Development and Application of Isobaric Labeling, ERLIC Separation, and Chemoproteomics Methods for Bioanalytical Research

By
Yusi Cui

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

Doctor of Philosophy
(Chemistry)

at the
University of Wisconsin-Madison
2021

Date of final oral examination: 06/18/2021

This dissertation is approved by the following members of the Final Oral Exam Committee:
Lingjun Li, Professor, Pharmacy and Chemistry
Joshua Coon, Professor, Chemistry
Weiping Tang, Professor, Pharmacy and Chemistry
Luigi Puglielli, Professor, Medicine

Acknowledgments

I would like to express the sincerest gratitude to my PhD supervisor Prof. Lingjun Li, whose continuous support helped me to accomplish this work. She recruited me into many interdisciplinary projects, which were my initial opportunities to explore bioanalytical research. She encouraged me to freely drive some method development projects with interest, from which I learned to become an independent scientist. She provided with not only academic but also career and life advice with wisdom.

I would like to acknowledge my committee members Prof. Joshua Coon, Prof. Weiping Tang and Prof. Luigi Puglielli, for their time and feedbacks to my thesis background exam, research proposal exam and this dissertation.

I would like to acknowledge my collaborators. Dr. Inca Dieterich from Puglielli lab, Dr. Deepa Acharya from Bugni lab and Dr. Ka Yang from Tang lab worked with me on interdisciplinary projects and taught me a lot from different areas. Dr. Andy Alpert from Poly LC Inc gave tremendous technical support and professional advice. I was very fortunate to work with those talented people and hope there are more fruitful collaborations between the labs to come.

Furthermore, I would like to thank my colleagues from Li lab. My mentor Dr. Qing Yu guided me into proteomics field five years ago. Dr. Dustin Frost provided advice for DiLeu labeling. Dr. Junfeng Huang supervised me on PTM research. Dr. Min Ma taught me bioinformatics methods. Dr. Yatao Shi helped me on instrument troubleshooting and gave me technical advice. My mentees Dylan Nicholas Tabang, Danqing Wang and Zishan Zhang provided tremendous help on various projects. Yang Liu, Zichuan Tian, Chris Sauer, Zihui Li, Qinying Yu, Meng Xu,

Ting-jia Gu have always been great friends and colleagues to spend time with. I am grateful to have

Finally, I would like to extend my thanks to my family. My grandparents and parents have given me unconditional love in my life. My fiancé Xinyi Li has been providing the greatest companion for the five years and the years to come.

Table of Contents

Development and Application of Isobaric Labeling, ERLIC Separation, and Chemoproteomics Methods for Bioanalytical Research

Yusi Cui
Under the supervision of Professor Lingjun Li
At the University of Wisconsin-Madison

Abstract

Life science research has entered the systems biology era. Proteomics is a necessary building block to complete the biological networks in modern bioanalytical research. Bottom-up proteomics where the extracted proteins are enzymatically digested into peptides followed by LC separation and tandem mass detection, is still the mainstream proteomics analysis method. Because of the high proteome coverage for complex sample analysis, bottom-up strategy provides great advantage of obtaining the exploratory picture of the proteome network structure, which serves as the first step for proteome dynamics and regulatory mechanistic studies. This dissertation is devoted to the development and improvement of LC-MS/MS based analytical methods for bottom-up proteomics, and their applications to biological, pharmaceutical, and medical research problems.

A portion of this work involved the application of isobaric labeling-based quantification approach, along with other omics techniques, to study the biosynthetic activation mechanism of the natural product keyicin from *Micromononspora* sp. WMMB235. Following this initial study, the quantification accuracy of dimethyl amino acid based isobaric tags was further improved by incorporating a synchronous precursor selection (SPS)-MS3 strategy. With the optimized strategy, proteome regulation of PCSK9 targeting compound HX261 in HepG2 cells was

accurately profiled and characterized. By combining *in situ* photo affinity labeling and click chemistry-based affinity enrichment, a chemoproteomics method was developed to identify the proximal targets of the drug. These innovative method developments provide useful tools for comprehensive understanding of drug action and regulation mechanism during early drug discovery process.

Furthermore, this dissertation also describes the integration of isobaric labeling, electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) and hybrid electron-transfer/higher-energy collision dissociation (EThcD) fragmentation strategy to study the intact N-glycoproteome in two dysfunctional mouse models of AT-1 mutations. The study proved AT-1 is essential to maintain proper organization and engagement of the ER secretory pathway. To expand the application of ERLIC enrichment to other PTMs besides N-glycosylation, the mobile phases were further optimized. Taking advantage of the observations that well-hydrated counterions have the effect of increasing retention of charged analytes in ERLIC, the selectivity of modified peptides in ERLIC separation was further improved. A large-scale identification of thousands of intact phosphopeptides, N-glycopeptides, and O-glycopeptides was achieved for complex mouse brain samples. The method developed in this study can greatly facilitate the PTM crosstalk analysis in various disease models. We anticipate that the technologies developed in this work will find versatile applications and benefit numerous areas of biomedical research.

List of Abbreviations and Acronyms

AGC	automatic gain control
AHL	N-acyl L-homoserine lactone
BCA	bicinchoninic acid
BGC	biosynthetic gene cluster
CVD	cardiovascular disease
DDA	data dependent acquisition
DGE	differential gene expression analysis
DIA	data independent acquisition
DiAla	N,N-dimethyl alanine
DiLeu	N,N-dimethyl leucine
DTT	dithiothreitol
ESI	electrospray ionization
ERLIC	electrostatic repulsion-hydrophilic interaction chromatography
FA	formic acid
GT	glycosyltransferase
HILIC	hydrophilic interaction chromatography
IAA	iodoacetamide
LC	liquid chromatography
LDL	low-density lipoprotein
MALDI	matrix assisted-laser desorption/ionization
MS	mass spectrometry
PCSK9	Proprotein Convertase Subtilisin/Kexin Type 9

QS	quorum sensing
SILAC	stable isotope labeling by amino acid
SPE	solid phase extraction
TFA	trifluoroacetic acid
TMT	tandem mass tag
ITRAQ	isobaric tags for relative and absolute quantitation

Chapter 1

Introduction and Research Summary

Introduction

Around 2000, systems biology emerged as a new paradigm of biological sciences. In contrast to the traditional molecular biology research with reductionism philosophy, which focuses on studying isolated molecules, the system-level understanding of biological systems is unprecedently advocated. The main tasks include understanding the system structures or the networks of biomolecules, the system dynamics or the changes of system over time, the control method of cell states, and the design method of the biological system construction.[1] The comprehensive understanding of biological models requires interdisciplinary integration of omics research, measurement technology and bioinformatics. Such integrated view of a biological system is significant to dissecting cell-regulation mechanisms and will greatly benefit mechanism-driven disease research and pharmaceutical research.

Proteins, as nucleic acids, carbohydrates, and lipids, are one of the most important macromolecular components of biological systems. It is also the component that is most actively involved in the work of living cells and participates in almost all cell activities.[2] Proteins perform versatile biological functions, including but not limited to providing mechanic support for cells, catalysis, transporting molecules, and signal transmission.[3] Systematically studying proteins is structurally and functionally necessary to have a comprehensive understanding of the biological networks. Proteome is the entire complement of proteins expressed by a genome, cell, tissue or organism,[4] and proteomics is the process to study proteome. Proteomics is conceptually and technically far more complex than its nucleic acid counterpart, genomics. Unlike the relatively simple goal of seeking for the complete sequencing of genome, proteomics not only involves revealing the protein sequences, but also the protein quantities, post-translational modifications,

protein-protein and protein-ligand interactions, etc. From the measurement technology aspect, protein analysis lacks the amplification characteristics of PCR for genomics, which is often sample amount limited. Fortunately, with decades of technological development, mass spectrometry (MS) has emerged as the most prevalent method for proteomic measurements.

Mass spectrometry is the technique to measure the mass-to-charge ratio of ionized gas phase molecules, with the instrument named as mass spectrometer that often consists of the ion sources, the mass analyzers, and the detectors. [5] MS was limited to analysis of small and thermostable compounds for long. It was not until the invention of two main-stream soft ionization methods-electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI), has MS expanded to large biomolecule analysis including proteins and peptides. [6] The integration of separation techniques such as liquid chromatography (LC) prior of MS analysis further improved the proteomics for complex samples such as whole yeast digest. The complete sequencing of genomes from different organisms[7–9] and the creation of computational algorithms enabled automatic processing of thousands of protein mass spectra. Collectively, bottom-up proteomics that combines advanced measurement technology and bioinformatics emerged.

There are different methods and logics of MS based proteomics, among which bottom-up proteomics (i.e. shotgun proteomics) is still the most common and widely used. Bottom-up proteomics design contains a basic framework where the extracted proteins are enzymatically digested into peptides followed by LC separation and tandem mass detection.[10] The MS is often conducted in the data dependent acquisition (DDA) mode, where only the precursor ions of top intensities are selected for tandem mass analysis. Because of the high proteome coverage of complex samples, it provides great advantage of obtaining an overall profile of the proteome network structure, which serves as the first step for dynamics and regulatory mechanism study. It

also has the limitation of low abundance protein analysis and run-to-run variations from the stochastic characteristics of DDA. Other useful methods include targeted proteomics, which monitors specific precursor to product ion transitions. Multiple transitions need to be developed as an assay for the quantitation of a protein *a priori*. This method provides higher sensitivity, reproducibility, and quantitation accuracy, which is useful as validation proteomics that complements the discovery proteomics. Data-independent acquisition (DIA) is an emerging strategy, where all precursor ions that fall in a specified mass range are fragmented in a systematic and unbiased fashion for product ion analysis via tandem mass detection. [11] DIA has high reproducibility and preserves the most analyte information, which enables retrospective analysis of the data. However, the data analysis for DIA is much more challenging and convoluted than DDA, where it often requires detailed peptide query parameters, sophisticated software tools and large informatics resources. In the context of this thesis, most of works were accomplished with bottom-up proteomics.

Built upon the bottom-up proteomics framework, numerous bioanalytical research directions have been explored. First, besides mapping the structure of proteome networks by identification, large-scale proteome quantitation is useful to study specific biological questions. Quantitative proteomics study is often hypothesis-driven. First, a perturbed proteome is generated via introducing genetic mutations, drug treatment, stress conditions, etc. Then the protein abundances from both the basal and the perturbed proteome are quantified and compared to test the hypothesis. From MS strategy aspect, there are three main types of relative quantitation methods for shotgun proteomics. Label free quantitation (LFQ) compares the area under the curve (AUC) or MS signal intensities of different LC-MS runs of different samples. [10] Stable isotope labeling strategy introduces stable isotopes to different samples, and the intensity ratios

of their precursor ions with certain mass shifts in the same MS1 spectra are used for quantitation. The labels can be introduced metabolically, such as in the stable isotope labeling by amino acid (SILAC) approach, or chemically, such as in dimethyl labeling. [12] In the isobaric labeling based quantitation, isobaric tag labeled peptides that have identical precursor mass will be distinguished and quantified with reporter ions generated at low mass region after fragmentation in a tandem mass spectrum. MS combined with isobaric labeling can achieve high throughput proteome comparison in one single LC-MS run, providing statistically reliable data that avoids run-to-run variations while saving instrument time.

The bottom-up proteomics has been widely used in modern bioanalytical research, including the field of systems biology, pharmaceutical and medical research. In this dissertation, I will discuss the method development to improve the accuracy of the isobaric labeling-based quantification by SPS-MS3 strategy. I will include the application of chemoproteomics method to facilitate drug discovery process. I will describe the development of electrostatic repulsion-hydrophilic interaction chromatography to facilitate large scale PTM analysis.

Research Summary

Chapter 1 provides the introduction of proteomics in systems biology era, the shotgun proteomics with the focus on quantification perturbed protein networks and proteome post-translational modifications. Furthermore, it summarizes the main analytical technical advances and key findings of each chapter.

Chapter 2 is a multiomics study for the biosynthesis activation mechanism of the natural product keyicin from *Micromononspora* sp. WMMB235. [13] WMMB235 has the biosynthetic

gene clusters *kyc* for keyicin. This normally silent cluster can be activated by coculture fermentation of WMMB235 with a *Rhodococcus* sp. WMMA185. Colorimetric assay of keyicin shows the keyicin production requires small molecules signals or quorum sensing (QS) between WMMA 185 and WMMB 235 instead of the direct cell contact of the two strains. Isobaric labeling based quantitative proteomics experiments and transcriptomics show a marginal number of biosynthetic proteins and transcripts coded by *kyc* cluster are upregulated after simulated coculturing. LC-MS based metabolomics shows that keyicin analogs and precursors are dynamically regulated during coculturing, which correlates with the regulation of *kyc* encoded proteins. The combined multiomics study successfully identified the biosynthetic bottleneck of keyicin in monoculture, which provides a method to exploit silent BGCs.

In Chapter 3, SPS-MS3 strategy was coupled with dimethyl amino acid based isobaric labeling strategy. The quantification accuracy was further improved by optimization of SPS-MS3 experimental parameters. The method was applied to study the proteomic regulation of a PCSK9 targeting compound HX261 in HepG2 cells. Potential drug target proteins were identified using a click-chemistry based photoaffnity labeling and biotin enrichment method coupled with LC-MS/MS.

Chapter 4 describes a quantitative N-glycoproteomic study enabled by an integrated method of isobaric labeling, ERLIC enrichment and EThcD MS fragmentation.[14] Nε-lysine acetylation in the ER is an essential component of the ER quality control machinery. A membrane transporter, AT-1/SLC33A1, has the function of transporting acetyl-CoA from cytosol to ER lumen. The acetylation of the nascent polypeptides within ER lumen will be modulated by AT-1 expression. Two dysfunctional models of AT-1 are developed: AT-1 sTg, a model of systemic AT-1 overexpression, and AT-1$^{S113R/+}$, a model of AT-1 haploinsufficiency. Quantitative proteomics

and N-glycoproteomics by DiLeu labeling and LC-MS/MS analysis are performed to investigate the dynamics of the models. Significant alterations in Golgi-based N-glycan modification, and a marked expansion of the lysosomal network were found from the proteomics results. The study proves AT-1 is essential to maintain proper organization and engagement of the ER secretory pathway.

The successful application of ERLIC for N-glycopeptide enrichment in Chapter 4 takes advantage of the electrostatic interaction in addition to the hydrophilic interaction in ERLIC. It inspired the follow-up study to expand the PTM enrichment target of ERLIC to phosphorylated peptides. Chapter 5 is the optimization of ERLIC mobile phases for simultaneous enrichment of both N-glyco and phosphopeptides.[15] Ten mobile phase compositions of ERLIC were systematically compared, showing selectivity difference over modified peptides. This study demonstrated good enrichment capability of ERLIC for glyco- and phospho- peptides in a complex tryptic digest. Although the study is performed at the solid-phase extraction (SPE) scale, the optimized condition can be adapted to the ERLIC HPLC system, where the high resolution in separating multiple PTMs will benefit large-scale MS-based PTM profiling and in-depth characterization.

Chapter 6 seeks to further improve the enrichment selectivity and PTM coverage of ERLIC coupled with LC-MS/MS analysis. Taking the advantage of counterion optimization in the mobile phases, the selectivity of modified peptides in ERLIC was further improved. Large scale identification of thousands of intact phosphopeptides, N-glycopeptides and O-glycopeptides was achieved. The method developed in the study can facilitate the PTM crosstalk analysis in disease models.

Chapter 7 provides the overall conclusions and future directions of this dissertation.

References

1. Kitano, H. Systems Biology: A Brief Overview. *Science* **295**, 1662–1664 (2002).

2. Patterson, S. D. & Aebersold, R. H. Proteomics: the first decade and beyond. *Nat Genet* **33**, 311–323 (2003).

3. JM, B., JL, T. & L., S. Protein Structure and Function - Biochemistry - NCBI Bookshelf. in (2002).

4. Wasinger, V. C. *et al.* Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis* **16**, 1090–1094 (1995).

5. Ruedi, A. & Matthias, M. Mass spectrometry-based proteomics. *Nature* **422**, 198 (2003).

6. Domon, B. & Aebersold, R. Mass Spectrometry and Protein Analysis. *Science* **312**, 212–217 (2006).

7. Goffeau, A. *et al.* Life with 6000 Genes. *Science* **274**, 546–567 (1996).

8. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

9. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science* **291**, 1304–1351 (2001).

10. Mayne, J. *et al.* Bottom-Up Proteomics (2013–2015): Keeping up in the Era of Systems Biology. *Anal Chem* **88**, 95–121 (2016).

11. Ludwig, C. *et al.* Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol* **14**, e8126 (2018).

12. Yates, J. R., Ruse, C. I. & Nakorchevsky, A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Biomed Eng* **11**, 49–79 (2009).

13. Acharya, D. *et al.* Omics Technologies to Understand Activation of a Biosynthetic Gene Cluster in Micromonospora sp. WMMB235: Deciphering Keyicin Biosynthesis. *Acs Chem Biol* **14**, 1260–1270 (2019).

14. Dieterich, I. A. *et al.* Acetyl-CoA flux from the cytosol to the ER regulates engagement and quality of the secretory pathway. *Sci Rep-uk* **11**, 2013 (2021).

15. Cui, Y. *et al.* Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides. *J Am Soc Mass Spectr* **30**, 2491–2501 (2019).

Chapter 2

Omics Technologies to Understand Activation of a Biosynthetic Gene Cluster in *Micromonospora* sp. WMMB235: Deciphering Keyicin Biosynthesis

Adapted from: Acharya, D.; Miller, I.; **Cui, Y**.; Braun, D. R.; Berres, M. E.; Styles, M. J.; Li, L.; Kwan, J.; Rajski, S. R.; Blackwell, H. E.; Bugni, T. S. Omics Technologies to Understand Activation of a Biosynthetic Gene Cluster in Micromonospora Sp. WMMB235: Deciphering Keyicin Biosynthesis. Acs Chem Biol 2019, 14 (6), 1260–1270. https://doi.org/10.1021/acschembio.9b00223.
Contribution: I performed the experiments and data analysis for proteomics quantitation experiment.

Abstract

DNA sequencing of a large collection of bacterial genomes reveals a wealth of orphan biosynthetic gene clusters (BGCs) with no identifiable products. BGC silencing, for those orphan clusters that are truly silent, rather than those whose products have simply evaded detection and cluster correlation, is postulated to result from transcriptional inactivation of these clusters under standard laboratory conditions. Here, we employ a multi-omics approach to demonstrate how interspecies inter-actions modulate the keyicin producing kyc cluster at the transcriptome level in cocultures of kyc-bearing *Micromonospora* sp. and a *Rhodococcus* sp. We further correlate coculture dependent changes in keyicin production to changes in transcriptomic and proteomic profiles and show that these changes are attributable to small molecule signaling consistent with a quorum sensing pathway. In piecing together the various elements underlying keyicin production in coculture, this study highlights how omics technologies can expedite future efforts to understand and exploit silent BGCs.

Introduction

One of the seminal findings in the arena of infectious disease over the last 20 years has been the realization that quorum sensing (QS) among microbial organisms plays a critical role in dictating how microbes govern themselves.[1–3] Microbial QS entails the generation of extracellular chemical signals, which accumulate in the local environment; once they reach a threshold concentration (and thus a “quorum” of cells has accumulated), the transcription of group-specific genes is activated. Ultimately, these QS-driven changes in transcription constitute

the expression of group-beneficial behaviors, including but not limited to virulence and biofilm formation. [4] Accordingly, it comes as no surprise that numerous campaigns to devise new "antivirulence" agents have targeted bacterial QS systems. [5-7] Generally speaking and informed by the pressing global need to identify new antimicrobial agents, most QS systems studied thus far have involved intraspecies interactions. Indeed, this makes perfect sense as intraspecies interactions needed to gain a "quorum" would be expected of "self-governing" mechanisms needed for a group of cells to achieve common goals.

That said, it is in this *intraspecies* realm that QS has most recently captured the imaginations of drug discovery scientists by influencing microbial secondary metabolic pathways. In short, it is now clear that QS mechanisms offer one means of activating (or "de-repressing") otherwise silent biosynthetic gene clusters (BGCs) leading to the production of new natural products. For instance, elegant work by Hertweck and co-workers revealed the critical role that LuxR-based QS plays in silencing the biosynthesis of thailandamide A lactone in wildtype *Burkholderia thailandensis*. [8] Subsequent work by Seyedsayamdost and co-workers recently revealed that the QS-controlled transcriptional regulator ScmR serves as a global gatekeeper of secondary metabolism in *Burkholderia thailandensis* E264[9] and repressor of numerous BGCs, whereas Greenberg and co-workers have shown that QS in *B.thailandensis* impacts biosynthetic gene clusters that code for the synthesis of malleobactin, malleilactone, quinolones, rhamnolipids, and others. [10] Importantly, all of these QS systems are of the LuxI/LuxR class that is typical of Gram-negative bacteria. [11] These systems consist of a LuxI-type synthase that produces a diffusible N-acyl L-homoserine lactone (AHL) signal, and a LuxR-type receptor that binds the AHL and activates transcription of QS-controlled genes. AHLs constitute the extracellular chemical

signals by which bacteria communicate en route to self-governance. [12-14] By extension, the relevance of interspecies associations to BGC activation processes holds tremendous promise and now constitutes an area of active investigation within our laboratory.

It is now clear that QS mechanisms offer one viable means by which to modulate the biosynthetic machineries housed within specific microbes. Specifically, the activation of otherwise silent or "orphan" BGCs is a particularly exciting application of QS pathways since it is now well established that microbial genetic diversity and possibilities far exceed previous expectations with respect to secondary metabolism and the natural product-based drug leads to which they give rise. For instance, a survey of only 1154 genomes revealed >10000 distinct biosynthetic gene clusters (BGCs), a number that is 10-fold greater than the *TOTAL number* of experimentally characterized BGC's currently in the MIBiG repository. [15-17] Alarmingly, no small molecule-to-BGC correlation can be made for the overwhelming majority of these clusters. [18-23]

Activation of silent BGCs (identified by genomics) has been achieved by (i) changing growth conditions, (ii) chemical elicitors, (iii) targeted genetic modifications, (iv) alterations to transcriptional machineries, and (v) heterologous expression methods;[22] although comprehensive correlations of natural product biosynthesis to changes in transcriptomics or proteomics have rarely been achieved. Genomic information alone, though useful from BGC mining initiatives, is insufficient to unveil and make available new secondary metabolites; transcriptomic, proteomic, and metabolomic data play indispensable roles in producing new structures from otherwise silent BGCs.

Coculturing different microbes has also been shown to activate silent BGCs,[24–28] and resulting metabolites are likely attributable to interspecies QS mechanisms that alter one or more of the

factors noted above. Consistent with this notion, we recently reported that coculturing of *Micromonospora* sp. WMMB235 and a *Rhodococcus* sp. WMMA185 enabled the production of a new glycosylated anthracycline constructed by a large type II PKS, keyicin. Neither bacterium, in isolation, produced keyicin. [29] Despite the excitement and significance of this finding, little is known about how these organisms synergize to generate keyicin.

Early studies into keyicin production revealed WMMB235 as the producer of keyicin (1, Figure 1a) in coculture [see Figure S10, Supporting Information for statistical analysis (triplicates) via PCA]. [29] This conclusion was supported primarily by two key findings. The first one was that only WMMB235 harbored a BGC (termed previously and herein kyc) able to code for all the machinery anticipated to be necessary for keyicin assembly; this was first illuminated upon PRISM[30] and AntiSmash[31−33] processing of the WMMB235 genome. Both analyses identified kyc as a large anthracycline type biosynthetic gene cluster housing several glycosyltransferases (GTs) envisioned as essential to keyicin assembly (Supporting Information). Second, fermentations in which the two microbial species were separated with a 0.2 μm cell impermeable membrane led, over time, to inhibited *Rhodococcus* sp. WMMA185 growth and increased keyicin production; this assay highlighted the antibacterial properties of keyicin as well as the absence of any required interspecies cell−cell contacts. [29]

We focus here on understanding how *Micromonospora* sp. WMMB235 and *Rhodococcus* sp. WMMA185 collaborate to produce keyicin via the application of genomics, transcriptomics, and proteomics technologies. We pay special attention to identifying biosynthetic bottleneck processes as well as keyicin analogs and how these findings might translate to other silent BGC systems. We also test the hypothesis that a LuxR-type receptor homologue, embedded within the keyicin BGC (kyc) dictates keyicin production in a fashion consistent with QS.

Results and Discussion

Keyicin Production Is Small-Molecule-Triggered. To unequivocally determine the mode of interaction between WMMB235 and WMMA185, we expanded on the results of previously reported two chamber fermentation assays. [29] We found that keyicin production, as detected by colorimetric analysis (λ_{max} = 470nm), from WMMB235 could be triggered simply by subjecting WMMB235 to supernatant from monocultured *Rhodococcus* sp. WMMA185. When inoculated into cell free media from a WMMA185 culture grown for 4 days, WMMB235 clearly generated **1** as reflected by production of keyicin's unique chromophore; the efficiency of keyicin production using WMMA185-derived supernatant was virtually identical to that seen in live coculture experiments, suggesting that nutrient depletion (by live WMMA185) exerts little to no influence upon the kyc machinery of WMMB235 (Figures S6 and S7, Supporting Information). This result clearly put to rest any possibility that WMMB235 and WMMA185 are involved in a dynamic communication system that requires both participants to be alive or active. Moreover, this experiment showed that WMMA185 produces a small molecule inducer of keyicin biosynthesis even in the absence of WMMB235, suggestive that QS may play an important, though not exclusive, role in triggering kyc BGC activation. Activation of keyicin biosynthesis by WMMB235 using only supernatants from monocultured WMMA185 also strongly affirms that keyicin biosynthesis is in response to a small molecule signal from WMMA185 and not a nutrient depletion or competition phenomenon.

Transcriptomic Activation of the kyc Cluster and Keyicin Production in Coculture. Early sequencing efforts made clear that keyicin assembly in coculture could be ascribed only to

WMMB235. [29, 34] To evaluate how kyc biosynthetic genes are impacted by the presence of WMMA185, we collected cells from days 2 and 5 of the cultures of WMMB235, WMMA185, and their coculture; LC/MS and colorimetric analyses (λ_{max} = 470 nm) revealed that keyicin was not produced in substantial quantities until day 4 of fermentation. Illumina sequencing of each mRNA collection enabled alignments of the resulting RNASeq data for the two genomes in order to parse transcript reads for WMMB235 (producer). The aggregate value of reads per kb/million reads aligning to annotated ORFs (RPKMO) of the gene clusters was calculated from the number of reads for each gene in the cluster that could be mapped to the genome. [35] This value was normalized to cluster length to ensure accurate representation of the smaller gene clusters in the genome relative to the large kyc cluster. Overall, the kyc cluster had RPKMO values of 1303.0 and 267 in monoculture on days 2 and 5, respectively. In coculture, the same RPKMO values were 1849.1 at day 2 and 2267.7 at day 5, consistent with significantly increased transcription of the kyc cluster in the presence of WMMA185 and the commensurate reduction of the same transcripts over time in monocultured WMMB235.

Further differential gene expression analyses (DGE) were conducted on day 5 data using EdgeR software[36] allowed us to quantitate the magnitude of differential expression of each WMMB235 gene in coculture relative to monoculture as a fold-change of read counts, along with the significance of this difference as adjusted p-value or false discovery rate (FDR). The volcano plot representative of this analysis (Figure 1b) revealed that, of all upregulated *orfs* within the complete WMMB235 genome, putative *kyc* cluster genes were among the most upregulated and had the lowest rates of false discovery. In fact, the vast majority of *orfs* within the *kyc* cluster showed at least a 2-fold (to the log2) increase in gene expression. These transcriptomic analyses suggested that the presence of WMMA185 induces the transcriptional

activation of the *kyc* cluster. We posit that, in the absence of WMMA185, WMMB235 channels resources to other metabolic machineries unrelated to the assembly of **1**. These findings are critical to expanding our understanding of the "omics" behind keyicin production as well as the potential role that QS plays in *kyc* cluster activation in *Micromonosopora* sp. WMMB235.

A Role for Quorum Sensing through LuxR in kyc Cluster Activation? Our ability to correlate *kyc* BGC expression within WMMB235 to the production of **1**, coupled with the realization that pathway-specific regulators often cluster within or proximal to BGCs,[37] inspired us to search the *kyc* cluster for regulatory gene candidates. This search revealed, among others, a luxR-type transcriptional regulator termed herein *kyc5*. In view of LuxR's well established role in QS in Gram-negative bacteria,[1,38,39] we posited that Kyc5 activation (via AHL exposure) may trigger keyicin production. Accordingly, we investigated the impact of established LuxR-selective ligands upon keyicin production. A library of 96 AHLs and related analogs (both natural products and synthetics) were screened for the ability to trigger keyicin production by monocultures of WMMB235.

A range of putative LuxR ligand concentrations were investigated, and even at the lowest concentration (1 nM), six compounds (Figure 2a) were found to activate keyicin production as detected by increases in absorbance at 470 nm, which is diagnostic for production of the keyicin aglycone. The fold change in production of **1** was calculated by measuring the absorbance of cell supernatants of WMMB235 spiked with inducers (in DMSO) compared to its monoculture treated with DMSO alone. AHL-triggered keyicin production by WMMB235 was not as pronounced as in the WMMB235/WMMA185 coculture system (Figure 2b) indicating that production of **1** is subject to more than just one regulatory element. This, combined with the absence of any decipherable *luxI* homologues in the WMMA185 genome,[40] suggests that the

LuxR pathway in the WMMA185/WMMB235 coculture system may respond to alternative (non-AHL) signals. Alternatively, kyc activation may be triggered by an altogether different mechanism. Although *luxI*/*luxR* are well studied in Gram-negative bacteria, it is only recently that these have been found not only in Gram-positive bacteria but also in other kingdom representatives.[41,42] In fact, several genomic studies across different species have found many QS-related *luxR* type genes that are unpaired to a cognate *luxI* to synthesize the signaling molecule and thus encode "orphan" LuxR receptors or "solos". [43, 44] This supports our hypothesis that an "orphan" LuxR in *Micromonospora* sp. may be involved in interspecies communication by interacting with the small molecule signal from *Rhodococcus* sp. It is altogether possible that keyicin production may require pathways in addition to, or even instead of, LuxR. For instance, efficient production of pyocyanin, a phenazine virulence factor produced by Pseudomonas aeruginosa calls upon a total of three separate, but interwoven regulatory systems. [45] The results of Figure 2 may reflect a similar scenario in which LuxR-type signaling plays an important but not exclusive role in *kyc* cluster activation. In the absence of other clear and readily testable regulatory elements related to *kyc* activation, we sought to better understand the proteomics and transcriptomics of the WMMB235/WMMA185 coculture.

Isobaric Tagging Reveals Important Proteomic Profiles Unique to WMMB235/WMMA185 Coculture. We applied a quantitative proteomics approach to evaluate WMMB235/WMMA185 cocultures to identify unique ele- ments of coculture that could be clearly correlated to kyc cluster activation and biosynthesis of 1. Proteomics initiatives were carried out on 5-day long and 8-day long fermentations in order to most accurately capture protein levels. To reduce the complexity of samples subjected to proteomics, we employed a simulated coculture system wherein the supernatant of WMMA185 (5 d fermentation) was used

as the WMMB235 growth medium. This enabled us to more confidently assign proteomic signatures to the keyicin producer and not Rhodococcus products. Established DiLeu tagging methods46 allowed us to multiplex all samples and generate quantitative data on proteins of WMMB235 origin.

A marginal number (i.e., 12) of putatively biosynthetic proteins coded for by the kyc cluster (~49 based on *kyc orfs*) were identified from the 8-day fermentation of WMMB235 (Figure 3); these included four glycosyltransferases (GTs) (Kyc12, Kyc20, Kyc25, and Kyc32), all of which are significantly upregulated in coculture, as well as the putative dehydratase (Kyc7), β-keto acyl synthase (Kyc14), epimerase (Kyc16), cytochrome P_{450} (Kyc26), and dTDP-4-dehydro-rhamnose 3,5-epimerase (Kyc28). The GTs Kyc12, Kyc20, and Kyc32 are all homologous to AknK, which is the GT responsible for adding the second and third 2-deoxy-L-fucose moieties during aclacinomycin biosynthesis (MIBiG, BGC0000191). [47] Kyc25, on the other hand, shares 61% identity with CosG, a GT known to transfer aminodeoxysugars like L-rhodosamine during biosynthesis of cosmomycin D (MIBiG, BGC0001074). 48 In depth sequence analyses for Kyc28 revealed 69% similarity to the sugar 3′-5′ epimerase SnogF (accession no. A0QSK5.1) involved in the deoxyhexose pathway required for nogalamycin assembly.[49] Interestingly, this protein was one of the few found in our initial proteomics study to be predominant in the coculture compared to the WMMB 235 monoculture. [29] Additionally, Kyc26 was reported to have 42% protein identity with SnogN, [29] which in turn shares similarity with AknT (43%) and CosT (39%) and is also considered to be involved in the deoxyhexose pathway, especially in the biosynthesis of nogalamine. [50] The closest homologue to Kyc16 is a putative NDP-sugar 4-ketoreductase encoded within the versipelostatin gene cluster (MIBiG, BGC0001204). Importantly, Kyc14 with 65% identity with AknC, also from the aclacinomycin biosynthetic

pathway, is the only *kyc orf* product found thus far that plays a role in the biosynthesis of the keyicin aglycone. Overall, four of the 12 proteins (33%) identified at the late 8 d fermentation are GTs suggesting that glycosylation likely takes place following aglycone assembly, that is, late in keyicin biosynthesis. As such, the GTs involved in keyicin production appear to function as true tailoring enzymes. Such a profile is consistent with other anthracycline biosynthetic studies where hydroxylated aglycone intermediates added exogenously to fermentation systems serve as efficient substrates for glycosylation. [51]

Coculture Dependent Proteomic Changes Correlate to kyc Cluster Metabolomics. The prominent changes in GT production found in coculture versus monoculture inspired us to investigate the prospect that keyicin analogs or precursors might be generated during coculture and other related keyicin-generating conditions but may have evaded detection. This hypothesis was further supported by the clear presence of many other compounds with distinct retention times in coculture extracts, all of which contained a chromophore with unique absorption at λ = 470 nm and MS/MS signals at m/z = 550.1696 and 586.1899 representative of the keyicin aglycone (Figure S4, Supporting Information). The relationship of these molecules with keyicin could be easily identified by subjecting the liquid chromatography tandem mass spectrometry (LC-MS/MS) analyses of coculture extracts collected over a fermentation period of 14 days to Global Natural Product Social (GNPS) Molecular Networking[52] and subsequent visualization by Cytoscape53 (Figure 4). By tracking the node representative of keyicin (m/z 805.347), we identified the subcluster that contained the keyicin analogs and intermediates. On mapping the AUC (Area Under Curve) for each of the parent masses identified in the cluster at each time point, we discovered that many of these signals initially increased in intensity and then gradually subsided with time consistent with the biosynthetic progression leading ultimately to keyicin and

away from incompletely glycosylated intermediates or precursors. For example, a doubly charged peak on the chromatogram corresponding to m/z values of 661.3050 and 645.7829 show a distinct temporal pattern. We propose that these signals represent differentially glycosylated analogs of keyicin. The m/z value of 645.7829 is consistent with decilonitrose,[29] the adduct resulting from the absence of keyicin's terminal 2-deoxy-fucose (S7, Figure 1). Additionally, the absence of both S3 and S7 (Figure 1) is likely reflected by the m/z signal at 661.305 (Figure S5, Supporting Information).

Having identified changes to the transcriptomic and proteomic profile as well as the players in keyicin production, and realizing that these changes likely invoke coculture-dependent changes that go beyond changes in *kyc* expression, we next sought to investigate *kyc*-specific transcriptional changes as well as those of the whole WMMB235 genome.

Impacts of Coculture on the WMMB235 Genome Revealed by Transcriptomics.

Transcriptomic evaluation of the Micromonospora sp. WMMB235 genome in WMMB235/ WMMA185 cocultures dramatically expanded what we know about *kyc* activation as well as the modulation of other WMMB235 embedded BGCs. These efforts also provided clarity into how *Rhodococcus*-derived small molecule induction can be correlated to changes in protein expression and production and commensurate biosynthesis of **1**. Notably, transcriptomics analysis of the *kyc* cluster revealed that effectively all *orfs* within the cluster show some level of overexpression. Only kyc4, 5, 30, 31, 39−42, and 53 showed less than a 4-fold increase in expression relative to WMMB235 monoculture (Figure 5). Interestingly, kyc30 and kyc31, both coding for transcriptionally suppressive regulators (Table S3, Supporting Information) were slightly suppressed under coculturing (FC < 1.0). With respect to kyc cluster-specific changes, it is clear that the overwhelming majority of *kyc* cluster elements in WMMB235 monoculture

suffer from limited transcription relative to coculture. That coculture driven transcriptomic enhancements are so much more dramatic than those seen at the proteomics level suggests that the production of **1** in WMMB235 monoculture is most likely limited or bottlenecked at a transcriptional level.

The clustering of biosynthetic genes on bacterial chromosomes enables computational approaches to identifying sources of new natural products.[54] Although the biological logic for defining and identifying a BGC is conserved, different algorithms approach BGC prediction differently. AntiSmash v3.0 data processing for the WMMB235 genome[34] revealed the presence of 50 putative BGCs, whereas PRISM processing of the same data set revealed the presence of 10 putative BGCs. The abundance of BGCs found by AntiSmash can be attributed to the low confidence/high novelty algorithm of ClusterFinder to identify BGCs. This probabilistic algorithm is optimized for detecting unknown types of gene clusters and consequently gives relatively high rates of false positives in the results. [15, 33, 54] As such, we restricted our transcriptomics analyses to only BGCs that resulted from PRISM; as expected, these same BGCs were also identified by AntiSmash processing of the WMMB235 genome.

In addition to their correlation of *kyc* cluster elements (i.e., transcripts and protein levels) to the production and structure of **1**, transcriptomics on the WMMB235 genome revealed that many other putative BGCs (annotated using PRISM) undergo transcriptomic changes in response to coculture with Rhodococcus sp. WMMA185. These include, as summarized in Table 1, several hybrid NRPS-type I PKS gene clusters (BGC4, 5, 6, 8), a type II PKS (BGC3), an AT-less type I PKS (BGC2), and clusters encoding a putative enediyne (BGC7) and lanthipeptide (BGC10). Impressively, of the 10 putative BGCs identified, nine are positively impacted by the presence of WMMA185 during fermentation, and of these, *kyc* was the dominantly impacted cluster (Figure

6). This transcriptomics finding is especially interesting since all 10 BGCs, except for *kyc* which has 79% similarity to aclacinomycin, have little similarity to known clusters in the MIBiG repository, making them orphan clusters (Table 1). Tentative mapping of BGC2−10 is shown in Figure 7, and prominent in these findings is the presence of luxR orfs embedded within BGC3 and 8. These findings bolster our hypothesis that new chemical scaffolds are yet to be found in WMMB235 and that coculturing or its emulation (e.g., via LuxR activation with synthetic AHLs or WMMA185-derived LuxR agonists) may enable a host of new natural product discoveries.

Coculturing WMMB235 with WMMA185 appeared to impair transcription for only two BGCs within the WMMB235 genome, BGC2, a type I PKS−NRPS hybrid with a trans-acyltransferase (AT) domain, and BGC9, which is a type I PKS. The closest homologous cluster for BGC2 is that of leinamycin[56] (MIBiG, BGC0001101) with only 15% similarity. This cluster, in particular, is interesting as its expression is enhanced early on during coculturing but is then slightly repressed by day 5. Perhaps most interesting about this finding is that, of the BGC2 orfs suppressed in coculture at day 5, those involving transport are the most strongly represented. Though further studies await, we envision that diminished transporter production with respect to BGC2 may represent some form of defense by way of restricted extracellular access.

BGC9 is the smallest BGC identified of the 10 found in the WMMB235 genome and is a type I PKS with only 7% similarity to any known BGC, chlorizidine A (MIBiG, BGC0001172) (Table 1). For BGC2, eight genes had >1 log negative fold change on day 5 (Table S4, Supporting Information), which notably included all the transporter genes. For BGC9, *orf* downregulation was prominent and consistent over the full course of coculture fermentation. BGC9 suppression, as reflected in Figure 6 was dramatic relative to all other BGCs noted. A point-by-point assessment of specific *orfs* within BGC9 that were negatively impacted by

coculturing is shown in Table S5 (Supporting Information). Unlike the BGC2 case, there is no one group or type of gene that is significantly downregulated although it is clear that elements of the PKS machinery for BGC9 are clearly impacted by coculturing. We posit that this may simply be a random event, or more enticingly, this may represent one means by which WMMB235 turns down production of a BGC9 encoded product. This may benefit the organism by allowing raw materials to be more wisely used given the competitive conditions of coculturing, or it may be a direct means of self- defense.

Transcriptomics of the WMMB235 genome made clear the ubiquity with which this organism appears to employ LuxR- based QS systems in embedded BGCs. BGC8 (Figures 6 and 7) was found to contain 184.203 kbp of information and to encode for a hybrid NRPS−type I PKS cluster with 22 modules and with 3-amino-5-hydroxy benzoic acid (AHBA) as a predicted substrate in one of the synthetase/ligase domains (Figure 7). A large group of natural products in the family of ansamycins, mitomycins, and saliniketals utilize AHBA as a precursor,[64] although BGC8 is only 35% similar to the closest ansamycin BGC of rifamycin (MIBiG, BGC0000137). More telling about BGC8 is that, in contrast to *kyc*, it contains two luxR genes, both of which code for products with 36% similarity to GdmRII (ABI93788.1). GdmRI and GdmRII are known homologues of LuxR proteins that positively regulate the production of geldanamycin in *Streptomyces hygroscopicus* 17997. [65] This suggests that BGC8, in addition to *kyc*, may also be regulated using small molecule inducers. Notably, although BGC8 is an orphan cluster, similar clusters are present in other *Micromonospora* sp. such as *Micromonospora* sp. strain B006. [66] Finally, it warrants noting that, besides *kyc* and BGC8, these transcriptomics studies revealed the presence of two *luxR* genes within the type II PKS-encoding BGC3. It is clear that WMMB235

embedded BGCs harbor the potential to exploit LuxR-based QS pathways, presumably to regulate secondary metabolism in response to assorted cellular challenges.

Changes in Global KEGG Categorization. Analysis of transcriptomics and genomics data for WMMB235 monocultures versus WMMB235/WMMA185 cocultures using Kyoto Encyclopedia of Genes & Genomes (KEGG) software revealed further insight into changes that occur within WMMB235 during coculture and that likely have a bearing on the expression of the *kyc* and other BGCs and their products.

As reflected in Figure 8, significant shifts are seen based on the duration of fermentation (2 vs 5 days) as well as the presence or absence of WMMA185. Particularly interesting are the significant increases in carbohydrate metabolism, energy metabolism, and metabolism of terpenoids and polyketides observed in coculture at day 5 (Figure 8, lane 4). Perhaps also noteworthy is the apparent reduction in translational capacity at day 5 in coculture relative to WMMB235 monoculture. Notably, these changes are reflective of altered gene expression with respect to the whole WMMB235 genome and most certainly encompass changes that have a bearing on keyicin production. Indeed, it is likely that the results of these KEGG studies can be understood, in part, by the transcriptomics changes depicted in Figure 6. In essence, the results shown in Figures 6 and 8 are clearly related; whereas Figure 6 conveys *kyc* cluster specific changes, Figure 8 provides a more global view of how WMMB235 genome readout and processing changes in response to coculturing with WMMA185.

Future Directions. In sum, the ability to track transcriptomic and proteomic information in relation to WMMB235/WMMA185 coculture and subsequent keyicin production sheds significant insight into the activation of *kyc*, an otherwise silent BGC. That keyicin production is also triggered by a panel of established QS ligands for LuxR supports the involvement of a

LuxR-based system in dictating whether WMMB235 can generate keyicin. The correlation of these omics data, QS results, and keyicin biosynthesis support the notion that, by comparing BGC host transcriptomes, proteomes, and metabolomes between monoculture and coculture scenarios, the identification of biosynthetic bottlenecks in monoculture as well as strategies by which to circumvent or overcome such bottlenecks is readily feasible. Our findings suggest that monocultured WMMB235 suffers from one or more transcriptionally based bottlenecks with respect to keyicin assembly. This logic is apparent when comparing transcriptomic and proteomic profiles of WMMB235 monoculture versus WMMB235/WMMA185 (or related) systems. At the same time, delineating possible regulatory differences in mono- and cocultures is envisioned to expand our understanding of microbial combinations able to activate cryptic BGCs.

