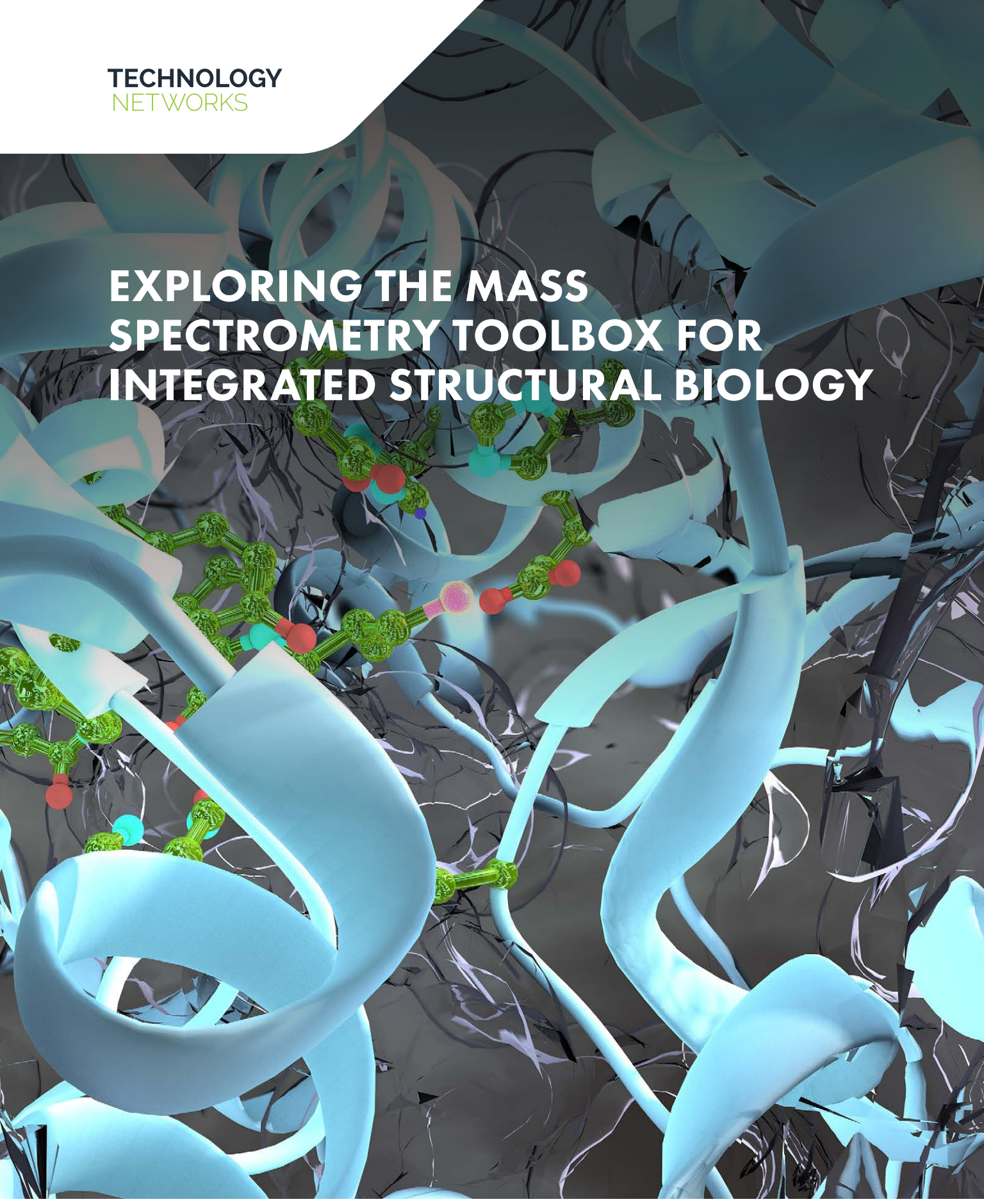
TECHNOLOGY
NETWORKS

EXPLORING THE MASS SPECTROMETRY TOOLBOX FOR INTEGRATED STRUCTURAL BIOLOGY



Brought to you by ThermoFisher SCIENTIFIC

CONTENTS

FOREWORD

Rosa Viner, PhD
Sr. Vertical Marketing Manager
Thermo Fisher Scientific

To deepen our knowledge of health and disease, it is important to appreciate the intricate structure and composition of proteins in cells and how proteins constantly interact to aid and complement one another in a global cellular context. To determine just how a minor shift in this subtle balance is sufficient to have unparalleled effects on the well-being of an individual is key to furthering human health and preventing disease.

Today, many scientists use mass spectrometry (MS) in a traditional manner. A basic peptide identification experiment or 'bottom-up' proteomics is broadly spread throughout the international community. Many scientists are content with this approach as answers are provided for the next step in their research or collaboration. Others may be hesitant to embrace new MS applications as the tried and tested approaches work, so why change a running system? Indeed, scientists may not even be aware that the field of MS is continually growing and evolving, and new instrumentation, reagents and software can revitalise classical techniques in the context of modern technology.

This has become particularly evident in the field of structural biology. With the advent of new reagents, methods and software, it is now possible to combine MS in an unprecedented way with traditional approaches such as protein crosslinking, affinity purification, limited proteolysis, and hydrogen-deuterium exchange. Newer, more sophisticated instrumentation can provide a more detailed analysis; but despite this, many laboratories are only fundamentally equipped with the necessary instrumentation required to advance into structural biological MS.

Modern-day MS not only complements existing approaches in structural biology; but more importantly, it offers a means to fine tune the data obtained with other more traditional methods. The level of molecular detail that can be provided by MS is unparalleled by any other singular technique or method currently available to structural biologists.

With this eBook, Thermo Fisher Scientific endeavours to introduce our existing and new customers to the innovative combination of MS with new reagents, robust workflows, and dedicated software. Thus, we can provide scientists with all the necessary tools to solve complex challenges in integrative structural biology.

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).

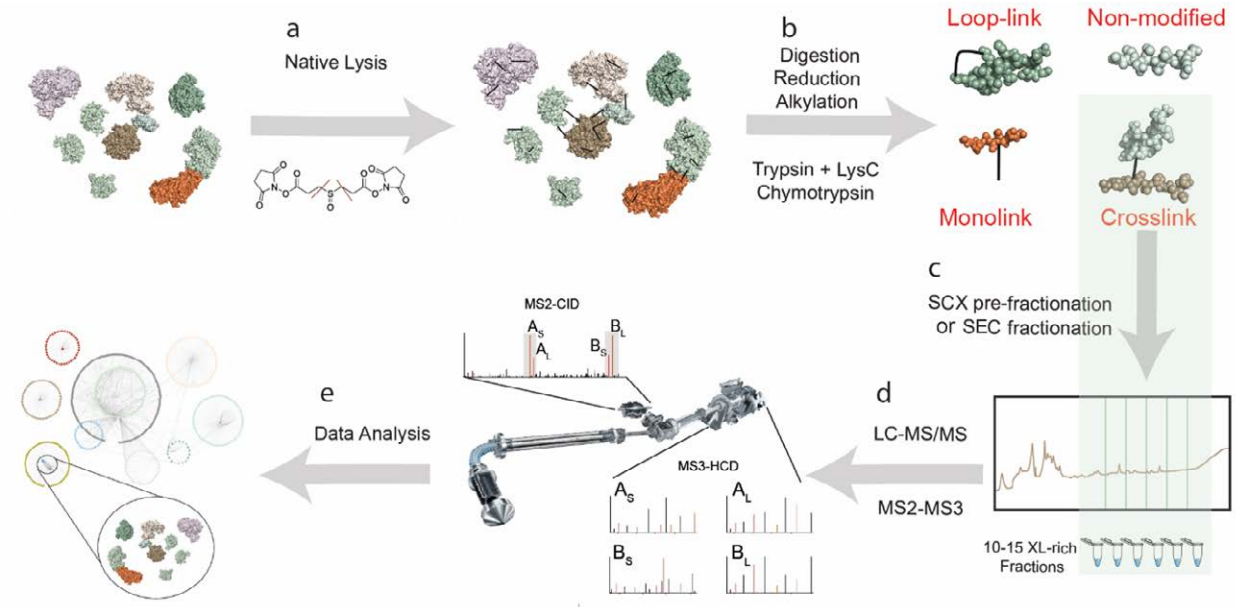


Figure 1. Generic workflow for XL-MS experiments. (a) Cells or tissue are lysed gently, leaving protein complexes intact. (b) After optimized incubation with the cross-linking reagent, and proteolytic digestion 4 peptide products can be formed. (c) Enrichment and pre-fractionation of XL-peptides using 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).

