function. This does not mean that the results are necessarily spurious, however, as Metformin is not known to be a highly specific small molecule and may interact with many other targets beyond those that are therapeutically relevant. However, determining the biological significance of these targets, if there is any, will require addition analysis.

4.4 ROTENONE

The final small molecule analyzed in the mitochondrial study was rotenone. Rotenone is a pesticide that is known to directly bind to iron-sulfur centers in complex I [2]. The protocol employed is described in detail in CHAPTER 2. Briefly, mouse mitochondria were extracted, lysed, incubated with 10 μM rotenone, partially digest with pronase, fractionated by molecular weight, digested with trypsin and analyzed by LC-MS/MS. The raw MS data was analyzed with SEQUEST, Mascot, and Protomap. Potential binding targets are identified by changes in proteolytic cleavage as indicated by shifts in migration from higher to lower molecular weight species between the experimental and control samples.

There were around 1500 proteins detected in the rotenone experiment and 8 proteins detected that had significant differences in migration between the rotenone and control samples (see Figure 4-5 for an overview of the dataset and Table 4-3 for a complete list of proteins). Based on the results, rotenone was found to interact with complex I as expected. However, rotenone also seems to interact with metal binding proteins. For example, both hemoglobin and cytochrome b5 have bound iron co-factors, while isocitrate dehydrogenase [NAD] subunit gamma 1, binds magnesium or manganese. Another of the results, Putative transferase CAF17 homolog, is involved in "the maturation of mitochondrial 4Fe-4S proteins functioning late in the iron-sulfur cluster assembly pathway" [6]. This is not necessarily surprising as rotenone is known to inhibit the electron transfer within complex I at the terminal iron-sulfur cluster [7]. Therefore, it is possible that rotenone has a particular affinity for iron-containing proteins. However,

further characterization would be required to confirm these results. Regardless of these potential interactions, it is clear that rotenone is highly specific especially when compared to promiscuous natural metabolites like α-KG. Indeed, DARTS-MS detected only eight putative targets, out of 1500 possible targets, demonstrating a highly selective analysis.

4.5 FIGURES

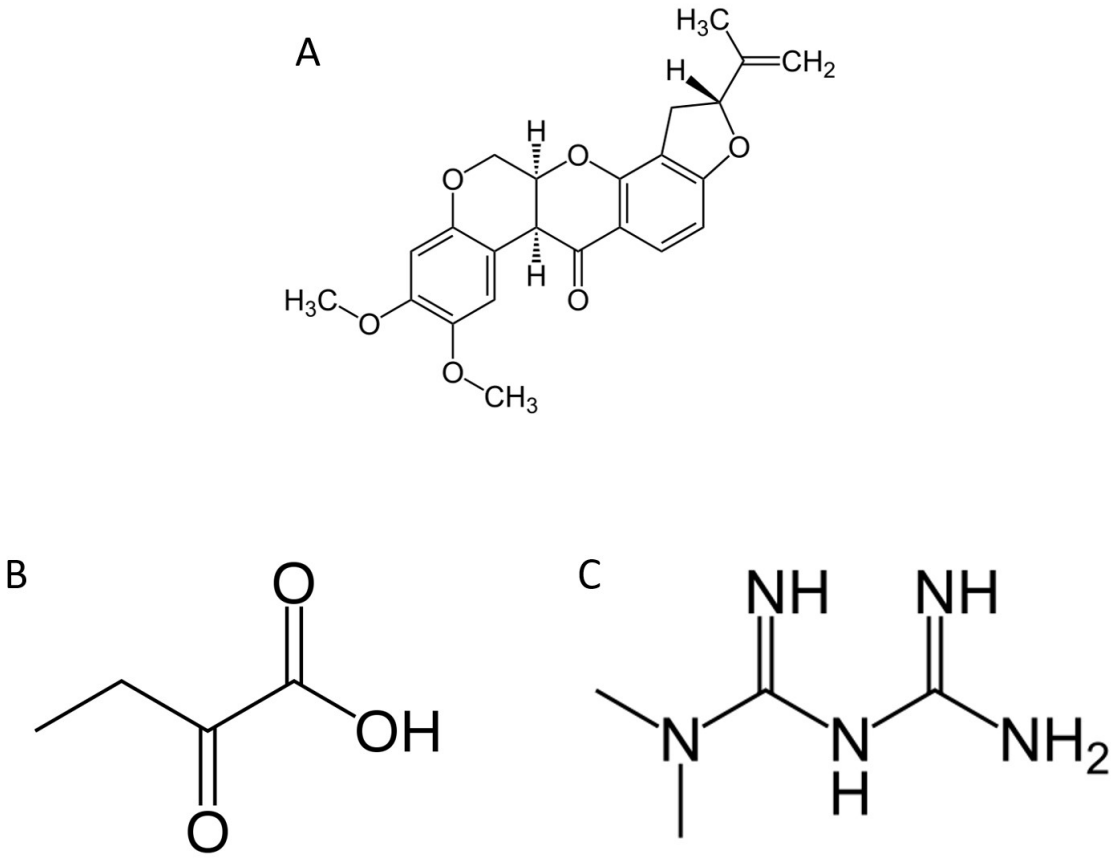


Figure 4-1: The structures of small molecules in mitochondria study are depicted

(A) Rotenone, (B) α-ketobutyric acid [α-KB], and (C) Metformin

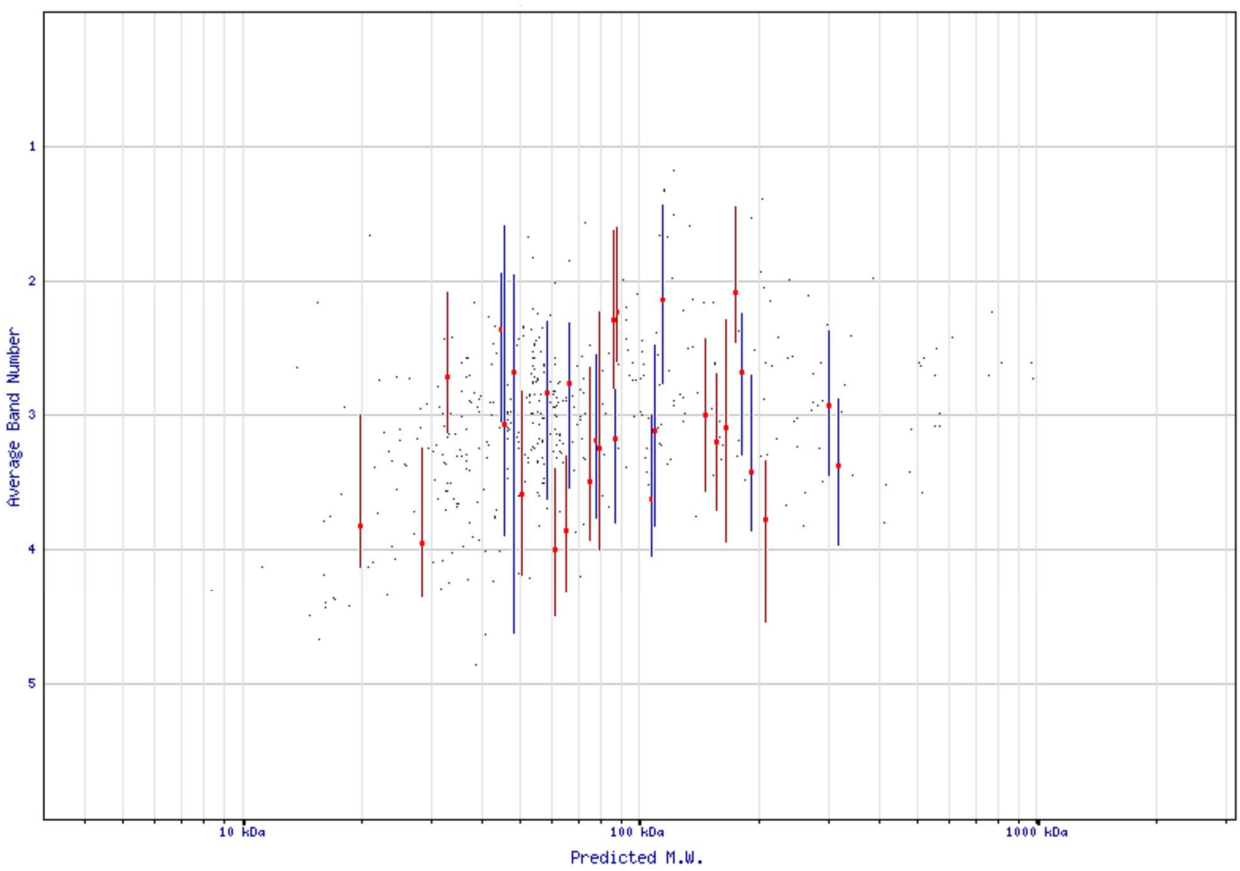


Figure 4-2: Shifts in molecular weight (MW) between α-KB and control samples indicating possible molecular targets. Each spot on the above plot represents a protein detected in both the experimental and control dataset. Red-spots with lines indicate proteins that have been protected from proteolysis by α-KB as evidenced by a shift in migration between control and experimental samples. Blue-spots with lines indicate proteins with an increased sensitivity to proteolysis. The size of the line indicating the magnitude of the shift. Black spots indicate no change in migration

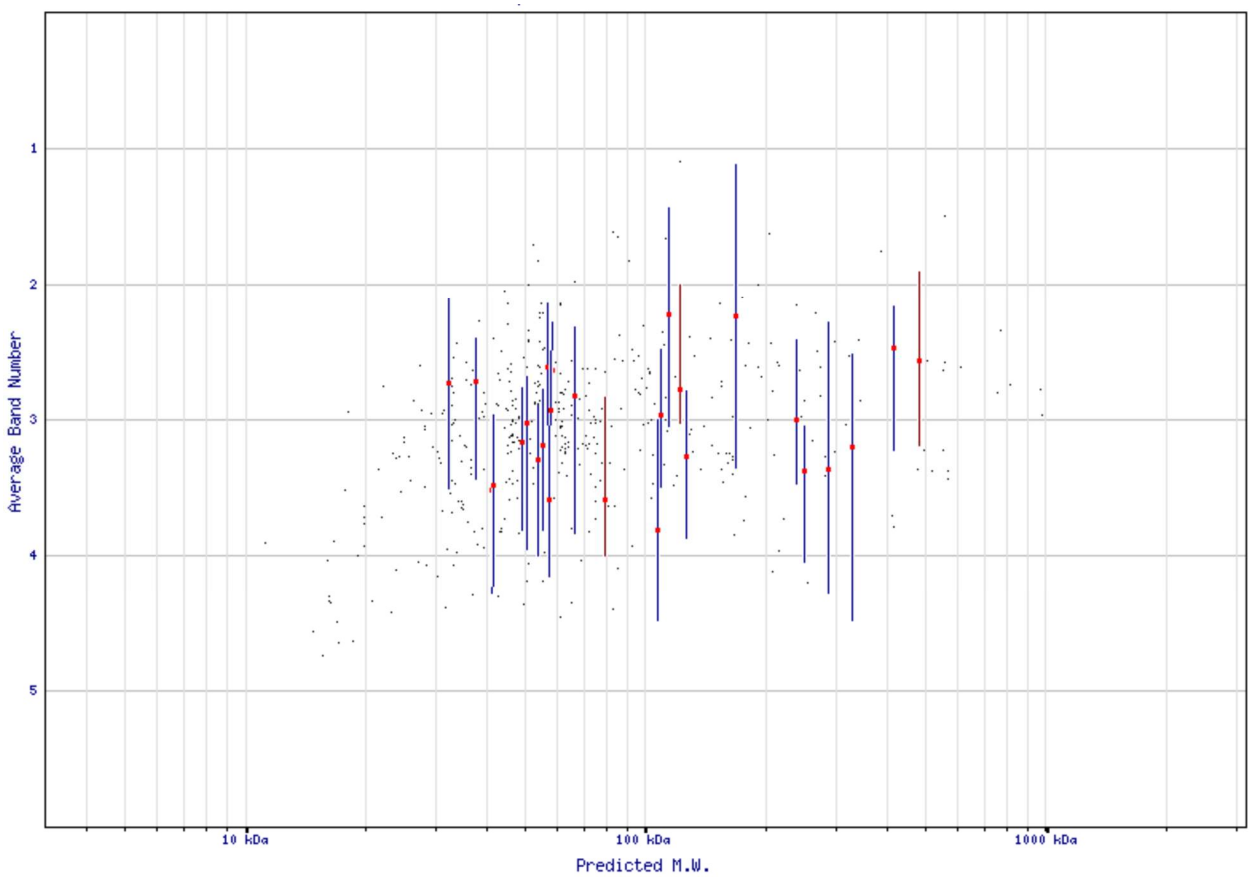


Figure 4-3: Shifts in molecular weight (MW) between metformin and control samples indicating possible molecular targets. Each spot on the above plot represents a protein detected in both the experimental and control dataset. Red-spots with lines indicate proteins that have been protected from proteolysis by metformin as evidenced by a shift in migration between control and experimental samples. Blue-spots with lines indicate proteins with an increased sensitivity to proteolysis. The size of the line indicating the magnitude of the shift. Black spots indicate no change in migration

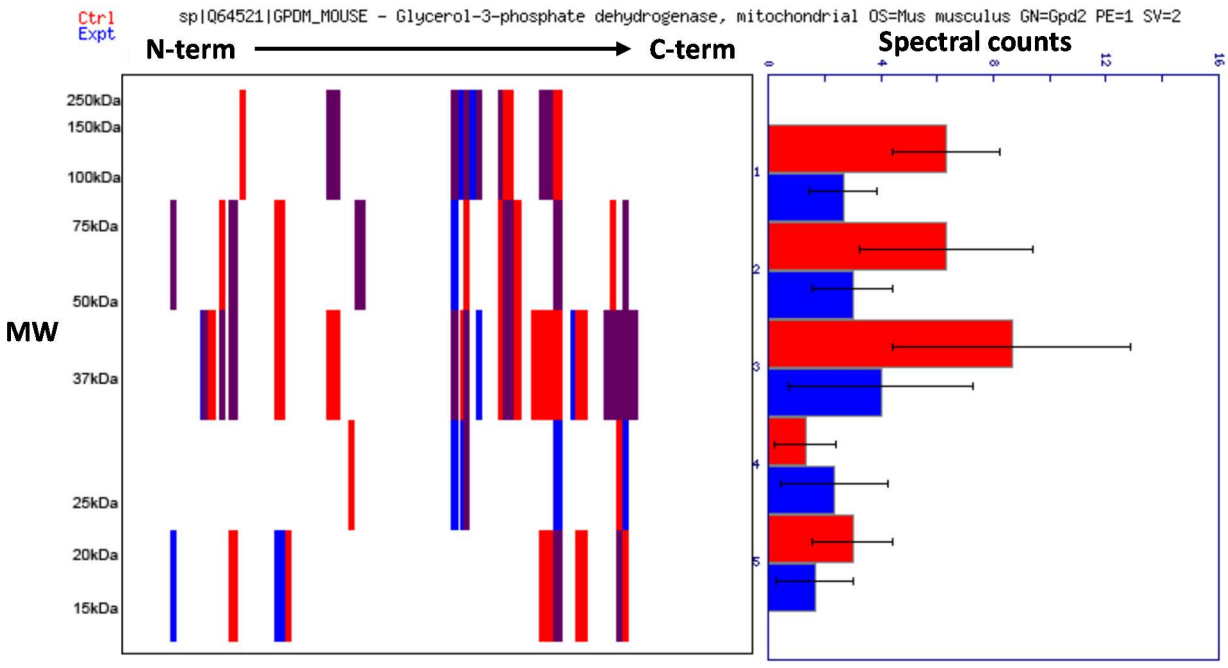


Figure 4-4: Results for glycerol-3-phosphate dehydrogenase in metformin DARTS-MS experiments All detected peptides related to glycerol-3-phosphate dehydrogenase in the metformin DARTS-MS experiments were integrated into peptographs using Protomap. In the left panel, sequence coverage for the protein is shown on the horizontal axis (from N-terminus on the left to C-terminus on the right) while migration of the partially digested proteins is shown on the vertical axis (with high MW on top and low MW on the bottom). The right panel shows the average number of spectral counts for the particular protein in each size exclusion fraction. Potential binding targets are identified by changes in proteolytic cleavage as indicated by shifts in migration from higher to lower MW species between the experimental (blue) and control (red) samples (purple is the overlap of both samples). In this case, it appears that metformin is sensitizing the protein to proteolysis.

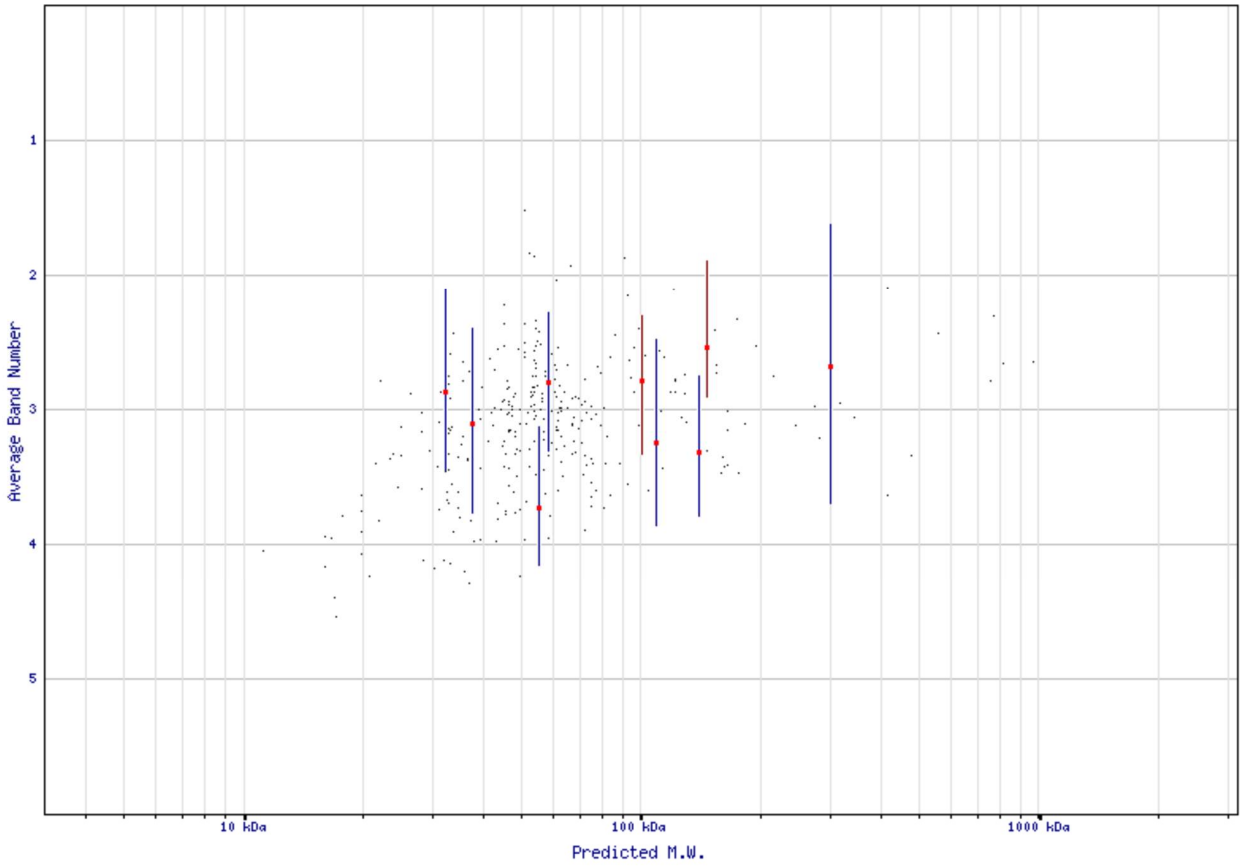


Figure 4-5: Shifts in molecular weight (MW) between rotenone and control samples indicating possible molecular targets. Each spot on the above plot represents a protein detected in both the experimental and control dataset. Red-spots with lines indicate proteins that have been protected from proteolysis by rotenone as evidenced by a shift in migration between control and experimental samples. Blue-spots with lines indicate proteins with an increased sensitivity to proteolysis. The size of the line indicating the magnitude of the shift. Black spots indicate no change in migration

4.6 TABLES

Table 4-1: Putative targets of α-KB based on shifts in migration. Each protein in the following table had a statistically significant difference between the α-KB samples and the control samples. The absolute value of the distance corresponds to the difference in average MW fraction between the experimental and control samples. Negative distances indicates a sensitization to proteolysis in the presence of the small molecule while a positive value corresponds to a protection from proteolysis.