Methods

Transcriptomics. WMMA185 and WMMB235 were grown in monoculture and coculture in triplicate as described previously. [29] Aliquots of 1.5 mL were taken from day 2 and day 5 for each sample and frozen at −80 °C. At the end of the experiment, the samples were thawed and centrifuged to collect the cell mass. Cells were lysed by freezing the samples in liquid nitrogen and mechanically breaking them in a mortar and pestle. The RNA was extracted using RNAeasy Plus Mini Kit according to the manufacturer's instructions (Supporting Information) and sent to UW-Madison Biotech Center for sequencing, quality control, and read mapping. RNAs used to generate transcriptomics data originated from mono- and coculture fermentations whose metabolomics analyses reproducibly adhered to expectation. Briefly, the rRNA was depleted

using Ribo-Zero rRNA removal kit (Epicenter), and TruSeq Total RNA v2 Illumina library was prepared. The samples were subjected to Illumina HiSeq 2500 at 1 × 100 bp read length. Extensive QC was conducted on the resultant sequencing data (Supporting Information) and showed high quality reads. The filtered RNA sequences were aligned with Bowtie 267 to contigs in the WMMB235 assembly using the end-to-end alignment options "-very-sensitive -no-discordant -no-unal". The WMMB235 assembly was annotated with Prokka68 and normalized reads per kbp of gene per million (RPKM) reads was calculated for each ORF annotated. Differential gene expression for RNA-Seq results from day 5 were analyzed using EdgeR with GLM after alignment of each of the two species, WMMB235 and WMMA185, to a "hybrid" genome created from both. Functional Kyoto Encyclopedia of Genes and Genomes (KEGG) categories were assigned to the predicted protein sequences for WMMB235 using MEGAN69 with previously described methods. [70] KEGG trees were uncollapsed two levels in MEGAN, and all assignments except for "organismal systems" and "human diseases" were exported to a csv file (with the columns "read name" and "KEGG name"). Calculated RPKM values and the MEGAN csv table were used to calculate proportions of the WMMB235 transcriptome that corresponded to each KEGG category.

Proteomics. WMMA185 was grown in 100 mL culture in triplicate in ASW-D media for a period of 5 days. The content of each of these culture flasks was then vacuum filtered through 0.2 μm PES filters (Thermo Scientific Nalgene Rapid-Flow Sterile Disposable Filter Units) and transferred to three new flasks. These were inoculated with WMMB235 and incubated in a shaker. Aliquots after 5 and 8 days of culture were taken and frozen in −80 °C freezer. The cells were lysed using a lysis buffer (10 mL) containing 8 M urea (4.8048 g), 50 mM Tris Base (60.57 mg), 5 mM CaCl2 (5.5 mg), 20 mM NaCl (17.5 mg), 1 EDTA-free Roche protease inhibitor

tablet (11836170001), 1 Roche PhosSTOP phosphatase inhibitor tablet (04906845001), and 25 μL of 12.1 N HCl (to make pH ≈ 8). To 100 μL of cell lysate, 500 μL of lysis buffer was added. This was vortexed and subsequently sonicated with a probe sonicator by applying 12 15s pulses at 50% amplitude, each followed by a 30 s rest period. Care was taken to ensure the sample was kept cold. The sample then underwent trypsin digestion using 2 μg of trypsin and incubating for 18 h at 37 °C. Subsequently, the samples were labeled using Dileu (Supporting Information), following a published protocol. [46] Labeled day 5 or day 8 bacterial peptides were combined as 6-plex mixtures. The mixtures were purified by strong cation exchange liquid chromatography (SCX LC) with a PolySULFOETHYL A column (200 mm × 2.1 mm, 5 μm, 300 Å, PolyLC, Columbia, MD). Eluates containing labeled peptides were collected with an FC-4 fraction collector (Rainin Dynamax) and dried under vacuum. Samples were then fractionated with a Kinetex C18 column (5 μm, 100 Å, Phenomenex, Torrance, CA) at pH = 10 into 8 fractions. Each fraction was dried under vacuum several times.

Peptides in each fraction was reconstituted in 0.1% formic acid (FA) and subjected to reversed phase LC-MS/MS analysis with an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) interfaced with a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific, San Jose, CA). Peptides were loaded onto a 75 μm inner diameter microcapillary column custom-packed with 15 cm of bridged ethylene hybrid C18 particles (1.7 μm, 130 Å, Waters). Labeled peptides were separated with a 120 min gradient from 3% to 30% ACN with 0.1% FA, followed by 10 min to 75% ACN, and then 10 min to 95% ACN. After that, the column was equilibrated at 3% ACN for 15 min to prepare for the next injection. Survey scans of peptide precursors from 350 to 1500 m/z were performed at a resolving power of 60000 and an AGC target of 2 × 105 with a maximum injection time of 100 ms. The top 20 intense

precursor ions were selected and subjected to the HCD fragmentation at a normalized collision energy of 27% followed by tandem MS acquisition at a resolving power of 30000 and an AGC target of 5 × 104, with a maximum injection time of 54 ms and a lower mass limit of 110 m/z. Precursors were subjected to a dynamic exclusion of 45 s with a 10 ppm mass tolerance.

Raw files were processed with PEAKS Studio (Bioinformatics Solutions Inc., Waterloo, ON, Canada). Trypsin was selected as the enzyme with the maximum two missed cleavages. Spectra were first annotated by de Novo sequencing then searched by PEAKS 7.0 against a transcriptome predicted protein database for WMB235, where the parent mass error tolerance was set to be 25.0 ppm and fragment mass tolerance was 0.3 Da. Fixed modifications included DiLeu labels on peptide N-termini and lysine residues (+145.12801 Da) and carbamidomethylation on cysteine residues (+57.02146 Da). Dynamic modifications included oxidation of methionine residues (+15.99492 Da) and deamidation of asparagine and glutamine residues (+0.98402 Da). Quantitation was performed with a reporter ion integration tolerance of 20 ppm with the peptide score threshold of 20.0. Protein quantitative ratios were calculated using unique peptides. Reporter ion ratios for protein groups were exported to Excel workbook, and Student t test was performed with biological triplicates. Proteins that had >50% fold change and $p < 0.05$ were filtered as significantly changed.

Metabolomics. WMMB235 and WMMA185 cultures grown in triplicate for transcriptomics analysis were allowed to grow for 14 days and were also used to collect aliquots of 1.5 mL for metabolomic analyses. The collected samples were processed using solid phase extraction and and analyzed using UHPLC/UV/qTOF-HRESI-MS/MS. [29, 71] Briefly, solubilized extracts in 10:1 H2O/MeOH were subjected to automated SPE using a Gilson GX-271 liquid handling system. Briefly, extracts were loaded onto EVOLUTE ABN SPE cartridges (25 mg absorbent mass, 1

mL reservoir volume; Biotage, S4 Charlotte, NC), and eluted with MeOH (500 μL) directly into an LC/MS-certified vial. LC/MS data were acquired using a Bruker MaXis ESI-qTOF mass spectrometer (Bruker, Billerica, MA) coupled with a Waters Acquity UPLC system (Waters, Milford, MA) operated by Bruker Hystar software. Chromatographic separations were achieved with a gradient of MeOH and H2O (containing 0.1% formic acid) on an RP C-18 column (Phenomenex Kinetex 2.6 μm, 2.1 mm × 100 mm; Phenomenex, Torrance, CA) at a flow rate of 0.3 mL/min. The method was as follows: 1−12 min (10%−97% MeOH in H2O) and 12−14 min (97% MeOH). Full scan mass spectra (m/z 150−1550) were measured in positive ESI mode. The mass spectrometer was operated using the following parameters: capillary, 4.5 kV; nebulizer pressure, 1.2 bar; dry gas flow, 8.0 L/min; dry gas temperature, 205 °C; scan rate, 2 Hz. Tune mix (ESI-L low concentration; Agilent, Santa Clara, CA) was introduced through a divert valve at the end of each chromatographic run for automated internal calibration. The full scan spectra were followed by MS/MS spectra acquisition at variable scan speed ranging from 0.5 to 2 Hz. CID energy varied linearly, 30, 25, and 20 eV for 500 m/z, 50, 40, and 35 eV for 1000 m/z, and 70, 50, and 45 eV for 2000 m/z for charge states of 1, 2, and 3, respectively. The precursor list was set to exclude precursor ions for 1.00 min after 3 spectra with the same precursor ion have been acquired. Bruker DataAnalysis 4.2 software was used for analysis of chromatograms and to convert MS/MS data from d files to mzXML. These files were then uploaded to the Mass Spectrometry Interactive Virtual Environment (MassIVE) server (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) and networked using the GNPS pipeline. [52] Parent ions with at least three fragments were considered in the network. A cosine similarity score of 0.7 for the fragmentation spectra was used. The resulting networks were visualized using Cytoscape 3.5.1 (www.cytoscape.org/cy3.html).[53] The network containing

parent ions representative of keyicin (m/z 805.347 and 797.35) was extracted and further analyzed. For each of the ions in the network, the AUC was calculated from the corresponding LC-MS data using the integrate method in the DataAnalysis software. These values were fed back into Cytoscape to color the nodes using continuous mapping color for each day.

AHL Induction Assay. Seed cultures of WMMB235 were grown in five portions of 10 mL of ASW-D media (Supporting Information). [3,72,73] After 3 days of culture, polypropylene square 96-deep well microplates (Enzyscreen, The Netherlands) containing 500 µL of ASW-D were inoculated with 15 µL of WMMB235, and 5 µL of AHL dissolved in DMSO at five concentrations was added to it in triplicate. Monocultures and coculture controls were also inoculated as described before. [71] The culture plates were incubated at 30 °C for 14 days and shaken at 300 rpm. Subsequently, the plates were centrifuged at 3000 rpm for 20 min (Eppendorf Centrifuge 5810R), the supernatants were transferred to a Corning Clear Polystyrene 96-Well Microplate, and the absorption at 470 nm was recorded using a BioTek Synergy microplate reader.

Associated Content

Supporting Information The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschem- bio.9b00223.

Parameters for AntiSmash and Transcriptomics studies as well as a summary of RNA concentrations yielded from RNAeasy Plus Mini Kit extractions, a listing of DiLeu reporter tags used to label each sample in the multiplex analyses, and schematics showing coculture interactions leading to kyc activation, sample quality control report for RNASeq reads, strategy

for DiLeu tagging in proteomics, MS/MS data sets and temporal changes to keyicin production (PDF)

Acknowledgments

This work was supported by funding from the University of Wisconsin Madison School of Pharmacy and from the University of Wisconsin Institute for Clinical and Translational Research funded through NIH/NCATS UL1TR000427. This work was also funded by the NIH through the administration of NIGMS Grants R01GM104192 (T.S.B.), R01GM109403 (H.E.B), and NIDDK R01DK071801 (to L.L.). We thank the Analytical Instrumentation Center at the University of Wisconsin Madison for the facilities to acquire spectroscopic data. Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531 to L.L.) and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin Madison. L.L. acknowledges a Vilas Distinguished Achievement Professorship and a Janis Apinis Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin Madison School of Pharmacy. I.M. was supported in part by a predoctoral fellowship from the American Foundation for Pharmaceutical Education. M.J.S. was supported in part by the National Institute of General Medical Sciences of the NIH under Award Number T32 GM008349 and by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1747503. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

(1) Bassler, B. L. (1999) How Bacteria Talk to Each Other: Regulation of Gene Expression by Quorum Sensing. Curr. Opin. Microbiol. 2, 582−587.

(2) Whiteley, M., Diggle, S. P., and Greenberg, E. P. (2017) Progress inandPromiseofBacterialQuorumSensingResearch.Nature551, 313−320.

(3) Welsh, M. A., and Blackwell, H. E. (2016) Chemical Probes of Quorum Sensing: From Compound Development to Biological Discovery. FEMS Microbiol. Rev. 40, 774−794.

(4) Papenfort, K., Silpe, J. E., Schramma, K. R., Cong, J. P., Seyedsayamdost, M. R., and Bassler, B. L. (2017) A Vibrio Cholerae Autoinducer−Receptor Pair That Controls Biofilm Formation. Nat. Chem. Biol. 13, 551−557.

(5)Rutherford,S.T.,andBassler,B.L.(2012)BacterialQuorum Sensing: Its Role in Virulence and Possibilities for Its Control. Cold Spring Harbor Perspect. Med. 2, a012427.

(6) Rasko, D. A., and Sperandio, V. (2010) Anti-Virulence Strategies to Combat Bacteria-Mediated Disease. Nat. Rev. Drug Discovery 9, 117−128.

(7) Njoroge, J., and Sperandio, V. (2009) Jamming Bacterial Communication: New Approaches for the Treatment of Infectious Diseases. EMBO Mol. Med. 1, 201−210.

(8) Ishida, K., Lincke, T., Behnken, S., and Hertweck, C. (2010) Induced Biosynthesis of Cryptic Polyketide Metabolites in a Burkholderia Thailandensis Quorum Sensing Mutant. J. Am. Chem. Soc. 132, 13966−13968.

(9) Mao, D., Bushin, L. B., Moon, K., Wu, Y., and Seyedsayamdost, M. R. (2017) Discovery of ScmR as a Global Regulator of Secondary Metabolism and Virulence in Burkholderia Thailandensis E264. Proc. Natl. Acad. Sci. U. S. A. 114, E2920−E2928.

(10) Majerczyk, C., Brittnacher, M., Jacobs, M., Armour, C. D., Radey, M., Schneider, E., Phattarasokul, S., Bunt, R., and Greenberg, E. P. (2014) Global Analysis of the Burkholderia Thailandensis Quorum Sensing-Controlled Regulon. J. Bacteriol. 196, 1412−1424.

(11) Schuster, M., Joseph Sexton, D., Diggle, S. P., and Peter Greenberg, E. (2013) Acyl-Homoserine Lactone Quorum Sensing: From Evolution to Application. Annu. Rev. Microbiol. 67, 43−63.

(12) Campbell, J., Lin, Q., Geske, G. D., and Blackwell, H. E. (2009) New and Unexpected Insights into the Modulation of LuxR-Type Quorum Sensing by Cyclic Dipeptides. ACS Chem. Biol. 4, 1051− 1059.

(13) Stacy, D. M., Welsh, M. A., Rather, P. N., and Blackwell, H. E. (2012) Attenuation of Quorum Sensing in the Pathogen Acinetobacter Baumannii Using Non-Native N -Acyl Homoserine Lactones. ACS Chem. Biol. 7, 1719−1728.

(14) Moore, J. D., Rossi, F. M., Welsh, M. A., Nyffeler, K. E., and Blackwell, H. E. (2015) A Comparative Analysis of Synthetic Quorum Sensing Modulators in Pseudomonas Aeruginosa : New Insights into Mechanism, Active Efflux Susceptibility, Phenotypic Response, and Next-Generation Ligand Design. J. Am. Chem. Soc. 137, 14626− 14639.

(15) Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., Pati, A., Godfrey, P. A., Koehrsen, M., Clardy, J., Birren, B. W., Takano, E., Sali, A., Linington, R. G., and Fischbach, M. A. (2014) Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. Cell 158, 412−421.

(16) Amos, G. C. A., Awakawa, T., Tuttle, R. N., Letzel, A.-C., Kim, M. C., Kudo, Y., Fenical, W., Moore, B. S., and Jensen, P. R. (2017) Comparative Transcriptomics as a Guide to

Natural Product Discovery and Biosynthetic Gene Cluster Functionality. Proc. Natl. Acad. Sci. U. S. A. 114, E11121−E11130.

(17) Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Chooi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., Hadjithomas, M., Haines, A. S., Helfrich, E. J., Hillwig, M. L., Ishida, K., Jones, A. C., Jones, C. S., Jungmann, K., Kegler, C., Kim, H. U., Kötter, P., Krug, D., Masschelein, J., Melnik, A. V., Mantovani, S. M., Monroe, E. A., Moore, M., Moss, N., Nützmann, H. W., Pan, G., Pati, A., Petras, D., Reen, F. J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N. J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A. K., Balibar, C. J., Balskus, E. P., Barona- Gomez,F.,Bechthold,A.,Bode,H.B.,Borriss,R.,Brady,S.F., Brakhage, A. A., Caffrey, P., Cheng, Y. Q., Clardy, J., Cox, R. J., De Mot, R., Donadio, S., Donia, M. S., van der Donk, W. A., Dorrestein, P. C., Doyle, S., Driessen, A. J., Ehling-Schulz, M., Entian, K. D., Fischbach, M. A., Gerwick, L., Gerwick, W. H., Gross, H., Gust, B., Hertweck, C., Höfte, M., Jensen, S. E., Ju, J., Katz, L., Kaysser, L., Klassen, J. L., Keller, N. P., Kormanec, J., Kuipers, O. P., Kuzuyama, T., Kyrpides, N. C., Kwon, H. J., Lautru, S., Lavigne, R., Lee, C. Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Mendez,C.,Metsa-Ketela,M.,Micklefield,J.,Mitchell,D.A.,Moore, B. S., Moreira, L. M., Müller, R., Neilan, B. A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M. S., Ostash, B.,

Payne, S. M., Pernodet, J. L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J. M., Salas, J. A., Schmitt, E. K., Scott, B., Seipke, R. F., Shen, B., Sherman, D. H., Sivonen, K., Smanski, M. J., Sosio, M., Stegmann, E., Süssmuth, R. D., Tahlan, K., Thomas, C. M., Tang, Y., Truman, A. W., Viaud, M., Walton, J. D., Walsh, C. T., Weber, T., van Wezel, G. P., Wilkinson, B., Willey, J.

M., Wohlleben, W., Wright, G. D., Ziemert, N., Zhang, C., Zotchev, S. B., Breitling, R., Takano, E., and Glöckner, F. O. (2015) Minimum Information about a Biosynthetic Gene Cluster. Nat. Chem. Biol. 11, 625−631.

(18) Udwary, D. W., Zeigler, L., Asolkar, R. N., Singan, V., Lapidus, A., Fenical, W., Jensen, P. R., and Moore, B. S. (2007) Genome Sequencing Reveals Complex Secondary Metabolome in the Marine Actinomycete Salinispora Tropica. Proc. Natl. Acad. Sci. U. S. A. 104, 10376−10381.

(19) Fenical, W., and Jensen, P. R. (2006) Developing a New Resource for Drug Discovery: Marine Actinomycete Bacteria. Nat. Chem. Biol. 2, 666−673.

(20) Abdelmohsen, U. R., Grkovic, T., Balasubramanian, S., Kamel, M. S., Quinn, R. J., and Hentschel, U. (2015) Elicitation of Secondary Metabolism in Actinomycetes. Biotechnol. Adv. 33, 798−811.

(21) Gross, H. (2007) Strategies to Unravel the Function of Orphan Biosynthesis Pathways: Recent Examples and Future Prospects. Appl. Microbiol. Biotechnol. 75, 267−277.

(22) Scherlach, K., and Hertweck, C. (2009) Triggering Cryptic Natural Product Biosynthesis in Microorganisms. Org. Biomol. Chem. 7, 1753−1760.

(23)Bentley,S.D.,Chater,K.F.,Cerdeño-Taŕraga,A.M.,Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J.,

and Hopwood, D. A. (2002) Complete Genome Sequence of the Model Actinomycete Streptomyces Coelicolor A3(2). Nature 417, 141−147.

(24) Rateb, M. E., Hallyburton, I., Houssen, W. E., Bull, A. T., Goodfellow, M., Santhanam, R., Jaspars, M., and Ebel, R. (2013) Induction of Diverse Secondary Metabolites in Aspergillus Fumigatus by Microbial Co-Culture. RSC Adv. 3, 14444−14450.

(25) Bertrand, S., Bohni, N., Schnee, S., Schumpp, O., Gindro, K., and Wolfender, J.-L. (2014) Metabolite Induction via Microorganism Co-Culture: A Potential Way to Enhance Chemical Diversity for Drug Discovery. Biotechnol. Adv. 32, 1180−1204.

(26) Kurosawa, K., Ghiviriga, I., Sambandan, T. G., Lessard, P. A., Barbara, J. E., Rha, C., and Sinskey, A. J. (2008) Rhodostreptomycins, Antibiotics Biosynthesized Following Horizontal Gene Transfer from Streptomyces Padanus to Rhodococcus Fascians. J. Am. Chem. Soc. 130, 1126−1127.

(27) Cueto, M., Jensen, P. R., Kauffman, C., Fenical, W., Lobkovsky, E., and Clardy, J. (2001) Pestalone, a New Antibiotic Produced by a Marine Fungus in Response to Bacterial Challenge. J. Nat. Prod. 64, 1444−1446.

(28) Pishchany, G., Mevers, E., Ndousse-Fetter, S., Horvath, D. J., Paludo, C. R., Silva-Junior, E. A., Koren, S., Skaar, E. P., Clardy, J., and Kolter, R. (2018) Amycomicin Is a Potent and Specific Antibiotic Discovered with a Targeted Interaction Screen. Proc. Natl. Acad. Sci. U. S. A. 115, 10124−10129.

(29) Adnani, N., Chevrette, M. G., Adibhatla, S. N., Zhang, F., Yu, Q., Braun, D. R., Nelson, J., Simpkins, S. W., McDonald, B. R., Myers, C. L., Piotrowski, J. S., Thompson, C. J., Currie, C. R., Li, L., Rajski, S. R., and Bugni, T. S. (2017) Coculture of Marine Invertebrate- Associated

Bacteria and Interdisciplinary Technologies Enable Biosynthesis and Discovery of a New Antibiotic, Keyicin. ACS Chem.Biol.12,3093−3102.

(30) Skinnider, M. A., Dejong, C. A., Rees, P. N., Johnston, C. W., Li, H., Webster, A. L. H., Wyatt, M. A., and Magarvey, N. A. (2015)Genomes to Natural Products PRediction Informatics for Secondary Metabolomes (PRISM). Nucleic Acids Res. 43, 9645−9662.

(31) Medema, M. H., Blin, K., Cimermancic, P., De Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., and Breitling, R. (2011) AntiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in Bacterial and Fungal Genome Sequences. Nucleic Acids Res. 39, W339−346.

(32) Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., and Weber, T. (2013) AntiSMASH 2.0–a Versatile Platform for Genome Mining of Secondary Metabolite Producers. Nucleic Acids Res. 41, W204−W212.

(33) Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., and Medema, M. H. (2015) AntiSMASH 3.0 a Comprehensive Resource for the Genome Mining of Biosynthetic Gene Clusters. Nucleic Acids Res. 43, W237−W243.

(34) Adnani, N., Braun, D. R., Mcdonald, B. R., Chevrette, M. G., Currie, C. R., and Bugni, T. S. (2017) Draft Genome Sequence of Micromonospora Sp. Strain WMMB235, a Marine Ascidian- Associated Bacterium. Genome Announce 5, e01369-16.

(35) Mandlik, A., Livny, J., Robins, W. P., Ritchie, J. M., Mekalanos, J. J., and Waldor, M. K. (2011) RNA-Seq-Based Monitoring of Infection-Linked Changes in Vibrio Cholerae Gene Expression. Cell Host Microbe 10, 165−174.

(36) Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010) edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. Bioinformatics 26, 139−140.

(37) Bibb, M. J. (2005) Regulation of Secondary Metabolism in Streptomycetes. Curr. Opin. Microbiol. 8, 208−215.

(38) Waters, C. M., and Bassler, B. L. (2005) Quorum Sensing : Communication in Bacteria. Annu. Rev. Cell Dev. Biol. 21, 319−346.

(39) Gerdt, J. P., Wittenwyler, D. M., Combs, J. B., Boursier, M. E., Brummond, J. W., Xu, H., and Blackwell, H. E. (2017) Chemical Interrogation of LuxR-Type Quorum Sensing Receptors Reveals New Insights into Receptor Selectivity and the Potential for Interspecies Bacterial Signaling. ACS Chem. Biol. 12, 2457−2464.

(40) Adnani, N., Braun, D. R., McDonald, B. R., Chevrette, M. G., Currie, C. R., and Bugni, T. S. (2016) Complete Genome Sequence of Rhodococcus Sp. Strain WMMA185, a Marine Sponge-Associated Bacterium. Genome Announc 4, e01406-16.

(41) Polkade, A. V., Mantri, S. S., Patwekar, U. J., and Jangid, K. (2016) Quorum Sensing: An Under-Explored Phenomenon in the Phylum Actinobacteria. Front. Microbiol. 7, 131.

(42) Santos, C. L., Correia-Neves, M., Moradas-Ferreira, P., and Mendes, M. V. (2012) A Walk into the LuxR Regulators of Actinobacteria: Phylogenomic Distribution and Functional Diversity. PLoS One 7, No. e46758.

(43) Venturi, V., and Ahmer, B. M. M. (2015) Editorial: LuxR Solos Are Becoming Major Players in Cell−Cell Communication in Bacteria. Front. Cell. Infect. Microbiol. 5, 89.

(44) Covaceuszach, S., Degrassi, G., Venturi, V., and Lamba, D. (2013) Structural Insights into a Novel Interkingdom Signaling Circuit by Cartography of the Ligand-Binding Sites of the Homologous Quorum Sensing LuxR-Family. Int. J. Mol. Sci. 14, 20578−20596.

(45) Higgins, S., Heeb, S., Rampioni, G., Fletcher, M. P., Williams, P.,andCamara,M.(2018)DifferentialRegulationofthePhenazine Biosynthetic Operons by Quorum Sensing in Pseudomonas Aeruginosa PAO1-N. Front. Cell. Infect. Microbiol. 8, 252.

(46) Frost, D. C., Greer, T., and Li, L. (2015) High-Resolution Enabled 12-Plex DiLeu Isobaric Tags for Quantitative Proteomics. Anal. Chem. 87, 1646−1654.

(47) Lu, W., Leimkuhler, C., Oberthür, M., Kahne, D., and Walsh, C. T. (2004) AknK Is an L-2-Deoxyfucosyltransferase in the Biosynthesis of the Anthracycline Aclacinomycin A. Biochemistry 43, 4548−4558.

(48) Garrido, L. M., Lombo, F., Baig, I., Nur-e-Alam, M., Furlan, R. L.A.,Borda,C.C.,Braña,A.,Mendez,C.,Salas,J.A.,Rohr,J.,and Padilla, G. (2006) Insights in the Glycosylation Steps during Biosynthesis of the Antitumor Anthracycline Cosmomycin: Characterization of Two Glycosyltransferase Genes. Appl. Microbiol. Biotechnol. 73, 122−131.

(49)Torkkell,S.,Kunnari,T.,Palmu,K.,Hakala,J.,Mantsala,P., and Ylihonko, K. (2000) Identification of a Cyclase Gene Dictating the C-9 Stereochemistry of Anthracyclines from Streptomyces Nogalater. Antimicrob. Agents Chemother. 44, 396−399.

(50) Torkkell, S., Kunnari, T., Palmu, K., Mae, P., Hakala, J., and Ylihonko, K. (2001) The Entire Nogalamycin Biosynthetic Gene Cluster of Streptomyces Nogalater : Characterization of a 20-Kb DNA Region and Generation of Hybrid Structures. Mol. Genet. Genomics 266, 276−288.

(51) Yoshimoto, A., Johdo, O., Takatsuki, Y., Ishikura, T., Sawa, T., Takeuchi, T., and Umezawa, H. (1984) New Anthracycline Antibiotics Obtained by Microbial Glycosidation of Beta-Isorhodo- mycinone and Alpha 2-Rhodomycinone. J. Antibiot. 37, 935−938.

(52) Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W. T., Crüsemann,M.,Boudreau,P.D.,Esquenazi,E.,Sandoval-Calderoń, M., Kersten, R. D., Pace, L. A., Quinn, R. A., Duncan, K. R., Hsu, C. C., Floros, D. J., Gavilan, R. G., Kleigrewe, K., Northen, T., Dutton, R. J., Parrot, D., Carlson, E. E., Aigle, B., Michelsen, C. F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B. T., Gerwick, L., Liaw, C. C., Yang, Y. L., Humpf, H. U., Maansson, M., Keyzers, R. A., Sims, A. C., Johnson, A. R., Sidebottom, A. M., Sedio, B. E., Klitgaard, A., Larson, C. B., Boya P, C. A., Torres- Mendoza, D., Gonzalez, D. J., Silva, D. B., Marques, L. M., Demarque, D. P., Pociute, E., O'Neill, E. C., Briand, E., Helfrich, E. J. N., Granatosky, E. A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J. J., Zeng, Y., Vorholt, J. A., Kurita, K. L., Charusanti, P., McPhail, K. L., Nielsen, K. F., Vuong, L., Elfeki, M., Traxler, M. F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R. R., Mascuch, S. J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P. G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A. M. C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B. M., Almaliti, J., Allard, P. M., Phapale, P., Nothias, L. F., Alexandrov, T., Litaudon, M., Wolfender, J. L., Kyle, J. E., Metz, T. O., Peryea, T., Nguyen, D. T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K. M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P. R., Palsson, B. O.,Pogliano,K.,Linington,R.G.,Gutieŕrez,M.,Lopes,N.P., Gerwick, W. H., Moore, B. S., Dorrestein, P. C., and Bandeira, N. (2016) Sharing and Community Curation of Mass

Spectrometry Data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. 34, 828−837.

(53) Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 13, 2498−2504.

(54) Medema, M. H., and Fischbach, M. A. (2015) Computational Approaches to Natural Product Discovery. Nat. Chem. Biol. 11, 639− 648.

(55) Chung, J.-Y., Fujii, I., Harada, S., Sankawa, U., and Ebizuka, Y. (2002) Expression, Purification, and Characterization of AknX Anthrone Oxygenase, Which Is Involved in Aklavinone Biosynthesis in Streptomyces Galilaeus. J. Bacteriol. 184, 6115−6122.

(56) Tang, G.-L., Cheng, Y.-Q., and Shen, B. (2004) Leinamycin Biosynthesis Revealing Unprecedented Architectural Complexity for a Hybrid Polyketide Synthase and Nonribosomal Peptide Synthetase. Chem. Biol. 11, 33−45.

(57) Zhang, W., Wang, L., Kong, L., Wang, T., Chu, Y., Deng, Z., and You, D. (2012) Unveiling the Post-PKS Redox Tailoring Steps in Biosynthesis of the Type II Polyketide Antitumor Antibiotic Xantholipin. Chem. Biol. 19, 422−432.

(58) Galm, U., Wang, L., Wendt-Pienkowski, E., Yang, R., Liu, W., Tao, M., Coughlin, J. M., and Shen, B. (2008) Vivo Manipulation of the Bleomycin Biosynthetic Gene Cluster in Streptomyces Verticillus ATCC15003 Revealing New Insights into Its Biosynthetic Pathway. J. Biol. Chem. 283, 28236−28245.

(59) Ogasawara, Y., and Liu, H. (2009) Biosynthetic Studies of Aziridine Formation in Azicemicins. J. Am. Chem. Soc. 131, 18066− 18068.

(60) Ge, H.-M., Huang, T., Rudolf, J. D., Lohman, J. R., Huang, S.- X., Guo, X., and Shen, B. (2014) Enediyne Polyketide Synthases Stereoselectively Reduce the β-Ketoacyl Intermediates to β-d- Hydroxyacyl Intermediates in Enediyne Core Biosynthesis. Org. Lett. 16, 3958−3961.

(61) August, P. R., Tang, L., Yoon, Y. J., Ning, S., Müller, R., Yu, T.- W., Taylor, M., Hoffmann, D., Kim, C.-G., Zhang, X., Hutchinson, C. R., and Floss, H. G. (1998) Biosynthesis of the Ansamycin Antibiotic Rifamycin: Deductions from the Molecular Analysis of the Rif Biosynthetic Gene Cluster of Amycolatopsis Mediterranei S699. Chem. Biol. 5, 69−79.

(62) Alvarez-Mico, X., Jensen, P. R., Fenical, W., and Hughes, C. C. (2013) Chlorizidine, a Cytotoxic 5 H -Pyrrolo[2,1- a]Isoindol-5- One-Containing Alkaloid from a Marine Streptomyces Sp. Org. Lett. 15, 988−991.

(63) Mantovani, S. M., and Moore, B. S. (2013) Flavin-Linked Oxidase Catalyzes Pyrrolizine Formation of Dichloropyrrole-Con- taining Polyketide Extender Unit in Chlorizidine A. J. Am. Chem. Soc. 135, 18032−18035.

(64) Kang, Q., Shen, Y., and Bai, L. (2012) Biosynthesis of 3,5- AHBA-Derived Natural Products. Nat. Prod. Rep. 29, 243−263.

(65) He, W., Lei, J., Liu, Y., and Wang, Y. (2008) The LuxR Family Members GdmRI and GdmRII Are Positive Regulators of Geldanamycin Biosynthesis in Streptomyces Hygroscopicus 17997. Arch. Microbiol. 189, 501−510.

(66) Braesel, J., Crnkovic, C. M., Kunstman, K. J., Green, S. J., Maienschein-Cline,M.,Orjala,J.,Murphy,B.T.,andEustaq́uio,A.S. (2018) Complete Genome of Micromonospora Sp. Strain B006 Reveals Biosynthetic Potential of a Lake Michigan Actinomycete. J. Nat. Prod. 81, 2057−2068.

(67) Langmead, B., and Salzberg, S. L. (2012) Fast Gapped-Read Alignment with Bowtie 2. Nat. Methods 9, 357−359.

(68) Seemann, T. (2014) Prokka: Rapid Prokaryotic Genome Annotation. Bioinformatics 30, 2068−2069.

(69) Huson, D. H., and Weber, N. (2013) Microbial Community Analysis Using MEGAN. Methods Enzymol. 531, 465−485.

(70) Miller, I. J., Vanee, N., Fong, S. S., Lim-Fong, G. E., and Kwan, J.C.(2016)LackofOvertGenomeReductionintheBryostatin- Producing Bryozoan Symbiont ′′ Candidatus Endobugula Sertula′′. Appl. Environ. Microbiol. 82, 6573−6583.

(71) Adnani, N., Vazquez-Rivera, E., Adibhatla, S., Ellis, G., Braun, D., and Bugni, T. (2015) Investigation of Interspecies Interactions within Marine Micromonosporaceae Using an Improved Co-Culture Approach. Mar. Drugs 13, 6082−6098.

(72) Styles, M. J., and Blackwell, H. E. (2018) Non-Native Autoinducer Analogs Capable of Modulating the SdiA Quorum Sensing Receptor in Salmonella Enterica Serovar Typhimurium. Beilstein J. Org. Chem. 14, 2651−2664.

(73) Eibergen, N. R., Moore, J. D., Mattmann, M. E., and Blackwell, H. E. (2015) Potent and Selective Modulation of the RhlR Quorum Sensing Receptor by Using Non-Native Ligands: An Emerging Target for Virulence Control in Pseudomonas Aeruginosa. ChemBioChem 16, 2348−2356.

Table 1. BGCs Identified Within the Mociromonospora sp. WMMB235 Genome as Annotated by PRISM[a]

BGC no.	cluster annotation	closest known homologous BGC[b]
1	keyicin	aclacinomycin(55) (72%) (BGC0000191)
2	AT-less PKS	leinamycin(56) (15%) (BGC0001101)
3	type II PKS	xantholipin(57) (16%) (BGC0000279)
4	NRPS-T1PKS	bleomycin(58) (12%) (BGC0000963)
5	NRPS-T1PKS	azicemicin(59) (13%) (BGC0000202)
6	NRPS-T1PKS	
7	enediyne	tiancimycin(60) (19%) (BGC0001378)
8	AHBA BGC	rifamycin(61) (35%) (BGC0000137)
9	T1PKS	chlorizidine A(62,63) (7%) (BGC0001172)
10	lanthipeptide	

[a]Assigned cluster numbers correlate to all subsequent tables and figures.
[b]Values in parentheses correspond to percentatge of genes similar to those in the WMMB235 embedded cluster and MiBIG number, respectively.

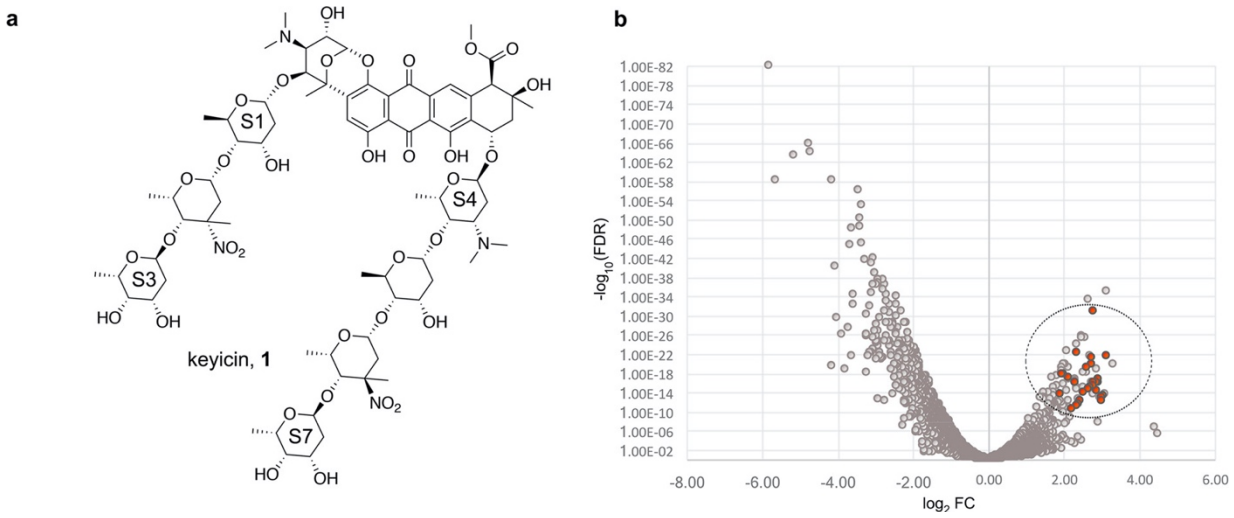

Figure 1. Structure of coculture-dependent polyketide keyicin 1 (a) and differential gene expression from WMMB235 genome in coculture with Rhodococcus sp. WMMA185 (b). Genes from the kyc gene cluster are indicated as red spheres within the circled (dashed lines) region.

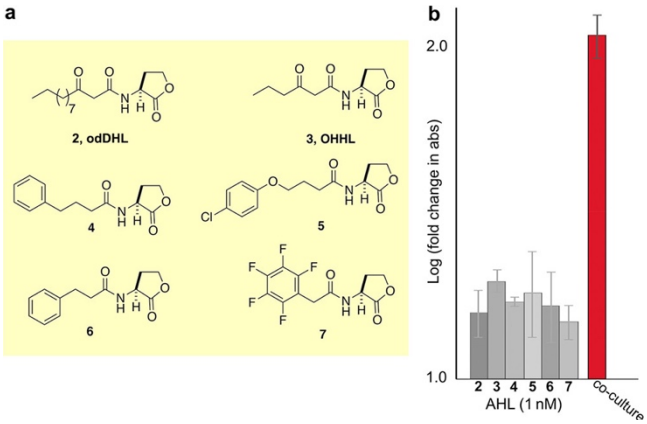

Figure 2. AHL inducers of keyicin. (a) Six out of 96 AHLs screened for kyc cluster activation and subsequent keyicin production: 2 and 3 are natural AHLs, whereas 4–7 are synthetic. (b) Increase in keyicin production shown as positive log fold change in the absorbance at 470 nm on treatment with AHLs compared to untreated monoculture; absorbance at 470 nm also enables detection of aglycone-containing precursors to 1. Coculture with WMMA185 shown as positive control (red bar). AHLs 2 and 3 are the native LuxR signals in P. aeruginosa and V. fischeri, respectively.(3)

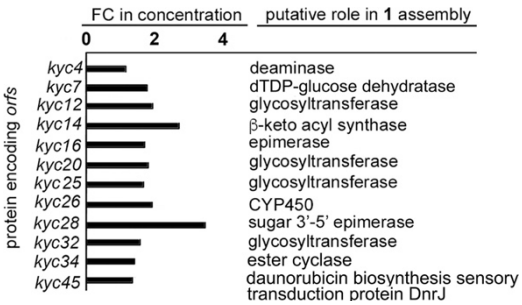

Figure 3. Summary of quantitative proteomics studies of WMMB235 fermented in WMMA185 supernatant (Rhodococcus cell free) for 8 d. FC, Fold change as compared to WMMB235 monoculture conditions. N = 3, P < 0.05.

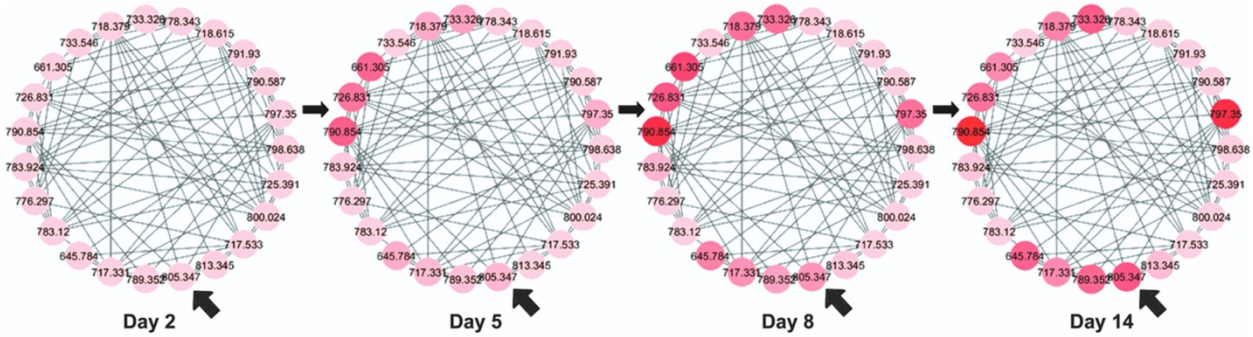

Figure 4. GNPS and Cytoscape visualization of keyicin analog masses (from LC-MS/MS of cocultured WMMB235) reflect varying extents of glycosylation over time (days 2, 5, 8, and 14). Continuous color mapping for each node in the network represents the relative concentrations of the species for which MS data is shown. Color intensities correlate to concentrations of each species for which MS data is acquired. The m/z signal for keyicin (805.347) is indicated at each time point with a thick diagonal arrow.

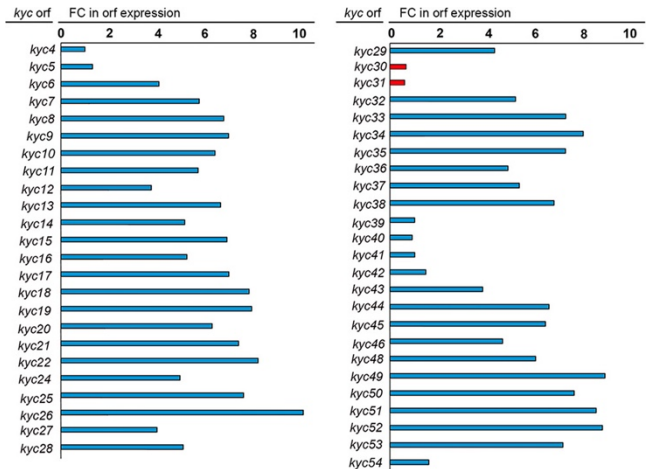

Figure 5. Summary of kyc cluster orf expression profiles in WMMB235/WMMA185 coculture compared to those generated in WMMB235 monoculture. Out of 49 orfs within the kyc cluster, only 6 undergo less than a 4-fold increase in expression and two (kyc30, 31 in red) appear to be suppressed in coculture. That these orfs appear to be dispersed at 3–4 different groupings within the kyc cluster suggests that kyc cluster regulation calls for more than just one global regulator. Beyond the earlier stated orf-to-function projections, a comprehensive listing of kyc genes and their putative roles in keyicin biosynthesis is provided in Table S3 of Supporting Information. FC, fold change. False Discovery Rate (q-value) for each gene expression change ≪ 0.01.

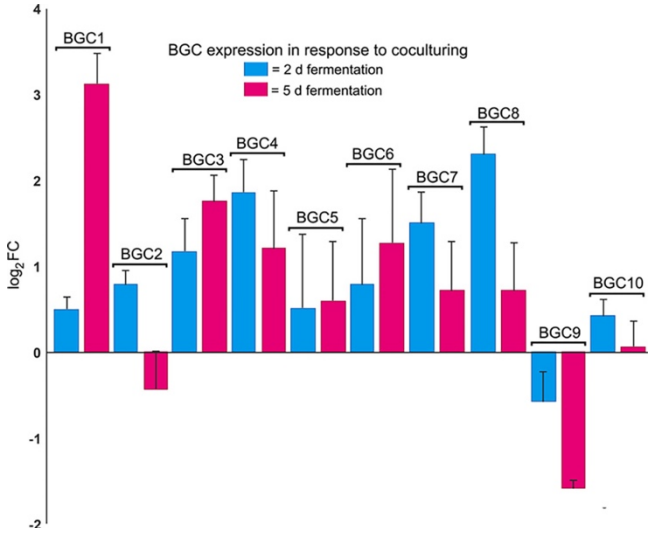

Figure 6. Global changes in BGC expression profiles in cocultured WMMB235 shown as logarithm of the fold change (FC) with base 2. The RPKMO over all the ORFs annotated by PRISM for each cluster were used to calculate the overall FCs. BGC numbers correlating to Table 1 are above each relevant bar, and expression profiles were obtained following 2 day (blue) or 5 day (purple) fermentations. N = 3.

