UNIVERSITY OF CALIFORNIA,
IRVINE

Expanding the Chemical Cross-linking Tool Kit for Cross-linking Mass Spectrometry

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mass Spectrometry

by

Craig Bryant Gutierrez

Dissertation Committee:
Professor Lan Huang, Chair
Professor Rongsheng Jin
Professor Todd Holmes

2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my committee chair, Professor Lan Huang, whose been a significant source of advice and support. Her deep knowledge of the field and desire to push the boundaries of technology conveyed a spirit of adventure in regard to research and scholarship. Without her guidance and persistent help this dissertation would not have been possible.

I would like to thank my committee members, Professor Todd Holmes and Professor Rongsheng Jin, whose advice helped to guide me in navigating both technical and conceptual issues that pertained to my projects.

I would like to thank my lab members, Dr. Clinton Yu, Dr. Xiaorong Wang, and Andrew Wheat, whose advice, assistance and encouragement helped me in both my personal and professional development. Without their support I would have never been able to accomplish my academic goals.

In addition, a thank you to Professor Howard Xu of California State University, Los Angeles, Dr. Steven Herring of Grifols biological Inc., and Dr. Kerry Dooriss of Grifols Biological Inc., for fostering my interest in research, and whose collect enthusiasm and passion for the biology had lasting effects.

Financial support was provided by the University of California, Irvine, CSB Training Grant T32GM108561 and Dr. Lan Huang's NIH grants RO1GM074830 and RO1GM106003.

Use of copywrited materials has been granted by ACS publishing and PNAS publishing.

VITA

Craig Bryant Gutierrez

EDUCATION

06/2010	B.S. Degree in Biological Science California State University, Fullerton, Fullerton, CA
06/2013	M.Bt. Degree in Biotechnology California State University, Fullerton, Fullerton, CA
12/2020	Ph.D. Degree in Biomedical Sciences: Proteomics and Biological Mass Spectrometry, School of Medicine, University of California, Irvine, Irvine, CA Advisor: Prof. Lan Huang. Expected date: November 2020.

AWARDS and HONORS

- University of California, Irvine Chemical and Structural Biology Training Grant Fellowship
  - July 2016-June 2018

- California State University, Fullerton Natural Sciences and Mathematics Scholarship
  - 2011

- Undergraduate B.S. Honors (2010): Cum Laude California State University, Fullerton Dean's List for the College of Natural Science and Mathematics
  - Every Semester 2005-2010

RESEARCH EXPERIENCE

09/2014-present: Graduate Student, Department of Physiology & Biophysics, University of California Irvine, Irvine, CA

- Involved in the development novel MS-cleavable cross-linking mass spectrometry platforms and reagents exploring various functionalities (e.g. residue reactivity, reagent lengths, MS-labile functional groups, enrichment strategies), to provide complementary information to existing technologies for improved integrative 3D modeling and protein-protein interaction network generation.
- Involved in the development of an Integrated Multiplexed Quantitative Cross-Linking Mass Spectrometry Platform to Determine Composition-Dependent Conformational Changes of Protein Complexes.

- Involved in the development of a novel cross-linking platform to define human interactomes in living cells using an integrative multifunctionality targeting (e.g. residue reactivity) approach that allows the capture of protein-protein interactions, and the identification of protein interaction interfaces.

06/2013-08/2014: Associate Research Scientist, Grifols Biologicals Inc., Department of Research & Development, Los Angeles, California

- Involved in the development of assays for the detection and quantification of various proteolytic enzymes in purified blood plasma
- Involved in the development of proprietary novel assays for quality control assessing clotting factor protein complex interactions in various GBI products

06/2012-06/2013: R&D Intern, Grifols Biologicals Inc., Department of Research & Development, Los Angeles, California

- Involved in the development of a semi-quantitative native western blotting assay for defining the binding relationship between human coagulation FVIII and von Williebrand Factor

03/2012-04/2013: Post-Biauriculate Researcher (Advisor: Prof. Howard Xu), Department Biological Sciences, California State University, Los Angeles, Los Angeles, CA

- Identification of plasmid associated virulence factors in Acinetobacter Baumannii.
- Utilization of high throughput antibiotic screening to identify potential inhibitors of the FabI virulence factor found in A. Baumannii and E. coli.

09/2010-2011: Sr. Laboratory Technician, Advanced Sterilization Products, Department of Microbiology Research & Development, Irvine, CA

- Testing and qualifying the ability experimental medical devices to decontaminate an area contaminated with HAI causing microbes (including both bacteria and Fungi).

09/2009-06/2010: Undergraduate Researcher (Advisor: Prof. Marcelo Tolmasky), Department Biological Sciences, California State University, Fullerton, Fullerton, CA

- Understanding mechanism of amikacin resistance by nuclease resistant external guide sequences in bacteria

FIELD OF STUDY

Mass Spectrometry

PUBLICATIONS

1. **Craig Gutierrez**, Leah J. Salituro, Xiaorong Wang, Sadie F Depeter, Clinton Yu, Scott D. Rychnovsky, Lan Huang. Enabling Photoactivated Cross-linking Mass Spectrometric Analysis of Protein Complexes by Novel MS-cleavable Cross-linkers. Submitted for Review. (2020)

2. Peng Chen, Ji Zheng, Zheng Liu, **Craig B. Gutierrez**, Lan Huang, Min Dong, Rongsheng. Structural basis for recognition of CSPG4 by diverse Clostridioides difficile toxin B. Submitted for Review. (2020)

3. **Craig Gutierrez**, Ilan E. Chemmama, Haibin Mao, Clinton Yu, Ignacia Echeverria, Sarah A. Block, Scott D. Rychnovsky, Ning Zheng, Andrej Sali, and Lan Huang. *PNAS* February 25, 2020 117 (8) 4088-4098

4. Peng Chen, Kwok-ho Lam, Zheng Liu, Frank A. Mindlin, Baohua Chen, **Craig B. Gutierrez**, Lan Huang, Yongrong Zhang, Therwa Hamza, Hanping Feng, Tsutomu Matsui, Mark E. Bowen, Kay Perry, Rongsheng Jin. Structure of the full-length Clostridium difficile toxin B. *Nat. Struct. Mol. Biol.* 2019.

5. Liliane Abuin, Lucia L. Prieto-Godino, Haiyun Pan, **Craig B. Gutierrez**, Lan Huang, Rongsheng Jin and Richard Benton. In vivo assembly and trafficking of olfactory Ionotropic Receptors. *BMC Biology*. 2019 17:34 https://doi.org/10.1186/s12915-019-0651-7

6. Claudio Iacobucci, Christine Piotrowski, Ruedi Aebersold, Bruno C. Amaral, Philip Andrews, Katja Bernfur, Christoph Borchers, Nicolas I. Brodie, James E. Bruce, Yong Cao, Stéphane Chaignepain, Juan D. Chavez, Stéphane Claverol, Jürgen Cox, Trisha Davis, Gianluca Degliesposti, Meng-Qiu Dong, Nufar Edinger, Cecilia Emanuelsson, Marina Gay, Michael Götze, Francisco Gomes-Neto, Fabio C. Gozzo, **Craig B. Gutierrez**, Caroline Haupt, Albert J. R. Heck, Franz Herzog, Lan Huang, Michael R. Hoopmann, Nir Kalisman, Oleg Klykov, Zdeněk Kukačka, Fan Liu, Michael J. MacCoss, Karl Mechtler, Ravit Mesika, Robert L. Moritz, Nagarjuna Nagaraj, Victor Nesati, Ana G. C. Neves-Ferreira, Robert Ninnis, Petr Novák, Francis J. O'Reilly, Matthias Pelzing, Evgeniy Petrotchenko, Lolita Piersimoni, Manolo Plasencia, Tara Pukala, Kasper D. Rand, Juri Rappsilber, Dana Reichmann, Carolin Sailer, Chris P. Sarnowski, Richard A. Scheltema, Carla Schmidt, David C. Schriemer, Yi Shi, J. Mark Skehel, Moriya Slavin, Frank Sobott, Victor Solis-Mezarino, Heike Stephanowitz, Florian Stengel, Christian E. Stieger, Esben Trabjerg, Michael Trnka, Marta Vilaseca, Rosa Viner, Yufei Xiang, Sule Yilmaz, Alex Zelter, Daniel Ziemianowicz, Alexander Leitner, and Andrea Sinz. *Analytical Chemistry* 2019 91 (11), 6953-6961 DOI: 10.1021/acs.analchem.9b00658

7. **Craig B. Gutierrez**, Sarah A. Block, Clinton Yu, Stephanie M. Soohoo, Alexander S. Huszagh, Scott D. Rychnovsky, Lan Huang. Developing a Novel Sulfoxide-containing MS-cleavable Homobifunctional Cysteine Reactive Cross-linker for Studying Protein-Protein Interactions. *Analytical Chemistry.* 2018 90(12):7600-7607. doi: 10.1021/acs.analchem.8b01287

8. **Craig B. Gutierrez**, Clinton Yu, Eric J. Novitsky, Alexander S. Huszagh, Scott D. Rychnovsky, and Lan Huang. Developing an Acidic Residue Reactive and Sulfoxide-Containing MS-Cleavable Homobifunctional Cross-Linker for Probing Protein–Protein

Interactions. *Analytical Chemistry* 2016 *88* (16), 8315-8322DOI: 10.1021/acs.analchem.6b02240

CONFERENCE PRESENTATIONS

1. **Craig Bryant Gutierrez**, Ilan Chemmama, Haibin Mao, Clinton Yu, Sarah Block, Scott Rychnovsky, Ning Zheng, Andrej Sali, Lan Huang **Developing an integrative XL-MS strategy to facilitate structural modeling of protein complexes.** 67th ASMS Conference for Mass Spectrometry and Allied Topics. Atlanta, GA. USA. 2019. ***(Poster)***

2. **Craig Bryant Gutierrez**, Sarah A. Block, Clinton Yu, Stephanie M. Soohoo, Alexander S. Huszagh, Scott Rychnovsky, Lan Huang. **Developing a Novel Sulfoxide-containing MS-cleavable Cysteine-Reactive Homobifunctional Cross-linker to Define Protein-Protein Interactions.** 66th ASMS Conference for Mass Spectrometry and Allied Topics. San Diego, CA. USA. 2018. ***(Poster)***

3. **Craig Bryant Gutierrez**, Haibin Mao, Clinton Yu, Alex Huszagh, Rosa Viner, Eric Novitsky, Scott Rychnovsky, Ning Zheng, Lan Huang. **Developing an Integrated Cross-Linking Mass Spectrometry Platform to Determine Composition-Dependent Conformational Changes of Protein Complexes.** 65th ASMS Conference for Mass Spectrometry and Allied Topics. Indianopolis, Indiana. USA. 2017. ***(Selected Oral Presentation)***

4. **Craig Bryant Gutierrez**, Clinton Yu, Alex Huszagh, Eric Novitsky, Scott Rychnovsky, Lan Huang. **Developing A Novel Acidic Residue Reactive and Sulfoxide-containing MS-cleavable homobifunctional Cross-linker for Studying Protein-protein Interactions.** 64th ASMS Conference for Mass Spectrometry and Allied Topics. San Antonio, TX. USA. 2016. ***(Poster)***

PROFESSIONAL SOCIETY MEMBERSHIPS

- American Society of Mass Spectrometry

ABSTRACT OF THE DISSERTATION

Expanding the Chemical Cross-linking Tool Kit for
Cross-linking Mass Spectrometry
by

Craig Gutierrez

Doctor of Philosophy in Mass Spectrometry

University of California, Irvine, 2020

Professor Lan Huang, Chair

Cross-linking mass spectrometry (XL-MS) is a powerful tool for studying protein-protein interactions and elucidating architectures of protein complexes. However, one of the inherent challenges in MS analysis of cross-linked peptides is their unambiguous identification. To facilitate this process, we have previously developed a series of amine-reactive sulfoxide-containing MS-cleavable cross-linkers. These MS-cleavable reagents have allowed us to establish a common robust XL-MS workflow that enables fast and accurate identification of cross-linked peptides using multistage tandem mass spectrometry (MS^n). Although amine reactive reagents targeting lysine residues have been successful, it remains difficult to characterize protein interaction interfaces with little or no lysine residues. To expand the coverage of protein interaction regions, we present here the development of a series novel sulfoxide-containing cross-linker that target acidic residues (dihydrazide sulfoxide (DHSO)), cystine residues (bismaleimidesulfoxide (BMSO)), and finally a heterobifunctional cross-linker that targets lysine on one end and a non-specifically targets residues (Succinimidyl diazirine sulfoxide (SDASO)) on the other. We demonstrate that cross-linkers create cross-linked peptides that display the same predictable and characteristic fragmentation pattern during collision induced dissociation as amine-reactive sulfoxide-containing MS-cleavable cross-linked peptides, thus permitting their simplified analysis and unambiguous identification by MS^n. Furthermore, we applied these linkers to either characterize the yeast 26S proteasome (SDASO) or the Cop9 signalosome (DSSO, DHSO and BMSO) demonstrating both the feasibility of SDASO's photocross-linking of large protein complexes for the first time and the ability of multi-chemistry data for the integrative structure modeling of protein complexes to determine the interaction and structural dynamics of CSN, respectively. Moreover, our platform targeting various chemistries with cross-linking is robust and captures interactions complementary to residue-specific reagents, providing the foundation for future applications of multi-chemistry targeting approach to studying protein complexes.

INTRODUCTION

Proteins rarely act alone in cells, instead, often functioning in the context of multi-subunit protein complexes. These macromolecular entities participate in complex protein-protein interaction (PPI) networks essential for controlling a diverse range of biological processes. Dysregulation of endogenous PPI networks can be detrimental to cell homeostasis and viability, and has been linked to multiple human diseases. Thus, protein interaction interfaces have become attractive targets for drug discovery [1,2]. Therefore, full characterization of PPIs can help not only understand the assembly, structure and function of protein complexes, but also identify PPIs critical in human pathologies, diagnostics and therapeutics. Structural analyses of these complexes by traditional biophysical structural techniques such as x-ray crystallography and nuclear magnetic resonance (NMR) are frequently utilized to elucidate their topologies. The vast majority of structural information to date has been contributed by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, cumulatively totaling over 98% of all Protein Data Bank (PDB) entries (89% and 9%, respectively). Unfortunately, many large and heterogeneous complexes are refractory to such methods, ushering the development of new hybrid structural strategies. Recent progress in cryo-electron microscopy (cryo-EM) detector technology and digital image processing algorithms now permit near-atomic resolution density maps for protein complex structure elucidation, but cryo-EM density maps still only represent approximately 1% of all PDB entries. The growing demand for complementary structural elucidation tools has ushered in the development of alternative methods for protein complex characterization.

Over the past decade, mass spectrometry (MS)-based structural tools increasingly been implemented, accelerated by the technical difficulties experienced by traditional biophysical

structure methods when attempting to resolve the dynamics of conformationally and compositionally heterogeneous protein complexes. Various MS methodologies such as covalent labeling/footprinting [3,4], hydrogen−deuterium exchange (HDX) [5], cross-linking mass spectrometry (XL-MS) [6-10], ion-mobility MS (IM-MS) [11], and native MS [12] are considered valuable tools, because of their ability to generate low-resolution characterizations of protein complexes which were difficult to impossible to obtain using traditional methodologies. These tools provide incomplete yet valuable pieces of information, having the advantages of versatility, sensitivity, and throughput as compared to traditional means; this allows these tools to supplement conventional analysis. These hybrid combinations of MS tool and traditional structural methodologies allows for the architectural elucidation of multimeric protein complexes which have remained recalcitrant to traditional methodologies alone. Over the past several years these MS-based strategies have been used extensively to complement and add to traditional structural methodologies resulting in an expansion of the structural biology knowledgebase changing how structural biology is perceived [13-30]

Owing to recent technological innovations, cross-linking mass spectrometry (XL-MS) has become a powerful and effective method for studying PPIs *in vitro* and *in vivo* [10,16,26,28,31-33]. In comparison to other MS structural tools, XL-MS is unique due to its capability to simultaneously capture PPIs from their native environment and uncover their physical interaction contacts, thus permitting the determination of both identity and connectivity of PPIs in cells [10,16,28,32,33]. In addition, identified cross-links provide distance constraints to facilitate three-dimensional modeling of protein complexes by refining existing high-resolution protein structures and/or complementing low-resolution structures to elucidate architectures of large protein complexes [6,9,26,31,33], which has remained recalcitrant to traditional methodologies alone.

XL-MS is orchestrated by a simple process driven workflow which is illustrated in Figure 0.1. First, proteins are reacted with the bifunctional cross-linking reagents which physically link proximal amino acid residues through the formation of covalent bonds. The crosslinked proteins are then enzymatically digested, then the resulting peptide mixtures are separated and analyzed via liquid chromatography−tandem mass spectrometric (LC-MS/MS) analysis. Subsequent database searching of MS data identifies cross-linked peptides and their linkage sites. Given that each cross-linking reagent carries a defined length, the resulting cross-links can be utilized as distance constraints for various applications, ranging from structure validation and integrative modeling [34-39] to de novo structure prediction [38,40,41].

While the principle concept of cross-linking XL-MS was conceived around a half century ago [42], only recently has technology matured to the point of practical application allowing it to become powerful structural tool for mapping protein−protein interactions in recent years. However, with all methodologies certain limitations needed to be worked around or overcome to make its practical application more feasible along with advances in technology. The application of XL-MS has faced three primary obstacles: (1) complex tandem mass spectrometry (MS/MS or MS2) fragmentation of cross-linked peptides; (2) low abundance of cross-linked peptides in complex peptide mixtures; (3) heterogeneity of cross-linked products. The latter two hinder effective MS detection of crosslinked peptides. These latter applications have been over come through development of chromatographic enrichment techniques and technologies, such as Peptide SEC [43] or enrichment handles on the linkers themselves [7,44-50], to aid in the extraction of XL peptides. The first hurdle which poses the greatest technical challenge, makes accurate identification of cross-linked peptides and unambiguous assignment of cross-linked sites difficult.

Along with software development, to address this challenging task to deconvolute complicated MS spectra, significant progress has been made to enhance the identification and characterization of cross-linked peptides resulting in the development of novel cleavable crosslinking reagents. [51]. These cleavable reagents carry a liable bond in the spacer region between the reactive ends. These liable bonds can be cleaved by various means depending upon the unique chemical properties each possesses, such as photo- [52], chemical- [53,54], and MS-induced cleavages [7,44-46,55-64]. MS-cleavable linkers have taken the lead in this category as fragmentation during MS-analysis enables easy correlation between separated cross-link peptide constituents and their respective parent ions. MS-cleavable cross-linking reagents have significantly facilitated MS analysis of cross-linked peptides in complex mixtures, due to their unique capability of eliminating the "n-square" problem and permitting effective sequencing of cross-linked peptides [10,51].

Various types of cleavable bonds have been incorporated into current MS-cleavable cross-linkers. Among them, collision-induced dissociation (CID)-cleavable bonds are most popular as they are present in the majority of existing MS-cleavable reagents. The two most commonly used CID-cleavable bonds are (1) C−S bonds adjacent to sulfoxide [46,60,64-66], sulfonium ion [56], or cyanuric ring [7] and (2) C−N bonds associated with D-P (Asp-Pro) bonds [47,67,68], Rink [45,48,55], urea [59,69,70], or quaternary diamine [71] structures. Depending on the strength of these cleavable bonds, the reagents can be implemented into MS^n and/or MS^2-based strategies to accurately identify of cross-linked peptides.

To enable robust MS-cleavability, the Huang lab have developed a series of sulfoxide-containing MS-cleavable cross-linking reagents (e.g. disuccinimidyl sulfoxide (DSSO)) [46,49,60,64-66]. The MS-labile C-S bonds adjacent to the sulfoxide can be preferentially fragmented prior to

peptide backbone cleavage upon collision-induced dissociation (CID), physically separating the two cross-linked peptide constituents for individual sequencing. Notably, this predictable fragmentation occurs independent of cross-linking chemistry, peptide charge and peptide sequence. These unique characteristics allow straightforward and unambiguous identification of cross-linked peptides by MS^n analysis coupled with conventional database searching tools. Sulfoxide-containing MS-cleavable cross-linkers have been successfully applied to not only study PPIs *in vitro* [60,72-74] and *in vivo* [46,75], but also to dissect structural dynamics of protein complexes [76-78].

To date, amine-reactive homobifunctional NHS ester cross-linkers have been the most popular reagents in XL-MS studies. This is due to the relatively high occurrence of lysines—particularly at the surfaces of protein structures—as well as the specificity and efficiency of amine-reactive chemistries. Although effective, these reagents alone cannot yield complete PPI maps, as profiling of interaction regions lacking lysines would be difficult. Unfortunately, only a limited number amino acid-specific cross-linkers (including the ones developed in this thesis) have been developed to complement lysine-reactive reagents, including carboxyl- [65,79,80], sulfhydryl- [66,81], arginine- [82], and multi-residue targeting ones [83-85], clearly expanding PPI coverage. In addition, integration of multiple cross-linkers has improved characterization of PPIs and increased the depth and accuracy of structural analysis [39,76,77,86], demonstrating the benefits of multi-chemistry based combinatory XL-MS approaches.

A specific cross-linking chemistry targeting other amino acid residues abundant at protein interaction sites would be ideal for complementing lysine targeting cross-linkers. While hydrophobic amino acid residues often constitute the cores, charged hydrophilic residues such as lysine, arginine, aspartic acid (Asp), and glutamic acid (Glu) often occupy surface-exposed

regions of protein complexes, making them ideal targets for mapping protein interactions. According to a recent SwissProt database release [87], aspartic and glutamic acids comprise roughly 12.2% of all amino acid residues, compared to the 5.8% of lysines. Therefore, acidic residues (i.e. aspartic and glutamic acids) represent high potential targets for cross-linking studies due to their abundance and prevalence at interaction interfaces. A recent study by Leitner et al. (2014) has demonstrated the feasibility of acidic residue-specific cross-linking chemistry to study protein interactions using non-cleavable homobifunctional dihydrazide cross-linkers in conjunction with the coupling reagent 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) [79]. This methodology is an improvement on the acidic residue cross-linking chemistry involving the coupling reagent 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) that requires the cross-linking reaction to occur at a pH of 5.5 [88]. In comparison, DMTMM coupling with dihydrazide cross-linkers is compatible with proteins at neutral pH (7.0-7.5) and therefore better suited for studying the structures of proteins and protein complexes under physiological conditions. However, this cross-linking strategy remains susceptible to the challenges associated with traditional cross-linking reagents in unambiguously identifying cross-linked peptides and their linkage sites. Due to the increased prevalence of Asp and Glu in protein sequences, the accurate and unambiguous identification of peptides containing non-cleavable dihydrazide cross-linked acidic residues would be intrinsically more complicated than the identification of lysine cross-linked peptides. Therefore, one of the main goals of the research detailed in Chapter 1 was to integrate dihydrazide reactive ends into the sulfoxide backbone system established with the creation of DSSO [60] to generate the new linker Dihydrazide Sulfoxide (DHSO) [65]. Chapter 1 verifies the

MS^n analysis pioneered in the Kao et al. (2014) worked for the new linker DHSO enabling the robust accurate identification of acidic residue targeted cross-linked peptides [65].

Apart from lysines and acidic residues, cysteines are useful alternatives for protein cross-linking due to several factors. First, the high specificity and efficiency of sulfhydryl chemistry has permitted its widespread adoption in a gamut of proteomics studies. Secondly, cysteine cross-linking can be more selective and informative due to the lower prevalence of cysteines compared to lysine and acidic residues. Therefore, developing new cysteine-targeting cross-linkers would be advantageous for profiling PPIs, and can further complement existing reagents. Among various types of cysteine-reactive reagents, maleimides are most widely utilized due to the specificity and efficiency of thiol-maleimide coupling [89]. In addition, maleimide moieties are also easily functionalized and reactions can occur at physiological pH in the absence of catalysts or heating, making these reagents well-suited for a variety of experimental uses. Such chemistry has been successfully employed in cross-linking studies to probe protein structures and identify interaction domains [89-91]. While homobifunctional maleimide cross-linkers are commercially available, their uses in XL-MS studies remain sparse. This is more likely due to difficulty in MS identification of cross-linked peptides. Therefore, as described in Chapter 2, we developed a homobifunctional cysteine targeting MS-cleavable crosslinker, Bismaleimide sulfoxide (BMSO), based on the previously developed sulfoxide backbone seen in both DSSO and DHSO [60,65,66]. However, despite these successes, mapping interaction regions lacking targetable residues by specific chemistry remained challenging. Therefore, a need still exists to develop a linker that can non-specifically target amino acid residues.

In recent years, photochemistry has shown great potential in capturing regions inaccessible to residue-specific cross-linkers due to its nonspecific reactivity [10,21,92,93]. Various

types of photoreactive reagents have been explored in XL-MS studies [81,94-101], almost all of which have been heterobifunctional cross-linkers with an amine-reactive specific end and a nonspecific end. Among the commonly used photoreactive groups, alkyl diazirine is most attractive due to its small size, long excitation wavelength, photostability, reactivity, and proven success in XL-MS studies [94,96-101]. Diazirines are activated by UV light to yield highly reactive carbenes, which then react with an X-H bond (X: C, N, O, S) of any proximal amino acids [96,97,99-102]. While promising, the indiscriminate nature of photocross-linking often results in highly complex and low abundance cross-linked products that complicate MS analysis and database searching, thus limiting its application predominantly to single proteins [96-100]. Therefore, to advance photo-reactive XL-MS studies for complex PPI mapping, it is essential to develop novel reagents that permit effective MS detection and accurate identification of photocross-linked peptides. Due to the need for an MS-cleavable nonspecific residue targeting crosslinker, we developed a series of heterobifunctional lysine to non-specific targeting MS-cleavable crosslinker, Succinimidyl Diazirine Sulfoxide (SDASO) -long (L), -medium (M), and -short (S), based on the previously developed sulfoxide backbone seen in DSSO DHSO, and BMSO [60,65,66], as described in Chapter 3.

While we had a primary interest in developing MS-cleavable crosslinkers targeting multiple chemistries, we also wanted to evaluate the performance of these linkers in a dynamic multi-subunit proteins complex. This would provide the opportunity to evaluate both the complementarity of targeting multiple chemistries but also the ability generates enough information to be able to model an in-solution protein complex. To do this we trained are focus on the COP9 signalosome (CSN). The CSN is an evolutionarily conserved and essential multi-subunit protein complex involved in diverse cellular and developmental processes in animals and plants

[103-105]. CSN functions as a deneddylase, specific for cleaving Nedd8 modification from cullin proteins, the key components of Cullin–RING ubiquitin E3 ligases (CRLs) [106-110]. CRLs represent the largest evolutionarily conserved superfamily of multi-subunit E3s [107,108] which embody ~30% of all human E3 proteins and coordinate degradation of ~20% of the proteins processed by the proteasome. The dynamic cycle of neddylation and deneddylation of cullins is a critical step in regulating the assembly and activity of CRLs [108,111,112]. In addition to enzymatic regulation of CRLs, CSN can inactivate CRLs non-catalytically by direct binding, preventing their association with E2 enzymes and ubiquitination substrates [113-116]. While abnormal CRL activity is frequently associated with various human diseases, multiple studies have also identified CSN as a positive regulator of oncogenes and negative regulator of tumor suppressors [117-121]. Moreover, elevated expression of CSN subunits has been found in a number of human tumors, often with poor prognosis [118,122]. Therefore, better understanding of CSN structure would provide new insights on their function and the regulation of CRLs associated with human pathology.

The canonical CSN complex (hereafter referred to as CSN) typically consists of eight subunits (CSN1-8) [103,105], including six different PCI (Proteasome lid-CSN-Initiation factor 3) domain-containing subunits (CSN1–CSN4, CSN7 and CSN8) and two MPN (MPR1/PAD1 amino-terminal) domain-containing proteins (CSN5 and CSN6). Among them, CSN5 is the catalytic subunit directly responsible for CSN deneddylase activity [106]. The CSN complex shares sequence similarities to the 19S proteasome lid subcomplex and the eukaryotic translation initiation complex eIF3, which also contain PCI and MPN domains [103,105]. The X-ray structure of recombinant human CSN has revealed that CSN5 and CSN6 MPN domains form a heterodimer, while the six remaining PCI subunits assemble into a horseshoe-shaped ring from which their arm-like α-helical domains project [123]. The PCI subunits provide a scaffold, primarily through CSN2

and CSN4, which facilitates the recruitment of neddylated CRLs. Meanwhile, the two MPN subunits are slightly juxtaposed, exposing the active MPN catalytic core in CSN5 [114,124-126]. All eight subunits are united in a helical bundle formed by their C-terminal carboxyl α-helices, which are stacked between the CSN5–CSN6 dimer and PCI ring. Interestingly, substrate-free CSN exists in an inactive, auto-inhibited state [124]. Structural and biochemical characterization of CSN-CRL complexes have revealed substrate-induced structural dynamics associated with CSN activation [114,124-127]. Binding of neddylated CRLs to CSN triggers substantial remodeling and extensive conformational changes of the complex, activating the isopeptidase activity of CSN5. Although the structural plasticity of CSN is important for CSN activation and function in regulating CRL activities in cells, it has not been well characterized due to limitations in existing technologies.

Recently, the 9th CSN subunit, CSN9 (a.k.a. CSNAP (CSN acidic protein)), has been discovered to complex with CSN1-8 stoichiometrically to form a 9-membered non-canonical CSN complex (a.k.a. CSN9-bound CSN, hereafter referred to as CSNn) [128]. As canonical CSN subunits (CSN1-8) have a one-to-one correspondence to the subunits of the 19S proteasome lid subcomplex [105,129], CSN9 is homologous to DSS1, the smallest component of the 19S lid. While CSN9 is not essential for the assembly and catalytic activity of CSN [128], a recent study has suggested that CSN9 reduces the affinity of CSN-CRL interactions, contributing to steric regulation of CRLs [116]. The depletion of CSN9 appears to have a global impact on CRL-associated activities, leading to altered reproductive capacity, suppressed DNA damage response, decreased viability and delayed cell cycle progression [116]. It has also been suggested that the C-terminus of CSN9 is important in its incorporation within the CSN complex, likely through interactions with CSN3, CSN5 and CSN6 [128]. However, due to its small size and highly disordered structure, it remains challenging to accurately determine interaction interfaces between CSN9 and CSN. As a result, no high-

resolution structures are available for the CSN9-bound CSN complex. Thus, alternative strategies to dissect the architecture of the non-canonical CSN complex and determine how CSN9 interacts with CSN1-8 are needed to help us uncover structural details underlying the functional importance of CSN9 in cells. Chapter 4 details the application of a multi-reside targeting cross-linking approach using DSSO, DHSO and BMSO to dissect the molecular architecture of both the CSN and CSNn.

In summary, the specific aims of this dissertation were to develop a series novel MS-cleavable crosslinkers targeting a variety of chemistries including: acidic-residues (DHSO, Chapter1), cysteine residues (BMSO, Chapter 2) and nonspecific residues (the SDASO series, Chapter 3). Once the first aims were accomplished, the final aim was to apply some of these novel linkers in order to characterize the dynamics of the Cop9 Signalosome with and without the newly discovered 9th subunit CSN9 (Chapter 4).

Chapter 1

Developing a Novel Acidic Residue Reactive and Sulfoxide-containing MS-cleavable Homobifunctional Cross-linker for Probing Protein-Protein Interactions

A recent study by Leitner et al. (2014) has demonstrated the feasibility of acidic residue-specific cross-linking chemistry to study protein interactions using non-cleavable homobifunctional dihydrazide cross-linkers in conjunction with the coupling reagent 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) [79]. This methodology is an improvement on the acidic residue cross-linking chemistry involving the coupling reagent 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) that requires the cross-linking reaction to occur at a pH of 5.5 [88]. In comparison, DMTMM coupling with dihydrazide cross-linkers is compatible with proteins at neutral pH (7.0-7.5) and therefore better suited for studying the structures of proteins and protein complexes under physiological conditions. However, this cross-linking strategy remains susceptible to the challenges associated with traditional cross-linking reagents in unambiguously identifying cross-linked peptides and their linkage sites. Due to the increased prevalence of Asp and Glu in protein sequences, the accurate and unambiguous identification of peptides containing non-cleavable dihydrazide cross-linked acidic residues would be intrinsically more complicated than the identification of lysine cross-linked peptides. To simplify MS analysis and facilitate the identification of acidic residue cross-linked peptides, we have developed a sulfoxide-containing MS-cleavable acidic residue-specific homobifunctional cross-linking reagent, dihydrazide sulfoxide (DHSO, a.k.a. 3,3'-sulfinyldi(propanehydrazide)). This reagent adopts the same MS-labile sulfoxide chemistry as our previously developed amine-reactive MS-cleavable cross-linkers (i.e. DSSO, DMDSSO and

Azide-A-DSBSO), thus enabling robust and unambiguous identification of cross-linked peptides via the same XL-MS^n workflow[46,60,64]. DHSO represents a novel class and the first generation of acidic residue-targeting cross-linking reagents with MS-cleavability. We expect that DHSO-based XL-MS strategies will become an invaluable tool in providing a complementary subset of cross-linking data towards a comprehensive structural elucidation of protein complexes by XL-MS.

Experimental Methods for this section can be found in Appendix A.

RESULTS

Design and Synthesis of a Novel Acidic Residue-targeting Sulfoxide-containing MS-cleavable Cross-linker

In order to facilitate accurate identification of acidic residue cross-linked peptides, we aimed to develop a novel MS-cleavable cross-linking reagent specific to Asp and Glu residues. This requires the incorporation of a functional group with robust MS-inducible cleavage sites located in the spacer region of the cross-linker. Previously, we successfully developed a novel class of amine-reactive, sulfoxide-containing MS-cleavable cross-linkers, i.e., DSSO [60], DMDSSO [64] and Azide-A-DSBSO [46] (Fig. 1.1A-C). The C-S bonds adjacent to the sulfoxide group(s) in these reagents have proven to be reliable labile bonds that fragment selectively and preferentially prior to the breakage of peptide backbone during collision induced dissociation. Additionally, such fragmentation is predictable and occurs independently of peptide charge and sequence. These unique features facilitate the simplified analysis of sulfoxide-containing cross-linked peptides and their unambiguous identification by MS^n [46,60,64]. Following the success of our MS-cleavable, amine-reactive cross-linkers, we designed a novel acidic residue-reactive, MS-cleavable homobifunctional dihydrazide cross-linker incorporating a sulfoxide group in the spacer region, i.e., dihydrazide sulfoxide (DHSO). DHSO is synthesized from DSSO with two additional

synthesis steps (Fig. 1.1D). As shown, DHSO is composed of two hydrazide reactive groups and two symmetrical C-S cleavable bonds flanking a central sulfoxide. The spacer length of DHSO is 12.4 Å (calculated between the terminal nitrogen atoms). In comparison to existing cross-linkers for XL-MS studies [46,60,64,79], DHSO carries a linker length well suited for defining interaction interfaces between and within protein complexes.

CID Fragmentation Patterns of DHSO Cross-linked Peptides

A previous study has shown that the reaction of hydrazide cross-linkers with acidic residues first requires activation of the terminal carboxyl groups of Asp (D) and Glu (E) side chains or protein C-termini [79]. The coupling reagent DMTMM has been demonstrated to be effective in activating carboxylic acid groups to form a reactive intermediate that can be displaced by nucleophilic attack from hydrazides under physiological pH [79] (Fig. 1.2A). Therefore, in this work, we have adopted DMTMM as the activating agent for DHSO cross-linking of acidic residues. Similar to lysine-reactive cross-linkers, DHSO cross-linking would result in the formation of three types of cross-linked peptides: dead-end (type 0), intra-link (type 1), and inter-link (type 2) modified peptides, among which inter-linked peptides provide the most informative data on the relative spatial orientation of cross-linked acidic residues [130]. Since all of the MS-cleavable, homobifunctional NHS esters we have previously developed display the same characteristic fragmentation patterns in MS^2 due to the cleavage of either of the two symmetric CID-cleavable C-S bonds adjacent to the sulfoxide functional group [46,60,64], we expect that DHSO cross-linked peptides will behave similarly during MS^n analysis even though their residue-targeting functional groups are different.

To elaborate this process, Figures 1.1E and 1.2B-C illustrate the predicted MS^2 fragmentation patterns of DHSO inter-linked, intra-linked and dead-end modified peptides,

respectively. For a DHSO inter-linked peptide α-β, the cleavage of one of the two symmetric C-S bonds would result in one of the two predicted peptide fragment pairs (i.e. α_A/β_S or α_S/β_A). The resulting α and β peptide fragments are modified by complementary cross-linker remnant moieties, i.e., alkene (A) or sulfenic acid (S). However, the sulfenic acid moiety can undergo dehydration to become a more stable unsaturated thiol moiety (i.e. T) (Fig. 1.2D). This conversion has been commonly observed in amine reactive, sulfoxide containing MS-cleavable cross-linked peptides, thus leading to the detection of α_A/β_T and α_T/β_A pairs instead as the four dominant MS^2 fragment ions [46,60,64]. Therefore, these two MS^2 fragment pairs (i.e. α_A/β_T and α_T/β_A) are expected for a DHSO cross-linked heterodimer as well (Fig. 1.1E), which can then be subjected to MS^3 analysis for unambiguous identification of cross-linked peptide sequences and cross-linking sites. For a DHSO intra-linked peptide α_{intra} in which proximal D or E amino acid residues are cross-linked within the same peptide, one peptide fragment (i.e. α_{A+T}) is expected in MS^2 analysis (Fig. 1.2B). In reality, this particular ion would represent two populations of ion species that have identical peptide sequences and m/z values, but transposed DHSO remnant-modified acidic residues. Lastly, a DHSO dead-end modified peptide α_{DN} would potentially fragment into two ion species during MS^2 analysis. Depending on the position of the cleaved C-S bond, α_A or α_T fragments would be observed, resulting in a pair of daughter ions detected during MS^2 (Fig. 1.2C). The distinct MS^2 fragmentation patterns of sulfoxide-containing MS-cleavable cross-linked peptides result in predictable mass relationships between parent ions and their respective fragments. These mass relationships are utilized as an additional verification of cross-linked peptide identification at the MS^2 level. Along with mass fingerprinting by MS^1 and peptide sequencing by MS^3, three lines of evidence can be obtained and integrated to accurately identify DHSO cross-linked peptides using

the identical MS^n workflow that has been developed for the analysis of DSSO, DMDSSO and DSBSO cross-linked peptides [46,60,64].

Characterization of DHSO cross-linked Model Peptides by MS^n Analysis

Despite the similarities in spacer arm structure to DSSO, it is necessary to verify whether DHSO cross-linked peptides indeed fragment as described above during MS^n analysis (Fig. 1.1E). Initial characterization of DHSO was performed on a synthetic peptide containing a single acidic residue, Ac-SR8 (Ac-SAKAYEHR). Inter-linked Ac-SR8 homodimer was detected as quadruply-charged (*m/z* 548.7623^{4+}) and quintuply-charged (*m/z* 439.2117^{5+}) ion species, respectively. Since the two peptide sequences of inter-linked homodimer are the same, only one pair of MS^2 fragment ions (i.e. α_A/α_T) would be expected. Indeed, MS^2 analysis of the quadruply-charged parent ion produced a pair of dominant fragment ions α_A/α_T (*m/z* $536.27^{2+}/552.26^{2+}$), demonstrating effective physical separation of the two cross-linked peptides as expected (Fig. 1.3A). Similarly, MS^2 analysis of the quintuply-charged parent ion (*m/z* 439.2117^{5+}) yielded a single pair of dominant fragment ions α_A/α_T (*m/z* $357.85^{3+}/552.26^{2+}$) as well (Fig. 1.3B), demonstrating the characteristic fragmentation independent of peptide charges as expected. Subsequent MS^3 analyses of the α_A (m/z 536.27^{2+}) fragment ion (Fig. 1.3C) resulted in series of y and b ions that unambiguously confirmed the peptide sequence as Ac-SAKAYE_AHR in which the glutamic acid was modified with a DHSO alkene (A) moiety. Similarly, MS^3 spectrum of the α_T fragment (m/z 552.26^{2+}) determined its identity as Ac-SAKAYE_THR, in which the glutamic acid was modified with a DHSO unsaturated thiol (T) moiety (Fig. 1.3D). Therefore, the cross-linked peptide was identified as [Ac-SAKAYE_AHR] inter-linked to [Ac-SAKAYE_THR] through E6 in both peptides. This result indicates that DHSO inter-linked peptides indeed display the same characteristic MS^n

fragmentation as sulfoxide-containing lysine inter-linked peptides, and can be analyzed using the same data analysis workflow as previously described [46,60,64].

Characterization of DHSO Cross-linked Model Proteins by MS^n Analysis

To evaluate the capability of DHSO for protein cross-linking *in vitro*, we used equine myoglobin and bovine serum albumin (BSA) as our model proteins. These two proteins contain above-average acidic residue content (16.3% and 13.6%, respectively), making them well suited for evaluating DHSO cross-linking. In addition, BSA was employed previously for acidic residue cross-linking by non-cleavable dihydrazides [79]. To identify DHSO cross-linked peptides in myoglobin and BSA, we have performed in-gel digestion of gel-separated DHSO cross-linked proteins or in solution digestion of DHSO cross-linked proteins followed by peptide SEC as illustrated (Fig. 1.4). The resulting peptides were subjected to LC MS^n analysis. Figure 1.5A displays the MS^1 spectrum of an exemplary inter-linked peptide (α-β) (*m/z* 517.2703^{5+}) identified from myoglobin. Its MS^2 analysis resulted in the detection of two peptide fragment pairs, i.e. α_A/β_T (*m/z* $429.74^{2+}/569.63^{3+}$) and α_T/β_A (*m/z* $445.72^{2+}/559.64^{3+}$) (Fig. 1.5B), characteristic for DHSO inter-linked heterodimers. MS^3 analysis of α_A (*m/z* 429.74^{2+}) (Fig. 1.5C) determined its sequence as ASE$_A$DLKK, in which the glutamic acid residue at the 3rd position from the N-terminus was modified with an alkene moiety. MS^3 analysis of β_T (m/z 569.63^{3+}) identified its sequence as VEAD$_T$IAGHGQEVLIR, with the aspartic acid residue at the 4th position from the N-terminus carrying an unsaturated thiol moiety (Fig. 1.5D). Collectively, the inter-linked peptide was unambiguously identified as [^{18}VEAD$_T$IAGHGQEVLIR^{32} cross-linked to ^{58}ASE$_A$DLKK^{64}], describing an inter-link formed between D21 and E60 of equine myoglobin.

Figure 1.6 displays MS^n analysis of a representative DHSO inter-linked BSA peptide, which was measured as a quadruply-charged ion (*m/z* 692.8475^{4+}) in MS^1 (Fig. 1.6A). Its MS^2

spectrum revealed two pairs of complementary MS^2 fragment ions, i.e. α_A/β_T and α_T/β_A (Fig. 1.6B), further demonstrating the robust fragmentation expected of DHSO inter-linked peptides. Together with MS^3 sequencing of MS^2 fragments α_A (*m/z* 616.32^{2+}) and β_T (m/z 760.36^{2+}) (Fig. 1.6C-D), this DHSO inter-linked peptide was unambiguously identified as [^{66}LVNE$_A$LTEFAK^{75} inter-linked to ^{89}SLHTLFGDE$_T$LCK^{100}], in which residue E69 cross-linked to residue E97 in BSA.

In addition to inter-linked peptides, intra-linked peptides were also observed as a result of DHSO cross-linking of our model proteins. For example, MS^2 fragmentation of an intra-linked myoglobin peptide (Fig. 1.7) produced a single fragment ion peak α_{A+T} (m/z 514.02^{4+}) 18 Da less than its parent ion, consistent with the expected fragmentation pattern described in Figure 1.2B following dehydration of the sulfenic acid moiety to an unsaturated thiol moiety. Analysis of the α_{A+T} ion in subsequent MS^3 analysis (Fig. 1.7C) yielded series of b and y ions permitting the unambiguous identification of two peptides sharing identical sequences but transposed alkene and unsaturated thiol moieties: ^{105}YLE$_A$FISD$_T$AIIHVLHSK^{119} and ^{05}YLE$_T$FISD$_A$AIIHVLHSK^{119}, indicating an intra-link between residues E106 and D110.

MS^2 fragmentation of a myoglobin dead-end modified peptide (*m/z* 604.3095^{3+}) resulted in the detection of a single pair of fragment ions α_A/α_T (*m/z* $559.30^{3+}/569.96^{3+}$) (Fig. 1.8), consistent with the expected fragmentation pattern described in Figure 1.2. These fragment ions were then identified by MS^3 analysis as ^{18}VE$_A$ADIAGHGQEVLIR^{32} and ^{18}VE$_T$ADIAGHGQEVLIR^{32} respectively, representing a dead-end cross-link located on E19 of myoglobin (Fig. 1.8C -D).

In total, LC-MS^n analysis of DHSO cross-linked myoglobin identified 33 unique inter-linked peptides, representing 32 unique E|D-E|D linkages (Table 1.1). Similarly, 62 unique DHSO inter-linked BSA peptides were identified, describing 69 unique E|D-E|D linkages (Table 1.2). Collectively, the results presented thus far indicate that DHSO can effectively cross-link acidic

residue containing peptides and proteins in the presence of DMTMM at neutral pH. More importantly, our results have demonstrated that DHSO cross-linked peptides indeed exhibit the same characteristic MS^2 fragmentation patterns as expected to allow their facile and accurate identification.

DHSO Cross-linking Maps of Myoglobin and BSA

In order to assess the efficacy and sequence coverage of DHSO cross-linking on our model proteins, we generated cross-linking maps of myoglobin and BSA based on their identified DHSO inter-linked peptides. The secondary structures of equine myoglobin comprise of eight α-helices and one short 3_{10} helix (PDB: 1DWR) (Fig. 1.9A). The globular nature of myoglobin suggests that many of the helices are in close proximity to one another in three-dimensional space. The DHSO cross-link map of myoglobin based on the 33 unique E|D-E|D linkages is illustrated in Figure 1.4B, describing numerous intra- and inter- secondary structure interactions (i.e. α1-α5, α1-α8, α2-α4, α3-α4, α3-α8, α4-α5, α4-α8, α6-α8, α7-α8, and α8-α8). To further evaluate the identified cross-links, we mapped the cross-linked residues onto the crystal structure of myoglobin and calculated the distances between their alpha carbons (Cα-Cα distances) (Figs. 1.9D and 1.9F). Considering the spacer length of the DHSO (12.4Å) and the distances contributed by D|E side chains (3.8Å|4.9Å, respectively), as well as backbone flexibility and structural dynamics, the theoretical upper limit for the Cα-Cα distances between DHSO cross-linked acidic residues is estimated to be ~30Å. 27 of the 32 myoglobin DHSO cross-links were mapped in the structure, with 26 having Cα-Cα distances < 30Å and one link slightly over the maximum distance at 31.1 Å. The remaining 5 linkages were not mapped on to the structure because they were identified as sites of oligomerization, in which identical residues or peptide sequences were cross-linked together.

Similarly, a DHSO cross-link map of BSA was generated based on the 69 unique E|D-E|D linkages (Fig. 1.10A). When mapped to a previously published BSA crystal structure (PDB: 4F5S), 64 out of 69 BSA linkages (93%) were calculated to have Cα-Cα distances below 30Å (Fig. 1.11A and 11C). Structural flexibility and/or oligomerization of BSA likely contribute to the other five identified linkages found to be > 30Å. As shown in Figure 1.10A, DHSO inter-links were distributed throughout the primary sequence of BSA, with regions of dense cross-link clusters identified in regions with higher α-helix density. This even distribution is likely due to the dispersion of aspartic acid and glutamic acid residues throughout the protein. Collectively, our results suggest that DHSO cross-linking yields cross-links within expected distance constraints useful for structural elucidation for computational modeling in the same way as lysine cross-linked data.

Comparison of MS-cleavable and non-Cleavable Acidic Residue Cross-linking

Previously, two non-cleavable acidic residue cross-linkers, i.e., adipic acid dihydrazide (ADH) and pimetic acid dihydrazide (PDH), were used for probing the structure of bovine serum albumin [79] which resulted in the identification of 27 and 35 unique acidic residue linkages, respectively [79]. A comparison of the linkage maps generated for DHSO, ADH, and PDH cross-linking of BSA (Fig. 1.10A-C) revealed a high degree of similarity in proximally cross-linked regions. Apart from covering interaction regions cross-linked by ADH and PDH, DHSO cross-linking resulted in 34 additional unique D|E-D|E linkages. These unique DHSO cross-links are generally clustered in regions of particularly high acidic residue density, such as the regions between D25 and D97, E250 and E344, and D405 to E494 (Fig. 1.10A-C). Limitations in bioinformatics software for analyzing non-cleavable cross-linked peptides have been previously noted [79], which made the accurate identification of acidic residue cross-linked peptides

considerably more challenging due to their higher frequency and corresponding increase in search space. In contrast, CID induced cleavage of DHSO cross-linked peptides during MS^2 significantly simplified subsequent peptide sequencing in MS^3. Given the same acidic residue reactive chemistry, the increase in identified cross-links using DHSO is mainly attributed to the simplified cross-link identification with improved accuracy afforded by MS-cleavability of DHSO cross-linked peptides. This ultimately facilitates unambiguous identification of individual linkages amidst peptides with multiple acidic residues in sequence. These results demonstrate the advantage of using DHSO, a MS-cleavable cross-linking reagent targeting acidic residues for probing protein-protein interactions over non-cleavable reagents.

Comparison of DHSO and DSSO Cross-linking

To assess the complementarity between acidic residue and primary amine cross-linking data, we examined the similarities and differences between DHSO and DSSO cross-linking of our selected model proteins. To this end, we also carried out LC-MS^n analyses of DSSO cross-linked myoglobin and BSA respectively. As summarized in Tables 1.3 and 1.4, 19 unique DSSO inter-linked myoglobin peptides and 33 unique DSSO inter-linked BSA peptides were identified. These linkages were then mapped onto their corresponding protein linear sequences (Figs. 1.9C and 1.10D) and crystal structures (Figs. 1.9E and Fig. 1.11B). As a result, all of myoglobin DSSO cross-links (Fig. 1.9F) and 94% of BSA DSSO cross-links corresponded to Cα-Cα distances ≤ 30Å (Fig. 1.11C). The two BSA cross-links that are outside the distance range may be a result of unexpected structural flexibility.

In the case of myoglobin, DSSO cross-linking identified several proximal helicase regions, such as α4-α5, α4-α7, α5-α8, α6-α8, and α5-3_{10}. In comparison, there is limited overlap between DHSO and DSSO cross-link maps except in regions containing α4-α5 and α6-α8 (Fig. 1.11A-C),

indicating that DHSO and DSSO cross-linking mapped different parts of interactions within myoglobin. The identified helicase interacting regions unique to DHSO or DSSO cross-linking correspond well with the number of cross-linkable residues and specific reactive chemistries. This is due to the fact that lysine and acidic residues are distributed unevenly across myoglobin sequence. For example, the N-terminal region of myoglobin (residues 1-41) spanning helices α1 through α3 contains only one lysine, but four glutamic acids and two aspartic acid residues. Therefore, profiling the interactions of the N-terminus within itself and with other parts of the protein will be difficult with amine-reactive cross-linking reagent such as DSSO. In contrast, acidic residue reactive cross-linker DHSO would be better suited for this purpose. Indeed, while DSSO was not able to cover this region as expected, DHSO cross-linking enabled the identification of 11 inter-linked peptides describing multiple interactions between the N-terminus and other parts of the protein (i.e. α1-α5, α1-α8, α2-α4, α3-α4, and α3-α8). While DHSO provided exclusive data from the lysine scarce N-terminus, the lysine-rich 3_{10} helix and many of the loop regions between the helical structures were better analyzed by DSSO due to the higher abundance of lysine residues in these regions. Together, these results demonstrate that acidic residue cross-linking can provide complementary structural information to that obtained using amine-reactive cross-linkers.

Interestingly, unlike myoglobin, DHSO and DSSO cross-linking of BSA have resulted in much more similar cross-linking profiles, meaning that similar interactions within BSA were identified (Figures .110A and 1.10D). This is most likely owing to the fact that BSA has more evenly dispersed distribution of lysine, aspartic acid, and glutamic acid residues throughout the protein sequence. Thus, combined usage of DHSO and DSSO can strengthen the validity of the cross-links identified by any of the two reagents individually. More importantly, this will generate

complementary structural information to facilitate a more comprehensive understanding of protein structures.

DISCUSSION

Here we report the development and characterization of a new acidic residue-targeting, sulfoxide-containing MS-cleavable cross-linker, dihydrazide sulfoxide (DHSO), which is a new derivative of our previously developed amine-reactive MS-cleavable reagent, DSSO [60]. Our analyses here have proven that DHSO cross-linked peptides possesses the same characteristics distinctive to peptides cross-linked by sulfoxide-containing amine reactive cross-linkers [46,60,64], thus permitting their fast and accurate identification by MS^n analysis. The unique features of DHSO will significantly facilitate cross-linking studies targeting acidic residues, which has been difficult in the past due to the large number of D|E present in protein sequences and complexity of their resulting cross-linked peptides for MS analysis. Comparison of DHSO and DSSO cross-linking confirms the need of expanding the coverage of protein interactions using cross-linkers targeting different residues, especially when the distribution of specific amino acids is uneven. In summary, this work further demonstrates the robustness and potential of our XL-MS technology based on sulfoxide-containing MS-cleavable cross-linkers and provides a viable analytical platform for the development of new MS-cleavable cross-linker derivatives to further define protein-protein interactions. The development of these new tools will aid in the goal of understanding the structural dynamics of protein complexes at the global scale in the future.

Chapter 2

Developing a Novel Sulfoxide-containing MS-cleavable Homobifunctional Cysteine Reactive Cross-linker for Studying Protein-Protein Interactions

Among various types of cysteine-reactive reagents, maleimides are most widely utilized due to the specificity and efficiency of thiol-maleimide coupling [89]. In addition, maleimide moieties are also easily functionalized and reactions can occur at physiological pH in the absence of catalysts or heating, making these reagents well-suited for a variety of experimental uses. Such chemistry has been successfully employed in cross-linking studies to probe protein structures and identify interaction domains [89-91]. While homobifunctional maleimide cross-linkers are commercially available, their uses in XL-MS studies remain sparse. This is more likely due to difficulty in MS identification of cross-linked peptides. During the development of lysine and acidic residue-targeting reagents, it has been demonstrated that MS-cleavable cross-linkers are powerful and effective in facilitating accurate identification of cross-linked peptides [10,51,52,56,59,70,131,132]. In recent years, we have successfully developed a suite of sulfoxide-containing, MS-cleavable cross-linkers (i.e. DSSO (Fig. 2.1A) [60], DMDSSO [64], Azide/Alkyne-A-DSBSO [46,49] and DHSO (Fig. 2.1B) [65]) that permit simplified and accurate identification of cross-linked peptides. The MS-labile C-S bonds adjacent to the can be selectively and preferentially fragmented prior to peptide backbone cleavage upon collision-induced dissociation (CID), physically separating the peptides for individual sequencing [46,60,64,65]. Notably, this robust and predictable fragmentation occurs independent of cross-linking chemistry, peptide charge and peptide sequence. These unique characteristics enable straightforward and unambiguous identification of cross-linked peptides by MSn analysis coupled with conventional database searching tools [46,52,56,59,60,64,65,70,131,132]. Sulfoxide-containing MS-cleavable cross-linkers have been successfully

applied to not only study PPIs in vitro [15,60,72-75,133,134] and in vivo [46,75], but also to quantify structural dynamics of protein complexes [64,135]. Given the robustness of sulfoxide-based cleavability, we aimed to design, synthesize and characterize a novel sulfoxide-containing MS-cleavable homobifunctional cysteine linker, namely, BMSO (Bis-maleimide sulfoxide) to facilitate the identification of cysteine cross-linked peptides. To the best of our knowledge, BMSO represents the first generation of cysteine-reactive cross-linking reagents with MS cleavability, which undoubtedly enhances our capability of mapping PPIs in the future.

Experimental Methods for this section can be found in Appendix B.

RESULTS

Development of a Novel Sulfoxide-containing MS-cleavable Cysteine-reactive Cross-linker

In order to improve the identification of cysteine cross-linked peptides, we sought to create a novel MS-cleavable cysteine-reactive homobifunctional cross-linking reagent by integrating the MS-cleavability of sulfoxide-containing cross-linkers [46,60,64,65] with maleimide chemistry. This resulted in the development of BMSO (bismaleimide sulfoxide or 3,3'-sulfinylbis(N-(2-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)ethyl)propanamide) (Fig. 2.1C). This cross-linker consists of two maleimide functional groups for cross-linking cysteines connected by a spacer arm containing a central sulfoxide group with two symmetric MS-cleavable C-S bonds. As shown, BMSO contains the same type of MS-cleavable bonds as DSSO [60] and DHSO [65] (Fig. 2.1A-B). The synthesis of BMSO only requires one additional amide bond forming step using DSSO as the starting material (Figure 2.1D). The core of DSSO is elaborated through the addition of the trifluoroacetate salt of 1-(2-aminoethyl) maleimide to install the cysteine-reactive moiety [136]. BMSO has a spacer arm of 24.2 Å, well within the distance range among cross-linking reagents that have been successfully applied for studying PPIs [10].

Predicted MS^2 Fragmentation of BMSO Cross-linked Peptides

Similar to other residue-specific cross-linkers, BMSO cross-linking is expected to result in three types of cross-linked peptides: dead-end (type 0), intra-link (type 1), and inter-link (type 2) modified peptides [130]. As inter-links (type 2) provide the most useful information pertaining to the relative spatial orientation of cross-linked cysteine residues, we have focused on the characterization of BMSO inter-linked peptides here. Figure 2.1E illustrates the thiol-maleimide coupling reaction between cysteine sulfhydryl (-SH) groups and the maleimide functional groups of BMSO, resulting in a closed-ring **s**uccin**i**midyl **t**hio**e**ther (SITE) bond by way of Michael addition [137]. The succinimidyl moiety of a SITE can then undergo irreversible hydrolysis in aqueous buffer, opening the ring to produce a stable **s**uccinamic **a**cid **t**hio**e**ther (SATE). To distinguish between these two forms, we have designated the closed-ring SITE inter-linked peptide as (α_c-β_c)—in which 'c' represents 'closed-ring SITE'—and the open-ring form of the same BMSO inter-linked peptide as (α_o-β_o), where 'o' represents 'open-ring SITE' (Fig. 2.1E). The mass difference between the fully closed and opened forms of BMSO cross-linked peptides is equivalent to the mass of $2H_2O$ (+32 Da).

Regardless of the thioester structures (SITE or SATE) covalently linked to the cysteine residues, BMSO cross-linked peptides share the same MS-cleavable bonds as previously reported sulfoxide-containing cross-linked peptides and are thus expected to yield the same characteristic fragmentation patterns that enable cross-link identification by MS^n analysis [46,60,64,65]. To illustrate, Figure 2.1F displays the predicted MS^2 fragmentation of a BMSO inter-linked heterodimer (α-β), with either SITE or SATE structures on cross-linked cysteines. Cleavage of either one of the two symmetric C-S bonds results in physical separation of the two cross-linked peptide constituents, resulting in the detection of peptide fragment ion pairs (i.e. α_A/β_S or α_S/β_A). These fragment

peptides are modified either with alkene (A) or sulfenic acid (S) moieties, remnants of BMSO following collision-induced dissociation. As previously noted for other sulfoxide-containing cross-linkers [46,60,64,65], the sulfenic moiety typically undergoes dehydration to become a more stable and dominant unsaturated thiol (T) moiety (Figure 2.2). Therefore, the fragmentation pairs for a BMSO cross-linked peptide α-β are expected to be α_A/β_T and α_T/β_A (Fig. 2.1F). Such characteristic CID-triggered cross-link fragmentation has been proven unique and robust to sulfoxide-containing cross-linkers, independent of cross-linking chemistry, peptide sequence and charge [46,60,64,65]. The resulting MS^2 fragment ions represent single peptide chains that can be subjected to individual MS^3 analyses, permitting unambiguous identification of both cross-linked peptide sequences and cross-linking sites.

Characterization of BMSO Cross-linked Model Peptides by MS^n Analysis

To evaluate BMSO cross-linking and establish an analytical workflow for the identification of BMSO cross-linked peptides, we first performed BMSO cross-linking using a synthetic cysteine-containing peptide Ac-LR9 (Ac-LADVCAHER). Due to this reaction being performed in DMSO, the major inter-linked product detected in MS^1 was an Ac-LR9 homodimer with closed-ring SITE, i.e. (α_c-α_c) (*m/z* 637.7849^{4+}) (Figure 2.3A). MS^2 analysis of this BMSO inter-linked homodimer generated a pair of dominant fragment ions α_{Ac}/α_{Tc} (*m/z* $625.29^{2+}/641.27^{2+}$) as expected for BMSO inter-linked homodimers (Figure 2.3B). Subsequent MS^3 analyses of α_{Ac} (m/z 625.29^{2+}) and α_{Tc} (m/z 641.27^{2+}) fragment ions yielded series of b and y sequencing ions identifying them as Ac-LADVC_{Ac}AHER and Ac-LADVC_{Tc}AHER, respectively (Fig. 2.4A-B), verifying a cysteine-cysteine BMSO linkage between two separate Ac-LR9 peptides.

