HAL
open science

Computational strategies and challenges for using native ion mobility mass spectrometry in biophysics and structural biology

Timothy M. Allison, Perdita Barran, Sarah Cianférani, Matteo T. Degiacomi, Valérie Gabelica, Rita Grandori, Erik G. Marklund, Thomas Menneteau, Lukasz G. Migas, Argyris Politis, et al.

To cite this version:

Timothy M. Allison, Perdita Barran, Sarah Cianférani, Matteo T. Degiacomi, Valérie Gabelica, et al.. Computational strategies and challenges for using native ion mobility mass spectrometry in biophysics and structural biology. Analytical Chemistry, 2020, 92 (16), pp.10872-10880. 10.1021/acs.analchem.9b05791 . hal-02960838

HAL Id: hal-02960838

https://hal.science/hal-02960838v1

Submitted on 16 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Computational strategies and challenges for using native ion mobility mass spectrometry in biophysics and structural biology

Journal:	*Analytical Chemistry*
Manuscript ID	ac-2019-05791x.R3
Manuscript Type:	Perspective
Date Submitted by the Author:	15-Jul-2020
Complete List of Authors:	Allison, Timothy; University of Oxford, Chemistry Barran, Perdita; University of Manchester, The School of Chemistry Cianférani, Sarah; Institut Pluridisciplinaire Hubert Curien (IPHC), Laboratoire de Spectrométrie de Masse Bio-Organique (LSMBO) Degiacomi, Matteo; Durham University, Chemistry Gabelica, Valérie; Inserm, U1212 ARNA (Inserm/CNRS/Univ. Bordeaux) Grandori, Rita; Universita degli Studi di Milano-Bicocca, Department of Biotechnology and Biosciences Marklund, Erik; Uppsala Universitet Teknisk-naturvetenskapliga fakulteten, Menneteau, Thomas; University College London Migas, Lukasz; Delft University of Technology, 3mE Politis, Argyris; King's College London, Chemistry Sharon, Michal; Weizmann Institute of Science, Biological Chemistry Sobott, Frank; University of Leeds, FBS - Astbury Thalassinos, Konstantinos; University College London, Institute of Structural and Molecular Biology Benesch, Justin; University of Oxford, Department of Chemistry

SCHOLARONE™
Manuscripts

Computational strategies and challenges for using native ion mobility mass spectrometry in biophysics and structural biology

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

[1] *School of Physical and Chemical Sciences, Biomolecular Interaction Centre, University of Canterbury, Christchurch 8140, New Zealand*

[2] *Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute of Biotechnology, School of Chemistry, University of Manchester, Manchester, M1 7DN, U.K.*

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

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

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

[6] *Department of Biotechnology and Biosciences, University of Milano-Bicocca, 20126, Milan, Italy*

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

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

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

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

[11] *Biomolecular & Analytical Mass Spectrometry, Department of Chemistry, University of Antwerp, 2020 Antwerp, Belgium.*

[12] *School of Molecular and Cellular Biology, University of Leeds, Leeds, LS2 9JT, U.K.*

[13] *Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, U.K.*

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

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

All authors contributed equally

*Address correspondence to JLPB: justin.benesch@chem.ox.ac.uk

Abstract

Native mass spectrometry (MS) allows the interrogation of structural aspects of macromolecules in the gas phase, under the premise of having initially maintained their solution-phase non-covalent interactions intact. In the more than 25 years since the first reports, the utility of native MS has become well established in the structural biology community. The experimental and technological advances during this time have been rapid, resulting in dramatic increases in sensitivity, mass range, resolution, and complexity of possible experiments. As experimental methods have improved, there have been accompanying developments in computational approaches for analysing and exploiting the profusion of MS data in a structural and biophysical context. In this perspective, we consider the computational strategies currently being employed by the community, aspects of best practice, and the challenges that remain to be addressed. Our article is based on discussions within the European Cooperation in Science and Technology Action on Native Mass Spectrometry and Related Methods for Structural Biology (EU COST Action BM1403), which involved participants from across Europe and North America. It is intended not as an in-depth review, but instead to provide an accessible introduction to and overview of the topic – to inform newcomers to the field and stimulate discussions in the community about addressing existing challenges. Our complementary perspective article (also in this issue) focuses on software tools available to help researchers tackle some of the challenges enumerated here.

Introduction

Native mass spectrometry (MS) involves the transfer of proteins and other macromolecules intact into the gas phase with minimal disruption to the non-covalent interactions that are present in their solvated form. This then allows a range of experiments to probe the macromolecules' higher-order structure, including their fold, assembly and non-covalent interactions [1-4]. Native MS has helped elucidate various aspects of biomolecular structure, including the subunit composition, stoichiometry and stability of complexes, as well as the dynamic behaviour they display. When combined with ion mobility (IM), where ions are separated based on their mobility through an inert buffer gas (kept at constant pressure and temperature) under a weak electric field, the size, in the form of a rotationally averaged collision cross section (CCS), of a macromolecule can be probed [5]. By virtue of being inherently dispersive, native IM-MS has a unique capability to characterize individual states in heterogeneous and dynamic systems, such as co-populated conformations or assembly states of complexes. Thus, native IM-MS has enabled a large number of insights into a diverse array of macromolecular systems, encompassing proteins, nucleic acids, carbohydrates and lipids, and combinations thereof [6-8].

Proteins and other macromolecules are typically dynamic, in that they populate a range of interconverting structures at equilibrium. Frequently, this heterogeneity is such that macromolecules are better described as structural ensembles (of conformations and/or assemblies), defined by the free-energy landscape accessible at given conditions. IM-MS is sensitive to some of this complexity, providing sparse data that can be a powerful descriptor of molecular states. These data on their own are not sufficient for characterizing molecular structure at atomic detail, but they can, in combination with other information, provide insight into the native state and surrounding free-energy landscape [9-10].

Native IM-MS is conducted in the absence of bulk solvent, a factor which may induce some structural changes in the molecules under analysis. Because the gas-phase structures of large biomolecules are dictated by numerous non-covalent interactions – many of which are far from the molecular surface – they typically retain the vast majority of their solution-phase character [11-12]. However, the removal of solvent and acquisition of charges alters the physico-chemical environment of the protein, and leads to some degree of restructuring into

different conformations, particularly for states that are intrinsically disordered or only marginally stable [13]. This provides an opportunity for experimental exploration of their free-energy landscape, albeit one reflecting – and dependent on – the gas-phase interaction strengths of residues involved [14].