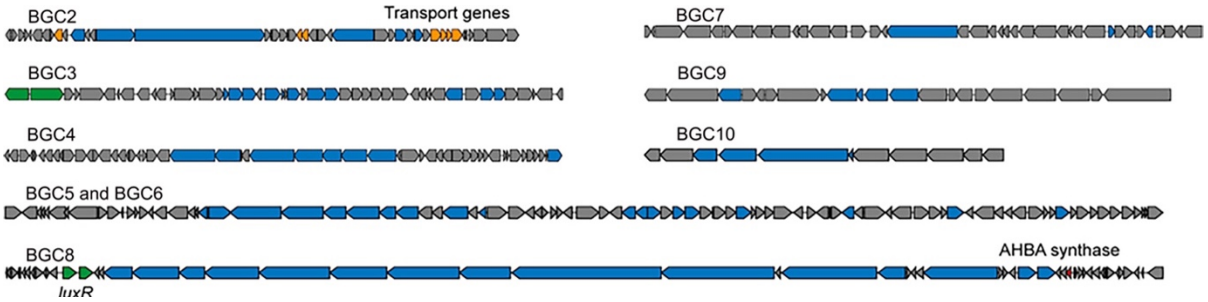

Figure 7. Early schematics of BGC2–10 (Table 1, Figure 6) from WMMB235 as annotated using both PRISM and AntiSmash. Blue ORFs indicate core biosynthetic operons. Transporter operons in BGC2 (yellow), *luxR* operons found in BGC3 and 8 (green), and AHBA synthase genes (red) in BGC8 are all highlighted.

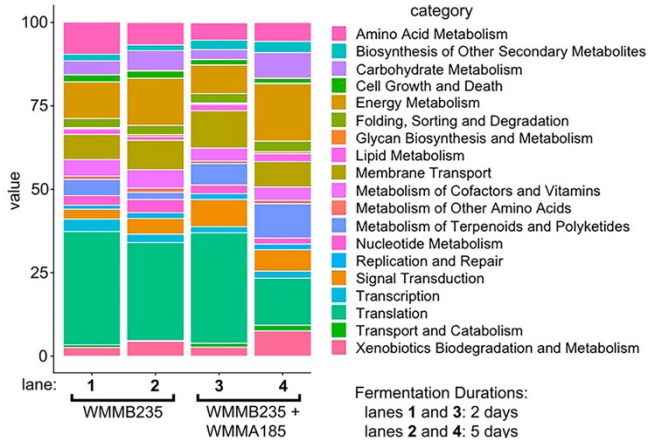

Figure 8. Summary of KEGG mapping for WMMB235 monoculture versus coculture with WMMA185. Lane contents are shown by combination of bracketing and lane coding below the categories listing. Coculturing and duration of fermentations both impact gene expression within WMMB235. Categories of function not abundant enough to depict graphically involved cell communication, cell motility, and signal molecules and interaction. All other categories are depicted in one or more of lanes 1–4.

Chapter 3

Synchronous Precursor Selection-MS3 and Chemoproteomics Assisted Compound Discovery Targeting Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9)

Contributions: This project is a collaboration between our lab and Prof. Weiping Tang's lab. Compounds were synthesized by Dr. Haibo Xie, cell samples with compound treatment, *in situ* labeling and affinity enrichment were performed by Dr. Ka Yang from Tang lab. Bottom-up proteomics and chemoproteomics sample preparation, MS-based analysis and data analysis were performed by myself in Li lab.

Abstract

Cardiovascular diseases (CVDs) have been one of the major death-causing diseases in the US and worldwide. Elevated blood cholesterol level is one of the major risk factors of atherosclerosis and CVDs. The most commonly used cholesterol lowering medication statins has their limitation of drug resistance and drug intolerance. Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9) is an emerging novel therapeutic target of CVDs that has been validated by various studies. By inhibiting PCSK9, low-density lipoprotein (LDL) Receptors can be rescued from PCSK9 binding and lysosome degradation, which will benefit cellular intake of LDL cholesterol and consequently lowering CVD risks. The existing monoclonal antibody drugs against PSCK9 has their limitations of low cross membrane ability and relatively high cost. To fulfill the gap of lacking efficient analytical method for discovering better small molecule drugs against PCSK9, mass spectrometry (MS) was used for its sensitivity, speed and ability to analyze complex drug treated cell model systems with high throughput. Isobaric labeling based LC-ESI-MS/MS quantitation was conducted to study the drug induced proteome remodeling in HepG2 cells. Synchronous precursor selection (SPS)-MS3 method was developed for DiLeu tags, improving the quantitation accuracy. Potential drug target proteins were identified using click-chemistry based photoaffnity labeling and biotin enrichment method coupled with LC-MS/MS. Overall, the MS based proteomics and chemoproteomics helped to optimize the choices and structures of the PCSK9 targeting lead compounds by more closely predicting their success or failure and identifying the ligand substrates. The study showed HX261 is a potential cost-effectice CVD treatment drug, which might benefit the vast majority of cholesterol-lowering medication consumers.

Introduction

Cardiovascular disease (CVD) is a general term for a broad range of heart and blood vessel diseases, the common forms of which include stroke, congenital cardiovascular defects, heart rhythm disorders, sudden cardiac arrest, coronary heart diseases, cardiomyopathy, peripheral artery diseases, etc.[1] CVDs have been one of the major death-causing diseases in the US and worldwide. According to the statistics, in 2013, the overall rate of death attributable to CVD was 222.9 per 100 000 Americans. In other words, more than 2200 Americans die of CVD each day, an average of 1 death every 40 seconds. Only 46.6% of adults have ideal cholesterol levels (untreated total cholesterol <200 mg/dL).[1] Many CVDs are related to a process called atherosclerosis, where the plaque that is made of cholesterol, fatty substances, cellular waste products, calcium and fibrin deposits in and clogs arteries. Therefore, elevated blood cholesterol level is one of the major risk factors of atherosclerosis and CVDs. More and more people are choosing to take cholesterol-lowering medications to reduce the risk of atherosclerotic CVD, among which statins is the most commonly used type of drugs,[2] for its ability to inhibit HMG-CoA reductase and block the pathway for synthesizing cholesterol in the liver. Along with the increasing usage, the resistance problem of statins is also coming into people's sight, where the low-density lipoprotein lipoprotein cholesterol (LDL-C) reducing effect fails to meet the expectation.[3] Various studies have shown that the proprotein convertase subtilisin/kexin type 9 (PCSK9) gene is of great interest as an emerging new therapeutic target for CVDs especially in the consideration of statin resistance.[4,5] PCSK9 is a 692-amino acid protein involved in the intracellular and extracellular regulation of LDL receptors (LDLR).[6] When binding to hepatic

LDL Receptor on the cell surface, it can reroute LDLR from endosome to lysosome and promote its degradation instead of recycling to cell surface for LDL-C intake.[5–7] Thus, strong evidence shows developing drugs against PCSK 9 target will assist statin therapy and further suppress the CVD risks. In fact, there has been several monoclonal antibody (mAb) drugs (alirocumab or SAR236553/REGN727 and evolucumab or AMG 145) in the market that were reported to reduce LDL-C level to 40-60%, which is better than mipomersen and lomitapide that target the messenger RNA coding ApoB-100 and the microsomal triglyceride transfer protein respectively.[8–10] However, mAb drugs have their own limitations such as lacking cross membrane abilities and the high cost, which makes the development of alternative small molecule drugs highly desirable. Tang lab has discovered a series of bis aromatic heterocycle ketones have high potency against PCSK9. This work aims at using mass spectrometry based quantitative proteomics as a powerful analytical method to assist the optimization of the highly potent lead compound.

MS when combined with multiplex isobaric labeling, where different experimental groups (control and compound treated) are labeled with tags that have identical precursor mass but will be distinguished at low mass region after fragmentation in a tandem mass spectrum, can achieve high throughput proteome comparison in one single LC-MS run, providing statistically reliable data that avoids run-to-run variations while saving instrument time. Besides the commercially available isobaric tags such as the tandem mass tag (TMT) and the isobaric tags for relative and absolute quantitation (iTRAQ), Li lab has developed cost-effective dimethyl amino acid based alternatives, such as dimethyl leucine (DiLeu) and dimethyl alanine (DiAla) with comparable performance.[11,12] Conventional method uses MS2 spectra for both identification and quantification, where it suffers from reporter ion ratio distortion that caused by co-isolation of

interfering precursor ions and the co-fragmentation at MS2 level. Synchronous precursor selection (SPS) MS3 strategy is to select multiple ions in MS2 spectra by ion trap excitation to undergo further fragmentation and quantified the reporter ions at MS3 level. SPS-MS3 is reported to be more accurate and reliable in TMT quantitation experiments.[13]

Quantitative proteomics can systematically validate the action of a novel chemical entity (NCE) by analyzing the regulated biological pathways with treatment. On the other hand, proteomics coupled with covalent labeling that can convert the ligand-substrate interactions into covalent interactions, facilitating the down-stream MS identification of the direct interactors of NCE *in vitro*. The chemoproteomics strategy have been found useful in mechanism-driven drug development processes. For example, Parker, Cravatt and coworkers reported a chemoproteomics platform to globally map the reversible small molecule fragment and protein interactions. They used a photoreactive group diazirine and an alkyne group to modify the drug fragments of interest, which constituted the fully functionalized fragment (FFF) probes that are able to photoaffinity label the interacted proteins upon UV radiation in vivo. After that, either a fluorescence dye tetramethylrhodamine (TAMRA)-azide or a biotin-N_3 was "clicked" onto the FFF probe labeled proteome followed by analysis via SDS-PAGE fluorescence imaging or LC-MS/MS after enrichment.[14]

In this study, SPS-MS3 method was optimized with DiLeu and DiAla tags for more accurate proteome quantitation in regards to PCSK9 targeting NCE treatment. With the proteomics assay screened most potent NCE, click chemistry based chemoproteomics strategy was developed to facilitate the identification of the protein substrates of NCE ligand. The results showed HX261 as a promising compound to inhibit PCSK9 and treat CVD. The integrated method can be adapted

to other systematical drug action and mechanism studies, potentially benefiting the rational drug design for other disease targets.

Experimental Section

Constructing the Interfered Proteome with HEK293 and Yeast.

To build an interfered proteome model for optimization of SPS-MS3 method, HEK293 cell line and yeast proteome were chosen. HEK293 cells were obtained from the American Type Culture Collection (ATCC) and lysed in the lysis buffer (8M urea, 50 mM Tris, pH=8, 5 mM $CaCl_2$, 20 mM NaCl, 1 EDTA-free Roche protease inhibitor tablet and 1 Roche PhosSTOP phosphatase inhibitor tablet) with a probe sonicator. Yeast lysate was obtained from Promega (Madison, WI). The trypsin digestion, desalting and DiLeu labeling followed previously described protocols. To be brief, protein concentrations from lysates were measured by the Bicinchoninic Acid Assay (Pierce, Rockford, IL) according to the manufacture's protocol. Proteins were reduced in 5 mM DTT for 1h at room temperature, followed by alkylation in 15 mM IAA for 30 min in the dark. After quenching the alkylation, an aliquot of proteins were digested with trypsin at 37 ºC for 18 hours, while the other will be digested with Lys-C at 37 ºC for 4 hours. Digestion was quenched by TFA and the digested peptides were then be desalted on a C18 SepPak cartridge (Waters, Milford, MA), and dried in vacuum before labeling.

1mg DiLeu or DiAla tags were mixed with DMTMM and NMM at 0.6×molar ratios to tags in anhydrous DMF. The mixtures were vortexed at room temperature for 30 min to activate the tags. After that, the digested peptides that are reconstituted in 0.5M TEAB were labeled by mixing the activated label solution at a 10:1 label to peptide mass ratio and vortexed at room

temperature for 2 h. The reaction was quenched by adding 5% NH_2OH to the final concentration of 0.25% and dried under vacuum.

To construct the labeled peptide mixture with the known ratio of isobaric tags, HEK293 peptides labeled with DiLeu 115, 116, 117 and 118 or DiAla 73, 74, 75, 76 were combined at the total peptide mass of 8 μg, 16 μg, 32 μg and 64 μg respectively, while the labeled yeasts peptides were combined at 30 μg for each isobaric tag channel. Peptides from the 2 species were then combined together with the same total peptide mass. Strong cation exchange liquid chromatography (SCX LC) was used to cleanup the excess tags with a PolySULFOETHYL A column (200mm × 2.1 mm, 5 μm, 300 Å, PolyLC, Columbia, MD). Elutes containing labeled peptides were collected with a FC-4 fraction collector (Rainin Dynamax) followed by C18 SepPak cartridge desalting, and drying under vacuum.

Optimizing SPS-MS3 Parameters for DiLeu and DiAla labeled Peptides.

Peptides were reconstituted in 0.1% Formic Acid (FA) and subjected to reversed phase LC-MS analysis with an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) interfaced with a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific). Mass spectrometry settings: Survey scans of peptide precursors from 300-1500 m/z were performed at a resolving power of 120k with an AGC target of 5×10^5 and maximum injection time of 150 ms. The top 15 precursors were selected for HCD or EThcD MS2 analysis with an isolation window of 2.0 Th, where the normalized collision energy and the ETD reaction time for each charge states were optimized at a resolving power of 15k, an AGC target of 2×10^5, a maximum injection time of 150 ms and a mass range of 150-2000 m/z. Precursors were subjected to dynamic exclusion for 45 s with a 10 ppm tolerance. In MS3 settings, top 6, 8 or 10 fragment ions in every MS2 spectrum were selected with an isolation window of 2 m/z to

undergo further HCD fragmentation with a normalized collision energy of 55 and generating MS3 spectra with a resolving power of 15 k for 4 plex or 60 k for 12 plex experiments, an AGC target of 5×10^4, a maximum injection time of 250 ms and a lower mass limit of 110 m/z for DiLeu labeling or 70 m/z for DiAla labeling were used.

Raw data were processed by Proteome Discoverer with embedded Sequest server. MS spectra were searched against the UniProt human proteome database. Static modifications included DiLeu or DiAla on lysine residues and peptide N-termini and carbamidomethylation on cysteine. Dynamic modifications included methylation and deamidation, which have been proven to be common modifications in DiLeu labeling experiments. The MS2 spectra were used for peptide identification, while the MS3 spectra in SPS-MS3 runs were used for reporter ion quantification. The median reporter ion ratios over the lightest reporter ion channel for HEK293 proteome were compared with the theoretical ratio. The group that had the closest ratios compared with the theoretical was defined as the optimal condition and the optimal parameters were used in the following proteomics assay to evaluate the drugs' potential on PSCK9 regulated proteome remodeling.

SPS-MS3 Quantitation for DiLeu and DiAla labeled Peptides from HX261 treated HepG2 Cells.

After the evaluation of SPS-MS3 experimental parameters, the proteomic regulations of HepG2 cells treated with HX261were evaluated in comparison with DMSO treated control group. Briefly, harvested HepG2 cells were homogenized, then lysed in 8M urea buffer with a probe sonicator. Tryptic peptides were prepared with the same workflow described in prior sections. Peptides from three DMSO treated groups were labeled with DiLeu 115a, 117a and 117c and the three HX261 treated groups were labeled with DiLeu 116a, 118a and 118c.

Labeled peptide mixtures were purified by SCX with a PolySULFOETHYL A column (200mm × 2.1 mm, 5 μm, 300 Å, PolyLC, Columbia, MD). Elutes containing labeled peptides were collected by a FC-4 fraction collector (Rainin Dynamax) and dried under vacuum. The cleaned samples were then fractionated with a Kinetex C18 column (5 μm, 100 Å, Phenomenex, Torrance, CA), and a binary mobile phase at pH=10 (mobile phase A is 10 mM aqueous ammonium formate and mobile phase B is 10 mM ammonium formate in 90% ACN). Gradient was set as following: 0-3 min 1% B phase; B phase linearly increased to 35% from 3-50 min, then quickly increased to 60%, 70% and 100 in 4 min, 4 min, and 2 min respectively, after which the column was washed by 100% B phase for 15 min. Eluents in the linear gradient were collected by 2-min intervals and combined into 12 fractions. Each fraction was dried under vacuum.

LC-MS/MS analysis was performed with SPS-MS3 method as described by previous section. Mass spectrometer was operated in a top 20 data-dependent acquisition mode with SPS-MS3 fragmentation. Survey scans of peptide precursors from *m/z* 350 to 1500 were performed at resolving power of 60K and AGC target of 2×10^5 with a maximum injection time of 100 ms. Tandem MS2 acquisition was at resolving power of 30K, AGC target of 1×10^4 and dynamic exclusion of 45 s of 20 ppm mass error tolerance. The top 20 intense precursor ions were selected and subjected to HCD fragmentation at a normalized collision energy of 27% and maximum injection time of 35 ms. The top 8 notches were co-selected for an SPS-MS3 scan with HCD collisional energy at 55%, resolving power at 30K, AGC target at 5×10^4 and maximum injectiontime of 54 ms.

Data were processed with Proteome Discoverer 2.1 with Sequest search as described in previous sections. Differentially regulated proteins upon drug treatment that have a p-value<0.05 from Student t test were reported for functional analysis.

Chemoproteomics with In situ Labeling of HX3-031 and Affinity Enrichment

HepG2 cells was cultured in 100mm-dish or 6-well plate to reach over 90% confluent. Culture medium was removed. Then cells were washed once with serum-free medium and treated with probe or vehicle in serum-free medium at 37°C for 40 mins. UV irradiation of 365nm was processed at 4°C for 10 mins. Then cells were washed twice with cold PBS and scraped. Cells was centrifuge at 3000xg for 3 mins and resuspend in PBS supplemented with 1% Triton-X100 and protease inhibitor cocktail (Roche, 1 tablet per 10 mL). After sonicated for 1 minute (3s-on, 3s-off and 50% Amp) at 4 °C, cells were centrifuged at 16,000xg at 4°C for 15 mins to remove cell debris and insoluble proteins. Soluble supernatant was collected and store at -80°C if needed.

Total protein concentration of lysate was adjusted to 0.5mg/mL with PBS. To 100uL lysate (50ug protein), 11uL freshly prepared "click-chemistry" mixture containing 0.1mM TBTA) (6 uL/sample, 1.7mMin 1:4 DMSO:t-BuOH), 1mM CuSO4 (2 uL/sample, 50 mM in H_2O), 12.5 mM Cy5-azide (1 uL/sample, 1.25 mM in DMSO), and 1 mM TCEP (2 uL/sample, 50 mM in H_2O) was added and gently vortexed for 1h. Resulting mixture was added 250uL cold MeOH/$CHCl_3$ (4:1) and 100uL cold PBS sequentially. Upper aqueous solution was removed after 2,000xg centrifugation for 10 mins at 4°C. Remaining organic solution was added with 400uL cold MeOH/$CHCl_3$ (4:1). Centrifugation at 2,000xg for 5 mins at 4°C was processed to pallet the protein. Removal of liquid and air-dry was followed by addition of 50uL 1X Laemmli Loading Dye. 25uL of boiled sample at 95-100°C for 5 mins was then subjected to 7.5-12% SDS-PAGE. In-gel fluorescent imaging of gel was generated by ChemiDoc MP Imaging

Systems (Bio-rad). To afford loading control image, gel was stained by Coomassie Brilliant Blue G-250 (Thermo Scientific) and destained in 50% MeOH and 10% acetic acid.

Total protein concentration of lysate was adjusted to 0.5mg/mL with PBS. To 2 mL lysate (1 mg protein), 210 uL freshly prepared "click-chemistry" mixture containing 0.1mM TBTA (120 uL/sample, 1.7mMin 1:4 DMSO:t-BuOH), 1mM CuSO4 (40 uL/sample, 50 mM in H_2O), 100 uM Biotin-PEG3-azide (10 uL/sample, 20 mM in DMSO), and 1 mM TCEP (40 uL/sample, 50 mM in H_2O) was added and gently vortexed for 1h. Resulting mixture was added 5 mL cold MeOH/$CHCl_3$ (4:1) and 2 mL cold PBS sequentially. Supernatant aqueous solution was removed after 2,000xg centrifugation for 5 mins at 4°C. Remaining organic solution was added with 10 mL cold MeOH/$CHCl_3$ (4:1) and then the liquid mixture was gently removed followed by addition of 10 mL cold MeOH/$CHCl_3$ (4:1). Centrifugation at 2,000xg for 10 mins at 4°C was processed to pallet the protein. Removal of liquid and third addition of 10 mL cold MeOH/$CHCl_3$ (4:1) was followed by 1-minute sonication (3s-on, 3s-off and 50% Amp) at 4 °C. Centrifugation at 2,000xg for 5 mins at 4°C was processed to pallet the protein again. Removal of liquid and air-dry was followed by resuspension in 400 uL 6M Urea (in PBS) contain 8 uL 10% SDS (in PBS) by 2-minute sonication (3s-on, 3s-off and 50% Amp). The resuspended solution was centrifuged at 2,000xg for 5 mins at room temperature to remove insoluble proteins. Supernatant was collected and add 40 uL of neutralized TCEP (1:1 mixture of 200 mM TCEP in PBS and 600 mM K_2CO_3). After 30 minutes incubation at 37 °C, the solution was add 56 uL 400 mM Iodoacetamide (in PBS) incubated for 30 minutes at room temperature protected from light. Resulting solution was then added 100 ul 10% SDS (in PBS) and 4.4 mL PBS. Diluted solution was added 100 uL Streptavidin beads slury. The streptavidin beads were collected by centrifugation (2,000xg, 5 min) and sequentially washed with 5 mL 0.2% SDS in DPBS once, 5

mL detergent-free PBS twice, and 5 mL LC-MS grade water twice to remove unbound protein, excess detergent, and small molecules. Finally, beads were transferred to 1.5 mL microtube(s) by 2X500 uL LC-MS grade water. To remove water, centrifugation (2,000xg, 5 min) was processed and wet beads were yielded and stored at -80 °C.

0.9 μL 1% ProteaseMax surfactant solution was added to 60 μL Streptavidin beads slurry. Bounded proteins were reduced in 5 mM dithiothreitol (DTT) at room temperature for 1h, followed by alkylation in 15 mM iodoacetamide (IAA) for 30 min in the dark. Alkylation was quenched by adding DTT to 5mM. Proteins were digested with trypsin (Promega, Madison, WI) at 1:100 (approximately) enzyme to protein ratio at 37 ºC for 18 hours. Digestion was quenched by adding trifluoroacetic acid (TFA) to a final concentration of 0.3%. The supernatant of digest was collected. The beads were then washed 3 times by 40 μL 10% FA, where the supernatants were collected and pooled. The pooled digest was desalted with Agilent OMIX C18 tips. Peptides were dried under vacuum.

On-beads digested peptides were reconstituted in 0.1% formic acid (FA) and subjected to reversed phase LC-MS/MS analysis with a Q Exactive mass spectrometer (Thermo Fisher Scientific, San Jose, CA) interfaced with a Waters nano-Acquity UPLC system (Milford, MA). Peptides were loaded onto a 75 μm inner diameter microcapillary column custom-packed with 15 cm of BEH C18 particles (1.7 μm, 130 Å, Waters). Peptides were separated with a 90 min gradient from 3% to 30% ACN with 0.1% FA, followed by 10 min with 75% ACN and then 10 min with 95% ACN. After that, the column was equilibrated at 3% ACN for 15 min to prepare for the next injection. The mass spectrometer was operated in a top-15 data-dependent acquisition mode. Survey scans of peptide precursors from m/z 300 to 1800 were performed at a resolving power of 70K and an AGC target of 2×10^5 with a maximum injection time of 100

ms. The top 15 intense precursor ions were selected and subjected to the stepped HCD fragmentation at a normalized collision energy of 27, 30, 33% followed by tandem MS acquisition at a resolving power of 17.5K and an AGC target of 5×10^4, with a maximum injection time of 200 ms and a lower mass limit of m/z 110. Precursors were subjected to a dynamic exclusion of 30s with a 10 ppm mass tolerance.

Raw files were processed with MaxQuant. Spectra were searched against the Uniprot *Homo sapiens* reviewed database with trypsin as the enzyme and maximum two missed cleavages. The parent mass error tolerance was set to be 20 ppm and fragment mass tolerance was 0.02 Da. Fixed modification was set as carbamidomethylation on cysteine residues (+57.02146 Da). Dynamic modification was set as oxidation of methionine residues (+15.99492 Da). Identifications were filtered to 1% peptide and protein FDR. Label free quantification was performed, and Student t test was performed on the LFQ intensities. The pulled down proteins that were upregulated in HX3-037 treated group were considered potential interactors of the compound.

Results and Discussions

Optimization of SPS-MS3 for DiLeu and DiAla labeling.

In order to adapt the SPS-MS3 method to dimethyl amino acid based isobaric tags DiLeu and DiAla, the model system of HEK293 proteome interfered by yeast proteome was made by mixing labeled samples with desired ratios. Labeled HEK293 digest aliquots were mixed at 1:2:4:8 ratio for DiLeu 115:116:117:118 channels and for DiAla 73:74:75/76 channels respectively. Labeled yeast digest aliquots were mixed at 1:1:1:1 for both DiLeu and DiAla

channels. Those two samples were then mixed at a 1/1 w/w ratio for SPS-MS3 and MS2 quantitation comparison. Different protease digestion choices including trypsin or Lys C, and different fragmentation strategies for MS2 including HCD and EThcD were investigated. Notch numbers of 6, 8 and 10 for SPS were compared. Fig. 1 shows the shows the distribution of experimental reporter ion ratios to the lightest reporter ion channels (115 m/z for DiLeu and 73 m/z for DiAla).

Generally, SPS-MS3 strategy was verified to greatly improve quantitation. The median reporter ion ratios of the HEK293 proteome is closer to the theoretical ratio 2:4:8 with SPS-MS3 method in comparison to MS2 method across all of the experimental conditions for DiLeu labeling and HCD fragmentation condition for DiAla labeling. Taken Lys-C digested and DiLeu labeled HEK peptide as an example, regular EThcD MS2 approach generated a distorted median ratio at 1:1.5:2.5:4.9, whereas with EThcD-HCD 10-notch SPSMS3, the ratio is restored to 1:1.8:3.2:6.4. It is also worth pointing out that the deficiency of EThcD fragmentation MS2 quantitation group for DiAla labeled samples is due to the low mass cut off in the ion trap where the first stage ETD fragmentation occurs. Ions that have m/z lower than 84 could not be restored and be analyzed by orbitrap mass analyzer. This finding, however, provides us another solution-SPS MS3, to couple the novel hybrid fragmentation technique with isobaric tags that have reporter ions m/z less than the lower mass limit of ion trap, which might benefit the quantitative study of protein post translational modifications (PTMs) such as phosphorylation and glycosylation.

Various notch numbers (6, 8 and 10) were evaluated to maintain the peptide signal intensity as well as the quantitation accuracy. As we expected, when increasing the number of precursors being synchronously selected for MS3 fragmentation, the ion intensity slightly increases, while

the quantitation accuracy departs further from the theoretical ratio because the signals from co-isolated ions also increase. From the aspect of protease choice, Lys-C outperformed trypsin across all experimental conditions in terms of both quantitation accuracy and reporter ion intensity, because Lys-C can cleave proteins on the C-terminal side of lysine residues, producing digested peptides that carry tags on both N and C termini. Thus, almost every b-, y- or c-, z-fragment ions in the MS2 spectra will carry tags that are suitable for SPS MS3 quantitation.

Table 1. and Table 2 summarized the proteome coverage of different optimization conditions of SPS-MS3 for DiLeu and DiAla. DiAla labeling yielded significantly decreased number of identifiable and quantifiable proteins. LysC digestion slightly outperforms trypsin digestion in most of the conditions except for the DiLeu labeled MS2 HCD groups.

Taking both quantification accuracy and proteome coverage into consideration, trypsin digestion, DiLeu labeling, MS2 HCD fragmentation and SPS-MS3 with 8 precursor selection notches were chosen as the optimal parameters for the follow-up experiments.

Quantitative Proteomics Study of the Action of HX261.

Tang lab has developed series of small molecule compounds that have **di**fluoro**b**enzo**i**midazolyl**i**ndolyl**k**etone (DFBIIK) structures to inhibit PSCK9. Fig. 2A shows the structure of two of the DFBIIK drug HX261 and HX235. In vitro ELISA assay showed HX261 has a sub-nanomolar IC_{50} at 0.15 nM against PCSK9. To further investigate the action of HX261 at systematic proteomics level, we applied the optimized SPS-MS3 strategy and performed 6-plex quantitative proteomics comparison of HX261 and DMSO treated HegG2 cells. Fig. 2B shows the volcano plot of the proteomics quantification result. Among the 6383 identified proteins, were significantly regulated upon HX261 treatment. The results were filtered after Student t test with the criteria of $p<0.05$ and $|\log2 \text{ fold change}|>1$. Interestingly, protein

regulation was shown to be closely related to aryl hydrocarbon receptor (AhR) pathway. For example, cytochrome P450 (CYP1A1) is among one of the significantly upregulated proteins. CYP1A1 is also known as aryl hydrocarbon hydroxylase (AHH), which plays important roles in metabolic degradation of natural AhR ligands. In the parallel transcriptomics quantitation experiment, more genes targeted by AhR were upregulated significantly, including CYP1A1, ALDH3A1, AHRR, MAF, CYP1A2, NQO1 and STC2.

Chemoproteomics Idetification of the Interactors of Probe HX3-031.

A "clickable" (being able to conduct click reaction) drug probe HX3-037 was made with an alkyne modified version of HX261. Taking the advantage of the intrinsic photo reactivity of HX261 itself, the only change made for HX3-037 is adding the alkyne group to the second aromatic ring of the drug molecule. Therefore, this "clickable" drug probe HX3-037 has similar size and conformation as HX261, which will ensure they have similar potency and similar targets. Another advantage of the probe is that it can covalently label target proteins upon UV radiation at 365nm in live cells, which means the labeled protein targets can closely reflect the inhibition action of drugs in the native physiological environment.

Fig. 3 illustrates the workflow of probing drug targets, with in situ labeling, affinity pull-down, on-beads digestion and LC-MS analysis. Label free quantification combined with statistical analysis facilitated identification of potential interactors of HX3-037that were pull-down by streptavidin beads. In Fig. 4, significantly enriched proteins from pulled-down MS analysis were highlighted. The most enriched proteins included scavenger receptor class B member 1 (SCARB1), carboxylesterase (CES1 and CES2), perilipin-2 (PLIN2), microsomal triglyceride transfer protein large subunit (MTTP), etc. Those proteins are involved in fatty acid and cholesterol metabolism pathways. For example, SCARB1 is a receptor different ligands such

as phospholipids, cholesterol ester, lipoproteins, phosphatidylserine and apoptotic cells. The enriched proteins are probable substrates of ligand HX261, although further experiments are required to study the link between the identified targets and the inhibition of PCSK9 expression in HepG2 cells.

Conclusions

In dimethyl amino acid tags based quantitative proteomics experiments, by applying SPS-MS3 approach, the co-isolation and co-fragmentation of interfering precursor ions was successfully reduced, which resulted in more accurate quantification than conventional MS2 approach. With the optimized method, the proteome regulation of PCSK9 targeting compound HX261 that involves AhR pathway was revealed.

The click reaction and MS based chemoproteomics strategy successfully probed proximal protein targets of probe HX3-037 during in vitro treatment. The identified targets actively involved in cholesterol and fatty acid metabolism pathways, which might be linked to PCSK9 inhibition characteristics of the compound.

Overall, we developed a new method that combines isobaric labeling and chemoproteomics for optimizing the choices and structures of the lead compound against PCSK9 by more closely predicting their success or failure, and closely studying the drug activities. This method might be expanded to other drug mechanism studies.

References

(1) Members, W. G.; Mozaffarian, D.; Benjamin, E. J.; Go, A. S.; Arnett, D. K.; Blaha, M. J.; Cushman, M.; Das, S. R.; Ferranti, S. de; Després, J.-P.; Fullerton, H. J.; Howard, V. J.; Huffman, M. D.; Isasi, C. R.; Jiménez, M. C.; Judd, S. E.; Kissela, B. M.; Lichtman, J. H.; Lisabeth, L. D.; Liu, S.; Mackey, R. H.; Magid, D. J.; McGuire, D. K.; Mohler, E. R.; Moy, C. S.; Muntner, P.; Mussolino, M. E.; Nasir, K.; Neumar, R. W.; Nichol, G.; Palaniappan, L.; Pandey, D. K.; Reeves, M. J.; Rodriguez, C. J.; Rosamond, W.; Sorlie, P. D.; Stein, J.; Towfighi, A.; Turan, T. N.; Virani, S. S.; Woo, D.; Yeh, R. W.; Turner, M. B.; Committee, A. H. A. S.; Subcommittee, S. S. Executive Summary. *Circulation* **2016**, *133* (4), 447–454. https://doi.org/10.1161/cir.0000000000000366.

(2) Gu, Q.; Paulose-Ram, R.; Burt, V. L.; Kit, B. K. Prescription Cholesterol-Lowering Medication Use in Adults Aged 40 and over: United States, 2003-2012. *Nchs Data Brief* **2014**, No. 177, 1–8.

(3) Reiner, Ž. Resistance and Intolerance to Statins. *Nutrition Metabolism Cardiovasc Dis* **2014**, *24* (10), 1057–1066. https://doi.org/10.1016/j.numecd.2014.05.009.

(4) Seidah, N. G.; Awan, Z.; Chrétien, M.; Mbikay, M. PCSK9. *Circ Res* **2014**, *114* (6), 1022–1036. https://doi.org/10.1161/circresaha.114.301621.

(5) Berge, K. E.; Ose, L.; Leren, T. P. Missense Mutations in the PCSK9 Gene Are Associated With Hypocholesterolemia and Possibly Increased Response to Statin Therapy. *Arteriosclerosis Thrombosis Vasc Biology* **2006**, *26* (5), 1094–1100. https://doi.org/10.1161/01.atv.0000204337.81286.1c.

(6) Gilles, L.; Barbara, S.; Benjamin, C.; P., K., John J.; Kees, H., G. The PCSK9 Decade: Thematic Review Series: New Lipid and Lipoprotein Targets for the Treatment of Cardiometabolic Diseases. *J Lipid Res* **2012**, *53* (12), 2515–2524. https://doi.org/10.1194/jlr.r026658.

(7) Dong, B.; Wu, M.; Li, H.; Kraemer, F. B.; Adeli, K.; Seidah, N. G.; Park, S. W.; Liu, J. Strong Induction of PCSK9 Gene Expression through HNF1α and SREBP2: Mechanism for the Resistance to LDL-Cholesterol Lowering Effect of Statins in Dyslipidemic Hamsters. *J Lipid Res* **2010**, *51* (6), 1486–1495. https://doi.org/10.1194/jlr.m003566.

(8) Pradeep, N.; Sekar, K. PCSK9 Inhibitors. *Cell* **2016**, *165* (5), 1037. https://doi.org/10.1016/j.cell.2016.05.016.

(9) P., G., Robert; R., D., Nihar; Payal, K.; J., R., William; Ransi, S.; Fannie, H.; Thomas, L.; Satishkumar, M.; B., H., Elaine; T., M., Shannon; E., A., Timothy; M., W., Scott; Robert, S.; S., S., Marc. Efficacy, Safety, and Tolerability of a Monoclonal Antibody to Proprotein Convertase Subtilisin/Kexin Type 9 in Combination with a Statin in Patients with Hypercholesterolaemia (LAPLACE-TIMI 57): A Randomised, Placebo-Controlled, Dose-Ranging, Phase 2 Study. *The Lancet* **2012**, *380*, 2007–2017. https://doi.org/https://doi.org/10.1016/S0140-6736(12)61770-X.

(10) J., K., Michael; P., G., Robert; J., R., Frederick; David, S.; Michael, B.; Gisle, L.; Fernando, C.; Ransi, S.; Patric, N.; Thomas, L.; Rob, S.; M., W., Scott; S., S., Marc. Efficacy and Safety of Longer-Term Administration of Evolocumab (AMG 145) in Patients With Hypercholesterolemia. *Circulation* **2014**, *129*, 234.

(11) Qing, Y.; Xudong, S.; Tyler, G.; B., L., Christopher; Craig, K., K.; Lingjun, L. Evaluation and Application of Dimethylated Amino Acids as Isobaric Tags for Quantitative Proteomics of the TGF-β/Smad3 Signaling Pathway. *J Proteome Res* **2016**, *15* (9), 3420–3431. https://doi.org/10.1021/acs.jproteome.6b00641.

(12) Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L. N , N -Dimethyl Leucines as Novel Isobaric Tandem Mass Tags for Quantitative Proteomics and Peptidomics. *Anal Chem* **2010**, *82* (7), 2817–2825. https://doi.org/10.1021/ac902778d.

(13) C., M., Graeme; P., N., David; P., J., Mark; Martin, W.; L., H.; K., E., Brian; Ramin, R.; Wilhelm, H.; P., G., Steven. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes Graeme C. McAlister, 1 David P. Nusinow, 1. *Anal Chem* **2014**, *86* (14), 7150–7158. https://doi.org/10.1021/ac502040v.

(14) G., P., C.; A., G.; Y., W.; E., C., B.; K., S.; M., J., C.; S., K., A.; L., C., C.; M., L., R.; R., J., S.; I., N.; E., S.; F., C., B. Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **2017**, *168* (3), 527-541 e29. https://doi.org/10.1016/j.cell.2016.12.029.

Tabel 1. Summary of Protein Identification and Quantification Coverage in SPS-MS3 Optimization Experiment

	Identified MS2	**Quantified MS3**	**Identified MS3_N6**	**Quantified MS3_N6**	**Identified MS3_N8**	**Quantified MS3_N8**	**Identified MS3_N10**	**Quantified MS3_N10**
DiLeu								
LysC_HCD	1561	1356	1000	902	1008	935	976	891
LysC_EThcD	938	884	429	408	395	377	427	406
Tryp_HCD	1760	1549	900	696	846	668	837	680
Tryp_EThcD	1043	994	489	450	483	456	457	438
DiAla								
LysC_HCD	1467	1238	869	735	982	883	1000	841
LysC_EThcD			713	563	718	589	701	595
Tryp_HCD	1211	1024	813	657	815	667	814	702
Tryp_EThcD			384	235	382	256	380	270

Table 2. The Ratio of Quantified Proteins to Identified Proteins

	MS2	MS3_N6	MS3_N8	MS3_N10
DiLeu				
LysC_HCD	86.9%	90.2%	92.8%	91.3%
LysC_EThcD	94.2%	95.1%	95.4%	95.1%
Tryp_HCD	88.0%	77.3%	79.0%	81.2%
Tryp_EThcD	95.3%	92.0%	94.4%	95.8%
DiAla				
LysC_HCD	84.4%	84.6%	89.9%	84.1%
LysC_EThcD		79.0%	82.0%	84.9%
Tryp_HCD	84.6%	80.8%	81.8%	86.2%
Tryp_EThcD		61.2%	67.0%	71.1%

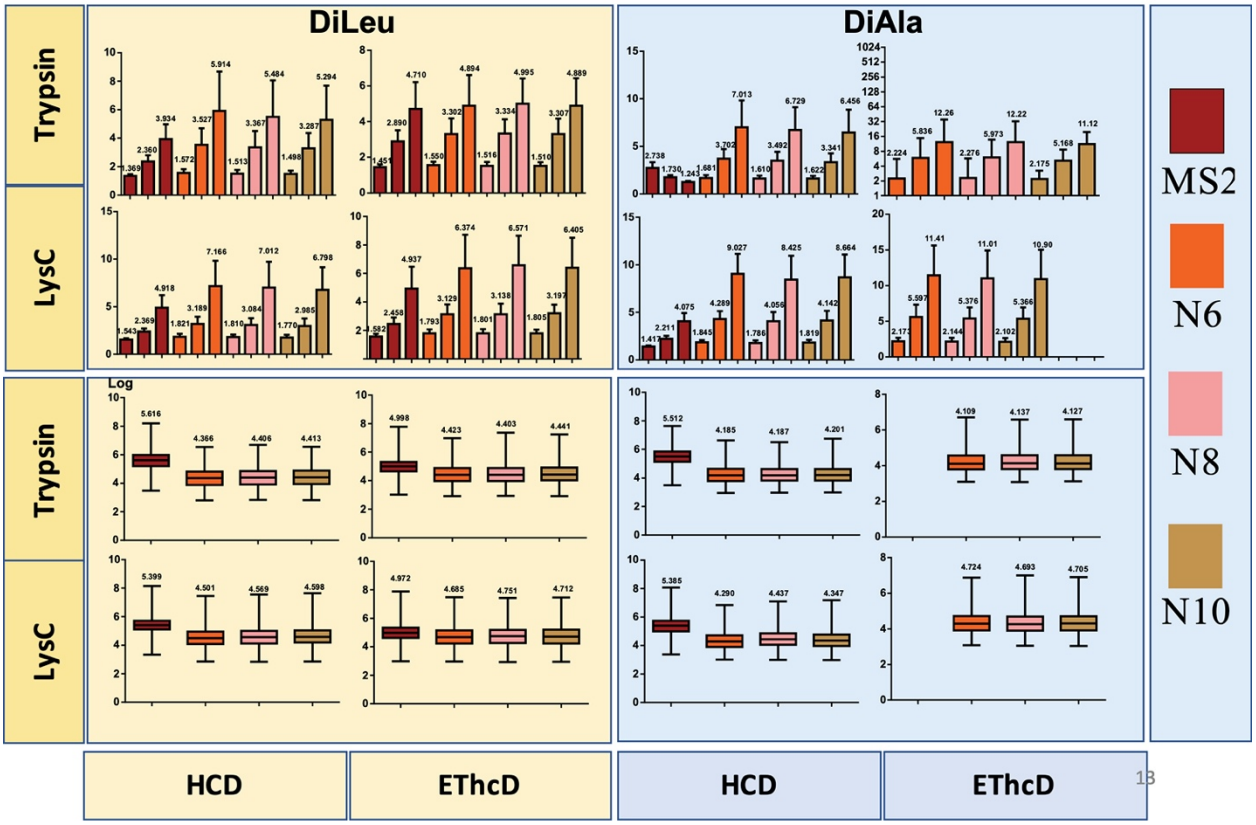

Figure 1. Relative reporter ion ratios and reporter ion intensity distributions in the SPS-MS3 optimization experiment.

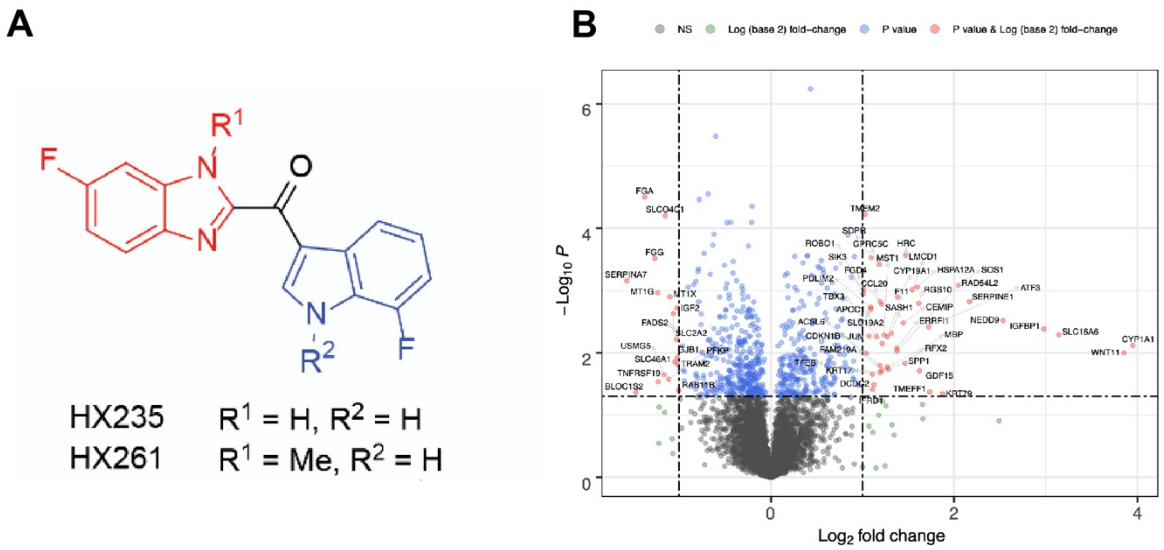

Figure 2. (A) The structures of HX261 and HX235. (B) The volcano plot of proteomics quantification result of HX261 treated HepG2 cells versus DMSO treated. Significantly regulated proteins were highlighted. (Student t test $p<0.05$, $|\log2 \text{ fold change}|>1$)

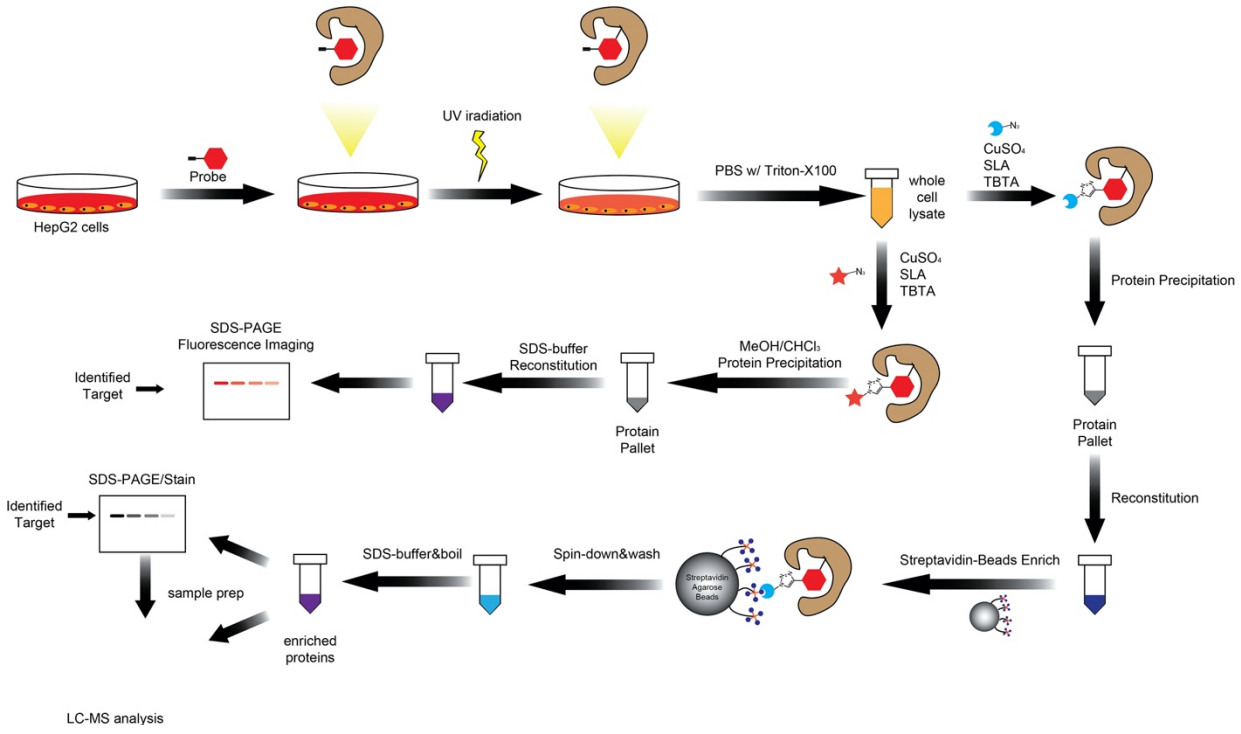

Figure 3. Workflow of chemoproteomics to pull down interactors of HX3-037 for MS analysis.

