

# The diverse and expanding role of mass spectrometry in structural and molecular biology

Philip Lössl<sup>1,2,†</sup>, Michiel van de Waterbeemd<sup>1,2,†</sup> & Albert JR Heck<sup>1,2,\*</sup>

## Abstract

The emergence of proteomics has led to major technological advances in mass spectrometry (MS). These advancements not only benefitted MS-based high-throughput proteomics but also increased the impact of mass spectrometry on the field of structural and molecular biology. Here, we review how state-of-the-art MS methods, including native MS, top-down protein sequencing, cross-linking-MS, and hydrogen–deuterium exchange-MS, nowadays enable the characterization of biomolecular structures, functions, and interactions. In particular, we focus on the role of mass spectrometry in integrated structural and molecular biology investigations of biological macromolecular complexes and cellular machineries, highlighting work on CRISPR–Cas systems and eukaryotic transcription complexes.

**Keywords** CRISPR–Cas; ligand binding; post-translational modifications; protein complexes; RNA polymerase

DOI 10.15252/embj.201694818 | Received 19 May 2016 | Revised 25 July 2016 | Accepted 7 October 2016 | Published online 26 October 2016

The EMBO Journal (2016) 35: 2634–2657

See the Glossary for abbreviations used in this article.

## Prelude—The coming of age of biomolecular mass spectrometry

The advent of mass spectrometry (MS) as an analytical technology dates back more than a century and was made possible by the groundbreaking work on cathode rays of the *physicist* J. J. Thomson (Thomson, 1897). Thomson not only discovered the electron but, together with F. W. Aston, also measured the masses of stable isotopes of elements (Thomson, 1911; Aston, 1920). To this end, Thomson employed electron impact ionization to charge the atoms and molecules. This allowed their mass-to-charge ( $m/z$ ) ratio to be measured by monitoring their trajectories in an electric or magnetic field—mass spectrometry was born and its inventor's ingenuity is nowadays acknowledged by terming the unit for  $m/z$  the “Thomson” (Th) (Cooks & Rockwood, 1991).

MS remained in the realm of isotope physics for about half a century before it entered the domain of *chemistry*. MS with its potential to identify substances based on accurate mass measurements sparked initially the interest of the oil industry in its attempt to find ways to characterize the low molecular weight compounds present in crude oil. This challenge was addressed by pioneers like F. H. Field, J. L. Franklin, and F. W. McLafferty. They introduced and applied chemical ionization (Munson *et al*, 1964; Munson & Field, 1966; Baldwin & McLafferty, 1973) as a novel technique to bring organic molecules into the gas phase, which had previously only been feasible by electron impact ionization (Smith, 1937). Owing to both ionization techniques, the field of organic MS started to bloom, for it was now possible to determine the masses of semi-volatile molecules up to about 500 Dalton (Da). Soon after, mass determination of not only the intact organic molecule but also its specific fragment ions was enabled by conceiving improved mass spectrometric instrumentation (Beynon *et al*, 1973; Paul, 1990) and new gas-phase fragmentation techniques, such as collision-induced dissociation (Haddon & McLafferty, 1968; Jennings, 1968; Kim *et al*, 1974). The related fragmentation mechanisms were elucidated by studying the unimolecular and bimolecular chemistry of organic molecular ions. These meticulous investigations aided the interpretation of the fragment ion mass spectra, eventually allowing to deduce details about a compound's molecular structure (Levsen & Schwarz, 1976). Over the years, about 250,000 unique organic molecules were characterized and catalogued in a library of reference mass spectra (Stein, 2016). This repository may be queried when seemingly unknown compounds are to be characterized by organic MS, which nowadays has become a core analytical technology in chemical, forensic, environmental, and atmospheric sciences.

More recently, MS entered the field of cellular, molecular, and structural *biology*. The dawn of biomolecular MS can be traced to the introduction of matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) in the 1980s, for which Koichi Tanaka and John Fenn, respectively, received the 2002 Nobel Prize in Chemistry (Tanaka *et al*, 1988; Fenn *et al*, 1989). These ionization technologies, which will be further explained later on, enabled the efficient transfer of large intact biomolecules (e.g. peptides and proteins) into the gas phase as charged ions—an essential prerequisite for MS analysis. Previously developed ionization

1 Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, The Netherlands

2 Netherlands Proteomics Center, Utrecht, The Netherlands

\*Corresponding author. Tel: +31 0302536797; E-mail: a.j.r.heck@uu.nl

†These authors contributed equally to this work

**Glossary**

<b>CID</b>	collision-induced dissociation
<b>Da</b>	Dalton
<b>ECD</b>	electron capture dissociation
<b>EM</b>	electron microscopy
<b>ESI</b>	electrospray ionization
<b>ETciD</b>	electron transfer/collision-induced dissociation
<b>ETD</b>	electron transfer dissociation
<b>ETHcD</b>	electron transfer/higher-energy collisional dissociation
<b>FTICR</b>	Fourier transform ion cyclotron resonance
<b>HCD</b>	higher-energy collisional dissociation
<b>HDX</b>	hydrogen/deuterium exchange
<b>IMS</b>	ion mobility spectrometry
<b><i>m/z</i></b>	mass-to-charge ratio
<b>MALDI</b>	matrix-assisted laser desorption/ionization
<b>MS1</b>	mass spectrometric analysis of non-fragmented precursor ions
<b>MS2</b>	mass spectrometric analysis of fragment ions generated by gas-phase dissociation of precursor ions
<b>MS</b>	mass spectrometry
<b>PTM</b>	post-translational modification
<b>SAXS</b>	small-angle X-ray scattering
<b>tandem-MS</b>	mass spectrometric experiment combining MS1 analysis and subsequent MS2 analysis of a selected <i>m/z</i> range
<b>Th</b>	Thomson
<b>TOF</b>	time-of-flight

methods, by contrast, were mostly limited to small, volatile compounds.

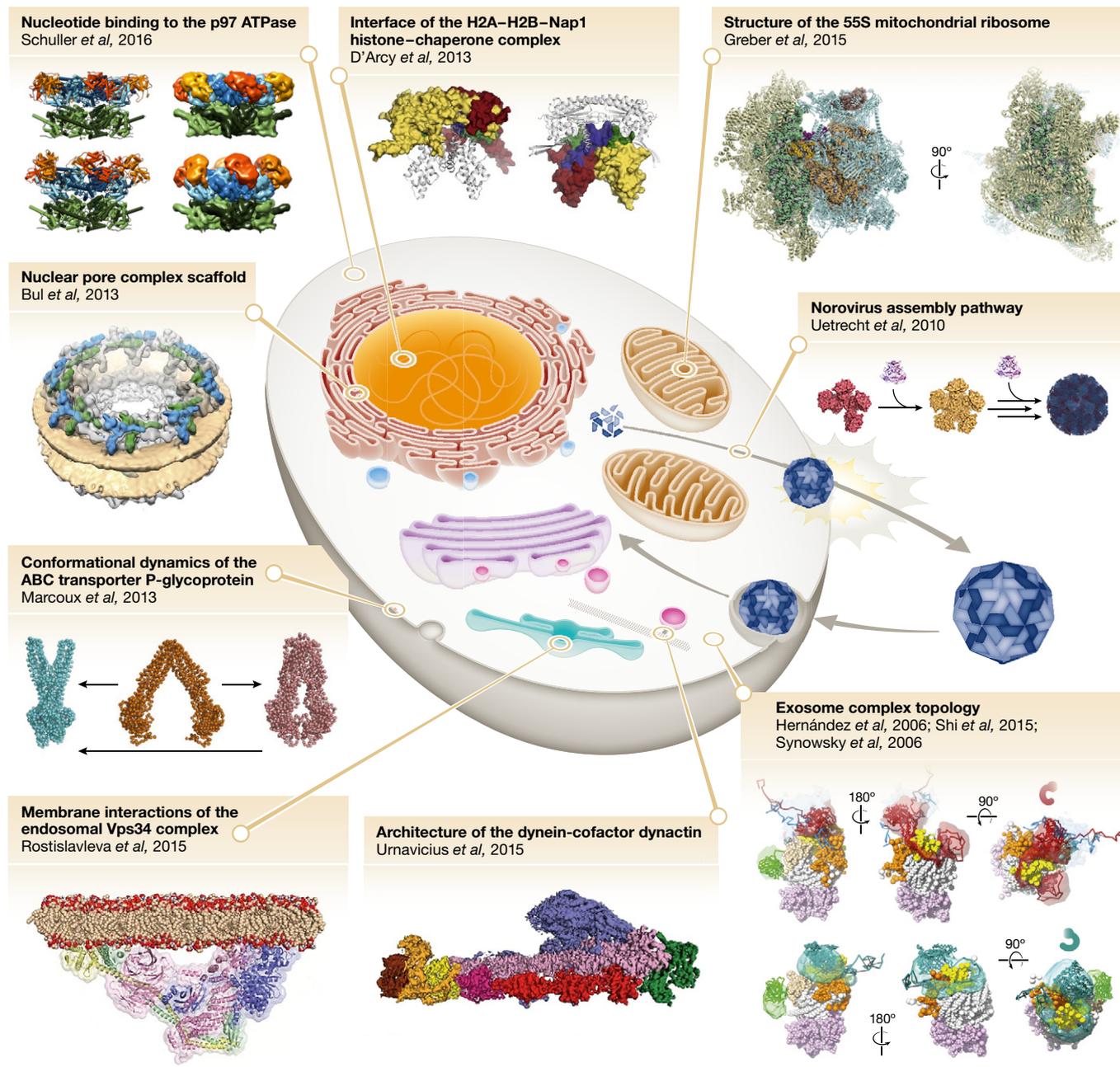
Only a few years after the MS analysis of peptides and proteins became feasible, first attempts were made to couple it with advanced separation techniques, such as miniaturized nanoflow liquid chromatography and electrophoresis (Henzel *et al*, 1993; Strupat *et al*, 1994; McCormack *et al*, 1997; Washburn *et al*, 2001). This development enhanced the throughput and dynamic range in the analysis of very complex peptide samples like those obtained from a proteolytic digest of a single protein, a protein mixture, or even all proteins within a whole-cell lysate. Nowadays, hundreds of thousands of peptides can be analyzed by combining orthogonal separation methods (e.g., reversed-phase, ion exchange and/or hydrophilic interaction chromatography) with modern mass spectrometers that perform high-speed tandem-MS experiments (see below for more details; Hebert *et al*, 2014). Interpreting these large amounts of data, especially deriving the peptide sequence from the acquired mass spectra and mapping it back onto the original proteins, posed a first major challenge but is meanwhile facilitated by sophisticated bioinformatics and statistics solutions (Cappadona *et al*, 2012; Bruce *et al*, 2013). Collectively, advanced peptide separation, rapid tandem-MS measurements, and elaborate bioinformatics analysis form the current workflow for high-throughput bottom-up proteomics (Zhang *et al*, 2013). This technique has become the standard for the system-wide analysis of all proteins present in a particular cell, body fluid, tissue, or organism (as reviewed by Cox & Mann, 2011; Bensimon *et al*, 2012; Altelaar *et al*, 2013; Richards *et al*, 2015). Recently, the progress of bottom-up proteomics has cumulated in first drafts of the human proteome (Kim *et al*, 2014; Wilhelm *et al*, 2014).

Bottom-up proteomics as a key driver for technological development has benefitted the field of biomolecular MS in general. First

and foremost, mass analyzers have become immensely more advanced, gaining performance in speed, sensitivity, selectivity, robustness, and, equally important, user-friendliness for non-specialists (Ens & Standing, 2005; Zubarev & Makarov, 2013). Next, advances in sample preparation and fractionation enabled more comprehensive and reproducible MS experiments (Tabb, 2013). Novel strategies for affinity purification (Li *et al*, 2015), protein digestion (Tsiatsiani & Heck, 2015), and peptide fragmentation (Madsen *et al*, 2010; Frese *et al*, 2012) may serve as illustrative examples here. Finally, more efficient and reliable analysis software (Cox & Mann, 2008; Bruce *et al*, 2013; Lee *et al*, 2015), public data repositories (Perez-Riverol *et al*, 2015), and tools to define and study protein interaction and signaling networks (Gavin *et al*, 2002; Hein *et al*, 2015; Huttlin *et al*, 2015) allow translating MS data into sound and meaningful biological information. These advancements significantly promoted the impact of MS on molecular and structural biology, a development that can be retraced using hydrogen–deuterium exchange (HDX)-MS of biomolecules as an example. HDX-MS has been around for about 25 years (Katta & Chait, 1991; Englander, 2006), arguably predating the advent of proteomics. However, due to limitations in hardware and analysis software, early HDX-MS studies were mostly limited to peptides, small proteins, or protein domains. Owing to the technological progress, HDX-MS can now be applied to study whole protein assemblies up to intact viruses (Wang *et al*, 2001; Lanman *et al*, 2004; Konermann *et al*, 2011; Bereszczak *et al*, 2013; Pirrone *et al*, 2015), making it a valuable tool for protein structure analysis, as we will discuss later on.

Today, MS methods are utilized in an increasing number of molecular and structural biology studies that aim for a better understanding of large biomolecular assemblies. A selection of these studies is presented in Fig 1, illustrating that biomolecular complexes from all cellular compartments and different kingdoms of life have been successfully probed by MS. In the course of this review, we will refer to several of the shown examples, as they mark the current frontiers of MS-based structural and molecular biology. Fundamentally, these approaches can be divided into peptide-centric and protein-centric strategies (Fig 2). Peptide-centric strategies comprise surface labeling approaches, such as HDX-MS and covalent labeling-MS, as well as cross-linking-MS and limited proteolysis-MS. All of these methods probe biomolecular structures in solution, utilizing the previously described bottom-up proteomics workflow to detect, at the peptide level, the results of the in-solution experiment (Fig 2, bottom). Protein-centric strategies, on the other hand, enable the mass spectrometric characterization of intact proteins and biomolecular assemblies, which can be additionally manipulated in the gas phase (Fig 2, top). Analytical targets of protein-centric MS range from small intact histones (10–20 kDa) over antibodies (150 kDa) to complex cellular machineries (500–20,000 kDa) such as proteasomes, ribosomes, or even full virus assemblies (Marcoux & Robinson, 2013; Snijder & Heck, 2014).

The applications of biomolecular MS (summarized in Fig 2) illustrate that a mass spectrometer may nowadays be considered a full-fledged “biochemical laboratory”, in analogy to a claim made 35 years ago that MS has become a “complete chemical laboratory” (Porter *et al*, 1981). In this review, we aim to familiarize non-mass spectrometrists with the inner workings and analytical opportunities



**Figure 1. MS-based molecular and structural biology reaches out over all cellular compartments and kingdoms of life.**

Shown are exemplary results of studies that employed biomolecular MS techniques. Clockwise (from bottom left): membrane interactions of the endosomal Vps34 complex (Rostislavleva *et al*, 2015), conformational dynamics of the ABC transporter P-glycoprotein (Marcoux *et al*, 2013), the organization of the nuclear pore complex scaffold (Bui *et al*, 2013), nucleotide binding to the p97 ATPase (Schuller *et al*, 2016), the interfaces of the H2A-H2B-Nap1 histone-chaperone complex (D'Arcy *et al*, 2013), the structure of the 55S mitochondrial ribosome (Greber *et al*, 2015), the norovirus assembly pathway (Utrecht *et al*, 2010a), the exosome complex topology (Hernández *et al*, 2006; Synowsky *et al*, 2006; Shi *et al*, 2015), and the architecture of the dynein cofactor dynactin (Urnavicius *et al*, 2015). Structural snapshots adapted from the referenced publications with permission.

of this MS-based “biochemical laboratory”. To this end, we will first introduce a few basic principles of biomolecular MS. Next, we will discuss the fundamental utilization of peptide-centric MS approaches as well as protein-centric MS under either denaturing or native-like conditions, highlighting selected applications to examine protein structures, functions, and interactions. Finally, we will exemplify how peptide- and protein-centric MS strategies have been

employed to answer central biological questions, often in hyphenation with more conventional molecular and structural biology methods, notably X-ray crystallography, NMR spectroscopy, and (cryo-)electron microscopy (EM). In particular, recent contributions of MS-based biochemical and structural analysis of the eukaryotic transcription machinery and prokaryotic CRISPR-Cas complexes will be highlighted.

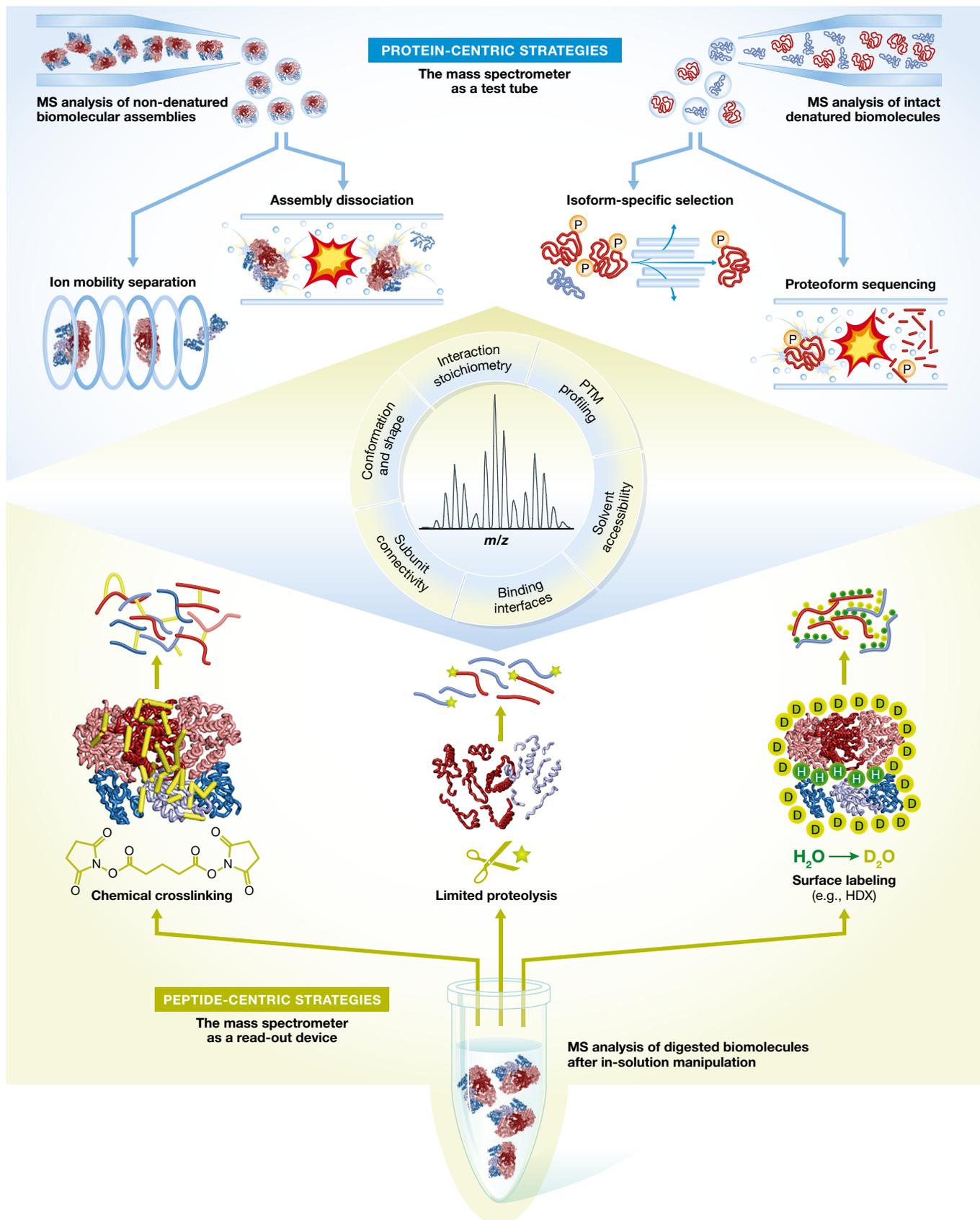


Figure 2.

