Published in final edited form as:

Anal Chem. 2019 January 02; 91(1): 142–155. doi:10.1021/acs.analchem.8b05014.

Evolution of Structural Biology through the Lens of Mass Spectrometry

Upneet Kaur, Danté T. Johnson, Emily E. Chea, Daniel J. Deredge, Jessica A. Espino, and Lisa M. Jones*

Department of Pharmaceutical Sciences, University of Maryland, Baltimore, Maryland 21201, United States

Since its inception in the early 20th century, mass spectrometry (MS) has become a significant method to analyze molecules. In the century since JJ Thomson's first use of MS in 1913, the application of this analytical technique has expanded to a wide variety of industries including pharmaceuticals, biotechnology, forensics, and environmental. Advancements in instrumentation have allowed for diversity in the type of molecules that can be analyzed by MS. The development of electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) in the 1980s has increased the size of molecules that can be studied, allowing for the analysis of proteins to be well in the scope of the technique and has increased the range of MS to biological applications. Although mass spectrometry has been used for many years to analyze proteins, mainly for protein sequencing, it is only through developments within roughly the last two decades that have enabled it to be used as a technique for studying protein structure.

METHODS FOR STRUCTURAL BIOLOGY

The analysis of protein structure is essential for understanding protein function and dysfunction. The field of structural biology has long been dominated by the high-resolution techniques X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR), which provide an atomic-level view of proteins. In recent years, cryo-EM has emerged as another powerful tool for structural biology. This technique enables the structural analysis of very large molecules (MDa range), and advancements in technology have allowed for higher resolution structures to be determined. The advantage of these three methods is their ability to provide high-resolution structural information on proteins, but they are also limited in their use. Many structural methods are limited in the size of molecules they can study, whereas mass spectrometry can study a wider range of molecular sizes (Figure 1). Multiple approaches including the use of protein digestion coupled to liquid chromatography (bottom-up proteomics) as well as the development of instrumentation with wider m/z ranges have enabled studies of larger biomolecular complexes. This gives MS a higher flexibility in providing structural information on isolated protein complexes as well as proteins in cells, tissues, and even organisms. MS-based methods also have the advantage

*Corresponding Author: ljones@rx.umaryland.edu; Tel: 410-706-3380; Fax: 410-706-7670.

The authors declare no competing financial interest.

that they can analyze heterogeneous proteins (post-translational modifications and varying conformers) that are difficult to study by other methods. Further, MS-based methods require substantially less protein (μg quantities) than other structural techniques. It has been established in recent years that the use of multiple structural techniques aids in fully characterizing proteins. This has led to the rise in the use of a combination of structural biology methods, including MS-based methods, to fully characterize proteins.

This review focuses on the MS-based structural proteomics methods native MS, ion mobility spectrometry (IMS), top-down proteomics, chemical cross-linking (XL-MS), and covalent labeling. This suite of methods has been used to study noncovalent complexes, to identify protein interaction sites, and for de novo structural modeling. Each of these methods have been extensively reviewed elsewhere with detailed descriptions of the history, advantages, and applications of each technique.^{1–8} Here, we focus on the most recent, within the last two years, advancements in each field. These advancements include developments in instrumentation, software, and methodology.

NATIVE MASS SPECTROMETRY

Biological pathways are coordinated by different macro-molecular interactions, which can be between different proteins, lipids, and nucleic acids. The characterization of these interactions is essential due to their potential as drug targets. Native MS is a well-established tool to characterize biomolecular structures and interactions under physiologically relevant conditions. In contrast to denaturing MS techniques where an organic solvent, often at an acidic pH, is used, native MS employs near native conditions to maintain the integrity of noncovalent interactions. In native MS, the intact macromolecules are directly ionized from a non-denaturing solvent, preserving noncovalent interactions in the gas phase and providing structural information including complex stoichiometry, assembly, and topology.¹ The ability to study large macromolecular complexes by native MS has been possible due to advancements in technology. The development of nanoelectrospray ionization (nESI), advancements in MS instrumentation including increased m/z ranges and lower pressures,^{9–11} and the improvement in sample preparation have increased the size of the analyte that is able to be structurally characterized by native MS. With these advancements, this method can potentially be implemented as a quality control step before analysis by EM or crystallography to determine the native state structure of the protein.

Using Native MS To Study Membrane Proteins.

Understanding membrane protein structures is crucial, because they play essential physiological roles and make up a majority of therapeutic targets. Membrane proteins have been challenging for biophysical studies because of low physiological expression levels, the insoluble nature of biological membranes, and their heterogeneity. Detergents have been used to study membrane protein structure for native MS but may give rise to the destabilization of protein structure, protein–protein interactions, and protein–ligand interactions. Recently, alternative membrane mimetics such as amphipols, lipid nanostructures, liposomes, and intact nanodiscs have been used to create biologically relevant approaches for native MS of membrane proteins.¹² Nanodiscs are nanoscale lipoprotein

particles consisting of a lipid bilayer surrounded by two membrane scaffold protein (MSP) belts. Nanodiscs have been shown to have extraordinary gas-phase stability when they are ionized by native ESI.^{13,14} To investigate the disassociation of nanodiscs in the gas phase, the collisional-induced dissociation (CID) energy or the multiphoton dissociation energy was increased.¹⁵ A shift of the nanodisc ions to lower m/z values showed that nanodisc complexes lost both mass and charge as they are activated. Lipid composition of heterogeneous nanodiscs was determined by employing lipids of slightly different masses. Nanodiscs were prepared with palmitoyl-oleoyl-phosphatidylcholine (POPC), palmitoyl-oleoyl-phosphatidylglycerol (POPG), and palmitoyl-oleoyl-phosphatidyl-serine (POPS) in different ratios. The nanodiscs displayed similar composition at low collisional energy, but at higher collisional energies, they displayed a polarity dependent depletion of certain lipids, suggesting that the chemistry of the lipid molecules played a crucial role in dissociation mechanisms.¹⁵

The integrity of intact membrane protein nanodiscs was assessed by using two membrane protein oligomers, trimeric AmtB and tetrameric AqpZ, in nanodiscs with different lipid compositions.^{15,16} Distinct features of the membrane protein nanodiscs showed variation as a function of collisional energy. At high collisional energy, the nanodisc complex disassociated into the lipids, membrane scaffold proteins (MSPs), and membrane protein monomers (Figure 2). At an intermediate collisional energy, the AmtB trimer was detected with nine lipids bound. At low collisional energy, the majority of the scaffold proteins and lipids were removed, leaving only the membrane protein oligomer and any lipids in contact with the protein surface. A challenge of this method is the m/z overlap between the MSP belts and the lipids. By designing multiple nanodiscs with different lipid compositions and/or MSP belts, their isobaric masses can be distinguished. One problem with this approach is the possible disruption of the protein complex and the time and cost of designing multiple nanodiscs for an experiment. Reid et al. performed a study with mutated MSP belts that resulted in subtle mass shifts to distinguish the MSP belts from the protein-bound lipids.¹⁶ These changes do not disrupt the interaction between protein-bound lipids and/or nanodisc assembly. By adding one to two amino acids to the N-terminus of the MSP belt, the nanodisc structure is preserved, and both the MSP belt and protein-bound lipids can be distinguished. This study demonstrates how native MS with membrane protein nanodiscs can be used to determine the stoichiometry of lipid-protein complexes.

Ultrahigh-Mass Range Instrumentation.

High-molecular-weight complexes have been successfully analyzed on time-of-flight mass spectrometers, but recent developments on Orbitrap mass spectrometers have extended the m/z range of these instruments to detect ion signals for biomolecular molecules up to MDa in size. Heck and co-workers modified a Q Exactive (QE) Plus mass spectrometer to a lower RF frequency of the ion optics to allow for better transmission characteristics of high- m/z ions.^{9,10} One limitation of high-mass ions is that as they fly from the ion optics through the bent flatapole, their high momentum causes rapid transfer of ions from atmospheric pressure to high vacuum, which causes ion instability and leads to sensitivity loss. The addition of in source trapping within the injection flatapole decreases the ion momentum prior to entering the bent flatapole, therefore increasing sensitivity.¹⁰ Modifications to the stacked ring ion

guide (SRIG) exit lens (E1), injection flatapole (E2), and injection flatapole lens (E3) allow for independent control of voltage and two operation modes, normal operation and trapping enabled. In normal operations, the voltage among the E1, E2, and E3 lenses are unmodified. In trapping mode, ions are subjected to a negative potential gradient change between the E1 and the E2, while the injection E3 potential is increased. The difference in voltage among the E1, E2, and E3 enables a trap in E2 for the ions. In addition, ions trapped in the E2 collide with background gas to decrease their momentum. The electrical potential between the ion optics are pulsed back to normal operation, and ions are allowed to exit E2 through E3 reaching the bent flatapole. These modifications allow for transmission of high m/z ions to the C-trap, but transmission of ions from the C-trap into the mass analyzer need to be modified as well. Under standard conditions, when ions are ejected from the C-trap to the analyzer, a deflector electrode turns the ions into the entrance slot of the Orbitrap. The deflector is pulsed with a voltage to minimize field perturbations and “block” any ions from entering the analyzer. The standard pulse is not long enough to allow large m/z ions flight from the C-trap to the analyzer, where the pulse time is increased to allow high m/z ions to enter the analyzer, keeping the Orbitrap “open” for longer. This ultrahigh-mass range (UHMR) instrument has enhanced detection of a homogeneous Cp180-mer virus-like protein, showing well-resolved ion signals up to 70 000 m/z (Figure 3).¹⁰ This mass range has not been achieved by other MS instruments.

