Software for the analysis and interpretation of native mass spectrometry data

Timothy M. Allison[1], Perdita Barran[2], Justin L.P. Benesch [3], Sarah Cianferani[4], Matteo T. Degiacomi[3, 5], Valerie Gabelica[6], Rita Grandori[7], Erik G. Marklund[8], Thomas Menneteau[9], Lukasz G. Migas[2], Argyris Politis[10], Michal Sharon[11], Frank Sobott[12, 13, 14], Konstantinos Thalassinos[9, 15] *

[1] *School of Physical and Chemical Sciences, Biomolecular Interaction Centre, University of Canterbury, Christchurch, New Zealand*

[2] *Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute of Biotechnology, School of Chemistry, University of Manchester, Manchester, U.K.*

[3] *Department of Chemistry, Chemistry Research Laboratory, University of Oxford, South Parks Road, OX1 3TA, U.K.*

[4] *Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), Université de Strasbourg, CNRS, IPHC UMR 7178, 67000 Strasbourg, France.*

[5] *Department of Physics, Durham University, South Road, Durham, DH1 3LE, U.K.*

[6] *University of Bordeaux, INSERM and CNRS, ARNA Laboratory, IECB site, 2 Rue Robert Escarpit, 33600, Pessac, France.*

[7] *Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy*

[8] *Department of Chemistry - BMC, Uppsala University, Box 576, 75123, Uppsala, Sweden.*

[9] *Division of Biosciences, Institute of Structural and Molecular Biology, University College of London, Gower Street, London, WC1E 6BT, U.K.*

[10] *Department of Chemistry, King's College London, 7 Trinity Street, SE1 1DB, London, U.K.*

[11] *Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 7610001, Israel*

[12] *Biomolecular & Analytical Mass Spectrometry, Department of Chemistry, University of Antwerp, Antwerp, Belgium.*

[13] *School of Molecular and Cellular Biology, University of Leeds, Leeds, U.K.*

[14] *Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, U.K.*

[15] *Department of Biological Sciences, Institute of Structural and Molecular Biology, Birkbeck, Malet Street, London, WC1E 7HX, U.K.*

All authors contributed equally

*Address correspondence to k.thalassinos@ucl.ac.uk

Abstract

The last few years have seen a dramatic increase in applications of native mass and ion mobility spectrometry, especially for the study of proteins and protein complexes. This increase has been catalysed by the availability of commercial instrumentation capable of carrying out such analyses. Like in most fields, however, the software to process the data generated from new instrumentation lags behind. Recently, a number of research groups have started addressing this by developing software, but further improvements are still required in order to realise the full potential of the datasets generated. Here we describe practical aspects as well as challenges in processing native mass spectrometry (MS) and ion mobility-MS datasets, and provide a brief overview of currently available tools. We then set out our vision of future developments that would bring the community together and lead to the development of a common platform to expedite future computational developments, provide standardised processing approaches and serve as a location for the deposition of data for this emerging field.

Introduction

A thorough understanding of biological processes at the molecular level requires a detailed study of proteins, their complexes, and their interactions with other classes of molecules such as DNA, lipids, carbohydrates and other small molecules. A large number of diseases arise from proteins being unable to carry out their physiological functions [1]. Since the development of soft ionisation techniques, in particular electrospray ionisation [2], mass spectrometry (MS) has rapidly been adapted for the study of proteins, protein-ligand and protein-protein complexes. Further advances in sample preparation and instrumental developments – such as the use of modified quadrupoles, lower voltage potentials, source temperatures, higher vacuum pressures and volatile aqueous solvents such as ammonium acetate – resulted in the field now referred to as native MS [3, 4].

In native MS, proteins and noncovalent complexes are prepared in an electrospray-compatible buffer, and are transferred to the gas phase in the form of ions, with instrument settings selected so that the perturbation of the solution structures is minimized [3, 5, 6]. An increasing number of studies confirm that, under carefully controlled conditions, the native-like structure of proteins and their complexes can be preserved [7-9]. Native MS experiments provide information regarding the conformation, composition, stoichiometry and interactions of proteins. The analyses are very fast and sensitive, and they make it possible to simultaneously characterise several species present in complex samples, even from mixtures isolated directly from complex biological matrices [10-12].

Incorporation of ion mobility (IM) separation into native MS experiments further enhances the information obtained on proteins and protein complexes. IM separates ions based on their interaction with a buffer gas as they travel through a collision cell under the influence of a weak electric field. The time an ion takes to traverse the IM cell is related to the charge and rotationally averaged collision cross-section (CCS or □) of the ion, the latter being a physical quantity related to the overall shape of the ion [13, 14], and, to a lesser degree, the ion mass. The IM-MS combination allows one to separate coexisting forms of a given protein/protein complex, even within the same charge state, that would otherwise be indistinguishable using either MS or IM alone [15].

While the field of native IM-MS is still relatively new, the last few years have seen a rapid expansion in the number of studies using this methodology, partially fuelled by the introduction of commercial instrumentation (**Figure 1)**.