TOP-DOWN MASS SPECTROMETRY: THE PROTEOMICS BIRD'S-EYE VIEW

Andrew Jan, PhD

Twenty thousand is the approximate number of human protein-coding genes. Although the exact figure is often revised, everyone agrees that these proteins exist in a vast array of forms, fittingly termed proteoforms (*1*). Proteoforms are nature's way of introducing variation and "decorating" proteins. They arise from alternative splicing, single amino acid variations, or post-translational modifications (PTMs), and can lead to significant functional consequences. Just think of cellular signaling pathways modulated by kinases and phosphorylation, the dozens of modifications found on histone proteins, or the influence of glycosylation on monoclonal antibody stability and structure.

While sequencing DNA helped to describe and understand the genome, many key questions focusing on the role of proteins, which are the *de facto* genome's molecular executors in any living organism, remain unanswered. Many of those are related to protein structure and proteoform, for example; what are the specific modifications present on key molecules? Are those linked to sequence variation or structural changes? How is the cross-talk between modifications modulated?

There is a range of analytical techniques available to study protein structure and proteoforms, such as cryo-electron microscopy, nuclear magnetic resonance and mass spectrometry (MS). It is one MS method however, top-down proteomics, that is making big waves in the field.

Top-down proteomics using MS is now powerful enough to rival the well-established bottom-up proteomics approach as a robust and relatively high-throughput method to probe these and many other proteoform and structure-related questions. Thanks to new developments in MS instrumentation, chromatographic separation and software solutions, the approach can deliver vital information about intact proteins and their PTMs. According to Dr Rosa Viner, Sr. Vertical Marketing Manager, Life Sciences at Thermo Fisher Scientific, "The use of top-down MS by pharmaceutical developers is already a reality, and now it is set to become an integral part of the structural biology lab".

Understand Top-down (From the Bottom-up)

To understand the top-down concept, it is first worth looking at the workhorse of proteomics, the bottom-up approach. Proteins are polymers; they are simply polypeptide chains joined by covalent bonds. As such, bottom-up takes place at the peptide level. It begins with extracting proteins from cells

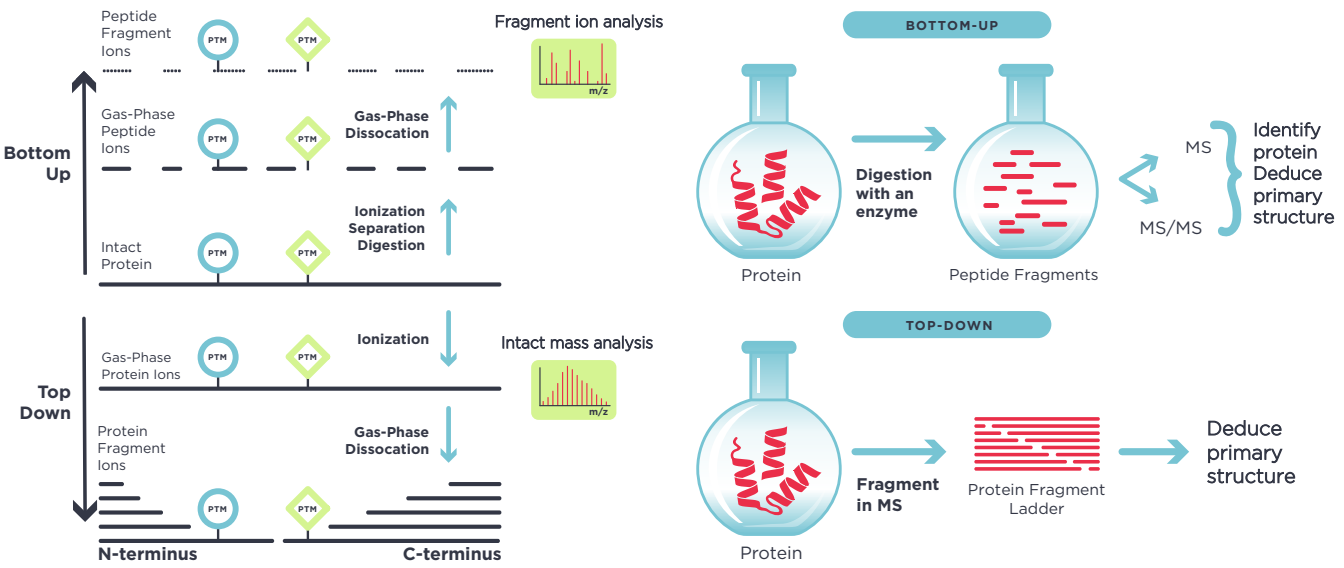


Figure 1. Bottom-up versus top-down approach in proteomics.

and chemically reducing them to break any unwanted covalent bonds. Then comes the tellingly named "shotgun" step where whole proteomes are fragmented into pieces using enzymes, such as trypsin or endoproteinase Lys-C, to digest proteins to peptides. The resulting mixture can be incredibly complex and often needs to be further separated using techniques such as reversed-phase chromatography or isoelectric focusing. To determine their sequence, peptides are then delivered to a mass spectrometer where they are fragmented, their tandem mass spectra obtained and later assembled *in silico*. This assembly process is not trivial and requires the researcher to compare the measured peptide's mass and fragmentation pattern to a database of protein sequences, deciding which one gives the best match.

In contrast, top-down practitioners dispense with digestion altogether and analyze entire proteins directly using MS (*2*, *3*). While bottom-up has been incredibly successful in measuring proteomes as diverse as those of viruses and humans, it has some drawbacks, especially when it comes to precisely characterizing proteoforms. With structural information being lost in the process, proteins have to be "inferred" based on often incomplete or imprecise peptide information. This makes it more difficult to quantify the abundance of positional isoforms or elucidate crucial information like the order of PTMs within a protein sequence.

"The big change in top-down proteoform characterization is the 'down' part.", says Neil Kelleher, the Glass Professor of Chemistry and Molecular Biosciences at Northwestern University and a leading figure in the field. "After [introducing intact proteins to the mass spectrometer and] 'weighing' each whole proteoform individually, top-down involves busting it apart using collisions with neutral gas, shooting it with electrons or photons of light", explains Kelleher. From then on, the experiment looks very similar to bottom-up; "The molecular weight of the resulting fragment ions (dozens or even hundreds of them) is measured and specialized software puts the proteoform back together with complete molecular specificity in most cases", adds Kelleher.

Most importantly for the researcher, the link between PTMs, sequence variations and individual proteoforms remains unbroken.

How Performance Powers Proteomics

Scientific progress often goes hand in hand with improvements in technology; top-down proteomics has a particular affinity for state of the art MS instrumentation. This demand in turn speeds up commercial development. Currently, a number of instruments suitable for top-down experiments are available, such as the high-end Orbitrap™ mass spectrometers from Thermo Fisher Scientific, points out Dr Luca Fornelli, a Postdoctoral Fellow in the Kelleher group at Northwestern University.