The large body of work developing and employing native IM-MS has indicated that a wealth of information is obtainable from such experiments, and principles of best experimental practice and standards have emerged[15-17]. Yet structural interpretation and translation of the data into structural biology information is often not straightforward. Here we give a perspective on the computational frameworks that must be put in place to address this challenge, and we describe the current thinking and state-of-the-art of the approaches that are being developed. Our article is intended to provide what we believe is a much-needed accessible introduction to and overview of the topic – to inform newcomers to the field and stimulate discussions about addressing existing challenges. We chart where we believe the field stands in terms of progress in five key computational themes and their interconnections (**Figure 1**), devoting one section of the text to each theme: 1) IM-MS data extraction and analysis, 2) CCS calculation, 3) computational modelling, 4) simulating charge distribution, and 5) gas-phase molecular dynamics (MD). Our thoughts are heavily influenced by discussions and contributions from the wider native IM-MS community, nucleated through the European Cooperation in Science and Technology Action on Native Mass Spectrometry and Related Methods for Structural Biology (EU COST Action BM1403), an international group focused on developing and applying new biomolecular MS methods to make the characterization of protein structure and dynamics more rapid and routine. We also refer interested readers to our companion perspective article, also in this issue, which details specific software that will aid users in extracting the most, and most reliable, information from their data.

Computational considerations in converting native IM-MS data into information

The cornerstone in using IM-MS data is to extract the raw data into a format from which it is possible to determine the key physical properties of the ions under investigation. At the most basic level, this comprises the mass, charge and mobility. All of these properties do not have single values but rather populate distributions, reflecting at least in part the heterogeneity of the system at hand (**Table 1**). The values and distributions these physical properties take represent the core information that can be obtained from native IM-MS experiments, and can then be used to infer meaningful structural, dynamical and functional insights into the proteins under study. In general, this is a well-advanced area of study for computational methods in native IM-MS (**Figure 1**); but key improvements remain to be made.

While instrument manufacturers' software typically allows the transformation of the measured mass-to-charge (*m/z*) spectrum onto a mass axis via the assignment of charge states, the frequent complexity of native MS data can make this process difficult. Charge-state assignment can be ambiguous for high charge states (where the difference between z and z+1 is small relative to z)[18] , and residual adducts are typical for large macromolecules [19]. Moreover, the samples under analysis themselves frequently contain multiple components, and can sometimes be extremely heterogeneous[20-22]. Another challenge is that peaks can be poorly resolved, due to the gentle nature of the ionization process employed. To overcome these challenges, both researchers and instrument vendors have developed software and algorithms tailored specifically to native MS data in order to aid users in their analysis (see our companion article for a comprehensive catalogue of available tools). Nevertheless, the limiting factor to obtaining high quality information remains with the data, and over-fitting is a risk that always needs to be considered.

While calibration of the *m/z* axis is straightforward, in order to transform the mobility information (typically acquired in the form of an arrival time distribution, ATD) into a CCS axis, a further calibration procedure is typically required [15]. In the overwhelming majority of cases, this is achieved using reference standards appropriate to the target analyte [23]. This process is sensitive to the conditions under which the experiments are performed, and care must be taken to minimize biases associated with the choice of solution and sampling

conditions, instrument settings, selection of standards, and the calibration procedures [15]. The information encoded in the CCS (and in CCS distributions, CCSDs) is often used to infer structural properties of a given analyte and can inform computational modelling and (in principle) molecular dynamics (MD) simulations. It also enables direct comparisons of molecular states without additional calibration and computational modelling – as systematic biases cancel when making relative measurements. Nonetheless in all these uses, an important (but underexplored) consideration is the appropriate incorporation of uncertainties associated with the native IM-MS measurement and its transformation into CCS.

The ATDs and corresponding CCSDs can differ considerably in profile and width, reflecting (after accounting for instrument-dependent resolving power and other effects [24]) the conformational heterogeneity of the analyte [25-26]. The width of these distributions can be exploited directly, or deconvolved into multiple Gaussian contributions in the case of feature-rich peak shapes [27]. IM-MS experiments can be data-rich, but objective deconvolution of complex ATDs into information of value remains challenging. The difficulty arises in having to decide on the number of conformational families present in the data, and the selection of appropriate width for each Gaussian. Higher resolution IM instrumentation and/or use of tandem IM-MS approaches might enable the separation and resolution of overlapping populations, at least for certain types of samples [28-31].

Calculating CCSs from structures and models

The translation of CCS data obtained during native IM-MS experiments into structural information involves several challenges, including determining how best to obtain the CCS values of the relevant reference structure of the computational model (generation of structures and models is outside the scope of this review). For instance, the user may wish to compare their experimental CCS to available atomic coordinates or to use the CCS to distinguish between various structural hypotheses. A number of approaches exist and selection of the most appropriate method depends on a multitude of factors including the chemical nature of the system under investigation, its shape and intrinsic dynamics, and experimental conditions such as the IM buffer gas [5]. A practical consideration is a trade-off between computational expediency and accuracy in CCS estimations: building a large number of models lets one screen a wider structural space, while performing higher accuracy calculations necessitates screening a smaller range of structures. As such, while we consider this topic to be quite advanced (**Figure 1**), a key challenge is to overcome these compromises.

In its most simplistic form, the CCS can be viewed as the rotationally averaged projected area ("shadow") of an object [32], plus a layer having a thickness related to the gas radius and its polarizability [5]. For any convex object, the projected area is equal to a quarter of its surface area [33]. This simple analytical relationship is useful when considering protein structure at an extremely coarse-grained level [34]. However, when considering protein structure at higher resolution, it is however clear that they are not convex, but feature cavities and protrusions that can lead to multiple collisions or occlude portions of the protein surface from collisions with the buffer gas [35]. On a finer scale, the surface roughness due to the amino acids that decorate the exterior influence the drag a protein experiences during the IM-MS experiment and severs the relation between surface area and projected area. Furthermore, the charge on the protein is inherently non-zero in ion mobility and is expected to impact on CCSs, modulated by the dipole moment and polarizability volume of the gas. The exact distribution of charge can in principle affect the mobility [36], but appears to have a minor effect on the CCSs of proteins [37-38]. For moderate charge states (i.e. the low amount of charge per unit mass typical in native mass spectra), the CCS appears to be relatively constant in He, but less so in N_2 [37-38]. How this phenomenon manifests itself for proteins of all sizes and shapes, and for other types of macromolecules, is currently not known, but

neglecting these effects is unlikely to be the major source of bias; more important perhaps are the perturbations the charges make to the structure (see below). Nevertheless, given the increasingly sophisticated questions that IM-MS is being used to answer, and the higher performance IM-MS instruments that have become available [39-40], considerable scope remains to ensure that local charges and interaction potentials are effectively accommodated in CCS calculations.

Different computational approaches (and implementations thereof) for estimating CCSs from structures exist, at differing levels of complexity and computational cost (see our companion paper, and others [5, 15]). The simplest and fastest approach is to consider a protein in terms of its area when projected from different viewpoints. Here the gas atoms are represented by hard spheres that are 'fired' through the sampling volume, and the projected area is calculated from the fraction of trajectories that collide with the protein. A bit more advanced, the exact hard spheres scattering model computes the angle of deflection of the gas to calculate the corresponding deflection (momentum transfer) for the ion. Both approaches ignore electrostatic interactions, and they ignore London dispersion forces acting at long range.

