Mass spectrometry: come of age for structural and dynamical biology
Justin LP Benesch¹ and Brandon T Ruotolo²

Over the past two decades, mass spectrometry (MS) has emerged as a bone fide approach for structural biology. MS can inform on all levels of protein organization, and enables quantitative assessments of their intrinsic dynamics. The key advantages of MS are that it is a sensitive, high-resolution separation technique with wide applicability, and thereby allows the interrogation of transient protein assemblies in the context of complex mixtures. Here we describe how molecular-level information is derived from MS experiments, and how it can be combined with spatial and dynamical restraints obtained from other structural biology approaches to allow hybrid studies of protein architecture and movements.

Addresses
¹ Department of Chemistry, Physical and Theoretical Chemistry Laboratory, South Parks Road, University of Oxford, Oxford OX1 3QZ, UK
² Department of Chemistry, University of Michigan, 930 N. University Avenue, Ann Arbor, MI 48109, USA

Corresponding authors: Benesch, Justin LP (justin.benesch@chem.ox.ac.uk) and Ruotolo, Brandon T (bruotolo@umich.edu)

Quantifying the oligomeric distribution of protein assemblies

In the early 1990s, studies were performed which demonstrated that protein assemblies could be transferred into the vacuum of the mass spectrometer without their dissociation [5], allowing the measurement of their mass with unprecedented precision and accuracy (Box 1). The information obtainable from just simple mass measurement of an intact protein complex is considerable, allowing the facile determination of oligomeric state, and the stoichiometry of ligand or cofactor binding. The unparalleled mass resolution of MS can however be confounded by the effects of multiple charging during nanoelectrospray ionization (nESI, the preferred method for ionizing non-covalent complexes [6]) potentially leading to proteins of significantly different masses overlapping in mass-to-charge (m/z) space.

One approach to overcome this challenge has been the development of user-guided software which facilitates the deconvolution of complex spectra [7,8]. As the peaks recorded for intact protein complexes are inherently broader than those obtained for small molecules and peptides, these approaches enable a more rigorous interpretation of such mass spectra [9,10]. Such software, therefore, represent an important step towards the fully automated MS analysis of proteins and protein assemblies.

Alternatively, an experimental strategy to simplify nESI data for complex mixtures is to reduce the charge states of the ionized components, and thereby increase the separation between adjacent peaks. Charge reduction can be achieved by using solution additives [11,12], or performing gas-phase chemistry [11,13,14]; however, currently the most commonly used strategy is tandem-MS with

Introduction

The majority of proteins exist and operate in the cell as multimeric assemblies, held together by non-covalent interactions of varying strength and lifetime [1]. These protein complexes underpin virtually all cellular processes, and therefore, understanding their interactions, structure, and dynamics is of critical importance for human health and medicine. Our knowledge of such molecular details is inexorably tied to the analytical approaches available, and their ability in overcoming the complexity of these macromolecules. However, many protein complexes of critical importance continue to confound individual structural biology techniques, often as a result of being present in low levels within mixtures, or displaying intrinsic dynamics. As such there is a growing interest in developing ‘hybrid’ strategies which combine the benefits of different technologies to characterize the most challenging protein assemblies [2].

One approach to overcome this challenge has been the development of user-guided software which facilitates the deconvolution of complex spectra [7,8]. As the peaks recorded for intact protein complexes are inherently broader than those obtained for small molecules and peptides, these approaches enable a more rigorous interpretation of such mass spectra [9,10]. Such software, therefore, represent an important step towards the fully automated MS analysis of proteins and protein assemblies.

Alternatively, an experimental strategy to simplify nESI data for complex mixtures is to reduce the charge states of the ionized components, and thereby increase the separation between adjacent peaks. Charge reduction can be achieved by using solution additives [11,12], or performing gas-phase chemistry [11,13,14]; however, currently the most commonly used strategy is tandem-MS with
Box 1 Measuring the mass and abundance of protein oligomers

Accuracy and precision

A comparison between the mass calculated from the primary sequence and that measured by means of nESI-MS is shown for a range of proteins and complexes (Chart 1). There is only a very small deviation between the experimental data (red points) and a 1:1 correlation between expected and measured mass (blue line). The small positive discrepancy results from the residual binding of solvent and buffer [6], and is less than 1.5% for all the assemblies shown here (blue points). Such adduction is the primary determinant of the width of the peaks in the mass spectra [73], and leads to an ‘effective resolving power’ which is lower than the instrumental limits of modern mass spectrometers. Nevertheless, with mass differences of ≈1% in 1 MDa routinely resolvable [73], the resolution remains far in excess of other mass-separative techniques.