"Top-down was the only technology that could show the cross-talk between genetic mutations and chemical modifications located far from each other on the protein sequence."
-Luca Fornelli

Why is instrument performance so important in top-down? While avoiding digestion means the analyte mixture entering the mass spectrometer is less complex, it also means that the protein polypeptides display a broad charge-state envelope when ionized. This gives rise to extremely complex spectra when fragmented. In addition, when a single protein is present in multiple proteoforms there is even greater complexity, making isolation of specific isoforms a lot more difficult. High mass resolution and mass accuracy are therefore crucial in the separation and identification of overlapping spectral peaks that can have tiny mass differences, down to 1 Da, approximately a mass of one proton, for some single amino acid variants, or even fractions of that for particular PTM combinations.

"The Thermo Scientific™ Orbitrap Fusion™ MS, and more recently the Thermo Scientific™ Orbitrap Fusion™ Lumos™ MS is a platform that combines the advantages of a quadrupole filter for efficient isolation of precursor ions and fast electronics with diverse fragmentation capabilities", explains Fornelli. "The result is a very fast instrument that is highly versatile and can be efficiently used both in large-scale studies - where the focus is on the number of proteoform identifications - and in targeted experiments aimed at the characterization of sub-populations of genetically related proteoforms using multiple ion fragmentation modes".

Breaking It All Down

Indeed, the ability to break apart large biomolecules is another sought-after feature. Methods to achieve efficient and extensive protein sequence coverage have been developed and characterized by leading groups in the field (*4*, *5*).

Joshua Coon, at the University of Wisconsin-Madison, USA, and his group developed a new implementation of the fragmentation method called electron transfer dissociation (ETD) (*4*). In classical ETD, a radical reagent anion, such as fluoranthene, is used to transfer an electron onto a positively-charged protein causing it to fragment. This process is akin to heating the

protein up with electrons, in contrast to other popular methods which bombard it with molecules of gas. For the classical ETD method to work, the protein ions must be efficiently captured and accumulated somewhere in the mass spectrometer. With their new approach, termed high-capacity ETD, Coon and colleagues were able to accumulate 3-fold more protein and reagent ions than previously possible on a commercial platform.

Figure 2. Ion beam path inside an Orbitrap mass spectrometer

In another study, Albert Heck at the University of Utrecht, together with a team from Thermo Fisher Scientific, benchmarked multiple protein fragmentation modes on the Orbitrap™ Fusion™ MS (*5*). They discovered that using a combination of different dissociation methods worked very well in increasing the polypeptide fragmentation efficiency. Those hybrid methods; ETD followed by higher-energy collisional dissociation (EThcD) or collision-induced dissociation (ETciD), tended to produce complementary fragment ions. This leads to increased sequence coverage, allowing researchers to pinpoint PTM localization with far greater confidence.

Heck and colleagues took advantage of this increased coverage to examine the protein aurora borealis (Bora), a key activator required for the onset of mitosis. By setting up an assay to monitor Bora phosphorylation by various kinases and collecting EThcD and ETciD data, they were able to disentangle multiple proteoforms of Bora. This allowed them to unambiguously localize *in vitro* phosphorylation sites (some previously not reported in the literature) and point to what were the preferred phosphorylation motifs of individual kinases.

"[Unsurprisingly], these and other recent developments in the field prove to be highly beneficial for intact protein analysis and top-down proteomics", explains Viner, who points out that the most sought-after features are usually incorporated in commercial instruments making them available to the majority of users and not just the specialist labs.

But the mass spectrometer is not the only piece of the puzzle; advances in a number of other areas are also hugely beneficial, especially when it comes to large-scale studies. Bioinformatics and software tools are one such area (*3*), as they are required to analyze and make sense of the growing amounts of data. Another one is the progress in prefractionation of intact proteins; "Protein separation, including liquid chromatography and other protein fractionation technologies such as GELFrEE [gel-eluted liquid fraction entrapment electrophoresis], is also pivotal for the success of top-down proteomics experiments", says Fornelli. "Multidimensional separation of proteoforms is certainly becoming very popular in our field", he adds.

Straight to the Top (Down)

As mass spectrometry developed from its original role as an analytical chemistry technique to become a gold standard analytical tool in biological and biomedical research, the scope and potential of top-down proteomics expanded. To bring momentum to this expansion, the Consortium for Top-Down Proteomics (CTDP) was founded in 2012. "The consortium is a not-for-profit (.org) with a mission to accelerate the comprehensive analysis of intact proteins and their complexes",

explains Kelleher, who has been the driving force behind the initiative. With hundreds of citations since its inception, the consortium has already caught many people's attention and elevated proteomics as a discipline.

Another prime example of top-down proteomics delivering significant impact is the recent work from the Kelleher group in cancer biology (*6*). Published in PNAS, the study describes a workflow combining optimized immunoprecipitation followed by MS to characterize KRAS proteoforms in colorectal cancer cell lines and primary tumors.

The *KRAS* gene is responsible for sending switch on/off signals that control cell proliferation. However, when mutated, *KRAS* has the potential to promote oncogenesis as the normal signaling pathways are disrupted. The oncogene is often found to be mutated in colorectal, as well as lung and pancreatic cancers.

In the PNAS study, the researches applied their MS approach to directly measure KRAS4b proteoforms arising from genetic mutations. They showed a direct link between the mutations and presence of certain PTMs, such as nitrosylation or carboxymethylation. Furthermore, they were able to quantify the amount of mutant versus wild type KRAS4b protein in actual tumor specimens.

"This study shows the potential of top-down proteomics in the context of cancer biology", says Fornelli, a co-author on the paper. "Top-down was the only technology that could show the cross-talk between genetic mutations and chemical modifications located far from each other on the protein sequence."

With studies like this, it is easy to see how top-down proteomics can bring considerable rewards – and right now, academics and industry leaders alike are optimistic about the future. "Proteomics 2.0 is what I call the next-generation type of proteomics technology we use and envision", says Kelleher. "We believe the human proteome can be 'domesticated' and the understanding of our molecular selves in health and disease improved"

References

1. L. M. Smith, N. L. Kelleher, *Science*. **359**, 1106–1107 (2018).
2. T. K. Toby *et al.*, *Annu. Rev. Anal. Chem. (Palo Alto. Calif)*. **9**, 499–519 (2016).
3. B. Chen *et al.*, *Anal. Chem.* **90**, 110–127 (2018).
4. N. M. Riley *et al.*, *J. Am. Soc. Mass Spectrom.* **27**, 520–531 (2016).
5. A. M. Brunner *et al.*, *Anal. Chem.* **87**, 4152–4158 (2015).
6. I. Ntai *et al. Proc. Natl. Acad. Sci. U. S. A.* **115**, 4140–4145 (2018).

GOING NATIVE: USING MASS SPECTROMETRY THAT STARTS WITH PROTEINS IN THE MOST NATURAL STATE

Mike May, PhD

Until about 30 years ago, scientists could only apply mass spectrometry (MS) to smaller molecules. To look at larger molecules, the process needed a gentler form of ionization, which came with electrospray ionization (ESI). Now many scientists use ESI-MS, and this technology can be applied to native MS. According to Aneika Leney, a postdoctoral researcher at the University of Utrecht, The Netherlands, and Albert Heck, Utrecht's chair of biomolecular mass spectrometry and proteomics, native MS is "the mass spectrometric analysis of biomolecules that are, prior to their ionization, in their most native-like state." *(1)*

As one of the scientists who coined the phrase native MS,*(2)* Heck says, "Native MS is as close as it can get to having the proteins in a functional state that resembles the state in the cell."

In discussing the best uses of this method, he asks: "Why would you want to denature proteins if you can study them in a more native state?" That is, where native MS is an option, use it. "People didn't go native before because they thought it was difficult and less sensitive, but with new mass analyzers and sample prep, native MS can be as sensitive as denaturing MS."