During LC-MS^n analysis, the fully open-ring Ac-LR9 homodimer cross-linked with BMSO was also detected, i.e. (α_o-α_o) (m/z 646.7885^{4+}) (Figure 2.3C). As expected, its MS^2 analysis

yielded the expected characteristic fragment ion pair α_{Ao}/α_{To} (*m/z* $634.30^{2+}/650.28^{2+}$). MS^3 analyses of these two MS^2 fragments allowed unambiguous peptide identification as Ac-LADVC$_{Ao}$AHER and Ac-LADVC$_{To}$AHER, respectively (Figure 2.5A-B). These results demonstrate that the state of the ring structures attached to cross-linked cysteines do not interfere with the characteristic MS^2 fragmentation of BMSO cross-linked peptides and their subsequent identification by MS^n analysis.

Apart from the fully closed and open-ring forms of BMSO inter-linked Ac-LR9 homodimers, an additional ion (α_c-α_o) (m/z 642.2861^{4+}) was detected, representing a half-hydrolyzed product containing two cross-linked peptides, one with closed-ring and the other with open-ring structures attached to cysteines. Its MS^n analysis further confirms that BMSO cross-linked peptides can be readily identified independent of the ring structures on cross-linked cysteines (Fig. 2.6). Nonetheless, the existence of different forms of the same BMSO cross-linked peptide not only increases sample complexity, but also decreases the abundance of each particular cross-link. Therefore, it would be most desirable to obtain a single form of BMSO cross-linked peptides for MS^n analysis. To this end, we examined several experimental conditions to favor complete SITE hydrolysis and thus generate the most stable form of BMSO cross-linked products– open-ring SATE structures. Incubation of BMSO inter-linked Ac-LR9 in 25 mM ammonium bicarbonate buffer overnight at 37^{o}C led to a nearly complete (98.7%) conversion of the fully closed-ring form (α_c-α_c) to the fully open-ring form (α_o-α_o) (Figure 2.7A-B). This indicates that SITE hydrolysis can be induced in order to minimize the heterogeneity of resulting cross-linked peptides. While BMSO cross-linking of standard peptides was carried out in DMSO, protein cross-linking and digestion are typically performed in physiological buffers. Therefore, we suspect that the majority of BMSO cross-linked products for protein samples may be in open-ring states.

However, similar procedures were carried out in following experiments to ensure homogenous cross-linked products for MS^n analysis.

Identification of BMSO Cross-linked Peptides of BSA

To evaluate BMSO cross-linking of proteins, we used bovine serum albumin (BSA) as our model protein since it has been previously used for characterizing cross-linking studies [65]. Importantly, BSA contains 35 cysteines out of a total of 607 amino acid residues (5.76%), well above the average cysteine content in the proteome (~1.2%) (http://www.uniprot.org). The general procedures for optimizing BMSO cross-linking and MS^n analysis of BMSO cross-linked proteins are very similar to those described for other sulfoxide-containing cross-linkers [46,60,64,65]. Briefly, BMSO cross-linking of BSA was optimized by titrating various ratios of protein to cross-linker concentrations, temperature and reaction time. The resulting cross-linked proteins were separated by SDS-PAGE, digested and subjected to LC-MS^n analysis. Figure 2.8 displays an exemplary MS^n analysis of a BMSO inter-linked peptide of BSA (α_o-β_o) (*m/z* 719.5689^{4+}). MS^2 fragmentation of this inter-linked peptide resulted in the production of two characteristic peptide fragment pairs: α_{Ao}/β_{To} (*m/z* $614.29^{2+}/815.84^{2+}$) and α_{To}/β_{Ao} (*m/z* $630.28^{2+}/799.86^{2+}$) as predicted for BMSO inter-linked heterodimers (Fig. 2.8B). Subsequent MS^3 analysis of α_{Ao} (*m/z* 614.29^{2+}) (Fig. 2.8C) determined its sequence as $SHC_{Ao}IAEVEK$, in which the cysteine in the 3rd position from the N-terminus was modified with an open-ring alkene moiety. MS^3 analysis of β_{To} (m/z 815.84^{2+}) identified its sequence as $YIC_{To}DNQDTISSK$, with the cysteine in the 3rd position from the N-terminus carrying an open-ring unsaturated thiol moiety (Fig. 2.8D). Collectively, the inter-linked peptide was identified as [^{286}YICDNQDTISSK^{298} cross-linked to ^{310}SHCIAEVEK^{318}], describing a fully open-ring inter-link formed between C289 and C312 of BSA.

Similarly, a total of 41 unique BMSO inter-linked peptides of BSA were identified by LC-MS^n analysis, representing 38 unique C-C linkages (Table 2.1). It is noted that nearly all modified peptides comprising BMSO cross-links were identified in the open-ring SATE state–with the exception of CASIQK, which was identified frequently with a closed-ring SITE. We suspect that the free amine group of the N-terminal cysteine may react with the 5-member SITE ring to form a more stable 6-member ring without any mass change, thereby preventing SITE hydrolysis. These results demonstrate that BMSO cross-linking is effective and that the heterogeneity of cross-linked products can be controlled (in other words, SITE hydrolysis can be stabilized by conversion to SATE products). More importantly, these results prove that BMSO cross-linked peptides exhibit the characteristic MS^n fragmentation patterns as expected for all sulfoxide-containing MS-cleavable cross-linked peptides [46,60,64,65], thus enabling their simplified and accurate identification using the same MS^n workflow as previously established for sulfoxide-containing cross-linkers.

BMSO Cross-linking Maps of BSA

To examine the efficacy and interaction coverage of BMSO cross-linking on our model protein, we first derived a 2-D cross-linking map using the unique C-C linkages identified (Fig. 2.9A). Considering the spacer arm length of BMSO (24.2 Å) and the distances contributed by cysteine side chains (2.8 Å), as well as backbone flexibility and structural dynamics, we estimated that the theoretical upper limit for the Cα-Cα distances between BMSO cross-linked cysteine residues is ~45 Å. To determine whether the identified BMSO cross-links correlate to residues with distances below the theoretical limit, we mapped them onto a published BSA crystal structure (PDB: 4F5S). All identified cross-links were able to be mapped to the crystal structure (Fig. 2.9B), with 97.4% (37 out of 38) having measurable Cα-Cα distances below 45 Å (Figure 2.9C, Table 2.1). This indicates that nearly all cross-links satisfy the expected distance constraints permitted

by the molecular structure of BMSO and that the captured cross-links correlate well with known BSA structure. In summary, our results suggest that BMSO cross-linking is effective for mapping protein-protein interactions.

Comparison of BMSO with DSSO and DHSO Cross-linking

To determine the complementarity of cysteine-reactive cross-linking with our previously developed amine-reactive and acidic residue-reactive cross-linkers, we have compared the cross-links identified using BMSO in this study to those previously reported using DSSO and DHSO (Fig. 2.10A-C) [65]. As aspartic/glutamic acid residues are most abundant in BSA, it is not surprising that acidic-residue cross-linking yielded the highest number of cross-links overall (69). Interestingly, although BSA has more lysines (9.8%) than cysteines (5.7%), we have identified more C-C linkages than K-K linkages (i.e. 43 vs. 33, respectively). Compared to XL-MS maps derived from DSSO and DHSO cross-linking data [65], BMSO cross-linking improves the overall coverage by identifying proximal regions unfavored by amine- and acidic residue-targeting cross-linkers, thereby complementing previous results. This observation is most evident when examining the spatial relationships of the centrally located helices (H4, H13, H17, and H22) relative to more peripheral helices in the 3-D structure of BSA (PDB: 4F5S). These regions poorly covered by DSSO and DHSO are better characterized by BMSO, which has identified a total of 17 C-C-linkages describing clusters of physical contacts between: 1) H4 with H10 (C99-C200, C99-C223) and H17 (C99-C312); 2) H13 with H7 (C288-C147), H10 (C288-C200, C288-C223), H22 (C288-C392), H26 (C288-C471, C288-C484), and H28 (C288-C510); 3) H17 with H6 (C312-C125); and 4) H22 with H10 (C392-C200), H21 (C392-C383, C392-C384), H26 (C392-C471; C392-C484), and H28 (C392-C510). The number of cysteines and their positions within the core as well as the length of BMSO are likely contributing factors that enable the detection of cross-links within these

regions. A large portion of identified cysteine cross-links within these clusters correspond to residues with Cα-Cα distances over 30 Å (11/17, 64.7%), which are more likely to be missed with cross-linkers containing shorter spacer arms (i.e. DSSO and DHSO). While BMSO appears to obtain more contacts in the center regions of BSA, DSSO and DHSO cross-linking have provided broader coverage of various regions, including both termini. This observation may also be correlated to the relative distribution of cysteines within the primary sequence and their roles in stabilizing the structure of BSA. While BSA is a single protein, our data suggests that the three distinct cross-linking chemistries can indeed facilitate a more comprehensive mapping of intramolecular BSA contacts. Therefore, we anticipate that such a combinatory XL-MS approach would be even more beneficial for detailed PPI profiling when applied to more complex samples such as large multiprotein assemblies and cell lysates.

DISCUSSION

Here we report the development and characterization of a novel sulfoxide-containing MS-cleavable cysteine reactive cross-linker, BMSO, derived from DSSO [60]. Using both a standard peptide and protein, we have demonstrated that BMSO cross-linking is efficient and that BMSO cross-linked peptides display the same characteristic MS-cleavability unique to other sulfoxide-containing cross-linked peptides [46,60,64,65]. Our results further illustrate that the same MS^n workflow can be applied for simplified and accurate identification of all sulfoxide-containing MS-cleavable cross-linked peptides regardless of cross-linking chemistries, thus facilitating data comparison and validation. However, unlike NHS esters (amine reactive) and dihydrazides (acidic residue reactive), maleimide chemistry yields two different forms of the cross-linked cysteines with either the SITE (closed-ring) structure or its hydrolyzed form, SATE (open-ring). Nevertheless, the irreversible hydrolysis of SITE can be pushed close to completion, minimizing sample

heterogeneity. Importantly, the ring states of BMSO cross-linked cysteines do not interfere with their identification by MS^n analysis. To the best of our knowledge, BMSO represents the first MS-cleavable cross-linking reagent, and we expect that it will significantly advance cross-linking studies targeting cysteine residues—which are currently underrepresented in XL-MS analysis. This comparison showing the complementary nature of BMSO, DSSO and DHSO cross-linking data further signifies the necessity and usefulness of multiple cross-linking chemistries to obtain high-density interaction maps with improved confidence, which will undoubtedly expand our capacity and capability of mapping PPIs at the systems-level in the future.

Chapter 3

Enabling Photoactivated Cross-linking Mass Spectrometric Analysis of Protein Complexes by Novel MS-cleavable Cross-linkers

MS-cleavable cross-linking reagents have significantly facilitated MS analysis of cross-linked peptides in complex mixtures, due to their unique capability of eliminating the "n-square" problem and permitting effective sequencing of cross-linked peptides [10,51]. To enable robust MS-cleavability, we have previously developed a series of sulfoxide-containing MS-cleavable cross-linking reagents (e.g. disuccinimidyl sulfoxide (DSSO) (Fig. 3.1A)) [46,49,60,64-66]. The MS-labile C-S bonds adjacent to the sulfoxide can be preferentially fragmented prior to peptide backbone cleavage upon collision-induced dissociation (CID), physically separating the two cross-linked peptide constituents for individual sequencing. Notably, this predictable fragmentation occurs independent of cross-linking chemistry, peptide charge and peptide sequence. These unique characteristics allow straightforward and unambiguous identification of cross-linked peptides by MS^n analysis coupled with conventional database searching tools. Sulfoxide-containing MS-cleavable cross-linkers have been successfully applied to not only study PPIs *in vitro* [60,72-74] and *in vivo* [46,75], but also to dissect structural dynamics of protein complexes [76,78,135]. Thus, to expedite the identification of photocross-linked peptides, we have developed three sulfoxide-containing MS-cleavable heterobifunctional NHS-diazirine cross-linkers with varied lengths, namely, SDASO (Succinimidyl diazirine sulfoxide)-L (long), -M (medium) and -S (short). These SDASO reagents represent the first class of MS-cleavable heterobifunctional photo-reactive cross-linkers. To illustrate their capabilities, we have characterized SDASO cross-linkers with a standard protein BSA, and applied them to map PPIs of affinity-purified yeast 26S proteasome. Our results demonstrate that MS-cleavability enables accurate identification of photocross-linked peptides and

that the SDASO-based XL-MS workflow is well-suited for probing PPIs in complex samples. In addition, comparison with residue-specific XL-MS data has determined that SDASO cross-linking is robust and captures PPIs complementary to existing reagents.

Experimental Methods for this section can be found in Appendix C.

RESULTS

Designs of MS-cleavable NHS-Diazirine Heterobifunctional Cross-linkers

In order to advance photoreactive cross-linkers for complex PPI mapping, we sought to create novel sulfoxide-containing MS-cleavable NHS-diazirine heterobifunctional cross-linking reagents to cross-link lysines to any nearby amino acids (AAs). It is noted that all of our previous sulfoxide-containing MS-cleavable cross-linkers are homobifunctional and carry two symmetric MS-cleavable C-S bonds adjacent to the central sulfoxide (Fig. 3.1 A, E) [46,60,64-66]. Due to the structural differences in reactive groups and their targeted residues, this symmetry is not retained in heterobifunctional cross-linkers. Recently, we have explored effects of spacer arm structures on MS-cleavability of sulfoxide-containing cross-linkers and identified an asymmetric spacer arm structure [138]} that maintains the characteristic and predictable fragmentation expected of symmetric sulfoxide-containing MS-cleavable cross-linkers [46,49,60,64-66]. This unique asymmetric spacer arm region carries a sulfoxide group that divides the spacer arm into two halves, i.e. a fixed half identical to DSSO with the sulfoxide and carbonyl group separated by '3' bond lengths, and a flexible half. Based on this design, we constructed three MS-cleavable heterobifunctional SDASO cross-linkers composed of a fixed NHS ester end and a flexible diazirine side with varying lengths from the center sulfoxide (i.e. long, 12.5 Å; medium, 10.2 Å; short, 7.7 Å), well within the distance range suited for studying PPIs[10] (Fig. 3.1B-D). The synthesis routes and chemical analyses of SDASOs were detailed here (Fig. 3.2 and Appendix C).

Fragmentation Characteristics of SDASO Cross-linked Peptides

Based on our recent studies on asymmetric sulfoxide-containing cross-linkers [139], only the C-S bond at the NHS ester end in SDASO should be preferentially cleaved during CID. Thus, a single pair of MS^2 fragment ions is expected for all three SDASO cross-linkers (Fig. 3.1F). For an SDASO inter-linked peptide (α-β), cleavage during CID physically separates the two cross-linked constituents and thus leads to the detection of two characteristic fragment ions (α_A/β_S) carrying remnants of SDASO. The α_A fragment contains a cross-linked lysine modified with the alkene (A) moiety, whereas the β_S fragment contains a photocross-linked amino acid modified with a sulfenic acid (S) moiety. Since the NHS ester side of all three SDASO reagents are identical to half of DSSO, the expected alkene moieties are the same as seen in DSSO cross-linked peptides (Fig. 3.1E). In contrast, the three SDASO cross-linkers yield three different sulfenic acid moieties due to spacer arm differences in the diazirine end (Fig. 3.1F). As previously noted for other sulfoxide-containing cross-linkers [46,60,64-66,139], the sulfenic acid moiety typically undergoes dehydration to become a more stable and dominant unsaturated thiol (T) moiety, leading to the detection of β_T (Fig. 3.3A). To examine whether SDASO cross-linked peptides produce the expected fragmentation, standard protein BSA was cross-linked by the three SDASO cross-linkers separately, and the resulting peptide digests were analyzed by LC MS^n. As illustrated (Fig. 3.4), each MS^n analysis of the same BSA peptides inter-linked by the three SDASO reagents yielded a dominant MS^2 fragment pair (α_A/β_T) as predicted. These resultant MS^2 fragment ions representing single peptide chains were then subjected to individual MS^3 analyses, permitting unambiguous identification of both cross-linked peptide sequences and cross-linking sites. As a result, the respective cross-links between BSA:K155 and BSA:E41 were identified for all SDASO linkers.

Similar to residue-specific cross-linkers, SDASO cross-linking can also result in dead-end and intra-linked peptides. For SDASO cross-linkers, two types of dead-end peptides are expected as both reactive ends can be hydrolyzed (Fig. 3.3B-C). For NHS ester dead-ends, the resulting fragment ions would carry thiol moieties (Fig. 3.3B), whereas the MS^2 fragment ion of diazirine dead-end peptides would be labeled with an alkene moiety (Fig. 3.3C). These predicted MS^2 fragmentations were demonstrated by respective SDASO dead-end peptides of BSA (Fig 3.5). Similarly, for SDASO intra-linked peptides, a single fragment would be expected, containing both an alkene and thiol modification (Fig. 3.3D). Exemplary MS^n spectra of the three SDASO intra-linked peptides of BSA further demonstrated the anticipated fragmentation (Fig 3.6).

Collectively, the three types of SDASO cross-linked peptides fragment as predicted during CID to generate characteristic and predictable MS^2 products, which enable their simplified and accurate identification by MS^n analysis in the same way as other sulfoxide-containing cross-linked peptides [46,60,64,138,140,141].

SDASO XL-MS Analysis of BSA

To evaluate the performance of the three SDASO cross-linkers, we first carried out XL-MS analyses of BSA with three biological replicates each. Based on the general workflow (Fig. 3.7), LC MS^n analyses resulted a total of 453 unique SDASO-L, 306 SDASO-M, 254 SDASO-S inter-linked BSA peptides, encompassing 427, 338, 306 unique K-X linkages, respectively (Tables 3.1-3.3). Here X represents any of the 20 common amino acids (AAs). Although the three SDASO cross-linkers produced similar amounts of XL-MS data, SDASO-S consistently generated the least number of cross-linked peptides (Fig. 3.8A). This is not entirely surprising as short linkers are more stringent on distance constraints between two cross-linkable residues, as seen in residue-specific linkers [8,77]. Overall, 37% of cross-linked peptide sequences and 29% of their

corresponding K-X linkages of BSA were found in common among the three SDASO cross-linkers (Fig. 3.8A-B). Our results indicate that the three SDASO linkers have similar efficiency in cross-linking BSA and mapped a considerable number of shared regions, but also yielded unique cross-linked peptides and sites.

To explore the interaction coverage of BSA by SDASO cross-linking, we derived both 2-D and 3-D XL-maps based on the identified K-X linkages (Fig. 3.8C-D). In comparison with our published XL-MS data of BSA using DSSO (amine-reactive), DHSO (acidic residue-reactive), and BMSO (cysteine-reactive) cross-linkers (Fig. 3.9A-D & Fig. 3.8E) [65,66], SDASO XL-MS resulted in the highest number of cross-linked peptides and contact sites, thus enabling the generation of the most extensive interaction coverages. As shown, interactions within the central core of BSA are broadly mapped by all types of linkers, while interactions at the N and C-termini of BSA are best profiled by the SDASO linkers (Fig. 3.8C-E & Fig. 3.9A-D). These results demonstrate that SDASO cross-linking is effective for mapping interactions of single proteins and generates structural information complementary to residue-specific cross-linkers.

Evaluation of SDASO Cross-links of BSA

Among the 20 common AAs that can be targeted by diazirine, arginine has the longest side-chain. Considering the spacer arm lengths of SDASOs (i.e. SDASO-L (12.5 Å), SDASO-M (10.2 Å), SDASO-S (7.7 Å)), side-chain lengths of lysine (6.3 Å) and arginine (7.1 Å), as well as backbone flexibility and structural dynamics, the theoretical upper limits for the Cα-Cα distances of SDASO cross-links between a lysine (NHS reactive end) and any AA (diazirine reactive end) would be ≤ 35 Å for SDASO-L and ≤ 30 Å for SDASO-M and SDASO-S. To validate the SDASO cross-links of BSA, we mapped all of the identified cross-links onto the crystal structure of BSA (PDB: 4F5S) (Fig. 3.8F). As a result, 100% of SDASO-L, 99% of SDASO-M, and 100% of

SDASO-S linkages were satisfied with most Cα-Cα distances well below their respective maximum thresholds (Fig. 3.8F, Table 3.3), supporting the validity of the SDASO cross-links. Notably, the average distances of SDASO cross-links also corresponded well with the linker lengths:15.8±5.8Å (SDASO-L), 15.1±5.8Å (SDASO-M), and 14.1±4.8 Å (SDASO-S). Although the spacer arm lengths are comparable, SDASO cross-links displayed higher satisfaction rates and lower average distances than those of DSSO and DHSO cross-links of BSA [65]. This may be due to the fact that amino acids other than arginine would result in distances less than the expected upper limits [101].

Because of the non-specificity, it is suspected that photoactivated reaction would lead to increased variance in cross-linked products compared to residue-specific cross-linkers. To test this, we first compared the sequences of identified SDASO cross-linked peptides of BSA for each linker without considering their site localization. Interestingly, all three linkers displayed similar reproducibility with considerably high overlaps (~64%) among their corresponding three biological replicates (Fig. 3.10A-C). When examining residue-to-residue (i.e. K-X) linkages, all three linkers also exhibited good reproducibility, with overlaps of 50% for SDASO-L, 42% for SDASO-M, and 37% for SDASO-S among their three respective biological replicates (Fig. 3.10D-F). Intriguingly, the observed residue-to-residue reproducibility of the three SDASO linkers is also quite comparable with cross-linkers with specific chemistries (i.e. DSSO, DHSO and BMSO) [65,66], indicating the robustness and reliability of SDASO cross-linking.

SDASO-based XL-MS Analysis of the Yeast 26S Proteasome Complex

To access the feasibility of photoactivated cross-linking for complex PPI mapping, we performed SDASO XL-MS analyses of affinity-purified yeast 26S proteasome complex. This 33-subunit protein degradation machine consists of two subcomplexes, the 19S regulatory particle

(RP) and 20S core particle (CP) [142]. The 19S RP contains 19 subunits that are assembled into the lid (i.e. Rpn3, Rpn5-9, Rpn11, Rpn12, Rpn15/Sem1) and base (Rpt1-6, Rpn1-2, Rpn10, Rpn13) subcomplexes, while the 20S CP is composed of 14 subunits (α1-7, β1-7) that form four stacked 7-member ring structures in the order of αββα. With three biological replicates for each linker, LC MS^n analyses of tryptic digests of SDASO cross-linked complexes resulted in the identification of 955 SDASO-L, 925 SDASO-M and 741 SDASO-S unique cross-linked peptides within the 26S proteasome (Tables 3.5-3.6), representing 1382 SDASO-L (725 inter-subunit and 657 intra-subunit), 1136 SDASO-M (648 inter-subunit and 488 intra-subunit) and 1001 SDASO-S (477 inter-subunit and 524 intra-subunit) unique K-X linkages (Table 3.8). As a result, 43% of SDASO-L, 52% of SDASO-M, and 60% of SDASO-S cross-linked peptide sequences (Fig. 3.11A-C), as well as 29% of SDASO-L, 37% of SDASO-M, and 38% of SDASO-S K-X linkages were found reproducible among their respective biological replicates (Fig. 3.11D-F), comparable to BSA data. These results further support the robustness of SDASO cross-linking. When comparing XL-MS data among the three linkers, we found that the number of SDASO cross-links of proteasomes increased with spacer arm lengths of the linkers, similar to BSA data. However, the resulting cross-link data among the three linkers shared considerably fewer in common for proteasomes than for BSA, with overlaps of 16% vs. 37% for cross-linked peptide sequences and of 11% vs. 29% for K-X linkages (Fig. 3.8A-B, 3.12A, & Fig. 3.13).These results suggest that spacer arm lengths of SDASO linkers play a more significant role in capturing interactions within protein complexes, most likely attributed to the presence of both inter-protein and intra-protein interactions. Thus, the use of the three SDASO linkers is beneficial not only for result cross-validation, but also for comprehensive PPI mapping of protein complexes.

As additional enzymatic digestions are known to increase sequence coverage in XL-MS analyses using residue-specific cross-linkers [43], we expected that similar results would be obtained for SDASO linkers. To test this, we performed chymotrypsin digestion of SDASO-L cross-linked proteasomes with three biological replicates. LC MS^n analyses of chymotryptic digests resulted in the identification of a total of 637 unique SDASO-L cross-linked peptides of the 26S proteasome (Tables 3.5 & 3.6), representing 1115 SDASO-L unique K-X linkages, comparable to the trypsin XL-MS data as described above (Table 3.8). While the reproducibility of XL-MS data was somewhat similar for both chymotryptic and tryptic digests of SDASO-L cross-linked proteasomes (Fig. 3.14A-B, 3.11A, 3.11D), their overlaps of cross-linked peptide sequences and K-X linkages were quite limited (~10%) (Fig. 3.14C-D). This confirms that additional enzymatic digestion could facilitate the expansion of PPI coverages. Thus, tryptic and chymotrypic datasets of SDASO-L were combined, yielding a total of 1711 unique SDASO-L K-X linkages for subsequent analyses (Table 3.8).

Validation of Proteasome Cross-links by Structural Mapping

It is known that the 26S proteasome is a dynamic entity and possesses multiple conformational states to fulfill its function [142,143]. To validate SDASO cross-links, we mapped the identified K-X linkages onto the four known structures of the yeast 26S proteasome that represent its progression through an ATP-driven functional cycle: s1 (PDB:4CR2), s2 (PDB:4CR3), s3 (PDB:4CR4) and s4 (PDB:5MPC) [144,145]. As a result, the average distance satisfaction rates of the identified K-X linkages across the four models for each linker were found to be very similar: 91% for SDASO-L (≤ 35 Å), 91% for SDASO-M (≤30 Å), 87 % for SDASO-S (≤30 Å), with an overall variation less than 1% (Fig. 3.12B, Fig. 3.15A-D). Additionally, we noticed a group of SDASO linkages that appeared to fit better with a subset of models (Table 3.9), suggesting the presence of

conformational heterogeneity in the sample. To examine this, we classified a total of 159 SDASO cross-links as structural state-specific, because they were satisfied only by one, two or three out of the four models. We then grouped these differentially satisfied cross-links into 14 state-specific combinations to infer the presence of preferred structural states. As illustrated in Fig. 3.12C, among all combinations, two major categories were detected for the three SDASO linkers, representing 82% of the total state-specific SDASO cross-links. One of them contained cross-links (54%) satisfied only by s1-s3 states but not by the s4 state, implying the presence of s1, s2 and/or s3 states in the purified proteasome. The other described cross-links (~28%) satisfied only by the s4 state, indicating presence of that state. These two groups of state-specific cross-links represent 28 protein interactions, half of which describe connectivity within the 20S CP. The remaining half embody interactions within the 19S, particularly concerning Rpn11 and Rpn1. The results correlate well with the fact that these regions are expected to undergo significant conformational changes during state conversions of the 26S proteasome [144,145].

When considering inter-subunit and intra-subunit cross-links separately, the latter has a slightly higher distance satisfaction when mapped to known structures (intra-subunit: SDASO-L: 98%, SDASO-M: 96%, and SDASO-S: 88% vs. inter-subunit: SDASO-L: 79%, SDASO-M: 86%, and SDASO-S: 86%) (Fig 3.16A-M). This is expected as inter-subunit interactions are typically more dynamic. Coincidentally, the majority of non-satisfied inter-subunit linkages also localized to the 19S RP (Fig. 3.17A-M), which is known to have diverse conformations [143]. Collectively, structural mapping supports the validity of the identified SDASO cross-links and suggests the existence of multiple states in our purified proteasome.

Comparison of SDASO XL-Maps of the 26S Proteasome

To further evaluate the performance of SDASO in complex PPI mapping, we generated 2-D XL-maps of the 26S proteasome based on unique K-X linkages identified by each SDASO linker (Fig. 4D). A total of 135 non-redundant PPIs (103 inter-subunit and 32 intra-subunit) within the 26S proteasome were determined based on 2487 K-X linkages identified by the three SDASO linkers, including 119 from SDASO-L (79 inter-subunit and 30 intra-subunit), 81 from SDASO-M (53 inter-subunit and 28 intra-subunit), and 61 from SDASO-S (32 inter-subunit and 29 intra-subunit) (Table 3.11). While ~20% of inter-subunit interactions were identified across all three linkers (21/103), each linker contributed unique interactions (SDASO-L 42/103, SDASO-M 16/103 and SDASO-S 5/103). The inter-subunit interactions of the 26S proteasome captured by each linker can be classified into 3 categories based on proteasome subcomplexes: 19S-19S (56 SDASO-L, 30 SDASO-M, and 14 SDASO-S), 19S-20S (7 SDASO-L, 7 SDASO-M, and 2 SDASO-S) and 20S-20S (16 each for SDASO-L, -M, and -S), as illustrated in Fig. 3.12E. The differences in the PPIs captured by SDASO linkers are most likely related to their spacer arm lengths. Nevertheless, these results indicate that SDASO cross-linking covers a diverse range of protein interactions and that each SDASO linker contributes to mapping the comprehensive interaction network within the 26S proteasome.

DSSO XL-MS Analysis of the 26S Proteasome

To better assess SDASO cross-link data, we performed a set of XL-MS experiments on the yeast 26S proteasome using DSSO for comparison. LC MS^n analyses identified a total of 1204 unique DSSO cross-linked peptides of proteasomes from two biological replicates, representing 1567 K-K linkages (748 inter-subunit and 819 intra-subunit) and describing 107 inter-subunit and 30 intra-subunit interactions (Tables 3.7 & 3.11). While the overlap (65%) of DSSO cross-linked peptide sequences between the two biological replicates was comparable to those of SDASO data

(57%~70%) (Supplementary Fig. 3.18A), the reproducibility of DSSO residue-to-residue (i.e. K-K) linkages was higher (~65%) (Supplementary Fig. 3.18B) than those of SDASO data (43%~52%). The increased variation in identified SDASO cross-link sites is expected as nonspecific cross-linking chemistry is inherently more variable. Nonetheless, these comparisons further demonstrate that SDASO cross-linking is robust on targetable interaction regions.

Next, we mapped DSSO cross-links onto the four conformational states (s1-s4) of the yeast 26S proteasome [144,145] and determined that on average ~75% of DSSO K-K linkages were satisfied (≤ 30Å) across all four models (Fig. 3.19A-B). Interestingly, a total of 114 DSSO cross-links were also found to be state-specific cross-links, as described above. However, the distribution of cross-links across the 14 state-specific combinations was somewhat different from SDASO data (Fig. 3.19C and Fig. 3.12C). In addition to the notable representations of s4 (30%) and s1-s3 states (15%) as seen in SDASO data, respective state-specific DSSO cross-links satisfied only by s1 state (~9%), s3 state (~8%), and s2-s3-s4 states (~14%) were markedly detected. These DSSO state-specific cross-links further support the presence of multiple conformational states of the 26S proteasome. Similar to SDASO data, intra-subunit DSSO cross-links were much better satisfied than inter-subunit linkages for all four models (intra: 89% vs. inter: 62%) (Fig. 3.19D, F-I), and most of the violating DSSO inter-subunit cross-links were attributed to the 19S RP (Fig. 3.19E, J-M). Taken together, DSSO XL-MS data corroborate well with SDASO results, confirming the structural heterogeneity of affinity-purified 26S proteasome and the dynamic nature of the 19S RP.

Comparison of SDASO and DSSO Cross-linking of Proteasomes

To delineate the interactions captured by residue-specific and nonspecific cross-linkers, we took the cross-links identified in at least two biological replicates from all of our XL-MS experiments and combined the XL data from the three SDASO linkers for further comparison. As

a result, we obtained a total of 2186 SDASO cross-links (959 inter-subunit, 1227 intra-subunit) and 1098 DSSO cross-links (649 inter-subunit, 449 intra-subunit) of the 26S proteasome (Table 3.11). From this data, 2-D and 3-D XL-maps were generated, displaying extensive connectivity among proteasome subunits (Fig 3.20A-C). In comparison, the most noticeable differences in the XL-maps are the increased density of inter-subunit cross-links within the 19S by DSSO (Fig 3.20A-B) and within the 20S by SDASO (Fig. 3.20A, C). When combined, SDASO and DSSO cross-links yielded a total of 118 inter-subunit (78 SDASO and 98 DSSO) and 33 intra-subunit (31 SDASO and 30 DSSO) protein-protein interactions of the 26S proteasome (Fig. 3.21A). While 85 interactions were shared by both types of linkers, 23 interactions were unique to SDASO (20 inter-subunit and 3 intra-subunit) and 43 interactions were only mapped by DSSO (42 inter-subunit and 1 intra-subunit). For the inter-subunit interactions, 47% were identified by both SDASO and DSSO, whereas 17% and 36% were revealed uniquely by SDASO and DSSO, respectively. For further examination, three types of interactions within the 26S proteasome were categorized: 19S-19S, 19S-20S and 20S-20S (Fig. 3.21B), and the 19S-containing interactions were further subdivided into 19S lid-lid, lid-base, base-base, 19S lid-20S and 19S base-20S interactions (Fig. 3.21C). In addition, SDASO and DSSO XL-PPI networks of the 26S proteasome were derived based on their respective cross-links (Fig. 3.21D-E). While both SDASO and DSSO uncovered mostly 19S-containing interactions, DSSO is more efficient than SDASO at defining these interactions i.e., 19S base-base (26 vs. 17), 19S lid-base (21 vs. 15), and 19S-20S (16 vs. 3) (Fig. 3.21C-E). In contrast, SDASO captured more 20S-20S interactions with significantly more cross-link coverage (20 SDASO XL-PPIs from 262 K-X linkages vs. 15 DSSO XL-PPIs from 37 K-K linkages) (Fig. 3.20C and Fig. 3.21B-E). Apart from inter-subunit interactions, 29 out of 33 intra-subunit interactions were captured by both types of linkers (Fig. 3.21A), but SDASO identified a

greater amount of contact sites within proteasome subunits relative to DSSO (Fig. 3.22A-B, Table 3.11).

In comparison, the 2-D XL-maps of the 20S CP demonstrate that SDASO provided broader coverage of both inter-subunit and intra-subunit interactions (Fig. 3.23A-B). For example, one of the inter-subunit interactions uniquely identified by SDASO was between subunit α4 and α5, described by 3 contact regions (α5:K66-α4:X(147-162), α4:K88-α5:X(128-131), and α4:K182-α5:X(233-234)) (Fig. 3.23C). Although there are several lysine residues in α4 (K146, K169, and K177) and α5 (K32 and K52) proximal to the interfaces identified by SDASO, DSSO was not able to capture this particular interaction (Fig. 3.23D). While both SDASO and DSSO identified intra-subunit interactions of α4 and α5 that were complementary, SDASO yielded denser connectivity within each protein (Fig. 3.23C, D). This type of observation is further exemplified by the inter-subunit and intra-subunit interactions of α1 and α2 (Fig. 3.23E, F). As shown, SDASO not only identified the same interaction regions as DSSO (i.e. SDASO: α1:X(98-120)-α2:K91 and α1:X(157-168)-α2:K50 vs. DSSO: α1:K107-α2:K91, α1:K167-α2:K50 and α1:K187-α2:K50), but also determined additional contacts (SDASO: α1:X(11-13)-α2:K17, α1:X(120-123)-α2:K98, α1:X(159-164)-α2:K166, and α1:X(161-166)-α2:K237). Taken together, these results indicate that SDASO is complementary to DSSO in mapping PPIs of protein complexes. In addition, SDASO appears to be better for mapping close contacts and stable interactions, whereas DSSO is better-suited for capturing dynamic interactions.

Identification of Proteasome Interacting Proteins

Besides interactions within the 26S proteasome, we also examined physical contacts with co-purified proteasome-interacting proteins (PIPs). Similarly considering only cross-links that were identified in at least two biological experiments from all of our XL-MS experiments, we

obtained a total of 125 unique SDASO cross-linked peptides (175 K-X linkages) and 90 unique DSSO cross-linked peptides (90 K-K linkages), representing 44 inter-protein and 4 intra-protein pair-wise interactions. This resulted in the identification of 24 PIPs (21 SDASO and 7 DSSO) with direct contacts to the 26S proteasome, including 22 known (https://thebiogrid.org/) and 2 novel ones (Fig. 3.21D-E), in which only 4 PIPs (Ecm29, Ubp6, Fzo1, and Rlf2) were found by both types of linkers (Fig. 3.21A, Table 3.11). The 4 shared PIPs were identified with a total of 17 PPIs, of which only 3 (Rpt2-Ubp6, Rpt3-Ecm29 and Rpn2-Rlf2) were captured by both SDASO and DSSO. Among the known PIPs, Ecm29 is a key regulator of the 26S proteasome, and human Ecm29 has been shown to interact with Rpt1, Rpt4, Rpt5, Rpn1 and Rpn10 by DSSO cross-linking [74]. Similarly, the interactions of yeast Ecm29 with Rpt1, Rpt4, Rpt5, and Rpn1 were confirmed by DSSO XL. In addition, Ecm29-Rpt3 and Ecm29-Rpn6 interactions from DSSO were identified for the first time. Furthermore, SDASO validated Ecm29-Rpt3 interaction and identified Ecm29-Rpt6 interaction (Fig. 3.24A). These results demonstrate extensive contacts between Ecm29 and the 26S proteasome, corroborating well with previous observation of its human orthologue [74]. Ubp6 is a proteasome-associated deubiquitinase that interacts with the 26S proteasome through Rpn1 [146,147]. While DSSO caught Ubp6-Rpt1 and Ubp6-Rpt2 interactions as reported [147], SDASO identified extensive interactions of Ubp6 with multiple subunits including Rpn1, Rpn2, Rpn8 and Rpt2 (Fig. 3.24B). Overall, SDASO XL-MS analyses identified higher number of PIPs than DSSO, illustrating its capability of capturing interacting proteins in affinity purified samples.

Relative Specificity of Diazirine Cross-linking

Although photoactivated diazirine chemistry can target any X-H bond, recent studies have suggested that the reaction shows preferences for acidic residues [99,102]. To examine this, we sought to determine whether any AA preference was observed in SDASO cross-linking of the 26S

proteasome. On average, ~26% of residues cross-linked by SDASO linkers were determined precisely at a single site, whereas the rest were localized at one out of two (~ 34%), three (~20%), or four and more (~20%) possible sites (Fig. 3.25A). Similar precisions in SDASO cross-linked site localization was also observed in BSA data (Fig. 3.25B), consistent with conventional diazirine linkers [101]. To prevent over-estimation due to site ambiguity, we calculated the weighted AA occurrence to assess the preference of diazirine labeling in the 26S proteasome based on localization certainty (Supplementary Fig. 3.25A, Table 3.10). Thus, among the 20 common AAs, glutamic acid clearly was the most favored by diazirine cross-linking, representing ~30% of the targeted residues for all three SDASO linkers (Fig. 3.26A). In comparison, four additional residues, i.e. alanine (7.2%), aspartic acid (6.8%), leucine (7.3%), and tyrosine (6.4%) were targeted relatively favorably by SDASOs, as they had an average frequency well above those of the remaining AAs (2.7%). The dominant preference of glutamic acid displayed by diazirine cross-linking in proteasome samples was also detected in BSA, in which ~25% of SDASO cross-linked sites were glutamic acids (Fig. 3.26B, Table 3.4). Interestingly, five relatively favorable diazirine cross-linked sites in BSA contained aspartic acid, histidine, threonine, valine, and tyrosine with an average frequency of 6.8~8.4%, in which only aspartic acid and tyrosine residues showed similar preference in proteasome samples. This discrepancy is more likely attributed to the occurrence of common AAs in close proximity to cross-linkable lysines at interaction interfaces within proteins of interest as well as MS detectability and identification of the resulting cross-linked peptides. Nonetheless, while diazirine reactivity is nonspecific, our results suggest that it preferably targets a subset of AAs with glutamic acid as its most favorite.

DISCUSSION

Here we report the development and characterization of three sulfoxide-containing MS-cleavable heterobifunctional photoactivated cross-linkers, SDASO-L, -M and -S. While built upon our previously developed amine-reactive DSSO [60], SDASO cross-linkers are distinctly different, representing the first generation of sulfoxide-containing MS-cleavable heterobifunctional cross-linkers. The unique designs of the SDASO linkers enable a single labile bond to be preferentially cleaved over peptide backbone, leading to only one pair of MS^2 fragment ions and enhancing analysis sensitivity [139]. Importantly, SDASO cross-linked peptides possess robust and predictable MS^2 fragmentation characteristics similar to sulfoxide-containing homobifunctional cross-linkers, thus permitting their fast and accurate identification using MS^n-based XL-MS workflow [46,60,64-66]. Although MS^2-based approaches have been widely used in XL-MS studies [10], it is important to note that MS^n analysis is critical for effective database searching to identify photocross-linked peptides and localize nonspecific cross-linked sites with speed and accuracy, especially for complex samples. Owing to their unique capabilities, the SDASO cross-linkers have been successfully employed to study PPIs of not only a single protein BSA, but also the affinity purified yeast 26S proteasome complex. To the best of our knowledge, this work represents the first application of photoactivated cross-linking on PPI mapping of large protein assemblies. The development of SDASO cross-linkers further demonstrates the robustness and potential of our XL-MS technology based on sulfoxide-containing MS-cleavable cross-linkers and provides a viable analytical platform for the expansion of new MS-cleavable reagents to generate a complete PPI map of cellular systems in the future.

Although photo-induced diazirine labeling is nonspecific, the observed reproducibility of cross-linked peptide sequences was comparable for SDASO and residue-specific cross-linkers [76], supporting the reliability of photoactivated cross-linked products. While all of the 20 common

AAs were detected as SDASO cross-linked sites in this work, SDASO displays preferential labeling of glutamic acids, corroborating well with previous reports on diazirine favoring acidic residues [99,102]. Although aspartic acids are in comparable abundance to glutamic acids in BSA and proteasomes, they were targeted noticeably less by SDASO. In comparison, acidic residue-reactive cross-linkers such as DHSO do not appear to have noticeable differences in reactivity towards these two AAs [65,76]. Therefore, the preferential labeling of glutamic acids over aspartic acids displayed by diazirine may be due to differences in physiochemical properties of their side-chains and short-lived photoactivated reaction. In addition to acidic residues, several AAs including tyrosine, valine, leucine, threonine and histidine have been detected as SDASO cross-linked sites more often than other AAs, in which tyrosine and histidine residues have exhibited favored carbene insertion in the past [102]. The preferred reactivity of SDASO cross-linkers towards a subset of AAs including ones that cannot be easily targeted by specific cross-linking chemistries is beneficial to XL-MS studies, as it helps enhance the analysis of the resulting photoactivated cross-linked peptides and expand PPI coverage.

The complementarity in PPI mapping among the three SDASO linkers appears to be much more pronounced in the XL-MS analyses of proteasomes than BSA, implying the benefits of variable linker lengths for complex PPI profiling. In comparison to residue-specific cross-linkers such as DSSO, DHSO and BMSO [65,66], SDASO XL-MS analyses of BSA has yielded the highest number of cross-linked peptides and the most comprehensive interaction maps. The high-density SDASO XL-maps of BSA illustrates the effectiveness of the heterobifunctional photocross-linkers for mapping a diverse range of interactions, which is in good agreement with previous reports [97,98]. Intriguingly, while SDASO XL-MS analysis of the yeast 26S proteasome identified extensive inter-subunit and intra-subunit interactions, the overall scopes of PPIs obtained from all three

SDASO linkers is only comparable to those by DSSO and other residue-specific cross-linkers [148]. Although DSSO produced a higher number of cross-linked peptides of the 26S proteasome than SDASO, comparisons of their cross-linked peptide sequences have revealed limited overlaps. Due to diazirine non-specificity, SDASO XL-maps of the 26S proteasome contain much more residue-to-residue connectivity. In addition, the three SDASO linkers have captured more interactions of the stable and compact 20S CP, but less of the dynamic and flexible 19S RP than DSSO. Since the spacer arm lengths of DSSO and SDASO linkers are similar, variance in PPI coverages is mostly attributed to cross-linkers' reactivity and kinetics [101]. Collectively, our results have demonstrated the value of SDASO photocross-linkers in probing PPIs of both simple and complex samples. The extensive SDASO XL-MS data have allowed us not only to obtain comprehensive XL-maps complementary to those of existing cross-linkers, but more importantly to better assess the reliability and capability of diazirine cross-linking in probing PPIs. Therefore, this work has established a solid foundation for future applications of photocross-linking in complex XL-MS studies.

Chapter 4

Structural Dynamics of the Human COP9 Signalosome Revealed by Cross-linking Mass Spectrometry and Integrative Modeling

In this work, we developed and employed a multi-chemistry cross-linking mass spectrometry approach enabled by three MS-cleavable cross-linkers to obtain comprehensive PPI maps of the CSN (CSN9-free) and CSNn (CSN9-bound) complexes to significantly improve precision and accuracy of their models. Based on our cross-link data, X-ray structures and comparative models of CSN subunits, we computed the complete integrative structures of CSN and CSNn at 16 Å and 22 Å precisions, respectively. The integrative structures have maintained the core architecture of the known X-ray structure of CSN (PDB 4D10), but importantly revealed additional conformations and configurations of CSN in solution that were absent in the static structure. The integrative structure of CSNn has defined the CSN9 binding site in a cleft formed among CSN1, CSN3 and CSN8, resulting in local subunit reorientations that more likely contribute to CSN9-dependent increase of CSN deneddylase activity in vitro. Collectively, this work not only provides new molecular features for us to better determine the structure dynamics of the CSN complex, but also reveals the structural basis underlying the role of CSN9 in CSN-mediated activities. Moreover, the integrated structural approach presented here is effective and can be generalized to define in-solution structures of dynamic protein complexes that remain inaccessible to other approaches.

Experimental Methods for this section can be found in Appendix D.

RESULTS

Multi-chemistry XL-MS Strategy for CSN Complexes

To define the architectures of CSN and CSNn complexes, we aimed to perform a comprehensive XL-MS analysis to maximize PPI mapping and to facilitate integrative structure modeling. To this end, we developed a combinatory XL-MS approach based on multiple MS-cleavable cross-linkers that carry specific but complementary cross-linking chemistries. Specifically, we selected three sulfoxide-containing MS-cleavable cross-linkers that target lysines (DSSO) [60], acidic residues (DHSO) [65], and cysteines (BMSO) [66], respectively. This combination is based on the critical roles of the selected reactive residues in protein structures, and the complementarity of the resulting cross-links for mapping PPIs. Both lysines and acidic residues are highly prevalent and often enriched at protein interaction interfaces, whereas cysteines are less abundant but can be more selective for targeting specific regions. In addition, no disulfide bonds have been reported for CSN subunits, indicating that cysteine cross-linking would be suited for structural analysis of CSN. Importantly, the usage of these reagents has shown to significantly improve the coverage of PPI mapping even for simple proteins [65,66]. The general workflow of our multi-chemistry XL-MS strategy is illustrated in Figure 4.1. CSN complexes were purified under reducing condition after co-expression in *E. Coli* (Table 4.1-4.2), which were catalytically active and used for all XL-MS experiments. It is noted that CSN7 exists as two functionally redundant homologs in mammalian cells, CSN7a and CSN7b [149]. Here, CSN7b was expressed and incorporated into CSN complexes for structural analysis. Each purified complex was first subjected to DSSO, DHSO and BMSO cross-linking separately (Fig. 4.1). The resulting cross-linked complexes were then enzymatically digested and separated to enrich cross-linked peptides by peptide size exclusion chromatography (SEC) [43]. The cross-links identified by LC MS^n analysis were then used for generating 2-D XL-maps to describe inter-subunit interactions and for integrative structure modeling.

Identification of CSN Cross-linked Peptides

To illustrate cross-link identification, representative MS^n spectra of DSSO, DHSO, and BMSO cross-linked peptides of CSN are displayed in Figure 4.2. As DSSO, DHSO, and BMSO cross-linked peptides all carry two symmetric MS-cleavable bonds adjacent to the central sulfoxide in linker regions, cleavage of either one during MS^2 analysis physically separates cross-linked peptide constituents (α and β), resulting in the detection of two characteristic fragment ion pairs modified with complementary cross-linker remnants (α_A/β_T & α_T/β_A), regardless of linker chemistries (Fig. 4.2A-C). MS^3 analyses of these characteristic MS^2 fragment ion pairs enabled accurate identification of their sequences (Fig. 4.2D-I). In combination with MS^1 and MS^2 data, DSSO, DHSO, and BMSO cross-linked peptides were identified unambiguously. In this work, we have performed at least 4 biological replicates for each XL-MS experiment. As a result, from all of XL-MS experiments, a total of 682 DSSO, 275 DHSO, and 456 BMSO unique cross-linked peptides of CSN (Tables 4.3-4.5), and a total of 856 DSSO, 723 DHSO, and 576 BMSO unique cross-linked peptides of CSNn (Tables 4.6-4.9) were identified, respectively. Based on the identified cross-linked peptides, residue-to-residue linkages were determined (Fig. 4.3). To ensure the validity of subsequent analyses, we decided to only use highly reproducible residue-to-residue linkages that have ≥60% occurrence among all biological replicates of each experiment. Thus, we obtained a total of 452 highly reproducible cross-links for CSN, including 214 K-K, 169 D/E-D/E and 69 C-C linkages, describing 205 inter-subunit (74 DSSO, 91 DHSO, and 40 BMSO) and 247 intra-subunit interactions (140 DSSO, 78 DHSO, 29 BMSO) (Table 4.10-4.12). For CSNn, a total of 544 highly reproducible cross-links were acquired with 269 K-K, 167 D/E-D/E and 108 C-C linkages, representing 244 inter-subunit (86 DSSO, 83 DHSO, and 75 BMSO) and 300 intra-

subunit interactions (183 DSSO, 84 DHSO, 33 BMSO) (Table 4.14-4.16). These high confidence cross-links were used for subsequent analyses (Fig. 4.3).

The CSN Interaction Topology

To define inter-subunit physical contacts, we generated experimentally-derived interaction topology maps of CSN complexes based on the highly reproducible cross-link data (Fig. 4.4, and Tables 4.10-4.17). As a result, extensive interaction networks were formulated comprising a total of 26 and 24 unique pairwise interactions for CSN and CSNn, respectively (Table 4.13 and 4.17). Among the three linkers, DSSO yielded the most connectivity within CSN, indicating lysine reactive reagents best-suited for general assessment of PPIs within CSN . While DHSO and BMSO identified less overall, they did yield additional subunit contacts. Specifically, DSSO alone identified five unique PPIs; in comparison, DHSO and BMSO yielded a total of seven unique PPIs (Table 4.13). To better assess linker-dependent interactions, we constructed DSSO, DHSO and BMSO PPI maps separately for each CSN complex (Fig. 4.5A-F). Since the majority of CSN subunits possess similar % of K, D/E and C residues in their primary sequences, the number of cross-links representing each inter-subunit interaction is more likely dependent on the number of cross-linkable residues at their interaction interfaces as well as the detectability of resulting cross-linked peptides. This is further illustrated by 2-D XL-maps (Fig. 4.5G-L). For example, for the two smallest subunits of CSN, CSN7b has a relatively high percentage of acidic residues and its interactions were mostly revealed by DHSO, whereas CSN8 interactions were better described by DSSO due to its relatively high percentage of lysines (Fig. 4.4A and Fig. 4.5A-C).

Similar to the CSN complex, all three linkers have yielded extensive and complementary cross-links to represent subunit interconnectivities of CSNn (Table 4.17). Importantly, 16 CSN9-containing cross-links have been identified (Table 4.15), demonstrating its physical interactions

within CSN at the residue level. Specifically, the C-terminal tail of CSN1 and several regions across CSN3 have been found to interact with CSN9. Since CSN9 is highly acidic with few lysine and no cysteine residues, only DHSO was able to capture CSN9 interactions within the CSNn complex. With the addition of CSN9, it appears that CSNn presented unique characteristics in its XL-maps in comparison to those of CSN (Fig. 4.4B and Fig. 4.5D-F). This suggests that CSN9 may induce local changes in the CSNn complex that impact cross-link formation. Collectively, our results have demonstrated the effectiveness and complementarity of our combinatory XL-MS strategy in mapping PPIs within CSN complexes. Integration of multi-chemistry cross-linking enabled not only cross-validation of inter-subunit interactions, but also expanded interaction coverage due to their distinct capabilities of uncovering interactions at specific protein regions.

Mapping of CSN Cross-links to the X-ray Structure

To assess whether the cross-links agreed with the X-ray structure, we first mapped the identified K-K, D/E-D/E and C-C linkages of CSN complexes to the existing CSN X-ray structure (3.8 Å, PDB 4D10) by determining their Cα-Cα spatial distances (Table 4.10-4.17). Considering linker spacer arm lengths (i.e. DSSO (10.1 Å), DHSO (14.3 Å) and BMSO (24.2 Å)), side chain lengths of targeted amino acids (i.e. K (5.4 Å), D/E (2.5/3.7 Å) and C (2.8 Å)), as well as side chain flexibility and dynamics, we have estimated the maximum Cα-Cα distances spanned by each linker: DSSO at 30 Å, DHSO at 30 Å, and BMSO at 45 Å. Thus, cross-links with distances above these thresholds were considered non-satisfying or violating. For inter-subunit interactions, 60% of DSSO cross-links of CSN were considered violating (Fig. 4.6A). This is surprising as usually less than 20% of lysine-reactive cross-links are violated when mapped onto existing high-resolution structures [26,74,150]. Similar discrepancies with the X-ray structure were observed for DHSO and BMSO data as 55% DHSO and 87% of BMSO inter-subunit cross-links were beyond

the expected thresholds (Fig. 4.6B-C). In contrast, most of intra-subunit cross-links of CSN were satisfied in the X-ray structure, with only 12% of DSSO, 15% of DHSO and 21% BMSO violating intra-subunit cross-links (Fig. 4.6A-C). Since the high-resolution structure of CSNn has not been resolved, we also mapped CSNn cross-links onto the same CSN structure. Similarly, a significant portion of inter-subunit cross-links of CSNn from all three linkers, i.e. 57% of DSSO, 52% of DHSO and 84% BMSO, were non-satisfying (Fig. 4.6D-F), whereas for the intra-subunit cross-links, only 10% of DSSO, 10% of DHSO and 23% BMSO were non-satisfying (Fig. 4.6D-F). The high proportion of violating inter-subunit cross-links is more likely due to the additional conformations that CSN complexes may adopt in solution beyond the one observed in the X-ray structure.

Integrative Structure Modeling of the CSN Complex

To determine CSN structure in solution, we performed integrative structure modeling using the previously described four-stage workflow (Fig. 4.7, Appendix D, and Table 12) [34,75,86,151-154]. The input information included the highly reproducible cross-link datasets (Tables 4.10-4.17), the X-ray structure of CSN (PDB 4D10), and two comparative models of CSN7b subunit domains based on the structure of the CSN7a subunit in the X-ray structure of CSN. The representation of the system used for modeling of CSN was chosen as follows. First, the helical bundle comprising segments from each of the eight subunits was constrained based on the X-ray structure. Second, the remaining structures of subunits CSN1-8 were represented by 15 rigid bodies, corresponding to different domains of the proteins (Appendix D, Tables 4.18-4.20, and Fig. 4.8H). Finally, short (4-13 residues) segments linking rigid bodies and regions missing in the X-ray structure (2-52 residues long) were modeled as flexible strings of 2-10 residue beads each. Next, we exhaustively sampled configurations of the 16 rigid bodies (i.e., the helical bundle and the 15 rigid bodies) that

satisfy the cross-links as well as sequence connectivity and excluded volume restraints, using a Monte Carlo method that started with a random initial structure. The modeling did not rely on any knowledge of the X-ray structure except for the shapes of the 16 rigid bodies. The sampling yielded 71,350 representative models that sufficiently satisfied the input restraints. The clustering of the ensemble identified a single distinct cluster containing the majority (76%) of the individual models (Fig. 4.8A-D), corresponding to the complete integrative structure of CSN in solution. The precision of the cluster corresponds to the variability among the clustered ensemble and defines the overall precision (uncertainty) of the integrative CSN structure (Fig. 4.8 and Fig. 4.9A), which was quantified by the average root-mean-square deviation (RMSD) with respect to the centroid of 16 Å (Appendix D). The centroid structure is the most similar structure to all the other structures in the cluster.

Validation of the Integrative Structure of the CSN Complex

To validate the integrative structure of CSN, we first assessed how well it satisfied the input cross-links used to compute it. The integrative structure of CSN satisfied 98% of the cross-links. The remaining 2% of the cross-links would be satisfied if the threshold was increased by 10Å (Fig. 4.8F). These violations can be rationalized by experimental uncertainty, coarse-grained representation of the complex, and/or finite structural sampling.

Next, we evaluated the integrative structure of CSN by cross-validation against different input cross-link datasets. Namely, we independently repeated integrative modeling described above with six different subsets of CSN cross-links (Table 4.10-4.13), including (i) DSSO only, (ii) DHSO only, (iii) BMSO only, (iv) DSSO and DHSO, (v) DSSO and BMSO, and (vi) DHSO and BMSO. The results were examined in three ways as follows. First, we gauged how well each of the six CSN model ensembles satisfied different subsets of the cross-links. All six models

satisfied more than 95% of all cross-links, whether or not they were used for modeling, thus increasing our confidence in modeling. Second, we showed that increasing the amount of input information improved the precision of the output model when sampling was exhaustive. This result is expected when the choice of model representation (here, the 16 rigid bodies) is appropriate for input information (here, mainly the cross-links) as encoded in the scoring function. In addition to validating the model and the data, the improved precision of the model resulting from increasing the number of cross-linking chemistries demonstrate the complementarity of the three cross-linking datasets (Fig. 4.9B-C). Specifically, the model precision increased from 37 Å for BMSO cross-links only to 16 Å for all three types of cross-links (i.e. DSSO+DHSO+BMSO). Third, we calculated the overlaps between the integrative structure ensemble using all cross-links and each of the six model ensembles based on a subset of cross-links. The overlap was quantified by the ratio of the distance between ensemble centroids to three times the sum of the ensemble precisions (Appendix D). The distance between two ensemble centroids is defined by their RMSD. The ensemble precision is defined by the RMSD from the centroid averaged over all models in the ensemble. In particular, two structural aspects were evaluated, including the tertiary structure of each individual subunit (a total of 8 subunits) as described by the intramolecular distances as well as the relative positions and orientations of all pairs of subunits (a total of 28 pairs) in the complex as described by the intermolecular distances. For each of the eight subunits and each of the 28 pairs of subunits, the integrative structure based on all cross-links overlapped with the integrative structures based on each of the six cross-link subsets (Fig. 4.10A-F). Therefore, these cross-validations further increased our confidence in the integrative structure of CSN.

Comparison of Integrative and X-ray Structures of CSN

To compare the integrative and X-ray structures of CSN, we first examined how well both structures satisfied our cross-link datasets and determined that the integrative structure did much better than the X-ray structure, for both intra-subunit (98% vs. 85%) and inter-subunit (99% vs. 39%) cross-links (Fig. 4.8F and Tables 4.14-4.16C). These results indicate that the integrative structure ensemble is a better representation of CSN conformations in solution than the X-ray structure.

Next, we inspected whether or not the integrative model preserved the core of the previously determined CSN structures, which contains three main features: 1) the PCI ring (in the order of CSN7-CSN4-CSN2-CSN1-CSN3-CSN8); 2) the CSN5-CSN6 dimer; and 3) a helical bundle consisting of a helix from each of the 8 subunits [124,149,155]. During our modeling, while the helical bundle was constrained as a rigid body (Figs. 4.9A, 4.11A and D), the order of the PCI ring and CSN5-CSN6 dimer were not enforced. However, the latter two features emerged from our simulation and resemble those in the X-ray structure (Figs. 4.8G, 4.9A and 4.11B). This preservation is important especially for the CSN5-CSN6 dimer as it is crucial for keeping CSN5 inactive in the absence of a substrate, and releasing CSN5 for activation upon substrate binding [114,124,125,156]. The CSN5-CSN6 dimer was well-represented by our cross-link data, resulting in the highest precisions among the 28 pairs of subunits in the integrative structure of CSN (16Å) (Figs. 4.8, 4.9A, 4.11B and D). Moreover, subunits CSN3 and CSN8 also adopted similar positions and orientations relative to other subunits in both the integrative and X-ray structures (Figs. 4.8, 4.9A and 4.11A), albeit the precision of the CSN3-CSN8 pair in the integrative structure was relatively low (25Å). In summary, the core of CSN integrative structure in solution is similar to previous X-ray and EM structures [124,149,155].