Quantifying oligomeric distributions

The unrivalled separative ability of MS allows for the identification of many different components within a mixture of complexes in solution. In cases where the molecules are of similar composition and have similar biophysical properties, the relative intensities recorded in MS data can be used to extract the relative populations of each oligomer. A comparison of the mass distribution of αB-crystallin obtained by multi-angle light scattering coupled to size-exclusion chromatography (SEC-MALS) (purple) [20] and MS (orange) [21**] shows excellent agreement (Chart 2). This demonstrates how the solution phase distribution of oligomers is faithfully maintained in the gas phase, and can therefore be quantified by means of MS.

MS offers dramatically improved mass accuracy and resolution of separation over SEC-MALS; however, revealing that for αB-crystallin oligomers comprising an even number of subunits are more prevalent than those with an odd number (i.e. more 28mer and 30mer than 29mer), a ~3% difference in mass not discernible using other approaches (Chart 2).

collision-induced dissociation (CID). In this approach, oligomer ions can be selected and activated such that monomers are removed, taking with them a disproportionate amount of charge [15]. This has the effect of both increasing the separation between charge states, and disrupting the congruence of peaks from species of different mass. This approach has allowed the separation of multiple stoichiometries of the ribosomal stalk complex which were overlapping in the original mass spectra [16], the confirmation of the accurate mass measurement of hepatitis B virus capsids in the 3–4 MDa range [17], and elucidation of the stoichiometry of small molecule binding to membrane protein assemblies [18**].

As well as enabling the identification of different components within heterogeneous mixtures, MS can be used to extract the relative abundances of these components by quantifying their intensities in the primary
and tandem mass spectra. This simple strategy is analogous to ‘spectrum counting’, a methodology widely used in quantitative proteomics [3]. Oligomeric distributions derived by means of nESI-MS match those obtained using other approaches very well, thus demonstrating how the solution-phase ensemble of oligomers can be faithfully maintained in the gas phase (Box 1).

Notably, however, MS offers separations of dramatically improved mass accuracy and resolution when compared with other approaches. In the case of the molecular chaperone HSP18.1 and its interaction with luciferase, over 300 complex stoichiometries were observed and their relative abundances quantified. Interestingly, complexes containing an even number of HSP18.1 subunits were found to be approximately 20% more abundant than those with an odd number [19**], a property echoed in the distribution of oligomers populated by αB-crystallin [20,21**]. In the latter case this ‘even preference’ was found to be regulated by both post-translocation modification and solution conditions [20,22], such that variations in the free energy of dimerization could be determined [21**]. Such subtleties in oligomeric distribution can currently only be resolved and quantified by MS, and demonstrate the utility of this approach for not only providing structural insight, but also simultaneously extracting the thermodynamics which govern protein assembly.

Blueprinting multi-protein assemblies

While studies of protein assemblies at equilibrium under native conditions can reveal the stoichiometries they populate, experiments employing solution conditions which perturb non-covalent interactions can provide complementary information on their composition, connectivity and architecture. Under ‘severe’ conditions, such as high concentrations of chemical denaturant or extremes of pH, multi-protein complexes can be completely disassembled and their component chains unfolded. A mass spectrum under these conditions will allow the determination of accurate masses for the individual components. Such information is often vital for unambiguously assigning protein stoichiometry, especially in the case of low-abundance endogenous proteins for which sequence database entries are often lacking proper annotation [23].

Between these extremes of solution conditions, which result in either complete preservation or disruption of quaternary structure, variations in ionic strength, adjustment of pH, or the addition of small amounts of organic solvent can lead to partial destabilization of the oligomers [24]. Under such conditions, MS data can reveal ‘sub-complexes’, non-covalently bound ‘building blocks’ of the assembly. When multiple such sub-complexes are identified they can be combined to elucidate the two-dimensional connectivity of subunits within the oligomer. This strategy was applied to elucidate the protein–protein interactions within two important molecular machines involved in ubiquitin-mediated proteolysis, the proteasome [25] and signalosome [26**]. Furthermore, analysis of which subcomplexes are preferentially formed allows the establishment of a hierarchy of assembly, which correlates qualitatively with the size of the subunit interfaces and the evolutionary pathway of the protein complex [27]. The thermodynamic quantities governing interface strength and protein assembly can also be directly determined by examining the equilibrium distribution of complexes, sub-complexes and subunits as a function of temperature [19**].