Waterbeemd et al., demonstrated the use of the QE-UHMR to study the prokaryotic 30S, 50S, and 70S ribosomal particles, enabling high-resolution MS analysis of large molecules from 0.8 to 2.3 MDa. Native MS of the ribosomal particles from *E. coli* revealed substoichiometric association of the small protein SRA.⁹ This demonstrates the high sensitivity of the instrument since LC/MS/MS studies indicated that only ~22% of the 30S particles are bound to SRA. In addition, the mass of the 30S particle established the presence of the RS1 protein. This protein, which is essential for translation, is highly dynamic and is therefore not present in any high-resolution structure. This indicates the ability of native MS to study dynamic complexes not possible by other structural biology methods. Native MS was able to detect the heterogeneity (i.e., ribosome-interacting proteins that are recruited at different stages of translation) within the ribosomal particles, which has been proven difficult by higher resolution structural methods. This type of high-resolution MS opens doors for studies of ribosomal-binding drugs, antibiotics, and initiation or elongation factors or other large complexes associated with disease states.

ION MOBILITY

Use of structural proteomics to provide high-throughput characterization of three-dimensional biomolecules has been enabled by enhancing the dynamic range and sensitivity of MS techniques. Improvements to instrumentation have helped accomplish these goals. These improvements included the development of ion mobility spectrometry (IMS), which separates gas-phase ions based on their size and charge. IMS compared to other MS techniques allows the study of cohabitating biomolecules in a physiologically relevant state. Different techniques for IMS include drift-time ion mobility spectrometry (DTIMS), traveling-wave ion mobility spectrometry (TWIMS), field-asymmetric ion mobility spectrometry (FAIMS), and trapped ion mobility spectrometry (TIMS). DTIMS, TWIMS,

and FAIMS each utilize electric fields to focus and/or propel ions toward a detector while flowing an inert gas like helium or nitrogen over the ions. Ions that are smaller and more compact can travel through the inert gas faster than larger and more elongated ions. TIMS performs similarly but utilizes the electric field to trap ions, and the inert gas pushes the ions toward the detectors, allowing larger ions to enter the detector first and smaller ions last. In all of these techniques, ions at the same charge state are separated based on their size and shape, which is represented through their orientationally averaged cross collision sections (CCS). The fundamentals of IMS and the differences between each technique have previously been described in detail.^{3,17,18} Here we present recent advancements in IM-MS to characterize antibodies.

Characterizing Antibodies with Ion Mobility.

The developments in structural proteomics have allowed large biomolecules, like antibodies, to be characterized successfully.^{19–26} For many years, monoclonal antibodies (mAbs) have been the top therapeutic candidates for various diseases including cancer, autoimmune diseases, and infections. Detection of small structural changes are necessary to monitor the quality of antibodies, compare isoforms, and characterize biosimilar monoclonal antibodies. Current IMS technology does not have the resolving power to distinguish small structural changes in larger proteins.²⁷ It has been shown that heating gas-phase ions with a small amount of collisional energy immediately before IMS has improved separation.²⁸ In 2014, Zhong et al. utilized collision-induced unfolding (CIU) to improve the detection of small structural changes in large biomolecules (Figure 4).²⁷ CIU traps gas-phase native proteins and slowly increases their internal energy with collisional energy of a background gas, like argon.²⁹ By limiting the extent of collisional energy, proteins unfold, but the integrity of the amide backbone is maintained.^{30–33} The gas-phase unfolding pattern, or the “fingerprint” plot of the protein, gives insight into protein stability. This has allowed small structural changes, like changes in disulfide bonds, differing glycosylation, or epitope location to be detected. By comparing CIU fingerprints, many different IgG subtypes have been distinguished that differ in either the number of disulfide bonds or locations between the heavy and light chains.²⁵ No other method has shown to distinguish these changes in the interdomain connectivity as quickly and easily by comparing the CIU fingerprints.

CIU has also been used to determine differing extents of mAb glycosylations and structural differences to aid in the comparison of biosimilar mAbs.³⁴ Recently, Remsima was the first biosimilar mAb to get FDA approval for clinical use, and in the next few years, patents of many other mAbs will expire. Developing biosimilar mAbs will provide significant cost savings to patients. To help streamline the comparison of future biosimilar mAbs, Pisupati et al. developed a template using native CIU IM-MS and other MS methods to evaluate and characterize the biosimilar product, Remsima to the original product, Remicade.

The use of CIU fingerprints to characterize epitope changes has recently been performed. Huang et al. observed changes in CIU fingerprints among three malarial antigen–antibody complexes.³⁵ PvDBP is a vaccine to prevent malaria with several inhibitory antibodies, where three of those inhibitory antibodies have known epitopes. The two antibodies, 2D10 and 2H2, have the same epitope, while the third antibody, 2C6, binds to a different epitope.

All three antibody/antigen complexes showed similar trends in their CIU fingerprints, but slight changes between 2D10/2H2 and 2C6 were observed. Complexes with 2D10 and 2H2 had identical drift-time transitions, while the 2C6 complex had shorter drift times. Also, the 2C6 complex required less energy to begin unfolding as compared to 2D10 and 2H2 complexes. The change in drift time and decreased stability indicated a more compact but less stable structure because of the different epitope location for the 2C6 complex.

Most recently, Tian and Ruotolo observed slight structural changes in antibody glycoforms with the use of CIU.²⁶ A standard antibody, IgG1, was used with varying glycoforms that were created through various reactions. By comparing each IgG1 glycoform's CIU fingerprint, they saw subtle decreased stabilities with each sugar molecule addition. This stability represents structural changes taking place in IgG1 because of the altered glycan. Continued work on characterizing mAb is desired to make this a common analytical tool in biotherapeutic development. With continued work, this could become a quick and relatively easy assay capable of characterizing mAb structure and identifying glycosylations that can be used for the analysis of biosimilar mAb and development of new glycoproteins.

IMS has become a valuable tool not only for structural proteomics but lipidomics,³⁶ metabolomics,³⁷ and as a secondary separation of complex systems.^{38–47} By pushing the limits of what IMS can do for scientists, developments will continue to grow, providing better conformer separation and CCS calculations. This will make IMS a bioanalytical tool that improves dynamic range with a high throughput, providing quick and descriptive information.

TOP-DOWN MS

In contrast to bottom-up proteomics, where enzymatic digestion is used to generate peptides for MS analysis, intact proteins are directly fragmented in top-down proteomics. This approach provides better detection of degradation products, sequence variants, and low-mass proteins. Top-down is especially useful for the identification of individual and combinations of post-translational modifications (PTMs). PTM analysis by bottom-up proteomics is hampered by low ionization efficiency of modified tryptic peptides and limited sequence coverage (50–90%) of the protein of interest compared to the 100% sequence coverage gained from top-down.⁴⁸ In addition, top-down fragmentation has the ability to detect conformer-specific PTMs. This spatial information is lost when proteins are digested into peptides.

With steady advances in the mass-spectrometry technologies, including instrumentation and software development, top-down proteomics has improved over the last 30 years, broadening its use to multiple laboratories. Developments in Fourier transform MS (FTMS) instrumentation, a widely used instrument for top-down analysis, has significantly increased the capabilities of top-down analysis. Recent FTMS instruments have the ability to reach resolutions as high as 10 000 000, increasing the breadth of information, including the ability to resolve large proteins that can be obtained from top-down mass spectrometry.⁴⁹ Further, advancements in software have led to increased proteome coverage and reduced data acquisition time.⁵⁰

Characterization of Proteoforms by Top-Down MS.

A major advantage of using top-down proteomics is the ability to identify proteoforms. Proteoforms encompass all of the molecular forms of a protein translated from a single gene.⁵¹ These different forms could arise from post-translational modifications, alternative gene splicing, or genetic variations (Figure 5). Studies have demonstrated that proteoforms generated from alternate splicing lead to functional diversity, including differences in interaction networks and subcellular localizations highlighting the importance of identifying different proteoforms.⁵² Top-down provides the advantage of the full characterization of the proteoform, detecting details such as sequence variants and post-translational modifications.⁵³ Other approaches for studying proteoforms and their complexes, which include yeast two hybrid analysis⁵² and proteoform detection through intact mass and lysine count,⁵⁴ are more time-consuming than top-down MS, and in the case of yeast, two hybrid analysis can result in false positives for complex binding.

