Integration of mass spectrometry data for structural biology

Hannah M. Britt[1], Tristan Cragnolini[1,2] and Konstantinos Thalassinos[1,2*]

1. Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London, WC1E 6BT, United Kingdom
2. Institute of Structural and Molecular Biology, Birkbeck College, University of London, London, WC1E 7HX, United Kingdom

*Corresponding author: k.thalassinos@ucl.ac.uk

Abstract

Mass spectrometry (MS) is increasingly being used to probe the structure and dynamics of proteins and the complexes they form with other macromolecules. There are now several specialised MS methods each with unique sample preparation, data acquisition and data processing protocols. Collectively these methods are referred to as structural MS and include, crosslinking-, hydrogen deuterium exchange-, hydroxyl radical footprinting- native, ion mobility- and top-down MS. Each of these provides a unique type of structural information, ranging from composition and stoichiometry through to residue level proximity and solvent accessibility. Structural MS has proved particularly beneficial in studying protein classes for which analysis by classic structural biology techniques proves challenging, such as glycosylated or intrinsically disordered proteins. To capture the structural details for a particular system, especially larger multiprotein complexes, more than one structural MS method with other structural and biophysical techniques is often required. Key to integrating these diverse data are computational strategies and software solutions to facilitate this process.

We provide a background to the structural MS methods and briefly summarise other structural methods and how these are combined with MS. We then describe current state of the art approaches for the integration of structural MS data for structural biology. We quantify how often these methods are used together and provide examples where such combinations have been fruitful. To illustrate the power of integrative approaches, we discuss progress in solving the structures of the proteasome and the nuclear pore complex. We also discuss how information from structural MS, particularly pertaining to protein dynamics is not currently utilised in integrative workflows and how such information can provide a more accurate picture of the systems studied. We conclude by discussing new developments in the MS and computational fields that will further enable in-cell structural studies.

Contents

1. Introduction

Initially the tool of physicists and analytical chemists, mass spectrometry (MS) expanded into the remit of the biosciences with the development of soft ionisation techniques for the analysis of biomolecules. These breakthroughs include the development of electrospray ionisation (ESI) by Yamashita and Fenn in 1984, followed by development of matrix-assisted laser desorption ionisation (MALDI) in 1985 by Karas and Hillenkamp.[1–3] For the first time, these soft ionisation techniques allowed peptides and biomolecules to be introduced into the gas phase and characterised. Early applications of MS in the biosciences focussed on characterising primary protein structure through the peptide sequencing work of Biemann and Roepstorff.[4–6] This interrogation of primary protein structure still proves important to this day, underpinning modern workflows in the fields of proteomics and peptide biomarker analysis, as well as in many of the structural MS workflows discussed in this review.

In addition to the primary amino acid sequence structure analysed by early MS approaches, it is widely acknowledged that proteins are able to adopt secondary structural elements, such as α-helices and β-sheets.[7] Ultimately, an intricate combination of these secondary structures within the backbone gives the protein a tertiary structure, which can be considered its three-dimensional conformation. Whilst some proteins remain largely fixed in a single native conformation, others transition between a number of fixed conformations, whilst others still are considered intrinsically disordered, rapidly interconverting between several transient conformations. This variation in protein conformation is further modified through biomolecular interactions with other proteins to form multi-protein complexes, or via binding of ligands such as small molecules or metal ions.

Protein structure and biomolecular interactions are closely linked to protein function, providing the basis for understanding key molecular mechanisms in the biology of life and those underpinning human health. As such, interrogating higher order protein structure and interactions with other molecules has become a significant area of biological research. Traditionally, the advanced study of higher order protein structure has been carried out using a widely accepted set of classic structural biology techniques.[7,8] The major experimental techniques within this toolkit are X-ray crystallography, nuclear magnetic resonance (NMR), small angle scattering (SAS), and over the last few years cryo-electron microscopy (cryo-EM).[9–12] The greatest benefit associated with these techniques is that three of them, X-ray crystallography, NMR, and cryo-EM, are able to visualise high resolution protein structures by providing atomic co-ordinates. As such, each of these methods has, to date, been fundamental in characterising the protein structures of multiple biological systems, including nucleic acid binding proteins, the ribosome, and protein ion channels.[9,13–17] However, these classic structural biology techniques are all limited by the requirement for a purified sample for analysis. Whilst expression and purification methodologies have been developed for these purposes, the requirement means that the techniques are intrinsically used for *in vitro* analyses rather than studying the proteins in their true cellular environment. Furthermore, X-ray crystallography, whilst providing high resolution structural detail, ultimately captures static structures. As a result, this technique is unable to sample the dynamics and kinetics of a protein system, factors which are vital for developing a true biological understanding of a protein. Similarly, the majority of cryo-EM studies have also focussed upon the analysis of static structures, although there are cases where the method has been applied to capture protein dynamics.[11,18–21] By contrast, NMR can routinely access information on protein dynamics, however, use of this methodology is limited by protein size and the complexity associated with interpreting the resulting data. SAS techniques similarly can inform upon protein conformational diversity, without the size limits associated with NMR, however the spatially averaging used by SAS results in resolution loss compared with other techniques. Finally, it should be noted that there are still classes of protein for which analysis using any of the classic structural biology techniques remains a challenge, notably highly dynamic structures such as intrinsically disordered proteins (IDPs), heavily glycosylated proteins which contain flexible modifications to the protein backbone, and proteins within a complex environment such as membrane proteins.

An additional suite of techniques increasingly included in the structural biology toolbox are those grouped under the title structural MS. Structural MS describes a set of methods which can be applied *in vitro* and *in cellulo* with relatively small amounts of material to inform upon biomolecule conformation and interactions.[22–26] These structural MS techniques have contributed to structural insights for a range of biomolecules, including DNA and RNA, however, this review will be dedicated to their use for interrogating protein structure.[27,28] The most widely used methods within the structural MS toolkit are affinity purification-MS (AP-MS), crosslinking-MS (XL-MS), hydrogen deuterium exchange-MS (HDX-MS), hydroxyl radical footprinting-MS (HRF-MS), native MS, ion mobility-MS (IM-MS) and top-down MS. While these techniques do not provide atomistic details, each of them provides a different type of structural information, ranging from composition and stoichiometry through to residue level proximity and solvent accessibility, that is not easily obtained otherwise. The diversity of structural information that can be obtained by applying this suite of techniques is one strength of these methodologies. There are also other benefits to the structural MS toolkits which earn it an important place in the field of structural biology. One such benefit is the applicability of structural MS to proteins which, as previously discussed, are traditionally challenging to study using classic high resolution structural techniques, such as IDPs, glycoproteins, and membrane proteins.[29–32] Furthermore, structural MS provides additional value in its ability to accurately sample the entire protein conformational landscape. This means that rather than visualising a static structure, the techniques can sample several conformers, along with the dynamics of the protein. This ability to obtain a broad conformational view is a useful feature of the structural MS techniques, shared by a handful of other experimental techniques such as SAS. Intrinsically incorporated within these structural MS workflows is their ability to report on protein post-translational modifications (PTMs) and interaction partners, along with comparative quantification. Classical structural biology techniques cannot easily characterise these features, which has left them somewhat overlooked to date. We argue, however, that they provide important context for understanding protein structure within the wider biological environment.

Although each method provides benefits, neither structural MS nor any of the classic structural biology techniques discussed so far are able to measure the full extent of protein structure, interactions, dynamics, and PTMs when applied alone. An excellent approach to address this is to integrate multiple complementary structural biology approaches. This collective approach facilitates building a more informative picture of biological systems of interest, whilst also filling any gaps in information that might result from focussing on use of only a single technique. We believe that within this approach, structural MS can provide unique parts of the structural biology puzzle. One such method for integrating MS with other techniques is to use the structural MS toolkit as input data for other techniques. In this case the protein landscape is surveyed using a single MS technique, and then these findings are applied to support a further structural biology tool. One example of this type of integration would be the use of HDX-MS to screen protein mutant structure to determine their suitability for analysis by high resolution X-ray crystallography. A further example would be the use of protein complex composition and stoichiometry obtained from native MS as input data for developing computational models of the protein system. An alternative method of integration involves collecting independent data from multiple structural techniques and combining the findings computationally in order to elucidate structure. By using this approach, it is possible to fit a cryo-EM density map or refine a computational model by filtering and scoring systems based on restraints obtained from structural MS techniques. For both methods described, to fruitfully integrate complementary structural techniques, computational strategies and software solutions are required to facilitate the process. Owing to the varied and non-atomic resolution of data obtained from structural MS, these strategies often require a great deal of computation. To date, a diverse range of approaches including binary classification and Bayesian modelling, have been applied to this challenge.

The purpose of this review is to explore how MS has so far contributed to the field of integrative protein structural biology, and how exciting developments in the field can further do so in the future. The workflows of widely used structural MS techniques will be described, along with the structural information which they provide for the purposes of integration. Publication abstracts in the field of protein structural MS have been analysed to provide a more objective view of which classic structural biology techniques are used in parallel with MS. A brief description of these other experimental techniques, along with examples of fruitful integration with MS will also be presented. In analysing the methods which facilitated data integration within these studies, a major focus will be on describing current state of the art computational approaches, rather than on processing of the raw MS data, as the later has already been the subject of several reviews. Finally, the contribution of structural MS within structural biology, and how this has developed in recent years, will be showcased using two systems which have benefited from an integrative approach, the proteasome, and the nuclear pore complex.

2. Structural mass spectrometry methods

Structural MS is an over-arching term used to describe an array of MS-based techniques which provide information on the conformation and interactions of biomolecules. Structural MS methods can be broadly sub-divided into two distinct categories, peptide-based approaches, and intact protein methods. The major techniques within each of these categories, and the information they provide for the purposes of integration for structural biology will be discussed. The visual depiction of each structural MS technique is shown in Figure 1.

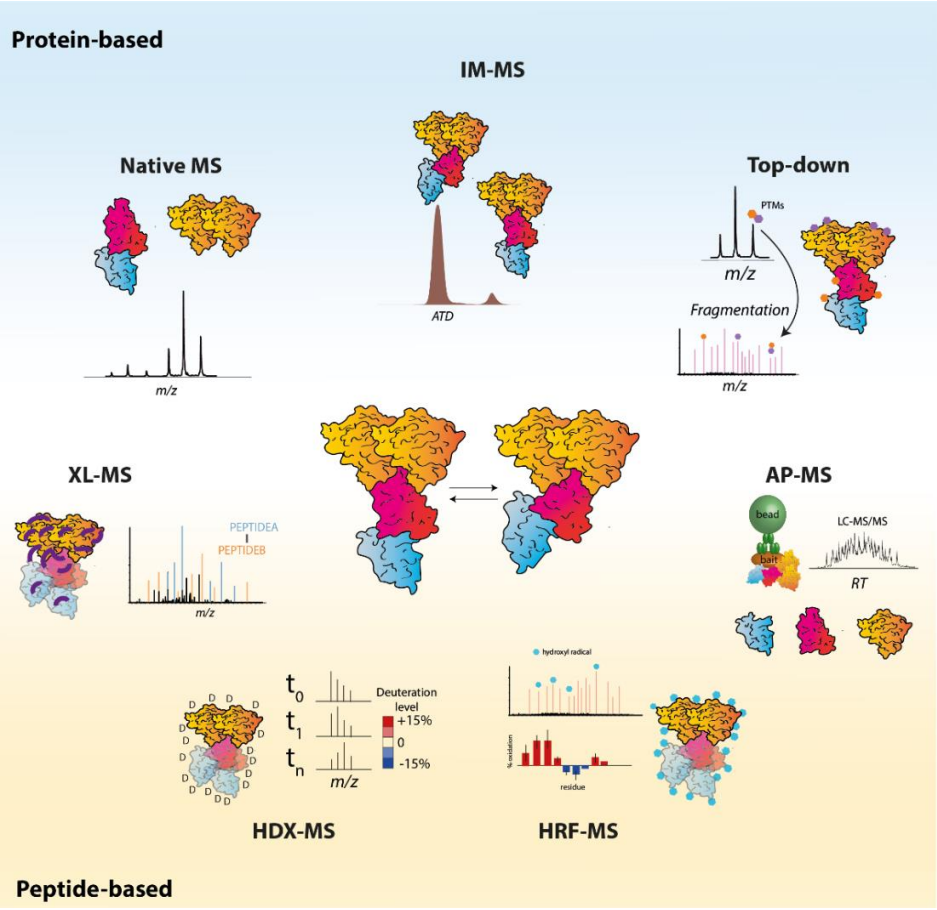

Figure 1: A visual summary of the seven structural MS techniques discussed within the review. The top panel of the figure describes the protein-based methods: native MS; IM-MS; top-down MS. The bottom panel summarises the peptide-based techniques: AP-MS; XL-MS; HDX-MS; HRF-MS.

2.1 Peptide-based methods

Peptide-based or bottom-up structural MS methods is an umbrella term for the group of techniques which approach protein analysis at the peptide level. These methods generally follow the same workflow, using some form of protein labelling or tagging, followed by enzymatic digestion of the protein into peptides, chromatographic separation of these peptides, and analysis by MS. Often the mass spectrometer is used as a readout for labelling events which take place in solution. As a result, the data from peptide-based methods provides a true reflection of the biological solution phase structure and dynamics of a protein. Since they work at the peptide level, and a chromatographic step precedes MS analysis, these bottom-up techniques cope well with sample complexity. These methods are therefore well suited for analysis of structural biology applications in which multiple proteins are present, such as those within native protein environments. Furthermore, the peptide-based techniques can often be applied and quantified under comparative conditions, allowing structural variation due to external stimuli to be studied. Four of the most common peptide-based MS techniques which are commonly integrated for structural biology applications are described.

2.1.1 Affinity purification-mass spectrometry (AP-MS)

The first peptide-based structural MS approach, which has gained momentum over the last two decades, is AP-MS.[33–35] The principle of AP-MS is that a protein of interest is captured, along with any bound interaction partners, through enrichment by a solid support bound ligand which the protein has affinity for. Using this method, the protein of interest can be captured from any matrix in which it is present, often cultured cells or homogenised tissue. There are several different enrichment strategies for AP-MS, and a detailed comparison of these is described in recent excellent reviews.[33] The most common affinity purification workflows, however, use either a targeted antibody for the endogenous protein, or an epitope tag translated into an affinity handle such as a His-tag or biotinylation. Following capture of the targeted protein, the solid support is washed to minimise the presence of unwanted proteins present due to non-specific binding. The final stages of AP-MS are enzymatic digestion using a protease, and analysis by MS. Often the MS stage of AP-MS is coupled to liquid chromatography (LC) in order to separate the resulting peptides, aiding in improved coverage and characterisation of identified proteins. At each stage of this AP-MS workflow it is vital to incorporate suitable biological replicates and control experiments, to eliminate incorrect complex assignment due to non-specific interactions.

AP-MS is a robust and well tested approach which provides excellent information in cases where the researcher is interested in gaining a detailed overview of the dynamic interaction landscape of a single target protein under a range of conditions. The AP-MS methodology enables characterisation of protein-protein interactions in a native context, facilitating identification of novel protein interactors and their stoichiometry. Since AP-MS can be applied across a range of biologically relevant matrices, such as cells or virus particles, information can be gained on a few to several hundred protein interactions in a single experiment.[36–38] Furthermore, the peptide-based nature of the MS analysis performed allows PTMs of protein interactors to be identified and characterised, providing insights into the functional and regulatory roles of these PTMs. The strengths of AP-MS can be further exploited by using a quantitative AP-MS approach, in which protein amounts are determined using, for example, internal standards or isotopic labelling.[39] Quantitative AP-MS facilitates comparison of protein-protein interactions under different external conditions, such as environmental stimuli or stage of development, giving a dynamic view of the interaction landscape for a given target protein. This method has been beneficial in providing a systematic map of protein-protein interactions within the human proteome.[36,37] Similarly, an AP-MS approach was successfully used to map the global interaction landscapes of both *S. cerevisiae* and *D. melanogaster*.[40–42] Several targeted studies have also been carried out using AP-MS, identifying key interactors for host cell proteins upon HIV infection, cellular prion proteins and deubiquitinating enzymes among others.[38,43–49] Finally, it is noted that AP-MS was key in early work to determine the structure of the nuclear pore complex, and its contribution

to this area will be discussed in detail in Section 4.2 of this review. These findings from AP-MS have created a signpost for interesting complexes on which to focus alternative structural biology techniques, whilst the AP-MS methodology has been the early inspiration and basis of several newer structural MS techniques, such as XL-MS. These methods, which will be discussed later in this review, build on AP-MS to provide more detailed protein information on protein structure and dynamics.

2.1.2 Crosslinking-mass spectrometry (XL-MS)

With the evolution of ever more sensitive mass spectrometers over the last 15 years, new structural MS techniques which build on the principles of AP-MS have come to the fore. One such method is XL-MS, a technique which combines chemical crosslinking of proteins with analysis by MS, which is used to study protein structure and dynamics.[50–53] The general XL-MS methodology involves introduction of a chemical crosslinking reagent into a sample of protein or proteinaceous mixture. A benefit of XL-MS is that this mixture can be under native conditions, such as cell culture or tissue, in addition to purified protein *in vitro* systems. The crosslinking reagent contains two reactive moieties able to covalently attach on to amino acid side chains of proteins, connected by a linker backbone of known length. The amino acids most commonly targeted are lysine residues, making use of *N*-hydroxysuccinimide (NHS) reactivity and amino acid hydrophilicity, which ensures they are often solvent exposed in a folded protein. However, reagents do exist which are able to crosslink additional amino acid functionalities, for example carboxyl groups, or diazirine based photoreactive crosslinkers which can attach to any amino acid. The second crosslinker component is linker length, often defined as a "molecular ruler", which provides a distance restraint between the two amino acid residues. The shortest linkers used for XL-MS are "zero-length" crosslinkers such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) or close proximity formaldehyde. These crosslinkers fix interactions to stabilise very closely bound protein complexes. Longer length crosslinkers, also known as structural crosslinkers, have several angstrom distance between the two reactive termini, enabling them to provide restraints which inform on residue level proximity within tertiary protein structure. Two common examples of this family of crosslinker are disuccinimidyl suberate (DSS) and bis(sulfosuccinimidyl)suberate (BS3).

Following on from the chemical crosslinking process, proteins are subjected to enzymatic digestion, harnessing one of the many protocols developed in the fields of proteomics and AP-MS. The resulting mixture of linear and crosslinked peptides is then fractionated prior to MS analysis. The most common XL-MS fractionation approaches are size exclusion chromatography (SEC) and strong cation exchange (SCX) chromatography, which exploit the intrinsic nature of crosslinked peptides being larger and more highly charged than their linear counterparts. Fractionation assists in overcoming one of the great challenges in XL-MS, the relatively low efficiency of the crosslinking reaction. The outcome of this low efficiency is that the resulting reaction mixture contains considerably more linear peptides than crosslinked peptides, making the detection of crosslinked peptides challenging. This inefficiency of the crosslinking reaction is particularly problematic when performing in cell XL-MS.[54] An additional approach to overcome this challenge is to use specially designed crosslinkers with additional functionality, such as MS cleavable crosslinkers disuccinimidyl sulfoxide (DSSO) and disuccinimidyl dibutyric urea (DSBU), enrichable crosslinkers such as PhoX, or the protein interaction reporter (PIR) technology pioneered by the Bruce lab.[55–58] Chromatographic separation prior to MS analysis can also help reduce suppression of the low abundance crosslinked peptides due to inefficient crosslinking, by staggering introduction of linear and crosslinked components into the mass spectrometer. More recent studies have further exploited this principle by incorporation of ion mobility into the XL-MS workflow, with the aim of minimising background signal from linear peptides compared to crosslinked counterparts.[55,59]

Workflows for the MS step in XL-MS have been developed on Time of Flight (ToF) and Fourier transform ion cyclotron resonance (FTICR) mass analysers, however the majority of studies use

Orbitrap systems.[59,60] The robust and reproducible workflows developed on Orbitraps, in combination with the advent of bespoke software, enables these systems to produce high sensitive and detailed crosslinking results.[61] Development of these workflows has been assisted through a 2019 community study of 32 XL-MS labs, the first attempt to evaluate current XL-MS approaches with the aim of establishing best practice guidelines for the field.[62] These XL-MS approaches have facilitated detailed studies of several protein systems, such as structure of the 70S ribosome and G-protein coupled receptors.[63,64] Furthermore XL-MS has been used to access information on complex composition and stoichiometry with a global rather than specific protein focus. Examples of successful global applications include elucidating the mitochondrial interactome, super-complexes in heart tissue, and the histone interaction landscape.[65–67] More recent developments in the XL-MS workflow have facilitated quantitative XL-MS approaches by exploiting classic quantitative proteomics methodologies. One such method is the use of isotope labelled crosslinkers, where for example a comparative XL-MS approach employed BS3 d0/d4 to successfully probe protein dynamics as a consequence of PTMs.[68] Alternative quantitative XL-MS approaches that have been applied to biological systems include stable isotope labelling by amino acids in cell culture (SILAC), and label free data independent acquisition.[69–71] These quantitative approaches allow the structural insights on proteins obtained by XL-MS to be compared under different external conditions or to examine the impact of protein mutations. As a result, comparative quantitative XL-MS can be used to probe the conformation dynamics of a protein system.

2.1.3 Hydrogen deuterium exchange-mass spectrometry (HDX-MS)

HDX-MS has its roots in the approach developed by Linderstrom-Lang and colleagues in the 1950s.[72–75] The concept is built upon the principle that backbone amide hydrogens found within a protein primary structure are able to continuously exchange with hydrogens in surrounding aqueous solution. Given that an amide hydrogen forms a component of every amino acid, with the exception of proline, this hydrogen exchange is possible along the length of the protein backbone. By contrast, whilst hydrogens on amino acid side chains are also able to exchange with solution hydrogens, the rate of this exchange is much faster than exchange on the protein backbone. HDX-MS exploits this principle of hydrogen exchange by using deuterated or “heavy” water (D_2O) in place of the aqueous solution, thus enabling backbone amide hydrogens in the protein to be exchanged for deuterium. Protein samples are then digested using a protease enzyme for analysis at the peptide level by MS, enabling identification of backbone areas which have undergone exchange by the mass difference between hydrogen and deuterium. Unlike other peptide-based methods, in which trypsin is the protease of choice, many HDX-MS studies use pepsin for digestion due to its higher efficiency in acidic conditions, which are necessary to minimise back-exchange.[76–78]

HDX-MS provides useful information for integrative structural biology applications because the rate of amide hydrogen exchange within any given residue of the protein backbone is dependent upon four key factors. These factors are pH, temperature, solvent accessibility and hydrogen bonding. Effects of the two environmental factors, pH and temperature, are tightly controlled during HDX-MS experiments in order to ensure any hydrogen exchange attributed to these factors is minimised. Typically, this is achieved by exposing the protein to the D_2O solvent under the desired temperature and pH conditions for a given amount of time, followed by a quenching step in which the pH is lowered to acidic levels of around 2.5, and the temperature to 0 ºC. These external conditions are maintained during the enzymatic digestion step and any chromatographic separation, since they reduce the backbone amide exchange rate to the minimum. These sample preparations steps are typically carried out in an automated fashion using commercially available or in-house equipment, in an attempt to maintain consistency across samples.[76,77] Newer approaches to tackle the effects of pH and temperature on hydrogen deuterium exchange rate include performing the exchange reaction in the gas phase, either at the mass spectrometer source or in a travelling wave ion guide, using for example ND_3.[79–81] These approaches, along with others which perform in solution exchange, also make use of

top-down fragmentation, thus avoiding back exchange during the enzymatic digestion step of the traditional methodology.

Having tightly controlled pH and temperature, hydrogen exchange is then reliant only upon solvent accessibility of the backbone and hydrogen bonding, both of which are innately tied to the structure and interactions of a given protein. Highly dynamic and exposed regions of a protein will exchange much more rapidly than areas buried deep within the protein folds, shielded by a ligand, or those which are involved in structural hydrogen bonding, for example to maintain α-helices or β-sheets. These differences will become apparent for a given protein or complex following interpretation of the HDX-MS data. It should be noted, however, that due to the nature of the methodology HDX-MS is most informative when applied to purified proteins or complexes, rather than complex cellular mixtures. The strengths of HDX-MS are most evident when examining proteins or complexes in a comparative or differential manner, for example by comparing mutants, or protein states such as apo and halo. Furthermore, the comparative process can be extended further by varying environmental conditions, or length of exchange time, in order to obtain more detailed information. Differences in structure due to modified folding or reduced solvent accessibility of protein regions resulting from these external factors are then identified in the HDX-MS results.[72] To date, HDX-MS has been applied to several systems, intrinsically disordered proteins, α-synuclein, α1-antitrypsin, prions and antibodies.[29,82–86] Several of these studies describe as being near amino acid resolution, such as those examining the stepwise protein folding of RNase H, intermediates of barrel protein folding, and dynamics of the multidrug resistance pump AcrB.[87–89] Examples of ligand binding systems studied include co-factor binding, lipid binding to membrane proteins, substrate-transporter binding, and receptor-ligand binding.[90–92]

2.1.4 Hydroxyl radical footprinting-mass spectrometry (HRF-MS)

In a similar fashion to HDX-MS and XL-MS, HRF-MS relies upon covalent labelling of biomolecules to probe protein conformation and ligand binding.[93–96] The method was originally applied to study the tertiary structure of DNA and RNA, before Chance and colleagues applied the process to investigate protein structure.[97–102] HRF-MS uses hydroxyl radicals ($OH^{\bullet}$) to irreversibly modify the side chains of solvent exposed amino acids within a protein. Hydroxyl radicals can be generated using a number of techniques, including laser based fast photochemical oxidation of proteins (FPOP) and Fenton chemistry using ethylenediaminetetraacetic acid (EDTA) and hydrogen peroxide, among others. Oxidative modifications can theoretically occur for any side chain, including neutral and hydrophobic amino acids, maximising the information obtainable across all residue types. Common HRF-MS modifications have been well characterised, taking into account amino acid side chain chemistries and reactivities, with common observations being addition of oxygen (+16 Da), carbonyl (+14 Da), or decarboxylation (-30 Da).[95] Characterised modifications are of known mass, therefore the changes caused by hydroxyl radicals can be detected following enzymatic digestion of the protein into peptides for tandem MS analysis. Given that oxidative modification of amino acids by HRF-MS is irreversible and stable, the practical challenges associated with LC-MS analysis are reduced compared to the similar structural method HDX-MS. Since hydroxyl radicals are similar to water molecules in their solvent properties, the extent of oxidative modification directly depends upon how solvent exposed the amino acid side chain is within a given protein conformation. Taking into account the reported variation in side chain reactivities previously discussed, along with known MS detection challenges, HRF-MS is believed to inform upon the structure of approximately 65% of a typical protein sequence.[93]

HRF-MS contributes to the field of structural biology by providing solvent accessibility information which can be used to map protein conformational folds, along with identifying interfaces between protein complex domains or bound ligands. By applying HRF-MS under different environmental conditions, such as in the presence and absence of a ligand, variation in the results can be used to identify environment-induced structural changes. As such, HRF-MS has been successfully used to

characterise hydrophobic collapse of α1-antitrypsin and lysozyme kinetics, among other examples.[103–106] Furthermore, time resolved HRF-MS has become popular in recent years, especially using laser based radical generation methods which can pulse on the microsecond timescale. These studies have been used to provide insight into protein dynamics and kinetics, the early examples being used to study the structural biology landscape of the calcium dependent activation of gelsolin, assembly of the G protein coupled receptor-G complex, and dynamics of barstar protein.[107–109] Unfortunately, despite the informative structural information obtainable from HRF-MS, its widespread adoption and integration has been limited by the complexity of data analysis due to residue reactivity differences. The analysis is further complicated by a lack of measure of radical dose due to the radical being scavenged prior to protein modification, for example by reducing agents, metal ion chelators, or buffer components. As such, automated HRF-MS systems, quantitation methods, and robust analysis software still lags behind comparable structural MS techniques. One area in which HRF-MS is progressing more quickly is moving the workflow in cell, with both the membrane permeable hydrogen peroxide and laser based FPOP methods being applied successfully. This approach facilitates protein labelling within organelles and, in the case of *C. elegans*, within whole organisms, enabling structural interrogation under native conditions.[110–112]

2.2 Protein-based methods

In addition to the peptide-based MS methods described above, there is a second category of MS techniques. This second suite of tools are often referred to as protein-based or intact protein methods as they perform the MS analysis directly on an intact protein or complex, often whilst trying to retain the native fold of the protein. Due to the focus on the intact protein, protein-based methods deal less well with sample complexity than the peptide-based methods and are commonly applied to purified systems of a single protein or complex.