**Figure 2. The broad scope of complementary biomolecular MS methods.**

In protein-centric approaches (top), non-digested biomolecules or biomolecular assemblies are analyzed. Typically, the sample is injected into the mass spectrometer via electrospray ionization using a conductive borosilicate capillary, as symbolized in the top panel. After injection, the analyte structure and composition can be further evaluated by several gas-phase manipulation strategies, some of which are schematically depicted. Ion mobility separation prior to mass measurement renders information about the analyte's conformation and shape. Dissociation of a non-denatured biomolecular assembly by collisional activation (symbolized as an explosion cartoon) allows inferring its stoichiometry and topological aspects. Mass selection can be performed for both denatured and non-denatured analytes, enabling the focused analysis of a specific species such as a protein isoform that carries a certain mutation or a defined number of PTMs. After mass selection, proteoform sequencing can be performed by fragmenting the protein isoforms in the gas phase (symbolized as an explosion cartoon), which allows to determine their amino acid sequence and thus locate mutated and/or modified residues. In peptide-centric approaches (bottom), proteins or protein complexes are manipulated in solution by traditional biochemical methods, such as chemical cross-linking, limited proteolysis, or surface labeling. Subsequently, the proteins are digested into peptides and liquid chromatography/tandem-MS is used as a highly accurate method to detect and identify the resulting peptides. A combination of the depicted peptide- and protein-centric approaches, as shown in the center, gives insights into a multitude of biochemical and structural properties of the studied protein assemblies.

**How to weigh a protein—basic principles of biomolecular MS**

MS can determine compound masses with an accuracy and precision that is unprecedented by any other analytical technique. Over the last 30 years, numerous strategies and applications of biomolecular MS have emerged and, concomitantly, many different types of mass spectrometers have been designed. Therefore, a generic instrumental platform for biomolecular MS does not exist. Most modern mass spectrometers, however, comprise a number of basic components (schematically depicted in Fig 3A) that are essential for the majority of biomolecular MS workflows. Initially, the analytes are separated from their carrier medium, usually an aqueous or organic phase, by transforming them into gaseous ions in the ion source. After ionization, the analytes, now carrying a charge, are transmitted into the high vacuum regions of the mass spectrometer, typically including a low-resolution mass analyzer and a collision cell (Fig 3A). On their way to the detector, the analytes pass through several electric and/or magnetic fields, which allow them to be mass-selected, activated, and separated from the remaining neutrals. At the final stage, the analyte mass-to-charge ratios are accurately and precisely measured by monitoring their motion through a high-resolution mass analyzer. In short, the most essential steps of a MS experiment are analyte ionization, mass determination, and selective manipulation. In this section, we provide a concise overview of these steps, focusing on principles and techniques that are commonly used in molecular and structural biology-related MS studies.

**Ionization principles**

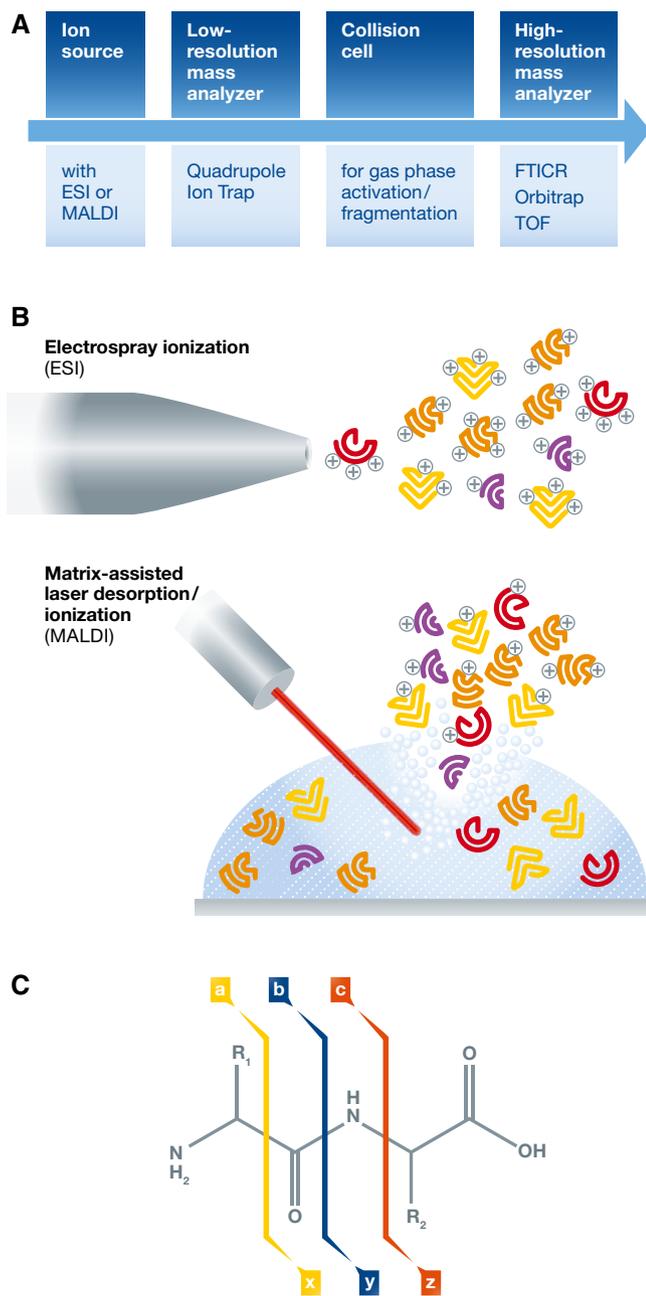
The ionization process must preserve the integrity of the analyte, which was long thought to be impossible for larger biomolecules because of their relatively labile peptide or sugar-phosphate backbone. The ionization methods MALDI and ESI both overcame this obstacle but follow fundamentally different mechanisms (Fig 3B). In MALDI, biomolecules are immersed in a crystalline matrix of dye-like molecules on a metal plate, wherefrom a laser induces the ionization and desorption of the analyte. In ESI, the analyte is dissolved in a volatile aqueous phase and inserted into a small conducting capillary. The analyte ionizes through the ultimate desolvation of minuscule charged droplets, which are created by applying an electrical current. Importantly, biomolecules become multiply charged during the ESI process, whereas mainly singly charged ions are generated during the MALDI process. MALDI has made substantial contributions to the emergence of biomolecular

MS and still has its unique application areas, for example, MALDI spatial imaging of biomolecules in tissue and cells (Amstalden van Hove *et al*, 2010; Schwamborn & Caprioli, 2010). However, ESI has become the method of choice for most biomolecular MS applications. This preference likely stems from the use of an aqueous phase for ESI sample preparation, which can be more easily coupled to chromatographic separation techniques. Moreover, advanced ESI technologies require minimal sample consumption (at nl/min rates), surpassing MALDI in terms of sensitivity and efficiency. ESI-based MS approaches are thus the focal point of this review; however, most of the herein discussed principles are also applicable to MALDI-MS.

**Mass analyzers**

One major criterion to distinguish mass analyzers is their mass resolving power. There are various types of quadrupole and ion-trap low-resolution mass analyzers that are mostly used for either targeted MS strategies (which will not be discussed herein, see Picotti & Aebersold, 2012), or for specific ion manipulation in tandem-MS experiments (ion activation and/or mass selection, see next subsection). However, accurate mass analysis of larger proteins and peptides is predominantly achieved using three different types of high-resolution mass analyzers: time-of-flight (TOF) tubes, Fourier transform ion cyclotron resonance (FTICR) traps, and Orbitraps (Marshall & Hendrickson, 2008). Although the term “high-resolution” mass analyzer is defined somewhat arbitrarily, the TOF, Orbitrap, and FTICR analyzers share one crucial feature—peptides of a typical protein digest, which usually comprise 5–50 amino acids and carry 2–7 charges, can be isotopically resolved, that is, mass separated into the different peptide isotopes. The existence of peptide isotopes is attributable to the natural occurrence of stable elemental isotopes, primarily  $^{13}\text{C}$  and  $^{15}\text{N}$ . Two adjacent peptide isotopes differ by 1 Da or, translated to the  $m/z$  scale, by 1 Th for a singly charged peptide, by 0.5 Th for a doubly charged peptide, etc. The regular isotope spacing, thus, readily allows determination of the peptide charge state, required for the calculation of its molecular weight.

The three high-resolution mass analyzers employ quite different principles of mass measurement. In TOF MS (Ens & Standing, 2005; Marshall & Hendrickson, 2008), ions travel a fixed distance within the TOF tube. The “flight time” of the ions can be used as a direct measure to determine their  $m/z$ . In FTICR MS (Marshall & Hendrickson, 2008), ions are trapped in a magnetic field and excited by an electric field oscillating at radio frequency. After removal of the excitation field, the ions rotate at an  $m/z$ -dependent frequency.



**Figure 3. Basic principles of biomolecular MS.**

(A) Essential components of a mass spectrometer. The ion source facilitates the transfer of the analyte into the gas phase. The low-resolution mass analyzer enables mass selection of specific analyte ion species and, in case of ion traps, may also be utilized for ion activation/fragmentation. Otherwise, gas-phase fragmentation, for example, using CID, HCD, or ETD, takes place in the collision cell. Finally, a high-resolution FTICR, TOF, or Orbitrap mass analyzer facilitates precise and accurate mass measurements. (B) Schematic representation of ESI and MALDI, the most commonly used ionization techniques in biomolecular MS. ESI produces multiply charged analyte ions (shown in yellow, orange, red, and purple) directly from a sample solution. In MALDI, a laser is used to ablate a mixture of matrix (shown in blue) and analyte molecules from a metal plate into the mass spectrometer, yielding predominantly singly charged ions. (C) Nomenclature of peptide fragment ions according to Roepstorff and Fohlman (1984) and Biemann (1990).

The rotational motion is detected by an electrode pair (forming the boundaries of the ICR cell) as an image current. The time-dependent image current can be transformed from the time domain to the  $m/z$  domain using Fourier transformation. Finally, Orbitrap MS applies principles that are related to FTICR, but Orbitraps do not require magnetic fields, which greatly simplifies instrument handling and maintenance (Zubarev & Makarov, 2013). Here, ions are trapped in an electric field generated between an outer barrel-like electrode and an inner spindle-like electrode. The ions oscillate along the inner electrode, again generating an image current. The oscillations are recorded and their  $m/z$ -dependent frequency is used to retrieve the  $m/z$  through Fourier transformation.

#### Gas-phase fragmentation techniques

Often, the  $m/z$  is not only determined for the intact analyte ion (MS1 level) but also for its fragment ions (MS2 level), which are generated in so-called tandem-MS experiments. Here, an intact analyte ion species of a specific mass is selected and isolated in the low-resolution mass analyzer and subsequently activated and fragmented. Gas-phase fragmentation (or activation) provides deeper insights into the analyte structure, for example, the amino acid sequence of a peptide. The analyte activation takes place either within the mass analyzer (typical for ion traps and FTICR cells) or in a collision cell situated between the low-resolution mass analyzer and the high-resolution mass analyzer (Fig 3A). Activation of peptides and denatured proteins typically leads to specific peptide backbone cleavage. Depending on which chemical bond within the peptide backbone is cleaved, this process generates a-/x-, b-/y- or c-/z-fragment ion series (Fig 3C). By definition, a-c describe N-terminal fragment ions and x-z describe C-terminal fragment ions (Roepstorff & Fohlman, 1984; Biemann, 1990; Steen & Mann, 2004). Series of these fragment ions detected at the MS2 level readily reveal the amino acid sequence of the analyzed peptide ion. If not peptides or denatured proteins, but native protein assemblies are subjected to gas-phase activation, efficient backbone fragmentation is generally not observed. Instead, activation may lead to (partial) unfolding and/or dissociation of non-covalently associated interaction partners. The release of these interaction partners and their mass measurement may provide insights into the quaternary structure of biomolecular complexes. Non-proteinogenic biomolecules, such as carbohydrates and nucleic acids, undergo their own characteristic fragmentation reactions, for which specific rules and nomenclatures have been established as well (Domon & Costello, 1988; Flora & Muddiman, 1998).

A diverse array of activation/fragmentation techniques is nowadays available. We confine this subsection to collision-induced dissociation (CID), higher-energy collisional dissociation (HCD, sometimes referred to as beam-type CID), and electron transfer dissociation (ETD), which represent some of the most popular ion activation strategies in biomolecular MS. For more information on alternative approaches, please refer to Bogdanov and Smith (2005); Brodbelt (2014); Zhou and Wysocki (2014); Zubarev (2004). Both CID and HCD depend on successive collisions between the analyte ions and inert gas molecules, for example, nitrogen, argon, or xenon, as they are accelerated in the collision cell. For peptide ions, CID- as well as HCD-based activation mainly gives rise to b- and y-type fragment ions in the MS2 spectra. CID, in particular, forces breakage of the analyte's most labile chemical bond. Applying CID

on a peptide bearing a labile post-translational modification (PTM) such as a phosphorylation or glycosylation therefore predominantly causes loss of the PTM rather than peptide backbone cleavage. This phenomenon is less prominent when HCD is used because it applies higher activation energies than CID. In ETD, analyte fragmentation is achieved through an electron transfer reaction from a radical anion toward the positively charged peptide (McLuckey & Stephenson, 1998; Syka *et al*, 2004). Due to this prompt reaction mechanism, ETD-based fragmentation is specifically targeted to the peptide backbone, generating mostly c- and z-type fragment ions, whereas PTMs remain bound to the peptide. A similar fragmentation pattern is produced by electron capture dissociation (ECD), which is specifically used in FTICR MS. As the ETD/ECD fragmentation pattern is complementary to CID/HCD, hybrid fragmentation approaches, such as ETcID and EThcD, can substantially improve the peptide sequence coverage (Swaney *et al*, 2007; Frese *et al*, 2012).

## Molecular biology meets mass spectrometry I—peptide-centric MS methods

The vast capabilities of peptide-centric bottom-up MS in identifying proteins and localizing amino acid modifications make it an ideal readout not only for standard proteomics workflows but also for approaches probing protein structures, conformations, and interactions by chemical or enzymatic in-solution modification. The most prominent examples of such approaches are chemical cross-linking, covalent and/or non-covalent surface labeling, and limited enzymatic proteolysis (Fig 2). All of these methods were introduced several decades ago, but their combination with bottom-up MS has substantially increased their impact on the field of structural biology. In the following subsections, we describe this development in more detail.

### Chemical cross-linking

Chemical cross-linking combined with bottom-up analysis (cross-linking-MS) can be used to study protein conformations as well as protein–protein and protein–nucleic acid interactions. This review will largely focus on protein–protein cross-linking, and a thorough overview of protein–nucleic acid cross-linking-MS applications has been published elsewhere (Schmidt *et al*, 2012).

