Published in final edited form as:

Proteomics. 2015 August ; 15(16): 2777–2791. doi:10.1002/pmic.201400517.

Exposing the subunit diversity and modularity of protein complexes by structural mass spectrometry approaches

Dror S. Chorev, Gili Ben-Nissan, and Michal Sharon

Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 7610001, Israel

Abstract

Although the number of protein-encoding genes in the human genome is only about 20,000, not far from the amount found in the nematode worm genome, the number of proteins that are translated from these sequences is larger by several orders of magnitude. A number of mechanisms have evolved to enable this diversity. For example, genes can be alternatively spliced to create multiple transcripts; they may also be translated from different alternative initiation sites. After translation, hundreds of chemical modifications can be introduced in proteins, altering their chemical properties, folding, stability and activity. The complexity is then further enhanced by the various combinations that are generated from the assembly of different subunit variants into protein complexes. This, in turn, confers structural and functional flexibility, and endows the cell with the ability to adapt to various environmental conditions. Therefore, exposing the variability of protein complexes is an important step towards understanding their biological functions. Revealing this enormous diversity, however, is not a simple task. In this review, we will focus on the array of mass spectrometry based strategies that are capable of performing this mission. We will also discuss the challenges that lie ahead, and the future directions toward which the field might be heading.

1 Introduction

Over a decade has passed since the human genome was fully sequenced, revealing the existence of approximately 20,000 protein-encoding genes [1, 2]. At the time, this finding was rather surprising, as much simpler organisms contained similar numbers of genes [3]. However, it was later realized that the combinatorial space in the human proteome is much larger, encompassing more than 1 million proteins [4, 5]. This enormous repertoire of proteins is the result of diversification mechanisms that exist at every step of the protein biosynthesis pathway, starting within the DNA, through RNA, and onwards to the post-translational level, all of which increase the proteins' functional diversity (Fig. 1).

At the DNA level, non-synonymous single nucleotide polymorphisms (nsSNPs) exist: these single-base changes lead to a modification in the amino acid sequence of the encoded protein [6, 7]. nsSNPs were shown to have an impact not only on the translated protein's post-translational modifications (PTMs), but also on its interactions pattern with other

proteins. Gene duplication constitutes an additional evolutionary process that enhances genetic variability [8, 9]. As a result, a single protein can be encoded by more than one gene. The duplicated gene is often free from pressures of selection; as a consequence, it accumulates mutations, leading to functional diversification of the translated protein.

Diversification is then further pursued by alternative splicing, a multiplication process that occurs at the RNA level (Fig. 1). Alternative splicing entails differential processing of the mRNA, producing several protein versions that differ from those encoded from the primary transcript. These protein isoforms may differ in structure, function, localization, or other properties [10]. In general, it is estimated that more than 90% of the genes undergo this process [10]; moreover, each gene in the human repertoire yields, on average, four protein isoforms [11]. This complexity is further enhanced by the multiple translation initiation sites on mRNA transcripts, enabling the translation of several versions of each protein from the same template [12]. The difference in the lengths of alternative N-termini of a protein can reach tens of codons, and may lead to different targeting of the variants produced.

Following protein synthesis, proteome diversity is facilitated by the covalent addition of functional groups, and proteolytic cleavage [13] (Fig. 1). These PTMs include phosphorylation, glycosylation, ubiquitination, nitrosylation, methylation, acetylation, lipidation, and proteolysis. In total, there are more than 400 documented types of PTMs [13], and it is estimated that about 5% of all human genes encode for modifying enzymes [14]. PTMs may affect the protein's activity state, localization, turnover, and interactions with other proteins. Moreover, considering that different PTMs display different physicochemical properties [13, 15], the same protein may exhibit different functions, depending on how it is modified [16]. Likewise, non-covalent interactions with co-factors and ligands may also induce structural changes in the protein, activate or inactivate it, influence its stability, and modulate its subcellular localization [17, 18].

These mechanisms of protein diversification generate millions of different proteins within the human body. In fact, recent estimates suggest that the total number of protein variants, termed "proteoforms," is somewhere between 10 million [19] and 1 billion [20, 21]. Nevertheless, the complexity does not end there. Considering that up to 80% of the proteome is predicted to exist in complexes or large macromolecular assemblies [22], the various "proteoforms" are expected to generate a multitude of compositionally distinct protein assemblies (Fig. 1). In addition, the same protein entity may belong to different modules at different times, and in different subcellular locations, and may differ in function, depending on its environment. Moreover, subunits or assembled states may form associations with other proteins/subcomplexes or complexes, providing a means of communication and cooperation between the many different functions and pathways within a cell. Overall, such modularity enables structural flexibility, creating novel and unique properties that enable response and adaption to varying cellular conditions [23–26].

While this modularity of protein assemblies yields clear biological benefits, its characterization is not a simple task. This is due to the multiple overlapping populations of protein complexes which might differ by relatively small shifts in mass, chemical character, and structural properties. In this review, we will focus on the ability of various structural

mass spectrometry (MS)-based approaches to provide insights into the functional and structural diversity of protein complexes (Fig. 2 and Box 1). As excellent reviews are already available on the basic use of MS for biomolecular analysis, and the technical challenges associated with such measurements, we will not discuss these aspects here [27–30]. We will, however, highlight how each of the MS methods described uncovers different aspects of the complexity hidden within protein complexes (Fig. 2, Box 1). These include approaches such as top-down and quantitative proteomics, which enable identification of different protein isoforms and PTM characteristics [5, 31, 32], hydrogen exchange and radical footprinting, which expose solvent accessible surface areas [33–37], and methods such as native MS [29, 38, 39], ion mobility separation [40–43], and cross-linking MS [44, 45], which yield information on the heterogeneity, topology and packing of protein assemblies. How this arsenal of MS tools is harnessed for exposing the multi-level diversity of protein complexes is detailed below.

Uncovering the multiple co-existing forms of a single protein

As described above, any given protein is subject to various pre- or post-translational modifications. Identification of these modifications has been traditionally carried out by bottom-up MS, which is based on proteolytic digestion of the proteins, and liquid chromatographic (LC) separation of the peptide mixture prior to MS and tandem MS (MS/MS) analysis [13, 15]. While this technique constitutes a very powerful tool for quantification and identification of PTMs, and in the past several years, tremendous progress has been made in developing proteomics technologies which enable the execution of numerous high-throughput, global PTM analyses [13, 15, 32], it is limited in its ability to distinguish between different proteoforms [46]. This is due to the fact that identical peptides can actually originate from different protein isoforms; furthermore, modifications or sequence variations may occur on distinct peptides, causing their relationship to one another to be lost, following digestion. This information can be conserved, however, by introducing intact proteins into the mass spectrometer, as is done in top-down MS analysis.

To date, measuring the intact mass of proteins has become routine; yet distinguishing between the different species of a single protein may be difficult, due to the small shifts in mass that may discriminate between them [47]. However, in recent years, the introduction of high-resolution MS instrumentation has greatly enhanced the ability to separate and measure miniscule mass differences [48–50]. In this respect, an elegant strategy that highlights the impressive heterogeneity within a single protein was recently described [51, 52]. It involves high-resolution electrospray ionization MS of intact ovalbumin under native conditions. Measurements were conducted on a modified Exactive Plus Orbitrap mass spectrometer [50], which enables high mass accuracy, and high resolving power of ions well over 2000 m/z . At least 59 different proteoforms were identified and semi-quantified (Fig. 3), using this instrument. This variety was shown to be largely attributable to multiple phosphorylation sites, and a glycosylation site that could be occupied by at least 45 different glycan structures. Similar results were also obtained from analyses of monoclonal antibodies, showing that even this relatively uniform molecule has over 20 different proteoforms [52]. It is unclear whether such a broad array of PTMs is a common

phenomenon; however, studies on proteins such as histones [53], tau [54] and Hsp90 [55] suggest that it widely occurs.

