CAPTURING AND VISUALIZING PROTEIN COMPLEX FLUX IN CELLULAR REAL-TIME

Keiryn L. Bennett, PhD

Most, if not all, biological systems are regulated to some degree by protein-protein interactions and the formation of protein molecular machines to perform specific tasks within a cell. Such biological regulation is a fundamental requirement for maintaining normal cellular activity. When complexes do not arrange correctly, the outcome can be a poorly- or non-functioning cell that can ultimately lead to a disease. To aid in the understanding of the function of a specific protein complex, it is important to carefully characterize the proteins that are involved in creating both working or misfunctioning machines and how these proteins are spatially arranged with respect to each other. Traditional, established methods for studying protein complexes are severely limited by the quantity of pure protein required for the analyses (X-ray crystallography, cryo-electron microscopy), limited size range (nuclear magnetic resonance spectroscopy), poor specificity (ultracentrifugation), and low mass resolution (gel electrophoresis). There is therefore a need for an alternative approach that is both specific and sensitive for characterizing protein-protein interactions.

Over the last two decades, native noncovalent protein complexes have been successfully studied by mass spectrometry (MS) (*1-3*). Despite the achievements of native MS, interactions between some proteins in these active, functional complexes are not directly captured by MS. Additionally, some proteins are only weakly-associated with a complex and are readily-released and lost during sample preparation prior to MS analysis. To overcome such shortcomings and capture all the proteins within a complex, it is possible to intervene via the introduction of a chemically-reactive crosslinker that covalently links parts of the complex via the interacting proteins. This can assist researchers in the study of higher-order protein structure and has the distinct advantage that noncovalently-associated proteins are stabilized by the inclusion of a crosslinker. There are a range of chemical crosslinkers available that can be used to effectively 'fix' or 'freeze' protein complexes in stasis. Dissociation of these 'frozen' complexes by digestion with a protease releases a series of covalently-linked peptides. Subsequent analysis of crosslinked peptides by MS (XL-MS) provides insight into the spatial distribution and orientation of the individual components in the original complex. With this knowledge, it is feasible to build virtual images of intact protein complexes and determine the relationship the complexes have with each other within the cell (Figure 1). The data provided by XL-MS can be utilized to complement and refine existing structural information on a protein; and when combined with *de novo* molecular modeling, infer a structure for completely uncharacterized proteins.

"It is a powerful technique for determining distance constraints, aiding molecular modeling, and providing information on protein structure and complex organization."
- Albert Heck

According to Albert Heck, Scientific Director of the Netherlands Proteomics Centre at Utrecht University, "Although XL-MS has been around for 20 years, it has definitely benefited from the general evolution of mass spectrometry and computational proteomics and is now going through a renaissance." Furthermore, "It is a powerful technique for determining distance constraints, aiding molecular modeling, and providing information on protein structure and complex organization." Heck continues, "For structural biology, XL-MS has huge potential to aid in determining and refining protein structure. XL-MS now outperforms some areas of NMR and X-ray crystallography because of the caveat that these techniques often require recombinantly-expressed proteins. At the same time, however, XL-MS aids the revolution in electron microscopy and even more so, electron tomography."

Early Obstacles Thwarted Large-scale Proteome-wide XL-MS

The non-cleavable crosslinking reagents that were commonly-used in the early days of XL-MS had several downstream challenges. These difficulties primarily arose when analyzing the extremely diverse and complicated tandem mass (MS^2)

spectra that were generated from the crosslinked peptides. Crosslinking reagents can react with proteins in several ways to produce many different peptide-crosslinker products (Figure 1b). Together such multiple possibilities and combinations thwarted straightforward data analysis and interpretation; particularly in the background of the much more abundant non-crosslinked peptides. Searching the data generated from standard, collision-induced dissociation (CID) of crosslinked peptides via traditional database approaches had three major problems. Both the search space and the data analysis time were markedly increased; as was the false discovery rate (FDR) that occurs via mismatching of the data to the peptides/proteins. Although some interim solutions were developed to address these issues, challenges persisted. As such, XL-MS approaches were usually developed on a case-by-case basis for isolated purified protein complexes and were far from broadly-applicable.