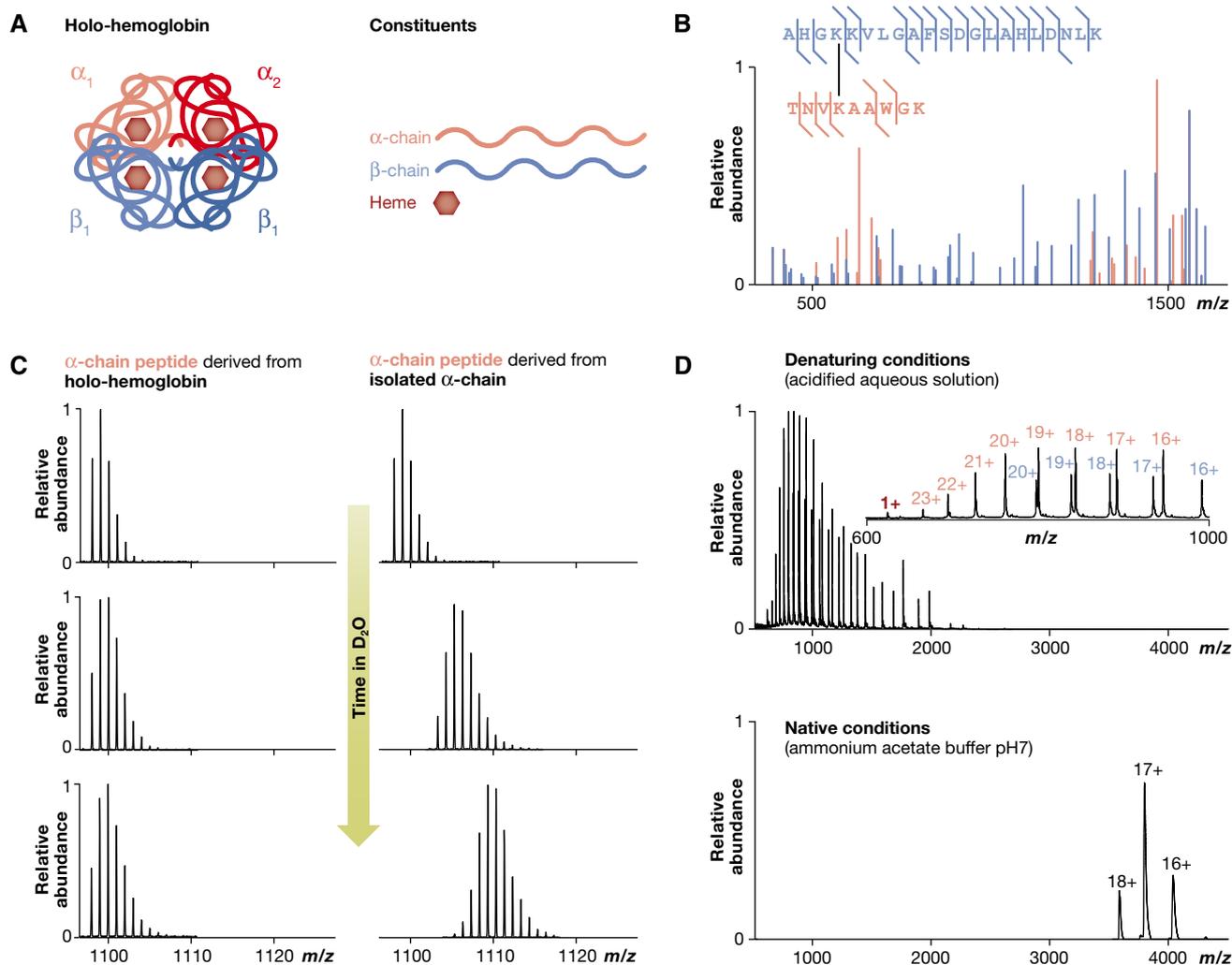
The fact that chemical cross-linking can capture intermolecular interactions of proteins in solution is known for at least 70 years (Fraenkel-Conrat & Olcott, 1946). The utilization of cross-linking for the structural probing of biomolecular systems was initially facilitated by the emergence of gel electrophoresis in the 1970s, as exemplified by topological studies on ribosomal protein complexes (Clegg & Hayes, 1974; Sun *et al*, 1974). Concomitantly, a wide array of cross-linking reagents has been explored and current cross-linking studies still largely rely on the same chemical principles (Sinz, 2003). In general, cross-linking reagents comprise a spacer arm of varying length connecting two functional groups, which are reactive toward specific amino acid residues. Presently, the most popular class of cross-linking reagents contains two N-hydroxysuccinimide ester functionalities that specifically target primary amino groups, that is, protein N-termini and lysine side chains. Other commercially available cross-linking reagents are reactive toward thiol groups

(i.e., Cys side chains) and carboxyl groups (i.e., Asp and Glu side chains) or contain non-specifically reactive photoactivatable moieties (Sinz, 2003; Petrotchenko & Borchers, 2010).

Figure 4 illustrates the experimental output produced by cross-linking-MS and other biomolecular MS experiments, to which we will recur in the following subsections. The respective data are exemplified using the simple model system human hemoglobin: a tetrameric protein complex consisting of two  $\alpha$ -chains, two  $\beta$ -chains, and four non-covalently bound heme groups (Fig 4A). Figure 4B shows a MS2 spectrum, acquired after mass selection and gas-phase fragmentation of two cross-linked hemoglobin peptides. The MS2 signals can be clearly annotated to fragment ions representing gas-phase fragmentation events at specific peptide backbone positions. Thus, MS allows sequencing of both peptides and localization of the cross-linker-modified residues. The conception of MS-based cross-link detection (Rappsilber *et al*, 2000; Young *et al*, 2000) has boosted the amount and quality of information that can be gained from a cross-linking experiment. Notably, two residues will only be cross-linked if their mutual distance can be bridged by the applied cross-linking reagent. Cross-links therefore impose distance constraints on the studied system, revealing binding interfaces (by locating inter-protein cross-links) and details about protein conformations (by locating intra-protein cross-links). Moreover, the abundance of cross-links under different experimental conditions can be relatively quantified, enabling comparisons between various biological states in respect of existing protein conformations and interactions (Fischer *et al*, 2013; Schmidt *et al*, 2013; Walzthoeni *et al*, 2015). Cross-linking data can be directly employed to guide computational protein homology modeling and protein–protein docking (Herzog *et al*, 2012; Kalisman *et al*, 2012; Lössl *et al*, 2014; Zeng-Elmore *et al*, 2014). Alternatively, they may be used to complement information from other structural biology approaches, most notably cryo-EM, thereby revealing the architecture of multi-protein complexes such as the nuclear pore complex scaffold (Bui *et al*, 2013), the 55S mitochondrial ribosome (Greber *et al*, 2015) or the INO80 chromatin remodeler complex (Tosi *et al*, 2013) (see also Fig 1). Of note, cross-linking data of protein homo-oligomers present a special case since links within and between the subunits cannot be readily distinguished. In most cases, the thorough annotation of homo-oligomeric cross-links critically depends on the availability of high-resolution three-dimensional protein structures (Kosinski *et al*, 2015).

Cross-linking-MS allows the unbiased structural probing of systems with—in theory—unlimited size and complexity, including cellular protein networks. To live up to this promise, however, cross-linking-MS had (and has) to master several obstacles concerning sample preparation and analysis, cross-link identification, and data interpretation. Here, we will focus on two major advances in these areas that have made cross-linking-MS a more powerful and versatile tool for molecular and structural biology research.

First, chemical cross-linking reactions generally proceed with low efficiency, thus generating a low amount of cross-linked protein relative to unreacted protein (Leitner *et al*, 2010). Beneficially, this largely prevents the detection of random protein contacts; however, it may also cause loss of information about specific interactions that are either short-lived or involve less abundant proteins. This especially becomes a limiting factor when cross-links are to be identified in complex mixtures, for example, whole-cell lysates, wherein



**Figure 4. Complementary biomolecular MS data exemplified using the human hemoglobin model system.**

(A) Schematic of native holo-hemoglobin and its constituents: two  $\alpha$ -chains, two  $\beta$ -chains, and 4 heme groups. (B) Cross-linking-MS. The displayed MS2 spectrum represents two linked peptides derived from cross-linked holo-hemoglobin. The ion signals, shown as blue and red sticks, correspond to specific fragments of the cross-linked peptides. These fragment ions enable the sequencing of both peptides and the localization of their linkage site, as indicated by the fragment ion map in the inset. Since two residues are only cross-linked if they are in close spatial proximity, the identified cross-link gives insights into the in-solution structure of holo-hemoglobin. (C) HDX-MS. Shown are the isotope distributions of the same peptide derived from either holo-hemoglobin (left panel) or the free hemoglobin  $\alpha$ -chain (right panel) after they were separately incubated in  $D_2O$ -containing buffer. The incubations were quenched at three different time points. The isotope distribution remains at the same  $m/z$  position, when the peptide is derived from holo-hemoglobin, whereas it gradually moves to higher  $m/z$ , when the peptide is derived from the free  $\alpha$ -chain. This shows that only in the free  $\alpha$ -chain, the peptide is able to take up the heavy deuterium isotope. Consequently, the peptide is solvent accessible in the free  $\alpha$ -chain, whereas it is solvent protected in holo-hemoglobin. (D) Protein-centric MS. Mass spectra of non-digested hemoglobin were acquired under denaturing (upper mass spectrum) and native (lower mass spectrum) conditions. Under denaturing conditions, all signals are concentrated in the low  $m/z$  region since the ions of the unfolded proteins are highly charged. The inset illustrates that the heme cofactor is present as a singly charged ion (brown), whereas the hemoglobin  $\alpha$ -chain (red) and  $\beta$ -chain (blue) are detected in several charge states. From these charge state envelopes, their accurate molecular masses can be derived ( $\alpha$ -chain =  $15,155 \pm 1$  Da,  $\beta$ -chain =  $15,895 \pm 1$  Da, heme =  $616.5$  Da). In the native mass spectrum, the signals are shifted to higher  $m/z$ , indicating that hemoglobin is detected in its folded state (see main text). The molecular weight can be calculated as  $64,588 \pm 41$  Da. Together with the constituent masses derived from the denaturing MS experiment, this result unambiguously evidences that holo-hemoglobin is an  $\alpha_2\beta_2$  heterotetramer with four bound heme ligands.

abundant proteins can “mask” low abundant cross-linked interactors. The key to a more sensitive detection of these less prevalent protein interactions is the enrichment of cross-links at either the protein or peptide level. Cross-link enrichment was first achieved using cross-linking reagents that harbor a biotin tag to facilitate affinity purification (Alley *et al.*, 2000; Trester-Zedlitz *et al.*, 2003). More recently, biotin-labeled cleavable cross-linkers, so-called protein interaction reporters, have been introduced (the general utility of cleavable cross-linkers is discussed in the next paragraph).

These cross-linkers were applied to enrich and identify cross-linked peptides from intact bacterial and human cells (Zheng *et al.*, 2011; Chavez *et al.*, 2013; Weisbrod *et al.*, 2013; Navare *et al.*, 2015). Similarly complex samples were probed by using a non-cleavable cross-linker with a removable biotin label (Tan *et al.*, 2016) as well as cleavable cross-linkers that can be biotinylated after the cross-linking reaction (Kaake *et al.*, 2014) or enriched by two-dimensional strong cation exchange chromatography (Buncherd *et al.*, 2014). Complementary to these cross-linker-based enrichment strategies,

several cross-linking-MS studies employed affinity-tagged proteins, enabling structural studies centered on a few proteins of interest. Based on such targeted approaches, detailed *in vivo* protein interaction networks of the 26S proteasome and the protein phosphatase 2A could be derived (Guerrero *et al*, 2006; Herzog *et al*, 2012). Recently, the structural organization of several endogenous protein complexes was examined by cross-linking-MS after affinity purification from transgenic GFP-tagged yeast strains and mice (Shi *et al*, 2015). Owing to the vast repository of these transgenic strains and organisms, this approach potentially opens an avenue to investigate virtually any protein system by *in vivo* cross-linking combined with protein-based affinity purification.

Second, cross-linking-MS had to devise efficient search engines for the identification of cross-linked peptides against very large peptide sequence databases. This presents a major challenge because all possible pairwise combinations of peptides need to be considered during the cross-link search. Searching for cross-linked peptides against a database generated from tens of proteins therefore is computationally as demanding as identifying linear peptides against the full human proteome (Liu *et al*, 2014). The identification of cross-links from more complex samples was initially facilitated by the search engine xQuest (Rinner *et al*, 2008), relying on the use of isotope-coded cross-linkers that “label” the cross-linked peptides in the mass spectra in order to decrease the search space. This, together with dedicated bioinformatics approaches, enabled cross-link identification against the full *E. coli* proteome (Rinner *et al*, 2008), which was later also achieved with the pLink search engine (Yang *et al*, 2012). Nowadays, xQuest and pLink have become the most widely used cross-link search engines to characterize the architecture of large protein complexes (Bui *et al*, 2013; Tosi *et al*, 2013; Cevher *et al*, 2014; Erzberger *et al*, 2014; Greber *et al*, 2014, 2015; Han *et al*, 2014c; Knutson *et al*, 2014; Shi *et al*, 2014).

A more fundamental approach to make cross-link identification more efficient focuses on cross-linking reagents, which are cleaved during the MS experiments (Soderblom & Goshe, 2006; Müller *et al*, 2010). Utilization of such MS-cleavable cross-linkers enables individual MS analysis of linear peptides, theoretically removing all limitations regarding the sample complexity. Consequently, MS-cleavable cross-linkers, which additionally contained an affinity tag, were applied in many of the above-mentioned cross-linking studies on intact cells (Zheng *et al*, 2011; Chavez *et al*, 2013; Weisbrod *et al*, 2013; Kaake *et al*, 2014; Navare *et al*, 2015). These studies employed sophisticated multi-dimensional chromatography and MS setups in combination with software tools tailor-made for the respective cross-linker. To simplify such advanced *in vivo* cross-linking experiments, we recently introduced the search engine XlinkX, the algorithm of which is compatible with any MS-cleavable cross-linker and with standard Orbitrap MS-based data acquisition workflows. In a proof of concept study, XlinkX was able to identify more than 2000 unique cross-links against a full human proteome database providing new insights into the interaction of the 80S ribosome with several associated proteins as they occur in the cellular environment (Liu *et al*, 2015).

### Surface labeling

Surface labeling is based on the principle that chemical probes preferably modify a biomolecule at its solvent-exposed parts, while amino acids that are buried, either in the folded protein core or by

an interacting protein, will not be affected. Any modification causes a defined mass shift, which can be detected at the peptide level by bottom-up MS analysis. Evidently, the information obtained from surface labeling is somewhat similar to the insights gained from cross-linking experiments. Cross-linking, however, readily identifies the interacting proteins, whereas surface labeling merely highlights protected regions. Translating surface labeling data into valuable structural information often requires probing of different protein states, for example, unfolded/folded, monomeric/oligomeric, or unbound/ligand-bound, in comparative or time-course labeling experiments. Such an analysis reveals which protein areas become solvent-exposed or buried and which regions remain unaffected, giving insights into biomolecular conformations and binding interfaces. Surface labeling can be performed in two different ways, that is, covalent labeling by amino acid side chain modifications and non-covalent labeling by hydrogen–deuterium exchange at the peptide backbone, better known as HDX-MS.

Covalent labeling has developed much in parallel with chemical cross-linking because the employed chemical principles partially overlap, as exemplified by the use of lysine-reactive and photoactivatable reagents (Klapper & Klotz, 1972; Bayley & Knowles, 1977). However, basically any covalent labeling strategy is compatible with bottom-up MS. Therefore, covalent labeling-MS can be tailored to target a wide array of amino acids (Mendoza & Vachet, 2009). Alternatively, it may be conducted in a non-specific fashion, for example, by photo-affinity labeling (Robinette *et al*, 2006) or oxidation with hydroxyl radicals (Maleknia & Downard, 2014). Hydroxyl radical probes are, for example, employed in a labeling approach named “fast photochemical oxidation of proteins” (FPOP) (Aye *et al*, 2005; Hambly & Gross, 2005). FPOP has been successfully used to monitor protein folding (Chen *et al*, 2010a), to map binding epitopes (Yan *et al*, 2014), and even to probe solvent-accessible protein regions *in vivo* while preserving the cellular integrity (Espino *et al*, 2015). Beneficially, FPOP seems to work equally efficient on proteins located in the cell membrane, cytoplasm, and nucleus.

The non-covalent labeling approach HDX is the best-known and most widely used strategy for biomolecular surface mapping. Hydrogen exchange as a means to probe biomolecular structures was already introduced in the late 1960s (Englander *et al*, 1972). It is based on the observation that hydrogens bound to N, O, or S exchange against solvent hydrogens, whereby peptide backbone hydrogens exchange with specific measurable rates while amino acid side chain hydrogens exchange too fast to be monitored (Englander *et al*, 1972). Soon after the emergence of biomolecular MS, HDX and MS were combined in pioneering studies, showing that this integrated approach gives insights into the solution and gas-phase conformation of proteins (Katta & Chait, 1993; Suckau *et al*, 1993). Most commonly, an HDX-MS analysis consists of three essential steps. First, a biomolecular assembly is transferred from H<sub>2</sub>O- to D<sub>2</sub>O-based buffers so that the solvent-exposed hydrogens will exchange to deuterium (see also Fig 2). Second, this exchange reaction is quenched at different time points by acidification and cooling of the solvent. Third, the reaction mix is subjected to bottom-up MS analysis to determine the deuterium uptake at the peptide level, allowing to derive HDX reaction kinetics that reveal solvent-accessible regions of the analyte. In Fig 4C, this is again illustrated for human hemoglobin. The mass spectra show peptides that have the same amino acid sequence but were derived from

either holo-hemoglobin (left panel) or the free hemoglobin  $\alpha$ -chain (right panel). Both samples were incubated in D<sub>2</sub>O-based buffer and analyzed at three different time points. The isotope envelope of the free  $\alpha$ -chain peptide moves continuously to higher  $m/z$  positions (right panel), evidencing that more and more hydrogens are replaced by heavy deuterium isotopes. In contrast, the holo-hemoglobin peptide remains at the same  $m/z$  position over time (left panel), showing that no deuterium is taken up. This suggests that the analyzed peptide is solvent accessible in the free  $\alpha$ -chain, but buried in holo-hemoglobin.

A main shortcoming of the standard HDX-MS workflow is that the bottom-up analysis is performed in H<sub>2</sub>O-based solvents, so the deuterium uptake may be partially mitigated by deuterium-to-hydrogen back exchange. In view of this obstacle, it is clear that increasing speed and efficiency of the bottom-up analysis has been a key to make HDX-MS a broadly applicable strategy (Konermann *et al*, 2011; Walters *et al*, 2012). This has been achieved by automating large parts of the HDX-MS analysis workflow. In a state-of-the-art HDX-MS experiment, the quenched reaction mix is subjected to on-column digestion with an immobilized acid-stable protease (mostly pepsin), peptides are separated using a miniaturized reversed-phase liquid chromatography system, and mass analysis is performed on a tandem-MS platform, all of which are often coupled online. Further advances—like online reduction of protein disulfide bonds prior to digestion (Trabjerg *et al*, 2015), ETD-based peptide fragmentation to monitor residue-specific deuterium uptake (Zehl *et al*, 2008), and sophisticated software tools for peptide mapping, data analysis, and structural interpretation (Slysz *et al*, 2009; Kan *et al*, 2011; Pascal *et al*, 2012; Rey *et al*, 2014)—have additionally streamlined the HDX-MS workflow.

