Integrative Mass Spectrometry Strategies for Mapping Protein Structural Variation and Modifications in Human Biofluids Under Disease Conditions

By

Zexin Zhu

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

Major in Pharmaceutical Science

at the

UNIVERSITY OF WISCONSIN-MADISON

2026

Date of final oral examination: Aug 27th, 2025

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

Lingjun Li, Professor, School of Pharmacy and Department of Chemistry

Jiaoyang Jiang, Associate Professor, School of Pharmacy

Jun Dai, Assistant Professor, School of Pharmacy

Luigi Puglielli, Professor, Department of Medicine

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Lingjun Li, for her unwavering support, guidance, and mentorship throughout my graduate studies. I still remember the summer of 2019, when I was introduced to Dr. Li by my acquaintance, Dr. Jianhui Zhu, who described her as the most knowledgeable yet easygoing professor she had ever met. During my first meeting with Dr. Li in her office, I was astonished by the "ocean of documents" — stacks of journals and publications that reflected a life devoted to science. Yet, when I spoke about my research interests, she could effortlessly retrieve relevant information within seconds. It was my great honor to join the Li Lab in my third year as an undergraduate and continue into my Ph.D. journey. That decision was truly life-changing. I know that many years from now, when I reflect on my career path while speaking to old friends or family members, I will remember that shining afternoon when Dr. Li welcomed me into her lab with open arms.

Throughout my six-year scientific research in the Rennebohm hall, Dr. Li's scientific insight, patience, and encouragement have been instrumental not only in shaping the direction of my research but also in fostering my growth as an independent scientist. I am especially thankful for the freedom and trust she provided, which allowed me to explore new ideas, develop innovative methods, and grow both intellectually and professionally. Her high standards and thoughtful feedback continuously pushed me to refine my work and think critically about the broader impact of my research.

I would like to extend my sincere gratitude to my committee members, Dr. Jun Dai, Dr. Jiaoyang Jiang, and Dr. Luigi Puglielli. Throughout every annual committee meeting and the preliminary examination, their thoughtful suggestions and innovative ideas have been invaluable to my development as a researcher. As experts from diverse scientific fields, their insights helped me

refine my experimental design, optimize workflow efficiency, and improve overall experimental outcomes. I am especially grateful to Dr. Dai for generously allowing me to use her cell culture hood and incubator during a critical phase of my in vivo cellular crosslinking experiments and subsequent histone structural analysis. Her support during that technical challenge was deeply appreciated. I also fondly recall Dr. Jiang's enthusiastic explanation of glycosylation pathways and her willingness to share ongoing lab progress when I had questions about O-glycans during my early graduate years. Last but not least, I am profoundly thankful to Dr. Puglielli for providing access to a large collection of valuable Alzheimer's disease (AD) samples and for his genuine interest in my structural biomarker investigations. Their collective mentorship and encouragement have had a lasting impact on my scientific journey.

I would like to express my sincere gratitude to Dr. Zhicheng Jin from the University of Wisconsin School of Medicine and Public Health, Department of Pathology and Laboratory Medicine, for providing rheumatoid arthritis (RA) serum and synovial fluid samples. His contributions were essential for my comparative studies of RA subgroups and for refining the manuscript on the systematic citrullinome profiling of plasma proteins. I am also deeply thankful to Dr. Gary Firestein from the University of California, San Diego (UCSD), for generously providing fibroblast-like synoviocyte (FLS) samples. These samples enabled me to investigate protein-level variations associated with joint inflammation and damage at the cellular level. I would also like to thank Dr. Jon Odorico for providing pancreatic tissue samples from individuals with type 1 diabetes, which made it possible to examine citrullination patterns in histone proteins with a systematic perspective. Their contributions paved the road to my academic achievements throughout my Ph. D career.

In addition, I am grateful to Dr. Christian Bleiholder for his warm welcome and for inviting me to participate in the Ion Mobility Summer School at Florida State University. It was an unforgettable week, offering hands-on experience with the structural relaxation approximation (SRA) model and an inspiring tour of his laboratory. I was especially impressed by the custom-built laser system integrated into a Trapped Ion Mobility Spectrometry (TIMS) and Time-of-Flight (TOF) mass spectrometer platform, which exemplified his team's cutting-edge instrumentation work.

Throughout my graduate research years, I have been supported and guided by many extraordinary mentors and lab members. Dr. Junfeng Huang was my first mentor when I joined the lab as an undergraduate student. As a postdoctoral researcher, he is extremely experienced in post-translational modification enrichment strategies and innovations with materials like Ti-IMAC and advanced mass spectrometry acquisition. He is kind, optimistic, and patient, and I am grateful that he introduced me to the field of mass spectrometry — from the very basics of column packing and sample preparation to the full proteomics workflow. My second mentor, Dr. Bin Wang, has also been an important figure in my Ph.D. development. He taught me everything about citrullination analysis and Lumos instrument maintenance. His extensive knowledge inspired my structural MS projects in crosslinking and Alzheimer’s biomarker discovery, which were critical for my subsequent research and publications. His energy and courage have given me tremendous emotional support, especially during times of frustration when facing research bottlenecks.

I am also thankful to Dr. Hua Zhang and Dr. Haiyan Lu for their continuous help and companionship. Together, we overcame many challenges, from orbitrap instrument failures to complex Alzheimer’s disease sample handling. Outside the lab, we enjoyed playing tennis, going

on weekend picnics, and exploring the city — they treated me like a brother, and their friendship has made my Ph.D. journey truly memorable. I would also like to thank Dr. Wei (Wilson) Li for his invaluable help in synthesizing the new XLeu crosslinker. I cherish our lunch breaks and the fun stories we shared. I am grateful as well to Dr. Zihui Li, Dr. Yatao Shi, Dr. Yuan Liu, Dr. Zhijun Zhu, Dr. Shuling Xu, Dr. Graham Delafield, Dr. Min Ma, Dr. Zicong Wang, and all the other members of the Li Lab for their tremendous help and support.

To my friends outside the lab, I would like to thank Han Zhang for three years of unwavering companionship. Living under the same roof, playing together every night after work, and preparing each other's lunches brought me the kind of friendship I will never forget. These support during difficult times means so much to me. I also appreciate Ye Sun, Dr. Peijing Jia, and Shujie Jin for all the badminton matches and the joy they brought to my routine. I must acknowledge the lasting happiness from my old friend Dr. Yilin Yang for his continued support. I cannot wait to reunite with him in our hometown.

Last but not least, I would like to express my deepest gratitude to my family. I am immensely thankful to my parents for their constant encouragement and for always respecting my decisions while offering the most valuable guidance for my career and life. I am especially grateful to my girlfriend, Xinyu Yu, for being the most loving person in the world and for giving me the strength and motivation to pursue my dreams. Together, we have raised our cats, Milktea and Adai, who fill our little family with endless happiness.

To everyone who has supported, inspired, and encouraged me along this journey — I am forever grateful.

Table of Contents

Integrative Mass Spectrometry Strategies for Mapping Protein Structural Variation and Modifications in Human Biofluids Under Disease Conditions

Zexin Zhu

Under the supervision of Professor Lingjun Li

University of Wisconsin-Madison

Abstract

Proteins are key building blocks of living creatures. These organic molecules may exhibit various structural modifications and changes in abundance under different conditions. They may undergo post-translational modifications (PTMs) catalyzed by multiple enzymes, and they can also experience structural alterations upon exposure to pathogens or stimulants associated with specific diseases. Mass spectrometry (MS) has evolved into a powerful and robust analytical tool for elucidating subtle mass shifts induced by PTMs. It enables fast and accurate analysis with high resolution, facilitating systematic and in-depth proteomic profiling.

This dissertation employs multiple MS-based strategies to characterize protein PTMs and perform structural analyses of essential biomarker proteins. Built upon our lab's previous work on the design and synthesis of a unique biotin-thiol tag that enables targeted analysis of protein citrullination, I applied this enrichment strategy to track citrullination variations in plasma proteins involved in autoimmune diseases. Through this approach, citrullination associated with rheumatoid arthritis (RA) and Sjogren syndrome (SS) was mapped across different biofluids and organs. Furthermore, this dissertation investigates the subsequent structural effects of this modification.

Chapter 1

Introduction and Research Summary

Introduction and research summary

Proteins are central to nearly every biological process, and their functions are dictated not only by their amino acid sequences but also by their three-dimensional structures and chemical modifications. Structural alterations and post-translational modifications (PTMs) expand the complexity of the proteome far beyond what is encoded in the genome, generating a diverse repertoire of proteoforms that underpin both health and disease. Among these PTMs, citrullination has emerged as a critical regulator of immune tolerance and chromatin dynamics, while conformational remodeling of proteins represents a key driver in neurodegenerative disorders. Despite advances in proteomics, accurately mapping these structural and chemical variations remains a major analytical challenge. To address this gap, my dissertation leverages mass spectrometry–based innovations to dissect protein structural dynamics and PTM-mediated regulation across autoimmune and neurodegenerative disease contexts, building bridges between molecular mechanisms and clinical phenotypes.

Chapter 1 provides an overview of the background, research objectives, and methodological foundation of the dissertation. The subsequent chapters are organized to reflect a progression from chemical modification–focused discovery to structural proteomics innovations and their biological applications. **Chapters 2 and 3** focus on citrullination identification in autoimmune and neuroinflammatory diseases, where we advanced acquisition strategies and uncovered novel biomarkers with mechanistic significance**. Chapters 4 through 7** shift the focus toward structural proteomics, particularly crosslinking mass spectrometry (XL-MS), to investigate biomarker-specific structural changes and dynamic protein–protein interaction

networks. This section also introduces methodological innovations, including the development of a new isobaric crosslinker (**Chapter 6**), and integrates XL-MS with complementary MS approaches to probe histone dimer stability under citrullination (**Chapter 7**). Finally, **Chapter 8** expands the scope of histone PTM studies by applying quantitative dimethyl labeling to characterize lysine methylation processes, offering a broader perspective on chromatin regulation. **Chapter 9** synthesizes the findings and outlines future directions. This structure was intentionally designed to move from PTM discovery in disease contexts to advanced structural interrogation, thereby providing both mechanistic insights and methodological contributions to the field.

Chapter 2 presents an application of citrullination identification in autoimmune disease. Citrullination, a post-translational modification (PTM) catalyzed by peptidylarginine deiminases (PADs), plays a critical role in the pathogenesis of rheumatoid arthritis (RA) and comorbid autoimmune conditions, such as Sjögren's syndrome (SS). However, the citrullinome landscape and its variation across these diseases remain largely unexplored. In this study, we conducted a systematic proteomic analysis of citrullinated proteins in serum and synovial fluid (SF) from RA and SS patients using biotin thiol tag-assisted mass spectrometry. We identified 95 unique citrullination sites across 59 proteins, predominantly in high-abundance plasma proteins such as immunoglobulins, apolipoproteins, and transferrins. Comparative analyses revealed a significant increase in citrullination levels in RA and SS samples compared to healthy controls, with both shared and distinct citrullination patterns between serum and SF. Functional enrichment analysis linked citrullinated proteins to

immune regulation and inflammatory pathways, reinforcing their role in RA pathogenesis. Notably, we identified a novel citrullination motif in cannabinoid receptor 2 (CNR2), which may disrupt receptor function and exacerbate inflammation, highlighting a potential therapeutic target. These findings provide new insights into the role of the citrullinome in RA and SS and suggest citrullination as a potential biomarker and therapeutic avenue for autoimmune disease management.

Chapter 3 introduces an upgraded chemoproteomic platform for DIA-based detection of citrullination profiling in AD. We report the development of a third-generation citrulline-specific probe, Bz2-PYA, featuring a benzil electrophile for irreversible capture, an alkyne handle for bioorthogonal conjugation, and a quaternary ammonium for MS signal enhancement. To ensure specificity, we implemented three safeguards: (i) dual-isotopic H6/D6 labeling for MS1 validation; (ii) ammonia elution to remove reversible non-citrulline adducts; and (iii) cobalt(II)-mediated blocking of guanidinium groups to suppress arginine cross-reactivity. The probe selectively labeled synthetic peptides, recombinant histone H2B, and complex proteomes with high specificity and sensitivity. Application to Alzheimer's disease (AD) and control sera, followed by CuAAC biotinylation, enrichment, and LC–MS/MS in both data-dependent and staggered data-independent acquisition (DIA) modes, enabled the identification of over 350 citrullination sites per sample with markedly reduced variability (CV = 9.7% in DIA *vs.* 40.6% in DDA). Comparative profiling revealed numerous AD-enriched citrullination events, particularly on proteins central to lipid transport and complement activation, including APOE, A2M, C3, and C4B. Functional enrichment

analyses highlighted immune and lipid-regulatory pathways, consistent with mechanisms of neuroinflammation and amyloid clearance in AD.

Chapter 4 summarizes structural proteomics strategies, focusing on two powerful MS-based techniques for peptide-level readout, namely limited proteolysis-mass spectrometry (LiP-MS) and cross-linking mass spectrometry (XL-MS). This review discusses the principles, features, and workflows of both methods. It also explores bioinformatics strategies and software tools used for interpreting data from these protein conformation readouts and how such data can be integrated with other computational tools. Furthermore, we provide a comprehensive summary of notable applications of LiP-MS and XL-MS in areas including neurodegenerative diseases, interactome studies, membrane proteins, and artificial intelligence-based structural analysis. Finally, we discuss factors that modulate protein conformational changes and highlight the remaining challenges in understanding the intricacies of protein structural dynamics using LiP-MS and XL-MS technologies.

Chapter 5 focuses on identifying key protein factors associated with AD progression by employing quantitative XL-MS to elucidate conformational changes in protein networks within cerebrospinal fluid (CSF). Using isotopically labeled crosslinkers BS^3d_0 and BS^3d_4, we reveal dynamic shifts in protein interaction networks during AD progression. Our comprehensive analysis highlights distinct alterations in protein-protein interactions within the Mild Cognitive Impairment (MCI) stage. This study underscores the potential of crosslinked peptides as indicators of AD-related conformational changes, including previously unreported site-specific binding between alpha-1-antitrypsin (A1AT) and Complement Component 3 (CO3).

Furthermore, we examine the structure of Apolipoprotein E (ApoE) and reveal modifications within its helical domains, suggesting their involvement in MCI pathogenesis. The quantitative approach provides insights into site-specific interactions and changes in crosslink abundance, offering a detailed understanding of the intricate protein-protein relationships underlying AD progression. These findings lay the foundation for developing potential diagnostic or therapeutic strategies aimed at mitigating the impact of AD.

Our research group is also devoted to developing innovative crosslinking methods. **Chapter 6** introduces a novel 4-plex isobaric crosslinker, XLeu, which enables simultaneous identification and quantitation of protein-protein interactions (PPIs) via cleavable reporter ions. Benchmarking against established crosslinkers (DSSO, BS^3, DSBSO) using bovine serum albumin demonstrated that XLeu provides labeling efficiency, structural coverage, and quantitative accuracy comparable to commercial reagents, while offering multiplexing capacity. Application of XLeu to human serum samples revealed PPI networks centered on transferrin and complement factors, highlighting their role as structural hubs. Importantly, condition-specific quantitation identified a high-confidence crosslink between complement C3 and C5 enriched in control females, supporting immune network remodeling in AD. Furthermore, age-stratified analysis uncovered a non-linear relationship between crosslink density and age, with the highest connectivity observed in individuals aged 65–80 years. Collectively, these results establish 4-plex XLeu as a robust reagent for multiplexed XL-MS, providing new opportunities to investigate condition- and age-dependent remodeling of PPI networks in complex proteomes.

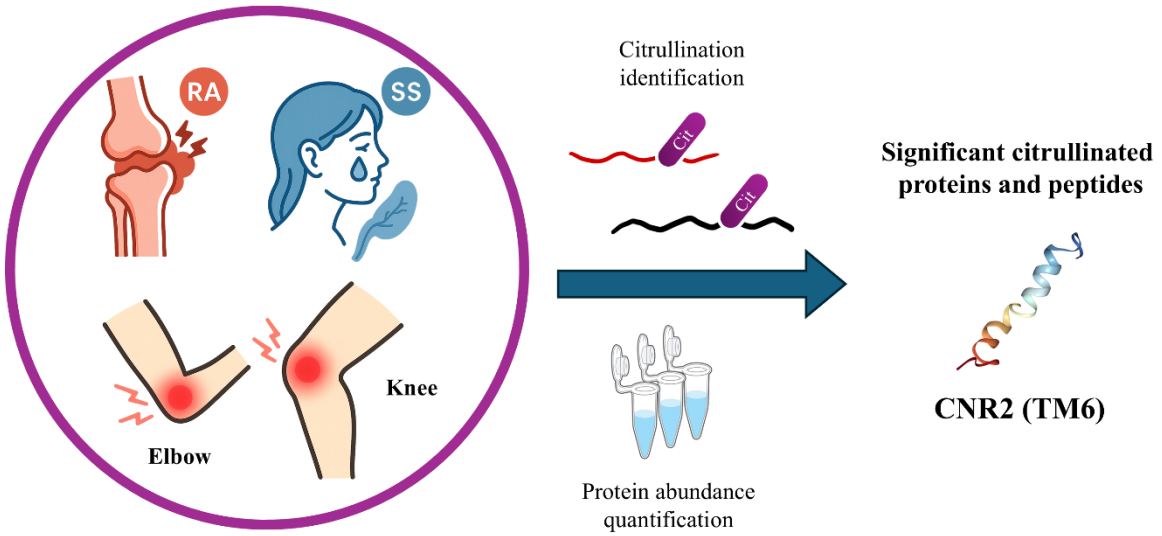
In **Chapter 7,** we applied an integrated mass spectrometry approach combining crosslinking mass spectrometry (XL-MS) and native ion mobility mass spectrometry (IM-MS) to investigate the structural impact of PAD4-mediated citrullination on the histone H2A–H2B dimer. XL-MS revealed that citrullination nearly doubled the number of inter-subunit crosslinks and shifted their distribution toward residues proximal to known citrullination sites. Structural mapping and distance filtering showed a transition from extended inter-dimer interactions in the unmodified state to more localized intra-dimer interactions following citrullination, suggesting condensation of the dimer interface. Native IM-MS confirmed this effect, as intact dimer ions were well-resolved in unmodified samples but substantially diminished or undetectable after citrullination, indicating a loss of dimer integrity. Further CIU analysis on an orthogonal platform revealed distinct unfolding patterns and a citrullination-specific CIU50 transition in citrullinated H2A, suggesting altered conformational stability. Additionally, the citrullinated monomers exhibited narrower drift-time distributions, indicative of more compact or aggregated gas-phase structures. These findings suggest that PAD4-mediated citrullination not only weakens histone dimer stability but also induces structural reorganization that favors localized interactions within the dimer while disrupting higher-order nucleosome architecture. This work demonstrates the power of combining covalent labeling and gas-phase ion mobility measurements to dissect the structural consequences of histone PTMs, providing mechanistic insights into chromatin remodeling at both the residue and complex levels.

In addition to citrullination studies, **Chapter 8** focuses on the histone PTM through another quantitative analysis. Histone lysine methylation plays important roles in chromatin-based biological processes such as transcription and DNA repair. In this work, stable isotope dimethyl labeling was employed to block free lysine residues while simultaneously enabling accurate relative quantitation. Dimethyl labeling provided ideal quantitative conditions, minimizing differences in charge state and retention time between un-, mono-, di-, and trimethylated tryptic peptides. Heavy isotopic dimethyl labeling (^{13}C, D_3) also enabled discrimination of free lysines from native mono-, di-, and trimethylated lysine peptides during LC-MS/MS analysis. Moreover, we investigated and verified a common side reaction in typical dimethyl labeling protocols -- arginine and lysine cyclization -- and strengthened our methodology by optimizing reaction conditions to eliminate this side reaction without compromising labeling efficiency. Using this improved method, we quantified changes in histone lysine methylation in MCF7 cells with or without CTR9 knockdown and revealed a significant increase of H3K27me3 upon loss of CTR9. This discovery guided further investigation into the regulatory mechanism of H3K27me3 by CTR9 and identified a potential therapeutic vulnerability for breast cancer treatment.

Chapter 9 summarizes the key findings and discoveries from the previous chapters and presents future directions of our lab's research.

Chapter 2

A Systematic Investigation of the Citrullinome and Its Variation in Human Rheumatoid Arthritis Tissues



Adapted from: **Zhu, Z.**; Chiang, H.; et al. A Systematic Investigation of the Citrullinome and Its Variation in Human Rheumatoid Arthritis Tissues. *JPR* **2025.** In revision. Zhu, Z, contributed to the experimental design, data collection, result analysis and manuscript preparation.

Abstract

Citrullination, a post-translational modification catalyzed by peptidylarginine deiminases (PADs), plays a critical role in the pathogenesis of rheumatoid arthritis (RA) and its comorbid autoimmune conditions, such as Sjögren's syndrome (SS). However, the citrullinome landscape and its variation across these diseases remain largely unexplored. In this study, we conducted a systematic proteomic analysis of citrullinated proteins in serum and synovial fluid (SF) from RA and SS patients using biotin thiol tag-assisted mass spectrometry. We identified 95 unique citrullination sites across 59 proteins, predominantly in high-abundance plasma proteins such as immunoglobulins, apolipoproteins, and transferrins. Comparative analyses revealed a significant increase in citrullination levels in RA and SS samples compared to healthy controls, with both shared and distinct citrullination patterns between serum and SF. Functional enrichment analysis linked citrullinated proteins to immune regulation and inflammatory pathways, reinforcing their role in RA pathogenesis. Notably, we identified a novel citrullination motif in cannabinoid receptor 2 (CNR2), which may disrupt receptor function and exacerbate inflammation, highlighting a potential therapeutic target. These findings provide new insights into the role of citrullinome in RA and SS and suggest citrullination as a potential biomarker and therapeutic avenue for autoimmune disease management.

Keywords: proteomics, post-translational modifications (PTM), crosslink mass spectrometry (XL-MS), rheumatoid arthritis, autoimmune disease

Introduction

Rheumatoid Arthritis (RA) is a chronic autoimmune disorder characterized by systemic inflammation, joint destruction, and the hallmark presence of anti-citrullinated protein antibodies (ACPAs)[1, 2]. These autoantibodies target citrullinated epitopes, which arise from the post-translational modification of arginine residues to citrulline—a process termed citrullination, catalyzed by peptidyl arginine deiminases (PADs)[3, 4]. Dysregulated citrullination is now recognized as a central driver of RA pathogenesis, fostering the breakdown of immune tolerance and promoting ACPA production, which in turn exacerbates synovial inflammation and tissue damage[5, 6]. Global epidemiological data reveal the staggering burden of RA, affecting approximately 17.6 million individuals as of 2020[7], with mounting evidence linking environmental triggers and genetic risk factors to aberrant citrullination and autoimmunity[8].

A critical nexus in RA progression lies in neutrophil extracellular traps (NETs), web-like structures released by activated neutrophils that are enriched with hypercitrullinated proteins. PAD2 and PAD4 isoforms within NETs amplify citrullination, generating neoantigens that perpetuate ACPA responses and sustain chronic inflammation[9, 10] . Notably, NET-derived citrullinated antigens have been shown to stimulate macrophage activation and synovial fibroblast invasiveness, directly contributing to joint erosion[11, 12]. Despite these advances, the citrullinome—the full spectrum of citrullinated proteins—in key biological matrices such as synovial fluid (SF) and serum remains inadequately mapped. This gap hinders the

identification of disease-specific citrullination patterns and their functional implications in immune dysregulation[13].

Mass spectrometry (MS) has revolutionized proteomic profiling of post-translational modifications (PTMs), offering unparalleled precision in site-specific identification. However, citrullination poses unique analytical challenges: its minimal mass shift (+0.984 Da) is often misannotated as deamidation or ^{13}C isotopic peaks[14]. To address this, our lab pioneered a biotin thiol tag derivatization strategy that selectively labels citrullinated residues, enabling high-confidence identification via bottom-up proteomics[15-17]. Coupled with multiplexed DiLeu isobaric labeling, this workflow permits quantitative tracking of citrullination dynamics during RA progression[16]. Nevertheless, a critical question persists—whether observed citrullination changes reflect true enzymatic activity or merely fluctuations in substrate protein abundance. Differentiation between protein abundance and citrullination profiles is essential to pinpoint PAD-driven pathology versus passive biomarker accumulation.

In this study, we present a dual-parallel proteomic approach to concurrently quantify protein expression and site-specific citrullination in paired SF and serum samples from RA patients. Synovial fluid, enriched with immune cells and inflammatory mediators, provides a direct window into the joint's pathological microenvironment[18], while serum reflects systemic immune activity and circulating citrullinated antigens. Notably, RA rarely exists in isolation; up to 30% of patients develop comorbid autoimmune conditions such as Sjögren's syndrome (SS), which shares overlapping serological features (e.g., ACPAs) and pathogenic

mechanisms, including dysregulated citrullination and neutrophil-driven inflammation[19, 20]. By extending our analysis to SS patients, we aim to disentangle RA-specific citrullination signatures from those common to multiple autoimmune disorders. This comparative approach not only enhances the specificity of RA biomarker discovery but also identifies shared pathways that may explain the clinical overlap between these diseases. By correlating citrullination profiles with protein abundance across both matrices and patient cohorts, we distinguish context-dependent citrullination events from constitutive modifications, thereby identifying nodes in inflammatory networks that are unique to RA or broadly relevant to autoimmunity.

Experimental Section

RA Human Samples and Chemical Reagents

The research involving human subjects described in this study has received approval from the University of Wisconsin Institutional Review Board. The biological samples comprised a total of 8 serum samples and 8 synovial fluid samples, collected from specific participants in the healthy control, RA or SS groups (**Supplemental Table S1**).

Protease Digestion

All samples were subjected to Albumin Depletion Kit (Thermo) to remove albumin. Protein concentrations were measured using the BCA Protein Assay Kit (Thermo). A total of 200 μg of protein from each sample was used separately for citrullination site detection and protein

abundance measurement, as outlined in the workflow. Samples were digested using the standard filter-aided sample preparation (FASP) procedure[21].

Biotin-Thiol Tag Derivatization of Citrullinated Peptides

The derivatization process was performed as previously described[15, 17]. A total of 300 µg of biotin- thiol tag was dissolved in 40 µl trifluoroacetic acid (TFA). One microliter of 2,3-butanedione solution was mixed with 114 µl 12.5% TFA solution. The mixture was vortexed in the dark at 37 °C for 6 hours and dried using SpeedVac. To remove excess tags, strong cation exchange (SCX) was performed using TopTips (PolyLC). The tips were equilibrated three times with 100 µl of loading buffer containing 50% acetonitrile (ACN), 0.2% formic acid (FA), and 10 mM ammonium formate. Peptides were resuspended in 300 µl of loading buffer, loaded onto the SCX tips, and washed eight times with 150 µl of loading buffer. Peptides were eluted three times with 100 µl of washing buffer containing 25% ACN and 0.4 M ammonium formate. All centrifugation steps were performed at 1600 rpm for 2 minutes. Flowthroughs were collected and dried in SpeedVac.

Enrichment of derivatized citrullinated peptides

A total of 100 µl of streptavidin agarose (Sigma) was washed five times with 1 ml of 1× phosphate-buffered saline (PBS). The tube containing the beads was vortexed and centrifuged at 3000 rpm for 2 minutes, and the supernatant was discarded. Peptide samples were resuspended in 1 ml of 1× PBS and incubated with the beads at room temperature for 2 hours with rotation. The agarose beads were subsequently washed three times with 1× PBS, three

times with 5% ACN in PBS, and ten times with pure water. Each wash used 1 ml of solution. Derivatized peptides were released using four rounds of elution with 300 µl of elution buffer containing 80% ACN, 0.2% TFA, and 0.1% FA. The first elution was performed at room temperature, while the remaining three were heated at 80 °C for 5 minutes with shaking. Eluents were combined and filtered through a 0.22 µm Millex^{TM} syringe filter (Sigma). Final products were dried using a SpeedVac.

LC-MS/MS analysis

A Dionex Ultimate 3000 nanoLC system was coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo). Samples were redissolved in 15 µl of 3% ACN and 0.1% FA in water, then transferred to HPLC autosampler vials. Mobile phase A consisted of 0.1% FA in water, and mobile phase B consisted of 0.1% FA in 80% ACN and 20% water. The flow rate was set to 0.3 µl/min, and the injection volume was 2 µl. The gradient was increased from 3.8% to 37.5% of phase B over a 100-minute separation using a homemade nano-C18 microcapillary column (75 µm inner diameter × 20 cm, 1.7 µm Ethylene Bridged Hybrid C18 packing material, Waters Corporation).

The following mass spectrometer parameters were used for data acquisition: Samples were ionized in positive-ion mode at 2500V. MS1 scans were acquired at 60K resolution over a range of 300 – 1500 *m/z*. The S-lens RF level was set to 30 %, the AGC target to 50%, and the maximum injection time to 100 ms. Data-dependent acquisition mode was performed in 3-second cycles between master scans. In the stepped higher-energy collisional dissociation

(HCD) method, collisional energies (CE) were set to 27, 30, and 33. MS2 scans were acquired at 30K resolution, with a first mass of 120 *m/z*, an isolation window is 1.6 *m/z*, and a maximum injection time of 54 ms. Each sample was analyzed in technical duplicates.

Data Processing

Raw data collected from the Orbitrap instrument were searched against the UniProt reviewed *Homo sapiens* database (2023) using MaxQuant, allowing up to three missed cleavages for Lys-C digestion. Variable modifications include Oxidation (+15.99491 Da) and biotin-labeled citrullination (+354.10718 Da), with diagnostic ions at 227.08487 Da and 304.11479 Da. Label-free quantification (LFQ) was enabled with advanced site intensity analysis. Both protein and peptide identifications were filtered at a 1% false discovery rate (FDR).

Results and Discussion

Citrullination levels in serum of RA and Sjogren's Syndrome (SS) are significantly upregulated compared to healthy controls

Citrullination is well-established as a critical factor in the clinical diagnosis and pathogenesis of RA[22]. Notably, RA has a high prevalence of co-occurrence with other autoimmune diseases, including SS, with over 32% of RA patients diagnosed with SS[23]. However, the citrullinome pattern and its variation between these associated diseases remain unclear. To address this gap, we analyzed serum and SF samples collected from RA and SS patients. Due to specimen availability and confidentiality constraints, we were unable to obtain serum and

SF samples from the same patients. Therefore, the analyses of serum and SF were conducted separately.

In this study, we employed label-free quantification instead of isobaric labeling to minimize the potential loss of low-abundance citrullinated peptides during additional strong cation exchange (SCX) cleanup steps[16]. Protein abundance was also evaluated in parallel to provide insights into protein-level variations between RA and healthy control samples. From mass spectrometry (MS) data and MaxQuant analysis, we identified 95 unique citrullination sites across 59 proteins (**Figure 2a, Table S2**). These sites predominantly occurred in high-abundance plasma proteins, including immunoglobulins, apolipoproteins, and transferrins. Most citrullination sites were shared between the RA and SS groups, with only four sites showing reduced levels during RA progression (**Figure S1a)**. Hierarchical clustering of eight serum samples, along with Venn diagram and heatmap analyses of citrullination patterns, revealed that both the number and abundance of citrullination sites were lowest in healthy control samples. In contrast, many high-abundance citrullination sites were found in both the RA and SS groups, suggesting a strong correlation between these diseases and the pathogenic role of citrullination. We further derived the average intensity-based absolute quantification (iBAQ) of serum proteins across the sample groups and examined the abundance and citrullination levels of four key plasma proteins. For immunoglobulin G (IgG) and ceruloplasmin (CERU), overall protein abundance remained relatively consistent across healthy controls, RA patients, and SS patients (**Figure 2c, d**). However, citrullination abundance in the RA and SS group was significantly elevated, with the RA+SS group

exhibiting the highest levels in IgG. This finding underscores the role of increased citrullination in linking RA with SS development, consistent with clinical evidence of elevated rheumatoid factor and anti-citrullinated protein antibodies during disease progression[24, 25]. Similarly, apolipoprotein A-I (APOA1) and transferrin (TRFE) demonstrated higher citrullination levels in RA and SS patients compared to healthy controls, with the RA+SS group displaying the highest abundance, paralleling their elevated protein levels. Among all proteins measured, the most prominently upregulated protein in serum was inter-alpha-trypsin inhibitor heavy chain 4 (ITIH4) (**Figure S1b**), corroborating its established role as an RA biomarker[26-28]. We also identified distinct site-specific citrullination in a related family member, ITIH2 (**Figure S1c**). These observations underscore the pathogenic relevance of citrullination in connecting RA and SS, while positioning these modifications as promising diagnostic targets for both autoimmune conditions.

Citrullination patterns in synovial fluid reveal distinctions between PTM upregulation and protein level changes in autoimmune disease

The eight distinct samples were categorized into four groups based on anatomical origin and clinical diagnosis (**Table S1**), as joint location is known to influence epigenetic composition [22, 29]. The four groups comprised elbow, swollen knee, knee effusion, and knee samples from confirmed rheumatoid arthritis (RA) patients, with two samples per group. Analysis of digested proteins revealed that, while chromatographic profiles and protein abundance rankings were similar across samples, the number of unique proteins identified in SF was markedly reduced (~50%) compared to serum (**Figure S2a, b**). This suggests a lower

prevalence of citrullinated proteins and distinct citrullinome profiles in SF. LFQ-based quantitative comparisons of citrullinated peptides from serum and SF further support divergent citrullination patterns between elbow and knee SF samples (**Figure 3a**). Specifically, elbow samples exhibited significantly fewer citrullinated sites than the average detected in knee samples, despite comparable overall protein levels across most categories (**Figure S3a**). For plasma proteins such as APOA1 and complement factor 3 (CO3), the extent of citrullination in elbow samples was markedly lower than in knee samples (**Figure S3b**).

Moreover, citrullination in SF highlights previously unreported traits directly associated with RA-affected joints. While 37 unique citrullination sites identified in serum were also found in SF knee samples, an additional 30 citrullination sites were uniquely detected in SF (**Figure 3b**). This includes nine citrullination sites on the fibrinogen alpha chain (FIBA): R38, R42, R160, R218, R443, R458, R459, R573 and R591 (**Figure 3c**). Many of these sites have been previously documented, reinforcing the observation of highly citrullinated fibrinogen in the inflamed synovium of RA patients[30]. In our study, we successfully identified additional citrullinated sites, including R459 and R591. In autoimmune diseases, fibrinogen levels are elevated in both serum and synovial fluid due to increased permeability of the synovial membrane[31, 32]. Under the action of PAD enzymes, increased citrullinated fibrinogen induces ACPA production and triggers inflammatory responses in the synovium[33]. Our experimental results confirmed elevated FIBA levels in both serum and SF and highlighted the predominant detection of citrullinated fibrinogen in SF knee samples. This evidence

underscores the pivotal role of citrullinated fibrinogen in RA pathology, potentially influencing fibrin structural changes and deposition that eventually lead to bone erosion[34, 35].

Certain proteins exhibited citrullination exclusively in RA serum. For instance, alpha-2-macroglobulin (A2MG) citrullination was predominantly observed in RA and SS serum, while absent in healthy controls (**Figure 3e**). Despite higher overall A2MG levels in serum than in SF, the lack of citrullination in protein-abundant control samples highlights its potential as a disease-specific modification (**Figure 3f**). As A2MG functions as a protease inhibitor, its oxidation and subsequent citrullination may impair cytokine regulation, contributing to inflammatory cascades in RA—a novel observation that warrants further exploration[36, 37]. Additional clinical and biological experiments are required to elucidate the interaction networks within the synovial environment.

To further interpret the system-level functions of these citrullinated proteins, we perform comprehensive enrichment analysis through Metascape and Gene Oncology (GO) analysis via DAVID informatics[38, 39]. Metascape reported numerous chronic pain and inflammation-related terms associated with RA (**Figure 4a)** , and highlighted links to inflammatory myositis—a common RA comorbidity[40]. GO analysis underscored molecular function (MF) terms such as adaptive immune response and B-cell receptor signaling pathway, while biological process (BP) terms included interleukin-8 binding, primarily induced by citrullinated autoantigens and NETs in the synovium[41]. In summary, these results emphasize the central role of citrullination in RA and related autoimmune conditions.

CNR2 Citrullination as a Potential Therapeutic Target in Rheumatoid Arthritis

Cannabinoid receptor 2 (CNR2) , a G-protein coupled receptor (GPCR) predominantly expressed by immune cells (e.g., macrophages, T cells) and peripheral nerve terminals, has emerged as a potential therapeutic target in rheumatoid arthritis[42]. Recent studies have demonstrated that CNR2 expression in RA synovial tissue is significantly upregulated compared to osteoarthritis controls, correlating with elevated inflammatory cytokines (e.g., TNF-α, IL-6)[43]. Notably, CNR2 agonists such as HU-308 exert potent anti-inflammatory effects in preclinical RA models, suppressing synovitis and attenuating joint damage[44, 45]. In this study, we identified a unique citrullination motif in CNR2. Citrullination occurred exclusively at two adjacent arginine residues, R238 and R242, within the same peptide sequence (**Figure 5a**). These sites were invariably co-modified (with no ambiguous assignments) and detected in both serum and synovial fluid (SF) samples from RA and Sjögren's syndrome (SS) patients, but were absent in healthy controls (**Figure 5b, c**). In serum, CNR2 citrullination levels (LFQ abundance) were elevated in SS and SS+RA cohorts, aligning with the progressive citrullination signature observed in previous sections and suggesting a link to heightened inflammatory burden.

The cryo-EM structure of CNR2 localizes R238 and R242 to the N-terminal region of transmembrane helix 6 (TM6)[46], a critical domain for G-protein coupling and inflammatory regulation (**Figure 5d**). While these residues do not directly engage agonists in the ligand-binding pocket, they flank D240, a conserved residue that forms an ionic lock with TM5 to stabilize the receptor's inactive conformation[47, 48]. Citrullination eliminates the positive

charge of arginine, potentially disrupting this salt bridge and impairing the TM5-TM6 conformational shift required for G-protein activation. We hypothesize that citrullination at R238/R242 uncouples CNR2 from Gi-α subunits, dampening downstream signaling pathways, including cAMP inhibition and MAPK activation[49]. This defect could compromise CNR2's anti-inflammatory function in RA, exacerbating synovial inflammation. Restoring CNR2 signaling by targeting these citrullination sites may represent a novel therapeutic strategy.

Conclusion

In summary, our comprehensive analysis of the citrullinome in RA and SS reveals a complex landscape of citrullination that plays a pivotal role in disease pathogenesis and progression. By employing a label-free quantitative proteomic approach, we identified unique citrullination sites across plasma proteins, with significant alterations observed in both serum and synovial fluid. Our data demonstrate that enhanced citrullination in RA and Sjögren's syndrome, highlighted by the marked upregulation of IgG and CERU and distinct site-specific modifications in FIBA within SF, underscores the potential of these modifications as both diagnostic biomarkers and therapeutic targets. Moreover, the identification of a unique citrullination motif in CNR2 opens new avenues for understanding the molecular mechanisms driving inflammation and offers promising prospects for novel intervention strategies. Together, these findings not only advance our understanding of the molecular

underpinnings of autoimmune diseases but also pave the way for future research aimed at developing targeted diagnostics and therapeutics in rheumatoid arthrits.

Data Availability

The mass spectrometry raw data, peak list and the mzid files has been uploaded through PRoteomics IDEntifications Database (PRIDE) and deposited in ProteomeXchange[50]. The dataset identifier number is PXD064027 for ProteomeXchange.

Acknowledgements

This research was supported, in part, by the National Institutes of Health Grants R21AG065728, R01AG052324, R01AG078794, and R01DK071801. Some of the mass spectrometers were acquired using NIH shared instrument grants S10 OD028473, S10 RR029531, and S10 OD025084. H.L. wishes to thank the funding support for a Postdoctoral Career Development Award provided by the American Society for Mass Spectrometry. L.L. acknowledges a Vilas Distinguished Achievement Professorship and 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. O'Neil, L. J.; Alpizar-Rodriguez, D.; Deane, K. D., Rheumatoid Arthritis: The Continuum of Disease and Strategies for Prediction, Early Intervention, and Prevention. *J Rheumatol* **2024,** *51* (4), 337-349.
2. Aggarwal, R.; Liao, K.; Nair, R.; Ringold, S.; Costenbader, K. H., Anti-citrullinated peptide antibody assays and their role in the diagnosis of rheumatoid arthritis. *Arthritis Rheum* **2009,** *61* (11), 1472-83.
3. Darrah, E.; Andrade, F., Rheumatoid arthritis and citrullination. *Curr Opin Rheumatol* **2018,** *30* (1), 72-78.
4. Fousert, E.; Toes, R.; Desai, J., Neutrophil Extracellular Traps (NETs) Take the Central Stage in Driving Autoimmune Responses. *Cells* **2020,** *9* (4).
5. van Venrooij, W. J.; Pruijn, G. J., How citrullination invaded rheumatoid arthritis research. *Arthritis Res Ther* **2014,** *16* (1), 103.
6. Mondal, S.; Thompson, P. R., Protein Arginine Deiminases (PADs): Biochemistry and Chemical Biology of Protein Citrullination. *Acc Chem Res* **2019,** *52* (3), 818-832.
7. Collaborators, G. B. D. R. A., Global, regional, and national burden of rheumatoid arthritis, 1990-2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. *Lancet Rheumatol* **2023,** *5* (10), e594-e610.
8. Klareskog, L.; Ronnelid, J.; Lundberg, K.; Padyukov, L.; Alfredsson, L., Immunity to citrullinated proteins in rheumatoid arthritis. *Annu Rev Immunol* **2008,** *26*, 651-75.
9. Chirivi, R. G. S.; van Rosmalen, J. W. G.; van der Linden, M.; Euler, M.; Schmets, G.; Bogatkevich, G.; Kambas, K.; Hahn, J.; Braster, Q.; Soehnlein, O.; Hoffmann, M. H.; Es, H.; Raats, J. M. H., Therapeutic ACPA inhibits NET formation: a potential therapy for neutrophil-mediated inflammatory diseases. *Cell Mol Immunol* **2021,** *18* (6), 1528-1544.
10. Khandpur, R.; Carmona-Rivera, C.; Vivekanandan-Giri, A.; Gizinski, A.; Yalavarthi, S.; Knight, J. S.; Friday, S.; Li, S.; Patel, R. M.; Subramanian, V.; Thompson, P.; Chen, P.; Fox, D. A.; Pennathur, S.; Kaplan, M. J., NETs are a source of citrullinated autoantigens and stimulate inflammatory responses in rheumatoid arthritis. *Sci Transl Med* **2013,** *5* (178), 178ra40.
11. McInnes, I. B.; Schett, G., Cytokines in the pathogenesis of rheumatoid arthritis. *Nature Reviews Immunology* **2007,** *7* (6), 429-442.
12. Carmona-Rivera, C.; Carlucci, P. M.; Moore, E.; Lingampalli, N.; Uchtenhagen, H.; James, E.; Liu, Y.; Bicker, K. L.; Wahamaa, H.; Hoffmann, V.; Catrina, A. I.; Thompson, P.; Buckner, J. H.; Robinson, W. H.; Fox, D. A.; Kaplan, M. J., Synovial fibroblast-neutrophil interactions promote pathogenic adaptive immunity in rheumatoid arthritis. *Sci Immunol* **2017,** *2* (10).
13. Tilvawala, R.; Nguyen, S. H.; Maurais, A. J.; Nemmara, V. V.; Nagar, M.; Salinger, A. J.; Nagpal, S.; Weerapana, E.; Thompson, P. R., The Rheumatoid Arthritis-Associated Citrullinome. *Cell Chem Biol* **2018,** *25* (6), 691-704 e6.

14. Rebak, A. S.; Hendriks, I. A.; Nielsen, M. L., Characterizing citrullination by mass spectrometry-based proteomics. *Philos Trans R Soc Lond B Biol Sci* **2023,** *378* (1890), 20220237.
15. Shi, Y.; Li, Z.; Wang, B.; Shi, X.; Ye, H.; Delafield, D. G.; Lv, L.; Ye, Z.; Chen, Z.; Ma, F.; Li, L., Enabling Global Analysis of Protein Citrullination via Biotin Thiol Tag-Assisted Mass Spectrometry. *Anal Chem* **2022,** *94* (51), 17895-17903.
16. Li, Z.; Wang, B.; Yu, Q.; Shi, Y.; Li, L., 12-Plex DiLeu Isobaric Labeling Enabled High-Throughput Investigation of Citrullination Alterations in the DNA Damage Response. *Anal Chem* **2022,** *94* (7), 3074-3081.
17. Wang, B.; Li, Z.; Shi, Y.; Zhu, Z.; Fields, L.; Shelef, M. A.; Li, L., Mass Spectrometry-Based Precise Identification of Citrullinated Histones via Limited Digestion and Biotin Derivative Tag Enrichment. *Anal Chem* **2024,** *96* (6), 2309-2317.
18. Smolinska, V.; Klimova, D.; Danisovic, L.; Harsanyi, S., Synovial Fluid Markers and Extracellular Vesicles in Rheumatoid Arthritis. *Medicina (Kaunas)* **2024,** *60* (12).
19. Harrold, L. R.; Shan, Y.; Rebello, S.; Kramer, N.; Connolly, S. E.; Alemao, E.; Kelly, S.; Kremer, J. M.; Rosenstein, E. D., Prevalence of Sjogren's syndrome associated with rheumatoid arthritis in the USA: an observational study from the Corrona registry. *Clin Rheumatol* **2020,** *39* (6), 1899-1905.
20. Molano-Gonzalez, N.; Olivares-Martinez, E.; Anaya, J. M.; Hernandez-Molina, G., Anti-citrullinated protein antibodies and arthritis in Sjogren's syndrome: a systematic review and meta-analysis. *Scand J Rheumatol* **2019,** *48* (2), 157-163.
21. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009,** *6* (5), 359-62.
22. Smolen, J. S.; Aletaha, D.; Barton, A.; Burmester, G. R.; Emery, P.; Firestein, G. S.; Kavanaugh, A.; McInnes, I. B.; Solomon, D. H.; Strand, V.; Yamamoto, K., Rheumatoid arthritis. *Nat Rev Dis Primers* **2018,** *4*, 18001.
23. Sebastian, A.; Madej, M.; Sebastian, M.; Butrym, A.; Woytala, P.; Halon, A.; Wiland, P., Prevalence and clinical presentation of lymphoproliferative disorder in patients with primary Sjogren's syndrome. *Rheumatol Int* **2020,** *40* (3), 399-404.
24. Nicolo, A.; Amendt, T.; El Ayoubi, O.; Young, M.; Finzel, S.; Senel, M.; Voll, R. E.; Jumaa, H., Rheumatoid factor IgM autoantibodies control IgG homeostasis. *Front Immunol* **2022,** *13*, 1016263.
25. Sieghart, D.; Konrad, C.; Swiniarski, S.; Haslacher, H.; Aletaha, D.; Steiner, G., The diagnostic and prognostic value of IgG and IgA anti-citrullinated protein antibodies in patients with early rheumatoid arthritis. *Front Immunol* **2022,** *13*, 1096866.
26. Joshi, L.; Chakraborty, D.; Kumar, V.; Biswas, S., ITIH4 in Rheumatoid Arthritis Pathogenesis: Network Pharmacology and Molecular Docking Analysis Identify CXCR4 as a Potential Receptor. *Pathophysiology* **2024,** *31* (3), 514-530.
27. Zhang, X. F.; Zhang, X. L.; Guo, L.; Bai, Y. P.; Tian, Y.; Luo, H. Y., The function of the inter-alpha-trypsin inhibitors in the development of disease. *Front Med (Lausanne)* **2024,** *11*, 1432224.

28. Kawaguchi, H.; Matsumoto, I.; Osada, A.; Kurata, I.; Ebe, H.; Tanaka, Y.; Inoue, A.; Umeda, N.; Kondo, Y.; Tsuboi, H.; Shinkai, Y.; Kumagai, Y.; Ishigami, A.; Sumida, T., Identification of novel biomarker as citrullinated inter-alpha-trypsin inhibitor heavy chain 4, specifically increased in sera with experimental and rheumatoid arthritis. *Arthritis Res Ther* **2018,** *20* (1), 66.
29. Ai, R.; Hammaker, D.; Boyle, D. L.; Morgan, R.; Walsh, A. M.; Fan, S.; Firestein, G. S.; Wang, W., Joint-specific DNA methylation and transcriptome signatures in rheumatoid arthritis identify distinct pathogenic processes. *Nat Commun* **2016,** *7*, 11849.
30. van Beers, J. J.; Raijmakers, R.; Alexander, L. E.; Stammen-Vogelzangs, J.; Lokate, A. M.; Heck, A. J.; Schasfoort, R. B.; Pruijn, G. J., Mapping of citrullinated fibrinogen B-cell epitopes in rheumatoid arthritis by imaging surface plasmon resonance. *Arthritis Res Ther* **2010,** *12* (6), R219.
31. Hugle, T.; Nasi, S.; Ehirchiou, D.; Omoumi, P.; So, A.; Busso, N., Fibrin deposition associates with cartilage degeneration in arthritis. *EBioMedicine* **2022,** *81*, 104081.
32. Knab, K.; Chambers, D.; Kronke, G., Synovial Macrophage and Fibroblast Heterogeneity in Joint Homeostasis and Inflammation. *Front Med (Lausanne)* **2022,** *9*, 862161.
33. Raijmakers, R.; van Beers, J. J.; El-Azzouny, M.; Visser, N. F.; Bozic, B.; Pruijn, G. J.; Heck, A. J., Elevated levels of fibrinogen-derived endogenous citrullinated peptides in synovial fluid of rheumatoid arthritis patients. *Arthritis Res Ther* **2012,** *14* (3), R114.
34. Kim, J. S.; Choi, M.; Choi, J. Y.; Kim, J. Y.; Kim, J. Y.; Song, J. S.; Ivashkiv, L. B.; Lee, E. Y., Implication of the Association of Fibrinogen Citrullination and Osteoclastogenesis in Bone Destruction in Rheumatoid Arthritis. *Cells* **2020,** *9* (12).
35. Bezuidenhout, J. A.; Venter, C.; Roberts, T. J.; Tarr, G.; Kell, D. B.; Pretorius, E., Detection of Citrullinated Fibrin in Plasma Clots of Rheumatoid Arthritis Patients and Its Relation to Altered Structural Clot Properties, Disease-Related Inflammation and Prothrombotic Tendency. *Front Immunol* **2020,** *11*, 577523.
36. Patterson, E. K.; Vanin Moreno, N.; Fraser, D. D.; Cepinskas, G.; Iida, T.; Berard, R. A., A Proteinase 3 Contribution to Juvenile Idiopathic Arthritis-Associated Cartilage Damage. *Pathophysiology* **2021,** *28* (3), 320-327.
37. Wu, S. M.; Pizzo, S. V., alpha(2)-Macroglobulin from rheumatoid arthritis synovial fluid: functional analysis defines a role for oxidation in inflammation. *Arch Biochem Biophys* **2001,** *391* (1), 119-26.
38. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **2019,** *10* (1), 1523.
39. Sherman, B. T.; Hao, M.; Qiu, J.; Jiao, X.; Baseler, M. W.; Lane, H. C.; Imamichi, T.; Chang, W., DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* **2022,** *50* (W1), W216-W221.

40. Naddaf, E.; Shelly, S.; Mandrekar, J.; Chamberlain, A. M.; Hoffman, E. M.; Ernste, F. C.; Liewluck, T., Survival and associated comorbidities in inclusion body myositis. *Rheumatology (Oxford)* **2022,** *61* (5), 2016-2024.
41. Sakkas, L. I.; Bogdanos, D. P.; Katsiari, C.; Platsoucas, C. D., Anti-citrullinated peptides as autoantigens in rheumatoid arthritis-relevance to treatment. *Autoimmun Rev* **2014,** *13* (11), 1114-20.
42. Fukuda, S.; Kohsaka, H.; Takayasu, A.; Yokoyama, W.; Miyabe, C.; Miyabe, Y.; Harigai, M.; Miyasaka, N.; Nanki, T., Cannabinoid receptor 2 as a potential therapeutic target in rheumatoid arthritis. *BMC Musculoskelet Disord* **2014,** *15*, 275.
43. Gui, H.; Liu, X.; Wang, Z. W.; He, D. Y.; Su, D. F.; Dai, S. M., Expression of cannabinoid receptor 2 and its inhibitory effects on synovial fibroblasts in rheumatoid arthritis. *Rheumatology (Oxford)* **2014,** *53* (5), 802-9.
44. Tian, N.; Yang, C.; Du, Y.; Chen, M.; Li, B.; Li, D.; Dai, S. M., Cannabinoid receptor 2 selective agonist ameliorates adjuvant-induced arthritis by modulating the balance between Treg and Th17 cells. *Front Pharmacol* **2025,** *16*, 1532518.
45. Richardson, D.; Pearson, R. G.; Kurian, N.; Latif, M. L.; Garle, M. J.; Barrett, D. A.; Kendall, D. A.; Scammell, B. E.; Reeve, A. J.; Chapman, V., Characterisation of the cannabinoid receptor system in synovial tissue and fluid in patients with osteoarthritis and rheumatoid arthritis. *Arthritis Res Ther* **2008,** *10* (2), R43.
46. Turcotte, C.; Blanchet, M. R.; Laviolette, M.; Flamand, N., The CB(2) receptor and its role as a regulator of inflammation. *Cell Mol Life Sci* **2016,** *73* (23), 4449-4470.
47. Li, X.; Chang, H.; Bouma, J.; de Paus, L. V.; Mukhopadhyay, P.; Paloczi, J.; Mustafa, M.; van der Horst, C.; Kumar, S. S.; Wu, L.; Yu, Y.; van den Berg, R.; Janssen, A. P. A.; Lichtman, A.; Liu, Z. J.; Pacher, P.; van der Stelt, M.; Heitman, L. H.; Hua, T., Structural basis of selective cannabinoid CB(2) receptor activation. *Nat Commun* **2023,** *14* (1), 1447.
48. Brust, C. A.; Swanson, M. A.; Bohn, L. M., Structural and functional insights into the G protein-coupled receptors: CB1 and CB2. *Biochem Soc Trans* **2023,** *51* (4), 1533-1543.
49. Hua, T.; Li, X.; Wu, L.; Iliopoulos-Tsoutsouvas, C.; Wang, Y.; Wu, M.; Shen, L.; Brust, C. A.; Nikas, S. P.; Song, F.; Song, X.; Yuan, S.; Sun, Q.; Wu, Y.; Jiang, S.; Grim, T. W.; Benchama, O.; Stahl, E. L.; Zvonok, N.; Zhao, S.; Bohn, L. M.; Makriyannis, A.; Liu, Z. J., Activation and Signaling Mechanism Revealed by Cannabinoid Receptor-G(i) Complex Structures. *Cell* **2020,** *180* (4), 655-665 e18.
50. Perez-Riverol, Y.; Bandla, C.; Kundu, D. J.; Kamatchinathan, S.; Bai, J.; Hewapathirana, S.; John, N. S.; Prakash, A.; Walzer, M.; Wang, S.; Vizcaino, J. A., The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res* **2025,** *53* (D1), D543-D553.

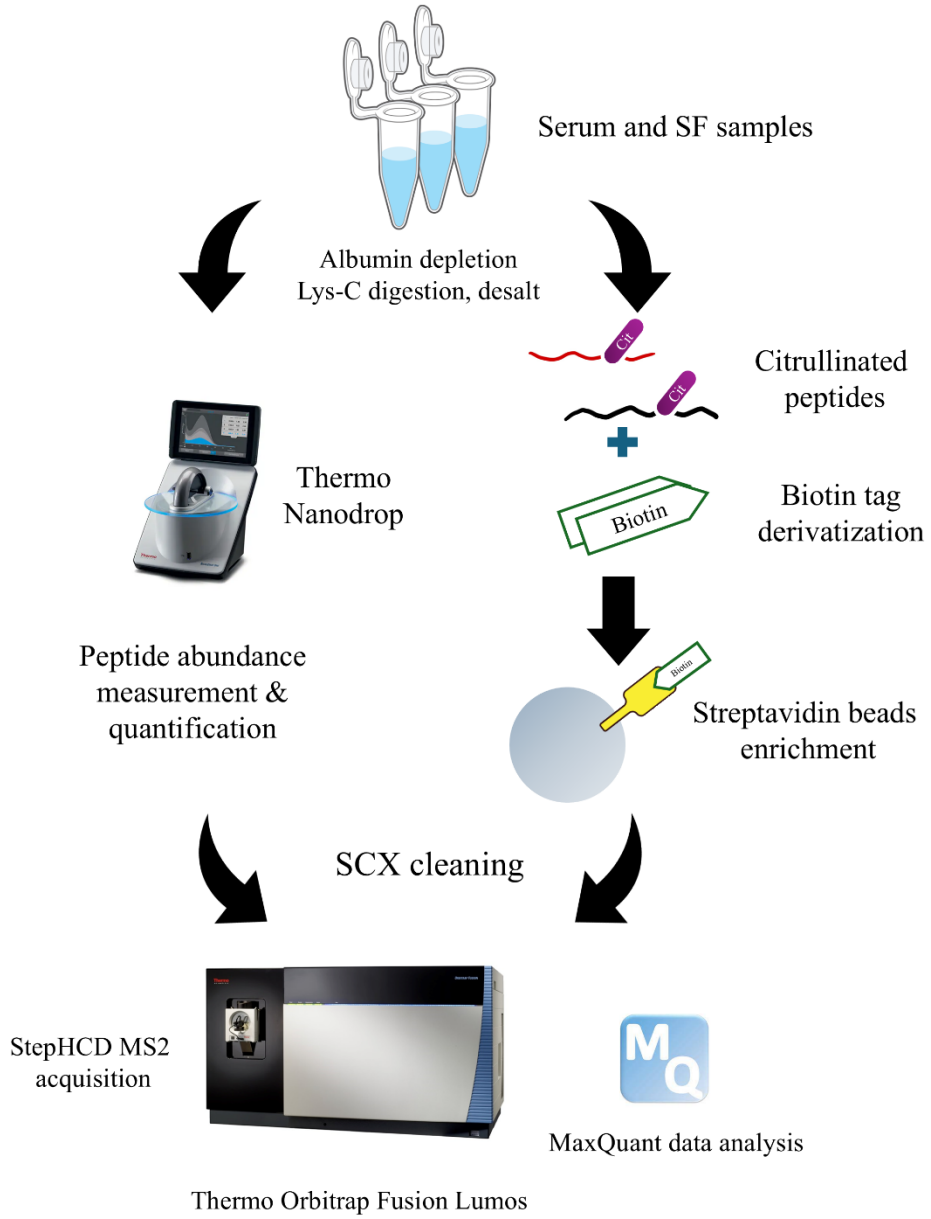


Figure 1. The experimental workflow for citrullination and protein abundance analysis. Human serum and SF samples were processed using the FASP method, followed by digestion into peptides. The resulting peptides were split into two equal portions. One portion was quantified using a Thermo NanoDrop to determine the relative abundance of serum/SF proteins. The other portion was used for citrullination site identification via biotin tag derivatization.

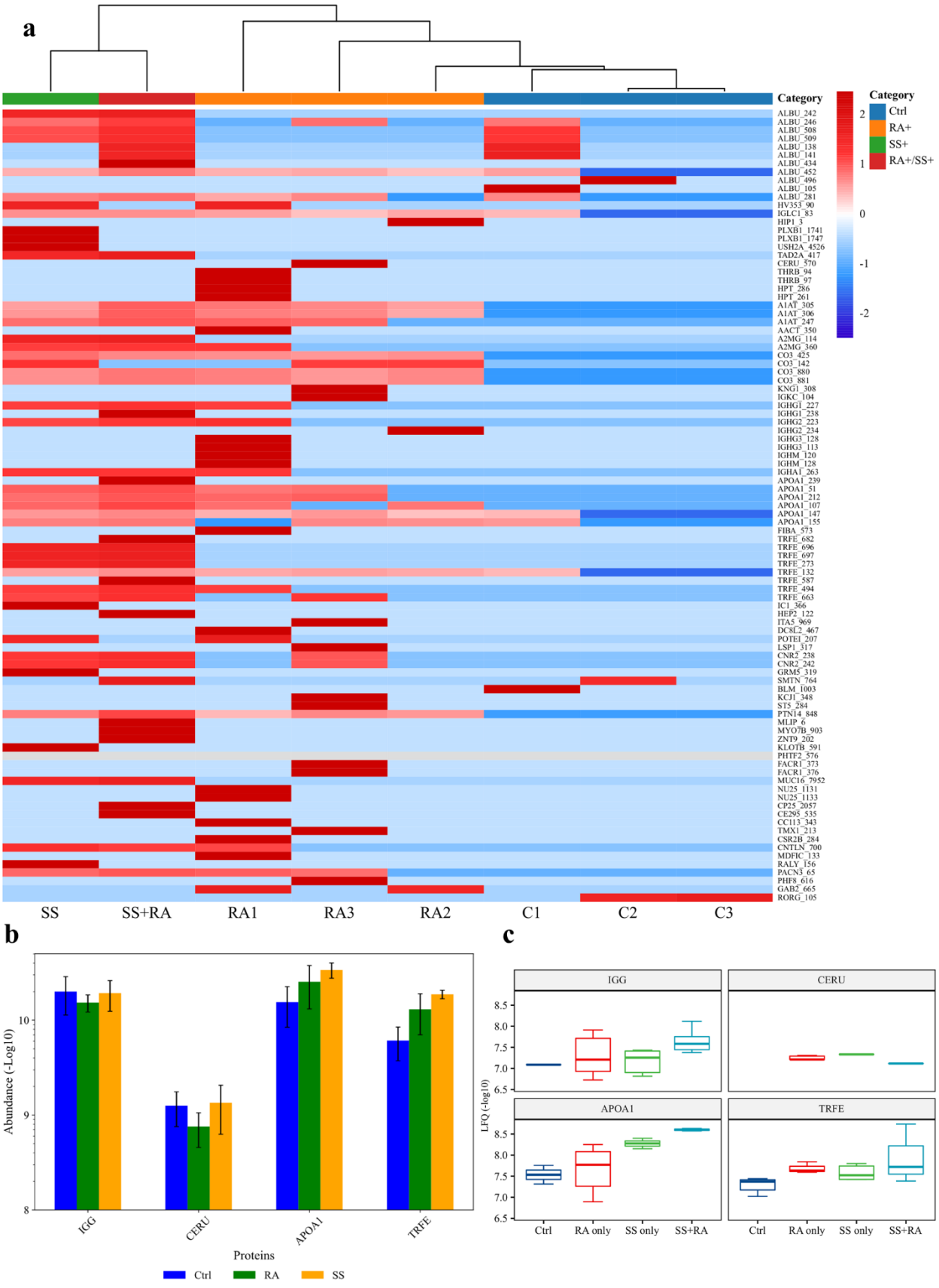

a
Category
Ctrl
RA+
SS+
RA+/SS+
2
1
0
-1
-2
SS
SS+RA
RA1
RA3
RA2
C1
C2
C3
b
Abundance (-Log10)
10
9
8
IGG
CERU
APOA1
TRFE
Proteins
Ctrl
RA
SS
c
IGG
CERU
APOA1
TRFE
LFQ (-log10)
8.5
8.0
7.5
7.0
6.5
Ctrl
RA only
SS only
SS+RA

Figure 2. Citrullination pattern variation between healthy control and autoimmune diseases (RA/SS) condition in serum samples. (a) Heatmap of LFQ citrullination sites across 8 different samples. (b) Protein abundance comparison of four essential plasma proteins in Ctrl, RA and SS condition. (c) The corresponding citrullination level quantification in four types of samples.

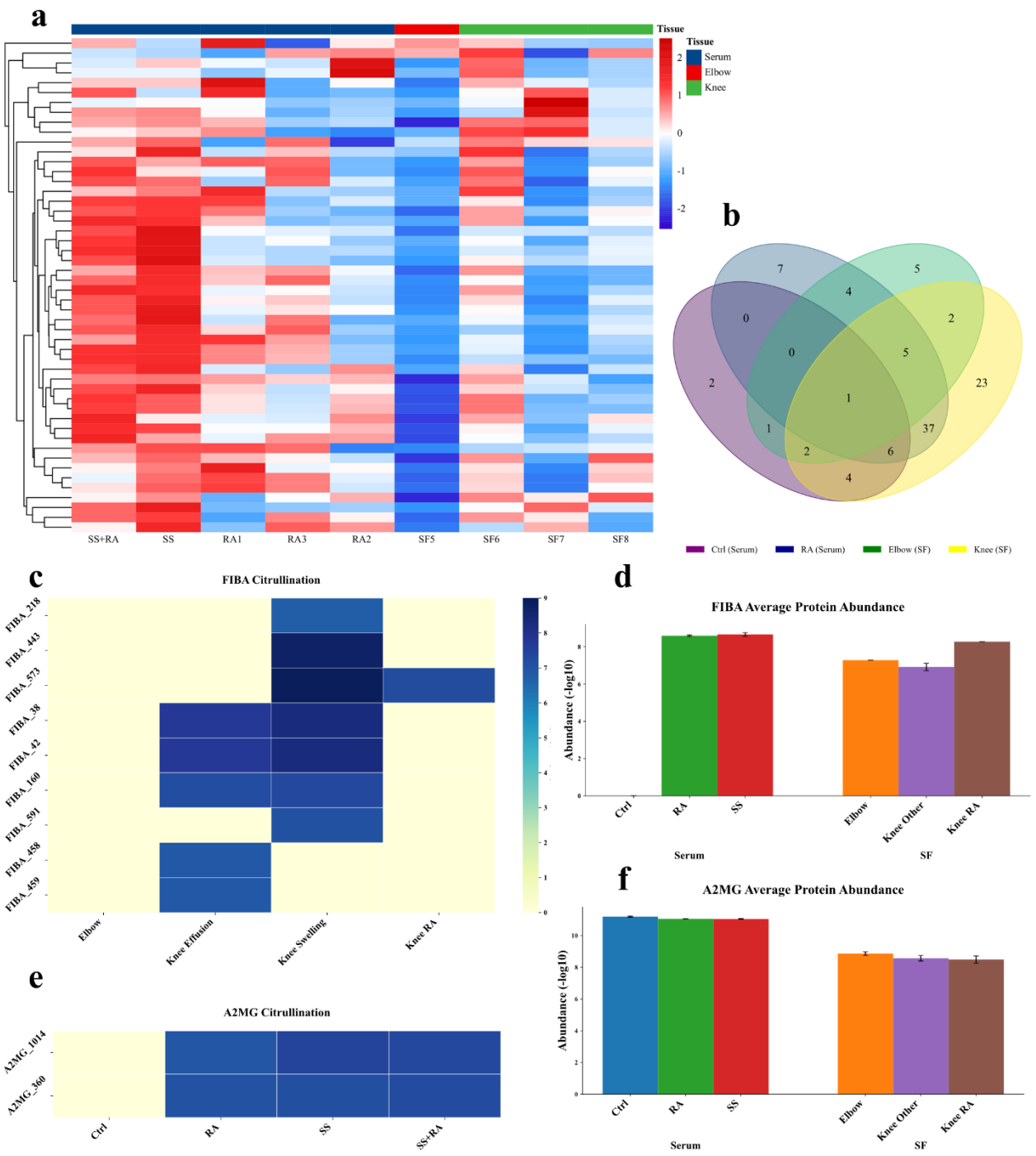


Figure 3. Citrullination pattern and corresponding protein level changes between serum and SF sample. (a) Heatmap clustering LFQ of citrullinated peptides. (b) Venn diagram of unique citrullinated protein in four categories. (c) FIBA and (e) A2MG site-specific citrullination patterns across different samples. (d) FIBA and (f) A2MG protein level comparison in serum and SF sample.

Figure 4. Enrichment analysis of citrullinated proteins in RA and associated disease. (a) Metascape analysis of disease-associated genes. (b) GO analysis provided by DAVID bioinformatics.

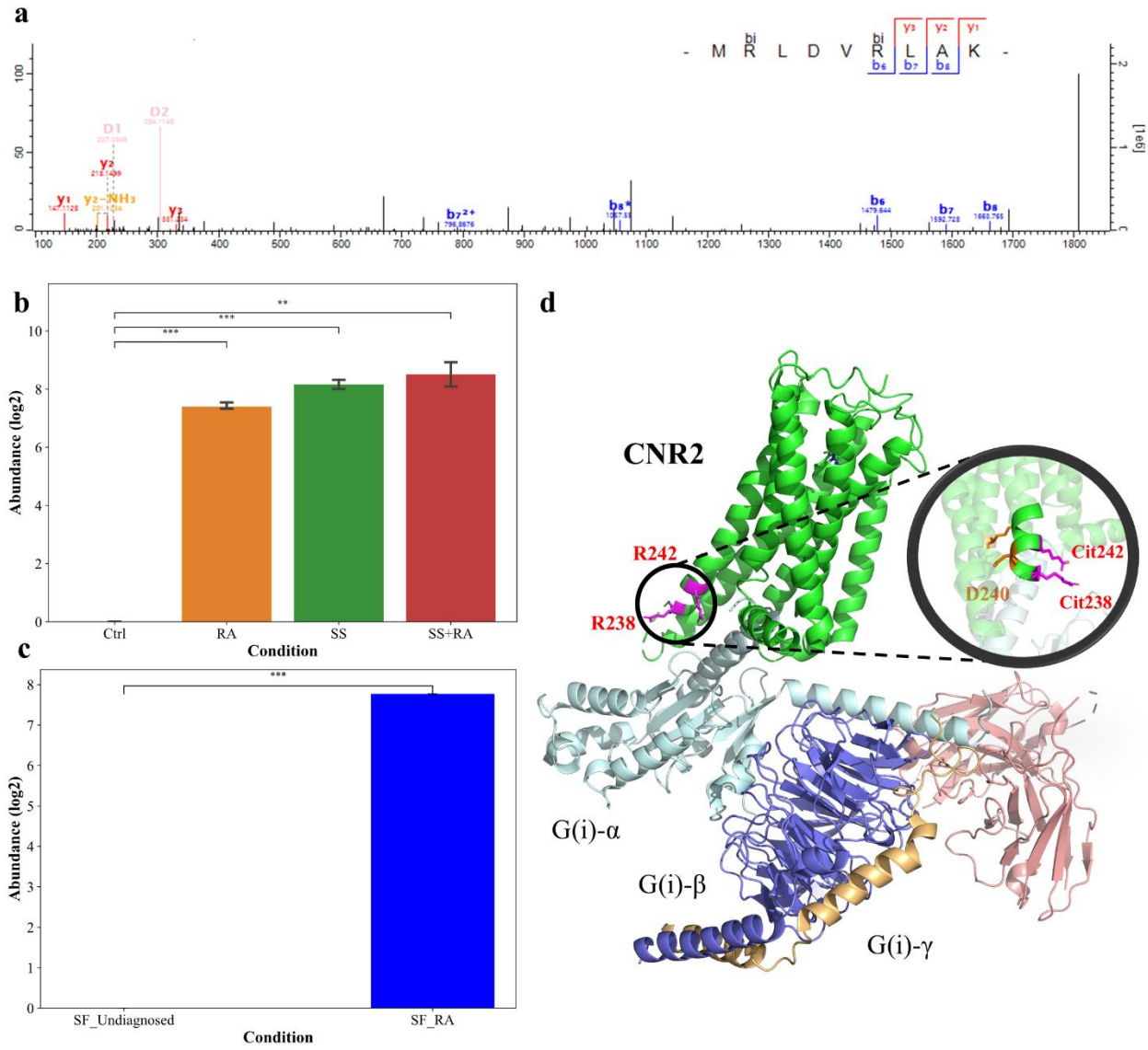


Figure 5. The unique citrullination upregulation of CNR2 in autoimmune diseases. (a) The annotated spectrum of citrullinated peptide of CNR2. (b) LFQ of the CNR2 R238 & R242 citrullination in serum and (c) SF sample through different conditions. (d) The 3D PyMOL structure of CNR2 with Gi complex (PDB: 8GUQ). The citrullinated arginine residues were highlighted in purple and the ionic lock between R131 (TM5) and D240 (TM6) were in orange.