Large-scale analysis of intact proteins still remains challenging in terms of proteome coverage, sensitivity, and throughput. However, recent advances in top-down proteomic techniques, involving both intact protein mass measurements and fragmentation of these ions within the mass spectrometer, have enabled such analyses [46]. For example, by utilizing a novel, four-dimensional separation system that combines both off- and on-line separation, a sufficient degree of separation enabling top-down analysis was achieved [56]. Moreover, this proteome-wide study [56], was dependent on high instrument performance for detection and identification of the intact proteins. In particular, high resolution and mass accuracy were essential for accurate assignment of peaks, given that species may differ by small mass shifts. For example, the difference between acetylation and trimethylation is only $m = 39$ mDa; phosphorylation versus sulfation ($m = 10$ mDa); or reduced/oxidized disulfide bonds ($m = 2$ Da) [46]. The sensitivity of the mass spectrometer was also a critical factor, considering that as the mass of a protein increases, it will display broad isotopic distributions, and an increased number of charge states (from electrospray ionization, ESI). These two effects cause the signal to split into multiple channels, reducing its intensity in any mass-to-charge ratio [46]. These challenges were overcome by using a high-resolution Fourier-transform mass spectrometer (7 T LTQ-FT-ICR) capable of attaining high resolving power (85,000 in m/z 400), in combination with software development for intact mass detection, interpretation, and visualization [56]. Overall, more than 1,000 gene products that generate more than 3,000 protein species were thus discovered. This study not only exposed the molecular complexity of the proteome, but also alluded to the ensembles of compositionally distinct protein complexes that may be generated from their assembly.

Whether all existing variants of a given protein subunit can be incorporated into a functional complex remains an unresolved issue. Similarly, the nature of the cross-talk between the different subunit variants that co-exist within a complex is not yet known; nor whether a correlation may be drawn between the types of modifications that occur on neighboring subunits.

To begin answering these questions, thereby exposing the diversity of subunits that comprise protein complexes, novel approaches are being developed, especially those that combine top-down and bottom-up proteomics. One such method relies on denaturing the protein complex, and separating its constituent subunits on a monolithic column prior to MS analysis [57]. Subsequently, the eluted flow is split into two fractions. One fraction is directed straight into a mass spectrometer for intact protein mass measurements, while the rest of the flow is fractionated, for subsequent bottom-up proteomic analysis. Measurement of the intact molecular weight enables identification of the distinct variants of a single subunit that are incorporated into the complex, possibly exposing the combinations of PTMs that might occur on a single protein, and the interplay between neighboring modifications. Proteomic analysis, on the other hand, makes it possible to define the protein sequence identity and PTMs, whereas the separation step prior to the proteomic analysis increases the odds of full sequence coverage and PTM identification. Subsequently, based on the elution profile, information from top-down and bottom-up proteomics is correlated. In the case of the

endogenous human COP9 signalosome, it was determined that almost all of the subunit components of the complex have multiple co-existing variants [57]. Unraveling how this variability in subunit composition changes between protein complexes isolated from different cellular compartments, tissues or organs constitutes the next step forward in understanding the dynamic regulatory program of protein complex formation.

Covalent modifications for detection of conformational changes upon post-translational modification or ligand binding

The function of protein complexes is adjusted according to external and internal cues, such as growth factors, cell-cycle checkpoints, DNA damage, oxygen tension and nutrient status [58]. Thus, functional modulation is often achieved by reversible covalent modifications of proteins, and non-covalent binding of cofactors, metabolites or ligands [18, 59]. These modifications and interactions lead to structural adjustments in protein conformation due to changes in physicochemical properties, which in turn modulate protein assembly activity, association with protein partners or cellular stability and localizations [60, 61]. During the past decade, several structural MS approaches have been developed, in order to probe these significant but often subtle conformational rearrangements. Specifically, these MS-based methods utilize covalently labeling strategies such as hydrogen-deuterium exchange, hydroxyl radical footprinting, or chemical cross-linking, prior to mass spectrometry analysis, as we describe below.

A recent MS-based method that was developed to detect the effects of PTMs on the structural architecture of protein complexes, was employed to probe the conformational changes induced in the chloroplast ATP synthase in response to phosphorylation [62, 63]. ATP synthases are membrane-bound rotary motors that are responsible for the generation of ATP, and maintenance of the cell's pH. Plant F_1F_0 -ATPase, located in the thylakoid membrane of the chloroplast, is composed of a membrane-spanning ring connected to a soluble, nucleotide-binding head by a central stalk that rotates under the pH gradient and a peripheral, non-rotating stator stalk (Fig. 4). Several phosphorylation sites within the soluble head, as well as the two stalks themselves, have been mapped within the complex; a comparative cross-linking strategy was subsequently applied to study the conformational changes induced upon dephosphorylation [62, 63].

In general, cross-linking MS involves the covalent linkage of two reactive groups of amino acid side chains in close proximity, using bifunctional chemical reagents [44, 45]. The distance bridged by the cross-linker is determined by the length of the carbon chain connecting the functional groups. Current chemical cross-linking reagents typically target the primary amine moieties found on lysine residues, due to their higher reactivity. Cysteine-specific cross-linking is also well-established; however, due to the low prevalence of cysteines in proteins, and their involvement in the formation of disulfide bonds, this approach usually does not yield sufficient structural information. Recently, cross-linking reagents that connect the less reactive acidic residues, aspartic and glutamic acid, under conditions that are compatible with the integrity of protein complexes, have been developed [48]. An additional strategy involves the use of non-specific photoreactive cross-linkers that react with the target proteins upon UV irradiation. To date, most photoreactive cross-linkers

are heterobifunctional reagents that, in addition, possess a different functional group (e.g., amine or sulfhydryl-reactive). This enables the cross-linking reaction to progress in a controlled, two-step reaction; thereby avoiding the formation of higher oligomers.

Identification of the cross-linked sites is usually undertaken in a bottom-up fashion that includes digestion of the cross-linked proteins, and LC-MS/MS analysis. Analyzing the MS data generated by cross-linking experiments is a challenging task, mainly because of the overwhelming numbers of possible cross-linked peptide combinations that should be considered. The current bottleneck in the field entails the development of specific software tools that would be capable of analyzing these complex datasets, including verification and validation steps, in a fully automated manner [44, 64]. Nevertheless, several different approaches have been developed in the last decade to assist in the automated identification of cross-linked peptides from MS data, and there are numerous examples demonstrating how this approach can then be used for fold detection, identification of protein–protein interactions, and characterization of the subunit architecture of protein complexes [44, 64]. In the ATP synthase study, in order to probe the structural changes induced by phosphorylation, various stable isotope–labeled reagents were used, independently, to cross-link the phosphorylated and dephosphorylated ATPase complexes (Fig. 4A). In this manner, dephosphorylation was shown to reduce the overall stability of the complex, and decrease the level of cross-linking between the soluble head and the two stalks, suggesting that phosphorylation regulates inactivation of the enzyme by preventing rotation of the soluble head. Moreover, dephosphorylation reduced cross-linking along the soluble head, inducing conformational fluctuations that enable nucleotide release (Fig. 4B and 4C) [63].