Nowadays, HDX-MS is applicable to biomolecular systems that are hardly tractable by other techniques. For example, HDX-MS has become increasingly popular to study the structural biology of membrane proteins, shedding light on interactions of membrane receptors (Chung *et al*, 2011; Shukla *et al*, 2014), conformational dynamics of membrane proteins (Vahidi *et al*, 2016), or protein–membrane interactions (Rostislavleva *et al*, 2015). Moreover, HDX-MS is a particularly powerful approach to study the interactions of chaperones and their folding substrates. Such investigations are complicated by the fact that specific and non-specific chaperone–substrate interactions and intra-substrate interactions, all of which change dynamically during the folding process, have to be discerned. This complex interplay of events is exemplified by the interaction between the histone H2A–H2B heterodimer and its chaperone Nap1, recently studied by HDX-MS (D'Arcy *et al*, 2013). In this study, Nap1 was shown to confine the conformational flexibility of H2A–H2B, transforming their partially disordered histone fold domains into a more folded conformation. At the same time, the Nap1/H2A–H2B binding interface was mapped (shown in Fig 1) and, concomitant with protein complex formation, distinct cooperative unfolding events within both Nap1 and H2A–H2B could be revealed. Another recent HDX-MS investigation focused on chaperone-dependent differences in the folding pathway of the TIM-barrel protein DapA (Georgescauld *et al*, 2014). The authors demonstrated that DapA folds in a slow cooperative manner in the absence of the bacterial chaperone complex GroEL/GroES. When present, GroEL/GroES catalyzes the separate folding of individual DapA segments, thereby accelerating the folding process without specifically interacting with

DapA. Intriguingly, a DapA homolog protein from a bacterium that lacks GroEL/GroES folds in a fast, segmental way even without the chaperone complex. These findings led to the hypothesis that segmental folding may be a general pathway for TIM-barrel proteins, providing an evolutionary route toward GroEL/GroES independence.

#### Limited proteolysis

Limited proteolysis, introduced more than 60 years ago (Linderstrøm-Lang, 1950), is one of the most established biochemical approaches to study the higher order structure of biomolecules (Hubbard, 1998). It is based on the concept that proteolysis of a folded protein is not only dependent on the amino acid sequence but also on the tertiary structure, with surface-exposed and flexible regions being the most proteolytically susceptible sites (Fontana *et al*, 1986). Limited proteolysis is therefore a popular method to separate stably folded protein domains from flexible regions, for example, to generate X-ray crystallography-compatible protein constructs (Hubbard, 1998).

The use of limited proteolysis in conjunction with MS was pioneered by Brian Chait and co-workers (Cohen *et al*, 1995). Subsequently, this approach gained increasing popularity, enabling protein–protein binding site mapping (Gervasoni *et al*, 1998) or even monitoring of virus capsid conformational dynamics (Bothner *et al*, 1998). Over the past decade, the application of limited proteolysis-MS has been somewhat overshadowed by HDX-MS and cross-linking-MS, yet it was recently brought back into the spotlight by an elegant proteome-wide study on protein structural transitions using limited proteolysis and targeted MS (Feng *et al*, 2014). The authors performed metabolic shift experiments on yeast cells, changing the carbon source from glucose to ethanol, and globally monitored the condition-dependent abundance of limited proteolysis products. Changes in the limited proteolysis patterns may reveal altered protein conformations, differing protein PTM profiles or protein–ligand binding events. Crucially, limited proteolysis was performed under native conditions using proteinase K and, subsequently, the sample was further proteolyzed under denaturing conditions using trypsin. This step ensured the bottom-up MS compatibility of this highly complex sample while still allowing to distinguish between normal tryptic peptides (cleaved C-terminally of lysines and arginines) and limited proteolysis products (with at least one proteinase K cleavage site). Comparative analysis of the limited proteolysis products and their abundances revealed metabolic shift-induced conformational changes in several enzymes involved in carbon metabolism.

Complementary to limited proteolysis, structural aspects of proteins can also be probed by the proteome-wide monitoring of thermal protein denaturation (Savitski *et al*, 2014; Reinhard *et al*, 2015). This approach is based on the rationale that structural changes, in particular ligand association, will affect the protein stability and thus its denaturation temperature. The thermal denaturation and the limited proteolysis approach represent another two examples where the combination of peptide-centric MS with in-solution structural probing enables the system-wide examination of protein structures, conformations, and binding events. Being conducted on intact cells or cell lysates, these experiments are particularly suitable to foster our understanding of *in vivo* biomolecular interactions, thus gradually unraveling the cellular interactome.

## Molecular biology meets mass spectrometry II—protein-centric MS methods

### Experimental approaches

Although peptide-centric MS is nowadays the most popular variant of biomolecular MS, the initial seminal work introducing ESI and MALDI focused largely on intact proteins with masses up to 130 kDa (Karas & Hillenkamp, 1988; Tanaka *et al*, 1988; Fenn *et al*, 1989). Shortly afterward, it was demonstrated that, especially by using ESI, also non-covalent assemblies could be preserved in the gas phase and analyzed intact by mass spectrometry (Ganem *et al*, 1991; Katta & Chait, 1991; Light-Wahl *et al*, 1993; Schwartz *et al*, 1994; Fitzgerald *et al*, 1996). This finding gave rise to a new field that was later called “native mass spectrometry” (van den Heuvel & Heck, 2004). In native MS, named in analogy to native gel electrophoresis, purified proteins or biomolecular assemblies are injected into the mass spectrometer under non-denaturing conditions. While the aforementioned early experiments were mainly performed on protein–ligand complexes and protein homooligomers, careful tuning of the mass spectrometer and the ionization conditions lead to successful analysis of more and more non-covalent biomolecular complexes, including impressive examples such as the ribosome (Rostom *et al*, 2000), intact viruses (Fuerstenau *et al*, 2001; Uetrecht *et al*, 2008), the substrate-loaded GroEL/ES chaperone complex (van Duijn *et al*, 2005), and endogenously produced eukaryotic exosome complexes (Hernández *et al*, 2006; Synowsky *et al*, 2006) (see also Fig 1). Most of these studies used ESI, ionizing the proteins from an aqueous ammonium acetate buffer. Beneficially, these solutions are volatile, minimizing the formation of biomolecule–buffer molecule adducts, which would impair accurate mass measurements. Ammonium acetate has only moderate buffer capacity at physiological pH, but can keep biomolecules in a native-like functionally active state. The retention of native-like biomolecular tertiary structures in the gas phase could eventually be evidenced by probing the gas-phase structure of non-denatured proteins and protein complexes using ion mobility spectrometry-MS (IMS-MS) (Ruotolo *et al*, 2005; van Duijn *et al*, 2009).

IMS-MS has been a niche technology for several years, mainly applied to study the gas-phase conformations of metal clusters, peptides, and small proteins (Clemmer & Jarrold, 1997). This changed rapidly with the introduction of a commercially available IMS-MS instrument in 2007 (Pringle *et al*, 2007). In the last ten years, IMS-MS has become a popular technique to study the overall structure of biomolecular assemblies, monitor changing protein conformations, or reduce sample complexity by complementing the mass measurement with another gas-phase separation dimension. All of these aspects have been described in several reviews (Clemmer & Jarrold, 1997; Uetrecht *et al*, 2010b; Niu *et al*, 2013; Lanucara *et al*, 2014) and protocols (Ruotolo *et al*, 2008). Therefore, we will provide only a brief explanation of the IMS principle. In essence, IMS separates gaseous ions based on their mobility through a buffer gas. This may allow the detection of different conformers of the same protein (Clemmer *et al*, 1995) or the separation of different oligomeric states of amyloidogenic proteins (Woods *et al*, 2013), which may exhibit the same  $m/z$  values but different ion mobilities. The ion mobility depends on a number of aspects including ion charge and size. Most importantly, ion mobility determination allows calculating rotationally averaged collisional cross-sections of

the analyte ions, which renders information on their shape. Changes in ion mobility detected by IMS, thus, indicate structural changes. Figure 1 shows two examples where such information was used to monitor ligand-induced conformational changes of an ABC transporter (Marcoux *et al*, 2013) and sequential steps of norovirus capsid assembly (Utrecht *et al*, 2010a).

Currently, the majority of the native (IMS-)MS experiments is performed on soluble proteins and protein complexes, as exemplified by studies on megadalton-sized assemblies like cargo-loaded bacterial nano-containers (Rurup *et al*, 2014) and the dynein cofactor dynactin (Urnavicius *et al*, 2015), the structure of which is shown in Fig 1. However, native MS may also be used to investigate membrane proteins, which can be directly infused after solubilization either in detergent micelles (Barrera *et al*, 2008) or by detergent-free methods (Hopper *et al*, 2013). Besides highly accurate mass measurements of biomolecular assemblies, native MS experiments can provide valuable information on stoichiometries and oligomeric states, especially when combined with gas-phase dissociation in native tandem-MS experiments and/or with complementary MS experiments under denaturing conditions. This is illustrated in Fig 4D wherein mass spectra of denatured and native hemoglobin are depicted. In the upper mass spectrum, hemoglobin was measured by ESI-MS under acidic denaturing conditions, disrupting all non-covalent interactions. Accordingly, the heme cofactor has dissociated from the  $\alpha$ - and  $\beta$ -chain, as evidenced by a signal corresponding to free heme (Fig 4D, inset in upper mass spectrum). Moreover, the  $\alpha$ - and  $\beta$ -protein chains are unfolded, exposing all their chargeable amino acids. This causes the emergence of highly charged  $\alpha$ - and  $\beta$ -chain ions during the ESI process, which are detected at relatively low  $m/z$ . Since these protein ions are substantially larger and higher charged than peptide ions, their isotope pattern is not resolved. However, their molecular weights can still be accurately determined based on the  $m/z$  differences between the differently charged  $\alpha$ - and  $\beta$ -chain ions (Fig 4D, inset in upper mass spectrum). The masses of the three hemoglobin components— $\alpha$ -chain ( $15,155 \pm 1$  Da),  $\beta$ -chain ( $15,895 \pm 1$  Da), and heme (616.5 Da)—are thus readily revealed by the denaturing MS analysis. In contrast, the native mass spectrum (Fig 4D, lower mass spectrum) contains only signals for one species, represented by three charge states in a relatively high  $m/z$  region. The shift to higher  $m/z$  is mainly due to the fact that the analyte takes up fewer charges than under denaturing conditions. Charging of the analyte is reduced in native MS because some chargeable sites are buried within the folded protein core and the pH of the spraying solution is closer to neutral. Based on the three charge states seen in the native mass spectrum, the species mass can be calculated as  $64.59 \pm 0.04$  kDa, which unambiguously corresponds to a protein complex consisting of two  $\alpha$ -chains, two  $\beta$ -chains, and four heme groups. Two mass spectra, each acquired in < 2 min, can thus correctly identify both the composition and binding stoichiometry of a biomolecular complex.

While the hemoglobin example illustrates (albeit for a simple case) the power of protein-centric MS, it also highlights that one level of information is still missing—the amino acid sequence. Sequencing intact proteins directly in a so-called top-down approach (Kelleher *et al*, 1999) rather than inferring their sequence from bottom-up MS data offers several advantages. Most importantly, top-down sequencing may go beyond sheer protein identification,

potentially revealing sequence variations, the position of PTMs, and even the interdependence among different mutations and modifications, all of which is barely feasible when using solely peptide-centric MS methods. The success of this strategy, however, heavily depends on a highly efficient fragmentation of the peptide backbone, which is generally more difficult to achieve for intact proteins than for short peptides. It is therefore not surprising that the development of top-down sequencing closely follows advances in peptide fragmentation techniques. While top-down sequencing of native proteins and protein complexes is still in its infancy, first successful top-down experiments on denatured protein samples were performed in the McLafferty laboratory around the turn of the millennium (Kelleher *et al*, 1999). This breakthrough was achieved on FTICR mass spectrometers mainly due to the invention of ECD fragmentation (Horn *et al*, 2000; Sze *et al*, 2002). Over the years, more fragmentation techniques such as ETD and UV photodissociation were introduced, allowing top-down sequencing to be performed on several types of instruments, including Orbitrap mass spectrometers (Chi *et al*, 2007; Fornelli *et al*, 2012; Shaw *et al*, 2013). Next to the progress in the field of gas-phase fragmentation, the development of powerful spectrum analysis software and statistical tools (Liu *et al*, 2012; Fellers *et al*, 2015; Cai *et al*, 2016) has greatly increased the scope of top-down protein analysis. Finally, more efficient protein extraction, separation, and fractionation methods were critical prerequisites for an in-depth analysis of complex samples by top-down MS (Sharma *et al*, 2007; Chen *et al*, 2008; Tran & Doucette, 2009; Han *et al*, 2014b).

It becomes apparent that, by now, many different variants of protein-centric MS have evolved, ranging from non-denaturing approaches that may be combined with gas-phase dissociation and ion mobility separation to denaturing approaches that even facilitate intact protein sequencing (Fig 2). Combining these approaches is, in our view, the key to a thorough understanding of biomolecular systems. Therefore, the following subsections will showcase examples that illustrate the diverse, often complementary, analytical angles provided by the different protein-centric MS approaches.

#### *Comprehensive analysis of post-translational modifications*

As phosphorylation is one of the most prevalent PTMs, the functional and structural characterization of protein phosphorylation is a major theme in molecular biology (Hunter, 1995; Johnson, 2009). Detection and site localization of protein phosphorylation sites using MS works without radioactive labeling and specific antibodies, in contrast to more traditional biochemical methods. Protein phosphorylation is typically analyzed with bottom-up proteomics approaches (Riley & Coon, 2016); however, the central role of phosphorylation in modulating protein conformation, activity, localization, and complex formation/dissociation has driven the development of low-throughput protein-centric MS approaches that are better suited to monitor these aspects.

Owing to the ability of native MS to capture non-covalent interactions, it appears to be a straightforward choice to monitor the effect of protein phosphorylation on biomolecular interactions. However, differentially phosphorylated protein isoforms were, for a long time, nearly impossible to distinguish due to limitations in mass resolving power. Notably, phosphorylation causes a mass shift of no more than 80 Da; therefore, phospho-isoforms differ in mass by generally < 0.1%. Resolving these subtle differences became possible with the

development of mass spectrometers that combined high resolution with a high mass range. Important early contribution in this area came from orthogonal Q-TOF mass spectrometers, modified to allow the transmission and detection of high mass ions (Sobott *et al*, 2002; van den Heuvel *et al*, 2006). More recently, the now-commercialized Orbitrap EMR instrument has really made an impact in this field (Rose *et al*, 2012; Snijder *et al*, 2014). This novel mass spectrometer is able to mass-resolve phospho-isoforms of proteins and protein complexes of several 100 kDa, as exemplified in a recent investigation of the interplay between protein phosphorylation and protein–protein or protein–ligand interaction dynamics (van de Waterbeemd *et al*, 2014). In this study, the phosphorylation and cyclic nucleotide binding of dimeric 150 kDa cGMP-dependent protein kinase (PKG) were simultaneously monitored by high-resolution native MS, showing that binding of cAMP or cGMP causes different PKG phosphorylation kinetics. In a second example, it was demonstrated that the binding and phosphorylation of the mitotic regulator Bora by the cycle kinase Aurora A proceed independently. Interestingly, all three investigated proteins—Aurora A, Bora, and PKG—existed in different phosphorylation states. The relative abundance of all these phospho-isoforms could be accurately determined by native MS, whereas complementary peptide-centric MS experiments were done to localize the phosphorylated residues (van de Waterbeemd *et al*, 2014). Since the coexisting phospho-isoforms are indistinguishable at the peptide level, the phosphorylated residues could not be allocated to specific phosphorylation states. This level of information was later accessed by combining highly specific ion isolation and complementary gas-phase fragmentation techniques in a top-down protein sequencing approach (Brunner *et al*, 2015). Top-down sequencing allowed to decipher the phospho-proteoforms of Bora resulting from phosphorylation by either Aurora A or Polo-like kinase 1 (Plk1), showing that both kinases target different Bora residues and generate distinct phosphorylation successions. Compared to these binary kinase/Bora systems, the tripartite Aurora A/Bora/Plk1 interplay is analytically even more challenging, as it is characterized by numerous mutual phosphorylation events with various implications on protein structure and function. Simultaneous probing of these often temporarily occurring effects was recently achieved by using an MS-based structural biology strategy, integrating native MS, cross-linking-MS, IMS-MS, top-down sequencing, and bottom-up proteomics (Lössl *et al*, 2016). Strikingly, it could be demonstrated that Aurora A and activated Plk1 hyper-phosphorylate Bora according to a defined sequence of residue-specific phosphorylations, thereby priming a substantial structural change of Bora, which eventually allows stable Plk1/Bora complex formation. This multipronged MS analysis, thus, provided mechanistic insights into the sequence of events accompanying the Aurora A/Bora-mediated Plk1 activation, which is essential for recovery from DNA damage-induced cell cycle arrest.

Another example that the seemingly small difference of one phosphorylation can have a profound effect on proteins was recently reported for calmodulin after phosphorylation by casein kinase 2 (Pan *et al*, 2016). The kinase was shown to phosphorylate calmodulin between one and four times in a specific order. To see whether any of these phosphorylation events influenced the calmodulin structure, the phosphorylated protein was incubated with deuterated buffer for a set amount of time, denatured and injected into the mass spectrometer. Subsequent selection and ETD

fragmentation of the distinct phospho-isoforms allowed the quantification of the deuterium uptake in specific regions. This innovative top-down HDX-MS experiment successfully proved that only a specific pair of phosphorylation events influenced the calmodulin structure substantially.

The in-depth characterization of PTMs using protein-centric MS is not limited to phosphorylation. An interesting example exhibiting more complex PTM patterns is presented by histone proteins bearing multiple modifications such as methylation (+14 Da), acetylation (+42 Da), and phosphorylation (+80 Da). Cross talk between these modifications has been examined for histone H3 using state-of-the-art top-down MS analysis (Zheng *et al*, 2016). Another protein modification that recently attracted a lot of attention in the biomolecular MS community is glycosylation. Unlike the aforementioned PTMs, glycosylation needs to be studied not only in respect of its location and abundance, but also with regard to the saccharide composition of the often very complex glycan trees. For this reason, glycosylation analysis requires a different set of MS strategies. Bottom-up MS analysis can be conducted to examine glycosylation patterns, whereby complementary fragmentation methods, for example, ETD and HCD, allow to derive both the peptide sequence and the glycan tree composition. While bottom-up MS reveals the specific glycan linkages as well as site-specific glycosylation differences, these data can be effectively complemented by protein-centric MS analysis. Both native MS and denaturing MS have been successfully used to profile complex glycosylation patterns even in the presence of other PTMs, as exemplified by the comprehensive analyses of chicken ovalbumin and interferon- $\beta$ 1, which were shown to consist of 59 and 138 different protein isoforms, respectively (Yang *et al*, 2013; Bush *et al*, 2016). As such, protein-centric MS may become a screening method to compare patient-derived plasma glycoproteins or production batches of protein therapeutics. Most therapeutic proteins, such as monoclonal antibodies, are well characterized regarding their primary sequence and specific glycosylation sites. However, the glycan tree composition at these sites can still differ substantially, depending on the source material (e.g., CHO cells, yeast or human cell lines) and its growth conditions. A protocol to investigate these glycan trees, in particular on monoclonal antibodies, with native MS has been reported 2 years ago (Rosati *et al*, 2014). This approach relies on the stepwise application of glycan-specific glycosidases to sequentially truncate the glycan trees. As a result of this procedure, certain proteoforms show glycan-specific mass losses, which can be immediately read out with native MS, allowing the step-by-step reconstruction of the glycosylation profile. As an additional benefit of native MS, this strategy can also be used for glycoprotein complexes. Examples of non-covalently associated glycoproteins characterized by native (IMS-) MS include glycosylated antibody-antigen complexes, multimeric glycoproteins (Dyachenko *et al*, 2015), glycosylated antibody-drug conjugates (Rosati *et al*, 2013; Marcoux *et al*, 2015), and glycoprotein complexes involved in complement activation (Diebold *et al*, 2014; Wang *et al*, 2016).