An early criticism of protein-based structural MS methods was that MS analyses proteins in the gas phase, whereas the native protein environment is in aqueous solution. This fact led the field of structural biology to question the suitability of MS to study proteins in their native state. As such, several fundamental studies were carried out, ultimately showing that protein structure and activity is preserved throughout protein-based MS analysis for species such as the tobacco mosaic virus and lysozyme.[113–118] Similarly, protein-based MS workflows incorporating ion mobility separation were able to show that, under controlled instrument conditions, native-like protein structure is preserved.[119–121] Furthermore, several modern structural MS workflows incorporate control experiments, for example gas phase trapping in IM-MS, in order to safeguard against attributing gas phase effects to solution phase proteins. However, despite this extensive work, it is acknowledged that there are cases in which analysis in the gas phase cannot be considered biologically informative. Proteins unsuitable for study in the gas phase include those which are susceptible to gas phase collapse, for example those identified as non-natural inverted structures in which hydrophobic regions present as solvent accessible, and some classes of IDP.[122,123] Additional limitations in intact-protein structural MS include protein size, buffer and environmental factors required for analysis. Despite these drawbacks, for many systems protein-based structural MS still provides an excellent methodology for in depth characterisation of structure and dynamics.

2.2.1 Native mass spectrometry

The term native MS was first coined in 2004, giving a new name to the series of early studies which employed native preparation of samples and soft introduction into the gas phase for MS.[24,124–127] Alternative terms for native MS, referenced in these early studies and still to this day, include non-denaturing, non-covalent and macromolecular. The principle of native MS is that prior to ionisation a protein or complex of interest is prepared in solution using non-denaturing aqueous solvent, with tightly controlled temperature, pH and ionic strength in order to retain its native characteristics. The protein or complex is then softly transferred into the gas phase using nano-electrospray ionisation

(nESI), such that it retains both its native conformation and biomolecular interactions. Following ionisation, the protein or complex ions proceed to separation based on their mass to charge ratio (*m/z*) in the mass analyser. In the early days of native MS, the mass analysis step proved challenging for higher molecular weight proteins, with analyte size being limited by the use of quadrupole mass analysers with upper *m/z* limits of only 2000 or 4000. This problem was addressed by development of higher *m/z* transmitting quadrupoles operating at a lower radiofrequency, combined with commercialisation of ToF analysers which have no theoretical upper mass range.[128–130] In the present day, multiple analyser types from diverse manufacturers are suitable for the application of a native MS workflow, with quadrupole ToF (QToF) and Orbitrap instrumentation remaining the most popular.

At the time of writing, native MS, like several other techniques in the field of structural biology, is limited to *in vitro* study. This means that sample preparation for native MS often requires protein purification from an expression system, and that the non-denaturing conditions employed by the technique are purely a mimic of biological conditions rather than being the true cellular environment. Native MS is far from alone in this respect within the field of structural biology, however, current efforts are focussing on better replicating the cellular environment for native MS analysis. Key examples of this are the development of a native liquid extraction surface analysis (LESA) workflow to extract native MS samples directly from tissue, and innovations in facilitating analysis of overexpressed proteins from crude cell lysates in prokaryotic and eukaryotic systems.[131,132]

The remit of native MS to *in vitro* systems has not limited its popularity within the structural MS toolkit as the method has been successfully applied to several protein systems. The technique has successfully informed upon protein composition and stoichiometry for oligomers of amyloidogenic proteins amylin and amyloid β, IDPs, glycoproteins, and even whole viral capsids.[29,31,133–135] These studies have proved particularly insightful given the challenges associated with using classic structural biology techniques to analyse these systems. Furthermore, considerable advances have been made in applying native MS to study membrane proteins, another traditionally challenging system, with the use of detergent micelles aiding analysis of ABC transporters, aquaporin Z and the ammonia channel AmtB.[30,136,137] These capabilities have also introduced the possibility of identifying specific protein-lipid interactions of biological importance, as described in the case of OmpF and outer mitochondrial membrane translocator protein.[138] The molecular mass of a protein obtained from native MS not only provides information on complexation of multiple proteins, it can also provide information on the modification state of the protein, including identification of PTMs, and binding of small molecule ligands or metal ions. Pairing additional functionality with native MS facilitates further characterisation of interactions, building a more detailed picture of the function and dynamics of a given biological system. Approaches such as complex dissociation using collision induced dissociation (CID) and performing analysis under different ligand concentrations can be paired with native MS to determine dissociation constant (K_D) values. Alternatively, interaction interfaces can be identified through comparative native MS of protein variants prepared through mutagenesis, or of multi-ligand studies as demonstrated for MurJ.[139]

2.2.2 Ion mobility-mass spectrometry (IM-MS)

IM-MS combines MS, normally native MS to retain native protein characteristics, with a technique first reported in 1898 called ion mobility spectrometry or ion mobility separation.[140] The ion mobility portion of IM-MS separates ions based upon their charge, mass, and physical shape, providing a measure of protein conformational envelope. There are multiple types of IM-MS, however, the most common ion mobility methods used for structural biology are drift tube ion mobility spectrometry (DTIMS) and travelling wave ion mobility spectrometry (TWIMS).[141–144] DTIMS pulses ions into a fixed length drift tube which is filled with an inert gas such as nitrogen or helium. A weak electric field is applied to the tube to pull ions through, with ions travelling through at different rates based on their size due to more extended structures being slowed down by an increased number of collisions.

Released commercially by Waters in 2006, and the most popular form of IM-MS for protein structure, TWIMS like DTIMS is based on a gas filled drift tube.[145] TWIMS propels ions through the tube using travelling waves, resulting in conformational separation due to differences in travelling speed due to both charge, size, and shape. In depth explanations of DTIMS and TWIMS, along with other forms of ion mobility spectrometry can be found elsewhere. The outcome of DTIMS or TWIMS IM-MS is an arrival time distribution (ATD) with gaussian peaks corresponding to distinct conformational families. These families may be comprised of a single protein conformer, or of multiple conformers with comparable ion mobility under the resolution accessible by the instrument.[146] Each ATD peak is also associated with an *m/z* spectrum for each peak, allowing identification of the protein species by molecular mass. The size of the conformation, as given by an ATD peak can be translated into a rotationally averaged collision cross section (CCS) either from first principles (for DTIMS) or by applying a calibration.[147–153] The CCS value obtained from IM-MS is essentially a conformational envelope for a given protein, providing low-resolution structural information.[154–157]

Many of the instrument requirements for IM-MS are similar to those discussed for native MS, such as soft nESI and high mass range analysers. The most popular analyser types used within structural IM-MS-enabled instruments are ToF or QToF, due to timescale compatibility with ion mobility separation compared with alternative high resolution analyser types such as Orbitraps. Instruments are also required to contain a DTIMS or TWIMS cell to facilitate IM-MS. Such instruments exist both commercially and those made *in house*, each with differing instrument geometries that facilitate varied experiments to be performed.[143] Using these instruments, IM-MS has been applied to the study of protein misfolding and aggregation, for example using amyloid β and β2-microglobulin, and to analyse viral systems.[158–162] The strength of these studies comes from the fact that IM-MS samples across the entire conformational landscape, identifying multiple conformational envelopes for a single protein, and therefore providing detailed structural oversight.

An additional strength of IM-MS is that the technique can give insights into the dynamics and stability of protein conformations. This is achieved by making use of the ability to activate ions in the gas phase, for example by using collisions, injection energies, or alternative methods.[118,163–168] One common workflow for this type of experiment monitors the ATD obtained from IM-MS as an applied voltage is increased in a stepwise fashion. As the voltage is increased the protein conformation being studied has access to increased activation energy, facilitating complex dissociation or structural changes. This facet of IM-MS methodology has found particular favour in biopharma, being used to characterise antibodies and protein-drug binding events.[157,169–171] Several non-biopharma uses of the methodology also exist, and in some cases the method has even been successfully applied to obtain kinetics information including potential energy diagrams.[172–174] These kinetics, and the dynamics associated with them, are a fruitful source of information for building an in depth understanding of the structural biology of a given protein.

2.2.3 Top-down mass spectrometry

Top-down MS, also called top-down proteomics, is a term used to describe a process in which the intact mass of a protein is measured, followed by fragmentation by tandem MS (MS/MS) or multi-stage MS (MS^n).[175–179] The term is used as a contrast to the bottom-up proteomics methodology on which many peptide-based structural MS approaches are based, in which MS analysis is performed after enzymatic protein digestion. Top-down MS is a powerful tool in protein characterisation, widely reported to give near to 100% sequence coverage, compared to the 50-90% reported for bottom-up proteomics. This improved sequence coverage is attributed to the fact that top-down MS performs all fragmentation within the mass spectrometer, reducing losses due to inequalities in ionisation efficiency or sample preparation stability for tryptic peptides. It is also for this reason the top-down MS is particularly beneficial for the study of proteoforms, including sequence variants, degradation or truncation products, and differences in PTMs such as glycosylation.[180,181] However, it must be noted

that the increase in sequence coverage attributed to top-down MS must be balanced with its limitations on the size of protein and complexity of the sample. The information obtained from top-down MS can be further strengthened through the method of sample preparation and introduction into the mass spectrometer.[181,182] One such method is denaturing top-down MS, in which precursor protein(s) are prepared in a denaturing solution such as one containing organic solvent or detergents, resulting in any native structure and interactions of the protein being lost. Although the denatured nature of the protein could be seen as a drawback, the process makes dissociation easier and thus can provide excellent coverage for PTM and proteoform characterisation. Furthermore, denaturing top-down MS can be used in tandem with separation techniques for sample introduction, such as reverse phase LC or capillary electrophoresis, which can facilitate analysis of a more complex multi-component protein mixture. The alternative sample introduction method is using nESI in a native MS workflow to maintain the native structure and interactions of the protein of interest. This form of top-down MS is sometimes referred to by the term complex-down, in cases where the analyte of interest is a protein complex rather than a single protein. In the case of native top-down MS, inducing dissociation in the protein backbone can be a more challenging endeavour, however, the results are fruitful since the fragmentation that is observed is considered to be backbone selective. As a result, native top-down fragmentation can inform upon how exposed protein regions are in a given conformation and can provide conformation-specific proteoform or PTM characterisation. The dissociation method selected for use during top-down MS also influences the information gained from the technique. Several protein fragmentation methods have been developed for this purpose, and these are summarised in Table 1.[183]

Acronym	Name	Brief Description
CID/CAD	Collision induced dissociation or collisionally activated dissociation	Dissociation is induced by an increase in ion internal energy caused by collision with inert gas molecules.
HCD	Higher-energy collisional dissociation	A term used in Orbitrap instruments to describe beam type CID.
ECD	Electron capture dissociation	Dissociation is induced by capture of one low energy electron from an electron beam by a cation.
ETD	Electron transfer dissociation	Dissociation is induced by capture of a low energy electron from a radical anion by a cation.
EID	Electron ionisation dissociation	Dissociation is induced by cation excitation using a high energy electron beam.
AI-ETD	Activated ion electron transfer dissociation	Uses infrared photoactivation concurrent to ETD.
EThcD	Electron transfer dissociation/ higher energy collisional dissociation	Hybrid ETD and HCD dissociation technique.
ETciD	Electron transfer dissociation/collision induced dissociation	Hybrid ETD and CID dissociation technique.
UVPD	Ultraviolet photodissociation	Dissociation is induced by an increase in internal energy due to capture of ultraviolet photons.
IRMPD	Infrared multiphoton photodissociation	Dissociation is induced by an increase in internal energy due to capture of infrared photons.
SID	Surface-induced dissociation	Ion is accelerated and collided with a surface to induce dissociation.

Table 1: A summary of the major dissociation techniques used in top-down MS.

Instrument and processing requirements for top-down MS are dependent upon both the dissociation technique being employed, and the number of dissociation steps required for analysis. As a result, top-down MS has been performed across a range of analyser types, including commercial QToF, FTICR, and Orbitrap analysers, along with several modified analogues. This variation in top-down

methodology has inspired two community-led interlaboratory studies, focussing on monoclonal antibodies and histone proteoforms respectively.[184,185] Additional successful examples of the application of top-down MS include characterisation of calcium binding to calmodulin, mutations in the Alzheimer's implicated protein Pin1, and phosphorylation sites of cell cycle kinases.[186–189] These studies exemplify how native and denaturing top-down MS can provide vital information on proteoforms and PTMs. Although some may consider these insights not to be within the remit of pure structural biology, we would argue that understanding these factors are vital in establishing a complete understanding of the protein landscape. Given that PTMs, in particular glycosylation, have been closely linked with protein folding and interactions, their consideration as the field of structural biology progresses is paramount. This has been particularly relevant under the present circumstances, where top-down MS and bottom-up proteomics have proved an important tool in characterising glycosylation of the SARS-CoV-2 Spike protein.[190,191]

2.3 Structural data obtained from MS methods

Structural data obtained from the MS techniques discussed up until this point can broadly be placed into many categories. For the purposes of this review article, ten of these categories have been defined: (i) composition; (ii) stoichiometry; (iii) conformational envelope; (iv) interactome; (v) interaction interfaces; (vi) solvent accessible surface; (vii) topology; (viii) conformational changes and dynamics; (ix) kinetics; (x) PTMs. After sampling the protein structural biology literature for publications which use MS, the proportion of occurrences in which a given MS technique is used to obtain structural data within each of these categories was determined. These data are summarised in Figure 2, and further details of how the information was obtained and processed can be found in the supplementary information of this review.

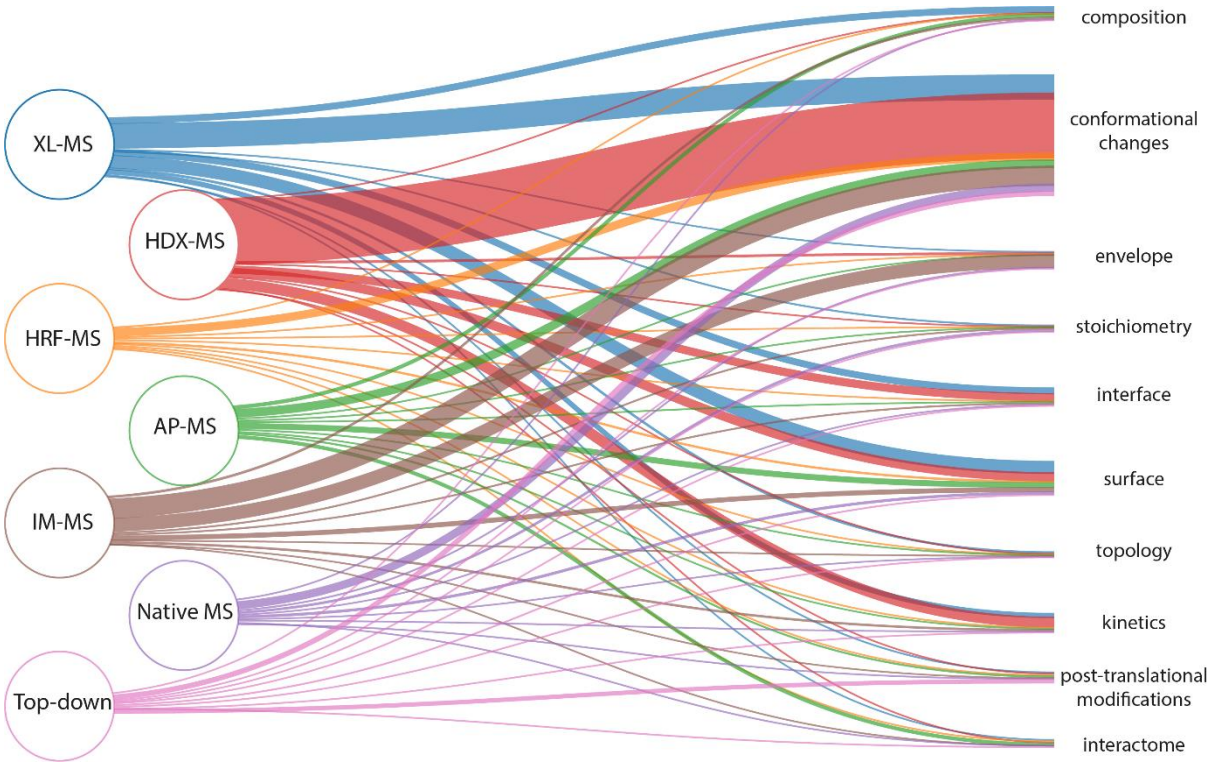

Figure 2: Schematic linking the structural MS techniques discussed in this review with the structural data each provides. Line width infers the proportion of publications on protein structure for a given MS technique which uses the technique for a given structural purpose. Details describing how the data underpinning this figure were collected can be found in the supplementary information.

Each of the seven structural MS techniques is found to report fundamental structural data on protein complex composition and stoichiometry, information which underpins all classic structural biology and modelling techniques. These findings are intrinsic to the *m/z* measurement which defines MS, however, the breadth of proteins over which they are obtained, global landscape compared to protein focussed, varies depending upon technique selected. Furthermore, the particular flavour of the MS technique must also be taken into consideration, as, for example, denaturing conditions from top-down analysis would not in fact inform upon complex composition and stoichiometry. The *m/z* measurement associated with MS analysis similarly facilitates PTM identification, and as such structural data of this kind is attributed across all the structural MS techniques studied. However, it should be noted that the techniques are not equal in this regard, with top-down MS clearly the preferred route for obtaining this information. The importance of PTM insights in understanding the context of protein structure has been discussed previously within this review.

Building upon the fundamental data discussed, several of the structural MS techniques are in addition able to probe protein interactions. XL-MS identifies interactions in a global fashion, with both structural and interaction crosslinking capturing multiple interactions across the protein landscape. As a result, XL-MS is considered a contributor to structural data on the interactome. However, the largest contributor to probing protein connectivity of the interactome was identified as AP-MS. Although based around a focussed affinity purification approach, the relatively long history of AP-MS means the technique has developed to incorporate several multi-bait experiments within a single study, thus building a more global interactome picture. By contrast, techniques such as HDX-MS and native MS also study protein interactions in a focussed fashion, however, at present they are limited to identifying interactions for a protein or subset of proteins. The methodologies associated with HDX-MS and XL-MS go beyond simply identifying an interaction, providing additional information on the interaction interface. HDX-MS achieves this through application in a comparative fashion and subsequent identification of differences in the level of deuteration. Structural XL-MS probes interaction interfaces through the “molecular ruler” portion of the chemical crosslinker. Using this ruler, XL-MS provides a distance restraint, from which residue level proximity is inferred.

Determining solvent accessible surface is another form of structural data which the findings presented in Figure 2 suggest is dominated by XL-MS and HDX-MS. HDX-MS achieves this using the same methodology discussed for determining interaction interfaces, by assessing level of deuteration on the protein backbone. Surface identification in XL-MS exploits the fact that chemical crosslinking of any type is only able to proceed if the amino acid being targeted is accessible to the crosslinking reagent. Given this fact, crosslinked residues identified by XL-MS can be determined to be solvent accessible in the structure or complex of interest. Building on this finding, recent work has applied similar principles to dead-end crosslinks, also known as monolinks.[192] In these cases the chemical crosslinker reacts with an amino acid residue on only one terminus and therefore cannot provide a distance restraint for residue level proximity, however, the findings have been successfully used to provide solvent accessibility information. Two little considered but potentially powerful contributors to determining solvent exposed surface are top-down MS and native MS. With application of the correct dissociation method, for example ECD, and native sample preparation, top-down MS selectively fragments exposed protein regions to provide a measurement of solvent accessibility.[183] Protein charge state distribution obtained from native MS alone can similarly be used as a measure of solvent accessible surface. The principle of this is that amino acid protonation is only possible for solved exposed basic residues, and this is reflected in the observed charge state distribution, most clearly demonstrated by the reported difference in the native MS spectra of folded and denatured proteins

Structural data relating to obtaining a conformational envelope for protein complexes and sub-complexes is dominated by findings from IM-MS. This finding is expected, as IM-MS is the sole

structural MS technique which can provide a direct measure of protein size and shape. It should be noted, however, that this measure is only low-resolution, being based on a rotationally averaged structure, and relies on the assumption that native protein conformation is retained into the gas phase. Nevertheless, identification of the conformation envelope through IM-MS also has great advantages. IM-MS samples across the entire conformational landscape, therefore multiple conformational envelopes can be identified for a single protein, providing detailed structural oversight of the system. This information can then be converted into widely accepted CCS values for comparison with theoretical CCS values calculated using one of the available software packages.[147,193–195] Some variation in method and accuracy of CCS determination, however, is still present within the field, and therefore must be considered when using this data type.

Arguably the greatest strength of structural MS is the data which all MS techniques are able to provide upon the conformational changes, dynamics, and kinetics of proteins and their complexes. Each of the structural MS techniques discussed is able to do this to some degree, however the findings unsurprisingly suggest that the major contributor of this data type is HDX-MS, followed by both XL-MS and IM-MS. IM-MS informs upon dynamics via the ion activation methodology previously discussed, in which energy is introduced to the system during MS analysis in order to facilitate conformational changes. Whilst this approach has benefits, the fact that the workflow occurs in the gas phase rather than in solution means that the observed conformational shifts must be considered with care, to ensure that the species are true native protein structures rather than gas phase artefacts. HDX-MS and XL-MS avoid this problem by performing the labelling step which informs upon dynamics in solution, and simply using gas phase MS as a readout. Results from differential HDX-MS, in which comparative analysis is performed in response to a stimulus, are therefore a particularly rich source of information on protein kinetics and dynamics.

3. Integration of MS data

3.1 Integration of multiple MS techniques

The simplest form of integration of MS for structural biology is the combination of multiple structural MS techniques to achieve a desired outcome. In an attempt to assess the prevalence of this approach, the data obtained from publication abstracts in the field of protein structural MS were mined for the co-occurrence of MS techniques. The resulting findings are displayed in Figure 3. Consideration of this figure highlights an important technical question on what is meant by the term of integration, as whilst techniques may appear to co-occur there are several workflows incorporated within this remit. Co-occurrence of techniques could for example indicate use of two techniques in an integrated experimental workflow, such as the use of native MS as a precursor to top-down MS. An alternative interpretation of co-occurrence is that multiple MS techniques were applied to a single biological system within a publication to provide complementary structural data. The findings from each technique are then considered together in order to come to a significant finding about the system, but the integration of the techniques is in the synthesis of an idea rather than direct experimental integration. A final interpretation of co-occurrence would be computational integration of data from multiple MS sources, however, it is speculated that this would be rare for MS techniques alone given that many such workflows require additional modelling or high-resolution structural data.

Taking into account the different interpretations of the term co-occurrence, in combination with the known structural data obtained from each MS technique, the frequency of multiple MS technique integration can be considered. Whilst a detailed explanation of all instances of co-occurrence would be too extensive for this review, key findings will be mentioned. It is noted, for example, that native MS, IM-MS and top-down MS all co-occur with each other at a high rate. As mentioned previously, this likely results from direct integration of experimental workflows. Similarly, top-down MS and HDX-MS are observed to co-occur, and it is widely reported that these techniques have been combined

experimentally in an attempt to avoid back exchange issues attributed to the bottom-up HDX-MS approach. The observed high level of co-occurrence between HDX-MS and IM-MS, however, is less likely due to an experimental integration, and rather to the use of each technique individually to study a single protein system. Among the structural MS techniques discussed in this review, HRF-MS appears to be used in the least integrative fashion, highlighted by reduced levels of co-occurrence in publication abstracts. It is unclear as to why this may be, given that HRF-MS experimental workflows and instrumentation are accessible, and the structural data obtained are highly complementary to other structural MS techniques. We therefore hypothesise that the reduced co-occurrence can be attributed to the challenges in HRF-MS data complexity discussed in Section 2.1.4. By contrast, XL-MS shows a high level of co-occurrence with all other structural MS techniques, and it is likely that this covers several variations in the form of integration. XL-MS is likely popular for integration due to the diverse range of structural data it provides, making its findings complementary to a diverse range of other structural MS techniques. The techniques with the greatest level of co-occurrence with XL-MS are AP-MS and HDX-MS. This provides an interesting finding, as AP-MS integration is likely through a combined experimental workflow, whereas HDX-MS integration would more accurately be attributed to non-experimental integration. Given that XL-MS and HDX-MS provide complementary structural data it is probable that their integration also occurs in combination with computational modelling and non-MS based experimental techniques.

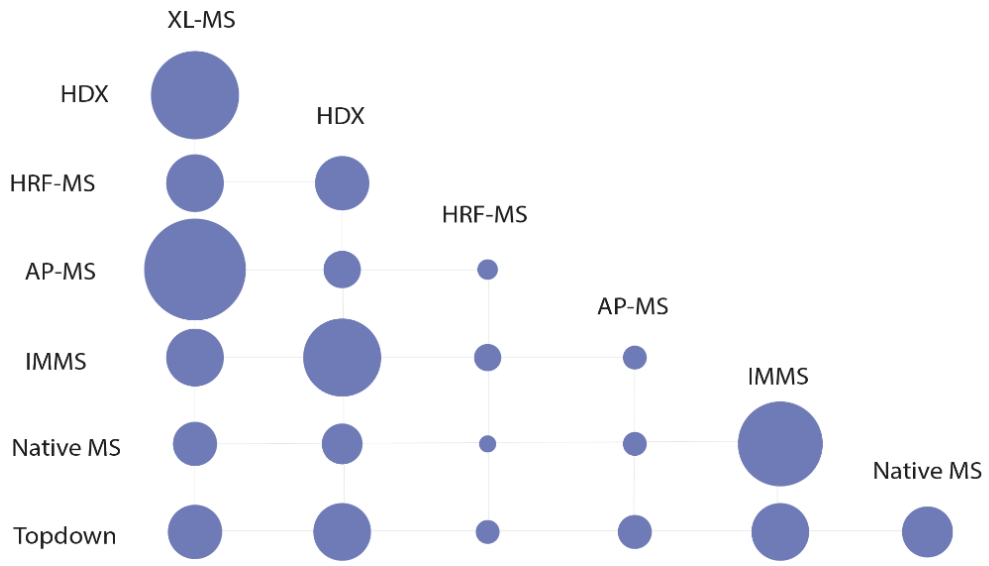

Figure 3: Data showing the number of times structural MS techniques co-appear in publications in the field of protein structural biology. Relative amounts are indicated by circle size. Details describing how the data underpinning this figure were collected can be found in the supplementary information.

3.2 Integration of MS with other experimental methods

MS data can further be integrated for structural biology applications by combining MS data with data obtained from other experimental techniques. Experimental methods in which this has been successfully achieved includes each of the four classical structural biology techniques, X-ray crystallography, cryo-EM, SAS and NMR. In addition, there are also several other experimental methods which have been used in tandem with structural MS approaches, for example biophysical techniques, microscopy and genetic methods. In this section we will discuss in brief these non-MS experimental methods, focussing on the structural data which they provide and examples of their successful integration with structural MS.

In order to evaluate the frequency of integrating non-MS experimental techniques with structural MS we once again reviewed co-occurrence of techniques in publication abstracts in the field of protein structural MS, with the results shown in Figure 4. It should be noted that, as discussed in Section 3.1, this study reports co-occurrence which can take a number of forms, only one of which is computation integration of the data. Data in Figure 4 highlights two key points for discussion when considering integrating MS with other experimental techniques. The first of these is that certain MS techniques are integrated to a far greater extent than others, for example XL-MS and HDX-MS are readily used with all other experimental techniques compared to the intact protein MS methods. There are several factors that could contribute to this finding, including the type of structural information provided by these MS techniques, the availability and deposition of data, the accessibility of computational methods to facilitate integration, or simply easy collaboration of research groups working in these areas.