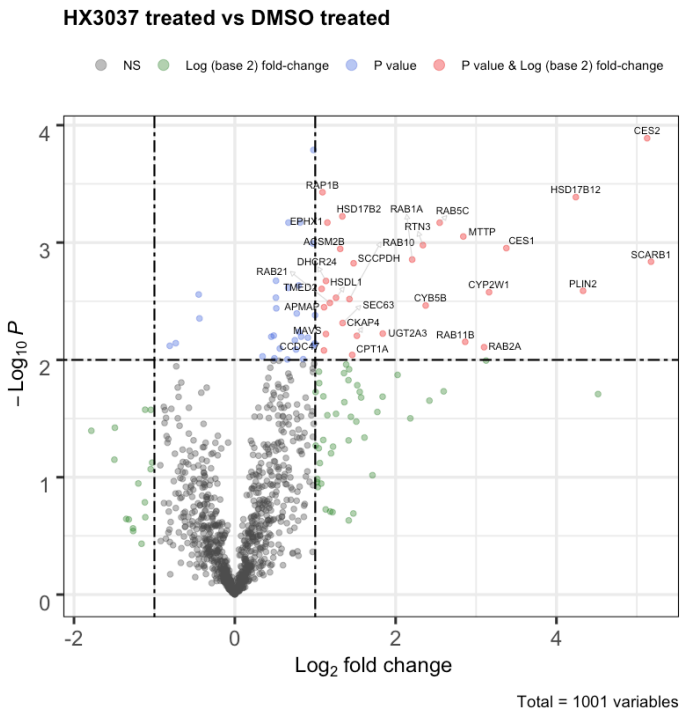

Figure 4. Volcano plot of label free quantification results of the streptavidin pulled-down proteins from HX3-037 probe and DMSO treated HepG2 cells.

Chapter 4

Acetyl-CoA Flux from the Cytosol to the ER Regulates Engagement and Quality of the Secretory Pathway

Adapted from: Dieterich, I. A.; **Cui, Y.**; Braun, M. M.; Lawton, A. J.; Robinson, N. H.; Peotter, J. L.; Yu, Q.; Casler, J. C.; Glick, B. S.; Audhya, A.; Denu, J. M.; Li, L.; Puglielli, L. Acetyl-CoA Flux from the Cytosol to the ER Regulates Engagement and Quality of the Secretory Pathway. Sci Rep-uk 2021, 11 (1), 2013. https://doi.org/10.1038/s41598-021-81447-6.
Contributions: I performed the proteomics and glycoproteomics experiments and analyzed the data.

Abstract

Nε-lysine acetylation in the ER is an essential component of the quality control machinery. ER acetylation is ensured by a membrane transporter, AT-1/SLC33A1, which translocates cytosolic acetyl-CoA into the ER lumen, and two acetyltransferases, ATase1 and ATase2, which acetylate nascent polypeptides within the ER lumen. Dysfunctional AT-1, as caused by gene mutation or duplication events, results in severe disease phenotypes. Here, we used two models of AT-1 dysregulation to investigate dynamics of the secretory pathway: AT-1 sTg, a model of systemic AT-1 overexpression, and AT-1$^{S113R/+}$, a model of AT-1 haploinsufficiency. The animals displayed reorganization of the ER, ERGIC, and Golgi apparatus. In particular, AT-1 sTg animals displayed a marked delay in Golgi-to-plasma membrane protein trafficking, significant alterations in Golgi-based N-glycan modification, and a marked expansion of the lysosomal network. Collectively our results indicate that AT-1 is essential to maintain proper organization and engagement of the secretory pathway.

Introduction

Nε-lysine acetylation in the Endoplasmic Reticulum (ER) has emerged as an essential component of the quality control (QC) machinery that maintains protein homeostasis (proteostasis) within the ER[1,2,3,4,5,6]. ER acetylation requires an ER membrane transporter, AT-1/SLC33A1, which translocates acetyl-CoA from the cytosol to the ER lumen, and two ER membrane-bound acetyltransferases, ATase1/NAT8B and ATase2/NAT8, which acetylate ER cargo proteins within the lumen of the organelle[1,3,7]. The acetylation of ER cargo nascent glycoproteins requires ATase1 and ATase2 to interact with the oligosaccharyl tranferase

complex (OST) and acetylate correctly folded nascent glycoproteins that are transiting across the ER membrane[1,2]. Studies performed with type I glycoproteins suggest that successful acetylation of the nascent polypeptide is necessary for successful engagement of the secretory pathway[2,8]. Supporting evidence comes from mouse models of increased or decreased ER acetylation[4,5,6].

Loss-of-function mutations or gene duplication events in AT-1/SLC33A1 are associated with severe disease phenotypes spanning from spastic paraplegia (heterozygous mutation) to developmental delay and premature death (homozygous mutations)[9,10,11] and intellectual disability with autistic-like traits and progeria-like dysmorphic features (gene duplication)[12,13]. These human disorders have been effectively recapitulated in mouse models. AT-1$^{S113R/S113R}$ mice, which lack AT-1 activity, die during embryogenesis, while AT-1$^{S113R/+}$ mice, a model of AT-1 haploinsufficiency, develop peripheral and central neuropathy as well as propensity to infections and cancer[4]. Mice with neuronal-specific overexpression of AT-1 (AT-1 nTg) develop an autistic-like phenotype[5] while mice with systemic overexpression (AT-1 sTg) develop a progeria-like phenotype[6].

In this study, we investigated the outcomes of dysregulated cytosol-to-ER acetyl-CoA flux within AT-1 sTg and AT-1$^{S113R/+}$ mice on the engagement and functional organization of the secretory pathway, as well as the overall quality of secreted glycoproteins (referred to as the secretome). We found that AT-1 sTg animals display enlarged cisternae within the rough ER, enlarged ER Golgi Intermediate Compartment (ERGIC), and a marked delay in protein trafficking to the cell surface, while AT-1$^{S113R/+}$ mice display contraction of the Golgi apparatus. The AT-1 sTg cellular phenotype was associated with significant alterations in Golgi-based N-glycan modification, highlighting changes in the overall quality of the secretome, together with marked expansion of the lysosomal network. Collectively our results indicate that AT-1 is

essential to maintain proper organization and engagement of the secretory pathway. Our results also highlight the possibility that the Golgi apparatus harbors a check point for QC of efficient post-translational N-glycosylation, allowing glycoproteins that fail QC to be diverted to the lysosomal compartment.

Results

Aberrant AT-1 activity affects secretory pathway-related processes

To dissect how changes in cytosol-to-ER flux of acetyl-CoA might affect the secretory pathway, we used mass spectrometry-based strategies[14] to examine the proteome and acetyl-proteome (stoichiometry of lysine acetylation) of AT-1 sTg and AT-1$^{S113R/+}$ mice. The effect of these genetic manipulations to the cytosol-to-ER flux of acetyl-CoA is reported elsewhere[4,6,14,15]. The analysis identified dynamic changes to protein abundance and acetylation across several essential and functional components of the secretory pathway (Fig. 1, Supplementary Table S1). Comparison of all proteomic datasets indicates that the changes imparted by AT-1 affect the secretory pathway more dramatically than other metabolic pathways14. Protein abundance changes found to be enriched in KEGG pathways at either level of regulation (proteome or acetyl-proteome) and in either model were used as the input for clustering. Therewithin, five clear sub-clusters emerged: transcription (spliceosome), translation initiation factors, ribosome, proteasome, and chaperones. Of the 117 proteins found to be enriched, 30 proteins showed significant changes in acetylation and 109 displayed significant changes in overall levels. Finally, 22 proteins showed changes at the level of both the proteome and the acetyl-proteome. The majority of changes imparted in acetylation stoichiometry were observed in two sub-

clusters: ribosomes, with 12 proteins, and chaperones, with 10 proteins. Overwhelmingly, 106 of the 117 secretory pathway proteins showed changes in the AT-1 sTg proteome, indicating a robust response to increased cytosol-to-ER acetyl-CoA flux. Of note, we found significant changes within the Sec61, Sec62 and Sec63 complex, which allows translocation of newly synthesized proteins into the ER[16]; Pdia3, Ero1l, Ero1lb and Calr, which ensure formation of disulfide bonds and folding of nascent glycoproteins within the calreticulin/calnexin cycle[17]; and Sec13 and Sec31, which make the outer layer of the coat that captures newly synthesized ER cargo proteins within COPII structures[18,19].

When taken together, these results indicate that acetyl-CoA homeostasis differentially impacts sub-clusters within the secretory pathway at the level of the proteome, acetyl-proteome, or both. However, in contrast to mitochondria- and lipid-related metabolic clusters[14], the secretory pathway is largely modulated at the level of the proteome.

Altered AT-1 activity results in morphological reorganization of the ER, ERGIC, and Golgi apparatus

Next, we used mouse embryonic fibroblasts (MEFs) from the above mouse models to determine whether identified proteomic changes impact the functional organization of the secretory pathway. Structure-illumination microscopy (SIM) revealed enlarged cisternae within the rough ER of AT-1 sTg MEFs (Fig. 2a). Successfully folded ER transiting proteins are captured at specialized ER structures referred to as ER exit sites (ERES) and then anterogradely transported by COPII-coated transport carriers, which emerge from transitional ER, to the ERGIC and then the Golgi apparatus[20,21,22]. To determine whether the expansion of the rough ER coincided with increased formation of ERES/COPII structures, we probed for several members

of the COPII coat, including Sec13, Sec31, Sec23, Sec24, and Sec16. However, we did not detect major differences in the number of individual puncta (Fig. 2b–f, Supplementary Fig. S1a-d), nor in the co-localization of Sec13 and Sec31 (Fig. 2g, Supplementary Fig. S1a), which reflect the formation of the outer COPII coat[19], or the co-localization Sec23 and Sec24 (Fig. 2h, Supplementary Fig. S1b), which reflect the formation of the inner COPII coat[23]. Similarly, there were no changes in co-localization of Sec31 and Sec 24 (Fig. 2i, Supplementary Fig. S1c) or Sec13 and Sec16 (Fig. 2j, Supplementary Fig. S1d), which reflect assembly of the inner coat to the outer coat as well as budding events at the transitional ER[24]. Overall, these results suggest that there are no major changes in number nor assembly of COPII cargo structures in either of the AT-1 models.

Next, we used SIM with anti-ERGIC-53 and anti-GM130 antibodies to determine structural changes at the level of the ERGIC and Golgi apparatus, respectively. The results showed a marked expansion of the ERGIC compartment in AT-1 sTg (Fig. 2k) and a significant reduction in volume and surface area of the Golgi apparatus in AT-1$^{S113R/+}$ (Fig. 2l). The above SIM findings were supported by the presence of numerous vesicles in close proximity of the ER and Golgi apparatus resembling ERGIC-structures in AT-1 sTg mice, and disorganized and somewhat smaller Golgi structures in AT-1$^{S113R/+}$ mice as visualized by electron microscopy (Fig. 2m).

Collectively, the above microscopy-based studies revealed expansion of the ER and ERGIC in AT-1 sTg and reduced Golgi apparatus in AT-1$^{S113R/+}$ mice. Overall, they support the conclusion that changes in cytosol-to-ER flux of acetyl-CoA, as determined by increased or reduced AT-1 activity, regulate the organization of the secretory pathway.

Altered AT-1 activity disrupts normal protein trafficking

To determine whether the above structural reorganization of the secretory pathway highlights changes in protein trafficking within the AT-1 models, we used a previously characterized inducible ER release system that employs tandem repeats of the conditional aggregation domain (CAD), FKBP (F_M), fused to a protein of interest[25,26]. The F_M domain dimerizes in the ER and inhibits transport out of the ER. This dimerization must be solubilized with a CAD ligand for the protein of interest to be released from the ER.

To examine the initial budding events at the ER (Fig. 3a, Exp. 1), we used 4×F_M-mCh-NL1[26], such that neuroligin bound to mCherry could be visualized as it exited the ER. Using this strategy, we did not observe differences in either cargo velocity (Fig. 3b) nor percent of cargo release (Fig. 3c) between WT and either AT-1 model. Consistent with the microscopy of COPII proteins (Fig. 2b–j, Supplementary Fig. S1a-d), these results indicate that the formation of COPII structures from transitional ER is overall preserved in both AT-1 models.

Next, to assess ER-to-Golgi trafficking (Fig. 3a, Exp. 2), we used the DsRedExpress2-FKBP(LV)-GalNAcT2-msGFP2 construct expressing a DsRed cargo protein and a GFP bound GalNAcT, which resides in the *trans*-Golgi. Prior to release, the distribution of DsRed in the ER and the GFP in the Golgi apparatus did not substantially overlap indicating successful sequestration of the two probes. By 2–4 min following CAD availability, all genotypes showed accumulation in the Golgi apparatus. However, at 18 min and beyond, the WT showed cargo dispersion from the Golgi apparatus, whereas the AT-1 sTg and AT-$1^{S113R/+}$ did not (Fig. 3d,e), highlighting a delay in exiting the Golgi apparatus across both AT-1 models.

Finally, to assess protein trafficking to the plasma membrane (Fig. 3a, Exp. 3), we used a 4xF_M-HaloTag-L1CAM construct containing a HaloTag, which is labeled with a cell

impermeable dye at the cell surface. DDS was added to initiate cargo release and cells were visualized every 10 min from 20 to 70 min post-release. In WT MEFs, the intensity of cell surface L1CAM expression steadily climbed with time (Fig. 3f,g), with the steepest slope from 20–50 min, and reaching maximum intensity from 50 to 70 min (Fig. 3f,g). By contrast, in AT-$1^{S113R/+}$ MEFs, L1CAM showed a delay in trafficking to the cell surface up to 40 min and then nearly returned to WT levels at 50–70 min (Fig. 3f,g). Most importantly, AT-1 sTg MEFs remained well below WT across the entire experimental settings (Fig. 3f,g) displaying a significant delay in L1CAM delivery to the cell surface.

The above results indicate that both AT-1 models can efficiently transport nascent glycoproteins from the ER to the Golgi apparatus but experience important alterations when transitioning through the Golgi apparatus. Particularly significant is the delay in Golgi-to-cell surface transport observed in AT-1 sTg mice. These data raise the question of whether defective post-translational processing of N-glycosylated cargo proteins within the Golgi apparatus is responsible for the defective transition and delivery.

Aberrant AT-1 models display defect in Golgi-dependent N-glycan modification and a shift in the quality of the secretome

The initial $GlcNAc_2Man_9Glc_3$ oligosaccharide structure that is added within the ER by the OST undergoes major modifications as glycoproteins move through the ER and Golgi apparatus. Specifically, the three terminal glucose are removed and the high-mannose structure is trimmed to allow final modification, which includes addition of fucose and galactose in the *cis/medial*-Golgi and sialic acid in the *trans*-Golgi and *trans*-Golgi network[27]. Importantly, the oligosaccharide chains define much of the functions and activity of glycoproteins[28,29,30].

Therefore, the complexity of the oligosaccharide chains can serve as a marker of efficient transition through the Golgi apparatus as well as a direct measure of the quality of the secretome (Supplementary Fig. S2).

In order to resolve the quality of the secretome, we developed a new integrated workflow that allows global analysis of the N-glycoproteome in the tissue. The approach uses sequential hydrophilic interaction chromatography (HILIC) for glycopeptide enrichment, coupled to electron-transfer higher-energy collision dissociation (EThcD)[31]. HILIC allows specific and enhanced enrichment of N-glycopeptides prior to LC–MS/MS analysis while EThcD allows highly confident site-specific characterization of intact N-glycopeptides by incorporating fragment ions that result from both glycan and peptide dissociation into one spectrum.

Both human[9,10,11,12,13] and mouse[4,5] data indicate that the nervous system is particularly vulnerable to changes in AT-1 activity. Furthermore, neurons respond to the overexpression of AT-1 by altering the expression of their proteome, and building more dendrites and synaptic terminals that heavily rely on the engagement of the secretory pathway[5]. Therefore, we decided to investigate the quality of the N-glycoproteome in the brain of AT-1 sTg and AT-1$^{S113R/+}$ mice.

We found significant changes in 175 and 255 glycoforms of the female (Fig. 4a) and male (Fig. 4b) AT-1 sTg cortex, and in 126 and 60 glycoforms of the female (Fig. 4c) and male (Fig. 4d) AT-1$^{S113R/+}$ cortex. A similar outcome was observed in the hippocampus with 180 and 112 glycoforms affected in female and male AT-1 sTg mice, and 179 and 214 glycoforms affected in female and male AT-1$^{S113R/+}$ mice (Supplementary Fig. S3a-d).

N-glycans were categorized based on their biological maturity, from least mature to most mature: high mannose, paucimannose, complex/hybrid, fucose, and sialic acid (Supplementary Fig. S2). Interestingly, most of the changes in glycosylation profile that we observed in both

cortex and hippocampus appeared to occur at the Golgi interface (Fig. 4e–g and Supplementary Fig. S3e-h; see also Supplementary Fig. S2). These alterations could not be attributed to immediate changes of the glycosylation machinery within the Golgi apparatus, as reflected by the mRNA levels of the fucose transporter (SLC35C1), fucosyltransferase (FUT8), sialic acid transporter (SLC35A1), sialyltransferases (ST3GAL3, ST6GAL1, ST6GAL2), galactose transporter (SLC35A2), nor the galactosyltransferases (B4GALT2, B4GALT4) (Supplementary Fig. S4).

In order to evaluate the heterogeneity of the glycoproteome, we quantified significantly changed glycans per glycosite (Fig. 4i–l). The AT-1 sTg male and AT-1$^{S113R/+}$ female showed the highest heterogeneity, with 23% and 27% (respectively) of glycosites having more than one glycan structure change significantly on a glycosite. AT-1$^{S113R/+}$ female also showed the highest proportion of multiple glycosites changing on a single protein (33% of the proteins showed more than one glycosite being affected). However, the heterogeneity of multiple glycosites on one protein being affected was ubiquitous across both models and sexes (Fig. 4m–p). These results were relatively consistent in the hippocampus (Supplementary Fig. S3i-p). Collectively, these findings illustrate the heterogeneous and complex nature of the N-glycan modifications caused by changes in AT-1 activity; multiple glycans on a glycosite are altered, as are the number of glycosites per protein.

To gain insight into glycosite specific changes, we examined glycan type distribution as they pertained to the number of sites affected per protein. AT-1 sTg female cortex displayed an even distribution of sialic acid, whereas fucose, high mannose, and complex/hybrid species showed a higher appearance on multiple sites per protein (Fig. 5a). These findings might highlight changes in specificity of individual glycosylation events imparted upon by defective transition of the

glycoproteins within the Golgi apparatus. Comparing the female cortex of AT-1 sTg to AT-$1^{S113R/+}$ mice, again the majority of fucose and complex/hybrid appear on more than one site per protein; however, sialic acid and high mannose show an even distribution between one site and more than one site per protein (Fig. 5c). AT-1 sTg male cortex shows that the majority of fucose and complex/hybrid glycans are on more than one glycosite per protein, whereas high mannose and sialic acid display an even distribution across number of sites (Fig. 5b). The very few glycan types in AT-$1^{S113R/+}$ male cortex show an even distribution of all glycan types except high mannose, which only exists on one site per protein (Fig. 5d). Overall similar results, with only minor differences, were observed in the hippocampus across models (Supplementary Fig. S5).

To understand the biological relevance of the changes observed within the entire secretome, all glycoforms were categorized into subcellular locations based on their GO Cellular Compartment terms. Across both AT-1 models and sexes, the subcellular locations which showed the greatest changes include vesicle, extracellular space, neural, and plasma membrane (Fig. 5e–h). Similar results were observed in the hippocampus (Supplementary Fig S5e,f,g,h). Albeit with some intrinsic differences, most of the significant changes observed across genetic models, brain regions, and sexes were accounted for by fucose, sialic acid and high mannose (Fig. 5e–h and Supplementary Fig S5e,f,g,h). Furthermore, across many of these analyses, we found a substantial amount of high mannose structures among proteins that are clustered within the plasma membrane and cell surface subgroups. Therefore, Golgi-dependent glycosylation events appear to be more affected and cellular compartments/pathways that depend on successful Golgi-to-plasma membrane transport appear to be more dramatically affected within our genetic models.

The adaptations addressed thus far indicate that similar pathways and subcellular locations are being affected in AT-1 sTg and AT-$1^{S113R/+}$. To determine whether there was concordance across all groups, we examined overlapping glycoforms according to their subcellular location, considering whether the glycoprotein had an increase or decrease expression when compared with WT. Interestingly, the most divergent (opposite) changes between AT-1 sTg and AT-$1^{S113R/+}$ included vesicle, extracellular, neural, and plasma membrane compartments, which, as highlighted before, depend on successful Golgi-to-plasma membrane transport (Fig. 6a). Conversely, the most convergent changes were found in the lysosome and Golgi subgroups (Fig. 6a). These differences and similarities were conserved across sex and brain region thus highlighting a specific biological response.

When analyzed for glycoprotein and glycosite, we observed 28% (female) and 17% (male) convergent changes in glycoproteins and 14% (female) and 10% (male) convergent changes in glycosites across the cortex of AT-1 sTg and AT-$1^{S113R/+}$ mice (Fig. 6b,c). A similar trend was observed in the hippocampus with a 35% (female) and 20% (male) overlap in altered glycoproteins, but only a 4% (female) and 12% (male) overlap in the glycosites affected (Fig. 6b,c). When we compared overlap between cortex with hippocampus, AT-1 sTg mice showed a 32% (female) and 19% (male) overlap in glycoproteins, and a 15% (female) and 11% (male) overlap in glycosites (Fig. 6b,c). Similarly, AT-$1^{S113R/+}$ mice showed a 24% (female) and 19% (male) overlap in glycoproteins, and a 13% (female) and 11% (male) overlap in glycosites.

Aberrant AT-1 models display expansion of the lysosomal network

Under normal conditions, a terminal mannose-6-phosphate (M6P) unit serves as a recognition signal for lysosomal-targeted enzymes. Sorting is mediated by two specific M6P receptors

(M6PRs), which bind to M6P in the Golgi apparatus and deliver lysosomal enzymes to the lysosomes[32,33]. The drastic expansion of high-mannose structures lacking a terminal glucose observed in AT-1 sTg mice can potentially disrupt normal M6PR-mediated functions resulting in mistargeting of non-lysosomal cargo glycoproteins from the Golgi apparatus to the lysosomes thus overloading the lysosomal network.

To determine whether this was indeed the case, we used EM and studied MEFs from AT-1 sTg and AT-1$^{S113R/+}$ mice. Consistent with the above hypothesis, we observed a striking expansion of the lysosomal network in the AT-1 sTg model, where the cytosol appeared loaded with electron dense organelles (Fig. 7a). To confirm the identity of these organelles, we used the lysosomal-specific probe, LysoTracker. We observed a marked increase in both number and area of lysosomes in AT-1 sTg MEFs when compared to WT MEFs, thus highlighting the expansion of the lysosomal network (Fig. 7b). Conversely, AT-1$^{S113R/+}$ MEFs displayed significantly smaller lysosomes than WT (Fig. 7b). No differences across genotypes were observed when we labeled the cells with a peroxisomal probe, used as negative control.

Finally, assessment of M6PR in AT-1 sTg MEFs revealed a significant expansion of both the cation-dependent (CD) and cation-independent (CI) networks suggesting increased cycling of CD-M6PR and CI-M6PR between the Golgi apparatus and the endosomal/lysosomal compartments (Fig. 7d,e).

Discussion

Both cell- and mouse-based studies indicate that the ER acetylation machinery is essential to maintain proteostasis within the ER. Essential features include: (1) a QC-like process where

acetylation of correctly folded glycoproteins allows them to successfully leave the ER and engage the secretory pathway; and (2) disposal of toxic protein aggregates through the induction of reticulophagy. The former process requires ATase1 and ATase2 to interact with the OST while the nascent polypeptide enters the organelle; the latter requires tight regulation of the acetylation status of ATG9A, which is essential for the engagement of FAM134B and SEC62 and the activation of the LC3β[6,34,35].

Acetyl-CoA, which serves as the donor for acetylation in the ER, is synthesized in the cytosol and must be transported across the ER membrane by AT-1. In this study, we used AT-1 sTg and AT-1$^{S113R/+}$ mice with increased and reduced AT-1 activity, respectively, to study how the secretory pathway adapts to changes in cytosol-to-ER flux of acetyl-CoA. Through the analysis of the proteome and acetyl-proteome, we determined that many proteins involved with essential ER features, such as protein biosynthesis and insertion within the ER, post-translational modification, folding, QC and ER-to-Golgi trafficking were altered. These changes were paralleled by significant reorganization of the secretory pathway. Specifically, AT-1 sTg mice displayed reorganization of the ER and expansion of the ERGIC network, while AT-1$^{S113R/+}$ mice displayed a contraction of the Golgi apparatus. Furthermore, these morphological changes were paralleled by evident changes in the trafficking of newly-synthesized glycoproteins out of the ER. Particularly significant was the delay in Golgi-to-plasma membrane transport observed in AT-1 sTg mice. Analysis of the N-glycoproteome revealed significant and widespread changes in both AT-1 sTg and AT-1$^{S113R/+}$ mice. Importantly, proteins that showed the most significant and divergent changes across genetic models were assigned to trafficking vesicles, extracellular space, neural, and plasma membrane compartments, all requiring post-translational maturation within the Golgi apparatus and successful post-Golgi transport across the secretory pathway.

Furthermore, the great majority of changes involved Golgi-specific glycan species thus highlighting a possible defect in the ability of the nascent glycoprotein to successfully transition in a progressive fashion through the N-glycan modification machinery of the Golgi apparatus. The apparent defects in Golgi-specific N-glycan modification as well as delayed Golgi-to-plasma membrane transition observed in AT-1 sTg mice was accompanied by a striking expansion of the M6PR and lysosomal networks, which likely involves mistargeting of glycoproteins with high-mannose structures lacking a terminal glucose by the M6PR.

Overexpression of AT-1 in cellular systems is accompanied by increased incorporation of O-propargyl-puromycin and azide-modified mannosamine, highlighting both increased protein biosynthesis and increased terminal sialylation[5,6]. Furthermore, neuronal overexpression of AT-1 caused a significant expansion of the proteome with 476 proteins found to be upregulated; an effect that was connected to the expansion of the dendritic and synaptic network[5]. The results described here suggest that in AT-1 sTg the expansion of the ER and the ERGIC, together with the increase in ER-to-Golgi transport of cargo material is not paralleled by compensatory changes within the Golgi apparatus to ensure post-translational N-glycosylation. It is also possible that the increased availability of acetyl-CoA within the ER, as caused by overexpression of AT-1 might affect the activity and specificity of the ATases thus resulting in incorrectly folded glycoproteins to be acetylated and forced to transition toward the Golgi apparatus. Indeed, individual acetyltransferases possess different *K*m values within the normal physiological range of acetyl-CoA, and changes or fluctuations in the concentrations of acetyl-CoA can reduce or promote the acetylation of specific lysine residues[1,36]. Furthermore, the analysis of the proteome identified significant changes within the chaperone sub-cluster, perhaps reflecting increased abundance of misfolded glycoproteins in the ER. If indeed the changes in post-translational

maturation of N-glycoproteins reflect the aberrant transition of incorrectly folded glycoproteins to the Golgi apparatus, we could argue that the delay in Golgi-to-plasma membrane transition and increased diversion of cargo material to the lysosomal compartment is part of a QC-like event within the Golgi apparatus attempting to prevent successful transport of incorrectly folded polypeptides to their final destination. This would explain why we observed minimal changes in the glycosylation status of Golgi resident proteins. Interestingly, down-regulation of the Golgi-based CMP-sialic acid and GDP-fucose transporters impedes successful Golgi-based glycosylation and prevents Golgi-to-cell surface transport of glycoconjugates[37]. This block in protein trafficking is paralled by induction of ER stress and inhibition of protein translation, suggesting that active crosstalk between the ER and the Golgi apparatus regulates dynamics of the secretory pathway[37].

Overall, these results support the conclusion that the cytosol-to-ER flux of acetyl-CoA can affect organelle dynamics across the secretory pathway with widespread consequences on Golgi-specific post-translational glycosylation and the quality of the secretome. Therefore, acetyl-CoA homeostasis, impinged upon by the ER acetylation machinery, has far reaching consequences in metabolic reprogramming across cellular compartments.

Methods

Animals

All animal experiments were carried out in accordance with the NIH Guide for the Care and Use of Laboratory Animals, and received ethical approval by the Institutional Animal Care and Use Committee of the University of Wisconsin-Madison. AT-1$^{S113R/+}$ animals were generated by

crossing mice carrying the Slc33a1-S113R mutation to WT animals[4]. AT-1$^{S113R/+}$ animals were studied between 2.5 to 3.5 months of age. Generation of AT-1 sTg mouse was achieved by crossing Rosa26:tTA mice with pTRE AT-1 mice to generate ROSA26:tTA;pTRE- AT-1 (AT-1 sTg mice)[6]. AT-1 sTg mice were studied at 2.5 to 3.5 months of age. Age-matched wild-type (WT) littermates were used as controls. Males were used for the proteome and acetyl-proteome, as described previously[14]. Males and females were used for all other experiments in this study.

Mouse embryonic fibroblast isolation

Isolation of mouse embryonic fibroblasts, MEFs, were described previously4. Briefly, embryos were collected from pregnant females at embryonic day 12.5 to 13.5. Embryos (without heads or visceral organs) were minced in sterile EDTA (0.25%; Mediatech), then incubated in a 37 °C CO_2 incubator for 30–45 min. Complete MEF media (DMEM-high glucose, 10% FBS, penicillin/streptomycin/glutamine mixture, fungizone) was used to quench the trypsin; tissue was further broken by gentle pipetting. Cells were spun at (1000 rpm for 5 min), and the supernatant was discarded. Cells were washed again in MEF media, then plated. Confluent cells were passaged at 1:4 dilutions by using trypsin–EDTA.

Stoichiometry of acetylation

Quantification of acetylation stoichiometry follows methods described previously, with the following modifications[14,38,39]. Liver protein samples (200 µg) from mitochondrial and cytosolic subcellular fractions were denatured in urea buffer (8 M urea (deionized), 100 mM ammonium bicarbonate (pH = 8.0), 5 mM DTT) and incubated at 60 °C for 20 min[40]. Cysteines were alkylated with 50 mM iodoacetamide, then incubated for 20 min. Two rounds of ~ 20 µmol

heavy isotopic D_6-acetic anhydride (Cambridge Isotope Laboratories) chemically acetylated the samples, which were then diluted to 2 M urea using 100 mM ammonium bicarbonate (pH = 8.0) before digestion with 1:100 trypsin at 37 °C for 4 h. Before a second digestion with 1:100 gluC, samples were diluted to 1 M urea. Chemically acetylated peptides were fractionated into 6 fractions using a Shimadzu LC-20AT HPLC system with a Phenomenex Gemini NX-C18 column (5 μm, 110Å, 150 × 2.0 mm). Data-independent acquisition (DIA) analysis was conducted by a Thermo Q-Exactive Orbitrap coupled to a Dionex Ultimate 3000 RSLC nano UPLC with a Waters Atlantic reversed phase column (100 μm × 150 mm). A spectral library containing all light and heavy acetyl-lysine feature pairs was generated to deconvolute and analyze the DIA spectra using the openly available MaxQuant (v1.6.1) software package. Spectral library samples were treated with C^{12}-acetic anhydride (Sigma), but were otherwise processed identically to the experimental samples and analyzed using data dependent acquisition (DDA) mass spectrometry analysis. A combined library was formed from DDA runs of both the mitochondrial and cytosolic fractions. Heavy acetyl fragment ion pairs were generated in silico, such that the spectral library would contain both the light (endogenous) acetylation peaks and the heavy (chemical) acetylation peaks. Spectronaut (v10) was used to process the experimental samples using the generated spectral library. The subcellular fraction experimental samples were processed separately using an in-house R script, which can be accessed through the GitHub link: [http://doi.org/10.5281/zenodo.3238525]; stoichiometry was calculated using the ratio of endogenous (light) fragment ion peak area over the total (endogenous and chemical) fragment ion peak area. All fractions were combined for downstream analysis. Proteins were filtered as significant changes if $P < 0.05$ compared to WT. The raw data, processed data, spectral library, and the analysis logs detailing the Spectronaut analyses settings have been deposited to the

ProteomeXchange Consortium via the MassIVE partner repository with the dataset identifier PXD014013 [http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD014013].

Quantitative proteomics

Quantification of the proteome was described previously[14]. Briefly, liver samples of cytosol, mitochondria, and nucleus were homogenized, then lysed (8 M urea, 50 mM Tris, pH = 8, 5 mM $CaCl_2$, 20 mM NaCl, 1 EDTA-free Roche protease inhibitor tablet and 1 Roche PhosSTOP phosphatase inhibitor tablet) with a probe sonicator. Crude lysates were centrifuged (14,000×*g*; 5 min), and supernatant protein concentrations were measured by Pierce BCA Protein Assay (Pierce, Rockford, IL). 400 μg of protein lysate was reduced in 5 mM dithiothreitol (DTT), followed by alkylation in 15 mM iodoacetamide (IAA), quenched by adding DTT to 5 mM, and diluted with Tris buffer (pH = 8) to 0.9 M urea. Proteins were digested with trypsin (Promega, Madison, WI), and quenched by adding trifluoroacetic acid (TFA) to a final concentration of 0.3% and desalted with C18 SepPak cartridges (Waters, Milford, MA). Peptides were vacuum dried and reconstituted in 0.5 M TEAB prior to labeling. Samples were assigned to two batches of 4-plex dimethylated leucine (DiLeu) tags each in biological duplicate. 4 mg of each DiLeu tags were suspended in anhydrous DMF, 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium tetrafluoroborate (DMTMM) and N-methylmorpholine (NMM) at 0.6 × molar ratios, then vortexed and centrifuged. The supernatant was mixed with 400 μg tryptic peptides for each condition. Peptides were labeled at a 10:1 label to peptide mass ratio, then vortexed and quenched by adding 5% NH_2OH to the final concentration of 0.25%. 4-plex mixtures were purified by strong cation exchange liquid chromatography (SCX LC) with a PolySULFOETHYL A column (200 mm × 2.1 mm, 5 μm, 300 Å, PolyLC, Columbia, MD). Labeled peptides were

collected cleaned, and fractionated with a Kinetex C18 column (5 μm, 100 Å, Phenomenex, Torrance, CA), and a binary mobile phase at pH = 10. Peptides were reconstituted in 0.1% formic acid (FA) and subjected to reversed phase LC–MS/MS analysis with an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) interfaced with a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific, San Jose, CA). Peptides were loaded onto a microcapillary column custom-packed with Bridged Ethylene Hybrid C18 particles (1.7 μm, 130 Å, Waters). Labeled peptide were separated with a 90 min gradient. Survey scans of peptide precursors from *m/z* 350 to 1500 were performed and an AGC target of 2×10^5 with a maximum injection time of 100 ms. The top 20 intense precursor ions were selected for HCD fragmentation. Raw files were processed with Proteome Discoverer 2.1 engine (Thermo Fisher Scientific, San Jose, CA) with Byonic search engine (Protein Metrics Inc, San Carlos, CA). Spectra were searched using the Uniprot *Mus musculus* database. DiLeu labels on peptide N-termini and lysine residues (+ 145.12801 Da) and carbamidomethylation on cysteine residues (+ 57.02146 Da) were considered fixed modifications. Identifications were filtered to 1% peptide and protein FDR. Proteome Discoverer was used for quantitation; only the PSMs that contained all reporter ion channels were considered. Reporter ion ratio values for protein groups were exported to Microsoft Excel and all fractions were combined for downstream analysis (see statistics section for processing). Significant proteins were determined using Fisher's method ($P < 0.05$). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD013736 [http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD013736].

DiLeu labeling and glycoproteomics

12-plex DiLeu[41] labeling was conducted as previously reported[14]. Briefly, dissected brain region samples of AT-1 sTg, AT-1$^{S113R/+}$ and WT mouse models were homogenized, then lysed in 8 M urea buffer with a probe sonicator. Lysate containing proteins was reduced in 5 mM dithiothreitol (DTT) at room temperature for 1 h, followed by alkylation in 15 mM iodoacetamide (IAA) for 30 min in the dark. Alkylation was quenched by adding DTT to 5 mM. The alkylated protein was diluted and then digested with trypsin (Promega, Madison, WI) at 1:50 enzyme to protein ratio at 37 °C for 18 h. Tryptic peptides were desalted with C18 SepPak cartridges (Waters, Milford, MA), dried under vacuum, and reconstituted in 0.5 M TEAB before labeling.

DiLeu tags were suspended in anhydrous DMF and combined with 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium tetrafluoroborate (DMTMM) and N-methylmorpholine (NMM) at 0.6 × molar ratios to tags. The mixture was vortexed at room temperature for 1 h. Following centrifugation, the supernatant was immediately mixed with tryptic peptides from one condition (10:1 tag to peptide w/w), and vortexed at room temperature for 2 h. The reaction was quenched by NH_2OH. Each batch of labeled peptides was combined respectively as 12-plex mixtures.

Glycopeptide enrichment

DiLeu labeled glycopeptides were enriched using in-house packed SAX-HILIC SPE tips following previously reported procedure with minor modification[42,43]. 3 mg of cotton wool was inserted into a 200 µL TopTip. SAX LP bulk material was dispersed in 1% TFA as a 10 mg/200 µL slurry and activated for 15 min under vigorous vortexing. After activation, 60 µL slurry was added to the spin-tip. Solvent was removed by centrifugation at 1200 rpm for 2 min,

after which the SAX material was packed at the top of the tip. The stationary phase was then conditioned by 300 µL 1% TFA and 300 µL loading buffer (80% ACN, 1% TFA), each repeated 3 times. 2 mg of DiLeu labeled peptides were aliquoted to 200 µg. Each aliquot was dissolved in 300 µL loading buffer and loaded onto the tips by centrifugation at 1200 rpm for 2 min; the flow through was collected and loaded again to ensure complete retention. The tips were then washed with 300 µL loading buffer for a total of 6 times, after which the four eluted fractions of 300 µL 70% ACN with 0.2% FA, 53% ACN with 0.2% FA, 30% ACN with 0.2% FA and 5% ACN with 0.2% FA were collected in four separate PE tubes. Corresponding fractions from the ten aliquots were pooled and dried under vacuum before MS analysis.

LC–MS/MS analysis of intact glycopeptides

Enriched glycopeptides in each fraction were reconstituted in 0.1% FA and subjected to reversed phase LC–MS/MS analysis with an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) interfaced with a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific, San Jose, CA). Peptides were separated on a 15 cm length, 75 µm i.d. custom-packed BEH C18 (1.7 µm, 130 Å, Waters) capillary column with an 80 min gradient from 0 to 30% ACN (0.1% FA). Mass spectrometer was operated in a top 20 data-dependent acquisition mode with HCD-product dependent-EThcD fragmentation[44]. Survey scans of peptide precursors from *m/z* 400 to 2000 were performed at resolving power of 120 K and AGC target of 4×10^5 with a maximum injection time of 150 ms. Tandem MS acquisition was at resolving power of 60 K, AGC target of 5×10^4 and dynamic exclusion of 12 s of 10 ppm mass tolerance. The top 20 intense precursor ions were selected and subjected to HCD fragmentation at a normalized collision energy of 33%. If signature oxomium ions (HexNAc 204.087 m/z,

HexNAcHex 366.140 m/z, HexNAc fragments 138.055 m/z and 168.065 m/z) of intact glycopeptides were detected by HCD survey scan, an EThcD hybrid fragmentation was triggered. ETD reaction time was set to 30, 20 or 10 ms when precursor charge states were z = 2, 3–5 or 6–7. HCD supplemental activation energy was 33%. Maximum ion injection times for HCD survey scan and EThcD scan are 125 and 250 ms.

Glycoproteome data processing

Raw files were processed with Byonic search engine (Protein Metrics Inc, San Carlos, CA) embedded within Proteome Discoverer 2.1 (Thermo Fisher Scientific, San Jose, CA). Spectra were searched against the SwissProt *Mus musculus* proteome database (August 13, 2016; 24,903 entries). Trypsin digestion missed cleavage was set < 3. The parent mass error tolerance was 10 ppm, and fragment mass tolerance was 0.01 Da. Fixed modifications were specified as carbamidomethylation (+ 57.02146 Da) on C residues and 12-plex DiLeu (+ 145.12801 Da) on peptide N-terminus and K. Dynamic modifications included oxidation of M (+ 15.99492 Da, rare1), deamidation (+ 0.984016 Da, rare1) of N or Q, and N-glycosylation (common1). Glycan modifications were specified as Byonic embedded mammalian N-glycan database (309 entries). Identifications were filtered to 1% protein FDR. Gene ontology annotation of glycoprotein and Student's t-test of glycopeptide quantitation results were performed using Perseus software[45]. Riley et al. reported useful tools to analyze large-scale site-specific glycoproteomics data[46]. Results were further processed by in-house written R scripts. Glycopeptides were exclusively categorized into five glycan type categories based on glycan composition: (1) sialic acid (containing NeuAc/NeuGc), (2) fucose (containing Fucose), (3) complex/hybrid (> 2 NeuAc), (4) high-mannose (2 NeuAc and > 5 Hex), and (5) paucimannose (2 NeuAc and < 5 Hex). The N-

glycoproteomics data have been deposited to the Proteome Xchange Consortium via the PRIDE partner repository with the data set identifier PXD019770 [http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD019770].

Quantitative post-acquisition data set analyses

The proteome and the acetyl-proteome were analyzed as described previously[14]. Cluster analysis was determined using KEGG pathways which arose during pathway analysis[47,48]. For the secretory pathway-related cluster, all proteins that were found in the following KEGG pathways in either of the AT-1 models and in either of the proteome or acetyl stoichiometry were included: protein processing in the ER, ribosome, proteasome, spliceosome, and RNA transport. These proteins were the input for a STRING analysis. Proteins with no interactions were hidden, with the minimum required interaction score set at high confidence (0.7). Active interaction sources included all sources except textmining.

To determine subcellular localization for glycoproteome data, proteins were categorized according to their Uniprot GO Cellular Component annotation. Proteins were placed inclusively, into the following categories as done previously[14]: “plasma membrane”; “other membrane” (includes GO terms with the word ‘membrane’ exclusive of plasma membrane); “neural” (includes GO terms that contain ‘axon’, ‘neuro’, or ‘myelin’); “ER” (includes any term that contains ‘endoplasmic’); “Golgi” (includes terms that contain ‘Golgi’ and do not contain ‘endoplasmic’); “cell surface” (includes any GO term that includes ‘surface’); “synapse”, “extracellular”, “vesicle”, “lysosome”, and “secreted” (include terms with their respective names); and “none listed” (includes any GO term that did not contain any of the other 11 subcellular groups used). The protein-glycan networks of significantly altered glycopeptides

($p < 0.05$) were created in R 3.6.0 using the igraph and ggnetwork library. Subcellular locations of glycopeptides were grouped from GO cellular component terms according to the same rules as in (Riley)[46]. Pairwise Euclidean distances between subcellular groups of significant glycopeptides from different quantitative experiments was calculated. Each GO group was considered as a 1707-dimension vector with each glycopeptide as a direction. The dimension vector was constructed by the union set of significantly altered glycopeptides across samples. The magnitude of the dimension is + 1 or − 1 for corresponding up- or down- regulated glycopeptides. Upset plots detailing the overlap between significantly altered glycoproteins and glycosites between datasets were created using the UpSetR package in R Studio. KEGG pathway enrichment and visualization of significant glycoproteins against *Mus Musculus* gene were performed with clusterProfiler library, Benjamini–Hochberg adjustment was performed to an FDR of 0.05[46].