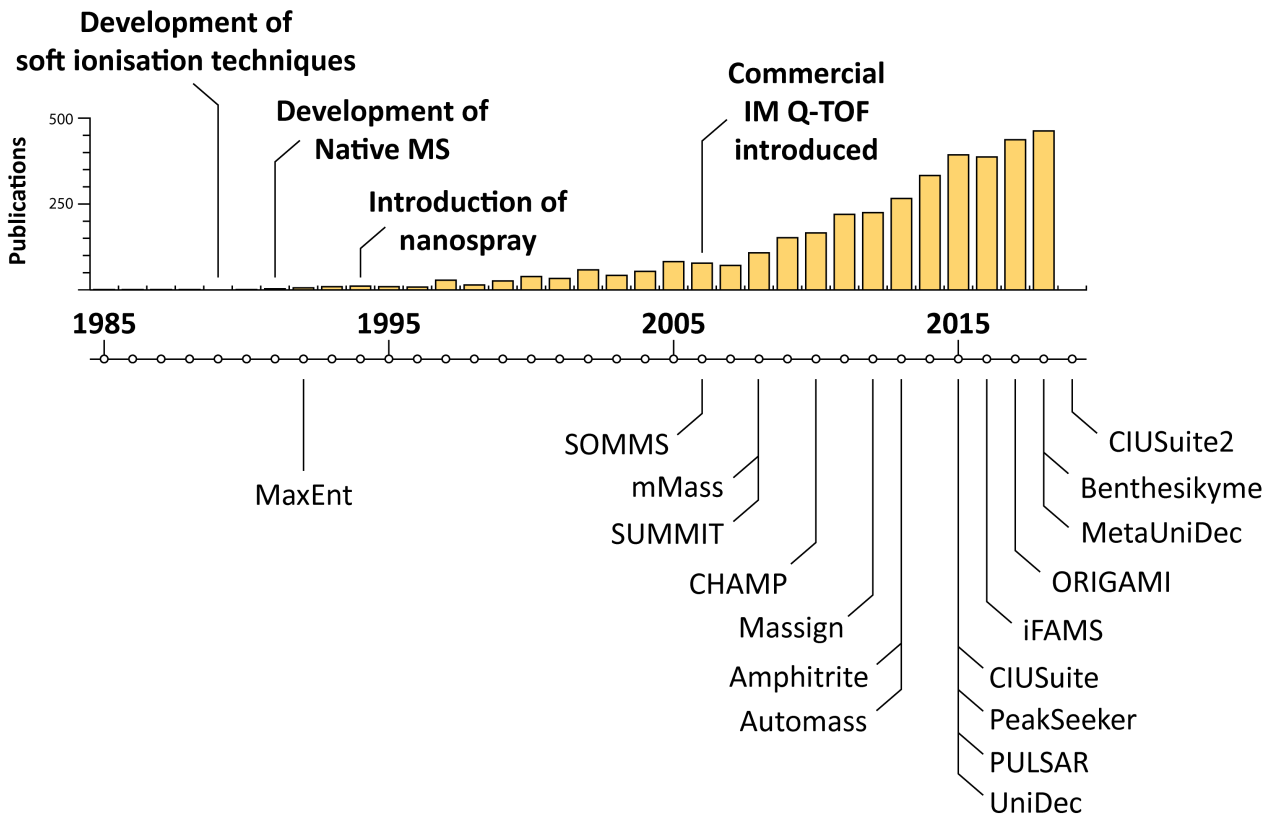


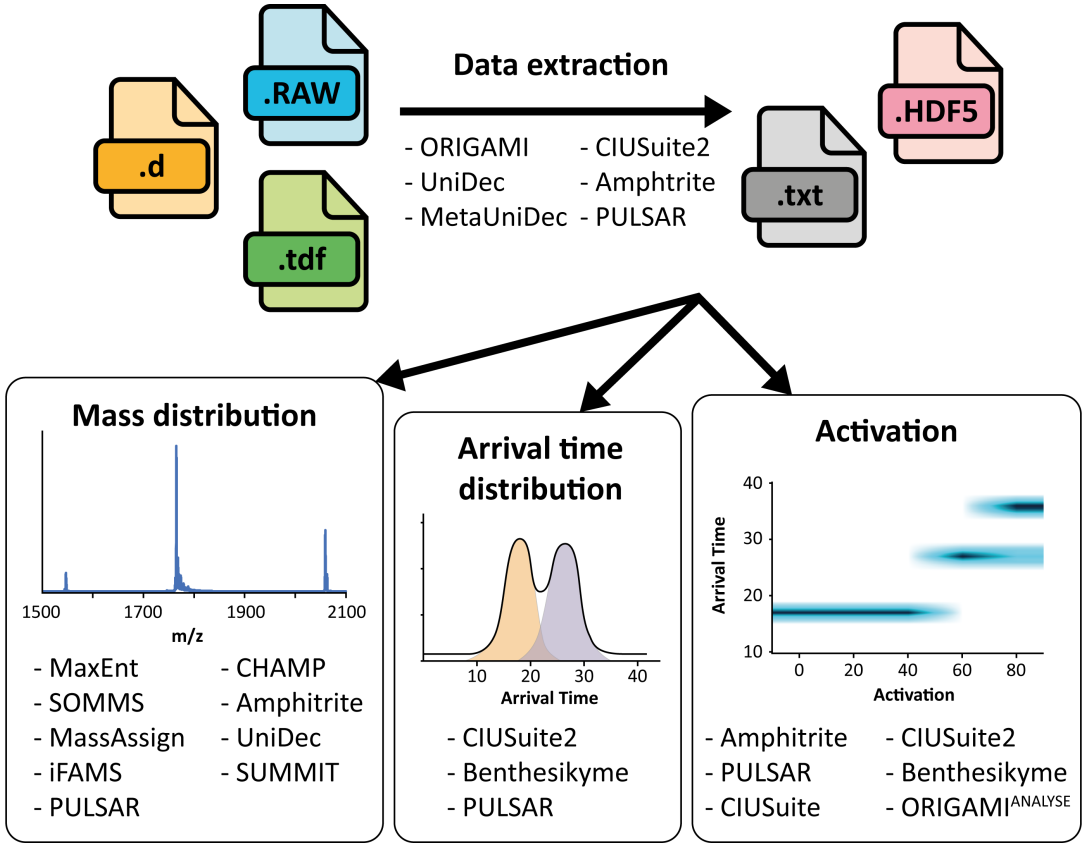
Figure 1*: Timeline of major developments in native MS (top) and associated supporting software (bottom). Publication metrics were generated from Web of Science in May 2019 by searching for topic: ("ion mobility" AND "mass spectrometry"). Software tools and their features are listed in Table 1.*

Even though the software offered by vendors are particularly useful for thorough processing of MS data, they provide only basic functionality and still require further manual processing routines, which are often slow, cumbersome and error-prone. MS software developed for the areas of proteomics, lipidomics, glycomics and metabolomics is more mature in nature, however such tools are not optimised for the types of analyses required for native IM-MS datasets. Specialised software providing automated processing approaches, apart from yielding considerable time savings, would increase the reproducibility of processing and analysis. It would also facilitate the exchange of results between laboratories and provide robust documentation of how experiments were carried out and how the data were processed.

Although in recent years significant progress has been made in the development of specialised native MS and IM-MS software [16-25] (**Figure 1 – Table 1**), further improvements are required to make full use of the data and cement the role of native MS and IM-MS as standard tools in structural biology investigations. Here we describe the challenges in processing native MS and IM-MS data, and provide an overview of existing software and their features. We then highlight current shortcomings and outline possible solutions along with our vision for the future of this field. We intend

to encourage the community to nucleate around a set of shared goals in order to capitalise on our collective expertise and enable more efficient innovation in this growing and very important area.

---------- Box 1 ----------



Box Figure 1: *From raw data to usable formats. Vendor data is converted to text or other file formats for downstream analyses. The types of analyses performed and the corresponding software offering such functionality are shown.*

MS-based methods yield raw spectral data stored in instrument manufacturers' proprietary formats. The supporting software programs supplied by these manufacturers can usually perform simple actions such as basic viewing and signal processing. Integration of vendor-processed spectral data into custom workflows and researcher software requires manual data export, making this route cumbersome and non-viable for automated and high-throughput analyses.

Software for bespoke analysis typically incorporates methods to read the raw data stored in these proprietary file formats (**Table 1**). Many of the early studies utilising native MS and IM-MS were conducted on Waters instruments, as they were the first to introduce a commercial IM instrument to the market[26]. For this reason, early custom processing software such as *Amphitrite*, *PULSAR*, and *CIUSuite2* have custom-built file readers that accept the raw files as input and can read them by calling the Waters dynamic linked library (cdt.dll). *CIUSuite2* more recently has also added functionality for reading Agilent raw files. Alternatively, programs such as *ORIGAMI*^{ANALYSE} utilise data extraction tools available within the Waters ion mobility software Driftscope. Due to the development of high-mass Orbitraps (made by Thermo Scientific) and the utility of such instruments for native MS applications, software such as *UniDec*, *MetaUniDec* and *ORIGAMI*^{ANALYSE} also accept Thermo Raw files. Alternatively, other software such as *Benthesikyme* and *iFAMS* use as input text files that have been exported by using the manufacturer's software.

While there are many open data formats available in the MS community, these have not been adopted by the native MS field. This is because the design of current open data formats has been chosen for more common applications of MS in various proteomic, lipidomic, and metabolomic fields. Depending on the type of analysis, be it MS or MS/MS,

eXtensible Markup Language (XML)-based formats such as *mzData, mzML, mzXML* are commonly used and interchanged. However, these tend to better cater to MS and not IM-MS metadata. In instances where IM is supported, the syntax tends to be too verbose in the full profile mode. Another open data format alternative is Unified Ion Mobility Files (UIMF) [27], which is based on a relational database management system. However, UIMF was mainly developed for IM proteomics applications and, thus, does not allow storing experimental data types such as those from activation experiments. So, overall none of the formats described above is ideal for storing data from native IM-MS experiments, which is why none of them has been broadly adopted as a standard in the community.

Recently, the open data format Hierarchical Data Format (HDF5) has been used by *MetaUniDec.* The ability to group together multiple types of data objects into a single file, offered by HDF5, appears to offer the right flexibility to store the types of native MS and IM-MS experiments described here. The HDF5 format also allows one to store many of the experimental parameters, crucial for reproducibility and transparency, that the community has recently put forward [14].