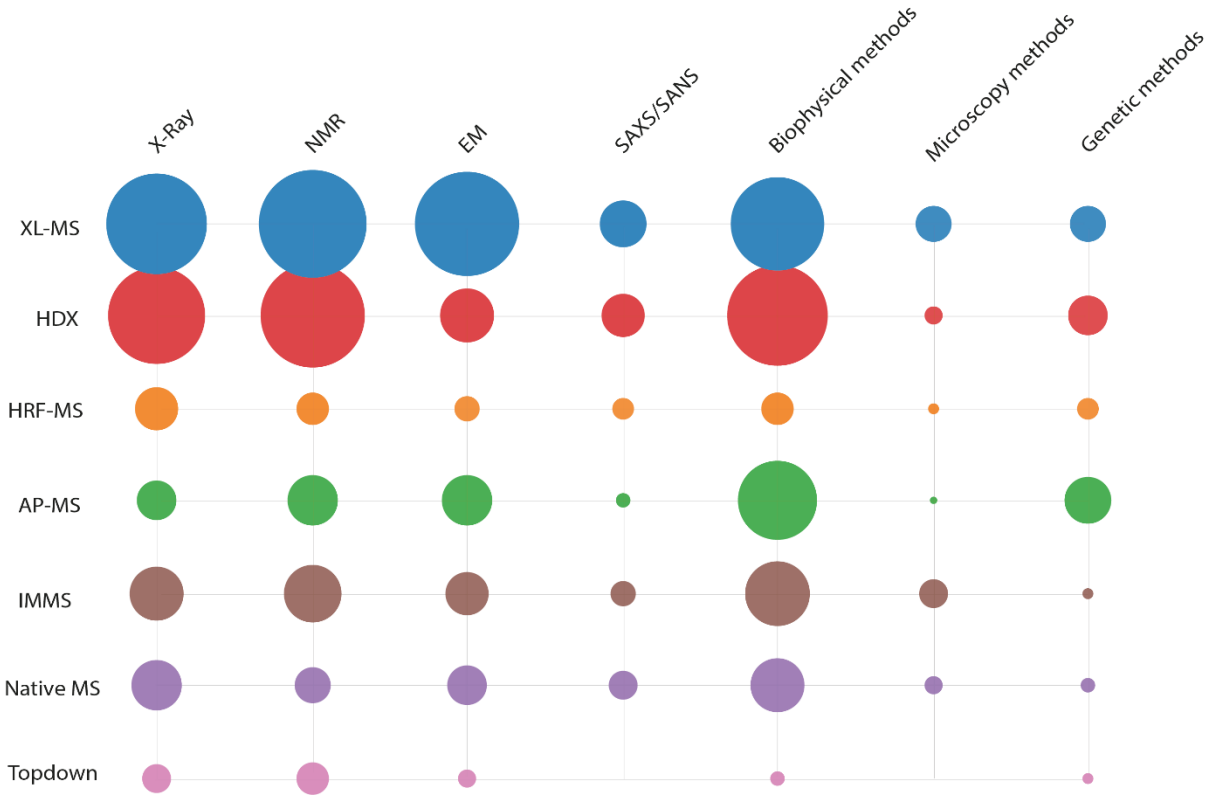

Figure 4: Data showing the number of times non-MS structural techniques appear together with structural MS techniques in publications in the field of protein structural biology. Relative amounts are indicated by circle size. Details describing how the data underpinning this figure were collected can be found in the supplementary information.

Data in Figure 4 also highlights the fact that some MS methods mesh particularly well with other experimental techniques, such that they exhibit high levels of co-occurrence. This finding can be attributed to the intrinsic nature of both the MS and non-MS experimental techniques, such that they provide complementary structural data for the purposes of integration. One example of this is the fact that native MS is commonly associated with X-ray crystallography or cryo-EM. This is because the stoichiometry of protein complexes can be reliably determined by native MS, whereas this cannot always be correctly obtained otherwise, for example due to crystal packing effects in X-ray crystallography. As such, the stoichiometry from native MS is used as starting information during the modelling of cryo-EM or X-ray crystallography density maps.[196] While complex stoichiometry is often set at the very start of an integrative study such as this, it is nevertheless particularly important, and

can have a profound effect on the outcome. Varying basic building blocks such as stoichiometry can vastly increase the complexity of building models, and in some cases the confidence in, for example, the number of a given type of subunit in a complex is high and does not warrant using it as a free parameter. As such, use of reliable stoichiometry information from native MS avoids a critical mistake in this initial assessment, preventing both fitting the wrong model to the data, or obtaining an incomplete or over-specified model that will be very hard to fit properly. Similarly, microscopy techniques are often coupled with XL-MS, which provides additional information regarding residue level contacts. Distance restraints obtained from XL-MS can be intuitively introduced in modelling and refinement pipelines as distance restraints and combined with data from cryo-EM, as will be discussed in detail for the proteasome and nuclear pore complex in Section 4.

3.2.1 Cryo-electron microscopy (Cryo-EM)

Cryo-EM imaging provides direct experimental evidence regarding the conformation of protein complexes. From a sample in solution, a thin film is deposited on a grid, and flash-frozen, to prevent the formation of large ice crystals that would preclude from observing the protein of interest. An electron beam is focused and passed through this thin sample, and the resulting pattern is captured by high-resolution electron detector, effectively producing images of the sample. High resolution structure is achievable in these images, although even low-resolution data can be useful, as it constrains the possible arrangement and orientation of the parts of a protein complex. The obtained images are next aligned and rotated, then combined to estimate the most likely shape to have produced those images. The resulting density can then be interpreted in term of structural models of varying precision depending on the resolution. A full cryo-EM data processing pipeline contains many steps, from contrast transfer function (CTF) estimation and correction, to ice thickness determination, to 2D and 3D classification and reconstruction. Further details on these types of cryo-EM data processing can be found elsewhere.[197]

Assignment of the observed density obtained from a cryo-EM experiment is non-trivial. Limitations in the orientation of the sample, and the resolution of the obtained data, means that it is often important to complement cryo-EM with other techniques. Without this complementary data it is possible to fit a complex in the wrong orientation in low-resolution envelopes. Furthermore, since large proteins and protein assemblies are dynamic, switching between multiple states, they can give rise to multiple alternative organisations and compositions. This process is stochastic, based on the kinetics of the interactions between sub-units. These dynamics are reflected in cryo-EM samples, which are often highly heterogeneous, and require careful preparation to produce interpretable images. Methods that provide information on flexible regions, and the kinetics of assembly, are therefore essential to provide context to the maps and structures generated from cryo-EM data. HDX-MS, for example, can be used to provide information on loop conformation and changes in flexibility.[198] XL-MS is also useful, as it provides additional information regarding the structure of a complex, and to ascertain the correct assembly of complexes observed via cryo-EM.[199] In this way, XL-MS may provide the additional information required to determine the topology of protein complexes observed in the map. AP-MS and native MS have also been used to validate the stoichiometry and composition of a complex observed by cryo-EM,[200] whilst native MS and IM-MS have been applied in combination with low-resolution EM to obtain a structural model of protein assemblies.[201]

3.2.2 X-ray crystallography

Most known protein structures have been solved using X-ray crystallography. From a highly concentrated solution of the system of interest, a stable periodic arrangement of the system can form a crystal. After crystallisation of the protein or protein complex of interest, and if the crystal(s) have become large enough, an X-ray beam is passed through the sample, and the resulting diffraction pattern is recorded. Because the observed patterns are a function of the spacing of elements within

the crystal, and its unit cell, computational methods can be used to evaluate the electron density in the crystal's unit cell.

The diffraction patterns of X-ray crystallography can be interpreted in term of electron density maps. Those maps are not in common use in integrated biology however: their interpretation in term of atomic coordinates is routinely done during the processing of the data, before its deposition in online data banks. Therefore, 3-dimensional atomic models are used in virtually all cases. This differs from cryo-EM due to the generally high resolution of X-ray data, and the impossibility to resolve the X-ray data if it is too low resolution, as the diffraction patterns degrade. The PDB provides access to a rich dataset of solved protein structures, and new structures can be proposed on the basis of homology with solved protein structures in this dataset.

Some complexes may not be amenable to crystallisation, even though its constituent proteins could be resolved. By using XL-MS (see Section 2.1.2), valuable information on the organisation of such complexes can be obtained.[199,202–204] Native MS (see Section 2.2.1) and collision-activated MS protocols (see Section 2.2.2) can also provide information regarding the relative stability of complexes, as well as information regarding the number of subunits in those complexes where the number or orientation of subunits is either variable, or unknown.[205]

3.2.3 Nuclear magnetic resonance (NMR)

NMR is a spectroscopic technique that has been long used to solve macromolecular structures. The nuclear spin of atoms in the sample of interest are polarised by a strong magnetic field; a second field is then applied, perturbing the alignment, and producing an electromagnetic wave that can be recorded. The intensity and frequency of this wave is dependent on the composition and conformation of the molecules in the sample. Nearby atoms may undergo a transfer of their spin polarization, which can be used to determine the overall structure of a biomolecular complex: J-couplings provide distance restraints between pairs of atoms, that can be used in a modelling framework to obtain likely structures. Solution NMR is generally used to solve biomolecular structures, and the inherent flexibility of the system of study can be captured in the information provided by NMR. Time-resolved NMR also provides kinetics regarding both conformational transition in proteins, as well as ligand binding, of particular importance to pharmacokinetics.[206] IM-MS can also provide kinetics information, and can be compared with NMR data, although they are usually on different time scale. [207]

HDX-MS can be used to determine unstructured regions and create constructs with better resolved NMR spectra.[208] Chung et al. used HDX-MS and NMR in combination to obtain information on the conformation and kinetics of lactalbumin. [209] HDX-MS and NMR have a certain synergy since NMR can be used to detect isotopically marked atoms. Both methods can also be employed on flexible and disordered proteins. While the processing of the resulting data can be complicated, the results above show promise. MS combined with arginine footprinting can also be used to assist in the spectrum assignment, as shown recently.[210]

3.2.4 Small angle scattering (SAS)

Small angle X-ray scattering (SAXS) is an experimental technique within the SAS family, whereby an X-ray beam is diffracted through a solution containing a sample of interest, and an absorption curve as a function of the angle is obtained. The intensity as a function of the angle is related to the scattering factors of the different components of the sample, as well as their relative spatial orientation. Because the particles are usually assumed to rotate freely in solution, with no preferred orientation, the curve is expressed as a rotational average of the scattering due to the full particle. Analysis of the variations in intensity as a function of the angle provides information on the overall shape of the system: a Fourier transformation of the SAXS data provides a representation of the intensity as a function of the

internal distance between parts of the particle. From this, potential shapes can be constructed that would produce a similar pattern.

SAXS thus provides information regarding the shape of the complexes present in the solution, but high-resolution is difficult to achieve, as the signal-to-noise becomes larger at the larger diffraction angles containing the signal describing the high-resolution details of the shape. SAXS however is an excellent candidate to be paired with XL-MS: the crosslinks can provide restraints to initially select candidate poses, which can then be scored with the SAXS data. This combined approach has been shown to produce superior results to either independently[211] and has been used to solve novel complexes.[212] Recent work on the nucleosome remodeler ISWI demonstrates how relatively low-resolution data can be efficiently combined to provide a well-resolved structure.[212] XL-MS data was carefully analysed to provide accurate contact points between subunits of the complex. This step included software modifications by the authors, to tune their crosslinking detection method to the study's setup. This type of development underlines how computational work is crucial to better interpret experimental data. The obtained crosslinking data was used in combination with homology models to obtain docked poses of the entire complex. Finally, SAXS was used to confirm that the overall complex conformation is likely correct, and the obtained structure was analysed and provided information regarding the regulation and conformational transition of the complex.

In addition to SAXS, there is a second comparable SAS technique called small angle neutron scattering (SANS). SANS records the intensity of a diffracted neutron beam as a function of the diffraction angle, in similar ways to SAXS. Neutron scatters better against light-element that will not interact strongly with an X-ray beam and can be combined with isotope labelling and can provide more detailed information regarding the atomic composition of the sample. SANS combined with native MS has been used to identify the oligomeric state, relative position of subunits, and overall assembly of large protein complexes.[213]

3.2.5 Biophysical methods

A large number of biophysical techniques have been employed to probe the structure and dynamics of protein complexes, often in conjunction with MS methods. We provide an overview of those techniques here and references as to their use.

Fluorescence (or Förster) resonance energy transfer (FRET)

FRET is a powerful technique to observe the dynamics of a system in solution. Fluorophores are covalently attached to the system, and their distance can be estimated by measuring the energy transfer happening between them. This gives direct insight into the dynamics of the system at hand, with minimal perturbation of its constituents. This information can be analysed to provide useful insights in the changes occurring over time in the system. When combined with MS, a more accurate picture of the system's structure and dynamics can be painted: studies combining ion mobility and FRET allow for the concomitant detection of the dynamics of complexes, as well as their overall shape.[214,215]

Dynamic light scattering (DLS)

In dynamic light scattering, light scatters on particles increases their diffusion, with larger particles diffusing more slowly. By observing this change in diffusion, the size of the particles can be inferred, as well as their aggregation propensity. The diffusion equation is dependent on temperature, and careful control of the temperature in the sample is therefore necessary. The sample otherwise requires very little adaptation, making this method applicable to a very large range of biomolecules, from small aggregating peptides to large biomolecular complexes, as well as protein-ligand interactions. [216] This is exemplified in a recent work combining MS and DLS to characterise amyloid fibrils intermediates.[217]

Size exclusion chromatography-multiple angle light scattering (SEC-MALS)

Multiple angle light scattering with size exclusion chromatography can be used to determine the mass of protein, and can be used to separate complexes with different oligomeric states.[218] Variation using other chromatography methods, such as ion-exchange, are also possible.[219] Antibody/antigen complexes were recently characterised by a combination of native MS and HDX-MS, and SEC-MALS.[220]

Electron paramagnetic resonance (EPR)

Electron paramagnetic resonance provides information on the distance between atoms. By exciting the spin of unpaired electrons, the distance between atoms can be determined. Functional groups with unpaired electrons usually need to be covalently attached to the protein of interest.[221]. The technique can be used in solution, and the absence of unpaired electron in proteins makes it easier to obtain a good signal. The technique has been used in conjunction with ESI-MS to characterise the binding modality of vanadium compounds to myoglobin.[222]

Surface plasmon resonance (SPR)

Surface Plasmon Resonance can be used to study the interaction of proteins with a surface, often coated with a compound of interest, often other biomolecules, to study their interaction.[223] The adsorption/desorption of molecules on the surface (often a gold film) alters the reflection of an incident light, and the reflected light (and its changes are captured by a detector.[224] SPR has been used in conjunction with native MS to characterise and identify proteins.[225]

Raman spectroscopy

Raman spectroscopy uses the spectrum of inelastically scattered light (Raman scattering) on a sample to deduce vibrational properties of the molecules within, such as bond lengths, angles, and other internal motions of the molecule. By its nature Raman spectroscopy provides information regarding both the structure and the fast kinetics of a molecule. Raman spectroscopy has been used in conjunction with MALDI MS to characterise the disulphide bond available for binding in an irradiated antibody.[226]

Analytical ultracentrifugation (AUC)

Analytical Ultracentrifugation uses the sedimentation equilibrium and velocity to estimate the mass, diffusion coefficient, and binding affinities of proteins and complexes. By rapidly spinning the sample, components of various mass and density will separate, with rates of separation dependent on the composition of the sample, and the interactions between its components. Computational models for the behaviour of the components in the field allow to estimate several of their properties.[227]

Terms	Description	Reference
Dynamic light scattering (DLS)	Light scatters on particles increases their diffusion, with larger particles diffusing more slowly. By observing this change in diffusion, the size of the particles can be inferred, as well as their aggregation propensity. The diffusion equation is dependent on temperature, and careful control of the temperature in the sample is therefore necessary. The sample otherwise requires very little adaptation, making this method applicable to a very large range of biomolecules, from small aggregating peptides to large biomolecular complexes, as well as protein-ligand interactions.	216
Fluorescence (or Förster) resonance energy transfer (FRET)	Fluorescent probes attached to the molecule provides distance information	214
Electron paramagnetic resonance (EPR)	Electron paramagnetic resonance provides information on the distance between atoms. By exciting the spin of unpaired electrons, the distance between atoms can be determined.	222
Circular dichroism (CD)	CD provides information regarding the secondary structure content of biomolecules. It can be resolved over time, allowing to monitor conformational changes.	

Raman spectroscopy	Raman spectroscopy can be used to determine the structure and dynamics of small molecules. The inelastic scattering of photons against the molecules in the sample, known as Raman scattering, generated by a laser results in a shift in the energy of the photons. This shift in energy provides data regarding the vibrational modes in the system.	226
Size exclusion chromatography-multiple angle light scattering (SEC-MALS)	After separating molecules using a size exclusion chromatography column, multiple angle light scattering is used to determine the mass and size of proteins, by determining the diffraction properties of molecules passing through with a laser probe and a light detector.	220
Analytical ultracentrifugation (AUC)	Analytical ultracentrifugation separates components by mass, and further information regarding their binding can be obtained by analytical modelling of the behaviour of the sample.	227
Surface plasmon resonance (SPR)	Surface Plasmon Resonance studies the adsorption/desorption of molecules on a surface by monitor the change in reflectivity of the surface, using an incident light. Information regarding the kinetics of adsorption desorption can be obtained.	223

Table 2: Summary of the biophysical methods commonly used with MS to obtain structural data.

3.2.6 Microscopy

Various microscopy methodologies are in use for the determination of cellular and protein structures. Most of those techniques provides shape information, akin to cryo-EM.

Micro-crystal electron diffraction (micro-ED)

Obtaining large crystals for protein and protein assemblies can be incredibly time consuming, and larger, more flexible assemblies may simply not be amenable to crystallisation on a large scale. [228–230] In micro-ED, a cryo-electron microscope is used to capture the diffraction patterns produced by small crystals in the samples. These patterns are then used to determine the electron density and with it the structure of the protein of interest, using computational methods similar to x-ray crystallography and standard cryo-EM single particle reconstruction. High resolution structures can thus be obtained for structures that would otherwise be unreachable. While the structures that can be solved for this method tend to be small, in some cases larger structures have also been obtained. As is the case for X-ray, the electron density is used less often than the coordinates of the model itself.

Atomic force microscopy

Atomic force microscopy (AFM) provides direct evidence for the structure of a sample by scanning across a surface using a probe. By measuring the force between the probe and the sample, the distance between them at a given point on the surface can be estimated, producing a map of the sample. The probe can also be used to estimate the force needed to pull away a molecule currently attached to the surface, which can also provide useful kinetic and structural insights: the shape of complexes and molecules adsorbed on the surface can be determined, and the change in force can be used to capture structural transitions. AFM has been combined with MS, to study the aggregation of amyloid fibrils,[159,217] to determine the size and composition of virus and protein complexes together with native-MS,[231,232] and to detect and characterise proteins at very low concentration.[233]

Fluorescence microscopy

By attaching a fluorescent probe to a protein of interest, important data can be gathered. Fluorescence microscopy is often used to determine whether two proteins are co-localised. The cellular localisation of a protein can also be obtained, and due to its non-invasive nature, time-resolved data can also be obtained. Data regarding its size and charge can also be gathered.[234] Fluorescence microscopy can also be used to determine the stoichiometry of protein components, by comparing the brightness of tagged proteins.[235] MS and fluorescence microscopy were used together to determine jointly the orientation and nature of antibodies patterned on a surface,[236] and to identify complex forming proteins.[237,238]

3.2.7 Genetic methods

Yeast-two hybrids have been a common method to establish whether two proteins form an interaction. By creating mutants of the proteins of interest, with an added domain that will result in the expression of a gene on contact between the two domains, it is possible to monitor whether the two proteins are interacting. Large protein interaction maps have been created by systematically modifying proteins in the genome of organisms of interest.[239] This type of information is incredibly useful, even if somewhat coarse, in integrative modelling: by establishing what proteins are likely in contact, the number of possible combinations that need to be explored to determine the most likely conformation of a complex can be vastly reduced. It may also lift ambiguity regarding the nature of subunits that are in contact.[240]

More precise information can be obtained by mutating specific residues in the protein sequence: site-directed mutagenesis provides a way to determine which residue, or ensemble of residues, are crucial to the stabilisation of an interface within a protein complex. This step itself can benefit from integrative methods, to decide which residues should be mutated, as it is generally impractical to test a very large number of mutants. More recently, work that combine large scale mutations with monitoring of the resulting phenotypic profile have been used to determine the architecture of protein complexes, showing the accuracy that can be reached with such methods.[241] These methods are often used in conjunction with structural MS methods, to relate structural changes in a protein conformation with its functional changes upon mutation,[242] or for validating the importance of a given structural interaction.[243]

3.3 Computational methods for data integration

The experimental methods described above provide diverse data regarding the structure, dynamics, and interactions of proteins and their complexes. To combine them effectively, computational methods are employed to generate models that are compatible with the data provided, while retaining reasonable stereochemical features. The greatest obstacle to this task is related to the extremely large number of possible conformations that a protein structure may have. A given experimental data set will generally be compatible with an ensemble of structures comprising many distinct conformations. To discern whether those conformations are occupied, and which ensemble might best fit the data, is a challenge which forms the subject of intense ongoing computational developments.

The common methods used for model generation, their advantages and drawbacks, and the way data from experimental protocols described above is harnessed to generate and improve models, are described below.

For the purposes of this review, modelling methods have been classified into three categories:

- Bayesian methods, where a probability is assigned to a given model, as well as its congruence to the data. Models are iteratively modified so as to increase both the confidence in the model, and its support by the data, by re-evaluating the match between them at each cycle.

- Direct constraint modelling with sampling methods such as molecular dynamics (MD), or Monte Carlo (MC), where the experimental data is represented as physical constraints (e.g. harmonic springs between particles), thereby generating models that automatically satisfy the constraints imposed.

- Modelling with filtering, where the experimental data is used to select models that best fit the data.

This broad classification is not restrictive: modelling pipelines will often combine methods belonging to those different categories, and sometimes cycle through them, to improve the quality of the generated models.

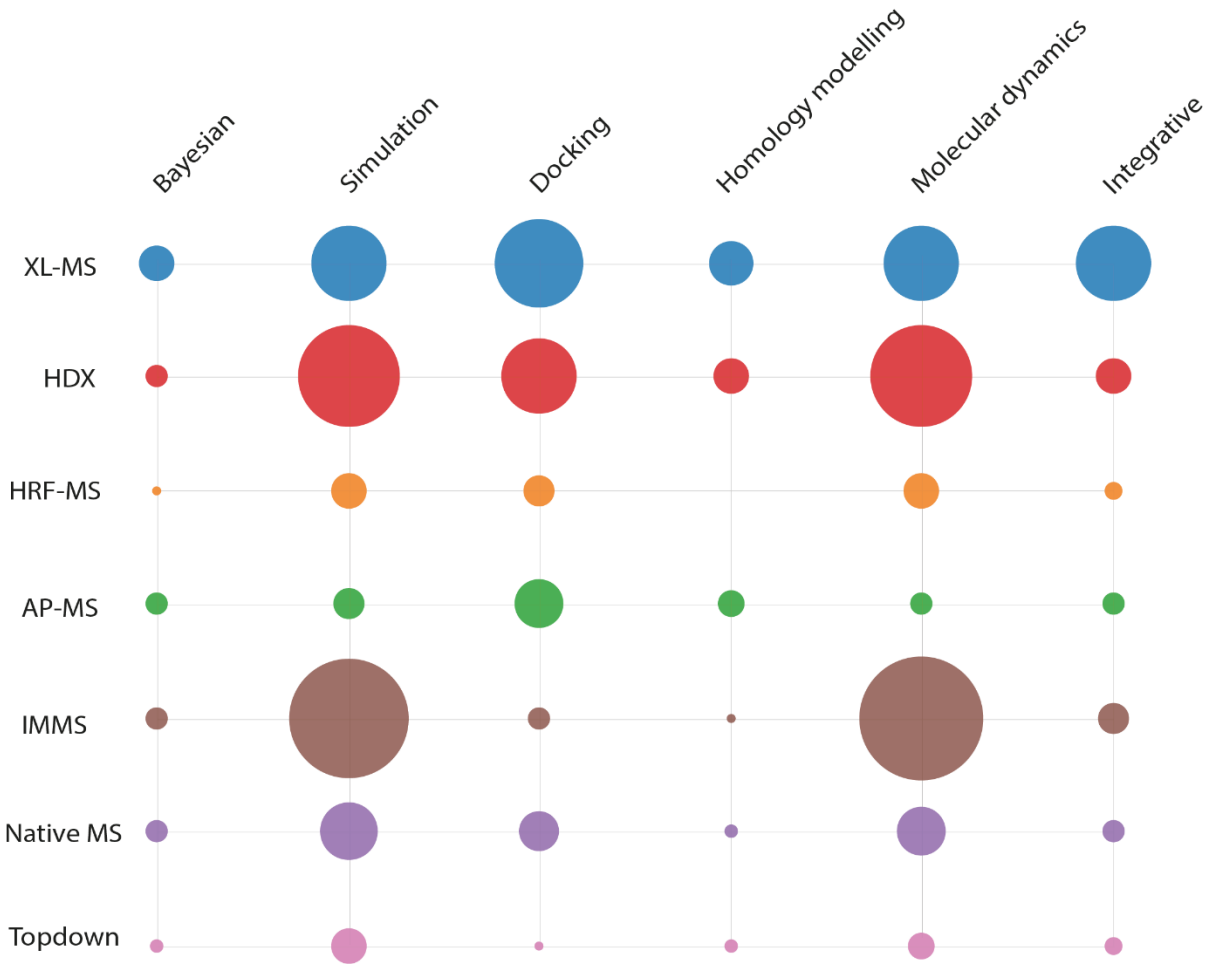

Figure 5: Plot showing the co-occurrence of computational and MS techniques, with larger dots corresponding to more common combinations.

In order to determine how these categories of computational methods pair with structural MS techniques, co-occurrence of the techniques was assessed, with the results shown in Figure 5. The computational methods presented in this Figure, further details of which can be found in Table 3, are expansions of the three previously defined categories of computational integration. From the first category, integrative and Bayesian are terms that refer to a type of approach rather than a specific modelling method, using multiple streams of experimental evidence for the former, while the latter characterises a type of probabilistic approach that attempt to rigorously incorporate new information on the probability of a given model as more data is provided. There are also a number of terms used in Figure 5 which fall into the direct constraint modelling category. Simulations, for example, is a very general term that may describe a wide range of such techniques. The simulations term includes MD and docking, both of which also appear as unique categories due to their popularity. The final defined category, modelling with filtering, is not explicitly featured in Figure 5. Whilst filtering of models is extremely common place in computational data integration, it was not easily amenable to detection

in the keyword analysis without being confused with other methods, and therefore was excluded from this search. Homology modelling is very specialised technique, that does not fall clearly into any of the above modelling categories, however it is included in the search due to its popularity.

Figure 5 shows how some experimental MS techniques are more naturally combined with specific computational methods. The distance restraints that XL-MS provide are naturally used in docking, and generally integrative modelling. On the other hand IM-MS and HDX-MS are more likely to be combined with MD simulation, which can be interpreted in term of conformational ensembles and flexibility. We will cover how those methods are combined below.

While the structural data obtained in the techniques described above can be used explicitly in an integrated pipeline to generate models, some of the information is often used in a more implicit manner, for example the stoichiometry of a complex may dictate how many units are used at the start of the pipeline, or for compounds of unknown structure, their sequence may be used to generate initial models.

Terms	Description	Reference
Bayesian inference	Given a distribution for the probability of a model given existing evidence, using Bayes' theorem, the probability for a model can be assigned, as well as the confidence in the evidence. This can be used to produce robust estimates of data and models together.	244,245
Simulation	A generic name for computational methods aimed at producing a physically realistic model of a system.	246,247
Docking	A method to determine likely configurations of a protein and a ligand or partner by sampling possible translation and orientation of the pair.	248–250
Homology modelling	A protein model building method whereby a model for the structure of a specific sequence is constructed using similar sequences that do have an experimentally obtained structure, that can be used to build a good initial model	251–253
Molecular dynamics (MD)	Simulation method based on the numerical integration of Newton's equation of motion; coupled with an energy function providing the physical description of the forces present in a system, this method will produce a trajectory of the system from a starting point.	254
Integrative modelling	A methodology whereby the modelling of a system is done using data from multiple experimental sources (see section 3), integrating them to help produce models.	255,256

Table 3: Definitions of common terms used in computational protein modelling.