Immunocytochemistry

Cultured primary MEF cells were transfected with a plasmid carrying ER3-mCherry fusion protein, KDEL, and ER signal peptide (Gift from Michael Davidson; Addgene Plasmid #55,041) or transfected with CellLight Golgi-GFP BacMam 2.0 (C10592, ThermoFisher Scientific) overnight. The coverslips were then fixed with paraformaldehyde for 10 min, (4%, 15,710, Electron Microscopy Sciences), and permeabilized with 0.1% Triton TM-X100 (Roche Applied Science) for 5 min, followed by an incubation in blocking buffer (10% BSA, 5% goat serum in PBS) for one hour. All antibodies were diluted in antibody dilution buffer (1% BSA, 5% goat serum in PBS) and stained with ERGIC-53 (E1031, Sigma Aldrich, 1:80), Sec13 (A303-980A, Bethyl Laboratories or H00006396-B02P, Novus Biologicals, 1:50), Sec16 (A300-648A, Bethyl

Laboratories. 1:50), Sec31A (sc-136233, SantaCruz Biotechnology, 1:50), Sec24C (A304-760A, Bethyl Laboratories, 1:50), Sec23 (ab99552, Abcam, 1:50), GM130 (610,823, BD Bioscience, 1:100), CI-M6PR (ab124767, Abcam, 1:200), or CD-M6PR (ab134153, Abcam, 1:50) for overnight incubation. Cells were washed three times in PBS and incubated with secondary antibodies AlexaFluor488 (A-11034, 1:1000) and AlexaFluor647 (A-21463 or A-21447, 1:1000) for one hour. Cells were washed three times in PBS; Fluoro-Gel II with DAPI (17985-50, Electron Microscopy Sciences) was used to mount the cover slips. Images were acquired using Structured Illumination Microscopy (Nikon SIM), and analyzed in Imaris imaging software (Bitplane, Oxford Instruments) using the Surfaces and Spots modules. Images of cells stained for CI-M6PR and CD-M6PR were acquired using both SIM and confocal microscopy (Nikon A1), and were analyzed using the surfaces module in Imaris imaging software. To stain for peroxisomes, CellLight Peroxisome-GFP, BacMam 2.0 (C10604, ThermoFisher) was used, and to stain for lysosomes, LysoTracker Green DND-26 (L7526, ThermoFisher) was used. Images were acquired using A1R-HD confocal microscope (Nikon) using a 20 × and 60 × water objective.

Electron microscopy

MEFs were grown on coverslips and fixed in 2.5% glutaraldehyde, 2.0% paraformaldehyde buffered in 0.1 M sodium phosphate buffer (PB) for 1 h at room temperature (RT). The samples were then rinsed 5 × 5 min in PB, and post-fixed in 1% osmium tetroxide and 1% potassium ferrocyanide in 0.1 M PB for 1 h at room temperature (RT), then rinsed in PB as before. Dehydration was performed in a graded ethanol series (35, 50, 70, 80, 90% for 5 min each step, 95% for 7 min, 100% for 3 × 7 min) at RT, and then transitioned in dry acetone 2 × 7 min at RT.

Fully dehydrated samples were infiltrated in increasing concentrations of Durcupan ACM (Sigma-Aldrich) and acetone mixtures in the following order: 20 min incubation in 25% PolyBed 812, 75% acetone, 20 min incubation in 50% PolyBed812, 50% acetone, 1 h incubation in 75% PolyBed 812, 25% acetone, and finally a 45 min incubation at 60 °C in 100% PolyBed 812. Embedding and polymerization took place in fresh Durcupan ACM for 24 h at 60 °C. The coverslips were removed with hydrofluoric acid and the samples were sectioned on a Leica EM UC6 ultramicrotome at 100 nm. The sections were collected on 300 mesh thin-bar Cu grids (EMS Hatfield, PA), and post-stained with uranyl acetate and lead citrate. The sectioned samples were viewed at 80 kV on a Philips CM120 transmission electron microscope and documented with an AMT BioSprint12 digital camera (Advanced Microscopy Techniques, Woburn, MA).

Live cell imaging

For all live cell imaging experiments, an ER inducible release system was used such that transfected MEFs expressed a fluorescently labeled protein to an aggregation domain that trapped the protein in the ER. Following a 48 h incubation period post transfection, cells were treated with a solubilizer to induce ER release. Solubilizing agents used were dependent on construct identity and are detailed in subsequent sections. To visualize early protein trafficking events from the ER, MEFs were transfected with the 4×F_M-mCh-NL1 plasmid (Gift from Matthew Kennedy). Before imaging, cells were treated with DDS (635,054, Takara) at a final concentration of 1 μM to induce release from the ER. Images were collected every 3 s for a duration of 3 min within 10 min of DDS induced solubilization with an A1R-HD confocal microscope (Nikon) using a 60 × water objective. Imaris imaging software with the Spots module was used to track vesicle movement. The maximum velocity of each cargo was tracked, and the

mean of maximum velocity was recorded for each cell. If cargo trafficked more than 2 μm in 3 min, the cargo was categorized as "released". The number of released cargo over total cargo in each cell was used to calculate percent release. To visualize trafficking from the ER to the Golgi, MEFs were transfected with the DsRedExpress2-FKBP(LV)-GalNAcT2-msGFP2 plasmid. As this aggregation domain is slightly different from the one used to measure ER budding, a different solubilizer, SLF (10007974, Cayman Chemical Co.), was added at a final concentration of 100 μM. Images were acquired with an ImageXpress Micro 4 (Molecular Devices) using a 40 × air objective every 2 min for a total of 30 min following solubilizer addition. Imaris imaging software was used to measure the colocalization of the green and red channels. Accumulation in the Golgi was quantified by measuring the sum of DsRed intensity as it overlapped with the GFP channel. To overcome cell-to-cell intensity variability, each cell was normalized to its maximum intensity over the 30 min. To visualize trafficking to the cell surface, MEFs were transfected with a plasmid containing 4×F_M-Halo-L1CAM. DDS was added to cells at 1 μM. Ten minutes prior to imaging, a cell impermeant halo dye was added to the cells at a final concentration of 100 nM. Cells were then washed three times with DMEM before imaging. DDS was added to cells at 1 μM. Full cell z-stack images were taken with a TI2 Spinning Disk Confocal Microscope (Nikon) using a 60 × oil objective every 10 min from 20 to 70 min post solubilizer addition. Fluorescence intensity was measured from max intensity projections using FIJI.

Real-time PCR

Real-time PCR (qPCR) was performed using the Roche 480 lightcycler and Sybr Green Real Time PCR Master Mix (04707416001, Roche). Expression levels were normalized against GAPDH levels, and are expressed as percent of control. PCR primers specific to each

gene are summarized in Supplementary Table S2 (Bio-Rad Laboratories; Prime PCR assay). The cycling parameters were as follows: 95 °C, 10 s; 60 °C, 20 s; and 72 °C, 30 s, for a total of 45 cycles. Primers are listed in Supplementary Table S2.

Statistics

Data analysis was performed in Graphpad Prism v 8.3.1 (GraphPad Software, Inc) and R v3.5.1. Data are expressed as mean ± standard deviation or SEM, as noted in each instance. For the proteomics, fold changes were computed within each DiLeu batch experiment. An F-test was used to test for equivalent variance among groups, and a Student's t test was performed assuming equal or unequal variance according to the results of the F-test. A final fold change was calculated by averaging the two experiments together, and the p-values of the two separate DiLeu experiments were combined using Fisher's method as implemented in the R package metap (R version 3.5.1). For the glycoproteomics, a Student's t-test was performed. For trafficking experiments, a two-way ANOVA was used with multiple comparisons reported at each time point. For all other analyses, comparison of the means was performed using a Student's t-test, or a one-way ANOVA. The following statistical significance was used: *$P < 0.05$; **$P < 0.005$; #$P < 0.0005$.

Data Availability

The proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the accession code PXD013736. The acetyl-proteomics data have been deposited to the ProteomeXchange Consortium via the MassIVE partner repository with the

accession code PXD014013. The R script that was used to process the acetyl-proteomics data has been deposited on Github with the identifier (search terms: AT1 Acetylation Stoich) (https://doi.org/10.5281/zenodo.3238525). The N-glycoproteomics data have been depositied to the Proteome Xchange Consortium via the PRIDE partner repository with the data set identifier PXD019770. The authors declare that all other data supporting the findings of this study are available within the paper and its supplementary tables.

Acknowledgements

This work was supported by the NIH (NS094154, AG053937 and AG057408 to L.P.; DK071801, AG052324, and GM108538 to L.L.; GM065386 to J.M.D; GM104010 to B.S.G.; GM134865 to A.A.); the Department of Veterans Affairs (I01 BX004202 to L.P.); and a core grant to the Waisman Center from NICHD-U54 HD090256. I.A.D. was supported by T32 AG000213; J.C.C. was supported by T32 GM007183. The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531 to L.L.) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. SIM imaging was performed at the Biochemistry Optical Core of the University of Wisconsin-Madison (Madison, WI). L.L. acknowledges a Vilas Distinguished Achievement Professorship and the Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Contributions

I.A.D., Y.C., M.M.B., A.J.L., N.H.R., J.L.P., and Q.Y. performed the experiments and analyzed the data. J.C.C. and B.S.C. provided reagents. A.A., J.M.D., L.L. and L.P. provided critical advice for the experiments. L.P. designed the overall study. L.P. and I.A.D. wrote the paper with input from all authors.

References

1. Farrugia, M. A. & Puglielli, L. Nepsilon-lysine acetylation in the endoplasmic reticulum—a novel cellular mechanism that regulates proteostasis and autophagy. *J. Cell Sci.* https://doi.org/10.1242/jcs.221747 (2018).

2. Ding, Y., Dellisanti, C. D., Ko, M. H., Czajkowski, C. & Puglielli, L. The endoplasmic reticulum-based acetyltransferases, ATase1 and ATase2, associate with the oligosaccharyl-transferase to acetylate correctly folded polypeptides. *J. Biol. Chem.* **289**, 32044–32055 (2014).

3. Jonas, M. C., Pehar, M. & Puglielli, L. AT-1 is the ER membrane acetyl-CoA transporter and is essential for cell viability. *J. Cell. Sci.* **123**, 3378–3388 (2010).

4. Peng, Y. *et al.* Deficient import of acetyl-CoA into the ER Lumen causes neurodegeneration and propensity to infections, inflammation, and cancer. *J. Neurosci.* **34**, 6772–6789 (2014).

5. Hullinger, R. *et al.* Increased expression of AT-1/SLC33A1 causes an autistic-like phenotype in mice by affecting dendritic branching and spine formation. *J. Exp. Med.* **213**, 1267–1284 (2016).

6. Peng, Y. *et al.* Increased transport of acetyl-CoA into the endoplasmic reticulum causes a progeria-like phenotype. *Aging Cell* https://doi.org/10.1111/acel.12820 (2018).

7. Ko, M. H. & Puglielli, L. Two endoplasmic reticulum (ER)/ER golgi intermediate compartment-based lysine acetyltransferases post-translationally regulate BACE1 levels. *J. Biol. Chem.* **284**, 2482–2492 (2009).

8. Mak, A. B. *et al.* Post-translational regulation of CD133 by ATase1/ATase2-mediated lysine acetylation. *J. Mol. Biol.* **426**, 2175–2182 (2014).

9. Huppke, P. *et al.* Mutations in SLC33A1 cause a lethal autosomal-recessive disorder with congenital cataracts, hearing loss, and low serum copper and ceruloplasmin. *Am. J. Hum. Genet.* **90**, 61–68 (2012).

10. Chiplunkar, S. *et al.* Huppke-Brendel syndrome in a seven months old boy with a novel 2-bp deletion in SLC33A1. *Metab Brain Dis.* **31**, 1195–1198 (2016).

11. Lin, P. *et al.* A missense mutation in SLC33A1, which encodes the acetyl-CoA transporter, causes autosomal-dominant spastic paraplegia (SPG42). *Am. J. Hum. Genet.* **83**, 752–759 (2008).

12. Sanders, S. J. *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863–885 (2011).

13. Prasad, A. *et al.* A discovery resource of rare copy number variations in individuals with autism spectrum disorder. *G3* **2**, 1665–1685 (2012).

14. Dieterich, I. A. *et al.* Acetyl-CoA flux regulates the proteome and acetyl-proteome to maintain intracellular metabolic crosstalk. *Nat. Commun.* **10**, 3929. https://doi.org/10.1038/s41467-019-11945-9 (2019).

15. Hullinger, R. & Puglielli, L. Molecular and cellular aspects of age-related cognitive decline and Alzheimer's disease. *Behav. Brain Res.* **322**, 191–205. https://doi.org/10.1016/j.bbr.2016.05.008 (2017).

16. Rapoport, T. A., Li, L. & Park, E. Structural and mechanistic insights into protein translocation. *Annu. Rev. Cell. Dev. Biol.* **33**, 369–390 (2017).

17. Feige, M. J. & Hendershot, L. M. Disulfide bonds in ER protein folding and homeostasis. *Curr. Opin. Cell Biol.* **23**, 167–175 (2011).

18. Lederkremer, G. Z. *et al.* Structure of the Sec23p/24p and Sec13p/31p complexes of COPII. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10704–10709 (2001).

19. Stagg, S. M. *et al.* Structure of the Sec13/31 COPII coat cage. *Nature* **439**, 234–238 (2006).

20. Zanetti, G., Pahuja, K. B., Studer, S., Shim, S. & Schekman, R. COPII and the regulation of protein sorting in mammals. *Nat. Cell. Biol.* **14**, 20–28 (2012).

21. Budnik, A. & Stephens, D. J. ER exit sites–localization and control of COPII vesicle formation. *FEBS Lett.* **583**, 3796–3803 (2009).

22. Peotter, J., Kasberg, W., Pustova, I. & Audhya, A. COPII-mediated trafficking at the ER/ERGIC interface. *Traffic* **20**, 491–503 (2019).

23. Bi, X., Corpina, R. A. & Goldberg, J. Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat. *Nature* **419**, 271–277 (2002).

24. Watson, P., Townley, A. K., Koka, P., Palmer, K. J. & Stephens, D. J. Sec16 defines endoplasmic reticulum exit sites and is required for secretory cargo export in mammalian cells. *Traffic* **7**, 1678–1687 (2006).

25. Rivera, V. M. *et al.* Regulation of protein secretion through controlled aggregation in the endoplasmic reticulum. *Science* **287**, 826–830 (2000).

26. Bowen, A. B., Bourke, A. M., Hiester, B. G., Hanus, C. & Kennedy, M. J. Golgi-independent secretory trafficking through recycling endosomes in neuronal dendrites and spines. *eLife* https://doi.org/10.7554/eLife.27362 (2017).

27. Hirschberg, C. B., Robbins, P. W. & Abeijon, C. Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annu. Rev. Biochem.* **67**, 49–69 (1998).

28. Millar, A. H. *et al.* The scope, functions, and dynamics of posttranslational protein modifications. *Annu. Rev. Plant Biol.* https://doi.org/10.1146/annurev-arplant-050718-100211 (2019).

29. Chang, I. J., He, M. & Lam, C. T. Congenital disorders of glycosylation. *Ann. Transl. Med.* **6**, 477. https://doi.org/10.21037/atm.2018.10.45 (2018).

30. Narimatsu, H. *et al.* Current technologies for complex glycoproteomics and their applications to biology/disease-driven glycoproteomics. *J. Proteome Res.* https://doi.org/10.1021/acs.jproteome.8b00515 (2018).

31. Yu, Q. *et al.* Electron-transfer/higher-energy collision dissociation (EThcD)-enabled intact glycopeptide/glycoproteome characterization. *J. Am. Soc. Mass Spectrom.* **28**, 1751–1764 (2017).

32. Hirst, J., Futter, C. E. & Hopkins, C. R. The kinetics of mannose 6-phosphate receptor trafficking in the endocytic pathway in HEp-2 cells: the receptor enters and rapidly leaves multivesicular endosomes without accumulating in a prelysosomal compartment. *Mol. Biol. Cell.* **9**, 809–816 (1998).

33. Wang, S. *et al.* A role of Rab29 in the integrity of the trans-Golgi network and retrograde trafficking of mannose-6-phosphate receptor. *PLoS ONE* **9**, e96242. https://doi.org/10.1371/journal.pone.0096242 (2014).

34. Pehar, M., Jonas, M. C., Hare, T. M. & Puglielli, L. SLC33A1/AT-1 protein regulates the induction of autophagy downstream of IRE1/XBP1 pathway. *J. Biol. Chem.* **287**, 29921–29930 (2012).

35. Peng, Y. *et al.* Improved proteostasis in the secretory pathway rescues Alzheimer's disease in the mouse. *Brain* **139**, 937–952 (2016).

36. Freiman, R. N. & Tjian, R. Regulating the regulators: Lysine modifications make their mark. *Cell* **112**, 11–17 (2003).

37. Xu, Y. X., Liu, L., Caffaro, C. E. & Hirschberg, C. B. Inhibition of Golgi apparatus glycosylation causes endoplasmic reticulum stress and decreased protein synthesis. *J. Biol. Chem.* **285**, 24600–24608 (2010).

38. Lindahl, A. J., Lawton, A. J., Baeza, J., Dowell, J. A. & Denu, J. M. Site-specific lysine acetylation stoichiometry across subcellular compartments. *Methods Mol. Biol.* **1983**, 79–106 (2019).

39. Baeza, J. *et al.* Revealing dynamic protein acetylation across subcellular compartments. *J. Proteome Res.* **19**, 2404–2418 (2020).

40. Baeza, J. *et al.* Stoichiometry of site-specific lysine acetylation in an entire proteome. *J. Biol. Chem.* **289**, 21326–21338 (2014).

41. Frost, D. C., Greer, T. & Li, L. High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. *Anal. Chem.* **87**, 1646–1654 (2015).

42. Cui, Y. *et al.* Finding the sweet spot in ERLIC mobile phase for simultaneous enrichment of N-glyco and phosphopeptides. *J. Am. Soc. Mass Spectrom.* **30**, 2491–2501 (2019).

43. Selman, M. H., Hemayatkar, M., Deelder, A. M. & Wuhrer, M. Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides. *Anal. Chem.* **83**, 2492–2499 (2011).

44. Yu, Q., Shi, X., Feng, Y., Kent, K. C. & Li, L. Improving data quality and preserving HCD-generated reporter ions with EThcD for isobaric tag-based quantitative proteomics and proteome-wide PTM studies. *Anal. Chim. Acta* **968**, 40–49 (2017).

45. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

46. Riley, N. M., Hebert, A. S., Westphall, M. S. & Coon, J. J. Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nat. Commun.* **10**, 1311. https://doi.org/10.1038/s41467-019-09222-w (2019).

47. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).

48. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkaa970 (2020).

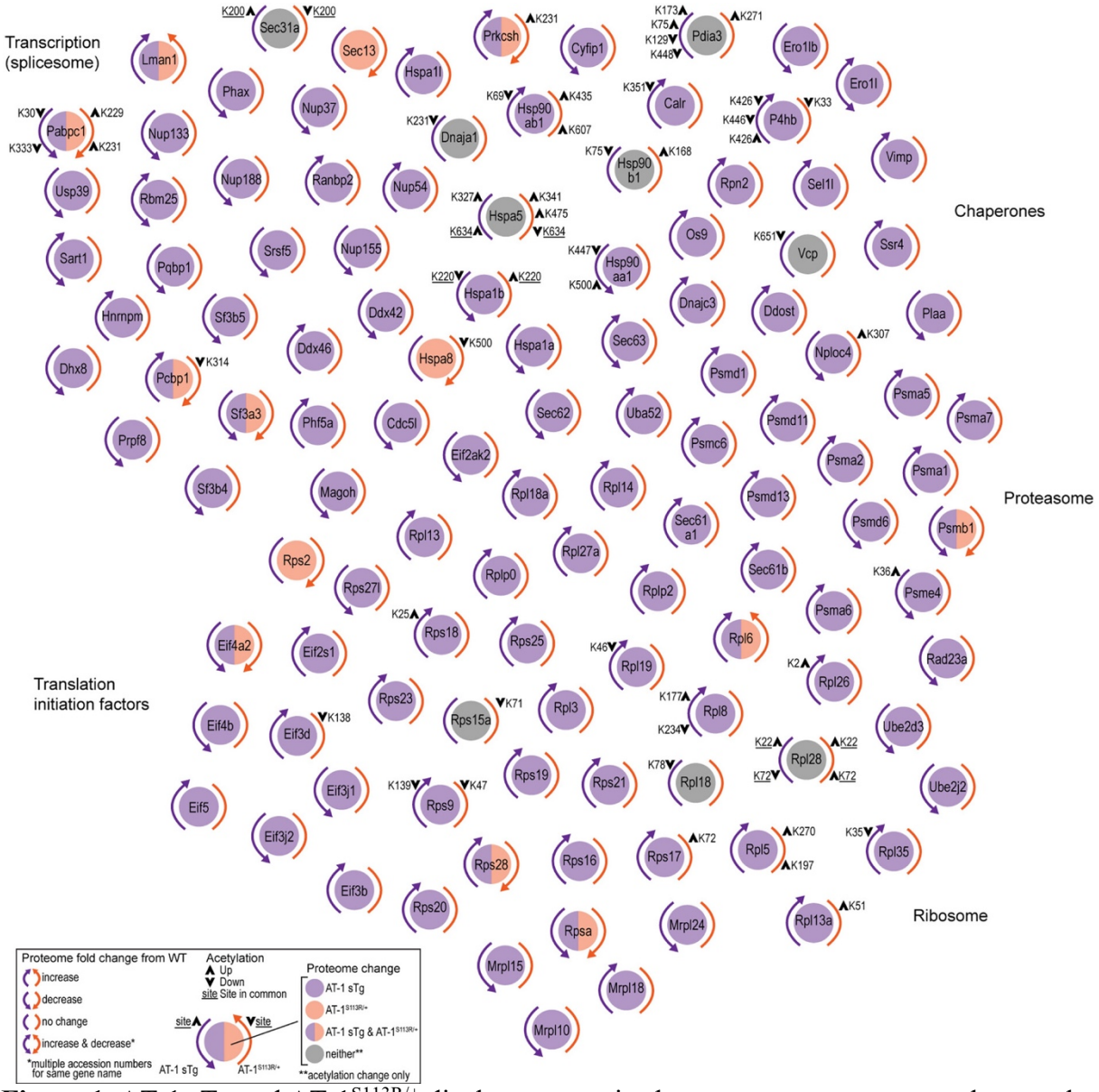

Figure 1. AT-1 sTg and AT-1$^{S113R/+}$ display proteomic changes across secretory pathway-related processes. Proteins with significantly changing abundance and acetylation stoichiometry identified across secretory pathway-related KEGG pathways from liver fractions are shown in clusters (n = 4 WT; n = 4 AT-1 sTg; n = 4 AT-1$^{S113R/+}$). Changes in proteome are indicated by a purple (AT-1 sTg), orange (AT-1$^{S113R/+}$), or grey (neither proteome changed) circle; corresponding colored arrows indicate increase, decrease, or no change in proteome compared with WT. Acetylation stoichiometry changes are designated with a black arrow on the corresponding AT-1 model in which it changed significantly. If a site changes significantly in both models, the site is underlined.

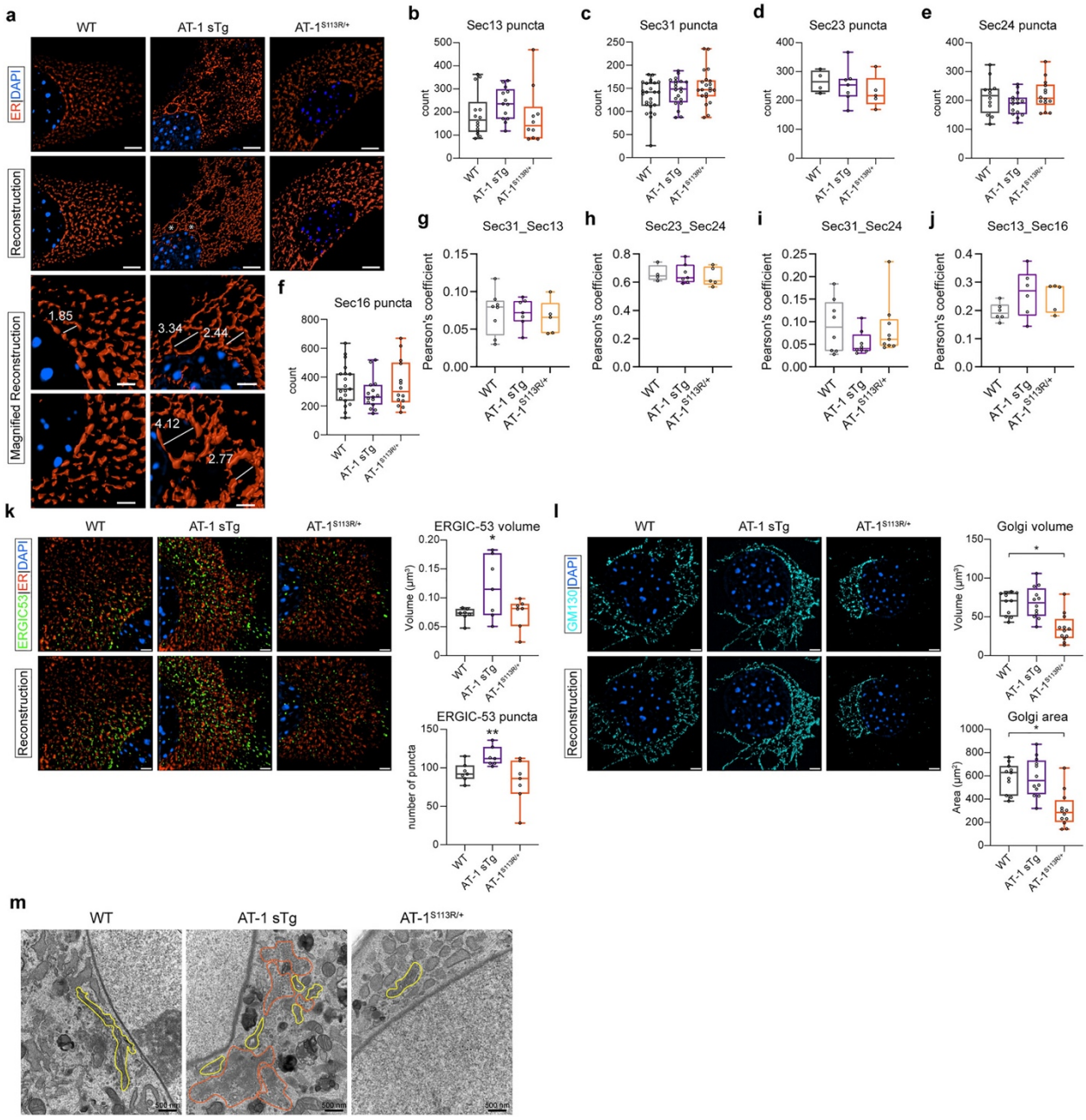

Figure 2. Morphological reorganization of ER, ERGIC and Golgi apparatus in aberrant AT-1 models. (**a**) Representative ER morphology in primary-cultured MEFs using ER3-mCherry (scale bar, 3 μm). White asterisk (*) indicates an enlarged cisternae within the perinuclear rough ER. High-magnification areas for WT and AT-1 sTg are shown (scale bar, 2 μm). (**b**–**f**) Quantification of puncta revealed using SIM microscopy in primary-cultured MEFs in (**b**) Sec13 puncta (n = 14 WT; n = 13 AT-1 sTg; n = 10 AT-1$^{S113R/+}$), (**c**) Sec31 puncta (n = 25 WT; n = 20 AT-1 sTg; n = 21 AT-1$^{S113R/+}$), (**d**) Sec23 puncta (n = 4 WT; n = 7 AT-1 sTg; n = 5 AT-1$^{S113R/+}$), (**e**) Sec24 puncta (n = 12 WT; n = 15 AT-1 sTg; n = 14 AT-1$^{S113R/+}$), (**f**) Sec16 puncta (n = 19 WT; n = 15 AT-1 sTg; n = 14 AT-1$^{S113R/+}$). (**g**–**j**) Pearson's coefficient of puncta revealed using SIM microscopy in primary-cultured MEFs in (**g**) Sec31 and Sec13 puncta (n = 8 WT; n = 7 AT-1 sTg; n = 5 AT-1$^{S113R/+}$), (**h**) Sec23 and Sec24 puncta (n = 4 WT; n = 7 AT-1 sTg; n = 5 AT-

$1^{S113R/+}$), (**i**) Sec31 and Sec24 puncta (n = 8 WT; n = 8 AT-1 sTg; n = 9 AT-$1^{S113R/+}$), (**j**) Sec13 and Sec16 puncta (n = 6 WT; n = 6 AT-1 sTg; n = 5 AT-$1^{S113R/+}$). (**k**) ERGIC morphology in primary-cultured MEFs using ERGIC-53 antibody (scale bar, 3 µm) and quantified using Imaris reconstruction of volume and number of puncta (n = 7 WT; n = 7 AT-1 sTg; n = 7 AT-$1^{S113R/+}$). (**l**) Golgi apparatus morphology in primary-cultured MEFs using GM130 antibody (scale bar, 3 µm) and quantified using Imaris reconstruction of volume and area (n = 11 WT; n = 12 AT-1 sTg; n = 12 AT-$1^{S113R/+}$). (**m**) Representative electron microscopy of MEFs from WT, AT-1 sTg, and AT-$1^{S113R/+}$. Yellow outlines Golgi structures and orange outlines disorganized secretory structures and vesicles. (scale bar, 500 nm). MEFs from multiple cells from 3 biologically independent animals for each genotype for (**b**–**l**). One-way ANOVA. *P < 0.05; **P < 0.005.

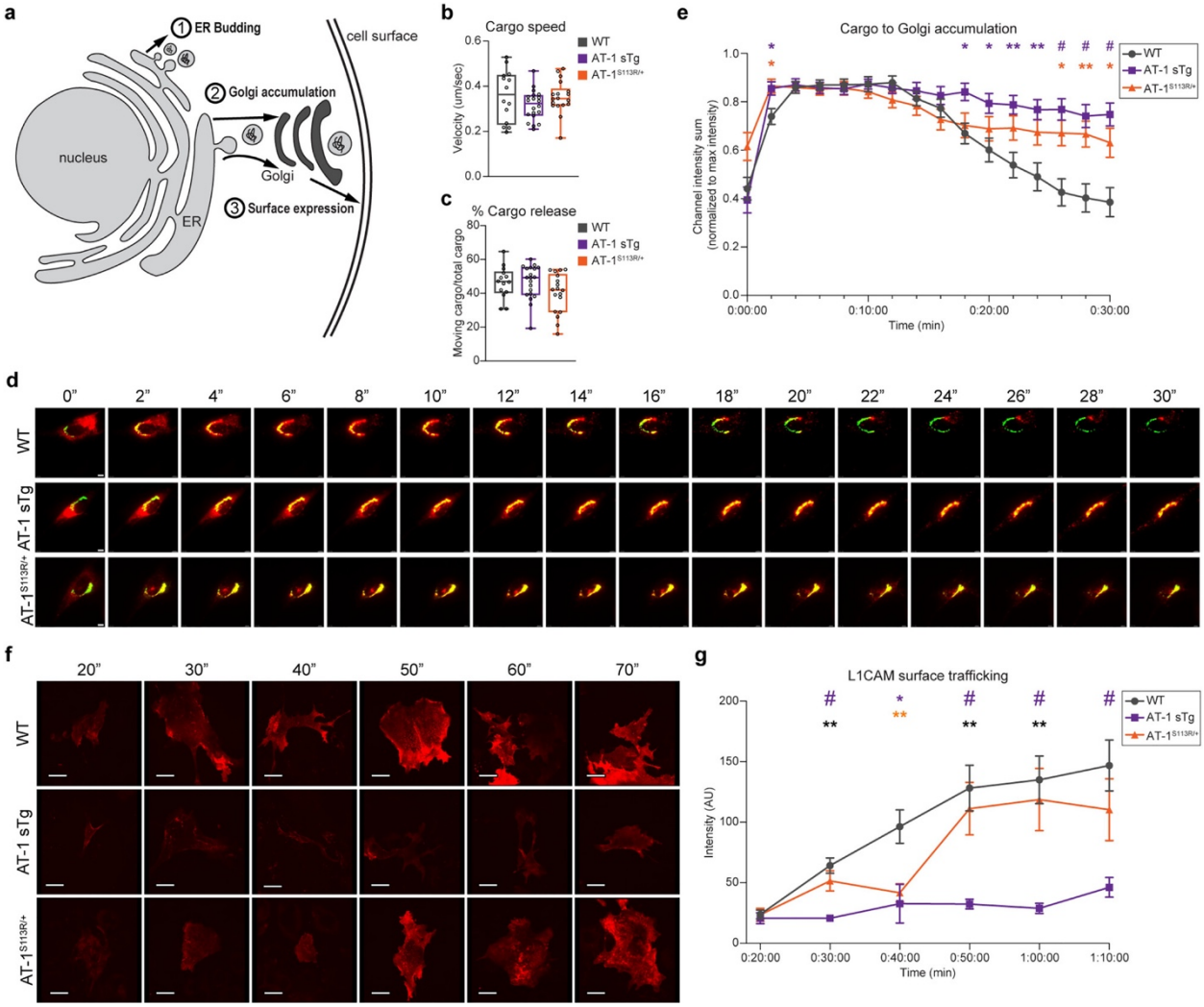

Figure 3. Aberrant AT-1 models demonstrate altered protein trafficking through the secretory pathway. (**a**) Schematic of three experiments used to assess protein trafficking in the secretory pathway. Experiment 1 assesses protein-laden carriers following release from the ER; solubilizer-dependent cargo fluoresces for visualization. Experiment 2 quantifies the time for ER cargo to accumulate in the Golgi apparatus and then egress; solubilizer-dependent DsRed cargo traffics to GFP-labeled Golgi. Experiment 3 quantifies the time for ER cargo to reach the cell

surface; solubilizer-dependent cargo with a HaloTag ligand is tracked for 70 min with HaloTag cell impermeable dye added prior to imaging. (**b,c**) Maximum speed of cargo release (**b**), and the percentage of cargo release (**c**), are tracked for 3 min after solubilizer induced cargo release in MEFs (MEFs from multiple cells from 3 biologically independent animals for each genotype n = 14 WT; n = 20 AT-1 sTg; n = 18 AT-$1^{S113R/+}$; MEF cells from 3 biologically independent animals for each group for (**b**,**c**). Student's T-test. (**d**) Representative cargo (red) accumulation in the Golgi (green) every 2 min, for 30 min in MEFs from WT, AT-1 sTg, and AT-$1^{S113R/+}$ (scale bar, 40 μm). (**e**) Cargo accumulation in the Golgi over 30 min in primary-cultured MEFs (MEFs from multiple cells from 3 biologically independent animals for each genotype at each time point, n = 27 WT; n = 22 AT-1 sTg; n = 23 AT-$1^{S113R/+}$). Mean with SEM are represented at each time point. (**f**) Representative surface trafficking of L1CAM in MEFs from WT, AT-1 sTg, and AT-$1^{S113R/+}$ (scale bar, 40 μm). (**g**) L1CAM intensity at the cell surface in MEFs (MEFs from multiple cells from 3 biologically independent animals for each genotype at each time point, n = 42–76 WT; n = 10–27 AT-1 sTg; n = 19–34 AT-$1^{S113R/+}$). Mean with SEM are represented at each time point. Two-way ANOVA with multiple comparisons at each time point for (**e**,**g**). Purple and orange statistics indicate differences between AT-1 sTg and WT, and between AT-$1^{S113R/+}$ and WT, respectively. Grey statistics indicate difference between AT-1 sTg and AT-$1^{S113R/+}$. *$P < 0.05$; **$P < 0.005$; $^{\#}P < 0.0001$.

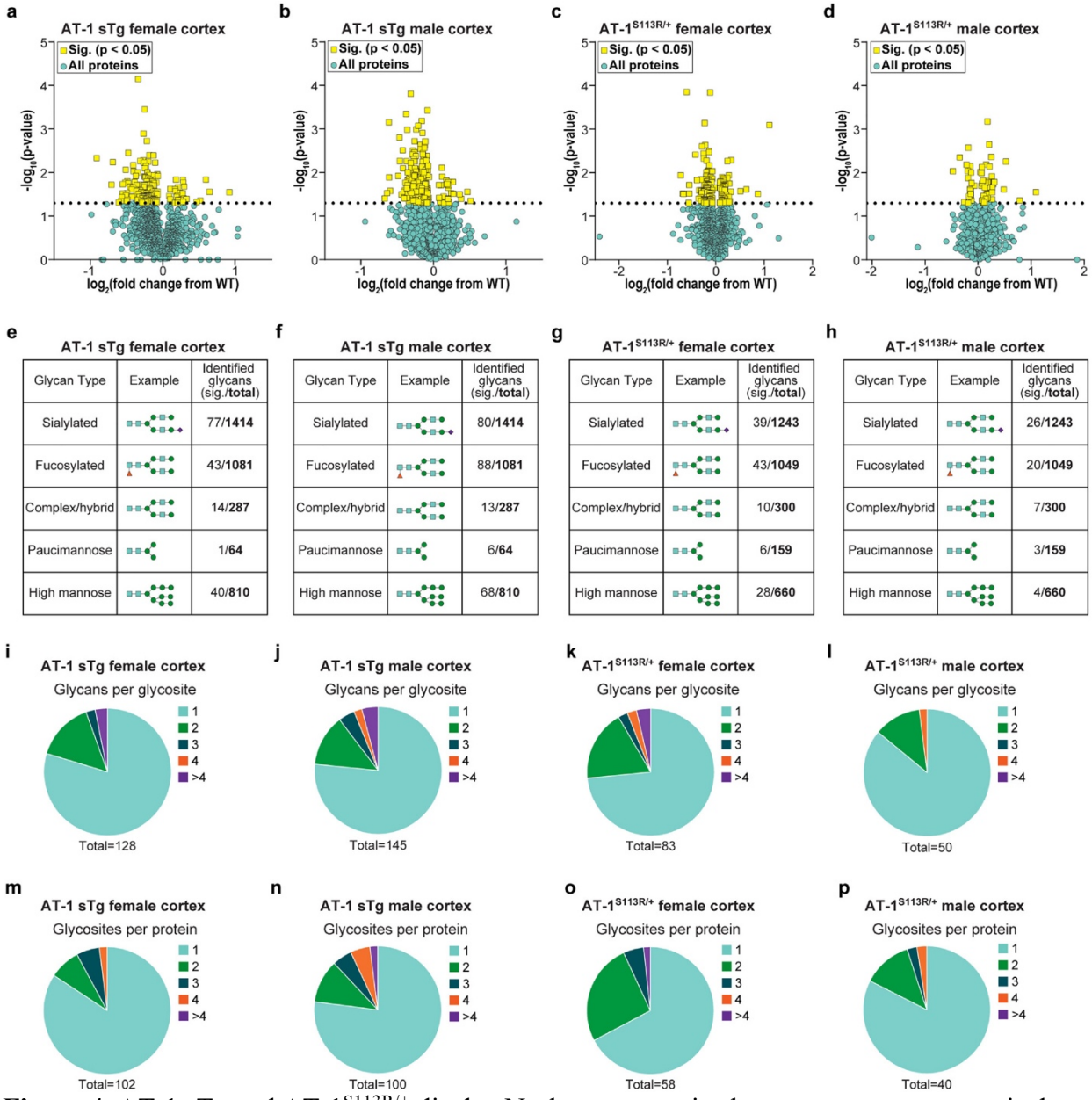

Figure 4. AT-1 sTg and AT-1$^{S113R/+}$ display N-glycoproteomic changes across many cortical glycoforms. (**a**,**b**) Volcano plot displaying all quantified glycoproteins in the cortex of AT-1 sTg (**a**) female (n = 3) and (**b**) male (n = 3), compared with age-matched WT littermates. Statistically significant proteins (175 in female; 255 in male) are highlighted in yellow, and all other proteins are designated in blue. Student's T-test, $P < 0.05$. (**c**,**d**) Volcano plot displaying all quantified glycoproteins in the cortex of AT-1$^{S113R/+}$ (**c**) female (n = 3) and (**d**) male (n = 3), compared with age-matched WT littermates. Statistically significant proteins (126 in female; 60 in male) are highlighted in yellow, and all other proteins are designated in blue. Student's T-test, $P < 0.05$. (**e**–**h**) Identified glycans are categorized into five glycan types, and are divided by significant over total identified in cortical (**e**) AT-1 sTg female, (**f**) AT-1 sTg male, (**g**) AT-1$^{S113R/+}$ female, and (**h**) AT-1$^{S113R/+}$ male. (**i**–**l**) Significant glycans per glycosite are identified in cortical (**i**) AT-1 sTg female, (**j**) AT-1 sTg male, (**k**) AT-1$^{S113R/+}$ female, and (**l**) AT-1$^{S113R/+}$ male. (**m**–**p**) Significant

glycosites per protein are identified in cortical (**m**) AT-1 sTg female, (**n**) AT-1 sTg male, (**o**) AT-$1^{S113R/+}$ female, and (**p**) AT-$1^{S113R/+}$ male.

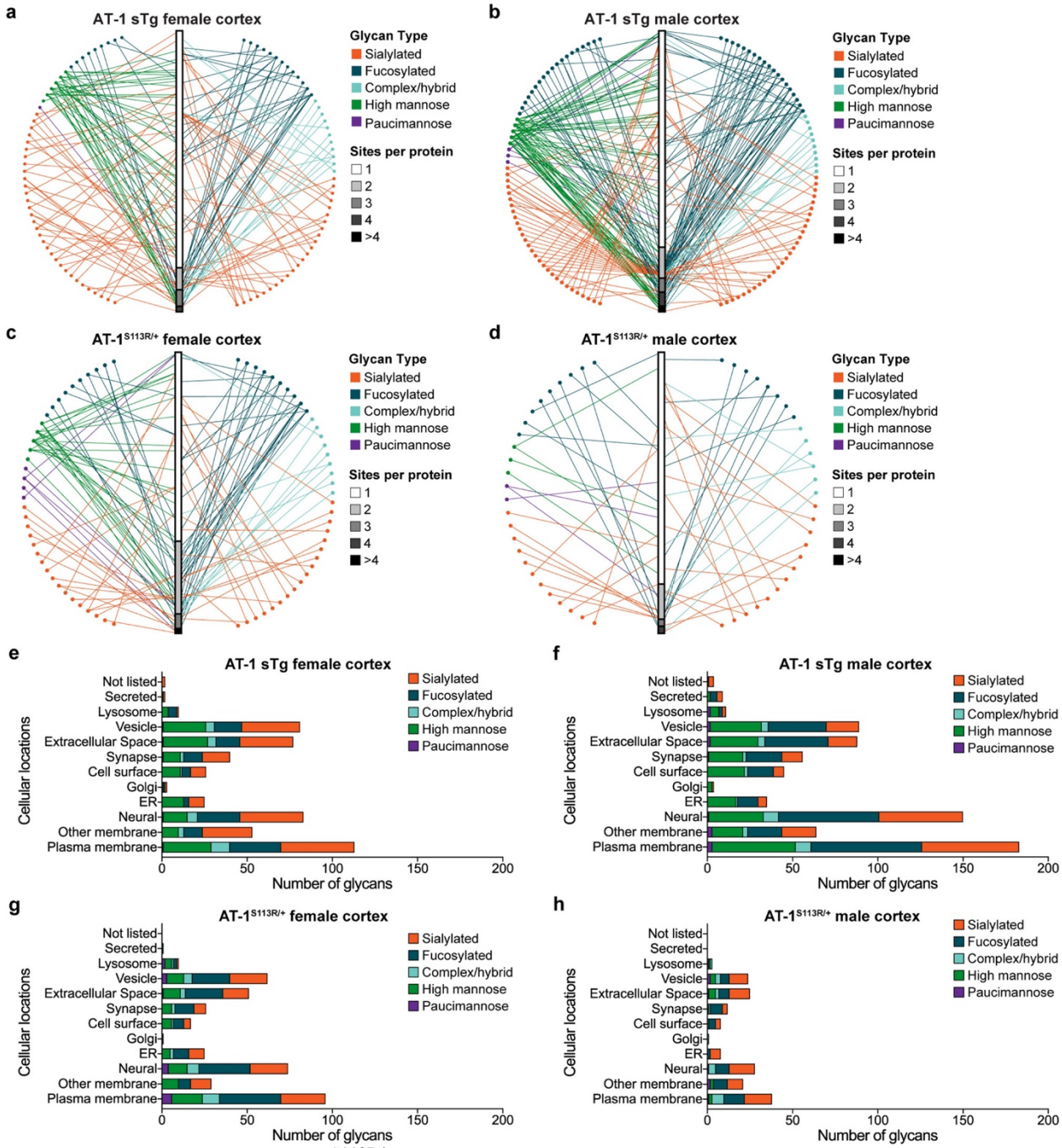

Figure 5. AT-1 sTg and AT-$1^{S113R/+}$ display a highly heterogeneous cortical secretome. (**a**–**d**) Network of significant glycan types plotted to glycoprotein for cortical (**a**) AT-1 sTg female, (**b**) AT-1 sTg male, (**c**) AT-$1^{S113R/+}$ female, and (**d**) AT-$1^{S113R/+}$ male. Circle represents glycan type, with each node representing a specific glycan tree. Glycans intersect with their protein, organized by the number of glycosites identified on that glycoprotein. (**e**–**h**) Glycosylation distribution according to subcellular localizations as defined by GO cellular component terms in

cortical (**e**) AT-1 sTg female, (**f**) AT-1 sTg male, (**g**) AT-1$^{S113R/+}$ female, and (**h**) AT-1$^{S113R/+}$ male.