In general, this comparative cross-linking strategy is not limited to probing the effects of phosphorylation events; rather, it may be applied to other types of PTMs, as well [62]. However, one prerequisite is that the stimulus can be controlled, *in vitro* or *in vivo*; i.e., the two states of the protein complex need to be accessible in their isolated forms, to enable cross-linking of the two states, independently. Moreover, the composition of the highly purified complexes should be known, in order to generate reliable insights. Like PTMs, the binding of cofactors or ligands to protein complexes may also induce structural transitions [65]. One example of such a scenario has been demonstrated for the Arp2/3 complex, which nucleates new branches of actin filaments [66]. This complex is tightly regulated by several activating factors, which include, among others, ATP and the WASp-family of nucleation-promoting factors [67]. To unravel the molecular mechanism underlying this activation, and probe the structural effects of nucleotide and WASp binding on Arp2/3, radical footprinting MS has been successfully applied [68]. Hydroxyl radical footprinting, when coupled with MS, reveals the protein's surface accessibility [36, 37]. In practice, surface-accessible amino acid side chains react with hydroxyl radicals to form stable covalent modifications, providing a snapshot of solvent exposure. Due to the small size of the hydroxyl radicals, which is comparable to that of water, and their nonspecific reactivity, the labeling reaction is a random process that depends only on the solvent-accessible surface, and the chemical properties of the exposed amino acids. Up to 14 of the 20 side chains can be modified in protein footprinting experiments, and there are about 12 possible types of side-chain oxidation products [69]. The most common net mass shift is +16 Da, due to the formation of

a hydroxyl group in side chains. Another common event is the formation of a carbonyl group, +14 Da, in many reactive residues (e.g., Leu, Ile, Arg, Val, Pro, Lys) [69].

Following the footprinting reaction, the modification sites are identified and quantified by the same proteolysis, chromatography, and mass spectrometry techniques employed in bottom-up proteomics. However, it should be mentioned that residues with low reactivity, such as Asp, Asn, Ala, Gly, and residues whose oxidation products are difficult to detect, such as Ser and Thr, often do not provide much information, and if there is no probe site in a peptide, it will not contribute structural information. Nevertheless, this method has been shown to be capable of providing resolution at the level of single side chains [36].

The essence of the footprinting approach is based on the fact that in proteins and protein complexes, the reactive residues buried within the protein or between interfaces are protected from oxidation. However, allosteric changes in protein conformation, such as those triggered by ligand binding can induce either protection from, or enhancement of, oxidation. Thus, changes in the oxidation level of each peptide can be used to determine conformation and conformational changes of the protein. Using this approach, it was discovered that ATP and WASp binding induce conformational alterations in two subunits, Arp2 and Arp3, without significant rearrangements of the other five Arp2/3 subunits (Fig. 5) [67].

To capture additional structural changes that occur within Arp2/3 upon nucleotide and WASp binding, specifically those involving subtle rearrangements or backbone contributions, a complementary MS approach known as hydrogen–deuterium exchange was used [70]. Similar to hydroxyl radical footprinting, hydrogen–deuterium exchange (HDX) identifies the surface accessibility of proteins [33–35]. The method exploits the fact that a covalently bonded hydrogen atom can exchange with deuterium in the surrounding solvent. The rate of hydrogen–deuterium exchange depends on hydrogen bonding and solvent accessibility. Hydrogens covalently bonded to carbon essentially do not exchange. The H/D exchange rates for side chains are usually too fast to be readily determined. Thus, of all exchangeable hydrogens present in a protein, only backbone amide hydrogens are measurable for H/D exchange studies. The HDX rate of amide hydrogens varies by many orders of magnitude, depending on pH, temperature, solvent accessibility, and involvement in hydrogen bonding. In general, solvent-exposed amide hydrogens can exchange with solvent relatively rapidly, while amide protons that are inaccessible to solvent or are participating in stable hydrogen will exchange with significantly slower kinetics. As amide protons are present in every amino acid except proline, this technique offers the potential to map structural changes along the entire length of the polypeptide chain.

Ligand binding can alter the HDX rate of a protein or peptide segment directly, due to steric blockage (i.e., shielding the segment from exposure to solvent), or indirectly, through a conformational rearrangement induced by ligand binding. The influence of the ligand on the extent of exchange can be determined by exposing the assembly in the presence and absence of ligand to a deuterated environment for a predefined time period, followed by immediate addition of a quench buffer (low pH, low temperature) to stop the reaction, and minimize back-exchange with nondeuterated solvent. Typically, deuterium incorporation is monitored as a function of exposure time, covering a range from minutes to hours and sometimes days,

during which the mass of the assembly gradually increases according to the protein complex structural dynamics. The regions that possess differential exchange kinetics, due to ligand binding, are identified by proteomic analysis, in which acid-stable proteases are used for sample digestion, and the uptake of deuterium is determined by measuring the increase in mass for each peptide that can be monitored. The spatial resolution of HDX MS measurements is limited to the size of the generated peptides, on the order of 5–10 residues. Overall, changes in the rate of HDX indicated that ATP binding causes conformational rearrangements of Arp2 and Arp3 that are transmitted allosterically to the remaining Arp2/3 subunits [70].

Probing large-scale structural transitions driven by cofactors and ligand binding

The binding of ligands, metabolites or cofactors to protein complexes can also induce large-scale movement, which ranges from domain motions, to substantial transitions between folded and unfolded conformations, or even between monomeric and polymeric states [71]. These structural reorientations can be probed by native MS methods that preserve the non-covalent interaction with the associated compound during ionization and transmission of the sample into the mass spectrometer [29, 38, 39]. Such measurements require the use of volatile buffers at physiological pH, careful control of electrospray ionization conditions, and manipulation of pressure gradients within the mass spectrometer, so as to maintain the native state of the assembly.

In general, native MS is a very powerful tool for determining the stoichiometry and connectivity between subunits, as well as in defining the overall architecture of the assembly. The major advantage of this approach is that in a single spectrum, all co-existing binding states of the protein/ligand complex can be resolved simultaneously [29, 38, 39]. Yet, this method cannot identify the direct binding region within the complex, nor the localization of specific binding interfaces. The mass range of the protein complex is not a limiting factor; however, the ability to resolve a relatively slight shift in mass associated with the binding of a small ligand or cofactor to a large protein complex, is particularly challenging.

Recently, the introduction of ethylenediamine diacetate (EDAA) as a volatile buffer for native MS analysis, was shown to provide very good desolvation and highly resolved peaks, making it possible to distinguish the small shifts in mass between apo GroEL (801 kDa) and its ATP-bound states [72]. In this study, which entailed a series of titration experiments, thus determining the distribution of ATP-bound forms of GroEL at each ATP concentration, the specific allosteric mechanism by which this complex functions could be identified.

The effect of cofactor binding on the folding state of the protein can be probed by combining native MS and ion mobility (IM) measurements [40–43]. In general, IM measures the time it takes for an ion to travel through a gas-filled mobility cell under the influence of a weak electric field [40–43]. Ions with a larger surface area will experience more collisions with the buffer gas and, as a result, will take longer to traverse the drift tube, in comparison to smaller, more compact ions of the same molecular mass, which will undergo fewer collisions with the buffer gas, and hence will display greater mobility and a shorter drift time. The recorded drift times can be converted into collision cross-sections (CCSs) which, in turn, can be related to the conformational and topological arrangements of the assembly.

While relatively low-resolution structural information is provided by IM-MS measurements, this methodology benefits from the relatively small quantities of sample that are required in comparison to high-resolution structural biology methods, and from its ability to simultaneously detect multiple components and structural populations which are often present in biological assemblies [40–43].