In the methods at the other end of the complexity spectrum however (several methods are found between these extremes), the short- and long-range interactions of the protein with the gas molecules are modelled explicitly, accounting for both the physico-chemical properties (polarizability, charge, Van der Waals interactions, and potentially internal degrees of freedom) of the gas and of the atoms in the protein, requiring numerical integration of gas-particle trajectories with numerous iterations for each such trajectory. While this more rigorous and explicit consideration of the physical processes underpinning the IM separation might provide more accurate CCSs for atomistic structure models, it does not readily lend itself to coarse-grained structural representations, whereas it is readily achievable to calculate the projected area of e.g. SAXS-derived bead models or iso-surfaces from electron microscopy [26, 41]. Consequently, the nature of the structure model can effectively narrow the repertoire of applicable methods for CCS calculation.

The difference in computational cost between these two extremes currently spans several orders of magnitude, with the most complex approaches taking hours to converge when applied to macromolecules. This renders them intractable for assessing the hundreds of thousands of models needed to explore adequately the roto-

translational space associated with structure modelling, or the thousands of frames from MD simulations. As a result, it is often only feasible to use simpler approaches, potentially compromising on the accuracy of the CCS estimation. However, in order to deduce ion shapes from IM-MS, what matters is not so much the accuracy of the absolute calculated values but rather how accurately they can be matched to experiment. For example, for large and globular proteins the simplest projection approximation method can be generally parameterised (i.e. scaled, or calibrated) to reproduce the results from the most computationally costly trajectory method with a relative error within 1% [26], and experimental drift-tube helium CCS values to within 3% RMSD [42]. In general, appropriate parameterization of the CCS calculation is as important as the underlying physical model that is being used [15], and one must pay attention to the type and size of system for which a given parameterization was developed, as well as to the type of experiment it was designed to match. For example, no simple parameterization has been thoroughly validated for proteins that are grossly convex, intrinsically disordered, or in extreme charge states. For smaller systems, the relative effect of surface interactions will be proportionally greater than for very large ones. For highly concave structures, a simple projection approach will not take into account "parachute" effects on ion friction. In all these cases, or whenever in doubt, more expensive methods are necessary for good accuracy [43-44].

Modelling protein structures using IM-MS data

Computational methods are needed to exploit native IM-MS data for validating or modelling three-dimensional protein structures. A typical workflow involves distinct steps: converting the experimental data acquired into modelling restraints, building models that sample the conformational space of individual proteins or protein assemblies, and evaluating the models in light of the data. Currently, there are two strategies for building models using MS and other related structural datasets. The first strategy filters models generated by computational methods based on their "goodness-of-fit" to the experimental datasets [45-48]. The second strategy samples models by directly integrating the experimentally derived restraints with an appropriate scoring function[a] into the computational workflow – i.e. using the restraint to optimise dynamically the model building [49-50]. Modelling work-flows that use IM-MS data have been established, but there remains considerable scope in refining and extending these to make maximum use of the opportunity available (**Figure 1**).

For modelling analysis, it is important to use appropriate "building blocks". In general, the individual subunits and or complexes can be represented as atomic coordinates (e.g. crystal structures, homology models), as coarse-grained models (e.g. spheroids), or as density maps. Furthermore, it can be important to consider multiple alternative starting structures to ensure that the space is suitably explored [51]. This is pertinent for proteins or complexes that are particularly flexible or are characterised by intrinsically dynamic regions, and where maybe only one particularly stable or abundant structure has been characterized previously e.g. by X-ray crystallography. In such cases, developing robust methods for building alternative starting structures for downstream model building becomes a critical aspect of the computational workflow.

An important aspect of any modelling pipeline is the consideration of the uncertainty introduced at each step of the analysis. First, one must consider ambiguity in the data caused by the limited resolving power of the instruments, the conformational heterogeneity of the protein (which manifests itself as a CCSD broader than

[a] A modelling restraint is defined as an assembly/protein feature (e.g. volume, shape, flexibility) quantified with respect to the data used to generate it. It represents the 'force' that glues the individual subunits and forms configurations consistent with the input data. The scoring function sums up all restraints and may be thought as the force field that enables to make up the assembly.

the instrumentation limit), and the possibility of low-quality data which can compromise the discriminatory ability of the CCS measurements [52-53].

There may also be large discrepancies between the experimentally measured and theoretically calculated CCS values if proteins undergo a significant degree of structural change upon transfer to the gas phase, and these discrepancies bring challenges for modelling. Side chains that are solvent-exposed in solution take advantage of the low permittivity of vacuum to collapse onto the surface by forming new interactions [54-56]. In the case of protein ions that are intrinsically malleable, e.g. hollow structures, those with hinges, or low charge states of intrinsically disordered proteins, these additional (non-native) non-covalent interactions can lead to unstimulated compaction of the overall protein structure [51, 56-61]. Gas-phase induced unfolding happens when the native intramolecular interactions are too weak compared to the repulsion between like charges, and is more likely to occur for high charge states (and at higher activation energies). Gas-phase structural changes require some energy barriers to be overcome, which in turn depends on the native interactions, on the charge state adopted during electrospray, on the internal energy uptake, and on the time spent in the mass spectrometer. Despite notable advances made [62-63], gas-phase structural changes remain hard to fully predict, and thus contribute to the uncertainty of the CCS calculation.

Uncertainty from computations that aim to match experimental data to structural models comprises contributions from the choice of representations [64-65], the completeness of the information available, the use of the appropriate scoring function, and the biases of individual sampling algorithms (e.g. if they don't accurately capture the data). Finally, measurable errors may be introduced by the post-processing step which typically scores models based on how well they match the input datasets, which may include clustering approaches for generating an ensemble of computational models. A final challenge comes in weighting the merits, and biases, of individual methods based on their ability to contribute to accurate models. As such, the final output of a combined experimental and modelling effort is best represented by an ensemble of structures that encapsulates the convolution of both the inherent conformational heterogeneity of the protein and the various sources of uncertainty in the IM-MS pipeline [48, 64]. Benchmarking studies have provided some ways of efficiently integrating the different methods by taking into account the relative uncertainty of the different methods [66-67],

such that it is becoming increasingly possible to bring together the individual techniques in a single workflow [68].