#### Protein–ligand binding kinetics and stoichiometries

Binding of ligands, such as cofactors, nucleotides, lipids, or drug molecules, is to some extent similar to PTMs, as both result in a characteristic mass shift of proteins and protein complexes. Unlike PTMs, however, ligands are in general non-covalently associated, so

their binding needs to be investigated by native MS. Native MS protein–ligand interaction studies are not limited to the mere detection of the binding event, but can also provide information about the binding stoichiometry (McCammon *et al*, 2004; Schuller *et al*, 2016), affinity (Clark & Konermann, 2004; El-Hawiet *et al*, 2012a), and cooperativity (Dyachenko *et al*, 2013; Lin *et al*, 2014). Over the past decade, such approaches have gained importance as small molecule screening studies in the pharmaceutical industry (Hofstadler & Sannes-Lowery, 2006; Vivat Hannah *et al*, 2010; Maple *et al*, 2014).

In an elegant example of MS-based ligand binding studies, the interaction between gangliosides (sialic acid glycosphingolipid conjugates) and human norovirus proteins has been investigated using three sophisticated native MS strategies (Han *et al*, 2014a). Initially, the authors used native “catch-and-release” ESI-MS (El-Hawiet *et al*, 2012b) to screen a carbohydrate mixture, resembling the oligosaccharide moiety of 17 gangliosides, against the 865 kDa oligomeric norovirus P-particle, a mimic of the capsid’s protruding spike structure. The resulting native mass spectra contained ions representing the most prominently formed P-particle–carbohydrate complexes. This convoluted signal was mass-selected and subjected to CID to release the bound carbohydrates, allowing the detection of any dissociated carbohydrate in the low  $m/z$  region. Second, a direct ESI-MS assay (Wang *et al*, 2003) was applied to confirm the identified ligands and to quantify their binding affinity toward a smaller dimeric version of the P-particle. Third, to verify the relevance of the P-particle model system, the derived kinetic constants were cross-validated by a “proxy protein” ESI-MS method (El-Hawiet *et al*, 2012a). In this approach, a 10.5 MDa norovirus-like particle and a small (16 kDa) “proxy protein” with known carbohydrate binding affinity were co-incubated to compete for carbohydrate binding. Since different carbohydrate binding states are much easier to distinguish for small proteins, the carbohydrate occupation of the “proxy protein” was measured at different norovirus-like particle concentrations. This provided an indirect readout to determine the norovirus-like particle–carbohydrate binding affinities, which were in good agreement with the results of the direct ESI-MS approach.

As mentioned above, native (IMS-)MS has also become a valuable method to study membrane proteins. The emerging view is that some of these proteins may preferentially interact with specific lipids and, especially in the case of membrane transporters, adopt situation-specific conformations. Intriguingly, both of these aspects can be probed by native IMS-MS, as illustrated in the following examples. First, a recent IMS-MS study elucidated the lipid-binding selectivity of three membrane proteins, the mechanosensitive channel of large conductance, aquaporin Z, and the ammonia channel (Laganowsky *et al*, 2014). The extent to which these membrane proteins are stabilized by different lipids was measured using IMS-MS, which gives information on both the protein shape and the protein mass. The former provides direct evidence for partial unfolding in the gas phase as a readout for protein stability, whereas the latter readily identifies the specific protein–lipid complex corresponding to the respective unfolding state. Beneficially, this allows separate interrogation of successive lipid-binding events, demonstrating how different synthetic and natural lipids or multiple lipid-binding events modulate the membrane protein stability. In all three cases, the highest stability was rendered by a class of lipids that was shown to be functionally significant for the respective protein. In

the second example, factors involved in the conformational transitions of membrane transporters have been investigated for the membrane-embedded mammalian drug efflux pump P-glycoprotein, probing the influence of the specific binding of lipids, nucleotides, and drugs (Marcoux *et al*, 2013). All three classes of small molecules were shown to bind independently as well as concomitantly to the P-glycoprotein; however, only synergistic binding triggered a significant shift in the conformational equilibrium, resembling the structural transition expected for an efflux process.

### Monitoring cellular machineries—the role of MS in integrated structural and molecular biology studies

So far, we have described several examples of biomolecular systems that were successfully probed by peptide- and/or protein-centric MS strategies (see also Fig 1). A few biomolecular assemblies have become recurring subjects of integrated structural and molecular biology studies, often involving biomolecular MS approaches next to more traditional methods such as X-ray crystallography and EM. In the final section of this review, we will focus on two such systems: the eukaryotic transcription machinery and the bacterial CRISPR–Cas immune system. Our understanding of these systems has substantially increased owing to studies that combined biomolecular MS with other structural and molecular biology techniques. Thus, this section aims to illustrate the added value of such integrated approaches highlighting specifically the niche of biomolecular MS therein.

#### Eukaryotic transcription complexes

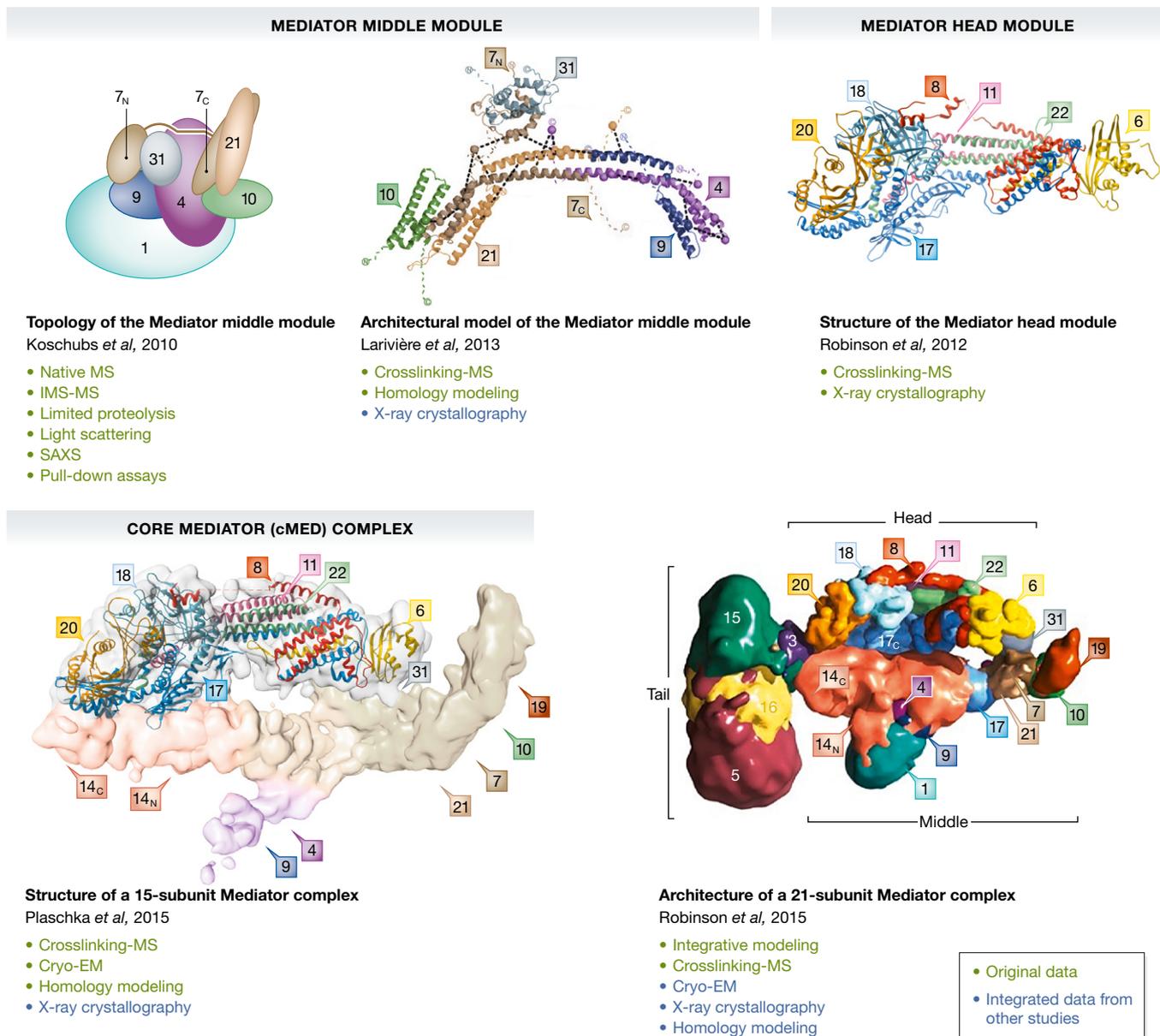
DNA-dependent RNA polymerases (Pol) are responsible for gene transcription in eukaryotic cells. In most eukaryotes, three of these multi-subunit enzymes are present, with Pol I synthesizing ribosomal RNA, Pol II producing messenger RNA, and Pol III making transfer RNA and small RNAs. Pol II, the most widely studied subtype, is a 514 kDa protein complex consisting of 12 different subunits (Bushnell & Kornberg, 2003). Several assembly states of Pol II have been characterized by X-ray crystallography, culminating in structural models of Pol II in complex with transcription factor (TF) IIB, TATA box binding protein, a DNA template, and an RNA synthesis product (Kostrewa *et al*, 2009; Liu *et al*, 2010; Sainsbury *et al*, 2012). However, Pol II engages in even more complex assemblies in the course of mRNA synthesis, recruiting other transcription factors, mRNA processing enzymes, and even supramolecular co-activators like the Mediator complex. Further complexity arises from the bound RNA transcript but also from heterogeneous PTMs, most notably on the Pol II C-terminal domain (Heidemann *et al*, 2013; Allen & Taatjes, 2015; Sainsbury *et al*, 2015). Similarly, Pol I and Pol III bind several transcription factors and general regulators (Vannini, 2013). Such massive, dynamically interacting ensembles are typically elusive to X-ray crystallography. Therefore, a number of these assemblies have recently been investigated by hybrid structural and molecular biology approaches, the key elements of which are cryo-EM and biomolecular MS. These strategies not only uncovered a wealth of structural information but also gave insights into the relationship between the architecture and biochemical function of RNA polymerase supercomplexes.

Pol II from yeast represents the first fully assembled cellular machinery probed by cross-linking-MS (Chen *et al*, 2010b). The architecture of Pol II was accurately reflected by cross-linking-MS, initiating its rise to an established structural biology method. Moreover, the Pol II–TFIIF binding interface could be mapped based on specific inter-protein cross-links. The Pol II–TFIIF complex model was further extended in a more comprehensive cross-linking-MS study, resulting in a structural model of the yeast core initiation complex that could be reconciled with previously obtained biochemical insights into the yeast pre-initiation complex (Mühlbacher *et al*, 2014). To obtain high-resolution structural models of this pre-initiation complex, cross-linking-MS was used in combination with cryo-EM in two independent investigations (Murakami *et al*, 2013; Plaschka *et al*, 2015). The more recent study from Plaschka *et al* (2015) even went beyond the pre-initiation complex and provided the structure of the pre-initiation complex bound to the co-activating Mediator core complex. The structure of this 1.2 MDa supercomplex was solved with subnanometer resolution using cryo-EM; however, this analysis did not reveal the subunit arrangement within the Mediator middle module. Instead, this part of the Mediator core complex could be topologically elucidated based on the cross-linking-MS distance constraints, demonstrating the benefits of integrating complementary structural biology approaches (Plaschka *et al*, 2015).

The Mediator complex is exemplary for biomolecular assemblies that are refractory to traditional structural biology methods, as it is conformationally highly flexible and compositionally diverse (Fig 5). The Mediator middle module was investigated early on by native MS, tandem-MS, and IMS-MS, which, in combination with light scattering, small-angle X-ray scattering (SAXS), and pull-down assays, revealed its overall shape and subunit topology (Koschubs *et al*, 2010). This model was later refined using cross-linking-MS and homology modeling (Larivière *et al*, 2013). Furthermore, a structure of the Mediator head module has been determined by X-ray crystallography and cross-linking-MS (Robinson *et al*, 2012). A full Mediator complex model was finally derived based on a comprehensive hybrid structural biology strategy combining X-ray crystallography, cryo-EM, computational modeling, and cross-linking-MS (Robinson *et al*, 2015). Very recently, this model was extended to a full Mediator–Pol II pre-initiation complex structure comprising 52 protein subunits (Robinson *et al*, 2016).

Moving on from transcription initiation to mRNA elongation and processing, native MS, cross-linking-MS, and cryo-EM proved once more to be a fruitful combination, uncovering the structural and biochemical basis for co-transcriptional mRNA capping (Martinez-Rucobo *et al*, 2015). The capping process, which modifies the 5' end of the newly synthesized mRNA, is performed in yeast by the Cet1 triphosphatase and the Ceg1 guanylyltransferase. Through native MS, a heterotetramer of these enzymes was shown to bind to Pol II. Next, the relevance of the formed capping and transcribing Pol II complex was proven by MS-based monitoring of the stepwise mRNA modification. This paved the road for cryo-EM and cross-linking-MS experiments that revealed the capping enzyme binding site at the Pol II mRNA exit tunnel.

In contrast to Pol II, eukaryotic Pol I and Pol III were only recently characterized by high-resolution structures (Engel *et al*, 2013; Fernández-Tornero *et al*, 2013; Hoffmann *et al*, 2015). However, some structural aspects of both complexes have



**Figure 5. Hybrid approaches gradually unravel the Mediator complex architecture.**

The Mediator complex is a challenging target for structural biology because it comprises several subcomplexes and highly flexible regions. First, a topological model of the Mediator middle module was derived with a hybrid approach employing native MS and IMS-MS. This could be extended to a more detailed model by means of cross-linking-MS and homology modeling. In parallel, a Mediator head module structure was derived by combining X-ray crystallography and cross-linking-MS results. Finally, hybrid structural biology approaches led to the characterization of a 15-subunit Mediator core complex and a 21-subunit Mediator complex. Structural images adapted with permission from the publications referenced in the figure.

previously been revealed by MS analysis. Native (tandem-)MS together with in-solution dissociation experiments confirmed that 10 of the 17 Pol III subunits form a stable core, whereas the remaining 7 subunits are more peripheral (Lorenzen et al, 2007). Three of these peripheral subunits form the C82/34/31 subcomplex, which could be accurately positioned on the Pol III core based on cross-linking-MS and biochemical data (Wu et al, 2012). The presence of Pol III subcomplexes was confirmed in another native MS study that also included IMS-MS to obtain topological information (Lane et al, 2011). This study additionally probed the assembly of Pol I, finding

that Pol I and Pol III exhibit similarities in their disassembly pathways. While these examples show how MS by itself can render structural insights, both native MS and cross-linking-MS were also integrated in hybrid structural biology approaches that revealed structural and functional features of Pol I and Pol III subunits and accessory factors. For instance, key aspects of the Pol III pre-initiation complex architecture were unveiled by cross-linking-MS and X-ray crystallography of the TFIIC complex (Male et al, 2015). Regarding the structural organization of Pol I, native MS proved that the Pol I subunits A49 and A34.5—the only ones for which no Pol II

homologs exist—form a stable heterodimer that associates with the other 12 Pol I subunits (Geiger *et al*, 2010). Subsequent crystallization of the A49/A34.5 heterodimer enabled structural comparisons to Pol II-associated transcription factors, uncovering key similarities between A49/A34.5 and TFIIF as well as TFIIE. Moreover, a cross-linking-MS-based Pol I model showed A49/A34.5 in positions similar to the TFIIF and TFIIE binding regions on Pol II (Jennebach *et al*, 2012), reinforcing the concept that some Pol I subunits act as stably associated transcription factors. Other factors for transcription initiation, however, are reversibly associated with Pol I. One of them, Rrn3, was studied by SAXS and native MS, which led to the conclusion that it forms dimers in solution but monomerizes during Pol I binding (Blattner *et al*, 2011). The Rrn3/Pol I interaction site could be mapped by cross-linking-MS, resulting in a model of the Pol I–Rrn3 initiation complex (Blattner *et al*, 2011).

#### Protein assemblies involved in the CRISPR–Cas immune system

Bacteria and archaea have evolved a variety of defense strategies to withstand viral infection, but most attention was in recent years drawn to the RNA-guided adaptive immune response through clustered interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins. The molecular mechanism behind the various CRISPR–Cas systems has been reviewed extensively (van der Oost *et al*, 2014; Wright *et al*, 2016). Briefly, CRISPR gene loci can accommodate short stretches of bacteriophage nucleic acid sequences (“spacer”) that are incorporated during viral infection. These spacers are transcribed into CRISPR RNA, which directs the Cas proteins toward matching viral nucleic acids, enabling their degradation.

CRISPR–Cas systems are divided into three main types. Type II systems, which contain CRISPR RNA and only one multifunctional Cas protein, have become highly popular tools for genome engineering (Mali *et al*, 2013; Wright *et al*, 2016). Type I and type III systems, however, comprise CRISPR RNA and multiple Cas proteins, forming ribonucleoprotein complexes of 350–450 kDa that represent a challenging target for structural characterization.