Plus, the analyte gets fewer charges on it in preparation for native MS, and that divides the whole signal over just a few different charge states instead of, say, a dozen. As a result, native MS provides a better signal-to-noise ratio, and fewer overlapping ion signals.

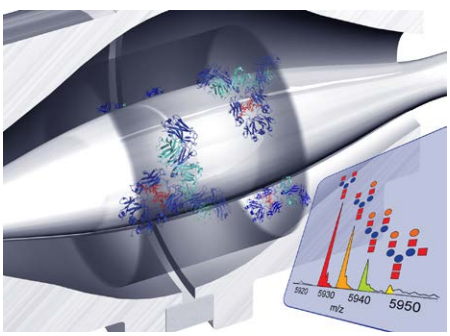


Figure 1. This schematic shows a glycoprotein (not to scale) in an Orbitrap mass spectrometer. (Image courtesy of Albert Heck.)

Native MS samples are analyzed in a gas phase, so the samples are only close to the native state prior to the ionization process. Still, this method can be used on a wide range of biomolecules, from nucleic acids and proteins to transport channels, as well as protein-drug interactions. Nonetheless, some scientists believe that the ionization turns a sample into a non-native form. That said, Leney and Heck noted that the "majority of the proteins studied by native MS to date have been shown to at least partly retain native-like structures in the gas phase," but they add that "much care still needs to be taken to avoid over-interpretation of native MS data since protein conformation can and will change upon transition from solution into the gas phase."

With careful techniques, though, native MS can reveal biological features that help explain basic biology and expand the potential applications.

"People didn't go native before because they thought it was difficult and less sensitive, but with new mass analyzers and sample prep, native MS can be as sensitive as denaturing MS."
-Albert Heck

Monitoring Modifications

The structure of a protein can vary extensively because of post-translational modifications (PTMs). One of the most common PTMs is glycosylation. Here, amino-acid side chains of a protein receive a carbohydrate, and that can arise in various forms. In N-glycosylation, for instance, the carbohydrate binds to a nitrogen atom, such as the amide nitrogen of an asparagine residue. In O- and C-glycosylation, a carbohydrate binds to amino acids via an oxygen or carbon atom, respectively.

Heck and his colleagues used native MS to study glycosylation in the human serum complement component C9.*(3)* Heck says that they examined this biomolecule because "it is part of the immune system's complement cascade that comes into action

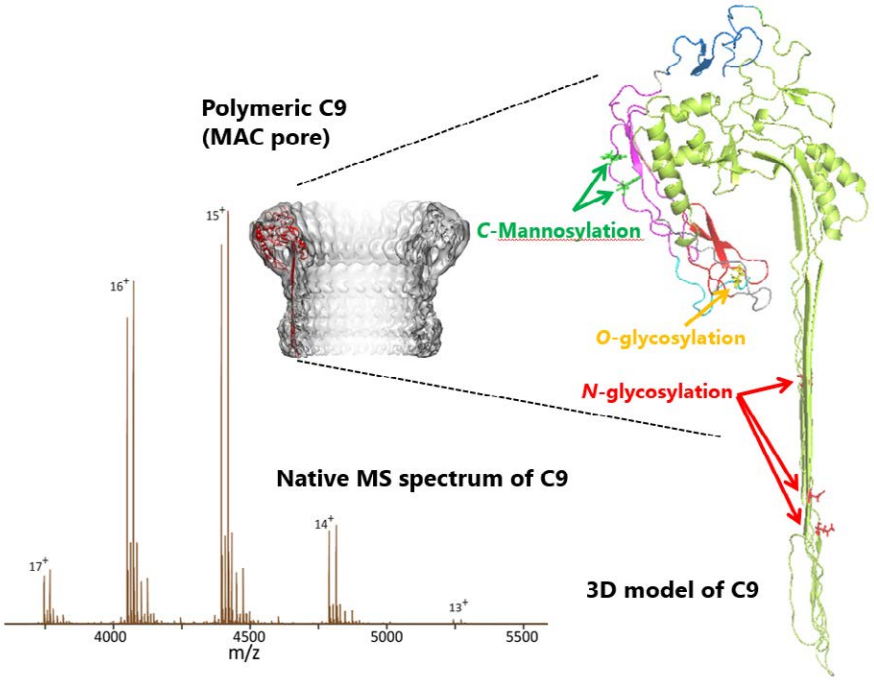


Figure 2. The membrane attack complex (MAC) kills pathogens by forming pores in their membranes (structure from Dudkina, N.V., et al. 2016. Structure of the poly-C9 component of the complement membrane attack complex. Nat. Commun. 7:1058). The major building block of the MAC complex is the protein C9, of which a single subunit is highlighted. Here, native MS analysis (Franc, V., Yang, Y., Heck, A.J.R. 2017. Proteoform profile mapping of the human serum complement component C9 revealing unexpected new features of N-, O-, and C-glycosylation. Anal. Chem. 89:3483–3491.) reveals N-, O- and C- glycosylation on the C9 protein, which possibly plays a role in the regulation of pore formation and pathogen recognition. (Image courtesy of Albert Heck.)

when we are attacked by pathogens—a bacteria or virus." In particular, C9 participates in forming the membrane attack complex, which creates pores in the membrane of foreign cells to kill them. As Heck says, "We're interested to know how we can boost the immune response when needed and reduce it when it's over active, and to do that we need as much structural detail as possible about the proteins, such as C9, in the cascade." C9 is modified by glycosylation in multiple ways. "It's a very complicated protein to analyze," Heck says.

Using the Thermo Scientific™ Exactive™ Plus Orbitrap instrument with extended mass range (EMR) that Heck and Alexander Makarov, Director of Global Research, Life Science Mass Spectrometry at Thermo Fisher Scientific, introduced in 2012,*(4)* Heck and his team uncovered various PTMs on the C9 protein extracted from human blood serum. Using native ESI-MS on C9, the scientists detected about 50 ion signals, each of which represents structurally and possibly functionally different C9's, being modified with different PTMs.

In the end, Heck and his colleagues reported three kinds of co-occurring C9 glycosylation—N, C and O—and validated 15 C9 proteoforms. They compared their results from native MS with tandem MS (MS^2) in a so-called peptide-centric approach, and the scientists wrote: "In total, we achieved more than 90% correlation between the native MS data and peptide-centric data."

But the reasons to use native MS go even deeper. With peptide-centric MS approaches, scientists often optimize it to look for either N-, O- or C-glycosylation, but when it's optimized for one that's the only one that it can detect. As Heck explains, "An optimized method for one is blind to the others, but with native MS you see them all in one go."

Despite the intriguing results from this work, Heck and his colleagues couldn't pinpoint a few of the glycosylation events. Plus, this work combined C9 samples from a pool of people. Now, Heck's team is working on glycosylation patterns from individuals. "It looks like these patterns can be different from person to person and even different in the same person, such as changing when you have an infection," Heck says.

Technically, the glycosylation pattern of C9 could be used to determine if a person has an infection. "It's feasible to take a sample and read it," Heck notes, "but the throughput is only about 10 people per day." As he adds, "So, to do it in a clinic, we have to make some steps."

Digging into Drug Discovery

In screening for new drugs, pharmaceutical scientists need efficient methods to test if a potential drug binds to a target, even if the binding is weak. Weak binding is especially prominent when testing fragments—smaller pieces of potentially complete chemicals that could turn into drugs. Keeping the potential protein target in its usual form is essential for understanding these interactions, and that creates another application for native MS.

To improve this application, Rebecca J. Burnley, a Senior Scientist at UCB Celltech in Slough, UK, and her colleagues tested an Exactive Plus EMR mass spectrometer as a platform to characterize non-covalent protein–small molecule interactions.*(5)* This method allowed the scientists to maintain the interactions between protein and ligands, revealing whether or not they bind to each other.