3.3.1 Probabilistic/Bayesian methods

Bayesian methods, and in general probabilistic models, will attempt to generate or update models that best fit the available data, by estimating the likelihood of a conformation, given the data. These approaches are very powerful, but require a probabilistic model to be developed for the estimation of the likelihood based on the provided data. While physical intuition of the system and apparatus used to generated the data, as well as previously developed models, may be used to guide this process, it is nevertheless time-consuming and difficult. Bayesian methods do have the advantage of taking all information into account, irrespective of source, but balancing its importance based on the current model and the rest of the information. By reweighing the evidence, and not just the model, a spurious restraint is much less likely to prevent convergence towards a high-quality model.[257] Alternatives to Bayesian schemes which address some of these challenges have also been proposed.[258]

A Bayesian framework provides a general approach to represent and combine heterogeneous, uncertain data, to provide an estimate for a model's likelihood. An algorithm to generate and improve new models is still required. Likelihood maximisation schemes are powerful, but maximisation in a very large parameter space can be impractical, especially if many local optima exist (which will usually be the case for large proteins, where the structure may be folded in alternate yet similar conformations), slowing down the search. Variational optimisation may be necessary to ensure the

convergence of the method: when a large number of parameters are available, there may be several combinations of parameters that give results of similar quality, rendering optimisation difficult; this in turn vastly increases the search space for the method, as good parameters need to be found for the models with different parameter numbers. Direct modelling with MD or MC is a popular choice, with many implementations readily available, as we will see in the next section.

3.3.2 Using mass spectrometry data as constraints

The most common approach to model experimental restraints, especially coupled with MD, is to represent the experimental data as a physical force, forcing the system towards conformations that are compatible with the data. This may be a distance restraint between parts of the structure, as can be obtained by XL-MS, NMR or FRET, an envelope in which the structure should fit, as is obtainable by cryo-EM, or SAS, or an order parameter such as a CCS obtained by IM-MS.[259] These methods rely on generating models with detailed 3-dimensional coordinates for all parts of the system involved. The various experimental constraints are represented as forces acting upon the system; coupled with a forcefield, representing the known inter-atomic forces acting upon the atoms (i.e., electrostatic forces), this provides a way to gauge the likelihood of a given conformation. This approach can be used to generate models of complexes from existing high-resolution models of subunits (often known as docking), and enhanced by the addition of distance restraint in the form of crosslinks.[202,203]

By either integrating those forces, following Newton's equation of motion, as in MD, or by generating new trial conformations and comparing their energy, as is common in MC, various states of the system of interest are explored, and the best-fitting ones are retained. Both MD and MC schemes are straightforward to implement, in their simplest form, and offer a great deal of flexibility in order to tune and improve their sampling properties. Extensive literature on those sampling algorithms can be found elsewhere.[246,260] Additional sampling methods have been refined over the years, with variations around the original MD and MC methods designed to improve their sampling properties, genetic algorithms and particle swarm optimisations for example, or to allow for the sampling in alternate conditions (for example, constant pH dynamics).

Simulations also offer the possibility to compare experimental data against ensembles of models generated with the above methods: because the signal from experimental data is often averaged across time, or many copies of the system of interest, it is not necessarily the case that the observed signal would correctly fit to any single model but may fit to their overall average.[244]

Cryo-EM 3-dimensional density maps, estimated from the initial images, are used as a force in many modern fitting routines.[261–263] Extensive work has been done on optimising the form and details of the parametrisation of those forces to reach a better agreement between maps and models.[264]

Molecular modelling is often deeply biased by the models initially provided: finding a different, well matching model is difficult, due to the complexity of the energy surface for large proteins. Strategies to mitigate this issue exist, and to improve the sampling of these methods, however, convergence remain an issue. Due to the high cost of some of those sampling methods, confidence intervals based on repeating the sampling are often absent from studies. Since the sampling is generally stochastic, results may be dependent on the run.

3.3.3 Using mass spectrometry data as a filter

Finally, the experimental data can be used in a filtering step, after models have been generated, for example by selecting only models whose structure is compatible with experimentally determined cross-links. This method has the advantage of requiring no modifications to the computational pipeline besides the filtering step and does not bias its output.

Because the filtering step does not usually entail repeated evaluation of the scoring function, as is the case during sampling, the use of more expensive scoring function is possible here, either using a more complete set of data, or a more accurate comparison between data and model. This was used in nuclear pore complex studies, where more complete comparisons against experimental data was only performed during model filtering steps, after initial sampling.[196]

Another advantage of a filtering step is to avoid overfitting: by considering a different source of data than that used to generate models, overfit models, that match very well the data but are otherwise unphysical, or simply incorrect, can be rejected. By using crosslinking data, it is possible to significantly improve the performance of a cryo-EM assembly fitting pipeline.[265] Conversely, crosslinking data can be first used in a docking or MD approach, to generate models that can then be matched against other experimental data, such as SAXS or NMR.

Filtering is most effective when used with multiple independent sources of data: by combining a cryo-EM scoring function with XL-MS scoring, Bullock et al. showed significantly improved performance in evaluating the quality of protein complexes.[265] On the other hand, combining experimental data providing similar information may not necessarily improve the quality of the scoring.[211]

Filtering is often used as a final step in a modelling pipeline, to ensure that the previously generated model fit well with the data.[266] This approach can also be used for validation purposes: when multiple models of the same complex have been proposed, the use of an independent source of information to assess which models best match the new evidence can be a powerful tool to rank the quality and likelihood of models. As an example, cryo-EM 3D reconstruction can be validated by comparing their predicted CCS to experimental values obtained using IM-MS.[194]

3.3.4 Combining computational approaches

Each of the methods described above has advantages: direct modelling makes it easy to sample conformations likely given the data, while a Bayesian framework may provide more accurate estimates of the likelihood of a conformation. Those methods are often combined, for example by integrating a direct modelling step (with the data represented as forces) as a sampling step in a Bayesian framework. This is exemplified in an automated modelling pipeline using crosslinking data,[15] where a Bayesian framework is used to weigh the crosslinking data, but sampling of conformations proceeds using direct simulation with distance restraints. Alternating a sampling and filtering step is also commonly done in the determination of large assemblies, so that computational time is expanded on models that are likely to be a good fit to the data, discarding lower quality models.[266] For larger assemblies, where the complexity of determining a conformation is high, those steps can be repeated several times, and in several independent runs, to ensure convergence of the procedure.

3.3.5 Data and method standardisation

As demonstrated in the previous sections, large quantities of heterogeneous data can be produced while studying the structural biology of protein complexes. For MS data, a range of instruments are in common usage, each with different manufacturers and modes of operation. Furthermore, different MS techniques produce different types of data, and that combined with the variety of existing processing software means that output data can take many forms. As a result the field of structural MS has experienced challenges in depositing open access data for others to access and use. Whilst deposition in databases is becoming more prevalent, the actual data deposition scheme remains done on a case-by-case basis. Recently, for example, standards for data deposition have been proposed in some sub-disciplines, such as for HDX-MS, IM-MS and XL-MS data.[76,144,267] In other cases community efforts are starting to address this problem.[268]

Beyond the field of structural MS, data deposition has become routine for published data acquired using other experimental techniques. For example, 3D coordinates of atomic models obtained by X-ray crystallography are found in the protein data bank (PDB),[269] while proteomics data is often found on the PRoteomics IDEntification database (PRIDE).[270] Cryo-EM data is available via the electron microscopy data bank (EMDB),[271] with the underlying electron microscopy data available through electron microscopy public image archive (EMPIAR).

Data deposition in such public databanks, and in particular with the original dataset, is extremely useful. As computational methods improve, deposited datasets can be processed again, providing useful benchmarks for the quality of a given method, as well as sometimes improving the interpretation of this data. The PDB-REDO and cryo-EM re-refinement system (CERES) projects have shown how those efforts provide informative, sometimes new data on both computational processing, and the output generated by those methods, since the latter is not independent from the former.[272] The deposition of previous data does not simply allow to use improved methods, but sometimes a reinterpretation of previous evidence, in light of new data, to obtain models that are more accurate to the deposited data.[205]

Integrative modelling is a relative newcomer to the field: as modelling methods have become more complex, and integrate more data from many sources, it has now become apparent that resources are necessary to keep track of the protocol and information used. PDB-Dev was created for this purpose and provide links to other existing databases for the experimental data used in a given modelling project.

Name	Description	Reference
PDB	The Protein Data Bank, used for high-resolution atomic models	269
EMDB	Electron Microscopy Data Bank, for cryo-EM 3D reconstruction. Integrated with the PDB, with many EM reconstructions accompanied with a fitted model	216
SASDB	The small-angle scattering biological data bank, is a database of experimental small-angle scattering data together with models, and relevant experimental details.	273
PDB-DEV	Atomic force microscopy provides direct evidence for the structure of a sample by scanning across a surface using a probe. Information on the structure of the sample can thus be obtained,	274
EMPIAR	Raw EM data, before 3d reconstruction	275
PRIDE	Proteomics and mass spectrometry databank	270

Table 4: Summary of current deposition services for structural biology data.

Many structural studies are accompanied with methodological developments, necessary to interpret the data. Those methods are generally implemented in programs, while the article text provides an overview of the steps involved. Unfortunately, it is often difficult, if not outright impossible to re-use or re-run these programs. Some code is simply not made available with the article, while in other cases the links that are provided end up deprecated. Change in both hardware and software can render programs near-unusable too. The spread of data deposition practice, and the ease of use of modern online code repositories has made it more common to find those method implementations, and potentially reuse them. The heterogeneity of starting data, as well as the lack of standard practices and formats, makes it difficult to rerun and compare different methods. As discussed in previous sections, such data reprocessing effort can be incredibly useful, providing large benefits to the community, and better standardisation in methodological developments and reporting would help in this respect.

4. Systems benefitting from integrative structural biology

Large, dynamic assemblies require a combination of the techniques described previously to be understood. Two of the largest, most important protein assemblies in the cell are the proteasome and nuclear pore complex. Both have been the subject of repeated integrative studies, that have revealed more and more information about their structure and function and have stimulated the development of new experimental and computational developments in integrative modelling. We will describe those developments, focusing on studies that made use of MS.

4.1 Proteasome

The proteasome is a large and heterogeneous protease complex which controls selective degradation of undesirable proteins.[276–282] This system forms the major mechanism for removal of damaged, misfolded, and regulatory proteins in eukaryotes, and therefore is of significant interest in biology.[283–285] Architecture of the standard proteasome complex is now known to include a catalytic core, referred to alone as the 20S proteasome or 20S core particle. This core contains four stacked rings, arranged as two β rings sandwiched between two outer α rings, with each ring comprising seven subunits.[286–289] In its enzymatically active form, termed the 26S proteasome, either one or two 19S regulatory particles attach terminally to the 20S core through interaction with the α rings. The 19S regulatory particles themselves consist of multiple subunits, which are grouped into the base region which interacts directly with the 20S core, and the lid. This description characterises the accepted view of the standard proteasome, however it is widely acknowledged that the system has additional heterogeneity. Examples include variation in 20S core subunits giving rise to immuno-, thymo-, and spermato-proteasomes, binding of alternative interactors such as PA28 in place of the 19S regulatory particles, and PTM of subunits.[290–295]

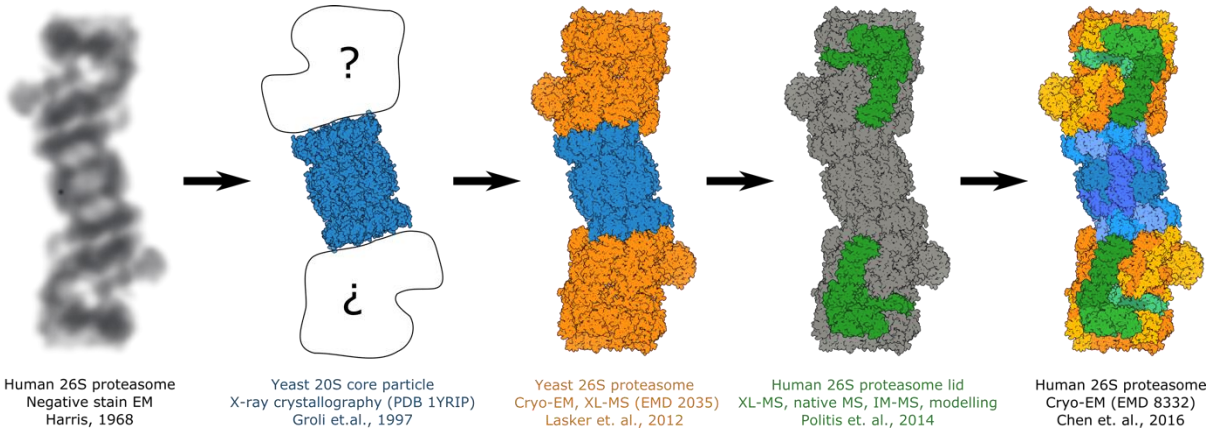

Figure 6: Evolution of structural knowledge for the proteasome. Initial negative stain EM revealed a barrel-like structure with a lid. The core proteasome was resolved by X-ray. An integrative approach, combining cryo-EM, XL-MS and existing knowledge regarding the proteasome structure was used to produce a model of the full proteasome. This model was later largely confirmed by cryo-EM studies resolving the full proteasome. Future work remains, regarding the kinetics of the proteasome complex, and its interaction with other cellular components.

Functional insights into the proteasome, including its close relationship with ubiquitination enzymes and its role in cellular homeostasis, have been the subject of numerous insightful studies.[282,296–298] In parallel with advances in the functional understanding of the proteasome, considerable focus has been given to building a structural understanding of the complex.[278,299] Structural studies in this area have proved challenging due to the complex and heterogeneous nature of proteasome components,

combined with its intrinsically dynamic landscape. As a result, structural characterisation of the proteasome is one of the best examples of where multiple structural biology techniques have been used, often in an integrative fashion, in order to advance the field.

Structural interrogation of the proteasome can be traced back to a seminal study from 1968, in which the first image was recorded showing a protein complex of unknown function.[300] This image, captured by negative stain EM, is now known to feature the 20S core particle. To this day, negative stain EM, proves an excellent technique for the structural visualisation of very large protein complexes and a useful tool in the integrative structural biology arsenal. The next structural steps in resolving the proteasome took place in the 1990s, when a series of x-ray crystallography studies provided atomic structures of the 20S core from archaeal and yeast.[301,302] These developments proved possible due to the relatively stable structure of the core particle compared to the 19S regulatory units, facilitating successful crystallisation. Given the dynamic nature of the regulatory particles of the enzymatically active 26S proteasome, the next big developments in proteasome structure came as a result of EM. Developments in this area successfully provided the molecular shape of 26S proteasome from *Drosophila* and *S. pombe*, however the resolution of these structures in the 19S region prevented subunit atomic co-ordinates being determined.[303,304]

Despite the developments in proteasome structural biology discussed up until this point, a complete structure with subunit organisation of the intact enzymatically active 26S proteasome still remained elusive.[305] As such, several research groups turned to integrative methods in hopes of solving the problem, combining cryo-EM data with chemical biology or structural MS approaches.[306–308] One example of this is the 2010 publication from Bohn *et. al.*, in which XL-MS was used to support cryo-EM in order to obtain a 9.1 Å resolution structure of the 26S proteasome from S. pombe.[309] In this case, the spatial restraints obtained from XL-MS provided information on the proximity between subunits, which was vital for fitting the density map obtained by cryo-EM. Similarly, a 2012 publication from Lasker *et. al.* built a pseudo atomic model of the entire 26S proteasome from yeast by using an approach with even greater structural biology integration.[310] The approach combined experimental and computational tools, taking complementary data from two yeast strains. The method was outlined in a four step workflow, which will we discussed in brief. The first step described involves gathering information from five structural biology approaches for use in structure determination: (i) a density map of *S. pombe* proteasome obtained at 8.4 Å from cryo-EM; (ii) distance restraints within the regulatory particle from XL-MS of S. pombe; (iii) cryo-EM density maps of 26S proteasome from subunit deletion strains of S. cerevisiae; (iv) the regulatory particle interactome derived from a combination of AP-MS, XL-MS and two-hybrid; (v) atomic structures of regulatory particle lid subunits from x-ray crystallography. These experimental approaches were next translated into spatial restraints for further use. Using these restraints, coarse grained representations were developed into refined models in three steps: (i) subunit localisation was performed based on XL-MS interaction restraints and cryo-EM density maps; (ii) models computed for regulatory particle lid subunits were fitted into the cryo-EM density map; (iii) MD flexible fitting was used to refine atomic co-ordinates of subunits. Finally the best scoring models based upon agreement with both atomic models and XL-MS distances were clustered, ultimately giving localisation probabilities for regulatory particle subunits.

Similarly to the ground-breaking steps forward made by Lasker *et. al.*, integrative approaches were also trailed to obtain structural definition within the lid of the human 19S regulatory particle.[310] Research carried out by Politis *et al.* successfully tackled this problem by integrating data from three structural MS techniques with computational modelling, which will now be discussed in brief.[311] The first step in their hybrid approach was to purify and isolate the 19S lid, for which an AP-MS approach was adopted. In brief, the entire 26S proteasome was first affinity purified from RPNX-3xFLAG cells before being exposed to high salt conditions to promote dissociation. Tagged lid subunits were then eluted and enriched, before MS and label free quantitation which confirmed successful enrichment of

the proteasome lid. This approach also confirmed protein identities, providing an overview of proteasome lid composition which is a vital step underpinning additional integrative approaches. Native MS was then performed on the affinity purified samples, identifying the intact lid and in doing so confirming its composition and the stoichiometry of protein subunits.[312] Native MS further revealed subcomplexes within the proteasome lid, notably one containing Rpn5/8/9/11 and a second consisting of Rpn3/7/Sem1. IM-MS experiments with CCS calibration were also performed on the intact lid and its subcomplexes, building on the native MS data in providing a conformational envelope for each complex. XL-MS was performed with DSS leading to identification of both interprotein and intra-protein interactions which support the proteasome lid subcomplexes identified by native MS data. High quality crosslinks, defined as those scoring an false discovery rate (FDR) of <0.05 in xProphet analysis and confirmed by manual inspection, were also identified within the XL-MS dataset.[313] These crosslinks provided distance restraints between residues within the identified proteasome lid subcomplexes. The described composition and stoichiometry information obtained from native MS and AP-MS were then used as subunit input for a MC search, which sampled the conformational space to generate over 10,000 models. The models were then scored using restraints based on IM-MS conformational envelope, native MS subcomplex characterisation, and XL-MS connectivity. Energy minimisation MD simulations were then applied to further prioritise structures. The top 1% of highest scoring models from this integrative MS and computational method were overlayed with a high resolution cryo-EM density map containing all proteasome lid subunits with the exception of smallest subunit Sem1. The model proved to be in agreement with the EM map, with a cross correlation coefficient of 0.74, and included additional information regarding Sem1 subunit placement.

High resolution structures of the entire human 26S proteasome were obtained between 2015 and 2018, following on from the integrative studies just described.[314–317] These cryo-EM data revealed six co-existing conformations of the 26S proteasome, at resolutions between 3.6 Å and 6.8 Å. The high resolution human 26S proteasome structures were closely followed in subsequent years by comparable cryo-EM conformations of the yeast 26S proteasome.[318] Since these evolutions, attention within the field of proteasome structural biology has begun to focus on additional features of the complex. New areas of interest include studying proteasome interactions and the structural changes they induce, such as native substrate binding, and targeting by drugs. Furthermore, greater attention is now being given to probing the dynamics and kinetics of the proteasome, to identifying structural assembly intermediates, to detailed characterisation of subunit composition, and to proteasome variation between cell lines.[278,295,319–321] Structural MS is making its presence known in these endeavours, with top-down MS facilitating characterisation of proteasome subunit PTMs, IM-MS probing core particle shape and stability, AP-MS identifying binding partners, and HDX-MS revealing differences between the 20S and immuno-20S core.[132,322–328] At the time of writing, however, MS techniques such as HRF-MS, remain open for further exploitation in this area. The renewed focus of the field on interactions and dynamics makes structural MS a perfect tool as these avenues are pursued. Furthermore, as discussed throughout this review, great power comes from combining insights from MS with other structural biology techniques in an integrative fashion. It is our belief that this approach will be key in providing a detailed picture of the proteasome going forward.

4.2 Nuclear pore complex

The nuclear pore complex (NPC) is a large protein assembly which is found on the nuclear envelope, and is the central element for mediating molecular transport in and out of the cell nucleus in eukaryotes. Its regulation is key to cellular function. Composed of over 30 different proteins, most present in several copies, the NPC is a very large and complex structure, whose elucidation has been the subject of intense study. The NPC is also a dynamic assembly, with a number of proteins binding conditionally to it. The complex as a whole can undergo a number of structural transitions, whilst portions of the proteins constituting it are also individually flexible. As such, structural biology of the NPC is complicated, and made even more so by the fact that its composition and structure vary greatly

between organisms. To date, the NPC structure has been partially resolved by integrative modelling techniques, using a combination of data from X-ray, cryo-EM, XL-MS, and other methods.[329] These studies have revealed that the overall NPC structure comprises three rings: nuclear, spoke, and cytoplasmic.[330] A coat extending beyond the nuclear and cytoplasmic ring is also present.[331]

Early work studying NPC structure focussed on obtaining information about the relative abundance of protein constituents within the complex. NPC proteins were successfully identified and quantified using an AP-MS approach. These fundamental studies were required to underpin much of the subsequent structural biology research carried out on the NPC. Having determined the NPC constituents, a model of the intact complex was necessary in order to accurately place proteins at their correct location within the NPC. The overall shape and space occupied by the non-flexible part of the complex was determined by negative stain EM.[332] Information from the other techniques, such as XL-MS, were then combined to fit the units of this complex together within the EM envelope. Unfortunately, the contents of this envelope could not initially be attributed to specific proteins, and the resolution did not directly allow exact determination of protein structures. However, this map did allow further work to be combined and validated, by reusing this envelope data to dock subcomplexes whose organisation was determined with much greater certainty. For example, subcomplexes were later purified and resolved with cryo-EM, and in conjunction with a higher-resolution envelope for the NPC, allowed several subcomplexes to be placed within the overall map.[330]

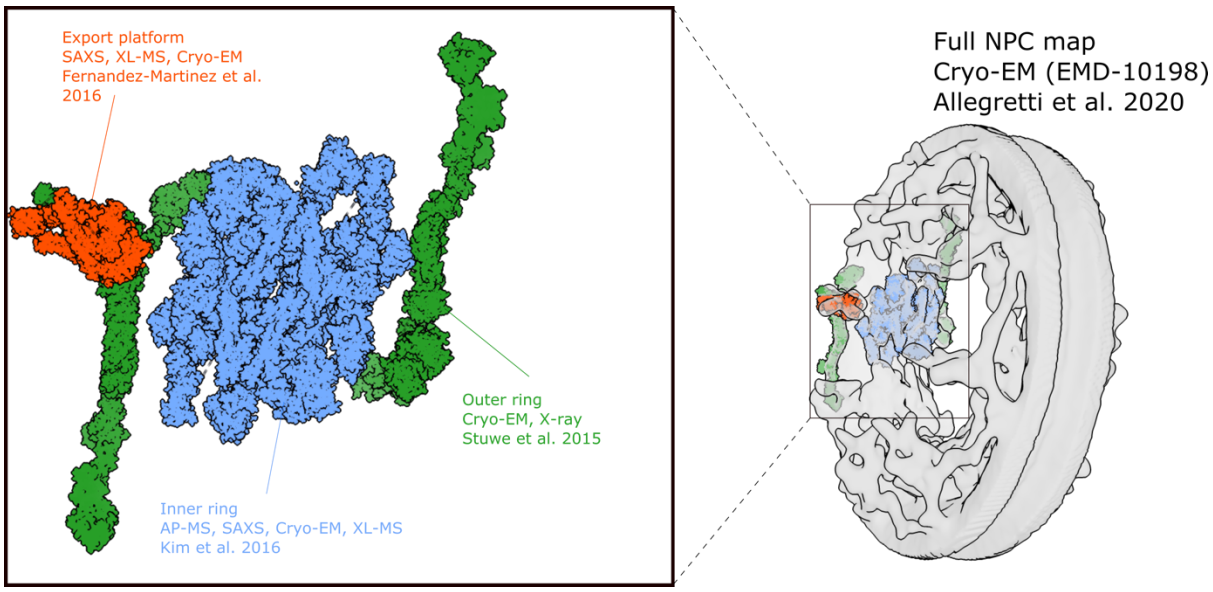

Figure 7: Representation of several components of the nuclear pore complex, and the integrative techniques that were used to resolve them. A full model for the NPC was proposed in Rout *et al.*[333]

Higher resolution X-ray structures for the coat Nup complex were later obtained and docked within the envelope, providing a higher-resolution view of this region of the NPC.[331] This approach also proved useful for the mRNA export platform assembly, which was determined and similarly docked into the overall NPC envelope[329] by combining data from negative stain EM and restraints from XL-MS, and computational sampling of possible conformations, a high-quality model that fit the obtained envelope was obtained. The authors employed a protocol that uses both sampling and filtering, as described above.[266] Crosslinking data was used to determine the topology and orientation between subunits in the complex. This demonstrates the importance of both “classical” MS techniques and more modern structural methodologies, which were together crucial to the determination of the overall mRNA export platform.

Finally, a full sub-nanometre structure of the entire *S. cerevisiae* NPC was presented. After confirming the stoichiometry of all components in the NPC by quantitative MS, the overall complex was assembled using cryo-EM, XL-MS, and SAXS. Coarse-grained representations of all the subunits present in the complex were determined using a combination of existing X-ray and NMR models from the PDB, predicted models were used for model lacking an experimentally solved structure, and SAXS profiles were used to validate those models. The XL-MS and cryo-ET data were used as restraint in a MC simulation. The resulting sampling was assessed for convergence and was then filtered to select high-quality models. Those models were then assessed against the data, and a further MD run (using Brownian dynamics) was performed to estimate the distribution of position of the Phe-Gly (FG) repeats extending from nucleoporins, and their interaction with nuclear transport factors (NTF).[196] The resulting NPC structure is a prime example of truly integrative work, with multiple mass spectrometric methods, biophysical experiments, and computational methodologies combined to produce a high-quality fit for a very large complex, that would not have been possible to achieve with a lesser combination of any of those techniques.

More recent work on the yeast NPC has shown this work can be extended and provide further information regarding the changes in such a large complex. By combining *in situ* fluorescence microscopy with cryo-electron microscopy, the authors could confirm whether components of the NPC could be successfully integrated in the NPC assembly, and derived information regarding the effect of a conformational change of Nup159 on the assembly dynamics of the larger NPC. With improved knowledge of the NPC structure, and improved methods, it is becoming possible to study larger cellular assemblies: In-situ cryo-electron tomography has also been used to show that the proteasome and the NPC, themselves still the subject of much study, have been shown to tether.[334]

5. Conclusions & future directions

5.1 New developments for integrative structural biology

5.1.1 Mass spectrometry developments

Developments in MS instrumentation, sample preparation, and data processing software have been fundamental in building the field of structural MS into what it is today.[268,335,336] Research focus on development in these areas remains strong, and it is therefore predictable that further advancements in the field of MS will continue to improve the technique's contribution to structural biology. One such development is in the field of IM-MS, with the expansion of tandem ion mobility approaches providing additional functionality compared to classic IM-MS.[337] Tandem DTIMS devices pioneered by the Smith group were among the early work in this field, initially being applied to enhance peptide fragmentation.[338–340] Building on these instruments, the tandem ion mobility methodology was commercialised in 2019 with a cyclic TWIMS IM-MS instrument.[341] The cyclic IM-MS platform is based upon the principles of several fundamental studies of tandem IM-MS, and as such facilitates both increased ion mobility resolution and conformer isolation. These abilities have already been proved as applicable to the study of native gas phase protein standards including Cytochrome C and Concanavalin A.[342,343] A further development in the field of tandem ion mobility is that of structures for lossless ion manipulations (SLIM) and its later analogues serpentine ultralong path with extended routing (SUPER) SLIM and multi-level SLIM.[344–347] These SLIM techniques are based upon travelling wave ion mobility, and allow ultrahigh resolution mobility separation of native protein conformations, ion trapping, and ion selection. Finally, trapped ion mobility spectrometry (TIMS) has also been applied in a tandem ion mobility workflow, facilitating both mobility selection of species, and ion activation.[348] These features have proved particularly beneficial in improving peptide fragmentation, increasing sequence coverage in bottom-up proteomics experiments.