While of great utility for mapping inter-subunit connectivity, solution-phase disruption experiments are often complemented by studies which induce dissociation in the gas phase. Though the biophysical factors governing the release of subunits from heteromeric protein oligomers during CID remain incompletely understood, those that are readily expelled are unlikely to be located in the core of the assembly [15]. This observation has been used to help elucidate the topology of the proteasome lid [28], and RNA polymerase III [29]. Furthermore, close examination of the energy profile of CID can also be used to infer the promters of protein assemblies [22,30].

The use of CID is, however, currently limited by its mechanistic underpinnings; in general exclusively monomers are expelled (irrespective of the oligomeric substructure), and only a small number thereof [15]. This pathway of gas-phase dissociation can be somewhat altered through the manipulation of the protein complex charge state selected for CID [12*,31,32], or by depositing the activation energy on a much faster timescale. One means to achieve the latter is surface-induced dissociation (SID), which, in the case of the heterohexamer toyocamycin nitrile hydratase, resulted in disassembly into its component non-covalently bound trimers [33*]. This is a particularly exciting result as it may prove SID to be a means for directly determining the building blocks of protein assemblies. Combined with the observation that gas-phase activation can lead to the fragmentation of individual monomers within a non-covalent assembly to give sequence information [12*,31,34,35*,36*], the possibility therefore emerges of gleaning information spanning from primary sequence to quaternary architecture from a single, rapid gas-phase measurement.

Obtaining three-dimensional spatial restraints

Complementary to experiments aimed at obtaining two-dimensional maps of protein–protein connectivity, there has been much interest in developing MS-based approaches for obtaining three-dimensional shape information on protein and their complexes. Spatial restraints for protein modeling can be derived from MS experiments in a number of ways. Perhaps the most intuitive approach involves chemical cross-linking of the protein.
Box 2 Measuring the size of protein oligomers

The use of IM-MS in structural biology is motivated by the general observation that the quaternary structure of proteins can be maintained in the gas phase, at least for the tens of milliseconds required for IM measurement [74], as evidenced by the correspondence between experimentally determined CCSs and those estimated from their atomic structures [75,76]. A comparison between theoretical CCSs, calculated in silico using a simple projection approximation (PA) algorithm [77], and CCS measurements for a range of proteins and complexes generated from native solution conditions [75] is shown in Chart 3. The error bars represent the variations in CCS over all charge states observed for the complexes (±2 standard deviations from the mean), and the data points are color-coded with the structures shown in the top panel. The experimental and theoretical values are very well correlated, in this case by \( CCS_{\text{exp}} = 1.14 \times CCS_{\text{PA}} \), with an RMSD of 3%. Though the precise relationship between theory and experiment depends on the specifics of how the theoretical CCS is determined, this data demonstrates that simple linear scaling can be used to relate measured and theoretical values, and therefore that IM-MS data can provide direct spatial restraints for protein topology models.

Alternative computational methods are available to estimate CCS, but are more computationally expensive and currently do not offer significantly greater accuracy for comparison with IM measurements of protein complexes than the scaled PA estimates shown in Chart 3 [78]. The PA approach has the additional advantage of supporting rapid coarse-grained protein complex representations that are of critical importance in cases where no atomic structures exist for comparison [53,79]. Protein complexes having flexible or labile structures can be difficult targets for CCS measurement, and ongoing efforts are aimed at developing generalized methods for stabilizing such structures in the absence of bulk solvent [11,80]. Additionally, the generation of models for complexes of low symmetry is currently an obstacle for IM data interpretation in the absence of structural information from other approaches. Current commercial instrumentation relies on CCS calibration with known standards [75], and has both high accuracy and resolution [81,82]. For native proteins and assemblies it appears that the apparent resolution achieved in state-of-the-art IM instrumentation is largely governed by the conformational heterogeneity of the proteins under investigation [82]. Therefore, while subtle differences in CCS (<2%) remain challenging to resolve, careful analysis of IM peak widths can allow the probing of conformational states and fluctuations of proteins [83].