Key Advancements in Large-scale Capture and Study of Protein Complexes

For a successful, integrated XL-MS approach that can be applied to a wide range of experiments, there are three key areas that must be addressed to empower researchers. Firstly, it is imperative that a robust, proteome-wide crosslinking strategy has been established via the development of new crosslinking reagents. Secondly, the necessary MS instrumentation and methods must be implemented to maximize the data quality. Finally, it is important that the data analysis and subsequent database searching can consolidate the crosslinked data generated by MS. Until recently, each of these steps was fraught with complications and such challenges made it difficult for XL-MS to be generically-amenable to a wide range of researchers. Advancements in MS-cleavable crosslinkers, MS fragmentation methods and data processing software have coalesced to significantly advance the field.

Ryan D. Bomgarden, Senior Staff Scientist at Thermo Fisher Scientific R&D, has been key in developing and supporting XL-MS workflows that are tailored to industry standards. "We want to learn from the expert laboratories to understand their latest and greatest methods," he explains, "From this, we can create and develop new reagents and standardize robust and reliable XL-MS workflows; not just for experts, but also for other non-expert customers." Bomgarden goes on to say, "Our aim is to open XL-MS for general use and broaden its application. We want to promote it to researchers within the broader structural biology community to show them that XL-MS is very complementary to their technology."

Heck nicely summarizes this: "XL-MS is now a straightforward method because the technology has developed to include MS-cleavable linkers and new software, thereby enabling whole cell crosslinking that can compete with, and complement, well-known and applied techniques such as BioID and classical interactome studies." Furthermore, Heck adds, "Nowadays, more and more scientists are becoming aware of XL-MS; especially structural biologists who have realized the benefits of the approach."

MS-cleavable Crosslinkers Simplify the MS^2 Data

One of the main advancements that has aided and simplified XL-MS in many research laboratories is the introduction of a new type of crosslinking reagents. MS-cleavable reagents are very similar in chemistry and reactivity to the traditional non-MS-cleavable products. The one major difference that has immensely simplified data interpretation is that the peptides are crosslinked with reagents that partially disintegrate in the gas phase during MS^2 (*4*). Two linear peptides are therefore produced from the same precursor ion, each containing part of the crosslinking reagent. This enables researchers to easily differentiate previously-crosslinked species from non-crosslinked species via specific diagnostic ions. The ions produced from the newly-generated linear peptides can then be individually isolated and further fragmented to determine the amino acid sequence and ultimately the identity of the crosslinked peptides. This MS method is referred to as an MS^3 experiment as it is basically fragmentation of a fragment ion. For downstream data interpretation, MS^3 has a major advantage. The generated spectra represent one of crosslinked peptides and are easily identified by standard peptide search.

Extracting Maximal Information from an XL-MS Experiment

When it comes to the way in which the mass spectrometer fragments the crosslinked peptides, Heck states that, "As crosslinked peptides are much more complex than linear peptides, more sophisticated MS methods are necessary." The concept of using multiple fragmentation pathways to generate different types of fragment ions increases the probability of correctly identifying both crosslinked peptides. As an example, fragmenting the same crosslinked peptide with CID followed immediately by electron transfer dissociation (ETD) and then combining the results from the complementary spectra can increase the total number of identified crosslinked peptides.

Along this theme, Frese and co-workers were the first to demonstrate that an MS fragmentation method termed electron-transfer/higher energy collision dissociation (EThcD) can be highly effective in fragmenting both unmodified, post-translationally-modified peptides and also crosslinked peptides (*5*). Combining the complementary ions that are generated from both the ETD and HCD processes into a single MS^2 spectrum markedly increased the accuracy of matching the data to peptides in the protein database (Figure 1d).

