

Novel Proteomic Approaches to Characterize Endogenous Membrane Proteins

By
Kyle Brown

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
(Chemistry)

At the
UNIVERSITY OF WISCONSIN-MADISON
2020

Date of final oral examination: 08/12/2020

The dissertation is approved by the following members of the Final Oral Committee:

Ying Ge, Professor, Cell and Regenerative Biology and Chemistry

Sam Gellman, Professor, Chemistry

Lingjun Li, Professor, Pharmacy and Chemistry

Song Jin, Professor, Chemistry

© Copyright by Kyle Brown 2020

All Rights Reserved

To my greatest role models and biggest supporters, Alex and Renee Brown

ACKNOWLEDGEMENTS

I am extremely thankful for the guidance and support of my mentor Professor Ying Ge. Ying's enthusiasm and passion have inspired me throughout my time in her group. Her trust and patience to let me conduct the research I am passionate about contributed greatly to my successes. Furthermore, the mentorship and opportunities Ying has given to me have enabled me to grow as a scientist. I would also like to thank my committee members, Professor Sam Gellman, Professor Lingjun Li, and Professor Song Jin who have provided encouragement and valuable advice on both my research and my future career. Dr. Arrietta Clauss, thank you for the helpful guidance and for always having an open door. Sue Martin, thank you for all the help with travel, planning coffee hours, and so much more.

Graduate school would not have been possible without the early support of my undergraduate mentors, Professor Lindsley and Professor Harden. Working in their groups inspired my early love for research. I appreciate their unwavering support, which has continued even in graduate school.

It has been a great privilege to get to work with so many incredible past and present members of the Ge group including Albert Chen, Tania Guardado, Bifan Chen, Leekyoung Hwang, Liming Wei, Ruixiang Sun, Yu Liang, Ziqing Lin, Yanlong Zhu, Zach Gregorich, Vege Cai, Yutong Jin, Trisha Tucholski, Abe Wu, Sam Knott, Tim Tiambeng, Stanford Mitchell, David Roberts, Elizabeth Bayne, Austin Carr, Eli Larson, Jake Melby, Ben Wancewicz, Sean McIlwain, Kent Wenger, Hannah Karp, Molly Wetzels, and Tim Aballo. The discussions, both scientific and extracurricular, have been a great source of learning and laughs. In particular, I want to thank Tania Guardado for “showing me the ropes” and teaching me that the simplest solution to a problem is often the best. This mentality has greatly aided me. Bifan Chen was a great deskmate and a positive role model who helped me to focus on the “bigger picture” and logically troubleshoot problems. Most importantly, Bifan is a great friend! David

R, you have been a great deskmate and your excitement for science inspires me. I cannot wait to keep working on exciting projects with you and to see your development as a scientist. I have gained a tremendous amount of knowledge from Dr. Andrew Alpert. He is always available to talk about challenges and new ideas regarding chromatography. Christian, Lucas, Morgan, and Andreas, I think you all have taught me more than I taught you. Your futures are so bright and I am so proud of what you have and will accomplish. It is been such a privilege to be your mentor. Thank you for all the hard work.

I am so thankful for my friends who made my graduate school experience some of the best times in my life. To Camille, Jessi, Margy, Leah, and Trisha: thank you for countless memories of tailgating, after-work beers, trips to Devil's lake, and so much more. I never would have imagined that I would have so much fun in graduate school or met life-long friends (or as Cam would say "my people"). To my roommate(s) of five years, Camille (also Ellie and Jack Bishop-Brown), thank you for being a supportive friend and such a fun person to live with. I will miss hanging out, watching hockey (I will always remember our Predator/Hawks games), grilling, and all the other great times we shared. I will miss all of my friends dearly, but cannot wait to see all of the amazing things they will accomplish moving forwards.

Finally, the unconditional love and support my family has been a rock throughout my life and particularly in graduate school. I am extremely fortunate to have met my girlfriend (and the best center mid buddy), Andi, while in graduate school. Over the years, we have made so many incredible memories (Chick-fil-a anniversary dinner, trips to Boise and Colorado, adventures all around Madison) and grown together. I am looking forward to our future together and am so grateful to always have you by my side. I would also like to thank Andi's family (especially Jenny and Steve) for providing such a warm, welcoming environment and a second home during my time in Wisconsin. Christmas tree hunting, Thanksgiving, and other

gatherings have been a blast. To my little sister, Lindsey, it has been incredible watching you grow up and handle adversity. I am so proud of what you have already accomplished. Thanks for being a great sister. I want to thank my grandparents (David and Jan) for their constant love and thoughtfulness. Finally, Renee and Alex (mom and dad) are my greatest role models. They inspired my love for science without pushing me towards it. Words cannot express how much I have missed you both (and Lindsey) while being away for school and how much your love and support have meant over the years. Thank you!

TABLE OF CONTENTS

TABLE OF CONTENTS	v
ABSTRACT	viii
LIST OF FIGURES	x
CHAPTER 1- Top-down Proteomics: Innovations, Challenges, and Applications in	
Basic and Clinical Research	1
1.1 Introduction	2
1.2 Methods for Top-down Proteomics	3
1.3 Application in Disease Research	14
1.4 Five-year Outlook	15
CHAPTER 2- A Photocleavable Surfactant for Top-down Proteomics	17
2.1 Abstract	18
2.2 Development and Characterization of a Photocleavable Surfactant for Top-down Proteomics	18
2.3 Supplemental Information	26
2.4 Acknowledgement	68
CHAPTER 3- High-throughput Proteomics Enabled by a Photocleavable Surfactant	69
3.1 Abstract	70
3.2 Introduction	70
3.3 Azo-enabled High-throughput Sample Preparation	72

3.4	Azo-enabled Membrane Proteomics	77
3.5	Azo for Integrated Top-down and Bottom-up Proteomics	80
3.6	Experimental Procedures and Supplemental Information	81
3.7	Acknowledgments	101
CHAPTER 4-	Photocleavable Surfactant-enabled Extracellular Matrix Proteomics	103
4.1	Abstract	104
4.2	Introduction	104
4.3.	Development of a High-throughput ECM Proteomic Approach	107
4.4.	Conclusion	113
4.5.	Supplemental Information	114
CHAPTER 5-	Novel Enrichment and Separation Tools for Top-down Membrane Proteomics	127
5.1	Abstract	128
5.2	Introduction	128
5.3	Experimental Section	131
5.4	Results and Discussion	136
5.5	Conclusion	149
5.6	Supplementary Information	150
CHAPTER 6-	Summary and Future Direction	160

CHAPTER 7- Thesis Chapter for Non-experts	163
7.1 Introduction	164
7.2 What are the challenges?	167
7.3 What is my research about?	169
7.4 What is our approach?	169
7.5 How does this help disease research?	172
7.6 What is the next step?	173
REFERENCES	174

ABSTRACT

Biological information flows as DNA is transcribed into mRNA and then translated into proteins. However, sequence variations from mutations and alternative splicing events combined with post-translational modification (PTMs) of proteins result in a diversity of protein forms (referred to as proteoforms) that can arise from a single gene. Mass spectrometry (MS)-based proteomics provides an unprecedented opportunity to understand the role of proteoforms in health and disease; however, many challenges remain. For example, despite their importance as drug targets (>50% of current drugs), membrane proteins are traditionally underrepresented using MS-based proteomics because of their lower expression level, hydrophobicity, and lack of established protocols. To address these challenges, I developed a novel photocleavable surfactant, Azo, which can effectively solubilize proteins and is compatible with MS analysis (**Chapter 2**). We demonstrated Azo-aided top-down proteomics (the analysis of intact proteins by MS) enabled the solubilization of important membrane proteins from biological samples, including heart tissues, for comprehensive characterization of their proteoforms. Moreover, Azo is simple to synthesize and can be used as a surfactant in polyacrylamide gel electrophoresis. We next incorporated the surfactant technology to facilitate high-throughput bottom-up proteomics (the analysis of digested proteins by MS) for more extensive proteome coverage and protein expression quantification. Furthermore, we established simple, high-throughput membrane and extracellular matrix proteomic methods using Azo (**Chapter 3-4**). Combining Azo-aided bottom-up and top-down proteomics, we established a powerful integrated strategy to extensively characterize proteins from biological samples. Finally, a novel membrane protein enrichment and multidimensional liquid chromatography separation strategy was developed to further expand the scope of MS-based top-down proteomics for characterizing the membrane proteoform landscape (**Chapter 4**). Future

development and applications of MS-based approaches for the characterization of membrane proteoforms are discussed in **Chapter 5**.

LIST OF FIGURES AND TABLES

Figure 1.1. Mass spectrometry-based top-down proteomics	3
Figure 1.2. Illustration of the different components of top-down proteomics.	16
Figure 2.1. Synthesis and characterization of a photo-cleavable anionic surfactant, sodium 4-hexylphenylazosulfonate (Azo).	20
Figure 2.2. Photocleavable Azo-enabled top-down membrane proteomics.	24
Figure 2.3. Negative ion ESI-FTICR mass spectrum of Azo.	43
Figure 2.4. ¹ H-NMR spectrum of Azo.	44
Figure 2.5. Polyacrylamide gel electrophoresis (PAGE) using Azo and SDS.	45
Figure 2.6. Effects of solvent conditions on Azo degradation kinetics.	46
Figure 2.7. Evaluation of surfactant-aided extraction of cardiac membrane proteins using western blot analysis.	47
Figure 2.8. Effect of UV exposure on protein mass spectrum.	48
Figure 2.9. The effect of the addition of reducing agents during Azo degradation.	49
Figure 2.10. The addition of methionine during Azo degradation alleviated oxidative modification.	50
Figure 2.11. Effect of Azo on relative quantitation by LC-MS using HEPES Cardiac Extract.	51
Figure 2.12. Evaluation of surfactant-aided extraction of cardiac tissue using SDS-PAGE and protein assay.	52
Figure 2.13. Direct MS comparison of common surfactants.	53
Figure 2.14. Comparison of surfactant effect on LC-MS analysis of cardiac tissue lysate.	54

Figure 2.15. Photocleavable Azo-enhanced top-down proteomics.	55
Figure 2.16. Comparison of cardiac tissue extraction using MS-compatible NH_4HCO_3 with or without Azo in a single RPLC/MS run.	56
Figure 2.17. Representative membrane proteins identified from human embryonic kidney (HEK) Cells.	57
Figure 2.18. Identification of succinate dehydrogenase cytochrome b560 subunit by top-down proteomics enabled by Azo.	58
Figure 2.19. Localization of palmitoylation (palm) on phospholamban (PLN) by online LC-MS/MS with CID.	59
Figure 2.20. Localization of trimethylation on ATP synthase c by online LC-MS/MS with CID.	60
Table 2.1. Protein solubility screening of a library of photo-cleavable surfactants.	62
Figure 3.1. Scheme for azo-enable high-throughput top-down and bottom-up proteomics.	72
Figure 3.2. Enhanced enzymatic digestion and MS analysis using a photocleavable surfactant, Azo.	73
Figure 3.3. Rapid and reproducible extraction, enzymatic digestion, and LC-MS/MS analysis of proteins from human embryonic kidney 293T (HEK293T) cells.	76
Figure 3.4. Analysis of integral membrane protein enabled by Azo.	78
Figure 3.5. Azo-enabled top-down proteomics analysis of cardiac tissue.	80
Figure 3.6. A Direct Comparison of the MS-compatibility of Azo and RapiGest™ (RG, also known as ALS).	90
Figure 3.7. Azo degradation evaluated by MS.	91

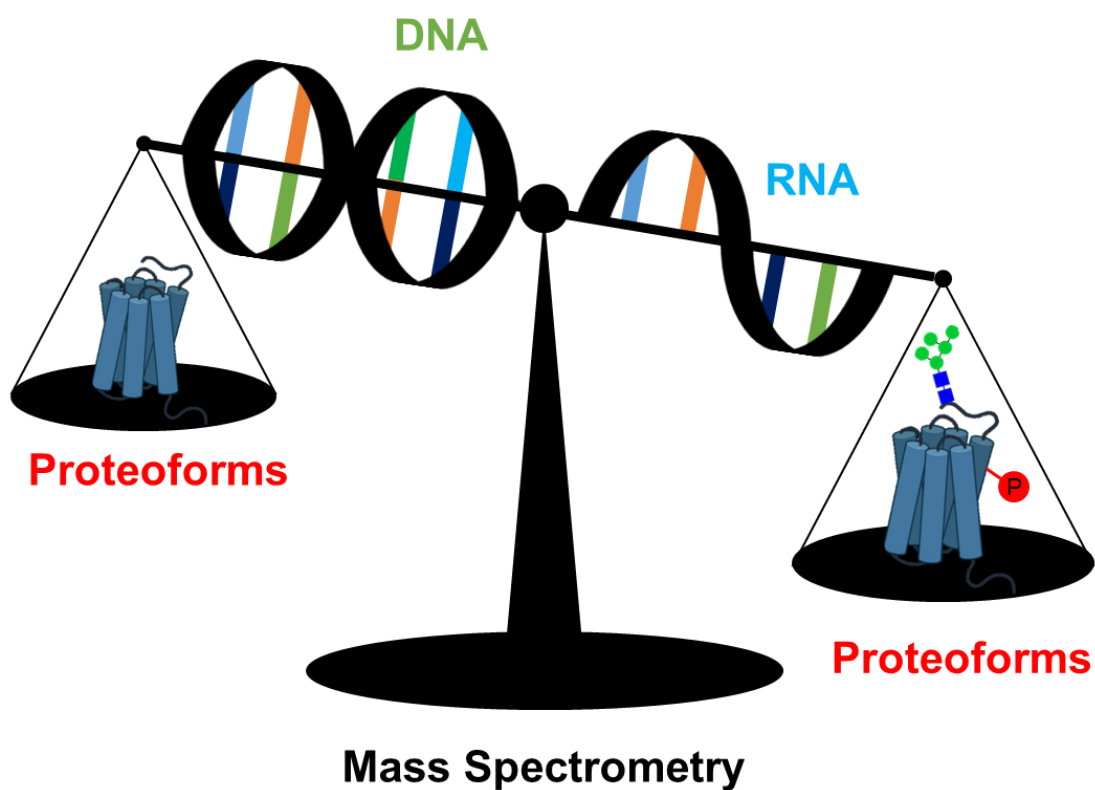
Figure 3.8. Azo-aided digestion for high-throughput targeted peptide mass fingerprinting.	92
Figure 3.9. A direct comparison between Azo and RapiGest™ (RG) in the bottom-up MS experiment with trypsin digestion of myoglobin followed by direct infusion electrospray-ionization MS analysis.	93
Figure 3.10. Enhanced digestion of standard proteins in the presence of Azo.	94
Figure 3.11. Rapid and reproducible extraction, enzymatic digestion, and LC-MS/MS analysis of proteins extracted from HEK293T cells.	95
Figure 3.12. Representative MS/MS spectra using Azo-aided digestion of enriched membrane proteins.	96
Figure 3.13. Azo-enabled bottom-up proteomics analysis of cardiac tissue.	97
Figure 3.14. Representative mass spectrum demonstrating proteoform complexity.	98
Figure 3.15. Intact protein LC-MS analysis of cardiac proteins extracted using Azo.	99
Figure 3.16. Top-down analysis of integral membrane protein cytochrome b-c1 complex subunit 10.	100
Figure 4.1. Schematic illustration of the Azo-enabled ECM proteomics strategy.	106
Figure 4.2. ECM proteomics of mammary tumor tissue.	108
Figure 4.3. Identification of PTMs in the ECM proteins.	113
Figure 4.4. Azo enabled two-step extraction for analysis of extracellular matrix (ECM) proteins.	119
Figure 4.5. Pictures were taken after tissue decellularization and washing steps.	120
Figure 4.6. 8% SDS-PAGE analysis of representative ECM extraction replicates of Decell extract 1 and Azo extract 2.	121

Figure 4.7. Scatter-plots of Log2 transformed relative ECM protein abundances in extraction replicates of Azo extract 2 from mammary mouse tumor tissue.	122
Figure 4.8. Composition of Decell extract 1 and Azo extract 2 from mammary mouse tumor tissue.	123
Figure 4.9. Comparison of Azo to the traditional ECM extraction buffers.	124
Figure 4.10. Comparison of 1- and 2-dimensional LC separation strategies for ECM proteomics.	125
Figure 4.11. Summary of PTM sites identified in ECM proteins from mammary mouse tumor tissue.	126
Figure 5.1. An integrated approach for top-down MS characterization of endogenous membrane proteins.	131
Figure 5.2. Cloud point extraction process flow and SEC fractionation of endogenous membrane proteins from HEK293T cells.	138
Figure 5.3. LC-MS of Large Proteins for HEK293T cells.	141
Figure 5.4. RPLC-MS analysis of enriched integral membrane proteins from HEK293T cells.	143
Figure 5.5. Representative online collisionally activated dissociation (CAD) of multipass integral membrane proteins.	145
Figure 5.6. Protein interactions captured utilizing a cloud point extraction.	146
Figure 5.7. Distribution of the number of transmembrane domains and cellular location of integral membrane proteins identified from HEK293T cells (a) and cardiac tissues (b).	148

Figure 5.8. Visualization of the cloud point extraction using human heart and human embryonic kidney cell (HEK) lysates.	150
Figure 5.9. Direct LC-MS analysis of proteins from membrane enriched-fraction.	151
Figure 5.10. SEC chromatogram comparing membrane proteins separated with a 4.6 and 9.4 mm column.	152
Figure 5.11. SEC chromatogram demonstrating the reproducibility of separation using three extraction replicates.	153
Figure 5.12. Effective high-throughput SEC separation.	154
Figure 5.13. SDS-PAGE analysis of SEC fractions collected from a 4.6 mm column after extraction using Tergitol NP-7 or Triton X-114.	155
Figure 5.14. Comparison of Triton X-114 and Tergitol NP-7 for membrane protein enrichment.	156
Figure 5.15. RPLC-MS analysis of SEC fraction 4 (Triton X-114) from HEK293T cells.	157
Figure 5.16. Azo for extraction of insoluble proteins.	158
Table S1. Collisional energy for protein fragmentation.	159
Figure 7.1. Scheme illustrating the process of DNA transcription into RNA, which is then translated to make proteins.	164
Figure 7.2. Proteins sequencing by mass spectrometry.	166
Figure 7.3. Examples of common protein modifications and the corresponding masses.	167
Figure 7.4. Diagram illustrates the challenge that arises from digesting proteins into peptides.	168

- Figure 7.5.** Diagram of the structure of a common soap, sodium dodecyl sulfate (SDS). 170
- Figure 7.6.** Photocleavable surfactant aid in breaking down the cell to extract and solubilizes proteins. 170
- Figure 7.7.** Illustration of reversed-phase liquid chromatography separation of proteins (top). 172
- Figure 7.8.** Illustration of a protein modification change that can be a factor contributing to a disease or result of a disease 173

Chapter 1. Top-down Proteomics: Innovations, Challenges, and Applications in Basic and Clinical Research



1.1. Introduction

Precise characterization of the function and dysfunction of proteins is the next great undertaking in the post-genomic era for basic and clinical research.^{1, 2} Mutations, alternative splicing events, and post-translational modifications (PTMs) result in a host of proteoforms (a term referring to all the protein products that arise from a single gene)³ that greatly diversifies the proteome.⁴ To understand the role of proteoforms in biological systems we need to unambiguously determine the role of different PTMs in regulating protein activity, localization, degradation, and their alteration in diseases.⁵ Moreover, the role of mutations and differential expression of protein isoform has been shown to drive molecular mechanisms contributing to disease necessitating a greater understanding of their characterization.^{5, 6}

Mass spectrometry (MS)-based proteomics has emerged as the most versatile and comprehensive strategy for protein characterization.^{7, 8} Typically, proteins are extracted from a sample of interest, enzymatically or chemically digested, separated by liquid chromatography (LC), analyzed by MS, and the results processed using bioinformatics to yield protein identification, expression-level quantification, and PTM information.⁹⁻¹¹ This “bottom-up” approach is common as peptides are biochemically more soluble and easier to separate. However, using peptides to infer information regarding proteins leads a loss in isoform and proteoform information because of many homologous sequence regions shared between different forms and sequence coverage is generally low.^{12, 13} For this reason, the “top-down” approach, which forgoes digestion to analyzed intact proteins, is the preferred method for achieving unambiguous, proteoform-resolved molecular detail (**Figure 1.1**).¹⁴ Top-down proteomics is a valuable tool for identifying proteins, localizing PTMs and sequence versions, and permits relative quantification of present proteoforms.¹⁵⁻¹⁹

Here, I will cover the latest advancements (particularly the last 2 years) in top-down proteomics, including sample preparation, protein separation, instrumentation, data acquisition and analysis, and important applications with an emphasis on work towards understanding the mechanisms underlying heart disease.

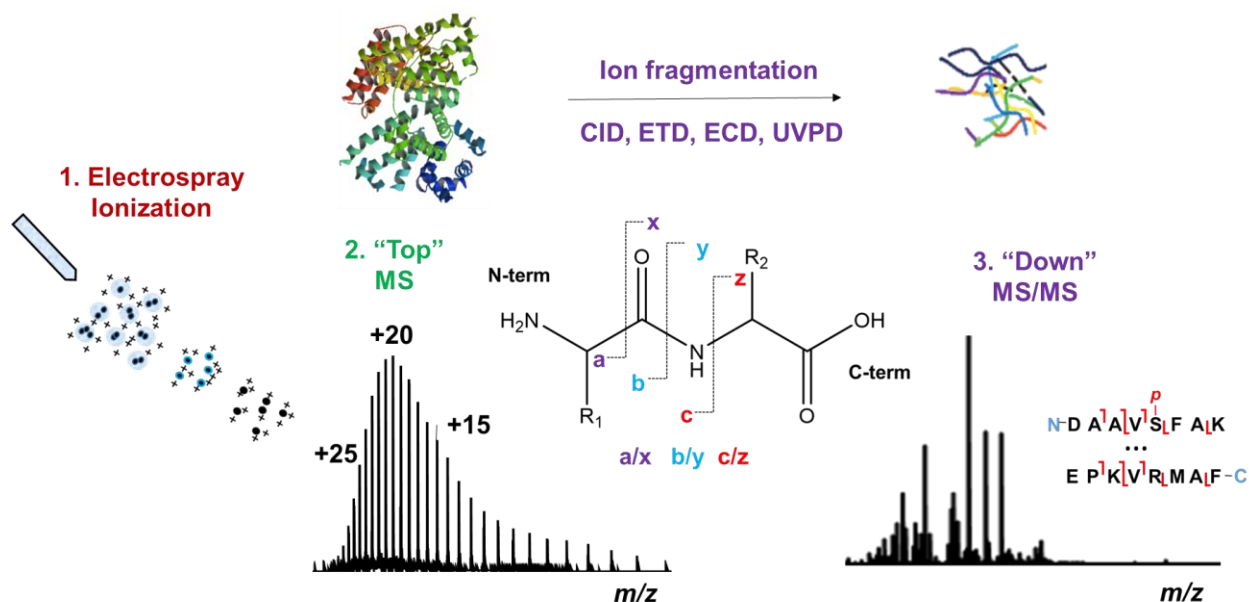


Figure 1.1. Schematic of top-down proteomics, which involves electrospray ionization, intact protein mass measurement, and protein sequencing by tandem MS (MS/MS).

1.2. Methods for Top-down Proteomics

1.2.1. Sample Preparation for Top-down Proteomics

Sample preparation is critical for reliable proteomics data. Commonly, protein extraction for biomedical research involves the use of Good's buffers, high salt (>100 mM), protease and phosphatase inhibitors, and surfactants like sodium dodecyl sulfate (SDS) or Triton X-100 for total protein solubilization.^{14, 20} However, these reagents are generally incompatible with MS analysis and must be removed before downstream analysis.²¹

In an effort to standardize sample preparation, a “best practices and benchmark” was prepared by the Consortium of Top-down Proteomics.²² Using their outlined approach, incompatible salts and other components are removed from target proteins (1-5) by ultracentrifugation filters to exchange MS-incompatible components for low concentration organic acid or volatile ammonium-based salts before analysis. On the other hand, precipitation is used to eliminate contaminants for the analysis of complex mixtures of proteins. Precipitation is often required when using surfactants, which are employed to lyse cells and solubilize hydrophobic membrane proteins but greatly (or completely) inhibit MS signal by outcompeting proteins for ionization or by forming harmful adducts.^{23, 24} However, precipitation can be time-consuming and lead to variability and sample loss, as some proteins are not readily re-solubilized.^{25, 26}

Previously, acid-cleavable surfactants like Rapigest,²⁷ ProteaseMAX,²⁸ and MaSDeS²⁹ were developed to enable protein solubilization for digestion and MS-analysis. However, these surfactants are incompatible with top-down proteomics; therefore, Brown et al. developed a photocleavable ionic surfactant, 4-hexylphenylazosulfonate (Azo).³⁰ Notably, Azo’s protein extraction efficiency was comparable to leading surfactants like SDS and it was compatible with top-down analysis after rapid UV-degradation. Azo was used to broadly solubilized proteins from cardiac tissue including important membrane proteins such as phospholamban,³¹ receptor expression enhancing protein 5,³² and every subunit of the ATP synthase complex.^{33, 34} Additionally, Azo was shown to enable rapid enzymatic digestion demonstrating its unique ability as an “all in one” surfactant for integrated top-down and bottom-up proteomics.³⁵

To improve the sensitivity and automation of top-down sample preparation, Zhu and coworkers adapted the nanoPOTS (nanodroplet processing in one pot for trace samples)³⁶ technology for top-down. They found that a combination of dodecyl- β -D-

maltopyranoside (DDM) with urea enabled the extraction and identification of ~170-620 proteoforms from ~70-770 HeLa cells.³⁷ The nanoPOTs technology shows incredible promise for highly sensitive top-down proteomics (potentially even single-cell top-down analysis), but the majority of identification were < 10 kDa. Moreover, the technology is relatively specialized.

Traditionally, proteomics is performed in denatured mode to study the primary protein sequence.³⁸ More recently, native top-down has been utilized to study protein structure.³⁹ This involves using non-denaturing conditions; often include 100-200 mM ammonium acetate, for analysis of intact protein and protein complexes. Valuable information regarding protein complex stoichiometry,⁴⁰ metal binding,^{41,42} protein interaction partners,⁴³ and more can be achieved using native MS. Analysis of membrane proteins by native MS generally require the use of nonionic surfactants like DDM to solubilize and stabilization.⁴⁴⁻⁴⁶ In this case, the entire protein-micelle complex is ionized, and the protein is liberated using high collisional energy (200 V).⁴⁵ Robinson and coworkers have successfully demonstrated native MS analysis for a number of application including most recently analyzing protein assemblies directly for native membranes.⁴⁷

1.2.2. Protein and Proteoform Enrichment for Top-down Proteomics

The proteome has a high dynamic range; therefore, the analysis of certain protein classes, low abundant proteoform, and target proteins often requires an enrichment step. Proteoform level enrichment is important for identifying and characterizing low abundant species. Enrichment techniques such as titanium oxide⁵² or IMAC⁵³ and lectin columns⁵⁴ are common for phosphopeptide and glycopeptide enrichment respectively. For intact proteoform enrichment and top-down analysis, the Ge group developed superparamagnetic cobalt ferrite NPs functionalized

with GAPT-Zn ligands for phosphoprotein enrichment before LC-MS/MS analysis for confident identification and localization of phosphorylation sites.⁵⁵ The synthetic scalability of these NP was improved using surface silanized magnetite particles coupled with the GAPT-Zn ligand.⁵⁶

For targeted protein characterization, affinity purification (AP) is often employed to purify proteins and characterize their modifications.⁴⁸⁻⁵⁰ Antibodies are the most common affinity ligand utilized for purification; however, they have inherent limitations. Most notably, batch-to-batch variation, high cost, and poor chemical stability, which collectively contribute to reproducibility issues.⁵¹ To address the limitation of antibody-based methods, Tiambeng and Roberts et al. developed functionalized magnetic nanoparticles (NPs) with a peptide affinity ligand to target cardiac troponin I (cTnI),⁵⁷⁻⁵⁹ an important cardiac biomarker.⁶⁰ The NP-peptide approach enabled the sensitive enrichment of cTnI (< 1 ng/mL) with high specificity and reproducibility, while at the same time depleted highly abundant proteins like human serum albumin. Top-down characterization of cTnI, enriched from serum, yielded proteoform-resolved molecular fingerprints of diverse cTnI proteoforms for evaluation of proteoform-pathophysiology relationships. Overall, these advancements demonstrate the great potential for improved enrichment of target proteins and proteoforms coupled to top-down proteomics.

1.2.3. Intact Protein Separation

Global proteoform characterization necessitates high-resolution protein separation before top-down analysis. However, front-end separation of intact proteins represents a major challenge for in-depth proteome coverage in part because of dramatic variation in physicochemical properties of proteins (size, hydrophobicity, charge, etc.).

Reversed-phase liquid chromatography (RPLC) is the predominant method for online sample desalting and separation coupled to MS. Typically a single-dimension of RPLC can only enable the identification of 100 intact proteoforms with limited coverage of larger species (>30 kDa).⁶¹ However, long (>1 m) were demonstrated to provide a peak capacity of > 400 and confidently identified ~900 proteoforms, but they operate at ultrahigh pressures (~14,000 psi).⁶² On the other hand, monolithic columns offer high-resolution protein separation with low backpressure.⁶³ Liang et al. demonstrated that bridged hybrid bis(triethoxysilyl)ethylene monolith with C8 functional groups afforded high-resolution separation of complex cardiac lysate including a large proteoform, α -actinin2 (103.77 kDa).⁶⁴

Capillary zone electrophoresis (CZE) has recently shown great promise as an alternative or complementary front-end separation due to its inherent sensitivity and peak capacity.⁶⁵ The development of sheathflow and sheathless interfaces have enabled fascicle coupling of CZE to MS.⁶⁵ However, the application of CZE for intact protein separation has been limited by the low loading capacity and narrow separation window. To address this issue, the Sun group optimized dynamic pH junction-based, using 10% acetic acid as the background electrolyte, to load ~1 μ L of sample and demonstrated the use of single-shot CZE-MS analysis to identify 600 proteoforms from *Escherichia coli*.⁶⁶

Offline hydrophilic interaction chromatography (HILIC) was previously utilized to separate intact subunits of mitochondria membrane proteins offline before intact protein MS analysis.⁶⁷ More recently, Gargano et al. developed online capillary HILIC for top-down proteomics.⁶⁸ Notably, they utilized a C4 tap column for loading (bypassing the traditional use of high organic solvent for sample composition) and doped propionic acid in the electrospray interface to overcome the signal suppression of trifluoroacetic acid in the mobile phase.⁶⁸

However, limited proteome coverage was achieved and significant work is still needed before the full potential HILIC for top-down proteomics is recognized.

Online LC-MS under nondenaturing conditions for separation before top-down analysis is gaining popularity particularly for the analysis of large biomolecule therapeutics. For example, online size exclusion chromatography (SEC) has been used for rapid online desalting of antibodies and protein complexes.⁶⁹ Similarly, ion exchange chromatography (IEX)-MS has been used to characterize protein conjugates.^{70, 71} Hydrophobic interaction chromatography (HIC)-MS shows great promise for characterizing intact protein directly coupled to top-down MS.⁷² Chen et al. demonstrated direct HIC-MS analysis for determination of the relative hydrophobicity, intact masses, and glycosylations of monoclonal antibodies (mAb) as well as gaining sequence and structural information using MS/MS.⁷³

While the continued development of online separation strategies is essential for enhancing the capabilities of top-down proteomics, no single separation mode can handle the complexity of the proteome. Therefore, multidimensional separation approaches are commonly employed for deeper proteome coverage.⁷⁴ Often large proteins, which have an inherently lower signal-to-noise, are underrepresented in top-down studies.⁷⁵ For this reason, gel-based techniques are commonly used to separate protein based on molecular weight. Gel-eluted liquid fraction entrapment electrophoresis (GELFrEE) is one of the most common pre-fractionation techniques owing to its high-resolution size separation. In 2012, the Kelleher group combined solution isoelectric focusing, GELFrEE, and RPLC to identify 3,000 proteoforms from 1043 protein groups including proteins up to 105 kDa.⁷⁶ More recently, PEPPI (passively eluting protein from polyacrylamide gel as intact species)-MS was introduced as an alternative gel-based separation technique.⁷⁷ Using PEPPI, good protein recovery, and high-resolution separation was achieved for top-down

proteomics of protein <50 kDa. However, little evidence was present that this technique could facilitate top-down proteomics of large proteins (>50 kDa). Overall, gel-based strategies suffer from the use of SDS, which requires sample clean-up that may result in protein loss and variability.²⁵

To bypass the use of SDS for separation, serial SEC (sSEC) was developed by coupling columns with differing pore sizes and using an MS-compatible mobile phase (1% formic acid). sSEC was used achieved high-resolution size-based separation of cardiac sarcomere proteins and when coupled with RPLC-MS enabled the detection of 4,000 unique proteomes up to 223 kDa.⁶¹ SEC-RPLC was later adapted for analysis of endogenous membrane proteins (6-115 kDa) with as many as 19 transmembrane domains (TMDs) by Brown et al. SEC was also coupled to CZE for identification of 5700 proteoforms from *Escherichia coli*.⁷⁸ Yu et al. took a different approach using offline fractionation by high pH RPLC followed by online low pH RPLC coupled to MS to enable the identification of 2778 proteoforms from 628 proteins.⁷⁹

Multidimensional separation strategies have become increasingly adopted for top-down proteomics in nondenaturing mode (i.e., native MS). For example, native SEC and CZE were coupled to identify 672 proteoforms, 144 proteins, and 23 protein complexes.⁸⁰ Using Native GELFrEE and IEX, Skinner et al. identified 125 intact endogenous complexes from mouse hearts and human cancer cell lines.⁸¹

Despite the advancements in the multidimensional separation strategies, performing offline fractionation remains labor-intensive, low throughput, and requires relatively large sample quantities. A few online approaches have been developed including using weak-cationic exchange-HILIC and RPLC characterization of hundreds of histone proteoforms from ~1.5 µg of protein.⁸² Additionally, online high pH RPLC coupled to low pH RPLC shows promise for intact protein

separation before top-down.⁸³ Moving forward the growth of online multidimensional separations will be critical for sensitive, high-throughput top-down proteomics.

1.2.4. Instrumentation and Fragmentation for Proteoform Identification and Characterization

Although several ionization techniques are available, ESI is predominately utilized for top-down proteomics as it is easy to interface with LC.⁸⁴ Moreover, ESI enables highly sensitive detection of analytes; however, the drawback is small molecule contaminants outcompete large molecules resulting in the need for highly pure samples.^{22, 75} Recently, submicron ESI emitter tips, developed by the Williams group, have pushed the boundaries of ESI sensitivity and even enabled the tolerance of high concentration of salts that mimicking physiological conditions (25 mM Tris, 150 mM KCl) by virtue of creating smaller droplets.⁸⁵ Another emerging technique is liquid extraction surface analysis (LESA), which directly sampling proteins from a solid surface for analysis of protein and protein complexes.^{86, 87}

A growing number of instrument platforms are now capable of top-down proteomic analysis. Traditionally, Fourier-transform ion cyclotron resonance (FTICR) mass spectrometers are the instrument of choice for top-down because of its capabilities to resolve very large species;⁸⁸ however, the slower scan speed can limit its capabilities for effective LC-MS/MS analysis and global protein characterization. Quadrupole time-of-flight (q-TOF) mass spectrometers are a common alternative instrument platform, as they are robust, affordable, and capable of analyzing large protein species.^{61, 89} Currently, benchtop Orbitrap mass spectrometers are the most common instrument for proteomics as they bypass the need for cryogenic cool (needed for the magnets of FTICR instruments) while offering high-resolution and fast acquisition speeds.^{90, 91} Moreover,

global top-down proteomics using the Orbitrap has enabled the identification and quantification of thousands of proteoforms.^{76, 92} However, top-down proteomics using an Orbitrap MS was limited to low-molecular-weight species as the signal of larger biomolecules rapidly decays due to collisions with the background gas.⁹³ Work by the Brodbelt and Kelleher group in tuning the Orbitrap instrument parameters has increased its scope for targeted and global analysis of larger proteoforms.^{90, 94} More recently, instrument improvements have led to enhanced large ion transmission and detection, especially with the implementation of ultra-high mass range (UHMR) Orbitraps that are capable of analyzing species at 70,000 m/z and mass as large as 800 kDa.^{95, 96}

Beyond new instrumentation, unique approaches have been implemented to improve the detection of intact proteins (particularly large protein) by generally reducing the complexity of the spectra. The simplest of these is using non-denaturing conditions, which inherently improves the detection of larger biomolecules by reducing the number of charge states.⁹⁷ Similarly, proton transfer reaction (PTR) MS has been used to reduce the average number of charge states for improved signal-to-noise and reduced spectral complexity for the identification of complex mixtures of proteins.⁹⁸ Arguably, one of the most important recent developments is charge detection mass spectrometry (CDMS) (also referred to as individual ion MS) for top-down proteomics (**Figure 4**).⁹⁹⁻¹⁰² CDMS analysis of intact proteins is achieved by selecting one or a few charged ions. Then the m/z and z of individual ions are simultaneously measured, which can then be converted into a mass measurement.¹⁰³ The major advantage of CDMS is the analysis of biomolecules including virus particles (2.47 MDa) and highly complex intact protein mixtures.^{100,}

101

1.2.5. Protein Fragmentation for Sequence Determination

A significant benefit of top-down proteomics is the comprehensive characterization of the protein primary sequence. Effective ion activation for MS/MS analysis is essential for characterizing proteoforms.¹⁰⁴ Collisionally activated dissociation (CAD) and higher-energy collisional dissociation (HCD) are the most prevalent, robust techniques for protein fragmentation especially for global top-down proteomics studies.^{92, 105} Using CAD or HCD, select ions are accelerated colliding with inert gas (nitrogen) breaking the most labile bonds. The result is a series of b/y ions from bond breakage between the carbonyl and amine groups (the “peptide bond”) along the peptide backbone. However, as CAD and HCD break the most labile bonds it may cleave labile PTMs (e.g., phosphorylation) and cause unwanted neutral loss of water and ammonium.¹⁰⁴ Generally, data acquisition in top-down on an LC-MS timescale is performed using data-dependent acquisition (DDA) wherein the topmost intense peaks in the initial MS scan are selected for fragmentation.¹⁰⁵

Electron capture dissociation (ECD)¹⁰⁶ and electron transfer dissociation (ETD),¹⁰⁷ on the other hand, are nonergodic techniques for protein fragmentation. Fragmentation by ECD and ETD cleave between the amine and alpha-carbon resulting in c/z ions. Breakages along the protein backbone occur more randomly enabling higher sequence coverage while preserving labile PTMs. However, the slower reactions kinetics, protein-to-protein variability, and charge state dependence of these techniques limit their application for global proteomic analysis.¹⁰⁴ Recently, electron ionization dissociation (EID), which uses high energy electrons for fragmentation, has shown promise for improved internal fragmentation for higher sequence coverage.¹⁰⁸

UVPD also preserved labile modification, but unlike ECD and ETD, it is charge state independent.¹⁹ This approach uses high-energy photons that translates into an array of protein fragment ions. Predominately, it breaks the bond between the alpha carbon and the carbonyl, a/x

ions, but also produces b/y and c/z ions. The diversity of bond breakages afforded by UVPD makes it ideal for deep sequence coverage, but higher numbers of protein identifications are achieved using HCD for global top-down analysis using LC-MS.¹⁰⁹

Hybrid dissociation technique like electron-transfer dissociation/higher-energy collisional dissociation (EThcD)^{110, 111} and activated ion-electron transfer dissociation (AI-ETD)^{18, 112} show great promise for protein characterization. Generally, these techniques involve activation of protein ions with collisional energy or photon bombardment before ion-ion reaction for improved fragmentation.¹⁰⁴ For example, Riley et al. demonstrated AI-ETD could identify 935 proteoforms from while HCD, ETD, and EThcD identified 1014 proteoforms, 915, and 817, respectively.¹¹³

Beyond primary sequence coverage, MS/MS analysis has been utilized to study protein complexes¹¹⁴ and protein-ligand binding.¹¹⁵ For example, Gault et al. demonstrated the use of MS⁴ analysis to identify both the proteoform state and the associated ligand structure.¹¹⁶ Using the 'Nativeomics' approach, endogenous ligands such as lipids and even drug binding could be identified as demonstrated for several membrane proteins.¹¹⁶

Overall, the growing number of different fragmentation tools available has enabled high protein sequence coverage, improved global proteoform characterization, and characterization of higher-order protein structure.¹⁰⁴

1.2.6. Data Acquisition and Analysis

After acquisition, complex spectra are generated from the fragmentation of protein ions resulting in many isotopic (possible overlapping) clusters. Thus, data analysis software is critical for interpreting spectra. Thorough High-Resolution Analysis of Spectra by Horn (THRASH) was

the first software to automate the deconvolution of high-resolution mass spectra and is still widely used.¹¹⁷ THRASH uses a theoretical isotopomer envelope based on the average elemental composition of amino acids (averagine)¹¹⁸ to compare to the experimental envelopes. A score is then assigned based on the overlap of the experimental and theoretical fit. On the other hand, MS-Deconv generates a large set of candidate isotopomer envelopes and scores sets of envelopes rather than a single.¹¹⁹ UniDec was developed to handle highly heterogeneous spectra, which has been critical for the interpretation of native protein mass spectra.¹²⁰ Recently, FLASHDeconv used pattern matching to quickly deconvolute MS with speeds orders of magnitudes higher than conventional approaches.¹²¹ Other deconvolution strategies have been developed including TopFD,¹²² pParseTD,¹²³ and SNAP (a vendor-specific tool). Recently, combining deconvolution results from a single data set using a simple voting ensemble method or random forest machine learning-enabled improved fragment identification and confidence compared to a single algorithm.¹²⁴ After deconvolution is used to create a fragment mass list, there are several software tools for protein identification. Examples of searching software include ProSight PC,¹²⁵ MS-Align,¹²⁶ TopPIC,¹²² pTop,¹²⁷ Mascot Top Down,¹²⁸ MS-TopDown,¹²⁹ Promex,¹³⁰ and Proteoform Suite.¹³¹ To take advantage of the unique aspects of several of these algorithms the Ge group develop MASH.¹³²

1.3. Application in Disease Research

MS-based proteomics has become a staple in biomedical research for the discovery of novel biomarkers and elucidation of disease mechanisms.¹³³ Commonly, proteomics is used to examine changes in protein expression level or PTMs between healthy and diseased states. For example, in the heart protein isoform switching and PTMs are known to regulate cardiac

contraction and relaxation, as well as play a role in the progression of heart failure.^{50, 134, 135} Additionally, decreased level of cTnI phosphorylation were found hearts with increased chronic heart failure⁵⁰ and cTnI released into the blood upon myocardial infarction is a common biomarker.¹³⁶ Using the previously discussed NP-based method, Tiambeng and Roberts et al. demonstrated different cTnI proteoforms could be enriched at a low concentration from serum and characterized by top-down MS, showing promised for improved clinical diagnosis.⁶⁰ Top-down analysis also revealed distinct sequence variations and phosphorylation sites of myosin light chain in ventricles versus atrium helping to better understand their function in cardiac physiology.¹³⁷ Using a label-free top-down quantification method¹³⁸ Cia et al. were able to simultaneously evaluate protein isoform expression and PTM changes of contractile proteins in human pluripotent stem cell-derived cardiomyocytes, an important model system for studying cardiac diseases, to assess maturation.¹³⁹

Top-down proteomics has been utilized to analyze a range clinical sample including serum, urine, or biopsy tissue enables the identification of protein profiles that characteristic of a certain disease generating a more reliable biomarker.¹⁴⁰⁻¹⁴² For example, PSA is a common maker of prostate cancer; however, benign prostatic hypertrophy and prostate cancer are indistinguishable using this maker. On the other hand, Petricoin et al. identified a unique protein profile that more accurately identified prostate cancer from serum.¹⁴³ Ultra-high-resolution MS using a 21 T FTICR enabled the diagnosis of hemoglobinopathy and β -thalassemia from blood with a 3 min assay.¹⁴⁴ Undoubtedly, the advancement of top-down is stimulated by the rise of large molecule therapeutics such as antibody-drug-conjugates as high-throughput methods for characterization and quality control are required.^{73, 145} As the number of large molecules therapeutics increase top-down will likely become more utilized in pharmaceutical companies.

1.4. Five-year Outlook

Top-down proteomics has broad-reaching applications for characterizing protein PTMs, isoforms, protein complex dynamics, protein-ligand binding, and more. Additionally, a growing number of commercial instrument platforms are amenable to top-down analysis making it easier for laboratories to adopt the strategy. Global identification of proteoforms is a challenge because of the limitations for protein separation and large protein detection. However, native MS and CDMS already demonstrate the incredible scope of intact biomolecule analysis by MS. Finally, top-down proteomics, integrated with other omics (e.g., genomics, metabolomics, etc.) strategies, as well as functional and structural techniques, holds great promise for understanding the mechanisms of various diseases and identifying a novel protein targets for therapeutic development.

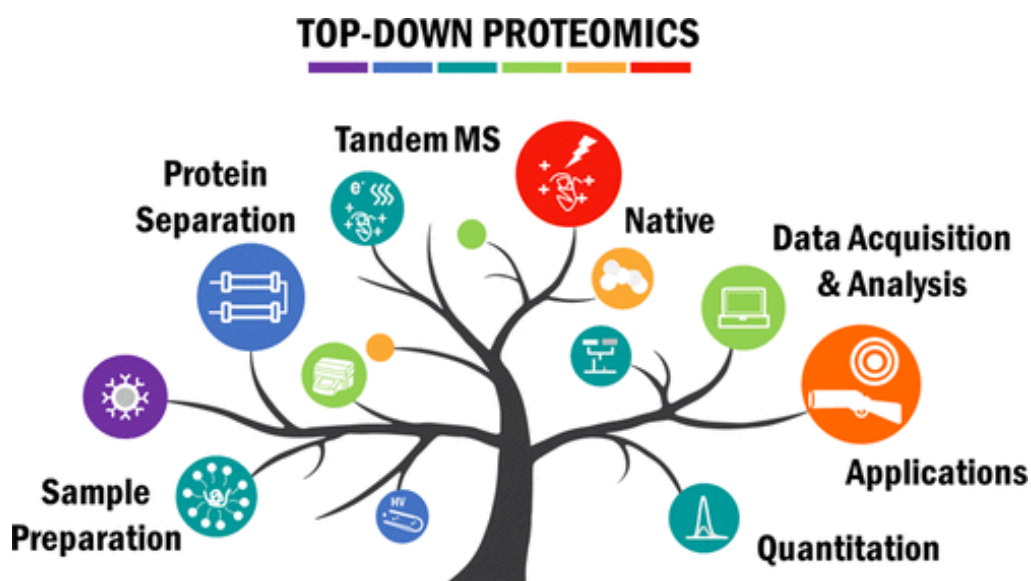


Figure 1.2. Illustration of the different components of top-down proteomics.¹⁴

Chapter 2. A Photocleavable Surfactant for Top-down Proteomics

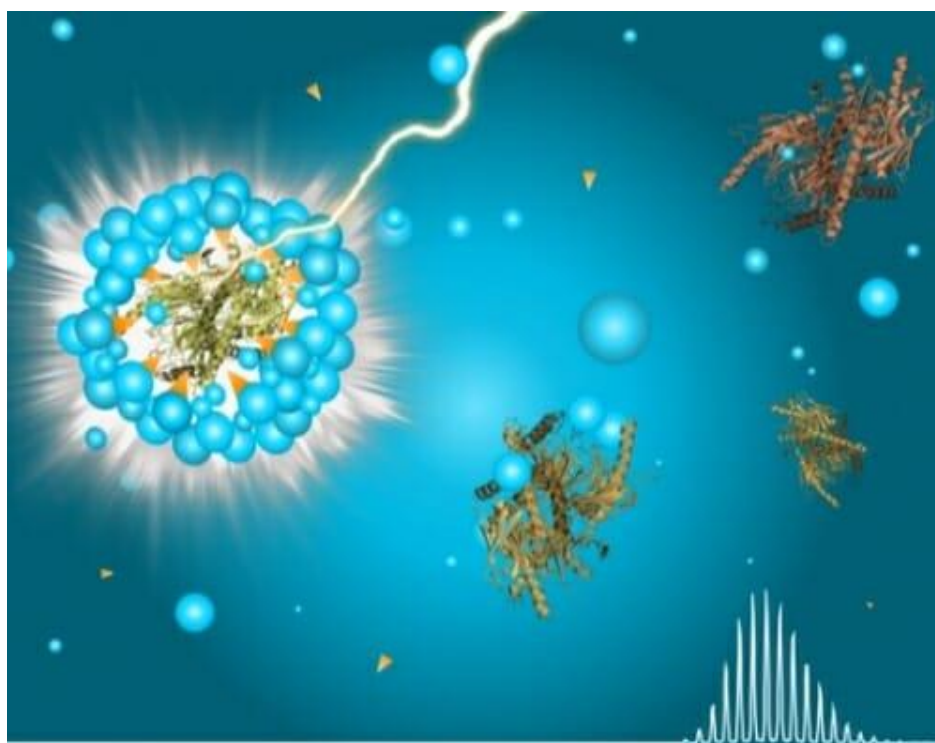


Image Credit: Stanford Mitchell

This chapter has been published and is adapted from:

Brown, K. A.; Chen, B.; Guardado-Alvarez, T. M.; Lin, Z.; Hwang, L.; Ayaz-Guner, S.; Jin, S.; Ge, Y., A photocleavable surfactant for top-down proteomics. *Nature Methods* **2019**.

2. 1. Abstract

We report the identification of a photo-cleavable anionic surfactant, 4-hexylphenylazosulfonate (Azo) that can be rapidly degraded upon UV irradiation, for top-down proteomics. Azo can effectively solubilize proteins with performance comparable to SDS and is mass spectrometry (MS)-compatible. Importantly, Azo-aided top-down proteomics enables the solubilization of membrane proteins for the comprehensive characterization of post-translational modifications. Moreover, Azo is simple to synthesize and can be used as a general SDS replacement in SDS-PAGE.

2.2. Development and Characterization of a Photocleavable Surfactant for Top-down Proteomics

A comprehensive analysis of “proteoforms” that arise from genetic variations and post-translational modifications (PTMs) is essential for deciphering biological systems at a functional level.¹⁴⁶ The conventional “bottom-up” proteomics analyzes peptides from protein digests which does not directly identify proteoforms and is suboptimal for characterizing PTMs and sequence variants¹⁴⁶. In contrast, top-down mass spectrometry (MS)-based proteomics analyzes intact proteins and is the most powerful method to comprehensively characterize proteoforms deciphering the PTMs together with sequence variations.^{6, 146, 147} However, despite significant promises, top-down proteomics still faces major challenges.¹⁴

One such challenge in top-down proteomics is protein solubility¹⁴, especially for membrane proteins, which comprise a large proportion of the proteome and play a critical role in many cellular functions and are important drug targets.^{46,24} To effectively extract proteins from

cells or tissues, surfactants (also known as detergents) are commonly included in the extraction buffer²⁴. Unfortunately, conventional ionic surfactants are not compatible with MS because they greatly suppress protein MS signal^{21, 24}. Therefore, surfactants need to be removed prior to MS analysis, which may result in protein loss and degradation^{25, 148}. Developing MS-compatible surfactants that can be quickly degraded into innocuous non-surfactant byproducts prior to MS analysis can help address the protein solubility challenge in top-down proteomics. Efforts have been made in developing various acid-labile surfactants that have been effective for bottom-up proteomics;^{27, 29, 149, 150} however, none have demonstrated direct compatibility with intact protein MS for top-down proteomics.

Distinct from the previous approaches using acid-labile surfactants^{27, 29, 149, 150}, herein we design and develop photo-cleavable surfactants by inserting a photo-cleavable moiety in between the hydrophilic head and hydrophobic tail that can be rapidly cleaved and degraded upon UV irradiation prior to MS analysis (**Fig. 1a**). Degradation via a photochemical reaction has the advantages of being simple, fast, and can be easily controlled by turning a UV lamp on and off¹⁵¹⁻¹⁵³. Our goal is to identify a strong photo-cleavable surfactant that can effectively solubilize proteins during sample preparation with similar performance to sodium dodecyl sulfate (SDS)¹⁴⁸, but is compatible with top-down proteomics.

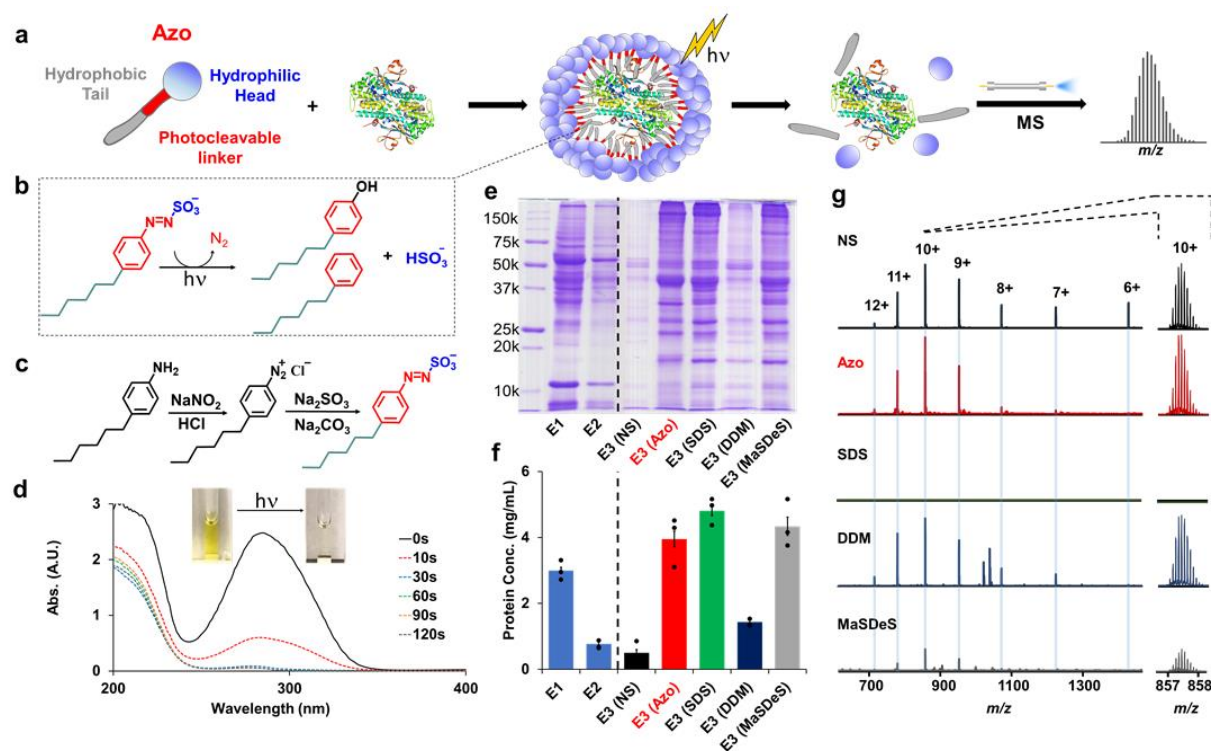


Figure 2.1. Synthesis and characterization of a photo-cleavable anionic surfactant, sodium 4-hexylphenylazosulfonate (Azo). (a) Scheme illustrating the use of Azo in solubilizing proteins, followed by rapid degradation with UV irradiation, and MS analysis of the intact proteins. Note that the molecules are not drawn to scale. (b) Degradation of Azo into 4-hexylphenol, 4-hexylbenzene, nitrogen, and hydrogen sulfate under UV irradiation. (c) Synthetic scheme for Azo. (d) UV-Vis spectra of Azo (0.1%) degradation as a function of time showing that Azo can be rapidly degraded upon UV irradiation at ambient temperature. (e) SDS-PAGE analysis and (f) protein assay for the evaluation of the effectiveness of surfactant aided protein extractions (E3) following the initial HEPES buffer extractions (E1 & E2) to deplete the cytosolic proteins. Error bars represent standard error of the mean for protein assay experiments ($n=3$). (g) Electrospray ionization (ESI)-MS analysis of Ubi with 0.1% surfactant showed the MS-compatibility of surfactants. The mass spectra were normalized to an intensity of $1.7E6$. NS, no surfactant (serving as a control); Azo; SDS, sodium dodecyl sulfate; DDM, n-dodecyl β -D-maltoside; MaSDeS, MS-compatible slowly degradable surfactant. Data are representative of three independent experiments.

We performed a systematic screening of many synthesized candidates (**Supplementary Note 1-3 and Supplementary Table 1**) and identified 4-hexylphenylazosulfonate¹⁵⁴ (**Fig. 1b and Supplementary Fig. 1-2**), hereafter referred to as “Azo”, to be the top-performing surfactant, as it not only is water-soluble, but also greatly improves protein extraction (**Supplementary Table**

1). Notably, Azo was simple to synthesize, requiring only two steps (**Fig. 1c**), and could be effectively purified by recrystallization, making it an ideal candidate for general use as a surfactant in biochemical applications. For instance, we have used Azo instead of SDS to perform polyacrylamide gel electrophoresis (PAGE) (**Supplementary Fig. 3**), demonstrating that Azo could be used as a SDS replacement in SDS-PAGE.

We further investigated the photo-degradation kinetics of the Azo dissociating into 4-hexylphenol, 4-hexylbenzene, nitrogen, and hydrogen sulfate¹⁵¹ (**Fig 1b**) upon irradiation with a 100 W high-pressure mercury lamp for 0, 10, 30, 60, 90, and 120 s using UV-Vis spectroscopy (**Fig. 1d**). By comparing several degradation conditions, we found that the presence of organic solvent and acid facilitates the rapid degradation of Azo (**Supplementary Fig. 4**).

Next, we examined the efficacy of Azo for solubilizing proteins from cardiac tissues using a direct side-by-side comparison with SDS, and its acid-labile mimic, MS-compatible slowly degradable surfactant (MaSDeS)²⁹, as well as dodecyl β -D-maltoside (DDM), a commonly used surfactant for native MS¹⁵⁵. The SDS-PAGE gel (**Fig. 1e**) and protein assay (**Fig. 1f**) shows that the addition of 0.5% Azo to the extraction buffer, labeled as E3(Azo), drastically improved the solubilization of proteins when compared to the control without surfactant, E3(NS), which barely solubilized proteins after the depletion of soluble proteins using HEPES buffer, E1 and E2. Overall, the anionic surfactants, Azo, SDS, and MaSDeS, are highly effective in solubilizing proteins compared to the non-ionic surfactant, DDM (**Fig. 1e,f**). Furthermore, a Western blot analysis confirmed the presence of common cardiac membrane proteins in E3(Azo), demonstrating the successful extraction of integral membrane proteins by Azo (**Supplementary Fig. 5**).

More importantly, the Azo surfactant is MS-compatible. We first performed direct infusion ESI-MS analysis using ubiquitin (Ubi) in the presence of 0.1% of a chosen surfactant, without an

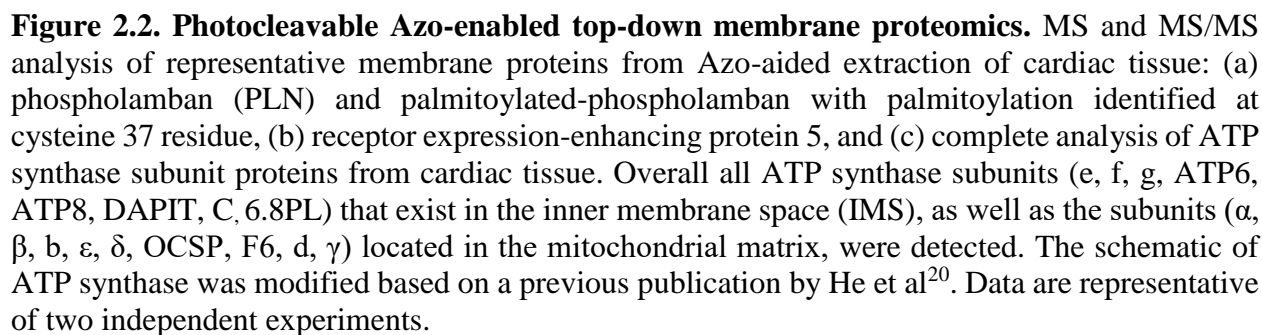
additional desalting step (**Fig. 1g**). The results showed the presence of 0.1% SDS completely suppressed the MS signal and 0.1% MaSDeS significantly suppressed the MS signal. In contrast, 0.1% DDM and 0.1% Azo yielded comparable MS signals, showing minimal signal suppression when compared to the control with no surfactant (**Fig. 1g**). We also examined the effect of the UV-irradiation on MS analysis of proteins and found that Azo had minimal effect on the qualitative and quantitative analysis of intact proteins in the presence of reducing agent, tris(2-carboxyethyl)phosphine hydrochloride (TCEP), or free methionine (**Supplementary Note 4, Supplementary Fig. 6-9**). Furthermore, we have performed a systematic comparison of Azo with a broader range of commonly used surfactants to evaluate their ability to solubilize proteins from the insoluble cardiac tissue pellets and subsequently assessed their MS-compatibility (**Supplementary Note 5 and Supplementary Fig. 10-12**). We concluded that among all the surfactants evaluated, Azo is the only surfactant that not only can effectively solubilize proteins but also is compatible with top-down MS analysis of intact proteins.