Determining how charges are distributed on the protein

Another challenge stems from considering how charge is distributed on a macromolecule. The locations of charges on a protein can considerably impact how experimental results should be interpreted, for example when interrogating that protein's interactions with other molecules. The charge distribution can also be of importance for computational approaches used in combination with MS. While the locations of charges do not appear critical for CCS calculations on large molecules, they remain an integral part of physical models used to describe the protein and help determine the system dynamics at the atomic level, thereby greatly influencing the accuracy of gas-phase MD simulations. This, of course, reflects the fact that, to a large extent, charge locations 'drive' structural dynamics, and *vice versa*. For macromolecules, charging in electrospray takes place via the protonation of basic sites, and deprotonation of acidic sites[b] . It should be noted that additional sites become available during electrospray due to their high gas-phase basicity or acidity [69], that Zwitterionic states are frequently stable in the gas phase [62, 70], and that, depending on solution conditions, charged buffer components can act as charge carriers.

The final charge configuration can be thought of as resulting from the balance between electrostatics – the attraction and repulsion between charged sites – acting at relatively long range, and short-range 'chemical' forces – the intrinsic propensities of different chemical moieties, such as amino acid side-chains, to accept (or donate) a proton. The playing field for the two forces is determined by the structure, and also by surrounding residues that can modulate the interaction between charges and solvate (de)protonated sites. Depending on the conditions under which the electrospray process generates charged particles, particularly the presence or absence of protic solvent and the time frame of ionization, the removal of solvent greatly affects the energetics of both the protonated and deprotonated form. Akin to pK_as in water solution, the gas-phase basicities describe the protonation equilibria in the gas phase for the isolated acidic and basic sites, but the higher the net charge of the system, the greater the influence of the electrostatics. However, because of a certain amount of kinetic trapping, the site might still carry some 'memory' of its protonation state in solution over the experimental

[b] Note that 'basic/acidic sites' is here used according to the Brønsted–Lowry definition, that is, their ability to accept or donate a proton. As such, aspartate and glutamate residues are basic *sites*, as they are corresponding bases to aspartic acid and glutamic acid, whereas they are typically considered to be acidic *residues* in biochemistry, regardless of protonation state.

time scales [71]. Experimentally pinpointing the location of charges is extremely difficult, and one cannot assume that protonation states simply carry over from solution to the gas phase, making computational approaches the primary means for mapping charge location. However, much work remains to be done (**Figure 1**).

The number of possible charge isomers grows rapidly with the number of (de)protonatable sites, meaning that a complete consideration of isomers is usually not feasible. It also means that any computational evaluation of a particular charge configuration must be fast to be of practical use, effectively restricting the modelling to force-field based methods or coarse-grained models. Higher-level methods can support charge-distribution calculations, however, by providing reliable salt-bridge energies, gauging the effects of polarization, etc. In lieu of complete enumeration of charge isomers, Monte-Carlo approaches, where protons are moved randomly between basic sites to generate new charge isomers, have been developed to more efficiently explore the landscape of possible charge isomers[58, 72]. While the details of how the energies are evaluated and how the charge isomers are sampled vary among the different approaches, they all compute isomer energy from the sum of the gas-phase basicities or proton affinities for all protonated sites, and the electrostatic interactions between charged sites and their surroundings (including other charged sites).

Hybrid MD - Monte-Carlo approaches have been developed for the combined search of conformer and charge-isomer space in the gas phase. These indicate that side chains have a propensity to fold onto the protein surface with consequent structure contraction and formation of new hydrogen bonds and salt bridges [70], a prediction for which experimental evidence is emerging [73]. Such structural rearrangements promote self-solvation and are compatible with maintenance of a native-like fold. An interesting feature in the emerging picture of folded protein ions in the gas phase is their ability to compensate for the energetic penalty of charge separation *in vacuo* with favourable, conformation-specific intramolecular interactions, in line with growing experimental and theoretical evidence [74-75]. Persistence of zwitterionic states in protein structures provides a rationale for conformational stability in the gas phase and conformational effects on charge-state distributions, and is a feature that simulation methods should accommodate. The interplay between charge and conformation means that even if the lowest-energy charge isomer can be identified for a crystal structure, relaxation of sidechains and higher conformational levels might shift the energy considerably [72]. Care must be taken to not let the rich structural detail in a crystal structure, obtained under considerably different conditions, bias the calculations

towards “incorrect” charge isomers. In totality, a successful computational approach for charge placement must be able to cover a vast number of possible charge isomers, evaluate the energy of each one very rapidly, and allow for some flexibility or dynamics in the macromolecule to accommodate the coupling between protonation and structure.

Gas-phase molecular dynamics harnessed to native IM-MS

The integration of native IM-MS experiments with MD simulations is highly desirable, as the two methods are complementary with respect to the resolution of the structural information they provide and the timescales that they operate on [9]. Solvent-free MD plays an important role in understanding the fundamentals of MS and for interpreting MS data [12, 56]. For example, the effects of solvent, temperature and charge on protein structure have been studied in this way, and there are numerous examples of system-specific investigations where MD has been used together with MS [9]. However, the computational hurdles to be overcome to attain truly satisfactory integrated IM-MS and MD workflows are several, and the field is still in its infancy (**Figure 1**). This is primarily because the most widespread MD methods have been developed mainly for condensed-phase calculations, which presents specific challenges when applying them to simulations in vacuum. For example, electrostatic interactions are significant over much longer distances in the absence of solvent which, if taken into account, slows down the calculations considerably, thus limiting the sampling and simulation timescales. Moreover, the commonly used force fields are designed to match the solution phase, and hence the effective polarization at the solution interface might not reflect gas-phase conditions. The magnitude of this inaccuracy is currently unquantified, however employing polarizable force fields could be a means to mitigate such errors at an additional computational cost [56].

Since the protein dynamics in the gas-phase are intertwined with charge isomerism, the accuracy of any MD simulation is limited by how realistically charges are placed on the protein. In addition to the combinatorial challenges in choosing a "correct" charge isomer, there may be several co-existing charge isomers, and protons could in principle transfer between sites in the gas phase (the "mobile proton model" [76]), following or promoting structural transitions [77]. As classical MD typically disallows the breakage or cleavage of chemical bonds, protonation dynamics cannot readily be incorporated into such simulations, and some structural processes may therefore be difficult to recapitulate. Recently there has been progress in accommodating proton mobility, with simulations being stopped at regular intervals and charges transferred at random towards charge isomers of lower energy [62-63, 78-79]. Current implementations of this approach are, however, not truly thermodynamic, in the sense that they do not strictly adhere to Boltzmann statistics, and consequently, they might be error-prone in quantifying how probable the different charge isomers are. Nevertheless, this

represents an important step towards accommodating the important role of charges in gas-phase MD, and future integration with popular MD software will be instrumental for the community. Combined quantum mechanics/molecular dynamics (QM/MM) would be a more accurate way to account for proton transfer [80]; although computationally much more costly than force field MD, it may prove valuable to IM-MS modelling in the future.