*Escherichia coli* type I Cascade was the first CRISPR–Cas system to be structurally characterized (Jore *et al*, 2011). Cascade is composed of five different Cas proteins and CRISPR RNA, the masses of which sum up to 184 kDa. When intact Cascade was analyzed by native MS, however, it exhibited a mass of approximately 405 kDa, showing that one or more components must be present in multiple copies. To decipher the Cascade binding stoichiometry, the authors first added a complementary single-stranded DNA probe and monitored the mass increase by native MS, demonstrating that only one CRISPR RNA is bound. Cascade was next subjected to gas-phase dissociation and tandem-MS analysis, monitoring the sequential loss of several Cas subunits. Since these results were not sufficient to infer the complete binding stoichiometry, additional in-solution dissociation experiments were performed, disrupting Cascade by adding low amounts of organic solvents. The resulting subcomplexes were analyzed by native (tandem-)MS analysis, which finally allowed to derive copy numbers for all Cas proteins in the intact Cascade assembly. By combining these results with two-dimensional EM data, a first topological model of Cascade could be proposed (Jore *et al*, 2011), which soon was confirmed by a more detailed cryo-EM reconstruction (Wiedenheft *et al*, 2011a). A similar approach combining native MS, EM, and SAXS was used

to probe the *P. aeruginosa* Csy complex, unveiling remarkable similarities to the Cascade quaternary structure despite the lack of obvious protein sequence homology (Wiedenheft *et al*, 2011b). This structural analogy was further demonstrated in a follow-up study that applied native IMS-MS and molecular modeling (van Duijn *et al*, 2012).

More recently, several type III CRISPR–Cas systems were subjected to hybrid structural investigations. First structural information was obtained on the *T. thermophilus* Cmr complex (Staals *et al*, 2013). This assembly was studied by deep sequencing to probe its CRISPR RNA content, native MS in combination with gas-phase and in-solution dissociation to determine its binding stoichiometry, and negative-stain EM to reconstruct a first 3D complex map, in which the subunits could be placed. Similar approaches combining deep sequencing, native MS, and EM resulted in structural models of Csm complexes isolated from bacteria (Staals *et al*, 2014) and archaea (Rouillon *et al*, 2013). Interestingly, all three studies showed that the respective ribonucleoprotein complex architectures resemble the *E. coli* Cascade complex. The same observation was made for the *P. furiosus* Cmr complex, which was modeled by using cross-linking-MS distance constraints to fit individual Cas protein crystal structures into a low-resolution EM map (Benda *et al*, 2014). Taken together, these results strongly suggest that type I and type III CRISPR–Cas systems share some common evolutionary ancestry.

An aspect that has long been neglected in structural studies of CRISPR–Cas systems is the precise mapping of the CRISPR RNA–protein binding interfaces. First attempts to pinpoint these interactions were made during characterization of the *T. thermophilus* Csm complex, in which contacts between all Csm subunits and their cognate RNA were revealed using UV-induced RNA–protein cross-linking and MS (Staals *et al*, 2014). This strategy was recently shown to be generally applicable for RNA interaction mapping of CRISPR-associated proteins (Hrle *et al*, 2014; Sharma *et al*, 2015), promising even more comprehensive structural maps of CRISPR–Cas systems in the future.

#### Concluding remarks

The continuous technological progress of MS provides opportunities to probe the structure and function of biomolecular systems with increasing analytical depth. Most importantly, MS experiments typically yield information that is complementary to the aspects monitored by traditional biochemical or structural biology approaches. While individual methods are often insufficient to understand highly complex and dynamically interacting biomolecular machineries, their characterization can be achieved by merging the unique benefits of diverse analytical techniques, as we have exemplified on CRISPR–Cas and transcription-related complexes. Closing in upon the *in vivo* architecture of such cellular key players, the focus of integrated structural studies is moving from recombinantly produced complexes to endogenously existing biomolecular assemblies, which also have become amenable to native MS and cross-linking-MS characterization. Evidently, the ultimate goal is to elucidate even more intricate systems, for example, cellular signaling pathways, organelles, and, eventually, entire cells. Here, in particular cross-linking-MS will likely prove as an ideal complement

to emerging *in vivo* and *in situ* technologies such as live-cell imaging, in-/on-cell NMR, and cryo-electron tomography. However, also other MS-based approaches (e.g., protein surface labeling and limited proteolysis) are extending their scope toward proteome-wide structural studies, allowing them to pull their weight in integrated analytical strategies. The future of MS-based approaches in structural and molecular biology is bright!

### Acknowledgements

We thank all members of our group, in particular those working with the here-described MS methods, for stimulating discussions and helpful comments. Moreover, we would like to acknowledge our national and international collaborators for continuously supporting our research. The work in the Heck laboratory is funded by the Roadmap Initiative Proteins@Work (project number 184.032.201) financed by The Netherlands Organisation for Scientific Research (NWO), and the MSMed Program (grant agreement number 686547) within the European Union's Horizon 2020 Framework Programme. Additional funding was received through the ManiFold project (grant agreement number 317371), embedded in the European Union 7th Framework Programme, and a Projectruimte grant (12PR3303-2) from Fundamenteel Onderzoek der Materie (FOM).

### Conflict of interest

The authors declare that they have no conflict of interest.

### References

- Allen BL, Taatjes DJ (2015) The mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* 16: 155–166
- Alley SC, Ishmael FT, Jones AD, Benkovic SJ (2000) Mapping protein–protein interactions in the bacteriophage T4 DNA polymerase holoenzyme using a novel trifunctional photo-cross-linking and affinity reagent. *J Am Chem Soc* 122: 6126–6127
- Altelaar AFM, Munoz J, Heck AJR (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* 14: 35–48
- Amstalden van Hove ER, Smith DF, Heeren RMA (2010) A concise review of mass spectrometry imaging. *J Chromatogr A* 1217: 3946–3954
- Aston FW (1920) Isotopes and atomic weights. *Nature* 105: 617–619
- Aye TT, Low TY, Sze SK (2005) Nanosecond laser-induced photochemical oxidation method for protein surface mapping with mass spectrometry. *Anal Chem* 77: 5814–5822
- Baldwin MA, McLafferty FW (1973) Direct chemical ionization of relatively involatile samples. Application to underivatized oligopeptides. *Org Mass Spectrom* 7: 1353–1356
- Barrera NP, Di Bartolo N, Booth PJ, Robinson CV (2008) Micelles protect membrane complexes from solution to vacuum. *Science* 321: 243–246
- Bayley H, Knowles JR (1977) Photoaffinity labeling. *Meth Enzymol* 46: 69–114
- Benda C, Ebert J, Scheltema RA, Schiller HB, Baumgärtner M, Bonneau F, Mann M, Conti E (2014) Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4. *Mol Cell* 56: 43–54
- Bensimon A, Heck AJR, Aebersold R (2012) Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem* 81: 379–405
- Bereszczak JZ, Rose RJ, van Duijn E, Watts NR, Wingfield PT, Steven AC, Heck AJR (2013) Epitope-distal effects accompany the binding of two distinct antibodies to hepatitis B virus capsids. *J Am Chem Soc* 135: 6504–6512
- Beynon JH, Cooks RG, Amy JW, Baitinger WE, Ridley TY (1973) Design and performance of a mass-analyzed ion kinetic energy (MIKE) spectrometer. *Anal Chem* 45: 1023A–1031A
- Biemann K (1990) Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Meth Enzymol* 193: 886–887
- Blattner C, Jennebach S, Herzog F, Mayer A, Cheung AC, Witte G, Lorenzen K, Hopfner KP, Heck AJR, Aebersold R, Cramer P (2011) Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes Dev* 25: 2093–2105
- Bogdanov B, Smith RD (2005) Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom Rev* 24: 168–200
- Bothner B, Dong XF, Bibbs L, Johnson JE, Siuzdak G (1998) Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *J Biol Chem* 273: 673–676
- Brodbeck JS (2014) Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem Soc Rev* 43: 2757–2783
- Bruce C, Stone K, Gulcicek E, Williams K (2013) Proteomics and the analysis of proteomic data: 2013 overview of current protein-profiling technologies. In *Current protocols in bioinformatics*, Baxevanis AD, Petsko GA, Stein LD, Stormo GD (eds), pp 13.1.1–13.1.31. Hoboken, NJ, USA: John Wiley & Sons, Inc
- Brunner AM, Lössl P, Liu F, Huguet R, Müller CW, Yamashita M, Zabrouskov V, Makarov A, Altelaar AFM, Heck AJR (2015) Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteiform characterization. *Anal Chem* 87: 4152–4158
- Bui KH, von Appen A, DiGuilio AL, Ori A, Sparks L, Mackmull M, Bock T, Hagen WJH, Andrés-Pons A, Glavy JS, Beck M (2013) Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* 155: 1233–1243
- Buncherd H, Roseboom W, de Koning LJ, de Koster CG, de Jong L (2014) A gas phase cleavage reaction of cross-linked peptides for protein complex topology studies by peptide fragment fingerprinting from large sequence database. *J Proteomics* 108: 65–77
- Bush DR, Zang L, Belov AM, Ivanov AR, Karger BL (2016) High resolution CZE-MS quantitative characterization of intact biopharmaceutical proteins: proteoforms of interferon- $\beta$ 1. *Anal Chem* 88: 1138–1146
- Bushnell DA, Kornberg RD (2003) Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription. *Proc Natl Acad Sci USA* 100: 6969–6973
- Cai W, Guner H, Gregorich ZR, Chen AJ, Ayaz-Guner S, Peng Y, Valeja SG, Liu X, Ge Y (2016) MASH Suite Pro: a comprehensive software tool for top-down proteomics. *Mol Cell Proteomics* 15: 703–714
- Cappadona S, Baker PR, Cutillas PR, Heck AJR, van Breukelen B (2012) Current challenges in software solutions for mass spectrometry-based quantitative proteomics. *Amino Acids* 43: 1087–1108
- Cevher MA, Shi Y, Li D, Chait BT, Malik S, Roeder RG (2014) Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nat Struct Mol Biol* 21: 1028–1034
- Chavez JD, Weisbrod CR, Zheng C, Eng JK, Bruce JE (2013) Protein interactions, post-translational modifications and topologies in human cells. *Mol Cell Proteomics* 12: 1451–1467
- Chen EI, McClatchy D, Park SK, Yates JR III (2008) Comparisons of mass spectrometry compatible surfactants for global analysis of the mammalian brain proteome. *Anal Chem* 80: 8694–8701
- Chen J, Rempel DL, Gross ML (2010a) Temperature jump and fast photochemical oxidation probe submillisecond protein folding. *J Am Chem Soc* 132: 15502–15504
- Chen ZA, Jawhari A, Fischer L, Buchen C, Tahir S, Kamenski T, Rasmussen M, Larivière L, Bukowski-Wills J, Nilges M, Cramer P, Rappsilber J (2010b)

- Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J* 29: 717–726
- Chi A, Bai DL, Geer LY, Shabanowitz J, Hunt DF (2007) Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry. *Int J Mass Spectrom* 259: 197–203
- Chung KY, Rasmussen SGF, Liu T, Li S, DeVree BT, Chae PS, Calinski D, Kobilka BK, Woods VL, Sunahara RK (2011) Conformational changes in the G protein Gs induced by the  $\beta_2$  adrenergic receptor. *Nature* 477: 611–615
- Clark SM, Konermann L (2004) Determination of ligand-protein dissociation constants by electrospray mass spectrometry-based diffusion measurements. *Anal Chem* 76: 7077–7083
- Clegg C, Hayes D (1974) Identification of neighbouring proteins in the ribosomes of *Escherichia coli*. A topographical study with the cross-linking reagent dimethyl suberimidate. *Eur J Biochem* 42: 21–28
- Clemmer DE, Hudgins RR, Jarrold MF (1995) Naked protein conformations: cytochrome c in the gas phase. *J Am Chem Soc* 117: 10141–10142
- Clemmer DE, Jarrold MF (1997) Ion mobility measurements and their applications to clusters and biomolecules. *J Mass Spectrom* 32: 577–592
- Cohen SL, Ferré-D'Amaré AR, Burley SK, Chait BT (1995) Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. *Protein Sci* 4: 1088–1099
- Cooks RG, Rockwood AL (1991) Letters to the editor The 'Thomson'. A suggested unit for mass spectroscopists. *Rapid Commun Mass Spectrom* 5: 93
- Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367–1372
- Cox J, Mann M (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* 80: 273–299
- D'Arcy S, Martin KW, Panchenko T, Chen X, Bergeron S, Stargell LA, Black BE, Luger K (2013) Chaperone Nap1 shields histone surfaces used in a nucleosome and can put H2A-H2B in an unconventional tetrameric form. *Mol Cell* 51: 662–677
- Diebold CA, Beurskens FJ, de Jong RN, Koning RI, Strumane K, Lindorfer MA, Voorhorst M, Ugurlar D, Rosati S, Heck AJR, van de Winkel JG, Wilson IA, Koster AJ, Taylor RP, Saphire EO, Burton Schuurman J, Gros P, Parren PW (2014) Complement is activated by IgG hexamers assembled at the cell surface. *Science* 343: 1260–1263
- Domon B, Costello CE (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate J* 5: 397–409
- van Duijn E, Bakkes PJ, Heeren RMA, van den Heuvel RHH, van Heerikhuizen H, van der Vies SM, Heck AJR (2005) Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nat Meth* 2: 371–376
- van Duijn E, Barbu IM, Barendregt A, Jore MM, Wiedenheft B, Lundgren M, Westra ER, Brouns SJJ, Doudna JA, van der Oost J, Heck AJR (2012) 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* 11: 1430–1441
- van Duijn E, Barendregt A, Synowsky S, Versluis C, Heck AJR (2009) Chaperonin complexes monitored by ion mobility mass spectrometry. *J Am Chem Soc* 131: 1452–1459
- Dyachenko A, Gruber R, Shimon L, Horovitz A, Sharon M (2013) Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proc Natl Acad Sci USA* 110: 7235–7239
- Dyachenko A, Wang G, Belov M, Makarov A, de Jong RN, van den Bremer ETJ, Parren PW, Heck AJR (2015) Tandem native mass-spectrometry on antibody-drug conjugates and submillion da antibody-antigen protein assemblies on an orbitrap EMR equipped with a high-mass quadrupole mass selector. *Anal Chem* 87: 6095–6102
- El-Hawiet A, Kitova EN, Arutyunov D, Simpson DJ, Szymanski CM, Klassen JS (2012a) Quantifying ligand binding to large protein complexes using electrospray ionization mass spectrometry. *Anal Chem* 84: 3867–3870
- El-Hawiet A, Shoemaker GK, Daneshfar R, Kitova EN, Klassen JS (2012b) Applications of a catch and release electrospray ionization mass spectrometry assay for carbohydrate library screening. *Anal Chem* 84: 50–58
- Engel C, Sainsbury S, Cheung AC, Kostrewa D, Cramer P (2013) RNA polymerase I structure and transcription regulation. *Nature* 502: 650–655
- Englander SW, Downer NW, Teitelbaum H (1972) Hydrogen exchange. *Annu Rev Biochem* 41: 903–924
- Englander SW (2006) Hydrogen exchange and mass spectrometry: a historical perspective. *J Am Soc Mass Spectrom* 17: 1481–1489
- Ens W, Standing KG (2005) Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules. *Meth Enzymol* 402: 49–78
- Erzberger JP, Stengel F, Pellarin R, Zhang S, Schaefer T, Aylett CHS, Cimermančić P, Boehringer D, Sali A, Aebersold R, Ban N (2014) Molecular architecture of the 40S-eIF1-eIF3 translation initiation complex. *Cell* 158: 1123–1135
- Espino JA, Mali VS, Jones LM (2015) In cell footprinting coupled with mass spectrometry for the structural analysis of proteins in live cells. *Anal Chem* 87: 7971–7978
- Fellers RT, Greer JB, Early BP, Yu X, Leduc RD, Kelleher NL, Thomas PM (2015) ProSight Lite: graphical software to analyze top-down mass spectrometry data. *Proteomics* 15: 1235–1238
- Feng Y, de Franceschi G, Kahraman A, Soste M, Melnik A, Boersema PJ, de Lauro PP, Nikolaev Y, Oliveira AP, Picotti P (2014) Global analysis of protein structural changes in complex proteomes. *Nat Biotechnol* 32: 1036–1044
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246: 64–71
- Fernández-Tornero C, Moreno-Morcillo M, Rashid UJ, Taylor NMI, Ruiz FM, Gruene T, Legrand P, Steuerwald U, Müller CW (2013) Crystal structure of the 14-subunit RNA polymerase I. *Nature* 502: 644–649
- Fischer L, Chen ZA, Rappsilber J (2013) Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J Proteomics* 88: 120–128
- Fitzgerald MC, Chernushevich I, Standing KG, Whitman CP, Kent SB (1996) Probing the oligomeric structure of an enzyme by electrospray ionization time-of-flight mass spectrometry. *Proc Natl Acad Sci USA* 93: 6851–6856
- Flora JW, Muddiman DC (1998) Comprehensive nomenclature for the fragment ions produced from collisional activation of peptide nucleic acids. *Rapid Commun Mass Spectrom* 12: 759–762
- Fontana A, Fassina G, Vita C, Dalzoppo D, Zamai M, Zamboni M (1986) Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* 25: 1847–1851
- Fornelli L, Damoc E, Thomas PM, Kelleher NL, Aizikov K, Denisov E, Makarov A, Tsybin YO (2012) Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation orbitrap FTMS. *Mol Cell Proteomics* 11: 1758–1767
- Fraenkel-Conrat H, Olcott HS (1946) Reaction of formaldehyde with proteins. II. Participation of the guanidyl groups and evidence of crosslinking. *J Am Chem Soc* 68: 34–37