A challenge, however, in characterizing proteoforms by top-down proteomics is the lack of high-resolution separation techniques for complex protein samples. To address this challenge, Wang et al. recently developed a method to acquire better separation of intact proteins and better proteoform sequence coverage.⁵⁵ A two-dimensional separation approach using “salt-free” high-pH reverse-phase liquid chromatography (RPLC) coupled to low-pH RPLC was compared to separation using conventional one-dimensional low-pH RPLC with top-down MS. For *E. coli*, a total of 365 proteins and 886 proteoforms were identified with the 2D method compared to 163 proteins and 328 proteoforms identified using 1D separation.⁵⁵ This indicates the efficacy of using 2D separations to increase proteoform coverage.

A recent study by Melania et al. provided the first proteoform-specific look at the rapidly evolving sequences of the venom proteome. Top-down analysis was used to detect the sequence variation and PTMs in *Ophiophagus hannah* (king cobra) venom. In initial top-down experiments, 17 proteoforms were identified. To improve experimental dynamic range and obtain better coverage of the venom proteome, they then employed off-line fractionation prior to LC/MS/MS analysis. Using two fractionation methods, GELFrEE and solution isoelectric focusing (sIEF), 64 and 113 proteoforms were identified, respectively. Further manual validation to annotate amino acid substitutions, alternative processing cleavage sites, and presence of additional PTMs, resulted in the identification of 53 additional proteoforms, for a total 184 unique proteoforms. The knowledge of sequence variants in snake venom is important for understanding venom lethality. The identifications of larger proteins (>50 kDa) and low-abundant proteins were hindered by this approach, which utilized denaturing conditions for LC/MS/MS. To obtain better coverage of the venom proteome, they employed native MS strategies, which preserved the macromolecular interactions between the toxins and allowed for the characterization of the complexes.⁵⁶

The advantage of coupling top-down with native ESI is that it bridges the gap between proteomics and structural biology for protein complexes. Specifically, many of the fragile noncovalent interactions that are critical in maintaining the three-dimensional structure of proteins are conserved before controlled fragmentation of complexes into the subunits and their backbone fragment ions. The detailed characterization and increased level of coverage

achieved with native top-down proteomics allows the study of complexes larger than 100 kDa, representing a powerful middle ground for both targeted and discovery proteomics.⁵³ Native top-down proteomics performs better with larger heterogeneous proteoform distributions because of the spread of the intensity over only a few charge states. This is in contrast to denaturing conditions, where a greater number of charge states are observed leading to signal dilution.

One goal of native top-down proteomics is to directly identify and characterize proteoforms and multiproteoform complexes (MPCs) in a single experiment. MPCs are specific protein complexes formed by different proteoforms (i.e., monomeric proteins and PTMs) from the same or different genes. Using native top-down MS, Skinner et al. identified three related MPCs of the human diphosphate kinase (NDPK) heterohexamer. The identification and characterization of three related MPCs in a single experiment, a feat that is difficult or impossible using other approaches, exemplifies that stoichiometry and precise subunit composition can be quickly determined using a native top-down approach to protein complex analysis.⁵⁷ Beyond sample preparation and separation strategies that retain native protein structure, the key to native top-down proteomics is using a three-tiered tandem MS approach to controllably disassemble protein complexes in the gas phase.⁵⁸ Using this approach, MPCs were identified, characterized, and scored. Also, stoichiometry was calculated using all sources of mass such as cofactors and PTMs.⁵⁷ This study highlights the fact that an unknown complex can potentially be identified by its subunits through native top-down MS. This platform expands the ability of MS to integrate proteomics and structural biology to provide insights into protein structure, functions, and regulation.

To provide even more structural information from top-down, Li et al. used complementary biophysical methods native top-down and ion mobility to elucidate the structural changes of α -synuclein induced by the binding of divalent metal ions (cobalt and manganese).⁵⁹ Mn-/Co-binding interactions can influence protein folding, which may be important factors for neurodegenerative diseases such as Parkinson's disease. The group was able to locate similar binding sites for both Co and Mn on the C-terminus of α -synuclein using native top-down mass spectrometry with IMS.⁵⁹ This demonstrates the ability of top-down proteomics to identify binding interaction sites.

CHEMICAL CROSS-LINKING

Chemical cross-linking mass spectrometry (XL-MS) has emerged as a powerful structural mass-spectrometry method with particular utility for the characterization of protein–protein interactions in the context of large macromolecular assemblies.⁶⁰ It involves the generation of intra or intermolecular covalent linkages between amino acid side chains and/or other functional groups of proteins, or with other biological macromolecules, using reactive small molecules of known dimensions and specific chemistries.⁵ The successful identification of cross-linked protein/peptides then directly reflects on spatial proximity, which reports on secondary, tertiary, and quaternary structures of proteins as well as transient protein/protein or protein/macromolecule interactions. Additionally, the accurate determination of the identity of the cross-linked side chains provide a measurable structural parameter in the form of distance constraints between two amino acids obtained from an in-solution native

structural ensemble. Although the use of chemical cross-linking for structural and/or system biology studies predates mass-spectrometry technology, the advent of mass spectrometry (MS) provided a powerful analytical platform to reliably identify the side chain involved in the cross-linking reaction in a high-throughput fashion.

Advancements in Cross-Linking Reagents.

The development of cross-linkers has generally focused toward charged or polar side chains, as they are more likely to be found to be solvent accessible and therefore more likely to undergo cross-linking. The most widely used and commercially available reagents are the homo-bifunctional *N*-hydrosuccinimide (NHS) esters, which preferentially react with the amine groups of lysines and the N-terminus. A drawback of using NHS esters is that they mainly target lysine residues, which are highly flexible, where the distance information obtained is not very stringent.⁶⁰ More recently, diazirines have been used for photo-cross-linking of proteins where cross-linking is induced by the formation of a reactive carbene after irradiation with UV-A light. Diazirines do not display a preference for reactivity with amino acids like NHS esters. Photoactivatable diazine-based amino acids, photoleucine and photomethionine, offer the additional advantage of conducting cross-linking in the cell. The short cross-linker length of the photoactivatable diazirines provides more stringent topological information on proteins.

Recently, cleavable cross-linkers under tandem MS conditions have been an area of interest for XL-MS, as they automate the complex data analysis. Protein interaction reporter (PIR)⁶¹ technology developed for XL-MS experiments takes advantage of using a MS-cleavable cross-linker, which has been used to study protein interactions with purified complexes and cell culture and has recently been performed in heart tissue. Iacobucci et al. presented the first MS-cleavable photothiol-diazirine reactive cross-linker for the study of protein structure (Figure 6).⁶² Diazirine-based photo-cross-linkers were shown to generate CID-MS/MS-cleavable cross-linkers that are advantageous to develop fully automated data analysis allowing for proteome-wide cross-linking studies. 1,3-Diallylurea (DAU) was used as the photoreactive cross-linker, where cysteine residues undergo a “click reaction” that yields stable alkyl sulfide products. The central urea bond is then cleaved upon collisional activation during tandem MS, generating characteristic product ions that aid in the automated cross-link identification. The DAU technology was applied to thiol-containing proteins, bMunc 13–2 and GCAP-2, to demonstrate the applicability of DAU in structural proteomics.

Structural Modeling in XL-MS.

XL-MS has specially provided tremendous benefits in the study of increasingly larger macromolecular assemblies. Structural characterization of a large macromolecular machine is usually difficult to attain by standard structural approaches alone. XL-MS is amenable to probing such large complexes and has provided invaluable input into integrative approaches that include cryoEM, crystallography, SAXS, native spray, and hydrogen–deuterium exchange. Such studies have resulted in detailed structural and architectural characterization of the 26S proteasome,⁶³ nuclear pores,⁶⁴ CRISPR⁶⁵ and CRISPR-cas,⁶⁶ ribosomes,^{67,68} TFIID,⁶⁹ and pore forming toxins,⁷⁰ among others. The power of XL-MS in probing the

architecture of large multi-subunit or multiprotein subunits is particularly evident in conjunction with cryo-EM. Technological advances have permitted dramatic improvement in resolution attained by cryo-EM, and megadalton-sized macromolecular complexes have been reported at near atomic resolution. The reconstruction of a high-resolution model from the cryo-EM-generated 3D volume requires fitting already available subcomponent structures within the larger complex or, alternatively, de novo modeling, if the resolution permits it. For that purpose, XL-MS has proven especially helpful in providing location, spatial orientation, and distance constraints to aid in the reconstruction of the high-resolution model.⁷¹ This is particularly true in the case of the limited resolution (local or global) of the 3D volume generated by cryo-EM, where de novo modeling and unambiguous fitting of known structures is difficult. Studies involving the use of XL-MS to augment cryo-EM structure determination include the structure of the 26S proteasome,⁶³ the Polycomb Repressor Complex 2,⁷² the chromatin remodeling complex,⁷³ chromosome segregation complexes,^{74–76} signalosomes,⁷⁷ the membrane embedded receptor,⁷⁸ nuclear pore complexes,^{64,79} eukaryotic ribosomal complexes,^{67,68,80} the mammalian mitochondrial complex,⁸¹ various RNA polymerase complexes,^{82–86} and various spliceosome complexes.^{87,88}

XL-MS in Cells.