Beyond MS developments in the area of IM-MS, advancements have taken place across other structural MS techniques. Examples of this include the development of more powerful data processing platforms, and application of the structural MS toolkit to study challenging biomolecules beyond proteins, such as G-quadruplexes of DNA.[27,28] In the field of top-down MS, novel fragmentation techniques are opening up improved characterisation of biomolecules. The application of SID to native protein complexes has, in particular, greatly improved the information on quaternary protein structure obtained within the field.[349–352] HDX-MS analysis is moving towards amino acid level resolution with new automated instrumentation allowing analysis on the millisecond timescale, as evidenced for glycogen phosphorylase, where the technique revealed the entropic mechanism of active site access.[353] Finally, the evolution of higher-resolution charge detection-mass spectrometry (CD-MS), is enabling the analysis of larger protein complexes than previous MS methodologies would allow. This single molecule MS approach has to date been applied to monitor amyloid fibril aggregation and viral assembly.[354–356] As with all structural biology techniques, improving the ability of MS to analyse protein structure within the native cellular environment is a key area of development, and will be discussed in detail in the following section.

5.1.2 Moving structural biology towards in cell analysis

Structural biology has to date been a field largely focussed on the study of purified *in vitro* protein systems, due to the increased challenges associated with performing analyses within the complex cellular environment. To study proteins in their true form, however, it is vital to capture the native structure, dynamics and biomolecular interactions of a protein *in vivo*. As a result, several structural biology techniques are moving towards in cell analysis. Correlative light and electron microscopy techniques allow the determination of large structures within a sample, from the NPC to entire cells.[333,357,358] Structural MS, and the peptide-based techniques in particular, are perfectly placed to deal with the problems of working within the complex native cellular environment. AP-MS, for example, has long been applied in a comparative fashion to proteins in cell, revealing their native interactome.[37,39–41] Impressively, XL-MS has also been making promising step forwards with in cell analysis. Kaake *et. al.* successfully developed an *in vivo* XL-MS platform to identify protein interactions in living cells, whilst the Bruce lab have applied their PIR technology to tissue systems including the mitochondria and mouse heart tissue.[65,67,322] The FPOP derivative of HRF-MS has had similar in cell success over recent years, with a 2019 paper from Espino and Jones applying technique *in vivo* to determine protein solvent accessibility in *C. elegans*.[110–112] Given the prevalent use of *C. elegans* as a model organism, development of this workflow has exciting potential for the future study of disease mechanisms. The development of *in vivo* structural MS, however, is not limited to the peptide-based techniques, as developments have also been made in an attempt to bring protein-based methods into cells. Focus in this area has largely been in the sample preparation arena, such that the methods are theoretically applicable for use with native MS, IM-MS and top-down MS. For example, the Sharon lab have pioneered analysis of overexpressed proteins from crude cell lysate using native MS, having success in both prokaryotic and eukaryotic systems.[359] In parallel, the Cooper lab have developed a native LESA workflow, which draws upon the principles of MS imaging to directly extract proteins across a tissue surface, retaining their native structure and interactions for analysis.[131] The native LESA technique has more recently been successfully combined with TWIMS IM-MS for the study of mouse kidney tissue samples.

5.1.3 Latest developments in machine learning

Machine learning methods are becoming increasingly popular, with new approaches being published regularly at the present time. By harnessing large datasets, machine learning allows the user to train predictors that will detect data features and combine them to predict a quantity of interest. With a large enough dataset, these methods are extremely powerful, since high-accuracy predictions can be

achieved without an accurate underlying physical model. Those also have drawbacks, since statistical methods can not invent data out of thin air: biases in the training dataset may prove difficult or impossible to overcome and use cases that are not well covered in the dataset will often be badly predicted. Nevertheless, machine learning methods are becoming more popular to process MS data, with newer methods able to provide high quality predictions for a number of tasks,[360,361] including MS peptide sequencing,[362–366] protein identification,[367–369] or CCS predictions.[370] On the structural modelling side, modern protein prediction methods are now achieving extremely accurate results, with state-of-the-art methods using deep neural networks (notably AlphaFold and RoseTTAFold) able to predict full protein structures at atomic resolution.[371–374] Despite the impressive results achieved using these approaches, flexible proteins and larger assemblies, in particular, will remain challenging. As such, we believe that better integration with experimental MS data will still play an important role in understanding the structure and dynamics of those complexes, in particular for cases where a single folded state does not represent the majority conformation.[375]

5.2 Getting more from current approaches

Throughout this article, we have discussed several methodologies and examples showing successful implementation of computational integration of MS data for structural biology. In general, these integrative workflows have relied on only a subset of the structural data which can be derived from MS techniques discussed in Section 2.3. Fruitful integration strategies have been developed to make use of information such as composition, stoichiometry, interface and distance restraints. However, at the time of writing, structural MS insights into conformational changes, dynamics, and kinetics are considerably less used in this fashion. Given that this is a core strength of many of the structural MS methodologies discussed, to leave them out of the structural biology toolkit creates a great deficit in information. The reticence towards integrating these forms of structural information may derive from struggles to mesh the fluid and flexible nature of these findings with the relatively static high-resolution images resulting from classic structural biology techniques. We argue, however, that since the native cellular environment of proteins is a fluid and dynamic system, these insights are vital to truly develop an in-depth understanding of protein structure and function. In addition, data on protein PTMs and proteoforms obtained from structural MS are also currently left out of integrative structural biology approaches. This finding is perhaps surprising, given that modification of proteins has long been intricately linked to folding, structure, and regulatory function. As the field of integrative structural biology moves forward, it must focus on how to incorporate these forms of structural data along with the classic techniques in order to build a truly comprehensive picture of the protein landscape. Better use of the data generated in structural MS is another avenue for progress: improved deconvolution and detection methods for the analysis of mass spectra can produce additional and more accurate data.[376] This also holds true for other structural methods, where algorithmic developments are contributing to improvement in the structural data generated. Algorithms for integrative modelling are also the subject of ongoing development, in particular for probabilistic approach combining heterogeneous data sources.

Beyond making greater use of the structural MS techniques discussed thus far, several additional structural MS tools exist from which insightful data for the purposes of integrative structural biology can be obtained, but which at this time are less widely used and rarely integrated. This set of techniques fall into the category of peptide-based methods, with many of them being derived from research groups with expertise in bottom-up proteomics. A subset of these techniques, and the structural information which can be obtained from them will be described. Cellular thermal shift assay paired with mass spectrometry (CETSA-MS) performs multiplexed bottom-up proteomics on soluble proteome fractions over a set of stepwise temperature increases.[377–379] This approach can provide information with potential for integration regarding protein dynamics and kinetics, as well as protein complex composition and stoichiometry. A second structural MS technique with integration potential is limited proteolysis coupled to MS (LiP-MS), in which the enzymatic digestion step of bottom-up

proteomics is two-fold.[380,381] The first digestion is performed under native conditions using a non-specific protease, and is followed by a second trypsin digest under denaturing conditions. LiP-MS therefore provides information on which protein regions are externally exposed within the native conformation, and in particular highlights variations in this structure due to ligand binding or external stimulus. A further technique is a covalent labelling method with close similarity to HRF-MS, but instead of using hydroxyl radical labelling, specific diethylpyrocarbonate (DEPC) labelling is used.[382] DEPC labels solvent exposed Cys, His, Lys, Thr, and Ser residues with a known mass, detectable by MS analysis, and therefore gives a measure of solvent accessibility for a protein structure. Two further techniques, which have found particular favour for the study of IDPs, also have their roots in the principles of covalent labelling, and in XL-MS.[29] These techniques, termed ligand footprinting-mass spectrometry (LiF-MS) and tag transfer XL-MS, both use diazirine chemistry to characterise biomolecular interactions.[383,384] The final technique to discuss is proximity labelling-MS, in which a protein of interested is expressed with a tag which, when activated, is able to promiscuously biotinylate other protein molecules within a known distance. There are several tags developed for this purpose, with the most common being APEX, APEX2, BioID and BASU.[385,386] Following biotinylation, modified proteins can be enriched and easily identified using MS, thus providing structural information on the protein interactome even within a complex cellular environment. The method and insights obtained from this technique parallel those of AP-MS, and as such proximity labelling-MS is in some cases included within this category, however, its ability to determine proximity in addition to direct interactions give the method additional integrative applications above AP-MS. It is acknowledged that several other structural MS techniques exist or are currently in development in addition to those discussed here. It is hoped that by highlighting these techniques within this review, their potential for further integration will be recognised going forward.

5.3 Final remarks

The development of soft ionisation methods ushered in a new era for MS in biological research. While initially used predominantly to interrogate the primary structure of proteins, many instrumental, sample preparation and computational developments gave rise to the structural MS methods described in this review. As already discussed, collectively these methods provide unique and complementary information regarding the structure, and most importantly, the dynamics of proteins and their complexes. Such information is not easily, or at all, obtainable by other established structural and biophysical methods, but it is nevertheless crucial for understanding how proteins function.

The last few years have seen exciting new developments in other structural biology techniques, and also in computational methods to integrate diverse pieces of data, an area we expect to experience further growth. As developments in structural MS, other structural biology techniques, and computational methods coalesce, we will move closer to probing the structure and dynamics of proteins on a proteome-wide scale and in their cellular environment. Such a prospect will truly unravel the elegant dance protein molecules perform within cells in order to carry out their biological function.

Author information

Hannah Britt graduated from Durham University in 2014 with an MSci degree in Natural Sciences (Biology and Chemistry). She remained in Durham to complete her PhD, where she used an integrative biophysical approach to probe reactivity at the membrane interface, under the supervision of Dr John Sanderson and Dr Jackie Mosely. In 2018 Hannah joined the Thalassinos lab in the Institute of Structural & Molecular Biology at University College London as a post-doctoral researcher, where she uses structural mass spectrometry to study the secondary envelopment of human cytomegalovirus. She has a particular interest in native and ion mobility-mass spectrometry, and their application to the analysis of disease-relevant glycoproteins.

Tristan Cragnolini obtained a Bsc in biochemistry from Université Paris-Est Créteil, followed by an Msc in Bioinformatics from Université Paris 7-Paris Diderot. He completed his PhD on the coarse-grain folding of nucleic acids under the supervision of Prof. Samuela Pasquali and Prof. Philippe Derreumaux. After a post-doctoral position under Prof. David Wales, he joined the Thalassinos and Topf labs at the Institute of Structural & Molecular Biology at University College London and Birkbeck College London, where he uses computational modelling to interpret cryo-EM and mass spectrometry data of biomolecules in terms of structural models.

Konstantinos Thalassinos is Professor of mass spectrometry at the Institute for Structural and Molecular Biology (ISMB), with a joint appointment at UCL and Birkbeck. He is also the Academic Lead of the Mass Spectrometry Science Technology Platform at UCL. He obtained his BSc in Genetics from the University of Leicester, followed by a Master's in Bioinformatics from the University of York. His PhD combined experimental and computational studies in mass spectrometry-based proteomics which he undertook at the University of Warwick under the supervision of Prof. Jim Scrivens. In 2010 he moved to the ISMB to take up a lectureship in mass spectrometry. In 2015 he was promoted to Senior Lecturer and in 2019 to Professor of mass spectrometry. His lab is using structural mass spectrometry approaches, especially ion mobility and crosslinking, to study the structure and dynamics of proteins and protein complexes, particularly those involved in protein misfolding diseases.

Supporting information

A single supporting information document is supplied with the manuscript. This file contains supplementary methods for the extraction and processing of data displayed in Figures 2-5 of this review.

Acknowledgements

H.B and T.C. are supported by the Wellcome Trust (209250/Z/17/Z). Mass spectrometry instrumentation in K.T.’s lab has been supported by Wellcome Trust grants 221521/Z/20/Z and 104913/Z/14/Z.

Abbreviations

AFM	Atomic force microscopy
AI-ETD	Activated ion electron transfer dissociation
AP-MS	Affinity purification-mass spectrometry
ATD	Arrival time distribution
AUC	Analytical ultracentrifugation
BS3	Bis(sulfosuccinimidyl)suberate
CAD	Collisionally activated dissociation
CCS	Collision cross section
CD	Circular dichroism

CD-MS	Charge detection-mass spectrometry
CERES	Cryo-EM re-refinement system
CETSA-MS	Cellular thermal shift assay-mass spectrometry
CID	Collision induced dissociation
Cryo-EM	Cryo-electron microscopy
CTF	Contrast transfer function
D_2O	Deuterated or “heavy” water
DEPC	Diethylpyrocarbonate
DLS	Dynamic light scattering
DSBU	Disuccinimidyl dibutyric urea
DSS	Disuccinimidyl suberate
DSSO	Disuccinimidyl sulfoxide
DTIMS	Drift tube ion mobility spectrometry
ECD	Electron capture dissociation
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EID	Electron ionisation dissociation
EMDB	Electron microscopy data bank
EMPIAR	Electron microscopy public image archive
EPR	Electron paramagnetic resonance
ESI	Electrospray ionisation
ETciD	Electron transfer dissociation/collision induced dissociation
ETD	Electron transfer dissociation
EThcD	Electron transfer dissociation/higher energy collisional dissociation
FDR	False discovery rate
FPOP	Fast photochemical oxidation of proteins
FRET	Fluorescence (or Förster) resonance energy transfer
FTICR	Fourier transform ion cyclotron resonance
HCD	Higher-energy collisional dissociation
HDX-MS	Hydrogen deuterium exchange-mass spectrometry
HRF-MS	Hydroxyl radical footprinting-mass spectrometry
IDP	Intrinsically disordered protein
IM-MS	Ion mobility-mass spectrometry
IRMPD	Infrared multiphoton photodissociation
LC	Liquid chromatography
LESA	Liquid extraction surface analysis
LiF-MS	Ligand footprinting-mass spectrometry
LiP-MS	Limited proteolysis coupled to MS
m/z	Mass to charge ratio
MALDI	Matrix-assisted laser desorption ionisation
micro-ED	Micro-crystal electron diffraction
MC	Monte Carlo
MD	Molecular dynamics
MS	Mass spectrometry
MS/MS	Tandem MS
MS^n	Multi-stage MS
nESI	Nano-electrospray ionisation
NHS	*N*-hydroxysuccinimide
NMR	Nuclear magnetic resonance
NPC	Nuclear pore complex
NTF	Nuclear transport factors

OH•	Hydroxyl radical
PDB	Protein data bank
PRIDE	PRoteomics IDEntification database
PTM	Post-translational modification
QToF	Quadrupole time of flight
SAS	Small angle scattering
SASDB	Small angle scattering biological data bank
SANS	Small angle neutron scattering
SAXS	Small angle X-ray scattering
SCX	Strong cation exchange
SEC	Size exclusion chromatography
SEC-MALS	Size exclusion chromatography-multiple angle light scattering
SID	Surface-induced dissociation
SILAC	Stable isotope labelling by amino acids in cell culture
SLIM	Structures for lossless ion manipulations
SPR	Surface plasmon resonance
SUPER	Serpentine ultralong path with extended routing
TIMS	Trapped ion mobility spectrometry
ToF	Time of Flight
TWIMS	Travelling wave ion mobility spectrometry
UVPD	Ultraviolet photodissociation
XL-MS	Crosslinking-mass spectrometry

References

(1) Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88*, 4451–4459.

(2) Fenn, J. B.; Mann, M.; Meng, C. K. A. I.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionisation for MS of Large Biomolecules. *Science* **1989**, *246*, 64–71.

(3) Karas, M.; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Anal. Chem.* **1988**, *60*, 2299–2301.

(4) Biemann, K. Mass Spectrometry of Peptides and Proteins. *Annu. Rev. Biochem.* **1992**, *61*, 977–1010.

(5) Biemann, K. Contributions of Mass Spectrometry to Peptide and Protein Structure. *Biomed. Environ. Mass Spectrom.* **1988**, *16*, 99–111.

(6) Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom.* **1984**, *11*, 601–601.

(7) Stollar, E. J.; Smith, D. P. Uncovering Protein Structure. *Essays Biochem.* **2020**, *64*, 649–680.

(8) Cerofolini, L.; Fragai, M.; Ravera, E.; Diebolder, C. A.; Renault, L.; Calderone, V. Integrative Approaches in Structural Biology: A More Complete Picture from the Combination of Individual Techniques. *Biomolecules* **2019**, *9*, 370–381.

(9) Purslow, J. A.; Khatiwada, B.; Bayro, M. J.; Venditti, V. NMR Methods for Structural Characterization of Protein-Protein Complexes. *Front. Mol. Biosci.* **2020**, *7*, 1–8.

(10) Shi, Y. A Glimpse of Structural Biology through X-Ray Crystallography. *Cell* **2014**, *159*, 995–1014.

(11) Murata, K.; Wolf, M. Cryo-Electron Microscopy for Structural Analysis of Dynamic Biological Macromolecules. *Biochim. Biophys. Acta - Gen. Subj.* **2018**, *1862*, 324–334.

(12) Blanchet, C. E.; Svergun, D. I. Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution. *Annu. Rev. Phys. Chem.* **2013**, *64*, 37–54.

(13) Wang, D.; Bushnell, D. A.; Westover, K. D.; Kaplan, C. D.; Kornberg, R. D. Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. *Cell* **2006**, *127*, 941–954.

(14) Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M.; Morgan-Warren, R. J.; Carter, A. P.; Vonrheln, C.; Hartsch, T.; Ramakrishnan, V. Structure of the 30S Ribosomal Subunit. *Nature* **2000**, *407*, 327–339.

(15) Schluenzen, F.; Tocilj, A.; Zarivach, R.; Harms, J.; Gluehmann, M.; Janell, D.; Bashan, A.; Bartels, H.; Agmon, I.; Franceschi, F.; et al. Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution. *Cell* **2000**, *102*, 615–623.

(16) Agirrezabala, X.; Samatova, E.; Klimova, M.; Zamora, M.; Gil-Carton, D.; Rodnina, M. V.; Valle, M. Ribosome Rearrangements at the Onset of Translational Bypassing. *Sci. Adv.* **2017**, *3*, 1–8.

(17) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* **2000**, *289*, 905–920.

(18) Loveland, A. B.; Demo, G.; Grigorieff, N.; Korostelev, A. A. Ensemble Cryo-EM Elucidates the Mechanism of Translation Fidelity. *Nature* **2017**, *546*, 113–117.

(19) Schnell, H. M.; Walsh, R. M.; Rawson, S.; Kaur, M.; Bhanu, M. K.; Tian, G.; Prado, M. A.; Guerra-Moreno, A.; Paulo, J. A.; Gygi, S. P.; et al. Structures of Chaperone-Associated Assembly Intermediates Reveal Coordinated Mechanisms of Proteasome Biogenesis. *Nat. Struct. Mol. Biol.* **2021**, *28*, 418–425.

(20) Frank, J.; Ourmazd, A. Continuous Changes in Structure Mapped by Manifold Embedding of Single-Particle Data in Cryo-EM. *Methods* **2016**, *100*, 61–67.

(21) Zhao, J.; Benlekbir, S.; Rubinstein, J. L. Electron Cryomicroscopy Observation of Rotational States in a Eukaryotic V-ATPase. *Nature* **2015**, *521*, 241–245.

(22) Benesch, J. L. P.; Ruotolo, B. T. Mass Spectrometry: Come of Age for Structural and Dynamical Biology. *Curr. Opin. Struct. Biol.* **2011**, *21*, 641–649.

(23) Kaur, U.; Johnson, D. T.; Chea, E. E.; Deredge, D. J.; Espino, J. A.; Jones, L. M. Evolution of Structural Biology through the Lens of Mass Spectrometry. *Anal. Chem.* **2019**, *91*, 142–155.

(24) Heck, A. J. R. Native Mass Spectrometry: A Bridge between Interactomics and Structural Biology. *Nat. Methods* **2008**, *5*, 927–933.

(25) Faini, M.; Stengel, F.; Aebersold, R. The Evolving Contribution of Mass Spectrometry to Integrative Structural Biology. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 966–974.

(26) Sharon, M. How Far Can We Go with Structural Mass Spectrometry of Protein Complexes? *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 487–500.

(27) Ghosh, A.; Largy, E.; Gabelica, V. DNA G-Quadruplexes for Native Mass Spectrometry in Potassium: A Database of Validated Structures in Electrospray-Compatible Conditions. *Nucleic Acids Res.* **2021**, *49*, 2333–2345.

(28) D'Atri, V.; Gabelica, V. DNA and RNA Telomeric G-Quadruplexes: What Topology Features Can

Be Inferred from Ion Mobility Mass Spectrometry? *Analyst* **2019**, *144*, 6074–6088.

(29) Beveridge, R.; Calabrese, A. N. Structural Proteomics Methods to Interrogate the Conformations and Dynamics of Intrinsically Disordered Proteins. *Front. Chem.* **2021**, *9*, 49–68.

(30) Barrera, N. P.; Zhou, M.; Robinson, C. V. The Role of Lipids in Defining Membrane Protein Interactions: Insights from Mass Spectrometry. *Trends Cell Biol.* **2013**, *23*, 1–8.

(31) Struwe, W. B.; Robinson, C. V. Relating Glycoprotein Structural Heterogeneity to Function-Insights from Native Mass Spectrometry. *Curr. Opin. Struct. Biol.* **2019**, *58*, 241–248.

(32) Banazadeh, A.; Veillon, L.; Wooding, K. M.; Zabet-moghaddam, M.; Mechref, Y. Recent Advances in Mass Spectrometric Analysis of Glycoproteins. *Electrophoresis* **2017**, *38*, 162–189.

(33) Bauer, A.; Kuster, B. Affinity Purification-Mass Spectrometry. *Eur. J. Biochem.* **2003**, *270*, 570–578.

(34) Meyer, K.; Selbach, M. Quantitative Affinity Purification Mass Spectrometry: A Versatile Technology to Study Protein–Protein Interactions. *Front. Genet.* **2015**, *6*, 237–243.

(35) Dunham, W. H.; Mullin, M.; Gingras, A. C. Affinity-Purification Coupled to Mass Spectrometry: Basic Principles and Strategies. *Proteomics* **2012**, *12*, 1576–1590.

(36) Ewing, R. M.; Chu, P.; Elisma, F.; Li, H.; Taylor, P.; Climie, S.; McBroom-Cerajewski, L.; Robinson, M. D.; O'Connor, L.; Li, M.; et al. Large-scale Mapping of Human Protein–Protein Interactions by Mass Spectrometry. *Mol. Syst. Biol.* **2007**, *3*, 89–105.

(37) Sardiu, M. E.; Cai, Y.; Jin, J.; Swanson, S. K.; Conaway, R. C.; Conaway, J. W.; Florens, L.; Washburn, M. P. Probabilistic Assembly of Human Protein Interaction Networks from Label-Free Quantitative Proteomics. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 1454–1459.

(38) Jäger, S.; Cimermancic, P.; Gulbahce, N.; Johnson, J. R.; McGovern, K. E.; Clarke, S. C.; Shales, M.; Mercenne, G.; Pache, L.; Li, K.; et al. Global Landscape of HIV-Human Protein Complexes. *Nature* **2012**, *481*, 365–370.

(39) Gingras, A. C.; Raught, B. Beyond Hairballs: The Use of Quantitative Mass Spectrometry Data to Understand Protein-Protein Interactions. *FEBS Lett.* **2012**, *586*, 2723–2731.

(40) Krogan, N. J.; Cagney, G.; Yu, H.; Zhong, G.; Guo, X.; Ignatchenko, A.; Li, J.; Pu, S.; Datta, N.; Tikuisis, A. P.; et al. Global Landscape of Protein Complexes in the Yeast Saccharomyces Cerevisiae. *Nature* **2006**, *440*, 637–643.

(41) Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; et al. Systematic Identification of Protein Complexes in Saccharomyces Cerevisiae by Mass Spectrometry. *Nature* **2002**, *415*, 180–183.

(42) Guruharsha, K. G.; Rual, J. F.; Zhai, B.; Mintseris, J.; Vaidya, P.; Vaidya, N.; Beekman, C.; Wong, C.; Rhee, D. Y.; Cenaj, O.; et al. A Protein Complex Network of Drosophila Melanogaster. *Cell* **2011**, *147*, 690–703.

(43) Schmitt-Ulms, G.; Legname, G.; Baldwin, M. A.; Ball, H. L.; Bradon, N.; Bosque, P. J.; Crossin, K. L.; Edelman, G. M.; DeArmond, S. J.; Cohen, F. E.; et al. Binding of Neural Cell Adhesion Molecules (N-CAMs) to the Cellular Prion Protein. *J. Mol. Biol.* **2001**, *314*, 1209–1225.

(44) Sowa, M. E.; Bennett, E. J.; Gygi, S. P.; Harper, J. W. Defining the Human Deubiquitinating Enzyme Interaction Landscape. *Cell* **2009**, *138*, 389–403.

(45) Malovannaya, A.; Lanz, R. B.; Jung, S. Y.; Bulynko, Y.; Le, N. T.; Chan, D. W.; Ding, C.; Shi, Y.; Yucer, N.; Krenciute, G.; et al. Analysis of the Human Endogenous Coregulator Complexome. *Cell* **2011**, *145*, 787–799.

(46) Ajuh, P.; Kuster, B.; Panov, K.; Zomerdijk, J. C. B. M.; Mann, M.; Lamond, A. I. Functional Analysis of the Human CDC5L Complex and Identification of Its Components by Mass Spectrometry. *EMBO J.* **2000**, *19*, 6569–6581.

(47) Van Leene, J.; Hollunder, J.; Eeckhout, D.; Persiau, G.; Van De Slijke, E.; Stals, H.; Van Isterdael, G.; Verkest, A.; Neirynck, S.; Buffel, Y.; et al. Targeted Interactomics Reveals a Complex Core Cell Cycle Machinery in Arabidopsis Thaliana. *Mol. Syst. Biol.* **2010**, *6*, 397–408.

(48) Bouwmeester, T.; Bauch, A.; Ruffner, H.; Angrand, P. O.; Bergamini, G.; Croughton, K.; Cruciat, C.; Eberhard, D.; Gagneur, J.; Ghidelli, S.; et al. A Physical and Functional Map of the Human TNF-α/NF-KB Signal Transduction Pathway. *Nat. Cell Biol.* **2004**, *6*, 97–105.

(49) Glatter, T.; Wepf, A.; Aebersold, R.; Gstaiger, M. An Integrated Workflow for Charting the Human Interaction Proteome: Insights into the PP2A System. *Mol. Syst. Biol.* **2009**, *5*, 237–248.

(50) Leitner, A.; Faini, M.; Stengel, F.; Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem. Sci.* **2016**, *41*, 20–32.

(51) Yu, C.; Huang, L. Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal. Chem.* **2018**, *90*, 144–165.

(52) Liu, F.; Heck, A. J. R. Interrogating the Architecture of Protein Assemblies and Protein Interaction Networks by Cross-Linking Mass Spectrometry. *Curr. Opin. Struct. Biol.* **2015**, *35*, 100–108.

(53) Tang, X.; Wippel, H. H.; Chavez, J. D.; Bruce, J. E. Crosslinking Mass Spectrometry (XL-MS): A Link between Structural Biology and Systems Biology. *Protein Sci.* **2021**, *30*, 773–784.

(54) Sinz, A. Crosslinking Mass Spectrometry Goes In-Tissue. *Cell Syst.* **2018**, *6*, 37–51.

(55) Steigenberger, B. A.; Pieters, R. J.; Heck, A. J. R.; Scheltema, R. A. PhoX: An IMAC-Enrichable Cross-Linking Reagent. *ACS Cent. Sci.* **2019**, *5*, 1514–1522.

(56) Sinz, A. Divide and Conquer: Cleavable Cross-Linkers to Study Protein Conformation and Protein–Protein Interactions. *Anal. Bioanal. Chem.* **2017**, *409*, 33–44.

(57) H. Ihling, C.; Springorum, P.; Iacobucci, C.; Hage, C.; Götze, M.; Schäfer, M.; Sinz, A. The Isotope-Labeled, MS-Cleavable Cross-Linker Disuccinimidyl Dibutyric Urea for Improved Cross-Linking/Mass Spectrometry Studies. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 183–189.

(58) Tang, X.; Bruce, J. E. A New Cross-Linking Strategy: Protein Interaction Reporter Technology for Protein-Protein Interaction Studies. *Mol. Biosyst.* **2010**, *6*, 939–947.

(59) James, J.; Cryar, A.; Thalassinos, K. An Optimization Workflow for the Analysis of Cross-Linked Peptides Using a Quadrupole Time of Flight Mass Spectrometer. *Anal. Chem.* **2019**, *91*, 1808–1814.

(60) Schmidt, A.; Kalkhof, S.; Ihling, C.; Cooper, D. M. F.; Sinz, A. Mapping Protein Interfaces by Chemical Cross-Linking and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Application to a Calmodulin/Adenylyl Cyclase 8 Peptide Complex. *Eur. J. Mass Spectrom.* **2005**, *11*, 525–534.