Finally, we computed the RMSD between the CSN X-ray and integrative structure centroids to assess whether the RMSD was larger than three times the precision of the integrative structure, as the resolution of the X-ray structure is much higher than that of the integrative structure. The crystallographic structures of 3 subunits (i.e. CSN2, CSN4, and CSN5) and 4 pairs of subunits (i.e., CSN2-CSN4, CSN2-CSN5, CSN4-CSN5, and CSN4-CSN6) were found to lie further than three times the integrative structure precision from the ensemble centroid (Fig. 4.11D), indicating significant differences in these regions between the two compared structures. The observed differences were further supported by the largest RMSDs measured in these regions between the X-ray and integrative structure centroid of CSN (Fig. 4.12A). The detected discrepancies are unlikely the result of integrative modeling uncertainty; instead, they likely reflect different functional states in solution and/or differences between the solution and X-ray structures. Specifically, the C-terminus of CSN4 interacts tightly with the C-terminus of CSN6 (precision of 20Å; Figs. 4.8G and 4.11D), opposite from CSN5 in the integrative structure (Fig. 4.9A and 4.11A). In contrast, CSN4 does not interact with CSN6 in the X-ray structure (Fig. 4.11D). The relative positions and orientations of CSN2, CSN4, CSN5, and CSN6 in the integrative structure were determined by satisfying all but one of the 47 inter-subunit cross-links. In contrast, the X-ray structure only satisfied 30 of these cross-links.

Although the arrangement order of CSN1, CSN2 and CSN3 remained unchanged, the N-terminus of CSN2 was found to wrap around CSN1 toward CSN3 in the integrative structure (Figs. 4.9A, 4.11A and C), whereas it projected outwards without contacting neither CSN1 nor CSN3 in the X-ray structure. The relative positions and orientations of CSN1, CSN2, CSN3, and CSN4 in the integrative structure were determined by satisfying all but one of the 98 inter-subunit cross-links. In contrast, the X-ray structure only satisfied 28 of these cross-links and none of the 16 cross-

links between CSN2 and CSN3. Taken together, the results demonstrate that integrative structure modeling of CSN based on our comprehensive cross-link data was able to not only recapitulate the core architecture common to all known CSN structures, but also uncover significant quaternary differences relative to the X-ray structure.

Integrative Structure Modeling of the CSNn Complex

To localize the CSN9 subunit and map its interactions with the CSN complex, we also performed integrative structure modeling of CSNn (CSN9-bound CSN), based primarily on 619 highly reproducible cross-links for CSNn from all three cross-linkers (Fig. 4.12, Appendix D, and Tables 4.21-4.23). Integrative structure modeling of CSNn was performed the same way as described above for CSN. The structure of CSN9, a 57 amino acid-long acidic protein, is unknown and cannot be modeled. Therefore, it was represented as a string of flexible beads corresponding to two residues each. The sampling of the CSNn complex yielded 125,750 representative models that sufficiently satisfied the input restraints. The clustering of the ensemble identified a single distinct cluster containing the majority (79%) of the individual models (Fig. 4.13), corresponding to the complete integrative structure of CSNn in solution. The precision of the cluster is 22 Å (Fig. 4.13A-D), which is sufficient to map all positions and relative orientations of CSN1-9 subunits (Figs. 4.11A, 4.13E, and 4.14A). Moreover, the integrative structure of CSNn satisfied 99% of the input cross-links (inter-subunit and intra-subunit). Importantly, the resulting structure of CSNn has precisely localized CSN9 at a cavity formed by the C-terminus of CSN1, all of CSN3 and CSN8 (Fig. 4.14A). The position of amino acid residues 20-57 of CSN9 was specified by satisfying all of the 16 CSN9-containing inter-subunit cross-links (Fig. 4.14A-B). It is noted that the exact position of the first 19 amino acid residues of CSN9 could not be accurately determined since cross-linked peptides involving this region were not identified. Regardless, we were able to

determine the interactions of CSN9 with CSN1-8 in the integrative structure. We consider a contact between CSN9 and any of the CSN1-8 subunit if the two subunits are within 12Å from each other; a contact is defined as an interaction if the contact frequency across the ensemble is at least 75% (Fig. 4.13G). As a result, CSN1 and CSN3 were found in the closest proximity to CSN9 across the ensemble and thus were identified as CSN9 interactors, corroborating well with our cross-link data. Therefore, the CSN9-CSN interactions have been precisely determined by integrative structure modeling (Figs. 4.13G and 4.14A-B), providing CSN9's binding cavity and its interactors.

Comparison of Integrative Structures of the Canonical and Non-canonical CSNs

To compare the two CSN complexes in light of their precisions, we then examined their structural differences among the conformations of single subunits and configurations of pairs of subunits by assessing whether the differences are larger than the sum of their precisions (Fig. 4.11D) and by computing the RMSD between their respective centroid (Fig. 4.12B). While a large portion of the two compared structures was similar, the conformation of 3 out of the 8 subunits (i.e. CSN2, CSN5, and CSN7) and 3 out of the 28 pairs of subunits (i.e., CSN2-CSN3, CSN2-CSN5, and CSN2-CSN7) had notable differences in these regions (Fig. 4.11D and 4.12B). Both the integrative structures of CSN and CSNn maintained similar core structures (i.e., ordering of the PCI ring, the CSN5-CSN6 dimer, and the helical bundle) (Fig. 4.11B). However, CSN2 changed its conformation and position relative to its neighbors (i.e., CSN3, CSN5, and CSN7) (Figs. 4.11A, 4.11C-D and 4.12B). Specifically, in the integrative structure of CSNn, CSN2 and CSN4 localize adjacent to one another, allowing the formation of the CSN9-binding cavity (Figs. 4.11D and 4.14). The conformation and relative position of the CSN2 subunit in the integrative structure of CSNn were determined by satisfying all 74 inter-subunit cross-links obtained for

CSNn. Therefore, our results suggest that CSN2 possesses structural plasticity, enabling its interaction with CSN1 and CSN3 to yield a more open configuration in CSN9-bound CSN than in CSN9-free CSN.

To explore the potential role of CSN9-mediated structural changes, we compared the integrative structures of CSN and CSNn to the cryo-EM structure of the CRL4A-bound CSN complex (at resolution of 6.4Å) [125]. Specifically, we assessed whether the structure of the CSN complexed with neddylated CRL4A overlapped with the two integrative structures. The structure of CRL4A-bound CSN differs from the integrative structure of CSN for one subunit (i.e., CSN2) and 2 pairs of subunits (i.e., CSN2-CSN4 and CSN2-CSN5) (Fig. 4.12C). In contrast, the structure of CRL4A-bound CSN has no significant differences with the integrative structure of CSNn (Fig. 4.12D). Similar comparisons were performed with the structure of CRL1-bound CSN (at resolution of 7.2Å) [126]. While the structure of CSN bound to neddylated CRL1 differs from the integrative structure of CSN for two subunits (i.e., CSN2 and CSN5) and three pairs of subunits (i.e., CSN2-CSN4, CSN2-CSN5, and CSN2-CSN6) (Fig. 4.12E), it has no significant differences with the integrative structure of CSNn (Fig. 4.12D). Collectively, these assessments suggest that CSN9-bound CSN is structurally similar to CRL-bound CSN [125,126]. Upon CSN9 binding, the integrative structure of CSNn displays local structural changes, mainly on the conformation and position of CSN2. Specifically, CSN2 moves closer to CSN4, causing CSN9-bound CSN to adopt a configuration resembling CRL-bound CSN[125,126].

Biochemical Validation of CSN9 Binding

In order to validate the interactions of CSN9 with the CSN complex revealed by XL-MS and structural modeling, we performed *in vitro* binding assays using purified CSN subunits. CSN9 only interacts with CSN1-2-3 and CSN1-2-3-8 subcomplexes, whereas no binding was detected

with CSN4-6-7, CSN4-6-7-5, or CSN4-6-7-5-8 subcomplexes (Fig. 4.15). These results confirm that CSN1 and CSN3 are present in the subcomplex required for CSN9 binding onto CSN. To understand the importance of CSN9, we have compared *in vitro* deneddylase activities of CSN and CSNn with neddylated Cullin 1 as the substrate. Similar results were obtained for the same assay performed at different time scales (Fig. 4.15A-D), demonstrating that CSNn displayed markedly increased activity over CSN and CSN9 can enhance CSN activity *in vitro*.

Quantitative Validation of the Structural Dynamics of the CSN complexes

To validate the observed structural differences between CSN models with and without CSN9, parallel reaction monitoring (PRM)-based targeted quantitation of CSN cross-links was utilized [131]. Since DHSO cross-linking yielded the most inter-subunit linkages best describing CSN9-induced structural changes, we individually cross-linked CSN and CSNn with DHSO for PRM experiments. To perform unbiased quantitative analysis, we generated a total of 341 PRM targets based on highly reproducible DHSO cross-linked peptides previously obtained from CSN and CSNn complexes (Table 4.10 and 4.15). Peptide quantitation was derived from the summation of peak areas of all transitions through Skyline software. As exemplified in Figure 4.17A, an intra-CSN4 cross-link (E306-E345) from both CSN and CSNn samples displayed similar abundance, indicating that this interaction is independent of CSN9. In contrast, a CSN2-CSN3 cross-link (CSN2:E63-CSN3:E333) was only observed in CSNn and not in CSN, suggesting a CSN9-induced conformational change. In total, 229 DHSO cross-linked peptides were quantified which represent 18 inter-subunit interactions (Table 4.24). As shown in Figure 4.17B, the vast majority of quantified cross-links remained unchanged between CSN and CSNn, confirming that CSN9 does not trigger major organizational changes within the CSN complex during its binding. This corroborates well with the modeling results as both of our CSN models satisfied 99% of DHSO

cross-links from both complexes. Apart from unchanged interactions, a total of 22 cross-linked peptides were found with significant changes (>2.5-fold, greater than 3σ) between the two compared complexes (Fig. 4.17B). Besides cross-links involving CSN9, 2 additional cross-linked peptides corresponding to 2 inter-subunit interactions (i.e. CSN4-CSN6 and CSN6-CSN7) have decreased CSN/CSNn ratios, suggesting that these cross-links are favored in CSNn. In contrast, 18 cross-linked peptides describing 7 inter-subunit interactions (CSN1-CSN2, CSN1-CSN3, CSN1-CSN5, CSN2-CSN3, CSN2-CSN7, CSN4-CSN5, and CSN6-CSN7) and 1 intra-CSN1 interaction have increased CSN/CSNn ratios, implying that these cross-links are preferably formed in CSN. Apart from CSN9-containing interactions, 5 quantifiable CSN2-CSN3 cross-links exhibited the most significant changes between the two compared complexes with CSN/CSNn ratios all greater than 10.2 (Fig. 4.17C), indicating that CSN2-CSN3 interactions were severely disrupted upon CSN9 binding. This is consistent with the structural differences between CSN and CSNn revealed by integrative modeling as these linkages were only satisfied by the CSN models (Fig. 4.17D). Since CSN1 closely interacts with CSN2, CSN3 and CSN9, the decreased abundance of CSN1-CSN2 and CSN1-CSN3 cross-links in CSNn supports the CSNn model, suggesting that the main body of CSN2 swings away CSN1 and CSN3 into a more open state. Collectively, PRM-based targeted quantitation of CSN cross-links strongly supports structural similarities and differences between the integrative models of the two CSN complexes.

DISCUSSION

In this work, we have developed a multi-chemistry XL-MS approach based on three distinct MS-cleavable cross-linkers (i.e. DSSO, DHSO and BMSO) to comprehensively map PPIs and facilitate integrative structure modeling of CSN complexes. The large number of cross-links identified in this work is highly complementary, allowing expanding PPI coverage and cross-

validating results. This approach enables us to obtain the most extensive intra-subunit and inter-subunit interaction maps of CSN (CSN9-free) and CSNn (CSN9-bound) complexes. It is noted that CSN9-containing interactions were only identified through DHSO cross-linking, not by DSSO and BMSO, signifying the need of multi-chemistry XL-MS to fully characterize PPIs of CSN complexes. Importantly, the combinatory XL-MS data enabled structural characterization of the CSN complexes with complete sequences and significantly enhanced the precision of integrative structure modeling, resulting in the precisions of 16 Å and 22 Å for CSN and CSNn, respectively. These are considerably higher than the precision of models from single and dual cross-linking chemistries (24Å-37Å). While lysine-to-lysine and acidic-to-acidic residue cross-links have been successfully applied for structural mapping and/or modeling [10,65,75,79,80], we demonstrated here that cysteine-to-cysteine cross-links are as effective for structure determination of protein complexes. This is illustrated by the fact that a single integrative structure (i.e., a single cluster of models) satisfies most of the BMSO cross-links, similarly to DSSO and DHSO cross-links (Figs. 4.8 and 4.9). In addition, we obtained highly overlapping model ensembles based on 7 different combinations of the three types of cross-link data, i.e. DSSO, DHSO and BMSO cross-links (Figs. 4.9C and 4.10), confirming the validity and coherence of our cross-link data. Therefore, coupling combinatory XL-MS based on multiple cross-linking chemistries with integrative structure modeling facilitates the determination of the interaction and structure dynamics of CSN complexes. The same strategy can be directly adopted for characterizing architectures of other dynamic protein complexes in solution.

During XL-MS analyses, we have found that although the majority of intra-subunit cross-links of CSN from all three linkers were satisfied by the known X-ray structure (PDB 4D10), most of inter-subunit cross-links were classified as violating. This implies that CSN has much more

flexible inter-subunit than intra-subunit interactions. Since X-ray crystallography only reveals static structures with single conformation, distance violation of cross-links suggests the presence of multiple conformations and configurations of CSN in solution. Similar results have been obtained for the CSNn complex, further confirming the interaction and structural plasticity of CSN complexes. While CSN is known to carry structural flexibility to allow its interaction with a diverse array of CRLs to regulate their activities [114,124-126], our XL-MS results provide additional evidence to support CSN structural heterogeneity in solution. Because of this, our cross-link dataset generated here is comprised of a wide range of possible conformations of CSN complexes. Therefore, to minimize complexity, only highly reproducible cross-link data was used to derive structural ensembles that represent major conformations of CSN complexes in solution. The integrative structures of CSN complexes have satisfied 98% of all of the cross-links obtained in this work, considerably better than the X-ray structure. This result further indicates that CSN contains additional accessible states other than the one determined by X-ray crystallography. In contrast to the observed conformational and configurational differences in inter-subunit interactions, the core structure of CSN is preserved. Indeed, we have found that the CSN model maintains overall configuration with the presence of the PCI ring and the positioning of CSN5-CSN6 dimer, apart from a rearrangement of CSN2 with respect to CSN1 and CSN3 and CSN4 positioning in the complex. The core structure of CSN has also been detected in the CSNn model, which was derived from a completely different set of cross-link data used for CSN modeling. As these core modules are crucial for the CSN assembly, structure and function [114,124-126,149,155], their determination by integrative modeling based primarily on cross-links further demonstrates the effectiveness of our approach and the validity of the determined integrative structures.

Here, we have determined that CSN9 predominantly interacts with CSN3 and CSN1, and is localized in a cavity formed by CSN1-3-8 in the CSNn structure. Although CSN3-CSN9 interaction has previously been shown biochemically [128], our results have identified interaction contacts between the two interactors. Importantly, we have identified CSN1 as an additional CSN9 interactor and determined CSN9 binding sites within the CSN complex. While it has been suggested that CSN9 may bind to CSN5 and CSN6 [128], no cross-links between CSN9 and CSN5 or CSN6 were identified and the integrative structure of CSNn shows that both subunits are much farther away from CSN9 than CSN1 and CSN3. Interestingly, CSN1, CSN3 and CSN8 form a connected submodule in the integrative and X-ray structures of CSN [124], and the assembled CSN1-CSN3-CSN8 subcomplex can be isolated in mammalian cells [157]. It is known that each CSN subunit has a corresponding homolog in the 9-subunit 19S lid complex [128,158]. Recently, the proteasome subunit DSS1/Rpn15, the homolog of CSN9, has been determined to interact with Rpn3 (homolog of CSN3) and Rpn7 (homolog of CSN1), which forms a subcomplex prior to the 19S lid assembly [159], corroborating well with the close interactions of CSN9 with CSN3 and CSN1. These results further indicate interaction similarities between the CSN and the 19S lid complexes.

Apart from similarities in organizational architectures in the CSN integrative and X-ray structures, we have observed structural differences between the integrative structures of CSN and CSNn that may contribute to CSN dynamics. One notable difference is the CSN2-CSN3 interaction and its relative location to CSN1 subunit. Specifically, in the CSN integrative structure, the CSN2 N-terminus wraps around CSN1 towards CSN3 and away from CSN4, whereby CSN2 is not readily available to interact with Cullin and Rbx1. This is of importance because CSN2 plays a major role along with CSN4 in stabilizing the CSN-CRL interaction when CSN binds to CRLs [114,124-126]. CSN1 has been shown to bind to the CRL4A adaptor DDB1, which is important in

stabilizing Cul4A and required for efficient deneddylation [125]. However, CSN1 involvement appears to be specific for CRL4 and not CRL2 and CRL3 complexes [125,127,160]. While CSN3 has not been shown to directly contact CRL components, overexpression of CSN3 leads to increased amounts of CSN in cells and downregulation of CSN3 causes the destruction of CSN and cell death [161]. Thus, we speculate that the observed changes of interactions among CSN1, CSN2 and CSN3 may represent one of the major conformations of CSN that is needed to interact with specific subsets of CRLs in cells.

While the integrative structures of CSN and CSNn have both maintained the core structure of CSN, CSN9 binding causes a major shift in CSN2 and its interactions with neighboring subunits that have been confirmed by quantitative XL-MS analysis. Given the critical importance of CSN2 in CSN-CRL interactions [114,124-126], we suspect that CSN9-induced structural changes may be associated with the augmented CSN *in vitro* deneddylase activity observed in this work. Comparative analysis has revealed that the major differences between canonical CSN (CSN9-free) and CRL-bound CSN lie in the relative position of CSN2 and its interaction with CSN5 (Fig. 4.12C and E), indicating that CSN2 has to undergo conformational changes to fulfill its role in facilitating CSN binding to CRLs [125,126]. Therefore, the observed structural alterations at CSN2 would be important for the formation of CSN-CRL complex, the prerequisite for subsequent deneddylation. The structure similarity between CSN9-bound CSN and CRL-bound CSN (Fig. 4.12D) strongly supports the biological relevance of CSN9-induced structural changes. Thus, these results prompt us to propose a structural model in which CSN9 causes the canonical CSN to adopt a configuration favorable for interacting with CRLs (Fig. 4.18). In the absence of CSN9, binding of neddylated CRL to CSN results in a series of conformational changes, among which the initial important steps involve the movement of N-terminal domains of CSN2 and CSN4 towards cullin [114,124-126]. These

rearrangements occur prior to the release and activation of CSN5. In contrast, the addition of CSN9 triggers CSN to undergo conformational changes by repositioning the N-terminus of CSN2 away from CSN3 but closer to CSN4 (Fig. 4.18). As the resulting conformation and configuration of CSN9-bound CSN are highly similar to those of CRL-bound CSN, we suspect that CSN9 may enhance the affinity (or recognition) between CSN and its substrate, neddylated CRLs, thus facilitating the assembly of CSN-neddylated CRL complex to enhance CSN activation and deneddylation of CRLs. In addition, the conformation of CSNn may also enable its faster release from deneddylated CRLs as reported [116]. In the absence of CSN9, the assembly/disassembly of CSN-CRL complex would more likely be much slower due to substantial conformational changes required for the activation of CSN upon binding to CRLs, thus leading to slower deneddylation rate. Therefore, the differences in the assembly/disassembly of CSN-CRL complexes more likely contribute to their interaction affinity, and slower disassembly of CSN-CRL complex could imply tighter interaction. In summary, CSN9-induced conformational changes related to CSN2, are biologically relevant, especially in preparing CSN for associating with neddylated CRLs, thereby contributing to augmenting deneddylation activity of CSN. The integrative structures of CSN complexes determined in this work have established a structural basis for us to further dissect condition-induced structural dynamics of CSN in the future, unraveling molecular insights into its activation, function and regulation under different physiological and pathological conditions.

CONCLUSION

In summary, we report the development and characterization of the first classes of MS-cleavable, homobifuntional hydrazide, homobifunctional maleimide, and heterobifunctional NHS-diazirine photoactivated cross-linkers. The unique designs of these linkers enable either a symmetrical (DHSO and BMSO) liable or a single labile (SDASO) bond to be preferentially cleaved over peptide backbone. In addition, we have demonstrated all of these linkers are effective for protein cross-linking and their cross-linked peptides can be analyzed using the same MS^n-based workflow developed for symmetric sulfoxide-containing cross-linkers [46,60,65,66]. Although MS^2-based approach has been widely used in XL-MS studies [10], it is important to note that MS^n analysis is preferred for cross-link identification especially for nonspecific cross-linkers to ensure the robustness and accuracy in cross-link identification especially for complex samples. First, the MS^n-based workflow enabled SDASO XL-MS analysis of the yeast 26S proteasome, demonstrating the feasibility of photocross-linking of large protein complexes for the first time. Secondly, the MS^n-based workflow tyring together multiple chemistries enabled the comprehensive mapping of PPIs and facilitated integrative structure modeling of CSN complexes. The large number of cross-links identified in this work is highly complementary, allowing expanding PPI coverage and cross-validating results. This approach enables us to obtain the most extensive intra-subunit and inter-subunit interaction maps of CSN (CSN9-free) and CSNn (CSN9-bound) complexes. It is noted that CSN9-containing interactions were only identified through DHSO cross-linking, not by DSSO and BMSO, signifying the need of multi-chemistry XL-MS to fully characterize PPIs of CSN complexes. Furthermore, the extensive XL-MS data captured at the simple protein level from BSA to the much more complex samples like the Cop9 Signalosome and 26S Proteasome have allowed us not only to obtain comprehensive XL-maps complementary to

those of existing cross-linkers, but more importantly to better assess the reliability and capability of multi chemistry residue targeting cross-linking in probing PPIs at different scales. Therefore, this work has provided the groundwork for future exploration of other chemistry reactions for XL-MS studies. The development of these cross-linkers further demonstrates the robustness and potential of our XL-MS technology based on sulfoxide-containing MS-cleavable cross-linkers and provides a viable analytical platform for the expansion of new MS-cleavable cross-linker derivatives to generate a complete PPI map of cellular systems in the future.

REFERENCES

1 Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001-1009, doi:10.1038/nature06526 (2007).

2 Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chemistry & biology* **21**, 1102-1114, doi:10.1016/j.chembiol.2014.09.001 (2014).

3 Mendoza, V. L. & Vachet, R. W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrometry Reviews* **28**, 785-815, doi:https://doi.org/10.1002/mas.20203 (2009).

4 Kiselar, J. G. & Chance, M. R. Future directions of structural mass spectrometry using hydroxyl radical footprinting. *J Mass Spectrom* **45**, 1373-1382, doi:10.1002/jms.1808 (2010).

5 Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society Reviews* **40**, 1224-1234, doi:10.1039/C0CS00113A (2011).

6 Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev* **25**, 663-682 (2006).

7 Petrotchenko, E. V., Serpa, J. J. & Borchers, C. H. An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics. *Molecular & cellular proteomics : MCP* **10**, M110.001420, doi:10.1074/mcp.M110.001420 (2011).

8 Leitner, A. *et al.* Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Molecular & cellular proteomics : MCP* **9**, 1634-1649, doi:10.1074/mcp.R000001-MCP201 (2010).

9 Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *Journal of structural biology* **173**, 530-540, doi:10.1016/j.jsb.2010.10.014 (2011).

10 Yu, C. & Huang, L. Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Analytical chemistry* **90**, 144-165, doi:10.1021/acs.analchem.7b04431 (2018).

11 Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K. & Heck, A. J. R. Ion mobility mass spectrometry of proteins and protein assemblies. *Chemical Society Reviews* **39**, 1633-1655, doi:10.1039/B914002F (2010).

12 Lorenzen, K. & Duijn, E. v. Native Mass Spectrometry as a Tool in Structural Biology. *Current Protocols in Protein Science* **62**, 17.12.11-17.12.17, doi:https://doi.org/10.1002/0471140864.ps1712s62 (2010).

13 Rajabi, K., Ashcroft, A. E. & Radford, S. E. Mass spectrometric methods to analyze the structural organization of macromolecular complexes. *Methods* **89**, 13-21, doi:https://doi.org/10.1016/j.ymeth.2015.03.004 (2015).

14 Politis, A. & Borysik, A. J. Assembling the pieces of macromolecular complexes: Hybrid structural biology approaches. *Proteomics* **15**, 2792-2803, doi:https://doi.org/10.1002/pmic.201400507 (2015).

15 Liu, F., Rijkers, D. T., Post, H. & Heck, A. J. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat Methods* **12**, 1179-1184, doi:10.1038/nmeth.3603 (2015).

16 Sinz, A., Arlt, C., Chorev, D. & Sharon, M. Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology. *Protein science : a publication of the Protein Society* **24**, 1193-1209, doi:10.1002/pro.2696 (2015).
17 Dailing, A., Luchini, A. & Liotta, L. Unlocking the secrets to protein-protein interface drug targets using structural mass spectrometry techniques. *Expert review of proteomics* **12**, 457-467, doi:10.1586/14789450.2015.1079487 (2015).
18 Holding, A. N. XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods* **89**, 54-63, doi:https://doi.org/10.1016/j.ymeth.2015.06.010 (2015).
19 Srinivasa, S., Ding, X. & Kast, J. Formaldehyde cross-linking and structural proteomics: Bridging the gap. *Methods* **89**, 91-98, doi:https://doi.org/10.1016/j.ymeth.2015.05.006 (2015).
20 Rivera-Santiago, R. F., Sriswasdi, S., Harper, S. L. & Speicher, D. W. Probing structures of large protein complexes using zero-length cross-linking. *Methods* **89**, 99-111, doi:https://doi.org/10.1016/j.ymeth.2015.04.031 (2015).
21 Piotrowski, C., Ihling, C. H. & Sinz, A. Extending the cross-linking/mass spectrometry strategy: Facile incorporation of photo-activatable amino acids into the model protein calmodulin in Escherichia coli cells. *Methods* **89**, 121-127, doi:https://doi.org/10.1016/j.ymeth.2015.02.012 (2015).
22 Hurt, E. & Beck, M. Towards understanding nuclear pore complex architecture and dynamics in the age of integrative structural analysis. *Current Opinion in Cell Biology* **34**, 31-38, doi:https://doi.org/10.1016/j.ceb.2015.04.009 (2015).
23 Faini, M., Stengel, F. & Aebersold, R. The Evolving Contribution of Mass Spectrometry to Integrative Structural Biology. *Journal of the American Society for Mass Spectrometry* **27**, 966-974, doi:10.1021/jasms.8b05300 (2016).
24 Chait, B. T., Cadene, M., Olinares, P. D., Rout, M. P. & Shi, Y. Revealing Higher Order Protein Structure Using Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **27**, 952-965, doi:10.1007/s13361-016-1385-1 (2016).
25 Chen, F. *et al.* Applying mass spectrometry to study non-covalent biomolecule complexes. *Mass Spectrometry Reviews* **35**, 48-70, doi:https://doi.org/10.1002/mas.21462 (2016).
26 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* **41**, 20-32, doi:10.1016/j.tibs.2015.10.008 (2016).
27 Lössl, P., van de Waterbeemd, M. & Heck, A. J. R. The diverse and expanding role of mass spectrometry in structural and molecular biology. *The EMBO Journal* **35**, 2634-2657, doi:https://doi.org/10.15252/embj.201694818 (2016).
28 Tran, B. Q., Goodlett, D. R. & Goo, Y. A. Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1864**, 123-129, doi:https://doi.org/10.1016/j.bbapap.2015.05.015 (2016).
29 Barysz, H. M. & Malmström, J. Development of Large-scale Cross-linking Mass Spectrometry. *Molecular & Cellular Proteomics* **17**, 1055, doi:10.1074/mcp.R116.061663 (2018).
30 Lipstein, N. *et al.* Presynaptic Calmodulin targets: lessons from structural proteomics. *Expert review of proteomics* **14**, 223-242, doi:10.1080/14789450.2017.1275966 (2017).

31 Fernandez-Martinez, J. *et al.* Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. *Cell* **167**, 1215-1228 e1225, doi:10.1016/j.cell.2016.10.028 (2016).
32 Sinz, A. Investigation of protein-protein interactions in living cells by chemical crosslinking and mass spectrometry. *Analytical and bioanalytical chemistry* **Jan 15. [Epub ahead of print]** (2010).
33 Bruce, J. E. In vivo protein complex topologies: sights through a cross-linking lens. *Proteomics* **12**, 1565-1575, doi:10.1002/pmic.201100516 (2012).
34 Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450**, 683-694, doi:10.1038/nature06404 (2007).
35 Chen, Z. A. *et al.* Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J* **29**, 717-726, doi:10.1038/emboj.2009.401 (2010).
36 Lasker, K. *et al.* Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1380-1387 (2012).
37 Senko, M. W. *et al.* Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Analytical chemistry* **85**, 11710-11714, doi:10.1021/ac403115c (2013).
38 Erzberger, J. P. *et al.* Molecular architecture of the 40SeIF1eIF3 translation initiation complex. *Cell* **158**, 1123-1135, doi:10.1016/j.cell.2014.07.044 (2014).
39 Shi, Y. *et al.* Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Molecular & cellular proteomics : MCP* **13**, 2927-2943, doi:10.1074/mcp.M114.041673 (2014).
40 Young, M. M. *et al.* High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci U S A.* **97**, 5802-5806 (2000).
41 Herzog, F. *et al.* Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* **337**, 1348-1352, doi:10.1126/science.1221483 (2012).
42 Davies, G. E. & Stark, G. R. Use of dimethyl suberimidate, a cross-linking reagent, in studying the subunit structure of oligomeric proteins. *Proceedings of the National Academy of Sciences of the United States of America* **66**, 651-656, doi:10.1073/pnas.66.3.651 (1970).
43 Leitner, A. *et al.* Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Molecular & cellular proteomics : MCP* **11**, M111 014126, doi:10.1074/mcp.M111.014126 (2012).
44 Soderblom, E. J., Bobay, B. G., Cavanagh, J. & Goshe, M. B. Tandem mass spectrometry acquisition approaches to enhance identification of protein-protein interactions using low-energy collision-induced dissociative chemical crosslinking reagents. *Rapid Commun Mass Spectrom* **21**, 3395-3408 (2007).
45 Luo, J., Fishburn, J., Hahn, S. & Ranish, J. An integrated chemical cross-linking and mass spectrometry approach to study protein complex architecture and function. *Molecular & cellular proteomics : MCP* **11**, M111 008318, doi:10.1074/mcp.M111.008318 (2012).
46 Kaake, R. M. *et al.* A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells. *Molecular & cellular proteomics : MCP* **13**, 3533-3543, doi:10.1074/mcp.M114.042630 (2014).

47 Zhang, H. *et al.* In vivo identification of the outer membrane protein OmcA-MtrC interaction network in Shewanella oneidensis MR-1 cells using novel hydrophobic chemical cross-linkers. *Journal of proteome research* **7**, 1712-1720 (2008).
48 Zhang, H. *et al.* Identification of protein-protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry. *Molecular & cellular proteomics : MCP* **8**, 409-420 (2009).
49 Burke, A. M. *et al.* Synthesis of two new enrichable and MS-cleavable cross-linkers to define protein-protein interactions by mass spectrometry. *Organic & biomolecular chemistry* **13**, 5030-5037, doi:10.1039/c5ob00488h (2015).
50 Weisbrod, C. R. *et al.* In vivo protein interaction network identified with a novel real-time cross-linked peptide identification strategy. *Journal of proteome research* **12**, 1569-1579, doi:10.1021/pr3011638 (2013).
51 Sinz, A. Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions. *Analytical and bioanalytical chemistry* **409**, 33-44, doi:10.1007/s00216-016-9941-x (2017).
52 Yang, L. *et al.* A Photocleavable and Mass Spectrometry Identifiable Cross-Linker for Protein Interaction Studies. *Analytical chemistry*, 663-682 (2010).
53 Petrotchenko, E. & Borchers, C. ICC-CLASS: isotopically-coded cleavable crosslinking analysis software suite. *BMC bioinformatics* **11**, 64 (2010).
54 Bennett, K. L. *et al.* Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping--a novel approach to assess intermolecular protein contacts. *Protein science : a publication of the Protein Society* **9**, 1503-1518, doi:10.1110/ps.9.8.1503 (2000).
55 Tang, X., Munske, G. R., Siems, W. F. & Bruce, J. E. Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. *Analytical chemistry* **77**, 311-318, doi:10.1021/ac0488762 (2005).
56 Lu, Y., Tanasova, M., Borhan, B. & Reid, G. E. Ionic reagent for controlling the gas-phase fragmentation reactions of cross-linked peptides. *Analytical chemistry* **80**, 9279-9287, doi:10.1021/ac801625e (2008).
57 Gardner, M. W. & Brodbelt, J. S. Preferential Cleavage of N−N Hydrazone Bonds for Sequencing Bis-arylhydrazone Conjugated Peptides by Electron Transfer Dissociation. *Analytical chemistry* **82**, 5751-5759, doi:10.1021/ac100788a (2010).
58 Trnka, M. J. & Burlingame, A. L. Topographic studies of the GroEL-GroES chaperonin complex by chemical cross-linking using diformyl ethynylbenzene: the power of high resolution electron transfer dissociation for determination of both peptide sequences and their attachment sites. *Molecular & cellular proteomics : MCP* **9**, 2306-2317 (2010).
59 Muller, M. Q., Dreiocker, F., Ihling, C. H., Schafer, M. & Sinz, A. Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. *Analytical chemistry* **82**, 6958-6968, doi:10.1021/ac101241t (2010).
60 Kao, A. *et al.* Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Molecular & cellular proteomics : MCP* **10**, M110.002212, doi:10.1074/mcp.M110.002212 (2011).
61 Chavez, J. D., Weisbrod, C. R., Zheng, C., Eng, J. K. & Bruce, J. E. Protein interactions, post-translational modifications and topologies in human cells. *Molecular & cellular proteomics : MCP* **12**, 1451-1467, doi:10.1074/mcp.M112.024497 (2013).

62 Falvo, F., Fiebig, L. & Schäfer, M. Presentation of a homobifunctional azo-reagent for protein structure analysis by collision-induced dissociative chemical cross-linking: Proof-of-principle in memory of Detlef Schröder. *International Journal of Mass Spectrometry* **s 354–355**, 26–32, doi:10.1016/j.ijms.2013.04.012 (2013).

63 Ihling, C., Falvo, F., Kratochvil, I., Sinz, A. & Schäfer, M. Dissociation behavior of a bifunctional tempo-active ester reagent for peptide structure analysis by free radical initiated peptide sequencing (FRIPS) mass spectrometry. *Journal of Mass Spectrometry* **50**, 396-406, doi:https://doi.org/10.1002/jms.3543 (2015).

64 Yu, C., Kandur, W., Kao, A., Rychnovsky, S. & Huang, L. Developing new isotope-coded mass spectrometry-cleavable cross-linkers for elucidating protein structures. *Analytical chemistry* **86**, 2099-2106, doi:10.1021/ac403636b (2014).

65 Gutierrez, C. B. *et al.* Developing an Acidic Residue Reactive and Sulfoxide-Containing MS-Cleavable Homobifunctional Cross-Linker for Probing Protein-Protein Interactions. *Analytical chemistry* **88**, 8315-8322, doi:10.1021/acs.analchem.6b02240 (2016).

66 Gutierrez, C. B. *et al.* Development of a Novel Sulfoxide-Containing MS-Cleavable Homobifunctional Cysteine-Reactive Cross-Linker for Studying Protein–Protein Interactions. *Analytical chemistry* **90**, 7600-7607, doi:10.1021/acs.analchem.8b01287 (2018).

67 Soderblom, E. J. & Goshe, M. B. Collision-induced dissociative chemical cross-linking reagents and methodology: Applications to protein structural characterization using tandem mass spectrometry analysis. *Analytical chemistry* **78**, 8059-8068, doi:10.1021/ac0613840 (2006).

68 Chakrabarty, J. K., Naik, A. G., Fessler, M. B., Munske, G. R. & Chowdhury, S. M. Differential Tandem Mass Spectrometry-Based Cross-Linker: A New Approach for High Confidence in Identifying Protein Cross-Linking. *Analytical chemistry* **88**, 10215-10222, doi:10.1021/acs.analchem.6b02886 (2016).

69 Hage, C., Falvo, F., Schäfer, M. & Sinz, A. Novel Concepts of MS-Cleavable Cross-linkers for Improved Peptide Structure Analysis. *Journal of the American Society for Mass Spectrometry* **28**, 2022-2038, doi:10.1021/jasms.8b05380 (2017).

70 Hage, C., Iacobucci, C., Rehkamp, A., Arlt, C. & Sinz, A. The First Zero-Length Mass Spectrometry-Cleavable Cross-Linker for Protein Structure Analysis. *Angew Chem Int Ed Engl* **56**, 14551-14555, doi:10.1002/anie.201708273 (2017).

71 Clifford-Nunn, B., Showalter, H. D. H. & Andrews, P. C. Quaternary Diamines as Mass Spectrometry Cleavable Crosslinkers for Protein Interactions. *Journal of the American Society for Mass Spectrometry* **23**, 201-212, doi:10.1021/jasms.8b04216 (2012).

72 Kao, A. *et al.* Mapping the structural topology of the yeast 19S proteasomal regulatory particle using chemical cross-linking and probabilistic modeling. *Molecular & cellular proteomics : MCP* **11**, 1566-1577, doi:10.1074/mcp.M112.018374 (2012).

73 Liu, F. & Heck, A. J. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Current opinion in structural biology* **35**, 100-108, doi:10.1016/j.sbi.2015.10.006 (2015).

74 Wang, X. *et al.* The proteasome-interacting Ecm29 protein disassembles the 26S proteasome in response to oxidative stress. *The Journal of biological chemistry* **292**, 16310-16320, doi:10.1074/jbc.M117.803619 (2017).

75 Wang, X. *et al.* Molecular Details Underlying Dynamic Structures and Regulation of the Human 26S Proteasome. *Molecular & cellular proteomics : MCP* **16**, 840-854, doi:10.1074/mcp.M116.065326 (2017).
76 Gutierrez, C. *et al.* Structural dynamics of the human COP9 signalosome revealed by cross-linking mass spectrometry and integrative modeling. *Proceedings of the National Academy of Sciences* **117**, 4088, doi:10.1073/pnas.1915542117 (2020).
77 Ding, Y. H. *et al.* Increasing the Depth of Mass-Spectrometry-Based Structural Analysis of Protein Complexes through the Use of Multiple Cross-Linkers. *Analytical chemistry* **88**, 4461-4469, doi:10.1021/acs.analchem.6b00281 (2016).
78 Yu, C. *et al.* Probing H₂O₂-mediated Structural Dynamics of the Human 26S Proteasome Using Quantitative Cross-linking Mass Spectrometry (QXL-MS). *Molecular & Cellular Proteomics* **18**, 954, doi:10.1074/mcp.TIR119.001323 (2019).
79 Leitner, A. *et al.* Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9455-9460, doi:10.1073/pnas.1320298111 (2014).
80 Zhang, X. *et al.* Carboxylate-Selective Chemical Cross-Linkers for Mass Spectrometric Analysis of Protein Structures. *Analytical chemistry* **90**, 1195-1201, doi:10.1021/acs.analchem.7b03789 (2018).
81 Iacobucci, C., Piotrowski, C., Rehkamp, A., Ihling, C. H. & Sinz, A. The First MS-Cleavable, Photo-Thiol-Reactive Cross-Linker for Protein Structural Studies. *Journal of the American Society for Mass Spectrometry* **30**, 139-148, doi:10.1021/jasms.8b05915 (2019).
82 Jones, A. X. *et al.* Improving mass spectrometry analysis of protein structures with arginine-selective chemical cross-linkers. *Nature communications* **10**, 3911, doi:10.1038/s41467-019-11917-z (2019).
83 Fioramonte, M. *et al.* XPlex: An Effective, Multiplex Cross-Linking Chemistry for Acidic Residues. *Analytical chemistry* **90**, 6043-6050, doi:10.1021/acs.analchem.7b05135 (2018).
84 Yang, B. *et al.* Proximity-enhanced SuFEx chemical cross-linker for specific and multitargeting cross-linking mass spectrometry. *Proceedings of the National Academy of Sciences* **115**, 11162, doi:10.1073/pnas.1813574115 (2018).
85 Shi, Y. *et al.* A strategy for dissecting the architectures of native macromolecular assemblies. *Nat Methods* **12**, 1135-1138, doi:10.1038/nmeth.3617 (2015).
86 Kim, S. J. *et al.* Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475-482, doi:10.1038/nature26003 (2018).
87 The UniProt, C. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic acids research* **41**, D43-D47, doi:10.1093/nar/gks1068 (2013).
88 Novak, P. & Kruppa, G. H. Intra-molecular cross-linking of acidic residues for protein structure studies. *Eur J Mass Spectrom (Chichester, Eng)* **14**, 355-365, doi:10.1255/ejms.963 (2008).
89 Gunnoo, S. B. & Madder, A. Chemical Protein Modification through Cysteine. *Chembiochem : a European journal of chemical biology* **17**, 529-553, doi:10.1002/cbic.201500667 (2016).
90 Geula, S., Naveed, H., Liang, J. & Shoshan-Barmatz, V. Structure-based analysis of VDAC1 protein: defining oligomer contact sites. *The Journal of biological chemistry* **287**, 2179-2190, doi:10.1074/jbc.M111.268920 (2012).

91 Knepp, A. M., Periole, X., Marrink, S. J., Sakmar, T. P. & Huber, T. Rhodopsin forms a dimer with cytoplasmic helix 8 contacts in native membranes. *Biochemistry* **51**, 1819-1821, doi:10.1021/bi3001598 (2012).
92 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* **57**, 6390-6396, doi:https://doi.org/10.1002/anie.201709559 (2018).
93 Schneider, M., Belsom, A. & Rappsilber, J. Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends in biochemical sciences* **43**, 157-169, doi:10.1016/j.tibs.2017.12.006 (2018).
94 Brodie, N. I., Makepeace, K. A., Petrotchenko, E. V. & Borchers, C. H. Isotopically-coded short-range hetero-bifunctional photo-reactive crosslinkers for studying protein structure. *J Proteomics* **118**, 12-20, doi:10.1016/j.jprot.2014.08.012 (2015).
95 Brodie, N. I., Petrotchenko, E. V. & Borchers, C. H. The novel isotopically coded short-range photo-reactive crosslinker 2,4,6-triazido-1,3,5-triazine (TATA) for studying protein structures. *J Proteomics* **149**, 69-76, doi:10.1016/j.jprot.2016.02.024 (2016).
96 Brodie, N. I., Popov, K. I., Petrotchenko, E. V., Dokholyan, N. V. & Borchers, C. H. Solving protein structures using short-distance cross-linking constraints as a guide for discrete molecular dynamics simulations. *Science Advances* **3**, e1700479, doi:10.1126/sciadv.1700479 (2017).
97 Belsom, A., Schneider, M., Fischer, L., Brock, O. & Rappsilber, J. Serum Albumin Domain Structures in Human Blood Serum by Mass Spectrometry and Computational Biology. *Molecular & cellular proteomics : MCP* **15**, 1105-1116, doi:10.1074/mcp.M115.048504 (2016).
98 Belsom, A., Mudd, G., Giese, S., Auer, M. & Rappsilber, J. Complementary Benzophenone Cross-Linking/Mass Spectrometry Photochemistry. *Analytical chemistry* **89**, 5319-5324, doi:10.1021/acs.analchem.6b04938 (2017).
99 Iacobucci, C. *et al.* Carboxyl-Photo-Reactive MS-Cleavable Cross-Linkers: Unveiling a Hidden Aspect of Diazirine-Based Reagents. *Analytical chemistry* **90**, 2805-2809, doi:10.1021/acs.analchem.7b04915 (2018).
100 Liu, J. *et al.* Photocaged Quinone Methide Crosslinkers for Light-Controlled Chemical Crosslinking of Protein-Protein and Protein-DNA Complexes. *Angewandte Chemie (International ed. in English)* **58**, 18839-18843, doi:10.1002/anie.201910135 (2019).
101 Ziemianowicz, D. S., Ng, D., Schryvers, A. B. & Schriemer, D. C. Photo-Cross-Linking Mass Spectrometry and Integrative Modeling Enables Rapid Screening of Antigen Interactions Involving Bacterial Transferrin Receptors. *Journal of proteome research* **18**, 934-946, doi:10.1021/acs.jproteome.8b00629 (2019).
102 Ziemianowicz, D. S., Bomgarden, R., Etienne, C. & Schriemer, D. C. Amino Acid Insertion Frequencies Arising from Photoproducts Generated Using Aliphatic Diazirines. *Journal of the American Society for Mass Spectrometry* **28**, 2011-2021, doi:10.1007/s13361-017-1730-z (2017).
103 Wei, N. & Deng, X. W. The COP9 signalosome. *Annu Rev Cell Dev Biol* **19**, 261-286 (2003).
104 Wolf, D. A., Zhou, C. & Wee, S. The COP9 signalosome: an assembly and maintenance platform for cullin ubiquitin ligases? *Nat Cell Biol* **5**, 1029-1033 (2003).

105 Wei, N., Serino, G. & Deng, X. W. The COP9 signalosome: more than a protease. *Trends in biochemical sciences* **33**, 592-600 (2008).
106 Cope, G. A. *et al.* Role of predicted metalloprotease motif of Jab1/Csn5 in cleavage of Nedd8 from Cul1. *Science* **298**, 608-611 (2002).
107 Deshaies, R. J. & Joazeiro, C. A. RING domain E3 ubiquitin ligases. *Annu Rev Biochem* **78**, 399-434, doi:10.1146/annurev.biochem.78.101807.093809 (2009).
108 Petroski, M. D. & Deshaies, R. J. Function and regulation of cullin-RING ubiquitin ligases. *Nature reviews. Molecular cell biology* **6**, 9-20 (2005).
109 Cope, G. A. & Deshaies, R. J. COP9 signalosome: a multifunctional regulator of SCF and other cullin-based ubiquitin ligases. *Cell* **114**, 663-671 (2003).
110 Lee, J. E. *et al.* The steady-state repertoire of human SCF ubiquitin ligase complexes does not require ongoing Nedd8 conjugation. *Molecular & cellular proteomics : MCP* **10**, M110 006460, doi:10.1074/mcp.M110.006460 (2011).
111 Skaar, J. R., Pagan, J. K. & Pagano, M. Mechanisms and function of substrate recruitment by F-box proteins. *Nature reviews. Molecular cell biology* **14**, 369-381, doi:10.1038/nrm3582 (2013).
112 Jia, L. & Sun, Y. SCF E3 ubiquitin ligases as anticancer targets. *Curr Cancer Drug Targets* **11**, 347-356 (2011).
113 Emberley, E. D., Mosadeghi, R. & Deshaies, R. J. Deconjugation of Nedd8 from Cul1 is directly regulated by Skp1-F-box and substrate, and the COP9 signalosome inhibits deneddylated SCF by a noncatalytic mechanism. *The Journal of biological chemistry* **287**, 29679-29689, doi:10.1074/jbc.M112.352484 (2012).
114 Enchev, R. I. *et al.* Structural basis for a reciprocal regulation between SCF and CSN. *Cell reports* **2**, 616-627, doi:10.1016/j.celrep.2012.08.019 (2012).
115 Fischer, E. S. *et al.* The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell* **147**, 1024-1039, doi:10.1016/j.cell.2011.10.035 (2011).
116 Füzesi-Levi, M. G. *et al.* CSNAP, the smallest CSN subunit, modulates proteostasis through cullin-RING ubiquitin ligases. *Cell Death & Differentiation* **27**, 984-998, doi:10.1038/s41418-019-0392-8 (2020).
117 Cope, G. A. & Deshaies, R. J. Targeted silencing of Jab1/Csn5 in human cells downregulates SCF activity through reduction of F-box protein levels. *BMC biochemistry* **7**, 1, doi:10.1186/1471-2091-7-1 (2006).
118 Lee, M. H., Zhao, R., Phan, L. & Yeung, S. C. Roles of COP9 signalosome in cancer. *Cell Cycle* **10**, 3057-3066, doi:10.4161/cc.10.18.17320 (2011).
119 Pan, Y. & Claret, F. X. Targeting Jab1/CSN5 in nasopharyngeal carcinoma. *Cancer letters* **326**, 155-160, doi:10.1016/j.canlet.2012.07.033 (2012).
120 Zhong, G., Li, H., Shan, T. & Zhang, N. CSN5 silencing inhibits invasion and arrests cell cycle progression in human colorectal cancer SW480 and LS174T cells in vitro. *International journal of clinical and experimental pathology* **8**, 2809-2815 (2015).
121 Zhang, H. *et al.* COPS5 inhibition arrests the proliferation and growth of serous ovarian cancer cells via the elevation of p27 level. *Biochemical and biophysical research communications* **493**, 85-93, doi:10.1016/j.bbrc.2017.09.070 (2017).
122 Richardson, K. S. & Zundel, W. The emerging role of the COP9 signalosome in cancer. *Mol Cancer Res* **3**, 645-653 (2005).
123 Fischer, E. S. *et al.* Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **512**, 49-53, doi:10.1038/nature13527 (2014).

124 Lingaraju, G. M. *et al.* Crystal structure of the human COP9 signalosome. *Nature* **512**, 161-165, doi:10.1038/nature13566 (2014).
125 Cavadini, S. *et al.* Cullin-RING ubiquitin E3 ligase regulation by the COP9 signalosome. *Nature* **531**, 598-603, doi:10.1038/nature17416 (2016).
126 Mosadeghi, R. *et al.* Structural and kinetic analysis of the COP9-Signalosome activation and the cullin-RING ubiquitin ligase deneddylation cycle. *eLife* **5**, doi:10.7554/eLife.12102 (2016).
127 Faull, S. V. *et al.* Structural basis of Cullin-2 RING E3 ligase regulation by the COP9 signalosome. *bioRxiv*, 483024, doi:10.1101/483024 (2018).
128 Rozen, S. *et al.* CSNAP Is a Stoichiometric Subunit of the COP9 Signalosome. *Cell reports* **13**, 585-598, doi:10.1016/j.celrep.2015.09.021 (2015).
129 Sharon, M. *et al.* Symmetrical modularity of the COP9 signalosome complex suggests its multifunctionality. *Structure* **17**, 31-40 (2009).
130 Schilling, B., Row, R. H., Gibson, B. W., Guo, X. & Young, M. M. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J Am Soc Mass Spectrom.* **14**, 834-850 (2003).
131 Chavez, J. D., Schweppe, D. K., Eng, J. K. & Bruce, J. E. In Vivo Conformational Dynamics of Hsp90 and Its Interactors. *Cell Chem Biol* **23**, 716-726, doi:10.1016/j.chembiol.2016.05.012 (2016).
132 Iacobucci, C., Hage, C., Schafer, M. & Sinz, A. A Novel MS-Cleavable Azo Cross-Linker for Peptide Structure Analysis by Free Radical Initiated Peptide Sequencing (FRIPS). *Journal of the American Society for Mass Spectrometry* **28**, 2039-2053, doi:10.1007/s13361-017-1744-6 (2017).
133 Scott, H. *et al.* Spatial Organization and Molecular Interactions of the Schizosaccharomyces pombe Ccq1-Tpz1-Poz1 Shelterin Complex. *Journal of molecular biology* **429**, 2863-2872, doi:10.1016/j.jmb.2017.08.002 (2017).
134 Liu, F., Lossl, P., Scheltema, R., Viner, R. & Heck, A. J. R. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nature communications* **8**, 15473, doi:10.1038/ncomms15473 (2017).
135 Yu, C. *et al.* Gln40 deamidation blocks structural reconfiguration and activation of SCF ubiquitin ligase complex by Nedd8. *Nature communications* **6**, 10053, doi:10.1038/ncomms10053 (2015).
136 Tang, F. *et al.* One-pot N-glycosylation remodeling of IgG with non-natural sialylglycopeptides enables glycosite-specific and dual-payload antibody-drug conjugates. *Organic & biomolecular chemistry* **14**, 9501-9518, doi:10.1039/c6ob01751g (2016).
137 Fontaine, S. D., Reid, R., Robinson, L., Ashley, G. W. & Santi, D. V. Long-Term Stabilization of Maleimide–Thiol Conjugates. *Bioconjugate Chemistry* **26**, 145-152, doi:10.1021/bc5005262 (2015).
138 Yu, C., Novitsky, E. J., Cheng, N. W., Rychnovsky, S. D. & Huang, L. Exploring Spacer Arm Structures for Designs of Asymmetric Sulfoxide-Containing MS-Cleavable Cross-Linkers. *Analytical chemistry* **92**, 6026-6033, doi:10.1021/acs.analchem.0c00298 (2020).
139 Yu, C., Novitsky, E. J., Cheng, N. W., Rychnovsky, S. D. & Huang, L. Exploring Spacer Arm Structures for Designs of Asymmetric Sulfoxide-Containing MS-Cleavable Cross-Linkers. *Analytical chemistry* **92**, 6026-6033, doi:10.1021/acs.analchem.0c00298 (2020).

140 Gutierrez, C. B. *et al.* Developing an Acidic Residue Reactive and Sulfoxide-Containing MS-Cleavable Homobifunctional Cross-Linker for Probing Protein-Protein Interactions. *Analytical chemistry*, doi:10.1021/acs.analchem.6b02240 (2016).
141 Gutierrez, C. B. *et al.* Development of a Novel Sulfoxide-Containing MS-Cleavable Homobifunctional Cysteine-Reactive Cross-Linker for Studying Protein-Protein Interactions. *Analytical chemistry* **90**, 7600-7607, doi:10.1021/acs.analchem.8b01287 (2018).
142 Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu Rev Biochem* **78**, 477-513 (2009).
143 Bard, J. A. M. *et al.* Structure and Function of the 26S Proteasome. *Annual Review of Biochemistry* **87**, 697-724, doi:10.1146/annurev-biochem-062917-011931 (2018).
144 Unverdorben, P. *et al.* Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 5544-5549, doi:10.1073/pnas.1403409111 (2014).
145 Wehmer, M. & Sakata, E. Recent advances in the structural biology of the 26S proteasome. *The International Journal of Biochemistry & Cell Biology* **79**, 437-442, doi:https://doi.org/10.1016/j.biocel.2016.08.008 (2016).
146 Leggett, D. S. *et al.* Multiple associated proteins regulate proteasome structure and function. *Mol Cell.* **10**, 495-507 (2002).
147 Aufderheide, A. *et al.* Structural characterization of the interaction of Ubp6 with the 26S proteasome. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 8626-8631, doi:10.1073/pnas.1510449112 (2015).
148 Mintseris, J. & Gygi, S. P. High-density chemical cross-linking for modeling protein interactions. *Proceedings of the National Academy of Sciences* **117**, 93, doi:10.1073/pnas.1902931116 (2020).
149 Dubiel, D., Rockel, B., Naumann, M. & Dubiel, W. Diversity of COP9 signalosome structures and functional consequences. *FEBS letters* **589**, 2507-2513, doi:10.1016/j.febslet.2015.06.007 (2015).
150 Liu, F., Lössl, P., Rabbitts, B. M., Balaban, R. S. & Heck, A. J. R. The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. *Molecular & Cellular Proteomics* **17**, 216, doi:10.1074/mcp.RA117.000470 (2018).
151 Rout, M. P. & Sali, A. Principles for Integrative Structural Biology Studies. *Cell* **177**, 1384-1403, doi:10.1016/j.cell.2019.05.016 (2019).
152 Russel, D. *et al.* Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLOS Biology* **10**, e1001244, doi:10.1371/journal.pbio.1001244 (2012).
153 Luo, J. *et al.* Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH. *Molecular cell* **59**, 794-806, doi:10.1016/j.molcel.2015.07.016 (2015).
154 Schneidman-Duhovny, D., Pellarin, R. & Sali, A. Uncertainty in integrative structural modeling. *Current opinion in structural biology* **28**, 96-104, doi:10.1016/j.sbi.2014.08.001 (2014).
155 Rockel, B., Schmaler, T., Huang, X. & Dubiel, W. Electron microscopy and in vitro deneddylation reveal similar architectures and biochemistry of isolated human and Flag-

mouse COP9 signalosome complexes. *Biochemical and biophysical research communications* **450**, 991-997, doi:10.1016/j.bbrc.2014.06.093 (2014).
156 Birol, M. *et al.* Structural and Biochemical Characterization of the Cop9 Signalosome CSN5/CSN6 Heterodimer. *PloS one* **9**, e105688, doi:10.1371/journal.pone.0105688 (2014).
157 Heusel, M. *et al.* Complex-centric proteome profiling by SEC-SWATH-MS. *bioRxiv*, 323683, doi:10.1101/323683 (2018).
158 Sharon, M., Taverner, T., Ambroggio, X. I., Deshaies, R. J. & Robinson, C. V. Structural organization of the 19S proteasome lid: insights from MS of intact complexes. *PLoS Biol* **4**, e267 (2006).
159 Bai, M. *et al.* In-depth Analysis of the Lid Subunits Assembly Mechanism in Mammals. *Biomolecules* **9**, 213, doi:10.3390/biom9060213 (2019).
160 Faull, S. V. *et al.* Structural basis of Cullin-2 RING E3 ligase regulation by the COP9 signalosome. *bioRxiv*, 483024, doi:10.1101/483024 (2018).
161 Peth, A., Berndt, C., Henke, W. & Dubiel, W. Downregulation of COP9 signalosome subunits differentially affects the CSN complex and target protein stability. *BMC biochemistry* **8**, 27 (2007).
162 Wang, X. *et al.* Mass spectrometric characterization of the affinity-purified human 26S proteasome complex. *Biochemistry* **46**, 3553-3565 (2007).
163 Yu, C. *et al.* Developing a New Quantitative Cross-linking Mass Spectrometry (XL-MS) Strategy to Define In Vivo Structural Dynamics of Protein Complexes. *The proceedings of 64th ASMS Conference on Mass Spectrometry and Allied Topics* **June 5-9** (2016).
164 Liang, J. *et al.* Chemical Synthesis of Diubiquitin-Based Photoaffinity Probes for Selectively Profiling Ubiquitin-Binding Proteins. *Angewandte Chemie International Edition* **56**, 2744-2748, doi:https://doi.org/10.1002/anie.201611659 (2017).
165 Sun, R. *et al.* Simple Light-Triggered Fluorescent Labeling of Silica Nanoparticles for Cellular Imaging Applications. *Chemistry – A European Journal* **23**, 13893-13896, doi:https://doi.org/10.1002/chem.201703653 (2017).
166 Ma, L. *et al.* Multidentate Polymer Coatings for Compact and Homogeneous Quantum Dots with Efficient Bioconjugation. *Journal of the American Chemical Society* **138**, 3382-3394, doi:10.1021/jacs.5b12378 (2016).
167 Wang, X., Yen, J., Kaiser, P. & Huang, L. Regulation of the 26S proteasome complex during oxidative stress. *Sci Signal* **3**, ra88 (2010).
168 Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).
169 Sali, A. *et al.* Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure* **23**, 1156-1167, doi:https://doi.org/10.1016/j.str.2015.05.013 (2015).
170 Skinner, O. S. *et al.* Top-down characterization of endogenous protein complexes with native proteomics. *Nature chemical biology* **14**, 36-41, doi:10.1038/nchembio.2515 (2018).
171 Robinson, P. J. *et al.* Molecular architecture of the yeast Mediator complex. *eLife* **4**, doi:10.7554/eLife.08719 (2015).
172 Upla, P. *et al.* Molecular Architecture of the Major Membrane Ring Component of the Nuclear Pore Complex. *Structure* **25**, 434-445, doi:10.1016/j.str.2017.01.006 (2017).
173 Webb, B. *et al.* Modeling of proteins and their assemblies with the Integrative Modeling Platform. *Methods Mol Biol* **1091**, 277-295, doi:10.1007/978-1-62703-691-7_20 (2014).

174 Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology* **234**, 779-815, doi:10.1006/jmbi.1993.1626 (1993).
175 Algret, R. *et al.* Molecular architecture and function of the SEA complex, a modulator of the TORC1 pathway. *Molecular & cellular proteomics : MCP* **13**, 2855-2870, doi:10.1074/mcp.M114.039388 (2014).
176 LoPiccolo, J. *et al.* Assembly and Molecular Architecture of the Phosphoinositide 3-Kinase p85α Homodimer. *The Journal of biological chemistry* **290**, 30390-30405, doi:10.1074/jbc.M115.689604 (2015).
177 Rieping, W., Habeck, M. & Nilges, M. Inferential Structure Determination. *Science* **309**, 303, doi:10.1126/science.1110428 (2005).
178 Shen, M. Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein science : a publication of the Protein Society* **15**, 2507-2524, doi:10.1110/ps.062416606 (2006).
179 Viswanath, S., Chemmama, I. E., Cimermancic, P. & Sali, A. Assessing Exhaustiveness of Stochastic Sampling for Integrative Modeling of Macromolecular Structures. *Biophys J* **113**, 2344-2353, doi:10.1016/j.bpj.2017.10.005 (2017).
180 Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic acids research* **41**, W349-357, doi:10.1093/nar/gkt381 (2013).
181 Rieping, W., Habeck, M. & Nilges, M. Inferential structure determination. *Science* **309**, 303-306, doi:10.1126/science.1110428 (2005).