To use this technology in drug screening, though, it needs to fit the pharmaceutical workflow, and part of that involves sensitivity. As Burnley and her colleagues wrote: "The Exactive Plus EMR was … found to show an improvement in sensitivity over the [time-of-flight] instruments typically used for native MS analysis."

By combining that MS platform with an automated ESI system (TriVersa NanoMate), Burnley's team also showed that the process could be fast enough and even automated to analyze hundreds of protein-ligand complexes. Plus, they noted that "ligand-binding to more heterogeneous samples, such as glycosylated proteins, can be studied."

Monitoring Membranes

In biological systems, membrane proteins play many crucial roles, such as interacting with analytes in the extracellular region. "For membrane proteins, native MS is uniquely suited—one of the only techniques available—to examine individual ligand-binding events," says Joseph Gault, postdoctoral researcher in the lab of Carol Robinson, a Dr. Lee's Professor of Physical and Theoretical Chemistry at the UK's University of Oxford.

In comparing native and other forms of MS to study membrane proteins, Gault says, "You have to be more careful using native MS." Normally surrounded by lipids, membrane proteins require a protective vehicle, such as a detergent micelle, to help transfer the protein from solution to the gas phase, and then that vehicle must be removed before getting to the MS detector—all while keeping the protein as native-like as possible. Gault says, "Using an Orbitrap platform, we released proteins from the micelle without destroying any of the native interactions."*(6)* He adds, "Natural interactions are necessary for functions like drug binding."

"In the past we couldn't get resolution of very large protein complexes with few charges. We have benefited from recent developments in native high-resolution mass spectrometry. We didn't realize what we weren't seeing before. It gives us a new view of our molecules and this is an exciting transformation."
-Carol Robinson

Gault and his colleagues tested the system on a wide range of membrane proteins with diverse structures and masses from 26 to 186 kilodaltons. With high-resolution native MS, Gault and his colleagues found that even when a protein binds to two lipids in a membrane and the mass difference is small, just 55 daltons, the results clearly separated the peaks.

Now, using a Thermo Scientific™ Q Exactive™ Orbitrap UHMR (Ultra-High Mass Range) MS, Robinson and Gault can analyze up to megadalton complexes, and they are working on approaches where a synthetic protective vehicle is not necessary. "Now, we can look at complete complexes that came straight from the cell—that have never seen detergent," Gault explains. The larger complexes extend the range of potential samples, and not using detergent makes it easier to keep a sample as native as possible.

This small collection of examples gives only a glimpse into the possibilities of native MS. As more scientists realize that proteins can be explored with native MS at higher sensitivity than expected and the preparation gets increasingly easy to perform, this technology will be used across an even wider range of labs and applications. By starting with samples in the most natural form that is currently possible and then analyzing them very precisely, scientists will reveal structural information about proteins and how they interact with each other, as well as other structures.

References

1. A.C. Leney, A.J. Heck, *J. Am. Soc. Mass Spectrom.* **28**, 5–13 (2017)
2. R.H. van den Heuvel, A.J. Heck, *Curr. Opin. Chem. Biol.* **8**, 519–526 (2004).
3. V. Franc *et al.*, *Anal. Chem.* **89**, 3483–3491 (2017).
4. R. J. Rose *et al.*, *Nat. Methods.* **9**, 1084–1086 (2012).
5. H. J. Maple *et al.*, *Rapid Commun. Mass Spectrom.* **28**, 1561–1568 (2014).
6. J. Gault *et al.*, *Nat. Methods.* **13**, 333–336 (2016).

HYDROGEN/DEUTERIUM EXCHANGE MS: A POWERFUL, REVOLUTIONARY TOOL IN STRUCTURAL BIOLOGY

Dayne West, PhD

Current deficiencies in protein structural information reflects the complexity and difficulties associated with the construction of atomic models to accurately characterize these macromolecules. As such, it can be challenging for scientists to gain insight into their physiological functions. Moreover, the behavior of proteins and their conformations can vary greatly depending on experimental conditions, making it difficult to study them with a single technique. The conformational tendencies and behaviors of proteins often require great time and energy to characterize which leads many scientists to ponder the extent to which they are willing to further their study of protein structure and dynamics, regardless of its impact on their respective fields.

The primary techniques used to analyze protein structures include X-ray crystallography, nuclear magnetic resonance (NMR), or electron microscopy on samples at cryogenic temperatures (cryo-EM). These methods use biochemical means to visualize proteins at single protein, amino acid, and even atomic level resolution in some instances. These are valuable techniques, as the resolution achieved allows for essential protein information to be obtained. Unfortunately, these structures are static, meaning that information on protein dynamics or mobility can be difficult, if not impossible, to elucidate.(*1*) Coupled with this, difficulties in determining optimal conditions to grow crystals suitable for diffraction, as well as the time-consuming nature of both data analysis and protein purification, make X-ray crystallography unsuitable for routine analysis. Furthermore, size limitations dictate that some proteins are too large and/or complex for NMR, especially when trying to determine a native, intact protein structure.(*2*)

While the techniques above are pillars in the field of structural biology, they have one significant flaw: they provide minimal information on protein dynamics. The structures divulged from techniques such as NMR and crystallography are static, snapshots of a position or conformation exhibited by a macromolecule at the time of crystallization and/or data collection. To fully unravel protein function, scientists must not only explore structure, but movement as well. Mass spectrometry (MS) is a powerful technique by which protein dynamics information can be elucidated. By integrating MS results with other protein chemical techniques, scientists can determine both protein structure and function. This combination of data and models from different experiments can reveal how a protein or protein complex works, as well as providing the structure and composition of these macromolecules.(*3*) By combining MS with the right technique, scientists can take advantage of a method's strengths and minimize its weaknesses, enabling them to illuminate the basic biology of proteins and how they can be used.

Despite advances in using and combining these techniques, scientists keep searching for improvements. According to Terry Zhang, an Application Specialist at Thermo Fisher Scientific, the goal of combining well-studied biochemical techniques in new and creative ways to develop a novel form of macromolecular structural analysis is, "growing in structural biology."

Zhang continues, "The combination of hydrogen-deuterium exchange with MS (HDX-MS) has been emerging as a strong tool for structural characterization of macromolecules, particularly proteins. The technique is placed in the middle range of resolution, and has no limitation on protein size, a key obstacle for many structural biology tools." Zhang describes HDX-MS as, "an efficient method for the study of protein dynamics, protein aggregation and degradation, as well as drug discovery and characterization of biosimiliars." With these many advantages, HDX-MS is emerging as a strong tool for structural characterization of macromolecules, particularly proteins.