Next, we assessed the utility of Azo for top-down proteomics in online reversed-phase chromatography (RPLC)-MS and RPLC-MS/MS experiments (**Supplementary Table 2**) with collision-induced dissociation (CID). Water-insoluble cardiac tissue pellets were extracted with 25 mM NH_4HCO_3 buffer either containing 0.5% Azo [E3(Azo)] or no surfactant [E3(NS)] (**Supplementary Fig. 13a**). Notably, both the SDS-PAGE gel (**Fig. 1e**) and the total ion current (**Supplementary Fig. 13b**) showed significant increases in protein concentration and MS signal, respectively, with the use of Azo when compared to no surfactant. In a *single* RPLC/MS run, we observed 663 *unique* proteoforms in E3(Azo), in contrast to E3(NS) where only 6 unique proteoforms were detected (**Supplementary Fig. 13c,d and Supplementary Table 4**). Moreover, we have detected a total of 2836 proteoforms based on accurate mass measurements

(**Supplementary Table 3**) from the combination of three LC-MS runs; among which 388 proteoforms were identified based on one-dimensional online RPLC-MS/MS data (**Supplementary Table 4**) representing 171 proteins (**Supplementary Table 5**) from mitochondria, nucleus, plasma membrane, cytoskeleton, endoplasmic reticulum, cytoplasm, and extracellular region (**Supplementary Fig. 13e,f and Supplementary Note 6**).

Importantly, we have observed various PTMs including acetylation, methylation, phosphorylation, and palmitoylation (**Supplementary Table 5-10**). In addition to the breadth in increased protein identifications, Azo also greatly improved the depth of the detection and revealed many proteins that were undetectable in the control sample (**Supplementary Fig. 14a-j**). For example, for the first time, Azo enabled the detection and identification of multiple proteoforms of an important Z-disk protein, calsarcin-1 (**Supplementary Fig. 14d**).

Because surfactants are beneficial for solubilization of peripheral and integral membrane proteins, we further show that Azo can effectively extract and enable the top-down proteomic analysis of membrane proteins from cardiac tissues (**Fig. 2 and Supplementary Fig. 15**) as well as human embryonic kidney (HEK) 293T cells (**Supplementary Fig. 16 and Supplementary Table 9-10**). Under the optimal UV-degradation conditions (which include organic solvent at low pH), many hydrophobic proteins were soluble post Azo-degradation. Using cardiac tissue as an example, we identified several important integral membrane proteins such as phospholamban (PLN), receptor-expressing enhancing protein, and succinate dehydrogenase cytochrome b560 with 1, 2, and 3 transmembrane domains (TMD), respectively (**Fig. 2a,b and Supplementary Fig. 17**). Notably, we not only detected intact unmodified PLN but also its highly abundant palmitoylated proteoform (**Fig. 2a**). We confidently localized the palmitoylation modification to cysteine 36 within the transmembrane region based on the unmodified b₃₃ ion and the



Moreover, we confidently identified 46 subunits of the electron transport chain (**Supplementary Table 6**) and 51 proteins with TMDs (**Supplementary Table 7**) directly from

cardiac tissue. Notably, all the subunits of the endogenous ATP synthase complex were identified with high mass accuracy (**Fig. 2c**). This enzyme, which plays a critical role in biological energy metabolism³⁴, includes a domain located in the inner mitochondrial membrane (IMS) (e, f, g, ATP6, ATP8, DAPIT, c, 6.8PL) as well as a domain in the mitochondrial matrix (α , β , b, ϵ , δ , OCSP, F₆, d, γ). In particular, Azo facilitated the identification of ATP6 (also known as ATP synthase subunit a; M_r 24952.55) with 6 TMD as well as the localization of a trimethylation to lysine 43 between 2 TMD of ATP synthase subunit c (**Supplementary Fig. 19**). Besides the small and intermediate size subunits (< 30 kDa), with the use of Azo, we were able to detect and identify the high molecular weight (> 50 kDa) ATP synthase subunits: ATP synthase α and β (**Fig. 2c**). We observed highly efficient CID fragmentation which preferentially cleaved in the transmembrane domain portions of the proteins (**Fig. 2a,b and Supplementary Fig. 17-19**), leading to confident protein identification of these integral membrane proteins and localization of PTMs in online RPLC-MS/MS experiments (**Supplementary Table 4-10**). Thus, Azo enables the detection and comprehensive characterization of these important cardiac membrane protein complexes, which opens up new opportunities to uncover their molecular basis in health and disease.

In summary, we have developed a photo-cleavable MS-compatible surfactant to increase protein solubility and enabled a general, high-throughput method for top-down proteomics. Among all the surfactants we have evaluated, we found that 4-hexylphenylazosulfonate (Azo) to be the only strong surfactant capable of effectively solubilizing proteins, including membrane proteins, without hindering downstream top-down MS analysis. Azo has the potential to further enhance top-down global proteomics when coupled with multidimensional separation, complementary fragmentation techniques, and improved data acquisition strategies¹⁴ (**Supplementary Note 6**). We expect that Azo will facilitate a myriad of proteomic studies for

understanding disease mechanisms and clinical diagnosis. Given surfactants' instrumental roles in biochemical research, we envision this photo-cleavable surfactant will have a broader impact beyond proteomics. Notably, because Azo can be easily synthesized and purified, it can be used as a cleavable SDS-replacement in general biochemical applications, for example, in SDS-PAGE.

2.3. Supplemental Information

2.3.1. *Materials and Reagents.*

All chemicals and reagents were used as received without further purification unless otherwise noted. Sodium nitrate (NaNO_3), sodium sulfite (Na_2SO_3), 4-nitrophenyl chloroformate, sodium carbonate, 4-n-hexylaniline, 4-n-octylaniline, 4-n-decylaniline, 4-n-dodecylaniline, N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC), 4-(bromomethyl)-3-nitrobenzoic acid, octylamine, decylamine, dodecylamine, N-ethyldiisopropylamine (EDIPA), piperidine, 1,4-butanedisulfone, anhydrous N,N-dimethylformamide (DMF), N,N,N',N'-Tetramethyloxy-(1H-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU), and dinitrophenylhydrazine (DNP) were obtained from TCI America (Portland, OR, USA). Fmoc-photolabile linker was purchased from Advanced Chemtech (Louisville, KY, USA). Tetrahydrofuran (THF), ammonium hydroxide (NH_4OH), hexafluoroisopropanol (HFIP), dichloromethane, heptane, acetone, trimethylamine (Et_3N), magnesium sulfate (MgSO_4), sodium carbonate and silica were purchased from Sigma-Aldrich Inc. (St. Louis, MO, USA). Extraction solutions were made in nanopure deionized (DI) water (H_2O) from Milli-Q water (Millipore, Corp., Billerica, MA, USA). HEPES, ammonium bicarbonate (NH_4HCO_3), sucrose, sodium fluoride (NaF), phenylmethanesulfonyl fluoride (PMSF), ethylenediaminetetraacetic acid (EDTA), [n-dodecyl](#) β -D-maltoside (DDM),

octyl β -D-glucopyranoside (OG), sodium dodecyl sulfate (SDS), digitonin (DGT), protease inhibitor cocktail, tri(2-carboxyethyl) phosphine hydrochloride (TCEP), dithiothreitol (DTT), 2-mercaptoethanol (2-ME), β -casein from bovine milk, ubiquitin from bovine erythrocytes (Ubi), bovine serum albumin (BSA), myoglobin from equine heart (Myo) and cytochrome c (Cyt_c) from equine heart, ribonuclease A (RNase A) and ribonuclease B (RNaseB) from bovine pancreas were purchased from Sigma-Aldrich Inc. (St. Louis, MO, USA). ProteaseMax™ (PM)¹⁵⁶ was obtained from Promega (Fitchburg, WI, USA). RapiGest™ (RG also known as ALS)^{27, 150} was purchased from Waters (Milford, MA, USA). Sodium orthovanadate, HPLC grade H₂O, acetonitrile (ACN), methanol (MeOH), ethanol (EtOH), optima LC-MS grade formic acid, optima LC-MS grade isopropanol (IPA), Pierce protein-free tris-buffered saline (TBS) blocking buffer, tween 20, and molecular weight cutoff (10 and 30 kDa MWCO) (0.5 mL) centrifugal filters, Coomassie blue R-250, and Dulbecco's modified eagle medium (DMEM) were purchased from Fisher Scientific (Waltham, MA). Goat Anti-Antigen: Rabbit IgG (H+L), Goat Anti-Antigen: mouse IgG (H+L), BCA protein assay, and Pierce 660 nm Protein Assay Reagent, Ionic Detergent Compatibility Reagent were purchased from Thermo Fisher ([Waltham, MA](#)). Protein Assay Dye Reagent Concentrate was purchased from BioRad (Hercules, CA). Voltage-dependent anion-selective channel (VDAC) antibody was purchased from Biovision (Milpitas, CA). Mitochondrial import receptor subunit (TOM20) was purchased from Santa Cruz Biotechnology (Dallas, Tx). Sodium potassium adenosine triphosphate (Na-K ATPase) and cadherin antibodies were purchased from Abcam (Cambridge, United Kingdom). Phospholamban antibody was purchased from Bioss (Woburn, MA). Fetal bovine serum (FBS) was purchased from Life Technologies (Carlsbad, CA). Mini-gels (12.5%) for SDS polyacrylamide gel electrophoresis (SDS-PAGE) were prepared in house using acrylamide/Bis-Acrylamide (37.5:1) 40% solution purchased from Hoefer (Holliston,

MA). MS-compatible degradable surfactant (MaSDeS) was synthesized by Promega and provided to us as a gift as described previously¹⁰.

2.3.2. *Synthesis of O-nitrobenzyl (ONB) Surfactant Family.*

Synthesis of 4-(hydroxyethyl)-3-nitrobenzoic acid. A solution of 500 mg of 4-(bromomethyl)-3-nitrobenzoic acid (1.92 mmol) and 814 mg of Na₂CO₃ (7.68 mmol) in 16 mL of a mixture of H₂O/acetone 1:1 (v/v) was refluxed for 5 h. The acetone was then evaporated and the resulting solution was washed with 9 mL of diethyl ether. After the wash, the solution was acidified with 18% hydrochloric acid until a precipitate was observed. The product was then extracted with ethyl acetate (3 × 12 mL). The concentrated organic layer was washed with H₂O (6 mL) and dried over MgSO₄. The dry organic layer was filtered and concentrated *in vacuo* to yield 74% of 4-(hydroxyethyl)-3-nitrobenzoic acid as a yellow solid²².

Synthesis of intermediate product I (C12) (Supplementary Note 1). Using a traditional EDC coupling, 270 mg of 4-(hydroxyethyl)-3-nitrobenzoic acid (1.37 mmol) was reacted with 0.32 mL of dodecylamine (1.37 mmol) to produce **1** (**n**= 10, **C12**) in a 44% yield.

Synthesis of intermediate product I (C8) (Supplementary Note 1). Using a traditional EDC coupling, 300 mg of 4-(hydroxyethyl)-3-nitrobenzoic acid (1.52 mmol) was reacted with 0.25 mL of octylamine (1.37 mmol) to produce **1** (**n**= 6, **C8**) in a 19% yield.

Synthesis of intermediate product I (C6) (Supplementary Note 1). Using a traditional EDC coupling, 300 mg of 4-(hydroxyethyl)-3-nitrobenzoic acid (1.52 mmol) was reacted with 0.20 mL of hexylamine (1.37 mmol) to produce **1** (**n**= 4, **C6**) in a 46% yield.

Synthesis of intermediate product IIs (Supplementary Note 1). 7.7 mmol of each intermediate product I was dissolved in 40 mL of THF and the solution was cooled to 0 °C. While stirring, 4-nitrophenylchloroformate was slowly added to the THF solution. Then 0.16 mL of pyridine was added dropwise over 20 min and the reaction was stirred for an additional 2 h. The reaction was then filtered. The final product was purified using a silica column that was packed using a solvent of 7:3 ratio of heptane: EtOH and an eluting solvent of a ratio of 4:1 heptane: EtOH²³.

Synthesis of ONB final product (Supplementary Note 1). 0.13 mmol of intermediate product II was dissolved in 2.3 mL of THF. In a separate container, 0.2 mmol of 3-aminopropane sulfonic acid sodium salt in 0.43 mL of water was added to the THF solution. The reaction was stirred overnight at 50 °C. The product was purified using a silica column with a mixture of dichloromethane:MeOH (1:5). The final product was confirmed by electrospray ionization mass spectrometry (ESI-MS). ESI mass spectra for the synthesized ligand molecules were obtained using a Waters (Micromass) LCT[®] mass spectrometer. ONB **C12** (C₂₄H₃₈N₃O₈SSNa), [M-Na+H+NH₄]⁺: calculated *m/z*: 547.6, observed *m/z*: 547.3; ONB **C8** (C₂₀H₃₀N₃O₈SSNa), [M-Na+H+NH₄]⁺: calculated *m/z*: 491.5, observed *m/z*: 491.3; ONB **C6** (C₁₈H₂₆N₃O₈SSNa), [M-Na+H+NH₄]⁺: calculated *m/z*: 463.5, observed *m/z*: 463.3.

2.3.3. *Synthesis of O-nitroveratryl (ONV) Surfactant Family.*

The ONV surfactants were synthesized following previously reported procedures¹⁵. Briefly, to a solution of Fmoc-ONV-COOH (0.57 mmol) and HBTU (0.69 mmol) in 3.5 mL of anhydrous DMF, EDIPA (1.2 mmol) was added dropwise. The solution was cooled on ice and

added to a solution of dodecylamine in 0.5 mL of ice-cold EtOH. After stirring for 30 min at 0 °C, the mixture was stirred overnight at room temperature (RT). The resulting precipitate was filtered and washed with DMF followed by *in vacuo* drying. Intermediate product I (**n= 10, C12**) was obtained as an amorphous white powder. A similar procedure was followed for **n= 8, C10**, and **n= 6, C8 (Supplementary Note 2)**.

Synthesis of NH₂-ONV-CH₂ (CH₂)_n CH₃ (Intermediate product II) (Supplementary Note 2). Piperidine was added dropwise to a solution of intermediate product I (0.6 mmol) in anhydrous DMF (3 mL) to reach a final concentration of 2 M. The solution was stirred at RT for 2 h, and then DMF was removed by evaporation. The residual was dissolved in MeOH and the resulting precipitate was removed by filtration. A pale yellow solid was obtained after evaporation of the filtered solution.

Synthesis of Sulfonate-ONV-CH₂ (CH₂)_n CH₃ (Final product) (Supplementary Note 2). 1,4-butanedisulfone (2.1 eq, 0.74 mmol) was added to a solution of intermediate product II (1.0 eq, 0.35 mmol) with Et₃N (2.0 eq) in ACN (2 mL) and then the flask was sealed. The mixture was stirred and heated to ~90 °C for 48 h. After removing the solvent by evaporation, a light yellow viscous oil was obtained quantitatively. The oil was suspended in water and a NH₄OH (aq) solution was added dropwise until pH ~8 was reached. The surfactant solutions were centrifuged. The final product was confirmed by ESI-MS. ONV **C12** (C₂₉H₅₀N₃O₈SN_a), [M-Na]⁻, calculated *m/z*: 600.3, observed *m/z*: 600.3; ONV **C10** (C₂₇H₄₆N₃O₈SN_a), [M-Na]⁻, calculated *m/z*: 572.3, observed *m/z*: 572.2. ONV **C8** (C₂₅H₄₂N₃O₈SN_a), [M-Na]⁻, calculated *m/z*: 544.3, observed *m/z*: 544.2.

2.3.4. Synthesis of the Azobenzene (AZO) Surfactant Family.

The AZO surfactant family was synthesized following similar procedures as previously described²⁴ (**Supplementary Note 3**). Specifically, 4 mmol of 4-n-hexylaniline (**n= 4, C6**) was stirred in a mixture of 4.8 mL of 10% hydrochloric acid and 8 mL of DI H₂O. Then 4 mmol of NaNO₂ dissolved in 4 mL of cold water, was added dropwise to this solution. During the addition of the NaNO₂, the solution was cooled to 10 °C. After the addition was completed (15 min), the solution was stirred for an additional 15 min at 5 °C. A similar procedure was carried out for 4-n-octylaniline (**n=6, C8**). For 4-n-decylaniline (**n= 8, C10**) and 4-n-dodecylaniline (**n= 10, C10**), the solution of 4-n-alkylaniline was heated to 70 °C and then cooled in an ice bath to 10 °C under vigorous stirring. NaNO₂ was added dropwise starting at 20 °C and concluded at 10 °C, followed by 15 min of stirring at 5 °C. For the coupling reaction, the freshly prepared diazonium salt was filtered into a stirring and cooled solution (T= 5-10 °C) of 8 mmol of Na₂SO₃ and 12 mmol of Na₂CO₃ in 20 mL of DI H₂O. To complete the precipitation of the surfactant, the solution was refrigerated at 4 °C overnight. The yellow compounds were purified by recrystallization with a yield of about 50% and no impurities were detected by NMR. Surfactant solutions were made by gently heating the surfactant at 37 °C then bringing to room temperature after no solid remained. Working concentration was 0.5%-1% in 25 mM NH₄HCO₃. Kraft temperature (a clear 1% surfactant solution) was previously reported at 24.5 °C⁷. A high-resolution mass spectrum of AZO (**C6**), referred to as Azo, (**Supplementary Figure 1**) was taken as follows: A solution of 1% Azo in 25 mM NH₄HCO₃ was diluted 1:100 in ACN (0.3% NH₄OH). The sample was directly injected into a 7 T linear ion trap/Fourier transform ion cyclotron resonance (LTQ/FT-ICR) mass spectrometer (LTQ/FT Ultra, Thermo Scientific, Bremen, Germany) with a nano-ESI source (Triversa NanoMate; Advion Bioscience, Ithaca, NY). A voltage of -1.4 kV was applied with 0.3 psi drying gas. 50 scans were averaged with 5 microscans in a scan. The mass range was set from

100 to 500 m/z . ESI-MS for Azo ($C_{12}H_{17}N_2O_3SNa$), $[M-Na]^+$, calculated m/z : 269.096, observed m/z : 269.098. A Hermes-Varian Mercury Plus 300 operating at 300 MHz was utilized for 1H -NMR spectroscopy with chemical shifts reported as ppm (parts per million). 1H NMR: δ 7.64 (2H, dd, -Ar-H), 7.37 (2H, d, -Ar-H), 2.67-2.48 (2H, m, -Ar-CH₂), 1.61 (2H, t, -Ar-CH₂CH₂), 1.28 (6H, t, -(CH₂)₃) .086 (3H, t, -(CH₂)₃-CH₃).

2.3.5. Tissue handling.

Swine hearts were excised from healthy Yorkshire domestic pigs, snap-frozen in liquid N₂, and stored under -80 °C before use. All homogenization and centrifugation steps were performed at 4 °C.

2.3.6. Protein Extraction and LC-MS Analysis of Cardiac Tissue.

The frozen tissue samples (~500 mg) were cut into small pieces and washed with PBS buffer containing protease inhibitors and reducing agent (5 mM DTT, 1 mM PMSF, 1x protease inhibitor cocktail). The tissue was then homogenized in HEPES buffer with both protease and phosphatase inhibitors (25 mM HEPES, 250 mM sucrose, 50 mM NaF, 1 mM PMSF, 2.5 mM EDTA, 1 mM Na₃VO₄, 1 mM PMSF, 5 mM DTT, 1x protease inhibitor cocktail) with a Polytron electric homogenizer (model PRO200, Pro scientific, Oxford, CT) set to the lowest speed as described previously¹⁰. The homogenate was centrifuged at $211,750 \times g$ using Beckman Ultracentrifuge and a Ti-80 rotor for 1 h. The supernatant after the first HEPES extraction was removed and saved as “E1” extraction. The HEPES extraction was repeated on the resulting pellet and saved as “E2”. After the second HEPES extraction, the tissue pellet was suspended in 25 mM

NH_4HCO_3 and evenly divided into smaller aliquots. In one aliquot, 25 mM NH_4HCO_3 buffer with no surfactant (NS) serving as controls was used in a 1:1 ratio (homogenate:buffer) and labeled as “E3(NS)” following incubation and centrifugation. In the other aliquots, surfactants (1% in 25 mM NH_4HCO_3) were individually added to the other aliquots in a 1:1 ratio (homogenate: surfactant) and labeled as “E3(Surfactant)” following incubation and centrifugation. Protein assays were performed using Pierce 660 nm Protein Assay Reagent with Ionic Detergent Compatibility Reagent (<https://www.thermofisher.com/order/catalog/product/22660>) for data presented in **Figure 1f** and BCA protein assay (with 5% SDS compatible) for data presented in **Supplementary Figure 10**.

Reversed-phase chromatography (RPLC) was performed with a nanoACQUITY M-Class UPLC system (Waters; Milford, MA, USA). Mobile phase A (MPA) contained 0.2% formic acid in H_2O , and mobile phase B (MPB) contained 49.9% ACN: 49.9% IPA: 0.2% formic acid. For each injection, 5 μL of the sample was loaded on a home-packed [250×0.250 mm or 0.5 mm, 5 μm , 1000 Å PLRP-S (Agilent Technology, Santa Clara, CA, USA)]. Samples eluted from the column were electrosprayed into a maXis II ETD Q-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) for online LC-MS and LC-MS/MS experiments. Endplate offset and capillary voltage were set at 500 and 4500 V, respectively. The nebulizer was set to 0.5 bar, and the dry gas flow rate was 4.0 L/min at 220 °C. The quadrupole low mass cutoff was set to 600 m/z during MS and 200 m/z during MS/MS. Mass range was set to 200-3,000 m/z and spectra were acquired at 1 Hz for LC-MS runs. For the top 3 data-dependent LC-MS/MS collision-induced dissociation (CID) runs, MS/MS spectra were acquired across 200-2,500 m/z at 2 - 4 Hz with active exclusion after 4 spectra. Targeted LC-MS/MS CID was performed at 1 Hz after determining the elution time frame from the targeted proteins.

20 μ L of cardiac tissue lysate with or without Azo (referred to as “Azo” or “NS”, respectively) was added 116 μ L H₂O, 2 μ L of HFIP (5%), 2 μ L trifluoroacetic acid (10%), 10 μ L TCEP (1 M), 50 μ L IPA, 50 μ L ACN. Reagents were added slowly and mixed throughout to avoid precipitation. The samples were transferred to a quartz cuvette and irradiated for 3 min using a 100 w high-pressure mercury lamp. The resulting samples were exchanged into 20% ACN: IPA (1% formic acid) with a 10 kDa MWCO centrifugal filter and adjusted to a final volume of 150 μ L. Proteins were separated using the following conditions 0-5 min 20% B, 5-25 min 20-60% B, 65-75 min 60-75% B, 75-80 min 75-95% B, 85-86 min 95-20%B, 86-95 min 20% B. The methods described here correspond to the data presented in **Figure 3, Supplementary Fig. 13-15, 17-19, and Supplementary Table 2-8.**

2.3.7. SDS-PAGE Comparing Azo with SDS, DDM, and MaSDeS.

An equal volume (7 μ L) of each extraction was subsequently resolved using 12.5% SDS-PAGE with a voltage of 50 V for 30 min and 120 V for approximately 75 min. Proteins were visualized using Coomassie Brilliant Blue R-250. The methods described here correspond to the data presented to **Figure 1e.**

2.3.8. Western Blot Comparing Azo with SDS, DDM, and MaSDeS.

Equal volumes of tissue lysate (10 μ L) were loaded and resolved on 12.5% SDS-PAGE gels. Proteins were transferred to a PVDF membrane, fast semi-dry blotter (Fisher Scientific, Waltham, MA), using 20 V for 12 h at 4 °C. The membrane was placed in a protein-free blocking buffer (Fisher Scientific, Waltham, MA) for 1 h at RT and incubated with primary antibodies for

1.5 h at RT. The membranes were then washed by using TBS with 0.1% tween five times before incubation with the secondary antibodies for 1.5 h (RT). After 5 washes with TBS with 0.1% Tween, the membranes were developed using enhanced chemiluminescence detection (Fisher Scientific, Waltham, MA). The methods described here correspond to the data presented in **Supplementary Figure 5**).

2.3.9. Comparison of the Top-down MS Compatibility of Azo with SDS, DDM, and MaSDeS.

Ubi was dissolved in a buffer containing 80: 5: 5: 10 IPA: H₂O: formic acid: 1% surfactant (Azo, SDS, DDM, or MaSDeS) with 10 mM DTT. The Azo sample was irradiated for 3 min. The MaSDeS sample was degraded for 24 h at RT. The samples were then directly injected into a 7 T linear ion trap/Fourier transform ion cyclotron resonance (LTQ/FTICR) mass spectrometer (LTQ/FT Ultra, Thermo Scientific, Bremen, Germany) with a nano-ESI sprayer (TriVersa NanoMate; Advion Bioscience, Ithaca, NY). A voltage of 1.4 kV vs the inlet was applied with 0.3 psi drying gas. 50 scans were collected with 5 microscans in one scan. The mass range was set from 600 to 2,000 *m/z*. The methods described here correspond to the data presented in **Figure 1g**.

2.3.10. Protein Extraction and LC-MS Analysis of Sarcoplasmic Reticulum (SR) and Mitochondria (Mit) Enrichment from Cardiac Tissue.

After cutting around 170 mg of tissue into small pieces, the tissue was homogenized in HEPES buffer containing both protease and phosphatase inhibitors (50 mM HEPES, 0.6 M KCl, 250 mM Sucrose, 500 mM NaF, 1 mM PMSF, 2 mM EDTA, 1 mM Na₃VO₄, 5 mM DTT, 25 µg/mL DGT, 1x protease inhibitor cocktail) with a Polytron electric homogenizer set to the lowest

speed (tissue) to deplete soluble proteins as described previously²⁵. The homogenate was centrifuged at $20,000 \times g$ using a Thermo Scientific Legend Micro 21R Ultracentrifuge. The supernatant was removed and labeled as “E1”. The pellet was suspended in the buffer (25 mM NH_4HCO_3 , 500 mM NaF, 1 mM PMSF, 2 mM EDTA, 1 mM Na_3VO_3 , 5 mM DTT, 25 $\mu\text{g}/\text{mL}$ digitonin, 1x protease inhibitor cocktail) to remove residual proteins and labeled as “E2”. The resulting tissue pellet was suspended in 25 mM NH_4HCO_3 and evenly divided into smaller aliquots, centrifuged at $20,000 \times g$, and the supernatant was removed. 25 mM NH_4HCO_3 buffer (NS) or 1% Azo in 25 mM NH_4HCO_3 was added to the aliquots respectively. After homogenization and incubation, the samples were centrifuged and the supernatant was collected.

50 μL of enriched sarcoplasmic reticulum and mitochondria lysate from cardiac tissue was diluted with 440 μL of 50: 48.5: 1: 0.5 IPA: H_2O : formic acid: HFIP and 10 μL of TCEP (1 M). The sample was irradiated for 3 min and concentrated to a final volume of 150 μL MWCO (10 kDa in run 1 or 30 kDa in run 2). Proteins were separated using the following gradient: 0-1 min 5% B, 1-5 min 5-30% B, 5-55 min 30-60% B, 55-57 min 60-95% B, 57-65 min 95%B, 65-67 min 95% B, 67-80 min 5% B. Column temperature was 35 °C. For ATP synthase subunit α , a single charge state was isolated and fragmented with 5, 10, 16, 18, 20 eV, respectively, using an isolation window of 3 m/z during targeted CID MS/MS experiments. The methods described here correspond to the data presented in **Figure 3, Supplementary Fig. 13-15, 17-19, and Supplementary Table 2-8**).

2.3.11. Protein Extraction and LC-MS Analysis of Endoplasmic Reticulum (ER) and Mitochondria (Mit) Enriched Lysate from Human Embryonic Kidney (HEK) 293T Cells.

Cells were grown on 10 cm plates in DMEM with 10% fetal bovine serum and 1x penicillin/streptomycin solution at 37 °C and 5% CO₂. Cells from two 10 cm plates were washed twice with PBS and lysed in 500 µL of buffer (10 mM Tris, 2 mM DTT, 1 mM PMSF, 50 µg/mL DGT, 1x protease inhibitor cocktail) using 50 strokes with Dounce homogenizer followed by 5 passages through a 27 G needle. Cells were then incubated for 10 min on ice, evenly divided into two aliquots, and centrifuged at $1,000 \times g$ (4 °C) to remove unbroken cells and the nuclei. The supernatant was mixed with 0.5 mL of sucrose (50%) and centrifuged at $21,000 \times g$ (4 °C). The pellet was washed with 1 mL of NH₄HCO₃ (E2). Finally, the pellets were dissolved in 100 µL of Azo (0.5% in 25 mM NH₄HCO₃) or 100 µL of 25 mM NH₄HCO₃ without surfactant (NS) serving as controls.

50 µL of enriched endoplasmic reticulum and mitochondria lysate from HEK cells was diluted with 400 µL of 50% IPA: 49% H₂O: 1% formic acid and 50 µL of TCEP (1M). The sample was irradiated for 3 min then concentrated and exchanged into 10:10:80 ACN: IPA: 1% formic acid in H₂O with a 10 kDa MWCO centrifugal filter. Protein were separated using the following gradient: 0-5 min 20% B, 5-65 min 20-95% B, 65-75 min 95% B, 75-76 min 20% B, 76-80 min 20% B. Column temperature was 50 °C. The methods described here correspond to the data presented in **Supplementary Fig. 16 and Supplementary Table 9-10**.

2.3.12. Azo-PAGE and SDS-PAGE Comparison.

Resolving gel was made using 1.62 mL water, 2.09 mL acrylamide, 1.25 mL Tris-base (1.5 M, pH 8.8), 0.05 APS (10%), and 0.002 mL TEMED. Stacking layer was made using 1.42 mL water, 0.33 mL acrylamide, Tris-base (1 M, pH 6.8), 0.02 APS (10%), and 0.002 mL TEMED.

2.5 µg of BSA, β-casein, and RNase A or 10 µg of cardiac myofilament extract²⁶ was separated on a 1 mm, 12.5% polyacrylamide gel running at 150V. Azo Loading Dye (2x) consisted of 100 µL Tris (1M pH 6.8), 10 mg Azo, 200 µL bromophenol blue (0.04% solution), 200 µL glycerol, 20 µL DTT (1M), and adjusted the volume to 1 mL with water. The azo running buffer was made using 1.5 g Tris base, 7.2 g Glycine, 2.5 g Azo, and adjusted the volume to 1 L with water. The SDS-PAGE comparison gel was run using the same condition except for 20 mg of SDS was used in the Loading Dye and 0.5 g of SDS was used in the running buffer. The methods described here correspond to the data presented in **Supplementary Figure 3**.

2.3.13. UV-Vis Degradation.

50 µL of 0.1% Azo in (a) H₂O, (b) 1% formic acid, (c) IPA, (d) 1% formic acid in IPA, (e) 2-ME in H₂O, and (f) 1% formic acid in IPA: H₂O, respectively, were irradiated with 100 W high-pressure mercury lamp (Nikon housing with Nikon HB-10101AF power supply; Nikon, Tokyo, Japan) for 0, 10, 30, 60, 90, and 120 s in a quartz cuvette. The samples were diluted to a final volume of 1 mL in H₂O. A UV-Vis spectrum was taken from each sample with a Varian Cary 50 UV-Visible spectrophotometer (background correction, medium scan rate, 600-200 nm). The methods described here correspond to the data presented in **Supplementary Figure 4**.

2.3.14. Evaluation of the Effect of Reducing Agents during Azo Degradation.

Standard proteins, Ubi, RNase A, Cyt_c, and BSA were dissolved in 49.5:49.5:1 H₂O: IPA: formic acid and kept on ice until analysis. Samples were irradiated with a 100 W lamp for 3 min. 5 µL of the sample was injected onto a trap column and eluted with 40:40:20 ACN: IPA: 1% formic acid

in H₂O after a 5 min wash with 2.5:2.5:95 ACN: IPA: 1% formic acid in H₂O. 50 mM of DTT, TCEP, and 2-ME was added to each Cytc samples prior to irradiation. Additionally, a sample of Cytc was kept at RT with no reducing agent and irradiated for 3 min with no reducing agent as a control (corresponding to **Supplementary Figure 6-7**). This method was repeated to test 10 mM TCEP and 33 mM methionine (corresponding to **Supplementary Figure 8**).

2.3.15. Protein Extraction and LC-MS Analysis for Evaluating the Effect of Azo on Relative Quantitation.

10 volumes of buffer (10 mM Tris, 500 mM NaF, 2 mM EDTA, 1 mM PMSF, 1 mM Na₃VO₄, 5 mM DTT) was added to swine heart tissue. The sample was homogenized with Teflon homogenizer, centrifuged at 16,000 × g, and the supernatant was collected. Protein extract was diluted to a final buffer containing 25% IPA, 25% ACN, 1% formic acid, 100mM TCEP, and 5 mM NH₄HCO₃ with or without 0.2% Azo. The sample was irradiated for 3 min and exchanged into a 10% ACN, 10% IPA, with 0.2% formic acid using a 10 kDa MWCO centrifugal filter. Protein were separated using the following gradient: 0-5 min 20% B, 5-30 min 20-65% B, 30-35 min 65% B, 35-36 min 20% B, 36-40 min 20% B. Column temperature was 60 °C. The methods described here correspond to the data presented in **Supplementary Figure 9**.

2.3.16. Comparison of the Top-down MS Compatibility of Azo to a Broader Range of Commonly used Surfactants.

Ubi was dissolved in a buffer containing 75: 10: 5: 10 MeOH: H₂O: formic acid: 1% surfactant (MaSDeS, PM, RG, NS, SDS, Azo, OG, DDM, DGT) with 10 mM TCEP. The Azo

sample was irradiated for 3 min. The acid-labile surfactants were incubated for 75 min (24 h for MaSDeS) at 37 °C. The samples were then directly injected into a 12 T Fourier transform ion cyclotron resonance (solariX) mass spectrometer (Bruker Daltonics, Bremen, Germany) with a nano-ESI sprayer (TriVersa NanoMate; Advion Bioscience, Ithaca, NY). A voltage of 1.4 kV vs the inlet was applied with 0.3 psi drying gas. 200 scans were averaged for each sample. The mass range was set from 600 to 2,000 m/z with a 512,000 word transient. The methods described here correspond to the data presented in **Supplementary Figure 11**.

2.3.17. Protein Extraction for Top-down LC-MS Compatibility Comparing Azo to MaSDeS, PM, RG, SDS, OG, DDM, DGT.

83.3 mg of swine cardiac tissue was homogenized in 1 mL of buffer (25 mM NH_4HCO_3 , 1 mM TCEP, and 1 mM PMSF). After centrifugation at 16,000 x g, the supernatant was collected and the protein concentration was adjusted to 2 mg/mL. To 15 μL of swine cardiac protein extract (2 mg/mL) was added 1.5 μL water, 6 μL methionine (25 mg/mL), 25 μL isopropanol, 5 μL TCEP (100 mM), and 2.5 μL formic acid. The Azo sample was irradiated for 3 min. The acid cleavable surfactants PM¹⁵⁶ and RG²⁷ (also known as ALS¹⁵⁰) were incubated at 37 °C for 1 h while MaSDeS was incubated at 37 °C for 24 h due to its slow degradation¹⁰. All samples without (NS) or with the surfactants (MaSDeS, PM, RG, SDS, Azo, OG, DDM, DGT) were buffer exchanged into 10 % ACN, 10 % IPA, and 1 % FA using a MWCO filter (3 x 100 μL) and adjusted to the original volume of 50 μL . Protein were separated using the following gradient: 0-5 min 20% B, 5-30 min 20-65% B, 30-35 min 65% B, 35-36 min 20% B, 36-40 min 20% B. Column temperature was 60

°C. Column temperature was 60 °C. The methods described here correspond to the data presented in **Supplementary Figure 12**.

2.3.18. Data Analysis.

Data were analyzed and processed in DataAnalysis 4.3 (Bruker Daltonics). A msalign file was created using the SNAP peak picking algorithm with the following parameters: quality factor (0.4); S/N (3); intensity threshold (500); retention window (1.5 min). The file contained the following information: precursor mass, precursor charge, precursor mass followed by the fragment masses, intensities, and charges. TopPIC²⁷ was utilized for intact protein identification based on protein spectrum matches searching against the UniProt *Sus scrofa* (released on Nov. 22nd, 2017; containing 26817 protein sequences) or *Homo sapiens* (released on Dec. 20th, 2017; containing 20244 reviewed protein sequences) database²⁸. Fragment mass tolerance was set to 15 ppm. All identifications were validated with statistically significant P and E values (<E-5) and satisfactory numbers of assigned fragment ions (>6). Sequence mass determination and validation were performed using Mash Suit Pro²⁹ or ProSight Lite³⁰. The corresponding MS and MS/MS data were summarized in **Supplementary Table S4-10**. UniProt gene ontology³¹ was used to determine the subcellular location of the identified proteins which were then graphed in Excel. String analysis software³² was used to create an interactome map of identified proteins belonging to the electron transport chain.

The proteoform maps were generated as follows: (1) LC-MS scans were averaged every min; (2) deconvoluted using Maximum Entropy algorithm (Resolution: 80,000; mass range: 5,000-60,000 Da); (3) mass list outputs were generated using SNAP peak picking (quality factor: 0.8,

S/N: 3, absolute intensity 1,000) as described previously³³. A graphic map was then generated in Microsoft Excel based on the first retention time and the monoisotopic mass generated from SNAP. The methods described here correspond to the data presented in **Supplementary Figure 13**.

2.3.19. Statistical Analysis.

For the protein solubility experiment (**Figure 1f**) comparing Azo, SDS, DDM, and MaSDeS, three independent protein assays (n=3) were performed to evaluate surfactant performance. Error bars represent standard error of the mean. For the broader protein solubility comparison (**Supplementary Fig. 10**), the data presented were based on three independent experiments (n=3). Error bars represent standard error of the mean. For LC-MS analysis (**Supplementary Fig. 9**), three separate samples (n=3) were prepared for each condition. Error bars represent standard error of the mean.

2.3.20. Supplementary Figures

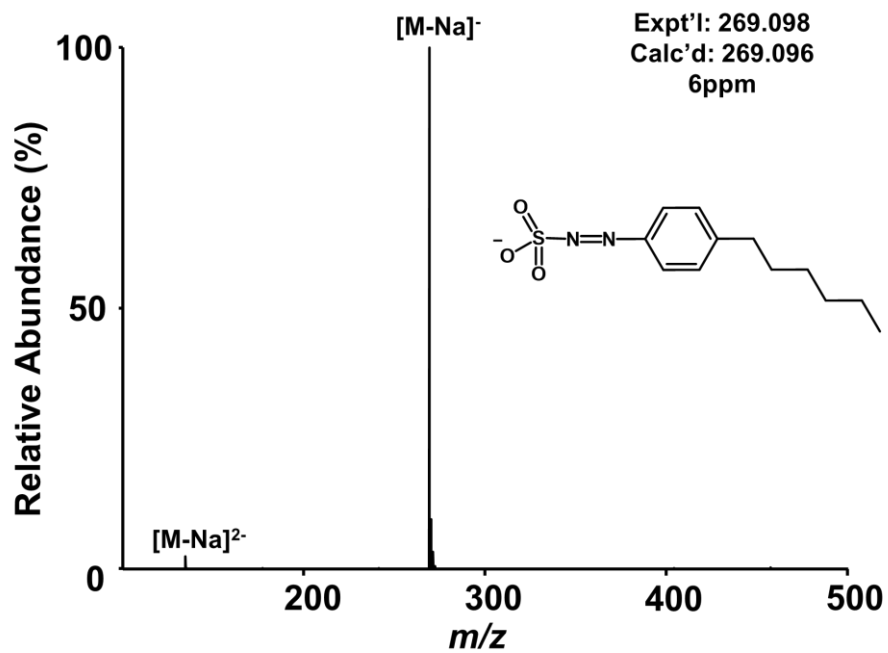


Figure 2.3. Negative ion ESI-FTICR mass spectrum of Azo. ESI-MS for Azo ($C_{12}H_{17}N_2O_3SNa$), $[M-Na]^-$, calculated m/z : 269.096, observed m/z : 269.098. Data is representative of four independent experiments.

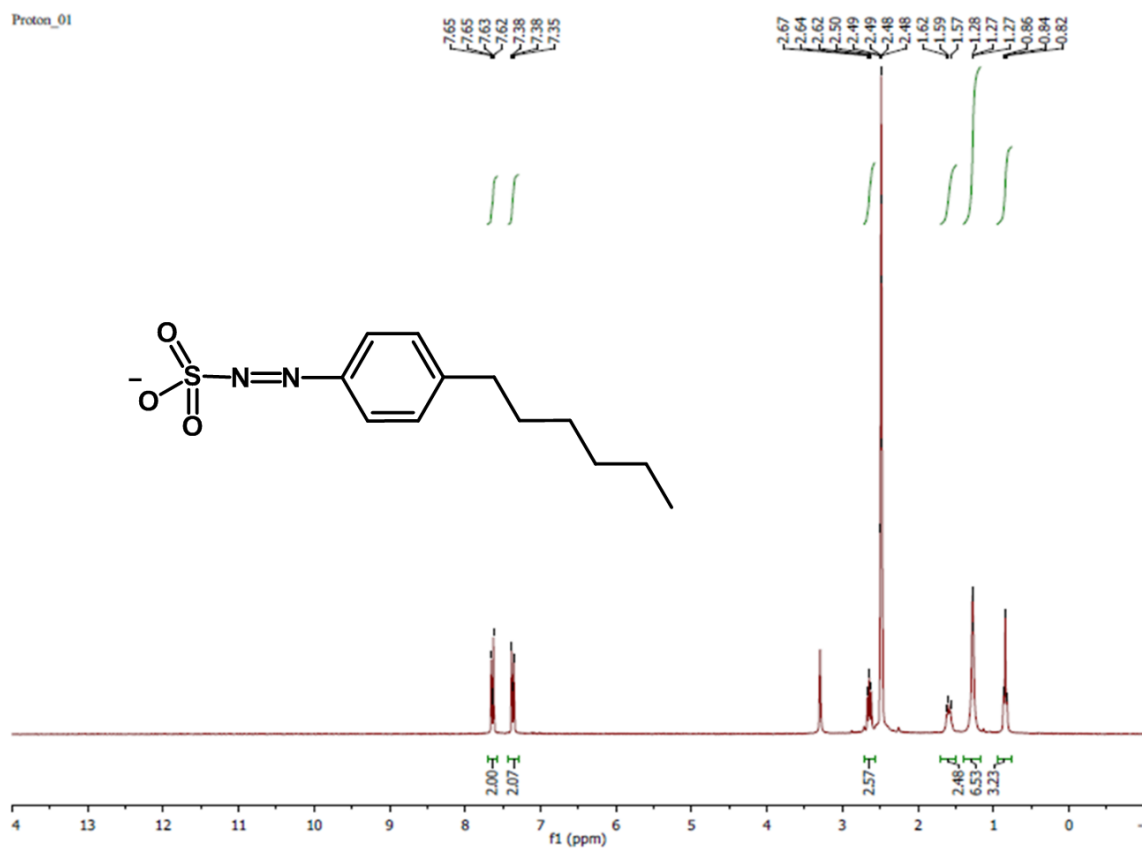


Figure 2.4. ^1H -NMR spectrum of Azo. ^1H NMR: δ 7.64 (2H, dd, -Ar-H), 7.37 (2H, d, -Ar-H), 2.67-2.48 (2H, m, -Ar-CH₂), 1.61 (2H, t, -Ar-CH₂CH₂), 1.28 (6H, t, -(CH₂)₃-CH₃). NMR analysis was performed one time to confirm the structure.

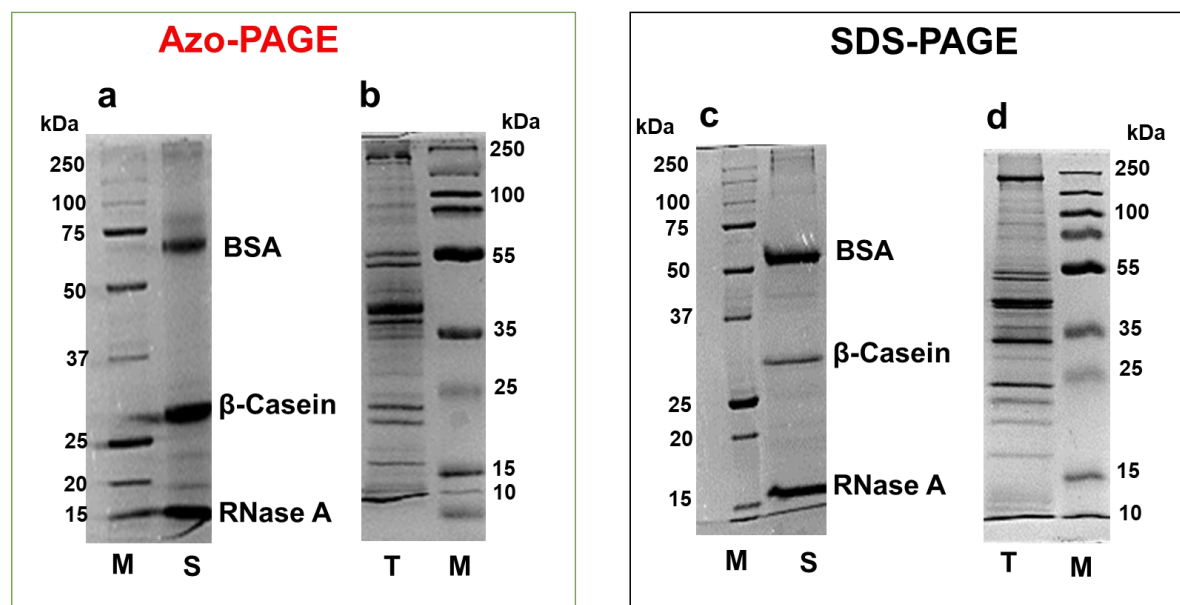


Figure 2.5. Polyacrylamide gel electrophoresis (PAGE) using Azo (a,b) and SDS (c, d). Azo-PAGE analysis of (a) 2.5 μ g bovine serum albumin (BSA), β -casein, and ribonuclease A (RNase A) and (b) myofilament cardiac tissue protein extract with Coomassie blue staining. SDS-PAGE analysis of (c) 2.5 μ g BSA, β -Casein, and RNase A and (d) myofilament cardiac tissue protein extract with Coomassie blue staining. M: molecular weight ladder; S: standard proteins, T, myofilament protein extract. Data are representative of two independent Azo-PAGE/SDS-PAGE comparisons.

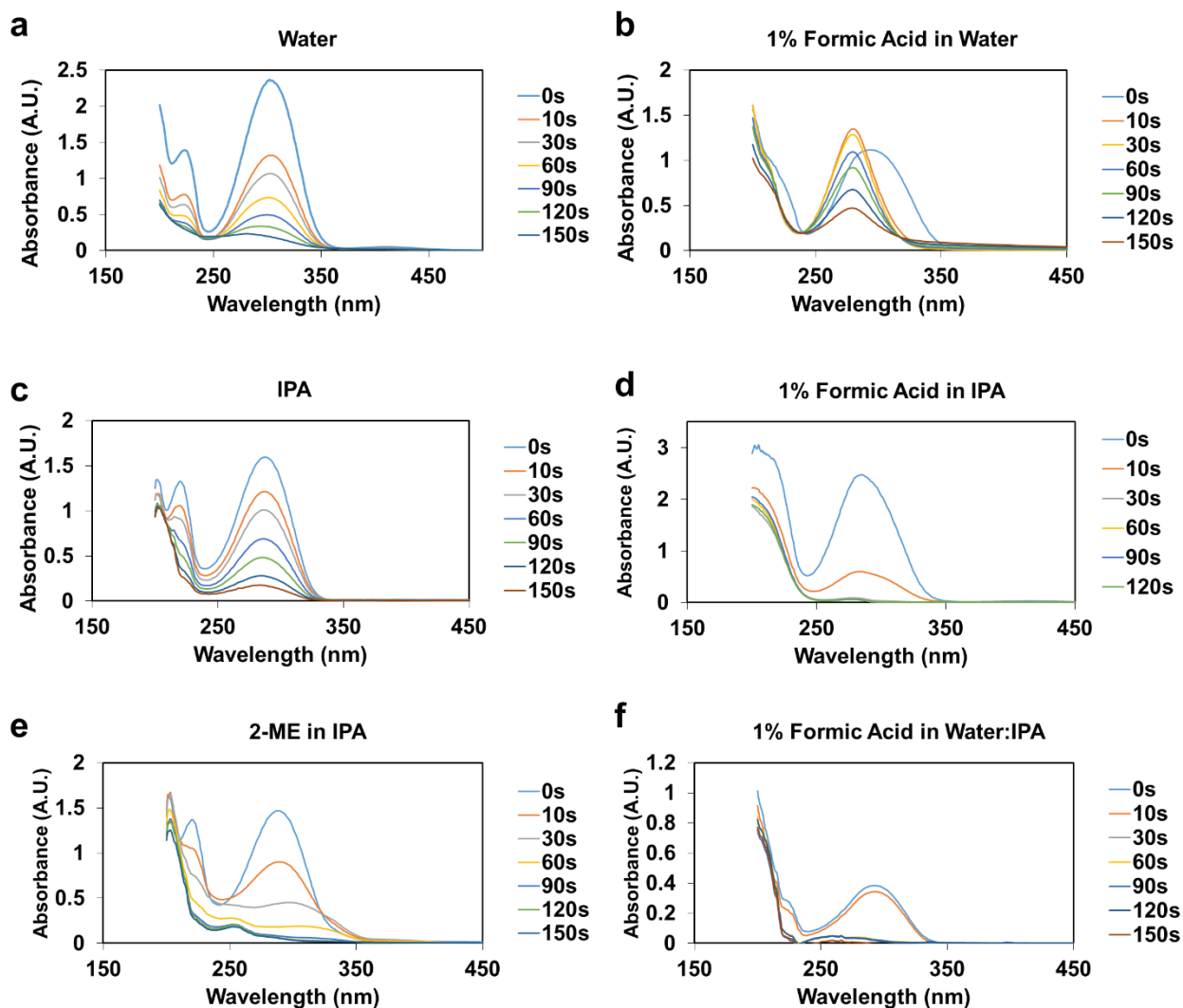


Figure 2.6. Effects of solvent conditions on Azo degradation kinetics. Azo (0.1%) was degraded in different solvent conditions; (a) water, (b) 1% formic acid in water, (c) IPA, (d) 1% formic acid in IPA, (e) 2-ME in IPA, (f) 1% formic acid in 50:50 mixture of water: IPA, respectively, to probe their effects on degradation kinetics. Overall the surfactant degraded rapidly particularly in the presence of organic solvent at low pH (d, f). Conditions in panel a and d have been repeated in three independent experiments. Conditions in b, c, e, and f are performed one time.

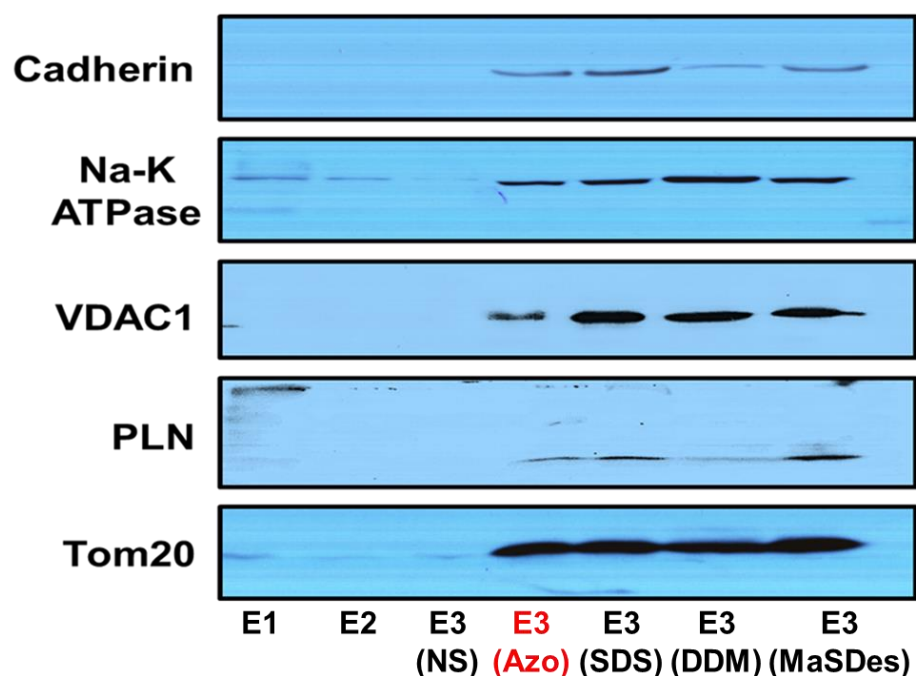


Figure 2.7. Evaluation of surfactant-aided extraction of cardiac membrane proteins using western blot analysis. Western blot analysis of common cardiac membrane proteins, cadherin, sodium-potassium adenosine triphosphate (Na-K ATPase), voltage-dependent anion-selective channel (VDAC), phospholamban (PLN), and mitochondrial import receptor subunit (TOM20). 12.5% SDS-PAGE was used to resolve 20 μ L of lysate extracted using 25 mM NH_4HCO_3 without surfactant (NS), or with the surfactant, 4-hexylphenylazosulfonate (Azo), sodium dodecyl sulfate (SDS), dodecyl β -D-maltoside (DDM), and MS-compatible degradable surfactant (MaSDeS), respectively. The addition of Azo in the buffer successfully aided in the extraction of these five important integral membrane proteins in cardiac tissue, with performance comparable to SDS and MaSDeS. Data are representative of two individual experiments for Cadherin, VDAC1, PLN, and TOM20.

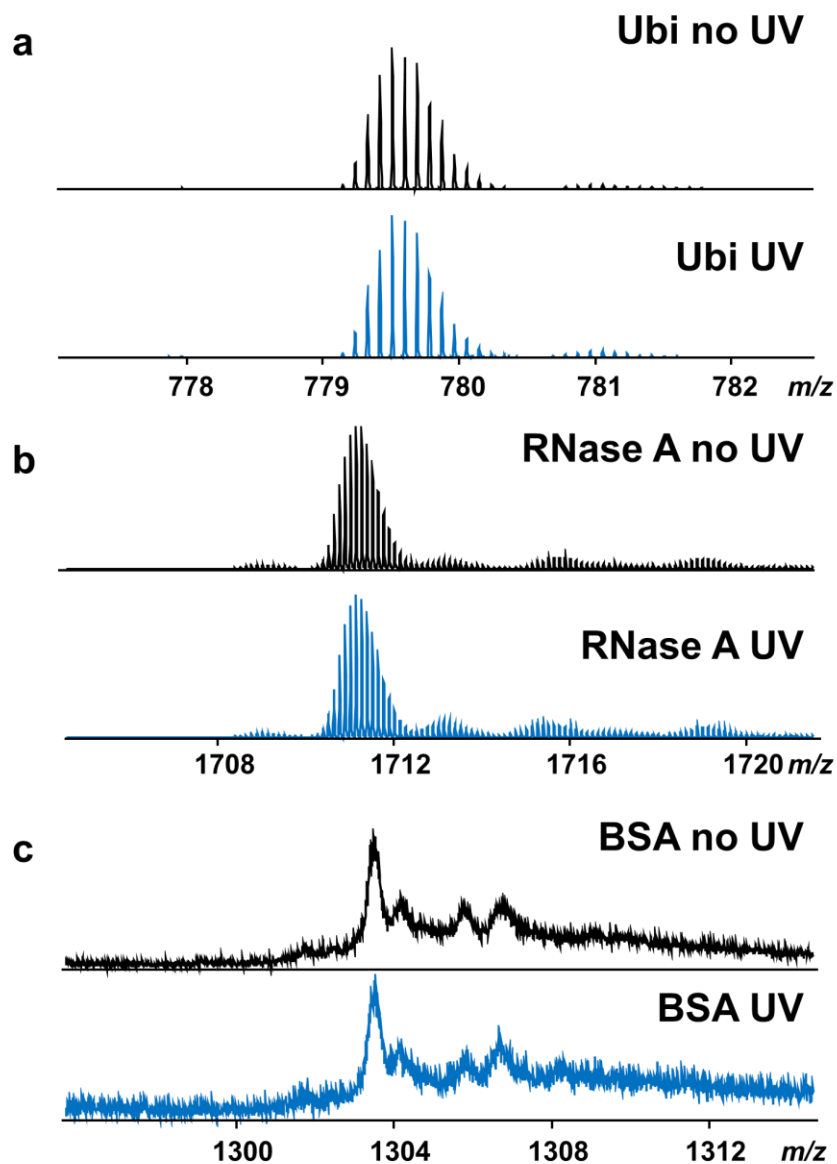


Figure 2.8. Effect of UV exposure on protein mass spectrum. (a) Ubi, (b) RNase A, (c) BSA were irradiated for 3 min before direct injection into the MS. Overall no change was observed after irradiation. The mass spectra were collected on a Bruker maXis II Q-TOF mass spectrometer. Data are presented from one dataset for three different proteins.

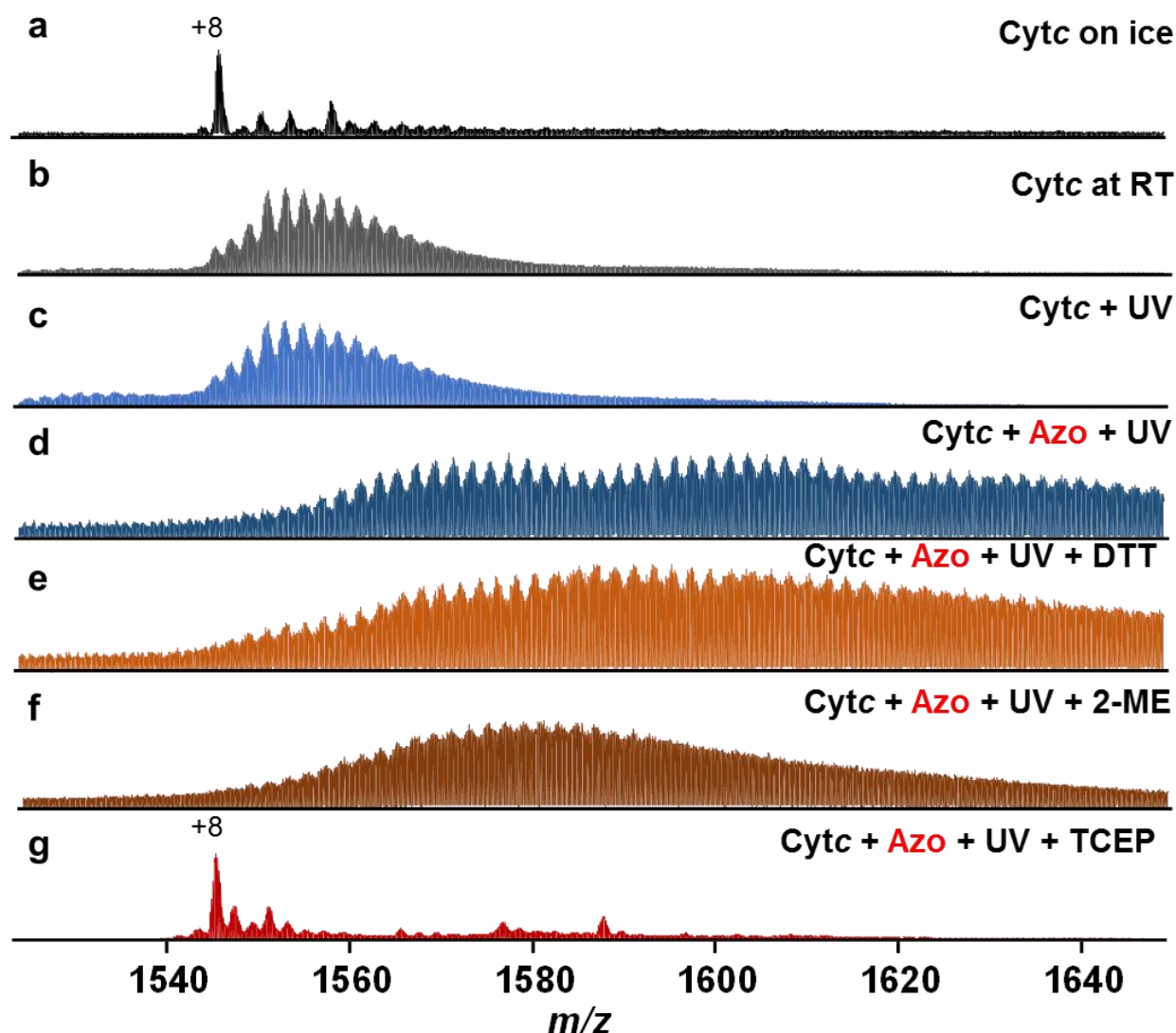


Figure 2.9. The effect of the addition of reducing agents during Azo degradation. A standard protein, Cyt c, was (a) incubated on ice, (b) incubated on room temperature (RT), (c) irradiated at RT, (d) irradiated at RT with Azo, (e) irradiated at RT with Azo and 50 mM DTT, (f) irradiated at RT with Azo and 50 mM 2-ME, and (g) irradiated at RT with Azo and 50 mM TCEP before LC-MS analysis. All the incubations were done for 3 min. Overall TCEP alleviated oxidative modification observed when incubated at RT or by exposing it to the radicals generated upon photolysis of the surfactant. Mass spectra were collected on a Bruker maXis II Q-TOF mass spectrometer. Data are presented based on one experimental design.