Protein	Distance
Carbonic anhydrase 5A, mitochondrial	-1
Chromodomain-helicase-DNA-binding protein 6	1
Cilia- and flagella-associated protein 74	-1
Desmoplakin	1
FERM domain-containing protein 4B	1.3
Fibroblast growth factor receptor 4	-1
Forkhead-associated domain-containing protein 1	-1
G patch domain-containing protein 8	-1.6
GDH/6PGL endoplasmic bifunctional protein	1
Interferon regulatory factor 7	-1.3
Kinesin-like protein KIF2C	-1.7
Major urinary protein 1	-1.1
Microtubule-associated serine/threonine-protein kinase 2	1.1
Nesprin-3	1
Partitioning defective 3 homolog	-1.1
Phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1 protein	1
Potassium voltage-gated channel subfamily H member 1	1.3
Protein CASP	-1.2
STE20/SPS1-related proline-alanine-rich protein kinase	-1.1
Syntaxin-binding protein 3	-1
Thioredoxin-dependent peroxide reductase, mitochondrial	-1.1
Transcription initiation factor TFIID subunit 1	-1.2
Zinc finger and SCAN domain-containing protein 10	-1.1

Table 4-2: Putative targets of metformin based on shifts in migration. Each protein in the following table had a statistically significant difference between the metformin samples and the control samples. The absolute value of the distance corresponds to the difference in average MW fraction between the experimental and control samples. Negative distances indicates a sensitization to proteolysis in the presence of the small molecule while a positive value corresponds to a protection from proteolysis.

Protein	Distance
Actin, aortic smooth muscle	1.2
Ankyrin repeat domain-containing protein 17	1.9
ATP-binding cassette sub-family A member 7	1
Beta-actin-like protein 2	1.2
Chromodomain-helicase-DNA-binding protein 7	1.9
Cilia- and flagella-associated protein 36	1
Death-inducer obliterator 1	1
DNA polymerase delta catalytic subunit	-1
E3 ubiquitin-protein ligase HUWE1	-1.2
FERM domain-containing protein 4B	1.6
Glucosylceramidase	1.1
Glycerol-3-phosphate dehydrogenase, mitochondrial	-1
Kinesin-like protein KIF2C	-1.1
Nesprin-3	1.4
Potassium voltage-gated channel subfamily H member 1	1
Probable carboxypeptidase PM20D1	1
Protein TALPID3	2.2
SAGA-associated factor 29 homolog	1.3
Transforming acidic coiled-coil-containing protein 2	1
UDP-glucuronosyltransferase 1-2	1
Xin actin-binding repeat-containing protein 2	1

Table 4-3: Putative targets of rotenone based on shifts in migration. Each protein in the following table had a statistically significant difference between the rotenone samples and the control samples. The value of the fold change corresponds to the difference in average MW fraction between the experimental and control samples. Values less than one indicates a sensitization to proteolysis in the presence of the small molecule while values greater than one corresponds to a protection from proteolysis.

Protein	Fold Change
Cytochrome b5	0.13
Cytochrome c oxidase subunit 4 isoform 1, mitochondrial	0.16
Hemoglobin subunit alpha	0.25
Isocitrate dehydrogenase [NAD] subunit gamma 1, mitochondrial	0.11
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 9	0.08
Putative transferase CAF17 homolog, mitochondrial	0.23
Serine/threonine-protein kinase SMG1	0.21
Stabilin-2	10.80

4.7 REFERENCES

1. Rena, G., Pearson, E.R. and Sakamoto, K., 2013. Molecular mechanism of action of metformin: old or new insights? Diabetologia, 56(9), pp.1898-1906.
2. Hopkins, J., 1991. Handbook of pesticide toxicology. Volume 1. General principles. Volume 2 & 3. Classes of pesticides: Edited by WJ Hayes, Jr and ER Laws, Jr. Academic Press, Inc., San Diego 1991. pp. 1-1576. ISBN 0-12-334161-2; 0-12-334162-0; 0-12-334163-9.
3. Fischer, J. and Ganellin, C.R. eds., Analogue-based Drug Discovery. Chemistry International - Newsmagazine for IUPAC, 32.4 2010. pp 12-15
4. Madiraju, A.K., Erion, D.M., Rahimi, Y., Zhang, X.M., Braddock, D.T., Albright, R.A., Prigaro, B.J., Wood, J.L., Bhanot, S., MacDonald, M.J. and Jurczak, M.J., 2014. Metformin suppresses gluconeogenesis by inhibiting mitochondrial glycerophosphate dehydrogenase. Nature, 510(7506), p.542.
5. Kirpichnikov, D., McFarlane, S.I. and Sowers, J.R., 2002. Metformin: an update. Annals of Internal Medicine, 137(1), pp.25-33.
6. Uniprot.org. (2018). Iba57 - Putative transferase CAF17 homolog, mitochondrial precursor - Mus musculus (Mouse) - Iba57 gene & protein. [online] Available at: https://www.uniprot.org/uniprot/Q8CAK1 [Accessed 19 Jul. 2018].
7. Heinz, S., Freyberger, A., Lawrenz, B., Schladt, L., Schmuck, G. and Ellinger-Ziegelbauer, H., 2017. Mechanistic investigations of the mitochondrial complex I inhibitor rotenone in the context of pharmacological and safety evaluation. Scientific Reports, 7, p.45465.

CHAPTER 5 : CONCLUSIONS AND FUTURE DIRECTIONS FOR DARTS-MS

5.1 CONCLUSION

While DARTS-MS is still a method in development, the addition of gel-free protein fractionation and LC-MS/MS to the original DARTS protocol has resulted in significant improvements to the sensitive and adaptability of the analytical platform. The results detailed in this dissertation clearly indicate that DARTS-MS is a useful tool for elucidating the molecular targets of small molecules with biological and medical significance. We have looked at the current limitation in drug target identification and the early development of DARTS in CHAPTER 1. We have seen how affinity-based target identification methods require a derivative small-molecule with unchanged bioactivity and binding specificity while affinity-based target identification methods require a measurable biological response. DARTS was developed to overcome these limitations and early results show promise. In particular we saw that immunophilin protein FKBP12 is selectively protected from proteolysis in the presence of rapamycin and FK506 while EF-1 alpha is protected in the presence of didemnin B.

These early experiments were limited to highly abundant proteins, however, and a more sensitive protocol was developed and detailed in CHAPTER 2. By replacing the one-dimensional gel readout with gel-free fractionation and LC-MS/MS analysis, significant improvements to the platform were possible. These improvements were demonstrated in CHAPTER 3 where we examined whole cell lysates for interactions with several aging-related metabolites. We detected over 100 putative, previously unknown

targets of α-KG (an order of magnitude improvement over the previous DARTS platform) and reaffirmed dozens of known targets. We also expanded the known protein-metabolite interaction network by investigating other α-keto acids implicated in longevity and detected significant overlap in the protein targets of these natural products.

In addition to improvements in sensitivity, we demonstrated the diverse utility of DARTS-MS in CHAPTER 4. We analyzed the pesticide rotenone and reaffirmed its interaction with complex I of the electron transport chain (ETC) in the mitochondria. We also investigated the antidiabetic medication Metformin, and identified several potentially biologically relevant targets including mitochondrial glycerophosphate dehydrogenase. By using subcellular fractionation, we were able to delve deeper into the proteome and identify more targets of α-KB that could be responsible for its pro-longevity effects in *C. elegans* and mice.

It is clear, however, that DARTS-MS can be applied to areas beyond the aging related metabolites and the drugs and pesticides detailed here. It is intended to be a general method for the efficient identification of protein targets of small molecules. Most drugs produce a biological response from binding to one or more proteins, thus developing methods to detect these protein targets is an area of intense research. However, the detection of small-molecule protein targets remains a daunting task due to the enormous diversity in structure, activity and mechanism of action of small molecules. We believe that DARTS-MS can make a significant impact in this area. Since DARTS-MS can be performed with native proteins under physiological conditions, it is not necessary to create small molecule derivatives or modify targets. The method is also completely unbiased and thus useful for identifying both on and off-target interactions, as

any interaction between the small molecule and target can be detected. Thus DARTS-MS has the potential to rapidly and accurately identify the molecular targets of novel therapeutic compounds as well as expand our understanding of the complex interactions of natural products. We have demonstrated that the system is useful for a wide array of small molecules, from highly promiscuous natural metabolites, to highly specific pesticides, to many drugs of varying specificity. As with any new protocol, however, there are areas that can be improved and the rest of this chapter will focus on potentially fruitful areas for future DARTS-MS experiments.

5.2 QUANTIFICATION

While strictly speaking quantification is not necessary for DARTS-MS to identify the protein targets of small molecules, there are numerous potential advantages. Most notably quantification could be used to decrease sample variability and increase analysis sensitivity. Many current quantitative proteomics strategies utilize stable-isotope labeling for this purpose, generally by combing samples from different conditions with differentiating mass labels allowing them to be run as a single sample. Although there are many possible stable-isotope labeling approaches that could be useful in combination with DARTS-MS, one of the most promising is Stable Isotope Labeling of Amino acids in Cell culture (SILAC) [1].

SILAC utilizes cells grown under two different conditions, a “light” unlabeled culture and “heavy” labeled culture. The two cultures are identical except that the heavy culture is grown in the presence of one or two isotopically labeled amino acids (e.g. ^{13}C labeled L-Lysine and L-Arginine). As proteins are synthesized, the heavy labeled amino acids are incorporated into proteins and, after several generations, all of a particular amino acid are replaced with its heavy labeled equivalent. Although the heavy labels do not affect the biology within the cell (or downstream procedures such as proteolysis, fractionation, or chromatography), they are readily distinguishable by mass spectrometry. Thus light and heavy labeled samples can be combined reducing the number of samples to prepare and analyze by 50% well as reducing the sample variability introduced by downstream procedures. Since the control and experimental samples are combined and only distinguishable by the heavy labeled amino acid, it is necessary to choose a labeled

amino acid that is present in every peptide (for example all tryptic or semi-tryptic peptides will have at least one lysine or arginine).

In addition to reducing sample variability and preparation time, SILAC would increase the number of proteins we can consider in our analysis. Currently DARTS-MS requires a high number of spectra per protein in order to establish a statistically confident measure of the molecular weight fraction in which a given protein elutes. This requires a tradeoff between identifying more proteins and collecting a sufficient number of spectra for proteins already identified. As a result, many proteins, including potential targets, are not considered in our data analysis because changes in the migration pattern between molecular weight fractions cannot be determined with statistical confidence. For example, the known α-KG binding proteins α-KG-dependent dioxygenase FTO (FTO), isocitrate dehydrogenase subunit beta (IDH3B), and D-3-phosphoglycerate dehydrogenase (PHGDH), all appear to be strongly protected in the presence of α-KG but fail to meet the significance cutoff ($p < 0.05$) in Protomap (Figure 5-1). Implementation of SILAC with DARTS-MS would allow for the relative quantitation of all proteins allowing for the direct comparison between experimental and control samples. This would eliminate the need for collecting spectra beyond those required to identify the protein and allow us to set much more stringent exclusion criteria for spectra previously detected.

Alternatively, one could use an N-terminal tag to quantify protein abundances. As in the normal DARTS-MS protocol, the experimental and control samples would be subjected to a limited proteolysis. Subsequently the N-termini of the proteins and fragments would be tagged with an isotope-labeled NHS-biotin tag (with a light isotope tag for the control sample and a heavy isotope tag for the drug sample). The samples

could then be combined and the isotope-labeled peptides could be recovered via a streptavidin column and digested with trypsin before LC-MS/MS analysis. The peak areas for the light and heavy isotope labels could then be compared with significant differences representing possible drug targets.

Another potentially advantageous addition to our current DARTS-MS platform that doesn't require stable-isotope labeling would be the use of data-independent acquisition (DIA) strategies. The MS^E data-independent scanning mode allows precursor and product ion data to be measured concurrently by alternating between low and high energies in the collision cells of tandem mass spectrometers [2, 3]. The data is deconvoluted into separate spectra (one for the precursor and another for the resulting product) and are assigned to the corresponding chromatographic retention time. This data can then be searched in a database and should provide higher sensitivity for low abundance proteins. This can also be coupled with traveling wave ion-mobility separation (IMS) that allows for increased selectivity of precursor- to product-ion alignment [4, 5].

5.3 OPTIMIZATION OF SAMPLE PREPARATION

To date, in the DARTS-MS procedure, most of the following are potentially variable or chosen somewhat empirically based on our previous results: small molecule (drug) concentration, protease concentration, the type of protease or mixture of proteases, protease incubation time, protease incubation conditions, source protein concentration, and the number of fractions during the gel-free fractionation step. While the technique is useful as is, it would be worthwhile to use a well-characterized model system (e.g., cell lysate from *E. coli*) to establish optimized procedures for DARTS-MS to maximize reproducibility for robust and unbiased results and to aid adoption in other labs. For example, as a first test, an *E. coli* lysate could be spiked with a known concentration of an exogenous target protein (e.g., human carbonic anhydrase). The small molecule should be a highly specific inhibitor of the exogenous target protein (although it may be interesting to look for off-target effects as well). Once a standard system is set up, one could run further optimization experiments using the same cell lysate, exogenous protein and small molecule, but varying (1) small molecule concentration, (2) the type of protease, and (3) protease concentration. One could also vary the amount of exogenous protein added to the sample to address the issue of sensitivity for low abundance proteins. Following standard rigorous experimental design principles, all analyses should be performed with technical and sample triplicate measurements to ensure robust and statistically relevant data.

1. The starting small molecule concentration should be based on physiologically relevant conditions, i.e., if a metabolite and the concentration one would expect to see in the cell *in vivo*. Assuming reasonable results at physiologically relevant conditions can be obtained, one could test other likely dosages (e.g., acute vs. chronic use, overdose, etc.).

2. Based on our results, pronase appears to be the most useful protease for DARTS-MS, although other protease mixtures have not been studied extensively. Given that pronase is able to break down virtually all proteins into individual amino acids at sufficiently high concentrations, it removes the potentially fundamental limitation of protease sensitivity from DARTS-MS. However, there are potential advantages in having a less effective protease since we are interested in protection from proteolysis and a fully proteolyzed sample will not have that information. Obvious potential candidates are trypsin and Lys-C (and Arg-N and Lys-N). Both are highly specific, commercially available and compatible with MS/MS (and database searching). It may be worthwhile to develop alternatives with these proteases to make the DARTS-MS protocol more adaptable.

3. Typical protease concentrations for a DARTS-MS reaction are between 1:100 and 1:1000 (protease to sample protein by weight). Optimizing protease concentrations could be achieved by running a serial dilution of the protease. For example, running the reaction at protease concentrations of 1:30, 1:100, 1:300, 1:1000, and 1:3000. Ideally, these concentrations would be tested with multiple small molecule concentrations. The small molecule should at least be concentrated enough to have a noticeable amount of protection at most of the protease concentrations.

Other sample preparation related procedures to be optimized include accommodation for membrane protein fractions, as membrane proteins represents 60% of the known existing and future drug targets [6]. The DARTS-MS sample processing procedure should have no difficulties to address membrane proteins; we envision detergents common to membrane proteomics studies, such as NP-40 (nonyl phenoxypolyethoxylethanol), would be added for solubilization prior to cell lysis.

5.4 FIGURES

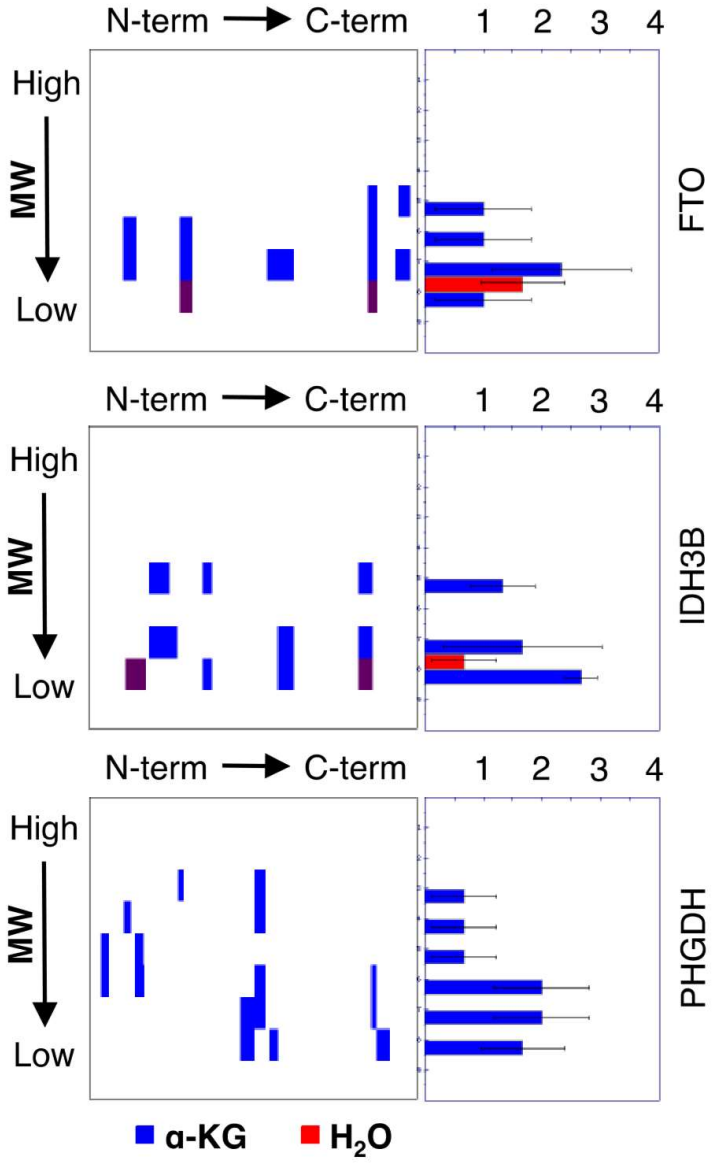


Figure 5-1: Gel-Free DARTS and spectral counting. Protomap analysis did not detect a significant difference for the known α-KG binding proteins α-KG-dependent dioxygenase FTO (FTO), isocitrate dehydrogenase subunit beta (IDH3B), and D-3-phosphoglycerate dehydrogenase (PHGDH) even though they appear to be highly protected.

5.5 REFERENCES

1. Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M., 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics*, *1*(5), pp.376-386.
2. Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P. and Geromanos, S.J., 2006. Absolute quantification of proteins by LCMS^{E} a virtue of parallel MS acquisition. *Molecular & Cellular Proteomics*, *5*(1), pp.144-156.
3. Geromanos, S.J., Vissers, J.P., Silva, J.C., Dorschel, C.A., Li, G.Z., Gorenstein, M.V., Bateman, R.H. and Langridge, J.I., 2009. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics*, *9*(6), pp.1683-1695.
4. Distler, U.; Kuharev, J.; Navarro, P.; Levin, Y.; Schild, H.; Tenzer, S. Drift Time-Specific Collision Energies Enable Deep-Coverage Data-Independent Acquisition Proteomics. Nature Methods 2014, 11, 167-170.
5. Distler, U., Kuharev, J., Navarro, P., Levin, Y., Schild, H. and Tenzer, S., 2013. Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nature Methods*, *11*(2), p.167.
6. Hopkins, A.L. and Groom, C.R., 2002. The druggable genome. *Nature Reviews Drug Discovery*, *1*(9), p.727.