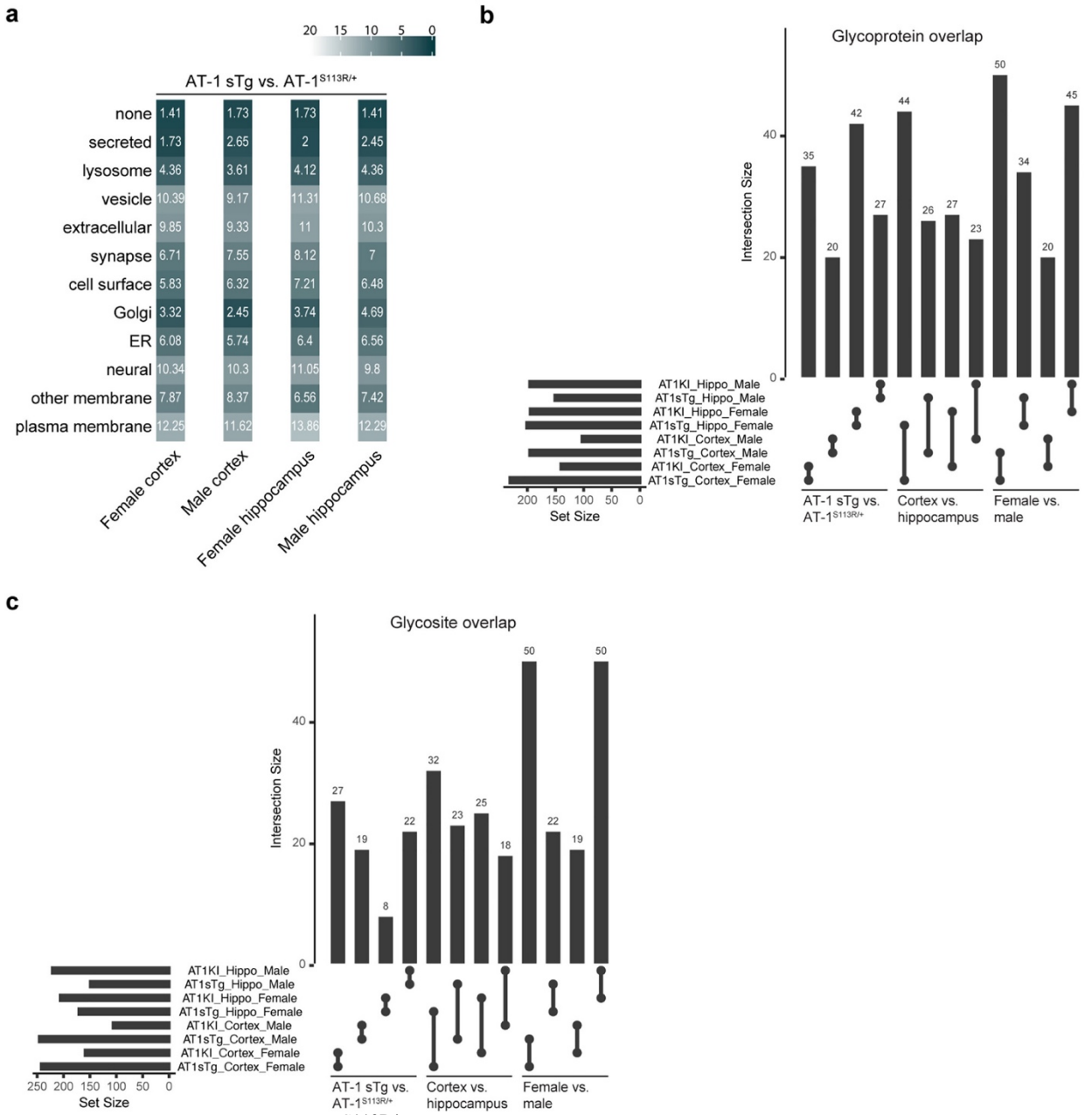

Figure 6. AT-1 sTg and AT-1$^{S113R/+}$ display both divergent and convergent N-glycoproteomic changes across subcellular locations. (**a**) Euclidean distances comparing the distance between significant glycoforms found in AT-1 sTg and AT-1$^{S113R/+}$, broken down by sex and brain region. Distances were calculated according to their subcellular locations, with higher similarity indicated with dark colors. (**b**,**c**) Significant glycoprotein (**b**) and glycosite (**c**) overlap is examined at the level of model (AT-1 sTg v. AT-1$^{S113R/+}$), brain region (cortex vs. hippocampus) and sex (female vs. male). The upper bar graph indicates the size of intersection and the left bar graphs shows total number of significant glycoproteins (**b**) or glycosites (**c**) in each dataset.

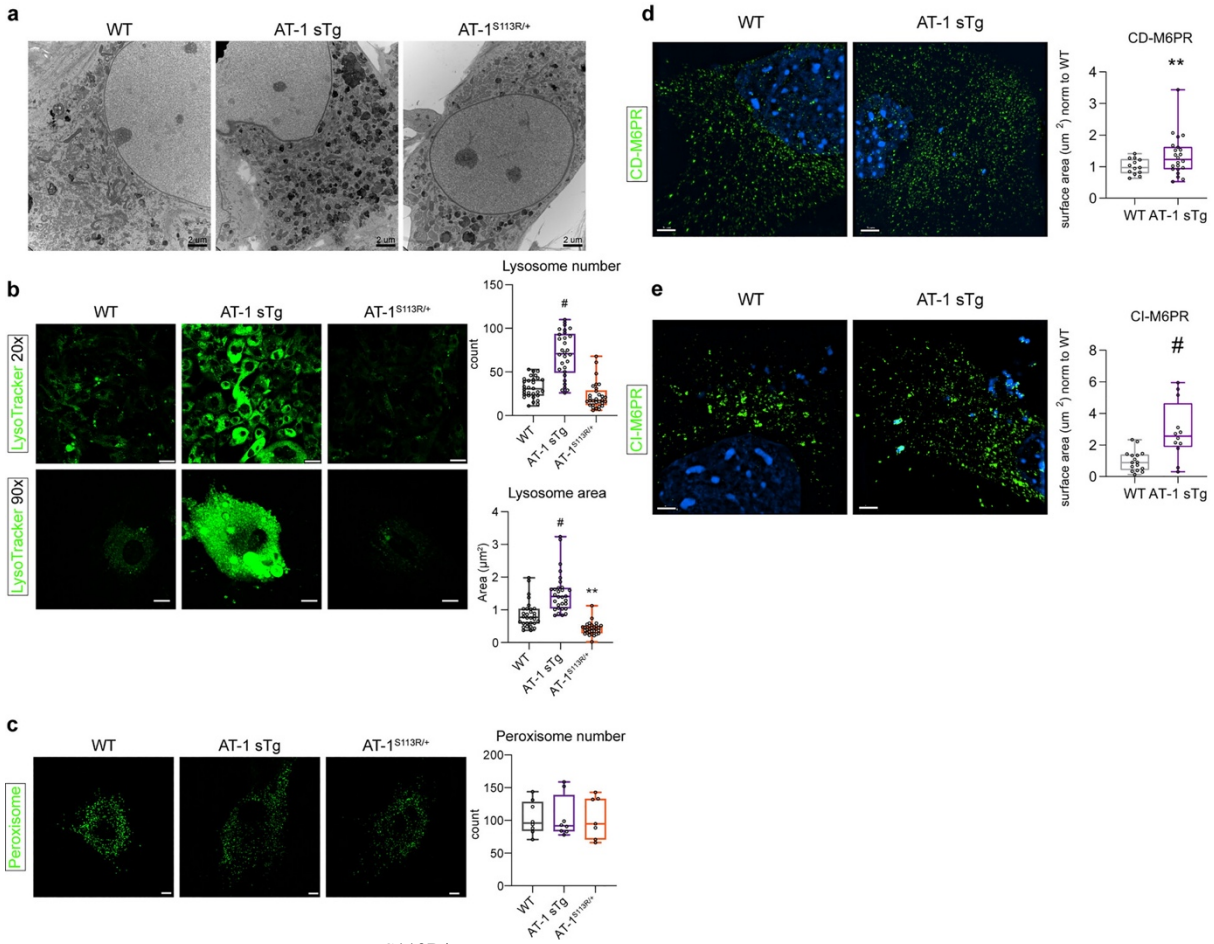

Figure 7. AT-1 sTg and AT-1$^{S113R/+}$ display alterations in the CD-M6PR, CI-M6PR, and lysosomal networks. (**a**) Representative electron microscopy of MEFs from WT, AT-1 sTg, and AT-1$^{S113R/+}$ (scale bar, 2 µm). (**b**) Lysosomal morphology in primary-cultured MEFs using LysoTracker stain (20 × scale bar, 30 µm; 90 × scale bar, 15 µm) and quantified using Imaris reconstruction of surface area and number of puncta (n = 30 WT; n = 30 AT-1 sTg; n = 30 AT-1$^{S113R/+}$). (**c**) Peroxisome morphology in primary-cultured MEFs using Peroxisome CellLight stain (scale bar, 10 µm) and quantified using Imaris reconstruction of number of puncta (n = 8 WT; n = 8 AT-1 sTg; n = 7 AT-1$^{S113R/+}$). (**d**) Representative CD-M6PR in primary cultured MEFs using CD-M6PR antibody (scale bar, 3 µm). Surface area was quantified using Imaris reconstruction (n = 13 WT; n = 22 AT-1 sTg). (**e**) Representative CI-M6PR in primary cultured MEFs using CI-M6PR antibody (scale bar, 3 µm). Surface area was quantified using Imaris reconstruction (n = 16 WT; n = 12 AT-1 sTg). Two-way ANOVA (**b**,**c**). Welchs' t test (**d**,**e**). **P < 0.005; $^{\#}$P < 0.0001.

Chapter 5

Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides

Adapted from: **Cui, Y.**; Yang, K.; Tabang, D. N.; Huang, J.; Tang, W.; Li, L. Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides. *J Am Soc Mass Spectr* **2019**, *30* (12), 2491–2501.

Abstract

Simultaneous enrichment of glyco- and phosphopeptides will benefit the studies of biological processes regulated by these posttranslational modifications (PTMs). It will also reveal potential crosstalk between these two ubiquitous PTMs. Unlike custom-designed multifunctional solid phase extraction (SPE) materials, operating strong anion exchange (SAX) resin in electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) mode provides a readily available strategy to analytical labs for enrichment of these PTMs for subsequent mass spectrometry (MS)-based characterization. However, the choice of mobile phase has largely relied on empirical rules from hydrophilic interaction chromatography (HILIC) or ion-exchange chromatography (IEX) without further optimization and adjustments. In this study, ten mobile phase compositions of ERLIC were systematically compared; the impact of multiple factors including organic phase proportion, ion pairing reagent, pH, and salt on the retention of glycosylated and phosphorylated peptides was evaluated. This study demonstrated good enrichment of glyco- and phosphopeptides from the nonmodified peptides in a complex tryptic digest. Moreover, the enriched glyco- and phosphopeptides elute in different fractions by orthogonal retention mechanisms of hydrophilic interaction and electrostatic interaction in ERLIC, maximizing the LC-MS identification of each PTM. The optimized mobile phase can be adapted to the ERLIC HPLC system, where the high resolution in separating multiple PTMs will benefit large-scale MS-based PTM profiling and in-depth characterization.

Keywords: Electrostatic repulsion-hydrophilic interaction chromatography (ERLIC); Enrichment; Glycosylation; Phosphorylation; Posttranslational modifications (PTMs); Strong anion exchange (SAX).

Introduction

Phosphorylation and glycosylation are two of the most ubiquitous and important protein post-translational modifications (PTMs) that can significantly alter protein tertiary structure, stability, turnover and protein-protein interactions [1]. It has been estimated that nearly one-third of human proteins are phosphorylated and more than one-half of mammalian proteins are glycosylated [2, 3]. They are highly associated with many biological processes, including signal transduction, enzyme activity regulation, protein folding and cell adhesion [4, 5], and related to various human diseases, including diabetes, Alzheimer's disease, autoimmunity, congenital disorders of glycosylation, and cancer [6–14]. It also has been demonstrated that there is complex crosstalk between the two PTMs in biological regulation. [15][16]. In order to study the PTM related biological processes and further dissect the molecular mechanisms underlying the diseases involving these two modifications, the ability to simultaneously analyze both phosphorylation and glycosylation at the proteome level is essential.

Bottom-up proteomics (i. e., shotgun proteomics), which couples liquid chromatography (LC) and tandem mass spectrometry to analyze proteolytic peptides, has become a main-stream technology that enables rapid profiling of proteins in complex samples [17–19]. The technology has undergone rapid development in the past few decades, and its high-throughput allows the whole proteome from eukaryotic cells to be analyzed within just a few hours [20–22]. To better adjust the

method for analyzing phosphate or glycan modified peptides, more sophisticated LC-MS experiments have been designed such as the 'multistage activation (MSA)' method [23], the electron-driven dissociation [24, 25], and the hybrid fragmentation strategies such as electron-transfer/higher-energy collision dissociation (EThcD) and activated ion electron-transfer dissociation (AI-ETD) that combine orthogonal types of activation on the modified precursor ions [26–33]. However, neither of the above strategies can avoid the increased duty cycle of MS/MS analysis, which will compromise the number of MS/MS events and reduce the large-scale profiling efficiency and overall coverage of modified peptides. Furthermore, MS^n and electron-driven fragmentation capabilities are currently only available on a limited number of instruments. On the other hand, stepped-energy of higher-energy collision dissociation (stepped HCD or stepped NCE) approach, which combines the fragment ions from three different collision energies and records these fragments in a single MS/MS spectrum, was found to increase the diversity of the fragment ions without significantly increasing the MS analysis duty cycle or sacrificing the throughput, which benefits large scale PTM analysis. For instance, improved phospho site localization through increased sequence coverage on HEK 293T cell lysate was reported by applying a stepped HCD approach [34]. A recent application of stepped NCE to the human serum proteome also proved its capability to fully characterize intact glycopeptides on a large scale, while simultaneously yielding structural information of the peptides and the covalently linked glycans via high and low collision energies [35].

Despite the technical improvements to MS instruments, it is still challenging to analyze phosphorylated or glycosylated peptides by direct LC-MS without reducing sample complexity, due to the sub-stoichiometric level abundance of the two PTMs and poor ionization efficiency in positive ion mode MS derived from additional negative charges of these PTMs. Therefore,

numerous enrichment techniques have been developed targeting phosphorylated or glycosylated peptides, including metal oxide affinity chromatography (MOAC) such as TiO_2 and ZrO_2 enrichment [36, 37], immobilized metal affinity chromatography (IMAC) [38–42], lectin affinity chromatography-based enrichment [43, 44], hydrazide chemistry based SPE method [45–48], boronic acid enrichment method [49–51] and hydrophilic interaction chromatography (HILIC) [52–55]. Most of the methods listed above were designed to target phosphorylation or glycosylation alone and each has its own limitations. MOAC and IMAC suffer from non-specific binding of aspartic acid and glutamic acid rich peptides. Lectin affinity chromatography exhibits biases towards specific glycan structures. Hydrazide chemistry-based approaches can enrich N-glycopeptides with high specificity and efficiency, but the release of enriched analytes relies on enzymatic digestion that makes intact glycopeptide analysis impossible. In recent years, novel MOAC and IMAC materials were synthesized with optimized morphology, ligand selection and density of functional groups to increase the selectivity and enrichment efficiency of both phosphorylated and glycosylated peptides [56][57] [58]. However, most of the applications of those novel enrichment materials were limited to MALDI-MS validation on mixtures of several phosphorylated and glycosylated peptide standards, while the in-depth analysis of the modified proteome was limited by the availability of the custom synthesized material to biological labs.

Ion exchange chromatography (IEX) such as strong cation exchange (SCX) or strong anion exchange (SAX) has been another widely used peptide and protein separation method in biological and analytical labs [59]. MS coupled with IEX separation has been successfully implemented in phosphorylated peptide analysis, due to the charge difference introduced by the phosphate group [60–62]. Electrostatic repulsion hydrophilic interaction chromatography (ERLIC) is a novel mode of separation utilizing IEX stationary phase with HILIC mobile phases (i.e.

analytes are eluted by decreasing organic solvent gradient), which was first introduced by Alpert [63]. In ERLIC, hydrophilic interaction and electrostatic interaction between analytes and stationary phase are superimposed; and one can manipulate the charge state of analytes by operating at optimal pH to maximize the electrostatic attraction between analytes of interest and charged stationary phase, while the electrostatic repulsion will antagonize the retention of interfering molecules that cannot be separated by conventional HILIC with neutral stationary phase [63]. Anion exchange-ERLIC can selectively separate phosphorylated peptides from non-modified counterparts because of the increased hydrophilicity and additional negative charge introduced by a phosphate group at the operating pH [64]. Glycosylated peptide enrichment by its hydrophilicity is also compatible with ERLIC, and it has been reported that ERLIC outperforms HILIC by neutral stationary phases on isobaric labeled glycopeptide enrichment, which facilitated further PTM relative quantification [65, 66]. Furthermore, ERLIC by both SCX and WAX has been investigated for the simultaneous, selective isolation of both phosphopeptides and glycopeptides from tryptic digests [67–69]. Despite the plentiful application of ERLIC, the mobile phase constitution for simultaneous enrichment of multiple PTMs has not been thoroughly explored.

In this work, we examined the potential of coupling ERLIC SPE with RPLC-MS to achieve large-scale phosphorylation and N-glycosylation analysis. Different mobile phase compositions, including organic phase proportion, ion pairing reagent, salt and pH, were tested and evaluated to maximize the enrichment performance of both N-glycosylated and phosphorylated tryptic peptides from MM.1S cell lysate (a model cell line for multiple myeloma) [70].

Experimental Section

Materials.

Poly SAX LP™ bulk material (12µm, pore size 300Å) was obtained from PolyLC (Columbia, MD). 0.22 µm Millex-GV Hydrophilic Durapore (PVDF) Membrane filter units were purchased from Merck Millipore (Tullagreen, Ireland). Iodoacetamide (IAA), Roche protease inhibitor tablets and Roche PhosSTOP phosphatase inhibitor tablet were purchased from Sigma-Aldrich (St. Louis, MO). Phosphoric acid, ammonium acetate, tris base, urea, potassium phosphate monobasic, sodium chloride and calcium chloride were purchased from Fisher Scientific (Pittsburgh, PA). C18 SepPak cartridges were purchased from Waters (Milford, MA). C18 OMIX tips were purchased from Agilent (Santa Clara, CA). Dithiothreitol (DTT) and sequencing grade trypsin were purchased from Promega (Madison, WI). Protein and peptide BCA assays were purchased from Pierce (Rockford, IL).

Cell Culture.

MM.1S cells were cultured in RPMI-1640 medium (Corning) supplemented with 10% FBS, 1% sodium pyruvate, 1% penicillin/streptomycin and 10mM HEPES at 37 °C in humidified atmosphere with 5% CO_2. When the MM.1S cells were confluent, they were harvested and then washed once with PBS to remove remaining media content. Resulting cell pellet(s) were stored under −80 °C until use.

Cell Lysis.

MM.1S cell pellets were homogenized in lysis buffer (8M urea, 50 mM Tris, pH=8, 5 mM $CaCl_2$, 20 mM NaCl, 1 EDTA-free Roche protease inhibitor tablet and 1 Roche PhosSTOP phosphatase inhibitor tablet) with a probe sonicator for 3 pulses at 60W, 20 kHz for 15 s, each followed by a 30 s pause period for cooling at 4ºC. Crude lysates were then centrifuged at 14000

xg for 5 min, after which the supernatant was collected and protein concentrations were measured by Pierce BCA Protein Assay according to the manufacturer's protocol.

Trypsin Digestion.

Lysate containing 2 mg protein was reduced in 5 mM DTT at room temperature for 1h, followed by alkylation in 15 mM IAA for 30 min in the dark. Alkylation was quenched by adding DTT to 5 mM. The resulting solution was then diluted with Tris buffer (pH=8) to 0.9 M urea and proteins were digested with trypsin at 1:50 enzyme to protein ratio at 37 ºC for 18 hours. Digestion was quenched by adding trifluoroacetic acid (TFA) to a final concentration of 0.3% and desalted with C18 SepPak cartridges. Digested peptide concentration was measured by Pierce Peptide BCA assay, peptides were aliquoted to 100 µg/tube and dried under vacuum before enrichment.

ERLIC Enrichment.

ERLIC enrichment of phospho- and N-glyco- peptides from MM.1S cell digest was performed in the custom packed spin-tips [71, 72]. As Figure 1B illustrates, 3mg of cotton wool was inserted into a 200 µL pipette tip. SAX LP bulk material was dispersed in 0.1% TFA as a 10 mg/ 200 µL slurry and activated for 15 min under vigorous vortexing. After activation, 60 µL slurry was added to the spin-tip. Solvent was removed by centrifugation at 1200 rpm for 2 min, leaving the SAX material packed above the cotton wool. The stationary phase was then conditioned by 150 µL ACN, 100 mM ammonium acetate, 1% TFA and the loading buffers, each repeated 3 times. 100 µg MM.1S cell digest was dissolved in 150 µL of the corresponding loading buffer and loaded onto the tips by centrifugation at 1200 rpm for 2 min, the flow-through was collected and loaded again to ensure complete retention. The tips were then washed with 150 µL loading buffers 6 times, after which four fractions were eluted with 300 µL 50/50 ACN/Water with 0.1%

formic acid (FA), 0.1% FA in water, 0.1% TFA in water and 300 mM KH_2PO_4 (pH = 2) and were collected in four separate centrifuge tubes. Each fraction was filtered by a 0.22 μm Millex-GV Hydrophilic Durapore (PVDF) Membrane filter unit, the KH_2PO_4 fractions were desalted by C18 OMIX tips according to manufacturer's protocol, after which all fractions were dried under vacuum before MS analysis.

NanoLC-MS/MS Analysis.

Peptides in each fraction were reconstituted in 0.1% FA, 3% ACN and subjected to reversed phase LC-MS/MS analysis with a Q-Exactive HF orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) interfaced with a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific, San Jose, CA). Peptides were loaded onto a 75 μm i.d. microcapillary column custom-packed with 15 cm of BEH C18 particles (1.7 μm, 130 Å, Waters). Peptides were separated with a 90 min gradient from 3% to 30% ACN with 0.1% FA, followed by 10 min to 75% ACN and then 10 min to 95% ACN. After that, the column was re-equilibrated with 3% ACN for 15 min to prepare for the next injection.

The mass spectrometer was operated in a top 15 data-dependent acquisition mode. Survey scans of peptide precursors from *m/z* 300 to 2000 were performed at a resolving power of 60K and an AGC target of 2×10^5 with a maximum injection time of 150 ms. The top 15 intense precursor ions were selected and subjected to the stepped HCD fragmentation at normalized collision energy of 22, 30, and 38% followed by tandem MS acquisition at a resolving power of 30K and an AGC target of 5×10^4, with a maximum injection time of 250 ms. Precursors were subjected to a dynamic exclusion of 15s with a 10 ppm mass tolerance.

Data Analysis

Raw files were processed with the Byonic search engine (Protein Metrics Inc, San Carlos, CA) embedded within Proteome Discoverer 2.1 (Thermo Fisher Scientific, San Jose, CA). Spectra were searched against the UniProt *Homo sapiens* proteome database (April 12, 2016; 16764 entries) with trypsin as the specific digestion enzyme and maximum two missed cleavages. The parent mass error tolerance was set to be 50 ppm and fragment mass tolerance was 0.02 Da. Fixed modifications are specified as carbamidomethylation on cysteine residues (+57.02146 Da). Dynamic modifications included oxidation of methionine residues (+15.99492 Da), phosphorylation on S, Y, and T, and N-glycosylation. It is reported that ERLIC can enrich for O-glycopeptides [73], but in this work, we only analyzed N-glycopeptides from each enrichment. Glycan modifications were specified as the common mammalian N-glycome (default N-glycome database in Byonic with removal of sodium mass, which contains overall 309 N-glycans), expanded with typical mannose-6-phosphate glycans including HexNAc (2)-Hex (4–9)-Phospho (1–2), HexNAc (3–4)-Hex (4–9)-Phospho (1–2), HexNAc (2) Hex (3–4) Phospho (1) and HexNAc (3) Hex (3–4) Phospho (1). Identifications were filtered at 1% false discovery rate (FDR).

Results and Discussion

Because multiple interactions such as hydrophilic interaction and electrostatic interaction are superimposed while operating ERLIC, selecting the best mobile phase is not straightforward. Here, we aim to examine the impact of organic phase proportion, ion pairing reagent, pH and salt content on the retention and elution of phospho- and N-glyco- peptides in ERLIC-RP-MS utilizing the new PolySAX LP material [64]. This work will benefit studies of the mechanism of

ERLIC and broaden its application to simultaneous phosphorylation and N-glycosylation analysis on a large scale.

Loading buffer affects enrichment efficiency.

To evaluate the impact of mobile phase composition on the retention of modified peptides, 10 different loading buffers were compared according to Table 1. Since HILIC was named and defined in 1990 [74], its mechanism of retention has aroused the interest of many scientists, among whom the partitioning theory is the most supported. In this theory, a water-rich layer is immobilized on the stationary phase due to the high density of hydrophilic functional groups, then the retention of solutes was equilibrated by the partition between the bulk eluent and the water-rich layer, which is different from conventional normal phase chromatography where the retention is predominantly governed by surface adsorption [75]. According to the empirical observations, mobile phase with a higher content of aqueous phase serves as a stronger eluent. Most of the pioneering studies utilizing ERLIC SPE for N-glycopeptide enrichment used 95% ACN for loading and washing without thorough justification [65, 66]. However, we found this condition to be too retentive in that many non-modified peptides with minimal hydrophilicity were co-retained with the modified peptides of interest. In Figure 2, the loading buffers with high (95%) ACN (group c, d and e) generally retain more total peptides than the low (80%) ACN buffer conditions, where the difference in peptide retention is as high as 10 times greater with the 95% ACN combinations (except for combination f, with 1% FA added), but the LC-MS identified significantly fewer modified peptides in high organic phase loading conditions. The identification of phosphorylated or N-glycosylated peptides was largely suppressed by the numerous co-eluting and co-fragmenting non-modified tryptic peptides that ionized better or

were more abundant. On the other hand, mobile phase with 80% ACN was adequate to retain the majority of the modified peptides while the non-modified peptides were removed during the washing step. By modulating the organic phase proportion in the mobile phase, the ERLIC enrichment specificity was greatly increased, benefiting the MS identification of peptides with phosphorylation and N-glycosylation.

Ion pairing has been reported as an effective strategy to increase the specificity in HILIC based enrichment of glycosylated peptides, and ion pairing HILIC has been successfully applied to mammalian and plant glycoproteomics in various forms including SPE and HPLC [52, 76–78]. Wimley-White water/octanol free energy scale of amino acids has shown theoretical evidence that peptides with charged moieties have lower hydrophobicity than their uncharged forms [77, 79], which is consistent with the experimental results in ion-pairing normal phase LC that the hydrophobicity of charged non-glycopeptides increase and the separation between glycopeptides and non-glycopeptides were achieved when efficient ion pairing reagents were added [77]. Ion pairing reagents such as monovalent salt ions including NaCl, LiCl and KCl or acids including acetic acid, formic acid and TFA will “neutralize” the charges of peptides by forming ion-pairs with the positively charged groups such as the N-termini, lysine and arginine residues, and the negatively charged groups including C-termini, aspartate and glutamate residues. With the non-modified peptides’ charges neutralized, their hydrophilicity is largely reduced, and the glycopeptide enrichment specificity is improved because the hydrophilic retention of analytes will predominantly depend on the existing glycans whose hydrophilicity is moderately affected by ion pairing reagents [52, 77, 78]. Moreover, a recent study systematically evaluated the impact of different salts on the retention of HILIC and ERLIC, where they found well-hydrated counterions of charged analytes usually increase the hydrophilic retention while poorly-hydrated

counterions decrease it [80]. The similar concept of counterion pairing was adapted to ERLIC here, with the choice of the poorly-hydrated TFA and weakly-hydrated FA as the ion pairing reagents being compared. On the other hand, the two acid modifiers resulted in different pH of the mobile phase and shifted the protonation-dissociation equilibrium of phosphate and other functional groups with acidic protons in the analytes, which would potentially influence the electrostatic interaction-based retention. As the most hydrophilic group in peptides except for basic residues, charged phosphate groups can increase the hydrophilicity of tryptic peptides [74], However, the phosphate group alone does not suffice to permit the separation of phosphopeptides from nonphosphopeptides as a set in HILIC. In ERLIC, the concurrent electrostatic interaction plays a significant role in phosphopeptide retention. Phosphate has a pK_{a1} of 2.15, which is lower than aspartic and glutamic residues and the C-terminus. With an appropriate pH in between them, the majority of peptides with acidic residues can be uncharged while the charged phosphopeptides can be strongly attracted by the positively charged SAX stationary phase. Thus the enrichment selectivity of phosphopeptides and highly hydrophilic glycopeptides over non-modified peptides can be enhanced with the selection of pH at 2–3.

As Figure 2B indicates, TFA serves as the best ion pairing reagent for selective N-glycosylated peptide retention, where use of loading buffer combinations of a, b, i and j, containing either 1% or 0.1% TFA resulted in more than 300 N-glycopeptide identifications. For phosphorylated peptides, the retention depends on hydrophilic interaction and electrostatic interaction synergistically. At high organic phase loading conditions, most of the phosphorylated peptides can be retained and identified regardless of the TFA content, because the strong hydrophilic interaction plays a predominate role. At low organic phase and stronger hydrophilic eluent loading conditions as in groups a, b and g-j, the electrostatic interaction became a more

significant component of the mechanism of phosphopeptide retention. At lower TFA or FA concentration, pH was higher and phosphate groups were more highly deprotonated and better attracted by the SAX resin. For example, in group b, 165 phosphopeptides were identified with high confidence whereas there were only 54 phospphopeptides identified in group a. Group h with even higher pH of 3 of 0.1% FA also showed great phosphopeptide retention where 171 phosphopeptides were identified in the total four fractions.

Conditioning is an essential preparation step before operating SAX resins. A highly concentrated salt buffer such as triethylammonium acetate is usually used to flush SAX resin for various reasons [65, 66]. Some of the main reasons include: (1) It can wash off the byproducts that came from SAX resin synthesis before the first time usage; (2) Well-hydrated counter ions can partition into the immobilized aqueous layer on the stationary phase and pulls the charged solutes into the aqueous layer thus increase the HILIC retention, while poorly hydrated counter ions have the opposite effect [80]. To prove the concept, we incorporated a control group i without flushing with salt in the experimental procedure before sample loading. Surprisingly, we found both N-glycosylated and phosphorylated peptide identifications from group i actually had a moderate increase over group a, which could arise from the more well-hydrated counterions that PolySAX LP has from manufacturer actually outperform acetate. It is worth mentioning that a large amount of 1% FA as the mobile phase modifier in groups f and g resulted in largely reduced retention of both modified and non-modified peptides comparing to groups e and h. We hypothesize that high concentration of FA partitioned in the immobilized aqueous layer can compete the electrostatic interaction between charged analytes and stationary phase to a large extent, while the retention of hydrophilic peptides is reduced by “salt out” effect [80]. In contrast, these effects are not observed with 1% TFA groups a, c and j, which is because TFA tends to

partition mostly into the organic mobile phase of ERLIC. Figure 2C shows the proportion of identified N-glycosylated or phosphorylated peptides within the total identified peptides for each combination. Group b is the optimal combination among the ten tested, with enrichment specificity of 24.8 % and 11.0 % for N-glycosylated and phosphorylated peptides, respectively.

Eluting profile reveals the key retention mechanism.

To further separate the modified peptides and simplify the MS based analysis, four crude elution fractions were collected with mobile phase conditions as indicated in Table 2. The peptide spectral matches (PSMs) of each modification-containing peptide are counted and the distributions of the PSMs among the eluting fractions are shown in Figure 3. Figure 3A summarizes total PSM distribution in four fractions. Figures 3B shows that the majority of N-glycosylated peptides elute in the first two fractions with the decreasing gradient of organic phase, indicating the retention mechanism for N-glycopeptides is predominately by hydrophilic interaction. On the other hand, in Figure 3C, phosphopeptides aggregate in the fractions 3 and 4, where the pH is lowered and the competitive salt is added to reduce the column interaction with the SAX resin. The results show that ERLIC can separate peptides with the two types of PTMs, which improves MS analysis and minimizes interference between peptides with the two modifications. Sialic acid containing N-glycopeptides have additional acidic groups of which the pK_a is 2.6. According to Figure 3D, the elution profiles of sialylated N-glycopeptides are consistent with those of the bulk N-glycopeptides from each condition in Figure 3B. The results indicate the predominant mechanism for ERLIC to retain sialic acid containing N-glycopeptides at pH 1~3 is through hydrophilic interaction. We hypothesize that if pH were further increased to >3, the electrostatic interaction would have more contribution.

Furthermore, we manually checked some PSMs of N-glyco-, phospho-, and sialylated N-glycopeptides from the Byonic search engine output. Figure 4 shows three representative spectra, where the HCD with stepped collisional energy of 22%, 30% and 38% resulted in sufficient peptide backbone fragments with clear modification site information and glycan fragments cleaved at the glycosidic linkage. The results support stepped HCD as being a highly efficient and simple MS fragmentation method for simultaneous profiling of multiple PTMs without tedious optimization when coupled with ERLIC fractionation and separation.

Conclusions

In this study, we demonstrated the great potential of ERLIC for simultaneous enrichment of phosphorylated and N-glycosylated peptides. Ten different loading buffer conditions were examined, thoroughly considering the factors of organic phase proportion, ion pairing reagents, pH and salt. While a high concentration of organic phase (95% ACN) has commonly been used in previous studies, it is outperformed here by 80% ACN in terms of hydrophilic interaction-based enrichment's selectivity and specificity for modified versus non-modified peptides. The use of a less well-hydrated ion pairing reagent benefits N-glycosylation enrichment by reducing non-specific retention. With weak hydrophilic eluent, phosphopeptides' retention is independent of ion-pairing reagents; with intermediate hydrophilic eluent, the pH of mobile phase has a stronger impact on electrostatic interaction-based retention of phosphorylated peptide, where pH between 2–3 ensures a decent extent of phosphate group dissociation and phosphopeptide retention, while the majority of acidic peptides remains to be uncharged, which enhances the enrichment selectivity of phosphopeptides by ERLIC. With an inappropriate selection of

counterions before or during the enrichment, the retention of modified peptides of interest can be largely hampered by competitive binding to charged resins and “salting out” effect. In the future, it is worth exploring the types of salt modifier that benefit the selective retention of N-glyco- and phospho- peptides in ERLIC.

The PSMs in the elution profiles of phosphorylated and N-glycosylated peptides showed a decent level of separation, where the majority of N-glycopeptides were eluted by disrupting hydrophilic interaction with decreasing organic phase gradient and most phosphopeptides were eluted by disrupting coulombic interaction with increased protonation and competitive salt content. The elution by multiple steps in the ERLIC SPE should be applicable to an HPLC system. We anticipate our optimized ERLIC SPE method will produce great resolution when adapted to HPLC and enable simultaneous enrichment and separation of N-glycosylated and phosphorylated peptides in complex mixtures, with improved coverage.

Acknowledgments

We would like to acknowledge Dr. Andrew Alpert from PolyLC Inc. for helpful discussions. This research was supported in part by the National Institutes of Health grants U01CA231081, R01 DK071801, R21 AG055377, and RF1 AG052324 (to LL). The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531 to LL) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. LL acknowledges a Vilas Distinguished Achievement Professorship and a Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the

Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

References

1. Ke M, Shen H, Wang L, Luo S, Lin L, Yang J, Tian R: Identification, Quantification, and Site Localization of Protein Posttranslational Modifications via Mass Spectrometry-Based Proteomics. In: Mirzaei H and Carrasco M (eds.) *Modern Proteomics -- Sample Preparation, Analysis and Practical Applications* pp. 345–382. Springer International Publishing, Cham; (2016)

2. Cohen P: The origins of protein phosphorylation. *Nat. Cell Biol* 4, E127–E130 (2002)

3. Apweiler R, Hermjakob H, Sharon N: On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta - Gen. Subj* 1473, 4–8 (1999)

4. Moremen KW, Tiemeyer M, Nairn AV: Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol* 13, 448–462 (2012)

5. Lee RT, Lauc G, Lee YC: Glycoproteomics: protein modifications for versatile functions. *EMBO Rep* 6, 1018–1022 (2005)

6. Ihara Y, Nukina N, Miura R, Ogawara M: Phosphorylated Tau Protein Is Integrated into Paired Helical Filaments in Alzheimer's Disease. *J. Biochem* 99, 1807–1810 (1986)

7. Maverakis E, Kim K, Shimoda M, Gershwin ME, Patel F, Wilken R, Raychaudhuri S, Ruhaak LR, Lebrilla CB: Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: a critical review. *J. Autoimmun* 57, 1–13 (2015)

8. Hakomori S: Glycosylation defining cancer malignancy: new wine in an old bottle. *Proc. Natl. Acad. Sci. USA* 99, 10231–10233 (2002

9. Schedin-Weiss S, Winblad B, Tjernberg LO: The role of protein glycosylation in Alzheimer disease. *FEBS J* 281, 46–62 (2014

10. Lassen PS, Thygesen C, Larsen MR, Kempf SJ: Understanding Alzheimer's disease by global quantification of protein phosphorylation and sialylated N-linked glycosylation profiles: A chance for new biomarkers in neuroproteomics? *J. Proteomics* 161, 11–25 (2017)

11. Karlsson HKR, Zierath JR, Kane S, Krook A, Lienhard GE, Wallberg-Henriksson H: Insulin-Stimulated Phosphorylation of the Akt Substrate AS160 Is Impaired in Skeletal Muscle of Type 2 Diabetic Subjects. *Diabetes* 54, 1692–1697 (2005)

12. Boucher J, Kleinridders A, Kahn CR: Insulin receptor signaling in normal and insulin-resistant states. *Cold Spring Harb. Perspect. Biol* 6, a009191.