By coupling IM with native MS, it was shown that the FAD cofactor is critical for maintaining the properly folded state of the homodimer NQO1 [73]. The broad distribution of charge states detected in the native MS spectrum of apo-NQO1, in comparison to the FAD-bound complex, suggested the existence of partially unfolded conformers (Fig. 6A and 6B). IM-MS measurements then indicated that holo-NQO1 occupies a compact structure, unlike the extended conformation of the apo form (Fig. 6C). Overall, the study demonstrated how FAD maintains the properly folded state of NQO1 that is required for sustaining its cellular stability.

Exposing the structural and functional modularity of protein complexes

During the past decade, a growing body of evidence indicated that protein complexes are not uniform in their composition and functional capacity, as was originally thought. Rather, a specific protein complex can actually encompass versatile subunit combinations, altering its composition according to the tissue or cellular compartment in which it is found, its exposure to various stimuli, or the specific time point during the cell cycle, in order to adapt to varying cellular conditions [24]. Structural mass spectrometry analysis, complemented with cellular assays, is emerging as a powerful strategy for exposing how this compositional variability regulates function [45].

A classic example demonstrating the structural plasticity and dynamics of protein assemblies entails studies focusing on the proteasome complex, the proteolytic machinery of the ubiquitin-proteasome system [74, 75]. The composition and cellular distribution of the different proteasome species were recently determined, by combining *in vivo* cross-linking, proteasome immuno-purification, and label-free quantitative proteomic MS analysis.

Generally, the aim of quantitative proteomic MS is to provide information on both the identity of proteins, and their amounts in one state or sample, relative to that of a second state or sample [76]. Unlike other methods for protein quantification label-free quantification does not employ a stable isotope-containing compound that will chemically bind to and thus label the protein; rather, it relies on the direct evaluation of MS signal intensities of naturally occurring peptides contained in a sample [76]. Using this approach, the study demonstrated that proteasome complexes are highly dynamic protein assemblies, whose activity is being regulated at different levels: by variations in the stoichiometry of bound regulators, in the composition of catalytic subunits and associated proteins, and in the rate of complex assembly [74, 75].

Another example that demonstrates how structural modularity influences the functional capacity of a protein complex concerns analysis of the Arp2/3 actin-nucleating complex [77]. By combining native mass spectrometry with various cell biology-based assays, this study revealed that the Arp2/3 complex displays a variable structural composition (Fig. 7),

which serves to regulate its subcellular localization. Specifically, it was shown that the actin-nucleating core of the complex (i.e., Arp2, Arp3, p34), interacts with different anchoring domains that direct the assembly localization to specific subcellular domains: If the monomeric protein vinculin binds the functional nucleating core of the complex, it is directed to focal adhesion regions, whereas if a trimeric complex composed of the canonical Arp2/3 subunits, p16, p20 and p41, is associated with the core the complex is directed to lamellipodia sites [78].

In the case of eukaryotic initiation factor 2B (eIF2B), which controls protein synthesis, the subunit arrangement within the complex was shown to facilitate functional regulation. Specifically, using native mass spectrometry, it was demonstrated that although the complex was generally thought to be a pentamer of five non-identical subunits (α - ϵ), it is actually a decamer; i.e., a dimer of eIF2B($\beta\gamma\delta\epsilon$) tetramers, which is stabilized by two copies of eIF2B α . The levels of eIF2B α dictate the cellular proportions of the eIF2B($\alpha\beta\gamma\delta\epsilon$) and eIF2B($\beta\gamma\delta\epsilon$) complexes, with important implications for the regulation of translation in individual cell types [79, 80].

An additional example of structural modularity is the CRISPR–Cas (clustered regularly interspaced short palindromic repeats–CRISPR-associated proteins) complex, a prokaryotic adaptive immune system that targets and degrades invading genetic elements [81]. CRISPR–Cas assemblies are highly diverse, probably due to the rapid evolution of immune systems: they are classified into three main types (I–III), characterized by the presence of a signature Cas protein: Cas3, Cas9, and Cas10 for types I, II, and III, respectively. Type I and Type II complexes target DNA, while the Type III-B is unique in that it targets RNA for degradation. Despite these differences, recent studies involving native MS and IM measurements in combination with electron microscopy have highlighted key similarities in the architecture of Type I and Type III complexes [82–84].

In these studies, subcomplexes were generated to determine the topology of the complexes, either in solution (by manipulating the pH, increasing the ionic strength, or adding organic solvents to the buffer), or in the gas phase (by collisional activation methods). MS and MS/MS analyses were then applied, to identify the composition of the subcomplexes generated. Subsequently, the list of assigned subcomplexes was used to build an interaction map. To explore the spatial arrangement of the subunits, IM-MS was used to measure the collision cross sections for both the intact complex and the various subcomplexes. These values were then used as structural constraints for computational molecular modeling approaches. Overall, although the Type III-associated proteins are phylogenetically distinct from those in the Type I systems, striking architectural similarities were observed between the two types of generated structures, suggesting that these complexes evolved from a common ancestor [82–84].

Yet, dissecting the static structural modularity of protein complexes is not enough, as diversification processes are often dynamic. For example, human immunoglobulin G4 (IgG4) monoclonal antibodies undergo a physiological reaction of Fab-arm exchange (FAE) in which half-molecules of two humanized IgG4 antibodies recombine to form a chimeric antibody with two different specificities [85]. Recently the *in-vitro* reaction was monitored

in real-time by combining native MS with IM [86]. The study took advantage of both, the fast time scale of MS analysis (ms), and the ability to simultaneously resolve the entire distribution of co-existing states in a single spectrum. Although, native MS has been used in various studies to follow the dynamics and kinetics of protein complex formation (see review [87]), this study also exploited the separation capabilities of the IM approach. Specifically, the various glycosylated species of the newly-formed chimeric antibodies were distinguished by IM separation as the mass differences of the diverse species was too small to be resolved by mass measurements per se. Overall, this study demonstrates how the formation of structural heterogeneity can be monitored in a time-resolved manner [86].

Summary and conclusions

Here, we have demonstrated how distinct structural MS approaches can expose the various levels of diversity hidden within the proteome (Fig. 1). We emphasized the different types of information that can be gained from each individual technique (Fig. 2, Box 1), and how their combination provides enhanced insights into the structural and functional complexity of protein assemblies. We anticipate that in the coming years, the urge to tackle this challenging, ever-changing topic will trigger further advances, through developments in methodological approaches, instrumentation, and computational analysis. Therefore, we would like to end this review by offering a perspective on the possible array of future developments the field may undergo.

i Computational tools for data analysis—While the fields of bottom-up and quantitative proteomics have been enhanced by rapid software development for automated data analysis, appropriate computational tools for the other MS-based approaches are still lacking. For example, native MS results are still largely interpreted and validated manually. Bottlenecks also exist in analyzing the complex datasets generated by chemical cross-linking, radical footprinting, and hydrogen exchange. In these approaches, the enormous number of possible combinations for peptide matches complicates interpretation of the data, and software that enables fully automated analysis of MS and MS/MS data, has not yet been developed. This also holds true for top-down proteomics, in which, due to the large numbers of modification sites and many possible PTMs potentially occurring on intact proteins, further advances are required for adequate computational analysis [88]. Thus, we anticipate that novel software platforms will be developed in the future to enable fast, global, and automated processing of raw data. Moreover, it is expected that modeling programs will continue to improve, toward the development of a single software platform that integrates the various structural restraints obtained by the different MS-based techniques for high-resolution generation of structural models [89, 90].