APPENDIX 1 : STRUCTURE OF *TETRAHYMENA* TELOMERASE REVEALS PREVIOUSLY UNKNOWN SUBUNITS, FUNCTIONS, AND INTERACTIONS

INTRODUCTION

The appendix includes several projects that I was involved with that are not directly related to the main thesis. The first is a reprint with permission of the publication “Structure of *Tetrahymena* telomerase reveals previously unknown subunits, functions, and interactions” from Science October 30, 2015, Vol. 350, Issue 6260, for which I was a contributing author. This project is primarily the work of postdocs and graduate students in Dr. Juli Feigon’s lab, however, I was personally responsible for the all mass spectrometry involved including preparation of the samples, running of the instrumentation and analysis/interpretation of the data.

RESEARCH ARTICLE SUMMARY

STRUCTURAL BIOLOGY

Structure of *Tetrahymena* telomerase reveals previously unknown subunits, functions, and interactions

Jiansen Jiang,* Henry Chan,* Darian D. Cash, Edward J. Miracco, Rachel R. Ogorzalek Loo, Heather E. Upton, Duilio Cascio, Reid O'Brien Johnson, Kathleen Collins, Joseph A. Loo, Z. Hong Zhou, Juli Feigon†

INTRODUCTION: Telomerase is a ribonucleoprotein complex that extends the telomere DNA at the 3′ ends of linear chromosomes, thereby counteracting the loss of DNA from replication and nucleolytic processing. Although telomerase is largely inactive in somatic cells, it is active in stem cells and highly active in most cancer cell lines, where its activity is necessary for the cells' immortal phenotype. Thus, telomerase is an important regulator of aging, tumorigenesis, and stem cell renewal. Telomerase uses a template contained within the integral telomerase RNA (TER) and a telomerase reverse transcriptase (TERT) to synthesize multiple copies of the G-strand telomere repeat, which is TTGGGG in *Tetrahymena*. Telomerase can synthesize telomere repeat DNA in vitro with only TERT and TER, but physiological function requires a variety of other proteins. Telomerase recruitment to telomeres is regulated by the cell cycle, where its activity requires interplay between telomere end-protection and telomerase proteins. Telomere end-maintenance also requires coordinated recruitment of telomerase and DNA polymerase α for synthesis of G and C strands, respectively.

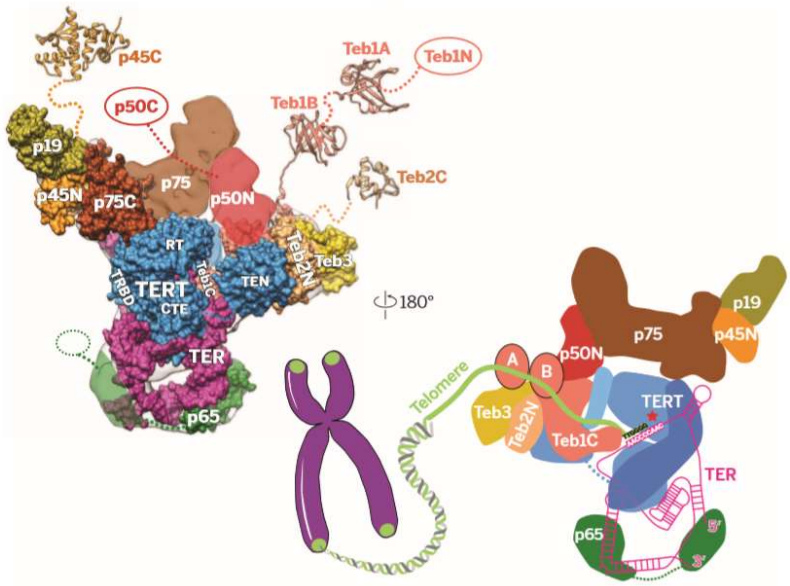
RATIONALE: A detailed mechanistic description of telomerase and its interaction at telomeres has been hampered by a lack of structural models, due to difficulties in obtaining samples of sufficient quantity and quality, as well as subunit complexity and flexibility and low sequence identity among subunits from different organisms. Unlike yeast and mammalian telomerase, *Tetrahymena* telomerase is constitutively assembled, making it possible to purify from cells and study all holoenzyme components in a stable complex. In addition to TERT and TER, the *Tetrahymena* telomerase holoenzyme contains the known proteins p65, p75, p45, p19, p50, and Teb1, a telomeric G-strand binding protein that is paralogous to the large subunit of heterotrimeric replication protein A (RPA). TER contains a template/pseudoknot domain enclosing a template with sequence complementarity to ~1.5 telomere repeats and a separate activating domain. Though TER is essential for activity, the physical arrangement of TER on TERT has remained largely unknown, as have details of protein subunit interactions.

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.aab4070

RESULTS: We report cryo–electron microscopy (cryo-EM) structures at ~9 Å resolution and pseudoatomic modeling of *Tetrahymena telomerase*, crystal structures of p19 and p45C, and a nuclear magnetic resonance structure of the TER pseudoknot. Two newly identified and functionally distinct RPA-related complexes, Teb1-Teb2-Teb3 (TEB) and p75-p45-p19, are connected to the TERT-TER-p65 catalytic core by p50. The presence of Teb2 and Teb3 is confirmed by mass spectrometry. p19 and p45 are structurally homologous to Ten1 and Stn1, respectively, which are found in the telomere end-binding complex CST (CTC1-STN1-TEN1) that recruits DNA polymerase α. The path of TER on TERT and the location of the TERT essential N-terminal domain (TEN) are revealed, providing mechanistic insights into telomerase function. A network of interactions between TEN, p50, and TEB regulates and enhances activity. p50-Teb1 appears to be structurally and functionally homologous to human TPP1-POT1. Negative-stain electron microscopy of labeled telomeric DNA bound to telomerase reveals the exit path from the template.

CONCLUSION: The structure of the *Tetrahymena* telomerase catalytic core and our identification of telomerase holoenzyme subcomplexes that are homologous to those found at mammalian, plant, and yeast telomeres provide new mechanistic insights and suggest commonalities of telomerase interaction, action, and regulation at telomeres. ■



Two views of the *Tetrahymena* telomerase holoenzyme. The top left image shows a front view of pseudoatomic models as spacefill superimposed on the cryo-EM map. Models of protein domains connected by flexible linkers and not visible in the cryo-EM map are illustrated as ribbons. The bottom right image shows a back-view schematic with the telomerase-bound telomeric G strand connected to a chromosome via double-stranded telomere DNA. Teb1AB domains are presumed to be ordered when DNA bound.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. E-mail: feigon@mbi.ucla.edu
Cite this article as J. Jiang *et al.*, *Science* 350, aab4070 (2015). DOI: 10.1126/science.aab4070

RESEARCH ARTICLE

STRUCTURAL BIOLOGY

Structure of *Tetrahymena* telomerase reveals previously unknown subunits, functions, and interactions

Jiansen Jiang,[1,2,3]* Henry Chan,[1]* Darian D. Cash,[1] Edward J. Miracco,[1]† Rachel R. Ogorzalek Loo,[4] Heather E. Upton,[5] Duilio Cascio,[1,6] Reid O'Brien Johnson,[1] Kathleen Collins,[5] Joseph A. Loo,[1,4,6] Z. Hong Zhou,[2,3] Juli Feigon[1,3,6]‡

Telomerase helps maintain telomeres by processive synthesis of telomere repeat DNA at their 3′-ends, using an integral telomerase RNA (TER) and telomerase reverse transcriptase (TERT). We report the cryo–electron microscopy structure of *Tetrahymena* telomerase at ~9 angstrom resolution. In addition to seven known holoenzyme proteins, we identify two additional proteins that form a complex (TEB) with single-stranded telomere DNA-binding protein Teb1, paralogous to heterotrimeric replication protein A (RPA). The p75-p45-p19 subcomplex is identified as another RPA-related complex, CST (CTC1-STN1-TEN1). This study reveals the paths of TER in the TERT-TER-p65 catalytic core and single-stranded DNA exit; extensive subunit interactions of the TERT essential N-terminal domain, p50, and TEB; and other subunit identities and structures, including p19 and p45C crystal structures. Our findings provide structural and mechanistic insights into telomerase holoenzyme function.

Telomerase is a ribonucleoprotein (RNP) complex that extends the telomere DNA at the 3′ ends of linear chromosomes, thereby counteracting the loss of DNA from replication and nucleolytic processing (*1*, *2*). Although telomerase is largely inactive in somatic cells, it is active in stem cells and highly active in most cancer cell lines, where its activity is necessary for their immortal phenotype (*3*–*5*). Thus, telomerase is an important regulator of aging, tumorigenesis, and stem cell renewal. Telomerase uses a template contained within the integral telomerase RNA (TER) and a telomerase reverse transcriptase (TERT) to synthesize multiple copies of the G-strand telomere repeat (TTGGGG in ciliates and TTAGGG in vertebrates). Telomerase recruitment to telomeres is regulated by the cell cycle, where its activity requires interplay between telomere end-protection and telomerase proteins (*6*). Telomere end-maintenance also requires coordinated recruitment of telomerase and DNA polymerase α for synthesis of the G and C strands, respectively (*7*, *8*). In humans, telomerase is recruited to telomeres by components of shelterin (*6*). Specifically, the TERT essential N-terminal (TEN) domain interacts with TPP1 (*9*–*11*), which, in complex with protection of telomeres 1 (POT1), is a processivity factor (*6*, *12*, *13*). The budding yeast telomerase holoenzyme subunit Est3 is a structural homolog of the oligosaccharide/oligonucleotide-binding (OB) fold of TPP1 (*14*) and, like TPP1, interacts with the TEN domain (*15*). In addition to recruiting telomerase, the human TPP1-POT1 complex recruits the replication protein A (RPA)–like CST (CTC1-STN1-TEN1) complex. RPA binds sequence-nonspecifically to single-stranded DNA (ssDNA) and plays a central role in DNA replication and repair through protein recruitment (*16*); CST complexes have been proposed as telomere-specific RPAs (*17*). CST stimulates DNA polymerase α for C-strand synthesis and has other diverse functions in different organisms (*8*, *18*–*20*). Mammalian CST acts as an inhibitor of telomerase action, determines telomeric 3′ overhang structure, and plays broader roles in telomere duplex replication and genome-wide replication restart (*8*, *18*, *20*, *21*). Budding yeast CST (Cdc13-Stn1-Ten1) subunit Cdc13 recruits the telomerase holoenzyme to telomeres via interaction of Cdc13 with telomerase subunit Est1, associated with Est3 (*22*–*24*).

Telomerase can synthesize telomere repeats in vitro with only TERT and TER, but physiological function requires a variety of other proteins (*25*, *26*). Unlike yeast and mammalian telomerase, *Tetrahymena* telomerase is constitutively assembled (*27*, *28*), making it possible to purify and study all holoenzyme components in a stable complex (*29*). In addition to TERT and TER, the *Tetrahymena* telomerase holoenzyme contains six other known proteins: p65, p75, p45, p19, p50, and Teb1 (*28*). TER contains a template/pseudoknot (t/PK) domain, which encloses a template with sequence complementarity to ~1.5 telomere repeats, and a separate activating domain (*30*). Although TER is essential for activity, the physical arrangement of TER on TERT has remained largely unknown. TERT comprises a telomerase RNA binding domain (TRBD), reverse transcriptase (RT), and C-terminal extension (CTE) that form a TERT "ring" (*31*), and a separate TEN domain (*32*) that is important for DNA handling and telomere repeat addition processivity (RAP) (*6*, *26*). p65 binds the TER activating domain [stem-loop 4 (SL4)], inducing a large bend in the RNA that facilitates assembly of TERT with TER to form the RNP catalytic core (*33*, *34*). Teb1 is a paralog of human RPA70, the large subunit of RPA (*16*, *28*). Direct single-stranded telomere DNA binding by Teb1 is necessary for telomerase recruitment to telomeres (*27*), where it may compete for binding with the *Tetrahymena* telomere end-binding Pot1a (*35*). p75, p45, and p19 form a ternary complex whose structure and function remain largely unknown (*28*). The 25 Å resolution negative-stain electron microscopy (EM) structure and individual subunit affinity labeling revealed the overall architecture of *Tetrahymena* telomerase, where TERT occupies the center of the holoenzyme with p65 bound to SL4 below (*29*). p50, which forms a central hub linking the TERT-TER-p65 catalytic core, p75-p45-p19, and Teb1, can greatly increase RAP and, with Teb1, can dramatically enhance the rate and processivity of long-product synthesis (*29*, *36*). A detailed mechanistic description of telomerase and its interaction at telomeres has been hampered by a lack of structural models, due to difficulties in obtaining samples of sufficient quantity and quality, as well as subunit complexity and flexibility and low sequence identity among subunits from different organisms. Here we report 9.4 and 8.9 Å cryo–electron microscopy (cryo-EM) structures of *Tetrahymena* telomerase holoenzyme. Combining the cryo-EM structures with data from x-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, negative-stain EM, and mass spectrometry, we discovered two RPA-related complexes: an RPA paralog TEB, comprising Teb1 and previously undetected proteins Teb2 and Teb3, and a CST complex comprising the previously structurally uncharacterized p75-p45-p19. Both are tethered to p50, a potential structural and functional homolog of TPP1, which in turn binds TERT. A pseudoatomic model of the TERT-TER-p65 catalytic core reveals the path of TER on TERT and the location of the TEN domain, and an exit path for the telomeric repeat DNA is proposed, together providing insights into the enzyme mechanism.

Cryo-EM reconstruction and overall structure

Telomerase holoenzyme endogenously assembled in *Tetrahymena thermophila* was affinity-purified

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA. [2]Department of Microbiology, Immunology, and Molecular Genetics, UCLA, Los Angeles, CA 90095, USA. [3]California Nanosystems Institute, UCLA, Los Angeles, CA 90095, USA. [4]Department of Biological Chemistry, UCLA, Los Angeles, CA 90095, USA. [5]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA. [6]UCLA–U.S. Department of Energy (DOE) Institute of Genomics and Proteomics, UCLA, Los Angeles, CA 90095, USA.
*These authors contributed equally to this work. †Present address: Moderna Therapeutics, Cambridge, MA 02141, USA.
‡Corresponding author. E-mail: feigon@mbi.ucla.edu

from a strain bearing a C-terminal 3×Flag (F) and tandem protein A (ZZ) tag on TERT (TERT-FZZ) (*29*). Cryo-EM specimens of *Tetrahymena* telomerase holoenzyme were prepared using "holey" carbon grids and imaged using a Gatan K2 Summit direct electron detection camera with drift correction (fig. S1). In addition to the preferred front view, particles showed various orientations, which are required for three-dimensional (3D) reconstructions (fig. S2 and methods). Using 40,754 particles, we obtained the intact structure of the *Tetrahymena* telomerase holoenzyme at an overall resolution of 9.4 Å (Fig. 1, A to C, and fig. S1). The negative-stain EM study revealed that the p75-p45-p19 subcomplex is conformationally dynamic (*29*), which is also apparent in the cryo-EM images (fig. S1, C and D). Therefore, to improve the resolution of the less flexible region, we used a soft mask to exclude the p75-p45-p19 subcomplex from the cryo-EM structure refinement. The resulting reconstruction has an overall resolution of 8.9 Å, with distinguishable secondary structure elements of proteins and RNA (Fig. 1, D to F, and figs. S1 and S3).

Using the features of the secondary structure elements, we were able to rigid-body fit the available atomic-resolution and homology models of protein domains and RNA helical elements (table S1) unambiguously into the 8.9 Å cryo-EM map, except for p75-p45-p19 homologs that were fit into the 9.4 Å cryo-EM map (Fig. 1 and fig. S3). All helical elements of TER—i.e., stem 1 (S1), stem-loop 2 (SL2), SL4, and the pseudoknot (PK)—are clearly visible with distinct grooves (Fig. 1F), and the path of the single-stranded regions can be approximately traced, although bases cannot be discerned. This allowed us to build a pseudoatomic model of the TERT-TER-p65 catalytic core and Teb1C. The locations of these subunit domains are the same as modeled in the negative-stain EM map within experimental resolution (*29*), with two exceptions: Teb1C [whose C terminus was correctly localized by Fab labeling (*29*)] is behind the TEN domain where the PK had been placed, whereas the PK is on the opposite side of TERT by the CTE. In addition, as discussed below, Teb1 forms a heterotrimer with two newly identified proteins, whose location was previously assigned to Teb1C. The p75-p45-p19 ternary complex, whose subunit boundaries could not be determined in the negative-stain EM map (*29*), is revealed to contain an RPA-like heterotrimer of OB-fold proteins with the domain structure of a CST complex. Overall the pseudoatomic models of the TERT-TER-p65 catalytic core, Teb1-Teb2-Teb3, and p75-p45-p19, which are all linked to p50 in the cryo-EM maps, reveal an intricate network of interactions between the subunits.

The TER t/PK domain encircles the TERT ring

Tetrahymena TERT TRBD-RT-CTE forms a ring-shaped structure and was modeled using the crystal structure of a partial *Tetrahymena* TRBD (residues 259 to 265 and 277 to 519) (*37*) and a homology model based on the RT-CTE in the *Tribolium* (flour beetle) TERT crystal structure (Fig. 2A) (*31*, *38*). *Tribolium* TERT lacks the TEN domain, and therefore the position of TEN relative to the TERT ring has remained uncertain. Fitting of the *Tetrahymena* TEN domain crystal structure (*32*) into the cryo-EM map (Fig. 1F) revealed its location on the active-site side of the TERT ring, stacked over the CTE (Fig. 2A and fig. S4A). Residues 640 to 740 of the insertion in fingers domain (IFD) in the RT [which are not modeled, as the *Tribolium* TERT IFD is much shorter than in *Tetrahymena* and other organisms (*39*, *40*)] appear to be near the TEN domain (Figs. 1B and 2A). The evolutionarily nonconserved sequence (residues 178 to 258) that connects TEN to the TERT ring has no available atomic structure and appears undefined in the cryo-EM maps. Consistent with this lack of a fixed conformation, the TEN domain can be assembled in trans with TER and the TERT ring to form an active catalytic core without linker residues 196 to 215 (*41*, *42*).

Tetrahymena TER comprises a circular t/PK domain, closed by a short S1, that contains the template, the 3′-flanking template recognition element (TRE), the template boundary element (TBE; also called the TERT binding element), SL2, and PK; and an activating domain (SL4), which is connected to S1 by a short single-stranded linker (Fig. 2B) (*30*). TER was modeled by fitting previously determined (SL2, p65xRRM:S4, L4) and newly determined (PK) (table S4) NMR structures into the cryo-EM density (see methods). The core t/PK domain encircles the TERT ring approximately perpendicularly, such that the template extends across the bottom of the RT domain on one side of the TERT ring, whereas the PK is on the other side next to the CTE (Fig. 2, C to E, and fig. S4). The bottom of S2 is adjacent to the TRBD CP motif, and the 3′ end of the TBE is also near the T motif (Fig. 2C and fig. S4), as predicted (*29*, *37*).

The PK, a conserved element of TER that has been proposed to contribute directly to catalysis (*43*), is on the opposite side of the TERT ring and far from the active site (~20 Å away) (Fig. 2, C and E, and fig. S4A). The PKs of human and yeast TERs contain conserved interactions that stabilize the PK fold through formation of base triples between loops and stems that are important for activity (*43*–*45*). The smaller *Tetrahymena* telomerase PK has two base triples in the NMR structure at 10°C and is not stably folded at higher temperatures in the absence of TERT (*46*). Based on the position of the PK on TERT, we conclude that, rather than contributing directly to catalysis, the correct PK fold is important for proper positioning of TER on TERT. It is plausible that the PK acts like a watchband ratchet clasp during catalytic core assembly, partially or completely unfolding to allow the TER t/PK domain to fit around the TERT ring and then locking onto the TERT ring by folding of the PK. An indirect effect on assembly may explain why mutations that either stabilize or destabilize the PK, at least for human TER, affect telomerase activity in vitro (*45*).