The transition from solution to the gas phase can also incur changes in the structure of the protein. Though these are often small in amplitude [81], they can significantly alter the contacts made between amino acids [56]. The absence of hydrogen bonding opportunities provided by a solvent likely makes the potential energy surface more rugged. This, together with the need to consider electrostatic interactions over long distances, means that MD might struggle to explore experimentally relevant parts of the conformational landscape [56, 61]. Experimental data from solution-phase methods are frequently used to restrain the MD simulations, facilitating the transition from the starting structure to the conformations that pertain to the question at hand. In principle, experimentally derived CCSs could be used in a similar fashion, but the considerable overhead required for continuously calculating the CCS during the simulation and comparing it with a given reference value has so far limited the use of CCS-based restraints [9]. Instead, other, more computationally expedient quantities, such as the radius of gyration or solvent accessible surface area (SASA), have been used as proxies for the CCS [44, 82-83]. Recent speed increases in CCS calculations might enable explicit CCS restraints, strengthening the link between simulation and experiments, especially for systems where non-globular structures or conformational transitions might complicate the relationship between proxies and CCSs.

Summary and outlook

Native IM-MS has the potential to significantly impact structural biology, analogous to the revolution that MS has enabled in proteomics. It is also clear that native MS-derived information benefits from being combined with results obtained from other, orthogonal techniques. These can be other MS-based approaches, such as chemical cross-linking, hydrogen-deuterium exchange, and covalent-labelling (footprinting) approaches, or other structural biology techniques altogether. The resulting "hybrid" strategies enable more accurate and confident structure modelling, particularly in the absence of high-resolution atomistic structures, and extend the validity of these models by sampling heterogeneous conformational and assembly space. However, in order to maximize the potential of native IM-MS, computational strategies that facilitate the translation of the raw data it produces into structural models with associated dynamics, as well as providing a deep understanding of the processes that occur between the protein in solution and its detection in the mass spectrometer, will be instrumental.

The development of computational strategies for native IM-MS centres on five inter-related themes (**Figure 1)**. The most mature theme at present is also the which fundamentally unlocks the potential of the technique: the conversion of native IM-MS data into information on ions' key physical properties, particularly mass, charge and mobility. Considerable progress has also been made in developing approaches to calculate CCSs from structures and models, and in three-dimensional modelling of protein structures using IM-MS data. There are two interconnected areas where progress is less advanced so far: determining how charges are distributed on proteins, and combining gas-phase MD with native IM-MS. Moreover, these two topics are mutually dependent, which means that limitations in the accuracy of estimations of one affect the other, and efforts to develop them further must be done hand-in-hand.

We imagine an era of structural proteomics where macromolecular structures can be computed in a high-throughput manner by exploiting native IM-MS data. Here, we have reviewed the key challenges to achieve this aim (**Box 1**). The high pace of activity in the field augurs well for these issues being resolved in the not-too-distant future. These efforts will benefit from the complementary perspectives of the structural MS

community, who bring insight into gas-phase effects derived from decades of study on small molecules, and computational structural biologists, who are aware of the priorities and sensitivities in modelling and MD. Success in this endeavour will ultimately enable deeper and more quantitative insights from harnessing MS data into understanding the structure, dynamics and interactions of biomolecules, impacting on our understanding of biological (mal)function as well.

Acknowledgements

We thank the EU COST Action BM1403 and members of WG1 (Native IM-MS) and WG4 (Computational Methods); and Dr Catherine Lichten for assistance in compiling this manuscript.

References

1. Calabrese, A. N.; Radford, S. E., Mass spectrometry-enabled structural biology of membrane proteins. *Methods* **2018,** *147*, 187-205.

2. Konermann, L.; Vahidi, S.; Sowole, M. A., Mass spectrometry methods for studying structure and dynamics of biological macromolecules. *Anal Chem* **2014,** *86* (1), 213-32.

3. Lossl, P.; van de Waterbeemd, M.; Heck, A. J., The diverse and expanding role of mass spectrometry in structural and molecular biology. *EMBO J* **2016,** *35* (24), 2634-2657.

4. Mehmood, S.; Allison, T. M.; Robinson, C. V., Mass spectrometry of protein complexes: from origins to applications. *Annu Rev Phys Chem* **2015,** *66*, 453-74.

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

6. Ben-Nissan, G.; Sharon, M., The application of ion-mobility mass spectrometry for structure/function investigation of protein complexes. *Curr Opin Chem Biol* **2018,** *42*, 25-33.

7. Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat Chem* **2014,** *6* (4), 281-94.

8. Stuchfield, D.; Barran, P., Unique insights to intrinsically disordered proteins provided by ion mobility mass spectrometry. *Curr Opin Chem Biol* **2018,** *42*, 177-185.

9. Marklund, E. G.; Benesch, J. L., Weighing-up protein dynamics: the combination of native mass spectrometry and molecular dynamics simulations. *Curr Opin Struct Biol* **2019,** *54*, 50-58.

10. Wyttenbach, T.; Pierson, N. A.; Clemmer, D. E.; Bowers, M. T., Ion mobility analysis of molecular dynamics. *Annu Rev Phys Chem* **2014,** *65*, 175-96.

11. Hilton, G. R.; Benesch, J. L., Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *J R Soc Interface* **2012,** *9* (70), 801-16.

12. Meyer, T.; Gabelica, V.; Grubmuller, H.; Orozco, M., Proteins in the gas phase. *Wires Comput Mol Sci* **2013,** *3* (4), 408-425.

13. Clemmer, D. E.; Russell, D. H.; Williams, E. R., Characterizing the Conformationome: Toward a Structural Understanding of the Proteome. *Acc Chem Res* **2017,** *50* (3), 556-560.

14. Chandler, S. A.; Benesch, J. L. P., Mass spectrometry beyond the native state. *Curr. Opin. Chem. Biol.* **2018,** *42*, 130-137.

15. 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, C. J., Jr.; 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 Spectrom Rev* **2019,** *38* (3), 291-320.

16. Kondrat, F. D.; Struwe, W. B.; Benesch, J. L., Native mass spectrometry: towards high-throughput structural proteomics. *Methods Mol Biol* **2015,** *1261*, 349-71.

17. Schachner, L. F.; Ives, A. N.; McGee, J. P.; Melani, R. D.; Kafader, J. O.; Compton, P. D.; Patrie, S. M.; Kelleher, N. L., Standard Proteoforms and Their Complexes for Native Mass Spectrometry. *J Am Soc Mass Spectrom* **2019,** *30* (7), 1190-1198.

18. McKay, A. R.; Ruotolo, B. T.; Ilag, L. L.; Robinson, C. V., Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. *J Am Chem Soc* **2006,** *128* (35), 11433-42.

19. Benesch, J. L.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V., Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem Rev* **2007,** *107* (8), 3544-67.