---------- Box 1 ----------

The dimensionality of native IM-MS data

To understand the challenges in processing native IM-MS data, we first describe the different dimensions of the data. By this, we refer to dimensions by which the ions are separated and we include ion intensity as an implicit dimension. We then describe what processing is required, available software for such processing, and the biological information gained. While native MS analyses can be interfaced with prior chromatographic separation [28, 29], this is not routine yet and so we do not include chromatographic retention time as a dimension.

1st dimension – mass to charge ratio (*m/z*)

The first dimension of separation is the mass-to-charge (*m/z*) ratio. Such information is obtained from native MS experiments of a protein sample even when heterogeneous or polydisperse. As the primary ionisation method for such analyses is electrospray, and in particular nano-electrospray, what is observed for each protein is a distribution of charges, referred to as a charge state distribution (CSD) [30]. Native mass spectra can therefore be complex, especially when multiple proteins are present in the sample, each resulting in their own CSD which can often overlap. Further complications arise as charge states are also typically not isotopically resolved, which impairs charge state assignment for each peak. CSDs are linked not only to the mass but also to the structural compactness of the detected ions [31-33].

2nd dimension – ion mobility separation

Incorporation of IM adds an extra dimension of separation, as each *m/z* species is separated according to its transit time through an ion mobility cell. What is recorded by the instrument is the arrival time, which is the time an ion spends in the mobility cell plus the time to travel from the exit of this cell to the instrument detector. The transit time in the mobility cell is related to the mobility of the ion, itself linked to its CCS.

3rd dimension – ion reactivity

A third dimension can be added by perturbing the structure of native-like ions by increasing their internal energy. The most frequently used activation methods include collision-induced dissociation (CID) [34, 35], surface-induced dissociation (SID) [36, 37] or ultraviolet photodissociation (UVPD) [38-40]. Other methods include electron capture dissociation (ECD) and electron transfer dissociation (ETD) [41]. In these cases, IM can monitor structural rearrangement by recording the change in the observed arrival time distribution (ATD) as a function of ion internal energy, or input energy.

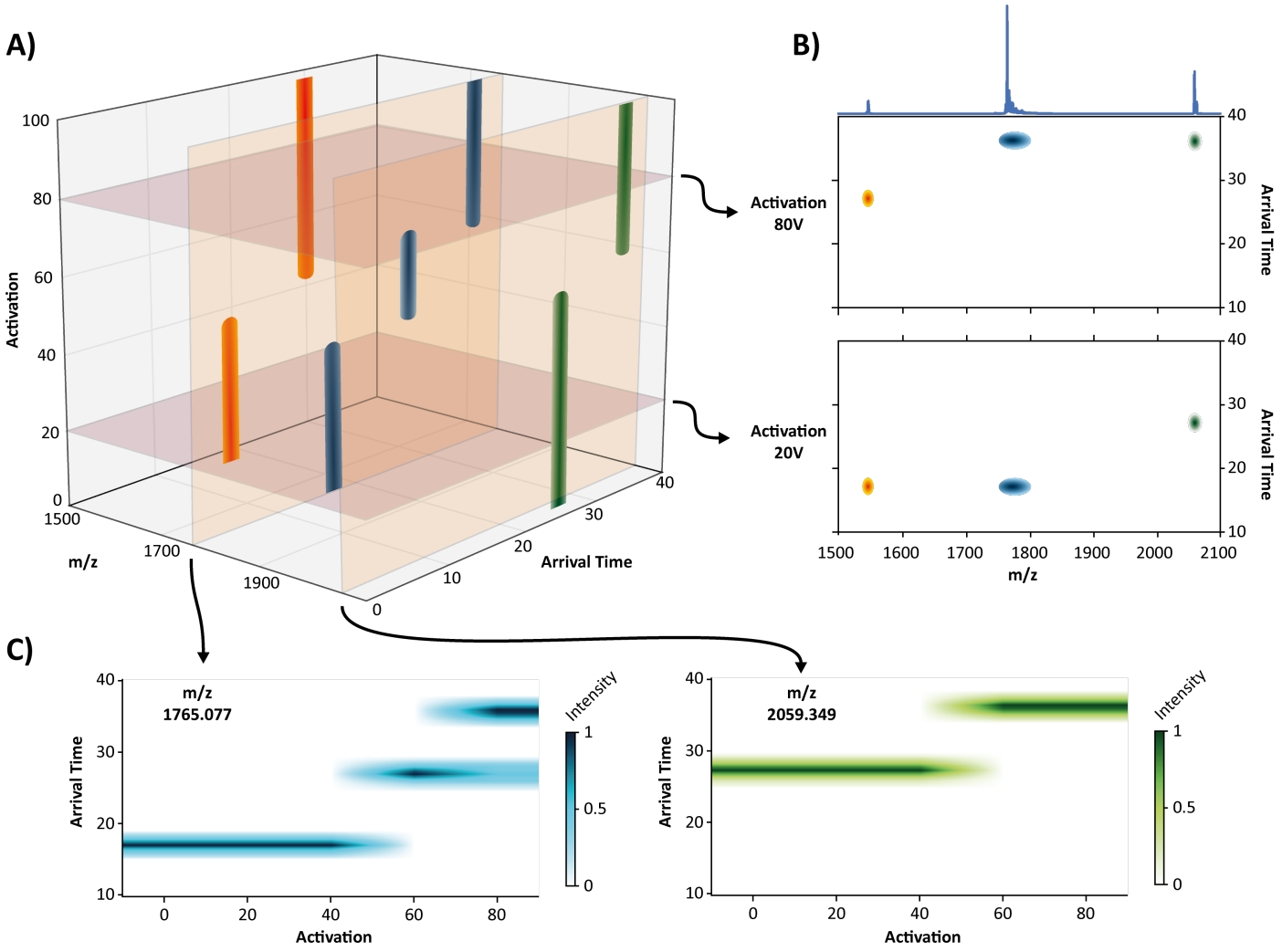


Figure 2: *A) Figure illustrating the different dimensions present in a native IM-MS experiment namely, m/z, ion mobility separation and ion reactivity. 2D plots of binary representations of this 3D dataset are shown in B) and C). B) shows the arrival time corresponding to each species in the native MS spectrum, at different activation voltages (20 and 80V). C) Evolution of arrival time distributions for selected m/z species due to activation conditions.*

Processing approaches

Like with mass spectra from other types of analyses, some standard processing of native mass spectra, such as smoothing and background subtraction, needs to take place. This is done after the raw data, almost always recorded in proprietary

data formats, are extracted into a more usable format (see **Box 1**). Once the data have been extracted and made available in an accessible format, the following processing steps can take place.

Peak assignment – from *m/z* to mass distribution

A simple way to calculate the mass from a CSD is to select the peak tops of what are presumed adjacent charge states of a given species and solve two equations to find mass and charge (see **Appendix**). This process, however, becomes difficult when multiple overlapping peak series are present in the spectrum. Initially, semi-automated deconvolution algorithms were created, with the introduction of *MaxEnt* and later with more sophisticated methods implemented in *SOMMS*, *Massign,* and *CHAMP*. More recently, programs such as *Amphitrite*, and *UniDec* have improved on previous methodologies. They simplify and speed up the deconvolution process by automatically picking, fitting and assigning charge states to protein and protein-complex samples. The peak series, containing each charge state and typically lacking isotopic resolution, are usually described by Gaussian distributions. However, other distributions, such as Lorentzian or mixed distributions, have also been used as the experimental peaks can have a pronounced tail due to adducts attachment. Most software simulate a spectrum and then use least squares optimisation to minimise the differences between the simulated and experimental data. *UniDec* instead uses a Bayesian deconvolution algorithm, while *iFAMS* uses a Fourier Transform-based algorithm for this optimisation.