The use of a bottom-up proteomics workflow with XL-MS enables the study of protein structure on the proteome-wide scale. The potential benefit of such expansion is the increased biological significance of reporting structural information within the cellular or tissue context, potentially bridging the gap between structural biology and systems biology. Taking advantage of the advancement in cross-linker technology, studies have characterized the network of interaction around Hsp90 in vivo,⁸⁹ protein communities in cell lysates,⁹⁰ the interactome in mitochondria⁹¹ and in bacteria,⁹² and the perturbation of the interactome upon viral infection of a cell.⁹³ Häupl et al., incorporated photomethionine and photoleucine amino acids into proteins in HeLa cell cultures to study affinity enriched protein complexes of interest.⁹⁴ Photoaffinity-labeled proteins were enriched, and protein complexes were covalently fixed by activation of photo-amino-acid diazirines with UV-A irradiation followed by sample digestion and MS analysis. The study highlighted the advantage of using photo-cross-linkers in cell cultures to identify protein–protein interactions. Finally, the most recent advances have performed XL-MS in tissue,⁹⁵ opening the potential of XL-MS in applications such as clinical diagnostics.

XL-MS in Heart Tissue.

XL-MS was performed in mouse heart tissue to identify protein–protein interactions relevant to organ-level disease states that cannot be obtained with cell culture studies. To obtain structural information, heart tissue was isolated from mice and subjected to chemical cross-linking with the protein interaction reporter⁶¹ cross-linker. Samples were either processed as whole-heart samples, or the mitochondria was isolated by subcellular fractionation. A real time MS3 technique termed ReACT was used to analyze in-tissue cross-linked peptide pairs. ReACT is an MS method that uses low-energy fragmentation to release the cross-linked peptides and the reporter ion. The mass of the reporter ion and the cross-linked peptides in the MS2 spectra are searched in real time, and the two released cross-linked peptides are

further fragmented. A total of 572 peptide pairs were detected from all the major sacromeric proteins, where the identified cross-linked pairs were used as probes for protein interactions for future XL-MS studies with diseased heart tissue. The mitochondria were extracted from heart tissue, because cardiomyocytes contain the highest concentration of mitochondria of any cell in the body, which would allow the analysis of oxidative phosphorylation complexes (OXPHOS) (Figure 7). Multiple links were identified between the complexes, suggesting the existence of larger supercomplex assemblies in agreement with single-particle EM results. In total, more than 2000 links were identified between the complexes, which can be used to compare supercomplex assemblies in normal and failing heart tissues to understand their role in mitochondrial function in heart disease.

COVALENT LABELING

Coupling covalent labeling techniques with MS has enabled the study of protein structure and protein–protein interactions. Covalent labeling (CL) is a protein surface modification technique that relies on the specific (diethylpyrocarbonate (DEPC))⁹⁶ and nonspecific (hydroxyl radicals, carbenes, and deuterium)^{6,97} labeling reagents that form new covalent bonds. Hydrogen–deuterium exchange coupled with MS (HDX-MS) relies on the exchange of backbone amide hydrogens with deuterium, leading to a mass increase in the regions that undergo exchange.^{97–101} This provides information about rigid/dynamic regions of the protein that can be used to study conformational dynamics of proteins. CL-MS techniques using hydroxyl radicals, DEPC, and carbenes rely on the formation of new covalent bonds with solvent accessible amino acid side chains providing information on the proteins surface structure.⁹⁶

DEPC, a specific covalent label, modifies Cys, His, Lys, Thr, and Ser residues, resulting in a +72.021 mass shift.¹⁰² There have been multiple applications using DEPC to study protein structure and protein–protein interactions such as studying protein-binding sites of amyloid inhibiting molecules.¹⁰³ The advantage of using a specific covalent label is that it reduces the complexity of data analysis. A recently developed nonspecific covalent label is the use of carbenes that are photoreactive.¹⁰⁴ The addition of diazirines to the solution followed by irradiation at near-ultraviolet wavelength (310–350 nm), which is outside the absorbance window of amino acids, results in the formation of reactive carbenes. Once formed, the carbene has an approximate lifetime of a few ns, which allows fast sampling of a protein state. Carbene footprinting has been successfully performed on a FPOP flow systems by photolysis of diazine precursors.¹⁰⁵ Recently, Schriemer et al. have used three different diazirines to understand the amino acid insertion frequency of aliphatic carbenes (Figure 8). The study successfully showed that the carbene precursors can access all 20 amino acids over a relatively narrow range of insertion frequencies.¹⁰⁶ To date, carbene footprinting is the only covalent labeling method that can modify all 20 amino acids.

Another nonspecific covalent labeling technique is hydroxyl radical footprinting (HRF) coupled with mass spectrometry to study protein structure.^{107,108} The generated hydroxyl radicals modify 17 of the 20 amino acid side chains on the protein surface, resulting, in most cases, in the addition of an oxygen providing information on solvent accessibility.⁶ One HRF method, fast photochemical oxidation of proteins (FPOP) utilizes a pulsed laser, which

is used to photolyze hydrogen peroxide on a microsecond time scale, which is faster than protein unfolding.^{109,110} FPOP has been applied to many purified protein complexes and systems, revealing a wealth of information on protein structure and protein–protein interactions.^{108,110–113}

De Novo Modeling with Covalent Labeling.

Similar to XL-MS, de novo modeling is also an emerging field within covalent labeling. Recent advancements in HRF, including normalization of radical reactivity, have aided the ability to use HRF for molecular modeling. The varying reactivity of residues with hydroxyl radicals has been a challenge in directly correlating HRF data with protein structure, because highly reactive residues that may not be as solvent accessible have a higher extent of modification compared to residues that are solvent accessible but are less reactive. When only comparing different states, i.e., ligand-free vs ligand-bound, these reactivity differences are not a major factor. However, direct correlation of the HRF extent of modification data to protein structure is hampered by these reactivity differences. Huang et al. developed an algorithm that correlates the experimental footprinting rate to a protection factor that is based on the hydroxyl radical reactivity of the modified residue. Normalization of footprinting rates provides more quantitative information on the structure allowing HRF data to be used for molecular modeling.¹¹⁴ Xie et al. used HRF data to distinguish between molecular models that were generated from high- and low-accuracy crystal structures.¹¹⁵ Recently, Aprahamian et al. used HRF-derived protection factors as a score term in Rosetta to obtain atomic resolution of proteins and predict tertiary structures.¹¹⁶ De novo models of four different proteins, calmodulin, cytochrome c, myoglobin, and lysozyme, were generated using Rosetta only and with HRF data as constraints. The incorporation of HRF data increased the RMSD of each structure (Figure 9). Atomic-level resolution was achieved for two structures, cytochrome c and myoglobin, upon the inclusion of HRF data as constraints. This is the first time that de novo structural modeling has been achieved from HRF mass-spectrometry-derived protection factors.

Similar to HRF modeling, other covalent labeling techniques such as DEPC coupled with mass spectrometry have been used to model dynamics of macromolecules. Schmidt et al. generated models from a restraint-based strategy and subjected those models to molecular dynamics simulations to investigate structure and functions of large complexes.¹¹⁷ The solvent accessibility information acquired from covalent labeling MS was converted into modeling restraints using an in-house developed code that estimated the solvent accessible surface area for all models generated using a sampling algorithm. The method successfully reconstructed the 3D assembly structure of three model proteins (tryptophan synthetase, carbonyl phosphate synthetase, and cATPase) with high accuracy and precision. Combining structural modeling and experimental data allows large complexes to be studied that are not feasible by traditional tools used to study protein complexes.

In-Cell FPOP.