Supplementary information

Figure S1. Citrullination and protein level variation in serum samples. **Figure S2.** The iBAQ curve of the serum and SF samples. **Figure S3.** Protein variation through four SF categories.

Table S1. Human serum and SF sample information. **Table S2.** List of all citrullination peptides identified in serum sample

Supplementary Figures

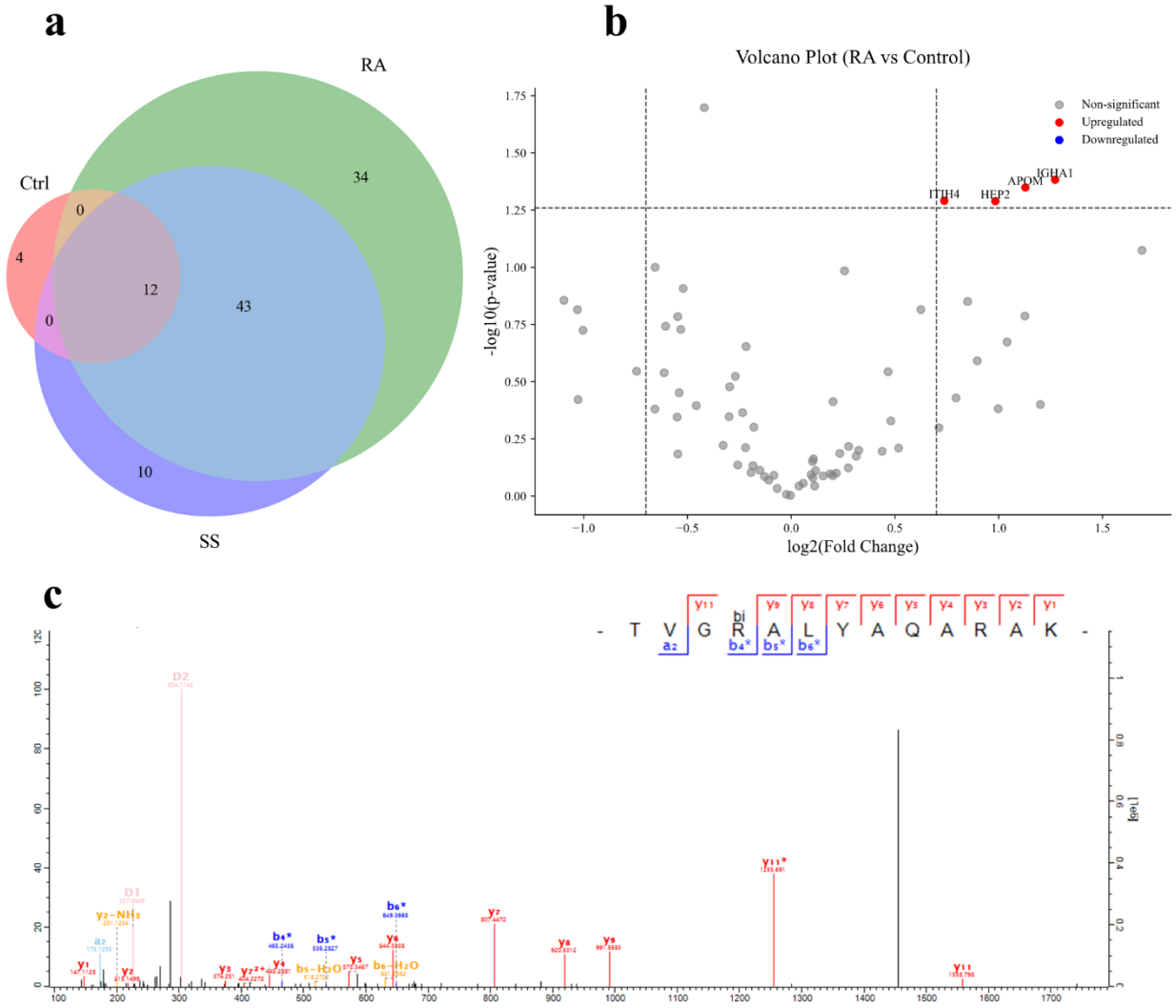


Figure S1. Citrullination and protein level variation in serum samples. (a) Venn diagram showing unique citrullination sites identified under each condition. (b) Volcano plot illustrating variation in serum protein levels. (c) MS/MS spectrum of a citrullinated ITIH peptide consistently detected across multiple samples. Biotin-tag-labeled citrullination sites are abbreviated as “bi” in sequence representation.

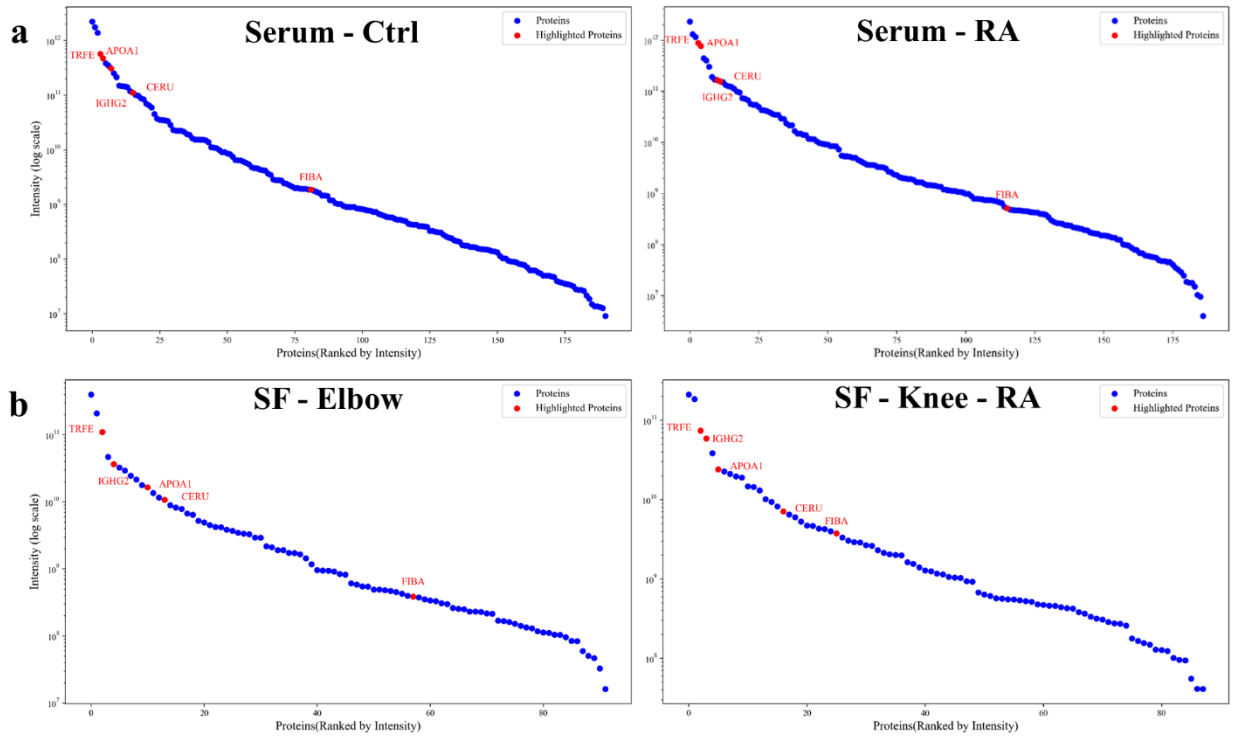


Figure S2. iBAQ curve of serum and synovial fluid (SF) samples. (a) iBAQ profile of serum samples from healthy controls and RA patients. (b) iBAQ profile of SF samples from elbow and knee joints in confirmed RA cases.

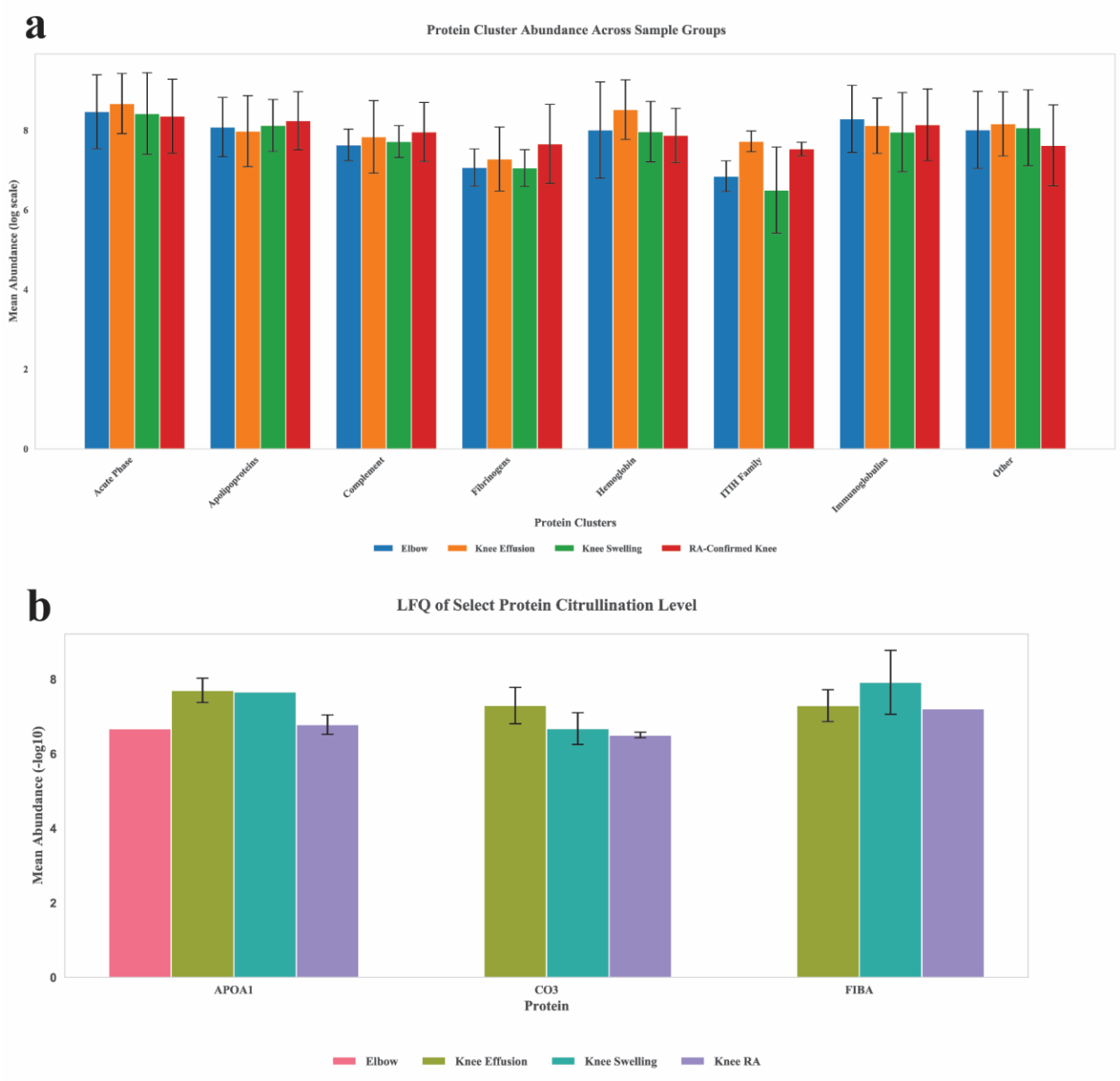


Figure S3. Protein variation across four synovial fluid (SF) categories. (a) Protein level variation in SF samples categorized by protein type. (b) LFQ-based quantification of citrullinated peptide levels in APOA1, CO3, and FIBA.

Supplementary Tables

Table S1. Human serum and SF sample information

Sample ID	**Abbreviation**	Diagnosis	Specimen	Group
RA1	**RA1**	Inflammatory polyarthritis	Serum	RA
RA4	**RA2**	Inflammatory arthropathy	Serum	RA
RA10	**RA3**	Rheumatoid arthritis	Serum	RA
RA3	**SS**	Sjogren syndrome	Serum	SS
RA2	**SS+RA**	Rheumatoid arthritis, Sjogren's disease	Serum	SS & RA
Control1	**C1**		Control Serum	Ctrl
Control2	**C2**		Control Serum	Ctrl
Control3	**C3**		Control Serum	Ctrl
RA8	**SF8**	Arthritis	SF, left knee	Knee_RA
RA15	**SF15**	Arthritis	SF, Right knee	Knee_RA
RA6	**SF6**	osteoarthritis	SF, Left knee	Knee_Effusion
RA13	**SF13**	osteoarthritis	SF, Left knee	Knee_Effusion
RA7	**SF7**	knee swelling, inflammation	SF, Right knee	Knee_Swelling
RA16	**SF16**	knee swelling, gout	SF, Right knee	Knee_Swelling
RA5	**SF5**	Elbow swelling, gout	SF, Left elbow	SF_elbow
RA17	**SF17**	Elbow swelling, gout	SF, right elbow	SF_elbow

Table S2. List of all citrullinated peptides identified in serum sample

Uniprot	Gene	Position	Name	Citrullinated peptides (Probability)	Score
P02768	ALBU	242	Serum albumin	AWAVAR(0.865)LSQRFPK	84.41
P02768	ALBU	246	Serum albumin	LSQR(1)FPK	134.26
P02768	ALBU	508	Serum albumin	CCTESLVNR(0.5)RPCFSALEVDETYVPK	146.1
P02768	ALBU	509	Serum albumin	CCTESLVNRR(0.5)PCFSALEVDETYVPK	146.1
P02768	ALBU	138	Serum albumin	DDNPNLPR(0.91)LVRPEVDVMCTAFHD NEETFLK	142.16
P02768	ALBU	141	Serum albumin	DDNPNLPRLVR(0.5)PEVDVMCTAFHDN EETFLK	142.16
P02768	ALBU	434	Serum albumin	FQNALLVR(1)YTK	69.345
P02768	ALBU	452	Serum albumin	KVPQVSTPTLVEVSR(1)NLGK	299.34
P02768	ALBU	105	Serum albumin	LCTVATLR(1)ETYGEMADCCAK	58.373
P02768	ALBU	281	Serum albumin	VHTECCHGDLLECADDR(1)ADLAK	238.72
P01767	HV353	90	Immunoglobulin heavy variable 3-53	GRFTISR(1)DNSK	104.97
P0CG04	IGLC1	83	Immunoglobulin lambda constant 1	SHR(1)SYSCQVTHEGSTVEK	190.8
O00291	HIP1	3	Huntingtin-interacting protein 1	MDR(1)MASSMK	40.724
O43157	PLXB1	1741	Plexin-B1	LLR(1)EDVEYR	53.756
O75445	USH2A	4526	Usherin	DR(1)TSPSAPSGMEPPKLQAR	46.543
O75478	TAD2A	417	Transcriptional adapter 2-alpha	LAQAR(1)ALIK	121.9
P00450	CERU	570	Ceruloplasmin	GSLHANGR(1)QK	107.09
P00734	THRB	94	Prothrombin	YTACETAR(0.5)TPRDK	58.079
P00734	THRB	97	Prothrombin	YTACETARTPR(0.5)DK	58.079
P00738	HPT	286	Haptoglobin	DYAEVGRVGYVSGWGR(1)NANFK	102.16
P00738	HPT	261	Haptoglobin	VSVNER(1)VMPICLPSK	77.08
P00738	HPT	311	Haptoglobin	YVMLPVADQDQCIR(1)HYEGSTVPEK	41.45
P01009	A1AT	305	Alpha-1-antitrypsin	FLENEDR(0.5)RSASLHLPK	74.505
P01009	A1AT	306	Alpha-1-antitrypsin	FLENEDRR(0.597)SASLHLPK	86.8
P01009	A1AT	247	Alpha-1-antitrypsin	R(1)LGMFNIQHCK	88.021
P01011	AACT	350	Alpha-1-antichymotrypsin	ADLSGITGAR(1)NLAVSQVVHK	81.477
P01023	A2MG	1014	Alpha-2-macroglobulin	AIGYLNTGYQR(1)QLNYK	104.88
P01023	A2MG	360	Alpha-2-macroglobulin	VDSHFR(1)QGIPFFGQVRLVDGK	71.882
P01024	CO3	425	Complement C3	LSINTHPSQKPLSITVR(1)TK	115.54
P01024	CO3	1042	Complement C3	R(1)QGALELIKK	123.95
P01024	CO3	880	Complement C3	R(0.5)RHQQTVTIPPK	92.939
P01024	CO3	881	Complement C3	RR(0.5)HQQTVTIPPK	92.939
P01042	KNG1	308	Kininogen-1	KAR(1)VQVVAGK	111.86
P01834	IGKC	104	Immunoglobulin kappa constant	SFNR(1)GEC	170.47
P01857	IGHG1	227	Immunoglobulin heavy constant gamma 1	GQPR(1)EPQVYTLPPSRDELTK	84.214
P01857	IGHG1	238	Immunoglobulin heavy constant gamma 1	GQPREPQVYTLPPSR(0.999)DELTK	59.55

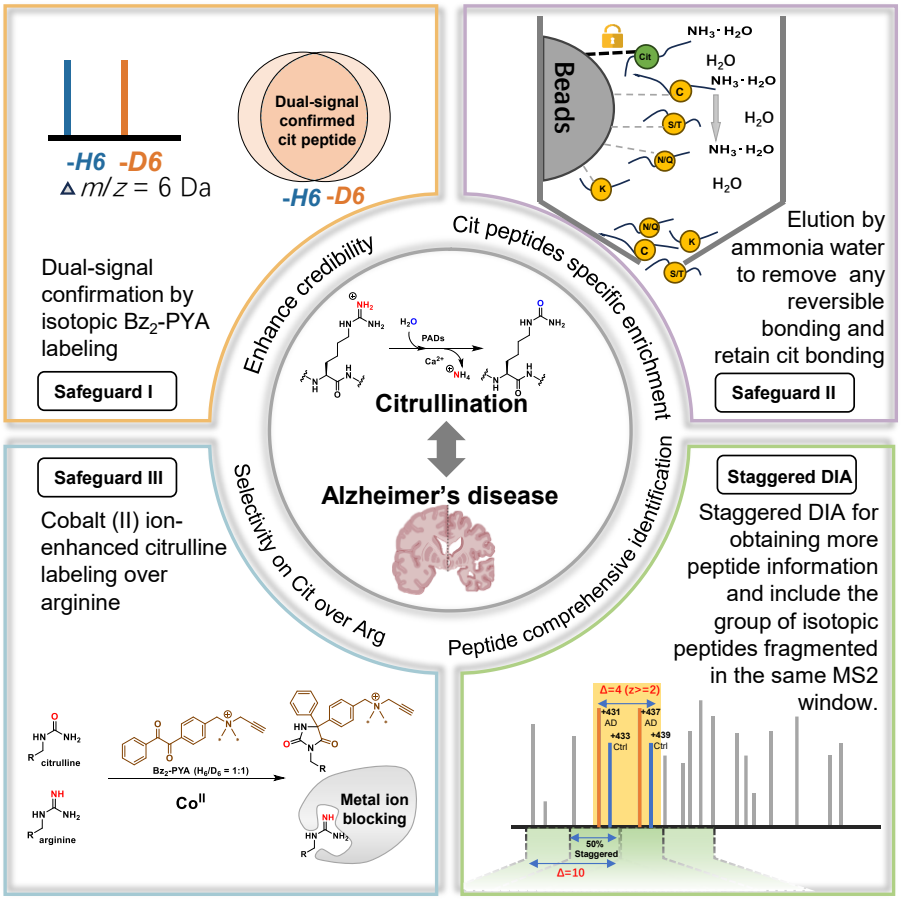
P01859	IGHG2	223	Immunoglobulin heavy constant gamma 2	GQPR(1)EPQVYTLPPSREEMTK	112.06
P01859	IGHG2	234	Immunoglobulin heavy constant gamma 2	GQPREPQVYTLPPSR(0.944)EEMTK	41.53
P01860	IGHG3	128	Immunoglobulin heavy constant gamma 3	SCDTPPPCPR(1)CPEPK	94.281
P01860	IGHG3	113	Immunoglobulin heavy constant gamma 3	TPLGDTTHTCPR(1)CPEPK	93.776
P01871	IGHM	120	Immunoglobulin heavy constant mu	VSVFVPPR(0.813)DGFFGNPRK	60.621
P01871	IGHM	128	Immunoglobulin heavy constant mu	VSVFVPPRDGFFGNPR(0.792)K	57.366
P01876	IGHA1	263	Immunoglobulin heavy constant alpha 1	DVLVR(1)WLQGSQELPREK	116.35
P02647	APOA1	239	Apolipoprotein A-I	AKPALEDLR(1)QGLLPVLESFK	87.792
P02647	APOA1	51	Apolipoprotein A-I	DSGR(1)DYVSQFEGSALGK	167.51
P02647	APOA1	212	Apolipoprotein A-I	LEALKENGGAR(1)LAEYHAK	120.23
P02647	APOA1	107	Apolipoprotein A-I	ETEGLR(1)QEMSK	199.04
P02647	APOA1	147	Apolipoprotein A-I	VEPLR(1)AELQEGARQK	84.653
P02647	APOA1	155	Apolipoprotein A-I	VEPLRAELQEGAR(1)QK	117.29
P02671	FIBA	573	Fibrinogen alpha chain	ESSSHHPGIAEFPSR(1)GK	108.71
P02787	TRFE	682	Serotransferrin	AVGNLR(1)K	103.74
P02787	TRFE	696	Serotransferrin	CSTSSLLEACTFR(0.5)RP	126.4
P02787	TRFE	697	Serotransferrin	CSTSSLLEACTFRR(0.5)P	126.4
P02787	TRFE	273	Serotransferrin	DCHLAQVPSHTVVAR(1)SMGGK	52.359
P02787	TRFE	132	Serotransferrin	KDSGFQMNQLR(1)GK	182.89
P02787	TRFE	587	Serotransferrin	DYELLCLDGTR(1)K	94.717
P02787	TRFE	494	Serotransferrin	INHCR(1)FDEFFSEGCAPGSK	148.29
P02787	TRFE	663	Serotransferrin	LHDR(1)NTYEK	185.37
P05155	IC1	366	Plasma protease C1 inhibitor	HR(1)LEDMEQALSPSVFK	61.942
P05546	HEP2	122	Heparin cofactor 2	SR(0.778)IQRLNILNAK	60.861
P08648	ITA5	969	Integrin alpha-5	ALKMPYR(0.5)ILPR	51.445
P0C7V8	DC8L2	467	DDB1- and CUL4-associated factor 8-like protein 2	GHR(1)NNTTVK	62.633
P0CG38	POTEI	207	POTE ankyrin domain family member I	KR(1)TALTK	112.71
P33241	LSP1	317	Lymphocyte-specific protein 1	STPSGKR(1)YK	48.794
P34972	CNR2	238	Cannabinoid receptor 2	MR(1)LDVRLAK	61.161

P34972	CNR2	242	Cannabinoid receptor 2	MRLDVR(1)LAK	61.161
P53814	SMTN	764	Smoothelin	TSASQAR(1)KAMIEK	44.367
P54132	BLM	1003	Bloom syndrome protein	LKR(1)LIMMEK	47.187
P78508	KCJ10	348	ATP-sensitive inward rectifier potassium channel 10	DSTVR(1)YGDPEKLK	57.288
P78524	ST5	284	Suppression of tumorigenicity 5 protein	ESSAVLSR(1)IQK	66.004
Q15678	PTN14	848	Tyrosine-protein phosphatase non-receptor type 14	EMMIR(1)NLEK	101.65
Q5VWP3	MLIP	6	Muscular LMNA-interacting protein	MELEKR(1)EK	62.582
Q6PIF6	MYO7B	903	Unconventional myosin-VIIb	R(1)SIYDTVTDTEMVEK	48.351
Q6PML9	ZNT9	202	Zinc transporter 9	KEAEIEYR(0.998)ER	70.028
Q86Z14	KLOTB	591	Beta-klotho	KQLEMLAR(1)MK	42.843
Q8WVX9	FACR1	373	Fatty acyl-CoA reductase 1	MTGR(0.5)SPRMMK	42.336
Q8WVX9	FACR1	376	Fatty acyl-CoA reductase 1	MTGRSPR(0.5)MMK	42.336
Q8WXI7	MUC16	7952	Mucin-16	R(1)GPEDVSWMSPPLLEK	62.104
Q92621	NU205	1131	Nuclear pore complex protein Nup205	VTSLNR(0.5)QR	90.308
Q9BV73	CP250	2057	Centrosome-associated protein CEP250	HQQER(1)EQLLEK	64.64
Q9C0D2	CE295	535	Centrosomal protein of 295 kDa	LR(1)LETDCFR	44.318
Q9H0I3	CC113	343	Coiled-coil domain-containing protein 113	MWER(1)KVEIAEMSLK	44.045
Q9H3N1	TMX1	213	Thioredoxin-related transmembrane protein 1	RR(0.998)PQPYPYPSK	64.224
Q9H8E8	CSR2B	284	Cysteine-rich protein 2-binding protein	EAAGFLDR(1)STSSTPVK	41.513
Q9NXG0	CNTLN	700	Centlein	VTELENR(1)LK	111.52
Q9P1T7	MDFIC	133	MyoD family inhibitor domain-containing protein	MHR(1)KIQSSLSVNSDISK	49.595
Q9UKM9	RALY	156	RNA-binding protein Raly	VTVPLVR(0.5)R	91.867

Q9UKS6	PACN3	65	Protein kinase C and casein kinase substrate in neurons protein 3	WR(1)GTVEK	116.73
Q9UPP1	PHF8	616	Histone lysine demethylase PHF8	R(1)VKSLSK	105.52

Chapter 3

Triple safeguard-enhanced chemoproteomic platform for reliable citrullination mapping with staggered DIA detection in Alzheimer's disease



Adapted from: Li, W.; **Zhu, Z.**; Fields, L.; Lu, K.H.; Lu, H.; Wang, Z.; Ibarra A.; Yao, Y., Nassar, A.; Li, L. Triple safeguard-enhanced chemoproteomic platform for reliable citrullination mapping with staggered DIA detection in Alzheimer's disease. *Nat. Commun.* 2025, To be submitted. **Zhu, Z** contributed to experimental design in DDA, data collection and manuscript preparation in AD sample comparison.

Abstract

Citrullination, the enzymatic conversion of arginine to citrulline, is a post-translational modification implicated in immune dysregulation, neurodegeneration, and autoimmunity. However, its global mapping remains challenging due to the chemical similarity between citrulline and arginine, the reversibility of probe adducts, and high false-positive rates in mass spectrometry (MS) detection. Here, we report the development of a third-generation citrulline-specific probe, Bz2-PYA, featuring a benzil electrophile for irreversible capture, an alkyne handle for bioorthogonal conjugation, and a quaternary ammonium for MS signal enhancement. To ensure specificity, we implemented three safeguards: (i) dual-isotopic H6/D6 labeling for MS1 validation; (ii) ammonia elution to remove reversible non-citrulline adducts; and (iii) cobalt(II)-mediated blocking of guanidinium groups to suppress arginine cross-reactivity. The probe selectively labeled synthetic citrullinated peptides, recombinant histone H2B, and complex proteomes with high specificity and sensitivity. Application to Alzheimer's disease (AD) and control sera, followed by CuAAC biotinylation, enrichment, and LC–MS/MS in both data-dependent and staggered data-independent acquisition (DIA) modes, enabled the identification of >350 citrullination sites per sample with markedly reduced variability (CV = 9.7% in DIA vs. 40.6% in DDA). Comparative profiling revealed numerous AD-enriched citrullination events, particularly on proteins central to lipid transport and complement activation, including APOE, A2M, C3, and C4B. Functional enrichment analyses highlighted immune and lipid-regulatory pathways, consistent with mechanisms of neuroinflammation and amyloid clearance in AD.

Introduction

Protein citrullination, also known as deimination, is a post-translational modification (PTM) in which peptidyl-arginine residues are enzymatically converted to citrulline by calcium-dependent protein arginine deiminases (PADs)[1]. This conversion eliminates a positive charge on the arginine side chain, often causing significant changes in protein structure, interactions, and function. Citrullination has long been implicated in autoimmune and inflammatory disorders such as rheumatoid arthritis (RA), where aberrantly citrullinated proteins trigger anti-citrullinated protein antibodies[2, 3]. Hypercitrullination of myelin basic protein is thought to drive myelin sheath destabilization[4, 5]. More recently, there is growing recognition that dysregulated citrullination plays a role in neurodegenerative diseases including prion disease and Alzheimer's disease (AD). In AD, studies have observed abnormal accumulation of citrullinated proteins in patient brains, particularly in the hippocampus. Neuronal PAD4 expression is also elevated in some AD cases and is associated with the production of autoantibodies against citrullinated neural proteins[6]. Such findings suggest that protein citrullination could contribute to AD pathogenesis, adding urgency to the development of tools for mapping citrullinated proteins in the context of neurodegeneration.

Despite the biological importance of citrullination, its comprehensive analysis remains challenging. Traditional antibody-based methods (e.g., Western blotting with anti-modified citrulline reagents or immunohistochemistry) can qualitatively detect protein citrullination but are low-throughput and cannot pinpoint modification sites[7]. Mass spectrometry (MS) offers a powerful alternative for site-specific and proteome-wide characterization of PTMs,

yet citrullination poses several analytical hurdles. First, citrullinated peptides are typically low in abundance and can be suppressed by the vast excess of unmodified peptides in digests. Enrichment strategies are therefore required to isolate citrullinated peptides from complex samples. Second, the mass shift introduced by converting arginine to citrulline is only +0.984 Da, which is exceedingly small. This $\Delta m \approx 1$ Da is easily confounded with natural isotopic ^{13}C peaks (+1.003 Da) and, more problematically, is identical to the mass increment of deamidation on asparagine/glutamine (+0.984 Da). As a result, conventional database searches struggle to distinguish true citrullination from spurious identifications of deamidated peptides, leading to high false-positive rates. The weak MS/MS fragmentation of citrulline-containing peptides (due to the loss of basicity) further complicates confident site localization. To address these challenges, chemical derivatization strategies have been proposed. For example, our lab developed biotin thiol tag using the reagent 2,3-butanedione, which can selectively react with the ureido group of citrulline, forming stable adducts that introduce a larger mass shift and a chemical handle for enrichment[7-9]. Using these methods, our group achieved the largest reported citrullinome to date, including hundreds of novel citrullination sites across diverse proteins. Despite these advances, drawbacks remain, including multiple sample-handling steps, potential peptide losses during enrichment, and the risk of arginine cross-reactivity under derivatization conditions.

Here, we introduce a triple-safeguard Bz2-PYA probe to overcome these limitations. The design incorporates: (i) dual isotopic labeling (H_6/D_6) to generate characteristic doublet peaks at MS1, ensuring confident identification of citrullinated peptides analogous to SILAC-like

internal validation; (ii) irreversible benzil adduction of the citrulline ureido group, coupled with ammonia elution to strip away reversible adducts from other nucleophilic amino acids; and (iii) cobalt(II) ion pre-incubation to selectively form coordination with guanidinium groups and prevent arginine modification. Together, these three safeguards significantly reduce false positive rates while enhancing labeling efficiency and specificity. Furthermore, by implementing the workflow with data-independent acquisition (DIA) MS, we maximize coverage, reproducibility, and quantification accuracy. DIA strategies enable fragmenting all ions within defined *m/z* windows, reducing stochastic undersampling inherent in data-dependent acquisition (DDA), and have been shown to improve peptide detection and reproducibility in PTM proteomics[10]. Our combined Bz2-PYA and DIA platform therefore enables robust and large-scale mapping of citrullination sites with unprecedented specificity, laying a foundation for systematic investigation of citrullination biology in AD and beyond.

Experimental Section

Materials and Chemical Reagents

Furan (≥98%), 2-methylfuran (2-MF, ≥98%), 2,5-dimethylfuran (2,5-DMF, ≥98%), and para-bromobenzyl mercaptan (BBM, ≥98%) were purchased from Aladdin Biotechnology Co., Ltd. (Shanghai, China). Pronase E, chymotrypsin, and rat tail collagen I were obtained from SalarBio (Beijing, China). Recombinant human protein-arginine deiminase 4 (PAD4, Cayman

Chemical, item No. 10500) was used for in vitro citrullination. All other reagents and solvents were of analytical or HPLC grade and sourced from commercial suppliers.

Optimization of Bz2-PYA Probe Labeling at the Protein Level

Recombinant histone H2B (2 μM) was incubated with PAD4 (0.2 μM) in 50 mM HEPES buffer (pH 8.0) containing 300 mM NaCl, 1 mM DTT, and 10% glycerol at 37 °C for 1 h. Cobalt(II) chloride (1.0 mM) and 8-hydroxyquinoline (1.0 mM) were added, followed by incubation at 50 °C for 10 min. The mixture was then adjusted to pH 12 with 4 mM NaOH and treated with varying concentrations of Bz2-PYA-H6 (500, 100, 10, and 1 μM) for 30 min at 37 °C. The reaction was quenched by neutralizing to pH 7.0 with HCl (4 mM). Proteins were precipitated with ice-cold methanol, washed twice, and re-dissolved in 20 μL of 1.2% SDS in PBS by sonication at 95 °C for 10 min.

Labeled proteins were conjugated with Fluor 488 Propyl Azide (100 μM, BroadPharm, Cat. No. BP-40668) via Cu(I)-catalyzed azide–alkyne cycloaddition (CuAAC) under standard conditions. Samples were boiled with 5× SDS loading buffer, separated by SDS-PAGE (12%, 150 V, 1 h), and imaged for fluorescence (Ex. 495 nm; Em. 519 nm). Gels were subsequently stained with Coomassie Brilliant Blue and imaged.

For probe selectivity validation testing, PAD4 incubation was omitted in the negative control group. To evaluate the contribution of cobalt(II) ions, reactions were performed in the presence or absence of 1 mM $CoCl_2$, using 100 μM Bz2-PYA-H6.

Selective Derivatization of Proteomes with Isotopically Mixed Bz2-PYA and Biotinylation

Serum samples from Alzheimer's disease (AD) patients and healthy donors (50 μL, 1 mg/mL, n = 3 per group) were diluted into 490 μL PBS. Samples were incubated with $CoCl_2$ (final concentration = 1.0 mM) and 8-hydroxyquinoline (final concentration = 1.0 mM) at 50 °C for 10 min. A premixed isotopic probe (Bz2-PYA-H6:D6 = 1:1, 5 μL of 2 mM stock) was added, and the pH was adjusted to 12. After incubation at ambient temperature for 1 h, the reaction mixture was neutralized.

CuAAC was performed with isotopically labeled DADPS biotin azide reagents (20 mM stock in DMSO; "light" H_2 version for AD samples, "heavy" D_2 version for control samples)[11]. For each reaction, 6 μL DADPS biotin azide, 2 μL TCEP (50 mM), 6 μL TBTA (1.7 mM in DMSO/t-butanol, 1:4), and 2 μL $CuSO_4$ (50 mM) were added and incubated for 1 h at ambient temperature. Excess reagents were removed using Amicon Ultra-0.5 centrifugal filters (Ultracel 3k, Cat. No. UPC500396). Proteins were reconstituted in 100 μL of 1.0% SDS in PBS, then diluted to 0.2% SDS with 400 μL PBS. Protein concentrations were measured by NanoDrop (Thermo Scientific, Cat. No. 13-400-526), and AD/control samples were mixed in a 1:1 ratio.

Protein Digestion and Enrichment

Streptavidin agarose resin slurry (50 µL, Pierce, Cat. No. 20349) was washed twice with PBS and resuspended in 500 µL PBS. Protein mixtures were incubated with the resin for 2 h at room temperature with rotation. Beads were pelleted (4,000 g, 1 min) and washed sequentially with PBS (2×), 10 mM ammonium hydroxide (2×), and water (2×).

Proteins bound to beads were reduced with 5 mM DTT (30 min, 24 °C) and alkylated with 10 mM iodoacetamide (30 min, 24 °C, dark). Beads were resuspended in 200 µL of 0.5 M urea in PBS and digested with Lys-C (1.5 µg) for 12 h at 37 °C with rotation. The supernatant was collected, and beads were washed with PBS and water; washes were combined with the digest ("trypsin fraction").

The biotin linker was cleaved with 200 µL of 2% formic acid in water (30 min, 24 °C), and beads were washed with 80% acetonitrile (400 µL). The cleavage eluates were pooled ("cleavage fraction"). Both fractions were dried in a SpeedVac (40 °C) and desalted using Sep-Pak C18 cartridges (Waters). Peptides were stored at –20 °C until LC–MS/MS analysis.

LC–MS/MS Analysis

Peptides were separated using a self-packed nano-C18 column (17 cm × 75 µm i.d., 1.7 µm Ethylene Bridged Hybrid C18 particles) with an integrated emitter. A Dionex Ultimate 3000 nanoLC was coupled to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). Samples were dissolved in 3% acetonitrile, 0.1% formic acid (FA) prior to injection. Mobile phase A was 0.1% FA in water; mobile phase B was 0.1% FA in acetonitrile. A 100-min linear gradient from 3–30% B at 0.3 µL/min was used.

Full MS scans were acquired from *m/z* 350–1500 at 60,000 resolution (at *m/z* 200), with an AGC target of 2×10^5 and a maximum injection time of 100 ms. Fragmentation was performed in data-dependent and staggered data-independent acquisition (DIA) modes as described below. Detailed instrument settings are provided in the Supporting Information[10].

Data Analysis

Raw data were processed using PEAKS Studio 11 (Bioinformatics Solutions Inc.). Searches were performed against the UniProt reviewed human database (February 2024). Enzyme specificity was set to Lys-C, with up to three missed cleavages. Fixed modification: carbamidomethylation of Cys (+57.02146 Da). Variable modifications included Met oxidation (+15.99492 Da), citrullination (+0.984 Da), and Bz2-PYA adducts (+431.2209, +433.2365, +437.2675, +439.2835). Peptide-spectrum matches were filtered to 1% FDR at peptide and protein levels. Citrullination site localization was considered confident at ≥0.75 probability.

Bioinformatic analyses (Sankey diagrams, arc plots, volcano plots, and heatmaps) were performed in R (v4.3.2) with packages ggplot2, ggraph, and ComplexHeatmap.

Spectral Library Construction and DIA Analysis

Spectra from DDA acquisition were analyzed via FragPipe (v22.0). The MSFragger search included the following parameters: precursor mass tolerance, 20 PPM; fragment mass tolerance, 0.02 Da; maximum variable modifications per peptide, 5; enzyme, strict trypsin. Fixed modifications were set to carbamidomethylation of cysteine (+57.02146 Da). Variable

modifications included oxidation of methionine (+15.9949), N-termini acetylation (+42.0106), and mass tags on arginine (+431.2209, +433.2365, +437.2675, +439.2835). Quantification was enabled through the IonQuant intensity extraction tool using default settings. Validation was performed using MSBooster, Percolator with a minimum probability of 0.5, Philospher with a protein FDR of 0.01. A spectral library was generated and converted as a pepXML format.

Analysis of DIA data was conducted in DIA-NN (v1.9.2). Prior to analysis, all spectral files were converted to .mzML format via MSConvert.7 Parameters were remained as default, including a Q-value threshold of 0.01, and the spectrum library compiled from DDA data was imported for reference. Corresponding proteomics annotations were retrieved via UniProt API.

Results and Discussion

Design and Characterization of the Citrulline-Capturing Probe Bz2-PYA

The Bz2-PYA probe was designed to selectively and irreversibly label citrulline residues via their alkyl urea group[12]. Its modular architecture (**Figure 1a**) incorporates: (i) a benzil-based electrophile for citrulline capture; (ii) an alkyne group for downstream CuAAC bioconjugation; and (iii) a quaternary ammonium group that enhances water solubility, provides inherent MS signal enhancement, and enables isotopic encoding[13].

Structural characterization of the model adduct formed between Bz2-PYA and N-ethylurea confirmed the covalent conjugation pattern. The conjugate (compound 14) exhibited a molecular ion [M]+ at *m/z* 376.211 without the need for protonation or adduct formation.

NMR (^{1}H and HMBC) analysis revealed that the alkyl substituent adopts the sterically favored distal position, consistent with the proposed mechanism.

Early iterations of probe design, including sulfonated benzil derivatives and pyridinium-substituted analogues, were either synthetically unstable or overly reactive[14, 15]. Relocating the cationic nitrogen to a benzylic position yielded a stable, water-soluble probe with tunable reactivity. Bromine was initially incorporated for isotopic tagging, but this was later replaced by deuterated methyl groups, providing a 6 Da isotopic pair with improved MS resolution.

Safeguard I – Dual Isotopic Signal Validation

Bz2-PYA was used as a premixed 1:1 isotopic pair (H6 and D6). Consequently, each labeled citrullinated peptide appeared as two MS1 peaks separated by 6 Da. Only peptides detected as isotopic pairs were considered valid identifications, thereby reducing false positives. The quaternary ammonium ensured strong MS response and maintained probe solubility at concentrations >500 mM, while the relatively low probe mass (306–312 Da) minimized spectral complexity.

Safeguard II – Irreversible Bonding to Citrulline and Reversible Adduct Removal

The adduction mechanism between citrulline's alkyl urea group and the benzil moiety is irreversible, driven by intramolecular cyclization and 1,2-aryl migration. In contrast, other nucleophilic amino acids form only reversible adducts, which can be displaced by ammonium hydroxide elution. This property allows selective retention of citrullinated peptides during streptavidin enrichment. Incubation of individual amino acids with Bz2-PYA confirmed that

reversible adducts (e.g., Lys, Ser, Thr, Tyr, Gln, His, etc.) were efficiently removed upon NH_4OH treatment, while citrulline adducts were retained[16]. Arginine displayed partial retention, prompting the introduction of Safeguard III.

Safeguard III – Transition Metal Blocking of Arginine Conjugation

Because guanidinium groups preferentially form coordination with transition metals through their lone pairs of electrons, cobalt(II) chloride was screened as a selective blocking agent. Model reactions with N-ethylurea and N-ethylguanidine showed that $CoCl_2$, $AgNO_3$, and $CeCl_3$ suppressed arginine labeling by Bz2-PYA, with $CoCl_2$ providing the most consistent selectivity (**Figure 2f**). This "metal blocking" strategy effectively prevented arginine adduction while preserving citrulline reactivity.

Validation at the Peptide and Protein Levels

Synthetic peptides containing citrulline (AALAELNQL[Cit]) or arginine (AALAELNQLR) were tested. Without $CoCl_2$, both residues were labeled; with $CoCl_2$ pre-incubation, only citrullinated peptides yielded Bz2-PYA adducts, while arginine labeling was suppressed. Dual-isotopic labeling generated paired MS1 signals separated by 6 Da, validating Safeguard I.

At the protein level, recombinant histone H2B was citrullinated by PAD4 and subsequently labeled with Bz2-PYA. Following CuAAC with Fluor 488 azide, SDS-PAGE and fluorescence imaging confirmed probe labeling only in citrullinated samples. Omission of $CoCl_2$ led to low-level background labeling of non-citrullinated protein, whereas $CoCl_2$ pre-

incubation eliminated this effect. Probe concentration and reaction time optimization revealed maximal labeling at 5 mM probe and 60 min incubation (**Figure 2h–i**).

Serum Proteome Labeling and Enrichment Workflow

AD and control serum samples (n = 4 each) were depleted of abundant albumin, labeled with isotopic Bz2-PYA, biotinylated with isotopically encoded DADPS reagents, digested, enriched on streptavidin beads, and selectively eluted (**Figure 3a–b**). Labeled peptides contained mass increments of +431.2209/+437.2675 Da (AD) and +433.2365/+439.2835 Da (control).

In DDA mode, labeled peptides appeared as four-peak clusters in MS1 spectra (dual Bz2-PYA isotopes × dual DADPS isotopes). To ensure these clusters were captured during fragmentation, a staggered DIA strategy was implemented, using 10 *m/z* windows with 50% overlap to guarantee co-fragmentation of isotopic pairs (**Figure 3d–e**).

Citrullination Mapping in AD and Control Serum

Representative MS1 and MS2 spectra confirmed citrullinated peptide identification by dual isotopic validation (**Figure 4a–b**). Enrichment efficiency was verified by comparing enriched versus flow-through fractions: citrullinated peptides were significantly enriched ($p < 0.05$), while non-citrullinated peptides predominated in the flow-through fractions (**Figure 5a**). Comparison of DDA and staggered DIA demonstrated that DIA identified 300–400 citrullination sites per sample, with lower variability (CV = 9.7%) than DDA (100–200 sites, CV = 40.6%) (**Figure 5b**).

Overlap analysis showed 128–166 sites common to AD and controls, 115–140 sites unique to AD, and 17–30 sites unique to controls (**Figure 5c**). AD-specific citrullination events were observed on key proteins, including Complement C3 (C3), Alpha-2-macroglobulin (A2M), Apolipoprotein E (APOE), Apolipoprotein A-I (APOA1), Apolipoprotein A-IV (APOA4), and Complement C4B (C4B).

Functional enrichment analysis revealed that citrullinated proteins in AD were predominantly associated with immune response (acute-phase response, complement activation, adaptive immunity) and lipid regulation (cholesterol transport, lipoprotein metabolism) (**Figure 6a–c**). Subcellular annotation indicated enrichment in extracellular vesicles and blood microparticles. Volcano plot analysis highlighted AD-specific citrullination sites, including IGLV3-21_R72, APOB_R2705 and R2019, C3_R1427, APSC_R139, A2M_R117, and APOA1_R147 (**Figure 6d**). Several citrullination sites exhibited elevation in AD relative to controls included A2M_R1271, ALB_R210, and C4B_R297.

Mapping citrullination sites onto protein structures showed preferential modification in flexible, surface-exposed regions. Notably, APOE (particularly the ε4 isoform), A2M, and C4B—all implicated in amyloid processing and neuroinflammation—exhibited the highest level of citrullinated proteins in AD serum (**Figure 7**).

Conclusion

In this study, we developed and validated a next-generation chemical probe, Bz2-PYA, for the selective and irreversible labeling of citrulline residues. By combining three safeguards—dual isotopic tagging, reversible adduct removal, and transition-metal blocking of arginine—we established a robust workflow that effectively overcomes longstanding barriers in citrullination mapping. Application to AD and control serum proteomes demonstrated that this approach not only enriches citrullinated peptides with high specificity but also enables reproducible, large-scale site identification through staggered DIA acquisition.

Our findings revealed extensive AD-associated citrullination events, particularly within immune- and lipid-related pathways, implicating proteins such as APOE, A2M, and complement components as potential functional nodes of dysregulated citrullination in disease progression. Beyond Alzheimer's disease, the platform offers broad utility for mapping citrullination in autoimmune and inflammatory contexts. Future work will expand multiplexing capacity and apply the method to larger clinical cohorts to assess the diagnostic and mechanistic value of citrullination signatures.

Collectively, this triple-safeguard strategy establishes Bz2-PYA as a versatile tool for citrullination proteomics and highlights its potential to illuminate disease-relevant biology with unprecedented specificity and confidence.

References

1. György, B.; Tóth, E.; Tarcsa, E.; Falus, A.; Buzás, E. I., Citrullination: A posttranslational modification in health and disease. *International Journal of Biochemistry & Cell Biology* **2006,** *38*, 1662-1677.
2. Wu, C. Y.; Yang, H. Y.; Luo, S. F.; Lai, J. H., From Rheumatoid Factor to Anti-Citrullinated Protein Antibodies and Anti-Carbamylated Protein Antibodies for Diagnosis and Prognosis Prediction in Patients with Rheumatoid Arthritis. *Int J Mol Sci* **2021,** *22* (2).
3. Tutturen, A. E.; Fleckenstein, B.; de Souza, G. A., Assessing the citrullinome in rheumatoid arthritis synovial fluid with and without enrichment of citrullinated peptides. *Journal of Proteome Research* **2014,** *13*, 2867-2873.
4. van Beers, J. J.; Schwarte, C. M.; Stammen-Vogelzangs, J.; Oosterink, E.; Bozic, B.; Pruijn, G. J., The rheumatoid arthritis synovial fluid citrullinome reveals novel citrullinated epitopes in apolipoprotein E, myeloid nuclear differentiation antigen, and beta-actin. *Arthritis Rheum* **2013,** *65* (1), 69-80.
5. Romero, V.; Fert-Bober, J.; Nigrovic, P. A.; Darrah, E.; Haque, U. J.; Lee, D. M.; van Eyk, J.; Rosen, A.; Andrade, F., Immune-mediated pore-forming pathways induce cellular hypercitrullination and generate citrullinated autoantigens in rheumatoid arthritis. *Sci Transl Med* **2013,** *5* (209), 209ra150.
6. Mondal, S.; Thompson, P. R., Protein Arginine Deiminases (PADs): Biochemistry and Chemical Biology of Protein Citrullination. *Acc Chem Res* **2019,** *52* (3), 818-832.
7. Shi, Y.; Li, Z.; Wang, B.; Shi, X.; Ye, H.; Delafield, D. G.; Lv, L.; Ye, Z.; Chen, Z.; Ma, F.; Li, L., Enabling Global Analysis of Protein Citrullination via Biotin Thiol Tag-Assisted Mass Spectrometry. *Anal Chem* **2022,** *94* (51), 17895-17903.
8. Li, Z.; Wang, B.; Yu, Q.; Shi, Y.; Li, L., 12-Plex DiLeu Isobaric Labeling Enabled High-Throughput Investigation of Citrullination Alterations in the DNA Damage Response. *Anal Chem* **2022,** *94* (7), 3074-3081.
9. Wang, B.; Li, Z.; Shi, Y.; Zhu, Z.; Fields, L.; Shelef, M. A.; Li, L., Mass Spectrometry-Based Precise Identification of Citrullinated Histones via Limited Digestion and Biotin Derivative Tag Enrichment. *Anal Chem* **2024,** *96* (6), 2309-2317.
10. Searle, B. C.; Pino, L. K.; Egertson, J. D.; Ting, Y. S.; Lawrence, R. T.; MacLean, B. X.; Villén, J.; MacCoss, M. J., Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nature Communications* **2018,** *9*, 5128.
11. Speers, A. E.; Cravatt, B. F., Profiling enzyme activities in vivo using click chemistry methods. *Chemistry & Biology* **2004,** *11* (4), 535-546.
12. Lin, Z.; Wang, X.; Bustin, K. A.; Shishikura, K.; McKnight, N. R.; He, L.; Suciu, R. M.; Hu, K.; Han, X.; Ahmadi, M.; Olson, E. J.; Parsons, W. H.; Matthews, M. L., Activity-Based Hydrazine Probes for Protein Profiling of Electrophilic Functionality in Therapeutic Targets. *ACS Central Science* **2021,** *7* (9), 1524-1534.

13. Setner, B.; Stefanowicz, P.; Szewczuk, Z., Quaternary ammonium isobaric tag for a relative and absolute quantification of peptides. *Journal of Mass Spectrometry* **2018,** *53* (2), 115-123.
14. Bicker, K. L.; Subramanian, V.; Chumanevich, A. A.; Hofseth, L. J.; Thompson, P. R., Seeing Citrulline: Development of a Phenylglyoxal-Based Probe To Visualize Protein Citrullination. *Journal of the American Chemical Society* **2012,** *134* (41), 17015-17018.
15. Lewallen, D. M.; Bicker, K. L.; Subramanian, V.; Clancy, K. W.; Slade, D. J.; Martell, J.; Dreyton, C. J.; Sokolove, J.; Weerapana, E.; Thompson, P. R., Chemical Proteomic Platform To Identify Citrullinated Proteins. *ACS Chemical Biology* **2015,** *10* (11), 2520-2528.
16. Holm, A.; Rise, F.; Sessler, N.; Sollid, L. M.; Undheim, K.; Fleckenstein, B., Specific modification of peptide-bound citrulline residues. *Analytical Biochemistry* **2006,** *352* (1), 68-76.

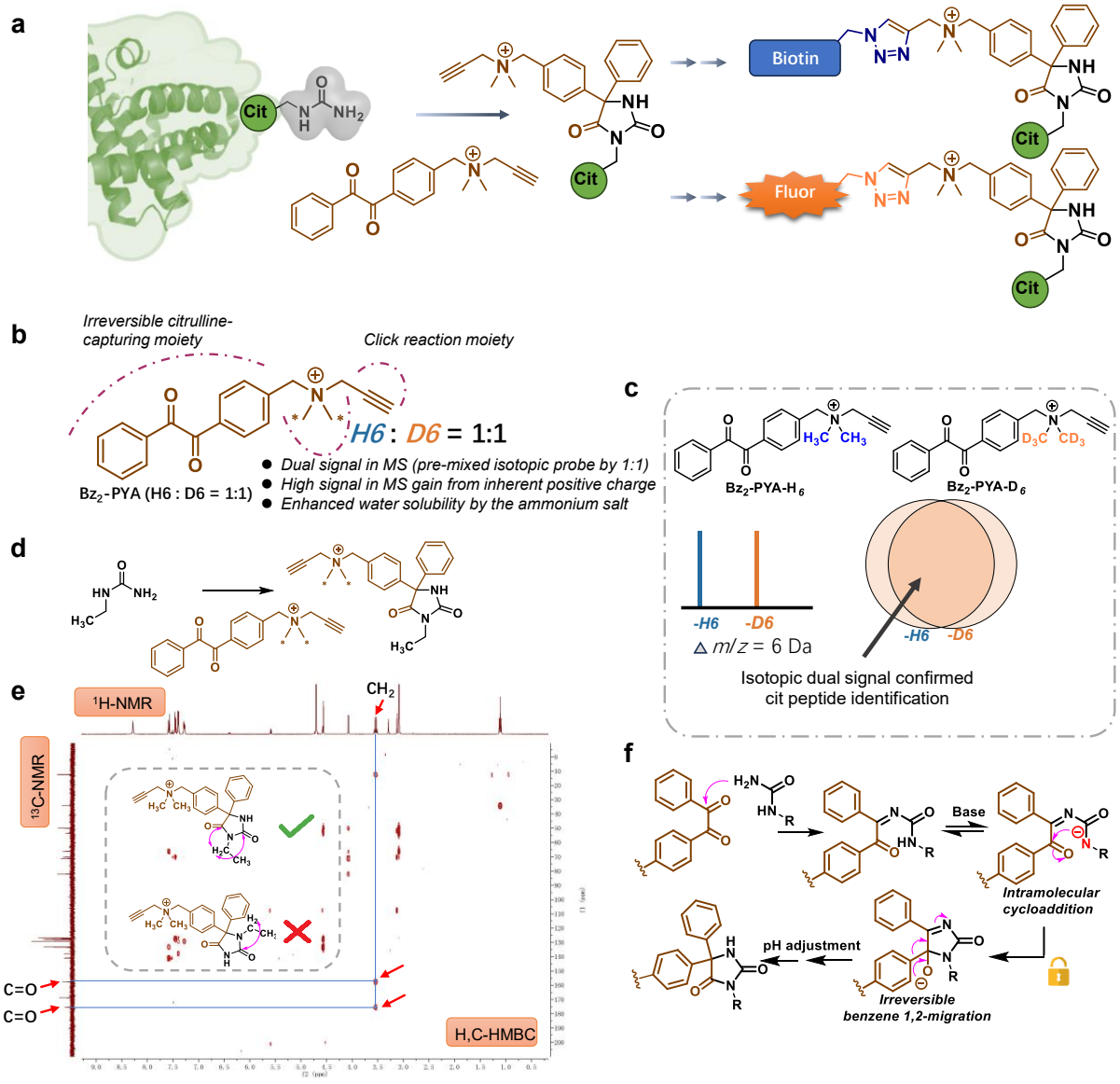


Figure 1. (a) Chemical reaction of probe Bz_2-PYA with citrullination residue, and further biotinylation for enrichment or fluorescence labeling. (b) Design features of the probe Bz_2-PYA for specific labeling of citrullination. (c) Isotopic dual-signal confirmation of citrullinated peptides with the pre-mixed Bz_2-PYA (-H_6: -D_6 = 1:1). (d) and (e) Structural confirmation of labeling reaction by model reaction of *N*-ethyl urea with Bz_2-PYA, and ^{1}H-NMR, ^{13}C-NMR and H,C-HMBC spectrum. (f) The mechanism of irreversible bonding between the benzil group and the ureido group to form a five-membered cyclic compound by benzene 1,2-migration.

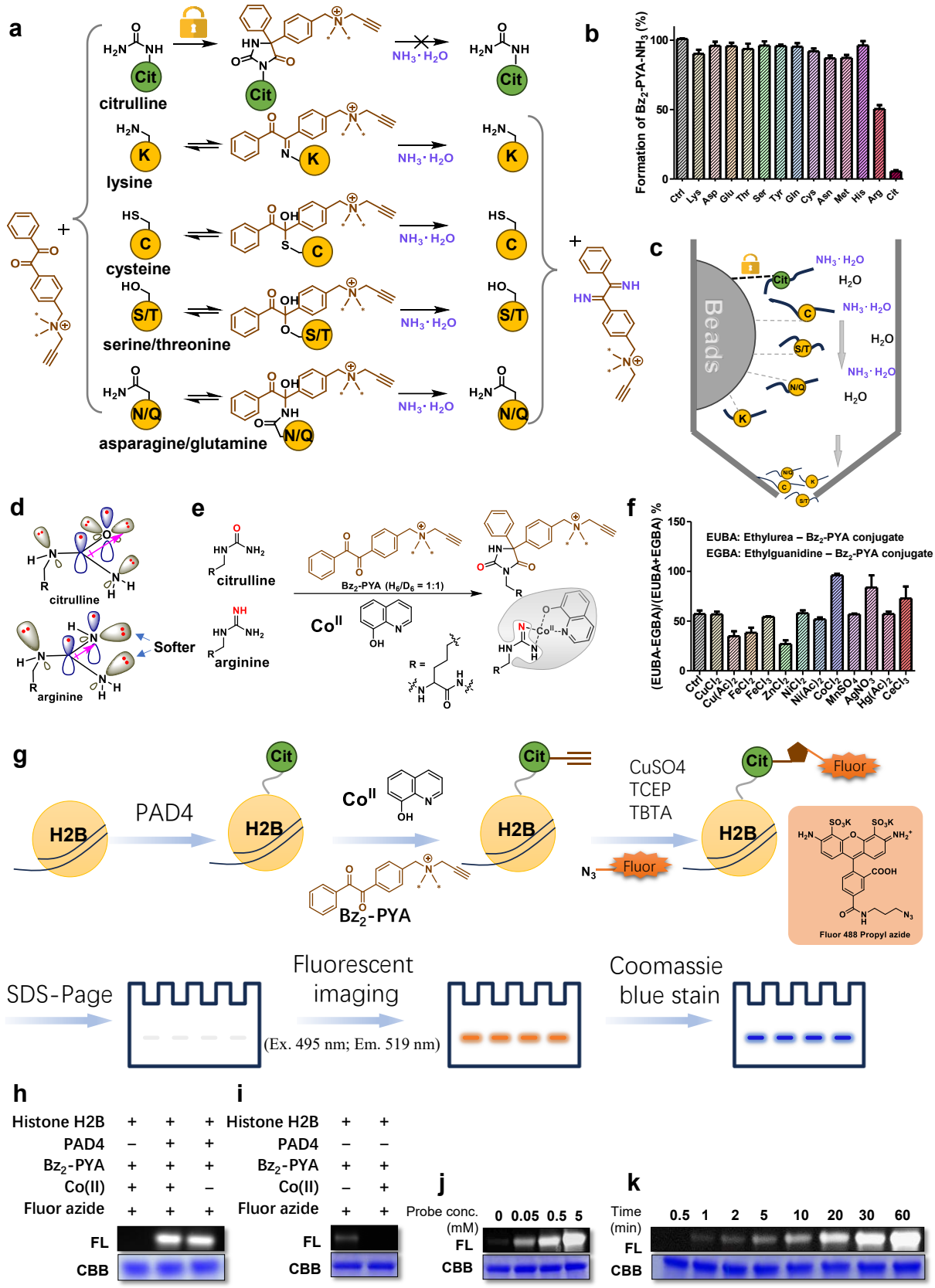


Figure 2. (a) Schematic diagram of probe Bz_2-PYA irreversible covalent bonding with citrulline residues and reversible bonding with other nucleophilic amino acids, which can be elution with

ammonia water. (b) Results of Bz_2-PYA-NH_3 formation in model reaction of probe Bz_2-PYA with various amino acids, followed by washing with excess ammonia water. (c) Schematic diagram of citrullinated peptides enrichment by streptavidin, and ammonia water elution to remove reversible bondings. (d) Molecular orbital diagram of citrulline and arginine residues. (e) Schematic diagram of citrulline selective bonding with Bz_2-PYA with cobalt(II) ion compete blocking with arginine. (f) Screening result of blocking effectiveness of common transition metal ion on *N*-ethylurea and *N*-ethylguanidine. (g) Workflow for the protein level validation experiment of human histone H2B citrullination selective labeling by probe Bz_2-PYA. Human histone H2B was citrullinated by PAD4 enzyme, following by probe Bz2-PYA labeling, fluorescent labeling by click reaction, SDS-page, fluorescent imaging, and Coomassie blue stain and imaging. (h) and (i) Fluorescent and Coomassie blue imaging results of histone H2B whether or not citrullinated by PAD4 enzyme or cobalt(II) ion pre-incubation. (j) and (k) Probe concentration and incubation time optimization by histone H2B.

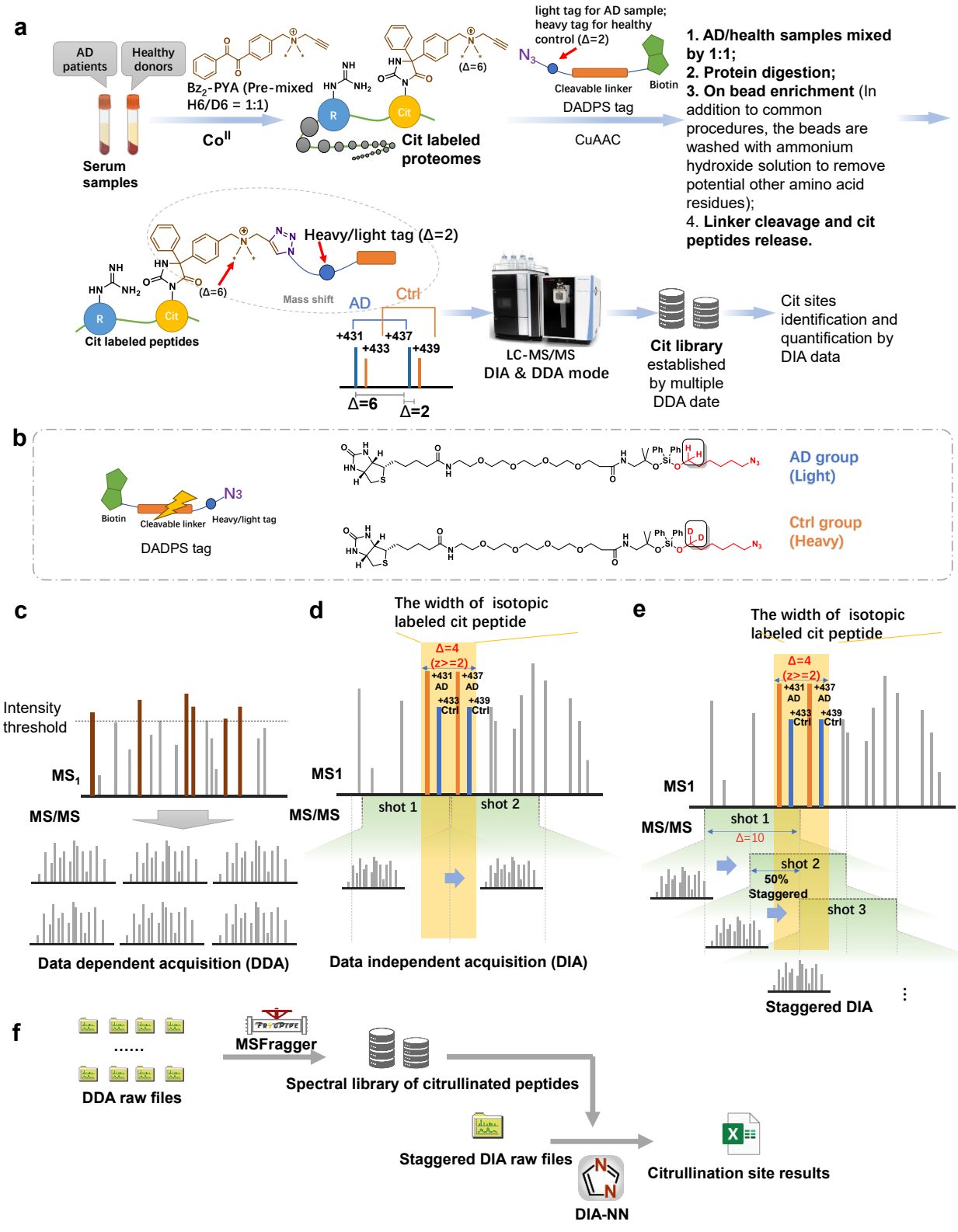


Figure 3. (a) Workflow of triple safeguard-enhanced citrullination mapping on AD/control serum. Briefly, human serum samples (n =4, AD/control) were treated with isotopic pre-mixed Bz_2-PYA (H_6 and D_6) in the optimized condition for citrulline selective and irreversible labeling, following by

biotinylation with DADPS (H_2 for AD and D_2 for control). AD and control sample were mixed into 1:1 ratio, then subject to trypsin digestion, on bead enrichment and elute with ammonium water to remove reversible amino acids, and cleavage of the linker for releasing enriched citrullinated peptides with mass increment of 431.2209 and 437.2675 for AD, and 433.2365 and 439.2835 for control. The samples were analyzed by Orbitrap Exploris 480 Mass Spectrometer in DDA mode and DIA mode. (b) Structure of light and heavy DADPS tag. (c) Schematic diagram shows data dependent acquisition (DDA) analysis. (d) Schematic diagram shows normal data independent acquisition (DIA) mode analysis will possible cut the isotopic labeled citrullinated peptides into different fragmentation window. (e) Schematic diagram shows the staggered DIA mode analysis to achieve covering a group of isotopic labeled citrullinated peptide in one fragmentation window, with shot window width of 10 *m*/*z* and 50% staggering. (f) Workflow of DDA spectral library based citrullination mapping in staggered DIA mode. A citrullination library with citrullinated peptide fragmentation ion information has been built by 46 DDA raw files, the spectral library was employed on the analysis of staggered DIA raw data for the purpose of deconvolution.

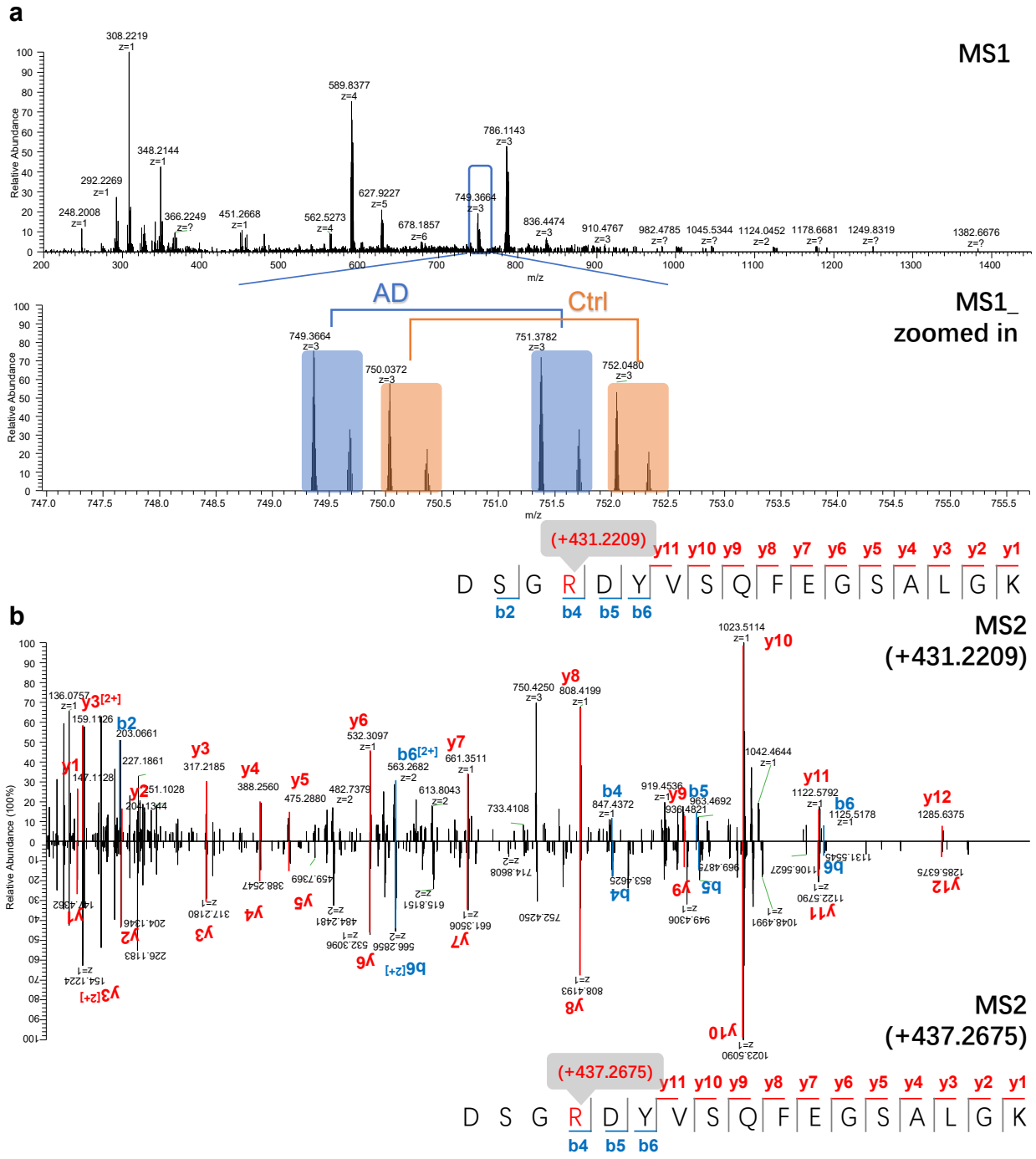


Figure 4. (a) An example of SM1 spectrum with zooming in of a group of isotopic labeled citrullinated peptide. (b) MS2 spectra of AD in SM1 (749.3664 and 751.3782) identified the peptide DSGRDYVSQFGSALGK with isotopic double confirmation.