Outside of the TER t/PK domain, only L4 in S1/SL4 comes into contact with TERT (Fig. 2, C, E, and F;

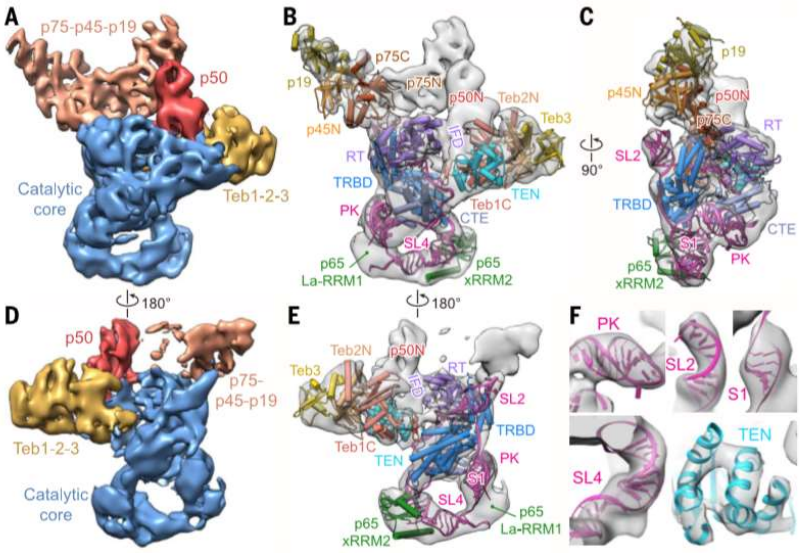


Fig. 1. Cryo-EM reconstructions of *Tetrahymena* telomerase holoenzyme. (**A**) Front view of the 9.4 Å cryo-EM map, with the catalytic core, Teb1C-Teb2N-Teb3 (TEB), p75C-p45N-p19 (CST), and p50N colored in blue, gold, copper, and red, respectively. (**B**) Front view of the 9.4 Å cryo-EM map (gray surface) with pseudoatomic models of the TERT-TER-p65 catalytic core, Teb1C-Teb2N-Teb3, and p75C-p45N-p19. (**C**) Side view of the cryo-EM map and pseudoatomic models shown in (B). (**D**) Back view of the 8.9 Å cryo-EM map colored as in (A). (**E**) Back view of the 8.9 Å cryo-EM map, showing pseudoatomic models of TERT-TER-p65 and Teb1C-Teb2N-Teb3. (**F**) Close-up views of fitting of TER helical domains PK, SL2, S1, and SL4 and TEN domain into the 8.9 Å cryo-EM map. TERT domains are TEN (cyan), TRBD (dark blue), RT (violet, with IFD labeled in violet), and CTE (light blue). Other proteins and TER are colored individually.

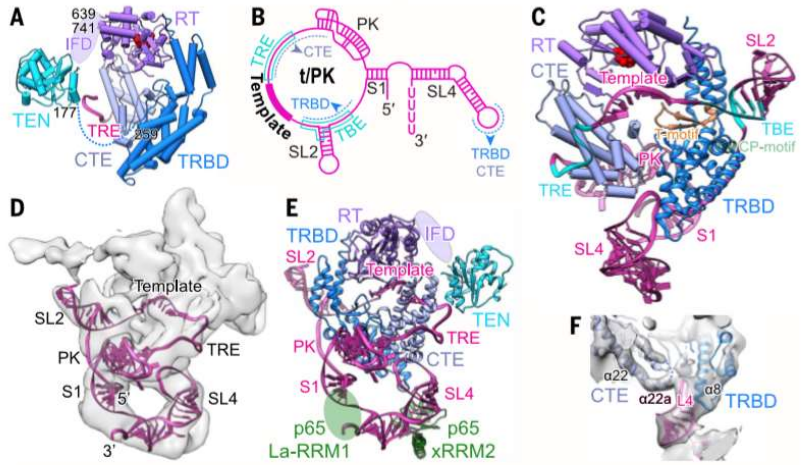


Fig. 2. Structure of the TERT-TER-p65 catalytic core. (**A**) Model structure of TERT, showing the putative IFD (light violet), active-site catalytic triad residues (red), linker (dotted blue line), and TER TRE. (**B**) Secondary structure of TER. Locations of CTE and TRBD on TER are indicated with dotted lines. (**C**) Pseudoatomic model of TERT ring-TER, showing TBE-template-TRE and L4 on TERT, viewed from the active-site side of TERT. (**D**) Front view of the 8.9 Å cryo-EM map, showing a pseudoatomic model of TER. (**E**) Front view of the pseudoatomic model TERT-TER-p65 catalytic core. The putative location of IFD is shown as violet oval. (**F**) A region of the 8.9 Å cryo-EM map and pseudoatomic model showing L4 at the interface of TRBD (helix α8) and CTE (helix α22a, residues 975 to 983). The numbering of *Tetrahymena* TERT secondary structure elements follows that of the *Tribolium* TERT structure.

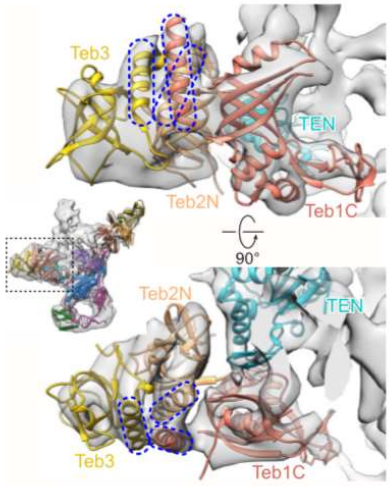


Fig. 3. Identification of two previously unknown holoenzyme proteins, Teb2 and Teb3. Shown are two views of the knob and Teb1C region of the 8.9 Å cryo-EM map with a model of Teb1C-Teb2N-Teb3 based on fitting of RPA70C-RPA32N-RPA14 into the cryo-EM map, followed by replacement of RPA70C with Teb1C, except for the RPA70C C-terminal α helix. The TEN domain is shown in cyan. The three-helix bundle between the C-terminal residues of Teb1C, Teb2N, and Teb3 (modeled from RPA) is highlighted with dashed lines. The inset shows the corresponding back view of the holoenzyme for reference.

and fig. S4A). The p65 La-RRM1 binds the 3′ polyU tail (*33*, *47*). La-RRM1 has low resolution in the cryo-EM map, probably due to flexibility or partial disassociation (fig. S1), but appears to be associated with both the 3′ tail and the 5′ end of S1, thereby possibly linking the 3′ and 5′ ends of TER (Fig. 2E). p65 C-terminal xRRM2 (*33*) binds to and bends S4, inserting L4 between the end of TRBD helix α8 and a short helix from the CTE (α22a) that does not exist in *Tribolium* TERT but has well-defined density in the 8.9 Å cryo-EM map (Fig. 2F). In complex with p65, SL4 stimulates activity, hierarchal assembly, and holoenzyme stability (*33*, *34*, *47*, *48*). L4 binds with high affinity to the TRBD, but its specificity for the CTE remains unknown. A hypothesis consistent with the above, as well as the location of L4 far from the active site and PK, proposes that L4 stabilizes a closed conformation of the TERT ring via specific interactions with TRBD and CTE.

The TER single-stranded region containing TRE, the template, and 3′TBE spans the active-site side of the TERT ring between the CTE and the TEN domain, across the RT, and over the TRBD T/CP pocket, respectively (Fig. 2C). The cryo-EM structure is of telomerase in the apo state before binding telomeric ssDNA. The TRE nucleotides, on the 3′ side of the template, pass between the CTE and the TEN domain and could contact the TEN domain. There is clear cryo-EM density for the 3′ half of the template only (fig. S3E), suggesting flexible positioning of the 5′ residues close to the TBE. Based on the distance from TER S2, the template appears to be positioned near where it would be at the end of telomere repeat synthesis—i.e., the 5′ end of the template is closest to the active site. The template also appears to be displaced by ~7 Å from its position at the active site in the crystal structure of *Tribolium* TERT in complex with an RNA-DNA hybrid helix (*38*), as proposed for the strand-separation step (*49*). The position of S2/TBE on TERT relative to the 5′ end of the template suggests a structural mechanism for template boundary definition. The single-stranded residues of the TBE, which flank S2, wrap on either side of the TERT ring (Fig. 2C and fig. S4B). At the end of telomere repeat synthesis, the S2 and TBE-TRBD interactions could act as an anchor to prevent residues beyond the template from being pulled into the active site (Fig. 2C and fig. S4B).

Teb1 forms an RPA-like complex with two newly identified proteins

Although Teb1 has four OB-fold domains (N, A, B, and C) (*50*), only the Teb1C domain and, in some cases, very weak density for Teb1B were apparently visible in the negative-stain EM class averages of *Tetrahymena* telomerase (*29*). Fitting of Teb1C into the cryo-EM map revealed density of unknown origin in the "knob" next to Teb1C (Fig. 3 and fig. S5A). An exhaustive analysis of the potential positions and fittings of other holoenzyme proteins revealed no known candidates for the knob density (fig. S5A). Because Teb1 is an RPA70 paralog, we investigated whether this part of the cryo-EM map might contain paralogs of the other two RPA subunits (RPA32 and RPA14). We found that the crystal structure of the RPA heterotrimeric core complex (RPA70C-RPA32N-RPA14) (*51*) fit particularly well into the cryo-EM map, with RPA70C positioned in the cryo-EM density for Teb1C and the RPA32N and RPA14 OB folds in the knob (Fig. 3). RPA trimerization requires the formation of a three–α-helix bundle from the C-termini of RPA70, RPA14, and RPA32N. The predicted C-terminal α helix of Teb1C is disordered in the crystal structure (*50*). However, in the cryo-EM map there is clear density of three α helices involved in the trimerization, with one of them near the C terminus of Teb1C, which we modeled in from the RPA70 structure (Fig. 3 and fig. S3H).

To confirm the presence of putative Teb1 heterotrimer proteins in the holoenzyme, we used liquid chromatography–tandem mass spectrometry (LC-MS/MS) to analyze telomerase purified from the TERT-FZZ strain. All seven of the known *Tetrahymena* telomerase proteins, plus two additional hypothetical proteins (TTHERM_001113129 and TTHERM_00439320) (www.ciliate.org), were detected with high confidence (table S2A). PSI-BLAST (Position-Specific Iterated Basic Local Alignment Search Tool) (*52*) searches of corrected cDNA sequences of these 31- and 14-kD proteins showed homology with predicted RPA32 and RPA14 homologs, respectively. Secondary structure predictions by Jpred4 (*53*) are consistent with an OB fold for the 14-kD protein and an N-terminal OB fold and C-terminal winged-helix (WH) domain for the 31-kD protein. Multiple sequence alignments of these domains with RPA32 and RPA14 orthologs show high similarity (fig. S6). We previously observed that telomerase purified from

a *Tetrahymena* strain containing N-terminally ZZF-tagged p50 (F-p50 telomerase) lacks the knob (fig. S5B) and has low RAP, which was attributed to the loss of Teb1 induced by the ZZF-tag on p50 (*29*). We analyzed F-p50 telomerase by LC-MS/MS and found that the two newly identified proteins and Teb1 are absent (table S2B), which confirms that these two additional proteins are located at the knob in the cryo-EM map and form a complex with Teb1. We conclude that there are two previously undetected proteins in the *Tetrahymena* telomerase holoenzyme that form a heterotrimer with Teb1, here named TEB, paralogous to the ssDNA binding RPA, except specific for telomeric G-strand DNA. In *Tetrahymena*, only the large subunit of RPA, Rfa1, has been identified (*54*). Transcript expression levels of the two newly discovered Teb1-binding proteins, here named Teb2 and Teb3, are much higher than other telomerase proteins and more similar to the level of Rfa1, as judged by expressed sequence tag abundance (*55*). These observations suggest that these two subunits may be shared between *Tetrahymena* RPA and TEB.

Multiple interactions between TEN, TEB, and p50 regulate telomerase activity

The TEN domain has been identified as a major determinant of telomerase activity (*6*, *42*). In addition to its potential TRE interaction (Fig. 2), the cryo-EM map reveals that TEN is adjacent to p50, Teb1C, Teb2N, and the IFD (Fig. 4, A and B). TEN, Teb1C, and p50 contact each other in a triangular arrangement. The Teb1C-TEN interaction is clearly defined in the pseudoatomic models (Fig. 4C). A Teb1 Phe^{590}→Ala^{590} (F590A)/F648A double mutant was previously shown to ablate purification of telomerase by Teb1-FZZ expressed in cells (*27*). These two residues of Teb1 are on the Teb1C-TEN interface (Fig. 4C), accounting for the loss of function. In the cryo-EM map, the density of TEN is also in contact with that of Teb2. TEN residues 77 to 87, which are missing in the crystal structure due to disorder (*32*), might form a structured interface with Teb2N in the holoenzyme (Fig. 4D). In vitro telomerase reconstitution activity assays, which lacked Teb2 and Teb3, showed that deletion of the Teb1C putative C-terminal α helix (ΔCTαH) had little effect on activity (*27*, *54*). However, purification of Teb1(ΔCTαH)-FZZ expressed in vivo did not recover any telomerase activity or holoenzyme subunits, indicating no in vivo telomerase assembly with Teb1(ΔCTαH) (*27*). Thus, Teb2 and Teb3, which interact with the Teb1 C-terminal α helix, likely stabilize association of Teb1 with telomerase.

To provide biochemical evidence for these interactions, we used in vitro reconstitution of complexes with individual TEB subunits expressed in rabbit reticulocyte lysate (RRL) to investigate whether Teb2-Teb3 had any effect on telomerase activity (Fig. 4E). Our activity assays were designed to be sensitive to improved holoenzyme assembly of Teb1 by the use of a very low level of Teb1 expressed in RRL, rather than a saturating concentration of purified bacterially expressed protein, and by purification of RNP from unbound Teb1 before the activity assay. Addition of Teb1 to the catalytic core plus p50 greatly increases RAP, as expected from previous studies (*29*, *36*). Coexpression of Teb2-Teb3 with Teb1 further increased overall activity, consistent with additional stabilization from synergistic interaction of Teb2 with the TEN domain and Teb1C. Addition of Teb2-Teb3 alone, without Teb1, provided no activity enhancement. Together, the cryo-EM structure, activity assays, and in vivo functional studies (*27*) discussed above support the notion that the TEN-TEB protein interaction network has crucial physiological importance.

The cryo-EM density of p50N appears to contact the TEN domain, as well as Teb1C, the IFD, and p75-p45-p19 (Figs. 1, B and E, and 4A). p50, which has no known sequence or structural homology to other proteins, has a 30-kD N-terminal domain (p50N) that is required for the high level of processive repeat synthesis conferred by p50 binding to the catalytic core (*36*). Only p50N is apparently visible in the class averages and 3D reconstructions of negative-stain EM images (*36*) and also in the cryo-EM maps. Like TPP1, p50 contacts the TEN domain (Fig. 4, A and B), and like TPP1-POT1 (*13*), p50-Teb1 greatly increases RAP (*36*). These data suggest that p50 could be a structural and functional ortholog to vertebrate TPP1. Although TPP1 was initially identified as a telomere binding protein complex with POT1 bound to the G-strand overhang to block telomerase access, it has emerged as also being the direct mediator of telomerase recruitment, activation, and homeostasis set-point regulation (*6*, *7*, *56*, *57*). We fit the structure of the TPP1 OB fold into the 8.9 Å cryo-EM map, but only the characteristic OB-fold β barrel matches well to the cryo-EM density (fig. S3, M to O). This is not surprising, due to the expected cofolding of p50 loops and helices with its interaction partners. Though the fitting does not provide definitive structural homology to TPP1, it does provide evidence that p50N is an OB fold.

The TEN domain has an extensive interface with p50 that apparently includes residues 122 to 127, which are missing in the crystal structure due to disorder (*32*) (Fig. 4B) and may cofold with p50. We tested whether other TEN domain interactions with p50 inferred from the pseudoatomic model and cryo-EM map were important

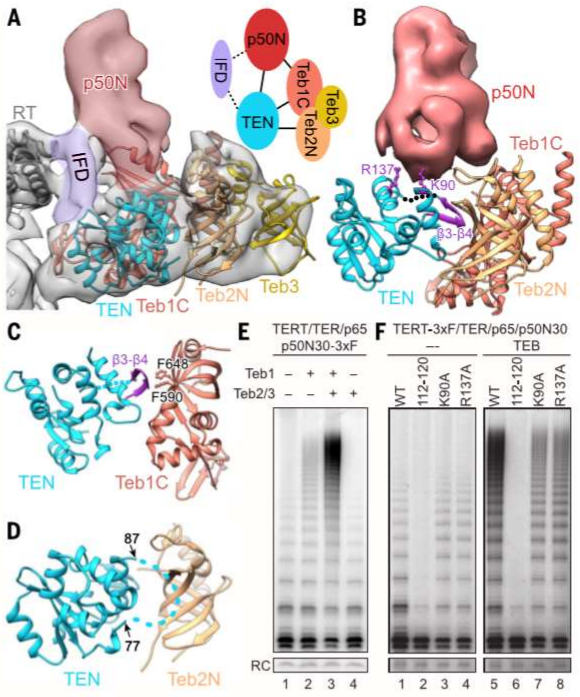


Fig. 4. Subunit interactions between p50, TEN, IFD, Teb1C, Teb2N, and Teb3. (**A**) Region of the 8.9 Å cryo-EM map showing density linking p50N, TEN, IFD, and TEB. The inset shows a schematic of the interactions. Dotted lines represent inferred interactions from the cryo-EM density and atomic models fitted to the cryo-EM map. (**B**) Interactions between p50N (red, cryo-EM density), TEN, Teb1C, and Teb2N. The TEN residue K90 side chain corresponding to the human TEN domain residue K78 that interacts with TPP1, R137, and the β3-β4 hairpin (residues 112 to 120) is highlighted in violet. (**C**) Interactions between TEN and Teb1C. Teb1C F590/F648 residues (orange stick) that together abrogate the Teb1C-TEN interaction and the TEN β3-β4 hairpin (violet) are indicated. TEN residues 122 to 127, disordered in the crystal structure, are represented by dotted lines in (B) and (C). (**D**) Interactions between TEN and Teb2N. TEN residues 77 to 87, disordered in the crystal structure, are shown as a dashed line. Pseudoatomic models of TEN, Teb1C, and Teb2N interactions are based on crystal structures of TEN, Teb1C, and RPA32N-RPA14 fit into the 8.9 Å cryo-EM map. (**E** and **F**) In vitro reconstitution telomerase activity assays for the effect of (E) TEB proteins and (F) TEN domain mutations on the catalytic core assembled with p50N30. TEN domain mutations were tested without (lanes 1 to 4) and with (lanes 5 to 8) TEB. WT, wild type; RC, recovery control.

for RNP activity stimulation by p50, in a manner similar to TEN interactions with TPP1 and Est3. A cluster of residues on human TPP1 (and Est3), called the TEL patch, has been identified as the surface of interaction with TEN (*9*–*11*, *14*, *56*, *57*). This interaction is essential for human telomerase recruitment to telomeres. Human TEN residues whose substitution disrupts the interaction of telomerase with TPP1 without greatly affecting catalytic activity have been identified; among these, a direct interaction between Lys^{78} (K78) and TPP1 has been demonstrated (*57*). The equivalent *Tetrahymena* TEN residue based on homology modeling of human TEN with the crystal structure of *Tetrahymena* TEN is K90, which is located near the density for p50 in the cryo-EM map (Fig. 4B). We used our in vitro telomerase reconstitution activity assay to investigate whether K90A at the putative TEN-p50 TEL patch interface imposes a defect in p50 stimulation of telomerase activity (Fig. 4F). K90A had a modest but notable effect on overall activity. The same results were obtained for TEN R137A (R, Arg), which is also located at the TEN-p50 interface in the cryo-EM map. These results support similar interactions between *Tetrahymena* TEN-p50 and human TEN-TPP1. TEN K90A and R137A did not abrogate TEB stimulation of high RAP (Fig. 4F), consistent with direct TEN domain interactions with TEB and p50-bridged TERT-TEB association. TEN residues 112 to 120 form a β hairpin, exclusive to *Tetrahymena* TEN, that inserts at the interface between p50, Teb1C, and Teb2N (Fig. 4B). An adjacent block sequence substitution $TEN_{(108\text{-}113}$NAAIRS) (N, Asn; I, Ile; S, Ser) abolishes p50 binding (*42*). In our experiment, we replaced the β hairpin with GSSG (G, Gly). This substitution abolished p50 activity stimulation, even with added TEB, which is consistent with disruption of both p50 and TEB interactions, but did not affect catalytic core activity. Taken together, these results validate the network of TEN-p50-Teb1C-Teb2N interactions proposed based on the cryo-EM structure (Fig. 4, A to D) and provide supporting evidence that p50 is a structural and functional paralog of TPP1.