- Frese CK, Altelaar AFM, van den Toorn H, Nolting D, Griep-Raming J, Heck AJR, Mohammed S (2012) Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal Chem* 84: 9668–9673
- Fuerstenau SD, Benner WH, Thomas JJ, Brugidou C, Bothner B, Siuzdak G (2001) Mass spectrometry of an intact virus. *Angew Chem Int Ed Engl* 40: 541–544
- Ganem B, Li YT, Henion JD (1991) Detection of noncovalent receptor-ligand complexes by mass spectrometry. *J Am Chem Soc* 113: 6294–6296
- Gavin A, Bösch M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon A, Cruciat C, Remor M, Höfert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T et al (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415: 141–147
- Geiger SR, Lorenzen K, Schrieck A, Hanecker P, Kostrewa D, Heck AJR, Cramer P (2010) RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol Cell* 39: 583–594
- Georgescauld F, Popova K, Gupta AJ, Bracher A, Engen JR, Hayer-Hartl M, Hartl FU (2014) GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. *Cell* 157: 922–934
- Gervasoni P, Staudenmann W, James P, Plückthun A (1998) Identification of the binding surface on  $\beta$ -lactamase for GroEL by limited proteolysis and MALDI-mass spectrometry †. *Biochemistry* 37: 11660–11669
- Greber BJ, Boehringer D, Leibundgut M, Bieri P, Leitner A, Schmitz N, Aebersold R, Ban N (2014) The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* 515: 283–286
- Greber BJ, Bieri P, Leibundgut M, Leitner A, Aebersold R, Boehringer D, Ban N (2015) Ribosome. The complete structure of the 55S mammalian mitochondrial ribosome. *Science* 348: 303–308
- Guerrero C, Tagwerker C, Kaiser P, Huang L (2006) An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified *in vivo* cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. *Mol Cell Proteomics* 5: 366–378
- Haddon WF, McLafferty FW (1968) Metastable ion characteristics. VII. Collision-induced metastables. *J Am Chem Soc* 90: 4745–4746
- Hambly DM, Gross ML (2005) Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *J Am Soc Mass Spectrom* 16: 2057–2063
- Han L, Tan M, Xia M, Kitova EN, Jiang X, Klassen JS (2014a) Gangliosides are ligands for human noroviruses. *J Am Chem Soc* 136: 12631–12637
- Han X, Wang Y, Aslanian A, Bern M, Lavallée-Adam M, Yates JR III (2014b) Sheathless capillary electrophoresis-tandem mass spectrometry for top-down characterization of *pyrococcus furiosus* proteins on a proteome scale. *Anal Chem* 86: 11006–11012
- Han Y, Luo J, Ranish J, Hahn S (2014c) Architecture of the *Saccharomyces cerevisiae* SAGA transcription coactivator complex. *EMBO J* 33: 2534–2546
- Hebert AS, Richards AL, Bailey DJ, Ulbrich A, Coughlin EE, Westphall MS, Coon JJ (2014) The one hour yeast proteome. *Mol Cell Proteomics* 13: 339–347
- Heidemann M, Hintermair C, Voß K, Eick D (2013) Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta* 1829: 55–62
- Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J, Buchholz F, Hyman AA, Mann M (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163: 712–723
- Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci USA* 90: 5011–5015
- Hernández H, Dziembowski A, Taverner T, Séraphin B, Robinson CV (2006) Subunit architecture of multimeric complexes isolated directly from cells. *EMBO Rep* 7: 605–610
- Herzog F, Kahraman A, Boehringer D, Mak R, Bracher A, Walzthoeni T, Leitner A, Beck M, Hartl FU, Ban N, Malmström L, Aebersold R (2012) Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* 337: 1348–1352
- van den Heuvel RHH, Heck AJR (2004) Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr Opin Chem Biol* 8: 519–526
- van den Heuvel RHH, van Duijn E, Mazon H, Synowsky SA, Lorenzen K, Versluis C, Brouns SJJ, Langridge D, van der Oost J, Hoyes J, Heck AJR (2006) Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* 78: 7473–7483
- Hoffmann NA, Jakobi AJ, Moreno-Morcillo M, Glatt S, Kosinski J, Hagen WJH, Sachse C, Müller CW (2015) Molecular structures of unbound and transcribing RNA polymerase III. *Nature* 528: 231–236
- Hofstadler SA, Sannes-Lowery KA (2006) Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes. *Nat Rev Drug Discov* 5: 585–595
- Hopper JTS, Yu YT, Li D, Raymond A, Bostock M, Liko I, Mikhailov V, Laganowsky A, Benesch JLP, Caffrey M, Nietlispach D, Robinson CV (2013) Detergent-free mass spectrometry of membrane protein complexes. *Nat Meth* 10: 1206–1208
- Horn DM, Ge Y, McLafferty FW (2000) Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins. *Anal Chem* 72: 4778–4784
- Hrle A, Maier LK, Sharma K, Ebert J, Basquin C, Urlaub H, Marchfelder A, Conti E (2014) Structural analyses of the CRISPR protein Csc2 reveal the RNA-binding interface of the type I-D Cas7 family. *RNA Biol* 11: 1072–1082
- Hubbard SJ (1998) The structural aspects of limited proteolysis of native proteins. *BBA Protein Struct Mol Enzymol* 1382: 191–206
- Hunter T (1995) Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell* 80: 225–236
- Huttlin EL, Ting L, Bruckner RJ, Gebreb F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, Dong R, Guarani V, Vaites LP, Ordureau A, Rad R, Erickson BK, Wühr M, Chick J, Zhai B, Kolipakkam D et al (2015) The BioPlex network: a systematic exploration of the human interactome. *Cell* 162: 425–440
- Jennebach S, Herzog F, Aebersold R, Cramer P (2012) Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res* 40: 5591–5601
- Jennings KR (1968) Collision-induced decompositions of aromatic molecular ions. *Int J Mass Spectrom Ion Phys* 1: 227–235
- Johnson LN (2009) The regulation of protein phosphorylation. *Biochem Soc Trans* 37: 627–641
- Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders APL, Dickman MJ, Doudna JA, Boekema EJ, Heck AJR, van der Oost J, Brouns SJJ (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* 18: 529–536
- Kaake RM, Wang X, Burke A, Yu C, Kandur W, Yang Y, Novtisky EJ, Second T, Duan J, Kao A, Guan S, Vellucci D, Rychnovsky SD, Huang L (2014) A new *in vivo* cross-linking mass spectrometry platform to define protein-protein interactions in living cells. *Mol Cell Proteomics* 13: 3533–3543

- Kalisman N, Adams CM, Levitt M (2012) Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *Proc Natl Acad Sci USA* 109: 2884–2889
- Kan Z, Mayne L, Chetty PS, Englander SW (2011) ExMS: data analysis for HX-MS experiments. *J Am Soc Mass Spectrom* 22: 1906–1915
- Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60: 2299–2301
- Katta V, Chait BT (1991) Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *J Am Chem Soc* 113: 8534–8535
- Katta V, Chait BT (1993) Hydrogen/deuterium exchange electrospray ionization mass spectrometry: a method for probing protein conformational changes in solution. *J Am Chem Soc* 115: 6317–6321
- Kelleher NL, Lin HY, Valaskovic GA, Aaserud DJ, Fridriksson EK, McLafferty FW (1999) Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J Am Chem Soc* 121: 806–812
- Kim KC, Uckotter M, Beynon JH, Cooks RG (1974) Collision-induced fragmentation of triatomic ions. *Int J Mass Spectrom Ion Phys* 15: 23–35
- Kim M, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabudhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LDN et al (2014) A draft map of the human proteome. *Nature* 509: 575–581
- Klapper MH, Klotz IM (1972) [46] Acylation with dicarboxylic acid anhydrides. In *Enzyme structure*, Part B. London, UK: Elsevier
- Knutson BA, Luo J, Ranish J, Hahn S (2014) Architecture of the *Saccharomyces cerevisiae* RNA polymerase I core factor complex. *Nat Struct Mol Biol* 21: 810–816
- Konermann L, Pan J, Liu Y (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem Soc Rev* 40: 1224–1234
- Koschubs T, Lorenzen K, Baumli S, Sandström S, Heck AJR, Cramer P (2010) Preparation and topology of the Mediator middle module. *Nucleic Acids Res* 38: 3186–3195
- Kosinski J, von Appen A, Ori A, Karius K, Müller CW, Beck M (2015) Xlink analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J Struct Biol* 189: 177–183
- Kostrewa D, Zeller ME, Armache K, Seizl M, Leike K, Thomm M, Cramer P (2009) RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* 462: 323–330
- Laganowsky A, Reading E, Allison TM, Ulmschneider MB, Degiacomi MT, Baldwin AJ, Robinson CV (2014) Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* 510: 172–175
- Lane LA, Fernández-Tornero C, Zhou M, Morgner N, Ptchelkine D, Steuerwald U, Politis A, Lindner D, Gvozdenovic J, Gavin AC, Müller CW, Robinson CV (2011) Mass spectrometry reveals stable modules in holo and apo RNA polymerases I and III. *Structure* 19: 90–100
- Langman J, Lam TT, Emmett MR, Marshall AG, Sakalian M, Prevelige PE (2004) Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. *Nat Struct Mol Biol* 11: 676–677
- Lanucara F, Holman SW, Gray CJ, Evers CE (2014) The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat Chem* 6: 281–294
- Larivière L, Plaschka C, Seizl M, Petrotchenko EV, Wenzek L, Borchers CH, Cramer P (2013) Model of the Mediator middle module based on protein cross-linking. *Nucleic Acids Res* 41: 9266–9273
- Lee DCH, Jones AR, Hubbard SJ (2015) Computational phosphoproteomics: from identification to localization. *Proteomics* 15: 950–963
- Leitner A, Walzthoeni T, Kahraman A, Herzog F, Rinner O, Beck M, Aebersold R (2010) Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics* 9: 1634–1649
- Levsen K, Schwarz H (1976) Collisional activation mass spectrometry?: a new probe for determining the structure of ions in the gas phase. *Angew Chem Int Ed Engl* 15: 509–519
- Li X, Wang W, Chen J (2015) From pathways to networks: connecting dots by establishing protein-protein interaction networks in signaling pathways using affinity purification and mass spectrometry. *Proteomics* 15: 188–202
- Light-Wahl KJ, Winger BE, Smith RD (1993) Observation of the multimeric forms of concanavalin A by electrospray ionization mass spectrometry. *J Am Chem Soc* 115: 5869–5870
- Lin HY, Kitova EN, Klassen JS (2014) Measuring positive cooperativity using the direct ESI-MS assay. Cholera toxin B subunit homopentamer binding to GM1 pentasaccharide. *J Am Soc Mass Spectrom* 25: 104–110
- Linderstrøm-Lang K (1950) Structure and enzymatic break-down of proteins. *Cold Spring Harb Symp Quant Biol* 14: 117–126
- Liu X, Bushnell DA, Wang D, Calero G, Kornberg RD (2010) Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* 327: 206–209
- Liu X, Sirotkin Y, Shen Y, Anderson G, Tsai YS, Ting YS, Goodlett DR, Smith RD, Bafna V, Pevzner PA (2012) Protein identification using top-down spectra. *Mol Cell Proteomics* 11: M111.008524
- Liu F, van Breukelen B, Heck AJR (2014) Facilitating protein disulfide mapping by a combination of pepsin digestion, electron transfer higher energy dissociation (ETHcD), and a dedicated search algorithm SlinkS. *Mol Cell Proteomics* 13: 2776–2786
- Liu F, Rijkers DTS, Post H, Heck AJR (2015) Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat Meth* 12: 1179–1184
- Lorenzen K, Vannini A, Cramer P, Heck AJR (2007) Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* 15: 1237–1245
- Lössl P, Kölbl K, Tänzler D, Nannemann D, Ihling CH, Keller MV, Schneider M, Zaucke F, Meiler J, Sinz A (2014) Analysis of nidogen-1/laminin  $\gamma$ 1 interaction by cross-linking, mass spectrometry, and computational modeling reveals multiple binding modes. *PLoS ONE* 9: e112886
- Lössl P, Brunner AM, Liu F, Leney AC, Yamashita M, Scheltema RA, Heck AJR (2016) Deciphering the interplay among multisite phosphorylation, interaction dynamics, and conformational transitions in a tripartite protein system. *ACS Cent Sci* 2: 445–455
- Madsen JA, Boutz DR, Brodbelt JS (2010) Ultrafast ultraviolet photodissociation at 193 nm and its applicability to proteomic workflows. *J Proteome Res* 9: 4205–4214
- Male G, von Appen A, Glatt S, Taylor NMI, Cristovao M, Groetsch H, Beck M, Müller CW (2015) Architecture of TFIIC and its role in RNA polymerase III pre-initiation complex assembly. *Nat Commun* 6: 7387
- Maleknia SD, Downard KM (2014) Advances in radical probe mass spectrometry for protein footprinting in chemical biology applications. *Chem Soc Rev* 43: 3244–3258
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* 339: 823–826
- Maple HJ, Scheibner O, Baumert M, Allen M, Taylor RJ, Garlish RA, Bromirski M, Burnley RJ (2014) Application of the exactive plus EMR for automated protein-ligand screening by non-covalent mass spectrometry. *Rapid Commun Mass Spectrom* 28: 1561–1568