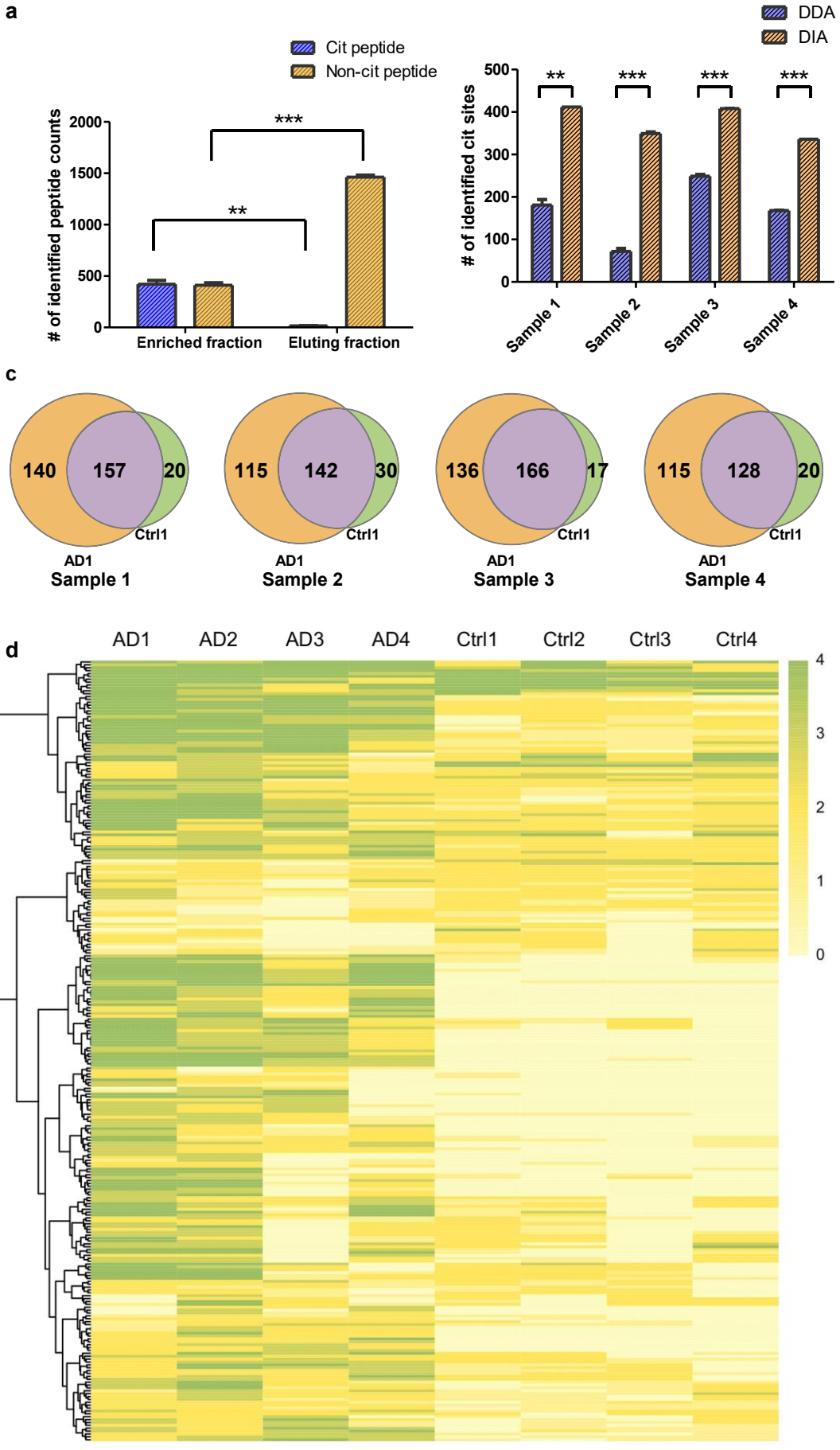

a
Cit peptide
Non-cit peptide
# of identified peptide counts
2000
1500
1000
500
0
***
**
Enriched fraction
Eluting fraction
DDA
DIA
# of identified cit sites
500
400
300
200
100
0
**
***
***
***
Sample 1
Sample 2
Sample 3
Sample 4
c
140
157
20
Ctrl1
AD1
Sample 1
115
142
30
Ctrl1
AD1
Sample 2
136
166
17
Ctrl1
AD1
Sample 3
115
128
20
Ctrl1
AD1
Sample 4
d
AD1
AD2
AD3
AD4
Ctrl1
Ctrl2
Ctrl3
Ctrl4
4
3
2
1
0

Figure 5. (a) The number of identified peptide counts of citrullinated peptide and non-citrullinated peptide from streptavidin enriched fraction and eluting fraction. (b) The number of identified citrullination sites from same four samples in DDA detection mode and staggered DIA mode. (c) Venn plot of identified citrullination sites from duplicated four sample of AD and control. (d) Heatmap of identified citrullination sites amount four AD and control.

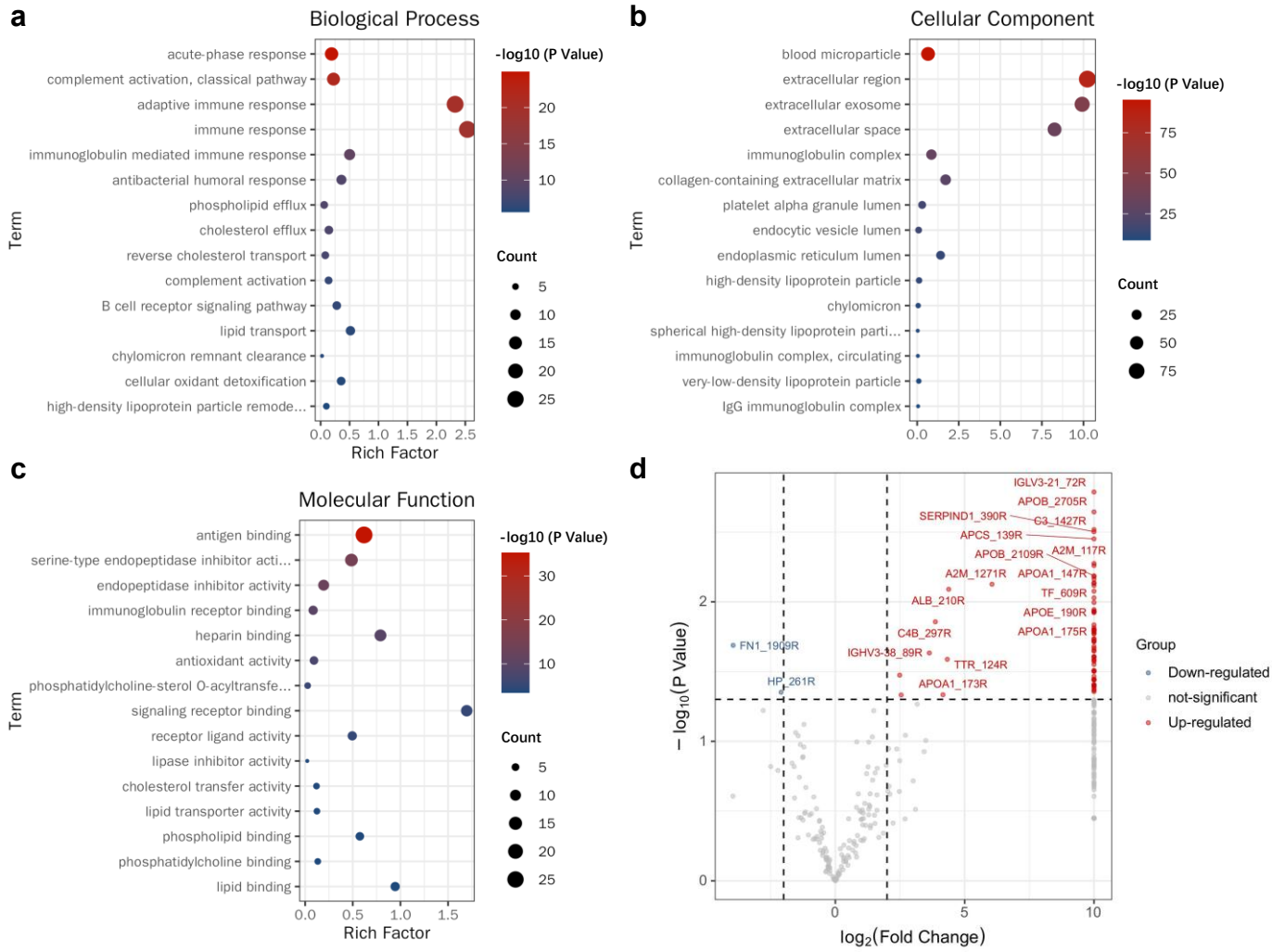


Figure 6. (a), (b), and (c) Bubble plot of the top 15 gene oncology enrichment analysis of citrullinated proteins involved in biological processes, cellular components, and molecular functions. (d) Volcano plot of citrullinated proteins and arginine sites. Citrullination sites found only in the AD sample and not in the control sample were set to a $\log_2$(fold change) value of 10.

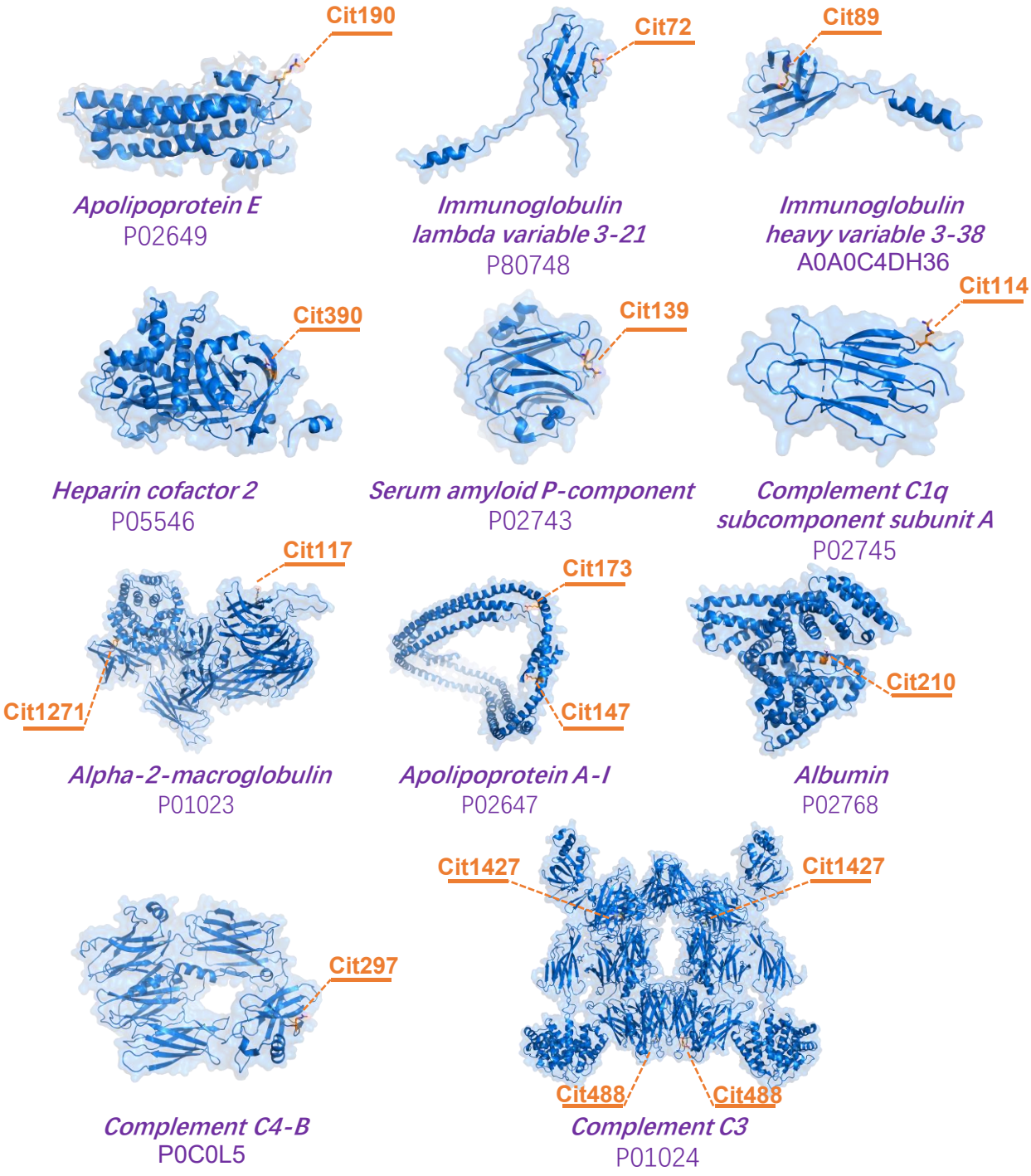


Figure 7. (a) Significantly upregulated citrullinated proteins and citrullination sites in AD compared to controls.

Chapter 4

Mass spectrometry structural proteomics enabled by limited proteolysis and cross-linking

Adapted from Lu, H.; **Zhu, Z**.; Fields, L.; Zhang, H.; Li, L. Mass spectrometry structural proteomics enabled by limited proteolysis and Cross-Linking. *Mass Spectrometry Reviews* **2024**, 43 (5), e22134. DOI: https://doi.org/10.1002/mas.21908. Zhu, Z. contributed to manuscript preparation in all crosslinking-related topics and representative application sections.

1. INTRODUCTION

Three-dimensional (3D) structural understanding is key to elucidating the biological function of a protein (Holm 2022), as it intricately regulates enzymatic activity in response to protein modifications, ligand binding, and protein-protein interactions (Na et al., 2020). Thus, the exploration of protein structure and function stands at the forefront of life science and represents an ever-expanding focus in the advancement of proteomics. Traditional methods used for the protein structure characterization at atomic resolution include X-ray crystallography (XRD), nuclear magnetic resonance spectroscopy (NMR), and cryo-electron microscopy (cryo-EM). However, the application of these methods for large-scale protein and protein complex characterization is limited, owing to stringent prerequisites on the purity, size, and amount of protein sample, and crystallization propensity (Kaur et al., 2019). Complementary to the aforementioned traditional methods, mass spectrometry (MS) enables the study of a wider range of molecular size, analysis of heterogeneous proteins (including post-translational modifications (PTMs) and varying conformers), and high-throughput identification and quantification of entire proteome in cells, tissues, and even organisms (Blackburn et al., 2022; Kaur et al., 2019).

MS-based structural proteomics in recent years has become a powerful method in the field of structural biology, making the characterization of protein structure at both the protein level and peptide level possible. MS-based structural proteomics methods can be categorized into protein-centric and peptide-centric strategies based on whether they focus on capturing protein or peptide information to acquire structural information (Na et al., 2020). Within

protein-centric strategies, native MS and ion mobility-mass spectrometry (IM-MS) are two common methods, which have shown remarkable advantages in studying protein structure, function, and assembly dynamics (Li et al., 2017; McCabe et al., 2021; Zimnicka 2023). For example, native MS can provide direct analysis of intact proteins and protein complexes in the gas phase while maintaining noncovalent interactions between proteins and their ligands (Barth et al., 2020), while IM-MS can discern protein structure and stability information based on mass, charge, and collision cross section (Christofi et al., 2023). Peptide-centric strategies are the gold standard for proteomics workflows, in which various forms of solution-phase chemistry driven by chemical or enzymatic reaction are used to introduce modifications to intact proteins and localize these modifications by MS analysis at the peptide level for the exploration of protein structure and interaction (Leitner 2016). The widely used peptide-centric methods for protein structure study include hydrogen/deuterium exchange (HDX) (Engen 2009; Masson et al., 2019), covalent labeling (CL) (Pan et al., 2022), thermal proteome profiling (TPP) (Franken et al., 2015), limited-proteolysis (LiP) (Schopper et al., 2017), and cross-linking (XL) (Yu et al., 2018) MS methods. HDX measures the exchange of labile amide hydrogens along the protein backbone with the heavier deuterium isotope, as protein conformation can affect the exchange rate of hydrogen and deuterium, monitoring the deuteration of proteins can reveal the aspects of conformation as well as changes of conformation perturbed by factors such as pH, PTMs, and protein interactions (Engen 2009; Trabjerg et al., 2018). CL is conceptually similar to HDX, in which irreversible modifications to reactive side chains are introduced to monitor the solvent

accessibility of side chains in proteins or protein complexes, wherein protein structure and interactions can be mapped by the differential reactivity of amino acid side chains (Bamberger et al., 2021; Kiselar et al., 2018; Mendoza et al., 2009). Different from HDX and CL, TPP is based on the principle that proteins denture and become insoluble when subjected to heat, which has been successfully applied in identification of drug targets (Mateus et al., 2022), monitoring ligand interactions with cellular membrane proteins (Reinhard et al., 2015), and studying protein-protein interactions (PPIs) (Mateus et al., 2020). TABLE 1 summarizes the advantages and limitations of the main peptide-centric structural MS proteomics methods.

The above peptide-centric methods provide invaluable and unique insights into protein conformation studies, making an immense contribution to the development of MS-based structural proteomics. A detailed introduction and representative applications of all peptide-centric methods is outside the scope of this review, and can be found in previous reviews (Engen 2009; Franken et al., 2015; Masson et al., 2019; Mateus et al., 2020; Pan et al., 2022; Trabjerg et al., 2018). Herein, our focus will be directed towards reviewing the methodologies of limited proteolysis-mass spectrometry (LiP-MS) and cross-linking mass spectrometry (XL-MS) and their role in characterizing protein structure, dynamics, and interactions. LiP-MS allows the detection of protein structural alterations in complex backgrounds and on a proteome-wide scale, while XL-MS provides crucial insights into PPIs and spatial proximity. Although HDX, CL, and TPP are powerful techniques for studying protein structure and function, LiP-MS and XL-MS offer several distinct advantages not

available with the other structural MS tools. For instance, HDX reveals solvent accessibility dynamics but does not directly probe conformational dynamics like LiP-MS. Similarly, CL and TPP are valuable for mapping specific protein modifications or interactions, but they do not provide detailed spatial information on protein structure in comparison to XL-MS. We will review (1) the principles, features, limitations, and considerations in sample preparation; (2) bioinformatics strategies and software tools used for interpreting data; (3) noteworthy applications in neurodegenerative diseases, interactome studies, membrane protein characterization, and artificial intelligence-based structural analysis; (4) factors that modulate protein conformational changes; and (5) challenges in understanding the intricacies of protein conformation changes through LiP-MS and XL-MS methods.

TABLE 1. The advantages and limitations of the main peptide-centric structural MS proteomics methods.

Methods	Primary Advantages	Main Limitations	Commonalities	Reference
HDX	Analyzing the conformation, dynamics and interactions of protein complex	A high or variable back-exchange severely limits sensitivity, reliability and the amount of information obtained		(Trabjerg et al., 2018)
CL	Irreversibility Covalent labeling of multiple amino acids	Labeling reagents might disturb protein structure and activity;		(Leitner 2016; Mendoza et al., 2009)
TPP	Measures protein precipitation due to heat-induced unfolding	Cannot provide specific structural change information	Requires careful control of experimental conditions to ensure accurate measurements; Compatible for validation and integration with other techniques	(Malinovska et al., 2023; Reinhard et al., 2015)
LiP	Applicable to the targeted or unbiased study; Detects proteome-wide structural changes; Enables comparative analysis of structural proteome profiles; Sensitivity to both pronounced and subtle structural transitions of proteins	To detect a structural change, the relevant protein region must be detected; Because multiple types of molecular events can be detected, mechanistic interpretation of a change typically requires orthogonal data		(Feng et al., 2014; Malinovska et al., 2023)
XL	Accurate identification of PPIs and neighboring residues; High toleration to various buffer conditions; Applicable to high complexity samples	Crosslinking solvent DMSO might affect the protein structure; Cell permeability for *in vivo* testing need to be approved		(Mendoza et al., 2009; O'Reilly et al., 2018; Yu et al., 2018)

2. METHODS: PRINCIPLES AND WORKFLOWS

2.1 Limited Proteolysis-Mass Spectrometry (LiP-MS)

2.1.1 Workflow

LiP-MS is a proteomics technique used to identify protein structural changes by pinpointing protein regions undergoing structural transitions at the peptide level. As illustrated in **Figure 1**, the workflow can be briefly described as follows: proteins are extracted from the sample of interest, such as cells, tissue, and biofluid, under native conditions and are evenly divided into LiP and trypsin (Try) samples. LiP samples are first subjected to limited enzymatic digestion by broad-specificity proteases (e.g., protease K and thermolysin) under native conditions. Subsequently, Try and LiP samples are denatured, followed by complete digestion to generate a peptide mixture amenable to bottom-up MS proteomics. Structurally informative proteolytic fragments are then used to identify the specific protein regions undergoing conformation changes (Schopper et al., 2017). Desalted peptides can be directly injected into liquid chromatography-mass spectrometry (LC-MS) for label-free quantification (LFQ) analysis under either data-dependent acquisition (DDA) or data independent acquisition (DIA) strategies (**Figure 1A**). To reduce systematic error from run-to-run variation and improve analytical throughput in a single LC-MS run, the LiP-MS method can be combined with isobaric labeling (**Figure 1B**). Recently, a workflow combining *N, N*-dimethyl leucine (DiLeu) isobaric labeling with LiP-MS (DiLeu-LiP-MS) was reported. The results indicated that the established DiLeu-LiP-MS method can be employed for high-throughput quantitation of structural protein changes in complex human serum samples (Lu et al., 2023a). In addition to using isobaric tags labeling to enhance the throughput in a single LC-MS run, some improvements can also be adopted during sample

preparation to substantially increase the number of samples that can be processed (Malinovska et al., 2023). Alternatively, off-line high-pH (HpH) reversed-phase fractionation can be conducted prior to LC-MS analysis to improve proteome coverage and enhance the detection of low-abundance proteins using either LFQ or isobaric labeling approaches. There are several factors that must be considered during sample preparation, which enable sensitive and accurate quantitation of LiP-MS results. It is worth noting that if larger size samples are divided into several batches using isobaric tags labeling strategy, the use of reference channels is highly recommended for batch normalization. Also, in LiP experiments, a tryptic control may be unnecessary if protein abundance changes are not expected, which is quite often the case for typical applications. Digestion time could affect the structural information obtained. Prolonged digestion time may not be beneficial, as it can lead to the generation of secondary cleavage products (Schopper et al., 2017; Malinovska et al., 2023). Therefore, systematic evaluation of digestion time variability and coverage is necessary. In addition to digestion time, optimization of detergents, digestion buffer, enzyme selection, and enzyme quantity can achieve an equitable approach for the proteolysis of easily digestible complexes with their more resistant counterparts in LiP-MS experiments, thereby enhancing the depth and accuracy of proteome-wide structural changes analysis. Two separate analyses need to be performed for both LFQ experiments and isobaric labeling data alike. Specifically, the Try sample is searched in a standard proteomics pipeline to obtain abundance changes for peptide-level LiP data normalization, while the LiP sample is searched by treating every peptide as an independent entity to obtain peptide-level abundance changes. A volcano plot provides a visualization tool for the results. Proteins associated with peptides exhibiting significant changes in abundance within the LiP samples can be regarded as structurally variant. These peptides with significant changes are subsequently

used for the identification of specific protein regions undergoing structural alterations (**Figure 1C**). In our workflow, unless explicitly noted otherwise, all parts are considered typical or standard practice. High-pH (HpH) fractionation can reduce sample complexity, thereby enhancing proteome coverage and improving the detection of low-abundance proteins. However, HpH fractionation may require more samples and longer instrument time. As isobaric tags labeling strategies can greatly reduce instrument time, HpH fractionation can be used in tandem with isobaric labeling to improve proteome coverage depth, while maintaining a reduced overall instrumentation time. Nevertheless, if instrument time is not a limiting factor, HpH fractionation remains applicable in LFQ. Therefore, HpH fractionation is classified as optional in our workflow.

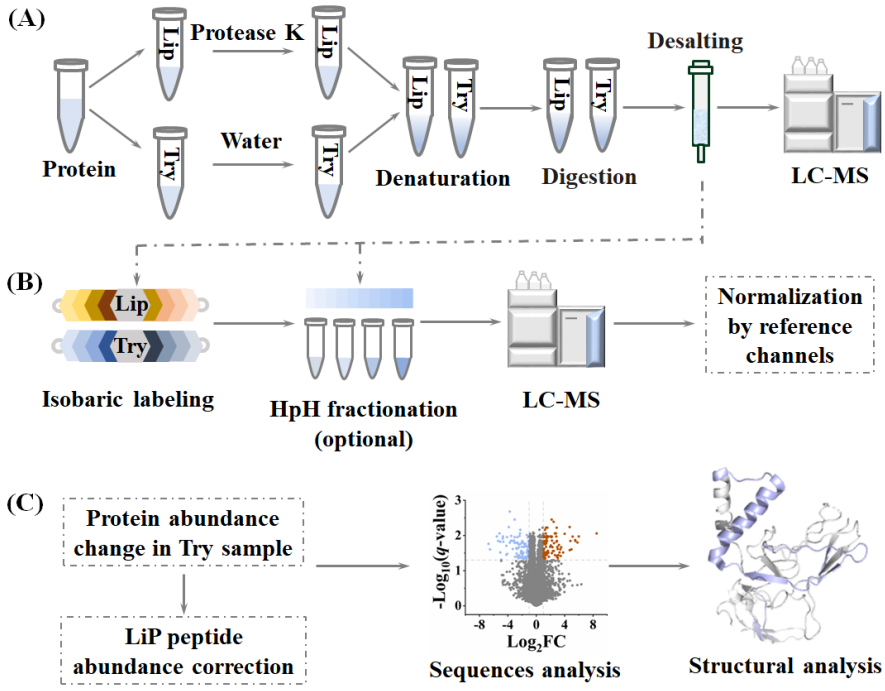


FIGURE 1. Schematic of LiP-MS workflow. (A) Preparation of limited proteolysis (LiP) and trypsin (Try) samples for LFQ analysis. Each proteome extract is divided into two samples: a Try sample and a LiP sample. The Try sample undergoes only trypsin digestion under denaturing

conditions, while the LiP sample first is exposed to limited enzymatic digestion by broad-specificity proteases (such as Protease K) under native conditions followed by complete trypsin digestion under denaturing conditions. (B) Isobaric labeling of desalted peptides accompanied by high-pH (HpH) fractionation (optional step). If larger size samples are divided into several batches when using the isobaric labeling strategy, the use of reference channels is highly recommended for batch normalization. (C) Data analysis for both LFQ and isobaric-labeled samples. Protein abundance changes are utilized as normalization factors for LiP peptide abundance correction. A volcano plot is used to visualize normalized peptide-level abundance changes. Proteins associated with significantly changed peptides in the LiP sample are deemed structurally variant, and these peptides are used to pinpoint the specific protein region undergoing structural changes.

Overall, the LiP-MS protocol has been evolving since its invention toward improved sensitivity and robustness, aiming to minimize experimental variability and maximize results reproducibility. This evolution includes updates on sample preparation protocols, which incorporates detailed optimizations of key parameters, including the enzyme/substrate (E:S) ratio and the incubation time of the LiP step (Wang et al., 2023). To deconvolute the fragment ion spectra, spectral libraries are employed in post-acquisition in silico data processing. While public repositories suffice for classical proteomics experiments, the presence of half-tryptic peptides in LiP-MS requires a study-specific library. This can be achieved by employing a DDA MS method on the same set of samples to generate an extensive spectral library for confident identifications and quantifications of DIA MS data. However, this strategy may lead to longer acquisition times when all samples are measured. To mitigate this, pooling samples across replicates can be used for library generation. Furthermore, MSstatsLiP, a comprehensive analysis package tailored for

integrated quality control, data processing, statistical analysis, and visualization of LiP-MS data, has been reported (Malinovska et al., 2023).

2.1.2 Advantages and limitations

LiP-MS has been successfully applied to global analysis of protein structural changes in complex samples, the study of protein-metabolite interactions (Piazza et al., 2018), and biomarker studies for neurodegenerative diseases (Lu et al., 2023a; Mackmull et al., 2022; Wang et al., 2023). These studies demonstrated the advantages of the LiP-MS methods, including (1) its applicability to various complex biological specimens, such as cells, serum and cerebrospinal fluid (CSF); (2) compatibility with both targeted MS detection methods (e.g., selected reaction monitoring [SRM] for one or several proteins of interest) and discovery-driven shotgun proteomics within their biological matrix; and (3) sensitivity to both pronounced and subtle structural transitions of proteins. Like any method, LiP-MS has some limitations: (1) the efficiency of proteolysis is sensitive to experimental conditions; (2) incomplete digestion from enzymatic cleavage can lead to missing potential structurally related information; (3) the sequencing and identification of half-tryptic peptides can be challenging; (4) difficulty in detection of low-abundance and insoluble of proteins as well as membrane proteins; and (5) lack of information on the nature of protein structural changes and artifactual structural alterations derived from protein extraction step (Schopper et al., 2017). Despite these limitations, LiP-MS remains a valuable tool for globally probing protein structure and dynamics, enabling researchers to gain valuable insights into the structural biology of proteins.

2.1.3 Sample preparation considerations

To date, a variety of protein sample sources have been used for LiP-MS analysis, including pure protein (Lu et al., 2023a), serum (Lu et al., 2023a), mammalian cells (Piazza et al., 2018), CSF (Mackmull et al., 2022; Shuken et al., 2022; Wang et al., 2023), and yeast lysates. In LiP-MS, mitigating the effect of confounding factors on variability and reproducibility demands meticulous attention to sample preparation. Different raw materials contain varying amounts and types of endogenous proteases. During protein extraction, endogenous proteases can lead to unwanted proteolysis. Therefore, protein extraction should be conducted at 4 °C and certain protease inhibitors (e.g., EDTA-free Roche cOmplete Protease Inhibitor Cocktail) can be added to the protein extraction buffer. Choice of a suitable buffer is paramount, ensuring compatibility with the protease used for limited proteolysis. Additionally, the choice of protease should be guided by the specific cleavage sites desirable to be studied. Protease K is commonly used, but other proteases (e.g., Thermolysin, Subtilisin, and Chymotrypsin) can be used and may be more suitable depending on the protein of interest (Schopper et al., 2017). Conducting pilot experiments to optimize enzyme-to-substrate ratio, incubation time, and temperature is crucial. Incubation time should be precisely controlled after protease addition for improved reproducibility. Ensuring the purity of protein sample is also important, since contaminants can interfere with the proteolysis process and affect the accuracy of the MS results. Taken together, these best practices must be followed in sample preparation to assure high-quality data acquisition to produce meaningful insights from the LiP-MS study.

2.2 Cross-Linking Mass Spectrometry (XL-MS)

Like LiP-MS, XL-MS has become an extensively used technology in MS structural proteomics studies. To cater to the demands of systems-wide studies on protein interactions within intricate

systems, several XL-MS workflows have been established and refined (Gotze et al., 2019; Iacobucci et al., 2018; Mendes et al., 2019). These workflows aim to increase the sensitivity, accuracy, and throughput of XL-MS experiments, enabling thorough analyses of PPIs across diverse biological systems. This section will present several workflows, including general XL-MS analysis, quantitative XL-MS, in vivo and in situ XL-MS, and the integration of trapped ion mobility spectrometry (TIMS) with XL-MS strategy.

2.2.1 General XL-MS analysis workflow

XL-MS has undergone rapid development in recent decades, emerging as a robust strategy for structural investigations. The fundamental principle of the workflow involves a chemical crosslinker which targets proximal residues and forms covalent bonds between them (Mendes et al., 2019; O'Reilly et al., 2018; Sinz 2003). After MS data acquisition and corresponding analysis, these conjugations provide critical insights into the 3D-strucuture within the specific proteins or the PPIs within the networks. The general workflow of XL-MS is shown in **Figure 2**. Crosslinkers typically consist of two reactive groups at the ends and one spacer linking in the center. Currently, the majority of the crosslinkers used are amine-reactive, comprising a N-hydroxysuccinimide (NHS) ester as a reactive group, targeting lysine residues only (Jin Lee 2008; Leitner et al., 2010). This prevalent crosslinker type will be the focal point of our discussion.

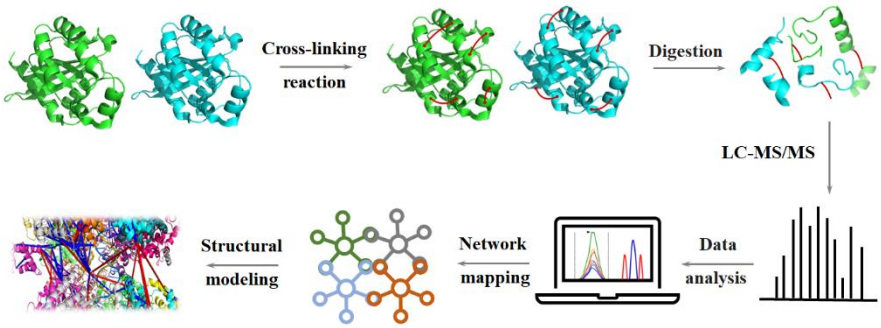


FIGURE 2. General workflow of proteome-wide XL-MS. The crosslinkers will be incubated with the samples in prior to the protein digestion by trypsin or Lys-C. After proper cleaning, the samples are analyzed through LC-MS/MS. The collected data will be analyzed and quantitated through various software tools.

As early as in 1979, ethylene glycolyl bis(succinimidyl succinate) was developed (Abdella et al., 1979), laying the groundwork for non-cleavable crosslinkers such as disuccinimidylsuberate (DSS) and bis(sulfosuccinimidyl)suberate (BS3), effective for the analysis of small assemblies (Piersimoni et al., 2022b). As a result of these crosslinking experiments, three types of crosslinks were identified. The interpeptide crosslinks (type 2) indicate that the target amino acid residues are in two different peptide sequences or two different proteins. The intrapeptide or "loop-linked" crosslinks (type 1) are created between proximal residues on the same peptide. Lastly, the "dead-end" crosslinks (type 0), also known as mono-links, occur when only one end of the crosslinker attaches to a peptide. In most experimental designs, type 2 interlinks are the most informative, providing direct evidence for PPIs and structural variations. While type 1 links can still provide useful perspective in specific protein inner structure, type 0 monolinks are usually discarded and some research groups have even developed corresponding strategies to filter them out (Steigenberger et al., 2020). Considering the field's shift towards conducting proteome wide

XL-MS studies, MS-cleavable cross-linkers become increasingly pivotal. The cleavable crosslinker usually contains a spacer that can be cleaved at specific site under MS collisional activation condition while generating characteristic fragment ions (Piersimoni et al., 2020). The advent of the cleavable crosslinker disuccinimidyl sulfoxide (DSSO) in 2011 further broadened the scope for investigating a more diverse range of protein complexes (Kao et al., 2011). DSSO contains two symmetric collision-induced dissociation (CID)-cleavage sites, enabling these peptides to be discriminated on the basis of their characteristic fragment patterns under MS/MS acquisition which are unique to cross-linking types (e.g. interlink, intralink, and monolink) (Kao et al., 2011; Sinz 2017). These diagnostic peaks facilitate faster and more reliable searching for the crosslinked peptides using contemporary analytical software tools (Gotze et al., 2015; Iacobucci et al., 2018).

2.2.2 Quantitative XL-MS strategy

Traditional XL-MS techniques have primarily focused on identifying crosslinked peptides and elucidating PPIs or protein structures. However, quantitative information is essential to distinguish major conformational changes across complex sample systems and multiple replicates, necessary to address the challenges imposed by the varied protein abundance levels across different protein types. Driven by advancements in quantitative proteomics, quantitative XL-MS(QXL-MS) has emerged as an unbiased way to assess the structural dynamics of protein complexes and quantitatively elucidate protein conformational changes under various conditions. The QXL-MS approach encompasses three principal strategies for crosslinking quantification, including LFQ, isotopic-labelling and isobaric labelling.

In LFQ, DDA MS is utilized in most cases, entailing quantification of the extracted ion intensities based on full scan (MS1) signals (Chen et al., 2019b). The crosslinked peptides are separated through the RPLC elution, and the extracted ion chromatograms are analyzed through the area under curve reflecting changes in signal abundances. The automated quantification may be processed by software including MaxQuant (Cox et al., 2014; Yılmaz et al., 2022). The advantage of LFQ lies in its MS1 signal-based strategy, which has almost zero exclusivity to specific types of crosslinker, nor to the strategies for the identification of crosslinked peptides (Chen et al., 2018). Meanwhile, this method also grapples with challenges related to diminished reproducibility between multiple replicates, an artifact of run-to-run variation, and the restricted abundance of crosslinked peptides, resulting in impeded efficacy of LFQ, particularly in scenarios involving complex mixtures (Muller et al., 2018). Some have addressed this through adoption of alternative strategies such as targeted methodologies, including parallel reaction monitoring (PRM) to facilitate LFQ analysis of these crosslink peptides with enhanced sensitivity and reliability (Yu et al., 2022). LFQ is one of the most widely used quantitative strategy in XL searching for its easy-handling and low cost in comparison to the other labeling methods involving expensive labeling reagents (Chen et al., 2018; Muller et al., 2018).

In contrast to LFQ, an isotopic labelling strategy can be employed for pairwise comparison of protein structural states. As shown in **Figure 3A**, isotopic labelling involves the use of “light” and “heavy” analogs of the same crosslinker in a 1:1 ratio, labeling proteins separately in two different conditions (Barth et al., 2021; Petrotchenko et al., 2005; Schmidt et al., 2014; Walzthoeni et al., 2015). For example, the BS3-d4 substitutes 4 deuterium atoms in the spacer of BS3, resulting in a 4 Da mass shift that can be identified by highlighting this MS1 doublet mass difference (Fischer et al., 2013). To ensure accurate quantitation, a “reverse labeled” set usually

is prepared, swapping the heavy analog for parallel analysis (Chen et al., 2019b). Furthermore, specialized searching software such as xiSearch, has been developed to utilize MS/MS feature detection of peptide pairs (Mendes et al., 2019). This powerful tool provides additional pairwise crosslink authenticity through MS/MS-based identification of crosslinked peptides. To ensure accurate quantitation, a "reverse labeled" set usually is prepared, swapping the heavy analog for parallel analysis (Chen et al., 2019b). This enables higher authenticity by providing a double confirmation, facilitating the subsequential quantification process automated by Skyline (Pino et al., 2020). Moreover, the incorporation of stable isotope labeling using amino acids in cell culture (SILAC) has enhanced quantitative crosslink analysis (Ong et al., 2002), providing robust quantitative measurement in large-scale interaction network investigations. This quantitative proteomic analysis of the protein interaction network has been proven a robust approach to evaluate the systematic changes in human cells and other complex samples (Chavez et al., 2011), which paves the way for the development of quantitative multiplexed crosslinker (Chavez et al., 2015, Chavez et al., 2020). Despite being a widely employed labeling strategy, the isotopically labeled crosslinking is constrained to samples with relatively low complexity (Kukacka et al., 2015; Zheng et al., 2016).

Conversely, the isobaric labeling strategy for crosslinking quantification frequently relies on multiplex isobaric labeling reagents, such as tandem mass tag (TMT) (Thompson et al., 2003; Yu et al., 2016). Commercially available kits expand the multiplex capacity up to 10-plex or even more, with each plex containing a unique reporter ion which is subjected to MS/MS fragmentation and identification (Isasa et al., 2015; Tarasova et al., 2015). Notably, the Quantitation of Multiplexed, Isobaric-labeled cross (X)-linked (QMIX) peptides strategy employed a duplex TMT reagent with DSSO-crosslinked peptides (**Figure 3B**), coupling isobaric

labeling with a cleavable crosslinker (Yu et al., 2016). Additionally, synchronous precursor selection (SPS)-based MS3 was implemented to increase the accuracy of quantitation. While this QMIX method does not increase the MS1-level sample complexity, it is noteworthy that the accuracy of quantification is contingent upon effective fragmentation. Up to four independent fragmentation spectra can be integrated into one quantification of the identified crosslinked peptide to improve the accuracy of quantitation (Chen et al., 2018). An alternative approach is the isobaric quantitative protein interaction reporter (iqPIR) technology (Chavez et al., 2020). This strategy simultaneously obtains both reporter ions and isotope-labeled crosslinked peptides in MS^2, capitalizing on the higher sensitivity and lower signal suppression from MS^2 compared to MS^1, along with the stable isotope signature fragments. This approach bypasses the issue of low sensitivity resulting from the compression of reporter ion signals at the MS^2 level.

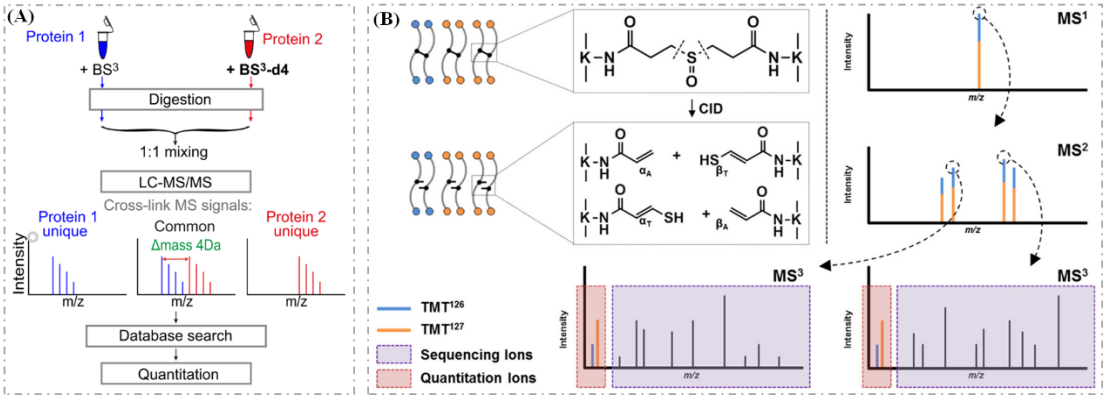


FIGURE 3. Workflow of different QXL-MS methods. (A) Isotopic labelling. Reprinted with permission from Chen. et al. ((Chen et al., 2016), copyright 2016 (Elsevier)). (B) Isobaric TMT tag labelling. Reprinted with permission from Yu. et al. ((Yu et al., 2016), copyright 2016 (American Chemical Society)).

2.2.3 *In vivo* and *in situ* XL-MS approaches

Over the past decade, significant investments have been made in advancing XL-MS studies, expanding its applicability from simple samples (e.g. purified proteins) to the investigation of interactomes and protein conformational changes within living cells. However, this transition from *in vitro* to *in vivo* crosslink analysis under native environment has faced several challenges, including identification of low-abundance crosslinked peptides, encountering a high rate of missed cleavages, and navigating the vast searching space during analysis (Matzinger et al., 2021). To address these challenges, the development of a membrane-permeable and enrichable crosslinker became imperative for system-wide studies within cells, facilitating the extraction of low-abundance crosslinked peptides from a substantial linear, or non-crosslinked, peptide background. A notable breakthrough comes from the selectively enrichable protein interaction reporter (PIR) linker. PIR, with its extended length (43 Å) and a stretched biotin tag in the spacer for streptavidin bead enrichment, represents a significant advancement in enrichable crosslinker products (Tang et al., 2005a). According to its documentation, PIR can traverse the cell membrane via an active transporter, providing over one thousand crosslinked peptide pairs in a reproducible pattern (Chavez et al., 2015). Subsequently, this strategy was further expanded to the exploration of *in vivo* testing in heat shock protein 90 (HSP90) and microtubules in Hela cells, evolving PIR into 6-plex isobaric labeled version (iqPIR), which enables quantitative analysis of six biological samples in a single acquisition run (Chavez et al., 2016; Chavez et al., 2019; Wippel et al., 2022). Another crosslinker fulfilling the requirements of membrane permeability, MS-cleavability, and capable of enrichment is the azide-tagged crosslinker, named azide-tagged, acid-cleavable disuccinimidyl bis-sulfoxide (Azide-A-DSBSO) (Burke et al., 2015; Kaake et al., 2014; Matzinger et al., 2020; Vellucci et al., 2010). Similarly, the Alkyne-A-

DSBSO-based *in vivo* XL-MS platform showed significant efficacy in profiling global PPIs and successfully delineating an intricate mapping of chaperone network in human living cells (Wheat et al., 2021). In recent years, more *in vivo* applications have been inspired by these methodologies, generating new types of crosslinkers like glycosidic-bond-based trehalose disuccinimidyl ester (TDS), with less bulky chemical structures and extraordinary biocompatibility (Chen et al., 2023a). Inspired by the groundbreaking developments in phosphoproteomics, an immobilized metal affinity chromatography (IMAC)-enrichable crosslinker PhoX, functionalized with a phosphonic acid group on the spacer region, was developed yielding pronounced robustness, specificity, and enrichment efficiency (Steigenberger et al., 2019).

The workflow generally remains the same in both *in vitro* and *in vivo* methods. The *in vitro* crosslinking method can be applied in extracted or purified proteins whereas the *in vivo* method is designed to target live cells and other complex tissue samples. Benefiting from the upgraded techniques listed above, the *in vivo* XL-MS may provide a greater number of crosslinked peptide pairs and PPIs from proteome-wide search in complex human samples. For instance, the Huang Group separately employed *in vivo* crosslinker DSBSO in human HEK293 cells for XL-MS analyses in 2014 and 2021, respectively (Kaake et al., 2014; Wheat et al., 2021). Over several years of development, the PPIs identification numbers have skyrocketed from hundreds to thousands of proteins, capturing proteins of low abundance and subnetwork PPIs. Moreover, *in vivo* crosslinking enables additional types of crosslinking, including glycan-protein interactions over cell membranes (Xie et al., 2021). The Lebrilla Group developed a technique binding an azide group on sialic acid and primary amine on polypeptides, allowing them to build the glycoprotein-protein crosslinking network from cell surface. Overall, the burgeoning trend of *in*

vivo crosslinking significantly advances the development of system-wide structural biology studies.

On the other hand, *in vivo* crosslinking techniques face challenges, such as cell distortion during crosslinker incubation (Michael et al., 2024). Scientists have attempted to develop *in situ* crosslinking methodology to reveal the architecture of proteins within cellular compartments. In 2015, the Rappsilber Lab first applied BS3 to a purified condensin complex, demonstrating the flexibility of the SMC2/SMC4 structure in mitotic chromosomes (Barysz et al., 2015). For small organelles, nanoparticles have been used to transport crosslinkers into mitochondria (Chen et al., 2023b). Additionally, low-temperature fixation and chemical permeabilization have been proposed to maximize *in situ* crosslinking while preserving cell structure (Michael et al., 2024). Altogether, *in situ* crosslinking broadens the application in multifaceted contexts and extends the investigation into proteins of cellular subcomplexes in greater depth.

2.2.4 TIMS combined with XL-MS

IM-MS has demonstrated significant potential in characterizing biomolecules and conformational transitions. In IM-MS, ions are separated based on their size and shape, characterized by their unique collision cross-section (CCS) (Uetrecht et al., 2010). When two peptides are crosslinked together, their chemical and physical properties may vary drastically. Leveraging parameter optimization, XL-MS has been employed using the Bruker TimsTOF Pro instrument, aiming at differentiation of various types of crosslinked peptides, such as mono-links and interlinks (Steigenberger et al., 2020). The TimsTOF Pro instrument enabled a new technology called parallel accumulation-serial fragmentation (PASEF), enabling a synchronized scanning of precursors with full sensitivity (Meier et al., 2015; Meier et al., 2018). Hereby, a new

acquisition strategy called collisional cross section assisted precursor selection PASEF (caps-PASEF) was proposed to utilize CCS information of these crosslinked peptides to facilitate an effective, *a priori* distinction between molecules of interest, effectively stratifying different types of crosslinks in the visualization graph of CCS versus *m/z* (Steigenberger et al., 2020). This innovation enables discrimination between useful crosslinking signals and unwanted linear "noise" peptides prior to the data analysis by software tools, further increasing the number in detected crosslink peptides while saving time required for software searching. In comparison, a revised workflow employing three urea-based crosslinkers, disuccinimidyl dibutyric urea (DSBU), disuccinimidyl dipropionic urea (DSPU), and disuccinimidyl diacetic urea (DSAU), were proposed, each with a different spacer arm distance (Ihling et al., 2021). The feasibility of these crosslinkers was tested across sample of escalating complexity, including bovine serum albumin (BSA), the *Escherichia coli* ribosome, and HEK293T cell nuclear lysates. The results indicated that CCS-based discrimination facilitated a rapid and highly sensitive search for unique crosslinking sites. This study highlighted the effectiveness of integrating XL-MS with ion mobility separation, not only for elucidating the 3D structures of single proteins but also for conducting system-wide protein interaction studies.

3. BIOINFORMATICS AND SOFTWARE

Careful sample preparation and a well-structured workflow are prerequisites for acquiring high-quality data in scientific research. In the realm of MS-based structural proteomics data analysis, bioinformatics and software are indispensable tools for processing, analyzing, and interpreting biological data. They play critical roles in better understanding the structure, function, dynamics,

and interactions of proteins. Both the LiP-MS and XL-MS methods described herein are particularly convenient because they can be approached as bottom-up proteomics in their respective data analysis workflows, maintaining use of available bottom-up proteomics software platforms (Kaur et al., 2018). Despite this, there are some major considerations for effective application of these methods.

3.1 LiP-MS data analysis pipeline and available resources

The database searching process that accompanies initial data analysis of LiP-MS data is very similar to that of bottom-up proteomics, where MaxQuant (Cox et al., 2008) or Proteome Discoverer (Orsburn 2021) are most commonly employed for raw data searching (Kim et al., 2023; Paukštytė et al., 2023). The major deviation of LiP-MS data analyses from bottom-up proteomics lies in the post-processing steps that follow extracting identifications via database searching. The complications here arise from the fact that each peptide must be handled as its own individual species. Typical experimental setups include Try sample undergoing only trypsin digestion under denaturing conditions and LiP sample firstly exposed to limited enzymatic digestion by broad-specificity proteases under native conditions followed by complete trypsin digestion under denaturing conditions. Structure-specific peptides generated from these steps provide insights into protein conformation, dynamics, and interactions (Lu et al., 2023a). Since LiP cleavages occur within the sequence of tryptic peptides, peptides produced in LiP-MS are shorter than classic tryptic peptides, resulting in increased hydrophilicities and earlier retention times during the reversed-phase LC separation. For this reason, chromatographic peak alignment against control samples should not be conducted, as it is incompatible with this experimental design. Additionally, each peptide in LiP-MS data should be handled as an independent species,

their abundances should not be aggregated to infer a protein's overall abundance. Instead, protein abundance should only be inferred from the control samples undergoing only trypsin digestion (Schopper et al., 2017).

Perhaps the most popular and more straightforward approach is to assess peptide intensities by comparing them to the fold change of their corresponding proteins (Cappelletti et al., 2021; Malinovska et al., 2023; Schopper et al., 2017). This step can be typically handled by a homebuilt script for automation purposes (Wang et al., 2023). Alternatively, some have used spectral counts to normalize for lack of trypsin specificity and endogenous protease activity (Feng et al., 2014). More sophisticated models, such as a multiple linear regression model, have been used to predict Try peptide intensities, LiP peptide intensities, and protein intensities while factoring in race, age, and other demographic information (Mackmull et al., 2022; Malinovska et al., 2023). A variation of this approach, which uses predicted protein intensities to adjust LiP peptide abundance, was recently included in the MSstatsLIP package (Malinovska et al., 2023). This is one of the only packages that exists for LiP-MS experimental analysis. It is postulated that the model normalization approach is advantageous for small datasets where only a few protein abundance changes are expected, given its noted increase in performance over the more classical approach of fold-change based normalization. However, standard error is incorporated into this algorithm, so if an extensive amount of significant structural changes are expected, the classical normalization approach is likely more appropriate (Malinovska et al., 2023). Another R package, protti, provides support for LiP-MS workflows, in which the exported results from any database searching platform can be scrutinized for quality control, data preparation, data analysis, and data interpretation (Quast et al., 2021). To complement these analyses, it is typical to use an alignment resource, such as EMBOSS' alignment tool Needle (Rice et al., 2000) to

determine and map LiP sites to known annotations, typically via UniProt (Apweiler et al., 2004; Bateman et al., 2023). The WebLogo (Crooks et al., 2004) resource can also be utilized to monitor amino acid specificity of proteases. Other resources are also available to visualize LiP data, either through automated mapping (Schopper et al., 2017) or through structural barcodes (Malinovska et al., 2023). Downstream analyses can also include protein functional inquires using Gene Ontology (GO) (Ashburner et al., 2000; Mackmull et al., 2022), classification of diseased samples using ProteinLasso (Huang et al., 2013; Mackmull et al., 2022), and PPIs analysis via the Search tool for the Retrieval of Interacting Gens/Proteins (STRING) (Mackmull et al., 2022; Szklarczyk et al., 2019).

3.2 XL-MS data analysis pipeline and available software

The handling of cross-linked data is notoriously complex due to the quick explosion of the search space, where for n peptides within a search library, the search space scales at n^2 (Crowder et al., 2023). This n^2 consideration is true for cross-linkers that react specifically with a limited number of amino acids. If this specificity is not defined, the search space expands to mn^2 or m^2n^2, if one or both ends of the crosslinker are unspecific, respectively, where m represents the average number of amino acids available to bind (Zhang et al., 2022). While a typical bottom-up proteomics search has a limited search space, when searching for cross-linked peptides, all combinations of two peptides, in addition to unlinked peptides must be considered (Liu et al., 2015; Yılmaz et al., 2022). To address this issue, many software tools have implemented filtering strategies to reduce the size of the input database, thereby shrinking the search space, possibly at the cost of reduced sensitivity (Crowder et al., 2023).

To avoid reducing the size of a database, some implement a simple brute-force enumeration approach, which calculates all possible pairs of peptides (peptide α and β) that equal the mass of a given precursor, including the linker. This approach is ideally very sensitive, accounting for all possible combinations, but is very computationally expensive (Crowder et al., 2023). The other common method uses a two-pass strategy, which first tries to identify the α peptide through the fragmentation spectrum, where the linker and peptide β are handled as an open modification. The detection of the α peptide is highly dependent on the quality of the MS/MS spectrum. The β spectrum is typically not sequenced in the same way that peptide α is, because this peptide typically produces fewer fragments. Sequencing peptides in this manner is referred to as "linearization" (Crowder et al., 2023).

3.2.1 XL-MS data identification strategies

A seemingly endless supply of software packages has been published for the identification of crosslinked peptides from MS datasets. Over the years, lists of these have been published by various groups in 2011 (Mayne et al., 2011), and 2014 (Tinnefeld et al., 2014). Contributing to the difficulty of surveying available packages, many have similar names, for example X!Link (Lee et al., 2007), X-Link (Taverner et al., 2002), XLink (Seebacher et al., 2006a), and X-Links (Anderson et al., 2007), each of which is a separate program with an independent algorithm. In this work, we aim to provide a comprehensive account of programs currently available at the time of writing. We hope this will help future developers from needing to re-invent a program that may already exist.

One of the major hurdles to applying and comparing these cross-linking software packages is that there is little standardization across file formats. This also complicates the pipeline of data

analysis through data visualization. CroCo is available for file conversion to bridge this gap, though is limited to only the output files of xQuest, Kojak, pLink, StavroX, and XiSearch, generating input files for DynamXL, xWalk, xiNET, xVis, and pLabel (Bender et al., 2020). Additionally, early database searching of cross-linked peptides yielded many false-positive identifications. Leveraging the searching tools from bottom-up proteomics, many of the existing quality-control metrics at play have been derived from more mature concepts in linear proteomics identification, particularly the target-decoy approach (Piersimoni et al., 2022a). Other niche tools for limiting false positives have emerged as well. One application presented a filter for mono- and intra-linked peptides, dubbed a mono- and intralink filter (mi-filter). Mono-linked peptides refer to a linker only attached to a single peptide, with the other end hydrolyzed, while intra-linked peptides represent a linker that is attached at both ends to the same peptide (Chen et al., 2022). From a structural standpoint, the program TopoLink was designed to determine the level of agreement between experimental conditions and the resulting structural model of cross-linked protein (Ferrari et al., 2019).

Early software in this area was dedicated to the searching of spectra from non-cleavable linkers, as MS-cleavable linkers are considerably more recent. These included numerous resources, in particular StavroX, xQuest, and Kojak. StavroX, one of the earliest programs, is an exhaustive search strategy, where all combinations of peptide α, linker, and peptide β are searched (Na et al., 2020). xQuest was designed specifically for handling isotopically-encoded non-cleavable cross-linkers (Na et al., 2020). Kojak is an open search strategy that applies a variation of the 2-pass approach, but where the fragmentation spectrum is referenced to sequence both peptides α and β (Crowder et al., 2023; Na et al., 2020).

For MS-cleavable linkers, MeroX remains one of the most popular resources, which uses signature ions to produce confidence crosslinked IDs (Liu et al., 2015). Previously MeroX only applied to MS-cleavable crosslinkers, but has since evolved to include the StavroX algorithm, including functionality for non-cleavable crosslinkers. The MeroX 2.0 algorithm begins by correcting monoisotopic signals, which they determined to be the source of many false positive identifications (Götze et al., 2019). Alternatively, XlinkX is a search engine for identification of crosslinked peptides that operates in two phases and is accompanied by a confidence score. This is achieved through revamping of the well-established target-decoy approach from traditional bottom-up proteomics (Liu et al., 2015; Yugandhar et al., 2020). Also, the combination of a hybrid data acquisition strategy to sequence crosslinked peptides at both MS^2 and MS^3 levels, with new algorithm XlinkX 2.0, which substantially enhancing the depth and accuracy in identifying protein cross-links (Liu et al., 2017). MaXLinker is an engine that uses an "MS^3 centric" approach which uses a "fraction of mis-identifications" metric to monitor the quality of identifications. The MS^3 centric approach replies on the MS3 spectra specifically for identification of peptides, using MS^2 spectra to provide additional context for identifications (Yugandhar et al., 2020). MS Annika initially was developed to identify crosslinked peptides in MS^2 spectra from various cleavable crosslinkers (Pirklbauer et al., 2021). In addition to the MS^2 level, newly released version (MS Annika 2.0) introduces a novel scoring function for peptides identification across multiple MS stages (Birklbauer et al., 2023).

As software evolves, many have shifted from focus on either non-cleavable or MS-cleavable linkers, to handling both. MaxLynx is a plugin to MaxQuant (Cox et al., 2008) that uses peak refinement and other specialized algorithms to improve identification of cross-linked peptides. The peak refinement can reidentify the monoisotopic peak of the isotope patterns that may have

been initially misassigned due to signal noise. MaxLynx is capable of searching both non-cleavable and MS-cleavable linkers (Yılmaz et al., 2022). pDeepXL is software designed for the prediction of MS/MS spectra from crosslinked peptides. This deep-learning model was generated using a transfer-learning technique, originally built from linear peptide PSMs, and trained on both cleavable and non-cleavable linkers. This software is designed to be integrated into the pLink2 search engine (Chen et al., 2021). Another emerging technology is XlinkProphet, which validates cross-linked peptide identifications with machine learning (Yılmaz et al., 2022).

3.2.2 XL-MS data visualization resources

Almost as abundant as identification software, there are a slew of resources available for visualization of XL-MS data analysis results. Unfortunately, no platform is necessarily one-size-fits-all, so these must be selected based on desired output (e.g. linear plots, 3D structural visualization) as well as the resource used for identification. If unsure of which data analysis platform is ideal for a given experimental design, perhaps CLMSVault, a web-based program that combines results from Kojak, pLink, Xi, and xQuest to determine a consensus of cross-link identification and performs subsequent visualization, is a starting point (Courcelles et al., 2017). Another popular web-based platform is xiView, developed by the Rappsilber Lab. It is another powerful visualization tool that can display PPI plots from crosslinking data (Graham et al., 2019). xiView works independently of search software to visualize annotated spectra, 2D networks, and 3D structures simultaneously. This versatile interface enables spectra-matching with elaborate annotations and allows for the uploads of 3D crystal structures from the protein data bank (PDB), facilitating the evaluation of crosslink distances for authenticity. If the desired output is 3D structure visualization in a software like PyMol, crisscrosslinkeR is a fitting choice.

CrisscrosslinkeR is a visualization tool for cross-linking that map protein-protein and protein-RNA interactions either at the sequence level or to the 3D structure via PyMol (Gail et al., 2021). Alternatively, PyXlinkViewer is a PyMOL plugin that enables automated display of inter- and intra-protein crosslinks (Schiffrin et al., 2020). Finally, XlinkAnalyzer is a UCSF Chimera plugin for a 3D molecular viewer (Kosinski et al., 2015), which allows for the visualization of various types of links within proteins and complexes, including inter-links, intra-links, and mono-links (Pettersen et al., 2004). For more flexibility, DynaXL is a visualization tool to observe ensemble structures of multi-domain proteins and protein complexes following crosslink identification with pLink (Gong et al., 2017). This tool can also be accessed via the Python based BioBox, capable of integrating into more customized workflows (Rudden et al., 2022).

To draw biological conclusions in a streamlined fashion, Cross-ID is a platform that works with the XlinkX export (via Proteome Discoverer) to visualize crosslinking results and incorporates additional analyses include gene ontology (GO) enrichment (Ashburner et al., 2000; de Graaf et al., 2019). xVis is a web-based utility that aims to combine biological information, such as evolutionary conservation, with XL-MS data processed via an established search engine. It includes capabilities to apply a variety of filters and subsequently generate visualizations which arrange subunits according to cross-link densities (Grimm et al., 2015). XLinkCyNET (Lima et al., 2021) is a plugin designed for Cytoscape (Shannon et al., 2003), a popular program for network visualization and analysis. This plugin leverages the existing infrastructure and provides a visual network for crosslinking data as well as integrates with other utilities within Cytoscape, including GeneMania (Montojo et al., 2014), DAVID (Dennis et al., 2003), ClueGO (Bindea et al., 2009), and BiNGO (Maere et al., 2005).

Alternatively, software suites that are more robust in modeling include DynamXL**,** a standalone software that specifically accounts for protein flexibility and dynamics in modeling (Degiacomi et al., 2017). IMProv (Ziemianowicz et al., 2021) is a program available via the Mass Spec Studio (Rey et al., 2014) that works to model crosslinked proteins dynamically through incorporation of additional datasets from other structural investigation methods, such as cryo-EM, XRD, and HDX-MS. Matched and non-accessible crosslink score (MNXL) and Jwalk are each complementary webservers for validation of protein models through incorporation of crosslinking-restraints. Jwalk estimates solvent-accessible surface distances (SASD) between crosslinked residues (Bullock et al., 2016; Bullock et al., 2018). While similar in principle, SASD is slightly different than Euclidean distances, which allows for a crosslinker to travel within a protein mass. The SASD calculations were initially introduced by Kahraman, et. al., at which point they were implemented in Xwalk, used for visualizing distances between two amnio acids. It should be noted that while in-principle Jwalk and Xwalk are very similar, users have noted that there are instances where Xwalk estimates the SASD based on a linker entering through a protein mass, it appears that Jwalk is more regularly maintained, and requires less intense computational power (Bullock et al., 2016). Additionally, Jwalk is equipped to estimate both SASD as well as the more traditional Euclidean distance. These distances can then be assessed for quality using MNXL. Jwalk and MNXL are available as standalone programs as well as via a webserver (Bullock et al., 2018). xiNET is a standalone platform for creating node-based visualizations from cross-linked MS datasets, particularly adept at including ambiguous crosslinks within the visualization. XiNET is available as a web-based architecture or can be locally installed (Combe et al., 2015). ProXL is a web-based, search engine-agnostic platform for visualization of crosslinking search results and generating sophisticated plots (Riffle et al., 2016;

Riffle et al., 2019). One advantage of ProXL is its accommodation of generic input formats, a relatively rare feature (de Graaf et al., 2019; Riffle et al., 2016).

Overall, with the abundance of resources available, particularly for XL-MS method, it is imperative that the user take time and care to determine which platform makes the most sense for the experiment at hand. **TABLE 2** provides a summary of all logged cross-linking software, including their most recent release date, novelty, and availability. To the best of our knowledge, all programs are available for free. Careful consideration should be given to the type of linker selected, whether the chosen program is suitable for that linker, and the input requirements of downstream visualization tools of interest.

TABLE 2. A summary of all logged cross-linking software, their most recent release date, novelty, and availability.

Software name	Release	Cleavable/non-cleavable	Novelty	Availability
Blinks (Hoopmann et al., 2010)	2010	Cleavable	Chromatographically identifies inter-linked peptides	https://brucelab.gs.washington.edu/software.html
Cleave-XL (Chakrabarty et al., 2020)	2020	Cleavable	Works for DUCCT dual-cleaved linkers	Available from authors upon request
CRIMP 2.0 (Crowder et al., 2023)	2023	Both	Uses a two-pass search to linearize both α and β peptides	Via Mass Spec Studio (https://www.msstudio.ca/crosslinking/)
Crux (McIlwain et al., 2010; McIlwain et al., 2014)	2023	Both	The program crafts a database that includes single peptides, dead-end products, self-loops, and cross-linked peptides, comparing to theoretical spectra in a SEQUEST-style manner.	https://crux.ms/
ECL2 (Yu et al., 2017)	2018	Non-cleavable	Searches all possible combinations while maintaining efficient computational usage through an additive scoring function	https://github.com/fcyu/ECLViewer2
ECL3.0 (Zhou et al., 2023)	2024	Both	The advancement over ECL2 uses a protein feedback method to compensate for the decreased fragmentation efficiency of non-cleavable cross-linked peptides.	https://github.com/yuweichuan/ECL-PF
Formaldeyde XL Analyzer (Tayri-Wilk et al., 2020)	2020	Both	Incorporates searching of formaldehyde cross-links.	https://github.com/Kalisman-Lab/Search_Formaldehyde_Cross-links

GPMAW (Peri et al., 2001)	2023	Non-cleavable	GPMAW is a web-based search and visualization program. It was one of the earliest programs available.	https://www.gpmaw.com/
Hetake (Holding et al., 2013)	2018	Both	Web-based analysis and automated 3D visualization via PyMol compatibility.	https://github.com/MRC-LMB-MassSpec/Hekate
Kojak (Hoopmann et al., 2010; Hoopmann et al., 2015; Hoopmann et al., 2023)	2023	Both	Applies the efficiency of database searching for cross-linkers with spectral processing and scoring algorithms.	https://kojak-ms.systemsbiology.net/
LinX (Kukačka et al., 2021)	2022	Non-cleavable	Designed for mixed-isotope cross-linking.	https://github.com/KukackaZ/LinX
Mango (Mohr et al., 2018)	2021	Cleavable	Extracts precursor masses from chimeric spectra, identifying at MS2 level.	https://github.com/jpm369/mango
MassMatrix (Xu et al., 2008)	2022	Non-cleavable	Designed for di-sulfide bonds.	https://www.massmatrix.bio/
MaXLinker (Yugandhar et al., 2020)	2020	Cleavable (any)	Applies to the MS2-MS3 XL-MS strategy. Begins search of peptides at MS3 level, moving to MS2 only to provide complimentary information	https://yulab.org/resources/MaxLinker/
MaxLynx (Yılmaz et al., 2022)	2023	Both	Introduces novel scoring algorithms for both noncleavable and cleavable linkers. Fit for routine data, also can accommodate ion mobility data	Via MaxQuant (https://maxquant.org/)
MeroX/StavoX (Gotze et al., 2015; Götze et al., 2012; Götze et al., 2019)	2020	Both (via StavroX for noncleavable)	Relies on signature fragment ions to reduce search space.	https://www.stavrox.com/

MetaMorpheusXL (Lu et al., 2018)	2024	Both	Mitigates need for signature ions and performs proteome-wide search in a short amount of time.	https://github.com/smith-chem-wisc/MetaMorpheus
MS Annika (Pirklbauer et al., 2021)	2021	Cleavable	Confident IDs at MS2 level, without need for MS3 spectra	Via ProteomeDiscoverer (https://www.thermofisher.com)
MS Annika 2.0 (Birklbauer et al., 2023)	2023	Cleavable	Implements ability to search MS3 spectra	Via ProteomeDiscoverer (https://www.thermofisher.com)
OpenPepXL (Netz et al., 2020)	2020	Non-cleavable	Searches entire space without reduction, using parallization to maintain manageable search times.	https://openms.de/openpepxl
Open Search Strategy (Slavin et al., 2020)	2020	Non-cleavable	Searches all possible combinations by comparing the spectra from the same sample with and without cross-linking.	http://biolchem.huji.ac.il/nirka/software.html
pLink2 (Chen et al., 2019a; Yang et al., 2012)	2022	Cleavable	Can perform exhaustive searches of mono-linked, loop-linked, inter-linked, and regular peptides.	http://pfind.org/software/pLink/
ProteinProspector (MS-Bridge) (Chu et al., 2004)	2024	Works with a set list of linkers.	Extensive web-based platform.	https://prospector.ucsf.edu/prospector/mshome.htm
SIM-XL (Lima et al., 2014)	2022	Works with a set list of linkers.	Uses reporter ions to achieve search space reduction.	http://patternlabforproteomics.org/sim-xl/
SpotLink (Zhang et al., 2022)	2022	Non-cleavable	Identifies cross-linked peptides linked by non-specific linkers with localization information.	https://github.com/DICP1810/SpotLink
xComb (Panchaud et al., 2010)	2017	Non-cleavable	From an input database of up to 30 proteins, generates a FASTA file of all cross-	https://github.com/GoodlettLab/xComb

			linked peptide combinations which can be input for MS search using any proteomics database searching software.	
Xilmass (Yılmaz et al., 2016)	2018	Non-cleavable	Explicitly encodes cross-linked residues in the database for searching.	https://compomics.github.io/projects/xilmass/wiki/Home
XiSearch (Mendes et al., 2019)	2023	Both	A program designed for the combination of cross-linking experiments and sequential digestion experiments.	https://github.com/Rappsilber-Laboratory/xiSEARCH
Xlink Identifier (Du et al., 2011)	2017	MS-Cleavable	Works with MS-cleavable linkers that feature an enrichment tag.	https://github.com/du-lab/Xlink-Identifier
XlinkX (Liu et al., 2015)	2024	Both	Novel application of target-decoy approach, includes ability to search MS3 spectra	Via ProteomeDiscoverer (https://www.thermofisher.com)
XLSearch (Ji et al., 2016)	2016	MS-Cleavable	Applies a probabilistic scoring model to produce more correct identifications.	https://github.com/COL-IU/XLSearch
Xolik (Dai et al., 2019)	2018	Non-cleavable	Performs an exhaustive search within a linear timescale while maintaining a strict precursor mass constraint.	https://bioinformatics.hkust.edu.hk/Xolik.html
xQuest/xProphet (Leitner et al., 2014)	2023	Specific for isotopically labeled non-cleavable linkers	Database search for isotopically labeled reagents.	https://gitlab.ethz.ch/leitner_lab/xquest_xprophet
xTract (Walzthoeni et al., 2015)	2015	Non-cleavable	Designed for quantitation of cross-linked peptides	https://www.nature.com/articles/nmeth.3631#MOESM512
ZXMiner (Rivera-Santiago et al., 2015)	2017	Cleavable	Specific for zero-length linkers	https://github.com/sirasris/ZXMiner

Deprecated software: The following software were designed for the evaluation of crosslinked proteins and peptides, but were found to be deprecated as of the time of this publication: X-Link (Taverner et al., 2002), VIRTUALMSLAB (de Koning et al., 2006), SearchXLinks (Wefing et al., 2006), MSX-3D (Heymann et al., 2008), MS2PRO (Kruppa et al., 2003), MS2Links (Kellersberger et al., 2004), CrossSearch (Nadeau et al., 2008), CLPM (Tang et al., 2005b), ASAP (Young et al., 2000), CrossWork (Rasmussen et al., 2011), doXLink (Seebacher et al., 2006b), XLPM (Ji et al., 2016), MS2Assign (Schilling et al., 2003), FindLink (Back et al., 2002), Pro-CrossLink (Gao et al., 2006)

4. REPRESENTATIVE APPLICATIONS

4.1 Neurodegenerative diseases

Neurodegenerative diseases are a broad category of neurological disorders featuring misfolding and aggregation of pathogenic proteins, most notably including Alzheimer's Disease (AD) and Parkinson's Disease (PD) (Raghunathan et al., 2022; Yin et al., 2023; Zhou et al., 2007). These diseases contribute to growing burdens worldwide and are the major cause of disability-adjusted life-year loss. Therefore, the discovery of robust biomarkers has proven to be more critical than ever to aid in early diagnosis, prognosis, and treatment efficacy of these neurodegenerative diseases (Dey et al., 2019; Hansson 2021). MS-based proteomics has emerged as a transformative tool, offering new avenues for identifying and quantifying potential biomarkers in an unbiased manner within the realms of biological and biochemical research, especially in neurodegenerative diseases (Bader et al., 2020; Bai et al., 2021; Dey et al., 2019). Notably, MS-based structural proteomics has evolved as a valuable tool in recent years, partly attributed to its broad applicability to a wide range of protein types, including structured, disordered, monomeric, and multimeric proteins (Beveridge et al., 2021). MS-based structural proteomics brings deeper insights into the pathophysiology of neurodegenerative diseases and introduces a new concept of structural biomarkers for these applications. As an illustration, LiP-MS technology identified 76 structurally altered proteins in the CSF between PD patients and healthy subjects. These proteins were enriched in processes that are dysregulated in PD, with some proteins exhibiting structural changes in the PD brain. This study highlights the significance of global analysis of the structural proteome in individuals with PD (Mackmull et al., 2022). CSF proteins and their structural alterations have been consistently implicated in aging and neurodegenerative diseases (Shuken et

al., 2022). LiP-MS was utilized to explore aging-related changes in the mouse CSF proteome, revealing six high-confidence protein candidates (Kng1, Itih2, Cd5l, Fbln1, Pla2g7 and Ywhaz/b) that exhibited structural alterations with aging. This discovery underscored the capability of LiP-MS to unveil age-related structural changes in CSF in the context of neurodegeneration. Similarly, LiP-MS has been utilized to characterize protein conformational changes in CSF collected from healthy controls, mild cognitive impairment (MCI), and AD cohorts (**Figure 4**). The study revealed hundreds of significantly up- and down-regulated peptides between MCI and AD, along with 53 structurally changed regions in 11 proteins with high confidence between the healthy control and AD groups, highlighting potential structural changes occurring during the progression of AD (Wang et al., 2023). To further enhance the throughput of LiP-MS within a single LC run, a DiLeu-LiP-MS method was developed and employed to effectively profile protein structural changes in AD serum. Among the identified proteins, twenty-three proteins were discovered to undergo structural changes, and 7 out of 23 proteins, including CO3, CO9, C4BPA, APOA1, APOA4, C1R, and APOA, exhibited a potential correlation with AD. These findings offer compelling evidence that support the utility of the DiLeu-LiP-MS method for high-throughput quantitation of structural protein changes, demonstrating significant potential to conduct large-scale and comprehensive quantitative analyses of protein conformational changes in diverse biological systems (Lu et al., 2023a). Similarly, the quantitative XL-MS has revealed the variation of interactions between AD biomarkers throughout the progression of AD (Zhu et al., 2024b). The study provides site-specific insights into previously unreported binding interactions between α-1-antitrypsin (A1AT) and complement component 3 (CO3), suggesting an influence on the coagulation cascade. Additionally, it describes structural transformations in APOE during the MCI stage, specifically

highlighting abnormal binding at Lysine 175, which may contribute to amyloid deposition. Taken together, these application examples demonstrate that both LiP-MS and XL-MS technologies not only can guide the discovery of novel structural biomarkers, but also can facilitate hypothesis generation regarding the underlying processes and mechanisms of neurodegenerative diseases.