Such analyses can provide the mass of a protein or protein complex, and are the gold standard in reporting on the oligomeric or ligand-binding state of a protein. Knowing the mass of a protein complex can also provide insights into its possible composition, especially if the masses of the individual components comprising the complex are known. Software such as *SUMMIT* [42] has been developed to do this. Measurements of mass changes due to small molecule or ligand binding can also be used to determine the stoichiometry of such binding events.

As well as measuring composition and stoichiometry, native MS can also be used to determine apparent binding affinities [43]. For example, one can determine the affinity of a ligand binding to a protein if one assumes that intensities reflect concentrations in solution of the bound and unbound species. These affinities are calculated by fitting the relative ion intensities of bound and unbound forms as a function of the ligand concentration. Accurate extraction of the intensities, or peak areas, of different bound forms is therefore required [44, 45]. Deconvolution of the data allows one to use the extracted peak areas for this purpose, which can provide more accurate data for fitting than the peak tops.

Software such as *UniDec* provide multiple models to fit data and enable the calculation of both micro- and macroscopic K_d values.

Extraction of arrival time distributions

Extraction of the arrival time distribution (ATD) for a particular ion (i.e. the ATD associated with a specified *m/z* range), is facilitated by the deconvolution of the mass spectrum (described above). From the deconvoluted spectrum, the *m/z* range corresponding to each peak is known, which can then be used to extract the corresponding range in the arrival time space (i.e. the ATD), and the charge *z* will be used to convert arrival times to CCSs. Once an ATD is extracted, its apex can be obtained, which is usually the main descriptor of an ATD. Further peak fitting, typically using one or multiple Gaussians, can also take place. This is useful as the width of an ATD reports on the conformational heterogeneity of an ion. In other cases, deconvolution of the ATDs can reveal the relative distribution of overlapping conformational families [17]. *CIUSuite2* performs this deconvolution in a semi-automated manner while programs like *Benthesikyme* do this using a genetic algorithm.

Monitoring changes in ATD features, e.g. the widths or relative populations of deconvoluted conformational families, can reveal insights into the effects of mutations or ligand binding events on protein structure [17]. Most importantly, combined information about folding and binding can be retrieved for each resolved component of the ATD profile. There are cases, however, where the ATD alone does not reveal significant changes. In these instances, activating particular charge states and monitoring changes in the ATDs as a function of activation energy provides further information [35].

Activation

Activation experiments usually involve stepwise increases in collisional energies. This elicits conformational changes in the protein ions, resulting in changes in the recorded arrival times in the IM domain [46]. For protein-ligand complexes, the ligand can dissociate either prior to or, more commonly, after protein unfolding. Such experiments can report on the strength of this protein-ligand interaction, and how far ligand binding depends on the preservation of local structure of a binding pocket in the protein. These changes are typically visualised by the generation of 'unfolding plots', which stack the extracted ATDs for a particular ion as a function of the applied activation energy. Several programs have been developed to generate these unfolding plots (*Amphitrite, PULSAR, CIUSuite, Benthesikyme, ORIGAMI*^{ANALYSE}). The

ATD for each voltage is usually normalised to the highest intensity and smoothed prior to being used to generate the two-dimensional plots.

CIUSuite and *CIUSuite2* have functions to identify features of unfolding trajectories that describe or 'fingerprint' a given protein's characteristics. Many software (*CIUSuite, CIUSuite2, PULSAR, ORIGAMI*^{ANALYSE}) can also generate difference and root-mean-square deviation plots to allow unfolding profiles to be compared. *Benthesikyme* provides summary plots that capture the changes in average arrival time and in the full width at half maximum of the ATD during activation experiments and uses these to compare samples.

The voltages at which transitions occur between discrete species in unfolding plots can be affected by the bound state of an ion. For example, ligand binding can induce conformational stability in the various observed forms, analogous to in-solution thermal shift unfolding/aggregation experiments. *PULSAR* implements an equilibrium unfolding model, similar to those used for solution studies. Fitting of unfolding plots to this model can be used to determine the activation levels where transitions are observed between discrete species in an unfolding plot. This information can be used to identify the mid-transition voltage between different conformational and/or association states of a protein. *CIUSuite2* identifies features from replicate experiments and models the transition between these features using a sigmoid function to identify the mid-transition voltage. Unfolding curves can be used to report on the effects of mutations and ligand binding (e.g. small molecules, metal ions, peptides, lipids) on protein stability, and to compare the structural properties of highly similar proteins.

Converting arrival time distributions to collision cross sections

IM measures the ATDs of individual ions, and these values can be used to obtain CCSs, either via first principles or via calibration approaches [14]. This step requires first the extraction of the ATD for a particular ion, *i.e.* the ATD associated with a specified *m/z* range. Prior deconvolution of the mass spectrum as described above aids with this step. ATDs can be further processed to smooth the data and to fit peaks (e.g. by Gaussians), in order to locate the centroids of the arrival time peaks. For experiments involving calibration, the peak centroids for a set of calibrant ions are identified and used to construct a calibration curve [14]. Some programs have the CCS values of selected calibrant ions pre-loaded, while others allow one to use calibrant values of choice. *Amphitrite, PULSAR* or *ORIGAMI*^{ANALYSE} automate many of the tasks required for calibration. The goal of the calibration is to provide an equation relating the arrival time to the CCS. Possible

improvements in the future would include reporting on the uncertainties associated with the calibrated values, in line with recently published recommendations [14]. Once a calibration has been performed, the ATD axis can be converted to a CCS axis.

Harnessing IM-MS data for structural biology

Once an experimental CCS has been obtained, it can be compared to the CCSs for molecular structures obtained by other experimental methods, such as X-ray crystallography or NMR, or models provided by computational methods. The development of methodologies and related software to estimate the CCS of a molecular structure constitutes a theoretical and technical challenge on its own, and is an area of active research. An in-depth discussion of these methodologies, software implementing them and how to parameterise this software is beyond the scope of this review. The reader is referred to a recent review on this topic [13]. Briefly, CCS calculation software originally focused on calculating the CCS from atomistic models given prior information about the radius, and sometimes partial charge, of each atom. More recently, however, efforts to calculate the CCS electron density maps represented as point clouds [47, 48] or super-coarse-grained models [49-54] have been proposed. These methods are especially useful when no atomic information about the proteins of interest is available (**see companion review**).

Comparison of experimental and theoretical CCS values has been used extensively to report on the gas-phase structure of ions and whether the ions are able [7, 9], or not [55], to retain their solution-phase structure. Recent work has focused on including CCSs in integrative modelling pipelines [53, 56-59].

Many of the data from native MS and IM-MS experiments can be highly complementary to data obtained by other structural biology methods, and can be exploited in molecular modelling scenarios. CCSs can be used to validate structures produced by, for example, homology modelling, coarse-grained, or molecular dynamics simulations [60]. MS and IM-MS data can also be fed directly into integrative modelling software programs to guide the refinement of the most likely conformation(s) of quaternary assemblies [61]. Data typically exploited in this context include (1) detailed stoichiometric information about the subunits and complex partners and (2) CCS distributions. Stoichiometric information is used to define how many subunits should be manipulated during a protein docking run. Information about subunit connectivity within a complex, calculated with programs such as *CHAMP*, as well as CCS data, are also used as restraints in modelling protocols [61-63].

Perspective and future directions

The recent growth in software for native MS has provided various tools, which have independently implemented many similar functionalities. There is now an opportunity for us as a community to consolidate best practice in the approaches and methods used for the preparation and processing of data. This is a necessary step to ensure the integrity and reproducibility of native IM-MS results across different instrument platforms and laboratories, as well as to define future data reporting standards. This will require establishing a set of benchmark data files that can be used to evaluate the capabilities of the available software and how they compare with respect to the biophysical quantities they report on.