Appendix A

Chapter 1 Experimental Procedures

Materials and Reagents

General chemicals were purchased from Fisher Scientific or VWR International. Bovine serum albumin (≥96% purity), myoglobin from equine heart (≥90% purity), and DMTMM (≥96% purity) were purchased from Sigma-Aldrich. Ac-SR8 peptide (Ac-SAKAYEHR, 98.22% purity) was custom ordered from Biomatik (Wilmington, DE).

Dihydrazide Sulfoxide (DHSO) Synthesis

Disuccinimidyl sulfoxide (DSSO) was synthesized as previously published 1. The 2-step synthesis scheme for DHSO from DSSO is depicted in Figure 1.1D. Briefly, tert-butyl carbazate (1.10 g, 8.32 mmol) was added to DSSO (1.41 g, 4.16 mmol) in DCM (50 mL). The resulting yellow solution was let stir at room temperature for 12 h, after which trifluoroacetic acid (2.20 mL, 28.7 mmol) was added. The resulting orange solution was let stir for 72 h before removing the solvent in vacuo. The resulting orange oil was dissolved in methanol, and then triethylamine was added. The resulting mixture was let stir for 20 mins, after which a white solid had precipitated. The solid was collected via centrifuge, and then stirred with fresh methanol for 20 mins. The solid was collected via centrifuge again, and this process of stirring with fresh methanol was repeated another two times. Drying the isolated white solid in vacuo afforded DHSO (0.375 g, 46%): mp 159–162 ºC; 1H NMR (500 MHz, DMSO-d6): δ 9.13 (s, 2H), 4.24 (s, 4H), 3.0 –2.97 (m, 2H), 2.83–2.75 (m, 2H), 2.43 (t, J = 7.5 Hz, 4H); 13C NMR (125 MHz, DMSO-d6): δ 169.3, 46.7, 26.2; IR (thin film): 3308, 3044, 1631, 1449, 1297, 1032 cm-1; HRMS (ESI) m/z calculated for C6H15N4O3S [M + H]+ 223.0865, found 223.0857.

DHSO Cross-linking of Synthetic Peptides

DHSO was synthesized as described (Fig. 1.1D). Synthetic peptide Ac-SR8 was dissolved in DMSO to 1 mM and cross-linked with DHSO in a 1:1 molar ratio of peptide to cross-linker in the presence of 1 equivalent of diisopropylethylamine and DMTMM. The resulting samples were diluted to 10 pmol/μL in 3% ACN/2% formic acid prior to MS^n analysis.

DHSO Cross-linking of Equine Myoglobin and Bovine Serum Albumin

50 μL of 50 μM BSA or 200 μM myoglobin in PBS buffer (pH 7.4) was reacted with DHSO in molar ratios of 1:5, 1:10, 1:20, and 1:30. The cross-linking reaction was initiated by adding equivalent concentrations of DHSO and DMTMM to protein solutions, reacted for 1 h at room temperature.

Digestion of DHSO Cross-linked Proteins

Cross-linked protein samples were subjected to either SDS-PAGE followed by in-gel digestion, or directly digested in solution prior to MS analysis [162]. For in-gel digestion, cross-linked proteins were separated by SDS-PAGE and visualized by Coomassie blue staining. The selected cross-linked gel bands were excised, reduced with TCEP for 30 min, alkylated with iodoacetamide for 30 min in the dark, and then digested with trypsin at 37°C overnight. Peptide digests were extracted, concentrated, and reconstituted in 3% ACN/2% formic acid for MSn analysis. For in-solution digestion, cross-linked proteins were first precipitated with TCA and then re-suspended in 8M urea buffer. Reduction and alkylation were performed prior to Lys-C/trypsin digestion as previously described 2. The resulting digests were desalted using Waters C18 Sep-Pak cartridges and fractionated by peptide size exclusion chromatography (SEC) based on the protocol by Leitner

et al. 2014. The fractions containing cross-linked peptides were collected for subsequent MSn analysis.

Liquid Chromatography-Multistage Tandem Mass Spectrometry (LC MS^n) Analysis

DHSO cross-linked peptides were analyzed by LC-MS^n utilizing an Easy-nLC 1000 (Thermo Fisher, San Jose, CA) coupled on-line to an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher, San Jose, CA) [46,64]. LC MS^n data extraction and database searching for the identification of DHSO cross-linked peptides were performed similarly as previously described [60]. First, each MSn experiment consists of one MS scan in FT mode (350-1400 m/z, resolution of 60,000 at m/z 400) followed by two data-dependent MS2 scans in FT mode (resolution of 7500) with normalized collision energy at 10% on the top two MS peaks with charges 4+ or higher, and three MS3 scans in the LTQ with normalized collision energy at 35% on the top three peaks from each MS2.

Data Analysis and Identification of DHSO Cross-linked Peptides

MSn Data extraction and analysis were performed in the same way as previously described [60]. MS3 data was subjected to a developmental version of Protein Prospector (v.5.16.0) for database searching, using Batch-Tag against SwissProt.2014.12.4. random.concat databases limited to either the Bos taurus or Equus caballus taxonomy with mass tolerances for parent ions and fragment ions set as ±20 ppm and 0.6 Da, respectively. Trypsin was set as the enzyme with four maximum missed cleavages allowed. Cysteine carbamidomethylation was set as a constant modification. A maximum of four variable modifications were also allowed, including protein N-terminal acetylation, methionine oxidation, and N-terminal conversion of glutamine to pyroglutamic acid. In addition, three defined modifications representing cross-linker fragment moieties on aspartic acid and glutamic acid were selected: alkene (A, C3H4N2, +68 Da), sulfenic

acid (S, C3H6N2SO, +118 Da), and unsaturated thiol (T, C3H4N2S, +100 Da) modifications. Initial acceptance criteria for peptide identification required a reported expectation value ≤ 0.1. The in-house program XL-Discoverer, a revised version of previously developed Link-Hunter, was used to validate and summarize cross-linked peptides based on MSn data and database searching [60].

Appendix B

Chapter 2 Experimental Procedures

Materials and Reagents

General chemicals were purchased from Fisher Scientific or VWR International. Bovine serum albumin (≥96% purity) was purchased from Sigma-Aldrich. Ac-LR9 peptide (Ac-LADVCAHER, 98% purity) was custom ordered from Biomatik (Wilmington, DE).

Synthesis and Characterization of BMSO

BMSO was synthesized as described in Figure 1.1D. Disuccinimidyl sulfoxide (DSSO) and the trifluoroacetate salt of 1-(2-aminoethyl) maleimide were synthesized as previously published [60,136] To a cooled (0 °C) solution of disuccinimidyl sulfoxide (468 mg, 1.21 mmol) and the maleimide salt (674 mg, 2.65 mmol) in H_2O (12 mL) was added 1M aq. $NaHCO_3$ (3.6 mL). After stirring for 12 h while allowing the reaction vessel to reach room temperature, the mixture was concentrated *in vacuo*. The crude material was then purified with column chromatography (30% MeOH in CH_2Cl_2) to afford BMSO as a colorless solid (251 mg, 47%): mp 90—98 °C; ^{1}H-NMR (600 MHz, $CDCl_3$) δ 6.76 (t, J = 5.5 Hz), 6.73 (s, 4H), 3.66—3.70 (app t, J = 5.7 Hz, 4H), 3.39—3.49 (m, 4H), 3.05—3.10 (m, 2H), 2.87—2.91 (m, 2H), 2.61 (app t, J = 7.1 Hz, 4H); ^{13}C-NMR (150 MHz, $CDCl_3$) δ170.92, 170.35, 134.2, 46.9, 38.7, 37.5, 28.8; IR: 3307, 3086, 1693, 1645, 1543, 1173, 1024 cm^{-1}; HRMS (ESI) *m/z* $[M + Na]^+$ Calcd for $C_{18}H_{22}N_4O_7SNa$ 461.1107; Found 461.1119.

BMSO Cross-linking of Synthetic Peptides

Synthetic peptide Ac-LR9 was dissolved in DMSO to 1 mM and cross-linked with BMSO in a 1:1 molar ratio of peptide to cross-linker. The resulting samples were diluted to 10 pmol/μL in 3% ACN/2% formic acid prior to MS^n analysis.

Preparation of BMSO Cross-linked Bovine Serum Albumin

50 μL of 50 μM BSA in PBS buffer (pH 7.4) was reacted with BMSO in molar ratios of 1:50 and 1:100. The cross-linking reaction was initiated by adding BMSO to protein solutions, reacted for 2 h at 37°C. Cross-linked protein samples were subjected to SDS-PAGE followed by in-gel digestion prior to MS^n analysis [60].

Digestion of BMSO Cross-linked Proteins

Cross-linked BSA was separated by SDS-PAGE and visualized by Coomassie blue staining. Selected cross-linked gel bands were excised and then digested with trypsin at 37 °C overnight. Peptide digests were extracted and concentrated and reconstituted in 25 mM ammonium bicarbonate and incubated at 37 °C overnight. The ammonium bicarbonate was then evaporated, and the peptides were reconstituted in 3% ACN/2% formic acid for MS^n analysis.

Liquid Chromatography-Multistage Tandem Mass Spectrometry (LC-MS^n) Analysis

BMSO cross-linked Ac-LR9 was analyzed by LC-MS^n utilizing an Easy-nLC 1000 (Thermo Fisher, San Jose, CA) coupled on-line to an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher, San Jose, CA) [60]. BMSO cross-linked peptides of BSA were analyzed by LC-MS^n utilizing a Dionex UltiMate™ 3000 (Thermo Fisher, San Jose, CA) coupled on-line to an Orbitrap Fusion™ Lumos™ mass spectrometer (Thermo Fisher, San Jose, CA). LC-MS^n data extraction and database searching for the identification of BMSO cross-linked peptides were performed similarly as previously described [60,75] A 15 cm x 75 μm Acclaim™ PepMap™ C18 Column was used to separate peptides over acetonitrile gradients of 1% to 25% at a flow rate of 300 nL/min. Two different types of acquisition methods were utilized to maximize the identification of BMSO cross-linked peptides: 1) top 4 data-dependent MS^3 and 2) targeted MS^3 acquisition optimized for

capturing DSSO cross-linked peptides by utilizing the mass difference between alkene- and thiol-modified ion pairs (31.9721 Da) [163].

Data Analysis and Identification of BMSO Cross-linked Peptides

MS^n data extraction and analysis were performed in the same way as previously described [60]. MS^3 data was subjected to a developmental version of Protein Prospector (v.5.19.1) for database searching, using Batch-Tag against SwissProt.2016.5.9.random.concat database limited to *Bos taurus* taxonomy (5998 entries) with mass tolerances for parent ions and fragment ions set as ±20 ppm and 0.6 Da, respectively. Trypsin was set as the enzyme with three maximum missed cleavages allowed. A maximum of four variable modifications were also allowed, including protein N-terminal acetylation, cysteine carbamidomethylation, methionine oxidation, and N-terminal conversion of glutamine to pyroglutamic acid. In addition, five defined modifications representing cross-linker fragment moieties on cysteine residues were selected: alkene closed-ring (Ac, C9H10N2O3, +194.0691 Da), alkene open-ring (Ao, C9H12N2O4, +212.0797 Da), sulfenic acid closed-ring/unsaturated thiol open-ring (Sc/To, C9H12N2O4S1, +244.0518 Da), sulfenic acid open-ring (So, C9H14N2O5S1, +262.0623 Da), and unsaturated thiol closed-ring (Tc, C9H10N2O3S1, +226.0412 Da) modifications. Initial acceptance criteria for peptide identification at the MS^3 level required a reported expectation value ≤ 0.15, which yielded a false discovery rate of 0.5%. The in-house program XL-Discoverer, a revised version of previously developed Xl-discoverer, was used to validate and summarize cross-linked peptides based on MS^n data and database searching [60,75]. Integration of MS^1, MS^2, and MS^3 spectral data identified 41 unique BMSO inter-linked BSA peptides (Table 2.1). Following integration of MS^n data, no cross-links were identified involving decoy proteins.

Appendix C

Chapter 3 Experimental Procedures

Materials and Reagents

All chemicals were purchased from Aldrich, Acros Organics, Alfa Aesar, TCI, VWR Internationa or Fisher Scientific and used without further purification. Deuterated solvents were purchased from Cambridge Isotope Laboratories. Solvents were purchased as ACS grade or better and as HPLC-grade and passed through a solvent purification system equipped with activated alumina columns prior to use.

Synthesis and Characterization of SDASO Cross-linkers

Three SDASO cross-linkers were designed, synthesized and analyzed in this work (Fig. 3.1), including SDASO-L, SDASO-M and SDASO-S. Their synthesis and characterization are described as follows. All reactions were carried out in oven-dried glassware, under an atmosphere of argon unless otherwise noted. Reactions were monitored by thin layer chromatography (TLC) or electrospray ionization mass spectrometry (ESI-MS). Thin layer chromatography (TLC) was carried out using glass plates coated with a 250 μm layer of 60 Å silica gel. TLC plates were visualized with a UV lamp at 254 nm, or by staining with potassium permanganate, cerium molybdate, or ninhydrin. ESI-MS was analyzed in positive mode with flow injection. Liquid chromatography was performed using a Teledyne ISCO CombiFlash® forced flow with an automated purification system on prepacked silica gel (SiO_2) columns or prepacked C18 columns. Proton NMRs were recorded at 500 MHz or 600 MHz using either a Bruker DRX500 (cryoprobe) or a Bruker AVANCE600 (cryoprobe) NMR, respectively. Carbon NMRs were recorded at 126 MHz or 151 MHz on the Bruker DRX500 or Bruker AVANCE600 NMR, respectively. All NMRs were taken at 25 °C. Chemical shifts (δ) are reported in parts per million (ppm) and referenced to

residual solvent peak at 7.26 ppm (^{1}H) or 77.16 ppm (^{13}C) for deuterated chloroform ($CDCl_3$), 2.50 ppm (^{1}H) or 39.52 ppm (^{13}C) for deuterated dimethyl sulfoxide (DMSO-d_6), 3.31 ppm (^{1}H) or 49.00 (^{13}C) for deuterated methanol (CD_3OD). NMR data are reported as the following: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad), coupling constants (J) in hertz (Hz), and integration. High-resolution mass spectrometry (HRMS) was performed using a Waters LCT Premier TOF spectrometer with ESI source.

1. SDASO-S

2-(3-Methyl-3H-diazirin-3-yl)ethyl 4-methylbenzenesulfonate (S1)

As shown in Fig. 3.2A, the staring material, 4-hydroxy-2-butanone (3.0 mL, 34.5 mmol, 1.0 equiv), was cooled to 0 °C and then 7N NH3 in MeOH (35 mL) was added dropwise. After stirring for 3 h at 0 °C, a solution of hydroxylamine-O-sulfonic acid (H2NOSO3H; 4.25 g, 37.5 mmol, 1.1 equiv) in MeOH (15 mL) was added dropwise then solution was allowed to warm to rt. After stirring for 17 h the reaction solution was concentrated in vacuo. Next, NEt3 (7.5 mL, 53.8, 1.6 equiv) was added to a solution of the residue in MeOH (25 mL) at 0 °C. After stirring for 10 min at 0 °C, iodine was added in aliquots until a dark brown color persisted in the solution. The reaction solution was diluted in EtOAc, washed with 1.0 M aq. HCl, sat. aq. Na2S2O3 solution, and brine. Finally, the combined organic phase was dried over MgSO4 and concentrated in vacuo. The resulting residue was subject to the next reaction without further purification.

This crude product was dissolved in pyridine (10.0 mL, 124 mmol, 3.6 equiv) and cooled to 0 °C. Next, p-toluenesulfonyl chloride (TsCl; 1.86 g, 10.0 mmol, 0.3 equiv) was added to the solution. After stirring for 1 h at rt, the reaction solution was diluted with EtOAc, washed with 1N HCl, sat. aq. NaHCO3 solution, and brine. The combined organic phase was then dried over MgSO4 and concentrated in vacuo. The resulting residue was purified by flash chromatography

(15% EtOAc in hexanes) to obtain tosylate S1 (971 mg, 11% over two steps). S1: 1H NMR (500 MHz, CDCl3): δ 7.82 (d, J = 8.5 Hz, 2H), 7.36 (d, J = 8.5 Hz, 2H), 3.96 (t, J = 6.4 Hz, 2H), 2.46 (s, 3H), 1.67 (t, J = 6.4 Hz, 2H), 1.00 (s, 3H); 13C NMR (126 MHz, CDCl3): δ 145.2, 132.9, 130.1, 128.1, 65.2, 34.3, 23.5, 21.8, 19.9. Spectral data were consistent with those previously reported for the compound [164].

Methyl 3-((2-(3-methyl-3H-diazirin-3-yl)ethyl)thio)propanoate (S2)

As shown in Fig. 3.2A, K2CO3 (477 mg, 3.5 mmol, 1.5 equiv) was added to a solution of tosylate S1 (579 mg, 2.3 mmol, 1.0 equiv,) and methyl-3-mercaptopropionate (0.50 mL, 4.6 mmol, 2 equiv) in MeOH (2.3 mL). After stirring for 3 h at rt, the reaction solution was filtered to remove solids. The filtrate was concentrated in vacuo and the resulting residue was purified by flash chromatography (20% EtOAc in hexanes) to obtain ester S2 (408 mg, 88%). S2: 1H NMR (500 MHz, CDCl3): δ 3.70 (s, 3H), 2.73 (t, J = 7.4 Hz, 2H), 2.56 (t, J = 7.4 Hz, 2H), 2.35 (t, J = 8.0 Hz, 2H), 1.61 (t, J = 8.0 Hz, 2H), 1.02 (s, 3H); 13C NMR (126 MHz, CDCl3): δ 172.3, 51.9, 34.64, 34.61, 27.1, 26.6, 25.3, 19.9; HRMS (ESI, MeOH) m/z: [M+Na]+ calcd for C8H14N2O2SNa 225.0669, found 225.0677.

3-((2-(3-Methyl-3H-diazirin-3-yl)ethyl)thio)propanoic acid (S3)

As shown in Fig. 3.2A, LiOH·H2O (79 mg, 1.9 mmol, 1.0 equiv) was added to a solution of ester S2 (382 mg, 1.9 mmol, 1.0 equiv) in THF:H2O (4:1, 12.7 mL). After stirring at rt for 1 h, an additional aliquot of LiOH·H2O (79 mg, 1.9 mmol, 1.0 equiv) was added. After stirring for an additional 2 h at rt, the reaction mixture was partitioned between hexanes and H2O. The aqueous phase was acidified to pH = 1 with 1N HCl and extracted with EtOAc five times. The combined organic layers were dried with MgSO4, and concentrated in vacuo. The crude product (acid S3) was subjected to the next step without further purification.

2,5-Dioxopyrrolidin-1-yl 3-((2-(3-methyl-3H-diazirin-3-yl)ethyl)thio)propanoate (S4)

As shown in Fig. 3.2A, N-hydroxysuccinimide (NHS-H; 219 mg, 1.9 mmol, 1.0 equiv) was added to a solution of acid S3 (358 mg, 1.9 mmol, 1.0 equiv) in CH2Cl2 (4.0 mL) followed by an addition of 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC·HCl; 401 mg, 2.1 mmol, 1.1 equiv). After stirring at rt for 16 h, the reaction solution was diluted in CH2Cl2, washed with H2O, then brine. The combined organic phase was dried with MgSO4 and concentrated in vacuo. The resulting residue was purified by flash chromatography (50% EtOAc in hexanes) to obtain NHS ester S4 (308 mg containing 5 wt% EtOAc by 1H NMR, 292 mg, 54% over two steps). S4: 1H NMR (500 MHz, CDCl3): δ 2.93–2.78 (m, 8H), 2.42 (t, J = 8.0, 2H), 1.65 (t, J = 8.0, 2H), 1.05 (s, 3H); 13C NMR (126 MHz, CDCl3): δ 169.0, 167.2, 34.6, 32.1, 26.9, 26.6, 25.7, 25.3, 19.9.

2,5-Dioxopyrrolidin-1-yl 3-((2-(3-methyl-3H-diazirin-3-yl)ethyl)sulfinyl)propanoate (SDASO-S)

Finally, as shown in Fig. 3.2A, 30% aq. H2O2 (72 μL, 0.70 mmol, 2.0 equiv) was added to a solution of NHS ester S4 (100 mg, 0.35 mmol, 1.0 equiv) in 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP; 0.9 mL) was added. After stirring at rt for 20 min, the reaction was quenched with dimethyl sulfide (DMS; 0.15 mL) and the mixture was allowed to stir for an additional 10 min. The reaction mixture was concentrated in vacuo, then partitioned between CHCl3 and H2O and the aqueous phase was extracted CHCl3. The combined organic phase was washed with H2O and the combined aqueous phase was again extracted with CHCl3. The combined organic phase was washed with brine, dried with MgSO4, and concentrated in vacuo to obtain SDASO-S as a white solid (75 mg, 93% purity by 1H NMR, 70 mg, 66%). SDASO-S: 1H NMR (500 MHz, CDCl3): δ 3.21–3.05 (m, 3H), 3.01–2.91 (m, 1H), 2.84 (s, 4H), 2.65–2.56 (m, 1H), 2.56–2.47 (m, 1H), 1.95–1.79 (m,

2H), 1.09 (s, 3H); 13C NMR (126 MHz, CDCl3): δ 168.8, 167.1, 47.0, 46.2, 28.0, 25.7, 24.9, 24.2, 19.8; HRMS (ESI, MeOH) m/z: [M+Na]+ calcd for C11H15N3O5SNa 324.0625, found 324.0636.

2. SDASO-M

3-(3-Methyl-3H-diazirin-3-yl)propyl 4-methylbenzenesulfonate (S5)

As shown in Fig. 3.2B, the staring material, 5-hydroxy-2-pentanone (1.5 mL, 14.6 mmol, 1.0 equiv), was cooled to 0 ℃ and 7N NH3 in MeOH (15 mL) was added dropwise. After stirring for 3 h at 0 ℃, a solution of H2NOSO3H (1.90 g, 16.8 mmol, 1.15 equiv) in MeOH (12.4 mL) was added dropwise and the solution was allowed to warm to rt. After stirring for 17.5 h the reaction solution was concentrated in vacuo. Next, NEt3 (3.3 mL, 23.4, 1.6 equiv) was added to a solution of the residue in MeOH (15 mL) at 0 ℃. After stirring for 10 min at 0 ℃, iodine was added in aliquots until a dark brown color persisted in the solution. The reaction solution was diluted in EtOAc, washed with 1.0 M aq. HCl, sat. aq. Na2S2O3 solution, and brine. The combined organic phase was dried over MgSO4 and concentrated in vacuo. The resulting residue was subject to the next reaction without further purification.

This crude product was dissolved in pyridine (4.2 mL, 52.6 mmol, 3.6 equiv) and cooled to 0 °C. Next, TsCl (835 mg, 4.38 mmol, 0.3 equiv) was added to the solution. After stirring for 1 h at rt, the reaction solution was diluted with EtOAc, washed with 1N HCl, sat. aq. NaHCO3 solution, and brine. The combined organic phase was dried over MgSO4 and concentrated in vacuo. The resulting residue was purified by flash chromatography (0% to 15% EtOAc in hexanes) to obtain tosylate S5 (102 mg, 3% over two steps). S5: 1H NMR (500 MHz, CDCl3): δ 7.75 (d, J = 8.2 Hz, 2H), 7.33 (d, J = 8.0 Hz, 2H), 3.97 (t, J = 6.2 Hz, 2H), 2.43 (s, 3H), 1.55–1.42 (m, 2H), 1.42–1.31 (m, 2H), 0.95 (s, 3H); 13C NMR (126 MHz, CDCl3): δ 145.0, 133.0, 130.0, 127.9, 69.5, 30.4, 25.1, 23.6, 21.7, 19.8.

Methyl 3-((3-(3-methyl-3H-diazirin-3-yl)propyl)thio)propanoate (S6)

As shown in Fig. 3.2B, K2CO3 (79 mg, 0.57 mmol, 1.5 equiv) was added to a solution of tosylate S5 (102 mg, 0.38 mmol, 1.0 equiv,) and methyl-3-mercaptopropionate (84 μL, 0.76 mmol, 2 equiv) in MeOH (0.4 mL). After stirring for 3 h at rt, the reaction mixture was filtered to remove solids. The filtrate was concentrated in vacuo and the resulting residue was purified by flash chromatography (0% to 20% EtOAc in hexanes) to obtain ester S6 (95 mg, containing 6 wt% of EtOAc and 17 wt% methyl-3-mercaptopropionate by 1H NMR; 73 mg, 89%). S6: 1H NMR (500 MHz, CDCl3): δ 3.69 (s, 3H), 2.81–2.72 (m, 2H), 2.59 (t, J = 7.3 Hz, 2H), 2.53–2.45 (m, 2H), 1.51–1.41 (m, 2H), 1.00 (s, 3H).

3-((3-(3-Methyl-3H-diazirin-3-yl)propyl)thio)propanoic acid (S7)

As shown in Fig. 3.2B, LiOH·H2O (16 mg, 0.38 mmol, 1.05 equiv) was added to a solution of ester S6 (73 mg, 0.36 mmol, 1.0 equiv) in THF:H2O (4:1, 2.5 mL). After stirring at rt for 1 h, an additional aliquot of LiOH·H2O (16 mg, 0.38 mmol, 1.05 equiv) was added. After stirring for an additional 2 h at rt, the reaction mixture was partitioned between hexanes and H2O. The aqueous phase was acidified to pH = 1 with 1N HCl and extracted with EtOAc five times. The combined organic layers were dried with MgSO4, and concentrated in vacuo. The crude product (acid S7) was subjected to the next step without further purification.

2,5-Dioxopyrrolidin-1-yl 3-((3-(3-methyl-3H-diazirin-3-yl)propyl)thio)propanoate (S8)

As shown in Fig. 3.2B, NHS-H (40 mg, 0.35 mmol, 1.0 equiv) was added to a solution of crude acid S7 (70 mg, 0.35 mmol, 1.0 equiv) in CH2Cl2 (0.7 mL) followed by an addition of EDC·HCl (74 mg, 0.39 mmol, 1.1 equiv). After stirring at rt for 15 h, the reaction solution was diluted in CH2Cl2, washed with H2O, then brine. The combined organic phase was dried with MgSO4 and concentrated in vacuo to obtain crude NHS ester S8 (56 mg). S8: 1H NMR (500 MHz,

CDCl3): δ 2.95–2.76 (m, 8H), 2.54–2.41 (m, 2H), 1.48–1.37 (m, 4H), 0.99 (s, 3H); 13C NMR (126 MHz, CDCl3): δ 169.1, 167.2, 33.3, 32.1, 31.6, 26.3, 25.7, 25.5, 24.0, 19.9.

2,5-Dioxopyrrolidin-1-yl3-((3-(3-methyl-3H-diazirin-3-yl)propyl)sulfinyl)propanoate (SDASO-M)

Finally, as shown in Fig. 3.2B, 30% aq. H2O2 (38 μL, 0.37 mmol, 2.0 equiv) was added to a solution of crude NHS ester S8 (56 mg, 0.19 mmol, 1.0 equiv) in HFIP (0.5 mL). After stirring at rt for 20 min, the reaction was quenched with dimethyl sulfide (DMS; 80 μL) and the solution was allowed to stir for an additional 10 min. The reaction mixture was concentrated in vacuo, then partitioned between CHCl3 and H2O and the aqueous phase was extracted CHCl3. The combined organic phase was washed with H2O and the combined aqueous phase was again extracted with CHCl3. The combined organic phase was washed with brine, dried with MgSO4, and concentrated in vacuo to obtain crude SDASO-M (15 mg,). SDASO-M: 1H NMR (500 MHz, CDCl3): δ 3.23¬¬–2.92 (m, 4H), 2.85 (s, 4H), 2.77–2.57 (m, 2H), 1.82–1.62 (m, 2H), 1.62–1.43 (m, 2H), 1.04 (s, 3H); 13C NMR (126 MHz, CDCl3): δ 168.8, 167.2, 51.7, 46.0, 33.4, 25.72, 25.69, 24.2, 19.7, 17.8.

3. SDASO-L

3-(3-Methyl-3H-diazirin-3-yl)propanoic acid (S9)

As shown in Fig. 3.2C, 7N NH3 in MeOH (14.8 mL) was added to a solution of levulinic acid (2.0 mL, 19.5 mmol, 1.0 equiv) in MeOH (4.5 mL, 4.30 M) cooled to 0 °C. After stirring for 3 h at 0 °C, a solution of H2NOSO3H (2.54 g, 22.4 mmol, 1.15 equiv) in MeOH (13 mL) was added and the reaction was allowed to warm to rt. The milky white solution was concentrated in vacuo after stirring for 19 h. Next, NEt3 (4.7 mL, 33.6 mmol, 1.5 equiv) was added to a solution of the white oil in MeOH (14 mL) at 0 °C. After stirring for 10 min at 0 °C, iodine was added in

aliquots (10.2 g, 40.3 mmol, 1.8 equiv) until a dark brown color persisted in the solution. The brown mixture was diluted in EtOAc, washed with 1.0 M aq. HCl, sat. aq. Na2S2O3 solution, and brine. The aqueous phases were combined and extracted three times with EtOAc. The combined organic phase was dried over Na2SO4 and concentrated in vacuo to obtain a yellow solid which was subject to the next reaction without further purification. Spectral data were consistent with those previously reported for the compound [165].

2,5-Dioxopyrrolidin-1-yl 3-(3-methyl-3H-diazirin-3-yl)propanoate (S10)

As shown in Fig. 3.2C, trifluoroacetic anhydride (TFAA; 4.1 mL, 29.4 mmol, 2.0 equiv) was added dropwise to a solution of acid S9 (1.88 g, 14.7 mmol, 1.0 equiv), NHS-H (3.38 g, 29.4 mmol, 2.0 equiv), and N,N-diisopropylethylamine (DIPEA; 10.2 mL, 58.7 mmol, 4.0 equiv) in DMF (75.0 mL, 0.20 M) cooled to 0 °C. The orange solution was stirred for 3 h at 0 °C, and then partitioned between EtOAc and brine. The organic phase was washed with brine five times, dried over Na2SO4, and concentrated in vacuo to obtain an orange oil. The oil was purified by flash chromatography (20 to 70% EtOAc in hexanes) to obtain NHS ester S10 as a tan powder (1.65 g, 50% over two steps). Spectral data were consistent with those previously reported for the compound [164].

N-(2-mercaptoethyl)-3-(3-methyl-3H-diazirin-3-yl)propenamide (S11)

As shown in Fig. 3.2C, NHS ester S10 (1.00 g, 4.44 mmol, 1.0 equiv) was added to a solution of 2-aminoethanethiol (343 mg, 4.44 mmol, 1.0 equiv) in CH2Cl2 (22.2 mL, 0.20 M). After stirring for 20 min at rt, the reaction solution was concentrated in vacuo. The crude product (thiol S11) was subject to the next step without further purification.

2,5-Dioxopyrrolidin-1-yl acrylate (S12)

As shown in Fig. 3.2C, NEt3 (6.30 mL, 45.0 mmol, 1.0 equiv) was added to a solution of NHS-H (5.18 g, 45.0 mmol, 1.0 equiv) in CH2Cl2 (90 mL, 0.50 M) cooled to 0 °C followed by an addition of acryloyl chloride (4.0 mL, 49.5 mmol, 1.1 equiv). The reaction was stirred for 3 h at 0 °C, and then vacuum filtered to collect the white precipitate. The filtrate was washed with H2O then brine, dried over Na2SO4, and concentrated in vacuo to obtain NHS ester S12 as a white solid (5.52 g, 72%). Spectral data were consistent with those previously reported for the compound [166].

2,5-Dioxopyrrolidin-1-yl3-((2-(3-(3-methyl-3H-diazirin-3-l) propanamido)ethyl)thio)propanoate (S13)

As shown in Fig. 3.2C, a solution of thiol S11 (832 mg, 4.44 mmol, 1.0 equiv) in CH2Cl2 (22.2 mL, 0.20 M) was added to NHS ester S12 (751 mg, 4.44 mmol, 1.0 equiv) followed by an addition of NEt3 (0.680 mL, 4.88 mmol, 1.1 equiv). After stirring at rt for 20 min, the reaction solution was washed with H2O then brine, dried with Na2SO4, and concentered in vacuo. The crude mixture was purified by flash chromatography (20 to 100% EtOAc in hexanes) to obtain sulfide S13 (887 mg, 56% over two steps). S13: 1H NMR (500 MHz, CDCl3): δ 6.32 (s, br, 1H), 3.42 (q, J = 6.0 Hz, 2H), 2.96–2.80 (m, 8H), 2.70 (t, J = 6.0 Hz, 2H), 2.00 (t, J = 7.7 Hz, 2H), 1.72 (t, J = 7.7 Hz, 2H), 1.00 (s, 3H); 13C NMR (125 MHz, CDCl3): δ 171.7, 169.2, 167.3, 38.4, 32.4, 32.2, 30.6, 30.1, 26.5, 25.7, 25.6, 20.0; HRMS (ESI, MeOH) m/z calcd for C14H20N4O5SNa (M+Na)+ 379.1047, found 379.1043.

2,5-Dioxopyrrolidin-1-yl3-((2-(3-(3-methyl-3H-diazirin-3-yl)propanamido)ethyl)sulfinyl) propanoate (SDASO-L)

Finally, as shown in Fig. 3.2C, 30% aq. H2O2 solution (0.11 mL, 1.11 mmol, 2.0 equiv) was added to a solution of sulfide S12 (200 mg, 0.561 mmol, 1.0 equiv) in HFIP (2.80 mL, 0.20 M). The reaction was stirred for 10 min at rt, and then quenched with DMS (0.20 mL) and the

mixture was concentrated in vacuo to obtain the cross linker SDASO (247 mg containing 23 wt% HFIP and 4 wt% DMSO; 181 mg, 87%). SDASO: 1H NMR (600 MHz, CDCl3): δ 6.77 (s, br, 1H), 3.85–3.65 (m, 2H), 3.20–3.04 (m, 4H), 2.95–2.80 (m, 4H), 2.63 (s, 4H), 2.04 (t, J = 7.6 Hz, 2H), 1.73 (t, J = 7.6 Hz, 2H), 1.01 (s, 3H); 13C NMR (151 MHz, CDCl3): δ 172.6, 168.9, 167.0, 51.3, 45.9, 34.4, 30.6, 30.0, 25.7, 25.6, 24.3, 19.9; HRMS (ESI, MeOH) m/z calcd for C14H20N4O6SNa (M+Na)+ 395.0996, found 395.0995.

Purification of the Yeast 26S Proteasome Complexes

RPN11-TAP strain was used for yeast proteasome purification as previously described [167]. The Yeast Strain was cultured in YEPD medium (1% yeast extract, 2% peptone, and 2% dextrose) at 30°C until stationary phase, then the cells were collected and washed with ice-cold water. The yeast cells were loaded into 20ml syringe and pushed into liquid nitrogen to get yeast frozen "noodles" which were ground in a Cryomill into frozen powder. The Yeast frozen powder was resuspended in a lysis buffer containing 50 mM Sodium phosphate (pH 8.0), 5 mM magnesium chloride, 1 mM ATP, 1× protease inhibitor (Sigma), then sonicated at 15 watts with 30s on and 30s off for three cycles and pellet was spun down at 15,000rpm for 15min. The supernatant was bound to IgG resin (MP Biomedical#55961) for 2 hours at 4^{o}C with rotation. The IgG resin was washed with 50 bed volume of wash buffer (50 mM Sodium phosphate (pH 7.4), 100mM NaCl, 5 mM magnesium chloride, 1 mM ATP, then 20 bed volume of TEV cleavage buffer (50 mM Sodium phosphate (pH 7.4), 10% glycerol, 1mM EDTA, 1mM DTT and 1 mM ATP and the bound proteasome was cleaved overnight with TEV at 4^{o}C. The resulting proteasome was concentrated with 30K cutoff Centricon devices (EMD Millipore UFC903024).

XL-MS Analysis of BSA and 26S Proteasomes

Protein cross-linking was performed similarly to previous studies with some modifications [66,97]. Briefly, for SDASO cross-linking of BSA, 50 µL of 50 µM protein solution in PBS buffer (pH 7.4) was reacted in triplicate with SDASO-L, SDASO-M or SDASO-S in molar ratio of 1:50, respectively, for 1 h at 25°C in the dark. The NHS reactive ends were quenched with the addition of ammonium bicarbonate at a 50-fold excess for 10 minutes at 25°C in the dark. Then NHS ester labeled proteins were transferred into Millipore Microcon Ultracel PL-30 (30-kDa filters) and washed 3 times with 300 µL PBS buffer. Diazirine cross-linking was activated by UV irradiation, which was carried out on ice ~5cm from the light source in an UV light chamber (Analytikjena™ UVP Cross-linker CL-1000L) and irradiated at 365 nm for 30 mins.

The yeast 26S proteasome was affinity purified as described [167] (Supplementary Methods) and cross-linked by SDASO linkers similarly as described above. Specifically, 100 µg of the 26S proteasome in PBS buffer (pH 7.4) was cross-linked in triplicate with 20 mM SDASO-L, SDASO-M and SDASO-S, respectively. In addition, 100 µg of the yeast 26S proteasome in PBS buffer (pH 7.4) was cross-linked with 2.5 mM or 5 mM DSSO for 1 hr at 25°C temp similarly as described [76], and the reactions were quenched with the addition of ammonium bicarbonate at a 50-fold excess for 10 minutes. Then cross-linked proteins were transferred into Millipore Microcon Ultracel PL-30 (30-kDa filters) for digestion.

Digestion of Cross-linked Proteins

The resulting cross-linked products were subjected to enzymatic digestion using a FASP protocol [168]. Briefly, cross-linked proteins on the FASP filters were reduced/alkylated, and digested with Lys-C/trypsin or chymotrypsin as described [60,76]. The resulting digests were desalted and cross-linked peptides were enriched by size-exclusion chromatography prior to LC MS^n analysis [43,65].

LC-MS^n Analysis and Identification of Cross-linked Peptides

Cross-linked peptides were analyzed by LC-MS^n using a Thermo Scientific™ Dionex UltiMate 3000 system online coupled with an Orbitrap Fusion Lumos™ mass spectrometer [76]. A 50 cm x 75 μm Acclaim™ PepMap™ C18 column was used to separate peptides over a gradient of 1% to 25% ACN in 106 mins for BSA, and in 166 mins for the 26S proteasome at a flow rate of 300 nL/min. MS^1 scans (375-1500 m/z, resolution at 120,000) was performed with the AGC target set to 4e5 in top speed mode with a cycle time of 5 s. For MS^n analysis, 3+ and up charged ions were selected for MS^2-CID in FT mode, followed by top 4 data-dependent MS^3 acquisition method [163]. A targeted MS^3 acquisition was also used for DSSO cross-linked peptides by utilizing the mass difference between alkene- and thiol-modified ion pairs (31.9721 Da) [163]. For MS^2 scans, the resolution was set to 30,000, the AGC target 5e4, the precursor isolation width was 1.6 m/z, and the maximum injection time was 100 ms for CID. The CID-MS^2 normalized collision energy was 25%. For MS^3 scans, CID was used with a collision energy of 35%, the AGC target was set to 2e4, and the maximum injection time was set to 120 ms. MS^n data were extracted using MSConvert (ProteoWizard 3.0.10738) and subjected to database searching using a developmental version of Protein Prospector (v.6.0.0). Cross-linked peptides were identified by the integration of MS^n data with database search results using the in-house software xl-Tools [163]

Identification of Cross-linked Peptides

MS^n data extraction and analysis were performed in the same way as previously described [76]. MS^3 data was subjected to a developmental version of Protein Prospector (v.6.0.0) for database searching, using Batch-Tag against limited to custom random concatenated databases. For BSA and the 26S proteasome data, the database consisted of BSA plus 493 *Saccharomyces cerevisiae* proteins (494 entries) with mass tolerances for parent ions and fragment ions set as ±20 ppm and

0.6 Da, respectively. Trypsin or Chymotrypsin was set as the enzyme with three or four maximum missed cleavages allowed, respectively. A maximum of four variable modifications were allowed, including cysteine carbamidomethylation, protein N-terminal acetylation, methionine oxidation, and N-terminal conversion of glutamine to pyroglutamic acid. In addition, three defined modifications representing alkene on uncleaved lysines, thiol and sulfenic fragment moieties on any AAs were selected for each respective SDASO cross-linker. Specifically, for SDASO-L cross-links: alkene (C_3H_2O; +54 Da), sulfenic acid ($C_7H_{13}NO_2S$; +175 Da), and thiol ($C_7H_{11}NOS$; +157 Da). For SDASO-M cross-links: alkene (C_3H_2O; +54 Da), sulfenic acid ($C_5H_{10}OS$; +118 Da), and thiol (C_5H_8S; +100 Da). For SDASO-S cross-links: alkene (C_3H_2O; +54 Da), sulfenic acid (C_4H_8OS; +104 Da), and thiol (C_4H_6S; +86 Da). For DSSO cross-links, three defined modifications on uncleaved lysines are: alkene (C_3H_2O; +54 Da), sulfenic acid ($C_3H_4O_2S$; +104 Da), and thiol (C_3H_2SO; +86 Da) [60]. Due to the conversion of the SDASO sulfenic acid moiety to the thiol moiety alongside backbone fragmentation during MS^3 analysis, we have incorporated such neutral loss in Batch-tag to facilitate the identification of sulfenic acid-modified peptides during database searching using Protein prospector. The in-house program xl-Tools was used to validate and summarize cross-linked peptides based on MS^n data and database searching [60,75] . Following integration of MS^n data, there were no decoy hits found in the final lists of identified cross-linked peptides for all XL-MS experiments except for the tryptic digests of SDASO-L cross-linked 26S proteasome with a FDR ≤ 0.08%. Raw data has been deposited at the PRIDE Archive proteomics data repository site: http://www.ebi.ac.uk/pride (Project accession: PXD022690, Username: reviewer_pxd022690@ebi.ac.uk, Password: LW6tdbel).

Analysis of the Identified Cross-links

Circular 2-D XL-maps were constructed using the CX-Cirus online application (http://cx-circos.cloudapp.net/) and linear 2-D XL-maps were created using the online application xiNET Crosslink Viewer (http://crosslinkviewer.org). 3-D maps were generated based on BSA (PDB: 4F5S), 26S proteasome structures (PDB:4CR2(s1), 4CR3(s2), 4CR4(s3), and 5MPD(s4)). The state-specific cross-links of the 26S proteasome were determined by mapping them onto s1-s4 state models, which are summarized in Table 3.9 (Note: blank spaces indicate linkage was not able to mapped to structure because o or more of the residues lacked mapped density within the structure).

Analysis of Amino Acid Preference for Diazirine Labeling

The unique K-X linkages identified for both BSA (Table 3.2) and 26S (Table 3.6) were used to determine diazirine labeling frequency at specific amino acids, in which only the peptide constituents labeled by diazirine were used for evaluation. The weighted occurrence values of diazirine labeled AAs were determined based on their localization precision, equal to 1 divided by the # of possible site locations (n) in the identified peptides. The resulting values for each AAs were summed to derive their labeling frequency by diazirine (Fig. 3.26).

Appendix D

Chapter 4 Experimental Procedures

Expression and Purification of CSN complexes

Eight of total nine subunits of human COP9 signalosome (CSN) complex, except CSN5, were over-expressed and purified from *E. coli*. Two three-subunit subcomplexes, CSN1-2-3 and CSN4-6-7, were prepared through co-expression. Briefly, CSN2 was subcloned into a modified pGEX4T1 (Amersham Biosciences) vector containing a glutathione S-transferase (GST) tag followed by a Tobacco etch virus (TEV) protease cleavage site, while both CSN1 and CSN3 were subcloned into a modified pET15b (Novagen) vector containing a chloramphenicol resistance cassette. After co-expression in BL21(DE3) (Novagen), the CSN1-2-3 formed a complex and was purified by glutathione-affinity chromatography. Following TEV cleavage, the CSN1-2-3 subcomplex was further purified by anion exchange and gel filtration chromatography. CSN4-6-7 was prepared in the same way. CSN8 and CSN9 were subcloned into the pGEX4T1 vector individually and subjected to the same purification procedure. Recombinant full length CSN5 inserted into a modified GTE vector (invitrogen). It has a GST tag which was removed during purification and was prepared from insect cells using baculovirus expression system. Two CSN complexes, with or without CSN9, were reconstituted by incubating the purified subcomplexes and individual subunits in equimolar ratio and polished by size exclusion chromatography. Neddylated Cul1-Rbx1 complex was prepared as described previously [163].

XL-MS Analysis of CSN Complexes

Affinity purified human CSN complex with or without CSN9 were cross-linked with DSSO, DHSO, or BMSO, respectively. Each CSN complex was reacted with a selected cross-linker at their optimized molar ratios (protein to linker) respectively: DSSO (1:250), BMSO

(1:400) and DHSO (1:30) [60,65,66]. DMTMM was used to activate acidic residues for DHSO cross-linking [65]. All reactions were performed for 1 h at room temperature. The resulting cross-linked proteins were digested by lys-C and trypsin. Cross-linked peptides were enriched by peptide SEC, analyzed by LC MS^n and identified through database searching as previously described [65,66].

LC MS^n Analysis of Cross-linked Peptides

LC MS^n analysis was performed using a Thermo Scientific™ Dionex UltiMate 3000 system online coupled with an Orbitrap Fusion Lumos™ mass spectrometer. A 50 cm x 75 μm Acclaim™ PepMap™ C18 column was used to separate peptides over a gradient of 1% to 25% ACN in 114 mins at a flow rate of 300 nL/min. Two different types of acquisition methods were utilized to maximize the identification of DSSO, DHSO and BMSO cross-linked peptides: (1) top 4 data-dependent MS^3 and (2) targeted MS^3 acquisition optimized for capturing DSSO, DHSO and BMSO cross-linked peptides by utilizing the mass difference between characteristic MS^2 fragment ions of DSSO cross-linked peptides (α−β) (i.e. $\Delta = \alpha_T - \alpha_A = \beta_T - \beta_A = 31.9721$ Da) [163].

Database Searching and Identification of Cross-linked Peptides

MS^n data extraction and analysis were performed in the same way as previously described[60]. MS^3 data was subjected to a developmental version of Protein Prospector (v.5.19.1) for database searching, using Batch-Tag against limited to custom random concatenated database consisting of CSN 1-9 (18 entries) with mass tolerances for parent ions and fragment ions set as ±20 ppm and 0.6 Da, respectively. Trypsin or Chymotrypsin was set as the enzyme with three or four maximum missed cleavages allowed, respectively. A maximum of four variable modifications were also allowed, including protein N-terminal acetylation, methionine oxidation, and N-terminal conversion of glutamine to pyroglutamic acid. Cysteine carbamidomethylation was selected as a

constant modification except when using BMSO. In addition, the previously defined modifications for either DSSO [60], DHSO [65] or BMSO [66] were also input into the search. For DSSO cross-links, three defined modification on uncleaved lysines were chosen, which included alkene (C_3H_2O; +54 Da), sulfenic acid ($C_3H_4O_2S$; +104 Da), and thiol (C_3H_2SO; +86 Da), representing cross-linker fragment moieties on lysine residues [60]. For DHSO cross-links, three defined modification were chosen, which included alkene ($C_3H_4N_2$; +68 Da), sulfenic acid ($C_3H_6N_2SO$; +118 Da), and thiol ($C_3H_4N_2S$; +100 Da), representing cross-linker fragment moieties on aspartic or glutamic acid residues [65]. For BMSO cross-links, five defined modifications representing cross-linker fragment moieties on cysteine residues were selected: alkene closed-ring ($C_9H_{10}N_2O_3$, +194.0691 Da), alkene open-ring ($C_9H_{12}N_2O_4$, +212.0797 Da), sulfenic acid closed-ring/unsaturated thiol open-ring ($C_9H_{12}N_2O_4S$, +244.0518 Da), sulfenic acid open-ring ($C_9H_{14}N_2O_5S$, +262.0623 Da), and unsaturated thiol closed-ring ($C_9H_{10}N_2O_3S$, +226.0412 Da) modifications on cysteine residues [66]. Initial acceptance criteria for peptide identification at the MS^3 level required a reported expectation value ≤ 0.18, which yielded a false discovery rate of 0.34%. The in-house program xl-Tools, a revised version of the previously developed Xl-Discoverer, was used to validate and summarize cross-linked peptides based on MS^n data and database searching [75,163]. Following integration of MS^n data, no cross-links were identified involving decoy proteins. Only cross-linked peptides that were identified in more than 60% biological replicates were reported (Table 4.10-4.17). Raw data has been deposited at the PRIDE Archive proteomics data repository site: https://www.ebi.ac.uk/pride/archive/ (reviewer86858@ebi.ac.uk)

PRM Targeted Quantitation of Cross-linked Peptides

341 PRM targets were obtained based on highly reproducible DHSO cross-linked peptides of CSN and CSNn complexes as summarized in Tables 4.10-4.17. For targeted analysis, the mass spectrometer was operated with the following settings: No survey scan collected, tMS2 resolving power 30,000, AGC target 5e4, maximum injection time 54 ms, isolation window 1 m/z, and CID normalized collision energy of 23%. A total of 341 cross-links were monitored over 3 separate targeted analyses for each sample, along with a set of 16 heavy-labeled AQUA peptides. Targeted analysis of AQUA peptides used the same settings as cross-link ions except were subjected to HCD with NCE of 30%. Transition lists based on expected cross-link fragmentation ions were generated and quantified using Skyline v.4.2.0.19072. Once exported, extracted intensities were normalized within sample sets using relative intensities of AQUA peptides based on quantified b and y ions.

In Vitro Deneddylation Assay

A mixture containing 5 μM Nedd8-Cul1-Rbx1 and 20 nM CSN was incubated in reaction buffer of 50 mM HEPES (pH 7.5), 150 mM NaCl and 1 mM TCEP. The reactions were carried out at room temperature and stopped by adding SDS-PAGE sample buffer at indicated time points, then analyzed by 9% SDS-PAGE and stained with Coomassie Blue.

Biochemical Validation of the CSN9 Interactors

Purified components of CSN, including CSN5, CSN8, subcomplex CSN1-2-3 and CSN4-6-7, were used for pull-down assay. His-GB1 fused CSN9 served as the bait protein. The prey samples (different combinations of CSN subunits) were mixed with His-GB1-CSN9 at molar ratio 2:1. After 10-minute incubation, His Mag Sepharose Ni beads (GE Healthcare) were added into the samples and suspended by gently tapping the sample tubes for 5 minutes to immobilize His-GB1-CSN9 and its binding partners. Then the beads were washed with 20 mM Tris-HCl (pH 8.0),

300 mM NaCl, 20 mM imidazole for 5 times. The beads were further eluted with 200 mM imidazole and the elution was analyzed on a 4-15% Mini-PROTEAN TGX Gel (Bio-RAD). To identify the binding partners of CSN9, all the purified CSN components were loaded on the same gel (lane 8 and 9).

Integrative Structure Modeling

Integrative structure modeling was carried out to determine the structures of the human canonical and non-canonical Cop9 signalosome complexes (Tables 4.18-4.23). All the relevant scripts, data, and results are available at https://salilab.org/CSN2019. The integrative structures of CSN and CSNn are deposited at PDB-dev (https://pdb-dev.wwpdb.org/) with accession code PDBDEV_00000037 and PDBDEV_00000038 respectively.

1. Integrative structure determination of the human canonical CSN complex

Integrative structure determination of the human canonical CSN complex proceeded through four stages (Fig. 4.7) [34,86,151,152,154,169]: (1) gathering, data, (2) representing subunits and translating data into spatial restraints, (3) configurational sampling to produce an ensemble of structures that satisfies the restraints, and (4) analyzing and validating the ensemble structures and data. The integrative structure modeling protocol (stage 2, 3, and 4) was scripted using the *Python Modeling Interface* (PMI) package, which is a library for modeling macromolecular assemblies based on our open-source *Integrative Modeling Platform* (IMP) package [152] version 2.8 (https://integrativemodeling.org). The current procedure is an updated version of previously described protocols [31,39,75,153,170-173]. Files containing the input data, scripts, and output results are available at https://salilab.org/CSN2019 as well as the nascent Protein Data Bank archive for integrative structures (https://pdb-dev.rcsb.rutgers.edu) accession code PDBDEV_00000037 (CSN) and accession code PDBDEV_00000038 (CSNn).

1.1 Integrative structure determination of the human canonical CSN complex

1.1.1 Stage 1: gathering data.

In total, 452 highly reproducible intra- and intermolecular DSSO, DHSO, and BMSO cross-links were identified using mass spectrometry (Table 4.10-4.13), which informed the spatial proximities among the 8 subunits of the CSN and their conformations.

Representations of individual subunits relied on (1) atomic structures of the CSN1-6 and CSN8 subunits determined by X-ray crystallography (PDB code 4D10)[124], and (2) comparative models of 2 domains of CSN7b built with MODELLER 9.17 [174] based on the known related structures of CSN7a [124].

1.2 Stage 2: representing subunits and translating data into spatial restraints

Information about the modeled system (above) can in general be used for defining the system's representation, defining the scoring function that guides sampling of alternative structural models, limiting sampling, filtering of representative structures obtained by sampling and final validation of the structures. Here, the CSN representation relies primarily on the atomic structure of CSN (PDB code 4D10). The CSN7b representation relies on comparative models for the following regions: (1) CSN7b$^{8\text{-}158}$ and CSN7b$^{163\text{-}212}$ (PDB code 4D10, sequence identity 60%, Z-DOPE score -0.90). The helical bundle comprising segments from each of the eight subunits was extracted and constrained based on the crystallographic structure. The scoring function relies on chemical cross-links, excluded volume, and sequence connectivity.

An optimal representation facilitates accurate formulation of spatial restraints as well as efficient and complete sampling representative solutions, while retaining sufficient detail without overfitting, so that the resulting models are maximally useful for subsequent biological analysis.

To maximize computational efficiency while avoiding using too coarse a representation, we represented the system in a multi-scale fashion. 16 rigid bodies consisting of multiple beads were defined for the comparative models of CSN7b and CSN (Fig. 4.8H). In a rigid body, the beads have their relative distances constrained during conformational sampling, whereas in a flexible string the beads are restrained by the sequence connectivity, excluded volume, and chemical cross-links [39,175,176].

Rigid bodies were coarse-grained using two resolutions, where beads represented either individual residues or segments of up to 10 residues. The coordinates of a 1-residue bead were those of the corresponding Cα atoms. The coordinates of a 10-residue bead were the center of mass of the 10 constituent 1-residue beads. Finally, the remaining regions without an atomic model were represented by a flexible string of beads encompassing up to 10 residues each.

With this representation in hand, we next encoded the spatial restraints into a scoring function based on the information gathered in Stage 1, as follows (Tables 4.10-4.13).

(1) *Cross-link restraints:* The 214 DSSO, 169 DHSO, and 69 BMSO cross-links (Tables 4.10-4.13) were used to construct the Bayesian scoring function [177] that restrained the distances spanned by the cross-linked residues [39].

(2) *Excluded volume restraints*: The excluded volume restraints were applied to each 10-residue bead, using the statistical relationship between the volume and the number of residues that it covered [34,176,178].

(3) *Sequence connectivity restraints*: We applied the sequence connectivity restraints, using a harmonic upper distance bound on the distance between consecutive beads in a subunit, with a threshold distance equal to twice the sum of the radii of the two connected beads. The bead radius

was calculated from the excluded volume of the corresponding bead, assuming standard protein density [39,86,175,176].

1.2 Stage 3: conformational sampling to produce an ensemble of structures that satisfies the restraints

We aimed to maximize the precision at which the sampling of solutions was exhaustive (Stage 4). We sampled the positions of the 16 rigid bodies of CSN. The search for representative models relied on Gibbs sampling, based on the Metropolis Monte Carlo algorithm. The initial positions of CSN were randomized. The Monte Carlo moves included random translation and rotation of rigid bodies (up to 1 Å and 0.01 radians, respectively), and random translation of individual beads in the flexible segments (up to 2 Å). A model was saved every 10 Gibbs sampling steps, each consisting of a cycle of Monte Carlo steps that moved every rigid body and flexible bead once. This sampling produced a total of 5,250,000 models from 70 independent runs, each starting from a different initial conformation of CSN, requiring ~1 day on 560 computational cores. For the most detailed specification of the sampling procedure, see the IMP modeling script (https://salilab.org/CSN2019). We only consider for further analysis. The sampling yielded 71,349 representative models that sufficiently satisfy the input restraints.

1.3 Stage 4: analyzing and validating the ensemble structures and data

Input information and output structures need to be analyzed to estimate structure precision and accuracy, detect inconsistent and missing information, and to suggest more informative future experiments. We used the analysis and validation protocol published earlier [179]. Assessment began with a test of the thoroughness of structural sampling, followed by structural clustering of the models and estimating their precision based on the variability in the ensemble of representative

structures, and quantification of the structure fit to the input information. These validations are based on the nascent wwPDB effort on archival, validation, and dissemination of integrative structure model [169]. We now discuss each one of these points in turn.

(1) ***Convergence of sampling***

We must first assess whether or not sampling found all representative solutions; this assessment is needed in particular for estimating the precision of the model ensemble consistent with the input data. We performed four tests, as follows.

The first convergence test confirmed that the scores of refined models do not continue to improve as more models are computed essentially independently (Fig. 4.8A).

The second convergence test confirmed that the representative models in independent sampling runs 1-35 (model sample 1) and 36-70 (model sample 2) satisfied the data equally well. While the non-parametric Mann-Whitney test indicates that the difference between the two score distributions is significant (p-value is less than 0.05), the magnitude of the difference is small, as demonstrated by the Cliff's *d* of 0.04 (Fig. 4.8B); thus, the two score distributions are effectively equal.

Next, we considered the representative structures themselves, not their scores as in the two tests described above. For stochastic sampling methods, thoroughness of sampling can be assessed by showing that multiple independent runs (e.g., using random starting configurations and different random number generator seeds, as is the case for model samples 1 and 2) do not result in significantly different models. We tested the similarity between model samples 1 and 2 in the following two ways.

The third convergence test relied on the χ^2-test for homogeneity of proportions between model samples 1 and 2 (Fig. 4.8C-D). The test involves clustering models from both samples, followed by comparing the proportions of models from each sample in each cluster. A comparison of two CSN structures considered all the beads representing subunits in the complex. The sampling precision is defined as the average bead RMSD between the models within the cluster and its corresponding centroid in the finest clustering for which each sample contributes models proportionally to its size (considering both significance and magnitude of the difference) and for which a sufficient proportion of all models occur in sufficiently large clusters. The sampling precision for our CSN modeling is 16Å (Fig. 4.7C-D). In particular, we computed a model at a precision of 16Å.

The fourth convergence test relied on a comparison of two localization probability density maps for CSN, obtained for models in samples 1 and 2 (Fig. 4.8E). A localization probability density map defines the probability of any voxel (here, 5x5x5 $Å^3$) being occupied by a specific protein in a set of model densities, which in turn are obtained by convolving superposed models with a Gaussian kernel (here, with a standard deviation of 3.4 Å, corresponding to a resolution equal to the cluster precision). The average cross-correlation coefficient between the two maps of CSN is 0.99, indicating that the position of CSN in the two samples is nearly identical at the model precision of 16 Å.

In conclusion, all four sampling tests indicate that the sampling was exhaustive at 16 Å precision. The caveat is that passing these tests is only necessary but not sufficient as evidence of thorough sampling; a positive outcome of the tests may be misleading if, for example, the landscape contains only a narrow, and thus difficult to find, pathway to the pronounced minimum corresponding to the *correct structure.*

(2) *Clustering* and structure precision

An ensemble of good-scoring structures needs to be analyzed in terms of the precision of its structural features [34,75,86,151]. The precision of a component position can be quantified by its variation in an ensemble of superposed good-scoring structures. It can also be visualized by the localization probability density for each of the components of the model. As described above, integrative structure determination of the CSN resulted in effectively a single solution, at the precision of 16 Å (Fig. 4.8D). This precision is sufficient to determine the structure of CSN.

(3) *Fit to input information*

An accurate structure needs to satisfy the input information used to compute it. Because the sampling was exhaustive at 16Å precision, overfitting is not a problem at this precision; all structures at this precision that are consistent with the data are provided in the ensemble.