Macromolecular crowding in the cell has shown to impact interactions of proteins and protein conformational states, which elucidates the importance of using a whole cell approach to study proteins.¹¹⁸ For HRF MS, FPOP has been successfully applied to African

green monkey kidney (Vero) cells to study protein structure and protein–protein interactions directly in the cellular environment.^{119,120} Hydrogen peroxide is permeable to cellular and organelle membranes, which enables in-cell FPOP (IC-FPOP) to oxidatively modify proteins in different cellular compartments. Since hydrogen peroxide was added to cells, cell viability assays were performed to demonstrate that FPOP was probing live cells. In the time scale of the FPOP experiment using 20 mM hydrogen peroxide, the majority of the cells (70%) were viable. To increase the number of oxidatively modified proteins, a flow system was developed to allow single-cell flow based on the principles of hydrodynamic focusing. Over 1300 proteins were oxidatively modified by modifications made to the flow system, which allowed equal cell exposure to laser irradiation.¹¹⁹ The modified proteins were present in 27 different compartments of the cell, verifying that hydrogen peroxide was able to penetrate various cellular organelles. The dynamic range of IC-FPOP was evaluated by comparing the oxidized proteins with their expression levels (transcripts per million, TPM) in human kidney cells. High-abundance proteins like actin (3154 TPMs) and low-abundance proteins like Protein Shroom 2 (4 TPMs) were oxidatively modified using the flow system. This shows that IC-FPOP has a large dynamic range.

In Vivo FPOP.

Cell culture studies provide structural information on proteins in their native cellular environment but do not take into consideration interactions at an organ level. Recently, FPOP coupled with MS has been expanded into an animal model, *Caenorhabditis elegans* (*C. elegans*), to study protein–protein interactions in vivo. Similar to IC-FPOP, a flow system was used for in vivo FPOP (IV-FPOP) for single-worm flow to avoid clumping of the worms and unequal exposure to laser irradiation.¹²¹ Initial studies indicate that several proteins in varying body systems in the worm were oxidatively modified. The development of IV-FPOP offers the opportunity to study a wide range of biological processes and disease states.

CONCLUSION

Here, we have presented recent advancements in MS-based structural proteomic methods to study noncovalent complexes and macromolecular interactions. Each of these different methods (native, IMS, top-down, chemical cross-linking, and covalent labeling) provide unique structural information that can be used to characterize proteins. With novel development in instrumentation, software, and methodology, the scope of molecules that can be studied by these methods has increased to large proteins and complexes. These developments have also enabled the ability to gain structural information not only from purified proteins but also from direct analysis of proteins in cells, tissues, and organisms. As these techniques are further developed, the characterization of macromolecular systems will be better understood and will provide information on disease states through structural-based MS.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Science Foundation MCB1701692 (to LMJ).

Biographies

Upneet Kaur graduated from University of Massachusetts Amherst with a B.S. in biochemistry and molecular biology. She joined Lisa Jones's lab in 2017 and has been working on studying protein interactions in cells.

Danté T. Johnson graduated from Louisiana State University with a B. S. in biological sciences with a minor in chemistry. She is pursuing her Ph.D in pharmaceutical sciences and joined Lisa Jones's lab in 2017. Her research focuses on the development of a new platform for protein folding studies.

Emily Chea graduated from Taylor University (Indiana, USA) with a BA. in chemistry. She joined Lisa Jones's group in 2015 and is working towards a Ph.D. in pharmaceutical sciences. Currently, her research focuses on IC-FPOP to study protein interactions in cells.

Daniel J. Deredge graduated from Louisiana State University with a Bachelor's of Science and Ph.D. in biochemistry where he studied solution thermodynamics of protein–DNA interaction. As a postdoctoral researcher at Case Western Reserve University and at University of Maryland School of Pharmacy, his studies have focused on structural and biophysical characterization of proteins using mass-spectrometry-based approaches. He is currently a research assistant professor.

Jessica A. Espino graduated from Indiana University – Purdue University Indianapolis with a B.S. in Forensic Science. She joined the pharmaceutical sciences program at the University of Maryland, Baltimore in 2016 and is working towards a Ph.D. in pharmaceutical sciences in Lisa Jones's lab. Currently, her research focuses on the development of in vivo FPOP to study protein interactions.

Lisa Jones completed her Ph.D. at Georgia State University (Atlanta, GA) where she trained in structural biology. As a postdoctoral researcher at the University of Alabama-Birmingham, she used mass-spectrometry-based structural methods to characterize virus particles. As a postdoctoral researcher at Washington University in St. Louis, she focused on using hydroxyl radical footprinting to characterize protein interaction sites. In 2012, she became an assistant professor at Indiana University-Purdue University Indianapolis. Currently, she is an assistant professor at the University of Maryland.

REFERENCES

- (1). Boeri Erba E; Petosa C Protein Sci. 2015, 24 (8), 1176–1192. [PubMed: 25676284]
- (2). Chen B; Brown KA; Lin Z; Ge Y Anal. Chem 2018, 90 (1), 110–127. [PubMed: 29161012]
- (3). Kanu AB; Dwivedi P; Tam M; Matz L; Hill HH Jr. J. Mass Spectrom 2008, 43 (1), 1–22. [PubMed: 18200615]
- (4). Sinz A Angew. Chem. Int. Ed 2018, 57 (22), 6390–6396.
- (5). Chavez JD; Bruce JE Curr. Opin. Chem. Biol 2019, 48, 8–18. [PubMed: 30172868]
- (6). Chea EE; Jones LM Analyst 2018, 143 (4), 798–807. [PubMed: 29355258]
- (7). Back JW; de Jong L; Muijsers AO; de Koster CG J. Mol. Biol 2003, 331 (2), 303–13. [PubMed: 12888339]

- (8). Kaur U; Meng H; Lui F; Ma R; Ogburn RN; Johnson JHR; Fitzgerald MC; Jones LM J. *Proteome Res* 2018, 17, 3614. [PubMed: 30222357]
- (9). van de Waterbeemd M; Fort KL; Boll D; Reinhardt-Szyba M; Routh A; Makarov A; Heck AJR *Nat. Methods* 2017, 14, 283. [PubMed: 28114288]
- (10). Fort KL; van de Waterbeemd M; Boll D; Reinhardt-Szyba M; Belov ME; Sasaki E; Zschoche R; Hilvert D; Makarov AA; Heck AJR *Analyst* 2018, 143 (1), 100–105.
- (11). Gupta K; Li J; Liko I; Gault J; Bechara C; Wu D; Hopper JTS; Giles K; Benesch JLP; Robinson CV *Nat. Protoc* 2018, 13, 1106. [PubMed: 29700483]
- (12). Marty MT; Hoi KK; Robinson CV *Acc. Chem. Res* 2016, 49 (11), 2459–2467. [PubMed: 27736086]
- (13). Marty MT; Zhang H; Cui W; Blankenship RE; Gross ML; Sligar SG *Anal. Chem* 2012, 84 (21), 8957–60. [PubMed: 23061736]
- (14). Hopper JT; Yu YT; Li D; Raymond A; Bostock M; Liko I; Mikhailov V; Laganowsky A; Benesch JL; Caffrey M; Nietlispach D; Robinson CV *Nat. Methods* 2013, 10 (12), 1206–8. [PubMed: 24122040]
- (15). Marty MT; Hoi KK; Gault J; Robinson CV *Angew. Chem., Int. Ed* 2016, 55 (2), 550–554.
- (16). Reid DJ; Keener JE; Wheeler AP; Zambrano DE; Diesing JM; Reinhardt-Szyba M; Makarov A; Marty MT *Anal. Chem* 2017, 89 (21), 11189–11192. [PubMed: 29048874]
- (17). Gabelica V; Marklund E *Curr. Opin. Chem. Biol* 2018, 42, 51–59. [PubMed: 29154177]
- (18). Cumeras R; Figueras E; Davis CE; Baumbach JI; Gracia I *Analyst* 2015, 140 (5), 1376–90. [PubMed: 25465076]
- (19). Tian Y; Ruotolo BT *Analyst* 2018, 143 (11), 2459–2468. [PubMed: 29736508]
- (20). Beck A; Debaene F; Diemer H; Wagner-Rousset E; Colas O; Van Dorsselaer A; Cianferani SJ *Mass Spectrom.* 2015, 50 (2), 285–297.
- (21). Damen CWN; Chen W; Chakraborty AB; van Oosterhout M; Mazzeo JR; Gebler JC; Schellens JHM; Rosing H; Beijnen JH *J. Am. Soc. Mass Spectrom* 2009, 20 (11), 2021–2033. [PubMed: 19744865]
- (22). Ehkirch A; Hernandez-Alba O; Colas O; Beck A; Guilleme D; Cianferani S *J. Chromatogr. B: Anal. Technol Biomed. Life Sci* 2018, 1086, 176–183.
- (23). Ferguson CN; Gucinski-Ruth AC *J. Am. Soc. Mass Spectrom* 2016, 27 (5), 822–33. [PubMed: 26988372]
- (24). Terral G; Beck A; Cianferani S *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci* 2016, 1032, 79–90.
- (25). Tian Y; Han L; Buckner AC; Ruotolo BT *Anal. Chem* 2015, 87 (22), 11509–11515. [PubMed: 26471104]
- (26). Tian Y; Ruotolo BT *Int. J. Mass Spectrom* 2018, 425, 1–9.
- (27). Zhong Y; Han L; Ruotolo BT *Angew. Chem., Int. Ed* 2014, 53 (35), 9209–12.
- (28). Jarrold MF; Honea EC *J. Am. Chem. Soc* 1992, 114 (2), 459–464.
- (29). Dixit SM; Polasky DA; Ruotolo BT *Curr. Opin. Chem. Biol* 2018, 42, 93–100. [PubMed: 29207278]
- (30). Benesch JLP; Aquilina JA; Ruotolo BT; Sobott F; Robinson CV *Chem. Biol* 2006, 13 (6), 597–605. [PubMed: 16793517]
- (31). Hyung S-J; Robinson CV; Ruotolo BT *Chem. Biol* 2009, 16 (4), 382–390. [PubMed: 19389624]
- (32). Shelimov KB; Clemmer DE; Hudgins RR; Jarrold MF *J. Am. Chem. Soc* 1997, 119 (9), 2240–2248.
- (33). Shelimov KB; Jarrold MF *J. Am. Chem. Soc* 1997, 119 (13), 2987–2994.
- (34). Pisupati K; Tian Y; Okbazghi S; Benet A; Ackermann R; Ford M; Saveliev S; Hosfield CM; Urh M; Carlson E; Becker C; Tolbert TJ; Schwendeman SP; Ruotolo BT; Schwendeman A *Anal. Chem* 2017, 89 (9), 4838–4846. [PubMed: 28365979]
- (35). Huang Y; Salinas ND; Chen E; Tolia NH; Gross ML *J. Am. Soc. Mass Spectrom* 2017, 28 (11), 2515–2518. [PubMed: 28875466]
- (36). Hinz C; Liggi S; Griffin JL *Curr. Opin. Chem. Biol* 2018, 42, 42–50. [PubMed: 29145156]