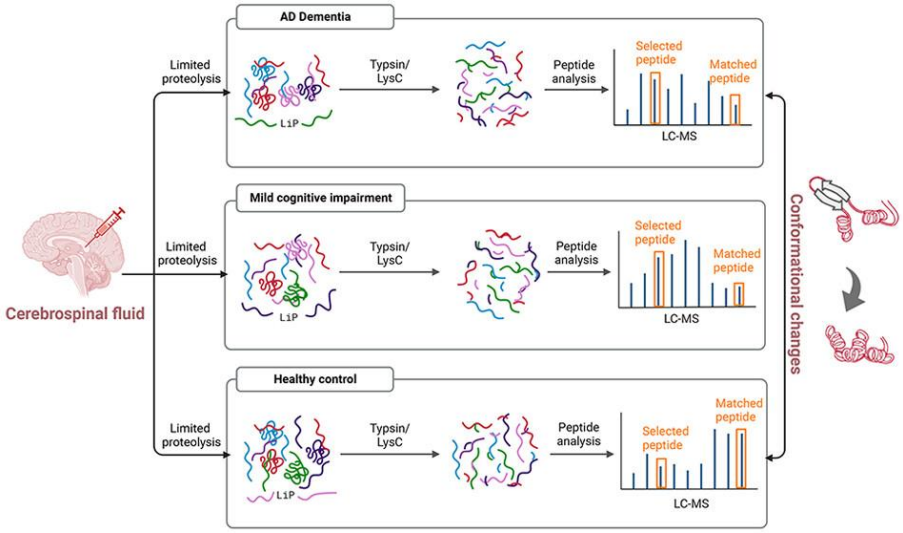


FIGURE 4. Application example of LiP-MS technology in structural proteomic profiling of CSF samples in different stages of AD. Stages represented here include AD dementia, mild cognitive impairment, and healthy controls. Reprinted with permission from Wang. et al. ((Wang et al., 2023), copyright 2023 (American Chemical Society)).

4.2 Interactome

The interactome encompasses a wide range of both intra- and intermolecular interactions. While it offers unparalleled insight into various molecular events within living systems, comprehensive analysis of the interactome poses a significant challenge (Chavez et al., 2022). MS has emerged as powerful tool for studying the interactome, owing to its superior sensitivity, high throughput, and capability to provide rich chemical information (Lu et al., 2023b). In particular, MS-based

technologies including affinity purification mass spectrometry (AP-MS) (Tracy et al., 2022) and proximity labelling techniques (such as engineered ascorbate peroxidase [APEX] (Rhee et al., 2013) and proximity-dependent biotin identification [BioID] (Chou et al., 2018)) have been successfully employed to study the interactome. However, these technologies often lack structural information on protein conformation and interaction. Therefore, it is necessary to develop complementary methods to elucidate the complexities and interactions of systems in native environment. With ongoing improvements in both instrumentation and software, XL-MS and LiP-MS are emerging techniques particularly well-suited for probing the interactome. This section will summarize and discuss some representative applications of XL-MS and LiP-MS to probe different types of interactions in a biological system, including PPIs, protein-metabolite interactions, and protein-drug/ligand interactions.

4.2.1 PPIs

PPIs are involved in various cellular processes and influence biological function; hence the study of PPIs is fundamental to deciphering functional interactions in intricate biological systems (Pelay-Gimeno et al., 2015). The feasibility of LiP-MS in characterizing PPIs *in situ* and detecting interaction interfaces across different classes of proteins has been demonstrated (Holfeld et al., 2023). Also, LiP-MS has confidently investigated the interaction between Calmodulin (CaM) and the rod cyclic nucleotide-gated (CNG) channel found in Rod outer segment membranes. The findings indicated that CaM induced accessibility changes, affecting not only the regions associated with the interaction with CaM binding sites CaM1, CaM2, and the coiled-coil of CNG channel subunit named CNGA1, but also extending to regions of CNGA1 that are in contact distance to another subunit CNGB (Barret et al., 2023).

Like LiP-MS, XL-MS has matured into a robust technology in high-throughput PPIs mapping. In a protocol paper utilizing an isotopically-labelled crosslinker BS3d_0/d_4, the 26S proteasome Lid complex with its intermediate assembly were studied (Chen et al., 2019b). Similarly, PPIs assessment of human complement proteins can be also elucidated using XL-MS. For instance, the QXL-MS workflow combined with an integrative modeling approach successfully revealed the structure of complement component, abbreviated by C3(H_2O), the hydrolytic and conformationally rearranged product of C3 (Chen et al., 2016). As XL-MS progresses into a powerful tool enabling large-scale proteome-wide studies, the number of unique site-specific pairwise crosslink identifications between different proteins is expected to surge from hundreds to thousands. In an illustrative application assessing protein conformation in mitochondria, crosslinker disuccinimidyl suberate (DSS) was employed and identified 5,518 distance restraints between protein residues in human mitochondria. This analysis underscored clusters of highly interactive proteins within the mitochondrial protein complex, providing critical insights into structural analyses of this cellular organelle (Ryl et al., 2020). In an alternative approach to the *in vivo* XL-MS testing (**Figure 5**), the MS-cleavable, cell permeable and enrichable crosslinker Alkyne-A-DSBSO was applied in HEK293 cells and exemplified a large PPIs network with 13,904 unique lysine-lysine linkage, 6,439 interactions among 2,484 proteins (Wheat et al., 2021). This work substantiates that both the scope and the depth of the XL application in PPIs are much broader with high efficiency and sensitivity by established strategies, showing great potential on proteome-wide XL-MS studies in any organisms.

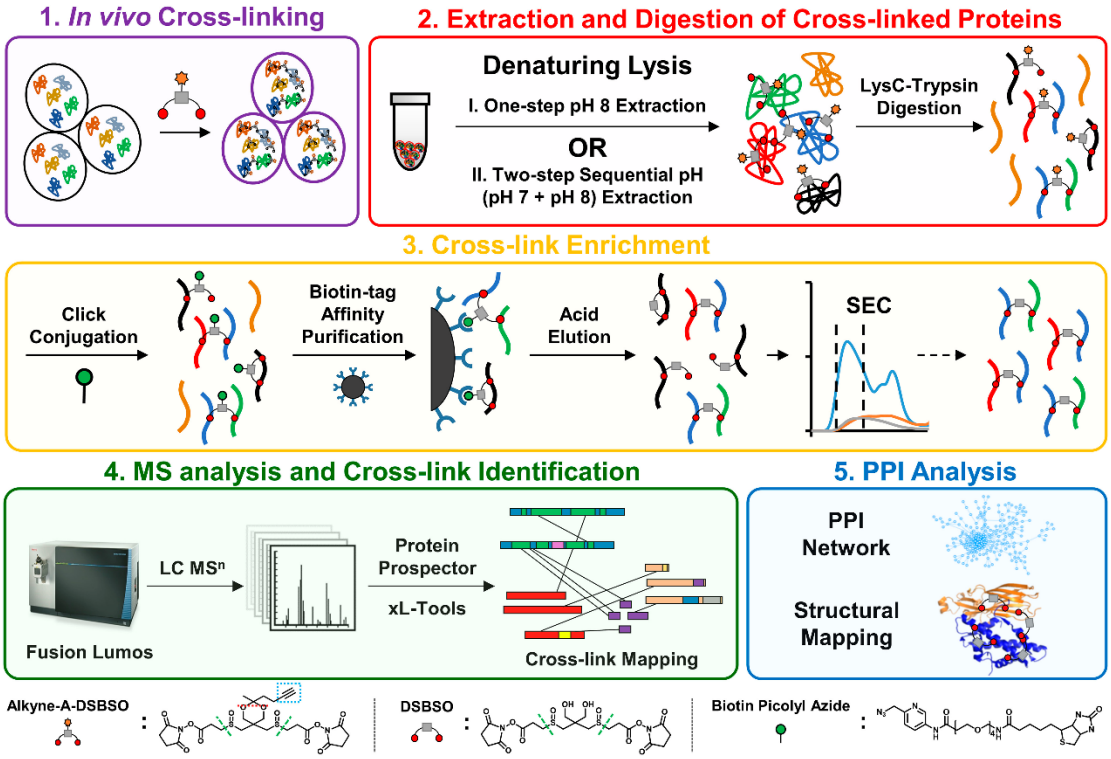


FIGURE 5. Application example of the Alkyne-A-DSBSO-based *in vivo* XL-MS platform for PPIs mapping. Reprinted with permission from Wheat et al. ((Wheat et al., 2021), copyright 2021 (National Academy of Sciences)).

In the path of structural biology development, XL-MS has been routinely coupled with other structural elucidation strategies for an integrative structure modeling workflow. For example, XL-MS has been applied as an auxiliary tool to define distance restraints in building cryo-EM maps (Shakeel et al., 2019). Supported by structural models from XRD and cryo-EM, over 100 crosslinks were matched between membrane attack complex (MAC) protein components (Khakzad et al., 2020), delineating a new quaternary assembly model in its native environment. Also, XL-MS has been combined with cellular cryo-electron tomography (cryo-ET) for analysis of *Mycoplasma pneumoniae*, through which 577 distinct PPIs at a 5% PPIs false-discovery rate (FDR) were identified (O'Reilly et al., 2020). This study underscores the potential of integrative

in-cell structural biology methods for unraveling the dynamic machineries of proteins within their native functional context. Another integrated study enabled by HDX-MS, XL-MS and molecular docking successfully characterized the binding interface and predicted the 3D structure of IL-7/IL-7Rα (Zhang et al., 2019a). HDX-MS data refined restrains derived from zero-length crosslinks for quaternary structure prediction in solution, generating a high-confidence model of the IL-7/IL-7Rα complex and successfully revealing a new binding region on the C-terminus of helix B of IL-7. Conclusively, these studies establish a framework for integrating XL-MS data with other structural biology techniques, such as HDX, XRD, and cryo-EM, to enable an improved understanding of protein structures through a holistic approach.

4.2.2 Protein-metabolite interactions

The study of protein-metabolite interactions is equally important as the study of PPIs. The interaction between proteins and metabolites plays a crucial role in governing various cellular processes, often triggering local or global alterations of protein structures. For example, metabolic activity is intricately connected to the T cell fate and function. LiP-MS has facilitated the proteome-wide investigation of structural alterations by the analysis of knockout T cell clones, revealing three structurally modified proteins (BAZ1B, PSIP1, and TSN). These proteins were found to sense L-arginine levels and contribute to the reprograming of T cells towards enhanced survival capacity (Geiger et al., 2016). LiP-MS was also used to systematically identify both known and novel protein-metabolite interactions in *Escherichia coli* (Piazza et al., 2018). The obtained data not only unveiled an extensive network comprising 1,678 interactions and 7,345 putative binding sites, but also provided functional principles of metabolite sensing, offering a valuable means to assess how changes in metabolite concentration within the cellular environment affect protein-metabolite interactions. This work motivated further investigation to

translate this prokaryotic workflow into eucaryotic investigations. To extend its applicability to the systematic identification of protein-metabolite interactions in eukaryotic proteomes, a machine learning-based workflow, termed as LiP-Quant, has been introduced to identify drug targets and binding sites in complex eukaryotic proteomes (**Figure 6**) (Piazza et al., 2020). The results suggested that LiP-Quant could estimate the half maximal effective concentration of compound binding sites in whole cell lysates, accurately distinguish drug binding to homologous proteins, and identify the previously unknown targets of a fungicide research compound. Overall, these advancements further highlight the significant promise of LiP-MS as an indispensable component of the chemoproteomics toolbox for analysis of protein-metabolite interactions.

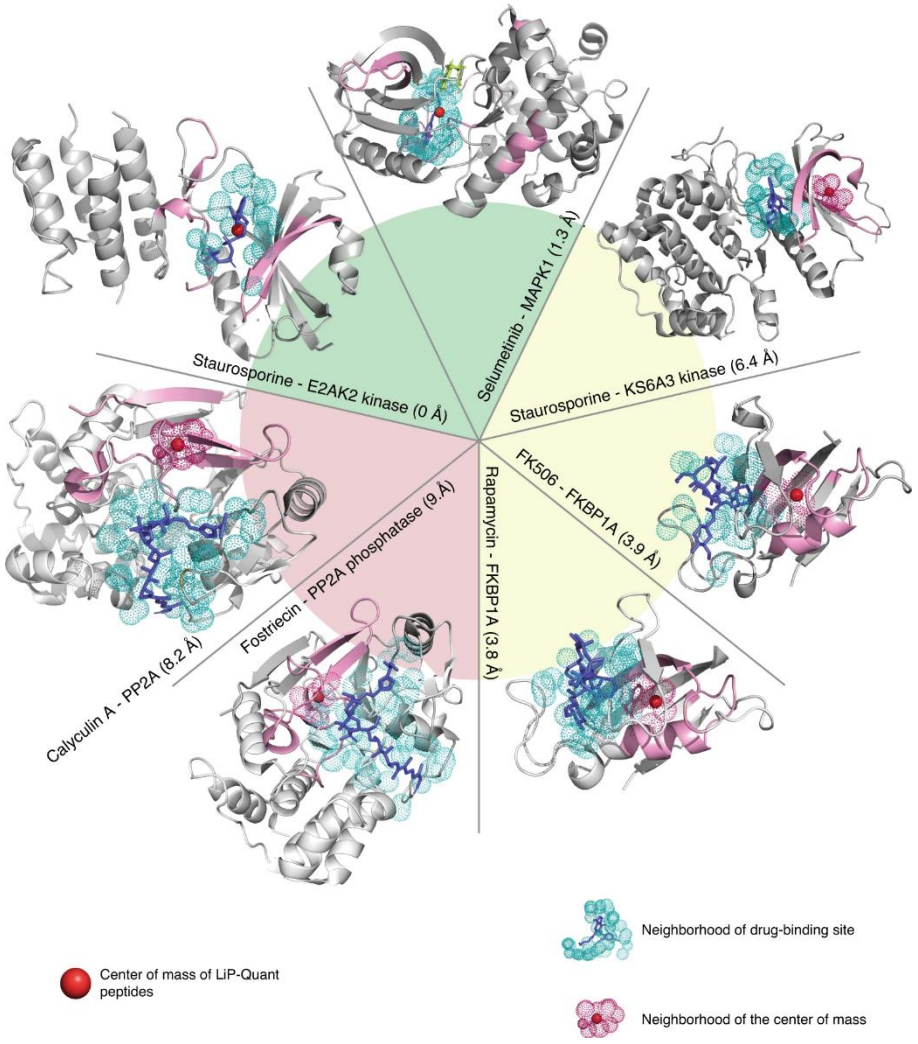


FIGURE 6. Application example of LiP-Quant in approximating drug binding sites positions. Reprinted with permission from Piazza et al. ((Piazza et al., 2020), copyright 2020 (Springer Nature)).

4.2.3 Protein-drug/ligand interactions

While the direct interaction of crosslinkers with small molecules has not yet become a prevailing strategy, the interaction between drug and protein interaction can be reflected by large-scale measurement of crosslinks. For example, crosslinker PIR/iqPIR has been utilized in a series of experiments. Treatment of HeLa cells with increasing concentrations of paclitaxel, a chemotherapeutic agent, successfully identified and visualized crosslinks positively correlating with paclitaxel treatment in tubulin, demonstrating drug-induced stabilization of microtubules *in vivo* (Chavez et al., 2019). Subsequently, HSP90 inhibitor 17-AAG treatment indicates rising level of crosslink number within the HS90B and HS90A homodimer, indicating a more compact conformation of HSP90 under drug effect (Chavez et al., 2020). Furthermore, increasing throughput using 6-plex isobaric labeling, researchers treated breast cancer MCF-7 cells with five different inhibitor drugs, revealing drug-induced interactome changes in breast cancer (**Figure 7A**) (Wippel et al., 2022). The multiplexing capability and quantitative accuracy of iqPIR make it particularly suitable for comparative studies, offering a deeper understanding of how protein interactions contribute to cellular function and regulation.

As protein structure data continues to expand, the study of molecular interactions between ligands and their target molecules gains increasing attention (Adasme et al., 2021). This importance stems from the essential role of characterizing interactions within protein-ligand complexes in structural bioinformatics research (Salentin et al., 2015). With the assistance of MS-based structural proteomics technologies, it is now feasible to study the conformational

properties of proteins and protein-ligand complexes in complex biological mixtures (Kaur et al., 2018). In a proof-of-concept study (**Figure 7B**), the interaction between a well-studied protein, bovine serum albumin (BSA), and the ligand bilirubin, which targets lysine residues, was examined under urea-based denaturation conditions by protein denaturation and QXL-MS (Mathay et al., 2023). The presented strategy was designed to complement other protein denaturation MS-based approaches, such as stability of proteins from rates of oxidation (SPROX), providing valuable reference points for studying protein-drug/ligand interactions on the cellular scale.

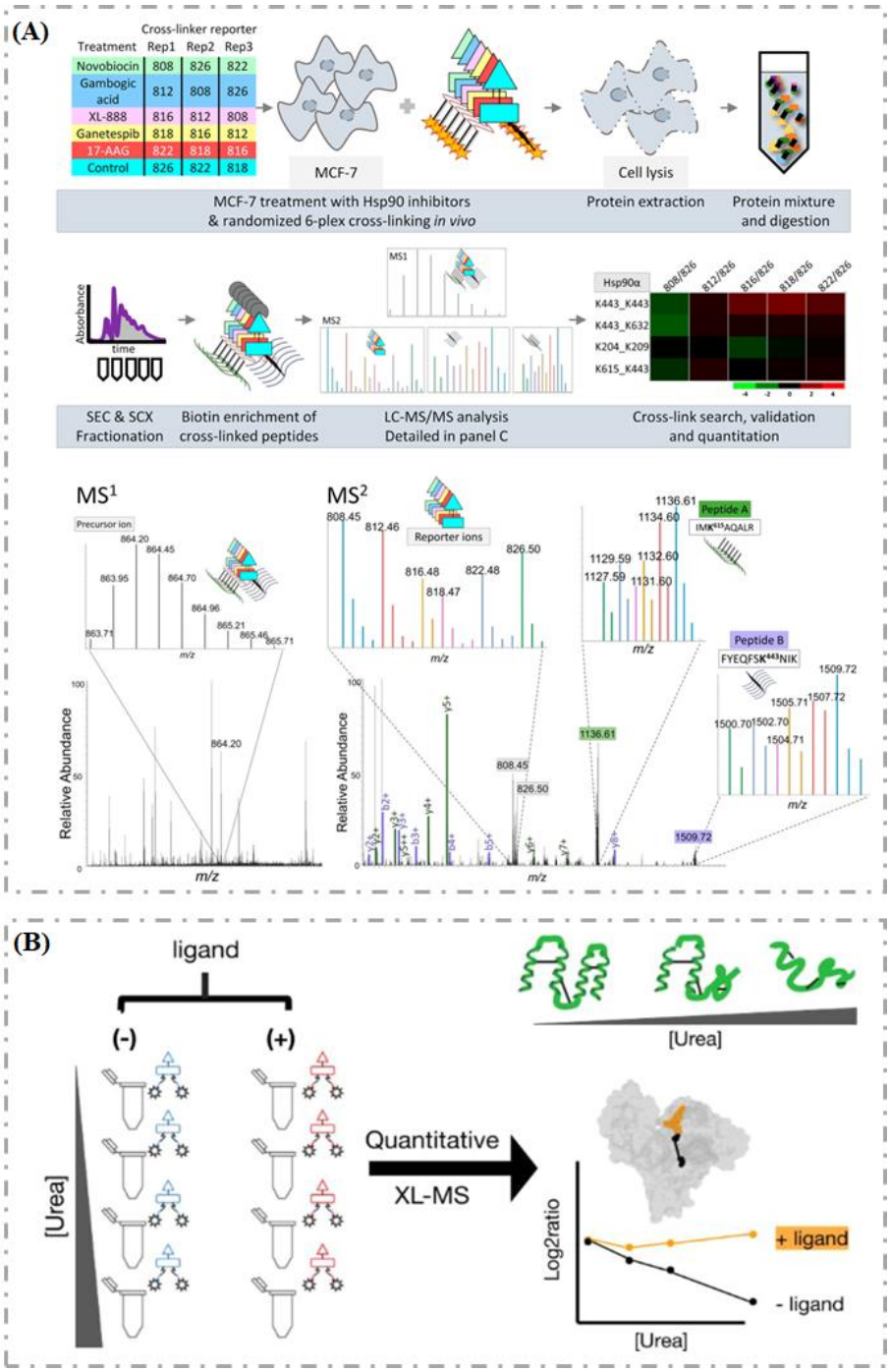


FIGURE 7. Application examples of QXL-MS in protein-drug/ligand interaction. (A) 6-plex iqPIR is demonstrated on MCF-7 breast cancer cells treated with five different Hsp90 inhibitors. Reprinted with permission from Wippel et al. ((Wippel et al., 2022), copyright 2022 (American Chemical Society)). (B) The study of protein-ligand interactions by protein denaturation and

QXL-MS. Reprinted with permission from Mathay et al. ((Mathay et al., 2023), copyright 2023 (American Chemical Society)).

4.3 Membrane proteins

Membrane proteins (MPs) constitute a substantial portion of therapeutic targets and play an important role in numerous essential cellular functions such as signal transduction, transport, intercellular joining, cell-cell recognition, cell adhesion, and catalysis (Yin et al., 2016). Despite their importance, the available structural information for MPs is relatively limited in comparison to water soluble proteins (Calabrese et al., 2018). Unlike the water soluble proteins, MPs need to be maintained in amphipathic agents for stabilization (Dafun et al., 2022). The continuous development of MS has provided crucial information in this arena, not only on the major compositions and primary structure of MPs complexes, but also their higher order structures, making MS a significant pillar for integrative structural biology.

In 2013, a covalent linkage technique through proximal cysteine was developed, targeting the G protein-coupled receptors (GPCRs) specifically at the CRF-R1 site (Xiang et al., 2013). Building upon this foundation, the Coin group has utilized crosslinking to investigate the GPCRs ligand-receptor interaction in the past few years. In 2020, they characterized the topology of β-arrestin, a ubiquitous regulator of GPCR signaling, through the crosslinking application of a chemical named BrEtY (Böttke et al., 2020). This study revealed the finger loop structure movement during the arrestin activation and exploited the cellular environment that cannot be accessed by cryogenic electron microscopy (cryoEM) or crystallography. Three years later, they optimized their photo-crosslinking strategy, featuring with more details on a class B GPCR, parathyroid hormone 1 receptor (PTH1R) activation along with the arrestin-2 (arr2) complex (Aydin et al., 2023). In this work, the 3D models of entire G-protein bound active site were generated under

the facilitation by the extensive pairwise crosslinking identification, especially in the proximal phosphorylation cluster. The advantage of their integrative modeling guided by crosslinking lies in capturing not a single snapshot of the complex but an average of overall conformations. Their work provides unprecedented insight into the PTH1R-arr2 interaction revealing GPCR dynamics and physiological environment within the live cell. Furthermore, XL-MS has unveiled dynamic features of GPCR structural regions that were not observable in crystal structures. For instance, the researchers employed seven crosslinkers and carefully evaluated their compatibility with receptor proteins. An AP-MS assay was adapted for assessment of crosslinking impact on the ligand-binding activity (Xia et al., 2020). Through this, they successfully identified the crosslinks at the specific regions on both extracellular and intracellular sides, providing more details than the molecular dynamics simulations could achieve. Finally, sub-organelle membrane topologies can also be validated using XL-MS. The Liu group has introduced a novel approach called cross-link assisted spatial proteomics (CLASP) strategy (Zhu et al., 2024a), which employs DSSO, DSBSO, and other enzymatic labeling approaches to elucidate mitochondrial sub-compartments and predict protein localization on both outer and inner membranes. Importantly, they have successfully validated their results through confocal fluorescence imaging. This work provides comprehensive maps of sub-organelle proteomes and membrane protein topologies, which are crucial for understanding cellular organization and function.

For other types of MPs, such as the F-type ATPase, the Robinson group has demonstrated that conformational changes in its subunits strongly suggest a regulatory role for phosphorylation in nucleotide binding (Schmidt et al., 2013). Their research employed an isotopically labeled crosslinker strategy using BS3-d0/d4, which elucidated multiple interactions between subunits and revealed differences between the untreated and dephosphorylated forms. Similarly, the

Raunser lab utilized a comparable approach with DSS-d0/d12 to investigate conformational changes in the ryanodine receptor, proposing a calcium-induced switch mechanism (Efremov et al., 2014). This study is another exemplary use of crosslinking mass spectrometry (XLMS) integrated with cryo-electron microscopy (cryoEM) to advance structural model investigations. In the context of the E. coli system, the Sinz lab provided functional insights into respiratory formate dehydrogenases (Fdh's) located in the periplasmic space, highlighting their role in stabilizing the cell envelope (Zorn et al., 2014). Overall, advances in structural biology have greatly facilitated the analysis of many previously inaccessible MPs, making structure-function studies and structure-based design of drugs targeting MPs more feasible.

4.4 Artificial intelligence-based structural analysis

Structural analysis of protein interactions typically involves platforms that integrate docking information with molecular modeling. For XL-MS, HADDOCK and ROSETTA are extraordinarily useful tools that incorporate distance restraints from crosslink data with Protein Data Bank (PDB) structures (Kahraman et al., 2013; van Zundert et al., 2016), leveraging algorithms and extensive calculations. With the rapid advancement of artificial intelligence (AI), structural prediction has entered a new era facilitated by AlphaFold (Jumper et al., 2021; Yang et al., 2023). Developed by DeepMind, AlphaFold2 can predict atomic-level 3D structures of proteins with high accuracy and resolution from sequence libraries and databases (Skolnick et al., 2021).

Many publications have described hybrid methods combining deep learning architecture with crosslinking data. In 2022, it was reported that covalent labeling data can enhance the computational framework of AlphaFold when combined with ROSETTA predictions based on experimental modifications of residues (Drake et al., 2022). For instance, in the case of actin/gs1,

the root-mean-square deviation (RMSD) of the best-scoring model significantly decreased from 11.2 Å to 1.6 Å when compared to the prediction made by AlphaFold2 alone. This improvement in RMSD was also illustrated in a 2024 paper combining AlphaFold2 ensemble with XL-MS data to model flexible protein structures (Manalastas-Cantos et al., 2024). The crosslink data also provides a benchmark for whether AlphaFold2 successfully provides accurate model proteins in their native intracellular conformations. One *in situ* crosslinking report states that 43% of AlphaFold2-predicted protein structures do not violate the Cα-to-Cα distances of 30 Å (McCafferty et al., 2023), underscoring its potential to bridge the gap between *in vitro* predictions and *in vivo* realities. Furthermore, the Rappsilber Group validated this statement by utilizing in-cell photo-crosslinking mass spectrometry combined with deep learning (Stahl et al., 2023), demonstrating that this approach not only enhances the accuracy of protein structure predictions but also provides a robust framework for studying protein interactions within the cellular milieu. Collectively, these studies underscore the transformative impact of integrating AlphaFold with XL-MS and other experimental techniques, paving the way for creating more accurate and comprehensive models of protein structure and function.

5. FACTORS THAT COULD LEAD TO PROTEIN CONFORMATIONAL CHANGES

Changes in protein conformation play an important role in facilitating diverse biochemical processes, encompassing signaling, phosphorylation, transport, and catalysis (Bhardwaj et al., 2009). Protein conformation change is influenced by various factors, including environmental changes, allosteric effects induced by binding events, post-translational modifications (PTMs),

and biomolecular interactions. This section will specifically focus on the factors affecting protein conformation changes, including allosteric and synergistic effects, PTMs and other contributors, such as physiological stress factors and environmental risk factors.

5.1 Allosteric and synergistic effects

Allosteric and synergistic effects are pivotal in influencing the conformation and function of proteins. Allosteric regulation, a universal phenomenon, occurs when an effector (such as ligand, PTMs and any other structural component) binds at an allosteric site, inducing conformational or dynamic changes at a distant functional active site. This binding can enhance or inhibit protein activity (Dokholyan 2016; Fauser et al., 2022; Sljoka 2021). For instance, PKM2, an important subtype of pyruvate kinase (PK), undergoes allosteric regulation during its transition between dimers and tetramers, influenced by both endogenous and exogenous activators and inhibitors (**Figure 8**) (Zhang et al., 2019b). Allosteric regulation enables precise control of protein function in response to specific signals. The investigation into the structural model of μ-opioid receptor confirmed that ligand binding leads to the conformational flexibility in the third intracellular loop (Serohijos et al., 2011).

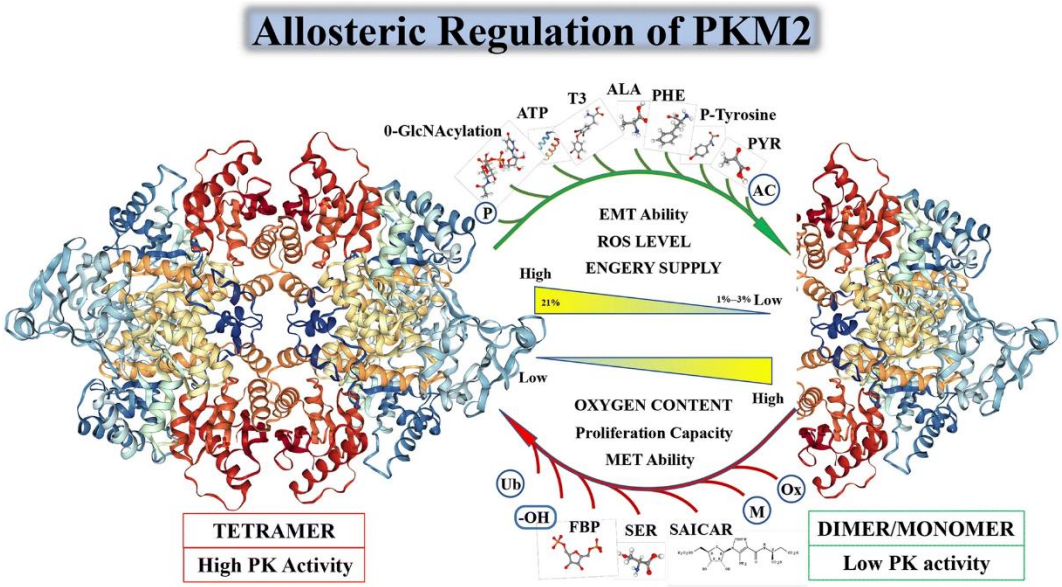


FIGURE 8. Allosteric regulation of PKM2 by various endogenous and exogenous activators and inhibitors. Reprinted with permission from Zhang et al. ((Zhang et al., 2019b)), copyright 2019 (Springer Nature).

In contrast, synergistic effects involve the cooperative interaction of multiple effectors to intricately fine-tune conformational changes, leading to enhanced functional outcome. For instance, a study uncovered a GPCR priming mechanism that resulted from the synergistic effects of two distinct G proteins (Gupte et al., 2017). A straightforward method employing HDX-MS quantified combinatorial effects in a dual-ligand kinase model, revealing multiple hotspots of synergistic allostery when two ligands bind to protein kinase 3-phosphoinositide-dependent protein kinase 1 (PDK1) (Ghode et al., 2020). Furthermore, the combination of solution NMR experiments with molecular dynamics simulations demonstrated that two key residues (Phe34 and Leu39) synergistically work to induce Trp 43 conformational change in protein DANCER-3 (Damry et al., 2019). The integration of allosteric and synergistic

mechanisms provides a sophisticated regulatory framework, allowing proteins to adapt to diverse environmental cues and ensuring precise cellular responses to maintain homeostasis.

5.2 PTMs

PTMs are classified as either reversible or irreversible and can occur at a single site or multiple sites. These modifications not only affect various protein behaviors and characteristics, including folding (Shental-Bechor et al., 2008), lifespan (Van et al., 2023), interaction (Li et al., 2013; Yang et al., 2020), enzyme activity, assembly and localization (WARDEN et al., 2001), but also alter protein conformation and function. For example, phosphorylation, one of the most common and significant PTMs, involves the covalent conjugation of phosphate groups to serine, threonine, or tyrosine in a protein by protein kinases, altering the protein's characteristics from hydrophobic to hydrophilic and enabling the protein to undergo conformational changes (Ardito et al., 2017). Glycosylation is a more complex PTM from a biochemical standpoint than phosphorylation, the effects of glycosylation on protein conformation can be quite heterogeneous, depending on the attached glycans moieties (Kumar et al., 2020). Taken together, the chemical diversity, vast number, and heterogeneity of PTMs poses a tremendous challenge in understanding protein conformation changes caused by PTMs, making it a complex and demanding area of research.

5.3 Other factors

Apart from allosteric and synergistic effects and PTMs, physiological stress such as pH, temperature, ionic strength, and mechanical pressure, as well as environmental risk factors such as heavy metals, pesticides, radiation, and air pollution can have an impact on protein conformation (Devi et al., 2022). Misfolding and aggregation of specific pathogenic protein are

the hallmark features of neurodegenerative diseases, such as amyloid-beta (Aβ) in AD, and alpha-synuclein (α-syn) in PD (Ross et al., 2004). Consequently, there are many previous studies that reported both physiological stress and environmental toxicants related to neurodegenerative disease, including in AD and PD (Agnihotri et al., 2020; Chin-Chan et al., 2015). Understanding of how these factors affect protein conformation is valuable to many fields, including biochemistry, molecular biology, biophysics, and medicine, which will not only help researchers to assess structural impact on living organisms, but also to gain molecular insights into disease mechanisms and develop effective therapeutic strategies.

6. CONCLUSIONS AND PERSPECTIVES

In summary, MS-based structural proteomics is an established, impactful tool for analyzing the structure of proteins and protein complexes. Specifically, two popular peptide-level readout MS techniques, LiP-MS and XL-MS, are particularly appropriate for these studies. LiP-MS elucidates protein structural dynamics and solvent accessibility, whereas XL-MS aids in deciphering protein spatial organization and interactions. Both techniques have provided structural insights into the conformation of proteins affected by allosteric effectors, PTMs, and the interactome, spanning from purified proteins to complex biological samples like cells, tissues, biofluid samples, and bacteria.

While LiP-MS and XL-MS have unveiled numerous protein structures and complexes, and provided valuable insights into structural proteomics, challenges persist, especially regarding limited sample quantities and structural heterogeneity. Thus, improvements are needed in several areas. Enhancing mass spectrometer sensitivity can facilitate the detection of low abundance peptides and prevent the loss of important information related to protein conformation changes.

Additionally, expanding the application of LiP-MS and XL-MS to map protein interactomes and monitor protein conformation changes within single cells and biological processes, including cellular differentiation and disease progression, holds significant promise. Sophisticated bioinformatics and modeling software breakthroughs are essential for extracting meaningful insights from complex and large-scale raw data. Integrating multiple types of experimental data, from complementary methods such as HDX-MS, cryo-EM and XRD, into modeling processes can propel structural proteomics into new frontiers. Furthermore, more research focusing on membrane protein structural dynamics is needed to reveal their diverse functions and interactions essential for life processes. Elaborating on the relationship between protein conformation changes with PTMs represents a demanding area of research. Addressing these areas will undoubtedly advance the field of MS-based structural proteomics and deepen our understanding of biological systems.

7. ACKNOWLEDGEMENTS

This review is dedicated to Professor Carlito Lebrilla for his notable contributions and substantial impact to various fields of mass spectrometry. Preparation of this manuscript was supported in part by National Institutes of Health (NIH) (R21AG065728, R01AG052324, R01AG078794, and R01DK071801). H.L. and H.Z. would like to thank the funding support for Postdoctoral Career Development Award provided by the American Society for Mass Spectrometry. L.F. was supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number T32GM008505 (Chemistry-Biology Interface Training Program). L.L. would like to acknowledge NIH grants S10OD028473, and

S10OD025084, as well as funding support from a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

8. REFERENCES

Abdella PM, Smith PK, Royer GP. A new cleavable reagent for cross-linking and reversible immobilization of proteins. *Biochem. Biophys. Res. Commun.* **1979**, 87: 734-742.

Adasme MF, Linnemann KL, Bolz SN, Kaiser F, Salentin S, Haupt VJ, Schroeder M. Plip 2021: Expanding the scope of the protein-ligand interaction profiler to DNA and rna. *Nucleic Acids Res.* **2021**, 49: W530-W534.

Agnihotri A, Aruoma OI. Alzheimer's disease and parkinson's disease: A nutritional toxicology perspective of the impact of oxidative stress, mitochondrial dysfunction, nutrigenomics and environmental chemicals. *J. Am. Coll. Nutr.* **2020**, 39: 16-27.

Anderson GA, Tolic N, Tang X, Zheng C, Bruce JE. Informatics strategies for large-scale novel cross-linking analysis. *J. Proteome Res.* **2007**, 6: 3412-3421.

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh L-SL. Uniprot: The universal protein knowledgebase. *Nucleic Acids Res.* **2004**, 32: D115–D119.

Ardito F, Giuliani M, Perrone D, Troiano G, Muzio LL. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (review). *Int. J. Mol. Med.* **2017**, 40: 271-280.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-

Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: Tool for the unification of biology. *Nature Genetics 2000 25:1* **2000**, 25.

Aydin Y, Böttke T, Lam JH, Ernicke S, Fortmann A, Tretbar M, Zarzycka B, Gurevich VV, Katritch V, Coin I. Structural details of a class b gpcr-arrestin complex revealed by genetically encoded crosslinkers in living cells. *Nat. Commun.* **2023**, 14: 1151.

Back JW, Notenboom V, de Koning LJ, Muijsers AO, Sixma TK, de Koster CGa, de Jong L. Identification of cross-linked peptides for protein interaction studies using mass spectrometry and 18o labeling. *Anal. Chem.* **2002**, 74: 4417–4422.

Bader JM, Geyer PE, Muller JB, Strauss MT, Koch M, Leypoldt F, Koertvelyessy P, Bittner D, Schipke CG, Incesoy EI, Peters O, Deigendesch N, Simons M, Jensen MK, Zetterberg H, Mann M. Proteome profiling in cerebrospinal fluid reveals novel biomarkers of alzheimer's disease. *Mol. Syst. Biol.* **2020**, 16: e9356.

Bai B, Vanderwall D, Li Y, Wang X, Poudel S, Wang H, Dey KK, Chen PC, Yang K, Peng J. Correction to: Proteomic landscape of alzheimer's disease: Novel insights into pathogenesis and biomarker discovery. *Mol. Neurodegener.* **2021**, 16: 55.

Bamberger C, Pankow S, Martinez-Bartolome S, Ma M, Diedrich J, Rissman RA, Yates JR, 3rd. Protein footprinting via covalent protein painting reveals structural changes of the proteome in alzheimer's disease. *J. Proteome Res.* **2021**, 20: 2762-2771.

Barret DCA, Schuster D, Rodrigues MJ, Leitner A, Picotti P, Schertler GFX, Kaupp UB, Korkhov VM, Marino J. Structural basis of calmodulin modulation of the rod cyclic nucleotide-gated channel. *Proc. Natl. Acad. Sci. U. S. A.* **2023**, 120: e2300309120.

Barth M, Schmidt C. Native mass spectrometry-a valuable tool in structural biology. *J. Mass Spectrom.* **2020**, 55: e4578.

Barth M, Schmidt C. Quantitative cross-linking of proteins and protein complexes. *Methods Mol. Biol.* **2021**, 2228: 385-400.

Barysz H, Kim JH, Chen ZA, Hudson DF, Rappsilber J, Gerloff DL, Earnshaw WC. Three-dimensional topology of the smc2/smc4 subcomplex from chicken condensin i revealed by cross-linking and molecular modelling. *Open Biol* **2015**, 5: 150005.

Bateman A, Martin M-J, Orchard S, Magrane M, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bye-A-Jee H, Cukura A, Denny P, Dogan T, Ebenezer T, Fan J, Garmiri P, da Costa Gonzales LJ, Hatton-Ellis E, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Joshi V, Jyothi D, Kandasaamy S, Lock A, Luciani A, Lugaric M, Luo J, Lussi Y, MacDougall A, Madeira F, Mahmoudy M, Mishra A, Moulang K, Nightingale A, Pundir S, Qi G, Raj S, Raposo P, Rice DL, Saidi R, Santos R, Speretta E, Stephenson J, Totoo P, Turner E, Tyagi N, Vasudev P, Warner K, Watkins X, Zaru R, Zellner H, Bridge AJ, Aimo L, Argoud-Puy G, Auchincloss AH, Axelsen KB, Bansal P, Baratin D, Batista Neto TM, Blatter M-C, Bolleman JT, Boutet E, Breuza L, Gil BC, Casals-Casas C, Echioukh KC, Coudert E, Cuche B, de Castro E, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gaudet P, Gehant S, Gerritsen V, Gos A, Gruaz N, Hulo C, Hyka-Nouspikel N, Jungo F, Kerhornou A, Le Mercier P, Lieberherr D, Masson P, Morgat A, Muthukrishnan V, Paesano S, Pedruzzi I, Pilbout S, Pourcel L, Poux S, Pozzato M, Pruess M, Redaschi N, Rivoire C, Sigrist CJA, Sonesson K, Sundaram S, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Zhang J. Uniprot: The universal protein knowledgebase in 2023. *Nucleic Acids Res.* **2023**, 51: D523-D531.

Bender J, Schmidt C. The croco cross-link converter: A user-centred tool to convert results from cross-linking mass spectrometry experiments. *Bioinformatics* **2020**, 36: 1296-1297.

Beveridge R, Calabrese AN. Structural proteomics methods to interrogate the conformations and dynamics of intrinsically disordered proteins. *Front. Chem.* **2021**, 9: 603639.

Bhardwaj N, Gerstein M. Relating protein conformational changes to packing efficiency and disorder. *Protein Sci.* **2009**, 18: 1230-1240.

Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z, Galon J. Cluego: A cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics (Oxford, England)* **2009**, 25: 1091-1093.

Birklbauer MJ, Matzinger M, Müller F, Mechtler K, Dorfer V. Ms annika 2.0 identifies cross-linked peptides in ms2–ms3-based workflows at high sensitivity and specificity. *J. Proteome Res.* **2023**, 22: 3009-3021.

Blackburn MR, Minkoff BB, Sussman MR. Mass spectrometry-based technologies for probing the 3d world of plant proteins. *Plant Physiol.* **2022**, 189: 12-22.

Böttke T, Ernicke S, Serfling R, Ihling C, Burda E, Gurevich VV, Sinz A, Coin IJEr. Exploring gpcr‐arrestin interfaces with genetically encoded crosslinkers. **2020**, 21: e50437.

Bullock JMA, Schwab J, Thalassinos K, Topf M. The importance of non-accessible crosslinks and solvent accessible surface distance in modeling proteins with restraints from crosslinking mass spectrometry. *Mol. Cell. Proteomics* **2016**, 15: 2491-2500.

Bullock JMA, Thalassinos K, Topf M. Jwalk and mnxl web server: Model validation using restraints from crosslinking mass spectrometry. *Bioinformatics* **2018**, 34: 3584-3585.

Burke AM, Kandur W, Novitsky EJ, Kaake RM, Yu C, Kao A, Vellucci D, Huang L, Rychnovsky SD. Synthesis of two new enrichable and ms-cleavable cross-linkers to define protein-protein interactions by mass spectrometry. *Org. Biomol. Chem.* **2015**, 13: 5030-5037.

Calabrese AN, Radford SE. Mass spectrometry-enabled structural biology of membrane proteins. *Methods* **2018**, 147: 187-205.

Cappelletti V, Hauser T, Piazza I, Pepelnjak M, Malinovska L, Fuhrer T, Li Y, Dörig C, Boersema P, Gillet L, Grossbach J, Dugourd A, Saez-Rodriguez J, Beyer A, Zamboni N, Caflisch A, de Souza N, Picotti P. Dynamic 3d proteomes reveal protein functional alterations at high resolution in situ. *Cell* **2021**, 184: 545-559.

Chakrabarty JK, Sadananda SC, Bhat A, Naik AJ, Ostwal DV, Chowdhury SM. High confidence identification of cross-linked peptides by an enrichment-based dual cleavable cross-linking technology and data analysis tool cleave-xl. *J. Am. Soc. Mass. Spectrom.* **2020**, 31: 173-182.

Chavez JD, Hoopmann MR, Weisbrod CR, Takara K, Bruce JEJPo. Quantitative proteomic and interaction network analysis of cisplatin resistance in hela cells. **2011**, 6: e19892.

Chavez JD, Schweppe DK, Eng JK, Zheng C, Taipale A, Zhang Y, Takara K, Bruce JE. Quantitative interactome analysis reveals a chemoresistant edgotype. *Nat. Commun.* **2015**, 6: 7928.

Chavez JD, Schweppe DK, Eng JK, Bruce JE. In vivo conformational dynamics of hsp90 and its interactors. *Cell Chem. Biol.* **2016**, 23: 716-726.

Chavez JD, Keller A, Zhou B, Tian R, Bruce JE. Cellular interactome dynamics during paclitaxel treatment. *Cell Rep.* **2019**, 29: 2371-2383 e2375.

Chavez JD, Keller A, Mohr JP, Bruce JE. Isobaric quantitative protein interaction reporter technology for comparative interactome studies. *Anal. Chem.* **2020**, 92: 14094-14102.

Chavez JD, Park SG, Mohr JP, Bruce JE. Applications and advancements of ft-icr-ms for interactome studies. *Mass Spectrom. Rev.* **2022**, 41: 248-261.

Chen J, Zhao Q, Gao H, Zhao LL, Chu HY, Shan YC, Liang Z, Zhang YK, Zhang LH. A glycosidic-bond-based mass-spectrometry-cleavable cross-linker enables in vivo cross-linking for protein complex analysis. *Angew. Chem. Int. Ed.* **2023a**, 62.

Chen X, Sailer C, Kammer KM, Fürsch J, Eisele MR, Sakata E, Pellarin R, Stengel F. Mono- and intralink filter (mi-filter) to reduce false identifications in cross-linking mass spectrometry data. *Anal. Chem.* **2022**, 94: 17751-17756.

Chen Y, Zhou W, Xia Y, Zhang W, Zhao Q, Li X, Gao H, Liang Z, Ma G, Yang K, Zhang L, Zhang Y. Targeted cross-linker delivery for the in situ mapping of protein conformations and interactions in mitochondria. *Nat. Commun.* **2023b**, 14: 3882.

Chen Z-L, Meng J-M, Cao Y, Yin J-L, Fang R-Q, Fan S-B, Liu C, Zeng W-F, Ding Y-H, Tan D, Wu L, Zhou W-J, Chi H, Sun R-X, Dong M-Q, He S-M, Chen Z-L, Meng J-M, Cao Y, Yin J-L, Fang R-Q, Fan S-B, Liu C, Zeng W-F, Ding Y-H, Tan D, Wu L, Zhou W-J, Chi H, Sun R-X, Dong M-Q, He S-M. A high-speed search engine plink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.* **2019a**, 10.

Chen Z-L, Mao P-Z, Zeng W-F, Chi H, He S-M. Pdeepxl: Ms/ms spectrum prediction for cross-linked peptide pairs by deep learning. *J. Proteome Res.* **2021**, 20: 2570-2582.

Chen ZA, Pellarin R, Fischer L, Sali A, Nilges M, Barlow PN, Rappsilber J. Structure of complement c3(h2o) revealed by quantitative cross-linking/mass spectrometry and modeling. *Mol. Cell. Proteomics* **2016**, 15: 2730-2743.

Chen ZA, Rappsilber J. Protein dynamics in solution by quantitative crosslinking/mass spectrometry. *Trends Biochem. Sci* **2018**, 43: 908-920.

Chen ZA, Rappsilber J. Quantitative cross-linking/mass spectrometry to elucidate structural changes in proteins and their complexes. *Nat. Protoc.* **2019b**, 14: 171-201.

Chin-Chan M, Navarro-Yepes J, Quintanilla-Vega B. Environmental pollutants as risk factors for neurodegenerative disorders: Alzheimer and parkinson diseases. *Front. Cell. Neurosci.* **2015**, 9: 124.

Chou CC, Zhang Y, Umoh ME, Vaughan SW, Lorenzini I, Liu F, Sayegh M, Donlin-Asp PG, Chen YH, Duong DM, Seyfried NT, Powers MA, Kukar T, Hales CM, Gearing M, Cairns NJ, Boylan KB, Dickson DW, Rademakers R, Zhang YJ, Petrucelli L, Sattler R, Zarnescu DC, Glass JD, Rossoll W. Tdp-43 pathology disrupts nuclear pore complexes and nucleocytoplasmic transport in als/ftd. *Nat. Neurosci.* **2018**, 21: 228-239.

Christofi E, Barran P. Ion mobility mass spectrometry (im-ms) for structural biology: Insights gained by measuring mass, charge, and collision cross section. *Chem. Rev.* **2023**, 123: 2902-2949.

Chu F, Shan S-o, Moustakas DT, Alber F, Egea PF, Stroud RM, Walter P, Burlingame AL, Chu F, Shan S-o, Moustakas DT, Alber F, Egea PF, Stroud RM, Walter P, Burlingame AL. Unraveling the interface of signal recognition particle and its receptor by using chemical cross-linking and tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101: 16454-16459.

Combe CW, Fischer L, Rappsilber J. Xinet: Cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **2015**, 14: 1137-1147.

Courcelles M, Coulombe-Huntington J, Cossette É, Gingras A-C, Thibault P, Tyers M. Clmsvault: A software suite for protein cross-linking mass-spectrometry data analysis and visualization. *J. Proteome Res.* **2017**, 16: 2645-2652.

Cox J, Mann M. Maxquant enables high peptide identification rates, individualized p.P.B.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, 26: 1367-1372.

Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed maxlfq. *Mol. Cell. Proteomics* **2014**, 13: 2513-2526.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. Weblogo: A sequence logo generator. *Genome Res.* **2004**, 14.

Crowder DA, Sarpe V, Amaral BC, Brodie NI, Michael ARM, Schriemer DC. High-sensitivity proteome-scale searches for crosslinked peptides using crimp 2.0. *Anal. Chem.* **2023**, 95: 6425-6432.

Dafun AS, Marcoux J. Structural mass spectrometry of membrane proteins. *Biochim. Biophys. Acta* **2022**, 1870: 140813.

Dai J, Jiang W, Yu F, Yu W. Xolik: Finding cross-linked peptides with maximum paired scores in linear time. *Bioinformatics* **2019**, 35: 251-257.

Damry AM, Mayer MM, Broom A, Goto NK, Chica RA. Origin of conformational dynamics in a globular protein. *Commun Biol* **2019**, 2: 433.

de Graaf SC, Klykov O, van den Toorn H, Scheltema RA. Cross-id: Analysis and visualization of complex xl–ms-driven protein interaction networks. *J. Proteome Res.* **2019**, 18: 642-651.

de Koning LJ, Kasper PT, Back JW, Nessen MA, Vanrobaeys F, Van Beeumen J, Gherardi E, de Koster CG, de Jong L. Computer‐assisted mass spectrometric analysis of naturally occurring and artificially introduced cross‐links in proteins and protein complexes. *The FEBS Journal* **2006**, 273: 281-291.

Degiacomi MT, Schmidt C, Baldwin AJ, Benesch JLP. Accommodating protein dynamics in the modeling of chemical crosslinks. *Structure* **2017**, 25: 1751-1757.e1755.

Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. David: Database for annotation, visualization, and integrated discovery. *Genome Biol.* **2003**, 4: P3-P3.

Devi S, Chaturvedi M, Fatima S, Priya S. Environmental factors modulating protein conformations and their role in protein aggregation diseases. *Toxicology* **2022**, 465: 153049.

Dey KK, Wang H, Niu M, Bai B, Wang X, Li Y, Cho JH, Tan H, Mishra A, High AA, Chen PC, Wu Z, Beach TG, Peng J. Deep undepleted human serum proteome profiling toward biomarker discovery for alzheimer's disease. *Clin. Proteomics* **2019**, 16: 16.

Dokholyan NV. Controlling allosteric networks in proteins. *Chem. Rev.* **2016**, 116: 6463-6487.

Drake ZC, Seffernick JT, Lindert S. Protein complex prediction using rosetta, alphafold, and mass spectrometry covalent labeling. *Nat. Commun.* **2022**, 13: 7846.

Du X, Chowdhury SM, Manes NP, Wu S, Mayer MU, Adkins JN, Anderson GA, Smith RD. Xlink-identifier: An automated data analysis platform for confident identifications of chemically cross-linked peptides using tandem mass spectrometry. *J. Proteome Res.* **2011**, 10: 923–931.

Engen JR. Analysis of protein conformation and dynamics by hydrogen/deuterium exchange ms. *Anal. Chem.* **2009**, 81: 7870-7875.

Fauser J, Leschinsky N, Szynal BN, Karginov AV. Engineered allosteric regulation of protein function. *J. Mol. Biol.* **2022**, 434: 167620.

Feng Y, De Franceschi G, Kahraman A, Soste M, Melnik A, Boersema PJ, de Laureto PP, Nikolaev Y, Oliveira AP, Picotti P. Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **2014**, 32: 1036-1044.

Ferrari AJR, Clasen MA, Kurt L, Carvalho PC, Gozzo FC, Martínez L. Topolink: Evaluation of structural models using chemical crosslinking distance constraints. *Bioinformatics* **2019**, 35: 3169-3170.

Fischer L, Chen ZA, Rappsilber J. Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J. Proteomics* **2013**, 88: 120-128.

Franken H, Mathieson T, Childs D, Sweetman GM, Werner T, Togel I, Doce C, Gade S, Bantscheff M, Drewes G, Reinhard FB, Huber W, Savitski MM. Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat. Protoc.* **2015**, 10: 1567-1593.

Gail EH, Shah AD, Schittenhelm RB, Davidovich C. Crisscrosslinker: Identification and visualization of protein–rna and protein–protein interactions from crosslinking mass spectrometry. *Bioinformatics* **2021**, 36: 5530-5532.

Gao Q, Song Xue, Catalin E. Doneanu, Scott A. Shaffer, David R. Goodlett a, Sidney D. Nelson*. Pro-crosslink. Software tool for protein cross-linking and mass spectrometry. *Anal. Chem.* **2006**, 78: 2145–2149.

Geiger R, Rieckmann JC, Wolf T, Basso C, Feng Y, Fuhrer T, Kogadeeva M, Picotti P, Meissner F, Mann M, Zamboni N, Sallusto F, Lanzavecchia A. L-arginine modulates t cell metabolism and enhances survival and anti-tumor activity. *Cell* **2016**, 167: 829-842.

Ghode A, Gross LZF, Tee WV, Guarnera E, Berezovsky IN, Biondi RM, Anand GS. Synergistic allostery in multiligand-protein interactions. *Biophys J* **2020**, 119: 1833-1848.

Gong Z, Liu Z, Dong X, Ding Y-H, Dong M-Q, Tang C. Protocol for analyzing protein ensemble structures from chemical cross-links using dynaxl. *Biophys. Rep.* **2017**, 3: 100-108.

Gotze M, Pettelkau J, Fritzsche R, Ihling CH, Schafer M, Sinz A. Automated assignment of ms/ms cleavable cross-links in protein 3d-structure analysis. *J. Am. Soc. Mass. Spectrom.* **2015**, 26: 83-97.

Gotze M, Iacobucci C, Ihling CH, Sinz A. A simple cross-linking/mass spectrometry workflow for studying system-wide protein interactions. *Anal. Chem.* **2019**, 91: 10236-10244.

Götze M, Pettelkau J, Schaks S, Bosse K, Ihling CH, Krauth F, Fritzsche R, Kühn U, Sinz A. Stavrox—a software for analyzing crosslinked products in protein interaction studies. *J. Am. Soc. Mass. Spectrom.* **2012**, 23: 76-87.

Götze M, Iacobucci C, Ihling CH, Sinz A. A simple cross-linking/mass spectrometry workflow for studying system-wide protein interactions. *Anal. Chem.* **2019**, 91: 10236-10244.

Graham M, Combe C, Kolbowski L, Rappsilber J. **2019**.

Grimm M, Zimniak T, Kahraman A, Herzog F. xvis : A web server for the schematic visualization and interpretation of crosslink-derived spatial restraints. *Nucleic Acids Res.* **2015**, 43: W362-W369.

Gupte TM, Malik RU, Sommese RF, Ritt M, Sivaramakrishnan S. Priming gpcr signaling through the synergistic effect of two g proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, 114: 3756-3761.

Hansson O. Biomarkers for neurodegenerative diseases. *Nat. Methods* **2021**, 27: 954-963.

Heymann M, Paramelle D, Subra G, Forest E, Martinez J, Geourjon C, Deléage G. Msx-3d: A tool to validate 3d protein models using mass spectrometry. *Bioinformatics* **2008**, 24: 2782-2783.

Holding AN, Lamers MH, Stephens E, Skehel JM. Hekate: Software suite for the mass spectrometric analysis and three-dimensional visualization of cross-linked protein samples. *J. Proteome Res.* **2013**, 12: 5923-5933.

Holfeld A, Schuster D, Sesterhenn F, Stalder P, Haenseler W, Barrio-Hernandez I, Ghosh D, Vowles J, Cowley SA, Nagel LJb. Systematic identification of structure-specific protein-protein interactions. *bioRxiv* **2023**: 2023.2002. 2001.522707.

Holm L. Dali server: Structural unification of protein families. *Nucleic Acids Res.* **2022**, 50: W210-W215.

Hoopmann MR, Weisbrod CR, Bruce JE. Improved strategies for rapid identification of chemically cross-linked peptides using protein interaction reporter technology. *J. Proteome Res.* **2010**, 9: 6323-6333.

Hoopmann MR, Zelter A, Johnson RS, Riffle M, MacCoss MJ, Davis TN, Moritz RL. Kojak: Efficient analysis of chemically cross-linked protein complexes. *J. Proteome Res.* **2015**, 14: 2190-2198.

Hoopmann MR, Shteynberg DD, Zelter A, Riffle M, Lyon AS, Agard DA, Luan Q, Nolen BJ, MacCoss MJ, Davis TN, Moritz RL. Improved analysis of cross-linking mass spectrometry data with kojak 2.0, advanced by integration into the trans-proteomic pipeline. *J. Proteome Res.* **2023**, 22: 647-655.

Huang T, Gong H, Yang C, He Z. Proteinlasso: A lasso regression approach to protein inference problem in shotgun proteomics. *Comput. Biol. Chem.* **2013**, 43: 46-54.

Iacobucci C, Gotze M, Ihling CH, Piotrowski C, Arlt C, Schafer M, Hage C, Schmidt R, Sinz A. A cross-linking/mass spectrometry workflow based on ms-cleavable cross-linkers and the merox software for studying protein structures and protein-protein interactions. *Nat. Protoc.* **2018**, 13: 2864-2889.

Ihling CH, Piersimoni L, Kipping M, Sinz A. Cross-linking/mass spectrometry combined with ion mobility on a timstof pro instrument for structural proteomics. *Anal. Chem.* **2021**, 93: 11442-11450.

Isasa M, Rose CM, Elsasser S, Navarrete-Perea J, Paulo JA, Finley DJ, Gygi SP. Multiplexed, proteome-wide protein expression profiling: Yeast deubiquitylating enzyme knockout strains. *J. Proteome Res.* **2015**, 14: 5306-5317.

Ji C, Li S, Reilly JP, Radivojac P, Tang H. Xlsearch: A probabilistic database search algorithm for identifying cross-linked peptides. *J. Proteome Res.* **2016**, 15: 1830–1841.

Jin Lee Y. Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes. *Mol. Biosyst.* **2008**, 4: 816-823.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with alphafold. *Nature* **2021**, 596: 583-589.

Kaake RM, Wang X, Burke A, Yu C, Kandur W, Yang Y, Novtisky EJ, Second T, Duan J, Kao A, Guan S, Vellucci D, Rychnovsky SD, Huang L. A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells. *Mol. Cell. Proteomics* **2014**, 13: 3533-3543.

Kahraman A, Herzog F, Leitner A, Rosenberger G, Aebersold R, Malmstrom L. Cross-link guided molecular modeling with rosetta. *Plos One* **2013**, 8: e73411.

Kao A, Chiu CL, Vellucci D, Yang Y, Patel VR, Guan S, Randall A, Baldi P, Rychnovsky SD, Huang L. Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol. Cell. Proteomics* **2011**, 10: M110 002212.

Kaur U, Meng H, Lui F, Ma R, Ogburn RN, Johnson JHR, Fitzgerald MC, Jones LM. Proteome-wide structural biology: An emerging field for the structural analysis of proteins on the proteomic scale. *J. Proteome Res.* **2018**, 17: 3614-3627.

Kaur U, Johnson DT, Chea EE, Deredge DJ, Espino JA, Jones LM. Evolution of structural biology through the lens of mass spectrometry. *Anal. Chem.* **2019**, 91: 142-155.

Kellersberger KA, Yu E, Kruppa GH, Young MM, Fabris D. Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation. *Anal. Chem.* **2004**, 76: 2438-2445.

Khakzad H, Happonen L, Tran Van Nhieu G, Malmstrom J, Malmstrom L. In vivo cross-linking ms of the complement system mac assembled on live gram-positive bacteria. *Front. Genet.* **2020**, 11: 612475.

Kim D, Lee MS, Kim ND, Lee S, Lee HS. Identification of α-amanitin effector proteins in hepatocytes by limited proteolysis-coupled mass spectrometry. *Chem. Biol. Interact.* **2023**, 386: 110778-110778.

Kiselar J, Chance MR. High-resolution hydroxyl radical protein footprinting: Biophysics tool for drug discovery. *Annu Rev Biophys* **2018**, 47: 315-333.

Kosinski J, von Appen A, Ori A, Karius K, Müller CW, Beck M. Xlink analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **2015**, 189: 177-183.

Kruppa GH, Schoeniger J, Young MM. A top down approach to protein structural studies using chemical cross‐linking and fourier transform mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, 17: 155-162.

Kukacka Z, Rosulek M, Strohalm M, Kavan D, Novak P. Mapping protein structural changes by quantitative cross-linking. *Methods* **2015**, 89: 112-120.

Kukačka Z, Rosůlek M, Jelínek J, Slavata L, Kavan D, Novák P. Linx: A software tool for uncommon cross-linking chemistry. *J. Proteome Res.* **2021**, 20: 2021-2027.

Kumar A, Narayanan V, Sekhar A. Characterizing post-translational modifications and their effects on protein conformation using nmr spectroscopy. *Biochemistry* **2020**, 59: 57-73.

Lee YJ, Lackner LL, Nunnari JM, Phinney BSJJopr. Shotgun cross-linking analysis for studying quaternary and tertiary protein structures. *J. Proteome Res.* **2007**, 6: 3908-3917.

Leitner A, Walzthoeni T, Kahraman A, Herzog F, Rinner O, Beck M, Aebersold R. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics* **2010**, 9: 1634-1649.

Leitner A, Walzthoeni T, Aebersold R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using lc-ms/ms and the xquest/xprophet software pipeline. *Nat. Protoc.* **2014**, 9: 120-137.

Leitner A. Cross-linking and other structural proteomics techniques: How chemistry is enabling mass spectrometry applications in structural biology. *Chem. Sci.* **2016**, 7: 4792-4803.

Li J, Lyu W, Rossetti G, Konijnenberg A, Natalello A, Ippoliti E, Orozco M, Sobott F, Grandori R, Carloni P. Proton dynamics in protein mass spectrometry. *J. Phys. Chem. Lett.* **2017**, 8: 1105-1112.

Li X, Foley EA, Kawashima SA, Molloy KR, Li Y, Chait BT, Kapoor TM. Examining post-translational modification-mediated protein-protein interactions using a chemical proteomics approach. *Protein Sci.* **2013**, 22: 287-295.

Lima DB, de Lima TB, Balbuena TS, Neves-Ferreira AGC, Barbosa VC, Gozzo FC, Carvalho PC. Sim-xl: A powerful and user-friendly tool for peptide cross-linking analysis. *J. Proteomics* **2014**, 129: 51-55.

Lima DB, Zhu Y, Liu F. Xlinkcynet: A cytoscape application for visualization of protein interaction networks based on cross-linking mass spectrometry identifications. *J. Proteome Res.* **2021**, 20: 1943-1950.

Liu F, Rijkers DTS, Post H, Heck AJR. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **2015**, 12: 1179-1184.

Liu F, Lossl P, Scheltema R, Viner R, Heck AJR. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.* **2017**, 8: 15473.

Lu H, Wang B, Liu Y, Wang D, Fields L, Zhang H, Li M, Shi X, Zetterberg H, Li L. Dileu isobaric labeling coupled with limited proteolysis mass spectrometry for high-throughput profiling of protein structural changes in alzheimer's disease. *Anal. Chem.* **2023a**, 95: 9746-9753.

Lu H, Zhang H, Li L. Chemical tagging mass spectrometry: An approach for single-cell omics. *Anal. Bioanal. Chem.* **2023b**, 415: 6901-6913.

Lu L, Millikin RJ, Solntsev SK, Rolfs Z, Scalf M, Shortreed MR, Smith LM. Identification of ms-cleavable and noncleavable chemically cross-linked peptides with metamorpheus. *J. Proteome Res.* **2018**, 17: 2370–2376.

Mackmull M-T, Nagel L, Sesterhenn F, Muntel J, Grossbach J, Stalder P, Bruderer R, Reiter L, van de Berg WDJ, de Souza N, Beyer A, Picotti P. Global, in situ analysis of the structural proteome in individuals with parkinson's disease to identify a new class of biomarker. *Nat. Struct. Mol. Biol.* **2022**, 29: 978–989.

Maere S, Heymans K, Kuiper M. Bingo: A cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **2005**, 21: 3448-3449.

Malinovska L, Cappelletti V, Kohler D, Piazza I, Tsai T-H, Pepelnjak M, Stalder P, Dörig C, Sesterhenn F, Elsässer F, Kralickova L, Beaton N, Reiter L, de Souza N, Vitek O, Picotti P. Proteome-wide structural changes measured with limited proteolysis-mass spectrometry: An advanced protocol for high-throughput applications. *Nat. Protoc.* **2023**, 18: 659-682.

Manalastas-Cantos K, Adoni KR, Pfeifer M, Märtens B, Grünewald K, Thalassinos K, Topf M. Modeling flexible protein structure with alphafold2 and crosslinking mass spectrometry. *Mol. Cell. Proteomics* **2024**, 23: 100724.