ii Analyzing the multi-level diversity of protein complexes within a single instrument—Characterizing the molecular variability of protein complexes could be greatly simplified if all stages of analysis could be performed by means of a single measurement, using one instrument. Such a methodology would require efficient use of tandem MS across a wide range of mass-to-charge ratios, wherein each stage of selection and fragmentation would expose a different layer of complexity. The resulting MS spectrum would depict the multiple, co-existing states of the intact assembly. MS/MS data would

reveal the heterogeneity existing within the individual subunits, while MS/MS/MS recordings could be used to define PTM configuration and subunit identity. Progress in this direction is already taking place [91, 92], paving the way towards a multi-level analysis, which would begin with a multi-component protein complex, and directly yield subunit identities and PTM maps.

iii Correlation between MS-based and cellular results—Naturally, a large number of structural MS methods require prior isolation of the protein complex that is being analyzed. Therefore, insights obtained by such investigations should encompass validation by various cell-based approaches, to correlate between the *in vitro* findings and the *in vivo* state. A promising direction could be the development of strategies that would enable a direct link between the MS analysis and the biological environment. Some progress is already being made, as exemplified in a recent study demonstrating how global protein structural transitions in response to metabolic perturbation can be captured by MS in unfractionated cell extracts [93]. This method couples limited proteolysis with targeted proteomic tools, enabling both pronounced and subtle protein conformational changes to be probed.

iv Coupling MS with other structural biology methods—Here, we have described how the field of structural MS is in a constant state of flux, and offered some novel approaches that could be developed for future, more thorough investigation. When examining challenging biological systems, however, coupling the information obtained from MS-based measurements with those obtained by other structural biology techniques, may be the only way forward. Advances toward an integrative mode of analysis require the development of designated platforms for hybrid acquisitions. Such developments are starting to take place, and a number of examples of integrated experiments have been reported, such as the correlation of native MS with electron microscopy [94], fluorescence measurements [95–97], and NMR [98]. We expect that the next era of technology will see novel, predesigned tools capable of linking MS with structural biology approaches, thereby transforming such hybrid analyses into standard routine.

References

- [1]. Venter JC, Adams MD, Myers EW, Li PW, et al. The sequence of the human genome. *Science*. 2001; 291:1304–1351. [PubMed: 11181995]
- [2]. International Human Genome Sequencing, C. Finishing the euchromatic sequence of the human genome. *Nature*. 2004; 431:931–945. [PubMed: 15496913]
- [3]. Pertea M, Salzberg SL. Between a chicken and a grape: estimating the number of human genes. *Genome Biol*. 2010; 11:206–213. [PubMed: 20441615]
- [4]. Brett D, Pospisil H, Valcarcel J, Reich J, Bork P. Alternative splicing and genome complexity. *Nat Genet*. 2002; 30:29–30. [PubMed: 11743582]
- [5]. Jensen ON. Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol*. 2006; 7:391–403. [PubMed: 16723975]
- [6]. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 2001; 409:928–933. [PubMed: 11237013]

- [7]. Yates CM, Sternberg MJ. The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. *J Mol Biol.* 2013; 425:3949–3963. [PubMed: 23867278]
- [8]. Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet.* 2004; 38:615–643. [PubMed: 15568988]
- [9]. Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc Biol Sci.* 2012; 279:5048–5057. [PubMed: 22977152]
- [10]. Wang ET, Sandberg R, Luo S, Khrebukova I, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature.* 2008; 456:470–476. [PubMed: 18978772]
- [11]. Melamud E, Moulton J. Structural implication of splicing stochasticity. *Nucleic Acids Res.* 2009; 37:4862–4872. [PubMed: 19528068]
- [12]. Bazykin GA, Kochetov AV. Alternative translation start sites are conserved in eukaryotic genomes. *Nucleic Acids Res.* 2011; 39:567–577. [PubMed: 20864444]
- [13]. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol.* 2003; 21:255–261. [PubMed: 12610572]
- [14]. Walsh, C. Post translational modification of proteins: Expanding nature's inventory. Roberts and Company Publishers; 2006.
- [15]. Jungblut PR, Holzhutter HG, Apweiler R, Schluter H. The speciation of the proteome. *Chem Cent J.* 2008; 2:16–25. [PubMed: 18638390]
- [16]. Jungblut PR, Holzhutter HG, Apweiler R, Schluter H. The speciation of the proteome. *Chem Cent J.* 2008; 2:16. [PubMed: 18638390]
- [17]. Wrabl JO, Gu J, Liu T, Schrank TP, et al. The role of protein conformational fluctuations in allostery, function, and evolution. *Biophys Chem.* 2011; 159:129–141. [PubMed: 21684672]
- [18]. Schmidt C, Robinson CV. Dynamic protein ligand interactions--insights from MS. *FEBS J.* 2014; 281:1950–1964. [PubMed: 24393119]
- [19]. Munoz J, Heck AJ. From the human genome to the human proteome. *Angew Chem Int Ed Engl.* 2014; 53:10864–10866. [PubMed: 25079383]
- [20]. Smith LM, Kelleher NL. Proteoform: a single term describing protein complexity. *Nat Methods.* 2013; 10:186–187. [PubMed: 23443629]
- [21]. Kelleher NL. A cell-based approach to the human proteome project. *J Am Soc Mass Spectrom.* 2012; 23:1617–1624. [PubMed: 22976808]
- [22]. von Mering C, Krause R, Snel B, Cornell M, et al. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature.* 2002; 417:399–403. [PubMed: 12000970]
- [23]. Basha E, O'Neill H, Vierling E. Small heat shock proteins and alpha-crystallins: dynamic proteins with flexible functions. *Trends Biochem Sci.* 2012; 37:106–117. [PubMed: 22177323]
- [24]. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature.* 1999; 402:C47–52. [PubMed: 10591225]
- [25]. Wang X, Yen J, Kaiser P, Huang L. Regulation of the 26S proteasome complex during oxidative stress. *Sci Signal.* 2010; 3:ra88. [PubMed: 21139140]
- [26]. Zovkic IB, Paulukaitis BS, Day JJ, Etikala DM, Sweatt JD. Histone H2A.Z subunit exchange controls consolidation of recent and remote memory. *Nature.* 2014; 515:582–586. [PubMed: 25219850]
- [27]. Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science.* 2006; 312:212–217. [PubMed: 16614208]
- [28]. Snijder J, Heck AJ. Analytical approaches for size and mass analysis of large protein assemblies. *Annu Rev Anal Chem.* 2014; 7:43–64.
- [29]. Benesch JL, Ruotolo BT, Simmons DA, Robinson CV. Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem Rev.* 2007; 107:3544–3567. [PubMed: 17649985]
- [30]. Watson, JT., Sparkman, OD. Introduction to mass spectrometry. Instrumentation, applications and strategies for data interpretation. John Wiley & Sons, Ltd; 2007.
- [31]. Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem.* 2011; 80:273–299. [PubMed: 21548781]