- (37). Schrimpe-Rutledge AC; Sherrod SD; McLean JA *Curr. Opin. Chem. Biol* 2018, 42, 160–166. [PubMed: 29287234]
- (38). Achour B; Al Feteisi H; Lanucara F; Rostami-Hodjegan A; Barber J *Drug Metab. Dispos* 2017, 45 (6), 666–675. [PubMed: 28373266]
- (39). Garabedian A; Benigni P; Ramirez CE; Baker ES; Liu T; Smith RD; Fernandez-Lima F J. *Am. Soc. Mass Spectrom* 2018, 29 (5), 817–826. [PubMed: 28889248]
- (40). Gombar R; Pitcher TE; Lewis JA; Auld J; Vacratsis PO J. *Proteomics* 2017, 157, 1–9. [PubMed: 28161412]
- (41). Kedia K; Wendler JP; Baker ES; Burnum-Johnson KE; Jarsberg LG; Stratton KG; Wright AT; Piehowski PD; Gritsenko MA; Lewinsohn DM; Sigal GB; Weiner MH; Smith RD; Jacobs JM; Nahid P *Tuberculosis* 2018, 112, 52–61. [PubMed: 30205969]
- (42). Konig S; Hadrian K; Schlatt S; Wistuba J; Thanos S; Bohm MRR, Topographic protein profiling of the age-related proteome in the retinal pigment epithelium of *Callithrix jacchus* with respect to macular degeneration. *J. Proteomics* 2018;
- (43). Orwoll ES; Wiedrick J; Jacobs J; Baker ES; Piehowski P; Petyuk V; Gao Y; Shi T; Smith RD; Bauer DC; Cummings SR; Nielson CM; Lapidus J *Aging Cell* 2018, 17 (2), e12717.
- (44). Pettit ME; Donnarumma F; Murray KK; Solouki T *Anal. Chim. Acta* 2018, 1034, 102–109. [PubMed: 30193623]
- (45). Ruland T; Wolbert J; Gottschalk MG; Konig S; Schulte-Mecklenbeck A; Minnerup J; Meuth SG; Gross CC; Wiendl H; Meyer Zu Horste G *Frontiers in neurology* 2018, 9, 407. [PubMed: 29922220]
- (46). Shliaha PV; Baird MA; Nielsen MM; Gorshkov V; Bowman AP; Kaszycki JL; Jensen ON; Shvartsburg AA *Anal. Chem* 2017, 89 (10), 5461–5466. [PubMed: 28406606]
- (47). Steevensz A; Gombar R; Vergilino R; Cristescu ME; Vacratsis PO *Proteomics* 2018, 18 (16), 1700460.
- (48). Kelleher NL; Lin HY; Valaskovic GA; Aaserud DJ; Fridriksson EK; McLafferty FW J. *Am. Chem. Soc* 1999, 121 (4), 806–812.
- (49). Bruker. Top-Down Proteomics: An Emerging Technique Driven by Cutting-Edge Mass Spectrometry; 2017; https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/literature/Articles/1852452_Case_Studie_Dr_Ge_ebook.pdf.
- (50). Fornelli L; Durbin KR; Fellers RT; Early BP; Greer JB; LeDuc RD; Compton PD; Kelleher NL J. *Proteome Res* 2017, 16 (2), 609–618. [PubMed: 28152595]
- (51). Smith LM; Kelleher NL; The Consortium for Top Down Proteomics. *Nat. Methods* 2013, 10 (3), 186–187. [PubMed: 23443629]
- (52). Yang X; Coulombe-Huntington J; Kang S; Sheynkman GM ; Hao T; Richardson A; Sun S; Yang F; Shen YA; Murray RR; Spirohn K; Begg BE; Duran-Frigola M; MacWilliams A; Pevzner SJ; Zhong Q; Trigg SA; Tam S; Ghamsari L; Sahni N ; Yi S; Rodriguez MD; Balcha D; Tan G; Costanzo M; Andrews B; Boone C; Zhou XJ; Salehi-Ashtiani K; Charlotiaux B; Chen AA; Calderwood MA; Aloy P; Roth FP; Hill DE; Iakoucheva LM; Xia Y; Vidal M *Cell* 2016, 164 (4), 805–17. [PubMed: 26871637]
- (53). Stoakes SF Top-Down Approach in Proteomics. *News Medical* [Online]. 2018; <https://www.news-medical.net/life-sciences/Top-Down-Approach-in-Proteomics.aspx>.
- (54). Shortreed MR; Frey BL; Scalf M; Knoener RA; Cesnik AJ; Smith LM J. *Proteome Res* 2016, 15 (4), 1213–1221. [PubMed: 26941048]
- (55). Tran JC; Zamborg L; Ahlf DR; Lee JE; Catherman AD ; Durbin KR; Tipton JD; Vellaichamy A; Kellie JF; Li M; Wu C; Sweet SM; Early BP; Siuti N; Leduc RD; Compton PD; Thomas PM; Kelleher NL *Nature* 2011, 480 (7376), 254–258. [PubMed: 22037311]
- (56). Melani RD; Skinner OS; Fornelli L; Domont GB; Compton PD; Kelleher NL *Mol. Cell. Proteomics* 2016, 15 (7), 2423–34. [PubMed: 27178327]
- (57). Skinner OS; Haverland NA; Fornelli L; Melani RD; Do Vale LHF; Seckler HS; Doubleday PF; Schachner LF; Srzentic K; Kelleher NL; Compton PD *Nat. Chem. Biol* 2018, 14 (1), 36–41. [PubMed: 29131144]