Hydrogen/Deuterium Exchange – Mass Spectrometry (HDX-MS)

As the name implies, (HDX) is a reversible chemical reaction in which a covalently bonded hydrogen atom is replaced by a deuterium atom. This is a well-known biochemical technique, particularly common in protein analysis and characterization. In HDX, the amide hydrogens on the protein's surface are the first to exchange with deuterium, and the less accessible hydrogens, those buried somewhere inside the protein structure, will be exchanged more slowly. This can tell you which parts of the protein are outside, and which are inside. "HDX is a

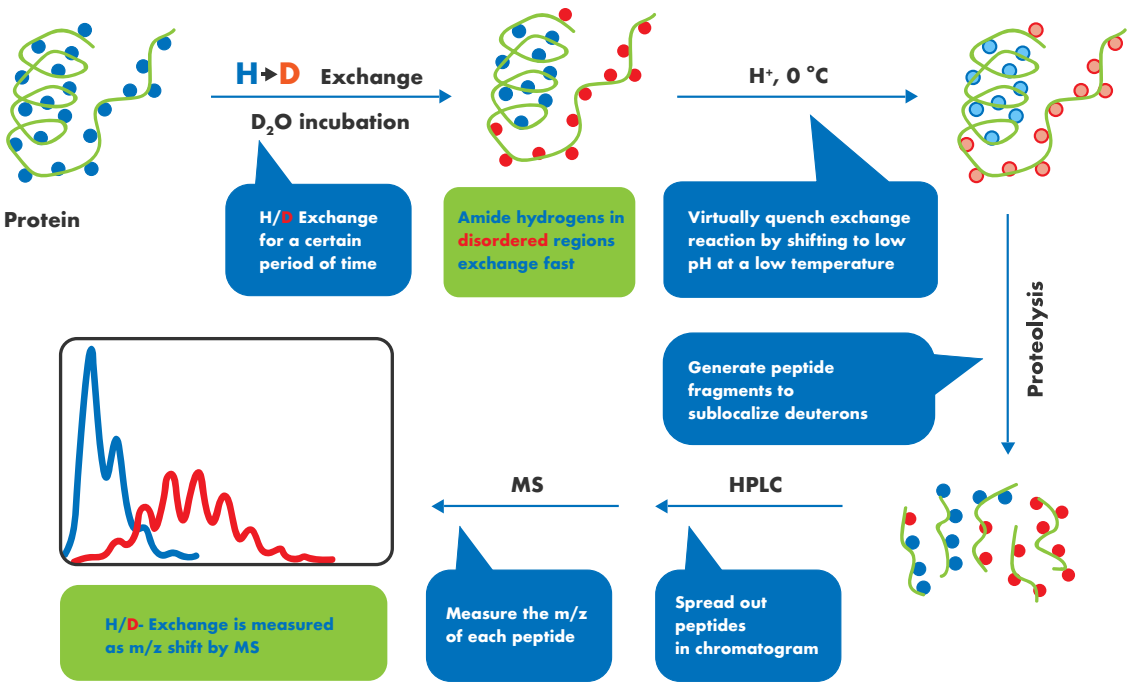


Figure 1. H/D-Exchange-MS Analysis of Proteins

common tool used to study conformations of antibodies," says Zhang, as well as the catalytic mechanism of powerful enzymes, such as carbonic anhydrases.*(4)* "The use of HDX-MS as a tool for protein characterization and structural analysis has been around for decades," says Dr Christoph Borchers, a professor at the University of Victoria, British Columbia, Canada with over 10 years of experience applying HDX-MS. However, according to Borchers, "it is only recently that advances in this technology, through improvements in MS and data processing, have made it powerful enough to study large protein complexes and even whole viruses.*(1)*"

There are two popular approaches to HDX-MS: "bottom-up" and "top-down." "Bottom-up" is a more traditional MS approach, involving digestion of proteins, followed by analysis using liquid chromatography-mass spectrometry (LC-MS). In contrast, the "top-down" approach utilizes a gas phase mechanism, in which fragmentation of the protein occurs through the introduction of electrons, either by electron transfer dissociation (ETD) or electron capture dissociation (ECD).*(2)* The "top-down" approach was not always the most practical option, as capturing of electrons proved rather costly, and was not compatible with all MS instrumentation. However, the development of ETD, a dissociation method similar to ECD that can be used on widely available commercial MS equipment, made this approach favorable. "The 'top-down' approach is preferable, as it requires less time and less manipulation of the sample" says Borchers.*(2)* Indeed, researchers using "top-down" do not have to worry about digestion and fragmentation of the protein beforehand, which can be time consuming for larger proteins. Borchers goes on to say, "Deuterium incorporation can be determined by MS at the amino acid level, which can allow a researcher to obtain a deep understanding of the structural features and dynamics associated with a protein at the monomeric, peptide level." Furthermore, fragments produced by the top down approach can be directly correlated to the intact protein, and the level of H/D back exchange is greatly reduced.*(2)* The exchange back of hydrogen ions for deuterium ions is "a major disadvantage" according to both Borchers and Zhang, as it defeats the purpose of incorporating deuterium if it readily exchanges with hydrogen. "We have overcome this challenge by performing the LC at subzero temperatures (-20 °C), resulting in a back exchange of just 2% in 10 minutes compared to up to around 30% at 4 °C," says Borchers regarding the improvements made in collecting reliable HDX "top down" MS data.

HDX-MS as an applied structural dynamics tool

According to Zhang, "One of the biggest upsides of HDX-MS is its ability to obtain data on conformational changes within proteins and protein complexes." There have been many developments in this area, focused on creating methodologies and workflows to bridge the analytical gap left by cryo-EM, x-ray crystallography and NMR.*(5)*

One area of research where HDX-MS has been heavily implemented is the study of antibodies. These proteins have been thoroughly used for the detection and treatment of various diseases, so characterization of chemical degradation is a primary concern. For example, critical conformational changes may impact the efficacy of antibody-based drugs developed by the biopharmaceutical industry.(*6*) In regard to antibody research, HDX-MS can operate as an analytical tool that can detect these minor conformational changes.(*7*) Not only can MS spectra be obtained for all created protein fragments, but according to Borchers, "Time points of deuteration can be achieved, allowing the determination of when the deuterium incorporation occurs, which can provide structural insight into the location and conformation of specific protein regions at the amino acid level."

"The use of HDX-MS has drawn interest from the pharmaceutical industry," says Borchers. "There has been a big push for biosimilars, pharmaceutical products that strongly mimic original biomolecules, and perform their same function." An example of this would be recombinant vs. native proteins. Whilst not identical, both proteins perform the same function, and recombinant proteins are often easier to obtain and produce in powerful host systems, such as *Escherichia coli* and yeast, compared to the intricate mammalian cell expression system. It is beneficial for pharmaceutical companies to begin utilizing the top-down approach, and develop a method to study subtle differences quickly to ensure these biosimilar products are structurally equipped and conformationally accurate. As previously mentioned, many structural methods are too time-consuming or limited in this regard.(*2*) However, the advantages of top-down HDX-MS, such as minimal sample manipulation, fast workflow, low back exchange, and simple data analysis, make it a simple but powerful method for comparative structural evaluation of not just intact antibodies, but other proteins as well.(*2*)

HDX-MS as the future of structural biology

Borchers believes that HDX-MS "has a very bright and prosperous future in structural biology and protein science." He goes on to say, "the addition of HDX-MS to the world of structural biology will add further validation to current methods. Not only can it identify the conformational changes associated with binding and/or polymerization, but the transition from ordered to disordered states." This is a huge plus in the field of neurodegenerative diseases, such as Alzheimer's disease, where protein aggregation is heavily associated.(*8*)

Commenting on the future of HDX-MS, Borchers says, "The market size for biosimilars is 20 to 50 billion dollars." Therefore, the use of HDX-MS will be greatly needed and utilized in the development of future therapeutics. The significant amount of financial support for an endeavor such as biosimilar product development, the ease and sensitivity at which we can study conformational changes, and the addition of an analytical tool to the field of structural biology are all indicative of the possible success and bright future of HDX-MS as a strong tool for protein characterization and analysis.

References

1. M. T. Mazhab-Jafari, J. L. Rubinstein, *Sci. Adv.* **2**, e1600725 (2016).
2. J. Pan *et al.*, *Anal Chem.* **87**, 5884-5890 (2017).
3. P.S. Chetty *et al.*, *PNAS* **114**, 968-973 (2017).
4. D. West *et al.*, *Biochemistry* **51**(**45**), 9156-9163 (2012).
5. T. Zhang *et al.*, *Thermo Scientific*, Poster Note 64466 (2015).
6. J. Pan *et al.*, *Biochimica et Biophysica Acta* **1864**, 1801-1808 (2016).
7. T. Zhang *et al.*, *Thermo Scientific*, Poster Note 64465 (2016).
8. D.R. Thal, M. Fandrich, *Acta Neuropathologica* **129**, 163-165 (2015).