(61) Liu, F.; Lössl, P.; Scheltema, R.; Viner, R.; Heck, A. J. R. Optimized Fragmentation Schemes and Data Analysis Strategies for Proteome-Wide Cross-Link Identification. *Nat. Commun.* **2017**, *8*, 15473–15480.

(62) Iacobucci, C.; Piotrowski, C.; Aebersold, R.; Amaral, B. C.; Andrews, P.; Borchers, C.; Brodie, N. I.; Bruce, J. E.; Chaignepain, S.; Chavez, J. D.; et al. First Community-Wide, Comparative Cross-Linking Mass Spectrometry Study. *Anal. Chem.* **2019**, *91*, 6953–6961.

(63) Tüting, C.; Iacobucci, C.; Ihling, C. H.; Kastritis, P. L.; Sinz, A. Structural Analysis of 70S Ribosomes by Cross-Linking/Mass Spectrometry Reveals Conformational Plasticity. *Sci. Rep.* **2020**, *10*, 12618–12630.

(64) Xia, L.; Ma, Z.; Tong, J.; Tang, Y.; Li, S.; Qin, S.; Lou, R.; Zhao, S.; Lei, X.; Shui, W. Evaluation of Chemical Cross-Linkers for in-Depth Structural Analysis of G Protein-Coupled Receptors through Cross-Linking Mass Spectrometry. *Anal. Chim. Acta* **2020**, *1102*, 53–62.

(65) Schweppe, D. K.; Chavez, J. D.; Lee, C. F.; Caudal, A.; Kruse, S. E.; Stuppard, R.; Marcinek, D. J.; Shadel, G. S.; Tian, R.; Bruce, J. E. Mitochondrial Protein Interactome Elucidated by Chemical Cross-Linking Mass Spectrometry. *Proc. Natl. Acad. Sci.* **2017**, *114*, 1732–1737.

(66) Fasci, D.; van Ingen, H.; Scheltema, R. A.; Heck, A. J. R. Histone Interaction Landscapes Visualized by Crosslinking Mass Spectrometry in Intact Cell Nuclei. *Mol. Cell. Proteomics* **2018**, *17*, 2018–2033.

(67) Chavez, J. D.; Lee, C. F.; Caudal, A.; Keller, A.; Tian, R.; Bruce, J. E. Chemical Crosslinking Mass Spectrometry Analysis of Protein Conformations and Supercomplexes in Heart Tissue. *Cell Syst.* **2018**, *6*, 136–141.

(68) Schmidt, C.; Robinson, C. V. A Comparative Cross-Linking Strategy to Probe Conformational Changes in Protein Complexes. *Nat. Protoc.* **2014**, *9*, 2224–2236.

(69) Chavez, J. D.; Eng, J. K.; Schweppe, D. K.; Cilia, M.; Rivera, K.; Zhong, X.; Wu, X.; Allen, T.; Khurgel, M.; Kumar, A.; et al. A General Method for Targeted Quantitative Cross-Linking Mass Spectrometry. *PLoS One* **2016**, *11*, 1–14.

(70) Müller, F.; Kolbowski, L.; Bernhardt, O. M.; Reiter, L.; Rappsilber, J. Data-Independent Acquisition Improves Quantitative Cross-Linking Mass Spectrometry. *Mol. Cell. Proteomics* **2019**, *18*, 786–795.

(71) Müller, F.; Graziadei, A.; Rappsilber, J. Quantitative Photo-Crosslinking Mass Spectrometry Revealing Protein Structure Response to Environmental Changes. *Anal. Chem.* **2019**, *91*, 9041–9048.

(72) Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R. Differential Hydrogen/Deuterium Exchange Mass Spectrometry Analysis of Protein-Ligand Interactions. *Expert Rev. Proteomics* **2011**, *8*, 43–59.

(73) Benhaim, M.; Lee, K. K.; Guttman, M. Tracking Higher Order Protein Structure by Hydrogen-Deuterium Exchange Mass Spectrometry. *Protein Pept. Lett.* **2018**, *26*, 16–26.

(74) Trabjerg, E.; Nazari, Z. E.; Rand, K. D. Conformational Analysis of Complex Protein States by Hydrogen/Deuterium Exchange Mass Spectrometry (HDX-MS): Challenges and Emerging Solutions. *TrAC - Trends Anal. Chem.* **2018**, *106*, 125–138.

(75) Hvidt, A.; Linderstrøm-Lang, K. Exchange of Hydrogen Atoms in Insulin with Deuterium Atoms in Aqueous Solutions. *BBA - Biochim. Biophys. Acta* **1954**, *14*, 574–575.

(76) Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Borchers, C.; Brier, S.; Bou-Assaf, G. M.;

Engen, J. R.; Englander, S. W.; Faber, J.; et al. Recommendations for Performing, Interpreting and Reporting Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) Experiments. *Nat. Methods* **2019**, *16*, 595–602.

(77) Fang, M.; Wang, Z.; Cupp-Sutton, K. A.; Welborn, T.; Smith, K.; Wu, S. High-Throughput Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) Coupled with Subzero-Temperature Ultrahigh Pressure Liquid Chromatography (UPLC) Separation for Complex Sample Analysis. *Anal. Chim. Acta* **2021**, *1143*, 65–72.

(78) Rey, M.; Yang, M.; Burns, K. M.; Yu, Y.; Lees-Miller, S. P.; Schriemer, D. C. Nepenthesin from Monkey Cups for Hydrogen/Deuterium Exchange Mass Spectrometry. *Mol. Cell. Proteomics* **2013**, *12*, 464–472.

(79) Sanguantrakun, N.; Chanthamontri, C.; Gross, M. L. Top-Down Analysis of In-Source HDX of Native Protein Ions. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 1151–1154.

(80) Mistarz, U. H.; Rand, K. D. Installation, Validation, and Application Examples of Two Instrumental Setups for Gas-Phase HDX-MS Analysis of Peptides and Proteins. *Methods* **2018**, *144*, 113–124.

(81) Rand, K. D.; Pringle, S. D.; Murphy, J. P.; Fadgen, K. E.; Brown, J.; Engen, J. R. Gas-Phase Hydrogen/Deuterium Exchange in a Traveling Wave Ion Guide for the Examination of Protein Conformations. *Anal. Chem.* **2009**, *81*, 10019–10028.

(82) Hodkinson, J. P.; Jahn, T. R.; Radford, S. E.; Ashcroft, A. E. HDX-ESI-MS Reveals Enhanced Conformational Dynamics of the Amyloidogenic Protein B2-Microglobulin upon Release from the MHC-1. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 278–286.

(83) Oyeyemi, O. A.; Sours, K. M.; Lee, T.; Kohen, A.; Resing, K. A.; Ahn, N. G.; Klinman, J. P. Comparative Hydrogen-Deuterium Exchange for a Mesophilic vs Thermophilic Dihydrofolate Reductase at 25 °c: Identification of a Single Active Site Region with Enhanced Flexibility in the Mesophilic Protein. *Biochemistry* **2011**, *50*, 8251–8260.

(84) Stephens, A. D.; Nespovitaya, N.; Zacharopoulou, M.; Kaminski, C. F.; Phillips, J. J.; Kaminski Schierle, G. S. Different Structural Conformers of Monomeric α-Synuclein Identified after Lyophilizing and Freezing. *Anal. Chem.* **2018**, *90*, 6975–6983.

(85) Hudgens, J. W.; Gallagher, E. S.; Karageorgos, I.; Anderson, K. W.; Filliben, J. J.; Huang, R. Y. C.; Chen, G.; Bou-Assaf, G. M.; Espada, A.; Chalmers, M. J.; et al. Interlaboratory Comparison of Hydrogen-Deuterium Exchange Mass Spectrometry Measurements of the Fab Fragment of NISTmAb. *Anal. Chem.* **2019**, *91*, 7336–7345.

(86) Tsutsui, Y.; Dela Cruz, R.; Wintrode, P. L. Folding Mechanism of the Metastable Serpin α 1-Antitrypsin. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 4467–4472.

(87) Hu, W.; Walters, B. T.; Kan, Z. Y.; Mayne, L.; Rosen, L. E.; Marqusee, S.; Englander, S. W. Stepwise Protein Folding at near Amino Acid Resolution by Hydrogen Exchange and Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 7684–7689.

(88) Jia, R.; Martens, C.; Shekhar, M.; Pant, S.; Pellowe, G. A.; Lau, A. M.; Findlay, H. E.; Harris, N. J.; Tajkhorshid, E.; Booth, P. J.; et al. Hydrogen-Deuterium Exchange Mass Spectrometry Captures Distinct Dynamics upon Substrate and Inhibitor Binding to a Transporter. *Nat. Commun.* **2020**, *11*, 6162–6171.

(89) Reading, E.; Ahdash, Z.; Fais, C.; Ricci, V.; Wang-Kan, X.; Grimsey, E.; Stone, J.; Malloci, G.; Lau, A. M.; Findlay, H.; et al. Perturbed Structural Dynamics Underlie Inhibition and Altered Efflux of the Multidrug Resistance Pump AcrB. *Nat. Commun.* **2020**, *11*, 5565–5575.

(90) Hebling, C. M.; Morgan, C. R.; Stafford, D. W.; Jorgenson, J. W.; Rand, K. D.; Engen, J. R. Conformational Analysis of Membrane Proteins in Phospholipid Bilayer Nanodiscs by Hydrogen Exchange Mass Spectrometry. *Anal. Chem.* **2010**, *82*, 5415–5419.

(91) Beveridge, R.; Migas, L. G.; Payne, K. A. P.; Scrutton, N. S.; Leys, D.; Barran, P. E. Mass Spectrometry Locates Local and Allosteric Conformational Changes That Occur on Cofactor Binding. *Nat. Commun.* **2016**, *7*, 12163–12171.

(92) West, G. M.; Chien, E. Y. T.; Katritch, V.; Gatchalian, J.; Chalmers, M. J.; Stevens, R. C.; Griffin, P. R. Ligand-Dependent Perturbation of the Conformational Ensemble for the GPCR β 2 Adrenergic Receptor Revealed by HDX. *Structure* **2011**, *19*, 1424–1432.

(93) Takamoto, K.; Chance, M. R. Radiolytic Protein Footprinting with Mass Spectrometry to Probe the Structure of Macromolecular Complexes. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 251–276.

(94) Chance, M. R.; Farquhar, E. R.; Yang, S.; Lodowski, D. T.; Kiselar, J. Protein Footprinting: Auxiliary Engine to Power the Structural Biology Revolution. *J. Mol. Biol.* **2020**, *432*, 2973–2984.

(95) Wang, L.; Chance, M. R. Structural Mass Spectrometry of Proteins Using Hydroxyl Radical Based Protein Footprinting. *Anal. Chem* **2011**, *83*, 7234–7241.

(96) Johnson, D. T.; Di Stefano, L. H.; Jones, L. M. Fast Photochemical Oxidation of Proteins (FPOP): A Powerful Mass Spectrometry–Based Structural Proteomics Tool. *J. Biol. Chem.* **2019**, *294*, 11969–11979.

(97) Schmitz, A.; Galas, D. J. The Interaction of RNA Polymerase and Lac Repressor with the Lac Control Region. *Nucleic Acids Res.* **1979**, *6*, 111–137.

(98) Galas, D. J.; Schmitz, A. DNAase Footprinting a Simple Method for the Detection of Protein-DNA Binding Specificity. *Nucleic Acids Res.* **1978**, *5*, 3157–3170.

(99) Mirzabekov, A. D.; Melnikova, A. F. Localization of Chromatin Proteins within DNA Grooves by Methylation of Chromatin with Dimethyl Sulphate. *Mol. Biol. Rep.* **1974**, *1*, 379–384.

(100) Kiselar, J. G.; Maleknia, S. D.; Sullivan, M.; Downard, K. M.; Chance, M. R. Hydroxyl Radical Probe of Protein Surfaces Using Synchrotron X-Ray Radiolysis and Mass Spectrometry. *Int. J. Radiat. Biol.* **2002**, *78*, 101–114.

(101) Maleknia, S. D.; Chance, M. R.; Downard, K. M. Electrospray-Assisted Modification of Proteins: A Radical Probe of Protein Structure. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2352–2358.

(102) Maleknia, S. D.; Brenowitz, M.; Chance, M. R. Millisecond Radiolytic Modification of Peptides by Synchrotron X-Rays Identified by Mass Spectrometry. *Anal. Chem.* **1999**, *71*, 3965–3973.

(103) Stocks, B. B.; Sarkar, A.; Wintrode, P. L.; Konermann, L. Early Hydrophobic Collapse of A1-Antitrypsin Facilitates Formation of a Metastable State: Insights from Oxidative Labeling and Mass Spectrometry. *J. Mol. Biol.* **2012**, *423*, 789–799.

(104) Sharp, J. S.; Becker, J. M.; Hettich, R. L. Analysis of Protein Solvent Accessible Surfaces by Photochemical Oxidation and Mass Spectrometry. *Anal. Chem.* **2004**, *76*, 672–683.

(105) Zhao, J.; Zhu, R.; Zhang, X.; Zhang, B.; Liu, Y.; Li, Y.; Wang, W.; Phillips, D. L. A Photoenhanced Oxidation of Amino Acids and the Cross-Linking of Lysozyme Mediated by Tetrazolium Salts. *Phys. Chem. Chem. Phys.* **2021**, *23*, 3761–3770.

(106) Wu, L.; Lapidus, L. J. Combining Ultrarapid Mixing with Photochemical Oxidation to Probe Protein Folding. *Anal. Chem.* **2013**, *85*, 4920–4924.

(107) Huang, W.; Ravikumar, K. M.; Chance, M. R.; Yang, S. Quantitative Mapping of Protein Structure by Hydroxyl Radical Footprinting-Mediated Structural Mass Spectrometry: A Protection Factor Analysis. *Biophys. J.* **2015**, *108*, 107–115.

(108) Chen, J.; Rempel, D. L.; Gross, M. L. Temperature Jump and Fast Photochemical Oxidation Probe Submillisecond Protein Folding. *J. Am. Chem. Soc.* **2010**, *132*, 15502–15504.

(109) Du, Y.; Duc, N. M.; Rasmussen, S. G. F.; Hilger, D.; Kubiak, X.; Wang, L.; Bohon, J.; Kim, H. R.; Wegrecki, M.; Asuru, A.; et al. Assembly of a GPCR-G Protein Complex. *Cell* **2019**, *177*, 1232–1242.

(110) Espino, J. A.; Jones, L. M. In Vivo Hydroxyl Radical Protein Footprinting for the Study of Protein Interactions in Caenorhabditis Elegans. *J. Vis. Exp.* **2020**, *158*, 1–8.

(111) Espino, J. A.; Jones, L. M. Illuminating Biological Interactions with in Vivo Protein Footprinting. *Anal. Chem.* **2019**, *91*, 6577–6584.

(112) Espino, J. A.; Zhang, Z.; Jones, L. M. Chemical Penetration Enhancers Increase Hydrogen Peroxide Uptake in C. Elegans for in Vivo Fast Photochemical Oxidation of Proteins. *J. Proteome Res.* **2020**, *19*, 3708–3715.

(113) Siuzdak, G.; Bothner, B.; Yeager, M.; Brugidou, C.; Fauquet, C. M.; Hoey, K.; Change, C.-M.; Chang, C. M. Mass Spectrometry and Viral Analysis. *Chem. Biol.* **1996**, *3*, 45–48.

(114) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Naked Protein Conformations: Cytochrome c in the Gas Phase. *J. Am. Chem. Soc.* **1995**, *117*, 10141–10142.

(115) Seo, J.; Hoffmann, W.; Warnke, S.; Bowers, M. T.; Pagel, K.; von Helden, G. Retention of Native Protein Structures in the Absence of Solvent: A Coupled Ion Mobility and Spectroscopic Study. *Angew. Chemie - Int. Ed.* **2016**, *55*, 14173–14176.

(116) Wyttenbach, T.; Bowers, M. T. Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility-Mass Spectrometry Environment. *J. Phys. Chem. B* **2011**, *115*, 12266–12275.

(117) Bakhtiari, M.; Konermann, L. Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures. *J. Phys. Chem. B* **2019**, *123*, 1784–1796.

(118) Valentine, S. J.; Anderson, J. G.; Ellington, A. D.; Clemmer, D. E. Disulfide-Intact and -Reduced Lysozyme in the Gas Phase: Conformations and Pathways of Folding and Unfolding. *J. Phys. Chem. B* **1997**, *101*, 3891–3900.

(119) CA, S.; K, T.; GR, H.; JH, S. Travelling Wave Ion Mobility Mass Spectrometry Studies of Protein Structure: Biological Significance and Comparison with X-Ray Crystallography and Nuclear Magnetic Resonance Spectroscopy Measurements. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3297–3304.

(120) Ruotolo, B. T.; Robinson, C. V. Aspects of Native Proteins Are Retained in Vacuum. *Curr. Opin. Chem. Biol.* **2006**, *10*, 402–408.

(121) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science* **2005**, *310*, 1658–1661.

(122) Sever, A. I. M.; Konermann, L. Gas Phase Protein Folding Triggered by Proton Stripping Generates Inside-Out Structures: A Molecular Dynamics Simulation Study. *J. Phys. Chem. B* **2020**, *124*, 3667–3677.

(123) Jhingree, J. R.; Bellina, B.; Pacholarz, K. J.; Barran, P. E. Charge Mediated Compaction and Rearrangement of Gas-Phase Proteins: A Case Study Considering Two Proteins at Opposing Ends of the Structure-Disorder Continuum. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1450–1461.

(124) Leney, A. C.; Heck, A. J. R. Native Mass Spectrometry: What Is in the Name? *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 5–13.

(125) Erba, E. B.; Petosa, C. The Emerging Role of Native Mass Spectrometry in Characterizing the Structure and Dynamics of Macromolecular Complexes. *Protein Sci.* **2015**, *24*, 1176–1192.

(126) Lorenzen, K.; Duijn, E. van. Native Mass Spectrometry as a Tool in Structural Biology. *Curr. Protoc. Protein Sci.* **2010**, *62*, 1–17.

(127) Konijnenberg, A.; Butterer, A.; Sobott, F. Native Ion Mobility-Mass Spectrometry and Related Methods in Structural Biology. *Biochim. Biophys. Acta - Proteins Proteomics* **2013**, *1834*, 1239–1256.

(128) Collings, B.; Douglas, D. An Extended Mass Range Quadrupole for Electrospray Mass Spectrometry. *Int. J. Mass Spectrom. Ion Process.* **1997**, *162*, 121–127.

(129) Winger, B. E.; Light-Wahl, K. J.; Ogorzalek Loo, R. R.; Udseth, H. R.; Smith, R. D. Observation and Implications of High Mass-to-Charge Ratio Ions from Electrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 536–545.

(130) Verentchikov, A. N.; Ens, W.; Standing, K. G. Reflecting Time-of-Flight Mass Spectrometer with an Electrospray Ion Source and Orthogonal Extraction. *Anal. Chem.* **1994**, *66*, 126–133.

(131) Hale, O. J.; Sisley, E. K.; Griffiths, R. L.; Styles, I. B.; Cooper, H. J. Native LESA TWIMS-MSI: Spatial, Conformational, and Mass Analysis of Proteins and Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 873–879.

(132) Vimer, S.; Ben-Nissan, G.; Morgenstern, D.; Kumar-Deshmukh, F.; Polkinghorn, C.; Quintyn, R. S.; Vasil'Ev, Y. V.; Beckman, J. S.; Elad, N.; Wysocki, V. H.; et al. Comparative Structural Analysis of 20S Proteasome Ortholog Protein Complexes by Native Mass Spectrometry. *ACS Cent. Sci.* **2020**, *6*, 573–588.

(133) Ashcroft, A. E. Mass Spectrometry and the Amyloid Problem-How Far Can We Go in the Gas Phase? *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1087–1096.

(134) Ashcroft, A. E. Mass Spectrometry-Based Studies of Virus Assembly. *Curr. Opin Virol.* **2019**, *36*, 17–24.

(135) Snijder, J. Extending the Boundaries of Native Mass Spectrometry to Study Virus Structure and Assembly, Utrecht University PhD Thesis, 2015.

(136) Bechara, C.; Robinson, C. V. Different Modes of Lipid Binding to Membrane Proteins Probed by Mass Spectrometry. *J. Am. Chem. Soc.* **2015**, *137*, 5240−5247.

(137) Barrera, N. P.; Robinson, C. V. Advances in the Mass Spectrometry of Membrane Proteins: From Individual Proteins to Intact Complexes. *Annu. Rev. Biochem.* **2011**, *80*, 247–271.

(138) Gault, J.; Liko, I.; Landreh, M.; Shutin, D.; Bolla, J. R.; Jefferies, D.; Agasid, M.; Yen, H. Y.; Ladds, M. J. G. W.; Lane, D. P.; et al. Combining Native and 'Omics' Mass Spectrometry to Identify Endogenous Ligands Bound to Membrane Proteins. *Nat. Methods* **2020**, *17*, 505–508.

(139) Bolla, J. R.; Sauer, J. B.; Wu, D.; Mehmood, S.; Allison, T. M.; Robinson, C. V. Direct Observation of the Influence of Cardiolipin and Antibiotics on Lipid II Binding to MurJ. *Nat. Chem.* **2018**, *10*, 363–371.

(140) Zeleny, J. On the Ratio of the Velocities of the Two Ions Produced in Gases by Röntgen Radiation; and on Some Related Phenomena. *London, Edinburgh, Dublin Philos. Mag. J. Sci.* **1898**, *46*, 120–154.

(141) Gabelica, V.; Marklund, E. Fundamentals of Ion Mobility Spectrometry. *Curr. Opin. Chem. Biol.* **2018**, *42*, 51–59.

(142) Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E. The Power of Ion Mobility-Mass Spectrometry for Structural Characterization and the Study of Conformational Dynamics. *Nat. Chem.* **2014**, *6*, 281–294.

(143) Cumeras, R.; Figueras, E.; Davis, C. E.; Baumbach, J. I.; Gràcia, I. Review on Ion Mobility Spectrometry Part 1: Current Instrumentation. *Analyst* **2015**, *140*, 1376–1390.

(144) Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.; et al. Recommendations for Reporting Ion Mobility Mass Spectrometry Measurements. *Mass Spectrom. Rev.* **2019**, *9999*, 1–30.

(145) Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. An Investigation of the Mobility Separation of Some Peptide and Protein Ions Using a New Hybrid Quadrupole/Travelling Wave IMS/Oa-ToF Instrument. *Int. J. Mass Spectrom.* **2007**, *261*, 1–12.

(146) Sivalingam, G. N.; Cryar, A.; Williams, M. A.; Gooptu, B.; Thalassinos, K. Deconvolution of Ion Mobility Mass Spectrometry Arrival Time Distributions Using a Genetic Algorithm Approach: Application to A1-Antitrypsin Peptide Binding. *Int. J. Mass Spectrom.* **2018**, *426*.

(147) Marklund, E. G.; Degiacomi, M. T.; Robinson, C. V.; Baldwin, A. J.; Benesch, J. L. P. Collision Cross Sections for Structural Proteomics. *Structure* **2015**, *23*, 791–799.

(148) May, J. C.; Morris, C. B.; McLean, J. A. Ion Mobility Collision Cross Section Compendium. *Anal. Chem.* **2017**, *89*, 1032–1044.

(149) Richardson, K.; Langridge, D.; Dixit, S. M.; Ruotolo, B. T. An Improved Calibration Approach for Travelling Wave Ion Mobility Spectrometry: Robust, High-Precision Collision Cross Sections. *Anal. Chem* **2021**, *93*, 3542–3550.

(150) Lee, J. W.; Davidson, K. L.; Bush, M. F.; Kim, H. I. Collision Cross Sections and Ion Structures: Development of a General Calculation Method via High-Quality Ion Mobility Measurements and Theoretical Modeling. *Analyst* **2017**, *142*, 4289–4298.

(151) Mortensen, D. N.; Susa, A. C.; Williams, E. R. Collisional Cross-Sections with T-Wave Ion Mobility Spectrometry without Experimental Calibration. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1282–1292.

(152) Sanders, J. D.; Grinfeld, D.; Aizikov, K.; Makarov, A.; Holden, D. D.; Brodbelt, J. S. Determination of Collision Cross-Sections of Protein Ions in an Orbitrap Mass Analyzer. *Anal. Chem.* **2018**, *90*, 5896–5902.

(153) Hinnenkamp, V.; Klein, J.; Meckelmann, S. W.; Balsaa, P.; Schmidt, T. C.; Schmitz, O. J. Comparison of CCS Values Determined by Traveling Wave Ion Mobility Mass Spectrometry and Drift Tube Ion Mobility Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 12042–12050.

(154) Thalassinos, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H.

Characterization of Phosphorylated Peptides Using Traveling Wave-Based and Drift Cell Ion Mobility Mass Spectrometry. *Anal. Chem.* **2009**, *81*, 248–254.

(155) Richardson, K.; Langridge, D.; Dixit, S. M.; Ruotolo, B. T. An Improved Calibration Approach for Traveling Wave Ion Mobility Spectrometry: Robust, High-Precision Collision Cross Sections. *Anal. Chem.* **2021**, *93*, 3542–3550.

(156) Smith, D.; Knapman, T. W.; Campuzana, I.; Malham, R. W.; Berryman, J. T.; Radford, S. E.; Ashcroft, A. E. Deciphering Drift Time Measurements from Travelling Wave Ion Mobility Spectrometry-Mass Spectrometry Studies. *Eur. J. Mass Spectrom.* **2009**, *15*, 113–130.

(157) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V. Ion Mobility-Mass Spectrometry Analysis of Large Protein Complexes. *Nat. Protoc.* **2008**, *3*, 1139–1152.

(158) Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J. E.; Ruotolo, B. T.; Robinson, C. V.; et al. Amyloid-β 2 Protein Oligomerization and the Importance of Tetramers and Dodecamers in the Aetiology of Alzheimer's Disease. *Nat. Chem.* **2009**, *1*, 326–331.

(159) Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T. Ion Mobilityg-Mass Spectrometry Reveals a Conformational Conversion from Random Assembly to β-Sheet in Amyloid Fibril Formation. *Nat. Chem.* **2011**, *3*, 172–177.

(160) Leney, A. C.; Pashley, C. L.; Scarff, C. A.; Radford, S. E.; Ashcroft, A. E. Insights into the Role of the Beta-2 Microglobulin D-Strand in Amyloid Propensity Revealed by Mass Spectrometry. *Mol. BioSyst.* **2014**, *10*, 412–420.

(161) Uetrecht, C.; Barbu, I. M.; Shoemaker, G. K.; Van Duijn, E.; Heck, A. J. R. Interrogating Viral Capsid Assembly with Ion Mobility-Mass Spectrometry. *Nat. Chem.* **2011**, *3*, 126–132.

(162) Nyon, M. P.; Segu, L.; Cabrita, L. D.; Lévy, G. R.; Kirkpatrick, J.; Roussel, B. D.; Patschull, A. O. M.; Barrett, T. E.; Ekeowa, U. I.; Kerr, R.; et al. Structural Dynamics Associated with Intermediate Formation in an Archetypal Conformational Disease. *Structure* **2012**, *20*, 504–512.

(163) Benesch, J. L. P. Collisional Activation of Protein Complexes: Picking Up the Pieces. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 341–348.

(164) Erba, E. B.; Ruotolo, B. T.; Barsky, D.; Robinson, C. V. Ion Mobility-Mass Spectrometry Reveals the Influence of Subunit Packing and Charge on the Dissociation of Multiprotein Complexes. *Anal. Chem.* **2010**, *82*, 9702–9710.

(165) Ruotolo, B. T.; Hyung, S.-J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. Ion Mobility–Mass Spectrometry Reveals Long-Lived, Unfolded Intermediates in the Dissociation of Protein Complexes. *Angew. Chemie* **2007**, *119*, 8147–8150.

(166) Bernstein, S. L.; Liu, D.; Wyttenbach, T.; Bowers, M. T.; Lee, J. C.; Gray, H. B.; Winkler, J. R. α-Synuclein: Stable Compact and Extended Monomeric Structures and PH Dependence of Dimer Formation. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1435–1443.

(167) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, *88*, 30–51.

(168) Macias, L. A.; Santos, I. C.; Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2020**, *92*, 227–251.

(169) Eyers, C. E.; Vonderach, M.; Ferries, S.; Jeacock, K.; Eyers, P. A. Understanding Protein–Drug Interactions Using Ion Mobility–Mass Spectrometry. *Curr. Opin. Chem. Biol.* **2018**, *42*, 167–

176.