The cryo-EM density that we attribute to the IFD appears to independently contact p50 and the TEN domain N terminus, which is disordered in the crystal structure (*32*) (Figs. 1B and 4A). The IFD is exclusive to TERTs and is important for the translocation step required for RAP (*39*, *40*). We suggest that the human TERT IFD may have parallel, as yet undetected, interactions with the TPP1 OB fold and TEN domains.

p75-p45-p19 is a CST complex

We determined a 2.3 Å x-ray crystal structure of p19 (table S3), which revealed an OB fold most structurally homologous to human Ten1 (Fig. 5A and fig. S7), suggesting the possibility that p75-p45-p19 might be a CST- or second RPA-like complex. We found that human RPA70C-RPA32N-RPA14, which fit the cryo-EM density of TEB as described above, also fit equally well in the density at the "tip" of the p75-p45-p19 subcomplex in the 9.4 Å cryo-EM map (Fig. 5B). For our model, RPA14 was then replaced with the p19 crystal structure, except for its unstructured C terminus, which forms an α helix in a three-helix bundle with the other two proteins in the RPA complex (Fig. 5B). p19 α2 and α3, which are lacking in RPA14, fit well into the cryo-EM density at the very end of the tip.

The positions of the C-terminal α helices of RPA70C and RPA14 correspond to the locations of the C termini of p75 and p19 determined by Fab labeling in negative-stain EM (Fig. 5B) (*29*). We used copurification and limited proteolysis to establish that p45 is composed of an N-terminal domain (p45N) that binds p19 and an independently folded C-terminal domain (p45C) (see methods). These results are consistent with the hypothesis that p45 is a Stn1 or RPA32 homolog. Previous attempts to locate the position of p45 in the holoenzyme by Fab labeling of its C terminus were unsuccessful, as Fab was apparently not visible in negative-stain EM class averages of telomerase holoenzyme (*29*). Based on the above, we hypothesized that p45C might be connected to p45N in the p75-p45-p19 subcomplex via a flexible linker, and thus Fab-bound p45C was not visible in holoenzyme class averages due to positional flexibility relative to the holoenzyme. We screened >2000 negative-stain EM particles of Fab-labeled p45-F telomerase. The majority had a cluster of three Fab (the 3×FLAG tag can bind up to three Fab) located in various positions ~100 Å from the location of p45N (Fig. 5D and fig. S8). The C-terminal domains of RPA32 and Stn1 are composed of WH and tandem WH (WH1-WH2) domains, respectively (*17*, *19*, *58*, *59*). We determined a 2.4 Å crystal structure of p45C (table S3) and found that it comprises the WH1-WH2 domains (Fig. 5C). The OB fold followed by two WH domains suggests that p45 is a Stn1 homolog (*17*, *19*, *58*, *59*). Structural alignments (fig. S7), sequence alignments (fig. S6), and phylogenetic cluster analysis (fig. S9) of p19 and p45 domains also support their identification as Ten1 and Stn1 orthologs, despite low sequence homology of CST proteins between kingdoms. In the 9.4 Å cryo-EM map, the remaining unmodeled density in the p75-p45-p19 complex, which appears to contain at least two additional OB folds and contacts p50, can only be attributed to p75 (Fig. 5B). Consistent with this observation, among p75-p45-p19, p75 is necessary and sufficient to bind p50 in vitro (*36*). There are distinct differences between p75 and Teb1 in terms of their interaction with p50 and their effect on telomerase activity. The cryo-EM maps show that p75 contacts p50 with a domain near its N terminus; in contrast, Teb1 binds p50 and the TEN domain with its C-terminal OB fold. How the potential p75 ortholog CTC1 interacts with telomerase is unknown, but it is interesting to note that the functionally divergent Cdc13 binds yeast telomerase holoenzyme protein Est1 through a recruitment domain adjacent to its N-terminal OB fold (*8*, *17*, *24*). In vitro reconstitution assays show that the addition of p75 (or p75-p45-p19) to the catalytic core plus p50 stimulates activity slightly but without an increase in

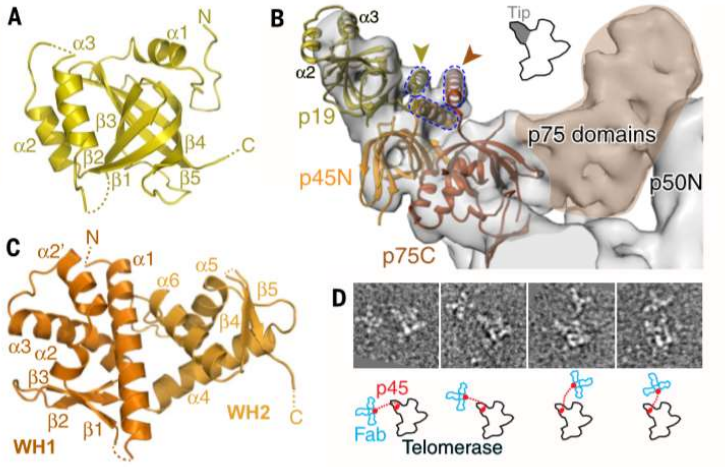


Fig. 5. Identification of p75-p45-p19 as a *Tetrahymena* CST complex. (**A**) Crystal structure of p19, an OB fold. (**B**) Model of p75C-p45N-p19 based on fitting of RPA70C-RPA32N-RPA14 into the 9.4 Å cryo-EM map, followed by replacement of RPA14 by p19, except for the RPA14 C-terminal helix. p19 α2 and α3 account for the density at the end of the tip. Gold and orange arrowheads point to the locations of the p19 and p75 C termini, respectively, previously determined by Fab labeling in negative-stain EM (*29*). The three-helix bundle between the C-terminal residues of p75, p45N, and p19 (modeled from RPA) is illustrated with dashed lines. (**C**) Crystal structure of p45C, a WH-WH domain. β5 is domain-swapped from a neighboring protein in the crystal lattice. (**D**) (Top) Negative-stain EM images of four typical p45-Fab–labeled telomerase holoenzyme particles, showing a cluster of 3 Fab bound to the C terminus of p45 at various locations near the holoenzyme. The side length of each image box is 44 nm. (Bottom) Corresponding outlines of telomerase and Fabs are shown in black and blue, respectively. Red dots indicate the p45C (attached to 3 Fab) and p45N (on telomerase) domains; dotted lines represent the linker.

RAP; in contrast, Teb1-p50 interaction stimulates high RAP (*29*, *36*). We conclude that p75-p45-p19 is structurally and functionally distinct from TEB and is most similar to a CST complex. CST has been identified in yeasts, plants, and mammals (*8*, *19*); this is the first evidence for the presence of CST in ciliates.

The telomeric DNA exits the template toward Teb1C

To obtain information on the path of telomere DNA on telomerase, we prepared a holoenzyme bound to a short telomeric DNA, biotin-5′-d(GTTGGG)$_2$GT$_L$T$_L$G$_L$T$_L$G$_L$G, where T$_L$ and G$_L$ are locked nucleic acid (LNA) nucleotides (*60*). For this DNA, six nucleotides (nts) should bind the template and 13 nts should extend out from its 3′ end. We then bound biotin with streptavidin, linking two telomerase holoenzymes together via the biotin binding sites in each streptavidin tetramer. Visualization of these dimers of the telomerase holoenzyme by negative-stain EM confirms that the telomere DNA is bound (Fig. 6, A to C). The class averages and 3D reconstructions show clearly visible density of streptavidin located near Teb1C-Teb2N and the putative location of Teb1B (*29*) on the backside of the telomerase holoenzyme (Fig. 6, B to D). These results reveal that telomeric DNA exits the template from the backside of telomerase and toward Teb1C (Fig. 6D). Furthermore, within the resolution (~30 Å) of negative-stain EM, there is no large-scale structural rearrangement of TERT between the apo structure without DNA and the structure with telomeric DNA bound to the template, consistent with the crystal structure of *Tribolium* TERT without TER (*31*) versus with an RNA-DNA hairpin mimicking a template-DNA hybrid (*38*).

Implications for telomerase catalytic activity and association with telomeres

The cryo-EM structures, p45C and p19 crystal structures, PK NMR structure, domain modeling, and mass spectrometry data presented here reveal a complex RNP composed of three ternary complexes tethered by p50: a TERT-TER-p65 catalytic core; an RPA paralog TEB, comprising Teb1-Teb2-Teb3; and a *Tetrahymena* CST complex comprising p75-p45-p19. TER wraps around the TERT ring and interacts with all four domains of TERT. The TEN domain is close to TER, the TERT ring IFD, p50, Teb1, and Teb2, signaling its importance in the catalytic cycle of telomerase. In addition to the protein domains visible in the cryo-EM map, there are several domains that are not observed due to flexible positioning or intrinsic disorder in the absence of binding partners; these are Teb1 N, A, and B; putative Teb2C, p45C, p50C, p65N; and possibly p75N (schematized in Fig. 7A). Teb1A and Teb1B bind telomere DNA (*54*), whereas the others (except for p65N) may recruit proteins involved in C-strand synthesis, nucleolytic end-processing, unwinding of G quadruplexes, telomere end-binding, and DNA repair, thereby coordinating telomere maintenance.

In the catalytic core, the TER interactions with TERT are exclusively from the t/PK domain and

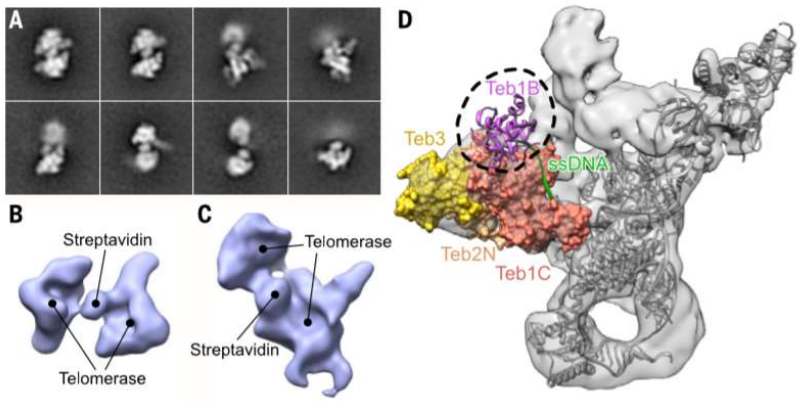


Fig. 6. Telomere ssDNA exits from the backside of the *Tetrahymena* telomerase holoenzyme. (**A**) Negative-stain EM class averages of the telomerase holoenzyme dimerized by primer-biotin-streptavidin-biotin-primer. The side length of each image box is 42 nm. (**B** and **C**) Random conical tilt (RCT) reconstruction of dimeric telomerase holoenzymes. One of two telomerase holoenzymes in (C) shows uninterpretable features due to flexible positioning of the two holoenzymes relative to each other. (**D**) A 9.4 Å cryo-EM map of the telomerase holoenzyme (gray surface) and the position of primer-attached streptavidin (black dashed circle), as identified by negative-stain EM RCT reconstruction in (B) and (C). Teb1B (purple) and ssDNA (green) exiting through Teb1C are modeled by fitting the crystal structure of the RPA:ssDNA complex (PDB ID 4GNX).

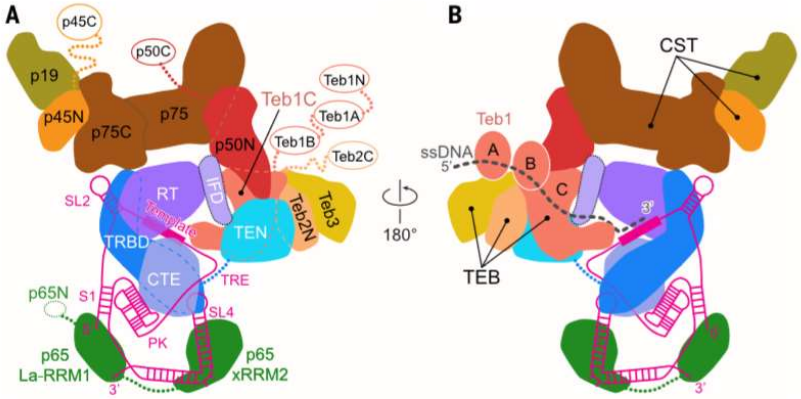


Fig. 7. Schematics of the complete *Tetrahymena* telomerase holoenzyme and DNA exit path. (**A**) Arrangement of the subunits and domains of the *Tetrahymena* telomerase catalytic core, TEB, and CST complexes tethered to p50, shown as front view. Domains connected by flexible linkers and not seen in the cryo-EM map are shown with oval outlines. (**B**) Arrangement of the holoenzyme with proposed path of telomere ssDNA, shown as back view. Teb1AB domains are presumed to be ordered when DNA-bound.

L4. Because a t/PK and a stem terminus element like L*4* are almost universally found in other organisms (*2*, *30*) and TERT is highly conserved, *Tetrahymena* telomerase provides general insights into TER-TERT interactions and assembly. Human TER has a much larger pseudoknot, but the region containing the tertiary stem-loop interactions is comparable in size and probably interacts with TERT at the same location near the CTE, and the human TER t/PK could similarly encircle the TERT ring. This would place the other end of the pseudoknot close to the TEN domain, as implicated biochemically (*41*). Vertebrate telomerase contains a conserved three-way junction, the CR4-CR5 domain, that binds the TRBD (*61*) and has a stem-loop (P6.1) that is critical for activity (*62*). The P6.1 loop probably inserts at the TRBD-CTE interface (*61*, *63*), similarly to *Tetrahymena* TER L*4*.

In the cryo-EM structure of apo telomerase presented here, the template is apparently positioned with the 5′ end near the active site. This places the 3′ end of the template close to Teb1C, where the telomeric DNA binds as it exits the holoenzyme. We suggest that this would facilitate (re)binding of the primer to the template 3′ end before translocation. Based on homology with the crystal structures of the RPA-ssDNA complex (*51*), a model of Teb1AB bound to ssDNA based on the Pot1AB-ssDNA complex (*50*), and the position

of the 5′ end of telomeric DNA on telomerase revealed by streptavidin labeling, we can model a possible path of the telomere DNA on telomerase (Fig. 7B). Although the DNA is shown going straight from the template to Teb1, it is possible that during the catalytic cycle the DNA strand may interact with TEN, TRBD, and/or CTE. Within the catalytic core, the TEN domain has been implicated to play a role in ssDNA handling and act as an anchor site (*32*, *64*) and/or in active-site use of the DNA–template RNA hybrid (*41*). The TEN domain is positioned so that the exiting end of the short DNA–template helix could contact its (unmodeled) N and C termini at the end of a complete telomere repeat synthesis, where this interaction could facilitate template DNA strand separation and help direct the DNA toward Teb1 (fig. S10).

The identification of two RPA related complexes in the *Tetrahymena* telomerase holoenzyme and their constitutive rather than cell-cycle–regulated association allows us to use studies of the *Tetrahymena* telomerase holoenzyme to inform models for interaction and function of human TPP1-POT1 and CST. Like p50-TEN, the interaction between human TPP1 and TEN is essential for bridging telomerase to telomeres (*9*, *11*, *56*, *57*). Although Teb1C is homologous to RPA70C, Teb1AB is more similar to POT1 (*50*), which suggests that TEB functions to temporarily prevent rebinding of the telomere-bound Tpt1-Pot1a complex (*65*). Consistent with the conformational flexibility of p75-p45-p19 around p50 (*29*), this CST complex could bind the telomeric DNA as it exits Teb1 to recruit DNA polymerase α for coordinated C-strand synthesis. In summary, the structure of the *Tetrahymena* telomerase catalytic core and identification of telomerase holoenzyme subcomplexes homologous to those found at mammalian, plant, and yeast telomeres provide new mechanistic insights and suggest commonalities of telomerase interaction, action, and regulation at telomeres.

Materials and methods

EM specimen preparation and data collection

Tetrahymena telomerase holoenzyme was purified following the previously described protocol (*29*), with some modifications. Briefly, 12 liters of cell culture of TERT-FZZ strain was used for tandem affinity purification following the established procedures (*29*). The final Flag elution was effected by incubating the telomerase-bound anti-Flag M2 affinity gel with 1.5 ml of elution buffer (20 mM HEPES·NaOH at pH 8.0, 50 mM NaCl, 1 mM $MgCl_2$, 1 mM TCEP·HCl, 0.025% IGEPAL CA-630, and 200 ng/μl of 3×Flag peptide) at 4°C. The eluate was incubated with 30 mg of Bio-Beads SM-2 absorbent (Bio-Rad) at 4°C for 2 hours by end-over-end rotation. The supernatant was then concentrated to 30 μl using a Microcon YM-10 centrifugal filter (Millipore).

For cryo-EM, 2.5 μl of the sample was applied to a glow-discharged Quantifoil R2/1 grid. The grid was blotted with filter paper to remove excess sample and was flash-frozen in liquid ethane with FEI Vitrobot Mark IV. The frozen-hydrated grids were loaded into an FEI Titan Krios electron microscope operated at 300 kV for automated image acquisition with Leginon (*66*). Micrographs were acquired with a Gatan K2 Summit direct electron detection camera operated in the electron-counting mode at a calibrated magnification of 36,764× (pixel size of 1.36 Å on the sample level) and defocus values ranging from −2.0 to −7.0 μm. A GIF Quantum LS Imaging Filter (Gatan) was installed between the electron microscope and the K2 camera, but the energy filter (slit) was not used. The dose rate on the camera was set to ~8 electrons (e^-) per pixel per second, and the total exposure time was 12 s fractionated into 48 frames of images with a 0.25-s exposure time for each frame. Frame images were aligned and averaged for correction of beam-induced drift using the graphics processing unit–accelerated motion correction program (*67*). The average images from all frames were used for defocus determination and particle picking, and those from the first 28 frames (corresponding to ~30 $e^-/Å^2$ of total dose on sample) were used for further data processing, including image classification and 3D structure refinement. A total of 4210 micrographs were used in the data processing.

Image processing

The defocus value of each cryo-EM micrograph was determined by CTFFIND (*68*), and the micrographs were corrected for contrast transfer function (CTF) by phase-flipping with the corresponding defocus and astigmatism values using Bsoft (*69*). A total of 478,698 particles were automatically picked using DoGpicker (*70*) and windowed out in dimensions of 256 pixels by 256 pixels using batchboxer in EMAN (*71*). The particles were binned to dimensions of 128 pixels by 128 pixels (pixel size of 2.72 Å) before the following processing. The binned particles were subjected to 2D and 3D classifications by RELION (*72*), following the recommended procedure (www2.mrc-lmb.cam.ac.uk/relion/). We performed two consecutive rounds of 2D classifications; in each round, the particles were classified into 100 classes for 25 iterations. After each round of 2D classification, class averages were visually inspected, and particles in the "bad" classes (i.e., those with fuzzy or uninterpretable features or fuzzy density of the p75-p45-p19 subcomplex) (fig. S1B) were removed. About 20% of the class averages were kept after each round of 2D classification, and 47,251 particles were selected for the following 3D classification. The previously obtained negative-stain EM reconstruction at 25 Å resolution (*29*) was low-pass filtered to 60 Å to serve as the initial model for 3D classification. The 3D classification generated five classes (fig. S1D), four of which showed an intact holoenzyme structure and were combined into a data set of 40,754 particles for 3D autorefinements by RELION as follows: First, a 3D autorefinement was performed using a spherical mask. The resolution was estimated to be 9.4 Å by the relion_postprocess program using the "gold-standard" Fourier shell correlation at a 0.143 criterion, with a soft mask for which the masking effect was corrected by phase randomization. Second, the density corresponding to the flexible p75-p45-p19 subcomplex was removed from the 9.4 Å cryo-EM map using the "Volume Erase" tool in UCSF Chimera (*73*), and the resulting map was used to generate a soft-edge mask by the relion_mask_create program. This soft-edge mask was then used in the other 3D autorefinement, in which the p75-p45-p19 subcomplex was excluded during the structure refinement and only included in the final reconstruction. The resolution of this focused refinement was 8.9 Å, estimated as described above. The cryo-EM maps were sharpened with B-factor and low-pass filtered to the stated resolution using the relion_postprocess program. The local resolution was calculated with ResMap (*74*), using two cryo-EM maps independently refined from halves of data.