First, a common data format will need to be agreed upon (see Box 1). With the advent of commercial IM-MS instruments, the need to move towards an open MS/IM-MS format is apparent. The interface between vendor proprietary software and post-processing software developed in the research community is not straightforward in all cases. Manufacturers need to concentrate support on the most common applications of their technology, but this can have the unfortunate consequence of impeding access to the raw data needed by those in the research community who develop their own specialised analytical tools. A common file format, into which vendor formatted data can be easily extracted, should be accompanied by common tools for reading and writing in that format, analogous to what the *msconvert* tool provides for the proteomics community [64].

It is also important that the file format records meta-data (i.e. experimental parameters, including sample preparation, and relevant instrumental parameters). These are important for the correct reproduction of experiments, and also for adherence to policies regarding the storage and management of research data [65]. A proposal outlining relevant information that needs to be captured in IM has recently been put forward [14].

A common file format would enable the creation of a data repository and encourage the deposition and meta-analysis of data. This would have significant benefits for computational researchers as well as experimentalists, and further support data sharing and transparency in data collection. Establishing a repository for native IM-MS data would provide opportunities to link these data with entries on other structural repositories such as the PDB [66], EMDataBank [67], PDB-Dev [68, 69], and pE-DB [70], as well as on proteomics databases such as PRIDE [71] and Uniprot [72]. We would also welcome the inclusion of native IM-MS outputs such as mass, ATD and CCS, under well-defined experimental conditions, among the parameters describing protein structures in these databases.

While these initiatives would consolidate capabilities that current software provide, developments in MS hardware continue to occur at a rapid pace. Examples are the development of higher resolution in MS and IM dimensions, tandem ion mobility analysers, and multistage activation and separation experiments [73, 74]. To accommodate these and any future innovations, a common programming library is needed that accomplishes common processing tasks such as reading raw files, smoothing, background subtraction, Gaussian envelope fitting, visualisation and data export. This would reduce duplications in software development, speed up development time and enable more focus on innovation in data analysis. This library would preferably be written in the Python programming language, due to its flexibility, accessibility, and existing codebase. It should be made available on code-hosting platforms such as GitHub or Bitbucket to allow for more sustainable software, better code testing, and greater collaboration between research groups. Using common tools for analysis should also improve the reproducibility of results. Having such a common codebase in a publicly accessible repository would also allow the community to further develop and extend the software together, and would allow any developments to be made available immediately to the community.

Once the software library and common file format are in place, a further development would be to create a software tool for accessing/visualising these files, similar to PRIDE Inspector, which is used to access proteomics datasets [75, 76]. Recent improvements in web technologies mean it is now possible to create a platform-agnostic, web-based front end for the visualisation of native IM-MS datasets. This front-end would be web-accessible via a browser and mobile devices, and could provide a barebones environment implementing common processing routines. A plug-in interface could allow the integration of algorithms for data processing. This front end could integrate access to CCS calculation software and other software that is currently needed for interpreting native MS and IM-MS data, such as new gas-phase molecular dynamics approaches. The latter is an underdeveloped area, but very important for fully exploiting the data; we anticipate a lot of algorithm and software innovations to take place in this field in the coming years. Currently, tools for calculating the theoretical CCS of molecular structures are used as standalone programs. Integration of these calculators with software for the processing of raw data would allow a more straightforward comparison between experimental observations and theoretical models, and would allow the parameters for both approaches to be stored in one common format, again assisting in transparency and data reproducibility.

We have set out ambitious aims with regards to data standardisation, the development of software tools and the creation of a common platform for the IM-MS field. We feel these steps are important for facilitating data sharing and collaboration, and for encouraging future innovation, transparency and reproducibility in this area.

Practically speaking, examples of how these aims could be achieved are the Collaborative Computational Projects (funded by the UK's Engineering and Physical Sciences Research Council) that exist in other structural biology fields, e.g. X-ray crystallography, NMR spectroscopy, electron microscopy. A similar initiative in IM-MS could also engage the broader structural mass spectrometry community – including those involved in cross-linking, covalent labelling or Hydrogen-Deuterium eXchange, as these techniques provide complementary information about protein structures and do not yet have a standardised format or dedicated data repository. A community-driven software framework will establish a solid platform for integrating native IM-MS data into the wider structural biology and biophysics communities.

Acknowledgements

We would like to acknowledge Michael Marty (University of Arizona), Sugyan Dixit and Daniel Polasky (University of Michigan) for helpful conversations, and Dr Catherine Lichten for her critical review of this manuscript. We thank the EU COST Action BM1403 and members of WG1 (Native IM-MS) and WG4 (Computational Methods).

References

1. Creighton, T. E., *Proteins: structures and molecular properties*. W. H. Freeman: New York, 1992.

2. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246* (4926), 64-71.

3. Heck, A. J., Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* **2008,** *5* (11), 927-33.

4. Sharon, M.; Robinson, C. V., The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem* **2007,** *76*, 167-93.

5. Hernandez, H.; Robinson, C. V., Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat Protoc* **2007,** *2* (3), 715-26.

6. Laganowsky, A.; Reading, E.; Hopper, J. T.; Robinson, C. V., Mass spectrometry of intact membrane protein complexes. *Nat Protoc* **2013,** *8* (4), 639-51.

7. Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H., Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun Mass Spectrom* **2008,** *22* (20), 3297-304.

8. Ruotolo, B. T.; Robinson, C. V., Aspects of native proteins are retained in vacuum. *Curr Opin Chem Biol* **2006,** *10* (5), 402-8.

9. Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V., Evidence for macromolecular protein rings in the absence of bulk water. *Science* **2005,** *310* (5754), 1658-61.

10. Konijnenberg, A.; Butterer, A.; Sobott, F., Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim Biophys Acta* **2013,** *1834* (6), 1239-56.

11. Chorev, D. S.; Baker, L. A.; Wu, D.; Beilsten-Edmands, V.; Rouse, S. L.; Zeev-Ben-Mordehai, T.; Jiko, C.; Samsudin, F.; Gerle, C.; Khalid, S.; Stewart, A. G.; Matthews, S. J.; Grünewald, K.; Robinson, C. V., Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry. *Science* **2018,** *362*, 829-834.

12. Ben-Nissan, G.; Chotiner, A.; Tarnavsky, M.; Sharon, M., Structural Characterization of Missense Mutations Using High Resolution Mass Spectrometry: A Case Study of the Parkinson's-Related Protein, DJ-1. *J Am Soc Mass Spectrom* **2016,** *27* (6), 1062-70.

13. Gabelica, V.; Marklund, E., Fundamentals of ion mobility spectrometry. *Curr Opin Chem Biol* **2018,** *42*, 51-59.

14. Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.; Campuzano, I. D. G.; Causon, T.; Clowers, B. H.; Creaser, C. S.; De Pauw, E.; Far, J.; Fernandez-Lima, F.; Fjeldsted, J. C.; Giles, K.; Groessl, M.; Hogan Jr, C. J.; Hann, S.; Kim, H. I.; Kurulugama, R. T.; May, J. C.; McLean, J. A.; Pagel, K.; Richardson, K.; Ridgeway, M. E.; Rosu, F.; Sobott, F.; Thalassinos, K.; Valentine, S. J.; Wyttenbach, T., Recommendations for reporting ion mobility Mass Spectrometry measurements. *Mass Spectrometry Reviews* **2019,** *38* (3), 291-320.

15. Beveridge, R.; Migas, L. G.; Das, R. K.; Pappu, R. V.; Kriwacki, R. W.; Barran, P. E., Ion Mobility Mass Spectrometry Uncovers the Impact of the Patterning of Oppositely Charged Residues on the Conformational Distributions of Intrinsically Disordered Proteins. *J Am Chem Soc* **2019,** *141* (12), 4908-4918.

16. Polasky, D. A.; Dixit, S. M.; Fantin, S. M.; Ruotolo, B. T., CIUSuite 2: Next-Generation Software for the Analysis of Gas-Phase Protein Unfolding Data. *Anal. Chem.* **2019,** *91* (4), 3147-3155.