In addition to such experiments where spatial restraints are determined on a local level, MS experiments can also inform on a global oligomeric level. Since the charge states of globular proteins and assemblies are correlated with solvent-exposed surface area, this information can be used as a general constraint to classify protein topology by MS [44], and to determine whether certain protein assemblies are in particularly extended or compact conformations [45]. A more explicit approach to assessing protein size is ion mobility spectrometry (IM), a technology that can be coupled directly to MS and separates protein assemblies according to their ability to traverse a pressurized ion guide under the influence of a weak electric field. The transit time of the ions is directly related to their size in terms of a rotationally averaged collision cross section (CCS) [46]. The correlation between the experimental CCSs and values calculated in silico from high-resolution structures of protein assemblies is excellent (Box 2). This demonstrates that proteins retain a ‘memory’ of their native quaternary structure in the gas phase, and thereby that IM measurements can be used to determine the native size of protein assemblies in solution.
Though a CCS represents only a single spatial restraint, the fact that size information can be obtained on all species separable in m/z makes IM-MS very attractive for the study of heterogeneous systems. As a result there has been considerable effort in studying early aggregates associated with protein deposition diseases by means of IM-MS. Investigations of amyloid-forming protein fragments [47], amyloid-β peptide [48,49], islet amyloid polypeptide [50], and β2-microglobulin [51] have shown the presence of not only multiple oligomeric states but also different conformations of each. The ability of IM-MS to separate both these sources of heterogeneity allows for the detailed characterization of oligomeric microstates, providing insight into the interplay between globular and extended oligomeric conformations on both the fibrilogenesis and cytotoxicity pathways.

Similarly, while the trajectory of virus assembly has been investigated by a variety of MS-based approaches, the topology of oligomeric intermediates has remained elusive. IM-MS has recently been applied to assess the shape of capsid-protein oligomers, generated by destabilizing intact norovirus and hepatitis B virus capsids [52]. By comparing CCS values with those calculated from atomic models, these oligomers were shown to be planar rather than globular, providing structural insight into disassembly and assembly of the capsid [52]. Measuring the size of oligomeric disassembly products can also be used to refine structural models of the corresponding intact complexes. For instance, IM measurements of two subcomplexes of the eukaryotic initiation factor 3 allowed for a two-dimensional blueprint of the complex to be refined into a partial topology model [53]. MS approaches, therefore, are capable of generating useful restraints, on both the local and global oligomeric levels, providing insight into the static structures of protein assemblies.

Monitoring dynamical motions of protein assemblies

The function of proteins is, however, not simply governed by their structures, but also by the motions they undergo. As these dynamics span secondary to quinary structure [54] and picoseconds to days [2], a variety of MS methods have been developed for their interrogation [55]. HDX-MS can reveal local fluctuations of individual protein chains [40], with amino-acid level information achievable on the minute timescale in real time [56]. Pulse-label HDX methods can achieve millisecond time resolution [40], and recent advances in pump/probe oxidative footprinting have shown the capability of accessing microsecond regimes [57]. Online approaches, in which intact proteins (rather than peptides) are injected into the mass spectrometer, have the benefit of enabling multiple species to be monitored in tandem [58]. The future combination of this with electron-mediated fragmentation raises the exciting possibility of obtaining dynamical information on co-populated, transient species at the residue level.

The separative capabilities of MS also renders it well suited to monitoring the assembly of protein complexes, simply by incubating components and obtaining mass spectra in real time. In this way the role of individual subunits and sections of sequence in governing the assembly of the 20S proteasome [59] and DNA clamp loader were elucidated [60]. A similar approach was employed to study the molecular chaperone action of HSP18.1, with the kinetics of target binding revealing a two-stage mechanism of protection [19].

Complementary to studying such assembly dynamics, the fluctuations in quaternary structure which proteins undergo at equilibrium can also be investigated by means of MS. Subunit exchange, the process in which monomers or other building blocks move between oligomers, can be studied in considerable detail by monitoring a sample containing ‘mass-labeled’ and unlabeled protein as a function of time [61]. The label can be of natural origins, such as the use of protein isoforms or homologues that differ in mass, or achieved through recombinant expression of the protein with heavy isotopes. In this way, the effect of small molecule binding on the rate of oligomeric dissociation in both transthryretin [62] and glucosamine-6-phosphate synthase [63] was elucidated. Furthermore, quantitative assessment of the quaternary dynamics suggested the presence of two conformations of HSP26 oligomers [64], revealed a pH dependence in the dissociation rates of different interfaces in αB-crystallin [21], and provided a rationale for the specificity of subunit ordering in pilus assembly [65]. In the case of two plant molecular chaperones, subunit exchange was found to proceed via the movement of dimers, revealing them as the building block of the oligomers [61]. Experiments such as these can, therefore, simultaneously inform both on the architecture of protein assemblies and their inherent dynamics.