Masson GR, Burke JE, Ahn NG, Anand GS, Borchers C, Brier S, Bou-Assaf GM, Engen JR, Englander SW, Faber J, Garlish R, Griffin PR, Gross ML, Guttman M, Hamuro Y, Heck AJR,

Houde D, Iacob RE, Jorgensen TJD, Kaltashov IA, Klinman JP, Konermann L, Man P, Mayne L, Pascal BD, Reichmann D, Skehel M, Snijder J, Strutzenberg TS, Underbakke ES, Wagner C, Wales TE, Walters BT, Weis DD, Wilson DJ, Wintrode PL, Zhang Z, Zheng J, Schriemer DC, Rand KD. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (hdx-ms) experiments. *Nat. Methods* **2019**, 16: 595-602.

Mateus A, Kurzawa N, Becher I, Sridharan S, Helm D, Stein F, Typas A, Savitski MM. Thermal proteome profiling for interrogating protein interactions. *Mol. Syst. Biol.* **2020**, 16: e9232.

Mateus A, Kurzawa N, Perrin J, Bergamini G, Savitski MM. Drug target identification in tissues by thermal proteome profiling. *Annu. Rev. Pharmacool. Toxicol.* **2022**, 62: 465-482.

Mathay M, Keller A, Bruce JE. Studying protein-ligand interactions by protein denaturation and quantitative cross-linking mass spectrometry. *Anal. Chem.* **2023**, 95: 9432-9436.

Matzinger M, Kandioller W, Doppler P, Heiss EH, Mechtler K. Fast and highly efficient affinity enrichment of azide-a-dsbso cross-linked peptides. *J. Proteome Res.* **2020**, 19: 2071-2079.

Matzinger M, Mechtler K. Cleavable cross-linkers and mass spectrometry for the ultimate task of profiling protein-protein interaction networks in vivo. *J. Proteome Res.* **2021**, 20: 78-93.

Mayne SLN, Patterton HG. Bioinformatics tools for the structural elucidation of multi-subunit protein complexes by mass spectrometric analysis of protein-protein cross-links. *Brief. Bioinform.* **2011**, 12: 660-671.

McCabe JW, Hebert MJ, Shirzadeh M, Mallis CS, Denton JK, Walker TE, Russell DH. The ims paradox: A perspective on structural ion mobility-mass spectrometry. *Mass Spectrom. Rev.* **2021**, 40: 280-305.

McCafferty CL, Pennington EL, Papoulas O, Taylor DW, Marcotte EM. Does alphafold2 model proteins' intracellular conformations? An experimental test using cross-linking mass spectrometry of endogenous ciliary proteins. *Commun. Biol.* **2023**, 6: 421.

McIlwain S, Draghicescu P, Singh P, Goodlett DR, Noble WS. Detecting cross-linked peptides by searching against a database of cross-linked peptide pairs. *J. Proteome Res.* **2010**, 9: 2488-2495.

McIlwain S, Tamura K, Kertesz-Farkas A, Grant CE, Diament B, Frewen B, Howbert JJ, Hoopmann MR, Käll L, Eng JK, MacCoss MJ, Noble WS. Crux: Rapid open source protein tandem mass spectrometry analysis. *J. Proteome Res.* **2014**, 13: 4488-4491.

Meier F, Beck S, Grassl N, Lubeck M, Park MA, Raether O, Mann M. Parallel accumulation-serial fragmentation (pasef): Multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J. Proteome Res.* **2015**, 14: 5378-5387.

Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, Goedecke N, Decker J, Kosinski T, Park MA, Bache N, Hoerning O, Cox J, Rather O, Mann M. Online parallel accumulation-serial fragmentation (pasef) with a novel trapped ion mobility mass spectrometer. *Mol. Cell. Proteomics* **2018**, 17: 2534-2545.

Mendes ML, Fischer L, Chen ZA, Barbon M, O'Reilly FJ, Giese SH, Bohlke-Schneider M, Belsom A, Dau T, Combe CW, Graham M, Eisele MR, Baumeister W, Speck C, Rappsilber J. An integrated workflow for crosslinking mass spectrometry. *Mol. Syst. Biol.* **2019**, 15: e8994.

Mendoza VL, Vachet RW. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* **2009**, 28: 785-815.

Michael ARM, Amaral BC, Ball KL, Eiriksson KH, Schriemer DC. **2024**.

Mohr JP, Perumalla P, Chavez JD, Eng JK, Bruce JE. Mango: A general tool for collision induced dissociation-cleavable cross-linked peptide identification. *Anal. Chem.* **2018**, 90: 6028-6034.

Montojo J, Zuberi K, Rodriguez H, Bader GD, Morris Q. Genemania: Fast gene network construction and function prediction for cytoscape. *F1000Research* **2014**, 3: 153-153.

Muller F, Fischer L, Chen ZA, Auchynnikava T, Rappsilber J. On the reproducibility of label-free quantitative cross-linking/mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **2018**, 29: 405-412.

Na S, Paek E. Computational methods in mass spectrometry-based structural proteomics for studying protein structure, dynamics, and interactions. *Computational and Structural Biotechnology Journal* **2020**, 18: 1391-1402.

Nadeau OW, Wyckoff GJ, Paschall JE, Artigues A, Sage J, Villar MT, Carlson GM. Crosssearch, a user-friendly search engine for detecting chemically cross-linked peptides in conjugated proteins. *Mol. Cell. Proteomics* **2008**, 7: 739-749.

Netz E, Dijkstra TMH, Sachsenberg T, Zimmermann L, Walzer M, Monecke T, Ficner R, Dybkov O, Urlaub H, Kohlbacher O. Openpepxl: An open-source tool for sensitive identification of cross-linked peptides in xl-ms. *Mol. Cell. Proteomics* **2020**, 19: 2157-2168.

O'Reilly FJ, Rappsilber J. Cross-linking mass spectrometry: Methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **2018**, 25: 1000-1008.

O'Reilly FJ, Xue L, Graziadei A, Sinn L, Lenz S, Tegunov D, Blötz C, Singh N, Hagen WJH, Cramer P, Stülke J, Mahamid J, Rappsilber J. In-cell architecture of an actively transcribing-translating expressome. *Science* **2020**, 369: 554-557.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, 1: 376-386.

Orsburn BC. Proteome discoverer—a community enhanced data processing suite for protein informatics. *Proteomes* **2021**, 9: 15.

Pan X, Vachet RW. Membrane protein structures and interactions from covalent labeling coupled with mass spectrometry. *Mass Spectrom. Rev.* **2022**, 41: 51-69.

Panchaud A, Singh P, Shaffer SA, Goodlett DR. Xcomb: A cross-linked peptide database approach to protein−protein interaction analysis. *J. Proteome Res.* **2010**, 9: 2508-2515.

Paukštytė J, López Cabezas RM, Feng Y, Tong K, Schnyder D, Elomaa E, Gregorova P, Doudin M, Särkkä M, Sarameri J, Lippi A, Vihinen H, Juutila J, Nieminen A, Törönen P, Holm L, Jokitalo E, Krisko A, Huiskonen J, Sarin LP, Hietakangas V, Picotti P, Barral Y, Saarikangas J. Global analysis of aging-related protein structural changes uncovers enzyme-polymerization-based control of longevity. *Mol. Cell* **2023**, 83: 3360-3376.

Pelay-Gimeno M, Glas A, Koch O, Grossmann TN. Structure-based design of inhibitors of protein-protein interactions: Mimicking peptide binding epitopes. *Angew. Chem. Int. Ed.* **2015**, 54: 8896-8927.

Peri S, Steen H, Pandey A. Gpmaw – a software tool for analyzing proteins and peptides. *Trends Biochem. Sci* **2001**, 26: 687-689.

Petrotchenko EV, Olkhovik VK, Borchers CH. Isotopically coded cleavable cross-linker for studying protein-protein interaction and protein complexes. *Mol. Cell. Proteomics* **2005**, 4: 1167-1179.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. Ucsf chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, 25: 1605-1612.

Piazza I, Kochanowski K, Cappelletti V, Fuhrer T, Noor E, Sauer U, Picotti P. A map of protein-metabolite interactions reveals principles of chemical communication. *Cell* **2018**, 172: 358-372 e323.

Piazza I, Beaton N, Bruderer R, Knobloch T, Barbisan C, Chandat L, Sudau A, Siepe I, Rinner O, de Souza N, Picotti P, Reiter L. A machine learning-based chemoproteomic approach to identify drug targets and binding sites in complex proteomes. *Nat. Commun.* **2020**, 11: 4200.

Piersimoni L, Sinz A. Cross-linking/mass spectrometry at the crossroads. *Anal. Bioanal. Chem.* **2020**, 412: 5981-5987.

Piersimoni L, Kastritis PL, Arlt C, Sinz A. Cross-linking mass spectrometry for investigating protein conformations and protein–protein interactions—a method for all seasons. *Chem. Rev.* **2022a**, 122: 7500-7531.

Piersimoni L, Kastritis PL, Arlt C, Sinz A. Cross-linking mass spectrometry for investigating protein conformations and protein-protein interactions horizontal line a method for all seasons. *Chem. Rev.* **2022b**, 122: 7500-7531.

Pino LK, Searle BC, Bollinger JG, Nunn B, MacLean B, MacCoss MJJMsr. The skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. **2020**, 39: 229-244.

Pirklbauer GJ, Stieger CE, Matzinger M, Winkler S, Mechtler K, Dorfer V. Ms annika: A new cross-linking search engine. *J. Proteome Res.* **2021**, 20: 2560-2569.

Quast J-P, Schuster D, Picotti P. Protti: An r package for comprehensive data analysis of peptide- and protein-centric bottom-up proteomics data. *Bioinform. adv.* **2021**, 2: 1-3.

Raghunathan R, Turajane K, Wong LC. Biomarkers in neurodegenerative diseases: Proteomics spotlight on als and parkinson's disease. *Int. J. Mol. Sci.* **2022**, 23: 9299.

Rasmussen MI, Refsgaard JC, Peng L, Houen G, Hojrup P. Crosswork: Software-assisted identification of cross-linked peptides. *J. Proteomics* **2011**, 74: 1871-1883.

Reinhard FBM, Eberhard D, Werner T, Franken H, Childs D, Doce C, Savitski MF, Huber W, Bantscheff M, Savitski MM, Drewes G. Thermal proteome profiling monitors ligand interactions with cellular membrane proteins. *Nat. Methods* **2015**, 12: 1129-1131.

Rey M, Sarpe V, Burns KM, Buse J, Baker Charles AH, van Dijk M, Wordeman L, Bonvin Alexandre MJJ, Schriemer David C. Mass spec studio for integrative structural biology. *Structure* **2014**, 22: 1538-1548.

Rhee HW, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, Ting AY. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **2013**, 339: 1328-1331.

Rice P, Longden I, Bleasby A. Emboss: The european molecular biology open software suite. *Trends Genet.* **2000**, 16: 276-277.

Riffle M, Jaschob D, Zelter A, Davis TN. Proxl (protein cross-linking database): A platform for analysis, visualization, and sharing of protein cross-linking mass spectrometry data. *J. Proteome Res.* **2016**, 15: 2863-2870.

Riffle M, Jaschob D, Zelter A, Davis TN. Proxl (protein cross-linking database): A public server, qc tools, and other major updates. *J. Proteome Res.* **2019**, 18: 759-764.

Rivera-Santiago RF, Sriswasdi S, Harper SL, Speicher DW. Probing structures of large protein complexes using zero-length cross-linking. *Methods* **2015**, 89: 99-111.

Ross CA, Poirier MA. Protein aggregation and neurodegenerative disease. *Nat. Med.* **2004**, 10 Suppl: S10-S17.

Rudden LSP, Musson SC, Benesch JLP, Degiacomi MT. Biobox: A toolbox for biomolecular modelling. *Bioinformatics* **2022**, 38: 1149-1151.

Ryl PSJ, Bohlke-Schneider M, Lenz S, Fischer L, Budzinski L, Stuiver M, Mendes MML, Sinn L, O'Reilly FJ, Rappsilber J. In situ structural restraints from cross-linking mass spectrometry in human mitochondria. *J. Proteome Res.* **2020**, 19: 327-336.

Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. Plip: Fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* **2015**, 43: W443-447.

Schiffrin B, Radford SE, Brockwell DJ, Calabrese AN. <scp>pyxlinkviewer</scp> : A flexible tool for visualization of protein chemical crosslinking data within the <scp>pymol</scp> molecular graphics system. *Protein Sci.* **2020**, 29: 1851-1857.

Schilling B, Row RH, Gibson BW, Guo X, Young MM. Ms2assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am. Soc. Mass. Spectrom.* **2003**, 14: 834-850.

Schmidt C, Robinson CV. A comparative cross-linking strategy to probe conformational changes in protein complexes. *Nat. Protoc.* **2014**, 9: 2224-2236.

Schopper S, Kahraman A, Leuenberger P, Feng Y, Piazza I, Muller O, Boersema PJ, Picotti P. Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. *Nat. Protoc.* **2017**, 12: 2391-2410.

Seebacher J, Mallick P, Zhang N, Eddes JS, Aebersold R, Gelb MH. Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing. *J. Proteome Res.* **2006a**, 5: 2270-2282.

Seebacher J, Mallick P, Zhang N, Eddes JS, Aebersold R, Gelb MH. Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing - pubmed. *J. Proteome Res.* **2006b**, 5: 2270-2282.

Serohijos AW, Yin S, Ding F, Gauthier J, Gibson DG, Maixner W, Dokholyan NV, Diatchenko L. Structural basis for mu-opioid receptor binding and activation. *Structure* **2011**, 19: 1683-1690.

Shakeel S, Rajendra E, Alcon P, O'Reilly F, Chorev DS, Maslen S, Degliesposti G, Russo CJ, He S, Hill CH, Skehel JM, Scheres SHW, Patel KJ, Rappsilber J, Robinson CV, Passmore LA. Structure of the fanconi anaemia monoubiquitin ligase complex. *Nature* **2019**, 575: 234-237.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, 13: 2498-2504.

Shental-Bechor D, Levy Y. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105: 8256-8261.

Shiio Y, Aebersold R. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nat. Protoc.* **2006**, 1: 139-145.

Shuken SR, Rutledge J, Iram T, Losada PM, Wilson EN, Andreasson KI, Leib RD, Wyss-Coray T. Limited proteolysis–mass spectrometry reveals aging-associated changes in cerebrospinal fluid protein abundances and structures. *Nat. Aging* **2022**, 2: 379-388.

Sinz A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.* **2003**, 38: 1225-1237.

Sinz A. Divide and conquer: Cleavable cross-linkers to study protein conformation and protein-protein interactions. *Anal. Bioanal. Chem.* **2017**, 409: 33-44.

Skolnick J, Gao M, Zhou H, Singh S. Alphafold 2: Why it works and its implications for understanding the relationships of protein sequence, structure, and function. *J. Chem. Inf. Model.* **2021**, 61: 4827-4831.

Slavin M, Tayri-Wilk T, Milhem H, Kalisman N. Open search strategy for inferring the masses of cross-link adducts on proteins. *Anal. Chem.* **2020**, 92: 15899-15907.

Sljoka A, Probing allosteric mechanism with long-range rigidity transmission across protein networks. In *Allostery: Methods and protocols*, Di Paola, L; Giuliani, A, Eds. Springer US: New York, NY, 2021; pp 61-75.

Stahl K, Graziadei A, Dau T, Brock O, Rappsilber J. Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. *Nat. Biotechnol.* **2023**, 41: 1810-1819.

Steigenberger B, Pieters RJ, Heck AJR, Scheltema RA. Phox: An imac-enrichable cross-linking reagent. *Acs Central Sci.* **2019**, 5: 1514-1522.

Steigenberger B, van den Toorn HWP, Bijl E, Greisch JF, Rather O, Lubeck M, Pieters RJ, Heck AJR, Scheltema RA. Benefits of collisional cross section assisted precursor selection (caps-pasef) for cross-linking mass spectrometry. *Mol. Cell. Proteomics* **2020**, 19: 1677-1687.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering C. String v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, 47: D607–D613.

Tang X, Munske GR, Siems WF, Bruce JE. Mass spectrometry identifiable cross-linking strategy for studying protein−protein interactions. *Anal. Chem.* **2005a**, 77: 311-318.

Tang Y, Chen Y, Lichti CF, Hall RA, Raney KD, Jennings SF. Clpm: A cross-linked peptide mapping algorithm for mass spectrometric analysis. *Bmc Bioinformatics* **2005b**, 6: S9-S9.

Tarasova IA, Surin AK, Fornelli L, Pridatchenko ML, Suvorina MY, Gorshkov MV. Ion coalescence in fourier transform mass spectrometry: Should we worry about this in shotgun proteomics? *Eur. J. Mass Spectrom.* **2015**, 21: 459-470.

Taverner T, Hall NE, O'Hair RA, Simpson RJ. Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. *J. Biol. Chem.* **2002**, 277: 46487-46492.

Tayri-Wilk T, Slavin M, Zamel J, Blass A, Cohen S, Motzik A, Sun X, Shalev DE, Ram O, Kalisman N. Mass spectrometry reveals the chemistry of formaldehyde cross-linking in structured proteins. *Nat. Commun.* **2020**, 11.

Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, Hamon C. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by ms/ms. *Anal. Chem.* **2003**, 75: 1895-1904.

Tinnefeld V, Sickmann A, Ahrends R. Catch me if you can: Challenges and applications of cross-linking approaches. *Eur. J. Mass Spectrom.* **2014**, 20: 99-116.

Trabjerg E, Nazari ZE, Rand KD. Conformational analysis of complex protein states by hydrogen/deuterium exchange mass spectrometry (hdx-ms): Challenges and emerging solutions. *TrAC, Trends Anal. Chem.* **2018**, 106: 125-138.

Tracy TE, Madero-Perez J, Swaney DL, Chang TS, Moritz M, Konrad C, Ward ME, Stevenson E, Huttenhain R, Kauwe G, Mercedes M, Sweetland-Martin L, Chen X, Mok SA, Wong MY, Telpoukhovskaia M, Min SW, Wang C, Sohn PD, Martin J, Zhou Y, Luo W, Trojanowski JQ, Lee VMY, Gong S, Manfredi G, Coppola G, Krogan NJ, Geschwind DH, Gan L. Tau interactome maps synaptic and mitochondrial processes associated with neurodegeneration. *Cell* **2022**, 185: 712-728.

Uetrecht C, Rose RJ, van Duijn E, Lorenzen K, Heck AJ. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **2010**, 39: 1633-1655.

Van V, Brown JB, O'Shea CR, Rosenbach H, Mohamed I, Ejimogu NE, Bui TS, Szalai VA, Chacon KN, Span I, Zhang F, Smith AT. Iron-sulfur clusters are involved in post-translational arginylation. *Nat. Commun.* **2023**, 14: 458.

van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, Bonvin AMJJ. The haddock2.2 web server: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **2016**, 428: 720-725.

Vellucci D, Kao A, Kaake RM, Rychnovsky SD, Huang L. Selective enrichment and identification of azide-tagged cross-linked peptides using chemical ligation and mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **2010**, 21: 1432-1445.

Walzthoeni T, Joachimiak LA, Rosenberger G, Röst HL, Malmström L, Leitner A, Frydman J, Aebersold R, Walzthoeni T, Joachimiak LA, Rosenberger G, Röst HL, Malmström L, Leitner A, Frydman J, Aebersold R. Xtract: Software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat. Methods* **2015**, 12: pages1185–1190.

Wang B, Zhong X, Fields L, Lu H, Zhu Z, Li L. Structural proteomic profiling of cerebrospinal fluids to reveal novel conformational biomarkers for alzheimer's disease. *J. Am. Soc. Mass. Spectrom.* **2023**, 34: 459-471.

WARDEN SM, RICHARDSON C, O'DONNELL J, Jr, STAPLETON D, KEMP BE, WITTERS LA. Post-translational modifications of the β-1 subunit of amp-activated protein kinase affect enzyme activity and cellular localization. *Biochem. J* **2001**, 354: 275-283.

Wefing S, Schnaible V, Hoffmann D. Searchxlinks. A program for the identification of disulfide bonds in proteins from mass spectra. *Anal. Chem.* **2006**, 78: 1235-1241.

Wheat A, Yu C, Wang X, Burke AM, Chemmama IE, Kaake RM, Baker P, Rychnovsky SD, Yang J, Huang L. Protein interaction landscapes revealed by advanced in vivo cross-linking-mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, 118.

Wippel HH, Chavez JD, Keller AD, Bruce JE. Multiplexed isobaric quantitative cross-linking reveals drug-induced interactome changes in breast cancer cells. *Anal. Chem.* **2022**, 94: 2713-2722.

Xia L, Ma Z, Tong J, Tang Y, Li S, Qin S, Lou R, Zhao S, Lei X, Shui W. Evaluation of chemical cross-linkers for in-depth structural analysis of g protein-coupled receptors through cross-linking mass spectrometry. *Anal. Chim. Acta* **2020**, 1102: 53-62.

Xiang Z, Ren H, Hu YS, Coin I, Wei J, Cang H, Wang L. Adding an unnatural covalent bond to proteins through proximity-enhanced bioreactivity. *Nat. Methods* **2013**, 10: 885-888.

Xie Y, Chen S, Li Q, Sheng Y, Alvarez MR, Reyes J, Xu G, Solakyildirim K, Lebrilla CBJCs. Glycan–protein cross-linking mass spectrometry reveals sialic acid-mediated protein networks on cell surfaces. *Chem. Sci.* **2021**, 12: 8767-8777.

Xu H, Zhang L, Freitas MA. Identification and characterization of disulfide bonds in proteins and peptides from tandem ms data by use of the massmatrix ms/ms search engine. *J. Proteome Res.* **2008**, 7: 138-144.

Yang B, Wu Y-J, Zhu M, Fan S-B, Lin J, Zhang K, Li S, Chi H, Li Y-X, Chen H-F, Luo S-K, Ding Y-H, Wang L-H, Hao Z, Xiu L-Y, Chen S, Ye K, He S-M, Dong M-Q, Yang B, Wu Y-J, Zhu M, Fan S-B, Lin J, Zhang K, Li S, Chi H, Li Y-X, Chen H-F, Luo S-K, Ding Y-H, Wang L-H, Hao Z, Xiu L-Y, Chen S, Ye K, He S-M, Dong M-Q. Identification of cross-linked peptides from complex samples. *Nat. Methods* **2012**, 9: 904–906.

Yang Y, He M, Wei T, Sun J, Wu S, Gao T, Guo Z. Photo-affinity pulling down of low-affinity binding proteins mediated by post-translational modifications. *Anal. Chim. Acta* **2020**, 1107: 164-171.

Yang Z, Zeng X, Zhao Y, Chen R. Alphafold2 and its applications in the fields of biology and medicine. *Signal Transduct. Tar.* **2023**, 8: 115.

Yılmaz Ş, Drepper F, Hulstaert N, Černič M, Gevaert K, Economou A, Warscheid B, Martens L, Vandermarliere E. Xilmass: A new approach toward the identification of cross-linked peptides. *Anal. Chem.* **2016**, 88: 9949-9957.

Yılmaz Ş, Busch F, Nagaraj N, Cox J. Accurate and automated high-coverage identification of chemically cross-linked peptides with maxlynx. *Anal. Chem.* **2022**, 94: 1608-1617.

Yin H, Flynn AD. Drugging membrane protein interactions. *Annu. Rev. Biomed. Eng.* **2016**, 18: 51-76.

Yin P, Gao Y, Chen R, Liu W, He C, Hao J, Zhou M, Kan H. Temperature-related death burden of various neurodegenerative diseases under climate warming: A nationwide modelling study. *Nat. Commun.* **2023**, 14: 8236.

Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger G. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, 97: 5802-5806.

Yu C, Huszagh A, Viner R, Novitsky EJ, Rychnovsky SD, Huang L. Developing a multiplexed quantitative cross-linking mass spectrometry platform for comparative structural analysis of protein complexes. *Anal. Chem.* **2016**, 88: 10301-10308.

Yu C, Huang L. Cross-linking mass spectrometry: An emerging technology for interactomics and structural biology. *Anal. Chem.* **2018**, 90: 144-165.

Yu C, Wang X, Huang L. Developing a targeted quantitative strategy for sulfoxide-containing ms-cleavable cross-linked peptides to probe conformational dynamics of protein complexes. *Anal. Chem.* **2022**, 94: 4390-4398.

Yu F, Li N, Yu W. Exhaustively identifying cross-linked peptides with a linear computational complexity. *J. Proteome Res.* **2017**, 16: 3942-3952.

Yugandhar K, Wang T-Y, Leung AK-Y, Lanz MC, Motorykin I, Liang J, Shayhidin EE, Smolka MB, Zhang S, Yu H. Maxlinker: Proteome-wide cross-link identifications with high specificity and sensitivity. *Mol. Cell. Proteomics* **2020**, 19: 554-568.

Zhang MM, Beno BR, Huang RY, Adhikari J, Deyanova EG, Li J, Chen G, Gross ML. An integrated approach for determining a protein-protein binding interface in solution and an evaluation of hydrogen-deuterium exchange kinetics for adjudicating candidate docking models. *Anal. Chem.* **2019a**, 91: 15709-15717.

Zhang W, Gong P, Shan Y, Zhao L, Hu H, Wei Q, Liang Z, Liu C, Zhang L, Zhang Y. Spotlink enables sensitive and precise identification of site nonspecific cross-links at the proteome scale. *Brief. Bioinform.* **2022**, 23: bbac316.

Zhang Z, Deng X, Liu Y, Liu Y, Sun L, Chen F. Pkm2, function and expression and regulation. *Cell Biosci.* **2019b**, 9: 52.

Zheng Q, Zhang H, Wu S, Chen H. Probing protein 3d structures and conformational changes using electrochemistry-assisted isotope labeling cross-linking mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **2016**, 27: 864-875.

Zhou C, Dai S, Lai S, Lin Y, Zhang X, Li N, Yu W. Ecl 3.0: A sensitive peptide identification tool for cross-linking mass spectrometry data analysis. *Bmc Bioinformatics* **2023**, 24: 351-351.

Zhou JY, Hanfelt J, Peng J. Clinical proteomics in neurodegenerative diseases. *Proteom. Clin. Appl.* **2007**, 1: 1342-1350.

Zhu Y, Akkaya KC, Ruta J, Yokoyama N, Wang C, Ruwolt M, Lima DB, Lehmann M, Liu F. Cross-link assisted spatial proteomics to map sub-organelle proteomes and membrane protein topologies. *Nat. Commun.* **2024a**, 15: 3290.

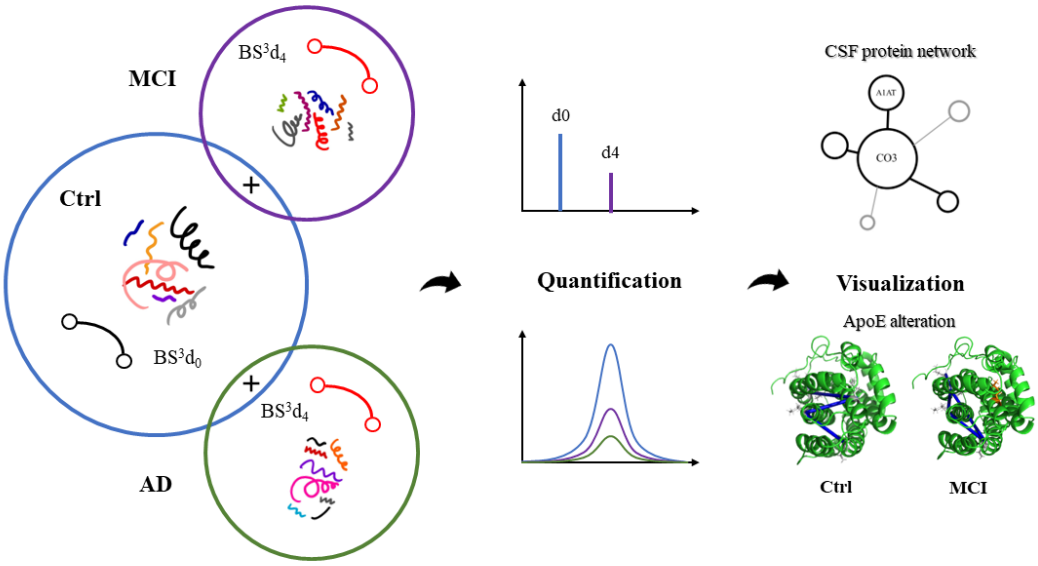
Zhu Z, Zhong X, Wang B, Lu H, Li L. Probing protein structural changes in alzheimer's disease via quantitative cross-linking mass spectrometry. *Anal. Chem.* **2024b**, 96: 7506-7515.

Ziemianowicz DS, Saltzberg D, Pells T, Crowder DA, Schräder C, Hepburn M, Sali A, Schriemer DC. Improv: A resource for cross-link-driven structure modeling that accommodates protein dynamics. *Mol. Cell. Proteomics* **2021**, 20: 100139-100139.

Zimnicka MM. Structural studies of supramolecular complexes and assemblies by ion mobility mass spectrometry. *Mass Spectrom. Rev.* **2023**: 1-34.

Chapter 5

Identification of cross-linked peptides in Alzheimer's disease by quantitative cross-linking mass spectrometry



Adapted from: **Z. Zhu**[#], X. Zhong[#], B. Wang[#], H. Lu, and L. Li. Probing Protein Structural Changes in Alzheimer's Disease Via Quantitative Cross-linking Mass Spectrometry. *Analytical Chemistry* **2024**, 96: 7506–7515. DOI: 10.1021/acs.analchem.4c00182. **Z. Zhu** contributed the experimental design, data collection, analysis, and manuscript preparation.

Abstract

Alzheimer's disease (AD) is a progressive neurological disorder, featured by abnormal protein aggregation in the brain, including the pathological hallmarks of amyloid plaques and hyperphosphorylated tau. Despite extensive research efforts, understanding the molecular intricacies driving AD development remains a formidable challenge. This study focuses on identifying key protein conformational changes associated with the progression in AD. To achieve this, we employed quantitative crosslinking mass spectrometry (XL-MS) to elucidate conformational changes in the protein networks in cerebrospinal fluid (CSF). By using isotopically labeled crosslinkers BS^3d_0 and BS^3d_4, we reveal a dynamic shift in protein interaction networks during AD progression. Our comprehensive analysis highlights distinct alterations in protein-protein interactions within Mild Cognitive Impairment (MCI) states. This study accentuates the potential of crosslinked peptides as indicators of AD-related conformational changes, including previously unreported site-specific binding between alpha-1-antitrypsin (A1AT) and Complement Component 3 (CO3). Furthermore, this work enables detailed structural characterization of Apolipoprotein E (ApoE) and reveals modifications within its helical domains, suggesting their involvement in MCI pathogenesis. The quantitative approach provides insights into site-specific interactions and changes in the abundance of crosslinked peptides, offering an improved understanding of the intricate protein-protein interactions underlying AD progression. These findings lay a foundation for the development of potential diagnostic or therapeutic strategies aimed at mitigating the negative impact of AD.

Keywords: crosslinking mass spectrometry (XL-MS), quantitative proteomics, Alzheimer's disease, protein-protein interaction, conformational changes, neurodegenerative disease

Introduction

As one of the most prominent public health concerns, Alzheimer's disease (AD) has affected more than 6.7 million people in the US and is ranked as the fifth leading cause of death among those aged 65 or older[1]. Rooted in synaptic homeostasis disruption and intricate signaling pathway dysfunction, AD is marked by progressive cognitive impairment. During the course of AD progression, mild cognitive impairment (MCI) has been reported as an intermediate state and about 15% of patients are diagnosed as MCI patients before AD dementia. The MCI stage manifests varied cognitive deficits, including memory loss, attention deficits, and language impairment[2]. Although the major pathological theory of amyloid beta (Aβ) fibrillation and plaque formation have been extensively documented, recent research suggests that Aβ modification only involves certain heterogeneous pathways and therapeutic advancements based on Aβ have been frustratingly slow[3-5]. In parallel, Tau phosphorylation has been implicated in the formation of the neurofibrillary tangles (NFTs) formation, which contributes to the collapse of cytoarchitecture and subsequent neuronal degradation[6]. Despite the significant attention to Tau topology, clinical progress is hampered by concerns regarding efficacy and toxicity[7]. Thus, the identification of alternative biomarker candidates is imperative for precision AD therapeutics. In particular, the transition from MCI to full-blown AD represents a critical juncture in AD's trajectory, underscoring the urgent need for effective interventions during this intermediate phase.

Mass spectrometry (MS) has emerged as an exceptional tool for the identification of protein structures and the characterization of conformational networks in recent decades. Among the various branches of MS techniques, crosslinking mass spectrometry (XL-MS) is an emerging approach that provides insights into structural dynamics[8, 9]. It involves the use of a chemical

reagent known as crosslinker, which forms new covalent bonds between specific amino acids. In general, crosslinkers consist of two reactive groups at the ends and a spacer arm bridging them. Each crosslinker has a defined distance threshold, within which only residues that are spatially proximate can be crosslinked[10]. When the data of crosslinked sites are obtained through MS acquisition, the compilation of the result can reveal structural information about the protein complex or the topology of the protein network. The primary advantage of XL-MS is the crosslinked peptides are stable and can withstand a wide range of buffer conditions[11]. Therefore, the application of the XL-MS is versatile, ranging from conformational changes in single protein or complex to *in vivo* systematic study of protein-protein network[12]. In addition, quantitative XL-MS (QXL-MS) enables direct comparison of crosslinking in different states of flexible regions. The incorporation of isotopically labeled crosslinkers provide extra separation between targeted peaks and allow for the detection of conformational changes through abundance quantification[13]. Numerous successful studies have demonstrated the elucidation of protein interactomes via chemical crosslinking strategy[14-16].

This study seeks to identify potential protein factors that can serve as indicators of AD prognosis. Hereby, we investigate the conformational changes in protein networks during AD development using human cerebrospinal fluids (CSF) samples. Leveraging the quantitative XL-MS (QXL-MS), our research provides critical insights into the dynamic shifts in protein-protein interaction networks as AD progresses. By visualizing the data, we aim to shed light on alternative biomarker candidates, thereby providing a hypothetical framework that elucidates the site-specific binding within the intricate protein structures. These new findings will facilitate the development of more precise diagnostic tools and therapeutic strategies for AD and related dementia.

Experimental Section

AD CSF Samples and Chemical Reagents

All human subjects involved in this study have received approval from the University of Wisconsin Institutional Review Board. The study comprised of 48 participants from the Wisconsin Alzheimer's Disease Research Center (ADRC), who were individually selected from healthy controls, MCI, and AD groups (**Supplemental Table S1**).

CSF Protein Crosslinking

The detailed workflow of the overall processing experiments is shown in **Figure 1**. The protein concentration of the CSF samples was determined via BCA protein assay kit from Thermo (Pittsburgh, PA). Approximately 100 µg protein of from each sample was used for the crosslinking incubation. The BS3 (bis[sulfosuccinimidyl] suberate) crosslinkers were dissolved in water to obtain a stock concentration of 50 mM and the final reaction concentration was 1 mM. The control sample was incubated with BS^3d_0, whereas the MCI or AD sample was incubated with BS^3d_4 (deuterium-labeled isotopomer) in 20 mM HEPES buffer for 1 hour (**Figure S1**) [17, 18]. The crosslinking reaction was terminated with 20 mM Tris buffer for 20 min. BS^3d_0 and BS^3d_4 cross-linked samples were mixed with a 1:1 ratio and subjected to Albumin Depletion Kit purchased from Pierce (Waltham, MA) to remove albumin.

Tryptic Proteolysis

The samples were digested with a filter-aided sample preparation (FASP) procedure[19]. In brief, the crosslinked samples were washed three times with 8 M urea using a 30 kDa molecular weight cutoff filter (Millipore). The samples were reduced with 20 mM dithiothreitol (DTT) for 45

minutes, alkylated with 20 mM iodoacetamide (IAA) in the dark for 45 minutes, and digested with trypsin (enzyme : protein = 1 : 40) from Promega (Madison, WI) overnight at 37°C. The digestion process is quenched with 10% Trifluoroacetic Acid (TFA) to reach a condition pH < 3. For in-solution digestion, the crosslinked protein sample was mixed with 8M urea at a 1:2 (vol/vol) ratio. For enhanced FASP (eFASP) , the samples were washed with 8M urea in 0.1% deoxycholic acid (DCA) from Sigma-Aldrich (St. Louis, MO)[20, 21]. All the other following steps were the same as described in FASP. The peptides were subsequently desalted through a Bond Elut OMIX C18 pipet tips with 80% ACN and 0.1% Formic Acid (FA) elution. The eluted sample was dried down in SpeedVac Vacuum Concentrator before MS acquisition.

LC-MS/MS analysis

Samples were run through Thermo Fusion Lumos Orbitrap Tribrid mass spectrometer. The sample was redissolved in 97% Optima water, 3% ACN, and 0.1% FA, and loaded to the Dionex UltiMate 3000 UPLC system. The separation was done via a homemade microcapillary column (75-μm inner diameter) packed with 18 cm of Bridged Ethylene Hybrid C18 particles (1.7 μm, 130 Å, Waters). LC separation is achieved via 100 minutes of gradient elution of 3% mobile phase B (pure ACN with 0.1% FA) to 95%, with a flow rate of 300 nL/min. The acquisition method is set under high-collisional dissociation (HCD) with a resolving power of 120k, scanning range from 400-1600 *m/z*, and maximum injection time of 50 ms. The data dependent mode was set to 3 seconds between master scans with a normalized collision energy of 30. MS2 acquisition was performed with an isolation window of 2 Da, resolving power of 30k, and standard AGC target.

Quantitative Structural Analysis and Visualization

Crosslink signals in the raw files were searched via Xisearch (version 1.7.6.7) against the UniProt *Homo sapiens* reviewed database. Both BS3 light (BS^3d_0) and BS3 heavy (BS^3d_4) were selected for searching[13, 22, 23]. For the identification parameters, trypsin was selected as enzyme digestion, and maximum 3 of miscleavage was chosen. Precursor tolerance is 20 ppm and False Discovery Rate (FDR) is controlled at < 5% at PSM level. To increase the efficiency of the algorithm running and filter the low-scored crosslinks interference, we only selected top 20 proteins from human proteome FASTA file that has most available crosslinks (**Table S2**). The quantitation work was performed through Skyline (version 22.2.0) with the library file in .ssl version imported from Xisearch[24]. All decoy crosslinks are excluded prior to the Skyline analysis. The dot product value (dotp) was set to above 0.9 and the peak area comparison of each target in the library list was manually checked to ensure more than 50% of the replicate in each set reached the dotp line. One set of BS^3d_0 (Ctrl group) and both sets of BS^3d_4 (MCI and AD) group were selected for the subsequent analysis and Skyline library construction. The peptide set remained consistent with the previous crosslinking search, while in the transition setting, the precursor charges ranged from 3 to 7. The fold change values of AD:Ctrl and MCI:Ctrl were calculated based on the quantitation result output via Skyline, and the p-value was derived from the analysis of the 16 replicates. The threshold for the p-value is determined over 0.05 and the fold change is equal to 1. The volcano plot is depicted to highlight the upregulation and downregulation of the crosslinked peptides. To visualize our data, including bar charts, bubble charts, and volcano plots, we used HiPlot websites and Python programming[25]. The .mgf file converted from raw data, FASTA file, and .mzML file from xiSearch was uploaded to xiView for 2D protein-protein network mapping[26]. For the 3D model for the protein structures, we employed PyMOL and ChimeraX 1.3[25, 27, 28]. Additionally, for

Gene Ontology analysis, the DAVID bioinformatics resources were applied with highlight of biological process, cellular component and molecular function[29].

Results and Discussion

Crosslinking MS reveals conformational variation within CSF protein networks during AD progression

Prior to the investigation of all human CSF samples, we compared the in-solution, eFASP, and FASP digestion methods to assess their compatibility with crosslinking proteins. Using BS^3d_0 labeled CSF from one control sample, the numbers of crosslinked spectral matches being identified were 126, 218, and 254, and the numbers of peptides were 80, 85, 109 in in-solution, eFASP, and FASP digestions methods, respectively. The performance of FASP and eFASP were comparable, but both outperformed in-solution digestion (**Figure S1**). Consequently, we opted to employ FASP for subsequent sample preparation.

Furthermore, the presence of albumin in substantial quantities within the CSF samples posed a significant challenge. Albumin is the most abundant protein in human plasma because of its high concentration and large molecular weight[30, 31]. It is rich in Lysine residues, which are specifically targeted by our selected crosslinkers. Both BS^3d_0 and BS^3d_4 are homobifunctional crosslinkers with spacer arms of the same length (11.4 Å) and sulfo-NHS reactive groups on both ends, which attach to the primary amine groups on Lysine residues and the N-termini of peptides. We observed that over 40% of the crosslinks were type-1 intralinks within the albumin itself. This raised concerns that without the removal of albumin, the presence of albumin-derived crosslinking peptides could obscure the detection of crosslinks involving other CSF proteins. Herein, we

subjected the CSF samples through albumin depletion resin to generate a more comprehensive view of interlinks among all potential protein biomarkers. The number of unique interlink matchups between different proteins after albumin depletion has been doubled in comparison to the same sample replicate without albumin depletion (**Figure S3**).

The analysis of our data revealed a noteworthy pattern: a significant decrease in the overall number of crosslinked spectral matches and unique crosslinked peptides from the control (Ctrl) group to MCI and further to the AD condition (**Figure 2a**). In the summary of the overall 16 sets of sample replicates, the Ctrl group exhibited almost double the number of discovered crosslinked spectral matches and unique crosslinked peptides compared to the AD group. Notably, the substantial proportion of observations within individual sets underscores the complexity of the conditions under investigation. This phenomenon was most pronounced in the Ctrl condition, where 66.5% of the XL matches were unique, while in the AD and MCI conditions, less than 50% of the matches were retained (**Figure 2b**). In terms of intersectional crosslinking in overlapped area, the greatest number of interlinks shared between Ctrl and MCI groups was observed, whereas the Ctrl-AD set exhibited the fewest preserved crosslinks. These findings highlight the significant alterations in crosslinking patterns between healthy controls and the AD condition, with MCI serving as an intermediate stage preserving a substantial number of interlinks from both conditions.

The identified crosslinked proteins were used for protein-protein interaction (PPI) analysis, which includes majority of plasma proteins: apolipoprotein (APOA1 and APOE), transferrin (TRFE), alpha-1-antitrypsin (A1AT), alpha-2-microglobulin (A2MG), complement components (CO3, CO5, CFAB), alpha-1-antichymotrypsin (AACT), antithrombin (ANT3), vitamin-D binding protein (VTDB) (**Figure 2c)**. We identified several representative biomarkers that are pertinent to Alzheimer's disease (AD) development. One such biomarker is Clusterin (CLUS), whose

conformational changes in AD have been previously reported in our earlier research[32]. To assess the reliability of our crosslinking map, we compared the crosslinking PPI mapping with the previously validated colocalization or co-expression traits using these successfully identified crosslinked proteins from the STRING database[33]. We manually input these crosslinked proteins and compared their PPI interaction shape with our xiView output (**Figure 2d-f**). As a result, more than 85% of the crosslinks and PPI have been previously confirmed by other studies or publications[34-36]. For instance, the direct interaction between transferrin and antithrombin has been verified in the previous publication via surface plasmon resonance (SPR) analysis, whereas a significant number of TRFE-ANT3 interlinks has been reported in both Ctrl and MCI group in our experiment[35]. This validation suggests that our crosslinking strategy delineates a genuine protein-protein interaction network in the selected CSF samples. From the visualization of the networks, xiView mapping also indicates that most of the PPI are centered around TRFE and complement component 3 (CO3) across all three different cases. The cluster of the interlinking between different proteins on these two anchors is even greater in MCI but in the AD group this TRFE-centered correlation is restrained because of fewer crosslinking matches and PPI. These findings strongly suggest that the protein interaction network is greatly affected during AD progression.

Quantitative analysis highlights distinctive site-specific protein structural changes as potential biomarkers for AD

While numerous publications have reported the CSF proteomic and corresponding pathological alterations during AD progression, the structural information within the protein complex or docking domains remains elusive[37-39]. Hence, our quantitative XL-MS workflow, executed through Skyline, presents a novel strategy highlighting site-specific interaction between CSF biomarkers. The crosslinking data were processed according to the methodology outlined in the Experimental

Section. The volcano plot serves as a rapid means of identifying crosslink abundance alterations across multiple sample replicates (**Figure 3a-b**). In the comparison between AD and Ctrl, nine crosslinks exhibited higher abundance in the AD group, while seven crosslinks were retained mostly in Ctrl group. Notably, the numbers shifted to 27 up- and 12 down-regulated crosslinks in the MCI-Ctrl comparison, highlighting the dynamic conformational changes in protein-protein interactions during disease progression from MCI to AD.

Among these significantly upregulated or downregulated crosslinks, many involve interlinks between diverse biomarkers, but the proportion of type 2 interlinks between different proteins is less than 40% of the overall crosslinks. Particularly noteworthy are the crosslink pairs that relate to CO3 and A1AT. In recent years, extensive research has focused on the interaction between CO3 and A1AT. Dysregulated complement activation has been demonstrated to have a strong impact on tau pathology and neurological damage[40-43]. A1AT may form covalent-binding aggregates along with VDTB and induce formation of the neutrophil[44, 45]. Meanwhile, this relative VDTB interaction is also successfully identified in the crosslinking network (**Figure 2d-e**). However, the nature of the aggregates between CO3 and A1AT remains poorly understood. In 2021, the structural insights into the complement complex were validated through *in vivo* crosslinking, supported by X-ray crystallography[46]. Our crosslink result proposes a hypothetical model detailing the site-specific binding between A1AT and CO3 (**Table S3**). Quantitative findings reveal a high degree of activity for A1AT at K246 in crosslinking with CO3 Lysine residues. This crosslink accounts for approximately 70% of the unique crosslinking matches, however, the composition of its crosslinking is very different from each other (**Table S4**). The crosslinks between A1AT K246 and CO3 K685 were identified in both Ctrl and MCI groups, but the intensity in MCI was significantly lower than that in Ctrl (**Figure 3c**). Conversely, the A1AT K246 – CO3 K428

crosslink demonstrated a preference towards MCI and AD cases, with the abundance in AD almost twice as high as in MCI. This indicates a dynamic change in the interaction between these two proteins during the progression of AD (**Figure 3d)**. It is noteworthy that these two crosslinked residues are located on two different chains, albeit in close proximity to each other in protein crystallography: K428 is on the conserved macroglobulin domain and K685 is on the anaphylatoxin (ANA) domain, which is the key of C3a fragment during cleavage[47] (**Figure 3e**). These findings further support the stepwise alteration in activation and subsequent signaling cascades throughout MCI and AD development.

To delve deeper into the functions that these crosslinked proteins are involved, we conducted both Gene Oncology Enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses (**Figure 4a, 4b**). These analyses revealed associations with serine-type endopeptidase inhibitor activity, the complement and coagulation cascade, and AD-specific terms, including the regulation of beta-amyloid clearance, tau protein binding, and the regulation of platelet or fibril formation. In summary, our findings demonstrate the involvement of protein biomarkers that are particularly pertinent to the progression of AD.

Intralinks reveals potential structural alteration in ApoE during MCI development

Apolipoprotein E (ApoE) plays a pivotal role in AD as a risk factor for neuropathology and pathogenesis of AD[48]. Within the CSF, ApoE serves as a critical regulator in multiple signaling pathways and the maintenance of the central nervous system (CNS)[49]. It has three main alleles E2, E3 and E4, which differed primarily at positions 112 and 158 in the amino acid sequence[50]. These structural variations between isoforms, especially in the major helix region on the C-terminal domain, are known to influence overall protein stabilization and functionality in the AD[50-52]. Consequently, our focus centers on identifying and examining variations in crosslinks within ApoE

during the development of MCI. Intralinks analysis provides insights indicating that specific interactions within the helix structures are retained in both conditions: ApoE K90-K113 and ApoE K251-K260. The K251-K260 interaction occurs within the C-terminal regions, which is implicated in oligomer formation and amyloid beta (Aβ) binding[37]. Our quantitative findings highlight the consistency in the abundance of this linkage across multiple sample sets for comparison (**Figure 5a**), suggesting the potential pathological binding between ApoE and heparan sulfate (HS) remains unaffected under MCI stage[53]. The underlying hypothesis that HS-assisted accumulation and induction with tau protein is not supported[54, 55]. Concurrently, differences in crosslink patterns between the control and MCI conditions become evident. In the control group, K175 exhibits robust crosslinking, seen in K113-K175 and K175-K260 interactions across multiple replicates. In contrast, these crosslinks are not discernible in the MCI condition through xiSearch, due to a pronounced decrease in abundance (**Figure 5b)**. Instead, alternate crosslinking sites remain active in the control group, including interactions like K90-K260 and K113-K260, detected across various samples. These observations collectively suggest a substantial structural transformation in ApoE during the course of MCI progression. Lysine 175, located within helix 4, a highly charged domain containing the low-density lipoprotein receptor (LDLR) binding site, emerges as a focal point of this conformational shift[51]. Our intralink analysis reveals that the conformational change at Lysine 175 may disrupt the interaction with helix 4 in ApoE (**Figure 5c-d).** This disruption could potentially contribute to abnormal binding processes, ultimately leading to amyloid deposition[56]. The alteration in the conformation of this region could also impact receptor interactions, potentially affecting cellular uptake and clearance mechanisms. Future work will be directed to substantiate our hypothesis using computational platforms such as HADDOCK[57]. Additionally, we aim to extract the ApoE protein for more precise 3D remodeling during AD progression.

Conclusion

This study employs XL-MS to investigate the intricate structural changes within CSF protein networks during the progression of MCI and AD. The comprehensive analysis reveals a dynamic shift in protein interaction networks as the pathological condition advances. The alterations in protein-protein interactions provide crucial insights into the proteomic changes associated with AD development. Furthermore, our investigation delves into the structural alterations of key biomarkers, with a particular focus on the interaction between A1AT and CO3, as well as the conformational changes within ApoE, shedding light on their potential involvement in AD pathogenesis. The observed site-specific interactions and changes in crosslink abundance offer a detailed understanding of the molecular intricacies underlying AD progression. The insights gained from this research not only open new avenues for the study of AD, but also provide potential directions for advancing our knowledge of neurodegenerative diseases and developing effective therapeutic strategies.

Data Availability

The mass spectrometry raw data, peak list and the mzid files has been uploaded through Mass Spectrometry Interactive Virtual Environment (MassIVE) and deposited in ProteomeXchange. The accession number is MSV000094259 for MassIVE and PXD050457 for ProteomeXchange.

Acknowledgements

This research was supported, in part, by the National Institutes of Health Grants R21AG065728, R01AG052324, R01AG078794, and R01DK071801. Some of the mass spectrometers were acquired using NIH shared instrument grants S10 OD028473, S10 RR029531, and S10 OD025084. H.L. wishes to thank the funding support for a Postdoctoral Career Development Award provided by the American Society for Mass Spectrometry. L.L. acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison, School of Pharmacy.

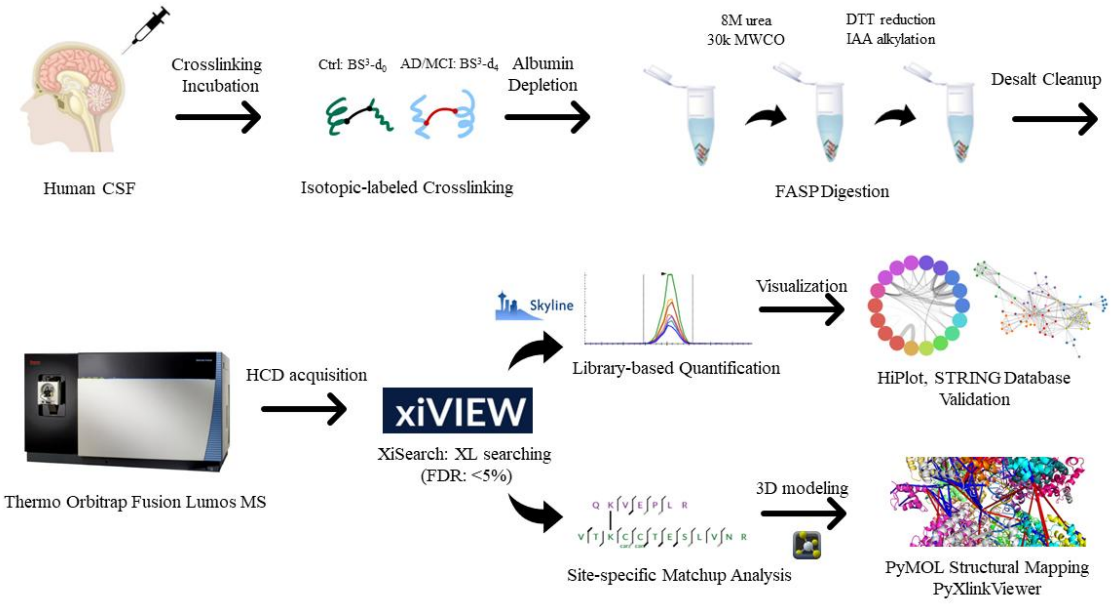


Figure 1. The workflow of the quantitative crosslinking MS for analysis of the protein structural changes and protein-protein network conformational changes in human CSF samples.

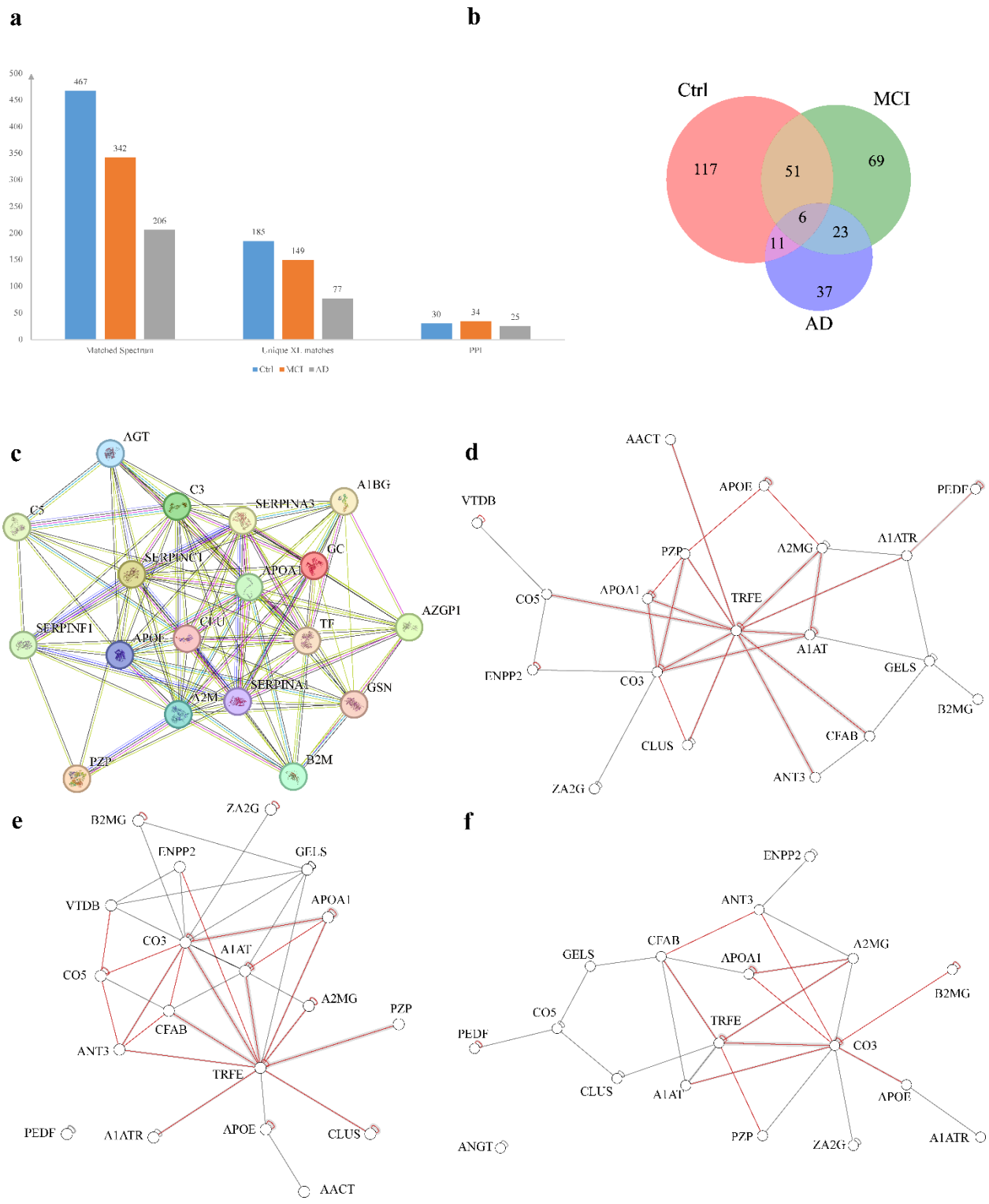


Figure 2. The experimental outcome of the crosslinking MS. (a) The comparison of the number of matched spectra, unique crosslinking matches and PPI for Ctrl, MCI and AD sample groups under XiSearch searching. (b) Venn diagram for the specific crosslinking matches in each condition. (c) The STRING database interaction diagram of the high-confident candidates. (d)

Visualization mapping of the pairwise unique crosslinking matches via xiView in Ctrl group and (e) MCI and (f) AD group. The crosslinking between the proteins with more than 3 unique site-specific matched pairs are highlighted.

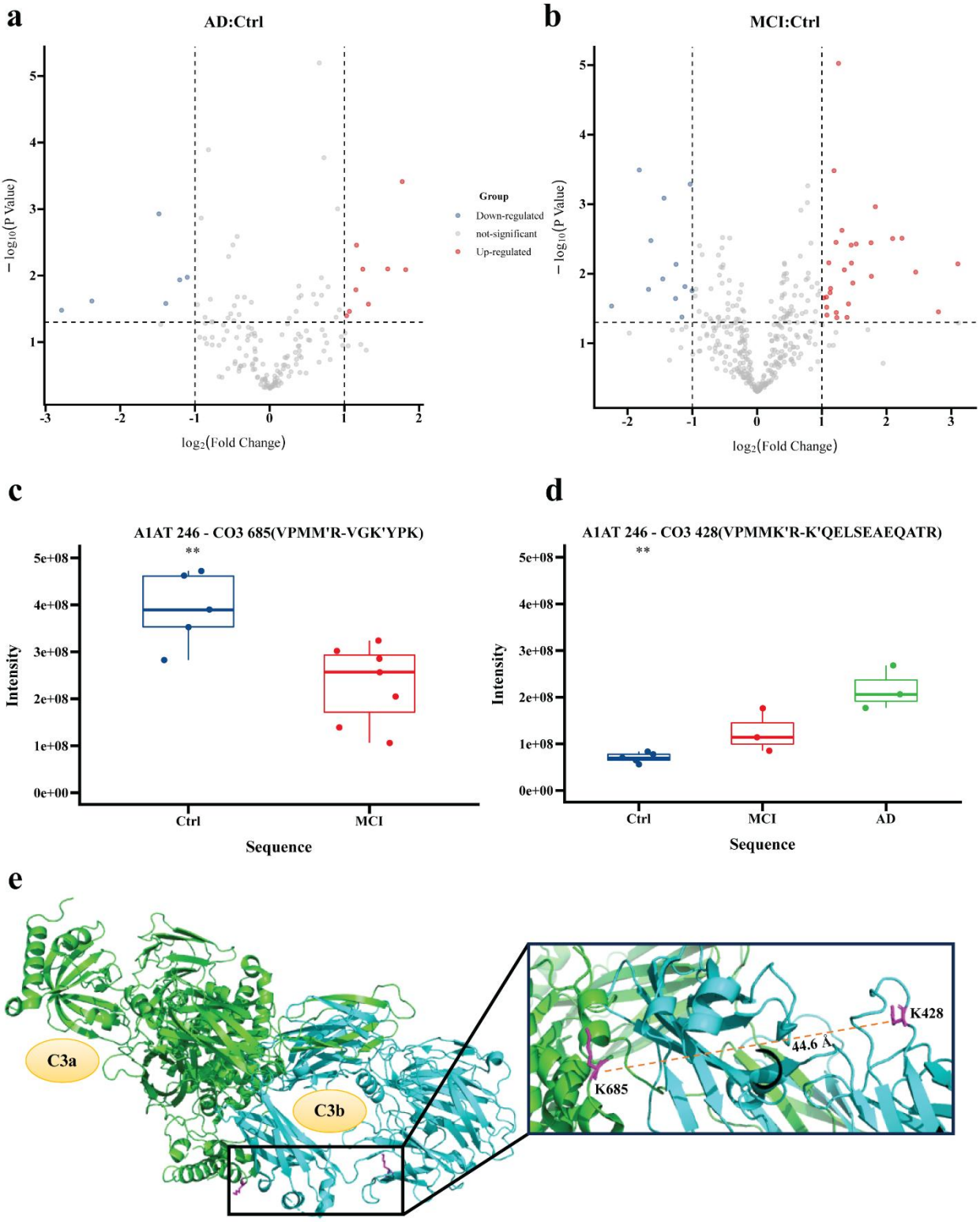


Figure 3. Quantitative crosslink matchups derived from Skyline. (a) the volcano plot of the

quantitative crosslinking pairs from Skyline in the AD:Ctrl set and (b) in the MCI:Ctrl set. In the case study of interlink between CO3 and A1AT, the abundance changes between (c) A1AT K246 and CO3 K685 and (d) A1AT K246 and CO3 K428 are provided. (e) The crystal structure of C3 (PDB: 2A73) is labeled with K685 and K428 in purple.

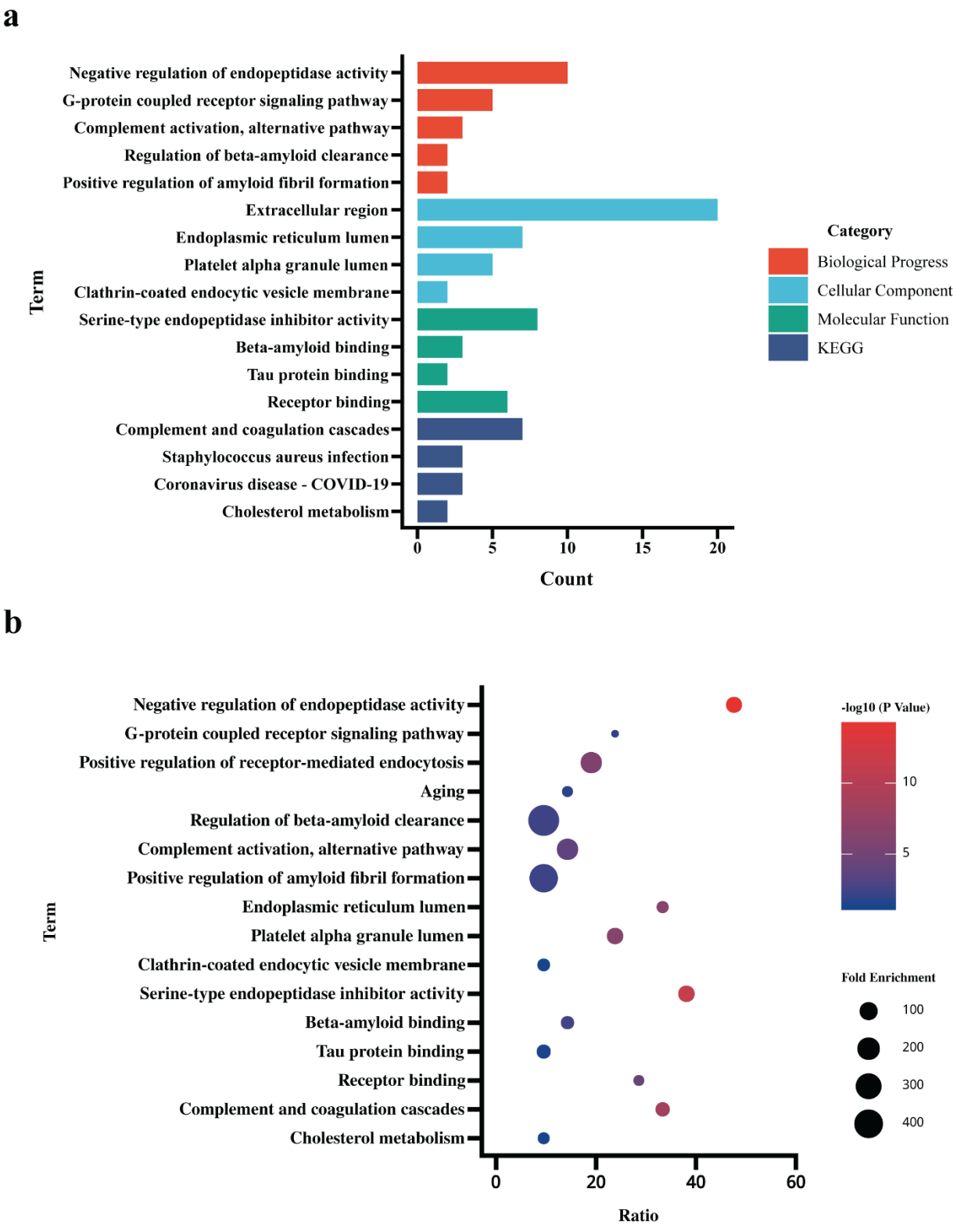


Figure 4. GO enrichment and KEGG pathway analysis based on identified crosslinked

biomarkers. (a) Bar chart of GO enrichment and KEGG pathway analysis of the significantly modified proteins during AD development. (b) Bubble chart of AD-related terms associated with crosslink represented structural changes.

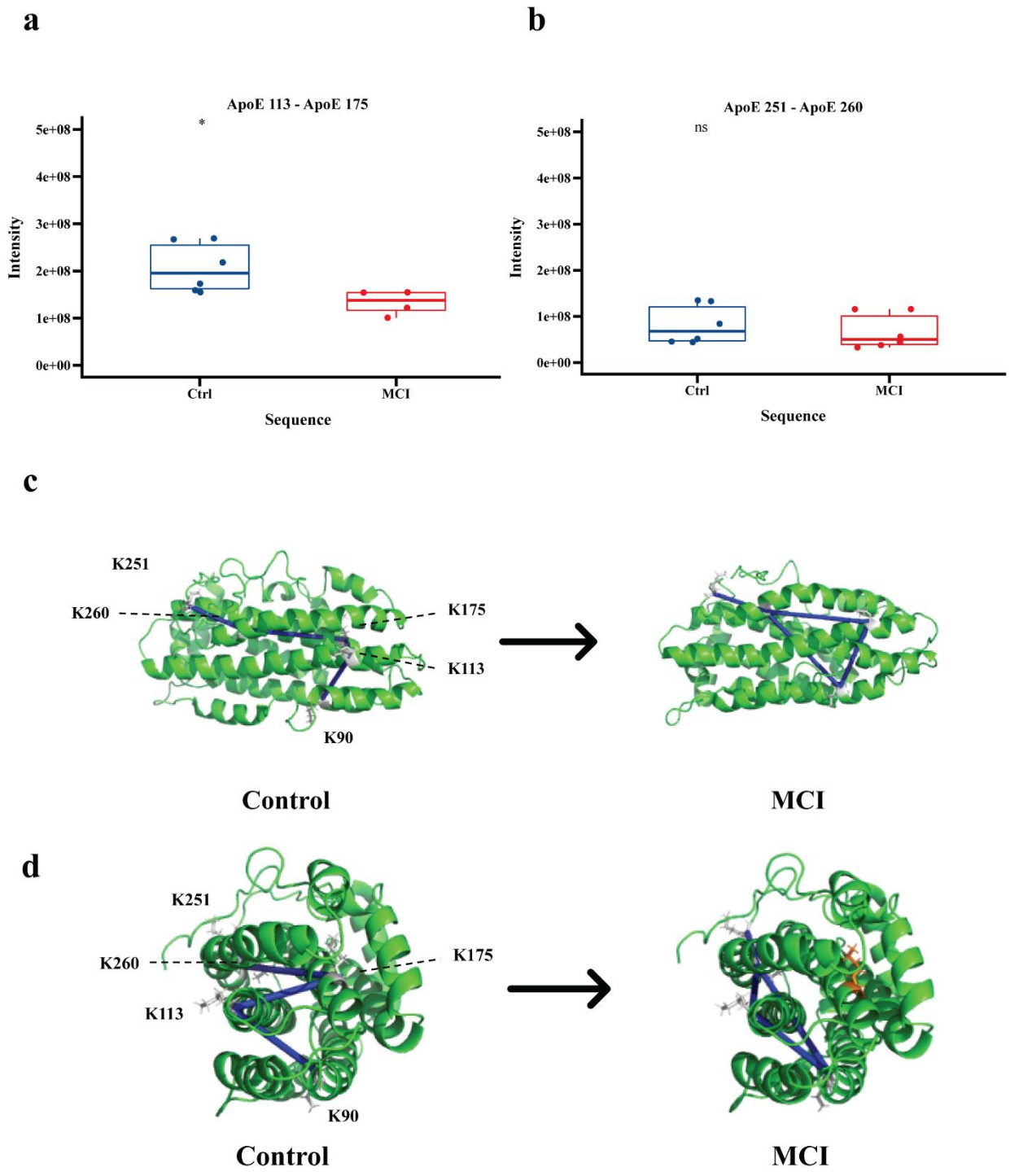


Figure 5. Intralinks within ApoE reveals a conformational change during MCI progression. Quantitative comparison of the abundance between healthy control and MCI sample in (a) APOE K251-K260 and (b) K113-K175. The overview (c) and sideview (d) of the demonstration of crosslink within PyMOL model (PDB: 2L7B), as K175 highlighted in orange.

References

1. 2023 Alzheimer's disease facts and figures. *Alzheimers Dement* **2023**.
2. Farias, S. T.; Mungas, D.; Reed, B. R.; Harvey, D.; DeCarli, C., Progression of mild cognitive impairment to dementia in clinic- vs community-based cohorts. *Arch Neurol* **2009,** *66* (9), 1151-7.
3. Mullane, K.; Williams, M., Alzheimer's disease (AD) therapeutics - 1: Repeated clinical failures continue to question the amyloid hypothesis of AD and the current understanding of AD causality. *Biochem Pharmacol* **2018,** *158*, 359-375.
4. Long, J. M.; Holtzman, D. M., Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. *Cell* **2019,** *179* (2), 312-339.
5. Chiti, F.; Dobson, C. M., Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **2006,** *75*, 333-66.
6. Xia, Y.; Prokop, S.; Giasson, B. I., "Don't Phos Over Tau": recent developments in clinical biomarkers and therapies targeting tau phosphorylation in Alzheimer's disease and other tauopathies. *Mol Neurodegener* **2021,** *16* (1), 37.
7. Congdon, E. E.; Sigurdsson, E. M., Tau-targeting therapies for Alzheimer disease. *Nat Rev Neurol* **2018,** *14* (7), 399-415.
8. Yu, C.; Huang, L., Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal Chem* **2018,** *90* (1), 144-165.
9. Liu, F.; Heck, A. J., Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr Opin Struct Biol* **2015,** *35*, 100-8.
10. Piersimoni, L.; Kastritis, P. L.; Arlt, C.; Sinz, A., Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions horizontal line A Method for All Seasons. *Chem Rev* **2022,** *122* (8), 7500-7531.
11. O'Reilly, F. J.; Rappsilber, J., Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat Struct Mol Biol* **2018,** *25* (11), 1000-1008.
12. Matzinger, M.; Mechtler, K., Cleavable Cross-Linkers and Mass Spectrometry for the Ultimate Task of Profiling Protein-Protein Interaction Networks in Vivo. *J Proteome Res* **2021,** *20* (1), 78-93.
13. Chen, Z. A.; Rappsilber, J., Quantitative cross-linking/mass spectrometry to elucidate structural changes in proteins and their complexes. *Nat Protoc* **2019,** *14* (1), 171-201.
14. Gomez-Ramirez, J.; Wu, J., Network-based biomarkers in Alzheimer's disease: review and future directions. *Front Aging Neurosci* **2014,** *6*, 12.
15. Rey, M.; Dhenin, J.; Kong, Y.; Nouchikian, L.; Filella, I.; Duchateau, M.; Dupre, M.; Pellarin, R.; Dumenil, G.; Chamot-Rooke, J., Advanced In Vivo Cross-Linking Mass Spectrometry Platform to Characterize Proteome-Wide Protein Interactions. *Anal Chem* **2021,** *93* (9), 4166-4174.
16. Wheat, A.; Yu, C.; Wang, X.; Burke, A. M.; Chemmama, I. E.; Kaake, R. M.; Baker, P.; Rychnovsky, S. D.; Yang, J.; Huang, L., Protein interaction landscapes revealed by advanced in vivo cross-linking-mass spectrometry. *Proc Natl Acad Sci U S A* **2021,** *118* (32), No. e2023360118.
17. Muller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O., Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Anal Chem* **2001,** *73* (9), 1927-34.
18. Sinz, A., Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J Mass Spectrom* **2003,** *38* (12), 1225-37.

19. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009,** *6* (5), 359-62.
20. Erde, J.; Loo, R. R.; Loo, J. A., Improving Proteome Coverage and Sample Recovery with Enhanced FASP (eFASP) for Quantitative Proteomic Experiments. *Methods Mol Biol* **2017,** *1550*, 11-18.
21. Lu, L.; Millikin, R. J.; Solntsev, S. K.; Rolfs, Z.; Scalf, M.; Shortreed, M. R.; Smith, L. M., Identification of MS-Cleavable and Noncleavable Chemically Cross-Linked Peptides with MetaMorpheus. *J Proteome Res* **2018,** *17* (7), 2370-2376.
22. Mendes, M. L.; Fischer, L.; Chen, Z. A.; Barbon, M.; O'Reilly, F. J.; Giese, S. H.; Bohlke-Schneider, M.; Belsom, A.; Dau, T.; Combe, C. W.; Graham, M.; Eisele, M. R.; Baumeister, W.; Speck, C.; Rappsilber, J., An integrated workflow for crosslinking mass spectrometry. *Mol Syst Biol* **2019,** *15* (9), e8994.
23. UniProt, C., UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **2023,** *51* (D1), D523-D531.
24. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010,** *26* (7), 966-8.
25. Li, J.; Miao, B.; Wang, S.; Dong, W.; Xu, H.; Si, C.; Wang, W.; Duan, S.; Lou, J.; Bao, Z.; Zeng, H.; Yang, Z.; Cheng, W.; Zhao, F.; Zeng, J.; Liu, X. S.; Wu, R.; Shen, Y.; Chen, Z.; Chen, S.; Wang, M.; Hiplot, C., Hiplot: a comprehensive and easy-to-use web service for boosting publication-ready biomedical data visualization. *Brief Bioinform* **2022,** *23* (4), No. bbac261.
26. Combe, C. W.; Fischer, L.; Rappsilber, J., xiNET: cross-link network maps with residue resolution. *Mol Cell Proteomics* **2015,** *14* (4), 1137-47.
27. Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E., UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **2018,** *27* (1), 14-25.
28. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **2021,** *30* (1), 70-82.
29. Sherman, B. T.; Hao, M.; Qiu, J.; Jiao, X.; Baseler, M. W.; Lane, H. C.; Imamichi, T.; Chang, W., DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* **2022,** *50* (W1), W216-W221.
30. LeVine, S. M., Albumin and multiple sclerosis. *BMC Neurol* **2016,** *16*, 47.
31. Menendez-Gonzalez, M.; Gasparovic, C., Albumin Exchange in Alzheimer's Disease: Might CSF Be an Alternative Route to Plasma? *Front Neurol* **2019,** *10*, 1036.
32. Foster, E. M.; Dangla-Valls, A.; Lovestone, S.; Ribe, E. M.; Buckley, N. J., Clusterin in Alzheimer's Disease: Mechanisms, Genetics, and Lessons From Other Pathologies. *Front Neurosci* **2019,** *13*, 164.
33. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A. L.; Fang, T.; Doncheva, N. T.; Pyysalo, S.; Bork, P.; Jensen, L. J.; von Mering, C., The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **2023,** *51* (D1), D638-D646.
34. Shah, A.; Kishore, U.; Shastri, A., Complement System in Alzheimer's Disease. *Int J Mol Sci* **2021,** *22* (24), 13467.

35. Tang, X.; Zhang, Z.; Fang, M.; Han, Y.; Wang, G.; Wang, S.; Xue, M.; Li, Y.; Zhang, L.; Wu, J.; Yang, B.; Mwangi, J.; Lu, Q.; Du, X.; Lai, R., Transferrin plays a central role in coagulation balance by interacting with clotting factors. *Cell Res* **2020,** *30* (2), 119-132.
36. Varma, V. R.; Varma, S.; An, Y.; Hohman, T. J.; Seddighi, S.; Casanova, R.; Beri, A.; Dammer, E. B.; Seyfried, N. T.; Pletnikova, O.; Moghekar, A.; Wilson, M. R.; Lah, J. J.; O'Brien, R. J.; Levey, A. I.; Troncoso, J. C.; Albert, M. S.; Thambisetty, M., Alpha-2 macroglobulin in Alzheimer's disease: a marker of neuronal injury through the RCAN1 pathway. *Mol Psychiatry* **2017,** *22* (1), 13-23.
37. Wang, B.; Zhong, X.; Fields, L.; Lu, H.; Zhu, Z.; Li, L., Structural Proteomic Profiling of Cerebrospinal Fluids to Reveal Novel Conformational Biomarkers for Alzheimer's Disease. *J Am Soc Mass Spectrom* **2023,** *34* (3), 459-471.
38. Maccarrone, G.; Milfay, D.; Birg, I.; Rosenhagen, M.; Holsboer, F.; Grimm, R.; Bailey, J.; Zolotarjova, N.; Turck, C. W., Mining the human cerebrospinal fluid proteome by immunodepletion and shotgun mass spectrometry. *Electrophoresis* **2004,** *25* (14), 2402-12.
39. Wenner, B. R.; Lovell, M. A.; Lynn, B. C., Proteomic analysis of human ventricular cerebrospinal fluid from neurologically normal, elderly subjects using two-dimensional LC-MS/MS. *J Proteome Res* **2004,** *3* (1), 97-103.
40. O'Brien, M. E.; Fee, L.; Browne, N.; Carroll, T. P.; Meleady, P.; Henry, M.; McQuillan, K.; Murphy, M. P.; Logan, M.; McCarthy, C.; McElvaney, O. J.; Reeves, E. P.; McElvaney, N. G., Activation of complement component 3 is associated with airways disease and pulmonary emphysema in alpha-1 antitrypsin deficiency. *Thorax* **2020,** *75* (4), 321-330.
41. O'Brien, M. E.; Murray, G.; Gogoi, D.; Yusuf, A.; McCarthy, C.; Wormald, M. R.; Casey, M.; Gabillard-Lefort, C.; McElvaney, N. G.; Reeves, E. P., A Review of Alpha-1 Antitrypsin Binding Partners for Immune Regulation and Potential Therapeutic Application. *Int J Mol Sci* **2022,** *23* (5), 2441.
42. Wu, T.; Dejanovic, B.; Gandham, V. D.; Gogineni, A.; Edmonds, R.; Schauer, S.; Srinivasan, K.; Huntley, M. A.; Wang, Y.; Wang, T. M.; Hedehus, M.; Barck, K. H.; Stark, M.; Ngu, H.; Foreman, O.; Meilandt, W. J.; Elstrott, J.; Chang, M. C.; Hansen, D. V.; Carano, R. A. D.; Sheng, M.; Hanson, J. E., Complement C3 Is Activated in Human AD Brain and Is Required for Neurodegeneration in Mouse Models of Amyloidosis and Tauopathy. *Cell Rep* **2019,** *28* (8), 2111-2123 e6.
43. Wetterling, T.; Tegtmeyer, K. F., Serum alpha 1-antitrypsin and alpha 2-macroglobulin in Alzheimer's and Binswanger's disease. *Clin Investig* **1994,** *72* (3), 196-9.
44. Lisowska-Myjak, B.; Skarzynska, E.; Jakimiuk, A., Links Between Vitamin D-Binding Protein, Alpha-1 Antitrypsin and Neutrophil Proteins in Meconium. *Cell Physiol Biochem* **2023,** *57* (1), 15-22.
45. Chen, Y. H.; Cheadle, C. E.; Rice, L. V.; Pfeffer, P. E.; Dimeloe, S.; Gupta, A.; Bush, A.; Gooptu, B.; Hawrylowicz, C. M., The Induction of Alpha-1 Antitrypsin by Vitamin D in Human T Cells Is TGF-beta Dependent: A Proposed Anti-inflammatory Role in Airway Disease. *Front Nutr* **2021,** *8*, 667203.
46. Khakzad, H.; Happonen, L.; Tran Van Nhieu, G.; Malmstrom, J.; Malmstrom, L., In vivo Cross-Linking MS of the Complement System MAC Assembled on Live Gram-Positive Bacteria. *Front Genet* **2020,** *11*, 612475.
47. Janssen, B. J.; Huizinga, E. G.; Raaijmakers, H. C.; Roos, A.; Daha, M. R.; Nilsson-Ekdahl, K.; Nilsson, B.; Gros, P., Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* **2005,** *437* (7058), 505-11.

48. Knopman, D. S.; Amieva, H.; Petersen, R. C.; Chetelat, G.; Holtzman, D. M.; Hyman, B. T.; Nixon, R. A.; Jones, D. T., Alzheimer disease. *Nat Rev Dis Primers* **2021,** *7* (1), 33.
49. Husain, M. A.; Laurent, B.; Plourde, M., APOE and Alzheimer's Disease: From Lipid Transport to Physiopathology and Therapeutics. *Front Neurosci* **2021,** *15*, 630502.
50. Chen, Y.; Strickland, M. R.; Soranno, A.; Holtzman, D. M., Apolipoprotein E: Structural Insights and Links to Alzheimer Disease Pathogenesis. *Neuron* **2021,** *109* (2), 205-221.
51. Frieden, C.; Garai, K., Structural differences between apoE3 and apoE4 may be useful in developing therapeutic agents for Alzheimer's disease. *Proc Natl Acad Sci U S A* **2012,** *109* (23), 8913-8.
52. Huang, R. Y.; Garai, K.; Frieden, C.; Gross, M. L., Hydrogen/deuterium exchange and electron-transfer dissociation mass spectrometry determine the interface and dynamics of apolipoprotein E oligomerization. *Biochemistry* **2011,** *50* (43), 9273-82.
53. Yamauchi, Y.; Deguchi, N.; Takagi, C.; Tanaka, M.; Dhanasekaran, P.; Nakano, M.; Handa, T.; Phillips, M. C.; Lund-Katz, S.; Saito, H., Role of the N- and C-terminal domains in binding of apolipoprotein E isoforms to heparan sulfate and dermatan sulfate: a surface plasmon resonance study. *Biochemistry* **2008,** *47* (25), 6702-10.
54. Snow, A. D.; Cummings, J. A.; Lake, T., The Unifying Hypothesis of Alzheimer's Disease: Heparan Sulfate Proteoglycans/Glycosaminoglycans Are Key as First Hypothesized Over 30 Years Ago. *Front Aging Neurosci* **2021,** *13*, 710683.
55. Zhang, G. L.; Zhang, X.; Wang, X. M.; Li, J. P., Towards understanding the roles of heparan sulfate proteoglycans in Alzheimer's disease. *Biomed Res Int* **2014,** *2014*, 516028.
56. Katsouri, L.; Georgopoulos, S., Lack of LDL receptor enhances amyloid deposition and decreases glial response in an Alzheimer's disease mouse model. *PLoS One* **2011,** *6* (7), e21880.
57. Dominguez, C.; Boelens, R.; Bonvin, A. M., HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **2003,** *125* (7), 1731-7.

Supplementary Information

Supplementary Figures

a **b**

Figure S1. Chemical structures of the crosslinker BS3d0 (a) and BS3d4 (b).

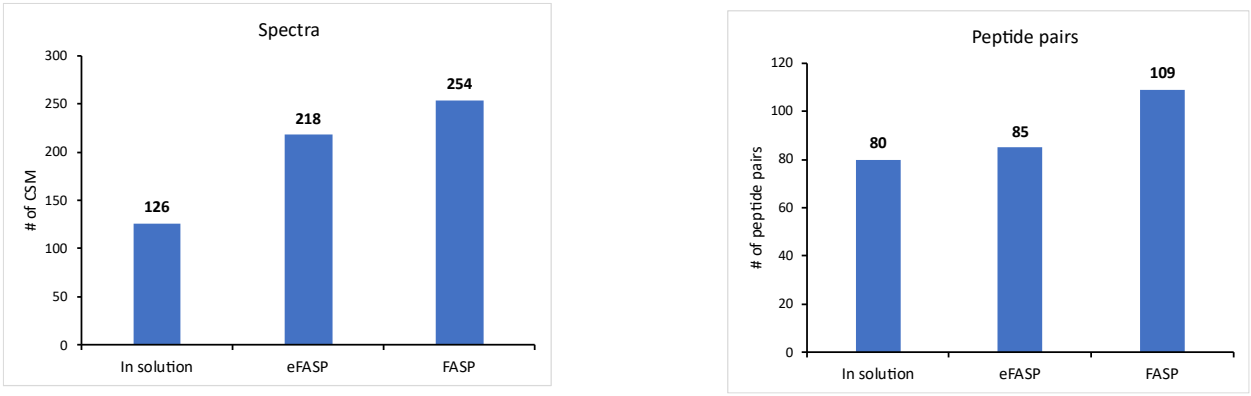


Figure S2. Comparison of outcomes of crosslinking from in solution, eFASP and FASP digestion methods. CSM means crosslinking spectrum matching. One peptide pair can have multiple numbers of CSM.

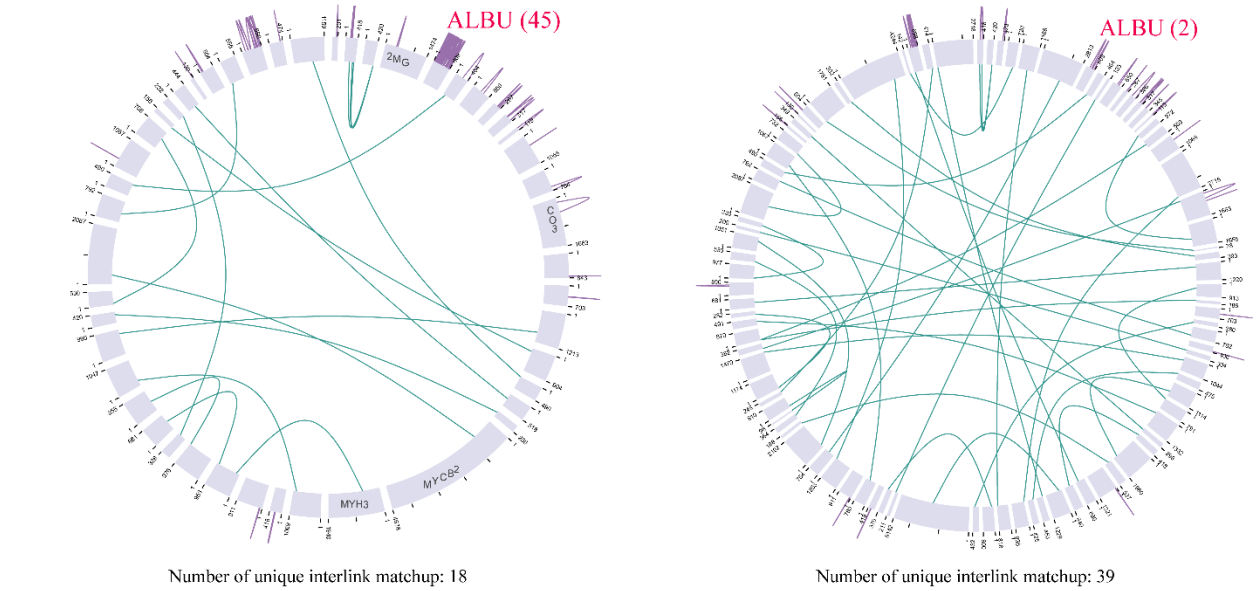


Figure S3. Comparison of the crosslink results with and without albumin depletion. The number in the bracket indicates the number of the identified crosslink pairs.

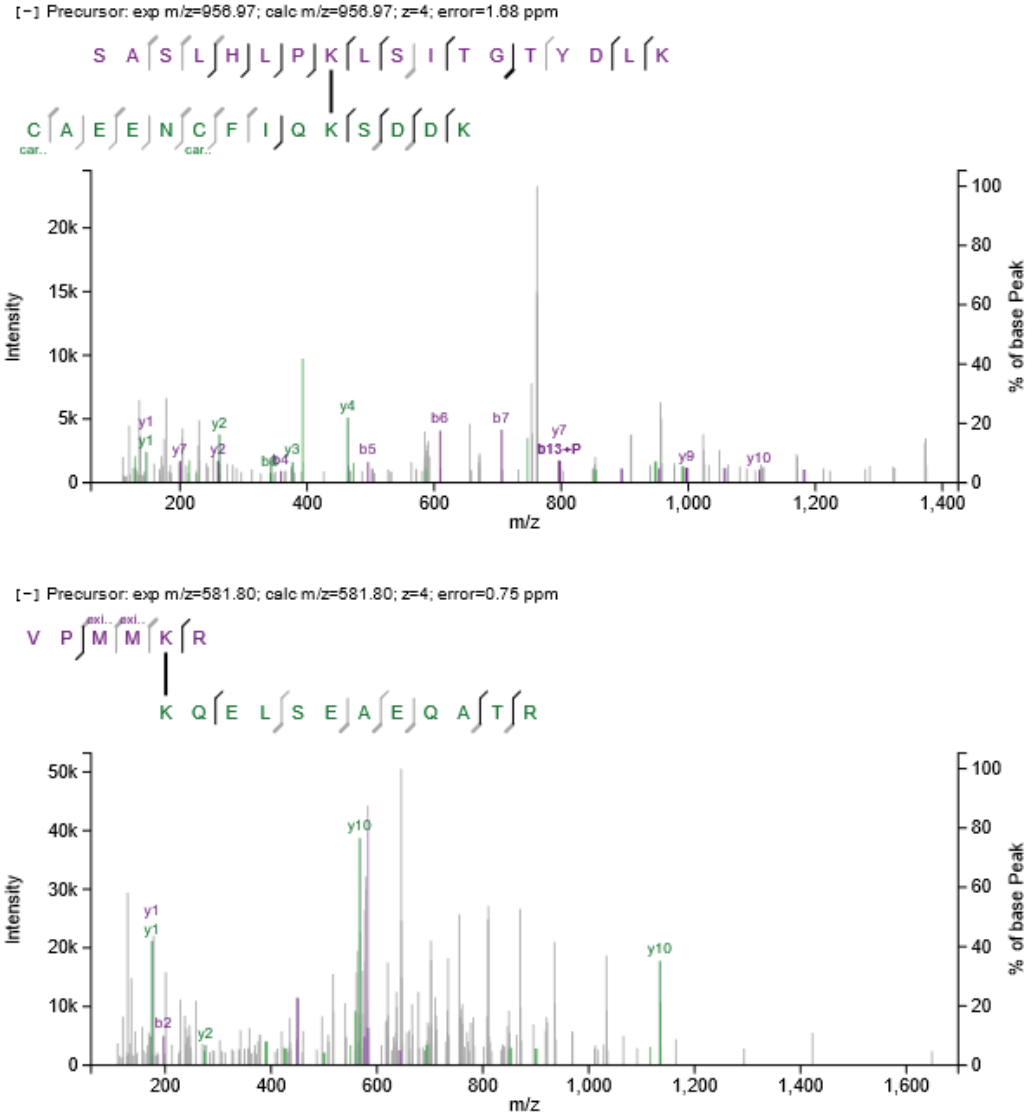


Figure S4. The different CSM matchup between A1AT and CO3 within one specific set. In the set of AD2 and Ctrl2, A1AT 304 - CO3 1522 from Ctrl group is shown on top and A1AT 246 - CO3 428 from AD is shown on the bottom.

Supplementary Tables

Table S1. Human CSF sample information

Name	Age	Gender	Biomarker group	Category
Ctrl1/2	69.55	Male	biomarker negative	Control
MCI1	65.24	Male	amyloid positive	MCI
AD1	65.85	Male	amyloid & tau positive	Dementia
Ctrl1/2	69.55	Male	biomarker negative	Control
MCI2	59.90	Male	amyloid & tau positive	MCI
AD2	72.99	Male	amyloid & tau positive	Dementia
Ctrl3	73.33	Male	biomarker negative	Control
MCI3	61.31	Male	amyloid & tau positive	MCI
AD3	62.45	Male	amyloid & tau positive	Dementia
Ctrl4	76.33	Male	biomarker negative	Control
MCI4	73.07	Male	amyloid & tau positive	MCI
AD4	63.61	Male	amyloid & tau positive	Dementia
Ctrl5	76.33	Male	biomarker negative	Control
MCI5	74.34	Male	amyloid & tau positive	MCI
AD5	76.91	Male	amyloid & tau positive	Dementia
Ctrl6	77.71	Male	biomarker negative	Control
MCI6	75.59	Male	amyloid & tau positive	MCI
AD6	84.58	Male	amyloid & tau positive	Dementia
Ctrl7	77.84	Male	biomarker negative	Control
MCI7	77.00	Male	amyloid & tau positive	MCI
AD7	79.01	Male	amyloid & tau positive	Dementia
Ctrl8	80.80	Male	amyloid & tau positive	Control
MCI8	83.27	Male	amyloid & tau positive	MCI
AD8	80.17	Male	amyloid & tau positive	Dementia
Ctrl9/10	80.91	Male	tau positive	Control
MCI9	79.09	Male	amyloid positive	MCI
AD9	63.34	Female	amyloid & tau positive	Dementia
Ctrl9/10	80.91	Male	tau positive	Control
MCI10	80.05	Male	amyloid & tau positive	MCI
AD10	65.48	Female	amyloid & tau positive	Dementia
Ctrl11	62.47	Female	tau positive	Control
MCI11	84.18	Male	amyloid & tau positive	MCI
AD11	67.77	Female	amyloid & tau positive	Dementia
Ctrl12	67.17	Female	biomarker negative	Control
MCI12	87.29	Male	amyloid & tau positive	MCI

AD12	93.06	Female	amyloid & tau positive	Dementia
Ctrl13	67.68	Female	amyloid & tau positive	Control
MCI13	77.66	Female	amyloid & tau positive	MCI
AD13	71.30	Female	amyloid & tau positive	Dementia
Ctrl14	69.70	Female	tau positive	Control
MCI14	79.38	Female	amyloid positive	MCI
AD14	71.51	Female	amyloid & tau positive	Dementia
Ctrl15	74.22	Female	biomarker negative	Control
MCI15	75.54	Female	amyloid positive	MCI
AD15	75.58	Female	amyloid & tau positive	Dementia
Ctrl16	74.83	Female	biomarker negative	Control
MCI16	77.33	Female	amyloid & tau positive	MCI
AD16	77.79	Female	amyloid & tau positive	Dementia

Table S2. Top 20 proteins in the searching database

Abbreviation	Protein Full Name	Uniprot No.
A1AT	Alpha-1-antitrypsin (SERPINA1)	P01009
A1ATR	Alpha-1-antitrypsin-related protein (SERPINA2)	P20848
A1BG	Alpha-1B-glycoprotein	P04217
A2MG	Alpha-2-macroglobulin	P01023
AACT	Alpha-1-antichymotrypsin	P01011
ANGT	Angiotensinogen	P01019
ANT3	Antithrombin-III	P01008
APOA1	Apolipoprotein A-I	P02647
APOE	Apolipoprotein E	P02649
B2MG	Beta-2-microglobulin	P61769
CFAB	Complement factor B	P00751
CO3	Complement C3	P01024
CO5	Complement C5	P01031
ENPP2	Ectonucleotide pyrophosphatase/phosphodiesterase family member 2 (Autotaxin)	Q13822
GELS	Gelsolin	P06396
PEDF	Pigment epithelium-derived factor	P36955
PZP	Pregnancy zone protein	P20742
TRFE	Serotransferrin	P02787
VTDB	Vitamin D-binding protein	P02774
ZA2G	Zinc-alpha-2-glycoprotein	P25311

Table S3. Summary of interlinks between A1AT and CO3

Sample category	XL peptide sequence	A1AT sites	CO3 sites
Ctrl	SASLHLP**K**LSITGTYDLKK-CAEENCFIQ**K**SDDK	314	1522
	VPMM**K**RK-SVQLTE**K**RMDK	246	678
	G**K**WERPFEVKK-RQGALELI**K**K	217	1050
	SASLHLP**K**LSITGTYDLKK-RQGALELI**K**K	314	1050
	FLEDV**K**K-A**K**DQLTCNK	159	1353
	VPMM**K**RK-VG**K**YPK	246	685
MCI	VPMM**K**RK-**K**QELSEAEQATR	246	428
AD	VPMM**K**RK-YELD**K**AFSDR	246	1436
	VPMM**K**RK-**K**QELSEAEQATR	246	428

Table S4. Summary of interlinks within ApoE

Sample category	XL peptide sequence	ApoE site 1	ApoE site 2
Ctrl	LDEV**K**EQVAEVR-A**K**LEEQAQQIR	251	260
	LS**K**ELQAAQAR-DADDLQ**K**R	113	175
	DADDLQ**K**R-A**K**LEEQAQQIR	175	260
	LS**K**ELQAAQAR-EL**K**AYK	113	90
MCI	LDEV**K**EQVAEVR-A**K**LEEQAQQIR	251	260
	EL**K**AYK- A**K**LEEQAQQIR	90	260
	EL**K**AYK- LS**K**ELQAAQAR	90	113
	LS**K**ELQAAQAR- A**K**LEEQAQQIR	113	260

Chapter 6

XLeu: A Dual-Function Crosslinker Enabling Cleavable Crosslinking and Isobaric Quantitation

Adopted from: **Zhu, Z**.; Wei, W.; Douglas, C. A.; Wang, Z.; Li, L. XLeu: A Dual-Function Crosslinker Enabling Cleavable Crosslinking and Isobaric Quantitation. In preparation. **Zhu, Z.** contributed the experimental design, data collection, analysis, and manuscript preparation.

Abstract

Protein–protein interactions (PPIs) are central to cellular function, and their dysregulation plays a key role in neurodegenerative diseases such as Alzheimer's disease (AD). Crosslinking mass spectrometry (XL-MS) has emerged as a powerful approach to map PPIs at residue-level resolution, yet existing reagents are limited in their ability to provide multiplexed quantitation across biological conditions. Here, we introduce a novel 4-plex isobaric crosslinker, XLeu, which enables simultaneous identification and quantitation of PPIs via cleavable reporter ions. Benchmarking against established crosslinkers (DSSO, BS^3, DSBSO) using bovine serum albumin demonstrated that XLeu provides labeling efficiency, structural coverage, and quantitative accuracy comparable to commercial reagents, while offering multiplexing capacity. Application of XLeu to human serum samples revealed PPI networks centered on transferrin and complement factors, highlighting their role as structural hubs. Importantly, condition-specific quantitation identified a high-confidence crosslink between complement component CO3 and CO5 enriched in control females, supporting immune network remodeling in AD. Furthermore, age-stratified analysis uncovered a non-linear relationship between crosslink density and age, with the highest connectivity observed in individuals between 65–80 years. Collectively, these results establish 4-plex XLeu as a robust reagent for multiplexed XL-MS, providing new opportunities to investigate condition- and age-dependent remodeling of PPI networks in complex proteomes.

Introduction

Crosslinking mass spectrometry (XL-MS) has become a key technology in structural proteomics, enabling the direct capture of residue-to-residue proximities that define protein conformations and protein–protein interactions (PPIs). This approach has revealed dynamic structural assemblies and interaction networks in both model proteins and complex biological systems. However, while identification of crosslinked peptides has advanced considerably, a critical need remains for quantitative strategies that can directly assess condition-dependent structural changes and interaction remodeling across multiple biological states[1, 2].

Quantitative XL-MS (QXL-MS) strategies have emerged to address this challenge by integrating structural information with quantitative proteomics. As summarized in recent reviews from our lab, QXL-MS approaches typically fall into three categories: label-free quantification (LFQ), isotopic labeling, and isobaric labeling[3]. LFQ relies on MS1 ion intensities of crosslinked peptides, providing broad applicability and low cost, but suffers from reproducibility issues and limited sensitivity in complex samples[4]. In contrast, isotopic labeling employs light/heavy crosslinker analogs to generate defined mass shifts detectable at the MS1 level. Pioneering work by the Rappsilber group established the conceptual framework for isotopic quantitative crosslinking, using approaches such as BS^3-d4 and isotope-coded reagents to track conformational differences across conditions[5, 6]. While robust for pairwise comparisons, isotopic methods are inherently limited to low-plex designs and often require laborious forward/reverse labeling schemes for accurate quantitation.

Isobaric labeling strategies extend multiplexing capacity by embedding barcode-like reporter ions within cleavable crosslinkers. Upon fragmentation, these reagents release reporter ions

whose relative intensities reflect the abundance of crosslinked peptides across multiple samples[7, 8]. These approaches enable higher throughput but often require complex multi-step workflows, such as tag removal or MS3 acquisition, to achieve accurate quantitation[9]. Furthermore, the expense and chemical instability of some isobaric reagents present barriers for routine adoption.

In this context, we developed the XLeu family of crosslinkers, beginning with a 4-plex design, as a proof-of-concept toward streamlined multiplex quantitative XL-MS. The key innovation of XLeu lies in its "two birds-one-stone" strategy: a single reagent that simultaneously incorporates MS2-cleavable bonds for structural identification and isobaric reporter groups for direct quantitation. Unlike conventional isobaric methods, XLeu automatically liberates its reporter ions during MS/MS without requiring additional tag-exchange steps, thereby simplifying sample preparation and increasing data acquisition efficiency. Moreover, its synthesis leverages cost-effective and chemically stable components, enabling broader accessibility.

The result is a reagent that provides direct abundance comparison, higher identification rates, and stable quantitative accuracy, while maintaining compatibility with existing XL-MS workflows. Here, we describe the synthesis, benchmarking, and application of 4-plex XLeu. We compare its performance with established commercial crosslinkers (DSSO, BS^3, DSBSO), validate its quantitative fidelity in purified proteins, and apply it to human serum samples to investigate disease- and age-related remodeling of PPI networks in Alzheimer's disease.

Experimental Section

AD Serum Samples and Chemical Reagents

All human subjects involved in this study have received approval from the University of Wisconsin Institutional Review Board. The study comprised of 16 participants from the Wisconsin Alzheimer's Disease Research Center (ADRC), who were individually selected from healthy controls and AD groups (**Supplemental Table S1**).

Synthesis of XLeu

All reactions were carried out under ambient atmosphere unless otherwise noted. N,N-Dimethylformamide (DMF), dichloromethane (DCM), ethyl acetate (EA), and trifluoroacetic acid (TFA) were used as received. N,N-Diisopropylethylamine (DIPEA) served as base. O-(7-azabenzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate (HATU) and O-(N-succinimidyl)-N,N,N′,N′-tetramethyluronium tetrafluoroborate (TSTU) were used as coupling/activation reagents (**Figure S1**).

Synthesis of compound 2. To a stirred solution of compound 1 (315 mg, 1.0 equiv) in MeOH/H_2O (4:1, 0.05–0.10 M) at room temperature was added aqueous formaldehyde (855uL, ~37 %, 4.0 equiv), followed by sodium cyanoborohydride ($NaBH_3CN$, 360mg, 3.0 equiv) in one portion. The mixture was stirred 2–4 h. The reaction was quenched with sat. $NaHCO_3$, concentrated to remove most methanol, diluted with water, and extracted with DCM (3×). The combined organic layers were washed with brine, dried (Na_2SO_4), filtered, and concentrated to give the dimethylated product compound 2, which was used without further purification.

Synthesis of compound 12. Compound 2 (210 mg, 1.38 mmol) and di-tert-butyl–protected linker 8 (377uL, 1.0 equiv) were dissolved in DCM 5mL. DIPEA (333 uL) was added, followed by HATU (520 mg). The reaction was stirred at r.t. for 3 h to afford the amide. The mixture was diluted with EA, washed successively with 0.1 M HCl, sat. $NaHCO_3$, and brine, dried, and concentrated. Crude material was subjected to silica gel column, elute with EA/hexane from 1:10 to 1:1 furnish the di-tert-butyl protected compound 12.

Synthesis of compound 11. Compound 12 (100 mg) was dissolved in DCM, and TFA was added (TFA/DCM 1:1 to 3:1, 0.02–0.05 M). The solution was stirred at r.t. for 0.5–2 h, compound 12 completely consumed by TLC monitoring. Volatiles were removed under reduced pressure, and residual TFA was co-evaporated with toluene (2–3×). The residue was dissolved in minimal DMF/H_2O and the mixture was adjusted to pH 7–8 with aqueous $NaHCO_3$ with ice bath. The aqueous phase was washed with DCM to remove nonpolar by-products, then concentrated (or lyophilized) to afford the compound 11.

Synthesis of XLeu (compound 6). Compound 11 (1.0 equiv) was dissolved in dry DMF (0.01–0.05 M) and cooled to r.t. DIPEA (3–4 equiv) was added, followed by TSTU (2.2–2.5 equiv, to activate both carboxylates). The mixture was stirred at r.t. for 1–2 h (LC–MS shows formation of the bis-NHS ester). The reaction was diluted with EA and carefully quenched with a small volume of cold water. Layers were separated; the organic phase was washed with ice-cold brine, dried (Na_2SO_4), filtered, and high-vacuum evaporated to give the XLeu as the bis-NHS ester.

Following completion of the reaction, the crude product was analyzed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI–MS). The spectrum displayed a dominant ion at m/z 497.354, corresponding to the $[M+H]^+$ species of the target XLeu bis-NHS ester (Figure S2). This peak was the most intense in the spectrum, indicating that the desired cross-linker was

obtained as the main reaction product, with only minor peaks corresponding to side products or residual impurities.

Crosslinking and Sample Preparation

The detailed experimental workflow is illustrated in Figure 1. Protein concentrations of human serum samples were determined using the BCA protein assay kit (Thermo Scientific, Pittsburgh, PA). For each sample, approximately 100 μg of protein was subjected to crosslinking with 4-plex XLeu crosslinkers dissolved in DMSO at a stock concentration of 50 mM. Male control samples were labeled with 118c, female control with 117a, male AD with 116b, and female AD with 115a. Based on these pairings, the 16 total samples were organized into four age groups: ⩾80, 73-80, 65-73, and ⩽65. Crosslinking reactions were performed in 20 mM HEPES buffer (pH 7.5) at room temperature for 1 h and quenched with 20 mM Tris buffer for 20 min. Following quenching, crosslinked samples were combined at a 1:1:1:1 ratio and subjected to albumin depletion using the Pierce Albumin Depletion Kit (Thermo Scientific, Waltham, MA). Samples were subsequently reduced with 20 mM dithiothreitol (DTT) for 45 min, alkylated with 20 mM iodoacetamide (IAA) in the dark for 45 min, and digested overnight at 37 °C with trypsin/LysC (enzyme:protein ratio 1:50, Promega, Madison, WI). The digestion was quenched with 10% trifluoroacetic acid (TFA) to pH < 3, followed by desalting on C18 cartridges using 80% acetonitrile (ACN) with 0.1% formic acid (FA) for elution. Eluates were dried under vacuum and stored at –20 °C until LC–MS/MS analysis. The validation process employed purified recombinant bovine serum albumin (BSA) to incubate with 4-plex crosslinker.

LC–MS/MS Analysis

Dried peptide samples were reconstituted in 97% Optima-grade water, 3% ACN, and 0.1% FA prior to LC–MS/MS analysis. Peptides were separated using a Dionex UltiMate 3000 UPLC system coupled to a Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer. Separation was performed on a homemade fused-silica microcapillary column (75 μm inner diameter) packed with 18 cm of Bridged Ethylene Hybrid C18 resin (1.7 μm, 130 Å, Waters). The gradient was run over 100 min, from 3% to 95% mobile phase B (ACN with 0.1% FA), at a flow rate of 300 nL/min. Data were acquired in a stepwise higher-energy collisional dissociation (HCD) mode with a resolving power of 120,000 for MS1 scans, an m/z range of 400–1600, and a maximum injection time of 50 ms. Data-dependent acquisition was performed with a 3 s cycle time between master scans using normalized collision energies of 21, 31, and 33 (low–high–high HCD strategy), adapted from protocols developed in consultation with Dr. Liu Fan's group. MS2 scans were acquired with an isolation window of 2 Da, resolving power of 30,000, and standard AGC target settings.

Quantitative Structural Analysis and Visualization

Raw files were searched using xiSearch (version 1.7.6.7) against the reviewed *Homo sapiens* UniProt database, with all 4-plex XLeu isobaric reporter ions enabled. Trypsin was set as the digestion enzyme, allowing up to three missed cleavages. Precursor mass tolerance was set to 20 ppm, and peptide-spectrum match (PSM) false discovery rate (FDR) was controlled below 5%. To improve computational efficiency and reduce interference from low-scoring matches, searches were restricted to the top 20 proteins with the greatest number of crosslinks in preliminary runs (**Table S2**). Since the protein composition in human serum differs significantly from that in CSF. Specifically, we recalibrated the protein abundance ranking using intensity-based absolute quantification (iBAQ) curve, informed by Thermo Nanodrop measurements and

basic proteomics. Quantification was performed using MaxQuant, with precursor charge states limited to 3–7 and peptide selection parameters matched to the xiSearch output. Statistical analyses were conducted with a p-value threshold of 0.05 and a fold-change cutoff of 1.0. Volcano plots were generated to visualize significantly up- or downregulated crosslinked peptides. Additional visualizations, including bar charts, bubble plots, and network diagrams, were prepared using HiPlot and custom Python scripts. The .mgf, .mzML, and FASTA files were also uploaded to xiView for two-dimensional protein–protein network mapping, while three-dimensional structural models were generated using PyMOL and ChimeraX (versions 1.3–1.4). Gene Ontology enrichment analyses for biological process, cellular component, and molecular function were conducted using the DAVID Bioinformatics Resources.

Results and Discussion

4-plex XLeu exhibits reliable labeling efficiency and quantitative accuracy corresponding to its reporter ion's abundance

The crosslinking performance of 4-plex XLeu was compared with three commercially available crosslinkers—DSSO, BS3, and azide-A-DSBSO—using purified bovine serum albumin as the model protein. Crosslinked spectrum matches (CSMs), unique protein–protein interactions (Unique PPs), and unique crosslinking sites were quantified for each reagent. DSSO yielded the highest total CSM count (154), followed by XLeu (87), DSBSO (34), and BS3 (23) (**Figure 2a**). Notably, 4-plex XLeu demonstrated a substantial number of unique protein–protein interactions (20) and unique sites (17), which are comparable to DSSO (34 PPIs, 16 sites) despite being a newly developed reagent. The corresponding xiView network maps further illustrate that XLeu

provided broad coverage across the albumin sequence, generating multiple inter- and intra-peptide crosslinks, with a mapping composition with DSSO and BS3 (**Figure 2b**), both of which have been widely adopted in structural proteomics for their robust spacer chemistry and crosslinking efficiency[10-12]. These results confirm that 4-plex XLeu achieves good labeling efficiency and delivers crosslink identifications with accuracy and breadth comparable to established commercial crosslinkers, while offering the added advantage of multiplex quantitation capability[13].

Protein dynamics within human serum are inherently variable, with fluctuations in protein abundance, interaction states, and accessibility that differ across individuals and conditions. As a result, crosslinking in such a complex biological background cannot capture every potential interaction in real time, and the resulting reporter ion signals may be influenced by these dynamic factors. To assess the quantitative fidelity of the 4-plex XLeu crosslinker, we first validated the reporter ion signals under controlled conditions using direct infusion ESI–MS (**Figure 3A**). Crosslinker standards were mixed at an equivalent ratio (1:1:1:1 for channels 115, 116, 117, and 118) prior to MS analysis. The resulting reporter ion intensities showed near-identical peak heights across all four channels, confirming that the isotopic labels in XLeu generate balanced signal output in an idealized, interference-free environment. We then evaluated the performance of XLeu in a complex biological background using purified BSA samples. Even in this heterogeneous proteome context, the reporter ion abundance remained consistent with the expected stoichiometry (**Figure 3C**), demonstrating that multiplex quantitation accuracy is preserved after crosslinking reactions and subsequent sample processing. Importantly, this abundance distribution strongly correlated with the crosslinking output metrics, including the number of crosslinked spectra matches, unique protein–protein interactions, and

unique crosslinked sites, as shown in **Figure 3B**. These results confirm that 4-plex XLeu maintains high quantitative precision—balancing controlled-condition accuracy with performance in real biological matrices.

These findings align with prior observations in quantitative crosslinking mass spectrometry, which emphasize that accurate MS1 or MS2 reporter ion quantitation relies on high-resolution acquisition and careful control of interference[2]. Furthermore, strategies employing isobaric labeling for crosslinking quantitation have demonstrated similar robustness in complex samples (e.g., isobaric crosslinkers like Qlinker)[14]. Together, these references reinforce that our approach with 4-plex XLeu is consistent with and builds upon established standards in the field of QXL-MS.

Capturing Proteome-Wide Interaction Networks in Serum Samples Collected from AD and Control Subjects

Blood serum represents a particularly challenging proteomic matrix for crosslinking mass spectrometry due to its extreme protein dynamic range, high abundance of albumin, and the presence of immune and transport proteins that undergo significant remodeling in Alzheimer's disease (AD)[15, 16]. As such, evaluating whether multiplexed crosslinking reagents like XLeu can reliably capture interaction networks in serum is critical for establishing their utility in clinical proteomics. **Figure 4** presents the crosslinking interaction networks detected in human serum using 4-plex XLeu labeling, with panels a–d corresponding to AD female (115), AD male (116), control female (117), and control male (118), respectively. Across all four groups, the majority of interlinks were centered on transferrin (TRFE, UniProt P02787) and extended into members of the complement system (e.g., CO3, CO5), reflecting their importance as structural hubs in serum proteome organization. This observation aligns with prior reports highlighting transferrin and

complement proteins as central nodes in serum protein–protein interaction networks and as biomarkers implicated in neuroinflammation and AD pathology[17-19]. Notably, comparison between AD and control samples revealed that crosslinking connectivity was modestly higher in controls, suggesting reduced structural stability or weaker interaction density in the AD background. This trend is consistent with our previously published cerebrospinal fluid (CSF) study, where diminished crosslinking density was observed in AD relative to controls, further supporting the conclusion that AD pathology is associated with destabilized protein–protein interaction networks across both CSF and serum compartments[20].

Building on our observation that transferrin and complement factors represented the central hubs of crosslinking networks in serum, we next examined specific spectra to validate these interactions at the peptide level. The annotated MS/MS spectrum revealed an important crosslink between complement component 3 (CO3) K1595 and complement component 5 (CO5) K1327 (**Figure 5)**. This crosslink was identified specifically in the sample 3514, with the highest reporter ion in channel 117 representing the female control group. This highlights the ability of reporter-ion-based multiplexed quantitation to directly associate individual protein–protein interactions with defined biological conditions. Importantly, the CO3–CO5 crosslink exhibited one of the strongest signal intensities among all detected crosslinks, reinforcing its biological significance and providing evidence that this interaction is both abundant and reproducibly detected in our dataset. Taken together, these results underscore that multiplex crosslinking with reporter ions not only validates structural hubs such as the complement cascade but also enables condition-specific mapping of protein–protein interactions, linking molecular connectivity directly to disease versus control states.

Finally, we evaluated whether age influenced the extent of crosslinking events detected in serum. The distribution of crosslinking metrics did not follow a straightforward linear pattern (**Figure 6**). Individuals between 65 and 80 years old exhibited the highest number of crosslinked spectrum matches (CSMs), type II interlinks, and protein–protein interactions (PPIs), suggesting that this age range may coincide with a period of increased crosslinking activity. In contrast, the >80 cohort showed a reduction in crosslink counts, indicating that advanced age does not necessarily correspond to greater crosslink density. However, due to the limited sample size and uneven representation across age groups, these observations should be interpreted with caution. Rather than supporting a definitive association, our results highlight the complexity of age-dependent remodeling in crosslinked proteomes and emphasize the importance of larger, stratified cohorts to determine whether specific crosslinking patterns can be robustly linked to age-related biology. For future studies, we intend to incorporate spectral library-based searches and related advanced methods to improve identification sensitivity and throughput. By extending our data acquisition in these ways, we expect to mitigate the sample size limitation and capture a much richer cross-link dataset. Ultimately, this will provide a stronger foundation for biological conclusions, moving closer to a more complete protein interactome mapping.

Conclusion

In this study, we developed and validated 4-plex XLeu, a next-generation isobaric crosslinker that integrates structural identification with multiplexed quantitation. Through systematic benchmarking, we demonstrated that XLeu achieves labeling efficiency and crosslink coverage comparable to leading commercial reagents while uniquely enabling simultaneous analysis of

four conditions. Application to human serum proteomics revealed central hubs in transferrin and complement proteins, condition-specific connectivity differences between AD and control groups, and reproducible peptide-level crosslinks such as C3–C5. Age-stratified analysis further uncovered that crosslinking density peaks in the 65–80 range but declines in extreme old age, underscoring the complexity of proteome remodeling during aging.

Together, these findings confirm that XLeu is not only a reliable tool for quantitative structural proteomics but also expands the analytical scope of XL-MS to capture subtle biological variations linked to disease and aging. Beyond serum studies, the multiplexing capacity of XLeu opens new avenues for investigating PPIs in tissue-specific proteomes, longitudinal clinical cohorts, and large-scale biomarker discovery efforts. Future work will focus on extending XLeu to higher-plex designs (e.g., 16-plex), integrating with ion mobility MS, and applying it to additional neurodegenerative disease models to further elucidate dynamic proteome organization.

References

1. Rappsilber, J.; Bruce, J. E.; Combe, C. W.; Fried, S. D.; Heck, A. J. R.; Iacobucci, C.; Leitner, A.; Mechtler, K.; Novak, P.; O'Reilly, F.; Schriemer, D. C.; Sinz, A.; Stengel, F.; Graziadei, A.; et al., A Roadmap for Improving Data Reliability and Sharing in Crosslinking Mass Spectrometry. *bioRxiv* **2025**.
2. Chen, Z. A.; Rappsilber, J., Protein Dynamics in Solution by Quantitative Crosslinking/Mass Spectrometry. *Trends Biochem. Sci.* **2018,** *43* (11), 908-920.
3. Lu, H.; Zhu, Z.; Fields, L.; Zhang, H.; Li, L., Mass Spectrometry Structural Proteomics Enabled by Limited Proteolysis and Cross-Linking. *Mass Spectrom Rev* **2024**.
4. Müller, F.; Rappsilber, J., On the Reproducibility of Label-Free Quantitative Cross-Linking/Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 568-577.
5. Fischer, L.; Rappsilber, J., Quantitative cross-linking/mass spectrometry using isotope-labeled cross-linkers. *Journal of Proteomics* **2013,** *88*, 120-128.
6. Graziadei, A.; Rappsilber, J., Leveraging Crosslinking Mass Spectrometry in Structural and Cell Biology. *Structure* **2022,** *30*, 11-22.

7. Yu, C.; et al., QMIX: Quantitation of Multiplexed Isobaric-labeled Cross-linked Peptides. *Molecular & Cellular Proteomics* **2016,** *15*, 2773-2785.
8. Chavez, J. D.; Keller, A.; Zhou, B.; Tian, R.; Bruce, J. E., Isobaric Quantitative Protein Interaction Reporter (iqPIR) Cross-Linkers Enable Multiplexed Quantitative Proteomics and Structural Biology. *Nature Biotechnology* **2020,** *38*, 1204-1212.
9. Walzthoeni, T.; Claassen, M.; Leitner, A.; Herzog, F.; Bohn, S.; Förster, F.; Beck, M.; Aebersold, R., False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nature Methods* **2015,** *12*, 1185-1190.
10. Liu, F.; Lössl, P.; Scheltema, R.; Viner, R.; Heck, A. J. R., Optimized Fragmentation Schemes and Data Analysis Strategies for Proteome-wide Cross-link Identification. *Nat. Commun.* **2017,** *8*, 15473.
11. Kalkhof, S.; Sinz, A., Chances and Pitfalls of Chemical Cross-linking with Amine-Reactive N-Hydroxysuccinimide Esters. *Anal. Bioanal. Chem.* **2008,** *392* (1-2), 305-312.
12. Iacobucci, C.; Piotrowski, C.; Rehkamp, A.; Ihling, C. H.; Sinz, A., Carboxylic Acid Diester-Based MS-Cleavable Cross-Linkers for Studying Protein–Protein Interactions. *Angew. Chem. Int. Ed.* **2019,** *58* (50), 17313-17318.
13. Wippel, H. H.; Rappsilber, J., Quantitative Chemical Cross-Linking with Mass Spectrometry (qXL-MS) Reveals Structural Dynamics in Proteins. *Anal. Chem.* **2021,** *93*, 12300-12310.
14. Luo, J.; Ranish, J., Isobaric Crosslinking Mass Spectrometry Technology for Studying Conformational and Structural Changes in Proteins and Complexes. *eLife* **2024,** *13*, e99809.
15. Anderson, N. L.; Anderson, N. G., The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics* **2002,** *1* (11), 845-867.
16. Hortin, G. L.; Sviridov, D.; Anderson, N. L., High-Abundance Polypeptides of the Human Plasma Proteome Comprising the Top 4 Logs of Polypeptide Abundance. *Clin. Chem.* **2008,** *54* (10), 1608-1616.
17. Faria, N. R.; et al., Transferrin as a Marker of Protein Interaction and Iron Homeostasis. *Biochim. Biophys. Acta* **2014,** *1840*, 3095-3103.
18. Hajjar, I.; et al., Serum Iron, Transferrin Saturation and Risk of Dementia. *J. Alzheimer's Dis.* **2016,** *49* (3), 725-732.
19. Morgan, B. P.; Harris, C. L., Complement, a Target for Therapy in Inflammatory and Degenerative Diseases. *Nat. Rev. Drug Discov.* **2015,** *14* (12), 857-877.
20. Zhu, Z.; Zhong, X.; Wang, B.; Lu, H.; Li, L., Probing Protein Structural Changes in Alzheimer's Disease via Quantitative Cross-linking Mass Spectrometry. *Anal Chem* **2024,** *96* (19), 7506-7515.

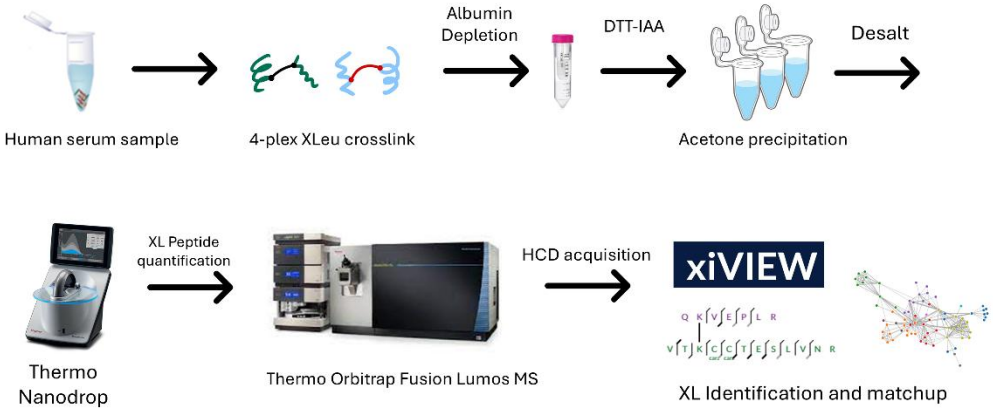


Figure 1. The workflow of the XLeu incubation in human serum testing.

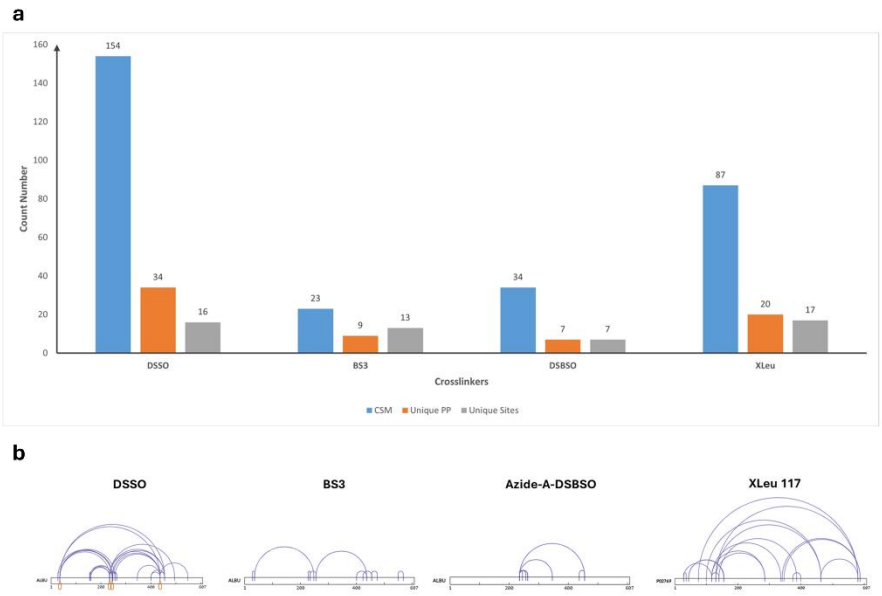


Figure 2. Crosslinking performance comparison to the other merchandized crosslinkers in BSA. a) The statistic number of CSM, unique PP and crosslinked sites of four crosslinkers. b) The 2D interlinking represented in xiView.

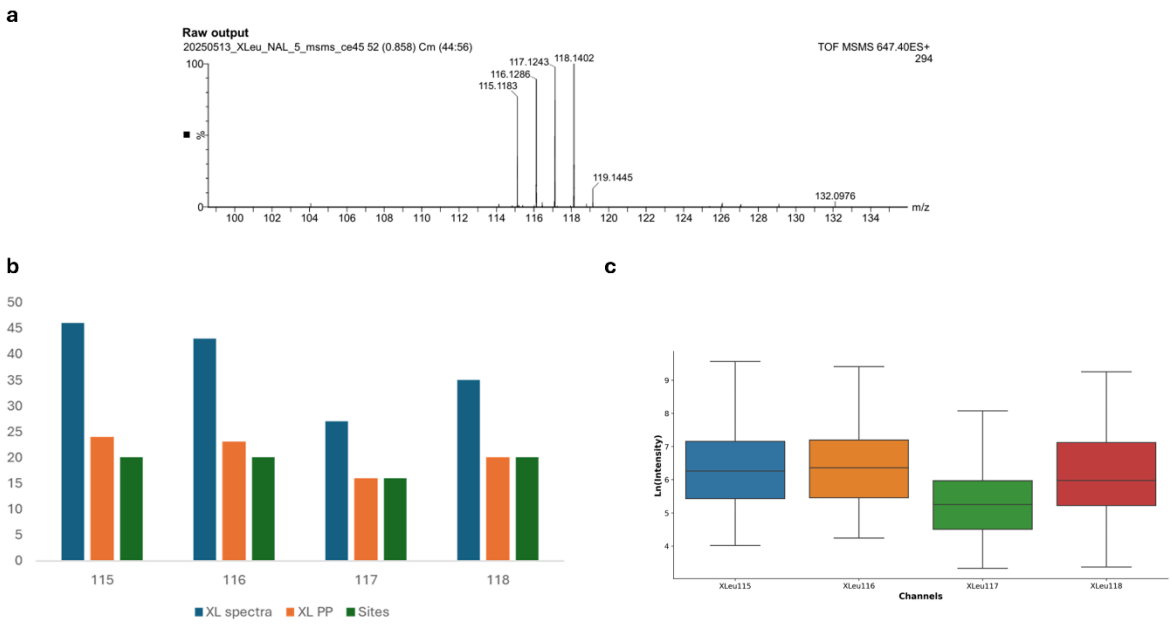


Figure 3. Quantitative performance evaluation of 4-plex XLeu crosslinker. a) Direct infusion ESI–MS reporter ion spectra for 4-plex XLeu channels 115, 116, 117, and 118 mixed at an equivalent 1:1:1:1 ratio prior to analysis. All four reporter ions displayed comparable peak intensities, confirming isotopic label balance under controlled conditions. (b) Quantitative crosslinking output from equivalent purified BSA samples showing the number of crosslinked spectra matches, unique protein–protein interactions, and unique crosslinked sites for each channel. (c) Box plot of reporter ion intensities from the same dataset illustrating that, even in a complex proteomic background, reporter ion abundance remains consistent with input ratios and correlates with crosslinking output metrics shown in panel B.

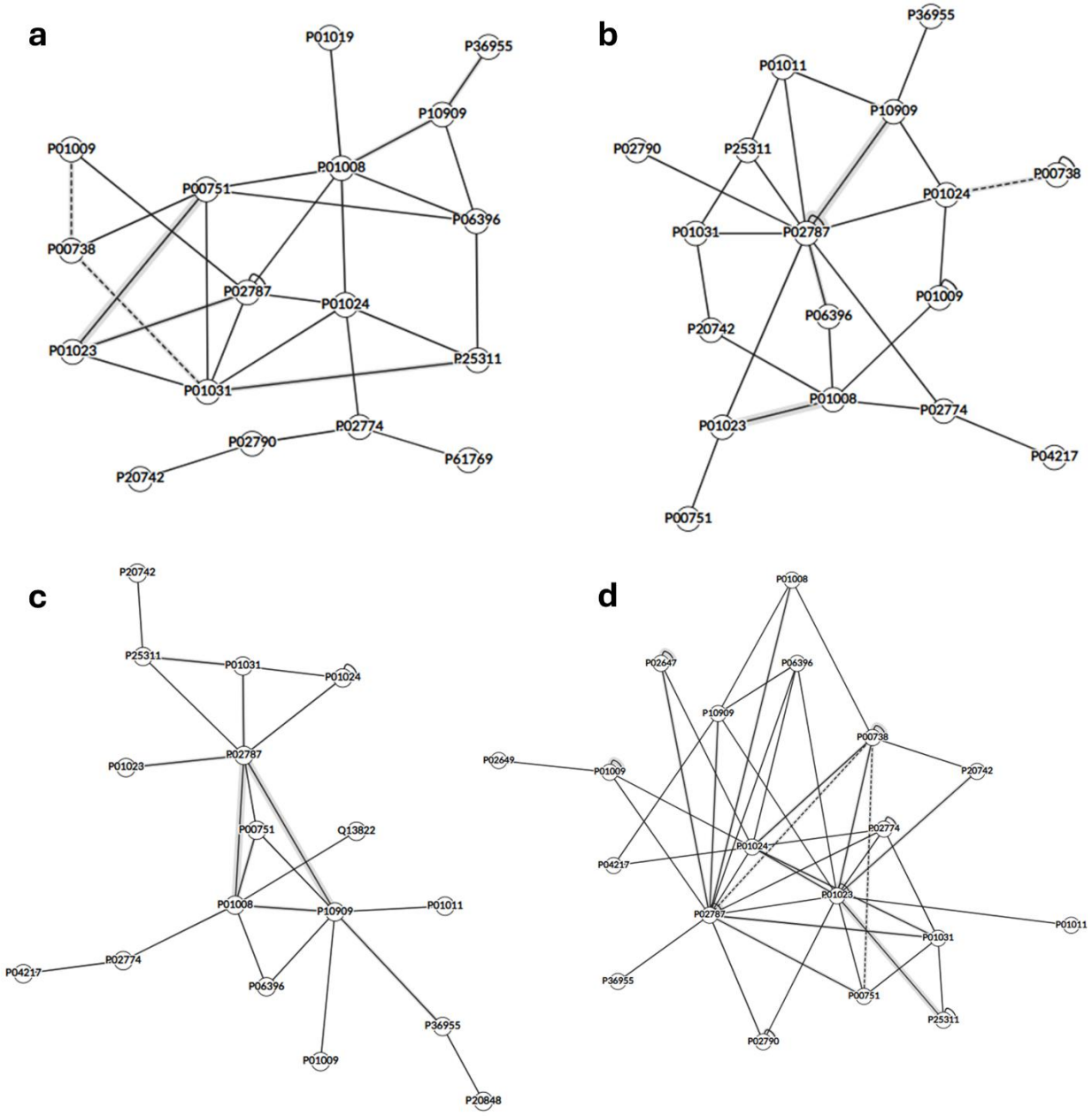


Figure 4. Crosslinking interaction networks in AD and control serum samples identified using 4-plex XLeu. Interaction maps were generated separately for (a) AD female (115), (b) AD male (116), (c) control female (117), and (d) control male (118).

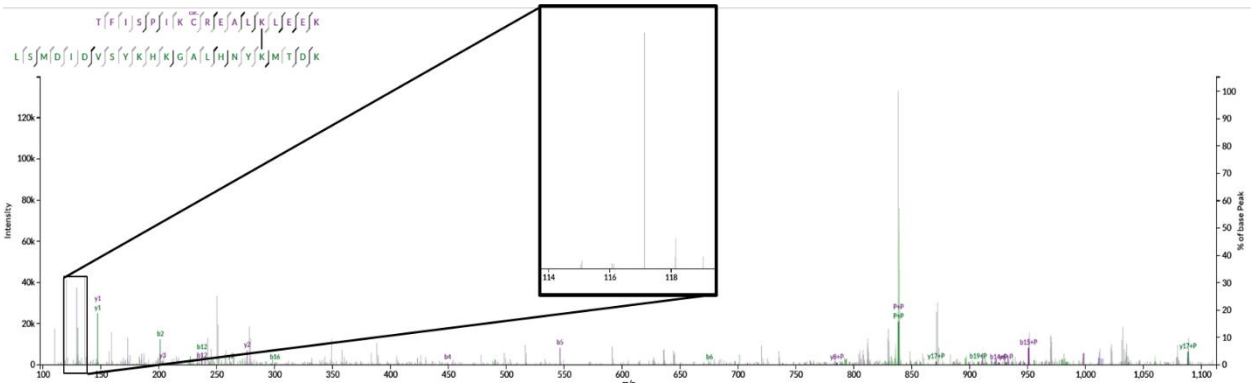


Figure 5. Annotated MS/MS spectrum of the inter-protein crosslink between complement cofactor 3 (CO3) and complement cofactor 5 (CO5). The spectrum highlights fragment ions supporting the interlinked peptides, with the reporter-ion region (m/z 114–118) shown in the zoom-in panel. The strong reporter-ion intensities confirm reliable identification and quantitation of this crosslink.

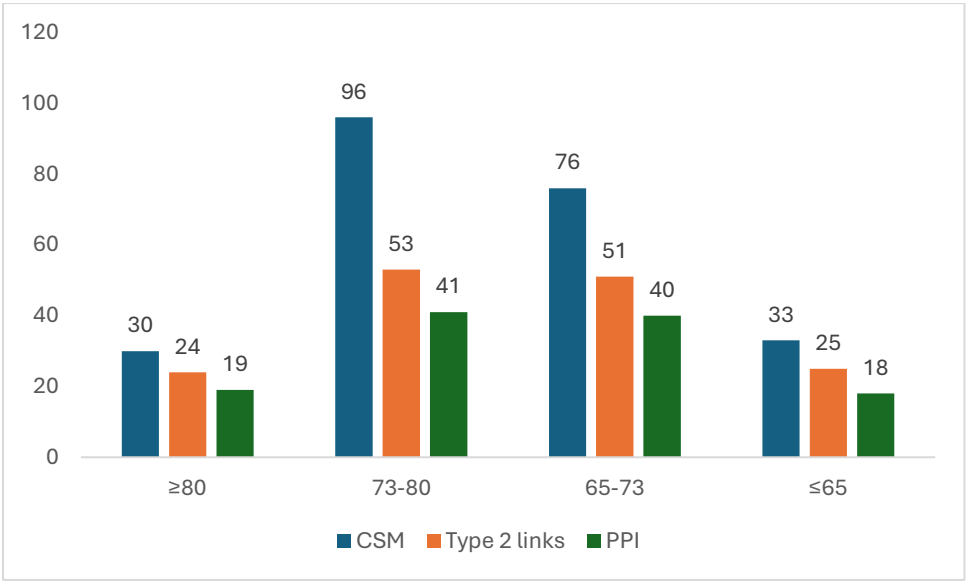


Figure 6. Age-dependent distribution of crosslinked spectrum matches (CSMs), type II interlinks, and protein–protein interactions (PPIs).

Supplementary information

Supplementary figures

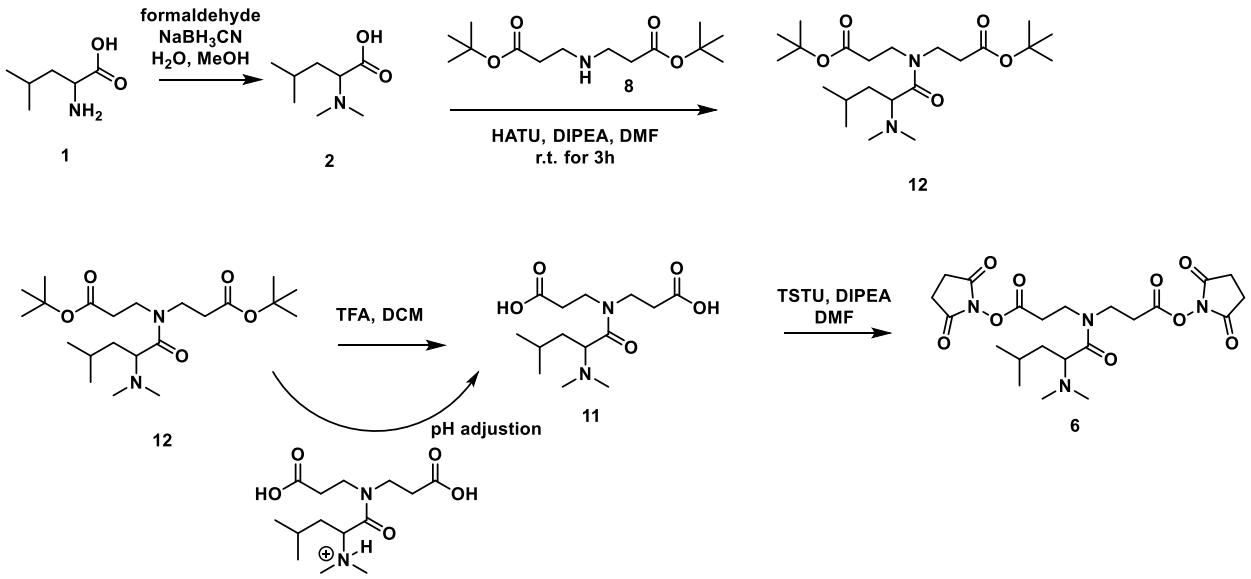


Figure S1. The organic chemistry pathway of the XLeu synthesis.

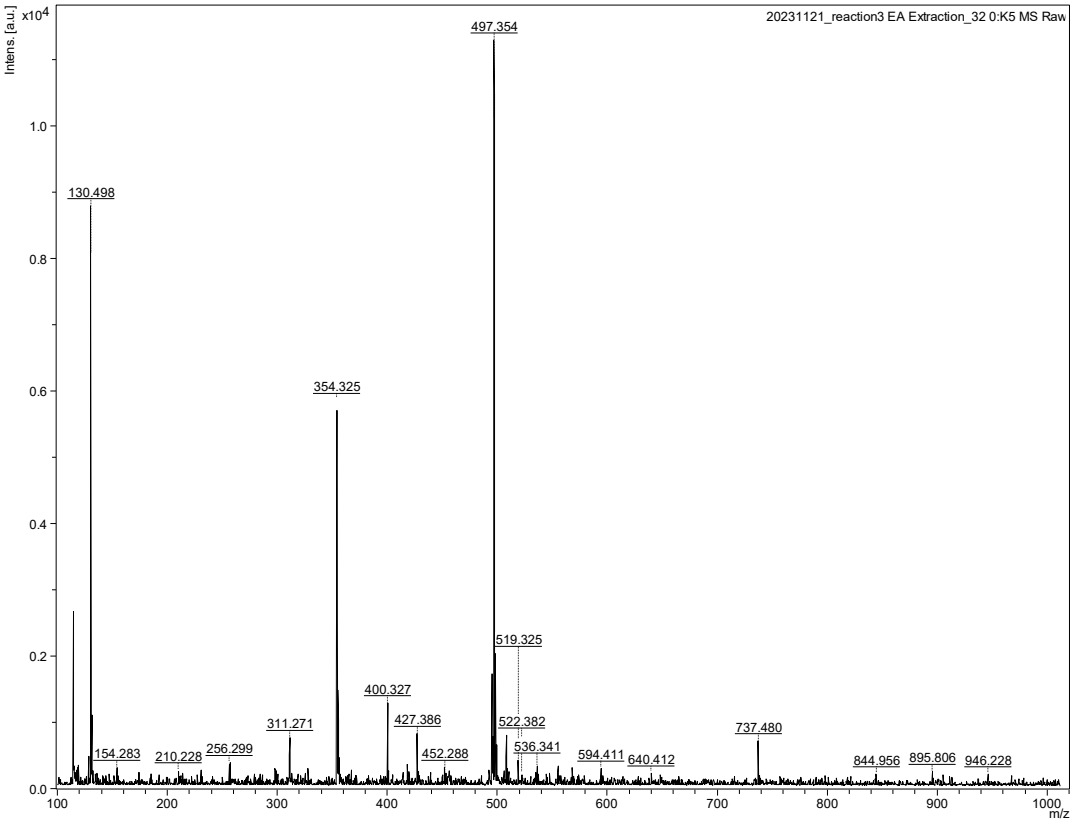


Figure S2. The final product of the XLeu in MALDI-MS.