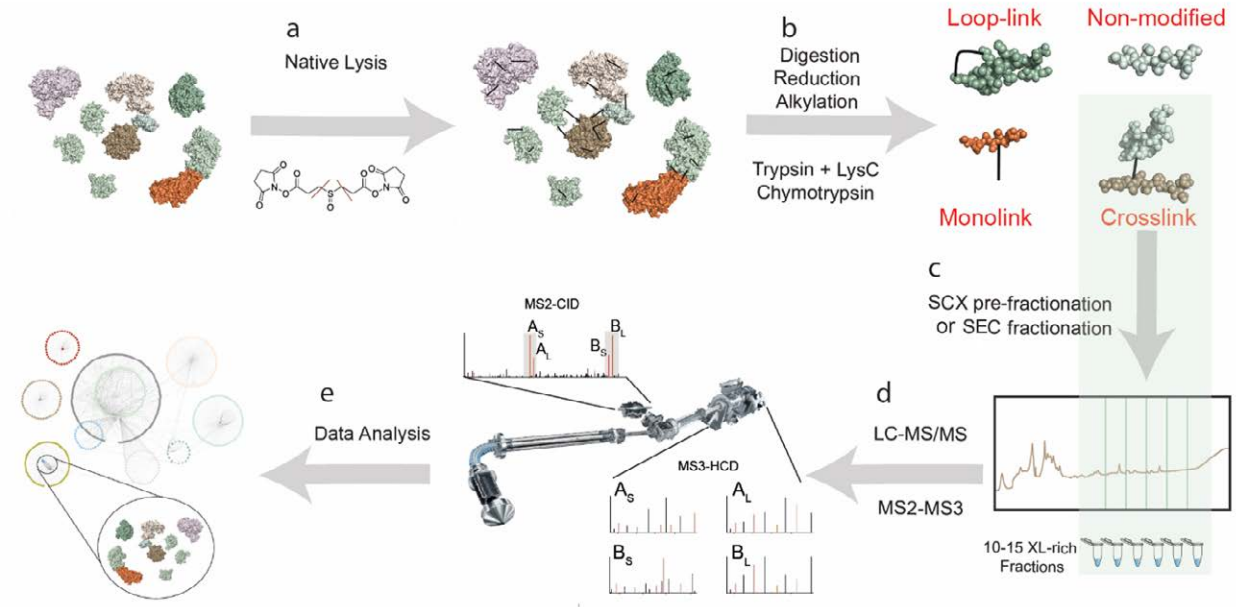

Figure 1. Generic workflow for XL-MS experiments. (a) Cells or tissue are lysed gently, leaving protein complexes intact. (b) After optimized incubation with the cross-linking reagent, and proteolytic digestion 4 peptide products can be formed. (c) Enrichment and pre-fractionation of XL-peptides using techniques like strong cation exchange (SCX) or size exclusion chromatography (SEC). (d) Advanced data acquisition techniques utilizing multiple steps of fragmentation techniques (CID, HCD) are used to identify the peptides. (e) The XlinkX node within Thermo Scientific™ Proteome Discoverer™ 2.2 software is used to identify the crosslinked peptides. The resulting data can consequently be integrated into structural modeling software (e.g. HADDOCK, I-TASSER, DisVis). Adapted from Klykov et al. Nature Protocols (2018) in press.

Resolving High Mismatch Rates and Slow Database Search Speed

The identification of crosslinked peptides by database searching has proven to be a major barrier in the adoption of XL-MS as a routine proteomic workflow. There have been many attempts to create specialized software and protein databases to ease the difficulties of analyzing the complicated data that is produced. According to Heck, "In the past 5-6 years, such software was developed in specific laboratories that was usually customized to the needs of the research group and the instruments in the laboratory." What was needed to advance XL-MS data analysis was a tool that is easy to use and generically-applicable to a range of experiments and workflows. The incorporation of the novel database search engine XlinkX (6) from the Heck laboratory into the Proteome Discoverer software platform offers that solution (Figure 1e). Heck points out that, "With the introduction of proteome-wide crosslinking using MS-cleavable crosslinkers and XlinkX, the previous high FDR problem has been largely resolved. A complicated crosslinking experiment is now transformed into a normal proteomic experiment and as such the conventional rules of FDR apply."