13. Freeze HH, Eklund EA, Ng BG, Patterson MC: Neurology of inherited glycosylation disorders. *Lancet Neurol* 11, 453–466 (2012)

14. Kanninen K, Goldsteins G, Auriola S, Alafuzoff I, Koistinaho J: Glycosylation changes in Alzheimer's disease as revealed by a proteomic approach. *Neurosci. Lett* 367, 235–240 (2004)

15. Wang X, Li D, Wu G, Bazer FW: Functional roles of fructose: Crosstalk between O-linked glycosylation and phosphorylation of Akt-TSC2-mtor cell signaling cascade in ovine trophectoderm cells. *Biol. Reprod* 95 (2016); 102:1–17

16. Hang Q, Isaji T, Hou S, Im S, Fukuda T, Gu J: Integrinα5 suppresses the phosphorylation of epidermal growth factor receptor and its cellular signaling of cell proliferation via N-glycosylation. *J. Biol. Chem* 290, 29345–29360 (2015)

17. Aebersold R, Mann M: Mass spectrometry-based proteomics. *Nature* 422, 198–207 (2003)

18. Domon B, Aebersold R: Mass spectrometry and protein analysis. *Science* 312, 212–217 (2006)

19. Yates JR, Ruse CI, Nakorchevsky A: Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu. Rev. Biomed. Eng* 11, 49–79 (2009)

20. Bekker-Jensen DB, Kelstrup CD, Batth TS, Larsen SC, Haldrup C, Bramsen JB, Sørensen KD, Høyer S, Ørntoft TF, Andersen CL, Nielsen ML, Olsen JV: An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst* 4, 587–599.e584 (2017)

21. Riley NM, Hebert AS, Coon JJ: Proteomics Moves into the Fast Lane. *Cell Syst* 2, 142–143 (2016)

22. Hebert AS, Richards AL, Bailey DJ, Ulbrich A, Coughlin EE, Westphall MS, Coon JJ: The One Hour Yeast Proteome. *Mol. Cell. Proteomics* 13, 339–347 (2014)

23. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ: A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal. Chem* 76, 3590–3598 (2004)

24. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF: Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A* 101, 9528–9533 (2004)

25. Zubarev RA: Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol* 15, 12–16 (2004)

26. Reiding KR, Bondt A, Franc V, Heck AJR: The benefits of hybrid fragmentation methods for glycoproteomics. *Trend. Anal. Chem* 108, 260–268 (2018)

27. Riley NM, Hebert AS, Dürnberger G, Stanek F, Mechtler K, Westphall MS, Coon JJ: Phosphoproteomics with Activated Ion Electron Transfer Dissociation. *Anal. Chem* 89, 6367–6376 (2017)

28. Glover MS, Yu Q, Chen Z, Shi X, Kent KC, Li L: Characterization of intact sialylated glycopeptides and phosphorylated glycopeptides from IMAC enriched samples by EThcD fragmentation: Toward combining phosphoproteomics and glycoproteomics. *Int. J. Mass Spectrom* 427, 35–42 (2018)

29. Yang Y, Liu F, Franc V, Halim LA, Schellekens H, Heck AJR: Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun* 7 (2016); 13397: 1–10

30. Yu Q, Shi X, Feng Y, Kent KC, Li L: Improving data quality and preserving HCD-generated reporter ions with EThcD for isobaric tag-based quantitative proteomics and proteome-wide PTM studies. *Anal. Chim. Acta* 968, 40–49 (2017)

31. Mommen GPM, Frese CK, Meiring HD, van Gaans-van den Brink J, de Jong APJM, van Els CACM, Heck AJR: Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc. Natl. Acad. Sci. U.S.A* 111, 4507–4512 (2014)

32. Yu Q, Wang B, Chen Z, Urabe G, Glover MS, Shi X, Guo L-W, Kent KC, Li L: Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. *J. Am. Soc. Mass Spectrom* 28, 1751–1764 (2017)

33. Yu Q, Canales A, Glover MS, Das R, Shi X, Liu Y, Keller MP, Attie AD, Li L: Targeted Mass Spectrometry Approach Enabled Discovery of O-Glycosylated Insulin and Related Signaling Peptides in Mouse and Human Pancreatic Islets. *Anal. Chem* 89, 9184–9191 (2017)

34. Diedrich JK, Pinto AFM, Yates JR III: Energy dependence of HCD on peptide fragmentation: Stepped collisional energy finds the sweet spot. *J. Am. Soc. Mass Spectrom* 24, 1690–1699 (2013)

35. Yang H, Yang C, Sun T: Characterization of glycopeptides using a stepped higher-energy C-trap dissociation approach on a hybrid quadrupole orbitrap. *Rapid Commun. Mass Sp* 32, 1353–1362 (2018)

36. Palmisano G, Lendal SE, Engholm-Keller K, Leth-Larsen R, Parker BL, Larsen MR: Selective enrichment of sialic acid–containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc* 5, 1974–1982 (2010)

37. Kweon HK, Håkansson K: Selective Zirconium Dioxide-Based Enrichment of Phosphorylated Peptides for Mass Spectrometric Analysis. *Anal. Chem* 78, 1743–1749 (2006)

38. Feng S, Ye M, Zhou H, Jiang X, Jiang X, Zou H, Gong B: Immobilized Zirconium Ion Affinity Chromatography for Specific Enrichment of Phosphopeptides in Phosphoproteome Analysis. *Mol. Cell. Proteomics* 6, 1656–1665 (2007)

39. Zhou H, Ye M, Dong J, Corradini E, Cristobal A, Heck AJR, Zou H, Mohammed S: Robust phosphoproteome enrichment using monodisperse microsphere–based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc* 8, 461–480 (2013)

40. Hong Y, Yao Y, Zhao H, Sheng Q, Ye M, Yu C, Lan M: Dendritic Mesoporous Silica Nanoparticles with Abundant Ti4+ for Phosphopeptide Enrichment from Cancer Cells with 96% Specificity. *Anal. Chem* 90, 7617–7625 (2018)

41. Iliuk AB, Martin VA, Alicie BM, Geahlen RL, Tao WA: In-depth analyses of kinase-dependent tyrosine phosphoproteomes based on metal ion functionalized soluble nanopolymers. *Mol. Cell. Proteomics* 9, 2162–2172 (2010)

42. Tao WA, Wollscheid B, O'Brien R, Eng JK, Li X. j., Bodenmiller B, Watts JD, Hood L, Aebersold R: Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. *Nat. Methods* 2, 591–598 (2005)

43. Zhu F, Trinidad JC, Clemmer DE: Glycopeptide Site Heterogeneity and Structural Diversity Determined by Combined Lectin Affinity Chromatography/IMS/CID/MS Techniques. *J. Am. Soc. Mass Spectrom* 26, 1092–1102 (2015)

44. Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, Kasai K.-i., Takahashi N, Isobe T: Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol* 21, 667–672 (2003)

45. Chen R, Jiang X, Sun D, Han G, Wang F, Ye M, Wang L, Zou H: Glycoproteomics Analysis of Human Liver Tissue by Combination of Multiple Enzyme Digestion and Hydrazide Chemistry. *J. Proteome Res* 8, 651–661 (2009)

46. Zhang H, Li X.-j., Martin DB, Aebersold R: Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol* 21, 660–666 (2003)

47. Tian Y, Zhou Y, Elliott S, Aebersold R, Zhang H: Solid-phase extraction of N-linked glycopeptides. *Nat. Protoc* 2, 334–339 (2007)

48. Zhou Y, Aebersold R, Zhang H: Isolation of N-Linked Glycopeptides from Plasma. *Anal. Chem* 79, 5826–5837 (2007)

49. Chen W, Smeekens JM, Wu R: A universal chemical enrichment method for mapping the yeast N-glycoproteome by MS. *Mol. Cell. Proteomics* 13, 1563–1572 (2014)

50. Xiao H, Chen W, Smeekens JM, Wu R: An enrichment method based on synergistic and reversible covalent interactions for large-scale analysis of glycoproteins. *Nat. Commun* 9 (2018); 1692: 1–12

51. Xiao H, Wu R: Simultaneous Quantitation of Glycoprotein Degradation and Synthesis Rates by Integrating Isotope Labeling, Chemical Enrichment, and Multiplexed Proteomics. *Anal. Chem* 89, 10361–10367 (2017)

52. Mysling S, Palmisano G, Højrup P, Thaysen-Andersen M: Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. *Anal. Chem* 82, 5598–5609 (2010)

53. Hägglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P: A New Strategy for Identification of N-Glycosylated Proteins and Unambiguous Assignment of Their Glycosylation Sites Using HILIC Enrichment and Partial Deglycosylation. *J. Proteome Res* 3, 556–566 (2004)

54. Calvano CD, Zambonin CG, Jensen ON: Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry. *J. Proteomics* 71, 304–317 (2008)

55. Yu L, Li X, Guo Z, Zhang X, Liang X: Hydrophilic Interaction Chromatography Based Enrichment of Glycopeptides by Using Click Maltose: A Matrix with High Selectivity and Glycosylation Heterogeneity Coverage. *Chem. – A Eur. J* 15, 12618–12626 (2009)

56. Xu D, Yan G, Gao M, Deng C, Zhang X: Highly selective SiO2–NH2@TiO2 hollow microspheres for simultaneous enrichment of phosphopeptides and glycopeptides. *Anal. Bioanal. Chem* 409, 1607–1614 (2017)

57. Zou X, Jie J, Yang B: Single-Step Enrichment of N-Glycopeptides and Phosphopeptides with Novel Multifunctional Ti4+-Immobilized Dendritic Polyglycerol Coated Chitosan Nanomaterials. *Anal. Chem* 89, 7520–7526 (2017)

58. Hong Y, Zhao H, Pu C, Zhan Q, Sheng Q, Lan M: Hydrophilic Phytic Acid-Coated Magnetic Graphene for Titanium(IV) Immobilization as a Novel Hydrophilic Interaction Liquid Chromatography-Immobilized Metal Affinity Chromatography Platform for Glyco- and Phosphopeptide Enrichment with Controllable Selectivity. *Anal Chem* 90, 11008–11015 (2018)

59. Di Palma S, Hennrich ML, Heck AJR, Mohammed S: Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J. Proteomics* 75, 3791–3813 (2012)

60. Taouatas N, Altelaar AFM, Drugan MM, Helbig AO, Mohammed S, Heck AJR: Strong Cation Exchange-based Fractionation of Lys-N-generated Peptides Facilitates the Targeted Analysis of Post-translational Modifications. *Mol. Cell. Proteomics* 8, 190–200 (2009)

61. Hennrich ML, Groenewold V, Kops GJPL, Heck AJR, Mohammed S: Improving Depth in Phosphoproteomics by Using a Strong Cation Exchange-Weak Anion Exchange-Reversed Phase Multidimensional Separation Approach. *Anal. Chem* 83, 7137–7143 (2011)

62. Mohammed S, Heck AJR: Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. *Curr. Opin. Biotechnol* 22, 9–16 (2011)

63. Alpert AJ: Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal. Chem* 80, 62–76 (2008)

64. Alpert AJ, Hudecz O, Mechtler K: Anion-exchange chromatography of phosphopeptides: weak anion exchange versus strong anion exchange and anion-exchange chromatography versus electrostatic repulsion-hydrophilic interaction chromatography. *Anal. Chem* 87, 4704–4711 (2015)

65. Totten SM, Feasley CL, Bermudez A, Pitteri SJ: Parallel Comparison of N-Linked Glycopeptide Enrichment Techniques Reveals Extensive Glycoproteomic Analysis of Plasma Enabled by SAX-ERLIC. *J. Proteome Res* 16, 1249–1260 (2017)

66. Yang W, Shah P, Hu Y, Toghi Eshghi S, Sun S, Liu Y, Zhang H: Comparison of Enrichment Methods for Intact N- and O-Linked Glycopeptides Using Strong Anion Exchange and Hydrophilic Interaction Liquid Chromatography. *Anal. Chem* 89, 11193–11197 (2017)

67. Tran TH, Hwang I-J, Park J-M, Kim J-B, Lee H-K: An Application of Electrostatic Repulsion Hydrophilic Interaction Chromatography in Phospho- and Glycoproteome Profiling of Epicardial Adipose Tissue in Obesity Mouse. *Mass Spectro. Lett* 3, 39–42 (2012)

68. Zhang H, Guo T, Li X, Datta A, Park JE, Yang J, Lim SK, Tam JP, Sze SK: Simultaneous Characterization of Glyco- and Phosphoproteomes of Mouse Brain Membrane Proteome with Electrostatic Repulsion Hydrophilic Interaction Chromatography. *Mol. Cell. Proteomics* 9, 635–647 (2010)

69. Hao P, Guo T, Sze SK: Simultaneous Analysis of Proteome, Phospho- and Glycoproteome of Rat Kidney Tissue with Electrostatic Repulsion Hydrophilic Interaction Chromatography. *PLoS One* 6, e16844 (2011)

70. Ge F, Tao S, Bi L, Zhang Z, Zhang XE: Proteomics: addressing the challenges of multiple myeloma. *Acta Bioch. Bioph. Sin* 43, 89–95 (2011)

71. Reiding KR, Blank D, Kuijper DM, Deelder AM, Wuhrer M: High-Throughput Profiling of Protein N-Glycosylation by MALDI-TOF-MS Employing Linkage-Specific Sialic Acid Esterification. *Anal. Chem* 86, 5784–5793 (2014)

72. Selman MHJ, Hemayatkar M, Deelder AM, Wuhrer M: Cotton HILIC SPE Microtips for Microscale Purification and Enrichment of Glycans and Glycopeptides. *Anal. Chem* 83, 2492–2499 (2011)

73. Darula Z, Sherman J, Medzihradszky KF: How to Dig Deeper? Improved Enrichment Methods for Mucin Core-1 Type Glycopeptides. *Mol. Cell. Proteomics* 11, 1–10 (2012)

74. Alpert AJ: Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr. A* 499, 177–196 (1990)

75. Hemström P, Irgum K: Hydrophilic interaction chromatography. *J. Sep. Sci* 29, 1784–1821 (2006)

76. Thannhauser TW, Shen M, Sherwood R, Howe K, Fish T, Yang Y, Chen W, Zhang S: A workflow for large-scale empirical identification of cell wall N-linked glycoproteins of tomato (Solanum lycopersicum) fruit by tandem mass spectrometry. *Electrophoresis* 34, 2417–2431 (2013)

77. Ding W, Hill JJ, Kelly J: Selective Enrichment of Glycopeptides from Glycoprotein Digests Using Ion-Pairing Normal-Phase Liquid Chromatography. *Anal. Chem* 79, 8891–8899 (2007)

78. Ding W, Nothaft H, Szymanski CM, Kelly J: Identification and Quantification of Glycoproteins Using Ion-Pairing Normal-phase Liquid Chromatography and Mass Spectrometry. *Mol. Cell. Proteomics* 8, 2170–2185 (2009)

79. Wimley WC, Gawrisch K, Creamer TP, White SH: Direct measurement of salt-bridge solvation energies using a peptide model system: implications for protein stability. *PNAS* 93, 2985 (1996)

80. Alpert AJ: Effect of salts on retention in hydrophilic interaction chromatography. *J. Chromatogr. A* 1538, 45–53 (2018)

Table 1. ERLIC loading buffer composition

Condition	Organic phase	Ion pairing regent	pH	Notes
a	80% ACN	1% TFA	1	
b	80% ACN	0.1% TFA	2	
c	95% ACN	1% TFA	1	
d	95% ACN	0.1% TFA	2	
e	95% ACN	0.1% FA	2.7	
f	95% ACN	1% FA	2	
g	80% ACN	1% FA	2	
h	80% ACN	0.1% FA	2.7	
i	80% ACN	1% TFA	1	Without ammonium acetate conditioning
j	80% ACN	1% TFA	1	Activated in 80% ACN 1% TFA

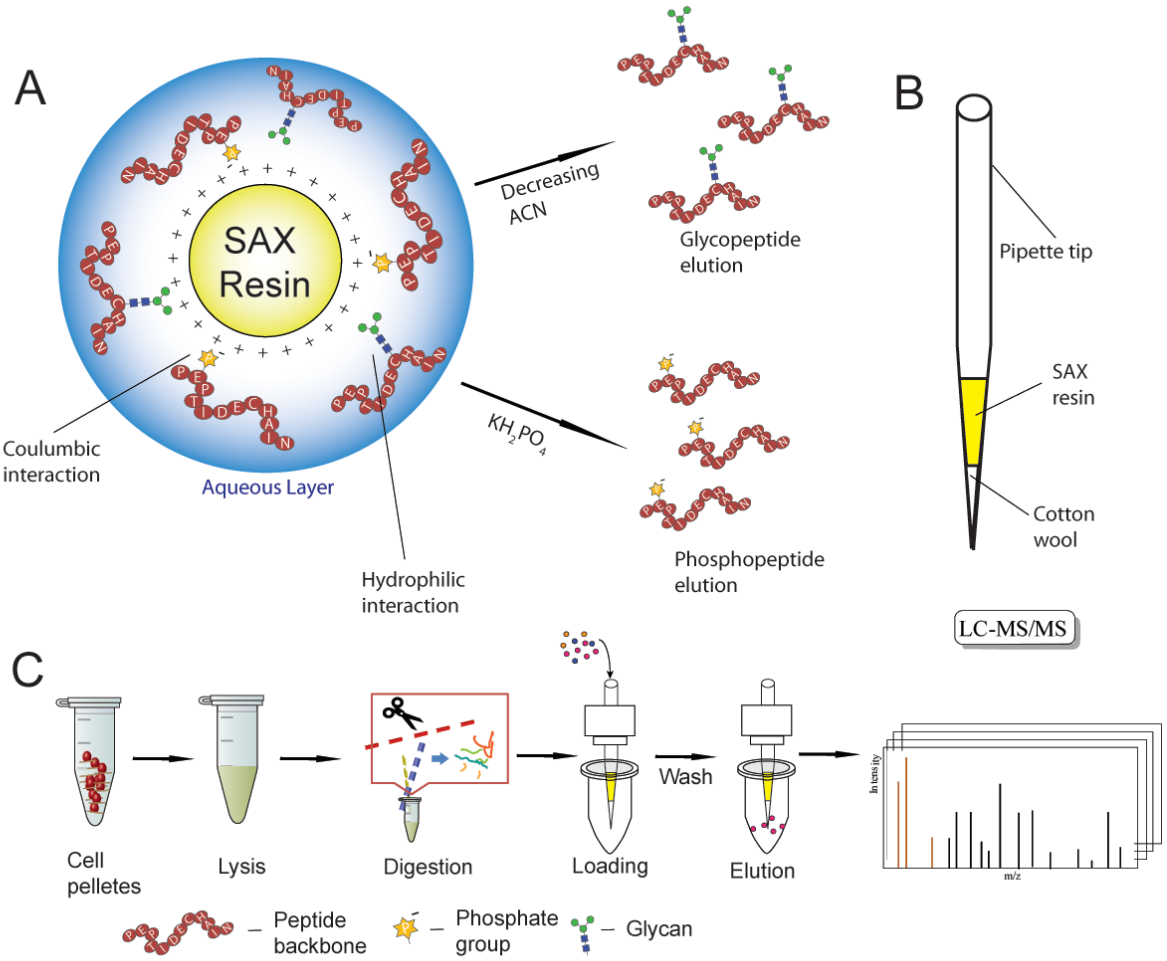

Figure 1. (A) Illustration of ERLIC enrichment and separation mechanism of phosphopeptides and glycopeptides; (B) Custom-made spin-tip composition; (C) Schematic workflow of spin-tip-based ERLIC SPE coupled with LC-MS/MS analysis on MM.1S cell lysate digest.

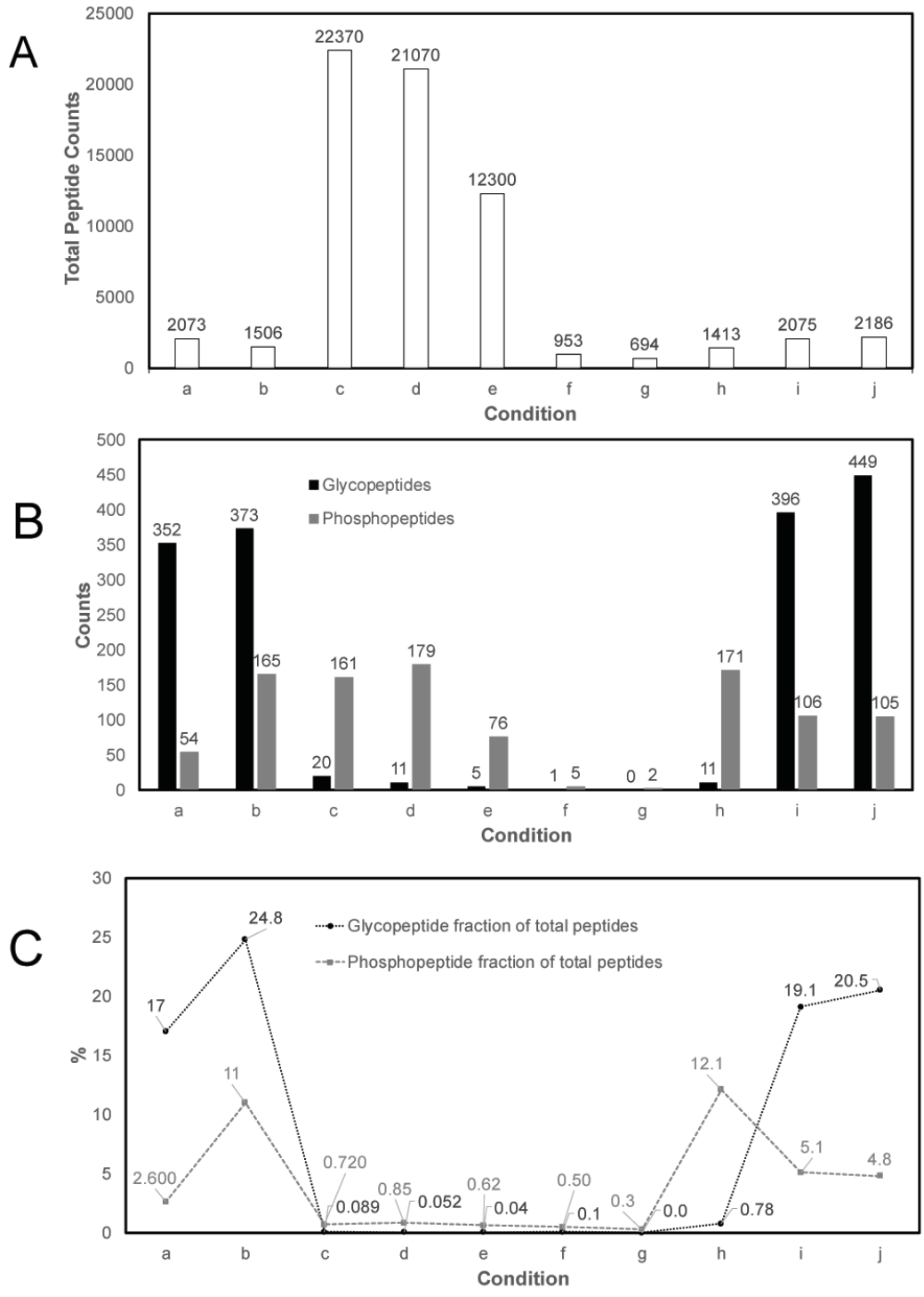

Figure 2. Identification results of peptide modifications in 100 µg of MM.1S cell lysate by ERLIC fractionation. ERLIC loading buffer conditions are provided in Table 1. (A) Total peptide identification; (B) Glycopeptide and phosphopeptide identification; (C) % specificity of ERLIC enrichment for each condition, calculated by dividing the number of modified peptides identified divided by the total number of peptides.

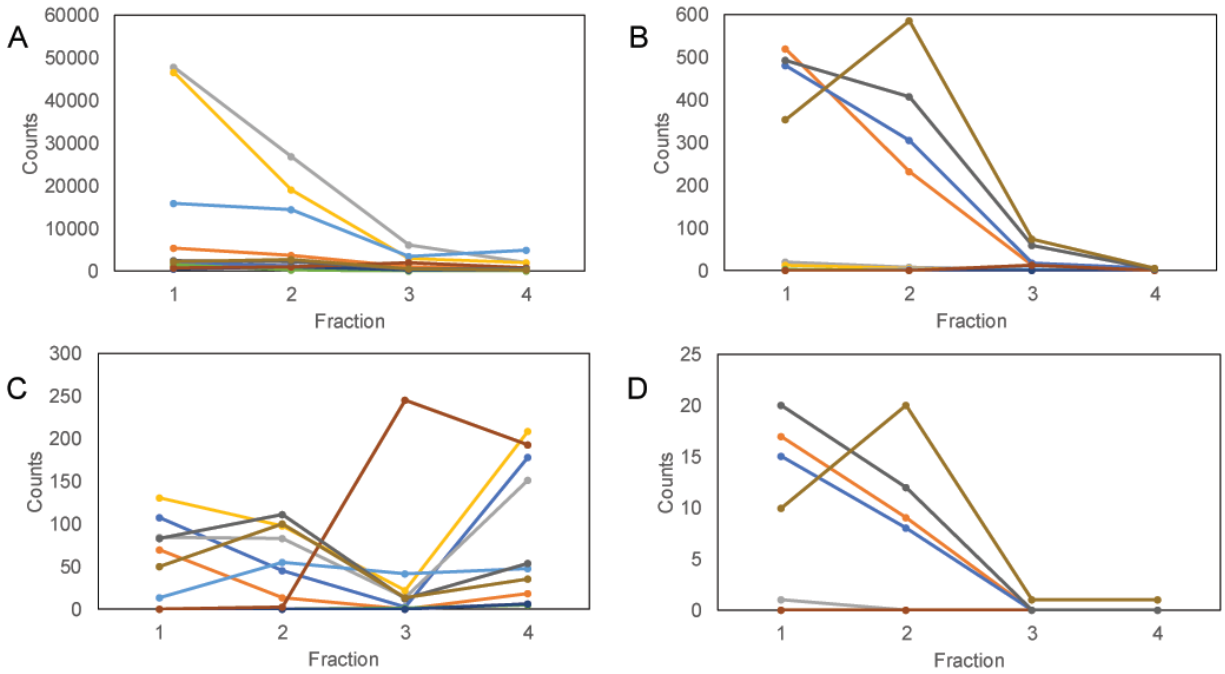

Figure 3. PSM distribution of modifications among fractions. ERLIC loading buffer conditions are given in Table 1. Eluting buffer conditions are given in Table 2. (A) total PSMs; (B) glycopeptide PSMs; (C) phosphopeptide PSMs; (D) Sialic acid containing glycopeptide PSMs.

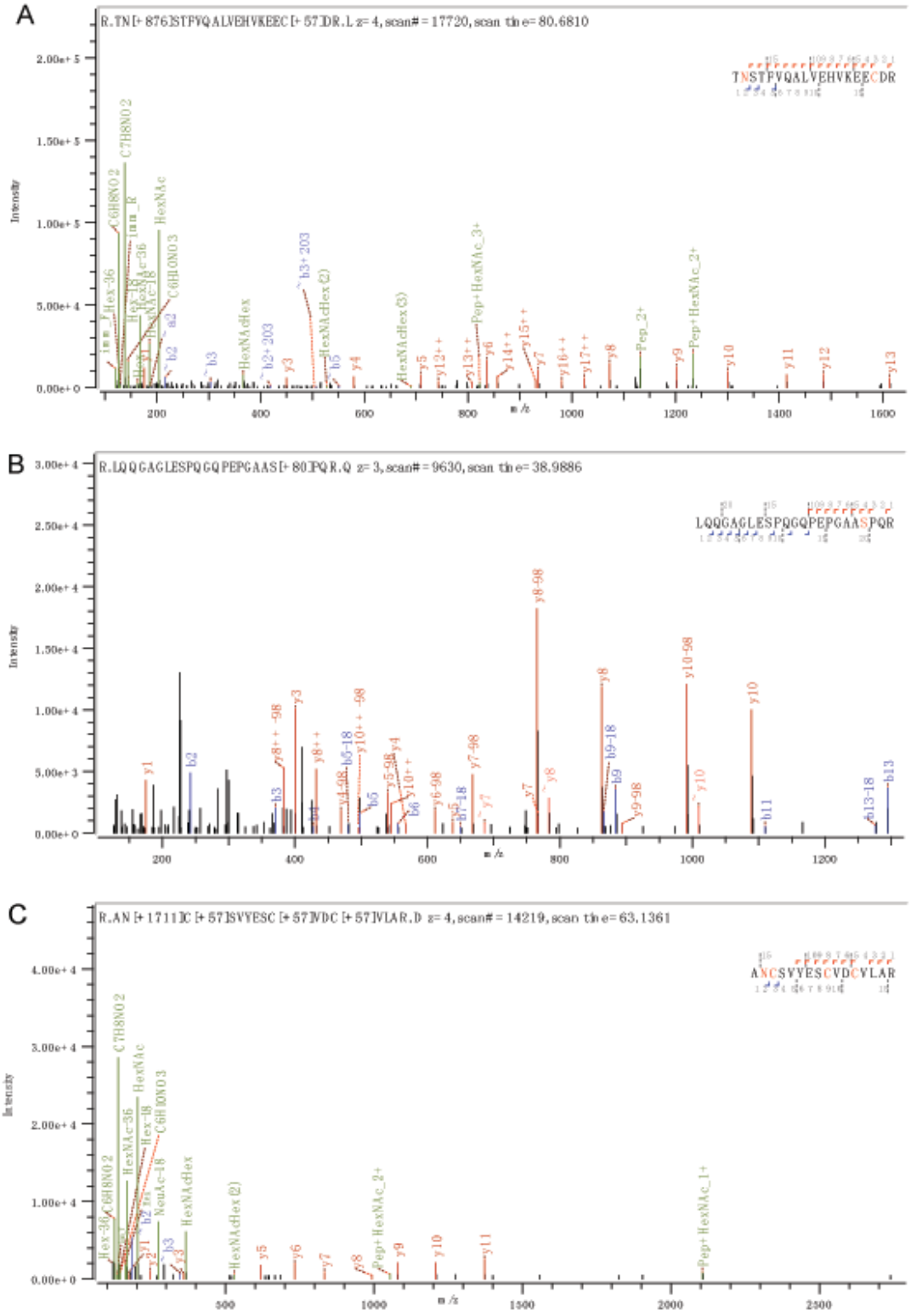

Figure 4. Representative stepped HCD MS/MS spectra with site-specific identifications of glycopeptides, phosphopeptides and those with sialic acid containing glycopeptides. (A) MS/MS

of N-glycopeptide TN[+876.32]STFVQALVEHVKEECDR (AA 214–232) from human prosaposin, spectra of +4 charged precursor at *m/z* 785.36; (B) MS/MS of phosphopeptide LQQGAGLESPQGQPEPGAAS[+79.97]PQR (AA 72–94) from human coiled-coil domain containing protein 86, spectra of +3 charged precursor at *m/z* 795.04; (C) MS/MS of sialylated glycopeptide AN[+2204.77]C[+57.02]SVYESC[+57.02]VDC[+57.02]VLAR (AA 495–510) from human semaphorin-4A, spectra of +4 charged precursor at *m/z* 1027.65; (I) *m/z* 150 to 410 of (C) with oxonium ions and NeuAc reporter ions resolved; (II) 7-fold zoom-in of (C) at *m/z* 410 to 2000 with most of the peptide backbone fragment ions being detected and resolved.

Chapter 6

Counterion Optimization Dramatically Improves Selectivity for Phosphopeptides and Glycopeptides in Electrostatic Repulsion-Hydrophilic Interaction Chromatography

Adapted from: **Cui, Y.**; Tabang, D. N.; Zhang, Z.; Ma, M.; Alpert, A. J.; Li, L. Counterion Optimization Dramatically Improves Selectivity for Phosphopeptides and Glycopeptides in Electrostatic Repulsion-Hydrophilic Interaction Chromatography. *Anal Chem* 2021, *93* (22), 7908–7916. https://doi.org/10.1021/acs.analchem.1c00615.

Abstract

A well-hydrated counterion can selectively and dramatically increase retention of a charged analyte in hydrophilic interaction chromatography. The effect is enhanced if the column is charged, as in electrostatic repulsion-hydrophilic interaction chromatography (ERLIC). This combination was exploited in proteomics for the isolation of peptides with certain post-translational modifications (PTMs). The best salt additive examined was magnesium trifluoroacetate. The well-hydrated Mg^{+2} ion promoted retention of peptides with functional groups that retained negative charge at low pH, while the poorly hydrated trifluoroacetate counterion tuned down the retention due to the basic residues. The result was an enhancement in selectivity ranging from 6- to 66-fold. These conditions were applied to a tryptic digest of mouse cortex. Gradient elution produced fractions enriched in peptides with phosphate, mannose-6-phosphate, and N- and O-linked glycans. The numbers of such peptides identified either equaled or exceeded the numbers afforded by the best alternative methods. This method is a productive and convenient way to isolate peptides simultaneously that contain a number of different PTMs, facilitating study of proteins with "crosstalk" modifications. The fractions from the ERLIC column were desalted prior to C-18-reversed phase liquid chromatography–tandem mass spectrometry analysis. Between 47–100% of the peptides with more than one phosphate or sialyl residue or with a mannose-6 phosphate group were not retained by a C-18 cartridge but were retained by a cartridge of porous graphitic carbon. This finding implies that the abundance of such peptides may have been significantly underestimated in some past studies.

Introduction

Advances in bottom-up proteomics depend on the identification of peptides with a variety of post-translational modifications (PTMs). Usually, an affinity chromatography method is used to isolate peptides with one type of PTM. Titanium dioxide or immobilized metal affinity chromatography (IMAC) is used for phosphopeptides, sometimes sequentially,[1,2] and can also be used for sialylated and phosphorylated glycopeptides.[3,4] Glycopeptides can be isolated using lectins or affinity materials based on boronic acids or esters.[5] These affinity methods all have shortcomings. Lectins and boronates only retain subsets of glycans. Titanium dioxide and IMAC also have difficulty either with retention or elution of all phosphopeptides and glycopeptides (especially regarding titanium dioxide and multiphosphorylated peptides). General-purpose methods tend to lack the selectivity necessary for effective identification of modified peptides of low abundance. Examples are attempts to isolate tryptic phosphopeptides using either conventional anion-exchange chromatography[6] or hydrophilic interaction chromatography (HILIC).[7] Neither method is selective enough to offer a satisfactory enrichment factor for phosphopeptides. An advance was introduced by Ding et al.[8] using trifluoroacetic acid (TFA) as an ion pairing agent in HILIC. Basic residues are normally the most hydrophilic residues in peptides.[9] TFA forms relatively hydrophobic ion pairs tenaciously with basic residues, which tunes down their contribution to retention in HILIC. The result is to render more important the contribution of the rest of the peptides to retention. In Ding's case, this involved the glycan portion of glycopeptides. While the selectivity was appreciably lower than that associated with an affinity material, it was still good enough to produce a clean separation of glycopeptides from nonglycopeptides on a neutral HILIC column. The weakness of the affinity for the glycan portion is actually an advantage; because the affinity interaction does not dominate the

chromatography, there is a wide dynamic range in the selectivity window. A HILIC column can separate glycopeptides with high resolution.[10]

The electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) variation of HILIC was introduced in 2008.[11] In the case of peptide separation, this usually involves an anion-exchange column being operated under HILIC conditions at a pH low enough to uncharge the carboxyl-residues of Asp and Glu residues. Peptides then have a net positive charge and are repelled by the stationary phase electrostatically. With sufficient organic solvents, the hydrophilic interaction can balance the repulsion, causing the peptides to elute in or near the void volume. An exception occurs with peptides containing groups that retain some negative charge at the low pH. This includes phosphate groups (pK_a ~2.1), sialyl groups (pK_a ~2.6), and isoaspartyl groups (pK_a ~3.1) resulting from deamidation of Asn. These groups confer electrostatic attraction in addition to the hydrophilic interaction, and peptides containing these groups can be pulled away from the rest of the peptides in a complex mixture, such as a digest. ERLIC being a variant of HILIC, peptides that lack PTMs but that are particularly polar will also be well retained. This applies as well for nonsialylated glycopeptides. While ERLIC was initially used for selective isolation of phosphopeptides, the well-retained fractions were soon found to be enriched in both phosphopeptides and glycopeptides.[12−15] Both glycosylation and phosphorylation of proteins have been implicated in various human diseases, including diabetes, Alzheimer's disease (AD), autoimmunity, and cancer. There is complex crosstalk between the two PTMs, including competitive or noncompetitive occupancy of the same or proximal sites, glycosylation of kinases, and phosphorylation of glycosylation-related enzymes. The simultaneous identification of peptides with different PTMs provides an opportunity to identify those crosstalk interactions, in which toggling between two (or more) modifications on nearby

residues controls physiological responses.[3,16,17] Other PTMs have been implicated in crosstalk as well, including acetylation, methylation, and SUMOylation.[18,19]

A recent study[20] has demonstrated that the hydration of counterions plays a key role in the retention of charged analytes in HILIC. This permits manipulation of retention and selectivity by the choice of the salt used in the mobile phase (MP). The effect is even more extreme in ERLIC. Retention and selectivity for peptides with negatively charged groups, such as phosphopeptides, increase several fold when Na^+ or NH_4^+ is replaced by the well-hydrated Mg^{+2} in ERLIC. In the present study, this increase in the retention of peptides with negatively charged PTMs is combined with the use of the poorly hydrated trifluoroacetate anion to tune down retention due to the rest of the peptide, as per Ding et al. The result is an unprecedented increase in selectivity for peptides with a number of important PTMs, to a degree that compares favorably with or exceeds the enrichment factors conferred by other methods in the literature, providing a convenient tool for PTM crosstalk analysis.

Experimental Section

Offline ERLIC Fractionation

Sample preparation followed previously described protocols with moderate adjustments.[14] The details of chromatography materials, MP preparation, cortex tryptic digest preparation, and liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis are provided in the Supporting Information.

Mouse cortex tryptic digest (1 mg) was fractionated using either a PolyWAX LP or PolySAX LP column (200 mm × 2.1 mm, 5 μm, 300 Å; PolyLC, Columbia, MD). Three sets of binary MP

combinations were used in a total of six fractionation experiments: (1) MP A1: 20 mM sodium methylphosphonate with 80% ACN, pH = 2.5; MP B1: 300 mM triethylammonium phosphate with 10% ACN, pH = 2.5; (2) MP A2: 20 mM magnesium trifluoroacetate with 80% ACN, pH = 2.5; MP B2: 300 mM triethylammonium trifluoroacetate with 10% ACN, pH = 2.5; and (3) MP A3: 20 mM ethylenediamine trifluoroacetate with 80% ACN, pH = 2.5; MP B3: 300 mM ammonium trifluoroacetate, 10% ACN, pH = 2.5. The gradient was as follows: 0–5 min, 0% B; 5–48 min, 0–100% B; 48–53 min, 100% B. The column was re-equilibrated at 100% A for 15 min between runs. The eluent was collected with a FC-4 fraction collector (Rainin Dynamax) at 1 min intervals as initial fractions. Between 6–30 min (MP1) and 7–30 min (MP2 and MP3), adjacent 2 min eluents were pooled into the first 12 fractions. Eluents between 30–42 min and 42–53 min were pooled into the last two fractions 13 and 14. Each fraction was dried under vacuum.

Solid-Phase Extraction (SPE)/Desalting

Every peptide fraction from the six experiments was desalted with OMIX C18 tips (Agilent, Santa Clara, CA). For the PTM crosstalk analysis using MP combination 2, C18 SPE flow-through was collected and run through an additional HyperCarb SPE (TopTips; PolyLC) procedure as per the manufacturer's instructions. Eluates from SPE were dried and stored at −80 °C prior to mass spectrometry (MS) analysis.

Data Analysis

Raw files were processed with the Byonic search engine (Protein Metrics Inc., San Carlos, CA) embedded within Proteome Discoverer 2.1 (Thermo Fisher Scientific, San Jose, CA). Spectra were searched against the SwissProt *Mus musculus* proteome database (August 13, 2016; 24,903 entries). Trypsin digestion-missed cleavages were set at <3. The parent mass error

tolerance was 10 ppm, and fragment mass tolerance was 0.01 Da. Fixed modifications were specified as carbamidomethylation (+57.02146 Da) on C residues. Dynamic modifications included oxidation of M (+15.99492 Da, rare1), deamidation (+0.984016 Da, rare1) of N or Q, N-glycosylation (common1), O-glycosylation (common1), and phosphorylation (common4). Glycan modifications were specified as the Byonic embedded mammalian N-glycan database (309 entries) expanded with typical mannose-6-phosphate glycans including HexNAc (2) Hex (4-9) Phospho (1-2) modification, HexNAc (3-4) Hex (4-9) Phospho (1-2), HexNAc (2) Hex (3-4) Phospho (1), and HexNAc (3) Hex (3-4) Phospho (1)). Identifications were filtered to 1% protein false discovery rate (FDR). Similarly, O-glycosylation (Byonic embedded mammalian O-glycan database with 89 O-glycans) was set as a common dynamic modification and searched separately.

Confident crosstalk interaction analysis relies on confident identification of PTM sites. To minimize false discoveries of modified peptides, post-Byonic search filtering was applied with in-house written scripts. Briefly, identifications were filtered to 1% PSM FDR. For method comparison, a Byonic score cutoff >100 was used. More stringent filtering of N-glycopeptides and phosphopeptides used a Byonic score cutoff >300, delta Byonic score >50, and |log10 Prob| >4 (the absolute value of the log10 posterior error probability). For the glycan–protein network, N-glycopeptides were exclusively categorized into six glycan type categories based on the glycan composition: (1) mannose-6-phosphate (containing phosphate on the glycan), (2) sialic acid (containing NeuAc/NeuGc), (3) fucose (containing fucose), (4) complex/hybrid (>2 HexNAc), (5) high-mannose (2 HexNAc and >5 Hex), and (6) paucimannose (2 HexNAc and <5 Hex). Figures were created in R 3.6.0 using ggplot2, igraph, networkD3, and ggnetwork library.

Results

Method Development with Peptide Standards

Figure S1 depicts the partitioning of a tryptic phosphopeptide and an unmodified tryptic peptide into the semi-immobilized aqueous layer in HILIC. A well-hydrated cation, such as Mg^{+2}, promotes retention of peptides through negatively charged counterions such as phosphate. A well-hydrated anion—here, methylphosphonate—does the same through the positively charged residues. Figure S2 depicts the situation when a poorly hydrated anion such as trifluoroacetate is used in place of methylphosphonate. While the phosphopeptide remains well retained via the association with the Mg^{+2} ion, no counterion confers similar retention upon the unmodified peptide. Figure 1A compares retention of synthetic peptides, with sequences similar to peptides resulting from digestion with trypsin, using various salt additives in the ERLIC mode. Substitution of magnesium trifluoroacetate for magnesium methylphosphonate results in a modest decrease in retention of the singly phosphorylated peptide but a much more dramatic decrease in retention of the unmodified peptide. The result is an increase in the selectivity ratio from 3.5:1 to 13:1 (the ratio of retention of a phosphopeptide to retention of the corresponding peptide with an Asp residue in place of the pSer). Ethylenediamine is perhaps the most hydrophilic volatile amine,[21] and so it was substituted for Mg^{+2} to implement the separation with a volatile MP. The result was a decrease in the selectivity ratio to 9:1. Subsequent study determined that ethylenediammonium trifluoroacetate was insufficiently volatile for use in MS. In retrospect, this is not surprising, since triethylammonium trifluoroacetate is an ionic liquid.[22]

In two chromatograms shown in Figure 1A, positional isomers were separated that had the phosphate group on different serine residues [the order of elution was the same in both runs].

With three of the isomers, the further the phosphate group was from the C-terminus, the earlier the elution. This is consistent with an earlier study[23] that showed that tryptic peptides tend to be oriented on an anion-exchange material with the C-terminus facing the stationary phase and the N-terminus facing away. Thus, the more remote the location of a phosphate group from the C-terminus, the less it can interact with the stationary phase, reducing its contribution to retention. The least well-retained positional isomer is an exception to this trend. In this case, the phosphoserine followed a proline residue in the primary sequence and was probably sequestered by it.

Figure 1B compares the separation of these peptide standards by both ERLIC and HILIC. ERLIC clearly affords better retention and selectivity of phosphopeptides than HILIC. It is also more sensitive to positional isomers, because HILIC does not feature the high degree of orientation of peptides that can be conferred by their interaction with a charged surface.

Complex tryptic digests contain a significant number of peptides with three or more Asp and Glu residues. Phosphopeptides are of much lower abundance. They must be physically separated from the peptides with three or more ordinary acidic residues, which would otherwise tend to suppress their ionization in MS.[6]Figure 2 shows that the conditions being introduced here have the necessary selectivity. Even the least well-retained singly phosphorylated peptide standard is separated from an analogous peptide in which the four serine residues have all been replaced by Asp residues. This comfortable separation of singly phosphorylated peptides from unmodified acidic peptides helps to address potential concerns about the utility of ERLIC.[24] A gradient to 1% unbuffered TFA and 30% ACN tunes down both the electrostatic attraction and hydrophilic interaction and permits the elution of peptides with two, three, or four phosphate residues.

The excellent selectivity for phosphopeptides enabled by ERLIC motivated us to compare it with the selectivity conferred by affinity materials. SPE cartridges of titanium dioxide are frequently used for isolation of phosphopeptides from complex digests. We packed a high-performance liquid chromatography (HPLC) column with titanium dioxide and evaluated it with gradient elution of our peptide standards. The results are shown in Figure S3. A peptide standard with no acidic residues was not retained. Addition of a single Asp residue resulted in a modest degree of retention with elution in a broad peak, because carboxyl-groups are weak Lewis bases and so interact weakly with titanium dioxide. The interaction of titanium dioxide with phosphate groups is much stronger because they are strong Lewis bases. It was necessary to run a gradient to ammonium hydroxide to elute any phosphopeptides. Both singly phosphorylated isomeric peptide standards then coeluted, along with the peptide standards with two, three, or four Asp residues. With continuation of the gradient, a peptide standard with two phosphate groups was resolved from the other acidic peptides. The interaction of titanium dioxide with Lewis bases is so strong that it blinds the material to any other aspect of the peptide's composition. These results are consistent with those from the only paper we could find in the literature where a titanium dioxide HPLC column was used with a similar gradient for peptides.[25] Peptides eluted in very wide peaks, and the best-retained peptides were difficult to elute at all. This accounts to some extent for the "lossy" nature of titanium dioxide regarding multiply phosphorylated peptides, as shown in Figure 2. Titanium dioxide has also been shown to be inferior to HILIC for isolation and identification of N-glycopeptides.[26,27] Phosphopeptides eluted more cleanly from an HPLC column packed with an Fe-IMAC material, but all in a single sharp peak within the gradient.[25]

Application of ERLIC Fractionation to the Mouse Cortex Phospho- and Glycoproteome

Proteins were extracted from mouse cortex and digested with trypsin, as described in the Experimental Section. The digest was fractionated via ERLIC using gradients with three alternative MP combinations:

- MP 1: Sodium methylphosphonate to triethylammonium (TEA) phosphate (TEAP).
- MP 2: Magnesium trifluoroacetate (Mg-TFA) to TEA-TFA.
- MP 3: Ethylenediammonium-TFA (EDA-TFA) to ammonium-TFA.

MP 1 was introduced in ref (11) and has since been widely used for the isolation of phospho- and glycopeptides. In ref (20), it was demonstrated that substitution of the Na^{+} ion by the better-hydrated Mg^{+2} ion as a counterion to phosphate groups resulted in much better retention of phosphopeptides in HILIC or ERLIC, while substitution with the nonhydrated tetramethylammonium ion resulted in a significant decrease in retention of phosphopeptides. This informed about the composition of MP 2, where combining Mg^{+2} and trifluoroacetate maximizes ERLIC selectivity for phospho- and glycopeptides over nonmodified counterparts. MP 3 was an attempt to develop a totally volatile MP that would permit direct ERLIC-MS analysis of peptides. In addition to the salt gradients, all three MPs featured a decreasing gradient of acetonitrile to tune down hydrophilic interaction.

Figure S4 compares the chromatograms obtained from 1 mg of mouse cortex digest with the three MPs. Ideally, in ERLIC the unmodified peptides elute in or near the void volume while the modified peptides are well retained and elute during the gradient. The profile associated with MP 2 appears to be the most promising in that regard. This seems to be substantiated in Figure 3, which compares the number of unmodified peptides identified using all three MPs with both the PolyWAX LP and PolySAX LP columns. With PolyWAX LP, the fewest identifications of unmodified tryptic peptides were obtained using MP 2. If most of the unmodified peptides do

elute in the same early fractions, then they would interfere with each other's ionization at the MS stage, reducing identifications.

Figure 3 also compares the two column materials and three MP combinations with peptides with PTMs. With MP 2 or 3, 5-6× more N- and O-linked glycopeptides were identified than with the MP 1 that has routinely been used with ERLIC. With phosphorylated peptides, a similar 6-fold increase in identifications was obtained when comparing MP2 to MP1 but not MP3. For peptides with mannose-6-PO_4 (M-6-P) groups, the number of identifications using MP 2 was 66-fold greater than with MP 1, a remarkable increase. In most cases, PolyWAX LP outperformed PolySAX LP. In view of these data, all further studies were performed using a PolyWAX LP column and MP 2. Redundant identifications of the same peptide in more than one fraction were deleted from the totals unless stated otherwise.

Figure S5B shows the distribution of modified and unmodified peptides among the 14 fractions collected during the gradient with the PolyWAX LP column. The tallies of peptides of a given type from all 14 segments are numbers larger than those in Figure 3. This is because the numbers are the combined totals of the C-18 desalted and porous graphitic carbon (PGC)-desalted peptide sets. Peptides identified in both sets were only counted once. The uniform distribution of peptides among the fractions in Figure S5B is an advantage of ERLIC.[12]Figure 4 focuses on the number of peptides with a given number of phosphate or sialyl residues, or an M-6-P group, in each fraction. Peptides with more phosphate and sialyl residues tend to elute later, as expected in HILIC in general. Also, this being a variant of HILIC, then in addition to peptides with PTMs, the latter fractions also contain a number of nonmodified peptides with numerous polar residues. These include

SGGNGGGGGGGGGGGGGYGGSGGGGGGAGVPSEGAAK,

SAAGSEKEEEPEEEEEEEEEYDEEEEEEDDDRPPK, and other extreme sequences.