Supplementary tables

Age Group	LID-nr	Crosslinker	age	gender	Tau	abeta	Ptau	AD stage
≥80	3583	115	80	f	784	378	98	AD
	3709	116	83	m	481	442	76	AD
	3444	117	86	f	178	821	28	non-AD
	3330	118	82	m	386	625	51	non-AD
73-80	3545	115	75	f	1010	329	137	AD
	3611	116	73	m	733	294	87	AD
	3486	117	76	f	303	877	42	non-AD
	3628	118	75	m	382	830	58	non-AD
65-73	3691	115	73	f	687	558	80	AD
	3522	116	71	m	644	442	75	AD
	3713	117	72	f	126	603	22	non-AD
	3482	118	68	m	168	806	25	non-AD
≤65	3511	115	56	f	541	464	80	AD
	3701	116	63	m	881	457	97	AD
	3514	117	64	f	342	1010	49	non-AD
	3478	118	61	m	248	1010	36	non-AD

Table S1. Human serum sample information

Abbreviation	Protein Full Name	Uniprot No.
A1AT	Alpha-1-antitrypsin (SERPINA1)	P01009
A1BG	Alpha-1B-glycoprotein	P04217
A2MG	Alpha-2-macroglobulin	P01023
AACT	Alpha-1-antichymotrypsin	P01011
ANGT	Angiotensinogen	P01019
ANT3	Antithrombin-III	P01008
APOA1	Apolipoprotein A-I	P02647
APOE	Apolipoprotein E	P02649
B2MG	Beta-2-microglobulin	P61769
CFAB	Complement factor B	P00751
CO3	Complement C3	P01024
CO5	Complement C5	P01031
GELS	Gelsolin	P06396
HEMO	Hemopexin	P02790
HPT	Haptoglobin	P00738
PEDF	Pigment epithelium-derived factor	P36955
PZP	Pregnancy zone protein	P20742
TRFE	Serotransferrin	P02787
VTDB	Vitamin D-binding protein	P02774
ZA2G	Zinc-alpha-2-glycoprotein	P25311

Table S2. Top 20 proteins in the crosslinking searching database

Chapter 7

A Dual-MS Approach to Explore the Citrullination Effect on Histone H2 Dimers

Adapted from: **Z. Zhu**, B. Wang, Z. Zhu, S. Xu, L. Li. A Dual-MS Approach to Explore the Citrullination Effect on Histone H2 Dimers. **2025.** In preparation. **Z. Zhu** contributed the experimental design, data collection, analysis, and manuscript preparation.

Abstract

Post-translational modifications (PTMs) such as citrullination play critical roles in regulating chromatin structure, yet their structural effects on histone assemblies remain challenging to define due to the dynamic nature of histone complexes. In this study, we applied an integrated mass spectrometry approach combining crosslinking mass spectrometry (XL-MS) and native ion mobility mass spectrometry (IM-MS) to investigate the structural impact of PAD4-mediated citrullination on the histone H2A–H2B dimer. XL-MS revealed that citrullination nearly doubled the number of inter-subunit crosslinks and shifted their distribution toward residues proximal to known citrullination sites. Structural mapping and distance filtering showed a transition from extended inter-dimer interactions in the unmodified state to more localized intra-dimer interactions following citrullination, suggesting a condensation of the dimer interface. Native IM-MS confirmed this effect, as intact dimer ions were well-resolved in unmodified samples but substantially diminished or undetectable after citrullination, indicating a loss of dimer integrity. Further CIU analysis on an orthogonal platform revealed distinct unfolding patterns and a citrullination-specific CIU50 transition in H2A, suggesting a change in conformational stability. Additionally, the citrullinated monomers exhibited narrower drift-time distributions, indicative of more compact or aggregated gas-phase structures. These findings suggest that PAD4-mediated citrullination not only weakens histone dimer stability but also induces structural reorganization that favors localized interactions within the dimer while disrupting higher-order nucleosome architecture. This work demonstrates the power of combining covalent labeling and gas-phase ion mobility measurements to dissect the structural consequences of histone PTMs, providing mechanistic insights into chromatin remodeling at both the residue and complex level.

Introduction

Histones are core structural proteins that organize DNA into chromatin and regulate essential nuclear functions, including transcription, replication, and repair[1]. Among the core histones, histone H2A and H2B form a heterodimer that associates with the H3–H4 tetramer to construct the nucleosome core particle, a fundamental unit of chromatin architecture[2, 3]. The dimerization of H2A–H2B is not only required for proper nucleosome assembly but also plays a key role in dynamic chromatin remodeling processes[4]. Subtle changes in H2 dimer structure and interaction can significantly alter nucleosome stability and DNA accessibility, implicating the dimer as a focal point in the regulation of gene expression, DNA repair, and epigenetic modulation[5]. Despite its central role, the transient and dynamic nature of H2 dimer formation presents a considerable challenge for structural characterization using traditional high-resolution methods.

Post-translational modifications (PTMs) of histones serve as critical epigenetic regulators that influence chromatin accessibility and gene expression. Among these, citrullination—the enzymatic deimination of arginine residues to citrulline by peptidylarginine deiminases (PADs)—has emerged as a key modulator of chromatin dynamics, particularly in inflammatory and developmental contexts[6, 7]. Citrullination alters the charge and hydrogen bonding potential of arginine side chains, thereby affecting nucleosome stability, histone-DNA interactions, and the recruitment of chromatin-associated proteins. Histone citrullination has been documented across multiple histone classes, including H3, H4, H2A, and H2B, and is often associated with transcriptional activation or chromatin decondensation[8, 9]. In particular, according to the research by Akashi group, citrullination of H2A and H2B has been linked to disrupted dimer formation and impaired nucleosome integrity[10, 11]. The structural consequences of these modifications remain incompletely defined, largely due to the challenges in capturing subtle conformational

shifts in dynamic histone assemblies. Advanced mass spectrometry techniques are well-suited to delineate these structural perturbations by simultaneously probing intermolecular contacts and conformational ensembles in citrullinated histone complexes.

Crosslinking mass spectrometry (XL-MS) has emerged as a powerful tool for mapping protein–protein interactions and structural conformations within native or near-native environments. By covalently capturing transient or flexible interactions through bifunctional crosslinkers, followed by tandem mass spectrometric identification of linked peptides, XL-MS provides valuable distance constraints that inform on protein topology and interface regions[12, 13]. This technique has been successfully applied to elucidate the subunit organization of macromolecular assemblies and to monitor conformational changes upon ligand binding or post-translational modifications. Complementary to XL-MS, native mass spectrometry (native MS) coupled with ion mobility separation (IM-MS) enables the interrogation of intact non-covalent complexes in the gas phase, preserving their stoichiometry and overall architecture. Ion mobility separation, which measures the collisional cross-section (CCS) of ions, provides an orthogonal dimension for probing global shape and conformational heterogeneity[14, 15]. As such, native IM-MS represents a keystone analytical strategy for deciphering the oligomeric states and structural stability of the H2A–H2B dimer under varying biological conditions. In this article, we present an application using crosslink to simulate the structural variation of H2 dimers after the catalysis by peptidylarginine deiminase 4 (PAD4), subsequentially bolster by the native MS validation. Together, these integrative approaches offer a multidimensional view of histone H2 dimerization, encompassing both covalent connectivity and gas-phase structural dynamics.

Experimental design

Histone dimer H2 sample citrullination

Recombinant H2A-H2B and individual monomers purchased from New England Biolab (Ipswich, MA). For the variant, H2A is type 3 (uniport ID Q7L7L0), whereas H2B is type 2E (Q16778). 20 ug of dimer is separately undergone buffer exchange through 20mM $CaCl_2$ solution through 3K molecular weight cutoff (MWCO) filter (Millipore). In the ion mobility testing batch, 50 mM ammonium acetate is included for dimer integrity and ESI direction infusion charging. 2 ug merchandised PAD4 from Cayman (Ann Arbor, MI) is induced to citrullinate dimer for 2 hours under room temperature. Parallel experiments were also performed on the histone octamer (Cayman) under the same conditions.

The synthesis of the biotin thiol tag was performed as previously described[16]. For derivatization, digested histone peptides were combined with 300 µg biotin thiol tag dissolved in 40 µL of 25% trifluoroacetic acid (TFA). A 2,3-butanedione solution was freshly prepared by diluting 1 µL 2,3-butanedione into 114 µL of 25% TFA, and 10 µL of this solution was added to initiate the derivatization reaction. The reaction mixture was vortexed and incubated in the dark at 37 °C for 6 h, then dried under vacuum.

To remove excess tag, strong cation exchange (SCX) cleanup was performed using TopTips (Poly LC). SCX tips were equilibrated three times with 100 µL loading buffer consisting of 50% acetonitrile (ACN), 0.2% formic acid (FA), and 10 mM ammonium formate. Derivatized peptides were resuspended in 300 µL loading buffer, loaded onto the SCX tips, and washed with 150 µL loading buffer eight times. Peptides were eluted with 50 µL of 25% ACN, 0.4 M

ammonium formate, repeated three times. All centrifugation steps were performed at 1500 rpm for 2 min, and eluates were dried in vacuo.

Enrichment of derivatized citrullinated peptides was carried out using streptavidin agarose (Sigma), following previously reported protocols. Beads were washed five times with PBS buffer to remove the storage solution. Peptides were resuspended in PBS and incubated with the beads at room temperature for 2 h with rotation. After incubation, the beads were washed sequentially with PBS (3×), 5% ACN in PBS (3×), and ultrapure water (10×) to remove non-specifically bound peptides. Targeted citrullinated peptides were eluted four times using 80% ACN containing 0.2% TFA and 0.1% FA. The first elution was performed at room temperature, and the remaining three were carried out at 95 °C for 5 min with shaking. Eluates were pooled and dried under vacuum prior to LC-MS analysis.

Survey scans of peptide precursors were acquired in the Orbitrap analyzer over an m/z range of 200–1500 at a resolving power of 60,000 at m/z 200, using an automatic gain control (AGC) target of 5×10^5 and a maximum injection time of 100 ms. For histone peptide analysis, the survey scan range was extended from 150–1500 m/z. Data-dependent MS/MS acquisition was carried out in a stepped higher-energy collisional dissociation (HCD) mode with a 3-second duty cycle, using stepped normalized collision energies of 27, 30, and 33 to optimize fragmentation coverage. MS2 scans were acquired in the Orbitrap at a resolving power of 30,000, with an AGC target of 5×10^4, maximum injection time of 54 ms, and a lower mass cutoff of 120 m/z. Precursors were isolated using a 1.6 Da window, and dynamic exclusion was enabled for 45 seconds within a 10 ppm mass tolerance to minimize redundant sequencing.

For data analysis, the biotin thiol tag-labeled citrullination of arginine was defined as a variable modification with a mass addition of +354.10718 Da. A neutral loss of 303.10752 Da

corresponding to tag cleavage, and two diagnostic fragment ions at 227.08487 Da and 304.11479 Da, were included in the search parameters. Peptide-spectrum matches (PSMs) were filtered to a 1% false discovery rate (FDR) at both peptide and protein levels. Reverse hits and potential contaminant matches were excluded. Additionally, site localization confidence threshold for citrullinated residues was set to $\geqslant 0.95$, and any peptides with C-terminal citrullination were removed from further consideration. All other search settings were kept at default values.

Crosslinking of H2 dimer

Both unmodified H2 dimer and citrullinated H2 was incubated with BS^3d_0 in 20 μM HEPES buffer for 1 hour. The crosslinking reaction was terminated with 50 mM Tris buffer for 20 min. Peptide samples were analyzed using a Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer coupled to a Dionex UltiMate 3000 UPLC system (Thermo Fisher Scientific). Prior to injection, samples were resuspended in 97% Optima water, 3% acetonitrile (ACN), and 0.1% formic acid (FA). Peptides were loaded onto a homemade analytical column (75 μm inner diameter) packed in-house with 18 cm of Bridged Ethylene Hybrid (BEH) C18 particles (1.7 μm, 130 Å pore size; Waters Corporation).

Liquid chromatography was performed using a 137-minute gradient elution, starting from 3% to 95% of mobile phase B (100% ACN with 0.1% FA), at a flow rate of 300 nL/min. The mass spectrometer operated in data-dependent acquisition (DDA) mode using high-energy collisional dissociation (HCD) fragmentation. Full MS scans were acquired in the Orbitrap with a resolving power of 120,000 at m/z 200, scanning from 400 to 1600 m/z, and a maximum injection time of 50 ms. The instrument was set to perform MS/MS scans on precursors selected within a 3-second cycle following each full MS scan. Fragment ion spectra (MS2) were collected with an isolation

window of 2.0 Da, a normalized collision energy (NCE) of 30, Orbitrap resolution of 30,000, and standard AGC target settings. Crosslink raw files were searched via pLink 2 against the UniProt reviewed database based on ID above[17]. The PyXlinkViewer plugin is employed to visualize the crosslinking mapping on the 3D protein structure[18].

Ion mobility testing of the H2 dimer stability

Ion mobility and collision-induced unfolding (CIU) analyses were performed on two platforms. For direct infusion native ion mobility-MS, 3 µL of each sample was introduced into a homemade pulled capillary emitter and infused into a Waters Synapt G2 mass spectrometer (Waters Corp.) operating in positive ion mode. Ionization was conducted at 1.5 kV, and 2 minutes of signal was collected under stable nano-electrospray conditions. Data acquisition was performed using MassLynx software, and deconvolution was carried out by correlating monoisotopic and charge-deconvoluted species to UniProt sequence entries.

For CIU, experiments were conducted on an Agilent 6560 IM-QTOF (Agilent Technologies, Santa Clara, CA) equipped with the CIU module. Samples were pre-equilibrated in the source for 30 seconds before CIU activation. Collision voltage increased from 0 to 300 V in 25 V steps every 0.1 minutes, and ion drift was monitored across the mobility dimension. The sheath gas temperature was set to 150 °C, and optimized ion optics were used to maintain non-covalent structure. CIU fingerprints—representing drift time as a function of collision energy—were visualized using CIUSuite 2 software from Ruotolo Lab, which enabled high-resolution heatmap generation, comparative stability profiling, and conformational cluster analysis across charge states and conditions[19, 20]. CIU feature plots and two-dimensional heatmaps were generated using built-in visualization tools and exported for downstream figure preparation and interpretation.

Result and discussion

Citrullination induced significant site-specific interaction between H2A and H2 monomers

Histones are intrinsically rich in lysine and arginine residues, which play critical roles in maintaining nucleosome integrity through electrostatic interactions. Citrullination, a post-translational modification catalyzed by peptidylarginine deiminase 4 (PAD4), converts arginine to citrulline by deaminating the side chain guanidinium group, resulting in loss of positive charge and thereby disrupting ionic interactions essential for chromatin compaction. Previous studies in our laboratory have established that histone proteins harbor multiple PAD4-catalyzed citrullination sites, particularly within flexible and solvent-accessible regions.

In this study, we prioritized the identification of PAD4-specific citrullination events occurring within the histone H2A–H2B dimer interface under structurally condensed conditions. This site-level characterization was pursued in preparation for downstream structural analysis via crosslinking mass spectrometry (XL-MS) and ion mobility-based stability profiling (**Figure 1**). To achieve this, we employed our previously developed biotin thiol tag strategy, which facilitates derivatization, affinity enrichment, and confident MS-based identification of citrullinated peptides in a bottom-up proteomics workflow[16, 21]. In total, seven citrullinated arginine residues were identified and mapped to histone H2 subunits, with the majority previously reported by our group[22]. Importantly, further comparative analysis revealed that four out of the seven sites are uniquely attributed to PAD4-mediated catalysis: R31 and R89 on the H2A, R80 and R87 on the H2B (**Figure 2**). This disclosure highlights their potential functional relevance within the histone assembly.

During XL-MS analysis, we observed multiple inter-subunit crosslinks within the unmodified native H2A–H2B dimers, primarily formed between the N-terminal region of H2A and the C-terminal region of H2B (**Figure 3a**). Upon PAD4-induced citrullination, the total number of inter-histone crosslinks significantly increased, and the number of the unique matchup pairs was almost doubled, as reflected by both the number of unique peptide pairs and spectral identifications obtained from database searches (**Figure 3b, c**). As presented in the Venn diagram, the number of unique H2A–H2B crosslinked pairs nearly doubled following citrullination, while only a small subset of native crosslinks was lost or altered. Further structural interpretation using 2D sequence mapping revealed prominent crosslinking hotspots, notably at lysine residues K37 and K96 on H2A and K86 and K109 on H2B. Over 55% of the newly established interlinks between H2A and H2B are related to these four "significant crosslinked sites". These sites exhibited marked increases in crosslinking extent in post-citrullination, suggesting enhanced local interaction or structural rearrangement. Notably, these enriched crosslinks are located in close proximity to PAD4-specific citrullination sites, implicating a potential local structural remodeling effect triggered by arginine citrullination within the histone dimer interface.

Proximal citrullination crosslinking sites reveal structural insights into dimer aggregation in histone complex

A critical step in crosslinking mass spectrometry (XL-MS)–based structural interpretation involves the manual mapping of identified crosslinked peptide pairs onto high-resolution 3D protein structures. This spatial validation is essential to confirm the structural plausibility of each crosslink, which must fall within the maximum distance constraints defined by the crosslinking reagent[23, 24]. To ensure structural accuracy, each inter-residue distance was carefully evaluated

against the theoretical Cα–Cα span of the crosslinker. A major analytical challenge in this context lies in distinguishing true intradimer crosslinks from possible interdimer interactions, especially in highly condensed chromatin states where histone dimers may cluster. To address this, we applied a distance cutoff of 32 Å for BS3-derived crosslinks, which reflects a compromise between published optimal working thresholds (~30 Å)[25] and the maximum reach (~35 Å)[26, 27] reported for this homobifunctional NHS ester. This threshold was used to manually curate and filter crosslinks assigned between H2A and H2B subunits, thereby enriching for structurally relevant contacts within individual dimers.

While the previous analysis of cross-linking patterns between H2A and H2B—both in unmodified and citrullinated states—has firmly established the intra-dimer interaction dynamics, these findings were derived primarily from experiments involving isolated H2A-H2B dimers. In this octameric environment, additional spatial constraints and interaction possibilities arise, raising the potential for cross-links not only within individual dimers but also between neighboring H2A-H2B dimers. To fully capture this topological complexity and distinguish these inter-dimer interactions from previously observed intra-dimer contacts, we extended our cross-linking analysis to the full octamer assembly. Though the total number of the cross-link between H2A-H2B were further truncated, the most essential crosslinks related to the proximal citrullination sites mentioned that were mentioned in the previous section are preserved. After filter with distance restraint of 32 Å and visualize the rest crosslinks in cryo-EM structure of the octamer[28], the resulting dataset revealed a more heterogeneous cross-linking pattern, reflective of the dense packing and rotational symmetry within the nucleosome core.

Our structural data revealed that, in unmodified nucleosomes, there exists an extensive crosslinking network bridging adjacent H2A–H2B dimers (**Figure 4A**). In contrast, following

PAD4-mediated citrullination, crosslinking becomes more localized within individual dimers (**Figure 4B**), suggesting that citrullination promotes structural decoupling of the two layers within the histone octamer. All H2B Lys 86 crosslinking has been now replaced by K109 whereas they still connect to H2AK37 and K96, further indicating the potential structural transformation. These findings provide new evidence that arginine citrullination may induce topological stratification within nucleosomes, with potential consequences for chromatin compaction and accessibility.

Ion mobility reveals affected stability of Histone H2 under citrullination

To investigate the structural consequences of citrullination on intact histone H2A–H2B dimers, we conducted ion mobility spectrometry (IMS) using two complementary platforms. In experiments performed on the Waters Synapt G2 system, native-like ionization conditions allowed us to preserve and resolve the intact dimer species in the gas phase, as indicated by a stratified drift time distribution distinct from the monomeric components (**Figure 5a**). This separation confirmed the preservation of noncovalent dimeric integrity in the unmodified histone complex. Following PAD4-mediated citrullination, IMS analysis revealed a clear mass shift in both H2A and H2B monomers, consistent with results from enrichment-based proteomics. However, the corresponding dimeric ion signal was substantially diminished or undetectable, suggesting that citrullination disrupts the ability of the monomers to reassemble into a stable dimeric form in the gas phase (**Figure 5b**). This observation indicates a marked reduction in dimer stability and implies that arginine citrullination compromises inter-subunit interactions, potentially through loss of electrostatic contacts essential for maintaining the native dimer interface.

In our ion mobility experiments using the Agilent 6560 system, the citrullinated H2A-H2B dimer exhibited extremely low signal intensity, posing a significant challenge for direct CIU50 analysis. As a result, we compared the unfolding behavior of citrullinated H2A in its monomeric form to its unmodified counterpart within the dimeric assembly. Notably, during collisional activation, both the dimer and monomer forms exhibited similar dissociation patterns characterized by a redistribution of charge states—where higher charge state signals diminished and lower charge state species became more prominent (Figure S2). Initial unfolding of the H2A-H2B dimer was first detectable at charge state +13 upon reaching 100 V, accompanied by a pronounced shift in drift time from ~30 ms to ~37 ms (**Figure 6a**), indicating a substantial increase in collision cross section (CCS) and the onset of a multi-stage unfolding process. In subsequent analyses, we found that H2B unfolding behavior remained consistent between its dimeric and monomeric forms under equivalent collisional energies and charge states (Figure S3), supporting the assumption that monomer unfolding in isolation faithfully reflects its behavior within the dimeric complex.

Strikingly, upon citrullination, an additional structural transition was observed in H2A at 200 V in the +12 charge state (**Figure 6b**), manifesting as a distinct CIU50 event. This suggests a unique citrullination-induced refolding process not seen in the unmodified protein. Furthermore, the drift-time distribution of citrullinated species showed a tighter clustering of signals compared to the broader distribution seen in the unmodified form, indicating a more compact or aggregated structural ensemble (**Figure 6c**). This observation aligns with our previous cross-linking results, which suggested that citrullinated monomer units within the dimer adopt a more localized, intra-dimer association pattern rather than forming extensive inter-dimer crosslinks.

Conclusion

In this study, we developed and applied a dual-mass spectrometry strategy—integrating crosslinking mass spectrometry (XL-MS) and native ion mobility spectrometry (IM-MS)—to dissect the structural impact of PAD4-mediated citrullination on the histone H2A–H2B dimer. We successfully identified PAD4-specific citrullination sites using biotin thiol derivatization and demonstrated that these modifications significantly enhance local crosslinking within dimers while reducing inter-dimer connectivity in the nucleosome context. Our structural mapping efforts revealed a spatial correlation between enriched crosslinks and proximal citrullination sites, suggesting localized conformational remodeling. Furthermore, native ion mobility analysis confirmed a substantial decrease in dimer stability and revealed unique unfolding transitions exclusive to the citrullinated forms. These results indicate that citrullination disrupts electrostatic integrity at the dimer interface, driving histone stratification and aggregation. Overall, this work underscores the power of integrated MS techniques to resolve structural perturbations driven by PTMs and provides novel insights into the molecular consequences of citrullination in chromatin biology.

References

1. Kornberg, R. D.; Lorch, Y., Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **1999,** *98* (3), 285-94.
2. Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **1997,** *389* (6648), 251-260.
3. Arents, G.; Burlingame, R. W.; Wang, B. C.; Love, W. E.; Moudrianakis, E. N., The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci U S A* **1991,** *88* (22), 10148-52.
4. Kouzarides, T., Chromatin modifications and their function. *Cell* **2007,** *128* (4), 693-705.
5. Clapier, C. R.; Iwasa, J.; Cairns, B. R.; Peterson, C. L., Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology* **2017,** *18* (7), 407-422.
6. Nakashima, K.; Arai, S.; Suzuki, A.; Nariai, Y.; Urano, T.; Nakayama, M.; Ohara, O.; Yamamura, K.; Yamamoto, K.; Miyazaki, T., PAD4 regulates proliferation of multipotent haematopoietic cells by controlling c-myc expression. *Nat Commun* **2013,** *4*, 1836.
7. Tanikawa, C.; Espinosa, M.; Suzuki, A.; Masuda, K.; Yamamoto, K.; Tsuchiya, E.; Ueda, K.; Daigo, Y.; Nakamura, Y.; Matsuda, K., Regulation of histone modification and chromatin structure by the p53-PADI4 pathway. *Nat Commun* **2012,** *3*, 676.
8. Christophorou, M. A.; Castelo-Branco, G.; Halley-Stott, R. P.; Oliveira, C. S.; Loos, R.; Radzisheuskaya, A.; Mowen, K. A.; Bertone, P.; Silva, J. C.; Zernicka-Goetz, M.; Nielsen, M. L.; Gurdon, J. B.; Kouzarides, T., Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* **2014,** *507* (7490), 104-8.
9. Wang, Y.; Wysocka, J.; Sayegh, J.; Lee, Y. H.; Perlin, J. R.; Leonelli, L.; Sonbuchner, L. S.; McDonald, C. H.; Cook, R. G.; Dou, Y.; Roeder, R. G.; Clarke, S.; Stallcup, M. R.; Allis, C. D.; Coonrod, S. A., Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science* **2004,** *306* (5694), 279-83.
10. Shimoyama, S.; Nagadoi, A.; Tachiwana, H.; Yamada, M.; Sato, M.; Kurumizaka, H.; Nishimura, Y.; Akashi, S., Deimination stabilizes histone H2A/H2B dimers as revealed by electrospray ionization mass spectrometry. *J Mass Spectrom* **2010,** *45* (8), 900-8.
11. Saikusa, K.; Shimoyama, S.; Asano, Y.; Nagadoi, A.; Sato, M.; Kurumizaka, H.; Nishimura, Y.; Akashi, S., Charge-neutralization effect of the tail regions on the histone H2A/H2B dimer structure. *Protein Sci* **2015,** *24* (8), 1224-31.
12. Sinz, A., Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein–Protein Interactions: Where Are We Now and Where Should We Go from Here? *Angewandte Chemie International Edition* **2018,** *57* (22), 6390-6396.
13. Leitner, A.; Faini, M.; Stengel, F.; Aebersold, R., Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends in Biochemical Sciences* **2016,** *41* (1), 20-32.
14. Benesch, J. L. P.; Ruotolo, B. T., Mass spectrometry: come of age for structural and dynamical biology. *Current Opinion in Structural Biology* **2011,** *21* (5), 641-649.
15. Leney, A. C.; Heck, A. J. R., Native Mass Spectrometry: What is in the Name? *Journal of the American Society for Mass Spectrometry* **2017,** *28* (1), 5-13.

16. Shi, Y.; Li, Z.; Wang, B.; Shi, X.; Ye, H.; Delafield, D. G.; Lv, L.; Ye, Z.; Chen, Z.; Ma, F.; Li, L., Enabling Global Analysis of Protein Citrullination via Biotin Thiol Tag-Assisted Mass Spectrometry. *Anal Chem* **2022,** *94* (51), 17895-17903.
17. Chen, Z. L.; Meng, J. M.; Cao, Y.; Yin, J. L.; Fang, R. Q.; Fan, S. B.; Liu, C.; Zeng, W. F.; Ding, Y. H.; Tan, D.; Wu, L.; Zhou, W. J.; Chi, H.; Sun, R. X.; Dong, M. Q.; He, S. M., A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat Commun* **2019,** *10* (1), 3404.
18. Schiffrin, B.; Radford, S. E.; Brockwell, D. J.; Calabrese, A. N., PyXlinkViewer: A flexible tool for visualization of protein chemical crosslinking data within the PyMOL molecular graphics system. *Protein Sci* **2020,** *29* (8), 1851-1857.
19. Dixit, S. M.; Polasky, D. A.; Ruotolo, B. T., Collision Induced Unfolding of Isolated Proteins in the Gas Phase: Past, Present, and Future. *Current Opinion in Chemical Biology* **2018,** *42*, 93-100.
20. Polasky, D. A.; Dixit, S. M.; Fantin, S. M.; Ruotolo, B. T., CIUSuite: A Quantitative Analysis Package for Collision Induced Unfolding Measurements of Gas-Phase Protein Ions. *Analytical Chemistry* **2019,** *91* (4), 3147-3155.
21. Li, Z.; Wang, B.; Yu, Q.; Shi, Y.; Li, L., 12-Plex DiLeu Isobaric Labeling Enabled High-Throughput Investigation of Citrullination Alterations in the DNA Damage Response. *Anal Chem* **2022,** *94* (7), 3074-3081.
22. Wang, B.; Li, Z.; Shi, Y.; Zhu, Z.; Fields, L.; Shelef, M. A.; Li, L., Mass Spectrometry-Based Precise Identification of Citrullinated Histones via Limited Digestion and Biotin Derivative Tag Enrichment. *Anal Chem* **2024,** *96* (6), 2309-2317.
23. Ding, Y. H.; Gong, Z.; Dong, X.; Liu, K.; Liu, Z.; Liu, C.; He, S. M.; Dong, M. Q.; Tang, C., Modeling Protein Excited-state Structures from "Over-length" Chemical Cross-links. *J Biol Chem* **2017,** *292* (4), 1187-1196.
24. Bullock, J. M. A.; Thalassinos, K.; Topf, M., Jwalk and MNXL web server: model validation using restraints from crosslinking mass spectrometry. *Bioinformatics* **2018,** *34* (20), 3584-3585.
25. Merkley, E. D.; Rysavy, S.; Kahraman, A.; Hafen, R. P.; Daggett, V.; Adkins, J. N., Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci* **2014,** *23* (6), 747-59.
26. Wang, X.; Cimermancic, P.; Yu, C.; Schweitzer, A.; Chopra, N.; Engel, J. L.; Greenberg, C.; Huszagh, A. S.; Beck, F.; Sakata, E.; Yang, Y.; Novitsky, E. J.; Leitner, A.; Nanni, P.; Kahraman, A.; Guo, X.; Dixon, J. E.; Rychnovsky, S. D.; Aebersold, R.; Baumeister, W.; Sali, A.; Huang, L., Molecular Details Underlying Dynamic Structures and Regulation of the Human 26S Proteasome. *Mol Cell Proteomics* **2017,** *16* (5), 840-854.
27. Fux, A.; Korotkov, V. S.; Schneider, M.; Antes, I.; Sieber, S. A., Chemical Cross-Linking Enables Drafting ClpXP Proximity Maps and Taking Snapshots of In Situ Interaction Networks. *Cell Chem Biol* **2019,** *26* (1), 48-59 e7.
28. Ariyoshi, M.; Makino, F.; Watanabe, R.; Nakagawa, R.; Kato, T.; Namba, K.; Arimura, Y.; Fujita, R.; Kurumizaka, H.; Okumura, E. I.; Hara, M.; Fukagawa, T., Cryo-EM structure of the CENP-A nucleosome in complex with phosphorylated CENP-C. *EMBO J* **2021,** *40* (5), e105671.

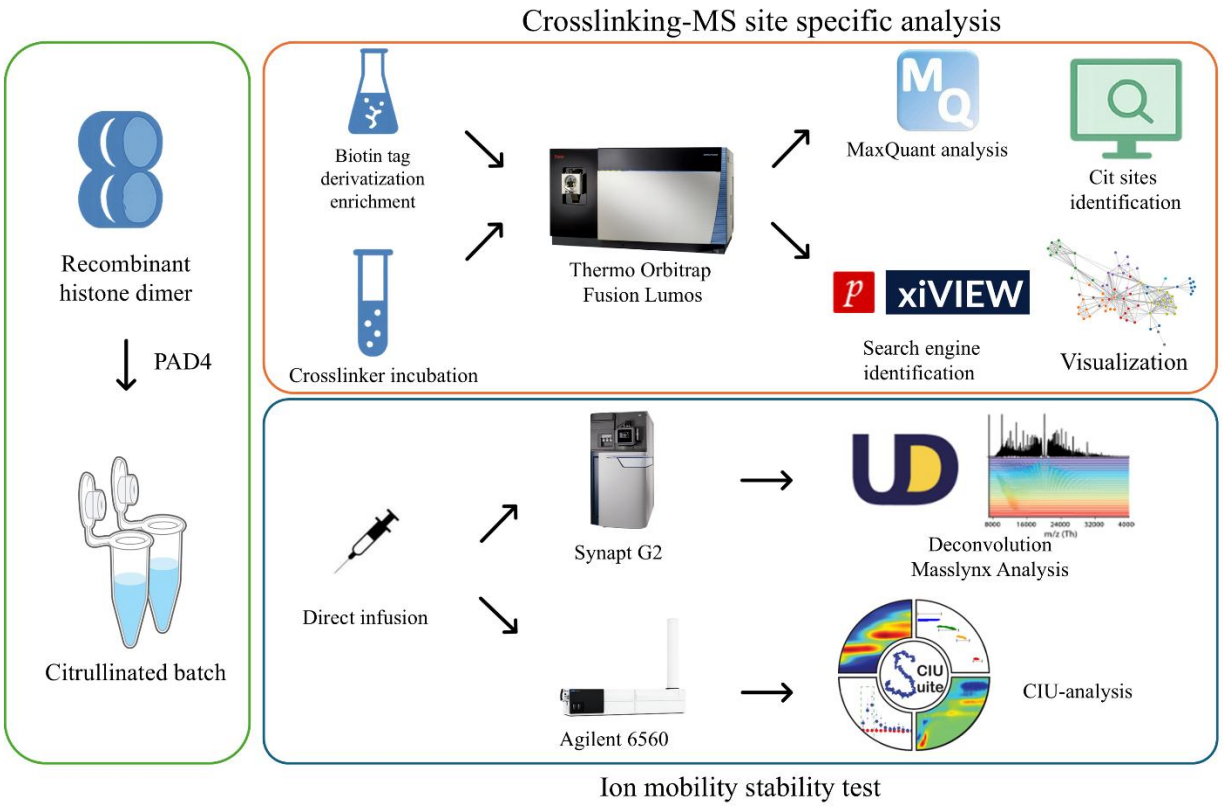


Figure 1. The workflow of the parallel structural analysis of citrullination effect to H2A-H2B interaction within dimer via crosslinking and ion mobility MS.

Figure 2. Identification of Histone H2A-H2B citrullination marks.

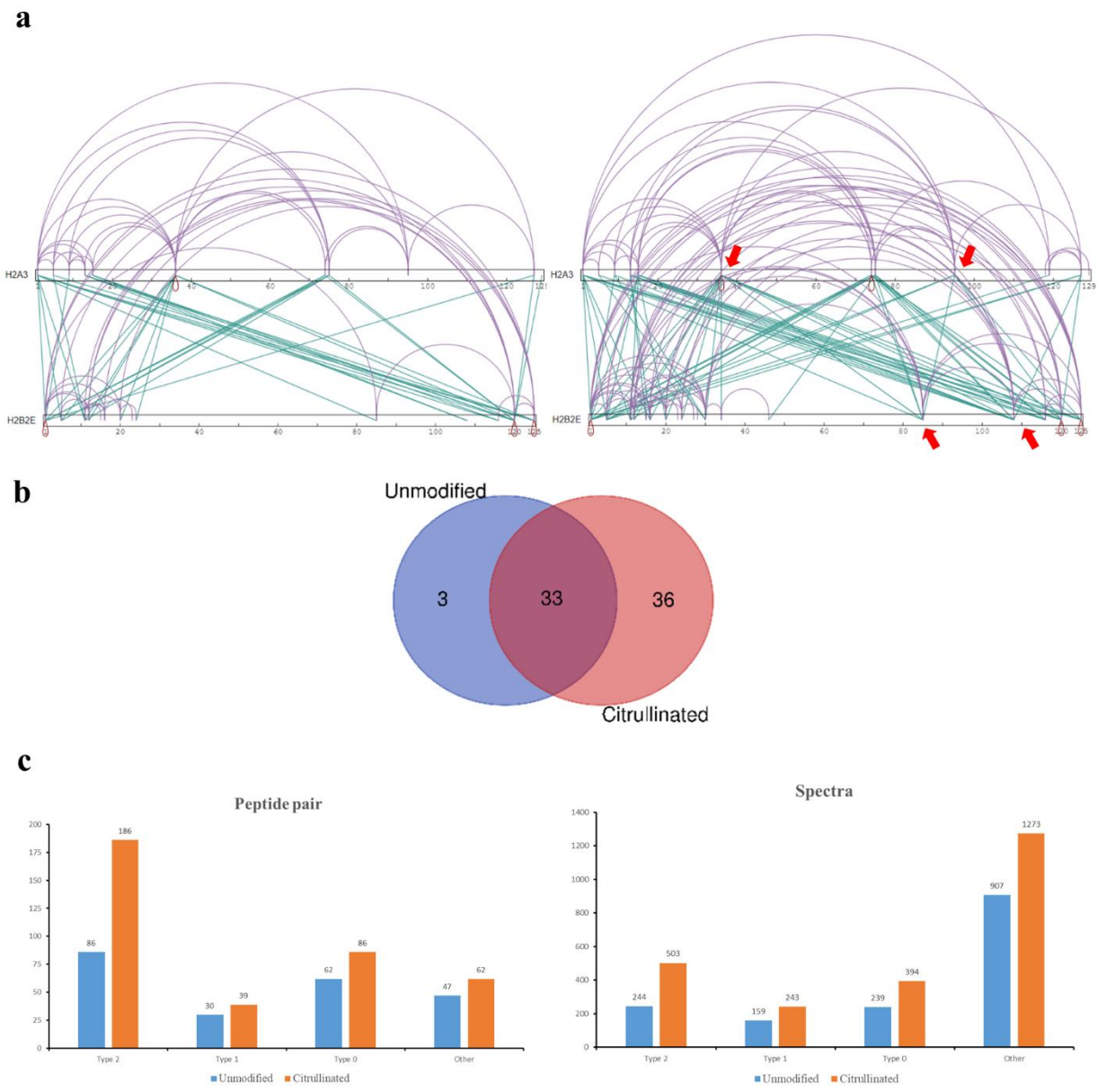


Figure 3. Crosslinking between H2A-H2B dimer. a) The unmodified (left) and the citrullinated (right) crosslinking 2D mapping visualized on xiView platform. The red arrowhead highlighted the significantly increased interlinked lysine residues between H2A-H2B. b) Venn diagram of the unique interlink matchup pairs between H2A and H2B. c)The statistic summary of all types of crosslink identified by searching software.

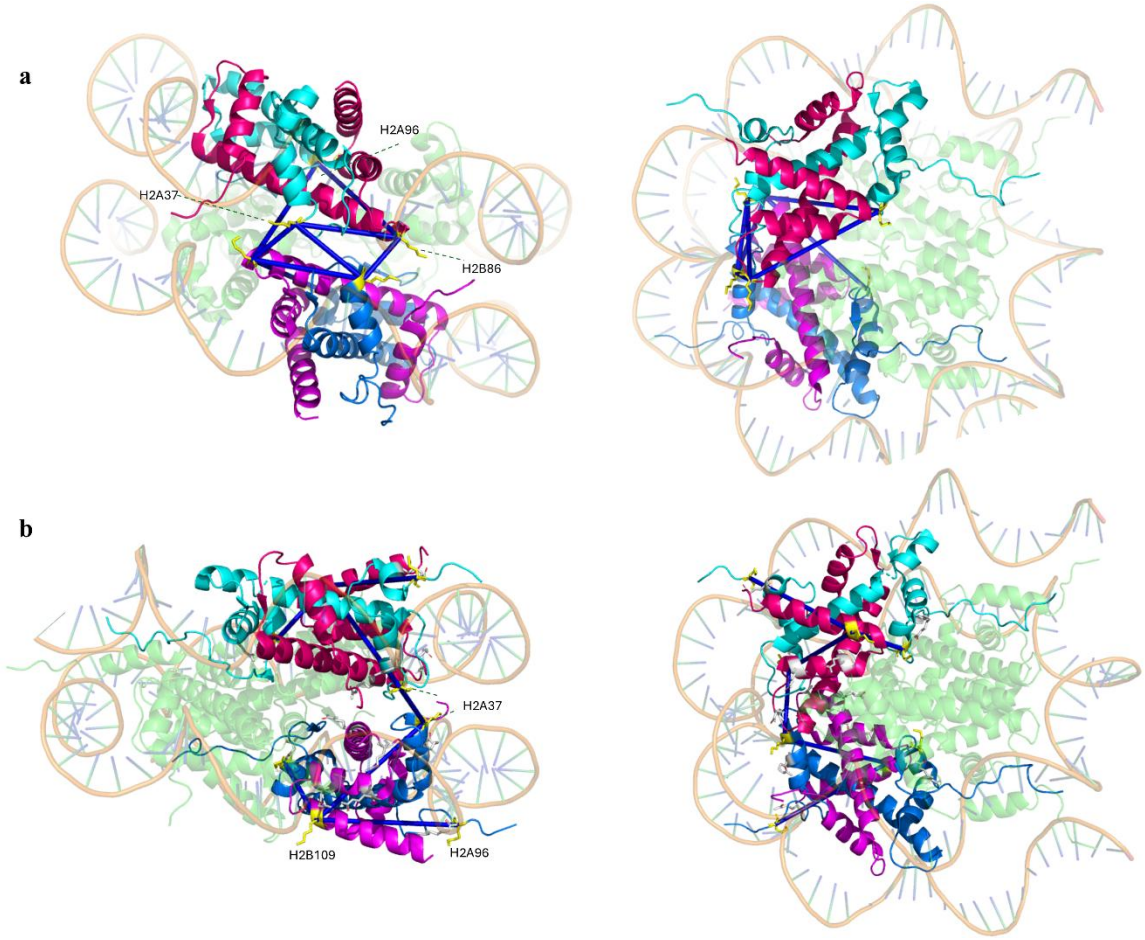


Figure 4. Visualization of histone H2 dimers in nucleosome complex 3D model. View from rear (left) and top (right) of a) unmodified and b) citrullinated nucleosome complex (PDB: 7BXT) with the filtered crosslink related to these "significant crosslinked sites", which highlighted in yellow. The red/purple are H2As and indigo/blue are H2Bs. The citrullinated lysine residues were colored in silver.

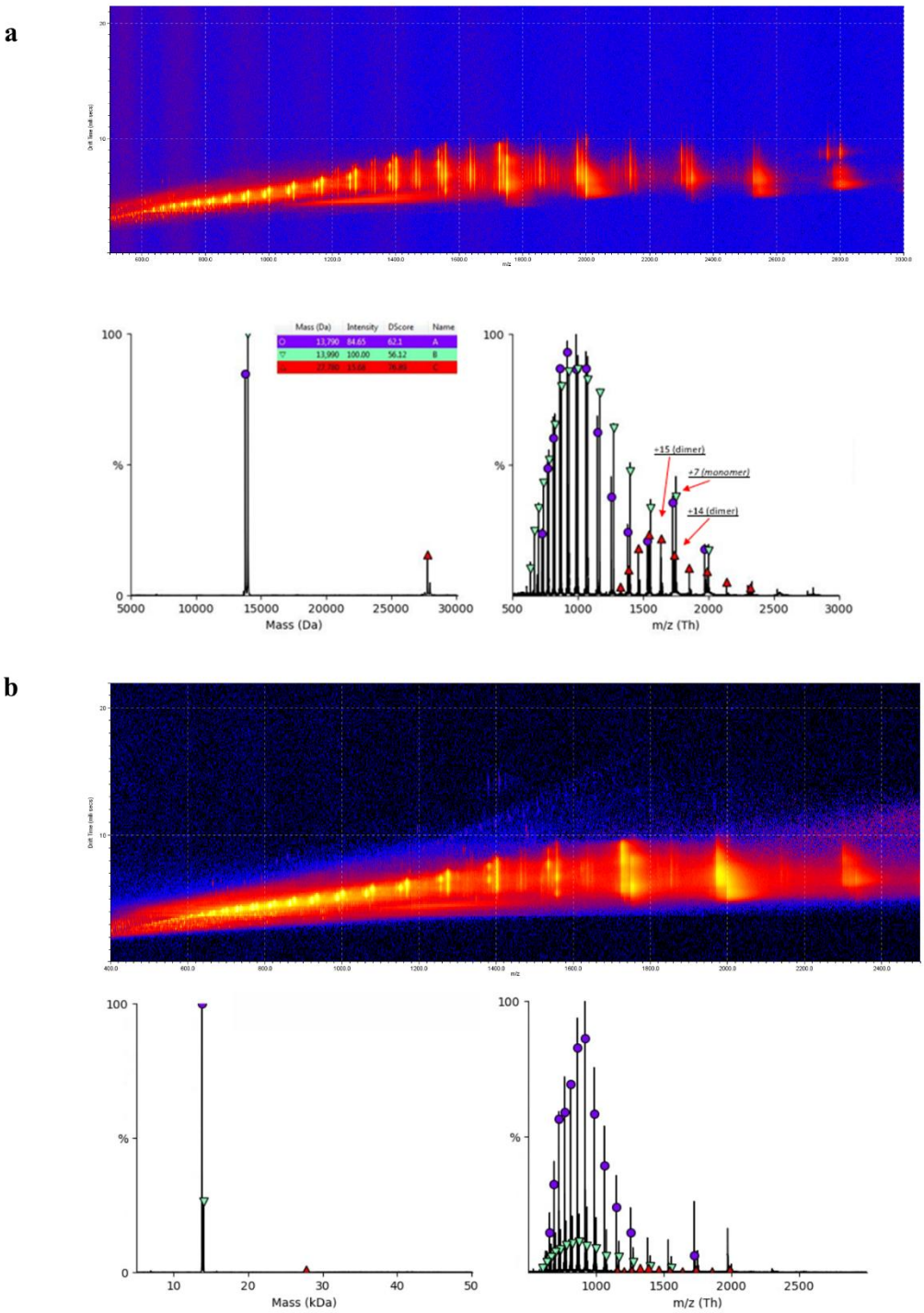


Figure 5. Recombinant H2 dimer stability effect by citrullination under Synapt G2. a) The unmodified and b) the citrullinated masslynx graph (top) with drift time variation and corresponding deconvoluted result (bottom).

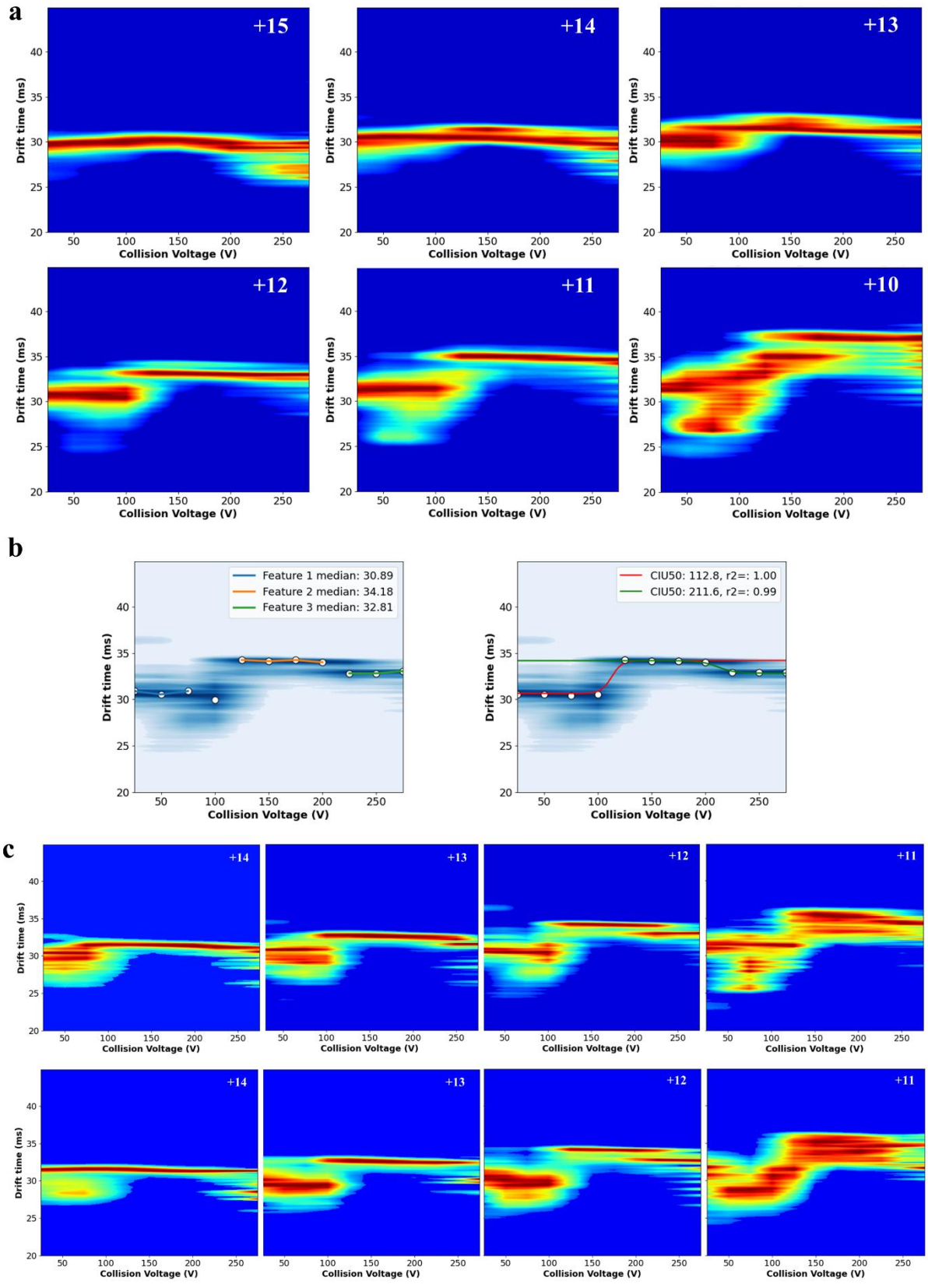

a
+15
+14
+13
+12
+11
+10
Drift time (ms)
Collision Voltage (V)
b
Feature 1 median: 30.89
Feature 2 median: 34.18
Feature 3 median: 32.81
CIU50: 112.8, r2=: 1.00
CIU50: 211.6, r2=: 0.99
c
+14
+13
+12
+11

Figure 6. CIU analysis of the H2 dimer and corresponding citrullinated monomers. a) The dissociation of the unmodified dimer in six different charge states. b) The CIU50 analysis of the H2A monomer in state +12. c) Comparison between the unmodified state of H2A in dimer (top) and citrullinated state of H2A (bottom) in four different charge states.

Supplementary Information

Supplementary figure

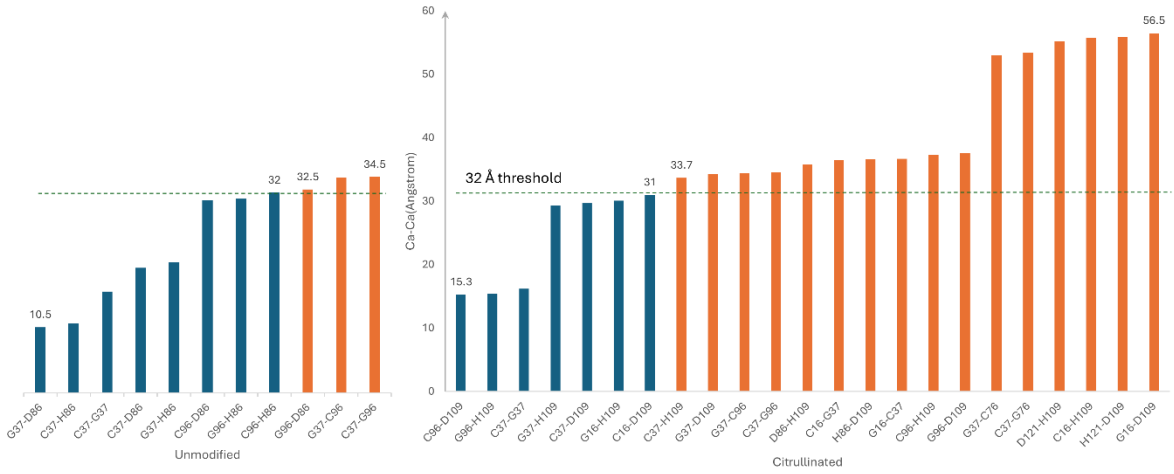


Figure S1. The filter of the potential interlink associated with "significant crosslinked sites'. According to the PDB structure, chain C, G are H2As whereas chain D,H are H2Bs. We preserved the crosslinks under distance restraint.

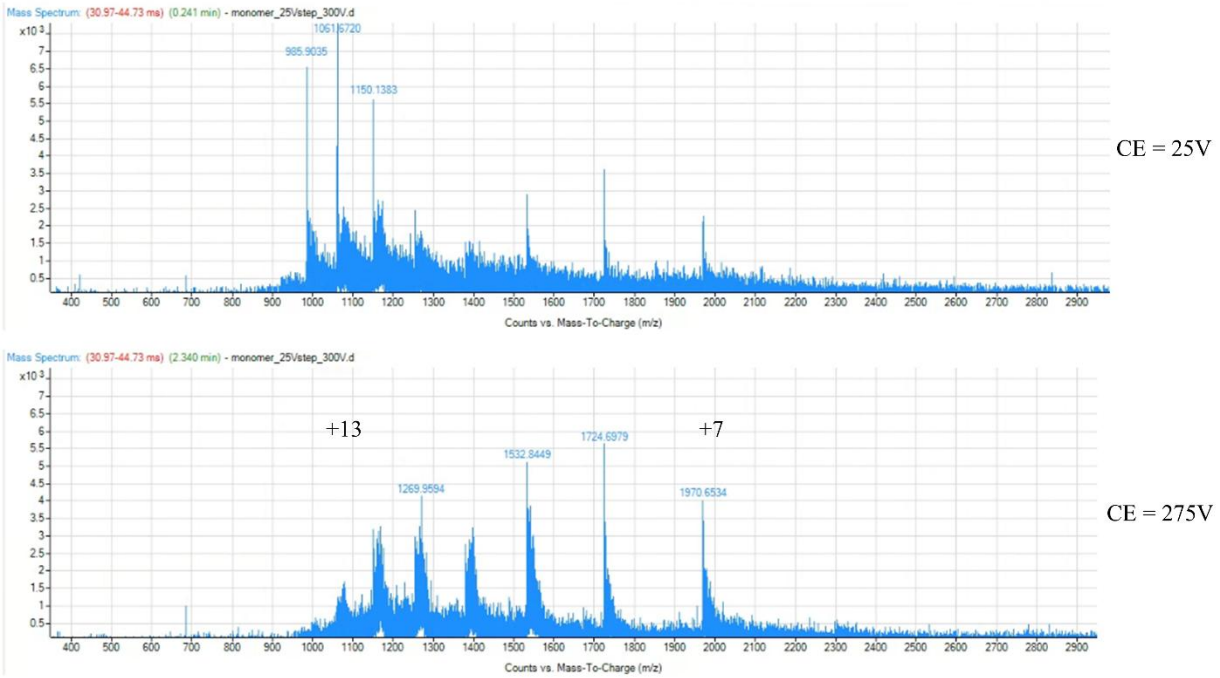


Figure S2. The spectrum of unmodified H2B in two different charge states.

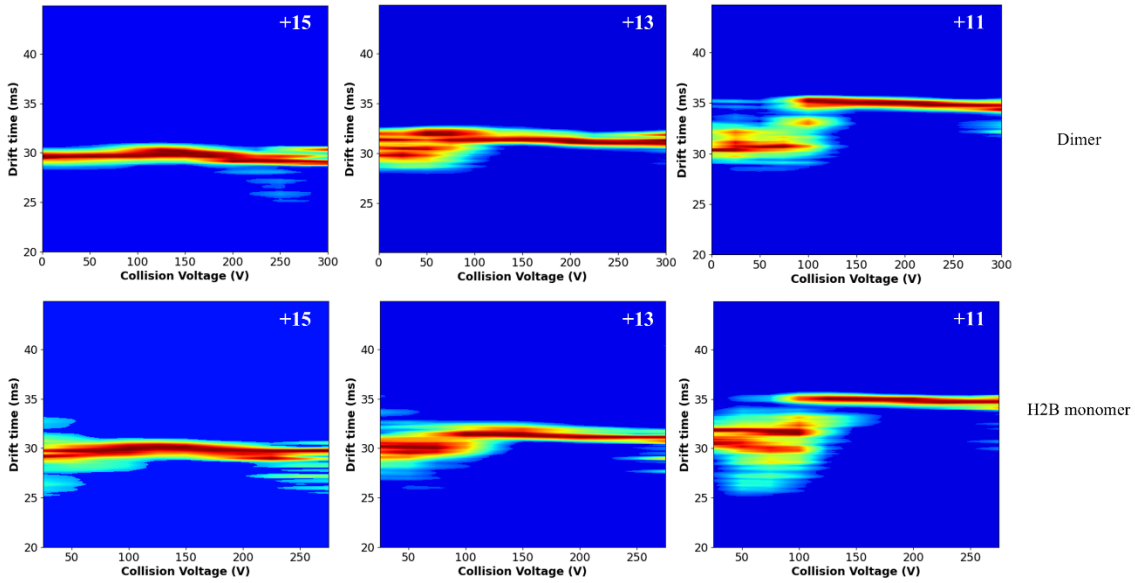
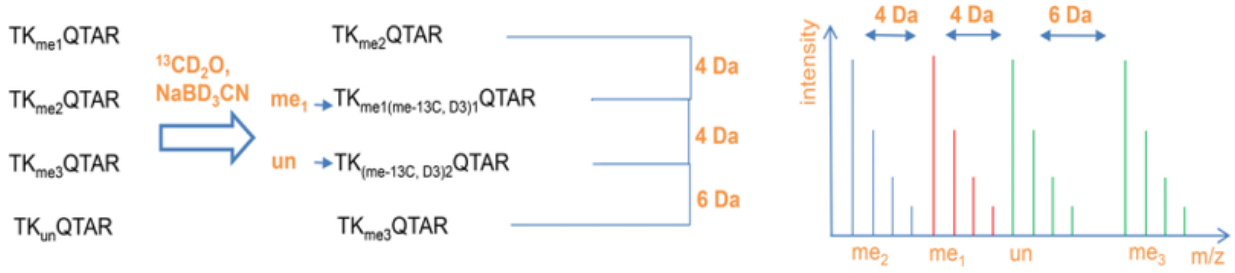


Figure S3. Comparison of the H2B unfolding within dimer and in its monomer state.

Chapter 8

Eliminating lysine-arginine cyclization during dimethyl labeling to improve accuracy of histone methylation stoichiometric analysis



Adapted from: Huang, J., Chan, N., **Zhu, Z.** et al. *Angewandte Chemie* **2025,** In Revision**. Zhu, Z.** Contributed to experimental design, data collection and manuscript preparation.

Abstract

Histone lysine methylation plays important roles in chromatin-based biological processes, such as transcription and DNA repair. Within this work, stable isotope dimethyl labeling was employed to block free lysine residues while simultaneously enabling accurate relative quantitation. As well, dimethyl labeling provided ideal quantitative conditions, inspiring least differences in charge state or retention time between un-, mono-, di-, and trimethylated tryptic peptides. Heavy isotopic dimethyl labeling (13C, D3) also enables the discrimination of free lysines from native lysine mono-, di-, and trimethylated peptides during LC-MS/MS analysis. Moreover, we further investigated and verified a common side reaction of typical dimethyl labeling protocols that is the arginine and lysine cyclization and strengthened our methodology by optimizing reaction conditions that eliminates this side reaction without undermining labeling efficiency. We used this improved method to quantify changes in histone lysine methylation in MCF7 cells with or without knocking down CTR9 and revealed a significant increase of H3K27me3 upon loss of CTR9. This discovery guided us to investigate the mechanism of regulation of H3K27me3 by CTR9 and find potential therapeutic vulnerability for breast cancer treatment.

Introduction

Histone lysine methylation is an important post-translational modification (PTM) that plays critical roles in numerous biological processes with abnormal histone lysine methylation having been associated with developmental defects and human diseases[1]. The primary amine of all lysine residues can be mono-, di-, or trimethylated and the extent of methylation at a single site is shown to generate unique protein function[2]. Moreover, histone lysine methylation can activate or

repress transcription depending on which sites are modified and to what degree[3]. Therefore, a comprehensive stoichiometric analysis of histone lysine methylation is necessary to fully elucidate its function[4]. As histones are lysine-rich and highly-hydrophilic proteins (**Figure S1**), trypsin digestion of histones results in small, hydrophilic peptides which often suffer from substantial losses during sample preparation and are difficult to detect in conventional reversed phase LC-MS[5]. Additionally, trypsin fails to cleave histone proteins when lysine residues are methylated, resulting in inaccurate estimations of site occupancy and methylation extent[6]. Previously, lysine propionylation labeling has been. developed to overcome this drawback and facilitate quantitative analysis of PTMs that frequently occur on the lysine residue of histones[7]. However, propionylation labeling of unmodified and monomethylated lysine residues causes a mismatch in hydrophobicity and charge state between labelled and nonlabelled di-/trimethylated peptides. This mismatch forces retention time and signal intensity differences that lead to inaccurate quantitation and miscalculation of site occupancy[8]. In this work, we developed a method that facilitates blocking of free lysine groups and discrimination of native lysine methylation, enables accurate calculation of the histone lysine methylation stoichiometry. Herein, we introduced dimethyl labeling to block free lysine (K) residues[9]. Dimethyl labeling as a reductive methylation of protein/peptide N-terminus and K ε-amino groups has long been a promising method for isotopic labelling of proteins/peptides in mild reaction conditions[10]. This labeling procedure can also add a methyl group on native mono-methylated K[11]. Therefore, after dimethyl labeling, un- and mono-methylated peptides are equivalent in composition to di-methylated peptides, and their hydrophobicity and charged state will be analogous to native di- and tri-methylated peptides (**Figure 1a**). Additionally, the mass shift that results from incorporation of heavy isotopic dimethyl labeling (13C, D3) of K residues enables unambiguous

distinction between these analytes and native mono-, di-, and trimethylated residues in LC-MS/MS analyses (**Figure 1b**).

Experimental design

In this work, we substituted protein level propionylation with dimethyl labeling. First, we compared the effects of dimethyl and propionylation labeling on standard peptides. Aliquots of histone H3 peptide (23-34) were subjected to dimethyl and propionylation labeling, mixed with equivalent amount of nonlabelled peptides, and subjected to LC-MS analysis. As shown in Figure 1C, dimethyl labeling resulted in negligible changes in analyte intensity, while propionylation labeling significantly enhanced the intensity of labelled peptides. MALDI-TOF analyses further demonstrate the adverse effects of propionylation labeling, displaying increased peptide intensity while suppressing dimethyl labelled and nonlabelled peptides when mixed with equivalent amount (**Figure S2-3**, **Table S1**). Though logical to conclude the increasing intensity of labelled peptides would be beneficial in the identification of highly-hydrophilic, low-abundance analytes in LC-MS-based proteome analyses[12], significant disparities in signal intensity adversely affect site occupancy accuracy in protein PTMs analysis. As these disparities are observed through propionylation labeling but not through dimethyl labeling, the latter strategy provides the most facile means for accurate quantitation of K methylation was selected as the method-of-choice for protein-level labeling moving forward.

However, we continued to leverage the increased hydrophobicity enabled through propionylation labeling by employing this strategy in peptide level after tryptic digestion (Figure S4). In this way, we enable unambiguous discrimination and quantitation of methylation site occupancy

through dimethyl labeling while increasing analyte retention and signal intensity in LC-MS analysis. Using histone H3 peptide (3-8) as an example, without N-terminal propionylation, the peptide is too hydrophilic to be retained by C18 HPLC column and was only detected at trace levels with a retention time 0-8 min (**Figure S5a**). After N-terminal propionylation, the peptide retention time changed to 19-20 min and displayed ideal peak shape, which demonstrates the improved peptide retention on C18 HPLC (**Figure S5b**). We then evaluated the most common and widely used dimethyl labeling protocol in comparison of all four modification types of standard histone H3 peptides (23-34)[13]. As shown in Figure S6, dimethyl labeling efficiency is almost 100% in all four peptides; however, a -6 Da side product was found in abundance during analysis of all four peptides on MALDI-TOF (**Figure S6**, **Figure 2**). We initially attributed the side products to the impurity of isotopic reagents. For this to be true, however, the side products of mono , di-, and tri-methylated peptides should be different from the unmodified peptide, stemming from the addition of different quantities of isotopic methyl groups; this truth does not agree with our observations.

Turning to the original references that detail the dimethyl labeling procedure, the authors report that N-terminal cyclization occur during this procedure[14]. Formaldehyde used during labeling reacts with the N-terminal amino group to form a reactive imine adduct, which then reacts with the less reactive amide bond to generate a methylene bridge. However, this side reaction forms 4-imidazolidinone that can only generate 2-3 Da mass shifts in different isotopic labeling procedures (Figure 3). To evaluate the side reactions of dimethyl labeling we studied the other two peptides, both of which contain one K in the middle of peptide sequence. Results demonstrated that peptide TSYKFESV generates -3 Da mass difference, which is in accordance with N-terminal cyclization side reaction (**Figure 4a**, **Figure S7a-b**). The other peptide

RPKPQQFFGLLM, generates 6 Da difference which is the same with histone H3 peptides (23 34) (**Figure 4b**, **Figure S7c-d**). We hypothesized that, in addition to the N-terminal cyclization, other amino acids may cyclize with K residue. This hypothesis is supported by Jiskoot et al., who indicated that arginine (R) residue readily cyclized with K residue[15]. The cyclization can form two isomers of triazine ring between K and R residues through dual methylene bridges (Figure 4c). The formation of first methylene bridge is similar to the N-terminal cyclization that formaldehyde reacts with primary amine of K to form the first intermediate imine, followed by nucleophilic attack of the guanidine group of R. Subsequently, secondary amine of K reacts with formaldehyde to generate another imine, which combines with either of the two-residue amines of guanidine group to form the stable triazine rings. The dual methylene bridges have -6 Da mass difference compared with dimethyl labeling. This side reaction has also been confirmed and used to staple native peptides for modulating the structure and properties of peptides[16].

We confirmed this side reaction by performing trypsin digestion. After dimethyl labeling, K residues of histone H3 peptide (23-24) are blocked, leaving R as the sole tryptic site (**Figure 5**). If R residue was cyclized, it would resist trypsin digestion. **Figure 5d** demonstrated that the normal dimethyl labelled peptide is fully cleaved after trypsin digestion. However, the side reaction peptide was still present. This demonstrated that K-R cyclization is the major side product of dimethyl labeling in histone K methylation analysis.

Fortunately, the side reaction can be avoided by optimizing reaction conditions. In standard dimethyl labeling protocols, the amine-free and trypsin digestion compatible reagent, triethyl ammonium bicarbonate (TEAB), is used as a labeling buffer with reaction pH 8, with formaldehyde added before the reductant, $NaBH_3CN$[9, 13]. Adding formaldehyde prior to other reagents facilitates the reaction of imine intermediates with guanidine groups without presence of

NaBH3CN, although the extent is low. Therefore, premixing formaldehyde and NaBH3CN before adding to the labeling buffer will allow the immediate reduction of imine and avoid its reaction with intermolecular guanidine group. As shown in Figure 6b, premixing formaldehyde and NaBH3CN resulted in a significant reduction of the side reaction peak. A 2-fold increase in the concentration of NaBH3CN also mitigated the formation of side products (**Figure 6c**). Moreover, Claridge et al. has reported that the formation of methylene bridges is more prevalent under basic condition[17]. Therefore, the TEAB buffer is not suitable for dimethyl labeling and the inclusion of a weak acid buffer may help to eliminate the side reaction without sacrificing the reaction efficiency[18]. We replaced the TEAB buffer with phosphate buffer at pH 6 and premix labeling reagents with 2-fold reductant prior to addition to the reaction solution. Results demonstrate that side reaction peaks are completely eliminated using optimized condition without undermining labeling efficiency (**Figure 6d**, **Figure S8**).

We then applied this optimized dimethyl labeling protocol to the quantitative histone methylation analyses. We used MCF7 as the model system and knocked down CTR9 (**Figure 7a**). CTR9 is a subunit in Paf1c, a complex regulating of RNA Pol II-mediated transcription. Global analyses of histone modifications revealed a significant increase of H3K27me3 upon loss of CTR9 (**Figure 7b**). This discovery guided us to investigate the mechanism of regulation of H3K27me3 by CTR9. We continually found that CTR9 depletion renders breast cancer cells adducted H3K27me3 for survival and hypersensitive to EZH2 (catalytic subunit of PRC2 for H3K27me3) inhibitors[19]. The mechanism of regulation of H3K27me3 by CTR9 is likely conserved across cell types, and CTR9-dependent response to EZH2 inhibitors provides therapeutic vulnerability for breast cancer treatment.

In conclusion, we introduced heavy stable isotope dimethyl labeling as a method of blocking free lysine residues to generate compositionally analogous tryptic peptides that may be readily distinguished from natively methylated lysine residues. After tryptic digestion, all peptide N-termini were labelled with propionylation to increase their hydrophobicity, thus, the sample loss of small and hydrophilic peptides was eliminated. Moreover, we have uncovered a substantial side reaction, K-R cyclization, that results from standard dimethyl labeling protocols. We optimized reaction conditions to overcome these issues and quantify changes in histone lysine methylation in MCF7 cells with or without knocking down CTR9. Our results showed a significant increase of H3K27me3 upon loss of CTR9. This discovery guided us to investigate the mechanism of regulation of H3K27me3 by CTR9 and find potential therapeutic vulnerability for breast cancer treatment.

Acknowledgements

This work was supported in part by grant funding from the NIH (R01AG052324, R01AG078794, and R01DK071801 to L.L., and R01 CA213293 and R01 CA236356 to W. X.), as well as NIH/NCI P30CA014520-UW Carbone Cancer Center Grant. The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531) and the University of Wisconsin-Madison, Office of the Vice Chancellor for Research and Graduate Education. LL 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.

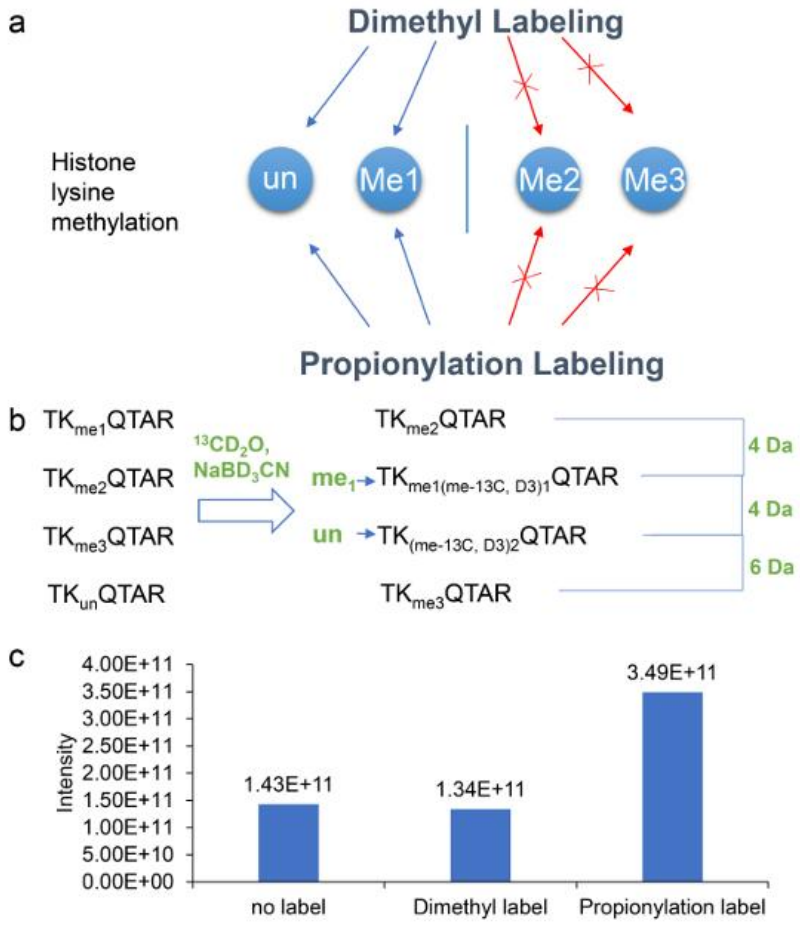


Figure 1. Comparison of dimethyl labeling and propionylation labelling. a) Both dimethyl labeling and propionylation labelling occurs on unmodified and monomethylated peptides. b) Heavy isotopic dimethyl labeling (13C, D3) enables the discrimination of free K with native mono-, di-, and tri-methylated K during LC-MS/MS analysis. c) LC-MS analysis of histone H3 peptide (23-34) with or without chemical labels.