Here, we consider a cluster satisfies a cross-link if the cross-linked distance in any individual model in the cluster is less than the maximum threshold (i.e., the maximum length of the assessed cross-linkers). In particular, the threshold is 30Å, 30Å, and 45Å for the DSSO, DHSO, and BMSO cross-linkers, respectively. The integrative CSN structure satisfied 98% of the input cross-links of the inter-subunit and intra-subunit cross-links; the remaining 2% of the cross-links are satisfied if the threshold is increased by 10Å (Fig. 4.8F). These violations can be rationalized by possible false positive cross-links, coarse-grained representation of the complex, and/or finite structural sampling.

The remainder of the restraints are harmonic, with a specified standard deviation. The dominant cluster generally satisfied at least 95% of restraints of each type (Table 4.24-4.25); a restraint is satisfied by a cluster of structures if the restrained distance in any structure in the cluster

(considering restraint ambiguity) is violated by less than 3 standard deviations, specified for the restraint. Most of the violations are small, and can be rationalized by local structural fluctuations, coarse-grained representations of some CSN domains and/or finite structural sampling.

(4) ***Cross-validation against the input cross-links***

We independently repeated integrative modeling described above with six different subsets of cross-links of CSN (Tables 4.10-4.13), including (i) DSSO only, (ii) DHSO only, (iii) BMSO only, (iv) DSSO and DHSO, (v) DSSO and BMSO, and (vi) DHSO and BMSO (Fig. 4.10A-F). See **Validation of the Integrative CSN Structure** in main text.

In conclusion, sampling is exhaustive, representation is appropriate for modeling, and both the tertiary and quaternary organization of the integrative structure are cross-validated, thus validating the integrative structure of CSN.

(5) ***Structural ensemble comparison***

To compare two structures (i.e., ensemble) taking into considerations their precisions, we calculated the overlap between the two structures. The overlap was quantified by the ratio of the distance between ensemble centroids to three times the sum of the ensemble precisions. For two normal distributions with the same standard deviation, an overlap of 1 corresponds to 0.13% of one distribution being overlapped by the other. Thus, an overlap greater than 1 indicates that the two model ensembles are different, given their precisions. The distance between two ensemble centroids is defined by their RMSD. The ensemble precision is defined by the RMSD from the centroid averaged over all models in the ensemble.

2.1 Integrative structure determination of the human non-canonical CSN complex

2.1 Stage 1: gathering data.

In total, 544 highly reproducible intra- and intermolecular DSSO, DHSO, and BMSO cross-links were identified using mass spectrometry (Table 4.14-4.17), which informed the spatial proximities among the 9 subunits of the CSNn and their conformations.

Representations of CSN1-8 subunits were identical to the representation used to determine the CSN structure above. Representation of CSN9 relied on secondary structure and disordered regions predicted by PSIPRED based on the CSN9 sequence [180].

2.2 Stage 2: representing subunits and translating data into spatial restraints

The CSN1-8 representation for CSNn structure was identical to the one used for CSN structure (Fig. 4.13H). Because the structure of CSN9 has not been previously determined nor can it be modeled based on its sequence, it was represented as a string of flexible beads corresponding to two residues each.

With this representation in hand, we next encoded the spatial restraints into a scoring function based on the information gathered in Stage 1, as follows (Tables 4.14-4.17).

(1) *Cross-link restraints:* The 269 DSSO, 83 DHSO, and 75 BMSO cross-links (Tables 4.14-4.17) were used to construct the Bayesian scoring function [177,181181181] that restrained the distances spanned by the cross-linked residues [39].

(2) *Excluded volume restraints*: The excluded volume restraints were applied to each 10-residue bead, using the statistical relationship between the volume and the number of residues that it covered [34,176,178].

(3) *Sequence connectivity restraints*: We applied the sequence connectivity restraints, using a harmonic upper distance bound on the distance between consecutive beads in a subunit, with a threshold distance equal to twice the sum of the radii of the two connected beads. The bead radius was calculated from the excluded volume of the corresponding bead, assuming standard protein density [34,176].

This sampling produced a total of 7,500,000 models from 100 independent runs, each starting from a different initial conformation of CSNn, requiring ~1 day on 800 computational cores. For the most detailed specification of the sampling procedure, see the IMP modeling script (https://salilab.org/CSN2019). We only consider for further analysis. The sampling yielded 125,750 representative models that sufficiently satisfy the input restraints.

2.3 Stage 4: analyzing and validating the ensemble structures and data

We assessed CSNn structures with identical tests for thoroughness of sampling, its fit to input information, and to data not used for modeling.

(1) *Convergence of sampling*

The first convergence test confirmed that the scores of refined models do not continue to improve as more models are computed essentially independently (Fig. 4.13A). The second convergence test confirmed that representative models in independent sampling runs 1-50 (model sample 1) and 51-100 (model sample 2) satisfied the data equally well (p-value is less than 0.05 and Cliff's *d* of 0.10) (Fig. 4.13B). The third convergence test validated that the CSNn structure is exhaustive at 22Å and a structure of CSNn was computed at a precision of 22Å (Fig. 4.13C-D). The fourth test confirmed that the CSNn structures in sample 1 and 2 are identical. In particular, the cross-correlation of the localization probability density maps for the two samples of CSNn is 0.91 (Fig.

4.13E). In conclusion, all four sampling tests indicate that the sampling was exhaustive at 22Å precision.

(2) ***Clustering and structure precision***

As described above, integrative structure determination of the CSNn resulted in effectively a single solution, at the precision of 22 Å (Fig. 4.13D). This precision is sufficient to determine the structure of CSNn and the binding mode of CSN9 on the CSNn,

(3) ***Fit to input information***

The structure of the integrative CSNn satisfied 99% of the input cross-links (inter-subunit and intra-subunit) (Fig. 4.13F). In conclusion the sampling is exhaustive, representation is appropriate for modeling, and the structure of CSNn satisfies data not used for modeling, thus validating the integrative structure of CSNn.

FIGURE 0.1

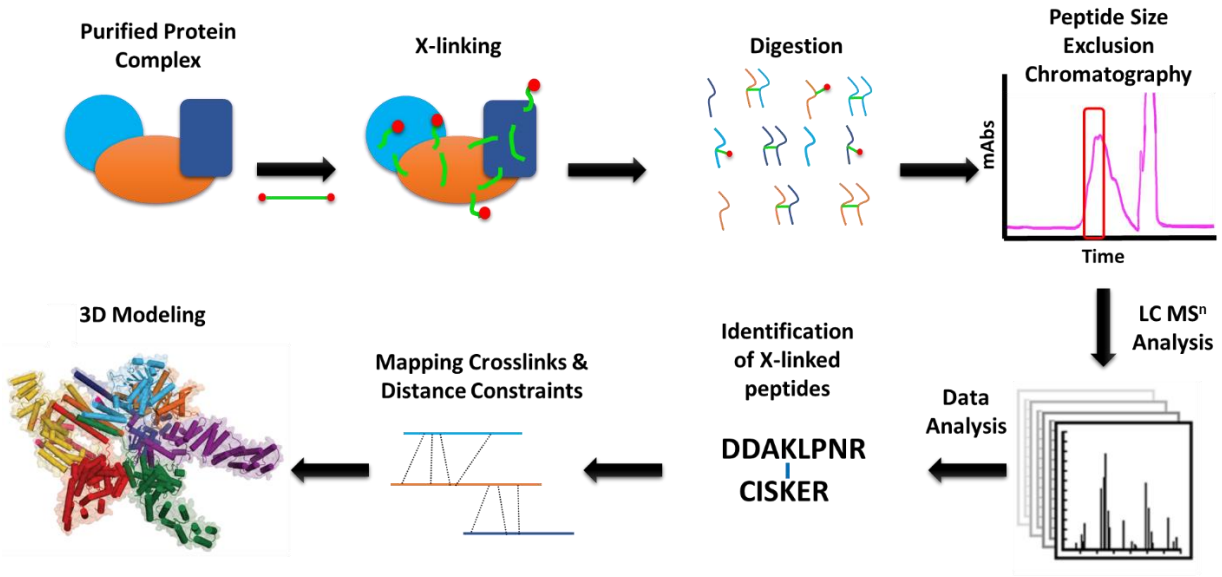

Figure 0.1. The general XL-MS Workflow

FIGURE 1.1

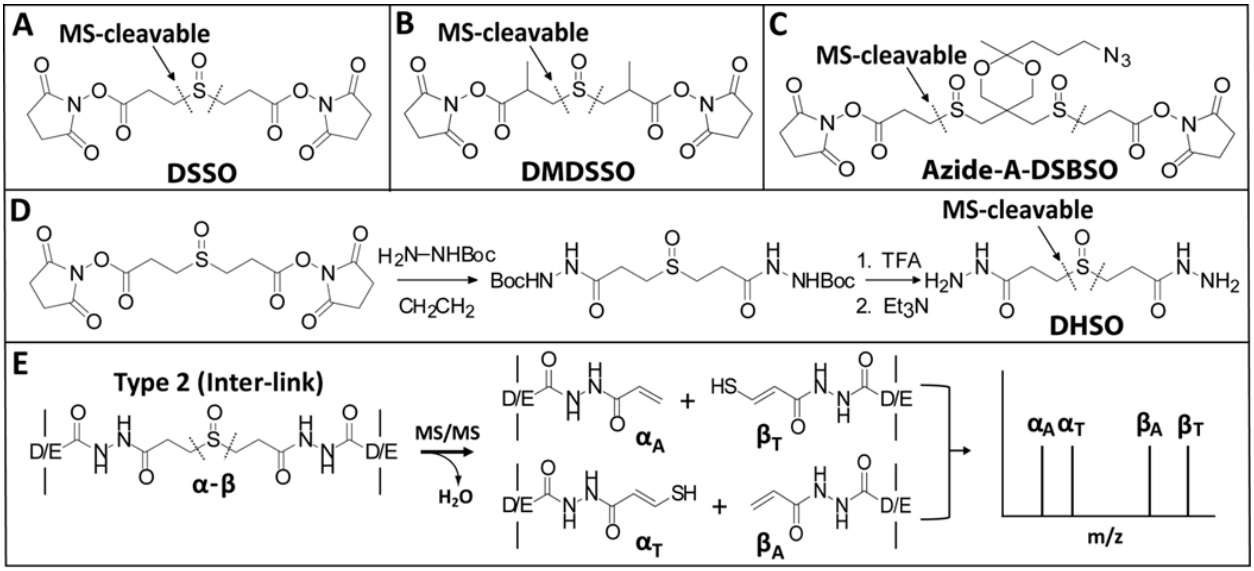

Figure 1.1. Sulfoxide containing MS-cleavable cross-linkers. (A) DSSO 14. (B) DMDSSO 15. and (C) Azide-A-DSBSO 16. (D) Synthesis scheme of MS-cleavable cross-linker DHSO. (E) Characteristic MS2 fragmentation of DHSO inter-linked heterodimer α-β.

FIGURE 1.2

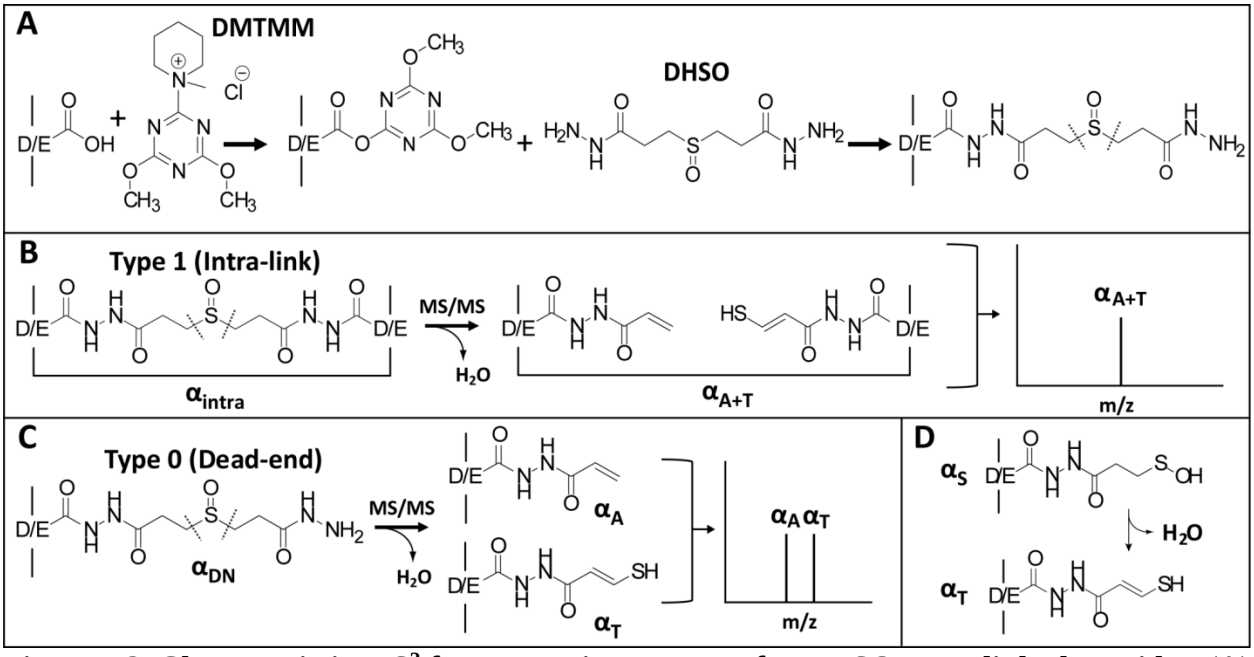

Figure 1.2. Characteristic MS^2 fragmentation patterns for DHSO cross-linked peptides. (A) The scheme of peptide cross-linking by DHSO in the presence of DMTMM. MS^2 fragmentation of (B) DHSO intra-linked peptide α_{intra}. (C) Dead-end modified peptide α_{DN}. (D) The conversion scheme of α_S to α_T. Note: S* (sulfenic acid moiety) can be converted to the more stable unsaturated thiol moiety (T) via water loss as shown in (D).

FIGURE 1.3

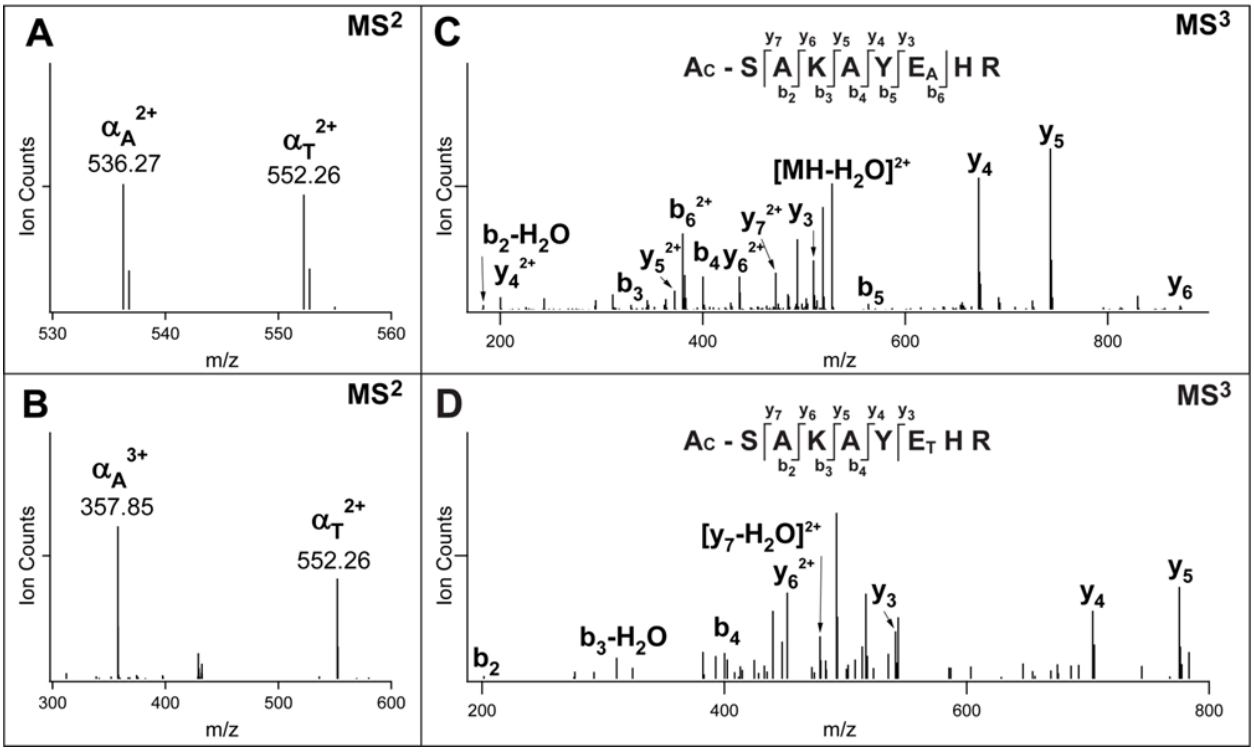

Figure 1.3. MS^n analysis of DHSO inter-linked Ac-SR8 homodimer. MS2 spectra of DHSO inter-linked Ac-SR8 at two different charge states: (A) [α−α]4+ (m/z 548.76234+) and (B) [α−α]5+ (m/z 439.21175+). MS3 spectra of MS2 fragment ions detected in (A): (C) αA (m/z 536.272+) and (D) βT (m/z 552.262+).

FIGURE 1.4

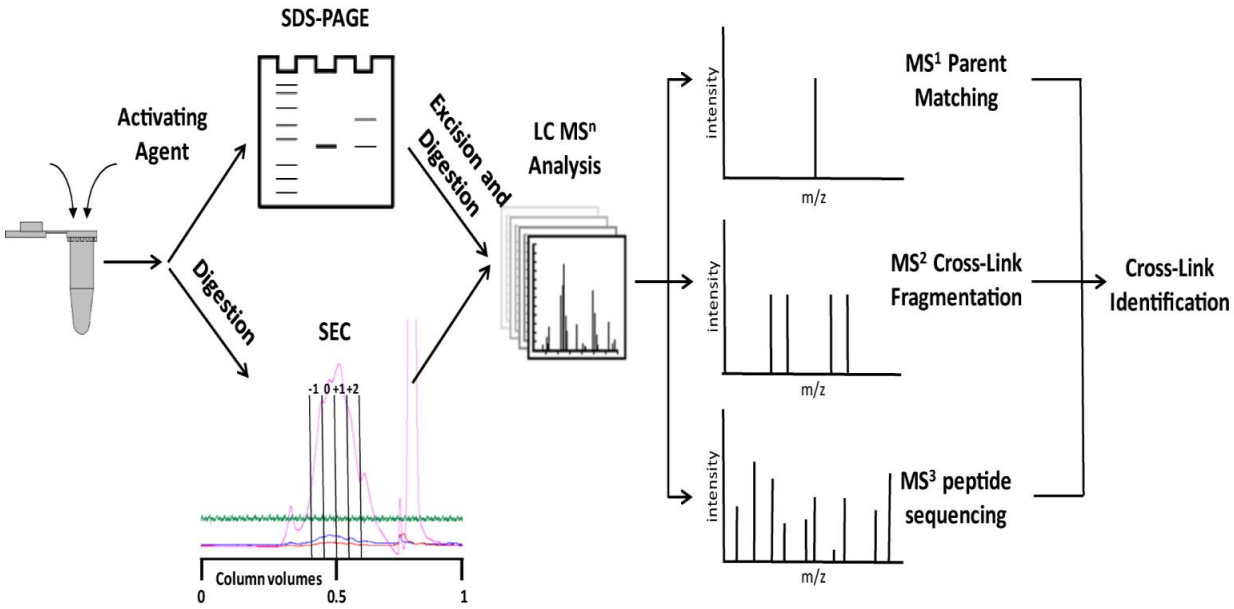

Figure 1.4. The general XL-MS workflow for the identification of cross-linked DHSO peptides from proteins.

FIGURE 1.5

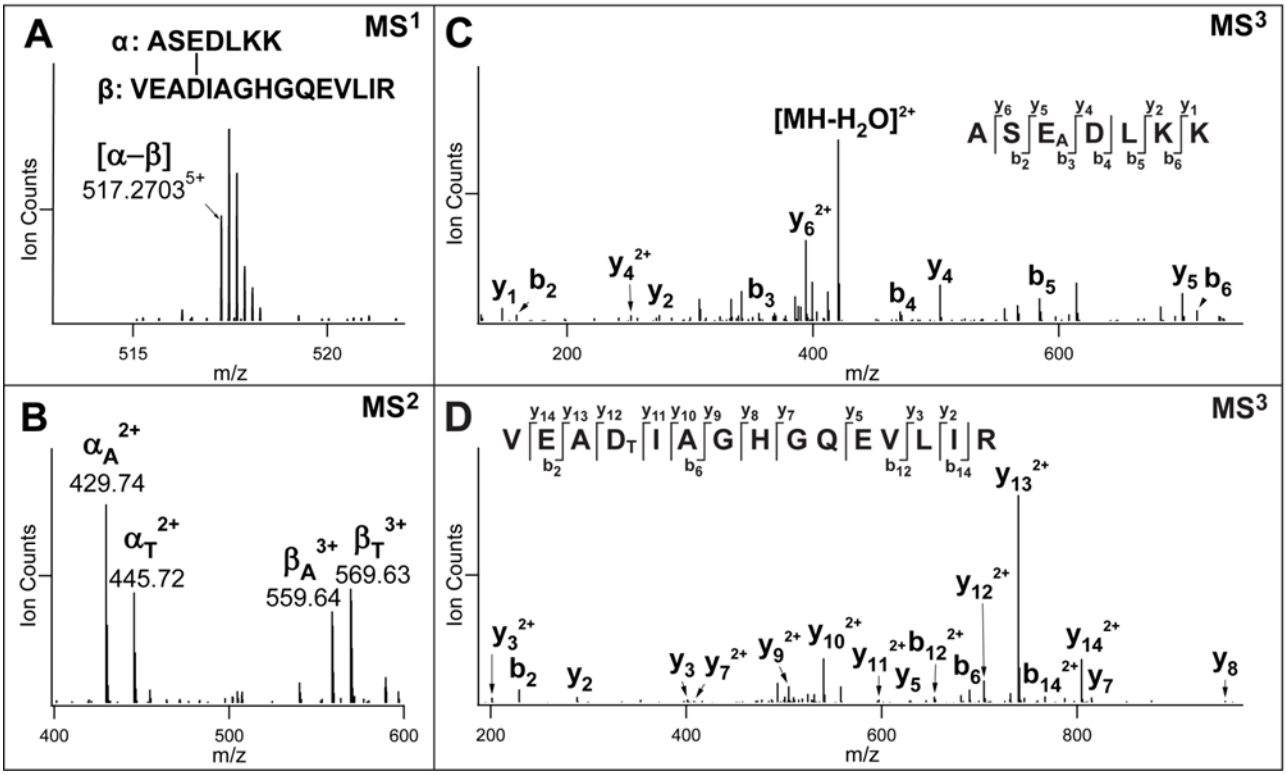

Figure 1.5. MS^n analysis of a representative DHSO inter-linked Myoglobin peptide. (A) MS spectrum of the inter-linked peptide α−β (m/z 517.27035+). (B) MS2 spectrum of the inter-linked peptide detected in (A). MS3 spectra of MS2 fragment ions: (C) αA (m/z 429.742+) and (D) βT (m/z 569.633+).

FIGURE 1.6

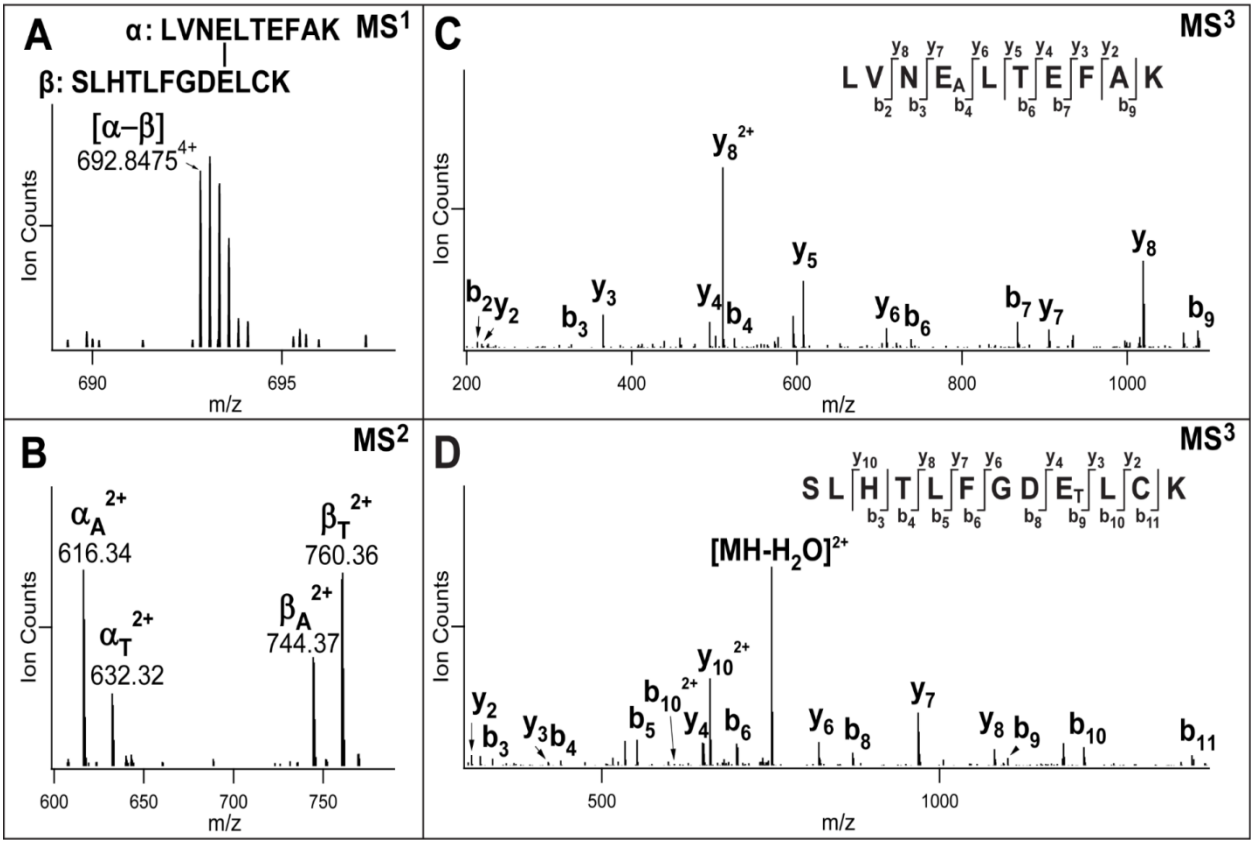

Figure 1.6. MS^n analysis of a representative DHSO inter-linked BSA peptide. (A) MS spectrum of a DHSO interlinked BSA peptide α–β (m/z 692.8475^{4+}). (B) MS^2 spectrum of the cross-linked peptide detected in (A). MS^3 spectra of MS^2 fragment ions (C) α_A (m/z 616.34^{2+}) and (D) β_T (m/z 760.36^{2+}).

FIGURE 1.7

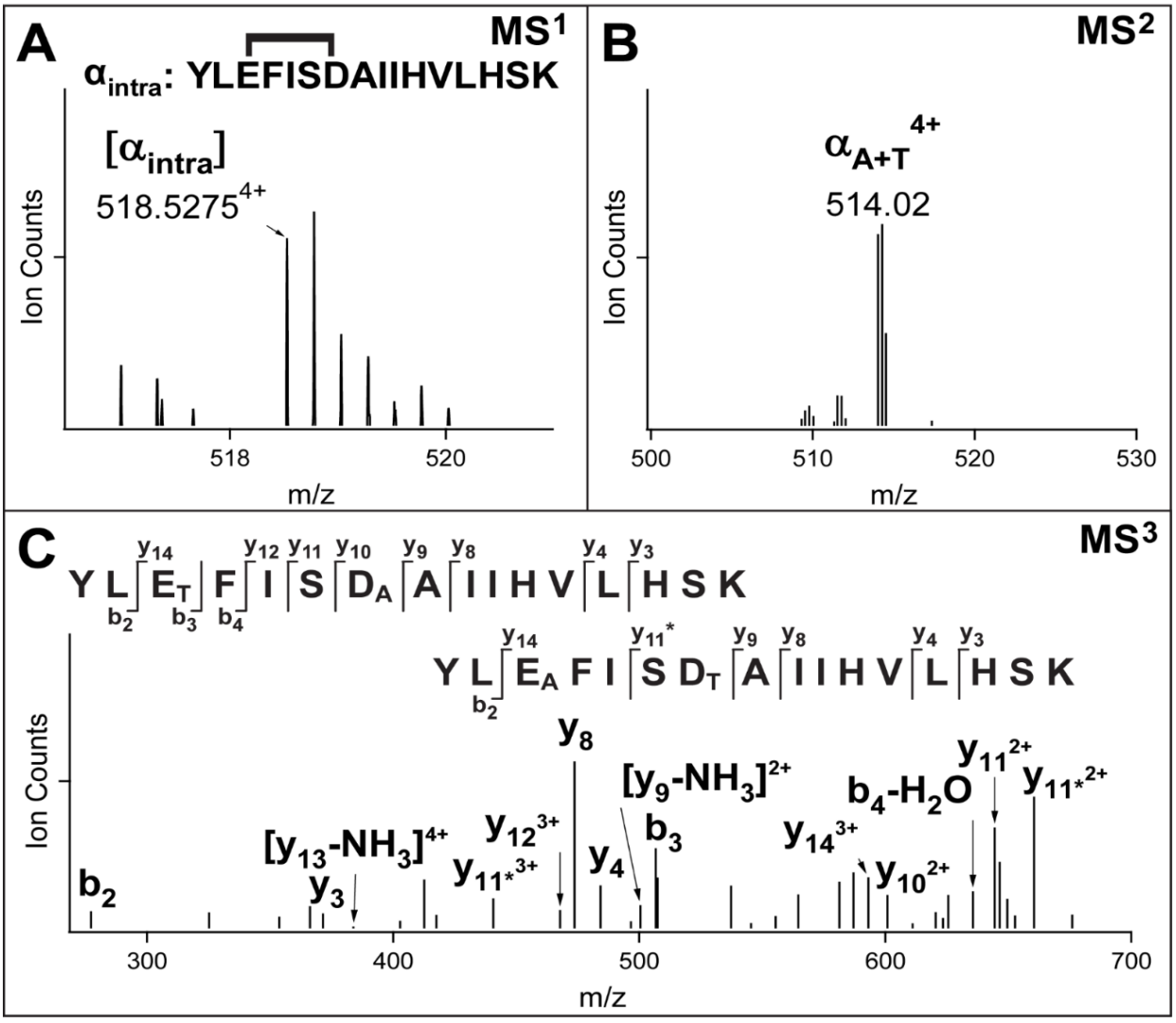

Figure 1.7. MS^n analysis of a DHSO intra-linked myoglobin peptide. (A) MS spectrum of a DHSO intra-linked myoglobin peptide α_{intra} (m/z 518.5275^{4+}). (B) MS^2 spectrum of the intra-linked peptide in (A). (C) MS^3 spectrum of the MS^2 fragment ion α_{A+T} (m/z 514.02^{4+}).

FIGURE 1.8

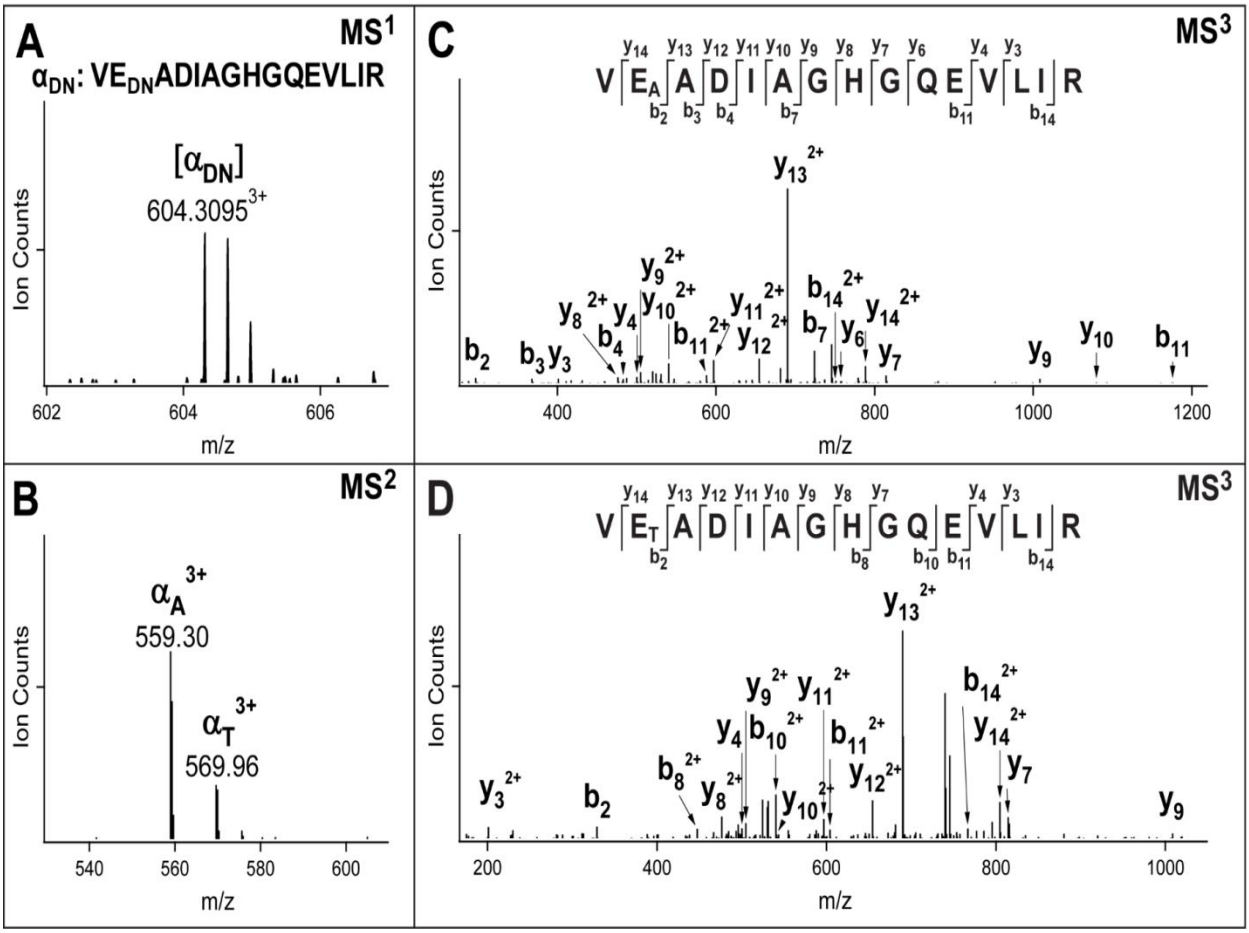

Figure 1.8. MS^n analysis of a DHSO dead-end modified myoglobin peptide. (A) MS spectrum of a DHSO dead-end modified myoglobin peptide α_{DN} (604.3095^{3+}). (B) MS^2 spectrum of the parent ion detected in (A). MS^3 spectra of MS^2 fragment ions (C) α_A (m/z 559.30^{3+}) and (D) α_T (m/z 569.96^{3+}).

FIGURE 1.9

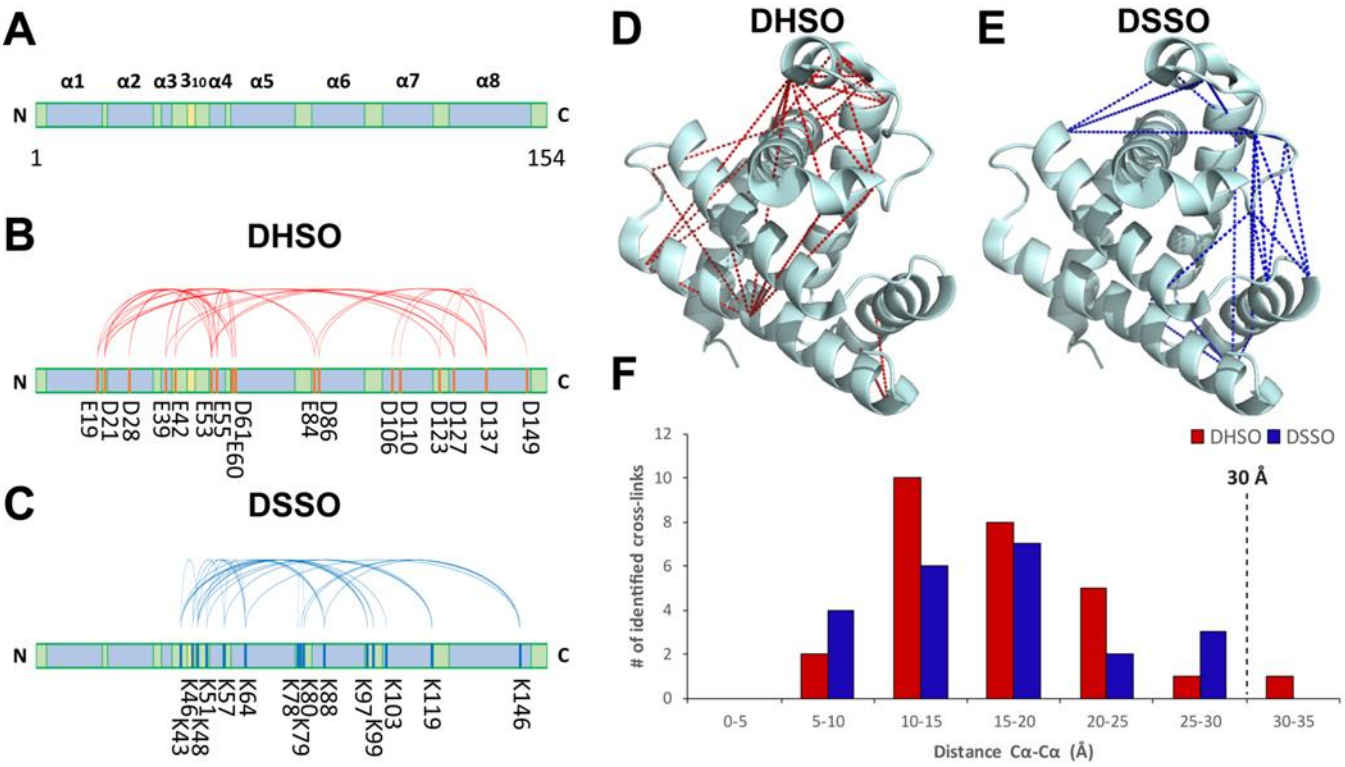

Figure 1.9. Myoglobin cross-link maps. (A) Myoglobin linear sequence showing locations of the 8 α-helices (blue) and 310 helix (yellow). (B) DHSO cross-link map on myoglobin linear sequence. (C) DSSO cross-link map on myoglobin linear sequence. (D) DHSO cross-link map on myoglobin crystal structure (PDB: 1DWR). (E) DSSO cross-link map on myoglobin crystal structure (PDB: 1DWR). (F) The distribution plot of identified linkages vs. their spatial distances of D|E-D|E for DHSO (red) or K-K for DSSO (blue) in myoglobin structure.

FIGURE 1.10

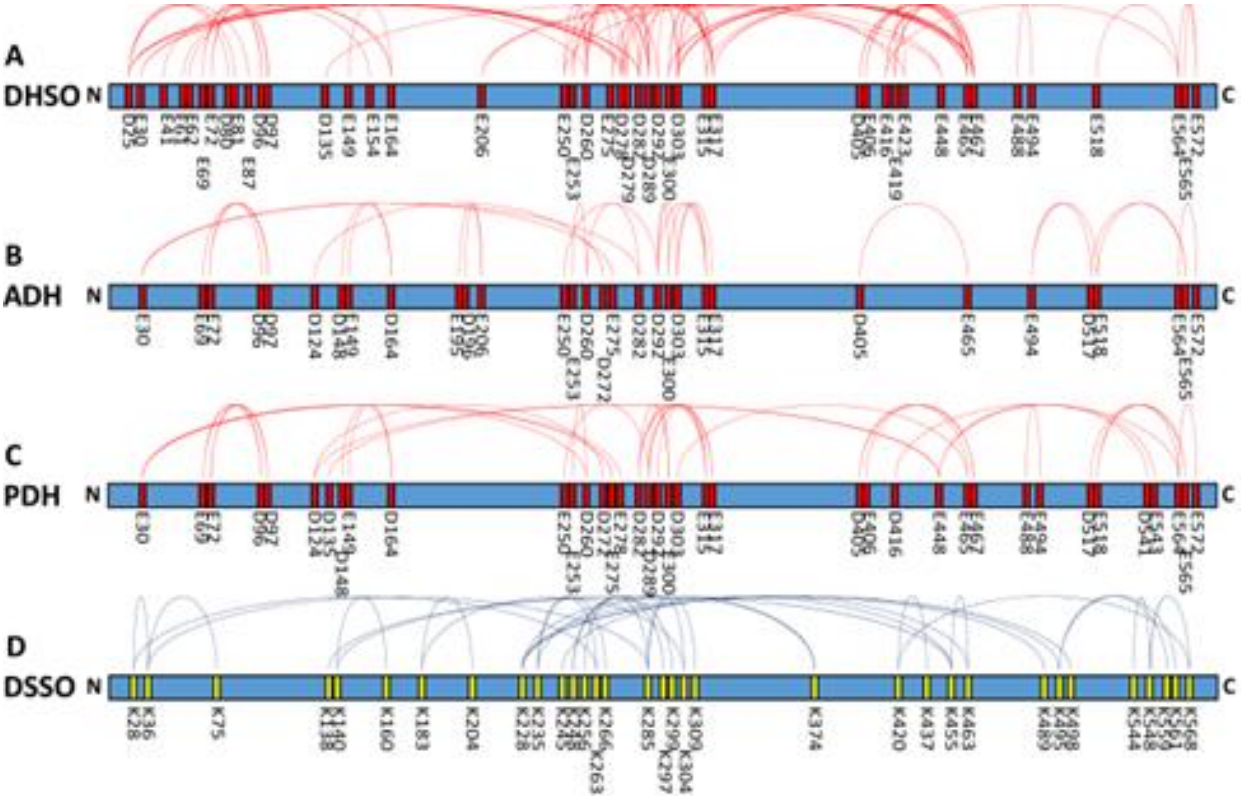

Figure 1.10. BSA cross-link maps on its linear sequence. (A) DHSO cross-link map. (B) ADH cross-link map. (C) PDH cross-link map. (D) DSSO cross-link map. Note: ADH and PDH cross-link maps are generated based on data obtained by Leitner et al. 2014.

FIGURE 1.11

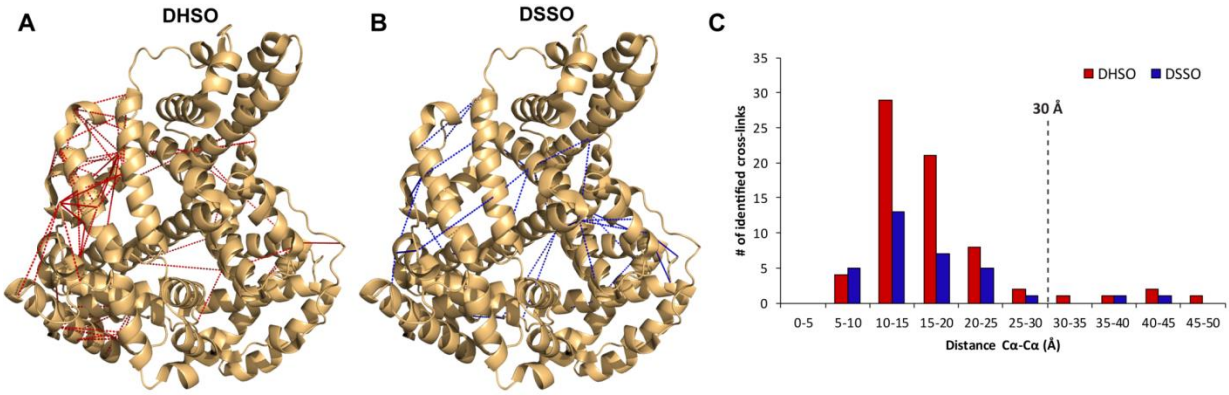

Figure 1.11. BSA cross-link maps on its crystal structure (PDB: 4F5S). (A) with DHSO cross-link map (red). (B) DSSO cross-link map (blue). (C) The distribution plot of identified linkages vs. their spatial distances between D|E-D|E for DHSO (red) or K-K for DSSO (blue) in BSA structure.

FIGURE 2.1

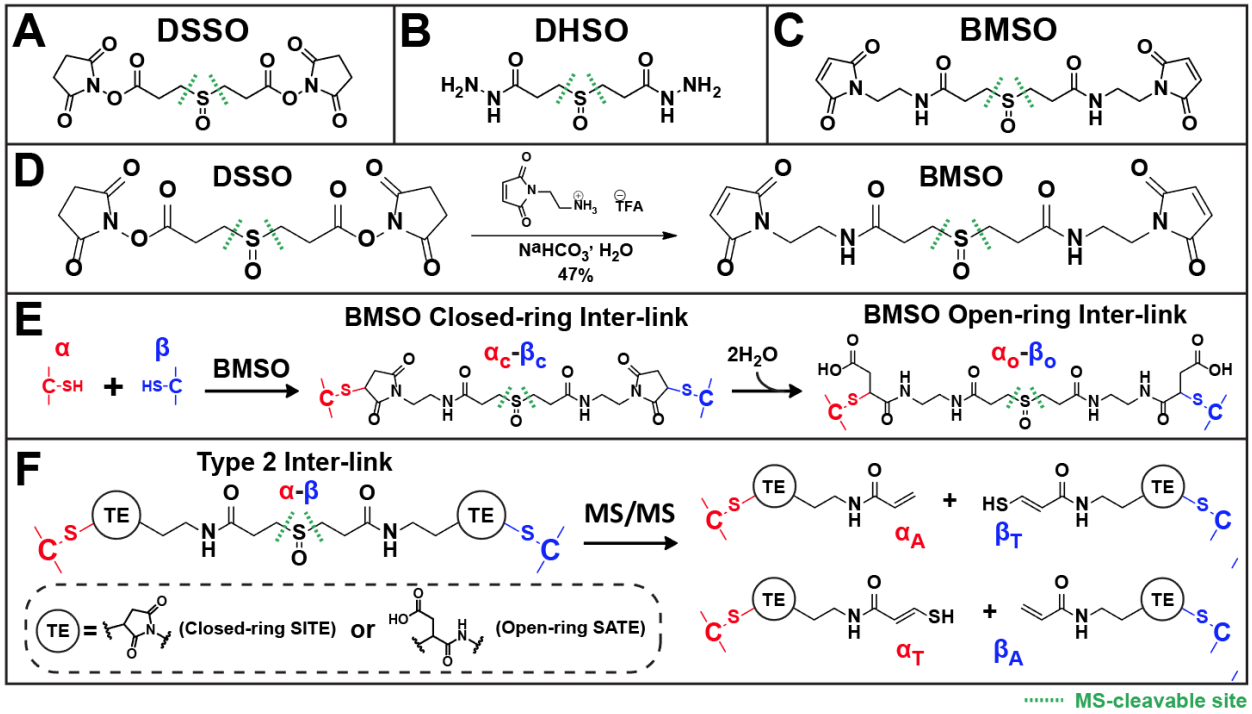

Figure 2.1. Design, synthesis and characteristics of the Sulfoxide-containing MS-cleavable Cysteine Reactive Cross-linker, BMSO. Molecular structures of (A) DSSO [60], (B) DHSO [65], and (C) BMSO [66]. (D) Synthesis scheme of BMSO. (E) BMSO cross-linking results in the formation of an inter-linked heterodimer (α-β), in which the closed-ring SITE structure on cross-linked cysteines can be converted to open-ring SATE structures upon hydrolysis. (F) Predicted characteristic MS^2 fragmentation of a BMSO inter-linked heterodimer α-β, in which thioester (TE) structures on cross-linked cysteines can be in the form of either closed-ring SITE or open-ring SATE. Note: c: closed-ring; o: open-ring.

FIGURE 2.2

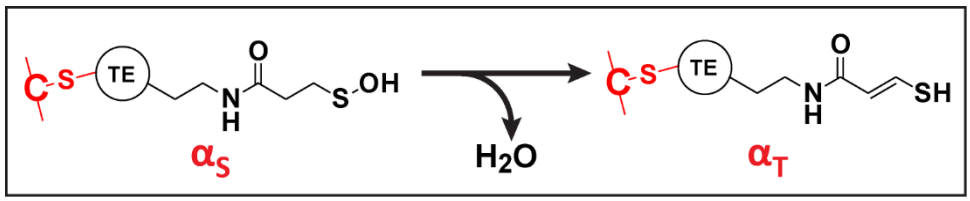

Figure 2.2. The conversion scheme of α_S to α_T, in which the sulfenic acid moiety (S) became the more stable unsaturated thiol moiety (T) via water loss.

FIGURE 2.3

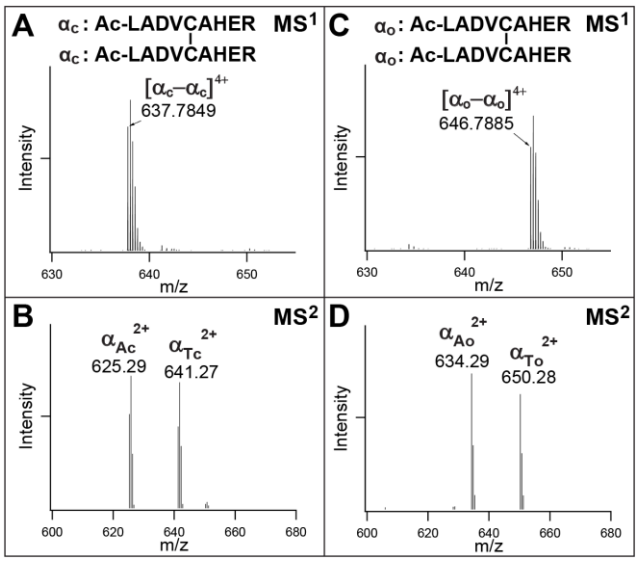

Figure 2.3. **MS analysis of the BMSO inter-linked Ac-LR9 homodimer (α−α).** (A) MS^1 spectrum of the inter-link with closed-ring form, $(\alpha_c–\alpha_c)^{4+}$ (m/z 637.7849^{4+}). (B) MS^2 spectrum of the $(\alpha_c–\alpha_c)^{4+}$ detected in (A), in which two dominant fragment ions, i.e. α_{Ac} (m/z 625.29^{2+}) and α_{Tc} (m/z 641.27^{2+}), were detected as predicted for homodimer inter-links. (C) MS^1 spectrum of the inter-link with open-ring form, $(\alpha_o–\alpha_o)^{4+}$ (m/z 646.7885^{4+}). (D) MS^2 spectrum of the $(\alpha_o–\alpha_o)^{4+}$ detected in (C), in which two dominant fragment ions, i.e. α_{Ao} (m/z 634.29^{2+}) and α_{To} (m/z 650.28^{2+}) were detected as expected. Note: c: closed-ring; o: open-ring; Ac/Tc: alkene/unsaturated thiol moieties with closed-ring SITE; Ao/To: alkene/unsaturated thiol moieties with open-ring SATE.

FIGURE 2.4

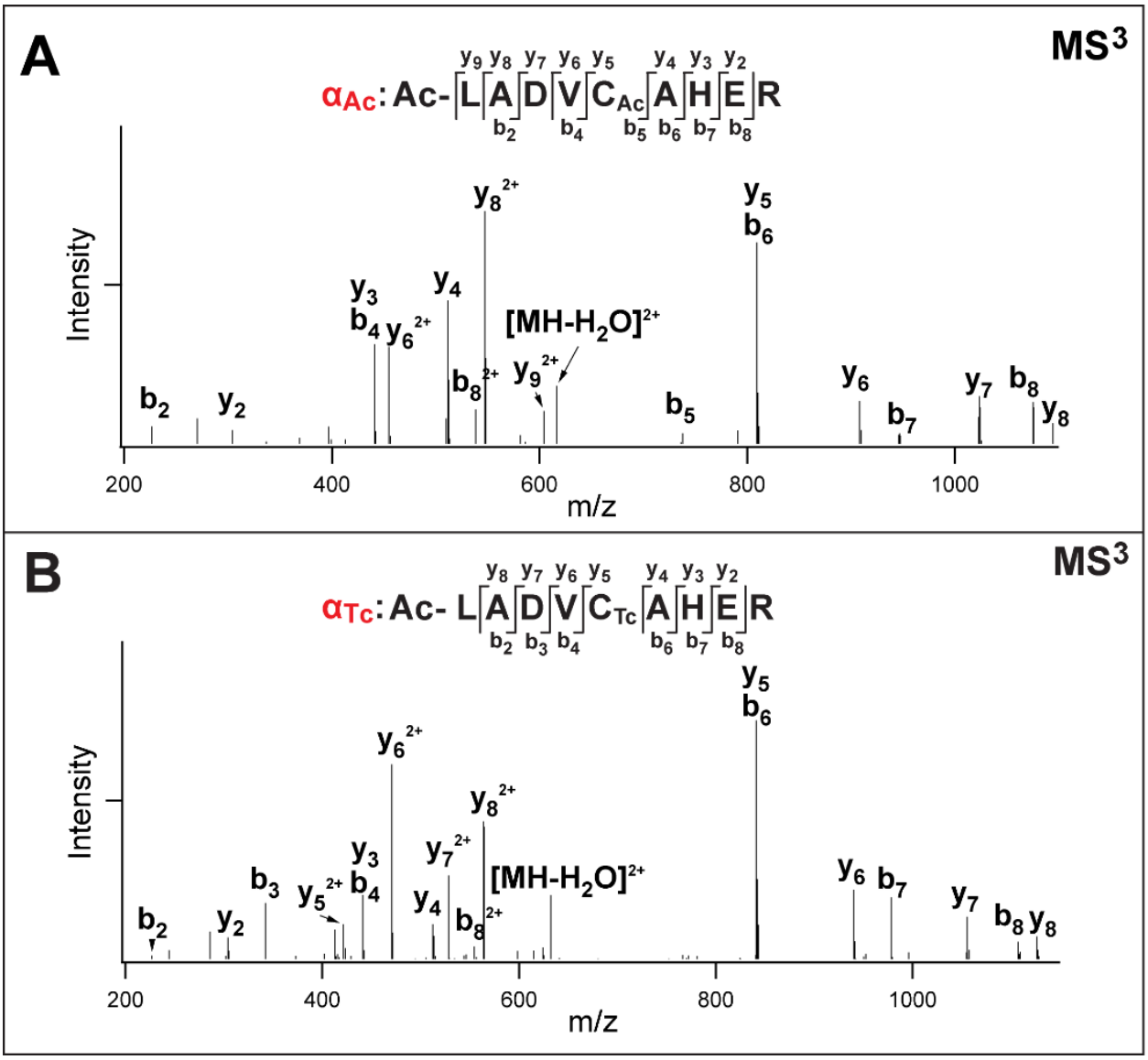

Figure 2.4. MS^3 Analysis of the BMSO inter-linked Ac-LR9 homodimer with closed-ring as shown in Figure 2.3A. MS^3 spectra of MS^2 fragment ions detected in Figure 2B: (A) α_{Ac} (m/z 625.29^{2+}) and (B) α_{Tc} (m/z 641.27^{2+}).

FIGURE 2.5

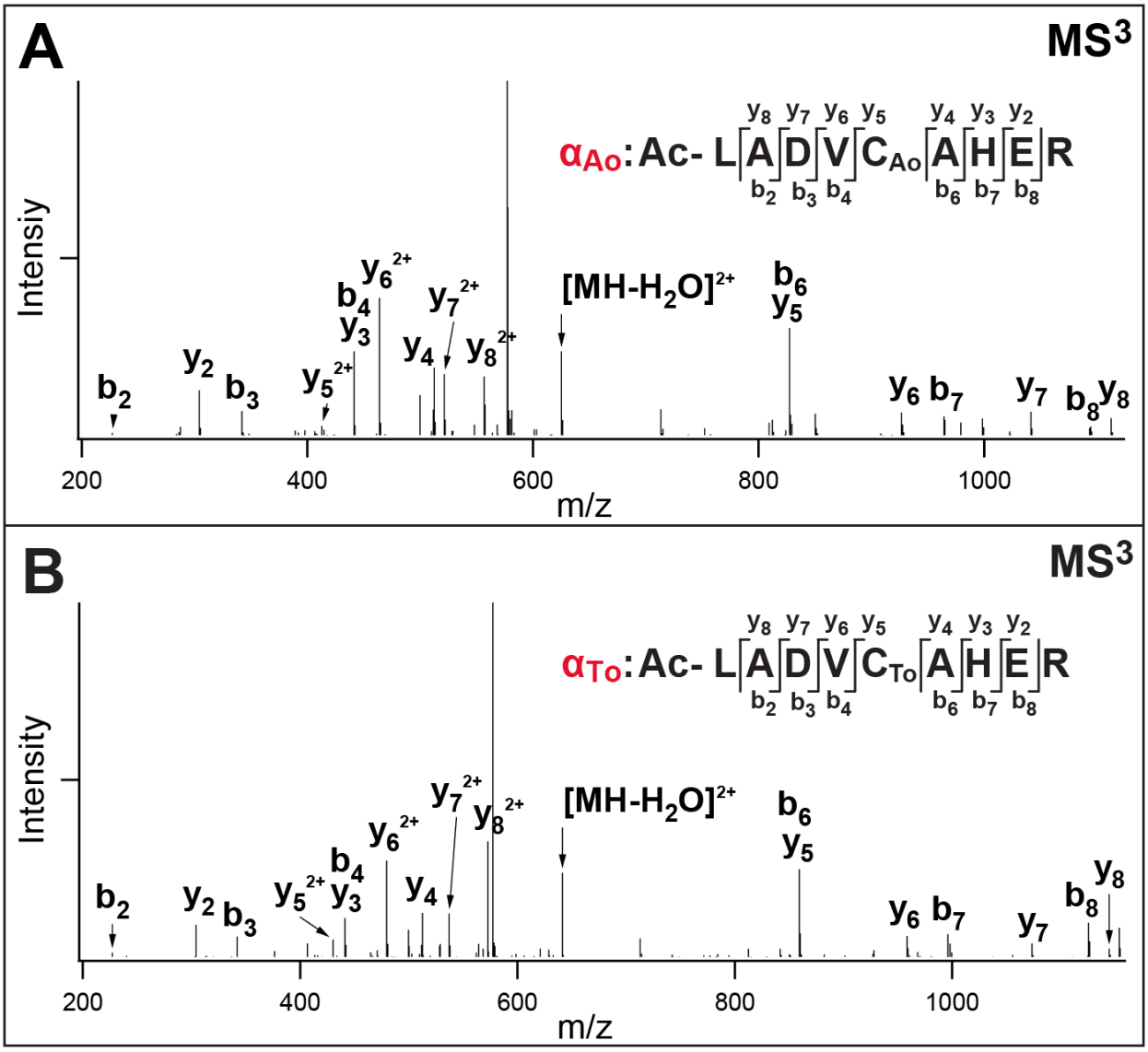

Figure 2.5. MS^3 Analysis of the BMSO inter-linked Ac-LR9 homodimer with open-ring as shown in Figure 2.3C. MS^3 spectra of MS^2 fragment ions detected in Figure 2D: (A) α_{Ao} (m/z 634.29^{2+}) and (B) α_{To} (m/z 650.28^{2+}).