Tetrahymena telomerase holoenzyme particles adopted a preferred front view in negative-stain EM grids (*29*) or in cryo-EM Quantifoil grids with holey carbon or coated with continuous carbon film, probably due to its irregular and flattened shape. We tried numerous sample and cryo-EM grid conditions to overcome this preferred orientation problem. The current protocol involving Bio-Beads had the most improved effect on the orientation distribution. However, ~60% of the particles were still in the front view configuration in the cryo-EM images. To assess the effect of preferred orientation on reconstruction resolution, the particles were divided into halves ("uniform" and "preferred") that showed more uniform or more preferred orientation distribution, respectively (fig. S2). Briefly, the particle orientation file (the "data.star" file generated by RELION 3D autorefinement) of 40,754 particles was analyzed to group particles with the same orientation (same AngleRot and AngleTilt). A threshold of particle count was set so that the particles above the threshold in each orientation group (where the particle count in that group was larger than the threshold) were sent to the preferred half, and the rest were kept in the uniform half. Separation of particles within an orientation group was random. These two halves of the data set were then subjected to 3D autorefinement, using the same parameters as those for the 8.9 Å reconstruction. Comparisons of orientation distributions and 3D reconstructions show that the uniform half had sufficient orientations to obtain a high-quality 3D reconstruction at an overall resolution better than 9.3 Å, and addition of the preferred half only improved the resolution modestly (fig. S2).

Fitting of atomic models into the cryo-EM maps

For the pseudoatomic model of TERT-TER-p65 and Teb1C, the x-ray crystal and NMR structures of *Tetrahymena* TERT TRBD [Protein Data Bank identification code (PDB ID) 2R4G] (*37*), TERT TEN (PDB ID 2B2A) (*32*), Teb1C (PDB ID 3U50) (*50*), TER SL2 (PDB ID 2FRL) (*29*), the model of p65 xRRM2:SL4 (*33*), and the homology model of *Tetrahymena* TERT RT-CTE domains based on *Tribolium* TERT (PDB IDs 3DU6 and 3KYL; TERT CTE residues 1058 to 1117 were not modeled due to low sequence homology to *Tribolium* TERT) (*29*, *31*, *38*)—which were previously assembled

together by fitting into the negative-stain EM 3D reconstruction (*29*)—were placed as a whole in an approximate position in the cryo-EM maps, followed by fitting the individual domain structures into the cryo-EM map using the "Fit in Map" function in UCSF Chimera (*73*). These domain structures or models fit into the cryo-EM map with subtle rotations and translocations, except for Teb1C, which was moved to the cryo-EM density behind TEN and next to TRBD and matched this with high consistency. The PK structure (PDB ID 2N6Q) was manually placed into the cryo-EM density matching its size and shape, as identified by exhaustive visual inspection of the cryo-EM map, and then fit into the density using UCSF Chimera. S1 was modeled as an ideal A-form helix and fit into cryo-EM density that matched its shape near the PK and SL4. The template was located by fitting the crystal structure of *Tribolium* TERT in complex with an RNA-DNA hybrid helix (PDB ID 3KYL) (*38*) into the cryo-EM map and finding unmodeled density near the RNA strand. The density assigned for the template is displaced by ~7 Å relative to that in the fitted *Tribolium* TERT. The remaining single-stranded regions of TER connecting the above fitted structure elements were modeled into the cryo-EM densities using Coot (*75*) as follows: First, ideal A-form single-stranded RNA fragments were approximately placed between the flanking high-resolution RNA structures. The bond angles of the RNA backbone were then manually adjusted until the backbone fit into the cryo-EM density that was assigned to the RNA after the protein subunits were fit into the cryo-EM map. Last, standard bond angles and lengths of the backbone of the single-stranded RNA fragments were achieved with the use of the "Regularize Zone" tool in Coot. Because the RNA bases were not distinguishable in the 8.9 Å cryo-EM map, only the backbone of the single-stranded regions was modeled. For the pseudoatomic model of Teb1-Teb2-Teb3, RPA70C:RPA32N:RPA14 from the crystal structure of RPA (PDB ID 1L1O) (*51*) was fit into the cryo-EM maps by the colores program of Situs, using the exhaustive 6D search algorithm (*76*). The structure of the RPA trimer fit with high consistency into the cryo-EM map in the density assigned for Teb1C and the adjacent knob. For the final model, RPA70C was replaced by Teb1C, which fits in the same density, except for the C-terminal helix that is disordered in the Teb1C crystal structure but is apparently well ordered in the cryo-EM maps. RPA32N and RPA14 in the model are paralogs of Teb2 and Teb3, respectively. During the Situs fitting of RPA70C:RPA32N:RPA14 to the 9.4 Å cryo-EM map, a second location, corresponding to the tip of the density assigned to the p75-p45-p19 subcomplex, was identified with high confidence, which suggests that it was also an RPA homolog. For the pseudoatomic model of p75C-p45N-p19, RPA14 was manually replaced by p19 (PDB ID 5DFM), except for its C-terminal helix that is disordered in the p19 crystal structure. The crystal structure of the TPP1 OB fold (PDB ID 2I46) (*13*) was placed in the cryo-EM density assigned to p50N and then fit using UCSF Chimera. The cryo-EM maps and the modeled protein and RNA structures were visualized with UCSF Chimera and PyMOL (*77*).

Structure determination of p19, p45C, and TER PK

Full-length p19 (*28*) was fused via a triple alanine linker to the wild-type maltose binding protein (MBP) vector as described in (*78*), expressed in *Escherichia coli*, and purified by amylose affinity and size exclusion chromatography (SEC). To obtain diffracting native crystals, a 2:1 mixture of 12 mg/ml of MBP-p19 and the reservoir solution (0.1 M sodium acetate at pH 4.6 and 2.0 M ammonium sulfate) was set up in a 24-well hanging drop format. Full-length p45 (*79*) was fused to His-tagged MBP via a tobacco etch virus (TEV) protease cleavage site, expressed in *E. coli*, and copurified with cells expressing 6×His-tagged p19 using Ni–nitrilotriacetic acid (NTA) affinity and SEC. p45C (residues 162 to 373) was identified via limited chymotrypsin proteolysis of full-length p45 bound to p19. p45C was fused to His-tagged MBP via the TEV protease cleavage site, expressed in *E. coli*, and purified by Ni-NTA affinity and SEC. To obtain diffracting native crystals, a 1:1 mixture of 14 mg/ml of p45C and the reservoir solution [0.1 M Na·HEPES at pH 7.5, 5% (v/v) MPD, and 10% (w/v) polyethylene glycol 6000) was set up in a 24-well hanging drop format. Data were collected at 100 K at Advanced Photon Source–Northeastern Collaborative Access Team (APS-NECAT) beamline 24-ID-C on a DECTRIS-PILATUS 6 M pixel detector. Data were indexed, integrated, and scaled using XDS/XSCALE (*80*). Structures were solved with the molecular replacement program PHASER (*81*), using initial models from SeMet data sets. Final models were iteratively built and refined in Coot (*75*) and PHENIX (*82*). In the crystal structure of p45C, β5 in WH2 is domain-swapped from a neighboring protein in the crystal lattice. This domain-swapped β5 was used to build a biological unit model of p45C (Fig. 5C). NMR assignments were obtained, and solution structure of the TER pseudoknot (nucleotides 68 to 100) was determined following established protocols for human and yeast TER pseudoknots (*44*, *45*, *83*). Sample conditions were ~1 mM RNA in 10 mM $NaPO_4$ at pH 6.3 and 50 mM KCl at 10°C. The structures were calculated with a total of 414 nuclear Overhauser effect distance, 171 dihedral angle, 82 H bond, and 79 residual dipolar coupling restraints using standard protocols in Xplor-NIH 2.9.8. Pairwise root mean square deviation for the 10 lowest-energy structures (out of 100) was 0.83 Å for all heavy atoms (table S4).

Mass spectrometry

Telomerase was purified from *Tetrahymena* TERT-FZZ and ZZF-p50 strains as described in (*29*), up to the final anti-FLAG resin wash step before elution. Telomerase was batch-washed on the resin three times by incubation in 1 ml of IGEPAL-free wash buffer at 4°C, rotating end-over-end for 10 min, and was eluted twice by incubation at room temperature with 50 μl of 0.1% trifluoroacetic acid for 30 min each time. For enzymatic digestion of the samples, 70 μl of deoxycholic acid (0.1% w/v) in 1 M NH_4HCO_3 was added to the elution. Cysteines were alkylated by adding iodoacetamide to 4 mM and incubating at room temperature. After 30 min, excess dithiothreitol was added to quench residual iodoacetamide, and trypsin (300 ng) was added to initiate digestion. The solution was incubated at 37°C overnight, after which it was acidified to pH 2 with trifluoroacetic acid, to precipitate the deoxycholic acid. The deoxycholic acid was extracted from the aqueous layer with three 200-μl aliquots of water-saturated ethyl acetate (*84*). The digested peptides were dried and resuspended in 1% formic acid. The resulting tryptic peptides were measured by LC-MS/MS, using an EASY-nLC 1000 (Thermo Scientific, Waltham, MA) coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Scientific) and an EASY-Spray nano–electrospray ionization source. Peptides were injected onto a 75 μm by 15 cm, 3 μm, 100 Å PepMap C18 reversed-phase LC column and separated using a linear gradient from 5% solvent B (0.1% formic acid in acetonitrile) and 95% solvent A (0.1% formic acid in water) to 50% solvent B in 2 hours at a constant flow rate of 300 nl/min. Eluted peptides were analyzed with a top-ten data-dependent acquisition method and identified using Proteome Discoverer (version 1.4; Thermo Scientific) coupled to MASCOT (version 2.4.1; Matrix Science, London). Orbitrap MS resolving power was set to 70,000 at a mass/charge ratio (m/z) of 200 for MS1 and 17,500 at a m/z of 200 for MS2. Tryptic peptides with up to two missed cleavages were searched against a *T. thermophila* protein database (www.ciliate.org) supplemented with common protein contaminant and enzyme sequences (27,046 sequences), with dynamic modifications for carbamidomethyl (C), oxidation (M), deamidation (N, Q), and peptide *N*-terminal cyclization to eliminate NH_3 (Q, carbamidomethyl-cys) or H_2O (E) (C, Cys; M, Met; Q, Gln; E, Glu). Precursor- and product-ion mass tolerances were set to 10 parts per million (or less) and 0.01 daltons, respectively.

Telomerase reconstitution and activity assays

Telomerase subcomplexes were assembled separately and then combined. The RNP catalytic core was assembled by TERT expression in RRL in the presence of purified bacterially expressed p65 and in vitro transcribed TER. Additional RRL synthesis reactions generated p50N30, Teb1, or coexpressed TEB complex subunits. One subunit was tagged with 3×FLAG (p50N30-F for TEB subunit function studies or TERT-F for TERT TEN domain mutational analysis). Subcomplexes were combined, bound to FLAG antibody resin, and assayed by primer extension as described previously (*29*), using radiolabeled deoxyguanosine triphosphate (dGTP) and primer $(GTTGGG)_3$ at a final concentration of 200 nM. Elongation time was 10 min unless otherwise indicated. Products were resolved by denaturing gel electrophoresis, including an end-labeled oligonucleotide added before product precipitation as a recovery control.

Preparation of DNA-LNA–bound telomerase linked by streptavidin and negative-stain EM

Three-repeat biotinylated LNA/DNA primer was purchased from Exiqon as biotin-5′-GTTGGGG-TTGGGGT$_L$T$_L$G$_L$T$_L$G$_L$G. Telomerase was purified as described (*29*) from TERT-FZZ strain up to the TEV protease cleavage step, during which 5 μM three-repeat primer and 50 μM dGTP were added to the immunoglobulin G resin. Immediately before the 10-min wash steps of the telomerase-bound flag resin, the resin was split, and one-half was incubated in 50 μl of wash buffer, whereas the other was incubated in 50 μl of wash buffer with 5 μM streptavidin, rotating end-over-end at 4°C for 30 min. Three wash steps were then performed as usual, after which the two resins were recombined and batch-eluted with 3×FLAG peptide, according to the original protocol. Negative-stain EM samples were prepared, imaged, and processed as previously described (*29*).

REFERENCES AND NOTES

1. E. H. Blackburn, C. W. Greider, J. W. Szostak, Telomeres and telomerase: The path from maize, *Tetrahymena* and yeast to human cancer and aging. *Nat. Med.* **12**, 1133–1138 (2006). doi: 10.1038/nm1006-1133; pmid: 17024208
2. E. H. Blackburn, K. Collins, Telomerase: An RNP enzyme synthesizes DNA. *Cold Spring Harbor Perspect. Biol.* **3**, a003558 (2011). doi: 10.1101/cshperspect.a003558; pmid: 20660025
3. M. Armanios, E. H. Blackburn, The telomere syndromes. *Nat. Rev. Genet.* **13**, 693–704 (2012). doi: 10.1038/nrg3246; pmid: 22965356
4. B. Bernardes de Jesus, M. A. Blasco, Telomerase at the intersection of cancer and aging. *Trends Genet.* **29**, 513–520 (2013). doi: 10.1016/j.tig.2013.06.007; pmid: 23876621
5. S. E. Artandi, R. A. DePinho, Telomeres and telomerase in cancer. *Carcinogenesis* **31**, 9–18 (2010). doi: 10.1093/carcin/bgp268; pmid: 19887512
6. J. Nandakumar, T. R. Cech, Finding the end: Recruitment of telomerase to telomeres. *Nat. Rev. Mol. Cell Biol.* **14**, 69–82 (2013). doi: 10.1038/nrm3505; pmid: 23299958
7. J. A. Stewart, M. F. Chaiken, F. Wang, C. M. Price, Maintaining the end: Roles of telomere proteins in end-protection, telomere replication and length regulation. *Mutat. Res.* **730**, 12–19 (2012). doi: 10.1016/j.mrfmmm.2011.08.011; pmid: 21945241
8. L. Y. Chen, J. Lingner, CST for the grand finale of telomere replication. *Nucleus* **4**, 277–282 (2013). doi: 10.4161/nucl.25701; pmid: 23851344
9. F. L. Zhong *et al.*, TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. *Cell* **150**, 481–494 (2012). doi: 10.1016/j.cell.2012.07.012; pmid: 22863003
10. A. N. Sexton, D. T. Youmans, K. Collins, Specificity requirements for human telomere protein interaction with telomerase holoenzyme. *J. Biol. Chem.* **287**, 34455–34464 (2012). doi: 10.1074/jbc.M112.394767; pmid: 22893708
11. J. Nandakumar *et al.*, The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature* **492**, 285–289 (2012). doi: 10.1038/nature11648; pmid: 23103865
12. H. Xin *et al.*, TPP1 is a homologue of ciliate TEBP-β and interacts with POT1 to recruit telomerase. *Nature* **445**, 559–562 (2007). doi: 10.1038/nature05469; pmid: 17237767
13. F. Wang *et al.*, The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**, 506–510 (2007). doi: 10.1038/nature05454; pmid: 17237768
14. T. Rao *et al.*, Structure of Est3 reveals a bimodal surface with differential roles in telomere replication. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 214–218 (2014). doi: 10.1073/pnas.1316453111; pmid: 24344315
15. J. M. Talley, D. C. DeZwaan, L. D. Maness, B. C. Freeman, K. L. Friedman, Stimulation of yeast telomerase activity by the ever shorter telomere 3 (Est3) subunit is dependent on direct interaction with the catalytic protein Est2. *J. Biol. Chem.* **286**, 26431–26439 (2011). doi: 10.1074/jbc.M111.228635; pmid: 21659533
16. A. Prakash, G. E. Borgstahl, The structure and function of replication protein A in DNA replication. *Subcell. Biochem.* **62**, 171–196 (2012). doi: 10.1007/978-94-007-4572-8_10; pmid: 22918586
17. K. A. Lewis, D. S. Wuttke, Telomerase and telomere-associated proteins: Structural insights into mechanism and evolution. *Structure* **20**, 28–39 (2012). doi: 10.1016/j.str.2011.10.017; pmid: 22244753
18. L. Y. Chen, S. Redon, J. Lingner, The human CST complex is a terminator of telomerase activity. *Nature* **488**, 540–544 (2012). doi: 10.1038/nature11269; pmid: 22763445
19. C. M. Price *et al.*, Evolution of CST function in telomere maintenance. *Cell Cycle* **9**, 3157–3165 (2010). doi: 10.4161/cc.9.16.12547; pmid: 20697207
20. P. Wu, H. Takai, T. de Lange, Telomeric 3′ overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. *Cell* **150**, 39–52 (2012). doi: 10.1016/j.cell.2012.05.026; pmid: 22748632
21. F. Wang *et al.*, Human CST has independent functions during telomere duplex replication and C-strand fill-in. *Cell Reports* **2**, 1096–1103 (2012). doi: 10.1016/j.celrep.2012.10.007; pmid: 23142664
22. N. F. Lue, E. Y. Yu, M. Lei, A popular engagement at the ends. *Nat. Struct. Mol. Biol.* **20**, 10–12 (2013). doi: 10.1038/nsmb.2483; pmid: 23288360
23. T. M. Tucey, V. Lundblad, Regulated assembly and disassembly of the yeast telomerase quaternary complex. *Genes Dev.* **28**, 2077–2089 (2014). doi: 10.1101/gad.246256.114; pmid: 25240060
24. R. J. Wellinger, V. A. Zakian, Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: Beginning to end. *Genetics* **191**, 1073–1105 (2012). doi: 10.1534/genetics.111.137851; pmid: 22879408
25. E. D. Egan, K. Collins, Biogenesis of telomerase ribonucleoproteins. *RNA* **18**, 1747–1759 (2012). doi: 10.1261/rna.034629.112; pmid: 22875809
26. J. D. Podlevsky, J. J. Chen, It all comes together at the ends: Telomerase structure, function, and biogenesis. *Mutat. Res.* **730**, 3–11 (2012). doi: 10.1016/j.mrfmmm.2011.11.002; pmid: 22093366
27. H. E. Upton, K. Hong, K. Collins, Direct single-stranded DNA binding by Teb1 mediates the recruitment of *Tetrahymena thermophila* telomerase to telomeres. *Mol. Cell. Biol.* **34**, 4200–4212 (2014). doi: 10.1128/MCB.01030-14; pmid: 25225329
28. B. Min, K. Collins, An RPA-related sequence-specific DNA-binding subunit of telomerase holoenzyme is required for elongation processivity and telomere maintenance. *Mol. Cell* **36**, 609–619 (2009). doi: 10.1016/j.molcel.2009.09.041; pmid: 19941821
29. J. Jiang *et al.*, The architecture of *Tetrahymena* telomerase holoenzyme. *Nature* **496**, 187–192 (2013). doi: 10.1038/nature12062; pmid: 23552895
30. C. A. Theimer, J. Feigon, Structure and function of telomerase RNA. *Curr. Opin. Struct. Biol.* **16**, 307–318 (2006). doi: 10.1016/j.sbi.2006.05.005; pmid: 16713250
31. A. J. Gillis, A. P. Schuller, E. Skordalakes, Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature* **455**, 633–637 (2008). doi: 10.1038/nature07283; pmid: 18758444
32. S. A. Jacobs, E. R. Podell, T. R. Cech, Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nat. Struct. Mol. Biol.* **13**, 218–225 (2006). doi: 10.1038/nsmb1054; pmid: 16462747
33. M. Singh *et al.*, Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. *Mol. Cell* **47**, 16–26 (2012). doi: 10.1016/j.molcel.2012.05.018; pmid: 22705372
34. M. D. Stone *et al.*, Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* **446**, 458–461 (2007). doi: 10.1038/nature05600; pmid: 17322903
35. N. K. Jacob, R. Lescasse, B. R. Linger, C. M. Price, *Tetrahymena* POT1a regulates telomere length and prevents activation of a cell cycle checkpoint. *Mol. Cell. Biol.* **27**, 1592–1601 (2007). doi: 10.1128/MCB.01975-06; pmid: 17158924
36. K. Hong *et al.*, *Tetrahymena* telomerase holoenzyme assembly, activation, and inhibition by domains of the p50 central hub. *Mol. Cell. Biol.* **33**, 3962–3971 (2013). doi: 10.1128/MCB.00792-13; pmid: 23918804
37. S. Rouda, E. Skordalakes, Structure of the RNA-binding domain of telomerase: Implications for RNA recognition and binding. *Structure* **15**, 1403–1412 (2007). doi: 10.1016/j.str.2007.09.007; pmid: 17997966
38. M. Mitchell, A. Gillis, M. Futahashi, H. Fujiwara, E. Skordalakes, Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat. Struct. Mol. Biol.* **17**, 513–518 (2010). doi: 10.1038/nsmb.1777; pmid: 20357774
39. N. F. Lue, Y. C. Lin, I. S. Mian, A conserved telomerase motif within the catalytic domain of telomerase reverse transcriptase is specifically required for repeat addition processivity. *Mol. Cell. Biol.* **23**, 8440–8449 (2003). doi: 10.1128/MCB.23.23.8440-8449.2003; pmid: 14612390
40. M. Xie, J. D. Podlevsky, X. Qi, C. J. Bley, J. J. Chen, A novel motif in telomerase reverse transcriptase regulates telomere repeat addition rate and processivity. *Nucleic Acids Res.* **38**, 1982–1996 (2010). doi: 10.1093/nar/gkp1198; pmid: 20044353
41. R. A. Wu, K. Collins, Human telomerase specialization for repeat synthesis by unique handling of primer-template duplex. *EMBO J.* **33**, 921–935 (2014). doi: 10.1002/embj.201387205; pmid: 24619002
42. B. Eckert, K. Collins, Roles of telomerase reverse transcriptase N-terminal domain in assembly and activity of *Tetrahymena* telomerase holoenzyme. *J. Biol. Chem.* **287**, 12805–12814 (2012). doi: 10.1074/jbc.M112.339853; pmid: 22367200
43. F. Qiao, T. R. Cech, Triple-helix structure in telomerase RNA contributes to catalysis. *Nat. Struct. Mol. Biol.* **15**, 634–640 (2008). doi: 10.1038/nsmb.1420; pmid: 18500353
44. D. D. Cash *et al.*, Pyrimidine motif triple helix in the *Kluyveromyces lactis* telomerase RNA pseudoknot is essential for function in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10970–10975 (2013). doi: 10.1073/pnas.1309590110; pmid: 23776224
45. C. A. Theimer, C. A. Blois, J. Feigon, Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell* **17**, 671–682 (2005). doi: 10.1016/j.molcel.2005.01.017; pmid: 15749017
46. M. Mihalusova, J. Y. Wu, X. Zhuang, Functional importance of telomerase pseudoknot revealed by single-molecule analysis. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20339–20344 (2011). doi: 10.1073/pnas.1017686108; pmid: 21571642
47. C. M. O'Connor, K. Collins, A novel RNA binding domain in *Tetrahymena* telomerase p65 initiates hierarchical assembly of telomerase holoenzyme. *Mol. Cell. Biol.* **26**, 2029–2036 (2006). doi: 10.1128/MCB.26.6.2029-2036.2006; pmid: 16507983
48. A. J. Berman, A. R. Gooding, T. R. Cech, *Tetrahymena* telomerase protein p65 induces conformational changes throughout telomerase RNA (TER) and rescues telomerase reverse transcriptase and TER assembly mutants. *Mol. Cell. Biol.* **30**, 4965–4976 (2010). doi: 10.1128/MCB.00827-10; pmid: 20713447
49. X. Qi *et al.*, RNA/DNA hybrid binding affinity determines telomerase template-translocation efficiency. *EMBO J.* **31**, 150–161 (2012). doi: 10.1038/emboj.2011.363; pmid: 21989387
50. Z. Zeng *et al.*, Structural basis for *Tetrahymena* telomerase processivity factor Teb1 binding to single-stranded telomeric-repeat DNA. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20357–20361 (2011). doi: 10.1073/pnas.1113624108; pmid: 22143754
51. E. Bochkareva, S. Korolev, S. P. Lees-Miller, A. Bochkarev, Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J.* **21**, 1855–1863 (2002). doi: 10.1093/emboj/21.7.1855; pmid: 11927569
52. S. F. Altschul *et al.*, Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997). doi: 10.1093/nar/25.17.3389; pmid: 9254694
53. C. Cole, J. D. Barber, G. J. Barton, The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36** (suppl. 2), W197–W201 (2008). doi: 10.1093/nar/gkn238; pmid: 18463136
54. B. Min, K. Collins, Multiple mechanisms for elongation processivity within the reconstituted *Tetrahymena* telomerase holoenzyme. *J. Biol. Chem.* **285**, 16434–16443 (2010). doi: 10.1074/jbc.M110.119172; pmid: 20363756
55. J. Xiong *et al.*, Transcriptome analysis of the model protozoan, *Tetrahymena thermophila*, using deep RNA sequencing. *PLOS ONE* **7**, e30630 (2012). doi: 10.1371/journal.pone.0030630; pmid: 22347391
56. A. N. Sexton *et al.*, Genetic and molecular identification of three human TPP1 functions in telomerase action: Recruitment, activation, and homeostasis set point regulation. *Genes Dev.* **28**, 1885–1899 (2014). pmid: 25128433