17. Sivalingam, G. N.; Cryar, A.; Williams, M. A.; Gooptu, B.; Thalassinos, K., Deconvolution of ion mobility mass spectrometry arrival time distributions using a genetic algorithm approach: Application to α1-antitrypsin peptide binding. *Int. J. Mass spectrom.* **2018,** *426*, 29-37.

18. Migas, L. G.; France, A. P.; Bellina, B.; Barran, P. E., ORIGAMI: A software suite for activated ion mobility mass spectrometry (aIM-MS) applied to multimeric protein assemblies. *Int. J. Mass spectrom.* **2018,** *427*, 20-28.

19. Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V., Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **2015,** *87* (8), 4370-4376.

20. Eschweiler, J. D.; Rabuck-Gibbons, J. N.; Tian, Y.; Ruotolo, B. T., CIUSuite: A Quantitative Analysis Package for Collision Induced Unfolding Measurements of Gas-Phase Protein Ions. *Anal. Chem.* **2015,** *87* (22), 11516-11522.

21. Allison, T. M.; Reading, E.; Liko, I.; Baldwin, A. J.; Laganowsky, A.; Robinson, C. V., Quantifying the stabilizing effects of protein–ligand interactions in the gas phase. *Nat Comms* **2015,** *6*, 1-10.

22. Tseng, Y.-H.; Uetrecht, C.; Yang, S.-C.; Barendregt, A.; Heck, A. J. R.; Peng, W.-P., Game-Theory-Based Search Engine to Automate the Mass Assignment in Complex Native Electrospray Mass Spectra. *Anal. Chem.* **2013,** *85* (23), 11275-11283.

23. Sivalingam, G. N.; Yan, J.; Sahota, H.; Thalassinos, K., Amphitrite: A program for processing travelling wave ion mobility mass spectrometry data. *Int. J. Mass spectrom.* **2013,** *345*, 54-62.

24. Stengel, F.; Baldwin, A. J.; Bush, M. F.; Hilton, G. R.; Lioe, H.; Basha, E.; Jaya, N.; Vierling, E.; Benesch, J. L. P., Dissecting Heterogeneous Molecular Chaperone Complexes Using a Mass Spectrum Deconvolution Approach. *Chem. Biol.* **2012,** *19* (5), 599-607.

25. Morgner, N.; Robinson, C. V., Massign: An Assignment Strategy for Maximizing Information from the Mass Spectra of Heterogeneous Protein Assemblies. *Anal. Chem.* **2012,** *84* (6), 2939-2948.

26. Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H., An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass spectrom.* **2007,** *261* (1), 1-12.

27. Shah, A. R.; Davidson, J.; Monroe, M. E.; Mayampurath, A. M.; Danielson, W. F.; Shi, Y.; Robinson, A. C.; Clowers, B. H.; Belov, M. E.; Anderson, G. A.; Smith, R. D., An efficient data format for mass spectrometry-based proteomics. *J. Am. Soc. Mass Spectrom.* **2010,** *21* (10), 1784-1788.

28. Ehkirch, A.; D'Atri, V.; Rouviere, F.; Hernandez-Alba, O.; Goyon, A.; Colas, O.; Sarrut, M.; Beck, A.; Guillarme, D.; Heinisch, S.; Cianferani, S., An Online Four-Dimensional HICxSEC-IMxMS Methodology for Proof-of-Concept Characterization of Antibody Drug Conjugates. *Anal Chem* **2018,** *90* (3), 1578-1586.

29. Ehkirch, A.; Goyon, A.; Hernandez-Alba, O.; Rouviere, F.; D'Atri, V.; Dreyfus, C.; Haeuw, J. F.; Diemer, H.; Beck, A.; Heinisch, S.; Guillarme, D.; Cianferani, S., A Novel Online Four-Dimensional SECxSEC-IMxMS Methodology for Characterization of Monoclonal Antibody Size Variants. *Anal Chem* **2018,** *90* (23), 13929-13937.

30. Kebarle, P.; Verkerk, U. H., Electrospray: From ions in solution to ions in the gas phase, what we know now. *Mass Spectrometry Reviews* **2009,** *28*, 898-917.

31. Frimpong, A. K.; Abzalimov, R. R.; Uversky, V. N.; Kaltashov, I. A., Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: conformational heterogeneity of alpha-synuclein. *Proteins* **2010,** *78* (3), 714-22.

32. Santambrogio, C.; Natalello, A.; Brocca, S.; Ponzini, E.; Grandori, R., Conformational Characterization and Classification of Intrinsically Disordered Proteins by Native Mass Spectrometry and Charge-State Distribution Analysis. *PROTEOMICS* **2019,** *19*, 1800060.

33. Simmons, D. A.; Konermann, L., Characterization of transient protein folding intermediates during myoglobin reconstitution by time-resolved electrospray mass spectrometry with on-line isotopic pulse labeling. *Biochemistry* **2002,** *41* (6), 1906-14.

34. Nyon, M. P.; Prentice, T.; Day, J.; Kirkpatrick, J.; Sivalingam, G. N.; Levy, G.; Haq, I.; Irving, J. A.; Lomas, D. A.; Christodoulou, J.; Gooptu, B.; Thalassinos, K., An integrative approach combining ion mobility mass spectrometry, X-ray crystallography, and nuclear magnetic resonance spectroscopy to study the conformational dynamics of alpha1 -antitrypsin upon ligand binding. *Protein Sci* **2015,** *24* (8), 1301-12.

35. Rabuck, J. N.; Hyung, S. J.; Ko, K. S.; Fox, C. C.; Soellner, M. B.; Ruotolo, B. T., Activation state-selective kinase inhibitor assay based on ion mobility-mass spectrometry. *Anal Chem* **2013,** *85* (15), 6995-7002.

36. Ma, X.; Lai, L. B.; Lai, S. M.; Tanimoto, A.; Foster, M. P.; Wysocki, V. H.; Gopalan, V., Uncovering the stoichiometry of Pyrococcus furiosus RNase P, a multi-subunit catalytic ribonucleoprotein complex, by surface-induced dissociation and ion mobility mass spectrometry. *Angew Chem Int Ed Engl* **2014,** *53* (43), 11483-7.

37. Quintyn, R. S.; Zhou, M.; Yan, J.; Wysocki, V. H., Surface-Induced Dissociation Mass Spectra as a Tool for Distinguishing Different Structural Forms of Gas-Phase Multimeric Protein Complexes. *Anal Chem* **2015,** *87* (23), 11879-86.

38. Theisen, A.; Yan, B.; Brown, J. M.; Morris, M.; Bellina, B.; Barran, P. E., Use of Ultraviolet Photodissociation Coupled with Ion Mobility Mass Spectrometry To Determine Structure and Sequence from Drift Time Selected Peptides and Proteins. *Anal Chem* **2016,** *88* (20), 9964-9971.

39. Morrison, K. A.; Clowers, B. H., Differential Fragmentation of Mobility-Selected Glycans via Ultraviolet Photodissociation and Ion Mobility-Mass Spectrometry. *J Am Soc Mass Spectrom* **2017,** *28* (6), 1236-1241.

40. Theisen, A.; Black, R.; Corinti, D.; Brown, J. M.; Bellina, B.; Barran, P. E., Initial Protein Unfolding Events in Ubiquitin, Cytochrome c and Myoglobin Are Revealed with the Use of 213 nm UVPD Coupled to IM-MS. *J Am Soc Mass Spectrom* **2019,** *30* (1), 24-33.

41. Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F., Radical solutions: Principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrometry Reviews* **2018,** *37*, 750-771.

42. Taverner, T.; Hernández, H.; Sharon, M.; Ruotolo, B. T.; Matak-Vinković, D.; Devos, D.; Russell, R. B.; Robinson, C. V., Subunit Architecture of Intact Protein Complexes from Mass Spectrometry and Homology Modeling †. *Accounts of Chemical Research* **2008,** *41*, 617-627.