20. Ben-Nissan, G.; Belov, M. E.; Morgenstern, D.; Levin, Y.; Dym, O.; Arkind, G.; Lipson, C.; Makarov, A. A.; Sharon, M., Triple-Stage Mass Spectrometry Unravels the Heterogeneity of an Endogenous Protein Complex. *Anal Chem* **2017,** *89* (8), 4708-4715.

21. Stengel, F.; Baldwin, A. J.; Painter, A. J.; Jaya, N.; Basha, E.; Kay, L. E.; Vierling, E.; Robinson, C. V.; Benesch, J. L., Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. *Proc Natl Acad Sci U S A* **2010,** *107* (5), 2007-12.

22. Wang, G.; de Jong, R. N.; van den Bremer, E. T. J.; Parren, P.; Heck, A. J. R., Enhancing Accuracy in Molecular Weight Determination of Highly Heterogeneously Glycosylated Proteins by Native Tandem Mass Spectrometry. *Anal Chem* **2017,** *89* (9), 4793-4797.

23. Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **2010,** *82* (22), 9557-9565.

24. Kune, C.; Far, J.; De Pauw, E., Accurate Drift Time Determination by Traveling Wave Ion Mobility Spectrometry: The Concept of the Diffusion Calibration. *Anal Chem* **2016,** *88* (23), 11639-11646.

25. Marchand, A.; Livet, S.; Rosu, F.; Gabelica, V., Drift Tube Ion Mobility: How to Reconstruct Collision Cross Section Distributions from Arrival Time Distributions? *Anal. Chem.* **2017,** *89* (23), 12674-12681.

26. Marklund, E. G.; Degiacomi, M. T.; Robinson, C. V.; Baldwin, A. J.; Benesch, J. L., Collision cross sections for structural proteomics. *Structure* **2015,** *23* (4), 791-9.

27. 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 alpha(1)-antitrypsin peptide binding. *International Journal of Mass Spectrometry* **2018,** *426*, 29-37.

28. Allen, S. J.; Eaton, R. M.; Bush, M. F., Analysis of Native-Like Ions Using Structures for Lossless Ion Manipulations. *Anal. Chem.* **2016,** *88* (18), 9118-9126.

29. Merenbloom, S. I.; Glaskin, R. S.; Henson, Z. B.; Clemmer, D. E., High-Resolution Ion Cyclotron Mobility Spectrometry. *Anal. Chem.* **2009,** *81* (4), 1482-1487.

30. Shepherd, D. A.; Marty, M. T.; Giles, K.; Baldwin, A. J.; Benesch, J. L. P., Combining tandem mass spectrometry with ion mobility separation to determine the architecture of polydisperse proteins. *International Journal of Mass Spectrometry* **2015,** *377*, 663-671.

31. Zhong, Y. Y.; Hyung, S. J.; Ruotolo, B. T., Characterizing the resolution and accuracy of a second-generation traveling-wave ion mobility separator for biomolecular ions. *Analyst* **2011,** *136* (17), 3534-3541.

32. Mack, E., Average cross-sectional areas of molecules by gaseous diffusion methods. *J. Am. Chem. Soc.* **1925,** *47*, 2468-2482.

33. Vouk, V., Projected Area of Convex Bodies. *Nature* **1948,** *162* (4113), 330-331.

34. Hewitt, D.; Marklund, E.; Scott, D. J.; Robinson, C. V.; Borysik, A. J., A Hydrodynamic Comparison of Solution and Gas Phase Proteins and Their Complexes. *J. Phys. Chem. B* **2014,** *118* (29), 8489-8495.

35. Marklund, E. G., Molecular self-occlusion as a means for accelerating collision cross-section calculations. *International Journal of Mass Spectrometry* **2015,** *386*, 54-55.

36. Young, M. N.; Bleiholder, C., Molecular Structures and Momentum Transfer Cross Sections: The Influence of the Analyte Charge Distribution. *J Am Soc Mass Spectrom* **2017,** *28* (4), 619-627.

37. Canzani, D.; Laszlo, K. J.; Bush, M. F., Ion Mobility of Proteins in Nitrogen Gas: Effects of Charge State, Charge Distribution, and Structure. *J Phys Chem A* **2018,** *122* (25), 5625-5634.

38. Shrivastav, V.; Nahin, M.; Hogan, C. J.; Larriba-Andaluz, C., Benchmark Comparison for a Multi-Processing Ion Mobility Calculator in the Free Molecular Regime. *J Am Soc Mass Spectrom* **2017,** *28* (8), 1540-1551.

39. Benigni, P.; Marin, R.; Molano-Arevalo, J. C.; Garabedian, A.; Wolff, J. J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F., Towards the analysis of high molecular weight proteins and protein complexes using TIMS-MS. *Int J Ion Mobil Spec* **2016,** *19* (2-3), 95-104.

40. 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,** *91* (12), 7554-7561.

41. 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.

42. Benesch, J. L.; Ruotolo, B. T., Mass spectrometry: come of age for structural and dynamical biology. *Curr Opin Struct Biol* **2011,** *21* (5), 641-9.

43. Bleiholder, C.; Wyttenbach, T.; Bowers, M. T., A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method. *International Journal of Mass Spectrometry* **2011,** *308* (1), 1-10.

44. Kulesza, A.; Marklund, E. G.; MacAleese, L.; Chirot, F.; Dugourd, P., Bringing Molecular Dynamics and Ion-Mobility Spectrometry Closer Together: Shape Correlations, Structure-Based Predictors, and Dissociation. *J Phys Chem B* **2018,** *122* (35), 8317-8329.

45. Baldwin, A. J.; Lioe, H.; Hilton, G. R.; Baker, L. A.; Rubinstein, J. L.; Kay, L. E.; Benesch, J. L., The polydispersity of alphaB-crystallin is rationalized by an interconverting polyhedral architecture. *Structure* **2011,** *19* (12), 1855-63.

46. D'Urzo, A.; Konijnenberg, A.; Rossetti, G.; Habchi, J.; Li, J.; Carloni, P.; Sobott, F.; Longhi, S.; Grandori, R., Molecular basis for structural heterogeneity of an intrinsically disordered protein bound to a partner by combined ESI-IM-MS and modeling. *J Am Soc Mass Spectrom* **2015,** *26* (3), 472-81.

47. Politis, A.; Park, A. Y.; Hyung, S. J.; Barsky, D.; Ruotolo, B. T.; Robinson, C. V., Integrating ion mobility mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes. *PLoS One* **2010,** *5* (8), e12080.

48. 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. *J Biol Chem* **2018,** *293* (51), 19511-19521.

49. Alber, F.; Forster, F.; Korkin, D.; Topf, M.; Sali, A., Integrating diverse data for structure determination of macromolecular assemblies. *Annu Rev Biochem* **2008,** *77*, 443-77.