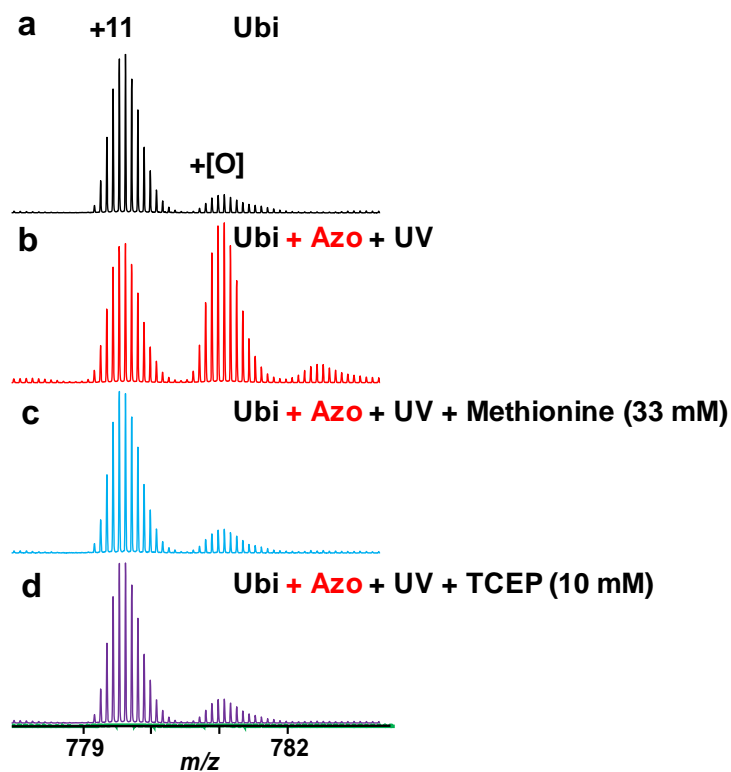


Figure 2.10. The addition of methionine during Azo degradation alleviated oxidative modification. A standard protein, ubiquitin (Ubi), was (a) incubated on ice, (b) irradiated at room temp with Azo (0.1%), (c) irradiated at room temp with Azo (0.1%) and Methionine (33 mM), (d) irradiated at room temp with Azo (0.1%) and TCEP (10 mM). Mass spectra were collected on a Bruker maXis II Q-TOF mass spectrometer. Data are repeated in two independent experiments with similar results.

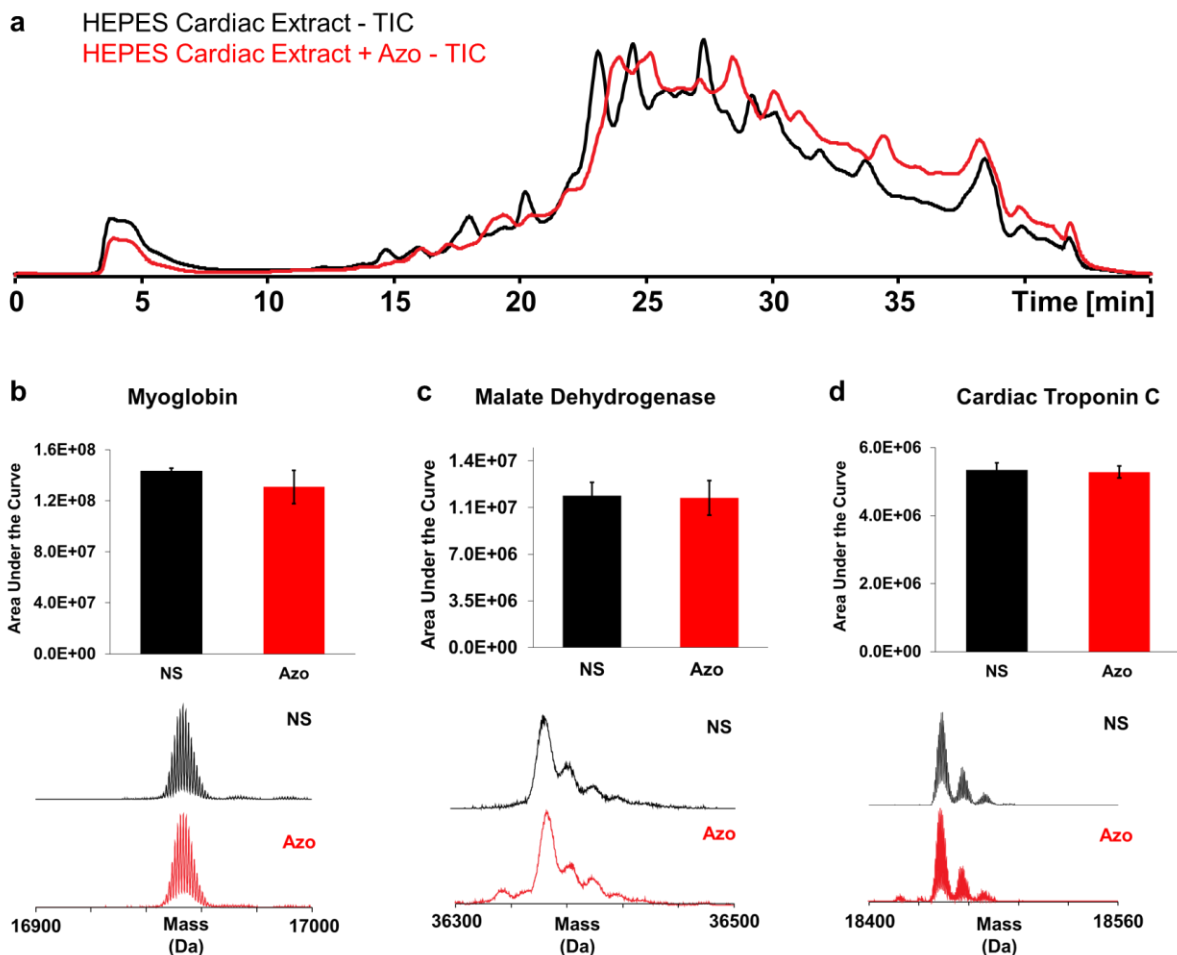


Figure 2.11. Effect of Azo on relative quantitation by LC-MS using HEPES Cardiac Extract. A sample of cardiac proteins extracted using HEPES buffer was analyzed by LCMS as well as a sample spiked with Azo (0.2%). (a) Total ion chromatogram for HEPES Cardiac Extract and Cardiac Extract + Azo. Extracted ion chromatograms of the top 5 charges states were performed and the area under the curve calculated for (b) myoglobin, (c) malate dehydrogenase, and (d) cardiac troponin C. Error bars represent standard error of the mean. Mass spectra were collected on a Bruker maXis II Q-TOF mass spectrometer. Data are representative of three independent experiments.

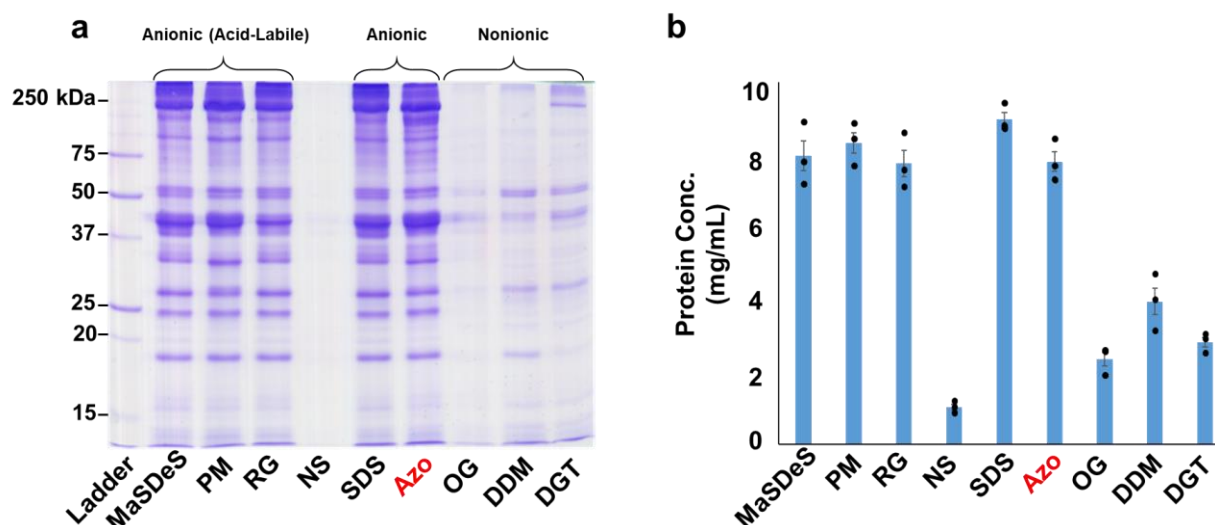


Figure 2.12. Evaluation of surfactant-aided extraction of cardiac tissue using SDS-PAGE and protein assay. Swine heart proteins were extracted using 0.5% surfactant in 25 mM NH_4HCO_3 following a HEPES extraction to remove water-soluble proteins. (a) MS-compatible degradable surfactant (MaSDeS), ProteaseMaxTM (PM), RapiGestTM (RG), 25mM NH_4HCO_3 with no surfactant (NS), sodium dodecyl sulfate (SDS), 4-hexylphenylazosulfonate (Azo), Octyl β -D-glucopyranoside (OG), dodecyl β -D-maltoside (DDM), digitonin (DGT) were tested. (b) Protein assay used to assess the extracted protein concentration facilitated by each surfactant. Data are presented based on three independent experiments ($n=3$). Error bars represent standard error of the mean.

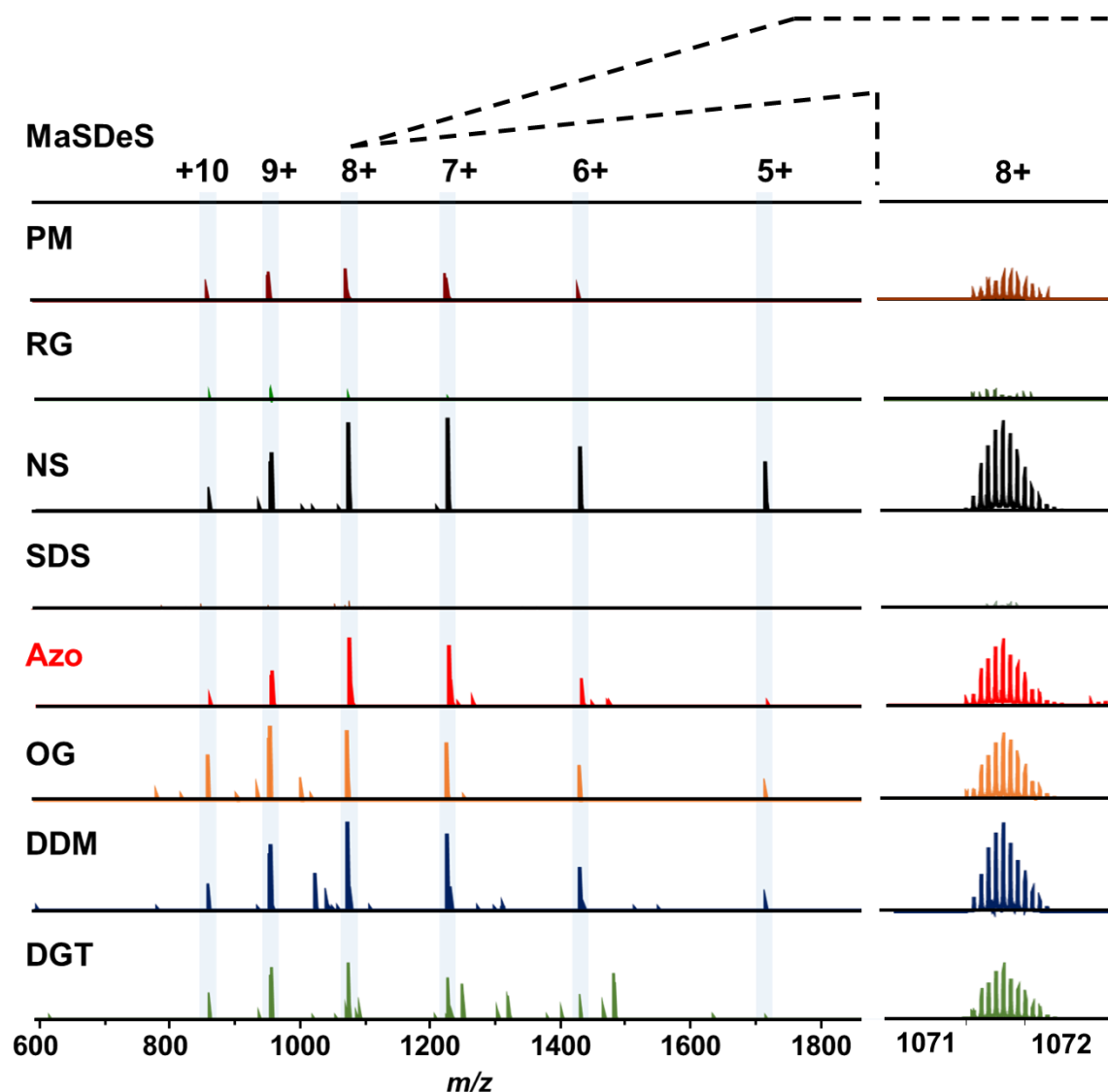


Figure 2.13. Direct MS comparison of common surfactants. MS-compatibility evaluation of Azo as compared to commonly used surfactants. 0.1% surfactant was added to ubiquitin in 25 mM NH_4HCO_3 solution and analyzed by high-resolution mass spectrometry and compared to controls with no surfactant (NS). Surfactants including MS-compatible slowly-degradable surfactant (MaSDeS), ProteaseMaxTM (PM)¹⁵⁶, RapiGestTM (RG), sodium dodecyl sulfate (SDS), 4-hexylphenylazosulfonate (Azo), Octyl β -D-glucopyranoside (OG), dodecyl β -D-maltoside (DDM), and digitonin (DGT) were tested. The mass spectra were collected on a Bruker 12 T FTICR mass spectrometer and normalized to the maximum intensity in the NS (control) spectrum of 8E10. Data are representative of three independent experiments.

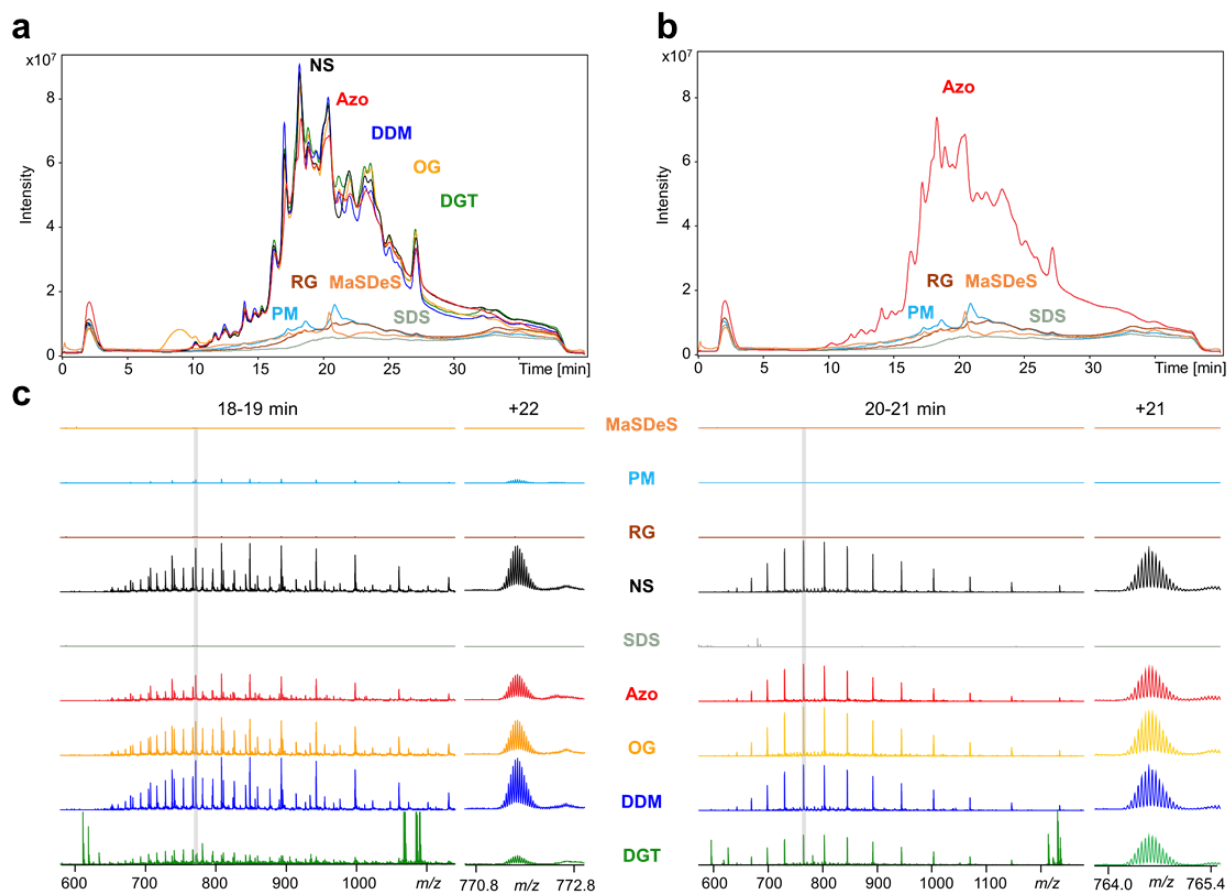


Figure 2.14. Comparison of surfactant effect on LC-MS analysis of cardiac tissue lysate. Cardiac proteins were extracted using NH_4HCO_3 buffer. Surfactants (final concentration of 0.1% in 25 mM NH_4HCO_3) were added to the sample and 25 mM NH_4HCO_3 buffer was added as a control (no surfactant, NS). The surfactants include MS-compatible slowly-degradable surfactant (MaSDeS), ProteaseMaxTM (PM), RapiGestTM (RG, also known as ALS), sodium dodecyl sulfate (SDS), 4-hexylphenylazosulfonate (Azo), Octyl β -D-glucopyranoside (OG), dodecyl β -D-maltoside (DDM), and digitonin (DGT). The acid- and photo-cleavable surfactants were degraded under acid or UV condition respectively. All the samples were buffer exchanged and analyzed by LC-MS. Mass spectra were collected on a Bruker maXis II Q-TOF mass spectrometer. (a) Total ion chromatograms comparing the MS signal of cardiac tissue lysate in the presence of different surfactants compared to control. (b) Total ion chromatograms comparing the MS signal of cardiac tissue lysate in the presence of different anionic surfactants. (c) Representative mass spectra from proteins eluted at 18-19 min showing multiple co-eluting proteins with intensity normalized to $1.36\text{E}4$. (d) Representative mass spectra of a protein eluted from 20-21 min with intensity normalized to $1.62\text{E}5$. Data are presented based on one series of LC-MS experiments with comparisons of Azo to multiple surfactants.

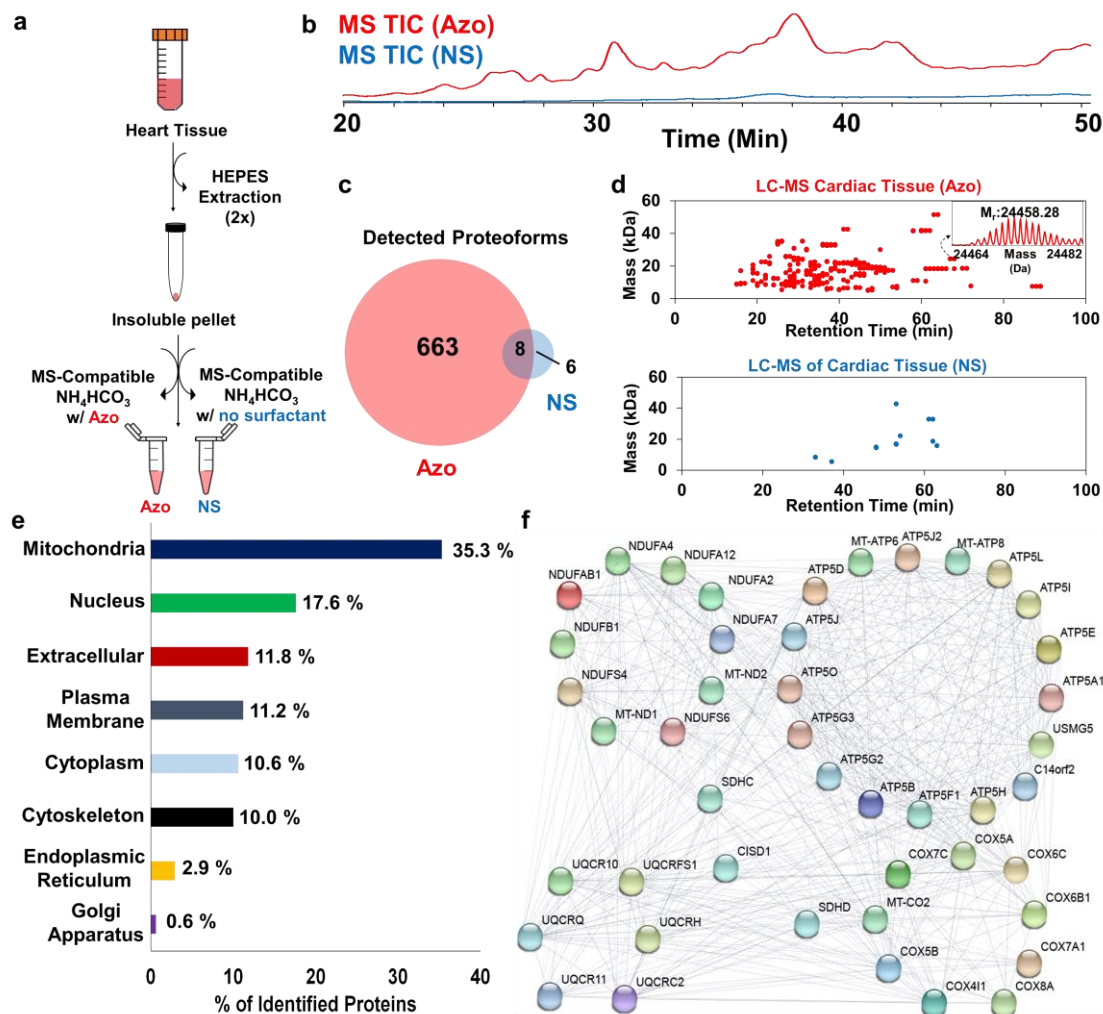


Figure 2.15. Photocleavable Azo-enhanced top-down proteomics. (a) Schematic depicting the sequential tissue extraction procedures. Proteins from cardiac tissue were extracted using HEPES buffer (2x) to deplete the cytosolic proteins, followed by an extraction using NH_4HCO_3 solution with or without 0.5% Azo. (b) Comparison of the total ion chromatogram (TIC) of equal injection volumes of the final cardiac extraction with or without 0.5% Azo. (c) Comparison of detected proteoforms in a single RPLC-MS run of proteins extracted with or without Azo. (d) Intact proteoform mass map from cardiac tissue extraction with Azo or without Azo (NS). (e) Subcellular location of proteins identified in cardiac tissue in the Azo-enhanced top-down proteomics combining 3 RPLC-MS/MS runs. (f) Interactome map of the identified proteins with Azo that belong to the electron transport chain. Figures here are based on the data in Supplementary Table 2-8. Data in c and d are representative of three independent experiments and data in e and f are the combined results of three independent LC-MS/MS experiments.

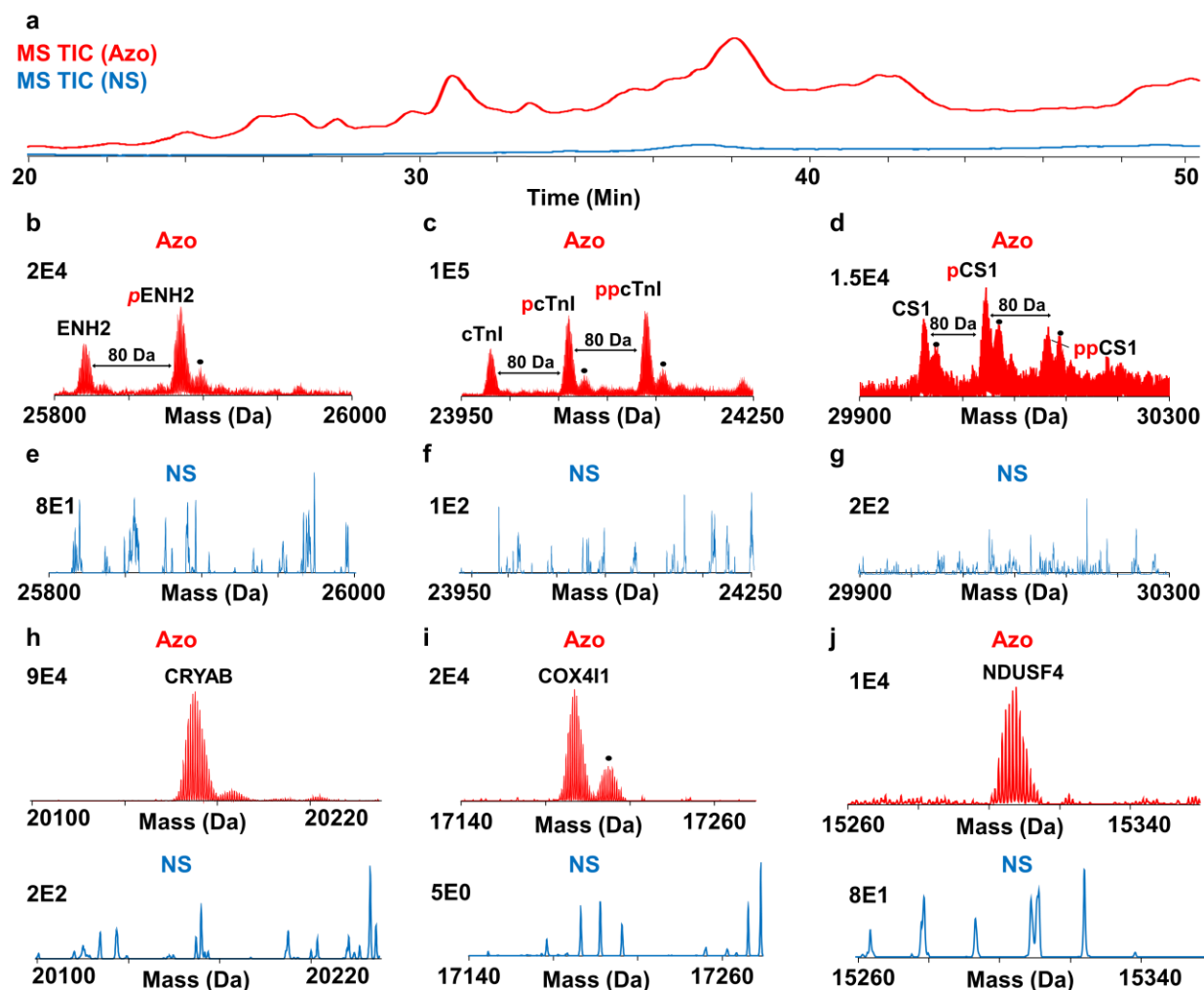
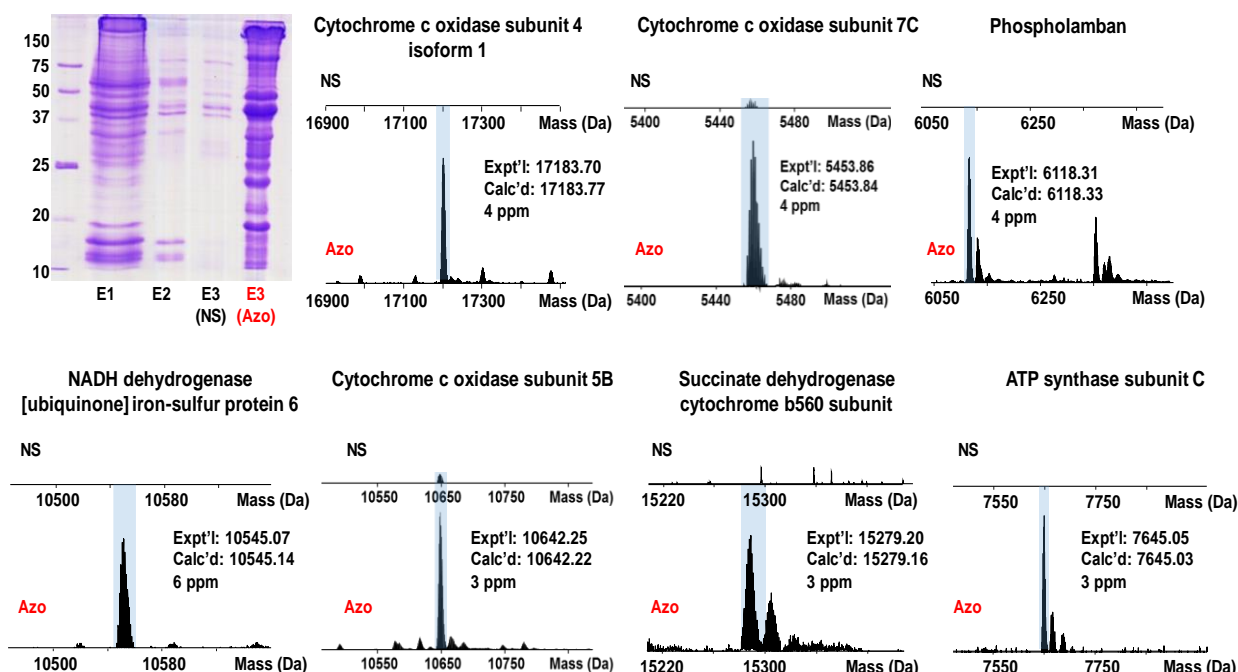


Figure 2.16. Comparison of cardiac tissue extraction using MS-compatible NH_4HCO_3 with or without Azo in a single RPLC/MS run. (a) Total ion chromatogram with and without Azo. Representative spectra includes (b) enigma homolog 2 (ENH2), (c) troponin I (cTnI), (d) calsarcin-1 (CS-1), (h) alpha-crystallin b chain (CRYAB), (i) cytochrome c oxidase subunit 4 isoform 1 (COX4I1), and (j) NADH dehydrogenase Fe-S protein 4 (NDUSF4). These proteins were detected in the Azo-extraction but not detected in the control without Azo. Mass spectra were collected on a Bruker maXis II Q-TOF mass spectrometer. NS: no surfactant (serving as controls). Data are representative of three independent experiments.



Supplementary Figure 15. Azo-enable membrane proteomics from cardiac tissue. (a) SDS-PAGE analysis of cardiac proteins extracted first in a high salt buffer twice and labeled as extraction 1 (E1) and extraction 2 (E2), respectively, followed by a third extraction (E3) in 25 mM NH_4HCO_3 with either no surfactant (NS) or 0.5% Azo. (b-h) Representative deconvoluted mass spectra from LC-MS analysis using an equal volume of extract (NS and Azo) confirming the enhancement of membrane protein signal that occurs by adding Azo to the extraction buffer. (b) Cytochrome c oxidase subunit 4 isoform 1. (c) Cytochrome c oxidase subunit 7C. (d) Phospholamban. (e) NADH dehydrogenase [ubiquinone] iron-sulfur protein 6. (f) Succinate dehydrogenase cytochrome b560 subunit. (h) ATP synthase subunit c. Mass spectra were collected on a Bruker maXis II Q-TOF mass spectrometer. Each spectrum was normalized to the intensity value corresponding to protein signal intensity in the Azo extraction. Data are representative of two independent experiments.

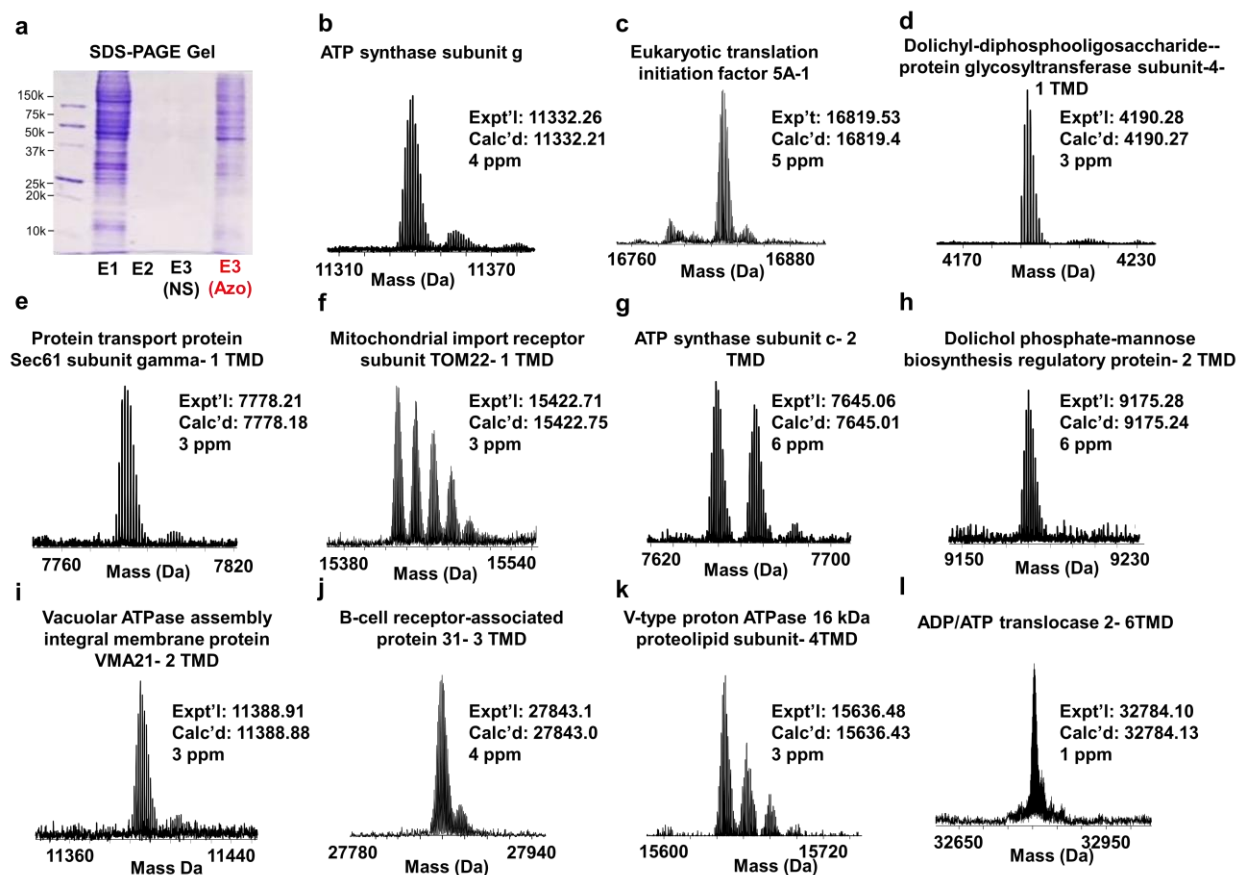


Figure 2.17. Representative membrane proteins identified from human embryonic kidney (HEK) Cells. (a) SDS-PAGE analysis of HEK cell proteins extracted with a Tris buffer and labeled as extraction 1 (E1). The pellets were next extracted with NH_4HCO_3 and labeled as extraction 2 (E2). Finally, a third extraction (E3) was performed using 25 mM NH_4HCO_3 with either no surfactant, E3 (NS), or 0.5% Azo, E3 (Azo). Representative membrane proteins identified in the E3 (Azo) extract by top-down proteomics include (b) ATP synthase subunit g, (c) eukaryotic translation initiation factor 5A, (d) dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit 4, (e) protein transport protein Sec61 subunit gamma, (f) mitochondrial import receptor subunit Tom22, (g) ATP synthase subunit c, (h) dolichol phosphate-mannose biosynthesis regulatory protein, (i) vacuolar ATPase assembly integral membrane protein VMA21, (j) B-cell receptor-associated protein 31, (k) V-type proton ATPase 16 kDa proteolipid subunit, and (l) ADP/ATP translocase 2. Mass spectra were collected on a Bruker maXis II Q-TOF mass spectrometer. Data are presented from one LC/MS/MS experiment.

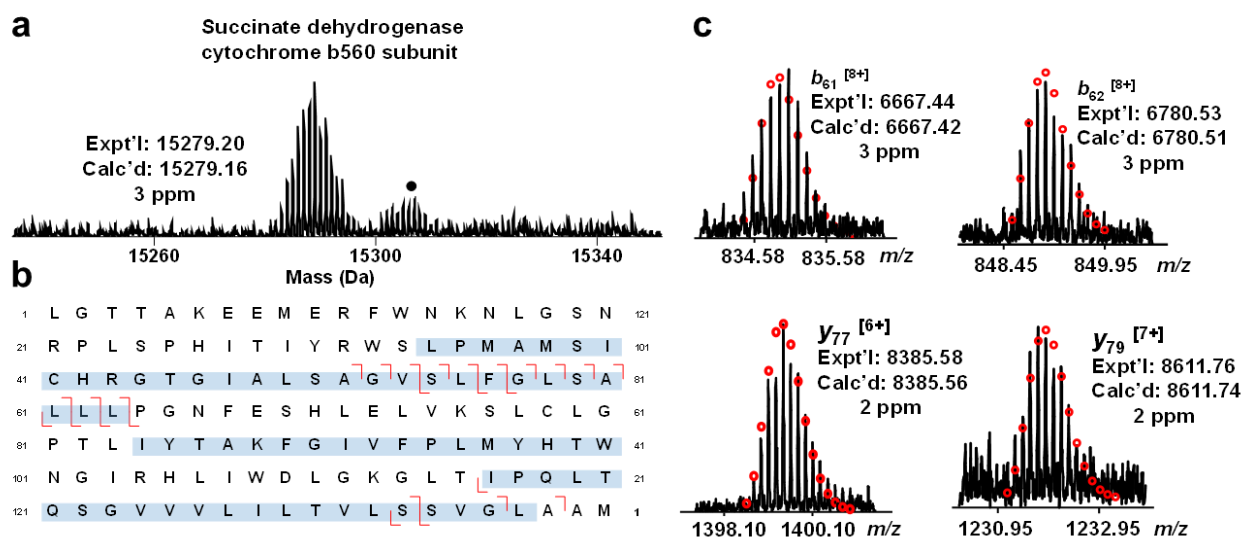


Figure 2.18. Identification of succinate dehydrogenase cytochrome b560 subunit by top-down proteomics enabled by Azo. (a) Deconvoluted mass spectrum of succinate dehydrogenase cytochrome b560 acquired with high mass accuracy using a Bruker maXis II Q-TOF mass spectrometer. (b) Sequence fragmentation map with blue regions indicating the three transmembrane domains. (c) Representative MS/MS fragments with collision-induced dissociation (CID). Data are representative of two independent experiments.

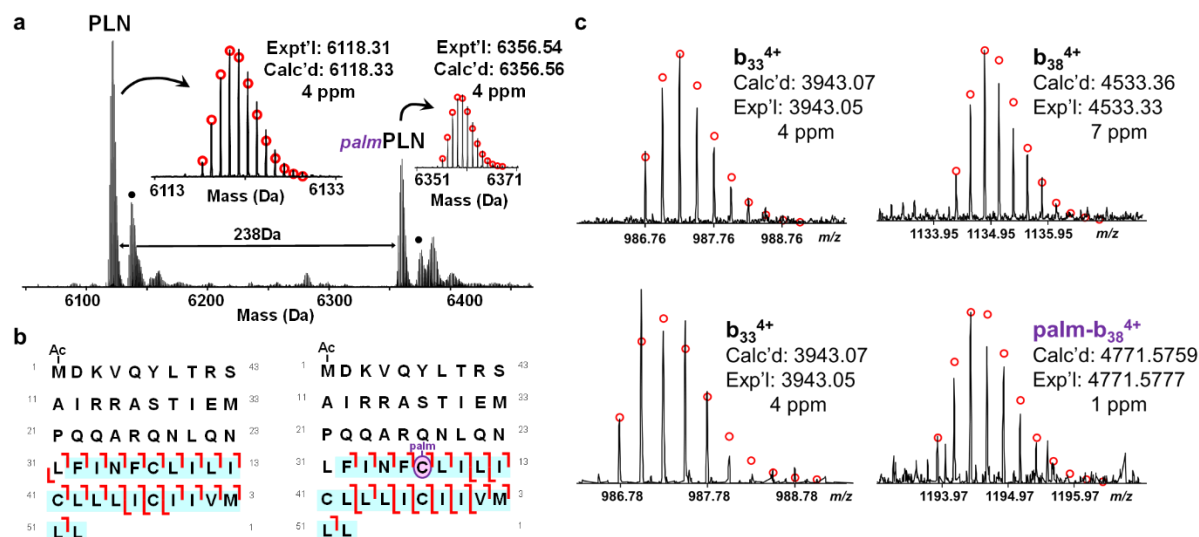


Figure 2.19. Localization of palmitoylation (palm) on phospholamban (PLN) by online LC-MS/MS with CID. (a) Deconvoluted mass spectrum showing high-accuracy detection of PLN and its highly abundant palmitoylated proteoform acquired on a Bruker maXis II Q-TOF mass spectrometer; (b) CID fragmentation map of both PLN and palm-PLN with blue regions designating the transmembrane domain; (c) representative fragment ions for the palmPLN showing the localization of the palmitoylation to the cysteine 36 residue. Red circles show theoretical isotopic distribution. Data are representative of two independent experiments.

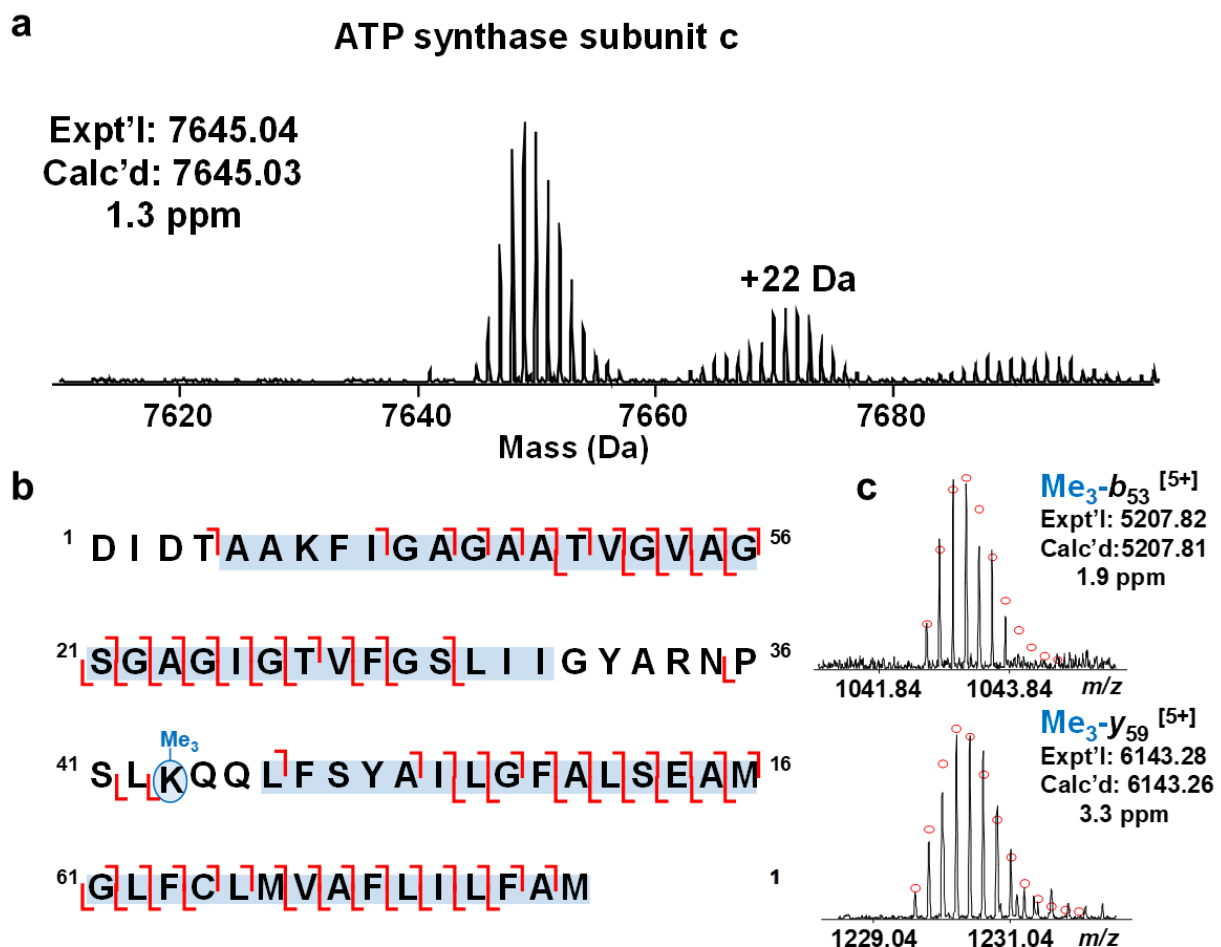


Figure 2.10. Localization of trimethylation on ATP synthase c by online LC-MS/MS with CID. (a) Deconvoluted mass spectrum showing the high-accuracy detection of ATP synthase subunit c acquired on a Bruker maXis II Q-TOF mass spectrometer; (b) CID fragmentation map with blue regions designating the two transmembrane domains; (c) representative fragment ions measured with high mass accuracy localizing the trimethylation (Me₃) to Lysine 43 residue in ATP synthase c. Red circles show theoretical isotopic distribution. Data are representative of two independent experiments.

2.3.22 Supplementary Tables

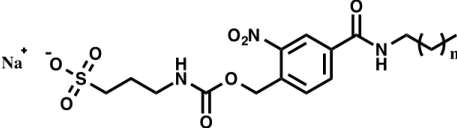
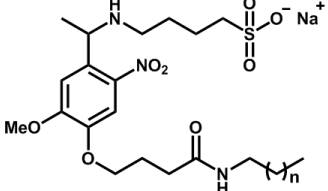
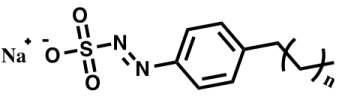
Surfactant			Water Solubility at RT	Protein Solubility
	ONB	C6	Slightly Soluble	Poor
		C8	Relatively Insoluble	Poor
		C12	Relatively Insoluble	Poor
	ONV	C8	Soluble	Fair
		C10	Soluble	Poor
		C12	Soluble	Poor
	AZO	C6	Soluble	Good
		C8	Soluble	Fair
		C10	Slightly Soluble	Poor
		C12	Slightly Soluble	Poor

Table 2.1. Protein solubility screening of a library of photo-cleavable surfactants. We synthesized and evaluated a library of photo-labile surfactants by inserting a photo-cleavable moiety in between the hydrophilic head and hydrophobic tail (**Fig. 1a**) aiming for rapid, controlled cleavage and subsequent loss of surface activity. Specifically, O-nitrobenzyl (ONB), O-nitroveratryl (ONV), and azobenzene (AZO) groups were chosen as the photo-cleavable linkers because of their rapid degradation kinetics¹⁵¹⁻¹⁵⁴. We mimicked the properties of the most widely used biological surfactant, SDS, by using a sulfonate head for the hydrophilic portion to maximize protein solubility. Various hydrocarbon chain lengths for the hydrophobic tails were synthesized to optimize the hydrophilic-lipophilic balance of the surfactants. Specifically, surfactants with ONB, ONV, and AZO cleavable groups were synthesized with varying chain lengths from 6-12 carbons (C6-12). Each synthesized surfactant was evaluated for its solubility in water first. Next, the surfactants were used to solubilize swine heart proteins after HEPES buffer was used to deplete the water-soluble species. Overall AZO (C6) or Azo was the only surfactant, which not only has excellent solubility in water at room temperature (RT) but can effectively solubilize proteins.

* Supplementary Tables 2-10 are presented in the Excel spreadsheets.

2.3.23 Determine the effect of the UV-irradiation on MS analysis of proteins

We have systematically examined the effect of the UV-irradiation on MS analysis of proteins. We found that UV irradiation alone neither introduced any adduct nor caused any alteration to the protein mass spectrum (**Supplementary Fig. 6**). Moreover, we demonstrated the inclusion of reducing agent, tris(2-carboxyethyl)phosphine hydrochloride (TCEP), could successfully minimize oxidation modifications possibly resulting from the exposure of the proteins to the radicals generated upon photolysis of the surfactant¹⁵¹ or heating at room temperature (**Supplementary Fig. 7**). Alternatively, free methionine could be added to minimize oxidation modification and preserve biologically relevant modifications such as disulfide and glutathionylation (**Supplementary Fig. 8**). We also observed that the usage of Azo had nearly no effect on the relative quantitation of intact proteins from a cardiac tissue lysate by reversed-phase liquid chromatography (RPLC)-MS (**Supplementary Fig. 9**).

2.3.24 A Systematic Comparison of Azo with Commonly Used Surfactants

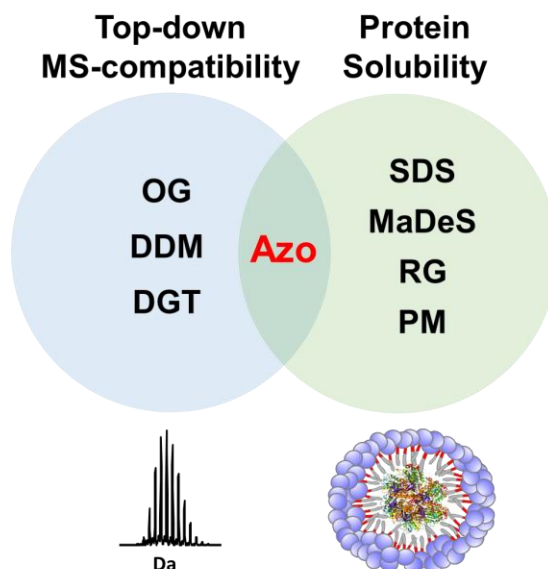
We performed a systematic comparison of Azo with a broader range of commonly used surfactants to evaluate their ability to solubilize proteins from the insoluble cardiac tissue pellets and subsequently assessed their MS-compatibility using both standard proteins and tissue lysates (**Supplementary Fig. 10-12**). The surfactants used for comparison include non-ionic surfactants, octyl β -D-glucopyranoside (OG) and dodecyl β -D-maltoside (DDM), digitonin (DGT), as well as anionic, acid-labile surfactants, RapiGestTM (RG)²⁷ (also known as ALS)¹⁵⁰, ProteaseMaxTM (PM)¹⁵⁶, and MS-compatible slowly-degradable surfactant (MaSDeS)²⁹, together with SDS and Azo. Non-ionic saccharide surfactants such as DDM and OG can be MS-compatible when used at

lower concentrations (0.01-0.1%)^{21, 155} but they are considered to be relatively mild with limited solubilization ability²⁰. On the other hand, anionic surfactants such as SDS and acid-labile SDS mimics are much stronger, capable of solubilizing and denaturing proteins with high efficiency²⁰. Acid-labile surfactants were originally developed for improving in-gel or in-solution digestion efficiency for bottom-up proteomics²⁹.

Utilizing SDS-PAGE (**Supplementary Fig. 10a**) and Bradford protein assay (Bio-Rad) analysis (**Supplementary Fig. 10b**), we observed that anionic surfactants including SDS, MaSDeS, PM, RG, and Azo, showed drastically better protein solubilization ability than the non-ionic surfactants, OG, DDM, and DGT. In particular, Azo effectively extracted proteins at a level similar to other leading anionic surfactants such as SDS, MaSDeS, RG, and PM.

Subsequently, we performed a comparison of the MS-compatibility of Azo with these leading surfactants (**Supplementary Fig. 11**). The MS analysis in the presence of 0.1% surfactants, showed that SDS and MaSDeS dramatically suppressed the MS signal of the intact Ubi; RG and PM significantly suppressed the MS signal of Ubi; whereas Azo and non-ionic surfactants, OG, DDM, and DGT showed minimal MS protein signal suppression.

Importantly, we also conducted a broader comparison of the LC-MS-compatibility of Azo with other commonly used surfactants using cardiac tissue lysate (**Supplementary Fig. 12**). The LC-MS analysis of the lysate with 0.1% surfactants demonstrated great top-down MS-compatibility of Azo, OG, DDM, and to a lesser extent DGT. On the other hand, significant signal suppression was observed in the presence of SDS and all acid cleavable surfactants, PM, RG, and MaSDeS. Overall, these broader comparison studies have demonstrated Azo's unique ability to effectively solubilize proteins without interfering with downstream top-down MS analysis.



2.3.25 Extended discussion

Our study provides a streamlined method for analyzing intact proteins that are difficult to extract and solubilize, thus previously inaccessible for top-down proteomic analysis. We have demonstrated the effectiveness of Azo-enabled top-down membrane proteomics, which significantly increased the number of membrane proteins that can be confidently identified in MS analysis and allowed for comprehensive characterization of membrane protein complexes.

In this study, although we have detected a total of 2836 proteoforms based on accurate mass measurements (**Supplementary Table 3**) from the combination of three LC/MS runs, we only identified 388 proteoforms based on one-dimensional online RPLC-MS/MS data (**Supplementary Table 4**) representing 171 proteins (**Supplementary Table 5**) from mitochondria, nucleus, plasma membrane, cytoskeleton, endoplasmic reticulum, cytoplasm, and extracellular region (**Supplementary Fig. 13e,f**). The low proportion of protein identification is not a result of proteins being modified by UV photolysis reactions or Azo-induced artefactual

modification as demonstrated (**Fig. 1g & Supplementary Fig. 6-8**). Instead, it is a result of the current limitation on protein separation, fragmentation, data acquisition, and identification algorithms in top-down proteomics¹⁴.

Top-down proteomics is inherently limited in its ability to identify proteins compared to bottom-up proteomics, despite its advantage in offering a deeper understanding of the existing proteoforms¹⁴. Even though we could detect the proteoforms based on their distinct accurate molecular masses, the fragments obtained from online MS/MS experiments are typically limited especially for large proteins. This challenge makes it difficult to unambiguously identify all of the proteins (proteoforms) detected using the currently available database search algorithm.¹⁵⁷ In particular, a single dimension of separation is not sufficient to achieve a large number of proteoform identification but does serve to demonstrate the potential of this surfactant to improve protein solubility and overall throughput.

Admittedly, a well-recognized challenge in top-down proteomics is the detection of large proteins due to the exponential decay in S/N in a mass spectrometer as a function of increasing molecular mass.⁷⁵ Top-down mass spectrometry is biased against larger protein species especially when a single dimension of separation is used; thus, the low molecular weight (MW) proteins may show up more readily even if they are in lower abundance.

Furthermore, the relatively lower number of identification could be due to the co-elution problems since only 1DLC is used in this study with limited separation power, which means that lower abundance and higher MW proteins could be insufficiently fragmented in online LC-MS/MS if co-eluting with higher abundance and lower MW species. We expect the incorporation of multi-dimensional (MD)LC separation of intact proteins reduces the co-elution of proteins, which enables the detection and identification of low abundance and/or high MW proteins towards a

deeper proteome coverage^{157, 158}. Nevertheless, the MDLC strategies need a significant amount of samples and a dramatically longer experiment time than 1DLC¹⁴. Moreover, RPLC could impose some concerns for analyzing hydrophobic proteins (e.g. too much retention on the column), although early work on intact integral membrane proteomics, namely the characterization of GPCR (7TMD) bacteriorhodopsin, was performed using a polymeric reversed-phase material, PLRP-S, with column heating as described in a previous study¹⁵⁹. In this study, we were able to observe the elution of very hydrophobic species such as ATP synthase subunit c also using the PLRP-S material with column heating. Notably, ATP synthase subunit c has a gravity score (an amino acid-based hydrophobicity scale that is particularly relevant for denatured proteins) of 1.14 when compared to Bacteriorhodopsin, 0.78, which illustrates that we can evaluate highly hydrophobic species using such RPLC method. However other separation modes may improve the analysis of hydrophobic species and it is important to actively investigate other chromatographic modes for membrane protein separation.

Lastly, while potential improvements could be made should complementary dissociation techniques be applied, we acknowledge that other dissociation techniques such as ultra-violet photodissociation (UVPD) have limited improvement over collision-induced dissociation (CID)/higher-energy collisional dissociation (HCD) in terms of the number of protein identifications via online LC-MS/MS and data-dependent acquisition mode¹⁰⁹. For membrane proteins, CID/HCD fragments transmembrane domains particularly well and we observed highly efficient CID fragmentation which preferentially cleaved in the transmembrane in agreement with previous studies^{159, 160}. On the other hand, from a protein characterization standpoint, the use of complementary dissociation techniques, such as UVPD, electron transfer dissociation (ETD)/electron capture dissociation (ECD) together with CID/HCD significantly improved

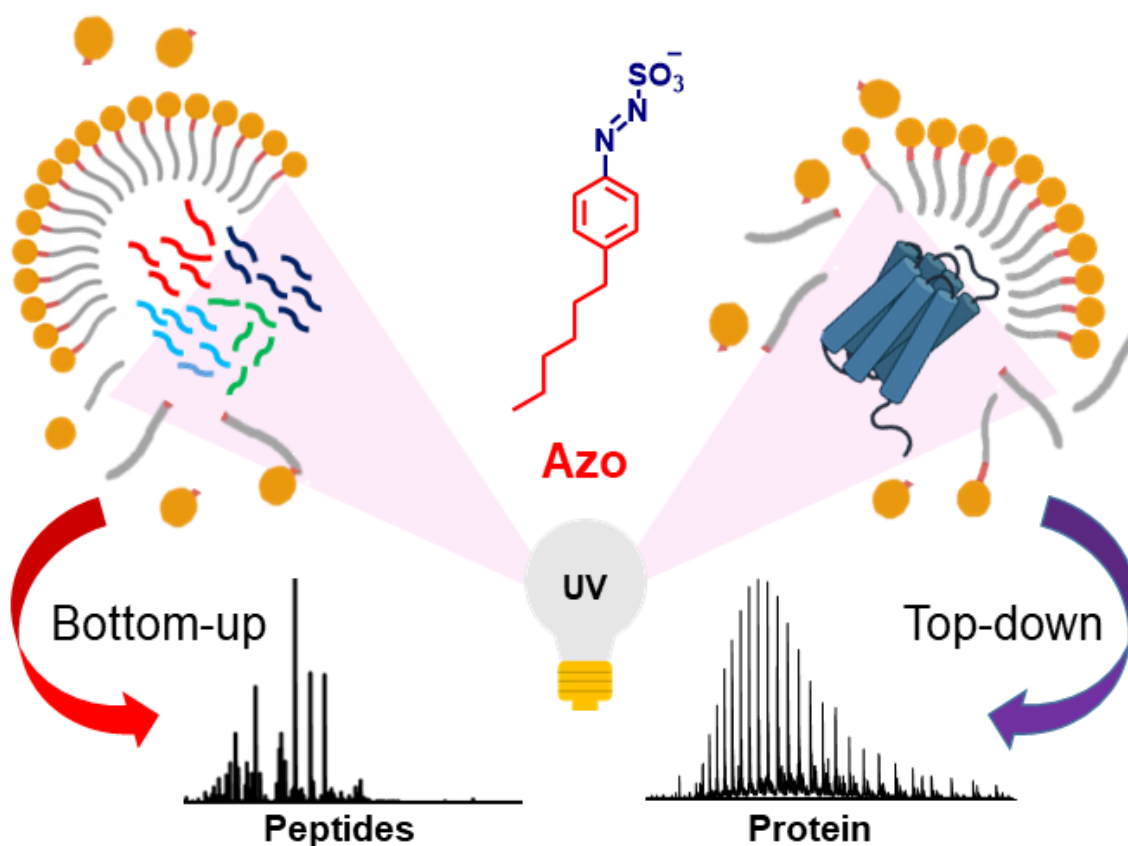
sequence coverage and could greatly aid proteoform characterization (e.g. labile PTM localization)^{14, 16, 109}. Nevertheless, this generally requires a more targeted approach to accumulate sufficient spectra, which is better suited to implement a targeted proteomics approach addressing a specific biological question.

Therefore, we believe Azo has the potential to improve global top-down proteomics studies when coupled to multidimensional separation method^{61, 158, 161}, combined with complementary fragmentation techniques such as UVPD and ETD/ECD for comprehensive protein characterization^{16, 109, 162}, and with further improvements in data acquisition and identification strategies^{130, 163}. We expect that Azo will facilitate a myriad of proteomic studies for understanding disease mechanisms and clinical diagnosis. Given surfactants' instrumental roles in biochemical research, we envision this photo-cleavable surfactant will have a broader impact beyond proteomics. Notably, Azo can be easily synthesized and purified, it can be used as a cleavable SDS-replacement in general biochemical applications, for example, in SDS-PAGE.

2.4. Acknowledgement

This research is supported by NIH R01 GM117058 (to S.J. and Y.G.). Y. G. would like to acknowledge R01 HL109810, R01 HL096971, R01 GM125085, and S10 OD018475 (to Y.G.). We thank A. Chen, E. Chang, and W. Tang for their assistance in the early stage of the project, S. Mitchell and T. Tucholski for the help with graphics, and T. Hacker for providing the swine hearts. We would like to thank Matt Willetts at Bruker for his assistance with DataAnalysis software. We would also like to acknowledge A. Carr, E. Bayne, and J. Melby for their help in testing the Supplementary Protocol to ensure reproducible results.

Chapter 3. High-throughput Proteomics Enabled by a Photocleavable Surfactant



This chapter has been published and is adapted from:

Brown, K. A.; Tucholski, T.; Eken, C.; Knott, S.; Zhu, Y.; Jin, S.; Ge, Y., High-Throughput Proteomics Enabled by a Photocleavable Surfactant. *Angewandte Chemie* **2020**, 132 (22), 8484-8488.

3.1 Abstract

Mass spectrometry (MS)-based proteomics provides unprecedented opportunities for understanding the structure and function of proteins in complex biological systems; however, protein solubility and sample preparation before MS analysis remains a bottleneck preventing high-throughput proteomics. Herein, we report for the first time a high-throughput bottom-up proteomics method enabled by a newly developed MS-compatible photocleavable surfactant, 4-hexylphenylazosulfonate (Azo)³⁰ that facilitates robust protein extraction, rapid enzymatic digestion (30 min compared to overnight), and subsequent MS-analysis following UV degradation. Moreover, we developed an Azo-aided bottom-up method for efficient analysis of integral membrane proteins, which are key drug targets and are generally underrepresented in global proteomic studies. Furthermore, we demonstrated the unique ability of Azo to serve as an “all-in-one” MS-compatible surfactant for both top-down and bottom-up proteomics, with streamlined workflows for high-throughput proteomics amenable to clinical applications.