57. J. C. Schmidt, A. B. Dalby, T. R. Cech, Identification of human TERT elements necessary for telomerase recruitment to telomeres. *eLife* **3**, e03563 (2014). doi: 10.7554/eLife.03563
58. J. Sun *et al.*, Stn1-Ten1 is an Rpa2-Rpa3-like complex at telomeres. *Genes Dev.* **23**, 2900–2914 (2009). doi: 10.1101/gad.1851909; pmid: 20008938
59. C. Bryan, C. Rice, M. Harkisheimer, D. C. Schultz, E. Skordalakes, Structure of the human telomeric Stn1-Ten1 capping complex. *PLOS ONE* **8**, e66756 (2013). doi: 10.1371/journal.pone.0066756; pmid: 23826127
60. B. Vester, J. Wengel, LNA (locked nucleic acid): High-affinity targeting of complementary RNA and DNA. *Biochemistry* **43**, 13233–13241 (2004). doi: 10.1021/bi0485732; pmid: 15491130
61. J. Huang *et al.*, Structural basis for protein-RNA recognition in telomerase. *Nat. Struct. Mol. Biol.* **21**, 507–512 (2014). doi: 10.1038/nsmb.2819; pmid: 24793650
62. J. L. Chen, K. K. Opperman, C. W. Greider, A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. *Nucleic Acids Res.* **30**, 592–597 (2002). doi: 10.1093/nar/30.2.592; pmid: 11788723
63. C. J. Bley *et al.*, RNA-protein binding interface in the telomerase ribonucleoprotein. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20333–20338 (2011). doi: 10.1073/pnas.1100270108; pmid: 22123986
64. A. J. Zaug, E. R. Podell, T. R. Cech, Mutation in TERT separates processivity from anchor-site function. *Nat. Struct. Mol. Biol.* **15**, 870–872 (2008). doi: 10.1038/nsmb.1462; pmid: 18641663
65. B. R. Linger, G. B. Morin, C. M. Price, The Pot1a-associated proteins Tpt1 and Pat1 coordinate telomere protection and length regulation in *Tetrahymena*. *Mol. Biol. Cell* **22**, 4161–4170 (2011). doi: 10.1091/mbc.E11-06-0551; pmid: 21900503
66. C. Suloway *et al.*, Automated molecular microscopy: The new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005). doi: 10.1016/j.jsb.2005.03.010; pmid: 15890530
67. X. Li *et al.*, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013). doi: 10.1038/nmeth.2472; pmid: 23644547
68. J. A. Mindell, N. Grigorieff, Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003). doi: 10.1016/S1047-8477(03)00069-8; pmid: 12781660
69. J. B. Heymann, D. M. Belnap, Bsoft: Image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007). doi: 10.1016/j.jsb.2006.06.006; pmid: 17011211
70. N. R. Voss, C. K. Yoshioka, M. Radermacher, C. S. Potter, B. Carragher, DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205–213 (2009). doi: 10.1016/j.jsb.2009.01.004; pmid: 19374019
71. S. J. Ludtke, P. R. Baldwin, W. Chiu, EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999). doi: 10.1006/jsbi.1999.4174; pmid: 10600563
72. S. H. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012). doi: 10.1016/j.jsb.2012.09.006; pmid: 23000701
73. E. F. Pettersen *et al.*, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
74. A. Kucukelbir, F. J. Sigworth, H. D. Tagare, Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014). doi: 10.1038/nmeth.2727; pmid: 24213166
75. P. Emsley, K. Cowtan, *Coot*: Model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004). doi: 10.1107/S0907444904019158; pmid: 15572765
76. P. Chacón, W. Wriggers, Multi-resolution contour-based fitting of macromolecular structures. *J. Mol. Biol.* **317**, 375–384 (2002). doi: 10.1006/jmbi.2002.5438; pmid: 11922671
77. The PyMOL Molecular Graphics System, Version 1.7.4, Schrödinger, LLC.
78. A. F. Moon, G. A. Mueller, X. Zhong, L. C. Pedersen, A synergistic approach to protein crystallization: Combination of a fixed-arm carrier with surface entropy reduction. *Protein Sci.* **19**, 901–913 (2010). pmid: 20196072
79. K. L. Witkin, K. Collins, Holoenzyme proteins required for the physiological assembly and activity of telomerase. *Genes Dev.* **18**, 1107–1118 (2004). doi: 10.1101/gad.1201704; pmid: 15131081
80. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66**, 133–144 (2010). doi: 10.1107/S0907444909047374; pmid: 20124693
81. A. J. McCoy *et al.*, *Phaser* crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007). doi: 10.1107/S0021889807021206; pmid: 19461840
82. P. V. Afonine *et al.*, Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallogr. D* **68**, 352–367 (2012). doi: 10.1107/S0907444912001308; pmid: 22505256
83. N. K. Kim *et al.*, Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. *J. Mol. Biol.* **384**, 1249–1261 (2008). doi: 10.1016/j.jmb.2008.10.005; pmid: 18950640
84. J. Erde, R. R. Loo, J. A. Loo, Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* **13**, 1885–1895 (2014). doi: 10.1021/pr4010019; pmid: 24552128

ACKNOWLEDGMENTS

This work was supported in part by grants from the NIH (GM048123) and NSF (MCB1022379) to J.F., NIH grant GM071940 to Z.H.Z., NIH grant GM103479 to J.A.L., NIH grant R01GM054198 to K.C., American Heart Association postdoctoral fellowship 14POST18870059 to J.J., NIH Ruth L. Kirschstein National Research Service Award (NRSA) postdoctoral fellowship GM101874 to E.J.M., NIH Ruth L. Kirschstein NRSA predoctoral training grant fellowship under GM007185 to H.C. and R.O.J., and NSF Graduate Research Fellowship under DGE-1106400 to H.E.U. We acknowledge support for the NMR core, crystallization core, and x-ray core facilities by DOE grant DE-FC0302ER63421 and the Electron Imaging Center for NanoMachines by NIH grants S10RR23057 and S10OD018111 and NSF DBI-133813. The APS of Argonne National Laboratory is supported by NIH grants P41 RR015301 and P41 GM103403. Use of the APS is supported by DOE under contract DE-AC02-06CH11357. Additional support for EM time was provided to J.F. by National Center for Advancing Translational Sciences UCLA Clinical and Translational Science Institute grant UL1TR000124. We thank M. Sawaya and M. Collazo for help with crystallography and M. Capel, K. Rajashankar, N. Sukumar, F. Murphy, and I. Kourinov of NECAT. The 9.4 and 8.9 Å cryo-EM density maps of *Tetrahymena* telomerase holoenzyme have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-6442 and EMD-6443, respectively. Coordinates and structure factors for the x-ray crystal structures of p19 and p45C have been deposited in the PDB with accession codes 5DFM and 5DFN, respectively. Coordinates and restraints for the solution NMR structure of the TER PK and chemical shifts have been deposited in the PDB (accession code 2N6Q) and Biological Magnetic Resonance Bank (accession code 25777), respectively. Nucleotide sequence data are available in the third-party annotation (TPA) section of the DNA Data Bank of Japan, European Molecular Biology Laboratory, and GenBank databases under the accession numbers TPA: BK009378 and BK009379.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/350/6260/aab4070/suppl/DC1
Figs. S1 to S10
Tables S1 to S4
References (*85*–*88*)

23 April 2015; accepted 1 September 2015
Published online 15 October 2015
10.1126/science.aab4070

Table S2A. Telomerase proteins identified from TERT-FZZ telomerase preparations by LC-MS/MS

			Replicate 1				Replicate 2				Replicate 3			
TTHERM ID	Protein	MW (kDa)	Uni. Pep[a]	Sig. Pep[b]	% Cov[c]	Score[d]	Uni. Pep[a]	Sig. Pep[b]	% Cov[c]	Score[d]	Uni. Pep[a]	Sig. Pep[b]	% Cov[c]	Score[d]
00112560	TERT	134.1	30	22	22	393	84	70	53	2440	37	22	30	370
000318539	p65	64.6	22	17	41	300	49	42	62	1447	36	20	50	358
01049190	p50	50.3	14	8	29	276	27	26	45	2539	11	7	29	147
0059040	p75	74.3	21	10	24	150	42	31	50	955	28	15	40	437
0083360	p45	43.8	17	11	34	201	30	27	59	1475	14	7	33	198
00658760	p19	19.5	7	3	38	88	9	8	50	223	1	1	8	39
00218760	Teb1	82.6	12	7	15	148	18	15	21	1479	5	2	8	251
00113129	Teb2	31.1	9	7	42	414	15	13	49	858	8	5	35	334
00439320	Teb3	14.0	8	7	44	406	9	8	51	507	5	3	40	409

Table S2B. Telomerase proteins identified from ZZF-p50 telomerase preparations by LC-MS/MS

			Replicate 1				Replicate 2			
TTHERM ID	Protein	MW (kDa)	Uni. Pep[a]	Sig. Pep[b]	% Cov[c]	Score[d]	Uni. Pep[a]	Sig. Pep[b]	% Cov[c]	Score[d]
00112560	TERT	134.1	15	3	12	40	40	33	31	616
000318539	p65	64.6	7	3	12	74	36	29	47	828
01049190	p50	50.3	16	8	38	118	30	24	60	1218
0059040	p75	74.3	24	15	33	256	25	19	37	334
0083360	p45	43.8	12	6	27	128	20	15	50	746
00658760	p19	19.5	7	4	33	79	9	7	44	241
00218760	Teb1	82.6	0	0	-	-	0	0	-	-
00113129	Teb2	31.1	0	0	-	-	1	1	-	-
00439320	Teb3	14.0	0	0	-	-	0	0	-	-

[a]Number of unique peptide sequences identified at ≤ 10 ppm mass accuracy
[b]Number of unique peptide sequences identified with E-value < 0.05 (>95% confidence that the match was not random)
[c]Percentage of theoretical sequence covered by identified peptides
[d]Mascot score employing all peptides with E-value < 0.05

APPENDIX 2 : CHARACTERIZED POST-TRANSLATIONAL PROTEOLYSIS THROUGH DEVELOPMENT OF NEW N-TERMINAL LABELING TECHNIQUES

INTRODUCTION

Protein post-translational degradation (i.e., the “degradome”) is an important process in a variety of diseases, such as cancer [1]. Current study of the degradome is often limited by the speed and capabilities of sequencing techniques. In particular, the identification of products from protein degradation can be enhanced by targeted analysis of the N-terminal sequence, but the N-terminus of the protein is frequently not detectable or difficult to characterize with current high-throughput sequencing techniques. Defining the amino termini of proteins at the proteome level is of paramount importance as it frequently determines cellular localization, activity and turn-over rate [2]. Furthermore, information on the N-terminus of proteins would aid tremendously in genome annotation and potentially disease biomarker identification.

In addition, we are using these techniques to better understanding proteolysis-induced degradation of proteins, such as astrocyte-specific glial fibrillary acidic protein (GFAP), found in cerebral spinal cord fluid (CSF) from patients with traumatic brain injury (TBI) [3]. GFAP is a cytoskeletal protein specifically expressed in astrocytes. It is believed that astrocyte injury may induce cleavage of cellular proteins through activation of caspases and other proteolytic enzymes.

PROCEDURE AND RESULTS

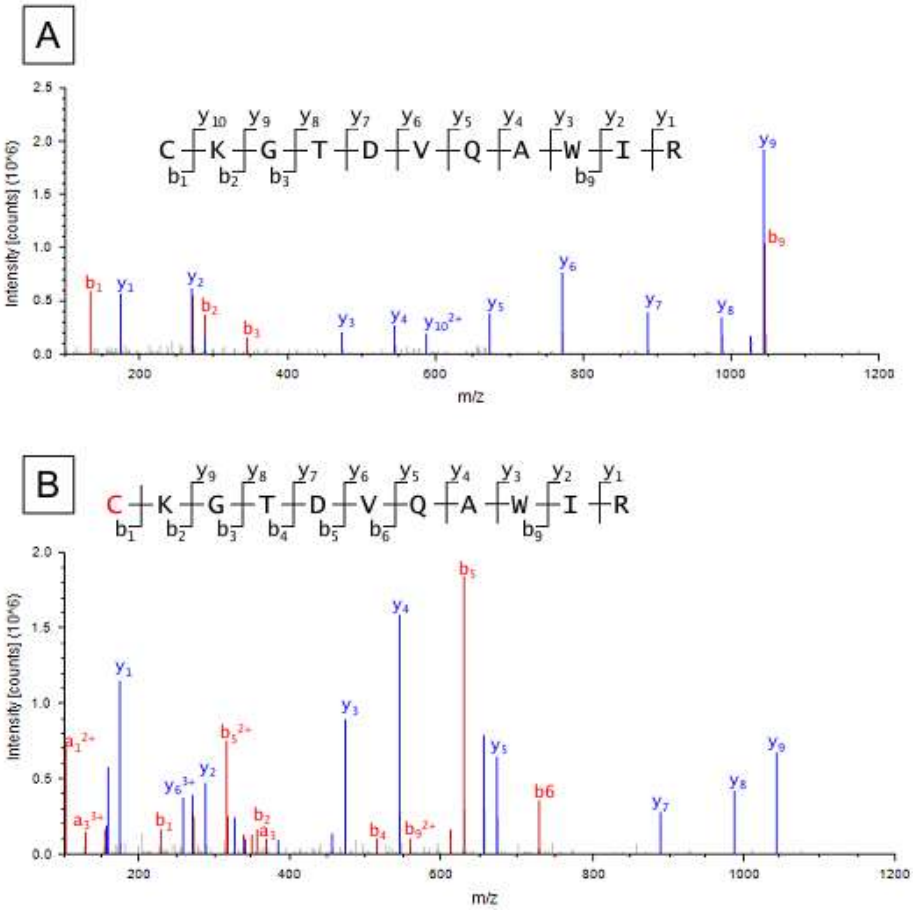
Two strategies with MS detection are utilized to investigate N-terminal labeling for biomarker characterization. The first is reductive glutaraldehydation in which glutaraldehyde is used to selectively modify amino groups on free N-termini [4]. This type of label also modifies free lysine residues. Here we compare it to a second labeling strategy that involves the addition of a thiol group at the N-terminus [5]. The thiol is used to bind the N-terminal peptide to a resin and pull it out of solution, thereby selecting only the peptides of interest. Together these techniques can be used in biomarker discovery projects, as well as serving as an important tool to address questions in biology [6, 7]. After labeling, samples were digested with trypsin and run on a Q-Exactive Orbitrap mass spectrometer. Undigested samples and affinity purified samples were also run on a Voyager-DE-STR MALDI-TOF instrument.