43. Hopper, J. T. S.; Robinson, C. V., Mass Spectrometry Quantifies Protein Interactions-From Molecular Chaperones to Membrane Porins. *Angewandte Chemie International Edition* **2014,** *53*, 14002-14015.

44. Dyachenko, A.; Gruber, R.; Shimon, L.; Horovitz, A.; Sharon, M., Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proceedings of the National Academy of Sciences* **2013,** *110*, 7235-7239.

45. Sokolovski, M.; Cveticanin, J.; Hayoun, D.; Korobko, I.; Sharon, M.; Horovitz, A., Measuring inter-protein pairwise interaction energies from a single native mass spectrum by double-mutant cycle analysis. *Nat Comms* **2017,** *8*, 212.

46. Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature Chemistry* **2014,** *6*, 281-294.

47. Degiacomi, M. T.; Benesch, J. L., EM intersectionIM: software for relating ion mobility mass spectrometry and electron microscopy data. *Analyst* **2016,** *141* (1), 70-5.

48. Alexeev, Y.; Fedorov, D. G.; Shvartsburg, A. A., Effective ion mobility calculations for macromolecules by scattering on electron clouds. *J Phys Chem A* **2014,** *118* (34), 6763-72.

49. Degiacomi, M. T., On the Effect of Sphere-Overlap on Super Coarse-Grained Models of Protein Assemblies. *J Am Soc Mass Spectrom* **2019,** *30* (1), 113-117.

50. Eschweiler, J. D.; Farrugia, M. A.; Dixit, S. M.; Hausinger, R. P.; Ruotolo, B. T., A Structural Model of the Urease Activation Complex Derived from Ion Mobility-Mass Spectrometry and Integrative Modeling. *Structure* **2018,** *26* (4), 599-606 e3.

51. Shi, Y.; Pellarin, R.; Fridy, P. C.; Fernandez-Martinez, J.; Thompson, M. K.; Li, Y.; Wang, Q. J.; Sali, A.; Rout, M. P.; Chait, B. T., A strategy for dissecting the architectures of native macromolecular assemblies. *Nat Methods* **2015,** *12* (12), 1135-8.

52. Politis, A.; Schmidt, C.; Tjioe, E.; Sandercock, A. M.; Lasker, K.; Gordiyenko, Y.; Russel, D.; Sali, A.; Robinson, C. V., Topological models of heteromeric protein assemblies from mass spectrometry: application to the yeast eIF3:eIF5 complex. *Chem Biol* **2015,** *22* (1), 117-28.

53. Hall, Z.; Politis, A.; Robinson, C. V., Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry. *Structure* **2012,** *20* (9), 1596-609.

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

55. Borysik, A. J.; Kovacs, D.; Guharoy, M.; Tornpa, P., Ensemble Methods Enable a New Definition for the Solution to Gas-Phase Transfer of Intrinsically Disordered Proteins. *J Am Chem Soc* **2015,** *137* (43), 13807-13817.

56. Politis, A.; Stengel, F.; Hall, Z.; Hernandez, H.; Leitner, A.; Walzthoeni, T.; Robinson, C. V.; Aebersold, R., A mass spectrometry-based hybrid method for structural modeling of protein complexes. *Nat Methods* **2014,** *11* (4), 403-406.

57. Eschweiler, J. D.; Farrugia, M. A.; Dixit, S. M.; Hausinger, R. P.; Ruotolo, B. T., A Structural Model of the Urease Activation Complex Derived from Ion Mobility-Mass Spectrometry and Integrative Modeling. *Structure* **2018,** *26*, 599-606.e3.

58. Wang, H.; Eschweiler, J.; Cui, W.; Zhang, H.; Frieden, C.; Ruotolo, B. T.; Gross, M. L., Native Mass Spectrometry, Ion Mobility, Electron-Capture Dissociation, and Modeling Provide Structural Information for Gas-Phase Apolipoprotein E Oligomers. *J. Am. Soc. Mass Spectrom.* **2019,** *30*, 876-885.

59. Santhanagopalan, I.; Degiacomi, M. T.; Shepherd, D. A.; Hochberg, G. K. A.; Benesch, J. L. P.; Vierling, E., It takes a dimer to tango: Oligomeric small heat shock proteins dissociate to capture substrate. *Journal of Biological Chemistry* **2018,** *293*, 19511-19521.

60. Hansen, K.; Lau, A. M.; Giles, K.; McDonnell, J. M.; Struwe, W. B.; Sutton, B. J.; Politis, A., A Mass-Spectrometry-Based Modelling Workflow for Accurate Prediction of IgG Antibody Conformations in the Gas Phase. *Angewandte Chemie International Edition* **2018,** *57*, 17194-17199.

61. Russel, D.; Lasker, K.; Webb, B.; Velazquez-Muriel, J.; Tjioe, E.; Schneidman-Duhovny, D.; Peterson, B.; Sali, A., Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol* **2012,** *10* (1), e1001244.

62. van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J., The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **2016,** *428* (4), 720-725.

63. Farabella, I.; Vasishtan, D.; Joseph, A. P.; Pandurangan, A. P.; Sahota, H.; Topf, M., TEMPy: a Python library for assessment of three-dimensional electron microscopy density fits. *J Appl Crystallogr* **2015,** *48* (Pt 4), 1314-1323.

64. Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.;

Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P., A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **2012,** *30* (10), 918-20.

65. Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; t Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B., The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016,** *3*, 160018.

66. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000,** *28* (1), 235-42.

67. Lawson, C. L.; Baker, M. L.; Best, C.; Bi, C.; Dougherty, M.; Feng, P.; van Ginkel, G.; Devkota, B.; Lagerstedt, I.; Ludtke, S. J.; Newman, R. H.; Oldfield, T. J.; Rees, I.; Sahni, G.; Sala, R.; Velankar, S.; Warren, J.; Westbrook, J. D.; Henrick, K.; Kleywegt, G. J.; Berman, H. M.; Chiu, W., EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res* **2011,** *39* (Database issue), D456-64.

68. Burley, S. K.; Kurisu, G.; Markley, J. L.; Nakamura, H.; Velankar, S.; Berman, H. M.; Sali, A.; Schwede, T.; Trewhella, J., PDB-Dev: a Prototype System for Depositing Integrative/Hybrid Structural Models. *Structure* **2017,** *25* (9), 1317-1318.

69. Vallat, B.; Webb, B.; Westbrook, J. D.; Sali, A.; Berman, H. M., Development of a Prototype System for Archiving Integrative/Hybrid Structure Models of Biological Macromolecules. *Structure* **2018,** *26* (6), 894-904 e2.

70. Varadi, M.; Kosol, S.; Lebrun, P.; Valentini, E.; Blackledge, M.; Dunker, A. K.; Felli, I. C.; Forman-Kay, J. D.; Kriwacki, R. W.; Pierattelli, R.; Sussman, J.; Svergun, D. I.; Uversky, V. N.; Vendruscolo, M.; Wishart, D.; Wright, P. E.; Tompa, P., pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Research* **2014,** *42*, D326-D335.

71. Martens, L.; Hermjakob, H.; Jones, P.; Adamski, M.; Taylor, C.; States, D.; Gevaert, K.; Vandekerckhove, J.; Apweiler, R., PRIDE: The proteomics identifications database. *PROTEOMICS* **2005,** *5*, 3537-3545.

72. Consortium, U., UniProt: a hub for protein information. *Nucleic acids research* **2015,** *43*, D204-12.

73. Eldrid, C.; Ujma, J.; Kalfas, S.; Tomczyk, N.; Giles, K.; Morris, M.; Thalassinos, K., Gas Phase Stability of Protein Ions in a Cyclic Ion Mobility Spectrometry Traveling Wave Device. *Anal Chem* **2019**.

74. Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M., A Cyclic Ion Mobility-Mass Spectrometry System. *Anal Chem* **2019**.

75. Perez-Riverol, Y.; Xu, Q. W.; Wang, R.; Uszkoreit, J.; Griss, J.; Sanchez, A.; Reisinger, F.; Csordas, A.; Ternent, T.; Del-Toro, N.; Dianes, J. A.; Eisenacher, M.; Hermjakob, H.; Vizcaino, J. A., PRIDE Inspector Toolsuite: Moving Toward a Universal Visualization Tool for Proteomics Data Standard Formats and Quality Assessment of ProteomeXchange Datasets. *Mol Cell Proteomics* **2016,** *15* (1), 305-17.

76. Wang, R.; Fabregat, A.; Rios, D.; Ovelleiro, D.; Foster, J. M.; Cote, R. G.; Griss, J.; Csordas, A.; Perez-Riverol, Y.; Reisinger, F.; Hermjakob, H.; Martens, L.; Vizcaino, J. A., PRIDE Inspector: a tool to visualize and validate MS proteomics data. *Nat Biotechnol* **2012,** *30* (2), 135-7.

77. Ferrige, A. G.; Seddon, M. J.; Jarvis, S.; Skilling, J.; Welch, J., The Application of Maxent to Electrospray Mass Spectrometry. In *Maximum Entropy and Bayesian Methods*, Springer Netherlands: Dordrecht, 1992; pp 327-335.

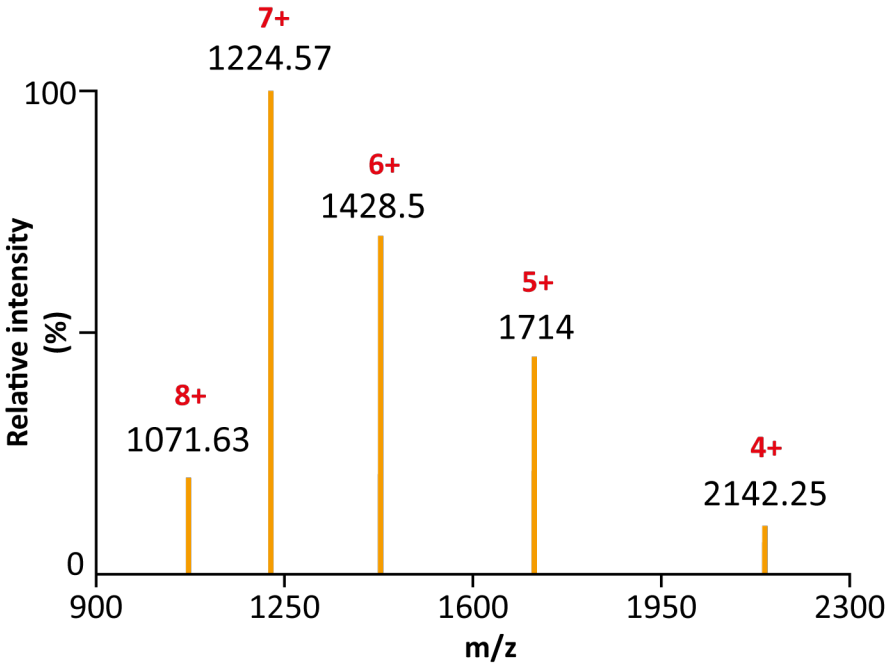
78. van Breukelen, B.; Barendregt, A.; Heck, A. J. R.; van den Heuvel, R. H. H., Resolving stoichiometries and oligomeric states of glutamate synthase protein complexes with curve fitting and simulation of electrospray mass spectra. *Rapid Communications in Mass Spectrometry* **2006,** *20*, 2490-2496.

79. Strohalm, M.; Hassman, M.; Košata, B.; Kodíček, M., mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Communications in Mass Spectrometry* **2008,** *22*, 905-908.

80. Lu, J.; Trnka, M. J.; Roh, S.-H.; Robinson, P. J. J.; Shiau, C.; Fujimori, D. G.; Chiu, W.; Burlingame, A. L.; Guan, S., Improved Peak Detection and Deconvolution of Native Electrospray Mass Spectra from Large Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2015,** *26*, 2141-2151.

81. Cleary, S. P.; Thompson, A. M.; Prell, J. S., Fourier Analysis Method for Analyzing Highly Congested Mass Spectra of Ion Populations with Repeated Subunits. *Anal. Chem.* **2016,** *88*, 6205-6213.

82. Reid, D. J.; Diesing, J. M.; Miller, M. A.; Perry, S. M.; Wales, J. A.; Montfort, W. R.; Marty, M. T., MetaUniDec: High-Throughput Deconvolution of Native Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2019,** *30*, 118-127.



$$\begin{cases} m/z = \dfrac{M + nH^+}{n} \\ m/z = \dfrac{M + (n+1)H^+}{n+1} \end{cases}$$

$$\begin{cases} 1224.57 = \dfrac{M + 1.008n}{n} \\ 1428.5 = \dfrac{M + 1.008(n+1)}{n+1} \end{cases}$$

$$\begin{cases} 1224.57n = M + 1.008n \\ 1428.5n + 1428.5 = M + 1.008n + 1.008 \end{cases}$$

$$\begin{cases} M = 1223.562n \\ 1427.492n = M + 1429.508 \end{cases}$$

$$1427.492n = 1223.562n + 1429.508$$

$$203.93n = 1429.508$$

$$n = 7 \qquad M = 8564.934$$

Appendix: *Example of simultaneous equations to solve the charge state and, from that, calculate the mass of a given species in the spectrum.*

Table 1: *List of the software available to process Native MS and/or IM-MS data.*

Name	Website	Interface	MS	IM-MS	Activated IM-MS	Accepted file formats	Language	Licence	Release year	Reference
MaxEnt	Contact Vendor	GUI	x			.raw (Waters)	N/A	Proprietary	1992	[77]
SOMMS	N/A	Command-line	x			.txt, Text	Perl	Free academic	2006	[78]
mMass	Link	GUI	x			Text, msD	Python/C	Free academic	2008	[79]
SUMMIT	N/A	N/A	x			N/A	N/A	N/A	2008	[42]
CHAMP	Request from authors	Command-line	x			.txt, Text	Python	Free	2010	[24]
Massign	Link	GUI	x			.txt, Text	LabView	Free academic	2012	[25]
Amphitrite	Link	GUI	x	x	x	.raw (Waters); .csv, Text	Python/C	Free GNU General Public Licence	2013	[23]
Automass	N/A	Command-line	x			N/A	N/A	N/A	2013	[22]
CIUSuite	Link	Command-line		x		.txt, Text	Python	Free academic	2015	[20]
PeakSeeker	Link	Command-line	x			.txt, Text	Python	Free	2015	[80]
PULSAR	Link	GUI		x		.raw (Waters)	Python/C	Free	2015	[21]
UniDec	Link	GUI; Command line	x			.raw (Waters); .txt, Text	Python/C	Free academic	2015	[19]
iFAMS	Link	Command line	x			.txt, Text	Python	Free academic	2016	[81]
ORIGAMI	Link	GUI	x	x	x	.raw (Waters); .raw (Thermo); .csv, Text	Python/JavaScript	Free academic	2017	[18]
Benthesikyme	Link	GUI; Command line		x	x	.txt, Text	JavaScript/Python/C /CUDA	Free	2018	[17]
MetaUniDec	Link	GUI; Command line	x		x	.txt, Text	Python/C	Free academic	2018	[82]
CIUSuite 2	Link	GUI; Command line	x	x	x	.raw (Waters); .txt, Text; .d (Agilent)	Python	Free academic	2019	[16]
Xtract & ReSpect	Contact Vendor	GUI	x			.raw (Thermo)	N/A	Proprietary		