- [32]. Maiolica A, Junger MA, Ezkurdia I, Aebersold R. Targeted proteome investigation via selected reaction monitoring mass spectrometry. *J Proteomics*. 2012; 75:3495–3513. [PubMed: 22579752]
- [33]. Wei H, Mo J, Tao L, Russell RJ, et al. Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications. *Drug Discov Today*. 2014; 19:95–102. [PubMed: 23928097]
- [34]. Konermann L, Pan J, Liu YH. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem Soc Rev*. 2011; 40:1224–1234. [PubMed: 21173980]
- [35]. Englander SW. Hydrogen exchange and mass spectrometry: A historical perspective. *J Am Soc Mass Spectrom*. 2006; 17:1481–1489.
- [36]. Kislar JG, Chance MR. Future directions of structural mass spectrometry using hydroxyl radical footprinting. *J Mass Spectrom*. 2010; 45:1373–1382. [PubMed: 20812376]
- [37]. Konermann L, Pan Y, Stocks BB. Protein folding mechanisms studied by pulsed oxidative labeling and mass spectrometry. *Curr Opin Struct Biol*. 2011; 21:634–640. [PubMed: 21703846]
- [38]. Sharon M. Biochemistry. Structural MS pulls its weight. *Science*. 2013; 340:1059–1060. [PubMed: 23723227]
- [39]. Heck AJ. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods*. 2008; 5:927–933. [PubMed: 18974734]
- [40]. Jurneczko E, Barran PE. How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst*. 2011; 136:20–28. [PubMed: 20820495]
- [41]. Ruotolo BT, Benesch JL, Sandercock AM, Hyung SJ, Robinson CV. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat Protoc*. 2008; 3:1139–1152. [PubMed: 18600219]
- [42]. Michaelevski I, Kirshenbaum N, Sharon M. T-wave ion mobility-mass spectrometry: basic experimental procedures for protein complex analysis. *J Vis Exp*. 2010:1985. [PubMed: 20729801]
- [43]. Lanucara F, Holman SW, Gray CJ, Eyers CE. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat Chem*. 2014; 6:281–294. [PubMed: 24651194]
- [44]. Sinz A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev*. 2006; 25:663–682. [PubMed: 16477643]
- [45]. Walzthoeni T, Leitner A, Stengel F, Aebersold R. Mass spectrometry supported determination of protein complex structure. *Curr Opin Struct Biol*. 2013; 23:252–260. [PubMed: 23522702]
- [46]. Catherman AD, Skinner OS, Kelleher NL. Top down proteomics: facts and perspectives. *Biochem Biophys Res Commun*. 2014; 445:683–693. [PubMed: 24556311]
- [47]. Lossel P, Snijder J, Heck AJ. Boundaries of mass resolution in native mass spectrometry. *J Am Soc Mass Spectrom*. 2014; 25:906–917. [PubMed: 24700121]
- [48]. Leitner A, Joachimiak LA, Unverdorben P, Walzthoeni T, et al. Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc Natl Acad Sci U S A*. 2014; 111:9455–9460. [PubMed: 24938783]
- [49]. Zhang H, Cui W, Wen J, Blankenship RE, Gross ML. Native electrospray and electron-capture dissociation FTICR mass spectrometry for top-down studies of protein assemblies. *Anal Chem*. 2011; 83:5598–5606. [PubMed: 21612283]
- [50]. Rose RJ, Damoc E, Denisov E, Makarov A, Heck AJ. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat Methods*. 2012; 9:1084–1086. [PubMed: 23064518]
- [51]. Yang Y, Barendregt A, Kamerling JP, Heck AJ. Analyzing protein micro-heterogeneity in chicken ovalbumin by high-resolution native mass spectrometry exposes qualitatively and semi-quantitatively 59 proteoforms. *Anal Chem*. 2013; 85:12037–12045. [PubMed: 24229052]
- [52]. Rosati S, van den Bremer ET, Schuurman J, Parren PW, et al. In-depth qualitative and quantitative analysis of composite glycosylation profiles and other micro-heterogeneity on intact monoclonal antibodies by high-resolution native mass spectrometry using a modified Orbitrap. *MAbs*. 2013; 5:917–924. [PubMed: 23995615]

- [53]. Sidoli S, Cheng L, Jensen ON. Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. *J Proteomics*. 2012; 75:3419–3433. [PubMed: 22234360]
- [54]. Martin L, Latypova X, Terro F. Post-translational modifications of tau protein: implications for Alzheimer's disease. *Neurochem Int*. 2011; 58:458–471. [PubMed: 21215781]
- [55]. Mollapour M, Neckers L. Post-translational modifications of Hsp90 and their contributions to chaperone regulation. *Biochim Biophys Acta*. 2012; 1823:648–655. [PubMed: 21856339]
- [56]. Tran JC, Zamdborg L, Ahlf DR, Lee JE, et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*. 2011; 480:254–258. [PubMed: 22037311]
- [57]. Rozen S, Tieri A, Ridner G, Stark AK, et al. Exposing the subunit diversity within protein complexes: a mass spectrometry approach. *Methods*. 2013; 59:270–277. [PubMed: 23296018]
- [58]. Seet BT, Dikic I, Zhou MM, Pawson T. Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol*. 2006; 7:473–483. [PubMed: 16829979]
- [59]. Lothrop AP, Torres MP, Fuchs SM. Deciphering post-translational modification codes. *FEBS Lett*. 2013; 587:1247–1257. [PubMed: 23402885]
- [60]. Xin F, Radivojac P. Post-translational modifications induce significant yet not extreme changes to protein structure. *Bioinformatics*. 2012; 28:2905–2913. [PubMed: 22947645]
- [61]. Perutz MF. Mechanisms of cooperativity and allosteric regulation in proteins. *Q Rev Biophys*. 1989; 22:139–237. [PubMed: 2675171]
- [62]. Schmidt C, Robinson CV. A comparative cross-linking strategy to probe conformational changes in protein complexes. *Nat Protoc*. 2014; 9:2224–2236. [PubMed: 25144272]
- [63]. Schmidt C, Zhou M, Marriott H, Morgner N, et al. Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation. *Nat Commun*. 2013; 4:1985–1995. [PubMed: 23756419]
- [64]. Leitner A, Walzthoeni T, Kahraman A, Herzog F, et al. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics*. 2010; 9:1634–1649. [PubMed: 20360032]
- [65]. Whitford, D. *Proteins: structure and function*. J. Wiley & Sons; 2005.
- [66]. Mullins RD, Heuser JA, Pollard TD. The interaction of Arp2/3 complex with actin: nucleation, high affinity pointed end capping, and formation of branching networks of filaments. *Proc Natl Acad Sci U S A*. 1998; 95:6181–6186. [PubMed: 9600938]
- [67]. Goley ED, Rodenbusch SE, Martin AC, Welch MD. Critical conformational changes in the Arp2/3 complex are induced by nucleotide and nucleation promoting factor. *Mol Cell*. 2004; 16:269–279. [PubMed: 15494313]
- [68]. Kiselar JG, Mahaffy R, Pollard TD, Almo SC, Chance MR. Visualizing Arp2/3 complex activation mediated by binding of ATP and WASp using structural mass spectrometry. *Proc Natl Acad Sci U S A*. 2007; 104:1552–1557. [PubMed: 17251352]
- [69]. Takamoto K, Chance MR. Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annu Rev Biophys Biomol Struct*. 2006; 35:251–276. [PubMed: 16689636]
- [70]. Zencheck WD, Xiao H, Nolen BJ, Angeletti RH, et al. Nucleotide- and activator-dependent structural and dynamic changes of arp2/3 complex monitored by hydrogen/deuterium exchange and mass spectrometry. *J Mol Biol*. 2009; 390:414–427. [PubMed: 19298826]
- [71]. Petsko, G., Ringe, D. *Protein structure and function (Primers in biology)*. OUP Oxford; 2008.
- [72]. Dyachenko A, Gruber R, Shimon L, Horovitz A, Sharon M. Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proc Natl Acad Sci U S A*. 2013; 110:7235–7239. [PubMed: 23589876]
- [73]. Moscovitz O, Tsvetkov P, Hazan N, Michaelevski I, et al. A mutually inhibitory feedback loop between the 20S proteasome and its regulator, NQO1. *Mol Cell*. 2012; 47:76–86. [PubMed: 22793692]
- [74]. Fabre B, Lambour T, Delobel J, Amalric F, et al. Subcellular distribution and dynamics of active proteasome complexes unraveled by a workflow combining in vivo complex cross-linking and quantitative proteomics. *Mol Cell Proteomics*. 2013; 12:687–699. [PubMed: 23242550]