MS in integrative structural and dynamical biology

In this article we have detailed some of the ways in which MS-based approaches can be used to obtain information regarding the structure and dynamics of protein assemblies. MS can inform from the primary to quinary levels, and on the sub-millisecond to hour timescales (Fig. 1). While other experimental and computational techniques can provide information of higher resolution, both in terms of space and time, MS has considerable benefits, three of which stand out as particularly important for structural biology. Firstly, MS is remarkably general in its applicability, allowing the study of complexes that range in terms of mass, size, solubility, flexibility, oligomeric composition, bound state, and dispersity. Additionally, it is a high-resolution separation technology that allows for
the identification, quantification, and interrogation of different components within a mixture without ensemble averaging of species present in solution. Finally, MS has very low limits of detection and quantification, enabling not only the analysis of small amounts of dilute sample extracted directly from cells but also the study of quaternary dynamics in real time.

These valuable qualities of MS make it ideally placed for integration with other structural biology approaches. Most frequently, MS data are used to provide information on oligomeric heterogeneity and stoichiometry, either as quality control, or to directly guide high-resolution analysis [66,67]. Recently, MS has been used as a purification method, allowing the deposition of mass-selected ions for electron and atomic force microscopy investigation [68]. Spatial restraints from MS can also be directly integrated with those from other structural techniques to facilitate the building of topological models of protein assemblies [69**,70]. Similarly, measurements of hierarchical protein fluctuations can be correlated between techniques, providing ‘dynamical restraints’ as to the organization of the component species [71**,**,72**]. Such integrated measurement and correlation of spatial and
Acknowledgements
The authors thank Zoe Hall and Gillian Hilton (both from University of Oxford); JLPB is a Royal Society University Research Fellow; and BTR acknowledges support from the National Institutes of Health (1-R01-GM-095832-01) and the University of Michigan.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:
- of special interest
- of outstanding interest


With Ref. [71**]. Back-to-back publications demonstrating how MS can be used to quantify oligomeric distributions and dynamics, and their interpretation in terms of interface free energies and microscopic rate constants. Correlation of these measurements with ones from NMR reveals a correlation between fluctuations in tertiary and quaternary structures, and thereby the regions of sequence responsible for protein assembly.


35. Yin S, Loo JA: Top-down mass spectrometry of supercharged native protein-ligand complexes. Int J Mass Spectrom 2011, 300:118-122. With Ref. [36]. Two studies which describe the use of MS and electron-capture dissociation to first determine the stoichiometry of protein-protein and protein-ligand interactions, and then provide information on the sequence level as to protein identity or the site of ligand binding.


43. Pan J, Han J, Borchers CH, Konemann L: Conformer-specific hydrogen exchange analysis of aβ(1-42) oligomers by top-down electron capture dissociation mass spectrometry. Anal Chem 2011, 83:5396-5399. HDX and electron-transfer dissociation MS study which demonstrates how residue-level solvent accessibility can be interrogated for different species separated in m/z, revealing insight into the fold of amyloid-β oligomers.


51. Smith DP, Radford SE, Ashcroft AE: Elongated oligomers in j2- microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. Proc Natl Acad Sci U S A 2010, 107:6794-6798. IM-MS and real-time MS studies enable the study of the structure and quaternary dynamics of the heterogeneous ensemble of oligomers formed by j2-microglobulin, and provide insights into the early stages of fibrillogenesis.

52. Uetrecht C, Barbu IM, Shoemaker GK, van Duijn E, Heck AJ: Interrogating viral capsid assembly with ion mobility-mass spectrometry. Nat Chem 2011, 3:126-132. Study that employs MS to accurately measure the mass of viral particles in excess 3 MDa, and, via destabilization of the capsids in solution, determine possible topologies of the oligomeric intermediates from IM-MS experiments.


MS is combined with other biophysical techniques to quantify the oligomeric distribution and dynamics of this polydisperse molecular chaperone.


Combined MS and electron microscopy study of Cascade, a protein assembly involved in the prokaryotic immune system, reveals its oligomeric shape and connectivity.


See annotation to Ref. [21**].

72. Carulla N, Zhou M, Arimon M, Gairi M, Giraldo E, Robinson CV, Dobson CM: Experimental characterization of disordered and ordered aggregates populated during the process of amyloid fibril formation. Proc Natl Acad Sci U S A 2009, 106:7828-7833. The kinetics of amyloid fibril formation monitored on the fibril, monomer, and residue levels by using electron microscopy, MS, and nuclear magnetic resonance spectroscopy, respectively. Temporal correlation of the experiments allows elucidation of the structural reorganizations which occur during fibrilogenesis.


With Refs. [81,82*], Three papers describing figures of merit and best practices for state-of-the-art commercial IM-MS instrumentation.


82. Zhong Y, Huang SJ, Ruotolo BT: Characterizing the resolution and accuracy of a second-generation travelling-wave ion mobility separator for biomolecular ions. Analyst 2011, 136:3534-3541. See annotation to Ref. [75*].