Virtues of PGC

PGC retains particularly small or polar peptides that are poorly retained by C-18 material.[28] It has also been shown to retain some tryptic phosphopeptides that are not retained by C-18 material.[29] Accordingly, when the ERLIC fractions were individually desalted with C-18 cartridges, the filtrates were also collected and passed through PGC cartridges. This resulted in a significant increase in peptide identifications. Figure 4 shows the number and distribution of peptides with various modifications among the 14 fractions, with peptides identified from the C-18-desalted fractions tallied in separate graphs from peptides identified from PGC-desalted fractions. Peptides with M-6-P groups are an exception, being compared in the same graph in stacked segments. Surprisingly, almost half of the peptides with this PTM were not retained by C-18 and so were identified from the PGC-desalted fractions. Most of these peptides eluted in later fractions in ERLIC. This suggests that they were unusually polar, which could account for the lack of retention by the C-18 material. Similarly, peptides with >1 phosphate and >1 sialyl residue tended to elute in later fractions, with about half retained only by PGC. The relative distribution is shown in more detail in Figure S5A. This is consistent with ref (30), which showed separation of sialic acid isomers using PGC-LC–MS.

In a search for properties that might account for some of these trends, the average number of residues per peptide was graphed for each ERLIC fraction. An interesting trend of note (Figure S6) was that peptides retained only by PGC tended to be three–six residues shorter than peptides that were retained from the same fraction by C-18.

PTM Crosstalk

Because the salt-enhanced ERLIC method described here enables simultaneous detection of peptides with four different types of PTMs with good coverage, then one can search for crosstalk between different modifications. MP2 provides the best performance in terms of PTM selectivity and coverage, and PGC enrichment complements C-18 for PTM identification. The crosstalk analysis is based on data acquired with ERLIC fractionation using MP2, followed by postfractionation desalting and enrichment with both C-18 and PGC materials. With this method and with a Byonic score cutoff of 100, we have identified 12,492 nonredundant phosphopeptides, 10,218 phosphorylation sites in 3078 proteins, 3819 N-glycopeptides, 1246 N-glycosylation sites in 788 proteins, 2730 O-glycopeptides, and 1979 O-glycosylation sites in 1033 proteins. M6P peptide PSMs were manually inspected, and spectra without phospho-oxonium ions were filtered, after which 145 unique M6P peptides were identified. Figure S11A shows the distribution of phospho-oxonium ion containing M6P peptide PSMs. In Figure S7, 46% of the phosphorylation sites identified have been reported in the Uniprot database while the rest are novel. 50% of the N-glycosylation sites identified here have been annotated in Uniprot. However, few of the O-glycosylation sites are listed in Uniprot; 99% of them are novel. Among the 4697 Uniprot-annotated phosphosites, most (67%) were assigned using “combinatorial evidence”, which comes from large-scale proteomics data or protein 3D structure data without manual curation; 32% were assigned using “sequence similarity”, which means that the evidence is propagated from related proteins with similar sequences. For 627 annotated N-glycosites, 77% were only assigned via “sequence analysis”, which means that the glycosites were predicted from the presence of the N-X-S/T sequon rather than by experimental observation. This study has provided substantiating experimental evidence for those annotated sites, as well as the novel PTM sites not annotated in Uniprot. We present one example: Neural cell adhesion molecule

1. Figure 5A displays the complex PTM profile of this protein, depicting the PTMs this study identified in the protein as well as the ones already listed in Uniprot. Compared with the PTM sites reported in Uniprot, (a) Four out of six reported N-glycosylation sites were detected here; (b) 7 out of 11 reported phosphorylation sites were detected here (S890, T929, T1001, and T1030 were not detected). The following modifications were also detected, which have not been reported in Uniprot: (a) An O-glycosylation site; (b) Eight additional phosphorylation sites (S788, S884, S885, S886, S950, S952, S1016, and S1113); (c) A novel mannose-6-phosphate site.

With such extensive PTM data available at the proteome level from the same sample, this method will be valuable in dissecting PTM characteristics and their interplay efficiently. Site-specific microheterogeneity is an important aspect of glycosylation and significantly complicates the analysis of glycoproteins.[31,32] This information is lost when glycans and deglycosylated peptides are analyzed separately but is retained in the present method because the glycopeptides are analyzed intact. For example, 384 proteins identified here as O-glycosylated contained more than one O-glycosite on the protein, with 331 O-glycosites modified by two or more alternative glycans. Similarly, 233 N-glycoproteins were ascertained to have more than one N-glycosite, with 592 N-glycosites being modified with two or more alternative glycans. According to the simple linear regression model in Figure S8A, N-glycosylation has a higher degree of microheterogeneity than O-glycosylation, with an average of three alternative N-glycans per site. Recently, Riley et al. reported a visualization method for large-scale proteomics data.[24] Here, an N-glycan-protein network is plotted in Figure S8B, which maps different types of glycans attached to identified glycoproteins. Sialylated glycans comprise the largest fraction of the N-glycans identified. The network also shows that sialylated, fucosylated, and complex/hybrid

glycans occur more frequently on proteins with multiple glycosylation sites, while high-mannose, paucimannose, and M-6-P glycans are more uniformly distributed. Figure 5B plots the overlap of protein substrates of phosphorylation and glycosylation. Five hundred seventy-four proteins contain both phosphorylation and N-glycosylation sites, 453 proteins contain phosphorylation and O-glycosylation, 271 proteins contain both N- and O-glycosylation, and 167 proteins contain all three PTMs. Among the comodified proteins, there is positive correlation between phosphorylation and O-glycosylation sites and between N- and O-glycosylation sites per protein. However, such correlation was not observed between phosphorylation and N-glycosylation, as in the heatmap of Figure 5C. Figure 5D,E visualizes the site numbers of different PTMs on the same protein directly, with some hypermodified proteins' accessions labeled.

PTMs are known to alter protein functions. When there are multiple PTMs on the same protein, then their interplay can be quite complex. Crosstalk between O-glycosylation and phosphorylation has been reported, because they both target protein S/T residues.[17,33] In general, Wang et al. described four potential categories of crosstalk between O-GlcNAcylation and phosphorylation.[34] Canvassing the present data for these four categories (Figure 6):

(1) Direct competition for occupancy of a single site. In the present study, 154 O-glycosylation sites were also found to be phosphorylated;

(2) Competition via steric hindrance by reciprocal modification at proximal sites in a polypeptide. Considering sites within ±10 amino acid residues as proximal, then in this study 4728 (46%) phosphorylation sites have proximal phosphorylation, 123 (1.2%) have proximal N-glycosylation, and 331 (3.2%) have proximal O-glycosylation. Similarly, 73 (5.9%) N-glycosylation sites have proximal phosphorylation, 55 (4.4%) have proximal N-glycosylation,

and 200 (16%) have proximal O-glycosylation. For O-glycosylation sites, the figures are 248 (13%), 241 (12%), and 509 (26%) for proximal phosphorylation, N-glycosylation, and O-glycosylation, respectively;

(3) Regulation of O-glycosylation cycling enzymes by phosphorylation of their catalytic subunits or of their regulatory/targeting subunits;

(4) Regulation of kinases or phosphatases by GlcNAcylation.

Regarding categories (3) and (4), by querying the present data for gene ontology molecular function annotation, 226 phosphoproteins, 94 O-glycosylated proteins, and 52 N-glycosylated proteins were annotated with phosphatase activity, and 223 phosphoproteins, 64 O-glycosylated proteins, and 44 N-glycosylated proteins had kinase activity. This suggests that phosphorylation-cycling enzymes can be self-phosphorylated or glycosylated. Conversely, 164 phosphoproteins, 57 O-glycosylated proteins, and 35 N-glycosylated proteins were involved in protein glycosylation processes.

Aberrant protein phosphorylation and glycosylation have been associated with neurodegenerative diseases.[35,36] Interestingly, 43 phosphoproteins, 29 O-glycosylated proteins, and 15 N-glycosylated proteins in the present data were annotated as related to AD by the KEGG database (Supporting Table S1). Those key modified proteins are potential targets for studying the role of PTM regulation in AD progression.

Cutoff Score for the Byonic Filter

It has been suggested that the cutoff score for the Byonic filter be set at 300 for nonmodified peptides and in the range 50–200 for N-glycopeptides as a compromise between accuracy and coverage.[37] A cutoff score of 150 was used in a recent study of the N-glycoproteome.[24] In the present study, the data in Figure 5 and Figure S6 were obtained using the stringent cutoff score

of 300, while a cutoff score of 100 was used in Figures 3 and 4, and Figure S5. To assess the impact of the stringency of the cutoff filter score, Figure S9 shows the number of peptides identified using Byonic cutoff scores of 0, 100, and 300. When a specific peptide was identified in more than one fraction ("redundancy"), then it was counted only once. A peptide was also counted only once if it was identified in oxidized or deamidated forms in addition to the original form. However, glycosylation variants of a peptide were counted individually. A change in the level of stringency tended to have the greatest effect on identification of peptides with the greatest number of modifications. Such peptides tend to be of low abundance and may also ionize less well compared to unmodified peptides. The impact was greatest on O-glycopeptides; identification of the nonsialylated peptides with Byonic scores >300 was only 14% of the unfiltered identifications.

Discussion

Attachment of phosphate and sialyl residues increases the polarity of peptides. PGC was more effective than a conventional C-18 material at capturing peptides decorated with more than one such residue. This raises the disturbing possibility that prior studies of peptides with these PTMs may have significantly underestimated the abundance of peptides with >1 phosphate residue, >1 sialyl residue, or any mannose-6-phosphate residues if they involved a desalting step with the C-18 material only. Supplementary Table 1 from ref (25) presents a relevant example. A tryptic digest of human epidermoid cells was desalted on a C-18 cartridge prior to affinity isolation of phosphopeptides and subsequent fractionation. Out of 5009 phosphopeptides identified, only 10

were triphosphopeptides. With our method, one might expect to identify 75–90 triphosphopeptides from brain tissue extract from a phosphopeptide set of that size.

If PGC complements the selectivity of the C-18 material for binding peptides, then why not mix both materials together in a single SPE desalting step? The answer is that while PGC binds many polar peptides that do not bind well to the C-18 material, it also binds many of the peptides that do bind to the C-18 material, and its affinity for such peptides may be greater than for the peptides with PTMs. This would saturate the binding capacity of the PGC, leaving it unable to bind the peptides that do not bind to the C-18 material. It is prudent to first deplete the sample of peptides that C-18 does bind. A similar situation was described in ref (1). There, a single IMAC cartridge bound mostly doubly phosphorylated peptides. The filtrate was then passed through an identical cartridge, and this time the singly phosphorylated peptides were bound because there were no multiply phosphorylated peptides to displace them.

Conclusions

The number of modified peptides identified in this study compares favorably with or exceeds the number reported using other methods. No specialized materials are required. This optimized version of ERLIC should be considered as a viable alternative to affinity enrichment methods involving titanium dioxide, IMAC, lectins, or boronic acids. Because peptides with different modifications can be enriched and identified simultaneously, this approach facilitates the study of PTM crosstalk and protein network interactions that are not convenient or possible to perform using the materials with higher degrees of affinity.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.1c00615.

Expanded experimental details of protein digestion and LC–MS/MS analysis; illustration of peptide retention with counterions added in HILIC; figures of chromatograms; figure of the distribution of modified peptides captured by C-18 material versus PGC material; figure of residue number per modified peptide analysis; figure of comparison of modification sites identified in this study and Uniprot reported evidence; figure of glycan heterogeneity analysis; figure of Byonic cutoff score optimization analysis; chromatogram before and after column passivation; figure of M6P peptide spectra manual inspection results; figure of reproducibility analysis; figure of literate comparisons; figure of example MS spectra; figure of EThcD O-glycopeptide analysis; tables of identified modified peptide distribution in ERLIC fractions; tables of link sizes of Figure 9; and table of proteins associated with AD and their PTMs identified in this study (PDF)

Acknowledgments

PolyWAX LP and PolySAX LP are trademarks of PolyLC Inc. TopTip is a trademark of Glygen Corp. (Columbia, MD, USA). HyperCarb is a trademark of ThermoFisher Scientific. Sachtopore is a trademark of Sachtleben Chemie GmbH (now part of Venator Materials PLC). This research was supported in part by the National Institutes of Health grants U01CA231081, R01 DK071801, and RF1 AG052324 (to L.L.). The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531 to L.L.) and

Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. L.L. acknowledges a Vilas Distinguished Achievement Professorship and a Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

References

1. Zhang, X.; Ye, J.; Jensen, O. N.; Roepstorff, P. Highly efficient phosphopeptide enrichment by calcium phosphate precipitation combined with subsequent IMAC enrichment. *Mol. Cell. Proteomics* **2007**, *6*, 2032– 2042, DOI: 10.1074/mcp.M700278-MCP200

2. Yue, X.; Schunter, A.; Hummon, A. B. Comparing multi-step IMAC and multi-step TiO_2 methods for phosphopeptide enrichment. *Anal. Chem.* **2015**, *87*, 8837– 8844, DOI: 10.1021/acs.analchem.5b01833

3. Glover, M. S.; Yu, Q.; Chen, Z.; Shi, X.; Kent, K. C.; Li, L. Characterization of intact sialylated glycopeptides and phosphorylated glycopeptides from IMAC enriched samples by ETHcD fragmentation: Toward combining phosphoproteomics and glycoproteomics. *Int. J. Mass Spectrom.* **2018**, *427*, 35– 42, DOI: 10.1016/j.ijms.2017.09.002

4. Čaval, T.; Zhu, J.; Tian, W.; Remmelzwaal, S.; Yang, Z.; Clausen, H.; Heck, A. J. R. Targeted Analysis of Lysosomal Directed Proteins and Their Sites of Mannose-6-phosphate Modification. *Mol. Cell. Proteomics* **2019**, *18*, 16– 27, DOI: 10.1074/mcp.RA118.000967

5. Xiao, H.; Chen, W.; Smeekens, J. M.; Wu, R. An enrichment method based on synergistic and reversible covalent interactions for large-scale analysis of glycopeptides. *Nat. Commun.* **2018**, *9*, 1692, DOI: 10.1038/s41467-018-04081-3

6. Alpert, A. J.; Hudecz, O.; Mechtler, K. Anion-Exchange Chromatography of Phosphopeptides: Weak anion exchange versus strong anion exchange and anion-exchange chromatography versus electrostatic repulsion–hydrophilic interaction chromatography. *Anal. Chem.* **2015**, *87*, 4704– 4711, DOI: 10.1021/ac504420c

7. McNulty, D. E.; Annan, R. S. Hydrophilic-interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol. Cell. Proteomics* **2008**, *7*, 971– 980, DOI: 10.1074/mcp.M700543-MCP200

8. Ding, W.; Nothaft, H.; Szymanski, C. M.; Kelly, J. Identification and quantification of glycoproteins using ion-pairing normal-phase liquid chromatography and mass spectrometry. *Mol. Cell. Proteomics* **2009**, *8*, 2170– 2185, DOI: 10.1074/mcp.M900088-MCP200

9. Alpert, A. J. Hydrophilic interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr. A* **1990**, *499*, 177– 196, DOI: 10.1016/S0021-9673(00)96972-3

10. Thannhauser, T. W.; Shen, M.; Sherwood, R.; Howe, K.; Fish, T.; Yang, Y.; Chen, W.; Zhang, S. A workflow for large-scale empirical identification of cell wall *N*-linked glycoproteins of tomato (*Solanum lycopersicum*) fruit by tandem mass spectrometry. *Electrophoresis* **2013**, *34*, 2417– 2431, DOI: 10.1002/elps.201200656

11. Alpert, A. J. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal. Chem.* **2008**, *80*, 62– 76, DOI: 10.1021/ac070997p

12. Zhang, H.; Guo, T.; Li, X.; Datta, A.; Park, J. E.; Yang, J.; Lim, S. K.; Tam, J. P.; Sze, S. K. Simultaneous characterization of glyco- and phosphoproteomes of mouse brain membrane proteome with electrostatic repulsion hydrophilic interaction chromatography. *Mol. Cell. Proteomics* **2010**, *9*, 635– 647, DOI: 10.1074/mcp.M900314-MCP200

13. Huyen Tran, T.; Hwang, I.-J.; Park, J.-M.; Kim, J. B.; Lee, H.-K. An application of electrostatic repulsion hydrophilic interaction chromatography in phospho- and glycoproteome profiling of epicardial adipose tissue in obesity mouse. *Mass Spectrom. Lett.* **2012**, *3*, 39– 42, DOI: 10.5478/MSL.2012.3.2.39

14. Cui, Y.; Yang, K.; Tabang, D. N.; Huang, J.; Tang, W.; Li, L. Finding the sweet spot in ERLIC mobile phase for simultaneous enrichment of N-glyco and phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2491– 2501, DOI: 10.1007/s13361-019-02230-6

15. Chen, R.; Williamson, S.; Fulton, K. M.; Twine, S. M.; Li, J. Simultaneous analysis of phosphopeptides and intact glycopeptides from secretome with mode switchable solid phase extraction. *Anal. Methods* **2019**, *11*, 5243– 5249, DOI: 10.1039/C9AY01756A

16. Zeidan, Q.; Hart, G. W. The intersections between O-GlcNacylation and phosphorylation: implications for multiple signaling pathways. *J. Cell Sci.* **2010**, *123*, 13– 22, DOI: 10.1242/jcs.053678

17. Leney, A. C.; El Atmioui, D.; Wu, W.; Ovaa, H.; Heck, A. J. R. Elucidating crosstalk mechanisms between phosphorylation and O-GlcNAcylation. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E7255– E7261, DOI: 10.1073/pnas.1620529114

18. Grimes, M.; Hall, B.; Foltz, L.; Levy, T.; Rikova, K.; Gaiser, J.; Cook, W.; Smirnova, E.; Wheeler, T.; Clark, N. R.; Lachmann, A.; Zhang, B.; Hornbeck, P.; Ma'ayan, A.; Comb, M. Integration of protein phosphorylation, acetylation, and methylation data sets to outline lung cancer signaling networks. *Sci. Signaling* **2018**, *11*, eaaq1087 DOI: 10.1126/scisignal.aaq1087

19. Hendriks, I. A.; Lyon, D.; Young, C.; Jensen, L. R.; Vertegaal, A. C. O.; Nielsen, M. L. Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. *Nat. Struct. Mol. Biol.* **2017**, *24*, 325– 336, DOI: 10.1038/nsmb.3366

20. Alpert, A. J. Effect of salts on retention in hydrophilic interaction chromatography. *J. Chromatogr. A* **2018**, *1538*, 45– 53, DOI: 10.1016/j.chroma.2018.01.038

21. Cohen, R. D.; Liu, Y.; Gong, X. Analysis of volatile bases by high performance liquid chromatography with aerosol-based detection. *J. Chromatogr. A* **2012**, *1229*, 172– 179, DOI: 10.1016/j.chroma.2012.01.036

22. Shmukler, L. E.; Gruzdev, M. S.; Kudryakova, N. O.; Fadeeva, Y. A.; Kolker, A. M.; Safonova, L. P. Thermal behavior and electrochemistry of protic ionic liquids based on triethylamine with different acids. *RSC Adv.* **2016**, *6*, 109664– 109671, DOI: 10.1039/C6RA21360J

23. Alpert, A. J.; Petritis, K.; Kangas, L.; Smith, R. D.; Mechtler, K.; Mitulović, G.; Mohammed, S.; Heck, A. J. R. Peptide orientation affects selectivity in ion-exchange chromatography. *Anal. Chem.* **2010**, *82*, 5253– 5259, DOI: 10.1021/ac100651k

24. Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nat. Commun.* **2019**, *10*, 1311, DOI: 10.1038/s41467-019-09222-w

25. Ruprecht, B.; Kock, H.; Medard, G.; Mundt, M.; Kuster, B.; Lemeer, S. Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns. *Mol. Cell. Proteomics* **2015**, *14*, 205– 215, DOI: 10.1074/mcp.M114.043109

26. Fang, P.; Wang, X.-J.; Xue, Y.; Liu, M.-Q.; Zeng, W.-F.; Zhang, Y.; Zhang, L.; Gao, X.; Yan, G.-Q.; Yao, J.; Shen, H.-L.; Yang, P.-Y. In-depth mapping of the mouse brain N-glycoproteome reveals widespread N-glycosylation of diverse brain proteins. *Oncotarget* **2016**, *7*, 38796– 38809, DOI: 10.18632/oncotarget.9737

27. Zhang, C.; Ye, Z.; Xue, P.; Shu, Q.; Zhou, Y.; Ji, Y.; Fu, Y.; Wang, J.; Yang, F. Evaluation of Different N-Glycopeptide Enrichment Methods for N-Glycosylation Sites Mapping in Mouse Brain. *J. Proteome Res.* **2016**, *15*, 2960– 2968, DOI: 10.1021/acs.jproteome.6b00098

28. Piovesana, S.; Montone, C. M.; Cavaliere, C.; Crescenzi, C.; La Barbera, G.; Laganà, A.; Capriotti, A. L. Sensitive untargeted identification of short hydrophilic peptides by high performance liquid chromatography on porous graphitic carbon coupled to high resolution mass spectrometry. *J. Chromatogr. A* **2019**, *1590*, 73– 79, DOI: 10.1016/j.chroma.2018.12.066

29. Alpert, A.J.; Gygi, S.P.; Shukla, A.K. *Desalting phosphopeptides by solid-phase extraction*. ASMS 2007, poster MP 438.

30. Chen, R.; Stupak, J.; Williamson, S.; Twine, S. M.; Li, J. Online porous graphic carbon chromatography coupled with tandem mass spectrometry for post-translational modification analysis. *Rapid Commun. Mass Spectrom.* **2019**, *33*, 1240– 1247, DOI: 10.1002/rcm.8459

31. An, H. J.; Froehlich, J. W.; Lebrilla, C. B. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr. Opin. Chem. Biol.* **2009**, *13*, 421– 426, DOI: 10.1016/j.cbpa.2009.07.022

32. Cao, L.; Qu, Y.; Zhang, Z.; Wang, Z.; Prytkova, I.; Wu, S. Intact glycopeptide characterization using mass spectrometry. *Exp. Rev. Proteomics* **2016**, *13*, 513– 522, DOI: 10.1586/14789450.2016.1172965

33. Wang, Z.; Gucek, M.; Hart, G. W. Cross-talk between GlcNAcylation and phosphorylation: Site-specific phosphorylation dynamics in response to globally elevated *O*-GlcNAc. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13793– 13798, DOI: 10.1073/pnas.0806216105

34. Yao, H.; Li, A.; Wang, M. Sysstematic Analysis and Prediction of *In Situ* Cross Talk of O-GlcNAcylation and Phosphorylation. *Biomed. Res. Int.* **2015**, *2015*, 279823 DOI: 10.1155/2015/279823

35. Kanninen, K.; Goldsteins, G.; Auriola, S.; Alafuzoff, I.; Koistinaho, J. Glycosylation changes in Alzheimer's disease as revealed by a proteomic approach. *Neurosci. Lett.* **2004**, *367*, 235– 240, DOI: 10.1016/j.neulet.2004.06.013

36. Barykin, E. P.; Mitkevich, V. A.; Kozin, S. A.; Makarov, A. A. Amyloid β Modification: A Key to the Sporadic Alzheimer's Disease?. *Front. Genet.* **2017**, *8*, 58, DOI: 10.3389/fgene.2017.00058

37. Lee, L. Y.; Moh, E. S. X.; Parker, B. L.; Bern, M.; Packer, N. H.; Thaysen-Andersen, M. Toward automated *N*-glycopeptide identification in glycoproteomics. *J. Proteome Res.* **2016**, *15*, 3904– 3915, DOI: 10.1021/acs.jproteome.6b00438'

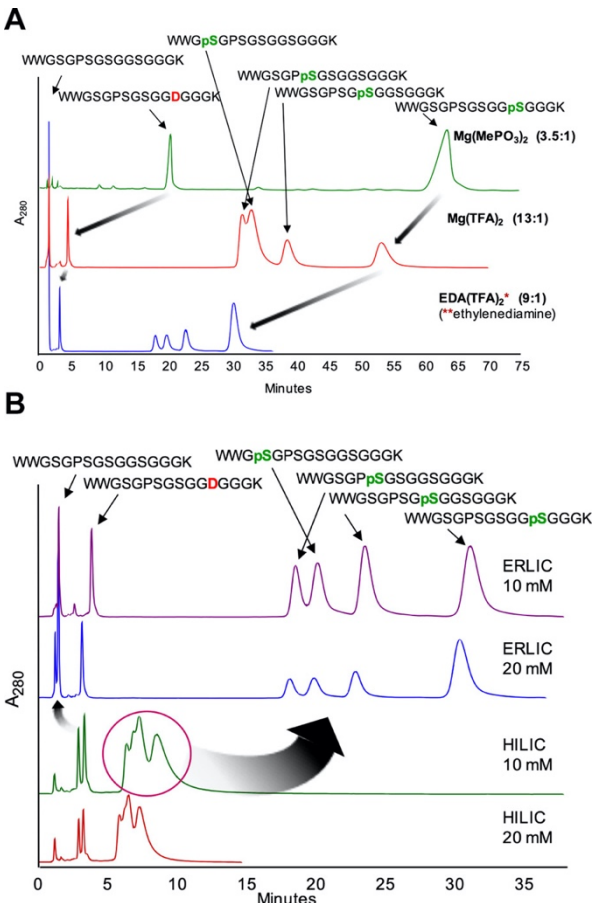

Figure 1. (A) Effect of counterions on ERLIC of phosphopeptides. Column: PolyWAX LP, 100 × 4.6-mm; 3-µm, 300-Å. MP: 20 mM salt as noted, pH 2.5, with 75% ACN. Flow rate: 1 mL/min. Detection: 280 nm. The ratio refers to retention of the last phosphopeptide to retention of the Asp-peptide. (B) HILIC vs ERLIC of phosphopeptides. HILIC column: PolyHYDROXYETHYL A, 100 × 4.6-mm; 3-µm, 300-Å. Salt: EDA-TFA (concentration as noted). All other variables as in (A).

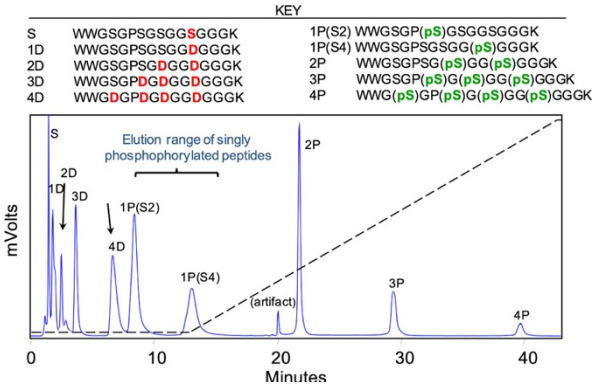

Figure 2. ERLIC of phosphopeptides and acidic analogues. Column, flow rate, and detection: As in Figure 1A. MP (A) 20 mM EDA-TFA, pH 2.1, with 77.5% ACN. MP (B) 1% TFA with 30% ACN. Gradient: 12′ delay, then 0–100% B in 30′.

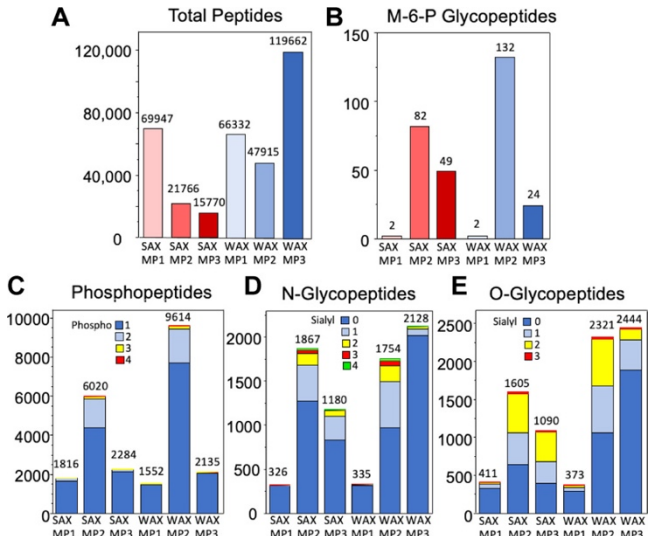

Figure 3. Peptide identification from 1 mg of mouse cortex after ERLIC fractionation and LC–MS/MS. Number of (A) total peptides; (B) mannose-6-phosphate glycopeptides; (C) phosphopeptides; (D) N-glycopeptides; (E) O-glycopeptides. Only fractions desalted with C-18 were analyzed, and redundant identifications were eliminated (i.e., if a peptide eluted in two adjacent fractions, then it was counted only once). The Byonic cutoff filter was 100.

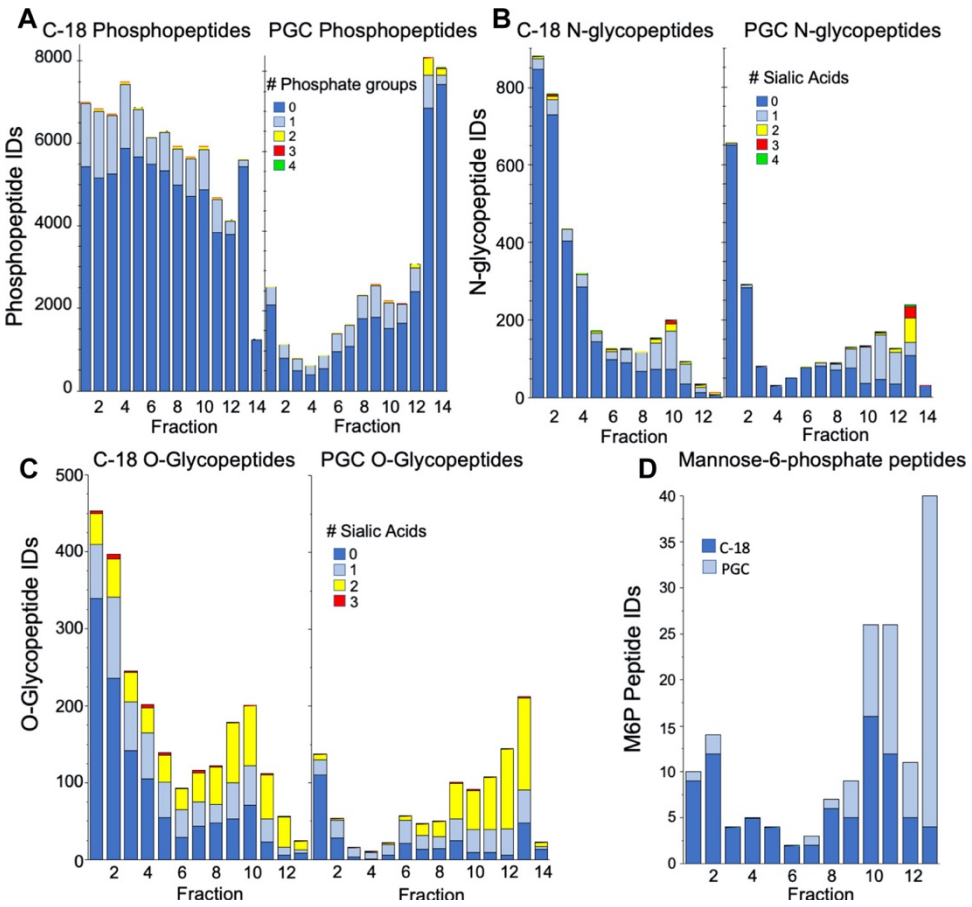

Figure 4. Distribution of modified peptide identifications among ERLIC fractions using MP2. Results from fractions desalted with C18 and with PGC are graphed separately except for M-6-P peptides. Redundant identifications were eliminated within each material data set but not between the C-18 and PGC datasets, because the characteristic of interest was the overall elution pattern rather than the absolute numbers. The Byonic cutoff filter was 100. The figures for each segment are listed in Tables S1–S4.

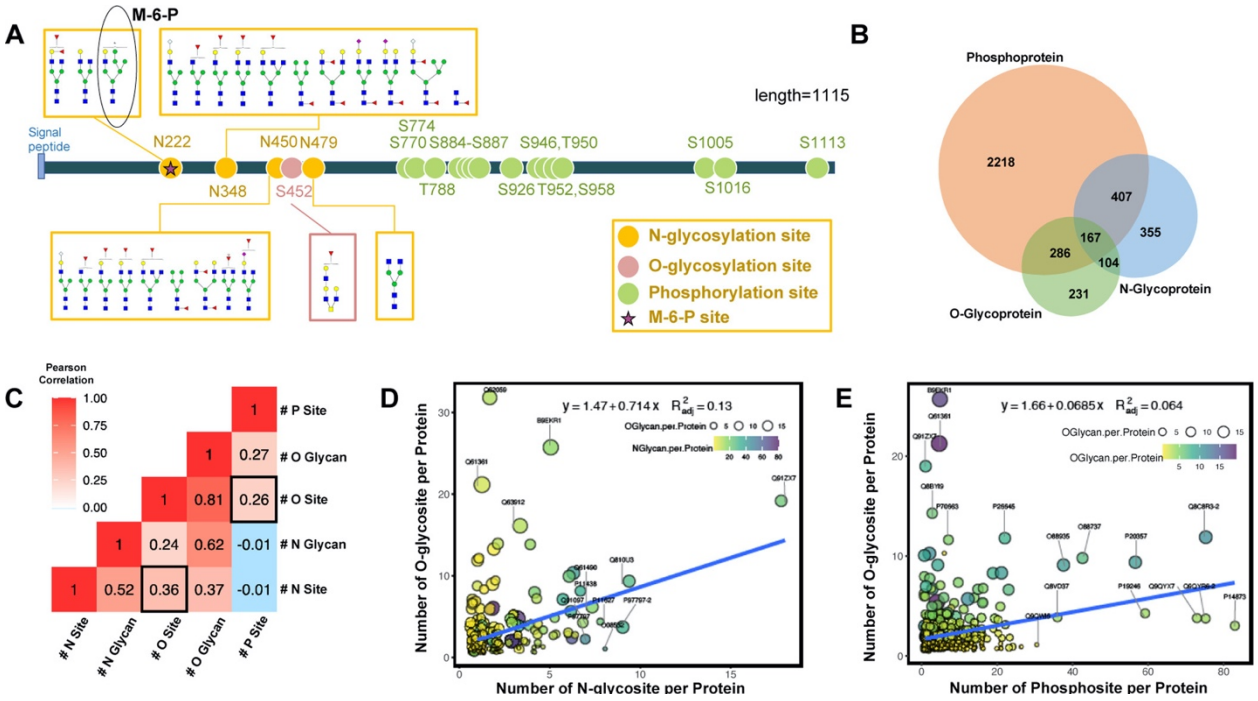

Figure 5. (A) PTMs in neural cell adhesion molecule 1 (Uniprot identifier P13595) [Gene: Ncam1]. A Byonic filter score of 300 was used. (B) Venn diagram showing coexistence of phosphorylation, N-glycosylation, and O-glycosylation at the protein level; (C) heatmap of the correlation of number of PTM sites and number of different glycans by protein. Only comodified proteins were included (for example, row 5, column 3 means the correlation between number of N- and O-glycosites for proteins that are both N- and O-glycosylated is 0.36). (D, E) Scatterplots of the number of modified sites within proteins that contain the two PTMs. The accessions of hypermodified proteins are labeled.

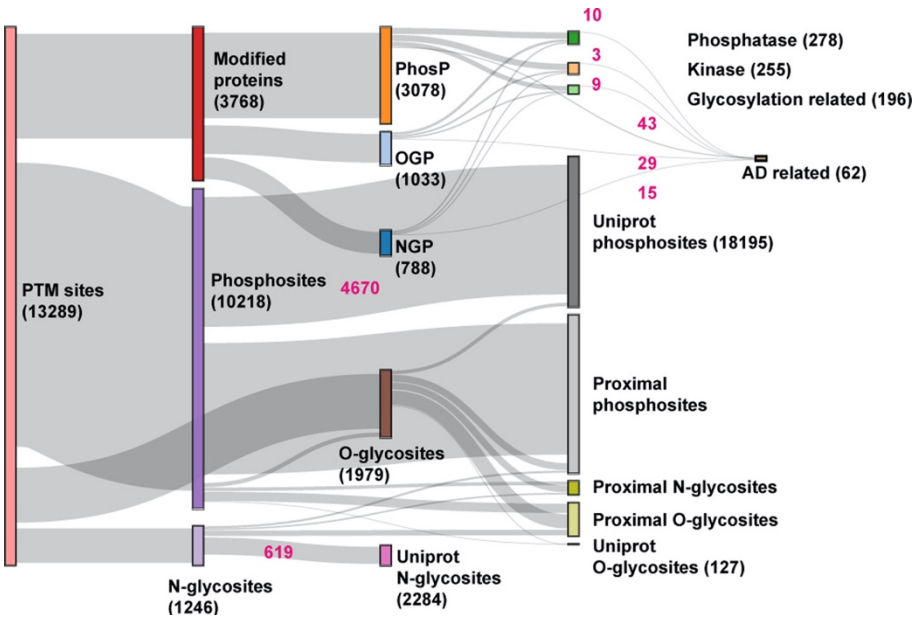

Figure 6. Sankey diagram of the PTM crosstalk network at the modification site-level and protein level. Numbers in black represent node sizes; numbers in pink represent link sizes. The complete sizes (numbers) of nodes and links are summarized in Tables S5 and S6. Abbreviations include PhosP (phosphoprotein), OGP (O-glycoprotein), and NGP (N-glycoprotein).

Chapter 7

Conclusions and Future Directions

In this dissertation, I developed numerous methodologies to improve the LC-MS/MS based proteomics analysis, including SPS-MS3 acquisition strategy, chemical labeling, and chromatography-based enrichment for PTM peptides.

In Chapter 2, DiLeu labeling based proteomics quantification successfully identified the activation of key biosynthetic proteins within *kyc* gene cluster during the coculture of *Micromonospora* sp. and a *Rhodococcus* sp. Along with colorimetric assay results of keyicin product, transcriptomics and metabolomics studies, we concluded that the multi-omics regulations are attributed to the small molecule signaling between the two bacteria strains. The BGC activation at both transcriptomic and proteomic levels resulted in the biosynthesis of keyicin by the normally silent BGC *kyc*.

Quantification accuracy of the DiLeu and DiAla labeling based quantification was improved by SPS-MS3 strategy as demonstrated in Chapter 3. Proteome regulation profile of the PCSK9 targeting compound HX261 in HepG2 cells was obtained, where AhR pathway was found to be activated. On the other hand, the development and application of the chemoproteomics method with *in situ* photoaffinity labeling and click chemistry enabled affinity purification coupled with MS LFQ successfully identified 29 potential protein interactors of the HX3-037 probe. Functional analysis showed those interactors are involved in cholesterol and fatty acid metabolisms. To draw clearer links between the identified interactors with the PCSK9 inhibition effect of the compound, models targeting each individual interactor can be developed to investigate in the future. For example, with the knock-down HepG2 cell model of each interactor, if the PCSK9 expression level also decreases as with compound treatment, then there is additional evidence that the PCSK9 inhibition of HX261 and HX3-037 is modulated through binding to the direct interactor protein.

Chapters 4-6 describe the development of electrostatic repulsion-hydrophilic interaction chromatography coupled with orbitrap MS to facilitate the large scale PTM analysis qualitatively and quantitatively. In the first application, DiLeu labeled intact N-glycopeptides from two AT-1 dysfunctional mouse models were enriched by ERLIC SPE and analyzed with EThcD hybrid fragmentation. Reproducible quantitation of 3-4 thousand intact N-glycopeptides was achieved across all brain regions and mouse models. The potential of using ERLIC to enrich peptides with phosphorylation and O-glycosylation was further explored. By utilization of counterions with different hydration numbers, the ERLIC retention of charged analytes can be modulated. We improved the selectivity of phosphopeptides, N- and O-glycopeptides in ERLIC by more than 5 fold. The method has the capability of simultaneously identifying different types of protein PTMs with site-localization and high coverage, which can be used for large scale PTM cross talk analysis, such as in biomarker discovery studies and investigations into disease progression mechanism. A follow-up study involves the investigation of the proteomic, phosphoproteomic and glycoproteomic changes in the APP/PS1 mouse model. Alzheimer's Disease (AD) is the leading cause of dementia globally. The neurological signaling cascade intrinsically involves extensive PTM regulations. PTMs such as phosphorylation and glycosylation might be key mediators that are altered in AD progression. Therefore, the comprehensive analysis of PTM regulations and crosstalk in AD brains might shed light on AD mechanism study and biomarker discovery.

Appendix I

Publications and Presentations

Publications

1. **Cui, Y.**, Tabang, D.N., Zhang, Z., Ma, M., Alpert, A.J., & Li, L. (2021) Counterion Optimization Dramatically Im- proves Selectivity for Phosphopeptides and Glycopeptides in Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC). Analytical Chemistry, 93 (22), 7908–7916. https://doi.org/10.1021/acs.analchem.1c00615.
2. **Cui, Y.**, Dieterich, I. A., Braun, M. M, Lawton., A. J., Robinson, N. H., Peotterm, J., Yu, Q., Casler, J. C., Glick, B. S., Audhya, A., Denu, J. M., Li, Li.,& Puglielli, L.(2021). The cytosol-to-ER flux of acetyl-CoA flux regulates engagement of the secretory pathway and the quality of the secretome. Scientific Reports, 11(1), 1-17.
3. **Cui, Y.**, Yang, K., Tabang, D. N., Huang, J., Tang, W., & Li, L. (2019). Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides. Journal of The American Society for Mass Spectrometry, 30(12), 2491-2501.
4. Dieterich, I. A., Lawton, A. J., Peng, Y., Yu, Q., Rhoads, T. W., Overmyer, K. A., **Cui, Y.**, ... & Li, L. (2019). Acetyl-CoA flux regulates the proteome and acetyl-proteome to maintain intracellular metabolic crosstalk. Nature Communications, 10(1), 1-11.
5. Acharya, D., Miller, I., **Cui, Y.**, Braun, D. R., Berres, M. E., Styles, M. J., ... & Bugni, T. S. (2019). Omics Technologies to Understand Activation of a Biosynthetic Gene Cluster in Micromonospora sp. WMMB235: Deciphering Keyicin Biosynthesis. ACS chemical biology, 14 (6), 1260-1270.
6. Huang, J., Dong, J., Shi, X., Chen, Z., **Cui, Y.**, Liu, X., ... & Li, L. (2019). Dual-Functional Tita- nium (IV) Immobilized Metal Affinity Chromatography Approach for

Enabling Large-Scale Profiling of Pro- tein Mannose-6-Phosphate Glycosylation and Revealing Its Predominant Substrates. Analytical Chemistry, 91(18), 11589-11597.

7. Liu, R., Wei, P., Keller, C., Orefice, N. S., Shi, Y., Li, Z., Huang, J., **Cui, Y.**,..., & Li, L. Integrated Label- Free and 10-Plex DiLeu Isobaric Tag Quantitative Methods for Profiling Changes in the Mouse Hypothalamic Neuropeptidome and Proteome: Assessment of the Impact of the Gut Microbiome. Analytical Chemistry, 92 (20), 14021-14030
8. **Cui, Y.**, Du, W., & Liang, G. (2016). Self-Assembly/Disassembly of Nanostructures Confers "Off/On" Signal for Molecular Imaging. ChemNanoMat, 2(5), 344-353.
9. Du, W., Yuan, Y., Wang, L., **Cui, Y.**, Wang, H., Xu, H., & Liang, G. (2015). Multifunctional bioconjugate for cancer cell-targeted theranostics. Bioconjugate Chemistry, 26(12), 2571-2578.
10. Book chapter: Du, W., **Cui, Y.**, & Liang, G. (2017). Applications of Organic-Inorganic Hybrid Nanoparticles for Cancer Imaging. In Hybrid Nanomaterials (pp. 110-155). CRC Press.
11. Zheng, Z., Sun, H., Hu, C., Li, G., Liu, X., Chen, P., **Cui,Y.**, ... & Liang, G. (2016). Using "On/Off" 19F NMR/Magnetic Resonance Imaging Signals to Sense Tyrosine Kinase/Phosphatase Activity in Vitro and in Cell Lysates. Analytical Chemistry, 88(6), 3363-3368.

Conference Presentations

1. Oral: **Cui, Y.**, Tabang, D.N., Alpert, A.J., & Li, L. Salt Optimization in ERLIC for Simultaneous and Highly Selective Enrichment of Peptides with Multiple PTMs, 39th

Interational Symposium and Exhibit on the Separation, Purification, and Characterization of Biologically Important Molecules, Baltimore, MD, July 2019.

2. **Cui, Y.**, Yang, K., Tabang, D.N., Huang, J., Tang, W., & Li, L. Finding the Sweet Spot in SAX-ERLIC Mobile Phase for Simultaneous Enrichment of Glyco and Phospho-peptides, 67th ASMS Conference, Atlanta, GA, June 2019.
3. **Cui, Y.**, Dieterich, I.A. Rhoads, T.W., Chen, Z., Anderson, R., Puglielli, L., & Li, L. Isobaric Labeling-based Quantitative Studies of Protein Expression and N-glycosylation of AT-1 sTg Mouse Model Reveal Molecular Basis of Aging, 66th ASMS Conference, San Diego, CA, June 2018.
4. **Cui, Y.**, Yu, Q., Li, L. Optimization of Multi-notch MS3 Technique for Accurate Quantification with Dimethyl Amino Acid-based Isobaric Tags, 65th ASMS Conference, Indianapolis, IN, June 2017.