- [75]. Fabre B, Lambour T, Garrigues L, Ducoux-Petit M, et al. Label-free quantitative proteomics reveals the dynamics of proteasome complexes composition and stoichiometry in a wide range of human cell lines. *J Proteome Res.* 2014; 13:3027–3037. [PubMed: 24804812]
- [76]. Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods.* 2012; 9:555–566. [PubMed: 22669653]
- [77]. Svitkina TM, Borisy GG. Arp2/3 complex and actin depolymerizing factor/cofilin in dendritic organization and treadmilling of actin filament array in lamellipodia. *J Cell Biol.* 1999; 145:1009–1026. [PubMed: 10352018]
- [78]. Chorev DS, Moscovitz O, Geiger B, Sharon M. Regulation of focal adhesion formation by a vinculin-Arp2/3 hybrid complex. *Nat Commun.* 2014; 5:3758. [PubMed: 24781749]
- [79]. Gordiyenko Y, Schmidt C, Jennings MD, Matak-Vinkovic D, et al. eIF2B is a decameric guanine nucleotide exchange factor with a gamma2epsilon2 tetrameric core. *Nat Commun.* 2014; 5:3902–3903. [PubMed: 24852487]
- [80]. Wortham NC, Martinez M, Gordiyenko Y, Robinson CV, Proud CG. Analysis of the subunit organization of the eIF2B complex reveals new insights into its structure and regulation. *FASEB J.* 2014; 28:2225–2237. [PubMed: 24532666]
- [81]. van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol.* 2014; 12:479–492. [PubMed: 24909109]
- [82]. Rouillon C, Zhou M, Zhang J, Politis A, et al. Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol Cell.* 2013; 52:124–134. [PubMed: 24119402]
- [83]. van Duijn E, Barbu IM, Barendregt A, Jore MM, et al. Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced short-palindromic-repeats (CRISPR)-associated protein complexes from *Escherichia coli* and *Pseudomonas aeruginosa*. *Mol Cell Proteomics.* 2012; 11:1430–1441. [PubMed: 22918228]
- [84]. Staals RH, Agari Y, Maki-Yonekura S, Zhu Y, et al. Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell.* 2013; 52:135–145. [PubMed: 24119403]
- [85]. van der Neut Kolfschoten M, Schuurman J, Losen M, Bleeker WK, et al. Anti-inflammatory activity of human IgG4 antibodies by dynamic Fab arm exchange. *Science.* 2007; 317:1554–1557. [PubMed: 17872445]
- [86]. Debaene F, Wagner-Rousset E, Colas O, Ayoub D, et al. Time resolved native ion-mobility mass spectrometry to monitor dynamics of IgG4 Fab arm exchange and "bispecific" monoclonal antibody formation. *Anal Chem.* 2013; 85:9785–9792. [PubMed: 24007193]
- [87]. Ben-Nissan G, Sharon M. Capturing protein structural kinetics by mass spectrometry. *Chem Soc Rev.* 2011; 40:3627–3637. [PubMed: 21547331]
- [88]. Gingras AC, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol.* 2007; 8:645–654. [PubMed: 17593931]
- [89]. Politis A, Stengel F, Hall Z, Hernandez H, et al. A mass spectrometry-based hybrid method for structural modeling of protein complexes. *Nat Methods.* 2014; 11:403–406. [PubMed: 24509631]
- [90]. Rey M, Sarpe V, Burns KM, Buse J, et al. Mass spec studio for integrative structural biology. *Structure.* 2014; 22:1538–1548. [PubMed: 25242457]
- [91]. Belov ME, Damoc E, Denisov E, Compton PD, et al. From protein complexes to subunit backbone fragments: a multi-stage approach to native mass spectrometry. *Anal Chem.* 2013; 85:11163–11173. [PubMed: 24237199]
- [92]. Rathore D, Dodds ED. Collision-induced release, ion mobility separation, and amino acid sequence analysis of subunits from mass-selected noncovalent protein complexes. *J Am Soc Mass Spectrom.* 2014; 25:1600–1609. [PubMed: 25001382]
- [93]. Feng Y, De Franceschi G, Kahraman A, Soste M, et al. Global analysis of protein structural changes in complex proteomes. *Nat Biotechnol.* 2014; 32:1036–1044. [PubMed: 25218519]
- [94]. Benesch JL, Ruotolo BT, Simmons DA, Barrera NP, et al. Separating and visualising protein assemblies by means of preparative mass spectrometry and microscopy. *J Struct Biol.* 2010; 172:161–168. [PubMed: 20227505]

- [95]. Daly S, Poussigue F, Simon AL, MacAleese L, et al. Action-FRET: Probing the molecular conformation of mass-selected gas-phase peptides with Forster resonance energy transfer detected by acceptor-specific fragmentation. *Anal Chem*. 2014; 86:8798–8804. [PubMed: 25073016]
- [96]. Iavarone AT, Patriksson A, van der Spoel D, Parks JH. Fluorescence probe of Trp-cage protein conformation in solution and in gas phase. *J Am Chem Soc*. 2007; 129:6726–6735. [PubMed: 17487969]
- [97]. Bian Q, Forbes MW, Talbot FO, Jockusch RA. Gas-phase fluorescence excitation and emission spectroscopy of mass-selected trapped molecular ions. *Phys Chem Chem Phys*. 2010; 12:2590–2598. [PubMed: 20200735]
- [98]. Baldwin AJ, Hilton GR, Lioe H, Bagneris C, et al. Quaternary dynamics of alphaB-crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus. *J Mol Biol*. 2011; 413:310–320. [PubMed: 21839749]
- [99]. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng*. 2009; 11:49–79. [PubMed: 19400705]
- [100]. Han X, Aslanian A, Yates JR 3rd. Mass spectrometry for proteomics. *Curr Opin Chem Biol*. 2008; 12:483–490. [PubMed: 18718552]
- [101]. Hart-Smith, G., Blanksby, SJ. *Mass Spectrometry in Polymer Chemistry*. Barner-Kowollik, C.Gruendling, T.Falkenhagen, J., Weidner, S., editors. Wiley-VCH Verlag & Co; Germany: 2012. p. 532

Box 1**The structural mass spectrometry toolbox**

A short description of each structural MS approach is provided. Instruments commonly used for every method are enclosed in parentheses, and a comparison of their performance and characteristics can be found in [99–101]