- Marcoux J, Robinson CV (2013) Twenty years of gas phase structural biology. *Structure* 21: 1541–1550
- Marcoux J, Wang SC, Politis A, Reading E, Ma J, Biggin PC, Zhou M, Tao H, Zhang Q, Chang G, Morgner N, Robinson CV (2013) Mass spectrometry reveals synergistic effects of nucleotides, lipids, and drugs binding to a multidrug resistance efflux pump. *Proc Natl Acad Sci USA* 110: 9704–9709
- Marcoux J, Champion T, Colas O, Wagner-Rousset E, Corvaia N, van Dorsselaer A, Beck A, Cianférani S (2015) Native mass spectrometry and ion mobility characterization of trastuzumab emtansine, a lysine-linked antibody drug conjugate. *Protein Sci* 24: 1210–1223
- Marshall AG, Hendrickson CL (2008) High-resolution mass spectrometers. *Annu Rev Anal Chem* 1: 579–599
- Martinez-Rucobo FW, Kohler R, van de Waterbeemd M, Heck AJR, Hemann M, Herzog F, Stark H, Cramer P (2015) Molecular basis of transcription-coupled pre-mRNA capping. *Mol Cell* 58: 1079–1089
- McCammon MG, Hernández H, Sobott F, Robinson CV (2004) Tandem mass spectrometry defines the stoichiometry and quaternary structural arrangement of tryptophan molecules in the multiprotein complex TRAP. *J Am Chem Soc* 126: 5950–5951
- McCormack AL, Schieltz DM, Goode B, Yang S, Barnes G, Drubin D, Yates JR III (1997) Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal Chem* 69: 767–776
- McLuckey SA, Stephenson JL (1998) Ion/ion chemistry of high-mass multiply charged ions. *Mass Spectrom Rev* 17: 369–407
- Mendoza VL, Vachet RW (2009) Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom Rev* 28: 785–815
- Mühlbacher W, Sainsbury S, Hemann M, Hantsche M, Neyer S, Herzog F, Cramer P (2014) Conserved architecture of the core RNA polymerase II initiation complex. *Nat Commun* 5: 4310
- Müller MQ, Dreiocker F, Ihling CH, Schäfer M, Sinz A (2010) Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. *Anal Chem* 82: 6958–6968
- Munson MSB, Franklin JL, Field FH (1964) High pressure mass spectrometric study of alkanes 1. *J Phys Chem* 68: 3098–3107
- Munson MSB, Field FH (1966) Chemical ionization mass spectrometry. I. General introduction. *J Am Chem Soc* 88: 2621–2630
- Murakami K, Elmlund H, Kalisman N, Bushnell DA, Adams CM, Azubel M, Elmlund D, Levi-Kalisman Y, Liu X, Gibbons BJ, Levitt M, Kornberg RD (2013) Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* 342: 1238724
- Navare AT, Chavez JD, Zheng C, Weisbrod CR, Eng JK, Siehnel R, Singh PK, Manoil C, Bruce JE (2015) Probing the protein interaction network of *Pseudomonas aeruginosa* cells by chemical cross-linking mass spectrometry. *Structure* 23: 762–773
- Niu S, Rabuck JN, Ruotolo BT (2013) Ion mobility-mass spectrometry of intact protein–ligand complexes for pharmaceutical drug discovery and development. *Curr Opin Chem Biol* 17: 809–817
- van der Oost J, Westra ER, Jackson RN, Wiedenheft B (2014) Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat Rev Micro* 12: 479–492
- Pan J, Zhang S, Borchers CH (2016) Protein species-specific characterization of conformational change induced by multisite phosphorylation. *J Proteomics* 134: 138–143
- Pascal BD, Willis S, Lauer MJ, Landgraf RR, West GM, Marciano D, Novick S, Goswami D, Chalmers MJ, Griffin PR (2012) HDX workbench: software for the analysis of H/D exchange MS data. *J Am Soc Mass Spectrom* 23: 1512–1521
- Paul W (1990) Electromagnetic traps for charged and neutral particles. *Rev Mod Phys* 62: 531–540
- Perez-Riverol Y, Alpi E, Wang R, Hermjakob H, Vizcaíno JA (2015) Making proteomics data accessible and reusable: current state of proteomics databases and repositories. *Proteomics* 15: 930–950
- Petrovchenko EV, Borchers CH (2010) Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom Rev* 29: 862–876
- Picotti P, Aebersold R (2012) Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. *Nat Meth* 9: 555–566
- Pirrone GF, Iacob RE, Engen JR (2015) Applications of hydrogen/deuterium exchange MS from 2012 to 2014. *Anal Chem* 87: 99–118
- Plaschka C, Larivière L, Wenzek L, Seizl M, Hemann M, Tegunov D, Petrovchenko EV, Borchers CH, Baumeister W, Herzog F, Villa E, Cramer P (2015) Architecture of the RNA polymerase II–Mediator core initiation complex. *Nature* 518: 376–380
- Porter CJ, Beynon JH, Ast T (1981) The modern mass spectrometer—a complete chemical laboratory. *Org Mass Spectrom* 16: 101–114
- Pringle SD, Giles K, Wildgoose JL, Williams JP, Slade SE, Thalassinis K, Bateman RH, Bowers MT, Scrivens JH (2007) An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int J Mass Spectrom* 261: 1–12
- Rappsilber J, Siniosoglou S, Hurt EC, Mann M (2000) A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal Chem* 72: 267–275
- Reinhard FBM, Eberhard D, Werner T, Franken H, Childs D, Doce C, Savitski MF, Huber W, Bantscheff M, Savitski MM, Drewes G (2015) Thermal proteome profiling monitors ligand interactions with cellular membrane proteins. *Nat Methods* 12: 1129–1131
- Rey M, Sarpe V, Burns KM, Buse J, Baker CAH, van Dijk M, Wordeman L, Bonvin AMJJ, Schriemer DC (2014) Mass spec studio for integrative structural biology. *Structure* 22: 1538–1548
- Richards AL, Merrill AE, Coon JJ (2015) Proteome sequencing goes deep. *Curr Opin Chem Biol* 24: 11–17
- Riley NM, Coon JJ (2016) Phosphoproteomics in the age of rapid and deep proteome profiling. *Anal Chem* 88: 74–94
- Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, Mueller M, Aebersold R (2008) Identification of cross-linked peptides from large sequence databases. *Nat Meth* 5: 315–318
- Robinette D, Neamati N, Tomer KB, Borchers CH (2006) Photoaffinity labeling combined with mass spectrometric approaches as a tool for structural proteomics. *Expert Rev Proteomics* 3: 399–408
- Robinson PJ, Bushnell DA, Trnka MJ, Burlingame AL, Kornberg RD (2012) Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc Natl Acad Sci USA* 109: 17931–17935
- Robinson PJ, Trnka MJ, Pellarin R, Greenberg CH, Bushnell DA, Davis R, Burlingame AL, Sali A, Kornberg RD (2015) Molecular architecture of the yeast Mediator complex. *eLife* 4: e08719
- Robinson PJ, Trnka MJ, Bushnell DA, Davis RE, Mattei PJ, Burlingame AL, Kornberg RD (2016) Structure of a Complete Mediator–RNA Polymerase II Pre-Initiation Complex. *Cell* 166: 1411–1422
- Roepstorff P, Fohlman J (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 11: 601
- Rosati S, van den Bremer ETJ, Schuurman J, Parren PWHI, Kamerling JP, Heck AJR (2013) 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* 5: 917–924
- Rosati S, Yang Y, Barendregt A, Heck AJR (2014) Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. *Nat Protoc* 9: 967–976
- Rose RJ, Damoc E, Denisov E, Makarov A, Heck AJR (2012) High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat Meth* 9: 1084–1086
- Rostislavleva K, Soler N, Ohashi Y, Zhang L, Pardon E, Burke JE, Masson GR, Johnson C, Steyaert J, Ktistakis NT, Williams RL (2015) Structure and flexibility of the endosomal Vps34 complex reveals the basis of its function on membranes. *Science* 350: aac7365–aac7365
- Rostom AA, Fucini P, Benjamin DR, Juenemann R, Nierhaus KH, Hartl FU, Dobson CM, Robinson CV (2000) Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proc Natl Acad Sci USA* 97: 5185–5190
- Rouillon C, Zhou M, Zhang J, Politis A, Beilsten-Edmands V, Cannone G, Graham S, Robinson CV, Spagnolo L, White MF (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol Cell* 52: 124–134
- Ruotolo BT, Giles K, Campuzano I, Sandercock AM, Bateman RH, Robinson CV (2005) Evidence for macromolecular protein rings in the absence of bulk water. *Science* 310: 1658–1661
- Ruotolo BT, Benesch JLP, Sandercock AM, Hyung S, Robinson CV (2008) Ion mobility–mass spectrometry analysis of large protein complexes. *Nat Protoc* 3: 1139–1152
- Rurup WF, Snijder J, Koay MST, Heck AJR, Cornelissen Jeroen JLM (2014) Self-sorting of foreign proteins in a bacterial nanocompartment. *J Am Chem Soc* 136: 3828–3832
- Sainsbury S, Niesser J, Cramer P (2012) Structure and function of the initially transcribing RNA polymerase II–TFIIB complex. *Nature* 493: 437–440
- Sainsbury S, Bernecky C, Cramer P (2015) Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* 16: 129–143
- Savitski MM, Reinhard FBM, Franken H, Werner T, Savitski MF, Eberhard D, Martínez Molina D, Jafari R, Dovega RB, Klaefer S, Kuster B, Nordlund P, Bantscheff M, Drewes G (2014) Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* 346: 1255784
- Schmidt C, Kramer K, Urlaub H (2012) Investigation of protein–RNA interactions by mass spectrometry—Techniques and applications. *J Proteomics* 75: 3478–3494
- Schmidt C, Zhou M, Marriott H, Morgner N, Politis A, Robinson CV (2013) Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation. *Nat Commun* 4: 1985
- Schuller JM, Beck F, Lössl P, Heck AJR, Förster F (2016) Nucleotide-dependent conformational changes of the AAA+ ATPase p97 revisited. *FEBS Lett* 590: 595–604
- Schwamborn K, Caprioli RM (2010) Molecular imaging by mass spectrometry — looking beyond classical histology. *Nat Rev Cancer* 10: 639–646
- Schwartz BL, Light-Wahl KJ, Smith RD (1994) Observation of noncovalent complexes to the avidin tetramer by electrospray ionization mass spectrometry. *J Am Soc Mass Spectrom* 5: 201–204
- Sharma S, Simpson DC, Tolić N, Jaitly N, Mayampurath AM, Smith RD, Paša-Tolić L (2007) Proteomic profiling of intact proteins using WAX-PLC 2-D separations and FTICR mass spectrometry. *J Proteome Res* 6: 602–610
- Sharma K, Hrle A, Kramer K, Sachsenberg T, Staals RH, Randau L, Marchfelder A, van der Oost J, Kohlbacher O, Conti E, Urlaub H (2015) Analysis of protein–RNA interactions in CRISPR proteins and effector complexes by UV-induced cross-linking and mass spectrometry. *Methods* 89: 138–148
- Shaw JB, Li W, Holden DD, Zhang Y, Griep-Raming J, Fellers RT, Early BP, Thomas PM, Kelleher NL, Brodbelt JS (2013) Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J Am Chem Soc* 135: 12646–12651
- Shi Y, Fernandez-Martinez J, Tjioe E, Pellarin R, Kim SJ, Williams R, Schneidman-Duhovny D, Sali A, Rout MP, Chait BT (2014) Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol Cell Proteomics* 13: 2927–2943
- Shi Y, Pellarin R, Fridy PC, Fernandez-Martinez J, Thompson MK, Li Y, Wang QJ, Sali A, Rout MP, Chait BT (2015) A strategy for dissecting the architectures of native macromolecular assemblies. *Nat Meth* 12: 1135–1138
- Shukla AK, Westfield GH, Xiao K, Reis RI, Huang L, Tripathi-Shukla P, Qian J, Li S, Blanc A, Oleskie AN, Dosey AM, Su M, Liang C, Gu L, Shan J, Chen X, Hanna R, Choi M, Yao XJ, Klink BU et al (2014) Visualization of arrestin recruitment by a G-protein-coupled receptor. *Nature* 512: 218–222
- Sinz A (2003) Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J Mass Spectrom* 38: 1225–1237
- Slysz GW, Baker CAH, Bozsa BM, Dang A, Percy AJ, Bennett M, Schriemer DC (2009) Hydra: software for tailored processing of H/D exchange data from MS or tandem MS analyses. *BMC Bioinformatics* 10: 162
- Smith LG (1937) Ionization and dissociation of polyatomic molecules by electron impact I. Methane. *Phys Rev* 51: 263–275
- Snijder J, Heck AJR (2014) Analytical approaches for size and mass analysis of large protein assemblies. *Annu Rev Anal Chem* 7: 43–64
- Snijder J, van de Waterbeemd M, Damoc E, Denisov E, Grinfeld D, Bennett A, Agbandje-McKenna M, Makarov A, Heck AJR (2014) Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by orbitrap mass spectrometry. *J Am Chem Soc* 136: 7295–7299
- Sobott F, Hernández H, McCammon MG, Tito MA, Robinson CV (2002) A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem* 74: 1402–1407
- Soderblom EJ, Goshe MB (2006) Collision-induced dissociative chemical cross-linking reagents and methodology: applications to protein structural characterization using tandem mass spectrometry analysis. *Anal Chem* 78: 8059–8068
- Staals RH, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K, Masuda A, Dohmae N, Schaap PJ, Doudna JA, Heck AJR, Yonekura K, van der Oost J, Shinkai A (2013) Structure and activity of the RNA-targeting type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell* 52: 135–145
- Staals RH, Zhu Y, Taylor DW, Kornfeld JE, Sharma K, Barendregt A, Koehorst JJ, Vlot M, Neupane N, Varossieau K, Sakamoto K, Suzuki T, Dohmae N, Yokoyama S, Schaap PJ, Urlaub H, Heck AJR, Nogales E, Doudna JA, Shinkai A et al (2014) RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Mol Cell* 56: 518–530
- Steen H, Mann M (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 5: 699–711
- Stein SE (2016) Mass spectra. In *NIST chemistry webbook: NIST standard reference database number 69*, Linstrom PJ, Mallard WG (eds). Gaithersburg, MD: National Institute of Standards and Technology, <http://webbook.nist.gov> (accessed: March 3, 2016)
- Strupat K, Karas M, Hillenkamp F, Eckerskorn C, Lottspeich F (1994) Matrix-assisted laser desorption ionization mass spectrometry of proteins

- electroblotted after polyacrylamide gel electrophoresis. *Anal Chem* 66: 464–470
- Suckau D, Shi Y, Beu SC, Senko MW, Quinn JP, Wampler FM, McLafferty FW (1993) Coexisting stable conformations of gaseous protein ions. *Proc Natl Acad Sci USA* 90: 790–793
- Sun TT, Bollen A, Kahan L, Traut RR (1974) Topography of ribosomal proteins of the *Escherichia coli* 30S subunit as studied with the reversible cross-linking reagent methyl 4-mercaptobutyrimidate. *Biochemistry* 13: 2334–2340
- Swaney DL, McAlister GC, Wirtala M, Schwartz JC, Syka JEP, Coon JJ (2007) Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* 79: 477–485
- Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA* 101: 9528–9533
- Synowsky SA, van den Heuvel RHH, Mohammed S, Pijnappel PWW, Heck AJR (2006) Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex. *Mol Cell Proteomics* 5: 1581–1592
- Sze SK, Ge Y, Oh H, McLafferty FW (2002) Top-down mass spectrometry of a 29-kDa protein for characterization of any post-translational modification to within one residue. *Proc Natl Acad Sci USA* 99: 1774–1779
- Tabb DL (2013) Quality assessment for clinical proteomics. *Clin Biochem* 46: 411–420
- Tan D, Li Q, Zhang MJ, Liu C, Ma C, Zhang P, Ding YH, Fan SB, Tao L, Yang B, Li X, Ma S, Liu J, Feng B, Liu X, Wang HW, He SM, Gao N, Ye K, Dong MQ, Lei X (2016) Trifunctional cross-linker for mapping protein-protein interaction networks and comparing protein conformational states. *eLife* 5: e12509
- Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T, Matsuo T (1988) Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 2: 151–153
- Thomson JJ (1897) Cathode rays. *Philos Mag* 44: 293–316
- Thomson JJ (1911) XXVI. Rays of positive electricity. *Philos Mag* 21: 225–249
- Tosi A, Haas C, Herzog F, Gilmozzi A, Berninghausen O, Ungewickell C, Gerhold CB, Lakomek K, Aebersold R, Beckmann R, Hopfner KP (2013) Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* 154: 1207–1219
- Trabjerg E, Jakobsen RU, Mysling S, Christensen S, Jørgensen TJD, Rand KD (2015) Conformational analysis of large and highly disulfide-stabilized proteins by integrating online electrochemical reduction into an optimized H/D exchange mass spectrometry workflow. *Anal Chem* 87: 8880–8888
- Tran JC, Doucette AA (2009) Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Anal Chem* 81: 6201–6209
- Trester-Zedlitz M, Kamada K, Burley SK, Fenyö D, Chait BT, Muir TW (2003) A modular cross-linking approach for exploring protein interactions. *J Am Chem Soc* 125: 2416–2425
- Tsiatsiani L, Heck AJR (2015) Proteomics beyond trypsin. *FEBS J* 282: 2612–2626
- Utrecht C, Versluis C, Watts NR, Roos WH, Wuite GJL, Wingfield PT, Steven AC, Heck AJR (2008) High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc Natl Acad Sci USA* 105: 9216–9220
- Utrecht C, Barbu IM, Shoemaker GK, van Duijn E, Heck AJR (2010a) Interrogating viral capsid assembly with ion mobility-mass spectrometry. *Nat Chem* 3: 126–132
- Utrecht C, Rose RJ, van Duijn E, Lorenzen K, Heck AJR (2010b) Ion mobility mass spectrometry of proteins and protein assemblies. *Chem Soc Rev* 39: 1633–1655
- Urnivicus L, Zhang K, Diamant AG, Motz C, Schlager MA, Yu M, Patel NA, Robinson CV, Carter AP (2015) The structure of the dynactin complex and its interaction with dynein. *Science* 347: 1441–1446
- Vahidi S, Bi Y, Dunn SD, Konermann L (2016) Load-dependent destabilization of the  $\gamma$ -rotor shaft in F<sub>1</sub>O<sub>1</sub> ATP synthase revealed by hydrogen/deuterium-exchange mass spectrometry. *Proc Natl Acad Sci USA* 113: 2412–2417
- Vannini A (2013) A structural perspective on RNA polymerase I and RNA polymerase III transcription machineries. *BBA-Gene Regul Mech* 1829: 258–264
- Vivat Hannah V, Atmanene C, Zeyer D, van Dorsselaer A, Sanglier-Cianféroni S (2010) Native MS: an “ESI” way to support structure- and fragment-based drug discovery. *Future Med Chem* 2: 35–50
- Walters BT, Ricciuti A, Mayne L, Englander SW (2012) Minimizing back exchange in the hydrogen exchange-mass spectrometry experiment. *J Am Soc Mass Spectrom* 23: 2132–2139
- Walzthoeni T, Joachimiak LA, Rosenberger G, Röst HL, Malmström L, Leitner A, Frydman J, Aebersold R (2015) xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat Meth* 12: 1185–1190
- Wang L, Lane LC, Smith DL (2001) Detecting structural changes in viral capsids by hydrogen exchange and mass spectrometry. *Protein Sci* 10: 1234–1243
- Wang W, Kitova EN, Klassen JS (2003) Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nano-electrospray Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 75: 4945–4955
- Wang G, de Jong RN, van den Bremer ETJ, Beurskens FJ, Labrijn AF, Ugurlar D, Gros P, Schuurman J, Parren PW, Heck AJR (2016) Molecular basis of assembly and activation of complement component C1 in complex with immunoglobulin G1 and antigen. *Mol Cell* 63: 135–145
- Washburn MP, Wolters D, Yates JR III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19: 242–247
- van de Waterbeemd M, Lössl P, Gautier V, Marino F, Yamashita M, Conti E, Scholten A, Heck AJR (2014) Simultaneous assessment of kinetic, site-specific, and structural aspects of enzymatic protein phosphorylation. *Angew Chem Int Ed Engl* 53: 9660–9664
- Weisbrod CR, Chavez JD, Eng JK, Yang L, Zheng C, Bruce JE (2013) *In vivo* protein interaction network identified with a novel real-time cross-linked peptide identification strategy. *J Proteome Res* 12: 1569–1579
- Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJJ, van der Oost J, Doudna JA, Nogales E (2011a) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477: 486–489
- Wiedenheft B, van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A, Westphal W, Heck AJR, Boekema EJ, Dickman MJ, Doudna JA (2011b) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci USA* 108: 10092–10097
- Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese J, Bantscheff M, Gerstmaier A et al (2014) Mass-spectrometry-based draft of the human proteome. *Nature* 509: 582–587

- Woods LA, Radford SE, Ashcroft AE (2013) Advances in ion mobility spectrometry-mass spectrometry reveal key insights into amyloid assembly. *Biochim Biophys Acta* 1834: 1257–1268
- Wright AV, Nuñez JK, Doudna JA (2016) Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* 164: 29–44
- Wu C, Herzog F, Jennebach S, Lin Y, Pai C, Aebersold R, Cramer P, Chen H (2012) RNA polymerase III subunit architecture and implications for open promoter complex formation. *Proc Natl Acad Sci USA* 109: 19232–19237
- Yan Y, Chen G, Wei H, Huang RY, Mo J, Rempel DL, Tymiak AA, Gross ML (2014) Fast photochemical oxidation of proteins (FPOP) maps the epitope of EGFR binding to adnectin. *J Am Soc Mass Spectrom* 25: 2084–2092
- Yang B, Wu Y, Zhu M, Fan S, Lin J, Zhang K, Li S, Chi H, Li Y, Chen H, Luo S, Ding Y, Wang L, Hao Z, Xiu L, Chen S, Ye K, He S, Dong M (2012) Identification of cross-linked peptides from complex samples. *Nat Meth* 9: 904–906
- Yang Y, Barendregt A, Kamerling JP, Heck AJR (2013) Analyzing protein micro-heterogeneity in chicken ovalbumin by high-resolution native mass spectrometry exposes qualitatively and semi-quantitatively 59 proteoforms. *Anal Chem* 85: 12037–12045
- Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger G (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci USA* 97: 5802–5806
- Zehl M, Rand KD, Jensen ON, Jørgensen TJD (2008) Electron transfer dissociation facilitates the measurement of deuterium incorporation into selectively labeled peptides with single residue resolution. *J Am Chem Soc* 130: 17453–17459
- Zeng-Elmore X, Gao X, Pellarin R, Schneidman-Duhovny D, Zhang X, Kozacka KA, Tang Y, Sali A, Chalkley RJ, Cote RH, Chu F (2014) Molecular architecture of photoreceptor phosphodiesterase elucidated by chemical cross-linking and integrative modeling. *J Mol Biol* 426: 3713–3728
- Zhang Y, Fonslow BR, Shan B, Baek M, Yates JR III (2013) Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* 113: 2343–2394
- Zheng C, Yang L, Hoopmann MR, Eng JK, Tang X, Weisbrod CR, Bruce JE (2011) Cross-linking measurements of *in vivo* protein complex topologies. *Mol Cell Proteomics* 10: M110.006841
- Zheng Y, Fornelli L, Compton PD, Sharma S, Canterbury J, Müller CW, Zabrouskov V, Fellers RT, Thomas PM, Licht JD, Senko MW, Kelleher NL (2016) Unabridged analysis of human histone H3 by differential top-down mass spectrometry reveals hypermethylated proteoforms from MMSET/NSD2 overexpression. *Mol Cell Proteomics* 15: 776–790
- Zhou M, Wysocki VH (2014) Surface induced dissociation: dissecting noncovalent protein complexes in the gas phase. *Acc Chem Res* 47: 1010–1018
- Zubarev RA (2004) Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotech* 15: 12–16
- Zubarev RA, Makarov A (2013) Orbitrap mass spectrometry. *Anal Chem* 85: 5288–5296



**License:** This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.