(170) Tian, Y.; Ruotolo, B. T. The Growing Role of Structural Mass Spectrometry in the Discovery and Development of Therapeutic Antibodies. *Analyst* **2018**, *143*, 2459–2468.

(171) Tian, Y.; Han, L.; Buckner, A. C.; Ruotolo, B. T. Collision Induced Unfolding of Intact Antibodies: Rapid Characterization of Disulfide Bonding Patterns, Glycosylation, and Structures. *Anal. Chem.* **2015**, *87*, 11509–11515.

(172) Silveira, J. A.; Fort, K. L.; Kim, D.; Servage, K. A.; Pierson, N. A.; Clemmer, D. E.; Russell, D. H. From Solution to the Gas Phase: Stepwise Dehydration and Kinetic Trapping of Substance p Reveals the Origin of Peptide Conformations. *J. Am. Chem. Soc.* **2013**, *135*, 19147–19153.

(173) Shi, L.; Holliday, A. E.; Glover, M. S.; Ewing, M. A.; Russell, D. H.; Clemmer, D. E. Ion Mobility-Mass Spectrometry Reveals the Energetics of Intermediates That Guide Polyproline Folding. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 22–30.

(174) Clemmer, D. E.; Russell, D. H.; Williams, E. R. Characterizing the Conformationome: Toward a Structural Understanding of the Proteome. *Acc. Chem. Res.* **2017**, *50*, 556–560.

(175) Cui, W.; Rohrs, H. W.; Gross, M. L. Top-down Mass Spectrometry: Recent Developments, Applications and Perspectives. *Analyst* **2011**, *136*, 3854–3864.

(176) Chen, B.; Brown, K. A.; Lin, Z.; Ge, Y. Top-Down Proteomics: Ready for Prime Time? *Anal. Chem.* **2018**, *90*, 110–127.

(177) Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A. Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 1149–1157.

(178) Armirotti, A.; Damonte, G. Achievements and Perspectives of Top-down Proteomics. *Proteomics* **2010**, *10*, 3566–3576.

(179) Fornelli, L.; Toby, T. K.; Schachner, L. F.; Doubleday, P. F.; Srzentić, K.; DeHart, C. J.; Kelleher, N. L. Top-down Proteomics: Where We Are, Where We Are Going? *J. Proteomics* **2018**, *175*, 3–4.

(180) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499–519.

(181) Siuti, N.; Kelleher, N. L. Decoding Protein Modifications Using Top-down Mass Spectrometry. *Nat. Methods* **2007**, *4*, 817–821.

(182) Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B.; et al. Best Practices and Benchmarks for Intact Protein Analysis for Top-down Mass Spectrometry. *Nat. Methods* **2019**, *16*, 587–594.

(183) Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F. Radical Solutions: Principles and Application of Electron-Based Dissociation in Mass Spectrometry-Based Analysis of Protein Structure. *Mass Spectrom. Rev.* **2018**, *37*, 750–771.

(184) Dang, X.; Scotcher, J.; Wu, S.; Chu, R. K.; Tolić, N.; Ntai, I.; Thomas, P. M.; Fellers, R. T.; Early, B. P.; Zheng, Y.; et al. The First Pilot Project of the Consortium for Top-down Proteomics: A Status Report. *Proteomics* **2014**, *14*, 1130–1140.

(185) Srzentić, K.; Fornelli, L.; Tsybin, Y. O.; Loo, J. A.; Seckler, H.; Agar, J. N.; Anderson, L. C.; Bai, D. L.; Beck, A.; Brodbelt, J. S.; et al. Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.*

2020, *31*, 1783–1802.

(186) Floris, F.; Chiron, L.; Lynch, A. M.; Barrow, M. P.; Delsuc, M.-A.; O'Connor, P. B. Application of Tandem Two-Dimensional Mass Spectrometry for Top-Down Deep Sequencing of Calmodulin. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1700–1705.

(187) Floris, F.; van Agthoven, M.; Chiron, L.; Soulby, A. J.; Wootton, C. A.; Lam, Y. P. Y.; Barrow, M. P.; Delsuc, M. A.; O'Connor, P. B. 2D FT-ICR MS of Calmodulin: A Top-Down and Bottom-Up Approach. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1531–1538.

(188) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein Characterization Using Top-down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135*, 12646–12651.

(189) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R. Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. *Anal. Chem.* **2015**, *87*, 4152–4158.

(190) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* **2020**, *369*, 330–333.

(191) Roberts, D. S.; Mann, M. W.; Melby, J. A.; Larson, E. J.; Zhu, Y.; Brasier, A. R.; Jin, S.; Ge, Y. Structural O-Glycoform Heterogeneity of the SARS-CoV-2 Spike Protein Receptor-Binding Domain Revealed by Native Top-Down Mass Spectrometry. *BioRxiv* **2021**.

(192) Sinnott, M.; Malhotra, S.; Madhusudhan, M. S.; Thalassinos, K.; Topf, M. Combining Information from Crosslinks and Monolinks in the Modelling of Protein Structures. *Structure* **2020**, *28*, 1061–1070.

(193) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 587–596.

(194) Degiacomi, M. T.; Benesch, J. L. P. EM∩IM: Software for Relating Ion Mobility Mass Spectrometry and Electron Microscopy Data. *Analyst* **2016**, *141*, 70–75.

(195) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996**, *100*, 16082–16086.

(196) Kim, S. J.; Fernandez-Martinez, J.; Nudelman, I.; Shi, Y.; Zhang, W.; Raveh, B.; Herricks, T.; Slaughter, B. D.; Hogan, J. A.; Upla, P.; et al. Integrative Structure and Functional Anatomy of a Nuclear Pore Complex. *Nature* **2018**, *555*, 475–482.

(197) Ognjenović, J.; Grisshammer, R.; Subramaniam, S. Frontiers in Cryo Electron Microscopy of Complex Macromolecular Assemblies. *Annu. Rev. Biomed. Eng.* **2019**, *21*, 395–415.

(198) Su, Z.; Wu, C.; Shi, L.; Luthra, P.; Pintilie, G. D.; Johnson, B.; Porter, J. R.; Ge, P.; Chen, M.; Liu, G.; et al. Electron Cryo-Microscopy Structure of Ebola Virus Nucleoprotein Reveals a Mechanism for Nucleocapsid-like Assembly. *Cell* **2018**, *172*, 966–978.

(199) Schneider, M.; Belsom, A.; Rappsilber, J. Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends Biochem. Sci.* **2018**, *43*, 157–169.

(200) Verbeke, E. J.; Mallam, A. L.; Drew, K.; Marcotte, E. M.; Taylor, D. W. Classification of Single Particles from Human Cell Extract Reveals Distinct Structures. *Cell Rep.* **2018**, *24*, 259-268.e3.

(201) Jore, M. M.; Lundgren, M.; Van Duijn, E.; Bultema, J. B.; Westra, E. R.; Waghmare, S. P.; Wiedenheft, B.; Pul, Ü.; Wurm, R.; Wagner, R.; et al. Structural Basis for CRISPR RNA-Guided DNA Recognition by Cascade. *Nat. Struct. Mol. Biol.* **2011**, *18*, 529–536.

(202) Politis, A.; Stengel, F.; Hall, Z.; Hernández, H.; Leitner, A.; Walzthoeni, T.; Robinson, C. V.; Aebersold, R. A Mass Spectrometry–Based Hybrid Method for Structural Modeling of Protein Complexes. *Nat. Methods* **2014**, *11*, 403–406.

(203) Hauri, S.; Khakzad, H.; Happonen, L.; Teleman, J.; Malmström, J.; Malmström, L. Rapid Determination of Quaternary Protein Structures in Complex Biological Samples. *Nat. Commun.* **2019**, *10*, 192–201.

(204) Rappsilber, J. The Beginning of a Beautiful Friendship: Cross-Linking/Mass Spectrometry and Modelling of Proteins and Multi-Protein Complexes. *J. Struct. Biol.* **2011**, *173*, 530–540.

(205) Leitner, A.; Joachimiak, L. A.; Bracher, A.; Mönkemeyer, L.; Walzthoeni, T.; Chen, B.; Pechmann, S.; Holmes, S.; Cong, Y.; Ma, B.; et al. The Molecular Architecture of the Eukaryotic Chaperonin TRiC/CCT. *Structure* **2012**, *20*, 814–825.

(206) Santambrogio, C.; Favretto, F.; D'Onofrio, M.; Assfalg, M.; Grandori, R.; Molinari, H. Mass Spectrometry and NMR Analysis of Ligand Binding by Human Liver Fatty Acid Binding Protein. *J. Mass Spectrom.* **2013**, *48*, 895–903.

(207) Pacholarz, K. J.; Garlish, R. A.; Taylor, R. J.; Barran, P. E. Mass Spectrometry Based Tools to Investigate Protein–Ligand Interactions for Drug Discovery. *Chem. Soc. Rev.* **2012**, *41*, 4335–4355.

(208) Sharma, S.; Zheng, H.; Huang, Y. J.; Ertekin, A.; Hamuro, Y.; Rossi, P.; Tejero, R.; Acton, T. B.; Xiao, R.; Jiang, M.; et al. Construct Optimization for Protein NMR Structure Analysis Using Amide Hydrogen/Deuterium Exchange Mass Spectrometry. *Proteins Struct. Funct. Bioinforma.* **2009**, *76*, 882–894.

(209) Chung, E. W.; Nettleton, E. J.; Morgan, C. J.; Groß, M.; Miranker, A.; Radford, S. E.; Dobson, C. M.; Robinson, C. V. Hydrogen Exchange Properties of Proteins in Native and Denatured States Monitored by Mass Spectrometry and NMR. *Protein Sci.* **1997**, *6*, 1316–1324.

(210) Lu, J.; Zhou, F.; Liu, W.; Yu, F. Mass Spectrometry Assisted Arginine Side Chains Assignment of NMR Resonances in Natural Abundance Proteins. *J. Biomol. NMR* **2020**, *74*, 173–181.

(211) Karaca, E.; Bonvin, A. M. J. J. On the Usefulness of Ion-Mobility Mass Spectrometry and SAXS Data in Scoring Docking Decoys. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 683–694.

(212) Harrer, N.; Schindler, C. E. M.; Bruetzel, L. K.; Forné, I.; Ludwigsen, J.; Imhof, A.; Zacharias, M.; Lipfert, J.; Mueller-Planitz, F. Structural Architecture of the Nucleosome Remodeler ISWI Determined from Cross-Linking, Mass Spectrometry, SAXS, and Modeling. *Structure* **2018**, *26*, 282–294.

(213) Sugiyama, M.; Yagi, H.; Ishii, K.; Porcar, L.; Martel, A.; Oyama, K.; Noda, M.; Yunoki, Y.; Murakami, R.; Inoue, R.; et al. Structural Characterization of the Circadian Clock Protein Complex Composed of KaiB and KaiC by Inverse Contrast-Matching Small-Angle Neutron Scattering. *Sci. Rep.* **2016**, *6*, 2–8.

(214) Daly, S.; MacAleese, L.; Dugourd, P.; Chirot, F. Combining Structural Probes in the Gas Phase - Ion Mobility-Resolved Action-FRET. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 133–139.

(215) Czar, M. F.; Jockusch, R. A. Sensitive Probes of Protein Structure and Dynamics in Well-Controlled Environments: Combining Mass Spectrometry with Fluorescence Spectroscopy.

Curr. Opin. Struct. Biol. **2015**, *34*, 123–134.

(216) Stetefeld, J.; Mckenna, S. A.; Patel, T. R. Dynamic Light Scattering : A Practical Guide and Applications in Biomedical Sciences. *Biophys. Rev.* **2016**, 409–427.

(217) Gladytz, A.; Lugovoy, E.; Charvat, A.; Häupl, T.; Siefermann, K. R.; Abel, B. Intermediates Caught in the Act: Tracing Insulin Amyloid Fibril Formation in Time by Combined Optical Spectroscopy, Light Scattering, Mass Spectrometry and Microscopy. *Phys. Chem. Chem. Phys.* **2015**, *17*, 918–927.

(218) Minton, A. P. Recent Applications of Light Scattering Measurement in the Biological and Biopharmaceutical Sciences. *Anal. Biochem.* **2016**, *501*, 4–22.

(219) Amartely, H.; Avraham, O.; Friedler, A.; Livnah, O.; Lebendiker, M. Coupling Multi Angle Light Scattering to Ion Exchange Chromatography (IEX-MALS) for Protein Characterization. *Sci. Rep.* **2018**, *8*, 1–9.

(220) Huang, R. Y. C.; Wang, F.; Wheeler, M.; Wang, Y.; Langish, R.; Chau, B.; Dong, J.; Morishige, W.; Bezman, N.; Strop, P.; et al. Integrated Approach for Characterizing Bispecific Antibody/Antigens Complexes and Mapping Binding Epitopes with SEC/MALS, Native Mass Spectrometry, and Protein Footprinting. *Anal. Chem.* **2020**, *92*, 10709–10716.

(221) Weickert, S.; Cattani, J.; Drescher, M. Intrinsically Disordered Proteins (IDPs) Studied by EPR and in-Cell EPR. *Electron Paramagn. Reson.* **2019**, *26*, 1–37.

(222) Sciortino, G.; Sanna, D.; Ugone, V.; Maréchal, J. D.; Garribba, E. Integrated ESI-MS/EPR/Computational Characterization of the Binding of Metal Species to Proteins: Vanadium Drug-Myoglobin Application. *Inorg. Chem. Front.* **2019**, *6*, 1561–1578.

(223) Smith, E. A.; Thomas, W. D.; Kiessling, L. L.; Corn, R. M. Surface Plasmon Resonance Imaging Studies of Protein-Carbohydrate Interactions. *J. Am. Chem. Soc.* **2003**, *125*, 6140–6148.

(224) Beseničar, M.; Maček, P.; Lakey, J. H.; Anderluh, G. Surface Plasmon Resonance in Protein-Membrane Interactions. *Chem. Phys. Lipids* **2006**, *141*, 169–178.

(225) Nedelkov, D.; Nelson, R. W. Surface Plasmon Resonance Mass Spectrometry: Recent Progress and Outlooks. *Trends Biotechnol.* **2003**, *21*, 301–305.

(226) Della Ventura, B.; Banchelli, M.; Funari, R.; Illiano, A.; De Angelis, M.; Taroni, P.; Amoresano, A.; Matteini, P.; Velotta, R. Biosensor Surface Functionalization by a Simple Photochemical Immobilization of Antibodies: Experimental Characterization by Mass Spectrometry and Surface Enhanced Raman Spectroscopy. *Analyst* **2019**, *144*, 6871–6880.

(227) Schuck, P. Analytical Ultracentrifugation as a Tool for Studying Protein Interactions. *Biophys. Rev.* **2013**, *5*, 159–171.

(228) Jones, C. G.; Martynowycz, M. W.; Hattne, J.; Fulton, T. J.; Stoltz, B. M.; Rodriguez, J. A.; Nelson, H. M.; Gonen, T. The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. *ACS Cent. Sci.* **2018**, *4*, 1587–1592.

(229) Nannenga, B. L.; Gonen, T. The Cryo-EM Method Microcrystal Electron Diffraction (MicroED). *Nat. Methods* **2019**, *16*, 369–379.

(230) Nannenga, B. L.; Shi, D.; Leslie, A. G. W.; Gonen, T. High-Resolution Structure Determination by Continuous-Rotation Data Collection in MicroEDED. *Nat. Methods* **2014**, *11*, 927–930.

(231) Weiss, V. U.; Bereszczazk, J. Z.; Havlik, M.; Kallinger, P.; Gösler, I.; Kumar, M.; Blaas, D.; Marchetti-Deschmann, M.; Heck, A. J. R.; Szymanski, W. W.; et al. Analysis of a Common Cold

Virus and Its Subviral Particles by Gas-Phase Electrophoretic Mobility Molecular Analysis and Native Mass Spectrometry. *Anal. Chem.* **2015**, *87*, 8709–8717.

(232) Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Barrera, N. P.; Morgner, N.; Wang, L.; Saibil, H. R.; Robinson, C. V. Separating and Visualising Protein Assemblies by Means of Preparative Mass Spectrometry and Microscopy. *J. Struct. Biol.* **2010**, *172*, 161–168.

(233) Ivanov, Y. D.; Pleshakova, T.; Malsagova, K.; Kozlov, A.; Kaysheva, A.; Kopylov, A.; Izotov, A.; Andreeva, E.; Kanashenko, S.; Usanov, S.; et al. Highly Sensitive Protein Detection by Combination of Atomic Force Microscopy Fishing with Charge Generation and Mass Spectrometry Analysis. *FEBS J.* **2014**, *281*, 4705–4717.

(234) Ruggeri, F.; Krishnan, M. Lattice Diffusion of a Single Molecule in Solution. *Phys. Rev. E* **2017**, *96*, 1–15.

(235) Viswanath, S.; Bonomi, M.; Kim, S. J.; Klenchin, V. A.; Taylor, K. C.; Yabut, K. C.; Umbreit, N. T.; Van Epps, H. A.; Meehl, J.; Jones, M. H.; et al. The Molecular Architecture of the Yeast Spindle Pole Body Core Determined by Bayesian Integrative Modeling. *Mol. Biol. Cell* **2017**, *28*, 3298–3314.

(236) Liu, F.; Dubey, M.; Takahashi, H.; Castner, D. G.; Grainger, D. W. Immobilized Antibody Orientation Analysis Using Secondary Ion Mass Spectrometry and Fluorescence Imaging of Affinity-Generated Patterns. *Anal. Chem.* **2010**, *82*, 2947–2958.

(237) Wine, R. N.; Dial, J. M.; Tomer, K. B.; Borchers, C. H. Identification of Components of Protein Complexes Using a Fluorescent Photo-Cross-Linker and Mass Spectrometry. *Anal. Chem.* **2002**, *74*, 1939–1945.

(238) Deng, C.; Xiong, X.; Krutchinsky, A. N. Unifying Fluorescence Microscopy and Mass Spectrometry for Studying Protein Complexes in Cells. *Mol. Cell. Proteomics* **2009**, *8*, 1413–1423.

(239) Costanzo, M.; VanderSluis, B.; Koch, E. N.; Baryshnikova, A.; Pons, C.; Tan, G.; Wang, W.; Usaj, M. M. M.; Hanchard, J.; Lee, S. D.; et al. A Global Genetic Interaction Network Maps a Wiring Diagram of Cellular Function. *Science* **2016**, *353*, 1381–1395.

(240) Dai, X.; Gong, D.; Lim, H.; Jih, J.; Wu, T. T.; Sun, R.; Zhou, Z. H. Structure and Mutagenesis Reveal Essential Capsid Protein Interactions for KSHV Replication. *Nature* **2018**, *553*, 521–525.

(241) Braberg, A. H.; Echeverria, I.; Bohn, S.; Cimermancic, P.; Shiver, A.; Alexander, R.; Xu, J.; Shales, M.; Dronamraju, R.; Jiang, S.; et al. Genetic Interaction Mapping Informs Integrative Determination of Biomolecular Assembly Structures. *Science* **2020**, *4910*, 1–69.

(242) Sriswasdi, S.; Harper, S. L.; Tang, H. Y.; Gallagher, P. G.; Speicher, D. W. Probing Large Conformational Rearrangements in Wild-Type and Mutant Spectrin Using Structural Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1801–1806.

(243) Martinez-Martin, N.; Marcandalli, J.; Huang, C. S.; Arthur, C. P.; Perotti, M.; Foglierini, M.; Ho, H.; Dosey, A. M.; Shriver, S.; Payandeh, J.; et al. An Unbiased Screen for Human Cytomegalovirus Identifies Neuropilin-2 as a Central Viral Receptor. *Cell* **2018**, *174*, 1158–1171.

(244) Allison, J. R. Using Simulation to Interpret Experimental Data in Terms of Protein Conformational Ensembles. *Curr. Opin. Struct. Biol.* **2017**, *43*, 79–87.

(245) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. Principles of Protein Structural Ensemble Determination. *Curr. Opin. Struct. Biol.* **2017**, *42*, 106–116.

(246) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: Cambridge, 2002.

(247) Frenkel, D. Simulations: The Dark Side. *Eur. Phys. J. Plus* **2013**, *128*, 10–31.

(248) Porter, K. A.; Desta, I.; Kozakov, D.; Vajda, S. What Method to Use for Protein–Protein Docking? *Curr. Opin. Struct. Biol.* **2019**, *55*, 1–7.

(249) Pagadala, N. S.; Syed, K.; Tuszynski, J. Software for Molecular Docking: A Review. *Biophys. Rev.* **2017**, *9*, 91–102.

(250) Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E. Receptor-Ligand Molecular Docking. *Biophys. Rev.* **2014**, *6*, 75–87.

(251) Leman, J. K.; Weitzner, B. D.; Lewis, S. M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R. F.; Aprahamian, M.; Baker, D.; Barlow, K. A.; Barth, P.; et al. Macromolecular Modeling and Design in Rosetta: Recent Methods and Frameworks. *Nat. Methods* **2020**, *17*, 665–680.

(252) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; et al. SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information. *Nucleic Acids Res.* **2014**, *42*, 252–258.

(253) Riaz, N.; Wolden, S. L.; Gelblum, D. Y.; Eric, J. Comparative Protein Structure Modeling Using MODELLER. **2016**, *118*, 6072–6078.

(254) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143.

(255) Koukos, P. I.; Bonvin, A. M. J. J. Integrative Modelling of Biomolecular Complexes. *J. Mol. Biol.* **2020**, *432*, 2861–2881.

(256) Russel, D.; Lasker, K.; Webb, B.; Velázquez-Muriel, J.; Tjioe, E.; Schneidman-Duhovny, D.; Peterson, B.; Sali, A. Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLoS Biol.* **2012**, *10*, 1–5.

(257) Ferber, M.; Kosinski, J.; Ori, A.; Rashid, U. J.; Moreno-Morcillo, M.; Simon, B.; Bouvier, G.; Batista, P. R.; Muller, C. W.; Beck, M.; et al. Automated Structure Modeling of Large Protein Assemblies Using Crosslinks as Distance Restraints. *Nat. Methods* **2016**, *13*, 515–520.

(258) Tamò, G.; Maesani, A.; Träger, S.; Degiacomi, M. T.; Floreano, D.; Peraro, M. D. Disentangling Constraints Using Viability Evolution Principles in Integrative Modeling of Macromolecular Assemblies. *Sci. Rep.* **2017**, *7*, 1–9.

(259) Landreh, M.; Sahin, C.; Gault, J.; Sadeghi, S.; Lee Drum, C.; Uzdavinys, P.; Drew, D.; Allison, T.; Degiacomi, M.; Marklund, E. Predicting the Shapes of Protein Complexes Through Collision Cross Section Measurements and Database Searches. *Anal. Chem.* **2020**, *92*, 12297–12303.

(260) Tuckerman., M. E. *Statistical Mechanics : Theory and Molecular Simulation*; Oxford University Press: Oxford, 2010.

(261) Trabuco, L. G.; Villa, E.; Mitra, K.; Frank, J.; Schulten, K. Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Structure* **2008**, *16*, 673–683.

(262) Trabuco, L. G.; Villa, E.; Schreiner, E.; Harrison, C. B.; Schulten, K. Molecular Dynamics Flexible Fitting: A Practical Guide to Combine Cryo-Electron Microscopy and X-Ray Crystallography. *Methods* **2009**, *49*, 174–180.

(263) Topf, M.; Lasker, K.; Webb, B.; Wolfson, H.; Chiu, W.; Sali, A. Protein Structure Fitting and

Refinement Guided by Cryo-EM Density. *Structure* **2008**, *16*, 295–307.

(264) Igaev, M.; Kutzner, C.; Bock, L. V.; Vaiana, A. C.; Grubmüller, H. Automated Cryo-EM Structure Refinement Using Correlation-Driven Molecular Dynamics. *Elife* **2019**, *8*, 1–33.

(265) Bullock, J. M. A.; Sen, N.; Thalassinos, K.; Topf, M. Modeling Protein Complexes Using Restraints from Crosslinking Mass Spectrometry. *Structure* **2018**, *26*, 1015–1024.

(266) Shi, Y.; Fernandez-Martinez, J.; Tjioe, E.; Pellarin, R.; Kim, S. J.; Williams, R.; Schneidman-Duhovny, D.; Sali, A.; Rout, M. P.; Chait, B. T. Structural Characterization by Cross-Linking Reveals the Detailed Architecture of a Coatomer-Related Heptameric Module from the Nuclear Pore Complex. *Mol. Cell. Proteomics* **2014**, *13*, 2927–2943.

(267) Leitner, A.; Bonvin, A. M. J. J.; Borchers, C. H.; Chalkley, R. J.; Chamot-Rooke, J.; Combe, C. W.; Cox, J.; Dong, M. Q.; Fischer, L.; Götze, M.; et al. Toward Increased Reliability, Transparency, and Accessibility in Cross-Linking Mass Spectrometry. *Structure* **2020**, *28*, 1259–1268.

(268) Allison, T. M.; Barran, P.; Benesch, J. L. P.; Cianferani, S.; Degiacomi, M. T.; Gabelica, V.; Grandori, R.; Marklund, E. G.; Menneteau, T.; Migas, L. G.; et al. Software Requirements for the Analysis and Interpretation of Native Ion Mobility Mass Spectrometry Data. *Anal. Chem.* **2020**, *92*, 10881–10890.

(269) Burley, S. K.; Berman, H. M.; Bhikadiya, C.; Bi, C.; Chen, L.; Costanzo, L. Di; Christie, C.; Duarte, J. M.; Dutta, S.; Feng, Z.; et al. Protein Data Bank: The Single Global Archive for 3D Macromolecular Structure Data. *Nucleic Acids Res.* **2019**, *47*, 520–528.

(270) Vizcaíno, J. A.; Côté, R. G.; Csordas, A.; Dianes, J. A.; Fabregat, A.; Foster, J. M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; et al. The Proteomics Identifications (PRIDE) Database and Associated Tools: Status in 2013. *Nucleic Acids Res.* **2013**, *41*, 1063–1069.

(271) Lawson, C. L.; Baker, M. L.; Best, C.; Bi, C.; Dougherty, M.; Feng, P.; Van Ginkel, G.; Devkota, B.; Lagerstedt, I.; Ludtke, S. J.; et al. EMDataBank.Org: Unified Data Resource for CryoEM. *Nucleic Acids Res.* **2011**, *39*, 456–464.

(272) Liebschner, D.; Afonine, P. V; Moriarty, N. W.; Poon, B. K.; Chen, V. B.; Adams, P. D. CERES : A Cryo-EM Re-Refinement System for Continuous Improvement of Deposited Models. *Acta Crystallogr. Sect. D Struct. Biol.* **2021**, *77*, 48–61.

(273) Kikhney, A. G.; Borges, C. R.; Molodenskiy, D. S.; Jeffries, C. M.; Svergun, D. I. SASBDB: Towards an Automatically Curated and Validated Repository for Biological Scattering Data. *Protein Sci.* **2020**, *29*, 66–75.

(274) Vallat, B.; Webb, B.; Westbrook, J. D.; Sali, A.; Berman, H. M. Development of a Prototype System for Archiving Integrative/Hybrid Structure Models of Biological Macromolecules. *Structure* **2018**, *26*, 894–904.

(275) Iudin, A.; Korir, P. K.; Salavert-Torres, J.; Kleywegt, G. J.; Patwardhan, A. EMPIAR: A Public Archive for Raw Electron Microscopy Image Data. *Nat. Methods* **2016**, *13*, 387–388.

(276) Tanaka, K. The Proteasome: Overview of Structure and Functions. *Proc. Japan Acad. Ser. B Phys. Biol. Sci.* **2009**, *85*, 12–36.

(277) Bard, J. A. M.; Goodall, E. A.; Greene, E. R.; Jonsson, E.; Dong, K. C.; Martin, A. Structure and Function of the 26S Proteasome. *Annu. Rev. Biochem.* **2018**, *87*, 697–724.

(278) Mao, Y. Structure, Dynamics and Function of the 26S Proteasome. *Subcell. Biochem.* **2021**, *96*, 1–151.

(279) Dwivedi, V.; Yaniv, K.; Sharon, M. Beyond Cells: The Extracellular Circulating 20S Proteasomes. *Biochim. Biophys. Acta - Mol. Basis Dis.* **2021**, *1867*, 166041–166050.