50. Thalassinos, K.; Pandurangan, A. P.; Xu, M.; Alber, F.; Topf, M., Conformational States of macromolecular assemblies explored by integrative structure calculation. *Structure* **2013,** *21* (9), 1500-8.

51. 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. *Angew Chem Int Ed Engl* **2018,** *57* (52), 17194-17199.

52. 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), 1596-1609.

53. Karaca, E.; Bonvin, A. M., On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys. *Acta Crystallogr D Biol Crystallogr* **2013,** *69* (Pt 5), 683-94.

54. Breuker, K.; McLafferty, F. W., Stepwise evolution of protein native structure with electrospray into the gas phase, 10(-12) to 10(2) s. *Proc Natl Acad Sci U S A* **2008,** *105* (47), 18145-52.

55. Loo, R. R.; Loo, J. A., Salt Bridge Rearrangement (SaBRe) Explains the Dissociation Behavior of Noncovalent Complexes. *J Am Soc Mass Spectrom* **2016,** *27* (6), 975-90.

56. van der Spoel, D.; Marklund, E. G.; Larsson, D. S. D.; Caleman, C., Proteins, Lipids, and Water in the Gas Phase. *Macromol. Biosci.* **2011,** *11* (1), 50-59.

57. Rolland, A. D.; Prell, J. S., Computational insights into compaction of gas-phase protein and protein complex ions in native ion mobility-mass spectrometry. *Trac-Trend Anal Chem* **2019,** *116*, 282-291.

58. Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V., Charge-State Dependent Compaction and Dissociation of Protein Complexes: Insights from Ion Mobility and Molecular Dynamics. *J. Am. Chem. Soc.* **2012,** *134* (7), 3429-3438.

59. Pacholarz, K. J.; Porrini, M.; Garlish, R. A.; Burnley, R. J.; Taylor, R. J.; Henry, A. J.; Barran, P. E., Dynamics of intact immunoglobulin G explored by drift-tube ion-mobility mass spectrometry and molecular modeling. *Angew Chem Int Ed Engl* **2014,** *53* (30), 7765-9.

60. Pagel, K.; Natan, E.; Hall, Z.; Fersht, A. R.; Robinson, C. V., Intrinsically disordered p53 and its complexes populate compact conformations in the gas phase. *Angew Chem Int Ed Engl* **2013,** *52* (1), 361-5.

61. Landreh, M.; Marklund, E. G.; Uzdavinys, P.; Degiacomi, M. T.; Coincon, M.; Gault, J.; Gupta, K.; Liko, I.; Benesch, J. L.; Drew, D.; Robinson, C. V., Integrating mass spectrometry with MD simulations reveals the role of lipids in Na(+)/H(+) antiporters. *Nat Commun* **2017,** *8*, 13993.

62. Konermann, L., Molecular Dynamics Simulations on Gas-Phase Proteins with Mobile Protons: Inclusion of All-Atom Charge Solvation. *The journal of physical chemistry. B* **2017,** *121* (34), 8102-8112.

63. Marchese, R.; Grandori, R.; Carloni, P.; Raugei, S., A computational model for protein ionization by electrospray based on gas-phase basicity. *J Am Soc Mass Spectrom* **2012,** *23* (11), 1903-10.

64. Eschweiler, J. D.; Frank, A. T.; Ruotolo, B. T., Coming to Grips with Ambiguity: Ion Mobility-Mass Spectrometry for Protein Quaternary Structure Assignment. *J Am Soc Mass Spectrom* **2017,** *28* (10), 1991-2000.

65. 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.

66. Schneidman-Duhovny, D.; Pellarin, R.; Sali, A., Uncertainty in integrative structural modeling. *Curr Opin Struct Biol* **2014,** *28*, 96-104.

67. Tamo, G.; Maesani, A.; Trager, S.; Degiacomi, M. T.; Floreano, D.; Dal Peraro, M., Disentangling constraints using viability evolution principles in integrative modeling of macromolecular assemblies. *Sci Rep* **2017,** *7* (1), 235.

68. 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.

69. Xia, H.; Attygalle, A. B., Untrapping Kinetically Trapped Ions: The Role of Water Vapor and Ion-Source Activation Conditions on the Gas-Phase Protomer Ratio of Benzocaine Revealed by Ion-Mobility Mass Spectrometry. *J Am Soc Mass Spectrom* **2017,** *28* (12), 2580-2587.

70. Marchese, R.; Grandori, R.; Carloni, P.; Raugei, S., On the zwitterionic nature of gas-phase peptides and protein ions. *PLoS Comput Biol* **2010,** *6* (5), e1000775.

71. Warnke, S.; Seo, J.; Boschmans, J.; Sobott, F.; Scrivens, J. H.; Bleiholder, C.; Bowers, M. T.; Gewinner, S.; Schollkopf, W.; Pagel, K.; von Helden, G., Protomers of benzocaine: solvent and permittivity dependence. *J Am Chem Soc* **2015,** *137* (12), 4236-42.

72. Wanasundara, S. N.; Thachuk, M., Theoretical investigations of the dissociation of charged protein complexes in the gas phase. *J Am Soc Mass Spectrom* **2007,** *18* (12), 2242-53.

73. McAlary, L.; Harrison, J. A.; Aquilina, J. A.; Fitzgerald, S. P.; Kelso, C.; Benesch, J. L. P.; Yerbury, J. J., Trajectory Taken by Dimeric Cu/Zn Superoxide Dismutase through the Protein Unfolding and Dissociation Landscape Is Modulated by Salt Bridge Formation. *Anal Chem* **2020,** *92* (2), 1702-1711.

74. Bakhtiari, M.; Konermann, L., Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures. *J Phys Chem B* **2019,** *123* (8), 1784-1796.

75. Bonner, J.; Lyon, Y. A.; Nellessen, C.; Julian, R. R., Photoelectron Transfer Dissociation Reveals Surprising Favorability of Zwitterionic States in Large Gaseous Peptides and Proteins. *J Am Chem Soc* **2017,** *139* (30), 10286-10293.

76. Boyd, R.; Somogyi, A., The mobile proton hypothesis in fragmentation of protonated peptides: a perspective. *J Am Soc Mass Spectrom* **2010,** *21* (8), 1275-8.

77. Mistarz, U. H.; Chandler, S. A.; Brown, J. M.; Benesch, J. L. P.; Rand, K. D., Probing the Dissociation of Protein Complexes by Means of Gas-Phase H/D Exchange Mass Spectrometry. *J Am Soc Mass Spectrom* **2019,** *30* (1), 45-57.

78. Konermann, L.; Metwally, H.; McAllister, R. G.; Popa, V., How to run molecular dynamics simulations on electrospray droplets and gas phase proteins: Basic guidelines and selected applications. *Methods* **2018,** *144*, 104-112.