3.2. Introduction

Mass spectrometry (MS)-based proteomics allows large-scale, quantitative measurements of proteins, and provides unprecedented insights into the structure and function of proteins in complex biological systems.^{7, 8, 164, 165} Currently, there are two major MS-based proteomic approaches: the bottom-up approach, where proteins are enzymatically digested and the resulting peptides measured by MS to infer the identity, quantity, and modification status of the proteins; and the top-down approach, where intact proteins are measured and their sequences determined for complete characterization of their proteoforms.^{3, 166-169} Bottom-up proteomics remains the most

widespread proteomic method due to the well-established experimental and computational tools⁹ developed over the last two decades, but it has limitations for comprehensive analysis of proteoforms.^{167, 169} On the other hand, significant strides have recently been made to advance top-down proteomics, which provides comprehensive analysis of proteoforms, owing to the recent successful developments of high-resolution mass spectrometers, protein separations, and user-friendly software for data analysis.^{3, 14, 167, 168} However, further method developments are urgently needed to increase the throughput in both bottom-up and top-down approaches, especially for time-sensitive clinical applications.

Sample preparation remains a bottleneck for obtaining high-quality MS data in a high-throughput fashion for both top-down and bottom-up proteomics.^{14, 148, 170, 171} Typically, surfactants (also known as detergents) are added to effectively solubilize proteins, especially membrane proteins, in cells and tissues.^{9, 25, 148, 155, 164} However, conventional surfactants that can solubilize cells and tissue well, such as sodium dodecyl sulfate (SDS), are incompatible with MS and need be removed before MS analysis, which is not only time consuming but may result in protein loss and degradation.^{25, 172, 173} An MS-compatible surfactant that can solubilize and denature proteins with comparable efficacy to SDS without the need for time-consuming sample clean-up is attractive for streamlining sample preparation for proteomics.¹⁷¹ To address this problem, MS-compatible surfactants that can be degraded into innocuous non-surfactant by-products have been developed.^{27, 156, 172, 174} Particularly, acid-labile ionic surfactants such as RapiGestTM^{27, 175, 176}, ProteaseMaxTM¹⁵⁶, and MaSDeS¹⁷⁴ can significantly improve the protein solubility and facilitate protein digestion. However, these acid-labile surfactants are degraded in highly acidic conditions post-digestion and often require an offline clean-up step, precluding automation.¹⁷² Moreover, strong acidic degradation may cause the loss of important post-

translational modifications that are unstable in acidic conditions (i.e. acid-labile phosphorylations¹⁷⁷ and glycosylations¹⁷⁸). Furthermore, some of these acid-labile surfactants generate amine products following degradation that interfere with the isobaric amine-reactive tandem mass tag labeling reagents, such as tandem mass tag (TMT) and isobaric tags for relative and absolute quantitation (iTRAQ).¹⁷² Most importantly, none of these acid-labile surfactants are compatible with top-down proteomics for the direct analysis of intact proteins, because they greatly suppressed intact protein MS signal as reported previously.³⁰ -Therefore, we aim to develop a high-throughput sample preparation method that is amenable to both top-down and bottom-up proteomics and overcome the limitations of the conventional acid cleavable surfactants.

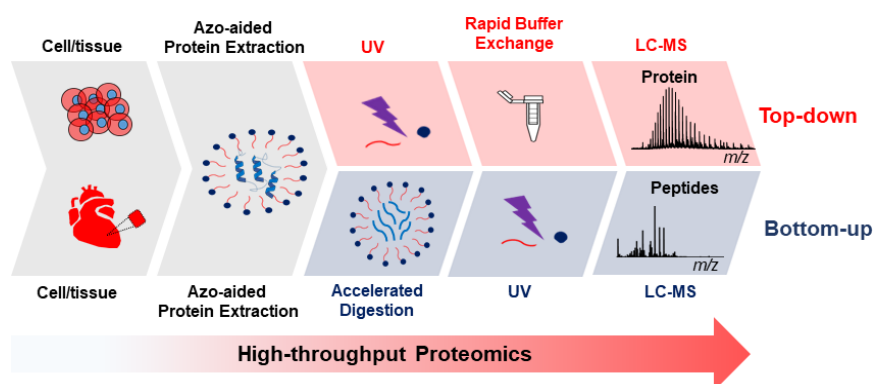


Figure 3.1. Scheme for azo-enable high-throughput top-down and bottom-up proteomics.

3.3. Azo-enabled High-throughput Sample Preparation

3.3.1. Rapid enzymatic digestion

Recently, we developed a photo-cleavable surfactant, 4-hexylphenylazosulfonate (Azo), for top-down proteomics.³⁰ Azo is straightforward to synthesize (requiring only two-step synthesis and simple purification),¹⁵⁴ effectively solubilizes proteins (including membrane proteins), and can be rapidly degraded before MS analysis.³⁰ Here, for the first time, we demonstrate that Azo

is fully compatible with bottom-up proteomics and uniquely serves as an “all-in-one” MS-compatible surfactant for both bottom-up and top-down proteomics, which greatly facilitates high-throughput sample handling before MS-analysis (Scheme 1). A direct comparison of the Azo and a leading acid-labile surfactant, RapiGest™ (RG, also known as ALS),^{27, 175, 176} clearly demonstrated Azo is compatible with direct-infusion electrospray MS analysis of intact proteins and peptides whereas RG greatly suppressed MS signal (Figure S1). Moreover, we demonstrate the ability of Azo to facilitate protein extraction, enable rapid enzymatic digestion, and rapidly degrade by UV before MS analysis for a streamlined bottom-up proteomics approach (Figure 1a).

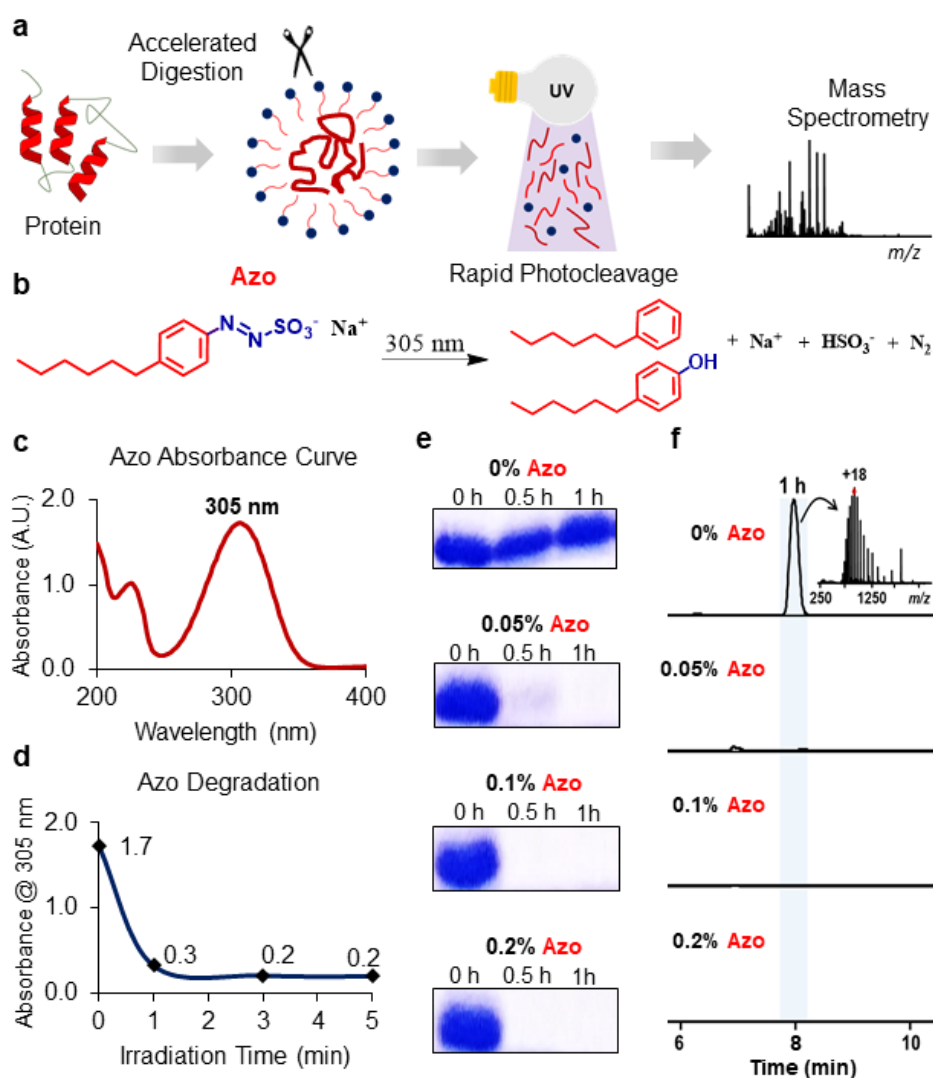


Figure 3.2. Enhanced enzymatic digestion and MS analysis using a photocleavable surfactant, Azo. (a) Scheme for Azo-aided bottom-up proteomics. (b) Degradation of Azo by UV irradiation. (c) UV-Vis spectrum for Azo highlighting maximal absorbance at 305 nm. (d) UV-Vis spectrum monitoring the rapid degradation of Azo (0.1% in water) as a function of irradiation time with a 100 W mercury lamp. (e) Digestion of myoglobin [0-1 h, Azo (0-0.2%)] was monitored by SDS-PAGE, stained with Coomassie Blue, and (f) LC-MS analysis using a Q-TOF mass spectrometer. Relative abundances of the extracted ion chromatograms were normalized to 5.5 E8

First, we demonstrated the rapid degradation of Azo by UV-Vis spectroscopy. The intact surfactant has a maximal absorbance at 305 nm^{30, 151} (Figure 1b-c); thus, degradation of Azo can be achieved using a mercury lamp (305 nm) in less than 5 min^{30, 179} (Figure 1d and S2). This facile surfactant degradation method makes Azo ideal for rapid sample processing.

Next, we evaluated Azo-aided in-solution digestion. Myoglobin, a trypsin resistant protein,²⁷ was digested in-solution without (0%, control) or in the presence of Azo (0.05%, 0.1%, or 0.2%). The efficiency of digestion was assessed by visualization of the remaining intact myoglobin through SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis (Figure 1e). After 0.5 h of trypsin digestion in the absence of surfactant, the intact protein band at ~17 kDa was clearly observed, suggesting poor digestion efficiency. In contrast, only a faint intact protein band was observed for sample digested in the presence of 0.05% Azo, and no intact protein band was observed in the presence of 0.1% and 0.2% Azo after 0.5 h of digestion indicating that Azo greatly improved digestion efficiency. Subsequently, we evaluated a longer digestion time (1 h) in the presence or absence of Azo. Even after 1 h digestion, a predominant intact protein was still observed without Azo, whereas only a minute amount of intact protein was visible in the presence of 0.05% Azo and no intact protein was detected for 0.1% or 0.2% Azo (Figure 1ef). This is conceivable, since without a denaturing agent (e.g. an anionic surfactant) the enzyme has limited access to the protein backbone, precluding digestion. These results indicate Azo facilitates the

denaturation of the protein, providing efficient digestion. Furthermore, we analyzed the Azo-aided myoglobin digest (0.1% Azo, 1 h) via direct electrospray ionization for peptide mass fingerprinting analysis without an additional clean-up procedure and observed 89% sequence coverage (Figure S3, bottom panel). On the other hand, analysis of the myoglobin sample digested in the absence of surfactant (0% Azo, 1 h) yielded minimal detectable peptides, and the MS signal was largely dominated by the intact protein (Figure S3, top panel). Next, we performed a direct comparison between Azo and RG for trypsin digestion of myoglobin and direct-infusion MS analysis, observing lower peptide signals for the sample with RG than that with Azo (Figure S4). Finally, to further confirm the rapid rate of enzymatic digestion in the presence of Azo, we digested a standard mixture of insulin, myoglobin, and carbonic anhydrase with (0.1% or 0.2%) and without Azo. We observed a similar improvement in digestion with Azo (Figure S5). Taken together, these results indicate proteins (even those resistant to trypsin digestion) can be rapidly and efficiently digested using Azo.

After demonstrating rapid enzymatic digestion using standard proteins, we further investigated Azo-aided digestion of complex cell lysates. Single-step Azo protein extraction was performed using human embryonic kidney 293T (HEK293T) cells to generate whole cell lysate. SDS-PAGE analysis-demonstrated reproducible extractions for three HEK293T samples (Figure 2a). The proteins were digested for 1 h or overnight to determine whether digestion time, the most time-consuming step in the bottom-up proteomics workflow, could be shortened significantly using Azo while maintaining reproducibility. After surfactant degradation by UV, a brief centrifugation step, and liquid chromatography (LC)-MS/MS analysis, the data were processed using MaxQuant¹⁸⁰ to determine the number of protein groups identified in each sample and to perform label-free quantitation (Table S1). Overall, 910 and 938 proteins were identified in all

three extraction replicates using overnight digestion (Figure 2b) and 1 h digestion (Figure 2c), respectively. Combined, we identified 1557 proteins using overnight digestion and 1572 proteins using a 1 h digestion with a 60 min LC gradient before MS analysis. This corresponds to an 81% overlap between the two conditions (Figure 2d), confirming 1 h is sufficient for effective protein digestion in the presence of Azo.

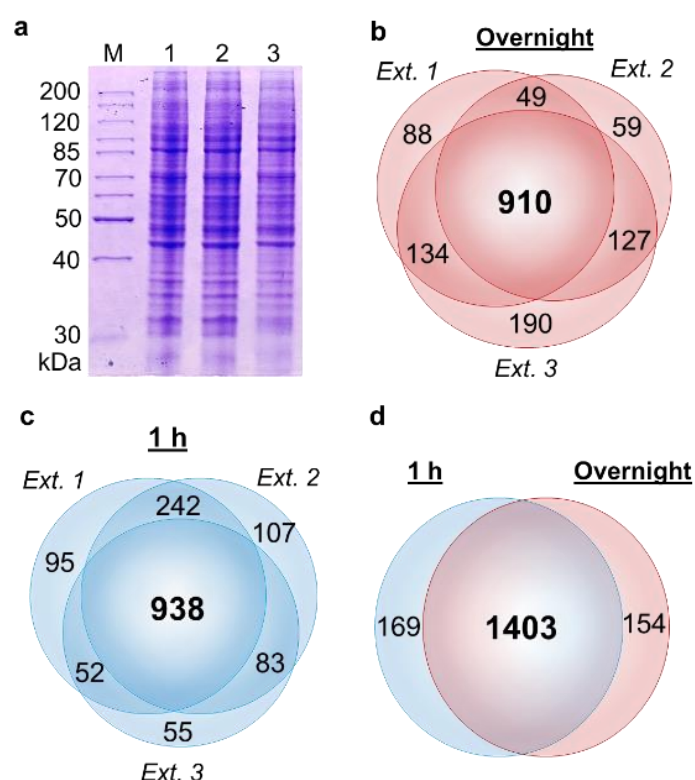


Figure 3.3. Rapid and reproducible extraction, enzymatic digestion, and LC-MS/MS analysis of proteins from human embryonic kidney 293T (HEK293T) cells (a) SDS-PAGE analysis demonstrates consistent extraction profiles across three extractions (marker (M), 1, 2, 3). Aliquots were taken from each extraction and digested overnight or for 1 h. After LC-MS/MS and MaxQuant identification, a high degree of overlap was observed for both the (b) overnight and (c) 1 h samples. Similarly (d) 1403 of the 1726 combined identified protein groups were observed in both digestion methods. Data collected using Q-TOF.

3.3.2. Benchmark comparison to FASP

Next, we compared the performance of Azo-aided digestion (30 min and overnight) with a gold-standard protocol for bottom-up proteomics using SDS, Filter Assisted Sample Preparation (FASP) developed by Mann and co-workers (overnight digestion).¹⁴⁸ Using three digestion replicates (proteins were extracted with Azo or SDS respectively) and a 60 min LC-MS analysis, we observed the greatest number of identifications using the Azo digestion (overnight) followed by the FASP (overnight) and Azo digestion (30 min) method (Figure S6 and Table S2). Significantly, we observed only 33 fewer protein identifications using a 30 min digestion protocol with Azo when compared to FASP, demonstrating a very robust, reproducible method that greatly reduces the sample preparation time by eliminating the multiple centrifugation steps and offline desalting (SI Discussion).

3.4. Azo-enabled Membrane Proteomics

Next, we demonstrate Azo's ability to aid in the analysis of integral membrane proteins, which are hydrophobic and generally expressed at lower levels, thus are underrepresented in broad proteomics studies.^{172, 181} Here, we focused on improving the bottom-up workflow for membrane proteomics by implementing a cloud-point extraction, using Triton X-114, to enrich membrane proteins from HEK293T cells with a single extraction (SI Discussion).^{182, 183} However, Triton surfactants are highly incompatible with MS analysis¹⁸⁴ and require removal before MS analysis. Subsequently, we used Azo to solubilize the precipitated protein pellets enriched with integral membranes for a more effective analysis of this important class of proteins. We observed more than a 6-fold increase in integral membrane protein identified using this method when compared to whole-cell extraction experiment (conceivably because highly abundant proteins that traditionally suppress the detection of integral membrane proteins were depleted) (Figure 3a and

Table S3). Similarly, a 4-fold increase in protein groups identified was achieved using 30 min digestions times for the enriched sample compared to the whole cell lysate (Figure 3a and Table S3). Importantly, we detected several proteins with over 10 transmembrane domains, demonstrating the ability of Azo to solubilize highly hydrophobic proteins and enable their MS analysis (Figure 3b). The results indicate longer digestion improves the identification of integral membrane proteins even with Azo, which can be rationalized by the lower abundance of lysine and arginine residue particularly in the transmembrane regions.¹⁸¹ Finally, we performed a global PTM discovery search using MetaMorephus¹⁸⁵ to localize modification on membrane proteins (Figure S7 and Table S3).

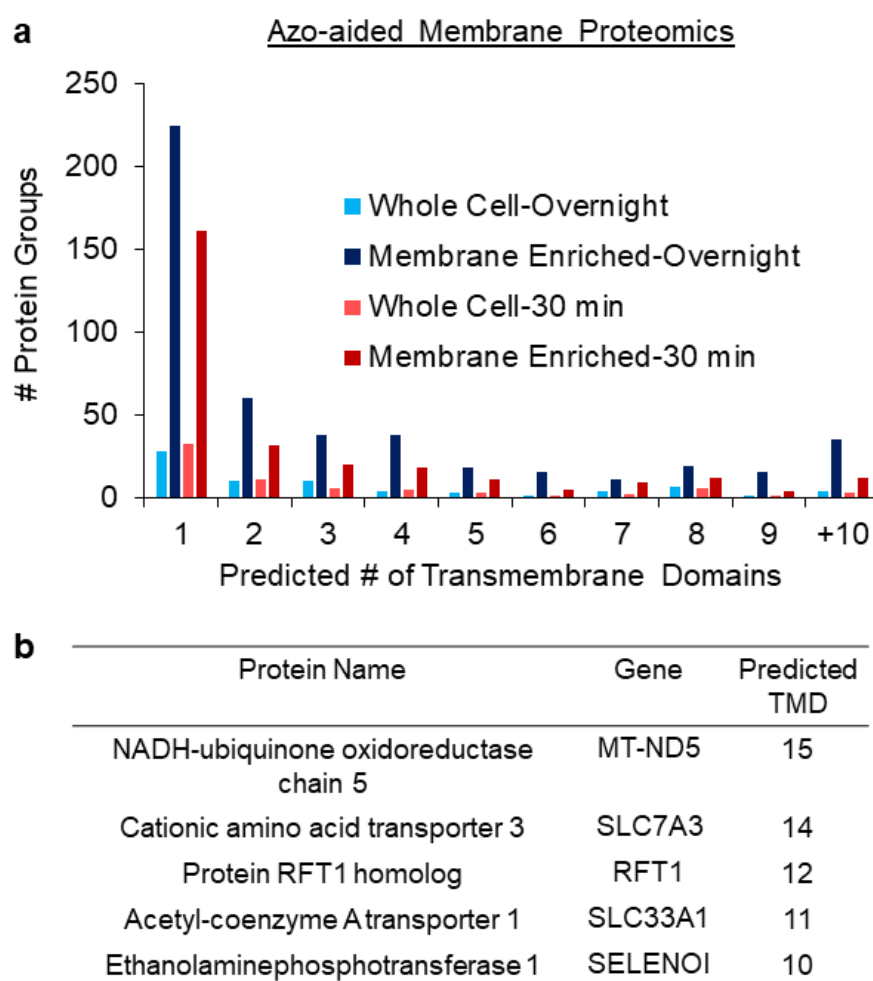


Figure 3.4. Analysis of integral membrane protein enabled by Azo. Proteins were extracted and membrane proteins enriched using a Triton x-114 cloud point extraction procedure,²⁸ then precipitated and solubilized again in Azo for in-solution digestion. Additionally, Azo was used for whole-cell protein extraction. Using an (a) overnight digestion a significantly higher number of integral membrane proteins were identified using the phase separation enrichment followed by Azo-aided digestion. Similarly, a 30 min digestion was performed demonstrating a high-throughput membrane protein analysis method. (b) A table of representative membrane proteins with a high number of transmembrane domains that were solubilized using Azo. Data represent a single LC-MS experiment.

3.5. Azo for Integrated Top-down and Bottom-up Proteomics

Here, we sought to demonstrate the unique application of Azo as an “all-in-one” MS-compatible surfactant for both top-down and bottom-up proteomics. Proteins were extracted, using Azo, from left ventricle heart tissue (after aqueous buffer extractions) and an aliquot was taken for bottom-up analysis. We observed that many of the protein identifications were involved in cellular and metabolic processes or localization (Figure S8a and Table S4). Moreover, we demonstrated the potential of this method to investigate disease-relevant proteins (e.g. sarcomeric proteins) and their interacting partners¹⁸⁶ which is critical for gaining a better understanding of their role in disease mechanisms (Figure S8b). Furthermore, an aliquot was taken for top-down proteomics which provides additional information about existing proteoforms (Figure 4 and Figure S9), thus improving the biological information that can be obtained from a sample. Overall LC-MS analysis of cardiac tissue illustrated the complexity of the Azo-extracted samples with many proteins and their respective proteoforms eluting simultaneously (Figure S10), demonstrating the promise of Azo for future applications in global top-down proteomics studies. Importantly, Azo also enables the extraction and analysis of intact integral membrane proteins (Figure S11).

In summary, we developed a novel approach using an anionic photocleavable surfactant, Azo, to enable high-throughput bottom-up proteomics. Azo can effectively extract and solubilize

proteins reproducibly, enables rapid enzymatic digestion, and is amenable to MS analysis without an additional desalting step which improves the throughput, permitting bottom-up analyses for a wide range of applications. Moreover, we have developed an Azo-aided bottom-up proteomics workflow for effective membrane protein solubilization. Furthermore, we have demonstrated the unique capability of Azo as an “all-in-one” MS-compatible surfactant for both top-down and bottom-up proteomics with great potential to provide a streamlined strategy for high throughput proteomics with far-reaching applications to fully realize the impact of proteomics in clinical diagnosis.

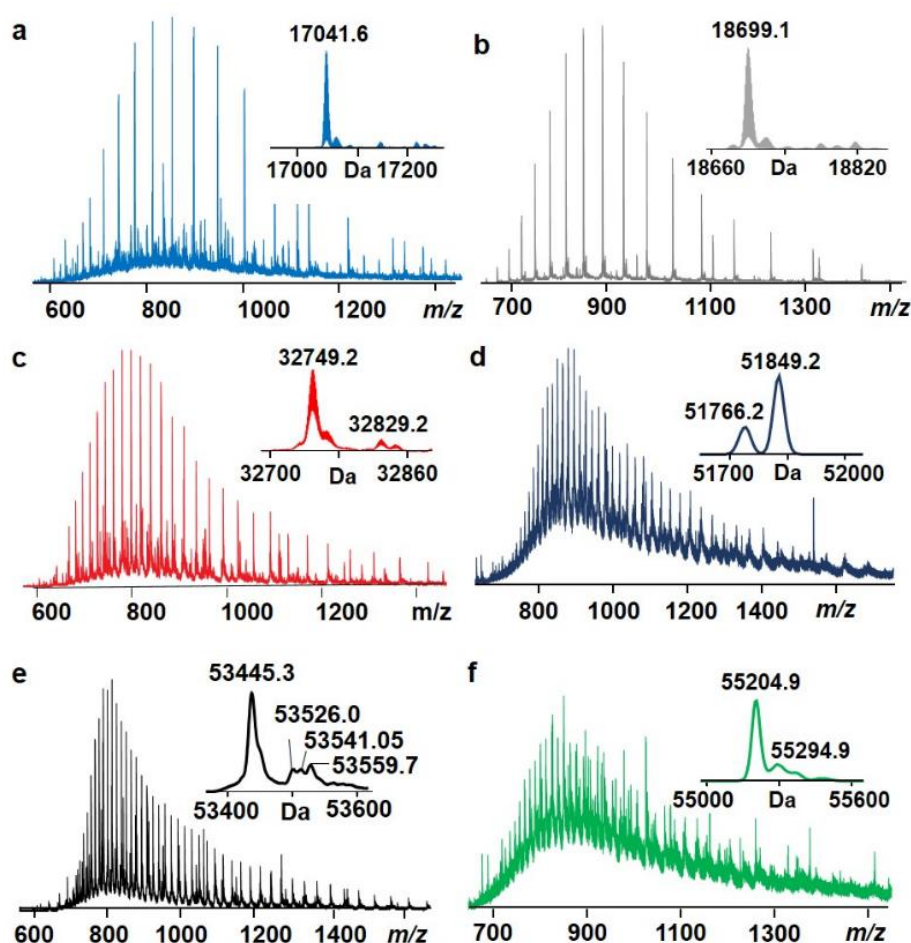


Figure 3.5. Azo-enabled top-down proteomics analysis of cardiac tissue. (a-f) Intact mass spectra of proteoforms analyzed by LC-MS. Data collected using Maxis II Q-TOF.

3.6. Experimental Procedures and Supplemental information

3.6.1. Chemicals and Materials

Chemicals and reagents were purchased from MilliporeSigma Inc. (St Louis, MO, USA) without further purification unless otherwise noted. Acrylamide gels were made in-house using acrylamide/bis-acrylamide (37:5:1) solution from Hoefer (Holliston, MA, USA). Halt protease/phosphatase inhibitor cocktail, tris(20carboxyethyl)phosphine (TCEP), BCA reagent, Fetal Bovine Serum (FBS), Dulbecco's Modified Eagle Medium (DMEM), phosphate-buffered saline (PBS), and Penicillin/Streptomycin were obtained from Thermofisher Scientific (Waltham, MA, USA). C18, 5 μ m materials were purchased from Restek (Bellefonte, PA, USA). Water, isopropanol (IPA), acetonitrile (ACN) (HPLC grade), and phenylmethylsulfonyl fluoride (PMSF) were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Trypsin Gold was purchased from Promega (Madison, WI, USA). RapiGest was purchased from Waters (Milford, MA). Azo was synthesized in-house as described previously.^{30, 154}

3.6.2. UV-Visible Spectroscopy Analysis of Azo Degradation

50 μ L of 0.1% Azo in 25 mM ammonium bicarbonate (ABC) was irradiated for 0, 1, 3, and 5 min using 100 W mercury lamp [Nikon housing with HB-10101 AFT power supply, Nikon (Tokyo, Japan)]. The sample was diluted with 950 μ L of water and a spectrum was taken using Varian Cary 50 UV-Visible (600 nm- 200 nm, background correction, fast scan rate). The resulting plots were generated in Excel. See **Figure 1b,c**.

3.6.3. MS Analysis of Azo Degradation

48 μ L of 0.1% Azo in 25 mM ABC or 50% IPA:49.5% Water: 0.5% formic acid was irradiated for 5 min using 100 W mercury lamp [Nikon housing with HB-10101 AFT power supply, Nikon (Tokyo, Japan)]. 2 μ L of NaOH was added to the sample. The sample was directly infused into Impact II Q-TOF mass spectrometer (Bruker Daltonics) using a capillary voltage of 4500 V and an endplate voltage of 500 V in negative mode. **See Figure S1.**

3.6.4. Digestion and Analysis of Standard Proteins

Myoglobin (25 μ g) was dissolved in 25 mM ABC with 0, 0.05, 0.1, 0.2% Azo (final volume was 50 μ L). A 10 μ L aliquot was removed for a 0 h time point and boiled in SDS loading dye. Trypsin (1:50 enzyme: protein) was added and the sample was incubated at 37 °C. A 10 μ L aliquot was removed at 0.5 h and 1 h and boiled in SDS loading dye to quench the reaction. The samples were loaded onto a 10% acrylamide gel, separated with 125 V, and stained with Coomassie Brilliant Blue for visualization. **See Figure 1e.**

To the remaining 20 μ L of the sample was added 1 μ L of formic acid and 19 μ L ACN after 1 h of digestion. The sample was irradiated and 2 μ L of protein was separated on a PLRP column (200 x 0.25 mm, 10 μ m, 1000 Å) using 0.2% formic acid in water and ACN:IPA for mobile phase A (MPA) and mobile phase B (MPB) respectively with a 15 min gradient from 20%-80% B. The sample was eluted from the column and infused into a maXis II ETD Q-TOF (Bruker Daltonics) using a capillary voltage of 4500 V and an endplate voltage of 500 V. MS scans were collected at 1 Hz. **See Figure 1d.**

Additionally, myoglobin digest (1 h, 0.1% Azo) was directly infused into a 12T solariX™ Fourier Transform Ion Cyclotron mass spectrometer (Bruker Daltonics) mass spectrometer using a nano-ESI source (Triversa NanoMate; Advion Bioscience, Ithaca, NY). A voltage of 1.5 kV was applied with 0.3 psi drying gas for peptide mass fingerprinting. Scans were collected at 512,000 word with a mass range of 200-3000 m/z . Data was analyzed in Mascot allowing for 2 missed cleavages with oxidation (M) as a variable modification. See **Figure S3**. The trypsin digestion and MS analysis were performed using 0.1% Azo or 0.1% RapiGest (according to the manufacturer's protocol) for a direct comparison. Samples were directly infused into a maXis II ETD Q-TOF (Bruker Daltonics) using a capillary voltage of 4500 V and an endplate voltage of 500 V. MS scans were collected at 1 Hz. See **Figure S4**.

Finally, 10 µg insulin, 20 µg of myoglobin, and 60 µg of carbonic anhydrase were digested in 0, 0.1, and 0.2% Azo for 1 h. After digestion, the samples were irradiated with UV and analyzed by LC-MS using a maXis II ETD Q-TOF as described above for myoglobin. **Figure S2**.

3.6.5. Protein Extraction, Digestion, and LC-MS/MS Analysis of HEK293T Cells

Human embryonic kidney 293T (HEK293T) cells (ATCC, Manassas, VA) were grown on 10-cm plates with 10% FBS and 1x Penicillin-Streptomycin at 37 °C with 5% CO₂. Plates (~90% confluent) were washed with PBS, pelleted at 500 × g , flash-frozen, and stored at -80 °C.

All steps were performed at 4 °C. Cells from a single plate were lysed with 250 µL of buffer (0.5% Azo, 25 mM ABC, 5 mM (TCEP), 5 mM ethylenediaminetetraacetic acid (EDTA), and 1 x Halt protease/phosphatase inhibitor cocktail) and sonicated for 5 s (at 20% amplitude) to shear the

DNA. The extract was centrifuged for 10 min at $20,000 \times g$ and the supernatant removed. Protein concentration was determined using BCA protein assay reagent with albumin as a standard. 25 μg was analyzed by SDS-PAGE as described above.

200 μg of protein (final volume was 50 μL and Azo concentration $\sim 0.1\%$ for digestion) was reduced with 5 mM dithiothreitol (DTT), alkylated with 15 mM iodoacetamide (IAA), and digested with trypsin (1:50) overnight or for 1 h at 37 °C. The solution was irradiated (5 min to ensure complete surfactant degradation) and the reaction was quenched with formic acid (precipitation observed upon acidification). The sample was centrifuged for 5 min at $20,000 \times g$ and the supernatant was subsequently collected.

10 μL of the sample was loaded onto a home-packed C18 column (250 x 0.25 mm) and separated using a 60 min gradient from 5-40% B. Mobile phases were 0.2% formic acid in water (A) and ACN (B). The column was heated to 60 °C and the flow rate was set to 5 $\mu\text{L}/\text{min}$. Peptides eluted from the column were infused into an Impact II Q-TOF mass spectrometer (Bruker Daltonics) using 4500 V and an endplate voltage of 500 V. MS scans were collected at 2 Hz and the top 30 most intense ions were selected for collision-induced dissociation (CID) using a 3 m/z window, 8-32 Hz (intensity 2500-25000) scan rate, and voltage energy scaled by the m/z and charge (see below). See **Figure 2**.

Mass (m/z)	Collision (eV)	Charge State (+)
300	20	2
300	20	3
400	21	2
400	21	3
500	23	2
500	21	3
600	27	2
600	24	3
700	30	2
700	24	3
800	38	2
800	35	3
900	43	2
900	41	3
1000	48	2
1000	43	3
1100	50	2
1100	45	3
1300	52	2
1300	50	3

3.6.6. Comparison of FASP and Azo-aided in-solution digestion

HEK293T cells from a single plate were lysed with 250 μ L of buffer (0.5% Azo or SDS, 25 mM ABC, 5 mM (TCEP), 1 mM ethylenediaminetetraacetic acid (EDTA), and 1 x Halt protease/phosphatase inhibitor cocktail) and sonicated for 5 s (at 20% amplitude) to shear the DNA. The extract was centrifuged for 10 min at $16,000 \times g$ and the supernatant removed. Protein concentration was determined using BCA protein assay reagent with albumin as a standard. 25 μ g was analyzed by SDS-PAGE as described above.

Subsequently, 200 μ g of protein from the Azo extract was digested and processed as described above. In parallel, 200 μ g of protein from the SDS was digested using the FASP method.¹⁴⁸ Briefly, the sample was reduced with 5 mM DTT and alkylated with 15 mM iodoacetamide (IAA). Next,

4 washes with 8 M Urea in 25 mM ABC (all spins were 20 min at $14,000 \times g$) were performed followed by 2 washes with 25 mM ABC. The sample was digested with trypsin (1:50) overnight at 37 °C. The sample was collected, desalted with a C18 Ziptip, dried in a speed-vac, and suspended in a final volume of 50 μ L of 0.2% formic acid in water (to match the Azo sample). LC-MS analysis was performed on 10 μ L of the sample using the parameters above with maXis II ETD Q-TOF (Bruker Daltonics). See **Figure S4**.

3.6.7. Membrane Protein Enrichment, Digestion, and LC-MS Analysis

HEK293T cells from a single 10-cm plate were lysed with 1 mL of buffer (2% Triton X-114, 25 mM ABC, 25 mM methionine, and 1 mM phenylmethylsulfonyl fluoride). After 1 h incubation at 4 °C, cells were centrifuged for 5 min at $20,000 \times g$ at 4 °C. The supernatant was removed and heated to 37 °C to induce the cloud point as described previously.¹⁸² The mixture was centrifuged at $3,000 \times g$ at room temperature (RT) and the top layer was removed. The bottom triton layer was washed with cold buffer (25 mM ABC and 25 mM methionine) and the cloud point procedure was repeated. The triton layer was precipitated with 4 v of acetone overnight. 0.1% Azo was added to solubilize the dried protein pellet and the sample was briefly sonicated in a water bath. The protein concentration was determined by BCA assay using albumin as a standard. 100 μ g of protein was reduced, alkylated, digested overnight, and analyzed as described above (a 90 min gradient from 5-45% was used for LC separation). See **Figure 3**.

3.6.8. Bottom-up and Top-down Analysis of Cardiac Tissue

Donor's hearts were received after surgical human heart transplant specimens in the University of Wisconsin Hospital and Clinics. All procedures were approved by the University of Wisconsin-Madison Institutional Review Board. Donor heart tissue was maintained in cardioplegic solution before dissection and the tissue was immediately snap-frozen in liquid nitrogen and stored at -80°C ^{157, 187}. All steps were performed at 4°C . About 170 mg of left ventricular tissue was cut into small pieces and homogenized in 1.5 mL of buffer (25 mM HEPES, 0.6 M KCl, 0.5 M NaF, 0.25 M sucrose, 2.5 mM methionine, 2 mM TCEP, 1 mM PMSF, 1 x protease inhibitor cocktail) with a Polytron electric homogenizer (model PRO200, Pro Scientific). The sample was centrifuged at $20,000 \times g$ and the supernatant discarded. This was repeated using in 0.5 mL of buffer (25 mM ABC, 2.5 mM methionine, 2 mM TCEP, 1 mM PMSF, 1 x protease inhibitor cocktail). Finally, the pellet was homogenized in 0.5 mL of buffer (1% Azo, 2.5 mM methionine, 2 mM TCEP). Protein concentration was determined by BCA assay using albumin as a standard.

A 200 μg aliquot was taken and diluted with 25 mM ABC (volume add such that the final volume would be 50 μL before LC-MS) and reduced with 5 mM DTT for 30 min at 37°C . The sample was alkylated with 15 mM IAA in the dark for 30 min. Trypsin was added (1:50) and the sample was incubated at 37°C for 30 min and 1 h. The sample was irradiated, quench with 1 μL formic acid, and centrifuged for 5 min at $20,000 \times g$ to remove insoluble material. The sample was analyzed as describe above for HEK293T cells; however, a maXis II ETD Q-TOF was used. **See Figure S7.**

A 25 μL aliquot was taken and dilute with 100 μL of methionine (250 mM), 375 μL dilution buffer (50% IPA, 49.4% water, 0.5 % formic acid, 0.1% hexfluoroisopropanol), and 1 μL TCEP (1 M). After the addition, the sample was briefly sonicated in a water bath to ensure protein

solubility was maintained. The sample was irradiated and added to a 30 kDa molecular weight cutoff (MWCO) filter. The sample was concentrated and washed 2x with 200 μ L of 89.5% water, 10% IPA, 10% ACN, 0.5% formic acid. For LC-MS analysis, 10 μ L of the sample was injected onto a PLRP column (200 x 0.5 mm, 5 μ m, 1000 Å) heated to 60 °C. Mobile phases consisted of 0.2% formic acid in water (MPA) and ACN:IPA 1:1 (MPB). A 60 min gradient from 20%-60% MPB followed by 2 min gradient from 60-97% was used at a flow rate of 15 μ L/min. The sample was eluted from the column and infused into maXis II ETD Q-TOF (Bruker Daltonics) using 4500 V and an endplate voltage of 500 V. MS scans were collected at 1 Hz. The top 3 ions were selected for MS/MS using a voltage gradient shown below. See **Figure 4, S8-10**.

m/z	Width (m/z)	Collision Energy (ev)	Charge state
500	5	35	1
500	5	20	8
500	5	18	15
1000	5	24	1
1000	5	22	8
1000	5	20	15
1300	8	32	1
1300	8	30	8
1300	8	28	15
2000	8	45	1
2000	8	43	8
2000	8	40	15

3.6.9. Data Analysis

For bottom-up LC-MS/MS experiments, data were analyzed using MaxQuant (Version 1.5.7.4) and MetaMorpheus¹⁸⁵ searching against the *homo sapiens* reviewed Uniprot database (3/12/2019). A 1% false discovery rate (FDR) was set for protein and peptide identifications. Carbamidomethylation of cysteine was set for fixed modification while N-terminal acetylation and methionine oxidation were set as variable modifications. Two missed cleavages were allowed. All the other settings were based on the default parameters for a Q-TOF instrument. See **Table S1-4**. The number of transmembrane domains was determined using the TMHMM server (2.0).¹⁸⁸

For top-down LC-MS experiments, data were analyzed using Bruker Compass Data Analysis (4.3). Deconvolution for the large proteins (see **Figure 4**) was performed using Maximum Entropy with a resolution of 1000¹⁵⁷ and the mass was determined using sum peak. Charge states determination (see **Figure S8**) was performed using the Sophisticated Numerical Annotation Procedure (SNAP) peak-picking algorithm. Protein identification was performed using TopPIC (1.1.2) (see **Figure S9**). Further, MS/MS characterization using both eTHRASH and TopFD deconvolution algorithms and peak picking was performed using in-house developed MASH software (1.0.0.16705).¹⁸⁹

3.7. Supplemental Information

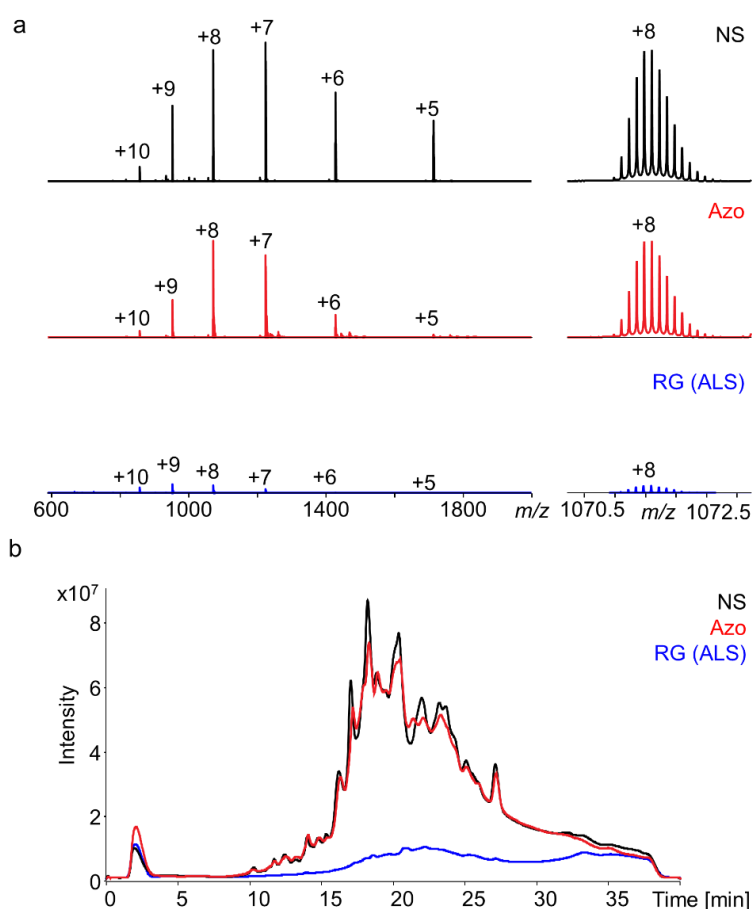


Figure 3.6. A Direct Comparison of the MS-compatibility of Azo and RapiGest™ (RG, also known as ALS). (a) 0.1% 4-hexylphenylazosulfonate (Azo) and RG (ALS) were added to ubiquitin and analyzed by high-resolution FTICR mass spectrometry and compared to a control with no surfactant (NS). The mass spectra were collected on a Bruker 12 T FTICR mass spectrometer and normalized to the maximum intensity in the NS (control) spectrum of 8E10 as described previously³⁰. (b) Total ion chromatogram comparing the MS signal of cardiac tissue lysate in the presence of different surfactants compared to no surfactant (NS) which serves as a control. Cardiac proteins extracted using NH_4HCO_3 buffer; Azo and ALS (0.1%) were added to the tissue lysate. The acid- and photocleavable surfactants were degraded under acid or UV, respectively. All samples were buffer exchanged and analyzed by LC-MS. Mass spectra were collected on a Bruker maXis II Q-TOF mass spectrometer.

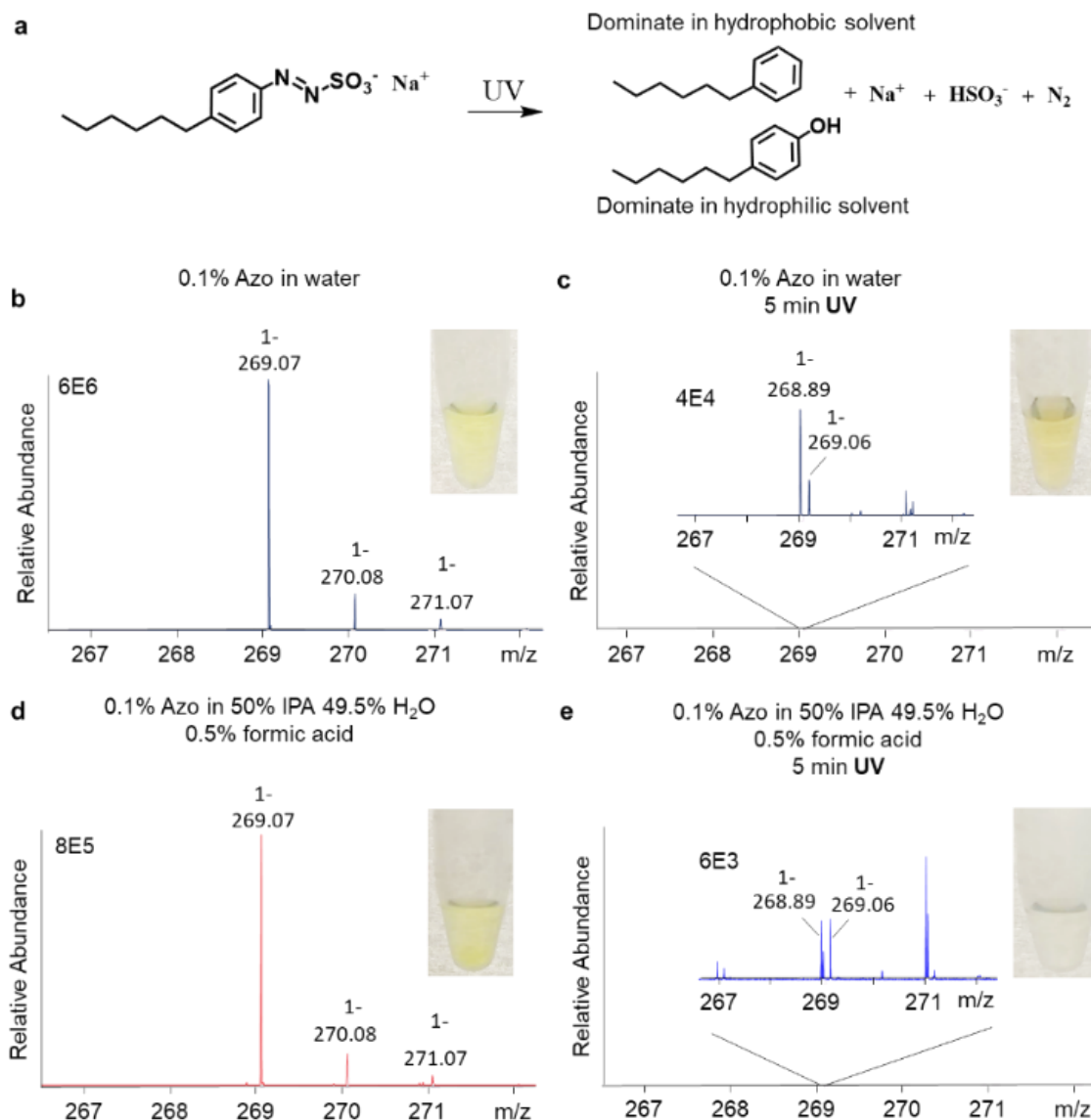


Figure 3.7. Azo degradation evaluated by MS. (a) Degradation of Azo upon UV irradiation. (b) Mass spectrum of 0.1% Azo in water. (c) Mass spectrum of 0.1% Azo in water after irradiation with a 100 W mercury lamp for 5 min. (d) MS of 0.1% Azo in 50 % isopropanol, 49.5% water, and 0.5% formic acid (e) MS of 0.1% Azo in 50 % isopropanol, 49.5% water, and 0.5% formic acid after 5 min of irradiation. The 4-hexylbenzene product is favored in the hydrophobic environment whereas the phenol is favored in the hydrophilic environment as described previously.^{179, 190}

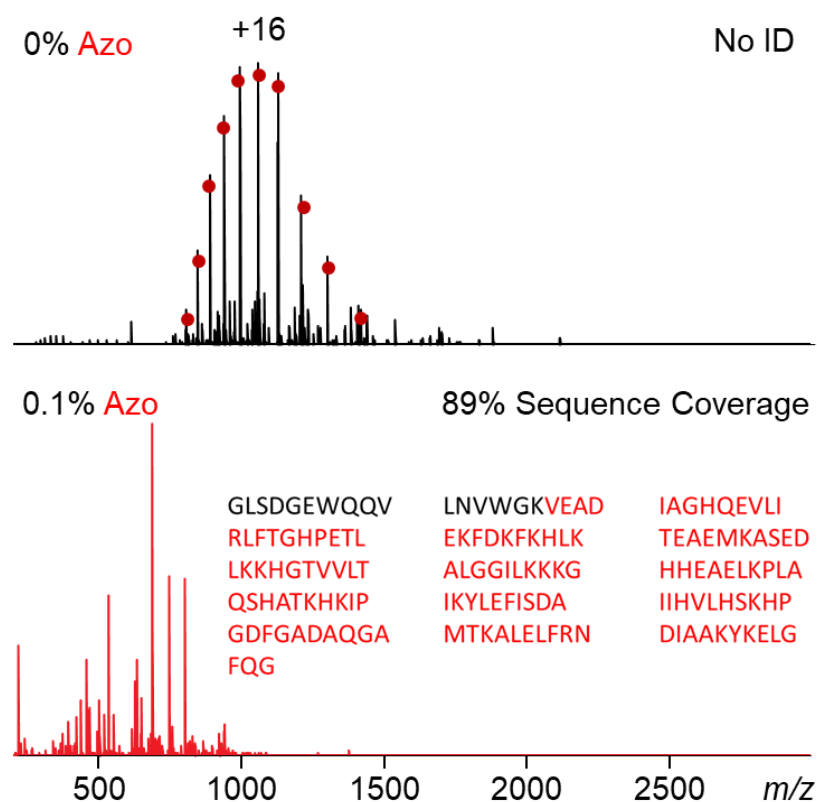


Figure 3.8. Azo-aided digestion for high-throughput targeted peptide mass fingerprinting.

Myoglobin was digested with trypsin in 0% (top) or 0.1% (bottom) Azo for 30 min. The sample was directly infused into a solariX FTICR (Bruker). Large intact myoglobin signal (indicated by red circles) can be observed with trypsin alone while complete digestion is observed with the aid of the surfactant. Overall no identification was determined for the digestion without surfactant whereas 89% sequence coverage was achieved using peptide mass fingerprinting when digestion was performed with the surfactant.

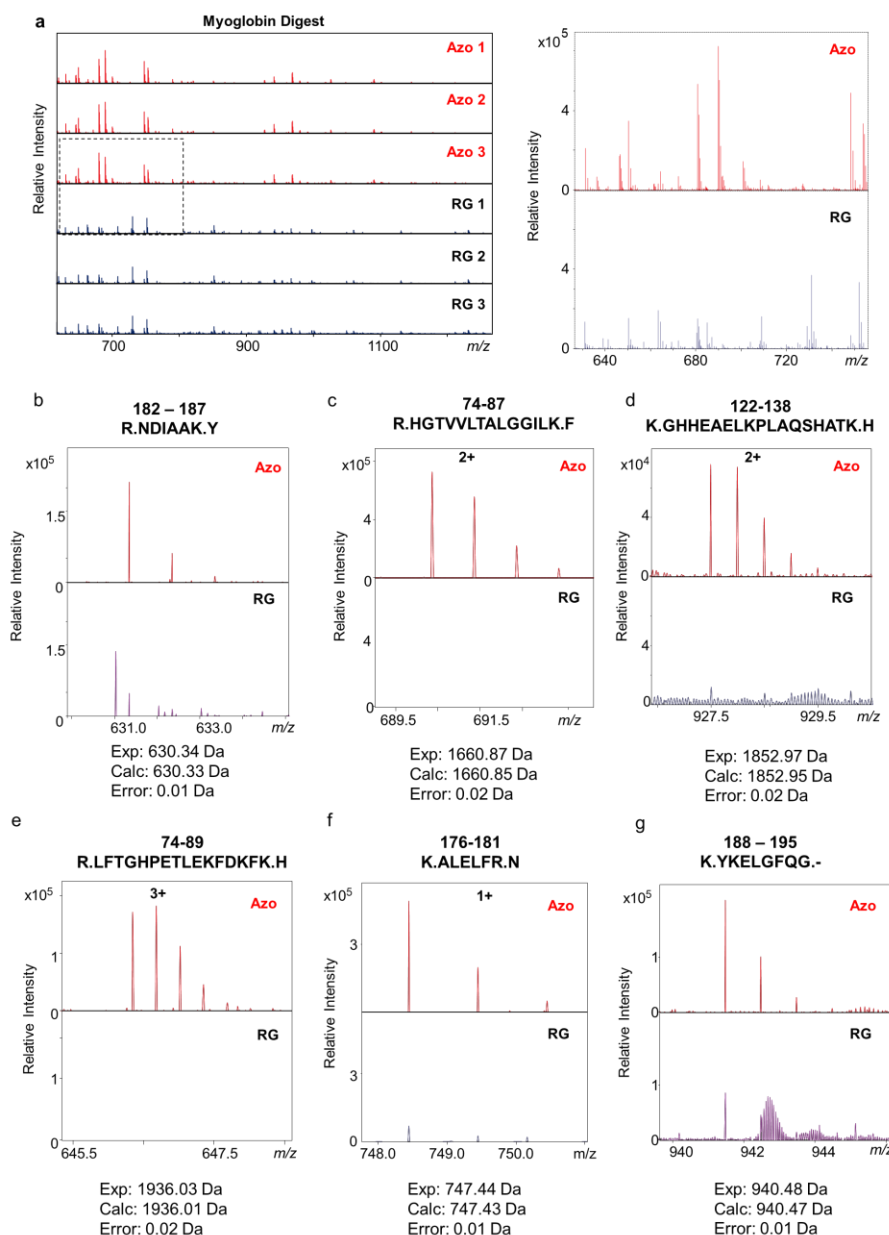


Figure 3.9. A direct comparison between Azo and RapiGest™ (RG) in the bottom-up MS experiment with trypsin digestion of myoglobin followed by direct infusion electrospray-ionization MS analysis. Myoglobin was digested with trypsin in 0.1% RG or 0.1% Azo using their standard conditions and analyzed directly by MS. (a) The signal intensity for three replicate digestions (Azo 1 to Azo 3 and RG1 to RG3) was normalized to 1E6. (b-g) Various representative tryptic peptides were identified and their signal intensities were normalized to compare the performance of the two surfactants. It clear that the signal intensity of these peptides digested with Azo is consistently higher than that using RG.

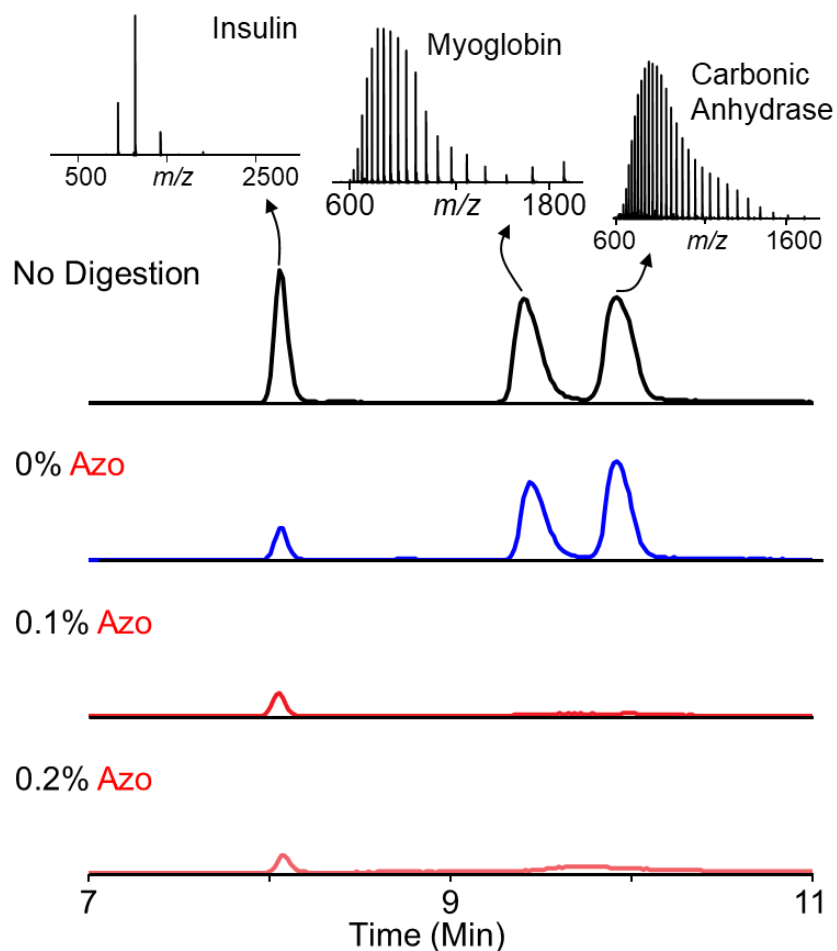


Figure 3.10. Enhanced digestion of standard proteins in the presence of Azo. Insulin, myoglobin, and carbonic anhydrase were digested using trypsin in the presence of 0, 0.1, and 0.2% Azo. The presence of intact protein was monitored by LC-MS using a maXis II ETD Q-TOF. Extracted ion chromatograms of intact proteins show increased digestion rate in the presence of Azo.

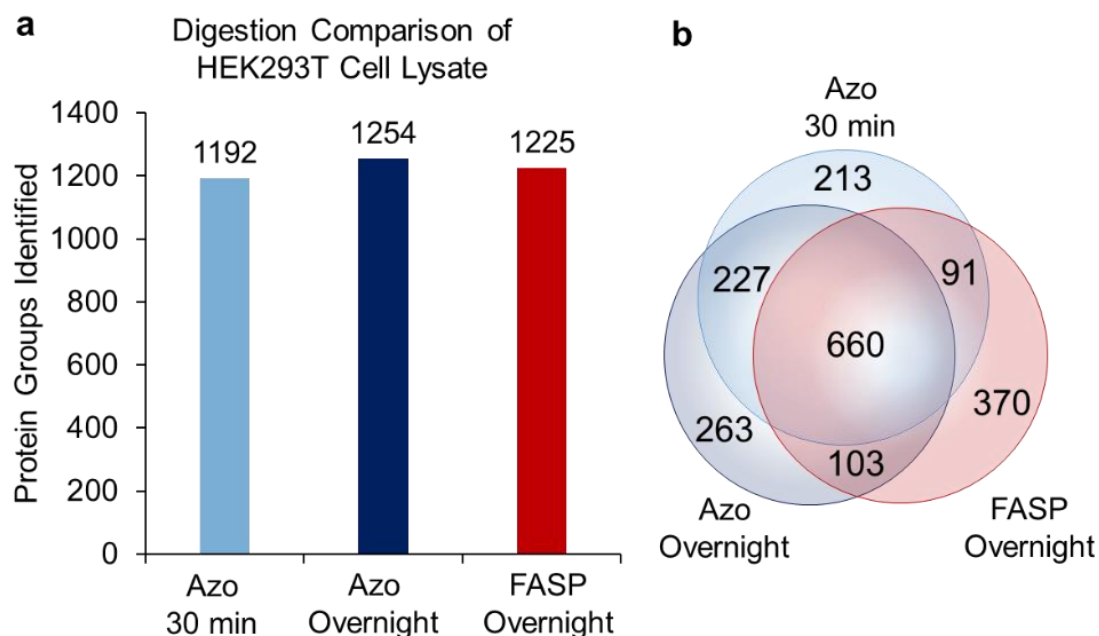


Figure 3.11. Rapid and reproducible extraction, enzymatic digestion, and LC-MS/MS analysis of proteins extracted from HEK293T cells. HEK293T proteins were extracted using Azo or SDS. Proteins were digested for 30 min or overnight using the Azo or FASP method. After LC-MS/MS analysis and MaxQuant processing (a) Azo with overnight digestion yielded the highest number of IDs followed by FASP overnight digestion and then Azo 30 min digestion. (b) Overall good overlap of identification was observed between the three methods. Data represent the combined results from three digestion and LC-MS/MS experiments. Data collected using Maxis II Q-TOF.