FIGURE 2.6

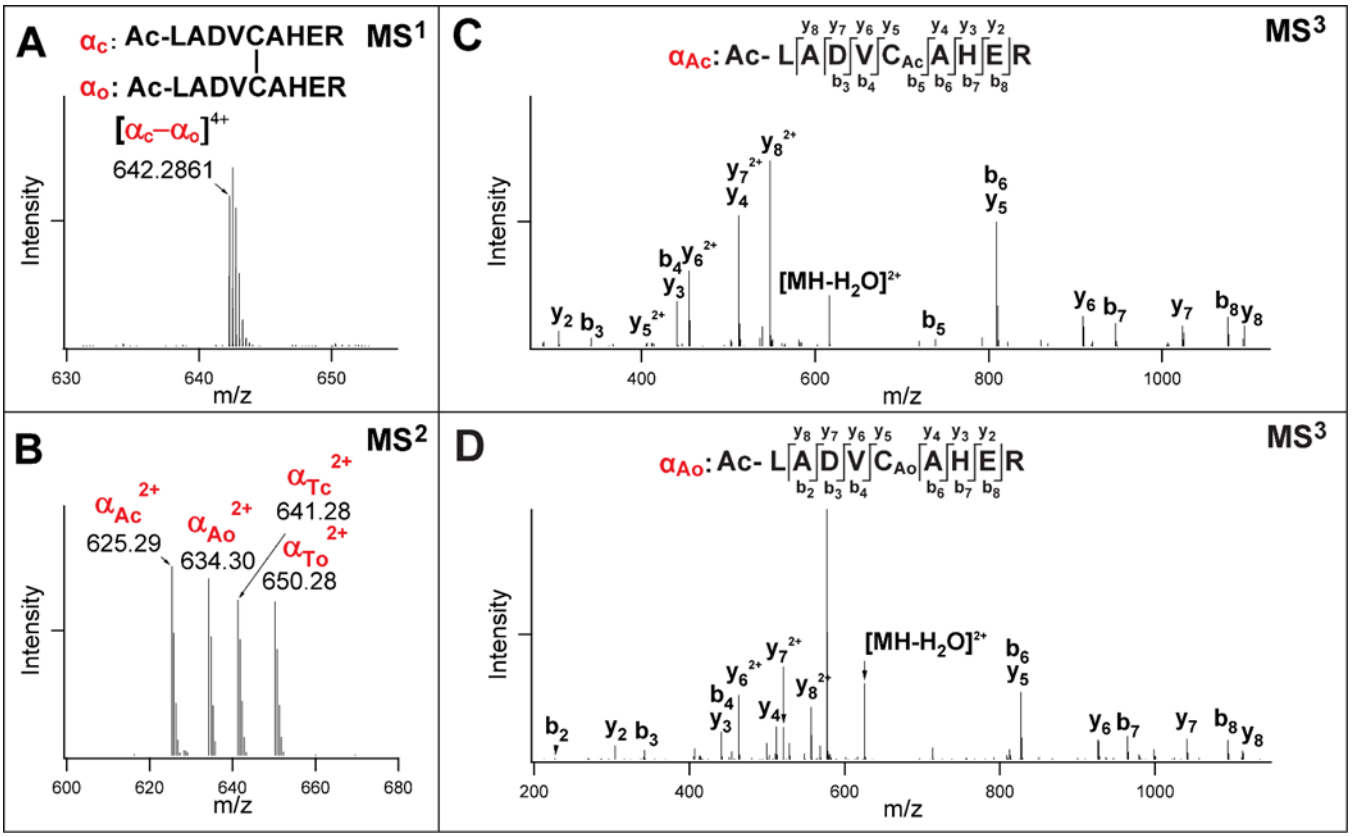

Figure 2.6. MS^n analysis of the BMSO inter-linked Ac-LR9 homodimer with one closed-ring and one open-ring structures (α_c–α_o). (A) MS^1 spectrum of the BMSO mixed-ring Ac-LR9 homodimer: $(\alpha_c–\alpha_o)^{4+}$ (m/z 642.2861^{4+}). (B) MS^2 analysis of $(\alpha_c–\alpha_o)^{4+}$ detected in (A) resulted in two fragment ion pairs: α_{Ac}/α_{To} (m/z 625.29^{2+}/650.28^{2+}) and α_{Tc}/α_{Ao} (m/z 641.28^{2+}/634.30^{2+}). This is expected as the ring states made the two identical sequences different, by modifying them with alkene and unsaturated moieties carrying either closed-ring or open-ring structures. Therefore, this makes the peptide ($\alpha_c–\alpha_o$) behave like heterodimer instead of homodimer. MS^3 analyses of (C) α_{Ac} (m/z 625.29^{2+}) and (D) α_{To} (m/z 650.28^{2+}) identified their sequences as Ac-LADVC$_{Ac}$AHER and Ac-LADVC$_{To}$AHER, respectively. Note: c: closed-ring; o: open-ring; Ac/Tc: alkene/unsaturated thiol moieties with closed-ring SITE; Ao/To: alkene/unsaturated thiol moieties with open-ring SATE.

FIGURE 2.7

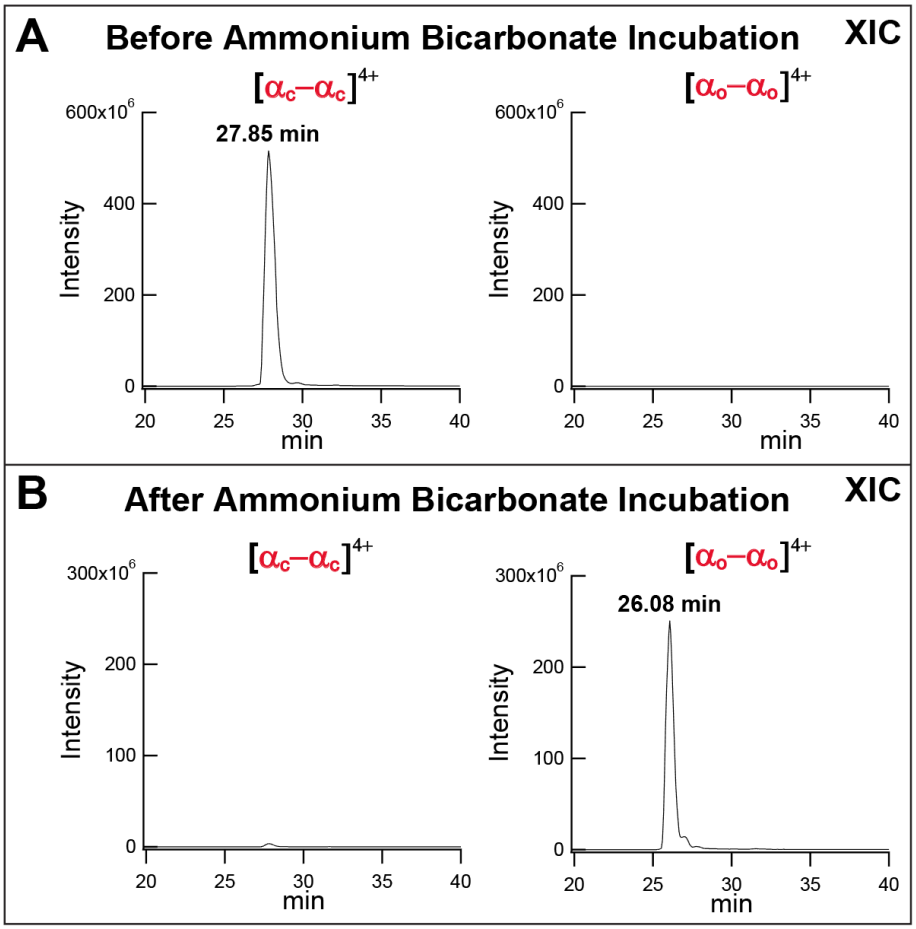

Figure 2.7. MS analysis of BMSO inter-linked Ac-LR9 homodimer before and after ammonium bicarbonate treatment. Extracted Ion Chromatograms display ion intensities of the two forms: (α_c–α_c) with closed-ring and (α_o–α_o) with open-ring before (A) and after (B) incubation with Ammonium Bicarbonate at 37°C overnight.

FIGURE 2.8

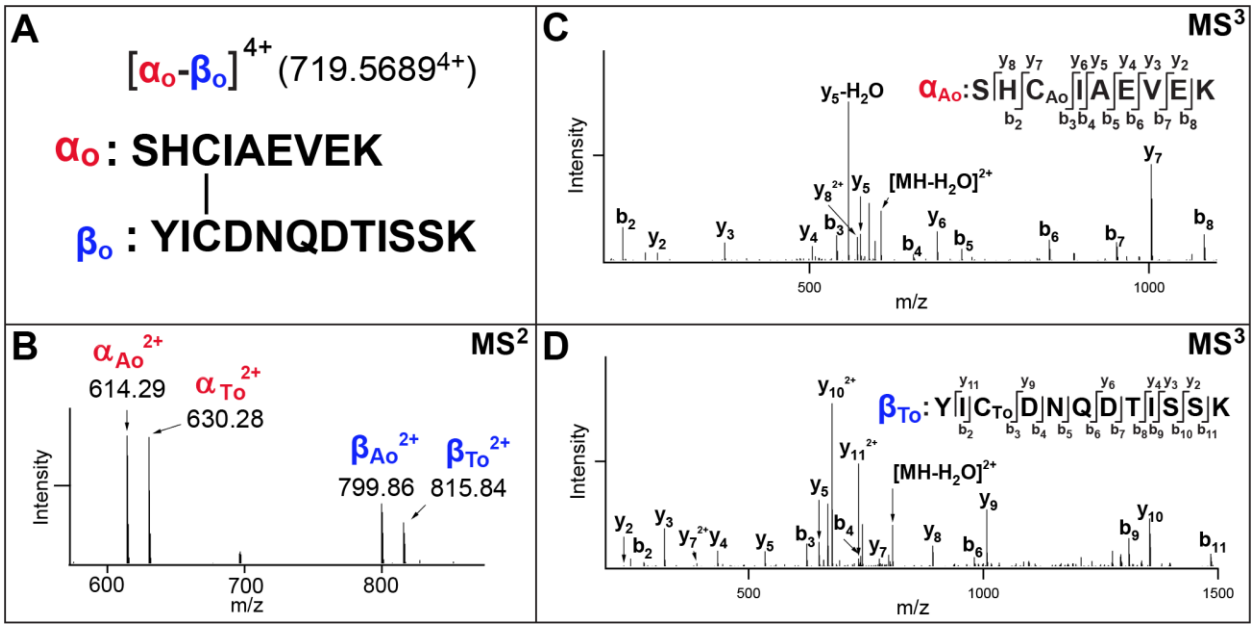

Figure 2.8. MS^n analysis of a representative BMSO inter-linked peptide of BSA. (A) The BMSO inter-linked peptide with open-ring SATE structures, (α_o–β_o) (m/z 719.5689^{4+}). (B) MS^2 spectrum of the (α_o–β_o), in which two characteristic fragment ion pairs were detected, i.e. α_{Ao}/ β_{To} (m/z 614.29^{2+}/815.84^{2+}) and α_{To}/ β_{Ao} (m/z 630.28^{2+}/799.86^{2+}). (C) MS^3 analysis of α_{Ao} (m/z 614.29^{2+}) identified the sequence as SHC$_{Ao}$IAEVEK, in which the cysteine residue was modified with alkene moiety carrying an open-ring SATE. (D) MS^3 analysis of β_{To} (m/z 815.84^{2+}) identified the sequence as YIC$_{To}$DNQDTISSK, in which the cysteine residue was modified with unsaturated thiol moiety carrying an open-ring SATE.

FIGURE 2.9

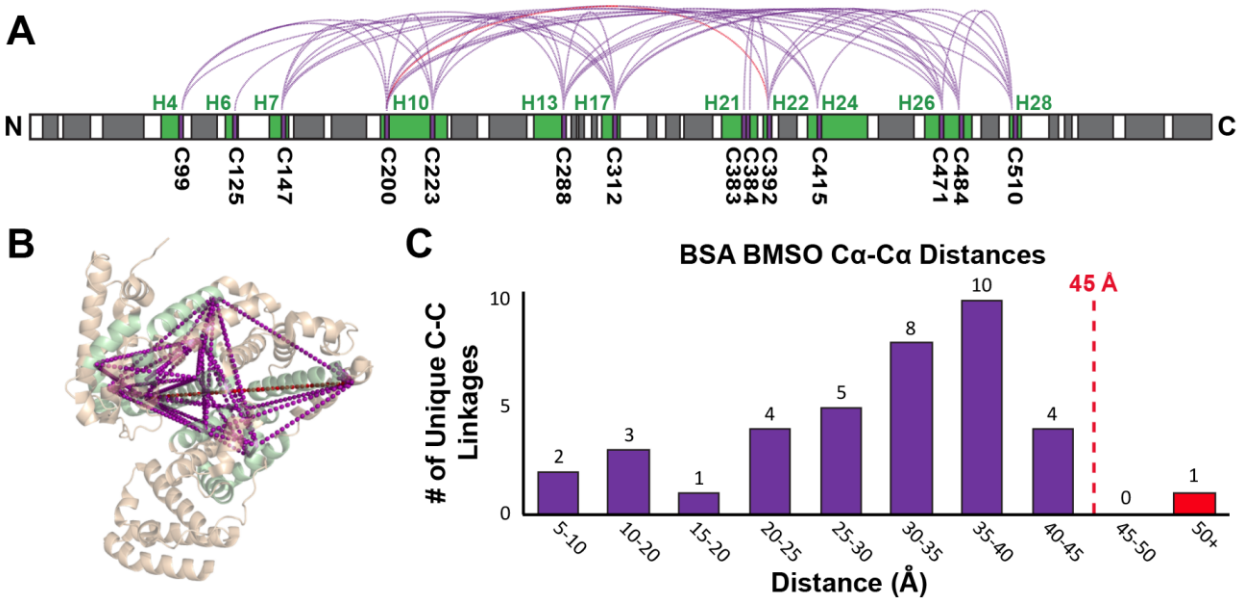

Figure 2.9. BMSO XL-MAP of BSA. (A) 2-D BMSO XL-map on BSA linear sequence. Helical secondary structures are designated by gray regions, green if containing cross-linked cysteines. (B) 3-D BMSO cross-link map on BSA crystal structure (PDB: 4F5S). Helices containing cross-linked cysteines are shown in green. (C) The distribution plot of identified C-C linkages vs. their spatial distances showing the number of linkages within (purple) and greater than (red) the expected distance constraint (< 45Å).

FIGURE 2.10

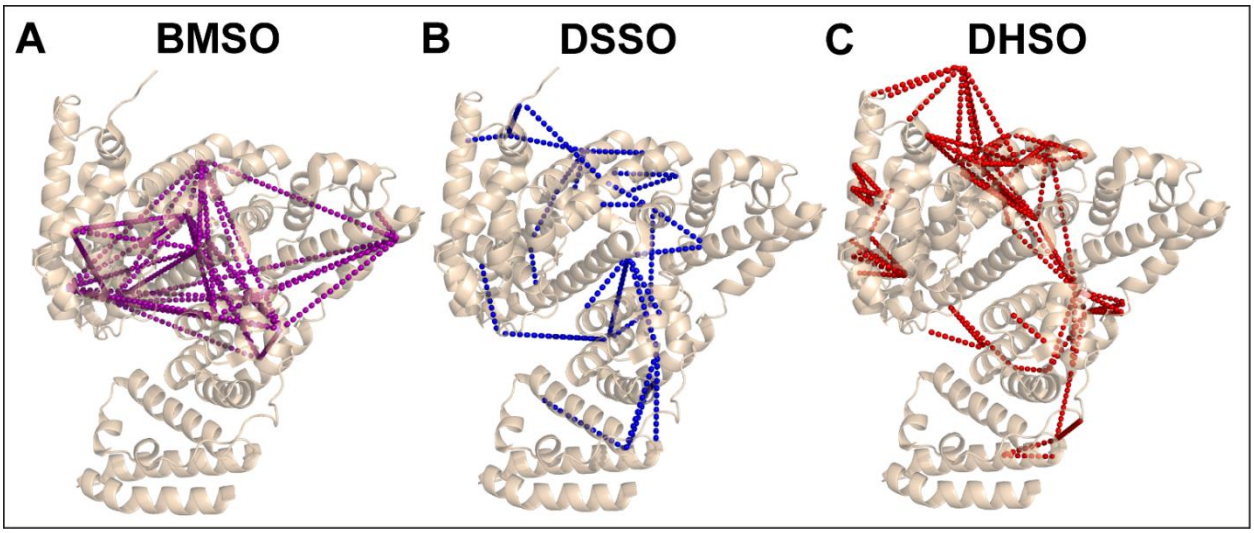

Figure 2.10. Comparison of XL-MS maps on BSA crystal structure (PDB: 4F5S) using (A) BMSO (purple). (B) DSSO (blue). (C) DHSO (red).

FIGURE 3.1

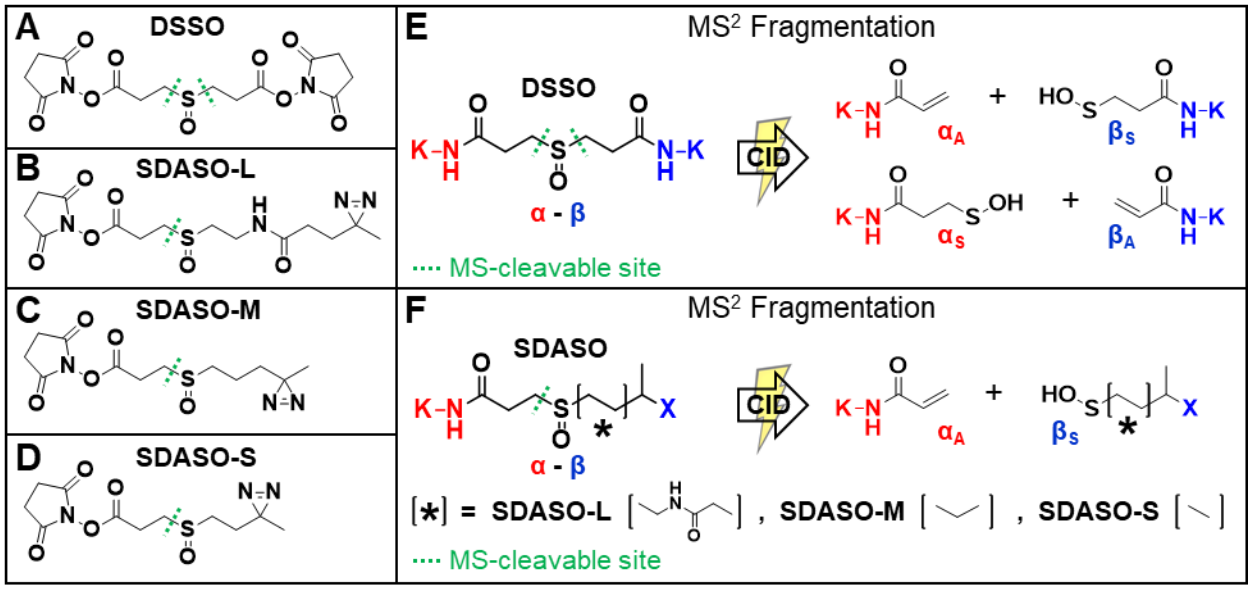

Figure 3.1. MS^2 fragmentation characteristics of sulfoxide-containing MS-cleavable cross-linkers. Molecular structures of (A) DSSO, (B) SDASO-L, (C) SDASO-M and (D) SDASO-S. (E) MS^2 fragmentation of a DSSO inter-linked peptide [α-β], representing the characteristics of sulfoxide-containing MS-cleavable cross-linkers with symmetric structures. CID cleavage of either one of the two MS-cleavable C-S bonds physically separates α and β peptide constituents into single peptide chains modified with either alkene (A) (i.e. α_A, β_A) or sulfenic acid (S) (i.e. α_S, β_S) moieties, the two complementary remnants of the cross-linker after cleavage. (F) MS^2 fragmentation of a SDASO inter-linked peptide [α-β], signifying the characteristics of sulfoxide-containing MS-cleavable heterobifunctional NHS-diazirine cross-linkers, namely, SDASO-L, -M and -S linkers (B-D). CID cleavage of the single MS-cleavable C-S bond in SDASO cross-linked peptides produces only one pair of cross-link fragment ions, α_A/β_S, in which α peptide is labeled by NHS ester, and β peptide is labeled by diazirine.

FIGURE 3.2

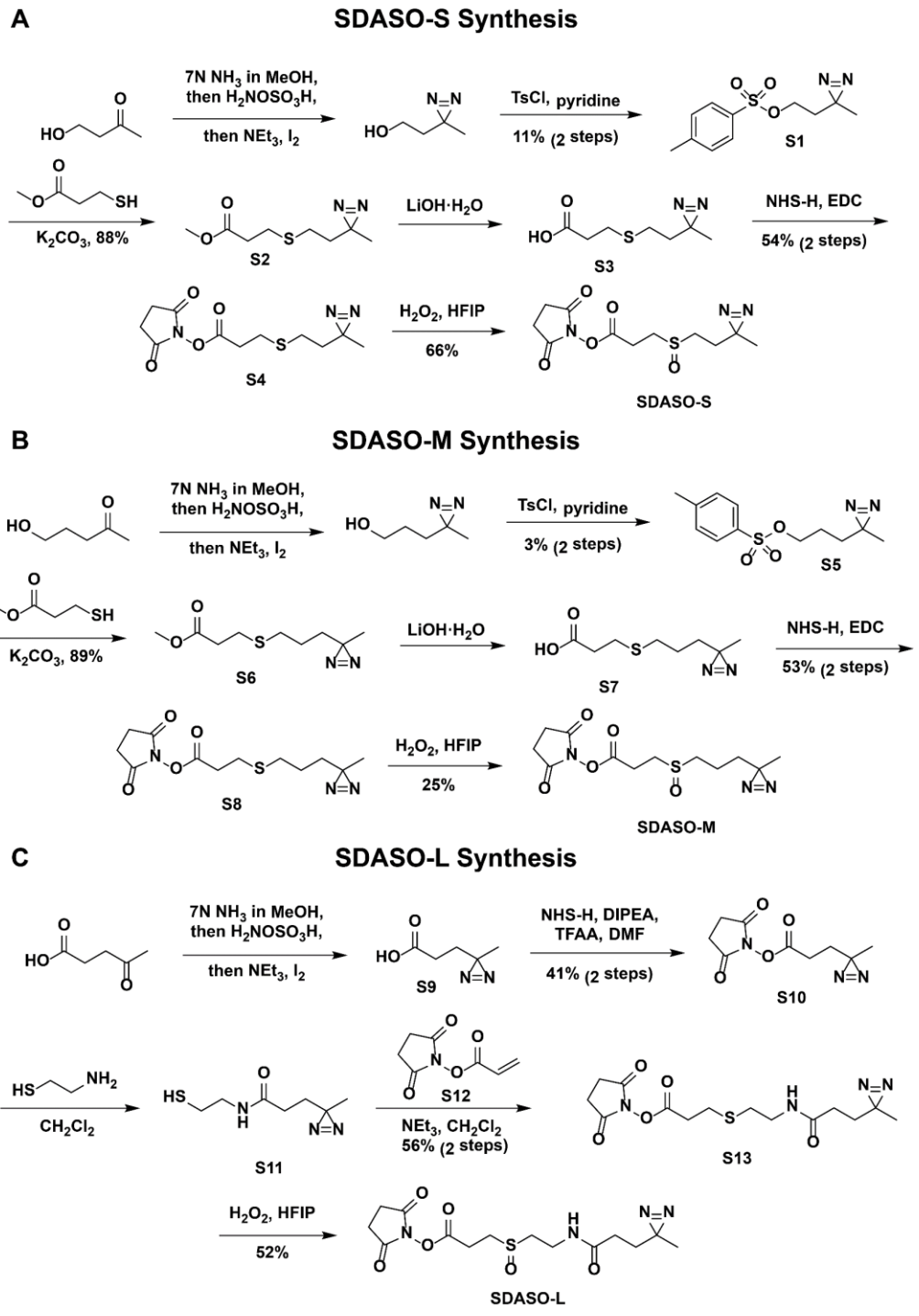

Figure 3.2. Synthesis pathways of the SDASO linkers: (A) SDASO-S. (B) SDASO-M. (C) SDASO-L.

FIGURE 3.3

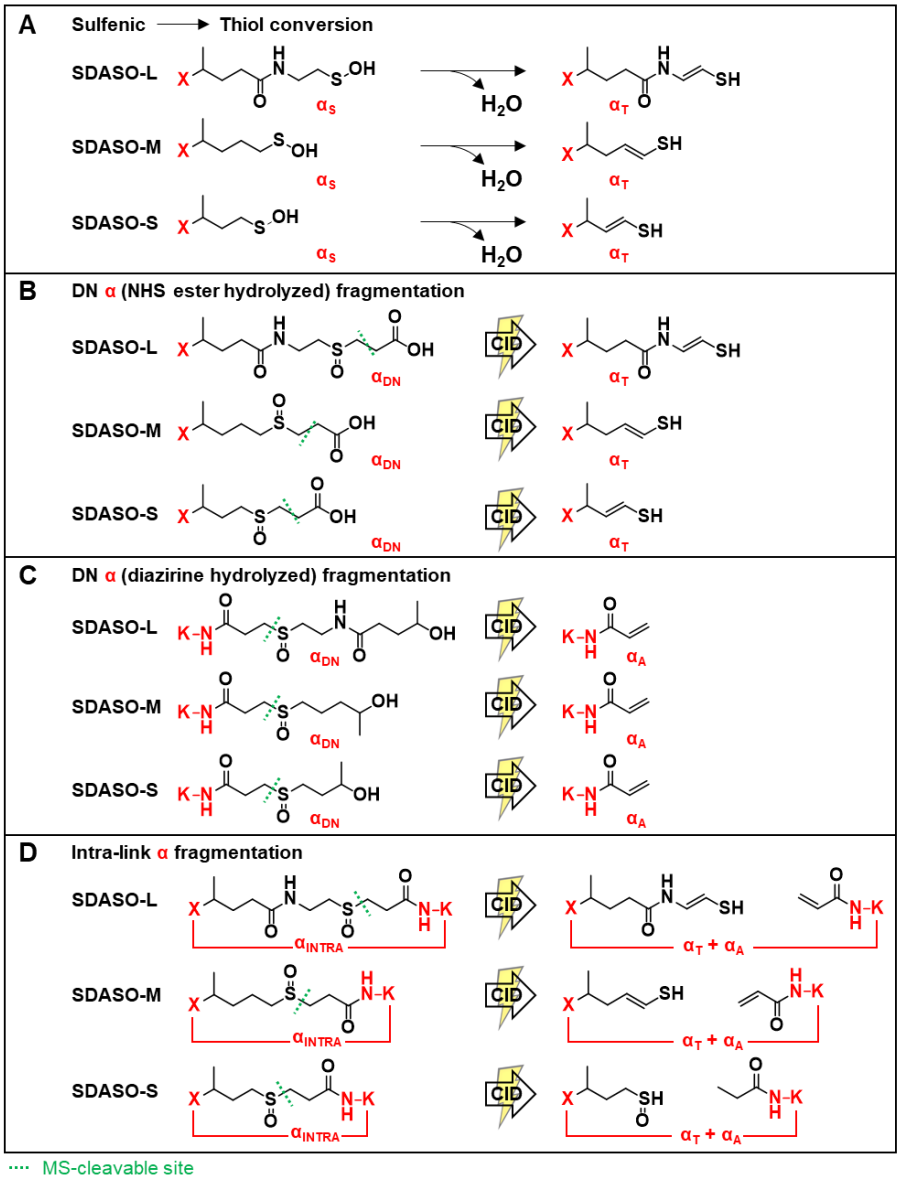

Figure 3.3. MS^2 fragmentation characteristics of SDASO linkers. (A) The conversion of sulfenic acid modified fragment α_S to unsaturated thiol modified fragment α_T. During MS^2-CID analysis, the sulfenic acid moiety loses water ($-H_2O$) to form the more stable unsaturated thiol (T) moiety, which is often detected as the dominant form. (B-C) Predicted MS^2 fragmentation of SDASO -L, -M and -S dead-end modified peptides (α_{DN}). Two types of dead-end products can be formed for SDASO linkers, depending on which reactive group is hydrolyzed. The products (α_{DN}) with a hydrolyzed NHS ester (B) or a hydrolyzed diazirine end (C) are illustrated. During MS^2-CID, the former only generates a sulfenic acid modified fragment α_S, whereas the latter only produces an alkene modified fragment α_A. (C) Predicted MS^2 fragmentation of SDASO-L, -M and -S intra-linked peptides (α_{intra}), yielding one fragment containing both alkene and sulfenic acid modifications (α_{A+S}).

FIGURE 3.4

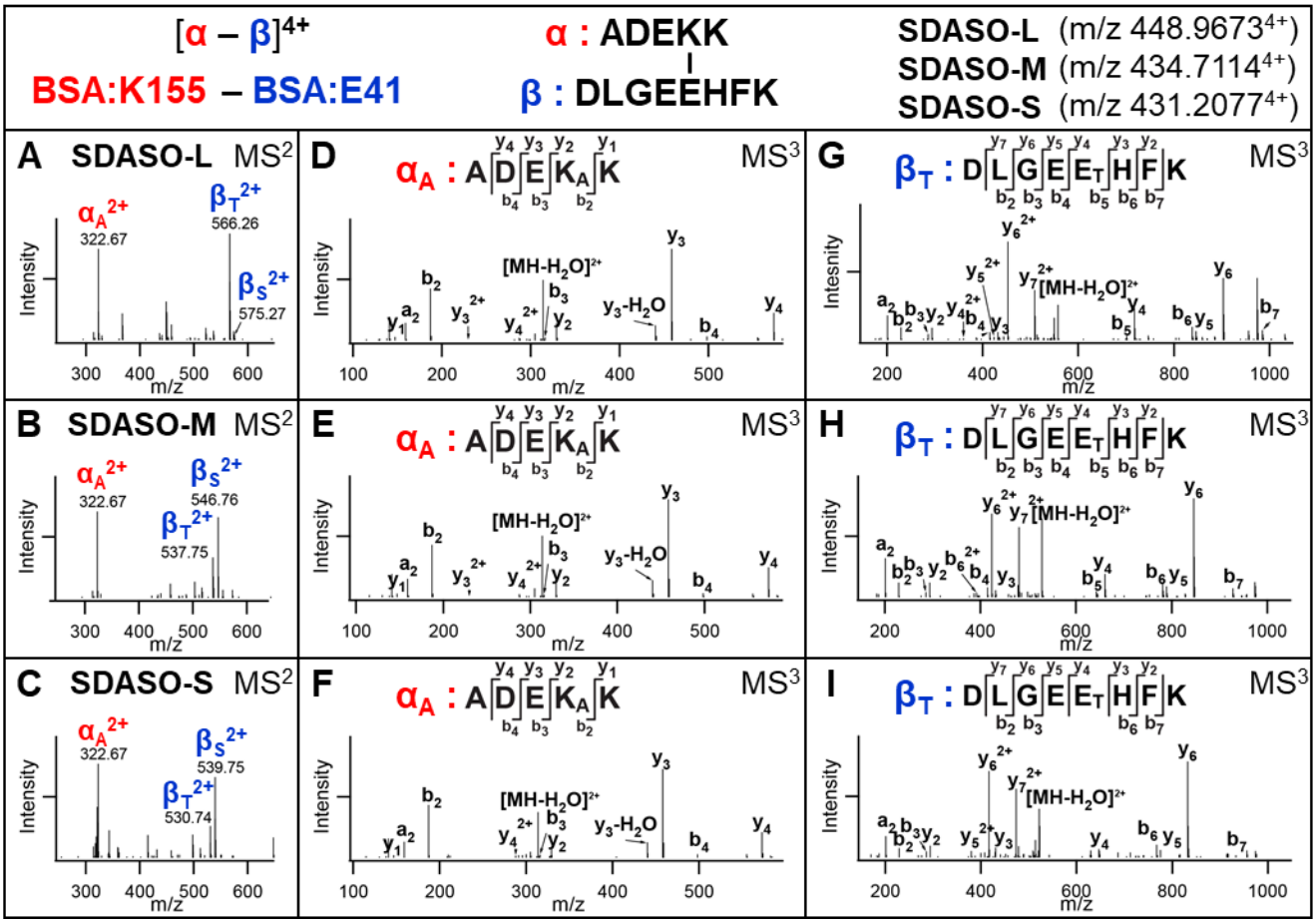

Figure 3.4. MS^n analyses of representative SDASO-L, SDASO-M and SDASO-S inter-linked peptides of BSA. MS^1 analyses determined the parent masses of the same peptides (α-β) cross-linked by SDASO-L (m/z 448.9673^{4+}), SDASO-M (m/z 434.7114^{4+}), SDASO-S (m/z 431.2077^{4+}), respectively. MS^2 spectra of the (A) SDASO-L, (B) SDASO-M, and (C) SDASO-S cross-linked peptides. MS^3 spectra of the SDASO-L MS^2 fragment ions: (D) α_A (m/z 322.67^{2+}) and (E) β_T (m/z 556.26^{2+}), the SDASO-M MS^2 fragment ions: (F) α_A (m/z 322.67^{2+}) and (G) β_T (m/z 537.75^{2+}), and the SDASO-S MS^2 fragment ions: (H) α_A (m/z 322.67^{2+}) and (I) β_T (m/z 530.74^{2+}). The selected BSA cross-linked peptide was identified as ^{152}ADEKK^{156} inter-linked to ^{37}DLGEEHFK^{44} by MS^3 analyses (A-I), in which the K155-E41 linkage was determined.

FIGURE 3.5

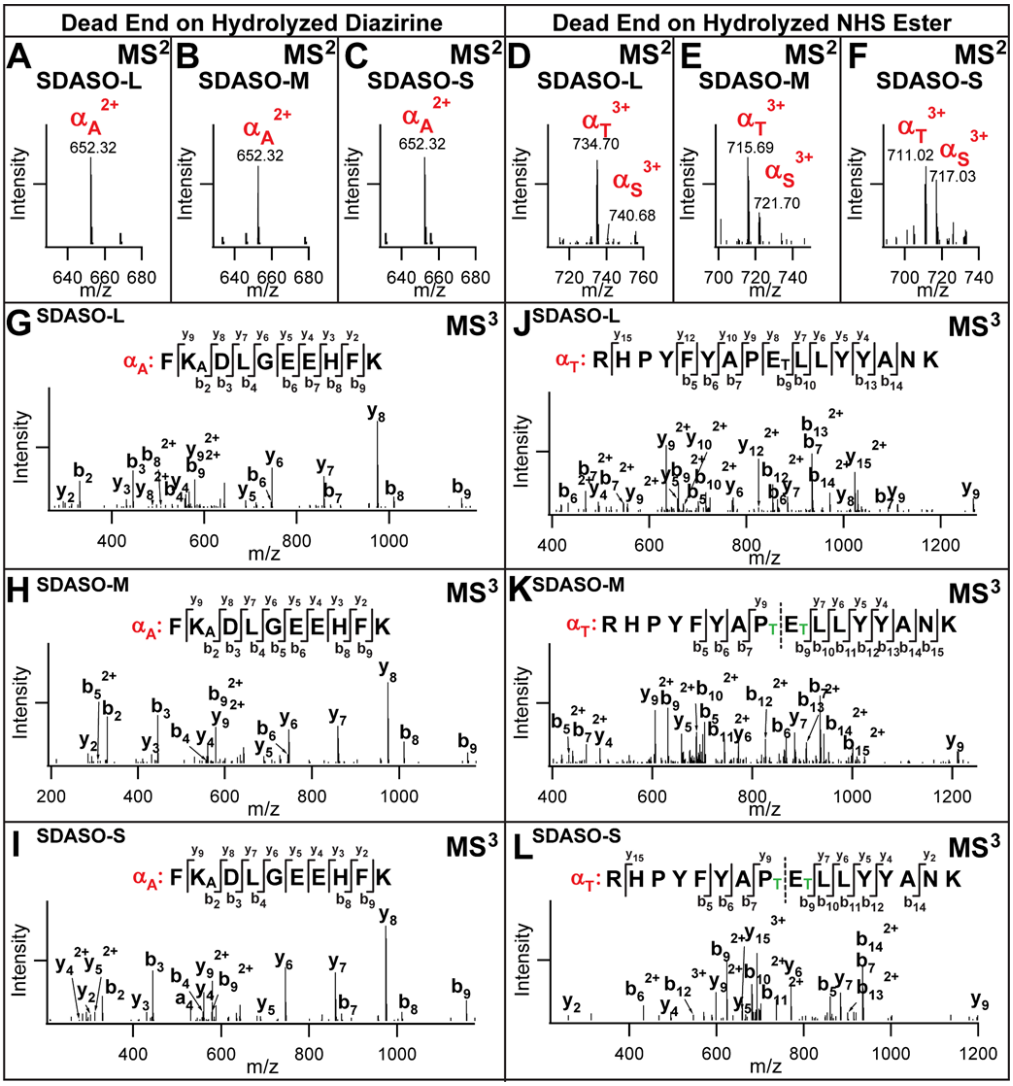

Figure 3.5. MS^n analyses of representative SDASO -L, -M and -S dead-end modified peptides of BSA. MS^2 spectra of the diazirine dead-end modified peptides α_{DN}: (A) SDASO-L: m/z 499.5673^{3+}, (B) SDASO-M: m/z 480.2411^{3+}, (C) SDASO-S: m/z 475.8982^{3+}. MS^2 spectra of the NHS ester dead-end modified peptides α_{DN} of BSA: (D) SDASO-L: m/z 765.0415^{3+}, (E) SDASO-M: m/z 745.7052^{3+}; (F) SDASO-S: m/z 741.0337^{3+}. MS^3 spectra of (G) α_A (m/z 652.32^{2+}) detected in (A), and (J) α_T (m/z 734.70^{2+}) detected in (D) from SDASO-L dead-end modified peptides. MS^3 spectra of (H) α_A (m/z 652.32^{2+}) detected in (B), and α_T (m/z 715.69^{2+}) detected in (E) from SDASO-M dead-end modified peptides. MS^3 spectra of (H) α_A (m/z 652.32^{2+}) detected in (C), and α_T (m/z 711.02^{2+}) detected in (F) from SDASO-S dead-end modified peptides. The diazirine dead-end modified peptides of BSA were determined as ^{35}FKDLGEEHFK^{44}, in which K36 was modified for all SDASO linkers (G, H, and I). The NHS ester dead-end modified peptides of BSA were identified as ^{168}RHPYFYAPELLYYANK^{183} (J), in which E176 was modified for SDASO-L, and P175 or E176 were modified for SDASO-M and SDASO-S (K and L).

FIGURE 3.6

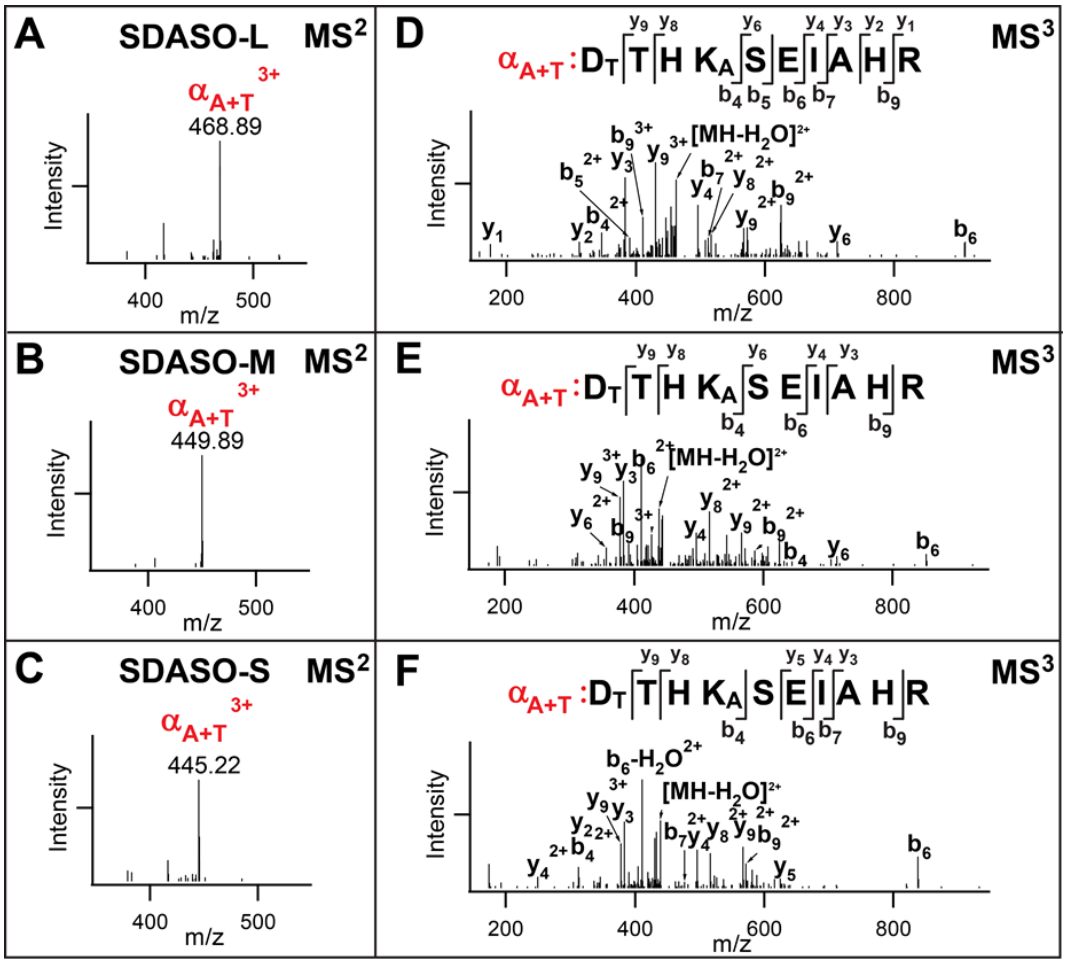

Figure 3.6. MS^n analyses of representative SDASO -L, -M and -S intra-linked peptides of BSA. MS^2 spectra of the intra-linked α_{intra} peptides (A) SDASO-L: m/z 474.8992^{3+}, (B) SDASO-M: m/z 455.8924^{3+}, (C) SDASO-S: m/z 451.2183^{3+}. MS^3 spectra of: (D) SDASO-L MS^2 fragment ion α_{A+T} (m/z 468.89^{3+}), (E) SDASO-M MS^2 fragment ion α_{A+T} (m/z 449.89^{3+}), (F) SDASO-S MS^2 fragment ion α_{A+T} (m/z 445.22^{3+}). The intra-linked peptides were determined as ^{25}DTHKSEIAHR^{34} with D25 linked to K28 for all three linkers.

FIGURE 3.7

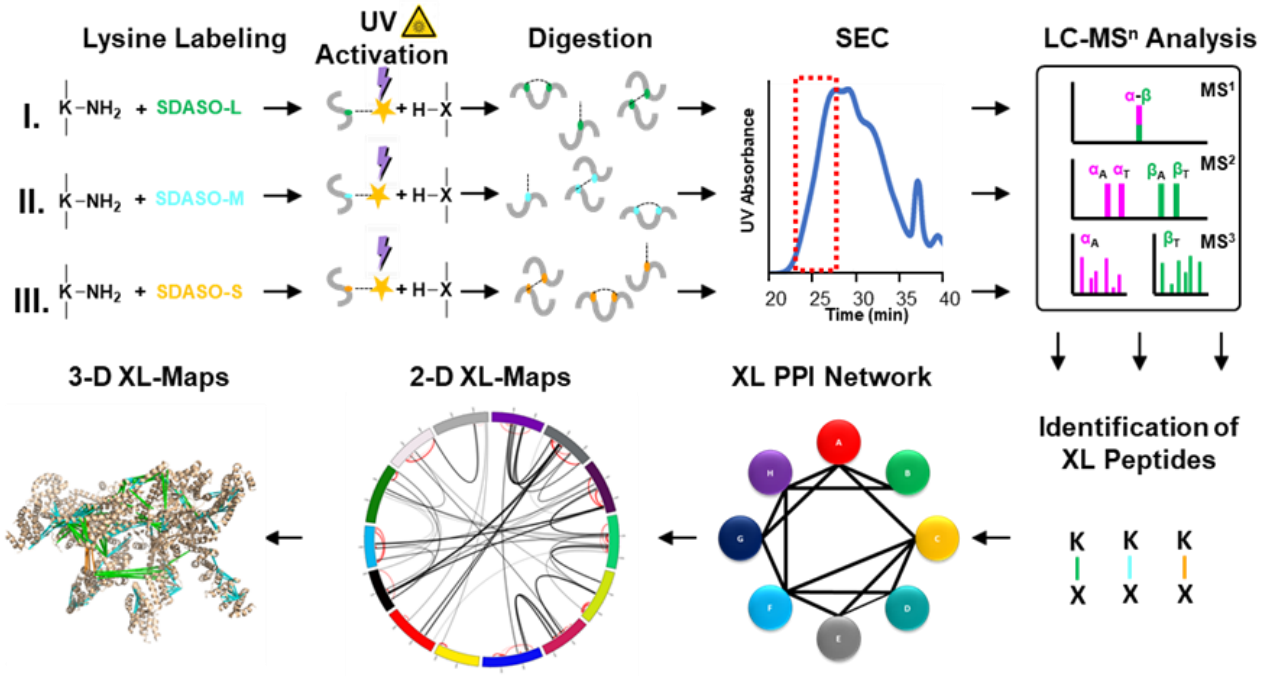

Figure 3.7. The SDASO XL-MS workflow. SDASO cross-linking involves two steps: 1) lysine labeling by NHS ester; 2) photoactivated diazirine cross-linking of any AAs upon UV irradiation. path I-SDASO-L, path II-SDASO-M, and path III-SDASO-S.

FIGURE 3.8

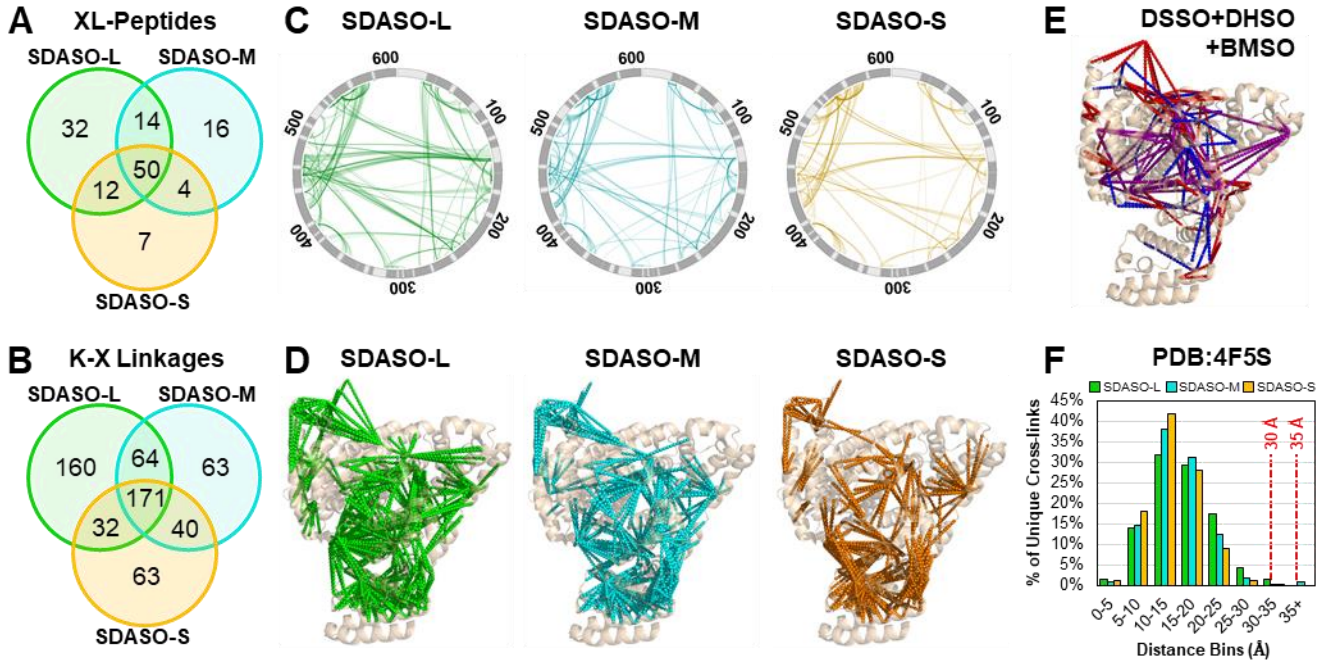

Figure 3.8. Comparisons of BSA cross-link data by the three SDASO linkers. Comparisons of (A) cross-linked peptide sequences and (B) residue-to-residue linkages of BSA obtained from SDASO-L, -M, and -S XL-MS experiments. (C) Circular 2-D SDASO XL-maps of BSA based on SDASO -L, -M and-S cross-links respectively. Helical secondary structures are designated by dark gray regions. (D) 3-D SDASO XL-maps of BSA on its crystal structure (PDB: 4F5S) based on SDASO-L, SDASO-M, and SDASO-S cross-links, respectively. (E) 3-D XL-map of BSA (PDB: 4F5S) generated based on the combined cross-links from DSSO (blue)+DHSO (red)+BMSO (purple) XL-MS experiments [Gutierrez, 2016; Gutierrez, 2018}. (F) Distance distribution plots of the identified SDASO cross-links to the BSA structure (PDB: 4F5S) (SDASO-L: ≤35Å, SDASO-M and -S: ≤ 30Å). Note: Colors schemes represent specific linkers: SDASO-L: light green, SDASO-M: light blue, and SDASO-S: gold orange.

FIGURE 3.9

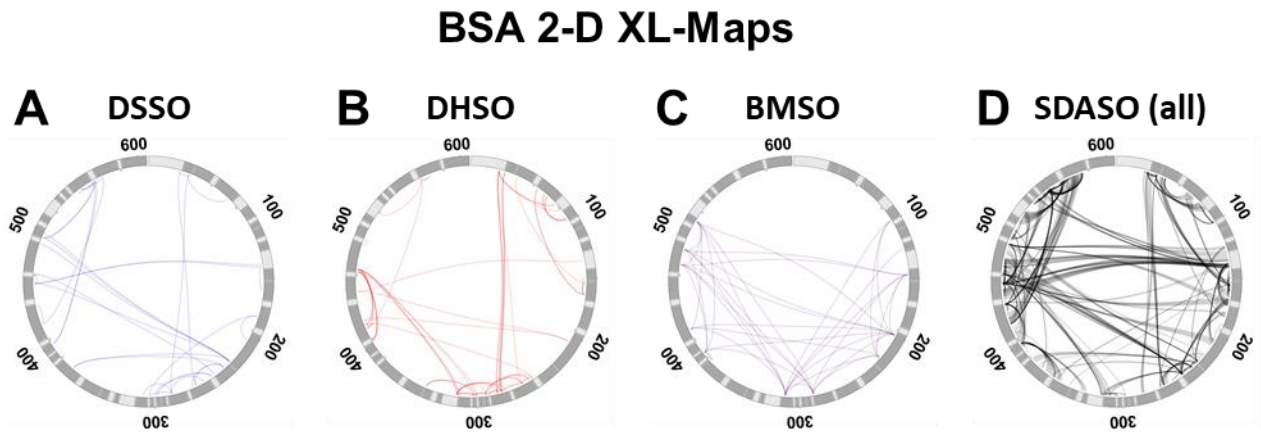

Figure 3.9. Circular 2-D XL-maps of BSA. The maps were generated based on cross-links of BSA using (A) DSSO [60], (B) DHSO [140], (C) BMSO [66] and (D) the three SDASO linkers in this work.

FIGURE 3.10

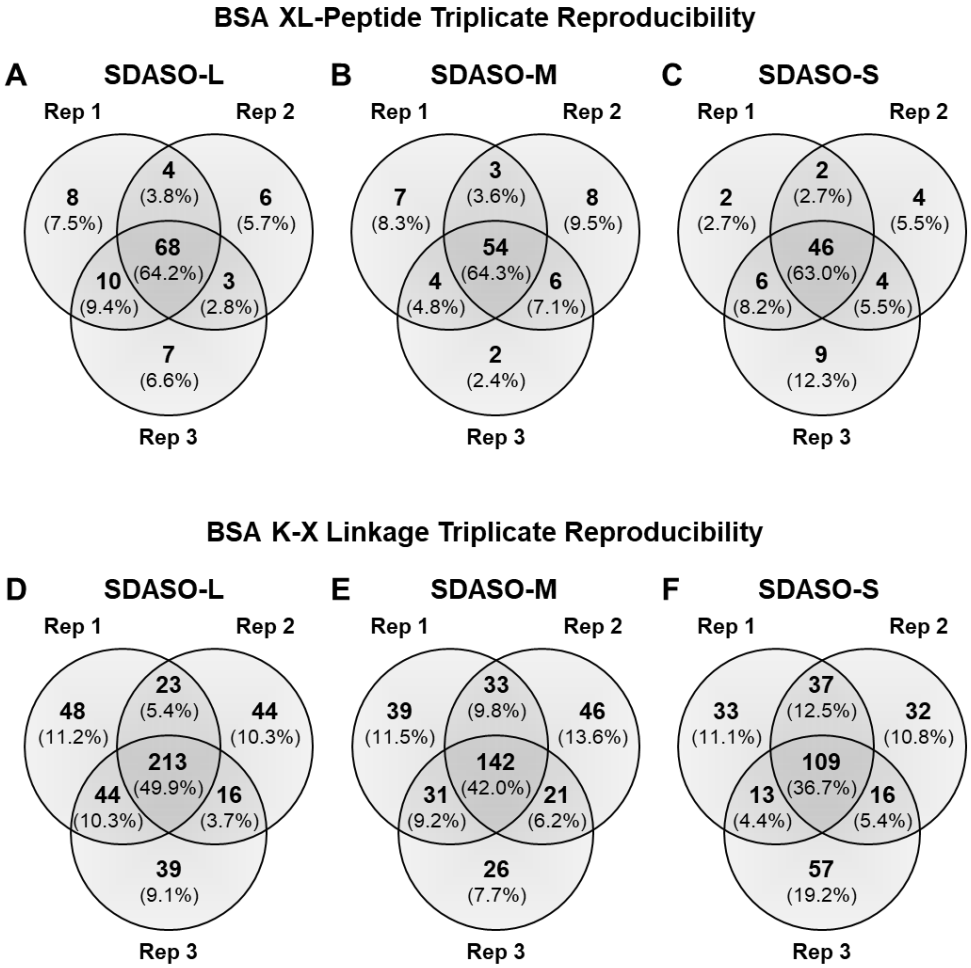

Figure 3.10. Reproducibility of SDASO XL-MS data for BSA. Comparisons of SDASO cross-linked peptide sequences among the three biological replicates for (A) SDASO-L, (B) SDASO-M, and (C) SDASO-S. Comparisons of SDASO K-X linkages among the three biological replicates for (D) SDASO-L, (E) SDASO-M, and (F) SDASO-S.

FIGURE 3.11

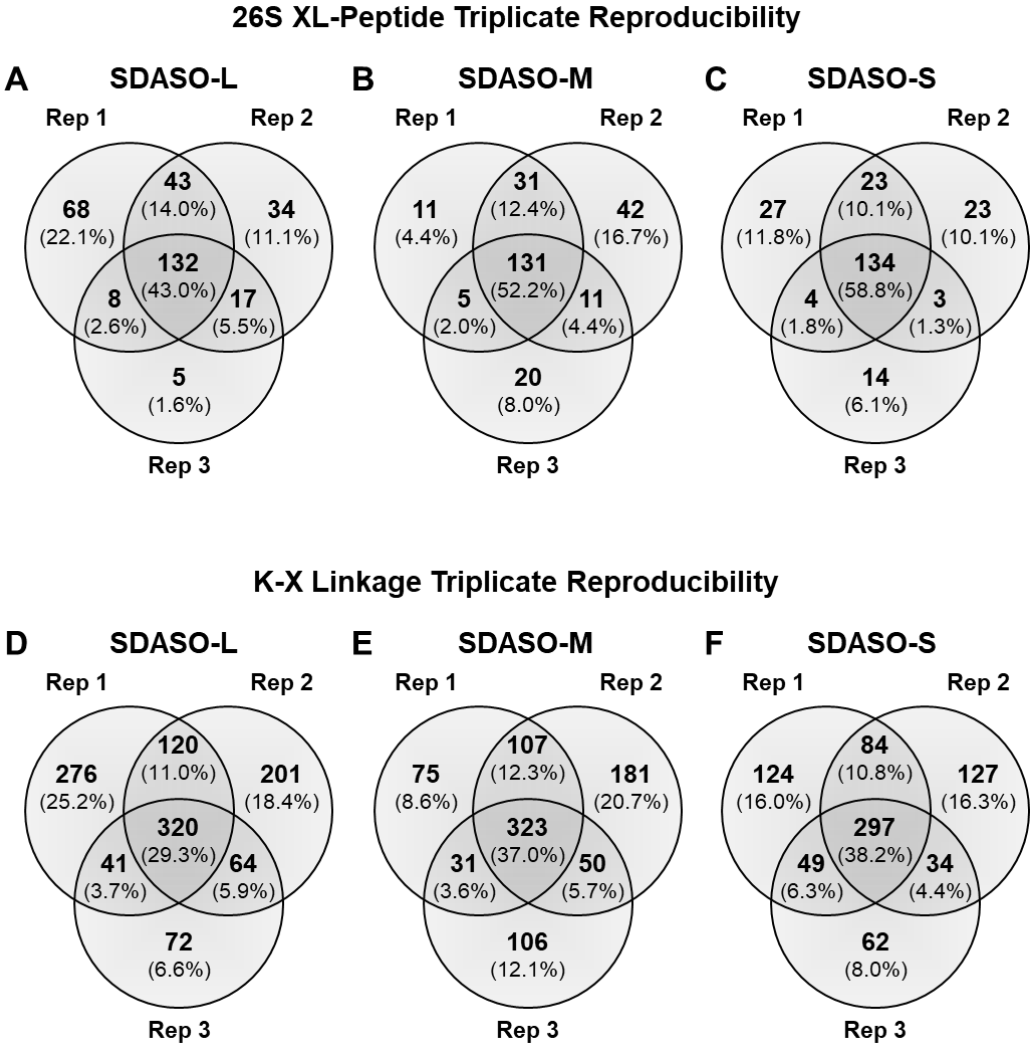

Figure 3.11. Reproducibility of SDASO XL-MS data for the 26S proteasome. Comparisons of SDASO cross-linked peptide sequences among the three biological replicates for (A) SDASO-L, (B) SDASO-M, and (C) SDASO-S. Comparisons of SDASO K-X linkages among the three biological replicates for (D) SDASO-L, (E) SDASO-M, and (F) SDASO-S. Note: the XL-MS data were generated from tryptic digests only.

FIGURE 3.12

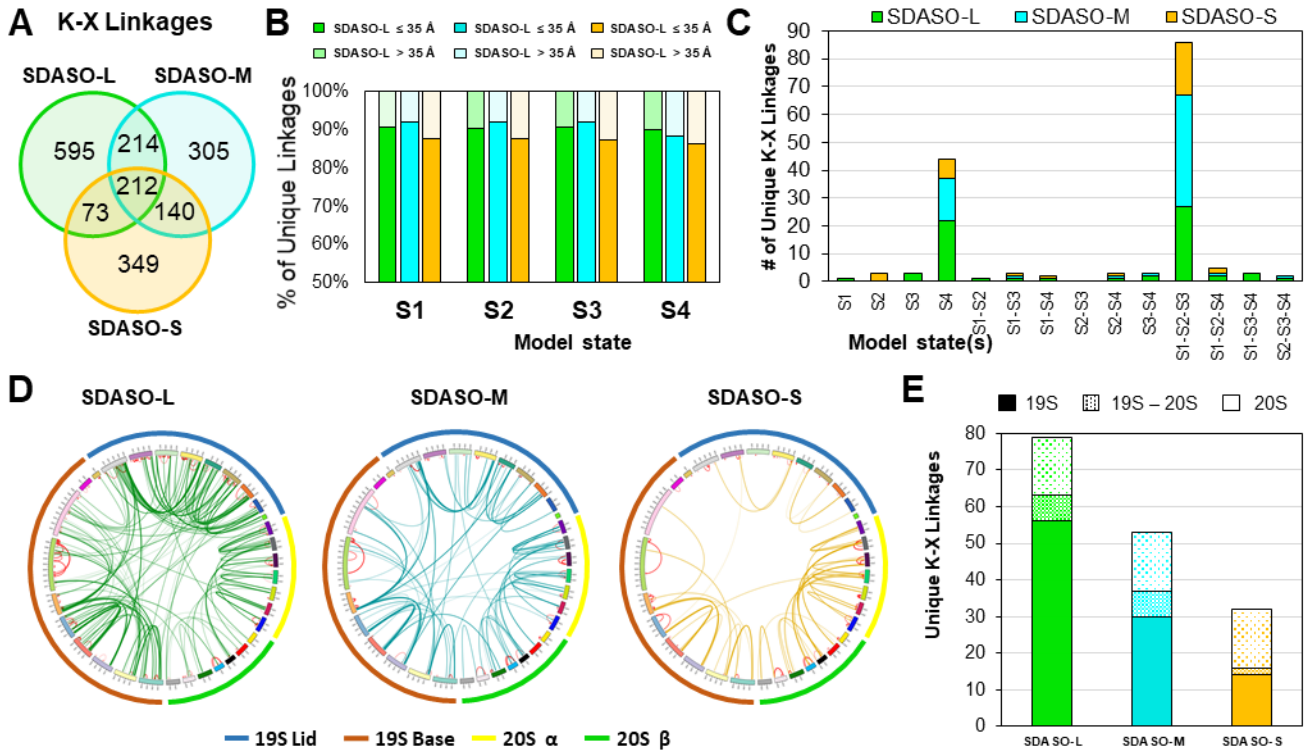

Figure 3.12. SDASO XL-MS data summary of the yeast 26S proteasome. (A) Comparisons of cross-linked peptide sequences and residue-to-residue linkages of the 26S proteasome obtained from SDASO-L, -M, and -S XL-MS experiments. (B) Respective distance satisfaction rates of SDASO -L, -M, and -S cross-links (SDASO-L: ≤ 35Å, SDASO-M and -S: ≤ 30Å) mapped onto the 4 model states of the yeast 26S (PDB: 4CR2 (s1), 4CR3 (s2), 4CR4 (s3), and 5MPD(s4)). (C) Distribution of unique state-specific K-X linkages of SDASO-L, -M, and -S across the 14 possible combinations of one, two or three out of the four states. (D) Circular 2-D XL-maps of the yeast 26S Proteasome for SDASO-L, -M, -S linkers. Intra-subunit linkages are shown red and inter-subunit linkages are illustrated based on the linkers: light green (SDASO-L), light blue (SDASO-M), or gold orange (SDASO-S). Outer circle represents the subunits within the two subcomplexes of the 26S proteasome, i.e. the 19S RP (Lid (blue) and base (dark orange)) and 20S CP (α ring (yellow) and β ring (green)). Proteasome subunits are color coded as described (Supplementary Table 3.11). (E) Distributions of SDASO-L, -M, and -S cross-links corresponding to three categories of inter-subunit interactions: 19S-19S, 19S-20S and 20S-20S (Table 3.11). Note: linker-specific color schemes: SDASO-L: light green, SDASO-M: light blue, and SDASO-S: gold orange.