- (58). Compton PD; Skinner O; Haverland N; Fornelli L; Doubleday P; Seckler H; Schachner L; Kelleher N FASEB J. 2017, 31 (1_supplement), 760.2.
- (59). Li H; Nguyen HH; Ogorzalek Loo RR; Campuzano IDG; Loo JA Nat. Chem 2018, 10 (2), 139–148. [PubMed: 29359744]
- (60). Gingras AC; Gstaiger M; Raught B; Aebersold R Nat. Rev. Mol. Cell Biol 2007, 8 (8), 645–54. [PubMed: 17593931]
- (61). Li H; Wongkongkathep P; Van Orden SL; Ogorzalek Loo RR; Loo JA J. Am. Soc. Mass Spectrom 2014, 25 (12), 2060–2068. [PubMed: 24912433]
- (62). Iacobucci C; Piotrowski C; Rehkamp A; Ihling CH; Sinz A, The First MS-Cleavable, Photo-Thiol-Reactive Cross-Linker for Protein Structural Studies. J. Am. Soc. Mass Spectrom 2018.
- (63). Lasker K; Forster F; Bohn S; Walzthoeni T; Villa E; Unverdorben P; Beck F; Aebersold R; Sali A; Baumeister W Proc. Natl. Acad. Sci. U. S. A 2012, 109 (5), 1380–7. [PubMed: 22307589]
- (64). Kosinski J; Mosalaganti S; von Appen A; Teimer R; DiGuilio AL; Wan W; Bui KH; Hagen WJ; Briggs JA; Glavy JS; Hurt E; Beck M Science 2016, 352 (6283), 363–5. [PubMed: 27081072]
- (65). Benda C; Ebert J; Scheltema RA; Schiller HB; Baumgartner M; Bonneau F; Mann M; Conti E Mol. Cell 2014, 56 (1), 43–54. [PubMed: 25280103]
- (66). Hrle A; Maier LK; Sharma K; Ebert J; Basquin C; Urlaub H; Marchfelder A; Conti E RNA Biol. 2014, 11 (8), 1072–82. [PubMed: 25483036]
- (67). Greber BJ; Boehringer D; Leitner A; Bieri P; Voigts-Hoffmann F; Erzberger JP; Leibundgut M; Aebersold R; Ban N Nature 2014, 505 (7484), 515–9. [PubMed: 24362565]
- (68). Greber BJ; Boehringer D; Leibundgut M; Bieri P; Leitner A; Schmitz N; Aebersold R; Ban N Nature 2014, 515 (7526), 283–6. [PubMed: 25271403]
- (69). Gupta K; Watson AA; Baptista T; Scheer E; Chambers AL; Koehler C; Zou J; Obong-Ebong I; Kandiah E; Temblador A; Round A; Forest E; Man P; Bieniossek C; Laue ED; Lemke EA; Rappsilber J; Robinson CV; Devys D; Tora L; Berger I eLife 2017, 6, e30395. [PubMed: 29111974]
- (70). Brauning B; Bertosin E; Praetorius F; Ihling C; Schatt A; Adler A; Richter K; Sinz A; Dietz H; Groll M Nat. Commun 2018, 9 (1), 1806. [PubMed: 29728606]
- (71). Schmidt C; Urlaub H Curr. Opin. Struct. Biol 2017, 46, 157–168. [PubMed: 29080436]
- (72). Ciferri C; Lander GC; Maiolica A; Herzog F; Aebersold R; Nogales E eLife 2012, 1, e00005. [PubMed: 23110252]
- (73). Nguyen VQ; Ranjan A; Stengel F; Wei D; Aebersold R; Wu C; Leschziner AE Cell 2013, 154 (6), 1220–31. [PubMed: 24034246]
- (74). Petrovic A; Keller J; Liu Y; Overlack K; John J; Dimitrova YN; Jenni S; van Gerwen S; Stege P; Wohlgemuth S; Rombaut P; Herzog F; Harrison SC; Vetter IR; Musacchio A Cell 2016, 167 (4), 1028–1040.e15. [PubMed: 27881301]
- (75). Hons MT; Huis In 't Veld PJ; Kaesler J; Rombaut P; Schleiffer A; Herzog F; Stark H; Peters JM Nat. Commun 2016, 7, 12523. [PubMed: 27549742]
- (76). Huis in't Veld PJ; Herzog F; Ladurner R; Davidson IF; Piric S; Kreidl E; Bhaskara V; Aebersold R; Peters JM Science 2014, 346 (6212), 968–72. [PubMed: 25414306]
- (77). Mosadeghi R; Reichermeier KM; Winkler M; Schreiber A; Reitsma JM; Zhang Y; Stengel F; Cao J; Kim M; Sweredoski MJ; Hess S; Leitner A; Aebersold R; Peter M; Deshaies RJ; Enchev RI eLife 2016, 5, e12102. [PubMed: 27031283]
- (78). Efremov RG; Leitner A; Aebersold R; Raunser S Nature 2015, 517 (7532), 39–43. [PubMed: 25470059]
- (79). von Appen A; Kosinski J; Sparks L; Ori A; DiGuilio AL; Vollmer B; Mackmull MT; Banterle N; Parca L; Kasttrit P; Buczak K; Mosalaganti S; Hagen W; Andres-Pons A; Lemke EA; Bork P; Antonin W; Glavy JS; Bui KH; Beck M Nature 2015, 526 (7571), 140–143. [PubMed: 26416747]
- (80). Greber BJ; Bieri P; Leibundgut M; Leitner A; Aebersold R; Boehringer D; Ban N Science 2015, 348 (6232), 303–8. [PubMed: 25837512]
- (81). Fiedorczuk K; Letts JA; Degliesposti G; Kaszuba K; Skehel M; Sazanov LA Nature 2016, 538 (7625), 406–410. [PubMed: 27595392]

- (82). Robinson PJ; Trnka MJ; Bushnell DA; Davis RE; Mattei PJ; Burlingame AL; Kornberg RD *Cell* 2016, 166 (6), 1411–1422.e16. [PubMed: 27610567]
- (83). Robinson PJ; Trnka MJ; Pellarin R; Greenberg CH; Bushnell DA; Davis R; Burlingame AL; Sali A; Kornberg RD *eLife* 2015, 4, e08719. [PubMed: 26402457]
- (84). Bernecky C; Herzog F; Baumeister W; Plitzko JM; Cramer P *Nature* 2016, 529 (7587), 551–4. [PubMed: 26789250]
- (85). Plaschka C; Lariviere L; Wenzek L; Seizl M; Hemann M; Tegunov D; Petrotchenko EV; Borchers CH; Baumeister W; Herzog F; Villa E; Cramer P *Nature* 2015, 518 (7539), 376–80. [PubMed: 25652824]
- (86). Xu Y; Bernecky C; Lee CT; Maier KC; Schwalb B; Tegunov D; Plitzko JM; Urlaub H; Cramer P *Nat. Commun* 2017, 8, 15741.
- (87). Agafonov DE; Kastner B; Dybkov O; Hofele RV; Liu WT; Urlaub H; Luhrmann R; Stark H *Science* 2016, 351 (6280), 1416–20. [PubMed: 26912367]
- (88). Bertram K; Agafonov DE; Dybkov O; Haselbach D; Leelaram MN; Will CL; Urlaub H; Kastner B; Luhrmann R; Stark H *Cell* 2017, 170 (4), 701–713.e11. [PubMed: 28781166]
- (89). Chavez JD; Schweppe DK; Eng JK; Bruce JE *Cell Chem. Biol* 2016, 23 (6), 716–26. [PubMed: 27341434]
- (90). Kastitis PL; O'Reilly FJ; Bock T; Li Y; Rogon MZ; Buczak K; Romanov N; Betts MJ; Bui KH; Hagen WJ; Hennrich ML; Mackmull MT; Rappsilber J; Russell RB; Bork P; Beck M; Gavin AC *Mol. Syst. Biol* 2017, 13 (7), 936. [PubMed: 28743795]
- (91). Schweppe DK; Chavez JD; Lee CF; Caudal A; Kruse SE; Stuppard R; Marcinek DJ; Shadel GS; Tian R; Bruce JE *Proc. Natl. Acad. Sci. U. S. A* 2017, 114 (7), 1732–1737. [PubMed: 28130547]
- (92). Zheng C; Yang L; Hoopmann MR; Eng JK; Tang X; Weisbrod CR; Bruce JE *Mol. Cell. Proteomics* 2011, 10 (10), M110.006841.
- (93). DeBlasio SL; Chavez JD; Alexander MM; Ramsey J; Eng JK; Mahoney J; Gray SM; Bruce JE; Cilia M J. *Virol* 2016, 90 (4), 1973–87. [PubMed: 26656710]
- (94). Haupl B; Ihling CH; Sinz A *Proteomics* 2017, 17, 1600459.
- (95). Chavez JD; Lee CF; Caudal A; Keller A; Tian R; Bruce JE *Cell Syst* 2018, 6 (1), 136–141.e5. [PubMed: 29199018]
- (96). Limpikirati P; Liu T; Vachet RW *Methods* 2018, 144, 79–93. [PubMed: 29630925]
- (97). Engen JR *Anal. Chem* 2009, 81 (19), 7870–5. [PubMed: 19788312]
- (98). Engen JR; Smith DL *Anal. Chem* 2001, 73 (9), 256A–265A.
- (99). Engen JR; Smith DL *Methods Mol. Biol* 2000, 146, 95–112. [PubMed: 10948498]
- (100). Xiao H; Kaltashov IA; Eyles SJ *J. Am. Soc. Mass Spectrom* 2003, 14 (5), 506–15. [PubMed: 12745220]
- (101). Kaltashov IA; Eyles SJ *Mass Spectrom. Rev* 2002, 21 (1), 37–71. [PubMed: 12210613]
- (102). Mendoza VL; Vachet RW *Mass Spectrom. Rev* 2009, 28 (5), 785–815. [PubMed: 19016300]
- (103). Liu T; Marcinko TM; Kiefer PA; Vachet RW *Anal. Chem* 2017, 89 (21), 11583–11591. [PubMed: 29028328]
- (104). Gomez GE; Mundo MR; Craig PO; Delfino JM *J. Am. Soc. Mass Spectrom* 2012, 23 (1), 30–42. [PubMed: 22006407]
- (105). Zhang B; Rempel DL; Gross ML *J. Am. Soc. Mass Spectrom* 2016, 27 (3), 552–5. [PubMed: 26679355]
- (106). Ziemianowicz DS; Bomgarden R; Etienne C; Schriemer DC *J. Am. Soc. Mass Spectrom* 2017, 28 (10), 2011–2021. [PubMed: 28799075]
- (107). Adilakshmi T; Lease RA; Woodson SA *Nucleic Acids Res.* 2006, 34 (8), e64. [PubMed: 16682443]
- (108). Hambly DM; Gross ML *J. Am. Soc. Mass Spectrom* 2005, 16 (12), 2057–63. [PubMed: 16263307]
- (109). Gau BC; Sharp JS; Rempel DL; Gross ML *Anal. Chem* 2009, 81 (16), 6563–6571. [PubMed: 20337372]
- (110). Chea EE; Jones LM *Protein Sci.* 2018, 27 (6), 1047–1056. [PubMed: 29575296]