79. Porrini, M.; Rosu, F.; Rabin, C.; Darre, L.; Gomez, H.; Orozco, M.; Gabelica, V., Compaction of Duplex Nucleic Acids upon Native Electrospray Mass Spectrometry. *ACS Cent Sci* **2017,** *3* (5), 454-461.

80. Li, J.; Lyu, W.; Rossetti, G.; Konijnenberg, A.; Natalello, A.; Ippoliti, E.; Orozco, M.; Sobott, F.; Grandori, R.; Carloni, P., Proton Dynamics in Protein Mass Spectrometry. *J Phys Chem Lett* **2017,** *8* (6), 1105-1112.

81. Meyer, T.; de la Cruz, X.; Orozco, M., An atomistic view to the gas phase proteome. *Structure* **2009,** *17* (1), 88-95.

82. Calvo, F.; Chirot, F.; Albrieux, F.; Lemoine, J.; Tsybin, Y. O.; Pernot, P.; Dugourd, P., Statistical analysis of ion mobility spectrometry. II. Adaptively biased methods and shape correlations. *J Am Soc Mass Spectrom* **2012,** *23* (7), 1279-88.

83. Chirot, F.; Calvo, F.; Albrieux, F.; Lemoine, J.; Tsybin, Y. O.; Dugourd, P., Statistical analysis of ion mobility spectrometry. I. Unbiased and guided replica-exchange molecular dynamics. *J Am Soc Mass Spectrom* **2012,** *23* (2), 386-96.

Figure 1

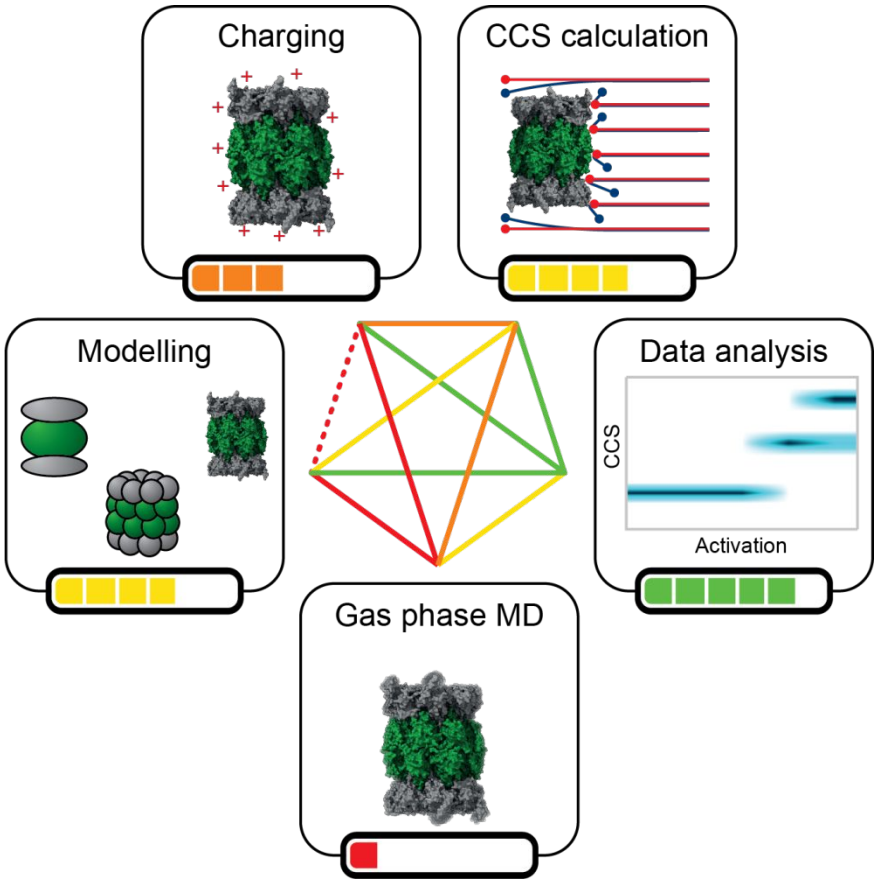


Fig.1: *Overview of progress in what we view as the five key computational themes for IM-MS. The "battery icons" depict our assessment of relative progress in each area, as described in the text. The central pentagram shows the links between these themes, and the edges are coloured according to how well the themes are currently integrated. The dashed line refers to a link whose relevance is not (yet) evident.*

Table 1

IM-MS Data	Information	Utility
Mass to charge ratio (*m*/*z*)	Mass	Stoichiometry
	Charge	Conformation/structure, surface properties
Charge state distribution	Charge distribution	Qualitative view of conformational diversity
Average arrival time*	Ion mobility	Conformation/structure (relative)
	CCS	Conformation/structure (absolute)
Arrival time distribution*	Ion mobility distribution	Relative conformational diversity
	CCSD	Absolute conformational diversity
IM-MS Experiment	**Data**	**Utility**
Stimulated unfolding (e.g. CIU)	Arrival time v activation	Quantify gas-phase stability
Stimulated dissociation (e.g. CID)	*m*/*z* v activation	Determine composition and stoichometry
Time course	Arrival time and *m*/*z* v time	Kinetics of assembly, disassembly and conformational exchange reactions
Titration	Arrival time and *m*/*z* v concentration	Solution stability (thermodynamics)

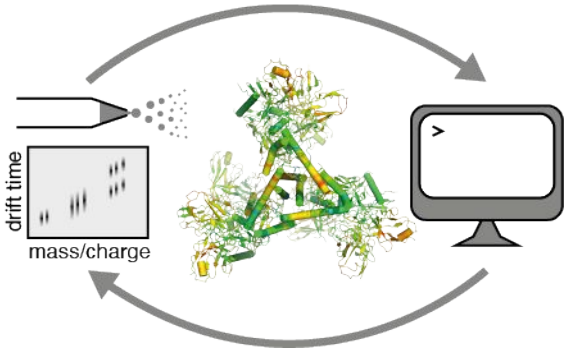
* or equivalent from non-drift tube or travelling wave instrument

Table.1: *Information content, and its utility for structural biology, in native IM-MS data and experimental workflows.*

Box 1

Key computational challenges for native IM-MS

- An improved understanding of structural changes upon desolvation, their case-specific amplitude, and how these changes can be predicted based on the solution structure.
- Knowledge of how important net charge and charge-site configurations are for MD, along with an understanding of how to accommodate them by robust charge placement and explicit allowance of charge mobility.
- Development of force fields and associated methods for solvent-free MD, and integration of solvent-free MD with on-the-fly CCS calculation.
- Quantitative accommodation of biases and uncertainty that may arise in raw native MS data or in its analysis and interpretation, and appropriate cross-validation strategies.
- Supporting and influencing MS experimental development, in terms of instruments, methodologies, and rigour.

Box. 1: *Outstanding and key computational challenges for the field of native IM-MS to overcome.*

For table of contents only