FIGURE 3.13

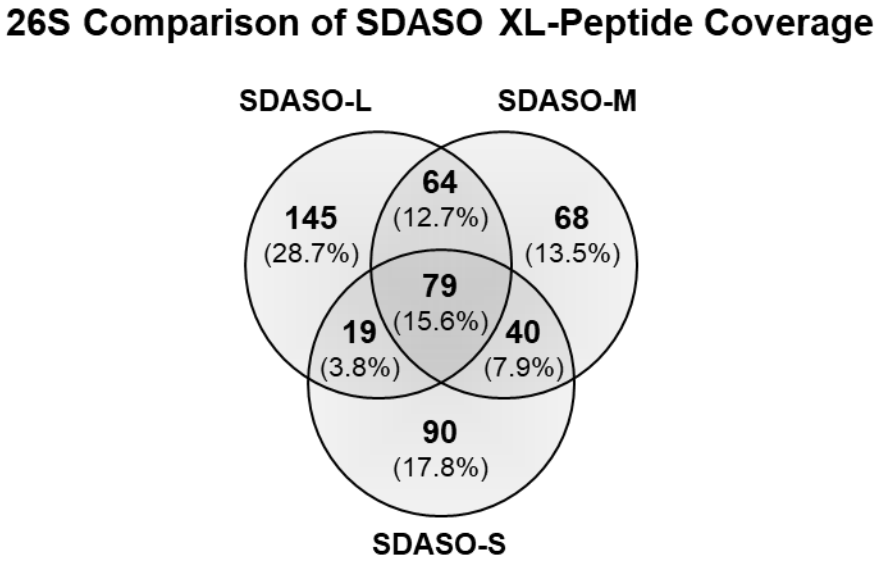

Figure 3.13. Comparison of SDASO XL-MS data for the 26S proteasome. Overlaps of the identified (A) cross-linked peptide sequences using SDASO-L, SDASO-M, and SDASO-S respectively. Note: the XL-MS data were generated from tryptic digests only.

FIGURE 3.14

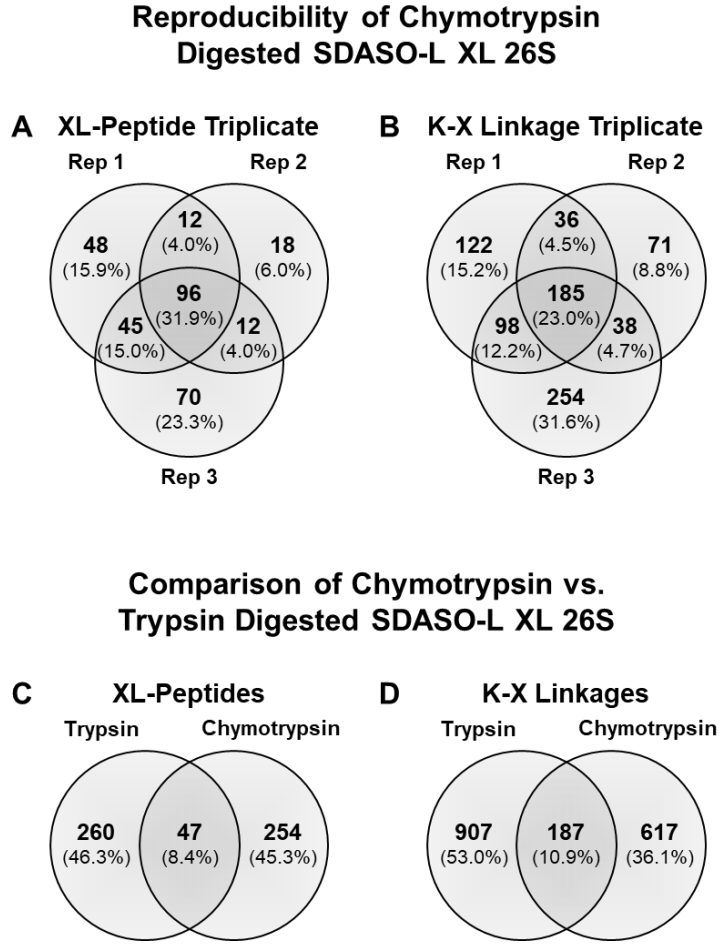

Figure 3.14. Reproducibility of SDASO-L XL-MS data from chymotryptic digests of the 26S proteasome and comparison of data from trypsin and chymotrypsin digests of the 26S proteasome. Comparisons of SDASO-L (A) cross-linked peptide sequences and (B) K-X linkages among the three biological replicates of chymotryptic digests. Overlaps of the identified SDASO-L (C) cross-linked peptide sequences and (D) K-X linkages using trypsin and chymotrypsin digestions respectively. SDASO-L: tryptic digests; SDASO-Lc: chymotryptic digests.

FIGURE 3.15

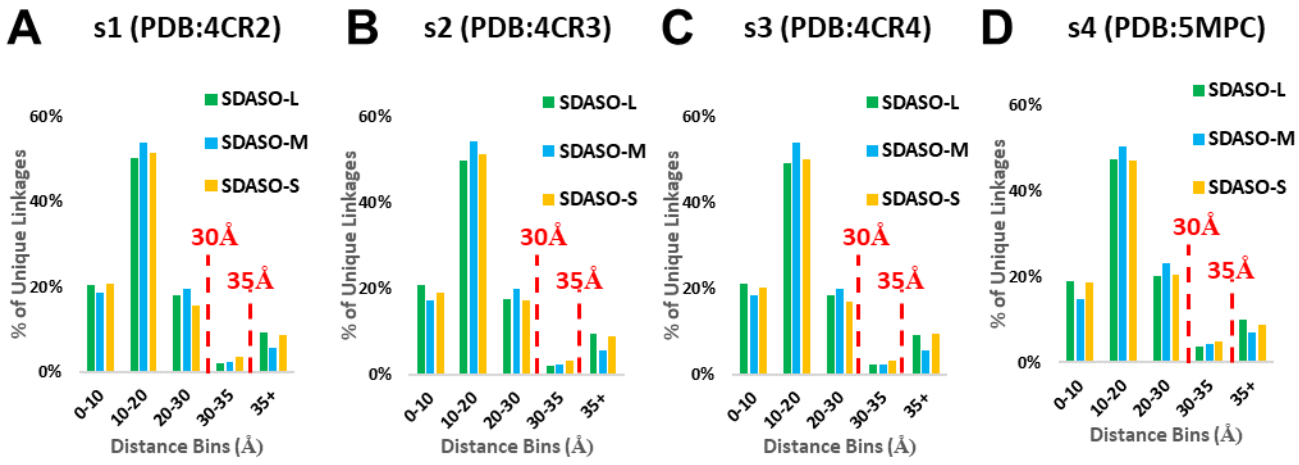

Figure 3.15. Distance distribution plots of SDASO cross-links mapped to the four known states of the yeast 26S proteasome structures (s1-s4). Respective cross-link distance distribution plots for SDASO-L, -M and -S cross-link data mapped onto (A) s1 (PDB: 4CR2), (B) s2 (PDB:4CR3), (C) s3 (PDB:4CR4) and (D) s4 (PDB:5MPC). Distance satisfaction thresholds are ≤30Å for SDASO-M and –S, and ≤35Å for SDASO-L.

FIGURE 3.16

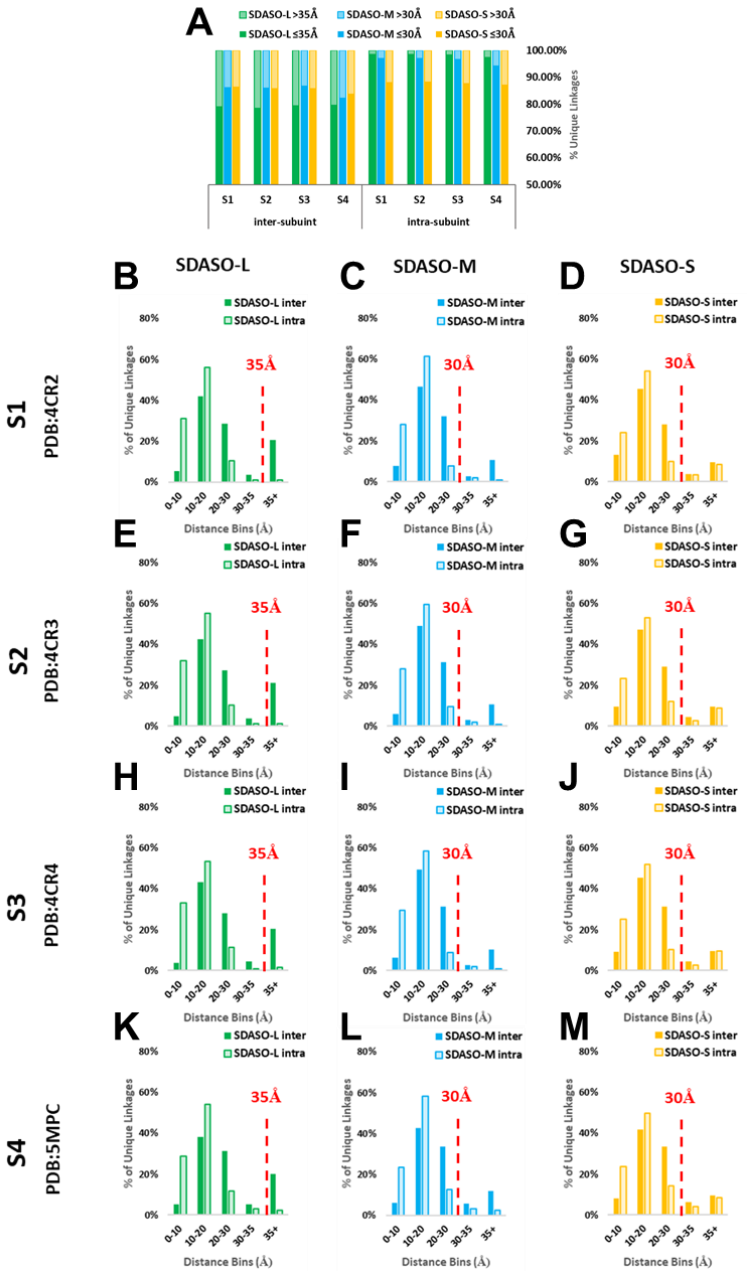

Figure 3.16. Distance distribution plots of the inter-subunit and intra-subunit K-X linkages mapped to the yeast 26S proteasome structures (s1-s4). (A) Average cross-link distance satisfaction rates across the four state models (s1: 4CR2, s2:4CR3, s3:4CR4, and s4:5MPD) for SDASO -L, -M, and –S, respectively. Respective distance distribution plots of SDASO inter-subunit and intra-subunit cross-links mapped onto (B-D) state s1 (PDB: 4CR2), (E-G) state s2 (PDB: 4CR3), (H-J) state s3 (PDB: 4CR4), and (K-M) state s4 (PDB: 5MPC) based on SDASO-L (B, E, H, K), SDASO-M (C, F, I, L) and SDASO-S (D, G, J, M) XL-MS data. Distance satisfaction thresholds are ≤30Å for SDASO-M and –S, and ≤35Å for SDASO-L.

FIGURE 3.17

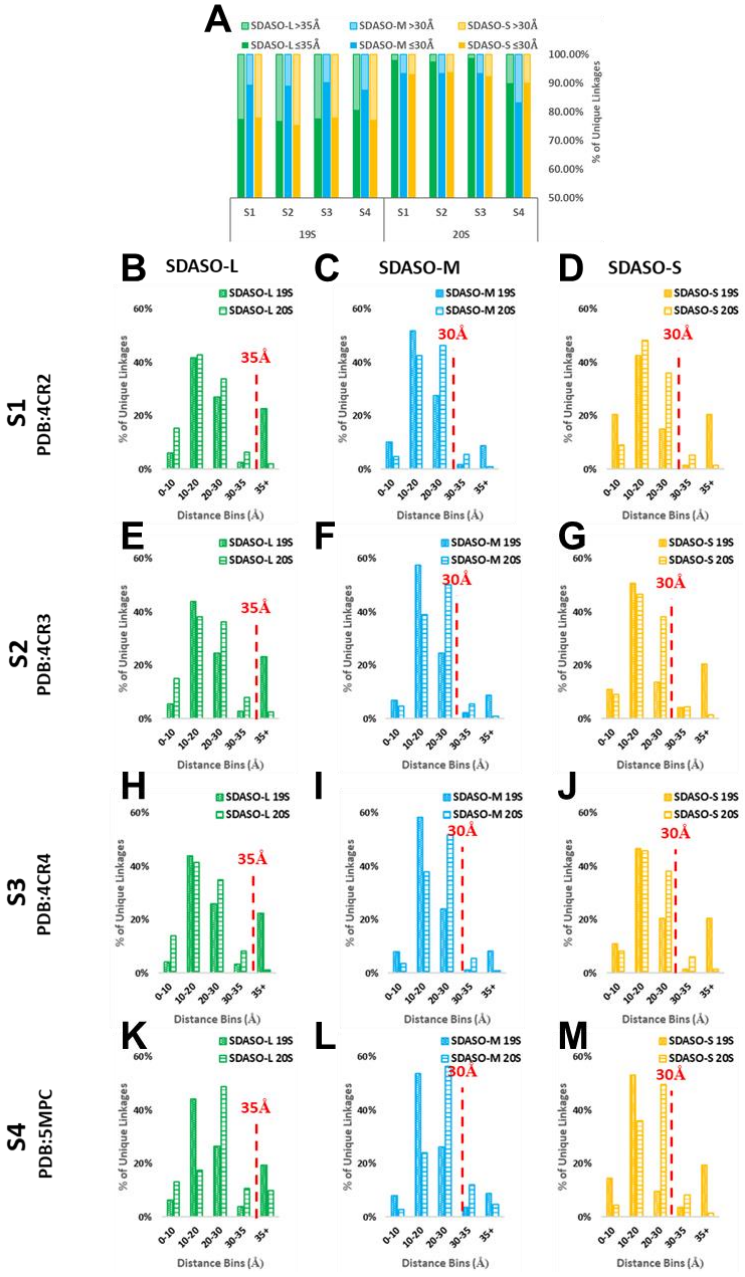

Figure 3.17 Distance distribution plots of the 19S RP and 20S CP K-X linkages mapped to the yeast 26S proteasome structures (s1-s4). (A) Average distance satisfaction rates across the four state models (s1: 4CR2, s2:4CR3, s3:4CR4, and s4:5MPD) for SDASO -L, -M, and -S K-X linkages of the 19S RP and 20S CP, respectively. Respective distance distribution plots of the 19S and 20S cross-links mapped onto (B-D) state s1 (PDB: 4CR2), (E-G) state s2 (PDB: 4CR3), (H-J) state s3 (PDB: 4CR4), and (K-M) state s4 (PDB: 5MPC) based on SDASO-L (B, E, H, K), SDASO-M (C, F, I, L) and SDASO-S (D, G, J, M) XL-MS data. Distance satisfaction thresholds are ≤30Å for SDASO-M and –S, and ≤35Å for SDASO-L.

FIGURE 3.18

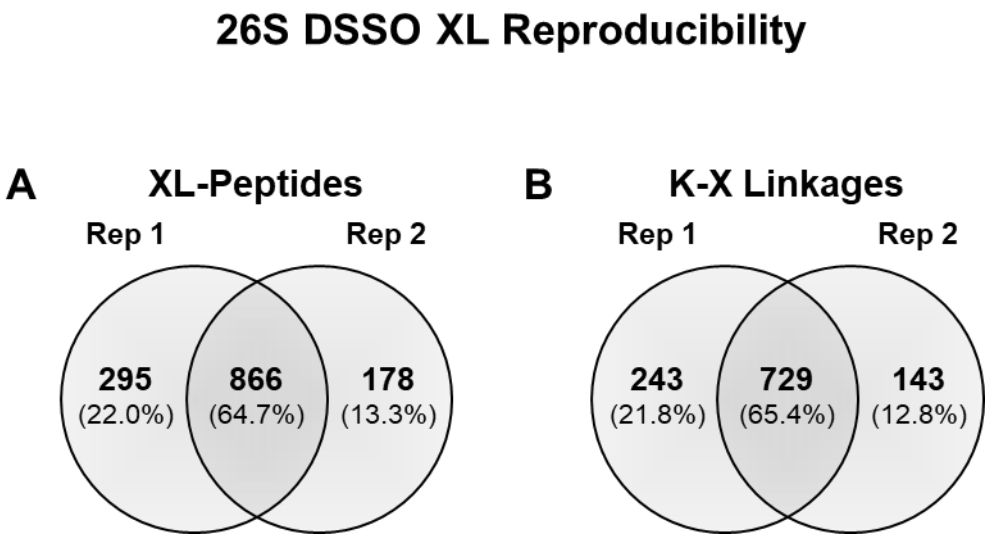

Figure 3.18. Reproducibility of DSSO XL-MS data for the 26S proteasome. Comparisons of DSSO (A) cross-linked peptide sequences and (B) K-K linkages among the two biological replicates of tryptic digests.

FIGURE 3.19

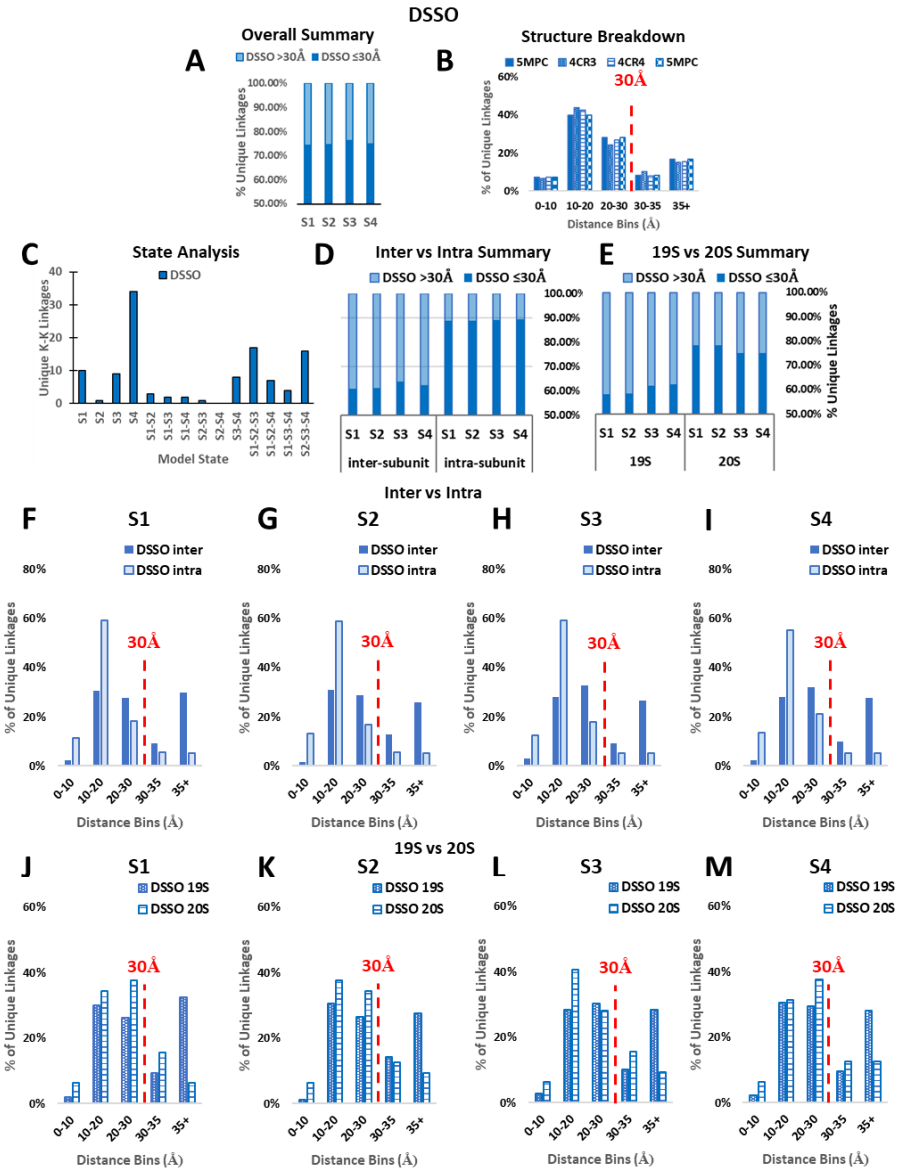

Figure 3.19. Distance distribution analysis of DSSO cross-links of the 26S proteasome. (A) Average distance satisfaction rates of DSSO cross-links across the four state models (s1: 4CR2, s2:4CR3, s3:4CR4, and s4:5MPD) (distance satisfaction threshold for DSSO cross-links ≤ 30Å). (B) Respective distance distribution plots of DSSO cross-links mapped to the four states of the 26S proteasome structures. (C) Distribution of state-specific K-K linkages among the 14 possible state combinations. (D) Average distance satisfaction rates of inter-subunit and intra-subunit K-K linkages across the four state models. (E) Average distance satisfaction rates of the 19S and 20S K-K linkages across the four state models. (F-I) Respective distance distribution plots of inter-subunit and intra-subunit DSSO cross-links mapped to the four state models: (F) s1 (PDB: 4CR2), (G) s2 (PDB: 4CR3), (H) s3 (PDB: 4CR4) and (I) s4 (PDB: 5MPC). (J-M) Respective distance distribution plots of the 19S and 20S cross-links mapped to the four state models: (J) s1 (PDB: 4CR2), (K) s2 (PDB: 4CR3), (L) s3 (PDB: 4CR4) and (M) s4 (PDB: 5MPC).

FIGURE 3.20

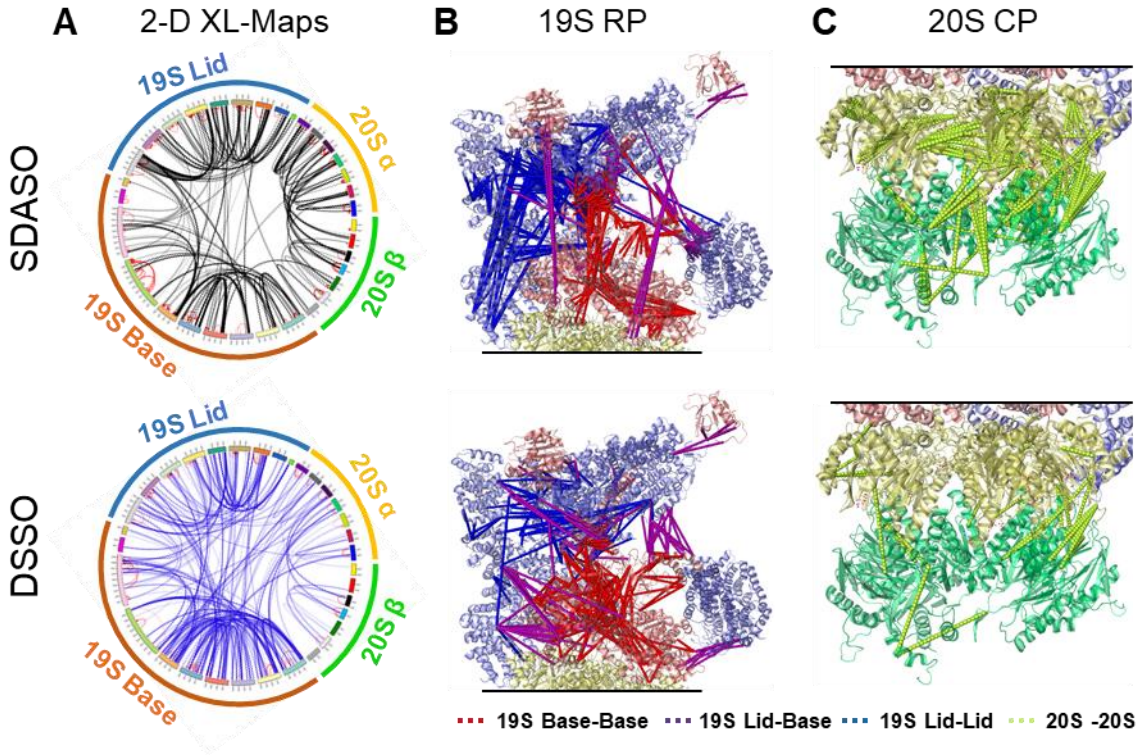

Figure 3.20. SDASO and DSSO XL-maps of the yeast 26S Proteasome. (A) Circular 2-D XL-maps of the 26S Proteasome derived from all three SDASO linkers (top) and DSSO (bottom). Intra-subunit linkages are shown red and inter-subunit linkages are colored black for SDASO and blue for DSSO. Outer circle represents the subunits within the two subcomplexes of the 26S proteasome, i.e. the 19S RP (Lid (blue) and base (dark orange)) and 20S CP (α ring (yellow) and β ring (green)). Subunits are color coordinated as shown in y Table 3.11. (B) 3D XL-maps of the 19S RP using SDASO (top) and DSSO (bottom) cross-links, in which 19S lid subunits are colored light blue and base subunits colored light red. Cross-links are also color coded: 19S lid-lid (blue lines), 19S base-base (red lines), 19S lid-base (purple). (C) 3D XL-maps of the 20S CP based on SDASO (top) and DSSO (bottom) cross-links, in which 20S α subunits are colored as light yellow, 20S β subunits as aqua green and 20S linkages as lime green. Note: high-resolution structure of the yeast 26S proteasome (PDB: 4CR2 (s1) was used for the maps in (B) and (C).

FIGURE 3.21

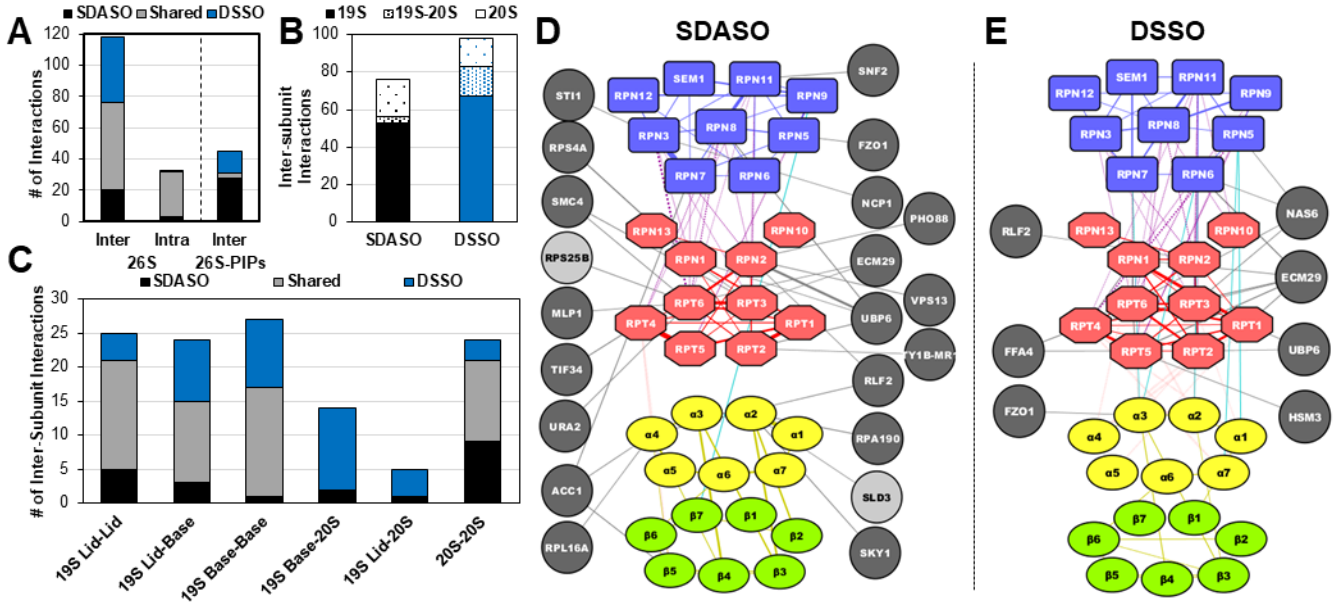

Figure 3.21. XL-PPI Analysis of the Yeast 26S proteasome. (A) Comparison of the total number of SDASO and DSSO XL-PPIs (inter- and intra-subunit) within the 26S Proteasome itself and with proteasome interaction proteins (PIPs). (B) Distribution of the total number of SDASO and DSSO inter-subunit interactions within the 26S proteasome. (C) Comparison of the distribution of SDASO and DSSO XL-PPIs among the six types of inter- and intra-subcomplex interactions within the 26S proteasome. (D) SDASO and (E) DSSO XL-PPI networks of the 26S proteasome and its interacting proteins, in which protein nodes are colored as follows: 19S lid subunits (light blue), 19S base subunits (light red), 20S α subunits (light yellow), 20S β subunits (green), known PIPs (dark gray), and novel PIPs (light gray). The edges are colored as: 19S lid-lid (blue), 19S base-base (red), 19S lid-base (purple), 19S lid-20S core (cyan), 19S base-20S core (pink), 20S-20S (gold), and 26S-PIP (black).

FIGURE 3.22

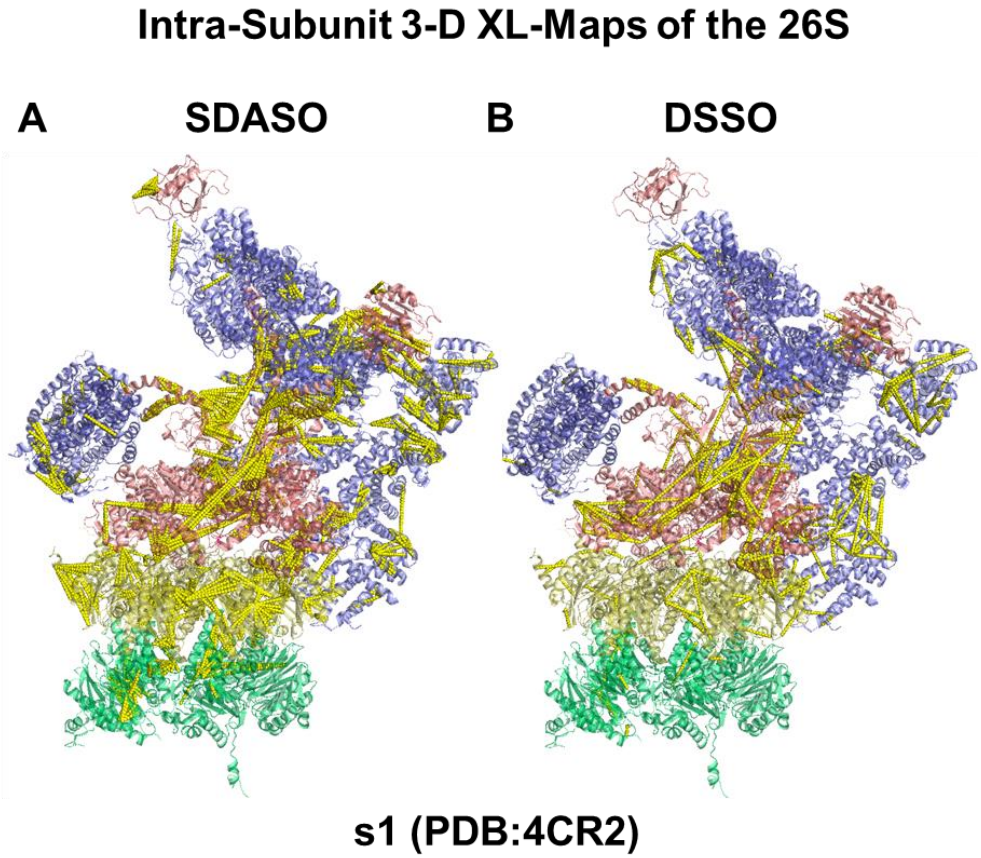

Figure 3.22. 3-D XL-Maps of SDASO and DSSO intra-subunit linkages of the 26S proteasome. (A) SDASO and (B) DSSO intra-subunit linkages (shown in yellow) mapped to s1 (PDB:4CR2).

FIGURE 3.23

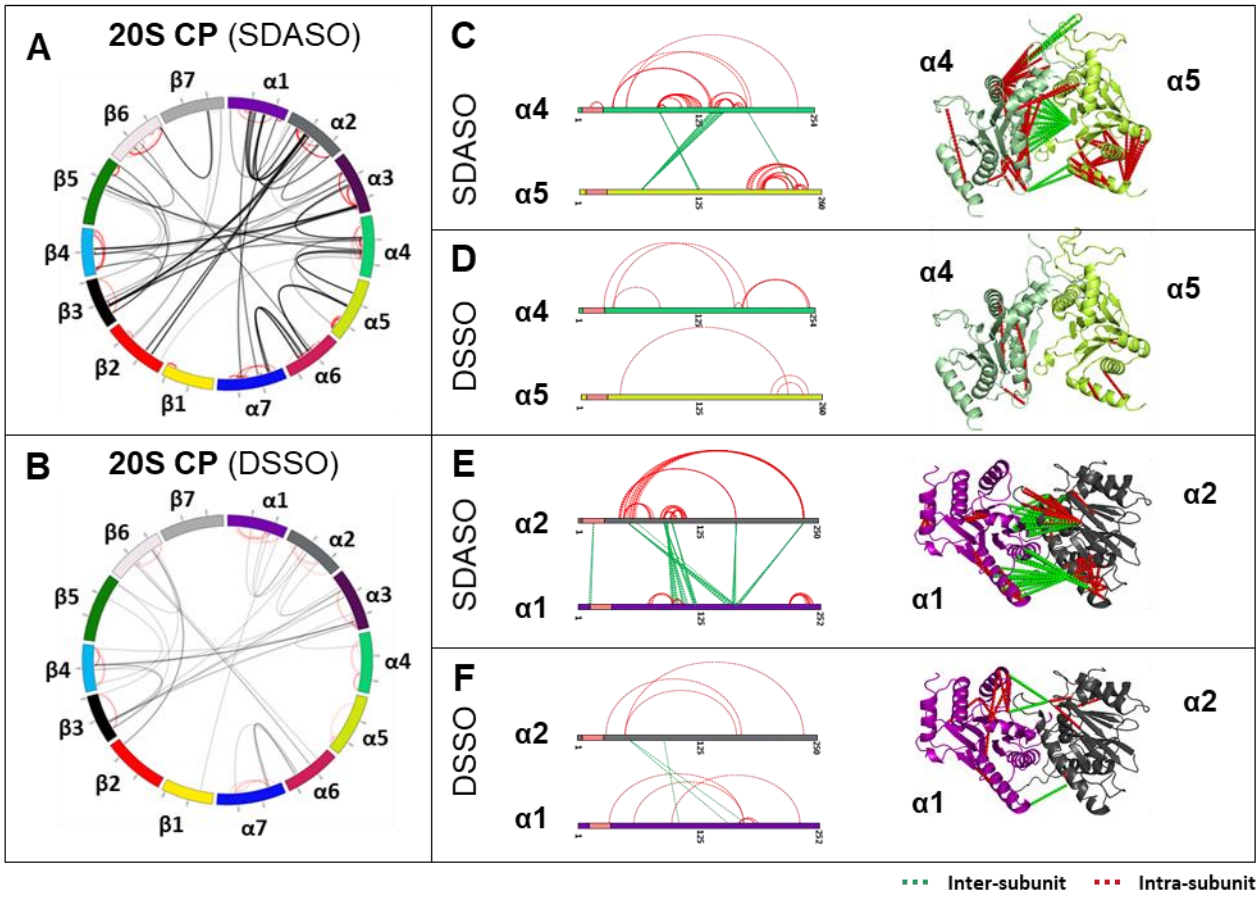

Figure 3.23. SDASO and DSSO XL-maps of the 20S CP. (A) Circular 2-D SDASO XL-map of the 20S CP. (B) Circular 2-D DSSO XL-map of the 20S CP. 2-D and 3-D XL-maps of α4-α5 interaction based on (C) SDASO and (D) DSSO cross-links. 2-D and 3-D XL-maps of α1-α2 interaction based on (E) SDASO and (F) DSSO cross-links. Note: high-resolution structure of the yeast 26S proteasome (PDB: 4CR2 (s1)) was used here. For 2-D XL-maps, inter-subunit linkages are colored black and intra-subunit linkages are colored red. For 3-D XL-maps, inter-subunit linkages are shown in green, while intra-subunit linkages are red.

FIGURE 3.24

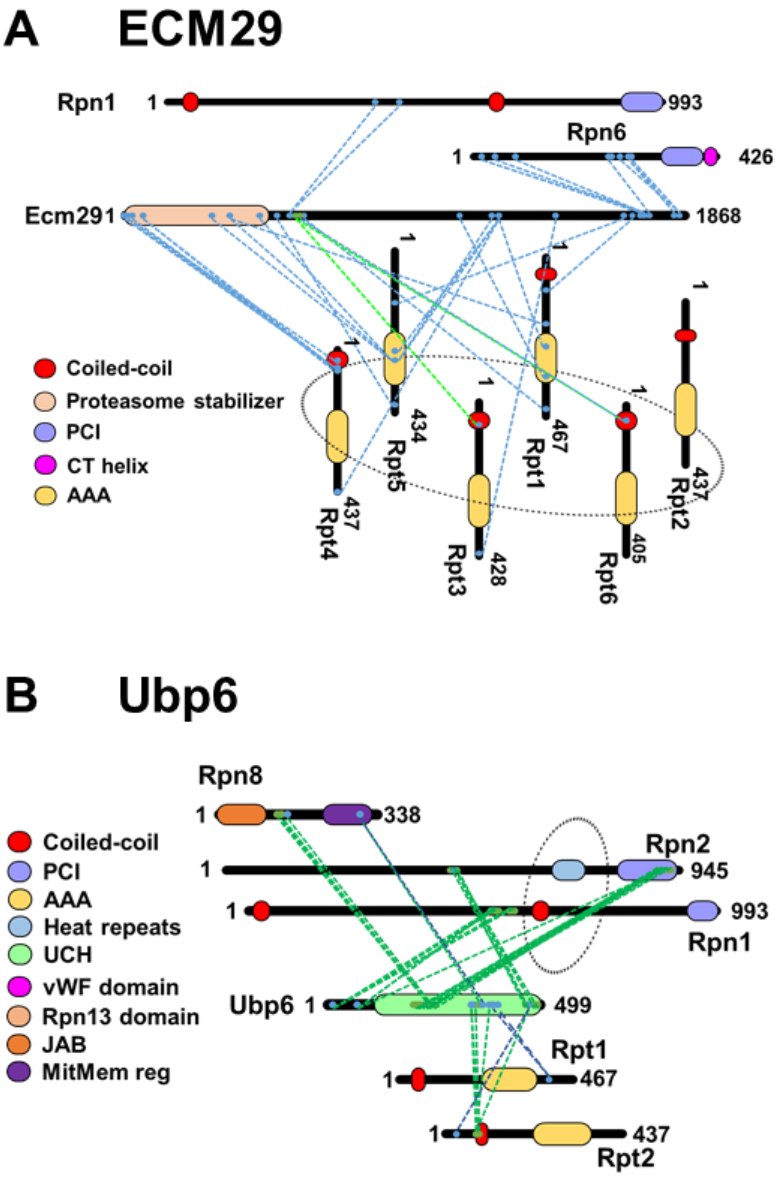

Figure 3.24. XL-maps of the Ecm29- and Ubp6-26S interactions. (A) **2**-D XL-map describing the interactions of Ecm29 with Rpn1, Rpn6, Rpt1, and Rpt3-6. (B) **2**-D XL-map illustrating the interactions of Ubp6 with Rpn1, Rpn2, Rpn8, Rpt1 and Rpt2. (Note: DSSO inter-subunit linkages are shown in blu and SDASO inter-subunit linkages are shown in green.)

FIGURE 3.25

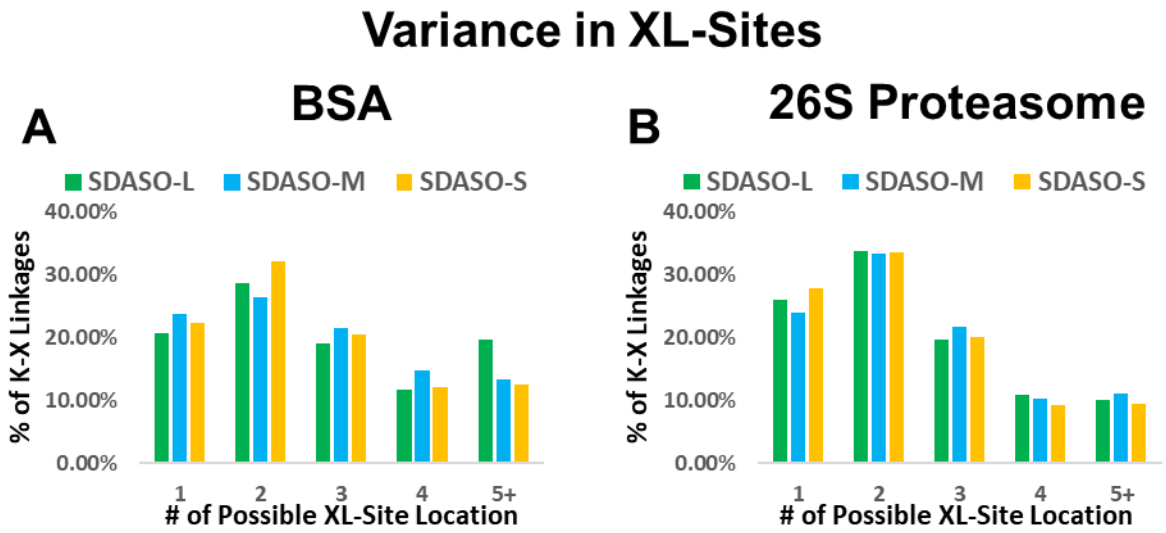

Figure 3.25. Variance in SDASO cross-linked sites. Distributions of the number of possible cross-linked site locations in the identified SDASO cross-links due to ambiguity in site identification for (A) BSA and (B) 26S Proteasome.

FIGURE 3.26

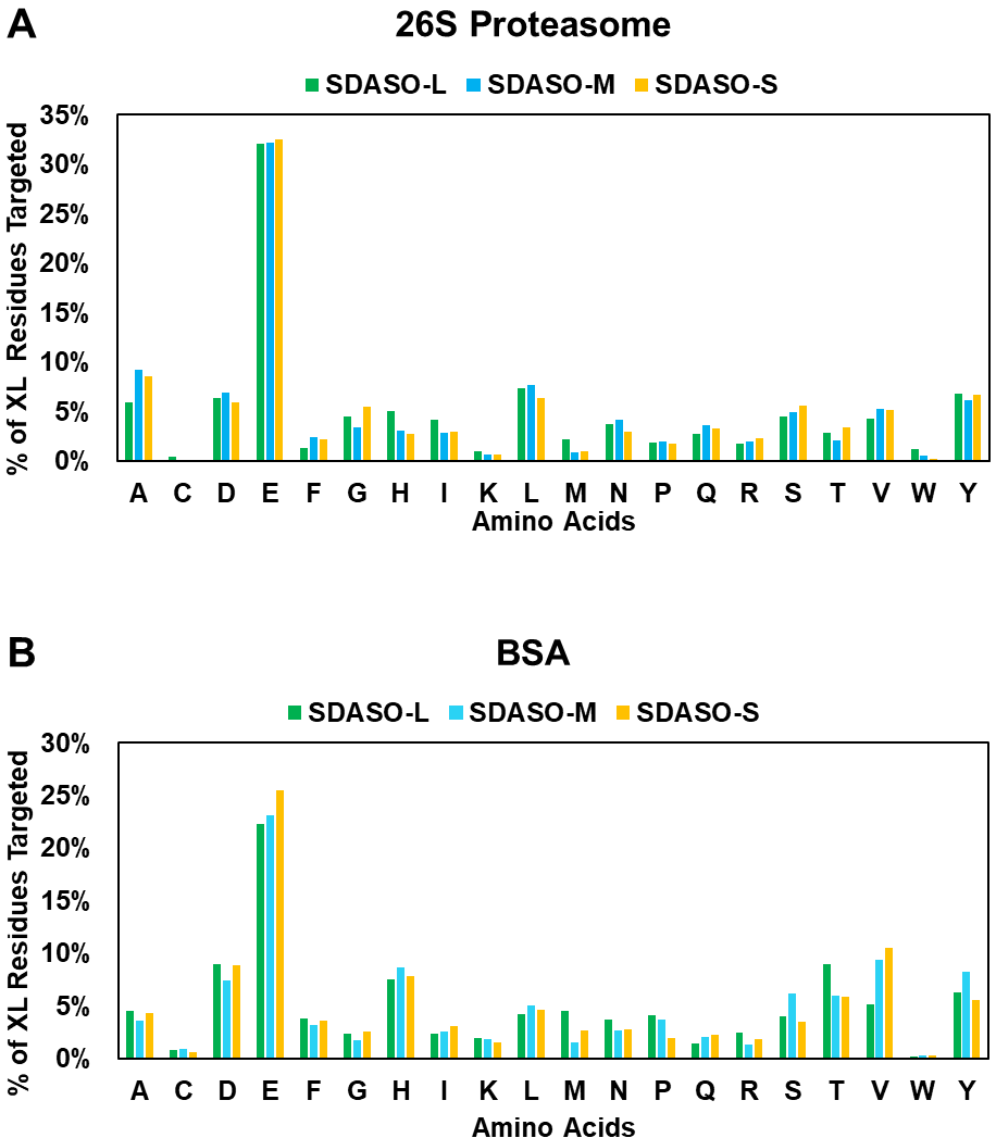

Figure 3.26. Respective distribution of amino acids targeted by diazirine based on the identified SDASO cross-links in (A) 26S proteasome and (B) BSA for SDASO -L, -M, and-S linkers.

FIGURE 4.1

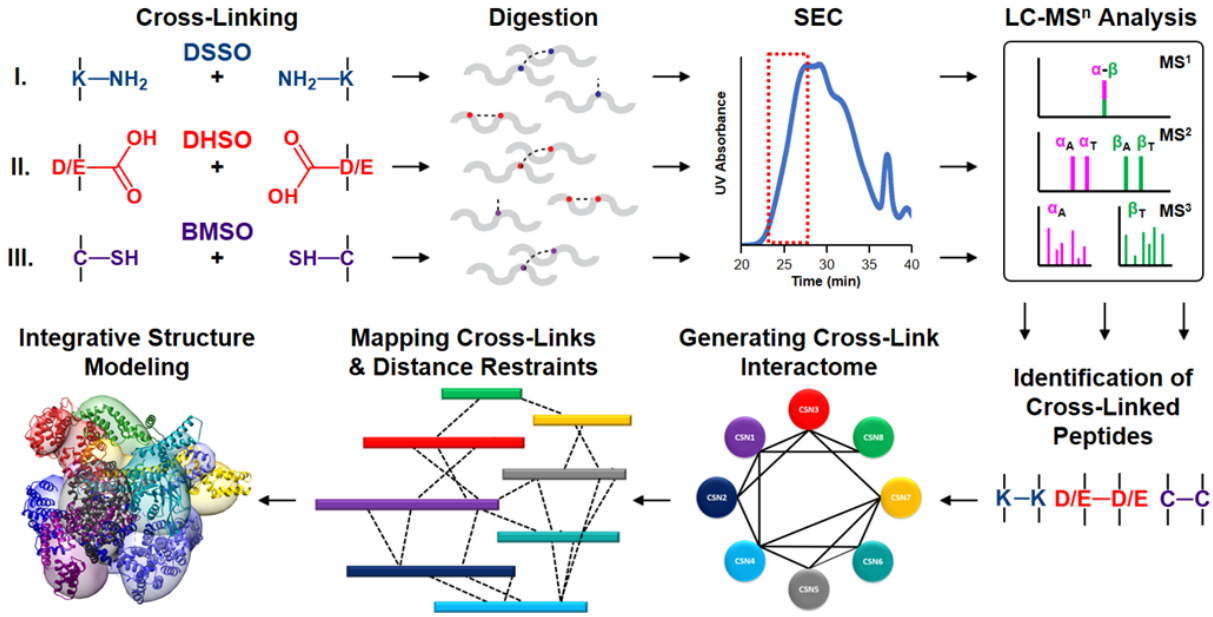

Figure 4.1. The multi-chemistry XL-MS workflow based on DSSO, DHSO and BMSO cross-linking. The two CSN complexes were independently cross-linked by each linker (path I (DSSO), II (DHSO), and III (BMSO)) and subsequently digested. Then cross-linked peptides were enriched from peptide mixtures by SEC, analyzed by LC-MS^n, and identified through database searching and result integration. The identified cross-linked peptides were then used for generating PPI networks, distance mapping and integrative structure modeling.

FIGURE 4.2

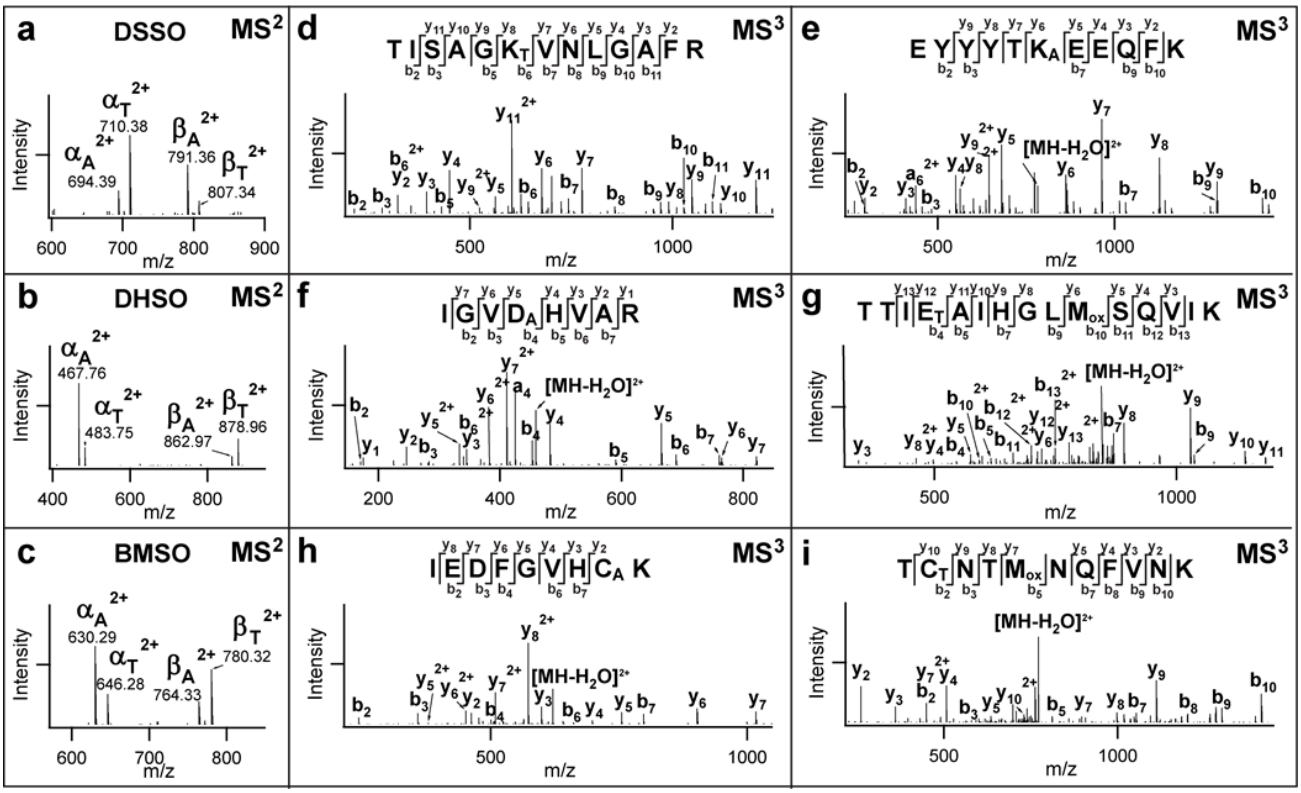

Figure 4.2. Representative MS^n analysis of a DSSO, DHSO and BMSO inter-linked peptide from CSN describing interactions between CSN5 and CSN6. MS^2 spectra of (a) DSSO (m/z 755.3713^{4+}), (b) DHSO (m/z 678.1147^{4+}), and (c) BMSO (m/z 710.0565^{4+}) cross-linked peptides (α–β), which fragmented similarly to yield two characteristic fragment ion pairs detected as α_A/β_T and α_T/β_A, respectively. This feature is unique to sulfoxide-containing MS-cleavable cross-linked peptides. MS^3 analyses of (c) α_T (m/z 710.38^{2+}) and (d) β_A (m/z 791.36^{2+}) fragment ions of the DSSO cross-link in (a) resulted in unambiguous identification of α_T as TISAGK$_T$VNLGAFR and of β_A as EYYYTK$_A$EEQFK, signifying a cross-link between CSN5:K180 and CSN6:K108. MS^3 analyses of (f) α_A (m/z 467.76^{2+}) and (g) β_T (m/z 878.96^{2+}) of the DHSO cross-link in (b) resulted in unambiguous identification of α_A as IGVD$_A$HVAR and of β_T as TTIE$_T$AIHGLM$_{ox}$SQVIK, signifying a cross-link between CSN5:E313 and CSN6:D201. MS^3 analyses of (h) α_A (m/z 630.29^{2+}) and (i) β_T (m/z 780.32^{2+}) of the BMSO cross-link in (c) resulted in unambiguous identification of α_A as IEDFGVHC$_A$K and of β_T as TC$_T$NTM$_{ox}$NQFVNK, signifying a cross-link between CSN5:C218 and CSN6:C299. K$_T$, E$_T$ and C$_T$ represent unsaturated thiol modified cross-link residue; whereas K$_A$, D$_A$ and C$_A$ represent alkene modified residues.

FIGURE 4.3

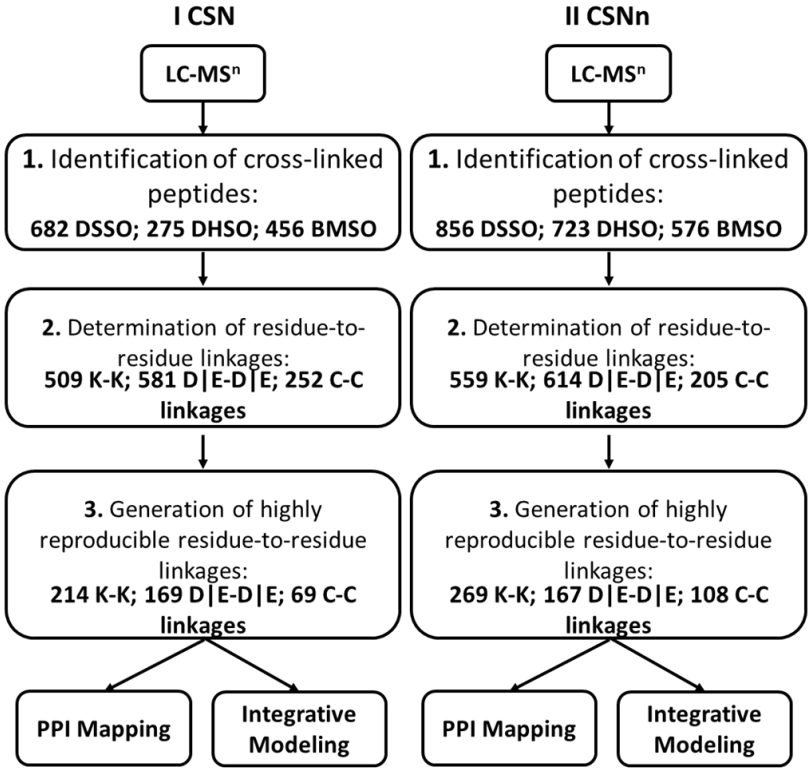

Figure 4.3. The highly reproducible residue-to-residue linkages were determined based on three steps: 1) identification of cross-linked peptides using LC MS^n (Table 4.3-4.5); 2) determination of residue-to-residue linkages (K-K for DSSO, D|E-D|E for DHSO, C-C for BMSO) based on the identified cross-linked peptides; 3) generation of highly reproducible residue-to-residue linkages that occur in ≥ 60% biological replicates (e.g. 3 out of 5 biological replicates) for each set of XL-MS experiments (Table 4.10-4.12), based on which PPI interaction topologies and integrative modeling were determined. XL-MS Data for CSN (path I) and CSNn (path II) were presented. It is noted that while the number of targetable residues in protein sequences attributes to the number of possible cross-links, the actual number of cross-linked peptides identified for each linker in XL-MS experiments mainly depends on: 1) the number of cross-linkable residues within the maximum distance range spanned by each linker; 2) the number of tryptic cross-linked peptides that can be detected and sequenced by mass spec analysis. Among the three selected linkers, BMSO has the longest spacer arm (24.2 Å) compared to DSSO (10.1 Å) and DHSO (12.4 Å). Thus, BMSO can bridge a significantly longer distance between targetable residues, which could contribute to increased number of cross-linkable residues in the structure and of cross-linked peptides identified.

FIGURE 4.4

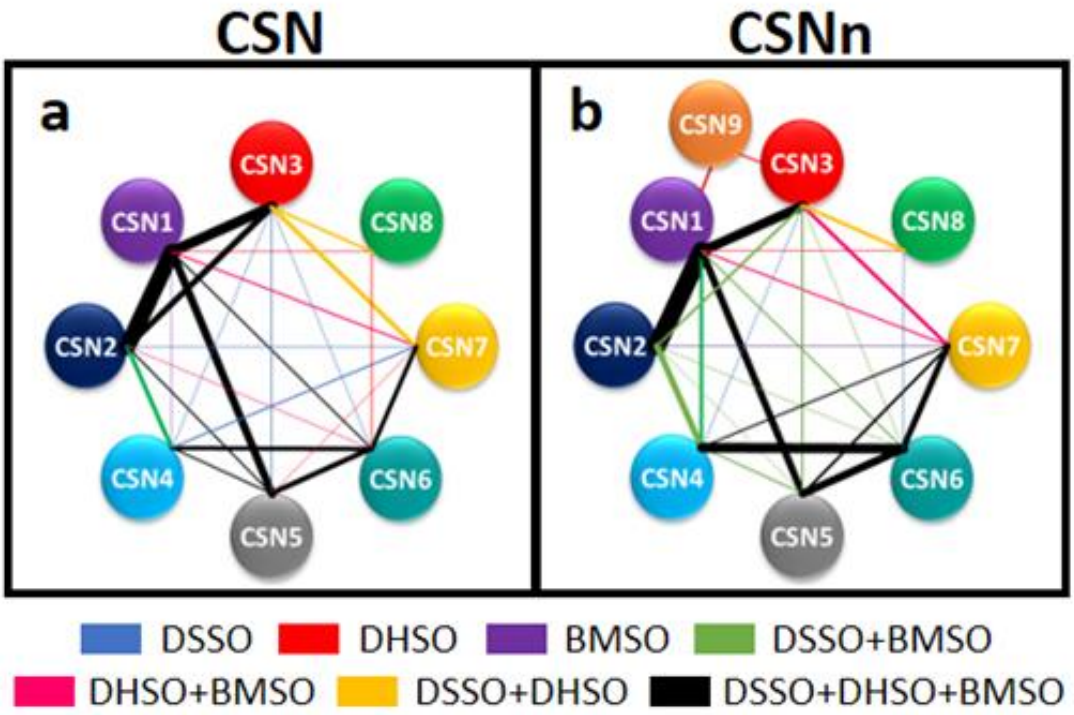

Figure 4.4. PPI maps of the CSN complexes based on cross-link data from all three linkers (DSSO, DHSO, BMSO). (a) CSN (CSN1-8). (b) CSNn (CSN1-9). Each CSN subunit was represented by colored nodes. The edges between two connected nodes were color-coded to describe PPIs resulted from individual or combinations of cross-linkers, i.e. blue-DSSO, red-DHSO, purple-BMSO, Lime-DSSO+BMSO, magenta-DHSO+BMSO, gold-DSSO+DHSO, black-DSSO+DHSO+BMSO. Edge thickness was determined by the total number of unique cross-links identified between the interactors.