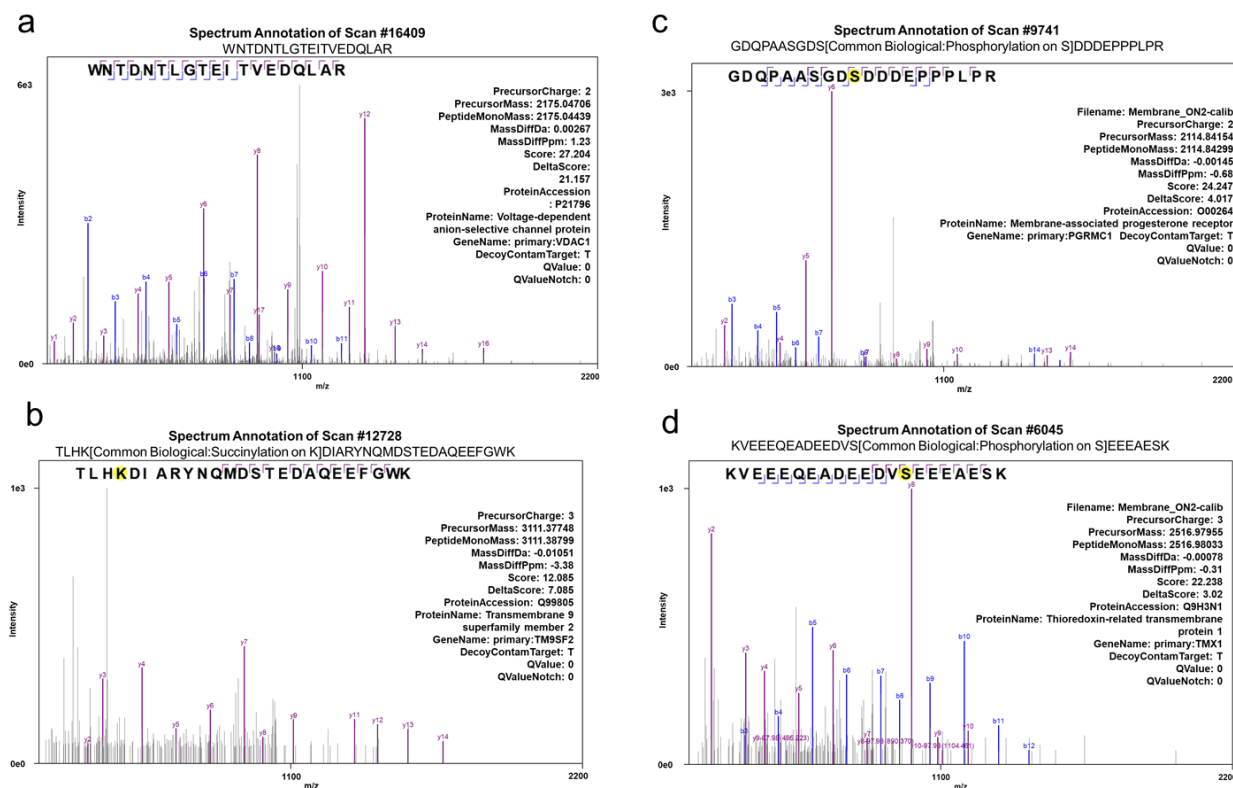


Figure 3.12. Representative MS/MS spectra using Azo-aided digestion of enriched membrane proteins. Peptides from (a) voltage-dependent anionic-selective channel protein, (b) succinylated transmembrane 9 superfamily member 2, (c) phosphorylated membrane-associated progesterone receptor, and (d) phosphorylated thioredoxin-related transmembrane protein 1 were identified and annotated using MetaMorpheus software.¹⁸⁵

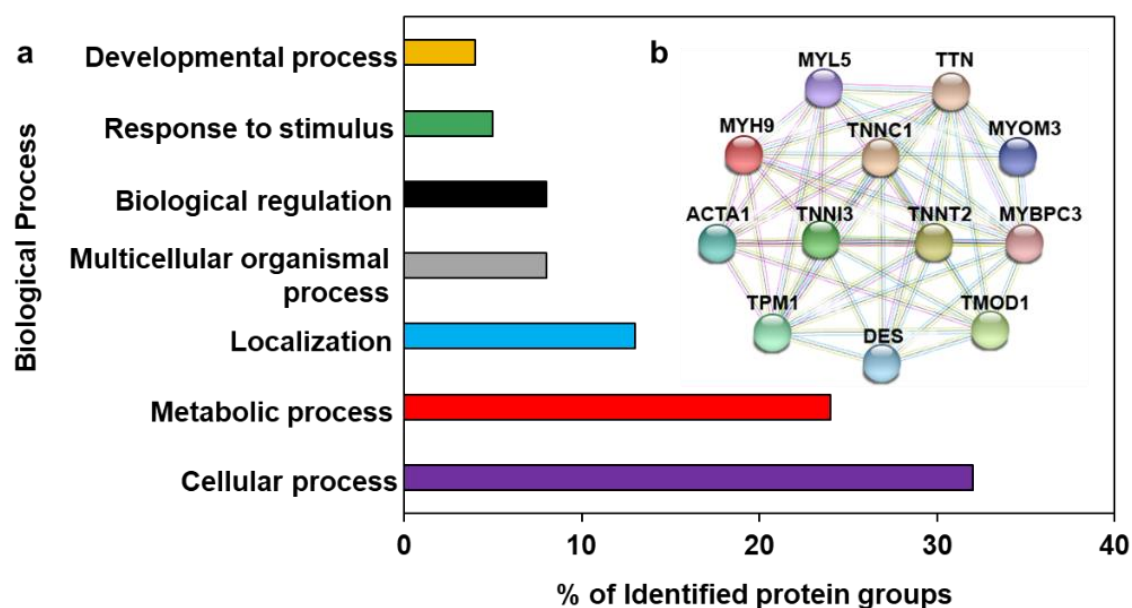


Figure 3.13. Azo-enabled bottom-up proteomics analysis of cardiac tissue. Human proteins identified from cardiac tissue using Azo-aided shotgun proteomics sorted based on their biologic processes. Representative interactome (using String analysis) of cardiac sarcomeric proteins identified by bottom-up proteomics. Data collected using a Maxis II Q-TOF.

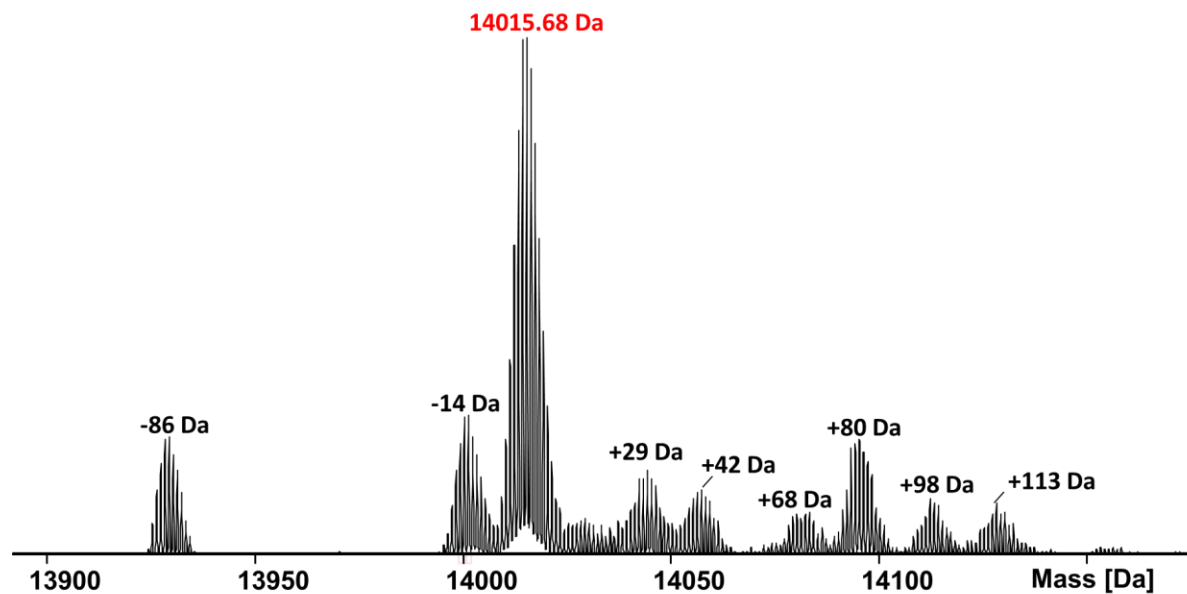


Figure 3.14. Representative mass spectrum demonstrating proteoform complexity. Mass spectrum from LC-MS analysis of cardiac tissue was deconvoluted using Maximum Entropy and masses determined using SNAP algorithm. Possible proteoforms were assigned based on their mass differences from the predominate 14015.68 Da proteoform peak.

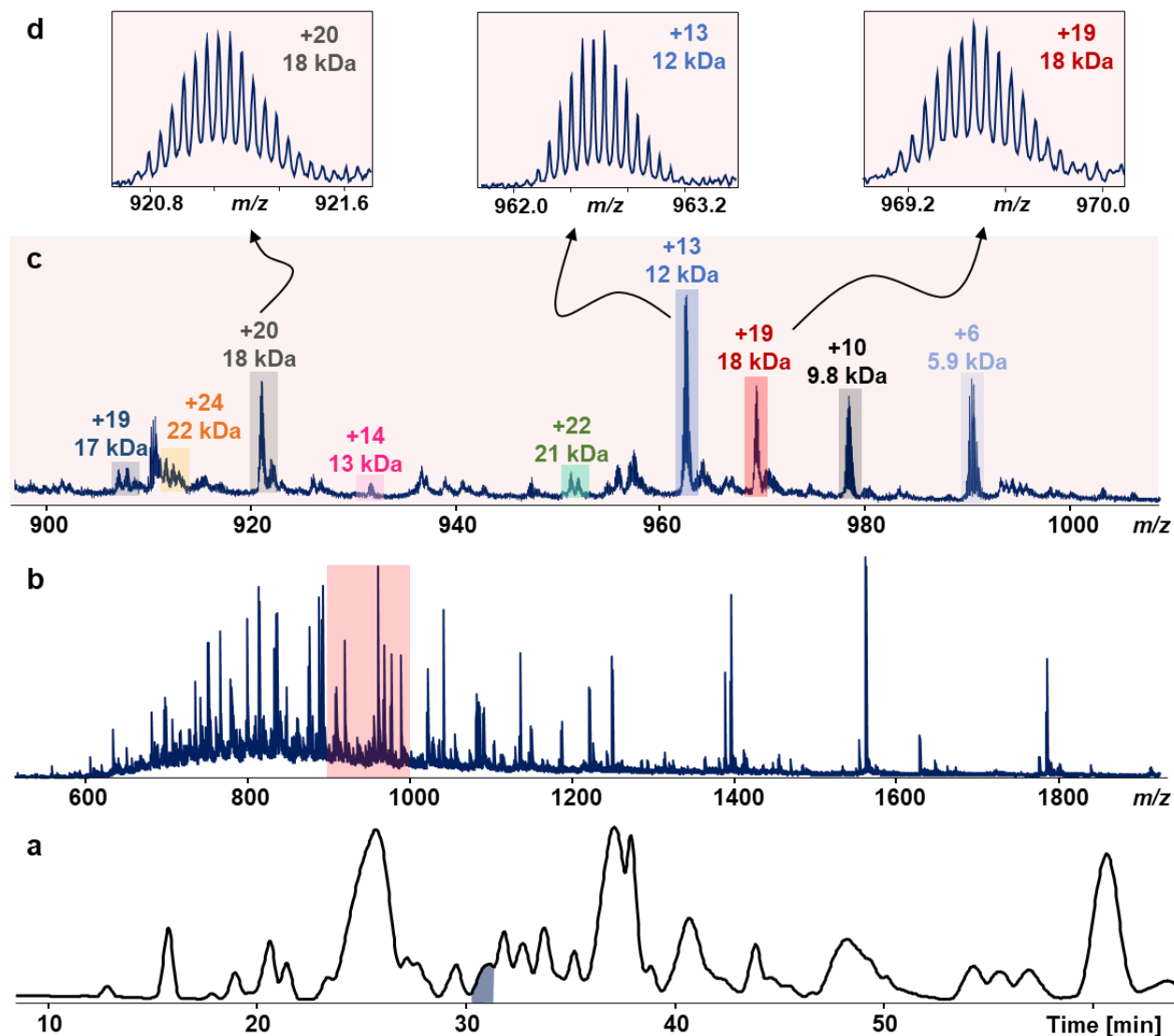


Figure 3.15. Intact protein LC-MS analysis of cardiac proteins extracted using Azo. (a) Base-peak chromatogram showing intact proteins from human cardiac tissue. (b) Representative intact mass spectrum showing co-eluting proteins from 22-5.9 kDa. (c) Zoom-in of 900-1000 m/z demonstrating the sample complexity. (d) Charges and masses determined by SNAP algorithm. Data collected using Maxis Q-TOF.

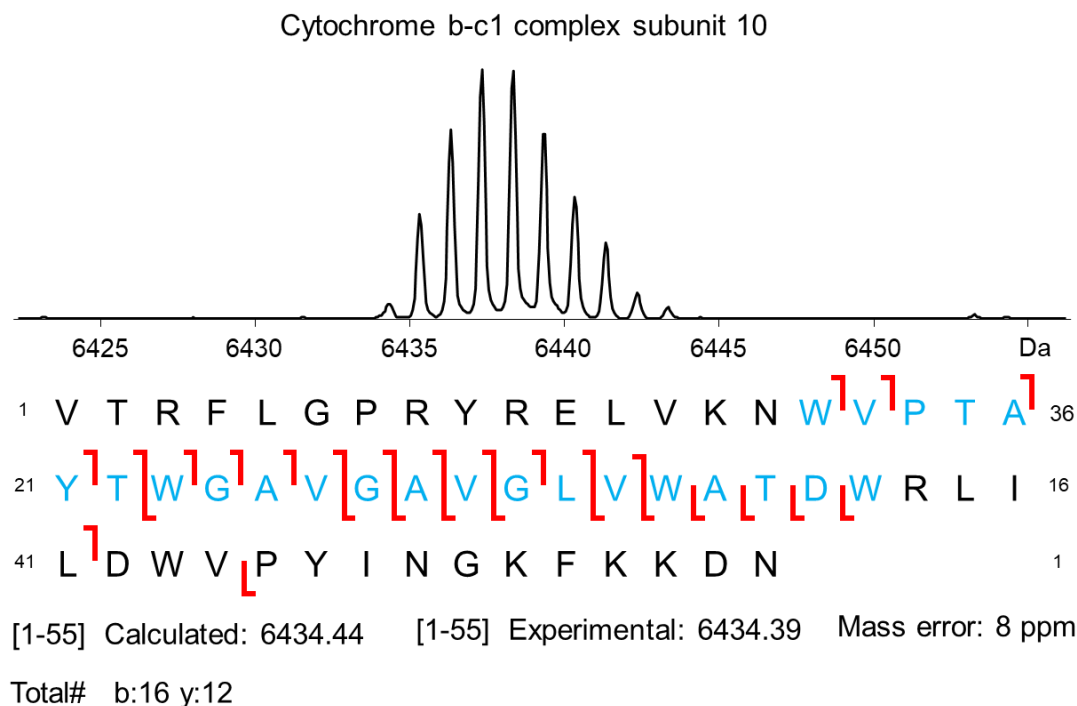


Figure 3.16. Top-down analysis of integral membrane protein cytochrome b-c1 complex subunit 10. Intact human cytochrome b-c1 complex subunit 10 with fragmentation map from online CID. The blue region indicates the transmembrane domain. Data collected using Maxis II Q-TOF.

* Table S1-S4 excel spreadsheets

3.7.1. Improvement of protein identifications

We note that the current number of protein identifications in this study was achieved using a one-dimensional LC-MS/MS analysis with an Impact II or maxis II Q-TOF mass spectrometer. Conceivably, we envision that with the use of advanced mass spectrometry technology^{191, 192} optimized for bottom-up proteomics and implementation of multidimensional separation techniques,^{8, 9, 193} the number of protein identifications will be further increased. However, considering that protein digestion is the most time-consuming step in bottom-up proteomics,^{194, 195}

shortening this step in the workflow without loss of digestion efficiency dramatically improves the throughput, which bottom-up analyses for a wide range of applications especially those requiring high-throughput analysis in clinical studies.

3.7.2. *Strategies for membrane protein analysis*

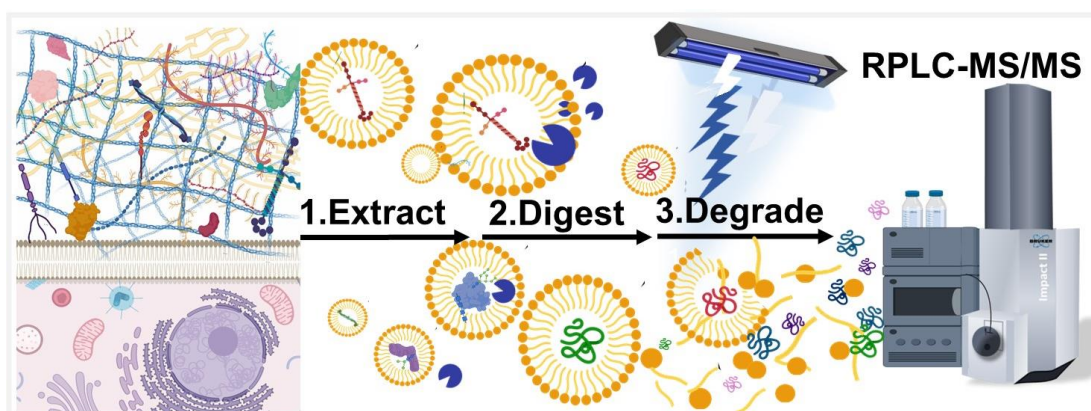
Previously, we demonstrated the use of multiple HEPES buffer extractions to deplete abundant soluble proteins before using Azo to solubilize the remaining insoluble proteins, thus enabling top-down analysis of integral membrane proteins.³⁰ Here, instead of multiple extractions, we implemented a cloud-point single extraction to focus on improving the bottom-up workflow for membrane proteomics by, using Triton X-114, to enrich membrane proteins from HEK293T cells.^{182, 183} Due to the MS-incompatibility of Triton surfactants¹⁸⁴ which require surfactant removal before MS analysis, thus precipitation/resolubilization is needed. Subsequently, we used Azo to solubilize the precipitated protein pellets enriched with integral membranes for a more effective analysis of this important class of proteins.

3.8. Acknowledgments

The financial support was provided by the NIH R01 grant GM117058 (to S.J. and Y. G.). Y.G. also would like to acknowledge NIH grants GM125085, HL109810, and HL096971 and the high-end instrument grant S10OD018475 (to Y.G.). K.A.B. would like to acknowledge support from the Training Program in Translational Cardiovascular Science, T32 HL007936-19. T.T. was supported by the NIH Chemistry Biology Interface Training Program, T32GM008505. S.K. was supported by the NSF Graduate Research Fellowship Program under Grant No. DGE-1747503 and

WiscAMP-DB program under grant No. HRD-1612530. We also would like to thank Dr. Rosa Viner for the helpful discussions.

Chapter 4. Photocleavable Surfactant-Enabled Extracellular Matrix Proteomics



This chapter has been published and is adapted from:

Knott, S.;* **Brown K.A.**;* Josyer, H.; Mitchell, S.; Carr, A.; Inman, D.; Millikin, R.; Smith, L; Friedl, A.; Ponik, S.; Ge, Y., Photocleavable Surfactant Enabled Extracellular Matrix Proteomics. Submitted *Co-first authors

4.1. Abstract

The extracellular matrix (ECM) provides an architectural meshwork that surrounds and supports cells. The dysregulation of heavily post-translationally modified ECM proteins directly contributes to various diseases. Mass spectrometry (MS)-based proteomics is an ideal tool to identify ECM proteins and characterize their post-translational modifications, but ECM proteomics remains extremely challenging due to the insoluble nature of the ECM. Herein, enabled by effective solubilization of ECM proteins using our recently developed photocleavable surfactant, Azo, we have developed a streamlined ECM proteomic strategy that allows fast tissue decellularization, efficient extraction and enrichment of ECM proteins, and rapid digestion prior to reversed-phase liquid chromatography (RPLC)-MS analysis. 212 and 280 unique ECM protein from mouse mammary tumors have been identified using 1D and 2D RPLC-MS/MS, respectively. Moreover, 585 (from 1DLC-MS/MS) and 3,623 (from 2DLC-MS/MS) post-translational modifications of ECM proteins, including glycosylation, phosphorylation, and hydroxylation, were identified and localized. This Azo-enabled ECM proteomics strategy will streamline the analysis of ECM proteins and promote the study of ECM biology.

4.2. Introduction

The extracellular matrix (ECM) consists of approximately 300 proteins including collagens, fibronectins, laminins, and proteoglycans that, together with ~900 associated factors, forms an architectural meshwork that provides stability for the surrounding cells.¹⁹⁶⁻²⁰⁰ Serving as a critical regulator of cell behaviors such as adhesion, migration, and proliferation, the ECM responds and communicates via biochemical cues to the intracellular milieu.¹⁹⁸ ECM proteins are

also known to be heavily post-translationally modified, most notably by glycosylation and hydroxylation.^{201,202} Dysregulation of ECM protein expression and post-translational modification (PTMs) directly contribute to disease progression and regulate ECM protein structures, functions, and interactions contributing to pathogenesis.¹⁹⁶⁻²⁰⁴ Particularly, the ECM is increasingly recognized as a major driver in tumor metastasis, which contributes to 90% of the cancer deaths.²⁰⁵⁻²⁰⁷ However, the biochemical characterization of ECM proteins remains a daunting task due to the insoluble nature of the ECM.²⁰⁸⁻²¹⁰

Advances in mass spectrometry (MS)-based proteomics make it an ideal tool to identify ECM proteins and characterize their PTMs occurring within the ECM microenvironment.^{197, 202, 211-216} Despite its immense potential, ECM proteomics remains extremely challenging mainly due to the dense network of cross-linked, fibrous proteins, which make it exceedingly difficult to solubilize, digest, and analyze by MS.²¹³ Moreover, the complexity and dynamic range of proteins present in the tissue lysates often result in underrepresentation of the important ECM sub-proteome in global proteomics studies.²¹⁷ Because of this significant bottleneck in sample preparation for high-throughput ECM proteomics, only a very limited number of ECM proteomics studies have been accomplished.^{197, 202, 211-216} The current protocols for ECM proteomic analysis typically have lengthy and labor-intensive workflows that often include multiple digestion steps, some including chemical digestion using toxic chemicals such as cyanogen bromide (CNBr).^{214, 215, 218}

To address these challenges, here we developed a photocleavable surfactant-enabled ECM proteomics strategy to streamline the enrichment, extraction, and digestion steps for ECM proteomics (Figure 1). Specifically, we established a new decellularization/extraction method, enabled by our recently developed photocleavable anionic surfactant, Azo,^{35, 219} for the efficient enrichment of ECM proteins, which eliminates the need for multiple digestion steps, and

minimizes the sample clean-up, before reversed-phased liquid chromatography (RPLC)-MS analysis. Our sample preparation method takes approximately 6 h (0.5 h for decellularization, 1.5 h for extraction of the insoluble pellet, and 3.5 h for reduction, alkylation, and trypsin digestion) compared to the conventional approaches that often take several days due to lengthy decellularization and multiple enzymatic digestions,^{214, 220} representing a significant improvement in the throughput of ECM proteomics. Specifically, we chose to analyze tumor tissues harvested from the transgenic mouse mammary tumor virus (MMTV) polyomavirus middle T (PyVT) mouse model as it represents a classic transgenic model for studying the microenvironment of metastatic human breast cancer²²¹⁻²²⁴ and extensive collagen deposition is a pathological hallmark of many cancers.²²⁵

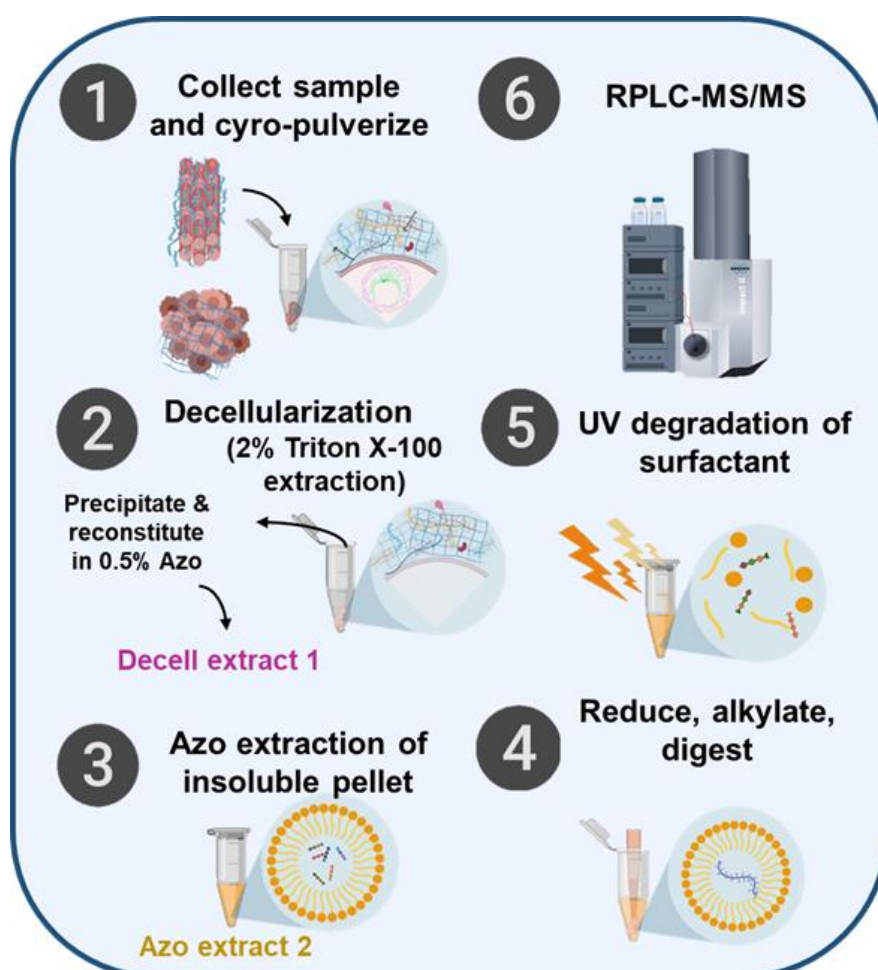


Figure 4.1. Schematic illustration of the Azo-enabled ECM proteomics strategy. (1) Tissues are collected, cryo-pulverized, and (2) undergone rapid decellularization in 2% Triton X-100. After centrifugation, the supernatants are collected, precipitated, and reconstituted in 0.5% Azo labeled as “Decell extract 1” for downstream analysis. (3) Azo was added to the insoluble tissue pellet to extract the remaining proteins, resulting in “Azo extract 2”. (4) Both Decell Extract 1 and Azo extract 2 were reduced, alkylated, and digested with trypsin. (5) Azo was degraded by UV-irradiation and (6) the digested peptides were analyzed by RPLC-MS/MS.

4.3. Development of a High-throughput ECM Proteomic Approach

For characterization of the mouse tumor tissue, we considered both the core ECM domain-containing regions as well as regulatory and secreted ECM associated proteins, which constitute “the matrisome” as defined previously.^{196, 197} The core matrisome was classified into three subcategories, collagens, glycoproteins, and proteoglycans, whereas regulators, secreted factors, and affiliated proteins were classified as associated matrisome proteins (Figure 2a). Characterizing the ensemble of core and associated matrisome proteins is critical to achieving a comprehensive understanding of ECM biology.^{196, 197} However, core matrisome proteins, like collagens, are generally considered extremely challenging to be analyzed by proteomics since they are essentially insoluble,²²⁶ and resistant to common extraction and enzymatic digestions.²²⁷

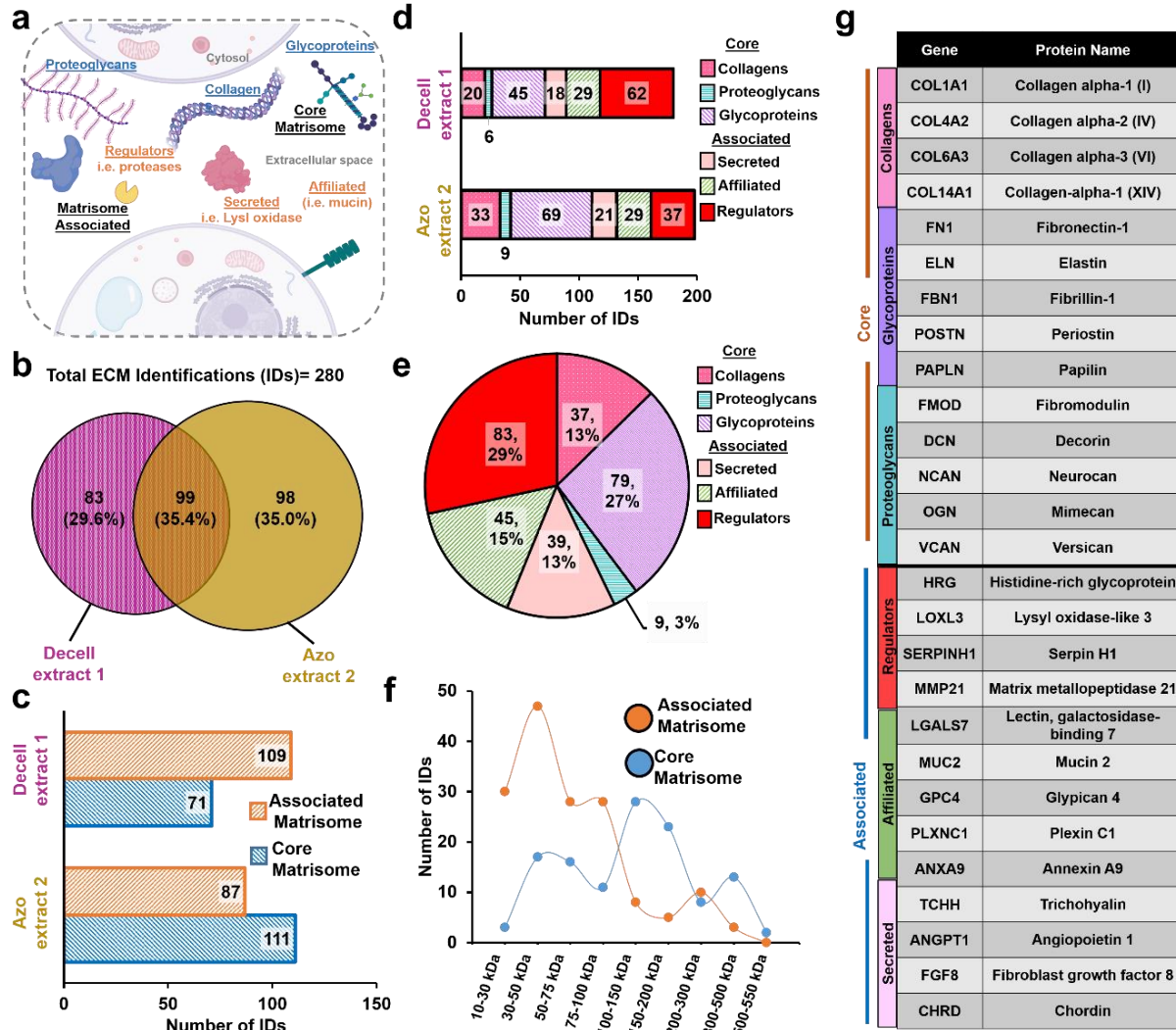


Figure 4.2. ECM proteomics of mammary tumor tissue. **a.** Schematic of matrisome ECM composition and annotation.^{196, 197} **b.** Venn diagrams showing the overlap of ECM protein identifications found in the Decell extract 1 and Azo extract 2 identified using 2D RPLC MSMS. **c.** Division distribution of ECM identifications in each extract. **d.** Category distribution of ECM identifications in each extract. **e.** Pie chart illustrating the composition of all ECM identifications. **f.** The molecular weight (MW) distribution of all ECM identifications. **g.** Table of representative core and associated matrisome proteins.

Typically, enrichment of the core matrisome proteins is obtained by decellularization,²²⁸ wherein the tissue is passively immersed in a buffer, such as SDS or Triton X-100, for several hours or days to remove soluble cellular material (proteins, lipids, metabolites, etc.), leaving an

intact ECM.^{214, 215, 229} However, since an intact scaffold is not required for proteomics, we mechanically homogenized the tissue in Triton X-100 buffer²³⁰ to dramatically increase the throughput. The samples were centrifuged and the supernatant was collected and labeled as “Decell extract 1”, and pellets of insoluble proteins remained (Figure S1). Commonly, proteins extracted during decellularization are discarded²¹⁴ presumably because few ECM proteins are extracted during this step. To investigate whether the decellularization extraction contains soluble ECM factors, the proteins in the Decell extract 1 were precipitated and reconstituted in 0.5% Azo for downstream LC-MS analysis (Figure S1).

Next, the insoluble tissue pellets were washed with ammonium bicarbonate (ABC) buffer (chosen for its MS-compatibility) to remove any remaining soluble proteins, surfactants, or salts (Figure S2). Azo was then used to extract the insoluble proteins yielding “Azo extract 2” for further LC-MS analysis. We observed excellent reproducibility using SDS-PAGE analysis across technical replicates (Figure S3). Interestingly, we qualitatively observed that Decell extract 1 was rich in lower molecular weight proteins, whereas Azo extract 2 contained many higher molecular weight species (Figure S3).

Subsequently, Decell extract 1 and Azo extract 2 were digested in-solution with trypsin for 2 h followed by rapid degradation of Azo³⁵ by UV light and subsequent analysis by RPLC-MS/MS. Recently we developed a high-throughput bottom-up proteomics method enabled by this photocleavable surfactant for robust protein extraction, rapid enzymatic digestion, and subsequent MS-analysis without additional sample clean-up following UV degradation.³⁵ Identification and relative protein abundance including reproducibility across replicates was determined with MetaMorpheus¹⁸⁵ and FlashLFQ²³¹ software using intensity-based normalization. Scatterplots of extraction replicates with normalized intensities of ECM peptide spectral matches (PSMs) were

plotted against each other demonstrating a highly reproducible method with Pearson correlation coefficients from 0.91 to 0.94 (Figure S4).

Using the normalized peptide intensities, we analyzed the relative abundance of core and associated matrisome proteins (Figure S5a and Table S1). Significantly, we observed the Azo extract 2 generally contained a higher abundance of core matrisome proteins such as collagen alpha-3 (VI) chain (COL6A3) and biglycan (BGN) (Figure S5b and Table S1). On the other hand, associated matrisome proteins, such as hemopexin (HPX) and regulator disintegrin and metalloproteinase domain 10 (ADAM10), were enriched in the Decell extract 1 (Figure S5b) (Table S1). Additionally, we evaluated the cellular locations of proteins in both extractions to better understand their protein compositions (Figure S5a). ER, mitochondrial, exosome, and membrane proteins were found in higher abundance in the Decell extract 1, whereas structural and nuclear proteins were primarily present in Azo extract 2 (Figure S5c). Overall, the results demonstrated successful protein fractionation at the extraction level and highlighted Azo's ability to solubilize and digest important ECM proteins, including notoriously insoluble, fibrillar proteins such as collagens²³² and elastin^{209, 226, 232} for a streamlined analysis.

To further assess Azo's extraction efficacy, we directly compared its performance to 8 M urea, a common reagent used for ECM protein extraction.^{31,32} Pulverized tumor tissue was decellularized with Triton-X-100, washed, and extracted with 8 M urea in 25 mM ABC, 0.5 % Azo in 25 mM ABC, or 25 mM ABC (serving as a control). SDS-PAGE analysis demonstrated Azo extracted a unique protein profile compared to the other conditions (Figure S6a). Using RPLC-MS/MS analysis, we found the relative abundance of collagen species, a common benchmark for ECM enrichment,^{197, 208} was significantly increased in the Azo extract compared to the urea or ABC alone (Figure S6b and Table S2). Hence, this demonstrated the superior

performance of Azo in solubilizing ECM proteins. Moreover, the use of photocleavable Azo eliminated the time-consuming desalting steps, required when using urea, and greatly improved the throughput of ECM proteomics.

Overall, we identified 212 ECM proteins using 1D RPLC-MS/MS. 66 and 83 proteins were uniquely identified in the Decell extract 1 and Azo extract 2 samples, respectively, and 63 proteins were identified in both (Figure S7a). Next, we investigated whether the addition of protein fractionation by high-pH RPLC could increase the proteome coverage. This 2D high-pH low-pH RPLC approach contributed 68 new unique identification (a 33% increase) compared to using 1D RPLC-MS/MS (Figure S7a). Although the increase in protein identification was moderate, we demonstrated our method could be easily combined with additional separation steps to increase proteome coverage.

Combined, we identified 280 unique ECM proteins from mouse mammary tumor tissue using this approach (Figure 2b and Table S3-4). Importantly, we observed both Decell extract 1 and Azo extract 2 contained many matrisome proteins. Of the ECM protein identifications, 99 ECM proteins were identified in both extracts, whereas 83 and 100 were uniquely identified in Decell extract 1 and Azo extract 2, respectively, which indicates that some proteins are preferentially solubilized and present in either Decell extract 1 or Azo extract 2 (Figure 2b).

We further investigated the protein compositions of each extract, observing that 71 core matrisome proteins and 109 associated proteins were identified in the Decell extract 1 whereas 111 core matrisome proteins and 87 associated proteins were identified in Azo extract 2 (Figure 2c). The results indicate enrichment of core matrisome proteins in the Azo extract 2, but also highlight the importance of analyzing the commonly disregarded decellularization fraction, Decell extract 1. Additionally, Decell extract 1 and Azo extract 2 contained similar numbers of proteoglycans,

secreted, and affiliated proteins (Figure 2e). On the other hand, ECM regulatory proteins were enriched in the Decell extract 1 whereas a large number of collagens and glycoproteins were identified in the Azo extract 2. All categories of ECM proteins were well represented, except proteoglycans, using our approach (Figure 2f). We observed that a larger number of higher molecular weight (>100 kDa) protein identified were core matrisome proteins than associated matrisome proteins (Figure 2d). A representative list of identified ECM proteins is shown in Figure 2g.

ECM proteins are known to be heavily post-translationally modified. Here, our Azo-aided workflow in combination with MetaMorpheus¹⁸⁵ software for global PTM discovery enabled the identification of a diverse range of ECM PTMs with potential biological relevance²⁰¹ including proline, asparagine, and lysine hydroxylation, acetylation, phosphorylation, and glycosylation (Figure S8). In total, 585 PTM sites were identified from mouse tumor tissue using 1D RPLC-MS/MS (Figure S8 and Table S5). Significantly, we observed a dramatic increase in PTM identifications, 3,623 total PTMs, with the 2DLC-MS/MS workflow (Figure S8 and Table S5), despite that the 2D approach provided only a moderate increase in the identifications of ECM proteins. Overall, proline hydroxylation, which is essential for the stability of collagen fibrils,²³² accounted for 78% of all identified modifications from mouse mammary tumor tissue (Figure S8). Figure 4 shows some representative examples of unambiguously localized PTMs, such as proline hydroxylation of collagen alpha-2 (I) chain (COL1A2) and asparagine β hydroxylation, on proteins like fibrillin-1 (FBN1) (Figure 4a,b). N-acetylglucosamine (HexNAc) was localized to a threonine residue of Host cell-factor-1 (Figure 4c). Additionally, we have localized a phosphorylation site to a serine residue of collagen alpha-1 chain (III) (COL3A1) (Figure 4d).

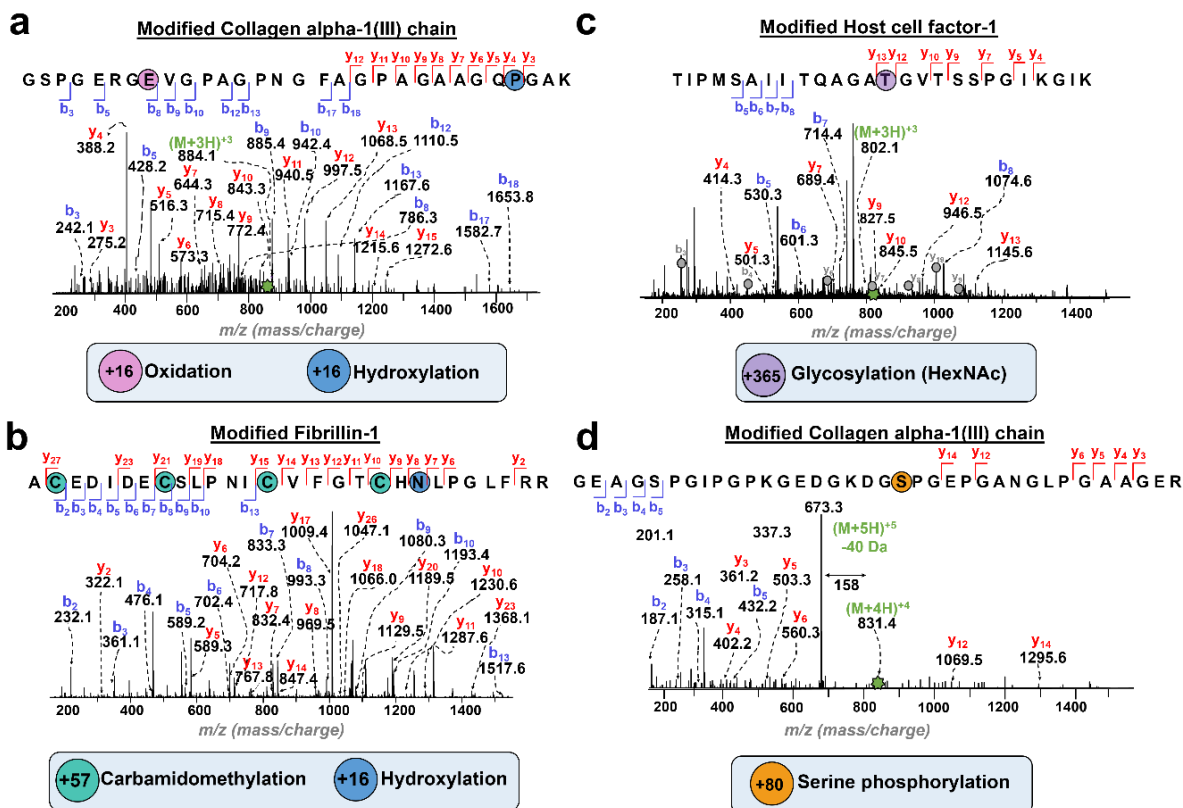


Figure 4.3. Identification of PTMs in the ECM proteins. Representative MS/MS spectra of identified ECM PTMs. **a** Collagen alpha-2 (I) chain peptide with proline (P) hydroxylation (+ 16 Da). Precursor ion, $[(M+2H)^{2+}]$, 788.9 m/z, Expt'l: 1575.80 Da, 0.8 ppm]. **b**. Fibrillin-1 peptide modified with cysteine (C) carbamidomethylation (+ 57 Da) and asparagine (N) hydroxylation (+16 Da). Precursor ion $[(M+3H)^{3+}]$, 1102.4 m/z, Expt'l: 3304.30 Da, 1.4 ppm]. **c**. Host cell-factor-1 (HCFC1) peptide with (N)-linked acetylhexosamine (HexNAc) glycosylation (+203 Da). Precursor ion $[(M+3H)^{3+}]$, 802.1 m/z, Expt'l: 2403.3 Da, 0.6 ppm]. HCFC1 co-eluted with a highly abundant peptide identified as splicing factor 3b subunit 2. Corresponding b and y ions are denoted with gray circles. **d**. Collagen alpha-1 (III) chain with localized serine (S) phosphorylation (+80 Da). The peptide precursor $[(M+3H)^{3+}]$, 884.1 m/z, Expt'l: 3,321.51 Da, 5.6 ppm].

4.4. Conclusion

In summary, for the first time, we developed a streamlined ECM proteomics method enabled by a photocleavable surfactant, Azo, which addressed several challenges in the conventional ECM proteomics workflows. Most notably, Azo facilitated robust extraction of

“insoluble” ECM proteins, aided in rapid trypsin digestion, and subsequently can be easily degraded by UV-irradiation before MS. This Azo-enabled ECM sample preparation including decellularization, ECM extraction, reduction, alkylation, and digestion takes about 6 h compared to several days or a week using the conventional approaches. This sample preparation does not require chemical digestion, multiple enzymes, or deglycosylation. Moreover, we demonstrated the importance of analyzing the commonly discarded decellularization fraction to identify soluble ECM and ECM-associated proteins. Using both 1D and 2D RPLC-MS/MS, we have established an ECM protein catalog consisting of 280 ECM proteins from mouse tumor tissue and have further identified 3,623 total PTMs for ECM proteins including hydroxylation, phosphorylation, and glycosylation. We envision this Azo-enabled ECM proteomics strategy will streamline the analysis of ECM proteins and promote an understanding of ECM biology in various diseases such as tumor metastasis.

4.5. Supplemental Information

4.5.1. Chemical and Materials

Chemicals and reagents were purchased from MilliporeSigma Inc. (St Louis, MO, USA) without further purification unless otherwise noted. Acrylamide gels (7%) were made in-house using acrylamide/bis-acrylamide (37:5:1) solution from Hoefer (Holliston, MA, USA). Criterion TGX Stain-Free precast gels (7.5%) were purchased from BioRad (Hercules, CA, USA). Halt protease/phosphatase inhibitor cocktail, tris (20-carboxyethyl) phosphine (TCEP), bicinchoninic acid (BCA) reagent were purchased from ThermoFisher Scientific (Waltham, MA, USA). Water (HPLC grade), acetonitrile (ACN) (HPLC grade), and phenylmethylsulfonyl fluoride (PMSF)

were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Trypsin Gold was purchased from Promega (Madison, WI, USA). Azo was synthesized in-house as described previously.²¹⁹

4.5.2. *Mouse Mammary Tumor Tissue*

Mice were maintained and bred at the University of Wisconsin under the approval of the University of Wisconsin Animal Use and Care Committee. Transgenic mice were created using the MMTV-PyVT^{+/+} for the development of metastatic mammary tumors. Mice were originally obtained from The Jackson Laboratory (Bar Harbor, ME, USA). After six to eight weeks, tumors were dissected, flash-frozen, and stored at -80 °C.

4.5.3. *ECM Protein Extraction*

Tissue was cryo-pulverized using a cellcrusher (Caheravirane Schull, Cork, Ireland). A rotor-stator homogenizer (Pro Scientific Inc., Oxford, CT) was used to further shear the frozen tissue samples on ice in the decellularization extraction buffer (2% Triton X-100, 25 mM ammonium bicarbonate [ABC], 0.25 mM PMSF, 0.25 mM Na₃VO₄, 2.5 mM ethylenediaminetetraacetic acid [EDTA], and 1x Halt protease/phosphatase inhibitor cocktail). The sample was centrifuged at 20,000 × *g* at 4 °C for 10 min, the supernatant was collected, and the proteins were precipitated with 4 volumes of acetone. The resulting protein pellet was allowed to air dry and dissolved in Azo buffer (0.5% Azo, 25 mM ABC, 0.25 mM PMSF, 0.25 mM Na₃VO₄, 2.5 mM EDTA, 1 mM 1,4-dithiothreitol [DTT]) using heat (10 min at 97 °C) and sonication (1 h sonication in a water bath). After centrifugation at 20,000 × *g* for 10 min, the supernatant was collected and labeled as “Decell extract 1”.

The tissue pellet resulting from the decellularization was washed with 25 mM ABC using aggressively vortexing. The sample was centrifuged again at $20,000 \times g$ for 5 min and the supernatant was discarded. After five washes, the insoluble tissue pellet was homogenized in Azo buffer and the DNA was sheared for 3 sec (20% amplitude) using a sonicator probe (ThermoFisher Scientific, Waltham, MA, USA). The solution was heated for 10 min at 97 °C, sonicated in a water bath for 1 h to solubilize pellet, and centrifuged for 10 min at $20,000 \times g$. The supernatant was removed and labeled as “Azo extract 2”. 8 M urea in 25 mM ABC or 25 mM ABC (serving as a control) was used to solubilize the pellet in second extraction step for a direct comparison with Azo. Urea extracts were dried in a SpeedVac, reconstituted in 1 % TFA and loaded onto Pierce™ C18 tips (ThermoScientific) before LC-MS. The standard washing and elution protocols from the manufacturer were used. After peptide elution with ACN, samples were dried in a vacuum centrifuge and reconstituted in 0.2% FA in water for RPLC-MS/MS.

4.5.4. Digestion and 1D-RPLC-MS/MS

Protein concentration was determined with a BCA protein assay according to the manufacture’s protocol using bovine serum albumin as a standard. For each digestion, 140 µg of protein was reduced with 15 mM DTT at 37 °C for 60 min, alkylated with 25 mM iodoacetamide (IAA) at room temperature for 30 min and then digested with trypsin at 37 °C for 2 h using a 1:50 enzyme-to-protein ratio. The reaction was quenched with 1 µL of formic acid (FA). The sample was UV-irradiated for 3 min with a 100 W mercury lamp (Nikon housing with Nikon HB-10101AF power supply; Nikon, Tokyo, Japan) to rapidly degrade the surfactant. Peptides were dried in a vacuum centrifuge and reconstituted in 0.2% FA in water. For the high pH fractions and unfractionated extracts, approximately, 30 µg of peptides were loaded and separated using a home-

packed C18, 5 μm , 250 mm x 0.25 mm column. C18, 5 μm materials were purchased from Restek (Bellefonte, PA, USA). The column was heated to 60 $^{\circ}\text{C}$ and 0.2% FA in water and 0.2% FA in ACN were used as mobile phases A (MPA) and B (MPB), respectively. Peptides were eluted over a range of 60 min with a step-wise gradient from 5-90% MPB. For the extraction comparison experiments, peptides were first loaded onto a capillary trap column 3 μm , 120 \AA UChrom C18 column purchased from nanolcms solutions (Gold River, CA, USA) column and then separated using the home-packed column as described above.

Eluting peptides were ionized via CaptiveSpray with acetonitrile NanoBooster and MS analysis was performed with a Bruker (Bruker Daltonics, Billerica, MA, USA) Impact II quadrupole-time of flight (Q-TOF) mass spectrometer. Auto MS/MS was used to select the top 30 most abundant peaks in each scan for collisionally activated dissociation (CAD).

4.5.5. SDS-PAGE

Laemmli buffer supplemented with 100 mM DTT was added to the extracts. Samples were boiled at 97 $^{\circ}\text{C}$ for 3 min and loaded onto a 7% (in-house precast) or 7% Bio-Rad TGX stain-free acrylamide gel, separated at 55 V for 30 min and 155 V until completion. Gels were stained with Coomassie Brilliant Blue R-250 for 20 min for visualization. Gels were destained overnight in methanol, acetic acid, and deionized water before imaging.

4.5.6. Offline High-pH Reverse-phase Fractionation

High-pH fractions were collected using an ACQUITY H-class UPLC system equipped with a UV detector and automated fraction collector (Waters, Milford, MA, USA). Following trypsin digestion, 300 µg peptides were loaded onto a Waters ACQUITY UPLC BEH C18, 1.7 µm, 50 mm x 2.1 mm column with MPA (98% water, 2% ACN pH 10) and separated using a step gradient of increasing concentrations of MPB (98% ACN, 2% water pH 10). Absorbance was monitored at 214 nm and 52 fractions were collected and subsequently labeled (F1, F2, etc.). A total of 5 runs were pooled and fraction concatenation was performed by combining equal distant fractions as previously described.^{233, 234} Finally, the fractions were dried in a vacuum centrifuge, reconstituted in 20 µL of 0.2% FA, and analyzed by RPLC-MS/MS as stated above.

4.5.7. Data Analysis

Bruker mass spectral files (.d) were first converted to mzML files using ProteoWizard's msconvert.⁴ Converted files were searched using an open-source bottom-up proteomics software tool MetaMorpheus version 0.0.301,¹⁸⁵ searching against the UniProt *Mus musculus* proteome databases downloaded on September 16th, 2019. Workflow parameters were set to default values: protease: trypsin; maximum missed cleavages: 2; minimum peptide length: 7; maximum peptide length: unspecified; initiator methionine behavior: variable; fixed modifications: common fixed; variable modifications: common variable; max modification isoforms: 1,024. We used a q-cutoff value of 0.01, corresponding to a 1% false discovery rate (FDR) for identifications. Only peptides with at least one confident peptide spectral match (PSM) were included. Masses reported for peptides are monoisotopic. Contaminant and decoy matches were removed from MetaMorpheus results. FlashLFQ²³¹ was used to normalize protein intensities for relative abundance comparisons between extractions. Data were uploaded to the PRIDE repository²³⁵ via ProteomeXchange.

4.5.8. Supplemental Figures

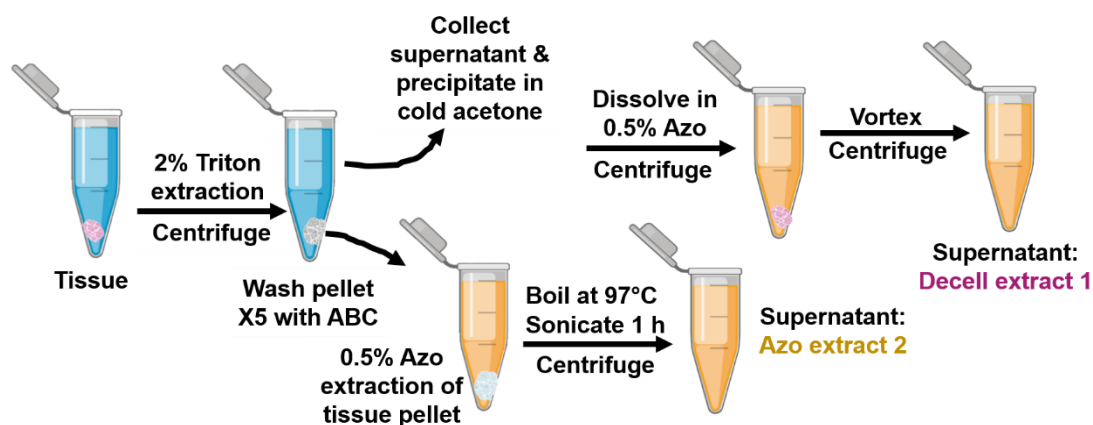


Figure 4.4. Azo enabled two-step extraction for analysis of extracellular matrix (ECM) proteins. The tissue is first rapidly decellularized with 2% Triton extraction buffer. After centrifugation, the supernatant is collected, precipitated in cold acetone, air dried, and reconstituted in Azo to yield extract 1, labeled as “Decell extract 1” The remaining insoluble pellet is washed in 25 mM ABC buffer and homogenized in Azo extraction buffer yielding extract 2, labeled “Azo extract 2”.

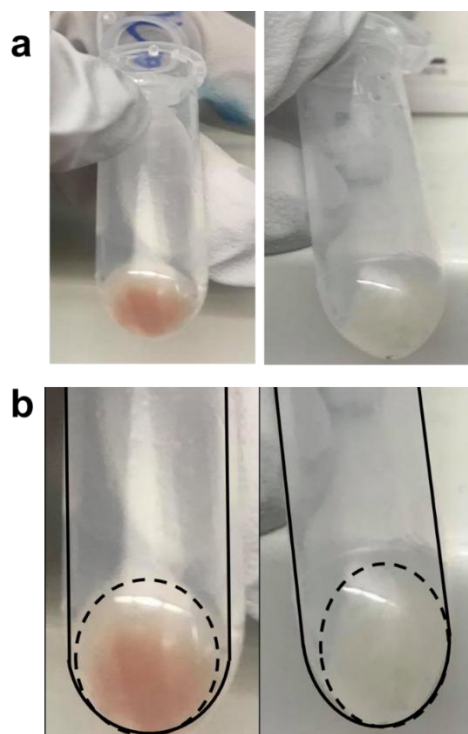


Figure 4.5. Pictures were taken after tissue decellularization and washing steps. a Insoluble pellet is pink in color after Triton X-100 decellularization and extraction of cellular proteins (“Decell extract 1”). The pellet is then washed with 25 mM ABC buffer, centrifuged to re-pellet, and the supernatant discarded. Washing is performed five times and the pellet becomes less pink with each subsequent wash, eventually leaving a transparent pellet (right). **b** Close-up and outline of insoluble pellet before (left) and after washing with ABC buffer (right). Azo was added to the decellularized pellet after washing to extract core ECM proteins (“Azo extract 2”).

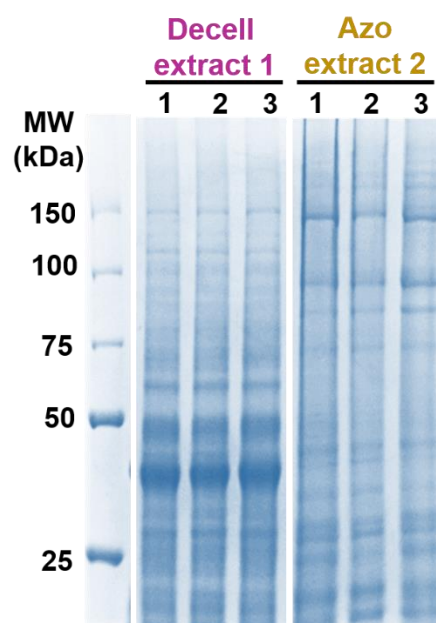


Figure 4.6. 8% SDS-PAGE analysis of representative ECM extraction replicates of Decell extract 1 and Azo extract 2.

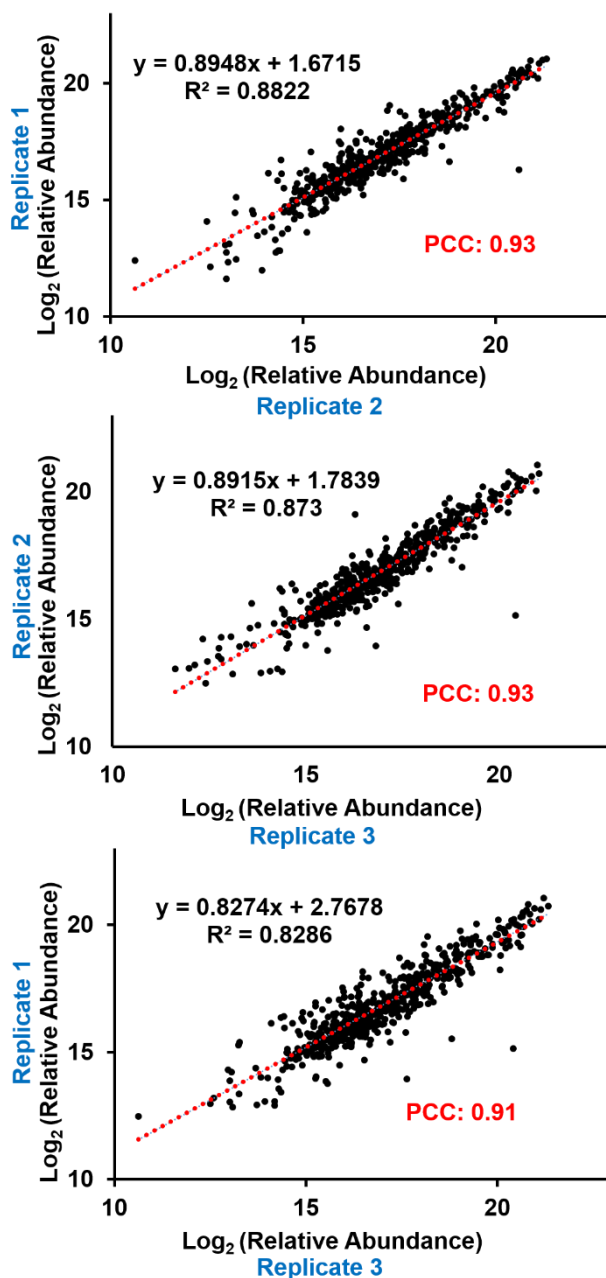


Figure 4.7. Scatter-plots of Log_2 transformed relative ECM protein abundances in extraction replicates of Azo extract 2 from mammary mouse tumor tissue. Only peptides identified in all three replicates were included and a 1% FDR cutoff was used. Pearson correlation coefficients (PCC) are shown in the bottom right.

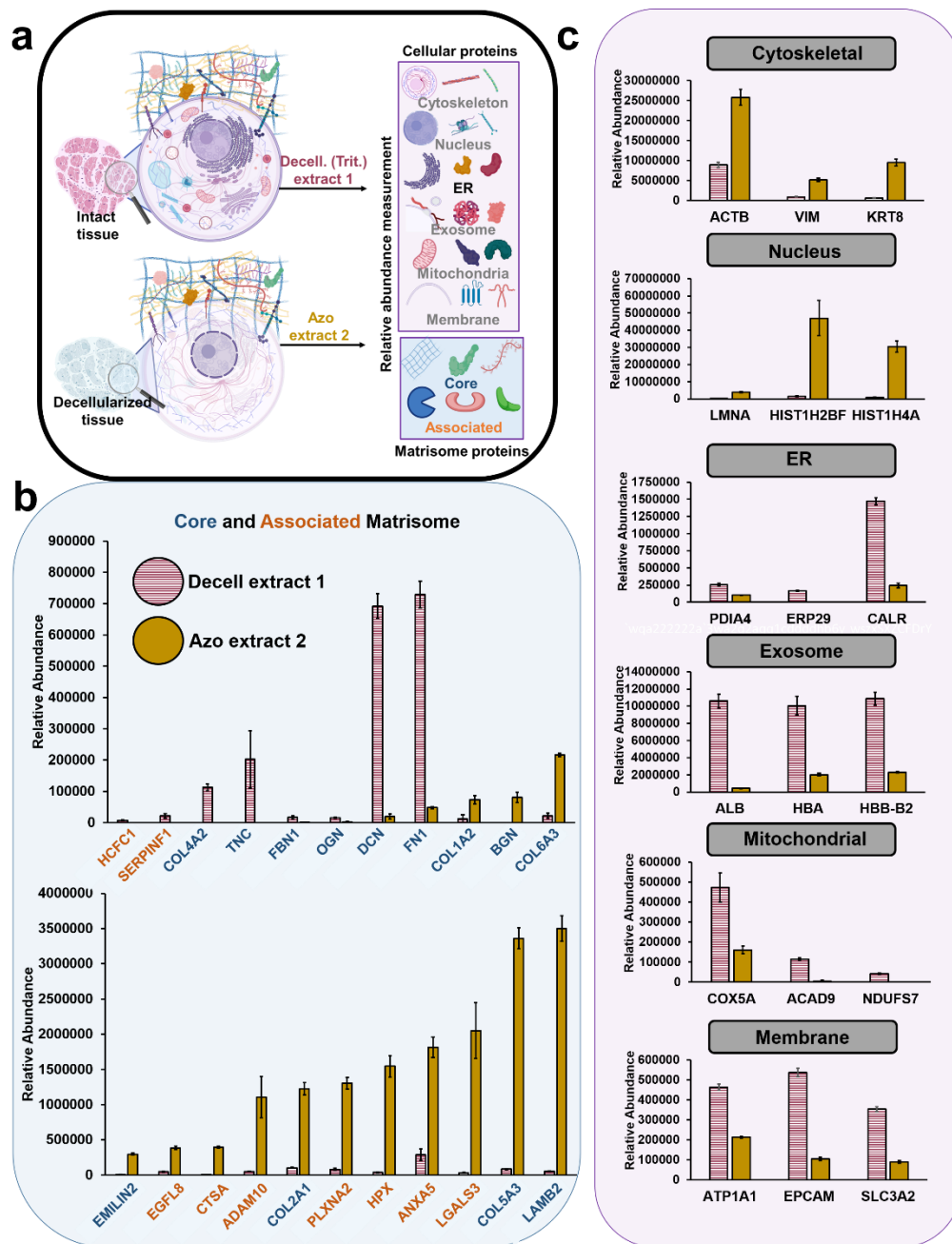


Figure 4.8. Composition of Decell extract 1 and Azo extract 2 from mammary mouse tumor tissue. **a** Intact and decellularized tissues were extracted and relative abundances of cellular and matrisome proteins were quantified in each extract. Identified protein intensities were first normalized with FlashLFQ software. The average across six replicates for each extract was calculated and plotted in the bar graph. Error bars represent standard error of the mean (SEM). **b**

Relative abundances of matrisome proteins in the respective extracts. **c** Relative abundances of cellular protein for representative cytoskeletal, nuclear, endoplasmic reticulum (ER), exosome, mitochondrial and membrane proteins identified.