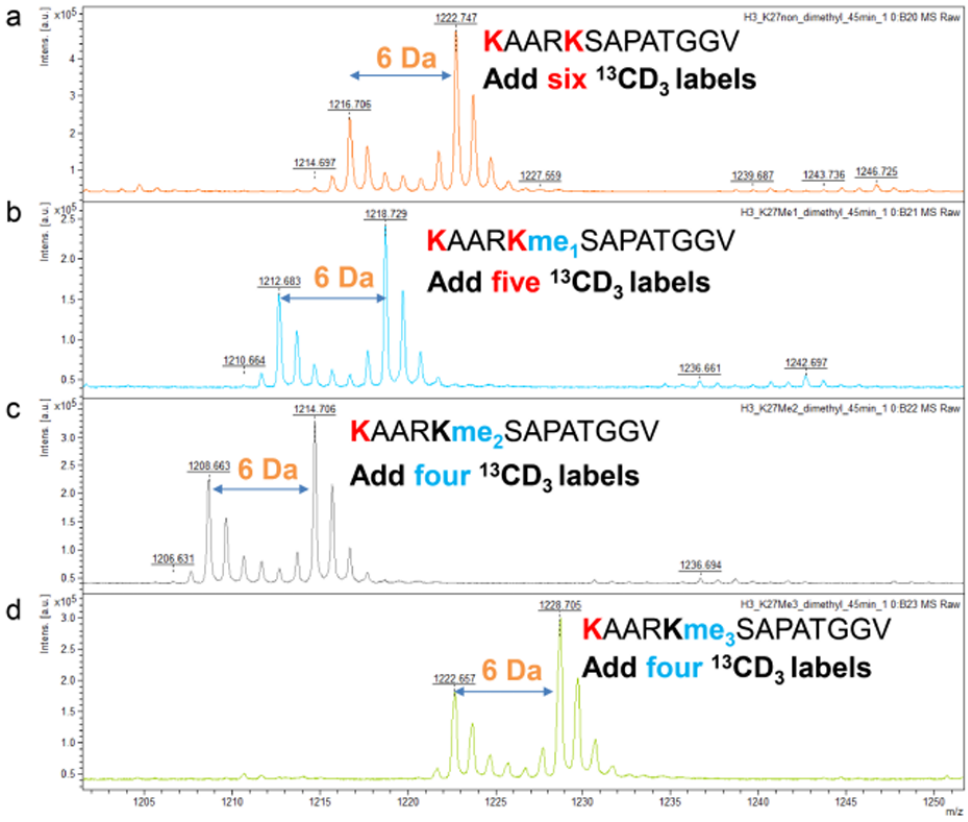


Figure 2. Evaluation of the widely used dimethyl labeling protocol by using all four modification types of standard histone H3 peptides (23-34). a) unmodified peptide b) mono-methylated peptide c) di-methylated peptide d) tri-methylated peptide.

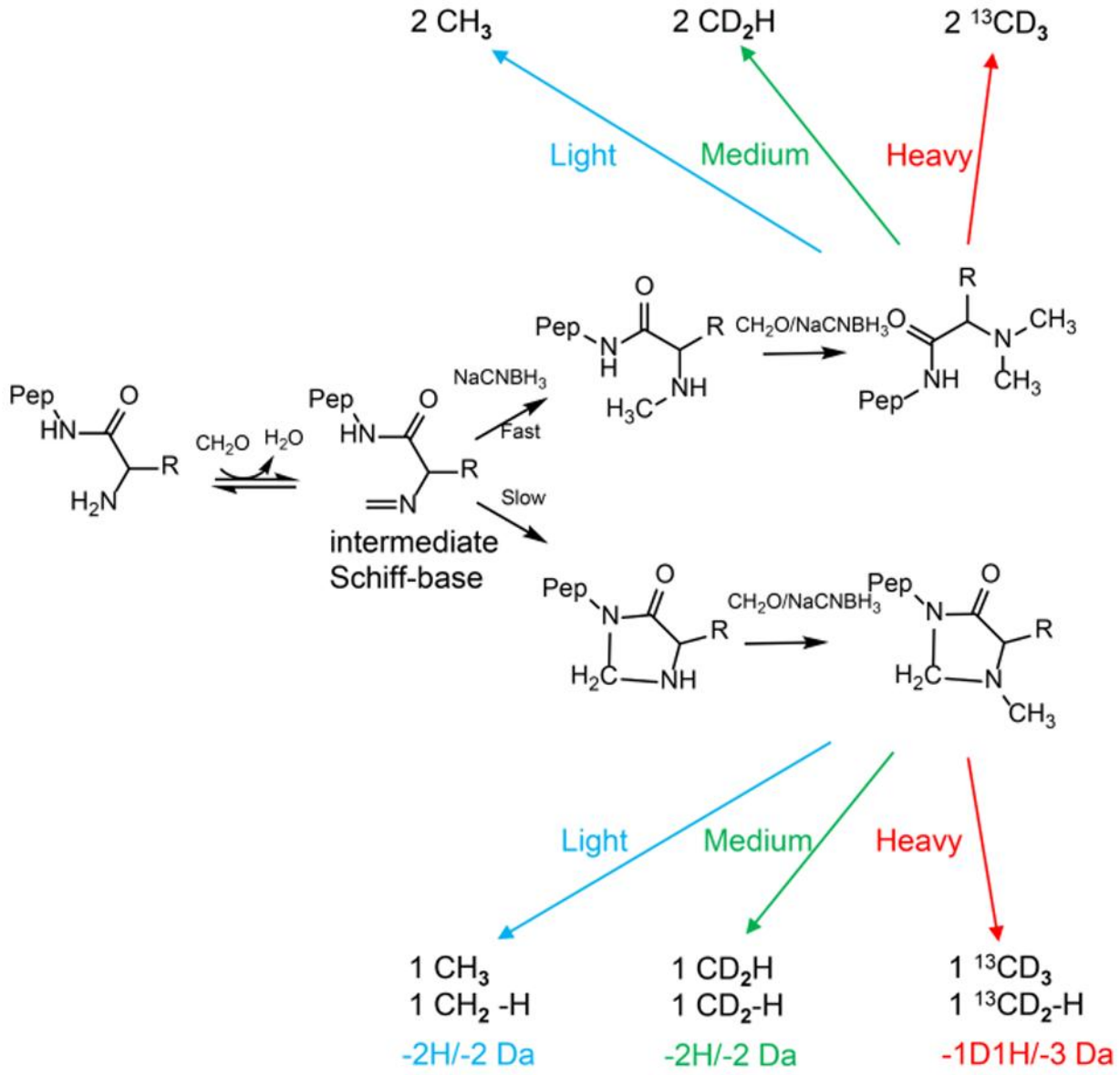


Figure 3. N-terminal cyclization (4-imidazolidinone) formed during dimethyl labeling procedure probably through the intermediate Schiff-base.

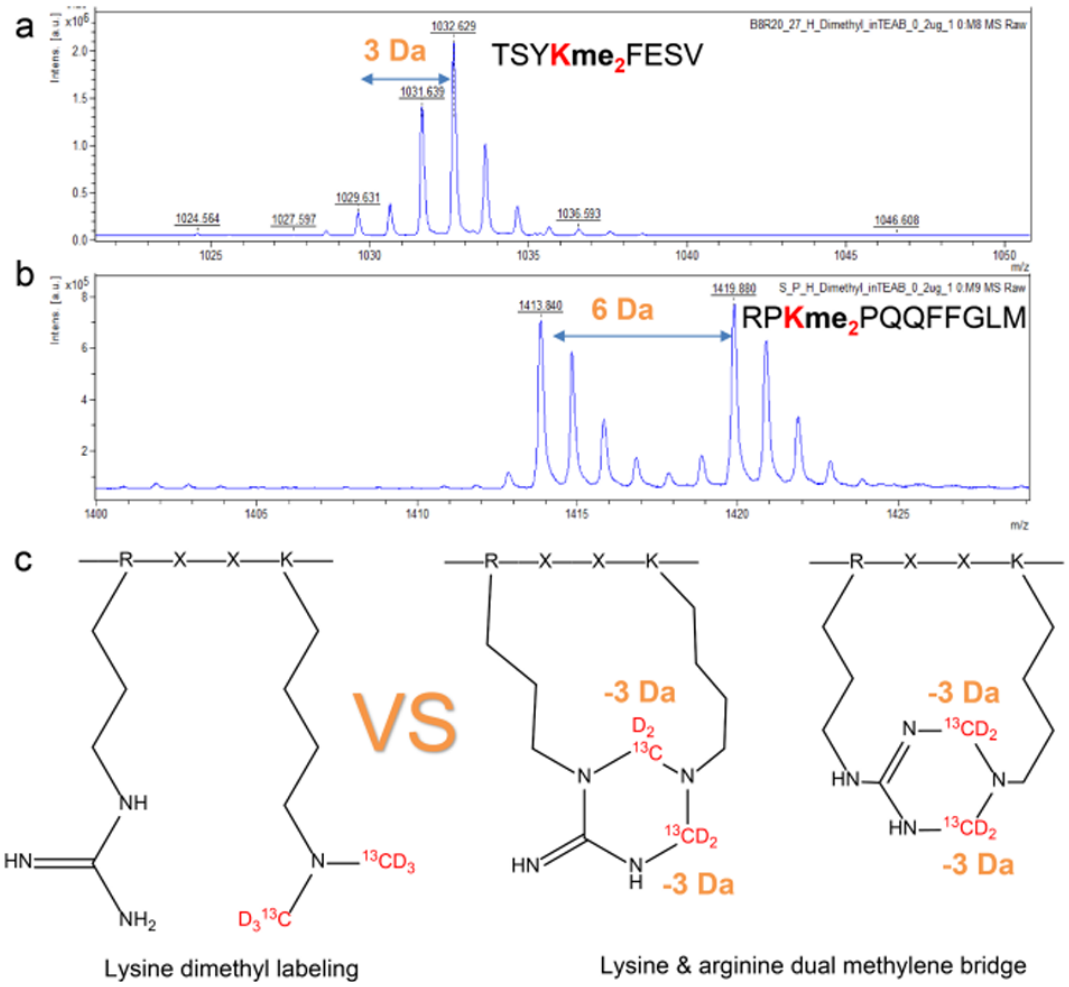


Figure 4. Evaluation of dimethyl labeling protocol with two unique peptides to determine the potential side reaction mechanism. MALDI-TOF spectra showing a) dimethyl labelled TSYKFESV with a -3 Da mass difference side product on and b) dimethyl labelled RPKPQQFFGLLM shows -6 Da mass difference side product. c) Structures of normal heavy dimethyl labelled K and side products K R cyclization.

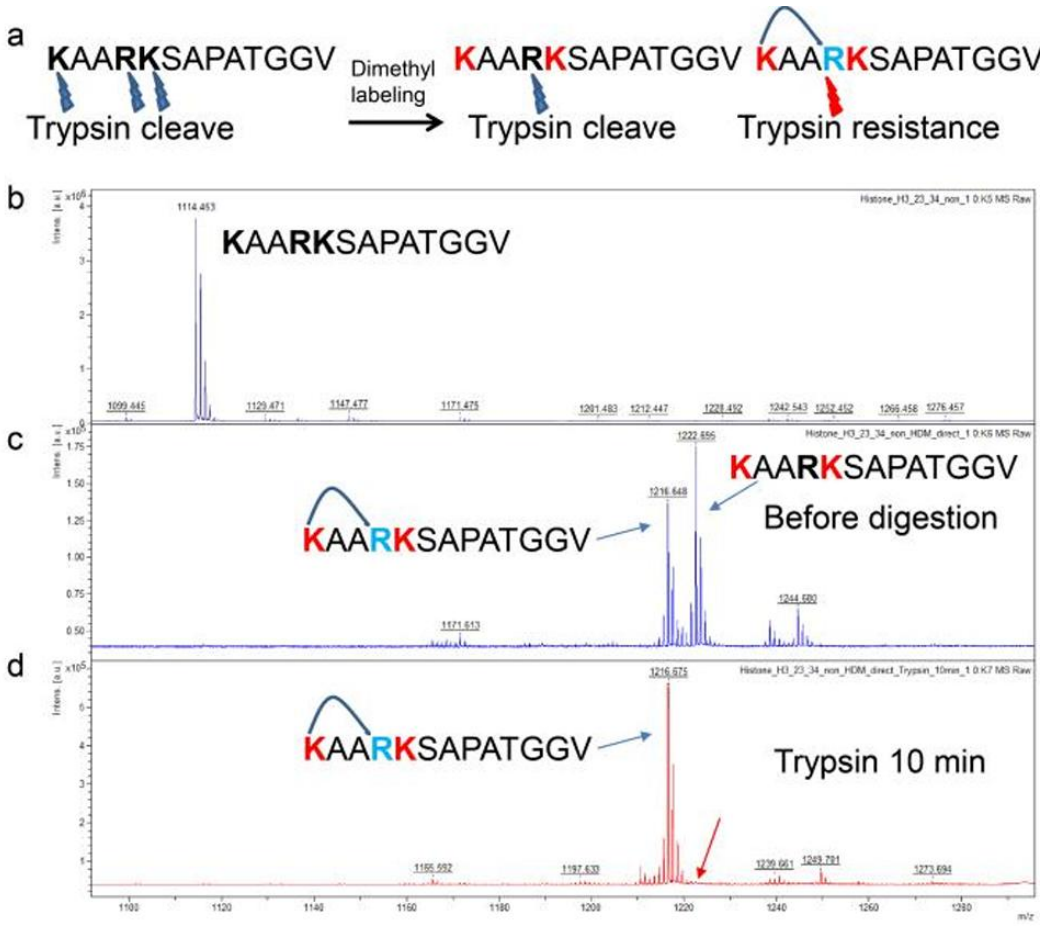


Figure 5. Confirmation of K-R side reaction via trypsin digestion. a) Trypsin cleavage site of histone H3 peptide (23-34) before and after dimethyl labeling. b) MALDI-TOF spectrum of histone H3 peptide (23-34) before dimethyl labeling. c) MALDI-TOF spectrum of dimethyl labelled histone H3 peptide (23-34) and its -6Da side product before trypsin digestion. d) Normal dimethyl labelled histone H3 peptide (23-34) is fully cleaved after digestion while the side product is still resistant to trypsin on MALDI-TOF spectrum.

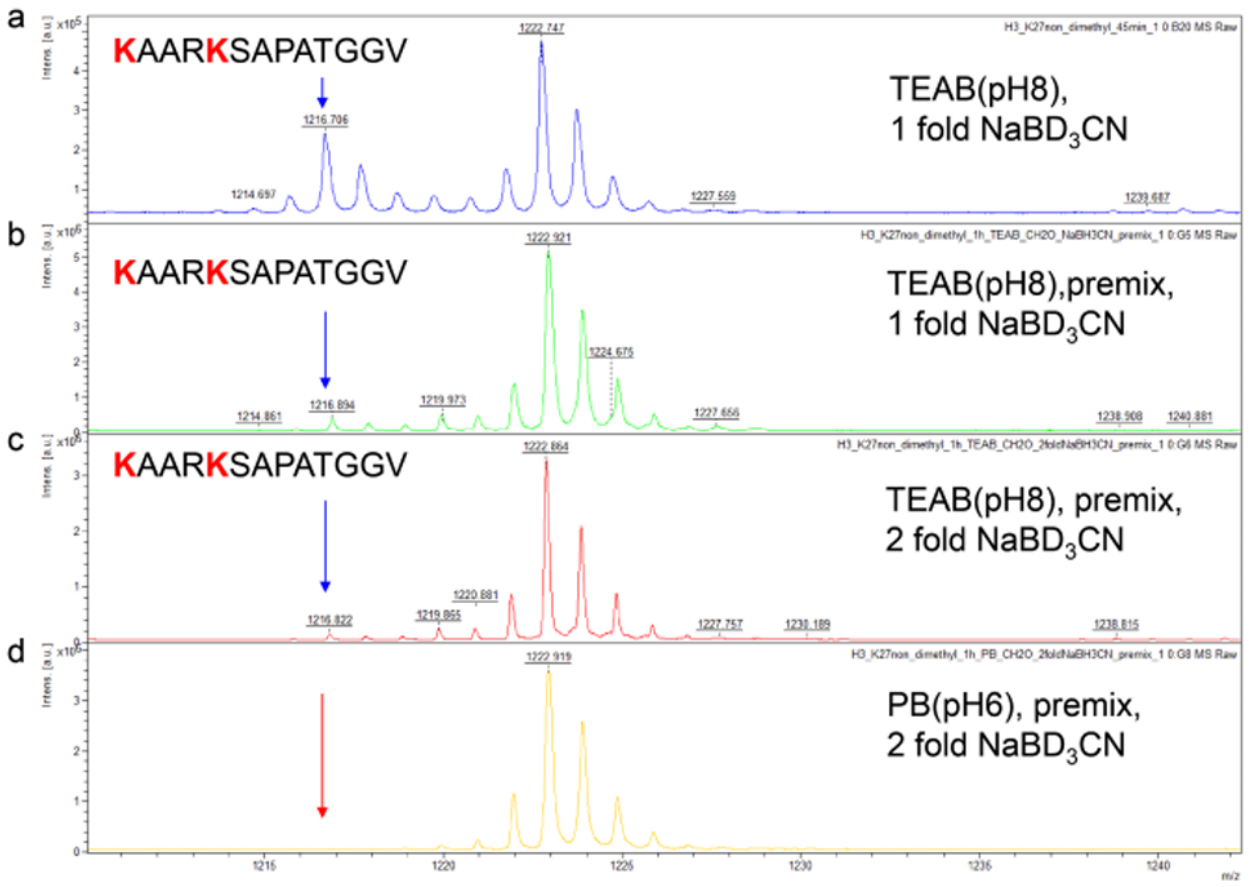
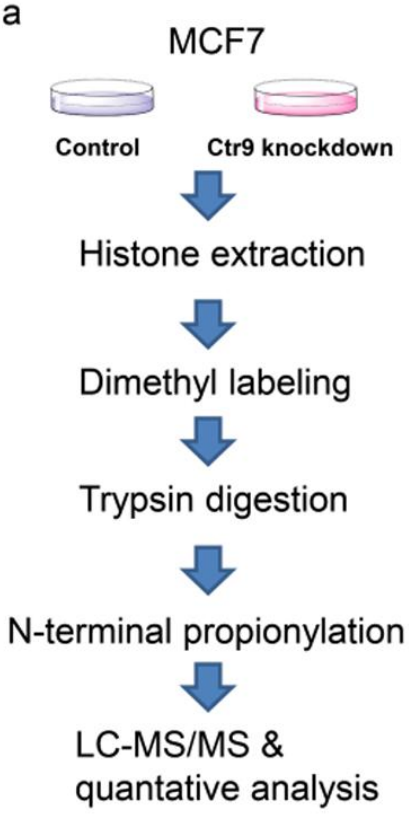


Figure 6. Side reaction can be avoided by optimizing reaction conditions. a) MALDI-TOF spectrum of histone H3 peptide (23-34) after standard dimethyl labeling procedure. b) Premix label reagents with reductant significantly reduced the side products. c) Increased concentration of reductant also helps to eliminate the side products. d) Weak acidic buffer is good to eliminate the side reaction without sacrificing the reaction efficiency.

a



b

Ctr9 knockdown/Control ratio

	Replicate 1	Replicate 2	Replicate 3	Average
H3K4unmodified	1.02	1.01	1.02	1.02
H3K4me1	0.84	0.88	0.83	0.85
H3K4me2	0.90	0.89	0.88	0.89
H3K4me3	0.98	0.92	1.74	1.13
H3K9unmodified	1.20	1.21	1.29	1.23
H3K9me1	0.97	1.01	1.01	1.00
H3K9me2	0.93	0.95	0.89	0.92
H3K9me3	0.86	0.81	0.82	0.83
H3K14unmodified	1.03	1.33	1.21	1.18
H3K14me1	0.85	0.64	0.67	0.71
H3K14me2	0.99	0.57	0.73	0.73
H3K14me3	0.73	0.61	0.52	0.61
H3K23unmodified	1.02	1.00	1.05	1.02
H3K23me1	1.35	1.11	1.06	1.17
H3K27unmodified	0.11	0.16	0.16	0.14
H3K27me1	0.69	0.82	0.76	0.75
H3K27me2	4.47	2.97	3.42	3.53
H3K27me3	5.71	4.68	4.95	5.09
H3K36unmodified	1.25	1.17	1.07	1.16
H3K36me1	0.84	1.12	0.90	0.94
H3K36me2	0.99	0.88	1.06	0.97
H3K36me3	0.27	0.45	0.54	0.41
H3K79unmodified	1.02	1.01	1.01	1.01
H3K79me1	0.92	0.96	0.95	0.94
H3K79me2	0.86	0.83	0.90	0.87

Figure 7. Global analyses of histone methylations in breast cancer cells with or without loss of RNA polymerase-associated protein CTR9. a) Scheme of the quantative histone lysine methylation analysis. b) Profiled histone H3 lysine methylation sites and ratio of lysine methylation change upon the loss of CTR9.

Reference

1. C. Martin, Y. Zhang, Nature Reviews Molecular Cell Biology **2005**, 6, 838-849.

2. K. Hyun, J. Jeon, K. Park, J. Kim, Exp. Mol. Med. **2017**, 49, e324 e324.

3. Joshua C. Black, C. Van Rechem, Johnathan R. Whetstine, Mol. Cell **2012**, 48, 491-507.

4. H. C. Beck, E. C. Nielsen, R. Matthiesen, L. H. Jensen, M. Sehested, P. Finn, M. Grauslund, A. M. Hansen, O. N. Jensen, Mol. Cell. Proteomics **2006**, 5, 1314-1325.

5. H. Huang, S. Lin, B. A. Garcia, Y. Zhao, Chem. Rev. **2015**, 115, 2376-2418.

6. B. M. Zee, R. S. Levin, B. Xu, G. LeRoy, N. S. Wingreen, B. A. Garcia, J. Biol. Chem. **2010**, 285, 3341-3350.

7. aS. Sidoli, N. V. Bhanu, K. R. Karch, X. Wang, B. A. Garcia, JoVE **2016**, e54112; bY. Hu, H. Jiang, B. Zhao, K. Yang, Z. Liang, L. Zhang, Y. Zhang, Analytical Methods **2021**, 13, 469-476; cM. D. Plazas-Mayorca, J. S. Bloom, U. Zeissler, G. Leroy, N. L. Young, P. A. DiMaggio, L. Krugylak, R. Schneider, B. A. Garcia, Mol. Biosyst. **2010**, 6, 1719-1729; dB. A. Garcia, S. Mollah, B. M. Ueberheide, S. A. Busby, T. L. Muratore, J. Shabanowitz, D. F. Hunt, Nat. Protoc. **2007**, 2, 933-938.

8. S. E. Kulevich, B. L. Frey, G. Kreitinger, L. M. Smith, Anal. Chem. 2010, 82, 10135-10142.

9. J.-L. Hsu, S.-Y. Huang, N.-H. Chow, S.-H. Chen, Anal. Chem. 2003, 75, 6843-6852.

10. aG. E. Means, R. E. Feeney, Biochemistry 1968, 7, 2192-2201; bY. Wu, F. Wang, Z. Liu, H. Qin, C. Song, J. Huang, Y. Bian, X. Wei, J. Dong, H. Zou, Chem. Commun. 2014, 50, 1708-1710; cJ. Huang, H. Qin, J. Dong, C. Song, Y. Bian, M. Dong, K. Cheng, F. Wang, D. Sun, L. Wang, M. Ye, H. Zou, J. Proteome Res. 2014, 13, 3896-3904; dJ. F. Huang, H. Q. Qin, Z. Sun, G. Huang, J. W. Mao, K. Cheng, Z. Zhang, H. Wan, Y. T. Yao, J. Dong, J. Zhu, F. J. Wang, M. L. Ye, H. F. Zou, Sci. Rep. 2015, 5, 10164; eJ.-L. Hsu, S.-H. Chen, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 2016, 374, 20150364.

11.M. J. Gidley, J. K. M. Sanders, Biochemical Journal 1982, 203, 331-334.

12. aV. Golghalyani, M. Neupärtl, I. Wittig, U. Bahr, M. Karas, J. Proteome Res. 2017, 16, 978-987; bC. U. Schräder, S. Moore, A. A. Goodarzi, D. C. Schriemer, Anal. Chem. 2018, 90, 9077-9084.

13. P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A. J. R. Heck, Nat. Protoc. 2009, 4, 484-494.

14. aB. Metz, G. F. A. Kersten, G. J. E. Baart, A. de Jong, H. Meiring, J. ten Hove, M. J. van Steenbergen, W. E. Hennink, D. J. A. Crommelin, W. Jiskoot, Bioconjugate Chemistry 2006, 17, 815 822; bR. Liao, Y. Gao, M. Chen, L. Li, X. Hu, Anal. Chem. 2018, 90, 13533-13540.

15. B. Metz, G. F. A. Kersten, P. Hoogerhout, H. F. Brugghe, H. A. M. Timmermans, A. de Jong, H. Meiring, J. t. Hove, W. E. Hennink, D. J. A. Crommelin, W. Jiskoot, J. Biol. Chem. 2004, 279, 6235-6243.

16. B. Li, H. Tang, A. Turlik, Z. Wan, X.-S. Xue, L. Li, X. Yang, J. Li, G. He, K. N. Houk, G. Chen, Angew. Chem. Int. Ed. 2021, 60, 6646 6652.

17. J. J. A. G. Kamps, R. J. Hopkinson, C. J. Schofield, T. D. W. Claridge, Communications Chemistry 2019, 2, 126.

18. H. Qin, F. Wang, Y. Zhang, Z. Hu, C. Song, R. a. Wu, M. Ye, H. Zou, Chem. Commun. 2012, 48, 6265-6267.

19. Ngai T. Chan, J. Huang, G. Ma, H. Zeng, K. Donahue, Y. Wang, L. Li, W. Xu, Nucleic Acids Res. 2022, 50, 1969-1992.

Supplementary information

Experimental Procedures

Cell lines and Cell Culture

MCF7 cell lines were obtained from American Type Culture Collection (ATCC) and maintained in Dulbecco's modified medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (VWR) and 1% Penicillin-Streptomycin (P/S) (Gibco). MCF7-tet-on-parental cells and MCF7-tet-on-shCTR9 cells were generated previously and were seeded on 3.5cm petri dishes and cultured in DMEM supplemented with 10% FBS in the absence or presence of 500 ng/mL Doxycycline (Dox). All cells were cultured at 37°C in a humidified atmosphere containing 5% CO_2. Doxycycline (Dox) was purchased from Clonetech and used at a final concentration of 500 ng/ml. Histone chemical derivation reagents triethylammonium bicarbonate buffer (TEAB), propionic anhydride, Formaldehyde-^{13}C, D_2 solution($^{13}CD_2O$), trifluoroacetic acid (TFA) and Sodium cyanoborodeuteride ($NaBD_3CN$) were purchased from Sigma Aldrich.

Histone extraction and purification

MCF7 cells with or without CTR9 knockdown were harvested after trypsinization. After washing with 1X PBS, cell pellets were resuspended in two volumes of lysis buffer [50 mM Tris-HCl pH7.4, 150 mM NaCl ,10% Glycerol and 0.05% NP-40]. Protease Inhibitors (1X protease inhibitor cocktail) and HDAC inhibitor (10 mM sodium butyrate) were added before use. The cell pellets were incubated on ice for 30 mins, followed by a brief sonication. After 15 mins of centrifugation at 13,000 rpm at 4°C, the supernatant was saved as whole cell lysate and the pellet

was used for histone extraction. Pellets were washed twice using NIB buffer [10 mM Tris-HCl pH7.5, 2 mM $MgCl_2$, 3 mM $CaCl_2$ and 1% NP40] containing 100 mM NaCl, and once with NIB buffer containing 400 mM NaCl. The pellet was finally resuspended in NIB buffer (400 mM NaCl) without NP40. For acid extraction of histones, double volume of 0.2N HCl was added and incubated overnight at 4°C. After centrifugation at 13,000 rpm for 15 mins at 4°C, solubilized histone in supernatant was dialyzed in ddH_2O overnight at 4°C using 10K MWCO Dialysis Tubing (Thermo). Histones were lyophilized and dissolved in ddH_2O.

Liquid Chromatography and Quantitative Histone Mass Spectrometry (LC-MS/MS)

Protein-level chemical derivatization of histones: this procedure was performed in three different method.

1) Propionylation method. Take 25 µg of purified histone samples and dissolve them in 80 µL 50 mM TEAB buffer (pH 8.0) and add 20µL 25% propionic anhydride buffer (v/v in ACN) in the sample, add 12 µL 28% NH_4OH to keep the pH of the sample reaction to be around 8.0. Put the sample tube on vortex for 20 min to label the lysine residue with histones with propionyl. Terminate the reaction by acidifying the sample with TFA.

2) Conventional dimethyl labeling method. Take 25 µg of purified histone samples and dissolve them in 100µL 50 mM TEAB buffer (pH 8.0). Sequentially add 4µL 4% $^{13}CD_2O$ (w/v) and 4µL 0.6 M $NaBD_3CN$ to the samples and vortex them at room temperature for 1h to label the free and mono-methylated lysine with heavy isotopic methyl. Terminate the reaction by acidifying the sample with TFA.

3) Optimized dimethyl labeling method. Take 25 μg of purified histone samples and dissolve them in 100μL 100 mM PB buffer (pH 6.0). Premix 4μL 4% $^{13}CD_2O$ (w/v) and 4μL 1.2 M $NaBD_3CN$ and then add to the samples and vortex them at room temperature for 1h to label the free and mono-methylated lysine with heavy isotopic methyl. Terminate the reaction by acidifying the sample with TFA.

Trypsin digestion. Transfer the chemical derivatization samples to 10K MWCO ultracentrifuge tube (Millipore) and centrifuge the samples for 15 min at 14,000 g at 4°C to remove the reaction reagents. Continually add 200 μL 100 mM TEAB buffer in the ultracentrifuge tube and centrifuge the samples for 15 min at 14,000 g at 4°C to wash histone sample for 3 times. Add 100 μL 100mM TEAB buffer and 1μg trypsin in the ultracentrifuge tube and put the samples at 37°C incubators for 16 h. Collect the digested sample by centrifuging the samples for 15 min at 14,000 g at 4°C. Wash the ultracentrifuge tube two times with 100μL 100 mM TEAB buffer and combine the follow-through with digested sample and dry down.

Peptide-level propionylation: Resolve the dry down sample with 80 μL 50 mM TEAB buffer and add 20μL 25% propionic anhydride buffer (v/v in ACN) in the sample, add 12 μL 28% NH_4OH to keep the pH of the sample reaction to be around 8.0. Put the sample tube on vortex for 20 min to label the N-terminal of digested histone peptides with propionyl 1) Terminate the reaction by acidifying the sample with TFA. After propionylation, the sample was desalted with Sep-Pak cartridge (Waters) and lyophilized.

LC-MS/MS for histone modification: Lyophilized histone peptides were resuspended in 0.1% formic acid (FA) and analyzed on a Dionex U3000 ultra performance liquid chromatography system coupled to a Q-Exactive HF quadrupole orbitrap mass spectrometer (Thermo Fisher Scientific). Peptide sample was injected (2 µl). A Waters BEH 300Å C18 reversed phase capillary column (150 mm x 75 µm, 1.7 µm) was used for separation. Water with 0.1% FA and acetonitrile with 0.1% FA were used as mobile phases A and B, respectively. The flow rate was set to 0.300 µL/min. 2 µL of sample was injected onto the column and separated over a 120-minute gradient as follows: 0-1 min 3-10% B; 1-90 min 10-35% B; 90-92 min 35-95% B; 92-102 min 95% B; 102-105 min 95-3% B; 105-120 min 3% B. The data was acquired under data dependent acquisition mode (DDA, top 20). Mass spectrometric conditions were as follows: spray voltage of 2.8 kV, no sheath and auxiliary gas flow; heated capillary temperature of 275°C, normalized high-energy collision dissociation (HCD) collision energy of 33%, resolution of 120, 000 for full scan, resolution of 60, 000 for MS/MS scan, automatic gain control of 2e5, maximum ion injection time of 100 ms, isolation window of 1.6, and fixed first mass of 110 m/z.

Data analysis and relative quantification of histone PTMs: Due to histones possess multiple post-translational modifications, and after heavy isotopic dimethylation and propionylation they are even highly modified and may contain several isoforms for same peptide sequence with different modifications, it is challenging to identify all different forms of histone modification peptides. Here in this work, the most frequently observed modifications including the mono-methylation, di-methylation, tri-methylation, and acetylation on lysine residue of histone H3 were analyzed. The same histone peptide possess different modifications can firstly be identified according to their difference in mass over charge (*m/z*). Then the peptide can be additionally

distinguished according to their retention time difference on the RP-HPLC column (trimethylated peptide ≈ unmodified (possesses two heavy isotopic methyl), mono (possesses one heavy isotopic methyl), and di-methylated peptide < acetylated peptides. To relatively quantify the abundance of histone PTMs, we used the area of each identified peptide peak in the MS chromatogram for comparison. Normally same peptide may have different charge state ions in LC-MS analysis. We only choose the highest intensity ions to measure their peak area. The total peak area of a histone peptide with all different PTM forms is regarded as 100%, and the percentage of each PTM on the peptide is calculated by dividing the area of the PTM peak area by the total peak area. To distinguish histone peptide isoform, we also investigate the MS/MS spectrum to calculate the ratios of b and/or y ions that were different between the two or more peptide isoforms, and the ratio is used to continually calculate the relative quantity of the peptide isoforms.

Supplementary figures

Figure S1. Histones are arginine & lysine-rich and super-hydrophilic proteins, trypsin digestion of histones results in very small and hydrophilic peptides which are difficult to be detected by conventional LC-MS/MS.

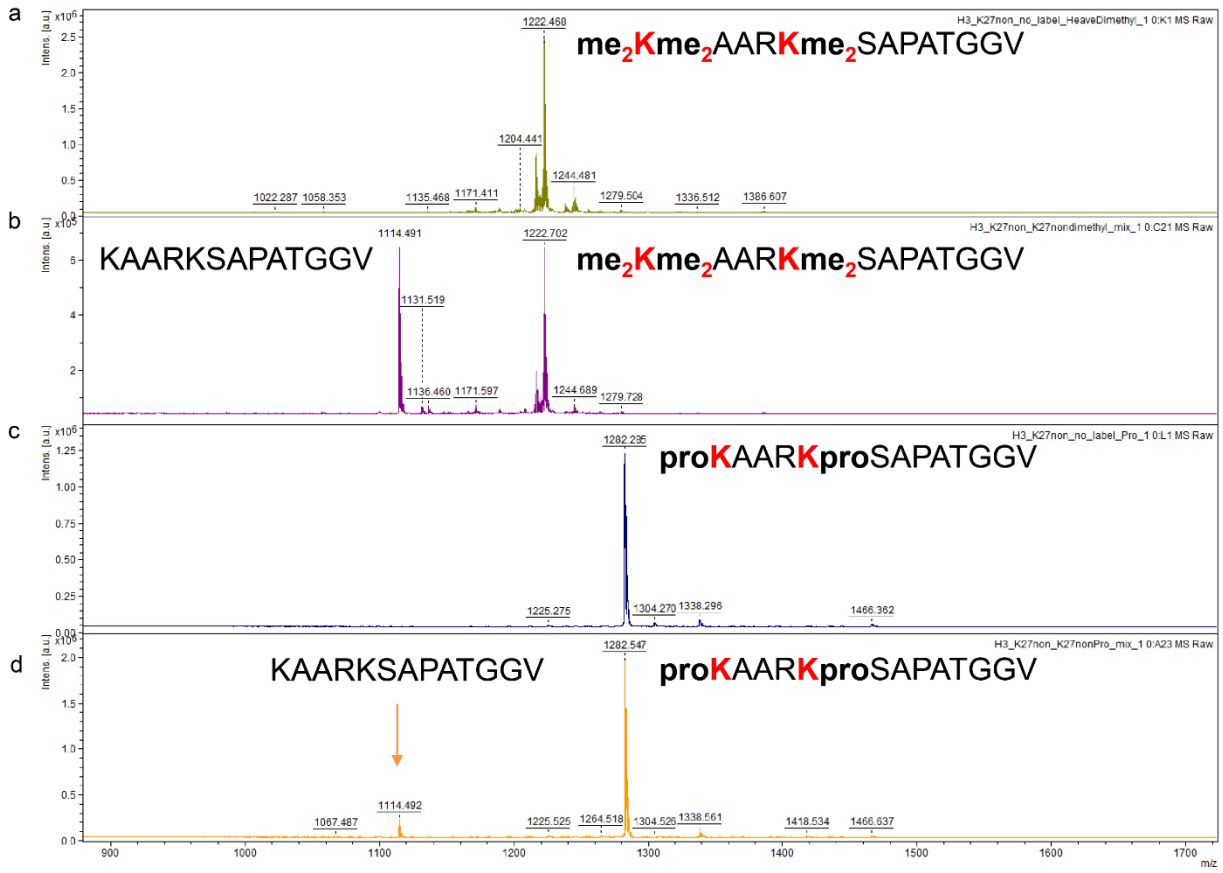


Figure S2. Comparison of MALDI-TOF spectra of histone H3 peptide (23-34) and its dimethyl or propionylation labeled ones. a) histone H3 peptide (23-34) with heavy dimethyl labels. b) Mix non-labeled histone H3 peptide (23-34) with aliquots of heavy dimethyl labeled one. c) histone H3 peptide (23-34) with propionylation label. d) Mix non-labeled histone H3 peptide (23-34) with aliquots of propionylation labeled ones.

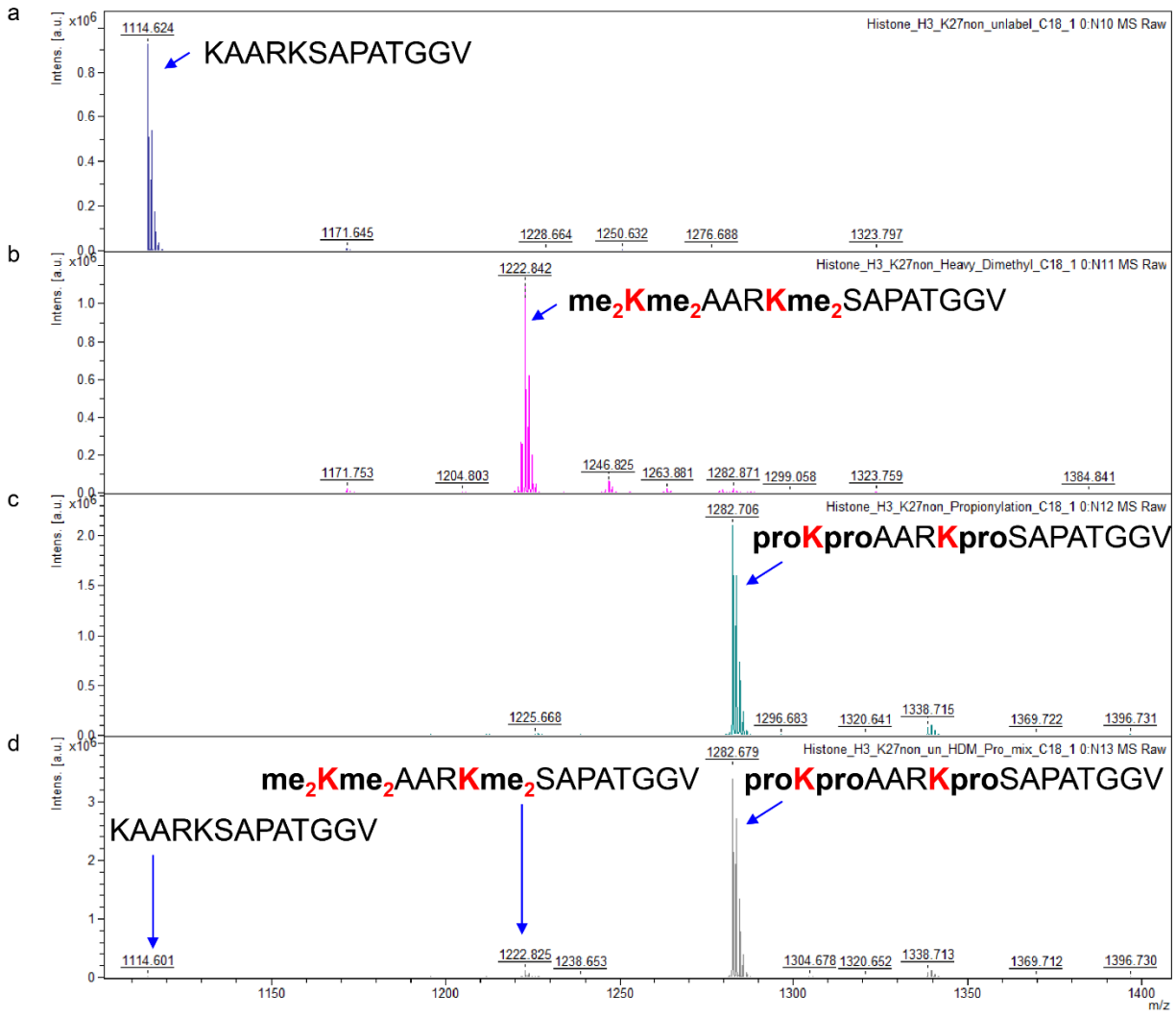


Figure S3. MALDI-TOF analysis of histone H3 peptide (23-34) with or without chemical labels. a) histone H3 peptide (23-34) without labels. b) histone H3 peptide (23-34) with heavy dimethyl label. c) histone H3 peptide (23-34) with propionylation label. d) Mix non-labeled histone H3 peptide (23-34) with aliquots of heavy dimethyl labeled and propionylation labeled ones.

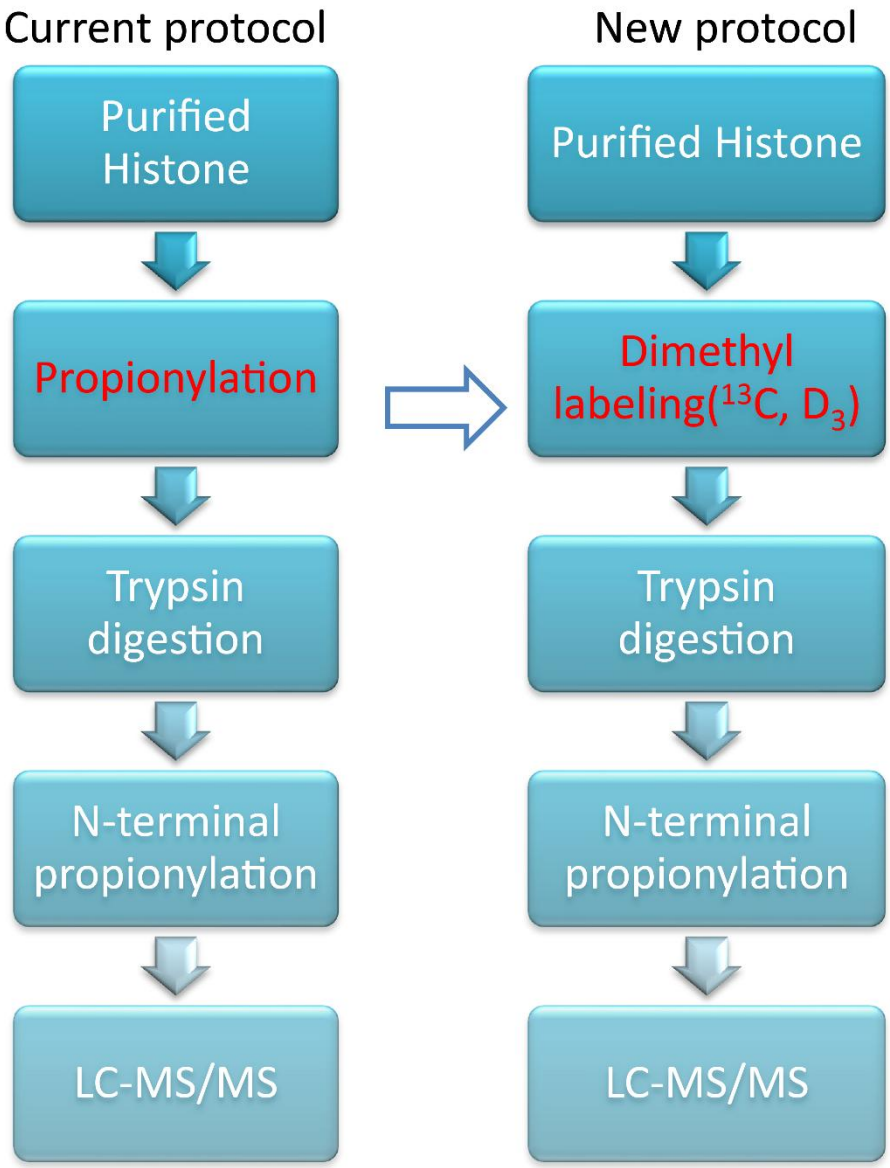


Figure S4. The new protocol replaced the protein level propionylation with dimethylation labeling, however, kept the peptide level propionylation procedure after tryptic digestion the same with current protocol.

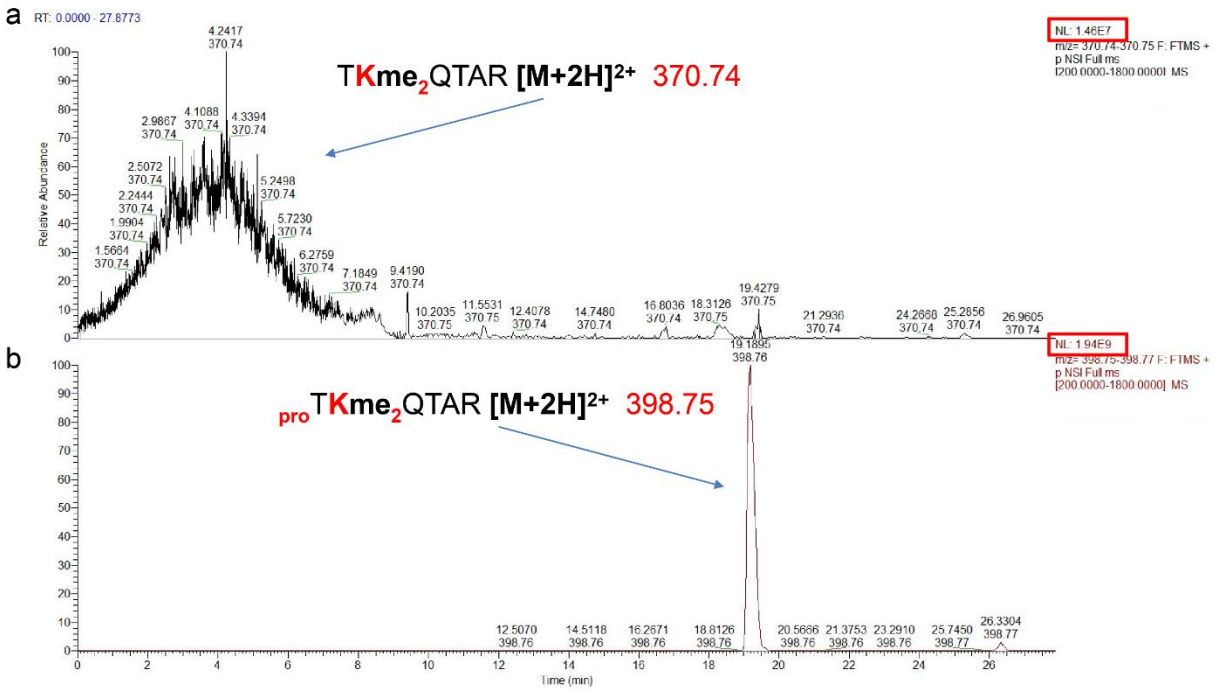


Figure S5. Peptide N-termini were labeled with propionylation to increase their hydrophobicity and eliminate the sample loss of small and hydrophilic peptides. a) LC-MS analysis of dimethyl labelled histone H3 peptide (3-8) without N-terminal propionylation labeling. b) LC-MS analysis of dimethyl labelled histone H3 peptide (3-8) with N-terminal propionylation labeling

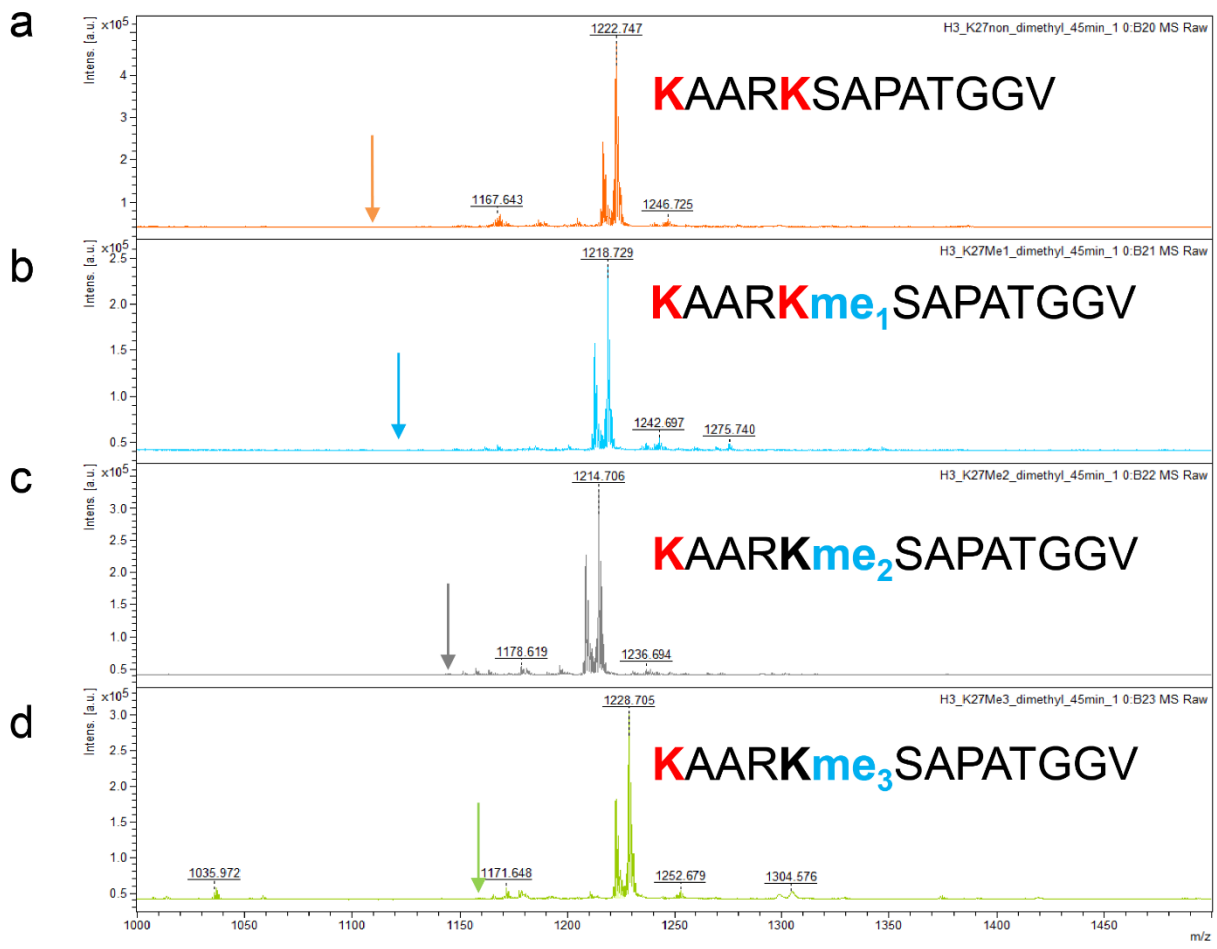


Figure S6. Evaluation of the widely used dimethyl labeling protocol by using all four modification types of standard histone H3 peptides (23-34). a) unmodified peptide. b) mono-methylated peptide. c) di-methylated peptide. d) tri-methylated peptide.

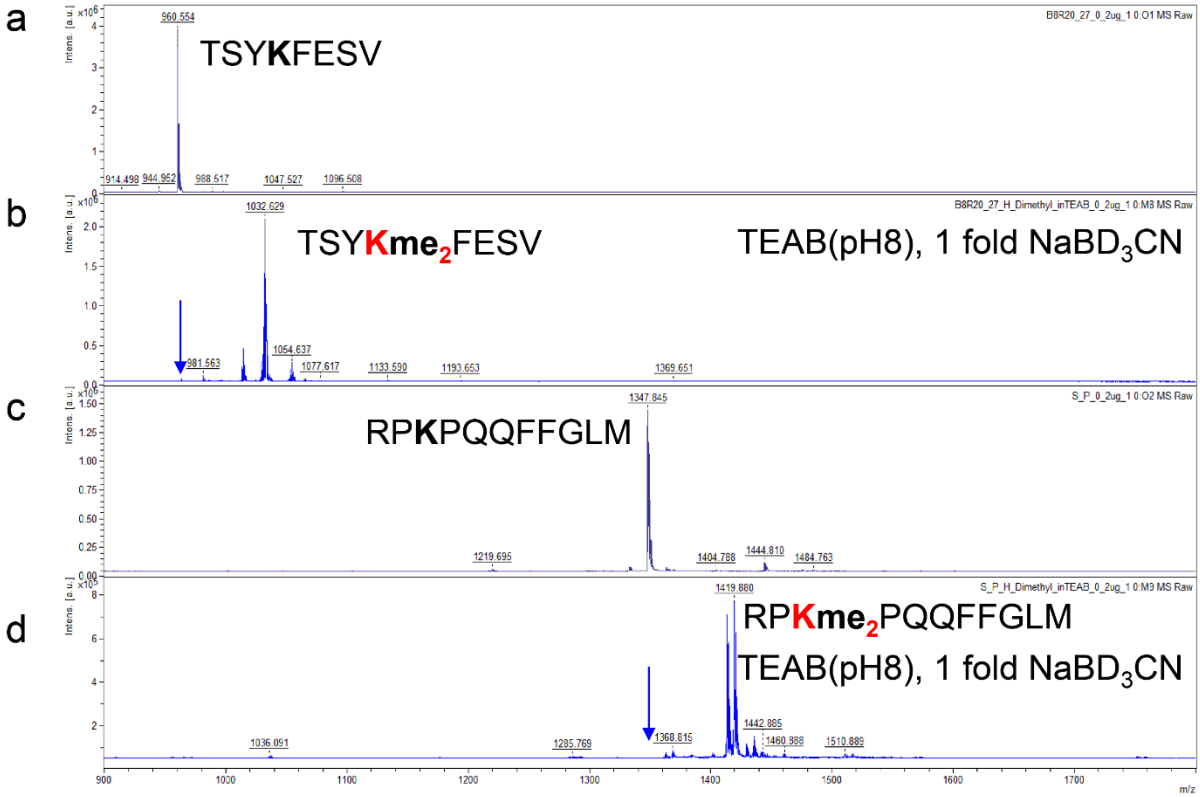


Figure S7. Evaluation of the widely used dimethyl labeling protocol by using the other two peptides TSYKFESV and RPKPQQFFGLM. a) unmodified TSYKFESV peptide. b) heavy dimethyl labeled TSYKFESV peptide. c) unmodified RPKPQQFFGLM peptide. d) heavy dimethyl labeled RPKPQQFFGLM peptide.

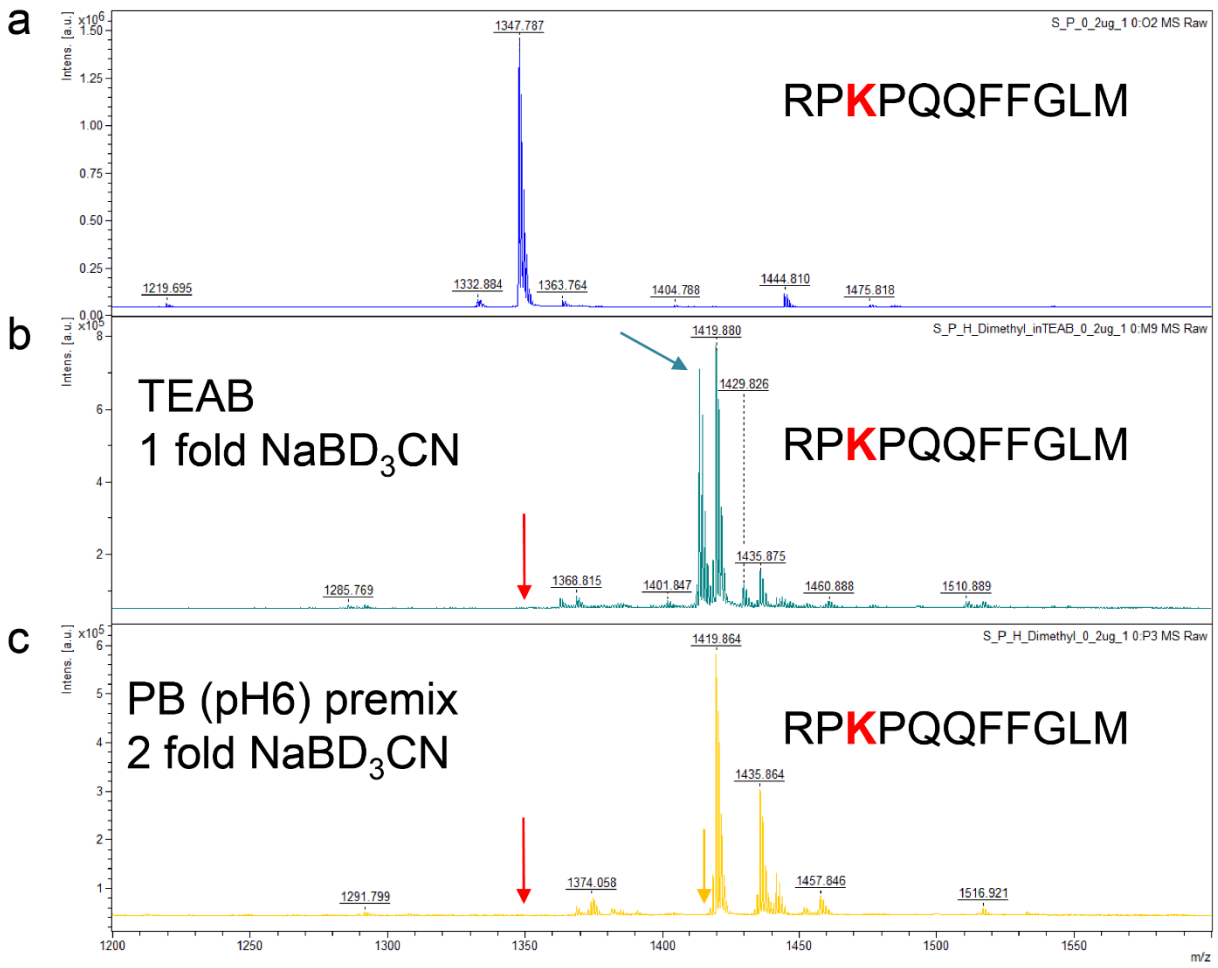


Figure S8. Side reaction can be avoided by optimizing reaction conditions without undermining the labeling efficiency. a) MALDI-TOF spectrum of RPKPQQFFGLM peptide. B) MALDI-TOF spectrum of RPKPQQFFGLM after standard dimethyl labeling procedure. c) MALDI-TOF spectrum of RPKPQQFFGLM after optimizing dimethyl labeling procedure.

Peptide	Modification method		m/z	peak intensity	Mean	SD
KAARKSAPATGG	Unlabeled		1114.624	9.21E+05	1.05E+06	1.20E+05
			1114.628	1.08E+06		
			1114.620	1.16E+06		
	Heavy Dimethyl		1222.842	1.01E+06	1.01E+06	1.42E+05
			1222.840	1.15E+06		
			1222.847	8.65E+05		
	Propionylation		1282.699	2.09E+06	2.27E+06	1.75E+05
			1282.681	2.44E+06		
			1282.699	2.28E+06		
	Mix of three	Unlabeled peak	1114.601	1.62E+04	1.70E+04	1.07E+04
			1114.619	6.66E+03		
			1114.622	2.80E+04		
		Heavy dimethyl peak	1222.825	1.12E+05	1.09E+05	5.02E+04
			1222.837	5.73E+04		
			1222.845	1.58E+05		
		Propionylation peak	1282.679	3.58E+06	2.68E+06	8.14E+05
			1282.685	1.99E+06		
			1282.694	2.47E+06		

Table S1. Export of MALDI-TOF intensities of histone H3 peptide (23-34) with or without chemical labels.

Chapter 9

Conclusions and future directions

Conclusions

This dissertation presents an integrated mass spectrometry framework that advances both methodological development and biological discovery in the study of protein structural variation and post-translational modifications. The body of work is intentionally structured to progress from focused characterization of citrullination in autoimmune and neurodegenerative diseases, to methodological innovations in structural proteomics, and finally to integrative analyses that connect molecular-level changes with biological mechanisms of disease.

In **Chapters 2 and 3**, citrullination was investigated as a critical yet underexplored post-translational modification in inflammatory and neurodegenerative contexts. Through biotin-thiol tag–assisted mass spectrometry and a newly designed citrulline-reactive probe, we mapped citrullination landscapes across rheumatoid arthritis, Sjögren's syndrome, and Alzheimer's disease, identifying modification hotspots, novel motifs, and potential biomarkers. These studies not only expand the citrullinome catalog but also demonstrate the value of tailored chemoproteomic platforms in uncovering subtle molecular alterations linked to disease.

The thesis then transitions into **Chapters 4–7,** which focus on advancing structural proteomics through crosslinking mass spectrometry (XL-MS) and complementary techniques. We benchmarked XL-MS alongside limited proteolysis-MS, developed an innovative 4-plex isobaric crosslinker (XLeu), and applied quantitative crosslinking to cerebrospinal fluid networks in Alzheimer's disease, revealing site-specific conformational changes in key proteins such as ApoE and complement factors. Further integration with native ion mobility-MS illuminated how PAD4-mediated citrullination destabilizes histone H2A–H2B dimers and remodels chromatin interfaces, offering mechanistic insight into epigenetic regulation under modification stress.

Chapter 8 broadened the focus on histone biology by employing stable isotope dimethyl labeling to quantify lysine methylation dynamics. This work provided refined methodology for PTM quantitation and uncovered regulatory mechanisms of histone methylation in breast cancer models, underscoring the broader utility of chemical labeling in chromatin biology.

Together, these studies highlight three core contributions: (i) establishing robust platforms for citrullination mapping in complex biological systems, (ii) innovating new tools and strategies for structural proteomics, and (iii) integrating PTM profiling with structural analyses to uncover molecular mechanisms underlying disease. Collectively, these advances demonstrate how mass spectrometry can bridge the gap between protein chemistry and disease pathology, from immune dysregulation in autoimmunity to network remodeling in Alzheimer's disease and chromatin instability under PTM regulation.

Looking forward, the work presented here opens several avenues for further exploration. Citrullination-specific chemoproteomic strategies can be extended to longitudinal patient cohorts to evaluate biomarker robustness and clinical utility. The multiplexed XL-MS methodology may be scaled to higher-plex crosslinkers, expanding throughput for systems-level analyses. Finally, the integrated investigation of histone modifications invites broader application of multi-dimensional MS approaches to unravel the interplay between diverse PTMs and chromatin organization. By advancing both methodology and application, this dissertation contributes to a deeper understanding of how protein modifications and structural changes shape disease processes and sets the foundation for future translational biomarker discovery and therapeutic targeting.

Future Directions

Building upon the development of the XLeu cross-linker presented in **Chapter 6**, several critical areas require further optimization and expansion before this platform can be fully leveraged for systematic protein–protein interaction (PPI) network analysis.

1. Expansion of Biological Replicates and Cohort Diversity

One of the most immediate challenges is the limited number of biological samples analyzed in preliminary studies. Although our current dataset is promising, it reveals substantial variability across age, gender, and disease status (AD vs. non-AD). The small sample size prevents the construction of a robust and generalizable PPI network. To address this limitation, future work will focus on preparing at least three biological replicates for each defined subgroup (age, gender, and disease category). This expanded dataset will provide the statistical power necessary to trace how disease-associated network changes are influenced by demographic and clinical variables. Establishing such a cohort-based library will form the foundation for quantitative and comparative PPI network studies.

2. Optimization of Crosslinking and Sample Preparation Strategies

Another priority is refining the experimental workflow for XLeu-based crosslinking. Current results indicate that crosslink spectrum matches (CSMs) and inter-protein crosslinks (PPs) are substantially fewer than those reported in cerebrospinal fluid (CSF) studies using established crosslinkers, such as those published in *Analytical Chemistry*. To increase coverage and yield, the following optimization strategies will be pursued:

Fractionation Approaches: Incorporate high-pH fractionation workflows, widely adopted by leading crosslinking laboratories, to enrich crosslinked peptides and enhance recovery of Type 2 interlinks.

Buffer and pH Conditioning: Systematically evaluate reaction pH and buffer composition to maximize labeling efficiency while preserving protein integrity.

Targeted Enrichment of Plasma Proteins: Develop enrichment strategies for high-value plasma proteins to focus structural analysis on the most biologically relevant PPIs.

3. Toward a Quantitative, Library-Based PPI Framework

Ultimately, the integration of optimized XLeu workflows with expanded sample sets will enable the construction of a library-based quantitative PPI network, capturing age-, gender-, and disease-dependent variations. This resource will serve as a reference for identifying condition-specific crosslinks, tracing disease progression, and potentially uncovering therapeutic targets in Alzheimer's disease and related disorders.

While the integration of crosslinking MS (XL-MS) and native IM-MS in **Chapter 7** provided important insights into PAD4-mediated citrullination effects on histone H2A–H2B dimers, several limitations remain. The intrinsic fragility of the dimer significantly hindered our ability to perform direct ion-mobility comparisons of unmodified versus citrullinated dimers. Despite carefully optimized direct-infusion conditions, the citrullinated complexes consistently dissociated either during transfer into the gas phase or upon low-level collisional excitation, preventing reliable CIU fingerprinting and drift-time (DT) analysis.

To overcome these challenges, future research will prioritize both sample preparation and instrumental strategies to stabilize the dimer complex:

1. Stabilization Approaches

Buffer optimization: Explore non-volatile additives (e.g., trehalose, glycerol, mild detergents) compatible with native MS to enhance stability of citrullinated complexes.

2. Instrumental and Methodological Refinements

Trapping and activation control: Utilize lower-voltage CIU ramping protocols, as well as controlled activation in different regions (trap vs. transfer cell), to capture intermediate unfolding states.

Orthogonal approaches: Incorporate complementary gas-phase methods (e.g., charge-detection MS, surface-induced dissociation) to resolve higher-order dissociation pathways missed in CIU.

Collectively, these refinements will allow us to more robustly track citrullination-dependent structural transitions in fragile histone dimers, ultimately providing higher-resolution mechanistic insight into chromatin destabilization.

Appendix

Publications

1. **Z. Zhu**[#], X. Zhong[#], B. Wang[#], H. Lu, and L. Li. Probing Protein Structural Changes in Alzheimer's Disease Via Quantitative Cross-linking Mass Spectrometry. *Analytical Chemistry* **2024**, 96: 7506–7515. (#Co-First authors)
2. **Z**. **Zhu,** H. Chiang, W. Li, H. Zhang, Z. Jin, L. Li. A Systematic Investigation of the Citrullinome and Its Variation in Human Rheumatoid Arthritis Tissues. *Journal of Proteomic Research* **2025.** In revision.
3. H. Lu, **Z**. **Zhu,** L. Fields, H. Zhang, L. Li. Mass spectrometry structural proteomics enabled by limited proteolysis and Cross‐Linking. *Mass Spectrometry Reviews* **2024**, 43 (5), e22134.
4. J. Huang, N. T. Chan, **Z. Zhu**, G. Delafield, D. Wang, S. Zhang, W. Xu, and L. Li, Eliminating lysine-arginine cyclization during dimethyl labeling to improve accuracy of histone methylation stoichiometric analysis," *Angew. Chem. Int. Ed*., **2025**. To be submitted
5. W. Li, **Z**. **Zhu,** L. Fields, H. Lu, Z. Wang, K. Lu, L. Li. Triple safeguard-enhanced chemoproteomic platform for reliable citrullination mapping with staggered DIA detection in Alzheimer's disease. *Nature Communications*. **2025.** In preparation.
6. **Z. Zhu,** W. Wei, C. A. Douglas, Z. Wang, L. Li. XLeu: An Innovative Type of Cleavable Crosslinker with Quantitative Isobaric Tag Labeling. **2025.** In preparation
7. **Z. Zhu,** B. Wang, Z. Zhu, S. Xu, L. Li. A Dual-MS Approach to Explore the Citrullination Effect on Histone H2 Dimers. **2025.** In preparation
8. B. Wang, X. Zhong, L. Field, X. H. Lu, **Z. Zhu,** L. Li, Structural proteome profiling in cerebrospinal fluid of Alzheimer's disease to reveals novel conformational biomarkers, *Journal of the American Society for Mass Spectrometry* **2023**, 34: 459–471.
9. J. Huang, D. Wang, R. D. Shipman, **Z. Zhu,** Y. Liu, and L. Li. Simultaneous enrichment and separation of neutral and sialyl glycopeptides of SARS-CoV-2 spike protein enabled by dual-functionalized Ti-IMAC material. *Analytical and Bioanalytical* Chemistry **2021**, 413 (29), 7295–7303.

Presentations

➢ American Society for Mass Spectrometry (ASMS) 72nd conference on Mass Spectrometry and Allied Topics (Baltimore, MD) June 2025

Poster Presentation: XLeu: An Innovative Type of Cleavable Crosslinker with Quantitative Isobaric Tag Labeling

➢ ASMS 71st conference (Anaheim, CA) June 2024

Poster Presentation: A Systematic Investigation of Citrullination and Corresponding Structural Effect in Human Rheumatoid Arthritis

➢ ASMS 71st conference (Houston, TX), June 2023

Poster Presentation: Characterization of Protein-Protein Interactions by Quantitative Cross-linking Mass Spectrometry in Alzheimer's Disease

➢ ASMS 70th conference (Minneapolis, MN), June 2022

➢ Chinese and American Society for Mass Spectrometry (CASMS) Annual Conference 2022 (Minneapolis, MN), June 2022

➢ 2022 Summer School on Protein Structure Elucidation with Ion Mobility, Florida State University, Dr. Bleiholder lab, (Tallahassee, FL), Mar 2022

Poster Presentation: Investigation of Citrullination Effect on Histone H2A-H2B Dimerization via Biotin Derivatization and Cross-linking Mass Spectrometry

➢ ASMS 69th conference (Philadelphia, PA), June 2021

Poster Presentation: Simultaneous enrichment and separation of neutral and sialyl glycopeptides of SARS-COV-2 spike protein enabled by dual-functionalized Ti-IMAC material