AFFINITY PURIFICATION MASS SPECTROMETRY: PROVIDING STRUCTURAL INSIGHTS INTO THE SOCIAL AND ANTI-SOCIAL BEHAVIOR OF PROTEINS

Keiryn L. Bennett, PhD

Affinity purification is a biochemical approach that was first described by Seraphin *et al. (1)* in 1999 (Figure 1a). Over the last two decades the methodology has been developed further and modified to encompass a diverse range of protein tags. When coupled to mass spectrometry, affinity purification (or AP-MS) enables scientists to extract intact, non-denatured protein complexes from cells and identify the constitutive protein components of these functional complexes. AP-MS has been successfully used to answer specific biological questions on how protein complexes are brought together, how such complexes interact with other protein complexes, and how these complexes may change with perturbations. Disruptions to the normal function of a protein complex may occur as a result of external stimuli, such as a drug or virus. And of course, in the context of disease, AP-MS can provide answers on how protein-protein interactions are altered in response to a specific genetic mutation and what effect this has on the role of the protein. Needless to say, this technology is extremely important in aiding our understanding of not only the stable, undisturbed, healthy human proteome; but how the cellular protein architecture is influenced and adjusted by a specific disease-induced effect. The subsequent architecture of all these proteins within the human proteome can aid us in better understanding the development of disease and ultimately how diseases progress.

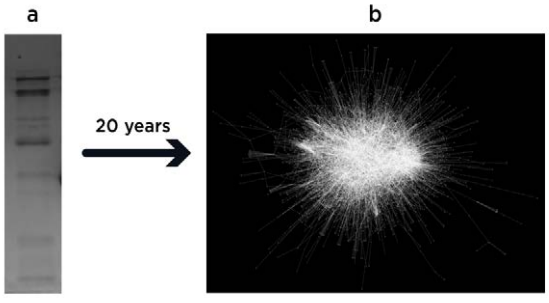


Figure 1. The progress of affinity purification mass spectrometry (AP-MS). In twenty years, AP-MS has evolved from the mapping and discovery of new, individual, functional protein complexes (a) to the mapping and study of entire protein interactomes (b).

A Task of Epic Proportions

A 2017 study led by Ed Huttlin, Steve Gygi and Wade Harper *(2)* from the Harvard Medical School showed how large-scale protein-protein interaction networks generated by AP-MS can be used to study the architecture of the entire human interactome (Figure 1b). The human genome is composed of approximately 20,000 individual protein-coding genes, many of which exist as multiple, alternatively-spliced forms and allelic variants. To create a comprehensive model of protein architecture that reveals how these individual protein assemblies can congregate into functional modules and networks is no small feat. The team achieved this task and compiled the information into a large data repository, termed BioPlex 2.0.,(3) thereby creating a resource that is vitally important for scientists and the general community alike. According to Gygi," Knowing the interactors of a given protein provides [spatial] context for the protein with respect to sub-cellular location within the cell. Further insight is therefore obtained for larger complexes and ultimately complete pathways."

"This approach can provide the missing pieces and alert the scientist to other events that they may not be aware of"
-Steve Gygi

Commenting on AP-MS for structural biologists, Gygi goes on to say that "AP-MS provides an approach where evidence can be obtained for the observation of a direct physical contact [between proteins] including secondary and tertiary binders." In terms of research in structural biology, "This approach can provide the missing pieces and alert the scientist to other events that they may not be aware of, for example, that the phosphorylation of a protein will ultimately affect the structure of the protein under investigation". Similarly, Huttlin says that "New hypotheses can be generated based on associations

amongst [protein] structural domains; moreover, patterns of connectivity can provide structural insight into large protein complexes, like the proteasome, which readily subdivides into its catalytic and regulatory components."

For the non-scientific community, research such as this provides new developments in understanding how a disease initially manifests and ultimately progresses. Huttlin compares the workings of a cell in the context of network biology with the social media interactions of an individual person by saying "With this approach we have essentially created a social network of the cell." A correctly functioning cell is comprised of socially-interacting proteins; whereas in a disease state, several proteins begin to exhibit anti-social behavior.

A Matter of Scale

With a task of such massive proportions, the first goal was to create a reference interactome that placed specific proteins into distinct molecular assemblies. Until now, most genome-wide experimental studies on human protein interactions have relied upon yeast two-hybrid technology or correlation profiling techniques; whilst prior AP-MS-based studies have targeted much narrower areas of the human proteome. Individually and combined, these earlier investigations have mapped only a proportion of the human interactome. Huttlin, Gygi and Harper have taken a tried and tested AP-MS approach to profile protein interactions in the context of a human cell with unprecedented depth and breadth. By targeting an unparalleled number of human proteins for AP-MS analysis using state-of-the-art LC-MS technology, they have been able to map a much larger cross-section of the interactome than had been collectively attempted by other groups and approaches. With this methodology, multiple protein communities have been identified that subsequently enabled them to define and discern several disease networks. To put serious numbers behind the data, from a total of 7,500 AP-MS experiments, Gygi, Huttlin and their fellow colleagues have discovered more than 56,000 interactions that contain more than 29,000 associations that were previously unknown. This information not only provides functional insight into hundreds of poorly-characterized proteins; but also enables prediction of the cellular localization of the proteins.

As with every advancement in scientific research, the approach adopted by Gygi, Huttlin and Harper is not without complications. As Gygi points out, "Dealing with highly-variable levels of bait expression from cell line to cell line can be challenging, as can working with membrane proteins." Similarly, Huttlin states that, "We have also routinely encountered other more technical challenges, including variability of affinity purification over time, intermittent problems with liquid chromatography and mass spectrometry, and so on." Therefore, it was imperative to the success of the project that "A rigid quality control process for both the culturing of the cells and for maintaining and operating the instrumentation was in place."

What Can Be Learnt from the Human Interactome?

This immense network of human protein-protein interactome data will enable other researchers to study protein interactions at a systems level. The data in BioPlex 2.0 *(3)* can lead to the generation of new hypotheses and the discovery of previously-unknown functions of proteins. Indeed, data mining of this network has already led scientists to discover new modalities of under-studied proteins. As pointed out by Huttlin, "Bioplex 2.0 provides a solid foundation to use and integrate other -omic data." In this way, scientists can glean deeper insights into their research by, "Incorporating additional data sets followed by data mining and correlation of the obtained protein networks". "Our future plans" says Gygi, "Are to begin to expand the data set by mapping the human interactome in additional disease-relevant cell lines. In this way, we can start to understand how the dynamic human interactome can alter and adapt as a consequence of specific disease phenotypes."

Bioplex 2.0 is freely-available to the community (http://bioplex.hms.harvard.edu).

References

1. G. Rigaut *et al.*, Nat. Biotechnol. 17, 1030–1032 (1999).
2. E. L. Huttlin *et al.*, Nature. 545, 505-509 (2017).
3. E. L. Huttlin *et al.*, Cell. 162, 425-440 (2015).

TRANSFORMING STRUCTURAL PROTEOMICS THROUGH LIMITED PROTEOLYSIS

Joanna Owens, PhD

Despite huge technological advances, one of the challenges facing the field of proteomics is that current technologies are unable to obtain a snapshot of all the proteins in a cell in their native conditions. But a new approach that combines a classic biochemical method, limited proteolysis (LiP) with modern-day mass spectrometry (MS) analysis, looks set to transform the field.

Limited proteolysis mass spectrometry (LiP-MS) works by exploiting differences in a protein's vulnerability to proteolysis when in certain conformations. A unique double-digestion step cleaves only the regions of the protein that are exposed under native conditions. When the resulting peptides are analyzed by MS, it reveals the structure of all protein molecules in the cell at a given moment in time.