- *Bottom-up proteomics* - The method is based on enzymatic digestion of the sample (typically with trypsin) into peptides, prior to MS analysis. The generated peptides may first be separated by liquid chromatography, followed by two stages of MS analysis. In the first (MS), the masses of the intact peptides are determined; in the second (MS/MS), these peptide ions are fragmented to produce information on the identity and sequence of the protein, as well as its PTMs [linear ion trap (LIT), quadrupole Time-of-Flight (QToF), linear trap quadrupole (LTQ)-Orbitrap, quadrupole (Q)-Orbitrap, quadrupole Fourier transform ion cyclotron resonance (Q-FTICR), LIT-FTICR].
- *Quantitative proteomics* - This method yields information not only on the identity of proteins within a sample, but also on their relative or absolute quantity. Relative quantitation methods use stable isotope or isobaric tag labeling to introduce signature mass tags to proteins/peptides, or label-free approaches, to compare protein or peptide abundance among samples. In the latter, the spiking of unlabeled samples with known concentrations of isotopically labeled synthetic peptides can yield absolute quantitation of target peptides [triple quadrupole (TQ), QToF, Q-Orbitrap, LTQ-Orbitrap].
- *Hydrogen/deuterium exchange (HDX) mass spectrometry* - This technique measures the rate of exchange between hydrogen and deuterium atoms, or vice versa. In general, hydrogens with a low degree of solvent accessibility, or those involved in stable hydrogen bonding, will undergo exchange much more slowly than those that do not satisfy either of these conditions. In practice, hydrogen/deuterium exchange labeling approaches are combined with proteolysis, under conditions that maintain the labeling information. Subsequently, MS measurements enable the determination of the specific location of the deuteration, due to the associated shift(s) in mass [QToF, Q-Orbitrap, LTQ-Orbitrap].
- *Radical footprinting mass spectrometry* - This method maps solvent-accessible regions by chemical labeling of the macromolecule surface. The most popular reagent for footprinting is the hydroxyl radical, which reacts with solvent-accessible amino acid side-chains; after digestion, the modified sites are identified by proteomic analysis, due to the associated shifts in mass [QToF, Q-Orbitrap, LTQ-Orbitrap].
- *Cross-linking mass spectrometry* - Chemical cross-linking enables formation of a set of structurally defined interactions, by covalently connecting pairs of

functional groups within a protein or a protein assembly. Typically, subsequent mass spectrometry analysis is performed in a bottom-up fashion that includes digestion of the cross-linked proteins, and proteomic identification of the cross-linked products. The location of the created cross-links imposes a distance constraint that serves as the basis for deducing low-resolution, three-dimensional structures [QToF, Q-Orbitrap, LTQ-Orbitrap, LIT-FTICR].

- *Top-down proteomics* - In the top-down approach, intact proteins are introduced into the gas phase, and then fragmented and analyzed within the mass spectrometer, yielding the molecular mass of the intact protein, as well as protein fragment ions. This method is used to characterize the specific molecular form of the protein that results from combinations of genetic variation, alternative splicing, and PTMs [FTICR, LTQ-Orbitrap, LTQ-FTICR].
- *Native mass spectrometry* - This approach is based on the ability to transfer intact protein complexes to the gas phase, while maintaining weak, non-covalent interactions between protein subunits and associated biomolecules such as DNA, cofactors and ligands. Structural insights into assembly composition, stoichiometry, and architecture are gained by extrapolating information from mass measurements of both the intact protein complex and smaller subcomplexes, generated either in solution or within the mass spectrometer [high-mass QToF, extended mass range (EMR) Orbitrap].
- *Ion-mobility mass spectroscopy* - The method measures the time it takes for an ion to travel through a tube densely filled with an inert gas. The ion's transit time is dependent not only on mass and charge, but also on the overall shape of the analyzed protein complex: an assembly with a large volume will experience more collisions with the gas, and consequently travel more slowly than a complex with the same mass, but a more compact structure. The measured drift times can then be converted into collision cross-sections which, in turn, can be related to the conformation of the analyzed assembly [hybrid quadrupole/ion mobility separator/orthogonal-TOF].

Figure 1. Protein diversity is generated by a combination of processes that occur at the DNA, RNA and protein levels.

On the DNA and RNA levels, protein diversity can arise from non-synonymous single nucleotide polymorphisms, gene duplications, and alternative splicing of RNA transcripts. The increase in complexity from the level of the genome to that of the proteome is further enhanced by alternative translation initiation sites, protein post-translational modifications, and interactions with co-factors and metabolites. An additional layer of complexity is added at the protein complex level, as protein subunits can assemble into multiple configurations, generating an array of complexes with various functionalities.

Figure 2. Different levels of information can be determined by distinct structural mass spectrometry approaches.

The diagram illustrates the various mass spectrometry-based methods (left side, highlighted in blue) and the structural information they provide (right side, illuminated in green). The lines in the center of the diagram connect a specific experimental approach with the relevant structural information that can be gained from it.

Figure 3. High-resolution mass spectrometry analysis of intact chicken ovalbumin exposes 59 different proteoforms.

In this study, variations caused by widespread glycosylations were identified and assigned using the spectrum of dephosphorylated ovalbumin. By comparing this spectrum (A) to unprocessed ovalbumin (B), proteoforms with either one or two phosphorylation sites could be detected. The signals in the gray box are multiplied by a factor of 10. The peak labeled with a red N* represents a proteoform known to be lacking N-terminal acetylation. The masses in (B) were decreased by 160 Da, reducing the mass of two phosphorylations, thus enabling a vertical comparison between the same proteoforms in both the phosphorylated and non-phosphorylated samples. Peaks labeled in purple are singly phosphorylated proteoforms. Adopted with permission from Yang, *et al.*, [51].

Figure 4. Post-translational modifications can alter the conformation of a protein complex, as demonstrated by comparative cross-linking analysis of the ATPase complex. (A) Schematic representation of the experimental design. The structural effect of phosphorylation was determined by the difference in relative abundance of cross-links formed in an untreated and dephosphorylated complex. (B) The structure of the interface between two adjacent α and β subunits comprising the soluble nucleotide-binding domain. Phosphosites are shown in yellow. The catalytic nucleotide-binding site (NBS) is shown in red. (C) A cartoon representing dephosphorylation-dependent conformational changes, resulting in nucleotide release from the nucleotide-binding domain (NBD). Cross-links that

decrease in number following dephosphorylation are shown as red dashed lines. Adapted with permission from Schmidt *et al.*, [63].

Figure 5. Probing the structural effects of nucleotide- and WASp binding on Arp2/3 by radical footprinting MS.

(A) Schematic representation of the crystal structure of the Arp2/3 complex. Each subunit is labeled with a different color. (B) Exposed domains within the complex that reacted with the hydroxyl radical are color-coded, according to their original color in (A). Amino acids whose oxidation rate was decreased upon binding to ATP (C) or WaSp (D) are highlighted. Adapted with permission from Kiselar *et al.*, [68].

Figure 6. The structural configuration of NQO1 is dependent on binding to its cofactor, FAD. Mass spectra of NQO1, in the presence of FAD (A), or after extraction of the FAD cofactor (apo-NQO1) (B). Apo-NQO1 displays a wide distribution of charge states (13^+ - 28^+), compared to holo-NQO1. This heterogeneous, highly charged population indicates the existence of partially unfolded conformers. Colored dots indicate the charge state series corresponding to dimeric NQO1, bound to different numbers of FAD molecules. Red stars represent ubiquitin, which was used as an internal standard. Black asterisks correspond to a nonrelated bacterial protein that is co-purified with NQO1. (C) Calculated CCS values

obtained from IM-MS experiments of apo- (red) and holo-NQO1 (blue). Theoretical CCS value for human NQO1, $3815 \text{ \AA}^2$, fits very well with the measured CCS value of holo-NQO1. The high charge states of apo-NQO1 display a significantly larger CCS value, compared to both the lower charge states (14^+ - 17^+) and holo-NQO1. Error bars represent SD for four different wave heights.

Figure 7. Native mass spectrometry reveals that multiple forms of the Arp2/3 complex co-exist. (A) A mass spectrum of hybrid Arp2/3 protein complexes purified from chicken gizzard extracts. The data indicate the co-existence of three different complexes: holo-Arp2/3 (Arp3-Arp2-p41-p34-p21-p20-p16), vinculin-Arp3-Arp2-p34-p21, and vinculin- α -actinin-Arp3-Arp2-p34. (B) A MS/MS spectrum showing the dissociation products of ions isolated at 6917 m/z , which correspond to the existence of a vinculin, Arp2, Arp3, p34, p21 complex, and the 7-mer Arp2/3. The labeling of the identified complexes is shown in (C).