FIGURE 4.5

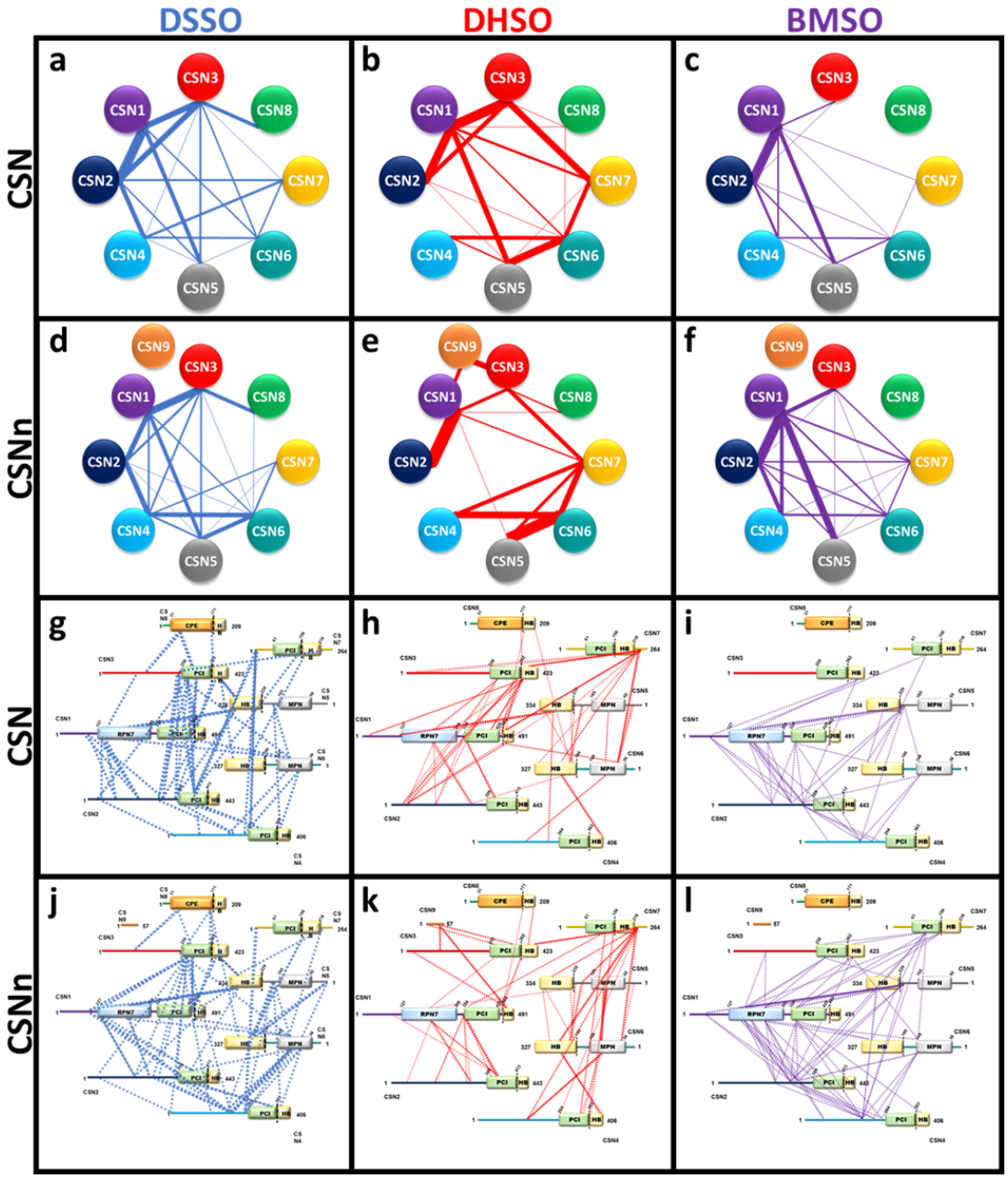

Figure 4.5. PPI maps of the CSN (a-c) and CSNn (d-f) complexes based on the selected cross-link data: (a, d) DSSO only; (b, e) DHSO only; (c,f) BMSO only. Each CSN subunit was represented by colored nodes as illustrated. Edges between two connected nodes were color-coded to describe PPIs determined by a specific cross-linker (blue-DSSO, red-DHSO, purple-BMSO). Edge thickness was determined based on the total number of linkages identified for each PPI. Respective 2-D XL-maps for CSN (g-i) and CSNn (j-l), illustrating residue-to-residue connectivity of inter-subunit interactions for the three linkers: DSSO (g,j), DHSO (h,k) and BMSO (i,l). Each subunit was displayed as a linear sequence containing their respective conserved domains.

FIGURE 4.6

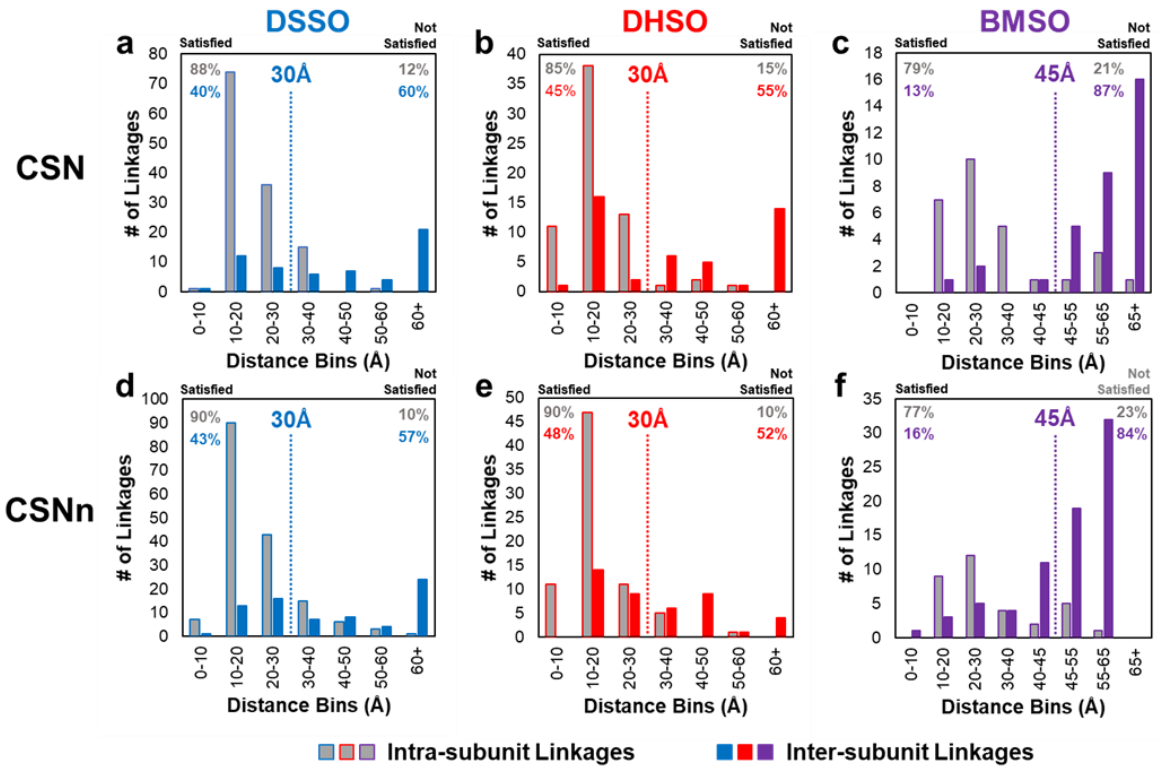

Figure 4.6. **Respective cross-link distance distribution plots for DSSO, DHSO or BMSO** cross-link data on the known CSN X-ray structure (PDB 4D10), displaying the number of linkages identified from both the CSN (a-c) and CSNn (d-f) complexes within and over the expected distance thresholds (30Å for DSSO and DHSO, and 45Å for BMSO). The greyed bars represent intra-subunit linkages, and the solid bars represent inter-subunit linkages. Cross-linkers were color coded: DSSO-blue, DHSO-red, and BMSO-purple. The % listed at the top left corners indicates % of cross-links (upper, intra-subunit; lower, inter-subunit) that were satisfied within the expected distances, whereas the % listed at the top right corners indicates % of cross-links (upper, intra-subunit; lower, inter-subunit) that were unsatisfied and outside the expected distances.

FIGURE 4.7

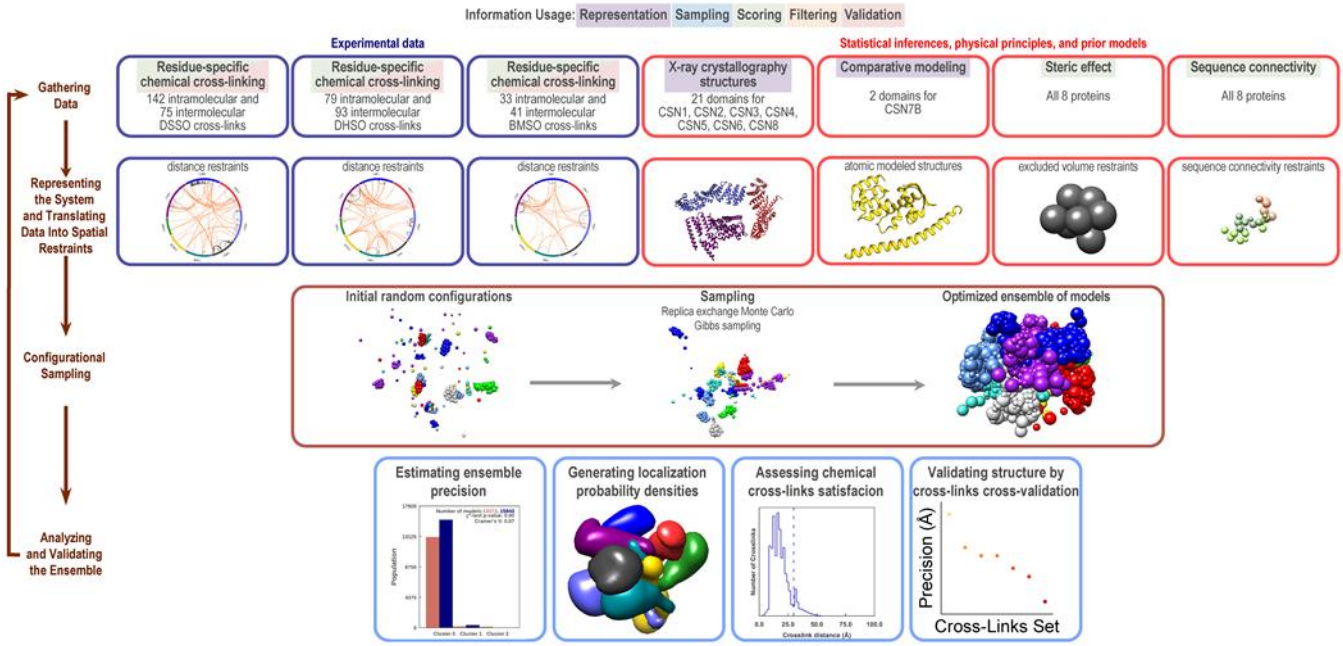

Figure 4.7. **The general scheme of integrative structure modeling.** Integrative structure determination of the human CSN complex, preceded through four stages: (1) gathering data, (2) representing subunits and translations data into spatial restraints, (3) configurational sampling to produce an ensemble of structures that satisfies the restraints, and (4) analyzing and validating the ensemble structures and data. The integrative structure modeling protocol (stages 2, 3, and 4) was scripted using the Python modeling interface package, which is a library for modeling macromolecular complexes based on our open-source *Integrative Modeling Platform* (IMP) package version 2.9 (https://integrativemodeling.org). Files containing the input data, scripts, and output results are available at https://github.com/salilab/CSN2019.

FIGURE 4.8

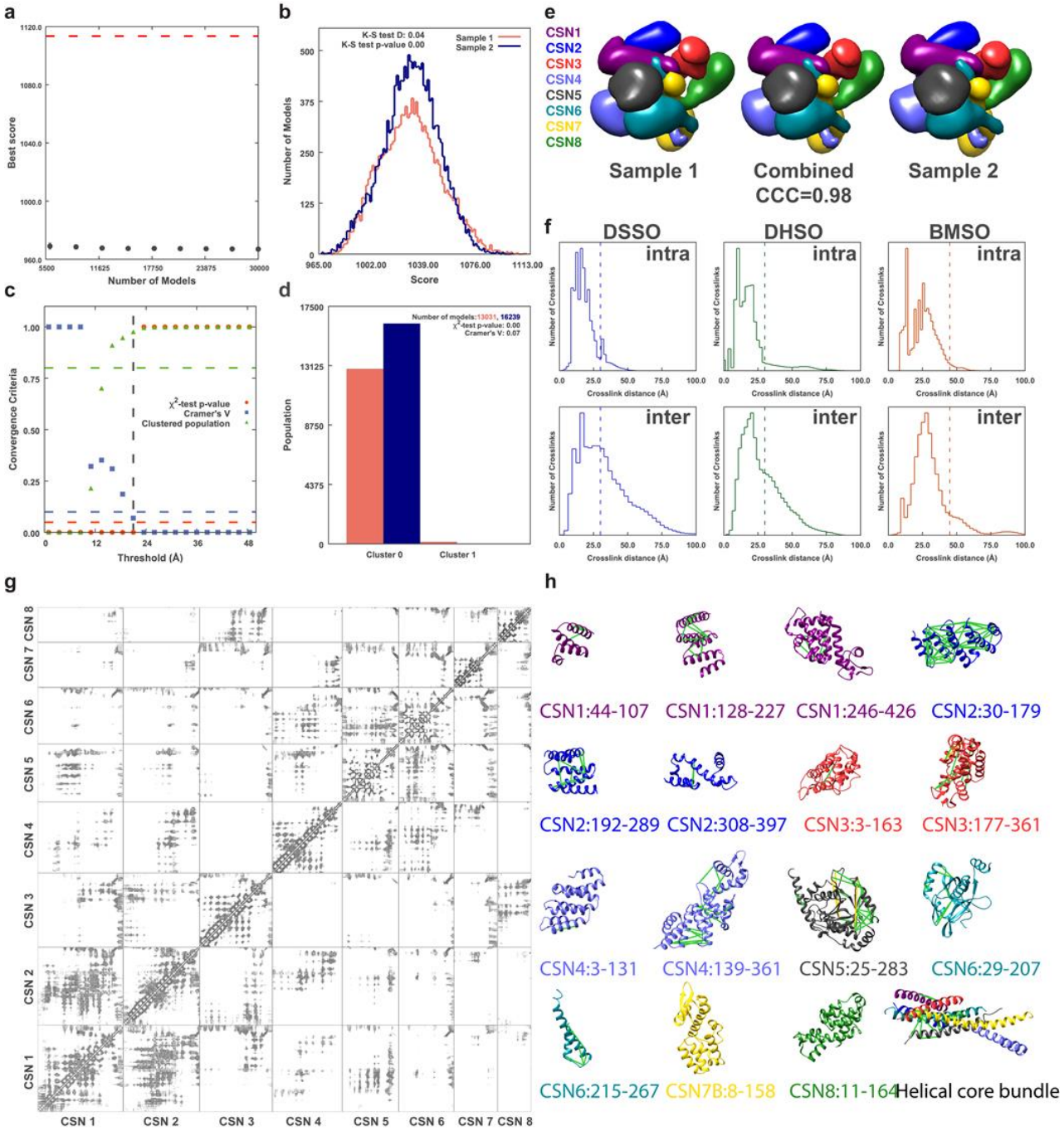

Figure 4.8. Validation of the canonical CSN structure. (a-e) Thoroughness of configurational sampling and model precision. (a) Convergence of the structure score for 30,000 randomly selected models of the canonical CSN complex out of the 54,702 clustered models; the scores do not continue to improve as more structures are computed, essentially independently of each other. (b) Distribution of scores for model samples 1 (red) and 2 (blue), comprising the 30,000 models (nsample1=13,123 and nsample2=16,876 structures). The non-parametric Kolmogorov–Smirnov two-sample test (two-sided) indicates that the difference between the two score distributions is insignificant, the magnitude of the difference is small, as demonstrated by the Kolmogorov–

Smirnov two-sample test statistic, D, of 0.04. Thus, the two score distributions are effectively equal. (c) Three criteria for determining the sampling precision (y axis), evaluated as a function of the RMSD. clustering threshold (x axis) (n=30,000 models). First, the P value is computed using the $\chi 2$ -test (one-sided) for homogeneity of proportions (red dots). Second, an effect size for the $\chi 2$ -test is quantified by the Cramer's V value (blue squares). Third, the population of structures in sufficiently large clusters (containing at least ten structures from each sample) is shown as green triangles. The vertical dotted grey line indicates the RMSD. clustering threshold at which two conditions are satisfied (Cramer's V (0.07)<0.10 (blue, horizontal dotted line) and the population of clustered structures (0.98)>0.80 (green, horizontal dotted line)), thus defining the sampling precision of 21Å. The three solid curves (in red, blue and green) were drawn through the points to help visualize the results. (d) Population of sample 1 and 2 structures in the two clusters obtained by threshold-based clustering using an RMSD. threshold of 21Å. The dominant cluster (cluster 1) contains 97% of the models. The precision of the dominant cluster defines the model precision, 16.4Å. (e) Comparison of localization probability densities of models from sample 1 (red) and sample 2 (blue) in the main cluster. The cross-correlation coefficient of 0.98 between the two samples localization probability densities suggests that the two samples are structurally identical. (f) Euclidean Cα–Cα distance distributions of all measured cross-links in the ensemble of solutions for each cluster. The y axis provides the normalized number of cross-links that were mapped onto the model. The dashed redline denotes the expected maximum reach of a cross-link. (g) Average contact maps calculated for the main cluster. Each square is a contact map calculated between a given pair of subunit of the CSN complex module proteins with border length proportional to the length of corresponding subunit sequences. The grey-shaded areas indicate observed interactions, with the grey-scale proportional to the fraction of models observing the corresponding interaction. (h) Illustration of the rigid bodies defined for modeling of the canonical CSN complex. The 16 rigid bodies used for modeling are shown in ribbon representation. The cross-links that map within each rigid body are visualized and are colored based on whether they were satisfied (green) or violated (gold) by our rigid body definition. 97% of these cross-links observed in the absence of CSN9 are satisfied by the rigid body definition.

FIGURE 4.9

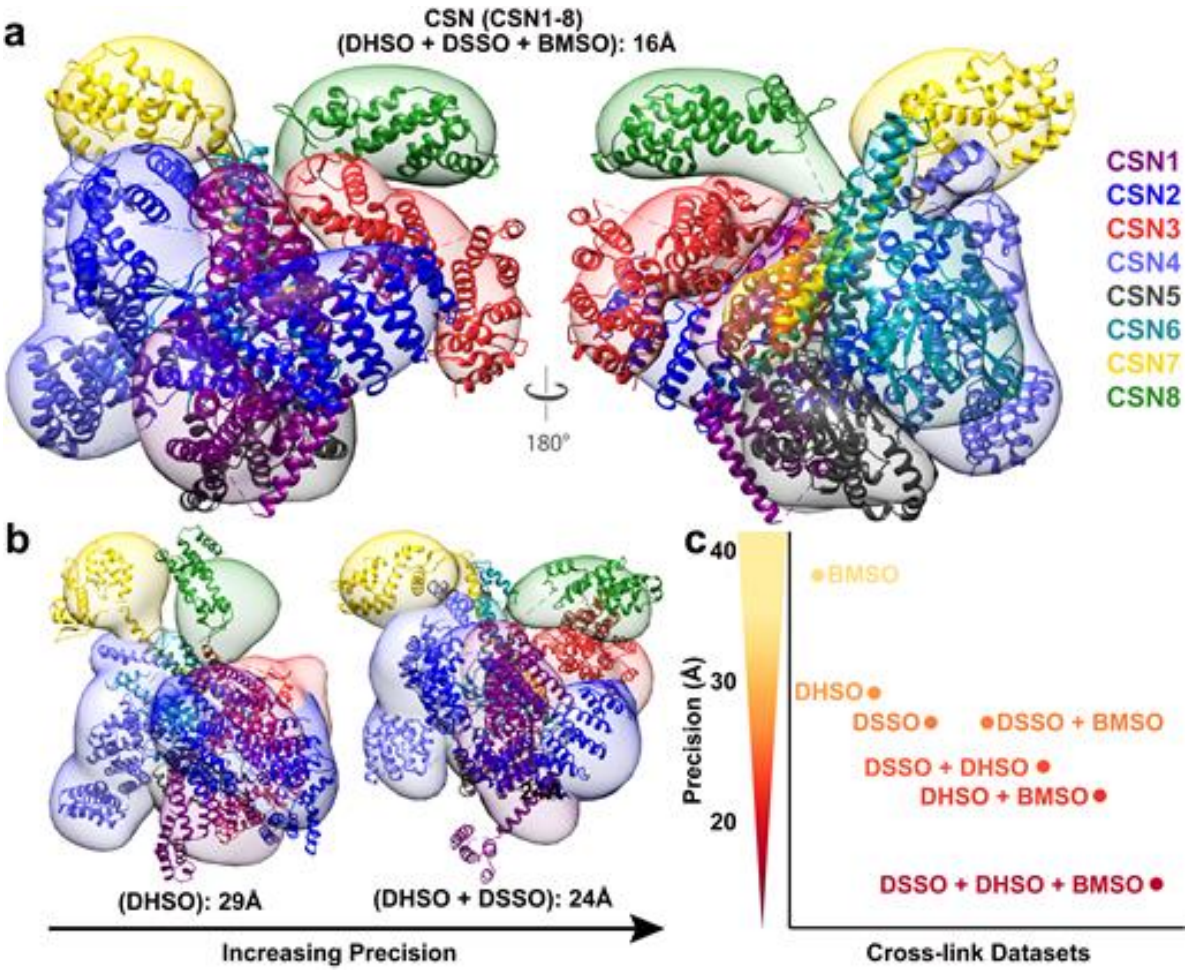

Figure 4.9. Integrative structures of CSN. (a) The integrative structure of CSN determined at 16Å precision when all three cross-link datasets (DSSO+DHSO+BMSO) were used for modeling. For each subunit, the localization probability density of the ensemble of models is shown with a representative structure (the centroid) from the ensemble embedded within it. (b) Integrative modeling of CSN determined using DHSO or DHSO+DSSO datasets yielded models determined at 29 Å and 24 Å precision respectively. (c) Graphical representation of determined model precisions with 7 combinations of our three cross-link datasets, illustrating that increasing the number of cross-linking chemistries (abscissa axis) for integrative structure modeling leads to increased model precision (ordinate axis). CSN subunit was color-coded as illustrated.

FIGURE 4.10

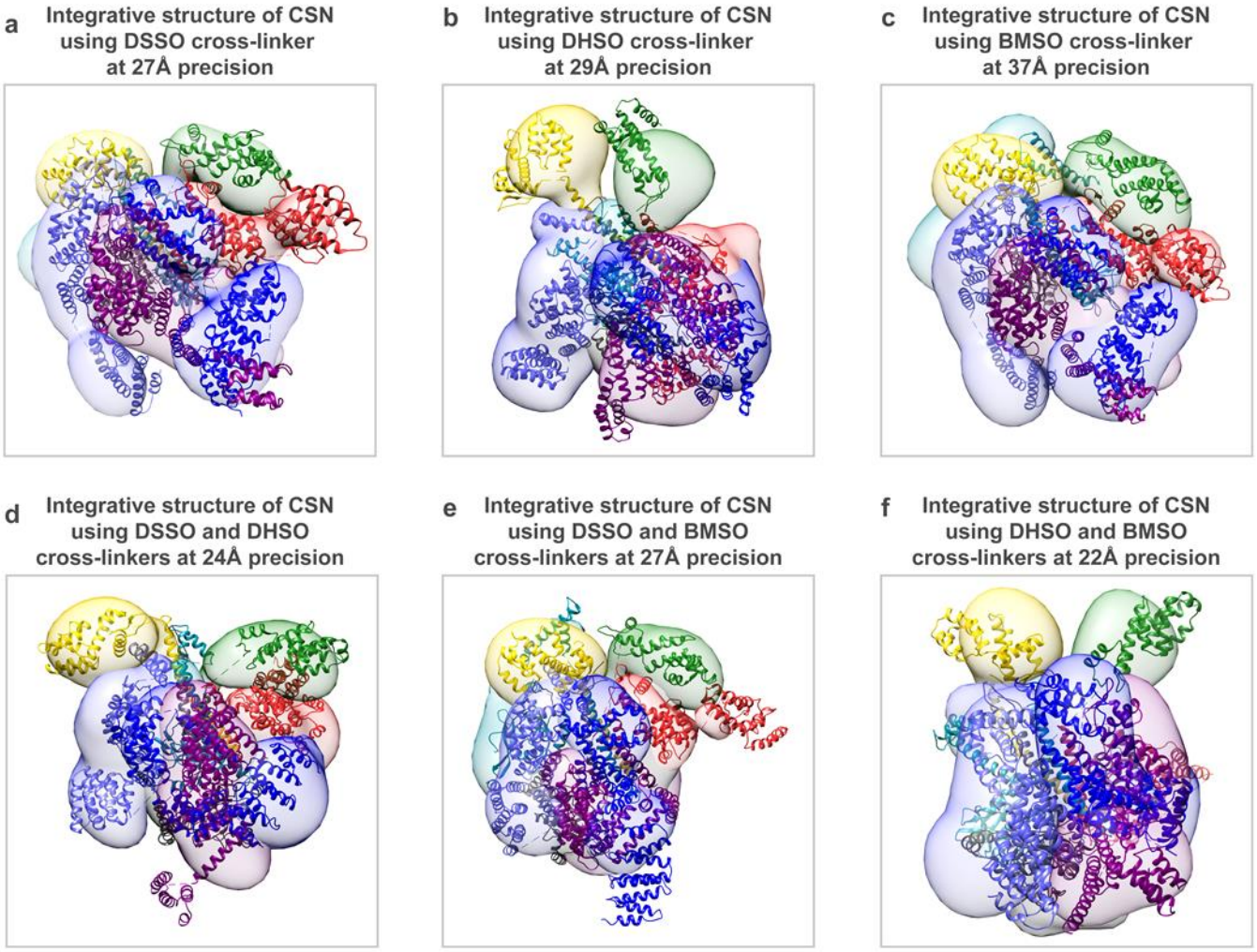

Figure 4.10. **Cross-validation of the canonical CSN complex integrative structure determination.** (a-f) Integrative structures of CSN obtained using different cross-links subset (i.e. (a) DSSO, (b) DHSO, (c) BMSO, (d) DSSO + DHSO, (e) DSSO + BMSO, (f) and DHSO + BMSO). The localization probability densities and the centroid of each ensemble of structures are shown, with the value of their respective precision.

FIGURE 4.11

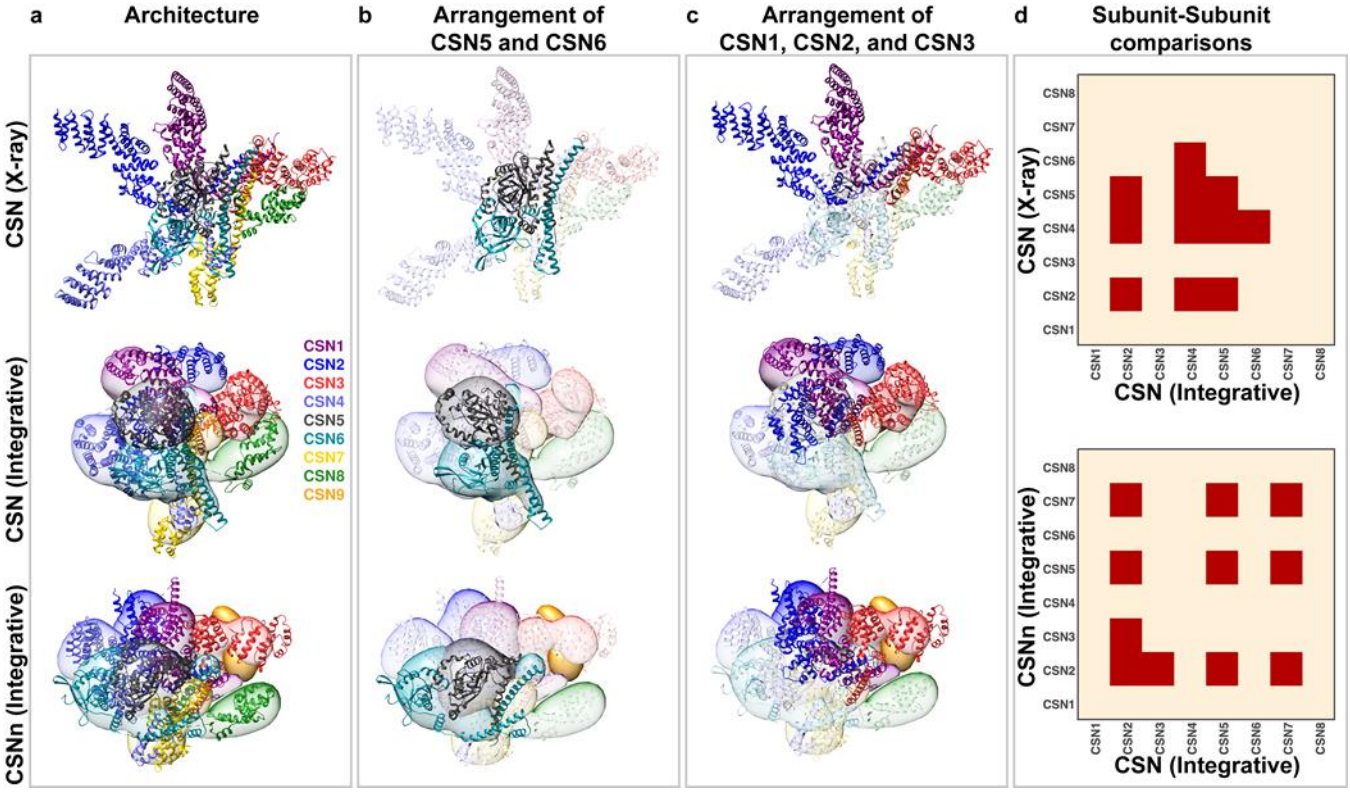

Figure 4.11. Comparison of integrative and X-ray structures of the CSN complexes. (a) Overall architectures of CSN: X-ray structure (PBD 4D10) (top row), CSN integrative structure (middle row) and CSNn integrative structure (bottom row). For each subunit in the integrative structures, the localization probability density of the ensemble of models is shown with a representative structure (the centroid) from the ensemble embedded within it. The CSN and CSNn structures show that the models adopt a more condensed state as compared to the X-ray structure, but they generally retain the overall architecture with only the helical bundle being constrained during modeling. (b) The arrangement of CSN5-CSN6 (MPN domain containing subunits) dimer was an emerging feature in integrative structures, however, a slight shift in the interface was observed in the CSNn model. (c) Models indicate that the arrangement of CSN1, CSN2 and CSN3 was altered in the presence of CSN9; CSN2 moved from a state interacting with CSN3 in CSN to an opened state in the CSNn model, resembling the overall architecture of the CSN X-ray structure. (d) Respective binary subunit-subunit comparison of the CSN integrative structure with the CSN X-ray structure (top) and the CSNn integrative structure (bottom row) respectively. The structures were compared by calculating their ensemble overlap; the overlap was quantified by the ratio of the distance between ensemble centroids to three times the sum of the ensemble precisions. Differences are shown in red. CSN subunit was color-coded as illustrated.

FIGURE 4.12

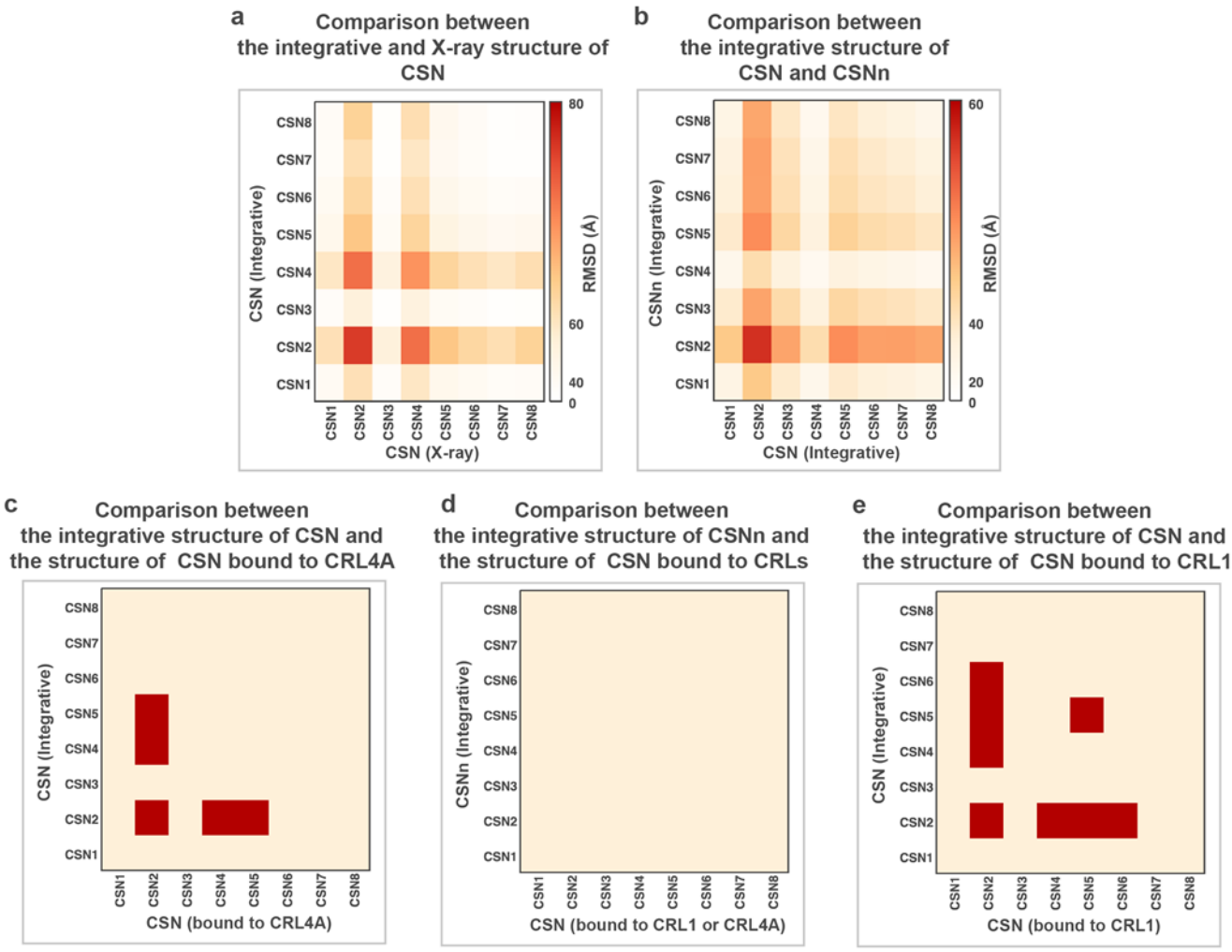

Figure 4.12. Subunit-subunit comparison among different CSN structures using RMSD between the centroid of each ensemble (a-b) and the overlap between structure ensembles (c-e). Subunit-subunit comparison between (a) the integrative and X-ray structures of CSN, (b) the integrative structures of CSN and CSNn, (c) the integrative structure of CSN and the cryo-EM Structure of CSN bound to neddylated CRL4A(27), (d) the integrative structure of CSNn and the cryo-EM structures of CSN bound to neddylated CRL4A or CRL1(27, 28), (e) the integrative structure of CSN and the cryo-EM structure of CSN bound to neddylated CRL1(28). When there was no overlap between the two structures, the differences were larger than three times precision of the respective structures as shown in red.

FIGURE 4.13

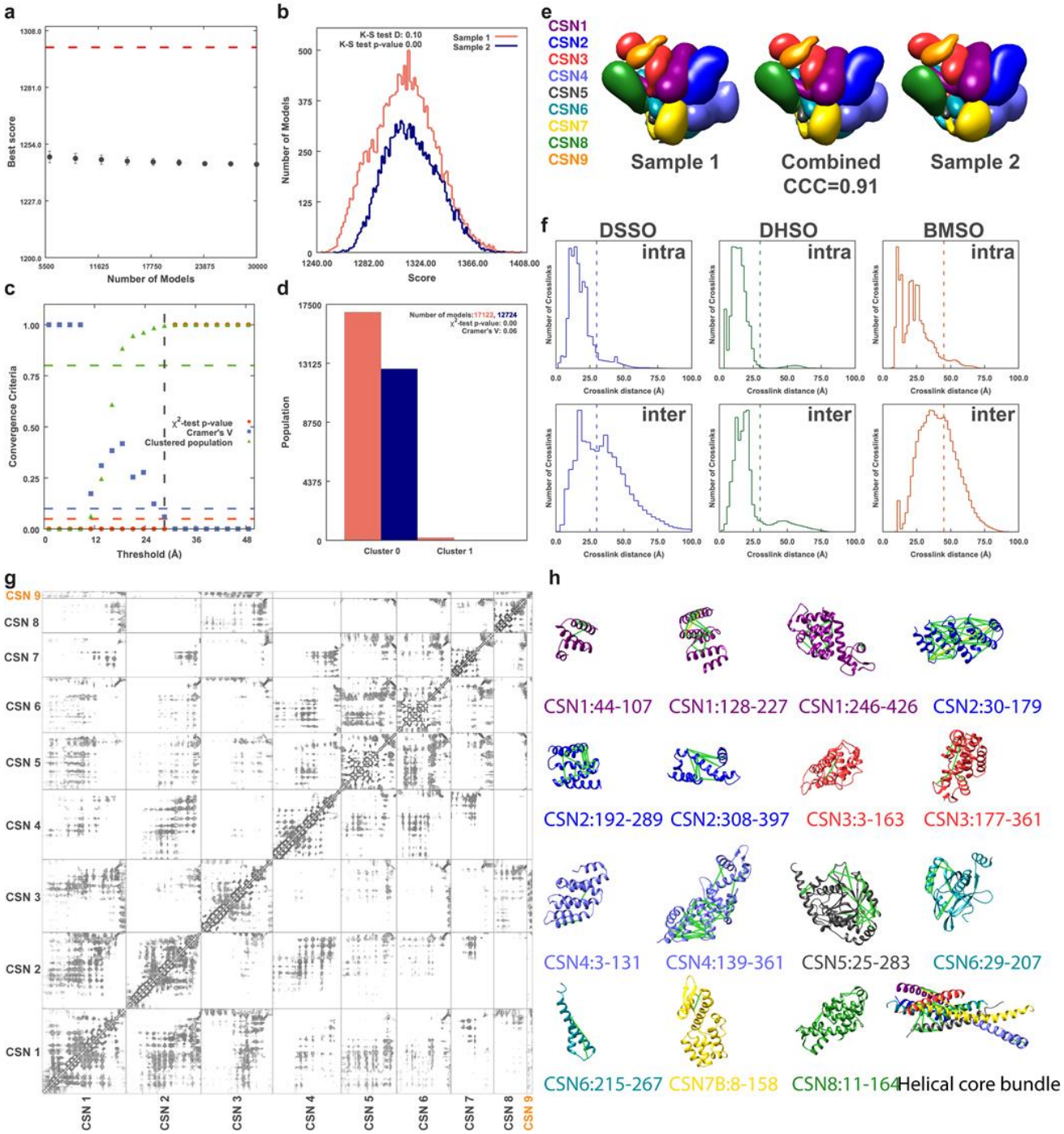

Figure 4.13. **Validation of the non-canonical CSN structure.** (a-e) Thoroughness of configurational sampling and model precision. (a) Convergence of the structure score for 30,000 randomly selected models of the non-canonical CSN complex out of the 99,902 clustered models; the scores do not continue to improve as more structures are computed, essentially independently of each other. (b) Distribution of scores for model samples 1 (red) and 2 (blue), comprising the 30,000 models (nsample1=17,252 and nsample2=12,747 structures). The non-parametric Kolmogorov–Smirnov two-sample test (two-sided) indicates that the difference between the two score distributions is insignificant, the magnitude of the difference is small, as demonstrated by

the Kolmogorov–Smirnov two-sample test statistic, D, of 0.10. Thus, the two score distributions are effectively equal. (c) Three criteria for determining the sampling precision (y axis), evaluated as a function of the RMSD. clustering threshold (x axis) (n=30,000 models). First, the P value is computed using the $\chi 2$ -test (one-sided) for homogeneity of proportions (red dots). Second, an effect size for the $\chi 2$ -test is quantified by the Cramer's V value (blue squares). Third, the population of structures in sufficiently large clusters (containing at least ten structures from each sample) is shown as green triangles. The vertical dotted grey line indicates the RMSD. clustering threshold at which two conditions are satisfied (Cramer's V (0.06)<0.10 (blue, horizontal dotted line) and the population of clustered structures (0.99)>0.80 (green, horizontal dotted line)), thus defining the sampling precision of 29Å. The three solid curves (in red, blue and green) were drawn through the points to help visualize the results. (d) Population of sample 1 and 2 structures in the two clusters obtained by threshold-based clustering using an RMSD. threshold of 29Å. The dominant cluster (cluster 1) contains 99% of the models. The precision of the dominant cluster defines the model precision, 21Å. (e) Comparison of localization probability densities of models from sample 1 (red) and sample 2 (blue) in the main cluster. The cross-correlation coefficient of 0.91 between the two samples localization probability densities suggests that the two samples are structurally identical. (f) Euclidean Cα–Cα distance distributions of all measured cross-links in the ensemble of solutions for each cluster. The *y* axis provides the normalized number of cross-links that were mapped onto the model. The *dashed red*line denotes the expected maximum reach of a cross-link. (g) Average contact maps calculated for the main cluster. Each square is a contact map calculated between a given pair of subunits of the CSNn complex module proteins with border length proportional to the length of corresponding subunit sequences. The grey-shaded areas indicate observed interactions, with the grey-scale proportional to the fraction of models observing the corresponding interaction. (h) Illustration of the rigid bodies defined for modeling of the non-canonical CSN complex. The 16 rigid bodies used for modeling are shown in ribbon representation. The cross-links that map within each rigid body are visualized and are colored based on whether they were satisfied (green) or violated (gold) by our rigid body definition. 99% of these cross-links observed in the presence of CSN9 are satisfied by the rigid body definition.

FIGURE 4.14

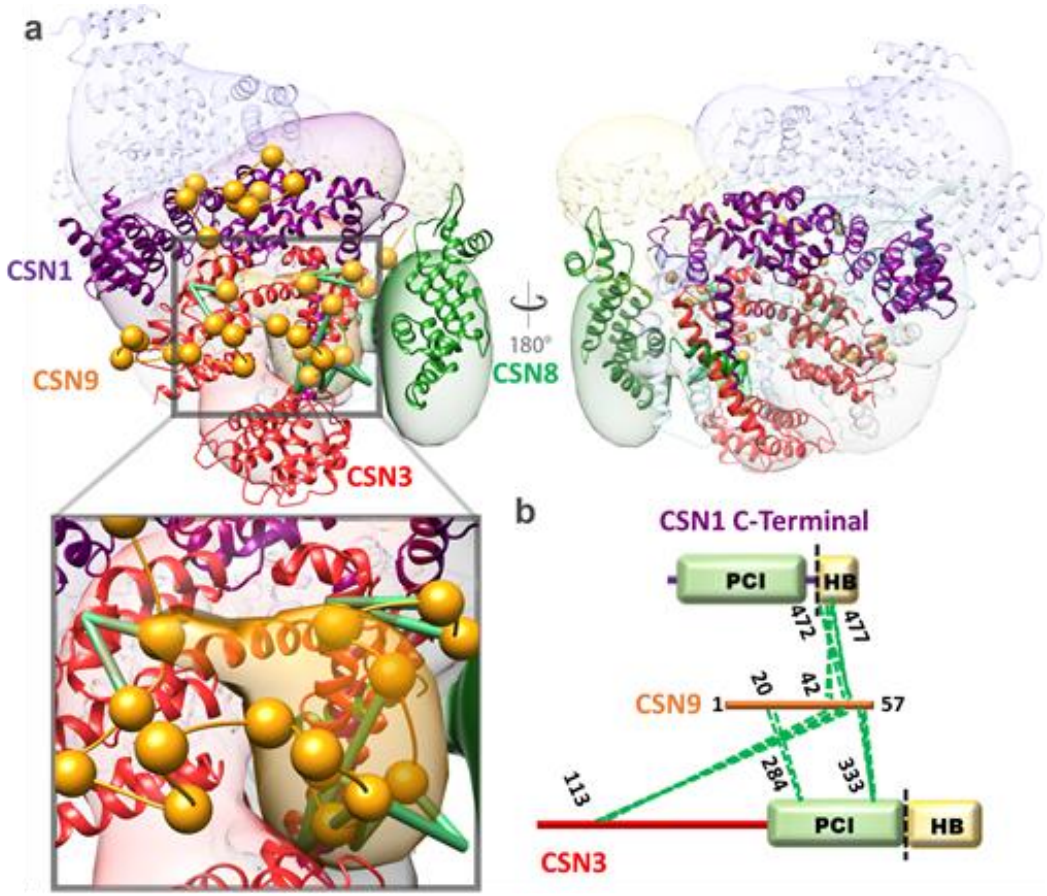

Figure 4.14. Binding of CSN9 in the CSNn integrative structure. (a) The integrative structure of CSNn determined at 22Å precision using all three cross-link datasets (DSSO+DHSO+BMSO). For each subunit, the localization probability density of the ensemble of models is shown with a representative structure (the centroid) from the ensemble embedded within it. The higher probable localization of CSN9, corresponding to its C-terminal, on the CSNn model is represented by the orange localization probability density, and a representative structure from the ensemble is show with spheres corresponding to 2-residue per beads connected by an extrapolated trace of the backbone. CSN9 primarily interacts with the main body of CSN3 (red) while its C-terminal tail also falls into the cavity between CSN1 (purple), CSN3 (red), and CSN8 (green). The inset displays a closer view of CSN9 interaction. Green lines represent CSN9-containing DHSO cross-links. (b) 2-D DHSO XL-map linking CSN9 to CSN1 and CSN3 at specific residues.

FIGURE 4.15

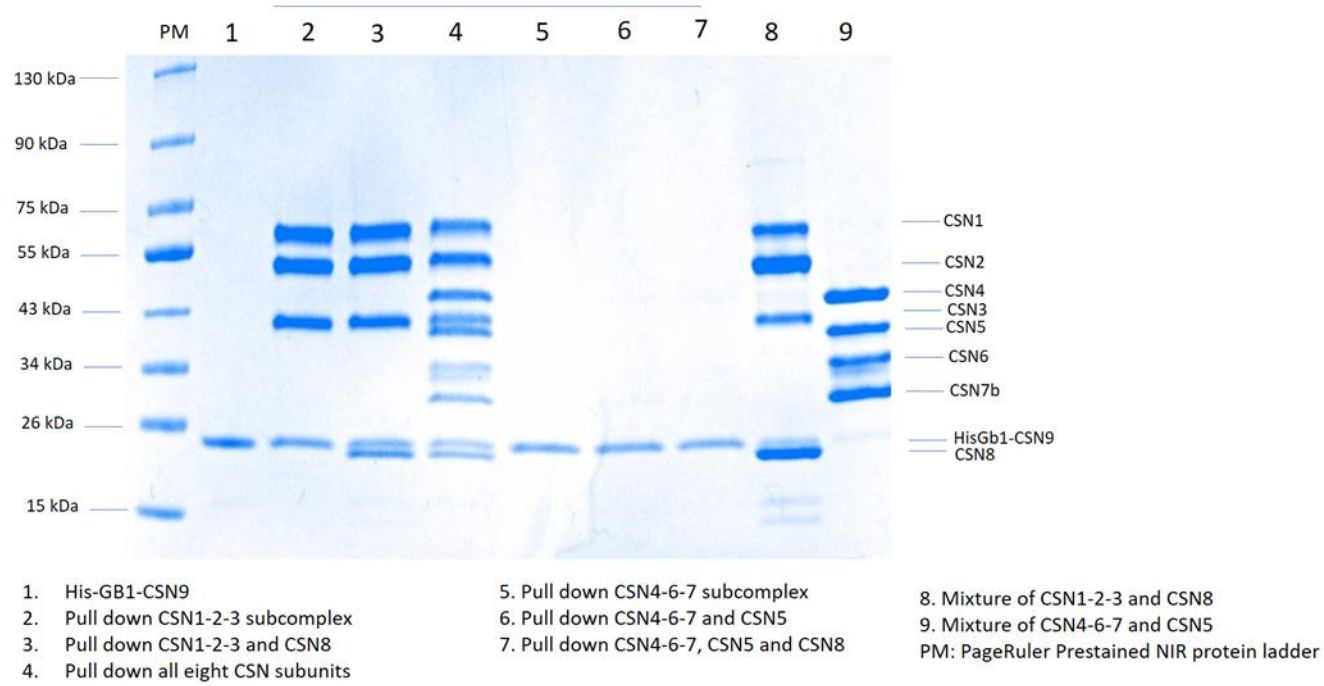

Figure 4.15. ***In vitro*** **pulldown assays to confirm the interactions of CSN9 with CSN1 and CSN3.** Six different CSN subcomplexes were used for CSN9 pulldown assay (Lane 2-7), in which His-GB1 fused CSN9 served as the bait protein to identify its binding partners. Lane 1, 8 and 9 display protein bands corresponding to CSN subunits (CSN1-9). The description of each lane is included.

FIGURE 4.16

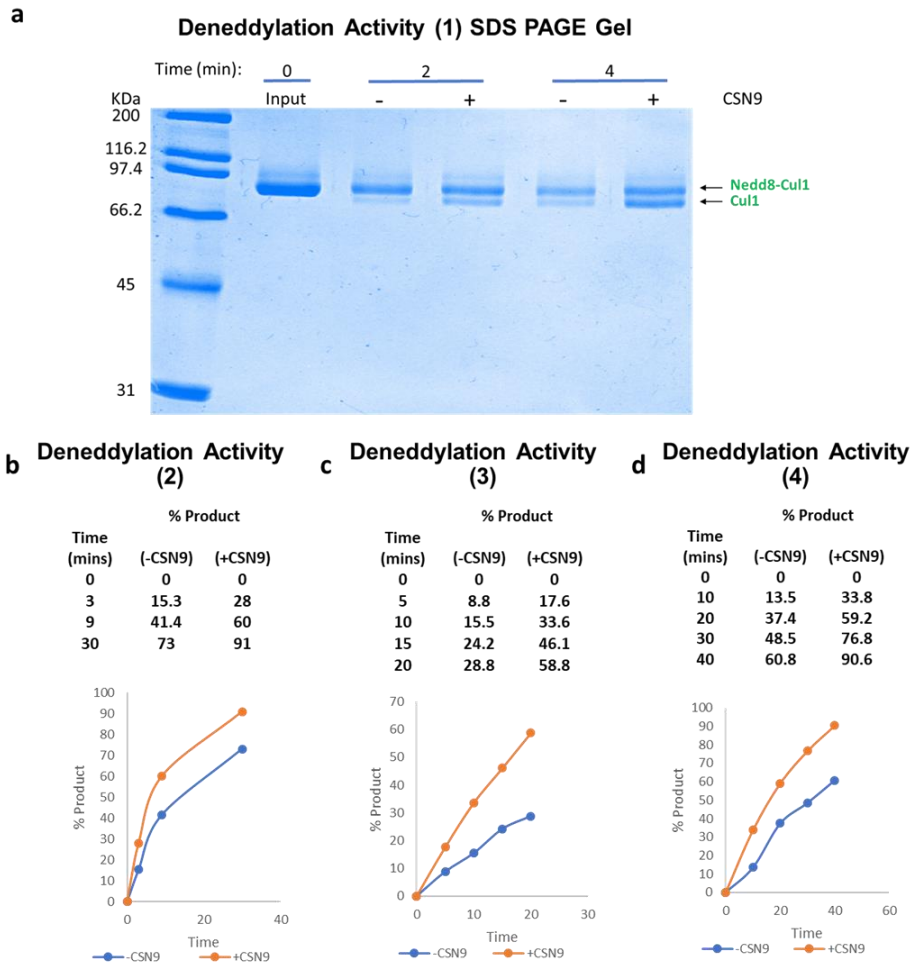

Figure 4.16. **CSN deneddylation activity in the presence or absence of CSN9.** (a) Purified CSN or CSNn was incubated with Nedd8-Cul1 over a 4 min period. Time points were collected at 2, and 4 minutes. SDS-PAGE analysis shows an increased rate of deneddylation for CSNn complex as compared to the CSN (CSN1-8) complex. (b-c) The same assay in (a) was performed three additional times with different time intervals to further demonstrate CSN9-induced activation of CSN deneddylase activity using neddylated Cul1 as the substrate. To illustrate the differences in the deneddylase activities of CSN in the absence and presence of CSN9, the percentage of un-neddylated Cul1 (the product) to total Cul1 at each time point was estimated based on protein band intensities on SDS-PAGE. Roughly, CSN9 can enhance CSN activity about 2-3 folds. Note: All assays were performed at room temperature with 5 μM Nedd8-Cul1 and 10 nM CSN in Buffer containing 20 mM HEPES (pH 7.5), 150 mM NaCl, and 1 mM TCEP.

FIGURE 4.17

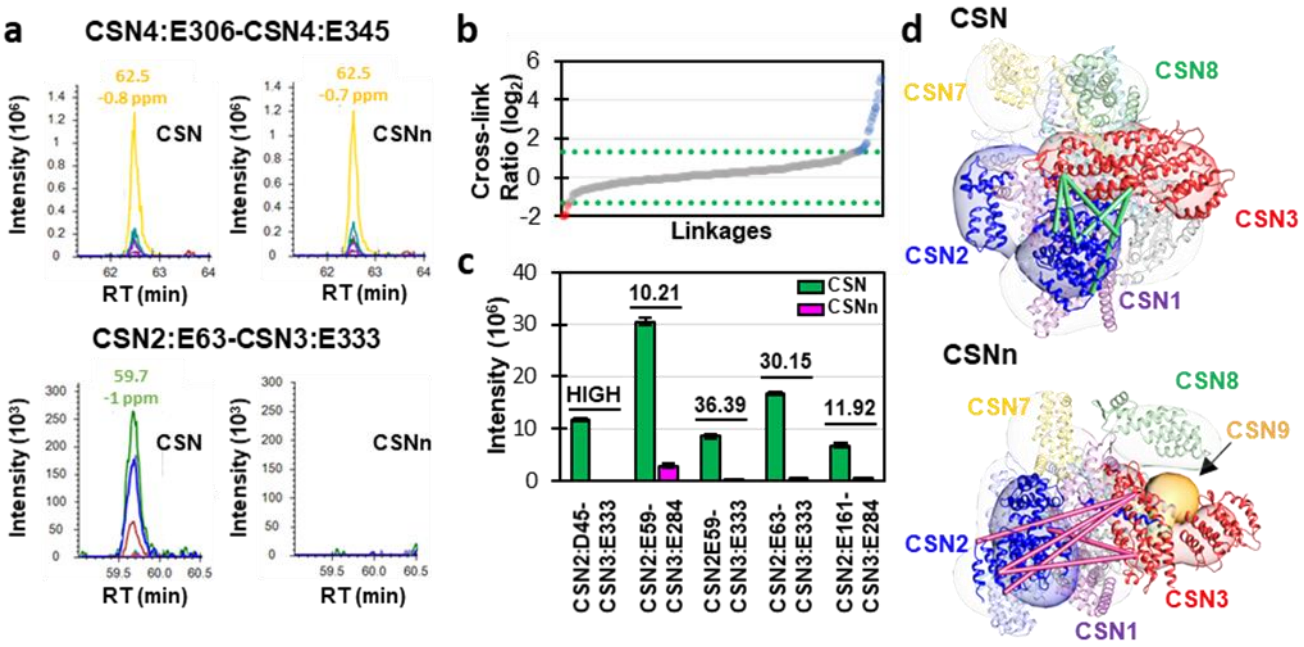

Figure 4.17. PRM-based targeted quantitation of DHSO cross-linked peptides to validate CSN9-induced structural changes in CSN. (a) Skyline outputs for PRM quantitation of a representative DHSO intra-subunit (CSN4:E306-CSN4:E345) (top) and an inter-subunit (CSN2:E63-CSN3:E333) (bottom) cross-linked peptides to compare their relative abundance in the CSN and CSNn complexes. Based on peak areas, the relative abundance ratio (CSN/CSNn) of the intra-subunit cross-link was determined as 1.11 (top), indicating no significant change. In contrast, the relative abundance of the inter-subunit cross-link (CSN/CSNn) was determined as 30.15 (bottom), suggesting a significant change. (b) The distribution of cross-link ratios (CSN/CSNn) of 229 DHSO cross-linked peptides (represented as log2 values) determined by PRM quantitation, in which only 22 cross-linked peptides displayed significant changes (>2.5-fold, greater than 3σ), including 4 with decreased ratios (red dots) and 18 with increased ratios (blue dots). The cross-link ratios (CSN/CSNn) describe the relative abundance of cross-linked peptides in the two compared complexes. (c) Abundance of five quantifiable CSN2-CSN3 cross-links (CSN2:D45-CSN3:E333, CSN2:E59-CSN3:E284, CSN2E59-CSN3:E333, CSN2:E63-CSN3:E333, and CSN2:E161-CSN3:E284) detected in the CSN and CSNn complexes. The underlined numbers shown represent relative abundance ratios (CSN/CSNn) of the selected cross-linked peptides between the two complexes, indicating that these interactions are favored in CSN. (d) The five cross-links shown in (c) were mapped on CSN and CSNn integrative structures. The linkages in CSN model (green) are satisfied within the expected distance (< 30Å), which are not satisfied in the CSNn model (magenta). Details on PRM quantitation of the cross-linked peptides are listed in Table 4.24.

FIGURE 4.18

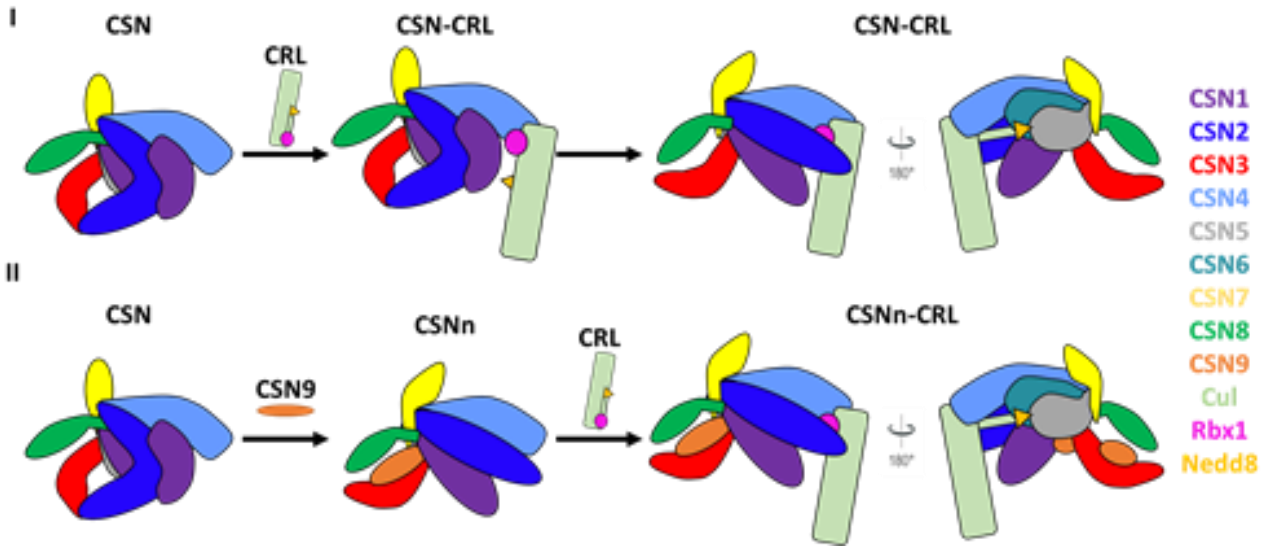

Figure 4.18. The proposed structural model of CSN9 binding to facilitate CSN interaction with neddylated CRLs. CSN and neddylated CRL subunits were color-coded as illustrated(I) CSN9-free CSN needs to undergo substantial conformational changes upon binding to a neddylated CRL. In comparison, (II) CSN9-bound CSN adopts a configuration better suited for CRL binding.