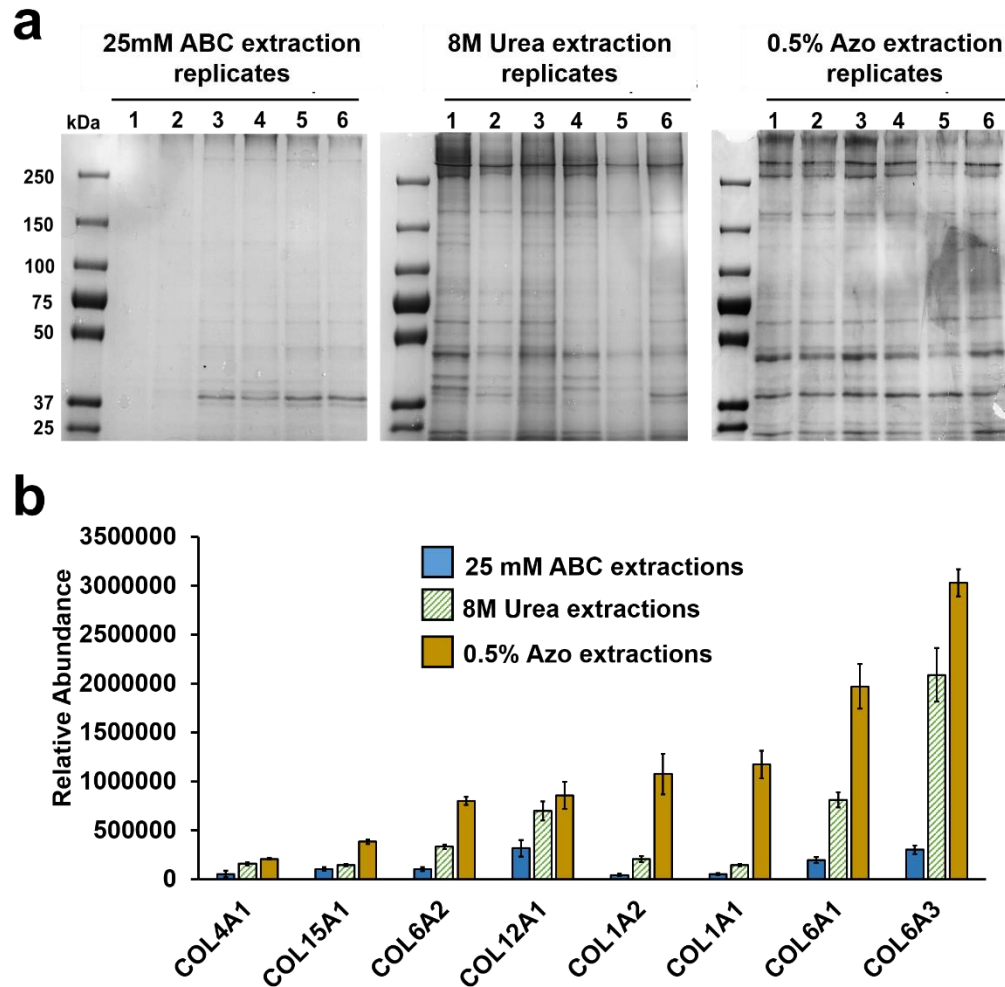


Figure 4.9. Comparison of Azo to the traditional ECM extraction buffers. Tissues were first decellularized with a 2% Triton extraction buffer to yield Decell extract 1 and pellets were subsequently homogenized in either 25 mM ABC (serving as a control), 8M urea, or 0.5% Azo extraction buffer. **a** SDS-PAGE analysis of six separate extraction replicates from mouse tissues (with equal loading volume) to evaluate the efficacy of the Azo extraction compared with the traditional methods. **b** Average normalized abundances of the collagens. Error bars represent the standard error of the mean (SEM).

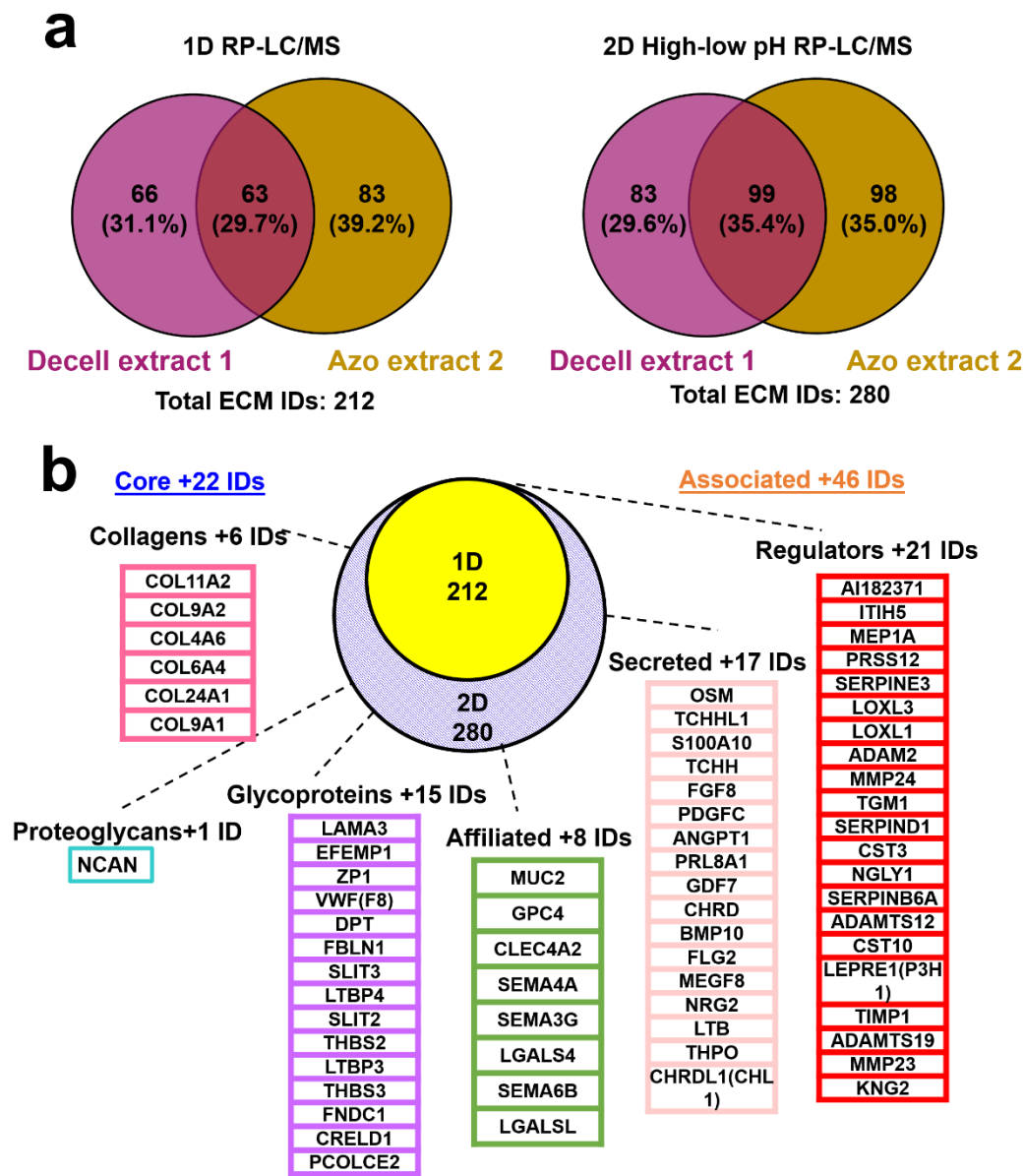


Figure 4.10. Comparison of 1- and 2-dimensional LC separation strategies for ECM proteomics. **a** Venn diagrams of the overlap in the ECM protein identifications of the Decell extract 1 and Azo extract 2 from mouse mammary tumor tissues. Extracts were analyzed with 1D RPLC-MS/MS at pH 2. For the 2D approach, high-pH RPLC at pH 10 was used to fractionate each extract into thirteen fractions. The fractions were then analyzed with RPLC-MS/MS at pH 2. **b** Core and associated matrisome protein identifications unique to the 2DLC approach.

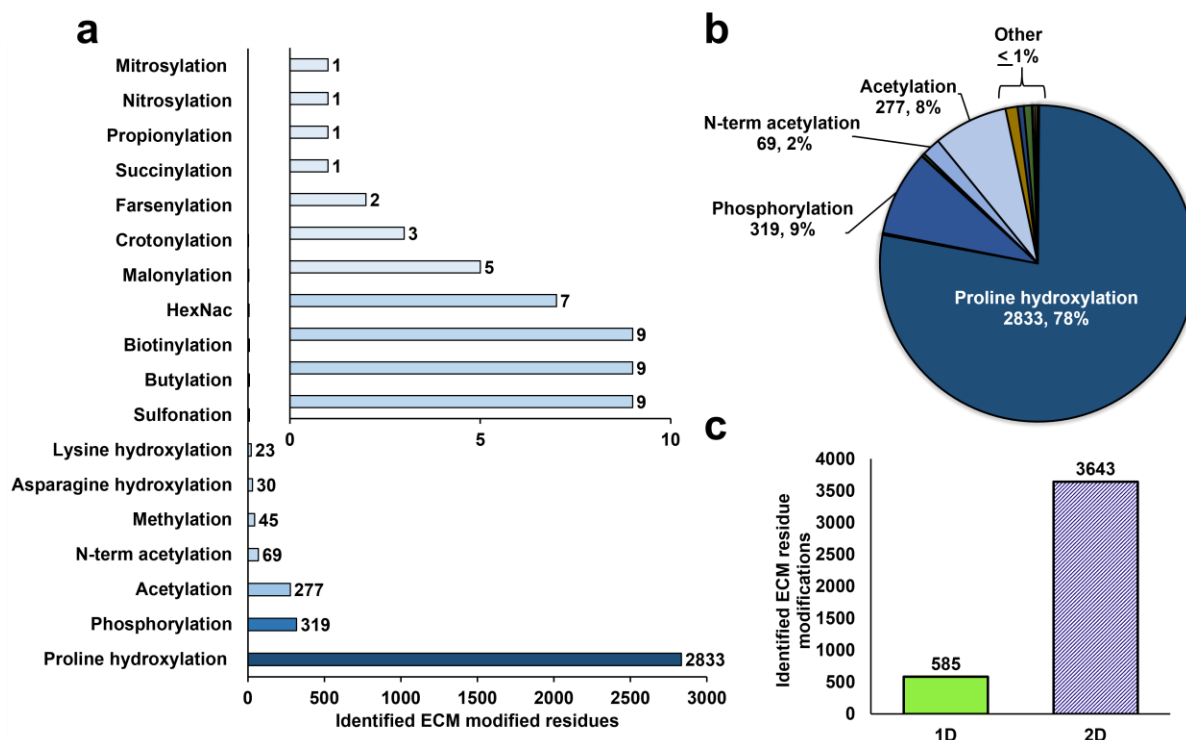
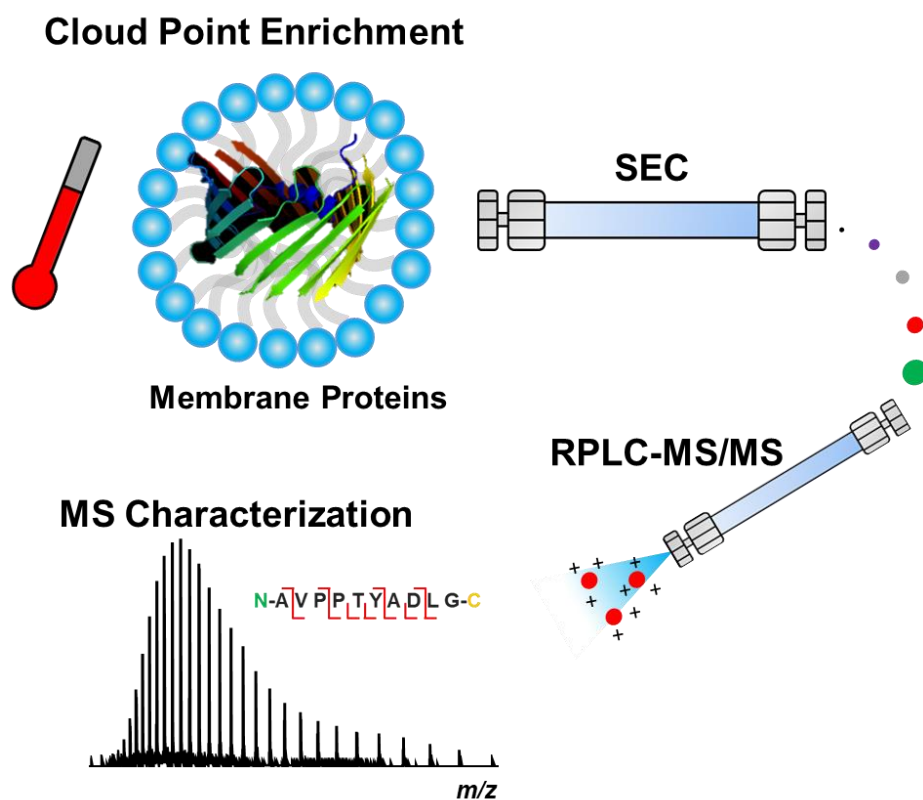


Figure 4.11. Summary of PTM sites identified in ECM proteins from mammary mouse tumor tissue. Peptide spectral matches (PSMs) of modified peptides were filtered using a 1% FDR and variable modifications were chosen. **a** Occurrence of the most abundant PTMs identified in the ECM proteins. Modifications resulting from sample handling including carbamidomethylation and oxidation were not included. **b** Occurrence of representative individual modifications identified in the 1D or 2D approach. The most abundant ECM modification, hydroxylation resulting in intramolecular and intermolecular cross-linking of collagen fibrils occurs at the modified residues. **c** Total ECM modifications identified using the 1D or 2D proteomics approach.

*Supplementary S1-S5 tables in excel files

Chapter 5. Novel Enrichment and Separation

Tools for Top-down Membrane Proteomics



This chapter is adapted from:

Brown, K.A.; Tucholski, T.; Alpert, A.; Eken, C.; Wesemann, L.; Kyrvasilis, A.; Jin, S.; Ge, Y., Top-down Proteomics of Endogenous Membrane Proteins Using a Temperature-Induced Cloud Point Enrichment and Multidimensional Liquid Chromatography-Mass Spectrometry. *In Revision*

5.1. Abstract

Although top-down proteomics has emerged as a powerful strategy to characterize proteins in biological systems, the analysis of endogenous membrane proteins remains challenging due to their low solubility, low abundance, and the complexity of the membrane sub-proteome. Here, we report a simple but effective enrichment and separation strategy for top-down proteomics of endogenous membrane proteins enabled by cloud point extraction and multidimensional liquid chromatography coupled to high-resolution mass spectrometry (MS). The cloud point extraction efficiently enriched membrane proteins using a single extraction, eliminating the need for time-consuming ultracentrifugation steps. Subsequently, size-exclusion chromatography (SEC) with an MS-compatible mobile phase (49% water, 40% isopropanol, 1% formic acid) was used to remove the residual surfactant and fractionate intact proteins (6-115 kDa). The fractions were separated further by reversed-phase liquid chromatography (RPLC) coupled with MS for protein characterization. This method was applied to human embryonic kidney cells and cardiac tissue lysates, to enable the identification of 188 and 124 endogenous integral membrane proteins respectively, some with as many as 19 transmembrane domains.

5.2. Introduction

Membrane proteins play critical roles in cellular functions and represent the largest class of therapeutic targets.^{24, 236-238} Despite accounting for one-third of protein-encoding genes, membrane proteins are generally under-characterized compared to other protein classes (e.g., <1% of structures in the Protein Database),²³⁸ because of the tremendous challenges in studying them. The proteomic analysis of these biomolecules is particularly difficult due to their high hydrophobicity (thus low solubility), low abundance in the proteome, and the complexity of the

membrane sub-proteome.^{181, 239} Characterization of biological systems at the level of proteoforms^{3, 4, 166} with top-down mass spectrometry (MS) has allowed the analysis of combinatorial post-translational modifications (PTMs) together with amino acid sequence variations and provided important insights into biology.^{14, 48, 49, 133} Given the combined technical challenges associated with membrane proteins and top-down proteomics,¹⁴ robust strategies targeting the membrane proteome remain underdeveloped.

Among the few top-down MS studies of membrane proteins, Whitelegge and colleagues first characterized purified membrane proteins to provide primary structural information, including PTMs.²⁴⁰⁻²⁴² Later, Kelleher and co-workers introduced the large-scale top-down analysis of membrane proteins from mitochondria using many lysis and ultracentrifugation steps followed by gel-eluted liquid fraction entrapment electrophoresis (GELFrEE),^{160, 243} and reversed-phase liquid chromatography (RPLC)-MS/MS with an additional prefractionation step of solution isoelectric focusing for deeper proteome coverage.¹⁰⁵ These important advances enabled the top-down MS analysis of membrane proteins, but the multi-step preparations are time-consuming and labor-intensive and generally, these studies have targeted the mitochondria membrane sub-proteome.^{67, 105, 243, 244} Recently, we established a simpler, high-throughput membrane proteomics approach enabled by a novel photocleavable surfactant, 4-hexylphenylazosulfonate (referred to as Azo) for both top-down and bottom-up MS.^{219, 245} Because Azo is a strong ionic surfactant that can solubilize all categories of proteins, many structural, nuclear, and extracellular proteins were also extracted in addition to the targeted membrane proteins, the overall fraction of membrane proteins among the enriched proteins was lower than desired (28% of identified proteins were integral membrane proteins). On the other hand, mild nonionic surfactants show little hydrophobic interaction with water-soluble proteins; therefore, they selectively interact with hydrophobic

membrane proteins.¹⁸² Thus, a method that utilizes a mild surfactant to target the membrane proteome with selectivity for integral membrane proteins²⁴⁶ would complement the Azo-enabled comprehensive proteomics strategy.

Herein, we aimed to develop a streamlined strategy for top-down membrane proteomics with a simple yet effective enrichment and separation of endogenous membrane proteins using cloud point extraction^{182, 247} from a complex cell and tissue lysates (**Scheme 1**). The cloud point extraction with nonionic surfactants, such as Triton X-114^{182, 247} or Tergitol NP-7,²⁴⁸ provided a phase separation with hydrophilic proteins found exclusively in the aqueous phase and integral membrane proteins recovered in the surfactant phase. However, because surfactants significantly suppress the MS signals,²³ the direct analysis of the enriched membrane proteins by RPLC-MS after cloud point extraction was not feasible. Therefore, to effectively remove the residual surfactant^{69, 249, 250} and simultaneously fractionate proteins based on size, we further developed a fast (<20 min) membrane protein-compatible SEC fractionation strategy (**Scheme 1**). RPLC-MS/MS analysis was used for protein identification and characterization (**Scheme 1**). Significantly, the integrated strategy enabled the enrichment and top-down proteomics analysis of large membrane proteins.

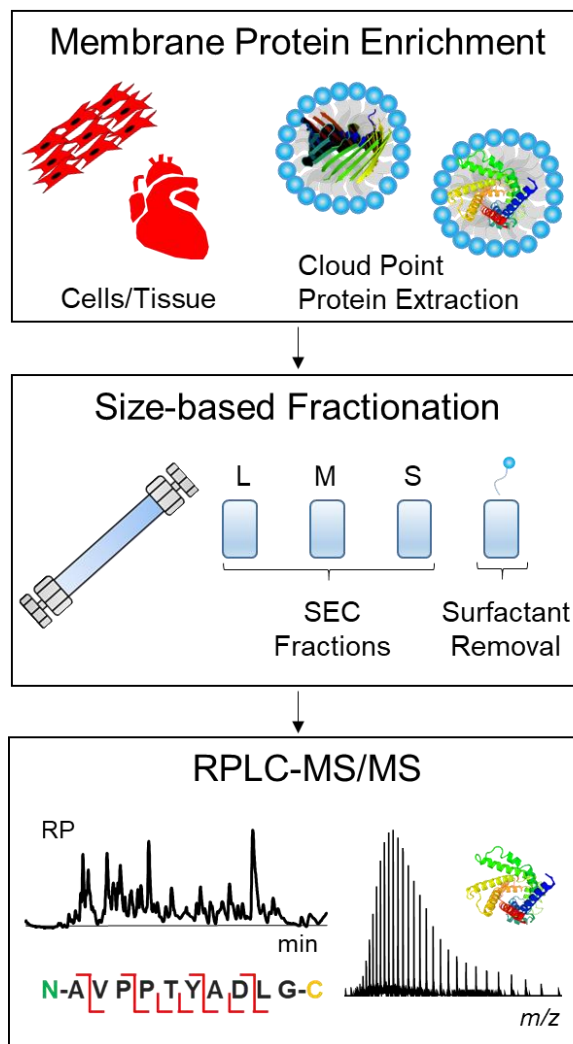


Figure 5.1. An integrated approach for top-down MS characterization of endogenous membrane proteins. A cloud point extraction procedure is performed using Triton X-114 or Tergitol NP-7 to extract and enrich endogenous membrane proteins from cells or tissues. Next, an SEC separation is performed for removal of detergent and protein fractionation. Finally, reversed-phase liquid chromatography-tandem mass spectrometry (RPLC-MS/MS) is used to obtain sequence information for protein identification.

5.3. Experimental Section

5.3.1. Chemicals and Materials

All chemicals and reagents were purchased from MilliporeSigma Inc. (St Louis, MO, USA) unless noted otherwise. HPLC grade water, isopropanol, formic acid, and acetonitrile (ACN) were

obtained from Fisher Scientific (Fair Lawn, NJ, USA). Acrylamide gels were made in-house using acrylamide/bis-acrylamide (37:5:1) solution from Hoefer (Holliston, MA, USA). Halt protease/phosphatase inhibitor cocktail, tris(2-carboxyethyl)phosphine (TCEP), bicinchoninic acid (BCA) reagent, Fetal Bovine Serum (FBS), Dulbecco's Modified Eagle's Medium (DMEM), phosphate-buffered saline (PBS), and Penicillin/Streptomycin were obtained from ThermoFisher Scientific (Waltham, MA, USA). Tergitol NP-7 was purchased from Spectrum Chemical (New Brunswick, NJ, USA). All solutions were prepared using HPLC grade water. PolyHYDROXYETHYL A columns were provided by PolyLC (Columbia, MD, USA). PLRP material was purchased from Agilent (Santa Clara, CA, USA) and was packed into capillaries of PEEK tubing from VICI (Houston, TX).

5.3.2. Cloud Point Protein Extraction

Human embryonic kidney cells (HEK) 293T cells (ATCC, Manassas, VA) were cultured on 10 cm plates (~90% confluent) using 10 % FBS and 1x Penicillin-Streptomycin at 37 °C with 5% CO₂. Plates were washed and harvested with PBS, pelleted using centrifugation at 500 × g, flash-frozen, and stored at -80 °C.

Myocardium (left and right ventricle) was obtained from non-failing hearts from brain-dead donors with no history of heart diseases but unsuitable for a heart transplant in the University of Wisconsin Hospital and Clinics and maintained in cardioplegic solution before dissection. The tissue was immediately snap-frozen in liquid nitrogen and stored at -80 °C as described previously.^{157, 187} All procedures were approved by the University of Wisconsin-Madison Institutional Review Board. Swine hearts were collected from the University of Wisconsin-

Madison meat science laboratory and flash-frozen immediately in liquid nitrogen and stored at -80 °C.

One 10-cm plate of HEK293T cells was lysed in 0.8 mL of buffer (25 mM ammonium bicarbonate, 0.5 M NaF, 10 mM methionine, 2 mM TCEP, 1 mM PMSF, and 1x Halt protease/phosphatase cocktail) using a 27 gauge needle (10x). 0.2 mL of Triton X-114 (Protein Grade) or Tergitol NP-7 was added and the solution was gently shaken at 4 °C for 20-60 min. After briefly clearing the sample using centrifugation ($15,000 \times g$, 4 °C), the supernatant was removed and incubated at 37 °C for 3 min, resulting in a cloudy bottom layer and a clear top layer. The sample was centrifuged ($3,000 \times g$, RT) for 2 min and the top layer was discarded. Similarly, ~50 mg of cardiac tissue was homogenized with a Polytron electric homogenizer, model PRO200 (Pro Scientific, Oxford, CT, USA) in 0.8 mL of buffer. Triton X-114 was added and the cloud point procedure was performed. 0.8 mL of fresh buffer was added to the bottom layer and the cloud point procedure performed again to further deplete hydrophilic proteins.

The proteins were precipitated using the chloroform:methanol:water method²⁵¹ to remove the surfactant. The protein pellet was dissolved in 50 μ L cold formic acid (80%)²⁶ supplemented with methionine (20 mM). The sample was briefly sonicated in a water bath to solubilize the pellet and diluted with 50 μ L of isopropanol and 25 μ L of water on ice. The protein concentration (~2-3 μ g/ μ L) was determined using BCA protein assay with bovine serum albumin (BSA) as a standard.

5.3.3. *Size-exclusion Chromatography*

SEC separation was performed using a 200 x 4.6 mm, 3 μ m, 1000 Å PolyHYDROXYETHYL A column on a Waters (Milford, MA, USA) Acquity H-Class HPLC

with a mobile phase of 40% isopropanol and 1% formic acid. The flow rate was set to 0.15 mL/min and the column was conditioned with two to four blank injections. 5 µg of BSA and ubiquitin (Ubi) from bovine erythrocytes were used as standards to ensure reproducibility. The elution was monitored by absorption at 280 nm. Next, 50 µL (~150 µg) of the sample was injected and four fractions were collected (8.5-11 min, 11-12 min, 12-13 min, and 13-15 min). Two collections were pooled and samples were concentrated using an Amicon 10 kDa molecular weight cutoff filter (EMD Millipore, Danvers, MA, USA). One volume of water was added and the fractions were concentrated to a final volume of ~60 µL. For SDS-PAGE analysis, 16 µL of the sample was dried via speed-vac and dissolved in 1x SDS loading buffer. The sample was loaded onto a 10% polyacrylamide gel (1 mm) and separated at 150 V for ~ 80 min. The gel was stained with Coomassie blue for visualization.

SEC separation was also performed using a 200 x 9.4 mm, 3 µm, 1000 Å PolyHYDROXYETHYL A column using a flow rate of 0.500 mL/min. 120 µL (~360 µg) of the sample was injected and eight fractions were collected using 1-minute intervals. One volume of water was added and the fractions were concentrated to a final volume of ~30 µL for MS analysis. For SDS-PAGE analysis, the sample was dried via speed-vac, dissolved in 1x SDS loading buffer, loaded onto a 10% polyacrylamide gel (1 mm), and separated at 150 V for ~ 80 min. The gel was stained with coomassie blue for visualization.

5.3.4. Reverse-phase Liquid Chromatography and Top-down Analysis

10 µL of each fraction was injected on a home-packed PLRP capillary (250 x 0.250 mm, 5 µm, 1000 Å) heated to 50 °C. The separation was performed with a mobile phase of water + 0.2%

formic acid (A) and acetonitrile: isopropanol (1:1) + 0.2% formic acid (B) using a flow rate of 5 μ L/min on a Waters nanoAcquity HPLC (M-Class). The gradient was varied to optimize separation and identify new proteoforms. The following conditions provided effective separation as demonstrated by Figure 3 and Figure S6: 0-5 min 20% B, 5-65 min 20-95% B, 65-70 min 95% B, 70-71 min 95%-20% B. Proteins eluting were infused into a maXis II ETD Q-TOF (Bruker Daltonics) via electrospray ionization with a capillary voltage of 4500 V and an endplate offset of 500 V. MS1 scans were collected at 1 Hz and the top 3 most intense ions were selected for collision-induced dissociation (CID) at 2-4 Hz (intensity 2000-20000) scan rate, and a collisional energy scaled by the m/z and charge (**Table S1**). Ions were excluded after 4 scans.

5.3.5. Data Analysis

DataAnalysis 4.3 (Bruker Daltonics) was used to analyze the raw MS spectra. For protein identification, a msalign file was created using the Sophisticated Numerical Annotation Procedure (SNAP) peak-picking algorithm with the following parameters: quality factor (0.4); signal-to-noise ratio (S/N) (3); intensity threshold (500); and a retention window (1.5 min) as described previously.²¹⁹ A compiled list of precursor mass-to-charge, precursor charge, and precursor mass followed by the fragment mass-to-charge, intensities, and charges were used to perform protein spectral matching searching against the *Homo sapiens* reviewed Uniprot²⁵² database (released on 2 July 2018) or the *Sus scrofa* Uniprot²⁵² database (released on 22 November 2017) using TopPIC (1.1.2)¹²². The fragment mass tolerance was set to 15 ppm, a satisfactory number of assigned fragment ions (>6) was required, and the false discovery rate was set to 1%. Sequence mass determination and validation and fragmentation mapping were performed using MASH Suite Pro¹⁸⁹. The proteoform mass distribution was determined using the following approach similarly

as described previously^{157, 219}: (1) LC-MS scans were averaged every minute; (2) deconvoluted using maximum entropy algorithm (resolution: 10,000; mass range: 5,000–150,000 Da); (3) mass list outputs were generated using Sum peak (S/N: 5 and absolute intensity >50,000). String analysis was performed using STRINGv11 online software using the highest confidence setting.²⁵³ A gene list was imported into Panther software (<http://www.pantherdb.org/>)²⁵⁴ for gene ontology analysis. The prediction of transmembrane domains was performed using TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>)²⁵⁵ and Uniprot²⁵². Prediction of subcellular localization was done using iLoc-Animal²⁵⁶ (<http://www.jci-bioinfo.cn/iLoc-Animal>). The proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE²⁵⁷ partner repository with the dataset identifier PXD019368.

5.4. Results and Discussion

5.4.1. *Cloud Point Enrichment of Endogenous Membrane Proteins.*

Effective extraction and enrichment of endogenous membrane proteins were accomplished with a cloud point extraction to provide a phase separation of integral membrane proteins. Here, cells or tissues were lysed in a cold (4 °C) nonionic surfactant, Triton X-114 or Tergitol NP-7,²⁴⁸ to extract and solubilize proteins. After a brief centrifugation step to remove insoluble material, the samples were brought to 37 °C, inducing phase separation with a surfactant-rich lower layer and an aqueous upper layer (see a photograph in **Figure S1**). The [temperature](#) at which the phases start to separate and the mixture becomes cloudy is the “cloud point” of a nonionic surfactant. Significantly, hydrophobic membrane proteins were sequestered in the surfactant-rich layer due to their interaction with the surfactant. Conversely, the soluble proteins remained

exclusively in the top aqueous layer and could be removed easily (**Figure 1a**).¹⁸² Overall, the cloud point extraction procedure takes as little as 30 minutes to complete using commercially available reagents and a standard benchtop centrifuge. Importantly, because Triton X-114 and Tergitol NP-7 surfactants are generally incompatible with MS analysis,^{23, 219} the bulk of the nonionic surfactant was then removed by protein precipitation.²⁵¹ The resulting protein pellet was rapidly resolubilized in cold formic acid^{26, 242} and quickly diluted with a mixture of isopropanol and water for downstream analysis. This approach is a major improvement over traditional membrane protein enrichment strategies that use multiple extractions or organelle purification that can take a day and often require specialized ultracentrifuges capable of speeds $>100,000 \times g$.²⁵⁸⁻²⁶¹

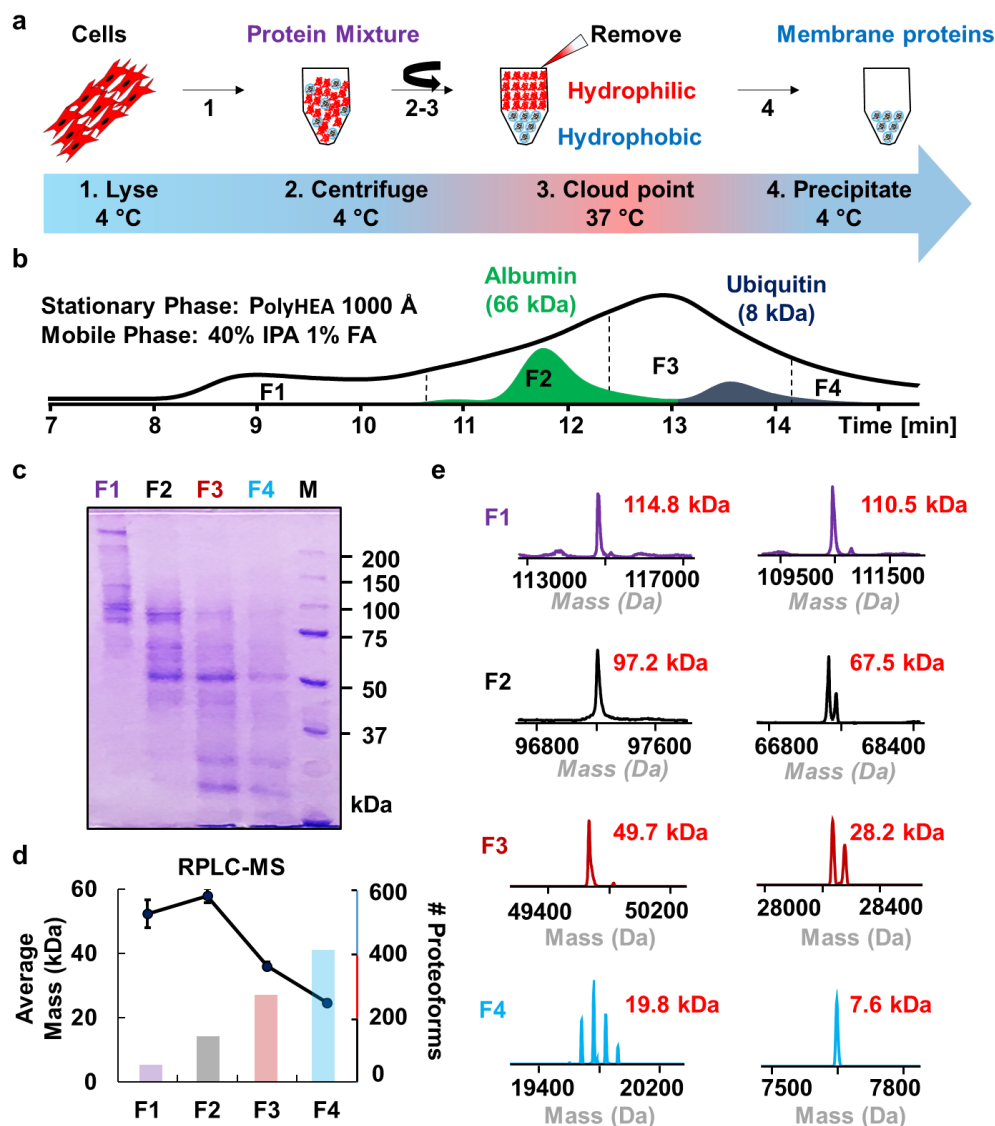


Figure 5.2. Cloud point extraction process flow and SEC fractionation of endogenous membrane proteins from HEK293T cells. (a) Illustration of the cloud point extraction for the enrichment of membrane proteins. (b) A representative chromatogram of membrane proteins separated by SEC with albumin (66 kDa) and ubiquitin (8 kDa) serving as elution reference standards. SEC was performed using a 1000Å PolyHYDROXYETHYL A (PolyHEA) column and a hydrophobic mobile phase consisting of 40% isopropanol (IPA) and 1% formic acid (FA). (c) SDS-PAGE analysis with coomassie blue visualization of SEC fractions 1-4 (F1, F2, F3, F4). (d) RPLC-MS analysis of SEC fractions. The average proteoform mass (error bars represent standard error of the mean) and the number of proteoforms detected were plotted for each fraction. (e) Representative deconvoluted (low resolution, maximum entropy)²⁶² mass spectra from each fraction. Data were collected on a Bruker maXis II QTOF mass spectrometer.

5.4.2. SEC Fractionation of Intact Membrane Proteins

After membrane protein enrichment, pre-fractionation is critical for increasing the proteome coverage. Previously, we used a single dimension of chromatography, which was insufficient for deep proteome coverage, particularly for larger proteins (>50 kDa),²¹⁹ because co-elution of small proteins causes signal suppression of large proteins. Thus, size-based fractionation is often required to access proteins >30 kDa by top-down MS.^{157, 263, 264} Furthermore, despite protein precipitation at the end of the cloud point extraction process, because the residual surfactant contamination suppressed protein signal²³ the direct analysis of the membrane-enriched sample by LC-MS was not feasible (**Figure S2**). This is conceivably why the cloud point extraction has not yet been used for top-down MS. Therefore, to increase proteoform detection (particularly high-molecular-weight species) and remove the residual surfactant, we further developed a membrane protein-compatible SEC fractionation strategy.

Previously, the advantage of MS-compatible SEC fractionation to improve the detection of intact soluble proteoforms up to 223 kDa from human heart tissue lysate has been demonstrated.^{157, 264} Here, we utilized an SEC column (4.6 mm diameter) with a neutral, polar stationary phase (PolyHYDROXYETHYL A), a 1000 Å matrix pore size, and a highly MS-compatible mobile phase, 49% water, 40% isopropanol, 1% formic acid, for SEC fractionation (**Figure 1b**). Previously, 1% formic acid was used as a mobile phase for fractionation of acid-soluble and water-soluble proteins;^{157, 264} but was insufficient for maintaining membrane protein solubility. However, we found incorporating isopropanol, a more hydrophobic solvent, facilitated membrane protein solubility in the absence of surfactants. We observed effective size-based fractionation of endogenous membrane proteins as confirmed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) visualized with coomassie blue (**Figure 1c**). Moreover, small molecule contaminations

(e.g., surfactants) were fully removed by SEC, eluting at the end of the run. SEC was highly orthogonal with regards to downstream RPLC-MS analysis.

5.4.3. SEC-RPLC for MS Analysis of a Complex Mixture of Intact Proteins.

Next, the SEC fractions were further separated by RPLC coupled with high-resolution MS. High-molecular-weight proteins were detected in the early fractions whereas low-molecular-weight species were found in the later fractions, consistent with SEC fractionation (**Figure 1d,e**). Fewer proteoforms were detected in the earlier high-molecular-weight fraction (**Figure 1d**), which can be accounted for by the decreased signal-to-noise ratio (S/N) of large proteins making them more difficult to detect.⁷⁵ Despite this challenge, we did detect several large proteoforms, including species >100 kDa, demonstrating the promise of this technology for high-molecular-weight proteoform characterization (**Figure 1e and Figure 2**). Next, we increased the column inner diameter (i.d.) to 9.4 mm to improve fractionation resolution (presumably by decreasing the wall effect).²⁴⁹ We found this method provided robust, reproducible fractionation (**Figure S3-5**). Increasing the column i.d. had the added benefit of increased sample loading capacity, thereby enabling a single 20-minute fractionation step and greater throughput, rather than pooling fractions from two runs using a 4.6 mm column. Overall, the SEC approach takes significantly less time (20 min) than GELFrEE, which typically takes 2-3 h for fractionation (based on the manufacturer's protocol) and requires further detergent removal prior to RPLC-MS.

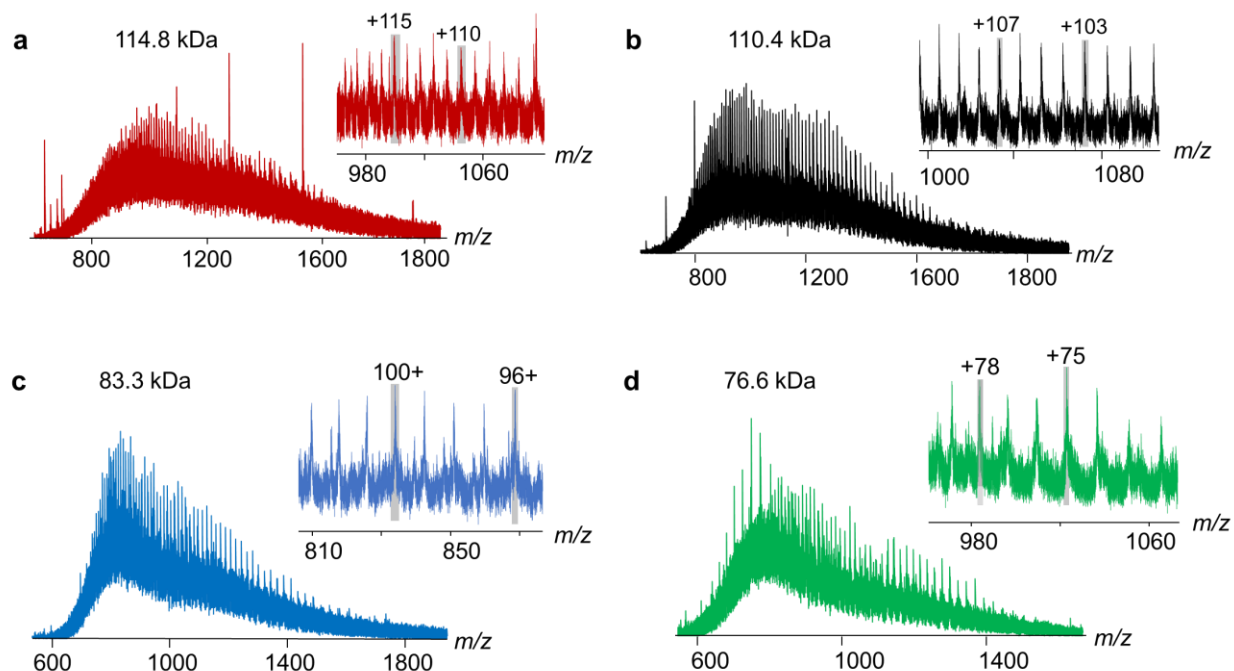


Figure 5.3. LC-MS of Large Proteins for HEK cells (a-d) Charge state distributions and deconvoluted mass spectra of representative high-molecular-weight proteins detected by RPLC-MS/MS from SEC fraction 1 using HEK293T cells. Data collected on a Bruker maXis II QTOF mass spectrometer.

5.4.4. Top-down Characterization of Endogenous Membrane Proteoforms Enabled by SEC-RPLC-MS/MS

With the establishment of a robust SEC-RPLC-MS method, we compared the performance of two common surfactants, Triton X-114 and Tergitol NP-7 (also known as Igepal), which undergo a phase separation at room temperature in water.²⁴⁸ After performing the cloud point extraction using human embryonic kidney (HEK) 293T cells, SEC fractionation (**Figure S6**), and a single RPLC-MS/MS analysis of the two low-molecular-weight fractions (3 and 4), we identified 176 protein using Triton X-114 versus 138 proteins using Tergitol NP-7 (215 total unique proteins) with a 1% false discovery rate (**Figure S7a,b**). Using UniProt²⁶⁵ to manually annotate the identifications, we found that 152 (70%) were membrane proteins and 116 (54%) were integral

membrane proteins (containing at least 1 transmembrane domain [TMD]), demonstrating effective enrichment of membrane proteins. Next, we performed gene ontology analysis²⁵⁴ based on molecular function and found many proteins involved in binding, catalytic activity, transporter activity, or structural functions, which are characteristics of membrane proteins (**Figure S7c**).

Generally, we observed that online RPLC using polymeric reversed-phase (PLRP) material^{242, 243} was compatible with separation of endogenous integral membrane proteins, permitting online characterizations of such proteins with transmembrane domains ranging from 2-19 (**Figure 3a-i and Figure S8**). Interestingly, we found that the GRAVY score (the measure of a protein's hydrophobicity based on the amino acid sequence)²⁶⁶ was a relatively good predictor of protein retention, with those with higher scores (more hydrophobic) eluting later (**Figure 3g-i and Figure S8**). This could be beneficial for future targeted studies, as GRAVY scores are easily determined.²⁶⁶ Additionally, as few integral membrane protein standards are commercially available, this platform is well suited for chromatography development towards the goal of improving membrane protein separation.

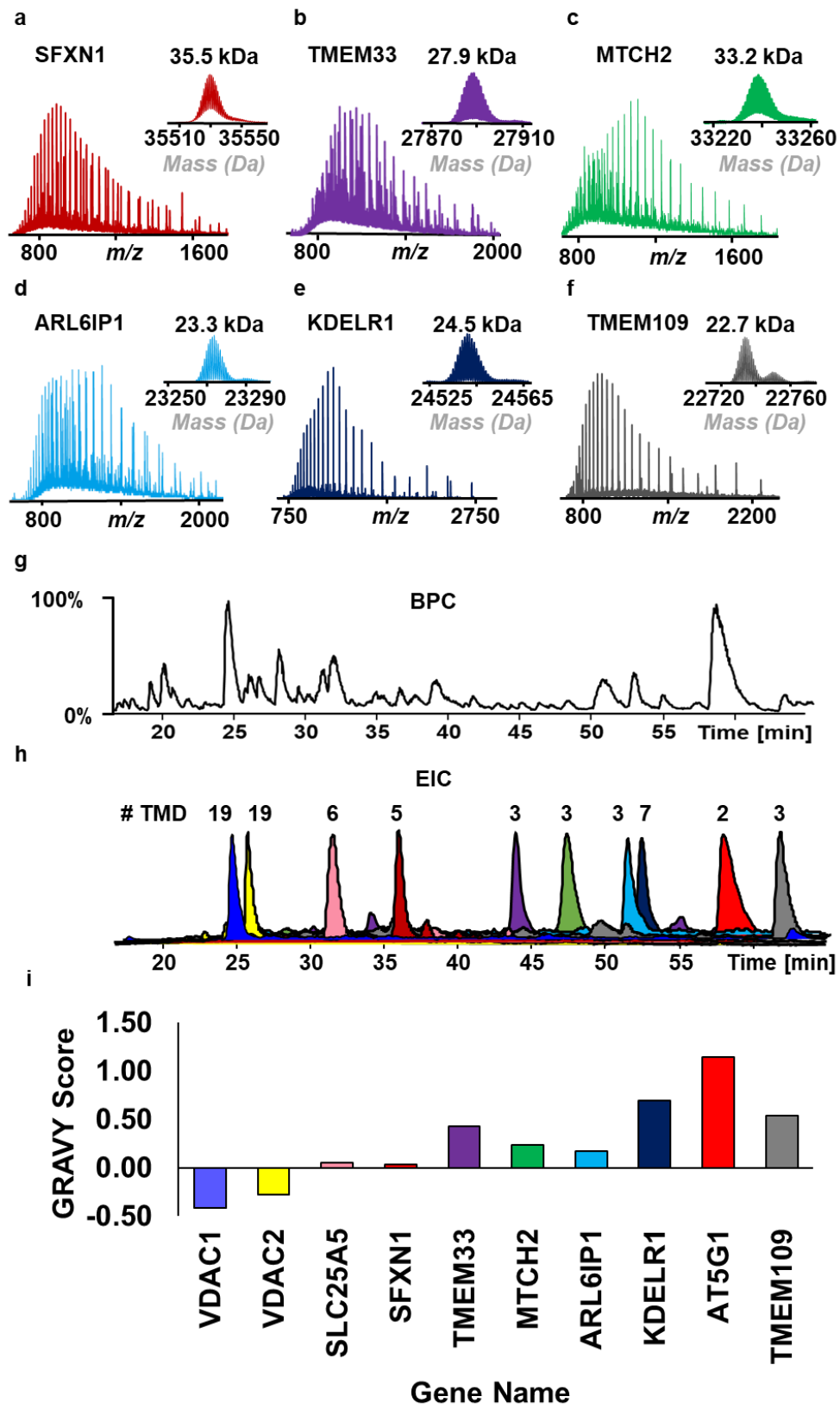


Figure 5.4. RPLC-MS analysis of enriched integral membrane proteins from HEK293T cells.

(a-f) Representative charge state distributions and deconvoluted mass spectra from SEC fraction 3. Proteins include sideroflexin-1 (SFXN1), transmembrane protein 33 (TMEM33), mitochondrial carrier homolog 2 (MTCH2), ADP ribosylation factor like GTPase 6 interacting protein 1 (ARL6IP1), ER lumen protein-retaining receptor 1 (KDELRL1), and transmembrane protein 109 (TMEM109). (g) Base peak chromatography (BPC) demonstrating effective membrane protein separation using PLRP stationary phase. (h) Extracted ion chromatograms (EIC) were generated to further demonstrate their effective separation using PLRP material. The number of transmembrane domains (#TMD) ranged from 2-19. Peak heights are not representative of protein MS signal intensity. (i) GRAVY score of representative proteins (indicated by gene name) in order of elution. Proteins with lower scores generally eluted earlier while proteins with higher scores eluted later. Data collected on a Bruker maXis II QTOF mass spectrometer.

Next, we attained good quality online MS/MS spectra using collisionally activated dissociation (CAD) allowing for confident identifications using TopPic¹²² for spectral matching and MASH Suite Pro¹⁸⁹ for validation (**Table S2-3 and ProteomeExchange^{257, 267} PXD017811**). Examples included multi-pass integral membrane proteins such as voltage-dependent anionic channel (VDAC) (Experimental M_r =30664.58 Da, 1.3 ppm error), which plays an important role in small molecule transport into and out of the mitochondria²⁶⁸ and has 19 TMD, as well as ADP/ATP translocase (Experimental M_r =32784.23 Da, 1.8 ppm error), which contains multiple sites of acetylation and has 6 TMD (**Figure 4**).

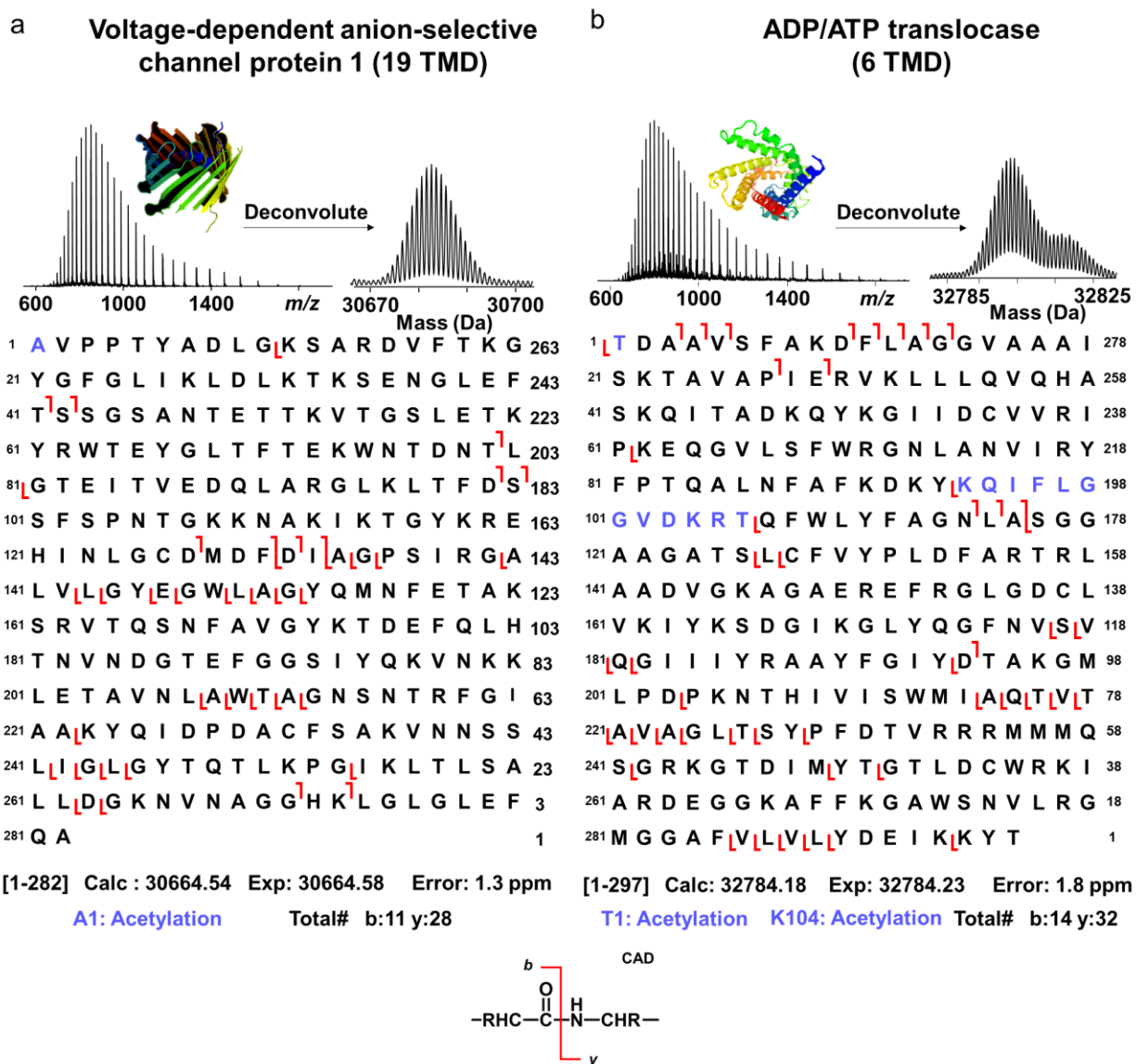


Figure 5.5. Representative online collisionally activated dissociation (CAD) of multipass integral membrane proteins. (a) Mass spectra of the voltage-dependent anion-selective channel protein 1 (VDAC) with fragmentation map. The protein was identified with 11 b ions and 28 y ions and with a mass accuracy of 1.3 ppm. Acetylation was localized to the N-terminus. (b) Mass spectra and fragmentation map of ADP/ATP translocase identified with 14 b ions and 32 y ions and with a mass accuracy of 1.8 ppm. Acetylation was localized to the N-terminus as well as residues 94-106. We hypothesize the site of modification is K104 based on the previous report.²⁶⁹ 3D protein structures were generated from the Swiss-Prot repository.²⁷⁰ Data collected on a Bruker maXis II QTOF mass spectrometer.

In addition to integral membrane proteins, we identified membrane-associated and soluble proteins. We reason that the use of a nonionic surfactant^{24, 182} in the extraction buffer preserves protein-protein interactions,²⁷¹ which can account for the presence of some hydrophilic proteins in the surfactant-rich layer. Examples include subunits of larger membrane protein complexes (**Figure 5a**) as well as proteins that are known to have biologically relevant interactions²⁷² with hydrophobic membrane proteins, such as tubulins (**Figure 5b**). Continued development of native separation^{73, 81} could permit this approach to probe protein-protein interactions and protein structure with MS.

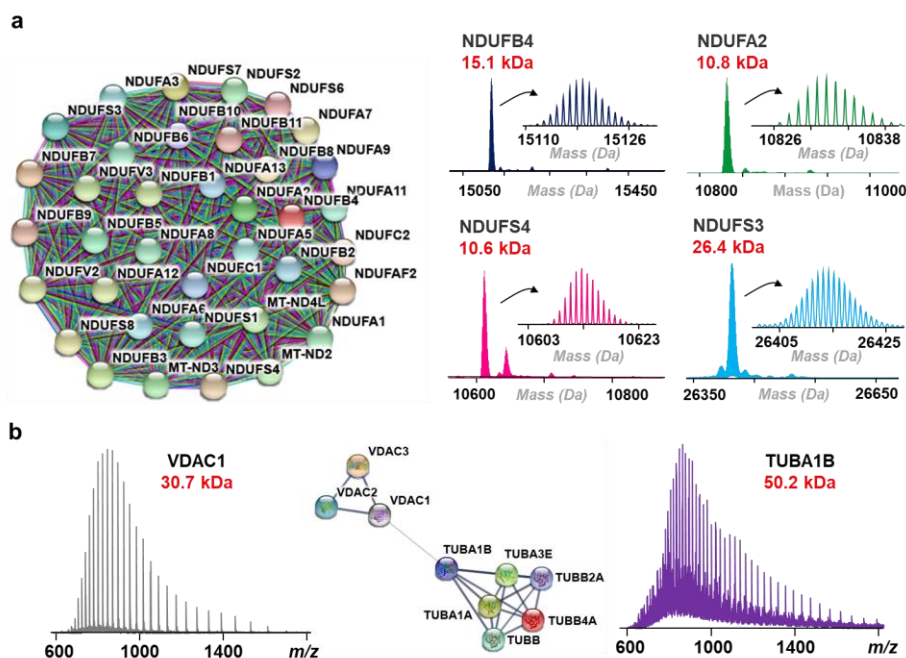


Figure 5.6. Protein interactions captured utilizing a cloud point extraction. (a) String interactome map²⁵³ of proteins belonging to the human NADH dehydrogenase [ubiquinone] complex identified from human cardiac tissue plus representative deconvoluted mass spectra of the protein complex subunits. (b) Mass spectra of voltage-dependent anion channel 1 (VDAC1) and tubulin alpha-1b (TUBA1B) along with a string interactome analysis of all voltage-dependent anion channels (VDAC) and tubulins identified from HEK293T cells. The interaction of VDAC and tubulin represents an ongoing area of neurodegeneration research and demonstrates the potential utility of this approach to probe endogenous interactions.²⁷² Data collected on a Bruker maXis II QTOF mass spectrometer.

To further demonstrate the broad applicability of this approach, we assessed our method using both HEK293T cells and cardiac tissues, identifying 1688 and 1781 proteoforms (Level 1-3 identification),²⁷³ respectively, with a 1% false discovery rate and at least 6 fragments using an error tolerance of 15 ppm (**Table S2-3**).¹²² Using TMHMM prediction software,²⁷⁴ we determined that 188 and 124 out of the set of 470 and 310 unique proteins identified were integral membrane proteins (**Figure 6**). Importantly, 108 and 58 of proteins were multipass-membrane proteins from the two unique protein sets (**Figure 6**). Analysis using iLoc-animal²⁵⁶ demonstrated a good distribution of protein subcellular locations, with many found in the cell membrane, endoplasmic reticulum, and mitochondrion and with a few proteins from the Golgi apparatus and nucleus. Future development will focus on improving the sequence coverage and the number of proteoform identifications by applying fragmentation techniques,^{19, 112} with a focus on larger membrane proteoforms which remain challenging to confidently identify despite quality mass spectra.²⁷⁵

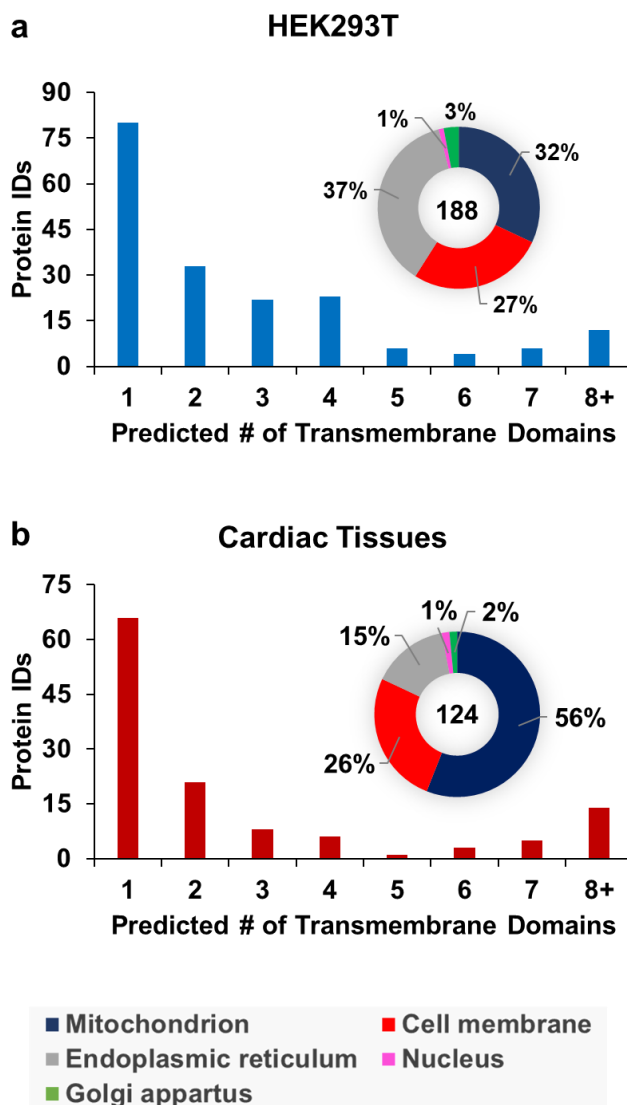


Figure 5.7. Distribution of the number of transmembrane domains²³⁷ and cellular location²⁵⁶ of integral membrane proteins identified from HEK293T cells (a) and cardiac tissues (b).

5.4.5. Strength and Weakness of the Cloud Point Extraction Approach for Top-down Membrane Proteomics.

Overall, cloud point extraction with Triton X-114 is a simple and effective method for general membrane protein enrichment. The mild nonionic surfactants (i.e. Triton X-114) show little hydrophobic interaction with water-soluble proteins but selectively interact with hydrophobic

membrane proteins.¹⁸² Moreover, Triton does not effectively solubilize nuclear,^{246, 276} contractile,⁵⁰ or extracellular matrix proteins.²⁷⁷ Therefore, the use of Triton X-114 allowed the enrichment of membrane sub-proteome with high specificity. However, a significant drawback to using Triton X-114 is the challenge of removing it prior to MS analysis and thus this cloud point extraction approach has not been used for top-down membrane proteomics previously. We observed that precipitation alone could not completely remove the surfactant even after multiple rounds. Here we have shown that SEC is essential for enabling top-down MS after cloud point enrichment, but this requirement could limit the throughput and sensitivity of the method in future applications. In contrast, the use of an MS-compatible surfactant such as Azo, which can be rapidly degraded by irradiation with UV light, enabled high-throughput proteomics.^{219, 245} Furthermore, mild nonionic surfactants do not provide comprehensive protein extraction.^{219, 246} Hence, a more comprehensive proteome extraction can be accomplished by using a strong ionic surfactant (such as Azo)²¹⁹ after treatment with Triton X-114 (**Figure S9**). Thus, in some sense, the cloud point extraction approach is complementary to our previously established Azo-enabled top-down proteomics. Combining the cloud point extraction approach with Azo enabled strategies could show even more promise for global proteomic studies by incorporating serial extractions to achieve deep proteome coverage.