New Hurdles to Overcome

Comparing the *status quo* of XL-MS to standard proteomics, Heck says, "At the moment, XL-MS is perhaps 10-15 years behind standard proteome-based MS and there is indeed room for improvement. For example, as many as 200,000 crosslinks probably exist in an XL-MS sample but we are only identifying 2,000 or so. This is analogous to proteomics 20 years ago when only the most abundant proteins were identified. We are currently at the same stage now in XL-MS where we are predominantly finding interactions within the most stable and abundant protein complexes." He goes on to say, "This is not a negative point for the technology, but rather reflects the current state-of-the-art of XL-MS which will definitely improve with advancements in reagents, instrumentation, workflows and software."

When asked about the current shortcomings of XL-MS, Heck responds by saying, "The limitations are more concerned with time and experience. The solutions are there, but they may not yet be the best or easiest for truly generic applications." He continues, "I don't believe that MS sensitivity is the major road block in advancing XL-MS, rather the bottlenecks are at the front end of the workflow. That is, separation of crosslinked peptides from the abundant background of non-crosslinked peptides." Heck further states that, "It is currently possible to find crosslinked peptides by XL-MS, but extensive fractionation is required; which is labor intensive." He is nonetheless optimistic about the progress of XL-MS: "The future is clear because it will become easier to enrich crosslinked peptides." Affinity isolation of the

crosslinked peptides will markedly decrease the complexity of the mixture and enhance the signal for modified peptides over the non-crosslinked counterparts. These thoughts are echoed by Bomgarden, who also believes that, "New workflows are still needed to increase the number of identified crosslinked peptides, and enrichment of the crosslinked peptides is necessary for *in vivo* crosslinking experiments."

With respect to quantitative crosslinking, Heck believes, "Technically, quantitative crosslinking proteomics should not be problematic." Again, he is confident that, "It is just a matter of time, but this will be key in studying the dynamics of protein complex change under certain conditions." Indeed, one of the new workflows currently being developed by Bomgarden and his colleagues is, "To combine crosslinked peptides with the tandem mass tag (TMT) reagents to quantitate changes in protein complex dynamics." He says that, "This capability is integrated into Proteome Discoverer 2.3 software."

The Future of XL-MS is Promising

Heck and Bomgarden whole heartedly agree that the long-term perspectives of XL-MS are extremely exciting and lie in the, "Clever combination of XL-MS with electron microscopy and molecular modeling." The immediate future also looks bright because more and more researchers will use XL-MS in their daily research to map *in vivo* protein complexes at both the specific, targeted protein-of-interest level and across entire proteomes. "The field is rapidly growing," says Heck, "And chemists are becoming involved in the development of new reagents for biology." He continues, "With their help, novel photo-activatable crosslinkers or reagents that enable enrichment of crosslinked peptides will be developed." Further advances in MS instrumentation and software solutions will also be central to elevating XL-MS to the next level. Bomgarden predicts that, "XL-MS will eventually extend beyond just crosslinking a cell to trap endogenous proteins and capture static interactomes." He believes, "When XL-MS is combined with higher multiplexed tags, we will be able to follow specific proteins/complexes in real-time. This will aid studies on their spatial and temporal interactions during their journey through the cell."

References

1. A. J. Heck, *Nat. Methods* **5**, 927-933 (2008).
2. A. Sinz *et al.*, *Protein Sci.* **24**, 1193-1209 (2015).
3. I. Liko *et al.*, *Curr. Opin. Struct. Biol.* **40**, 136-144 (2016).
4. C. Frese *et al.*, *J. Proteome Res.* **12**, 1520-1525 (2013).
5. F. Liu *et al.*, *Nat. Methods* **12**, 1179-1184 (2015).