Overcoming Challenges with Conventional Proteomics Approaches

The main hurdles to analyzing protein structures in their biological environment are the complexity of biological samples, and that the cellular milieu has little resemblance to the simplified experimental setups typically used for *in vitro* structural biology studies. This means it is difficult to detect dynamic yet often subtle changes to protein structures, which occur on very short timescales and cover a broad range of length scales.

To overcome this problem, Professor Paola Picotti from ETH Zurich decided they needed a new approach. "When we developed LiP-MS, there were basically no other technologies to probe the structural features and conformational changes of proteins in a complex biological matrix and on a proteome-wide scale," she explained. "Other *in situ* or *ex vivo* structural approaches had specific limitations."

These other approaches include nuclear magnetic resonance (NMR) and fluorescent resonance energy transfer (FRET). Both require labeling the protein of interest, which is likely to alter the function of the protein. Another method, cross-linking MS, can derive structural information for many proteins simultaneously, but it struggles with the complexity of cell and tissue extracts. And although cryo-electron tomography can tackle a variety of protein complexes and assemblies *in situ*, it has far from proteome-wide coverage.

By contrast, the unique power of Picotti's method is that it enables probing of protein structural changes directly from very complex cell and tissue extracts on a proteome-wide scale: "It allows you to pinpoint regions involved in a structural change," she explains. "This enables the identification of proteins that undergo a structural change, for example, as a consequence of specific cellular perturbations or development of disease."

"When we developed LiP-MS, there were basically no other technologies to probe the structural features and conformational changes of proteins in a complex biological matrix and on a proteome-wide scale,"
- Paola Picotti

The LiP-MS Approach

Limited proteolysis itself is not a new technique. It has been applied in a variety of studies using purified proteins or simple reconstituted protein systems since the 1960s, says Picotti. "The novelty of this approach is that it enables 'LiPping' a whole proteome and directly from a complex biological extract and analyzing the resulting complex proteolytic mixtures."

The double-digestion step and the switch from native to denaturing conditions is key to the approach, says Picotti. In the first step, cell lysate is treated with a broad specificity protease, such as proteinase K, under native conditions at a low enzyme-to-substrate ratio, generating large protein fragments. The lysates are then switched to denaturing conditions and a complete digestion by trypsin generates peptides suitable for MS analysis. A control sample is treated with trypsin only. Initial MS analysis reveals fully trypsin-digested peptides with much lower abundance in the double-digested samples compared to the trypsin-only samples. These will contain LiP cleavage sites and are therefore further analyzed by selected reaction

monitoring (SRM) MS. This allows the identification of peptides specific to a given protein conformation.

In this way, the method can be used as a screening tool to detect both subtle and significant changes in protein structure during disease development or to detect when the protein of interest is bound to a metabolite or small molecule drug.

Advantages over Conventional Methods

Dr Rosa Viner, Sr. Vertical Marketing Manager, Life Sciences at Thermo Fisher Scientific, thinks the LiP-MS method is an elegant approach: "I like that it uses native cells, so it captures real cell events and it's universal."

Another key advantage, she says, is the lack of protein labelling required. "Other techniques for studying the interactome - for example, bio-ID – need to introduce biotin, and biotin is a large molecule. Any molecule introduced into the protein is going to change its conformation and have some effect on function. With the LiP-MS approach you can really start to look at protein-protein interactions, protein-small molecule interactions and different post-translational modifications."

"We suddenly and quite unexpectedly became able to probe a variety of structural changes of high biological and biomedical relevance"
-Paola Picotti

"Such large-scale analysis will require continued development of the MS capabilities to match," she points out. "We are always pushing the boundaries, we want to be faster, more sensitive, more user-friendly, more robust and more stable."

Transforming Proteomics in Disease and Drug Discovery Research

The development of the LiP-MS approach has had a major impact on Professor Picotti's research and future directions. "We suddenly and quite unexpectedly became able to probe a variety of structural changes of high biological and biomedical relevance, such as those associated with allostery, drug binding, pathological protein aggregation, protein denaturation or receptor activation."

They demonstrated this with three landmark studies. In the first *(1)* they validated the new approach by showing that it could detect a conformational change in the α-Syn protein, which is unfolded under normal physiological conditions but switches to a β-sheet conformation and aggregates into fibrils in Parkinson's disease. They also demonstrated detection of a more subtle structural transition – the switch from the holo- to apo- form of myoglobin. This was the first demonstration of the feasibility of analyzing protein structural changes on a global proteomic scale against a complex biological background.

In the second study. *(2)* they used the approach to probe for the determinants of thermal sensitivity on a proteome-wide scale for proteomes from four different organisms. The results showed that temperature-induced cellular collapse is due to the loss of a subset of proteins with key functions and shed light on the evolutionary conservation of protein thermostability.

In their most recent paper, having discovered that LiP-MS enables the discovery of protein-small molecule interactions, Picotti's team exploited this principle to generate the largest map to date of cellular protein-small molecule interactions. *(3)* The map revealed functional and structural principles of chemical communication and shed light on the prevalence and mechanisms of enzyme promiscuity.

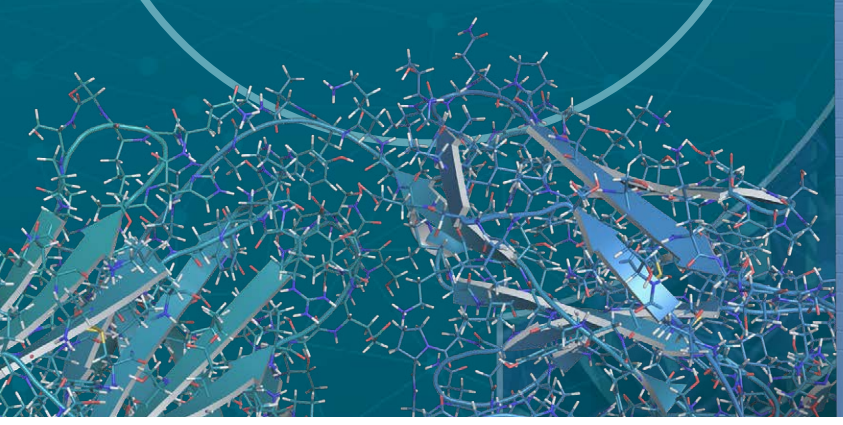
Picotti believes that LiP-MS has significant potential beyond these studies: "The capability of LiP to identify protein-small molecule interactions has important implications for drug target deconvolution, as it translates into the ability to identify targets or off-target effects of drugs directly from complex biological extracts," Picotti says. "And the detection of structurally altered proteins upon disease development presents the intriguing possibility that protein structures could serve as disease biomarkers and be directly monitored in patient cohorts. LiP-MS has opened up a range of exciting possibilities, and we are only just beginning to explore them."

References

1. Y. Feng *et al.*, Nat. Biotechnol. 32, 1036-44 (2014).
2. P. Leuenberger *et al.*, Science. 355, eaai7825 (2017).
3. I. Piazza *et al.*, Cell. 172, 358-372 (2018).

thermoscientific

Go beyond with native mass spectrometry



Improve the pursuit for disease understanding and biotherapeutic effectiveness

Detailed structural insights from the new Thermo Scientific™ Q Exactive™ UHMR Hybrid Quadrupole-Orbitrap™ Mass Spectrometer enhance our understanding of protein function, disease mechanisms, potential drug targets, and biotherapeutics. Sophisticated advancements in sensitivity and mass resolution at high *m/z* help resolve small mass differences to characterize intact biomolecular assemblies and complexes, and reveal previously unknown substoichiometric details.

Gain deeper insight **thermofisher.com/qexactiveuhmr**

ThermoFisher SCIENTIFIC