5.5. Conclusion

In summary, we have developed a simple but effective enrichment and separation strategy for top-down proteomics of endogenous membrane proteins from complex cell and tissue lysates. The cloud point extraction enables effective membrane protein enrichment with a single extraction, including many integral membrane proteins containing as many as 19 transmembrane domains.

The subsequent SEC fractionation removes the surfactants from the extraction step and permits the fractionation and detection of proteoforms with molecular weights ranging from 6-115 kDa. Finally, we have demonstrated the effectiveness of online RPLC-MS/MS for membrane protein separation and characterization. We anticipate this simple approach can be widely utilized for membrane protein characterization as an integral part of a generalizable toolset for the development of top-down proteomics methods for endogenous membrane proteins.

5.6. Supplementary Information

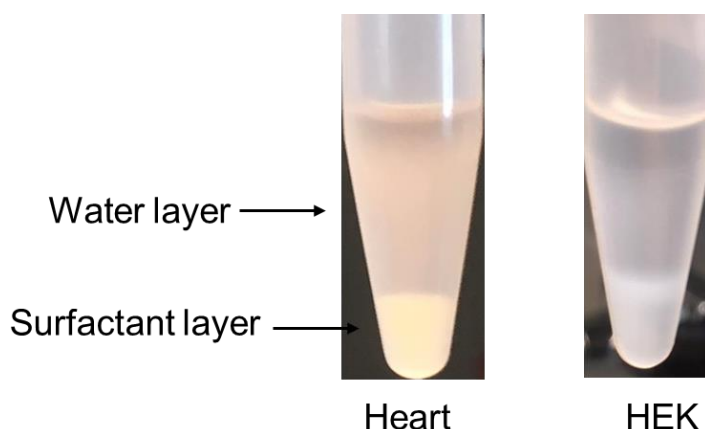


Figure 5.8. Visualization of the cloud point extraction using human heart and human embryonic kidney cell (HEK) lysates.

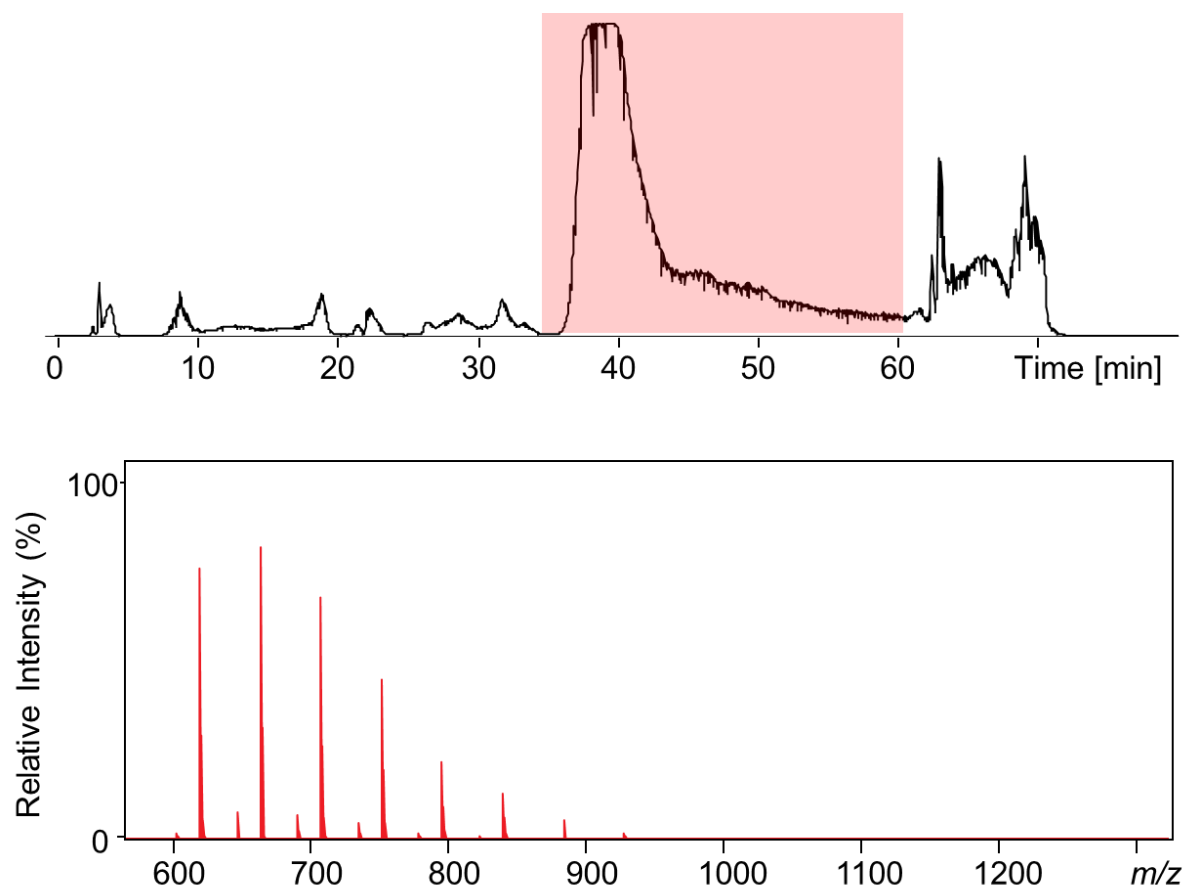


Figure 5.9. Direct LC-MS analysis of proteins from membrane enriched-fraction. Despite the use of protein precipitation (chloroform:methanol:water), a large surfactant contamination peak is observed.

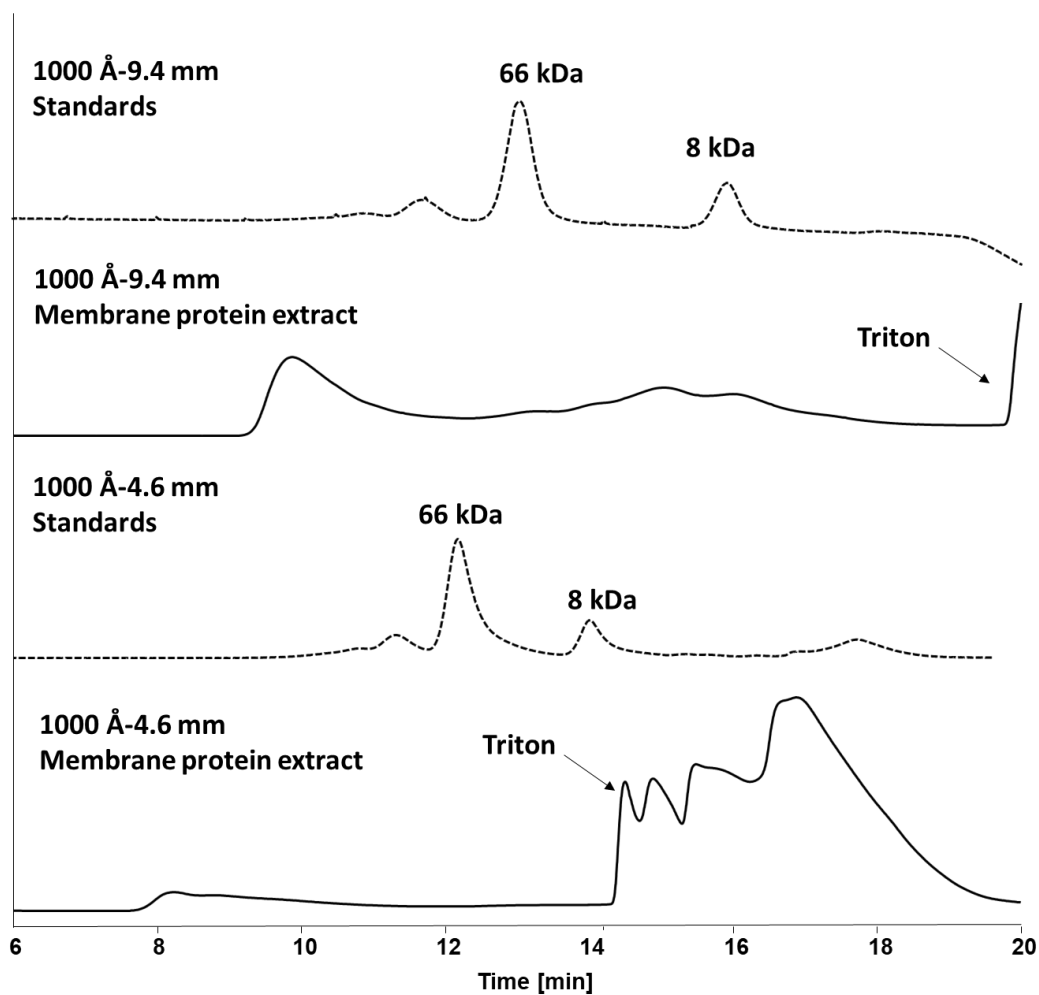


Figure 5.10. SEC chromatogram comparing membrane proteins separated with a 4.6 and 9.4 mm column. Albumin (66 kDa) and ubiquitin (8 kDa) were used as standards. The 9.4 mm column afforded greater separation between the residual Triton X-114 contamination and the low-molecular-weight proteins.

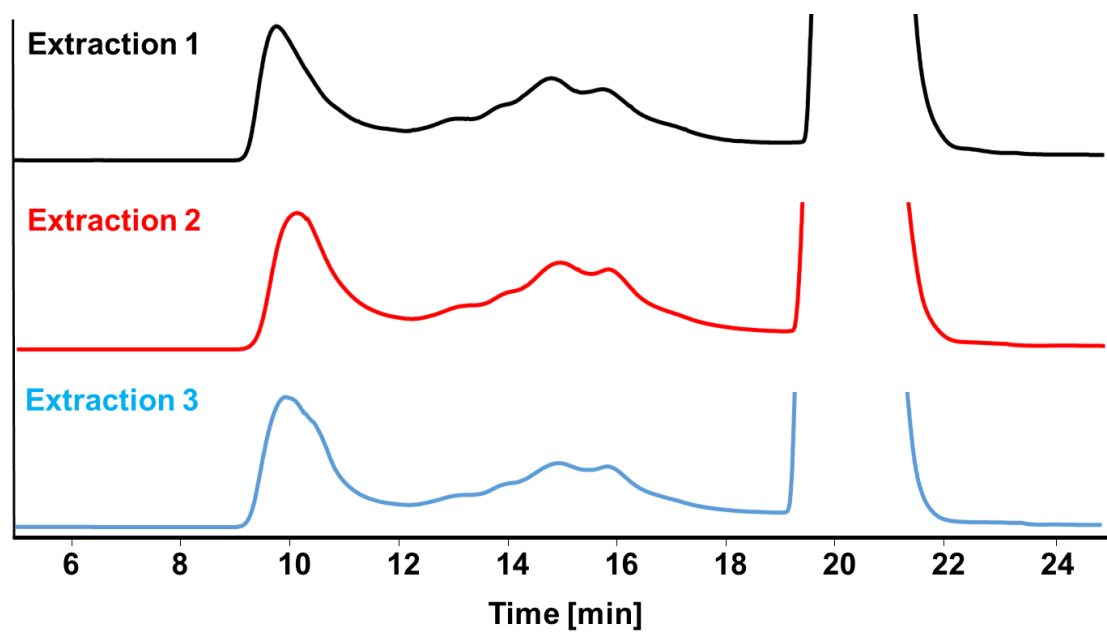


Figure 5.11. SEC chromatogram demonstrating the reproducibility of separation using three extraction replicates.

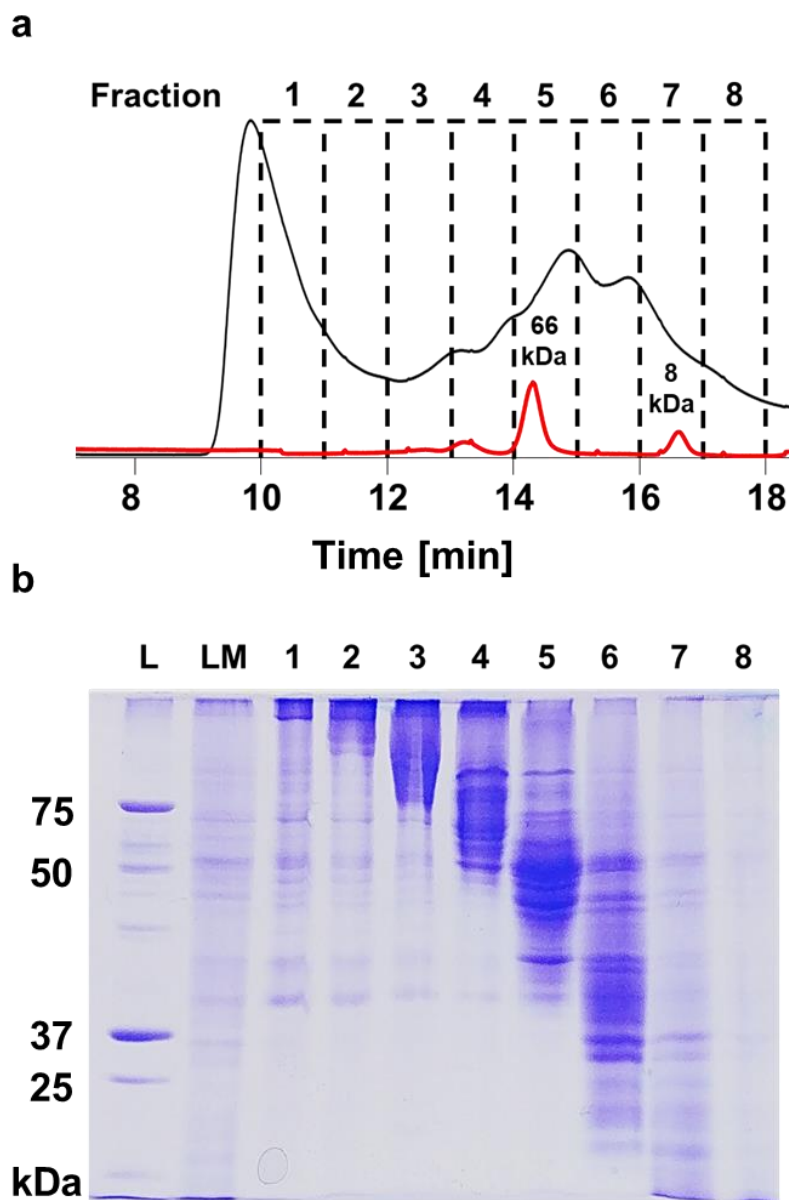


Figure 5.12. Effective high-throughput SEC separation. (a) SEC chromatogram of membrane proteins extracted with Triton X-114 and separated with a 9.4 mm column. Albumin and ubiquitin were used as a reference [*red trace*]. (b) SDS-PAGE analysis of SEC fractions, demonstrating good protein separation.

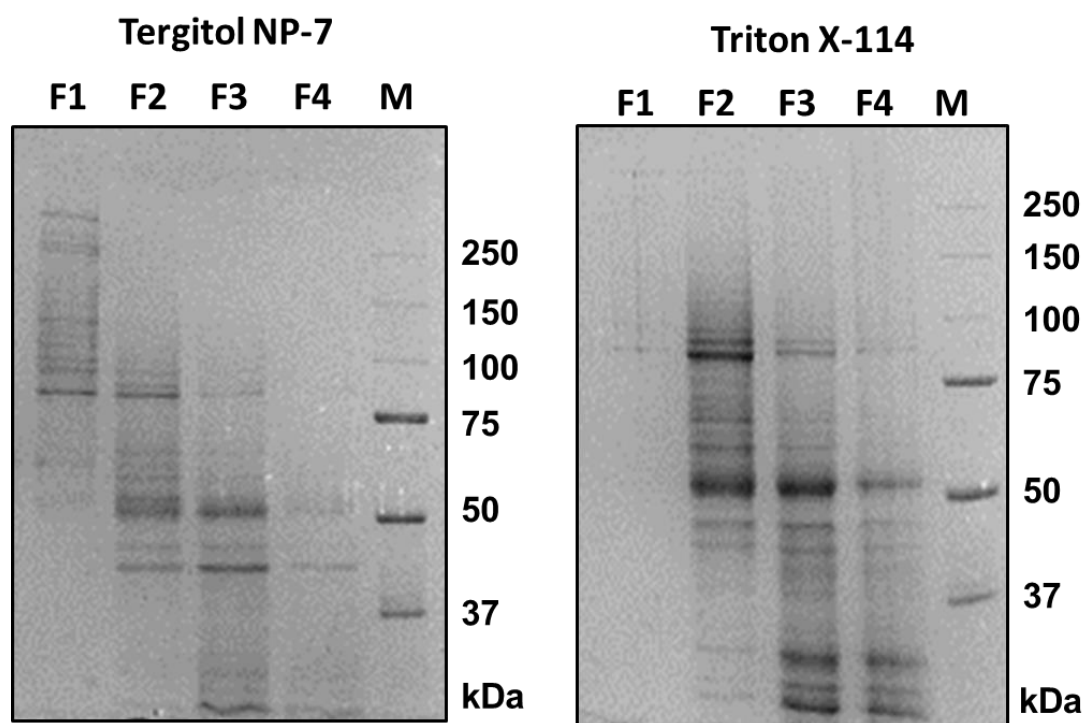


Figure 5.13. SDS-PAGE analysis of SEC fractions collected from a 4.6 mm column after extraction using Tergitol NP-7 or Triton X-114.

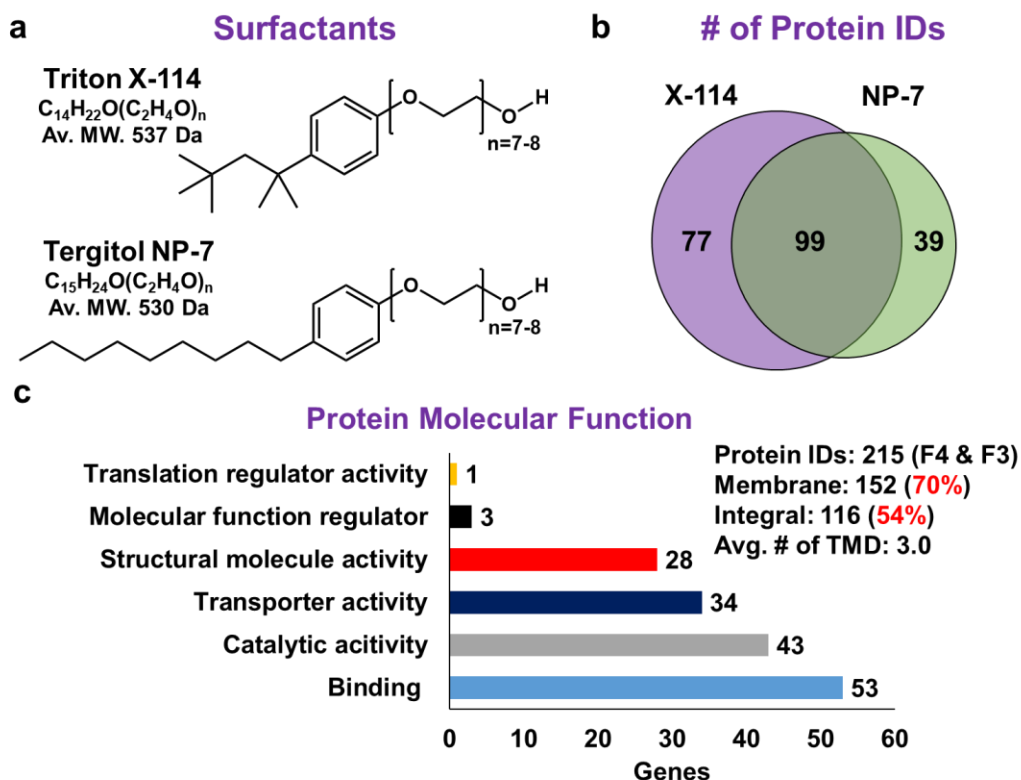


Figure 5.14. Comparison of Triton X-114 and Tergitol NP-7 for membrane protein enrichment. (a) Chemical structure, molecular formula, and average molecular weight (Av. MW.) for Triton X-114 and Tergitol NP-7, which undergo phase separation in water at room temperature (20–25 °C).²⁴⁸ (b) Comparison of proteins identified by SEC-RPLC-MS/MS. (c) Gene ontology analysis²⁵⁴ of protein identifications categorized based on molecular function. Many of the proteins identified perform binding, catalysis, transport, and structural roles, which are some of the characteristic functions of membrane proteins. Data represents a single LC-MS/MS analysis for each fraction. Data collected on a Bruker maXis II QTOF mass spectrometer.

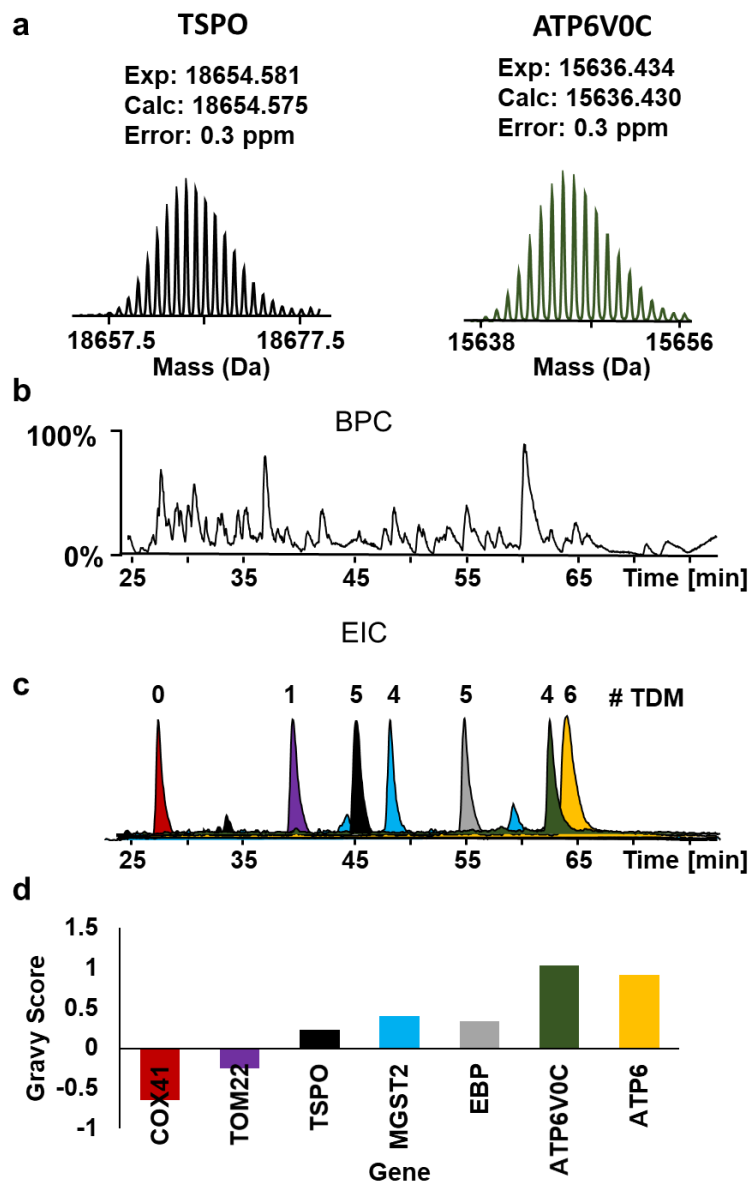


Figure 5.15. RPLC-MS analysis of SEC fraction 4 (Triton X-114) from HEK293T cells. (a) Representative deconvoluted mass spectra of Translocator protein (TSPO) and V-type proton ATPase 16 kDa proteolipid subunit (ATP6V0C), demonstrating good mass accuracy. (b) Base peak ion chromatogram (BPC) and (c) extracted ion chromatograms (EIC) demonstrating effective membrane protein separation. The peak heights for the EIC are not representative of MS signal intensity. (d) GRAVY score of representative proteins with respect to the order of elution. Data collected on a Bruker maXis II QTOF mass spectrometer.

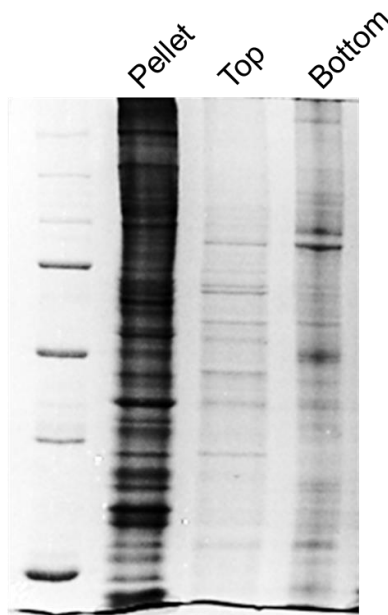


Figure 5.16. Azo for extraction of insoluble proteins. Cloud point extraction with Triton X-114 was performed using heart tissue. After centrifugation, the supernatant was removed and heated forming a top water layer and a bottom surfactant layer. Azo (0.5%) was added to the insoluble pellet to extract the remaining proteins. An equal volume of extract was separated by SDS-PAGE and visualized with coomassie blue. The results demonstrate that a significant amount of protein remains after Triton X-114 treatment.

Table 5.1. Collisional energy for protein fragmentation

m/z	Width	Collision	Charge State
500	5	35	1
500	5	20	8
500	5	18	15
1000	5	24	1
1000	5	22	8
1000	5	20	15
1300	8	32	1
1300	8	30	8
1300	8	28	15
2000	8	45	1
2000	8	43	8
2000	8	40	15

Table S2. Summary of HEK293T proteoforms identified using RPLC-MS/MS

– See excel Table S2

Table S3. Summary of cardiac proteoforms identified using RPLC-MS/MS

-See excel Table S3

Chapter 6. Conclusions and Future Directions

In summary, the projects described in **Chapters 2-4** provide new methods for mass spectrometry (MS)-based proteomic analysis of challenging protein groups such as membrane and extracellular matrix (ECM), making their characterization more accessible and higher throughput. Using our MS-compatible photocleavable surfactant, Azo, insoluble proteins were easily extracted from important biological samples, such as heart tissues, at high-concentrations and the UV-cleavable linker enabled facile surfactant removal before MS analysis. We believe there are many future applications that could utilize the Azo technology for both targeted and global top-down proteomics. For example, current efforts include in-depth characterization of proteins such as phospholamban to determine their role in regulation in cardiac diseases.³¹ Azo's ability to aid in rapid enzymatic digestion of proteins for high-throughput bottom-up proteomics analysis has great potential for clinical diagnosis where limited sample preparation steps and speed is critical. Moreover, the ECM proteomics method described in **Chapter 4** has great potential for elucidating the role of ECM proteins in cardiac regeneration and cancer metastases.²⁰¹ Future work will adapt the ECM proteomics for top-down proteomics to fully characterize ECM proteoforms. Likely, this will require size-based separation techniques such as size-exclusion chromatography (SEC), as ECM proteins are relatively large in molecular weight.

The implementation of cloud point extraction in **Chapter 5** greatly facilitated enrichment (up to 70%) of membrane proteins. The development of a membrane protein compatible SEC method provided rapid protein separation (~20 min) and surfactant removal before RPLC-MS/MS enabling detection of protein as large as 115 kDa, which shows great promise for expanding the scope of top-down membrane proteomics. Future development of online SEC coupled to reversed-phase liquid chromatography (RPLC) will improve throughput and widespread adoption of this method. Additionally, separation modes like hydrophilic interaction chromatography (HILIC)⁶⁷,

^{68, 82, 244} and recent developments in capillary zone electrophoresis (CZE)⁷⁸ show great promise as complementary or alternative separation modes in the future. Furthermore, combining the cloud point extraction with a subsequent Azo extraction appears to be a highly effective “1-2 punch”, facilitating global protein extraction as well as protein fractionation at the extraction level.

While the cloud point enrichment provides broad membrane protein enrichment, extremely low abundant membrane proteins such as certain transmembrane receptors or ion channels remain elusive. Thus, the development of more targeted approaches is required for MS-based characterization. Polyacrylamide-gel-based prefractionation (PEPPI)-MS⁷⁷ represents an intriguing technique to isolate proteins directly from SDS-PAGE gels, but it remains to be demonstrated for effective analysis of high-molecular-weight proteins. Conceivably, Azo could replace SDS for protein separation thereby could alleviate the potential interference from SDS that remains even after sample cleanup thereby improving MS analysis of larger proteoforms.

In the past 5 years, significant advances have greatly expanded the scope and throughput of top-down membrane proteomics. We optimistically envision continued development in sample preparation, protein separation, and instrumentation will likely improve our ability to globally characterize membrane proteoforms and expand the implementation top-down analysis, especially when integrated with other omics (genomics, metabolomics, etc.) and structural biology techniques (cryo-EM, NMR, X-ray crystallography).

Chapter 7. Developing New Soaps to Access

Hard-to-Study Proteins

Kyle Brown, Ge Group

7.1 Introduction

In many cases, it can be challenging to provide enough detail and context in a scientific publication to explain the work to non-experts while maintaining conciseness (otherwise every publication would be a book). Moreover, small advancements in a particular area of science, while important, may only be of interest to those in that field. Therefore, it is practical for us to tailor our publication, to a certain degree, to our expertise. On the other hand, we must engage the broader scientific and non-scientific communities to explain the broader importance of our work. Fostering widespread interest and faith in science is vital for its future success. To this end, I have written a chapter of my thesis explaining my research to an audience of non-experts to make it more accessible than a traditional scientific publication. I hope that my family, friends, and anyone curious in understanding more about science will be able to learn something about what I have accomplished over the course of my PhD. I'd like to thank the Wisconsin Initiative for Science Literacy at the University of Wisconsin-Madison for supporting this chapter and for their important work of promoting the communication of science outside the academic setting.

I have always been interested in understanding how the world around me works. In particular, I am fascinated by the complexity we observe in biology. There is some much going on in our bodies that we cannot begin to explain or understand. It can be daunting to know where to

begin when trying to explain why people have certain traits or get sick or any number of interesting questions regarding our nature. For this reason, I decided to get PhD focusing on life science.

The simplest explanation of our approach to better understanding human biology is we try to study the most fundamental components of humans and work our way backward. We can remember from our early biology classes that humans are composed of various organs, which are made up of cells (the basic unit of all life). An important part of our cells is DNA, the genetic blueprint of an individual. The role of DNA is to code for RNA that makes protein out of chains of amino acids called peptides. Each protein has a specific function, or set of functions, within cells that keeps us alive and allows us to respond to our environment. As the building blocks of cells, proteins act as the molecular machines that carry out many functions. They represent an important starting point towards understanding cells, organs, and human biology functions as a whole system.

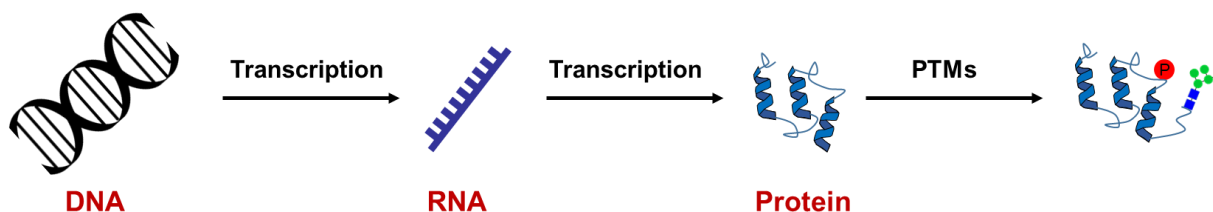


Figure 7.1. Scheme illustrating the process of DNA transcription into RNA, which is then translated to make proteins. These important biomolecules perform many vital functions in cells and post-translational modification (PTMs) of proteins can change their functions giving rise to much of the biological complexity we observe in cells.

What is not as commonly known is that proteins can modify by with specific chemical groups, which can dramatically change their function (**Figure 7.1**). The chemical groups that are added to proteins are called post-translational modification (PTMs). A single gene, therefore, can produce many different protein forms, called proteoforms, giving us a lot of our biological

complexity. We often think of PTMs as the body's barcode system, where a protein is produced and then labeled with different codes resulting in them going to different locations in the cell or performing different roles.

The study of proteins, referred to as proteomics, seeks to understand the function of proteins in our cells and determine how they are regulated by PTMs and other biological events. In particular, we tend to study proteins by comparing two conditions, like health versus disease. Identifying changes from the normal state, for instance, lower abundance of a protein or the presence/absence of modification enables us to identify markers of disease progression and potential targets for new therapeutics.

A common tool used to study proteins is mass spectrometry, which can be thought of as a “molecular-scale”. Mass spectrometry allows us to identify proteins in a sample, determine the abundance, and evaluate their modified all by simply measuring their mass (measured in Daltons [Da] which is equivalent to $\sim 10^{-27}$ pounds). To study proteins, we first have to extract them from the sample of interest. Then we separate the proteins so that we can analyze them individually. The proteins are then ionized by a process called electrospray ionization (ESI). In ESI, a protein solution is guided through a narrow tip and a high voltage (1-4 kV) is applied to produce small, highly charged droplets. The droplets progressively evaporate (by adding heat) until the protein(s) break free of the droplets as charged gas ions. Finally, the ions are analyzed using a mass spectrometer to determine their mass (**Figure 7.2**).

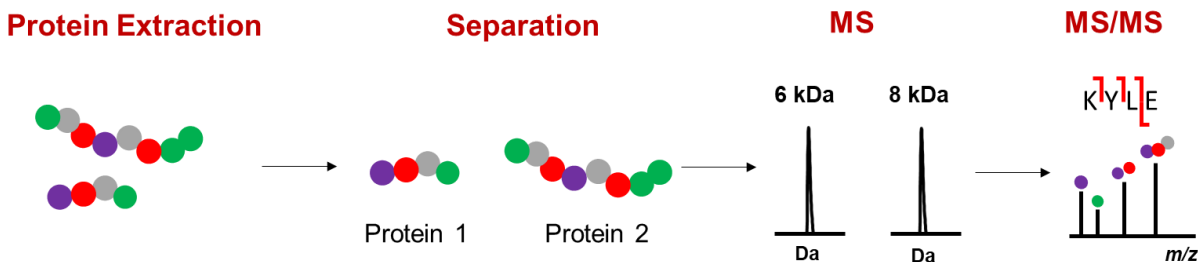


Figure 7.2. Proteins sequencing by mass spectrometry. Proteins are extracted from cells, separated, weighed using mass spectrometry. Finally, proteins are fragmented in the mass spectrometer the pieces weighed to gain sequence information for more confident protein identification.

As we discussed earlier, proteins are made up of a unique sequence of amino acids and therefore have a unique mass. The theoretical mass of all proteins is known since their amino acid sequences are determined by DNA, and we have sequenced the entire human genome. Unfortunately, modifications, mutation, and similarities in protein sequences can result in proteins having almost identical masses or at least result in their masses differing from the predicted. Therefore, we sequence proteins, determine their amino acid composition, by fragmenting them in the mass spectrometer (**Figure 7.2**). Protein fragmentation (also referred to as tandem MS or MS/MS) is performed by colliding the protein ions with inert gas molecules (e.g., nitrogen) and measuring the mass of the resulting pieces to put together the puzzle so to speak. Importantly, if the mass of the protein, or the mass of its fragments, is different than the expected value, we can match it to a modification. For example, if a protein mass is 80 Da more than the predicted value then we know it has been modified in a phosphate group. Examples of common protein modifications and their masses can be seen in **Figure 7.3**. In summary, mass spectrometry allows us to identify, sequence, and determine the modification of proteins.

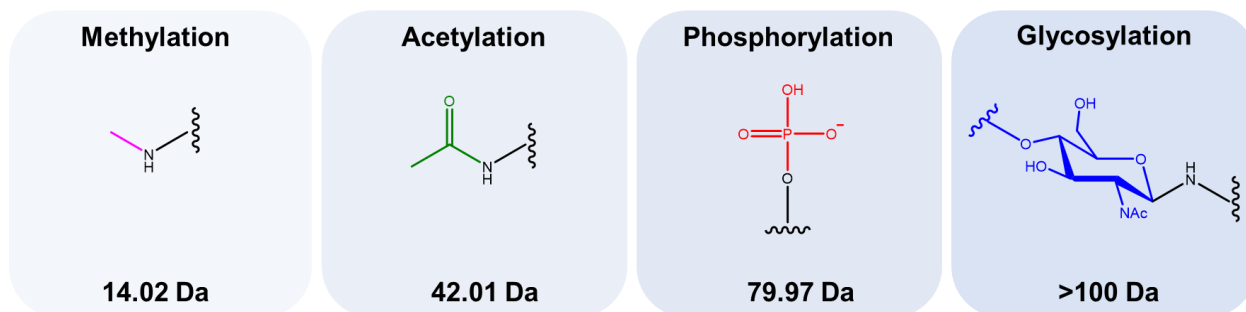


Figure 7.3. Examples of common protein modifications and the corresponding masses.

7.2 What are the challenges?

Studying proteins gives us a window into what is happening to the cells at a very fundamental level; however, there are many challenges. For example, there are roughly 20,000 different proteins that can be mutated, truncated, and modified by many chemical groups resulting in a huge number of possible protein forms. The result of this protein level diversity is a huge number of possibilities and differences even between individuals. With so many proteins, we can't study all of them at once; so we have to separate them away from each other to study them a few at a time. However, there is a wide range of proteins with different properties (size, charge, hydrophobicity, etc.), so one method cannot effectively separate them all. This is further complicated by the fact that some proteins are produced in very high numbers and others in much lower. However, lower abundance does not mean they are less important.

One common way to handle this complexity is to digesting (chopping up) proteins into peptides. Digestion makes the mixture more homogeneous, as the resulting peptides will have similar sizes and charges. The smaller size of peptides compared to proteins makes it easier to separate and measure by mass spectrometry. This approach, called bottom-up proteomics, allows

us to study thousands of proteins from a sample and is particularly useful for identifying changes in protein abundance.

However, there are significant drawbacks to studying peptides rather than intact proteins. When proteins are digested you can't always put the sequence back in the correct order to get the true state of the proteins. The confusion regarding the true protein state that results from digestion is best illustrated in **Figure 7.4**. As you can see, even if we identify all the peptides associated with the protein, we lose information about how the protein was modified. The exact state of a protein can be critical for understanding its function and role in disease; therefore, we want to precisely know its sequence and modification state.

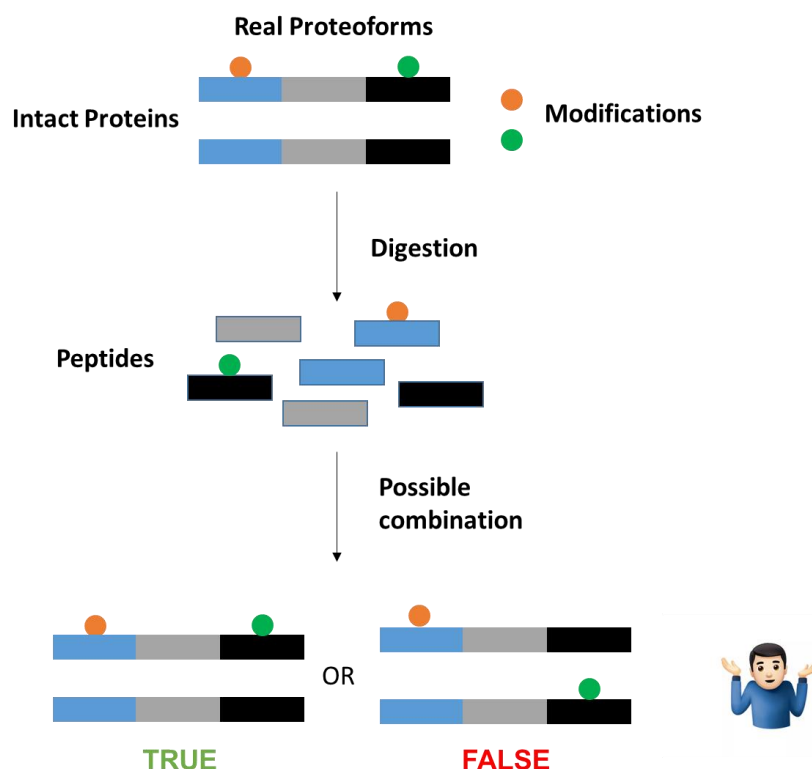


Figure 7.4. Diagram illustrates the challenge that arises from digesting proteins into peptides. Although we can identify all the peptides, we don't know whether the proteins exist as a doubly modified species, two single modified, or some combination.

7.3 What is my research about?

At this point, we've established that it is important to study proteins and that there is a great benefit to looking at them intact, but this makes them more difficult to separate and analyze by mass spectrometry. My research focuses on developing methods to better extract, separate, and sequence intact proteins to gain a better understanding of their role in diseases. More specifically, I have focused on an important class of protein located in the cell membrane. Membrane proteins account for about 50% of drug targets and one-third of our genes code for them. However, working with membrane proteins presents unique challenges as they are very hydrophobic (insoluble in water like oil) and there are generally fewer copies per cell relative to other types of proteins.

7.4 What is our approach?

To study membrane proteins with mass spectrometry, we must first extract them into a water solution from cell or tissue samples. Just like in everyday life, we use soaps (also called detergents or surfactants) to remove difficult or insoluble material. This is because soaps are amphiphilic meaning they have hydrophobic (water-hating) tail groups and a hydrophilic (water-loving) head group (**Figure 7.5**). Hydrophobic proteins, like those found in membranes, are generally insoluble in water, but can readily be dissolved with soap. As seen in **Figure 7.5b**, soap naturally forms micelles that keep the hydrophobic portions of proteins away from water thereby allowing them to be soluble. Not only do soaps help solubilize membrane proteins, but they also break down cells helping release all the proteins within.

Figure 7.6. Photocleavable surfactant aid in breaking down the cell to extract and solubilizes proteins. Afterward, it can be easily degraded with UV light allowing for protein characterization by mass spectrometry.

With an improved method for extracting proteins, I next developed different ways to separate proteins. Protein separation is generally accomplished using high-performance liquid chromatography (HPLC). A protein solution is introduced to a solid stationary phase that can be made of a variety of materials, but most commonly consists of silica particles functionalized with carbon chains for a technique called reversed-phase liquid chromatography (**Figure 7.7**). The proteins interact with the stationary phase with varying degrees of strength. A mobile phase is then used to elute (wash away) proteins off the stationary phase. Thus, proteins that have the weakest interaction come off first followed by the proteins with stronger interaction come off last. An example we commonly use to demonstrate this is putting dye on a coffee filter then introducing it to water. You can then see the dye start to separate into its different components based on differing interactions with the paper and the water (**Figure 7.7**).

Another common HPLC technique to separate proteins is called size-exclusion chromatography (SEC). Proteins are introduced to a porous, inert stationary phase. Bigger proteins will not be able to enter the pores and therefore elute earlier while smaller proteins will enter the pores and elute later. SEC allows us to sort protein based on their size making it easier to study more proteins from a sample. My research combines multiple different separations techniques, like SEC and reversed-phase, and optimizing them for membrane protein to study hundreds of proteins from a sample after using the photocleavable surfactant to extract them.

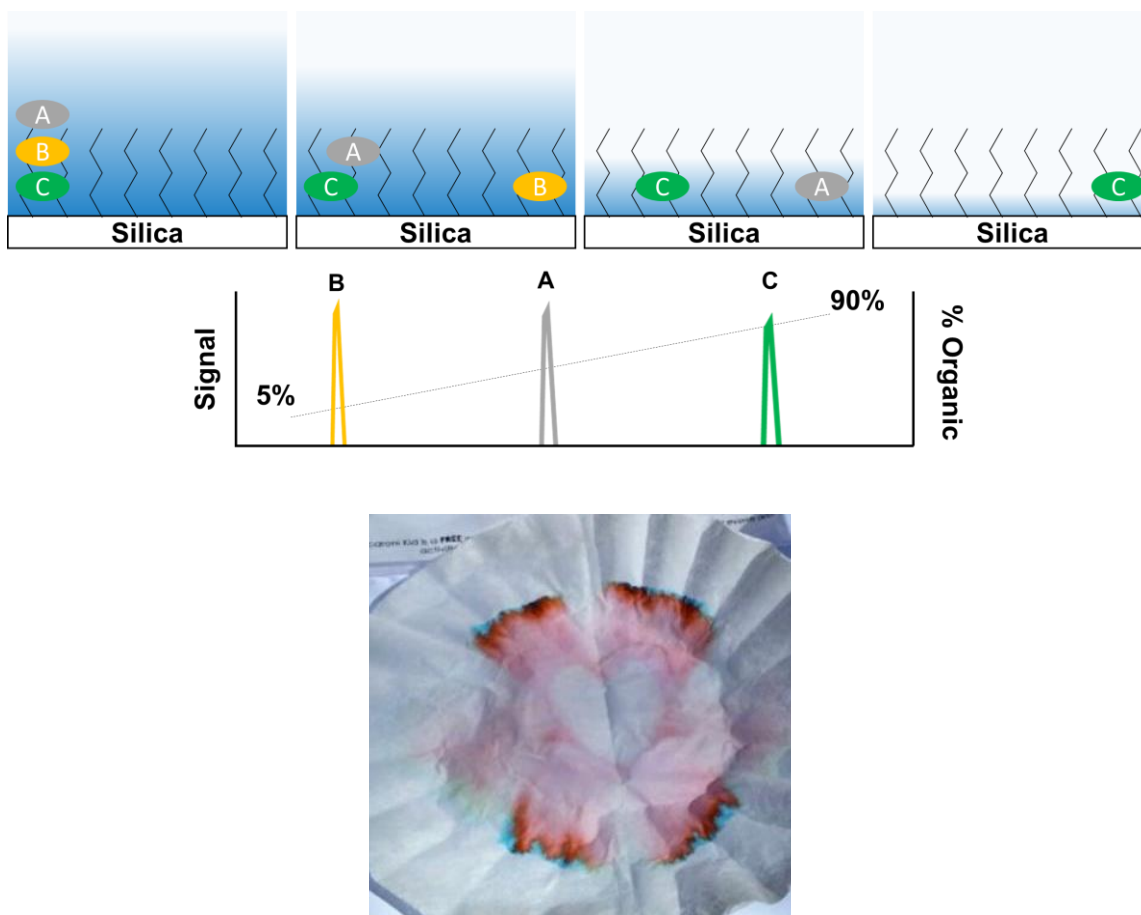


Figure 7.7. Illustration of reversed-phase liquid chromatography separation of proteins (top). Example of separating dyes using a coffee filter. This is a common demonstration of how chromatography works. Image credit: Michelle Melnik. Paper Chromatography Coffee Filters. <https://deerfield.macaronikid.com/articles/5826504224d8f4da3aafb36e/paper-chromatography-coffee-filters>. (Accessed May 28, 2020).

7.5 How does this help disease research?

Our lab focus on studying proteins involved in cardiac diseases. We want to understand how proteins change as cardiac disease progress or after a major cardiac event (e.g., a heart attack). A better understanding of protein roles in cardiac disease will allow us to identify the important proteins that cause the dysfunction or the proteins that are most affected by the disease resulting in the long-term health consequence.

The technology I developed helps extract and separate from heart tissues. This can be from animals where we have simulated a specific heart disease or even from donated patient samples. We then look at how proteins are changing before and after the disease (**Figure 7.8**). This knowledge allows us to determine what role certain proteins play in causing disease and target them to detect the disease early on or develop better treatments. Because there is variability from person to person, being able to sequence an individual's proteome could allow us, in the future, to design person-specific treatment plans that enable precision medicine. Since proteins are the building blocks and control so much of the molecular functions, they are the truest representation of what is going on and therefore could be the best predictors.



Figure 7.8. Illustration of a protein modification change that can be a factor contributing to a disease or result of a disease.

7.6 What is the next step?

The technology I developed in my PhD research for extracting and separating membrane proteins can be applied to any disease to help discover how they are being altered. Since membrane proteins are common drug targets, it is critical to be able to determine their role in diseases. Although I was able to improve the separation of intact proteins enabling the identification of hundreds from a given sample, there is still significant work towards the ultimate goal of being

able to sequence the entire human proteome. I anticipate further development in protein separation and mass spectrometry analysis will significantly improve our ability to characterize protein changes.

Moving forward, I hope to continue to do quality research and communicate it to non-experts. For example, I believe working with journalists to publish summaries of important research (e.g., our work with the photocleavable surfactant <https://phys.org/news/2019-04-approach-ready-access-hard-to-study-proteins.html>), with a non-expert target audience, is crucial for engaging broader communities. Additionally, I believe outreach programs like science fairs are a great way to disseminate information especially to younger students, and I look forward to continuing to participate in them throughout my career.

During my time in graduate school, I have had a great experience learning about proteins, diseases, and much more. I am very excited to move on to a postdoc position where I will continue to study proteins and their role in diseases. I am very thankful for the mentorship I have received and the friends I have made!

References

1. Smith, L. M.; Kelleher, N. L., Proteoforms as the next proteomics currency. *Science* **2018**, 359 (6380), 1106-1107.
2. Uhlén, M.; Fagerberg, L.; Hallström, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigartyo, C. A.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P. H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; von Heijne, G.; Nielsen, J.; Pontén, F., Proteomics. Tissue-based map of the human proteome. *Science* **2015**, 347 (6220), 1260419.
3. Smith, L. M.; Kelleher, N. L.; The Consortium for Top Down, P.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla*, L.; Kruppa, G.; Zubarev, R.; Rontree, J.; Chamot-Rooke, J.; Garavelli, J.; Heck, A.; Loo, J.; Penque, D.; Hornshaw, M.; Hendrickson, C.; Pasa-Tolic, L.; Borchers, C.; Chan, D.; Young*, N.; Agar, J.; Masselon, C.; Gross*, M.; McLafferty, F.; Tsybin, Y.; Ge, Y.; Sanders*, I.; Langridge, J.; Whitelegge*, J.; Marshall, A., Proteoform: a single term describing protein complexity. *Nature Methods* **2013**, 10, 186.
4. Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schluter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlen, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschlagel, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B., How many human proteoforms are there? *Nat Chem Biol* **2018**, 14 (3), 206-214.
5. Gregorich, Z. R.; Ge, Y., Top-down proteomics in health and disease: Challenges and opportunities. *Proteomics* **2014**, 14 (10), 1195-1210.
6. Cai, W. X.; Tucholski, T. M.; Gregorich, Z. R.; Ge, Y., Top-down Proteomics: Technology Advancements and Applications to Heart Diseases. *Expert Review of Proteomics* **2016**, 13 (8), 717-730.
7. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003**, 422 (6928), 198-207.
8. Yates, J. R.; Ruse, C. I.; Nakorchevsky, A., Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* **2009**, 11, 49-79.
9. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., 3rd, Protein analysis by shotgun/bottom-up proteomics. *Chemical reviews* **2013**, 113 (4), 2343-2394.
10. Eng, J. K.; McCormack, A. L.; Yates, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **1994**, 5 (11), 976-989.
11. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* **2001**, 19 (3), 242-247.
12. Toby, T. K.; Fornelli, L.; Kelleher, N. L., Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annual review of analytical chemistry* **2016**, 9 (1), 499-519.
13. Schaffer, L. V.; Millikin, R. J.; Miller, R. M.; Anderson, L. C.; Fellers, R. T.; Ge, Y.; Kelleher, N. L.; LeDuc, R. D.; Liu, X.; Payne, S. H.; Sun, L.; Thomas, P. M.; Tucholski, T.; Wang, Z.; Wu, S.; Wu, Z.; Yu, D.; Shortreed, M. R.; Smith, L. M., Identification and Quantification of Proteoforms by Mass Spectrometry. *PROTEOMICS* **2019**, 19 (10), 1800361.

14. Chen, B.; Brown, K. A.; Lin, Z.; Ge, Y., Top-Down Proteomics: Ready for Prime Time? *Analytical Chemistry* **2018**, *90* (1), 110-127.
15. Wu, Z.; Jin, Y.; Chen, B.; Gugger, M. K.; Wilkinson-Johnson, C. L.; Tiambeng, T. N.; Jin, S.; Ge, Y., Comprehensive Characterization of the Recombinant Catalytic Subunit of cAMP-Dependent Protein Kinase by Top-Down Mass Spectrometry. *J Am Soc Mass Spectrom* **2019**, *30* (12), 2561-2570.
16. Lin, Z.; Guo, F.; Gregorich, Z. R.; Sun, R.; Zhang, H.; Hu, Y.; Shanmuganayagam, D.; Ge, Y., Comprehensive Characterization of Swine Cardiac Troponin T Proteoforms by Top-Down Mass Spectrometry. *J Am Soc Mass Spectrom* **2018**, *29* (6), 1284-1294.
17. Zhang, H.; Ge, Y., Comprehensive Analysis of Protein Modifications by Top-Down Mass Spectrometry. *Circulation-Cardiovascular Genetics* **2011**, *4* (6), 711-+.
18. Riley, N. M.; Westphall, M. S.; Coon, J. J., Activated Ion-Electron Transfer Dissociation Enables Comprehensive Top-Down Protein Fragmentation. *Journal of Proteome Research* **2017**, *16* (7), 2653-2659.
19. Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S., Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *Journal of the American Chemical Society* **2013**, *135* (34), 12646-12651.
20. Speers, A. E.; Wu, C. C., Proteomics of Integral Membrane Proteins Theory and Application. *Chemical Reviews* **2007**, *107* (8), 3687-3714.
21. Loo, R. R.; Dales, N.; Andrews, P. C., Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein science : a publication of the Protein Society* **1994**, *3* (11), 1975-83.
22. Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B.; Lantz, C.; Jung, W.; Johnson, K. R.; Koller, A.; Wolff, J. J.; Campuzano, I. D. G.; Auclair, J. R.; Ivanov, A. R.; Whitelegge, J. P.; Paša-Tolić, L.; Chamot-Rooke, J.; Danis, P. O.; Smith, L. M.; Tsybin, Y. O.; Loo, J. A.; Ge, Y.; Kelleher, N. L.; Agar, J. N., Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat Methods* **2019**, *16* (7), 587-594.
23. Loo, R. R. O.; Dales, N.; Andrews, P. C., Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein Science* **1994**, *3* (11), 1975-1983.
24. Speers, A. E.; Wu, C. C., Proteomics of integral membrane proteins--theory and application. *Chem Rev* **2007**, *107* (8), 3687-714.
25. Kachuk, C.; Doucette, A. A., The benefits (and misfortunes) of SDS in top-down proteomics. *Journal of Proteomics* **2018**, *175*, 75-86.
26. Doucette, A. A.; Vieira, D. B.; Orton, D. J.; Wall, M. J., Resolubilization of Precipitated Intact Membrane Proteins with Cold Formic Acid for Analysis by Mass Spectrometry. *Journal of Proteome Research* **2014**, *13* (12), 6001-6012.
27. Yu, Y. Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S.; Gebler, J. C., Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal Chem* **2003**, *75* (21), 6023-8.
28. Saveliev, S. V.; Woodroffe, C. C.; Sabat, G.; Adams, C. M.; Klaubert, D.; Wood, K.; Urh, M., Mass spectrometry compatible surfactant for optimized in-gel protein digestion. *Anal Chem* **2013**, *85* (2), 907-14.
29. Chang, Y.-H.; Gregorich, Z. R.; Chen, A. J.; Hwang, L.; Guner, H.; Yu, D.; Zhang, J.; Ge, Y., New Mass-Spectrometry-Compatible Degradable Surfactant for Tissue Proteomics. *J. Proteome Res.* **2015**, *14* (3), 1587-1599.
30. Brown, K. A.; Chen, B.; Guardado-Alvarez, T. M.; Lin, Z.; Hwang, L.; Ayaz-Guner, S.; Jin, S.; Ge, Y., A photocleavable surfactant for top-down proteomics. *Nature Methods* **2019**.
31. MacLennan, D. H.; Kranias, E. G., Phospholamban: a crucial regulator of cardiac contractility. *Nat Rev Mol Cell Biol* **2003**, *4* (7), 566-77.

32. Lee, S.-H.; Hadipour-Lakmehsari, S.; Murthy, H. R.; Gibb, N.; Miyake, T.; Teng, A. C. T.; Cosme, J.; Yu, J. C.; Moon, M.; Lim, S.; Wong, V.; Liu, P.; Billia, F.; Fernandez-Gonzalez, R.; Stagljar, I.; Sharma, P.; Kislinger, T.; Scott, I. C.; Gramolini, A. O., REEP5 depletion causes sarco-endoplasmic reticulum vacuolization and cardiac functional defects. *Nature Communications* **2020**, *11* (1), 965.
33. Dautant, A.; Meier, T.; Hahn, A.; Tribouillard-Tanvier, D.; di Rago, J.-P.; Kucharczyk, R., ATP Synthase Diseases of Mitochondrial Genetic Origin. *Frontiers in physiology* **2018**, *9*, 329-329.
34. He, J.; Ford, H. C.; Carroll, J.; Douglas, C.; Gonzales, E.; Ding, S.; Fearnley, I. M.; Walker, J. E., Assembly of the membrane domain of ATP synthase in human mitochondria. *Proc. Natl. Acad. Sci.* **2018**.
35. Brown, K. A.; Tucholski, T.; Eken, C.; Knott, S.; Zhu, Y.; Jin, S.; Ge, Y., High-Throughput Proteomics Enabled by a Photocleavable Surfactant. *Angewandte Chemie* **2020**, *132* (22), 8484-8488.
36. Zhu, Y.; Piehowski, P. D.; Zhao, R.; Chen, J.; Shen, Y.; Moore, R. J.; Shukla, A. K.; Petyuk, V. A.; Campbell-Thompson, M.; Mathews, C. E.; Smith, R. D.; Qian, W.-J.; Kelly, R. T., Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nature Communications* **2018**, *9* (1), 882.
37. Zhou, M.; Uwugiaren, N.; Williams, S. M.; Moore, R. J.; Zhao, R.; Goodlett, D.; Dapic, I.; Paša-Tolić, L.; Zhu, Y., Sensitive Top-Down Proteomics Analysis of a Low Number of Mammalian Cells Using a Nanodroplet Sample Processing Platform. *Analytical Chemistry* **2020**, *92* (10), 7087-7095.
38. Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A., Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches. *Journal of the American Society for Mass Spectrometry* **2019**, *30* (7), 1149-1157.
39. Leney, A. C.; Heck, A. J., Native Mass Spectrometry: What is in the Name? *J Am Soc Mass Spectrom* **2017**, *28* (1), 5-13.
40. Blackwell, A. E.; Dodds, E. D.; Bandarian, V.; Wysocki, V. H., Revealing the Quaternary Structure of a Heterogeneous Noncovalent Protein Complex through Surface-Induced Dissociation. *Analytical Chemistry* **2011**, *83* (8), 2862-2865.
41. Crittenden, C. M.; Novelli, E. T.; Mehaffey, M. R.; Xu, G. N.; Giles, D. H.; Fies, W. A.; Dalby, K. N.; Webb, L. J.; Brodbelt, J. S., Structural Evaluation of Protein/Metal Complexes via Native Electrospray Ultraviolet Photodissociation Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2020**, *31* (5), 1140-1150.
42. Wongkongkathep, P.; Han, J. Y.; Choi, T. S.; Yin, S.; Kim, H. I.; Loo, J. A., Native top-down mass spectrometry and ion mobility MS for characterizing the cobalt and manganese metal binding of α -synuclein protein. *Journal of the American Society for Mass Spectrometry* **2018**, *29* (9), 1870-1880.
43. Yen, H.-Y.; Hoi, K. K.; Liko, I.; Hedger, G.; Horrell, M. R.; Song, W.; Wu, D.; Heine, P.; Warne, T.; Lee, Y.; Carpenter, B.; Plückthun, A.; Tate, C. G.; Sansom, M. S. P.; Robinson, C. V., PtdIns(4,5)P₂ stabilizes active states of GPCRs and enhances selectivity of G-protein coupling. *Nature* **2018**, *559* (7714), 423-427.
44. Reading, E.; Liko, I.; Allison, T. M.; Benesch, J. L.; Laganowsky, A.; Robinson, C. V., The role of the detergent micelle in preserving the structure of membrane proteins in the gas phase. *Angew Chem Int Ed Engl* **2015**, *54* (15), 4577-81.
45. Laganowsky, A.; Reading, E.; Hopper, J. T.; Robinson, C. V., Mass spectrometry of intact membrane protein complexes. *Nat Protoc* **2013**, *8* (4), 639-51.
46. Barrera, N. P.; Robinson, C. V., Advances in the mass spectrometry of membrane proteins: from individual proteins to intact complexes. *Annu Rev Biochem* **2011**, *80*, 247-71.
47. Chorev, D. S.; Baker, L. A.; Wu, D.; Beilsten-Edmands, V.; Rouse, S. L.; Zeev-Ben-Mordehai, T.; Jiko, C.; Samsudin, F.; Gerle, C.; Khalid, S.; Stewart, A. G.; Matthews, S. J.; Grünwald, K.; Robinson, C. V., Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry. *Science* **2018**, *362* (6416), 829.