- (111). Gau BC; Chen J; Gross ML *Biochim. Biophys. Acta, Proteins Proteomics* 2013, 1834 (6), 1230–8.
- (112). Kiselar JG; Janmey PA; Almo SC; Chance MR *Mol. Cell. Proteomics* 2003, 2 (10), 1120–32. [PubMed: 12966145]
- (113). Wang L; Chance MR *Anal. Chem* 2011, 83 (19), 7234–41. [PubMed: 21770468]
- (114). Huang W; Ravikumar KM; Chance MR; Yang S *Biophys. J* 2015, 108 (1), 107–15. [PubMed: 25564857]
- (115). Huang TY; Moser DK; Hsieh YS; Gau BS; Chiang FT; Hwang SL *journal of nursing research: JNR* 2013, 21 (1), 49–58. [PubMed: 23407337]
- (116). Aprahamian ML; Chea EE; Jones LM; Lindert S *Anal. Chem* 2018, 90 (12), 7721–7729. [PubMed: 29874044]
- (117). Schmidt C; Macpherson JA; Lau AM; Tan KW; Fraternali F; Politis A *Anal. Chem* 2017, 89 (3), 1459–1468. [PubMed: 28208298]
- (118). Gershenson A; Gierasch LM *Curr. Opin. Struct. Biol* 2011, 21 (1), 32–41. [PubMed: 21112769]
- (119). Rinas A; Mali VS; Espino JA; Jones LM *Anal. Chem* 2016, 88 (20), 10052–10058. [PubMed: 27681498]
- (120). Espino JA; Mali VS; Jones LM *Anal. Chem* 2015, 87 (15), 7971–8. [PubMed: 26146849]
- (121). Espino JA; Jones LM *Biophys. J* 2017, 112, 191a.

Figure 1.

Varying resolution of different biophysical techniques. NMR and cryo-EM are mainly applicable for macromolecules, and light microscopy and electron tomography are mainly applicable in tissues, cells, and organelles. Mass spectrometry has a dynamic range providing structural information from tissues to macromolecules.

Figure 2.

MS analysis of AmtB nanodisc. Panels A and B show the representative mass spectrum and deconvoluted spectrum. Four species are identified and highlighted, including AmtB with a large number of lipids (red), AmtB with ionic contact lipids (yellow), AmtB monomer (green), and MSP (blue). Panel C combines the mass spectra at different collision voltages, and panel D is a summation across all collision voltages. Reprinted from Marty, M. T.; Hoi, K. K.; Robinson, C. V. *Acc Chem Res* 2016, 49 (11), 2459–2467 (ref 12). Copyright 2016 American Chemical Society.

Figure 3.

Analysis of MDa virus-like particles on the modified mass spectrometer. During an MS1 scan, the charge-reduced Cp180 (a) shows a well-resolved charge state envelope centered on ~30 000 m/z . To further shift ions to higher m/z values and demonstrate the full capability of the instrument, the noncharge-reduced Cp180 (b) charge state envelope was subjected to increasing HCD collision energies. At a collision energy of 250 V, the MS2 products extend to 50 000 m/z (c). At maximal HCD, 300 V (d), the production mass spectrum shows further fragmentation of the Cp180 assembly and ions up 70 000 m/z . The mass resolution is high enough to baseline-resolve these different dissociation products even at this high m/z value (d, left inset). At the highest m/z , the mass resolution is greater than 500 (d, right inset). Reprinted from Fort, K. L.; van de Waterbeemd, M.; Boll, D.; Reinhardt-Szyba, M.; Belov,

M. E.; Sasaki, E.; Zschoche, R.; Hilvert, D.; Makarov, A. A.; Heck, A. J. R. *Analyst* 2018, 143 (1), 100–105 (ref 10), with permission of The Royal Society of Chemistry.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.

Illustration of the collision-induced unfolding analysis workflow for intact antibodies. (a) Selected antibody ions are unfolded through collisional heating, resulting in increased drift times; (b) drift-time data for a single protein charge state are tracked at each collision energy; (c) a collision-induced unfolding “fingerprint” is projected as a contour plot, where intensities for the features observed are denoted by a color-coded axis. Once compiled, fingerprint data are compared using custom software in order to detect differences. Reprinted from Tian, Y.; Han, L.; Buckner, A. C.; Ruotolo, B. T. *Analytical Chemistry* 2015, 87 (22), 11509–11515 (ref 25). Copyright 2015 American Chemical Society.

Figure 5.

Illustration of a proteoform: A term describing the complexity of a single-molecule protein species including genetic variation, alternative splicing of RNA, single-nucleotide polymorphisms (SNPs) in regions of genes coding for amino acids, and post-translational modifications that are explicitly defined in a specific combination.

Figure 6.

Reactivity of the DAU cross-linker with thiol groups, i.e., cysteine residues in proteins. As for other urea-based cross-linkers, two pairs of amine and isocyanate product ions are generated in (+)-ESI collisional activation (CID or HCD) experiments. Reprinted from Iacobucci, C.; Piotrowski, C.; Rehkamp, A.; Ihling, C. H.; Sinz, A. *J Am Soc Mass Spectrom* 2018 (ref 62).

Figure 7.

Cross-linking-derived model for respirasome supercomplex $CI_2CIII_2CIV_2$. (a) Cryo-EM-derived structure of respirasome $CI_2CIII_2CIV_2$ (PDB: 5GUP) with cross-links identifying interactions between CI (gold ribbon) and CIII (purple ribbon) (NDUA2 K13, K75, and K98 linked to QCR2 K250), CIII homodimer (QCR2 K159 linked to QCR2 K159), and CI and CIV (teal-blue ribbon) (COX5A K189 linked to NDUA9 K68) displayed. Cross-linked sites are shown as space-filled residues. Residues connected by red lines agree with the structure, while residues connected by a yellow line exceed the maximum cross-linkable distance (42 Å). (b) Structure of a circular representation of the respirasome $CI_2CIII_2CIV_2$, which agrees with all observed cross-linked sites. Reprinted from Chavez, J. D.; Lee, C. F.; Caudal, A.; Keller, A.; Tian, R.; Bruce, J. E. *Cell Syst* 2018, 6 (1), 136–141 e5 (ref 95). Copyright 2018 Cell.

Figure 8.

Overview of the photolytic covalent labeling workflow, for peptide labeling. (a) Proteolytic digestion of protein generates a population of unlabeled peptides. (b) Equilibration of peptides with substituted diazirines in aqueous solution. Diazirine labeling reagents used in this study include 3,3'-azibutan-1-ol, 3,3'-azibutyl-1 ammonium, and 4,4'-azipentan-1-oate. (c) Photolysis at $\lambda = 355$ nm to generate reactive carbenes that insert into chemically accessible regions of peptide. Photolysis is constrained to a sub-microliter volume in a windowed UV-transparent capillary that supports flash-freezing in liquid nitrogen. (d) Localization of carbene insertion sites within peptides using high-resolution MS/MS data analyzed in the Mass Spec Studio software package. Reprinted from Ziemianowicz, D. S.; Bomgarden, R.; Etienne, C.; Schriemer, D. C. *J Am Soc Mass Spectrom* **2017**, 28 (10), 2011–2021 (ref 106). Copyright 2018 Journal of The American Society for Mass Spectrometry.

Figure 9.

(a) Rosetta score versus RMSD to the native structure plots for 20 000 models generated using Rosetta ab initio for each of the four benchmark proteins. The top scoring model is represented as a star on each plot. (b) The top scoring models from the Rosetta score versus RMSD distributions in A (color) superimposed on the respective native model (gray). (c) Rosetta score + hrf_ms_labeling versus RMSD to the native structure plots for each of the four benchmark proteins after rescoring with the new score term. The top scoring model is represented as a star on each plot. (d) The top scoring models from the Rosetta score + hrf_ms_labeling rescoring distributions in C (color) superimposed on the respective native model (gray). Reprinted from Aprahamian, M. L.; Chea, E. E.; Jones, L. M.; Lindert, S. 2018, 90 (12), 7721–7729 (ref 116). Copyright 2018 American Chemical Society.