(280) Collins, G. A.; Goldberg, A. L. The Logic of the 26S Proteasome. *Cell* **2017**, *169*, 792–806.

(281) Coux, O.; Tanaka, K.; Goldberg, A. L. Structure and Functions of the 20S and 26S Proteasomes. *Annu. Rev. Biochem.* **1996**, *65*, 801–847.

(282) Xie, Y. Structure, Assembly and Homeostatic Regulation of the 26S Proteasome. *J. Mol. Cell Biol.* **2010**, *2*, 308–317.

(283) Mayer, R. J. The Meteoric Rise of Regulated Intracellular Proteolysis. *Nat. Rev. Mol. Cell Biol.* **2000**, *1*, 145–148.

(284) Voges, D.; Zwickl, P.; Baumeister, W. The 26S Proteasome: A Molecular Machine Designed for Controlled Proteolysis. *Annu. Rev. Biochem.* **1999**, *68*, 1015–1068.

(285) Ciechanover, A.; Kwon, Y. T. a. Degradation of Misfolded Proteins in Neurodegenerative Diseases: Therapeutic Targets and Strategies. *Exp. Mol. Med.* **2015**, *47*, 147–162.

(286) Löwe, J.; Stock, D.; Jap, B.; Zwickl, P.; Baumeister, W.; Huber, R. Crystal Structure of the 20S Proteasome from the Archaeon T. Acidophilum at 3.4 Å Resolution. *Science* **1995**, *268*, 533–539.

(287) Arrigo, A. P.; Tanaka, K.; Goldberg, A. L.; Welch, W. J. Identity of the 19S “prosome” Particle with the Large Multifunctional Protease Complex of Mammalian Cells (the Prot Easome). *Nature* **1988**, *331*, 192–194.

(288) Groll, M.; Ditzel, L.; Löwe, J.; Stock, D.; Bochtler, M.; Bartunik, H. D.; Huber, R. Structure of 20S Proteasome from Yeast at 2.4 Å Resolution. *Nature* **1997**, *386*, 463–471.

(289) Dong, Y.; Zhang, S.; Wu, Z.; Li, X.; Wang, W. L.; Zhu, Y.; Stoilova-McPhie, S.; Lu, Y.; Finley, D.; Mao, Y. Cryo-EM Structures and Dynamics of Substrate-Engaged Human 26S Proteasome. *Nature* **2019**, *565*, 49–55.

(290) Ferrington, D. A.; Gregerson, D. S. Immunoproteasomes: Structure, Function, and Antigen Presentation. *Prog. Mol. Biol. Transl. Sci.* **2012**, *109*, 75–112.

(291) Dahlmann, B.; Ruppert, T.; Kuehn, L.; Merforth, S.; Kloetzel, P. M. Different Proteasome Subtypes in a Single Tissue Exhibit Different Enzymatic Properties. *J. Mol. Biol.* **2000**, *303*, 643–653.

(292) Murata, S.; Sasaki, K.; Kishimoto, T.; Niwa, S. I.; Hayashi, H.; Takahama, Y.; Tanaka, K. Regulation of CD8+ T Cell Development by Thymus-Specific Proteasomes. *Science* **2007**, *316*, 1349–1353.

(293) Uechi, H.; Hamazaki, J.; Murata, S. Characterization of the Testis-Specific Proteasome Subunit A4s in Mammals. *J. Biol. Chem.* **2014**, *289*, 12365–12374.

(294) Murata, S.; Takahama, Y.; Kasahara, M.; Tanaka, K. The Immunoproteasome and Thymoproteasome: Functions, Evolution and Human Disease. *Nat. Immunol.* **2018**, *19*, 923–931.

(295) Stadtmueller, B. M.; Hill, C. P. Proteasome Activators. *Mol. Cell* **2011**, *41*, 8–19.

(296) Aiken, C. T.; Kaake, R. M.; Wang, X.; Huang, L. Oxidative Stress-Mediated Regulation of Proteasome Complexes. *Mol. Cell. Proteomics* **2011**, *10*, 1–11.

(297) Weissman, A. M.; Shabek, N.; Ciechanover, A. The Predator Becomes the Prey: Regulating the

Ubiquitin System by Ubiquitylation and Degradation. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 605–620.

(298) Komander, D.; Rape, M. The Ubiquitin Code. *Annu. Rev. Biochem.* **2012**, *81*, 203–229.

(299) Bard, J. A. M.; Goodall, E. A.; Greene, E. R.; Jonsson, E.; Dong, K. C.; Martin, A. Structure and Function of the 26S Proteasome. *Annu. Rev. Biochem.* **2018**, *87*, 697–724.

(300) Harris, J. R. Release of a Macromolecular Protein Component from Human Erythrocyte Ghosts. *Biochim. Biophys. Acta - Biomembr.* **1968**, *150*, 534–537.

(301) Groll, M.; Ditzel, L.; Löwe, J.; Stock, D.; Bochtler, M.; Bartunik, H. D.; Huber, R. Structure of 20S Proteasome from Yeast at 2.4 Å Resolution. *Nature* **1997**, *386*, 463–471.

(302) Löwe, J.; Stock, D.; Jap, B.; Zwickl, P.; Baumeister, W.; Huber, R. Crystal Structure of the 20S Proteasome from the Archaeon T. Acidophilum at 3.4 Å Resolution. *Science* **1995**, *268*, 533–539.

(303) Peters, J. M.; Cejka, Z.; Harris, J. R.; Kleinschmidt, J. A.; Baumeister, W. Structural Features of the 26 S Proteasome Complex. *J. Mol. Biol.* **1993**, *234*, 932–937.

(304) Nickell, S.; Beck, F.; Scheres, S. H. W.; Korinek, A.; Förster, F.; Lasker, K.; Mihalache, O.; Sun, N.; Nagy, I.; Sali, A.; et al. Insights into the Molecular Architecture of the 26S Proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11943–11947.

(305) Förster, F.; Unverdorben, P.; Śledź, P.; Baumeister, W. Unveiling the Long-Held Secrets of the 26S Proteasome. *Structure* **2013**, *21*, 1551–1562.

(306) Lander, G. C.; Estrin, E.; Matyskiela, M. E.; Bashore, C.; Nogales, E.; Martin, A. Complete Subunit Architecture of the Proteasome Regulatory Particle. *Nature* **2012**, *482*, 186–191.

(307) Beck, F.; Unverdorben, P.; Bohn, S.; Schweitzer, A.; Pfeifer, G.; Sakata, E.; Nickell, S.; Plitzko, J. M.; Villa, E.; Baumeister, W.; et al. Near-Atomic Resolution Structural Model of the Yeast 26S Proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 14870–14875.

(308) Da Fonseca, P. C. A.; He, J.; Morris, E. P. Molecular Model of the Human 26S Proteasome. *Mol. Cell* **2012**, *46*, 54–66.

(309) Bohn, S.; Beck, F.; Sakata, E.; Walzthoeni, T.; Beck, M.; Aebersold, R.; Förster, F.; Baumeister, W.; Nickell, S. Structure of the 26S Proteasome from Schizosaccharomyces Pombe at Subnanometer Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20992–20997.

(310) Lasker, K.; Förster, F.; Bohn, S.; Walzthoeni, T.; Villa, E.; Unverdorben, P.; Beck, F.; Aebersold, R.; Sali, A.; Baumeister, W. Molecular Architecture of the 26S Proteasome Holocomplex Determined by an Integrative Approach. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1380–1387.

(311) Politis, A.; Stengel, F.; Hall, Z.; Hernández, H.; Leitner, A.; Walzthoeni, T.; Robinson, C. V.; Aebersold, R. A Mass Spectrometry-Based Hybrid Method for Structural Modeling of Protein Complexes. *Nat. Methods* **2014**, *11*, 403–406.

(312) Sharon, M.; Taverner, T.; Ambroggio, X. I.; Deshaies, R. J.; Robinson, C. V. Structural Organization of the 19S Proteasome Lid: Insights from MS of Intact Complexes. *PLoS Biol.* **2006**, *4*, 1314–1323.

(313) Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. Identification of Cross-Linked Peptides from Large Sequence Databases. *Nat. Methods* **2008**, *5*, 315–318.

(314) Chen, S.; Wu, J.; Lu, Y.; Ma, Y. B.; Lee, B. H.; Yu, Z.; Ouyang, Q.; Finley, D. J.; Kirschner, M. W.; Mao, Y. Structural Basis for Dynamic Regulation of the Human 26S Proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 12991–12996.

(315) Huang, X.; Luan, B.; Wu, J.; Shi, Y. An Atomic Structure of the Human 26S Proteasome. *Nat. Struct. Mol. Biol.* **2016**, *23*, 778–785.

(316) Schweitzer, A.; Aufderheide, A.; Rudack, T.; Beck, F.; Pfeifer, G.; Plitzko, J. M.; Sakata, E.; Schulten, K.; Förster, F.; Baumeister, W. Structure of the Human 26S Proteasome at a Resolution of 3.9 Å. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 7816–7821.

(317) Dong, Y.; Zhang, S.; Wu, Z.; Li, X.; Wang, W. L.; Zhu, Y.; Stoilova-McPhie, S.; Lu, Y.; Finley, D.; Mao, Y. Cryo-EM Structures and Dynamics of Substrate-Engaged Human 26S Proteasome. *Nature* **2019**, *565*, 49–55.

(318) Wehmer, M.; Rudack, T.; Beck, F.; Aufderheide, A.; Pfeifer, G.; Plitzko, J. M.; Förster, F.; Schulten, K.; Baumeister, W.; Sakata, E. Structural Insights into the Functional Cycle of the ATPase Module of the 26S Proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 1305–1310.

(319) Park, S.; Li, X.; Kim, H. M.; Singh, C. R.; Tian, G.; Hoyt, M. A.; Lovell, S.; Battaile, K. P.; Zolkiewski, M.; Coffino, P.; et al. Reconfiguration of the Proteasome during Chaperone-Mediated Assembly. *Nature* **2013**, *497*, 512–516.

(320) Schrader, J.; Henneberg, F.; Mata, R. A.; Tittmann, K.; Schneider, T. R.; Stark, H.; Bourenkov, G.; Chari, A. The Inhibition Mechanism of Human 20S Proteasomes Enables Next-Generation Inhibitor Design. *Science* **2016**, *353*, 594–598.

(321) Toste Rêgo, A.; da Fonseca, P. C. A. Characterization of Fully Recombinant Human 20S and 20S-PA200 Proteasome Complexes. *Mol. Cell* **2019**, *76*, 138–147.

(322) Kaake, R. M.; Kao, A.; Yu, C.; Huang, L. Characterizing the Dynamics of Proteasome Complexes by Proteomics Approaches. *Antioxidants Redox Signal.* **2014**, *21*, 2444–2456.

(323) Fabre, B.; Lambour, T.; Garrigues, L.; Amalric, F.; Vigneron, N.; Menneteau, T.; Stella, A.; Monsarrat, B.; Van den Eynde, B.; Burlet-Schiltz, O.; et al. Deciphering Preferential Interactions within Supramolecular Protein Complexes: The Proteasome Case. *Mol. Syst. Biol.* **2015**, *11*, 1–15.

(324) Ben-Nissan, G.; Vimer, S.; Tarnavsky, M.; Sharon, M. Structural Mass Spectrometry Approaches to Study the 20S Proteasome. *Methods Enzymol.* **2019**, *619*, 179–223.

(325) Lesne, J.; Locard-Paulet, M.; Parra, J.; Zivković, D.; Menneteau, T.; Bousquet, M. P.; Burlet-Schiltz, O.; Marcoux, J. Conformational Maps of Human 20S Proteasomes Reveal PA28- and Immuno-Dependent Inter-Ring Crosstalks. *Nat. Commun.* **2020**, *11*, 1–15.

(326) Huang, L.; Burlingame, A. L. Comprehensive Mass Spectrometric Analysis of the 20S Proteasome Complex. *Methods Enzymol.* **2005**, *405*, 187–236.

(327) Bousquet-Dubouch, M. P.; Baudelet, E.; Guérin, F.; Matondo, M.; Uttenweiler-Joseph, S.; Burlet-Schiltz, O.; Monsarrat, B. Affinity Purification Strategy to Capture Human Endogenous Proteasome Complexes Diversity and to Identify Proteasome-Interacting Proteins. *Mol. Cell. Proteomics* **2009**, *8*, 1150–1164.

(328) Claverol, S.; Burlet-Schiltz, O.; Girbal-Neuhauser, E.; Gairin, J. E.; Monsarrat, B. Mapping and Structural Dissection of Human 20 S Proteasome Using Proteomic Approaches. *Mol. Cell. Proteomics* **2002**, *1*, 567–578.

(329) Fernandez-Martinez, J.; Kim, S. J.; Shi, Y.; Upla, P.; Pellarin, R.; Gagnon, M.; Chemmama, I. E.;

Wang, J.; Nudelman, I.; Zhang, W.; et al. Structure and Function of the Nuclear Pore Complex Cytoplasmic MRNA Export Platform. *Cell* **2016**, *167*, 1215–1228.

(330) Bui, K. H.; von Appen, A.; DiGuilio, A. L.; Ori, A.; Sparks, L.; Mackmull, M.-T.; Bock, T.; Hagen, W.; Andrés-Pons, A.; Glavy, J. S.; et al. Integrated Structural Analysis of the Human Nuclear Pore Complex Scaffold. *Cell* **2013**, *155*, 1233–1243.

(331) Stuwe, T.; Correia, A. R.; Lin, D. H.; Paduch, M.; Lu, V. T.; Kossiakoff, A. A.; Hoelz, A. Architecture of the Nuclear Pore Complex Coat. *Science* **2015**, *347*, 1148–1152.

(332) Beck, M.; Lůí, V.; Förster, F.; Baumeister, W.; Medalia, O. Snapshots of Nuclear Pore Complexes in Action Captured by Cryo-Electron Tomography. *Nature* **2007**, *449*, 611–615.

(333) Allegretti, M.; Zimmerli, C. E.; Rantos, V.; Wilfling, F.; Ronchi, P.; Fung, H. K. H.; Lee, C. W.; Hagen, W.; Turoňová, B.; Karius, K.; et al. In-Cell Architecture of the Nuclear Pore and Snapshots of Its Turnover. *Nature* **2020**, *586*, 796–800.

(334) Albert, S.; Schaffer, M.; Beck, F.; Mosalaganti, S.; Asano, S.; Thomas, H. F.; Plitzko, J. M.; Beck, M.; Baumeister, W.; Engel, B. D. Proteasomes Tether to Two Distinct Sites at the Nuclear Pore Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 13726–13731.

(335) Wales, T. E.; Eggertson, M. J.; Engen, J. R. Considerations in the Analysis of Hydrogen Exchange Mass Spectrometry Data. *Methods Mol. Biol.* **2013**, *1007*, 263–288.

(336) Iacobucci, C.; Götze, M.; Sinz, A. Cross-Linking/Mass Spectrometry to Get a Closer View on Protein Interaction Networks. *Curr. Opin. Biotechnol.* **2020**, *63*, 48–53.

(337) Eldrid, C.; Thalassinos, K. Developments in Tandem Ion Mobility Mass Spectrometry. *Biochem. Soc. Trans.* **2020**, *48*, 2457–2466.

(338) Tang, K.; Li, F.; Shvartsburg, A. A.; Strittmatter, E. F.; Smith, R. D. Two-Dimensional Gas-Phase Separations Coupled to Mass Spectrometry for Analysis of Complex Mixtures. *Anal. Chem.* **2005**, *77*, 6381–6388.

(339) Shvartsburg, A. A.; Li, F.; Tang, K.; Smith, R. D. Characterizing the Structures and Folding of Free Proteins Using 2-D Gas-Phase Separations: Observation of Multiple Unfolded Conformers. *Anal. Chem.* **2006**, *78*, 3304–3315.

(340) Koeniger, S. L.; Merenbloom, S. I.; Valentine, S. J.; Jarrold, M. F.; Udseth, H. R.; Smith, R. D.; Clemmer, D. E. An IMS-IMS Analogue of MS-MS. *Anal. Chem.* **2006**, *78*, 4161–4174.

(341) Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S. D.; Richardson, K.; Langridge, D.; Green, M. R. A Cyclic Ion Mobility – Mass Spectrometry System. *Anal. Chem.* **2019**, *91*, 8564–8573.

(342) Eldrid, C.; Ujma, J.; Kalfas, S.; Tomczyk, N.; Giles, K.; Morris, M.; Thalassinos, K. Gas Phase Stability of Protein Ions in a Cyclic Ion Mobility Spectrometry Travelling Wave Device. *Anal. Chem.* **2019**, *91*, 7554–7561.

(343) Eldrid, C.; Ben-Younis, A.; Ujma, J.; Britt, H.; Cragnolini, T.; Kalfas, S.; Cooper-Shepherd, D.; Tomczyk, N.; Giles, K.; Morris, M.; et al. Cyclic Ion Mobility−Collision Activation Experiments Elucidate Protein Behavior in the Gas Phase. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1545–1552.

(344) Webb, I. K.; Garimella, S. V. B.; Tolmachev, A. V.; Chen, T. C.; Zhang, X.; Norheim, R. V.; Prost, S. A.; LaMarche, B.; Anderson, G. A.; Ibrahim, Y. M.; et al. Experimental Evaluation and Optimization of Structures for Lossless Ion Manipulations for Ion Mobility Spectrometry with Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2014**, *86*, 9169–9176.

(345) Deng, L.; Webb, I. K.; Garimella, S. V. B.; Hamid, A. M.; Zheng, X.; Norheim, R. V.; Prost, S. A.; Anderson, G. A.; Sandoval, J. A.; Baker, E. S.; et al. Serpentine Ultralong Path with Extended Routing (SUPER) High Resolution Traveling Wave Ion Mobility-MS Using Structures for Lossless Ion Manipulations. *Anal. Chem.* **2017**, *89*, 4628–4634.

(346) Hollerbach, A. L.; Li, A.; Prabhakaran, A.; Nagy, G.; Harrilal, C. P.; Conant, C. R.; Norheim, R. V.; Schimelfenig, C. E.; Anderson, G. A.; Garimella, S. V. B.; et al. Ultra-High-Resolution Ion Mobility Separations over Extended Path Lengths and Mobility Ranges Achieved Using a Multilevel Structures for Lossless Ion Manipulations Module. *Anal. Chem.* **2020**, *92*, 7972–7979.

(347) Wojcik, R.; Nagy, G.; Attah, I. K.; Webb, I.; Garimella, S. V. B.; Weitz, K. K.; Hollerbach, A.; Monroe, M. E.; Ligare, M. R.; Nielson, F. F.; et al. SLIM Ultrahigh Resolution Ion Mobility Spectrometry Separations of Isotopologues and Isotopomers Reveal Mobility Shifts Due to Mass Distribution Changes. *Anal. Chem.* **2019**, *91*, 11952–11962.

(348) Liu, F. C.; Ridgeway, M. E.; Park, M. A.; Bleiholder, C. Tandem Trapped Ion Mobility Spectrometry. *Analyst* **2018**, *143*, 2249–2258.

(349) Beardsley, R. L.; Jones, C. M.; Galhena, A. S.; Wysocki, V. H. Noncovalent Protein Tetramers and Pentamers with "n" Charges Yield Monomers with n/4 and n/5 Charges. *Anal. Chem.* **2009**, *81*, 1347–1356.

(350) Galhena, A. S.; Dagan, S.; Jones, C. M.; Beardsley, R. L.; Wysocki, V. H. Surface-Induced Dissociation of Peptides and Protein Complexes in a Quadrupole/Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2008**, *80*, 1425–1436.

(351) Wysocki, V. H.; Joyce, K. E.; Jones, C. M.; Beardsley, R. L. Surface-Induced Dissociation of Small Molecules, Peptides, and Non-Covalent Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 190–208.

(352) Wysocki, V. H.; Jones, C. M.; Galhena, A. S.; Blackwell, A. E. Surface-Induced Dissociation Shows Potential to Be More Informative Than Collision-Induced Dissociation for Structural Studies of Large Systems. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 903–913.

(353) Kish, M.; Smith, V.; Subramanian, S.; Vollmer, F.; Lethbridge, N.; Cole, L.; Bond, N. J.; Phillips, J. J. Allosteric Regulation of Glycogen Phosphorylase Solution Phase Structural Dynamics at High Spatial Resolution. *BioRxiv* **2019**.

(354) Contino, N. C.; Pierson, E. E.; Keifer, D. Z.; Jarrold, M. F. Charge Detection Mass Spectrometry with Resolved Charge States. *J. Am. Soc. Mass Spectrom* **2013**, *24*, 101–108.

(355) Elliott, A. G.; Merenbloom, S. I.; Chakrabarty, S.; Williams, E. R. Single Particle Analyzer of Mass: A Charge Detection Mass Spectrometer with a Multi-Detector Electrostatic Ion Trap. *Int. J. Mass Spectrom.* **2017**, *414*, 45–55.

(356) Todd, A. R.; Barnes, L. F.; Young, K.; Zlotnick, A.; Jarrold, M. F. Higher Resolution Charge Detection Mass Spectrometry. *Anal. Chem.* **2020**, *92*, 11357–11364.

(357) Arnold, J.; Mahamid, J.; Lucic, V.; De Marco, A.; Fernandez, J. J.; Laugks, T.; Mayer, T.; Hyman, A. A.; Baumeister, W.; Plitzko, J. M. Site-Specific Cryo-Focused Ion Beam Sample Preparation Guided by 3D Correlative Microscopy. *Biophys. J.* **2016**, *110*, 860–869.

(358) Seacrist, C. D.; Kuenze, G.; Hoffmann, R. M.; Moeller, B. E.; Burke, J. E.; Meiler, J.; Blind, R. D. Integrated Structural Modeling of Full-Length LRH-1 Reveals Inter-Domain Interactions Contribute to Receptor Structure and Function. *Structure* **2020**, *28*, 830–846.

(359) Vimer, S.; Ben-Nissan, G.; Sharon, M. Direct Characterization of Overproduced Proteins by Native Mass Spectrometry. *Nat. Protoc.* **2020**, *15*, 236–265.

(360) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput. *Nat. Methods* **2020**, *17*, 41–44.

(361) Kantz, E. D.; Tiwari, S.; Watrous, J. D.; Cheng, S.; Jain, M. Deep Neural Networks for Classification of LC-MS Spectral Peaks. *Anal. Chem.* **2019**, *91*, 12407–12413.

(362) Zhou, X. X.; Zeng, W. F.; Chi, H.; Luo, C.; Liu, C.; Zhan, J.; He, S. M.; Zhang, Z. PDeep: Predicting MS/MS Spectra of Peptides with Deep Learning. *Anal. Chem.* **2017**, *89*, 12690–12697.

(363) Ma, B. Novor: Real-Time Peptide de Novo Sequencing Software. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1885–1894.

(364) Gessulat, S.; Schmidt, T.; Zolg, D. P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; et al. Prosit: Proteome-Wide Prediction of Peptide Tandem Mass Spectra by Deep Learning. *Nat. Methods* **2019**, *16*, 509–518.

(365) Tran, N. H.; Zhang, X.; Xin, L.; Shan, B.; Li, M. De Novo Peptide Sequencing by Deep Learning. *Proc. Natl. Acad. Sci.* **2017**, *114*, 8247–8252.

(366) Qiao, R.; Tran, N. H.; Shan, B.; Xin, L.; Zhang, X.; Ghodsi, A.; Liu, C.; Li, M.; Chen, X. Deep Learning Enables de Novo Peptide Sequencing from Data-Independent-Acquisition Mass Spectrometry. *Nat. Methods* **2018**, *16*, 63–66.

(367) Elias, J. E.; Gibbons, F. D.; King, O. D.; Roth, F. P.; Gygi, S. P. Intensity-Based Protein Identification by Machine Learning from a Library of Tandem Mass Spectra. *Nat. Biotechnol.* **2004**, *22*, 214–219.

(368) Sinitcyn, P.; Daniel Rudolph, J.; Cox, J. Computational Methods for Understanding Mass Spectrometry–Based Shotgun Proteomics Data. *Annu. Rev. Biomed. Data Sci.* **2018**, *1*, 207–234.

(369) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data. *Nat. Methods* **2016**, *13*, 731–740.

(370) Meier, F.; Köhler, N. D.; Brunner, A.; Wanka, J. H.; Voytik, E.; Strauss, M. T.; Theis, F. J.; Mann, M. Deep Learning the Collisional Cross Sections of the Peptide Universe from a Million Experimental Values. *Nat. Commun.* **2021**, *12*, 1185–1196.

(371) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A. W. R.; Bridgland, A.; et al. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577*, 706–710.

(372) Wang, S.; Peng, J.; Ma, J.; Xu, J. Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. *Sci. Rep.* **2016**, *6*, 18962–18974.

(373) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; et al. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *10*, 1–12.

(374) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; on ne be rg er, O. R.; hr yn un ya su vu na ko ol, K. T.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**.

(375) Joseph, A. P.; Polles, G.; Alber, F.; Topf, M. Integrative Modelling of Cellular Assemblies. *Curr. Opin. Struct. Biol.* **2017**, *46*, 102–109.

(376) Trnka, M. J.; Baker, P. R.; Robinson, P. J. J.; Burlingame, A. L.; Chalkley, R. J. Matching Cross-Linked Peptide Spectra: Only as Good as the Worse Identification. *Mol. Cell. Proteomics* **2014**, *13*, 420–434.

(377) Hashimoto, Y.; Sheng, X.; Murray-Nerger, L. A.; Cristea, I. M. Temporal Dynamics of Protein Complex Formation and Dissociation during Human Cytomegalovirus Infection. *Nat. Commun.* **2020**, *11*, 1–20.

(378) Hall, D. R.; Peng, H. Characterizing Physical Protein Targets of Chemical Contaminants with Chemical Proteomics: Is It Time to Fill a Crucial Environmental Toxicology Knowledge Gap? *Comp. Biochem. Physiol. Part D Genomics Proteomics* **2020**, *34*, 1–6.

(379) Miettinen, T. P.; Peltier, J.; Härtlova, A.; Gierliński, M.; Jansen, V. M.; Trost, M.; Björklund, M. Thermal Proteome Profiling of Breast Cancer Cells Reveals Proteasomal Activation by CDK 4/6 Inhibitor Palbociclib. *EMBO J.* **2018**, *37*, 1–19.

(380) Kaur, U.; Meng, H.; Lui, F.; Ma, R.; Ogburn, R. N.; Johnson, J. H. R.; Fitzgerald, M. C.; Jones, L. M. Proteome-Wide Structural Biology: An Emerging Field for the Structural Analysis of Proteins on the Proteomic Scale. *J. Proteome Res.* **2018**, *17*, 3614–3627.

(381) Schopper, S.; Kahraman, A.; Leuenberger, P.; Feng, Y.; Piazza, I.; Müller, O.; Boersema, P. J.; Picotti, P. Measuring Protein Structural Changes on a Proteome-Wide Scale Using Limited Proteolysis-Coupled Mass Spectrometry. *Nat. Protoc.* **2017**, *12*, 2391–2410.

(382) Limpikirati, P.; Liu, T.; Vachet, R. W. Covalent Labeling-Mass Spectrometry with Non-Specific Reagents for Studying Protein Structure and Interactions. *Methods* **2018**, *144*, 79–93.

(383) Parker, B. W.; Goncz, E. J.; Krist, D. T.; Statsyuk, A. V.; Nesvizhskii, A. I.; Weiss, E. L. Mapping Low-Affinity/High-Specificity Peptide–Protein Interactions Using Ligand-Footprinting Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 21001–21011.

(384) Horne, J. E.; Walko, M.; Calabrese, A. N.; Levenstein, M. A.; Brockwell, D. J.; Kapur, N.; Wilson, A. J.; Radford, S. E. Rapid Mapping of Protein Interactions Using Tag-Transfer Photocrosslinkers. *Angew. Chemie Int. Ed.* **2018**, *57*, 16688–16692.

(385) Liu, X.; Salokas, K.; Weldatsadik, R. G.; Gawriyski, L.; Varjosalo, M. Combined Proximity Labeling and Affinity Purification−mass Spectrometry Workflow for Mapping and Visualizing Protein Interaction Networks. *Nat. Protoc.* **2020**, *15*, 3182–3211.

(386) Nguyen, T. M. T.; Kim, J.; Doan, T. T.; Lee, M. W.; Lee, M. APEX Proximity Labeling as a Versatile Tool for Biological Research. *Biochemistry* **2020**, *59*, 260–269.

Table of contents (TOC) graphic