Proteins are often post-translationally modified as a result of proteolysis, yielding truncated products on both the N- and C-termini of the original protein. For certain disease models, the identification of protein truncation products may be a viable and selective biomarker. By labeling the N-terminus of such proteins, these protein fragments are more detectable via MS. A glutaraldehyde N-terminal tag aids MS detection by improving ionization of the peptide (and thereby increasing its signal in the MS) while the addition of a thiol tag allows for the isolation of N-terminal fragments (thereby removing background noise and increasing abundance of relevant peptides). Preliminary data, utilizing both MALDI-TOF and hybrid quadrupole-Orbitrap (Q-Exactive) instruments, have demonstrated efficient labeling of several test proteins and protein fragments.

Additionally, short (10-15 aa) peptide sequences have been successfully labeled and analyzed. Furthermore, glutaraldehyde labeled peptides have more N-terminal fragments, especially a- and b-ions, and an increased number of multiply charged ions (Appendix Figure 2-1).

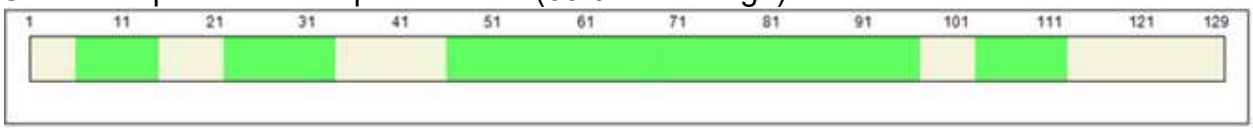
The addition of a thiol tag has also been shown to improve sequence coverage of the N-terminus of proteins and protein fragments as it allows for the isolation of N-terminal peptides (thereby removing background noise and increasing abundance of relevant peptides). We conclude that both reductive glutaraldehydation and reactive thiol labeling techniques increase the rate of detection of the N-terminus of proteins and potentially proteolysis sites, but do not improve sequence coverage in other areas (Appendix Figure 2-2). Currently, the same techniques are being applied to biology projects to aid in genome annotation and other projects related to human diseases.

FIGURES

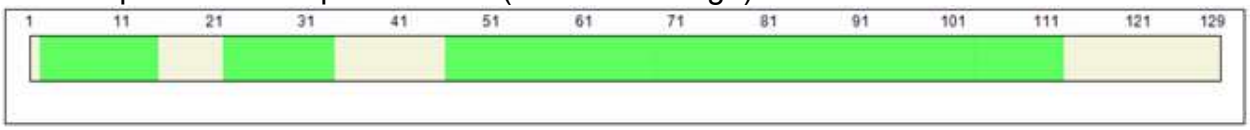


Appendix Figure 2-1: LC-MS/MS of a typical peptide modified with a N-terminal glutaraldehyde. A peptide identified in both unlabeled [A] and labeled [B] protein mixtures. The N-terminal residues in the labeled sample have been modified with a glutaraldehyde. This causes an increased abundance of N-terminal fragments, including a-ions, increased number of multiply charged ions and, in some cases, increased peak intensity. The peptide shown is from the C-terminus of lysozyme.

Unlabeled protein in complex mixture (63.6% coverage)



Labeled protein in complex mixture (71.3% coverage)



Appendix Figure 2-2: Improvement in sequence coverage due to N-terminal labeling. Both labeling strategies typically only improve sequence coverage at N-termini. Lysozyme detected in complex sample mixtures (labeled and unlabeled) is shown here. Average sequence coverage from 3 independent runs.

REFERENCES

1. López-Otín, C. and Overall, C.M., 2002. Protease degradomics: a new challenge for proteomics. *Nature Reviews Molecular Cell Biology*, *3*(7), p.509.

2. Meinnel, T., Serero, A. and Giglione, C., 2006. Impact of the N-terminal amino acid on targeted protein degradation. *Biological Chemistry*, *387*(7), pp.839-851.

3. Shen, S., Loo, R.R.O., Wanner, I.B. and Loo, J.A., 2014. Addressing the needs of traumatic brain injury with clinical proteomics. *Clinical Proteomics*, *11*(1), p.11.

4. Russo, A., Chandramouli, N., Zhang, L. and Deng, H., 2008. Reductive glutaraldehydation of amine groups for identification of protein N-termini. *Journal of Proteome Research*, *7*(9), pp.4178-4182.

5. Kim, J.S., Dai, Z., Aryal, U.K., Moore, R.J., Camp, D.G., Baker, S.E., Smith, R.D. and Qian, W.J., 2013. Resin-assisted enrichment of N-terminal peptides for characterizing proteolytic processing. *Analytical Chemistry*, *85*(14), pp.6826-6832.

6. Han, H. and McLuckey, S.A., 2009. Selective covalent bond formation in polypeptide ions via gas-phase ion/ion reaction chemistry. *Journal of the American Chemical Society*, *131*(36), pp.12884-12885.

7. Kleifeld, O., Doucet, A., Prudova, A., auf dem Keller, U., Gioia, M., Kizhakkedathu, J.N. and Overall, C.M., 2011. Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nature Protocols*, *6*(10), p.1578.

APPENDIX 3 : IMPROVEMENTS IN LC-ESI-MS OF LARGE INTACT PROTEINS BY ADDITION OF VARIOUS SUPERCHARGING AGENTS AS MOBILE PHASE ADDITIVES

INTRODUCTION

The use of trifluoroacetic acid (TFA) as an ion pairing agent provides superior chromatographic resolving power, but it can reduce ESI-MS sensitivity for on-line LC-MS. However, we have found that addition of some supercharging reagents to the mobile phase can rescue the loss of sensitivity from using TFA as an ion pairing agent [1]. Chromatographic performance for protein and peptide separations can be maintained, and LC-MS sensitivity is increased.

Supercharging in general has many advantages [2-4]. Multiple charging of molecules extends the effective mass range of the mass analyzer in direct proportion to the number of charges per ion. Furthermore, multiply charged ions with lower mass-to-charge (m/z) values are easier to detect than singly charged ions of the same mass. Putting more charges on proteins and peptides not only reduces the m/z requirement of the mass analyzer but also results in more efficient dissociation and sequence coverage in tandem (MS/MS) mode. In this study, we also investigate the role of Brønsted basicity (solution phase basicity) on the charge state distributions of proteins eluted with 0.1% TFA or 0.1% formic acid mobile phases, and to see whether adding particular supercharging reagents will rescue the TFA-suppressed ionization.

PROCEDURES

For LC-MS experiments, protein solutions (0.6-17 μM) were injected into an Agilent C3 column (pore size 300 Å) with dimensions 2.1 mm x 150 mm. LC-MS analysis was performed on an Agilent 1200 LC-System connected on-line with an Agilent 6220 ESI-TOF mass spectrometer (Santa Clara, CA). All mass spectra were acquired in positive (+ESI) mode. Solvent A was either 0.1% TFA or 0.1% formic acid in water. Solvent B was 90% acetonitrile (ACN) containing either 0.1% TFA or 0.1% formic acid and water. The flow rate was 200 μl/min. Supercharging reagents (0.1%, v/v) were added to both aqueous and organic components of the mobile phase containing either 0.1% TFA or 0.1% formic acid.

Pyrazole experiments were performed on a Waters Synapt G2-Si equipped with nanoESI. Myoglobin (10μM) mixed with varying concentrations (w/v) of substituted pyrazoles was prepared in 1:1 CH_3CN:H_2O (v:v) with TFA or formic acid. All samples were sprayed with metalcoated borosilicate capillaries (Au/Pd-coated, 1μm i.d.) with flow rates between 10-40nL/min.

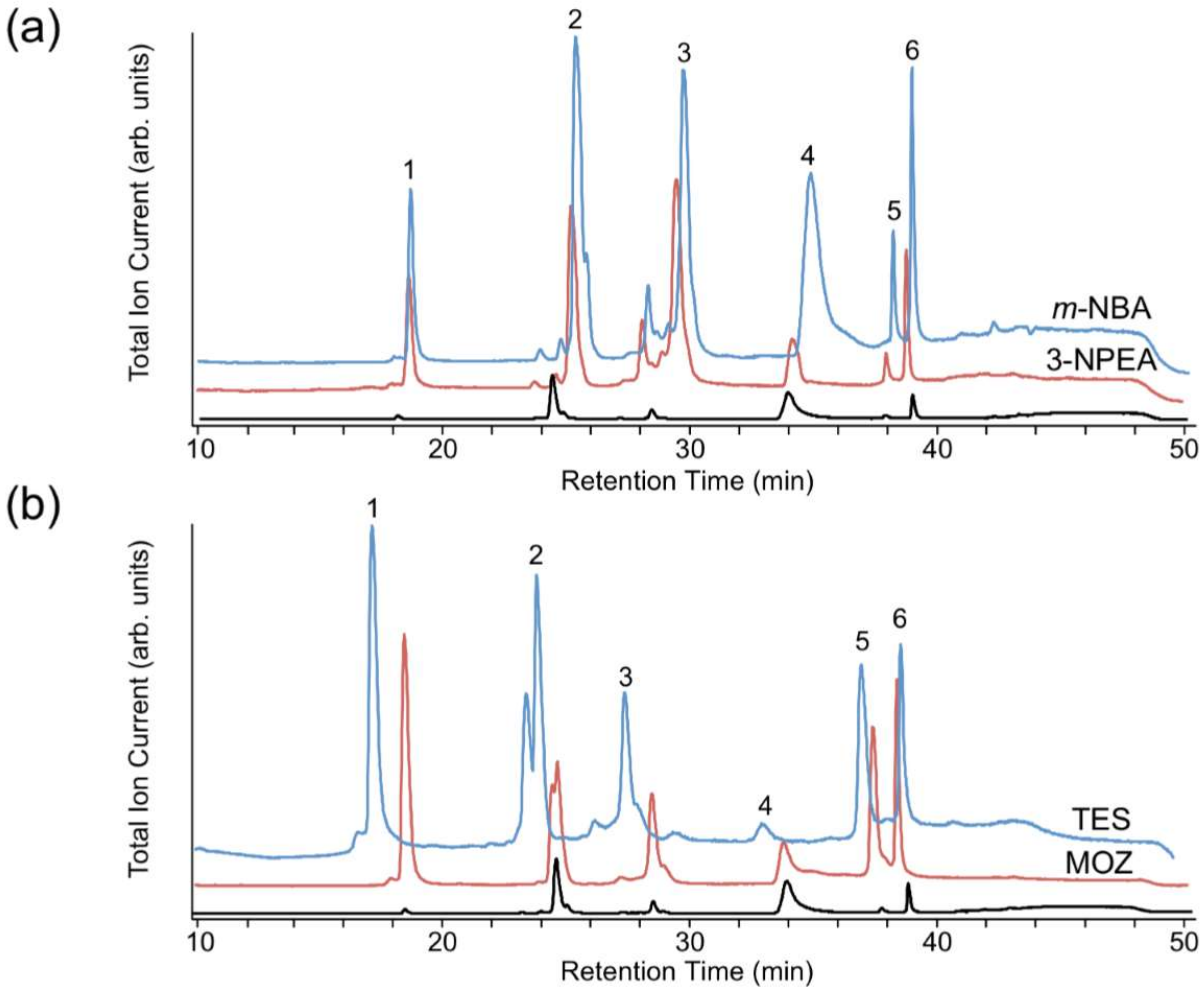
RESULTS

Adding supercharging reagents to the LC mobile phase can increase the chromatographic plate number and improve the peak shape of proteins such as 8 kDa ubiquitin, 13-14 kDa lysozyme and ribonuclease A, and 29 kDa carbonic anhydride. Previously, we reported on the success of using supercharging reagents such as

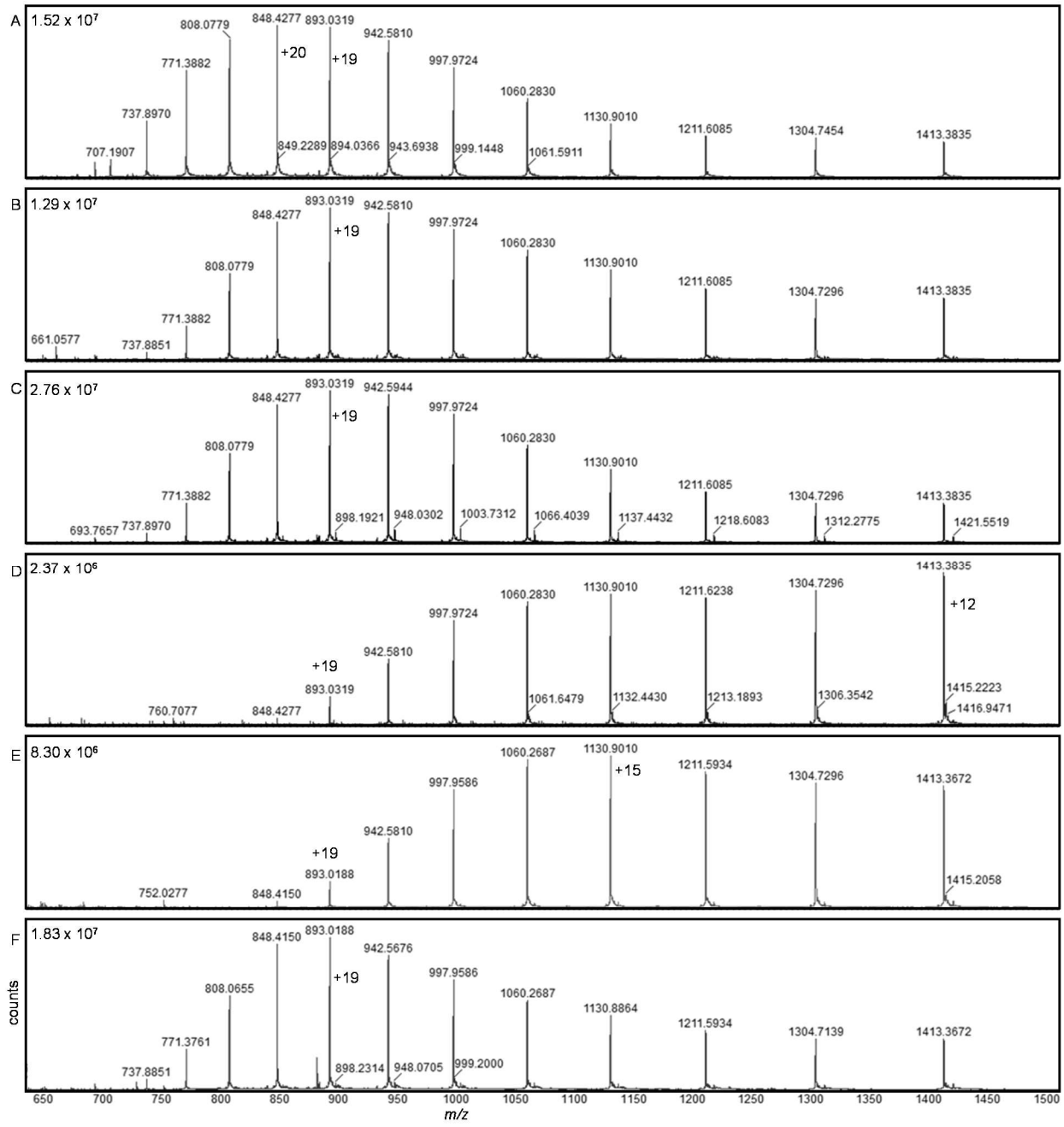
propylene carbonate (PC), 3-nitrophenethylalcohol (NPEA), 2-methyl-2-oxazoline (MOZ), methyl methanesulfonate (MMS), N,N,N'N'tetraethylsulfamide (TES), and m-nitro benzyl alcohol (m-NBA) for protein LC-MS with 0.1% TFA in the mobile phase (relative to using formic acid). The increase in protein charge was observed in mobile phases containing both formic acid and TFA, but there was a marked increase in the observed total ion chromatogram (TIC) signal in the presence of supercharging reagents, especially in the presence of TFA. The suppression of ionization generally observed in the presence of TFA was partially rescued in the presence of supercharging reagents, with the added benefit of improved separation efficiency, or plate number (N), and a lowering of the limit of detection for proteins (Appendix Figure 3-1).

We are probing the mechanism of supercharging [4]. If charge allocation in + ion mode ESI is viewed as a competition between depositing protons on solvent clusters versus analyte, then additives that reduce the ability of solvent and other solution components to accept protons should increase analyte charge. The idea predicts that adding involatile, very weak Brønsted bases to ESI solutions will increase charge. Previous results [1] followed those predictions. Pyrazole compounds of varying basicity (5-nitropyrazole, 1-nitropyrazole, 4-nitropyrazole, phenylpyrazole, and pyrazole) were compared for their ability to supercharge in the presence of either 0.1% trifluoroacetic acid (TFA) or 0.1% formic acid. The substituted pyrazoles that are less basic than H_2O ($pKBH^+$ below ca. -1.7) increase analyte charge, consistent with the idea that increased analyte charging arises from less charge lost to solvent because the supercharger disrupts solvent H-bonding (reduces "solvent" basicity). However, it is not yet apparent if substituted pyrazoles will rescue ion suppression caused by TFA.

FIGURES



Appendix Figure 3-1: LC-MS of protein mixture containing (1) 17 μM RNase A, (2) 13 μM ubiquitin, (3) 7.7 μM lysozyme, (4) 15 μM BSA, (5) 6.5 μM myoglobin, and (6) 7.6 μM carbonic anhydrase with 0.1% TFA and (a) 0.1% m-NBA and 0.1% 3-NPEA, and (b) 0.1% MOZ and 0.1% TES. No supercharging agent was added to the measurements represented by the bottom trace in both (a) and (b).



Appendix Figure 3-2: Spectra of myoglobin with related pyrazole compounds: samples contain 10 µM myoglobin in 50% ACN with 0.1% TFA supplemented with: [A] 0.2% (w/v) 5-nitropyrazole (pK_{BH+}= -4.66), [B] 0.2% (w/v) 1-nitropyrazole (pK_{BH+}= -4.21), [C] 0.2% (w/v) 4-nitropyrazole (pK_{BH+}= -2.0), [D] 0.2% (w/v) Phenylpyrazole (pK_{BH+}= 0.43), [E] 0.2% (w/v) Pyrazole (pK_{BH+}= 2.5), [F] Control.

REFERENCES

1. Nshanian, M., Lakshmanan, R., Chen, H., Loo, R.R.O. and Loo, J.A., 2018. Enhancing sensitivity of liquid chromatography–mass spectrometry of peptides and proteins using supercharging agents. *International Journal of Mass Spectrometry*, *427*, pp.157-164.

2. Lomeli, S.H., Peng, I.X., Yin, S., Loo, R.R.O. and Loo, J.A., 2010. New reagents for increasing ESI multiple charging of proteins and protein complexes. *Journal of the American Society for Mass Spectrometry*, *21*(1), pp.127-131.

3. Lomeli, S.H., Yin, S., Loo, R.R.O. and Loo, J.A., 2009. Increasing charge while preserving noncovalent protein complexes for ESI-MS. *Journal of the American Society for Mass Spectrometry*, *20*(4), pp.593-596.

4. Loo, R.R.O., Lakshmanan, R. and Loo, J.A., 2014. What protein charging (and supercharging) reveal about the mechanism of electrospray ionization. *Journal of the American Society for Mass Spectrometry*, *25*(10), pp.1675-1693.