48. Ntai, I.; Fornelli, L.; DeHart, C. J.; Hutton, J. E.; Doubleday, P. F.; LeDuc, R. D.; van Nispen, A. J.; Fellers, R. T.; Whiteley, G.; Boja, E. S.; Rodriguez, H.; Kelleher, N. L., Precise characterization of KRAS4b proteoforms in human colorectal cells and tumors reveals mutation/modification cross-talk. *Proceedings of the National Academy of Sciences* **2018**, *115* (16), 4140.
49. Ge, Y.; Rybakova, I. N.; Xu, Q.; Moss, R. L., Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. *Proceedings of the National Academy of Sciences* **2009**, *106* (31), 12658.
50. Zhang, J.; Guy, M. J.; Norman, H. S.; Chen, Y. C.; Xu, Q.; Dong, X.; Guner, H.; Wang, S.; Kohmoto, T.; Young, K. H.; Moss, R. L.; Ge, Y., Top-down quantitative proteomics identified phosphorylation of cardiac troponin I as a candidate biomarker for chronic heart failure. *J Proteome Res* **2011**, *10* (9), 4054-65.
51. Baker, M., Reproducibility crisis: Blame it on the antibodies. In *Nature*, England, 2015; Vol. 521, pp 274-6.
52. Thingholm, T. E.; Jorgensen, T. J.; Jensen, O. N.; Larsen, M. R., Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nat Protoc* **2006**, *1* (4), 1929-35.
53. Villén, J.; Gygi, S. P., The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature Protocols* **2008**, *3* (10), 1630-1638.
54. Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **2003**, *21* (6), 660-6.
55. Chen, B.; Hwang, L.; Ochowicz, W.; Lin, Z.; Guardado-Alvarez, T. M.; Cai, W.; Xiu, L.; Dani, K.; Colah, C.; Jin, S.; Ge, Y., Coupling functionalized cobalt ferrite nanoparticle enrichment with online LC/MS/MS for top-down phosphoproteomics. *Chem Sci* **2017**, *8* (6), 4306-4311.
56. Roberts, D. S.; Chen, B.; Tiambeng, T. N.; Wu, Z.; Ge, Y.; Jin, S., Reproducible large-scale synthesis of surface silanized nanoparticles as an enabling nanoproteomics platform: Enrichment of the human heart phosphoproteome. *Nano Research* **2019**, *12* (6), 1473-1481.
57. Zhang, J.; Guy, M. J.; Norman, H. S.; Chen, Y.-C.; Xu, Q.; Dong, X.; Guner, H.; Wang, S.; Kohmoto, T.; Young, K. H.; Moss, R. L.; Ge, Y., Top-Down Quantitative Proteomics Identified Phosphorylation of Cardiac Troponin I as a Candidate Biomarker for Chronic Heart Failure. *Journal of Proteome Research* **2011**, *10* (9), 4054-4065.
58. La Vecchia, L.; Mezzena, G.; Zanolla, L.; Paccanaro, M.; Varotto, L.; Bonanno, C.; Ometto, R., Cardiac troponin I as diagnostic and prognostic marker in severe heart failure. *The Journal of Heart and Lung Transplantation* **2000**, *19* (7), 644-652.
59. Dong, X.; Sumandea, C. A.; Chen, Y.-C.; Garcia-Cazarin, M. L.; Zhang, J.; Balke, C. W.; Sumandea, M. P.; Ge, Y., Augmented Phosphorylation of Cardiac Troponin I in Hypertensive Heart Failure. *Journal of Biological Chemistry* **2012**, *287* (2), 848-857.
60. Tiambeng, T. N.; Roberts, D. S.; Brown, K. A.; Zhu, Y.; Chen, B.; Wu, Z.; Mitchell, S. D.; Guardado-Alvarez, T. M.; Jin, S.; Ge, Y., Nanoproteomics enables proteoform-resolved analysis of low-abundance proteins in human serum. *Nat Commun* **2020**, *11* (1), 3903.
61. Cai, W.; Tucholski, T.; Chen, B.; Alpert, A. J.; McIlwain, S.; Kohmoto, T.; Jin, S.; Ge, Y., Top-Down Proteomics of Large Proteins up to 223 kDa Enabled by Serial Size Exclusion Chromatography Strategy. *Anal Chem* **2017**, *89* (10), 5467-5475.
62. Shen, Y.; Tolić, N.; Piehowski, P. D.; Shukla, A. K.; Kim, S.; Zhao, R.; Qu, Y.; Robinson, E.; Smith, R. D.; Paša-Tolić, L., High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics. *J Chromatogr A* **2017**, *1498*, 99-110.
63. Simone, P.; Pierri, G.; Foglia, P.; Gasparrini, F.; Mazzocanti, G.; Capriotti, A. L.; Ursini, O.; Ciogli, A.; Laganà, A., Separation of intact proteins on γ-ray-induced polymethacrylate monolithic columns: A

highly permeable stationary phase with high peak capacity for capillary high-performance liquid chromatography with high-resolution mass spectrometry. *J Sep Sci* **2016**, *39* (2), 264-71.

64. Liang, Y.; Jin, Y.; Wu, Z.; Tucholski, T.; Brown, K. A.; Zhang, L.; Zhang, Y.; Ge, Y., Bridged Hybrid Monolithic Column Coupled to High-Resolution Mass Spectrometry for Top-Down Proteomics. *Analytical chemistry* **2019**, *91* (3), 1743-1747.

65. Gomes, F. P.; Yates, J. R., 3rd, Recent trends of capillary electrophoresis-mass spectrometry in proteomics research. *Mass Spectrom Rev* **2019**, *38* (6), 445-460.

66. Lubeckyj, R. A.; McCool, E. N.; Shen, X.; Kou, Q.; Liu, X.; Sun, L., Single-Shot Top-Down Proteomics with Capillary Zone Electrophoresis-Electrospray Ionization-Tandem Mass Spectrometry for Identification of Nearly 600 Escherichia coli Proteoforms. *Anal Chem* **2017**, *89* (22), 12059-12067.

67. Carroll, J.; Fearnley, I. M.; Walker, J. E., Definition of the mitochondrial proteome by measurement of molecular masses of membrane proteins. *Proceedings of the National Academy of Sciences* **2006**, *103* (44), 16170.

68. Gargano, A. F. G.; Roca, L. S.; Fellers, R. T.; Bocxe, M.; Domínguez-Vega, E.; Somsen, G. W., Capillary HILIC-MS: A New Tool for Sensitive Top-Down Proteomics. *Anal Chem* **2018**, *90* (11), 6601-6609.

69. VanAernum, Z. L.; Busch, F.; Jones, B. J.; Jia, M.; Chen, Z.; Boyken, S. E.; Sahasrabudhe, A.; Baker, D.; Wysocki, V. H., Rapid online buffer exchange for screening of proteins, protein complexes and cell lysates by native mass spectrometry. *Nature Protocols* **2020**, *15* (3), 1132-1157.

70. Muneeruddin, K.; Nazzaro, M.; Kaltashov, I. A., Characterization of intact protein conjugates and biopharmaceuticals using ion-exchange chromatography with online detection by native electrospray ionization mass spectrometry and top-down tandem mass spectrometry. *Anal Chem* **2015**, *87* (19), 10138-45.

71. Muneeruddin, K.; Bobst, C. E.; Frenkel, R.; Houde, D.; Turyan, I.; Sosic, Z.; Kaltashov, I. A., Characterization of a PEGylated protein therapeutic by ion exchange chromatography with on-line detection by native ESI MS and MS/MS. *Analyst* **2017**, *142* (2), 336-344.

72. Chen, B.; Peng, Y.; Valeja, S. G.; Xiu, L.; Alpert, A. J.; Ge, Y., Online Hydrophobic Interaction Chromatography-Mass Spectrometry for Top-Down Proteomics. *Anal Chem* **2016**, *88* (3), 1885-91.

73. Chen, B.; Lin, Z.; Alpert, A. J.; Fu, C.; Zhang, Q.; Pritts, W. A.; Ge, Y., Online Hydrophobic Interaction Chromatography-Mass Spectrometry for the Analysis of Intact Monoclonal Antibodies. *Analytical Chemistry* **2018**, *90* (12), 7135-7138.

74. Regnier, F. E.; Kim, J., Proteins and Proteoforms: New Separation Challenges. *Analytical chemistry* **2018**, *90* (1), 361-373.

75. Compton, P. D.; Zamdborg, L.; Thomas, P. M.; Kelleher, N. L., On the scalability and requirements of whole protein mass spectrometry. *Anal Chem* **2011**, *83* (17), 6868-74.

76. Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L., Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **2011**, *480* (7376), 254-8.

77. Takemori, A.; Butcher, D. S.; Harman, V. M.; Brownridge, P.; Shima, K.; Higo, D.; Ishizaki, J.; Hasegawa, H.; Suzuki, J.; Yamashita, M.; Loo, J. A.; Loo, R. R. O.; Beynon, R. J.; Anderson, L. C.; Takemori, N., PEPMI-MS: Polyacrylamide-Gel-Based Prefractionation for Analysis of Intact Proteoforms and Protein Complexes by Mass Spectrometry. *Journal of Proteome Research* **2020**.

78. McCool, E. N.; Lubeckyj, R. A.; Shen, X.; Chen, D.; Kou, Q.; Liu, X.; Sun, L., Deep Top-Down Proteomics Using Capillary Zone Electrophoresis-Tandem Mass Spectrometry: Identification of 5700 Proteoforms from the Escherichia coli Proteome. *Anal Chem* **2018**, *90* (9), 5529-5533.

79. Yu, D.; Wang, Z.; Cupp-Sutton, K. A.; Liu, X.; Wu, S., Deep Intact Proteoform Characterization in Human Cell Lysate Using High-pH and Low-pH Reversed-Phase Liquid Chromatography. *J Am Soc Mass Spectrom* **2019**, *30* (12), 2502-2513.

80. Shen, X.; Kou, Q.; Guo, R.; Yang, Z.; Chen, D.; Liu, X.; Hong, H.; Sun, L., Native Proteomics in Discovery Mode Using Size-Exclusion Chromatography-Capillary Zone Electrophoresis-Tandem Mass Spectrometry. *Anal Chem* **2018**, *90* (17), 10095-10099.
81. Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Do Vale, L. H. F.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentić, K.; Kelleher, N. L.; Compton, P. D., Top-down characterization of endogenous protein complexes with native proteomics. *Nature Chemical Biology* **2018**, *14* (1), 36-41.
82. Gargano, A. F. G.; Shaw, J. B.; Zhou, M.; Wilkins, C. S.; Fillmore, T. L.; Moore, R. J.; Somsen, G. W.; Paša-Tolić, L., Increasing the Separation Capacity of Intact Histone Proteoforms Chromatography Coupling Online Weak Cation Exchange-HILIC to Reversed Phase LC UVPD-HRMS. *J Proteome Res* **2018**, *17* (11), 3791-3800.
83. Baghdady, Y. Z.; Schug, K. A., Online Comprehensive High pH Reversed Phase × Low pH Reversed Phase Approach for Two-Dimensional Separations of Intact Proteins in Top-Down Proteomics. *Anal Chem* **2019**, *91* (17), 11085-11091.
84. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246* (4926), 64.
85. Susa, A. C.; Xia, Z.; Williams, E. R., Small Emitter Tips for Native Mass Spectrometry of Proteins and Protein Complexes from Nonvolatile Buffers That Mimic the Intracellular Environment. *Analytical Chemistry* **2017**, *89* (5), 3116-3122.
86. Kocurek, K. I.; Stones, L.; Bunch, J.; May, R. C.; Cooper, H. J., Top-Down LESA Mass Spectrometry Protein Analysis of Gram-Positive and Gram-Negative Bacteria. *Journal of the American Society for Mass Spectrometry* **2017**, *28* (10), 2066-2077.
87. Mikhailov, V. A.; Griffiths, R. L.; Cooper, H. J., Liquid extraction surface analysis for native mass spectrometry: Protein complexes and ligand binding. *International Journal of Mass Spectrometry* **2017**, *420*, 43-50.
88. Valeja, S. G.; Kaiser, N. K.; Xian, F.; Hendrickson, C. L.; Rouse, J. C.; Marshall, A. G., Unit mass baseline resolution for an intact 148 kDa therapeutic monoclonal antibody by Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* **2011**, *83* (22), 8391-5.
89. Mamyurin, B. A., Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry* **2001**, *206* (3), 251-266.
90. Fornelli, L.; Durbin, K. R.; Fellers, R. T.; Early, B. P.; Greer, J. B.; LeDuc, R. D.; Compton, P. D.; Kelleher, N. L., Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J Proteome Res* **2017**, *16* (2), 609-618.
91. Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J., The one hour yeast proteome. *Mol Cell Proteomics* **2014**, *13* (1), 339-47.
92. Durbin, K. R.; Fornelli, L.; Fellers, R. T.; Doubleday, P. F.; Narita, M.; Kelleher, N. L., Quantitation and Identification of Thousands of Human Proteoforms below 30 kDa. *Journal of proteome research* **2016**, *15* (3), 976-982.
93. Makarov, A.; Denisov, E., Dynamics of ions of intact proteins in the Orbitrap mass analyzer. *J Am Soc Mass Spectrom* **2009**, *20* (8), 1486-95.
94. Shaw, J. B.; Brodbelt, J. S., Extending the isotopically resolved mass range of Orbitrap mass spectrometers. *Anal Chem* **2013**, *85* (17), 8313-8.
95. Fort, K. L.; van de Waterbeemd, M.; Boll, D.; Reinhardt-Szyba, M.; Belov, M. E.; Sasaki, E.; Zschoche, R.; Hilvert, D.; Makarov, A. A.; Heck, A. J. R., Expanding the structural analysis capabilities on an Orbitrap-based mass spectrometer for large macromolecular complexes. *Analyst* **2017**, *143* (1), 100-105.
96. Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R., High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nature Methods* **2012**, *9* (11), 1084-1086.

97. Kafader, J. O.; Melani, R. D.; Schachner, L. F.; Ives, A. N.; Patrie, S. M.; Kelleher, N. L.; Compton, P. D., Native vs Denatured: An in Depth Investigation of Charge State and Isotope Distributions. *Journal of the American Society for Mass Spectrometry* **2020**, *31* (3), 574-581.
98. Huguet, R.; Mullen, C.; Srzentić, K.; Greer, J. B.; Fellers, R. T.; Zabrouskov, V.; Syka, J. E. P.; Kelleher, N. L.; Fornelli, L., Proton Transfer Charge Reduction Enables High-Throughput Top-Down Analysis of Large Proteoforms. *Analytical Chemistry* **2019**, *91* (24), 15732-15739.
99. Kafader, J. O.; Durbin, K. R.; Melani, R. D.; Des Soye, B. J.; Schachner, L. F.; Senko, M. W.; Compton, P. D.; Kelleher, N. L., Individual Ion Mass Spectrometry Enhances the Sensitivity and Sequence Coverage of Top-Down Mass Spectrometry. *Journal of Proteome Research* **2020**, *19* (3), 1346-1350.
100. Kafader, J. O.; Melani, R. D.; Durbin, K. R.; Ikwaagwu, B.; Early, B. P.; Fellers, R. T.; Beu, S. C.; Zabrouskov, V.; Makarov, A. A.; Maze, J. T.; Shinholt, D. L.; Yip, P. F.; Tullman-Ercek, D.; Senko, M. W.; Compton, P. D.; Kelleher, N. L., Multiplexed mass spectrometry of individual ions improves measurement of proteoforms and their complexes. *Nature Methods* **2020**, *17* (4), 391-394.
101. Wörner, T. P.; Snijder, J.; Bennett, A.; Agbandje-McKenna, M.; Makarov, A. A.; Heck, A. J. R., Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. *Nature Methods* **2020**, *17* (4), 395-398.
102. Harper, C. C.; Elliott, A. G.; Oltrogge, L. M.; Savage, D. F.; Williams, E. R., Multiplexed Charge Detection Mass Spectrometry for High-Throughput Single Ion Analysis of Large Molecules. *Analytical Chemistry* **2019**, *91* (11), 7458-7465.
103. Kafader, J. O.; Beu, S. C.; Early, B. P.; Melani, R. D.; Durbin, K. R.; Zabrouskov, V.; Makarov, A. A.; Maze, J. T.; Shinholt, D. L.; Yip, P. F.; Kelleher, N. L.; Compton, P. D.; Senko, M. W., STORI Plots Enable Accurate Tracking of Individual Ion Signals. *Journal of The American Society for Mass Spectrometry* **2019**, *30* (11), 2200-2203.
104. Macias, L. A.; Santos, I. C.; Brodbelt, J. S., Ion Activation Methods for Peptides and Proteins. *Analytical Chemistry* **2020**, *92* (1), 227-251.
105. Catherman, A. D.; Durbin, K. R.; Ahlf, D. R.; Early, B. P.; Fellers, R. T.; Tran, J. C.; Thomas, P. M.; Kelleher, N. L., Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Molecular & cellular proteomics : MCP* **2013**, *12* (12), 3465-3473.
106. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society* **1998**, *120* (13), 3265-3266.
107. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101* (26), 9528.
108. Zenaidee, M.; Lantz, C.; Perkins, T.; Fu, J.; Jung, W.; Loo, R. R. O.; Loo, J. A., Internal Fragments Generated by Electron Ionization Dissociation Enhances Protein Top-down Mass Spectrometry. **2020**.
109. Cleland, T. P.; DeHart, C. J.; Fellers, R. T.; VanNispen, A. J.; Greer, J. B.; LeDuc, R. D.; Parker, W. R.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S., High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. *J Proteome Res* **2017**, *16* (5), 2072-2079.
110. Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R., Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. *Analytical Chemistry* **2015**, *87* (8), 4152-4158.
111. Gomes, F. P.; Diedrich, J. K.; Saviola, A. J.; Memili, E.; Moura, A. A.; Yates, J. R., 3rd, ETHcD and 213 nm UVPD for Top-Down Analysis of Bovine Seminal Plasma Proteoforms on Electrophoretic and Chromatographic Time Frames. *Anal Chem* **2020**, *92* (4), 2979-2987.

112. Riley, N. M.; Westphall, M. S.; Coon, J. J., Sequencing Larger Intact Proteins (30-70 kDa) with Activated Ion Electron Transfer Dissociation. *Journal of the American Society for Mass Spectrometry* **2018**, *29* (1), 140-149.
113. Riley, N. M.; Sikora, J. W.; Seckler, H. S.; Greer, J. B.; Fellers, R. T.; LeDuc, R. D.; Westphall, M. S.; Thomas, P. M.; Kelleher, N. L.; Coon, J. J., The Value of Activated Ion Electron Transfer Dissociation for High-Throughput Top-Down Characterization of Intact Proteins. *Analytical Chemistry* **2018**, *90* (14), 8553-8560.
114. Greisch, J.-F.; Tamara, S.; Scheltema, R. A.; Maxwell, H. W. R.; Fagerlund, R. D.; Fineran, P. C.; Tetter, S.; Hilvert, D.; Heck, A. J. R., Expanding the mass range for UVPD-based native top-down mass spectrometry. *Chemical Science* **2019**, *10* (30), 7163-7171.
115. Gault, J.; Liko, I.; Landreh, M.; Shutin, D.; Bolla, J. R.; Jefferies, D.; Agasid, M.; Yen, H.-Y.; Ladds, M. J. G. W.; Lane, D. P.; Khalid, S.; Mullen, C.; Remes, P. M.; Huguet, R.; McAlister, G.; Goodwin, M.; Viner, R.; Syka, J. E. P.; Robinson, C. V., Combining native and 'omics' mass spectrometry to identify endogenous ligands bound to membrane proteins. *Nature Methods* **2020**, *17* (5), 505-508.
116. Gault, J.; Liko, I.; Landreh, M.; Shutin, D.; Bolla, J. R.; Jefferies, D.; Agasid, M.; Yen, H. Y.; Ladds, M.; Lane, D. P.; Khalid, S.; Mullen, C.; Remes, P. M.; Huguet, R.; McAlister, G.; Goodwin, M.; Viner, R.; Syka, J. E. P.; Robinson, C. V., Combining native and 'omics' mass spectrometry to identify endogenous ligands bound to membrane proteins. *Nat Methods* **2020**, *17* (5), 505-508.
117. Horn, D. M.; Zubarev, R. A.; McLafferty, F. W., Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *Journal of the American Society for Mass Spectrometry* **2000**, *11* (4), 320-332.
118. Senko, M. W.; Beu, S. C.; McLafferty, F. W., Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *Journal of the American Society for Mass Spectrometry* **1995**, *6* (4), 229-233.
119. Liu, X.; Inbar, Y.; Dorrestein, P. C.; Wynne, C.; Edwards, N.; Souda, P.; Whitelegge, J. P.; Bafna, V.; Pevzner, P. A., Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Molecular & cellular proteomics : MCP* **2010**, *9* (12), 2772-2782.
120. Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V., Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Analytical Chemistry* **2015**, *87* (8), 4370-4376.
121. Jeong, K.; Kim, J.; Gaikwad, M.; Hidayah, S. N.; Heikaus, L.; Schlüter, H.; Kohlbacher, O., FLASHDeconv: Ultrafast, High-Quality Feature Deconvolution for Top-Down Proteomics. *Cell Syst* **2020**, *10* (2), 213-218.e6.
122. Kou, Q.; Xun, L.; Liu, X., TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **2016**, *32* (22), 3495-3497.
123. Yuan, Z. F.; Liu, C.; Wang, H. P.; Sun, R. X.; Fu, Y.; Zhang, J. F.; Wang, L. H.; Chi, H.; Li, Y.; Xiu, L. Y.; Wang, W. P.; He, S. M., pParse: a method for accurate determination of monoisotopic peaks in high-resolution mass spectra. *Proteomics* **2012**, *12* (2), 226-35.
124. McIlwain, S. J.; Wu, Z.; Wetzel, M.; Belongia, D.; Jin, Y.; Wenger, K.; Ong, I. M.; Ge, Y., Enhancing Top-Down Proteomics Data Analysis by Combining Deconvolution Results through a Machine Learning Strategy. *Journal of the American Society for Mass Spectrometry* **2020**, *31* (5), 1104-1113.
125. LeDuc, R. D.; Taylor, G. K.; Kim, Y. B.; Januszzyk, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L., ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res* **2004**, *32* (Web Server issue), W340-5.
126. Liu, X.; Hengel, S.; Wu, S.; Tolić, N.; Pasa-Tolić, L.; Pevzner, P. A., Identification of Ultramodified Proteins Using Top-Down Tandem Mass Spectra. *Journal of Proteome Research* **2013**, *12* (12), 5830-5838.

127. Sun, R.-X.; Luo, L.; Wu, L.; Wang, R.-M.; Zeng, W.-F.; Chi, H.; Liu, C.; He, S.-M., pTop 1.0: A High-Accuracy and High-Efficiency Search Engine for Intact Protein Identification. *Analytical Chemistry* **2016**, *88* (6), 3082-3090.
128. Karabacak, N. M.; Li, L.; Tiwari, A.; Hayward, L. J.; Hong, P.; Easterling, M. L.; Agar, J. N., Sensitive and Specific Identification of Wild Type and Variant Proteins from 8 to 669 kDa Using Top-down Mass Spectrometry. *Molecular & Cellular Proteomics* **2009**, *8* (4), 846.
129. Frank, A. M.; Bandeira, N.; Shen, Z.; Tanner, S.; Briggs, S. P.; Smith, R. D.; Pevzner, P. A., Clustering millions of tandem mass spectra. *J Proteome Res* **2008**, *7* (1), 113-22.
130. Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; Moore, R. J.; Liu, T.; Petyuk, V. A.; Tolić, N.; Paša-Tolić, L.; Smith, R. D.; Payne, S. H.; Kim, S., Informed-Proteomics: open-source software package for top-down proteomics. *Nature Methods* **2017**, *14*, 909.
131. Cesnik, A. J.; Shortreed, M. R.; Schaffer, L. V.; Knoener, R. A.; Frey, B. L.; Scalf, M.; Solntsev, S. K.; Dai, Y.; Gasch, A. P.; Smith, L. M., Proteoform Suite: Software for Constructing, Quantifying, and Visualizing Proteoform Families. *J Proteome Res* **2018**, *17* (1), 568-578.
132. Cai, W. X.; Guner, H.; Gregorich, Z. R.; Chen, A. J.; Ayaz-Guner, S.; Peng, Y.; Valeja, S. G.; Liu, X. W.; Ge, Y., MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics. *Molecular & Cellular Proteomics* **2016**, *15* (2), 703-714.
133. Cai, W.; Tucholski, T. M.; Gregorich, Z. R.; Ge, Y., Top-down Proteomics: Technology Advancements and Applications to Heart Diseases. *Expert review of proteomics* **2016**, *13* (8), 717-730.
134. Peng, Y.; Gregorich, Z. R.; Valeja, S. G.; Zhang, H.; Cai, W. X.; Chen, Y. C.; Guner, H.; Chen, A. J.; Schwahn, D. J.; Hacker, T. A.; Liu, X. W.; Ge, Y., Top-down Proteomics Reveals Concerted Reductions in Myofilament and Z-disc Protein Phosphorylation after Acute Myocardial Infarction. *Mol Cell Proteomics* **2014**, *13* (10), 2752-2764.
135. de Tombe, P. P.; Solaro, R. J., Integration of cardiac myofilament activity and regulation with pathways signaling hypertrophy and failure. *Ann Biomed Eng* **2000**, *28* (8), 991-1001.
136. Katrukha, A. G.; Bereznikova, A. V.; Esakova, T. V.; Pettersson, K.; Lövgren, T.; Severina, M. E.; Pulkki, K.; Vuopio-Pulkki, L.-M.; Gusev, N. B., Troponin I is released in bloodstream of patients with acute myocardial infarction not in free form but as complex. *Clinical Chemistry* **1997**, *43* (8), 1379.
137. Gregorich, Z. R.; Cai, W.; Lin, Z.; Chen, A. J.; Peng, Y.; Kohmoto, T.; Ge, Y., Distinct sequences and post-translational modifications in cardiac atrial and ventricular myosin light chains revealed by top-down mass spectrometry. *Journal of molecular and cellular cardiology* **2017**, *107*, 13-21.
138. Lin, Z.; Wei, L.; Cai, W.; Zhu, Y.; Tucholski, T.; Mitchell, S. D.; Guo, W.; Ford, S. P.; Diffey, G. M.; Ge, Y., Simultaneous Quantification of Protein Expression and Modifications by Top-down Targeted Proteomics: A Case of the Sarcomeric Subproteome. *Molecular & Cellular Proteomics* **2019**, *18* (3), 594-605.
139. Cai, W.; Zhang, J.; de Lange, W. J.; Gregorich, Z. R.; Karp, H.; Farrell, E. T.; Mitchell, S. D.; Tucholski, T.; Lin, Z.; Biermann, M.; McIlwain, S. J.; Ralphe, J. C.; Kamp, T. J.; Ge, Y., An Unbiased Proteomics Method to Assess the Maturation of Human Pluripotent Stem Cell-Derived Cardiomyocytes. *Circ Res* **2019**, *125* (11), 936-953.
140. Azad, N. S.; Rasool, N.; Annunziata, C. M.; Minasian, L.; Whiteley, G.; Kohn, E. C., Proteomics in Clinical Trials and Practice. *Molecular & Cellular Proteomics* **2006**, *5* (10), 1819.
141. Petricoin, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Mills, G. B.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A., Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **2002**, *359* (9306), 572-7.
142. Toby, T. K.; Fornelli, L.; Srzentić, K.; DeHart, C. J.; Levitsky, J.; Friedewald, J.; Kelleher, N. L., A comprehensive pipeline for translational top-down proteomics from a single blood draw. *Nature Protocols* **2019**, *14* (1), 119-152.

143. Petricoin, E. F., 3rd; Ornstein, D. K.; Paweletz, C. P.; Ardekani, A.; Hackett, P. S.; Hitt, B. A.; Velasco, A.; Trucco, C.; Wiegand, L.; Wood, K.; Simone, C. B.; Levine, P. J.; Linehan, W. M.; Emmert-Buck, M. R.; Steinberg, S. M.; Kohn, E. C.; Liotta, L. A., Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst* **2002**, *94* (20), 1576-8.
144. He, L.; Rockwood, A. L.; Agarwal, A. M.; Anderson, L. C.; Weisbrod, C. R.; Hendrickson, C. L.; Marshall, A. G., Diagnosis of Hemoglobinopathy and β -Thalassemia by 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometry and Tandem Mass Spectrometry of Hemoglobin from Blood. *Clin Chem* **2019**, *65* (8), 986-994.
145. Shaw, J. B.; Liu, W.; Vasil'ev, Y. V.; Bracken, C. C.; Malhan, N.; Guthals, A.; Beckman, J. S.; Voinov, V. G., Direct Determination of Antibody Chain Pairing by Top-down and Middle-down Mass Spectrometry Using Electron Capture Dissociation and Ultraviolet Photodissociation. *Analytical Chemistry* **2019**, *92* (1), 766-773.
146. Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schlüter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlén, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschläger, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B., How many human proteoforms are there? *Nat. Chem. Biol.* **2018**, *14*, 206.
147. Siuti, N.; Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **2007**, *4*, 817-821.
148. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009**, *6* (5), 359-62.
149. Chen, E. I.; Cociorva, D.; Norris, J. L.; Yates, J. R., Optimization of Mass Spectrometry Compatible Surfactants for Shotgun Proteomics. *J. Proteome Res.* **2007**, *6* (7), 2529-2538.
150. Meng, F.; Cargile, B. J.; Patrie, S. M.; Johnson, J. R.; McLoughlin, S. M.; Kelleher, N. L., Processing Complex Mixtures of Intact Proteins for Direct Analysis by Mass Spectrometry. *Analytical Chemistry* **2002**, *74* (13), 2923-2929.
151. Bradley, M.; Vincent, B.; Warren, N.; Eastoe, J.; Vesperinas, A., Photoresponsive Surfactants in Microgel Dispersions. *Langmuir* **2006**, *22* (1), 101-105.
152. Hwang, L.; Guardado-Alvarez, T. M.; Ayaz-Gunner, S.; Ge, Y.; Jin, S., A Family of Photolabile Nitroveratryl-Based Surfactants That Self-Assemble into Photodegradable Supramolecular Structures. *Langmuir* **2016**, *32* (16), 3963-3969.
153. Kim, M. S.; Diamond, S. L., Photocleavage of o-nitrobenzyl ether derivatives for rapid biomedical release applications. *Bioorg. Med. Chem. Lett.* **2006**, *16* (15), 4007-10.
154. Dunkin, I. R.; Gittinger, A.; Sherrington, D. C.; Whittaker, P., Synthesis, characterization and applications of azo-containing photodestructible surfactants. *Journal of the Chemical Society, Perkin Transactions 2* **1996**, (9), 1837-1842.
155. Laganowsky, A.; Reading, E.; Hopper, J. T. S.; Robinson, C. V., Mass Spectrometry of Intact Membrane Protein Complexes. *Nature protocols* **2013**, *8* (4), 639-651.
156. Saveliev, S. V.; Woodroffe, C. C.; Sabat, G.; Adams, C. M.; Klaubert, D.; Wood, K.; Urh, M., Mass Spectrometry Compatible Surfactant for Optimized In-Gel Protein Digestion. *Analytical Chemistry* **2013**, *85* (2), 907-914.
157. Cai, W.; Tucholski, T.; Chen, B.; Alpert, A. J.; McIlwain, S.; Kohmoto, T.; Jin, S.; Ge, Y., Top-Down Proteomics of Large Proteins up to 223 kDa Enabled by Serial Size Exclusion Chromatography Strategy. *Analytical Chemistry* **2017**, *89* (10), 5467-5475.

158. Valeja, S. G.; Xiu, L. C.; Gregorich, Z. R.; Guner, H.; Jin, S.; Ge, Y., Three Dimensional Liquid Chromatography Coupling Ion Exchange Chromatography/Hydrophobic Interaction Chromatography/Reverse Phase Chromatography for Effective Protein Separation in Top-Down Proteomics. *Analytical Chemistry* **2015**, *87* (10), 5363-5371.
159. Whitelegge, J. P.; Gundersen, C. B.; Faull, K. F., Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. *Protein science : a publication of the Protein Society* **1998**, *7* (6), 1423-1430.
160. Skinner, O. S.; Catherman, A. D.; Early, B. P.; Thomas, P. M.; Compton, P. D.; Kelleher, N. L., Fragmentation of Integral Membrane Proteins in the Gas Phase. *Analytical Chemistry* **2014**, *86* (9), 4627-4634.
161. Xiu, L. C.; Valeja, S. G.; Alpert, A. J.; Jin, S.; Ge, Y., Effective Protein Separation by Coupling Hydrophobic Interaction and Reverse Phase Chromatography for Top-down Proteomics. *Analytical Chemistry* **2014**, *86* (15), 7899-7906.
162. Riley, N. M.; Mullen, C.; Weisbrod, C. R.; Sharma, S.; Senko, M. W.; Zabrouskov, V.; Westphall, M. S.; Syka, J. E.; Coon, J. J., Enhanced Dissociation of Intact Proteins with High Capacity Electron Transfer Dissociation. *J Am Soc Mass Spectrom* **2016**, *27* (3), 520-31.
163. Durbin, K. R.; Fellers, R. T.; Ntai, I.; Kelleher, N. L.; Compton, P. D., Autopilot: an online data acquisition control system for the enhanced high-throughput characterization of intact proteins. *Anal Chem* **2014**, *86* (3), 1485-92.
164. Hopper, J. T. S.; Robinson, C. V., Mass Spectrometry Quantifies Protein Interactions—From Molecular Chaperones to Membrane Porins. *Angewandte Chemie International Edition* **2014**, *53* (51), 14002-14015.
165. Snijder, J.; Rose, R. J.; Veessler, D.; Johnson, J. E.; Heck, A. J. R., Studying 18 MDa Virus Assemblies with Native Mass Spectrometry. *Angewandte Chemie International Edition* **2013**, *52* (14), 4020-4023.
166. Smith, L. M.; Kelleher, N. L., Proteoforms as the next proteomics currency. *Science* **2018**, *359* (6380), 1106.
167. Gregorich, Z. R.; Chang, Y. H.; Ge, Y., Proteomics in heart failure: top-down or bottom-up? *Pflugers Archiv-European Journal of Physiology* **2014**, *466* (6), 1199-1209.
168. Cai, W.; Tucholski, T. M.; Gregorich, Z. R.; Ge, Y., Top-down Proteomics: Technology Advancements and Applications to Heart Diseases. *Expert Rev Proteomics* **2016**, *13* (8), 717-30.
169. Chait, B. T., Mass Spectrometry: Bottom-Up or Top-Down? *Science* **2006**, *314* (5796), 65.
170. Waas, M.; Bhattacharya, S.; Chuppa, S.; Wu, X.; Jensen, D. R.; Omasits, U.; Wollscheid, B.; Volkman, B. F.; Noon, K. R.; Gundry, R. L., Combine and conquer: surfactants, solvents, and chaotropes for robust mass spectrometry based analyses of membrane proteins. *Anal Chem* **2014**, *86* (3), 1551-9.
171. Zhang, Z.; Wu, S.; Stenoien, D. L.; Pasa-Tolic, L., High-throughput proteomics. *Annu Rev Anal Chem (Palo Alto Calif)* **2014**, *7*, 427-54.
172. Zhang, X., Less is More: Membrane Protein Digestion Beyond Urea-Trypsin Solution for Next-level Proteomics. *Molecular & cellular proteomics : MCP* **2015**, *14* (9), 2441-2453.
173. Botelho, D.; Wall, M. J.; Vieira, D. B.; Fitzsimmons, S.; Liu, F.; Doucette, A., Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation. *J Proteome Res* **2010**, *9* (6), 2863-70.
174. Chang, Y. H.; Gregorich, Z. R.; Chen, A. J.; Hwang, L.; Guner, H.; Yu, D.; Zhang, J.; Ge, Y., New mass-spectrometry-compatible degradable surfactant for tissue proteomics. *J Proteome Res* **2015**, *14* (3), 1587-99.
175. Yu, Y. Q.; Gilar, M.; Gebler, J. C., A complete peptide mapping of membrane proteins: a novel surfactant aiding the enzymatic digestion of bacteriorhodopsin. *Rapid Commun Mass Spectrom* **2004**, *18* (6), 711-5.

176. Yu, Y. Q.; Gilar, M.; Kaska, J.; Gebler, J. C., A rapid sample preparation method for mass spectrometric characterization of N-linked glycans. *Rapid Commun Mass Spectrom* **2005**, *19* (16), 2331-6.
177. Hauser, A.; Penkert, M.; Hackenberger, C. P. R., Chemical Approaches to Investigate Labile Peptide and Protein Phosphorylation. *Accounts of Chemical Research* **2017**, *50* (8), 1883-1893.
178. Wu, D.; Struwe, W. B.; Harvey, D. J.; Ferguson, M. A. J.; Robinson, C. V., N-glycan microheterogeneity regulates interactions of plasma proteins. *Proceedings of the National Academy of Sciences* **2018**, *115* (35), 8763.
179. Mezger, T.; Nuyken, O.; Meindl, K.; Wokaun, A., Light decomposable emulsifiers: application of alkyl-substituted aromatic azosulfonates in emulsion polymerization. *Progress in Organic Coatings* **1996**, *29* (1), 147-157.
180. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **2008**, *26*, 1367.
181. Helbig, A. O.; Heck, A. J.; Slijper, M., Exploring the membrane proteome--challenges and analytical strategies. *J Proteomics* **2010**, *73* (5), 868-78.
182. Bordier, C., Phase separation of integral membrane proteins in Triton X-114 solution. *J Biol Chem* **1981**, *256* (4), 1604-7.
183. Pavic, T.; Gudelj, I.; Keser, T.; Pucic-Bakovic, M.; Gornik, O., Enrichment of hydrophobic membrane proteins using Triton X-114 and subsequent analysis of their N-glycosylation. *Biochim Biophys Acta* **2016**, *1860* (8), 1710-5.
184. Loo, R. R.; Dales, N.; Andrews, P. C., Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein science : a publication of the Protein Society* **1994**, *3* (11), 1975-1983.
185. Solntsev, S. K.; Shortreed, M. R.; Frey, B. L.; Smith, L. M., Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *Journal of Proteome Research* **2018**, *17* (5), 1844-1851.
186. Szklarczyk, D.; Morris, J. H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N. T.; Roth, A.; Bork, P.; Jensen, L. J.; von Mering, C., The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* **2017**, *45* (D1), D362-d368.
187. Cai, W.; Hite, Z. L.; Lyu, B.; Wu, Z.; Lin, Z.; Gregorich, Z. R.; Messer, A. E.; McIlwain, S. J.; Marston, S. B.; Kohmoto, T.; Ge, Y., Temperature-sensitive sarcomeric protein post-translational modifications revealed by top-down proteomics. *J Mol Cell Cardiol* **2018**, *122*, 11-22.
188. Möller, S.; Croning, M. D. R.; Apweiler, R., Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **2001**, *17* (7), 646-653.
189. Cai, W.; Guner, H.; Gregorich, Z. R.; Chen, A. J.; Ayaz-Guner, S.; Peng, Y.; Valeja, S. G.; Liu, X.; Ge, Y., MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics. *Mol Cell Proteomics* **2016**, *15* (2), 703-14.
190. Nuyken, O.; Voit, B., The photoactive diazosulfonate group and its role in polymer chemistry. *Macromolecular Chemistry and Physics* **1997**, *198* (8), 2337-2372.
191. Meier, F.; Brunner, A. D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M. A.; Bache, N.; Hoerning, O.; Cox, J.; Rather, O.; Mann, M., Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol Cell Proteomics* **2018**, *17* (12), 2534-2545.
192. Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J.-J.; Brademan, D. R.; Westphall, M. S.; Coon, J. J., Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Analytical Chemistry* **2018**, *90* (15), 9529-9537.
193. Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P., Multidimensional Separations-Based Shotgun Proteomics. *Chemical Reviews* **2007**, *107* (8), 3654-3686.

194. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R.; Iii, Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnology* **1999**, *17*, 676.
195. Proc, J. L.; Kuzyk, M. A.; Hardie, D. B.; Yang, J.; Smith, D. S.; Jackson, A. M.; Parker, C. E.; Borchers, C. H., A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin. *Journal of Proteome Research* **2010**, *9* (10), 5422-5437.
196. Shao, X.; Taha, I. N.; Clauser, K. R.; Gao, Y.; Naba, A., MatrisomeDB: the ECM-protein knowledge database. *Nucleic Acids Research* **2020**, *48* (D1), D1136-D1144.
197. Naba, A.; Clauser, K. R.; Hoersch, S.; Liu, H.; Carr, S. A.; Hynes, R. O., The Matrisome: In Silico Definition and In Vivo Characterization by Proteomics of Normal and Tumor Extracellular Matrices. *Molecular & Cellular Proteomics* **2012**, *11* (4), M111.014647.
198. Hynes, R. O., The Extracellular Matrix: Not Just Pretty Fibrils. *Science* **2009**, *326* (5957), 1216-1219.
199. Mouw, J. K.; Ou, G.; Weaver, V. M., Extracellular matrix assembly: a multiscale deconstruction. *Nature Reviews Molecular Cell Biology* **2014**, *15* (12), 771-785.
200. Gilkes, D. M.; Semenza, G. L.; Wirtz, D., Hypoxia and the extracellular matrix: drivers of tumour metastasis. *Nature Reviews Cancer* **2014**, *14* (6), 430-439.
201. Isra; Naba, A., Exploring the extracellular matrix in health and disease using proteomics. *Essays in Biochemistry* **2019**, *63* (3), 417-432.
202. Yang, C.; Park, A. C.; Davis, N. A.; Russell, J. D.; Kim, B.; Brand, D. D.; Lawrence, M. J.; Ge, Y.; Westphall, M. S.; Coon, J. J.; Greenspan, D. S., Comprehensive Mass Spectrometric Mapping of the Hydroxylated Amino Acid residues of the $\alpha 1(V)$ Collagen Chain. *Journal of Biological Chemistry* **2012**, *287* (48), 40598-40610.
203. Rienks, M.; Papageorgiou, A.-P.; Frangogiannis, N. G.; Heymans, S., Myocardial Extracellular Matrix. *Circulation Research* **2014**, *114* (5), 872-888.
204. Lindsey, M. L.; Iyer, R. P.; Zamilpa, R.; Yabluchanskiy, A.; DeLeon-Pennell, K. Y.; Hall, M. E.; Kaplan, A.; Zouein, F. A.; Bratton, D.; Flynn, E. R.; Cannon, P. L.; Tian, Y.; Jin, Y.-F.; Lange, R. A.; Tokmina-Roszyk, D.; Fields, G. B.; De Castro Brás, L. E., A Novel Collagen Matricryptin Reduces Left Ventricular Dilation Post-Myocardial Infarction by Promoting Scar Formation and Angiogenesis. *Journal of the American College of Cardiology* **2015**, *66* (12), 1364-1374.
205. Conklin, M. W.; Eickhoff, J. C.; Riching, K. M.; Pehlke, C. A.; Eliceiri, K. W.; Provenzano, P. P.; Friedl, A.; Keely, P. J., Aligned Collagen Is a Prognostic Signature for Survival in Human Breast Carcinoma. *The American Journal of Pathology* **2011**, *178* (3), 1221-1232.
206. Henke, E.; Nandigama, R.; Ergün, S., Extracellular Matrix in the Tumor Microenvironment and Its Impact on Cancer Therapy. *Frontiers in Molecular Biosciences* **2020**, *6*.
207. Naba, A.; Clauser, K. R.; Lamar, J. M.; Carr, S. A.; Hynes, R. O., Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. *eLife* **2014**, *3*.
208. Barrett, A. S.; Wither, M. J.; Hill, R. C.; Dzieciatkowska, M.; D'Alessandro, A.; Reisz, J. A.; Hansen, K. C., Hydroxylamine Chemical Digestion for Insoluble Extracellular Matrix Characterization. *Journal of Proteome Research* **2017**, *16* (11), 4177-4184.
209. Eyre, D. R.; Paz, M. A.; Gallop, P. M., Cross-Linking in Collagen and Elastin. *Annual Review of Biochemistry* **1984**, *53* (1), 717-748.
210. Frantz, C.; Stewart, K. M.; Weaver, V. M., The extracellular matrix at a glance. *Journal of Cell Science* **2010**, *123* (24), 4195-4200.
211. Byron, A.; Humphries, J. D.; Humphries, M. J., Defining the extracellular matrix using proteomics. *International Journal of Experimental Pathology* **2013**, *94* (2), 75-92.
212. Merl-Pham, J.; Basak, T.; Knüppel, L.; Ramanujam, D.; Athanason, M.; Behr, J.; Engelhardt, S.; Eickelberg, O.; Hauck, S. M.; Vanacore, R.; Staab-Weijnitz, C. A., Quantitative proteomic profiling of

extracellular matrix and site-specific collagen post-translational modifications in an in vitro model of lung fibrosis. *Matrix Biology Plus* **2019**, *1*, 100005.

213. Lindsey, M. L.; Jung, M.; Hall, M. E.; Deleon-Pennell, K. Y., Proteomic analysis of the cardiac extracellular matrix: clinical research applications. *Expert Review of Proteomics* **2018**, *15* (2), 105-112.
214. Naba, A.; Pearce, O. M. T.; Del Rosario, A.; Ma, D.; Ding, H.; Rajeeve, V.; Cutillas, P. R.; Balkwill, F. R.; Hynes, R. O., Characterization of the Extracellular Matrix of Normal and Diseased Tissues Using Proteomics. *Journal of Proteome Research* **2017**, *16* (8), 3083-3091.
215. Hill, R. C.; Calle, E. A.; Dzieciatkowska, M.; Niklason, L. E.; Hansen, K. C., Quantification of Extracellular Matrix Proteins from a Rat Lung Scaffold to Provide a Molecular Readout for Tissue Engineering. *Molecular & Cellular Proteomics* **2015**, *14* (4), 961-973.
216. Barallobre-Barreiro, J.; Didangelos, A.; Schoendube, F. A.; Drozdov, I.; Yin, X.; Fernández-Caggiano, M.; Willeit, P.; Puntmann, V. O.; Aldama-López, G.; Shah, A. M.; Doménech, N.; Mayr, M., Proteomics Analysis of Cardiac Extracellular Matrix Remodeling in a Porcine Model of Ischemia/Reperfusion Injury. *Circulation* **2012**, *125* (6), 789-802.
217. Liu, H.; Sadygov, R. G.; Yates, J. R., A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Analytical Chemistry* **2004**, *76* (14), 4193-4201.
218. Goddard, E. T.; Hill, R. C.; Barrett, A.; Betts, C.; Guo, Q.; Maller, O.; Borges, V. F.; Hansen, K. C.; Schedin, P., Quantitative extracellular matrix proteomics to study mammary and liver tissue microenvironments. *The International Journal of Biochemistry & Cell Biology* **2016**, *81*, 223-232.
219. Brown, K. A.; Chen, B.; Guardado-Alvarez, T. M.; Lin, Z.; Hwang, L.; Ayaz-Guner, S.; Jin, S.; Ge, Y., A photocleavable surfactant for top-down proteomics. *Nature Methods* **2019**, *16* (5), 417-420.
220. Garcia-Puig, A.; Mosquera, J. L.; Jiménez-Delgado, S.; García-Pastor, C.; Jorba, I.; Navajas, D.; Canals, F.; Raya, A., Proteomics Analysis of Extracellular Matrix Remodeling During Zebrafish Heart Regeneration. *Molecular & Cellular Proteomics* **2019**, *18* (9), 1745-1755.
221. Rennhack, J. P.; To, B.; Swiatnicki, M.; Dulak, C.; Ogrodzinski, M. P.; Zhang, Y.; Li, C.; Bylett, E.; Ross, C.; Szczepanek, K.; Hanrahan, W.; Jayatissa, M.; Lunt, S. Y.; Hunter, K.; Andrechek, E. R., Integrated analyses of murine breast cancer models reveal critical parallels with human disease. *Nature Communications* **2019**, *10* (1).
222. García-Mendoza, M. G.; Inman, D. R.; Ponik, S. M.; Jeffery, J. J.; Sheerar, D. S.; Van Doorn, R. R.; Keely, P. J., Neutrophils drive accelerated tumor progression in the collagen-dense mammary tumor microenvironment. *Breast Cancer Research* **2016**, *18* (1).
223. Provenzano, P. P.; Inman, D. R.; Eliceiri, K. W.; Beggs, H. E.; Keely, P. J., Mammary Epithelial-Specific Disruption of Focal Adhesion Kinase Retards Tumor Formation and Metastasis in a Transgenic Mouse Model of Human Breast Cancer. *The American Journal of Pathology* **2008**, *173* (5), 1551-1565.
224. Provenzano, P. P.; Inman, D. R.; Eliceiri, K. W.; Knittel, J. G.; Yan, L.; Rueden, C. T.; White, J. G.; Keely, P. J., Collagen density promotes mammary tumor initiation and progression. *BMC Medicine* **2008**, *6* (1), 11.
225. Xu, S.; Xu, H.; Wang, W.; Li, S.; Li, H.; Li, T.; Zhang, W.; Yu, X.; Liu, L., The role of collagen in cancer: from bench to bedside. *Journal of Translational Medicine* **2019**, *17* (1).
226. Bailey, A. J., Collagen and elastin fibres. *J Clin Pathol Suppl (R Coll Pathol)* **1978**, *12*, 49-58.
227. Rosenblum, G.; Van Den Steen, P. E.; Cohen, S. R.; Bitler, A.; Brand, D. D.; Opdenakker, G.; Sagi, I., Direct Visualization of Protease Action on Collagen Triple Helical Structure. *PLoS ONE* **2010**, *5* (6), e11043.
228. Lindsey, M. L.; Hall, M. E.; Harmancey, R.; Ma, Y., Adapting extracellular matrix proteomics for clinical studies on cardiac remodeling post-myocardial infarction. *Clinical Proteomics* **2016**, *13* (1).
229. Suna, G.; Wojakowski, W.; Lynch, M.; Barallobre-Barreiro, J.; Yin, X.; Mayr, U.; Baig, F.; Lu, R.; Fava, M.; Hayward, R.; Molenaar, C.; White, S. J.; Roleder, T.; Milewski, K. P.; Gasior, P.; Buszman, P. P.; Buszman, P.; Jahangiri, M.; Shanahan, C. M.; Hill, J.; Mayr, M., Extracellular Matrix Proteomics Reveals

Interplay of Aggrecan and Aggrecanases in Vascular Remodeling of Stented Coronary Arteries. *Circulation* **2018**, *137* (2), 166-183.

230. Vavken, P.; Joshi, S.; Murray, M. M., TRITON-X is most effective among three decellularization agents for ACL tissue engineering. *Journal of Orthopaedic Research* **2009**, *27* (12), 1612-1618.

231. Millikin, R. J.; Solntsev, S. K.; Shortreed, M. R.; Smith, L. M., Ultrafast Peptide Label-Free Quantification with FlashLFQ. *Journal of Proteome Research* **2018**, *17* (1), 386-391.

232. Shoulders, M. D.; Raines, R. T., Collagen Structure and Stability. *Annual Review of Biochemistry* **2009**, *78* (1), 929-958.

233. Wang, Y.; Yang, F.; Gritsenko, M. A.; Wang, Y.; Clauss, T.; Liu, T.; Shen, Y.; Monroe, M. E.; Lopez-Ferrer, D.; Reno, T.; Moore, R. J.; Klemke, R. L.; Camp, D. G.; Smith, R. D., Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *PROTEOMICS* **2011**, *11* (10), 2019-2026.

234. Batth, T. S.; Francavilla, C.; Olsen, J. V., Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *Journal of Proteome Research* **2014**, *13* (12), 6176-6186.

235. Vizcaíno, J. A.; Côté, R. G.; Csordas, A.; Dienes, J. A.; Fabregat, A.; Foster, J. M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; O'Kelly, G.; Schoenegger, A.; Ovelheiro, D.; Pérez-Riverol, Y.; Reisinger, F.; Ríos, D.; Wang, R.; Hermjakob, H., The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Research* **2012**, *41* (D1), D1063-D1069.

236. Hauser, A. S.; Attwood, M. M.; Rask-Andersen, M.; Schiöth, H. B.; Gloriam, D. E., Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov* **2017**, *16* (12), 829-842.

237. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L., Predicting transmembrane protein topology with a hidden markov model: application to complete genomes¹¹Edited by F. Cohen. *Journal of Molecular Biology* **2001**, *305* (3), 567-580.

238. Doerr, A., Membrane protein structures. *Nature Methods* **2009**, *6* (1), 35-35.

239. Wu, C. C.; Yates, J. R., 3rd, The application of mass spectrometry to membrane proteomics. *Nat Biotechnol* **2003**, *21* (3), 262-7.

240. Whitelegge, J. P.; Zhang, H.; Aguilera, R.; Taylor, R. M.; Cramer, W. A., Full subunit coverage liquid chromatography electrospray ionization mass spectrometry (LCMS+) of an oligomeric membrane protein: cytochrome b(6)f complex from spinach and the cyanobacterium *Mastigocladus laminosus*. *Mol Cell Proteomics* **2002**, *1* (10), 816-27.

241. Kar, U. K.; Simonian, M.; Whitelegge, J. P., Integral membrane proteins: bottom-up, top-down and structural proteomics. *Expert review of proteomics* **2017**, *14* (8), 715-723.

242. Whitelegge, J. P.; Gundersen, C. B.; Faull, K. F., Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. *Protein Science* **1998**, *7* (6), 1423-1430.

243. Catherman, A. D.; Li, M.; Tran, J. C.; Durbin, K. R.; Compton, P. D.; Early, B. P.; Thomas, P. M.; Kelleher, N. L., Top down proteomics of human membrane proteins from enriched mitochondrial fractions. *Anal Chem* **2013**, *85* (3), 1880-8.

244. Carroll, J.; Altman, M. C.; Fearnley, I. M.; Walker, J. E., Identification of membrane proteins by tandem mass spectrometry of protein ions. *Proceedings of the National Academy of Sciences* **2007**, *104* (36), 14330.

245. Brown, K.; Tucholski, T.; Eken, C.; Knott, S.; Zhu, Y.; Jin, S.; Ge, Y., High-throughput Proteomics Enabled by a Photocleavable Surfactant. *Angewandte Chemie International Edition* **2020**, *n/a* (n/a).

246. Schuck, S.; Honsho, M.; Ekroos, K.; Shevchenko, A.; Simons, K., Resistance of cell membranes to different detergents. *Proc Natl Acad Sci U S A* **2003**, *100* (10), 5795-800.

247. Donoghue, P. M.; Hughes, C.; Vissers, J. P.; Langridge, J. I.; Dunn, M. J., Nonionic detergent phase extraction for the proteomic analysis of heart membrane proteins using label-free LC-MS. *Proteomics* **2008**, *8* (18), 3895-905.

248. Duraimurugan, D.; John, I., Performance Evaluation of Tergitol NP-7 and Triton X-114 for the Removal of Crystal Violet Using Cloud-point Extraction. *Chemical and Biochemical Engineering Quarterly* **2016**, *30*, 189-198.
249. Regnier, F. E.; Gooding, K. M., High-performance liquid chromatography of proteins. *Analytical Biochemistry* **1980**, *103* (1), 1-25.
250. Waitt, G. M.; Xu, R.; Wisely, G. B.; Williams, J. D., Automated in-line gel filtration for native state mass spectrometry. *J Am Soc Mass Spectrom* **2008**, *19* (2), 239-45.
251. Wessel, D.; Flugge, U. I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **1984**, *138* (1), 141-3.
252. The UniProt Consortium, UniProt: the universal protein knowledgebase. *Nucleic acids research* **2017**, *45* (D1), D158-D169.
253. Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N. T.; Morris, J. H.; Bork, P.; Jensen, L. J.; Mering, C. V., STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **2019**, *47* (D1), D607-d613.
254. Mi, H.; Muruganujan, A.; Thomas, P. D., PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* **2013**, *41* (Database issue), D377-86.
255. Moller, S.; Croning, M. D.; Apweiler, R., Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **2001**, *17* (7), 646-53.
256. Lin, W.-Z.; Fang, J.-A.; Xiao, X.; Chou, K.-C., iLoc-Animal: a multi-label learning classifier for predicting subcellular localization of animal proteins. *Molecular BioSystems* **2013**, *9* (4), 634-644.
257. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Perez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaino, J. A., The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **2019**, *47* (D1), D442-d450.
258. Smith, S. M., Strategies for the purification of membrane proteins. *Methods Mol Biol* **2011**, *681*, 485-96.
259. Lai, X., Reproducible method to enrich membrane proteins with high purity and high yield for an LC-MS/MS approach in quantitative membrane proteomics. *Electrophoresis* **2013**, *34* (6), 809-17.
260. Lund, R.; Leth-Larsen, R.; Jensen, O. N.; Ditzel, H. J., Efficient Isolation and Quantitative Proteomic Analysis of Cancer Cell Plasma Membrane Proteins for Identification of Metastasis-Associated Cell Surface Markers. *Journal of Proteome Research* **2009**, *8* (6), 3078-3090.
261. Okamoto, T.; Schwab, R. B.; Scherer, P. E.; Lisanti, M. P., Analysis of the association of proteins with membranes. *Curr Protoc Cell Biol* **2001**, *Chapter 5*, Unit 5.4.
262. Ferrige, A. G.; Seddon, M. J.; Jarvis, S.; Skilling, J.; Aplin, R., Maximum entropy deconvolution in electrospray mass spectrometry. *Rapid communications in mass spectrometry* **1991**, *5* (8), 374-377.
263. Tran, J. C.; Doucette, A. A., Multiplexed Size Separation of Intact Proteins in Solution Phase for Mass Spectrometry. *Analytical Chemistry* **2009**, *81* (15), 6201-6209.
264. Tucholski, T.; Knott, S. J.; Chen, B.; Pistono, P.; Lin, Z.; Ge, Y., A Top-Down Proteomics Platform Coupling Serial Size Exclusion Chromatography and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* **2019**, *91* (6), 3835-3844.
265. The UniProt, C., UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **2018**, *47* (D1), D506-D515.
266. Eichacker, L. A.; Granvogl, B.; Mirus, O.; Muller, B. C.; Miess, C.; Schleiff, E., Hiding behind hydrophobicity. Transmembrane segments in mass spectrometry. *J Biol Chem* **2004**, *279* (49), 50915-22.

267. Deutsch, E. W.; Csordas, A.; Sun, Z.; Jarnuczak, A.; Perez-Riverol, Y.; Ternent, T.; Campbell, D. S.; Bernal-Llinares, M.; Okuda, S.; Kawano, S.; Moritz, R. L.; Carver, J. J.; Wang, M.; Ishihama, Y.; Bandeira, N.; Hermjakob, H.; Vizcaino, J. A., The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res* **2017**, *45* (D1), D1100-d1106.
268. Hiller, S.; Abramson, J.; Mannella, C.; Wagner, G.; Zeth, K., The 3D structures of VDAC represent a native conformation. *Trends Biochem Sci* **2010**, *35* (9), 514-21.
269. Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M., Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **2009**, *325* (5942), 834-40.
270. Bienert, S.; Waterhouse, A.; de Beer, Tjaart A. P.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T., The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research* **2016**, *45* (D1), D313-D319.
271. Garavito, R. M.; Ferguson-Miller, S., Detergents as tools in membrane biochemistry. *J Biol Chem* **2001**, *276* (35), 32403-6.
272. Rovini, A., Tubulin-VDAC Interaction: Molecular Basis for Mitochondrial Dysfunction in Chemotherapy-Induced Peripheral Neuropathy. *Front Physiol* **2019**, *10*, 671.
273. Smith, L. M.; Thomas, P. M.; Shortreed, M. R.; Schaffer, L. V.; Fellers, R. T.; LeDuc, R. D.; Tucholski, T.; Ge, Y.; Agar, J. N.; Anderson, L. C.; Chamot-Rooke, J.; Gault, J.; Loo, J. A.; Paša-Tolić, L.; Robinson, C. V.; Schlüter, H.; Tsybin, Y. O.; Vilaseca, M.; Vizcaino, J. A.; Danis, P. O.; Kelleher, N. L., A five-level classification system for proteoform identifications. *Nature Methods* **2019**, *16* (10), 939-940.
274. Sonnhammer, E. L.; von Heijne, G.; Krogh, A., A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **1998**, *6*, 175-82.
275. Schaffer, L. V.; Tucholski, T.; Shortreed, M. R.; Ge, Y.; Smith, L. M., Intact-Mass Analysis Facilitating the Identification of Large Human Heart Proteoforms. *Analytical Chemistry* **2019**, *91* (17), 10937-10942.
276. Dahl, K. N.; Kahn, S. M.; Wilson, K. L.; Discher, D. E., The nuclear envelope lamina network has elasticity and a compressibility limit suggestive of a molecular shock absorber. *Journal of Cell Science* **2004**, *117* (20), 4779.
277. Mirzarafie, A.; Grainger, R. K.; Thomas, B.; Bains, W.; Ustok, F. I.; Lowe, C. R., A fast and mild decellularization protocol for obtaining extracellular matrix. *Rejuvenation Res* **2014**, *17* (2), 159-60.
278. Li, H.; Nguyen, H. H.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A., An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nature Chemistry* **2018**, *10*, 139.
279. Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Do Vale, L. H. F.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentić, K.; Kelleher, N. L.; Compton, P. D., Top-down characterization of endogenous protein complexes with native proteomics. *Nature Chemical Biology* **2017**, *14*, 36.
280. Bolla, J. R.; Agasid, M. T.; Mehmood, S.; Robinson, C. V., Membrane Protein-Lipid Interactions Probed Using Mass Spectrometry. *Annu Rev Biochem* **2019**, *88*, 85-111.
281. Reid, D. J.; Keener, J. E.; Wheeler, A. P.; Zambrano, D. E.; Diesing, J. M.; Reinhardt-Szyba, M.; Makarov, A.; Marty, M. T., Engineering Nanodisc Scaffold Proteins for Native Mass Spectrometry. *Anal Chem* **2017**, *89* (21), 11189-11192.
282. Li, J.; Richards, M. R.; Bagal, D.; Campuzano, I. D. G.; Kitova, E. N.; Xiong, Z. J.; Privé, G. G.; Klassen, J. S., Characterizing the Size and Composition of Saposin A Lipoprotein Picodiscs. *Analytical Chemistry* **2016**, *88* (19), 9524-9531.
283. Lippens, J. L.; Nshanian, M.; Spahr, C.; Egea, P. F.; Loo, J. A.; Campuzano, I. D. G., Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry as a Platform for Characterizing Multimeric Membrane Protein Complexes. *J Am Soc Mass Spectrom* **2018**, *29* (1), 183-193.

284. Hopper, J. T. S.; Yu, Y. T.-C.; Li, D.; Raymond, A.; Bostock, M.; Liko, I.; Mikhailov, V.; Laganowsky, A.; Benesch, J. L. P.; Caffrey, M.; Nietlispach, D.; Robinson, C. V., Detergent-free mass spectrometry of membrane protein complexes. *Nature Methods* **2013**, *10*, 1206.